

# Tensor and Its Tucker Core: the Invariance Relationships

Bo JIANG <sup>\*</sup>      Fan YANG <sup>†</sup>      Shuzhong ZHANG <sup>‡</sup>

November 8, 2016

## Abstract

In [13], Hillar and Lim famously demonstrated that “multilinear (tensor) analogues of many efficiently computable problems in numerical linear algebra are NP-hard”. Despite many recent advancements, the state-of-the-art methods for computing such ‘tensor analogues’ still suffer severely from the curse of dimensionality. In this paper we show that the Tucker core of a tensor however, retains many properties of the original tensor, including the CP rank, the border rank, the tensor Schatten quasi norms, and the Z-eigenvalues. When the core tensor is smaller than the original tensor, this property leads to considerable computational advantages as confirmed by our numerical experiments. In our analysis, we in fact work with a generalized Tucker-like decomposition that can accommodate any full column-rank factor matrices.

**Keywords:** Tucker decomposition, CP decomposition, border rank, tensor Schatten quasi norm, tensor eigenvalues.

**AMS subject classifications:** 15A69, 15A18, 15A03

---

<sup>\*</sup>Research Center for Management Science and Data Analytics, School of Information Management and Engineering, Shanghai University of Finance and Economics, Shanghai 200433, China. Email: isyebojiang@gmail.com. The research of this author was supported in part by National Natural Science Foundation of China (Grant 11401364) and Program for Innovative Research Team of Shanghai University of Finance and Economics.

<sup>†</sup>Institute for Computational & Mathematical Engineering, Stanford University, Stanford, CA 94305, USA. Email: fanfyang@stanford.edu.

<sup>‡</sup>Department of Industrial and Systems Engineering, University of Minnesota, Minneapolis, MN 55455, USA. Email: zhangs@umn.edu. The research of this author was supported in part by National Science Foundation (Grant CMMI-1462408).

# 1 Introduction

A tensor is a multidimensional extension of matrices, which has recently attracted a surge of research attention due to its wide applications in computer vision [39], psychometrics [12, 5], diffusion magnetic resonance imaging [10, 3, 34], quantum entanglement problem [14] and tensor-structured numerical methods for multi-dimensional PDEs [20, 19]. We refer the interested reader to the surveys [23, 21] on these subjects.

To study the spectral theory of tensors, various notions of tensor decompositions, eigenvalues and norms have been proposed. Unfortunately, unlike many of their matrix counter-parties, most tensor problems are computationally intractable [13]. Therefore, the numerical algorithms that aim to globally solve those problems are often time-consuming. For instance, the approach proposed in [8] to compute all Z-eigenvalues of a tensor is based on the so-called SOS (sum of squares) approach, which leads to a series of Semidefinite Programs with fast-growing sizes. Thus, it is naturally desirable that the same computational task would be performed on a tensor with smaller size. In this paper we establish that many of the aforementioned properties of a tensor carryover to its Tucker core, which is typically much smaller. To start off, let us introduce an extended notion of Tucker decomposition.

**Definition 1.** Consider an  $N$ -way tensor  $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$ . The equation

$$\mathcal{X} = \mathcal{G} \times_1 \mathbf{A}^{(1)} \times_2 \dots \times_N \mathbf{A}^{(N)} \triangleq \llbracket \mathcal{G}; \mathbf{A}^{(1)}, \dots, \mathbf{A}^{(N)} \rrbracket, \quad (1)$$

is called a size- $(J_1, \dots, J_N)$  Tucker decomposition of  $\mathcal{X}$ , and  $\mathcal{G} \in \mathbb{C}^{J_1 \times J_2 \times \dots \times J_N}$  is called a core tensor associated with this decomposition, and  $\mathbf{A}^{(n)} \in \mathbb{C}^{I_n \times J_n}$  is the  $n$ -th factor matrix for  $n = 1, \dots, N$ , where  $\times_n$  is the mode- $n$  (matrix) product. Moreover, a Tucker decomposition is said to be independent if each of the factor matrix has full column rank; a Tucker decomposition is said to be orthonormal if each of the factor matrix has orthonormal columns.

Note that the conventional Tucker decomposition corresponds to the orthonormal Tucker decomposition, which is also known as the higher-order SVD (HOSVD) in the literature. De Lathauwer, De Moor and Vandewalle [26] proposed an algorithm to compute such a decomposition. In a sequel, the same authors soon later proposed the higher-order orthogonal iteration (HOOI) in [27] to accommodate for inexact Tucker decomposition. In this paper, our analysis will be performed on the above-defined general independent Tucker decomposition unless specified otherwise.

The CP decomposition of a tensor is another important notion of tensor-decomposition, which leads to the definition of CP rank. The fact that the CP decomposition of core tensor is useful to decompose the original tensor itself has already been observed (see [4] and Section 5.3 of [23]). Consequently, the CP rank of a tensor equals that of its Tucker core follows from this observation.

Moreover, De Lathauwer et al. [26] showed that the Frobenius norm remains invariant for a given tensor and its core. However, those results are scattered in the literature, and often they are implicit. In this paper, we aim to establish the equivalence between a tensor and its Tucker core in a systematic fashion, including the CP rank and the Frobenius norm, and also other forms of tensor ranks, Z-eigenvalues and tensor Schatten quasi norms. In addition to tensor decompositions, the study of tensor eigenvalues became popular after the seminal papers of Qi [31] and Lim [28]. Furthermore, recently the tensor nuclear norm was used by Yuan and Zhang [41] in tensor completion to capture the low-rank structure; the regression bound obtained in [41] is better than that induced by the mode- $n$  matricization. Those properties allow us to propose the following scheme to compute the rank, the norms and the eigenvalues of a tensor, as long as these properties are invariant between the tensor and its Tucker core. As a first step, one computes Tucker decomposition of a given tensor. Then, the computations are performed on the smaller Tucker core. Finally, the computed quantity is transformed back to the original tensor. As we shall see later, the size of the Tucker core may be no more than 2 for some structured tensors, regardless the size of the original tensor. The savings on the computational time gained by following this scheme could be significant when the size of the core is considerably smaller compared to the original tensor, which is the case for many specific instances encountered in our numerical experiments.

The remainder of this paper is organized as follows. In the next section, we introduce the tensor notations and operations that we shall use in this paper. Then we discuss the invariance of the tensor ranks, the norms and the eigenvalues in Sections 3, 4 and 5 respectively. Section 6 discusses the implications of the invariance properties, the advantages of size reduction, and an error estimation of the new computational scheme. Finally, we apply our scheme to compute all the Z-eigenvalues of symmetric tensors. Our numerical results show that our strategy leads to a significant reduction in computational time on a set of testing instances.

## 2 Notations and preliminaries

Throughout this paper, we use the boldface lowercase letters, the capital letters, and the calligraphic letters to denote vectors, matrices, and tensors, respectively. For example, a vector (always a column vector unless otherwise stated)  $\mathbf{x}$ , a matrix  $\mathbf{A}$ , and a tensor  $\mathcal{X}$ . We use  $\|\cdot\|$  to denote the Euclidean norm of the vectors. For a matrix  $\mathbf{A}$ ,  $\sigma_{\max}(\mathbf{A})$  and  $\sigma_{\min}(\mathbf{A})$  denote the largest and smallest singular value of  $\mathbf{A}$ , while  $\|\mathbf{A}\|_2$  denotes its spectral norm:

$$\|\mathbf{A}\|_2 = \max_{\|\mathbf{x}\|=1} \|\mathbf{A}\mathbf{x}\| = \sigma_{\max}(\mathbf{A}).$$

We use lowercase subscripts to denote its components; e.g.  $x_i$  is the  $i$ -th entry of vector  $\mathbf{x}$ ,  $a_{ij}$  is the  $(i, j)$ -th entry of matrix  $\mathbf{A}$ , and  $x_{i_1 \dots i_n}$  is the  $(i_1, \dots, i_n)$ -th entry of  $n$ -th order tensor  $\mathcal{X}$ . Moreover,

we use the superscripts with bracket to refer a sequence of variables; e.g., a sequence of  $N$  matrices is denoted by  $\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \dots, \mathbf{A}^{(N)}$ .

The *mode- $n$  (matrix) product* denoted by “ $\times_n$ ” of a tensor  $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$  and a matrix  $\mathbf{A} \in \mathbb{C}^{J \times I_n}$  results in a tensor of size  $I_1 \times \dots \times I_{n-1} \times J \times I_{n+1} \times \dots \times I_N$  such that

$$[\mathcal{X} \times_n \mathbf{A}]_{i_1 \dots i_{n-1} j i_{n+1} \dots i_N} = \sum_{i_n=1}^{I_n} x_{i_1 \dots i_{n-1} i_n i_{n+1} \dots i_N} \cdot a_{j i_n}.$$

In the meanwhile, for a given tensor  $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$ , the *mode- $n$  matricization* denoted by  $\mathbf{X}_{(n)}$  is a mapping from tensor to matrix. In particular, the  $(i_1, i_2, \dots, i_N)$ -th entry of  $\mathcal{X}$  corresponds to the  $(i_n, j)$ -th entry of  $\mathbf{X}_{(n)}$ , where

$$j = 1 + \sum_{k=1, k \neq n}^N (i_k - 1) J_k, \quad J_k = \prod_{m=1, m \neq n}^{k-1} I_m.$$

Some properties relating the mode- $n$  product and mode- $n$  matricization are summarized in the following proposition, which will be used later. Interested reader is referred to [1] for more information on tensor multiplications.

**Proposition 1.** *For any  $N$ -way tensor  $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$  and matrix  $\mathbf{U} \in \mathbb{C}^{J \times I_n}$ , their mode- $n$  product satisfies*

$$(\mathcal{X} \times_n \mathbf{U})_{(n)} = \mathbf{U} \mathbf{X}_{(n)}.$$

Moreover, for any given matrices  $\mathbf{A} \in \mathbb{C}^{J_m \times I_m}$ ,  $\mathbf{B} \in \mathbb{C}^{J_l \times I_l}$ , if  $m \neq l$  then we have

$$(\mathcal{X} \times_m \mathbf{A}) \times_l \mathbf{B} = (\mathcal{X} \times_l \mathbf{B}) \times_m \mathbf{A};$$

if  $m = l$  and suppose the matrix multiplication is compatible, then we have

$$(\mathcal{X} \times_m \mathbf{A}) \times_m \mathbf{B} = \mathcal{X} \times_m (\mathbf{B} \mathbf{A}).$$

The *outer product* denoted by “ $\circ$ ” of two tensors  $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_{N_1}}$  and  $\mathcal{Y} \in \mathbb{C}^{I_{N_1+1} \times \dots \times I_{N_1+N_2}}$  is a tensor of size  $I_1 \times \dots \times I_{N_1+N_2}$  such that

$$[\mathcal{X} \circ \mathcal{Y}]_{i_1 \dots i_{N_1+N_2}} = x_{i_1 \dots i_{N_1}} \cdot y_{i_{N_1+1} \dots i_{N_1+N_2}}.$$

In particular,  $\mathcal{X}$  is a rank-1 tensor if it can be written as an outer products of vectors; e.g.

$$\mathcal{X} = \mathbf{a}^{(1)} \circ \mathbf{a}^{(2)} \circ \dots \circ \mathbf{a}^{(N)}.$$

The *inner product* of two tensors  $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$  and  $\mathcal{Y} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$  is denoted by

$$\langle \mathcal{X}, \mathcal{Y} \rangle = \sum_{i_1=1}^{I_1} \dots \sum_{i_N=1}^{I_N} x_{i_1 \dots i_N} \cdot y_{i_1 \dots i_N}.$$

For any tensor  $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$ , the *Frobenius norm* of tensor  $\mathcal{X}$  is defined as

$$\|\mathcal{X}\|_F \triangleq \sqrt{\langle \mathcal{X}, \mathcal{X} \rangle}.$$

More discussions on tensor operations can be found in [24]. The notion of tensor decomposition is central to the study of tensors. Let us now formally introduce the so-called CP decomposition and the CP rank, where ‘CP’ is a further abbreviation from CANDECOMP (canonical decomposition) by Carroll and Chang [5] and PARAFAC (parallel factorization) by Harshman [12] in early 1970’s.

**Definition 2.** For an  $N$ -way tensor  $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$ , a CP decomposition is a representation of  $\mathcal{X}$  as a sum of  $r$  rank-1 tensors,

$$\mathcal{X} = \sum_{t=1}^r \mathbf{a}^{(1,t)} \circ \dots \circ \mathbf{a}^{(N,t)},$$

where  $\mathbf{a}^{(n,t)} \in \mathbb{C}^{I_n}$ . The CP rank of  $\mathcal{X}$ , denoted by  $\text{rank}_{CP}(\mathcal{X})$ , is the minimum integer  $r$  such that a size- $r$  CP decomposition is possible.

Another important notion of tensor decomposition is the so-called Tucker decomposition proposed by Tucker [38], as introduced in Definition 1 (though in a slightly more general format). Likewise, this decomposition also leads to another notion of tensor rank.

**Definition 3.** For an  $N$ -way tensor  $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$ , its Tucker Rank, denoted by  $\text{rank}_T(\mathcal{X})$ , is an  $N$ -dimensional vector

$$\text{rank}_T(\mathcal{X}) = (\text{rank}(\mathbf{X}_{(1)}), \dots, \text{rank}(\mathbf{X}_{(N)})), \quad (2)$$

where  $\mathbf{X}_{(n)}$  is the mode- $n$  matricization of  $\mathcal{X}$  for  $n = 1, \dots, N$ , and  $\text{rank}(\cdot)$  denotes the regular matrix rank.

In fact, many of the operations we have discussed about can be represented by Tucker decomposition. For example, a rank-one tensor

$$\mathcal{X} = \mathbf{a}^{(1)} \circ \mathbf{a}^{(2)} \circ \dots \circ \mathbf{a}^{(N)} = \llbracket 1; \mathbf{a}^{(1)}, \mathbf{a}^{(2)}, \dots, \mathbf{a}^{(N)} \rrbracket$$

where 1 is the scalar one; mode- $n$  matrix product

$$\mathcal{X} \times_n \mathbf{A} = \llbracket \mathcal{X}; \mathbf{I}_{I_1}, \dots, \mathbf{I}_{I_{n-1}}, \mathbf{A}, \mathbf{I}_{I_{n+1}}, \dots, \mathbf{I}_{I_N} \rrbracket,$$

where  $\mathbf{I}_k$  is the unit matrix of dimension  $k \times k$ ; the inner product of a tensor  $\mathcal{G}$  and a symmetric rank-one tensor (which defines a polynomial):

$$\langle \mathcal{G}, \underbrace{\mathbf{x} \circ \dots \circ \mathbf{x}}_N \rangle = \llbracket \mathcal{G}; \underbrace{\mathbf{x}^T, \dots, \mathbf{x}^T}_N \rrbracket.$$

The CP decomposition can also be viewed as a special Tucker decomposition with

$$\boldsymbol{\mathcal{X}} = \sum_{t=1}^r \mathbf{a}^{(1,t)} \circ \dots \circ \mathbf{a}^{(N,t)} = \llbracket \boldsymbol{\mathcal{I}}; \mathbf{A}^{(1)}, \dots, \mathbf{A}^{(N)} \rrbracket,$$

where  $\mathbf{A}^{(n)} = [\mathbf{a}^{(n,1)}, \dots, \mathbf{a}^{(n,r)}]$  for each  $n = 1, \dots, N$ , and  $\boldsymbol{\mathcal{I}}$  is an  $N$ -way unit tensor, with all zero elements except the diagonal elements are ones.

### 3 The invariance of the tensor ranks

There are certain correspondence between the CP decomposition of a given tensor and that of its *orthonormal* Tucker core [19, 4, 20, 23]. In particular, performing Tucker decomposition and further CP decomposition on the Tucker core is called two-level rank decomposition in [20]. Approximating the original tensor by the two-level decomposition is discussed in [22, 19]. Possible computational savings gained by exploiting this relation were discussed in [37]. Moreover, as a direct consequence of this correspondence, the CP rank of a tensor equals to that of its orthonormal Tucker core, which is also known as the CANDELINC Theorem; see [6].<sup>1</sup>

In this section we aim to show that in fact several notions on the rank of a tensor carryover to that of its *independent* Tucker core. Besides, we provide a unified treatment on the independent Tucker decomposition in such a way that the technical results presented in this section will facilitate our analysis in later discussions.

#### 3.1 Invariance of the CP rank under independent Tucker decomposition

Let us start with the CP rank.

**Theorem 2.** *For any given tensor  $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$  with independent Tucker decomposition  $\boldsymbol{\mathcal{X}} = \llbracket \boldsymbol{\mathcal{G}}; \mathbf{X}^{(1)}, \dots, \mathbf{X}^{(N)} \rrbracket$ , we have  $\text{rank}_{CP}(\boldsymbol{\mathcal{X}}) = \text{rank}_{CP}(\boldsymbol{\mathcal{G}})$ .*

Before proving Theorem 2, we shall first show the following lemma.

**Lemma 3.** *For an  $N$ -way tensor  $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$  with independent Tucker decomposition*

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{G}} \times_1 \mathbf{A}^{(1)} \times_2 \dots \times_N \mathbf{A}^{(N)}, \tag{3}$$

*there exist  $\mathbf{B}^{(1)}, \mathbf{B}^{(2)}, \dots, \mathbf{B}^{(N)}$  such that*

$$\boldsymbol{\mathcal{G}} = \boldsymbol{\mathcal{X}} \times_1 \mathbf{B}^{(1)} \times_2 \dots \times_N \mathbf{B}^{(N)} \text{ and } \boldsymbol{\mathcal{G}} \in \mathbb{C}^{R_1 \times R_2 \times \dots \times R_N}.$$

---

<sup>1</sup>We would like to thank Nikos Sidiropoulos for his insightful comments and information on the topic in private communications.

Also, for any  $n = 1, 2, \dots, N$ , we have

$$\mathcal{G} = \mathcal{G} \times_n \mathbf{A}^{(n)} \times_n \mathbf{B}^{(n)}, \quad \mathcal{X} = \mathcal{X} \times_n \mathbf{B}^{(n)} \times_n \mathbf{A}^{(n)}.$$

**Proof:** Since  $\mathbf{A}^{(n)}$  is a tall matrix (columns are linearly independent),  $\mathbf{A}^{(n)H} \mathbf{A}^{(n)}$  is invertible. By letting

$$\mathbf{B}^{(n)} = (\mathbf{A}^{(n)H} \mathbf{A}^{(n)})^{-1} \mathbf{A}^{(n)H}$$

we have  $\mathbf{B}^{(n)} \mathbf{A}^{(n)} = (\mathbf{A}^{(n)H} \mathbf{A}^{(n)})^{-1} \mathbf{A}^{(n)H} \mathbf{A}^{(n)} = \mathbf{I}_{I_n}$ . As a result,

$$\mathcal{G} \times_n \mathbf{A}^{(n)} \times_n \mathbf{B}^{(n)} = \mathcal{G} \times_n (\mathbf{B}^{(n)} \mathbf{A}^{(n)}) = \mathcal{G} \times_n \mathbf{I}_{R_n} = \mathcal{G}.$$

Moreover, since  $\mathcal{X} = \mathcal{G} \times_1 \mathbf{A}^{(1)} \times_2 \dots \times_N \mathbf{A}^{(N)}$ , applying  $\times_1 \mathbf{B}^{(1)} \times_2 \dots \times_N \mathbf{B}^{(N)}$  on both sides of (3) yields:

$$\begin{aligned} \mathcal{X} \times_1 \mathbf{B}^{(1)} \times_2 \dots \times_N \mathbf{B}^{(N)} &= \mathcal{G} \times_1 \mathbf{A}^{(1)} \times_2 \dots \times_N \mathbf{A}^{(N)} \times_1 \mathbf{B}^{(1)} \times_2 \dots \times_N \mathbf{B}^{(N)} \\ &= \mathcal{G} \times_1 (\mathbf{B}^{(1)} \mathbf{A}^{(1)}) \times_2 \dots \times_N (\mathbf{B}^{(N)} \mathbf{A}^{(N)}) = \mathcal{G}. \end{aligned}$$

In a similar vein, applying  $\times_n \mathbf{B}^{(n)} \times_n \mathbf{A}^{(n)}$  on both sides of (3) yields

$$\mathcal{X} \times_n \mathbf{B}^{(n)} \times_n \mathbf{A}^{(n)} = \mathcal{G} \times_1 \mathbf{A}^{(1)} \times_2 \dots \times_n \mathbf{A}^{(n)} \mathbf{B}^{(n)} \mathbf{A}^{(n)} \dots \times_N \mathbf{A}^{(N)} = \mathcal{X}.$$

□

Note that the above lemma leads to an exact independent Tucker decomposition of a given tensor  $\mathcal{X}$ . First, for each  $n$  one performs a mode- $n$  matricization on  $\mathcal{X}$  to get  $\mathbf{X}_{(1)}, \dots, \mathbf{X}_{(N)}$ . Then, for each mode- $n$  one computes a matrix factorization such that  $\mathbf{X}_{(n)} = \mathbf{A}^{(n)} \mathbf{C}^{(n)}$  and  $\mathbf{A}^{(n)}$  has full column rank. Finally, letting  $\mathbf{B}^{(n)} = \mathbf{A}^{(n)} ((\mathbf{A}^{(n)})^H \mathbf{A}^{(n)})^{-1}$ , we have

$$\begin{aligned} \left( \mathcal{X} \times_n (\mathbf{A}^{(n)})^H \times_n \mathbf{B}^{(n)} \right)_{(n)} &= \left( \mathcal{X} \times_n (\mathbf{A}^{(n)} ((\mathbf{A}^{(n)})^H \mathbf{A}^{(n)})^{-1} (\mathbf{A}^{(n)})^H) \right)_{(n)} \\ &= \mathbf{A}^{(n)} ((\mathbf{A}^{(n)})^H \mathbf{A}^{(n)})^{-1} (\mathbf{A}^{(n)})^H \mathbf{X}_{(n)} \\ &= \mathbf{A}^{(n)} ((\mathbf{A}^{(n)})^H \mathbf{A}^{(n)})^{-1} (\mathbf{A}^{(n)})^H \mathbf{A}^{(n)} \mathbf{C}^{(n)} \\ &= \mathbf{A}^{(n)} \mathbf{C}^{(n)} = \mathbf{X}_{(n)}. \end{aligned}$$

Moreover, due to the one-to-one correspondence between a tensor and its mode matricization we conclude that  $\mathcal{X} \times_n (\mathbf{A}^{(n)})^H \times_n \mathbf{B}^{(n)} = \mathcal{X}$ . Now, by letting  $\mathcal{G} = \mathcal{X} \times_1 (\mathbf{A}^{(1)})^H \times_2 \dots \times_N (\mathbf{A}^{(N)})^H$ , an exact independent Tucker decomposition  $[[\mathcal{G}; \mathbf{B}^{(1)}, \dots, \mathbf{B}^{(N)}]]$  of  $\mathcal{X}$  follows.

**Lemma 4.** For any  $N$ -way tensor

$$\mathcal{X} = \sum_{t=1}^r \mathbf{a}^{(1,t)} \circ \mathbf{a}^{(2,t)} \circ \dots \circ \mathbf{a}^{(N,t)}$$

assuming the multiplications are compatible we have

$$\mathcal{X} \times_n \mathbf{B} = \sum_{t=1}^r \mathbf{a}^{(1,t)} \circ \dots \circ \mathbf{a}^{(n-1,t)} \circ (\mathbf{B} \mathbf{a}^{(n,t)}) \circ \mathbf{a}^{(n+1,t)} \circ \dots \circ \mathbf{a}^{(N,t)}.$$

**Proof:** Denote  $\mathbf{A}^{(n)} = [\mathbf{a}^{(n,1)}, \dots, \mathbf{a}^{(n,r)}]$ . We have  $\mathcal{X} = \llbracket \mathcal{I}; \mathbf{A}^{(1)}, \dots, \mathbf{A}^{(N)} \rrbracket$ , and so

$$\begin{aligned} \mathcal{X} \times_n \mathbf{B} &= \llbracket \mathcal{I}; \mathbf{A}^{(1)}, \dots, \mathbf{A}^{(N)} \rrbracket \times_n \mathbf{B} \\ &= \llbracket \mathcal{I}; \mathbf{A}^{(1)}, \dots, \mathbf{A}^{(n-1)}, \mathbf{B}\mathbf{A}^{(n)}, \mathbf{A}^{(n+1)}, \dots, \mathbf{A}^{(N)} \rrbracket \\ &= \sum_{t=1}^r \mathbf{a}^{(1,t)} \circ \dots \circ \mathbf{a}^{(n-1,t)} \circ \left( \mathbf{B}\mathbf{a}^{(n,t)} \right) \circ \mathbf{a}^{(n+1,t)} \circ \dots \circ \mathbf{a}^{(N,t)}, \end{aligned}$$

which completes the proof.  $\square$

Now we are ready to prove Theorem 2.

**Proof of Theorem 2:** Suppose the core tensor  $\mathcal{G}$  has a CP rank  $r$  associated with the decomposition:

$$\mathcal{G} = \sum_{t=1}^r \mathbf{b}^{(1,t)} \circ \mathbf{b}^{(2,t)} \circ \dots \circ \mathbf{b}^{(N,t)}.$$

Then Lemma 4 suggests that

$$\begin{aligned} \mathcal{X} &= \mathcal{G} \times_1 \mathbf{A}^{(1)} \times_2 \dots \times_N \mathbf{A}^{(N)} \\ &= \sum_{t=1}^r \left( \mathbf{A}^{(1)} \mathbf{b}^{(1,t)} \right) \circ \dots \circ \left( \mathbf{A}^{(N)} \mathbf{b}^{(N,t)} \right), \end{aligned}$$

which is a valid rank-1 decomposition of  $\mathcal{X}$  with  $r$  rank-1 terms, implying that  $\text{rank}_{CP}(\mathcal{X}) \leq r = \text{rank}_{CP}(\mathcal{G})$ .

On the other hand, since the Tucker decomposition is independent, Lemma 3 holds and there exist  $\mathbf{B}^{(1)}, \mathbf{B}^{(2)}, \dots, \mathbf{B}^{(N)}$  such that

$$\mathcal{G} = \mathcal{X} \times_1 \mathbf{B}^{(1)} \times_2 \dots \times_N \mathbf{B}^{(N)}.$$

Applying the same argument, we have  $\text{rank}_{CP}(\mathcal{G}) \leq \text{rank}_{CP}(\mathcal{X})$ . This completes the proof for Theorem 2.  $\square$

Since an orthonormal Tucker decomposition is independent, as a direct consequence of above theorem, we conclude that if  $\mathcal{G}$  is the core tensor of  $\mathcal{X}$  under the orthonormal Tucker decomposition then  $\text{rank}_{CP}(\mathcal{X}) = \text{rank}_{CP}(\mathcal{G})$ .

### 3.2 Tensors with a symmetric structure

In this subsection, we establish similar results for the *symmetric* tensors. Formally speaking, a tensor is *symmetric* if the length along all the directions are equal, and the elements are invariant under any permutation of the indices, i.e.

$$x_{i_{\sigma(1)} i_{\sigma(2)} \dots i_{\sigma(N)}} = x_{i_1 i_2 \dots i_N}$$

where  $\sigma(\cdot)$  is any given permutation function of  $\{1, 2, \dots, N\}$ . Symmetric tensors are well studied; see e.g. [7]. Parallel to the definitions in the preceding sections, we have:

**Definition 4.** For an  $N$ -way symmetric tensor  $\mathcal{X} \in \mathbb{C}^{I \times I \times \dots \times I}$ , a symmetric CP decomposition of size  $r$  is to represent  $\mathcal{X}$  as follows

$$\mathcal{X} = \sum_{t=1}^r \mathbf{a}^{(t)} \circ \dots \circ \mathbf{a}^{(t)}.$$

The symmetric rank of  $\mathcal{X}$ , denoted by  $\text{rank}_S(\mathcal{X})$ , is the minimal integer  $r$  such that a size- $r$  symmetric CP decomposition exists.

**Definition 5.** For an  $N$ -way symmetric tensor  $\mathcal{X} \in \mathbb{C}^{I \times I \times \dots \times I}$ , an exact Tucker decomposition is called symmetric if all the factor matrices are identical; i.e., the Tucker decomposition is of the form

$$\mathcal{X} = \llbracket \mathcal{G}^s; \mathbf{X}, \mathbf{X}, \dots, \mathbf{X} \rrbracket.$$

As before, a symmetric Tucker decomposition is said to be independent if the factor matrix has full column rank; a symmetric Tucker decomposition is said to be orthonormal if the factor matrix has orthonormal columns.

We remark that for an  $N$ -way symmetric tensor  $\mathcal{X} \in \mathbb{C}^{I \times I \times \dots \times I}$ , its matricizations of all modes are identical to each other, i.e.,  $\mathcal{X}_{(m)} = \mathcal{X}_{(n)}$  for all  $m, n = 1, \dots, N$ . Based on this observation, an independent symmetric Tucker decomposition of a given symmetric tensor  $\mathcal{X}$  can be constructed as follows. First, perform a symmetric matrix factorization such that  $\mathbf{X} = \mathbf{A}\mathbf{A}^H$  and  $\mathbf{A}$  has full column rank. Then, construct

$$\mathcal{G} = \mathcal{X} \times_1 \mathbf{A}^H \times_2 \dots \times_N \mathbf{A}^H,$$

and  $\mathbf{B} = \mathbf{A}(\mathbf{A}^H \mathbf{A})^{-1}$ . Following a similar argument for asymmetric tensors, it can be verified that  $\llbracket \mathcal{G}; \mathbf{B}, \dots, \mathbf{B} \rrbracket$  is an independent symmetric Tucker decomposition of  $\mathcal{X}$ . Now we present the invariance of symmetric CP rank.

**Theorem 5.** For any symmetric tensor  $\mathcal{X} \in \mathbb{C}^{I \times I \times \dots \times I}$  and its independent symmetric decomposition

$$\mathcal{X} = \llbracket \mathcal{G}^s; \mathbf{X}, \mathbf{X}, \dots, \mathbf{X} \rrbracket,$$

the core tensor  $\mathcal{G}^s \in \mathbb{C}^{J \times J \times \dots \times J}$  is also symmetric, and  $\text{rank}_S(\mathcal{G}^s) = \text{rank}_S(\mathcal{X})$ .

**Proof:** By Lemma 3 we have

$$\mathcal{G}^s = \mathcal{X} \times_1 \mathbf{B} \times_2 \dots \times_N \mathbf{B}, \tag{4}$$

where  $\mathbf{B} = (\mathbf{X}^H \mathbf{X})^{-1} \mathbf{X}^H \in \mathbb{C}^{J \times I}$ . Now we show that  $\mathcal{G}^s$  is symmetric. For any  $(j_1 j_2 \dots j_N)$  and any permutation function  $\sigma(\cdot)$  of  $\{1, 2, \dots, N\}$ , by definition of the mode product, we have

$$\begin{aligned}
g_{j_1 j_2 \dots j_N}^s &= \sum_{i_1, \dots, i_N=1}^I x_{i_1 i_2 \dots i_N} \cdot b_{i_1 j_1} b_{i_2 j_2} \dots b_{i_N j_N} \\
&= \sum_{i_1, \dots, i_N=1}^I x_{i_1 i_2 \dots i_N} \cdot b_{i_{\sigma(1)} j_{\sigma(1)}} b_{i_{\sigma(2)} j_{\sigma(2)}} \dots b_{i_{\sigma(N)} j_{\sigma(N)}} \\
&= \sum_{i_1, \dots, i_N=1}^I x_{i_{\sigma(1)} i_{\sigma(2)} \dots i_{\sigma(N)}} \cdot b_{i_{\sigma(1)} j_{\sigma(1)}} b_{i_{\sigma(2)} j_{\sigma(2)}} \dots b_{i_{\sigma(N)} j_{\sigma(N)}} \\
&= g_{j_{\sigma(1)} j_{\sigma(2)} \dots j_{\sigma(N)}}^s,
\end{aligned}$$

where the third equality follows from the fact that  $\mathcal{X}$  is symmetric.

Now that  $\mathcal{G}^s$  is symmetric, we may assume its symmetric CP rank to be  $r$ , with the decomposition

$$\mathcal{G}^s = \sum_{t=1}^r \underbrace{\mathbf{b}^t \circ \dots \circ \mathbf{b}^t}_N.$$

Invoking Lemma 4 yields

$$\begin{aligned}
\mathcal{X} &= \mathcal{G}^s \times_1 \mathbf{X} \times_2 \dots \times_N \mathbf{X} \\
&= \sum_{t=1}^r \underbrace{(\mathbf{X} \mathbf{b}^t) \circ \dots \circ (\mathbf{X} \mathbf{b}^t)}_N,
\end{aligned}$$

which is a symmetric rank-1 decomposition of  $\mathcal{X}$ . This implies that  $\text{rank}_S(\mathcal{X}) \leq \text{rank}_S(\mathcal{G}^s)$ . Noting that we can also decompose  $\mathcal{G}^s$  in the form of (4), by the same argument we have  $\text{rank}_S(\mathcal{G}^s) \leq \text{rank}_S(\mathcal{X})$ , leading to  $\text{rank}_S(\mathcal{X}) = \text{rank}_S(\mathcal{G}^s)$ .  $\square$

## 4 Invariance of the tensor norms and the border rank

In this section, we study the invariance properties of various tensor norms and the so-called border rank. These relationships enable us to measure the error between the tensor and an approximative decomposition.

### 4.1 Invariance of tensor norms under independent Tucker decomposition

We first consider the tensor Frobenius norm.

**Theorem 6.** For any given tensor  $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$  with independent Tucker decomposition  $\mathcal{X} = \llbracket \mathcal{G}; \mathbf{X}^{(1)}, \dots, \mathbf{X}^{(N)} \rrbracket$ , we have that

$$\alpha \|\mathcal{G}\|_F \leq \|\mathcal{X}\|_F \leq \beta \|\mathcal{G}\|_F, \quad (5)$$

where  $\beta = \|\mathbf{X}^{(1)}\|_2 \|\mathbf{X}^{(2)}\|_2 \dots \|\mathbf{X}^{(N)}\|_2$  and  $\alpha = \beta / (\prod_{n=1}^N \kappa(\mathbf{X}^{(n)}))$  with  $\kappa(\mathbf{X}^{(n)})$  being the condition number of the matrix  $\mathbf{X}^{(n)}$  for all  $n$ .

The following lemma presents a bound on the Frobenius norm of mode- $n$  matrix multiplication, which leads to the proof of Theorem 6.

**Lemma 7.** For any given tensor  $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$  and matrix  $\mathbf{A} \in \mathbb{C}^{J_n \times I_n}$ , we have

$$\|\mathcal{X} \times_n \mathbf{A}\|_F \leq \|\mathbf{A}\|_2 \|\mathcal{X}\|_F.$$

**Proof:** The claimed inequality is equivalent to

$$\|\mathbf{A} \mathbf{X}_{(n)}\|_F^2 \leq \|\mathbf{A}\|_2^2 \|\mathbf{X}_{(n)}\|_F^2.$$

To show this, let  $\mathbf{X}_{(n,k)}$  be the  $k$ -th column of  $\mathbf{X}_{(n)}$ . We have

$$\begin{aligned} \|\mathbf{A} \mathbf{X}_{(n)}\|_F^2 &= \sum_k \|\mathbf{A} \mathbf{X}_{(n,k)}\|_F^2 \\ &\leq \sum_k \|\mathbf{A}\|_2^2 \|\mathbf{X}_{(n,k)}\|_F^2 \\ &= \|\mathbf{A}\|_2^2 \left( \sum_k \|\mathbf{X}_{(n,k)}\|_F^2 \right) \\ &= \|\mathbf{A}\|_2^2 \|\mathbf{X}_{(n)}\|_F^2, \end{aligned}$$

where the inequality is due to the consistency of the matrix spectral norm and the Euclidean vector norm.  $\square$

**Proof of Theorem 6:** Since  $\mathcal{X} = \llbracket \mathcal{G}; \mathbf{X}^{(1)}, \dots, \mathbf{X}^{(N)} \rrbracket$ , by Lemma 7,

$$\|\mathcal{X}\|_F = \|\llbracket \mathcal{G}; \mathbf{X}^{(1)}, \dots, \mathbf{X}^{(N)} \rrbracket\|_F \leq \|\mathcal{G}\|_F \|\mathbf{X}^{(1)}\|_2 \|\mathbf{X}^{(2)}\|_2 \dots \|\mathbf{X}^{(N)}\|_2.$$

Thus, we may simply let  $\beta = \|\mathbf{X}^{(1)}\|_2 \|\mathbf{X}^{(2)}\|_2 \dots \|\mathbf{X}^{(N)}\|_2$  and obtain the upper bound in (5). On the other hand, by Lemma 3 it follows that

$$\mathcal{G} = \llbracket \mathcal{X}; \mathbf{Y}^{(1)}, \dots, \mathbf{Y}^{(N)} \rrbracket,$$

with  $\mathbf{Y}^{(n)} = \left( (\mathbf{X}^{(n)})^H \mathbf{X}^{(n)} \right)^{-1} (\mathbf{X}^{(n)})^H$ . Therefore, from Lemma 7 one has

$$\|\mathcal{G}\|_F = \|\llbracket \mathcal{X}; \mathbf{Y}^{(1)}, \dots, \mathbf{Y}^{(N)} \rrbracket\|_F \leq \|\mathcal{X}\|_F \|\mathbf{Y}^{(1)}\|_2 \|\mathbf{Y}^{(2)}\|_2 \dots \|\mathbf{Y}^{(N)}\|_2.$$

Now we let  $\alpha = \frac{1}{\|\mathbf{Y}^{(1)}\|_2 \|\mathbf{Y}^{(2)}\|_2 \cdots \|\mathbf{Y}^{(N)}\|_2}$ , and it remains to prove  $\beta/\alpha = \prod_{n=1}^N \kappa(\mathbf{X}^{(n)})$ . To this end, for any  $n$ , suppose the matrix  $\mathbf{X}^{(n)}$  has SVD:  $\mathbf{X}^{(n)} = \mathbf{U}^{(n)} \boldsymbol{\Sigma} (\mathbf{V}^{(n)})^H$ . Then

$$\begin{aligned} \mathbf{Y}^{(n)} &= \left( (\mathbf{X}^{(n)})^H \mathbf{X}^{(n)} \right)^{-1} (\mathbf{X}^{(n)})^H \\ &= \left( \mathbf{V}^{(n)} \boldsymbol{\Sigma}^2 (\mathbf{V}^{(n)})^H \right)^{-1} \mathbf{V}^{(n)} \boldsymbol{\Sigma} (\mathbf{U}^{(n)})^H \\ &= \mathbf{V}^{(n)} \boldsymbol{\Sigma}^{-2} (\mathbf{V}^{(n)})^H \mathbf{V}^{(n)} \boldsymbol{\Sigma} (\mathbf{U}^{(n)})^H \\ &= \mathbf{V}^{(n)} \boldsymbol{\Sigma}^{-1} (\mathbf{U}^{(n)})^H. \end{aligned}$$

Therefore,  $\|\mathbf{Y}^{(n)}\|_2 = \frac{1}{\sigma_{\min}(\mathbf{X}^{(n)})}$  for all  $n$  and  $\alpha = \prod_{n=1}^N \sigma_{\min}(\mathbf{X}^{(n)})$ . Consequently,

$$\frac{\beta}{\alpha} = \frac{\prod_{n=1}^N \|(\mathbf{X}^{(n)})\|_2}{\prod_{n=1}^N \sigma_{\min}(\mathbf{X}^{(n)})} = \frac{\prod_{n=1}^N \sigma_{\max}(\mathbf{X}^{(n)})}{\prod_{n=1}^N \sigma_{\min}(\mathbf{X}^{(n)})} = \prod_{n=1}^N \kappa(\mathbf{X}^{(n)}).$$

□

Note that when  $\beta = \alpha$ , the estimation (5) on the Frobenius norm of  $\mathcal{X}$  becomes exact. In general, the ratio  $\beta/\alpha$  measures the quality of approximating the Frobenius norm of  $\mathcal{X}$  by that of  $\alpha \mathcal{G}$ . When the factor matrix is orthonormal, the associated condition number is 1, and then the Frobenius norm of the tensor and its Tucker core are equal.

**Corollary 8.** *For an orthonormal Tucker decomposition, its factor matrix  $\mathbf{X}^{(n)}$  is orthonormal. Thus  $\|\mathbf{X}^{(n)}\|_2 = \|(\mathbf{X}^{(n)})^H\|_2 = 1$  for  $n = 1, 2, \dots, N$ , and we have  $\|\mathcal{X}\|_F = \|\mathcal{G}\|_F$ .*

Remark that the above result was shown by De Lathauwer et al. as Property 8 in [26].

## 4.2 The quasi- $p$ norm and the tensor nuclear norm

We proceed to other tensor norms in this subsection.

**Definition 6.** *For any tensor  $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$ , the tensor  $p$ -quasi norm for  $0 < p \leq 1$ <sup>2</sup> of  $\mathcal{X}$  is defined as*

$$\|\mathcal{X}\|_p \triangleq \inf \left\{ \left( \sum_{s=1}^r |\lambda_s|^p \right)^{1/p} : \mathcal{X} = \sum_{s=1}^r \lambda_s \mathbf{x}_s^{(1)} \circ \mathbf{x}_s^{(2)} \circ \cdots \circ \mathbf{x}_s^{(N)}, \right. \\ \left. \|\mathbf{x}_s^{(n)}\| = 1, \forall s = 1, 2, \dots, r, \forall n = 1, 2, \dots, N \right\}. \quad (6)$$

---

<sup>2</sup>In a private conversation, Lek-Heng Lim pointed out to us that  $\|\mathcal{X}\|_p$  trivially equals to zero for any tensor  $\mathcal{X}$  for any  $p > 1$ .

When  $N = 2$ , it reduces to the Schatten  $p$ -quasi norm for matrix, which plays an important role in low-rank matrix optimization [16]. When  $p = 1$ , the above definition corresponds to the tensor nuclear norm, which was originally proposed in Grothendieck [11] and Schatten [35]. Recently, this tensor nuclear norm was applied in Yuan and Zhang [41] to analyze the statistical properties of Tensor completion problem. Friedland and Lim [9] showed that computing the nuclear norm of a given tensor is NP-hard. Interestingly, for the quasi- $p$  norm, similar bounds between a given tensor and its core tensor hold true.

**Theorem 9.** For any given tensor  $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$  with independent Tucker decomposition  $\mathcal{X} = \llbracket \mathcal{G}; \mathbf{X}^{(1)}, \dots, \mathbf{X}^{(N)} \rrbracket$ , we have

$$\alpha \|\mathcal{G}\|_p \leq \|\mathcal{X}\|_p \leq \beta \|\mathcal{G}\|_p,$$

where  $\beta = \|\mathbf{X}^{(1)}\|_2 \|\mathbf{X}^{(2)}\|_2 \dots \|\mathbf{X}^{(N)}\|_2$  and  $\alpha = \beta / (\prod_{n=1}^N \kappa(\mathbf{X}^{(n)}))$  with  $\kappa(\mathbf{X}^{(n)})$  being the condition number of matrix  $\mathbf{X}^{(n)}$  for all  $n$ .

**Proof:** For any  $\epsilon > 0$  find CP decomposition of tensor  $\mathcal{G} = \sum_{s=1}^r \lambda_s \mathbf{g}_s^{(1)} \circ \mathbf{g}_s^{(2)} \circ \dots \circ \mathbf{g}_s^{(N)}$  with  $\|\mathbf{g}_s^{(n)}\| = 1$  satisfying  $\|\mathcal{G}\|_p \geq (\sum_{s=1}^r |\lambda_s|^p)^{1/p} - \epsilon$ , denote the index set  $S = \{1 \leq s \leq r \mid \mathbf{X}^{(i)} \mathbf{g}_s^{(i)} = 0 \text{ for some } i, 1 \leq i \leq N\}$ . Then from Lemma 4, we have

$$\begin{aligned} \mathcal{X} &= \sum_{s=1}^r \lambda_s \mathbf{g}_s^{(1)} \circ \mathbf{g}_s^{(2)} \circ \dots \circ \mathbf{g}_s^{(N)} \times_1 \mathbf{X}^{(1)} \times_2 \mathbf{X}^{(2)} \times_3 \dots \times_N \mathbf{X}^{(N)} \\ &= \sum_{s=1}^r \lambda_s (\mathbf{X}^{(1)} \mathbf{g}_s^{(1)}) \circ (\mathbf{X}^{(2)} \mathbf{g}_s^{(2)}) \circ \dots \circ (\mathbf{X}^{(N)} \mathbf{g}_s^{(N)}) \\ &= \sum_{s \in S} \lambda_s (\mathbf{X}^{(1)} \mathbf{g}_s^{(1)}) \circ (\mathbf{X}^{(2)} \mathbf{g}_s^{(2)}) \circ \dots \circ (\mathbf{X}^{(N)} \mathbf{g}_s^{(N)}) \\ &= \sum_{s \in S} \prod_{n=1}^N \|\mathbf{X}^{(n)} \mathbf{g}_s^{(n)}\| \cdot \lambda_s \left( \frac{\mathbf{X}^{(1)} \mathbf{g}_s^{(1)}}{\|\mathbf{X}^{(1)} \mathbf{g}_s^{(1)}\|} \right) \circ \dots \circ \left( \frac{\mathbf{X}^{(N)} \mathbf{g}_s^{(N)}}{\|\mathbf{X}^{(N)} \mathbf{g}_s^{(N)}\|} \right), \end{aligned}$$

which is a valid CP decomposition of  $\mathcal{X}$ . Moreover, by the definition of tensor  $p$ -quasi norm,

$$\begin{aligned} \|\mathcal{X}\|_p &\leq \left( \sum_{s \in S} \prod_{n=1}^N \|\mathbf{X}^{(n)} \mathbf{g}_s^{(n)}\|^p \cdot |\lambda_s|^p \right)^{1/p} \leq \left( \sum_{s \in S} \prod_{n=1}^N (\|\mathbf{X}^{(n)}\|_2^p \|\mathbf{g}_s^{(n)}\|^p) \cdot |\lambda_s|^p \right)^{1/p} \\ &= \left( \sum_{s=1}^r \prod_{n=1}^N (\|\mathbf{X}^{(n)}\|_2^p \|\mathbf{g}_s^{(n)}\|^p) \cdot |\lambda_s|^p \right)^{1/p} = \prod_{n=1}^N \|\mathbf{X}^{(n)}\|_2 \left( \sum_{s=1}^r |\lambda_s|^p \right)^{1/p} = \prod_{n=1}^N \|\mathbf{X}^{(n)}\|_2 \cdot (\|\mathcal{G}\|_p + \epsilon), \end{aligned}$$

where the second equality is due to  $\|\mathbf{g}_s^{(n)}\| = 1$  for all  $n$  and  $s$ . Since  $\epsilon > 0$  can be chosen arbitrarily, we have  $\|\mathcal{X}\|_p \leq \prod_{n=1}^N \|\mathbf{X}^{(n)}\|_2 \cdot \|\mathcal{G}\|_p$ . On the other hand, due to the nature of independent Tucker decomposition and Lemma 3, one has that  $\mathcal{G} = \llbracket \mathcal{X}; \mathbf{Y}^{(1)}, \dots, \mathbf{Y}^{(N)} \rrbracket$ . By repeating above argument,

we have  $\|\mathcal{G}\|_p \leq \prod_{n=1}^N \|\mathbf{Y}^{(n)}\|_2 \cdot \|\mathcal{X}\|_p$ . The rest of the proof follows similarly from that of Theorem 6.  $\square$

Similar to the analysis in the previous subsection, when the Tucker decomposition is orthonormal then the  $p$ -quasi norm of a tensor and that of its core are equal.

**Corollary 10.** *For an orthonormal Tucker decomposition, we have  $\|\mathcal{X}\|_p = \|\mathcal{G}\|_p$ .*

### 4.3 The border rank

Although the CP rank is a natural extension of matrix rank, one undesirable theoretical property of this definition is that the best rank- $r$  approximation may not even exist (see [23] for more details). In particular, a rank- $r$  tensor may be approximated arbitrarily close by a sequence of tensors whose CP ranks are strictly less than  $r$ . To get around this, Bini [2] proposes the concept of the border rank, which is defined as follows.

**Definition 7.** *For an  $N$ -way tensor  $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$ , its border rank, denoted by  $\text{rank}_B(\mathcal{X})$ , is defined as*

$$\text{rank}_B = \min \{r \mid \forall \epsilon > 0, \exists \mathcal{E} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}, \text{ s.t. } \|\mathcal{E}\|_F \leq \epsilon \text{ and } \text{rank}_{CP}(\mathcal{X} + \mathcal{E}) \leq r\}.$$

In other words, the border rank of a given tensor is the minimum CP-rank of tensors that can be found in any neighborhood of the given tensor. By the bound on the tensor norms, we now establish the equality of the border rank between a given tensor and its core.

**Theorem 11.** *For any given tensor  $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$  with independent Tucker decomposition  $\mathcal{X} = \llbracket \mathcal{G}; \mathbf{X}^{(1)}, \dots, \mathbf{X}^{(N)} \rrbracket$ , we have  $\text{rank}_B(\mathcal{X}) = \text{rank}_B(\mathcal{G})$ .*

**Proof:** Assume  $\text{rank}_B(\mathcal{G}) = r$ , we want to show  $\text{rank}_B(\mathcal{X}) \leq \text{rank}_B(\mathcal{G}) = r$ . By definition of the border rank, for any  $\epsilon > 0$ , there exists  $\mathcal{E} \in \mathbb{C}^{J_1 \times J_2 \times \dots \times J_N}$  such that  $\|\mathcal{E}\|_F \leq \epsilon / \prod_{n=1}^N \sigma_{\max}(\mathbf{X}^{(n)})$  and  $\text{rank}_{CP}(\mathcal{G} + \mathcal{E}) \leq r$ . Now, construct

$$\mathcal{T} = \mathcal{E} \times_1 \mathbf{X}^{(1)} \times_2 \dots \times_N \mathbf{X}^{(N)},$$

and consider the tensor  $\mathcal{X} + \mathcal{T} = (\mathcal{G} + \mathcal{E}) \times_1 \mathbf{X}^{(1)} \times_2 \dots \times_N \mathbf{X}^{(N)}$ . Obviously, by Theorem 2 we have

$$\text{rank}_{CP}(\mathcal{G} + \mathcal{E}) = \text{rank}_{CP}(\mathcal{X} + \mathcal{T}) \leq r. \quad (7)$$

Moreover, according to Theorem 6 we also have

$$\|\mathcal{T}\|_F = \prod_{n=1}^N \sigma_{\max}(\mathbf{X}^{(n)}) \cdot \|\mathcal{E}\|_F \leq \epsilon. \quad (8)$$

Combining (7) and (8) implies that  $\text{rank}_B(\mathcal{X}) \leq \text{rank}_B(\mathcal{G})$ .

Now we recall that for independent Tucker decomposition,  $\mathcal{G} = \llbracket \mathcal{X}; \mathbf{Y}^{(1)}, \dots, \mathbf{Y}^{(N)} \rrbracket$ , with  $\mathbf{Y}^{(n)} = ((\mathbf{X}^{(n)})^H \mathbf{X}^{(n)})^{-1} (\mathbf{X}^{(n)})^H$  for all  $n$  (see Lemma 3 for more details). The inequality  $\text{rank}_B(\mathcal{X}) \geq \text{rank}_B(\mathcal{G})$  follows similarly from the argument above, which establishes the desired equality.  $\square$

## 5 Invariance of tensor eigenvalues

In this section, we focus on the real-field and investigate the invariance properties of various notions of tensor eigenvalues.

### 5.1 Invariance of the Z-eigenvalues

Let us consider the class of real-valued symmetric tensors, and denote the symmetric rank-one tensor  $\underbrace{\mathbf{x} \circ \dots \circ \mathbf{x}}_N$  as  $\mathbf{x}^{\circ N}$ . For any symmetric tensor  $\mathcal{T} \in \mathbb{R}^{I \times \dots \times I}$  and  $(N-1)$ -way rank-1 tensor  $\mathbf{x}^{\circ(N-1)}$ ,  $\mathcal{T}(\mathbf{x}^{\circ(N-1)})$  denotes an  $I$  dimensional vector such that

$$(\mathcal{T}(\mathbf{x}^{\circ(N-1)}))_{i_N} \triangleq \sum_{i_1, i_2, \dots, i_{N-1}=1}^I T_{i_1 i_2 \dots i_{N-1} i_N} \cdot x_{i_1} x_{i_2} \dots x_{i_{N-1}}.$$

With this notion in place, the Z-eigenvalue and Z-eigenvector of a tensor are defined as follows.

**Definition 8.** For an  $N$ -way symmetric tensor  $\mathcal{T} \in \mathbb{R}^{I \times \dots \times I}$ , if there exists a number  $\lambda \in \mathbb{R}$  and a nonzero vector  $\mathbf{x} \in \mathbb{R}^I$  such that

$$\mathcal{T}(\mathbf{x}^{\circ(N-1)}) = \lambda \mathbf{x}, \quad \mathbf{x}^T \mathbf{x} = 1. \quad (9)$$

Then  $\lambda$  is called the Z-eigenvalue of  $\mathcal{T}$ , and  $\mathbf{x}$  is called the corresponding Z-eigenvector.

The Z-eigenvalues were first studied by Qi [31] and Lim [28] independently. The relationship between the Z-eigenvalues of a symmetric tensor and that of its core is described as follows.

**Theorem 12.** For any given  $N$ -way symmetric tensor  $\mathcal{T} \in \mathbb{R}^{I \times I \times \dots \times I}$  with exact independent symmetric Tucker decomposition  $\mathcal{T} = \llbracket \mathcal{G}^s; \mathbf{X}, \mathbf{X}, \dots, \mathbf{X} \rrbracket$ , construct

$$\hat{\mathcal{G}}^s = \mathcal{G}^s \times_1 (\mathbf{X}^T \mathbf{X})^{1/2} \times_2 \dots \times_N (\mathbf{X}^T \mathbf{X})^{1/2}.$$

Then any Z-eigenvalues of  $\hat{\mathcal{G}}^s$  are also Z-eigenvalues of  $\mathcal{T}$  while any non-zero Z-eigenvalue of  $\mathcal{T}$  are also Z-eigenvalues of  $\hat{\mathcal{G}}^s$ .

**Proof:** Since the Tucker decomposition is independent, by Lemma 3,

$$\mathcal{G}^s = \mathcal{T} \times_1 \mathbf{Y} \times_2 \cdots \times_N \mathbf{Y},$$

with  $\mathbf{Y} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ . Furthermore, due to Theorem 5,  $\mathcal{G}^s$  is symmetric and so is  $\hat{\mathcal{G}}^s$ . Let  $\lambda$  be a Z-eigenvalue of  $\hat{\mathcal{G}}^s$  with Z-eigenvector  $\mathbf{a}$ . We have

$$\begin{aligned} \lambda \mathbf{a} &= \hat{\mathcal{G}}^s(\mathbf{a}^{\circ(N-1)}) = \hat{\mathcal{G}}^s \times_1 \mathbf{a}^T \times_2 \cdots \times_{N-1} \mathbf{a}^T \\ &= \mathcal{G}^s \times_1 (\mathbf{X}^T \mathbf{X})^{1/2} \times_2 \cdots \times_N (\mathbf{X}^T \mathbf{X})^{1/2} \times_1 \mathbf{a}^T \times_2 \cdots \times_{N-1} \mathbf{a}^T \\ &= \mathcal{T} \times_1 \mathbf{Y} \times_2 \cdots \times_N \mathbf{Y} \times_1 \mathbf{a}^T (\mathbf{X}^T \mathbf{X})^{1/2} \times_2 \cdots \times_{N-1} \mathbf{a}^T (\mathbf{X}^T \mathbf{X})^{1/2} \times_N (\mathbf{X}^T \mathbf{X})^{1/2} \\ &= \mathcal{T} \times_1 \mathbf{a}^T (\mathbf{X}^T \mathbf{X})^{-1/2} \mathbf{X}^T \times_2 \cdots \times_{N-1} \mathbf{a}^T (\mathbf{X}^T \mathbf{X})^{-1/2} \mathbf{X}^T \times_N (\mathbf{X}^T \mathbf{X})^{-1/2} \mathbf{X}^T. \end{aligned}$$

Applying  $\times_N \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1/2}$  to both sides and invoking Lemma 3 yield

$$\begin{aligned} \mathcal{T}((\mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1/2} \mathbf{a})^{\circ(N-1)}) &= \mathcal{T} \times_1 \mathbf{a}^T (\mathbf{X}^T \mathbf{X})^{-1/2} \mathbf{X}^T \times_2 \cdots \times_{N-1} \mathbf{a}^T (\mathbf{X}^T \mathbf{X})^{-1/2} \mathbf{X}^T \\ &= \lambda \mathbf{a} \times_N \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1/2} = \lambda \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1/2} \mathbf{a}. \end{aligned}$$

Furthermore, we have  $\mathbf{a}^T (\mathbf{X}^T \mathbf{X})^{-1/2} \mathbf{X}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1/2} \mathbf{a} = \mathbf{a}^T \mathbf{a} = 1$ , thus  $\lambda$  is an eigenvalue of  $\mathcal{T}$  with corresponding eigenvector  $\mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1/2} \mathbf{a}$ .

On the other hand, suppose  $\mu$  is a Z-eigenvalue of  $\mathcal{T}$  associated with the Z-eigenvector  $\mathbf{b}$ . Following a similar argument, one has

$$\begin{aligned} &\hat{\mathcal{G}}^s \left( ((\mathbf{X}^T \mathbf{X})^{-1/2} \mathbf{X}^T \mathbf{b})^{\circ(N-1)} \right) \\ &= \hat{\mathcal{G}}^s \times_1 (\mathbf{b}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1/2}) \times_2 \cdots \times_{N-1} (\mathbf{b}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1/2}) \\ &= \mu (\mathbf{X}^T \mathbf{X})^{-1/2} \mathbf{X}^T \mathbf{b}. \end{aligned}$$

By applying  $\times_N \mathbf{b}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1/2}$  to both sides of the above equality, we have

$$\hat{\mathcal{G}}^s \times_1 (\mathbf{b}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1/2}) \times_2 \cdots \times_N (\mathbf{b}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1/2}) = \mu \mathbf{b}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{b}.$$

Moreover, it is easy to verify that

$$\begin{aligned} &\hat{\mathcal{G}}^s \times_1 (\mathbf{b}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1/2}) \times_2 \cdots \times_N (\mathbf{b}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1/2}) \\ &= \mathcal{G}^s \times_1 \mathbf{b}^T \mathbf{X} \times_2 \cdots \times_N \mathbf{b}^T \mathbf{X} \\ &= \mathcal{T} \times_1 \mathbf{Y} \times_2 \cdots \times_N \mathbf{Y} \times_1 \mathbf{b}^T \mathbf{X} \times_2 \cdots \times_N \mathbf{b}^T \mathbf{X} \\ &= \mathcal{T} \times_1 \mathbf{b}^T \times_2 \cdots \times_N \mathbf{b}^T = \mu \mathbf{b} \times_N \mathbf{b}^T = \mu, \end{aligned}$$

where the third equality is due to Lemma 3. If  $\mu \neq 0$ , then  $\mathbf{b}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{b} = 1$ , meaning that  $\mu$  is the Z-eigenvalue of core tensor  $\mathcal{G}^s$  with the associated Z-eigenvector  $(\mathbf{X}^T \mathbf{X})^{-1/2} \mathbf{X}^T \mathbf{b}$ .  $\square$

We remark that if the Tucker decomposition is orthonormal then  $\hat{\mathcal{G}}^s = \mathcal{G}^s$ , and the above theorem states that all the Z-eigenvalues (except zero) of a symmetric tensor equal to the Z-eigenvalues of its core. This motivates us to focus on the eigenvalues of the core tensor, which however may miss a zero eigenvalue. Fortunately, the following result tells us that when the size of the core is strictly less than the size of the original tensor, then zero eigenvalue is always present.

**Proposition 13.** *Suppose a given  $N$ -way symmetric tensor  $\mathcal{T} \in \mathbb{R}^{I \times I \times \dots \times I}$  has an exact independent symmetric Tucker decomposition  $\mathcal{T} = \llbracket \mathcal{G}^s; \mathbf{X}, \mathbf{X}, \dots, \mathbf{X} \rrbracket$  such that  $\mathcal{G}^s \in \mathbb{R}^{J \times J \times \dots \times J}$ . If  $I > J$ , then 0 is an Z-eigenvalue.*

**Proof:** We note the the factor matrix  $\mathbf{X} \in \mathbb{R}^{I \times J}$  in the independent Tucker decomposition. Since  $J < I$ , there exists a non-zero vector  $\mathbf{a}$  such that  $\mathbf{X}^T \mathbf{a} = 0$ . Thus we have

$$\mathcal{T} \times_1 \mathbf{a}^T \cdots \times_{N-1} \mathbf{a}^T = \mathcal{G}^s \times_1 (\mathbf{a}^T \mathbf{X}) \cdots \times_{N-1} (\mathbf{a}^T \mathbf{X}) \times_N \mathbf{X} = 0,$$

implying that  $\mathbf{a}$  is a Z-eigenvector corresponding to the Z-eigenvalue 0 of the tensor  $\mathcal{T}$ .  $\square$

Invariance of the Z-eigenvalues has interesting implications regarding the nonnegativity properties as well. A commonly used notion of nonnegativity is the *positive semidefinite (PSD)* tensor:

$$\mathcal{T}(\mathbf{x}^{\circ(2N)}) \geq 0, \forall \mathbf{x} \in \mathbb{R}^I,$$

where  $\mathcal{T}$  is symmetric and has degree  $2N$ . Since all the Z-eigenvalues of  $\mathcal{T}$  correspond to all the KKT points of the polynomial optimization:  $\min_{\|\mathbf{x}\|_2=1} \mathcal{T}(\mathbf{x}^{\circ(2N)})$ , we have the following result as a consequence of Theorem 12:

**Corollary 14.** *For any given  $2N$ -way symmetric tensor  $\mathcal{T} \in \mathbb{R}^{I \times I \times \dots \times I}$  with exact independent symmetric Tucker decomposition  $\mathcal{T} = \llbracket \mathcal{G}^s; \mathbf{X}, \mathbf{X}, \dots, \mathbf{X} \rrbracket$ ,  $\mathcal{T}$  is PSD if and only if  $\mathcal{G}^s$  is PSD.*

The dual of the class of PSD tensors is the *sum of powers (SOP)* tensors (see [17]):

$$\mathcal{T} = \sum_{k=1}^m (\mathbf{x}^k)^{\circ(2N)}, \text{ where } m \text{ a positive integer and } \mathbf{x}^k \in \mathbb{R}^I, \forall k = 1, \dots, m.$$

Similarly, Theorem 2 also implies:

**Corollary 15.** *For any given  $2N$ -way symmetric tensor  $\mathcal{T} \in \mathbb{R}^{I \times I \times \dots \times I}$  with exact independent symmetric Tucker decomposition  $\mathcal{T} = \llbracket \mathcal{G}^s; \mathbf{X}, \mathbf{X}, \dots, \mathbf{X} \rrbracket$ ,  $\mathcal{T}$  is SOP if and only if  $\mathcal{G}^s$  is SOP.*

We remark that verifying whether a tensor is PSD or SOP are in general NP-hard problems [13, 17]. A famous result of Hilbert states that a 4th order tertiary polynomial is PSD if and only if it is a sum of squares, where the latter condition can be verified easily. This implies that if the core of a 4th order symmetric tensor has size no more than 3, then one can easily verify if it is PSD or not.

## 5.2 Invariance of the M-eigenvalue

In this subsection, we consider tensors with a “less” symmetric structure, termed *partial symmetry*. In particular, for a four-way tensor  $\mathcal{T} \in \mathbb{C}^{N \times M \times N \times M}$ , we call it *partial symmetric* if

$$t_{ijkl} = t_{kjil} = t_{ilkj} = t_{klij}, \text{ for } i, k = 1, 2, \dots, N; j, l = 1, 2, \dots, M.$$

Similarly, a Tucker decomposition is called *partial symmetric* if it is of the form

$$\mathcal{T} = \llbracket \mathcal{G}^{ps}; \mathbf{A}, \mathbf{B}, \mathbf{A}, \mathbf{B} \rrbracket.$$

Below we introduce the notion of M-eigenvalue and M-eigenvector proposed in Qi, Dai and Han in [32].

**Definition 9.** For a four-way partial symmetric tensor  $\mathcal{T} \in \mathbb{R}^{N \times M \times N \times M}$ , if there exist two numbers  $\lambda$  and  $\mu \in \mathbb{R}$ , two nonzero vectors  $\mathbf{x} \in \mathbb{R}^N$  and  $\mathbf{y} \in \mathbb{R}^M$  such that

$$\mathcal{T}(\cdot, \mathbf{y}, \mathbf{x}, \mathbf{y}) = \lambda \mathbf{x}, \quad \mathbf{x}^T \mathbf{x} = 1$$

$$\mathcal{T}(\mathbf{x}, \mathbf{y}, \mathbf{x}, \cdot) = \mu \mathbf{y}, \quad \mathbf{y}^T \mathbf{y} = 1$$

where  $\mathcal{T}(\cdot, \mathbf{y}, \mathbf{x}, \mathbf{y}) = \sum_{k=1}^N \sum_{l=1}^M t_{ijkl} y_j x_k y_l$  and  $\mathcal{T}(\mathbf{x}, \mathbf{y}, \mathbf{x}, \cdot) = \sum_{i,k=1}^N \sum_{j=1}^M t_{ijkl} x_i y_j x_k$ . Then  $\lambda$  and  $\mu$  are called the M-eigenvalues of  $\mathcal{T}$ , while  $\mathbf{x}$  and  $\mathbf{y}$  are called the corresponding M-eigenvectors.

**Theorem 16.** For any four-way partial symmetric tensor  $\mathcal{T} \in \mathbb{R}^{N \times M \times N \times M}$  with its exact independent partial symmetric Tucker decomposition  $\mathcal{T} = \llbracket \mathcal{G}^{ps}; \mathbf{A}, \mathbf{B}, \mathbf{A}, \mathbf{B} \rrbracket$ , construct

$$\hat{\mathcal{G}}^{ps} = \times_1 (\mathbf{A}^T \mathbf{A})^{1/2} \times_2 (\mathbf{B}^T \mathbf{B})^{1/2} \times_3 (\mathbf{A}^T \mathbf{A})^{1/2} \times_4 (\mathbf{B}^T \mathbf{B})^{1/2}.$$

Then any M-eigenvalues of  $\hat{\mathcal{G}}^{ps}$  are also M-eigenvalues of  $\mathcal{T}$  while any non-zero M-eigenvalues of  $\mathcal{T}$  are also M-eigenvalues of  $\hat{\mathcal{G}}^{ps}$ .

**Proof:** Similar to the proof of Theorem 5, one can show that  $\mathcal{G}^{ps}$  and  $\hat{\mathcal{G}}^{ps}$  are both partial symmetric. Therefore, the M-eigenvalues of  $\hat{\mathcal{G}}^{ps}$  are well-defined. The rest of the proof is similar to that of Theorem 12, and is omitted here for brevity.  $\square$

Similar to the symmetric case, if the partial symmetric Tucker decomposition is orthonormal then  $\hat{\mathcal{G}}^{ps} = \mathcal{G}^{ps}$ , and the equivalence of the M-eigenvalues (except for 0) between a partial symmetric tensor and its Tucker core can be established. The following proposition demonstrates when 0 is always an eigenvalue of the original tensor.

**Proposition 17.** Suppose a given four-way partial symmetric tensor  $\mathcal{T} \in \mathbb{R}^{I_1 \times I_2 \times I_1 \times I_2}$  has an exact independent partial symmetric Tucker decomposition  $\mathcal{T} = \llbracket \mathcal{G}^{ps}; \mathbf{A}, \mathbf{B}, \mathbf{A}, \mathbf{B} \rrbracket$  such that  $\mathcal{G}^{ps} \in \mathbb{R}^{J_1 \times J_2 \times J_1 \times J_2}$ . Either  $J_1 < I_1$  or  $J_2 < I_2$  implies the existence of a zero M-eigenvalue.

The proof is almost identical to that of Proposition 13, and is omitted here.

## 6 Miscellaneous discussions and error estimations

The invariance properties that we have established immediately lead to possible enhancements of many existing bounds. For instance, it is well known that for  $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$  with  $I_1 \leq I_2 \leq \dots \leq I_N$ , it holds that  $\text{rank}_{CP}(\mathcal{X}) \leq I_1 I_2 \dots I_{N-1}$ . Recently, Hu [15] showed that the tensor nuclear norm  $\|\mathcal{X}\|_1$  is upper bounded by  $I_1 I_2 \dots I_{N-1} \cdot \|\mathbf{X}_{(N)}\|_*$ , where  $\|\cdot\|_*$  denotes the nuclear norm of a matrix, and the bound is tight when  $N = 3$ . Now, all these bounds can be sharpened by means of the Tucker rank.

Suppose  $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$  is an  $N$ -way tensor with  $\text{rank}_T(\mathcal{X}) = (R_1, \dots, R_N)$ . Without loss of generality, assume that  $R_1 \leq R_2 \leq \dots \leq R_N$ . Then

- $\text{rank}_{CP}(\mathcal{X}) \leq R_1 R_2 \dots R_{N-1}$ ;
- $\|\mathcal{X}\|_1 \leq R_1 R_2 \dots R_{N-1} \cdot \|\mathbf{X}_{(N)}\|_*$ .

As discussed earlier, tensor related computations such as the CP decompositions, norms or eigenvalues are mostly NP hard [13]. Moreover, exact solution methods, such as the SOS (Sum of Squares) approach to tensor eigenvalue computation (see [8]), are often very sensitive to the size of the underlying tensor. At the same time, the Tucker decomposition involves only matrix operations, hence easy computable. Therefore it is natural to consider a reduction scheme where the tensor computation is only carried out for its core. Before proceeding, we set out to explore if size of the Tucker core of a tensor is indeed typically smaller than the size of the tensor itself. To this end, we find it compelling to test the size reduction on some well studied specific instances of tensors. Below is a summary of our experimental results.

- (*A tensor case studied in [25].*)

This specific tensor is in  $\mathbb{C}^{3 \times 3 \times 3}$ , corresponding to the following polynomial

$$\begin{aligned} \mathcal{T}(\mathbf{x}^4) = & 81x_0^4 + 17x_1^4 + 626x_2^4 - 144x_0x_1^2x_2 + 216x_0^3x_1 - 108x_0^3x_2 + 216x_0^2x_1^2 + 54x_0^2x_2^2 + \\ & 96x_0x_1^3 - 12x_0x_2^3 - 52x_1^3x_2 + 174x_1^2x_2^2 - 508x_1x_2^3 + 72x_0x_1x_2^2 - 216x_0^2x_1x_2. \end{aligned}$$

The dimension of the tensor is 3 while the size of its core tensor is 2.

- (*A tensor case studied in [29].*)

This specific tensor is in  $\mathbb{C}^{5 \times 5 \times 5}$ , with its components given by

$$\mathcal{T}_{i_1 i_2 i_3} = i_1 i_2 i_3 - i_1 - i_2 - i_3 \quad (0 \leq i_1, i_2, i_3 \leq 4).$$

The dimension of the tensor is 5 while the size of its core tensor is 2.

- (Another tensor case studied in [29].)

This specific tensor is in  $\mathbb{C}^{5 \times 5 \times 5 \times 5}$ , with its components given by

$$\mathcal{T}_{i_1 i_2 i_3 i_4} = \tan(i_1 i_2 i_3 i_4) \quad (0 \leq i_1, i_2, i_3, i_4 \leq 4).$$

The dimension of the tensor is 5 while the size of the core tensor is 4.

For some special classes of tensors, it might be possible to estimate the size of its Tucker core.

**Proposition 18.** *Consider  $N$ -th order tensor  $\mathcal{X}$  with a separate structure:*

$$\mathcal{X}_{i_1 \dots i_N} = \sum_{n=1}^N f_n(i_n). \quad (10)$$

*Then the size of its Tucker core is no more than  $(2, \dots, 2)$ . Consider the following  $N$ -th order symmetric real tensor  $\mathcal{T}$  (see [30] or Example 4.12 in [8]):*

$$\mathcal{T}_{i_1 \dots, i_N} = \sin(i_1 + \dots + i_N).$$

*The size of its Tucker core is no more than 2.*

*Proof.* For tensor  $\mathcal{X}$ , it suffices to show that the rank of any mode- $n$  matricization  $\mathbf{X}_{(n)}$  is no more than 2. Note that  $\mathbf{X}_{(n)}$  can be specified componentwise by  $(\mathbf{X}_{(n)})_{i_n, j} = f_n(i_n) + \sum_{k \neq n} f_k(i_k)$  where  $j = 1 + \sum_{k=1, k \neq n}^N (i_k - 1) J_k$  with  $J_k = \prod_{m=1, m \neq n}^{k-1} I_m$ . By constructing vectors  $\mathbf{a} = [f_n(i_n)]$  and  $\mathbf{b} = [\sum_{k \neq n} f_k(i_k)]$ , it follows that  $\mathbf{X}_{(n)} = \mathbf{a} \mathbf{e}^T + \mathbf{e} \mathbf{b}^T$ . Consequently,  $\text{rank}(\mathbf{X}_{(n)}) \leq 2$  and the first half of the conclusion is proved.

To study tensor  $\mathcal{T}$ , due to the symmetric property, without loss of generality it suffices to consider the mode-1 matricization  $\mathbf{T}_{(1)}$  such that

$$\begin{aligned} (\mathbf{T}_{(1)})_{i_1, j} &= \sin(i_1 + \dots + i_N) \\ &= \sin(i_1) \cos(i_2 + \dots + i_N) + \cos(i_1) \sin(i_2 + \dots + i_N) \text{ with } j = 1 + \sum_{k=2}^N (i_k - 1) n^{k-1}, \end{aligned}$$

where  $n$  is the length of the tensor along each direction. Simply letting

$$\mathbf{a} = [\sin(i_1)], \quad \mathbf{b} = [\cos(i_2 + \dots + i_N)], \quad \mathbf{c} = [\cos(i_1)] \text{ and } \mathbf{d} = [\sin(i_2 + \dots + i_N)]$$

yields that

$$\mathbf{T}_{(1)} = \mathbf{a} \mathbf{b}^T + \mathbf{c} \mathbf{d}^T,$$

proving the second half of the proposition. □

As observed above, Tucker core of size no more than 2 is not uncommon. Actually, Examples 2-5 provided in the next section all belong to this category. We shall remark here that there are specific techniques available to solve tensor problems of size 2. For example, computing the Z-eigenvalues of a size 2 tensor is equivalent to finding the common roots of two bivariate polynomials, and a numerical procedure for solving the latter problem was discussed in [36].

Now let us turn to the issue of estimating the error projected on the original tensor while working with a Tucker core approximately. Obviously, errors may occur in the process of Tucker decomposition; so the core tensor that we deal with may not be the true core tensor. The question is: Will the errors expand very quickly? We shall discuss the case for the CP decomposition here. The answer is negative.

**Proposition 19.** *For a given tensor  $\mathcal{T} \in \mathbb{C}^{I_1 \times I_2 \times \dots \times I_N}$ , its independent (but not necessarily exact) Tucker decomposition  $\mathcal{T} = \llbracket \mathcal{G}; \mathbf{A}^{(1)}, \dots, \mathbf{A}^{(N)} \rrbracket$  has the error*

$$Err_1 = \|\mathcal{T} - \mathcal{G} \times_1 \mathbf{A}^{(1)} \cdots \times_N \mathbf{A}^{(N)}\|_F.$$

Now we perform a CP decomposition on  $\mathcal{G}$  and get

$$\tilde{\mathcal{G}} = \sum_{t=1}^r \mathbf{a}^{(1,t)} \circ \dots \circ \mathbf{a}^{(N,t)}$$

with  $Err_2 = \|\mathcal{G} - \tilde{\mathcal{G}}\|_F$ . Then

$$\tilde{\mathcal{T}} = \tilde{\mathcal{G}} \times_1 \mathbf{A}^{(1)} \times_2 \cdots \times_N \mathbf{A}^{(N)}$$

is a CP decomposition of  $\mathcal{T}$  with an error estimation

$$\|\mathcal{T} - \tilde{\mathcal{T}}\|_F \leq Err_1 + Err_2 \prod_{n=1}^N \|\mathbf{A}^{(n)}\|_2.$$

**Proof:** By Lemma 4,

$$\begin{aligned} \tilde{\mathcal{G}} \times_1 \mathbf{A}^{(1)} \times_2 \cdots \times_N \mathbf{A}^{(N)} &= \sum_{t=1}^r \mathbf{a}^{(1,t)} \circ \dots \circ \mathbf{a}^{(N,t)} \times_1 \mathbf{A}^{(1)} \times_2 \cdots \times_N \mathbf{A}^{(N)} \\ &= \sum_{t=1}^r \mathbf{A}^{(1)} \mathbf{a}^{(1,t)} \circ \dots \circ \mathbf{A}^{(N)} \mathbf{a}^{(N,t)}, \end{aligned}$$

which is indeed a CP decomposition of  $\mathcal{T}$ . Moreover, the error of this decomposition is:

$$\begin{aligned} \|\mathcal{T} - \tilde{\mathcal{T}}\|_F &= \|\mathcal{T} - \tilde{\mathcal{G}} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \cdots \times_N \mathbf{A}^{(N)}\|_F \\ &= \|\mathcal{T} - \mathcal{G} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \cdots \times_N \mathbf{A}^{(N)} + (\mathcal{G} - \tilde{\mathcal{G}}) \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \cdots \times_N \mathbf{A}^{(N)}\|_F \\ &\leq \|\mathcal{T} - \mathcal{G} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \cdots \times_N \mathbf{A}^{(N)}\|_F + \|(\mathcal{G} - \tilde{\mathcal{G}}) \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \cdots \times_N \mathbf{A}^{(N)}\|_F \\ &\leq Err_1 + Err_2 \prod_{n=1}^N \|\mathbf{A}^{(n)}\|_2, \end{aligned}$$

where the last inequality is due to Lemma 7. □

We remark that when the Tucker decomposition is exact and orthonormal, then the above proposition reduces to Lemma 2.5 in [22]. This proposition also suggests that one may indeed choose to work with a smaller Tucker core, and the resulting approximative CP decomposition will have a controllable error bound, thanks to an additive rate of the error accumulations.

## 7 Numerical experiments

The goal of this section is to experiment if a reduced Tucker core helps to solve the tensor problem overall. The answer is, interestingly: *it depends*. If we apply the standard Alternating Least Squares (ALS) approach to find the CP decomposition, then [37] reported that more ALS iterations may be required on a compressed core tensor. This is perhaps not very surprising, because in some cases the computational complexity does not necessarily go down with the size per se. Since we are not aware of a standard solver to compute the CP decomposition exactly to do the comparison, we choose to experiment with an ‘easier’ computational object: the Z-eigenvalues and Z-eigenvectors of a symmetric tensor.

Most papers focussed on the computation of the largest or the smallest Z-eigenvalue; see [33, 30, 18]. Most recently, Cui, Dai and Nie [8] proposed an algorithm that can find every Z-eigenvalue of a given tensor. Their idea is to formulate the problem of computing each Z-eigenvalue as a polynomial optimization problem, and then resort to the SOS method that in principle can globally solve any polynomial optimization to optimality. In our numerical experiments, we used the method in [8] to compute all the Z-eigenvalues and Z-eigenvectors for the original tensor and for its Tucker core tensor. In particular, we record both the running time of Algorithm 3.6 in [8] applied to some specific instances and the running time of the same algorithm on the corresponding Tucker core tensor plus the time consumed by the Tucker decomposition. Our code is based on that in [8] with some slight modifications and parameter tunings.

In the following, we choose five testing examples to do this experiment, and report the corresponding numerical results. All of our experiments are run using MATLAB R2013a on a MacBook with an Intel dual core CPU at 1.3 GHz  $\times 2$  and 4 GB of RAM, under an OS X 10.9.5 operating system.

**Example 1.** (*Example 2 in [40]*)

Consider the symmetric tensor  $\mathcal{X} \in \mathbb{R}^{5 \times 5 \times 5 \times 5}$  such that

$$\mathcal{X}(\mathbf{x}^{\circ 4}) = (x_1 + x_2 + x_3 + x_4)^4 + (x_2 + x_3 + x_4 + x_5)^4.$$

The dimension of the original tensor is 5 while the dimension of its core is 2. It took 30.03 seconds to compute directly on the tensor itself while the computation on its core (including the Tucker

decomposition) took only 3.29 seconds. The computed eigenvalues and eigenvectors of the original tensor are in Table 1 and that of the core tensor are in Table 2.

Eigenvalues	Eigenvectors				
24.500	-0.267	-0.535	-0.535	-0.535	-0.267
0.500	0.707	-0.000	-0.000	-0.000	-0.707
0.000	-0.073	0.618	-0.751	0.211	-0.067

Table 1: Eigenvalues and Eigenvectors of the Original Tensor in Example 1

Eigenvalues	Eigenvectors	
24.500	-1.000	0.000
0.500	-0.000	-1.000
0.000	-0.354	-0.935

Table 2: Eigenvalues and Eigenvectors of the Core Tensor in Example 1

**Example 2.** (Example 3.5 in [30])

Consider the symmetric tensor  $\mathcal{X} \in \mathbb{R}^{n \times n \times n}$  such that

$$\mathcal{X}_{ijk} = \frac{(-1)^i}{i} + \frac{(-1)^j}{j} + \frac{(-1)^k}{k} \quad (1 \leq i, j, k \leq n).$$

For the case  $n = 8$ , the dimension of the original tensor is 8 while the dimension of its core is 2. The direct computation on the original tensor took 17670.71 seconds, and the computation on its core took 2.66 seconds. The resulting eigenvalues and eigenvectors of the original tensor and its core tensor are in Table 3 and Table 4 respectively.

Eigenvalues	Eigenvectors							
14.436	-0.687	-0.066	-0.411	-0.169	-0.356	-0.204	-0.332	-0.221
8.586	-0.225	0.579	0.132	0.445	0.203	0.400	0.234	0.378
0.000	-0.335	0.283	0.252	0.256	-0.030	-0.751	-0.013	0.337
-14.436	-0.687	-0.066	-0.411	-0.169	-0.356	-0.204	-0.332	-0.221
-8.586	-0.225	0.579	0.132	0.445	0.203	0.400	0.234	0.378

Table 3: Eigenvalues and Eigenvectors of the Original Tensor in Example 2

**Example 3.** (Example 4.14 in [8])

Eigenvalues	Eigenvectors	
14.436	-0.985	-0.175
8.586	0.488	-0.873
-0.000	0.344	0.939
-14.436	-0.985	-0.175
-8.586	0.488	-0.873

Table 4: Eigenvalues and Eigenvectors of the Core Tensor in Example 2

Consider the symmetric tensor  $\mathcal{X} \in \mathbb{R}^{n \times n \times n \times n \times n}$ :

$$\mathcal{X}_{i_1 i_2 i_3 i_4 i_5} = \ln(i_1) + \ln(i_2) + \ln(i_3) + \ln(i_4) + \ln(i_5) \quad (1 \leq i_1, i_2, i_3, i_4, i_5 \leq n).$$

For the case  $n = 4$ , the dimension of the original tensor is 4 while the dimension of its core is 2. The direct computation on the original tensor took 186.58 seconds, and the computation on its core took 5.23 seconds. The eigenvalues and eigenvectors of the original tensor and its core tensor are in Table 5 and Table 6 respectively.

Eigenvalues	Eigenvectors			
132.307	0.403	0.484	0.532	0.566
0.707	-0.905	-0.308	0.041	0.289
0.001	0.565	0.254	-0.022	-0.785
-132.307	0.403	0.484	0.532	0.566
-0.707	-0.905	-0.308	0.041	0.289

Table 5: Eigenvalues and Eigenvectors of the Original Tensor in Example 3

Eigenvalues	Eigenvectors	
132.307	1.000	-0.000
0.707	-0.329	-0.944
-0.000	0.127	-0.992
-132.307	1.000	-0.000
-0.707	-0.329	-0.944

Table 6: Eigenvalues and Eigenvectors of the Core Tensor in Example 3

**Example 4.** (Example 4.12 in [8]) Consider the symmetric tensor  $\mathcal{X} \in \mathbb{R}^{n \times n \times n \times n}$ :

$$\mathcal{X}_{i_1 i_2 i_3 i_4} = \sin(i_1 + i_2 + i_3 + i_4) \quad (1 \leq i_1, i_2, i_3, i_4 \leq n).$$

For the case  $n = 4$ , the dimension of the original tensor is 4 while the dimension of its core is 2. The direct computation on the original tensor took 74.75 seconds, and the computation on its core took 5.48 seconds. The eigenvalues and eigenvectors of the original tensor and its core tensor are in Table 7 and Table 8 respectively.

Eigenvalues	Eigenvectors			
4.632	0.500	-0.133	-0.644	-0.563
2.991	-0.347	-0.766	-0.482	0.246
0.000	0.623	-0.587	0.512	0.068
-5.645	-0.629	-0.485	0.105	0.598
-2.525	0.083	-0.621	-0.755	-0.194

Table 7: Eigenvalues and Eigenvectors of the Original Tensor in Example 4

Eigenvalues	Eigenvectors	
4.632	-0.821	0.571
2.991	-0.485	-0.874
-5.645	0.195	-0.981
-2.525	-0.969	-0.247

Table 8: Eigenvalues and Eigenvectors of the Core Tensor in Example 4

To conclude, the computational time spent on finding the Z-eigenvalues and Z-eigenvectors of a tensor can be substantially reduced if we turn to its Tucker core tensor instead. Table 9 summarizes the recorded computational times for the above examples 1 – 5.

**Acknowledgements.** We would like to thank Chunfeng Cui for sharing with us the codes on computing all Z-eigenvalues, and we thank Shmuel Friedland, Lek-Heng Lim, Jiawang Nie and Nikos Sidiropoulos for the fruitful discussions on the topics related to this paper.

<i>Example</i>	<i>Tensor Order</i>	<i>Tensor Size</i>	<i>Core Size</i>	<i>CPU for Tensor</i>	<i>CPU for Core</i>
1	4	5	2	30.03s	3.29s
2	3	8	2	17670.70s	2.66s
3	5	4	2	186.58s	5.23s
4	4	4	2	74.75s	5.48s

Table 9: Computational Time Comparison

## References

- [1] B. W. Bader and T. G. Kolda. Algorithm 862: Matlab tensor classes for fast algorithm prototyping. *ACM Trans. Math. Software*, 32:635–653, 2006.
- [2] D. Bini. *The Role of Tensor Rank in the Complexity Analysis of Bilinear Forms*. ICIAM07, Zürich, Switzerland, 2007.
- [3] L. Bloy and R. Verma. On computing the underlying fiber directions from the diffusion orientation distribution function. In *International Conference on Medical Image Computing and Computer-assisted Intervention-part I*, pages 1–8, 2008.
- [4] R. Bro and C. A. Andersson. Improving the speed of multi-way algorithms: Part ii. *Compression, Chemometrics and Intelligent Laboratory Systems*, 42:105–113, 1998.
- [5] J. D. Carroll and J. J. Chang. Analysis of individual differences in multidimensional scaling via an n-way generalization of "eckart-young" decomposition. *Psychometrika*, 35:283–319, 1970.
- [6] J. D. Carroll, S. Pruzansky, and J. B. Kruskal. Candelinc: A general approach to multidimensional analysis of many-way arrays with linear constraints on parameters. *Psychometrika*, 45:3–24, 1980.
- [7] P. Comon, G. Golub, L.H. Lim, and B. Mourrain. Symmetric tensors and symmetric tensor rank. *SIAM J. Matrix Anal. Appl.*, 30(3):1254–1279, 2008.
- [8] C. Cui, Y. Dai, and J. Nie. All real eigenvalues of symmetric tensors. *SIAM J. Matrix Anal. Appl.*, 35:1582–1601, 2014.
- [9] S. Friedland and L.-H. Lim. Computational complexity of tensor nuclear norm. *Working Paper*, 2014.
- [10] A. Ghosh, E. Tsigaridas, M. Descoteaux, P. Comon, B. Mourrain, and R. Deriche. A polynomial based approach to extract the maxima of an antipodally symmetric spherical function and its application to extract fiber directions from the orientation distribution function in diffusion mri. In *Computational Diffusion MRI Workshop (CDMRI08)*, New York, 2008.
- [11] A. Gröthendieck. Produits tensoriels topologiques et espaces nucléaires. *Mem. Amer. Math. Soc.*, 16:140 pages, 1955.
- [12] R. A. Harshman. Foundations of the parafac procedure: Models and conditions for an "explanatory" multi-modal factor analysis. *UCLA Working Paper*, 1969.

- [13] C.J. Hillar and L.-H. Lim. Most tensor problems are NP-hard. *J. ACM*, 60(6):Art. 45, 39 pages, 2013.
- [14] J. J. Hilling and A. Sudbery. The geometric measure of multipartite entanglement and the singular values of a hypermatrix. *J. Math. Phys.*, 51(7):165–169, 2010.
- [15] S. Hu. Relations of the nuclear norms of a tensor and its matrix flattenings. *Linear Algebra and its Applications*, 478:188–199, 2015.
- [16] S. Ji, K.-F. Sze, Z. Zhou, A. M.-C. So, and Y. Ye. Beyond convex relaxation: A polynomial-time non-convex optimization approach to network localization. In *Proceedings of the 32nd IEEE International Conference on Computer Communications (INFOCOM 2013)*, 2013.
- [17] B. Jiang, Z. Li, and S. Zhang. On cones of nonnegative quartic forms. *Found. Comput. Math. to appear*, 2015.
- [18] B. Jiang, S. Ma, and S. Zhang. Tensor principal component analysis via convex optimization. *Math. Program.*, 150:423–457, 2015.
- [19] V. Khoromskaia. *Numerical Solution of the Hartree-Fock Equation by Multilevel Tensor-structured Methods*. PhD thesis, TU Berlin, 2010.
- [20] B. N. Khoromskij. Structured rank- $(r_1, \dots, r_d)$  decomposition of function-related tensors in  $\mathbb{R}^d$ . *Comp. Meth. in Appl. Math.*, 6:194–220, 2006.
- [21] B. N. Khoromskij. Tensor numerical methods for high-dimensional PDEs: Basic theory and initial applications. *ESAIM: Proceedings and Surveys*, 48:1–28, 2015.
- [22] B. N. Khoromskij and V. Khoromskaia. Low rank Tucker-type tensor approximation to classical potentials. *Central European J. Math.*, 5:523–550, 2007.
- [23] T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM Review*, 51(3):455–500, 2009.
- [24] T. G. Kolda and T. Gibson. Multilinear operators for higher-order decompositions. Technical report, Sandia National Laboratories, 2006.
- [25] J. M. Landsberg. Tensors: Geometry and applications. *Graduate Studies in Mathematics*, 2012.
- [26] L. De Lathauwer, B. De Moor, and J. Vandewalle. A multilinear singular value decomposition. *SIAM J. Matrix Anal. Appl.*, pages 1253–1278.
- [27] L. De Lathauwer, B. De Moor, and J. Vandewalle. On the best rank-1 and rank- $(r_1, r_2, \dots, r_n)$  approximation of higher-order tensors. *SIAM J. Matrix Anal. Appl.*, pages 1324–1342.

- [28] L.-H. Lim. Singular values and eigenvalues of tensors: A variational approach. In *Computational Advances in Multi-Sensor Adaptive Processing, 2005 1st IEEE International Workshop on*, pages 129–132. IEEE, 2005.
- [29] J. Nie. Generating polynomials and symmetric tensor decompositions. *Found. Comput. Math.*, to appear, 2015.
- [30] J. Nie and L. Wang. Semidefinite relaxations for best rank-1 tensor approximations. *SIAM J. Matrix Anal. Appl.*, 35:1155–1179, 2014.
- [31] L. Qi. Eigenvalues of a real supersymmetric tensor. *J. Symbolic Comput.*, 40:1302–1324, 2005.
- [32] L. Qi, H.-H. Dai, and D. Han. Conditions for strong ellipticity and m-eigenvalues. *Front. Math. China*, 4:349–364, 2009.
- [33] L. Qi, F. Wang, and Y. Wang. Z-eigenvalue methods for a global polynomial optimization problem. *Math. Program.*, 118:301–316, 2009.
- [34] L. Qi, G. Yu, and E. X. Wu. Higher order positive semi-definite diffusion tensor imaging. *SIAM J. Imaging Sci.*, 3(3):416–433, 2010.
- [35] R. Schatten. *A Theory of Cross-Spaces*. Princeton University Press, NJ, 1950.
- [36] L. Sorber, M. Van Barel, and L. De Lathauwer. Numerical solution of bivariate and polyanalytic polynomial systems. *SIAM J. Numer. Anal.*, pages 1551–1572.
- [37] G. Tomasi and R. Bro. A comparison of algorithms for fitting the parafac model. *Comput. Statist. Data Anal.*, 50:1700–1734, 2006.
- [38] L. R. Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31:279–311, 1966.
- [39] H. Wang and N. Ahuja. Compact representation of multidimensional data using tensor rank-one decomposition. *International Conference on Pattern Recognition*, 1:44–47, 2014.
- [40] J. Xie and A. Chang. On the z-eigenvalues of the signless laplacian tensor for an even uniform hypergraph. *Numer. Linear Algebra Appl.*, 20(6):1030–1045, 2013.
- [41] M. Yuan and C.-H. Zhang. On tensor completion via nuclear norm minimization. *Found. Comput. Math.*, to appear.