arXiv:1909.07570v3 [math.NA] 21 Mar 2022

Nonnegative Canonical Tensor Decomposition with Linear Constraints: nnCANDELINC

Boian Alexandrov[a], Derek DeSantis[b], Gianmarco Manzini[c], Erik Skau[d]

[a] *Group T-1, Theoretical Division, Los Alamos National Laboratory, New Mexico, USA*
[b] *Group CCS-2, Computer, Computational, and Statistical Sciences Division, Los Alamos National Laboratory, New Mexico, USA*
[c] *Group T-5, Theoretical Division, Los Alamos National Laboratory, New Mexico, USA*
[d] *Group CCS-3, Computer, Computational, and Statistical Sciences Division, Los Alamos National Laboratory, New Mexico, USA*

Abstract

There is an emerging interest in tensor factorization applications in big-data analytics and machine learning. To speed up the factorization of extra-large datasets, organized in multidimensional arrays (aka tensors), easy to compute compression-based tensor representations, such as Tucker and Tensor Train formats, are used to approximate the initial large-tensor. Further, tensor factorization is used to extract latent features that can facilitate discoveries of new mechanisms and signatures hidden in the data, where the explainability of the latent features is of principal importance. Nonnegative tensor factorization extracts latent features that are naturally sparse and parts of the data, which makes them easily interpretable. However, to take into account available domain knowledge and subject matter expertise, additional constraints often need to be imposed, which lead us to Canonical decomposition with linear constraints (CANDELINC), a Canonical Polyadic Decomposition with rank deficient factors. In CANDELINC, Tucker compression is used as a pre-processing step, which leads to a larger residual error but to more explainable latent features. Here, we propose a nonnegative CANDELINC (nnCANDELINC) accomplished via a specific nonnegative Tucker decomposition; we refer to as minimal or canonical nonnegative Tucker. We derive several results required to understand the specificity of nnCANDELINC, focusing on the difficulties of preserving the nonnegative rank to its Tucker core and comparing the real-valued to the nonnegative case. Finally, we demonstrate nnCANDELINC performance on synthetic and real-world examples.

Keywords: Nonnegative Tucker, Minimal cones, Nonnegative rank, Nonnegative multirank, Nonnegative CANDELINC, linear constraints, data compression

---

1. Introduction

Large amounts of high-dimensional data are constantly generated by sensor networks; large-scale scientific experiments; massive computer simulations; complex engineering activities; electronic communications; social networks, and many other sources [1]. Utilizing such big-data for decision making, emergency response, and data-driven science requires understanding the processes underlying the data [2]. High-dimensional data are naturally organized in tensors (i.e., multi-dimensional arrays). Tensor factorization is a cutting-edge factor analysis that can serve for latent features extraction, dimensional reduction, blind source separation, data mining, pattern recognition, subspace learning, data fusion, compression, and many other applications [3, 4]. A tensor factorization's main objective is to decompose high-dimensional data into factor matrices and one, or in the case of tensor networks [5], several core-tensors of a smaller size.

The number of the tensor entries scales exponentially with tensor dimension, which leads to exponential scaling of the burden of any tensor computation, in terms of storage and floating point operations. This phenomenon is known as the *curse of dimensionality.* One way to speed up tensor calculations and decrease the needed storage is to use stable compression-based representations of the large initial tensors, and then to extract the needed information from the compressed data [6]. Some of the proposed stable compression-based formats are Tucker [7] (related to the multirank of a tensor [8]) and Tensor Train (TT) formats [5], which need $O(dnr + r^d)$ and $O(dnr + (d-2)r^3)$ parameters, respectively, vs. $O(n^d)$ entries of the full tensor (here d is the tensor dimension, n is the number of entries in each dimension, and r is the Tucker/TT ranks used in compression). Canonical Polyadic Decomposition [9, 10](CPD), related to the rank of the tensor [8], also offers a good compression, however, computation of the tensor rank is an NP-hard problem [11], and ill-conditioned decompositions and ill-posed optimization problems often remain unsolved [12].

Another problem is that the existing datasets are formed by directly observable quantities, while the underlying processes (features or variables) usually remain unobserved, hidden, or latent [13]. This necessitates the ability to identify and extract *explainable* latent features needed to identify essential signatures that are manifestation of the processes and causalities hidden in large high-dimensional datasets. Imposing various constraints on the factors, reflecting available prior information, usually helps to mitigate this problem.

Many types of real-world data (e.g., density, energy, spectral power, population, pixels, probabilities, frequencies of appearance, etc.) are naturally nonnegative and the extracted features will lose their meaning if the nonnegativity is not preserved. Tensor factorizations with nonnegative constraint extract nonnegative latent features formed by only positive combinations, which favors parts based sparse representation where extracted features are parts of the original data [14]. Importantly, because the extracted features are parts of the original data they are easy to understand and interpret which makes the nonnegative factorization invaluable for scientific applications [15]. Classical tensor decompositions corresponding to nonnegative tensor ranks are nonnegative Canonical Polyadic Decomposition (nnCPD) [15] and nonnegaive Tucker Decomposition (nnTD) [15]. In Tucker, the minimum dimensions of the core tensor are often called *multirank* and the concept of *nonnegative multirank* in nnTD is introduced in Section 3.

In addition to nonnegativity, various other constraints on the decomposition are often needed to take into account the available domain knowledge and subject matter expertise and extract explainable and meaningful latent features. The canonical decomposition with linear constraints (CANDELINC) [16] is one of these decompositions. A preprocessing step in CANDELINC is Tucker compression, which often leads to a larger residual error but also to interpretable latent features [17].

In this work, we derive formulation of nonnegative CANDELINC (nnCANDELINC). This is accomplished via nnTDs we refer to as *minimal nonnegative Tucker Decompositions*: A minimal TD/nnTD is where the Tucker core has the smallest shape possible (Definitions 2.8, 3.10). In Section 2, we discuss the well-known fact that for real valued tensors minimal Tuckers always exist [8], and preserve the rank of the original tensor to the Tucker core (Theorem 2.12). We then relate the CPD to the minimal Tucker, which leads to CANDELINC (Theorem 2.13). The previous Theorem guarantees CANDELINC will successfully find a rank factorization of the tensor. The nonnegative counterpart to CANDELINC faces greater challenges however. In Section 3, we discuss nnCPD with rank deficiency (aka PARALIND for real valued tensors [18, 19]), and its relation to the minimal nnTD. We show that a minimal nnTD need not exist (Example 3.12) and even if it does exist, it need not preserve the rank to the core (Example 3.13). However under some mild conditions, some minimal nnTD will preserve the rank (Theorem 3.15). Unfortunately, these conditions do not guarantee that every minimal nnTD will preserve the rank (Example 3.16). This naturally leads to the discussion of when the nonnegative rank is preserved (Theorem 3.17) and when we can overcome the challenges just discussed. We therefore loosen our requirement on the shape of the Tucker core. This leads us naturally to the definition of a canonical nnTD (Definition 3.19). We show that every nonnegative tensor has a canonical nnTD which preserves the rank to the core (Theorem 3.20). Finally, in Section 4, we perform numerical experiments with nnCANDELINC on synthetic and real-world datasets. We consider two different algorithms for nnCANDELINC: (i) Performing, first nnTD compression, and then nnCPD on the core, and (ii) First nnCPD, and then reconstruction of the linear dependence of the extracted factors by Nonnegative Matrix Factorization (NMF). We also investigate the effect of choosing the nonnegative canonical vs. nonnegative minimal multirank.

2. Decompositions of Real Valued Tensors

In this section, we review some of the basics of real-valued tensors decompositions. For notational simplicity, we consider only 3-way tensors, although the analysis is valid for d-way tensors. A detailed presentation of the basic results can be found in [15, 8, 20, 3]. We will begin with a few formal definitions on tensors and tensor decompositions. This will include the concept of a minimal subspace, which will motivate our definition of a minimal Tucker decomposition. We will then go on to provide some results on rank perseverance to the core of a minimal Tucker decomposition. Precise notation for $n-$mode multiplication and unfolding used throughout the text can be found in the Appendix.

Definition 2.1. *For vectors $a^{(1)} \in \mathbb{R}^{N_1}$, $a^{(2)} \in \mathbb{R}^{N_2}$, $a^{(3)} \in \mathbb{R}^{N_3}$, the tensor product is the 3-way tensor $a^{(1)} \otimes a^{(2)} \otimes a^{(3)}$ given by*

$$\left(a^{(1)} \otimes a^{(2)} \otimes a^{(3)}\right)_{i,j,k} = a_i^{(1)} a_j^{(2)} a_k^{(3)}.$$

The tensor $a^{(1)} \otimes a^{(2)} \otimes a^{(3)}$ is referred to as a rank-1*, elementary, or decomposable tensor. For U_i subspace of $\mathbb{R}^{N_i}$, the* tensor product space *$U_1 \otimes U_2 \otimes U_3$ consists of all linear combinations of elementary tensors where $a^{(i)} \in U_i$.*

The tensor product space $\mathbb{R}^{N_1} \otimes \mathbb{R}^{N_2} \otimes \mathbb{R}^{N_3}$ is isomorphic to the linear space of 3-way arrays $\mathbb{R}^{N_1 \times N_2 \times N_3}$. Thus for ease of notation, we will often write $\mathcal{X} \in \mathbb{R}^{N_1 \times N_2 \times N_3}$ for a real 3-way tensor of dimension $N_1 \times N_2 \times N_3$, with components $\mathcal{X} = (\mathcal{X}_{i,j,k})$, for i, j, k ranging from 1 to N_1, N_2, and N_3, respectively. Every $\mathcal{X} \in \mathbb{R}^{N_1} \otimes \mathbb{R}^{N_2} \otimes \mathbb{R}^{N_3}$ can be written as $\mathcal{X} = \sum_{i,j,k} \mathcal{X}_{i,j,k} e_i^{(1)} \otimes e_j^{(2)} \otimes_k^{(3)}$, where, $\{e_i^{(1)}\}$, $\{e_j^{(2)}\}$, and $\{e_k^{(3)}\}$ are the canonical basis vectors of $\mathbb{R}^{N_1}$, $\mathbb{R}^{N_2}$, and $\mathbb{R}^{N_3}$, respectively. However, every tensor $\mathcal{X} \in \mathbb{R}^{N_1 \times N_2 \times N_3}$ can be decomposed in many different ways. And perhaps most significant is the decomposition as a weighted sum of rank-1 tensors:

Definition 2.2. *For every tensor $\mathcal{X} \in \mathbb{R}^{N_1 \times N_2 \times N_3}$, there exists a sufficiently large positive integer r such that $\mathcal{X}$ may be written as*

$$\mathcal{X} = \sum_{n=1}^{r} \lambda_n a_n^{(1)} \otimes a_n^{(2)} \otimes a_n^{(3)}, \tag{1}$$

where $\lambda_n \in \mathbb{R}$ and $a^{(i)} \in \mathbb{R}^{N_i}$ are unit vectors. Such a decomposition is a polyadic decomposition*. The* rank *of a tensor is defined as the smallest integer number r of rank-1 terms for which a polyadic decomposition exists, or*

$$\mathit{rank}(\mathcal{X}) = \min\left\{r \mid \mathcal{X} = \sum_{n=1}^{r} \lambda_n a_n^{(1)} \otimes a_n^{(2)} \otimes a_n^{(3)}, \lambda_n \in \mathbb{R},\, a_n^{(i)} \in \mathbb{R}^{N_i},\, i = 1, 2, 3\right\}. \tag{2}$$

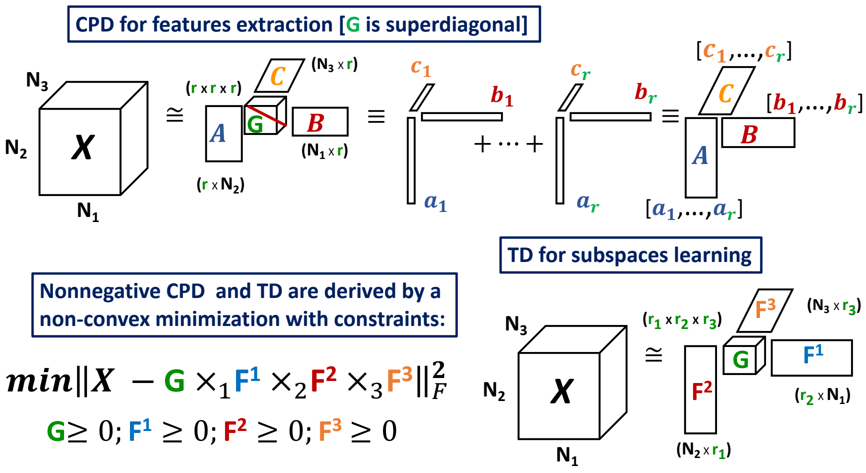


Figure 1: Two classical tensor decompositions: A) Canonical Polyadic Decomposition (CPD) of a 3-dimensional tensor $\mathcal{X}$ of size $N_1 \times N_2 \times N_3$ into a superdiagonal core tensor $\mathcal{G} \equiv \mathcal{D}$ of size $r \times r \times r$ and three matrix factors, A, B, and C. B) Tucker Decomposition (TD) of a 3-dimensional tensor $\mathcal{X}$ into a dense core tensor $\mathcal{G}$ of size $r_1 \times r_2 \times r_3$ and three matrix factors, $F^{(1)}$, $F^{(2)}$, and $F^{(3)}$.

A corresponding decomposition is called a Canonical Polyadic Decomposition (CPD) of $\mathcal{X}$.

Collecting the vectors $a_n^{(i)}$ into factor matrices $A^{(i)} = \left[\, a_1^{(i)} \mid \dots \mid a_r^{(i)} \,\right]$ and the coefficients λ_n into a superdiagonal tensor $\mathcal{D}$ allows us to represent CPD as the product of a superdiagonal tensor $\mathcal{D}$ and factor matrices, or

$$\mathcal{X} = \mathcal{D} \times_1 A^{(1)} \times_2 A^{(2)} \times_3 A^{(3)}, \tag{3}$$

as seen in Figure 1, panel A. Here the $n-$mode multiplication, $\times_n$ is defined in the Appendix (Definition Appendix A.1). In general, $\mathcal{X}$ does not require the full ambient space $\mathbb{R}^{N_1} \otimes \mathbb{R}^{N_2} \otimes \mathbb{R}^{N_3}$ to represent it. Indeed, $\mathcal{X}$ can be contained in the tensor product of subspaces $U_1 \otimes U_2 \otimes U_3$ where U_i is a subspace of $\mathbb{R}^{N_i}$. This is the concept behind a Tucker Decomposition:

Definition 2.3. *The* Tucker Decomposition *(TD) is a weighted tensor product decomposition of the form*

$$\mathcal{X} = \sum_{n_1,n_2,n_3=1}^{r_1,r_2,r_3} \mathcal{G}_{n_1,n_2,n_3} f_{n_1}^{(1)} \otimes f_{n_2}^{(2)} \otimes f_{n_3}^{(3)}, \tag{4}$$

where the vectors $f^{(i)} \in \mathbb{R}^{N_i}$, for $i = 1,2,3$, and the core tensor $\mathcal{G} \in \mathbb{R}^{r_1} \otimes \mathbb{R}^{r_2} \otimes \mathbb{R}^{r_3}$.

Tucker decomposition factorizes tensor $\mathcal{X}$ into the product of a tensor core $\mathcal{G}$ and three factor matrices $F^{(i)} = \left[f_1^{(i)} | \dots | f_{r_i}^{(i)} \right] \in \mathbb{R}^{N_i \times r_i}$, for $i = 1,2,3$, as seen in Figure 1 panel B. Similarly to (3), we can reformulate (4) as

$$\mathcal{X} = \mathcal{G} \times_1 F^{(1)} \times_2 F^{(2)} \times_3 F^{(3)} \,. \tag{5}$$

For a tensor $\mathcal{X} \in \mathbb{R}^{N_1} \times \mathbb{R}^{N_2} \times \mathbb{R}^{N_3}$, the matrix factors $F^{(i)}$ in a Tucker decomposition are associated with such subspaces of $\mathbb{R}^{N_i}$. Given a matrix $F \in \mathbb{R}^{N \times r}$ we let $\mathsf{col}(F)$ denote the column space of F. Then the following is a direct consequence of Equation 4:

Proposition 2.4. *Given three matrices $F^{(1)}$, $F^{(2)}$, and $F^{(3)}$, a tensor $\mathcal{X}$ admits the Tucker decomposition $\mathcal{X} = \mathcal{G} \times_1 F^{(1)} \times_2 F^{(2)} \times_3 F^{(3)}$ if and only if $\mathcal{X} \in \mathsf{col}(F^{(1)}) \otimes \mathsf{col}(F^{(2)}) \otimes \mathsf{col}(F^{(3)})$.*

Every Tucker decomposition of a three-way tensor is linked to three integer numbers, namely, r_1, r_2, and r_3, from $\mathcal{G} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$. The immediate question is: what are the permissible or minimal values of r_1, r_2, and r_3 such that there exists a Tucker decomposition with a core tensor of these dimensions? The smallest such values would describe the maximal permissible lossless compression within the shape of the tensor. This information is encoded in the concept of the *minimal subspaces* and *minimal tensor multirank*[8].

Definition 2.5. *Given a tensor $\mathcal{X} \in \mathbb{R}^{N_1} \otimes \mathbb{R}^{N_2} \otimes \mathbb{R}^{N_3}$, the* minimal subspaces associated with *$\mathcal{X}$ are subspaces $U_i^{\min} \subset \mathbb{R}^{N_i}$ such that $\mathcal{X} \in U_1^{\min} \otimes U_2^{\min} \otimes U_3^{\min}$ and if $\mathcal{X} \in U_1 \otimes U_2 \otimes U_3$ then $U_i^{\min} \subset U_i$.*

We remark that minimal subspaces always exist and are unique. Indeed, one can show that

$$(U_1 \otimes U_2 \otimes U_3) \bigcap (U_1' \otimes U_2' \otimes U_3') = \bigotimes_{i=1}^{3} U_i \cap U_i' \tag{6}$$

for any collection of subspaces U_i, $U_i' \subset \mathbb{R}^{N_i}$ [8]. It then follows that

$$U_1^{\min} \otimes U_2^{\min} \otimes U_3^{\min} = \bigcap \{U_1 \otimes U_2 \otimes U_3 : \mathcal{X} \in U_1 \otimes U_2 \otimes U_3\}.$$

Eq. (6) also shows that the minimal subspaces can be found coordinatewise rather than simultaneously. Hence, if $U_i^{\min}$ are the minimal subspaces found such that

$$\mathcal{X} \in U_1^{\min} \otimes \mathbb{R}^{N_2} \otimes \mathbb{R}^{N_3}; \quad \mathcal{X} \in \mathbb{R}^{N_1} \otimes U_2^{\min} \otimes \mathbb{R}^{N_3}; \quad \mathcal{X} \in \mathbb{R}^{N_1} \otimes \mathbb{R}^{N_2} \otimes U_3^{\min},$$

then by Eq. (6), we have $\mathcal{X} \in U_1^{\min} \otimes U_2^{\min} \otimes U_3^{\min}$. Associated with the minimal subspaces of $\mathcal{X}$ is the concept of the *i-th minimal multirank* of $\mathcal{X}$.

Definition 2.6. *The i-*th minimal multirank *of a tensor $\mathcal{X}$, denoted by $\mu rank_i(\mathcal{X})$, is the dimension of the i-th minimal subspace $U_i^{\min}$. The* minimal multilinear rank *of $\mathcal{X}$ is the triple of dimensions*

$$\mu rank(\mathcal{X}) = \big(\mu rank_1(\mathcal{X}),\, \mu rank_2(\mathcal{X}),\, \mu rank_3(\mathcal{X})\big).$$

We note that the i-th minimal multirank of $\mathcal{X}$ does not depend on the j-th tensor coordinate for $j \neq i$. Formally, the first minimal multirank of $\mathcal{X}$ is given by

$$\mu\mathrm{rank}_1(\mathcal{X}) = \min\big\{ \dim(U_1) \mid \mathcal{X} \in U_1 \otimes \mathbb{R}^{N_2} \otimes \mathbb{R}^{N_3}, U_1 \subset \mathbb{R}^{N_1} \big\}$$

with analogous definitions for the second and the third minimal multiranks. For any Tucker decomposition $\mathcal{X} = \mathcal{G} \times_1 F^{(1)} \times_2 F^{(2)} \times_3 F^{(3)}$, it holds that $\mathrm{rank}(F^{(i)}) \geq \mu\mathrm{rank}_i(\mathcal{X})$, $i = 1, 2, 3$, since by Proposition 2.4 the span of the columns of matrix factor $F^{(i)}$ must contain the corresponding minimal subspace $U_i^{\min}$, i.e., $U_i^{\min} \subset \mathrm{span}(F^{(i)})$. We record the following well known connection between the minimal multirank of $\mathcal{X}$ and its unfoldings (for definition of unfolding see Appendix, Definition Appendix A.3).

Proposition 2.7. *Given a tensor $\mathcal{X}$, $U_i^{\min} = \mathit{col}(\mathrm{unfold}_i(\mathcal{X}))$ and $\mu rank_i(\mathcal{X}) = rank(\mathrm{unfold}_i(\mathcal{X}))$ [8].*

In general, a Tucker decomposition does not satisfy the identity $\mathrm{rank}(F^{(i)}) = \mu\mathrm{rank}_i(\mathcal{X})$. This fact motivates us to introduce the notion of *minimal* Tucker Decomposition (*minimal* TDs) in the next definition, which is a Tucker decomposition with core dimensions corresponding to the *minimal multirank*.

Definition 2.8. *Consider tensor $\mathcal{X} \in \mathbb{R}^{N_1 \times N_2 \times N_3}$. We say that the Tucker decomposition $\mathcal{X} = \mathcal{G} \times_1 F^{(1)} \times_2 F^{(2)} \times_3 F^{(3)}$ is* minimal *if the dimensions of the core tensor $\mathcal{G}$ are equal to the minimal multiranks, i.e., $\mathcal{G} \in \mathbb{R}^{\mu rank_1(\mathcal{X}) \times \mu rank_2(\mathcal{X}) \times \mu rank_3(\mathcal{X})}$ and $F^{(i)} \in \mathbb{R}^{N_i \times \mu rank_i(\mathcal{X})}$, $i = 1, 2, 3$.*

In a minimal TD, $F^{(i)} \in \mathbb{R}^{N_i \times \mu\mathrm{rank}_i(\mathcal{X})}$ implies that $\mathrm{rank}(F^{(i)}) \leq \mu\mathrm{rank}_i(\mathcal{X})$. However by the discussion above, $\mathrm{rank}(F^{(i)}) \geq \mu\mathrm{rank}_i(\mathcal{X})$, so that $\mathrm{rank}(F^{(i)}) = \mu\mathrm{rank}_i(\mathcal{X})$. This is rather different than the case of the loading matrices in a CPD, where rank deficiency can occur. The following simple example demonstrates that a CPD need not be minimal TD.

Example 2.9. *Let $\mathcal{X} \in \mathbb{R}^{2,2,2}$ be the rank 2 tensor*

$$\mathcal{X} = \left[\begin{array}{cc|cc} 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \end{array}\right].$$

It is not hard to see that $\mu rank_1(\mathcal{X}) = 1$, while $\mu rank_2(\mathcal{X}) = \mu rank_3(\mathcal{X}) = 2$. Hence any minimal TD will satisfy $\mathcal{G} \in \mathbb{R}^{1,2,2}$. However since $\mathcal{X}$ is rank 2, the rank decomposition will have the shape $\mathcal{D} \in \mathbb{R}^{2,2,2}$. Thus, a CPD of $\mathcal{X}$ does not need to be a minimal Tucker decomposition.

2.1. Real Rank Preservation to Minimal Tucker Core

Given a Tucker decomposition $\mathcal{X} = \mathcal{G} \times_1 F^{(1)} \times_2 F^{(2)} \times_3 F^{(3)}$, we say that the rank is preserved to the core if $\mathrm{rank}(\mathcal{X}) = \mathrm{rank}(\mathcal{G})$. Not every TD needs to preserve the rank to the Tucker core, as the following simple example illustrates:

Example 2.10. *Let $\mathcal{X} \in \mathbb{R}^{2,2,2}$ be the rank 1 tensor*

$$\mathcal{X} = \left[\begin{array}{cc|cc} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{array}\right].$$

Then $\mathcal{X}$ can be decomposed using a rank 2 core as

$$\mathcal{X} = \left[\begin{array}{cc|cc} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{array}\right] \times_1 \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

One always has the following simple rank and minimal Tucker relationships which we record in the following lemma:

Lemma 2.11. *For any Tucker decomposition $\mathcal{X} = \mathcal{G} \times_1 F^{(1)} \times_2 F^{(2)} \times_3 F^{(3)}$, $rank(\mathcal{X}) \leq rank(\mathcal{G})$. Moreover, a minimal Tucker decomposition always exists for real factorizations.*

Proof. Suppose $\mathcal{X} = \mathcal{G} \times_1 F^{(1)} \times_2 F^{(2)} \times_3 F^{(3)}$ is a Tucker decomposition of $\mathcal{X}$. Then consider a CPD of $\mathcal{G}$ given by

$$\mathcal{G} = \mathcal{D}_{\mathcal{G}} \times_1 A_{\mathcal{G}}^{(1)} \times_2 A_{\mathcal{G}}^{(2)} \times_3 A_{\mathcal{G}}^{(3)}. \tag{7}$$

Then, substituting the CPD of $\mathcal{G}$ into the Tucker decomposition of $\mathcal{X}$ yields

$$\begin{aligned}\mathcal{X} &= \mathcal{G} \times_1 F^{(1)} \times_2 F^{(2)} \times_3 F^{(3)} \\ &= \mathcal{D}_{\mathcal{G}} \times_1 A_{\mathcal{G}}^{(1)} \times_2 A_{\mathcal{G}}^{(2)} \times_3 A_{\mathcal{G}}^{(3)} \times_1 F^{(1)} \times_2 F^{(2)} \times_3 F^{(3)} \\ &= \mathcal{D}_{\mathcal{G}} \times_1 \big(F^{(1)} A_{\mathcal{G}}^{(1)}\big) \times_2 \big(F^{(2)} A_{\mathcal{G}}^{(2)}\big) \times_3 \big(F^{(3)} A_{\mathcal{G}}^{(3)}\big).\end{aligned}$$

The last right-hand side is a polyadic decomposition of $\mathcal{X}$ with $\operatorname{rank}(\mathcal{G})$ summands proving that $\operatorname{rank}(\mathcal{X}) \le \operatorname{rank}(\mathcal{G})$.

It is also not hard to see that a minimal Tucker always exists for real factorizations. Let $F^{(i)}$ be basis matrices for $U_i^{\min}$. Then by definition,

$$\mathcal{X} \in U_1^{\min} \otimes U_2^{\min} \otimes U_3^{\min} = \mathsf{col}(F^{(1)}) \otimes \mathsf{col}(F^{(2)}) \otimes \mathsf{col}(F^{(3)}).$$

By Proposition 2.4, there exists a $\mathcal{G}$ such that $\mathcal{X} = \mathcal{G} \times_1 F^{(1)} \times_2 F^{(2)} \times_3 F^{(3)}$. As $F^{(i)} \in \mathbb{R}^{\mu\mathrm{rank}_i(\mathcal{X}) \times N_i}$, this is a minimal TD. □

Lemma 2.11 demonstrates that a minimal Tucker decomposition can be constructed by choosing basis matricies for $U_i^{\min}$. However, this is how all real minimal Tucker decompositions are formed. Indeed if $\mathcal{X} = \mathcal{G} \times_1 F^{(1)} \times_2 F^{(2)} \times_3 F^{(3)}$ is a minimal Tucker, then by Proposition 2.4, $\mathcal{X} \in \mathsf{col}(F^{(1)}) \otimes \mathsf{col}(F^{(2)}) \otimes \mathsf{col}(F^{(3)})$. Since $\operatorname{rank}(F^{(i)}) = \mu\mathrm{rank}_i(\mathcal{X})$ and $F^{(i)} \in \mathbb{R}^{\mu\mathrm{rank}_i(\mathcal{X}) \times N_i}$, this implies $F^{(i)}$ are basis matrices for $U_i^{\min}$.

While a minimal Tucker decomposition always exists, what we are interested in is the preservation of the rank of $\mathcal{X}$ to the minimal core $\mathcal{G}$. The next theorem establishes that minimal Tucker decompositions do always preserve the rank to the core. While we believe this result is known, it does not appear to be written explicitly down in the literature. Hence, we record it alongside its proof:

Theorem 2.12. *Given a real tensor $\mathcal{X}$, any minimal Tucker decomposition $\mathcal{X} = \mathcal{G} \times_1 F^{(1)} \times_2 F^{(2)} \times_3 F^{(3)}$ satisfies rank($\mathcal{X}$) = rank($\mathcal{G}$).*

Proof. The proof is constructive. Let $U_i^{\min}$ be the i-th minimal subspace of $\mathcal{X}$ according to the definition of the minimal multirank. For every $i = 1, 2, 3$, we choose the basis matrix $F^{(i)} \in \mathbb{R}^{\mu\mathrm{rank}_i(\mathcal{X}) \times N_i}$ such that $\operatorname{span}(F^{(i)}) = U_i^{\min}$. By construction, $F^{(i)}$ is a full column rank matrix, its rank being equal to $\mu\mathrm{rank}_i(\mathcal{X})$. Hence, $F^{(i)}$ admits the (left) Moore-Penrose pseudo-inverse $F^{(i)\dagger} \in \mathbb{R}^{N_i \times \mu\mathrm{rank}_i(\mathcal{X})}$, so that $F^{(i)\dagger} F^{(i)} = I$. Moreover, the matrix $P^{(i)} := F^{(i)} F^{(i)\dagger}$ is the projection onto the column space of $F^{(i)}$ (which is $U_i^{\min}$). Let

$$\mathcal{G} = \mathcal{X} \times_1 F^{(1)\dagger} \times_2 F^{(2)\dagger} \times_3 F^{(3)\dagger}. \tag{8}$$

We show that $\mathcal{X} = \mathcal{G} \times_1 F^{(1)} \times_2 F^{(2)} \times_3 F^{(3)}$. Indeed, by redistributing we verify that

$$\mathcal{G} \times_1 F^{(1)} \times_2 F^{(2)} \times_3 F^{(3)} = \mathcal{X} \times_1 P^{(1)} \times_2 P^{(2)} \times_3 P^{(3)}.$$

Hence, it suffices to show that $\mathcal{X} \times_i P^{(i)} = \mathcal{X}$ for each i. This happens if and only if $\mathsf{col}(\mathrm{unfold}_i(\mathcal{X})) \subset \mathsf{col}(P^{(i)}) = \mathsf{col}(F^{(i)}) = U_i^{\min}$. By Proposition 2.7, these are equal. Therefore, $\mathcal{G} \times_1 F^{(1)} \times_2 F^{(2)} \times_3 F^{(3)} = \mathcal{X}$.

To prove the identity $\operatorname{rank}(\mathcal{X}) = \operatorname{rank}(\mathcal{G})$, it suffices to show that $\operatorname{rank}(\mathcal{X}) \ge \operatorname{rank}(\mathcal{G})$. Let

$$\mathcal{X} = \mathcal{D}_{\mathcal{X}} \times_1 A_{\mathcal{X}}^{(1)} \times_2 A_{\mathcal{X}}^{(2)} \times_3 A_{\mathcal{X}}^{(3)} \tag{9}$$

be a CPD of $\mathcal{X}$ with matrix factors $A_{\mathcal{X}}^{(i)} \in \mathbb{R}^{N_i \times r}$ and superdiagonal core tensor $\mathcal{D}_{\mathcal{X}} \in \mathbb{R}^{r \times r \times r}$, where $r = \text{rank}(\mathcal{X})$. Then, starting from (8) and using (9), a straightforward calculation yields:

$$\begin{aligned}
\mathcal{G} &= \mathcal{X} \times_1 F^{(1)^\dagger} \times_2 F^{(2)^\dagger} \times_3 F^{(3)^\dagger} \\
&= \mathcal{D}_{\mathcal{X}} \times_1 A_{\mathcal{X}}^{(1)} \times_2 A_{\mathcal{X}}^{(2)} \times_3 A_{\mathcal{X}}^{(3)} \times_1 \left(F^{(1)^\dagger}\right) \times_2 \left(F^{(2)^\dagger}\right) \times_3 \left(F^{(3)^\dagger}\right) \\
&= \mathcal{D}_{\mathcal{X}} \times_1 \left(F^{(1)^\dagger} A_{\mathcal{X}}^{(1)}\right) \times_2 \left(F^{(2)^\dagger} A_{\mathcal{X}}^{(2)}\right) \times_3 \left(F^{(3)^\dagger} A_{\mathcal{X}}^{(3)}\right).
\end{aligned}$$

The last right-hand side is a polyadic decomposition of $\mathcal{G}$ with $\text{rank}(\mathcal{X})$ summands, proving that $\text{rank}(\mathcal{G}) \leq \text{rank}(\mathcal{X})$. □

2.2. Rank Deficiency in CPD Factors and the CANDELINC Solution

In this subsection, we discuss the challenges that arise from rank deficient factors in a CPD and how the CANDELINC method [18, 3] can provide a suitable decomposition. Theorem 2.12 establishes that one can always construct minimal Tucker decompositions that will preserve the rank. The following theorem relates the uniqueness of a CPD to the minimal TDs:

Theorem 2.13 (Ranks of CPD factors related to minimal TD)**.** *Let $\mathcal{X} = \mathcal{D} \times_1 A^{(1)} \times_2 A^{(2)} \times_3 A^{(3)}$ be a CPD of $\mathcal{X}$, then $\mu rank_i(\mathcal{X}) \leq rank(A^{(i)})$. Furthermore, if the CPD is unique, then $\mu rank_i(\mathcal{X}) = rank(A^{(i)})$. In this case, $\mathsf{col}(A^{(i)}) = U_i^{\min}$.*

Proof. Let $\mathcal{X} = \mathcal{D} \times_1 A^{(1)} \times_2 A^{(2)} \times_3 A^{(3)}$ be a CPD of $\mathcal{X}$. Since $\mathcal{D} \times_1 A^{(1)} \times_2 A^{(2)} \times_3 A^{(3)}$ is a Tucker decomposition, by definition of minimal multirank we have $\mu\text{rank}_i(\mathcal{X}) \leq \text{rank}(A^{(i)})$. Now suppose that $\mathcal{X}$ has a unique CPD, and consider a minimal TD of the form $\mathcal{X} = \mathcal{G} \times_1 F^{(1)} \times_2 F^{(2)} \times_3 F^{(3)}$. We recall that $\text{rank}(F^{(i)}) = \mu\text{rank}_i(\mathcal{X})$ by definition. On its turn, the Tucker core admits the CPD $\mathcal{G} = \mathcal{D}_{\mathcal{G}} \times_1 B^{(1)} \times_2 B^{(2)} \times_3 B^{(3)}$. Substituting the CPD of $\mathcal{G}$ in the TD of $\mathcal{X}$, we obtain the alternative CPD $\mathcal{X} = \mathcal{D}_{\mathcal{G}} \times_1 F^{(1)}B^{(1)} \times_2 F^{(2)}B^{(2)} \times_3 F^{(3)}B^{(3)}$. Since we assume that the CPD is unique, with appropriate scalings and permutations, which are rank-preserving operations, we obtain that $A^{(i)} = F^{(i)}B^{(i)}$. Therefore, $\text{rank}(A^{(i)}) \leq \text{rank}(F^{(i)}) = \mu\text{rank}_i(\mathcal{X})$. It follows from Proposition 2.7 and rank arguments that $U_i^{\min} = \mathsf{col}(\text{unfold}_i(\mathcal{X})) = \mathsf{col}(A^{(i)})$. □

Theorem 2.13 suggests why a direct CPD computation can be algorithmically problematic. Let $\mathcal{X} = \mathcal{D} \times_1 A^{(1)} \times_2 A^{(2)} \times_3 A^{(3)}$ be the unique CPD of a rank r tensor with $\mu\text{rank}_i(\mathcal{X}) = r_i$. If $r_i < r$, as is the case with probability 1 for many shaped tensors [21, 22], then $A^{(i)}$ is a rank deficient matrix by Theorem 2.13. Indeed, $A^{(i)}$ is an $(N_i \times r)$-sized matrix with only r_i linearly independent columns. Algorithmically, finding rank deficient matrices without an explicit rank constraint for tensors of the form $\mathcal{X} = \widetilde{\mathcal{X}} + \mathcal{E}$ is challenging, as the rank deficient subspaces of the factors of $\widetilde{\mathcal{X}}$ can always be expanded to accommodate some of the noise, $\mathcal{E}$.

The proof of Theorem 2 suggests a more suitable method for computing the CPD of $\mathcal{X}$. First compute a minimal TD of $\mathcal{X}$ (which will preserve the rank); then compute a CPD of the TD core (which will lack rank deficiency); and finally substitute the CPD of the TD core into the TD and obtain a CPD of the original tensor. Bro *et al.* followed this strategy in their construction of the PARALIND models, cf. [18], and Carroll *et al.* followed this strategy in their construction of the CANDELINC models, cf. [16]. Formally, if $\mathcal{X} = \mathcal{G} \times_1 F^{(1)} \times_2 F^{(2)} \times_3 F^{(3)}$ is a minimal TD, and $\mathcal{G} = \mathcal{D}_{\mathcal{G}} \times_1 A_{\mathcal{G}}^{(1)} \times_2 A_{\mathcal{G}}^{(2)} \times_3 A_{\mathcal{G}}^{(3)}$ is the CPD of the Tucker core, then each factor $A_{\mathcal{G}}^{(i)}$ is a full column rank matrix, avoiding the algorithmic problems previously discussed. A simple substitution yields a CPD of $\mathcal{X}$ where each loading matrix is rank factored, i.e., $\mathcal{X} = \mathcal{D}_{\mathcal{G}} \times_1 \left(F^{(1)} A_{\mathcal{G}}^{(1)}\right) \times_2 \left(F^{(2)} A_{\mathcal{G}}^{(2)}\right) \times_3 \left(F^{(3)} A_{\mathcal{G}}^{(3)}\right)$, and we have explicitly the linear constraints of the CPD factors.

3. Nonnegative Decompositions of Nonnegative Tensors

Following [23, 24], we now present the nonnegative counterparts to the discussion for real tensors above. This theory necessarily depends on some basic knowledge of nonnegative matrix factorizations. For the unfamiliar reader, we have provided some background information in the Appendix. Throughout, we let $\mathbb{R}_+$ denote the nonnegative real numbers. All of the basic definitions from real tensors will carry over to nonnegative with some appropriate adaptations. While real rank factorizations fundamentally rely on subspaces, nonnegative factorizations are concerned with the nonnegative analog of subspaces - polyhedral cones.

Definition 3.1. *A convex cone is a subset $C \subset \mathbb{R}_+^N$ that is closed under addition of vectors and $\mathbb{R}_+$ scalar multiplication. Given $W \subset \mathbb{R}_+^N$, the non-negative span of W defines a cone. A subset of the cone $W \subset C \subset \mathbb{R}_+^N$ is a generating set if its span is equal to C. The order of the cone $C \subset \mathbb{R}_+^N$, denoted $\mathcal{O}(C)$, is the size of a minimal generating set. A cone is polyhedral if $\mathcal{O}(C) < \infty$. Given a nonnegative matrix $W \in \mathbb{R}_+^{N,R}$, we define the cone of the matrix W to be*

$$\mathit{cone}(W) = \{Wh : h \in \mathbb{R}_+^R\} \subset \mathbb{R}_+^N.$$

Every polyhedral cone $C \subset \mathbb{R}_+^N$ is cone(W) for some nonnegative matrix $W \in \mathbb{R}_+^{N,R}$. Furthermore, every polyhedral cone can be equivalently described as the intersection of half spaces [25]. With the precise definition of cone, we can now define the analogous tensor product space of cones, and the associated nonnegative tensor decompositions.

Definition 3.2. *For vectors $a^{(1)} \in \mathbb{R}_+^{N_1}$, $a^{(2)} \in \mathbb{R}_+^{N_2}$, $a^{(3)} \in \mathbb{R}_+^{N_3}$, the tensor product is the 3-way tensor $a^{(1)} \otimes a^{(2)} \otimes a^{(3)}$ given by*

$$\left(a^{(1)} \otimes a^{(2)} \otimes a^{(3)}\right)_{i,j,k} = a_i^{(1)} a_j^{(2)} a_k^{(3)}.$$

The tensor $a^{(1)} \otimes a^{(2)} \otimes a^{(3)}$ is referred to as a nonnegative rank-1*, elementary, or decomposable tensor. For $\mathcal{C}_i$ a polyhedral cone of $\mathbb{R}_+^{N_i}$, the* tensor product space *$\mathcal{C}_1 \otimes \mathcal{C}_2 \otimes \mathcal{C}_3$ consists of all nonnegative linear combinations of elementary tensors where $a^{(i)} \in \mathcal{C}_i$.*

Analogous to the real case, every tensor $\mathcal{X} \in \mathbb{R}^{N_1 \times N_2 \times N_3}$ can be decomposed in different ways. The definitions of polyadic and Tucker decompositions for tensors will translate with the appropriate nonnegative adjustments.

Definition 3.3. *For every tensor $\mathcal{X} \in \mathbb{R}_+^{N_1 \times N_2 \times N_3}$, there exists a sufficiently large positive integer r such that $\mathcal{X}$ may be written as*

$$\mathcal{X} = \sum_{n=1}^{r} \lambda_n a_n^{(1)} \otimes a_n^{(2)} \otimes a_n^{(3)}, \tag{10}$$

where $\lambda_n \in \mathbb{R}_+$ and $a^{(i)} \in \mathbb{R}_+^{N_i}$ are unit vectors. Such a decomposition is a nonnegative polyadic decomposition*. The* nonnegative rank *of a tensor is defined as the smallest integer number r of rank-1 terms for which a polyadic decomposition exists, or*

$$\mathit{rank}(\mathcal{X}) = \min\left\{ r \;\middle|\; \mathcal{X} = \sum_{n=1}^{r} \lambda_n a_n^{(1)} \otimes a_n^{(2)} \otimes a_n^{(3)}, \lambda_n \in \mathbb{R}_+,\, a_n^{(i)} \in \mathbb{R}_+^{N_i},\, i = 1,2,3 \right\}. \tag{11}$$

A corresponding decomposition is called a nonnegative Canonical Polyadic Decomposition (nnCPD) *of $\mathcal{X}$. For brevity, if the nonnegative qualifier is clear from context we may omit it when discussing various nonnegative ranks.*

It is immediately clear that for tensors $\operatorname{rank}_+(\mathcal{X}) \geq \operatorname{rank}(\mathcal{X})$, as the nnCPD is also a polyadic decomposition. Analogous to the real case, $\mathcal{X}$ does not require the full ambient space of $\mathbb{R}_+^{N_1} \otimes \mathbb{R}_+^{N_2} \otimes \mathbb{R}_+^{N_3}$ to represent it. It is possible that $\mathcal{X}$ can be contained in the tensor product of cones $\mathcal{C}_1 \otimes \mathcal{C}_2 \otimes \mathcal{C}_3$ where $\mathcal{C}_i$ is a polyhedral cone of $\mathbb{R}_+^{N_i}$. This once again motivates the concept of a nonnegative Tucker decomposition:

Definition 3.4. *A* nonnegative Tucker decomposition *(nnTD) of a nonnegative tensor is a nonnegative weighted tensor product decomposition of the form,*

$$\mathcal{X} = \sum_{n_1,n_2,n_3=1}^{r_1,r_2,r_3} \mathcal{G}_{n_1,n_2,n_3} f_{n_1}^{(1)} \otimes f_{n_2}^{(2)} \otimes f_{n_3}^{(3)} \ , \tag{12}$$

where the vectors $f_{n_i}^{(i)} \in \mathbb{R}_+^{N_i}$, for $i = 1,2,3$, and the core tensor $\mathcal{G}_{n_1,n_2,n_3} \in \mathbb{R}_+^{r_1} \otimes \mathbb{R}_+^{r_2} \otimes \mathbb{R}_+^{r_3}$.

The factors of an nnTD are associated with nonnegative cones, and are inherently tied to the tensor belonging to the tensor product space of these cones:

Proposition 3.5. *Given nonnegative matrices $F^{(1)}$, $F^{(2)}$, and $F^{(3)}$, a tensor admits an nnTD: $\mathcal{X} = \mathcal{G} \times_1 F^{(1)} \times_2 F^{(2)} \times_3 F^{(3)}$ if and only if $\mathcal{X} \in \operatorname{cone}(F^{(1)}) \otimes \operatorname{cone}(F^{(2)}) \otimes \operatorname{cone}(F^{(3)})$.*

One important subtle difference between polyhedral cones and subspaces is that cone intersection does not commute with the tensor product. That is, if $\mathcal{C}_i, \mathcal{C}_i' \subset \mathbb{R}_+^{N_i}$ are cones for $i = 1,2,3$, then

$$(\mathcal{C}_1 \otimes \mathcal{C}_2 \otimes \mathcal{C}_3) \bigcap (\mathcal{C}_1' \otimes \mathcal{C}_2' \otimes \mathcal{C}_3') \neq \bigotimes_{i=1}^{3} \mathcal{C}_i \cap \mathcal{C}_i'.$$

Example 3.6 shows that we cannot simply take the "smallest" cones via intersection as we could with subspaces.

Example 3.6. *Consider a $3 \times 3 \times 2$ nonnegative tensor with the unfoldings,*

$$\operatorname{unfold}_1(\mathcal{X}) = \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 2 & 1 & 1 & 1 \\ 1 & 1 & 2 & 1 & 1 & 1 \end{array}\right],$$

$$\operatorname{unfold}_2(\mathcal{X}) = \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 2 & 2 & 1 & 1 & 1 \end{array}\right],$$

$$\operatorname{unfold}_3(\mathcal{X}) = \left[\begin{array}{ccc|ccc|ccc} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 \\ 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 \end{array}\right].$$

One can easily verify that:

$$\mathcal{X} \in \mathcal{C}_1 \otimes \mathbb{R}_+^3 \otimes \mathbb{R}_+^2 \quad \textit{where} \quad \mathcal{C}_1 = \operatorname{cone}(W^{(1)}) \quad \textit{and} \quad W^{(1)} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 1 \end{bmatrix},$$

$$\mathcal{X} \in \mathbb{R}_+^3 \otimes \mathcal{C}_2 \otimes \mathbb{R}_+^2 \quad \textit{where} \quad \mathcal{C}_2 = \operatorname{cone}(W^{(2)}) \quad \textit{and} \quad W^{(2)} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 1 \end{bmatrix},$$

$$\mathcal{X} \in \mathbb{R}_+^3 \otimes \mathbb{R}_+^3 \otimes \mathcal{C}_3 \quad \textit{where} \quad \mathcal{C}_3 = \operatorname{cone}(W^{(3)}) \quad \textit{and} \quad W^{(3)} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Recall that in a linear system $AX = B$, if A has full column rank then there exists a unique solution X. Consequently, by taking unfoldings, one finds that if $\mathcal{X} = \mathcal{G} \times_1 W^{(1)} \times_2 W^{(2)} \times_3 W^{(3)}$ and each $W^{(i)}$ has full column rank, then there is a unique solution for $\mathcal{G}$. Note in our example, each $W^{(i)}$ is full column rank. Therefore, there is a unique core $\mathcal{G}$ with the loading matrices $W^{(i)}$. One can show $\mathrm{unfold}_1(\mathcal{G}) = \left[\begin{array}{cc|cc} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & -1 \end{array}\right]$. *Since $\mathcal{G}$ is not nonnegative, by Proposition 3.5 conclude that $\mathcal{X} \notin \mathcal{C}_1 \otimes \mathcal{C}_2 \otimes \mathcal{C}_3$. However, if instead the cones corresponding to* $\bar{W}^{(1)} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$, $\bar{W}^{(2)} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$, *and* $\bar{W}^{(3)} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ *were chosen, then $\mathcal{X} \in \bar{C}^{(1)} \otimes \bar{C}^{(2)} \otimes \bar{C}^{(3)}$.*

Example 3.6 shows that we cannot take the intersection of cones to produce a "minimal" cone. Therefore, we make the following mode-wise definition:

Definition 3.7. *Given a nonnegative tensor $\mathcal{X} \in \mathbb{R}_+^{N_1} \otimes \mathbb{R}_+^{N_2} \otimes \mathbb{R}_+^{N_3}$, a* minimal 1-mode nonnegative cone*, denoted by $\mathcal{C}_1^{\min}$, is a cone such that $\mathcal{X} \in \mathcal{C}_1^{\min} \otimes \mathbb{R}_+^{N_2} \otimes \mathbb{R}_+^{N_3}$ and if $\mathcal{X} \in \mathcal{C}_1 \otimes \mathbb{R}_+^{N_2} \otimes \mathbb{R}_+^{N_3}$ for some cone $\mathcal{C}_1$, then $\mathcal{O}(\mathcal{C}_1^{\min}) \leq \mathcal{O}(\mathcal{C}_1)$ (we recall that $\mathcal{O}(\mathcal{C})$ is the minimum number of vectors generating $\mathcal{C}$). We define the* 2-mode *and* 3-mode *minimal cones analogously.*

Unlike minimal subspaces the minimal cones are generally not unique, and they are defined mode wise because different mode cones are not necessarily interchangeable. We define the *minimal nonnegative multirank* of a nonnegative tensor as the minimum number of extreme rays of minimal nonnegative cones along each axis.

Definition 3.8. *The* i-th minimal nonnegative multilinear rank *or* i-th minimal nonnegative multirank *of a tensor $\mathcal{X}$, denoted $\mu rank_{+,i}(\mathcal{X})$ is defined as $\mathcal{O}(\mathcal{C}_i^{\min})$. The* minimal nonnegative multilinear rank *of $\mathcal{X}$ is the triple of orders:*

$$\mu rank_+(\mathcal{X}) = (\mu rank_{+,1}(\mathcal{X}), \mu rank_{+,2}(\mathcal{X}), \mu rank_{+,3}(\mathcal{X})).$$

As before, we note that the i'th minimal nonnegative multilinear rank does not depend on the j'th tensor coordinate for $j \neq i$. Concretely, we can compute the first minimal nonnegative multirank as

$$\mu\mathrm{rank}_{+,1}(\mathcal{X}) = \min\left\{\mathcal{O}(\mathcal{C}_1) \mid \mathcal{X} \in \mathcal{C}_1 \otimes \mathbb{R}_+^{N_2} \otimes \mathbb{R}_+^{N_3}, \mathcal{C}_1 = \mathrm{cone}(W^{(1)}) \subset \mathbb{R}_+^{N_1}\right\}$$

and, similarly, for $\mu\mathrm{rank}_{+,2}(\mathcal{X})$ and $\mu\mathrm{rank}_{+,3}(\mathcal{X})$. As with the real case, it follows directly that for any nTD $\mathrm{rank}_+(F^{(i)}) \geq \mu\mathrm{rank}_{+,i}(\mathcal{X})$. Additionally, we have an analogous nonnegative statement to Proposition 2.7:

Proposition 3.9. *For any nonnegative tensor $\mathcal{X}$, $\mu rank_{+,i}(\mathcal{X}) = rank_+(\mathrm{unfold}_i(\mathcal{X}))$.*

Proof. Without loss of generality we prove this for $i = 1$ through proving the inequality in both directions. Let $\mu\mathrm{rank}_{+,1}(\mathcal{X}) = k$, then there exists a nonnegative cone $\mathcal{C}^{(1)}$ with k extreme rays such that $\mathcal{X} \in \mathcal{C}^{(1)} \otimes \mathbb{R}_+^{N_2} \otimes \mathbb{R}_+^{N_3}$. Thus $\mathcal{X}$ admits a decomposition of the form $\mathcal{X} = \mathcal{D}_{identity} \times_1 A^{(1)} \times_2 A^{(2)} \times_3 A^{(3)}$ where the columns of $A^{(i)}$ are contained by their respective cones, $\mathcal{C}^{(1)}, \mathbb{R}_+^{N_2}, \mathbb{R}_+^{N_3}$. Assemble the extreme rays of $\mathcal{C}^{(1)}$ into a matrix $W^{(1)} \in \mathbb{R}^{N_1 \times k}$ so that $\mathcal{C}^{(1)} = \mathrm{cone}(W^{(1)})$. Then $A^{(1)} = W^{(1)} H^{(1)}$ for some $H^{(1)} \geq 0$, and with substitution we have

$$\mathcal{X} = \mathcal{D}_{identity} \times_1 W^{(1)} H^{(1)} \times_2 A^{(2)} \times_3 A^{(3)}.$$

Through distributing and applying unfoldings we have

$$\text{unfold}_1(\mathcal{X}) = W^{(1)}\,\text{unfold}_1(\mathcal{D}_{identity} \times_1 H^{(1)} \times_2 A^{(2)} \times_3 A^{(3)}),$$

which proves $\text{rank}_+(\text{unfold}_1(\mathcal{X})) \leq \text{rank}_+(W^{(1)}) \leq k = \mu\text{rank}_{+,1}(\mathcal{X})$.

Let $\text{rank}_+(\text{unfold}_1(\mathcal{X})) = k$. Since $\mathcal{X}$ is nonnegative, $\text{unfold}_1(\mathcal{X})$ admits a nonnegative decomposition as $\text{unfold}_1(\mathcal{X}) = W^{(1)}H^{(1)}$. Since each column of $H^{(1)}$ is nonnegative and is associated with a fiber of the tensor, we write the decomposition $\mathcal{X} = \sum_{j=1}^{N_2}\sum_{k=1}^{N_3} W^{(1)}H_{:,j,k} \otimes e_j \otimes e_k$. This demonstrates that $\mathcal{X} \in \text{cone}(W^{(1)}) \otimes \mathbb{R}_+^{N_2} \otimes \mathbb{R}_+^{N_3}$, so $\mu\text{rank}_{+,i}(\mathcal{X}) \leq \text{rank}_+(\text{unfold}_i(\mathcal{X}))$. □

Propositions 2.7 and 3.9 highlight a key difference between the real and nonnegative TD. In the real case, one had that the minimal subspace was obtained via the unfolding. In the nonnegative case, the unfolding does not result in a minimal cone. From the definition of the unfolding, one has

$$\mathcal{X} = \sum_{j=1}^{N_2}\sum_{k=1}^{N_3} \text{unfold}_1(\mathcal{X})_{:,m(j,k)} \otimes e_j \otimes e_k,$$

where $m(j,k) = j + N_2(k-1)$, and e_j and e_k are the j-th and the k-th vector of the canonical basis of $\mathbb{R}_+^{N_2}$ and $\mathbb{R}_+^{N_3}$, respectively. Hence,

$$\mathcal{X} \in \text{cone}(\text{unfold}_1(\mathcal{X})) \otimes \mathbb{R}_+^{N_2} \otimes \mathbb{R}_+^{N_3}.$$

However, it may be the case that $\mathcal{O}(\text{cone}(\text{unfold}_1(\mathcal{X}))) > \mu\text{rank}_{1,+}(\mathcal{X})$. Indeed from Proposition 3.9, $\mu\text{rank}_{+,i}(\mathcal{X}) = \text{rank}_+(\text{unfold}_i(\mathcal{X}))$ and in general $\mathcal{O}(\text{cone}(A)) > \text{rank}_+(A)$ for many nonnegative matrices A since the nonnegative rank is equal to the order of the minimal cone that contains the data.

Just as in the real case, we are interested when the nonnegative TD has no degeneracy in the loading matrices $F^{(i)}$. When a tensor is simultaneously contained in the tensor product of minimal nonnegative cones, we call the corresponding nonnegative TD a *minimal nnTD*.

Definition 3.10. *An nnTD: $\mathcal{X} = \mathcal{G}\times_1 F^{(1)}\times_2 F^{(2)}\times_3 F^{(3)}$ of a tensor $\mathcal{X}$ is a* minimal nnTD *whenever the core dimensions are equal to the minimal nonnegative multiranks, i.e., when $F^{(i)} \in \mathbb{R}^{N_i\times \mu rank_{+,i}(\mathcal{X})}$ and $\mathcal{G} \in \mathbb{R}_+^{\mu rank_{+,1}(\mathcal{X})\times \mu rank_{+,2}(\mathcal{X})\times \mu rank_{+,3}(\mathcal{X})}$.*

We note that since $F^{(i)} \in \mathbb{R}^{N_i\times \mu\text{rank}_{+,i}(\mathcal{X})}$, and $\text{rank}_+(F^{(i)}) \geq \mu\text{rank}_{+,i}(\mathcal{X})$, it follows that minimal nnTDs satisfy $\text{rank}_+(F^{(i)}) = \mu\text{rank}_{+,i}(\mathcal{X})$. While Example 3.6 showed that one cannot take intersections to achieve minimal cones, the next simple result connects the equivalence of simultaneous minimal cones and a minimal nnTD:

Proposition 3.11. *A nonnegative tensor $\mathcal{X}$ has a minimal nnTD if and only if there exists minimal cones $\mathcal{C}_i^{\min}$ $i = 1, 2, 3$ for $\mathcal{X}$ such that $\mathcal{X} \in \mathcal{C}_1^{\min} \otimes \mathcal{C}_2^{\min} \otimes \mathcal{C}_3^{\min}$.*

Proof. Suppose that $\mathcal{X}$ has a minimal nTD $\mathcal{X} = \mathcal{G} \times_1 F^{(1)} \times_2 F^{(2)} \times_3 F^{(3)}$ and let $\mathcal{C}_i = \text{cone}(F^{(i)})$. By Proposition 3.5, $\mathcal{X} \in \mathcal{C}_1 \otimes \mathcal{C}_2 \otimes \mathcal{C}_3$. Since $F^{(i)} \in \mathbb{R}^{N_i\times\mu\text{rank}_{+,i}(\mathcal{X})}$ we have $\mathcal{O}(\mathcal{C}_i) \leq \mu\text{rank}_{+,i}(\mathcal{X})$. However for any matrix A, one has that $\mathcal{O}(\text{cone}(A)) \geq \text{rank}_+(A)$. Thus $\mathcal{O}(\mathcal{C}_i) \geq \text{rank}_+(F^{(i)}) = \mu\text{rank}_{+,i}(\mathcal{X})$. Combining these two inequalities, we see that $\mathcal{O}(\mathcal{C}_i) = \mu\text{rank}_{+,i}(\mathcal{X})$, so that $\mathcal{C}_i$ are minimal order. Since $\mathcal{X} \in \mathcal{C}_1 \otimes \mathcal{C}_2 \otimes \mathcal{C}_3$, we clearly have $\mathcal{X} \in \mathcal{C}_1 \otimes \mathbb{R}_+^{N_2} \otimes \mathbb{R}_+^{N_3}$ so that $\mathcal{C}_1$ is a minimal cone; likewise for $\mathcal{C}_2$ and $\mathcal{C}_3$.

Conversely suppose that $\mathcal{X}$ has minimal cones $\mathcal{C}_i^{\min}$ (namely $\mathcal{O}(\mathcal{C}_i^{\min}) = \mu\text{rank}_{+,i}(\mathcal{X})$) such that $\mathcal{X} \in \mathcal{C}_1^{\min} \otimes \mathcal{C}_2^{\min} \otimes \mathcal{C}_3^{\min}$. Let $F^{(i)}$ be the matrix whose columns are the extreme rays of $\mathcal{C}_i^{\min}$. Then $F^{(i)} \in \mathbb{R}^{N_i \times \mu\text{rank}_{+,i}(\mathcal{X})}$ and $\text{cone}(F^{(i)}) = \mathcal{C}_i^{\min}$. Since

$$\mathcal{X} \in \mathcal{C}_1^{\min} \otimes \mathcal{C}_2^{\min} \otimes \mathcal{C}_3^{\min} = \text{cone}(F^{(1)}) \otimes \text{cone}(F^{(2)}) \otimes \text{cone}(F^{(3)}),$$

by Proposition 3.5 there exists a $\mathcal{G}$ such that $\mathcal{X} = \mathcal{G} \times_1 F^{(1)} \times_2 F^{(2)} \times_3 F^{(3)}$. By construction, this nnTD is a minimal nnTD. $\square$

We further remark from the proof of Proposition 3.11 that the minimal cones associated with a minimal nnTD are found by considering the extreme rays of the cone.

3.1. Nonnegative Rank Preservation to Tucker Core

Analogous to the real case it is natural to ask if a minimal nnTD always exists, or under what conditions does an nnTD exist? For instance, if $\mu\text{rank}_+(\mathcal{X}) = (r_1, r_2, r_3)$, then does there exist nonnegative cones $\mathcal{C}^{(i)}$ with number of extreme rays equal to r_i such that $\mathcal{X} \in \mathcal{C}^{(1)} \otimes \mathcal{C}^{(2)} \otimes \mathcal{C}^{(3)}$? Example 3.12 demonstrates a tensor can fail to have a minimal nnTD:

Example 3.12. *Consider the* $4 \times 4 \times 3$ *nonnegative tensor with the unfoldings,*

$$\text{unfold}_1(\mathcal{X}) = \left[\begin{array}{cccc|cccc|cccc} 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 2 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 2 & 0 & 2 \end{array}\right],$$

$$\text{unfold}_2(\mathcal{X}) = \left[\begin{array}{cccc|cccc|cccc} 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 2 & 1 & 1 & 2 \\ 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 2 \end{array}\right],$$

$$\text{unfold}_3(\mathcal{X}) = \left[\begin{array}{cccc|cccc|cccc|cccc} 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 2 & 1 & 1 & 2 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 2 \end{array}\right].$$

Suppose there exists a minimal nTD

$$\mathcal{X} = \mathcal{G} \times_1 F^{(1)} \times_2 F^{(2)} \times_3 F^{(3)}$$

with $\mathcal{G} \in \mathbb{R}_+^{\mu rank_{+,1}(\mathcal{X}) \times \mu rank_{+,2}(\mathcal{X}) \times \mu rank_{+,3}(\mathcal{X})}$. *From the decomposition*

$$\text{unfold}_2(\mathcal{X}) = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix} \left[\begin{array}{cccc|cccc|cccc} 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 2 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \end{array}\right]$$

it can be verified that $\mu rank_{+,2}(\mathcal{X}) = 3$, *and therefore* $F^{(2)} \in \mathbb{R}_+^{4\times 3}$. *From the decomposition*

$$\text{unfold}_1(\mathcal{X}) = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix} \left[\begin{array}{cccc|cccc|cccc} 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \end{array}\right]$$

it can be verified that $\mu rank_{+,1}(\mathcal{X}) = 3$, and therefore $F^{(1)} \in \mathbb{R}_+^{4\times 3}$. This decomposition, and the corresponding tensor decomposition $\mathcal{X} = \mathcal{H} \times_1 W$, where

$$W = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix}, \text{unfold}_1(\mathcal{H}) = \left[\begin{array}{cccc|cccc|cccc} 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \end{array}\right],$$

is unique ([26] Theorem 6), which implies that with proper permutation and scaling $W = F^{(1)}$, and $\mathcal{H} = \mathcal{G} \times_2 F^{(2)} \times_3 F^{(3)}$. Note that the $\mu rank_{+,2}(\mathcal{H}) = 4$ since the second unfolding

$$\text{unfold}_2(\mathcal{H}) = \left[\begin{array}{ccc|ccc|ccc} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 \end{array}\right],$$

contains Example Appendix B.2 as a submatrix. However, by Proposition 3.9

$$\begin{aligned} \mu rank_{+,2}(\mathcal{H}) &= \mu rank_{+,2}(\mathcal{G} \times_2 F^{(2)} \times_3 F^{(3)}) \\ &= rank_+(\text{unfold}_2(\mathcal{G} \times_2 F^{(2)} \times_3 F^{(3)}) \\ &= rank_+(F^{(2)}\,\text{unfold}_2(\mathcal{G} \times_3 F^{(3)})) \leq 3. \end{aligned}$$

This is a contradiction, so the supposition that there exists a minimal nnTD is false.

A further question is: if the nnTD: $\mathcal{X} = \mathcal{G} \times_1 F^{(1)} \times_2 F^{(2)} \times_3 F^{(3)}$ does exist, is the nonnegative rank of $\mathcal{X}$ preserved to the nnTD, $\mathcal{G}$, that is, is $\text{rank}_+(\mathcal{X}) = \text{rank}_+(\mathcal{G})$? Example 3.13 demonstrates that even when the minimal nnTD does exist, the nonnegative rank of the tensor is not necessarily preserved to the core.

Example 3.13. *Let $\mathcal{X} = \mathcal{D}_{identity} \times_1 A_{\mathcal{X}}^{(1)} \times_2 A_{\mathcal{X}}^{(2)} \times_3 A_{\mathcal{X}}^{(3)}$, where $\mathcal{D}_{identity}$ is the diagonal identity tensor and*

$$A_{\mathcal{X}}^{(1)} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \quad A_{\mathcal{X}}^{(2)} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad A_{\mathcal{X}}^{(3)} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Kruskal's theorem [27, 28] proves that $rank_+(\mathcal{X}) = 4$, and the nnCPD of the tensor is unique. Using Proposition 3.9, one can show that $\mu rank_+(\mathcal{X}) = (3, 4, 4)$. For example, from the first unfolding of $\mathcal{X}$, we have

$$\text{unfold}_1(\mathcal{X}) = \left[\begin{array}{cccc|cccc|cccc|cccc} 2 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 2 & 2 & 1 & 1 & 2 & 2 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 1 & 2 & 2 & 2 & 1 & 2 & 2 & 2 & 1 & 1 & 1 & 1 & 1 \end{array}\right].$$

The rank of $\text{unfold}_1(\mathcal{X})$ *is* 3*, and this matrix admits a nonnegative decomposition,*

$$\text{unfold}_1(\mathcal{X}) = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix} \left[\begin{array}{cccc|cccc|cccc|cccc} 2 & 2 & 1 & 1 & 2 & 2 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

Using these decompositions, one can show that $\mathcal{X}$ admits a minimal nnTD of the form $\mathcal{X} = \mathcal{G} \times_1 F^{(1)} \times_2 I \times_3 I$ where

$$\text{unfold}_1(\mathcal{G}) = \left[\begin{array}{cccc|cccc|cccc|cccc} 2 & 2 & 1 & 1 & 2 & 2 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

and

$$F^{(1)} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix}.$$

Now, suppose to the contrary that rank$_+(\mathcal{X})$ = rank$_+(\mathcal{G})$. Then, let $\mathcal{G} = \mathcal{D}_{\mathcal{G}} \times_1 A_{\mathcal{G}}^{(1)} \times_2 A_{\mathcal{G}}^{(2)} \times_3 A_{\mathcal{G}}^{(3)}$ be an nCPD of $\mathcal{G}$. Since the nCPD of $\mathcal{X}$ is unique, we have up to permutation and nonnegative scaling that

$$A_{\mathcal{X}}^{(1)} = F^{(1)} A_{\mathcal{G}}^{(1)}. \tag{13}$$

From Example Appendix B.2 we know that rank$_+(A_{\mathcal{X}}^{(1)}) = 4$. But then

$$\textit{rank}_+(A_{\mathcal{X}}^{(1)}) > 3 = \textit{rank}_+(F^{(1)}) \geq \textit{rank}_+(F^{(1)} A_{\mathcal{G}}^{(1)}),$$

which is a contradiction. Therefore, rank$_+(\mathcal{X}) \neq$ rank$_+(\mathcal{G})$.

Example 3.13 highlights a key difference between the real and nonnegative minimal Tucker decompositions. By Theorem 2.13, if the CPD is unique then the column space of the CPD loading matrices will recover the minimal subspaces. However when the nonnegative CPD is unique, the loading matrices can still fail to capture the minimal cone ($\mathcal{C}_{\min}^{(1)}$ and $A_{\mathcal{X}}^{(1)}$ in previous example). In particular, the nnCPD cannot be a minimal Tucker decomposition in this case. This causes problems with preservation of the rank to the core of the tensor. It turns out that this issue in Example 3.13 is always hold. Namely, when $\mathcal{X}$ has a unique nnCPD and a Tucker has a factor with nonnegative rank smaller then the loading matrix factor in the nnCPD, nonnegative rank cannot be preserved:

Theorem 3.14. *Let $\mathcal{X}$ be a nonnegative tensor with unique nCPD $\mathcal{X} = \mathcal{D}_{identity} \times_1 A_{\mathcal{X}}^{(1)} \times_2 A_{\mathcal{X}}^{(2)} \times_3 A_{\mathcal{X}}^{(3)}$. Suppose that $\mathcal{X}$ has a nTD $\mathcal{X} = \mathcal{G} \times_1 F^{(1)} \times_2 F^{(2)} \times_3 F^{(3)}$ where rank$_+(A^{(i)}) >$ rank$_+(F^{(i)})$ for some $i = 1, 2, 3$. Then rank$_+(\mathcal{X}) \neq$ rank$_+(\mathcal{G})$.*

Proof. Without loss of generality, let $i = 1$, i.e. $\text{rank}_+(A^{(1)}) > \text{rank}_+(F^{(1)})$. Suppose to the contrary that $\text{rank}_+(\mathcal{X}) = \text{rank}_+(\mathcal{G})$. Let $\mathcal{G} = \mathcal{D}_{\mathcal{G}} \times_1 A_{\mathcal{G}}^{(1)} \times_2 A_{\mathcal{G}}^{(2)} \times_3 A_{\mathcal{G}}^{(3)}$ be an nCPD of $\mathcal{G}$. Then both $\mathcal{X} = \mathcal{D}_{identity} \times_1 A_{\mathcal{X}}^{(1)} \times_2 A_{\mathcal{X}}^{(2)} \times_3 A_{\mathcal{X}}^{(3)}$ and

$$\mathcal{X} = \mathcal{G} \times_1 F^{(1)} \times_2 F^{(2)} \times_3 F^{(3)} = \mathcal{D}_{\mathcal{G}} \times_1 (F^{(1)} A_{\mathcal{G}}^{(1)}) \times_2 (F^{(1)} A_{\mathcal{G}}^{(2)}) \times_3 (F^{(1)} A_{\mathcal{G}}^{(3)})$$

are rank decompositions of $\mathcal{X}$. Since the nCPD of $\mathcal{X}$ is unique, up to permutation and nonnegative scaling one has

$$A_{\mathcal{X}}^{(1)} = F^{(1)} A_{\mathcal{G}}^{(1)}.$$

Thus, $\text{rank}_+(A_{\mathcal{X}}^{(1)}) = \text{rank}_+(F^{(1)} A_{\mathcal{G}}^{(1)})$. However, by assumption

$$\text{rank}_+(A_{\mathcal{X}}^{(1)}) > \text{rank}_+(F^{(1)}) \geq \text{rank}_+(F^{(1)} A_{\mathcal{G}}^{(1)}),$$

a contradiction. $\square$

Theorem 3.14 gives condition on when the rank is *not* preserved based on the CPD. We believe that for a large class of nonnegative tensors where compression is achieved, this implies that the rank is not preserved. However, deriving precises probabilistic statements is challenging due to the non-stochastic relationship between loading matrices and random tensors.

Examples 3.12 and 3.13 demonstrate the subtleties of the nonnegative factorizations compared to the real valued. First, the minimal nnTD can fail to exist. Second, even if it exists, the nonnegative rank of the minimal nnTD core may not be equal to the nonnegative rank of $\mathcal{X}$. The following Theorem provides sufficient conditions for a minimal nnTD to exist, and for the nonnegative rank of the tensor to be preserved to the core of the minimal nnTD. We note that because of Theorem 3.14, a rank requirement for nnCPD loading matrices is required. The following is the nonnegative analog of Theorem 2.12.

Theorem 3.15. *Suppose a nonnegative tensor $\mathcal{X}$ has an nCPD: $\mathcal{X} = \mathcal{D}_{\mathcal{X}} \times_1 A_{\mathcal{X}}^{(1)} \times_2 A_{\mathcal{X}}^{(2)} \times_3 A_{\mathcal{X}}^{(3)}$ with $\text{rank}_+(A_{\mathcal{X}}^{(i)}) = \mu\text{rank}_{+,i}(\mathcal{X})$ for $1 \leq i \leq 3$. Then a minimal nTD: $\mathcal{X} = \mathcal{G} \times_1 F^{(1)} \times_2 F^{(2)} \times_3 F^{(3)}$ exists such that $\text{rank}_+(\mathcal{X}) = \text{rank}_+(\mathcal{G})$.*

Proof. Since $\text{rank}_+(A_{\mathcal{X}}^{(i)}) = \mu\text{rank}_{+,i}(\mathcal{X})$ for $1 \leq i \leq 3$, each $A_{\mathcal{X}}^{(i)}$ has a nonnegative decomposition $A_{\mathcal{X}}^{(i)} = W^{(i)} H^{(i)}$. Substituting into the nCPD and distributing

$$\begin{aligned}
\mathcal{X} &= \mathcal{D}_{\mathcal{X}} \times_1 A_{\mathcal{X}}^{(1)} \times_2 A_{\mathcal{X}}^{(2)} \times_3 A_{\mathcal{X}}^{(3)} \\
&= \mathcal{D}_{\mathcal{X}} \times_1 W^{(1)} H^{(1)} \times_2 W^{(2)} H^{(2)} \times_3 W^{(3)} H^{(3)} \\
&= (\mathcal{D}_{\mathcal{X}} \times_1 H^{(1)} \times_2 H^{(2)} \times_3 H^{(3)}) \times_1 W^{(1)} \times_2 W^{(2)} \times_3 W^{(3)} \\
&= \mathcal{G} \times_1 W^{(1)} \times_2 W^{(2)} \times_3 W^{(3)}
\end{aligned}$$

where $\mathcal{G} = \mathcal{D}_{\mathcal{X}} \times_1 H^{(1)} \times_2 H^{(2)} \times_3 H^{(3)}$. The core $\mathcal{G}$ is a nonnegative tensor with shape equal to the nonnegative minimal multiranks of $\mathcal{X}$, so $\mathcal{X} = \mathcal{G} \times_1 W^{(1)} \times_2 W^{(2)} \times_3 W^{(3)}$ is a minimal nTD. To prove $\text{rank}_+(\mathcal{X}) = \text{rank}_+(\mathcal{G})$, it once again suffices to show that $\text{rank}_+(\mathcal{X}) \geq \text{rank}_+(\mathcal{G})$. However from the constructed decomposition $\mathcal{G} = \mathcal{D}_{\mathcal{X}} \times_1 H^{(1)} \times_2 H^{(2)} \times_3 H^{(3)}$ we know $\text{rank}_+(\mathcal{G}) \leq \text{rank}_+(\mathcal{X})$. $\square$

Theorem 3.15 demonstrates that a minimal nnTD exists that will preserve the rank. However contrary to Theorem 2.12, it does not state that every minimal nnTD will preserve the rank to the core. This is yet another fundamental challenge one must surmount in the nonnegative case - not every minimal nnTD will necessarily preserve the rank. The following example illustrates this issue:

Example 3.16. *Let $\mathcal{X} \in \mathbb{R}_+^{3,3,3}$ be the tensor given by*

$$\mathcal{X} = \left[\begin{array}{ccc|ccc|ccc} 2 & 8 & 3 & 1 & 5 & 2 & 2 & 8 & 3 \\ 4 & 15 & 5 & 2 & 8 & 3 & 4 & 15 & 5 \\ 2 & 6 & 2 & 1 & 3 & 1 & 2 & 6 & 2 \end{array}\right].$$

Then $\mathcal{X}$ has an nnCPD given by

$$A_{\mathcal{X}}^{(1)} = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 1 & 0 \\ 2 & 1 & 1 \end{bmatrix}, \quad A_{\mathcal{X}}^{(2)} = \begin{bmatrix} 1 & 1 & 0 \\ 2 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \quad A_{\mathcal{X}}^{(3)} = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 2 & 1 \\ 1 & 1 & 0 \end{bmatrix}.$$

One can check that $\text{rank}_+(A_{\mathcal{X}}^{(1)}) = 2 = \mu\text{rank}_{+,1}(\mathcal{X})$ and $\text{rank}_+(A_{\mathcal{X}}^{(2)}) = \text{rank}_+(A_{\mathcal{X}}^{(3)}) = 3 = \mu\text{rank}_{+,2}(\mathcal{X}) = \mu\text{rank}_{+,3}(\mathcal{X})$. Thus, $\mathcal{X}$ satisfies the hypothesis of Theorem 3.15. We will now show that there are two minimal nTDs

$$\mathcal{X} = \mathcal{G}_1 \times_1 F_1^{(1)} \times_2 F_1^{(2)} \times_3 F_1^{(3)} = \mathcal{G}_2 \times_1 F_2^{(1)} \times_2 F_2^{(2)} \times_3 F_2^{(3)}$$

with rank($\mathcal{X}$) = rank$_+(\mathcal{G}_1)$ < rank$_+(\mathcal{G}_2)$. Therefore, not every minimal nnTD can preserve the rank to the Tucker core. Indeed, one can check

$$\mathcal{G}_1 = \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 2 & 1 \end{array}\right], \quad F_1^{(1)} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}, \quad F_1^{(2)} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \quad F_1^{(3)} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

and

$$\mathcal{G}_2 = \left[\begin{array}{ccc|ccc} 0 & 2 & 1 & 1 & 3 & 1 \\ 0 & 1 & 1 & 2 & 7 & 2 \\ 0 & 0 & 0 & 1 & 3 & 1 \end{array}\right], \quad F_2^{(1)} = \begin{bmatrix} 1 & 2 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}, \quad F_2^{(2)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad F_2^{(3)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

result in minimal nTDs of $\mathcal{X}$. From the decomposition

$$\mathcal{G}_1 = \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{array}\right] + \left[\begin{array}{ccc|ccc} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] + \left[\begin{array}{ccc|ccc} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{array}\right]$$

we see that rank$_+(\mathcal{G}_1) \leq 3$. Since rank$_+(\mathcal{X}) = 3$ by above, we have that rank$_+(\mathcal{G}_1) = 3$. We now show that rank$_+(\mathcal{G}_2) \neq 3$. Indeed by Kruskal's Theorem [27], $\mathcal{X}$ has a unique nnCPD. By Proposition 1 of [29], rank$_+(\mathcal{G}_1)$ = rank$_+(\mathcal{X}) = 3$ if and only if $A_{\mathcal{X}}^{(i)} \subset \operatorname{cone}(F_2^{(i)})$. Since $[1,0,1] \notin \operatorname{cone}(F_2^{(1)})$, we see that the rank cannot be preserved to the core $\mathcal{G}_2$.

Cohen *et al.* [29] (see Proposition 1) provide some necessary and sufficient conditions for the nonnegative rank of a tensor to persist to the core of an nnTD under some geometric hypothesis. We remark that their theorem, as stated, requires uniqueness of the nnCPD along with a full column rank condition on the factors of the nnCPD. However the full column rank is not needed, and the uniqueness of the nnCPD is only required for one direction. Namely, that if the nnCPD is unique and the rank of the tensor is preserved to the core, then the (unique) nnCPD factors are contained inside the cones from the nnTD loading matrices. We made the equivalent converse statement in Theorem 3.17. It too, requires uniqueness of the nnCPD. However, since one half of our 'if and only if' does not require uniqueness, we have opted to separate the two conditions.

3.2. Nonnegative Rank Deficiency in nnCPD Factors

We now discuss the difficulties associated from nonnegative rank deficent factors in a nnCPD. We recap the work above to discuss the challenges facing a nonnegative analog of CANDELINC - in particular, the issues surrounding the existence of a min nnTD. Following the subsection on real valued CANDELINC above, we begin by exploring the relations between the nonnegative minimal multirank and the nonnegative ranks of nnCPD factors. The following theorem relates the uniqueness of a nnCPD to the minimal nnTDs:

Theorem 3.17. *Let $\mathcal{X} = \mathcal{D} \times_1 A^{(1)} \times_2 A^{(2)} \times_3 A^{(3)}$ be an nnCPD, then μrank$_{+,i}(\mathcal{X}) \leq$ rank$_+(A^{(i)})$. Furthermore, if the nnCPD is unique and there exists a minimal nnTD with rank$_+(\mathcal{X})$ = rank$_+(\mathcal{G})$ then μrank$_{+,i}(\mathcal{X})$ = rank$_+(A^{(i)})$*

Proof. That $\mu\mathrm{rank}_{+,i}(\mathcal{X}) \leq \mathrm{rank}_+(A^{(i)})$ follows from the fact that an nnCPD is a nnTD. Now suppose that $\mathcal{X}$ has a unique nCPD, and consider a minimal nTD of the form $\mathcal{X} = \mathcal{G} \times_1 F^{(1)} \times_2 F^{(2)} \times_3 F^{(3)}$. We recall that $\mathrm{rank}_+(F^{(i)}) = \mu\mathrm{rank}_{+,i}(\mathcal{X})$ by definition. On its turn, the Tucker core admits an nCPD $\mathcal{G} = \mathcal{D}_{\mathcal{G}} \times_1 B^{(1)} \times_2 B^{(2)} \times_3 B^{(3)}$. Substituting the nCPD of $\mathcal{G}$ in the nTD of $\mathcal{X}$, we obtain the alternative nCPD $\mathcal{X} = \mathcal{D}_{\mathcal{G}} \times_1 F^{(1)} B^{(1)} \times_2 F^{(2)} B^{(2)} \times_3 F^{(3)} B^{(3)}$. Since we assume that the nCPD is unique, with appropriate nonnegative scalings and permutations, which are nonnegative rank-preserving operations, we obtain that $A^{(i)} = F^{(i)} B^{(i)}$, which proves that $\mathrm{rank}_+(A^{(i)}) \leq \mathrm{rank}_+(F^{(i)}) = \mu\mathrm{rank}_{+,i}(\mathcal{X})$. $\square$

Theorem 3.17 illuminates an algorithmic challenge of computing a nCPD directly. Suppose $\mathcal{X} = \mathcal{D} \times_1 A^{(1)} \times_2 A^{(2)} \times_3 A^{(3)}$ has a unique nCPD with $\text{rank}_+(\mathcal{X}) = r$ and minimal multiranks $\mu\text{rank}_i(\mathcal{X}) = r_i$. In practice, many shaped tensors have $r_i < r$. Thus, $A^{(i)}$ is a rank deficient matrix by Theorem 3.17, which can be an algorithmically challenging task to overcome.

This motivates the need for nonnegative version of CANDELINC. Concretely, if $\mathcal{X} = \mathcal{G} \times_1 F^{(1)} \times_2 F^{(2)} \times_3 F^{(3)}$ is a minimal nnTD such that $\text{rank}_+(\mathcal{G}) = \text{rank}_+(\mathcal{X})$, and $\mathcal{G} = \mathcal{D}_\mathcal{G} \times_1 A_\mathcal{G}^{(1)} \times_2 A_\mathcal{G}^{(2)} \times_3 A_\mathcal{G}^{(3)}$ is the CPD of the Tucker core, then each factor $A_\mathcal{G}^{(i)}$ is a full column rank matrix This suggests that, under some conditions, a stable way of computing nnCPD is to first compute a minimal nnTD, e.g., using approximate NMF, then compute nCPD on the nonnegative Tucker core, and finally substitute the nnCPD of the core-tensor in the nnTD to obtain the final nnCPD of the original tensor.

Unfortunately, the previous work highlights some of the major challenges a nonnegative CANDELINC must overcome. Indeed a min nnTD need not exist (Example 3.12), and even when it does, it need not preserve the rank to the core (Example 3.13). Furthermore, a tensor may have a minimal nnTD which preserves the rank, but this does not mean all minimal nnTD will preserve the rank to the core (Example 3.16).

These issues indicate that minimal Tucker decompositions, while desirable, are perhaps not feasible in the nonnegative case. To overcome this hurdle, we redirect our interest to nonnegative Tucker decompostions we call *canonical*. In the next subsection, we define the canonical Tucker decomposition and show that one always exists which preserves the rank to the core.

3.3. The Canonical Multirank and Canonical Tucker

Ultimately, we desire a decomposition that 1) preserves the nCPD rank to the core and 2) does not require extraction of rank deficient matrices. While minimal nnTD's are devoid of rank deficiency in the loading matrices, they may not preserve the rank to the core (if they exist at all). This conversation leads us to desire a less strict type of Tucker decomposition which not only preservers the rank, but whose loading matrices are also devoid of rank deficiency. We therefore make the following definition:

Definition 3.18. *Let $\mathcal{X} \in \mathbb{R}_+^{N_1 \times N_2 \times N_3}$ have a unique nnCPD given by $\mathcal{X} = \mathcal{D} \times_1 A^{(1)} \times_2 A^{(2)} \times_3 A^{(3)}$. The i-th canonical (nonnegative) multirank of $\mathcal{X}$, denoted $\kappa_i(\mathcal{X})$, is the rank of the i-th nnCPD factor $rank_+(A^{(i)})$. The canonical (nonnegative) multirank of $\mathcal{X}$ is the triple*

$$\kappa(\mathcal{X}) = (\kappa_1(\mathcal{X}), \kappa_2(\mathcal{X}), \kappa_3(\mathcal{X})) = (rank_+(A^{(1)}), rank_+(A^{(2)}), rank_+(A^{(3)})).$$

Corresponding to the canonical multirank we have a canonical Tucker decomposition.

Definition 3.19. *Consider tensor $\mathcal{X} \in \mathbb{R}_+^{N_1 \times N_2 \times N_3}$ with unique nnCPD. We say that the Tucker decomposition $\mathcal{X} = \mathcal{G} \times_1 F^{(1)} \times_2 F^{(2)} \times_3 F^{(3)}$ is* canonical *if the dimensions of the core tensor $\mathcal{G}$ are equal to the canonical multiranks, i.e., $\mathcal{G} \in \mathbb{R}_+^{\kappa_1(\mathcal{X}) \times \kappa_2(\mathcal{X}) \times \kappa_3(\mathcal{X})}$ and $F^{(i)} \in \mathbb{R}_+^{N_i \times \kappa_i(\mathcal{X})}$, $i = 1, 2, 3$.*

By definition $\mu\text{rank}_{i,+}(\mathcal{X}) \leq \kappa_i(\mathcal{X})$. Therefore, a canonical Tucker is a less restrictive shape constraint than a minimal Tucker. Unlike minimal Tuckers, one can always find a canonical Tucker that will preserve the rank:

Theorem 3.20. *Let $\mathcal{X} \in \mathbb{R}_+^{N_1 \times N_2 \times N_3}$ have unique nnCPD. Then there exists a canonical nnTD which preserves the rank.*

Proof. Let $\mathcal{X} \in \mathbb{R}_+^{N_1 \times N_2 \times N_3}$ have unique nnCPD given by $\mathcal{X} = \mathcal{D} \times_1 A^{(1)} \times_2 A^{(2)} \times_3 A^{(3)}$. For each $i = 1, 2, 3$, consider the rank factorizations $A^{(i)} = W^{(i)} H^{(i)}$. Subbing these factorizations into the nnCPD:

$$\begin{aligned} \mathcal{X} &= \mathcal{D} \times_1 A^{(1)} \times_2 A^{(2)} \times_3 A^{(3)} \\ &= \left(\mathcal{D} \times_1 H^{(1)} \times_2 H^{(2)} \times_3 H^{(3)}\right) \times_1 W^{(1)} \times_2 W^{(2)} \times_3 W^{(3)}. \end{aligned}$$

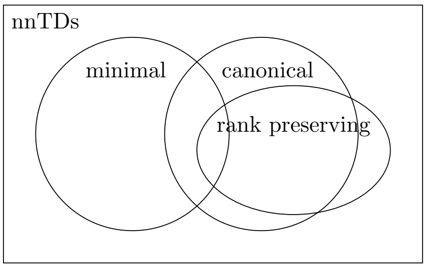


Figure 2: Diagram of nonnegative Tucker decompositions for tensors with a unique nnCPD.

Let $\mathcal{G} := \mathcal{D} \times_1 H^{(1)} \times_2 H^{(2)} \times_3 H^{(3)}$. By construction, $W^{(i)} \in \mathbb{R}_+^{N_i \times \kappa_i(\mathcal{X})}$ so that $\mathcal{X} = \mathcal{G} \times_1 W^{(1)} \times_2 W^{(2)} \times_3 W^{(3)}$ is a canonical Tucker. Furthermore, $\text{rank}_+(\mathcal{G}) \leq \text{rank}_+(\mathcal{X})$ so that $\text{rank}_+(\mathcal{G}) = \text{rank}_+(\mathcal{X})$. $\square$

Theorem 3.20 states that there exists a rank preserving canonical Tucker. However, Example 3.16 shows that not every canonical Tucker can preserve the rank. Unlike its real counterparts, the shape of a nonnegative Tucker decomposition does not guarantee that all such factorizations will preserve rank.

In minimal nnTD, the rank of the loading matrix $F^{(i)}$ is equal to $\mu\text{rank}_{i,}(\mathcal{X})$ so that the matrix is not degenerate. Note that in a canonical nnTD $\mathcal{X} = \mathcal{G} \times_1 F^{(1)} \times_2 F^{(2)} \times_3 F^{(3)}$, one has that $\mu\text{rank}_{i,+}(\mathcal{X}) \leq \text{rank}_+(F^{(i)}) \leq \kappa_i(\mathcal{X})$. If one selects a canonical nnTD which preserves the rank to the core, then $\text{rank}_+(F^{(i)}) = \kappa_i(\mathcal{X})$ so that once again, the matrix is not degenerate. The challenge here lies instead in the selection of the correct cones that contains the minimal cones needed to preserve the rank.

Proposition 3.21. *Let $\mathcal{X} \in \mathbb{R}_+^{N_1 \times N_2 \times N_3}$ have unique nnCPD given by $\mathcal{X} = \mathcal{D} \times_1 A^{(1)} \times_2 A^{(2)} \times_3 A^{(3)}$. Suppose $\mathcal{X} = \mathcal{G} \times_1 F^{(1)} \times_2 F^{(2)} \times_3 F^{(3)}$ is a canonical Tucker such that $R = \text{rank}_+(\mathcal{X}) = \text{rank}_+(\mathcal{G})$. Then $\text{rank}_+(F^{(i)}) = \kappa_i(\mathcal{X})$.*

Proof. From the preceding comments, we have that $\text{rank}_+(F^{(i)}) \leq \kappa_i(\mathcal{X})$. Since $\text{rank}_+(\mathcal{X}) = \text{rank}_+(\mathcal{G})$ from the canonical nnTD, we have that $\mathcal{G}$ admits a nonnegative rank R nnCPD

$$\mathcal{G} = \mathcal{D} \times_1 B^{(1)} \times_2 B^{(2)} \times_3 B^{(3)}.$$

Substituting into the canonical nnTD, we two rank R decompositions of $\mathcal{X}$ given by

$$\begin{aligned} \mathcal{X} &= \mathcal{D} \times_1 A^{(1)} \times_2 A^{(2)} \times_3 A^{(3)} \\ &= \mathcal{D} \times_1 (F^{(1)}B^{(1)}) \times_2 (F^{(2)}B^{(2)}) \times_3 (F^{(3)}B^{(3)}). \end{aligned}$$

Thus up to scaling and permutation, $F^{(i)}B^{(i)} = A^{(i)}$ so that $\text{rank}_+(F^{(i)}) \geq \text{rank}_+(A^{(i)}) = \kappa_i(\mathcal{X})$. $\square$

What Theorem 3.20 and Proposition 3.21 indicate is that a nonnegative canonical Tucker decomposition gives just enough breathing room to provide existence of a not rank degenerate Tucker as seen in Figure 2. In the diagram, we see the relation that if a minimal nnTD also preserves the rank, then the minimal multiranks are also the canonical multiranks. There are rank preserving nnTDs that are neither minimal, nor canonical. These can be obtained, for instance, by taking a rank preserving canonical and adding appropriate zeros. As noted above, not every nonnegative canonical TD can preserve the rank, and identifying the correct extensions of the minimal cones to higher order cones to construct a rank preserving nnTD is a challenge. In the next section, we compare the performance of different nonnegative CANDELINC algorithms to demonstrate these challenges on synthetic and real data.

4. Numerical Experiments

Exact decompositions of tensors from real or experimental data are typically not attainable, so we solve for approximations with optimization problems. To find a nnCANDELINC approximation of a tensor $X \in \mathbb{R}_+^{n_1,n_2,n_3}$, given a set of multirank dimensions $[m_1, m_2, m_3]$ and a tensor rank r, a typical Frobenius norm optimization problem is,

$$\begin{array}{ll} \underset{F^{(i)},B^{(i)}}{\text{minimize}} & \left\| \mathcal{X} - \mathcal{I} \times_1 F^{(1)}B^{(1)} \times_2 F^{(2)}B^{(2)} \times_3 F^{(3)}B^{(3)} \right\|_{\mathcal{F}}^2 \\ \text{Subject to:} & F^{(i)} \in \mathbb{R}_+^{n_i,m_i} \text{ for } 1 \leq i \leq 3 \\ & B^{(i)} \in \mathbb{R}_+^{m_i,r} \text{ for } 1 \leq i \leq 3 \; . \end{array} \tag{14}$$

A similar optimization problem, without the nonnegativity constraints, is suitable for CANDELINC. To compute CANDELINC decompositions, theory informs us of two procedures to find approximate decompositions using the readily available tools of TD, CPD and SVD. One procedure follows as: first compute a TD followed by a CPD on the core. Alternatively: first compute a CPD followed by SVDs on each of the CPD factors. Theoretically, both of these two procedures provide equally valid CANDELINC decompositions under perfect conditions. In practice the first procedure is often preferred as it is typically less computationally expensive with the cheap dimension reduction via Tucker compression before the more expensive CPD step. In this section, we explore two procedures to compute nnCANDELINC decompositions. We discuss the theory, benefits and drawbacks, and demonstrate their performance on synthetic and real data.

4.1. Algorithms and Scoring

While a multitude of methods can be applied to the nnCANDELINC optimization in Equation 14, we concentrate on procedures that can be built using readily available tools. Namely, we are interested in procedures that compute nnCANDElINC decompositions using the sub-procedures: nnTD, nnCPD, and NMF.

Algorithm 1 computes a nnCANDELINC decomposition by first nnTD compression, reducing the dimension, and then by nnCPD on the resulting core. Much of the previously discussed theory informs us of potential problems with this procedure in selecting various multiranks and ranks for the nnTD and nnCPD dimensions. Namely, it is necessary, but not sufficient, for the nnTD to be computed with dimensions greater or equal to the canonical multiranks, not the minimal multiranks, to obtain a nnCANDELINC decomposition that expresses the tensors rank.

Algorithm 1 nnCANDELINC algorithm using nnTD and nnCPD.

Require: $\mathcal{X} \in \mathbb{R}_+^{N_1,N_2,N_3}, m \in \mathbb{N}^3, r \in \mathbb{N}$
Ensure: $F^{(i)} \in \mathbb{R}_+^{N_i,m_i}, B^{(i)} \in \mathbb{R}_+^{m_i,r}$
  $\mathcal{G}, F^{(1)}, F^{(2)}, F^{(3)} \leftarrow \text{nnTD}(X, (m_1, m_2, m_3))$
  $B^{(1)}, B^{(2)}, B^{(3)} \leftarrow \text{nnCPD}(\mathcal{G}, r)$

Alternatively, Algorithm 2 computes a nnCANDELINC decomposition by first computing nnCPD, followed by NMF on each of the nnCPD factors. Here, theory informs us that it is both necessary and sufficient to use the canonical multiranks as the latent dimensions in their respective NMFs.

To demonstrate the performance of the Algorithms we utilize the functions

Algorithm 2 nnCANDELINC algorithm using nnCPD and NMF.

Require: $\mathcal{X} \in \mathbb{R}_+^{N_1,N_2,N_3}, m \in \mathbb{N}^3, r \in \mathbb{N}$
Ensure: $F^{(i)} \in \mathbb{R}_+^{N_i,m_i}, B^{(i)} \in \mathbb{R}_+^{m_i,r}$
$A^{(1)}, A^{(2)}, A^{(3)} \leftarrow \text{nCPD}(X, r)$
$F^{(1)}, B^{(1)} \leftarrow \text{NMF}(A^{(1)}, m_1)$
$F^{(2)}, B^{(2)} \leftarrow \text{NMF}(A^{(2)}, m_2)$
$F^{(3)}, B^{(3)} \leftarrow \text{NMF}(A^{(3)}, m_3)$

tensorly.decomposition.non_negative_tucker

and

tensorly.decomposition.non_negative_parafac

from the freely available high-level API for tensor decomposition methods in python, TensorLy [30], and

sklearn.decomposition.NMF

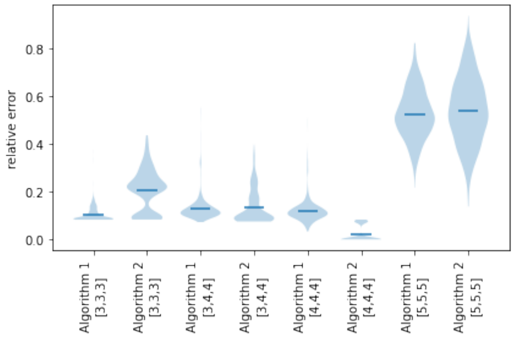
from scikit-learn [31]. In all experiments, random initializations are used with each call and a constant 5000 iterations are used for each sub-optimization to ensure reasonable convergence, with no early termination criteria.

To evaluate the resulting decompositions of these algorithms we utilize two different scores. The congruence [32] between two rank one tensors, $\mathcal{X} = a_1 \otimes b_1 \otimes c_1$ and $\mathcal{Y} = a_2 \otimes b_2 \otimes c_2$ is,

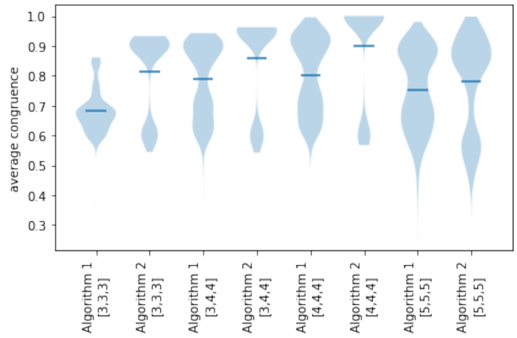
$$\text{cong}(\mathcal{X}, \mathcal{Y}) = \cos(X, Y) = \frac{a_1^\top \cdot a_2}{\|a_1\|_2 \|a_2\|_2} \cdot \frac{b_1^\top \cdot b_2}{\|b_1\|_2 \|b_2\|_2} \cdot \frac{c_1^\top \cdot c_2}{\|c_1\|_2 \|c_2\|_2}.$$

The mean congruence of all rank one factors is relevant for two rank r tensors after the appropriate permutation of the factors is applied to maximize the mean congruence [33]. We apply the mean congruence to the appropriate products, $F^{(i)}B^{(i)}$, in the nnCANDELINC decompositions. For a tensor $\mathcal{X}$, the Frobenius norm is defined as the square root of the sum of the squares or $||\mathcal{X}||_F = \sqrt{\sum_{i,j,k} \mathcal{X}_{i,j,k}^2}$. The relative reconstruction error (we call further relative error) of the decomposition is the ratio of the Frobenius norm of the residual and the Frobenius norm of the tensor or matrix. In addition to the mean congruence, we utilize the relative reconstruction error to evaluate the quality of the nnCANDELINC decompositions.

To evaluate these algorithms and relate them to theory we apply them to both synthetic and real datasets. Our first investigation of a synthetic tensor highlights the importance of using the nonnegative canonical multiranks and not the minimal multiranks. We additionally construct a large number nonnegative tensors with pre-determined nonnegative canonical multiranks, and show the performance of each algorithm at recovering the factors to evaluate them in more generic situations. For the first real dataset, we apply the nnCANDELINC algorithms to extract the nCPD features of a well-known fluorescence data that has been previously analyzed in the PhD Thesis of Bro [34]. Next, we apply nnCANDELINC to a computer generated $3D$ dataset with nonnegative rank deficient nCPD factors that represents a microphase separation of block copolymers as a function of temperature and was analyzed in [35].



(a) Violin plots of relative errors of decompositions.



(b) Violin plots of congruence scores of decompositions.

Figure 3: Violin plots and means of the results of nnCANDELINC decompositons on the tensor from Example 3.13 using the various nonnegative multiranks in Algorithms 1 and 2.

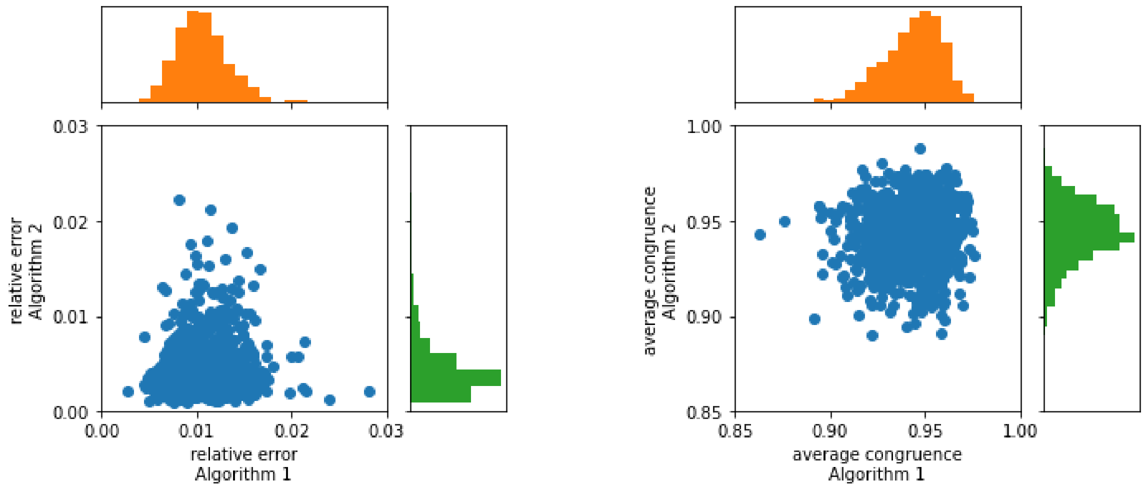
4.2. Various Multiranks

We first investigate the efficacy of Algorithms 1 and 2 when various multiranks are used for the nnTD and NMF dimensions. Example 3.13 provides an instance where the nonnegative rank, minimal multiranks, and canonical multiranks are all known. We evaluate nnCANDELINC on the proposed in the Example 3.13 tensor with four different effective nonnegative multiranks:

- $[3, 3, 3]$ which are less than the minimal multiranks of the tensor,
- $[3, 4, 4]$ which are the minimal multiranks which is known not to preserve the rank to the core,
- $[4, 4, 4]$ which are the canonical multiranks where it is feasible that the rank is preserved to the core,
- $[5, 5, 5]$ which are greater than the latent dimensions needed everywhere.

To evaluate the algorithms we decompose the Example 3.13 tensor 1000 times with each set of assumed multiranks, each starting from random initial conditions.

Figure 3 reports violin plots of the relative errors and average congruence scores of the resulting decompositions. In Figure 3a we see mild relative errors for the three smallest assumed nonnegative multiranks $[3, 3, 3], [3, 4, 4]$, and $[4, 4, 4]$. The multiranks $[3, 3, 3]$ and $[3, 4, 4]$ are not sufficient to obtain a rank revealing nnCANDELINC decomposition resulting in moderate relative errors. Using the canonical nonnegative multiranks, $[4, 4, 4]$ is expectedly more successful with Algorithm 2. This is unsurprising since for both Algorithm 1 and 2 the canonical multiranks are necessary to obtain a low relative error, but for Algorithm 2 the use of the canonical multiranks is both necessary and sufficient. Using an assumed multirank of $[5, 5, 5]$ leads to disastrous performance with both algorithms, on the surface this is surprising since the problem has more degrees of freedom than necessary to perfectly reconstruct the tensor. Figure 3b reflects the mean and standard deviations of the average congruences for each scheme. There is little correlation between the relative error performances and the congruence scores. By the congruence score measure, Algorithm 2 shows better scores than Algorithm 1 on this tensor.



(a) Relative errors of decompositions of synthetic dataset.

(b) Average congruence scores of decompositions of synthetic dataset.

Figure 4: Scatterplots and projected histograms from results of nnCANDELINC decompositons using Algorithms 1 and 2 on randomly generated synthetic dataset. Each point corresponds to a tensor whose location indicates how well each algorithm performed.

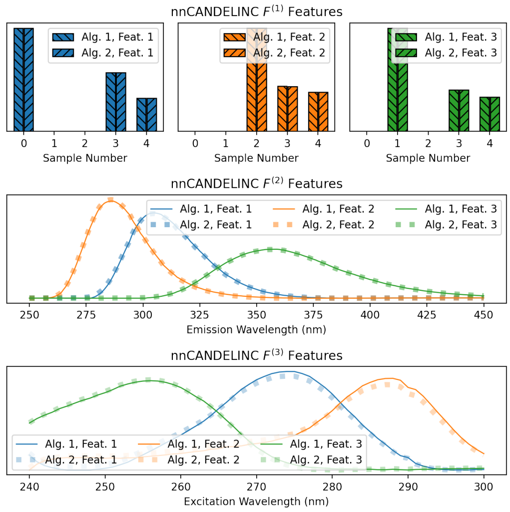
4.3. Randomly Generated Tensors

To evaluate the performance of the nnCANDELINC Algorithms on a more varied dataset, we randomly generate synthetic tensors that have a unique nnCPD with nonnegative rank deficient factors and where the canonical multiranks are known. To do this, we randomly generate our nnCPD factors as a product of two smaller nonnegative matrices, and confirm that they satisfy the suppositions of Kruskal's theorem [36]. For specified dimensions N_1, N_2, N_3, nonnegative canonical multiranks r_1, r_2, r_3, and rank r, we randomly sample from a uniform distribution the factors $F^{(i)} \in \mathbb{R}_+^{N_i \times r_i}$ and $B^{(i)} \in \mathbb{R}_+^{r_i \times r}$ to construct the decomposition $\mathcal{X} = \mathcal{I} \times_1 (F^{(1)}B^{(1)}) \times_2 (F^{(1)}B^{(2)}) \times_3 (F^{(1)}B^{(3)})$. We ensure that factors satisfy the Kruskal rank criteria, Kruskal-rank$(F^{(1)}B^{(1)})$ + Kruskal-rank$(F^{(2)}B^{(2)})$ + Kruskal-rank$(F^{(3)}B^{(3)}) \leq 2r + 2$.

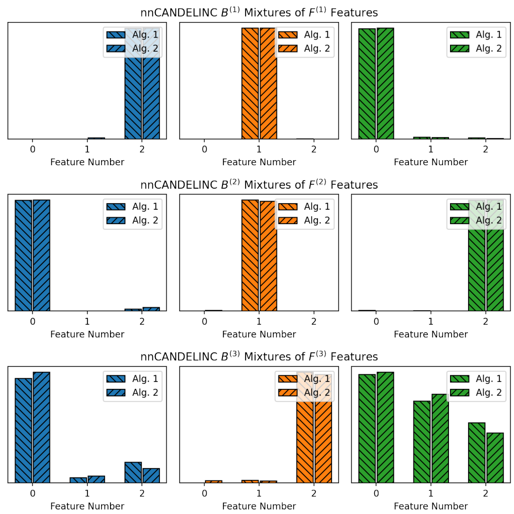
Here we report the results of 1000 randomly generated tensors, each of size $N_1 = N_2 = N_3 = 40$ with a nonnegative rank of 5 and canonical multiranks of $[3, 4, 5]$. Figure 4 depicts scatterplots of the resulting relative errors and congruence scores using each algorithm for each of the 1000 tensors. Both algorithms yield a low relative error and high average congruence score with Algorithm 2 demonstrating better scores than Algorithm 1.

4.4. Fluorescence Data Decompositions

The experimental fluorescence dataset includes five samples, each with different amounts of amino acids of three types: tyrosine, tryptophan and phenylalanine dissolved in buffered water. We consider this data to demonstrate the different algorithm performances when there is no strong linear dependence of the factors. The fluorescence in these samples has been excited by UV irradiation at wavelengths, $\lambda \in (240 - 300\text{nm})$. The UV-emission was measured by the spectrofluorometer at wavelengths $\lambda \in [250, 450]$nm by sampling at 1 nm intervals. The experimental data formed a $3D$ array with size $5 \times 61 \times 201$. If we assume that each amino acid gives a nonnegative linear contribution



(a) Comparison of features $F^{(i)}$ as extracted from the fluorescence tensor using both Algorithms 1 and 2



(b) Comparison of activations $B^{(i)}$ of the features $F^{(i)}$ from the fluorescence tensor using both Algorithms 1 and 2

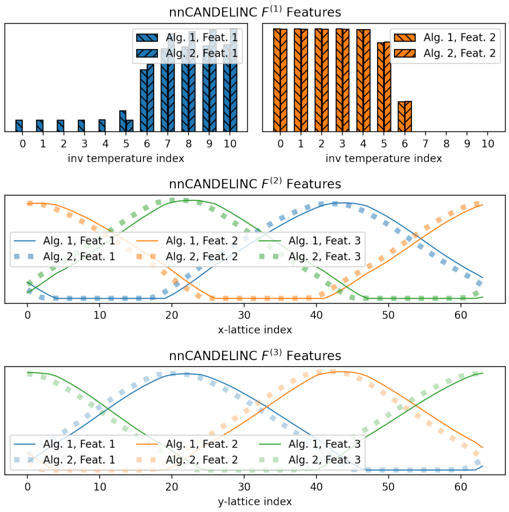
Figure 5: Comparisons of the features, $F^{(i)}$, and mixings, $B^{(i)}$, obtained from nnCANDELINC Algorithms 1 and 2 on the experimental fluorescence data of size $(5 \times 61 \times 201)$ with nonnegative multiranks $[3, 3, 3]$, and nonnegative rank r=3.

to the fluorescence data-tensor, than the measured fluorescence, i.e., the output, $\mathcal{X}$, is three-linear, and its components, $\mathcal{X}_{i,j,k}$ are,

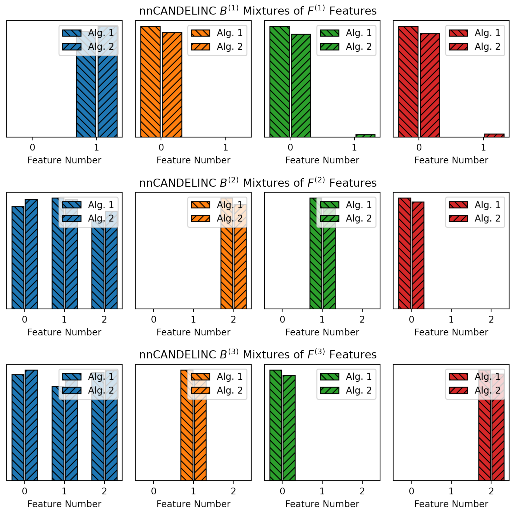
$$\mathcal{X}_{i,j,k} = \sum_{n=1}^{r} A^{(1)}_{i,n}\, A^{(2)}_{j,n}\, A^{(3)}_{k,n} + \epsilon_{i,j,k}.$$

Here, $A^{(1)}_{i,n}$ is linearly related to the concentration of the n^{th} fluorophore dissolved in the i^{th} sample; $A^{(2)}_{j,n}$ to the relative emission of n^{th} fluorophore at wavelength λ_j; $A^{(3)}_{k,n}$ to the relative amount of UV light absorbed by n^{th} fluorophore at excitation λ_k, and $\epsilon_{i,j,k}$ denotes the error. Although the above formula represents an ideal physical situation, it has been shown that for small concentrations of amino acids it is a valid approximation [34]. Here, we apply the nnCANDELINC Algorithms described in the previous sections, compare their results, and validate that the final decompositions coincides with the previously well-known results.

The scaled and appropriately permuted results presented in Figure 5 show minor differences between the resulting decompositions obtained from Algorithm 1 with 2.79% relative error, and Algorithm 2 with 2.51%. Figure 5a depicts the features extracted along the sample, emission, and excitation axes. We see virtually no difference in the sample and emission extracted features, with only slight deviations occurring in the excitation features. Similarly in Figure 5b the mixtures of the sample and emission features are virtually identical, while there are slight deviations in the mixtures of the excitation features between the two algorithms. A comparison with previously extracted features from the same data [34] confirms that both nnCANDELINC algorithms are producing correct results. It is also worth mentioning that the utilization of the nonnegative TD in Algorithm 1 does not results in a superdiagonal core-tensor $\mathcal{G}$, and the products of the final factors of both



(a) Comparison of features $F^{(i)}$ as extracted from the copoloymers tensor using both Algorithms 1 and 2



(b) Comparison of activations $B^{(i)}$ of the features $F^{(i)}$ from the copolymers tensor using both Algorithms 1 and 2

Figure 6: Performances of nnCANDELINC Algorithms 1 and 2 on computer generated data with size $(11 \times 64 \times 64)$, nonnegative multiranks $[2, 3, 3]$ and rank $r = 4$, representing phase separation with temperature in a system of copolymers.

nnCANDELINC algorithms are indistinguishable from those obtained by a direct application of CPD [34].

4.5. Decomposition of data generated by physics-based computer simulations

Here, we use nnCANDELINC to analyze a $3D$ data-tensor describing phase separation in a system of blocks copolymers whose evolution with temperature has been introduced and analyzed in a previous work [35]. We chose this system because of the natural nonnegativity of the data, the already known nonnegative rank, $r = 4$, and the fact that the extracted factors have a rank deficiency demonstrated in the previous analysis.

The multivariate function describing the phase separation is the order parameter of the system, $\Delta(T, f_A, x, y)$, which in this case is a function of: (a) temperature, T, (b) length f_A of the A-type blocks, and (c) the spatial coordinates, (x, y), of the 2-dimensional 64×64 lattice-space of the system. The order parameter, $\Delta(T, f_A, x, y)$, is simply the spatial density of the A-type blocks on the lattice, and therefore the data is inherently nonnegative. For A-type blocks with a fixed length, f_A, the order parameter is represented by 3-dimensional data: $\Delta(T, f_A, x, y) \equiv \Delta(T, x, y)$, and the tensor $\Delta_{n,m,l}$ that we analyze here has size $11 \times 64 \times 64$.

The nonnegative ranks r_i of each unfolding of the tensor $\Delta(T, x, y)$ has been previously estimated [35] and the nonnegative multirank has been determined to be, $\mu\text{rank}_+(\mathcal{X}) = [2, 3, 3]$. With this nonnegative minimal multirank we applied both nnCANDELINC algorithms to $\mathcal{X}$ and compare the results.

In Figure 6 we present the components of factors $F^{(i)}$ and $B^{(i)}$ from both Algorithm 1 with a relative error of 11.29%, and Algorithm 2 with an error of 10.02%. In Figure 6a the extracted features from the algorithms vary slightly, along the temperature axis we see relative shifts between the feature extracted by the algorithms, with similar shifts seen in the x-lattice and y-lattice axes. These shifts result in slight differences of the mixing of these features seen in Figure 6b. The nonnegative rank deficiencies become clear with the found combinations of features to represent the four rank one tensors needed for an nnCPD.

Appendix A. Notation and Operations

Here we list the precise operations used throughout the paper. A useful operation often used is the multiplication of a tensor by a matrix along a specific dimension, or *n-mode multiplication*.

Definition Appendix A.1. *The* 1-mode multiplication *between a tensor* $\mathcal{X} \in \mathbb{R}^{N_1 \times N_2 \times N_3}$ *and a matrix* $A \in \mathbb{R}^{M \times N_1}$ *is defined as*

$$(\mathcal{X} \times_1 A)_{i,j,k} = \sum_{l=1}^{N_1} \mathcal{X}_{l,j,k} A_{i,l}.$$

We define the 2*-mode and* 3*-mode multiplication analogously.*

For $i \neq j$ mode multiplications are commutative: $(\mathcal{X} \times_i A^{(i)}) \times_j A^{(j)} = (\mathcal{X} \times_j A^{(j)}) \times_i A^{(i)}$, and a matrix multiplication can be distributed through mode multiplication: $\mathcal{X} \times_i AB = (\mathcal{X} \times_i B) \times_i A$.

Definition Appendix A.2. *A* mode-i tensor fiber *of* $\mathcal{X}$ *is a one dimensional vector obtained by fixing all but the* i^{th} *index in the tensor. We let* $\mathcal{X}_{:,n,m}$*,* $\mathcal{X}_{n,:,m}$*,* $\mathcal{X}_{n,m,:}$ *denote the* n^{th}, m^{th} *mode-*1*, mode-*2 *and mode-*3 *tensor fibers, respectively. For* $i = 1, 2, 3$*,* $\text{unfold}_i(\mathcal{X})$ *denotes an* i-mode unfolding*, which rearranges all the mode-i fibers of a tensor into columns of a* $N_i \times N_j N_k$ *matrix, for* $i \neq j \neq k$*.*

Definition Appendix A.3. *For* $i = 1, 2, 3$*,* $\text{unfold}_i(\mathcal{X})$ *denotes an* i-mode unfolding*, which rearranges all the mode-i fibers of a tensor into columns of a* $N_i \times N_j N_k$ *matrix, for* $i \neq j \neq k$*.*

Each unfolding has an inverse mapping, which rearranges the columns of a matrix as fibers of a tensor. Consider the tensor $\mathcal{X} = \mathcal{Y} \times_1 A^{(1)} \times_2 A^{(2)} \times_3 A^{(3)}$. A particularly useful relation between unfoldings and mode multiplications is

$$\text{unfold}_i(\mathcal{X}) = A^{(i)}\, \text{unfold}_i(\mathcal{Y} \times_j A^{(j)} \times_k A^{(k)}), \tag{A.1}$$

for $i \neq j \neq k$.

Appendix B. Basics of NMF

Nonnegative matrix factorization (NMF) decomposes a nonnegative matrix $V \in \mathbb{R}_+^{N \times M}$, into a product of two nonnegative matrices $W \in \mathbb{R}_+^{N \times r}$ and $H \in \mathbb{R}^{r \times M}$. The geometric interpretation of a nonnegative decomposition, $V = WH$, is that that each column of V is a conic combination of the columns of W. With this geometric interpretation, computing an NMF is identical to searching for a polyhedral cone C which contains the columns of V, and is contained in the nonnegative orthant, $V \subset C \subset \mathbb{R}_+^N$. Of particular interest are cones with a minimum number of extreme rays, which correspond to the nonnegative rank.

Definition Appendix B.1. *The nonnegative rank of a matrix is defined as*

$$\mathit{rank}_+(V) := \min \left\{ r \;\middle|\; V = \sum_{n=1}^{r} w_n \otimes h_n,\, w_n \geq 0,\, h_n \geq 0 \right\}.$$

If $\mathrm{rank}_+(V) = r$ then there is a set of r nonnegative extreme rays $\{w_1, \dots, w_r\}$ such that every column of V is a conic combination of these extreme rays. When $\{w_1, \dots, w_r\}$ are assembled into the nonnegative matrix W, and the conic combinations are specified by a nonnegative matrix H, this corresponds to the nonnegative matrix factorization $V = WH$.

The nonnegative rank of a matrix has several well-known properties. For example, if V is an $(N_1 \times N_2)$-sized matrix, then $\mathrm{rank}(V) \leq \mathrm{rank}_+(V) \leq \min(N_1, N_2)$ [37]. A case illustrating the inequality between rank and nonnegative rank can be seen in the following Example Appendix B.2, which is mentioned in [37] as a private communication from H. Robbins.

Example Appendix B.2. *Consider the nonnegative matrix:*

$$V = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \tag{B.1}$$

and note that $v_1 + v_4 = v_2 + v_3$ where v_i is the i^{th} column of V. This linear dependence between the columns proves that rank$(V) = 3$. Also it was proved in [37] that the rank$_+(V) = 4$. This example demonstrates a case when rank$(V) <$ rank$_+(V)$. □

In general, computing the nonnegative rank of a nonnegative matrix $V \in \mathbb{R}^{N_1 \times N_2}$ is an NP-hard problem [15], and even providing a reliable estimate can be quite hard.

Acknowledgments

This work was supported by the LDRD program of Los Alamos National Laboratory under project number 20190020DR. Los Alamos National Laboratory is operated by Triad National Security, LLC, for the National Nuclear Security Administration of U.S. Department of Energy (Contract No. 89233218CNA000001). We would also like to thank the anonymous reviewer for their helpful comments and suggestions. This study does not have any conflicts to disclose.

References

[1] Hickey A. 2019. *Zettabytes of data hog up space and resources.*

[2] Franke B., Plante J.F., Roscher R., Lee E.A., Smyth C., Hatefi A., et al. Statistical inference, learning and models in big data. International Statistical Review. 2016;**84**(3):371–389.

[3] Kolda T., and Bader B. Tensor decompositions and applications. SIAM Review. 2009;**51**(3):455–500.

[4] Cichocki A., Phan A.H., Zhao Q., Lee N., Oseledets I., Sugiyama M., et al. Tensor networks for dimensionality reduction and large-scale optimization: Part 2 applications and future perspectives. Foundations and Trends® in Machine Learning. 2017;**9**(6):431–673.

[5] Oseledets I.V. Tensor-train decomposition. SIAM Journal on Scientific Computing. 2011;**33**(5):2295–2317.

[6] Vervliet N. 2018. *Compressed sensing approaches to large-scale tensor decompositions.* . KU Leuven. (PhD Thesis).

[7] Tucker L.R. Some mathematical notes on three-mode factor analysis. Psychometrika. 1966;**31**(3):279–311.

[8] Hackbusch W. Tensor spaces and numerical tensor calculus. vol. 42 of Springer series in computational mathematics. Heidelberg: Springer; 2012.

[9] Hitchcock F.L. The expression of a tensor or a polyadic as a sum of products. Journal of Mathematics and Physics. 1927;**6**:164–189.

[10] Harshman R.A. Foundation of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis. In: UCLA Working Papers in Phonetics. vol. 16. University Microfilms, Ann Arbor, Michigan, No. 10,085; 1970. p. 1–84.

[11] Håstad J. Tensor rank is NP-Complete. In: International Colloquium on Automata, Languages, and Programming. Springer; 1989. p. 451–460.

[12] De Silva V., and Lim L.H. Tensor rank and the ill-posedness of the best low-rank approximation problem. SIAM Journal on Matrix Analysis and Applications. 2008;**30**(3):1084–1127.

[13] Everett B. An introduction to latent variable models. Springer Science & Business Media; 2013.

[14] Lee D.D, and Seung H.S. Learning the parts of objects by non-negative matrix factorization. Nature. 1999;**401**(6755):788.

[15] Cichocki A., Zdunek R., Phan A.H., and Amari Si. Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation. Wiley Publishing; 2009.

[16] Carroll J.D., Pruzansky S., and Kruskal J.B. CANDELINC: A general approach to multidimensional analysis of many-way arrays with linear constraints on parameters. Psychometrika. 1980;**45**(1):3–24.

[17] Bro R., and Andersson C.A. Improving the speed of multiway algorithms: Part II: Compression. Chemometrics and intelligent laboratory systems. 1998;**42**(1-2):105–113.

[18] Bro R., Harshman R.A., Sidiropoulos N.D., and Lundy M.E. Modeling multi-way data with linearly dependent loadings. Journal of Chemometrics: A Journal of the Chemometrics Society. 2009;**23**(7-8):324–340.

[19] Wei W., Le X., Xiaofei Z., and Jianfeng L. Compressed Sensing PARALIND Decomposition-Based Coherent Angle Estimation for Wireless Communications and Mobile Computing. 2019;**2019**.

[20] Cohen J. 2016. *Environmental multiway data mining.* . Université de Grenobles Alpes, France. (Ph.D. Thesis).

[21] Qi Y. A Very Brief Introduction to Nonnegative Tensors from the Geometric Viewpoint. Mathematics. 2018;**6**(11):230.

[22] Qi Y., Comon P., and Lim L.H. Semialgebraic geometry of nonegative tensor rank. SIAM Journal of Matrix Analysis and Applications. 2016;**37**(4):1556–1580.

[23] Landsberg J.M. Tensors: geometry and applications. Representation theory. 2012;**381**(402):3.

[24] Comon P. Tensors: a brief introduction. IEEE Signal Processing Magazine. 2014;**31**(3):44–53.

[25] Bertsimas D., and Tsitsiklis J.N. Introduction to linear optimization. vol. 6. Athena Scientific Belmont, MA; 1997.

[26] Gillis N. Sparse and unique nonnegative matrix factorization through data preprocessing. Journal of Machine Learning Research. 2012;**13**(Nov):3349–3386.

[27] Kruskal J.B. Three-ways arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. Linear Algebra and Applications. 1977;**18**(2):95–138.

[28] Domanov I., and De Lathauwer L. On the uniqueness of the canonical polyadic decomposition of third-order tensors — Part I: Basic results and uniqueness of one factor matrix. SIAM Journal on Matrix Analysis and Applications. 2013;**34**(3):855–875.

[29] Cohen J.E., Comon P., and Gillis N. Some theory on non-negative Tucker decomposition. In: International Conference on Latent Variable Analysis and Signal Separation. Springer; 2017. p. 152–161.

[30] Kossaifi J., Panagakis Y., Anandkumar A., and Pantic M. Tensorly: Tensor learning in python. The Journal of Machine Learning Research. 2019;**20**(1):925–930.

[31] Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research. 2011;**12**:2825–2830.

[32] Tomasi G., and Bro R. A comparison of algorithms for fitting the PARAFAC model. Computational Statistics & Data Analysis. 2006;**50**(7):1700–1734.

[33] Battaglino C., Ballard G., and Kolda T.G. A practical randomized CP tensor decomposition. SIAM Journal on Matrix Analysis and Applications. 2018;**39**(2):876–901.

[34] Bro R. 1998. *Multi-way analysis in the food industry. Models, algorithms, and applications.* . Royal Veterinary and Agricultural University, Denmark. (Ph.D. Thesis).

[35] Alexandrov B.S., Stanev V.G., Vesselinov V.V., and Rasmussen K.Ø. Nonnegative tensor decomposition with custom clustering for microphase separation of block copolymers. Statistical Analysis and Data Mining: The ASA Data Science Journal. 2019;.

[36] Kruskal J.B. Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. Linear algebra and its applications. 1977;**18**(2):95–138.

[37] Cohen J.E., Rothblum U.G., and RUTCOR–Rutgers Center for Operational Research. Nonnegative ranks, decompositions and factorizations of nonnegative matrices. Linear Algebra and its Applications. 1993;**190**:149–168.