SIAM J. MATRIX ANAL. APPL.
Vol. 47, No. 3, pp. 1210–1252

IDENTIFIABILITY OF NONNEGATIVE TUCKER DECOMPOSITIONS—PART I: THEORY*

SUBHAYAN SAHA†, GIOVANNI BARBARINO‡, AND NICOLAS GILLIS†

Abstract. Tensor decompositions have become a central tool in data science, with applications in areas such as data analysis, signal processing, and machine learning. A key property of many tensor decompositions, such as the canonical polyadic decomposition, is identifiability: the factors are unique, up to trivial scaling and permutation ambiguities. This allows one to recover the groundtruth sources that generated the data. The Tucker decomposition (TD) is a central and widely used tensor decomposition model. However, it is in general not identifiable. In this paper, we study the identifiability of the nonnegative TD (nTD). By adapting and extending identifiability results of nonnegative matrix factorization, we provide uniqueness results for nTD. Our results require the nonnegative matrix factors to have some degree of sparsity (namely, satisfy the separability condition, or the sufficiently scattered condition), while the core tensor only needs to have some slices (or linear combinations of them) or unfoldings with full column rank (but does not need to be nonnegative). Under such conditions, we derive several procedures, using either unfoldings or slices of the input tensor, to obtain identifiable nTDs by minimizing the volume of unfoldings or slices of the core tensor.

Key words. Tucker decomposition, nonnegavitity, identifiability, separability, sufficiently scattered condition, volume minimization

MSC codes. 15A69, 15A23

DOI. 10.1137/25M1761914

1. Introduction.

Tensors are important algebraic objects that appear in different branches of science such as mathematics, physics, computer science, and chemistry. For our purpose, tensors can be viewed as multidimensional arrays with entries from the underlying field of real numbers $\mathbb{R}$. Following this, order-1 tensors are vectors and order-2 tensors are matrices. For a tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$, its order is the number of dimensions, d, also referred to as *modes,* while the dimension of the kth mode is n_k for $k \in [d] := \{1, 2, \ldots, d\}$.

Tensor decompositions have generated significant interest in recent years due to their applications in different fields such as signal processing, computer vision, chemometrics, neuroscience, and others; see, e.g., [33, 45] for comprehensive surveys, and [9] for a very recent book. One of the central motivations for studying such decompositions and applying them in real-world problems comes from the fact that some models of tensor decompositions are identifiable under very mild assumptions. One such model is the canonical polyadic decomposition (CPD) that decomposes a given tensor as the *minimal* sum of *rank-one* tensors. Such a decomposition is essentially unique, meaning that is unique up to permutations and scalings of the rank-one factors, if the

*Received by the editors May 20, 2025; accepted for publication (in revised form) February 17, 2026; published electronically July 13, 2026.
https://doi.org/10.1137/25M1761914
Funding: The authors acknowledge the support by the European Union (ERC consolidator, eLinoR, no 101085607).
†Department of Mathematics and Operational Research, University of Mons, Rue de Houdain 9, 7000 Mons, Belgium (subhayan.saha@umons.ac.be, nicolas.gillis@umons.ac.be).
‡Departamento de Ciencias Integradas, Facultad de Ciencias Experimentales, Universidad de Huelva, campus El Carmen, Huelva 21007, Spain (giovanni.barbarino@dci.uhu.es). GB is member of the Research Group GNCS (Gruppo Nazionale per il Calcolo Scientifico) of INdAM (Istituto Nazionale di Alta Matematica).

input tensor satisfies some *mild* conditions [37, 15, 29, 36]. As a consequence, there is a lot of research on developing algorithms for the CPD model [7, 49, 27, 35, 22, 32, 31].

In this paper, we focus on another widely-used model of tensor decomposition called the Tucker decomposition (TD) which is a generalization of the CPD. It is a form of *higher-order principal component analysis* which decomposes a tensor into a core tensor and multiple matrices such that the given tensor can be obtained by transforming the core tensor by a matrix along each mode.

Let us focus on order-3 tensors for simplicity. Given an order-3 tensor $\mathcal{G} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$ and matrices $U_i \in \mathbb{R}^{n_i \times r_i}$ for all $i \in [3]$, the following multilinear transformation,

$$\mathcal{T} = (U_1, U_2, U_3).\mathcal{G} \quad \in \quad \mathbb{R}^{n_1 \times n_2 \times n_3}, \tag{1}$$

is defined as

$$\mathcal{T}_{j_1,j_2,j_3} = \sum_{i_t \in [r_t] \text{ for all } t \in [3]} (U_1)_{j_1,i_1} (U_2)_{j_2,i_2} (U_3)_{j_3,i_3} \mathcal{G}_{i_1,i_2,i_3}. \tag{2}$$

The decomposition $\mathcal{T} = (U_1, U_2, U_3).\mathcal{G}$ is called an (r_1, r_2, r_3)-TD of $\mathcal{T}$. Note that for order-2 tensors, that is, for matrices, an (r_1, r_2)-TD corresponds to a matrix trifactorization of the form

$$X = U_1 G U_2^\top, \tag{3}$$

where $U_1 \in \mathbb{R}^{n_1 \times r_1}$, $U_2 \in \mathbb{R}^{n_2 \times r_2}$, and $G \in \mathbb{R}^{r_1 \times r_2}$ is the core matrix.

CPD is a special case of TD when $r_i = r$ for all i while $\mathcal{G}$ is a diagonal tensor, that is, $\mathcal{G}_{i_1,i_2,\dots,i_d} \neq 0$ only when $i_1 = i_2 = \cdots = i_d$. In contrast to CPD, TD of an arbitrary tensor is not essentially unique as it suffers from rotation ambiguities. In fact, for an (r_1, r_2, r_3)-TD of $\mathcal{T} = (U_1, U_2, U_3).\mathcal{G}$, and for any invertible matrices $R_i \in \mathbb{R}^{r_i \times r_i}$ for $i \in [3]$, $(U_1R_1, U_2R_2, U_3R_3).\mathcal{G}'$ is another (r_1, r_2, r_3)-TD for $\mathcal{T}$, where $\mathcal{G}' = (R_1^{-1}, R_2^{-1}, R_3^{-1}).\mathcal{G}$. Hence, one needs to impose additional constraints or regularization on the TD in order to get uniqueness.

Outline and contribution of the paper. In this paper, we focus on nonnegativity constraints on the U_i's. We refer to such decompositions as nonnegative Tucker decompositions (nTDs). We will use NTD, as done in the literature, only when the core tensor, $\mathcal{G}$, is also imposed to be nonnegative. However, in most of our results, we will not need this assumption. If a given tensor has an nTD of a certain dimension, we are interested in the following two questions:

- Under what conditions is such a decomposition *unique*?
- How can we (approximately) *compute* such a decomposition?

Part I of the paper, "Identifiability of nTD: Theory", focuses on the first question, and is divided into 7 sections:

- Section 2 presents the notation, definitions and the preliminary results necessary to present our contributions; in particular, it presents identifiability results of nonnegative matrix factorization (NMF) that are the starting point of our work.
- Section 3 summarizes previous works on the identifiability of nTDs. At the end of this section, we summarize our main identifiability results; see Table 1. (Skipping directly to Table 1 allows the interested reader familiar with tensor notation to have a quick overview of our results.)

- Section 4 deals with order-2 nTDs, that is, nonnegative matrix trifactorizations; see (3). We generalize identifiability results from the NMF literature to order-2 nTDs. Identifiability is obtained via sparsity conditions on the factors. Our first result, Theorem 4.1, relies on the separability condition (Definition 2.5) which requires the factors to contain a diagonal matrix as a submatrix. Our second result, Theorem 4.2, relies on the sufficiently scattered condition (SSC, Definition 2.7), a relaxation of separability, which requires the factors to contain rows sufficiently spread in the nonnegative orthant, and allows one to retrieve the ground truth factors using volume minimization of the core matrix.
- Section 5 deals with order-3 nTDs. We propose 5 procedures to compute order-3 nTDs, with identifiability guarantees (Theorems 5.4, 5.11, 5.14, 5.16, 5.18). They rely on the SSC for the factors, but use different conditions on the core tensor, $\mathcal{G}$. The proofs rely on the identifiability of NMF (section 2.2), and of order-2 nTDs (section 4). The first procedure relies on unfoldings of the input tensor, while the other four on its slices.
- Section 6 generalizes the results for order-3 from section 5 to any order.
- In sections 5 and 6, we sometimes need that the Kronecker product of two or more factors satisfies the SSC. In section 7, we explore sufficient conditions for this property to hold.

In this paper, we only consider identifiability result for exact decompositions. Robustness analysis in the presence of noise is a study of further research. Part II of this paper [44] will focus on the second question: it will provide efficient algorithms to compute identifiable nTDs, and illustrate their effectiveness on synthetic and real data sets.

2. Notation, definitions, and preliminaries.

Let us introduce the notation, definitions, and results from the literature used in this paper.

2.1. Tensors.

We first fix the notion of uniqueness that we consider in this paper.

DEFINITION 2.1 (essential uniqueness of TD). *Let $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ be an order-d tensor which has an $(r_1, \ldots, r_d)$-TD of the form $\mathcal{T} = (U_1^{\#}, \ldots, U_d^{\#}).\mathcal{G}^{\#}$. Let f be an objective function and $\mathcal{S}$ be a feasible set for $(U_1, \ldots, U_d, \mathcal{G})$, and define the optimization problem*

$$\min_{U_1,\ldots,U_d,\mathcal{G}} f(U_1, \ldots, U_d, \mathcal{G}) \quad \text{such that} \quad (U_1, \ldots, U_d, \mathcal{G}) \in \mathcal{S}. \tag{4}$$

Note that f and $\mathcal{S}$ will depend on $\mathcal{T}$. Then the decomposition $\mathcal{T} = (U_1^{\#}, \ldots, U_d^{\#}).\mathcal{G}^{\#}$ with $(U_1^{\#}, \ldots, U_d^{\#}, \mathcal{G}^{\#}) \in \mathcal{S}$ is an essentially unique TD with respect to $(f, \mathcal{S})$ if for any optimal solution of (4), *$(U_1^*, \ldots, U_d^*, \mathcal{G}^*)$, there exist permutation matrices Π_i and diagonal matrices D_i such that $U_i^{\#} = U_i^* \Pi_i D_i$ for all $i \in [d]$ and $\mathcal{G}^{\#} = (D_1^{-1}\Pi_1^\top, \ldots, D_d^{-1}\Pi_d^\top).\mathcal{G}^*$. For simplicity, we will also say that $\mathcal{T} = (U_1^{\#}, \ldots, U_d^{\#}).\mathcal{G}^{\#}$ is essentially unique w.r.t.* (4).

Since we consider nTDs in this paper, we will always have

$$\mathcal{S} \subseteq \{(U_1, \ldots, U_d, \mathcal{G}) \mid U_i \geq 0 \text{ for } i \in [d]\}.$$

Let us now define slices and unfoldings of a tensor.

DEFINITION 2.2 (slices of an order-3 tensor). *Let $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ be an order-3 tensor. Then the n_3 slices along the third mode, denoted by $\mathcal{T}_k^{(3)} \in \mathbb{R}^{n_1 \times n_2}$ for $k \in [n_3]$, are matrices given by*

$$\left(\mathcal{T}_k^{(3)}\right)_{i,j} = \mathcal{T}_{i,j,k} \quad \text{for all } i,j. \tag{5}$$

One can similarly define the slices along the first and second modes and denote them by $\mathcal{T}_i^{(1)} \in \mathbb{R}^{n_2 \times n_3}$ and $\mathcal{T}_j^{(2)} \in \mathbb{R}^{n_1 \times n_3}$, respectively.

DEFINITION 2.3 (unfoldings of an order-3 tensor). *Let $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ be an order-3 tensor. Then the unfolding of $\mathcal{T}$ along the third mode is given by the matrix $\mathcal{T}_{(3)} \in \mathbb{R}^{n_1 n_2 \times n_3}$, where*

$$(\mathcal{T}_{(3)})_{(i,j),k} = \mathcal{T}_{i,j,k} \quad \text{for all } i,j,k,$$

where we use the notation $(i,j) = i + (j-1)n_1$. One can similarly define the unfoldings along the first and second modes and denote them by $\mathcal{T}_{(1)} \in \mathbb{R}^{n_2 n_3 \times n_1}$ and $\mathcal{T}_{(2)} \in \mathbb{R}^{n_1 n_3 \times n_2}$, respectively.

These notions can be extended to higher-order tensors; see section 6.

Let us define the Kronecker product between two matrices.

DEFINITION 2.4 (Kronecker products). *Let $A \in \mathbb{R}^{m \times M}$ and $B \in \mathbb{R}^{n \times N}$ be matrices with columns $a_1, \dots, a_M$ and $b_1, \dots, b_N$, respectively. Then the Kronecker product between A and B, denoted by $A \otimes B$, is the $mn \times MN$ matrix with column at position $p = j + (i-1)N$ given by*

$$(A \otimes B)_{:,p} = a_i \otimes b_j = \operatorname{vec}(a_i b_j^\top) = [a_i b_j(1); a_i b_j(2); \dots; a_i b_j(n)] \in \mathbb{R}^{mn},$$

where vec *vectorizes a matrix columnwise.*

2.2. NMF: concepts and identifiability. To prove our identifiability results for nTD, we will heavily rely on tools and results from the NMF literature. Recall that, in the exact case, NMF is the following problem: given a nonnegative matrix $X \in \mathbb{R}_+^{m \times n}$ and a factorization rank r, find $W \in \mathbb{R}_+^{m \times r}$ and $H \in \mathbb{R}_+^{n \times r}$ such that $X = WH^\top$. Note that NMF decompositions are typically not unique unless additional constraints or regularization are used; we refer the interested reader to the survey [18] and the book [23, Chapter 4] for more details.

Cones. For any matrix $A \in \mathbb{R}^{m \times n}$, the convex cone generated by the columns of A, denoted cone(A), is given by the set of *conic* combinations, that is, linear combinations with nonnegative weights, of the columns of A:

$$\operatorname{cone}(A) = \{x \mid x = Ay \text{ for } y \in \mathbb{R}^n, y \geq 0\}.$$

The dual cone of A, denoted by cone$^*(A)$, is given by

$$\begin{aligned} \operatorname{cone}^*(A) &= \left\{y \mid x^\top y \geq 0 \text{ for all } x \in \operatorname{cone}(A)\right\} \\ &= \left\{y \mid z^\top A^\top y \geq 0 \text{ for all } z \geq 0\right\} = \left\{y \mid A^\top y \geq 0\right\}. \end{aligned} \tag{6}$$

Separability. The notion of separability dates back to the hyperspectral community under the pure-pixel assumption [11]. The terminology was introduced by Donoho and Stodden [16], and it was later used by Arora et al. [6] to obtain unique and polynomial-time solvable NMF problems; see [23, Chapter 7] for a survey on separable NMF.

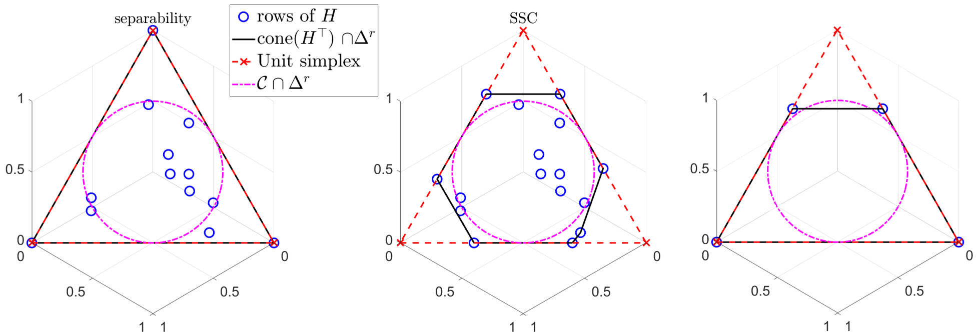

FIG. 1. *Comparison of separability (left) and the SSC (middle) for the matrix H whose columns lie on the probability simplex, $\Delta^r = \{x \mid x \geq 0, e^\top x = 1\}$, in the case $r = 3$. On the left, separability requires the columns of H to contain the unit vectors, that is, $H(:,\mathcal{K}) = I_r$ for some $\mathcal{K}$. In the middle, the SSC requires $\mathcal{C} \subseteq \text{cone}(H)$. On the right, the SSC is not satisfied but the NMF solution is unique, illustrating a case when the SSC is not necessary. Figure adapted from* [1].

DEFINITION 2.5. *A matrix $H \in \mathbb{R}_+^{n\times r}$ is called separable if there exists an index set $\mathcal{K} \subseteq [n]$, where $|\mathcal{K}| = r$ such that $H(\mathcal{K},:)$ is a diagonal matrix with positive diagonal elements.*

Equivalently, a matrix H is *separable* if the convex cone generated by its rows spans the entire nonnegative orthant, that is, $\text{cone}(H^\top) = \mathbb{R}_+^r$; see Figure 1 for an illustration when $r = 3$.

We say that X admits a separable NMF (W,H) of size r if there exists a decomposition of X of the form $X = WH^\top$ of size r such that $H \in \mathbb{R}^{n\times r}$ is a separable matrix. By the scaling degree of freedom, this implies that there exists a decomposition $X = W'H'^\top$, where $W' = X(:,\mathcal{K})$, that is, the columns of W are a subset of the columns of X.

THEOREM 2.6 ([6]). *Let $X = WH^\top \in \mathbb{R}^{m\times n}$ be a separable NMF of size*[1] *$r = \text{rank}(X)$. Then for any other separable NMF (W_*, H_*) of size r of X, there exists a permutation matrix Π and diagonal matrix D such that $W_* = WD\Pi$ and $H_* = HD^{-1}\Pi$.*

The proof of Theorem 2.6 directly follows from the fact that the extreme rays of $\text{cone}(X)$ coincide with that of $\text{cone}(W)$ under the separability assumption. Given a matrix X and rank r, there exist fast and robust polynomial-time algorithms to compute a separable NMF: they identify the extreme rays of $\text{cone}(X)$ [23, Chapter 7]; we will generalize such algorithms to order-2 nTDs in Part II of this paper [44].

SSC. The separability assumption is relatively strong. To relax it, a crucial notion is the SSC, which was introduced in [28].

DEFINITION 2.7 (SSC[2]). *A matrix $H \in \mathbb{R}_+^{n\times r}$ with $r \geq 2$ satisfies the SSC if the following two conditions hold:*

1. *SSC1: $\mathcal{C} = \{x \in \mathbb{R}_+^r \mid e^\top x \geq \sqrt{r-1}\|x\|_2\} \subseteq \text{cone}(H^\top)$, where e is the vector of all ones of appropriate dimension.*

[1] This can be relaxed to $\text{cone}(X)$ having r rays. We use the rank here for simplicity. Note also that W does not need to be nonnegative; see Remark 2.9.

[2] Slight variants of the SSC exist in the literature. Refer to Section 4.2.3.1 in [23] for a more detailed description and the relation between these variants.

2. *SSC2:* $\text{cone}^*(H^\top) \cap \boldsymbol{bd}(\mathcal{C}^*) = \{\lambda e_k \mid \lambda \geq 0$ *and* $k \in [r]\}$, *where* $\boldsymbol{bd}$ *denotes the border of a cone,* $\mathcal{C}^* = \{x \in \mathbb{R}^r \mid e^\top x \geq \|x\|_2\}$, *and* e_k *is the* k*th unit of appropriate dimension.*

SSC1 requires that $\text{cone}(H^\top)$ contains the ice-cream cone that is tangent to every facet of the nonnegative orthant; see Figure 1 for an illustration. Equivalently, by using the definition of a dual cone (6), one can rewrite the condition in SSC1 as $\text{cone}^*(H^\top) \subseteq \mathcal{C}^*$. SSC2 is typically satisfied if SSC1 is, and allows one to avoid pathological cases; see [23, Chapter 4.2.3] for more details.

Note that separability implies the SSC, that separability and the SSC coincide when $r \leq 2$, and that the SSC implies that a matrix is full rank.

To leverage the SSC, we need a notion of volume. Given a matrix $W \in \mathbb{R}^{m\times r}$ with $\text{rank}(W) = r$, the quantity $\det(W^\top W)$ is a measure of the volume of the columns of W; more precisely, it is the volume of the convex hull of the columns of W and the origin in the linear subspace spanned by the columns of W. Combining volume minimization of W under the SSC of H leads to unique NMFs. In fact, let us consider the following optimization problem, referred to as a minimum-volume (min-vol) NMF:

$$(7) \quad \min_{W\in\mathbb{R}^{m\times r}, H\in\mathbb{R}^{n\times r}} \det(W^\top W) \quad \text{such that} \quad X = WH^\top, H^\top e = e, \text{ and } H \geq 0.$$

We have the following identifiability result.

Theorem 2.8 (identifiability of min-vol NMF [17]). *Let* $X \in \mathbb{R}^{m\times n}$ *admit the decomposition* $X = W_\# H_\#^\top$, *where* $H_\# \in \mathbb{R}_+^{r\times n}$ *satisfies the SSC and* $r = \text{rank}(X)$. *Then* $X = W_\# H_\#^\top$ *is essentially unique w.r.t.* (7).

In simple terms, the SSC of H in an NMF $X = WH^\top$ with $H\top e = e$ implies that there exist no other factorization where the first factor has a smaller volume. This intuition was introduced in the hyperspectral unmixing literature [13]. Note that the first variant of identifiability for min-vol NMF under the SSC was proved in [21, 39] under the constraint $He = e$, which is not without loss of generality (w.l.o.g.) since normalizing the rows of H makes the column of X belong to the convex hull of the columns of W. Later it was relaxed to the normalization $H^\top e = e$ in [17], and to $W^\top e = e$ in [38].

Remark 2.9 (nonnegativity of X and W). The nonnegativity of X and W is not necessary for Theorems 2.6 and 2.8 to hold. Hence, there is a slight abuse of language when referring to separable NMF and min-vol NMF since, in such decompositions, W could potentially have negative entries. The reason is that these models appeared in the NMF literature, hence authors kept the name NMF, although using the term semi-NMF would have been more appropriate. We refer the interested reader to the discussions in [23, Chapter 4] for more details.

Necessity of the SSC condition. Let us show an example where the SSC is not necessary for NMF to have a unique solution. Let

$$H = \begin{bmatrix} 0 & 0 & \omega & 0 \\ 1 & 0 & 0 & \omega \\ 0 & 1 & 1-\omega & 1-\omega \end{bmatrix}.$$

For $\frac{\sqrt{2}}{3} < \omega \leq 1$, one can show that H does not satisfy the SSC condition; see Figure 1 for an illustration. Let $X = I_r H$, and let (W', H') be another NMF of X, where the columns of W' and H' have been normalized to the ℓ_1 norm. Since $\text{conv}(X) \subseteq$

conv(W'), two columns of W' must be $e_2 = (0,1,0)$ and $e_3 = (0,0,1)$. W.l.o.g., take them to be the first two columns. Let the third column be $c = (c_1, c_2, c_3) \in \mathbb{R}^3_+$. The third column of X, $(\omega, 0, 1-\omega)$, is formed as a nonnegative linear combination of c and e_3, implying $c_3 = 0$. Similarly, one can conclude that $c_2 = 0$ and, hence, $c_1 = 1$. This illustrates that the SSC is not necessary for identifiability of NMF.

However, SSC is among the weakest assumptions to guarantee uniqueness of NMF, and related matrix factorization problems. Moreover, minimizing the volume of W leads to a certain degree of sparsity of H [23, section 4.3.4.3], and the SSC has been used successfully in many different applications such as hyperspectral unmixing, spectroscopy, topic modeling, and audio source separation [23, section 4.3.4].

3. Previous works on the identifiability of nTD.

As for NMF, the nonnegativity constraints in nTD allow for a meaningful interpretation of the factors which may have a probabilistic or a physical meaning. In fact, nTD has been used in many applications, including clustering [40], hyperspectral image denoising and compression [8, 25], audio pattern extraction [41, 46], image fusion [54], community detection in multi-layer networks [52, 2], topic modeling [4], and EEG signal analysis [42, 53]; see also [12] and the references therein. We will present several applications in Part II of this paper [44], as well as standard algorithms to compute nTDs. Let us now briefly describe previous results on the identifiability of nTD.

Connections between NTD and NMF. Zhou et al. [56] have explored the connections between the identifiability of NTD and NMF. They stated that if there exists a unique NMF of the unfoldings along all modes of the tensor, then the NTD of the given tensor is essentially unique. Moreover, they also look at the so-called population NTD model where one factor matrix is the identity matrix. In that case, they showed that uniqueness of the NMF of the unfolding along the mode corresponding to the factor matrix which is the identity translates to uniqueness of the NTD as well.

As we will see in the rest of this section, further research in this direction has focused on imposing stricter conditions on the factors and the core tensor on the underlying NTD of the given tensor in order to achieve identifiability results.

Separability. Agterberg and Zhang [2] studied the identifiability of order-3 nTD where $U_i \in \mathbb{R}^{n_i \times r_i}_+$ are separable for $i \in [3]$. Their motivation comes from the interpretability of this model as a tensor version of the mixed-membership stochastic blockmodel [3]. Each factor, U_i, provides the membership of r_i communities; more precisely, $U_i(:,j)$ is the membership vector of the jth community along the ith mode, with the magnitude of the entry representing the intensity of membership within that community. The core tensor, $\mathcal{G}$, provides the relationship between the communities; $\mathcal{G}(i,j,k)$ indicates the magnitude of the interactions between three communities: the ith in mode 1, the jth in mode 2 and the kth in mode 3. Separability of the U_i's in this context requires that, for each community, there exists a pure node that only belongs to that community. Agterberg and Zhang [2] proved the following result.

THEOREM 3.1 ([2, Proposition 1]). *Let $\mathcal{T} = (U_1, U_2, U_3).\mathcal{G} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ be an order-3 tensor where $U_i \in \mathbb{R}^{n_i \times r_i}_+$ are separable for $i \in [3]$. If, for all $k \in [3]$, the unfolding of the tensor $\mathcal{T}$ along the kth mode, that is, $\mathcal{T}_{(k)}$, has rank r_k, then for any other decomposition of the form $\mathcal{T} = (U_1', U_2', U_3').\mathcal{G}'$, where the U_i' are separable for $i \in [3]$, there exist permutation matrices Π_1, Π_2, Π_3 and diagonal matrices D_1, D_2, D_3 such that $U_k = U_k' D_k \Pi_k$ for all $k \in [3]$ and $\mathcal{G} = (\Pi_1^\top D_1^{-1}, \Pi_2^\top D_2^{-1}, \Pi_3^\top D_3^{-1}).\mathcal{G}'$.*

The proof essentially follows the identifiability of separable NMFs (Theorem 2.6), and can actually be generalized to any order; see Theorem A.1 in Appendix A.

Moreover, in section A.2, we show that one can replace the rank assumption on the unfoldings along all modes with a rank assumption on the unfolding along only one mode (although this requires some restrictions on the dimensions of the core tensor).

SSC. Sun and Huang [47] studied the identifiability of nTD, where $U_i \in \mathbb{R}_+^{n_i \times r_i}$ satisfy the SSC for $i \in [d]$. They considered the following optimization problem,

(8)
$$\max_{\mathcal{G}, U_1, \ldots, U_d} \sum_{i=1}^{d} \text{logdet}(U_i^\top U_i) \text{ such that } U_i \geq 0, U_i^\top e = e \text{ for } i \in [d], \mathcal{T} = (U_1, \ldots, U_d).\mathcal{G},$$

where they maximize the volume of the factors of the nTD. They claimed that if there exists a decomposition of the core tensor $\mathcal{T}$ such that all U_i's satisfy the SSC, then the nTD of $\mathcal{T}$ obtained from an optimal solution to (8) recovers the U_i's, up to permutation [47, Theorem 1]. Unfortunately, in this conference version of the paper [47], the authors do not discuss the crucial assumptions regarding the core tensor (nor provide a proof). Hence, we provide here a complete result, whose proof follows directly from max-vol NMF identifiability on each unfolding [50, Theorem 7.1].

THEOREM 3.2. *Let $\mathcal{T} = (U_1^{\#}, \ldots, U_d^{\#}).\mathcal{G}^{\#}$, where the $U_i^{\#} \in \mathbb{R}^{n_i \times r_i}$ satisfy the SSC, and the unfoldings of $\mathcal{G}^{\#}$ along all modes $k \in [d]$, denoted by $G_{(k)}^{\#}$, have rank r_k. Then $\mathcal{T} = (U_1^{\#}, \ldots, U_d^{\#}).\mathcal{G}^{\#}$ is essentially unique w.r.t.* (8).

The *d*-persistent topic models. Anandkumar et al. [4] discussed the notion of identifiability of the d-persistent topic models. For a d-persistent topic, the $2d$th-order moment of the model, denoted M_d, is an unfolding of an order-$2d$ tensor which has a decomposition of the form

$$M_{2d} = (A^{\odot d})\, G\, (A^{\odot d})^\top,$$

where $A^{\odot d}$ is the Khatri–Rao product of the matrix A taken d times and G is the covariance matrix of the *hidden variables*. This is a special case of a TD, and a generalization of the CPD. They showed that the columns of A are identifiable under some nontrivial conditions. Using this tri-NMF structure, one could further enquire the identifiability of the nonnegative version of the $2d$-persistent topic model using the framework introduced in this paper. This is left for future work.

How our work goes beyond the state of the art. In this paper, we will exploit the structure of the slices of the tensor and of its unfoldings, and will *minimize the volume of the core tensor*. Our modeling approach has several advantages:

- When using unfoldings, instead of enforcing conditions on the unfoldings of the tensor along all the modes, as in previous works (in particular Theorems 3.2 and 3.1, and in [56]), we restrict our attention to the unfolding along just a single mode, leading to stronger identifiability results; see section 5.1 for order-3 tensors, and section 6.1 for higher orders.
- When using slices, the dimensions of the matrix factorization subproblems that we will solve will be significantly smaller compared to previous works—we will only need to consider the factorization of $d-1$ slices. We will show that, under certain restrictions on the factors and the core tensor, the slices have an essentially unique factorization which will imply the identifiability of the underlying nTD; see section 5.2 for order-3 tensors, and section 6.2 for higher orders.

TABLE 1

Conditions on the U_i's, $\mathcal{G}$, and $\mathcal{T}$ to have an identifiability procedure for order-3 nTD. The matrix $\mathcal{G}_{(3)}$ is the third mode unfolding of $\mathcal{G}$. The matrix $\mathcal{T}_i^{(k)}$ is the ith slice along the kth mode. Determ. means deterministic, rand. means randomized, $\mathcal{V}^{(j)}$ is the linear span of the slices along mode j, defined as $\mathcal{V}^{(j)} := \{\sum_{i=1}^{n_j} \alpha_i T_i^{(j)} \mid \alpha_i \in \mathbb{R}_+\}$, max-rank$(\mathcal{V}^{(j)})$ is the maximum r such that there exists $A \in \mathcal{V}^{(j)}$ with $\operatorname{rank}(A) = r$.

$\mathcal{T} = (U_1, U_2, U_3).\mathcal{G}$	$\mathcal{G}$	U_i's	$\mathcal{T}$	
Procedure 0 (Theorem 5.4)	$\operatorname{rank}(\mathcal{G}_{(3)}) = r_3 = r_1 r_2$	$U_1 \otimes U_2$ is SSC U_3 is SSC		determ.
[47] (Theorem 3.2)	$\operatorname{rank}(\mathcal{G}_{(i)}) = r_i \quad \forall i$	U_i's are SSC		deterem.
Procedure 1 (Theorem 5.11)	$r_3 \le r_1 = r_2 = r$	U_i's are SSC	$\exists i_2$ s.t. $\operatorname{rank}(\mathcal{T}_{i_2}^{(2)}) = r_3$ $\exists i_3$ s.t. $\operatorname{rank}(\mathcal{T}_{i_3}^{(3)}) = r_1$	determ.
Procedure 2 (Theorem 5.14)	$r_3 \le r_1 = r_2 = r$	U_i's are SSC	max-rank$(\mathcal{V}^{(3)}) = r$ max-rank$(\mathcal{V}^{(2)}) = r_3$	rand.
Procedure 3 (Theorem 5.16)	$\sqrt{r_3} \le r_1 = r_2$ $\operatorname{rank}(\mathcal{G}_{(3)}) = r_3$	U_i's are SSC	$\exists i$ s.t. $\operatorname{rank}(\mathcal{T}_i^{(3)}) = r_1$	determ.
Procedure 4 (Theorem 5.18)	$\sqrt{r_3} \le r_1 = r_2$ $\operatorname{rank}(\mathcal{G}_{(3)}) = r_3$	U_i's are SSC	max-rank$(\mathcal{V}^{(3)}) = r$	rand.

Table 1 summarizes the main cases in which we can prove identifiability for order-3 nTDs, where the roles of the modes are interchangeable. These results generalize to any order; see section 6.

Comparison with [47]. From a theoretical point of view, Theorem 3.2 proceeds via applying the identifiability of max-vol NMF for the unfoldings along each of the modes, and each unfolding of $\mathcal{T}$, denoted by $\mathcal{T}_{(j)}$, needs to have rank r_j. This translates to the fact that the multilinear rank of the core tensor $\mathcal{G}$ is full across every mode, that is, $\operatorname{rank}(\mathcal{G}_{(j)}) = r_j$ for all $j \in [3]$. From a practical point of view, the objective function can be hard to optimize and tune. In fact, maximizing $\operatorname{logdet}(\cdot)$ is much harder than minimizing it, because $\operatorname{logdet}(\cdot)$ is easily majorized with a quadratic [20]; see [50] for a discussions on this issue. Moreover, in the presence of noise, one needs to properly weight the d logdet terms in the objective, along with the reconstruction error, $\|\mathcal{T} - (U_1, \ldots, U_d).\mathcal{G}\|$. In [47], the authors propose an algorithm using successive convex approximations. The numerical schemes for our procedures will be more easily implementable as we will minimize the volume of $\mathcal{G}$ using a single regularization term. Part II of this paper [44] will discuss further the algorithmic details of the various approaches.

Note that Procedure 2 (resp., 4) are randomized versions of Procedure 1 (resp., 3) with weaker assumptions: instead of requiring two (resp. one) slices to have maximum rank, we just need the span of all the slices to have maximum rank. If the given tensor has certain guarantees on the slices along a certain mode, say mode 3, the crucial requirement for Procedures 14 is to have the dimensions along the other two modes to be equal, that is, $r_1 = r_2 = r$. For Procedures 1 and 2, the dimension of $\mathcal{G}$ along the third mode, that is, r_3, can be at most r, whereas for Procedures 3 and 4, it can be up to r^2. If the dimensions along modes 1 and 2 are not equal, one can resort to Procedure 0 but under the condition that $r_3 = r_1 r_2$. Finally, our approaches also have algorithmic advantages over previous works. We defer the details to Part II of this paper [44].

4. Order-2 nTD also known as nonnegative matrix trifactorization. Given a data matrix $X \in \mathbb{R}_+^{m\times n}$, order-2 nTD decomposes X into a product of three factor matrices $U_1 \in \mathbb{R}_+^{m\times r}$, $G \in \mathbb{R}_+^{r\times r}$, and $U_2 \in \mathbb{R}_+^{n\times r}$ such that $X = U_1 G U_2^\top$, which is also known as nonnegative matrix trifactorization [14]. When G is the identity matrix, order-2 nTD boils down to NMF. Otherwise, order-2 nTD can never be identifiable; in fact, we have

$$U_1\, G\, U_2^\top \;=\; U_1\, I_r\, (U_2 G^\top)^\top \;=\; (U_1 G)\, I_r\, U_2^\top .$$

Therefore, we must use additional constraints and/or regularization to make order-2 nTD identifiable. In practice, people used sparsity constraints; see, e.g., [14, 51, 55, 34].

Surprisingly, although order-2 nTDs have been extensively used in practice, there exists, to the best of our knowledge, no identifiability results in this case. This section aims to fill in this gap. We discuss the case of separability in section 4.1, and the case of the SSC in section 4.2.

4.1. Order-2 nTD under separability. The notion of separable NMF can be extended from two to three factors and such decompositions have been extensively studied in the matrix factorization literature in the symmetric case. Given a symmetric matrix $A \in \mathbb{R}^{m\times m}$, one can consider its decomposition $A = WSW^\top$, sometimes referred to as trisymmetric NMF. This decomposition occurs in community detection in graphs, and is closely related to the stochastic block models: the matrix W contains the community information, and the matrix S the interactions between the communities; see, e.g., [18] and the references therein. Arora et al. [5] gave a provably robust polynomial-time algorithm for trisymmetric NMF with identifiability guarantees. They used it in the context of topic modeling, where A is the word cooccurence matrix, the matrix W is the word-by-topic matrix such that $W(i,k)$ is the probability for word i to be picked under the topic k, and the matrix S is a topic-by-topic matrix that accounts for the interactions between the topics. In topic modeling, the following separability assumption makes sense: for each topic, there exists an anchor word, that is, a word that is only used by that topic. Following Definition 2.5, this is equivalent to requiring that the matrix $W^\top$ is separable, that is, there exists $\mathcal{K}$ such that $W(\mathcal{K},:)$ is a diagonal matrix. Moreover, to eliminate the scaling degree of freedom, one can assume w.l.o.g. that $W^\top e = e$.

In this section, we present a generalization of the symmetric model to the nonsymmetric case. In the noiseless setting, we assume that there exist matrices $U_1 \in \mathbb{R}_+^{n_1\times r_1}$, $G \in \mathbb{R}^{r_1\times r_2}$, $U_2 \in \mathbb{R}_+^{n_2\times r_2}$ such that $X = U_1 G U_2^\top$. Moreover, we assume that U_1 and U_2 are separable. This model can be interpreted in the following way from the point of community detection: the matrices U_1 and U_2 are the membership indicator matrices, where $U(j_1,k_1)$ is the membership indicator of item j_1 for community k_1 and $U_2(j_2,k_2)$ is the membership indicator of item j_2 for community k_2. The entry $G(k_1,k_2)$ is the strength between the communities k_1 and k_2. Since $U_1 \neq U_2$ in general, U_1 and U_2 correspond to different communities. For example, if X is a movie-by-user matrix (e.g., $X(i,j) = 1$ if user j has watched movie i, and $X(i,j) = 0$ otherwise), U_1 will correspond to communities of movies while U_2 to communities of users. Separability of U_1 (and similarly for U_2) requires that, for each community (that is, for each column of U_1), there exist a node that is only associated to that community. Such nodes are referred to as pure nodes or anchor nodes.

In the following theorem, we show that when $r_1 = r_2 = r = \operatorname{rank}(X)$ and under the separability assumption, such a decomposition is *essentially unique.*

THEOREM 4.1. *Let $X = U_1 G U_2^\top$, where $U_1 \in \mathbb{R}_+^{n_1 \times r}$ and $U_2 \in \mathbb{R}_+^{n_2 \times r}$ are separable matrices, and $r = \operatorname{rank}(X)$. For any other decomposition of $X = U_1^* G^* U_2^{*\top}$ of size r, where U_1^* and U_2^* are separable, there exist permutation matrices Π_1, Π_2 and diagonal matrices D_1, D_2 such that $U_i^* = U_i D_i \Pi_i$ for $i \in [2]$, and $G^* = \Pi_1^\top D_1^{-1} G D_2^{-1} \Pi_2$.*

Proof. Let $X = WH^\top$ with $W = U_1 G$ and $H = U_2$ be an NMF of X of size r, and $X = W_* H_*^\top$ with $W_* = U_1^* G^*$ and $H = U_2^*$ be an another one. Since, U_2 and U_2^* are separable matrices, using Theorem 2.6, there exist a permutation matrix Π_2 and a diagonal matrix D_2 such that $U_2^* = U_2 D_2 \Pi_2$. Using the same argument on $X^\top = U_2 G^\top U_1^\top = U_2^* G^{*\top} U_1^{*\top}$, we obtain that $U_1^* = U_1 D_1 \Pi_1$ for some permutation matrix Π_1 and a diagonal matrix D_1. Finally, $X = U_1 G U_2^\top = U_1^* G^* U_2^{*\top}$ implies that $G^* = \Pi_1^\top D_1^{-1} G D_2^{-1} \Pi_2$ since all factor matrices (U_1, U_2, U_2^*, U_2^*) are separable, hence have rank r. ∎

Note that this result is not applicable when $r_1 \neq r_2$. This is because G is rank deficient, since $\operatorname{rank}(G) \leq \min(r_1, r_2)$. However, it can be easily extended if cone(X) has r_1 rays and cone$(X^\top)$ has r_2 rays; similarly as for separable NMF (see Footnote 1). However, to keep it simple, we restrict our analysis to the full-rank case.

In part II of this paper [44], we will extend the ideas from [5] to give a polynomial-time robust algorithm for computing separable order-2 nTDs.

4.2. Order-2 nTD under the SSC. Separability of both U_1 and U_2 is rather strong. For example, in terms of community detection, it requires all communities to have a pure node. In order to provide a relaxation to the notion of separability, we now generalize min-vol NMF under the SSC (Theorem 2.8) for order-2 nTD.

We propose the following formulation: Given the matrix $X \in \mathbb{R}^{n_1 \times n_2}$ and the factorization rank $r = \operatorname{rank}(X)$,

$$\min_{U_i \in \mathbb{R}^{n_i \times r} i \in [2], G \in \mathbb{R}^{r \times r}} |\det(G)| \quad \text{such that} \quad X = U_1 G U_2^\top, U_i^\top e = e, U_i \geq 0 \text{ for } i \in [2]. \tag{9}$$

We refer to this problem as min-vol order-2 nTD. The constraints $U_i^\top e = e$ for $i \in [2]$ remove the scaling ambiguity. Let us provide an identifiability result for (9) under the SSC. This is a generalization of the symmetric case studied in [26, 19], where X is assumed to be a symmetric matrix and $U_1 = U_2$.

THEOREM 4.2. *Let $X = U_1^\# G^\# U_2^{\#\top}$ be a matrix of rank r where $U_i^\# \in R^{n_i \times r}$ satisfy the SSC and $U_i^{\#\top} e = e$ for $i \in [2]$. Then $X = U_1^\# G^\# U_2^{\#\top}$ is essentially unique w.r.t. (9).*

Proof. Since $\operatorname{rank}(X) = r$ and $X = U_1^\# G^\# U_2^{\#\top} = U_1^* G^* U_2^{*\top}$, we have $\operatorname{rank}(U_1^\#) = \operatorname{rank}(U_1^*) = \operatorname{rank}(U_2^*) = \operatorname{rank}(U_2^\#) = r$. Hence there exist invertible matrices $A, B \in \mathbb{R}^{r \times r}$ such that $U_1^* = U_1^\# A$, $U_2^* = U_2^\# B$, and $G^* = A^{-1} G^\# B^{-\top}$. Since (U_1^*, G^*, U_2^*) is an optimal solution to (9), it follows that

$$|\det(G^*)| = \left|\det\left(A^{-1} G^\# B^{-1}\right)\right| \leq \left|\det\left(G^\#\right)\right|. \tag{10}$$

This gives us that

$$|\det(A)| \; |\det(B)| \geq 1. \tag{11}$$

Moreover, since $(U_1^\#, G^\#, U_2^\#)$ and (U_1^*, G^*, U_2^*) are feasible solutions to (9),

$$e = U_1^{*\top} e = A^\top U_1^{\#\top} e = A^\top e \;\text{ and }\; e = U_2^{*\top} e = B^\top U_2^{\#\top} e = B^\top e. \tag{12}$$

Moreover, $U_1^* = U_1^{\#} A \geq 0$ and $U_2^* = U_2^{\#} B \geq 0$. Following the definition of dual cones in (6), this gives us that for all $i,j \in [r]$,

$$A[:,j] \in \text{cone}^* \left(U_1^{\#\top}\right) \text{ and } B[:,i] \in \text{cone}^* \left(U_2^{\#\top}\right), \tag{13}$$

where $A[:,j]$ denotes the jth column of A. Since $U_1^{\#}$ and $U_2^{\#}$ satisfy the SSC, we have for all $i,j \in [r]$

$$A[:,j], B[:,i] \in \mathcal{C}^* = \left\{x \in \mathbb{R}_+^r \mid e^\top x \geq \|x\|_2\right\}. \tag{14}$$

Then, using (14) and (12), we get

$$|\det(A)| \leq \prod_{j=1}^{r} \|A(:,j)\|_2 \leq \prod_{j=1}^{r} e^\top A(:,j) = 1, \tag{15}$$

and similarly for B. Using (11) and (15), we obtain $|\det(A)| = |\det(B)| = 1$. Since the equality holds in (15), $A(:,j), B(i,:) \in \mathbf{bd}(\mathcal{C}^*)$. This, along with (13) and the definition of SSC, gives us that for all $i,j \in [r]$,

$$A[:,j], B[i,:] \in \{\lambda e_k | \lambda \geq 0 \text{ and } k \in [r]\}.$$

Since $A^\top e = e$, $B^\top e = e$ and $|\det(A)| = |\det(B)| = 1$, A and B are permutation matrices. ∎

The proof of the previous theorem implies that identifiability is actually achieved by any feasible solution of (9) whose objective function value is smaller than that of the ground truth. Let us state this result explicitly as it will be useful later on.

COROLLARY 4.2.1. *Let $X = U_1^{\#} G^{\#} U_2^{\#\top}$ be a matrix of rank r, where $U_1^{\#}$ and $U_2^{\#}$ satisfy the SSC and $U_i^{\#\top} e = e$ for $i \in [2]$. Let $U_1^* \in \mathbb{R}_+^{m\times r}, U_2^* \in \mathbb{R}_+^{n\times r}$, and $G^* \in \mathbb{R}^{r\times r}$ be matrices such that X has a decomposition of the form $X = U_1^* G^* U_2^{*\top}$, where $U_i^{*\top} e = e$ for $i \in [2]$ and $|\det(G^*)| \leq \left|\det\left(G^{\#}\right)\right|$. Then there exist permutation matrices Π_1, Π_2 such that $U_1^* = U_1^{\#}\Pi_1$, $U_2^* = U_2^{\#}\Pi_2$, and $G^* = \Pi_1^\top G^{\#} \Pi_2$.*

Can the above result be extended to the case when $r_1 \neq r_2$? As discussed in the previous section, this corresponds to a *rank-deficient* case. The characterization from (3) does not work because $\det(G) = 0$. For $r_1 \leq r_2$, one could also try to minimize $\det\left(G^\top G\right)$ instead, and the proof of Theorem 4.2 works for that function when $r_1 = r_2$, since $\det\left(G^\top G\right) = \det(G)^2$ for a square matrix G. The proof does not extend directly for the rank-deficient case: in fact, there exist matrices A, B such that

$$\det(B^{-\top} G^\top A^{-\top} A^{-1} G B^{-1}) < \det(G^\top G)$$

while $\det(A)\det(B) < 1$. Take for example $B = 1, A = \left[\begin{smallmatrix} 1 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{smallmatrix}\right]$, $G = \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$. We have $C := A^{-1}G = \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$, $\det(C^\top C) = 1 = \det(G^\top G)$ but $\det(A)^2 = \frac{1}{4} < 1$. Note that here A satisfies $A^\top e = e$.

It would be possible to relax the condition $r_1 \neq r_2$ using the approach from [1] that relies on sparsity conditions, and would not require G to be square and full rank. This is a topic for further research.

5. Order-3 nTDs. In this section, we provide 5 different approaches to obtain identifiable order-3 nTDs. Each approach relies on different assumptions and optimization models. We first consider unfoldings for which we provide a natural approach using directly order-2 nTD identifiability under the SSC (Theorem 4.2); see section 5.1. Then we propose 4 approaches relying on slices; see section 5.2.

These 5 approaches can be generalized to higher-order tensors, which we explain in section 6. We start with order-3 tensors to make the presentation simpler and help the reader follow the line of thoughts more easily.

From now on, we focus on the SSC using minimum-volume minimization. However, all identifiability results presented in this section directly specialize to the separable case, since separability implies the SSC. To avoid redundancy and keep the paper more concise, we do not explicitly state these identifiability results in the separable case. The reason we specified the identifiability for separable order-2 nTD (Theorem 4.1) is twofold:

- This provided an example of identifiability for nTD under separability, and
- we will design a polynomial-time and robust algorithm for separable order-2 nTD in Part II of this paper [44]. It generalizes the algorithm for the symmetric case proposed in [5]. We will use this algorithm to initialize higher-order nTD algorithms.

Throughout this section, we will assume the tensor follows the following assumption.

Assumption 5.1. The order-3 tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ satisfies the following conditions:

1. $\mathcal{T}$ has an nTD of the form $\mathcal{T} = (U_1^\#, U_2^\#, U_3^\#).\mathcal{G}^\#$, where $U_i^\# \in \mathbb{R}_+^{n_i \times r_i}$ for all $i \in [3]$ and $\mathcal{G}^\# \in \mathbb{R}^{r_1 \times r_2 \times r_3}$, and no columns of the $U_i^\#$'s are zero.
2. To simplify the presentation and remove the scaling ambiguity, we assume w.l.o.g. that $(U_i^\#)^\top e = e$ for $i \in [3]$.

5.1. Using unfoldings. Let us start with the most natural approach: reduce an order-3 nTD to an equivalent order-2 nTD using unfoldings. For this, let us recall the following property.

PROPERTY 5.2 (unfoldings). *Let $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ be an order-3 tensor with $\mathcal{T} = (U_1, U_2, U_3).\mathcal{G}$, where $\mathcal{G} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$. The unfolding of $\mathcal{T}$ along the third mode is given by*

$$\mathcal{T}_{(3)} = (U_1 \otimes U_2)\, \mathcal{G}_{(3)}\, U_3^\top,$$

where $\mathcal{G}_{(3)}$ is the unfolding of $\mathcal{G}$ along the third mode. By symmetry, the same property applies to unfoldings along the first and second modes.

Property 5.2 allows us to apply directly Theorem 4.2 for an order-2 nTD to obtain an identifiable order-3 nTD. However, this requires a very strong condition, namely, $r_3 = r_1 r_2$, because Theorem 4.2 only applies to order-2 nTD, where the two inner ranks are equal to each other. Let us state the assumptions needed to obtain a first identifiability result for order-3 nTD.

Assumption 5.3. The order-3 tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ satisfies Assumption 5.1 and

1. $\operatorname{rank}(\mathcal{G}_{(3)}^\#) = r_3 = r_1 r_2$;
2. $(U_1^\# \otimes U_2^\#)$ and $U_3^\#$ satisfy the SSC.

Note that, for simplicity of the presentation, this assumption arbitrarily unfolds the tensor in the third mode. Of course, by symmetry of the problem, one can apply the same reasoning to unfoldings in the first and second modes. We will discuss the condition $(U_1^\# \otimes U_2^\#)$ satisfies the SSC in section 7.

Note that Assumption 5.3.1 on the dimensions of the core tensor is rather strong. But this is the first natural assumption required to generalize the matrix case to the tensor setting. We will relax this assumption for the procedures proposed in the next sections.

By Property 5.2, $\mathcal{T}_{(3)} = (U_1 \otimes U_2)\,\mathcal{G}_{(3)}\,U_3^\top$, and, under Assumption 5.3, we can recover, up to permutations, $U_1 \otimes U_2$, $\mathcal{G}_{(3)}$, and U_3 using order-2 nTDs (Theorem 4.2). It remains to recover U_1 and U_2 from their Kronecker product, $U_1 \otimes U_2$, which we show how to do below (taking into account the unknown permutation). Let us describe this procedure:

Procedure 0*: Unique order-*3 *nTD under Assumption* 5.3

1. Computation of U_3^*: Solve the following min-vol order-2 nTD problem

$$\min_{G, U_{1,2}, U_3} \left|\det\left(G\right)\right| \text{ such that } \mathcal{T}_{(3)} = U_{1,2} G U_3^\top, \qquad U_{1,2}^\top e = e, U_3^\top e = e, \text{ and } (U_{1,2}, U_3) \geq 0, \tag{16}$$

   to obtain an optimal solution $(G^*, U_{1,2}^*, U_3^*)$.
2. Computation of U_1^*, U_2^*: Find a permutation $\Pi_{1,2}^*$ such that $U_{1,2}^*\Pi_{1,2}^*$ admits a decomposition $U_{1,2}^*\Pi_{1,2}^* = U_1^* \otimes U_2^*$ with $U_1^{*\top} e = e$ and $U_2^{*\top} e = e$; see Lemma 5.6 and Theorem 5.5 below.
3. Computation of $\mathcal{G}^*$: Compute the matrix $\mathcal{G}_{(3)}^* = (\Pi_{1,2}^*)^\top G^*$ and fold it to form the order-3 tensor $\mathcal{G}^*$.

Theorem 5.4 shows that, under Assumption 5.3, a decomposition computed by Procedure 0 is essentially unique.

THEOREM 5.4. *Let $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ be an order-3 tensor satisfying Assumption 5.3, with the corresponding (r_1, r_2, r_3)-nTD given by $(U_1^\#, U_2^\#, U_3^\#).\mathcal{G}^\#$, where $r_3 = r_1 r_2$. Then for any other decomposition $(U_1^*, U_2^*, U_3^*).\mathcal{G}^*$ obtained with Procedure 0, there exist permutation matrices Π_1, Π_2, Π_3 such that $U_i^* = U_i^\# \Pi_i$ for all $i \in [3]$ and $\mathcal{G}^* = (\Pi_1^\top, \Pi_2^\top, \Pi_3^\top).\mathcal{G}^\#$.*

We defer the proof to the end of this section, and first describe the lemmas required to prove it. The second step of Procedure 0 requires one to find a permutation $\Pi_{1,2}^*$ such that the following problem has a feasible solution: Given $U_{1,2}^*\Pi_{1,2}^* \in \mathbb{R}^{n_1 n_2 \times r_1 r_2}$, find U_1^*, U_2^* such that $U_i^{*\top} e = e$, $U_i^* \geq 0$ for $i \in [2]$ such that $U_{1,2}^*\Pi_{1,2}^* = U_1^* \otimes U_2^*$. Theorem 5.5 and Lemma 5.6 below imply that for any such permutation, U_1^* and U_2^* obtained by solving this problem will be permutations of $U_1^\#$ and $U_2^\#$, respectively.

Trying all possible permutations of $[r_1 r_2]$ for $\Pi_{1,2}^*$ is of course impractical. We will explain in Part II of this paper [44] how this can be done efficiently. In a nutshell, this can be done by looking at $U_{1,2}^*$ columnwise, which requires one to check at most $r_1 r_2$ possibilities. Moreover, this issue disappears when using the all-at-once procedure described after Lemma 5.6.

THEOREM 5.5. *Let X be an $n_1 n_2 \times r_1 r_2$ matrix and with $X = U_1 \otimes U_2$, where $U_1 \in \mathbb{R}^{n_1 \times r_1}$ and $U_2 \in \mathbb{R}^{n_2 \times r_2}$. Such a decomposition can be computed trivially, and is unique if $U_1^\top e = U_2^\top e = e$.*

Proof. From Definition 2.4, we get that the (i,j)th column of X is given by

$$X_{:,(i,j)} = U_1(:,i) \otimes U_2(:,j) \quad \text{for all } i \in [r_1] \text{ and } j \in [r_2].$$

Rewriting $X_{:,(i,j)}$ as an $r_1 \times r_2$ matrix, denoted by $\text{mat}(X_{:,(i,j)})$, we get that

$$\text{mat}(X_{:,(i,j)}) = U_1(:,i)U_2(:,j)^\top \quad \text{for all } i \in [r_1] \text{ and } j \in [r_2]. \tag{17}$$

Note that $\text{mat}(X_{:,(i,j)}) \neq 0$ for all (i,j) since the columns of U_1 and U_2 are nonzero (their entries sum to one). Rank-one matrix decompositions are unique, up to scaling, and easy to compute. In fact, given two vectors u, v and $A = uv^\top$ with $A(i,j) \neq 0$ for some (i,j), all rank-one approximations of $A = u'v'^\top$ have the form $u' = \alpha A(:,j)$ and $v' = \beta A(i,:)$ for any α, β such that $\alpha\beta = A(i,j)^{-1}$. Since $U_i^\top e = e$ for $i \in [1,2]$, the scaling degree of freedom is absent, and the columns of U_1 and U_2 are uniquely determined using (17). Another way to see uniqueness is to notice that $\text{mat}(X_{:,(i,j)})e = U_1(:,i)$, and $\text{mat}(X_{:,(i,j)})^\top e = U_2(:,j)$. ∎

In Part II of this paper [44], we will design an algorithm for an optimization version of this problem that minimizes $\|X - U \otimes V\|_F^2$ under the above constraints. Note that, in the unconstrained case, this problem can be solved via an SVD computation [48].

LEMMA 5.6. *Let X be an $n_1n_2 \times r_1r_2$ matrix with a decomposition of the form $X = U_1^\# \otimes U_2^\#$, where $U_1^\# \in \mathbb{R}^{n_1 \times r_1}$ and $U_2^\# \in \mathbb{R}^{n_2 \times r_2}$ are full column rank matrices. Let $Y = X\Pi$, where Π is a permutation matrix of dimension r_1r_2, and $Y = U_1^* \otimes U_2^*$, where $U_1^* \in \mathbb{R}^{n_1 \times r_1}$ and $U_2^* \in \mathbb{R}^{n_2 \times r_2}$ are full column rank matrices. Moreover, $(U_1^*)^\top e = (U_1^\#)^\top e = e$ and $(U_2^*)^\top e = (U_2^\#)^\top e = e$. Then there exist permutation matrices Π_1 and Π_2 such that $U_1^* = U_1^\#\Pi_1$ and $U_2^* = U_2^\#\Pi_2$.*

Proof. Let π be the permutation map on the indices of the columns of X induced by multiplication by the permutation matrix Π on the right. Then for all $(i,j) \in [r_1r_2]$, there exist i', j' such that $\pi(i,j) = (i',j')$. Then

$$(U_1^* \otimes U_2^*)_{:,(i',j')} = (U_1^*)_{:,i'} \otimes (U_2^*)_{:,j'} = Y_{:,(i',j')} = X_{:,(i,j)} = (U_1^\#)_{:,i} \otimes (U_2^\#)_{:,j}.$$

Using Theorem 5.5, this gives us that $(U_1^*)_{:,i'} = (U_1^\#)_{:,i}$ and $(U_2^*)_{:,j'} = (U_2^\#)_{:,j}$ for all $(i,j) \in [r_1r_2]$. If now there exists $i'' \neq i'$ such that $\pi(i,k) = (i'',k')$ for some k, k', then $(U_1^*)_{:,i} = (U_1^\#)_{:,i'} = (U_1^\#)_{:,i''}$ which is not possible since $U_1^\#$ has full column rank. As a consequence, π maps (i,k) to $(i',\cdot)$ independently from k, thus inducing a permutation on the indexes in the first position. The same reasoning can be repeated for the indexes in the second position, and conclude that $\pi = \pi_1 \otimes \pi_2$, where $\pi(i,j) = (\pi_1(i), \pi_2(j))$.

In terms of matrices, this can be written as $\Pi = \Pi_1 \otimes \Pi_2$, where $\Pi_1 \in S_{r_1}$, $\Pi_2 \in S_{r_2}$ are permutation matrices and $U_1^* = U_1^\#\Pi_1$, $U_2^* = U_2^\#\Pi_2$. ∎

Proof of Theorem 5.4. By Assumption 5.3, $(U_1^\# \otimes U_2^\#)$ and $U_3^\#$ satisfy the SSC and $\text{rank}(\mathcal{G}_{(3)}^\#) = r_3 = r_1r_2$, hence, solving (16) will find matrices $(U_{1,2}^*, G^*, U_3^*)$ such that there exist permutation matrices $\Pi_{1,2}, \Pi_3$ such that $U_{1,2}^* = (U_1^\# \otimes U_2^\#)\Pi_{1,2}$, $U_3^* = U_3^\#\Pi_3$, and $G^* = \Pi_{1,2}^\top G_{(3)}^\#\Pi_3$ by Theorem 4.2.

Since, at the end of step 2, we have recovered matrices U_1^*, U_2^* such that

$$(U_1^* \otimes U_2^*) = U_{1,2}^*\Pi_{1,2}^* = (U_1^\# \otimes U_2^\#)\Pi_{1,2}\Pi_{1,2}^*.$$

Lemma 5.6 (see below), shows that this implies that there exist permutation matrices Π_1, Π_2 such that $U_i^* = U_i^\#\Pi_i$ for all $i \in [2]$. This also gives us that $\Pi_{1,2}^*\Pi_{1,2} = \Pi_1 \otimes \Pi_2$. Then we have that $\mathcal{G}_{(3)}^* = (\Pi_{1,2}^*)^\top G^* = (\Pi_1 \otimes \Pi_2)^\top G_{(3)}^\#\Pi_3$. On folding this matrix to form an order-3 tensor, we can conclude that $\mathcal{G}^* = (\Pi_1^\top, \Pi_2^\top, \Pi_3^\top).\mathcal{G}^\#$. ∎

All-at-once procedure via penalization. Instead of the two-step approach described in Procedure 0, one can consider the following all-at-once procedure for computing the nTD via penalization: for any $\lambda > 0$, solve the following min-vol order-2 nTD problem,

$$\begin{aligned} \min_{\mathcal{G}_{(3)},U_{1,2},U_1,U_2,U_3} & \left|\det\left(\mathcal{G}_{(3)}\right)\right| + \lambda \|U_{1,2} - U_1 \otimes U_2\|_F^2 \\ & \text{such that } \mathcal{T}_{(3)} = U_{1,2}\mathcal{G}_{(3)}U_3^\top, U_i^\top e = e \text{ and } U_i \geq 0 \text{ for all } i \in [3], \end{aligned} \tag{18}$$

to obtain an optimal solution $(\mathcal{G}^*_{(3)}, U_1^*, U_2^*, U_3^*)$.

The following theorem shows that, under Assumption 5.3, an nTD obtained by solving (18) is essentially unique.

THEOREM 5.7. *Let $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ be an order-3 tensor satisfying Assumption 5.3, with the corresponding (r_1, r_2, r_3)-nTD given by $(U_1^\#, U_2^\#, U_3^\#, \mathcal{G}^\#)$, where $r_3 = r_1 r_2$. Then for any other decomposition $(U_1^*, U_2^*, U_3^*).\mathcal{G}^*$ obtained as an optimal solution of (18), there exist permutation matrices Π_1, Π_2, Π_3 such that $U_i^\# = U_i^* \Pi_i$ for all $i \in [3]$ and $\mathcal{G}^\# = (\Pi_1^\top, \Pi_2^\top, \Pi_3^\top).\mathcal{G}^*$.*

Proof. We first want to show that there exists a feasible solution to the optimization problem in (18). Since $(U_1^\#, U_2^\#, U_3^\#).\mathcal{G}^\#$ is the nTD induced by Assumption 5.3, using Property 5.2, we get that

$$\mathcal{T}_{(3)} = U_{1,2}^\# \mathcal{G}_{(3)}^\# (U_3^\#)^\top, \text{ where } U_{1,2}^\# = U_1^\# \otimes U_2^\#. \tag{19}$$

It follows that $(U_1^\#, U_2^\#, U_3^\#, \mathcal{G}_{(1,2)}^\#)$ is indeed a feasible solution to the optimization problem in (18) for all λ. Let $(U_1^*, U_2^*, U_3^*, \mathcal{G}_{(3)}^*)$ be an optimal solution to (18). Then

$$\begin{aligned} |\det(\mathcal{G}^*_{(3)})| & \leq |\det(\mathcal{G}^*_{(3)})| + \lambda \|U_{1,2}^* - U_1^* \otimes U_2^*\|_F^2 \\ & \leq |\det(\mathcal{G}^\#_{(3)})| + \lambda \|U_{1,2}^\# - U_1^\# \otimes U_2^\#\|_F^2 = |\det(\mathcal{G}^\#_{(3)})|. \end{aligned} \tag{20}$$

The last equality follows from the fact that $U_{1,2}^\# = U_1^\# \otimes U_2^\#$ as stated in (19). Since by Assumption 5.3, $U_1^\# \otimes U_2^\#$ and $U_3^\#$ satisfy the SSC, then $\text{rank}(U_1^\# \otimes U_2^\#) = r_1 r_2 = r_3 = \text{rank}(U_3^\#)$. This implies $\text{rank}(\mathcal{T}_{(3)}) = r_3$, and using Corollary 4.2.1, there exist permutation matrices Π_{12}, Π_3 such that

$$U_{1,2}^* = U_{1,2}^\# \Pi_{12}, U_3^* = U_3^\# \Pi_3 \text{ and } G^*_{(3)} = (\Pi_{12})^\top G^\#_{(3)} \Pi_3. \tag{21}$$

From this, it follows that $|\det(\mathcal{G}^*_{(3)})| = |\det(G^\#_{(3)})|$ and, hence, using (20), $U_{1,2}^* = U_1^* \otimes U_2^*$.

Since $U_{1,2}^* = U_1^* \otimes U_2^* = (U_1^\# \otimes U_2^\#)\Pi_{12}$, using Proposition 5.6, there exist permutation matrices Π_1, Π_2 such that $U_1^* = U_1^\# \Pi_1$ and $U_2^* = U_2^\# \Pi_2$. This also gives us that $\mathcal{G}^*_{(1,2)} = (\Pi_1 \otimes \Pi_2)^\top \mathcal{G}^\#_{(1,2)} \Pi_3$. Rewriting this in tensor form gives $\mathcal{G}^* = (\Pi_1^\top, \Pi_2^\top, \Pi_3^\top).\mathcal{G}^\#$. ∎

5.2. Using slices. Previously we provided an identifiability result for order-3 nTD using unfoldings. However, this result is not very satisfactory as it applies only when one rank of the order-3 nTD is equal to the product of the other two. In this section, we provide four more practical identifiability results relying on slices of the input tensor. For that, let us recall the following property.

PROPERTY 5.8 (slices). *Let $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ be an order-3 tensor with $\mathcal{T} = (U_1, U_2, U_3).\mathcal{G}$, where $\mathcal{G} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$. The jth slice of $\mathcal{T}$ along the third mode is given by*

$$\mathcal{T}_j^{(3)} = U_1 \left(\sum_{k \in [r_3]} (U_3)_{j,k} \mathcal{G}_k^{(3)} \right) U_2^\top,$$

where $\mathcal{G}_k^{(3)}$ is the kth slice of $\mathcal{G}$ along the third mode. By symmetry, the same property applies to the slices along the first and second modes.

5.2.1. Approach 1: Two modes of the input tensor have one slice with maximum rank. In this section, we give identifiability results for order-3 nTD under the assumption that there exist two modes of the given tensor that have one slice with maximum possible rank. We first state the assumptions explicitly and then in Remark 5.10 we explain why these slices have maximum rank.

Assumption 5.9. The order-3 tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ satisfies Assumption 5.1 and

1. $r_3 \leq r = r_1 = r_2$;
2. $U_i^{\#}$ satisfies the SSC for $i \in [3]$;
3. there exists $i_2 \in [n_2]$ and $i_3 \in [n_3]$ such that $\text{rank}(\mathcal{T}_{i_3}^{(3)}) = r$ and $\text{rank}(\mathcal{T}_{i_2}^{(2)}) = r_3$.

Note that, for simplicity of the presentation, this assumption arbitrarily chooses modes 2 and 3 to have at least one slice of maximum rank. Of course, by symmetry of the problem, any combination of two modes can be chosen.

Remark 5.10 (why maximal rank?) Since $\mathcal{T} = (U_1^{\#}, U_2^{\#}, U_3^{\#}).\mathcal{G}^{\#}$, using Property 5.8, any slice $\mathcal{T}_i^{(3)}$ has a decomposition of the form $\mathcal{T}_i^{(3)} = U_1 S_i U_2^\top$ for some r-by-r matrix S_i and, hence, $\text{rank}(\mathcal{T}_i^{(3)}) \leq r$. Therefore Assumption 5.9 requires that there exists $i_3 \in [n_3]$ such that $\mathcal{T}_{i_3}^{(3)}$ attains the maximum possible rank, r. The same reasoning applies to $\mathcal{T}_{i_2}^{(2)}$ for some $i_2 \in [n_2]$. This justifies the choice of the title for this section.

By Property 5.8, $\mathcal{T}_j^{(3)} = U_1 (\sum_{k \in [r_3]} (U_3)_{j,k} \mathcal{G}_k^{(3)}) U_2^\top$, and, under Assumption 5.9, we can recover, up to permutations, U_1 and U_2 using order-2 nTDs (Theorem 4.2). It remains to recover U_3 and $\mathcal{G}$, which can be done via an identifiable min-vol NMF formulation (Theorem 2.8). Let us describe this two-step procedure.

<u>Procedure 1</u>*: Unique order-*3 *nTD under Assumption* 5.9

1. Computation of U_1^*, U_2^*: Solve the following min-vol order-2 nTD problem

(22)
$$\min_{U_1, U_2, S_{i_3}} |\det(S_{i_3})| \text{ such that } \mathcal{T}_{i_3}^{(3)} = U_1 S_{i_3} U_2^\top, U_i^\top e = e, \text{ and } U_i \geq 0 \text{ for } i \in [2],$$

   to obtain an optimal solution $(U_1^*, U_2^*, S_{i_3}^*)$.
2. Computation of U_3^*: Form the matrix $T_{i_2}^* = (U_1^*)^\dagger \mathcal{T}_{i_2}^{(2)}$, where $U^\dagger$ denotes the Moore–Penrose inverse of U. Then solve the following min-vol NMF problem

$$\min_{U_3, S_{i_2}} \det(S_{i_2}^\top S_{i_2}) \quad \text{such that} \quad T_{i_2}^* = S_{i_2} U_3^\top, U_3^\top e = e, \text{ and } U_3 \geq 0, \tag{23}$$

   to obtain an optimal solution $(U_3^*, S_{i_2}^*)$.
3. Computation of $\mathcal{G}^* = ((U_1^*)^\dagger, (U_2^*)^\dagger, (U_3^*)^\dagger).T$.

Procedure 1 assumes the knowledge of the indices for which the slices with maximal rank are known (namely, i_2 and i_3). From a practical point of view, if this information is not available, one can check this rank condition for the slices along each mode. Note that we only need to find two of them with maximal rank, without having to look at all slices, and then solve the corresponding optimization problems.

Theorem 5.11 shows that, under Assumption 5.9, a decomposition computed by this procedure is essentially unique.

THEOREM 5.11. *Let $\mathcal{T} \in \mathbb{R}^{n_1\times n_2\times n_3}$ be an order-3 tensor satisfying Assumption 5.9 with the corresponding (r,r,r_3)-nTD given by $(U_1^\#,U_2^\#,U_3^\#,\mathcal{G}^\#)$. Then for any other decomposition $(U_1^*,U_2^*,U_3^*).\mathcal{G}^*$ obtained with Procedure 1, there exist permutation matrices Π_1,Π_2,Π_3 such that $U_i^* = U_i^\#\Pi_i$ for all $i \in [3]$ and $\mathcal{G}^* = (\Pi_1^\top, \Pi_2^\top,\Pi_3^\top).\mathcal{G}^\#$.*

Proof. Let us first show that the nTD $(U_1^\#,U_2^\#,U_3^\#,\mathcal{G}^\#)$ of $\mathcal{T}$ provides a feasible solution to (22) and (23). By Assumption 5.9, the slices $\mathcal{T}_{i_2}^{(2)}$ and $\mathcal{T}_{i_3}^{(3)}$ have rank r and $r_3 \le r$, respectively. Using Property 5.8, we have

$$\mathcal{T}_{i_3}^{(3)} = U_1^\# D_{i_3}^\# (U_2^\#)^\top, \tag{24}$$

where $S_{i_3}^\# = \sum_{k\in[r_3]}(U_3^\#)_{i_3,k}\mathcal{G}_k^\#$ and $\mathcal{G}_k^\#$ is the kth slice along the third mode of $\mathcal{G}^\#$. This shows that $(U_1^\#, D_{i_3}^\#, (U_2^\#)^\top)$ is a feasible solution of (22), the first step of Procedure 1. Moreover, using Property 5.8, it also follows that $\mathcal{T}_{i_2}^{(2)} = U_1^\# D_{i_2}^\# (U_3^\#)^\top$. Since $U_1^\#$ satisfies the SSC, $\operatorname{rank}(U_1^\#) = r$. Hence, $(U_1^\#)^\dagger\mathcal{T}_{i_2}^{(2)} = D_{i_2}^\#(U_3^\#)^\top$ which shows that $(D_{i_2}^\#,(U_3^\#)^\top)$ is a feasible solution of (23), the second step of Procedure 1.

Let us now show that Procedure 1 recovers an essentially unique nTD. Let $(U_1^*,U_2^*,S_{i_3}^*)$ be an optimal solution of (22). Since $U_1^\#$ and $U_2^\#$ satisfy the SSC and $\operatorname{rank}(\mathcal{T}_{i_3}^{(3)}) = r$ by Assumption 5.9, Theorem 4.2 applies: there exist permutation matrices Π_1,Π_2 such that $U_1^* = U_1^\#\Pi_1$, $U_2^* = U_2^\#\Pi_2$, and $S_{i_3}^* = \Pi_1^\top S_{i_3}^\#\Pi_2$. Let $\mathcal{T}_{i_2}^\# = (U_1^\#)^\dagger\mathcal{T}_{i_2}^{(2)}$ and $\mathcal{T}_{i_2}^* = (U_1^*)^\dagger\mathcal{T}_{i_2}^{(2)}$. Since $U_1^* = U_1^\#\Pi_1$, $\mathcal{T}_{i_2}^* = \Pi_1^\top\mathcal{T}_{i_2}^\#$. Hence the problems (23) constructed from $U_1^\#$ or from U_1^* are the same, up to permutations. By Assumption 5.9, $U_3^\#$ satisfies the SSC, while $\operatorname{rank}(\mathcal{T}_{i_2}^*) = \operatorname{rank}(\mathcal{T}_{i_2}^{(2)}) = r_3$ because $U_1^* = U_1^\#\Pi_1$ satisfies the SSC and hence is full rank. Therefore Theorem 4.2 applies to (23): $U_3^* = U_3^\#\Pi_3$ for some permutation matrix Π_3.

Moreover, since $U_i^\#$ have full column rank, we have that $(U_i^\#)^\dagger U_i^\# = I_{r_i}$ for all $i \in [3]$. Using this, we already have that $\mathcal{G}^\# = ((U_1^\#)^\dagger,(U_2^\#)^\dagger,(U_3^\#)^\dagger).T$. From this, we can conclude that

$$\begin{aligned}\mathcal{G}^* &= ((U_1^*)^\dagger,(U_2^*)^\dagger,(U_3^*)^\dagger).\mathcal{T} = (\Pi_1^\top,\Pi_2^\top,\Pi_3^\top).((U_1^\#)^\dagger,(U_2^\#)^\dagger,(U_3^\#)^\dagger).T \\ &= (\Pi_1^\top,\Pi_2^\top,\Pi_3^\top).\mathcal{G}^\#.\end{aligned}$$

Finally, since all factors involved have full column rank, $\mathcal{G}^*$ computed by the third step of Procedure 1 gives the desired result:

$$\begin{aligned}\mathcal{T} = (U_1^\#,U_2^\#,U_3^\#).\mathcal{G}^\# &= (U_1^*\Pi_1^\top,U_2^*\Pi_2^\top,U_3^*\Pi_3^\top).\Big((\Pi_1,\Pi_2,\Pi_3)\mathcal{G}^*\Big) \\ &= (U_1^*,U_2^*,U_3^*).\mathcal{G}^*.\end{aligned}$$ ∎

5.2.2. Approach 2: The spans of slices of the core tensor along two modes have maximum possible rank. In the previous section, we used a single maximum-rank slice of $\mathcal{T}$. In this section, we show how to combine slices to obtain a stronger identifiability result. We need the following definition.

DEFINITION 5.12 (maximal rank of a matrix space). *We define a subspace of matrices* $\mathcal{V} \subseteq \mathbb{R}_+^{r\times r}$ *to have maximal rank* r *if there exists* $A \in \mathcal{V}$ *such that* $\operatorname{rank}(A) = r$.

For any family of matrices $\mathcal{A} = \{A_1, \dots, A_n\}$, we define their *linear span* over a field $\mathbb{K}$ as $\text{lin-span}_{\mathbb{K}}(\mathcal{A}) := \{\sum_{i=1}^n v_i A_i \mid v_i \in \mathbb{K}\}$.

Assumption 5.13. The order-3 tensor $\mathcal{T} \in \mathbb{R}^{n_1\times n_2\times n_3}$ satisfies Assumptions 5.1, 5.9.1–5.9.2 (namely, $r_3 \le r = r_1 = r_2$ and $U_i^{\#}$ satisfies the SSC for $i \in [3]$), and

1. the set $\mathcal{V}^{(3)} := \text{lin-span}_{\mathbb{R}_+}\{\mathcal{G}_1^{(3)}, \dots, \mathcal{G}_{r_3}^{(3)}\}$ has maximal rank r and the set $\mathcal{V}^{(2)} := \text{lin-span}_{\mathbb{R}_+}\{\mathcal{G}_1^{(2)}, \dots, \mathcal{G}_r^{(2)}\}$ has maximal rank r_3.

Assumption 5.13 is a relaxation of Assumption 5.9. In fact, Assumption 5.9 requires $\operatorname{rank}(\mathcal{T}_j^{(3)}) = r$ for some j. Using Property 5.8,

$$\mathcal{T}_j^{(3)} = U_1^{\#} \underbrace{\Big(\sum_{k\in[r_3]} (U_3^{\#})_{j,k} \mathcal{G}_k^{\#(3)} \Big)}_{=:S} U_2^{\#\top}$$

so that $S \in \mathcal{V}^{(3)}$ has rank r. Similarly, one can show that $\mathcal{V}^{(2)}$ has maximal rank r_3.

Given an order-3 tensor $\mathcal{T}$ that satisfies Assumption 5.13, let us consider the following randomized procedure for computing a decomposition.

<u>*Procedure* 2</u>*: Unique order-*3 *nTD under Assumption* 5.13

1. Let $\mathcal{T}_\alpha = \sum_{i=1}^{n_3} \alpha_i \mathcal{T}_i^{(3)}$ and $\mathcal{T}_\beta = \sum_{i=1}^{n_2} \beta_i \mathcal{T}_i^{(2)}$ be two random linear combinations from a continuous distribution of the slices along the third and second mode of the tensor $\mathcal{T}$, respectively.
2. Computation of U_1^*, U_2^*: Solve the following min-vol order-2 nTD problem

(25)
$$\min_{U_1,U_2,S_\alpha} |\det(S_\alpha)| \text{ such that } \mathcal{T}_\alpha = U_1 S_\alpha U_2^\top, U_i^\top e = e, \text{ and } U_i \ge 0 \text{ for } i \in [2],$$

to obtain an optimal solution $(U_1^*, U_2^*, S_{i_3}^*)$.

3. Computation of U_3^*: For the matrix $T_\beta^* = (U_1^*)^\dagger \mathcal{T}_\beta$ and solve the following min-vol NMF problem

$$\min_{U_3,S_\beta} \det(S_\beta^\top S_\beta) \text{ such that } T_\beta^* = S_\beta U_3^\top, U_3^\top e = e, \text{ and } U_3 \ge 0, \tag{26}$$

to obtain an optimal solution (U_3^*, S_β^*).

4. Compute $\mathcal{G}^* = ((U_1^*)^\dagger, (U_2^*)^\dagger, (U_3^*)^\dagger).T$.

Theorem 5.14 shows that, under Assumption 5.13, a decomposition computed by this procedure is essentially unique with probability 1. The proof proceeds by showing that the set of points $\alpha = (\alpha_1, \dots, \alpha_n)$ and $\beta = (\beta_1, \dots, \beta_n)$ for which Procedure 2 does not return the unique decomposition are the roots of two polynomials and hence they form a set of Lebesgue measure zero.

THEOREM 5.14. *Let* $\mathcal{T} \in \mathbb{R}^{n_1\times n_2\times n_3}$ *be an order-*3 *tensor satisfying Assumption* 5.13 *with the corresponding* (r, r, r_3)*-nTD given by* $(U_1^{\#}, U_2^{\#}, U_3^{\#}).\mathcal{G}^{\#}$*. Then, with probability* 1*, for any decomposition* $(U_1^*, U_2^*, U_3^*).\mathcal{G}^*$ *obtained from Procedure* 2*, there exist permutation matrices* Π_1, Π_2, Π_3 *such that* $U_i^* = U_i^{\#}\Pi_i$ *for all* $i \in [3]$ *and* $\mathcal{G}^* = (\Pi_1^\top, \Pi_2^\top, \Pi_3^\top).\mathcal{G}^{\#}$.

Proof. Let us first show that the nTD $(U_1^{\#}, U_2^{\#}, U_3^{\#}, \mathcal{G}^{\#})$ of $\mathcal{T}$ provides a feasible solution to (25) and (26) with probability 1. Let $\mathcal{T}_j^{(i)}$ be the jth slice of $\mathcal{T}$ along mode i. From Property 5.8, we have

$$\mathcal{T}_j^{(3)} = U_1^{\#} S_j^{(3)} (U_2^{\#})^\top,$$

where $S_j^{(3)} = \sum_{k\in[r_3]} (U_3^{\#})_{j,k} \mathcal{G}_k^{\#}$ for all $j \in [n_3]$ and $\mathcal{G}_k^{\#}$ is the kth slice along the third mode of $\mathcal{G}^{\#}$. Using this, we obtain

$$\mathcal{T}_\alpha = U_1^{\#} S_\alpha (U_2^{\#})^\top, \tag{27}$$

where

$$S_\alpha = \sum_{j\in[r]} \alpha_j S_j^{(3)} = \sum_{k\in[r]} \sum_{j\in[n_3]} \alpha_j (U_3^{\#})_{j,k} \mathcal{G}_k^{\#} = \sum_{k\in[r]} \langle \alpha, (U_3^{\#})_{\cdot,k}\rangle \mathcal{G}_k^{\#}.$$

Define the n-variate polynomial

$$P(x) = \det\left(\sum_{k\in[r]} \langle x, (U_3^{\#})_{\cdot,k}\rangle \mathcal{G}_k^{\#}\right) = \det\left(\sum_{k\in[r]} ((U_3^{\#}))^\top x)_k \mathcal{G}_k^{\#}\right).$$

Let us show that $P(x) \not\equiv 0$. Since, $\mathcal{V}^{(3)}$ is a nonsingular subspace, there exists a vector $t = (t_1, \ldots, t_r)$ such that $\det(\sum_{i\in[r]} t_i \mathcal{G}_i^{\#}) \neq 0$. Moreover, since $U_3^{\#} \in \mathbb{R}^{n_3\times r}$ satisfies the SSC, it has full column rank. Hence, $(U_3^{\#})^\dagger U_3^{\#} = I_r$. Taking $x_0 = ((U_3^{\#})^\dagger)^\top t$, we obtain

$$\begin{aligned} P(x_0) &= \det\left(\sum_{k\in[r]} (((U_3^{\#})^\dagger)^\top x_0)_k \mathcal{G}_k^{\#}\right) = \det\left(\sum_{k\in[r]} ((U_3^{\#}))^\top ((U_3^{\#})^\dagger)^\top t)_k \mathcal{G}_k^{\#}\right) \\ &= \det\left(\sum_{k\in[r]} t_k \mathcal{G}_k^{\#}\right) \neq 0. \end{aligned}$$

Hence, if α is picked at random, $\operatorname{rank}(S_\alpha) = r$ with probability 1. Moreover, since, U_1, U_2 satisfy the SSC, $\operatorname{rank}(U_1^{\#}) = \operatorname{rank}(U_2^{\#}) = r$. This gives us that $\operatorname{rank}(\mathcal{T}_\alpha) = r$ with probability 1. Hence, using (27), $(U_1^{\#}, S_\alpha, (U_2^{\#})^\top)$ is a feasible solution to (25).

Since $\mathcal{V}^{(2)}$ is a matrix subspace with maximal rank r_3, a similar argument gives us that $\operatorname{rank}(S_\beta) = r_3$ with probability 1. Moreover, using a similar argument to (27), we also get that $\mathcal{T}_\beta = U_1^{\#} S_\beta (U_3^{\#})^\top$, where $S_\beta = \sum_{k\in[r]} \langle \beta, (U_2^{\#})_{\cdot,k}\rangle \mathcal{G}_k^{\#}$. Since, $U_1^{\#}$ and $U_3^{\#}$ satisfy the SSC, $\operatorname{rank}(U_1^{\#}) = r$ and $\operatorname{rank}(U_3^{\#}) = r_3$. We conclude that $(S_\beta, (U_3^{\#})^\top)$ is a solution of (25).

Now following the same proof as that of Theorem 5.11, it follows that the (r, r, r_3)-nTD returned by Procedure 2 is essentially unique with probability 1. ∎

5.2.3. Approach 3: One slice along one mode has full rank and the unfolding of the core tensor along that mode has full column rank. In the previous two sections, we have focused on the case where some information related to slices along two modes is available. What if we do not have information related to slices along two modes, but just along one? W.l.o.g., in the rest of this section, we assume that the information is available along the third mode. The same results hold with appropriate modifications for the first and second modes as well.

Assumption 5.15. The order-3 tensor $\mathcal{T} \in \mathbb{R}^{n_1\times n_2\times n_3}$ satisfies Assumption 5.1 and
- $\sqrt{r_3} \leq r = r_1 = r_2$;
- $U_i^{\#}$ satisfies the SSC for all $i \in [3]$;
- there exists $i^* \in [n_3]$ such that $\operatorname{rank}(\mathcal{T}_{i^*}^{(3)}) = r$;
- $\operatorname{rank}(\mathcal{G}_{(3)}^{\#}) = r_3$.

As for Procedure 1, we can recover $U_1^\#$ and $U_2^\#$, up to permutations, using min-vol nTD, since $\mathcal{T}_{i^*}^{(3)} = U_1^\# S U_2^\#$ for some full-rank square matrix S. To recover $U_3^\#$ and $\mathcal{G}^\#$, we solve another min-vol NMF problem for the projection of all the slices $\mathcal{T}_i^{(3)}$ using $U_1^\#$ and $U_2^\#$, as described in the procedure below. Note that here we use all the slices of $\mathcal{T}_i^{(3)}$, while only two were used in Procedure 1.

Procedure 3: *Unique order-3 nTD under Assumption* 5.15

1. Computation of U_1^*, U_2^*: Solve the following min-vol order-2 nTD problem

$$\min_{U_1,U_2,S_{i^*}} |\det(S_{i^*})| \text{ such that } \mathcal{T}_{i^*}^{(3)} = U_1 S_{i^*} U_2^\top, U_i^\top e = e, \text{ and } U_i \geq 0 \text{ for } i \in [2] \tag{28}$$

   to obtain an optimal solution (U_1^*, U_2^*, S_j^*).

2. Computation of $\mathcal{G}_{(3)}^*, U_3^*$: Form the matrix S^* whose ith column is given by $\text{vec}((U_1^*)^\dagger \mathcal{T}_i^{(3)} ((U_2^*)^\top)^\dagger)$ for all $i \in [n_3]$ and solve the min-vol NMF problem

$$\min_{\mathcal{G}_{(3)},U_3} \det(\mathcal{G}_{(3)}^\top \mathcal{G}_{(3)}) \text{ such that } S^* = \mathcal{G}_{(3)} U_3^\top, U_3^\top e = e, \text{ and } U_3 \geq 0, \tag{29}$$

   to obtain an optimal solution $(\mathcal{G}_{(3)}^*, U_3^*)$.

Theorem 5.16 shows that, under Assumption 5.15, a decomposition computed by this procedure is essentially unique with probability 1.

THEOREM 5.16. *Let $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ be an order-3 tensor satisfying Assumption 5.15 with the corresponding (r, r, r_3)-nTD given by $(U_1^\#, U_2^\#, U_3^\#, \mathcal{G}^\#)$. Then for any other decomposition $(U_1^*, U_2^*, U_3^*).\mathcal{G}^*$ obtained with Procedure 3, there exist permutation matrices Π_1, Π_2, Π_3 such that $U_i^* = U_i^\# \Pi_i$ for all $i \in [3]$ and $\mathcal{G}^* = (\Pi_1^\top, \Pi_2^\top, \Pi_3^\top).\mathcal{G}^\#$.*

Proof. Let us first show that the existence of an nTD $(U_1^\#, U_2^\#, U_3^\#).\mathcal{G}^\#$ of the input tensor $\mathcal{T}$ that satisfies Assumption 5.15 also proves the existence of a solution to the optimization problems in (28) and (29). Using Property 5.8, the slices along the third mode have the form

$$\mathcal{T}_i^{(3)} = U_1^\# S_i^\# (U_2^\#)^\top, \qquad \text{where } S_i^\# = \sum_{k \in [r_3]} (U_3^\#)_{i,k} \mathcal{G}_k^{(3)},$$

where $\mathcal{G}_k^{(3)}$ is the kth slice of $\mathcal{G}^\#$ along the third mode. Hence $(U_1^\#, S_{i^*}^\#, U_2^\#)$ is a feasible solution to (28). Let $S^\#$ be the $r_1 r_2 \times n_3$ matrix such that the ith column of $S^\#$ is given by

$$\text{vec}((U_1^\#)^\dagger \mathcal{T}_i^{(3)} ((U_2^\#)^\top)^\dagger) = \text{vec}(S_i^\#) \quad \text{ for all } i \in [n_3].$$

Then for all $j \in [r_1]$ and $k \in [r_2]$,

$$\begin{aligned}(S^\#)_{jk,i} = (S_i^\#)_{j,k} &= \sum_{m \in [r_3]} (U_3^\#)_{i,m} (\mathcal{G}_m^{(3)})_{j,k} = \sum_{m \in [r_3]} (\mathcal{G}_{(3)}^\#)_{jk,m} (U_3^\#)_{m,i}^\top \\ &= (\mathcal{G}_{(3)}^\# . (U_3^\#)^\top)_{jk,i}.\end{aligned} \tag{30}$$

Let us now show that the solution returned by Procedure 3 is essentially unique. Let $(U_1^*, U_2^*, S_{i^*}^*)$ be an optimal solution to (28). The assumptions of Theorem 4.2 are satisfied, by Assumption 5.15, hence there exist permutation matrices Π_1, Π_2 such that $U_1^* = U_1^\# \Pi_1$, $U_2^* = U_2^\# \Pi_2$, and $S_{i^*}^* = \Pi_1^\top S_{i^*}^\# \Pi_2$.

Let $S^{\#}$ be the matrix with the ith column given by $\text{vec}((U_1^{\#})^{\dagger}\mathcal{T}_i^{(3)}((U_2^{\#})^{\top})^{\dagger})$ and S^* be the matrix with the jth column given by $\text{vec}((U_1^*)^{\dagger}\mathcal{T}_j^{(3)}((U_2^*)^{\top})^{\dagger})$ for all $j \in [n_3]$. We have the following relation between their ith columns:

$$\begin{aligned}\text{vec}(S_i^*) &= \text{vec}((U_1^*)^{\dagger}\mathcal{T}_i^{(3)}((U_2^*)^{\top})^{\dagger}) = \text{vec}(\Pi_1^{\top}(U_1^{\#})^{\dagger}\mathcal{T}_i^{(3)}((U_2^{\#})^{\top})^{\dagger}\Pi_2) \\ &= \text{vec}(\Pi_1^{\top} S_i^{\#}\Pi_2) = (\Pi_1 \otimes \Pi_2)^{\top}\text{vec}(S_i^{\#}).\end{aligned} \tag{31}$$

In matrix form, this gives $S^* = (\Pi_1 \otimes \Pi_2)^{\top} S^{\#}$. This shows that S^* is equal to $S^{\#}$, up to a permutation of the rows. Finally, using Theorem 2.8 on (29): there exists a permutation matrix Π_3 such that $\mathcal{G}_{(3)}^* = (\Pi_1^{\top} \otimes \Pi_2^{\top})\mathcal{G}_{(3)}^{\#}\Pi_3$ and $U_3^* = U_3^{\#}\Pi_3$, which gives us the desired result. ∎

5.2.4. Approach 4: The span of slices along one mode has maximum rank and the unfolding of the core tensor along that mode has full column rank. Similarly as was done between Approaches 1 and 2, we can relax the maximum-rank assumption on one slice of $\mathcal{T}$ along the third mode to their span.

Assumption 5.17. The order-3 tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ satisfies Assumptions 5.1, 5.15(.1, .2, .4) (namely, $\sqrt{r_3} \le r = r_1 = r_2$, $U_i^{\#}$ satisfies the SSC for $i \in [3]$, $\text{rank}(\mathcal{G}_{(3)}^{\#}) = r_3$), and $\mathcal{V}^{(3)} = \text{lin-span}_{\mathbb{R}_+}\{\mathcal{G}_1^{(3)}, \ldots, \mathcal{G}_r^{(3)}\}$ has maximal rank r.

Given an order-3 tensor $\mathcal{T}$ that satisfies Assumption 5.17, let us consider the following procedure for computing a decomposition.

<u>Procedure 4</u>*: Unique order-*3 *nTD under Assumption* 5.17

1. Pick $\alpha^{(i)} = \alpha_1^{(i)}, \ldots, \alpha_{n_3}^{(i)}$ independently at random from a continuous distribution for all $i \in [n_3]$ and define the matrix $A \in \mathbb{R}^{n_3 \times n_3}$ whose columns are give by the $\alpha^{(i)}$'s, that is, $A_{:i} = \alpha^{(i)}$ for $i \in [n_3]$. Let $\mathcal{T}_{\alpha^{(i)}} = \sum_{j=1}^{n_3} \alpha_j^{(i)}\mathcal{T}_j^{(3)}$ be n_3 random linear combination of the slices along the third mode of the tensor $\mathcal{T}$.
2. Computation of U_1^*, U_2^*: Solve the following min-vol order-2 nTD problem
$$\min_{U_1, U_2, S_1} |\det(S_1)| \text{ such that } \mathcal{T}_{\alpha^{(1)}} = U_1 S_1 U_2^{\top},\ U_i^{\top} e = e, \text{ and } U_i \ge 0 \text{ for } i \in [2], \tag{32}$$
to obtain an optimal solution $(U_1^*, U_2^*, S_\alpha^*)$.
3. Computation of $\mathcal{G}_{(3)}^*, U_3^*$: Let S^* be the matrix with the ith column given by $\text{vec}((U_1^*)^{\dagger}\mathcal{T}_{\alpha^{(i)}}((U_2^*)^{\top})^{\dagger})$, and solve the min-vol NMF problem
$$\min_{\mathcal{G}_{(3)}, U_3} \det(\mathcal{G}_{(3)}^{\top}\mathcal{G}_{(3)}) \text{ such that } S' = S^* A^{-1} = \mathcal{G}_{(3)} U_3^{\top} \text{ and } U_3^{\top} e = e, U_3 \ge 0, \tag{33}$$
to obtain an optimal solution $(\mathcal{G}_{(3)}^*, U_3^*)$.

Theorem 5.18 shows that, under Assumption 5.17, a decomposition computed by this procedure is essentially unique, with probability 1. The proof proceeds by showing that the set of points $\alpha = (\alpha^{(1)}, \ldots, \alpha^{(n_3)})$ for which Procedure 4 does not return the unique decomposition are the roots of a polynomial and hence they form a set of Lebesgue measure zero.

Theorem 5.18. *Let $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ be an order-*3 *tensor satisfying Assumption* 5.17 *with the corresponding (r, r, r_3)-nTD given by $(U_1^{\#}, U_2^{\#}, U_3^{\#}, \mathcal{G}^{\#})$. Then for any other decomposition $(U_1^*, U_2^*, U_3^*).\mathcal{G}^*$ obtained with Procedure* 4*, there exist permutation matrices Π_1, Π_2, Π_3 such that $U_i^{\#} = U_i^*\Pi_i$ for all $i \in [3]$ and $\mathcal{G}^{\#} = (\Pi_1^{\top}, \Pi_2^{\top}, \Pi_3^{\top}).\mathcal{G}^*$ with probability* 1.

Proof. Let us first show that the existence of an nTD $(U_1^\#, U_2^\#, U_3^\#).\mathcal{G}^\#$ of $\mathcal{T}$ that satisfies Assumption 5.17 provides a feasible solution to (32) and (33) with probability 1. As done in (27), we have

$$\mathcal{T}_{\alpha^{(i)}} = U_1^\# S_i^\# (U_2^\#)^\top, \quad \text{where } S_i^\# = \sum_{k\in[r]} \langle \alpha^{(i)}, (U_3^\#)_{\cdot,k}\rangle \mathcal{G}_k^{(3)} \text{ for all } i \in [n_3]. \tag{34}$$

Since $\mathcal{V}^{(3)}$ has maximal rank r, following the proof of Theorem 5.14, $\operatorname{rank}(\mathcal{T}_{\alpha_1}) = r$ with probability 1.

Since $S^\#$ is the matrix with its ith column defined as $\operatorname{vec}((U_1^\#)^\dagger \mathcal{T}_{\alpha^{(i)}}((U_2^\#)^\top)^\dagger)$, using (34), we get that the ith column of $S^\#$ is given by $\operatorname{vec}(S_i^\#)_{p,q} = \sum_{k\in[r]}((U_3^\#)^\top \alpha_t)_k(\operatorname{vec}(\mathcal{G}_k^{(3)}))_{p,q}$. Rewriting this in matrix form gives us $vec(S_i^\#) = \mathcal{G}_{(3)}^\#(U_3^\#)^\top \alpha^{(i)}$, where $\mathcal{G}_{(3)}^\#$ is the unfolding of the core tensor $\mathcal{G}^\#$ along the 3rd mode. Finally, this gives us that

$$S^\# = \mathcal{G}_{(3)}^\# (U_3^\#)^\top A. \tag{35}$$

Since A is invertible with probability 1 and $\operatorname{rank}(\mathcal{G}_{(3)}^\#) = r_3$, $(\mathcal{G}_{(3)}^\#, (U_3^\#)^\top)$ is a solution for the second step of the optimization problem in (32).

The remainder of the proof that shows that the solution returned by Procedure 4 is essentially unique is similar to the proof of Theorem 5.16. ∎

6. Higher-order nTDs. In this section, we generalize the results for order-3 nTDs of the previous section to any order. The reasoning is the same but the notation gets slightly heavier.

As for order-3 nTDs, we will assume the tensor follows the following assumption.

Assumption 6.1. The order-d tensor $\mathcal{T} \in \mathbb{R}^{n_1\times\cdots\times n_d}$ satisfies the following conditions:

1. $\mathcal{T}$ has an nTD of the form $\mathcal{T} = (U_1^\#, \ldots, U_d^\#).\mathcal{G}^\#$, where $U_i^\# \in \mathbb{R}_+^{n_i\times r_i}$ for all $i \in [d]$ and $\mathcal{G}^\# \in \mathbb{R}^{r_1\times\cdots\times r_d}$, and no columns of the $U_i^\#$'s are zero.
2. We assume w.l.o.g. that $(U_i^\#)^\top e = e$ for $i \in [d]$.

6.1. Using unfoldings. Let $\mathcal{I} = \{j_1, \ldots, j_k\} \subset [d]$. For any tuple $(i_1, \ldots, i_d)$, we denote $i_\mathcal{I} = (i_{j_1}, \ldots, i_{j_k})$, and $i_{\mathcal{I}^-}$ the tuple with the rest of the $(d-k)$ elements, that is, $\mathcal{I}^- = [d]\backslash\mathcal{I}$. For an order-$d$ tensor $\mathcal{T}$, we define the unfolding along the mode $\mathcal{I}$, denoted by $\mathcal{T}_\mathcal{I}$, as the $\prod_{j\in\mathcal{I}^-} n_j \times \prod_{j\in\mathcal{I}} n_j$ matrix with its $(i_{\mathcal{I}^-}, i_\mathcal{I})$ entry given by $\mathcal{T}_{i_1,\ldots,i_d}$. Similarly to Property 5.2, we have the following property for any unfolding.

Property 6.2 (unfoldings). *Let $\mathcal{T} = (U_1, \ldots, U_d).\mathcal{G}$, where $U_i \in \mathbb{R}^{n_i\times r_i}$ for all $i \in [d]$ and $\mathcal{G} \in \mathbb{R}^{r_1\times\cdots\times r_d}$. The unfolding along mode $\mathcal{I}$ of $\mathcal{T}$ can be factorized as*

$$\mathcal{T}_\mathcal{I} = (\otimes_{j\in\mathcal{I}^-} U_j)\, \mathcal{G}_\mathcal{I}\, (\otimes_{j\in\mathcal{I}} U_j)^\top.$$

Assuming $\otimes_{j\in\mathcal{I}^-} U_j$ and $\otimes_{j\in\mathcal{I}} U_j$ satisfy the SSC, we will be able to recover them using min-vol order-2 nTD (Theorem 4.2), while we will be able to recover the individual U_i's by using Kronecker factorizations of these matrices.

Assumption 6.3. The tensor $\mathcal{T} = (U_1^\#, \ldots, U_d^\#).\mathcal{G}$ satisfies Assumption 6.1 and there exists a nonempty set $\mathcal{I} \subsetneq [d]$ such that the following conditions are satisfied:

- $\otimes_{i\in S} U_i^\#$ satisfy the SSC for $S \in \{\mathcal{I}, \mathcal{I}^-\}$.
- $\operatorname{rank}(\mathcal{G}_\mathcal{I}) = \prod_{i\in\mathcal{I}} r_i = \prod_{j\in\mathcal{I}^-} r_j$.

Hence adapting Procedure 0 for order-3 nTDs, we propose the following procedure to compute an identifiable order-d nTD under Assumption 6.3. We call it Procedure d.0 as it generalizes Procedure 0 for order-d tensors.

<u>Procedure d.0</u>: *Unique order-d nTD under Assumption* 6.3

1. Solve the following min-vol order-2 nTD problem,

$$\min_{U_{\mathcal{I}}, U_{\mathcal{I}^{-1}}, G} |\det(G)| \text{ such that } \mathcal{T}_{\mathcal{I}} = U_{\mathcal{I}^-} G U_{\mathcal{I}}^\top, \tag{36}$$
$$U_{\mathcal{I}}^\top e = e, U_{\mathcal{I}^-}^\top e = e, \text{ and } (U_{\mathcal{I}}, U_{\mathcal{I}^-}) \geq 0,$$

   to obtain an optimal solution $(U_{\mathcal{I}}^*, U_{\mathcal{I}^{-1}}^*, G^*)$.
2. Computation of $(U_1^*, \dots, U_d^*)$: Find permutations $\Pi_{\mathcal{I}}^*, \Pi_{\mathcal{I}^-}^*$ such that $U_S^* \Pi_S^*$ admits a decomposition $U_S^* \Pi_S^* = \otimes_{j \in S} U_j^*$ such that $U_j^{*\top} e = e$ for all $j \in S0$, where $S \in \{\mathcal{I}, \mathcal{I}^-\}$; see Corollary 6.4.1 and Theorem 5.5.
3. Computation of $\mathcal{G}^*$: Compute the matrix $\mathcal{G}_{\mathcal{I}}^* = (\Pi_{\mathcal{I}^-}^*)^\top G^* \Pi_{\mathcal{I}}^*$ and fold it to form the order-d tensor $\mathcal{G}^*$.

Following the proof of Theorem 5.4 for order-3 nTD, one can show that if the given tensor satisfies Assumption 6.3, then Procedure d.0 returns an *essentially* unique nTD.

THEOREM 6.4. *Let $\mathcal{T} \in \mathbb{R}^{n_1 \times \cdots \times n_d} = (U_1^\#, \dots, U_d^\#).\mathcal{G}^\#$ satisfy Assumption* 6.3. *Then for any other decomposition $(U_1^*, \dots, U_d^*).\mathcal{G}^*$ of $\mathcal{T}$ obtained by Procedure d.0, there exist permutation matrices Π_i for all $i \in [d]$ such that $U_i^* = U_i^\# \Pi_i$ for all $i \in [d]$ and $\mathcal{G}^* = (\Pi_1^\top, \dots, \Pi_d^\top)\mathcal{G}^\#$.*

Proof. We first want to show that there exists a feasible solution to the optimization problem in (36). Since $(U_1^\#, \dots, U_d^\#).\mathcal{G}^\#$ is the nTD induced by Assumptions 6.3, using Lemma 6.2, we get that

$$\mathcal{T}_{\mathcal{I}} = U_{\mathcal{I}^-}^\# \mathcal{G}_{\mathcal{I}}^\# (U_{\mathcal{I}}^\#)^\top, \text{ where } U_S^\# = \otimes_{i \in S} U_i^\# \text{ for all } S \in \{\mathcal{I}, \mathcal{I}^-\}. \tag{37}$$

By Assumption 6.3, since $\otimes_{i \in S} U_i^\#$ satisfy *SSC* for $S \in \{\mathcal{I}, \mathcal{I}^-\}$ and $\operatorname{rank}(\mathcal{G}_{\mathcal{I}}^\#) = \prod_{i \in \mathcal{I}} r_i = \prod_{j \in \mathcal{I}^-} r_j$, solving (36) will find matrices $(U_{\mathcal{I}}^*, U_{\mathcal{J}}^*, G^*)$ such that there exist permutation matrices Π'_S, where $U_S^* = (\otimes_{i \in S} U_i^\#)\Pi'_S$ for all $S \in \{\mathcal{I}, \mathcal{I}^-\}$ and $G^* = (\Pi'_{\mathcal{I}^-})^\top \mathcal{G}_{\mathcal{I}}^\# \Pi_{\mathcal{I}}$ by Theorem 4.2.

At the end of step 2, we have recovered matrices U_i^* for all $i \in [d]$ and permutation matrices $\Pi_{\mathcal{I}}^*, \Pi_{\mathcal{I}^-}^*$ such that

$$\otimes_{i \in S} U_i^* = U_S^* \Pi_S^* = (\otimes_{i \in S} U_i^\#) \Pi_S \Pi_S^* \text{ for all } S \in \{\mathcal{I}, \mathcal{I}^-\}.$$

Using Corollary 6.4.1 (see below), this implies that there exist permutation matrices Π_i for all $i \in [d]$ such that $U_i^* = U_i^\# \Pi_i$ for all $i \in [d]$. Note that, from this one can further deduce that $\Pi_S \Pi_S^* = \prod_{i \in S} \Pi_i$ for all $S \in \{\mathcal{I}, \mathcal{I}^-\}$.

Rewriting $\mathcal{G}_{\mathcal{I}}^* = (\Pi_{\mathcal{I}^-}^*)^\top G^* \Pi_{\mathcal{I}}^* = (\otimes_{i \in \mathcal{I}^-} \Pi_i)^\top \mathcal{G}_{\mathcal{I}}^\# (\otimes_{i \in \mathcal{I}} \Pi_i)$ in the tensor structure gives us that $\mathcal{G}^* = (\Pi_1^\top, \dots, \Pi_d^\top).\mathcal{G}^\#$. ∎

In order to compute the permutations in step 2 of Procedure d.0, Lemma 5.6 can be extended to a Kronecker product of $d \geq 2$ matrices. This follows from the following facts:

- If $U_i^\top e = e$ for $i \in [d]$, then for any $k \leq d$, $(\otimes_{i \in [k]} U_i)^\top e = e$.
- For matrices A, B, $\operatorname{rank}(A \otimes B) = \operatorname{rank}(A) \operatorname{rank}(B)$.

Let $(\otimes_{i \in [d]} U_i^*) = Y = X\Pi = (\otimes_{i \in [d]} U_i^\#)\Pi$. Then one can apply Lemma 5.6 to the two factors $\otimes_{i \in [d-1]} U_i^*$ and U_d^*, which satisfy the conditions of Lemma 5.6 due to the above facts. This gives us that $\otimes_{i \in [d-1]} U_i^* = (\otimes_{i \in [d-1]} U_i^\#)\Pi'$ for some permutation Π', and we can then proceed inductively. Hence, we get the following corollary.

COROLLARY 6.4.1. *Let X be an $\prod_{i=1}^d n_i \times \prod_{i=1}^d r_i$ matrix with a decomposition of the form $X = U_1^{\#} \otimes \cdots \otimes U_d^{\#}$, where $U_i^{\#} \in \mathbb{R}^{n_i \times r_i}$ are full column rank matrices for all $i \in [d]$. Let $Y = X\Pi$, where Π is a permutation matrix of dimension $\prod_{i=1}^d r_i$, and $Y = U_1^* \otimes \cdots \otimes U_d^*$, where $U_i^* \in \mathbb{R}^{n_i \times r_i}$ are full column rank matrices for all $i \in [d]$. Moreover, $(U_i^*)^\top e = (U_i^{\#})^\top e = e$ for all $i \in [d]$. Then there exist permutation matrices Π_i such that $U_i^* = U_i^{\#} \Pi_i$ for all $i \in [d]$.*

All-at-once approach. Similarly as was done in section 5.1, instead of Procedure $d.0$, one can consider the following all-at-once procedure for computing the nTD via penalization: for any $\lambda_1, \lambda_2 > 0$, solve the following min-vol order-2 NTD problem:

(38)
$$\min_{U_i \in \mathbb{R}_+^{n_i \times r_i} \text{ for } i \in [d],\ \mathcal{G}_\mathcal{I}} |\det(\mathcal{G}_\mathcal{I})| + \lambda_1 \left(\|U_{\mathcal{I}^-} - \bigotimes_{i \in \mathcal{I}^-} U_i\|_F^2 \right) + \lambda_2 \left(\|U_\mathcal{I} - \bigotimes_{i \in \mathcal{I}} U_i\|_F^2 \right)$$
$$\text{such that } \mathcal{T}_\mathcal{I} = U_{\mathcal{I}^-} \det(\mathcal{G}_\mathcal{I}) U_\mathcal{I}^\top, U_i^\top e = e, \text{ and } U_i \geq 0 \text{ for } i \in [d],$$

where $\mathcal{G}_\mathcal{I} \in \mathbb{R}^{\Pi_{j \in \mathcal{I}^-} r_j \times \Pi_{i \in \mathcal{I}} r_i}$, to obtain an optimal solution $(U_1^*, \ldots, U_d^*, \mathcal{G}_\mathcal{I}^*)$. By a result similar to Theorem 5.7, one can also show that, under Assumption 6.3, for all $\lambda_1, \lambda_2 > 0$, an nTD obtained by solving (38) is essentially unique.

6.2. Using slices. Let us define slices of higher-order tensors.

DEFINITION 6.5 (slices for higher order). *Let $\mathcal{T} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ be an order-d tensor. For $i < j \in [d]$ and $k_p \in [n_p]$ for $p \neq i, j$, we define the $[i,j]$-slices of dimension $n_i \times n_j$ as*

$$(\mathcal{T}^{[i,j]}_{k_1,\ldots,k_{i-1},k_{i+1},\ldots,k_{j-1},k_{j+1},\ldots,k_d})_{k_i,k_j} = \mathcal{T}_{k_1,\ldots,k_d} \quad \text{for } k_i \in [n_i] \text{ and } k_j \in [n_j]. \tag{39}$$

Extending Definition 2.2 of the slices of order-3 tensors directly to order-d would lead to the notation $\mathcal{T}^{(1,\ldots,i-1,i+1,\ldots,j-1,j+1,\ldots,d)}_{k_1,\ldots,k_{i-1},k_{i+1},\ldots,k_{j-1},k_{j+1},\ldots,k_d}$. However, we use the notation from (39) for simplicity of the exposition.

Remark 6.6. The slices have been defined in this way (by fixing $d-2$ indices) for order-d tensors in order to still preserve the fact that each $[i,j]$-slice is indeed an $n_i \times n_j$ matrix. We will see in section 6.2.2 that one could relax this notion of slices by fixing $j \leq d-2$ indices which would result in an order-$(d-j)$ subtensor and consider the slices to be a suitable flattening of the subtensor.

Property 5.8 can be generalized to order-d tensors in the following way.

LEMMA 6.7. *Let $\mathcal{T}$ be an order-d tensor such that there exists a decomposition of the form $\mathcal{T} = (U_1, \ldots, U_d).\mathcal{G}$. Then the $[1,2]$-slices of $\mathcal{T}$ are given by*

$$\mathcal{T}^{[1,2]}_{k_3,\ldots,k_d} = U_1 S^{[1,2]}_{k_3,\ldots,k_d} U_2^\top,$$

where

$$S^{[1,2]}_{k_3,\ldots,k_d} = \sum_{t_i \in [r_i] \text{ for } i \in \{3,\ldots,d\}} \left(\prod_{i=3}^d (U_i)_{k_i,t_i} \right) \mathcal{G}^{[1,2]}_{t_3,\ldots,t_d}$$

and $\mathcal{G}^{[1,2]}_{t_3,\ldots,t_d}$ are the $[1,2]$-slices of the tensor $\mathcal{G}$.

6.2.1. Approach 1: $d-1$ modes of the input tensor have one slice with maximum rank. The idea from section 5.2.1 can be generalized to any order by computing one min-vol order-2 nTD to recover U_1 and U_2, and $d-2$ min-vol MF to recover the other U_i's. The following assumptions generalize Assumption 5.9, that is, there exist at least one slice along d different directions such that each of the slices achieves maximum possible rank.

Assumption 6.8. The order-d tensor $\mathcal{T} \in \mathbb{R}^{n_1\times\cdots\times n_d} = (U_1^{\#},\ldots,U_d^{\#}).\mathcal{G}^{\#}$ satisfies Assumption 6.1, and

1. $U_i^{\#}$ satisfies the SSC for $i \in [d]$;
2. for all $i \in \{2,\ldots,d\}$, there exist indices $k^{(i)} = \{k_1^{(i)},\ldots,k_{d-2}^{(i)}\}$ such that $\operatorname{rank}(\mathcal{T}_{k^{(i)}}^{[1,i]}) = r_i$.

Let us now describe how to identify order-d nTDs satisfying Assumption 6.8, following the same idea as in Procedure 1.

Procedure d.1 : *Unique order-d nTD under Assumption* 6.8

1. Computation of U_1^*, U_2^*: Solve the following min-vol order-2 nTD problem,

$$\min_{U_1,U_2,S_2} |\det(S_2)| \text{ such that } \mathcal{T}_{k^{(2)}}^{[1,2]} = U_1 S_2 U_2^\top, U_i^\top e = e, \text{ and } U_i \geq 0 \text{ for } i \in [2], \tag{40}$$

   to obtain an optimal solution (U_1^*, U_2^*, S_1^*).
2. Computation of U_i^*: For all $i \in [d]\setminus\{1,2\}$, let $\mathcal{T}'_{k^{(i)}} = (U_1^*)^\dagger \mathcal{T}_{k^{(i)}}^{[1,i]}$ and solve the following min-vol NMF problem,

$$\min_{U_i,S_i} \det(S_i^\top S_i) \text{ such that } \mathcal{T}'_{k^{(i)}} = S_i U_i^\top, U_i^\top e = e, \text{ and } U_i \geq 0, \tag{41}$$

   to obtain an optimal solution (U_i^*, S_i^*).
3. Computation of $\mathcal{G}^* = ((U_1^*)^\dagger,\ldots,(U_d^*)^\dagger).\mathcal{T}$.

As for order-3 nTD in Theorem 5.11, we now show that, under Assumption 6.8, a decomposition computed by Procedure d.1 is essentially unique.

THEOREM 6.9. *Let $\mathcal{T} = (U_1^{\#},\ldots,U_d^{\#}).\mathcal{G}^{\#}$ satisfy Assumption 6.8. Then for any other decomposition $(U_1^*,\ldots,U_d^*).\mathcal{G}^*$ of $\mathcal{T}$ obtained with Procedure d.1, there exist permutation matrices Π_i such that $U_i^* = U_i^{\#}\Pi_i$ for all $i \in [d]$ and $\mathcal{G}^* = (\Pi_1^\top,\ldots,\Pi_d^\top).\mathcal{G}^{\#}$.*

Proof. First, we show that if there exists an nTD $(U_1^{\#},\ldots,U_d^{\#}).\mathcal{G}^{\#}$ of $\mathcal{T}$ that satisfies Assumptions 6.8, then there exist feasible solutions to the optimization problems in (40) and (41). Using Lemma 6.7, $\mathcal{T}_{k^{(2)}}^{[1,2]} = U_1^{\#} S_{k^{(2)}}^{[1,2]} (U_2^{\#})^\top$ such that

$$S_{k^{(2)}}^{[1,2]} = \sum_{t_i\in[r_i] \text{ for } i\in[d]\setminus\{1,2\}} \left(\prod_{j\in[d]\setminus\{1,2\}} (U_j^{\#})_{k_j^{(2)},t_j} \right) \mathcal{G}_{t_3,\ldots,t_d}^{[1,2]}, \tag{42}$$

then $\mathcal{G}_{t_3,\ldots,t_d}^{[1,2]}$ are the $[1,2]$-slices of $\mathcal{G}^{\#}$. Hence, $(U_1^{\#}, U_2^{\#}, S_{i_3,i_4}^{(3,4)})$ is a feasible solution to the optimization problem in (40).

Then using Lemma 6.7, we have $\mathcal{T}_{k^{(i)}}^{[1,i]} = U_1^{\#} S_{k^{(i)}}^{[1,i]} (U_i^{\#})^\top$ for all $i \in [d]\setminus\{1,2\}$. Since $U_1^{\#}$ satisfies the SSC, $\operatorname{rank}(U_1^{\#}) = r$ and, hence,

$$\mathcal{T}'_{k^{(i)}} = (U_1^{\#})^\dagger \mathcal{T}_{k^{(i)}}^{[1,i]} = S_{k^{(i)}}^{[1,i]} (U_i^{\#})^\top. \tag{43}$$

Using the fact that $\operatorname{rank}(\mathcal{T}_{k^{(i)}}^{(1,i)}) = r_3 \leq r$, we get $\operatorname{rank}(\mathcal{T}'_{k^{(i)}}) = r_3$. It follows that $(S_{k^{(i)}}^{[1,i]}, U_i^{\#})$ is indeed a feasible solution to (41) for all $i \in [d]\setminus\{1,2\}$.

Now we want to show that a solution $(U_1^*,\dots,U_d^*).\mathcal{G}^*$ returned by the procedure is essentially unique. Let (U_1^*,S_2^*,U_2^*) be an optimal solution to (40). Since $\operatorname{rank}(\mathcal{T}_{k^{(2)}}^{[1,2]})=r$, following the proof of Theorem 4.2, there exist permutation matrices Π_1,Π_2 such that $U_1^*=U_1^{\#}\Pi_1$, $U_2^*=U_2^{\#}\Pi_2$, and $S_1^*=\Pi_1^\top S_{k^{(2)}}^{[1,2]}\Pi_2$.

Let $\mathcal{T}_{k^{(i)}}^*=(U_1^*)^\dagger(\mathcal{T}_{k^{(i)}}^{[1,i]})$. Then using the previous result, we also have that $\mathcal{T}_{k^{(i)}}^*=\Pi_1^\top\mathcal{T}'_{k^{(i)}}$, where $\mathcal{T}'_{k^{(i)}}$ is as mentioned in (43). Let (S_i^*,U_i^*) be an optimal solution to the optimization problem in (41) when run on $\mathcal{T}_{k^{(i)}}^*$ for all $i\in[d]\setminus\{1,2\}$. Then $(\Pi_1^\top S_{k^{(i)}}^{[1,i]},U_i^*)$ is also an optimal solution to the second step of the optimization problem in (41) when run on $T_{k^{(i)}}^*$ for all $i\in[d]\setminus\{1,2\}$. Again following the proof of Theorem 2.8, we can conclude that there exists a permutation matrix Π_i such that $S_i^*=\Pi_1^\top S_{k^{(i)}}^{[1,i]}\Pi_i$ and $U_i^*=U_i^{\#}\Pi_i$ for all $i\in[d]\setminus\{1,2\}$. Moreover, since $U_i^{\#}$ have full column rank, we have that $(U_i^{\#})^\dagger U_i^{\#}=I_{r_i}$ for all $i\in[d]$. Using this, we already have that $\mathcal{G}^{\#}=((U_1^{\#})^\dagger,\dots,(U_d^{\#})^\dagger).\mathcal{T}$, and we conclude that

$$\begin{aligned}\mathcal{G}^*&=((U_1^*)^\dagger,\dots,(U_d^*)^\dagger).T=(\Pi_1^\top,\dots,\Pi_d^\top).\Big(((U_1^{\#})^\dagger,\dots,(U_d^{\#})^\dagger).T\Big)\\&=(\Pi_1^\top,\dots,\Pi_d^\top).\mathcal{G}^{\#}.\end{aligned}$$

Finally, using, the following relation

$$\mathcal{T}=(U_1^{\#},\dots,U_d^{\#}).\mathcal{G}^{\#}=(U_1^*\Pi_1^\top,\dots,U_d^*\Pi_d^\top).\Big((\Pi_1,\dots,\Pi_d)\mathcal{G}^*\Big)=(U_1^*\dots,U_d^*).\mathcal{G}^*,$$

the decomposition returned by the procedure is an $(r,r,r_3,\dots,r_d)$-nTD of the tensor $\mathcal{T}$. ∎

Remark 6.10 (variant for Assumption 6.8.2). Let $S\subset\binom{[d]}{2}$, where $|S|\geq d$. We say S *covers* $[d]$ if for all $j\in[d]$, there exists $i\in[d]$ such that $(i,j)\in S$ or $(j,i)\in S$. Then Assumption 6.8.2 can be replaced with the following one: There exists S which covers $[d]$ such that for all $(i,j)\in S$, there exists $k_1^{(i,j)},\dots,k_{d-2}^{(i,j)}$ such that $\mathcal{T}^{(i,j)}_{k_1^{(i,j)},\dots,k_{d-2}^{(i,j)}}$ achieves maximum possible rank. The restriction on the dimensions of the core tensor needs to be imposed accordingly. And then, one can devise a procedure similar to Procedure $d.1$ to compute the factors and the core tensor.

6.2.2. Approach 3 for order-d tensors: Fully generalized slices. In this section, we generalize Approach 3 from section 5.2.3 to higher orders. Recall that, in that approach, we had rank guarantees on a slice along a certain mode and on the unfoldings of the core tensor along the same mode. For any two indices $i\neq j$, $[i,j]$-slices for an order-d tensor $\mathcal{T}\in\mathbb{R}^{n_1\times\cdots\times n_d}$ are defined in Definition 6.5 as $n_i\times n_j$ matrices formed by fixing $d-2$ indices of the tensor and letting the other two indices vary. However, one can generalize this definition by fixing a subset of indices $\mathcal{J}\subsetneq[d]$. The resulting slices are subtensors of order $(d-|\mathcal{J}|)$. Then, given another subset $\mathcal{K}\subsetneq[d]\setminus\mathcal{J}$, we call the slices to be the $\mathcal{K}$-unfoldings of the corresponding subtensors (following the definition of unfoldings of order-d tensors).

Let $[d]=\mathcal{I}\sqcup\mathcal{J}\sqcup\mathcal{K}$, where $\sqcup$ denotes a partition into disjoint nonempty subsets, that is, $[d]=\mathcal{I}\cup\mathcal{J}\cup\mathcal{K}$, $\mathcal{I}\neq\varnothing,\mathcal{J}\neq\varnothing,\mathcal{K}\neq\varnothing$, $\mathcal{I}\cap\mathcal{J}=\varnothing$, $\mathcal{I}\cap\mathcal{K}=\varnothing$, and $\mathcal{J}\cap\mathcal{K}=\varnothing$. For any index $(i_1,\dots,i_d)$, we define i_S to be the tuple with elements i_k, where $k\in S$ for all $S\in\{\mathcal{I},\mathcal{J},\mathcal{K}\}$. Then for any order-$d$ tensor $\mathcal{T}\in\mathbb{R}^{n_1\times\cdots\times n_d}$, we can define the $(\mathcal{J},\mathcal{K})$-slices of $\mathcal{T}$ to be the $\prod_{i\in\mathcal{I}}n_i\times\prod_{k\in\mathcal{K}}n_k$ matrices $(\mathcal{T}_{i_\mathcal{J}}^{(\mathcal{J})})_{i_\mathcal{I},i_\mathcal{K}}=T_{i_1,\dots,i_d}$. Note that ideally one should refer to these slices by $\mathcal{T}_{i_\mathcal{J}}^{(\mathcal{J},\mathcal{K})}$ since $\mathcal{K}$ is necessary to fix the corresponding unfoldings. But here we will just refer to them as $\mathcal{T}_{i_\mathcal{J}}^{(\mathcal{J})}$ because the subset of modes along which the unfolding will take place will be clear from context.

Note that since $\mathcal{I},\mathcal{J},\mathcal{K} \neq \varnothing$, for the order-3 tensor case, this forces $|\mathcal{I}| = |\mathcal{J}| = |\mathcal{K}| = 1$. Then in that case, the $(\mathcal{J},\mathcal{K})$-slices correspond to the actual slices of the tensor.

The following structural result holds for $(\mathcal{J},\mathcal{K})$-slices of $\mathcal{T}$.

LEMMA 6.11. *Let $\mathcal{T} \in \mathbb{R}^{n_1\times\cdots\times n_d}$ be an order-d tensor with TD $\mathcal{T} = (U_1,\ldots,U_d).\mathcal{G}$, where $U_i \in \mathbb{R}^{n_i\times r_i}$ for all $i \in [d]$ and $\mathcal{G} \in \mathbb{R}^{r_1\times\cdots\times r_d}$. Then, for any $\mathcal{I},\mathcal{J},\mathcal{K}$ such that $[d] = \mathcal{I} \sqcup \mathcal{J} \sqcup \mathcal{K}$, the $(\mathcal{J},\mathcal{K})$-slices denoted by $\mathcal{T}^{(\mathcal{J})}_{i_\mathcal{J}}$ can be decomposed into*

$$\mathcal{T}^{(\mathcal{J})}_{i_\mathcal{J}} = \left(\bigotimes_{i\in\mathcal{I}} U_i\right) D^{(\mathcal{J})}_{i_\mathcal{J}} \left(\bigotimes_{k\in\mathcal{K}} U_k\right)^\top, \tag{44}$$

where

$$D^{(\mathcal{J})}_{i_\mathcal{J}} = \sum_{i'_j\in[r_j] \text{ for all } j\in\mathcal{J}} \left(\prod_{k\in\mathcal{J}} (U_k)_{i_k,i'_k}\right) \mathcal{G}^{(\mathcal{J})}_{i'_\mathcal{J}},$$

and $i'_\mathcal{J}$ is the tuple with i'_j for all $j \in \mathcal{J}$ and $\mathcal{G}^{(\mathcal{J})}_{i_\mathcal{J}}$ are the $(\mathcal{J},\mathcal{K})$-slices of $\mathcal{G}$.

Moreover, if $D^{(\mathcal{J})}$ is the $(\prod_{i\in\mathcal{I}} r_i)(\prod_{k\in\mathcal{K}} r_k) \times (\prod_{j\in\mathcal{J}} r_j)$ matrix with its columns given by $\operatorname{vec}(D^{(\mathcal{J})}_{i_\mathcal{J}})$, then $D^{(\mathcal{J})}$ can be decomposed into

$$D^{(\mathcal{J})} = \mathcal{G}_\mathcal{J} \left(\bigotimes_{j\in\mathcal{J}} U_j\right)^\top. \tag{45}$$

Proof. Let $\mathcal{I} \sqcup \mathcal{J} \sqcup \mathcal{K}$ be a partition of $[d]$ into disjoint nonempty subsets, and for a tuple of indices $(i_1,\ldots,i_d)$, let $i_\mathcal{I}, i_\mathcal{J}, i_\mathcal{K}$ be the tuples of indices induced by this partition. Following the definition of the multilinear transformation operation, we have that

$$\begin{aligned}(\mathcal{T}^{(\mathcal{J})}_{i_\mathcal{J}})_{i_\mathcal{I},i_\mathcal{K}} = T_{i_1,\ldots,i_d} &= \sum_{i'_t\in[r_t] \text{ for all } t\in[d]} \prod_{k\in[d]} (U_k)_{i_k,i'_k} \mathcal{G}_{j_1,\ldots,j_d} \\ &= \left(\bigotimes_{i\in\mathcal{I}} U_i\right)_{i_\mathcal{I},i'_\mathcal{I}} (D^{(\mathcal{J})}_{i_\mathcal{J}})_{i'_\mathcal{I},i'_\mathcal{K}} \left(\bigotimes_{i\in\mathcal{K}} U_i\right)_{i_\mathcal{K},i'_\mathcal{K}},\end{aligned}$$

where

$$D^{(\mathcal{J})}_{i_\mathcal{J}} = \sum_{i'_t\in[r_t] \text{ for all } t\in\mathcal{J}} \left(\prod_{k\in\mathcal{J}} (U_k)_{i_k,i'_k}\right) (\mathcal{G}^{(\mathcal{J})}_{i'_\mathcal{J}}), \tag{46}$$

and $i'_\mathcal{I}, i'_\mathcal{J}, i'_\mathcal{K}$ is the partition of the tuple $(i'_1,\ldots,i'_d)$ induced by the disjoint subsets of $[d]$.

Since $D^\mathcal{J}$ is a matrix of dimension $(\prod_{i\in\mathcal{I}} r_i)(\prod_{k\in\mathcal{I}} r_k) \times (\prod_{j\in\mathcal{J}} r_j)$, we will denote its entries with the indices $(i_\mathcal{I}, i_\mathcal{K}), i_\mathcal{J}$. If $D^\mathcal{J}$ is the matrix with columns given by $\operatorname{vec}(D^{(\mathcal{J})}_{i_\mathcal{J}})$, then using (46), we get that

$$
\begin{aligned}
(D^{\mathcal{J}})_{(i_{\mathcal{I}},i_{\mathcal{K}}),i_{\mathcal{J}}} &= (D^{(\mathcal{J})}_{i_{\mathcal{J}}})_{(i_{\mathcal{I}},i_{\mathcal{K}})} \\
&= \sum_{i'_t\in[r_t] \text{ for all } t\in\mathcal{J}} \left(\prod_{k\in\mathcal{J}} (U_k)_{i_k,i'_k}\right)(\mathcal{G}^{(\mathcal{J})}_{i'_{\mathcal{J}}})_{i_{\mathcal{I}},i_{\mathcal{K}}} \\
&= \sum_{i'_t\in[r_t] \text{ for all } t\in\mathcal{J}} (\mathcal{G}_{\mathcal{J}})_{(i_{\mathcal{I}},i_{\mathcal{K}}),i'_{\mathcal{J}}} \left(\bigotimes_{j\in\mathcal{J}} U_j\right)^{\top}_{i'_{\mathcal{J}},i_{\mathcal{J}}} \\
&= \left(\mathcal{G}_{\mathcal{J}}\cdot\left(\bigotimes_{j\in\mathcal{J}} U_j\right)^{\top}\right)_{(i_{\mathcal{I}},i_{\mathcal{K}}),i_{\mathcal{J}}}. \qquad \square
\end{aligned}
$$

In the rest of this section, we generalize the ideas from section 5.2.3 to higher orders.

Assumption 6.12. The order-d tensor $\mathcal{T}\in\mathbb{R}^{n_1\times\cdots\times n_d} = (U_1^{\#},\dots,U_d^{\#}).\mathcal{G}^{\#}$ satisfies Assumption 6.1, and

1. there exists a partition $\mathcal{I}\sqcup\mathcal{J}\sqcup\mathcal{K}$ of $[d]$ such that $\bigotimes_{i\in S} U_i^{\#}$ satisfies the SSC for $S\in\{\mathcal{I},\mathcal{J},\mathcal{K}\}$;
2. there exists $i_{\mathcal{J}}$ such that $\operatorname{rank}(\mathcal{T}^{(\mathcal{J})}_{i_{\mathcal{J}}}) = \prod_{i\in\mathcal{I}} r_i = \prod_{k\in\mathcal{K}} r_k = r$;
3. the unfolding of the core tensor along the mode $\mathcal{J}$ satisfies $\operatorname{rank}(\mathcal{G}^{\#}_{\mathcal{J}}) = \prod_{j\in\mathcal{J}} r_j \le r^2$.

Given an order-d tensor $\mathcal{T}$ that satisfies Assumption 6.12, let us consider the following procedure for computing a decomposition.

<u>*Procedure d*.3</u>*: Unique order-d nTD under Assumption* 6.12

1. Solve the following min-vol order-2 nTD problem,

$$\min_{U_{\mathcal{I}},U_{\mathcal{K}},S_1} |\det(S_1)| \text{ such that } \mathcal{T}^{(\mathcal{J})}_{i_{\mathcal{J}}} = U_{\mathcal{I}}S_1U_{\mathcal{K}}^{\top}, U_i^{\top}e = e, \text{ and } U_i\ge 0 \text{ for } i\in\{\mathcal{I},\mathcal{K}\}, \tag{47}$$

   to obtain an optimal solution $(U^*_{\mathcal{I}},U^*_{\mathcal{K}},S^*_1)$.
2. Form the matrix S^* to be the matrix with $i_{\mathcal{J}}$th column as $\operatorname{vec}(U^{*\dagger}_{\mathcal{I}}\mathcal{T}^{(\mathcal{J})}_{i_{\mathcal{J}}}(U^{*}_{\mathcal{K}}{}^{\top})^{\dagger})$, and solve the min-vol NMF problem

$$\min_{G,U_{\mathcal{J}}} \det(G^{\top}G) \text{ such that } S^* = GU_{\mathcal{J}}^{\top}, U_{\mathcal{J}}^{\top}e = e, U_{\mathcal{J}}\ge 0, \tag{48}$$

   to obtain an optimal solution $(G^*,U^*_{\mathcal{J}})$.
3. Compute U_i^* for all $i\in[d]$: Find permutations $\Pi^*_{\mathcal{I}},\Pi^*_{\mathcal{J}},\Pi^*_{\mathcal{K}}$ such that $U^*_S\Pi^*_S$ admits a decomposition $U^*_S\Pi^*_S = \otimes_{j\in S}U^*_j$ such that $U_j^{*\top}e = e$ for all $j\in S$ where $S\in\{\mathcal{I},\mathcal{J},\mathcal{K}\}$; see Corollary 6.4.1 and Theorem 5.5.
4. Computation of $\mathcal{G}^*$: Compute the matrix $\mathcal{G}^*_{(\mathcal{J})} = (\Pi^*_{\mathcal{I}}\otimes\Pi^*_{\mathcal{K}})^{\top}G^*\Pi^*_{\mathcal{J}}$ and fold the matrix to compute the order-d tensor $\mathcal{G}^*$.

One can then prove a generalization of Theorem 5.16 to show that under Assumption 6.12, a decomposition computed by Procedure d.3 is essentially unique.

THEOREM 6.13. *Let $\mathcal{T}$ be an order-d tensor such that Assumption* 6.12 *is satisfied and let $(U_1^{\#},\dots,U_d^{\#}).\mathcal{G}^{\#}$ be the decomposition induced by these assumptions. Then for any other decomposition $(U_1^*,\dots,U_d^*).\mathcal{G}^*$ returned by the procedure, there exist permutation matrices Π_i for $i\in[d]$ such that $U_i^* = U_i^{\#}\Pi_i$ for all $i\in[d]$ and $\mathcal{G}^* = (\Pi_1^{\top},\dots,\Pi_4^{\top}).\mathcal{G}^{\#}$.*

Proof. First, we show that the existence of an nTD $(U_1^\#,\dots,U_d^\#).\mathcal{G}^\#$ of the input tensor $\mathcal{T}$ that satisfies Assumptions 6.12 also proves the existence of a feasible solution to the optimization problems in (47) and (48). Using Lemma 6.7, the $(\mathcal{J},\mathcal{K})$-slices have the form $\mathcal{T}_{i_\mathcal{J}}^{(\mathcal{J})} = U_\mathcal{I}^\# S_{i_\mathcal{J}}^{(\mathcal{J})} (U_\mathcal{K}^\#)^\top$ for all $j_3 \in [n_3], j_4 \in [n_4]$ such that

$$S_{i_\mathcal{J}}^{(\mathcal{J})} = \sum_{i'_j \in [r_j] \text{ for all } j \in \mathcal{J}} \left(\prod_{j\in\mathcal{J}} (U_j^\#)_{i_j, i'_j} \right) \mathcal{G}_{i'_\mathcal{J}}^{(\mathcal{J})}, \tag{49}$$

where $i'_\mathcal{J}$ is the tuple of indices i'_j for all $j \in \mathcal{J}$, $\mathcal{G}_{i'_\mathcal{J}}^{(\mathcal{J})}$ are the $(\mathcal{J},\mathcal{K})$-slices of $\mathcal{G}^\#$, and $U_S^\# = \otimes_{j\in S} U_j^\#$, where $S \in \{\mathcal{I},\mathcal{K}\}$. Hence, $(U_\mathcal{I}^\#, U_\mathcal{J}^\#, S_{i_\mathcal{J}}^{(\mathcal{J})})$ is a feasible solution to the optimization problem in (47). If $S^{(\mathcal{J})}$ is the matrix with its $i_\mathcal{J}$ column as $\mathrm{vec}(S_{i_\mathcal{J}}^{(J)})$, rewriting (49) in matrix form, and using Lemma 6.7, we get that $S^{(\mathcal{J})} = \mathcal{G}_\mathcal{J}(U_\mathcal{J}^\#)^\top$, where $U_\mathcal{J}^\# = \bigotimes_{j\in\mathcal{J}} U_j$ and $\mathcal{G}_{(\mathcal{J})}^\#$ is the unfolding of $\mathcal{G}^\#$ along the $\mathcal{J}$ mode (following the definition at the beginning of section 6.1). Then $(\mathcal{G}_\mathcal{J}^\#, U_\mathcal{J}^\#)$ is a feasible solution to the optimization problem in (48).

Now we want to show that a solution $(U_1^*,\dots,U_d^*).\mathcal{G}^*$ returned by the procedure is essentially unique.

Let $(U_\mathcal{I}^*, U_\mathcal{K}^*, S_\mathcal{J}^*)$ be another distinct optimal solution to the first step of the optimization problem in (47). Then following the proof of Theorem 4.2, we already get that there exist permutation matrices $\Pi'_\mathcal{I}, \Pi'_\mathcal{K}$ such that $U_S^* = U_S^\# \Pi'_S$, where $U_S^\# = \otimes_{i\in S} U_i^\#$ for $S \in \{\mathcal{I},\mathcal{K}\}$ and $S_\mathcal{J}^* = (\Pi'_\mathcal{I})^\top S_i^\# \Pi'_\mathcal{J}$. Let $S^\#$ be the matrix with $i_\mathcal{J}$th column given by $\mathrm{vec}((U_\mathcal{I}^\#)^\dagger \mathcal{T}_{i_\mathcal{J}}^{(\mathcal{J})} ((U_\mathcal{K}^\#)^\top)^\dagger)$. Now, following the proof of Theorem 5.16 (refer to (31)), we get that

$$S^* = (\Pi'_\mathcal{I} \otimes \Pi'_\mathcal{K})^\top S^\#. \tag{50}$$

Let $(G^*, U_\mathcal{J}^*)$ be another distinct optimal solution to the second step of the optimization problem in (48) when run on S^*. Using (50), $((\Pi'_\mathcal{I} \otimes \Pi'_\mathcal{K})^\top \mathcal{G}_\mathcal{J}^\#, U_\mathcal{J}^*)$ is also an optimal solution to the optimization problem in (48) when run on S^*. Then using Theorem 4.2, we get that there exists a permutation matrix $\Pi_\mathcal{J}$ such that $G^* = (\Pi'_\mathcal{I} \otimes \Pi'_\mathcal{K})^\top \mathcal{G}_\mathcal{J}^\# \Pi'_\mathcal{J}$ and $U_\mathcal{J}^* = (\otimes_{j\in\mathcal{J}} U_j^\#)\Pi'_\mathcal{J}$.

At the end of step 3 of Procedure d.3, we have recovered matrices U_i^* for all $i \in S$ and permutation matrix Π_S^*, where $S \in \{\mathcal{I},\mathcal{J},\mathcal{K}\}$ such that

$$\otimes_{j\in S} U_j^* = U_S^* \Pi_S^* = (\otimes_{j\in S} U_j^\#)\Pi'_S \Pi_S^*.$$

Corollary 6.4.1 shows that there exist permutation matrices $\Pi_1,\dots,\Pi_d$ such that $U_i^* = U_i^\# \Pi_i$ for $i \in [d]$. It also follows that $\Pi'_S \Pi_S^* = \otimes_{i\in S} \Pi_i$ for $S \in \{\mathcal{I},\mathcal{J},\mathcal{K}\}$. This gives us that

$$\mathcal{G}_\mathcal{J}^* = (\otimes_{i\in\mathcal{I}\cup\mathcal{K}} \Pi_i)^\top \mathcal{G}_\mathcal{J}^\# (\otimes_{j\in\mathcal{J}} \Pi_i)$$

using which we can conclude that $\mathcal{G}^* = (\Pi_1^\top,\dots,\Pi_d^\top).\mathcal{G}^\#$. ◻

Two-step procedure. Instead of using Procedure d.3, one can also consider the following two-step procedure:

1. For any $\lambda_1, \lambda_2, \lambda_3 > 0$, solve the following optimization problem,

$$\min_{U_i \text{ for } i\in\mathcal{I}\cup\mathcal{K}, S_1} |\det(S_1)| + \lambda_1 \|U_\mathcal{I} - \otimes_{i\in\mathcal{I}} U_i\|_F^2 + \lambda_2 \|U_\mathcal{J} - \otimes_{k\in\mathcal{K}} U_k\|_F^2$$
$$\text{such that } \mathcal{T}_{i_\mathcal{J}}^{(\mathcal{J})} = U_\mathcal{I} S_1 U_\mathcal{K}^\top, U_i^\top e = e, \text{ and } U_i \geq 0 \text{ for } i \in \mathcal{I}\cup\mathcal{K},$$

to obtain an optimal solution U_i^* for all $i \in \mathcal{I}\cup\mathcal{K}$ and S_1^*.

2. Form the matrix S^* to be the matrix with the $((i_{\mathcal{I}}, i_{\mathcal{K}}), i_{\mathcal{J}})$th column given by

$$\operatorname{vec}\left((\otimes_{i \in \mathcal{I}} U_i^*)^\dagger \mathcal{T}_{i_{\mathcal{J}}}^{(\mathcal{J})} ((\otimes_{k \in \mathcal{K}} U_k^*)^\top)^\dagger \right),$$

and solve

$$\min_{\mathcal{G}_{\mathcal{J}}, U_j \text{ for } j \in \mathcal{J}} \det(\mathcal{G}_{\mathcal{J}}^\top \mathcal{G}_{\mathcal{J}}) + \lambda_3 ||U_{\mathcal{J}} - \otimes_{j \in \mathcal{J}} U_j||_F^2$$
$$\text{such that } S^* = \mathcal{G}_{\mathcal{J}} U_{\mathcal{J}}^\top,\ U_j^\top e = e, \text{ and } U_j \geq 0 \text{ for all } j \in \mathcal{J},$$

to obtain an optimal solution U_i^* for $i \in \mathcal{J}$ and $\mathcal{G}_{\mathcal{J}}^*$.

By a result similar to Theorem 5.7, one can show that, under Assumption 6.12, for any $\lambda_1, \lambda_2, \lambda_3 > 0$, an nTD obtained with the above two-step procedure is essentially unique.

6.2.3. Using random linear combinations of slices. The idea of section 5.2.2, to take random linear combinations of slices instead of just one, generalizes in a similar way to order-d tensors. Note that an order-d tensor has $\prod_{i=1}^{d-2} n_i$ many $(d-1, d)$-slices. So if we are just given access to the entire tensor, taking a random linear combination of the slices indeed requires looking at all the entries of the tensor which is computationally heavy. We could of course use a tradeoff and select a subset of slices. One could also look at other computational models where the underlying assumptions is that a succinct representation of the tensor is given as input and in those settings, one could access the random linear combination of slices at a much lower cost; see [30, 43] for more detailed discussions.

7. Kronecker product of two matrices and the SSC.

In quite a few cases, our identifiability results relied on the Kronecker product of two or more factors to satisfy the SSC, namely, for order-3 nTDs using unfoldings in Theorem 5.4, and for higher-order nTDs in Theorems 6.4 for unfoldings and 5.16 for slices. In this section, we discuss sufficient conditions for the Kronecker product of two factors to satisfy the SSC.

As discussed in section 5, for order-3 nTDs, one has similar identifiability results when the factor matrices individually satisfy the SSC, or when the Kronecker product of two of them satisfies the SSC. One can then ask the following question.

QUESTION 7.1 (SSCness of Kronecker product of SSC matrices). *Let $U_i \in \mathbb{R}^{n_i \times r_i}$ satisfy the SSC for $i = 1, 2$. Does $U_1 \otimes U_2 \in \mathbb{R}^{n_1 n_2 \times r_1 r_2}$ satisfy the SSC?*

Unfortunately, the answer is negative as soon as $\min(r_1, r_2) \geq 3$ and $\max(r_1, r_2) \geq 4$ (Theorem 7.2), otherwise it is positive (Corollary 7.5.2). The negative result is as follows, the proof of which is deferred to section 7.2.

THEOREM 7.2. *For $\min\{r_1, r_2\} \geq 3$ and $\max\{r_1, r_2\} \geq 4$, there exist U_1 and U_2 satisfying SSC while $U_1 \otimes U_2$ does not satisfy SSC1.*

On the other hand, when either U_1 or U_2 is separable and the other is SSC, their Kronecker product is SSC (this will be a consequence of Theorem 7.5).

Since separability is much more restrictive than SSC (see section 2.2), we now define a condition in-between separability and SSC, that will allow us to prove SSCness of the Kronecker product between two matrices satisfying this condition.

DEFINITION 7.3 (expanded sufficiently scattered condition (p-SSC)). *Let $U \in \mathbb{R}_+^{n \times r}$, $r \geq 2$, and $p \geq 1$. The matrix U satisfies the p-SSC if*

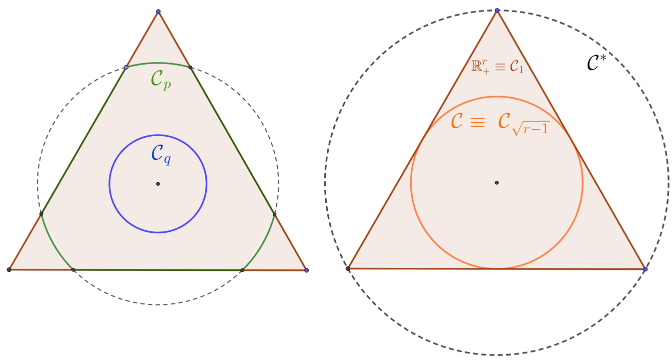

FIG. 2. *Visualization of p-SSC in the case $r=3$ on the plane $\{x \in \mathbb{R}^r \mid e^\top x = 1\}$. On the left, the cones $\mathcal{C}_p$ with $1 < p < \sqrt{r-1}$ and $\mathcal{C}_q$ with $\sqrt{r-1} < q < \sqrt{r}$. On the right, the cone $\mathcal{C} \equiv \mathcal{C}_{\sqrt{r-1}}$ and the nonnegative orthant $\mathbb{R}^r_+ \equiv \mathcal{C}_1$.*

$$\mathcal{C}_p = \left\{x \in \mathbb{R}^r_+ \;\middle|\; e^\top x \geq p\|x\|_2\right\} \subseteq \operatorname{cone}\left(U^\top\right).$$

Intuitively, p-SSC of U requires cone$(U^\top)$ to contain a larger cone than the one required by SSC1 when $p < \sqrt{r-1}$, as shown in Figure 2. For $p = 1$, p-SSC is equivalent to separability because $\mathcal{C}_1 = \mathbb{R}^r_+$. For any $p < \sqrt{r-1}$, p-SSC implies the SSC, because $\mathcal{C}_q \subseteq \mathcal{C}_p$ for any $q \geq p$, and $\mathcal{C} = \mathcal{C}_{\sqrt{r-1}}$, that is, p-SSC with $p = \sqrt{r-1}$ is equivalent to SSC1; this is why we need p smaller than $\sqrt{r-1}$ to have SSC2 and hence SSC.

We can now ask the following question.

QUESTION 7.4 (SSCness of Kronecker product of p-SSC matrices). *Let $U_i \in \mathbb{R}^{n_i \times r_i}$ satisfy the p_i-SSC for $i = 1, 2$. For what values of the p_i's does $U_1 \otimes U_2 \in \mathbb{R}^{n_1 n_2 \times r_1 r_2}$ satisfy the SSC?*

Before providing an answer to this question, let us discuss further the p-SSC condition. It is linked to the so-called uniform pixel purity level γ defined in [39] as follows:

$$\gamma := \sup\left\{s \leq 1 \;\middle|\; B_s \cap \Delta^r \subseteq \operatorname{conv}\left(U^\top\right)\right\}, \qquad \text{where} \qquad B_s = \{x \in \mathbb{R}^r \mid \|x\| \leq s\},$$

where U is assumed to be a row-stochastic matrix (w.l.o.g.). Since cone$(B_s \cap \Delta^r) = \mathcal{C}_{1/s}$, it is immediate to see that a row-stochastic matrix U satisfies p-SSC if and only if its uniform pixel purity level is at least $\gamma \geq 1/p$.

Moreover,

- in [20], it is proved that U satisfies the SSC according to Definition 2.7 if and only if it satisfies p-SSC for some $p^2 < r-1$ when $r > 2$;
- as a consequence, min-vol NMF (7) is identifiable if H satisfies p-SSC for some $p^2 < r-1$ when $r > 2$;
- when $r = 2$, the separability conditions, SSC and 1-SSC coincide.

7.1. Kronecker product of expanded SSC matrices. The following theorem provides an answer to Question 7.4. It will also allow us to answer Question 7.1.

THEOREM 7.5. *Let $U_i \in \mathbb{R}^{n_i \times r_i}$ satisfy the SSC and p_i-SSC for $i = 1, 2$. Then $U_1 \otimes U_2 \in \mathbb{R}^{n_1 n_2 \times r_1 r_2}$ satisfies the SSC when*

$$1 \leq \sqrt{\frac{r_1 - p_1^2}{p_1^2(r_1 - 1)}} + \sqrt{\frac{r_2 - p_2^2}{p_2^2(r_2 - 1)}}.$$

Proof. First of all, notice that since U_i satisfies the SSC, $p_i^2 \le r_i - 1$ and, if $r_i > 2$, then $p_i^2 < r_i - 1$.

The condition SSC1 on $U_1 \otimes U_2$ is equivalent to the condition

$$(U_1 \otimes U_2)v \ge 0 \implies e^\top v \ge \|v\|,$$

and by recasting v to a matrix $V \in \mathbb{R}^{r_1 \times r_2}$, that is, $V = \mathrm{mat}(v)$, it can be written as

$$U_1 V U_2^\top \ge 0 \implies e^\top V e \ge \|V\|_F.$$

Let us suppose from now on that $U_1 V U_2^\top \ge 0$. Since U_1 satisfies p_1-SSC with $p_1^2 \le r_1 - 1$, $\mathcal{C} \subseteq \mathcal{C}_{p_1} \subseteq \mathrm{cone}(U_1^\top)$ and $\mathrm{cone}^*(U_1^\top) \subseteq \mathcal{C}^* = \mathcal{C}_1$. Since U_2 satisfies p_2-SSC,

$$\begin{aligned} U_1 V U_2^\top \ge 0 &\implies V U_2^\top x \in \mathrm{cone}^*(U_1^\top) \;\; \forall x \ge 0 \implies V u_2 \in \mathrm{cone}^*(U_1^\top) \;\; \forall u_2 \in \mathrm{cone}(U_2^\top) \\ &\implies e^\top V u_2 \ge \|V u_2\| \;\; \forall u_2 \in \mathrm{cone}(U_2^\top) \implies e^\top V c \ge \|Vc\| \;\; \forall c \in \mathcal{C}_{p_2}. \end{aligned} \tag{51}$$

Take now $\alpha_i \ge 0$ as the largest constant such that $\alpha_i e + e_1 \in \mathcal{C}_{p_i}$. Since $\alpha_i e + e_1 \ge 0$, we have $e^\top(\alpha_i e + e_1) = p_i \|\alpha_i e + e_1\|$, and by squaring each term,

$$\begin{aligned} (r_i \alpha_i + 1)^2 = p_i^2(\alpha_i^2 r_i + 2\alpha_i + 1) = \frac{p_i^2}{r_i}(\alpha_i r_i + 1)^2 + p_i^2 - \frac{p_i^2}{r_i} \\ \implies (\alpha_i r_i + 1)^2 = p_i^2 \frac{r_i - 1}{r_i - p_i^2}. \end{aligned} \tag{52}$$

Moreover, notice that for any e_j, $e^\top(e - e_j) = r_i - 1 \ge p_i \sqrt{r_i - 1} = p_i \|e - e_j\|$ since by hypothesis $p_i \le \sqrt{r_i - 1}$. As a consequence, $e - e_j \in \mathcal{C}_{p_i}$ and $\alpha_i e + e_j \in \mathcal{C}_{p_i}$. From (51), we find that

$$U_1 V U_2^\top \ge 0 \implies e^\top V(\alpha_2 e + e_j) \ge \|V(\alpha_2 e + e_j)\|, \quad e^\top V(e - e_j) \ge \|V(e - e_j)\| \quad \forall j \tag{53}$$

and, thus,

$$\begin{aligned} (\alpha_2^2 + \alpha_2)\|Ve\|^2 + (\alpha_2 + 1)\|Ve_j\|^2 &= \|V(\alpha_2 e + e_j)\|^2 + \alpha_2 \|V(e - e_j)\|^2 \\ &\le \left(e^\top V(\alpha_2 e + e_j)\right)^2 + \alpha_2 \left(e^\top V(e - e_j)\right)^2 \\ &= (\alpha_2^2 + \alpha_2)\left(e^\top V e\right)^2 + (\alpha_2 + 1)\left(e^\top V e_j\right)^2 \quad \forall j \end{aligned}$$

or, written in a simpler way,

$$\alpha_2 \|Ve\|^2 + \|Ve_j\|^2 \le \alpha_2 \left(e^\top V e\right)^2 + \left(e^\top V e_j\right)^2 \quad \forall j.$$

By summing the above relations over all indices j, we get a bound on the Frobenius norm of V:

$$\|V\|_F^2 = \sum_j \|Ve_j\|^2 \le \alpha_2 r_2 \left(e^\top V e\right)^2 + \|e^\top V\|^2 - \alpha_2 r_2 \|Ve\|^2. \tag{54}$$

By applying the same reasoning by swapping U_1 and U_2, we obtain the symmetric relation

$$\|V\|_F^2 \le \alpha_1 r_1 \left(e^\top V e\right)^2 + \|Ve\|^2 - \alpha_1 r_1 \|e^\top V\|^2. \tag{55}$$

Since $e \in \mathcal{C}_{p_2}$, from (51) we get that $e^\top V e \geq \|Ve\|$ and by exchanging U_1, U_2 we also get the symmetric relation $e^\top V e \geq \|e^\top V\|$. We can thus rewrite the conditions (54) and (55) with a change of variables as

$$\begin{cases} x = \|V\|_F^2 / \left(e^\top V e\right)^2, \\ y = \|Ve\|^2 / \left(e^\top V e\right)^2, \\ z = \|e^\top V\|^2 / \left(e^\top V e\right)^2, \end{cases} \implies \begin{cases} x \leq \alpha_2 r_2 + z - \alpha_2 r_2 y = f_2(y,z), \\ x \leq \alpha_1 r_1 + y - \alpha_1 r_1 z = f_1(y,z), \\ x, y, z \geq 0, \quad y, z \leq 1, \end{cases}$$

and, in particular,

$$x \leq \max_{0 \leq y, z \leq 1} \min\{f_2(y,z), f_1(y,z)\}. \tag{56}$$

Notice that if $f_2(y,z) < f_1(y,z)$, then to raise the value of the minimum, one can increment z to $\overline{z}$ until either $f_2(y,\overline{z}) = f_1(y,\overline{z})$ or $\overline{z} = 1$ and, in the second case, one can decrease y to $\overline{y}$ until again $f_2(\overline{y},1) = f_1(\overline{y},1)$ or $\overline{y} = 0$. Since $f_1(0,1) = 0 < 1 + \alpha_2 r_2 = f_2(0,1)$, we conclude that, by this procedure, we will always reach a point $(\overline{y},\overline{z})$ such that $f_2(\overline{y},\overline{z}) = f_1(\overline{y},\overline{z})$. The same reasoning can be repeated when $f_2(y,z) > f_1(y,z)$ to conclude that we have to check only $f_1(y,z) = f_2(y,z)$. Since

$$\begin{aligned} &f_1(y,z) = f_2(y,z) \iff (1+\alpha_1 r_1) z = (1+\alpha_2 r_2) y + \alpha_1 r_1 - \alpha_2 r_2 \\ &\implies f_1(y,z) = f_2(y,z) = 1 + \frac{\alpha_1\alpha_2 r_1 r_2 - 1}{1+\alpha_1 r_1}(1-y) = 1 + \frac{\alpha_1\alpha_2 r_1 r_2 - 1}{1+\alpha_2 r_2}(1-z), \end{aligned}$$

we have

$$\begin{aligned} \max_{0 \leq y,z \leq 1} \min\{f_2(y,z), f_1(y,z)\} &= \max_{\substack{0 \leq y, z \leq 1 \\ f_1(y,z) = f_2(y,z)}} f_1(y,z) \leq \max_{0 \leq y \leq 1} 1 \\ &\quad + \frac{\alpha_1\alpha_2 r_1 r_2 - 1}{1+\alpha_1 r_1}(1-y) \\ &= \max\left\{1, 1 + \frac{\alpha_1\alpha_2 r_1 r_2 - 1}{1+\alpha_1 r_1}\right\} = \max\left\{1, \alpha_1 r_1 \frac{\alpha_2 r_2 + 1}{\alpha_1 r_1 + 1}\right\}. \end{aligned}$$

Putting together (56), (52), and the hypothesis, we get

$$\begin{aligned} x &\leq \max_{0 \leq y,z \leq 1} \min\{f_2(y,z), f_1(y,z)\} \leq \max\left\{1, \alpha_1 r_1 \frac{\alpha_2 r_2 + 1}{\alpha_1 r_1 + 1}\right\} \\ &= \max\left\{1, 1 + \frac{\alpha_1 r_1(\alpha_2 r_2 + 1) - (\alpha_1 r_1 + 1)}{\alpha_1 r_1 + 1}\right\} \\ &= \max\left\{1, 1 + (\alpha_2 r_2 + 1)[1 - (\alpha_2 r_2 + 1)^{-1} - (\alpha_1 r_1 + 1)^{-1}]\right\} \\ &= \max\left\{1, 1 + \sqrt{p_2^2 \frac{r_2 - 1}{r_2 - p_2^2}}\left[1 - \sqrt{\frac{r_2 - p_2^2}{p_2^2(r_2-1)}} - \sqrt{\frac{r_1 - p_1^2}{p_1^2(r_1 - 1)}}\right]\right\} = 1. \end{aligned}$$

Recalling now the variable change for x, we conclude that $\|V\|_F^2 \leq (e^\top V e)^2$. By (51), we have that $e^\top V e \geq \|Ve\| \geq 0$, so finally $e^\top V e \geq \|V\|_F$, that is, $U_1 \otimes U_2$ satisfies SSC1.

To prove that $U_1 \otimes U_2$ satisfies SSC2, let $v \in \operatorname{cone}^*(U_1^\top \otimes U_2^\top) \cap \mathbf{bd}(\mathcal{C}^*)$ be a nonzero vector. Recasting v as the matrix $V \in \mathbb{R}^{r_1 \times r_2} = \operatorname{mat}(v)$, we can rewrite the condition as $U_1 V U_2^\top \geq 0$ and $e^\top V e = \|V\|_F$. Since we already proved that $e^\top V e \geq \|V\|_F$, we can retrace all the formulas of the proof for SSC1 and substitute the inequality signs with equality ones. In particular, (53) becomes

$$\|V(e-e_j)\| = e^\top V(e-e_j) \quad \text{for all } j. \tag{57}$$

Let now us analyze the concave function $g(z) := e^\top Vz - \|Vz\|$. We know that $g(e-e_j)=0$ for all j and that $g(u_2)\geq 0$ for all $u_2 \in \mathrm{cone}(U_2^\top)$ due to (51).

Since $\mathrm{cone}(U_2^\top)$ is an SSC nonnegative cone, then $\mathcal{C} \subseteq \mathrm{cone}(U_2^\top) \subseteq \mathbb{R}_+^{r_2}$. The point $e-e_j$ belongs to the boundary of all of the three conesl and, in particular, it lies on the relative interior part of the facet $F_j := \{x \in \mathbb{R}_+^{r_2} \mid x_j = 0\}$ of $\mathbb{R}_+^{r_2}$. Since $\mathrm{cone}(U_2^\top)$ is polyhedral, and F_j is tangent to $\mathcal{C}$ in the point $e-e_j$, then necessarily $\mathrm{cone}(U_2^\top)$ must possess a facet $G_j \subseteq F_j$ such that $e-e_j$ lies on the interior part of G_j. Therefore, $g(z)$ is a concave nonnegative function on G_j that has a minimum in the interior point $e-e_j$, so $g(z)|_{G_j} \equiv 0$. It means that $(e^\top Vz)^2 = \|Vz\|^2$ or, equivalently, $z^\top V^\top(ee^\top - I)Vz = 0$ for all $z \in G_j$. This is a quadratic function, that is, in particular, analytic, taking the value zero in an open set $G_j \subseteq \{x_j = 0\}$, thus implying that $z^\top V^\top(ee^\top - I)Vz = 0$ for all z such that $z_j = 0$ by analytic continuation. Moreover, for a quadratic function to be zero on r_2 distinct hyperplanes, necessarily $z^\top V^\top(ee^\top - I)Vz \equiv 0$ implies $V^\top V = V^\top ee^\top V$ or $r_2 = 2$. Doing the same reasoning by inverting U_1 and U_2 brings us to the conclusion that $y^\top V(ee^\top - I)V^\top y = 0$ for all y such that $y_j = 0$ and, in particular, either $VV^\top = Vee^\top V^\top$ or $r_1 = 2$.

If $r_1 = r_2 = 2$, then both U_1 and U_2 are separable and thus $U_1VU_2^\top \geq 0 \implies V \geq 0$ and $e^\top Ve = \|V\|_F \implies V = \lambda e_i e_j^\top$ for some indices i,j and $\lambda \geq 0$.

Otherwise the matrix V must have rank-1 and can be written as $V = pq^\top$ and, since $V \neq 0$, both p,q are nonzero. Since $z^\top V^\top(ee^\top - I)Vz = 0$ for all z that have at least a zero entry, if $q_j \neq 0$ and $z = e_j$, we get

$$(z^\top q)^2((e^\top p)^2 - \|p\|^2) = 0 \implies |e^\top p| = \|p\|,$$

and, if needed by changing the sign of both p,q, we can conclude that $e^\top p = \|p\|$ and $p \in \mathbf{bd}(\mathcal{C}^*)$. Since $e \in \mathrm{cone}(U_1^\top)$, then by (51) for any $u_2 \in \mathrm{cone}(U_2^\top)$, we have

$$e^\top Vu_2 \geq \|Vu_2\| \implies e^\top p \cdot q^\top u_2 \geq \|p\| \cdot |q^\top u_2| \implies q^\top u_2 \geq |q^\top u_2| \geq 0$$

and, therefore, $q \in \mathrm{cone}^*(U_2^\top)$ and, in particular, $q^\top e \geq 0$. Notice that from the symmetric reasoning, we still have that $y^\top V(ee^\top - I)V^\top y = 0$ for all y that have at least a zero entry, so $|e^\top q| = \|q\|$, and from above we conclude that $e^\top q = \|q\|$ and $q \in \mathbf{bd}(\mathcal{C}^*) \cap \mathrm{cone}^*(U_2^\top)$. Analogously, we also find that $p \in \mathbf{bd}(\mathcal{C}^*) \cap \mathrm{cone}^*(U_1^\top)$.

Finally, since U_1 and U_2 satisfy the SSC2, then both p and q are positive multiples of canonical basis vectors. The matrix $V = pq^\top$ is thus a zero matrix with the exception of a single positive entry, and the original vector v will thus also be a positive multiple of some e_i, proving that $U_1 \otimes U_2$ satisfies the SSC2. ◻

Theorem 7.5 is important because it gives precise bounds on the expansion factors p_1, p_2 for the SSC conditions on U_1 and U_2 to achieve the SSC condition on their product $U_1 \otimes U_2$. An immediate corollary is that if one of the two is separable and the other satisfies the SSC, then the product also satisfies the SSC.

Corollary 7.5.1. *Let $U_1 \in \mathbb{R}^{n_1 \times r_1}$ satisfy the SSC and let $U_2 \in \mathbb{R}^{n_2 \times r_2}$ satisfy p_2-SSC for any*

$$p_2^2 \leq \frac{r_2}{1 + (r_2 - 1)f(r_1)}, \quad f(r_1) = \left(1 - \frac{1}{r_1 - 1}\right)^2.$$

Then $U_1 \otimes U_2 \in \mathbb{R}^{n_1 n_2 \times r_1 r_2}$ satisfies the SSC.

In particular, if U_1 satisfies the SSC, and U_2 is separable, then $U_1 \otimes U_2$ satisfies the SSC.

Proof. Since U_1 satisfies the SSC, then it is p_1-SSC for some $p_1^2 \le r_1 - 1$. If we now analyze the condition on p_1 and p_2 given by Theorem 7.5, we find that

$$\sqrt{\frac{r_1 - p_1^2}{p_1^2(r_1-1)}} + \sqrt{\frac{r_2 - p_2^2}{p_2^2(r_2-1)}} \ge \frac{1}{r_1 - 1} + \sqrt{\frac{r_2[1+(r_2-1)f(r_1)] - r_2}{r_2(r_2-1)}}$$
$$= \frac{1}{r_1-1} + \sqrt{f(r_1)} = 1$$

and thus $U_1 \otimes U_2$ satisfies the SSC. Moreover, since $0 \le f(r_1) < 1$, then the upper bound on p_2 in the hypothesis is at least 1. In particular, if $p_2 = 1$ then U_2 is separable and $U_1 \otimes U_2$ satisfies the SSC. ∎

We can now turn to Question 7.1 and give a positive answer when the dimension r_i is low enough.

COROLLARY 7.5.2. *Let (r_1, r_2) with $r_1 = 2$ or $r_2 = 2$ or $r_1 = r_2 = 3$. For any $U_1 \in \mathbb{R}^{n_1 \times r_1}$ and $U_2 \in \mathbb{R}^{n_2 \times r_2}$ satisfying the SSC, their Kronecker product $U_1 \otimes U_2$ satisfies the SSC.*

Proof. Let $r_1 = 2$. Since U_1 satisfies SSC, then U_1 is also separable, and the results follows from Corollary 7.5.1. The same argument can be repeated for $r_2 = 2$.

The only case left is $r_1 = r_2 = 3$. Since U_1, U_2 satisfy the SSC, then $p_1^2, p_2^2 \le 2$ and the condition in Theorem 7.5 reads as

$$\sqrt{\frac{r_1 - p_1^2}{p_1^2(r_1-1)}} + \sqrt{\frac{r_2 - p_2^2}{p_2^2(r_2-1)}} = \sqrt{\frac{3 - p_1^2}{2p_1^2}} + \sqrt{\frac{3 - p_2^2}{2p_2^2}} \ge \sqrt{\frac{1}{4}} + \sqrt{\frac{1}{4}} = 1,$$

thus proving that $U_1 \otimes U_2$ satisfies the SSC. ∎

Sadly, the same result does not hold for all other choices of dimensions (r_1, r_2), as shown in the next section.

7.2. Negative answer to Question 7.1 for $r_1, r_2 \ge 3$ and $\max\{r_1, r_2\} \ge 4$. Notice that the definitions for SSC and p-SSC only talk about cones, thus we can always normalize all the nonzero rows of U into stochastic vectors. For this reason, from now on we focus only on row stochastic matrices.

The condition SSC1 on $U_1 \otimes U_2$ is equivalent to the condition

$$(U_1 \otimes U_2)v \ge 0 \implies e^\top v \ge \|v\|,$$

and by recasting v to a matrix $V \in \mathbb{R}^{r_1 \times r_2}$, that is, $V = \mathrm{mat}(v)$, it can be written as

$$U_1 V U_2^\top \ge 0 \implies e^\top V e \ge \|V\|_F.$$

As a consequence, to have a negative answer to Question 7.1; it is enough to look for matrices $V \in \mathbb{R}^{r_1 \times r_2}$ for which $U_1 V U_2^\top \ge 0$ and $e^\top V e < \|V\|_F$ when both U_1 and U_2 satisfy the SSC.

LEMMA 7.6. *Let $U_i \in \mathbb{R}_+^{n_i \times r_i}$, where $\mathrm{cone}(U_i^\top) \subseteq \mathcal{C}_{p_i}$ for $i = 1, 2$, $r_1 \le r_2$, and*

$$\left(\frac{1}{p_1^2} - \frac{1}{r_1}\right)\left(\frac{1}{p_2^2} - \frac{1}{r_2}\right) < \frac{1}{r_2 r_1}\frac{r_1 - 1}{r_1 r_2 - 1}, \tag{58}$$

then the product $U_1 \otimes U_2$ does not satisfy SSC1.

Proof. Call $c_i = \frac{1}{p_i^2} - \frac{1}{r_i}$ and notice that we can always consider U_i row stochastic. If u_i is any row of U_i, then

$$\begin{aligned}\operatorname{cone}(U_i^\top) \subseteq \mathcal{C}_{p_i} \implies & 1 = (e^\top u_i)^2 \geq p_i^2 \|u_i\|^2 = p_i^2(\|u_i - e/r_i\|^2 + \|e/r_i\|^2) \\ \implies & \|u_i - e/r_i\|^2 \leq \frac{1}{p_1^2} - \frac{1}{r_i} = c_i.\end{aligned}$$

The hypotheses (58) can thus be rewritten as

$$\begin{aligned}&\max_{i\in[n_1]} \|U_1(i,:)^\top - e/r_1\|^2 \leq c_1, \qquad \max_{j\in[n_2]} \|U_2(j,:)^\top - e/r_2\|^2 \leq c_2, \\ &c_1 c_2 < \frac{1}{r_2 r_1}\frac{r_1 - 1}{r_1 r_2 - 1}.\end{aligned} \tag{59}$$

Let now $V = \lambda E/\sqrt{r_1 r_2} + B \in \mathbb{R}^{r_1 \times r_2}$, where E is the matrix of all ones, $\lambda \geq \sqrt{c_1 c_2 r_1 r_2}$, and B is any real matrix satisfying

$$Be = 0, \quad e^\top B = 0, \quad \|B\| = 1, \quad \|B\|_F^2 = r_1 - 1.$$

This can be realized as $B = Q_1 Q_2^\top$, where $Q_1 \in \mathbb{R}^{r_1 \times (r_1 - 1)}$, $Q_2 \in \mathbb{R}^{r_2 \times (r_1 - 1)}$ are matrices with orthonormal columns such that $e^\top Q_1 = 0$, $e^\top Q_2 = 0$. We can thus compute

$$\begin{aligned}U_1 V U_2^\top &= \frac{\lambda}{\sqrt{r_1 r_2}} U_1 E U_2^\top + U_1 B U_2^\top = \frac{\lambda}{\sqrt{r_1 r_2}} E + \left(U_1 - \frac{E}{r_1}\right) B \left(U_2^\top - \frac{E}{r_2}\right) \\ &\geq \frac{\lambda}{\sqrt{r_1 r_2}} E - \left(\max_{i\in[n_1]} \|U_1(i,:)^\top - e/r_1\| \max_{j\in[n_2]} \|U_2(j,:)^\top - e/r_2\|\right) \|B\| E \\ &= \left(\frac{\lambda}{\sqrt{r_1 r_2}} - \sqrt{c_1 c_2}\right) E \geq 0, \\ (e^\top V e)^2 &= \lambda^2 r_1 r_2, \\ \|V\|_F^2 &= \left\|\frac{\lambda}{\sqrt{r_1 r_2}} E\right\|_F^2 + \|B\|_F^2 = \lambda^2 + r_1 - 1.\end{aligned}$$

Hence $U_1 V U_2^\top \geq 0$, and $e^\top V e < \|V\|_F$ holds if and only if $\lambda^2 < \frac{r_1 - 1}{r_1 r_2 - 1}$. Since $\lambda \geq \sqrt{c_1 c_2 r_1 r_2}$, then for

$$c_1 c_2 < \frac{1}{r_1 r_2}\frac{r_1 - 1}{r_1 r_2 - 1}$$

we can find a λ such that $U_1 V U_2^\top \geq 0$ and $e^\top V e < \|V\|_F$, that is, the matrix $U_1 \otimes U_2$ does not satisfy the SSC1. ∎

Using Lemma 7.6, we can now prove Theorem 7.6.

Proof of Theorem 7.6. Suppose w.l.o.g. that $r_1 \leq r_2$. Let us look for U_1, U_2, c_1, c_2 satisfying (59) with U_1, U_2 SSC, so that $U_1 \otimes U_2$ automatically does not satisfy SSC1 by Lemma 7.6.

If U_1 satisfies SSC1, then there exists a column u_1 of $U_1^\top$ that is also an extreme ray for $\operatorname{cone}(U_1^\top)$ that does not lie in $\mathcal{C}$, and thus

$$c_1 \geq \left\|u_1 - \frac{e}{r_1}\right\|^2 = \|u_1\|^2 - \frac{1}{r_1} > \frac{1}{r_1 - 1} - \frac{1}{r_1} = \frac{1}{r_1(r_1 - 1)}.$$

Analogously, there exists a column u_2 of $U_2^\top$ that is an extreme ray of cone$(U_2^\top)$ and respects $c_2 \geq \|u_2 - e/r_2\|^2 > 1/(r_2(r_2-1))$. Therefore a necessary condition for the existence of c_1, c_2 satisfying (59) is

$$\frac{1}{r_2(r_2-1)}\frac{1}{r_1(r_1-1)} < c_1c_2 < \frac{1}{r_2r_1}\frac{r_1-1}{r_1r_2-1} \iff r_1r_2 - 1 < (r_1-1)^2(r_2-1). \tag{60}$$

It is easy to see that for $r_1 \geq 3$ and $r_2 \geq 4$, the last relation of (60) is satisfied. If we impose $c_ir_i(r_i-1) = \lambda$ for $i = 1, 2$, then c_1, c_2 satisfy (60) whenever

$$1 < \lambda^2 < \frac{\frac{1}{r_2r_1}\frac{r_1-1}{r_1r_2-1}}{\frac{1}{r_2(r_2-1)}\frac{1}{r_1(r_1-1)}} = \frac{(r_2-1)(r_1-1)^2}{r_1r_2-1},$$

which holds, for example, for

$$\lambda = \sqrt[4]{\frac{(r_2-1)(r_1-1)^2}{r_1r_2-1}}.$$

To conclude, we need to build U_i satisfying the SSC such that any row u_i and, in particular, any vertex of cone$(U_i^\top)$, satisfy $\|u_i - e/r_i\|^2 \leq c_i$. In [10], it is proved that for any $\epsilon > \delta > 0$ there exists a polyhedral cone $\mathcal{G}_\epsilon^r$ such that

$$\mathcal{L}_\delta^r \subseteq \mathcal{G}_\epsilon^r \subseteq \mathcal{L}_\epsilon^r, \qquad \mathcal{L}_\gamma^r := \{(x,t) \in \mathbb{R}^r \mid (1+\gamma)t \geq \|x\|\}.$$

If $Q \in \mathbb{R}^{r\times r}$ is any orthogonal matrix mapping e_1 into $e/\sqrt{r}$, and $D \in \mathbb{R}^{r\times r}$ a diagonal matrix with all diagonal elements equal to $\frac{1}{\sqrt{r-1}}$ except for the first equal to 1, then it can be easily proved that $QD\mathcal{L}_0^r = \mathcal{C}$ and, more generally, that

$$y \in QD\mathcal{L}_\gamma^r, \quad e^\top y = 1 \implies \|y - e/r\|^2 \leq \frac{(1+\gamma)^2}{r(r-1)} \implies QD\mathcal{L}_\gamma^r \cap \mathbb{R}_+^r = \mathcal{C}_{\sqrt{\frac{r(r-1)}{(1+\gamma)^2+r-1}}}.$$

If now we set ϵ such that $(1+\epsilon)^2 = \lambda$ and cone$(U_i^\top) = QD\mathcal{G}_\epsilon^{r_i} \cap \mathbb{R}_+^{r_i}$, we conclude that

$$\mathcal{L}_\delta^{r_i} \subseteq \mathcal{G}_\epsilon^{r_i} \subseteq \mathcal{L}_\epsilon^{r_i} \implies \mathcal{C} \subsetneq \mathcal{C}_{\sqrt{\frac{r_i(r_i-1)}{(1+\delta)^2+r_i-1}}} \subseteq \operatorname{cone}(U_i^\top) \subseteq QD\mathcal{L}_\epsilon^{r_i}.$$

In particular, any row u_i of U_i satisfies $\|u_i - e/r_i\|^2 \leq \frac{\lambda}{r_i(r_i-1)} = c_i$, and U_i is SSC since it is p_i-SSC with $p_i^2 = \frac{r_i(r_i-1)}{(1+\delta)^2+r_i-1} < r_i - 1$. □

The last result does not present an explicit construction for U_1 and U_2 that negates Question 7.1, but it proves its existence. The proof relies on the construction of [10, Proposition 3.1] that also gives us a lower bound on the number of rows for an SSC matrix $U_i \in \mathbb{R}^{n_i\times r_i}$. Recall that by Theorem 7.6, we need that $\mathcal{C} = \mathcal{C}_{\sqrt{r_i-1}} \subseteq$ cone$(U_i^\top) \subseteq \mathcal{C}_{p_i}$. If $\sqrt{r_i-1} - p_i$ is quite small, then $\mathcal{C}_{p_i}$ and $\mathcal{C}_{\sqrt{r_i-1}}$ are very close and every point on the boundary of $\mathcal{C}_{p_i}$ must be close enough to an extreme ray of cone$(U_i^\top)$. Once we have an upper bound δ on this distance, we find that the union of the balls with center a vertex of cone$(U_i^\top)$ and radius δ must cover the surface of $\mathcal{C}_{p_i}$. As a consequence, we can compute a lower bound on the number of balls (and thus vertices) needed for the covering. Following the proof of Theorem 7.2, one obtains

$$n_i = \left(\frac{1}{\epsilon}\right)^{\Omega(r_i)} = \left(\frac{1}{\lambda-1}\right)^{\Omega(r_i)} = \left(\sqrt[4]{\frac{(r_2-1)(r_1-1)^2}{r_1r_2-1}} - 1\right)^{-\Omega(r_i)}. \tag{61}$$

Theorem 7.6 thus ensures that whenever $\min\{r_1, r_2\} \geq 3$ and $\max\{r_1, r_2\} \geq 4$, we can always generate U_1 and U_2 that satisfy the SSC, and whose Kronecker product does not satisfy the SSC1.

7.3. Numerical experiment. As shown in Theorem 7.2, for $\min(r_1, r_2) \geq 3$ and $\max(r_1, r_2) \geq 4$, the Kronecker product of two sufficiently scattered matrices with r_1 and r_2 columns does not necessarily satisfy the SSC. However, this is a worse case scenario where n_1 and n_2 need to be very large in the construction of our counterexamples (see (61)), and we have observed experimentally that this Kronecker product often satisfies the SSC. In fact, when we started to study this question, we were trying to find counter examples numerically, generating randomly matrices satisfying the SSC, but we were never able to generate a counterexample with such a random generation process. In particular, we randomly generated U_1 and U_2 of dimensions 20×4, each of which have two nonzeros per row, as done in [24, section 5.1]. Using the algorithm to check the SSC criterion [24, Algorithm 1], in 13 out of the 20 cases, both U_1 and U_2 satisfy the SSC. In each of these 13 cases, their Kronecker product, $U_1 \otimes U_2 \in \mathbb{R}^{400 \times 16}$, also satisfied the SSC.

7.4. Future Directions. Since the Kronecker product of randomly generated sufficiently scattered matrices appear to often satisfy the SSC, it would be interesting to explain this behavior theoretically (e.g., using some well-defined generation process for U_1 and U_2, and quantifying the probability for their Kronecker product to satisfy the SSC). However, this is a hard question because checking the SSC is NP-hard in general [28].

For tensors of order larger than 3, we might need the product of more than two factors to satisfy the SSC; see Theorems 6.4 and 6.9. Theorem 7.5 cannot be used in this context: it only guarantees the Kronecker product to be SSC, not p-SSC, which would be needed to use Theorem 7.5 recursively. Ideally, we would need to show that the Kronecker product between two matrices, one satisfying p_1-SSC and the other satisfying p_2-SSC, satisfy p-SSC for some p depending on p_1 and p_2 (note that, for $p_1 = p_2 = 1$, $p = 1$; this is the separable case.) This is a topic for further research.

8. Conclusion. In this paper, we provided new, stronger, identifiability results for nTDs, $\mathcal{T} = (U_1, \ldots, U_d).\mathcal{G}$, where $U_i \geq 0$ for $i \in [d]$. They relied on two main conditions:

1. The factor matrices, U_i for $i \in [d]$, need to satisfy some sparsity conditions, namely, the separability condition (Definition 2.5), the sufficiently scattered condition (the SSC, Definition 2.7), or their product must satisfy the SSC (see section 7).
2. The core tensor, $\mathcal{G}$, needs to have some slices or unfoldings of maximum rank, and its dimensions need to satisfy some conditions (e.g., it must be square for order-2 nTDs).

Under such conditions, our identifiability results relied on minimizing some measure of the volume of the core tensor, while combining two main ingredients: the identifiability of min-vol NMF (Theorem 2.8 from [17]), and of min-vol order-2 nTDs (which is new; see Theorem 4.2). For order-3 tensors, these results are summarized in Table 1, and they generalize to any order (section 6).

In Part II of this paper [44], we will discuss how to compute these identifiable solutions, and compare the different variants, both in terms of computational cost and practical effectiveness to recover the hidden factors. We will compare our algorithms to the state of the art on synthetic and real data sets.

Appendix A. Identifiability of NTD under separability assumptions. Since separability implies the SSC, our results apply to the more constrained case when the U_i's are separable. In this section, we explicitly state two such results for completeness.

A.1. Generalization of Proposition 1 from [2]. One can consider a generalization of Theorem 3.1, and give an identifiability result for the following model:

$$\mathcal{T} = (U_1, \dots, U_d).\mathcal{G} \quad \text{with } U_i \in \mathbb{R}_+^{n_i \times r_i} \text{ separable for } i \in [d], \text{ and } \mathcal{G} \in R^{r_1 \times \cdots \times r_d}. \tag{62}$$

In the following theorem, we show that the identifiability result from [2] can be generalized to any order-d tensor.

THEOREM A.1 (generalization of Theorem 3.1). *Let $\mathcal{T} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ be an order-d tensor that admits a decomposition of the form* (62)*. Moreover if for all $k \in [d]$, the unfolding of the tensor $\mathcal{T}$ along the kth mode, denoted by $\mathcal{T}_{(k)}$, has rank r_k, then for any other decomposition of the form $\mathcal{T} = (U_1^*, \dots, U_d^*).\mathcal{G}^*$, where U_i^* are separable matrices, there exist permutation matrices Π_i such that $U_i^* = U_i \Pi_i$ and $\mathcal{G}^* = (\Pi_1^\top, \dots, \Pi_d^\top).\mathcal{G}$.*

Proof. Recall from Property 6.2 that the unfolding of a tensor $\mathcal{T}$ along the mode k, denoted by $\mathcal{T}_{(k)}$ has a decomposition of the form

$$\mathcal{T}_{(k)} = \left(\bigotimes_{j \neq k} U_j \right) \mathcal{G}_{(k)} U_k^\top. \tag{63}$$

Using Theorem 2.6, we get that for any other decomposition of $\mathcal{T}_{(k)}$ of the form $\mathcal{T}_{(k)} = W_k^* (U_k^*)^\top$ such that U_k^* is a separable matrix, there exists a permutation matrix Π_k and diagonal matrix D_k such that $W_k^* = (\bigotimes_{j \neq k} U_j) \mathcal{G}_{(k)} D_k \Pi_k$ and $U_k^* = U_k D_k^{-1} \Pi_k$ for all $k \in [d]$. Moreover, since $U_k \in [0,1]^{n_k \times r_k}$, D_k is also the identity matrix of size r_k.

Since U_k^* and $U_k^\#$ are separable for all $k \in [d]$, they have full column rank. Then $(U_k^*)^\dagger U_k^* = (U_k^\#)^\dagger U_k^\# = I_{r_k}$ and from this, we can conclude that

$$\mathcal{G}^* = ((U_1^*)^\dagger, \dots, (U_d^*)^\dagger).\mathcal{T} = (\Pi_1^\top, \dots, \Pi_d^\top).((U_1^\#)^\dagger, \dots, (U_d^\#)^\dagger).\mathcal{T} = (\Pi_1^\top, \dots, \Pi_d^\top).\mathcal{G}^\#. \qquad \square$$

A.2. Unfoldings along one mode in the separable case. *Assumption* A.2. Let $\mathcal{T}$ be an order-d tensor such that it has a TD of the form $\mathcal{T} = (U_1^\#, \dots, U_d^\#).\mathcal{G}$, where

- $U_i^\# \in \mathbb{R}_+^{n_i \times r_i}$ is separable for all $i \in [d]$;
- there exists $\mathcal{I} \subset [d]$ such that $\operatorname{rank}(\mathcal{G}_{\mathcal{I}}) = \prod_{i \in \mathcal{I}} r_i = \prod_{j \in \mathcal{I}^-} r_j$.

Following the proof of Theorem 6.4 and using the fact that if U_1, U_2 are separable matrices, then $U_1 \otimes U_2$ is separable (this is a direct consequence of Corollary 7.5.1), one can show that if the given tensor satisfies Assumption A.2, then Procedure d.0 returns an essentially unique nTD. Note, however, that using separability allows one to design more effective algorithms (running in polynomial time with provable guarantees, even in the presence of noise); see Part II of this paper [44].

REFERENCES

[1] M. ABDOLALI AND N. GILLIS, *Simplex-structured matrix factorization: Sparsity-based identifiability and provably correct algorithms*, SIAM J. Math. Data Sci., 3 (2021), pp. 593–623, https://doi.org/10.1137/20M1354982.

[2] J. Agterberg and A. R. Zhang, *Estimating higher-order mixed memberships via the $\ell_{2,\mathrm{inf}}$ tensor perturbation bound*, J. Amer. Statist. Assoc., 120 (2024), pp. 1–11, https://doi.org/10.1080/01621459.2024.2404265.

[3] E. M. Airoldi, D. Blei, S. Fienberg, and E. Xing, *Mixed membership stochastic blockmodels*, J. Mach. Learn. Res., 9 (2008), pp. 1981–2014.

[4] A. Anandkumar, D. J. Hsu, M. Janzamin, and S. M. Kakade, *When are overcomplete topic models identifiable? Uniqueness of tensor Tucker decompositions with structured sparsity*, J. Mach. Learn. Res., 16 (2015), pp. 2643–2694.

[5] S. Arora, R. Ge, Y. Halpern, D. Mimno, A. Moitra, D. Sontag, Y. Wu, and M. Zhu, *A practical algorithm for topic modeling with provable guarantees*, JMLR Workshop Conf. Proc., 28 (2013), pp. 280–288.

[6] S. Arora, R. Ge, R. Kannan, and A. Moitra, *Computing a nonnegative matrix factorization–provably*, in ACM Symposium on Theory of Computing, ACM, New York, 2012, pp. 145–162.

[7] B. W. Bader, T. G. Kolda, *MATLAB Tensor Toolbox Version* 2.6, (2015) https://github.com/andrewssobral/tensor_toolbox.

[8] X. Bai, F. Xu, L. Zhou, Y. Xing, L. Bai, and J. Zhou, *Nonlocal similarity based nonnegative Tucker decomposition for hyperspectral image denoising*, IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens, 11 (2018), pp. 701–712, https://doi.org/10.1109/JSTARS.2018.2791718.

[9] G. Ballard and T. G. Kolda, *Tensor Decompositions for Data Science*, Cambridge University Press, Cambridge, 2025.

[10] A. Ben-Tal and A. Nemirovski, *On polyhedral approximations of the second-order cone*, Math. Oper. Res., 26 (2001), pp. 193–205, https://doi.org/10.1287/moor.26.2.193.10561.

[11] J. W. Boardman, F. A. Kruse, and R. O. Green, *Mapping Target Signatures via Partial Unmixing of AVIRIS Data*, NASA technical report N95-33736, 1995, pp. 23–26.

[12] A. Cichocki, R. Zdunek, A. H. Phan, and S. I. Amari, *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*, Wiley, New York, 2009.

[13] M. D. Craig, *Minimum-volume transforms for remotely sensed data*, IEEE Trans. Geosci. Remote Sens., 32 (1994), pp. 542–552, https://doi.org/10.1109/36.297973.

[14] C. Ding, T. Li, W. Peng, and H. Park, *Orthogonal nonnegative matrix t-factorizations for clustering*, in ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, New York, 2006, pp. 126–135.

[15] I. Domanov and L. De Lathauwer, *Canonical polyadic decomposition of third-order tensors: Relaxed uniqueness conditions and algebraic algorithm*, Linear Algebra Appl., 513 (2017), pp. 342–375, https://doi.org/10.1016/j.laa.2016.10.019.

[16] D. Donoho and V. Stodden, *When does non-negative matrix factorization give a correct decomposition into parts?*, in Advances in Neural Information Processing Systems, Curran Associaters, Red Hook, NY, 16 2004, pp. 1141–1148.

[17] X. Fu, K. Huang, and N. D. Sidiropoulos, *On identifiability of nonnegative matrix factorization*, IEEE Signal Process. Lett., 25 (2018), pp. 328–332, https://doi.org/10.1109/LSP.2018.2789405.

[18] X. Fu, K. Huang, N. D. Sidiropoulos, and W. K. Ma, *Nonnegative matrix factorization for signal and data analytics: Identifiability, algorithms, and applications*, IEEE Signal Process. Mag., 36 (2019), pp. 59–80, https://doi.org/10.1109/MSP.2018.2877582.

[19] X. Fu, K. Huang, N. D. Sidiropoulos, Q. Shi, and M. Hong, *Anchor-free correlated topic modeling*, IEEE Trans. Pattern Anal. Mach. Intell., 41 (2019), pp. 1056–1071, https://doi.org/10.1109/TPAMI.2018.2827377.

[20] X. Fu, K. Huang, B. Yang, W. K. Ma, and N. D. Sidiropoulos, *Robust volume minimization-based matrix factorization for remote sensing and document clustering*, IEEE Trans. Signal Process., 64 (2016), pp. 6254–6268, https://doi.org/10.1109/TSP.2016.2602800.

[21] X. Fu, W. K. Ma, K. Huang, and N. D. Sidiropoulos, *Blind separation of quasi-stationary sources: Exploiting convex geometry in covariance domain*, IEEE Trans. Signal Process., 63 (2015), pp. 2306–2320, https://doi.org/10.1109/TSP.2015.2404577.

[22] 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 Process. Mag., 37 (2020), pp. 78–94, https://doi.org/10.1109/MSP.2020.3003544.

[23] N. Gillis, *Nonnegative Matrix Factorization*, SIAM, Philadelphia, 2020.

[24] N. Gillis and R. Luce, *Checking the sufficiently scattered condition using a global non-convex optimization software*, IEEE Signal Process. Lett., 31 (2024), pp. 1610–1614, https://doi.org/10.1109/LSP.2024.3411521.

[25] S. Hassanzadeh and A. Karami, *Compression and noise reduction of hyperspectral images using non-negative tensor decomposition and compressed sensing*, Eur. J. Remote Sens., 49 (2016), pp. 587–598, https://doi.org/10.5721/EuJRS20164931.

[26] K. Huang, X. Fu, and N. D. Sidiropoulos, *Anchor-free correlated topic modeling: Identifiability and algorithm*, Adv. Neural Inform. Process. Syst., 29 (2016), pp. 1794–1802.

[27] K. Huang, N. D. Sidiropoulos, and A. P. Liavas, *A flexible and efficient algorithmic framework for constrained matrix and tensor factorization*, IEEE Trans. Signal Process., 64 (2016), pp. 5052–5065, https://doi.org/10.1109/TSP.2016.2576427.

[28] K. Huang, N. D. Sidiropoulos, and A. Swami, *Non-negative matrix factorization revisited: Uniqueness and algorithm for symmetric decomposition*, IEEE Trans. Signal Process., 62 (2013), pp. 211–224, https://doi.org/10.1109/TSP.2013.2285514.

[29] P. Koiran, *An efficient uniqueness theorem for overcomplete tensor decomposition*, in ACM-SIAM Symposium on Discrete Algorithms, SIAM, Philadelphia, 2025, pp. 1909–1932.

[30] P. Koiran and S. Saha, *Absolute reconstruction for sums of powers of linear forms: Degree* 3 *and beyond*, Comput. Complex., 32 (2023), 8, https://doi.org/10.1007/s00037-023-00239-8.

[31] P. Koiran and S. Saha, *Undercomplete Decomposition of Symmetric Tensors in Linear Time, and Smoothed Analysis of the Condition Number*, preprint, https://arxiv.org/abs/2403.00643, 2024.

[32] P. Koiran and S. Saha, *Complete decomposition of symmetric tensors in linear time and polylogarithmic precision*, Theoret. Comput. Sci., 1037 (2025), 115159, https://doi.org/10.1016/j.tcs.2025.115159.

[33] T. G. Kolda and B. W. Bader, *Tensor decompositions and applications*, SIAM Rev, 51 (2009), pp. 455–500, https://doi.org/10.1137/07070111X.

[34] C. Kolomvakis, A. Vandaele, and N. Gillis, *Boolean matrix tri-factorization*, in IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, Piscataway, NJ, 2025, pp. 1–5.

[35] J. Kossaifi, Y. Panagakis, A. Anandkumar, and M. Pantic, *Tensorly: Tensor learning in python*, J. Mach. Learn. Res., 20 (2019), pp. 1–6, http://jmlr.org/papers/v20/18-277.html.

[36] P. K. Kothari, A. Moitra, and A. S. Wein, *Overcomplete tensor decomposition via Koszul-Young flattenings*, in IEEE Symposium on Foundations of Computer Science, IEEE, Piscataway, NJ, 2025, pp. 1874–1885.

[37] J. B. Kruskal, *Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics*, Linear Algebra Appl., 18 (1977), pp. 95–138, https://doi.org/10.1016/0024-3795(77)90069-6.

[38] V. Leplat, N. Gillis, and A. M. Ang, *Blind audio source separation with minimum-volume beta-divergence NMF*, IEEE Trans. Signal Process., 68 (2020), pp. 3400–3410, https://doi.org/10.1109/TSP.2020.2991801.

[39] C. H. Lin, W. K. Ma, W. C. Li, C. Y. Chi, and A. Ambikapathi, *Identifiability of the simplex volume minimization criterion for blind hyperspectral unmixing: The no-pure-pixel case*, IEEE Trans. Geosci. Remote Sens., 53 (2015), pp. 5530–5546, https://doi.org/10.1109/TGRS.2015.2424719.

[40] Q. Liu, L. Lu, and Z. Chen, *Non-negative Tucker decomposition with graph regularization and smooth constraint for clustering*, Pattern Recognit., 148 (2024), 110207, https://doi.org/10.1016/j.patcog.2023.110207.

[41] A. Marmoret, J. E. Cohen, N. Bertin, and F. Bimbot, *Uncovering Audio Patterns in Music with Nonnegative Tucker Decomposition for Structural Segmentation*, in Proceedings of the 21st International Society for Music Information Retrieval Conference (ISMIR), 2020, pp. 788–794.

[42] Z. Rošt'áková, R. Rosipal, S. Seifpour, and L. J. Trejo, *A comparison of non-negative Tucker decomposition and parallel factor analysis for identification and measurement of human EEG rhythms*, Meas. Sci. Rev., 20 (2020), pp. 126–138, https://doi.org/10.2478/msr-2020-0015.

[43] S. Saha, *Algebraic and Numerical Algorithms for Symmetric Tensor Decompositions*, Thesis, Ecole Normale Supérieure de Lyon Lyon, France, 2023.

[44] S. Saha, G. Barbarino, and N. Gillis, *Identifiability of Nonnegative Tucker Decompositions —Part* II: *Algorithms and Applications*, in preparation.

[45] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, *Tensor decomposition for signal processing and machine learning*, IEEE Trans. Signal Process., 65 (2017), pp. 3551–3582, https://doi.org/10.1109/TSP.2017.2690524.

[46] J. B. Smith, Y. Kawasaki, and M. Goto, *Unmixer: An interface for extracting and remixing loops*, in International Society for Music Information Retrieval (ISMIR), ISMIR, 2020. pp. 824–831.

[47] Y. Sun and K. Huang, *Volume-regularized nonnegative Tucker decomposition with identifiability guarantees*, in IEEE International Conference on Acoustics, Speech and Signal Processing, Piscataway, NJ, 2023.

[48] C. F. Van Loan and N. Pitsianis, *Approximation with Kronecker Products*, Springer, Dordrecht, Netherlands, 1993, pp. 293–314.

[49] N. Vervliet, O. Debals, L. Sorber, M. Van Barel, and L. De Lathauwer, *Tensorlab* 3.0 — *Numerical optimization strategies for large-scale constrained and coupled matrix/tensor factorization,* in 2016 50th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 2016, pp. 1733–1738, https://doi.org/10.1109/ACSSC.2016.7869679.

[50] O. Vu Thanh, *Low-rank Matrix Factorizations with Volume-based Constraints and Regularizations*, Ph.D. thesis, University of Mons, Mons, Belgium, 2024.

[51] H. Wang, F. Nie, H. Huang, and F. Makedon, *Fast nonnegative matrix tri-factorization for large-scale data co-clustering*, in International Joint Conference on Artificial Intelligence, https://www.ijcai.org/Proceedings/11/Papers/261.pdf (2011).

[52] M. Wang and Y. Zeng, *Multiway clustering via tensor block models*, Adv. Neural Inform. Process. Syst., 32 (2019), pp. 715–725.

[53] W. Yin, Y. Qu, Z. Ma, and Q. Liu, *HyperNTF: A hypergraph regularized nonnegative tensor factorization for dimensionality reduction*, Neurocomputing, 512 (2022), pp. 190–202.

[54] M. Zare, M. S. Helfroush, K. Kazemi, and P. Scheunders, *Hyperspectral and multispectral image fusion using coupled non-negative Tucker tensor decomposition*, Remote Sens., 13 (2021), 2930, https://doi.org/10.3390/rs13152930.

[55] Y. Zhang and D. Y. Yeung, *Overlapping community detection via bounded nonnegative matrix tri-factorization*, in ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, New York, 2012, pp. 606–614.

[56] G. Zhou, A. Cichocki, Q. Zhao, and S. Xie, *Efficient nonnegative Tucker decompositions: Algorithms and uniqueness*, IEEE Trans. Image Process., 24 (2015), pp. 4990–5003, https://doi.org/10.1109/TIP.2015.2478396.