Some theory on Non-negative Tucker Decomposition

Jérémy Cohen, Pierre Comon, Nicolas Gillis

Some theory on Non-negative Tucker Decomposition

Jeremy E. Cohen*, Pierre Comon, and Nicolas Gillis

Departement of Mathematics and Operational Research,
Rue de Houdain 9, Faculté polytechnique, Université de Mons
Gipsa-lab,
11 rue des mathematiques, 38210 St Martin D'Hères, France
{jeremy.cohen}@umons.ac.be

Abstract. Some theoretical difficulties that arise from dimensionality reduction for tensors with non-negative coefficients is discussed in this paper. A necessary and sufficient condition is derived for a low non-negative rank tensor to admit a non-negative Tucker decomposition with a core of the same non-negative rank. Moreover, we provide evidence that the only algorithm operating mode-wise, minimizing the dimensions of the features spaces, and that can guarantee the non-negative core to have low non-negative rank requires identifying on each mode a cone with possibly a very large number of extreme rays. To illustrate our observations, some existing algorithms that compute the non-negative Tucker decomposition are described and tested on synthetic data.

Keywords: Non-negative Tucker Decomposition, Non-negative Canonical Polyadic Decomposition, dimensionality reduction, Non-negative Matrix Factorization

Notation

The following notation will be used: bold calligraphic letters $\boldsymbol{\mathcal{T}}$ for tensors, bold uppercase letters $\boldsymbol{U}$ for matrices or linear operators, and bold lowercase letters $\boldsymbol{a}$ for vectors. Here tensors are real-valued vectors in $\mathbb{R}^K \otimes \mathbb{R}^L \otimes \mathbb{R}^M$ or multilinear operators in $\mathbb{R}^{K\times R_1} \otimes \mathbb{R}^{L\times R_2} \otimes \mathbb{R}^{M\times R_3}$ with K, L, M, R_i integers and the product $\otimes$ is a tensor product [1], which implies $\lambda\boldsymbol{x} \otimes \boldsymbol{y} = \boldsymbol{x} \otimes \lambda\boldsymbol{y} = \lambda(\boldsymbol{x} \otimes \boldsymbol{y})$. Rank-one linear operators acting on tensors are denoted as $\boldsymbol{U} \otimes \boldsymbol{V} \otimes \boldsymbol{W}$, where the tensor product is the canonical tensor product for linear applications inherited from the tensor product of vectors, and by definition, $(\boldsymbol{U} \otimes \boldsymbol{V} \otimes \boldsymbol{W})(\boldsymbol{U}_2 \otimes \boldsymbol{V}_2 \otimes \boldsymbol{W}_2) = \boldsymbol{U}\boldsymbol{U}_2 \otimes \boldsymbol{V}\boldsymbol{V}_2 \otimes \boldsymbol{W}\boldsymbol{W}_2$. Also, for two-way arrays, $(\boldsymbol{U} \otimes \boldsymbol{V})\boldsymbol{T} = \boldsymbol{U}\boldsymbol{T}\boldsymbol{V}^t$. The Kronecker product [2] is denoted by $\boxtimes$ and is one possible expression of a tensor product in $\mathbb{R}^{KLM}$. Further discussion on notations can be found in [3].

* Research funded by ERC advanced grant "DECODA" no. 320594, ERC starting grant"COLORAMAP" no. 679515, and F.R.S.-FNRS incentive grant for scientific research n° F.4501.16.

1 Tensor decomposition models

In this section we quickly survey two tensor decomposition models, namely the Tucker Decomposition (TD) and the Canonial Polyadic Decomposition (CPD) [4].

1.1 Tucker Decompositions

Given a tensor $\boldsymbol{\mathcal{T}} \in \mathbb{R}^K \otimes \mathbb{R}^L \otimes \mathbb{R}^M$, the TD finds so-called factor matrices $\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W}$ of respective sizes $K \times R_1$, $L \times R_2$ and $M \times R_3$ defining bases onto which the tensor can be expressed mode-wise:

$$\boldsymbol{\mathcal{T}} = (\boldsymbol{U} \otimes \boldsymbol{V} \otimes \boldsymbol{W})\, \boldsymbol{\mathcal{G}}, \tag{1}$$

where $\boldsymbol{\mathcal{G}}$ is a tensor of coefficient often called the core of the TD. In other words, the span of $\boldsymbol{U}$ contains all the columns in $\boldsymbol{\mathcal{T}}$. This is interesting for dimensionality reduction if R_1 is strictly smaller than K. The same observation holds for the two other modes. TD has been first investigated by Hitchcock in 1927 [4], and is now a widely used data mining model [5]. The main drawback is that there are infinitely many solution to decompose $\boldsymbol{\mathcal{T}}$, so that it may not be possible to recover the ground truth for $\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W}, \boldsymbol{\mathcal{G}}$ from the data $\boldsymbol{\mathcal{T}}$ solely.

Similarly to the matrix factorization problem, in the hope to restore identifiability of the parameters, the Non-negative Tucker Decomposition (NTD) was introduced recently [6]:

$$\begin{cases} \boldsymbol{\mathcal{T}} = (\boldsymbol{U} \otimes \boldsymbol{V} \otimes \boldsymbol{W})\, \boldsymbol{\mathcal{G}}, \\ \boldsymbol{\mathcal{G}} \geq 0,\ \boldsymbol{U} \geq 0,\ \boldsymbol{V} \geq 0,\ \boldsymbol{W} \geq 0, \end{cases} \tag{2}$$

but NTD was later shown to be unique up to permutation and scaling ambiguities if and only if NMF of each unfolding is unique, which is a very strong assumption and may not be verified in practice [7]. Imposing non-negativity constraints however improves the interpretability of the results of the Tucker Decomposition in some applications, see for instance [7] for an application in neuroscience.

Again to reduce the set of solutions to (2), a Sparse Non-negative Tucker Decomposition (SNTD) was suggested by Morup *et. al.* [6] in which the factors matrices and the core are also constrained to be sparse. As we will show below, imposing sparsity on the factors may not be sufficient to restore identifiability.

Note that other constraints have been imposed on the factors of TD in the literature, notably orthogonality constraints and slice-orthogonality on the core [8].

1.2 Canonical Polyadic Decomposition

Maybe the most widely used tensor decomposition model is the Canonical Polyadic Decomposition (CPD) also called PARAFAC. It is similar to TD in the sense that a basis is sought on each mode, but in CPD the core is required to be diagonal, which makes CPD a much more constrained model than TD:

$$\boldsymbol{\mathcal{T}} = (\boldsymbol{A} \otimes \boldsymbol{B} \otimes \boldsymbol{C})\, \boldsymbol{\mathcal{I}}_R, \tag{3}$$

where $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ are respectively of sizes $K \times R$, $L \times R$ and $L \times R$ and R is the minimal integer so that (3) holds. CPD is often unique (up to permutations and scaling ambiguities) under mild conditions on the factors often verified in practice. A very common assumption is that R is much smaller than the dimensions of the data, in which case $\boldsymbol{\mathcal{T}}$ is said to be a low rank tensor. CPD has been used in many applications ranging from chemometrics to social sciences [9].

In those applications, it often makes sense to look for non-negative factors. The Non-negative CPD (NCPD) [10] can then be used instead as a decomposition model:

$$\begin{cases} \boldsymbol{\mathcal{T}} = (\boldsymbol{A} \otimes \boldsymbol{B} \otimes \boldsymbol{C})\boldsymbol{\mathcal{I}}_R, \\ \boldsymbol{A} \geq 0, \ \boldsymbol{B} \geq 0, \ \boldsymbol{C} \geq 0, \end{cases} \tag{4}$$

where R is now called the non-negative rank of $\boldsymbol{\mathcal{T}}$ if it is the smallest integer so that (4) holds. It is denoted $\text{rank}_+(\boldsymbol{\mathcal{T}})$.

2 Propagating non-negativity and non-negative rank through NTD

In the following, we show that NTD may not propagate the low non-negative rank of the original tensor $\boldsymbol{\mathcal{T}}$, and that to ensure $\boldsymbol{\mathcal{G}}$ has the same non-negative rank as $\boldsymbol{\mathcal{T}}$, it is sufficient to identify the rays of a particular cone. We also show that no mode-wise procedure with $R_1 = R_2 = R_3 = R$ can guarantee non-negative rank propagation.

2.1 Elements of cone theory

First let us define some basic tools of cone theory that we shall use later in this section, most of which can be found in [11]. We start with a possible definition of the cone generated by columns of a matrix $\boldsymbol{U}$:

Definition 1. *The cone generated by the columns of a matrix* $\boldsymbol{U} \in \mathbb{R}^{K \times R_1}$ *is the set* $\text{cone}(\boldsymbol{U}) = \{\boldsymbol{U}x, \ x \in \mathbb{R}_+^{R_1}\}$.

Another important notion is the extreme rays of a cone, intuitively the generating set of all elements in the cone:

Definition 2. *A vector* y *in* $\text{cone}(\boldsymbol{U})$ *spans an extreme ray if there does not exists* $x, z \in \text{cone}(\boldsymbol{U}) \backslash \text{cone}(y)$ *such that* $y = x + z$.

Moreover, a cone is said to be simplicial if and only if all the extreme rays are linearly independent. Clearly, given a full column rank matrix $\boldsymbol{U}$ in $\mathbb{R}^{K \times R_1}$ with R_1 strictly smaller than K, then $\text{cone}(\boldsymbol{U})$ is simplicial and the columns of $\boldsymbol{U}$ are the extreme rays.

A set of interest for what follows is $\mathcal{H}(\boldsymbol{U}) = \text{span}(U) \cap \mathbb{R}_+^K$, namely the intersection of the non-negative orthant with the span of the columns of matrix $\boldsymbol{U} \geq 0$. It can be seen that $\mathcal{H}(\boldsymbol{U})$ is a cone [12], and its number of extreme rays is between R_1 and $O(C_K^{R_1})$ [11] (the upper bound is attained by cones whose slices are cyclic polytopes with many vertices). This means that $\mathcal{H}(\boldsymbol{U})$ may be a cone with a very large number of extreme rays. Note however that $\mathcal{H}(\boldsymbol{U}) \subset \text{cone}(\boldsymbol{I})$ which has K rays and corresponds to a trivial factorization ($\boldsymbol{U} = \boldsymbol{I}\boldsymbol{U}$).

2.2 Working hypotheses

In this paper we wish to explore the properties of the NTD. In particular, we found the case of a low non-negative rank tensor $\mathcal{T}$ of particular interest; see below. These results are meant as a first step in the understanding of NTD so we allow ourselves to make restrictive hypotheses. Note however that these hypotheses are often verified in real-life applications, for instance in fluorescence spectroscopy or neuroimaging.

Here are the working hypotheses that we need in order to establish the results presented in the remainder of this section:

- **H1**: $\mathcal{T}$ is non-negative, i.e. all entries of $\mathcal{T}$ are greater or equal to 0.
- **H2**: $\mathcal{T}$ admits a unique NCPD with factors $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$. [1]

All three hypotheses are required for results presented in subsection 2.3 to hold, but only **H1** is used in section 2.4.

2.3 Propagating the non-negative rank to the core

Our goal in this subsection is to study the propagation of the non-negative rank of $\mathcal{T}$ to the core $\mathcal{G}$ in (2). A property enjoyed by Tucker Decomposition is that the rank of $\mathcal{T}$ and the rank of $\mathcal{G}$ are always equal in the exact setting provided factors $\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W}$ admit left inverses. This may not be the case however for the non-negative rank and the NTD. First, we give a necessary and sufficient condition for the two non-negative ranks to match:

Proposition 1. [2] *Let $\mathcal{T}$ be a $K \times L \times M$ non-negative tensor of non-negative rank R satisfying **H1,H2**. Let $\mathcal{T} = (\boldsymbol{U} \otimes \boldsymbol{V} \otimes \boldsymbol{W})\,\mathcal{G}$ be a NTD with $\mathcal{G}$ of size $R_1 \times R_2 \times R_3$ so that $\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W}$ admit left inverses. Then $R = rank_+(\mathcal{G})$ if and only if $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ belongs respectively to cone($\boldsymbol{U}$), cone($\boldsymbol{V}$) and cone($\boldsymbol{W}$). Moreover, if $\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W}$ do not admit left inverses, then there exists a core $\mathcal{G}'$ of non-negative rank R such that $\mathcal{T} = (\boldsymbol{U} \otimes \boldsymbol{V} \otimes \boldsymbol{W})\,\mathcal{G}'$, where $\mathcal{G} - \mathcal{G}'$ belongs to the null space of $\boldsymbol{U} \otimes \boldsymbol{V} \otimes \boldsymbol{W}$.*

Proof. First suppose that $\text{rank}_+(\mathcal{G}) = R$. Then there exists $\boldsymbol{A}_c, \boldsymbol{B}_c, \boldsymbol{C}_c$ such that:

$$\mathcal{T} = (\boldsymbol{U} \otimes \boldsymbol{V} \otimes \boldsymbol{W})\,(\boldsymbol{A}_c \otimes \boldsymbol{B}_c \otimes \boldsymbol{C}_c)\,\mathcal{I}_R \tag{5}$$

so that $\mathcal{T}$ admits a NCPD with factors $\boldsymbol{U}\boldsymbol{A}_c, \boldsymbol{V}\boldsymbol{B}_c, \boldsymbol{W}\boldsymbol{C}_c$. Because the NCPD of $\mathcal{T}$ is unique, we can conclude that $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{A}_c$ and similarly on the other modes. Conversely, first note that Equation (5) after developping shows that $\text{rank}_+(\mathcal{G}) \geq R$. Moreover, because $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ belong to the cones spanned by

[1] ERRATUM 2020/03/31: Hypothesis 3, stating that the factors needed to be full column rank, was unnecessary and has been removed. Also, Hypothesis 2 required the tensor rank to be smaller than the dimensions, which is also unnecessary and has been removed. Thanks to Erik Skao and Derek DeSantis for spotting this mistake.

[2] ERRATUM 2020/03/31: The proposition has been corrected, so that the statement and the proof still hold even with weaker assumptions H1 and H2.

$\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W}$, there exist $\boldsymbol{A}_c, \boldsymbol{B}_c, \boldsymbol{C}_c$ non-negative $R_i \times R$ matrices so that $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{A}_c$, $\boldsymbol{B} = \boldsymbol{V}\boldsymbol{B}_c$ and $\boldsymbol{C} = \boldsymbol{W}\boldsymbol{C}_c$. These non-negative coefficient matrices are factors in a NCPD of $\boldsymbol{\mathcal{G}}$ because $\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W}$ are left invertible. From this, we get that $\text{rank}_+(\boldsymbol{\mathcal{G}}) \leq R$ which concludes the proof. If the factors of the NTD are not invertible, then simply set $\boldsymbol{\mathcal{G}}' = (\boldsymbol{A}_c \otimes \boldsymbol{B}_c \otimes \boldsymbol{C}_c)\boldsymbol{\mathcal{I}}_R$.

In the next section, some algorithms designed for NTD and SNTD will be tested to check whether this condition is verified or not in practice. But in a theoretical perspective, it is natural to wonder whether matrices $\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W}$ can be found solely from $\boldsymbol{\mathcal{T}}$ so that the necessary and sufficient condition from proposition 1 is always verified. Since this problem can be cast mode-wise, it is closely related to recent uniqueness results obtained for Non-negative Matrix Factorization [13]. In what follows, we study mode-wise approaches to this problem. These involve the unfoldings of the data tensor, which are the columns/rows/fibers stacked into matrices. Using the matricization suggested in [3], the unfoldings of $\boldsymbol{\mathcal{T}}$ can be expressed as follows:

$$\begin{aligned} \boldsymbol{T}_1 &= \boldsymbol{A}\left(\boldsymbol{B} \odot \boldsymbol{C}\right)^T = \boldsymbol{U}\boldsymbol{G}_1\left(\boldsymbol{V} \boxtimes \boldsymbol{W}\right)^T, \\ \boldsymbol{T}_2 &= \boldsymbol{B}\left(\boldsymbol{A} \odot \boldsymbol{C}\right)^T = \boldsymbol{V}\boldsymbol{G}_2\left(\boldsymbol{U} \boxtimes \boldsymbol{W}\right)^T, \\ \boldsymbol{T}_3 &= \boldsymbol{C}\left(\boldsymbol{A} \odot \boldsymbol{B}\right)^T = \boldsymbol{W}\boldsymbol{G}_3\left(\boldsymbol{U} \boxtimes \boldsymbol{V}\right)^T, \end{aligned} \tag{6}$$

where $\odot$ is the Khatri-Rao product, that is, the column-wise Kronecker product.

Now, how can we guarantee that, say on the first mode, $\text{cone}(\boldsymbol{U})$ contains $\boldsymbol{A}$? A first (non mode-wise) solution is to constrain the core to be diagonal and actually look for the NCPD instead of the NTD. In the following, we restrict our preliminary study to the case where $\text{span}(\boldsymbol{U}) = \text{span}(\boldsymbol{A})$. In that case it is possible to choose $\boldsymbol{U}$ as extreme rays of $\mathcal{H}(\boldsymbol{A})$. By definition, $\mathcal{H}(\boldsymbol{A})$ is the largest cone in the intersection of the non-negative orthant and the column space of $\boldsymbol{A}$ containing $\boldsymbol{T}_1$. It also contains $\boldsymbol{A}$ since $\boldsymbol{A}$ belongs to the non-negative orthant. This means that extreme rays $\boldsymbol{U}$ of $\mathcal{H}(\boldsymbol{A})$ can be used in the NTD to ensure that the non-negative rank is preserved using Proposition 1.

However finding the extreme rays of $\mathcal{H}(\boldsymbol{A})$ is likely to be of little interest in practice since the number of extreme rays needed can be larger than K. Yet a special and easy case is when the non-negative matrix factorization of each unfolding is unique, then any cone spanning the unfolding on one mode also spans the NCPD factor on that mode.

In the light of the previous paragraph, a more interesting question is the following: can we design a procedure to find a simplicial cone $\text{cone}(\boldsymbol{U})$ with $R_1 = R$ extreme rays (i.e. of order R) which always contains $\boldsymbol{A}$? If a solution to this problem is found, then in theory it would be possible to compress the non-negative tensor $\boldsymbol{\mathcal{T}}$ into $\boldsymbol{\mathcal{G}}$ and to only compute the NCPD on $\boldsymbol{\mathcal{G}}$.

Such a procedure needs to compute a maximal volume cone. Indeed, suppose the procedure outputs a set $\boldsymbol{U}$ of extreme rays, and suppose there exists a larger cone $\boldsymbol{U}'$ also enclosing $\boldsymbol{T}_1$, then because the only requirement for $\boldsymbol{A}$ in this problem is that $\boldsymbol{T}_1$ belongs to $\text{cone}(\boldsymbol{A})$, then possibly $\boldsymbol{U}' = \boldsymbol{A}$ and $\text{cone}(\boldsymbol{U})$ may not contain the columns of $\boldsymbol{A}$.

However, the largest simplicial one of order R may not be unique, and this provides a counter example to the idea that the largest cone of order R could always contain the columns of $\boldsymbol{A}$ (see Figure 1).

Sketch of a counter example. Let us build a matrix $\boldsymbol{A}$ in $\mathcal{R}^{4\times 3}$ and data $\boldsymbol{\mathcal{T}}_1$ so that there will be at least two largest cones $\mathcal{H}_3$ containing $\boldsymbol{T}_1$ with three extreme rays in the cone $\mathcal{H}$ defined by intersection of the span of $\boldsymbol{A}$ and the non-negative orthant. We set $\boldsymbol{A}^T = \begin{bmatrix} 1\,1\,0\,0 \\ 0\,1\,1\,0 \\ 0\,0\,1\,1 \end{bmatrix}$. Using theorem 9.1.1 from [12], we know that a non-zero vector from $\mathcal{H}$ belongs to an extreme ray if and only if it is as sparse as possible, and that for each set of 0 indices there is only one extreme ray. Here this can be applied in a straightforward manner, since in the span of $\boldsymbol{A}$ there can be no vector with three zeros. This means that the extreme rays can only have two zeros among 4 coefficients, and we thus need to check which combinations among the 6 belong to the span of $\boldsymbol{A}$ and the non-negative orthant. Since $\boldsymbol{A}$ has a simple structure, it is easy to check that $\mathcal{H}$ has 4 extreme rays, containing the columns of $\boldsymbol{A}$ and $e = [1, 0, 0, 1]^T$. Finally, the problem admits a rotational symmetry and it is easy to build $\boldsymbol{T}_1$ as a smaller cone contained in both cone($[\boldsymbol{a}_1, \boldsymbol{a}_2, e]$) and cone($A$), see Figure 1.

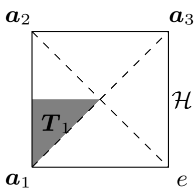

Fig. 1. A case where symmetry gives birth to two maximal volume cones with 3 extreme rays. The figure is the projection of the cones and data on the subspace $\{\boldsymbol{x} \in \mathbb{R}^4 | \sum_{i=1}^{4} x_i = 1\}$.

This maximal volume cone $\mathcal{H}_R$ of order R is actually what SNTD computes since SNTD imposes minimal ℓ_1 norm on the factors, meaning they should be as close as possible to the border of $\mathcal{H}$. Whether SNTD actually manages to compute cones containing the factors or not is investigated in the simulation section.

As a partial conclusion here, the only procedure that computes factors independently on each mode that can guarantee the propagation of the non-negative rank and under the constraint span($\boldsymbol{A}$) = span($\boldsymbol{U}$) is the computation of $\mathcal{H}(\boldsymbol{A})$. This provides evidence that using NTD as a preprocessing step for NCPD is difficult, but we cannot conclude that it is impossible since there may exist procedures working globally on the tensor (not mode-wise) or increasing the dimension of the column space of $\boldsymbol{U}$ that can guarantee non-negative rank propagation other than NCPD.

2.4 Propagating non-negativity to the core

A very fundamental question to answer for computing the NTD is whether some conditions can be imposed on the factor matrices $\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W}$ to ensure that $\boldsymbol{\mathcal{G}}$ has non-negative entries. At first glance, a natural condition to impose on the factors is that their cones contain the columns of the unfoldings of $\boldsymbol{\mathcal{T}}$. Clearly this condition is necessary, otherwise from (6), the product of the core and the Kronecker product of two factors has to contain negative entries, which itself is possible only if either the core or the factors contain negative entries.

However contrary to what is trivial for NMF, finding cones containing the columns of the unfoldings in each mode does not guarantee a non-negative core. We do not provide in this communication a simple counter-example, but we made this observation after running some numerical experiments reported in the next section, and this was confirmed by simulations run by the reviewers for this communication. This means that computing NMF on each unfolding to obtain $\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W}$ and infer the core by inverting a linear system may not yield a non-negative core.

3 Simulations

In this section we run numerical tests to support the previous theoretical discussion and provide evidence that neither NTD nor SNTD propagates non-negative rank in practice, and that computing NMF on each mode does not ensure obtaining a non-negative core.

3.1 Some algorithms for NTD and NMF

There has been a few algorithms reported in the literature to compute NTD. In the following simulations we make use of Hierarchical Alternating Least Squares (HALS) by Phan *et. al.* [14].

HALS is based on coordinate descent, where the set of variables is alternatively each columns of $\boldsymbol{U}$,$\boldsymbol{V}$ and $\boldsymbol{W}$, because with respect to these columns the underlying constrained least squares optimization problem admits a closed-form solution. For computing NMF, we used an algorithm based on the same idea from Gillis *et. al.* called accHALS for accelerated HALS [15]. Again, other algorithms exist for computing NMF and NTD that offer at least the same performances, but the goal here is not to compare state-of-the-art algorithms.

To compute SNTD, we used the algorithms from Morup *et. al.* which was the first algorithm designed for SNTD in the literature [6]. It is based on multiplicative updates, which are known to be slow for least square problems.

3.2 Some tests on the outputs of algorithms

Settings For both experiments, the tensors are rank $R = 3$ non-negative tensors build using the NCPD from factors drawn from the uniform distribution over $[0, 1]$, with sizes $K = L = M = 20$. The NCPD factors are normalized column-wise using the ℓ_2 norm. No noise is added. We set $R_1 = R_2 = R_3 = R$ - we will study the case where the number of components used in NTD are larger than the true rank in a longer communication. The maximal number of iterations is set to 1000 for the HALS algorithm, and to 3000 for the multiplicative algorithm solving SNTD - which we will abusively denote by SNTD. For accHALS applied on each unfolding, the maximal number of iterations is set to 1000. We chose the number of maximal iterations large enough so that convergence is always reached. In SNTD, the sparsity coefficient on the core is set to 0, and set to 10^{-3} on factors.

HALS and accHALS compute exact NTD and NMF up to around 10^{-8} relative error on the reconstructed tensor when no noise is added on the data and a good initialization is provided. We chose to initialize with High Order Singular Value Decomposition [8] to start in the right subspace on each mode. For SNTD, relative error with respect to the norm of the original data is of order of magnitude 10^{-4} in the following simulations.

Experiment 1: Number of negative entries in the core computed by mode-wise accHALS In this first experiment, NTD is computed using NMF on each unfolding of a hundred tensors. We plot the number of negative entries in $\boldsymbol{\mathcal{G}}$ obtained by an unconstrained linear system. We also plot the percentage of negative coefficients in $\boldsymbol{U}^\dagger \boldsymbol{T}_1$, $\boldsymbol{V}^\dagger \boldsymbol{T}_2$ and $\boldsymbol{W}^\dagger \boldsymbol{T}_3$, † denoting the left pseudo inverse. If the unfoldings are contained in the cones spanned by $\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W}$, then there should be no negative coefficients in these products.

Results reported in Figure 2 show that although the unfoldings are indeed almost contained in the cone of computed factor matrices, the core $\boldsymbol{\mathcal{G}}$ obtained contains a high number of negative entries. Moreover, the negative entries have a non-negligible intensity. This observation supports the idea that spanning the columns of the unfoldings is not a sufficient condition to ensure non-negativity of the core.

Experiment 2: Estimation of the span of factors and propagation of non-negative rank In this second experiment, 10 different tensors are decomposed using the NTD model using HALS and mode-wise accHALS and the SNTD model using the multiplicative algorithm. We check that the span of factors from the known NCPD and the computed NTD or SNTD are the same by comparing the norm of projected columns of $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ on the subspaces spanned by NTD factors. Moreover, we want to show that NCPD factors are not contained in the cones spanned by $\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W}$. The latter is checked by computing the amount of negative entries in products of the form $A_c = \boldsymbol{U}^\dagger \boldsymbol{A}$. Results are presented in Figure 3.

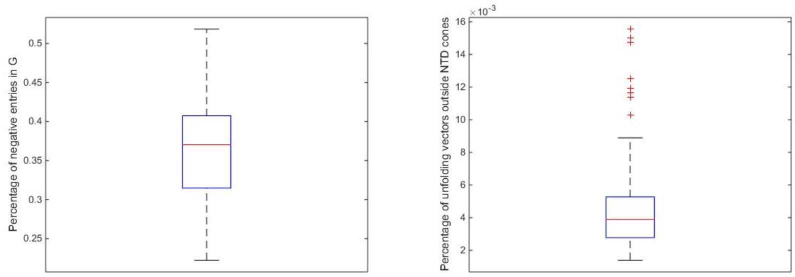

Fig. 2. Left: Percentage of negative entries in the core $\boldsymbol{\mathcal{G}}$ estimated by three NMF. Right: Percentage of negative values in the coefficients of vectors of unfoldings of $\boldsymbol{\mathcal{T}}$ in subspaces spanned by $\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W}$.

We observe that although the spans of factors from NCPD and NTD are the same (with a small variation for SNTD), the necessary and sufficient condition from Proposition 1 is not verified in this example. This means that neither NTD nor SNTD propagate the non-negative rank.

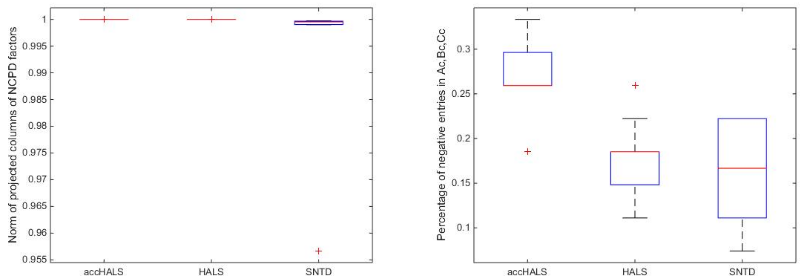

Fig. 3. Left: Average norm of the projected column of factors $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ onto the subspaces spanned by estimated $\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W}$. Right: Percentage of negative entries in the coefficients $\boldsymbol{A}_c, \boldsymbol{B}_c, \boldsymbol{C}_c$ so that $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{A}_c$ among all coefficients.

4 Conclusion

Non-negative Tucker Decomposition is a relatively unexplored research topic among constrained tensor decomposition models. We have shown in this paper that choosing the maximum volume cone generating the data does not necessarily restore identifiability of the factor matrices. We have also illustrated on some numerical experiments that choosing cones containing the unfoldings of the tensor on each mode does not necessarily yield a non-negative core, and that both algorithms computing the NTD and its sparse counterpart fail at preserving the low non-negative rank of the tensor, leaving little hope for designing a compression scheme based on NTD for large tensors with low non-negative rank. Such a procedure would require to choose non-trivial $\boldsymbol{U}$ so that $\text{cone}(\boldsymbol{U}) \supseteq \mathcal{H}(\boldsymbol{A})$, and similarly on the other modes.

5 Acknowledgements

The authors wish to thank the reviewers as well as the editor for very precious technical comments on a first version of this communication.

References

1. L. Schwartz, Y. Bamberger, and J.-P. Bourguignon, "Les tenseurs," 1977.
2. J. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Transactions on circuits and systems*, vol. 25, no. 9, pp. 772–781, 1978.
3. J. E.Cohen, "Environmental multiway data mining," Ph.D. dissertation, Universite Grenoble Alpes, 2016.
4. F. L. Hitchcock, "The expression of a tensor or a polyadic as a sum of products," *Journal of Mathematics and Physics*, vol. 6, no. 1, pp. 164–189, 1927.
5. E. E. Papalexakis, C. Faloutsos, and N. D. Sidiropoulos, "Tensors for data mining and data fusion: Models, applications, and scalable algorithms," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 8, no. 2, p. 16, 2016.
6. M. Mørup, L. K. Hansen, and S. M. Arnfred, "Algorithms for sparse nonnegative tucker decompositions," *Neural computation*, vol. 20, no. 8, pp. 2112–2131, 2008.
7. G. Zhou, A. Cichocki, Q. Zhao, and S. Xie, "Efficient nonnegative tucker decompositions: Algorithms and uniqueness," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 4990–5003, 2015.
8. L. De Lathauwer, B. De Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM journal on Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1253–1278, 2000.
9. E. Acar and B. Yener, "Unsupervised multiway data analysis: A literature survey," *IEEE transactions on knowledge and data engineering*, vol. 21, no. 1, pp. 6–20, 2009.
10. R. Bro, "Parafac. tutorial and applications," *Chemometrics and intelligent laboratory systems*, vol. 38, no. 2, pp. 149–171, 1997.
11. G. M. Ziegler, *Lectures on polytopes.* Springer Science & Business Media, 1995, vol. 152.
12. P. D'Alessandro, *Conical Approach to Linear Programming.* CRC Press, 1997.
13. H. Laurberg, M. G. Christensen, M. D. Plumbley, L. K. Hansen, and S. H. Jensen, "Theorems on positive data: On the uniqueness of nmf," *Computational intelligence and neuroscience*, vol. 2008, 2008.
14. A. H. Phan and A. Cichocki, "Extended hals algorithm for nonnegative tucker decomposition and its applications for multiway analysis and classification," *Neurocomputing*, vol. 74, no. 11, pp. 1956–1969, 2011.
15. N. Gillis and F. Glineur, "Accelerated multiplicative updates and hierarchical als algorithms for nonnegative matrix factorization," *Neural Computation*, vol. 24, no. 4, pp. 1085–1105, 2012.