OPERATOR NORM INEQUALITIES BETWEEN TENSOR UNFOLDINGS ON
THE PARTITION LATTICE

MIAOYAN WANG\textsuperscript{1}, KHANH DAO DUC\textsuperscript{1}, JONATHAN FISCHER\textsuperscript{2}, AND YUN S. SONG\textsuperscript{1,2,3}

\textsuperscript{1}Department of Mathematics, University of Pennsylvania
\textsuperscript{2}Department of Statistics, University of California, Berkeley
\textsuperscript{3}Computer Science Division, University of California, Berkeley

Abstract. Interest in higher-order tensors has recently surged in data-intensive fields, with a wide
range of applications including image processing, blind source separation, community detection,
and feature extraction. A common paradigm in tensor-related algorithms advocates unfolding
(or flattening) the tensor into a matrix and applying classical methods developed for matrices.
Despite the popularity of such techniques, how the functional properties of a tensor changes upon
unfolding is currently not well understood. In contrast to the body of existing work which has
focused almost exclusively on matricizations, we here consider all possible unfoldings of an order-\textit{k}
tensor, which are in one-to-one correspondence with the set of partitions of \{1, \ldots, \textit{k}\}. We derive
general inequalities between the \(l^p\)-norms of arbitrary unfoldings defined on the partition lattice.
In particular, we demonstrate how the spectral norm \((p = 2)\) of a tensor is bounded by that of its
unfoldings, and obtain an improved upper bound on the ratio of the Frobenius norm to the spectral
norm of an arbitrary tensor. For specially-structured tensors satisfying a generalized definition
of orthogonal decomposability, we prove that the spectral norm remains invariant under specific
subsets of unfolding operations.

1. Introduction

Tensors of order 3 or greater, known as higher-order tensors, have recently attracted increased
attention in many fields across science and engineering. Methods built on tensors provide powerful
tools to capture complex structures in data that lower-order methods may fail to exploit. Among
numerous examples, tensors have been used to detect patterns in time-course data \[7, 17, 22, 29\]
and to model higher-order cumulants \[1, 2, 14\]. However, tensor-based methods are fraught with
challenges. Tensors are not simply matrices with more indices; rather, they are mathematical
objects possessing multilinear algebraic properties. Indeed, extending familiar matrix concepts
such as norms to tensors is non-trivial \[12, 18\], and computing these quantities has proven to be
NP-hard \[4–6\].

The spectral relations between a general tensor and its lower-order counterparts have yet to
be studied. There are generally two types of approaches underlying many existing tensor-based
methods. The first approach flattens the tensor into a matrix and applies matrix-based techniques
in downstream analyses, notably higher-order SVD \[3, 10\] and TensorFace \[25\]. Flattening is computationally convenient because of the ubiquity of well-established matrix-based methods, as well
as the connection between tensor contraction and block matrix multiplication \[19\]. However, matricization leads to a potential loss of the structure found in the original tensor. This motivates the
key question of how much information a flattening retains from its parent tensor.

The second approach either handles the tensor directly or unfolds it into objects of order-3 or
higher. Recent work on bounding the spectral norm of sub-Gaussian tensors reveals that solving for

2010 Mathematics Subject Classification. 15A60; 15A69; 06B99; 05A18.
Key words and phrases. higher-order tensors; general unfoldings, partition lattice, operator norm, orthogonality.
the convex relaxation of tensor rank by unfolding is suboptimal [24]. Interestingly, in the context of tensor completion, unfolding a higher-order tensor into a nearly cubic tensor requires smaller sample sizes than matricization [28]. These results are probabilistic in nature and merely focus on a particular class of tensors. Assessing the general impact of unfolding operations on an arbitrary tensor and the role of the tensor’s intrinsic structure remains challenging.

The primary goal of this paper is to study the effect of unfolding operations on functional properties of tensors, where an unfolding is any lower-order representation of a tensor. We study the operator norm of a tensor viewed as a multilinear functional because this quantity is commonly used in both theory and applications, especially in tensor completion [15,28] and low-rank approximation problems [23,27]. Given an order-\(k\) tensor, we represent each possible unfolding operation using a partition \(\pi\) of \([k] = \{1, \ldots, k\}\), where a block in \(\pi\) corresponds to the set of modes that should be combined into a single mode. Each unfolding is a rearrangement of the elements of the original tensor into a tensor of lower order. Here we study the \(L^p\) operator norms of all possible tensor unfoldings, which together define what we coin a “norm landscape” on the partition lattice. A partial order relation between partitions enables us to find a path between an arbitrary pair of unfoldings, which together define what we coin a “norm landscape” on the partition lattice. A partial order relation between partitions enables us to find a path between an arbitrary pair of unfoldings, which together define what we coin a “norm landscape” on the partition lattice. A partial order relation between partitions enables us to find a path between an arbitrary pair of unfoldings, which together define what we coin a “norm landscape” on the partition lattice.

The remainder of this paper is organized as follows. In Section 2, we introduce some notation, and relate the spectral norm and the general \(L^p\)-norm landscape over all possible tensor unfoldings. We conclude in Section 6 with discussions about our findings and avenues of future work.

2. Higher-order tensors and their operator norms

An order-\(k\) tensor \(\mathcal{A} = [a_{i_1 \ldots i_k}] \in \mathbb{F}^{d_1 \times \cdots \times d_k}\) over a field \(\mathbb{F}\) is a hypermatrix with dimensions \((d_1, \ldots, d_k)\) and entries \(a_{i_1 \ldots i_k} \in \mathbb{F}\), for \(1 \leq i_n \leq d_n\), \(n = 1, \ldots, k\). In this paper, we focus on real tensors, \(\mathbb{F} = \mathbb{R}\). The total dimension of \(\mathcal{A}\) is denoted by \(\text{dim}(\mathcal{A}) = \prod_{n=1}^k d_n\).

The vectorization of \(\mathcal{A}\), denoted \(\text{Vec}(\mathcal{A})\), is defined as the operation rearranging all elements of \(\mathcal{A}\) into a column vector. For ease of notation, we use the shorthand \([n]\) to denote the \(n\)-set \(\{1, \ldots, n\}\) for \(n \in \mathbb{N}_+\), and sometimes write \(\otimes_{n=1}^k x_n = x_1 \otimes \cdots \otimes x_k\) when space is limited. We use \(S^{d-1} = \{x \in \mathbb{R}^d : \|x\|_2 = 1\}\) to denote the \((d-1)\)-dimensional unit sphere, and \(I_d\) to denote the \(d \times d\) identity matrix.

For any two tensors \(\mathcal{A} = [a_{i_1 \ldots i_k}], \mathcal{B} = [b_{i_1 \ldots i_k}] \in \mathbb{R}^{d_1 \times \cdots \times d_k}\) of identical order and dimensions, their inner product is defined as

\[
\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{i_1, \ldots, i_k} a_{i_1 \ldots i_k} b_{i_1 \ldots i_k},
\]

while the tensor Frobenius norm of \(\mathcal{A}\) is defined as

\[
\|\mathcal{A}\|_F = \sqrt{\langle \mathcal{A}, \mathcal{A} \rangle} = \sqrt{\sum_{i_1, \ldots, i_k} |a_{i_1 \ldots i_k}|^2},
\]

both of which are analogues of standard definitions for vectors and matrices.
Following [12], we define the covariant multilinear matrix multiplication of a tensor $A \in \mathbb{R}^{d_1 \times \cdots \times d_k}$ by matrices $M_1 = (m_{i_1j_1}^{(1)}) \in \mathbb{R}^{d_1 \times s_1}, \ldots, M_k = (m_{ijk}^{(k)}) \in \mathbb{R}^{d_k \times s_k}$ as

$$A(M_1, \ldots, M_k) = \left[ \sum_{i_1=1}^{d_1} \cdots \sum_{i_k=1}^{d_k} a_{i_1 \ldots i_k} m_{i_1j_1}^{(1)} \cdots m_{ijk}^{(k)} \right],$$

which results in an order-$k$ tensor in $\mathbb{R}^{s_1 \times \cdots \times s_k}$. This operation multiplies the $n^{th}$ mode of $A$ by the matrix $M_n$ for all $n \in [k]$. Just as a matrix may be multiplied in up to two modes by matrices of consistent dimensions, an order-$k$ tensor can be multiplied by up to $k$ matrices in $k$ modes. In the case of $k = 2$, $A$ is a matrix and $A(M_1, M_2) = M_1^T A M_2$. Sometimes we are interested in multiplying by vectors rather than matrices, in which case we obtain the $k$-multilinear functional $A : \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_k} \rightarrow \mathbb{R}$ given by

$$A(x_1, \ldots, x_k) = \sum_{i_1=1}^{d_1} \cdots \sum_{i_k=1}^{d_k} a_{i_1 \ldots i_k} x_1^{(1)} \cdots x_k^{(k)} = \langle A, x_1 \otimes \cdots \otimes x_k \rangle,$$

where $x_n = (x_1^{(n)}, \ldots, x_d^{(n)})^T \in \mathbb{R}^{d_n}$, $n \in [k]$. Note that multiplying by a vector in $r$ modes in the manner defined in (1) reduces the order of the output tensor to $k - r$, whereas multiplying by matrices leaves the order unchanged. Although the coordinate representation of a tensor as a hypermatrix provides a concrete description, viewing it instead as a multilinear functional provides a coordinate-free, basis-independent perspective which allows us to better characterize the spectral relations among different tensor unfoldings.

We define the operator norm, or induced norm, of a tensor $A$ using the associated $k$-multilinear functional (1).

**Definition 2.1** (Lim [12]). Let $A \in \mathbb{R}^{d_1 \times \cdots \times d_k}$ be an order-$k$ tensor. For any $1 \leq p \leq \infty$, the $l^p$-norm of the multilinear functional associated with $A$ is defined as

$$\|A\|_p = \sup \left\{ \frac{A(x_1, \ldots, x_k)}{\|x_1\|_p \cdots \|x_k\|_p} : x_n \neq 0, x_n \in \mathbb{R}^{d_n}, n \in [k] \right\},$$

(2)

where $\|x_n\|_p$ denotes the vector $l^p$-norm of $x_n$.

**Remark 2.2.** The special case of $p = 2$ is called the spectral norm, frequently denoted $\|A\|_\sigma$. By (2), $\|A\|_\sigma$ is the maximum value obtained as the inner product of the tensor $A$ with a rank-1 tensor, $x_1 \otimes \cdots \otimes x_n$, of Frobenius norm 1 and of the same dimensions. This point of view provides an equivalent definition of $\|A\|_\sigma$ as determining the best rank-1 tensor approximation to $A$, and we note that the rank-1 constraint becomes weaker as more unfolding is applied. See Section 4 for further details.

Because we restrict all entries of $A$ and $\{x_n\}_{n=1}^k$ to be real, we need not take the absolute value of $A(x_1, \ldots, x_k)$ as in [12]. It is worth mentioning that the notion of tensor $l^p$-norms defined by (2) are not extensions of the classical matrix $l^p$-norms when $p \neq 2$. To see this, recall that for an $m \times n$ matrix $A$, one usually defines the $l^p$ operator norm as

$$\|A\|_p = \sup \left\{ \frac{\|Ax\|_p}{\|x\|_p} : x \neq 0, x \in \mathbb{R}^n \right\},$$

(3)

In general (2) and (3) are not equal even for matrices, illustrated by the following example:
Example 2.3. Let \( A = [a_{ij}] \) be the 2 \( \times \) 2 matrix \( A = \begin{bmatrix} 1 & 1 \\ 0 & 4 \end{bmatrix} \) and consider \( p = 1 \). Solving (3), we have
\[
\|A\|_1 = \sup \left\{ \frac{\|Ax\|_1}{\|x\|_1} : x \neq 0, x \in \mathbb{R}^2 \right\} = \sup_j \sum_i |a_{ij}| = 5.
\]
However, instead using (2) gives
\[
\|A\|_1 = \sup \left\{ \frac{x^T A x_2}{\|x_1\|_1 \|x_2\|_1} : x_n \in \mathbb{R}^2, x_n \neq 0, n \in [2] \right\}
\]
\[
= \sup \left\{ x^T A x_2 : x_n \in \mathbb{R}^2, \|x_n\|_1 = 1, n \in [2] \right\}
\]
\[
= 4,
\]
which is neither the classical matrix \( l^1 \)-norm (equal to 5), nor the entry-wise \( l^1 \)-norm (equal to \( \sum_{i,j} |a_{ij}| = 6 \)).

Throughout this paper, we adopt Definition 2.1 and always use \( \| \cdot \|_p \) to denote the \( l^p \)-norm defined therein, even for matrices. In fact, (3) defines an operator norm by viewing the matrix as a linear operator from \( \mathbb{R}^{d_2} \) to \( \mathbb{R}^{d_1} \), whereas (2) defines an operator norm in which the matrix defines a bilinear functional from \( \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \) to \( \mathbb{R} \). These two definitions are equivalent when \( p = 2 \), but otherwise represent two different operators and result in two distinct operator norms. To be consistent with our treatment of tensors as \( k \)-multilinear functionals, we formulate matrices as bilinear functionals.

For a given tensor, its \( l^p \)-norm and \( l^q \)-norm mutually control each other, and the comparison bound is polynomial in the total dimension of the tensor, \( \text{dim}(A) = \prod_{n=1}^k d_n \).

Proposition 2.4 (\( l^p \)-norm vs. \( l^q \)-norm). Let \( A \in \mathbb{R}^{d_1 \times \cdots \times d_k} \) be an order-\( k \) tensor. Suppose \( \| \cdot \|_p \) and \( \| \cdot \|_q \) are two norms defined in (2) with \( q \geq p \geq 1 \). Then,
\[
\|A\|_p \leq \|A\|_q \leq \text{dim}(A)^{\frac{1}{p} - \frac{1}{q}} \|A\|_p.
\]

Proof. Starting from Definition 2.1, we have
\[
\|A\|_q = \sup \left\{ \frac{\mathcal{A}(x_1, \ldots, x_k)}{\|x_1\|_q \cdots \|x_k\|_q} : x_n \neq 0, x_n \in \mathbb{R}^{d_n}, n \in [k] \right\}
\]
\[
= \sup \left\{ \mathcal{A}(x_1, \ldots, x_k) \times \frac{\|x_1\|_p \cdots \|x_k\|_p}{\|x_1\|_q \cdots \|x_k\|_q} : x_n \neq 0, x_n \in \mathbb{R}^{d_n}, n \in [k] \right\}.
\]

For any \( q \geq p \geq 1 \), the equivalence of vector norms tells us
\[
\|x\|_q \leq \|x\|_p \leq d^{\frac{1}{q} - \frac{1}{p}} \|x\|_q, \text{ for all } x \in \mathbb{R}^d.
\]
Applying (5) to \( x_n \), for \( n \in [k] \), gives
\[
1 \leq \prod_{n=1}^k \frac{\|x_n\|_p}{\|x_n\|_q} \leq \left( \prod_{n=1}^k d_n \right)^{\frac{1}{p} - \frac{1}{q}}.
\]
Inserting (6) into (4) and noting \( \text{dim}(A) = \prod_{n=1}^k d_n \), we find
\[
\|A\|_p \leq \|A\|_q \leq \text{dim}(A)^{\frac{1}{p} - \frac{1}{q}} \|A\|_p,
\]
which completes the proof. \( \Box \)
3. Partitions and general tensor unfoldings

Any higher-order tensor can be transformed into different lower-order tensors by modifying its indices in various ways. The most common transformations are \(n\)-mode flattenings, or matricizations, which rearrange the elements of an order-\(k\) tensor into a \(d_1 \times \cdots \times d_k\) matrix. For example, the \(n\)-mode matricization of a tensor \(A \in \mathbb{R}^{d_1 \times \cdots \times d_k}\) is obtained by mapping the fixed tensor index \((i_1, \ldots, i_k)\) to the matrix index \((i_n, m)\), where

\[
m = 1 + \prod_{a \in [k], a \neq n} (i_a - 1) J_a, \quad \text{where } J_a = \prod_{t \in [a-1], t \neq n} d_t.
\]

Recently there has been much interest in studying the relationship between tensors and their matrix flattenings [9]. We present here a more general analysis by considering all possible lower-order tensor unfoldings rather than just matricizations. Using the blocks of a partition of \([k]\) to specify which modes are combined into a single mode of the new tensor, we establish a one-to-one correspondence between the set of all partitions of \([k]\) and the set of lower-order tensor unfoldings. The partition lattice then describes the underlying relationship between possible tensor unfoldings of a tensor \(A\).

For any \(k \in \mathbb{N}_+\), a partition \(\pi\) of \([k]\) is a collection \(\{B_1^\pi, B_2^\pi, \ldots, B_\ell^\pi\}\) of disjoint, nonempty subsets (or blocks) \(B_i^\pi\) satisfying \(\bigcup_{i=1}^\ell B_i^\pi = [k]\). The set of all partitions of \([k]\) is denoted \(P_{[k]}\). We use \(|\pi|\) to denote the number of blocks in \(\pi\), and \(|B_i^\pi|\) to denote the number of elements in \(B_i^\pi\). We say that a partition \(\pi\) is a level-\(\ell\) partition if \(|\pi| = \ell\). The set of all level-\(\ell\) partitions of \([k]\) is denoted by \(P_{[k]}^{\leq \ell}\), which is a set of \(S(k, \ell)\) elements, where \(S(k, \ell)\) is the Stirling number of the second kind.

The following partial order naturally relates partitions satisfying a basic compatibility constraint and the resulting structure plays a key role in our work.

**Definition 3.1** (Partition Lattice). A partition \(\pi_1 \in P_{[k]}\) is called a refinement of \(\pi_2 \in P_{[k]}\) if each block of \(\pi_1\) is a subset of some block of \(\pi_2\); conversely, \(\pi_2\) is said to be a coarsening of \(\pi_1\). This relationship defines a partial order, expressed as \(\pi_1 \leq \pi_2\), and we say that \(\pi_1\) is finer than \(\pi_2\) while \(\pi_2\) is coarser than \(\pi_1\). If either \(\pi_1 \leq \pi_2\) or \(\pi_2 \leq \pi_1\), then \(\pi_1\) and \(\pi_2\) are comparable. According to this partial order, the least element of \(P_{[k]}\) is \(0_{[k]} := \{\{1\}, \ldots, \{k\}\}\), while the greatest element is \(1_{[k]} := \{\{1, \ldots, k\}\}\). Equipped with this notion, \(P_{[k]}\) generates a partition lattice by connecting any two comparable partitions that differ by exactly one level. An example is illustrated in Figure 1. Henceforth, \(P_{[k]}\) may represent either the set of all partitions of \([k]\) or the partition lattice it generates depending on context.

It is clear that many partitions are not comparable, including every pair of distinct partitions at the same level. To consider arbitrary partitions in tandem, we require an extension of this partial order. In general, for any two partitions \(\pi_1, \pi_2 \in P_{[k]}\), we define their greatest lower bound \(\pi_1 \land \pi_2\) as

\[
\pi_1 \land \pi_2 := \sup\{\pi \in P_{[k]} : \pi \leq \pi_1, \pi \leq \pi_2\}.
\]

More concretely, \(\pi_1 \land \pi_2\) consists of the collection of all nonempty intersections of blocks in \(\pi_1\) and \(\pi_2\) and is unique for a given pair \((\pi_1, \pi_2)\).

**Example 3.2.** Figure 1 illustrates \(P_{[4]}\) in lattice form. Recall that an edge connects two partitions if and only if they are comparable and their levels differ by exactly one. Partitions are comparable if and only if there exists a non-reversing path between them. To clarify the components of Definition 3.1, take \(\pi_1 = \{\{1, 4\}, \{2, 3\}\}\) and \(\pi_2 = \{\{1, 3, 4\}, \{2\}\}\). Then \(\pi_1 \land \pi_2 = \{\{1, 4\}, \{2\}, \{3\}\}\), \(0_{[4]} = \{\{1\}, \{2\}, \{3\}, \{4\}\}\), and \(1_{[4]} = \{\{1, 2, 3, 4\}\}\).

The following definition generalizes the concept of \(n\)-mode flattenings defined in (7) to general unfoldings induced by arbitrary partitions \(\pi \in P_{[k]}\).
Example 3.5. Consider an order-4 tensor $A = [a_{ijkl}] \in \mathbb{R}^{2 \times 2 \times 2 \times 2}$. We provide a subset of the possible tensor unfoldings to elucidate both the manner in which the operation works and the natural association with partitions.

- For $\pi = \{\{1, 2\}, \{3\}, \{4\}\}$, $\text{Unfold}_\pi(A)$ is an order-3 tensor of dimensions $(4, 2, 2)$ with entries given by
  \[(\text{Unfold}_\pi(A))_{1kl} = a_{11kl}, \quad (\text{Unfold}_\pi(A))_{2kl} = a_{12kl},\]
  
  \[(\text{Unfold}_\pi(A))_{3kl} = a_{21kl}, \quad (\text{Unfold}_\pi(A))_{4kl} = a_{22kl},\]
  
  for all $(k, l) \in [2] \times [2]$. 

---

**Definition 3.3** (General Tensor Unfolding). Let $A \in \mathbb{R}^{d_1 \times \cdots \times d_k}$ be an order-$k$ tensor and $\pi = \{B_1^\pi, \ldots, B_k^\pi\} \in \mathcal{P}_k$. The partition $\pi$ defines a mapping $\phi_\pi: [d_1] \times \cdots \times [d_k] \to [\prod_{j \in B_1^\pi} d_j] \times \cdots \times [\prod_{j \in B_k^\pi} d_j]$ such that

\[
\phi_\pi(i_1, \ldots, i_k) = (m_1, \ldots, m_\ell),
\]

where

\[
m_j = 1 + \prod_{r \in B_j^\pi} (i_r - 1)J_r, \quad \text{where} \quad J_r = \prod_{l \in B_j^\pi \setminus l < r} d_l, \quad \text{for all} \quad j \in [\ell].
\]

Clearly, $\phi_\pi$ is a one-to-one mapping, so its inverse $\phi_\pi^{-1}$ is well defined. Thus $\phi_\pi$ induces an unfolding action $A \mapsto \text{Unfold}_\pi(A)$ such that

\[
(\text{Unfold}_\pi(A))_{(m_1, \ldots, m_\ell)} = (A)_{\phi_\pi^{-1}(m_1, \ldots, m_\ell)},
\]

for all $(m_1, \ldots, m_\ell) \in [\prod_{j \in B_1^\pi} d_j] \times \cdots \times [\prod_{j \in B_k^\pi} d_j]$, or, equivalently,

\[
(\text{Unfold}_\pi(A))_{(i_1, \ldots, i_k)} = (A)_{(i_1, \ldots, i_k)},
\]

for all $(i_1, \ldots, i_k) \in [d_1] \times \cdots \times [d_k]$. Thus $\text{Unfold}_\pi(A)$ is an order-$\ell$ tensor of dimensions $(\prod_{i \in B_1^\pi} d_i, \ldots, \prod_{i \in B_k^\pi} d_i)$, and we call it the tensor unfolding of $A$ induced by $\pi$.

**Remark 3.4.** By the definition of $\phi_\pi$ in (8), $\text{Unfold}_{0[k]}(A) = A$ and $\text{Unfold}_{1[k]}(A) = \text{Vec}(A)$.

---

![Partition Lattice](image.png)

**Figure 1.** The partition lattice $\mathcal{P}_4$. 

---

A⇒→
• For \( \pi = \{1, 2\} \), \( \{3, 4\} \)}, \( \text{Unfold}_\pi(A) \) is a 4 \times 4 matrix

\[
\text{Unfold}_\pi(A) = \begin{pmatrix}
a_{1111} & a_{1112} & a_{1121} & a_{1122} \\
a_{1211} & a_{1212} & a_{1221} & a_{1222} \\
a_{2111} & a_{2112} & a_{2121} & a_{2122} \\
a_{2211} & a_{2212} & a_{2221} & a_{2222}
\end{pmatrix}.
\]

• For \( \pi = \{1, 2, 3\} \), \( \{4\} \)}, \( \text{Unfold}_\pi(A) \) is a 8 \times 2 matrix

\[
\text{Unfold}_\pi(A) = \begin{pmatrix}
a_{1111} & a_{1112} \\
a_{1121} & a_{1122} \\
a_{1211} & a_{1212} \\
a_{1221} & a_{1222} \\
a_{2111} & a_{2112} \\
a_{2121} & a_{2122} \\
a_{2211} & a_{2212} \\
a_{2221} & a_{2222}
\end{pmatrix}.
\]

**Remark 3.6.** There are different conventions to order the elements within each transformed mode. In principle, the ordering of elements within each transformed mode is irrelevant, so we do not explicitly spell out their orderings hereafter.

**Remark 3.7.** The unfolding operation leaves the Frobenius norm unchanged; that is, \( \|A\|_F = \|\text{Unfold}_\pi(A)\|_F \) for all \( \pi \in \mathcal{P}_k \). More generally, the inner product remains invariant under all unfoldings: \((A, B) = \langle \text{Unfold}_\pi(A), \text{Unfold}_\pi(B) \rangle \) for all \( \pi \in \mathcal{P}_k \), where \( A, B \in \mathbb{R}^{d_1 \times \cdots \times d_k} \) are two order-\( k \) tensors of the same dimensions.

### 4. Operator norm inequalities on the partition lattice

In this section, we compare the operator norms of different unfoldings of a tensor, in particular relative to that of the original tensor. We first focus on the spectral norm (\( p = 2 \)) and then discuss extensions to general \( p \)-norms.

Recall that for an order-\( k \) tensor \( A \in \mathbb{R}^{d_1 \times \cdots \times d_k} \), its spectral norm is defined as

\[
\|A\|_\sigma = \sup \left\{ A(x_1, \ldots, x_k) : \|x_n\|_2 = 1, x_n \in \mathbb{R}^{d_n}, n \in [k] \right\}.
\]

The maximization of the polynomial form \( A(x_1, \ldots, x_k) \) on the unit sphere is closely related to the best, in the least-square sense, rank-1 tensor approximation. Specifically, for an order-\( k \) tensor \( A \in \mathbb{R}^{d_1 \times \cdots \times d_k} \), the problem of determining the spectral norm is equivalent to finding a scalar \( \lambda \) and a rank-1 norm-1 tensor \( x_1 \otimes \cdots \otimes x_k \) that minimize the function

\[
f(x_1, \ldots, x_k) = \|A - \lambda x_1 \otimes \cdots \otimes x_k\|_F.
\]

The corresponding value of \( \lambda \) is equal to \( \|A\|_\sigma \). Since \( x_1 \otimes \cdots \otimes x_k \) must have the same order and dimensions as \( A \), the rank-1 condition becomes less strict the more unfolded the tensor. In particular, for the vectorized unfolding, \( \text{Vec}(A) \), the best rank-1 tensor approximation is simply \( \text{Vec}(A) \) itself. For unfoldings into a matrix, \( \text{Mat}(A) \), the best rank-1 approximation is the outer product of the leading left and the right singular vectors of \( \text{Mat}(A) \). For higher-order unfoldings, the closed form of the best rank-1 approximation is not known in general. Nevertheless, the set of the rank-1 tensors over which the supremum is taken in (9) becomes more restricted. This observation implies that the spectral norm of a tensor unfolding decreases as the order of the unfolded tensor increases; that is, the spectral norm preserves the partial order on partitions.

**Proposition 4.1** (Monotonicity). For all partitions \( \pi_1, \pi_2 \in \mathcal{P}_k \) satisfying \( \pi_1 \leq \pi_2 \),

\[\|\text{Unfold}_{\pi_1}(A)\|_\sigma \leq \|\text{Unfold}_{\pi_2}(A)\|_\sigma.\]
In particular, we have the global extrema

\[ \|A\|_\sigma = \left\| \text{Unfold}_{[k]}(A) \right\|_\sigma = \min_{\pi \in \mathcal{P}[k]} \|\text{Unfold}_\pi(A)\|_\sigma, \]

\[ \|A\|_F = \left\| \text{Unfold}_{[k]}(A) \right\|_\sigma = \max_{\pi \in \mathcal{P}[k]} \|\text{Unfold}_\pi(A)\|_\sigma. \]

Proof. Suppose \( \pi_1, \pi_2 \in \mathcal{P}[k] \), and \( \pi_1 \) is a one-step refinement of \( \pi_2 \). Without loss of generality, assume \( \pi_2 \) is obtained by merging two blocks \( B_1, B_2 \in \pi_1 \) into a single block \( B \in \pi_2 \). Let \( \text{Unfold}_{\pi_1}(A) \in \mathbb{R}^{d_1 \times \cdots \times d_{\ell}} \) denote the tensor unfolding induced by \( \pi_1 = \{B_1, \ldots, B_\ell\} \). Then, \( \text{Unfold}_{\pi_2}(A) \) is a \((d_1 \cdot d_2, d_3, \ldots, d_\ell)\)-dimensional tensor. By definition of the spectral norm,

\[ \|\text{Unfold}_{\pi_1}(A)\|_\sigma = \sup \{ \langle \text{Unfold}_{\pi_1}(A), x_1 \otimes \cdots \otimes x_\ell \rangle : x_n \in S^{d_n-1}, n \in [\ell] \} \]

\[ = \sup \{ \langle \text{Unfold}_{\pi_2}(A), \text{Vec}(x_1 \otimes x_2) \otimes \cdots \otimes x_\ell \rangle : x_n \in S^{d_n-1}, n \in [\ell] \} \]

\[ \leq \sup \{ \langle \text{Unfold}_{\pi_2}(A), y \otimes x_3 \otimes \cdots \otimes x_\ell \rangle : y \in S^{d_1 \cdot d_2 - 1}, x_n \in S^{d_n-1}, n = 3, \ldots, \ell \} \]

\[ = \|\text{Unfold}_{\pi_2}(A)\|_\sigma, \]

where the third line comes from the fact that the set \( \{\text{Vec}(x_1 \otimes x_2) : (x_1, x_2) \in S^{d_1-1} \times S^{d_2-1}\} \) is contained in the set \( \{y : y \in S^{d_1 \cdot d_2-1}\} \).

In general, if \( \pi_1 \leq \pi_2 \), we can obtain \( \text{Unfold}_{\pi_2}(A) \) from \( \text{Unfold}_{\pi_1}(A) \) by a series of single unfoldings. Hence, applying the above arguments to these successive unfoldings gives the desired result. \( \square \)

The following lemma will play a key role in proving our main results:

**Lemma 4.2 (One-Step Inequality).** For \( 1 < \ell \leq k \), let \( B \in \mathbb{R}^{d_1 \times \cdots \times d_{\ell}} \) be an order-\( \ell \) tensor unfolding of an order-\( k \) tensor unfolding of an order-\( \ell \) tensor unfolding of a tensor unfolding induced by merging blocks \( B_i \) and \( B_j \) for some \( i, j \in [\ell] \). Then,

\[ \min (d_i, d_j)^{-1/2} \|C\|_\sigma \leq \|B\|_\sigma \leq \|C\|_\sigma. \]

Proof. The upper bound follows readily from Proposition 4.1. To prove the lower bound, without loss of generality, assume \( C \) corresponds to the merging of blocks \( B_{\ell-1} \) and \( B_{\ell} \) so that \( C \) is a \((d_1, \ldots, d_{\ell-2}, d_{\ell-1} \cdot d_\ell)\)-dimensional tensor. Note that \( S^{d_{\ell-1} \times \cdots \times S^{d_{\ell-2}-1} \times S^{d_{\ell-1} \cdot d_\ell}} \) is a compact set, so the supremum in \( C \) is attained in that set for \( C \). Then, there exists \( (x_1^*, \ldots, x_{\ell-2}^*, y^*) \in S^{d_1-1} \times \cdots \times S^{d_{\ell-2}-1} \times S^{d_{\ell-1} \cdot d_\ell} \) such that

\[ \|C\|_\sigma = \langle C, x_1^* \otimes \cdots \otimes x_{\ell-2}^* \otimes y^* \rangle. \]

Define \( C^* = C(x_1^*, \ldots, x_{\ell-2}^*, I_{d_{\ell-1}}, I_{d_\ell}) \in \mathbb{R}^{d_{\ell-1} \cdot d_\ell} \). By the self duality of the Frobenius norm in \( \mathbb{R}^{d_{\ell-1} \cdot d_\ell} \), we can rewrite \( C^* \) as

\[ \|C\|_\sigma = \sup \{ \langle C^*, y \rangle : y \in S^{d_{\ell-1} \cdot d_\ell} \} = \|C^*\|_F. \]

On the other hand, if we define \( \text{Mat}(C^*) := B(x_1^*, \ldots, x_{\ell-2}^*, I_{d_{\ell-1}}, I_\ell) \in \mathbb{R}^{d_{\ell-1} \times d_\ell} \), then \( C^* \) is simply the vectorization of \( \text{Mat}(C^*) \). Hence, by Remark 3.7, we obtain

\[ \|C^*\|_F = \|\text{Mat}(C^*)\|_F. \]

Using the definition of \( \text{Mat}(C^*) \), we can write

\[ \|B\|_\sigma \geq \sup \{ \langle B, x_1^* \otimes x_2^* \otimes \cdots \otimes x_{\ell-2}^* \otimes x_\ell \rangle : (x_{\ell-1}, x_\ell) \in S^{d_{\ell-1} \times S^{d_\ell}} \} \]

\[ = \sup \{ \langle \text{Mat}(C^*), x_{\ell-1} \otimes x_\ell \rangle : (x_{\ell-1}, x_\ell) \in S^{d_{\ell-1} \times S^{d_\ell}} \} \]

\[ = \|\text{Mat}(C^*)\|_\sigma. \]
Recall that \( \text{Mat}(C^*) \) is a matrix of size \( d_{\ell-1} \times d_\ell \), meaning [8]
\[
\| \text{Mat}(C^*) \|_\sigma \geq \min (d_{\ell-1}, d_\ell)^{-1/2} \| \text{Mat}(C^*) \|_F.
\]
(15)

Using (14) in conjunction with (15), (13), and (12), we then obtain
\[
\| B \|_\sigma \geq \min (d_{\ell-1}, d_\ell)^{-1/2} \| \text{Mat}(C^*) \|_F
= \min (d_{\ell-1}, d_\ell)^{-1/2} \| C^* \|_F
= \min (d_{\ell-1}, d_\ell)^{-1/2} \| C \|_\sigma,
\]
which proves the lower bound. \( \square \)

**Remark 4.3.** Both bounds in the one-step inequality (10) are sharp. The sharpness of the upper bound will be discussed in Section 5. For the lower bound, consider an order-\( \ell \) tensor unfolding \( B = \sum_{i=1}^{d_1} e_{1,i} \otimes e_{2,i} \otimes D \in \mathbb{R}^{d_1 \times d_2 \times \cdots \times d_\ell} \), where \( \{ e_{1,i} : i \in [d_1] \} \) is the standard orthonormal basis of \( \mathbb{R}^{d_1} \), \( \{ e_{2,i} : i \in [d_1] \} \) is a set of \( d_1 \) standard basis vectors in \( \mathbb{R}^{d_2} \), and \( D \) is an arbitrary \( (d_3, \ldots, d_\ell) \)-dimensional tensor. Assume \( d_1 \leq d_2 \) and consider the two blocks \( B_1 = \{1\}, B_2 = \{2\} \). By the unfolding operation specified in Lemma 4.2 with \( B_1 \) and \( B_2 \), \( C = \left( \sum_{i=1}^{d_1} \text{Vec}(e_{1,i} \otimes e_{2,i} \otimes D) \right) \otimes D \in \mathbb{R}^{d_1 \times d_2 \times d_3 \times \cdots \times d_\ell} \), which is an order-(\( \ell - 1 \)) tensor unfolding by merging the first two modes of \( B \) into a single mode. It follows that \( \| B \|_\sigma = \| D \|_\sigma \) and \( \| C \|_\sigma = \sqrt{d_1} \| D \|_\sigma \), and therefore the left-hand-side inequality in (10) is saturated.

More generally, we can establish inequalities relating the spectral norms of two tensor unfoldings corresponding to two arbitrary partitions \( \pi_1, \pi_2 \in \mathcal{P}_k \) that are not necessarily comparable. To do so, we must first introduce the following definition.

**Definition 4.4.** Given an order-\( k \) tensor \( A \in \mathbb{R}^{d_1 \times \cdots \times d_k} \), we define the map \( \text{dim}_A : \mathcal{P}_k \times \mathcal{P}_k \to \mathbb{N}_+ \) as
\[
\text{dim}_A(\pi_1, \pi_2) = \prod_{B \in \pi_1} \max_{B' \in \pi_2} D_A(B, B'), \quad \text{where} \quad D_A(B, B') = \prod_{n \in B \cap B'} d_n,
\]
where \( \pi_1, \pi_2 \in \mathcal{P}_k \).

We label this quantity as \( \text{dim}_A(\cdot, \cdot) \) because it involves a product of a subset of the dimensions of \( A \), and \( \text{dim}_A(\pi_1, \pi_2) \leq \text{dim}(A) \) with equality only when \( \pi_1 = \pi_1 \cap \pi_2 \). Intuitively, \( \text{dim}_A(\pi_1, \pi_2) \) reflects the overlap between the unfoldings induced by \( \pi_1 \) and \( \pi_1 \cap \pi_2 \). Example 4.6 presents a concrete illustration.

**Remark 4.5.** We set \( D_A(B, B') = 0 \) when \( B \cap B' = \emptyset \). This does not affect the multiplication in Definition 4.4 because a block of \( \pi_1 \) cannot be disjoint from every block of \( \pi_2 \). Note that \( D_A(B, B') = D_A(B', B) \), but in general \( \text{dim}_A(\pi_1, \pi_2) \neq \text{dim}_A(\pi_2, \pi_1) \).

**Example 4.6.** To illustrate the above map, let \( A \in \mathbb{R}^{d_1 \times d_2 \times d_3 \times d_4} \) and consider the partitions \( \pi_1 = \{\{1, 2\}, \{3, 4\}\} \) and \( \pi_2 = \{\{1, 2, 3\}, \{4\}\} \), for which \( \pi_1 \cap \pi_2 = \{\{1, 2\}, \{3\}\} \). From Definition 4.4,
\[
D_A(\{1, 2\}, \{1, 2, 3\}) = d_1 d_2
D_A(\{3, 4\}, \{1, 2, 3\}) = d_3
D_A(\{1, 2\}, \{4\}) = 0
D_A(\{3, 4\}, \{4\}) = d_4.
\]

Then,
\[
\text{dim}_A(\pi_1, \pi_2) = \max \{ D_A(\{1, 2\}, \{1, 2, 3\}), D_A(\{1, 2\}, \{4\}) \}
\times \max \{ D_A(\{3, 4\}, \{1, 2, 3\}), D_A(\{3, 4\}, \{4\}) \}
= d_1 d_2 \max(d_3, d_4).
Exchanging arguments, we find
\[
\dim_A(\pi_2, \pi_1) = \max \left\{ D_A(\{1, 2, 3\}, \{1, 2\}), D_A(\{1, 2, 3\}, \{3, 4\}) \right\} \\
\times \max \left\{ D_A(\{4\}, \{1, 2\}), D_A(\{4\}, \{3, 4\}) \right\}
\]
\[
= d_4 \max\{d_1d_2, d_3\}.
\]

**Remark 4.7.** As in Lemma 4.2, if \(\pi_2\) is a one-step coarsening of \(\pi_1\) obtained by merging two blocks \(B_i^{\pi_1}\) and \(B_j^{\pi_1}\) of dimensions \(d_i\) and \(d_j\) into a single block \(B_{1}^{\pi_2}\), then
\[
D_A(B_i^{\pi_1}, B_j^{\pi_1}) = d_i,
\]
\[
D_A(B_i^{\pi_1}, B_j^{\pi_1}) = d_j.
\]

Hence, Lemma 4.2 can be written as
\[
\min \left\{ D_A(B_1^{\pi_1}, B_1^{\pi_2}), D_A(B_2^{\pi_1}, B_2^{\pi_2}) \right\}^{-1/2} \|\text{Unfold}_{\pi_2}(A)\|_\sigma \leq \|\text{Unfold}_{\pi_1}(A)\|_\sigma \leq \|\text{Unfold}_{\pi_2}(A)\|_\sigma.
\]

Having introduced Definition 4.4, we can now state our main result on how the spectral norms of two arbitrary unfoldings of a tensor are related:

**Theorem 4.8** (Spectral norm inequalities). Let \(A \in \mathbb{R}^{d_1 \times \cdots \times d_k}\) be an arbitrary order-\(k\) tensor, and \(\pi_1, \pi_2\) any two partitions in \(\mathcal{P}_{[k]}\). Then,
\[
\left[ \frac{\dim(A)}{\dim_A(\pi_2, \pi_1)} \right]^{-1/2} \|\text{Unfold}_{\pi_1}(A)\|_\sigma \leq \|\text{Unfold}_{\pi_2}(A)\|_\sigma \leq \left[ \frac{\dim(A)}{\dim_A(\pi_2, \pi_1)} \right]^{1/2} \|\text{Unfold}_{\pi_1}(A)\|_\sigma.
\]

**Remark 4.9.**

(a) Note that \(\pi_1\) and \(\pi_2\) need not be comparable.

(b) If \(d_n = d\) for all \(n \in [k]\), then the result reduces to
\[
d^{-c_1/2} \|\text{Unfold}_{\pi_1}(A)\|_\sigma \leq \|\text{Unfold}_{\pi_2}(A)\|_\sigma \leq d^{c_2/2} \|\text{Unfold}_{\pi_1}(A)\|_\sigma,
\]
where \(c_1 = (k - \sum_{B \in \pi_1} \max_{B' \in \pi_2} |B \cap B'|)\) and \(c_2 = (k - \sum_{B \in \pi_2} \max_{B' \in \pi_1} |B \cap B'|)\).

(c) For \(k = 4\), the above inequalities in (b) are sharp. For example, consider the tensor \(A = I_4 \otimes I_4\), and partitions \(\pi = \{\{1, 2\}, \{3, 4\}\}\) and \(\pi' = \{\{1\}, \{2\}, \{3\}, \{4\}\}\), for which we have \(\|\text{Unfold}_{\pi}(A)\|_\sigma = d\) and \(\|\text{Unfold}_{\pi'}(A)\|_\sigma = 1\). If \(\pi_1 = \pi\) and \(\pi_2 = \pi'\), then \(c_1 = 2\) and \(d^{-c_1/2} \|\text{Unfold}_{\pi_1}(A)\|_\sigma = \|\text{Unfold}_{\pi_1}(A)\|_\sigma\). On the other hand, if \(\pi_1 = \pi'\) and \(\pi_2 = \pi\), then we have \(c_2 = 2\) and \(\|\text{Unfold}_{\pi_2}(A)\|_\sigma = d^{c_2/2} \|\text{Unfold}_{\pi_1}(A)\|_\sigma\). This particular tensor is further discussed later in Example 4.15.

**Proof of Theorem 4.8.** The main idea is to apply Lemma 4.2 to some appropriate sequence of partitions connecting \(\pi_1\) and \(\pi_2\) in the partition lattice. To do so, we consider Unfold\(_{\pi_1 \wedge \pi_2}(A)\) and compare its spectral norm to that of Unfold\(_{\pi_1}(A)\). Since \(\pi_1 \wedge \pi_2 \leq \pi_1\), from Proposition 4.1 we have
\[
\|\text{Unfold}_{\pi_1 \wedge \pi_2}(A)\|_\sigma \leq \|\text{Unfold}_{\pi_1}(A)\|_\sigma.
\]
Let \(\pi_1 = \{B_i^{\pi_1} : i \in [\ell_1]\}\) and \(\pi_2 = \{B_j^{\pi_2} : j \in [\ell_2]\}\), and note that \(\pi_1 \wedge \pi_2 = \{B_i^{\pi_1} \cap B_j^{\pi_2} : i \in [\ell_1], j \in [\ell_2]\}\). Now, order the blocks in \(\pi_2\) such that
\[
D_A(B_i^{\pi_1}, B_j^{\pi_2}) \geq D_A(B_i^{\pi_1}, B^*_{m_1}) \geq \cdots \geq D_A(B^*_{m_1}, B^*_{m_2}),
\]
and define \(B_{ij} = B_i^{\pi_1} \cap B^*_{m_2}\). Consider a sequence \((T_1, T_2, \ldots, T_{\ell_2 - 1})\) of unfoldings of \(A\) where \(T_j\), for \(1 \leq j \leq \ell_2 - 1\), is obtained from the tensor Unfold\(_{\pi_1 \wedge \pi_2}(A)\) by an unfolding operation
corresponding to merging the blocks $B_{1,1}, \ldots, B_{1,j+1}$ into a single block. Using Lemma 4.2 and (17), we obtain

$$\|\text{Unfold}_{\pi_1 \wedge \pi_2}(A)\|_\sigma \geq \min\{D_A(B_1^{\pi_1}, B_{m_1}^{\pi_2}), D_A(B_1^{\pi_1}, B_{m_2}^{\pi_2})\}^{-1/2} \|\mathcal{T}_1\|_\sigma$$

Similarly for all $i \in [\ell_2 - 2]$,

$$\|\mathcal{T}_i\|_\sigma \geq \min\{D_A(B_1^{\pi_1}, B_{m_{i+1}}^{\pi_2}), \prod_{j=1}^{i+1} D_A(B_1^{\pi_1}, B_{m_j}^{\pi_2})\}^{-1/2} \|\mathcal{T}_{i+1}\|_\sigma$$

Combining these inequalities gives

$$\|\text{Unfold}_{\pi_1 \wedge \pi_2}(A)\|_\sigma \geq \prod_{j=2}^{\ell_2} D_A(B_1^{\pi_1}, B_{m_j}^{\pi_2})^{-1/2} \|\mathcal{T}_{\ell_2-1}\|_\sigma$$

We can iterate the same line of argument with $\{B_1^{\pi_1} \cap B_{m_j}^{\pi_2} : j \in [\ell_2]\}$ for $i = 2, \ldots, \ell_1$ to obtain

$$\|\text{Unfold}_{\pi_1 \wedge \pi_2}(A)\|_\sigma \geq \prod_{i=1}^{\ell_1} \left[ \prod_{j=1}^{\ell_2} D_A(B_1^{\pi_1}, B_{m_j}^{\pi_2}) \right]^{-1/2} \|\text{Unfold}_{\pi_1}(A)\|_\sigma$$

Together with (16), this last inequality means

$$\text{(18) } \|\text{Unfold}_{\pi_1 \wedge \pi_2}(A)\|_\sigma \leq \|\text{Unfold}_{\pi_1}(A)\|_\sigma \leq \left[ \prod_{n \in [k]} \frac{d_n}{\dim_A(\pi_1, \pi_2)} \right]^{1/2} \|\text{Unfold}_{\pi_1 \wedge \pi_2}(A)\|_\sigma .$$

By symmetry,

$$\|\text{Unfold}_{\pi_1 \wedge \pi_2}(A)\|_\sigma \leq \left[ \prod_{n \in [k]} \frac{d_n}{\dim_A(\pi_2, \pi_1)} \right]^{1/2} \|\text{Unfold}_{\pi_1 \wedge \pi_2}(A)\|_\sigma .$$

Finally, combining (18) and (19) completes the proof. \hfill \square

We may immediately establish several corollaries of Theorem 4.8.
Corollary 4.10. All order-$k$ tensors $A \in \mathbb{R}^{d_1 \times \cdots \times d_k}$ satisfy

$$\|A\|_F \leq \left[ \frac{\text{dim}(A)}{\max_{n \in [k]} d_n} \right]^{1/2} \|A\|_{\sigma}.$$ 

**Proof.** Taking $\pi_1 = 0_{[k]}$ and $\pi_2 = 1_{[k]}$ in Theorem 4.8 yields the result. \qed

Corollary 4.10 gives the worst-case ratio of the Frobenius norm to the spectral norm for an arbitrary tensor. This ratio is sharper than the bound recently found by Friedland and Lim [5, Lemma 5.1], namely $\|A\|_F \leq \text{dim}(A)^{1/2} \|A\|_{\sigma}$.

We now give a set of inequalities comparing the spectral norms of unfoldings at level $\ell$ to that at either level $k$ or level 1. For ease of exposition, we assume $d_n = d$ for all $n \in [k]$.

**Corollary 4.11 (Bottom-Up Inequality).** Let $A \in \mathbb{R}^{d \times \cdots \times d}$ be an order-$k$ tensor with the same dimension $d$ in all modes. For all levels $1 \leq \ell \leq k$ and partitions $\pi \in \mathcal{P}_k^{\ell}$,

$$d^{-(k-\ell)/2} \max_{\pi \in \mathcal{P}_k^{\ell}} \|\text{Unfold}_\pi(A)\|_{\sigma} \leq \|A\|_{\sigma} \leq \min_{\pi \in \mathcal{P}_k^{\ell}} \|\text{Unfold}_\pi(A)\|_{\sigma}.$$ 

**Proof.** Take $\pi_2 = 0_{[k]}$ in Theorem 4.8. \qed

**Remark 4.12.** The existing work that is most closely related to our own is that of Hu’s [9], in which the author bounds the nuclear norm of a tensor by that of its matricization. Since the nuclear norm and spectral norm are dual to each other in tensor space, many of our results apply to the nuclear norm as well. Particularly, letting $\ell = 2$ in the bottom-up inequality (20) reproduces Hu’s results.

**Corollary 4.13 (Top-Down Inequality).** Let $A \in \mathbb{R}^{d \times \cdots \times d}$ be an order-$k$ tensor with the same dimension $d$ in all modes. For all levels $1 \leq \ell \leq k$ and partitions $\pi \in \mathcal{P}_k^{\ell}$,

$$d^{-(k - \max_{i \in [\ell]} |B^\pi_i|)/2} \|A\|_F \leq \|\text{Unfold}_\pi(A)\|_{\sigma} \leq \|A\|_F.$$ 

**Proof.** Take $\pi_1 = 1_{[k]}$ in Theorem 4.8. \qed

**Corollary 4.14.** Let $A \in \mathbb{R}^{d \times \cdots \times d}$ be an order-$k$ tensor with the same dimension $d$ in all modes. For all levels $1 \leq \ell \leq k$,

$$d^{-(k-\lceil k/\ell \rceil)/2} \|A\|_F \leq \min_{\pi \in \mathcal{P}_{k}^{\ell}} \|\text{Unfold}_\pi(A)\|_{\sigma}.$$ 

**Proof.** Note that the minimum of the maximal block sizes across all level-$\ell$ partitions of $[k]$ is $\min_{\pi \in \mathcal{P}_{k}^{\ell}} \max_{B \in \pi} |B| = \lceil k/\ell \rceil$, and apply Corollary 4.13. \qed

The above corollaries bound the amount by which norms can vary over a specific level $\ell$. They imply that the ratios $\|\text{Unfold}_\pi(A)\|_{\sigma} / \|A\|_{\sigma}$ and $\|\text{Unfold}_\pi(A)\|_{\sigma} / \|A\|_F$ fall in the intervals $[1, d^{(k-\ell)/2}]$ and $[d^{-(k-\lceil k/\ell \rceil)/2}, 1]$, respectively. Therefore, in the worst case, $\|\text{Unfold}_\pi(A)\|_{\sigma}$ only recovers $\|A\|_{\sigma}$ or $\|A\|_F$ at poly($d$) precision. Note that the factor $d^{(k-\ell)/2}$ has an exponent linear in the difference between the orders of the original tensor and its flattening. This means that the potential deviation between their spectral norms depends only on the difference in their orders rather than the actual orders themselves, and that the deviation accumulates in multiplicative fashion with a loss of $\sqrt{d}$ in precision at each level. In contrast, the factor $d^{-(k-\lceil k/\ell \rceil)/2}$ depends on more than just the gap between $k$ and $\ell$, with a larger impact for unfoldings with orders close to $k$.

We provide a low-order example that reaches the poly($d$) scaling factor in Corollary 4.11.
Example 4.15. Consider the order-4 tensor \( A = I_d \otimes I_d \). Straightforward calculation shows that \( \|A\|_\sigma = 1 \). Furthermore, by symmetry, the spectral norm of an unfolding induced by any partition in \( \mathcal{P}_4 \) must fall into one of the following five representative cases:

- For \( \pi_1 = \{\{1\}, \{2\}, \{3, 4\}\} \in \mathcal{P}_4 \), \( \|\text{Unfold}_{\pi_1}(A)\|_\sigma = \sqrt{d} \).
- For \( \pi_2 = \{\{1\}, \{3\}, \{2, 4\}\} \in \mathcal{P}_4 \), \( \|\text{Unfold}_{\pi_2}(A)\|_\sigma = 1 \).
- For \( \pi_3 = \{\{1, 2\}, \{3, 4\}\} \in \mathcal{P}_4 \), \( \|\text{Unfold}_{\pi_3}(A)\|_\sigma = d \).
- For \( \pi_4 = \{\{1\}, \{2, 3\}, \{4\}\} \in \mathcal{P}_4 \), \( \|\text{Unfold}_{\pi_4}(A)\|_\sigma = \sqrt{d} \).
- For \( \pi_5 = \{\{1, 3\}, \{2, 4\}\} \in \mathcal{P}_2 \), \( \|\text{Unfold}_{\pi_5}(A)\|_\sigma = 1 \).

Therefore,

\[
\max_{\pi \in \mathcal{P}_4} \|\text{Unfold}_\pi(A)\|_\sigma = d^{(4-3)/2} \|A\|_\sigma \quad \text{and} \quad \min_{\pi \in \mathcal{P}_4} \|\text{Unfold}_\pi(A)\|_\sigma = \|A\|_\sigma .
\]

\[
\max_{\pi \in \mathcal{P}_2} \|\text{Unfold}_\pi(A)\|_\sigma = d^{(4-2)/2} \|A\|_\sigma \quad \text{and} \quad \min_{\pi \in \mathcal{P}_2} \|\text{Unfold}_\pi(A)\|_\sigma = \|A\|_\sigma .
\]

We conclude this section by generalizing Theorem 4.8 to \( l^p \)-norms.

Theorem 4.16 (\( l^p \)-norm inequalities). Let \( A \in \mathbb{R}^{d_1 \times \cdots \times d_k} \) be an arbitrary order-\( k \) tensor, and \( \pi_1, \pi_2 \) any two partitions in \( \mathcal{P}_k \). Then,

(a) For any \( 1 \leq p \leq 2 \),

\[
\frac{[\dim(A)]^{-1/p}}{[\dim(A(\pi_1, \pi_2))^{-1/2}} \|\text{Unfold}_{\pi_1}(A)\|_p \leq \|\text{Unfold}_{\pi_2}(A)\|_p \leq \frac{[\dim(A)]^{1/p}}{[\dim(A(\pi_2, \pi_1))^{1/2}} \|\text{Unfold}_{\pi_1}(A)\|_p .
\]

(b) For any \( 2 \leq p \leq \infty \),

\[
\frac{[\dim(A)]^{1/p-1}}{[\dim(A(\pi_1, \pi_2))^{-1/2}} \|\text{Unfold}_{\pi_1}(A)\|_p \leq \|\text{Unfold}_{\pi_2}(A)\|_p \leq \frac{[\dim(A)]^{1-1/p}}{[\dim(A(\pi_2, \pi_1))^{1/2}} \|\text{Unfold}_{\pi_1}(A)\|_p .
\]

Proof. We only prove (a) since (b) follows similarly. For any given \( 1 \leq p \leq 2 \), taking \( q = 2 \) in Proposition 2.4 implies that the bound between the \( l^p \)-norm and spectral norm depends only on the total dimension of the tensor, \( \dim(A) = \prod_{n \in [k]} d_n \). Because the total dimension is invariant under any unfolding operation, we have

\[
(21) \quad \dim(A)^{\frac{1}{2} - \frac{1}{p}} \|\text{Unfold}_\pi(A)\|_\sigma \leq \|\text{Unfold}_\pi(A)\|_p \leq \|\text{Unfold}_\pi(A)\|_\sigma .
\]

for all \( \pi \in \mathcal{P}_k \). Combining (21) with Theorem 4.8 gives the desired results. \( \square \)

5. Orthogonal decomposability and norm equality on upper cones

We have seen that the unfolding operation may change the spectral norm by up to a poly(\( d \)) factor for an arbitrary \( A \). This is undesirable in many flattening-based algorithms, such as [3, 25]. However, for some specially-structured tensors, the operator norm on the partition lattice may not change much either globally or locally. We demonstrate such a behavior for the following class of tensors:

Definition 5.1 (\( \pi \)-orthogonal decomposable). Let \( A \in \mathbb{R}^{d_1 \times \cdots \times d_k} \) be an order-\( k \) tensor and consider any partition \( \pi \in \mathcal{P}_k \). Then \( A \) is called \( \pi \)-orthogonal decomposable, or \( \pi \)-OD, over \( \mathbb{R} \) if it admits the decomposition

\[
A = \sum_{n=1}^{r} \lambda_n a_1^{(n)} \otimes \cdots \otimes a_k^{(n)},
\]
where \( \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r \geq 0 \), and the set of vectors \( \{a_i^{(n)} \in \mathbb{R}^{d_i} : i \in [k], n \in [r] \} \) satisfies
\[
\langle \otimes_{i \in B} a_i^{(n)}, \otimes_{i \in B} a_i^{(m)} \rangle = \delta_{nm},
\]
for all \( B \in \pi \) and all \( n, m \in [r] \).

A concept similar to \( \pi \)-OD, referred to as biorthogonal eigentensor decomposition [28], is introduced in the tensor completion literature when \( k = 3 \) and \( \pi = \{\{1\}, \{2, 3\}\} \). Informally speaking, \( \pi \)-OD imposes an orthogonality constraint on every block of singular vectors.

**Remark 5.2 (0[k]-OD).** When \( \pi = 0[k] \) in Definition 5.1, we obtain the special case of 0[k]-OD tensors, which admit the decomposition (22) while satisfying
\[
\langle a_i^{(n)}, a_i^{(m)} \rangle = \delta_{nm},
\]
for all \( i \in [k] \) and all \( n, m \in [r] \).

The definition of 0[k]-OD tensors generalizes the definition of orthogonal decomposable tensors presented in [21], as we require neither symmetry nor equality of dimension across modes. In fact, a 0[k]-OD tensor \( \mathcal{A} \) is a diagonalizable tensor [3], meaning that the core tensor output from higher-order SVD is superdiagonal (i.e., entries are zero unless \( i_1 = \cdots = i_k \)).

**Lemma 5.3.** Consider an order-\( k \) tensor \( \mathcal{A} \in \mathbb{R}^{d_1 \times \cdots \times d_k} \).

(a) Let \( \pi_1, \pi_2 \in \mathcal{P}_r [k] \) and \( \pi_1 \subseteq \pi_2 \). If \( \mathcal{A} \) is \( \pi_1 \)-OD, then \( \mathcal{A} \) is \( \pi_2 \)-OD.

(b) Let \( \pi \in \mathcal{P}_r [k] \) and \( \pi \neq 1[k] \). If \( \mathcal{A} \) is \( \pi \)-OD, then
\[
\|\text{Unfold}_\pi (\mathcal{A})\|_\sigma = \lambda_1.
\]

**Proof.** Part (a): For any two finite sets of vectors \( \{x_i\} \) and \( \{y_i\} \) for which \( x_i, y_i \in \mathbb{R}^{d_i} \), we have
\[
\langle \otimes_{i \in B} x_i, \otimes_{j \in B} y_j \rangle = \prod_{i \in B} \langle x_i, y_i \rangle,
\]
for all \( B \subseteq [k] \). Suppose \( B \in \pi_2 \). If \( \pi_1 \subseteq \pi_2 \), then there exist subsets \( C_1, \ldots, C_m \subseteq \pi_1 \) such that \( C_1 \cup \cdots \cup C_m = B \). So,
\[
\langle \otimes_{i \in B} x_i, \otimes_{j \in B} y_j \rangle = \prod_{i \in B} \langle x_i, y_i \rangle = \prod_{a=1}^m \prod_{i \in C_a} \langle x_i, y_i \rangle = \prod_{a=1}^m \langle \otimes_{i \in C_a} x_i, \otimes_{j \in C_a} y_j \rangle,
\]
which implies that \( \mathcal{A} \) is \( \pi_2 \)-OD if \( \mathcal{A} \) is \( \pi_1 \)-OD.

Part (b): Suppose \( \pi \in \mathcal{P}_r [k] \) is of the form \( \pi = \{B_i^\tau : i \in [\ell]\} \). Note that \( \pi \neq 1[k] \) implies \( \ell \geq 2 \). Letting \( \tau = \{B_i^\tau, (B_i^\tau)^c\} \) where \( (B_i^\tau)^c \) denotes the complement of \( B_i^\tau \) with respect to \([k]\), we have \( \pi \in \mathcal{P}_r [k] \) and \( \pi \leq \tau \). By Lemma 5.3(a), \( \mathcal{A} \) is \( \tau \)-OD, so \( \mathcal{A} \) admits a decomposition of the form
\[
\mathcal{A} = \sum_{n=1}^r \lambda_n a_1^{(n)} \otimes \cdots \otimes a_k^{(n)},
\]
where
\[
\langle \otimes_{i \in B_1} a_i^{(n)}, \otimes_{j \in B_1} a_j^{(m)} \rangle = \delta_{nm} \quad \text{and} \quad \langle \otimes_{i \in (B_1^\tau)^c} a_i^{(n)}, \otimes_{j \in (B_1^\tau)^c} a_j^{(m)} \rangle = \delta_{nm}
\]
for all \( n, m \in [r] \).

Now define \( x_n = \text{Vec}(\otimes_{i \in B_1} a_i^{(n)}) \) and \( y_n = \text{Vec}(\otimes_{i \in (B_1^\tau)^c} a_i^{(n)}) \) for all \( n \in [r] \). By (26), both \( \{x_n\} \) and \( \{y_n\} \) are sets of orthonormal vectors. By the definition of Unfold\( r \)(\( \mathcal{A} \)), (25) implies
\[
\text{Unfold}_r (\mathcal{A}) = \sum_{n=1}^r \lambda_n x_n y_n^T,
\]
which is simply the matrix SVD of Unfold$_\pi$(A). Hence \(\|\text{Unfold}_\pi(A)\|_\sigma = \lambda_1\). Using monotonicity (c.f., Proposition 4.1), we have
\[
(27) \quad \|\text{Unfold}_\pi(A)\|_\sigma \leq \|\text{Unfold}_\tau(A)\|_\sigma = \lambda_1.
\]
Conversely, by the definition of the spectral norm, we have
\[
\|\text{Unfold}_\tau(A)\|_\sigma \geq \left\langle \text{Unfold}_\tau(A), \text{Vec}\left(\otimes_{i \in B_1^\tau} a_i^{(1)}\right) \otimes \cdots \otimes \text{Vec}\left(\otimes_{i \in B_r^\tau} a_i^{(n)}\right)\right\rangle 
= \sum_{n=1}^r \lambda_n \left\langle \text{Vec}\left(\otimes_{i \in B_1^\tau} a_i^{(n)}\right) \otimes \cdots \otimes \text{Vec}\left(\otimes_{i \in B_r^\tau} a_i^{(n)}\right)\right\rangle,
\]
\[
= \sum_{n=1}^r \lambda_n \prod_{j \in [r]} \left\langle \text{Vec}\left(\otimes_{i \in B_j^\tau} a_i^{(n)}\right), \text{Vec}\left(\otimes_{i \in B_j^\tau} a_i^{(1)}\right)\right\rangle 
= \sum_{n=1}^r \lambda_n \delta_{n,1} = \lambda_1
\]
where the third line comes from (24) and the last line follows from the fact that \(A\) is \(\pi\)-OD. Combining (27) and (28), we conclude \(\|\text{Unfold}_\pi(A)\|_\sigma = \lambda_1\).

**Remark 5.4.** The condition \(\pi \neq 1_{[k]}\) in Lemma 5.3 is needed for (23) to hold. In fact, consider a \(2 \times 2\) matrix \(A = 2e_1 \otimes e_1 + e_1 \otimes e_2\), where \(\{e_i, i \in [2]\}\) is the canonical basis of \(\mathbb{R}^2\). Then, \(A\) is \(1_{[k]}\)-OD with \(\lambda_1 = 2\), but \(\|\text{Unfold}_{1_{[k]} \cdot OD}(A)\|_\sigma = \|A\|_F = \sqrt{5} \neq 2\).

**Theorem 5.5** (Norm equality on upper cones). If \(A\) is \(\pi\)-OD, then for any partition in the upper cone \(U_\pi = \{\tau \in \mathcal{P}_{[k]} : \pi \leq \tau < 1_{[k]}\}\) of \(\pi\), we have
\[
\|\text{Unfold}_\tau(A)\|_\sigma = \|\text{Unfold}_\pi(A)\|_\sigma.
\]

**Proof.** If \(A\) is \(\pi\)-OD, then by Lemma 5.3(b), we have \(\|\text{Unfold}_\pi(A)\|_\sigma = \lambda_1\). Given any \(\tau \geq \pi\), by Lemma 5.3(a), \(A\) is also \(\tau\)-OD. Again, using Lemma 5.3(b), we have \(\|\text{Unfold}_\tau(A)\|_\sigma = \lambda_1\). Therefore, \(\|\text{Unfold}_\tau(A)\|_\sigma = \|\text{Unfold}_\tau(A)\|_\sigma = \|\text{Unfold}_\pi(A)\|_\sigma\).

Theorem 5.5 states that the spectral norm is invariant for \(\pi\)-OD \(A\) under any unfolding induced by the partitions in the upper cone \(U_\pi\) of \(\pi\). This lies in contrast with the poly\((d)\) factor we have seen for unstructured tensors.

**Corollary 5.6.** If \(A\) is \(0_{[k]}\)-OD, then for all partitions \(\pi \neq 1_{[k]}\), we have
\[
\|\text{Unfold}_\pi(A)\|_\sigma = \|A\|_\sigma.
\]

Corollary 5.6 implies that for \(0_{[k]}\)-OD tensors, the operator norm is invariant under any unfolding operations except vectorization. Lastly, \(\pi_1, \pi_2 \in U_{\pi_1 \wedge \pi_2}\) implies the following corollary:

**Corollary 5.7.** Let \(\pi_1, \pi_2 \in \mathcal{P}_{[k]}\). If \(A\) is \((\pi_1 \wedge \pi_2)\)-OD, then
\[
\|\text{Unfold}_{\pi_1}(A)\|_\sigma = \|\text{Unfold}_{\pi_2}(A)\|_\sigma.
\]

6. **Discussion**

In this paper, we presented a new framework representing all possible tensor unfoldings by the partition lattice and established a set of general inequalities quantifying the impact of tensor unfoldings on the operator norms of the resulting tensors. We showed that the comparison bounds scale polynomially in the dimensions \(\{d_n\}\) of the tensor, with powers depending on the corresponding partition and block sizes for any pair of tensor unfoldings being compared. As a direct consequence,
we demonstrated how the operator norm of a general tensor is lower and upper bounded by that of its unfoldings.

In general, an unfolding operation may inflate the operator norm by up to a poly($d$) factor, as seen in Corollary 4.11. Note that the quantity dim($A$) plays a key role in the worst-case inflation factor and is a manifestation of the curse of dimensionality. Specifically, dim($A$) can be quite large as the mode dimensions and tensor order increase, with particular sensitivity to the latter. In such settings, our main result seems to bode poorly for flattening-based algorithms; however, we believe that it should be interpreted with caution because our comparison bounds deal with arbitrary tensors rather than those often sought in applications. In fact, $\pi$-OD tensors permit much tighter bounds in which some unfoldings, including certain matricizations, leave the operator norm relatively unaffected. In practice, $\pi$-OD tensors, or those within a small neighborhood around $\pi$-OD tensors, arise widely in statistical and machine learning applications [1,11,16,26].

Additionally, our work enables us to compare different unfoldings at the same level $\ell$. Recent work on problems featuring nuclear-norm regularization has shown that not all $n$-mode flattenings are equally preferable [13]. Indeed, as illustrated in Example 4.15, the operator norm of level-2 unfoldings (i.e. matricizations) can be quite different. Recently, several algorithms have been proposed to account for this behavior. For example, Tomioka et al. [23] consider a weighted sum of the norms of all single-mode matricizations. Other techniques include two-mode matricization [26] and square matricization [15,20], in which the original tensor is reshaped into a matrix by flattening along multiple modes. Our work provides general bounds to evaluate the effectiveness of such schemes. In particular, the results presented here are used in the theoretical analysis of a two-mode higher-order SVD algorithm proposed recently [26].

We have not attempted to characterize the degree to which operator norm relations on the partition lattice restrict the original tensor. Essentially, this is a converse problem asking whether $\pi$-OD is a necessary condition for Theorem 5.5 and Corollary 5.6 in addition to being sufficient. If not, it would be useful to determine the extent to which such equalities inform us about the intrinsic structure of the original tensor. From a practical standpoint, norm comparisons between different matricizations are relatively simple, but the optimal manner in which to use this information to learn about the original tensor remains unknown.

In closing, we emphasize that while this work focuses on theory rather than computational tractability, it possesses practical implications as well. Because direct calculation of the operator norm of a level-$\ell$ tensor is generally computationally prohibitive for $\ell \geq 3$, exploiting level-2 unfoldings may be attractive when the unfolding effect is small enough. Alternatively, for more precise calculations, a number of approximation algorithms exist for higher-order tensor problems [23,27] at the cost of increased computation. Given that the trade-off between accuracy and computation is often unavoidable, our work may be of help in finding an appropriate application-specific balance when working with higher-order tensors.

ACKNOWLEDGEMENTS

This research is supported in part by a Math+X Research Grant from the Simons Foundation, a Packard Fellowship for Science and Engineering, and an NIH training grant T32-HG000047.

REFERENCES

[1] A. Anandkumar, R. Ge, D. Hsu, S. M. Kakade, and M. Telgarsky, Tensor decompositions for learning latent variable models, The Journal of Machine Learning Research 15 (2014), no. 1, 2773–2832.
[2] J.-F. Cardoso, Eigen-structure of the fourth-order cumulant tensor with application to the blind source separation problem, International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 1990, pp. 2655–2658.
[3] L. De Lathauwer, B. De Moor, and J. Vandewalle, A multilinear singular value decomposition, SIAM Journal on Matrix Analysis and Applications 21 (2000), no. 4, 1253–1278.
[4] S. Friedland and L.-H. Lim, Computational complexity of tensor nuclear norm, SIAM Journal on Optimization 4 (2016), no. 26, 2378–2393.
17

[5] ______, Nuclear norm of higher-order tensors, Mathematics of Computation (to appear).
[6] C. J. Hillar and L.-H. Lim, Most tensor problems are NP-hard, Journal of the ACM 60 (2013), no. 6, Article 45.
[7] P. D. Hoff, Multilinear tensor regression for longitudinal relational data, The Annals of Applied Statistics 9 (2015), no. 3, 1169–1193.
[8] R. A. Horn and C. R. Johnson, Matrix Analysis, Cambridge University Press, 2012.
[9] S. Hu, Relations of the nuclear norm of a tensor and its matrix flattenings, Linear Algebra and its Applications 478 (2015), 188–199.
[10] P. M Kroonenberg, Three-Mode Principal Component Analysis: Theory and Applications, Vol. 2, DSWO Press, 1983.
[11] V. Kuleshov, A. Chaganty, and P. Liang, Tensor factorization via matrix factorization, Proceedings of the 18th International Conference on Artificial Intelligence and Statistics (AISTATS), 2015, pp. 507–516.
[12] L.-H. Lim, Singular values and eigenvalues of tensors: a variational approach, Proceedings of the IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2005, pp. 129–132.
[13] J. Liu, P. Musialski, P. Wonka, and J. Ye, Tensor completion for estimating missing values in visual data, IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (2013), no. 1, 208–220.
[14] P. McCullagh, Tensor notation and cumulants of polynomials, Biometrika 71 (1984), no. 3, 461–476.
[15] C. Mu, B. Huang, J. Wright, and D. Goldfarb, Square deal: Lower bounds and improved relaxations for tensor recovery, Proceedings of the 31st International Conference on Machine Learning (ICML), 2014, pp. 73–81.
[16] M. Nickel, V. Tresp, and H.-P. Kriegel, A three-way model for collective learning on multi-relational data, Proceedings of the 28th International Conference on Machine Learning (ICML), 2011, pp. 809–816.
[17] L. Omberg, G. H. Golub, and O. Alter, A tensor higher-order singular value decomposition for integrative analysis of DNA microarray data from different studies, Proceedings of the National Academy of Sciences 104 (2007), no. 47, 18371–18376.
[18] L. Qi, Eigenvalues of a real supersymmetric tensor, Journal of Symbolic Computation 40 (2005), no. 6, 1302–1324.
[19] S. Ragnarsson and C. F. Van Loan, Block tensor unfoldings, SIAM Journal on Matrix Analysis and Applications 33 (2012), no. 1, 149–169.
[20] E. Richard and A. Montanari, A statistical model for tensor PCA, Advances in Neural Information Processing Systems (NIPS) 27, 2014, pp. 2897–2905.
[21] E. Robeva, Orthogonal decomposition of symmetric tensors, SIAM Journal on Matrix Analysis and Applications 37 (2016), 86–102.
[22] J. Sun, D. Tao, and C. Faloutsos, Beyond streams and graphs: dynamic tensor analysis, Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2006, pp. 374–383.
[23] R. Tomioka, K. Hayashi, and H. Kashima, Estimation of low-rank tensors via convex optimization, arXiv:1010.0789 (2010).
[24] R. Tomioka and T. Suzuki, Spectral norm of random tensors, arXiv:1407.1870 (2014).
[25] M. A. O. Vasilescu and D. Terzopoulos, Multilinear analysis of image ensembles: Tensorfaces, Proceedings of European Conference on Computer Vision (ECCV), 2002, pp. 447–460.
[26] M. Wang and Y. S. Song, Orthogonal tensor decompositions via two-mode higher-order SVD (HOSVD), arXiv:1612.03839 (2016).
[27] Y. Yu, H. Cheng, and X. Zhang, Approximate low-rank tensor learning, 7th NIPS Workshop on Optimization for Machine Learning, 2014.
[28] M. Yuan and C.-H. Zhang, On tensor completion via nuclear norm minimization, Foundations of Computational Mathematics (2015), 1–38.
[29] L. Zhao and M. J. Zaki, Tricluster: an effective algorithm for mining coherent clusters in 3D microarray data, Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data, 2005, pp. 694–705.