BOUNDS ON THE SPECTRAL NORM AND THE NUCLEAR NORM OF A TENSOR BASED ON TENSOR PARTITIONS[*]

ZHENING LI[†]

Abstract. It is known that computing the spectral norm and the nuclear norm of a tensor is NP-hard in general. In this paper, we provide neat bounds for the spectral norm and the nuclear norm of a tensor based on tensor partitions. The spectral norm (respectively, the nuclear norm) can be lower and upper bounded by manipulating the spectral norms (respectively, the nuclear norms) of its subtensors. The bounds are sharp in general. When a tensor is partitioned into its matrix slices, our inequalities provide polynomial-time worst-case approximation bounds for computing the spectral norm and the nuclear norm of the tensor.

1. Introduction. Given any tensor $\mathcal{T}$ that is well partitioned (a formal definition called a regular partition is given in Definition 2.5) into any set of subtensors $\{\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_m\}$, its spectral norm $\|\mathcal{T}\|_\sigma$ and its nuclear norm $\|\mathcal{T}\|_*$ are bounded as follows:

$$(1) \quad \|(\|\mathcal{T}_1\|_\sigma, \|\mathcal{T}_2\|_\sigma, \ldots, \|\mathcal{T}_m\|_\sigma)\|_\infty \leq \|\mathcal{T}\|_\sigma \leq \|(\|\mathcal{T}_1\|_\sigma, \|\mathcal{T}_2\|_\sigma, \ldots, \|\mathcal{T}_m\|_\sigma)\|_2,$$

$$(2) \quad \|(\|\mathcal{T}_1\|_*, \|\mathcal{T}_2\|_*, \ldots, \|\mathcal{T}_m\|_*)\|_2 \leq \|\mathcal{T}\|_* \leq \|(\|\mathcal{T}_1\|_*, \|\mathcal{T}_2\|_*, \ldots, \|\mathcal{T}_m\|_*)\|_1,$$

where $\|\cdot\|_p$ stands for the L_p norm of a vector for $1 \leq p \leq \infty$. These intuitive bounds are tight, in general. In this paper we will prove these inequalities and show some preliminary applications.

The spectral norm and the nuclear norm of a matrix are evidently important in many branches of mathematics as well as in various practical applications; see, e.g., [10, 5]. They are easy to compute from the singular value decompositions (SVD) since the matrix spectral norm is the largest singular value and the matrix nuclear norm is the sum of all the singular values. The matrix nuclear norm is known to be dual to the matrix spectral norm. In recent years, due to the surge of research on studying various tensor data problems, the use of the tensor spectral norm and tensor nuclear norm are widely seen, in particular, in tensor decompositions (see, e.g., [12]) and tensor completions (see, e.g., [4]). Lim [14] and Qi [18] independently defined the eigenvalue and the singular value of a tensor in 2005. It turns out that the spectral norm of a tensor is nothing but the largest singular value of the tensor [14]. Similar to the matrix case, the tensor nuclear norm is also the dual norm to the tensor spectral norm [15].

On the computational side, the spectral norm of a tensor is NP-hard to compute [7, 9]. It is equivalent to finding the best rank-one approximation of a tensor [12], as well as the sphere constrained homogenous polynomial optimization problem [7].

[†]Department of Mathematics, University of Portsmouth, Portsmouth PO1 3HF, United Kingdom (zheningli@gmail.com).

Various practical methods (e.g., higher order SVD [1]) as well as some approximation methods [7, 20] were proposed. However, computing the nuclear norm of a tensor is less understood. Recently, Friedland and Lim [3] showed that the computational complexity of the tensor nuclear norm is also NP-hard. They also proposed simple lower and upper bounds for the tensor spectral norm and the tensor nuclear norm [3, Lemma 9.1], which are implied by a special case of our main results in this paper. Derksen [2] discussed the nuclear norm of various tensors of interests based on orthogonality. A general lower bound [2, Theorem 1.4] on the tensor nuclear norm was proposed. It should be noted that the formal definition of the tensor nuclear norm (see Definition 2.2) is different than the one from the nuclear norms of its flattening matrices that is practically used in tensor completions [4, 17]. In fact, matrix flattenings of tensors have been a major tool to study the tensor spectral norm and nuclear norm both in theory and in practice. He, Li, and Zhang [7] essentially applied the matrix flattening to obtain a worst-case approximation bound for the tensor spectral norm. Hu [11] recently proved that the nuclear norm of the matrix flattening of the underlying tensor is a lower bound of the tensor nuclear norm, from which an upper bound of the tensor nuclear norm was also established.

Although matrix flattenings of tensors are widely used, to the best of our knowledge, the study on the tensor spectral norm and nuclear norm from tensor partitions is not to be found in the literature. The general bounds (1) and (2) established in this paper look through the tensor norms from a new perspective. As an immediate application, they provide easy computable worst-case approximation bounds for the tensor spectral norm and nuclear norm. The approximation bound for the tensor nuclear norm is currently the best. We believe these inequalities will have potential both in theory and in practice.

2. Preparation. Throughout this paper, the usual lower case letters (e.g., x), the boldface lower case letters (e.g., $\boldsymbol{x} = (x_i)$), the capital letters (e.g., $X = (x_{ij})$), and the calligraphic letters (e.g., $\mathcal{X} = (x_{i_1 i_2 \ldots i_d})$) denote scalars, vectors, matrices, and higher order (order three or more) tensors, respectively. We denote $\mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ to be the space of dth order real tensors of dimension $n_1 \times n_2 \times \cdots \times n_d$. The same notations apply for the vector space and the matrix space when $d = 1$ and $d = 2$, respectively. $\mathbb{N}$ denotes the set of positive integers.

2.1. Tensor norms. Associated with the tensor space $\mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ is the Frobenius inner product

$$\langle \mathcal{A}, \mathcal{B} \rangle := \sum_{i_1=1}^{n_1} \sum_{i_1=2}^{n_2} \cdots \sum_{i_d=1}^{n_d} a_{i_1 i_2 \ldots i_d} b_{i_1 i_2 \ldots i_d},$$

where $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$. The induced Frobenius norm is naturally defined as

$$\|\mathcal{A}\|_2 := \sqrt{\langle \mathcal{A}, \mathcal{A} \rangle},$$

which is also known as the Hilbert–Schmidt norm in the literature [15]. When $d = 1$ the Frobenius norm reduces 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) for $1 \le p \le \infty$ by looking at a tensor as a vector, as follows:

$$\|\mathcal{A}\|_p = \left(\sum_{i_1=1}^{n_1} \sum_{i_1=2}^{n_2} \cdots \sum_{i_d=1}^{n_d} |a_{i_1 i_2 \ldots i_d}|^p \right)^{1/p}.$$

One should note that the Hölder p-norm defined above is different than the matrix p-norm that is defined from the operator [8].

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

$$\mathcal{T} = \boldsymbol{x}^1 \otimes \boldsymbol{x}^2 \otimes \cdots \otimes \boldsymbol{x}^d.$$

It is easy to verify that $\|\mathcal{T}\|_2 = \prod_{k=1}^d \|\boldsymbol{x}^d\|_2$.

A mode-k product of a vector $\boldsymbol{x} \in \mathbb{R}^{n_k}$ on a tensor $\mathcal{T} = (t_{i_1 i_2 \dots i_d}) \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$, denoted by $\mathcal{T} \times_k \boldsymbol{x}$, is a $(d-1)$th order tensor in $\mathbb{R}^{n_1 \times \cdots \times n_{k-1} \times n_{k+1} \times \cdots \times n_d}$, whose $(i_1, \dots, i_{k-1}, i_{k+1}, \dots, i_d)$th entry is

$$\sum_{j=1}^{n_k} t_{i_1 \dots i_{k-1} j i_{k+1} \dots i_d} x_j$$

for all $(i_1, \dots, i_{k-1}, i_{k+1}, \dots, i_d)$. This is the same mode-$k$ product of a matrix on a tensor widely used in the literature (see, e.g., [12]) by looking at the vector $\boldsymbol{x}$ as a $1 \times n_k$ matrix.

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

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

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.

DEFINITION 2.2. *For a given tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$, the nuclear norm of $\mathcal{T}$, denoted by $\|\mathcal{T}\|_*$, is defined as*

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

The tensor nuclear norm is the dual norm to the tensor spectral norm, and vice versa.

LEMMA 2.3. *For a given tensor $\mathcal{T}$, it follows that*

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

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

One may find the proof of Lemma 2.3 in [16, 2]. It can be seen that the spectral norm and the nuclear norm of a tensor are generalizations of the spectral norm and the nuclear norm of a matrix, respectively. It is worth mentioning that both the spectral norm and the nuclear norm of a vector are nothing but the L_2 norm of the vector, which fact is helpful in explaining our main theorem in some special cases. Unlike the matrix spectral norm and the matrix nuclear norm that can be computed in polynomial time, NP-hardness to compute the tensor spectral norm and the tensor nuclear norm was proved by He, Li, and Zhang [7] and Friedland and Lim [3], respectively.

2.2. Tensor partitions. A matrix can be partitioned into submatrices, the same can be applied for a tensor. One important class of tensor partitions, the block tensor, was already studied by Ragnarsson and Van Loan [19], which is a straightforward generalization of the block matrix. In fact, we focus on a more general class of tensor partitions than the block tensor studied in [19]. Let us specifically introduce these concepts of tensor partitions.

DEFINITION 2.4. *A partition $\{\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_m\}$ is called a tensor partition of a tensor $\mathcal{T}$ if*
- *every $\mathcal{T}_j$ $(j = 1, 2, \ldots, m)$ is a subtensor of $\mathcal{T}$,*
- *every pair of subtensors $\{\mathcal{T}_i, \mathcal{T}_j\}$ with $i \neq j$ has no common entry of $\mathcal{T}$, and*
- *every entry of $\mathcal{T}$ belongs to one of the subtensors in $\{\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_m\}$.*

Given a dth order tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$, 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

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

We will use the same notation to cut a matrix and cut a vector. In particular, for a first order tensor, a vector $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{x} = \boldsymbol{x}_1 \vee_1 \boldsymbol{x}_2$ is exactly same as $\boldsymbol{x}^{\mathrm{T}} = (\boldsymbol{x}_1^{\mathrm{T}}, \boldsymbol{x}_2^{\mathrm{T}})$. The mode subscript of $\vee$ in a tensor cut is sometimes omitted for clearer presentations. For example, $\mathcal{T} = \mathcal{T}_1 \vee \mathcal{T}_2$ 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 $\vee_k$ is applied, the dimensions of its associated two tensors must be the same in every mode except mode k.

Let us now present the main definition in the paper by recursion.

DEFINITION 2.5. *$\{\mathcal{T}\}$ is called the 1-regular partition of a tensor $\mathcal{T}$. For $m \in \mathbb{N}$ with $m \geq 2$, a partition $\{\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_m\}$ is called an m-regular partition of a tensor $\mathcal{T}$ if there exist two tensors $\mathcal{F}_1, \mathcal{F}_2$ and $\ell \in \mathbb{N}$ with $1 \leq \ell \leq m - 1$, such that*
- *$\mathcal{T} = \mathcal{F}_1 \vee \mathcal{F}_2$,*
- *$\{\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_\ell\}$ is an ℓ-regular partition of $\mathcal{F}_1$, and*
- *$\{\mathcal{T}_{\ell+1}, \mathcal{T}_{\ell+2}, \ldots, \mathcal{T}_m\}$ is an $(m - \ell)$-regular partition of $\mathcal{F}_2$.*

In other words, an m-regular partition of $\mathcal{T}$ can be written by $\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_m$ and $m-1$ $\vee$'s with the help of $m-2$ pairs of brackets. For example, in the middle subfigure of Figure 1, $\{A_1, A_2, \ldots, A_5\}$ is a 5-regular partition of a second order tensor (matrix) A, written as

$$A = (A_1 \vee_2 (A_2 \vee_1 A_3)) \vee_1 (A_4 \vee_2 A_5),$$

where the row and the column of a matrix is denoted by mode 1 and mode 2, respectively. In plain language for an m-regular partition, one can assemble a tensor

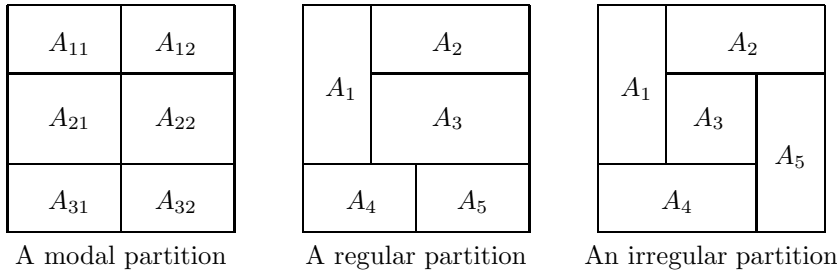| A modal partition | A regular partition | An irregular partition |

FIG. 1. *Tensor partitions of a second order tensor (matrix).*

from subtensors step by step $m-1$ times, where in each step stacking two subtensors remains a subtensor. From the original tensor point of view, an m-regular partition can be obtained by applying the mode-k tensor cut $m-1$ times sequentially, where k may vary in each cut. In the following discussion, an m-regular partition is often called a regular partition under no ambiguity.

To clarify different concepts of tensor partitions, we now discuss a special type of regular partitions called block tensors in the literature (e.g., [19]), which we will name as modal partitions. Given a tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$, the indices of its mode k can be partitioned into r_k nonempty sets, i.e., for $k = 1, 2, \ldots, d$,

$$\{1, 2, \ldots, n_k\} = \mathbb{I}_1^k \cup \mathbb{I}_2^k \cup \cdots \cup \mathbb{I}_{r_k}^k.$$

For simplicity, we assume that the indices in $\mathbb{I}_i^k$ are consecutive and the $\mathbb{I}_i^k$'s are monotonically increasing as i increases, since this can be done essentially via indices relabeling without affecting tensor norms.

DEFINITION 2.6. *The tensor partition* $\{\mathcal{T}_{j_1 j_2 \ldots j_d} : 1 \le j_k \le r_k, \, k = 1, 2, \ldots, d\}$ *is called a modal partition of a tensor* $\mathcal{T} = (t_{i_1 i_2 \ldots i_d}) \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$, *where*

$$\mathcal{T}_{j_1 j_2 \ldots j_d} := \left((t_{i_1 i_2 \ldots i_d})_{i_k \in \mathbb{I}_{j_k}^k, \, i=1,2,\ldots,d} \right).$$

An example of modal partition for the matrix case (block matrix) is the left subfigure of Figure 1. Essentially in a modal partition, all the tensor cuts must cut the original tensor $\mathcal{T}$ simultaneously, while in a regular partition the tensor cuts can be applied on subtensors sequentially.

It is obvious that a modal partition is a special type of regular partition, and the latter is a special type of tensor partition. It is easy to see that for any first order tensor (vector), these three partitions are the same. However, this is not true for a second or higher order tensor; see Figure 1 for the matrix case. For an irregular partition such as the right subfigure of Figure 1, one cannot find the first tensor cut, either mode-1 cut or mode-2 cut, to cut the whole matrix into two without breaking any submatrix A_j.

Finally, we remark that a subtensor $\mathcal{T}_j$ in a regular partition of a tensor $\mathcal{T} = \{\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_m\}$ may not have the same order as the original tensor $\mathcal{T}$. If the dimension of some mode of $\mathcal{T}_j$ is reduced to one due to a tensor cut on that mode, the order of $\mathcal{T}_j$ is automatically reduced by one. This is the same for any tensor partition. In an extreme case, a tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ can be regularly partitioned into $\prod_{k=1}^{d} n_k$ number of scalars (zeroth order tensors), which is also a modal partition.

3. The bounds of tensor norms. This section is devoted to the proof and discussion of our main results in the paper, described as follows.

THEOREM 3.1. *If* $\{\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_m\}$ *is a regular partition of a tensor* $\mathcal{T}$, *then*

$$(3) \quad \|(\|\mathcal{T}_1\|_\sigma, \|\mathcal{T}_2\|_\sigma, \ldots, \|\mathcal{T}_m\|_\sigma)\|_\infty \le \|\mathcal{T}\|_\sigma \le \|(\|\mathcal{T}_1\|_\sigma, \|\mathcal{T}_2\|_\sigma, \ldots, \|\mathcal{T}_m\|_\sigma)\|_2,$$

$$(4) \quad \|(\|\mathcal{T}_1\|_*, \|\mathcal{T}_2\|_*, \ldots, \|\mathcal{T}_m\|_*)\|_2 \le \|\mathcal{T}\|_* \le \|(\|\mathcal{T}_1\|_*, \|\mathcal{T}_2\|_*, \ldots, \|\mathcal{T}_m\|_*)\|_1.$$

3.1. Proof of the main results. Let us set out to prove Theorem 3.1. The following observations are essential.

- For any regular partition, the tensor $\mathcal{T}$ can be cut sequentially by applying a mode-k tensor cut $m-1$ times.

- The L_p norm of a vector has a certain additive property for $1 \leq p \leq \infty$, i.e., if $\boldsymbol{x} = \boldsymbol{x}^1 \vee \boldsymbol{x}^2 \in \mathbb{R}^{n_1 + n_2}$ with $\boldsymbol{x}^1 \in \mathbb{R}^{n_1}$ and $\boldsymbol{x}^2 \in \mathbb{R}^{n_2}$, then

$$\|(\|\boldsymbol{x}^1\|_p, \|\boldsymbol{x}^2\|_p)\|_p = \|\boldsymbol{x}\|_p.$$

According to these properties, it is not hard to see that (3) can be obtained from Lemma 3.2 (the spectral norm case for $m = 2$) and (4) can be obtained from Lemma 3.3 (the nuclear norm case for $m = 2$), both by induction.

LEMMA 3.2. *If $\mathcal{T} = \mathcal{A} \vee \mathcal{B}$, then* $\max\{\|\mathcal{A}\|_\sigma, \|\mathcal{B}\|_\sigma\} \leq \|\mathcal{T}\|_\sigma \leq \sqrt{\|\mathcal{A}\|_\sigma^2 + \|\mathcal{B}\|_\sigma^2}$.

Proof. Without loss of generality, we suppose $\mathcal{T} = \mathcal{A} \vee_1 \mathcal{B}$. Let $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$, $\mathcal{A} \in \mathbb{R}^{\ell_1 \times n_2 \times \cdots \times n_d}$, and $\mathcal{B} \in \mathbb{R}^{\ell_2 \times n_2 \times \cdots \times n_d}$ with $\ell_1 + \ell_2 = n_1$.

Let $\|\mathcal{A}\|_\sigma = \langle \mathcal{A}, \boldsymbol{x}^1 \otimes \boldsymbol{x}^2 \otimes \cdots \otimes \boldsymbol{x}^d \rangle$, where $\|\boldsymbol{x}^k\|_2 = 1$ for $k = 1, 2, \ldots, d$. Since $\boldsymbol{x}^1 \in \mathbb{R}^{\ell_1}$ and $\ell_1 \leq n_1$, we may let $\boldsymbol{x}^0 = \boldsymbol{x}^1 \vee \boldsymbol{0} \in \mathbb{R}^{n_1}$. Clearly $\|\boldsymbol{x}^0\|_2 = \|\boldsymbol{x}^1\|_2 = 1$. Therefore, by Definition 2.1

$$\begin{aligned}
\|\mathcal{T}\|_\sigma &\geq \langle \mathcal{T}, \boldsymbol{x}^0 \otimes \boldsymbol{x}^2 \otimes \cdots \otimes \boldsymbol{x}^d \rangle \\
&= \langle \mathcal{A} \vee_1 \mathcal{B}, (\boldsymbol{x}^1 \vee \boldsymbol{0}) \otimes \boldsymbol{x}^2 \otimes \cdots \otimes \boldsymbol{x}^d \rangle \\
&= \langle \mathcal{A}, \boldsymbol{x}^1 \otimes \boldsymbol{x}^2 \otimes \cdots \otimes \boldsymbol{x}^d \rangle + \langle \mathcal{B}, \boldsymbol{0} \otimes \boldsymbol{x}^2 \otimes \cdots \otimes \boldsymbol{x}^d \rangle \\
&= \|\mathcal{A}\|_\sigma.
\end{aligned}$$

Similarly we have $\|\mathcal{T}\|_\sigma \geq \|\mathcal{B}\|_\sigma$, proving the lower bound of $\|\mathcal{T}\|_\sigma$.

For the upper bound, one important observation is that

$$\|\mathcal{F}\|_\sigma = \max_{\|\boldsymbol{x}\|_2 \leq 1} \|\mathcal{F} \times_1 \boldsymbol{x}\|_\sigma$$

for any tensor $\mathcal{F}$. This is because

$$\begin{aligned}
\|\mathcal{F}\|_\sigma &= \max_{\|\boldsymbol{x}^k\|_2 \leq 1, \, k=1,2,\ldots,d} \langle \mathcal{F}, \boldsymbol{x}^1 \otimes \boldsymbol{x}^2 \otimes \cdots \otimes \boldsymbol{x}^d \rangle \\
&= \max_{\|\boldsymbol{x}^1\|_2 \leq 1} \max_{\|\boldsymbol{x}^k\|_2 \leq 1, \, k=2,3,\ldots,d} \langle \mathcal{F}, \boldsymbol{x}^1 \otimes \boldsymbol{x}^2 \otimes \cdots \otimes \boldsymbol{x}^d \rangle \\
&= \max_{\|\boldsymbol{x}^1\|_2 \leq 1} \max_{\|\boldsymbol{x}^k\|_2 \leq 1, \, k=2,3,\ldots,d} \langle \mathcal{F} \times_1 \boldsymbol{x}^1, \boldsymbol{x}^2 \otimes \boldsymbol{x}^3 \otimes \cdots \otimes \boldsymbol{x}^d \rangle \\
&= \max_{\|\boldsymbol{x}^1\|_2 \leq 1} \|\mathcal{F} \times_1 \boldsymbol{x}^1\|_\sigma.
\end{aligned}$$

Therefore $\|\mathcal{T}\|_\sigma = \max_{\|\boldsymbol{x}\|_2 \leq 1} \|\mathcal{T} \times_1 \boldsymbol{x}\|_\sigma$. Let $\boldsymbol{x} = \boldsymbol{y} \vee \boldsymbol{z}$, where $\boldsymbol{y} \in \mathbb{R}^{\ell_1}$, and $\boldsymbol{z} \in \mathbb{R}^{\ell_2}$. Let further $y = \|\boldsymbol{y}\|_2$, $z = \|\boldsymbol{z}\|_2$, $\boldsymbol{y} = y\boldsymbol{y}^0$, and $\boldsymbol{z} = z\boldsymbol{z}^0$. We then have

$$\begin{aligned}
\|\mathcal{T}\|_\sigma &= \max_{\|\boldsymbol{x}\|_2 \leq 1} \|\mathcal{T} \times_1 \boldsymbol{x}\|_\sigma \\
&= \max_{\|\boldsymbol{y}\|_2^2 + \|\boldsymbol{z}\|_2^2 \leq 1} \|(\mathcal{A} \vee_1 \mathcal{B}) \times_1 (\boldsymbol{y} \vee \boldsymbol{z})\|_\sigma \\
&= \max_{\|\boldsymbol{y}\|_2^2 + \|\boldsymbol{z}\|_2^2 \leq 1} \|\mathcal{A} \times_1 \boldsymbol{y} + \mathcal{B} \times_1 \boldsymbol{z}\|_\sigma \\
&\leq \max_{\|\boldsymbol{y}\|_2^2 + \|\boldsymbol{z}\|_2^2 \leq 1} (\|\mathcal{A} \times_1 \boldsymbol{y}\|_\sigma + \|\mathcal{B} \times_1 \boldsymbol{z}\|_\sigma) \\
&= \max_{y^2 + z^2 \leq 1, \, \|\boldsymbol{y}^0\|_2 \leq 1, \, \|\boldsymbol{z}^0\|_2 \leq 1} (\|\mathcal{A} \times_1 (y\boldsymbol{y}^0)\|_\sigma + \|\mathcal{B} \times_1 (z\boldsymbol{z}^0)\|_\sigma) \\
&= \max_{y^2 + z^2 \leq 1} \left(\max_{\|\boldsymbol{y}^0\|_2 \leq 1} y\|\mathcal{A} \times_1 \boldsymbol{y}^0\|_\sigma + \max_{\|\boldsymbol{z}^0\|_2 \leq 1} z\|\mathcal{B} \times_1 \boldsymbol{z}^0\|_\sigma \right) \\
&= \max_{y^2 + z^2 \leq 1} (y\|\mathcal{A}\|_\sigma + z\|\mathcal{B}\|_\sigma) \\
&= \sqrt{\|\mathcal{A}\|_\sigma^2 + \|\mathcal{B}\|_\sigma^2}. \qquad \square
\end{aligned}$$

LEMMA 3.3. *If* $\mathcal{T} = \mathcal{A} \vee \mathcal{B}$, *then* $\sqrt{\|\mathcal{A}\|_*^2 + \|\mathcal{B}\|_*^2} \leq \|\mathcal{T}\|_* \leq \|\mathcal{A}\|_* + \|\mathcal{B}\|_*$.

Proof. Without loss of generality, we suppose $\mathcal{T} = \mathcal{A} \vee_1 \mathcal{B}$. Let $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$, $\mathcal{A} \in \mathbb{R}^{\ell_1 \times n_2 \times \cdots \times n_d}$, and $\mathcal{B} \in \mathbb{R}^{\ell_2 \times n_2 \times \cdots \times n_d}$ with $\ell_1 + \ell_2 = n_1$.

The upper bound follows easily from the triangle inequality, since

$$\begin{aligned}
\|\mathcal{T}\|_* &= \|\mathcal{A} \vee_1 \mathcal{B}\|_* \\
&= \|\mathcal{A} \vee_1 (0 \cdot \mathcal{B}) + (0 \cdot \mathcal{A}) \vee_1 \mathcal{B}\|_* \\
&\leq \|\mathcal{A} \vee_1 (0 \cdot \mathcal{B})\|_* + \|(0 \cdot \mathcal{A}) \vee_1 \mathcal{B}\|_* \\
&= \|\mathcal{A}\|_* + \|\mathcal{B}\|_*.
\end{aligned}$$

The lower bound is obtained from the dual norm point of view (Lemma 2.3). In particular, for any $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$, let $\mathcal{X} = \mathcal{Y} \vee_1 \mathcal{Z}$, where $\mathcal{Y} \in \mathbb{R}^{\ell_1 \times n_2 \times \cdots \times n_d}$ and $\mathcal{Z} \in \mathbb{R}^{\ell_2 \times n_2 \times \cdots \times n_d}$, i.e., $\mathcal{Y}$ and $\mathcal{Z}$ are in the same tensor spaces as $\mathcal{A}$ and $\mathcal{B}$, respectively. By Lemma 3.2 we have that

$$\|\mathcal{Y}\|_\sigma^2 + \|\mathcal{Z}\|_\sigma^2 \leq 1 \implies \|\mathcal{X}\|_\sigma \leq 1.$$

Therefore, according to Lemma 2.3

$$\begin{aligned}
\|\mathcal{T}\|_* &= \max_{\|\mathcal{X}\|_\sigma \leq 1} \langle \mathcal{T}, \mathcal{X} \rangle \\
&\geq \max_{\|\mathcal{Y}\|_\sigma^2 + \|\mathcal{Z}\|_\sigma^2 \leq 1} \langle \mathcal{A} \vee_1 \mathcal{B}, \mathcal{Y} \vee_1 \mathcal{Z} \rangle \\
&= \max_{\|\mathcal{Y}\|_\sigma^2 + \|\mathcal{Z}\|_\sigma^2 \leq 1} (\langle \mathcal{A}, \mathcal{Y} \rangle + \langle \mathcal{B}, \mathcal{Z} \rangle).
\end{aligned}$$

Let $y = \|\mathcal{Y}\|_\sigma$, $z = \|\mathcal{Z}\|_\sigma$, $\mathcal{Y} = y\mathcal{Y}_0$, and $\mathcal{Z} = z\mathcal{Z}_0$. The above inequality further leads to

$$\begin{aligned}
\|\mathcal{T}\|_* &\geq \max_{\|y\mathcal{Y}_0\|_\sigma^2 + \|z\mathcal{Z}_0\|_\sigma^2 \leq 1} (\langle \mathcal{A}, y\mathcal{Y}_0 \rangle + \langle \mathcal{B}, z\mathcal{Z}_0 \rangle) \\
&= \max_{y^2 + z^2 \leq 1, \, \|\mathcal{Y}_0\|_\sigma \leq 1, \, \|\mathcal{Z}_0\|_\sigma \leq 1} (\langle \mathcal{A}, y\mathcal{Y}_0 \rangle + \langle \mathcal{B}, z\mathcal{Z}_0 \rangle) \\
&= \max_{y^2 + z^2 \leq 1} \left(\max_{\|\mathcal{Y}_0\|_\sigma \leq 1} y\langle \mathcal{A}, \mathcal{Y}_0 \rangle + \max_{\|\mathcal{Z}_0\|_\sigma \leq 1} z\langle \mathcal{B}, \mathcal{Z}_0 \rangle \right) \\
&= \max_{y^2 + z^2 \leq 1} (y\|\mathcal{A}\|_* + z\|\mathcal{B}\|_*) \\
&= \sqrt{\|\mathcal{A}\|_*^2 + \|\mathcal{B}\|_*^2}. \qquad \square
\end{aligned}$$

3.2. Discussion on the main results. Let us discuss some special cases of Theorem 3.1. When a tensor $\mathcal{T}$ is partitioned entrywise into $\prod_{k=1}^d n_k$ number of scalars, they reduce to the following inequalities on tensor norms:

$$\|\mathcal{T}\|_\infty \leq \|\mathcal{T}\|_\sigma \leq \|\mathcal{T}\|_2 \leq \|\mathcal{T}\|_* \leq \|\mathcal{T}\|_1.$$

When a tensor $\mathcal{T}$ is partitioned into mode-k vector fibers, say

$$\{\boldsymbol{t}_i \in \mathbb{R}^{n_k} : i = 1, 2, \ldots, m\},$$

where $m = \prod_{1 \leq j \leq d, \, j \neq k} n_j$, the lower bound of $\|\mathcal{T}\|_\sigma$ is tightened to

$$\|(\|\boldsymbol{t}_1\|_2, \|\boldsymbol{t}_2\|_2, \ldots, \|\boldsymbol{t}_m\|_2)\|_\infty \leq \|\mathcal{T}\|_\sigma,$$

and the upper bound of $\|\mathcal{T}\|_*$ is tightened to

$$\|\mathcal{T}\|_* \le \|(\|\boldsymbol{t}_1\|_2, \|\boldsymbol{t}_2\|_2, \ldots, \|\boldsymbol{t}_m\|_2)\|_1 \,.$$

We will discuss the partition into matrix slices in the next section.

All the bounds in Theorem 3.1 are sharp, in general, in the sense that for any given tensor space and one of the four inequalities in Theorem 3.1, there exists a tensor in that space such that the inequality becomes an equality. It can be easily observed that both the lower bound (3) of $\|\mathcal{T}\|_\sigma$ and the upper bound (4) of $\|\mathcal{T}\|_*$ are tight if all but one of the subtensors in the regular partition $\{\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_m\}$ are zero tensors.

The upper bound (3) of $\|\mathcal{T}\|_\sigma$ and the lower bound (4) of $\|\mathcal{T}\|_*$ are certainly more significant. One interesting example for both bounds being tight is the following identity.

PROPOSITION 3.4. *Any regular partition $\{\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_m\}$ of a rank-one tensor $\mathcal{T}$ satisfies*

(5) $\quad \|(\|\mathcal{T}_1\|_\sigma, \|\mathcal{T}_2\|_\sigma, \ldots, \|\mathcal{T}_m\|_\sigma)\|_2 = \|\mathcal{T}\|_\sigma = \|\mathcal{T}\|_* = \|(\|\mathcal{T}_1\|_*, \|\mathcal{T}_2\|_*, \ldots, \|\mathcal{T}_m\|_*)\|_2 \,.$

Proof. Let $\mathcal{T} = (t_{i_1 i_2 \ldots i_d}) \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$. We observe that the maximum partition $\{t_{i_1 i_2 \ldots i_d} \in \mathbb{R} : 1 \le i_k \le n_k,\, k = 1, 2, \ldots, d\}$ of $\mathcal{T}$ is a further regular partition of any regular partition $\{\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_m\}$ of $\mathcal{T}$, i.e., $\mathcal{T}_j$ can be regularly partitioned into a subset of $\{t_{i_1 i_2 \ldots i_d} \in \mathbb{R} : 1 \le i_k \le n_k,\, k = 1, 2, \ldots, d\}$ for any $1 \le j \le m$. Noticing that any scalar $x \in \mathbb{R}$ has $\|x\|_\sigma = \|x\|_* = |x|$, by applying the upper bound of (3) for $\mathcal{T}$ and each $\mathcal{T}_j$ $(1 \le j \le m)$, one has
(6)

$$\|\mathcal{T}\|_\sigma \le \|(\|\mathcal{T}_1\|_\sigma, \|\mathcal{T}_2\|_\sigma, \ldots, \|\mathcal{T}_m\|_\sigma)\|_2 \le \left(\sum_{i_1=1}^{n_1} \sum_{i_1=2}^{n_2} \cdots \sum_{i_d=1}^{n_d} \|t_{i_1 i_2 \ldots i_d}\|_\sigma^2 \right)^{1/2} = \|\mathcal{T}\|_2,$$

and by applying the lower bound of (4) one also has
(7)

$$\|\mathcal{T}\|_2 = \left(\sum_{i_1=1}^{n_1} \sum_{i_1=2}^{n_2} \cdots \sum_{i_d=1}^{n_d} \|t_{i_1 i_2 \ldots i_d}\|_*^2 \right)^{1/2} \le \|(\|\mathcal{T}_1\|_*, \|\mathcal{T}_2\|_*, \ldots, \|\mathcal{T}_m\|_*)\|_2 \le \|\mathcal{T}\|_*.$$

On the other hand, as $\mathcal{T}$ is rank-one, $\mathcal{T}$ can be written as

$$\mathcal{T} = \lambda \boldsymbol{x}^1 \otimes \boldsymbol{x}^2 \otimes \cdots \otimes \boldsymbol{x}^d,$$

where $\|\boldsymbol{x}^k\|_2 = 1$ for $k = 1, 2, \ldots, d$. By Definition 2.1 we have that $\|\mathcal{T}\|_\sigma \ge \lambda$ and by Definition 2.2 we have that $\|\mathcal{T}\|_* \le \lambda$. Therefore

$$\|\mathcal{T}\|_\sigma \ge \lambda \ge \|\mathcal{T}\|_*.$$

Combining the above inequality with (6) and (7) leads to the final identity (5). $\quad\square$

Before concluding this section, we post a conjecture that any tensor partition (not necessarily regular) of a tensor satisfies the inequalities in Theorem 3.1.

CONJECTURE 3.5. *If $\{\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_m\}$ is a tensor partition of a tensor $\mathcal{T}$, then*

$$\|(\|\mathcal{T}_1\|_\sigma, \|\mathcal{T}_2\|_\sigma, \ldots, \|\mathcal{T}_m\|_\sigma)\|_\infty \le \|\mathcal{T}\|_\sigma \le \|(\|\mathcal{T}_1\|_\sigma, \|\mathcal{T}_2\|_\sigma, \ldots, \|\mathcal{T}_m\|_\sigma)\|_2,$$
$$\|(\|\mathcal{T}_1\|_*, \|\mathcal{T}_2\|_*, \ldots, \|\mathcal{T}_m\|_*)\|_2 \le \|\mathcal{T}\|_* \le \|(\|\mathcal{T}_1\|_*, \|\mathcal{T}_2\|_*, \ldots, \|\mathcal{T}_m\|_*)\|_1.$$

The lower bound of the spectral norm and the upper bound of the nuclear norm are straightforward from the proofs of Lemmas 3.2 and 3.3. Moreover, the upper bound of the spectral norm is equivalent to the lower bound of the nuclear norm by applying a similar argument in the proof of Lemma 3.3 from the dual norm point of view. Therefore, essentially we need to settle only one bound in the conjecture. We believe both bounds hold in general.

4. Approximating tensor norms. The bounds in Theorem 3.1 give rise to estimates of the tensor spectral norm and the tensor nuclear norm which are both NP-hard to compute in general. The flexibility on *any* regular partition of a tensor provides a useful tool for dealing with tensors that have specific structures in practice. Nevertheless, it is also important to study the approximation of tensor norms for a general tensor. Here in this section, we denote the tensor space to be $\mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ with $d \geq 3$, and suppose without loss of generality that $n_1 \leq n_2 \leq \cdots \leq n_d$.

DEFINITION 4.1. *A tensor norm $\|\cdot\|_\theta$ can be approximated with an approximation bound $\alpha \geq 1$ if there exists a polynomial-time approximation algorithm that computes a quantity $q_{\mathcal{T}}$ for any tensor instance $\mathcal{T}$, such that*

$$q_{\mathcal{T}} \leq \|\mathcal{T}\|_\theta \leq \alpha q_{\mathcal{T}}.$$

The closer α is to one, the better the approximation bound.

By virtue of Theorem 3.1 and the gaps among the L_1, L_2, and L_∞ norms of a vector, it is easy to see the following approximation bound for the spectral norm and the nuclear norm of a tensor.

COROLLARY 4.2. *If $\{\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_m\}$ is a regular partition of a tensor $\mathcal{T}$ and the tensor spectral norm $\|\mathcal{T}_j\|_\sigma$ (respectively, the tensor nuclear norm $\|\mathcal{T}_j\|_*$) can be computed in polynomial time for all $1 \leq j \leq m$, then the tensor spectral norm $\|\mathcal{T}\|_\sigma$ (respectively, the tensor nuclear norm $\|\mathcal{T}\|_*$) can be approximated with an approximation bound $\sqrt{m}$.*

One should note that the approximation bound is for the *worst-case* scenario. The actual bound working in practice may be much better than a theoretical proved bound. The approximation bound for the tensor spectral norm is well studied in the field of approximation methods for polynomial optimization [13]. He, Li, and Zhang [7] proposed a first approximation algorithm for the multilinear form optimization model that exactly computes the tensor spectral norm:

$$\begin{aligned} \max \quad & \mathcal{T}(\boldsymbol{x}^1, \boldsymbol{x}^2, \ldots, \boldsymbol{x}^d) \\ \text{s.t.} \quad & \|\boldsymbol{x}^k\|_2 = 1, \ k = 1, 2, \ldots, d, \end{aligned}$$

where $\mathcal{T}(\boldsymbol{x}^1, \boldsymbol{x}^2, \ldots, \boldsymbol{x}^d) := \langle \mathcal{T}, \boldsymbol{x}^1 \otimes \boldsymbol{x}^2 \otimes \cdots \otimes \boldsymbol{x}^d \rangle$. A polynomial-time approximation algorithm with an approximation bound $\sqrt{\prod_{k=1}^{d-2} n_k}$ (for $d = 3$ it is $\sqrt{n_1}$) was proposed which was essentially constructed from the matrix flattening of a tensor. So [20] improved the approximation bound to $O(\sqrt{\prod_{k=1}^{d-2} \frac{n_k}{\log n_k}})$ (for $d = 3$ it is $O(\sqrt{\frac{n_1}{\log n_1}})$) by relating the problem to computing the L_2 diameter of a certain convex body. The algorithm is complicated and is only of theoretical interest. He et al. [6] proposed a simple randomized algorithm to compute the tensor spectral norm with the same approximation bound to that of So [20]. The approximation bound for the tensor nuclear norm is not addressed until recently by Hu [11], where an approximation bound $\sqrt{\prod_{k=1}^{d-2} n_k}$ (for $d = 3$ it is $\sqrt{n_1}$) of the tensor nuclear norm was proposed based on the matrix flattening of a tensor.

According to our main result, one useful regular partition is to cut a tensor into matrix slices since both the matrix spectral norm and the matrix nuclear norm can be computed in polynomial time. By virtue of Corollary 4.2, the smallest number of matrix slices in a regular partition of a tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ is $\prod_{k=1}^{d-2} n_k$, i.e., the regular partition

$$\left\{ T_{i_1 i_2 \ldots i_{d-2}} := \left((t_{i_1 i_2 \ldots i_d})_{i_{d-1} i_d} \right) \in \mathbb{R}^{n_{d-1} \times n_d} : 1 \le i_k \le n_k, \, k = 1, 2, \ldots, d-2 \right\}.$$

Therefore, Theorem 3.1 and Corollary 4.2 immediately imply the following approximation bounds for the tensor spectral norm and the tensor nuclear norm.

COROLLARY 4.3. *The spectral norm of a tensor* $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ *with* $n_1 \le n_2 \le \cdots \le n_d$ *can be approximated by*

$$\max_{1 \le i_k \le n_k, \, k=1,2,\ldots,d-2} \| T_{i_1 i_2 \ldots i_{d-2}} \|_\sigma$$

with an approximation bound $\sqrt{\prod_{k=1}^{d-2} n_k}$. *The nuclear norm of* $\mathcal{T}$ *can be approximated by*

$$\left(\sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \cdots \sum_{i_{d-2}=1}^{n_{d-2}} \| T_{i_1 i_2 \ldots i_{d-2}} \|_*^2 \right)^{1/2}$$

with an approximation bound $\sqrt{\prod_{k=1}^{d-2} n_k}$. *In particular, for the case* $d = 3$, *both bounds are* $\sqrt{n_1}$.

We remark that the approximation bounds in Corollary 4.3 are established through matrix slices, which are very easy to compute and different from all other methods in the literature [7, 20, 6, 11]. In fact, Corollary 4.3 not only provides easy computable bounds for the tensor spectral norm and the tensor nuclear norm, but also (with the help of Theorem 3.1) can generate a rank-one tensor that approximates $\|\mathcal{T}\|_\sigma$ from below and a rank-one decomposition of $\mathcal{T}$ that approximates $\|\mathcal{T}\|_*$ from above with the approximation bound $\sqrt{\prod_{k=1}^{d-2} n_k}$, according to the following algorithms.

ALGORITHM 4.4. *Find a rank-one tensor that approximates the spectral norm of a given tensor from below.*
- *Input: A tensor* $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ *with* $n_1 \le n_2 \le \cdots \le n_d$.
1. *Compute* $(j_1, j_2, \ldots, j_{d-2}) = \arg\max_{1 \le i_k \le n_k, \, k=1,2,\ldots,d-2} \| T_{i_1 i_2 \ldots i_{d-2}} \|_\sigma$.
2. *Find the left singular vector* $\boldsymbol{x}$ *and the right singular vector* $\boldsymbol{y}$ *corresponding to the largest singular value of the matrix* $T_{j_1 j_2 \ldots j_{d-2}}$.
3. *Compute* $\mathcal{X} = \boldsymbol{e}_{j_1} \otimes \boldsymbol{e}_{j_2} \otimes \cdots \otimes \boldsymbol{e}_{j_{d-2}} \otimes \boldsymbol{x} \otimes \boldsymbol{y}$, *where* $\boldsymbol{e}_j$ *is the vector whose jth entry is one and other entries are zeros.*
- *Output: A rank-one tensor* $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ *with* $\|\mathcal{X}\|_2 = 1$.

PROPOSITION 4.5. *The rank-one tensor* $\mathcal{X}$ *output from Algorithm 4.4 satisfies*

$$\langle \mathcal{T}, \mathcal{X} \rangle \le \|\mathcal{T}\|_\sigma \le \sqrt{\prod_{k=1}^{d-2} n_k} \langle \mathcal{T}, \mathcal{X} \rangle.$$

Proof. If we look into matrix slices of $\mathcal{X}$ by fixing all but the last two indices

$$\left\{ X_{i_1 i_2 \ldots i_{d-2}} := \left((x_{i_1 i_2 \ldots i_d})_{i_{d-1} i_d} \right) \in \mathbb{R}^{n_{d-1} \times n_d} : 1 \le i_k \le n_k, \, k = 1, 2, \ldots, d-2 \right\},$$

we find that the only nonzero matrix is $X_{j_1 j_2 \ldots j_{d-2}} = \boldsymbol{x} \otimes \boldsymbol{y}$. Therefore

$$
\begin{aligned}
\langle \mathcal{T}, \mathcal{X} \rangle &= \langle T_{j_1 j_2 \ldots j_{d-2}}, X_{j_1 j_2 \ldots j_{d-2}} \rangle \\
&= \langle T_{j_1 j_2 \ldots j_{d-2}}, \boldsymbol{x} \otimes \boldsymbol{y} \rangle \\
&= \| T_{j_1 j_2 \ldots j_{d-2}} \|_\sigma \\
&= \max_{1 \le i_k \le n_k, \, k=1,2,\ldots,d-2} \| T_{i_1 i_2 \ldots i_{d-2}} \|_\sigma.
\end{aligned}
$$

The claim follows immediately from Corollary 4.3.   □

ALGORITHM 4.6. *Find a rank-one decomposition of a given tensor that approximates its nuclear norm from above.*
- *Input: A tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ with $n_1 \le n_2 \le \cdots \le n_d$.*
1. *Compute SVD for the matrix*

$$
T_{i_1 i_2 \ldots i_{d-2}} = \sum_{i_{d-1}=1}^{n_{d-1}} \lambda_{i_1 i_2 \ldots i_{d-1}} \boldsymbol{x}_{i_1 i_2 \ldots i_{d-1}} \otimes \boldsymbol{y}_{i_1 i_2 \ldots i_{d-1}}
$$

   for all $1 \le i_k \le n_k$, $k = 1, 2, \ldots, d-2$. If the rank of any matrix $T_{i_1 i_2 \ldots i_{d-2}}$ is strictly less than n_{d-1}, add some zero singular values.
2. *Let*

$$
\mathcal{T} = \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \cdots \sum_{i_{d-1}=1}^{n_{d-1}} \lambda_{i_1 i_2 \ldots i_{d-1}} \boldsymbol{e}_{i_1} \otimes \boldsymbol{e}_{i_2} \otimes \cdots \otimes \boldsymbol{e}_{i_{d-2}} \otimes \boldsymbol{x}_{i_1 i_2 \ldots i_{d-1}} \otimes \boldsymbol{y}_{i_1 i_2 \ldots i_{d-1}},
$$

   where $\boldsymbol{e}_j$ is the vector whose jth entry is one and the other entries are zeros.
- *Output: A rank-one decomposition of $\mathcal{T}$.*

PROPOSITION 4.7. *The rank-one decomposition of $\mathcal{T}$ output from Algorithm 4.6 satisfies*

$$
\sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \cdots \sum_{i_{d-1}=1}^{n_{d-1}} \lambda_{i_1 i_2 \ldots i_{d-1}} \Bigg/ \sqrt{\prod_{k=1}^{d-2} n_k} \le \| \mathcal{T} \|_* \le \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \cdots \sum_{i_{d-1}=1}^{n_{d-1}} \lambda_{i_1 i_2 \ldots i_{d-1}}.
$$

Proof. First, for any given $(i_1, i_2, \ldots, i_{d-2})$ with $1 \le i_k \le n_k$ for $k = 1, 2, \ldots, d-2$, we have

$$
\sum_{i_{d-1}=1}^{n_{d-1}} \lambda_{i_1 i_2 \ldots i_{d-1}} = \| T_{i_1 i_2 \ldots i_{d-2}} \|_*.
$$

By applying Theorem 3.1 with the regular partition of matrix slices

$$
\left\{ T_{i_1 i_2 \ldots i_{d-2}} = \left((t_{i_1 i_2 \ldots i_d})_{i_{d-1} i_d} \right) \in \mathbb{R}^{n_{d-1} \times n_d} : 1 \le i_k \le n_k, \, k = 1, 2, \ldots, d-2 \right\},
$$

it follows that

$$
\| \mathcal{T} \|_* \le \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \cdots \sum_{i_{d-2}=1}^{n_{d-2}} \| T_{i_1 i_2 \ldots i_{d-2}} \|_* = \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \cdots \sum_{i_{d-1}=1}^{n_{d-1}} \lambda_{i_1 i_2 \ldots i_{d-1}}
$$

and

$$\|\mathcal{T}\|_* \geq \left(\sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \cdots \sum_{i_{d-2}=1}^{n_{d-2}} \|T_{i_1 i_2 \ldots i_{d-2}}\|_*^{\,2} \right)^{1/2}$$

$$\geq \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \cdots \sum_{i_{d-2}=1}^{n_{d-2}} \|T_{i_1 i_2 \ldots i_{d-2}}\|_* \Bigg/ \sqrt{\prod_{k=1}^{d-2} n_k}$$

$$= \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \cdots \sum_{i_{d-1}=1}^{n_{d-1}} \lambda_{i_1 i_2 \ldots i_{d-1}} \Bigg/ \sqrt{\prod_{k=1}^{d-2} n_k} \; . \qquad \square$$

5. Concluding remarks. We study the spectral norm and the nuclear norm of a tensor from a new perspective. We introduce a new concept of tensor partition, named as a regular partition, which generalizes the concept of block tensor in the literature. For any regular partition of a given tensor into subtensors, the spectral norm (respectively, the nuclear norm) of the tensor can be lower and upper bounded by manipulating the spectral norms (respectively, the nuclear norms) of its subtensors. Some tensor norm and matrix norm inequalities can be cast into the framework of our new inequalities. As an application, when a tensor is partitioned into its matrix slices, these results lead to polynomial-time worst-case approximation bounds for computing the spectral norm and the nuclear norm of the tensor. All the bounds are sharp in general.

We believe these inequalities will have great potential in various applications. The flexibility on any regular partition of a tensor provides a useful tool for dealing with tensors that have specific structures in practice. For example, if a tensor is too large to estimate its spectral norm or nuclear norm directly, one may regularly partition it into small pieces and apply the bounds in the paper to estimate. It is worth mentioning that by restricting the order of a tensor to two, the matrix case, the relationship between the matrix nuclear norm and its submatrices' is not seen elsewhere, to the best of our knowledge.

One interesting question is whether our main bounds (1) and (2) hold for any tensor partition. This is posted as a conjecture (Conjecture 3.5) in the paper. Two bounds of them are actually true, and for the other two, one of them implies the other. Therefore, essentially only one bound in Conjecture 3.5 needs to be settled.

When computing the main bounds from the norms of subtensors in the paper, we do not consider positions of these subtensors in the original tensor, i.e., these norms of subtensors are treated in the same way as components of a vector. If these norms of subtensors form a smaller tensor by taking their positions in the original tensor into account, in particular in a modal partition, is there any further connection between the norm of the smaller tensor to the norm of the original tensor? This would be an interesting question for further study.

REFERENCES

[1] L. DE LATHAUWER, B. DE MOOR, AND J. VANDEWALLE, *A multilinear singular value decomposition*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1253–1278.

[2] H. DERKSEN, *On the nuclear norm and the singular value decomposition of tensors*, Found. Comput. Math., 16 (2016), pp. 779–811.

[3] S. FRIEDLAND AND L.-H. LIM, *Nuclear Norm of Higher-Order Tensors*, preprint, arXiv:1410.6072, 2016.

[4] S. Gandy, B. Recht, and I. Yamada, *Tensor completion and low-n-rank tensor recovery via convex optimization*, Inverse Problems, 27 (2011), 025010.

[5] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, MD, 1996.

[6] S. He, B. Jiang, Z. Li, and S. Zhang, *Probability bounds for polynomial functions in random variables*, Math. Oper. Res., 39 (2014), pp. 889–907.

[7] S. He, Z. Li, and S. Zhang, *Approximation algorithms for homogeneous polynomial optimization with quadratic constraints*, Math. Program., 125 (2010), pp. 353–383.

[8] N. J. Higham, *Estimating the matrix p-norm*, Numer. Math., 62 (1992), pp. 539–555.

[9] C. J. Hillar and L.-H. Lim, *Most tensor problems are NP-hard*, J. ACM, 60 (2013), 45.

[10] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, New York, 1985.

[11] S. Hu, *Relations of the nuclear norm of a tensor and its matrix flattenings*, Linear Algebra Appl., 478 (2015), pp. 188–199.

[12] T. G. Kolda and B. W. Bader, *Tensor decompositions and applications*, SIAM Rev., 51 (2009), pp. 455–500.

[13] Z. Li, S. He, and S. Zhang, *Approximation Methods for Polynomial Optimization: Models, Algorithms, and Applications*, Springer, New York, 2012.

[14] L.-H. Lim, *Singular values and eigenvalues of tensors: A variational approach*, in Proceedings of the IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, IEEE, New York, 2005, pp. 129–132.

[15] L.-H. Lim, *Tensors and hypermatrices*, in Handbook of Linear Algebra, Chapter 15, 2nd ed., L. Hogben, ed., CRC Press, Boca Raton, FL, 2013.

[16] L.-H. Lim and P. Comon, *Blind multilinear identification*, IEEE Trans. Inform. Theory, 60 (2014), pp. 1260–1280.

[17] J. Liu, P. Musialski, P. Wonka, and J. Ye, *Tensor completion for estimating missing values in visual data*, IEEE Trans. Pattern Anal. Mach. Intell., 35 (2013), pp. 208–220.

[18] L. Qi, *Eigenvalues of a real supersymmetric tensor*, J. Symbolic Comput., 40 (2005), pp. 1302–1324.

[19] S. Ragnarsson and C. F. Van Loan, *Block tensors and symmetric embeddings*, Linear Algebra Appl., 438 (2013), pp. 853–874.

[20] A. M.-C. So, *Deterministic approximation algorithms for sphere constrained homogeneous polynomial optimization problems*, Math. Program., 192 (2011), pp. 357–382.