

Personalized Coupled Tensor Decomposition for Multimodal Data Fusion: Uniqueness and Algorithms

Ricardo A. Borsoi, Konstantin Usevich, David Brie, Tülay Adali

Abstract—Coupled tensor decompositions (CTDs) perform data fusion by linking factors from different datasets. Although many CTDs have been already proposed, current works do not address important challenges of data fusion, where: 1) the datasets are often heterogeneous, constituting different “views” of a given phenomena (multimodality); and 2) each dataset can contain *personalized* or dataset-specific information, constituting distinct factors that are not coupled with other datasets. In this work, we introduce a personalized CTD framework tackling these challenges. A flexible model is proposed where each dataset is represented as the sum of two components, one related to a common tensor through a multilinear measurement model, and another specific to each dataset. Both the common and distinct components are assumed to admit a polyadic decomposition. This generalizes several existing CTD models. We provide conditions for specific and generic uniqueness of the decomposition that are easy to interpret. These conditions employ *uni-mode* uniqueness of different individual datasets and properties of the measurement model. Two algorithms are proposed to compute the common and distinct components: a semi-algebraic one and a coordinate-descent optimization method. Experimental results illustrate the advantage of the proposed framework compared with the state of the art approaches.

Index Terms—Coupled tensor decomposition, personalized learning, shared and distinct components, uniqueness.

I. INTRODUCTION

Coupled tensor decompositions (CTDs) perform data fusion by linking factors among different datasets, benefiting from the complementary information across modalities [1], [2]. CTD has recently drawn increasing interest in various disciplines, particularly in data mining and multimodal data fusion [1]–[5]. Applications include social network data analysis [6] and link prediction [7], multimodal image fusion [8]–[11], multi-subject fMRI data fusion [12], [13], and several problems in array signal processing [14]–[18]. CTDs are also important in methods based on high-order statistics for blind system identification [19], joint blind source separation from multiple datasets [20], and for blind deconvolution of underdetermined systems [21]. Moreover, related approaches coupling both matrix and tensor data have shown great potential in

This work has been partially supported by the French National Research Agency under grants ANR-23-CE23-0024, ANR-23-CE94-0001, and by the National Science Foundation, under grant NSF 2316420.

R. Borsoi, K. Usevich, and D. Brie are with the Université de Lorraine, CNRS, CRAN, Vandoeuvre-lès-Nancy, France. e-mail: {firstname.lastname}@univ-lorraine.fr.

T. Adali is with the Department of Computer Science and Electrical Engineering, University of Maryland Baltimore County, Baltimore, MD 21250, USA. e-mail: {lastname}@umbc.edu.

metabolomics, combining nuclear magnetic resonance and spectrometry or spectroscopy measurements [22], and in neuroimaging, fusing electroencephalogram (EEG) with functional magnetic resonance imaging (fMRI) data [23], [24], or fMRI with magnetoencephalography (MEG) data [25]. The flexibility of CTDs allow them to address data fusion both in the *multiset* setting, where data is collected using the same modality over different conditions or observation times (e.g., fMRI datasets of different subjects), and also in the *multimodal* setting, where information is collected about the same phenomenon using different modalities or sensors (e.g., EEG and fMRI datasets) [26].

One of the main advantages of CTDs is that the coupled decomposition can be unique even when uniqueness does not hold for the decomposition of any of the individual tensors and matrices [27]. This occurs, for example, when coupling images with different spectral and spatial resolutions [9]. Thus, CTDs can exploit advantages of each data modality, being able to leverage weaker uniqueness results of each dataset to provide uniqueness to the full decomposition.

Shared and distinct features: In many applications, the following aspect of CTDs is important: even if the modes of all coupled tensors express the same “view” of the data (e.g., spatial dimensions, subjects, time, etc.), there is often content unique to each dataset in addition to the content shared with other measurements. This can occur due to datasets being captured from multiple subjects or from experiments running in different conditions, which provides more diversity but also incurs factors of variation which do not occur in all measurements [28]. For instance, in multitask fMRI data fusion this allows one to account for task-specific information [13]. Moreover, accurately identifying underlying shared and distinct factors from EEG and fMRI data reveals biologically meaningful components (biomarkers) which allow for differentiating between patients with schizophrenia and healthy controls [29]. Subspace-based representation learning of face images can be made more accurate by considering one subspace common to all persons along with others unique to each individual [30]. Moreover, when fusing hyperspectral images (HSI) with multispectral images (MSI) of different spatial/spectral resolutions, such models account for the distinct information present in each image due to differences between their acquisition conditions (e.g., cloud cover, illumination or atmosphere) [10], [11], [31], addressing the so-called inter-image variability problem. In what follows, we consider a general framework of so-called *personalized* decompositions,

which include some components that are common to all datasets, and other which are distinct to each dataset.

Traditional CTDs fail to consider shared and distinct features in the datasets, which motivated the development of more flexible models. For example, in [4] the authors proposed a flexible decomposition framework in which the contribution of the coupled components to each dataset is weighted by a set of sparse coefficients. This allows components not to contribute to some datasets. Other works proposed to “uncouple” a subset of the columns of one of the factor matrices. For instance, the authors in [32] perform nonnegative tensor factorization with common and distinct parts in one factor to achieve representation learning for classification tasks. A similar model was used in [33] for the analysis of critical events in data (from, e.g., social networks). More recently, this approach has been considered for coupled matrix decomposition [34], and for the coupled block term decomposition (BTD) in multimodal image fusion [10], [11]. A different approach uses flexible couplings, in which the factor matrices of different tensors or matrices are not constrained to be equal, but only to belong to a Euclidean ball [35], [36]. However, previous approaches suffer from one of two limitations: they either lack flexibility in the observation model or do not have established uniqueness guarantees. In this work we will consider a general personalized CTD framework, addressing both these issues.

Uniqueness of CTDs: Uniqueness of matrix and tensor decompositions is of fundamental importance in most applications as it allows the recovered factors to be directly interpreted when there is a match between the decomposition and the data generating model [37]. The uniqueness of the coupled canonical polyadic decomposition (CPD) was first studied in [38], [39]. The authors derived uniqueness guarantees based on the Kruskal condition for the case when one of the factors was coupled for all tensors.

More recent works established uniqueness for more flexible data acquisition models and different tensor decompositions. The recovery of an order-3 tensor admitting a CPD from measurements acquired according to sampling schemes acting separately on each mode (such as sampling entries, fibers or slabs) was studied in [40]. Conditions for the recovery of the full tensor were given depending on the uniqueness of individual tensors in the decomposition. This approach is closely related to [9] in the context of hyperspectral and multispectral image fusion, and to [41] for spectrum cartography in the case when the observed tensors have missing data, in which a BTD model is considered. However, these problems do not require the uniqueness of all factors of the decomposition; instead, only the recoverability of the unobserved “high-resolution” tensor is necessary.

Although the basic problem of fusing HSIs and MSIs does not require the uniqueness of all factor matrices, recent work addressed the inter-image variability problem by considering flexible CTDs containing both shared and distinct factors. In this case, conditions ensuring the recoverability of the true high resolution image tensor based on degraded measurements were obtained for image tensors admitting both an LL1 BTD [11] and a Tucker decomposition [10]. However, those

works assume a very specific measurement model which also contains no variability (i.e., distinct components) for one of the measured tensors.

Other related work studied the recoverability of principal component analysis [42] and dictionary learning [43], [44] with shared and dataset-specific factors using incoherence assumptions between the shared and distinct components. The uniqueness of coupled decomposition with structured matrices accounting for shared and unique factors was studied in [34].

Contribution: In this paper, we propose a general personalized CTD framework for multimodal data fusion. A distinctive feature of our approach is that we clearly show how it is possible to leverage the so-called *uni-modal uniqueness* of each tensor [45], [46], which can be satisfied under milder conditions when compared to the traditional uniqueness requirements, to provide uniqueness guarantees for the full decomposition. The main contributions of this paper are:

1) Our coupling framework generalizes several existing ones, including [9], [40] (which do not have distinct components), [10], [11], which uses a very specific measurement model and have a distinct component in only one dataset. Although [34] provides uniqueness results for the CPD with shared and distinct components, it does not consider a general measurement/degradation model or heterogeneity between datasets. However, the results in [34] are based on a different low-rank model and cannot be directly compared to the results presented in this paper.

2) The general multilinear measurement model comprises a wide range of possible practical applications in, e.g., hyperspectral and neuroimaging (e.g., fMRI, MRI, MEG) data fusion problems.

3) Uniqueness results which are easy to interpret and provide insight on the role that “weaker” (uni-mode) uniqueness of each measured tensor and the measurement model have on the uniqueness of the full decomposition. An important feature of the proofs is that they are constructive and directly motivate a semi-algebraic decomposition method.

4) Semi-algebraic (motivated by the constructive uniqueness results) and optimization-based algorithms are proposed to compute the decomposition. The solution provided by the semi-algebraic algorithm also serves as a principled approach to initialize optimization-based methods.

A general measurement model is presented in Section III where each measured tensor is represented as a sum of two terms. The first term is linked to a *common component* through an arbitrary multilinear measurement/degradation model, which couples the different datasets. The second (uncoupled) component is *distinct* to each measured tensor representing personalized or dataset-specific information. The common and distinct tensors are assumed to admit CPDs. In Section IV we demonstrate that this decomposition is generically unique (i.e., the common and distinct components can be recovered) under mild conditions. The developed uniqueness conditions are also highly interpretable: informally, they show that recovery of the common and distinct components is possible as long as 1) one of the low-resolution measured tensors is fully unique, and 2) for each mode j , there exists a measured tensor that is mode- j unique and whose mode- j

factor is measured at full resolution. Obtaining uniqueness of the full decomposition from “weaker” uniqueness results of each individual tensor highlights the benefits of data fusion. A semi-algebraic algorithm inspired by the uniqueness results as well as an optimization approach based on alternating least squares (ALS) are proposed in Section V. Experiments with synthetic data and real hyperspectral images illustrate the advantage of the proposed framework compared to competing algorithms. In addition, a preliminary conference version of this work containing experimental results for multi-subject and multitask fMRI data fusion based on a more constrained version of the model proposed in Section III was presented in [13]. These results further demonstrate the practical value of the proposed model.

II. BACKGROUND

A. Definitions and notation

Scalars, vectors and matrices are denoted by plain font (x or X), lowercase bold font (\mathbf{x}) and uppercase bold font (\mathbf{X}), respectively. Order-3 tensors are represented by calligraphic bold font (\mathcal{X}). The (i, j) -th element of a matrix X is denoted by $[X]_{i,j}$, while the (i, j, k) -th element of a tensor \mathcal{X} is denoted by $[\mathcal{X}]_{i,j,k}$. We also use the notation $a : b$, for a and b being positive integers to compute submatrices of a given matrix. For example, $[X]_{1:a, 1:b}$ denotes the first a rows and b columns of X . We denote by \otimes and by \odot the Kronecker and the Khatri-Rao products [47]. The outer product between three vectors is defined as $\mathbf{a} \circ \mathbf{b} \circ \mathbf{c}$, $[\mathbf{a} \circ \mathbf{b} \circ \mathbf{c}]_{i,j,k} = a_i b_j c_k$, and constitutes a rank-1 tensor. We denote the left pseudoinverse of matrix X by X^\dagger . A mode- k fiber of a tensor \mathcal{X} is the vector obtained by fixing all but one mode of \mathcal{X} , while a slice is a two dimensional subset obtained by fixing one mode of \mathcal{X} . For $K \in \mathbb{N}_+$, we also use $[K]$ to represent the set of integers $\{1, \dots, K\}$. Note that although we consider the case of third order tensors to keep the presentation cleaner, the proposed framework can be extended to higher order tensors as well.

Definition 1. The Kruskal rank (also called K -rank) of a matrix X , denoted by $\text{kr}(X)$, is the largest number r such that every set of r columns of X are linearly independent.

Definition 2. The mode- k matricization of a tensor $\mathcal{T} \in \mathbb{R}^{N_1 \times N_2 \times N_3}$, denoted by $\mathbf{T}_{\langle k \rangle}$, arranges its mode- k fibers to be the rows of matrix $\mathbf{T}_{\langle k \rangle} \in \mathbb{R}^{N_k N_m \times N_k}$ such that the n_k -th column of $\mathbf{T}_{\langle k \rangle}$ consists of the vectorization of the slice of \mathcal{T} obtained by fixing the index of the k -th mode of \mathcal{T} as n_k .

Definition 3. The mode- k product between a tensor \mathcal{T} and a matrix B is denoted by $\mathcal{T} \times_k B$ and consists of multiplying every mode- k fiber of \mathcal{T} by B . It can be written using the mode- k matricization as $\mathbf{U} = \mathcal{T} \times_k B \Leftrightarrow \mathbf{U}_{\langle k \rangle} = \mathbf{T}_{\langle k \rangle} B^\top$.

Definition 4. The full multilinear product consists of mode- k products between a tensor \mathcal{T} and matrices B_k for each mode $k \in \{1, 2, 3\}$, and can be expressed as $[\mathcal{T}; B_1, B_2, B_3] = \mathcal{T} \times_1 B_1 \times_2 B_2 \times_3 B_3$.

B. Canonical Polyadic Decomposition

The polyadic decomposition expresses a tensor $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ as a sum of rank-1 terms:

$$\mathcal{X} = \sum_{r=1}^R \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r = [\mathbf{A}, \mathbf{B}, \mathbf{C}], \quad (1)$$

where matrices $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_R]$, $\mathbf{B} = [\mathbf{b}_1, \dots, \mathbf{b}_R]$, $\mathbf{C} = [\mathbf{c}_1, \dots, \mathbf{c}_R]$ are called the factor matrices. When R is minimal, this is called the canonical polyadic decomposition (CPD), and R is called the (CP) rank of \mathcal{X} .

A fundamental property of the CPD is that it is *essentially unique* under mild conditions. More precisely, a decomposition $\mathcal{X} = [\mathbf{A}, \mathbf{B}, \mathbf{C}]$ is essentially unique if every alternative decomposition $\mathcal{X} = [\mathbf{A}', \mathbf{B}', \mathbf{C}']$ satisfies:

$$\mathbf{A}' = \mathbf{A} \Pi \Lambda_A, \quad (2)$$

$$\mathbf{B}' = \mathbf{B} \Pi \Lambda_B, \quad (3)$$

$$\mathbf{C}' = \mathbf{C} \Pi \Lambda_C, \quad (4)$$

where Π is a permutation matrix and $\Lambda_A, \Lambda_B, \Lambda_C$ are diagonal matrices satisfying $\Lambda_A \Lambda_B \Lambda_C = \mathbf{I}$. This means that the factors of the decomposition are unique up to a permutation and scaling of their columns. We also say that one of the factor matrices of the CPD, say, \mathbf{C} , is essentially unique if (4) is satisfied, but nothing is required of the remaining factor matrices \mathbf{A} and \mathbf{B} .

Several works investigated the uniqueness of the CPD. One of the earliest conditions was obtained by Kruskal:

Theorem 1. [48], [49] The CPD of an order-3 tensor $\mathcal{X} = [\mathbf{A}, \mathbf{B}, \mathbf{C}]$ with rank R is essentially unique if

$$\text{kr}(\mathbf{A}) + \text{kr}(\mathbf{B}) + \text{kr}(\mathbf{C}) \geq 2R + 2. \quad (5)$$

Note that Kruskal’s condition is sufficient but not necessary for a CPD to be unique. Recent work has investigated more relaxed uniqueness conditions [50]. In particular, it has been shown that conditions for the CPD of generic tensors to be unique can be mild, especially for tensors that are “tall” in at least one of the modes [51]. We also note that CPD can be computed algebraically even when the factor matrices are not full rank [52]. We refer to a tensor which satisfies Kruskal’s uniqueness condition as *fully unique* in order to specify the case in which all factor matrices (from the three modes) are unique up to permutation and scaling ambiguities and to clearly distinguish it from the uni-mode uniqueness condition discussed in the following.

C. Coupled decompositions and uni-mode uniqueness

Given a set of tensors $\mathcal{X}_t, t = 1, \dots, T$, coupled decompositions represent each \mathcal{X}_t using a low-rank model but constrain some of the factors to be the same for different t . For instance, a coupled PD of $\{\mathcal{X}_t\}_t$ can be expressed as

$$\mathcal{X}_t = [\mathbf{A}_t, \mathbf{B}_t, \mathbf{C}], \quad t = 1, \dots, T, \quad (6)$$

where mode-1 and mode-2 factors $\mathbf{A}_t \in \mathbb{R}^{I_t \times R}$ and $\mathbf{B}_t \in \mathbb{R}^{J_t \times R}$ are individual to each tensor \mathcal{X}_t , while the mode-3 factor $\mathbf{C} \in \mathbb{R}^{K \times R}$ is shared among all tensors.

The uniqueness of decompositions in the form of (6) was studied in [38], [39]. It was shown that conditions for the coupled decomposition to be unique can be much milder than those guaranteeing the uniqueness of every tensor. This is an important advantage of coupled decompositions, which can be unique even when some of $\{\mathcal{X}_t\}_t$ are matrices [2].

Such relaxed uniqueness conditions are strongly related to the so-called *uni-mode uniqueness* of tensors, which guarantee the unique recovery of only a single factor of the decomposition. Such results will be important to guarantee uniqueness of the coupled decompositions later in the paper. The uni-mode uniqueness of the CPD has been studied in [45], [46]. We recall the condition obtained in [45]:

Theorem 2. [45, Theorem 2.1] *Given the CP decomposition of an order-3 tensor $\mathcal{X} = [\mathbf{A}, \mathbf{B}, \mathbf{C}]$ with rank R , if \mathbf{C} has no zero columns and*

$$\text{kr}(\mathbf{A}) + \text{kr}(\mathbf{B}) + \text{rank}(\mathbf{C}) \geq 2R + 2 \quad (7)$$

holds, then the third factor matrix \mathbf{C} is essentially unique.

Other conditions for the uni-modal and full uniqueness of the CPD were later studied considering the compound matrix framework in [46], [50]. In particular, the uni-mode uniqueness condition given in Expression 1.12 of [46] shows that, under some mild additional conditions on factors \mathbf{A} and \mathbf{B} , the condition (7) from Theorem 2 can be substituted by the alternative condition $\text{rank}(\mathbf{C}) + \min(\text{kr}(\mathbf{A}), \text{kr}(\mathbf{B})) \geq R + 1$.

The uni-mode uniqueness of the CPD plays a fundamental role in the uniqueness of coupled decomposition models. Also, the uni-mode uniqueness results for a single CPD have been extended to the coupled case (6) to provide more relaxed conditions to recover the factor \mathbf{C} [38, Theorem 4.6]. Moreover, the coupled CPD (6) can be transformed into an LL1 BTD model (see [38, Section 4.3]). Thus, (6) inherits the uniqueness of the LL1 BTD [53], which guarantees the recovery of \mathbf{C} . These results on the uniqueness of \mathbf{C} serve as the basis to derive uniqueness guarantees for the coupled CPD (6) even when none of the individual tensors is unique [38]. Thus, uni-mode uniqueness has a significant impact in the development of recovery conditions in data fusion. In particular, it is possible to recover a “high-resolution” tensor from the coupled decomposition of multiple “low-resolution” ones, where the low-resolution tensors only need to ensure the uniqueness of a high-resolution factor matrix by being uni-mode unique, but not necessarily fully unique.

III. PERSONALIZED COUPLED TENSOR DECOMPOSITION

We consider the problem of recovering a common order-3 tensor $\mathbf{C} \in \mathbb{R}^{M_1 \times M_2 \times M_3}$ and distinct order-3 tensors $\mathbf{D}_k \in \mathbb{R}^{N_{k,1} \times N_{k,2} \times N_{k,3}}$ from a set of possibly degraded measurements $\mathbf{Y}_k \in \mathbb{R}^{N_{k,1} \times N_{k,2} \times N_{k,3}}$, $k = 1, \dots, K$, which are acquired according to the following measurement model:

$$\mathbf{Y}_k = \mathcal{P}_k(\mathbf{C}) + \mathbf{D}_k, \quad (8)$$

where \mathcal{P}_k is a separable operator of the form:

$$\mathcal{P}_k(\mathbf{C}) = \mathbf{C} \times_1 \mathbf{P}_{k,1} \times_2 \mathbf{P}_{k,2} \times_3 \mathbf{P}_{k,3}, \quad (9)$$

with $\mathbf{P}_{k,j} \in \mathbb{R}^{N_{k,j} \times M_j}$ being matrices of rank $\min\{N_{k,j}, M_j\}$ which describe the way in which the j -th mode of \mathbf{C} is measured in (or contributes to) \mathbf{Y}_k . The common tensor \mathbf{C} represents information/content that is shared among all datasets, linking them together, while the distinct tensors \mathbf{D}_k represent information unique to each dataset. The model in (8) and (9) is illustrated in Figures 1 and 2.

This general model can be related to different applications. For instance, in multimodal image fusion, tensors \mathbf{Y}_k can represent multichannel images acquired from the same scene, \mathbf{C} is a latent high-resolution image, operators $\mathbf{P}_{k,j}$ represent spatial and spectral degradations during image acquisition, and \mathbf{D}_k contains inter-image variations/changes [8], [10], [11]. When \mathbf{Y}_k represents fMRI data from multiple tasks, \mathbf{C} contains functional brain networks that are active during data acquisition for all tasks, whereas \mathbf{D}_k contains task-specific functional networks [13]. This model could also be used, e.g., to separate images of faces from multiple subjects (represented in \mathbf{Y}_k) represent common (\mathbf{C}) and subject specific (\mathbf{D}_k) components as in [30], or to perform fusion of EEG and fMRI data (represented in \mathbf{Y}_k) accounting for components that might be present only in a single data modality (\mathbf{D}_k) [4], [29].

Our aim in this work is to recover \mathbf{C} and $\{\mathbf{D}_k\}_k$ from $\{\mathbf{Y}_k\}_k$ by considering a low-rank CPD tensor model for \mathbf{C} and \mathbf{D}_k , which is described in the following. Assuming that the common and distinct components admit CPDs, we can write the tensors in (8) and (9) as:

$$\mathbf{C} = [\mathbf{C}_1, \mathbf{C}_2, \mathbf{C}_3], \quad (10)$$

$$\mathbf{D}_k = [\mathbf{D}_{k,1}, \mathbf{D}_{k,2}, \mathbf{D}_{k,3}], \quad (11)$$

$$\mathcal{P}_k(\mathbf{C}) = [\mathbf{P}_{k,1}\mathbf{C}_1, \mathbf{P}_{k,2}\mathbf{C}_2, \mathbf{P}_{k,3}\mathbf{C}_3], \quad (12)$$

where $\mathbf{C}_j \in \mathbb{R}^{M_j \times R}$ and $\mathbf{D}_{k,j} \in \mathbb{R}^{N_{k,j} \times L_k}$, $j \in \{1, 2, 3\}$ are the factor matrices. Note that the model in (12) consists in a polyadic decomposition with linear constraints on the factor matrices. This is directly connected to the CANDELINC decomposition [54]. However, unlike the traditional CANDELINC, we do not require matrices $\mathbf{P}_{k,j}$ to be full column rank. In such cases, recovery of the common factors \mathbf{C}_j can still be achieved in the present case due to the coupling between the different tensors. The tensor \mathbf{Y}_k also admits a polyadic decomposition with factor matrices equal to

$$[\mathbf{P}_{k,j}\mathbf{C}_j, \mathbf{D}_{k,j}], \quad (13)$$

for $j \in \{1, 2, 3\}$. Note, however, that although \mathbf{C} and \mathbf{D}_k are assumed to admit CPDs, the polyadic decomposition of \mathbf{Y}_k with the factor matrices given in (13) is not necessarily canonical, since the number of components might not be minimal. This model is fairly general, and encompasses as particular cases several applications as illustrated by the following examples.

Example 1. (fMRI decomposition) *Multi-subject fMRI data of the brain can be represented in the form of spatial feature maps, which are obtained by performing voxel-wise regression of the raw time-series data [55]. Such feature representations can be ordered as tensors, whose modes correspond to the spatial dimensions (voxels), to the subject index, and to sets*

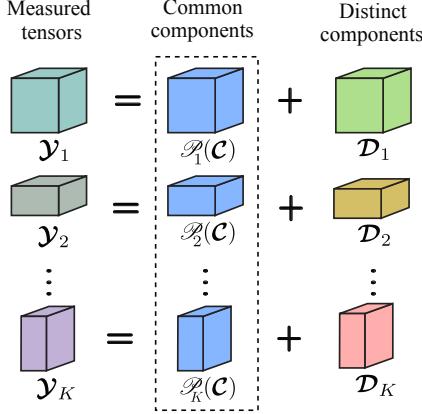


Figure 1. Illustration of the model in (8). A latent tensor \mathcal{C} is common in all measurements and acquired through an operator \mathcal{P}_k while tensors \mathcal{D}_k are distinct to each measurement, leading to a personalized decomposition.

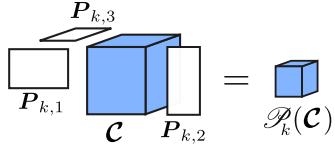


Figure 2. Illustration of the multilinear measurement model in (9).

of features corresponding to different data acquisitions within a task. By acquiring data while subjects are performing certain tasks (e.g., recognizing number sequences displayed on a screen, or sound tunes), different tensors can be constructed, one for each task. Then, by decomposing them according to the model (8), (9), with the matrices $\mathbf{P}_{k,j}$ being identity matrices, it is possible to recover the brain activations which are common to all tasks in \mathcal{C} , and those which are specific to each of the tasks in \mathcal{D}_k [13].

Remark 1. Although the fMRI voxels are matricized in Example 1 (i.e., they are considered a single mode of the tensor), each of the three spatial dimensions of the fMRI volume can be considered as additional tensor modes, as investigated in [56]. Although this would require extending the proposed model to higher order tensors (the measurements would have order 5), it allows the use of more general low-rank models which can reduce the amount of unknowns to be estimated.

Example 2. (MRI fusion) A high-resolution MRI tensor \mathcal{C} (whose modes represent the three spatial dimensions) can be computed efficiently by fusing three acquisitions, each with a smaller resolution (i.e., a smaller number of slices) in one of the modes, which can make acquisition time smaller [57], [58]. The acquisition model can be written as in (8), (9), in which $\mathbf{P}_{k,j}$ consist in either the identity matrix (for modes observed in full resolution), or downsampling and reweighting matrices (for modes observed in lower resolution). The variability between the acquisitions, represented in \mathcal{D}_k , can originate from patient motion, which might not be perfectly compensated between the different acquisitions [59].

Example 3. (Spectral image fusion) Spectral images can be represented as order-3 tensors, with two modes corresponding

to spatial dimensions (pixels), and one mode to the spectral dimension (bands). A high resolution image \mathcal{C} can be recovered by fusing two images, one with low spectral resolution (an MSI), and one with low spatial resolution (an HSI) [9]. These images can be represented as degraded versions of \mathcal{C} according to model (8), (9), where matrices $\mathbf{P}_{k,j}$ are either identity matrices (for the modes observed with high resolution), or degradation matrices represent the spatial or spectral responses of the sensors in a given mode. Inter-image variability between the different acquisitions, represented in \mathcal{D}_k , originates from illumination or other acquisition differences between the images [10], [60].

The model in (8)–(12) considers that all factors of \mathcal{C} are coupled among the different datasets through the operators \mathcal{P}_k . However, when fusing heterogeneous datasets such as EEG and fMRI data [61] or simulated and real metabolomics datasets [62], where some tensor modes such as the time dimension might not be necessarily aligned among datasets, a more flexible approach in which such modes of the tensor are left “un-coupled” (i.e., the coupling is enforced only in one or two of the modes) might be desirable. We will introduce such additional flexibility when presenting the optimization-based CTD framework later in Section V-B.

In the following, we first study the uniqueness of this decomposition in Section IV. Then, in Section V we develop two algorithms to compute it, one being semi-algebraic, and another being based on an alternating optimization approach.

IV. UNIQUENESS OF THE DECOMPOSITION

In this section, we provide both deterministic and generic conditions under which the proposed coupled decomposition model is unique, allowing the recovery of the common and distinct tensors \mathcal{C} and $\{\mathcal{D}_k\}_k$ from the measurements $\{\mathcal{Y}_k\}_k$. Those results will also motivate the development of a semi-algebraic algorithm to compute the decomposition.

A. Deterministic recoverability conditions

Considering the uniqueness and uni-mode uniqueness results for the CPD presented in Section II-C, the following result concerning the recoverability of the proposed coupled CPD-based model from Section III can be stated.

Theorem 3. Consider the measurement model (8), (9) and the CPD model (10), (11), (12), (13). Then, if the following conditions are satisfied:

- A1: There exists an $\eta \in [K]$ such that \mathcal{Y}_η is fully unique.
- A2: For each $j \in [3]$ there exists $\xi_j \in [K]$ such that \mathcal{Y}_{ξ_j} is mode- j unique and $\mathbf{P}_{\xi_j,j}$ has full column rank.
- A3: There exists at least one $j \in [3]$ such that $\xi_j \neq \eta$. Moreover, for all such j , the Kruskal rank of matrix $[\mathbf{P}_{\eta,j}\mathbf{C}_j, \mathbf{P}_{\eta,j}\mathbf{P}_{\xi_j,j}^\dagger\mathbf{D}_{\xi_j,j}, \mathbf{D}_{\eta,j}]$ is bigger than one.

then the common and distinct tensors \mathcal{C} and $\{\mathcal{D}_k\}_k$ can be uniquely recovered from $\{\mathcal{Y}_k\}_k$.

Proof. Consider a decomposition for the coupled CPD model in (8)–(13) given by $\{\mathbf{C}_j, \mathbf{D}_{k,j}\}_{k,j}$. Suppose that there exists

an alternative decomposition $\{\hat{\mathbf{C}}_i, \hat{\mathbf{D}}_{k,i}\}_{i,k}$ that also satisfies equations (8)–(13).

Step 1 (linking the recovered factors). Let us take a $j \in [3]$ satisfying $\xi_j \neq \eta$. Such a j is guaranteed to exist due to hypothesis A3. The uniqueness of \mathbf{Y}_η and mode- j uniqueness of \mathbf{Y}_{ξ_j} from hypotheses A1 and A2 mean that their decompositions $\{\hat{\mathbf{C}}_i, \hat{\mathbf{D}}_{\eta,i}\}_i$ and $\{\hat{\mathbf{C}}_j, \hat{\mathbf{D}}_{\xi_j,j}\}_j$ satisfy:

$$[\mathbf{P}_{\eta,i} \hat{\mathbf{C}}_i, \hat{\mathbf{D}}_{\eta,i}] = [\mathbf{P}_{\eta,i} \mathbf{C}_i, \mathbf{D}_{\eta,i}] \Pi_\eta \Lambda_{\eta,i}, \quad (14)$$

which holds for $i \in [3]$, and

$$[\mathbf{P}_{\xi_j,j} \hat{\mathbf{C}}_j, \hat{\mathbf{D}}_{\xi_j,j}] = [\mathbf{P}_{\xi_j,j} \mathbf{C}_j, \mathbf{D}_{\xi_j,j}] \Pi_{\xi_j} \Lambda_{\xi_j,j}. \quad (15)$$

where Π_{ξ_j} and Π_η are permutation matrices, and $\Lambda_{\xi_j,j}$ and $\Lambda_{\eta,i}$ are diagonal matrices, with $\Lambda_{\eta,1} \Lambda_{\eta,2} \Lambda_{\eta,3} = \mathbf{I}$.

Moreover, hypothesis A2 means that $\mathbf{P}_{\xi_j,j}$ has full column rank. This allows us to relate the common part of the mode- j factor matrices of the tensor in measurements ξ_j and η as

$$\mathbf{P}_{\eta,j} \hat{\mathbf{C}}_j = [\mathbf{P}_{\eta,j} \mathbf{P}_{\xi_j,j}^\dagger] \mathbf{P}_{\xi_j,j} \hat{\mathbf{C}}_j. \quad (16)$$

Using (14) and (15) on the left- and right-hand side of equation (16), respectively, we obtain:

$$(\mathbf{P}_{\eta,j} \mathbf{C}_j \Pi_\eta^{(1)} + \mathbf{D}_{\eta,j} \Pi_\eta^{(2)}) \Lambda_{\eta,j}^{(1)} \quad (17)$$

$$= \mathbf{P}_{\eta,j} \mathbf{P}_{\xi_j,j}^\dagger (\mathbf{P}_{\xi_j,j} \mathbf{C}_j \Pi_{\xi_j}^{(1)} + \mathbf{D}_{\xi_j,j} \Pi_{\xi_j}^{(2)}) \Lambda_{\xi_j,j}^{(1)} \quad (18)$$

$$= (\mathbf{P}_{\eta,j} \mathbf{C}_j \Pi_{\xi_j}^{(1)} + \mathbf{P}_{\eta,j} \mathbf{P}_{\xi_j,j}^\dagger \mathbf{D}_{\xi_j,j} \Pi_{\xi_j}^{(2)}) \Lambda_{\xi_j,j}^{(1)}, \quad (19)$$

where

$$\Pi_\rho = \begin{bmatrix} \Pi_\rho^{(1)} & \Pi_\rho^{(3)} \\ \Pi_\rho^{(2)} & \Pi_\rho^{(4)} \end{bmatrix}, \quad \Lambda_{\rho,j} = \begin{bmatrix} \Lambda_{\rho,j}^{(1)} & \mathbf{0} \\ \mathbf{0} & \Lambda_{\rho,j}^{(2)} \end{bmatrix}, \quad (20)$$

denote the partitions of Π_ρ and $\Lambda_{\rho,j}$ into blocks of appropriate dimensions, for $\rho \in \{\xi_j, \eta\}$. Matrices $\Pi_\rho^{(i)}$, $i = 1, \dots, 4$ are blocks of a permutation matrix, thus having at most one element equal to one in each row or column, with the remaining ones being equal to zero.

Step 2 (solving the ambiguities). In order to proceed, we need to show that (19) yields a permutation that allows for the correct recovery of the common part. We can now show that the permutations will be block diagonal, and will not mix the common and distinct parts of the factor matrices. Again, let us consider a $j \in [3]$ satisfying hypothesis A3, such that $\xi_j \neq \eta$. Note that the columns of the matrices in the left- and right-hand side of (17)–(19) can be expressed as:

$$\alpha(\mathbf{a} + \mathbf{b}) = \beta(\mathbf{a}' + \mathbf{c}) \neq \mathbf{0}, \quad (21)$$

where \mathbf{a} and \mathbf{b} are either the zero vector or columns of $\mathbf{P}_{\eta,j} \mathbf{C}_j$ and $\mathbf{D}_{\eta,j}$, respectively, and, similarly, \mathbf{a}' and \mathbf{c} are either the zero vector or columns of $\mathbf{P}_{\eta,j} \mathbf{C}_j$ and $\mathbf{P}_{\eta,j} \mathbf{P}_{\xi_j,j}^\dagger \mathbf{D}_{\xi_j,j}$. Moreover, since Π_ρ is a permutation matrix, only one vector at each side of (21) is different from zero; thus, one vector among $\{\mathbf{a}, \mathbf{b}\}$, and one vector among $\{\mathbf{a}', \mathbf{c}\}$ is equal to zero. Scalars α and β are scalings related to $\Lambda_{\eta,j}^{(1)}$ and $\Lambda_{\xi_j,j}^{(1)}$ and different from zero.

Note that (21) implies that $\mathbf{a} + \mathbf{b}$ and $\mathbf{a}' + \mathbf{c}$ are proportional. However, since the Kruskal rank of

$[\mathbf{P}_{\eta,j} \mathbf{C}_j, \mathbf{P}_{\eta,j} \mathbf{P}_{\xi_j,j}^\dagger \mathbf{D}_{\xi_j,j}, \mathbf{D}_{\eta,j}]$ is bigger than one due to hypothesis A3, no two columns of the matrix are proportional, which implies that (21) can only be satisfied if $\mathbf{b} = \mathbf{c} = \mathbf{0}$. Since this holds for all columns of (17)–(19), we have that $\Pi_\eta^{(2)} = \mathbf{0}$ and $\Pi_{\xi_j}^{(2)} = \mathbf{0}$. This implies that the blocks $\Pi_\eta^{(1)}$ and $\Pi_{\xi_j}^{(1)}$ are themselves permutation matrices, and that Π_{ξ_j} and Π_η are block diagonal.

Step 3 (linking the three modes). Note that \mathbf{Y}_η is fully unique due to hypothesis A1, and the above results showed that Π_η is block diagonal. We can also show that Π_{ξ_j} has a block diagonal structure for $j \in [3]$, not mixing common and distinct parts of the factor matrices. Specifically, if $\xi_j \neq \eta$, then by the previous arguments we have directly that Π_{ξ_j} is block diagonal with the desired structure. Otherwise, if $\xi_j = \eta$, then obviously $\Pi_{\xi_j} = \Pi_\eta$, thus being block diagonal with the same structure. It remains to show that the permutations and scalings are consistent across the three modes.

To proceed, consider ξ_j , for each $j \in [3]$. Due to the block diagonal structure of Π_{ξ_j} , the mode- j uniqueness in (15), the coupling between η and ξ_j in mode j , we have

$$\mathbf{P}_{\xi_j,j} \hat{\mathbf{C}}_j = \mathbf{P}_{\xi_j,j} \mathbf{C}_j \Pi_{\xi_j}^{(1)} \Lambda_{\xi_j,j}^{(1)}, \quad (22)$$

which, using the full column rank of $\mathbf{P}_{\xi_j,j}$, leads to the following condition:

$$\hat{\mathbf{C}}_j = \mathbf{C}_j \Pi_{\xi_j}^{(1)} \Lambda_{\xi_j,j}^{(1)}. \quad (23)$$

Due to the uniqueness of \mathbf{Y}_η , we can use the block diagonal structure of Π_η with equations (14) and (23) to obtain

$$\begin{aligned} \mathbf{P}_{\eta,j} \hat{\mathbf{C}}_j &= \mathbf{P}_{\eta,j} \mathbf{C}_j \Pi_\eta^{(1)} \Lambda_{\eta,j}^{(1)} \\ &= \mathbf{P}_{\eta,j} \mathbf{C}_j \Pi_{\xi_j}^{(1)} \Lambda_{\xi_j,j}^{(1)}. \end{aligned} \quad (24)$$

Note that due to necessary conditions for the uniqueness of the CPD, \mathbf{Y}_η being unique implies that $\mathbf{P}_{\eta,j} \mathbf{C}_j$ does not have proportional columns [50]. Since $\Lambda_{\xi_j,j}^{(1)}$ and $\Lambda_{\eta,j}^{(1)}$ are diagonal and $\Pi_\eta^{(1)}$ and $\Pi_{\xi_j}^{(1)}$ are permutation matrices, this means that equation (24) is only satisfied when $\Pi_{\xi_j}^{(1)} = \Pi_\eta^{(1)}$ and $\Lambda_{\xi_j,j}^{(1)} = \Lambda_{\eta,j}^{(1)}$. This holds for all $j \in [3]$. Thus, since $\Lambda_{\eta,1}^{(1)} \Lambda_{\eta,2}^{(1)} \Lambda_{\eta,3}^{(1)} = \mathbf{I}$, the reconstructed tensor satisfies

$$\begin{aligned} \hat{\mathbf{C}} &= [\hat{\mathbf{C}}_1, \hat{\mathbf{C}}_2, \hat{\mathbf{C}}_3] \\ &= [\mathbf{C}_1 \Pi_\eta^{(1)} \Lambda_{\eta,1}^{(1)}, \mathbf{C}_2 \Pi_\eta^{(1)} \Lambda_{\eta,2}^{(1)}, \mathbf{C}_3 \Pi_\eta^{(1)} \Lambda_{\eta,3}^{(1)}] \\ &= \mathbf{C}, \end{aligned} \quad (25)$$

which proves the recovery of \mathbf{C} . The recovery of \mathbf{D}_k follows directly from (8), which concludes the proof. \square

Remark 2. Note that the conditions in Theorem 3 also imply the uniqueness of the factors of the common tensor \mathbf{C} . Moreover, conditions for the uniqueness of the factors of the distinct tensors \mathbf{D}_k can be obtained by analyzing each tensor separately using Kruskal's condition, since $\mathbf{D}_k = \mathbf{Y}_k - \mathcal{P}_k(\mathbf{C})$.

The conditions in Theorem 3 are mild and interpretable. Condition A1 is equivalent to one of the measured tensors \mathbf{Y}_η being fully unique. However, it does not require its factors to be measured in full resolution (i.e., $\mathbf{P}_{\eta,j}$ need not be

full column rank). The full uniqueness of one tensor is used to ensure consistent permutations and scalings among the recovered common factors in the three modes.

Intuitively, condition A2 means that for each mode $j \in [3]$ we need one of the measured tensors to be mode- j unique and to have the corresponding degradation matrix $\mathbf{P}_{i,j}$ with full column rank. This means that the mode- j factor matrix (containing both the common and distinct components) can be recovered in full resolution from this tensor.

Finally, condition A3 is key to allow us to relate factor matrices containing both common and distinct components from two different measured tensors in order to find out which of the components are the common ones.

B. Generic recoverability conditions

Based on Theorem 3 we can also derive generic uniqueness results which give explicit conditions based on the rank of the decomposition (i.e., the ranks of the common and distinct tensors, given by R and L_k , respectively), on the measurement matrices $\mathbf{P}_{k,j}$ and on the tensor dimensions that guarantee that a generic CTD following the model presented in Section III is essentially unique. This result is given in the following theorem.

Theorem 4. *Consider the measurement model (8), (9) and the CPD model (10), (11), (12), (13). Assume that all factors $\{\mathbf{C}_j, \mathbf{D}_{k,j}\}_{k,j}$ are drawn from some joint absolutely continuous distribution w.r.t. the Lebesgue measure. Then, if the following conditions are satisfied:*

A4: (uniqueness of η): *There exists an $\eta \in [K]$ such that*

$$\sum_{j \in [3]} \min\{\text{rank}(\mathbf{P}_{\eta,j}), R + L_\eta\} \geq 2(R + L_\eta) + 2. \quad (26)$$

A5: (uni-mode uniqueness of ξ_j): *For each $j \in [3]$ there exist a $\xi_j \in [K]$ such that $\mathbf{P}_{\xi_j,j}$ is full column rank and the following condition is satisfied:*

$$\min(N_{\xi_j,j}, \min(M_j, R) + L_{\xi_j}) + \sum_{i \neq j} \min\{\text{rank}(\mathbf{P}_{\xi_j,i}), R + L_{\xi_j}\} \geq 2(R + L_{\xi_j}) + 2. \quad (27)$$

A6: *There exists at least one $j \in [3]$ such that $\xi_j \neq \eta$. then the common and distinct tensors \mathbf{C} and $\{\mathbf{D}_k\}_k$ can be uniquely recovered from $\{\mathbf{Y}_k\}_k$ with probability one.*

Before proceeding with the proof of Theorem 4, we need a couple auxiliary results, provided in the following lemmas.

Lemma 1. *Let the elements of matrices $\mathbf{X}_1 \in \mathbb{R}^{R_1 \times S_1}, \dots, \mathbf{X}_K \in \mathbb{R}^{R_K \times S_K}$ be drawn according to a joint distribution that is absolutely continuous with respect to the Lebesgue measure in $\mathbb{R}^{R_1 S_1 \dots R_K S_K}$, and matrices $\mathbf{Q}_1 \in \mathbb{R}^{T \times R_1}, \dots, \mathbf{Q}_K \in \mathbb{R}^{T \times R_K}$ be deterministic. If there exist a full row rank matrix $\mathbf{E} \in \mathbb{R}^{\tilde{T} \times T}$ such that matrices $\mathbf{EQ}_1, \dots, \mathbf{EQ}_K$ also have full row rank, then the matrix $\mathbf{W} = [\mathbf{Q}_1 \mathbf{X}_1, \dots, \mathbf{Q}_K \mathbf{X}_K]$ has Kruskal rank at least $\min\{\tilde{T}, \sum_k R_k\}$ with probability one.*

Proof. Let us denote by $\{w_1, \dots, w_i\}$ a subset of i columns of matrix \mathbf{W} . Due to matrix \mathbf{E} having full row rank, if

$\{\mathbf{Ew}_1, \dots, \mathbf{Ew}_i\}$ are linearly independent vectors, this implies that $\{w_1, \dots, w_i\}$ will also be linearly independent. Since this holds for any subset of columns of \mathbf{W} , it implies that $\text{kr}(\mathbf{W}) \geq \text{kr}(\mathbf{EW})$.

We now aim to compute the Kruskal rank of $\mathbf{EW} = [\mathbf{EQ}_1 \mathbf{X}_1, \dots, \mathbf{EQ}_K \mathbf{X}_K]$. Since \mathbf{EQ}_i is assumed to be of full row rank for all i , this implies that if $\{\mathbf{X}_1, \dots, \mathbf{X}_K\}$ are jointly absolutely continuous, $\{\mathbf{EQ}_i \mathbf{X}_i\} \cup \{\mathbf{X}_j : j \neq i\}$ will be jointly absolutely continuous too (see Lemma 1 of [9]). Thus, noting that the above result holds when applying the transformation \mathbf{EQ}_i to each block of variables, we can proceed recursively for each of the blocks $i = 1, \dots, K$ to show that the elements of matrix \mathbf{EW} are also jointly absolutely continuous. This, on the other hand, implies that $\text{rank}(\mathbf{EW}) = \min(\tilde{T}, \sum_k R_k)$ with probability one. Since the rank and Kruskal rank of generic matrices coincide, the result follows. \square

Note that this lemma can be used to lower bound the generic Kruskal rank of the factor matrices of \mathbf{Y}_k .

Corollary 1. *Suppose \mathbf{C}_j and $\mathbf{D}_{k,j}$ are distributed according to a joint absolutely continuous measure, then*

$$\text{kr}([\mathbf{P}_{k,j} \mathbf{C}_j, \mathbf{D}_{k,j}]) \geq \min\{\text{rank}(\mathbf{P}_{k,j}), R + L_k\}.$$

Proof. Let us take $\mathbf{Q}_1 = \mathbf{P}_{k,j}$, $\mathbf{Q}_2 = \mathbf{I}$, and consider $\mathbf{E} \in \mathbb{R}^{\text{rank}(\mathbf{P}_{k,j}) \times N_{k,j}}$ to be the matrix which selects the $\text{rank}(\mathbf{P}_{k,j})$ linearly independent rows of $\mathbf{P}_{k,j}$. Clearly, $\text{rank}(\mathbf{E}) = \text{rank}(\mathbf{EQ}_1) = \text{rank}(\mathbf{EQ}_2) = \text{rank}(\mathbf{P}_{k,j})$. Applying Lemma 1 gives the desired result. \square

Lemma 2. *Consider $\mathbf{X}_1 \in \mathbb{R}^{M \times R}$ and $\mathbf{X}_2 \in \mathbb{R}^{N \times L}$ distributed according to a joint absolutely continuous measure, and $\mathbf{Q} \in \mathbb{R}^{N \times M}$ deterministic and full column rank. Then, for $\mathbf{Z} = [\mathbf{Q} \mathbf{X}_1, \mathbf{X}_2]$, $\text{rank}(\mathbf{Z}) = \min(N, \min(M, R) + L)$, with probability one.*

Proof. Consider the square matrix $\tilde{\mathbf{Q}} = [(\mathbf{Q}^\dagger)^\top, \mathbf{Q}_\perp^\top]^\top$, where \mathbf{Q}^\dagger is the left pseudoinverse of \mathbf{Q} and \mathbf{Q}_\perp its orthogonal complement. Since \mathbf{Q} is full column rank, $\tilde{\mathbf{Q}}$ is invertible, and thus $\text{rank}(\mathbf{Z}) = \text{rank}(\tilde{\mathbf{Z}})$, where

$$\tilde{\mathbf{Z}} = \tilde{\mathbf{Q}} \mathbf{Z} = \begin{bmatrix} \mathbf{X}_1 & \tilde{\mathbf{X}}_{2,1} \\ \mathbf{0} & \tilde{\mathbf{X}}_{2,2} \end{bmatrix}, \quad (28)$$

where $\tilde{\mathbf{X}}_{2,1} = \mathbf{Q}^\dagger \mathbf{X}_2$ and $\tilde{\mathbf{X}}_{2,2} = \mathbf{Q}_\perp \mathbf{X}_2$. Note that since $\tilde{\mathbf{Q}}$ is deterministic, \mathbf{X}_1 , $\tilde{\mathbf{X}}_{2,1}$, and $\tilde{\mathbf{X}}_{2,2}$ are jointly absolutely continuous. We will consider two cases depending on the shape of \mathbf{X}_1 .

Case $R \geq M$: In this case, \mathbf{X}_1 is generically full row rank, and thus by results related to the Schur complement [63, fact 6.5.6] the rank of the matrix is equal to

$$\text{rank}(\tilde{\mathbf{Z}}) = \text{rank}(\mathbf{X}_1) + \text{rank}(\tilde{\mathbf{X}}_{2,2}),$$

with the right hand side equal generically to $M + \min(N - M, L) = \min(N, M + L)$.

Case $R \leq M$: Denote $T = \min(N, R + L)$. We know that the rank is bounded by the dimensions of the matrix, i.e., $\text{rank}(\tilde{\mathbf{Z}}) \leq T$. To prove that the equality holds, we use the standard argument that is often used for proving generic

(holding with probability 1) properties [64]. We first show that there exists a matrix $\tilde{\mathbf{Z}}_0$ with the block-triangular form (28) that has rank T . Such a matrix can be constructed as

$$\tilde{\mathbf{Z}}_0 = \begin{bmatrix} \mathbf{I}_T & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}. \quad (29)$$

Note that this matrix is still in the same form as (28), although the dimensions of the blocks are different. Then the existence of such $\tilde{\mathbf{Z}}_0$ implies that $\text{rank}(\tilde{\mathbf{Z}}) = T$ with probability 1. Below we explain in detail why it is the case (although it can be also deduced from the semicontinuity of the rank function).

By construction, $\det([\tilde{\mathbf{Z}}_0]_{1:T, 1:T}) = 1 \neq 0$, which is a multivariate polynomial in $\tilde{\mathbf{Z}}$. Since the zero set (set of solutions) of a nonzero polynomial equation is necessarily of Lebesgue measure zero [65], we conclude that $\det([\tilde{\mathbf{Z}}]_{1:T, 1:T}) \neq 0$ with probability 1 for a matrix of the form (28), hence $\text{rank}(\tilde{\mathbf{Z}}) = T$ with probability 1. Combining the two cases gives the desired result. \square

Considering the results in Lemmas 1 and 2, as well as the deterministic uniqueness result in Theorem 3, we can proceed to prove the generic uniqueness case of Theorem 4.

Proof. **Step 1 (full uniqueness).** Let us consider the factor matrix of the j -th mode of the k -th tensor. Due to the joint absolute continuity of \mathbf{C}_j and $\mathbf{D}_{k,j}$, we can use Corollary 1 to show that the Kruskal rank of the j -th factor matrix of \mathbf{Y}_k is lower bounded as

$$\text{kr}([\mathbf{P}_{k,j} \mathbf{C}_j, \mathbf{D}_{k,j}]) \geq \min\{\text{rank}(\mathbf{P}_{k,j}), R + L_k\}$$

with probability one. Thus, for the tensor indexed by η , condition (26) of A4 leads to

$$\begin{aligned} \sum_{j \in [3]} \text{kr}([\mathbf{P}_{\eta,j} \mathbf{C}_j, \mathbf{D}_{\eta,j}]) &\geq \sum_{j \in [3]} \min\{\text{rank}(\mathbf{P}_{\eta,j}), R + L_\eta\} \\ &\geq 2(R + L_\eta) + 2, \end{aligned} \quad (30)$$

which implies that the CPD of \mathbf{Y}_η is essentially unique. This means that condition A1 of Theorem 3 is satisfied with probability one.

Step 2 (uni-mode uniqueness). Moreover, for each ξ_j , $j \in [3]$, for which $\mathbf{P}_{\xi_j,j}$ is full column rank due to hypothesis A5, we can use Lemma 2 to show that, with probability one, the rank of the j -th factor matrix of \mathbf{Y}_{ξ_j} is equal to

$$\text{rank}([\mathbf{P}_{\xi_j,j} \mathbf{C}_j, \mathbf{D}_{\xi_j,j}]) = \min(N_{\xi_j,j}, \min(M_j, R) + L_{\xi_j}),$$

and from Corollary 1 it has no zero columns. This, along with condition (27) of hypothesis A5, leads to

$$\begin{aligned} \text{rank}([\mathbf{P}_{\xi_j,j} \mathbf{C}_j, \mathbf{D}_{\xi_j,j}]) + \sum_{i \neq j} \text{kr}([\mathbf{P}_{\xi_j,i} \mathbf{C}_i, \mathbf{D}_{\xi_j,i}]) \\ \geq \min(N_{\xi_j,j}, \min(M_j, R) + L_{\xi_j}) \\ + \sum_{i \neq j} \min\{\text{rank}(\mathbf{P}_{\xi_j,i}), R + L_{\xi_j}\} \\ \geq 2(R + L_{\xi_j}) + 2, \end{aligned} \quad (31)$$

for all $j \in [3]$. Combining this result with Theorem 2 means that the mode- j factor matrix of \mathbf{Y}_{ξ_j} is essentially unique. This implies that condition A2 of Theorem 3 is satisfied with probability one.

Step 3 (auxiliary results to solve the ambiguities). Using hypothesis A6, let us choose a $\xi_j \in [K]$ satisfying $\xi_j \neq \eta$. Since $\text{rank}(\mathbf{P}_{\eta,j}) \geq 2$ due to assumption A4, we can find an $\mathbf{E} \in \mathbb{R}^{T \times N_{\eta,j}}$, $\tilde{T} \geq 2$ such that $\mathbf{E}\mathbf{P}_{\eta,j}$ has full row rank. Moreover, since $\mathbf{P}_{\xi_j,j}$ has full column rank due to assumption A5, $\text{rank}(\mathbf{P}_{\xi_j,j}^\dagger) = M_j$, and due to the Sylvester rank inequality, we have

$$\begin{aligned} \text{rank}(\mathbf{E}\mathbf{P}_{\eta,j}\mathbf{P}_{\xi_j,j}^\dagger) &\geq \text{rank}(\mathbf{E}\mathbf{P}_{\eta,j}) + \text{rank}(\mathbf{P}_{\xi_j,j}^\dagger) - M_j \\ &= \text{rank}(\mathbf{E}\mathbf{P}_{\eta,j}), \end{aligned} \quad (32)$$

which implies $\mathbf{E}\mathbf{P}_{\eta,j}\mathbf{P}_{\xi_j,j}^\dagger$ also has full row rank. Thus, due to the joint absolute continuity of $\mathbf{C}_j, \mathbf{D}_{\xi_j,j}, \mathbf{D}_{\eta,j}$ (a consequence of $\eta \neq \xi_j$), using Lemma 1 we have that with probability one the Kruskal rank of $[\mathbf{P}_{\eta,j} \mathbf{C}_j, \mathbf{P}_{\eta,j} \mathbf{P}_{\xi_j,j}^\dagger \mathbf{D}_{\xi_j,j}, \mathbf{D}_{\eta,j}]$ is bigger than one. This holds for any $\xi_j \neq \eta$. Thus, condition A3 of Theorem 3 is satisfied with probability one. Since A1, A2 and A3 are satisfied, Theorem 3 implies that tensors \mathbf{C} and $\{\mathbf{D}_k\}_k$ can be uniquely recovered from $\{\mathbf{Y}_k\}_k$ with probability one. \square

The generic uniqueness result in Theorem 4 allows us to more clearly illustrate how considering conditions based on uni-mode uniqueness can lead to more relaxed conditions for the uniqueness of the full CTD. This can be illustrated by the following example.

Example 4. Let us consider a common tensor \mathbf{C} , with size $7 \times 11 \times 9$, and distinct components \mathbf{D}_k , with CP ranks of $R = 5$ and $L_k = 5$ for all k , whose factor matrices are drawn from some joint continuous distribution. Suppose we have $K = 3$ measured tensors $\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3$, with sizes $10 \times 5 \times 7, 5 \times 12 \times 7$ and $5 \times 7 \times 10$, respectively. The measured tensors satisfy the model (8), (9) for some full rank matrices $\mathbf{P}_{k,j}$ of appropriate dimensions. This CTD is generically unique according to Theorem 4. However, only tensor \mathbf{Y}_2 is guaranteed to be fully unique, satisfying condition (26). Tensors \mathbf{Y}_1 and \mathbf{Y}_3 are only uni-mode unique, satisfying (27) for modes $j = 1$ and $j = 3$, respectively. While our result based on the uni-mode uniqueness allows us to consider ranks $R + L_k \leq 10$, if we were to require \mathbf{Y}_1 and \mathbf{Y}_3 to be fully unique instead (satisfying (26)), uniqueness would only hold for ranks satisfying $R + L_1 \leq 8$ and $R + L_3 \leq 9$.

The works [9], [34] consider the uniqueness of models related to the one in Section III. However, these results are not directly comparable to ours. Flexible coupled decompositions with shared and distinct components were considered in [34], but without a general measurement model like (8), (9). Moreover, the uniqueness results in [34] are based on algebraic and topological properties of matrices and a graph constructed from the measured tensors and factor matrices, whereas the conditions in Theorem 4 are directly based on the dimensions of the tensor and on the ranks of $\mathbf{P}_{k,j}$. The results in [9], which considers the spectral image fusion problem of Example 3, are easier to compare to ours. However, although both our results and those in [9] require one of the measured tensors to be fully unique, [9] does not consider distinct components (i.e., $\mathbf{D}_k = \mathbb{0}$) and exploits the fact that several $\mathbf{P}_{k,j}$ are identity

matrices. This leads to conditions for the recovery of the “high resolution” factor matrices that can be slightly weaker the uni-mode uniqueness used in our results.

V. COMPUTING THE DECOMPOSITION

This section presents algorithms to compute the coupled CP decomposition according to the model presented in Section III. Two approaches are proposed. The first, presented in Section V-A, is a semi-algebraic strategy, while the second, presented in Section V-B uses an optimization algorithm.

A. A semi-algebraic solution

Algebraic algorithms can provide accurate solutions in low-noise settings and constitute good quality initializations to optimization strategies [66]. Theorem 3 provides a means of devising a semi-algebraic algorithm. It requires the CPD of at least two and at most four of the measured tensors, and solutions to linear algebra and assignment problems, as will be detailed in the sequence. The approach is referred to as semi-algebraic because except for the computation of the CPDs, which might involve the use of optimization algorithms, the rest of the procedure is based on simple algebraic operations. To this end, let us assume the conditions in Theorem 3 to hold.

Step 0 (measurements selection): Select an $\eta \in [K]$ and $\xi_j \in [K]$, $j \in [3]$ such that \mathcal{Y}_η is fully unique, \mathcal{Y}_{ξ_j} is mode- j unique, $\mathbf{P}_{\xi_j,j}$ has full column rank, and the cardinality of the set $\{\eta, \xi_1, \xi_2, \xi_3\}$ is at least two. This ensures there exists a $\xi_j \neq \eta$. Note that there can be several possible choices of η and ξ_j satisfying these conditions, introducing a measure of user choice. First, since we will compute CPDs of the tensors indexed by $\{\eta, \xi_1, \xi_2, \xi_3\}$, the smaller the cardinality of this set the smaller the number of CPDs to compute. Moreover, we can also prioritize choices of η or ξ_j for tensors that are easier to decompose. For example, for tensors of rank $R + L_k$, typically the larger their dimensions $N_{k,j}$ the easier it is to compute its decomposition accurately (i.e., larger tensors might have less collinearity between the columns of the factor matrices, reducing convergence issues due to “swamping” [67]). In addition, if \mathcal{Y}_k are contaminated by noise, we might prioritize measurements with higher signal-to-noise ratio (SNR).

Step 1 (recovery of factors separately): Compute the (unique) rank- $R + L_\eta$ CPD of \mathcal{Y}_η . Denote the recovered factor matrices by $\mathbf{U}_{\eta,i}$, $i \in [3]$.

Step 2 (recover a different uni-mode factor): Pick j such that $\xi_j \neq \eta$. Since \mathcal{Y}_{ξ_j} is mode- j unique by assumption, compute its CPD with rank $R + L_{\xi_j}$, and denote the recovered (unique) mode- j factor matrix by $\mathbf{U}_{\xi_j,j}$, and the remaining factors (not necessarily unique) by $\mathbf{U}_{\xi_j,i}$, $i \neq j$.

Step 3 (correcting ambiguities): We find the common part of the factor by searching for an optimal assignment between the transformed columns of $\mathbf{U}_{\eta,j}$ and $\mathbf{U}_{\xi_j,j}$. More precisely, we first map the factors to the same column space using the transformation

$$\mathbf{V}_{\xi_j,j} = \mathbf{P}_{\eta,j} \mathbf{P}_{\xi_j,j}^\dagger \mathbf{U}_{\xi_j,j},$$

recalling that $\mathbf{P}_{\xi_j,j}$ is left-invertible by assumption. Now we find an optimal correspondence between R columns of

matrices $\mathbf{V}_{\xi_j,j}$ and $\mathbf{U}_{\eta,j}$. To do this, we first construct a matrix $\mathbf{Z} \in \mathbb{R}^{R+L_\eta \times R+L_{\xi_j}}$ with the normalized inner product between each pairs of columns form these matrices, that is,

$$Z_{n,m} = \frac{|\langle \mathbf{u}_n, \mathbf{v}_m \rangle|}{\|\mathbf{u}_n\|_2 \|\mathbf{v}_m\|_2}, \quad (33)$$

where \mathbf{u}_n is the n -th column of $\mathbf{U}_{\eta,j}$, and \mathbf{v}_m is the m -th column of $\mathbf{V}_{\xi_j,j}$. Then, we find a partial permutation that best aligns the common columns among these matrices, as the binary matrix $\Delta \in \{0, 1\}^{R+L_\eta \times R+L_{\xi_j}}$ that solves:

$$\max_{\Delta_{r,s} \in \{0,1\}} \sum_{r,s} \Delta_{r,s} Z_{r,s} \quad (34a)$$

$$\text{s.t. } \mathbf{1}^\top \Delta \mathbf{1} = R, \quad \Delta \mathbf{1} \leq \mathbf{1}, \quad \Delta^\top \mathbf{1} \leq \mathbf{1}. \quad (34b)$$

Note that this is an unbalanced assignment problem with fixed cardinality R , which can be solved efficiently [68].

From the solution $\hat{\Delta}$ to problem (34) we define two matrices: \mathbf{M}_η , equal to the identity matrix with the columns corresponding to the indices of the all-zero rows of $\hat{\Delta}$ removed (size $R \times R + L_\eta$), and \mathbf{M}_{ξ_j} , which is the matrix $\hat{\Delta}$ with the all-zero rows removed (of size $R \times R + L_{\xi_j}$). Now, we can select the common parts and correct the permutations for all factors of tensors η and ξ_j :

$$\hat{\mathbf{X}}_{\eta,i}^C = \mathbf{U}_{\eta,i} \mathbf{M}_\eta, \quad (35)$$

$$\hat{\mathbf{X}}_{\xi_j,i}^C = \mathbf{U}_{\xi_j,i} \mathbf{M}_{\xi_j}. \quad (36)$$

for $i \in \{1, 2, 3\}$. Finally, we compute $\hat{\mathbf{C}}_j$ by compensating the scaling ambiguities as $\hat{\mathbf{C}}_j = \mathbf{P}_{\xi_j,j}^\dagger \hat{\mathbf{X}}_{\xi_j,j}^C \mathbf{\Lambda}_j$, where $\mathbf{\Lambda}_j$ is the diagonal matrix solving

$$\mathbf{\Lambda}_j = \underset{\text{diagonal matrix } \mathbf{\Lambda}}{\arg \min} \quad \|\hat{\mathbf{X}}_{\eta,j}^C - \mathbf{P}_{\eta,j} \mathbf{P}_{\xi_j,j}^\dagger \hat{\mathbf{X}}_{\xi_j,j}^C \mathbf{\Lambda}\|_F. \quad (37)$$

This gives us the j -th common factor with permutation scaling ambiguities matching those of tensor η . The solution to (37) can be computed in a straightforward manner for each diagonal element of $\mathbf{\Lambda}_j$. Moreover, in a setting where $\hat{\mathbf{X}}_{\eta,j}^C$ and $\hat{\mathbf{X}}_{\xi_j,j}^C$ are not computed exactly, the solution to (37) will result in an approximate scaling that best matches the two factor matrices.

Step 4 (computing the remaining factors): Let us now consider the other modes, $\ell \neq j$. Let us consider the measurements available for which the remaining modes are uni-mode unique and observed with full resolution, $\{\xi_\ell, \ell \neq j\}$. Depending on the value of ξ_j and η , we have two cases:

- 1) If $\xi_\ell \in \{\eta, \xi_j\}$, then $\hat{\mathbf{X}}_{\xi_\ell,\ell}^C$ was computed in (35) or (36) and the mode- ℓ factor can be computed as $\hat{\mathbf{C}}_\ell = \mathbf{P}_{\xi_\ell,\ell}^\dagger \hat{\mathbf{X}}_{\xi_\ell,\ell}^C \mathbf{\Lambda}_\ell$, with $\mathbf{\Lambda}_\ell$ computed similarly to (37).
- 2) If $\xi_\ell \notin \{\eta, \xi_j\}$, then let us compute the rank $R + L_{\xi_\ell}$ CPD of \mathcal{Y}_{ξ_ℓ} , and denote the recovered mode- ℓ factor matrix by $\mathbf{U}_{\xi_\ell,\ell}$. Then, we repeat the same procedure as in the **Step 3**, to find the optimal assignment between the columns of matrices $\hat{\mathbf{C}}_j$ and $\mathbf{P}_{\xi_\ell,\ell}^\dagger \hat{\mathbf{X}}_{\xi_\ell,\ell}^C$, which will give us the corresponding matrices \mathbf{M}_{ξ_ℓ} and $\mathbf{\Lambda}_\ell$. This way, we can recover $\hat{\mathbf{C}}_\ell$ as $\hat{\mathbf{C}}_\ell = \mathbf{P}_{\xi_\ell,\ell}^\dagger \mathbf{U}_{\xi_\ell,\ell} \mathbf{M}_{\xi_\ell} \mathbf{\Lambda}_\ell$, which matches the scaling ambiguities. This gives us the ℓ -th common factor with permutation scaling ambiguities matching those of $\hat{\mathbf{C}}_j$.

Algorithm 1: Semi-algebraic algorithm

```

1 Inputs: Data  $\mathcal{Y}_k$ , operators  $\mathbf{P}_{k,j}$ , ranks  $R, L_k$ , indices of unique
   tensor  $\eta$  and mode- $i$  unique tensors  $\xi_i, i \in [3]$ .
2 Compute the CPD of  $\mathcal{Y}_\eta$ , with factors  $\mathbf{U}_{\eta,i}$ , for  $i \in [3]$ .
3 Take  $j \in [3]$  such that  $\xi_j \neq \eta$ , compute the mode- $j$  CPD factor
   matrix of  $\mathcal{Y}_{\xi_j}$ , denoted by  $\mathbf{U}_{\xi_j,j}$ .
4 Compute the optimal partial assignment between the columns of  $\mathbf{U}_\eta$ 
   and  $\mathbf{U}_{\xi_j}$  according to Step 3, solving (34)–(37), and compute the
   factor matrix  $\hat{\mathbf{C}}_j = \mathbf{P}_{\xi_j,j}^\dagger \hat{\mathbf{X}}_{\xi_j,j}^C \mathbf{\Lambda}_j$ .
5 for  $\ell \in \{1, 2, 3\} \setminus \{j\}$  do
6   if  $\xi_\ell \in \{\eta, \xi_j\}$  then
7      $\hat{\mathbf{C}}_\ell = \mathbf{P}_{\xi_\ell,\ell}^\dagger \hat{\mathbf{X}}_{\xi_\ell,\ell}^C \mathbf{\Lambda}_\ell$ .
8   else
9     Compute rank  $R + L_{\xi_\ell}$  CPD of  $\mathcal{Y}_{\eta_\ell}$  and denote the
     recovered mode- $\ell$  factor matrix by  $\mathbf{U}_{\xi_\ell,\ell}$ .
10    Compute the optimal partial assignment between the
     columns of  $\hat{\mathbf{C}}_j$  and  $\mathbf{P}_{\xi_\ell,\ell}^\dagger \hat{\mathbf{X}}_{\xi_\ell,\ell}^C$  and the corresponding
     matrices as in line 4, and compute the factor  $\mathbf{C}_\ell$ 
     according to Step 4.
11   end
12 end
13 return  $\mathbf{C}_{\text{alg}} = [\hat{\mathbf{C}}_1, \hat{\mathbf{C}}_2, \hat{\mathbf{C}}_3]$ ,  $\mathcal{D}_k = \mathcal{Y}_k - \mathbf{C}_{\text{alg}}, k \in [K]$ .

```

Step 5 (reconstruct the tensor): With the permutations and scalings compatible across the three modes, we can reconstruct the common tensor as $\hat{\mathcal{C}} = [\hat{\mathbf{C}}_1, \hat{\mathbf{C}}_2, \hat{\mathbf{C}}_3]$. The distinct components $\hat{\mathcal{D}}_k$ are recovered as the CPD of $\mathcal{Y}_k - \mathcal{P}_k(\hat{\mathcal{C}})$.

Note that, following the same arguments as in the proof of Theorem 3, it can be shown that this recovers the true tensor.

B. An optimization-based solution

Let us consider the case when the tensors are measured with noise. The CTD can be formulated as an optimization problem by seeking the decomposition satisfying the model (8), (10), (11) and (12) which is the closest to the measurements according to some criterion. However, the computational complexity can increase substantially when the number of measured tensors K becomes large. To allow the computational complexity to be reduced, we will propose a more flexible formulation in which one can control which factors of the common tensor are coupled to a given measurement in the optimization problem. This is related to factorization frameworks with both shared and unshared factors [1], [34]. To define this more precisely, let us introduce some additional notation. First, we can write the polyadic decomposition of tensor \mathcal{Y}_k as:

$$\mathcal{Y}_k = \sum_{m \in \{C, D\}} [\mathbf{X}_{k,1}^m, \mathbf{X}_{k,2}^m, \mathbf{X}_{k,3}^m], \quad (38)$$

where $\mathbf{X}_{k,j}^C$ and $\mathbf{X}_{k,j}^D$ are related to the factor matrices in (11) and (12) through:

$$\mathbf{X}_{k,j}^C = \mathbf{P}_{k,j} \mathbf{C}_j, \quad \mathbf{X}_{k,j}^D = \mathbf{D}_{k,j}. \quad (39)$$

Let $\Gamma_j \subseteq [K]$ for $j \in [3]$ be the sets defining which coupling constraints are enforced, i.e., $k \in \Gamma_j$ means that, for mode- j factor \mathbf{C}_j of the common tensor, the coupling (39) in the k -th measurement is included as a constraint in the optimization problem (leaving the possibility to relax some of the constraints). Having more flexibility in the definition

of couplings was shown to be useful in prior works such as in double coupled factorizations [20], [24], where the factor matrix of a given measurement tensor can be coupled to one among multiple common factors, instead of only a single global factor. The flexible coupled tensor decomposition considered in this work can be formulated as follows:

$$\min_{\Phi} \sum_{k=1}^K \left\| \mathcal{Y}_k - \sum_{m \in \{C, D\}} [\mathbf{X}_{k,1}^m, \mathbf{X}_{k,2}^m, \mathbf{X}_{k,3}^m] \right\|_F^2 \quad (40a)$$

$$\text{s.t. } \mathbf{X}_{k,j}^C = \mathbf{P}_{k,j} \mathbf{C}_j, \quad k \in \Gamma_j, \quad j \in [3], \quad (40b)$$

where $\Phi = \{\mathbf{C}_j, \mathbf{X}_{k,j}^C, \mathbf{X}_{k,j}^D\}$. Note that the factor matrices $\mathbf{D}_{k,j}$ are not included in the optimization problem since they are equal to $\mathbf{X}_{k,j}^D$ due to (39), and thus redundant. The constraint (40b) links the mode- j factor of the common tensor \mathbf{C} among all measured datasets indexed in Γ_j . Note that we do not require knowledge of $\mathbf{P}_{k,j}$ for $k \notin \Gamma_j$. We also remark that although the loss function in (40a) is based on the squared loss, other loss functions can also be considered, such as the Kullback-Leibler divergence or the logistic loss, which can be more natural for count or binary tensors (see, e.g., [69]–[71]).

Remark 3. Note that following the same reasoning as for the semi-algebraic algorithm developed in Section V-A, supposing the conditions in Theorem 3 are satisfied, is is desired that: 1) all common factors of the unique tensor are coupled, leading to $\eta \in \bigcap_{j \in [3]} \Gamma_j$; and 2) the uni-mode unique factors are coupled, i.e., for each mode $j \in [3]$, $\xi_j \in \Gamma_j$. Moreover, for all k there must be at least one mode j such that $k \in \Gamma_j$, otherwise there is no coupling linking tensor \mathcal{Y}_k . However, the uniqueness results in Theorems 3 and 4 concern the model (8)–(12) and would have to be extended in order to rigorously address the flexible coupling case in (40).

Remark 4. There is a trade-off between accuracy and computational cost in the selection of Γ_j . In case the measurements \mathcal{Y}_k are contaminated by noise, one might want to couple as many of the measurements as possible to aid in the estimation of the common factor \mathbf{C}_j , since this increases the amount of available data and can improve accuracy. However, including more measurements in the coupling (i.e., increasing the cardinality of Γ_j) increases the computation cost, since faster solvers are available for the case of $|\Gamma_j| \leq 2$ as will be discussed in detail in Appendix A.

To solve (40), different iterative approaches can be considered, such as the structured data fusion (SDF) framework [1], among others [66]. In this work, we consider an ALS solution, which is a common iterative approach to compute tensor decompositions since each of the updates has an easy-to-find solution [72]. We present the overall procedure in Algorithm 2. The detailed derivations for the solution of the different optimization subproblems are presented in Appendix A.

VI. NUMERICAL EXPERIMENTS

We evaluate the performance of the proposed decomposition methods using synthetic and real data. As a quantitative

Algorithm 2: ALS solution to problem (40)

```

1 Inputs: Data  $\mathcal{Y}_k$ , operators  $\mathbf{P}_{k,j}$ , ranks  $R, L_k$ , couplings  $\Gamma_j$ .
2 Initialize the factors (e.g., randomly or using Algorithm 1) ;
3 while not converged do
4   Compute  $\mathbf{C}_1$  by solving (48) ;
5   Compute  $\mathbf{X}_{k,1}^C$  using (51) and (53) ;
6   Compute  $\mathbf{C}_2$  by solving (49) ;
7   Compute  $\mathbf{X}_{k,2}^C$  using (51) and (54) ;
8   Compute  $\mathbf{C}_3$  by solving (50) ;
9   Compute  $\mathbf{X}_{k,3}^C$  using (51) and (55) ;
10  Compute  $\mathbf{X}_{k,1}^D, \mathbf{X}_{k,2}^D$  and  $\mathbf{X}_{k,3}^D$  using (57), (58) and (59) ;
11 end
12 return  $\mathcal{C} = [\mathbf{C}_1, \mathbf{C}_2, \mathbf{C}_3], \mathcal{D}_k = [\mathbf{X}_{k,1}^D, \mathbf{X}_{k,2}^D, \mathbf{X}_{k,3}^D], k = [K]$ .

```

metric, we considered the normalized root mean square error (NRMSE) between the estimated and the true common component, defined as $\text{NRMSE} = \|\hat{\mathcal{C}} - \mathcal{C}\|_F / \|\mathcal{C}\|_F$. The accurate recovery of \mathcal{C} also means an accurate reconstruction of $\mathcal{D}_k \approx \mathcal{Y}_k - \mathcal{P}_k(\mathcal{C})$ if the SNR is not too low. In all experiments, we computed the average NRMSE results over 20 Monte Carlo runs. The CPDs required by the semi-algebraic algorithms were computed using Tensorlab [73].

A. Synthetic data

1) *Performance as a function of the SNR*: This experiment evaluates the performance of the proposed semi-algebraic and optimization algorithms when decomposing noisy data.

Data generation: We generated synthetic data following model (8), (9), (10), and (11). To this end, we generated common tensor \mathcal{C} , distinct tensors \mathcal{D}_k and $K = 3$ measurements $\mathcal{Y}_1, \mathcal{Y}_2, \mathcal{Y}_3$, all with the same sizes and rank values as in Example 4. The factor matrices of \mathcal{C} and \mathcal{D}_k (generated according to (10) and (11)) were randomly sampled from a standard Gaussian distribution. To generate matrices $\mathbf{P}_{i,j}$ (of appropriate dimensions), we sampled each of their elements from an uniform distribution over the interval $[0, 1]$. The measurements $\mathcal{Y}_1, \mathcal{Y}_2$ and \mathcal{Y}_3 were generated as

$$\mathcal{Y}_k = \mathcal{P}_k(\mathcal{C}) + \mathcal{D}_k + \mathcal{N}_k, \quad (41)$$

where \mathcal{N}_k denotes white Gaussian noise, with SNR given by

$$\text{SNR} = 10 \log_{10} \left(\frac{\mathbb{E}\{\|\mathcal{P}_k(\mathcal{C}) + \mathcal{D}_k\|_F^2\}}{\mathbb{E}\{\|\mathcal{N}_k\|_F^2\}} \right), \quad (42)$$

with $\mathbb{E}\{\cdot\}$ denoting the expectation operator. Note that from the SNR in (42) the noise variance is adjusted according to the energy of the full tensor $\mathcal{P}_k(\mathcal{C}) + \mathcal{D}_k$, which is different from the energy of the common tensor \mathcal{C} evaluated in the NRMSE. As explained in Example 4 this CTD is generically unique in the noiseless case, but only tensor \mathcal{Y}_2 is guaranteed to be fully unique, with \mathcal{Y}_1 and \mathcal{Y}_3 being only uni-mode unique. Moreover, the rank of each measured tensor is larger than the dimension of two of its modes, making this a numerically challenging example.

Algorithms setup: We compared the semi-algebraic algorithm with the optimization algorithm following two different initialization options, with “*init. 1*” being the solution from the semi-algebraic method, and “*init. 2*” being a uniform random

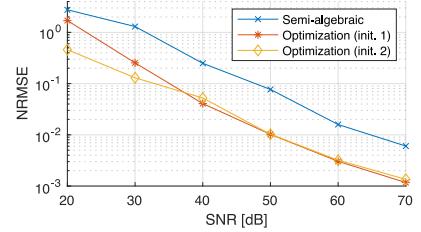


Figure 3. NRMSE for different SNRs for the example with synthetic data.

initialization. We ran the optimization algorithms for at most 1000 ALS iterations. To mitigate the random effect of the initialization, we ran all the algorithms using 50 independent initializations (including the initialization of the CPD used in the semi-algebraic approach) and picked the solution which had the lowest reconstruction errors, $\sum_k \|\mathcal{Y}_k - \hat{\mathcal{Y}}_k\|_F^2$, with $\hat{\mathcal{Y}}_k$ being the reconstructed tensor.

Results: The NRMSE results are shown in Figure 3. It can be seen that the errors are very low for high SNRs (indicating a correct recovery of the true factorization) but increase approximately linearly with the noise intensity. Moreover, although the semi-algebraic solution achieved low NRMSEs for high SNRs, its performance was consistently worse than that of the optimization-based solutions. The two initializations options (*init. 1* and *init. 2*) performed similarly for high (≥ 40 dB) SNRs, but the random initialization (*init. 2*) achieved lower NRMSEs compared to the semi-algebraic one (*init. 1*) for lower SNRs, particularly for the case of 20 dB. However, since the NRMSE values for this SNR are large, they do not provide a meaningful way to compare the quality of different solutions; thus, their interpretation must be performed with proper care.

2) *Ablation experiment 1*: This experiment evaluates the performance of the proposed algorithms when there is variability in the common tensor between different datasets.

Experimental setup: We consider the same general setting as in Section VI-A1 with an SNR fixed at 30 dB. We generate the measured tensors as

$$\mathcal{Y}_k = \mathcal{P}_k(\mathcal{C}_k(\alpha)) + \mathcal{D}_k + \mathcal{N}_k, \quad (43)$$

where $\mathcal{C}_k(\alpha)$ represents a “noisy” version of the common tensor in the k -th dataset, computed as $\mathcal{C}_k(\alpha) = (1 - \alpha)\mathcal{C} + \alpha\mathcal{E}_k$. Tensor \mathcal{C} , which is the same for all datasets, is generated randomly as in Section VI-A1, while \mathcal{E}_k , which is generated in the same way as \mathcal{C} , is independent for each dataset. The contribution of each term is controlled by $\alpha \in [0, 1]$. We decompose \mathcal{Y}_k using the proposed optimization-based algorithm with random initialization (*init. 2*) for different values of α .

Results: The NRMSE, defined for this example as $\text{NRMSE} = \frac{1}{K} \sum_k \|\mathcal{C}_k(\alpha) - \hat{\mathcal{C}}\|_F / \|\mathcal{C}_k(\alpha)\|_F$ for the recovery of $\mathcal{C}_k(\alpha)$, is shown as a function of α (averaged over 20 Monte Carlo runs) in Figure 4. It can be seen that the NRMSE increases with α , which is expected since this reduces the amount of information shared among datasets. Nonetheless, the performance degradation occurs smoothly, indicating the proposed method is not overly sensitive to this type of mismodeling effect when α is small.

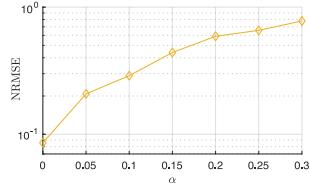


Figure 4. NRMSE (between $\mathcal{C}_k(\alpha)$ and $\hat{\mathcal{C}}$) of the solutions from the optimization algorithm (*init. 2*) as a function of α .

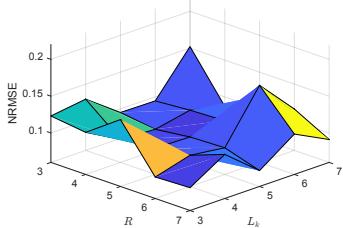


Figure 5. NRMSE of the solution given by the proposed optimization-based algorithm (*init. 2*) as a function of the ranks of the decomposition R and L_k . The true ranks of the tensors are given by $R^{\text{true}} = 5$ and $L_k^{\text{true}} = 5$.

3) *Ablation experiment 2*: This experiment evaluates the performance of the proposed algorithm when the rank of the decomposition is misspecified.

Experimental setup: We generate sets of measurements \mathcal{Y}_k following the same setup as in Section VI-A1, fixing the SNR at 30 dB. Thus, the data are generated from a (noisy) CTD model with ranks $R^{\text{true}} = 5$ and $L_k^{\text{true}} = 5$. We then aim to recover the common tensor \mathcal{C} using the optimization-based algorithm with random initialization (*init. 2*), but specifying different values of ranks in the range $R, L_k \in \{3, 4, 5, 6, 7\}$ to evaluate the impact of rank misspecification.

Results: The NRMSE ($\|\mathcal{C} - \hat{\mathcal{C}}\|_F / \|\mathcal{C}\|_F$) as a function of R and L_k (averaged over 20 Monte Carlo runs) is shown in Figure 5. It can be seen that as the differences between the ranks specified to the algorithm and the true ones increase, so does the NRMSE. Nonetheless, even when the specified ranks are considerably different than the true ones the NRMSE performance is still reasonable (e.g., for $R = 3, L_k = 3$ the NRMSE increases 30% compared to the optimal value). This indicates that the proposed method is not overly sensitive to the choice of the ranks.

4) *Computation time and comparison to SDF*: This experiment compares the execution times of the proposed algorithms, and provides comparisons to the SDF framework of [1].

Experimental setup: We consider the same general experimental setup as in Section VI-A1 while fixing the SNR at 30 dB. The SDF framework [1] was implemented to solve problem (40) using Tensorlab [73]. All algorithms were implemented on MatlabTM and run on a MacBook Pro M1 with 16GB RAM.

Results: The NRMSE and execution times (averaged over 20 Monte Carlo runs) can be seen in Table I. It can be seen that the proposed optimization-based approach (with both random and semi-algebraic initializations) and SDF obtain very similar NRMSEs, whereas the semi-algebraic approach present a higher NRMSE. In terms of computation times, the fastest approach was Algorithm 2 with random initialization (*init. 2*), followed by the semi-algebraic solution, whose exe-

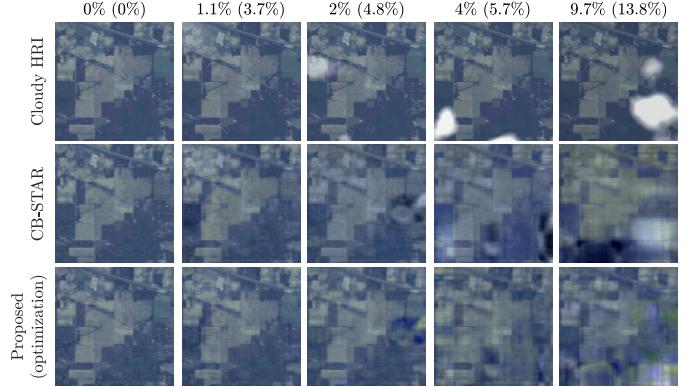


Figure 6. Illustration of common component \mathcal{C} and of the high resolution image reconstructed by the proposed method and by CB-STAR (the best performing competing method). The cloudy image \mathcal{X}_1 (underlying the HSI) for the different cloud contamination levels is shown on top (the top-left corresponds to the ground truth). The percentages on top read as CC% (CP%).

cution is slightly slower due to an iterative optimization procedure being used to compute the CPDs. The SDF framework, although very flexible to tackle different types of problems, presented the longest execution time, significantly higher than the proposed algorithms.

Table I
NRMSE AND COMPUTATION TIME COMPARISON WITH SYNTHETIC DATA.

	Semi-algebraic	Opt. (<i>init. 1</i>)	Opt. (<i>init. 2</i>)	SDF [1]
NRMSE	0.9088	0.0770	0.0767	0.0765
Time [s]	2.776	6.531	1.026	23.553

B. Real data

For the experiments with real data, we consider the problem of imaging a given scene subject to cloud contamination. Two distinct acquisitions of this scene are obtained, one yielding an HSI and one an MSI, each of which is subject to some random cloud contamination (which are independent and identically distributed across the images). This results in image-specific variability affecting both the HSI and the MSI. The objective is to recover an uncontaminated high-resolution image (HRI) of the scene by fusing the acquired HSI and MSI.

Experimental setup: We consider an underlying high-resolution image $\mathcal{C} \in \mathbb{R}^{145 \times 145 \times 200}$ as the Indian Pines image, with 145×145 pixels and 200 spectral bands. Cloud cover maps (i.e., percentage of clouds per pixel) were then generated randomly using a nonlinear transformation of a Gaussian random field, which led to realistic cloud distribution (see the first row of Figure 6) and allowed the intensity and spatial extent of the clouds to be adjusted. Specifically, we generated two cloud cover maps, S_1 and S_2 ($\in \mathbb{R}^{145 \times 145}$) to represent the cloud cover underlying the HSI and MSI, respectively. Then, cloud-corrupted images \mathcal{X}_1 and \mathcal{X}_2 were generated as a convex combination between the ground image and a cloud spectral signature as

$$[\mathcal{X}_k]_{x,y,\lambda} = [\mathcal{C}]_{x,y,\lambda} (1 - [S_k]_{x,y}) + g_\lambda [S_k]_{x,y}, \quad (44)$$

for $k \in [2]$, where $g_\lambda \in \mathbb{R}$ is the cloud spectral reflectance at band index λ , which was computed from a real cloud signature presented in [74], and x, y is the pixel indices.

Table II

AVERAGE NRMSE RESULTS FOR FUSING AN HSI AND AN MSI WITH DIFFERENT AMOUNTS OF CLOUD CONTAMINATION ON BOTH IMAGES.

Cloud cover (CC) %	0%	1.1%	2%	4%	9.7%
Corrupted pixels (CP) %	0%	3.7%	4.8%	5.7%	13.8%
STEREO	0.035	0.082	0.146	0.222	0.386
SCOTT	0.038	0.088	0.147	0.280	0.436
CT-STAR	0.054	0.094	0.154	0.291	0.450
CB-STAR	0.037	0.081	0.100	0.182	0.371
Proposed (semi-algebraic)	0.079	0.102	0.125	0.189	0.297
Proposed (optimization)	0.052	0.055	0.073	0.134	0.251

Note that decomposing \mathcal{X}_1 and \mathcal{X}_2 into common and distinct components is challenging since, due to the model (44), neither the common nor the distinct components will have low CP rank. Thus, the data do not really follow the model in (10), (11), leading to an approximation problem rather than an exact decomposition.

From the cloudy high-resolution images \mathcal{X}_1 and \mathcal{X}_2 we generated the HSI and the MSI, respectively, using degradation operators $\mathbf{P}_{k,j}$ as described in Example 3 of Section III. We used a spatial decimation factor of four for the HSI and the spectral response of the Sentinel-2A sensor for the MSI (leading to a spectral decimation factor of 10 for the MSI). The HSI $\mathcal{Y}_1 \in \mathbb{R}^{36 \times 36 \times 200}$ and the MSI $\mathcal{Y}_2 \in \mathbb{R}^{145 \times 145 \times 10}$ were then generated as

$$\mathcal{Y}_k = \mathcal{P}_k(\mathcal{X}_k) + \mathcal{N}_k \quad (45)$$

for $k \in \{1, 2\}$, where \mathcal{P}_k is as defined in (9) and \mathcal{N}_k are white Gaussian noise tensors whose variance is adjusted to obtain an SNR of 30 dB, which is here defined as

$$\text{SNR} = 10 \log_{10} \left(\frac{\mathbb{E}\{\|\mathcal{P}_k(\mathcal{X}_k)\|_F^2\}}{\mathbb{E}\{\|\mathcal{N}_k\|_F^2\}} \right). \quad (46)$$

For more details on the measurement model for this application, see, e.g., [8], [9], [75].

We adjusted the cloud cover generation setup to obtain images with a desired amount of cloud contamination, which was kept the same for the HSI and MSI and was measured using two metrics. The first metric is the cloud cover (CC) percent, defined as the mean percentage of clouds per pixel and computed as $\text{CC} = \frac{1}{Z} \sum_{k,x,y} [\mathcal{S}_k]_{x,y}$, in which $Z = 2 \times 145^2$ is a normalization equal to the total number of terms the summation. The second metric is the percent of corrupted pixels (CP), defined as the mean percentage of pixels with more than 15% of cloud presence, and computed as $\text{CP} = \frac{1}{Z} \sum_{k,x,y} \iota([\mathcal{S}_k]_{x,y} > 0.15)$, where $\iota(\cdot)$ is an indicator function and Z is defined as before. Note that the CC and CP metrics shown in the experiments are also averaged among all Monte Carlo runs.

Algorithms setup: We compare the proposed method with the state-of-the-art CTD approaches STEREO [9], SCOTT [8], and with CT-STAR and CB-STAR [10], the latter two which consider an image specific component in one of the images (the MSI). The ranks of all methods were adjusted to maximize performance in a separate independently generated dataset. We implement the optimization algorithm initialized with the algebraic approach, and considered up to 50 ALS iterations. For this example, we implemented a modification on the semi-algebraic algorithm to simplify it and improve its

robustness inspired by TenRec in [9]. Specifically, we follow the steps 1 through 4 detailed in Section V-A, and exploited the fact that $\mathbf{P}_{2,1} = \mathbf{P}_{2,2} = \mathbf{I}$ to compute $\hat{\mathbf{C}}_1$ and $\hat{\mathbf{C}}_2$ from the common parts of the factors obtained through the CPD of the MSI \mathcal{Y}_2 . However, to compute the final factor $\hat{\mathbf{C}}_3$ we use a different approach by just solving the linear regression problem on the HSI as $\min_{\mathbf{C}_3} \|\mathcal{Y}_1 - [\mathbf{P}_{1,1} \hat{\mathbf{C}}_1, \mathbf{P}_{1,2} \hat{\mathbf{C}}_2, \mathbf{C}_3]\|_F$. This showed better performance than recovering \mathbf{C}_3 through the separate CPD of the HSI \mathcal{Y}_1 .

Results: Quantitative results are shown in Table II along with the CC and CP percentages. Visual results of the two overall best performing methods (the proposed optimization-based algorithm and CB-STAR) corresponding to the run with the median performance difference between them are shown in Figure 6. It can be seen that the proposed methods perform slightly worse when there are no clouds, which is natural as the models in the competing methods (without personalized components) match the dataset better. However, as soon as cloud cover is present (even for 1.1%) the proposed optimization-based approach gets considerably lower NRMSE than the other algorithms. The semi-algebraic algorithm performed considerably worse than the optimization one for low CC percentages, with their relative performances getting better as the cloud contamination increases. The visual results also corroborate these findings, with the reconstruction by CB-STAR showing more artifacts compared to the proposed method reconstruction. The artifacts also get increasingly strong as the cloud contamination increases, being subtle for a CC of 1.1% but very strong for a CC of 9.7%.

VII. CONCLUSIONS AND PERSPECTIVES

In this paper, a personalized CTD framework was proposed. A flexible model was presented representing the datasets as the sum of two components, each admitting a low-rank CPD. The first component was linked to a common tensor through a multilinear measurement model, while the second component was distinct and captured dataset-specific information. This model generalized several existing CPDs. The generic uniqueness of the decomposition was shown to hold under mild conditions that highlighted the influence of the measurement model and of “weaker” uni-mode uniqueness of individual datasets on the uniqueness of the full decomposition. Assuming the ranks of the common and distinct components to be known, two algorithms were proposed to compute the decomposition. The first was semi-algebraic, performing well for high SNRs but being less robust in low-SNR settings. The second was based on an optimization procedure and showed better performance, especially on noisy and real data. Experiments illustrated the advantage of the proposed framework compared with the state of the art methods.

An important issue related to the practical use of the proposed CTD framework is the estimation of its ranks, R and L_k . While several strategies have been investigated to estimate the number of common components in statistical models involving multiple datasets (see, e.g., [76]–[78]), these approaches are based on a very different set of model assumptions. Thus, in future work we aim to investigate the order selection problem

for CTDs with common and distinct components from both theoretical and experimental perspectives.

APPENDIX A SOLUTION TO THE OPTIMIZATION PROBLEM (40)

Optimization w.r.t. C_j : Let us first consider the solution for C_1 . Incorporating the equality constraint (40b) directly in the cost function, this problem can be written as:

$$\min_{C_1} \sum_{k \in \Gamma_1} \left\| \tilde{\mathbf{Y}}_k - [\mathbf{P}_{k,1} \mathbf{C}_1, \mathbf{X}_{k,2}^C, \mathbf{X}_{k,3}^C] \right\|_F^2 \quad (47)$$

where $\tilde{\mathbf{Y}}_k = \mathbf{Y}_k - [\mathbf{X}_{k,1}^D, \mathbf{X}_{k,2}^D, \mathbf{X}_{k,3}^D]$. Note that this is equivalent to a coupled CP decomposition without distinct components. Using the properties of the mode-1 matricization, it can be shown that the critical points of (47) satisfy

$$\sum_{k \in \Gamma_1} (\mathbf{J}_{k,1}^\top \mathbf{J}_{k,1}) \mathbf{C}_1^\top (\mathbf{P}_{k,1}^\top \mathbf{P}_{k,1}) - \mathbf{J}_{k,1}^\top \tilde{\mathbf{Y}}_{k\langle 1 \rangle} \mathbf{P}_{k,1} = \mathbf{0}, \quad (48)$$

where $\mathbf{J}_{k,1} = \mathbf{X}_{k,3}^C \odot \mathbf{X}_{k,2}^C$. Note that the product $\mathbf{J}_{k,1}^\top \mathbf{J}_{k,1}$ can be computed using mixed product properties (see, e.g., [47, eq. 2.2]). Since (48) is linear in $\mathbf{C}_1 \in \mathbb{R}^{M_1 \times R}$, it can be solved with a complexity of at most $O((M_1 R)^3)$ operations. However, since this consists in a generalization of the Sylvester equation, when $|\Gamma_1| \leq 2$ specialized solvers can be much faster: the complexity of solving the generalized Sylvester equation in this case is $O(M_1^3 + R^3)$ operations when using the Hessenberg-Schur or Bartels-Stewart methods [79]–[82].

Optimizing the cost function w.r.t. C_2 and C_3 can be done similarly by using the mode-2 and mode-3 tensor matricizations, which leads to the following equations:

$$\sum_{k \in \Gamma_2} (\mathbf{J}_{k,2}^\top \mathbf{J}_{k,2}) \mathbf{C}_2^\top (\mathbf{P}_{k,2}^\top \mathbf{P}_{k,2}) - \mathbf{J}_{k,2}^\top \tilde{\mathbf{Y}}_{k\langle 2 \rangle} \mathbf{P}_{k,2} = \mathbf{0}, \quad (49)$$

$$\sum_{k \in \Gamma_3} (\mathbf{J}_{k,3}^\top \mathbf{J}_{k,3}) \mathbf{C}_3^\top (\mathbf{P}_{k,3}^\top \mathbf{P}_{k,3}) - \mathbf{J}_{k,3}^\top \tilde{\mathbf{Y}}_{k\langle 3 \rangle} \mathbf{P}_{k,3} = \mathbf{0}, \quad (50)$$

where $\mathbf{J}_{k,2} = \mathbf{X}_{k,3}^C \odot \mathbf{X}_{k,1}^C$ and $\mathbf{J}_{k,3} = \mathbf{X}_{k,2}^C \odot \mathbf{X}_{k,1}^C$.

Optimization w.r.t. $\mathbf{X}_{k,j}^C$: The solution to this problem depends on whether $\mathbf{X}_{k,j}^C$ is included in the set of equality constraints (40b). If $k \in \Gamma_j$, then the solution is direct:

$$\mathbf{X}_{k,j}^C = \mathbf{P}_{k,j} \mathbf{C}_j, \quad k \in \Gamma_j, \quad j \in [3]. \quad (51)$$

On the other hand, if $k \notin \Gamma_j$, the optimization problem for the mode $j = 1$ will be:

$$\min_{\mathbf{X}_{k,1}^C} \left\| \tilde{\mathbf{Y}}_k - [\mathbf{X}_{k,1}^C, \mathbf{X}_{k,2}^C, \mathbf{X}_{k,3}^C] \right\|_F^2, \quad (52)$$

where $\tilde{\mathbf{Y}}_k = \mathbf{Y}_k - [\mathbf{X}_{k,1}^D, \mathbf{X}_{k,2}^D, \mathbf{X}_{k,3}^D]$. The solution to this problem is given by

$$\mathbf{X}_{k,1}^C = [(\mathbf{X}_{k,3}^C \odot \mathbf{X}_{k,2}^C)^\dagger \tilde{\mathbf{Y}}_{k\langle 1 \rangle}]^\top, \quad k \notin \Gamma_1. \quad (53)$$

The same steps can be used to compute the solutions for modes $j = 2$ and $j = 3$:

$$\mathbf{X}_{k,2}^C = [(\mathbf{X}_{k,3}^C \odot \mathbf{X}_{k,1}^C)^\dagger \tilde{\mathbf{Y}}_{k\langle 2 \rangle}]^\top, \quad k \notin \Gamma_2, \quad (54)$$

$$\mathbf{X}_{k,3}^C = [(\mathbf{X}_{k,2}^C \odot \mathbf{X}_{k,1}^C)^\dagger \tilde{\mathbf{Y}}_{k\langle 3 \rangle}]^\top, \quad k \notin \Gamma_3. \quad (55)$$

Optimization w.r.t. $\mathbf{X}_{k,j}^D$: This problem is given by

$$\min_{\mathbf{X}_{k,1}^D, \mathbf{X}_{k,2}^D, \mathbf{X}_{k,3}^D} \left\| \bar{\mathbf{Y}}_k - [\mathbf{X}_{k,1}^D, \mathbf{X}_{k,2}^D, \mathbf{X}_{k,3}^D] \right\|_F^2, \quad (56)$$

where $\bar{\mathbf{Y}}_k = \mathbf{Y}_k - [\mathbf{X}_{k,1}^C, \mathbf{X}_{k,2}^C, \mathbf{X}_{k,3}^C]$. Problem (56) consists of computing a rank- L_k CPD of $\bar{\mathbf{Y}}_k$. Following the ALS strategy, we update $\mathbf{X}_{k,j}^D$ sequentially over $j \in [3]$. Using the mode- j matricization and proceeding in the same way as in the previous problem, the solutions are given by:

$$\mathbf{X}_{k,1}^D = [(\mathbf{X}_{k,3}^D \odot \mathbf{X}_{k,2}^D)^\dagger \bar{\mathbf{Y}}_{k\langle 1 \rangle}]^\top, \quad (57)$$

$$\mathbf{X}_{k,2}^D = [(\mathbf{X}_{k,3}^D \odot \mathbf{X}_{k,1}^D)^\dagger \bar{\mathbf{Y}}_{k\langle 2 \rangle}]^\top, \quad (58)$$

$$\mathbf{X}_{k,3}^D = [(\mathbf{X}_{k,2}^D \odot \mathbf{X}_{k,1}^D)^\dagger \bar{\mathbf{Y}}_{k\langle 3 \rangle}]^\top. \quad (59)$$

REFERENCES

- [1] L. Sorber, M. Van Barel, and L. De Lathauwer, “Structured data fusion,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 4, pp. 586–600, 2015.
- [2] E. Acar, M. A. Rasmussen, F. Savorani, T. Næs, and R. Bro, “Understanding data fusion within the framework of coupled matrix and tensor factorizations,” *Chemometrics and Intelligent Laboratory Systems*, vol. 129, pp. 53–63, 2013.
- [3] D. Lahat, T. Adali, and C. Jutten, “Multimodal data fusion: an overview of methods, challenges, and prospects,” *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1449–1477, 2015.
- [4] E. Acar, E. E. Papalexakis, G. Gürdeniz, M. A. Rasmussen, A. J. Lawaetz, M. Nilsson, and R. Bro, “Structure-revealing data fusion,” *BMC bioinformatics*, vol. 15, no. 1, pp. 1–17, 2014.
- [5] C. Chatzichristos, S. Van Eynhoven, E. Kofidis, and S. Van Huffel, “Chapter 10 - Coupled tensor decompositions for data fusion,” in *Tensors for data processing*, Y. Liu, Ed. Elsevier, 2022, pp. 341–370.
- [6] E. E. Papalexakis, C. Faloutsos, T. M. Mitchell, P. P. Talukdar, N. D. Sidiropoulos, and B. Murphy, “Turbo-SMT: Accelerating coupled sparse matrix-tensor factorizations by 200x - Supplementary material,” in *Proceedings of the 2014 SIAM International Conference on Data Mining*. SIAM, 2014, pp. 118–126.
- [7] B. Ermiş, E. Acar, and A. T. Cemgil, “Link prediction in heterogeneous data via generalized coupled tensor factorization,” *Data Mining and Knowledge Discovery*, vol. 29, no. 1, pp. 203–236, 2015.
- [8] C. Prévost, K. Usevich, P. Comon, and D. Brie, “Hyperspectral super-resolution with coupled tucker approximation: Recoverability and SVD-based algorithms,” *IEEE Transactions on Signal Processing*, vol. 68, pp. 931–946, 2020.
- [9] C. I. Kanatsoulis, X. Fu, N. D. Sidiropoulos, and W.-K. Ma, “Hyperspectral super-resolution: A coupled tensor factorization approach,” *IEEE Transactions on Signal Processing*, vol. 66, no. 24, pp. 6503–6517, 2018.
- [10] R. A. Borsoi, C. Prévost, K. Usevich, D. Brie, J. C. M. Bermudez, and C. Richard, “Coupled tensor decomposition for hyperspectral and multispectral image fusion with inter-image variability,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 3, pp. 702–717, 2021.
- [11] C. Prévost, R. A. Borsoi, K. Usevich, D. Brie, J. C. M. Bermudez, and C. Richard, “Hyperspectral super-resolution accounting for spectral variability: Coupled tensor LL1-based recovery and blind unmixing of the unknown super-resolution image,” *SIAM Journal on Imaging Sciences*, vol. 15, no. 1, pp. 110–138, 2022.
- [12] L.-D. Kuang, Z.-M. He, J. Zhang, and F. Li, “Coupled canonical polyadic decomposition of multi-group fMRI data with spatial reference and orthonormality constraints,” *Biomedical Signal Processing and Control*, vol. 80, p. 104232, 2023.
- [13] R. A. Borsoi, I. Lehmann, M. A. B. S. Akhonda, V. Calhoun, K. Usevich, D. Brie, and T. Adali, “Coupled CP tensor decomposition with shared and distinct components for multi-task fMRI data fusion,” in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Rhodes, Greece: IEEE, 2023, pp. 1–5.
- [14] M. Sørensen and L. De Lathauwer, “Multidimensional harmonic retrieval via coupled canonical polyadic decomposition–Part I: Model and identifiability,” *IEEE Transactions on Signal Processing*, vol. 65, no. 2, pp. 517–527, 2016.

[15] ——, “Multidimensional harmonic retrieval via coupled canonical polyadic decomposition—Part II: Algorithm and multivariate sampling,” *IEEE Transactions on Signal Processing*, vol. 65, no. 2, pp. 528–539, 2016.

[16] ——, “Multiple invariance ESPRIT for nonuniform linear arrays: A coupled canonical polyadic decomposition approach,” *IEEE Transactions on Signal Processing*, vol. 64, no. 14, pp. 3693–3704, 2016.

[17] M. Sørensen, I. Domanov, and L. De Lathauwer, “Coupled canonical polyadic decompositions and multiple shift invariance in array processing,” *IEEE Transactions on Signal Processing*, vol. 66, no. 14, pp. 3665–3680, 2018.

[18] H. Zheng, Z. Shi, C. Zhou, M. Haardt, and J. Chen, “Coupled coarray tensor CPD for DOA estimation with coprime L-shaped array,” *IEEE Signal Processing Letters*, vol. 28, pp. 1545–1549, 2021.

[19] F. Van Eeghem, O. Debals, N. Vervliet, and L. De Lathauwer, “Coupled and incomplete tensors in blind system identification,” *IEEE Transactions on Signal Processing*, vol. 66, no. 23, pp. 6137–6147, 2018.

[20] X.-F. Gong, Q.-H. Lin, F.-Y. Cong, and L. De Lathauwer, “Double coupled canonical polyadic decomposition for joint blind source separation,” *IEEE Transactions on Signal Processing*, vol. 66, no. 13, pp. 3475–3490, 2018.

[21] Y. Yu and A. P. Petropulu, “PARAFAC-based blind estimation of possibly underdetermined convolutive MIMO systems,” *IEEE Transactions on Signal Processing*, vol. 56, no. 1, pp. 111–124, 2007.

[22] E. Acar, R. Bro, and A. K. Smilde, “Data fusion in metabolomics using coupled matrix and tensor factorizations,” *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1602–1620, 2015.

[23] E. Karahan, P. A. Rojas-Lopez, M. L. Bringas-Vega, P. A. Valdés-Hernández, and P. A. Valdés-Sosa, “Tensor analysis and fusion of multimodal brain images,” *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1531–1559, 2015.

[24] C. Chatzichristos, E. Kofidis, W. Van Paesschen, L. De Lathauwer, S. Theodoridis, and S. Van Huffel, “Early soft and flexible fusion of electroencephalography and functional magnetic resonance imaging via double coupled matrix tensor factorization for multisubject group analysis,” *Human brain mapping*, vol. 43, no. 4, pp. 1231–1255, 2022.

[25] I. Belyaeva, B. Gabrielson, Y.-P. Wang, T. W. Wilson, V. D. Calhoun, J. M. Stephen, and T. Adali, “Learning spatiotemporal brain dynamics in adolescents via multimodal MEG and fMRI data fusion using joint tensor/matrix decomposition,” *IEEE Transactions on Biomedical Engineering*, 2024.

[26] T. Adali, M. Akhonda, and V. D. Calhoun, “ICA and IVA for data fusion: An overview and a new approach based on disjoint subspaces,” *IEEE Sensors Letters*, vol. 3, no. 1, pp. 1–4, 2019.

[27] L. De Lathauwer and E. Kofidis, “Coupled matrix-tensor factorizations—the case of partially shared factors,” in *Proc. 51st Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2017, pp. 711–715.

[28] A. K. Smilde, I. Måge, T. Naes, T. Hankemeier, M. A. Lips, H. A. Kiers, E. Acar, and R. Bro, “Common and distinct components in data fusion,” *Journal of Chemometrics*, vol. 31, no. 7, p. e2900, 2017.

[29] E. Acar, C. Schenker, Y. Levin-Schwartz, V. D. Calhoun, and T. Adali, “Unraveling diagnostic biomarkers of schizophrenia through structure-revealing fusion of multi-modal neuroimaging data,” *Frontiers in neuroscience*, vol. 13, p. 416, 2019.

[30] S. J. Prince and J. H. Elder, “Probabilistic linear discriminant analysis for inferences about identity,” in *Proc. 11th international conference on computer vision (ICCV)*. IEEE, 2007, pp. 1–8.

[31] X. Wang, R. A. Borsoi, C. Richard, and J. Chen, “Deep hyperspectral and multispectral image fusion with inter-image variability,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, 2023.

[32] W. Liu, J. Chan, J. Bailey, C. Leckie, and K. Ramamohanarao, “Mining labelled tensors by discovering both their common and discriminative subspaces,” in *Proceedings of the 2013 SIAM International Conference on Data Mining*. SIAM, 2013, pp. 614–622.

[33] X. Wen, Y.-R. Lin, and K. Pelechrinis, “Pairfac: Event analytics through discriminant tensor factorization,” in *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, 2016, pp. 519–528.

[34] M. Sørensen and N. D. Sidiropoulos, “Multi-set low-rank factorizations with shared and unshared components,” *IEEE Transactions on Signal Processing*, vol. 68, pp. 5122–5137, 2020.

[35] C. Schenker, J. E. Cohen, and E. Acar, “A flexible optimization framework for regularized matrix-tensor factorizations with linear couplings,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 3, pp. 506–521, 2020.

[36] R. C. Farias, J. E. Cohen, and P. Comon, “Exploring multimodal data fusion through joint decompositions with flexible couplings,” *IEEE Transactions on Signal Processing*, vol. 64, no. 18, pp. 4830–4844, 2016.

[37] T. Adali, F. Kantar, M. A. B. S. Akhonda, S. Strother, V. D. Calhoun, and E. Acar, “Reproducibility in matrix and tensor decompositions: Focus on model match, interpretability, and uniqueness,” *IEEE Signal Processing Magazine*, vol. 39, no. 4, pp. 8–24, 2022.

[38] M. Sørensen and L. D. De Lathauwer, “Coupled canonical polyadic decompositions and (coupled) decompositions in multilinear rank- $(L_{r,n}, L_{r,n}, 1)$ terms—part I: Uniqueness,” *SIAM Journal on Matrix Analysis and Applications*, vol. 36, no. 2, pp. 496–522, 2015.

[39] M. Sørensen, I. Domanov, and L. De Lathauwer, “Coupled canonical polyadic decompositions and (coupled) decompositions in multilinear rank- $(L_{r,n}, L_{r,n}, 1)$ terms—part II: Algorithms,” *SIAM Journal on Matrix Analysis and Applications*, vol. 36, no. 3, pp. 1015–1045, 2015.

[40] C. I. Kanatsoulis, X. Fu, N. D. Sidiropoulos, and M. Akçakaya, “Tensor completion from regular sub-Nyquist samples,” *IEEE Transactions on Signal Processing*, vol. 68, pp. 1–16, 2019.

[41] G. Zhang, X. Fu, J. Wang, X.-L. Zhao, and M. Hong, “Spectrum cartography via coupled block-term tensor decomposition,” *IEEE Transactions on Signal Processing*, vol. 68, pp. 3660–3675, 2020.

[42] N. Shi and R. A. Kontar, “Personalized PCA: Decoupling shared and unique features,” *Journal of machine learning research*, vol. 25, pp. 1–82, 2024.

[43] G. Liang, N. Shi, R. A. Kontar, and S. Fattah, “Personalized dictionary learning for heterogeneous datasets,” in *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

[44] N. Shi, R. A. Kontar, and S. Fattah, “Heterogeneous matrix factorization: When features differ by datasets,” *arXiv preprint arXiv:2305.17744*, 2023.

[45] X. Guo, S. Miron, D. Brie, and A. Stegeman, “Uni-mode and partial uniqueness conditions for CANDECOMP/PARAFAC of three-way arrays with linearly dependent loadings,” *SIAM Journal on Matrix Analysis and Applications*, vol. 33, no. 1, pp. 111–129, 2012.

[46] 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 Journal on Matrix Analysis and Applications*, vol. 34, no. 3, pp. 855–875, 2013.

[47] T. G. Kolda and B. W. Bader, “Tensor decompositions and applications,” *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.

[48] J. B. Kruskal, “Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics,” *Linear algebra and its applications*, vol. 18, no. 2, pp. 95–138, 1977.

[49] A. Stegeman and N. D. Sidiropoulos, “On Kruskal’s uniqueness condition for the Candecomp/Parafac decomposition,” *Linear Algebra and its applications*, vol. 420, no. 2-3, pp. 540–552, 2007.

[50] I. Domanov and L. De Lathauwer, “On the uniqueness of the canonical polyadic decomposition of third-order tensors—Part II: Uniqueness of the overall decomposition,” *SIAM Journal on Matrix Analysis and Applications*, vol. 34, no. 3, pp. 876–903, 2013.

[51] L. Chiantini and G. Ottaviani, “On generic identifiability of 3-tensors of small rank,” *SIAM Journal on Matrix Analysis and Applications*, vol. 33, no. 3, pp. 1018–1037, 2012.

[52] I. Domanov and L. D. Lathauwer, “Canonical polyadic decomposition of third-order tensors: Reduction to generalized eigenvalue decomposition,” *SIAM Journal on Matrix Analysis and Applications*, vol. 35, no. 2, pp. 636–660, 2014.

[53] L. De Lathauwer, “Decompositions of a higher-order tensor in block terms—part II: Definitions and uniqueness,” *SIAM Journal on Matrix Analysis and Applications*, vol. 30, no. 3, pp. 1033–1066, 2008.

[54] J. Douglas Carroll, S. Pruzansky, and J. B. Kruskal, “CANDELINC: A general approach to multidimensional analysis of many-way arrays with linear constraints on parameters,” *Psychometrika*, vol. 45, pp. 3–24, 1980.

[55] Y. Levin-Schwartz, V. D. Calhoun, and T. Adali, “Quantifying the interaction and contribution of multiple datasets in fusion: Application to the detection of schizophrenia,” *IEEE transactions on medical imaging*, vol. 36, no. 7, pp. 1385–1395, 2017.

[56] C. Chatzichristos, E. Kofidis, M. Morante, and S. Theodoridis, “Blind fMRI source unmixing via higher-order tensor decompositions,” *Journal of neuroscience methods*, vol. 315, pp. 17–47, 2019.

[57] C. Prévost and F. Odille, “Non-local tensor sparse coding for multi-image super-resolution in magnetic resonance imaging,” 2023.

[58] ——, “Multi-frame super-resolution MRI using coupled low-rank tucker approximation,” in *Proc. 30th European Signal Processing Conference (EUSIPCO)*. IEEE, 2022, pp. 1263–1267.

[59] W. Shi, H. Xu, C. Sun, J. Sun, Y. Li, X. Xu, T. Zheng, Y. Zhang, G. Wang, and D. Wu, "AFFIRM: affinity fusion-based framework for iteratively random motion correction of multi-slice fetal brain MRI," *IEEE Transactions on Medical Imaging*, vol. 42, no. 1, pp. 209–219, 2022.

[60] R. A. Borsoi, T. Imbiriba, J. C. M. Bermudez, C. Richard, J. Chanussot, L. Drumetz, J.-Y. Tourneret, A. Zare, and C. Jutten, "Spectral variability in hyperspectral data unmixing: A comprehensive review," *IEEE geoscience and remote sensing magazine*, vol. 9, no. 4, pp. 223–270, 2021.

[61] E. Acar, Y. Levin-Schwartz, V. D. Calhoun, and T. Adali, "Tensor-based fusion of EEG and fMRI to understand neurological changes in schizophrenia," in *Proc. IEEE international symposium on circuits and systems (ISCAS)*. IEEE, 2017, pp. 1–4.

[62] L. Li, H. C. Hoefsloot, B. M. Bakker, D. Horner, M. A. Rasmussen, A. Smilde, and E. Acar, "Longitudinal metabolomics data analysis informed by mechanistic models," *bioRxiv*, 2024, <https://doi.org/10.1101/2024.08.13.607724>.

[63] D. S. Bernstein, *Matrix mathematics: theory, facts, and formulas*, 2nd ed. Princeton university press, 2009.

[64] Q. Lyu and X. Fu, "Nonlinear multiview analysis: Identifiability and neural network-assisted implementation," *IEEE Transactions on Signal Processing*, vol. 68, pp. 2697–2712, 2020.

[65] R. Caron and T. Traylor, "The zero set of a polynomial," University of Windsor, Windsor, ON, Canada, Tech. Rep. WSMR 05-02, 2005.

[66] X. Fu, N. Vervliet, L. De Lathauwer, K. Huang, and N. Gillis, "Computing large-scale matrix and tensor decomposition with structured factors: A unified nonconvex optimization perspective," *IEEE Signal Processing Magazine*, vol. 37, no. 5, pp. 78–94, 2020.

[67] M. Rajih, P. Comon, and R. A. Harshman, "Enhanced line search: A novel method to accelerate PARAFAC," *SIAM journal on matrix analysis and applications*, vol. 30, no. 3, pp. 1128–1147, 2008.

[68] L. Ramshaw and R. E. Tarjan, "On minimum-cost assignments in unbalanced bipartite graphs," HP Labs, Palo Alto, CA, USA, Tech. Rep. HPL-2012-40R1, Tech. Rep., 2012.

[69] D. Hong, T. G. Kolda, and J. A. Duersch, "Generalized canonical polyadic tensor decomposition," *SIAM Review*, vol. 62, no. 1, pp. 133–163, 2020.

[70] W. Pu, S. Ibrahim, X. Fu, and M. Hong, "Stochastic mirror descent for low-rank tensor decomposition under non-euclidean losses," *IEEE Transactions on Signal Processing*, vol. 70, pp. 1803–1818, 2022.

[71] C. Prévost and V. Leplat, "Nonnegative block-term decomposition with the β -divergence: Joint data fusion and blind spectral unmixing," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Rhodes Island, Greece: IEEE, 2023, pp. 1–5.

[72] P. Comon, X. Luciani, and A. L. De Almeida, "Tensor decompositions, alternating least squares and other tales," *Journal of Chemometrics: A Journal of the Chemometrics Society*, vol. 23, no. 7-8, pp. 393–405, 2009.

[73] N. Vervliet, O. Debals, L. Sorber, M. Van Barel, and L. De Lathauwer, (2016, Mar.) Tensorlab 3.0. Available online. [Online]. Available: <https://www.tensorlab.net>

[74] B.-C. Gao, W. Han, S. C. Tsay, and N. F. Larsen, "Cloud detection over the Arctic region using airborne imaging spectrometer data during the daytime," *Journal of Applied Meteorology*, vol. 37, no. 11, pp. 1421–1429, 1998.

[75] R. A. Borsoi, T. Imbiriba, and J. C. M. Bermudez, "Super-resolution for hyperspectral and multispectral image fusion accounting for seasonal spectral variability," *IEEE transactions on image processing*, vol. 29, pp. 116–127, 2019.

[76] M. Akhonda, B. Gabrielson, S. Bhinge, V. D. Calhoun, and T. Adali, "Disjoint subspaces for common and distinct component analysis: Application to the fusion of multi-task fMRI data," *Journal of neuroscience methods*, vol. 358, p. 109214, 2021.

[77] S. Bhinge, Y. Levin-Schwartz, and T. Adali, "Estimation of common subspace order across multiple datasets: Application to multi-subject fMRI data," in *Proc. 51st Annual Conference on Information Sciences and Systems (CISS)*. IEEE, 2017, pp. 1–5.

[78] T. Hasija, T. Marrinan, C. Lameiro, and P. J. Schreier, "Determining the dimension and structure of the subspace correlated across multiple data sets," *Signal Processing*, vol. 176, p. 107613, 2020.

[79] R. H. Bartels and G. W. Stewart, "Solution of the matrix equation $AX+XB=C$," *Communications of the ACM*, vol. 15, no. 9, pp. 820–826, 1972.

[80] J. D. Gardiner, A. J. Laub, J. J. Amato, and C. B. Moler, "Solution of the Sylvester matrix equation $AXB^T + CXD^T = E$," *ACM Transactions on Mathematical Software (TOMS)*, vol. 18, no. 2, pp. 223–231, 1992.

[81] G. Golub, S. Nash, and C. Van Loan, "A Hessenberg-Schur method for the problem $ax+xb=c$," *IEEE Transactions on Automatic Control*, vol. 24, no. 6, pp. 909–913, 1979.

[82] V. Simoncini, "Computational methods for linear matrix equations," *SIAM Review*, vol. 58, no. 3, pp. 377–441, 2016.



Ricardo Borsoi (Member, IEEE) received the Doctorate degree from the Federal University of Santa Catarina, Brazil, and from the Côte d'Azur University, France, in 2021. He is currently a permanent research scientist in the French National Center for Scientific Research (CNRS) in the CRAN laboratory, Nancy, France. His research focuses on statistical inference, tensor decompositions and inverse problems, with applications in hyperspectral and medical image analysis.



Konstantin Usevich (Member, IEEE) obtained a Ph.D in 2011, from St. Petersburg State University, Russia. He held several postdoctoral positions in UK, Belgium, and France, before joining the CNRS in 2017. He is currently a permanent CNRS researcher at CRAN laboratory, Nancy, France. His research interests linear and multilinear algebra, optimization, signal and image processing, and machine learning. He has been a member of the editorial board of SIAM Journal on Matrix Analysis and Applications since 2018.



Workshop and general chair of GRETSI 2022 conference.



Tülay Adali (Fellow, IEEE) is a Distinguished University Professor at the University of Maryland Baltimore County (UMBC), Baltimore, MD. Prof. Adali has been active within the IEEE. She served as the Chair of the IEEE Brain Technical Community in 2023, and the Signal Processing Society (SPS) Vice President for Technical Directions 20192022. She is currently the editor-in-chief of the IEEE Signal Processing Magazine. Her roles in conference organization include general/technical chair position for the IEEE Machine Learning for Signal Processing (MLSP) and Neural Networks for Signal Processing (NNNSP) Workshops 20012009, 2014, and 2023, and program/technical chair position for the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) in 2017 and 2026 among other roles. She was the Chair of the NNNSP/MLSP Technical Committee, 20032005 and 20112013, and served on numerous boards and technical committees of the SPS. Prof. Adali is a Fellow of the IEEE, AIMBE, and AAIA, a Fulbright Scholar, an IEEE SPS Distinguished Lecturer, and the UMBC Presidential Research Professor for 2024-2027. She is the recipient of SPS Meritorious Service Award, Humboldt Research Award, IEEE SPS Best Paper Award, SPIE Unsupervised Learning and ICA Pioneer Award, the University System of Maryland Regents' Award for Research, and the NSF CAREER Award. Her current research interests are in statistical signal processing and machine learning, with applications to neuroimaging data analysis.