



On norm compression inequalities for partitioned block tensors

Zhening Li¹ · Yun-Bin Zhao²

Received: 5 February 2019 / Revised: 4 October 2019 / Accepted: 9 January 2020 /

Published online: 18 February 2020

© The Author(s) 2020

Abstract

When a tensor is partitioned into subtensors, some tensor norms of these subtensors form a tensor called a norm compression tensor. Norm compression inequalities for tensors focus on the relation of the norm of this compressed tensor to the norm of the original tensor. We prove that for the tensor spectral norm, the norm of the compressed tensor is an upper bound of the norm of the original tensor. This result can be extended to a general class of tensor spectral norms. We discuss various applications of norm compression inequalities for tensors. These inequalities improve many existing bounds of tensor norms in the literature, in particular tightening the general bound of the tensor spectral norm via tensor partitions. We study the extremal ratio between the spectral norm and the Frobenius norm of a tensor space, provide a general way to estimate its upper bound, and in particular, improve the current best upper bound for third order nonnegative tensors and symmetric tensors. We also propose a faster approach to estimate the spectral norm of a large tensor or matrix via sequential norm compression inequalities with theoretical and numerical evidence. For instance, the complexity of our algorithm for the matrix spectral norm is $O(n^{2+\epsilon})$ where ϵ ranges from 0 to 1 depending on the partition and the estimate ranges correspondingly from some close upper bound to the exact spectral norm.

Keywords Norm compression inequality · Spectral norm · Tensor partition · Block tensor · Rank-one approximation

This work was partially supported by the Engineering and Physical Sciences Research Council (EPSRC) under Grant #EP/K00946X/1.

✉ Zhening Li
zheningli@gmail.com

Yun-Bin Zhao
y.zhao.2@bham.ac.uk

¹ School of Mathematics and Physics, University of Portsmouth, Portsmouth, Hampshire PO1 3HF, UK

² School of Mathematics, University of Birmingham, Birmingham, West Midlands B15 2TT, UK

Mathematics Subject Classification 15A69 · 15A60 · 15A45 · 90C59

1 Introduction

With the advances in data collection and storage capabilities, massive multiway (tensor) data are being generated in a wide range of emerging applications [25]. Multilinear algebra and tensor computations are playing more and more important roles in dealing with multiway data in recent years. Computing tensor norms are evidently essential in many tensor computation problems. However, most tensor norms are NP-hard to compute [19], such as the tensor spectral norm [17] and the tensor nuclear norm [12]. As a useful method to approximate matrix norms via block matrices, the computation of tensor norms via block tensors is straightforward and becomes increasingly important within the field of numerical linear algebra [9,29,41,42]. When a tensor is partitioned into subtensors, not necessarily having the same size, some tensor norms of these subtensors form a tensor called a norm compression tensor. Norm compression inequalities for tensors focus on the relation of a norm of this compressed tensor to the norm of the original tensor. These inequalities straightforwardly provide a handy tool to bound and estimate norms of large tensors via norms of smaller subtensors.

In the case of matrices, tensors of order two, norm compression inequalities have been well studied since Bhatia and Kittaneh [6]. Such inequalities have several applications in, for instance, quantum information theory [2,5] and covariance estimation [7]. An overview of several norm compression inequalities for matrices can be found in [2] and references therein. One important result is due to King [23]: If a matrix M is partitioned into 2×2 blocks $\begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}$, then

$$\begin{aligned} \|M\|_{p_s} &\geq \left\| \begin{pmatrix} \|M_{11}\|_{p_s} & \|M_{12}\|_{p_s} \\ \|M_{21}\|_{p_s} & \|M_{22}\|_{p_s} \end{pmatrix} \right\|_{p_s} & 1 \leq p \leq 2, \\ \|M\|_{p_s} &\leq \left\| \begin{pmatrix} \|M_{11}\|_{p_s} & \|M_{12}\|_{p_s} \\ \|M_{21}\|_{p_s} & \|M_{22}\|_{p_s} \end{pmatrix} \right\|_{p_s} & 2 \leq p \leq \infty, \end{aligned} \quad (1)$$

where $\|\cdot\|_{p_s}$ stands for the Schatten p -norm of a matrix, i.e., the L_p -norm of the vector consisting of all the singular values of the matrix. However, there exists an example [3] of a partitioned 3×3 block matrix such that inequalities of type (1) fail to hold. A conjecture of this type to hold for $2 \times m$ blocks was proposed and its several special cases were proven by Audenaert [3]. There are two notable special cases for $2 \times m$ blocks to hold, namely when the matrix M is positive semidefinite [23], and when the blocks of M are all diagonal matrices [24].

Among all Schatten p -norms of a matrix, three of them are particularly important, namely the spectral norm ($p = \infty$), the Frobenius Norm ($p = 2$), and the nuclear norm ($p = 1$). The Schatten 2-norm coincides with the Frobenius norm of a matrix, and this makes the corresponding norm compression inequality trivial, which is actually an equality. In fact, the norm compression inequality of type (1) holds for any $m_1 \times m_2$ blocks when $p = \infty$. This result, to the best of our knowledge, was not studied in the literature, and it is a special case of the main result in this paper. For higher order

(order three or higher) tensors, the Schatten p -norms are not well defined unless $p = \infty, 1$ [12], corresponding to the tensor spectral norm and nuclear norm, respectively. As both the spectral norm and the nuclear norm of a tensor are NP-hard to compute while that of a matrix can be computed in polynomial time, matrix unfoldings have become a major approach dealing with various problems involving these tensor norms, no matter in theory and in practice. Relations of these norms of a tensor and its matrix unfoldings have been studied in [12,17,21]. A generalization of such relations under tensor unfoldings has been studied by Want et al. [48].

Block tensors are becoming increasingly important. They have been used in large tensor factorizations [39], tensor decompositions [36], tensor optimization [47], and imaging processing [10]. Ragnarsson and Van Loan [41] developed an infrastructure that supports reasoning about block tensor computations. They [42] further applied block tensors to symmetric embeddings of tensors. Extending block tensors, Li [29] proposed more general concepts of tensor partitions and provided bounds of the spectral norm and the nuclear norm of a tensor via norms of subtensors in a regular partition of the tensor. The results were further generalized to the spectral p -norm and the nuclear p -norm of a tensor and to arbitrary partitions of the tensor [9]. This paper explores the structure of block tensors instead of treating subtensors merely as elements in [9,29] and proposes more accurate estimation of the spectral norm of a tensor, albeit block tensors are special but most common types of regular partitions [29] and arbitrary partitions [9]. It is worth mentioning that bounds of the tensor spectral p -norm have been extensively studied in the literature [16–18,20,38,44,48], in particular in the area of polynomial optimization [30].

In this paper, we study norm compression inequalities for tensors. We prove that for any block partitioned tensor, no matter how many blocks, the spectral norm of its norm compressed tensor is an upper bound of the spectral norm of the original tensor. The result can be generalized to a wider class of tensor spectral norms. These norm compression inequalities improve many existing bounds of tensor spectral norms in the literature, including the recent bounds via tensor partitions studied in [9,29]. We discuss two important applications of our results. The first one is on the extremal ratio between the spectral norm and the Frobenius norm of a tensor space. We provide a general methodology to compute upper bounds of this ratio, and in particular to improve the current best upper bound for third order nonnegative tensors and symmetric tensors. The second one is to estimate the spectral norm of a large tensor or matrix via sequential norm compression inequalities. Some numerical evidence is provided to justify our methodology.

This paper is organized as follows. We start with the preparation of various notations, definitions and properties of tensor spectral norms in Sect. 2. In Sect. 3, we present our main result on norm compression inequalities for tensors, and in Sect. 4, we discuss how our main inequalities lead to various other bounds of tensor spectral norms in the literature. For applications, the study of the extremal ratio between the spectral norm and the Frobenius norm of a tensor space is presented in Sect. 5, and estimating the tensor and the matrix spectral norms is discussed in Sect. 6.

2 Preparation

Throughout this paper, we uniformly use the lower case letters (e.g., x), the boldface lower case letters (e.g., $\mathbf{x} = (x_i)$), the capital letters (e.g., $X = (x_{ij})$), and the calligraphic letters (e.g., $\mathcal{X} = (x_{i_1 i_2 \dots i_d})$) to denote scalars, vectors, matrices, and higher order (order three or more) tensors, respectively. Denote $\mathbb{R}^{n_1 \times n_2 \times \dots \times n_d}$ to be the space of d -th order real tensors of dimension $n_1 \times n_2 \times \dots \times n_d$. The same notations apply for a vector space and a matrix space when $d = 1$ and $d = 2$, respectively. Unless otherwise specified, the order of a general tensor in this paper is always denoted by d and the dimension of its mode- k is always denoted by n_k for $k = 1, 2, \dots, d$. Given a d -th order tensor space $\mathbb{R}^{n_1 \times n_2 \times \dots \times n_d}$, we denote $\mathbb{I}^k := \{1, 2, \dots, n_k\}$ to be the index set of mode- k for $k = 1, 2, \dots, d$. Trivially, $\mathbb{I}^1 \times \mathbb{I}^2 \times \dots \times \mathbb{I}^d$ becomes the index set of the entries of a tensor in $\mathbb{R}^{n_1 \times n_2 \times \dots \times n_d}$. Denote \mathbb{N} to be the set of positive integers and denote $\mathbb{P} = [1, \infty]$, the interval where the L_p -norm of a vector is well defined when $1 \leq p \leq \infty$.

The Frobenius inner product of two tensors $\mathcal{U}, \mathcal{V} \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_d}$ is defined as:

$$\langle \mathcal{U}, \mathcal{V} \rangle := \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \dots \sum_{i_d=1}^{n_d} u_{i_1 i_2 \dots i_d} v_{i_1 i_2 \dots i_d}.$$

Its induced Frobenius norm is naturally defined as $\|\mathcal{T}\|_2 := \sqrt{\langle \mathcal{T}, \mathcal{T} \rangle}$. When $d = 1$, the Frobenius norm is reduced to the Euclidean norm of a vector. In a similar vein, we may define the L_p -norm of a tensor (also known as the Hölder p -norm [33]) for $p \in \mathbb{P}$ by looking at a tensor as a vector, as follows:

$$\|\mathcal{T}\|_p := \left(\sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \dots \sum_{i_d=1}^{n_d} |t_{i_1 i_2 \dots i_d}|^p \right)^{\frac{1}{p}}. \tag{2}$$

A rank-one tensor, also called a simple tensor, is a tensor that can be written as outer products of vectors

$$\mathcal{T} = \mathbf{x}^1 \otimes \mathbf{x}^2 \otimes \dots \otimes \mathbf{x}^d.$$

It is easy to verify that $\|\mathcal{T}\|_p = \prod_{k=1}^d \|\mathbf{x}^k\|_p$ for all $p \in \mathbb{P}$. When $d = 2$, a rank-one tensor is reduced to the well known concept of a rank-one matrix.

The spectral norm of a tensor is an important measure of the tensor.

Definition 2.1 For a given tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_d}$, the spectral norm of \mathcal{T} , denoted by $\|\mathcal{T}\|_\sigma$, is defined as

$$\|\mathcal{T}\|_\sigma := \max \left\{ \left\langle \mathcal{T}, \mathbf{x}^1 \otimes \mathbf{x}^2 \otimes \dots \otimes \mathbf{x}^d \right\rangle : \|\mathbf{x}^k\|_2 = 1, \quad k = 1, 2, \dots, d \right\},$$

and the nuclear norm of \mathcal{T} , denoted by $\|\mathcal{T}\|_*$, is defined as

$$\|\mathcal{T}\|_* := \min \left\{ \sum_{i=1}^r |\lambda_i| : \mathcal{T} = \sum_{i=1}^r \lambda_i \mathbf{x}_i^1 \otimes \mathbf{x}_i^2 \otimes \dots \otimes \mathbf{x}_i^d, \|\mathbf{x}_i^k\|_2 = 1 \right. \\ \left. \text{for all } k = 1, 2, \dots, d \text{ and } i = 1, 2, \dots, r \in \mathbb{N} \right\}. \tag{3}$$

Essentially, $\|\mathcal{T}\|_\sigma$ is the maximal value of the Frobenius inner product between \mathcal{T} and a rank-one tensor whose Frobenius norm is one. Computing the tensor spectral norm is also known as the Euclidean spherical constrained multilinear form maximization problem [30]. The tensor nuclear norm is the dual norm to the tensor spectral norm, and vice versa [11,34], i.e.,

$$\|\mathcal{T}\|_\sigma = \max_{\|\mathcal{X}\|_* \leq 1} \langle \mathcal{T}, \mathcal{X} \rangle \text{ and } \|\mathcal{T}\|_* = \max_{\|\mathcal{X}\|_\sigma \leq 1} \langle \mathcal{T}, \mathcal{X} \rangle.$$

Apart from L_p -norms defined via tensor entries in (2), there is another set of norms for matrices called Schatten p -norms, defined by the L_p -norm of the vector consisting all singular values of a matrix. In particular, the spectral norm and the nuclear norm of a matrix is nothing but the Schatten ∞ -norm and the Schatten 1-norm of the matrix, respectively. In fact, the study of norm compression inequalities for matrices in the literature is mostly on Schatten p -norms, including the spectral norm and the nuclear norm as special cases. However, one cannot get a Schatten p -norm for tensors in the manner defining the nuclear norm of a tensor in (3). If the L_1 -norm expression $\sum_{i=1}^r |\lambda_i|$ in (3) is replaced by an L_p -norm expression $(\sum_{i=1}^r |\lambda_i|^p)^{\frac{1}{p}}$ for any $1 < p < \infty$, the minimum is always zero [12].

One can actually extend the tensor spectral norm and tensor nuclear norm in Definition 2.1 as follows.

Definition 2.2 For a given tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_d}$ and a vector $\mathbf{p} = (p_1, p_2, \dots, p_d) \in \mathbb{P}^d$, the spectral \mathbf{p} -norm of \mathcal{T} , denoted by $\|\mathcal{T}\|_{\mathbf{p}_\sigma}$, is defined as

$$\|\mathcal{T}\|_{\mathbf{p}_\sigma} := \max \left\{ \langle \mathcal{T}, \mathbf{x}^1 \otimes \mathbf{x}^2 \otimes \dots \otimes \mathbf{x}^d \rangle : \|\mathbf{x}^k\|_{p_k} = 1, \quad k = 1, 2, \dots, d \right\}.$$

and the nuclear \mathbf{p} -norm of \mathcal{T} , denoted by $\|\mathcal{T}\|_{\mathbf{p}_*}$, is defined as

$$\|\mathcal{T}\|_{\mathbf{p}_*} := \min \left\{ \sum_{i=1}^r |\lambda_i| : \mathcal{T} = \sum_{i=1}^r \lambda_i \mathbf{x}_i^1 \otimes \mathbf{x}_i^2 \otimes \dots \otimes \mathbf{x}_i^d, \|\mathbf{x}_i^k\|_{p_k} = 1 \right. \\ \left. \text{for all } k = 1, 2, \dots, d \text{ and } i = 1, 2, \dots, r \in \mathbb{N} \right\}.$$

In particular, the spectral $(2, 2, \dots, 2)$ -norm and the nuclear $(2, 2, \dots, 2)$ -norm of a tensor are the usual spectral norm and nuclear norm of the tensor, respectively. The

tensor spectral \mathbf{p} -norm was firstly defined at the same time as the tensor spectral norm by Lim [32] in 2005. Computation of the spectral \mathbf{p} -norm for nonnegative tensors was discussed in [13]. When $p_1 = p_2 = \dots = p_d = p$, the tensor spectral p -norm and nuclear p -norm were studied in [9,12], which are denoted by $\|\cdot\|_{p_\sigma}$ and $\|\cdot\|_{p_*}$, respectively. Similar to the tensor spectral norm and nuclear norm, the tensor norms in Definition 2.2 are primal-dual pairs.

Lemma 2.3 *For given d -th order tensors \mathcal{T} and \mathcal{X} in a same tensor space and $\mathbf{p} \in \mathbb{P}^d$, it follows that*

$$\langle \mathcal{T}, \mathcal{X} \rangle \leq \|\mathcal{T}\|_{p_\sigma} \|\mathcal{X}\|_{p_*},$$

and further

$$\|\mathcal{T}\|_{p_\sigma} = \max_{\|\mathcal{X}\|_{p_*} \leq 1} \langle \mathcal{T}, \mathcal{X} \rangle,$$

$$\|\mathcal{T}\|_{p_*} = \max_{\|\mathcal{X}\|_{p_\sigma} \leq 1} \langle \mathcal{T}, \mathcal{X} \rangle.$$

This duality can be proved similarly to the case when all p_k 's are equal [9, Lemma 2.5] and is thus omitted.

The spectral \mathbf{p} -norm and nuclear \mathbf{p} -norm of a tensor are in general very difficult to compute. For the computational complexity for various \mathbf{p} 's and orders of the tensor, one is referred to Friedland and Lim [12]. It is worth mentioning that computing these norms for a rank-one tensor admits a closed form.

Proposition 2.4 *If a tensor \mathcal{T} is rank-one, say $\mathcal{T} = \mathbf{x}^1 \otimes \mathbf{x}^2 \otimes \dots \otimes \mathbf{x}^d$, then for any $\mathbf{p} \in \mathbb{P}^d$,*

$$\|\mathcal{T}\|_{p_\sigma} = \prod_{k=1}^d \|\mathbf{x}^k\|_{q_k},$$

$$\|\mathcal{T}\|_{p_*} = \prod_{k=1}^d \|\mathbf{x}^k\|_{p_k},$$

where $\frac{1}{p_k} + \frac{1}{q_k} = 1$ for $k = 1, 2, \dots, d$.

The proof is left to interested readers. Therefore, $\|\mathcal{T}\|_{p_\sigma}$ can be taken as the maximal value of the Frobenius inner product between \mathcal{T} and a rank-one tensor whose spectral \mathbf{q} -norm is one.

3 Norm compression inequalities for tensors

To study norm compression inequalities for tensors, we first introduce tensor partitions. One important class of tensor partitions, block tensors, was proposed and studied

by Ragnarsson and Van Loan [41]. It is a straightforward generalization of block matrices. Li [29] proposed three types of tensor partitions, namely, modal partitions (an alternative name for block tensors), regular partitions, and tensor partitions, with the later generalizing the former. A more general class of partitions, called arbitrary partitions, was proposed and studied by Chen and Li [9].

Norm compression inequalities are established on block tensors, which are also called modal partitions as they are constructed by partitions of the index sets of tensor modes. Given a tensor $\mathcal{T} = (t_{i_1 i_2 \dots i_d}) \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_d}$, for every $1 \leq k \leq d$, the indices of its mode k can be partitioned into r_k nonempty sets, i.e.,

$$\mathbb{I}^k = \{1, 2, \dots, n_k\} = \bigcup_{j=1}^{r_k} \mathbb{I}_j^k \quad \text{and} \quad \mathbb{I}_i^k \cap \mathbb{I}_j^k = \emptyset \text{ if } i \neq j.$$

For simplicity, we assume that the indices in \mathbb{I}_i^k are consecutive and \mathbb{I}_i^k 's monotonically increase as i increases, since this can be done easily via indices relabeling without affecting tensor norms. Any $(\mathbb{J}_1, \mathbb{J}_2, \dots, \mathbb{J}_d)$ where $\mathbb{J}_k \subseteq \mathbb{I}^k$ for $k = 1, 2, \dots, d$ uniquely defines a subtensor of \mathcal{T} by only keeping indices in \mathbb{J}_k for mode k of \mathcal{T} , i.e.,

$$\mathcal{T}(\mathbb{J}_1, \mathbb{J}_2, \dots, \mathbb{J}_d) = \left((t_{i_1 i_2 \dots i_d})_{i_k \in \mathbb{J}_k, k=1,2,\dots,d} \right) \in \mathbb{R}^{|\mathbb{J}_1| \times |\mathbb{J}_2| \times \dots \times |\mathbb{J}_d|}.$$

Definition 3.1 The partition $\left\{ \mathcal{T} \left(\mathbb{I}_{j_1}^1, \mathbb{I}_{j_2}^2, \dots, \mathbb{I}_{j_d}^d \right) : 1 \leq j_k \leq r_k, k = 1, 2, \dots, d \right\}$ is called a modal partition of a tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_d}$, where $\{\mathbb{I}_1^k, \mathbb{I}_2^k, \dots, \mathbb{I}_{r_k}^k\}$ is a partition of \mathbb{I}^k for $k = 1, 2, \dots, d$.

In our words, a partitioned block tensor is a tensor that has been modal partitioned. Trivially for $d = 2$, a block matrix can be obtained by a modal partition, i.e., partitions of row indices and column indices. We remark that a subtensor in a modal partition of a tensor may not possess the same order of the original tensor. If some \mathbb{I}_j^k contains only one index, it causes the disappearance of mode k and reduces the order of the subtensor by one. However, we still treat this subtensor as a d -th order tensor by keeping the dimension of mode k being one. For instance, we can always treat a scalar as a one-dimensional vector, or a one-by-one matrix.

In order to present the proof of our main result (Theorem 3.3) clearly as well as to provide a better picture of modal partitions, we now discuss tensor cuts. Given a d -th order tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_d}$, a mode- k tensor cut, cuts the tensor \mathcal{T} at mode k into two subtensors \mathcal{T}_1 and \mathcal{T}_2 , denoted by $\mathcal{T} = \mathcal{T}_1 \vee_k \mathcal{T}_2$, where

$$\mathcal{T}_1 \in \mathbb{R}^{n_1 \times \dots \times n_{k-1} \times \ell_1 \times n_{k+1} \times \dots \times n_d}, \quad \mathcal{T}_2 \in \mathbb{R}^{n_1 \times \dots \times n_{k-1} \times \ell_2 \times n_{k+1} \times \dots \times n_d}, \quad \text{and} \quad \ell_1 + \ell_2 = n_k.$$

The same notation can be used to cut a matrix and cut a vector. In particular, for a first order tensor, a vector $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x} = \mathbf{x}_1 \vee_1 \mathbf{x}_2$ is exactly same as $\mathbf{x}^T = (\mathbf{x}_1^T, \mathbf{x}_2^T)$. The mode subscript of \vee in a tensor cut is sometimes omitted for clearer presentations. For instance, $\mathcal{T} = \mathcal{T}_1 \vee \mathcal{T}_2$ implies that there exists $k \in \mathbb{N}$ such that $\mathcal{T} = \mathcal{T}_1 \vee_k \mathcal{T}_2$. Obviously, the operation \vee is not commutative and associative in general. Once the

notation \bigvee_k is applied, the dimensions of its associated two tensors must be the same in every mode except mode k . With this handy notation, a block tensor via a modal partition (Definition 3.1) can be simply written as

$$\mathcal{T} = \bigvee_{j_1=1}^{r_1} \bigvee_{j_2=1}^{r_2} \cdots \bigvee_{j_d=1}^{r_d} \mathcal{T} \left(\mathbb{I}_{j_1}^1, \mathbb{I}_{j_2}^2, \dots, \mathbb{I}_{j_d}^d \right).$$

The following norm compression identity for vectors is straightforward.

Lemma 3.2 *If a vector $\mathbf{x} = \bigvee_{j=1}^r \mathbf{x}_j \in \mathbb{R}^n$, then for any $p \in \mathbb{P}$,*

$$\|\mathbf{x}\|_p = \left\| \bigvee_{j=1}^r \|\mathbf{x}_j\|_p \right\|_p.$$

Proof Denote the vector

$$\mathbf{y} = \bigvee_{j=1}^r \|\mathbf{x}_j\|_p = (\|\mathbf{x}_1\|_p, \|\mathbf{x}_2\|_p, \dots, \|\mathbf{x}_r\|_p)^T.$$

We have

$$\|\mathbf{y}\|_p = \left(\sum_{j=1}^r \|\mathbf{x}_j\|_p^p \right)^{\frac{1}{p}} = \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}} = \|\mathbf{x}\|_p.$$

□

Though in practice one is often interested in the spectral norm of a tensor rather than general spectral p -norms of the tensor, we present our main result for the general case. It obviously applies to the tensor spectral norm when $\mathbf{p} = (2, 2, \dots, 2)$.

Theorem 3.3 *If $\mathcal{T} = \bigvee_{j_1=1}^{r_1} \bigvee_{j_2=1}^{r_2} \cdots \bigvee_{j_d=1}^{r_d} \mathcal{T} \left(\mathbb{I}_{j_1}^1, \mathbb{I}_{j_2}^2, \dots, \mathbb{I}_{j_d}^d \right)$ is a modal partition of a tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ where $\{\mathbb{I}_1^k, \mathbb{I}_2^k, \dots, \mathbb{I}_{r_k}^k\}$ is a partition of \mathbb{I}^k for $k = 1, 2, \dots, d$, then for any $\mathbf{p} \in \mathbb{P}^d$,*

$$\begin{aligned} \|\mathcal{T}\|_{\mathbf{p}_\sigma} &= \left\| \bigvee_{j_1=1}^{r_1} \bigvee_{j_2=1}^{r_2} \cdots \bigvee_{j_d=1}^{r_d} \mathcal{T} \left(\mathbb{I}_{j_1}^1, \mathbb{I}_{j_2}^2, \dots, \mathbb{I}_{j_d}^d \right) \right\|_{\mathbf{p}_\sigma} \\ &\leq \left\| \bigvee_{j_1=1}^{r_1} \bigvee_{j_2=1}^{r_2} \cdots \bigvee_{j_d=1}^{r_d} \left\| \mathcal{T} \left(\mathbb{I}_{j_1}^1, \mathbb{I}_{j_2}^2, \dots, \mathbb{I}_{j_d}^d \right) \right\|_{\mathbf{p}_\sigma} \right\|_{\mathbf{p}_\sigma}. \end{aligned} \tag{4}$$

Proof Let $\|\mathcal{T}\|_{\mathbf{p}_\sigma} = \langle \mathcal{T}, \mathbf{x}^1 \otimes \mathbf{x}^2 \otimes \cdots \otimes \mathbf{x}^d \rangle$ with $\|\mathbf{x}^k\|_{p_k} = 1$ for $k = 1, 2, \dots, d$ (by compactness, these \mathbf{x}^k 's must exist). Denote

$$\mathcal{T}_{j_1 j_2 \dots j_d} = \mathcal{T} \left(\mathbb{I}_{j_1}^1, \mathbb{I}_{j_2}^2, \dots, \mathbb{I}_{j_d}^d \right) \quad \forall 1 \leq j_k \leq r_k, \quad k = 1, 2, \dots, d,$$

which leads to

$$\mathcal{T} = \prod_{j_1=1}^{r_1} \prod_{j_2=1}^{r_2} \cdots \prod_{j_d=1}^{r_d} \mathcal{T}_{j_1 j_2 \dots j_d}.$$

Further, we denote

$$\mathbf{x}_{j_k}^k = \mathbf{x}^k (\mathbb{1}_{j_k}^k) \quad \forall 1 \leq j_k \leq r_k, \quad k = 1, 2, \dots, d.$$

Obviously we have

$$\mathbf{x}^k = \bigvee_{j_k=1}^{r_k} \mathbf{x}_{j_k}^k \quad \forall k = 1, 2, \dots, d.$$

First, for every (j_1, j_2, \dots, j_d) , if none of the vectors $\mathbf{x}_{j_1}^1, \mathbf{x}_{j_2}^2, \dots, \mathbf{x}_{j_d}^d$ is a zero vector, then

$$\begin{aligned} & \left\langle \mathcal{T}_{j_1 j_2 \dots j_d}, \mathbf{x}_{j_1}^1 \otimes \mathbf{x}_{j_2}^2 \otimes \cdots \otimes \mathbf{x}_{j_d}^d \right\rangle \\ &= \left\langle \mathcal{T}_{j_1 j_2 \dots j_d}, \frac{\mathbf{x}_{j_1}^1}{\|\mathbf{x}_{j_1}^1\|_{p_1}} \otimes \frac{\mathbf{x}_{j_2}^2}{\|\mathbf{x}_{j_2}^2\|_{p_2}} \otimes \cdots \otimes \frac{\mathbf{x}_{j_d}^d}{\|\mathbf{x}_{j_d}^d\|_{p_d}} \right\rangle \prod_{k=1}^d \|\mathbf{x}_{j_k}^k\|_{p_k} \\ &\leq \|\mathcal{T}_{j_1 j_2 \dots j_d}\|_{p_\sigma} \prod_{k=1}^d \|\mathbf{x}_{j_k}^k\|_{p_k}. \end{aligned} \tag{5}$$

The above inequality trivially holds even if some of $\mathbf{x}_{j_1}^1, \mathbf{x}_{j_2}^2, \dots, \mathbf{x}_{j_d}^d$ are zero vectors, and thus it holds in general.

Next, by the norm compression identity for vectors in Lemma 3.2, we have

$$\left\| \bigvee_{j_k=1}^{r_k} \|\mathbf{x}_{j_k}^k\|_{p_k} \right\|_{p_k} = \left\| \bigvee_{j_k=1}^{r_k} \mathbf{x}_{j_k}^k \right\|_{p_k} = \|\mathbf{x}^k\|_{p_k} = 1 \quad \forall k = 1, 2, \dots, d. \tag{6}$$

Therefore, we have

$$\begin{aligned} \|\mathcal{T}\|_{p_\sigma} &= \left\langle \mathcal{T}, \mathbf{x}^1 \otimes \mathbf{x}^2 \otimes \cdots \otimes \mathbf{x}^d \right\rangle \\ &= \left\langle \prod_{j_1=1}^{r_1} \prod_{j_2=1}^{r_2} \cdots \prod_{j_d=1}^{r_d} \mathcal{T}_{j_1 j_2 \dots j_d}, \left(\bigvee_{j_1=1}^{r_1} \mathbf{x}_{j_1}^1 \right) \otimes \left(\bigvee_{j_2=1}^{r_2} \mathbf{x}_{j_2}^2 \right) \right. \\ &\quad \left. \otimes \cdots \otimes \left(\bigvee_{j_d=1}^{r_d} \mathbf{x}_{j_d}^d \right) \right\rangle \\ &= \left\langle \prod_{j_1=1}^{r_1} \prod_{j_2=1}^{r_2} \cdots \prod_{j_d=1}^{r_d} \mathcal{T}_{j_1 j_2 \dots j_d}, \bigvee_{j_1=1}^{r_1} \bigvee_{j_2=1}^{r_2} \cdots \bigvee_{j_d=1}^{r_d} \mathbf{x}_{j_1}^1 \otimes \mathbf{x}_{j_2}^2 \otimes \cdots \otimes \mathbf{x}_{j_d}^d \right\rangle \end{aligned}$$

$$\begin{aligned}
 &= \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} \cdots \sum_{j_d=1}^{r_d} \left\langle \mathcal{T}_{j_1 j_2 \dots j_d}, \mathbf{x}_{j_1}^1 \otimes \mathbf{x}_{j_2}^2 \otimes \cdots \otimes \mathbf{x}_{j_d}^d \right\rangle \\
 &\leq \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} \cdots \sum_{j_d=1}^{r_d} \left(\|\mathcal{T}_{j_1 j_2 \dots j_d}\|_{p_\sigma} \prod_{k=1}^d \|\mathbf{x}_{j_k}^k\|_{p_k} \right) \\
 &= \left\langle \bigvee_{j_1=1}^{r_1} \bigvee_{j_2=1}^{r_2} \cdots \bigvee_{j_d=1}^{r_d} \|\mathcal{T}_{j_1 j_2 \dots j_d}\|_{p_\sigma}, \bigvee_{j_1=1}^{r_1} \bigvee_{j_2=1}^{r_2} \cdots \bigvee_{j_d=1}^{r_d} \left(\prod_{k=1}^d \|\mathbf{x}_{j_k}^k\|_{p_k} \right) \right\rangle \\
 &= \left\langle \bigvee_{j_1=1}^{r_1} \bigvee_{j_2=1}^{r_2} \cdots \bigvee_{j_d=1}^{r_d} \|\mathcal{T}_{j_1 j_2 \dots j_d}\|_{p_\sigma}, \left(\bigvee_{j_1=1}^{r_1} \|\mathbf{x}_{j_1}^1\|_{p_1} \right) \otimes \left(\bigvee_{j_2=1}^{r_2} \|\mathbf{x}_{j_2}^2\|_{p_2} \right) \right. \\
 &\quad \left. \otimes \cdots \otimes \left(\bigvee_{j_d=1}^{r_d} \|\mathbf{x}_{j_d}^d\|_{p_d} \right) \right\rangle \\
 &\leq \left\| \bigvee_{j_1=1}^{r_1} \bigvee_{j_2=1}^{r_2} \cdots \bigvee_{j_d=1}^{r_d} \|\mathcal{T}_{j_1 j_2 \dots j_d}\|_{p_\sigma} \right\|_{p_\sigma},
 \end{aligned}$$

where the first inequality is due to (5), and the last inequality is due to (6) and Definition 2.2. □

We remark that for $d = 2$, the case of matrices, Theorem 3.3 was not studied in the literature, to the best of our knowledge. The tightness of the norm compression inequality in Theorem 3.3 is in general hard to establish. We list some special cases below albeit some of them are trivial:

1. $d = 1$ which corresponds to the case of vectors essentially established in Lemma 3.2;
2. $\mathbf{p} = (1, 1, \dots, 1)$ for which the spectral \mathbf{p} -norm of a tensor is simply the L_∞ -norm of the tensor or the largest absolute-valued entry of the tensor;
3. All but one of the subtensors are zero tensors;
4. The original tensor \mathcal{T} is rank-one.

The last case actually includes the first case as a special one. Its proof can be obtained by using Proposition 2.4 and is left to interested readers.

For the nuclear \mathbf{p} -norm of a tensor, like the nice dual bounds of the nuclear norm shown in [29] and the nuclear p -norm shown in [9], one hopes to establish a dual inequality to (4) as follows:

$$\begin{aligned}
 \|\mathcal{T}\|_{\mathbf{p}_*} &= \left\| \bigvee_{j_1=1}^{r_1} \bigvee_{j_2=1}^{r_2} \cdots \bigvee_{j_d=1}^{r_d} \mathcal{T} \left(\mathbb{I}_{j_1}^1, \mathbb{I}_{j_2}^2, \dots, \mathbb{I}_{j_d}^d \right) \right\|_{\mathbf{p}_*} \\
 &\geq \left\| \bigvee_{j_1=1}^{r_1} \bigvee_{j_2=1}^{r_2} \cdots \bigvee_{j_d=1}^{r_d} \left\| \mathcal{T} \left(\mathbb{I}_{j_1}^1, \mathbb{I}_{j_2}^2, \dots, \mathbb{I}_{j_d}^d \right) \right\|_{\mathbf{p}_*} \right\|_{\mathbf{p}_*}. \tag{7}
 \end{aligned}$$

Unfortunately, this does not hold in general; see the example below. This actually makes the norm compression inequalities for tensors more interesting, and the result in Theorem 3.3 more valuable.

Example 3.4 Let $M = \begin{pmatrix} -1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \in \mathbb{R}^{3 \times 3}$ and a modal partition of M be the 3×3 entry-wise partition, resulting the nuclear norm compressed tensor to be $|M| = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$. It follows that

$$\|M\|_* = 2\sqrt{3} < 4 = \||M|\|_*,$$

disproving (7) when $d = 2$ and $\mathbf{p} = (2, 2)$.

As far as we are aware, the only known nontrivial case for (7) to hold is for 2×2 blocked matrices with $\mathbf{p} = (2, 2)$ due to King [23], i.e., (1) for $p = 1$. There is a general case, though sort of trivial, for (7) to hold as an equality when $\mathbf{p} = (1, 1, \dots, 1)$ for which the nuclear \mathbf{p} -norm of a tensor becomes the L_1 -norm of the tensor.

To conclude this section, we provide some insights on (7) although we are unable to prove any general result. We believe that (7) holds for nonnegative tensors, i.e., tensors having all nonnegative entries. Another interesting question is to find the smallest $\tau > 0$ such that

$$\begin{aligned} \|\mathcal{T}\|_{\mathbf{p}_*} &= \left\| \prod_{j_1=1}^{r_1} \prod_{j_2=1}^{r_2} \cdots \prod_{j_d=1}^{r_d} \mathcal{T} \left(\mathbb{I}_{j_1}^1, \mathbb{I}_{j_2}^2, \dots, \mathbb{I}_{j_d}^d \right) \right\|_{\mathbf{p}_*} \\ &\geq \tau \left\| \prod_{j_1=1}^{r_1} \prod_{j_2=1}^{r_2} \cdots \prod_{j_d=1}^{r_d} \left\| \mathcal{T} \left(\mathbb{I}_{j_1}^1, \mathbb{I}_{j_2}^2, \dots, \mathbb{I}_{j_d}^d \right) \right\|_{\mathbf{p}_*} \right\|_{\mathbf{p}_*} \end{aligned}$$

holds in general. Numerical evidence shows that τ may not be a universal constant, and can depend on dimensions of the tensor space.

4 Improved bounds on tensor and matrix norms

In this section, we discuss how our norm compression inequalities improve other known bounds on tensor and matrix norms in the literature.

4.1 Norm compression inequalities for matrices

As mentioned in Sect. 1, norm compression inequalities for matrices were studied mainly for the Schatten p -norms, which unfortunately do not hold for general $r_1 \times r_2$ blocks [3]. However, there do exist two related results for general $r \times r$ blocks. For the relevancy to our results, we present them using our notations. Let

$T = \bigvee_{i=1}^r \bigvee_{j=1}^r T_{ij}$ be a modal partition of a matrix T . One result is due to Bhatia and Kittaneh [6, Theorem 1]:

$$r^{-2} \sum_{i=1}^r \sum_{j=1}^r \|T_{ij}\|_{\sigma}^2 \leq \|T\|_{\sigma}^2 \leq \sum_{i=1}^r \sum_{j=1}^r \|T_{ij}\|_{\sigma}^2, \tag{8}$$

and the other is due to Bebendorf [4, Lemma 2.14]:

$$\|T\|_{\sigma} \leq \left(\max_{1 \leq i \leq r} \sum_{j=1}^r \|T_{ij}\|_{\sigma} \right)^{\frac{1}{2}} \left(\max_{1 \leq j \leq r} \sum_{i=1}^r \|T_{ij}\|_{\sigma} \right)^{\frac{1}{2}}. \tag{9}$$

A basic inequality between the spectral norm and the Frobenius norms of a matrix (see, e.g., [14]) state that

$$\|T\|_{\sigma} \leq \|T\|_2.$$

Therefore, according to (4) when $d = 2$ and $\mathbf{p} = (2, 2)$, we get

$$\|T\|_{\sigma} \leq \left\| \bigvee_{i=1}^r \bigvee_{j=1}^r \|T_{ij}\|_{\sigma} \right\|_{\sigma} \leq \left\| \bigvee_{i=1}^r \bigvee_{j=1}^r \|T_{ij}\|_{\sigma} \right\|_2 = \left(\sum_{i=1}^r \sum_{j=1}^r \|T_{ij}\|_{\sigma}^2 \right)^{\frac{1}{2}},$$

providing a tighter upper bound of $\|T\|_{\sigma}$ than that in (8).

To see how our inequality (4) improves the upper bound of $\|T\|_{\sigma}$ in (9), we need to use a classical result due to Schur [43], which states that for any matrix $T = (t_{ij}) \in \mathbb{R}^{n_1 \times n_2}$,

$$\|T\|_{\sigma} \leq \left(\max_{1 \leq i \leq n_1} \sum_{j=1}^{n_2} |t_{ij}| \right)^{\frac{1}{2}} \left(\max_{1 \leq j \leq n_2} \sum_{i=1}^{n_1} |t_{ij}| \right)^{\frac{1}{2}}. \tag{10}$$

By (4) when $d = 2$ and $\mathbf{p} = (2, 2)$, we get

$$\|T\|_{\sigma} \leq \left\| \bigvee_{i=1}^r \bigvee_{j=1}^r \|T_{ij}\|_{\sigma} \right\|_{\sigma} \leq \left(\max_{1 \leq i \leq r} \sum_{j=1}^r \|T_{ij}\|_{\sigma} \right)^{\frac{1}{2}} \left(\max_{1 \leq j \leq r} \sum_{i=1}^r \|T_{ij}\|_{\sigma} \right)^{\frac{1}{2}},$$

providing a tighter upper bound of $\|T\|_{\sigma}$ than that in (9).

We remark that the result (10) of Schur [43] has been generalized to tensors by Hardy, Littlewood, and Pólya [15, Theorem 273] and so we can easily apply the norm compression inequality (4) to generalize inequality (9) from matrices to tensors. The detail is left to interested readers.

4.2 Bounds on tensor norms via partitions

Li [29] first proposed bounds on the tensor spectral norm based on tensor partitions. Specifically, if $\{\mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_m\}$ is a regular partition of a tensor \mathcal{T} , then

$$\|(\|\mathcal{T}_1\|_\sigma, \|\mathcal{T}_2\|_\sigma, \dots, \|\mathcal{T}_m\|_\sigma)\|_\infty \leq \|\mathcal{T}\|_\sigma \leq \|(\|\mathcal{T}_1\|_\sigma, \|\mathcal{T}_2\|_\sigma, \dots, \|\mathcal{T}_m\|_\sigma)\|_2. \tag{11}$$

This result was later generalized to the most general class of partitions and to any tensor spectral p -norm by Chen and Li [9], i.e., if $\{\mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_m\}$ is an arbitrary partition of a tensor \mathcal{T} and $p \in \mathbb{P}$ with $\frac{1}{p} + \frac{1}{q} = 1$, then

$$\|(\|\mathcal{T}_1\|_{p_\sigma}, \|\mathcal{T}_2\|_{p_\sigma}, \dots, \|\mathcal{T}_m\|_{p_\sigma})\|_\infty \leq \|\mathcal{T}\|_{p_\sigma} \leq \|(\|\mathcal{T}_1\|_{p_\sigma}, \|\mathcal{T}_2\|_{p_\sigma}, \dots, \|\mathcal{T}_m\|_{p_\sigma})\|_q. \tag{12}$$

Here we do not introduce regular partitions and arbitrary partitions of a tensor, but only mention that they are more general than modal partitions. However, modal partitions are the most commonly seen partitions in practice.

Corollary 4.1 *If $\mathcal{T} = \bigvee_{j_1=1}^{r_1} \bigvee_{j_2=1}^{r_2} \dots \bigvee_{j_d=1}^{r_d} \mathcal{T}_{j_1 j_2 \dots j_d}$ is a modal partition of a tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_d}$ and $p \in \mathbb{P}$ with $\frac{1}{p} + \frac{1}{q} = 1$, then*

$$\|\mathcal{T}\|_{p_\sigma} \leq \left\| \bigvee_{j_1=1}^{r_1} \bigvee_{j_2=1}^{r_2} \dots \bigvee_{j_d=1}^{r_d} \|\mathcal{T}_{j_1 j_2 \dots j_d}\|_{p_\sigma} \right\|_{p_\sigma} \leq \left(\sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} \dots \sum_{j_d=1}^{r_d} \|\mathcal{T}_{j_1 j_2 \dots j_d}\|_{p_\sigma}^q \right)^{\frac{1}{q}}. \tag{13}$$

Proof The first inequality in (13) is exactly (4) if we let $\mathbf{p} = (p, p, \dots, p)$. To see why one upper bound is tighter than the other upper bound in (13), we only need to apply the upper bound of (12) since $\left\{ \|\mathcal{T}_{j_1 j_2 \dots j_d}\|_{p_\sigma} : 1 \leq j_k \leq r_k, r = 1, 2, \dots, d \right\}$ is a modal partition of the tensor $\bigvee_{j_1=1}^{r_1} \bigvee_{j_2=1}^{r_2} \dots \bigvee_{j_d=1}^{r_d} \|\mathcal{T}_{j_1 j_2 \dots j_d}\|_{p_\sigma}$. \square

The above result obviously improves the upper bound in (12) and when $p = 2$ improves the upper bound in (11). These improvements are made by considering positions of $\|\mathcal{T}_{j_1 j_2 \dots j_d}\|_{p_\sigma}$'s in the norm compressed tensor $\bigvee_{j_1=1}^{r_1} \bigvee_{j_2=1}^{r_2} \dots \bigvee_{j_d=1}^{r_d} \|\mathcal{T}_{j_1 j_2 \dots j_d}\|_{p_\sigma}$ rather than treating all of them as entries of a vector.

4.3 Bounds on norms of matrix unfoldings

Matrix unfoldings of a tensor has been one of the main tools in tensor computations, partially because that most tensor problems are NP-hard [19] while corresponding matrix problems are much easier. For instance, computing the spectral norm and the nuclear norm of a tensor is NP-hard [12,17] while that of a matrix can be done in polynomial time. The relation between norms of matrix unfoldings of a tensor and

norms via certain partitions of the tensor has been investigated by Chen and Li [9]. As discussed in Sect. 4.2, the norm compression inequality in Theorem 3.3 improves (12) in [9]. Consequently, bounds of the tensor spectral p -norm can be improved in various ways by applying some specific partitions of the tensor. Here we discuss one particular instance of this kind to appreciate the applicability of our general approach. It could be of special interest to bounding the spectral norm of a large matrix analogous to the discussion in Sect. 6.

Let $\mathcal{T} \in \mathbb{R}^{n \times n \times n \times n}$ be a fourth order tensor. The traditional matrix unfolding of \mathcal{T} unfolds \mathcal{T} to an $n \times n^3$ matrix, and it can be done via four different modes. Square matrix unfoldings, i.e., unfolding \mathcal{T} to an $n^2 \times n^2$ matrix, has been appeared frequently recently, in particular in studying the largest eigenvalue of a fourth order tensor [22,37]. Let $T_{13,24} \in \mathbb{R}^{n^2 \times n^2}$ be the square matrix unfolding of \mathcal{T} by grouping modes 1 and 3 of \mathcal{T} into the row and modes 2 and 4 of \mathcal{T} into the column of $T_{13,24}$. It is well known (see, e.g., Wang et al. [48]) that $\|\mathcal{T}\|_\sigma \leq \|T_{13,24}\|_\sigma$. Let

$$T_{ij} = \mathcal{T}(\{i\}, \{j\}, \{1, 2, \dots, n\}, \{1, 2, \dots, n\}) \in \mathbb{R}^{1 \times 1 \times n \times n} \quad \forall 1 \leq i, j \leq n,$$

and so we obtain a modal partition of \mathcal{T} ,

$$\mathcal{T} = \bigvee_{i=1}^n \bigvee_{j=1}^n T_{ij}.$$

As T_{ij} is essentially a matrix, we use $T_{ij} \in \mathbb{R}^{n \times n}$ to denote it. An important observation is that

$$T_{13,24} = \bigvee_{i=1}^n \bigvee_{j=1}^n T_{ij}.$$

In fact, we also have $T_{14,23} = \bigvee_{i=1}^n \bigvee_{j=1}^n (T_{ij})^T$, and other square matrix unfoldings of \mathcal{T} can also be modal partitioned similarly. The above discussion can be clearly verified by the follow example.

Example 4.2 Let $\mathcal{T} = (t_{ijkl}) \in \mathbb{R}^{2 \times 2 \times 2 \times 2}$, and we have

$$T_{13,24} = \left(\begin{array}{cc|cc} t_{1111} & t_{1112} & t_{1211} & t_{1212} \\ t_{1121} & t_{1122} & t_{1221} & t_{1222} \\ \hline t_{2111} & t_{2112} & t_{2211} & t_{2212} \\ t_{2121} & t_{2122} & t_{2221} & t_{2222} \end{array} \right) = \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix} \in \mathbb{R}^{4 \times 4},$$

$$T_{14,23} = \left(\begin{array}{cc|cc} t_{1111} & t_{1121} & t_{1211} & t_{1221} \\ t_{1112} & t_{1122} & t_{1212} & t_{1222} \\ \hline t_{2111} & t_{2121} & t_{2211} & t_{2221} \\ t_{2112} & t_{2122} & t_{2212} & t_{2222} \end{array} \right) = \begin{pmatrix} (T_{11})^T & (T_{12})^T \\ (T_{21})^T & (T_{22})^T \end{pmatrix} \in \mathbb{R}^{4 \times 4}.$$

Let us now apply Theorem 3.3, and we obtain

$$\|\mathcal{T}\|_\sigma \leq \|T_{13,24}\|_\sigma \leq \left\| \bigvee_{i=1}^n \bigvee_{j=1}^n \|T_{ij}\|_\sigma \right\|_\sigma \leq \left(\sum_{i=1}^n \sum_{j=1}^n \|T_{ij}\|_\sigma^2 \right)^{\frac{1}{2}}. \tag{14}$$

The bound for both $\|\mathcal{T}\|_\sigma$ and $\|T_{13,24}\|_\sigma$ by the norm compression inequality, $\left\| \bigvee_{i=1}^n \bigvee_{j=1}^n \|T_{ij}\|_\sigma \right\|_\sigma$, improves $\left(\sum_{i=1}^n \sum_{j=1}^n \|T_{ij}\|_\sigma^2 \right)^{\frac{1}{2}}$, which is an instance of [9, Theorem 4.7]. In fact, the bound $\left\| \bigvee_{i=1}^n \bigvee_{j=1}^n \|T_{ij}\|_\sigma \right\|_\sigma$ can be computed in polynomial time. In practice, if a given matrix $T_{13,24}$ is very large, (14) can be used to lower and upper bound its spectral norm by the spectral norm of \mathcal{T} and the norm compressed matrix $\bigvee_{i=1}^n \bigvee_{j=1}^n \|T_{ij}\|_\sigma$. We will discuss this further in Sect. 6.

To conclude this section, we remark that the variety of modal partitions of a tensor provides various specific needs, such as norms of tensor unfoldings, i.e., unfolding a given tensor to a tensor of a lower order [48]. The usefulness of norm compression inequalities for tensors (Theorem 3.3) is based on the following ground fact: For any tensor unfolding (including matrix unfolding) of a given tensor, there exists a modal partition of the tensor, which is also a modal partition of the tensor unfolding.

5 Extremal ratio between the spectral and Frobenius norms

Given a tensor space $\mathbb{R}^{n_1 \times n_2 \times \dots \times n_d}$, the extremal ratio between the spectral norm and the Frobenius norm is defined as

$$\tau(\mathbb{R}^{n_1 \times n_2 \times \dots \times n_d}) := \min \left\{ \frac{\|\mathcal{T}\|_\sigma}{\|\mathcal{T}\|_2} : \mathcal{T} \neq 0, \mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_d} \right\}. \tag{15}$$

This natural question is easy when $d = 2$ (matrices) but becomes very difficult when $d \geq 3$. The concept was proposed by Qi [40], known as the best rank-one approximation ratio of a tensor space, although Kühn and Peetre [28] studied this ratio much earlier for some small n_i 's when $d = 3$. For a nonzero tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_d}$, the following projection problem

$$\max \{ \langle \mathcal{T}, \mathcal{X} \rangle : \|\mathcal{X}\|_2 = 1, \mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_d} \} = \|\mathcal{T}\|_2, \tag{16}$$

and obviously attains optimum when $\mathcal{X} = \frac{\mathcal{T}}{\|\mathcal{T}\|_2}$ by the Cauchy–Schwartz inequality. However, if we consider the projection on rank-one tensors, then

$$\max \{ \langle \mathcal{T}, \mathcal{X} \rangle : \|\mathcal{X}\|_2 = 1, \text{rank}(\mathcal{X}) = 1, \mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_d} \} = \|\mathcal{T}\|_\sigma. \tag{17}$$

Therefore, as an optimization problem (16), (17) becomes its convex relaxation. One is often interested in this relaxation gap, which is exactly the extremal ratio

$\tau(\mathbb{R}^{n_1 \times n_2 \times \dots \times n_d})$. In the case of matrices, similar problems are known to be equivalence constants for matrix norms; see, e.g., Tonge [45]. In this sense, the problem is to determine the largest $\tau > 0$ such that $\tau \|\mathcal{T}\|_2 \leq \|\mathcal{T}\|_\sigma$ for all $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_d}$. On the other hand, the closest gap between the two norms is one since $\|\mathcal{T}\|_\sigma \leq \|\mathcal{T}\|_2$ and the equality holds if and only if \mathcal{T} is rank-one.

As an application of this extremal ratio, one obtains an interpretation as a perturbed steepest descent method and can deduce a rate of convergence using bounds of the extremal ratio (see [46, Theorem 2] for details). Since the time it was posted as a conjecture [40, Sect. 7], deciding this extremal ratio for a general tensor space has been a challenge task. Without loss of generality, we assume that $2 \leq n_1 \leq n_2 \leq \dots \leq n_d$ holds in this section. Some known values of $\tau(\mathbb{R}^{n_1 \times n_2 \times \dots \times n_d})$ are: $\tau(\mathbb{R}^{n_1 \times n_2}) = \frac{1}{\sqrt{n_1}}$, $\tau(\mathbb{R}^{2 \times n_2 \times n_3}) = \frac{1}{\sqrt{2n_2}}$ for even n_2 [26], and $\tau(\mathbb{R}^{n_1 \times n_2 \times \dots \times n_d}) = \frac{1}{\sqrt{n_1 n_2 \dots n_{d-1}}}$ for $n_d = 2, 4, 8$ [31]. Note that

$$\tau(\mathbb{R}^{n_1 \times n_2 \times \dots \times n_d}) \geq \frac{1}{\sqrt{n_1 n_2 \dots n_{d-1}}} \tag{18}$$

is a naive lower bound, but can be obtained in various cases; see [31] for recent development on the topic. In the space of symmetric tensors, i.e.,

$$\tau(\mathbb{R}_{\text{sym}}^{n^d}) := \min \left\{ \frac{\|\mathcal{T}\|_\sigma}{\|\mathcal{T}\|_2} : \mathcal{T} \neq 0, \mathcal{T} \in \mathbb{R}_{\text{sym}}^{n^d} \right\},$$

where $\mathbb{R}_{\text{sym}}^{n^d}$ denotes the set of d -th order symmetric real tensors, this extremal ratio has also been studied in [1,26,40,49]. In particular, it was recently shown in [1] that the naive bound (18) is tight for symmetric tensors only if $n_d = 2$. One may also consider the extremal ratio for nonnegative tensors, i.e.,

$$\tau(\mathbb{R}_+^{n_1 \times n_2 \times \dots \times n_d}) := \min \left\{ \frac{\|\mathcal{T}\|_\sigma}{\|\mathcal{T}\|_2} : \mathcal{T} \neq 0, \mathcal{T} \in \mathbb{R}_+^{n_1 \times n_2 \times \dots \times n_d} \right\},$$

where $\mathbb{R}_+^{n_1 \times n_2 \times \dots \times n_d}$ denotes the set of d -th order nonnegative tensors.

In this section, we provide a general tool investigating upper bounds of this extremal ratio by the norm compression inequality in Theorem 3.3. According to (15), the value of $\frac{\|\mathcal{T}\|_\sigma}{\|\mathcal{T}\|_2}$ for any nonzero \mathcal{T} provides an upper bound of the extremal ratio. By recursively constructing modal partitions, we obtain the main result in this section.

Theorem 5.1 *If $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_d}$ is a nonnegative tensor with $\frac{\|\mathcal{T}\|_\sigma}{\|\mathcal{T}\|_2} = \tau$, then there exists a nonnegative tensor $\mathcal{T}_m \in \mathbb{R}^{n_1^m \times n_2^m \times \dots \times n_d^m}$ satisfying $\frac{\|\mathcal{T}_m\|_\sigma}{\|\mathcal{T}_m\|_2} = \tau^m$ for any $m \in \mathbb{N}$. If \mathcal{T} is further symmetric, then \mathcal{T}_m is also symmetric.*

Proof Let $\mathcal{T} = (t_{i_1 i_2 \dots i_d})$ where the entry $t_{i_1 i_2 \dots i_d} \geq 0$ for all $1 \leq i_k \leq n_k, k = 1, 2, \dots, d$, in other words, $\mathcal{T} = \bigvee_{i_1=1}^{n_1} \bigvee_{i_2=1}^{n_2} \dots \bigvee_{i_d=1}^{n_d} t_{i_1 i_2 \dots i_d}$. \mathcal{T}_m is defined recursively as follows,

$$\mathcal{T}_1 := \mathcal{T}, \mathcal{T}_{m+1} := \bigvee_{i_1=1}^{n_1} \bigvee_{i_2=1}^{n_2} \cdots \bigvee_{i_d=1}^{n_d} (t_{i_1 i_2 \dots i_d} \mathcal{T}_m) \quad m \geq 1.$$

It is easy to see that $\mathcal{T}_m \in \mathbb{R}^{n_1^m \times n_2^m \times \dots \times n_d^m}$ is nonnegative, and is symmetric if \mathcal{T} is symmetric. The Frobenius norm of \mathcal{T}_m satisfies

$$\begin{aligned} \|\mathcal{T}_{m+1}\|_2^2 &= \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \cdots \sum_{i_d=1}^{n_d} \|t_{i_1 i_2 \dots i_d} \mathcal{T}_m\|_2^2 \\ &= \|\mathcal{T}_m\|_2^2 \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \cdots \sum_{i_d=1}^{n_d} t_{i_1 i_2 \dots i_d}^2 \\ &= \|\mathcal{T}_m\|_2^2 \|\mathcal{T}\|_2^2, \end{aligned}$$

and so we get

$$\|\mathcal{T}_m\|_2 = \|\mathcal{T}\|_2^m.$$

For the spectral norm of \mathcal{T}_m , by noticing $t_{i_1 i_2 \dots i_d} \geq 0$ and applying (4) in Theorem 3.3,

$$\begin{aligned} \|\mathcal{T}_{m+1}\|_\sigma &\leq \left\| \bigvee_{i_1=1}^{n_1} \bigvee_{i_2=1}^{n_2} \cdots \bigvee_{i_d=1}^{n_d} \|t_{i_1 i_2 \dots i_d} \mathcal{T}_m\|_\sigma \right\|_\sigma \\ &= \|\mathcal{T}_m\|_\sigma \left\| \bigvee_{i_1=1}^{n_1} \bigvee_{i_2=1}^{n_2} \cdots \bigvee_{i_d=1}^{n_d} t_{i_1 i_2 \dots i_d} \right\|_\sigma \\ &= \|\mathcal{T}_m\|_\sigma \|\mathcal{T}\|_\sigma, \end{aligned}$$

and we obtain

$$\|\mathcal{T}_m\|_\sigma \leq \|\mathcal{T}\|_\sigma^m. \tag{19}$$

On the other hand, let $\mathcal{X} = \mathbf{x}^1 \otimes \mathbf{x}^2 \otimes \dots \otimes \mathbf{x}^d$ with $\mathbf{x}^k \in \mathbb{R}^{n_k}$ and $\|\mathbf{x}^k\|_2 = 1$ for $k = 1, 2, \dots, d$, such that $\langle \mathcal{T}, \mathcal{X} \rangle = \|\mathcal{T}\|_\sigma$, i.e., \mathcal{X} is a best rank-one approximation of \mathcal{T} . It is easy to see that \mathcal{X} must be nonnegative as \mathcal{T} is nonnegative. Recursively construct \mathcal{X}_m as that for \mathcal{T}_m as follows,

$$\mathcal{X}_1 := \mathcal{X}, \mathcal{X}_{m+1} := \bigvee_{i_1=1}^{n_1} \bigvee_{i_2=1}^{n_2} \cdots \bigvee_{i_d=1}^{n_d} (x_{i_1 i_2 \dots i_d} \mathcal{X}_m) \quad m \geq 1.$$

Similar to \mathcal{T}_m , we have $\|\mathcal{X}_m\|_2 = \|\mathcal{X}\|_2^m = 1$. Moreover, \mathcal{X}_m is actually rank-one. To see why, we notice that \mathcal{X}_1 is rank-one and so $x_{i_1 i_2 \dots i_d} = \prod_{k=1}^d x_{i_k}^k$. If \mathcal{X}_m is rank-one, say $\mathcal{X}_m = \mathbf{y}^1 \otimes \mathbf{y}^2 \otimes \dots \otimes \mathbf{y}^d$, then

$$\begin{aligned} \mathcal{X}_{m+1} &= \bigvee_{i_1=1}^{n_1} \bigvee_{i_2=1}^{n_2} \cdots \bigvee_{i_d=1}^{n_d} \left(\left(\prod_{k=1}^d x_{i_k}^k \right) \mathbf{y}^1 \otimes \mathbf{y}^2 \otimes \cdots \otimes \mathbf{y}^d \right) \\ &= \left(\bigvee_{i_1=1}^{n_1} (x_{i_1}^1 \mathbf{y}^1) \right) \otimes \left(\bigvee_{i_2=1}^{n_2} (x_{i_2}^2 \mathbf{y}^2) \right) \otimes \cdots \otimes \left(\bigvee_{i_d=1}^{n_d} (x_{i_d}^d \mathbf{y}^d) \right) \end{aligned}$$

is rank-one. By the constructions of \mathcal{T}_m and \mathcal{X}_m , we have

$$\begin{aligned} \langle \mathcal{T}_{m+1}, \mathcal{X}_{m+1} \rangle &= \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \cdots \sum_{i_d=1}^{n_d} \langle t_{i_1 i_2 \dots i_d} \mathcal{T}_m, x_{i_1 i_2 \dots i_d} \mathcal{X}_m \rangle \\ &= \langle \mathcal{T}_m, \mathcal{X}_m \rangle \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \cdots \sum_{i_d=1}^{n_d} (t_{i_1 i_2 \dots i_d} \cdot x_{i_1 i_2 \dots i_d}) \\ &= \langle \mathcal{T}_m, \mathcal{X}_m \rangle \langle \mathcal{T}, \mathcal{X} \rangle, \end{aligned}$$

implying that $\langle \mathcal{T}_m, \mathcal{X}_m \rangle = \langle \mathcal{T}, \mathcal{X} \rangle^m$. Therefore, by that \mathcal{X}_m is rank-one and $\|\mathcal{X}_m\|_2 = 1$,

$$\|\mathcal{T}_m\|_\sigma \geq \langle \mathcal{T}_m, \mathcal{X}_m \rangle = \langle \mathcal{T}, \mathcal{X} \rangle^m = \|\mathcal{T}\|_\sigma^m,$$

which combined with (19) leads to

$$\|\mathcal{T}_m\|_\sigma = \|\mathcal{T}\|_\sigma^m.$$

Finally, the constructed \mathcal{T}_m satisfies

$$\frac{\|\mathcal{T}_m\|_\sigma}{\|\mathcal{T}_m\|_2} = \frac{\|\mathcal{T}\|_\sigma^m}{\|\mathcal{T}\|_2^m} = \tau^m$$

for any $m \in \mathbb{N}$. □

We remark that the construction of \mathcal{T}_m in the proof of Theorem 5.1 is essentially from the Kronecker products of m copies of \mathcal{T} . For the sake of simplicity, we do not introduce more notations at this point.

Let us now apply Theorem 5.1 to get an improved upper bound for both $\tau(\mathbb{R}_+^{n \times n \times n})$ and $\tau(\mathbb{R}_{\text{sym}}^{n \times n \times n})$. First we introduce the following example.

Example 5.2 Let $\mathcal{U} \in \mathbb{R}^{2 \times 2 \times 2}$ be nonnegative and symmetric where $u_{112} = u_{121} = u_{211} = 1$ and other entries are zeros. We have $\|\mathcal{U}\|_\sigma = \frac{2}{\sqrt{3}}$, $\|\mathcal{U}\|_2 = \sqrt{3}$, and $\frac{\|\mathcal{U}\|_\sigma}{\|\mathcal{U}\|_2} = \frac{2}{3}$.

The calculation of $\|\mathcal{U}\|_\sigma$ can be easily obtained using the fact that the best rank-one approximation of a symmetric tensor can be obtained at a symmetric rank-one tensor; see e.g., [8,49]. Since \mathcal{U} is symmetric, we have

$$\begin{aligned} \|\mathcal{U}\|_\sigma &= \max_{\|x^k\|_2=1} \langle \mathcal{U}, x^1 \otimes x^2 \otimes x^3 \rangle = \max_{\|y\|_2=1} \langle \mathcal{U}, y \otimes y \otimes y \rangle = \max_{\|y\|_2=1} 3y_1^2 y_2 \\ &= \max_{-1 \leq y_2 \leq 1} 3(1 - y_2^2) y_2 = \frac{2}{\sqrt{3}}. \end{aligned}$$

We believe that $\frac{\|\mathcal{T}\|_\sigma}{\|\mathcal{T}\|_2} \geq \frac{2}{3}$ for any nonnegative tensor $\mathcal{T} \in \mathbb{R}_+^{2 \times 2 \times 2}$, i.e. $\tau(\mathbb{R}_+^{2 \times 2 \times 2}) = \frac{2}{3}$ from some numerical evidence although we are unable to verify it theoretically.

Theorem 5.3 *It holds that*

$$\frac{1}{n} \leq \tau(\mathbb{R}_+^{n \times n \times n}), \tau(\mathbb{R}_{\text{sym}}^{n \times n \times n}) \leq \frac{1.5}{n^{\frac{\ln 1.5}{\ln 2}}} \leq O\left(\frac{1}{n^{0.584}}\right).$$

Proof The lower bound is listed for reference only, which is the naive one in (18) but currently the best known one. For the upper bound, let $\mathcal{U} \in \mathbb{R}^{2 \times 2 \times 2}$ be the tensor in Example 5.2 and let $m \in \mathbb{N}$ such that $2^m \leq n < 2^{m+1}$. By Theorem 5.1, there exists $\mathcal{U}_m \in \mathbb{R}^{2^m \times 2^m \times 2^m}$, both nonnegative and symmetric, such that

$$\begin{aligned} \tau(\mathbb{R}_{\text{sym}}^{2^m \times 2^m \times 2^m}) &\leq \frac{\|\mathcal{U}_m\|_\sigma}{\|\mathcal{U}_m\|_2} = \left(\frac{\|\mathcal{U}\|_\sigma}{\|\mathcal{U}\|_2}\right)^m = \left(\frac{2}{3}\right)^m = \frac{3}{2} \cdot 2^{(m+1) \log_2 \frac{2}{3}} \\ &< \frac{3}{2} \cdot n^{\log_2 \frac{2}{3}} = \frac{1.5}{n^{\frac{\ln 1.5}{\ln 2}}}. \end{aligned}$$

Finally, by the obvious fact that $\tau(\mathbb{R}_{\text{sym}}^{n \times n \times n})$ is nonincreasing as n increases, we get

$$\tau(\mathbb{R}_{\text{sym}}^{n \times n \times n}) \leq \tau(\mathbb{R}_{\text{sym}}^{2^m \times 2^m \times 2^m}) \leq \frac{1.5}{n^{\frac{\ln 1.5}{\ln 2}}}.$$

The above argument obviously showed the upper bound of $\tau(\mathbb{R}_+^{n \times n \times n})$ as well. □

The upper bound in Theorem 5.3 improves the existing one $O\left(\frac{1}{n^{0.5}}\right)$ and remains the best for general $\mathbb{R}_+^{n \times n \times n}$ and $\mathbb{R}_{\text{sym}}^{n \times n \times n}$, to the best of our knowledge. It may be possible to investigate other small-size nonnegative tensors, say $\mathcal{T} \in \mathbb{R}_+^{3 \times 3 \times 3}$, to obtain new upper bounds of $\tau(\mathbb{R}_+^{n \times n \times n})$ using Theorem 5.1. We are not sure if this can beat the one in Theorem 5.3 even though finding $\tau(\mathbb{R}_+^{3 \times 3 \times 3})$ is already hard. In general, this tool can be certainly used to find better upper bounds of the extremal ratio for fourth or higher order tensors.

6 Estimating the spectral norm

A straightforward application of norm compression inequalities is to estimate a norm of a large tensor via norms of small subtensors in a modal partition. This is because computing the spectral norm of a tensor is NP-hard in general while computing that of small tensors can be done quite efficiently and accurately. Even for a matrix, computing

the spectral norm could be costly when its size gets very large. Estimating matrix norms is an important topic in matrix computations. Most methods in the literature are based on random sampling [35] and power method [27]. In this section, we conduct a preliminary study on norm compression approach using the matrix spectral norm as an example, to provide a picture on how *fast* the method runs theoretically and how *good* the approximation is numerically.

As we know, computing the spectral norm of an $n \times n$ matrix requires $O(n^3)$ operations, which is essentially the complexity of singular value decompositions. For simplicity, we do not consider numerical errors that could bring some sort of $\log \frac{1}{\epsilon}$. Suppose that we have a (large) matrix $T \in \mathbb{R}^{n \times n}$ and need to compute some norm $\|T\|$. The exact computation requires αn^s operations where $\alpha > 0$ and $s > 2$ are two universal constants. The following algorithm provides an estimation of $\|T\|$ based on a norm compression hierarchy.

Algorithm 6.1 *Approximating a matrix norm via a norm compression hierarchy.*

- *Input:* A matrix $T \in \mathbb{R}^{n \times n}$, a level of hierarchy $m \geq 1$, and a factorization of $n = \prod_{k=1}^m n_k$ with $n_k \geq 2$ for $k = 1, 2, \dots, m$.
 1. Set $T^0 = T$ and $\ell = 1$.
 2. Denote

$$\mathbb{I}_j^\ell = \{(j-1)n_{m-\ell+1} + 1, (j-1)n_{m-\ell+1} + 2, \dots, jn_{m-\ell+1}\} \quad \forall 1 \leq j \leq \prod_{k=1}^{m-\ell} n_k,$$

and compute the level- ℓ norm compression matrix $T^\ell \in \mathbb{R}^{\prod_{k=1}^{m-\ell} n_k \times \prod_{k=1}^{m-\ell} n_k}$ whose (i, j) -th entry

$$t_{ij}^\ell = \left\| T^{\ell-1} \left(\mathbb{I}_i^\ell, \mathbb{I}_j^\ell \right) \right\| \quad \forall 1 \leq i, j \leq \prod_{k=1}^{m-\ell} n_k.$$

3. If $\ell = m$, stop; otherwise increase ℓ by one and go to the previous step.
- *Output:* An approximation $T^m \in \mathbb{R}$.

Remark that the level- m norm compression matrix T^m is actually one-by-one, i.e., a scalar. It is not difficult to see that T^ℓ is a norm compression matrix of $T^{\ell-1}$ for $\ell = 1, 2, \dots, m$. Therefore, according to Theorem 3.3, we have the following property for Algorithm 6.1.

Proposition 6.2 *For the matrix spectral p -norm where $p \in \mathbb{P}^2$, the matrices computed in Algorithm 6.1 satisfy*

$$\|T\|_{p_\sigma} \leq \|T^1\|_{p_\sigma} \leq \dots \leq \|T^{m-1}\|_{p_\sigma} = T^m.$$

Obviously, the higher the hierarchy, the less accurate the approximation is. However, the approximation is closely related to the way to factorize the dimension n . Let us first

study how to choose n_k 's properly in order to *optimize* the computational complexity of Algorithm 6.1. The exact computation requires αn^s operations for an $n \times n$ matrix. The dimension of T^ℓ is $\prod_{k=1}^{m-\ell} n_k \times \prod_{k=1}^{m-\ell} n_k$, and obtaining every entry of T^ℓ requires to compute the norm of an $n_{m-\ell+1} \times n_{m-\ell+1}$ matrix, which can be done in $\alpha n_{m-\ell+1}^s$ operations. Therefore, the complexity of Algorithm 6.1 is

$$\begin{aligned} \sum_{\ell=1}^m \left(\alpha n_{m-\ell+1}^s \left(\prod_{k=1}^{m-\ell} n_k \right)^2 \right) &= \alpha \sum_{\ell=1}^m \left(n_{m-\ell+1}^{s-2} \prod_{k=1}^{m-\ell+1} n_k^2 \right) \\ &= \alpha \sum_{\ell=1}^m \left(n_\ell^{s-2} \prod_{k=1}^{\ell} n_k^2 \right). \end{aligned} \tag{20}$$

In particular, if all the n_k 's are the same, we have $n = n_1^m$ and the complexity in (20) is bounded by the highest term in the summation, i.e., $O(n_1^{s-2+2m}) = O\left(n^{\frac{2m+s-2}{m}}\right)$. This complexity tends to $n^{2+\epsilon}$ for any $s > 2$.

Proposition 6.3 *If the complexity to compute a norm of a general $n \times n$ matrix is $O(n^s)$ for some $s > 2$, then the complexity of Algorithm 6.1 is $O\left(n^{\frac{2m+s-2}{m}}\right)$ if $n_1 = n_2 = \dots = n_m$.*

For a fixed m , the level of hierarchy, in order to minimize the complexity of Algorithm 6.1, one should let every term in the summation in (20) be in the same order of magnitude. To this end, we have for any $1 \leq \ell \leq m - 1$,

$$n_\ell^{s-2} \prod_{k=1}^{\ell} n_k^2 = n_{\ell+1}^{s-2} \prod_{k=1}^{\ell+1} n_k^2 \implies n_{\ell+1} = n_\ell \frac{s-2}{s} \implies n_\ell = n_1 \left(\frac{s-2}{s}\right)^{\ell-1}.$$

This leads to

$$n = \prod_{k=1}^m n_k = \prod_{k=1}^m n_1 \left(\frac{s-2}{s}\right)^{k-1} = n_1^{\frac{s}{2}} \left(1 - \left(\frac{s-2}{s}\right)^m\right).$$

Plugging the expression of n and n_k in terms of n_1 into (20), we obtain an optimal complexity

$$\alpha \sum_{\ell=1}^m \left(n_1^{(s-2)\left(\frac{s-2}{s}\right)^{\ell-1}} \prod_{k=1}^{\ell} n_1 2^{\left(\frac{s-2}{s}\right)^{k-1}} \right) = \alpha m n_1^s = \alpha m n^{\frac{2}{1 - \left(\frac{s-2}{s}\right)^m}}.$$

Summarizing the above discussion, we have the following result.

Theorem 6.4 *If the complexity to compute a norm of a general $n \times n$ matrix is $O(n^s)$ for some $s > 2$, then the complexity of Algorithm 6.1 is $f_m(n) := O\left(mn^{\frac{2}{1 - \left(\frac{s-2}{s}\right)^m}}\right)$*

by choosing $n_k = O\left(n^{\frac{2\left(\frac{s-2}{s}\right)^{k-1}}{s\left(1-\left(\frac{s-2}{s}\right)^m\right)}}\right)$ for $k = 1, 2, \dots, m$, i.e., $n = \prod_{k=1}^m n_k$ with $n_{k+1} = O\left(n_k^{\frac{s-2}{s}}\right)$.

For whatever $s > 2$, this complexity can be $O(n^{2+\epsilon})$ even for small m . For instance, for the matrix spectral norm where $s = 3$, the complexity $f_m(n)$ and the way to choose a best factorization of n for some small m 's are listed as follows:

$$\begin{aligned} f_1(n) &= O(n^3) && \text{when } n = n^1, \\ f_2(n) &= O(n^{2.25}) && \text{when } n = n^{0.75} \cdot n^{0.25}, \\ f_3(n) &= O(n^{2.077}) && \text{when } n = n^{0.692} \cdot n^{0.231} \cdot n^{0.077}, \\ f_4(n) &= O(n^{2.025}) && \text{when } n = n^{0.675} \cdot n^{0.225} \cdot n^{0.075} \cdot n^{0.025}. \end{aligned}$$

We emphasize that the main purpose of Algorithm 6.1 is not to replace or compare with existing methods of matrix norm computations. It even cannot work without these methods as it needs them to compute norms of submatrices. The main job of Algorithm 6.1 is to speed up these methods, as illustrated in the above derivation of complexities, and in the same time to maintain good approximability, which is shown below numerically.

We now conduct some preliminary numerical tests to see the performance of Algorithm 6.1. In the first set of tests, we choose some $n \times n$ matrices for $n = 10^4$, whose dimension is reasonably large yet computationally tractable by MATLAB in a personal computer. According to Algorithm 6.1, if the level m is set to one, it is simply the spectral norm of the original matrix. Using the optimal complexity setting in Theorem 6.4, for $m = 2$, we should set $(n_1, n_2) = (1000, 10)$, and for $m = 3$, we should set $(n_1, n_2, n_3) = (625, 8, 2) \approx (586, 8.39, 2.03)$ in order to make $n_1 n_2 n_3 = n$. As mentioned in the above, the complexity for 2 levels is $O(n^{2.25})$ and that for 3 levels is $O(n^{2.077})$. For comparison, some classical upper bounds of the matrix spectral norm are used [14, Sect. 2.3]:

$$\|T\|_\sigma \leq \|T\|_2, n\|T\|_\infty, \sqrt{t_1 t_\infty}, \sqrt{n} t_1, \sqrt{n} t_\infty,$$

where

$$t_1 := \max_{1 \leq j \leq n} \sum_{i=1}^n |t_{ij}| \quad \text{and} \quad t_\infty := \max_{1 \leq i \leq n} \sum_{j=1}^n |t_{ij}|.$$

Two types of data matrices are tested. The first type is randomly generated matrices, which include i.i.d. uniform distributions on $[0, 1]$, i.i.d. standard normal distributions, the absolute values of i.i.d. standard normal distributions, i.i.d. Bernoulli distributions, rank- r matrices obtained by sum of i.i.d. rank-one matrices for $r = 1, 10, 100$, as well as low-rank matrices plus some noises. The second type is covariance matrices defined by exponential covariance function $\exp\left(\frac{-|i-j|}{s}\right)$ and by squared exponential

Table 1 Spectral norm and its upper bounds for $n \times n$ matrices when $n = 10^4$

	$\ T\ _\sigma$ (1 level)	2 levels	3 levels	$\ T\ _2$	$\sqrt{t_1 t_\infty}$	$n\ T\ _\infty$	$\sqrt{nt_1}$	$\sqrt{nt_\infty}$
Uniform	1.000	1.030	1.095	1.155	1.022	2.000	102.1	102.2
Normal	1.000	28.20	45.00	50.00	41.10	295.5	4100	4114
Absolute normal	1.000	1.055	1.159	1.253	1.028	7.197	102.8	102.8
Bernoulli	1.000	1.089	1.275	1.414	1.038	2.000	103.8	103.9
Rank-1	1.000	1.000	1.000	1.000	3.179	15.87	309.4	326.9
Rank-10	1.000	1.259	1.538	1.661	3.453	17.68	346.9	344.5
Rank-100	1.000	2.267	3.396	3.761	4.625	25.69	467.1	458.1
Rank-1 + noise	1.000	1.006	1.021	1.034	3.260	16.51	325.2	327.3
Rank-10 + noise	1.000	1.259	1.540	1.663	3.649	19.83	351.8	379.1
$\exp\left(\frac{- i-j }{1}\right)$	1.000	1.335	1.434	52.95	1.000	4621	100.0	100.0
$\exp\left(\frac{- i-j }{2}\right)$	1.000	1.306	1.409	36.03	1.000	2449	100.0	100.0
$\exp\left(\frac{- i-j }{10}\right)$	1.000	1.053	1.107	15.82	1.000	499.6	100.0	100.0
$\exp\left(\frac{- i-j }{100}\right)$	1.000	1.001	1.002	4.992	1.001	50.05	100.1	100.1
$\exp\left(\frac{- i-j }{1000}\right)$	1.000	1.000	1.000	1.648	1.062	5.345	106.2	106.2
$\exp\left(\frac{-(i-j)^2}{2 \cdot 1^2}\right)$	1.000	1.470	1.632	53.11	1.000	3989	100.0	100.0
$\exp\left(\frac{-(i-j)^2}{2 \cdot 2^2}\right)$	1.000	1.429	1.535	37.55	1.000	1995	100.0	100.0
$\exp\left(\frac{-(i-j)^2}{2 \cdot 10^2}\right)$	1.000	1.069	1.146	16.79	1.000	398.9	100.0	100.0
$\exp\left(\frac{-(i-j)^2}{2 \cdot 100^2}\right)$	1.000	1.001	1.002	5.299	1.001	39.91	100.0	100.0
$\exp\left(\frac{-(i-j)^2}{2 \cdot 1000^2}\right)$	1.000	1.000	1.000	1.697	1.040	4.150	104.0	104.0

covariance function $\exp\left(\frac{-(i-j)^2}{2s^2}\right)$ for several s . The spectral norm of the original matrix is computed and scaled to one for easy reference. In the results shown in Table 1, each entry for random matrices is the average of 10 randomly generated instances. It clearly indicates that the bounds by norm compression hierarchies, no matter 2 levels or 3 levels, are good estimations of the true spectral norm, and are in general better than other classical upper bounds.

In the second set of tests, we enlarge the number of levels in Algorithm 6.1 in order to investigate the effect of level increasing to the change of spectral norms. According to Proposition 6.2, the spectral norms of the norm compression hierarchy increase as levels increase in Algorithm 6.1. For this purpose, we test $n \times n$ matrices with $n = 3^8 = 6561$, i.e., $n_1 = n_2 = \dots = n_8 = 3$. The same two types of data matrices are tested, and their spectral norms are scaled to one. The results of the second set are shown in Table 2, where each entry for random matrices is the average of 10 randomly generated instances. We find that increment of compression levels has few effect to

Table 2 Spectral norms of the norm compression hierarchy for $n \times n$ matrices when $n = 3^8$

	$\ T\ _\sigma$	$\ T^1\ _\sigma$	$\ T^2\ _\sigma$	$\ T^3\ _\sigma$	$\ T^4\ _\sigma$	$\ T^5\ _\sigma$	$\ T^6\ _\sigma$	$\ T^7\ _\sigma = T^8$
Uniform	1.000	1.076	1.082	1.083	1.083	1.083	1.083	1.083
Normal	1.000	34.01	34.50	34.55	34.56	34.56	34.56	34.56
Absolute normal	1.000	1.136	1.151	1.153	1.153	1.153	1.153	1.153
Bernoulli	1.000	1.231	1.244	1.245	1.245	1.245	1.245	1.245
Rank-1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Rank-9	1.000	1.456	1.477	1.479	1.480	1.480	1.480	1.480
Rank-81	1.000	2.937	2.981	2.986	2.987	2.987	2.987	2.987
Rank-1 + noise	1.000	1.067	1.070	1.071	1.071	1.071	1.071	1.071
Rank-3 + noise	1.000	1.169	1.179	1.180	1.181	1.181	1.181	1.181
Rank-9 + noise	1.000	1.397	1.417	1.419	1.420	1.420	1.420	1.420
$\exp\left(\frac{- i-j }{1}\right)$	1.000	1.277	1.636	1.994	2.353	2.709	3.046	3.281
$\exp\left(\frac{- i-j }{2}\right)$	1.000	1.120	1.466	1.855	2.243	2.629	2.995	3.249
$\exp\left(\frac{- i-j }{10}\right)$	1.000	1.007	1.058	1.301	1.695	2.090	2.463	2.723
$\exp\left(\frac{- i-j }{100}\right)$	1.000	1.000	1.001	1.006	1.048	1.263	1.632	1.893
$\exp\left(\frac{- i-j }{1000}\right)$	1.000	1.000	1.000	1.000	1.000	1.004	1.030	1.147
$\exp\left(\frac{-(i-j)^2}{2 \cdot 1^2}\right)$	1.000	1.080	1.322	1.650	1.982	2.312	2.624	2.842
$\exp\left(\frac{-(i-j)^2}{2 \cdot 2^2}\right)$	1.000	1.149	1.411	1.689	1.967	2.243	2.504	2.686
$\exp\left(\frac{-(i-j)^2}{2 \cdot 10^2}\right)$	1.000	1.006	1.045	1.216	1.542	1.891	2.223	2.454
$\exp\left(\frac{-(i-j)^2}{2 \cdot 100^2}\right)$	1.000	1.000	1.001	1.005	1.039	1.190	1.489	1.721
$\exp\left(\frac{-(i-j)^2}{2 \cdot 1000^2}\right)$	1.000	1.000	1.000	1.000	1.001	1.004	1.027	1.110

the change of spectral norms, while the benefit of level increasing lies in decreasing of the complexity of Algorithm 6.1, which can be seen from (20).

The estimation of spectral norms of random matrices generating from i.i.d. normal distributions is not good by the norm compression hierarchy, observed both from Tables 1 and 2. Perhaps it could be the nature of these matrices, as other classical upper bounds are not good as well (Table 1). For nonnegative matrices, low-rank matrices and covariance matrices, the bounds obtained by Algorithm 6.1 are very good, and the algorithm enjoys a low complexity.

To conclude this section, we remark that Algorithm 6.1 can be easily extended to nonsquare matrices ($\mathbb{R}^{n_1 \times n_2}$ with $n_1 \neq n_2$), to higher order tensors ($\mathbb{R}^{n_1 \times n_2 \times \dots \times n_d}$ with $d \geq 3$), to different factorizations of n_k 's of a tensor, and to modal partitions having different sizes of subtensors. We do not extensively expand here. The least message is that estimating the matrix spectral norm via norm compression hierarchies

can be done in $O(n^{2+\epsilon})$ operations with a good accuracy. An important message is that for any tensor or matrix norm computation methods, no matter existed or being developed, the approach via norm compression hierarchies can speed up these methods while keep good approximability.

To briefly conclude the whole paper, we proposed norm compression inequalities for partitioned block tensors. For any spectral p -norm, the norm of a norm compression tensor is an upper bound of the norm of the original tensor. By applying these inequalities, various bounds of tensor and matrix spectral norms in the literature can be improved. Norm compression inequities for tensors have been shown good potential in studying the extremal ratio between the spectral norm and the Frobenius norm of a tensor space, and estimating tensor and matrix norms via norm compression hierarchies. We think this is a promising start and the research can be further extended both in theory and in applications.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Agrachev, A., Kozhasov, K., Uschmajew, A.: Chebyshev polynomials and best rank-one approximation ratio. *SIAM J. Matrix Anal. Appl.* (2020) (to appear)
2. Audenaert, K.M.R.: A norm compression inequality for block partitioned positive definite matrices. *Linear Algebra Appl.* **413**, 155–176 (2006)
3. Audenaert, K.M.R.: On a norm compression inequality for $2 \times N$ partitioned block matrices. *Linear Algebra Appl.* **428**, 781–795 (2008)
4. Bebendorf, M.: *Hierarchical Matrices: A Means to Efficiently Solve Elliptic Boundary Value Problems*. Springer, Berlin (2008)
5. Bennett, C.H., Shor, P.W.: Quantum information theory. *IEEE Trans. Inf. Theory* **44**, 2724–2748 (1998)
6. Bhatia, R., Kittaneh, F.: Norm inequalities for partitioned operators and an application. *Math. Ann.* **287**, 719–726 (1990)
7. Cai, T.T., Yuan, M.: Adaptive covariance matrix estimation through block thresholding. *Ann. Stat.* **40**, 2014–2042 (2012)
8. Chen, B., He, S., Li, Z., Zhang, S.: Maximum block improvement and polynomial optimization. *SIAM J. Optim.* **22**, 87–107 (2012)
9. Chen, B., Li, Z.: On the tensor spectral p -norm and its dual norm via partitions. *Comput. Optim. Appl.* (2020). <https://doi.org/10.1007/s10589-020-00177-z>
10. Chen, L., Liu, Y., Zhu, C.: Iterative block tensor singular value thresholding for extraction of lowrank component of image data. In: *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1862–1866 (2017)
11. Derksen, H.: On the nuclear norm and the singular value decomposition of tensors. *Found. Comput. Math.* **16**, 779–811 (2016)
12. Friedland, S., Lim, L.-H.: Nuclear norm of higher-order tensors. *Math. Comput.* **87**, 1255–1281 (2018)
13. Gautier, A., Tudisco, F., Hein, M.: A unifying Perron–Frobenius theorem for nonnegative tensors via multihomogeneous maps. *SIAM J. Matrix Anal. Appl.* **40**, 1206–1231 (2019)

14. Golub, G.H., Van Loan, C.F.: *Matrix Computations*, 4th edn. Johns Hopkins University Press, Baltimore (2013)
15. Hardy, G.H., Littlewood, J.E., Pólya, G.: *Inequalities*. Cambridge University Press, Cambridge (1934)
16. He, S., Jiang, B., Li, Z., Zhang, S.: Probability bounds for polynomial functions in random variables. *Math. Oper. Res.* **39**, 889–907 (2014)
17. He, S., Li, Z., Zhang, S.: Approximation algorithms for homogeneous polynomial optimization with quadratic constraints. *Math. Program.* **125**, 353–383 (2010)
18. He, S., Li, Z., Zhang, S.: Approximation algorithms for discrete polynomial optimization. *J. Oper. Res. Soc. China* **1**, 3–36 (2013)
19. Hillar, C.J., Lim, L.-H.: Most tensor problems are NP-hard. *J. ACM* **60**, 45 (2013)
20. Hou, K., So, A.M.-C.: Hardness and approximation results for L_p -ball constrained homogeneous polynomial optimization problems. *Math. Oper. Res.* **39**, 1084–1108 (2014)
21. Hu, S.: Relations of the nuclear norm of a tensor and its matrix flattenings. *Linear Algebra Appl.* **478**, 188–199 (2015)
22. Jiang, B., Ma, S., Zhang, S.: Tensor principal component analysis via convex optimization. *Math. Program.* **150**, 423–457 (2015)
23. King, C.: Inequalities for trace norms of 2×2 block matrices. *Commun. Math. Phys.* **242**, 531–545 (2003)
24. King, C., Nathanson, M.: New trace norm inequalities for 2×2 blocks of diagonal matrices. *Linear Algebra Appl.* **389**, 77–93 (2004)
25. Kolda, T.G., Bader, B.W.: Tensor decompositions and applications. *SIAM Rev.* **51**, 455–500 (2009)
26. Kong, X., Meng, D.: The bounds for the best rank-1 approximation ratio of a finite dimensional tensor space. *Pac. J. Optim.* **11**, 323–337 (2015)
27. Kuczynski, J., Woźniakowski, H.: Estimating the largest eigenvalue by the power and Lanczos algorithms with a random start. *SIAM J. Matrix Anal. Appl.* **13**, 1094–1122 (1992)
28. Kühn, T., Peetre, J.: Embedding constants of trilinear Schatten–von Neumann classes. *Proc. Estonian Acad. Sci. Phys. Math.* **55**, 174–181 (2006)
29. Li, Z.: Bounds on the spectral norm and the nuclear norm of a tensor based on tensor partitions. *SIAM J. Matrix Anal. Appl.* **37**, 1440–1452 (2016)
30. Li, Z., He, S., Zhang, S.: *Approximation Methods for Polynomial Optimization: Models, Algorithms, and Applications*. Springer, New York (2012)
31. Li, Z., Nakatsukasa, Y., Soma, T., Uschmajew, A.: On orthogonal tensors and best rank-one approximation ratio. *SIAM J. Matrix Anal. Appl.* **39**, 400–425 (2018)
32. Lim, L.-H.: Singular values and eigenvalues of tensors: a variational approach. In: *Proceedings of the IEEE International Workshop on Computational Advances in Multi-sensor Adaptive Processing*, vol. 1, pp. 129–132 (2005)
33. Lim, L.-H.: Tensors and hypermatrices. In: Hogben, L. (ed.) *Handbook of Linear Algebra*, 2nd edn. CRC Press, Boca Raton (2013)
34. Lim, L.-H., Comon, P.: Blind multilinear identification. *IEEE Trans. Inf. Theory* **60**, 1260–1280 (2014)
35. Magdon-Ismail, M.: A note on estimating the spectral norm of a matrix efficiently (2011). [arXiv:1104.2076](https://arxiv.org/abs/1104.2076)
36. Mueller-Smith, C., Spasojević, P.: Column-wise symmetric block partitioned tensor decomposition. In: *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 2956–2960 (2016)
37. Nie, J., Wang, L.: Semidefinite relaxations for best rank-1 tensor approximations. *SIAM J. Matrix Anal. Appl.* **35**, 1155–1179 (2014)
38. Nikiforov, V.: Combinatorial methods for the spectral p -norm of hypermatrices. *Linear Algebra Appl.* **529**, 324–354 (2017)
39. Phan, A.H., Cichocki, A.: Block decomposition for very large-scale nonnegative tensor factorization. In: *Proceedings of the IEEE International Workshop on Computational Advances in Multi-sensor Adaptive Processing*, vol. 3, pp. 316–319 (2009)
40. Qi, L.: The best rank-one approximation ratio of a tensor space. *SIAM J. Matrix Anal. Appl.* **32**, 430–442 (2011)
41. Ragnarsson, S., Van Loan, C.F.: Block tensor unfoldings. *SIAM J. Matrix Anal. Appl.* **33**, 149–169 (2012)
42. Ragnarsson, S., Van Loan, C.F.: Block tensors and symmetric embeddings. *Linear Algebra Appl.* **438**, 853–874 (2013)

43. Schur, I.: Bemerkungen zur Theorie der beschränkten Bilinearformen mit unendlich vielen Veränderlichen. *J. Reine Angew. Math.* **140**, 1–28 (1911)
44. So, A.M.-C.: Deterministic approximation algorithms for sphere constrained homogeneous polynomial optimization problems. *Math. Program.* **192**, 357–382 (2011)
45. Tonge, A.: Equivalence constants for matrix norms: a problem of goldberg. *Linear Algebra Appl.* **306**, 1–13 (2000)
46. Uschmajew, A.: Some results concerning rank-one truncated steepest descent directions in tensor spaces. In: *Proceedings of the International Conference on Sampling Theory and Applications*, pp. 415–419 (2015)
47. Vannieuwenhoven, N., Meerbergen, K., Vandebril, R.: Computing the gradient in optimization algorithms for the CP decomposition in constant memory through tensor blocking. *SIAM J. Sci. Comput.* **37**, C415–C438 (2015)
48. Wang, M., Dao Duc, K., Fischer, J., Song, Y.S.: Operator norm inequalities between tensor unfoldings on the partition lattice. *Linear Algebra Appl.* **520**, 44–66 (2017)
49. Zhang, X., Ling, C., Qi, L.: The best rank-1 approximation of a symmetric tensor and related spherical optimization problems. *SIAM J. Matrix Anal. Appl.* **33**, 806–821 (2012)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.