Tensor similarity in two modes

Frederik Van Eeghem, Otto Debals, Lieven De Lathauwer, *Fellow, IEEE*

***Abstract*—Multi-way datasets are widespread in signal processing and play an important role in blind signal separation, array processing and biomedical signal processing, among others. One key strength of tensors is that their decompositions are unique under mild conditions, which allows the recovery of features or source signals. In several applications, such as classification, we wish to compare factor matrices of the decompositions. Though this is possible by first computing the tensor decompositions and subsequently comparing the factors, these decompositions are often computationally expensive. In this paper, we present a similarity method that indicates whether the factors in two modes are essentially equal without explicitly computing them. Essential equality conditions, which ensure the theoretical validity of our approach, are provided for various underlying tensor decompositions. The developed algorithm provides a computationally efficient way to compare factors. The method is illustrated in a context of emitter movement detection and fluorescence data analysis.**

***Index Terms*—tensor, similarity, classification, canonical polyadic decomposition, block term decomposition**

I. Introduction

The increasing digitalization of society and advances in data acquisition lead to a growing amount of multiway datasets. These datasets can be naturally stored in tensors, which are higher-order generalizations of vectors and matrices. Apart from offering natural data structures, higher-order tensors are useful tools for analysis as well. One favorable property is that their decompositions are unique under mild conditions, which contrasts with matrix decompositions that require additional constraints to attain uniqueness.

Initially driven by applications in psychometrics and chemometrics, tensors have made their entrance in signal processing and machine learning from the 90s onwards [1], [2]. A variety of applications in these domains have been tackled using tensors. In signal processing, tensors have proven to be extremely useful in blind signal separation problems, where the underlying sources are estimated from a set of mixtures. To solve these problems, additional assumptions have to be imposed on the sources. Examples include statistical independence [3], low-rank conditions [4], and many more [5], [6], [7]. Another example application in signal processing is emitter localization [8], [9]. In machine learning, tensors have been used in recommender systems and topic modeling (see [2] and references therein), face recognition [10], classification [11] and for learning latent variable models [12], to name a few.

Many of these applications rely on computing the factors of a tensor decomposition. However, in several applications, such as classification, we are interested in comparing the factors rather than in their explicit values. Because tensor decompositions are computationally expensive and can be ill-conditioned, strategies that are able to compare factors without explicitly computing them are of interest. In [13], a subspace-based method is derived that allows one to determine whether all factors of two tensors are equal without explicitly computing them. For third-order tensors for instance, this method can determine whether all three factor matrices are equal. However, in several applications the relevant information is contained in just two modes of a tensor. For instance, tensors used in emitter localization contain the location information in the first two modes and the (continuously changing) source information in the third mode [8], [4]. Consequently, we are only interested in the first two modes if we wish to verify whether the emitter locations have changed. Another example can be found in chemometrics, where third-order tensors contain excitation-emission information of chemical compounds [14], [15], [6]. If the goal is to check whether sets of mixtures consist of the same compounds, it suffices to compare the factors in the first two modes, as we will illustrate in Section VI.

In this paper, we develop a subspace-based method that is able to compare two third-order tensors in just two modes without imposing strong constraints on the third mode. More explicitly, the factor matrices in the first two modes are compared up to trivial indeterminacies such as column permutation and scaling. Note that this can be interpreted as the comparison of latent variables without explicitly computing them. The difference with the method from [13] lies in specialization. In [13], the underlying decomposition of two tensors is unknown. Several subspaces of the tensors are then compared to determine which decomposition they admit and whether the terms are the same. In certain applications, the underlying decomposition is known. Our method exploits this prior knowledge, which allows one to check just one subspace equality to determine whether the factors in two modes are equal. We also provide the theoretical foundations that ensure the subspace comparisons are equivalent to comparing the factor matrices.

A few related papers can be found in the literature. For

Frederik Van Eeghem is supported by an Aspirant Grant from the Research Foundation Flanders (FWO). Otto Debals is supported by an Aspirant Grant from the Institute of Science and Technology (IWT). This research is funded by (1) Research Council KU Leuven: C1 project c16/15/059-nD; (2) FWO: projects: G.0830.14N (Block term decompositions), G.0881.14N (Tensor based data similarity); (3) Belgian Federal Science Policy Office: IUAP P7/19 (DYSCO, Dynamical systems, control and optimization, 2012-2017); (4) EU: The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC Advanced Grant: BIOTENSORS (no. 339804). This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information.

Frederik Van Eeghem, Otto Debals and Lieven De Lathauwer are with both the Group of Science, Engineering and Technology, KU Leuven Kulak, E. Sabbelaan 53, B-8500 Kortrijk, Belgium and with the Department of Electrical Engineering (ESAT), KU Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium. (e-mail: frederik.vaneeghem@kuleuven.be, otto.debals@kuleuven.be, lieven.delathauwer@kuleuven.be)

instance, the results of [13] have been used to develop tensorial kernels in [16]. Another related paper is [17], where tensor canonical correlation analysis (TCCA) is presented.

As will be explained in Section III, the similarity method represents the comparison of two full third-order tensors as one vector of principal angles. This resulting vector can then easily be used in classification algorithms such as support vector machines, neural networks or decision trees. In this sense, our method can be interpreted as a dimensionality reduction strategy. This allows us to classify entire tensors based on the similarity of their underlying decompositions in two modes.

The remainder of this section introduces the notations used throughout this text. In Section II we introduce the relevant tensor decompositions with their properties. The core idea of the method is subsequently introduced in Section III, together with the algorithm and computational remarks. Sections IV and V then provide the theoretical details for full and partial similarity. The results are illustrated by a numerical experiment and two applications in Section VI.

Notations

Scalars are represented by italic lowercase letters (e.g., x), column vectors by bold lowercase letters (e.g., $\mathbf{x}$), matrices by bold uppercase letters (e.g., $\mathbf{X}$) and tensors by calligraphic letters (e.g., $\mathcal{X}$). Subscripts are used to denote subsets of vectors, matrices and tensors. For instance, a_n represents the nth entry of the vector $\mathbf{a}$ while $\mathbf{b}_n$ represents the nth column of the matrix $\mathbf{B}$ and t_{ijk} represents the entry with indices (i, j, k) of the tensor $\mathcal{T}$. The Kronecker product of two matrices $\mathbf{A} \in \mathbb{C}^{I\times J}$ and $\mathbf{B} \in \mathbb{C}^{K\times L}$ is denoted by $\mathbf{A} \otimes \mathbf{B} \in \mathbb{C}^{IK\times JL}$. The Khatri–Rao product of partitioned matrices $\mathbf{A} = [\mathbf{A}_1, \ldots, \mathbf{A}_N]$ and $\mathbf{B} = [\mathbf{B}_1, \ldots, \mathbf{B}_N]$ is given by

$$\mathbf{A} \odot \mathbf{B} = [\mathbf{A}_1 \otimes \mathbf{B}_1, \ldots, \mathbf{A}_N \otimes \mathbf{B}_N].$$

Unless explicitly stated otherwise, all partitionings in this paper will be column-wise (i.e., $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_N]$). The outer product is denoted by $\circ$. The transpose and the conjugate transpose are denoted by $\cdot^{\mathrm{T}}$ and $\cdot^{\mathrm{H}}$, respectively. The column space of a matrix $\mathbf{A}$ is denoted by $\operatorname{span}\{\mathbf{A}\}$. Column-wise vectorization of a matrix is denoted by $\operatorname{Vec}\{\mathbf{A}\}$.

II. Tensor decompositions

We briefly review three relevant tensor decompositions: the canonical polyadic decomposition (CPD), the block term decomposition (BTD) in multilinear rank-$(L, L, 1)$ terms, and the BTD in multilinear rank-$(L, L, \cdot)$ terms.

A. Canonical polyadic decomposition

A rank-1 tensor $\mathcal{A} \in \mathbb{C}^{I_1\times\cdots\times I_N}$ of order N is defined as the outer product of N nonzero vectors $\mathbf{u}^{(n)} \in \mathbb{C}^{I_n}$, i.e., $\mathcal{A} = \mathbf{u}^{(1)}\circ\cdots\circ\mathbf{u}^{(N)}$. The rank of a tensor $\mathcal{T}$ is then defined as the minimum number of rank-1 terms that yield $\mathcal{T}$ in a linear combination. Such a (minimal) linear combination of rank-1 terms is called a canonical polyadic decomposition (CPD).

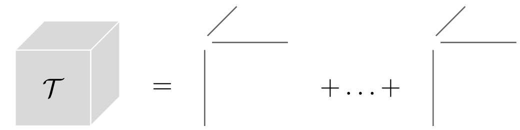


Figure 1: Graphical illustration of a CPD.

Mathematically, the CPD of a rank-R tensor $\mathcal{T} \in \mathbb{C}^{I_1\times\cdots\times I_N}$ can be written as

$$\mathcal{T} = \sum_{r=1}^{R} \lambda_r \mathbf{u}_r^{(1)} \circ \cdots \circ \mathbf{u}_r^{(N)}. \tag{1}$$

The factor vectors $\mathbf{u}_r^{(n)} \in \mathbb{C}^{I_n}$ are often concatenated in factor matrices $\mathbf{U}^{(n)} = \left[\mathbf{u}_1^{(n)}, \ldots, \mathbf{u}_R^{(n)}\right] \in \mathbb{C}^{I_n\times R}$. Following the notation of [1], (1) can be concisely written as

$$\mathcal{T} = \left[\!\left[\boldsymbol{\lambda}; \mathbf{U}^{(1)}, \ldots, \mathbf{U}^{(N)}\right]\!\right], \tag{2}$$

with $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_R] \in \mathbb{C}^R$. A graphical representation of the CPD is shown in Figure 1.

One of the key strengths of the CPD is its essential uniqueness under mild conditions. By essential uniqueness, we mean that the decomposition is unique up to two trivial indeterminacies. First, the different rank-1 terms can be arbitrarily permuted. Second, each vector within a rank-1 term can be scaled as long as the other vectors of the rank-1 term are counterscaled. Essential uniqueness conditions for CPDs have been extensively studied in literature, see e.g. [18], [19], [2]. Various algorithms to compute CPDs have been developed as well, see e.g. [2], [20] and references therein.

In this paper, we will mainly rely on uniqueness conditions of third-order tensors of which the third factor matrix has full column rank. Uniqueness conditions for this case are well-known and one set of conditions that we will use in this paper is given in Theorem 1. In this theorem, $\mathcal{C}_k(\mathbf{A})$ denotes the kth compound matrix of an $I\times R$ matrix $\mathbf{A}$, which is the $\binom{I}{k}\times\binom{R}{k}$ matrix containing the determinants of all $k\times k$ submatrices of $\mathbf{A}$ [21]. The determinants are arranged such that the submatrix index sets are in lexicographic order.

Theorem 1. *Consider the polyadic decomposition of* $\mathcal{X} \in \mathbb{C}^{I_1\times I_2\times I_3}$ *as defined in* (2)*. If*

$$\begin{cases} \mathbf{U}^{(3)} \in \mathbb{C}^{I_3\times R} \text{ has full column rank,} \\ \mathcal{C}_2\left(\mathbf{U}^{(1)}\right) \odot \mathcal{C}_2\left(\mathbf{U}^{(2)}\right) \text{ has full column rank,} \end{cases}$$

then the rank of $\mathcal{X}$ *is R and the CPD of* $\mathcal{X}$ *is essentially unique [22], [23], [21], [19].*

Let us consider the unfolding of a tensor to a matrix. A third order tensor $\mathcal{T} = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \in \mathbb{C}^{I\times J\times K}$ can be unfolded in several ways. We will use $\mathbf{T}_{IJ\times K}$ to denote the matrix of which the kth column is the column-wise vectorization of the kth frontal slice of $\mathcal{T}$. This matrix can be written in terms of the CPD factors as follows:

$$\mathbf{T}_{IJ\times K} = (\mathbf{B} \odot \mathbf{A})\mathbf{C}^{\mathrm{T}} \in \mathbb{C}^{IJ\times K}.$$

Similar expressions exist for higher-order tensors and for unfoldings to other modes [1].

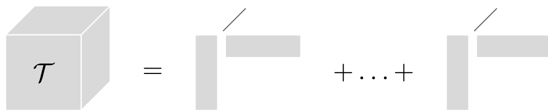


Figure 2: Graphical illustration of a BTD in multilinear rank-$(L, L, 1)$ terms.

B. Block term decomposition in rank-$(L, L, 1)$ terms

A block term decomposition (BTD) in multilinear rank-$(L, L, 1)$ terms is a generalization of a CPD that allows low-rank contributions in the first two modes of each term [24]. Mathematically, the BTD of a third-order tensor $\mathcal{T} \in \mathbb{C}^{I\times J\times K}$ in rank-$(L, L, 1)$ terms can be written as

$$\mathcal{T} = \sum_{r=1}^{R} \left(\mathbf{A}_r \mathbf{B}_r^{\mathrm{T}}\right) \circ \mathbf{c}_r,$$

in which $\mathbf{A}_r \in \mathbb{C}^{I\times L}$, $\mathbf{B}_r \in \mathbb{C}^{J\times L}$ and $\mathbf{c}_r \in \mathbb{C}^{K}$. Note that a connection can be made to the PARALIND-decomposition [25]. A graphical representation of a BTD in rank-$(L, L, 1)$ is shown in Figure 2. Essential uniqueness conditions for this decomposition are given in [24], [7], [26]. The terms can again be arbitrarily permuted, but the scaling indeterminacy is slightly more complex than before. Within each term, one can postmultiply $\mathbf{A}_r$ by any nonsingular $\mathbf{G}_r \in \mathbb{C}^{L\times L}$ as long as $\mathbf{B}_r$ is postmultiplied by $\left(\mathbf{G}_r^{-1}\right)^{\mathrm{T}}$. Moreover, the factors within each rank-$(L, L, 1)$ term can be scaled and counterscaled as long as their product remains the same. Algorithms for this type of decomposition can be found in [20], [27], [25].

The matrix unfolding $\mathbf{T}_{IJ\times K}$ is given by

$$\mathbf{T}_{IJ\times K} = \left[\left(\mathbf{B}_1 \odot \mathbf{A}_1\right) \mathbf{1}_L, \ldots, \left(\mathbf{B}_R \odot \mathbf{A}_R\right) \mathbf{1}_L\right] \cdot \mathbf{C}^{\mathrm{T}}.$$

C. Block term decomposition in rank-$(L, L, \cdot)$ terms

Mathematically, the BTD in multilinear rank-$(L, L, \cdot)$ terms of a tensor $\mathcal{T} \in \mathbb{C}^{I\times J\times K}$ is given by

$$\mathcal{T} = \sum_{r=1}^{R} \mathcal{D}_r \cdot_1 \mathbf{A}_r \cdot_2 \mathbf{B}_r,$$

in which $\mathcal{D}_r \in \mathbb{C}^{L\times L\times K}$ (with mode-1 rank equal to L and mode-2 rank equal to M) and in which $\mathbf{A}_r \in \mathbb{C}^{I\times L}$ and $\mathbf{B}_r \in \mathbb{C}^{J\times L}$ have full column rank. Note that the mode-n rank of a tensor is defined as the dimension of the vector space spanned by the mode-n vectors of the tensor. A graphical representation of this decomposition is given in Figure 3.

Again, essential uniqueness conditions can be found in [24]. The different terms can be arbitrarily permuted and the factor matrices $\mathbf{A}_r$ and $\mathbf{B}_r$ can be postmultiplied by nonsingular matrices $\mathbf{X}_r$ and $\mathbf{Y}_r$ respectively, as long as the core tensor $\mathcal{D}_r$ is replaced by $\mathcal{D}_r \cdot_1 \mathbf{X}_r^{-1} \cdot_2 \mathbf{Y}_r^{-1}$. Algorithms to compute this type of decomposition can for instance be found in [27], [28].

III. Problem statement and algorithm

A. Problem statement

As explained in the introduction, the comparison of tensors in the first two modes consists of verifying whether their factors in these modes are equal up to trivial indeterminacies. The

Figure 3: Graphical illustration of a BTD in multilinear rank-$(L, L, \cdot)$ terms.

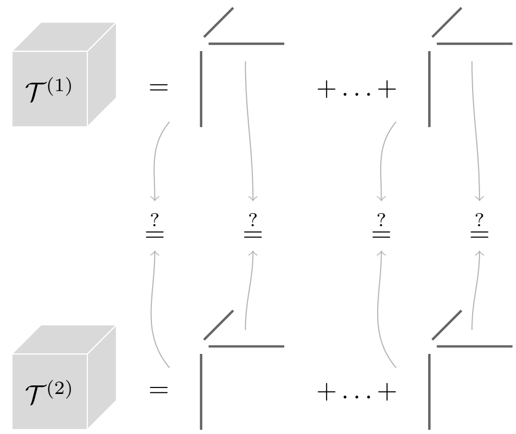


Figure 4: The goal of the paper is to verify whether the factor matrices of $\mathcal{T}^{(1)}$ and $\mathcal{T}^{(2)}$ in the first two modes are equal without explicitly computing the decompositions and using only linear algebra, illustrated here for the case in which both tensors admit a CPD.

factors will obviously depend on the type of decomposition the tensors admit. Let two tensors $\mathcal{T}^{(1)}$ and $\mathcal{T}^{(2)} \in \mathbb{C}^{I\times J\times K}$ both admit the same type of decomposition, which may be a CPD, BTD in rank-$(L, L, 1)$ terms or a BTD in rank-$(L, L, \cdot)$ terms. In general, we can compactly write the tensor decompositions as follows:

$$\mathcal{T}^{(1)} = \left[\!\left[\mathcal{G}^{(1)}; \mathbf{A}, \mathbf{B}, \mathbf{C}\right]\!\right],$$
$$\mathcal{T}^{(2)} = \left[\!\left[\mathcal{G}^{(2)}; \mathbf{D}, \mathbf{E}, \mathbf{F}\right]\!\right],$$

in which the structure of the core tensors $\mathcal{G}^{(1)}$ and $\mathcal{G}^{(2)}$ depends on the underlying decomposition [11]. The details of this compact representation, such as the structure of the core tensors, can be found in [11], [24] but are not that important for our purposes. The main goal in this paper is to compare the matrices $\mathbf{A}$ and $\mathbf{D}$, and $\mathbf{B}$ and $\mathbf{E}$ up to the trivial indeterminacies associated with the underlying decompositions. This is represented graphically for the CPD case in Figure 4. Throughout the rest of this text, we will denote equality of the factors in two modes up to trivial indeterminacies by essential equality in two modes. For tensors that do not admit unique decompositions, essential equality means that there exists a pair of decompositions of the tensors that are essentially equal.

B. Algorithm overview

To check essential equality in two modes without explicitly computing the factors, we turn to subspace comparisons. More specifically, the remainder of the text will show that under certain decomposition-dependent conditions, $\mathcal{T}^{(1)}$ and $\mathcal{T}^{(2)} \in \mathbb{C}^{I\times J\times K}$ are essentially equal in two modes if and only if

$$\operatorname{span}\left\{\mathbf{T}^{(1)}_{IJ\times K}\right\} = \operatorname{span}\left\{\mathbf{T}^{(2)}_{IJ\times K}\right\}.$$

This calls for a measure to compare subspaces. One way to characterize subspace similarity consists of using a set of principal angles. To draw conclusions from these angles, further processing is often needed. This can be done by a variety of popular machine learning algorithms, ranging from simple thresholding to more advanced support vector machines and clustering.

The decomposition-dependent conditions for the method play a fundamental role, as they ensure that the underlying decompositions are essentially unique. This ensures that the compared factors are unique and allows the use of subspaces to compare the underlying factor matrices. Further relaxation of the conditions to non-unique decompositions is briefly considered as well in Section IV-A.

A high-level overview of the procedure is given in Algorithm 1. The mathematical and computational details are provided in the subsequent sections. More specifically, the full mathematical proofs of the assumptions are given in Sections IV and V, and the computational details are discussed in Section III-C. This presentation was chosen to highlight the simplicity and accessibility of the presented algorithm without clouding the view with mathematics.

Algorithm 1: High-level overview of the procedure to check tensor similarity in two modes.

Data: tensors $\mathcal{T}^{(1)}$ and $\mathcal{T}^{(2)} \in \mathbb{C}^{I\times J\times K}$
Result: Decision whether $\mathbf{A} \stackrel{?}{=} \mathbf{D}$ and $\mathbf{B} \stackrel{?}{=} \mathbf{E}$
1) Check if all assumptions hold (in the generic case)
2) Unfold $\mathcal{T}^{(1)}$ and $\mathcal{T}^{(2)}$ to $\mathbf{T}^{(1)}_{IJ\times K}$ and $\mathbf{T}^{(2)}_{IJ\times K}$
3) Select an appropriate value N for the number of components
4) Find basis $\mathbf{U} \in \mathbb{C}^{IJ\times N}$ for dominant span of $\mathbf{T}^{(1)}_{IJ\times K}$
5) Find basis $\mathbf{V} \in \mathbb{C}^{IJ\times N}$ for dominant span of $\mathbf{T}^{(2)}_{IJ\times K}$
6) Compute N principal angles between $\mathbf{U}$ and $\mathbf{V}$
7) Classify the principal angles

C. Computational remarks

In this section, additional information is provided for the various steps in Algorithm 1. These remarks should clarify how the method can be used in practice.

1) Verifying conditions: Theorems 2 to 6 in the following sections specify which conditions must hold such that Algorithm 1 yields unambiguous results. If the factor matrices are known, the verification of the assumptions is straightforward. However, the goal of our similarity method is to avoid computing the factors explicitly. Here, we show that it is still possible to verify the conditions without explicitly computing the factors $\mathbf{A}, \mathbf{B}, \mathbf{D}$ and $\mathbf{E}$.

One of the recurring assumptions is a full-column-rank factor matrix. Take the example of a rank-R CPD of $\mathcal{T} \in \mathbb{C}^{I\times J\times K}$, given by $\mathcal{T} = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$. To check whether $\mathbf{C} \in \mathbb{C}^{K\times R}$ has full column rank, one can consider the unfolded tensor $\mathbf{T} = (\mathbf{B} \odot \mathbf{A})\,\mathbf{C}^{\mathrm{T}} \in \mathbb{C}^{IJ\times K}$. By simply checking whether the rank of $\mathbf{T}$ equals R, we know that $\mathbf{C}$ must have full column rank as long as $R \leq K$ and $R \leq IJ$.

Considering the same tensor $\mathcal{T}$ defined above, another recurring condition that has to be checked is whether $\mathcal{C}_2(\mathbf{A}) \odot \mathcal{C}_2(\mathbf{B})$ has full column rank. Note that this is a $\binom{I}{2}\binom{J}{2} \times \binom{R}{2}$ matrix. To check whether this matrix has full column rank without computing $\mathbf{A}$ and $\mathbf{B}$, we use a result from [29] which transforms a CPD-admitting tensor to a matrix $\mathbf{Q}_2(\mathcal{T}) \in \mathbb{C}^{\binom{I}{2}\binom{J}{2}\times L}$ that can be written as

$$\mathbf{Q}_2(\mathcal{T}) = [\mathcal{C}_2(\mathbf{A}) \odot \mathcal{C}_2(\mathbf{B})]\, \mathcal{Q}_2(\mathbf{C})^{\mathrm{T}}.$$

The details concerning the construction of $\mathbf{Q}_2(\mathcal{T})$ can be found in [29]. The important part for our purposes is that $\mathcal{C}_2(\mathbf{A}) \odot \mathcal{C}_2(\mathbf{B})$ is one of the factors of the matrix $\mathbf{Q}_2(\mathcal{T})$. By checking whether the rank of $\mathbf{Q}_2(\mathcal{T})$ equals $\binom{R}{2}$, we can determine whether $\mathcal{C}_2(\mathbf{A}) \odot \mathcal{C}_2(\mathbf{B})$ has full column rank, provided that $\binom{R}{2} \leq \binom{I}{2}\binom{J}{2}$ and $\binom{R}{2} \leq L$.

The previous approach checked the conditions deterministically. Alternatively, one can turn to generic conditions. For instance, a generic matrix is known to have full column rank if it has at least as many rows as columns. For the rank-R tensor $\mathcal{T} = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \in \mathbb{C}^{I\times J\times K}$, this means that $\mathbf{C}$ generically has full column rank if $R \leq K$. Similarly, in [19] it is stated that the matrix $\mathcal{C}_2(\mathbf{A}) \odot \mathcal{C}_2(\mathbf{B})$ has full column rank in the generic case when it has at least as many rows as columns. Generic uniqueness conditions for essential uniqueness of a BTD in rank-$(L, L, 1)$ terms and a BTD in rank-$(L, L, \cdot)$ terms can be found in [24], [19].

2) Selecting an appropriate value for the number of components: Once the tensors have been reshaped to matrices $\mathbf{T}^{(1)}_{IJ\times K}$ and $\mathbf{T}^{(2)}_{IJ\times K}$, the next step is to determine the number of (dominant) components N. Note that for underlying CPDs or BTDs in rank-$(L, L, 1)$ terms, N equals the number of terms R. However, this is not the case for a BTD in R rank-$(L, L, \cdot)$ terms, where the number of important components is given by $N = RL^2$. An appropriate value for N can either follow directly from prior knowledge of the application or can be estimated using linear algebra. The latter is done by checking the number of dominant singular values of $\mathbf{T}_{IJ\times K}$. Apart from some special cases, such as the case where the number of components N is bigger than IJ or K, this approach gives a reasonable estimate for N.

If the number of components is incorrectly estimated, the principal angles will be affected. More specifically, if the estimated number of components N exceeds the actual number of components M, there will generally be M zero-valued principal angles and $N - M$ nonzero principal angles since the extra information only contains noise. If the estimated number of components N is lower than the actual number of components M, the situation is slightly more involved. Assume without loss of generality that $N = M - 1$. This implies that we only get a partial view of the full desired subspaces. Even if both partial subspaces stem from the same subspace, there will be nonzero principal angles because the partial subspaces do not contain exactly the same information.

For example, if $N = 3$ and $M = 4$, there will be two zero-valued principal angles and 1 nonzero-valued principal angle. We will illustrate this point numerically in Section VI.

3) Constructing a basis for a dominant subspace: Given a subspace such as span $\left\{\mathbf{T}^{(1)}_{IJ\times K}\right\}$, the goal is to construct an orthonormal basis for its N-dimensional dominant subspace. To do this, compute the left singular vectors of $\mathbf{T}^{(1)}_{IJ\times K}$ corresponding to the N dominant singular values. The dominant subspace can also be obtained by updating a previous subspace using tracking techniques (e.g., [30]). Once the basis vectors are found, store them as columns in a matrix $\mathbf{U} \in \mathbb{C}^{IJ\times N}$. Note that apart from the method based on singular value decompositions given here, there are many ways to find a basis for a subspace (see e.g., [31]).

4) Principal angles: One way to characterize subspace similarity consists of using principal angles. When two matrices $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{I\times N}$ are compared, the principal angles $\{\theta_1, \ldots, \theta_N\} \in [0, \pi/2]$ form a set of minimized angles between both subspaces. Mathematically, the principal angles can be defined recursively as (e.g., [32]):

$$\theta_i \overset{\text{def}}{=} \min\left\{\arccos\left(\frac{|\mathbf{a}^{\mathrm{T}}\mathbf{b}|}{||\mathbf{a}||\,||\mathbf{b}||}\right) \,\middle|\, \mathbf{a} \in \text{span}\{\mathbf{A}\}, \ldots \right.$$
$$\left. \mathbf{b} \in \text{span}\{\mathbf{B}\}, \mathbf{a} \perp \mathbf{a}_j, \mathbf{b} \perp \mathbf{b}_j \quad \forall j \in \{1, \ldots, i-1\}\right\},$$

in which $\mathbf{a}_j$ and $\mathbf{b}_j$ are the vectors corresponding to θ_j. Note that all principal angles are (approximately) zero if the subspaces are (approximately) equal. The definition also implies that if span $\{\mathbf{A}\} \cap$ span $\{\mathbf{B}\}$ is Q-dimensional, there are exactly Q zero-valued principal angles. Computation of principal angles can for instance be done using the singular value decomposition, see the algorithms in [33]. Note that the principal angles can be used in noisy conditions as well. The effect of noise on the principal angles will be illustrated numerically in Section VI. By simply setting a threshold, similar tensors can still be identified by checking whether the principal angles are sufficiently small. The appropriate threshold depends on both the application and the noise level. Note that more advanced machine learning techniques than thresholding can be used as well. For instance, the principal angles can be fed into a neural network, support vector machine or decision tree to determine whether the factors are sufficiently similar.

IV. Full equality

Full essential equality in two modes entails that we check whether the full factor matrices in two modes are essentially equal. Equality conditions are provided for various underlying decompositions of the tensors. More specifically, theorems are given for underlying CPDs, BTDs in rank-$(L, L, 1)$ terms and BTDs in rank-$(L, L, \cdot)$ terms.

A. CPD

Let two tensors $\mathcal{T}^{(1)}, \mathcal{T}^{(2)} \in \mathbb{C}^{I\times J\times K}$ admit a rank-R CPD:

$$\begin{aligned}\mathcal{T}^{(1)} &= [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!], \\ \mathcal{T}^{(2)} &= [\![\mathbf{D}, \mathbf{E}, \mathbf{F}]\!].\end{aligned} \tag{3}$$

We now wish to assess when $\mathbf{A}$ and $\mathbf{D} \in \mathbb{C}^{I\times R}$, and $\mathbf{B}$ and $\mathbf{E} \in \mathbb{C}^{J\times R}$ are equal up to column scaling and permutation. Mathematically, we want to verify whether there exists an $s \in \{1, \ldots, R\}$ for each $r \in \{1, \ldots, R\}$ such that

$$\mathbf{a}_r\mathbf{b}_r^{\mathrm{T}} = \alpha_s \mathbf{d}_s \mathbf{e}_s^{\mathrm{T}},$$

in which α_s captures the possible scaling differences. Essential equality conditions are formulated in the following theorem:

Theorem 2. *Let two tensors $\mathcal{T}^{(1)}, \mathcal{T}^{(2)}$ admit a rank-R CPD as in* (3). *If*

$$\begin{aligned} &(i) && \mathbf{C} \textit{ and } \mathbf{F} \textit{ have full column rank}, \\ &(ii) && \mathcal{C}_2(\mathbf{A}) \odot \mathcal{C}_2(\mathbf{B}) \textit{ has full column rank}, \end{aligned} \tag{4a}$$

then $\mathcal{T}^{(1)}$ and $\mathcal{T}^{(2)}$ are essentially equal in the first two modes if and only if span $\left\{\mathbf{T}^{(1)}_{IJ\times K}\right\}$ = span $\left\{\mathbf{T}^{(2)}_{IJ\times K}\right\}$.

Proof. Because both $\mathcal{T}^{(1)}$ and $\mathcal{T}^{(2)}$ admit a CPD, their unfoldings can be written as

$$\begin{aligned}\mathbf{T}^{(1)}_{IJ\times K} &= (\mathbf{B} \odot \mathbf{A})\,\mathbf{C}^{\mathrm{T}}, \\ \mathbf{T}^{(2)}_{IJ\times K} &= (\mathbf{E} \odot \mathbf{D})\,\mathbf{F}^{\mathrm{T}},\end{aligned}$$

as explained in Section II-A.

We first show that span $\left\{\mathbf{T}^{(1)}_{IJ\times K}\right\}$ = span $\left\{\mathbf{T}^{(2)}_{IJ\times K}\right\}$ is a necessary condition. Assume $\mathcal{T}^{(1)}$ and $\mathcal{T}^{(2)}$ are essentially equal in the first two modes, i.e., that $\mathbf{D} = \mathbf{A}\mathbf{\Lambda}^{(1)}\mathbf{\Pi}$ and $\mathbf{E} = \mathbf{B}\mathbf{\Lambda}^{(2)}\mathbf{\Pi}$ with $\mathbf{\Lambda}^{(1)}, \mathbf{\Lambda}^{(2)}$ diagonal scaling matrices and $\mathbf{\Pi}$ a permutation matrix. The matrix unfoldings of $\mathcal{T}^{(1)}$ and $\mathcal{T}^{(2)}$ then become

$$\begin{aligned}\mathbf{T}^{(1)}_{IJ\times K} &= (\mathbf{B} \odot \mathbf{A})\,\mathbf{C}^{\mathrm{T}}, \\ \mathbf{T}^{(2)}_{IJ\times K} &= (\mathbf{B} \odot \mathbf{A})\,\mathbf{\Lambda}^{(1)}\mathbf{\Lambda}^{(2)}\mathbf{\Pi}\mathbf{F}^{\mathrm{T}}.\end{aligned}$$

Because $\mathbf{C}$ and $\mathbf{F}$ are assumed to have full column rank and $\mathbf{\Lambda}^{(1)}, \mathbf{\Lambda}^{(2)}, \mathbf{\Pi}$ are nonsingular, it immediately follows that span $\left\{\mathbf{T}^{(1)}_{IJ\times K}\right\}$ = span $\left\{\mathbf{T}^{(2)}_{IJ\times K}\right\}$.

Conversely, assume span $\left\{\mathbf{T}^{(1)}_{IJ\times K}\right\}$ = span $\left\{\mathbf{T}^{(2)}_{IJ\times K}\right\}$ holds to show that this subspace equality is also sufficient. Because $\mathbf{C}$ and $\mathbf{F}$ have full column rank, the subspace equality can be written as

$$(\mathbf{B} \odot \mathbf{A})\left(\mathbf{M}^{(1)}\right)^{\mathrm{T}} = (\mathbf{E} \odot \mathbf{D})\left(\mathbf{M}^{(2)}\right)^{\mathrm{T}}, \tag{5}$$

with $\mathbf{M}^{(1)}, \mathbf{M}^{(2)} \in \mathbb{C}^{R\times R}$ nonsingular matrices. Both sides of (5) represent an unfolded CPD of the same tensor. Consequently, we can write

$$\mathcal{G} = \left[\!\!\left[\mathbf{A}, \mathbf{B}, \mathbf{M}^{(1)}\right]\!\!\right] = \left[\!\!\left[\mathbf{D}, \mathbf{E}, \mathbf{M}^{(2)}\right]\!\!\right] \in \mathbb{C}^{I\times J\times R}. \tag{6}$$

It follows from Theorem 1 that the CPD in equation (6) is essentially unique since $\mathbf{M}^{(1)}$ is nonsingular and $\mathcal{C}_2(\mathbf{A}) \odot \mathcal{C}_2(\mathbf{B})$ has full column rank as assumed in (4a). Consequently, the factor matrices $\mathbf{A}$ and $\mathbf{D}$, and $\mathbf{B}$ and $\mathbf{E}$ are equal up to the trivial indeterminacies stated in Section II-A. ☐

Most results in this paper concern the comparison of factor matrices of tensors with essentially unique decompositions. Results for the non-unique case can be obtained as well. The following theorem shows that uniqueness of the CPDs of $\mathcal{T}^{(1)}$ and $\mathcal{T}^{(2)}$ is not required to compare their factors in the first two modes.

Theorem 3. *Let two tensors $\mathcal{T}^{(1)}, \mathcal{T}^{(2)}$ admit a (possibly non-unique) rank-R CPD as in* (3) *and let us construct the stacked tensor $\mathcal{T}^{(\text{stack})} \in \mathbb{C}^{I\times J\times 2K}$ by concatenating $\mathcal{T}^{(1)}$ and $\mathcal{T}^{(2)}$ along the third mode. If $\mathcal{T}^{(\text{stack})}$ admits a rank-R CPD, then $\mathcal{T}^{(1)}$ and $\mathcal{T}^{(2)}$ are essentially equal in the first two modes.*

Proof. If $\mathcal{T}^{(\text{stack})} \in \mathbb{C}^{I\times J\times 2K}$ admits a rank-R CPD, it can be written as

$$\mathcal{T}^{(\text{stack})} = \left[\!\!\left[\mathbf{U}, \mathbf{V}, \begin{bmatrix}\mathbf{W}^{(1)}\\ \mathbf{W}^{(2)}\end{bmatrix}\right]\!\!\right],$$

with $\mathbf{U} \in \mathbb{C}^{I\times R}$, $\mathbf{V} \in \mathbb{C}^{J\times R}$ and $\mathbf{W}^{(1)}, \mathbf{W}^{(2)} \in \mathbb{C}^{K\times R}$. Since $\mathcal{T}^{(\text{stack})}$ is constructed by concatenating $\mathcal{T}^{(1)}$ and $\mathcal{T}^{(2)}$ along the third mode, it immediately follows that

$$\mathcal{T}^{(1)} = \left[\!\!\left[\mathbf{U}, \mathbf{V}, \mathbf{W}^{(1)}\right]\!\!\right], \quad \mathcal{T}^{(2)} = \left[\!\!\left[\mathbf{U}, \mathbf{V}, \mathbf{W}^{(2)}\right]\!\!\right].$$

These expressions show that there exist rank-R decompositions of $\mathcal{T}^{(1)}$ and $\mathcal{T}^{(2)}$ with essentially equal factor matrices in the first two modes. □

B. BTD in rank-$(L, L, 1)$ terms

Let two tensors $\mathcal{T}^{(1)}, \mathcal{T}^{(2)} \in \mathbb{C}^{I\times J\times K}$ admit a BTD in rank-$(L, L, 1)$ terms:

$$\mathcal{T}^{(1)} = \sum_{r=1}^{R} \left(\mathbf{A}_r\mathbf{B}_r^{\mathrm{T}}\right) \circ \mathbf{c}_r, \tag{7}$$

$$\mathcal{T}^{(2)} = \sum_{r=1}^{R} \left(\mathbf{D}_r\mathbf{E}_r^{\mathrm{T}}\right) \circ \mathbf{f}_r, \tag{8}$$

with $\mathbf{A}_r, \mathbf{D}_r \in \mathbb{C}^{I\times L}$ and $\mathbf{B}_r, \mathbf{E}_r \in \mathbb{C}^{J\times L}$.

We now wish to assess when $\mathbf{A}_r$ and $\mathbf{D}_r$, and $\mathbf{B}_r$ and $\mathbf{E}_r$ are equal for all values of r up to the trivial indeterminacies associated with the BTD in rank-$(L, L, 1)$ terms. We again denote this by essential equality in the first two modes. Mathematically, we want to verify whether there exists an $s \in \{1, \ldots, R\}$ for each $r \in \{1, \ldots, R\}$ such that

$$\mathbf{A}_r\mathbf{B}_r^{\mathrm{T}} = \alpha_s\mathbf{D}_s\mathbf{E}_s^{\mathrm{T}},$$

in which α_s captures the possible scaling differences. Equality conditions for this case are formulated in the following theorem:

Theorem 4. *Consider two tensors $\mathcal{T}^{(1)}, \mathcal{T}^{(2)}$ admitting a BTD in rank-$(L, L, 1)$ terms as defined in* (7) *and* (8). *If following conditions hold*

(i) $\mathbf{C}$ *and* $\mathbf{F}$ *have full column rank,*

(ii) *The BTD* $\mathcal{G} = \sum_{r=1}^{R} \left(\mathbf{A}_r\mathbf{B}_r^{T}\right) \circ \mathbf{m}_r^{(1)}$ *from* (23) *is essentially unique,* (9a)

then $\mathcal{T}^{(1)}$ and $\mathcal{T}^{(2)}$ are essentially equal in the first two modes if and only if span $\left\{\mathbf{T}^{(1)}_{IJ\times K}\right\} =$ *span* $\left\{\mathbf{T}^{(2)}_{IJ\times K}\right\}$.

Proof. The proof is conceptually similar to the proof of Theorem 2 and is given in Appendix A. □

C. BTD in rank-$(L, L, \cdot)$ terms

Let two tensors $\mathcal{T}^{(1)}, \mathcal{T}^{(2)} \in \mathbb{K}^{I\times J\times K}$ admit a BTD in rank-$(L, L, \cdot)$ terms:

$$\mathcal{T}^{(1)} = \sum_{r=1}^{R} \mathcal{S}_r^{(1)} \cdot_1 \mathbf{A}_r \cdot_2 \mathbf{B}_r, \tag{10}$$

$$\mathcal{T}^{(2)} = \sum_{r=1}^{R} \mathcal{S}_r^{(2)} \cdot_1 \mathbf{C}_r \cdot_2 \mathbf{D}_r, \tag{11}$$

with $\mathbf{A}_r, \mathbf{D}_r \in \mathbb{K}^{I\times L}$ and $\mathbf{B}_r, \mathbf{E}_r \in \mathbb{K}^{J\times L}$. These tensors can be unfolded as

$$\mathbf{T}^{(1)}_{IJ\times K} = [\mathbf{B}_1 \otimes \mathbf{A}_1, \ldots, \mathbf{B}_R \otimes \mathbf{A}_R] \cdot \mathbf{S}^{(1)},$$
$$\mathbf{T}^{(2)}_{IJ\times K} = [\mathbf{D}_1 \otimes \mathbf{C}_1, \ldots, \mathbf{D}_R \otimes \mathbf{C}_R] \cdot \mathbf{S}^{(2)},$$

with $\mathbf{S}^{(1)}, \mathbf{S}^{(2)} \in \mathbb{K}^{RL^2\times K}$ matrices consisting of the matricized versions of $\mathcal{S}_r^{(1)}$ and $\mathcal{S}_r^{(2)}$, respectively. The main question remains: when are $\mathbf{A}$ and $\mathbf{C}$, and $\mathbf{B}$ and $\mathbf{D}$ equal up to the trivial indeterminacies of a rank-$(L, L, \cdot)$ BTD. Conditions for this essential equality problem in two modes are given in

Theorem 5. *Consider two tensors $\mathcal{T}^{(1)}, \mathcal{T}^{(2)}$ admitting a BTD in rank-$(L, L, \cdot)$ terms as defined in* (10) *and* (11). *If following conditions hold*

(i) $\mathbf{S}^{(1)}$ *and* $\mathbf{S}^{(2)}$ *have full row rank,*

(ii) *The BTD* $\mathcal{G} = \sum_{r=1}^{R} \mathcal{M}_r^{(1)} \cdot_1 \mathbf{A}_r \cdot_2 \mathbf{B}_r$ *from* (25) *is essentially unique,* (12a)

then $\mathcal{T}^{(1)}$ and $\mathcal{T}^{(2)}$ are essentially equal in the first two modes if and only if span $\left\{\mathbf{T}^{(1)}_{IJ\times K}\right\} =$ *span* $\left\{\mathbf{T}^{(2)}_{IJ\times K}\right\}$.

Proof. The proof is conceptually similar to the proof of Theorem 2 and is given in Appendix B. □

V. Partial equality: the CPD case

In the previous section, full essential equality in two modes was considered. Here, we discuss the case in which only a subset of the terms of two CPDs are essentially equal in two modes. In terms of the factors, this is the case where a subset of the factor vectors are essentially equal.

A. Verifying partial similarity

Let two tensors $\mathcal{T}^{(1)}, \mathcal{T}^{(2)} \in \mathbb{C}^{I\times J\times K}$ admit a rank-R CPD:

$$\mathcal{T}^{(1)} = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!], \quad \mathcal{T}^{(2)} = [\![\mathbf{D}, \mathbf{E}, \mathbf{F}]\!], \tag{13}$$

with $\mathbf{A}, \mathbf{D} \in C^{I\times R}$, $\mathbf{B}, \mathbf{E} \in C^{J\times R}$ and $\mathbf{C}, \mathbf{F} \in C^{K\times R}$. The following theorem shows that under certain conditions, the number of zero-valued principal angles equals the number of terms of $\mathcal{T}^{(1)}$ and $\mathcal{T}^{(2)}$ that are essentially equal in two modes.

Theorem 6. *Consider two tensors* $\mathcal{T}^{(1)}, \mathcal{T}^{(2)} \in \mathbb{C}^{I \times J \times K}$ *admitting a rank-*R *CPD as defined in* (13)*. If*

$$(i) \quad \mathbf{C} \text{ and } \mathbf{F} \text{ have full column rank } R, \tag{14a}$$

$$(ii) \quad \mathcal{C}_2(\mathbf{B}) \odot \mathcal{C}_2(\mathbf{A}) \text{ has full column rank}, \tag{14b}$$

$$(iii) \quad \mathcal{C}_2(\mathbf{E}) \odot \mathcal{C}_2(\mathbf{D}) \text{ has full column rank}, \tag{14c}$$

$$(iv) \quad \text{One basis for span}\left\{\mathbf{T}^{(1)}_{IJ\times K}\right\} \cap \text{span}\left\{\mathbf{T}^{(2)}_{IJ\times K}\right\} \text{ admits a unique (unfolded) rank-}Q\text{ CPD}, \tag{14d}$$

then there are Q *essentially equal terms in two modes if and only if* Q *principal angles between the unfolded tensors* $\mathbf{T}^{(1)}_{IJ\times K}$ *and* $\mathbf{T}^{(2)}_{IJ\times K}$ *are zero, with* $Q \leq R$*.*

Proof. Consider the unfolded matrices $\mathbf{T}^{(1)}_{IJ\times K}$ and $\mathbf{T}^{(2)}_{IJ\times K} \in \mathbb{C}^{IJ\times K}$ of the tensors $\mathcal{T}^{(1)}$ and $\mathcal{T}^{(2)}$, which can be written as

$$\mathbf{T}^{(1)}_{IJ\times K} = (\mathbf{B} \odot \mathbf{A})\,\mathbf{C}^{\mathrm{T}}$$
$$\mathbf{T}^{(2)}_{IJ\times K} = (\mathbf{E} \odot \mathbf{D})\,\mathbf{F}^{\mathrm{T}}.$$

Since both $\mathbf{C}$ and $\mathbf{F}$ are assumed to have full column rank in (14a), it follows that $(\mathbf{B} \odot \mathbf{A})$ and $(\mathbf{E} \odot \mathbf{D})$ form bases for the column spaces of $\mathbf{T}^{(1)}_{IJ\times K}$ and $\mathbf{T}^{(2)}_{IJ\times K}$, respectively.

$\Leftarrow$) Assume there are Q zero-valued principal angles. This implies that the common subspace $\text{span}\left\{\mathbf{T}^{(1)}_{IJ\times K}\right\} \cap \text{span}\left\{\mathbf{T}^{(2)}_{IJ\times K}\right\}$ is exactly Q-dimensional. Let the columns of $\mathbf{G} \in \mathbb{C}^{IJ\times Q}$ contain a basis for this subspace. We then have

$$\mathbf{G} = (\mathbf{B} \odot \mathbf{A})\,\mathbf{M}^{(1)} = (\mathbf{E} \odot \mathbf{D})\,\mathbf{M}^{(2)}, \tag{15}$$

in which $\mathbf{M}^{(1)}, \mathbf{M}^{(2)} \in \mathbb{C}^{R\times Q}$ have full column rank. Because we assume in (14d) that $\mathbf{G}$ admits an unfolded rank-Q CPD, we can write

$$\mathbf{G} = (\mathbf{V} \odot \mathbf{U})\,\mathbf{W}^{\mathrm{T}}, \tag{16}$$

with $\mathbf{U} \in \mathbb{C}^{I\times Q}$, $\mathbf{V} \in \mathbb{C}^{J\times Q}$ and $\mathbf{W} \in \mathbb{C}^{Q\times Q}$. Because $\mathbf{W}$ has full rank, we can use (15) and (16) to write

$$(\mathbf{V} \odot \mathbf{U}) = (\mathbf{B} \odot \mathbf{A})\,\mathbf{N}^{(1)} = (\mathbf{E} \odot \mathbf{D})\,\mathbf{N}^{(2)},$$

with $\mathbf{N}^{(1)} = \mathbf{M}^{(1)}\left(\mathbf{W}^{\mathrm{T}}\right)^{-1} \in \mathbb{C}^{R\times Q}$ and $\mathbf{N}^{(2)} = \mathbf{M}^{(2)}\left(\mathbf{W}^{\mathrm{T}}\right)^{-1} \in \mathbb{C}^{R\times Q}$. The qth column of $(\mathbf{V} \odot \mathbf{U})$ is given by

$$\mathbf{v}_q \otimes \mathbf{u}_q = \sum_{r=1}^{R} n^{(1)}_{rq}\,(\mathbf{b}_r \otimes \mathbf{a}_r) = \sum_{r=1}^{R} n^{(2)}_{rq}\,(\mathbf{e}_r \otimes \mathbf{d}_r)\,.$$

A column-wise reshape of these equations to $I \times J$ matrices then gives

$$\mathbf{u}_q \circ \mathbf{v}_q = \sum_{r=1}^{R} n^{(1)}_{rq}\,(\mathbf{a}_r \circ \mathbf{b}_r) = \sum_{r=1}^{R} n^{(2)}_{rq}\,(\mathbf{d}_r \circ \mathbf{e}_r)\,. \tag{17}$$

Conditions (14b) and (14c) are sufficient for the CPD uniqueness of $\mathcal{T}^{(1)}$ and $\mathcal{T}^{(2)}$ when (14a) holds. If these conditions are satisfied, then following necessary and sufficient conditions hold as well [23]:

$$(i) \quad \mathbf{C} \text{ and } \mathbf{F} \text{ have full column rank}, \tag{18a}$$

$(\cdot)$ For every $\mathbf{w}$ that has at least two nonzero entries we have

$$(ii) \quad \text{rank}\left(\mathbf{K}\left(\mathbf{w}\right)\right) > 1 \text{ and} \tag{18b}$$

$$(iii) \quad \text{rank}\left(\mathbf{L}\left(\mathbf{w}\right)\right) > 1, \tag{18c}$$

in which (18a) is the same condition as (14a) and the functions $\mathbf{K}(\mathbf{w})$ and $\mathbf{L}(\mathbf{w})$ are defined as $\mathbf{K}(\mathbf{w}) = \sum_{r=1}^{R} w_r\,(\mathbf{a}_r \circ \mathbf{b}_r)$ and $\mathbf{L}(\mathbf{w}) = \sum_{r=1}^{R} w_r\,(\mathbf{d}_r \circ \mathbf{e}_r)$. It immediately follows from conditions (18b) and (18c) that equation (17) can never be satisfied if the linear combinations are nontrivial, i.e., if there is more than one term in the linear combination. Consequently, for each $q \in \{1, \ldots, Q\}$ we have

$$\mathbf{v}_q \otimes \mathbf{u}_q = n^{(1)}_{rq}\,(\mathbf{b}_r \otimes \mathbf{a}_r) = n^{(2)}_{sq}\,(\mathbf{e}_s \otimes \mathbf{d}_s)\,,$$

for certain $r, s \in \{1, \ldots, R\}$. We thus have Q columns of $(\mathbf{B} \odot \mathbf{A})$ and $(\mathbf{E} \odot \mathbf{D})$ that are equal up to scaling, which implies that there are Q essentially equal terms in the first two modes.

$\Rightarrow$) Conversely, assume that exactly Q terms are essentially equal in the first two modes. In this case, the bases can be written as

$$\mathbf{B} \odot \mathbf{A} = [\mathbf{B}_{\text{eq}} \odot \mathbf{A}_{\text{eq}},\, \mathbf{B}_{\text{diff}} \odot \mathbf{A}_{\text{diff}}] \tag{19}$$

$$\mathbf{E} \odot \mathbf{D} = [(\mathbf{B}_{\text{eq}} \odot \mathbf{A}_{\text{eq}})\,\mathbf{\Lambda}\mathbf{\Pi},\, \mathbf{E}_{\text{diff}} \odot \mathbf{D}_{\text{diff}}]\,, \tag{20}$$

in which the equal parts of the factor matrices are denoted by $\mathbf{A}_{\text{eq}} \in \mathbb{C}^{I\times Q}$ and $\mathbf{B}_{\text{eq}} \in \mathbb{C}^{J\times Q}$. The $Q \times Q$ matrices $\mathbf{\Lambda}$ and $\mathbf{\Pi}$ are used to express possible column scaling and permutation, respectively. Because the principal angles are defined as a set of minimized angles between subspaces, it follows immediately from (19) and (20) that there will be at least Q zero-valued principal angles between the subspace of $\mathbf{T}^{(1)}_{IJ\times K}$ and $\mathbf{T}^{(2)}_{IJ\times K}$. We now prove by contradiction that there are no more than Q zero-valued principal angles. Assume there are $Q + S$ zero-valued principal angles, with $S > 0$. This implies that the null space of $[\mathbf{B} \odot \mathbf{A}, \mathbf{E} \odot \mathbf{D}]$ is $(Q + S)$-dimensional. Note that this null space equals $\text{span}\left\{\mathbf{T}^{(1)}_{IJ\times K}\right\} \cap \text{span}\left\{\mathbf{T}^{(2)}_{IJ\times K}\right\}$. Because assumption (14d) states that the null space admits an unfolded rank-Q CPD, this situation cannot occur. We can conclude that under the assumptions (14a) to (14d) there will be exactly Q zero-valued principal angles when Q terms are essentially equal in the first two modes. □

Note that in the conditions and proof of Theorem 6, verifying whether a matrix space admits a unique CPD without explicitly computing the factors can be done using the techniques mentioned in III-C1.

B. Computing the common terms

When two tensors are partially equal in two modes, the common terms can be extracted without having to compute the other terms. This approach is especially advantageous when there are just a few common terms. The procedure to compute these common terms is given below.

First, compute the R-dimensional dominant column spaces of $\mathbf{T}^{(1)}_{IJ\times K}$ and $\mathbf{T}^{(2)}_{IJ\times K}$ and store them in the matrices $\mathbf{T}^{(1)}_{\text{dom}}$ and $\mathbf{T}^{(2)}_{\text{dom}}$. These matrices form a basis for $(\mathbf{B}\odot\mathbf{A})$ and $(\mathbf{E}\odot\mathbf{D})$, respectively. Next, construct the matrix $\mathbf{Z} = \left[\mathbf{T}^{(1)}_{\text{dom}}, \mathbf{T}^{(2)}_{\text{dom}}\right] \in \mathbb{K}^{IJ\times 2R}$. Assuming without loss of generality that there are U common terms in the first two modes, let the columns of $\mathbf{N}\in\mathbb{C}^{2R\times U}$ form a basis for the null space of $\mathbf{Z}$. This matrix $\mathbf{N}$ can then be partitioned as

$$\mathbf{N} = \begin{bmatrix}\mathbf{R}\\ \mathbf{S}\end{bmatrix} \in \mathbb{K}^{2R\times U},$$

in which $\mathbf{R},\mathbf{S}\in\mathbb{K}^{R\times U}$. Now construct the matrix $\mathbf{W}$ as

$$\mathbf{W} = \mathbf{T}^{(1)}_{\text{dom}}\mathbf{R} = -\mathbf{T}^{(2)}_{\text{dom}}\mathbf{S}.$$

This matrix contains a basis for the subspace spanned by the common columns of $(\mathbf{B}\odot\mathbf{A})$ and $(\mathbf{E}\odot\mathbf{D})$. Consequently, it holds that

$$\mathbf{W} = (\mathbf{B}_{\text{comm}}\odot\mathbf{A}_{\text{comm}})\,\mathbf{M}^{\text{T}}, \tag{21}$$

in which $\mathbf{M}\in\mathbb{C}^{U\times U}$ is a nonsingular matrix, and $\mathbf{A}_{\text{comm}}\in\mathbb{C}^{I\times U}$ and $\mathbf{B}_{\text{comm}}\in\mathbb{C}^{J\times U}$ denote the common subsets of $\mathbf{A}$ and $\mathbf{D}$, and $\mathbf{B}$ and $\mathbf{E}$, respectively. Computation of the (unfolded) CPD of $\mathbf{W}$ in (21) then yields the columns of $\mathbf{A}$ and $\mathbf{B}$ of the common terms of $\mathcal{T}^{(1)}$ and $\mathcal{T}^{(2)}$.

C. Computing the distinct terms

The previous section considered how the common terms can be extracted from the subspaces without computing the full decompositions. However, difficulties arise when trying to extract terms that are distinct. A full treatment of ways to compute these distinct terms is outside the scope of the paper, but we briefly mention one approach that improves upon simply computing both CPDs. Start by extracting the common terms as described in the previous section. Next, use this result as partial initialization for the computation of the full CPDs. This may result in faster convergence for optimization-based methods.

VI. Numerical experiments and applications

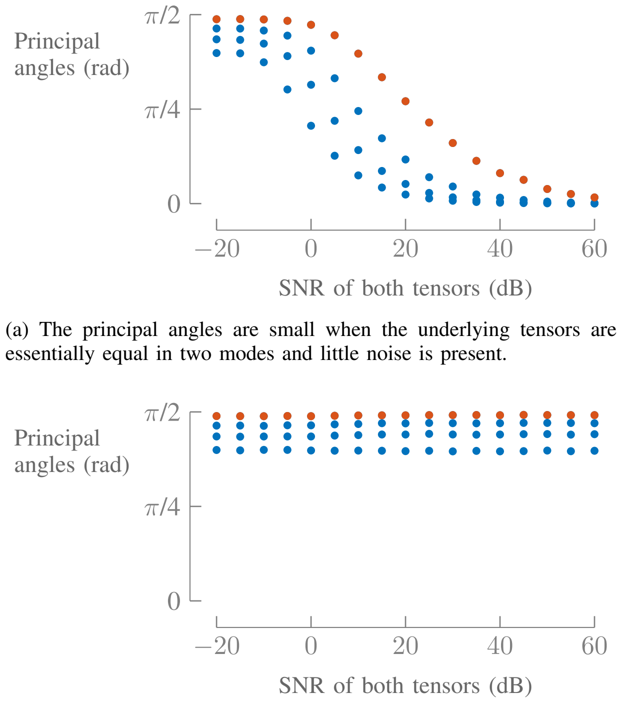
The developed algorithm will be illustrated using both synthetic and real-life data. First, the influence of noise on the principal angles is briefly discussed and the results are compared with a method that explicitly computes the factors. Next, the method is applied to emitter movement detection and fluorescence spectroscopy.

A. Influence of noise

In Section III, we stated that the principal angles can still be used in noisy conditions. Using numerical simulations, we show how they are affected when noise is present.

Let two tensors $\mathcal{T}^{(1)}$ and $\mathcal{T}^{(2)}\in\mathbb{R}^{10\times10\times10}$ admit a rank-4 CPD. The tensors are constructed as

$$\mathcal{T}^{(1)} = [\![\mathbf{A},\mathbf{B},\mathbf{C}]\!]\,,$$
$$\mathcal{T}^{(2)} = [\![\mathbf{A\Gamma},\mathbf{B\Lambda},\mathbf{F}]\!]\,,$$



(a) The principal angles are small when the underlying tensors are essentially equal in two modes and little noise is present.

(b) The principal angles are systematically large when the underlying tensors are not essentially equal.

Figure 5: Effects on the principal angles of varying noise levels on tensors that are essentially equal in two modes (a) and not essentially equal in two modes (b). For visual clarity, the largest principal value is shown in red.

in which the entries of $\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{F}\in\mathbb{R}^{10\times4}$ are pseudo-randomly drawn from a standard normal distribution. The matrices $\mathbf{\Gamma}$ and $\mathbf{\Lambda}\in\mathbb{R}^{4\times4}$ are diagonal with diagonal entries also drawn from a standard normal distribution. These tensors are clearly essentially equal in the first two modes. We now add Gaussian noise to both tensors such that various signal-to-noise ratios (SNR) are obtained. At each noise level, the principal angles are computed following the procedure outlined in Section III.

Figure 5a shows the mean principal angles between the unfolded tensors $\mathcal{T}^{(1)}$ and $\mathcal{T}^{(2)}$ over 1000 experiments at each noise level. It is clear that the principal angles become smaller if the noise level decreases. This behavior matches the error on the factor matrices if the decompositions would be explicitly computed, as illustrated in Figure 6. For the CPD computation, the default nonlinear least-squares algorithm of Tensorlab was used [34].

By contrast, Figure 5b shows the principal angles between the unfoldings of two randomly generated $10\times10\times10$ tensors admitting a rank-4 CPD. The figure shows that all principal angles are large if the underlying tensors are not essentially equal in two modes, no matter what the noise level is. Again, this matches the expected behavior of the method that explicitly decomposes the tensors. Comparing figures 5a and 5b, we can conclude that the principal angles are able to distinguish between essentially equal and unequal tensors if the noise level is not too high.

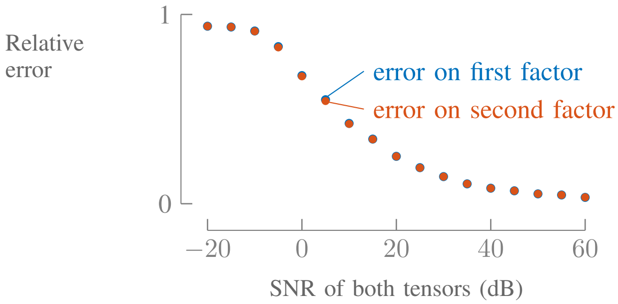


Figure 6: The relative errors on the first two factor matrices when the CPDs are explicitly computed show similar behavior as the principal angles. Note that the errors on both factors almost coincide.

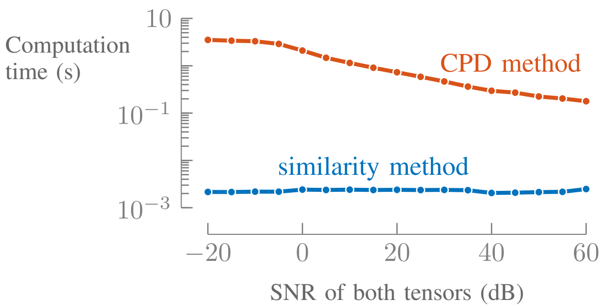
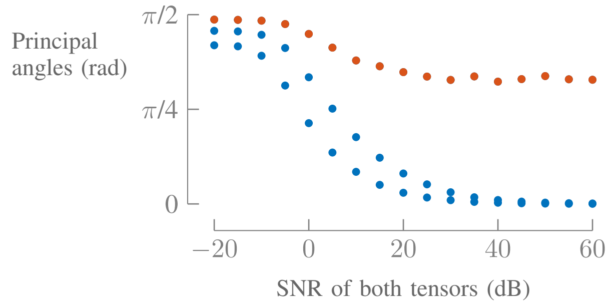
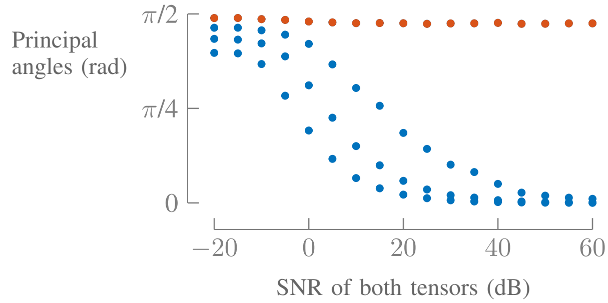


Figure 7: The similarity method is orders of magnitude faster than the method computing the CPDs and is independent of the noise level.



(a) The estimated number of components is lower (3) than the actual number of components (4).



(b) The estimated number of components is higher (4) than the actual number of components (3).

Figure 8: Effects on the principal angles of an incorrectly estimated number of components, which is either chosen too low (a) or too high (b) for tensors that are essentially equal in two modes. For visual clarity, the largest principal value is shown in red.

One of the main merits of our similarity method is its speed. Figure 7 shows that the similarity method is about two orders of magnitude faster than the approach that explicitly computes the factor matrices. Note that the speed of the latter depends on the noise level because the convergence of the CPD algorithm is generally slower in the presence of noise. By contrast, the speed of our similarity method is not affected. Similarly, our method is unaffected by the condition of the problem.

B. Incorrectly estimated number of components

In Section III-C2, we stated that the principal angles were affected if the number of components is incorrectly estimated. Using numerical simulations, we show how they are affected in a few examples.

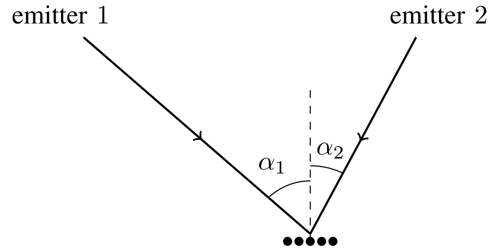
Let two tensors $\mathcal{T}^{(1)}$ and $\mathcal{T}^{(2)} \in \mathbb{R}^{10\times10\times10}$ admit a rank-4 CPD. The tensors are constructed as

$$\mathcal{T}^{(1)} = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]\,,$$
$$\mathcal{T}^{(2)} = [\![\mathbf{A\Gamma}, \mathbf{B\Lambda}, \mathbf{F}]\!]\,,$$

in which the entries of $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{F} \in \mathbb{R}^{10\times4}$ are pseudo-randomly drawn from a standard normal distribution. The matrices $\mathbf{\Gamma}$ and $\mathbf{\Lambda} \in \mathbb{R}^{4\times4}$ are diagonal with diagonal entries also drawn from a standard normal distribution. These tensors are clearly essentially equal in the first two modes. We now add Gaussian noise to both tensors such that various signal-to-noise ratios (SNR) are obtained. At each noise level, the principal angles are computed following the procedure outlined in Section III, but with the estimated number of components equal to $N = 3$.

Figure 8a shows the mean results of 300 experiments at each noise level, which clearly indicate that one of the principal angles remains nonzero. This is because the available 3-dimensional subspaces only contain a part of the actual 4-dimensional subspace, and those parts are not exactly the same. In this example, two principal angles are approximately zero for high SNRs, whereas the one that captures the differences between the partial subspaces remains large.

Similarly, let two tensors $\mathcal{T}^{(1)}$ and $\mathcal{T}^{(2)} \in \mathbb{R}^{10\times10\times10}$ admit a rank-3 CPD. The tensors are constructed in exactly the same way as before, with the exeption that the number of components is different now. The principal angles are computed following the procedure outlined in Section III, but with the estimated number of components equal to $N = 4$, so the number of components is overestimated by one. The mean results of 300 experiments at each noise level are shown in Figure 8b. As explained in Section III-C2, the fourth principal angle does not involve subspace information, which is fully contained in the first three angles. Therefore, the fourth one only involves noise and is approximately equal to $\pi/2$.



Antenna array

Figure 9: Direction of arrival estimation tries to retrieve the values α_r using solely the information obtained at the antenna array.

C. Application: Emitter movement detection

One possible application of our similarity measure can be found in emitter localization. Using an antenna array that receives data from several emitters, various methods are able to pinpoint the direction from which each emitter is transmitting, called the direction of arrival (DOA) [35], [36], [37], [38]. The direction-of-arrival problem is illustrated in Figure 9. Knowledge of this direction may be useful in military applications such as passive radar or sonar, or in telecommunications.

In [8], a tensor-based method is presented that captures the location information in the first two modes of a third-order tensor. We briefly repeat this method to clarify the relation between tensor decompositions and direction-of-arrival estimation. Consider a uniform linear array (ULA) of N uniformly spaced omnidirectional antennas as depicted in Figure 9. If this array receives signals from R narrowband emitters located in the far field, the received mixture $\mathbf{x}(t) \in \mathbb{C}^N$ at time instance t is given by

$$\mathbf{x}(t) = \mathbf{M}\mathbf{s}(t),$$

in which $\mathbf{s}(t) \in \mathbb{C}^R$ contains the emitter signals and $\mathbf{M} \in \mathbb{C}^{N\times R}$ is the mixing matrix with $m_{nr} = \theta_r^{n-1}$ and $\theta_r = e^{-2\pi i\Delta\sin(\alpha_r)\lambda^{-1}}$. Δ is the known inter-element spacing of the ULA, λ is the wavelength, and the angle α_r to the normal is the direction of arrival of the rth emitter and lies between $-\pi/2$ and $\pi/2$. It is shown in [8] that by rearranging the received mixture $\mathbf{x}$(t) into a matrix and stacking the obtained matrices at several time instances t, a tensor is obtained that admits a CPD. The factors in the first two modes contain the mixture information and the factor in the third mode contains the emitter data information. Using the first two factor matrices the matrix $\mathbf{M}$ can be reconstructed, from which the directions of arrival α_r can be extracted. A similar approach for large-scale applications has been studied in [4].

Continuously monitoring the emitter locations requires the recomputation of the DOAs at each time step. However, this approach is computationally expensive and is cumbersome if emitters remain stationary for long time intervals. Moreover, this expensive computation also implies higher power usage, which may be an important issue for (remote) battery-driven computation units. By using our similarity method, one can use just a fraction of the computational effort to check whether any emitter has moved. When this is the case, the DOAs can be recomputed using existing techniques.

For the experiment, an ULA of 100 antennas was simulated receiving data from two emitters that move over time. The transmitted data consist of multi-modulus samples with continuous phase shift keying and are constructed by randomly sampling from a uniform distribution over concentric circles in the complex plane with radii 1, 2 or 3. Gaussian noise was added to the antenna outputs such that an SNR of 20 dB was obtained. In Figure 10, results for the simulated direction of arrival problem are shown. The top figure shows the true direction angles of the emitters. The middle figure then shows the estimated direction angles. This estimation is done using a sliding window of length 10 as follows. First, the method from [8] is used to estimate the starting angles using the first 10 samples. In doing so, a tensor is constructed by a simple reshaping operation, which is stored in $\mathcal{A}$. Next, the sliding window is moved one time instance and a new tensor $\mathcal{B}$ is constructed using the same reshaping procedure as in [8]. However, this tensor is not decomposed, but the similarity method is used to check whether $\mathcal{A}$ and $\mathcal{B}$ are essentially equal in the first two modes. If this is the case, the previously computed direction angles are retained. If at least one of the principal angles is different from zero, the new direction angles are computed by decomposing $\mathcal{B}$ and we replace $\mathcal{A}$ by $\mathcal{B}$ such that all subsequent tensors are compared to the one that has been decomposed last. In this experiment, we consider a principal angle to be different from zero if it is larger than 0.2 rad to take noise effects into account. The principal angles computed using this procedure are shown at the bottom of Figure 10. The vertical lines represent time instances at which the direction angles have been explicitly recomputed. The figure shows that this only happens when one of the emitters has moved, which is exactly the goal of this approach.

On average, checking the principal angles takes about $3.65 \cdot 10^{-4}$ s at each time instance, whereas recomputing the direction angles takes 0.0649 s, which is a difference of two orders of magnitude.

D. Application: Fluorescence spectrography

Fluorescence spectroscopy data has been extensively analyzed in literature, often in an attempt to retrieve the pure analyte spectra making up the mixture, see [14] and references therein. The traditional way to do this involves stacking *several* excitation-emission matrices $\mathbf{E} \in \mathbb{C}^{K\times N}$, which gives a tensor admitting a low-rank CPD, see [14], [15] and references therein. The excitation-emission matrices are obtained by exciting the mixtures at K different excitation wavelengths and measuring the spectrum of the emitted light at N different emission wavelengths.

The real-life dataset[1] that we will use consists of two samples containing different relative concentrations of tyrosine (Tyr), tryptophane (Trp) and phenylalanine (Phe). Apart from these mixtures, measured excitation-emission spectra of the pure analytes are available as well. The excitation and emission

[1]Available at http://www.models.life.ku.dk/amino_acid_fluo

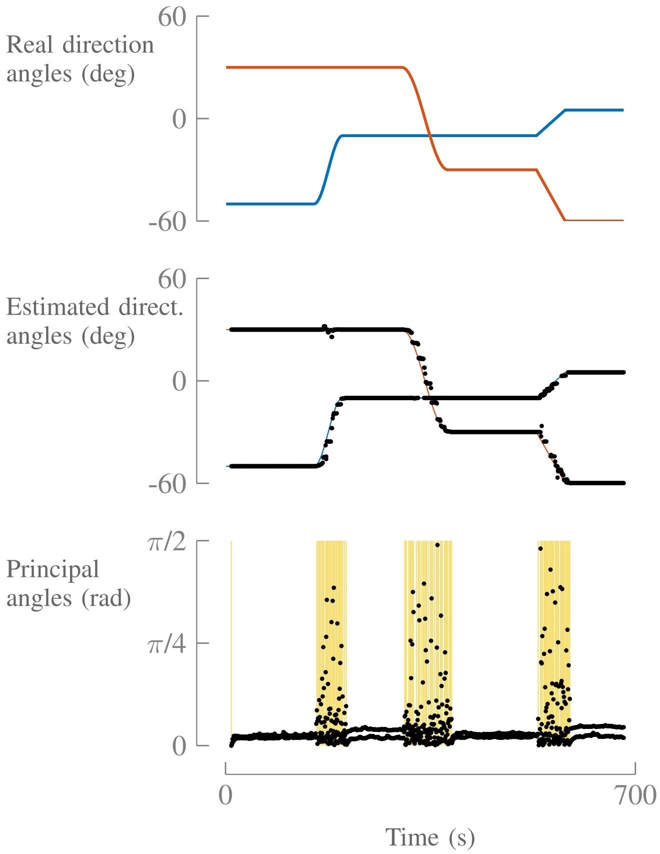


Figure 10: Using essential equality in two modes one can detect whether any emitter has displaced, which leads to lighter computational loads when there is no displacement.

wavelengths are $240-300$ nm and $250-450$ nm respectively, measured in steps of 1 nm. The obtained excitation-emission matrices thus have dimensions 201×61. A more detailed description of the data can be found in [14]. Here, we transform a *single* excitation-emission matrix to a tensor using Löwnerization [6]. In this method, a tensor is constructed from matrix data by transforming the data columns to Löwner matrices, which are then stacked in a third order tensor. It has been shown in [6] that the resulting tensor for our dataset can be approximated using a BTD in rank-$(L, L, 1)$ terms. The goal is to determine whether two mixtures consist of the same components without explicitly identifying them.

In a first experiment, we start by tensorizing the two available mixtures using Löwnerization along the emission mode. It has been shown in [6] that the resulting tensors $\mathcal{T}^{(1)}$ and $\mathcal{T}^{(2)} \in \mathbb{C}^{101\times100\times61}$ admit an approximate BTD in three rank-$(L, L, 1)$ terms with $L = 2$. This implies that each of the analytes contributes one rank-$(2, 2, 1)$ term. Though we know that the mixtures consist of three components, one could extract the number of components from the data as well. Figure 11 shows the first 8 singular values of $\mathbf{T}^{(1)}_{10100\times61}$ and $\mathbf{T}^{(2)}_{10100\times61}$, from which it immediately follows that the matrices can be well-approximated with 3 components. The principal angles between both matrices are 0.037, 0.064 and 0.328. As expected, these small values indicates that both

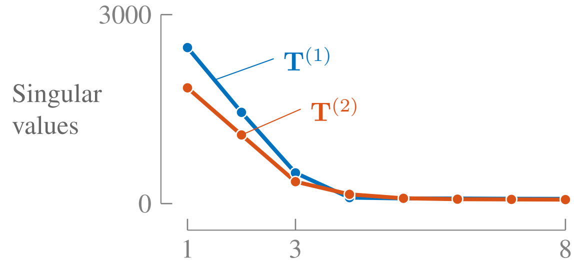


Figure 11: The first 3 singular values of $\mathbf{T}^{(1)}_{10100\times61}$ and $\mathbf{T}^{(2)}_{10100\times61}$ contain most of the information.

Table I: The principal angles clearly indicate which mixtures consist of the same components.

Mixtures to compare	Principal angles (rad)	
Mix1 and Mix2	0.0664	1.3032
Mix1 and Mix3	0.0217	0.0674
Mix2 and Mix3	0.1198	1.2892

mixtures consist of the same components. These values differ from zero due to measurement noise and the presence of Rayleigh scatter in the data [14].

For the second experiment, we use the excitation-emission matrices of the separate analytes to contruct synthetic mixtures of two analytes as follows:

$$\begin{aligned}
\mathrm{EE}\,(\mathrm{Mix1}) &= \mathrm{EE}\,(\mathrm{Trp}) + \mathrm{EE}\,(\mathrm{Tyr})\,,\\
\mathrm{EE}\,(\mathrm{Mix2}) &= 0.4\ \mathrm{EE}\,(\mathrm{Tyr}) + 0.8\ \mathrm{EE}\,(\mathrm{Phe})\,,\\
\mathrm{EE}\,(\mathrm{Mix3}) &= 0.8\ \mathrm{EE}\,(\mathrm{Trp}) + 0.4\ \mathrm{EE}\,(\mathrm{Tyr})\,,
\end{aligned}$$

in which we use $\mathrm{EE}\,(\mathrm{x})$ as shorthand to denote the excitation-emission matrix of x. The excitation-emission matrices of all three mixtures are then tensorized using Löwnerization as in the previous experiment. The principal angles resulting from the comparison of the mixtures are given in Table I. The table shows small principal angles between the tensors associated with mixtures 1 and 3, which indeed consist of the same components. Note that the other comparisons show one small and one large principal angle, which corresponds to the mixtures having one component in common. Though we did not explicitly generalize Theorem 6 to check partial equality of CPDs to BTDs, we expect similar results for the latter. This seems to be confirmed by the data in Table I.

VII. Conclusions

We have introduced a method to compare underlying factors of tensor decompositions in two modes without explicitly computing the factors. A step-by-step guide was provided detailing how to proceed in practice. Theoretical theorems were given for CPDs, BTDs in rank-$(L, L, 1)$ terms and BTDs in rank-$(L, L, \cdot)$ terms. Additionally, the partial similarity in two modes of CPDs was considered and we provided a result for non-unique CPDs. The method was illustrated using artificial data and two applications were tackled. The experiments have shown that the method is able to efficiently determine whether the factors in two modes are essentially equal.

APPENDIX A
PROOF OF THEOREM 4

Because both $\mathcal{T}^{(1)}$ and $\mathcal{T}^{(2)}$ admit a BTD in rank-$(L, L, 1)$ terms, their unfoldings can be written as

$$\mathbf{T}^{(1)}_{IJ\times K} = [(\mathbf{B}_1 \odot \mathbf{A}_1)\,\mathbf{1}_L, \ldots, (\mathbf{B}_R \odot \mathbf{A}_R)\,\mathbf{1}_L] \cdot \mathbf{C}^{\mathrm{T}},$$
$$\mathbf{T}^{(2)}_{IJ\times K} = [(\mathbf{E}_1 \odot \mathbf{D}_1)\,\mathbf{1}_L, \ldots, (\mathbf{E}_R \odot \mathbf{D}_R)\,\mathbf{1}_L] \cdot \mathbf{F}^{\mathrm{T}},$$

as explained in Section II-B.

We first show that $\operatorname{span}\left\{\mathbf{T}^{(1)}_{IJ\times K}\right\} = \operatorname{span}\left\{\mathbf{T}^{(2)}_{IJ\times K}\right\}$ is a necessary condition. Assume $\mathcal{T}^{(1)}$ and $\mathcal{T}^{(2)}$ are essentially equal in the first two modes, i.e., that $\mathbf{D}_r = \mathbf{A}_r\mathbf{G}_r$ and $\mathbf{E}_r = \mathbf{B}_r\left(\mathbf{G}_r^{-1}\right)^{\mathrm{T}}$ for each $r \in \{1, \ldots, R\}$ with $\mathbf{G}_r$ nonsingular matrices. Also, the different terms may be permuted and we assume without loss of generality that any further scaling is captured in the factors $\mathbf{C}$ and $\mathbf{F}$. Because $(\mathbf{B}_r \odot \mathbf{A}_r)\,\mathbf{1}_L$ is simply the vectorized matrix product $\mathbf{A}_r\mathbf{B}_r^{\mathrm{T}} = \mathbf{A}_r\mathbf{G}_r\mathbf{G}_r^{-1}{\mathbf{B}_r}^{\mathrm{T}} = \mathbf{D}_r\mathbf{E}_r^{\mathrm{T}}$, it follows that

$$\mathbf{T}^{(1)}_{IJ\times K} = [(\mathbf{B}_1 \odot \mathbf{A}_1)\,\mathbf{1}_L, \ldots, (\mathbf{B}_R \odot \mathbf{A}_R)\,\mathbf{1}_L]\,\mathbf{C}^{\mathrm{T}},$$
$$\mathbf{T}^{(2)}_{IJ\times K} = [(\mathbf{B}_1 \odot \mathbf{A}_1)\,\mathbf{1}_L, \ldots, (\mathbf{B}_R \odot \mathbf{A}_R)\,\mathbf{1}_L]\,\mathbf{\Pi}\mathbf{F}^{\mathrm{T}},$$

in which the $\mathbf{\Pi}$ accounts for the possible column permutation. Since $\mathbf{C}$ and $\mathbf{F}$ are assumed to have full column rank and $\mathbf{\Pi}$ is nonsingular, it immediately follows that $\operatorname{span}\left\{\mathbf{T}^{(1)}_{IJ\times K}\right\} = \operatorname{span}\left\{\mathbf{T}^{(2)}_{IJ\times K}\right\}$.

Conversely, assume $\operatorname{span}\left\{\mathbf{T}^{(1)}_{IJ\times K}\right\} = \operatorname{span}\left\{\mathbf{T}^{(2)}_{IJ\times K}\right\}$ holds to show that this subspace equality is also sufficient. Because $\mathbf{C}$ and $\mathbf{F}$ have full column rank, the subspace equality can be written as

$$\begin{aligned}&[(\mathbf{B}_1 \odot \mathbf{A}_1)\,\mathbf{1}_L, \ldots, (\mathbf{B}_R \odot \mathbf{A}_R)\,\mathbf{1}_L]\left(\mathbf{M}^{(1)}\right)^{\mathrm{T}}\\ &= [(\mathbf{E}_1 \odot \mathbf{D}_1)\,\mathbf{1}_L, \ldots, (\mathbf{E}_R \odot \mathbf{D}_R)\,\mathbf{1}_L]\left(\mathbf{M}^{(2)}\right)^{\mathrm{T}}, \qquad (22)\end{aligned}$$

with $\mathbf{M}^{(1)}, \mathbf{M}^{(2)} \in \mathbb{C}^{R\times R}$ nonsingular matrices. Both sides of equation (22) represent an unfolded BTD in rank-$(L, L, 1)$ terms of the same tensor. Consequently, we can write

$$\begin{aligned}\mathcal{G} &= \sum_{r=1}^{R}\left(\mathbf{A}_r\mathbf{B}_r^{\mathrm{T}}\right) \circ \mathbf{m}_r^{(1)} \qquad (23)\\ &= \sum_{r=1}^{R}\left(\mathbf{D}_r\mathbf{E}_r^{\mathrm{T}}\right) \circ \mathbf{m}_r^{(2)} \in \mathbb{C}^{I\times J\times R}.\end{aligned}$$

It follows from (9a) that the BTD in rank-$(L, L, 1)$ terms in equation (23) is essentially unique. Consequently, the factor matrices $\mathbf{A}$ and $\mathbf{D}$, and $\mathbf{B}$ and $\mathbf{E}$ are equal up to the trivial indeterminacies stated in Section II-B.

APPENDIX B
PROOF OF THEOREM 5

We first show that $\operatorname{span}\left\{\mathbf{T}^{(1)}_{IJ\times K}\right\} = \operatorname{span}\left\{\mathbf{T}^{(2)}_{IJ\times K}\right\}$ is a necessary condition. Assume $\mathcal{T}^{(1)}$ and $\mathcal{T}^{(2)}$ are essentially equal in the first two modes, i.e., that $\mathbf{C}_r = \mathbf{A}_r\mathbf{G}_r$ and $\mathbf{D}_r = \mathbf{B}_r\left(\mathbf{G}_r^{-1}\right)^{\mathrm{T}}$ for each $r \in \{1, \ldots, R\}$ with $\mathbf{G}_r \in \mathbb{C}^{L\times L}$ nonsingular matrices. Also, the different terms may be permuted and we assume without loss of generality that any further scaling is captured in the factors $\mathbf{C}$ and $\mathbf{F}$. Because $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$ have full row rank, bases for the column spaces of $\mathbf{T}^{(1)}_{IJ\times K}$ and $\mathbf{T}^{(2)}_{IJ\times K}$ are given by

$$\operatorname{span}\left\{\mathbf{T}^{(1)}_{IJ\times K}\right\} = \operatorname{span}\left\{[\mathbf{B}_1 \otimes \mathbf{A}_1, \ldots, \mathbf{B}_R \otimes \mathbf{A}_R]\right\},$$
$$\begin{aligned}\operatorname{span}\left\{\mathbf{T}^{(2)}_{IJ\times K}\right\} = \operatorname{span}\Big\{\Big[&\left(\mathbf{B}_1\left(\mathbf{G}_1^{-1}\right)^{\mathrm{T}}\right) \otimes (\mathbf{A}_1\mathbf{G}_1), \ldots,\\ &\left(\mathbf{B}_R\left(\mathbf{G}_R^{-1}\right)^{\mathrm{T}}\right) \otimes (\mathbf{A}_R\mathbf{G}_R)\Big]\Big\}.\end{aligned}$$

Using the Kronecker product property $(\mathbf{UV}) \otimes (\mathbf{WX}) = (\mathbf{U} \otimes \mathbf{W})(\mathbf{V} \otimes \mathbf{X})$, the second basis becomes

$$\operatorname{span}\left\{\mathbf{T}^{(2)}_{IJ\times K}\right\} = \operatorname{span}\left\{[\mathbf{B}_1 \otimes \mathbf{A}_1, \ldots, \mathbf{B}_R \otimes \mathbf{A}_R]\right\},$$

from which it immediately follows that $\operatorname{span}\left\{\mathbf{T}^{(1)}_{IJ\times K}\right\} = \operatorname{span}\left\{\mathbf{T}^{(2)}_{IJ\times K}\right\}$.

Conversely, assume $\operatorname{span}\left\{\mathbf{T}^{(1)}_{IJ\times K}\right\} = \operatorname{span}\left\{\mathbf{T}^{(2)}_{IJ\times K}\right\}$ holds to show that this condition is sufficient as well. Because $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$ have full row rank, the subspace equality can be written as

$$\begin{aligned}&[\mathbf{B}_1 \otimes \mathbf{A}_1, \ldots, \mathbf{B}_R \otimes \mathbf{A}_R]\left(\mathbf{M}^{(1)}\right)^{\mathrm{T}}\\ &= [\mathbf{D}_1 \otimes \mathbf{C}_1, \ldots, \mathbf{D}_R \otimes \mathbf{C}_R]\left(\mathbf{M}^{(2)}\right)^{\mathrm{T}}, \qquad (24)\end{aligned}$$

with $\mathbf{M}^{(1)}, \mathbf{M}^{(2)} \in \mathbb{C}^{RL^2\times RL^2}$ nonsingular matrices. Both sides of equation (24) represent an unfolded BTD in rank-$(L, L, \cdot)$ terms of the same tensor. Consequently, we can write

$$\begin{aligned}\mathcal{G} &= \sum_{r=1}^{R}\mathcal{M}_r^{(1)} \cdot_1 \mathbf{A}_r \cdot_2 \mathbf{B}_r, \qquad (25)\\ &= \sum_{r=1}^{R}\mathcal{M}_r^{(2)} \cdot_1 \mathbf{C}_r \cdot_2 \mathbf{D}_r \in \mathbb{C}^{I\times J\times RL^2}.\end{aligned}$$

Since it was assumed in (12a) that this BTD is essentially unique, the factor matrices $\mathbf{A}$ and $\mathbf{D}$, and $\mathbf{B}$ and $\mathbf{E}$ are equal up to the trivial indeterminacies stated in Section II-B.

REFERENCES

[1] T. Kolda and B. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.

[2] 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.*, vol. 65, no. 13, pp. 3551–3582, 2017.

[3] P. Comon and C. Jutten, Eds., *Handbook of blind source separation, independent component analysis and applications*. Academic Press, 2010.

[4] M. Boussé, O. Debals, and L. De Lathauwer, "A tensor-based method for large-scale blind source separation using segmentation," *IEEE Trans. Signal Process.*, vol. 65, no. 2, pp. 346–358, 2017.

[5] A. Cichocki, R. Zdunek, A. H. Phan, and S. Amari, *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*. John Wiley & Sons, 2009.

[6] O. Debals, M. Van Barel, and L. De Lathauwer, "Löwner-based blind signal separation of rational functions with applications," *IEEE Trans. Signal Process.*, vol. 64, no. 8, pp. 1909–1918, 2016.

[7] 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.*, vol. 32, no. 4, pp. 1451–1474, 2011.

[8] N. D. Sidiropoulos, R. Bro, and G. B. Giannakis, "Parallel factor analysis in sensor array processing," *IEEE Trans. Signal Process.*, vol. 48, no. 8, pp. 2377–2388, 2000.

[9] D. Nion and N. D. Sidiropoulos, "Tensor algebra and multidimensional harmonic retrieval in signal processing for MIMO radar," *IEEE Trans. Signal Process.*, vol. 58, no. 11, pp. 5693–5705, 2010.

[10] M. A. O. Vasilescu and D. Terzopoulos, "Multilinear analysis of image ensembles: Tensorfaces," in *European Conference on Computer Vision*. Springer, 2002, pp. 447–460.

[11] A. Cichocki, C. Mandic, A. H. Phan, C. Caifa, G. Zhou, Q. Zhao, and L. De Lathauwer, "Tensor decompositions for signal processing applications. from two-way to multiway component analysis," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 145–163, Mar. 2015.

[12] A. Anandkumar, R. Ge, D. J. Hsu, S. M. Kakade, and M. Telgarsky, "Tensor decompositions for learning latent variable models." *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 2773–2832, 2014.

[13] L. De Lathauwer, "Characterizing higher-order tensors by means of subspaces," 2011, Tech. Report 11–32, ESAT–STADIUS, KU Leuven, Belgium.

[14] R. Bro, "PARAFAC. Tutorial and applications," *Chemometrics and intelligent laboratory systems*, vol. 38, no. 2, pp. 149–171, 1997.

[15] A. Smilde, R. Bro, and P. Geladi, *Multi-way analysis: applications in the chemical sciences*. John Wiley & Sons, 2005.

[16] M. Signoretto, L. De Lathauwer, and J. Suykens, "A kernel-based framework to tensorial data analysis," *Neural networks*, vol. 24, no. 8, pp. 861–874, 2011.

[17] T.-K. Kim, S.-F. Wong, and R. Cipolla, "Tensor canonical correlation analysis for action classification," in *IEEE Conference on Computer Vision and Pattern Recognition.* IEEE, 2007, pp. 1–8.

[18] I. Domanov and L. De Lathauwer, "Canonical polyadic decomposition of third-order tensors: relaxed uniqueness conditions and algebraic algorithm," *Linear Algebra and its Applications*, vol. 513, pp. 342–375, 2017.

[19] 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.*, vol. 34, no. 3, pp. 876–903, 2013.

[20] 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.*, vol. 23, no. 2, pp. 695–720, 2013.

[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.*, vol. 34, no. 3, pp. 855–875, 2013.

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

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

[24] ——, "Decomposition of a Higher-Order Tensor in Block Terms — Part II: Definitions and Uniqueness," *SIAM J. Matrix Anal. Appl.*, vol. 30, pp. 1033–1066, 2008.

[25] R. Bro, R. A. Harshman, N. D. Sidiropoulos, and M. E. Lundy, "Modeling multi-way data with linearly dependent loadings," *Journal of Chemometrics*, vol. 23, no. 7-8, pp. 324–340, 2009.

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

[27] L. De Lathauwer and D. Nion, "Decompositions of a Higher-Order Tensor in Block Terms — Part III: Alternating least squares algorithms," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1067–1083, 2008.

[28] L. Sorber, M. Van Barel, and L. De Lathauwer, "Structured data fusion," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 4, pp. 586–600, June 2015.

[29] I. Domanov and L. De Lathauwer, "Canonical polyadic decomposition of third-order tensors: reduction to generalized eigenvalue decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 35, no. 2, pp. 636–660, 2014.

[30] B. Yang, "Projection approximation subspace tracking," *IEEE Trans. Signal Process.*, vol. 43, no. 1, pp. 95–107, 1995.

[31] I. Markovsky, *Low rank approximation: algorithms, implementation, applications*. Springer Science & Business Media, 2011.

[32] A. Björck and G. H. Golub, "Numerical methods for computing angles between linear subspaces," *Mathematics of computation*, vol. 27, no. 123, pp. 579–594, 1973.

[33] A. V. Knyazev and M. E. Argentati, "Principal angles between subspaces in an A-based scalar product: algorithms and perturbation estimates," *SIAM Journal on Scientific Computing*, vol. 23, no. 6, pp. 2008–2040, 2002.

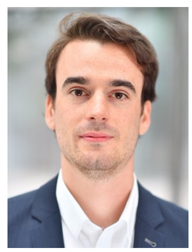
[34] N. Vervliet, O. Debals, L. Sorber, M. Van Barel, and L. De Lathauwer, "Tensorlab 3.0," Mar. 2016, available online. [Online]. Available: http://www.tensorlab.net

[35] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Process. Mag.*, vol. 13, no. 4, pp. 67–94, 1996.

[36] P. Stoica and K. C. Sharman, "Maximum likelihood methods for direction-of-arrival estimation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 7, pp. 1132–1143, 1990.

[37] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE transactions on antennas and propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[38] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on acoustics, speech, and signal processing*, vol. 37, no. 7, pp. 984–995, 1989.



Frederik Van Eeghem obtained his M.Sc. degree in Mathematical Engineering from KU Leuven, Belgium, in 2014. Currently, he works as a technology consultant and is a Ph.D. candidate affiliated with the Group Science, Engineering and Technology of Kulak, KU Leuven and with the STADIUS Center for Dynamical Systems, Signal Processing, and Data Analytics of the Electrical Engineering Department of KU Leuven. His research interests include blind system identification, independent component analysis and tensor decompositions.

Otto Debals obtained the M.Sc. degree in mathematical engineering from KU Leuven, Belgium, in 2013 and a Ph.D. degree from KU Leuven in 2017. Currently, he works as a management consultant.

Lieven De Lathauwer received the Ph.D. degree from the Faculty of Engineering, KU Leuven, Belgium, in 1997. From 2000 to 2007 he was Research Associate with the Centre National de la Recherche Scientifique, France. He is currently Professor with KU Leuven. He is affiliated with both the Group Science, Engineering and Technology of Kulak and with the Stadius Center for Dynamical Systems, Signal Processing and Data Analytics of the Electrical Engineering Department (ESAT). He is Associate Editor of the SIAM Journal on Matrix Analysis and Applications and has served as Associate Editor for the IEEE Transactions on Signal Processing. He is Fellow of the IEEE and Fellow of SIAM. His research concerns the development of tensor tools for engineering applications.