OPERATOR NORM INEQUALITIES BETWEEN TENSOR UNFOLDINGS ON THE PARTITION LATTICE

Miaoyan Wang[1], **Khanh Dao Duc**[1], **Jonathan Fischer**[2], and **Yun S. Song**[1,2,3]

[1]Department of Mathematics, University of Pennsylvania

[2]Department of Statistics, University of California, Berkeley

[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-k tensor, which are in one-to-one correspondence with the set of partitions of $\{1, \ldots, k\}$. 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.

Key words and phrases

higher-order tensors; general unfoldings; partition lattice; operator norm; orthogonality

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 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 π of $[k] = \{1, \ldots, k\}$, where a block in π 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 $\mathbb{P}$ 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 and establish our main inequalities relating their operator norms. For specially-structured tensors satisfying a generalized definition of orthogonal decomposability, we show that the spectral norm ($p = 2$) remains invariant under unfolding operations corresponding to a specific subset of partitions. To our knowledge, our results represent the first attempt to provide a full picture of the norm landscape over all possible tensor unfoldings.

The remainder of this paper is organized as follows. In Section 2, we introduce some notation, and relate the spectral norm and the general $\mathbb{P}$-norm of a tensor. We then describe in Section 3 general tensor unfoldings defined on the partition lattice. In Section 4, we present our main results on the inequalities between the operator norms of any two tensor unfoldings and describe how the norm landscape changes over the partition lattice. In Section 5, we generalize the notion of orthogonal decomposable tensors and prove that the

spectral norm is invariant within a specific set of 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 $\mathscr{A} = [\![a_{i_1 \ldots i_k}]\!] \in \mathbb{F}^{d_1 \times \cdots \times d_k}$ 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 $\mathscr{A}$ is denoted by $\dim(\mathscr{A}) = \prod_{n=1}^{k} d_n$.

The vectorization of $\mathscr{A}$, denoted $\mathrm{Vec}(\mathscr{A})$, is defined as the operation rearranging all elements of $\mathscr{A}$ 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 $\mathscr{A} = [\![a_{i_1 \ldots i_k}]\!]$, $\mathscr{B} = [\![b_{i_1 \ldots i_k}]\!] \in \mathbb{R}^{d_1 \times \cdots \times d_k}$ of identical order and dimensions, their inner product is defined as

$$\langle \mathscr{A}, \mathscr{B} \rangle = \sum_{i_1, \ldots, i_k} a_{i_1 \ldots i_k} b_{i_1 \ldots i_k},$$

while the tensor Frobenius norm of $\mathscr{A}$ is defined as

$$\|\mathscr{A}\|_F = \sqrt{\langle \mathscr{A}, \mathscr{A} \rangle} = \sqrt{\sum_{i_1, \ldots, i_k} |a_{i_1 \ldots i_k}|^2},$$

both of which are analogues of standard definitions for vectors and matrices.

Following [12], we define the *covariant multilinear matrix multiplication* of a tensor $\mathscr{A} \in \mathbb{R}^{d_1 \times \cdots \times d_k}$ by matrices $M_1 = (m_{i_1 j_1}^{(1)}) \in \mathbb{R}^{d_1 \times s_1}, \ldots, M_k = (m_{i_k j_k}^{(k)}) \in \mathbb{R}^{d_k \times s_k}$ as

$$\mathscr{A}(M_1, \ldots, M_k) = [\![\sum_{i_1=1}^{d_1} \cdots \sum_{i_k=1}^{d_k} a_{i_1 \ldots i_k} m_{i_1 j_1}^{(1)} \cdots m_{i_k j_k}^{(k)}]\!],$$

which results in an order-k tensor in $\mathbb{R}^{s_1 \times \cdots \times s_k}$. This operation multiplies the nth mode of $\mathscr{A}$ 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 $k = 2$, $\mathscr{A}$ is a matrix and $\mathscr{A}(M_1, M_2) = M_1^T \mathscr{A} M_2$. Sometimes we are interested in multiplying by vectors rather than matrices, in which case we obtain the k-multilinear functional $\mathscr{A} : \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_k} \to \mathbb{R}$ given by

$$\mathscr{A}(\boldsymbol{x}_1,\dots,\boldsymbol{x}_k)=\sum_{i_1=1}^{d_1}\cdots\sum_{i_k=1}^{d_k}a_{i_1\dots i_k}x_{i_1}^{(1)}\cdots x_{i_k}^{(k)}=\langle\mathscr{A},\boldsymbol{x}_1\otimes\cdots\otimes\boldsymbol{x}_k\rangle, \tag{1}$$

where $\boldsymbol{x}_n=\left(x_1^{(n)},\dots,x_{d_n}^{(n)}\right)^T\in\mathbb{R}^{d_n}$, $n\in[k]$. 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 $\mathscr{A}$ using the associated k-multilinear functional (1).

Definition 2.1 (Lim [12])

Let $\mathscr{A}\in\mathbb{R}^{d_1\times\cdots\times d_k}$ be an order-k tensor. For any $1\leq p\leq\infty$, the $\mathbb{P}$-norm of the multilinear functional associated with $\mathscr{A}$ is defined as

$$\|\mathscr{A}\|_p=\sup\left\{\frac{\mathscr{A}(\boldsymbol{x}_1,\dots,\boldsymbol{x}_k)}{\|\boldsymbol{x}_1\|_p\cdots\|\boldsymbol{x}_k\|_p}:\boldsymbol{x}_n\neq 0,\boldsymbol{x}_n\in\mathbb{R}^{d_n},n\in[k]\right\}=\sup\left\{\mathscr{A}(\boldsymbol{x}_1,\dots,\boldsymbol{x}_k):\|\boldsymbol{x}_n\|_p=1,\boldsymbol{x}_n\in\mathbb{R}^{d_n},n\in[k]\right\},$$

$$\tag{2}$$

where $\|\boldsymbol{x}_n\|_p$ denotes the vector $\mathbb{P}$-norm of $\boldsymbol{x}_n$.

Remark 2.2

The special case of $p=2$ is called the *spectral norm*, frequently denoted $\|\mathscr{A}\|_\sigma$. By (2), $\|\mathscr{A}\|_\sigma$ is the maximum value obtained as the inner product of the tensor $\mathscr{A}$ with a rank-1 tensor, $\boldsymbol{x}_1\otimes\cdots\otimes\boldsymbol{x}_n$, of Frobenius norm 1 and of the same dimensions. This point of view provides an equivalent definition of $\|\mathscr{A}\|_\sigma$ as determining the best rank-1 tensor approximation to $\mathscr{A}$, 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 $\mathscr{A}$ and $\{\boldsymbol{x}_n\}_{n=1}^k$ to be real, we need not take the absolute value of $\mathscr{A}(\boldsymbol{x}_1,\dots,\boldsymbol{x}_k)$ as in [12]. It is worth mentioning that the notion of tensor $\mathbb{P}$-norms defined by (2) are *not* extensions of the classical matrix $\mathbb{P}$-norms when $p\neq 2$. To see this, recall that for an $m\times n$ matrix $\boldsymbol{A}$, one usually defines the $\mathbb{P}$ operator norm as

$$\|\boldsymbol{A}\|_p=\sup\left\{\frac{\|\boldsymbol{A}\boldsymbol{x}\|_p}{\|\boldsymbol{x}\|_p}:\boldsymbol{x}\neq\boldsymbol{0},\boldsymbol{x}\in\mathbb{R}^n\right\}. \tag{3}$$

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 2×2 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

$$\|A\|_1 = \sup\left\{\frac{x_1^T A x_2}{\|x_1\|_1 \|x_2\|_1} : x_n \in \mathbb{R}^2, x_n \neq 0, n \in [2]\right\} = \sup\left\{x_1^T A x_2 : x_n \in \mathbb{R}^2, \|x_n\|_1 = 1, n \in [2]\right\} = 4,$$

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, $\dim(\mathscr{A}) = \prod_{n=1}^k d_n$.

Proposition 2.4 (l^p-norm vs. l^q-norm)

Let $\mathscr{A} \in \mathbb{R}^{d_1 \times \cdots \times d_k}$ be an order-k tensor. Suppose $\|\cdot\|_p$ and $\|\cdot\|_q$ are two norms defined in (2) with $q \geq p \geq 1$. Then,

$$\|\mathscr{A}\|_p \leq \|\mathscr{A}\|_q \leq \dim(\mathscr{A})^{\frac{1}{p} - \frac{1}{q}} \|\mathscr{A}\|_p.$$

Proof—Starting from Definition 2.1, we have

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

$$= \sup\left\{\frac{\mathscr{A}(x_1, \ldots, x_k)}{\|x_1\|_p \cdots \|x_k\|_p} \times \frac{\|x_1\|_p \cdots \|x_k\|_p}{\|x_1\|_q \cdots \|x_k\|_q} : x_n \neq 0, x_n \in \mathbb{R}^{d_n}, n \in [k]\right\}. \tag{4}$$

For any $q \geq p \geq 1$, the equivalence of vector norms tells us

$$\|\boldsymbol{x}\|_q \leq \|\boldsymbol{x}\|_p \leq d^{\frac{1}{p} - \frac{1}{q}} \|\boldsymbol{x}\|_q, \quad \text{for all } \boldsymbol{x} \in \mathbb{R}^d. \quad (5)$$

Applying (5) to $\boldsymbol{x}_n$, for $n \in [k]$, gives

$$1 \leq \prod_{n=1}^{k} \frac{\|\boldsymbol{x}_n\|_p}{\|\boldsymbol{x}_n\|_q} \leq \left(\prod_{n=1}^{k} d_n\right)^{\frac{1}{p} - \frac{1}{q}}. \quad (6)$$

Inserting (6) into (4) and noting $\dim(\mathscr{A}) = \prod_{n=1}^{k} d_n$, we find

$$\|\mathscr{A}\|_p \leq \|\mathscr{A}\|_q \leq \dim(\mathscr{A})^{\frac{1}{p} - \frac{1}{q}} \|\mathscr{A}\|_p,$$

which completes the proof.

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 $d_n \times \prod_{i \neq n} d_i$ matrix. For example, the n-mode matricization of a tensor $\mathscr{A} \in \mathbb{R}^{d_1 \times \cdots \times d_k}$ is obtained by mapping the fixed tensor index $(i_1, \ldots, i_k)$ to the matrix index (i_n, m), where

$$m = 1 + \prod_{a \in [k], a \neq n} (i_a - 1) J_a, \quad \text{where } J_a = \prod_{l \in [a-1], l \neq n} d_l. \quad (7)$$

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 $\mathscr{A}$.

For any $k \in \mathbb{N}_+$, a partition $\boldsymbol{\pi}$ of $[k]$ is a collection $\{B_1^\pi, B_2^\pi, \ldots, B_\ell^\pi\}$ of disjoint, nonempty subsets (or blocks) B_i^π satisfying $\cup_{i=1}^\ell B_i^\pi = [k]$. The set of all partitions of $[k]$ is denoted $\mathscr{P}_{[k]}$. We use $|\boldsymbol{\pi}|$ to denote the number of blocks in $\boldsymbol{\pi}$, and $|B_i^\pi|$ to denote the number of elements in B_i^π. We say that a partition $\boldsymbol{\pi}$ is a level-ℓ partition if $|\boldsymbol{\pi}| = \ell$. The set of all level-$\ell$ partitions of $[k]$ is denoted by $\mathscr{P}_{[k]}^\ell$, 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 $\pi_1 \in \mathscr{P}_{[k]}$ is called a *refinement* of $\pi_2 \in \mathscr{P}_{[k]}$ if each block of π_1 is a subset of some block of π_2; conversely, π_2 is said to be a *coarsening* of π_1. This relationship defines a partial order, expressed as $\pi_1 \preceq \pi_2$, and we say that π_1 is *finer* than π_2 while π_2 is *coarser* than π_1. If either $\pi_1 \preceq \pi_2$ or $\pi_2 \preceq \pi_1$, then π_1 and π_2 are *comparable*. According to this partial order, the least element of $\mathscr{P}_{[k]}$ is $\mathbf{0}_{[k]} := \{\{1\}, \ldots, \{k\}\}$, while the greatest element is $\mathbf{1}_{[k]} := \{\{1, \ldots, k\}\}$. Equipped with this notion, $\mathscr{P}_{[k]}$ generates a partition lattice by connecting any two comparable partitions that differ by exactly one level. An example is illustrated in Figure 1. Henceforth, $\mathscr{P}_{[k]}$ 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 $\pi_1, \pi_2 \in \mathscr{P}_{[k]}$, we define their *greatest lower bound* $\pi_1 \wedge \pi_2$ as

$$\pi_1 \wedge \pi_2 := \sup\{\pi \in \mathscr{P}_{[k]} : \pi \leq \pi_1, \pi \leq \pi_2\}.$$

More concretely, $\pi_1 \wedge \pi_2$ consists of the collection of all nonempty intersections of blocks in π_1 and π_2 and is unique for a given pair (π_1, π_2).

Example 3.2

Figure 1 illustrates $\mathscr{P}_{[4]}$ 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 $\pi_1 \wedge \pi_2 = \{\{1, 4\}, \{2\}, \{3\}\}$, $\mathbf{0}_{[4]} = \{\{1\}, \{2\}, \{3\}, \{4\}\}$, and $\mathbf{1}_{[4]} = \{\{1, 2, 3, 4\}\}$.

The following definition generalizes the concept of *n*-mode flattenings defined in (7) to general unfoldings induced by arbitrary partitions $\pi \in \mathscr{P}_{[k]}$.

Definition 3.3 (General Tensor Unfolding)

Let $\mathscr{A} \in \mathbb{R}^{d_1 \times \cdots \times d_k}$ be an order-k tensor and $\pi = \{B_1^\pi, \ldots, B_\ell^\pi\} \in \mathscr{P}_{[k]}$. The partition π defines a mapping $\phi_\pi : [d_1] \times \cdots \times [d_k] \to [\prod_{j \in B_1^\pi} d_j] \times \cdots \times [\prod_{j \in B_\ell^\pi} d_j]$ such that

$$\phi_\pi(i_1, \ldots, i_k) = (m_1, \ldots, m_\ell), \quad (8)$$

where

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

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

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

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

$$(\text{Unfold}_\pi(\mathscr{A}))_{\phi_\pi(i_1, \ldots, i_k)} = (\mathscr{A})_{(i_1, \ldots, i_k)},$$

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

Remark 3.4

By the definition of ϕ_π in (8), $\text{Unfold}_{\mathbf{0}_{[k]}}(\mathscr{A}) = \mathscr{A}$ and $\text{Unfold}_{\mathbf{1}_{[k]}}(\mathscr{A}) = \text{Vec}(\mathscr{A})$.

Example 3.5

Consider an order-4 tensor $\mathscr{A} = [\![a_{ijkl}]\!] \in \mathbb{R}^{2 \times 2 \times 2 \times 2}$. 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 $\pi = \{\{1, 2\}, \{3\}, \{4\}\}$, $\text{Unfold}_\pi(\mathscr{A})$ is an order-3 tensor of dimensions (4, 2, 2) with entries given by

$$(\text{Unfold}_\pi(\mathscr{A}))_{1kl} = a_{11kl}, (\text{Unfold}_\pi(\mathscr{A}))_{2kl} = a_{12kl},$$

$$(\text{Unfold}_\pi(\mathscr{A}))_{3kl} = a_{21kl}, (\text{Unfold}_\pi(\mathscr{A}))_{4kl} = a_{22kl},$$

for all $(k, l) \in [2] \times [2]$.

- For $\pi = \{\{1, 2\}, \{3, 4\}\}$, $\text{Unfold}_\pi(\mathscr{A})$ is a 4×4 matrix

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

- For $\pi = \{\{1, 2, 3\}, \{4\}\}$, $\text{Unfold}_\pi(\mathscr{A})$ is a 8×2 matrix

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

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, $\|\mathscr{A}\|_F = \|\mathrm{Unfold}_\pi(\mathscr{A})\|_F$ for all $\pi \in \mathscr{P}_{[k]}$. More generally, the inner product remains invariant under all unfoldings: $\langle \mathscr{A}, \mathscr{B} \rangle = \langle \mathrm{Unfold}_\pi(\mathscr{A}), \mathrm{Unfold}_\pi(\mathscr{B}) \rangle$ for all $\pi \in \mathscr{P}_{[k]}$, where $\mathscr{A}, \mathscr{B} \in \mathbb{R}^{d_1 \times \cdots \times d_k}$ 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 $\mathbb{P}$-norms.

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

$$\|\mathscr{A}\|_\sigma = \sup \left\{ \mathscr{A}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) : \|\boldsymbol{x}_n\|_2 = 1, \boldsymbol{x}_n \in \mathbb{R}^{d_n}, n \in [k] \right\}. \quad (9)$$

The maximization of the polynomial form $\mathscr{A}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k)$ 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 $\mathscr{A} \in \mathbb{R}^{d_1 \times \cdots \times d_k}$, the problem of determining the spectral norm is equivalent to finding a scalar λ and a rank-1 norm-1 tensor $\boldsymbol{x}_1 \otimes \cdots \otimes \boldsymbol{x}_k$ that minimize the function

$$f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) = \|\mathscr{A} - \lambda \boldsymbol{x}_1 \otimes \cdots \otimes \boldsymbol{x}_k\|_F.$$

The corresponding value of λ is equal to $\|\mathscr{A}\|_\sigma$. Since $\boldsymbol{x}_1 \otimes \cdots \otimes \boldsymbol{x}_k$ must have the same order and dimensions as $\mathscr{A}$, the rank-1 condition becomes less strict the more unfolded the tensor. In particular, for the vectorized unfolding, $\mathrm{Vec}(\mathscr{A})$, the best rank-1 tensor approximation is simply $\mathrm{Vec}(\mathscr{A})$ itself. For unfoldings into a matrix, $\mathrm{Mat}(\mathscr{A})$, the best rank-1 approximation is the outer product of the leading left and the right singular vectors of $\mathrm{Mat}(\mathscr{A})$. 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 π_1, $\pi_2 \in \mathscr{P}_{[k]}$ satisfying $\pi_1 \preceq \pi_2$,

$$\left\|\mathrm{Unfold}_{\pi_1}(\mathscr{A})\right\|_\sigma \leq \left\|\mathrm{Unfold}_{\pi_2}(\mathscr{A})\right\|_\sigma.$$

In particular, we have the global extrema

$$\left\|\mathscr{A}\right\|_\sigma = \left\|\mathrm{Unfold}_{0_{[k]}}(\mathscr{A})\right\|_\sigma = \min_{\pi \in \mathscr{P}_{[k]}} \left\|\mathrm{Unfold}_\pi(\mathscr{A})\right\|_\sigma,$$

$$\left\|\mathscr{A}\right\|_F = \left\|\mathrm{Unfold}_{1_{[k]}}(\mathscr{A})\right\|_\sigma = \max_{\pi \in \mathscr{P}_{[k]}} \left\|\mathrm{Unfold}_\pi(\mathscr{A})\right\|_\sigma.$$

Proof—Suppose π_1, $\pi_2 \in \mathscr{P}_{[k]}$, and π_1 is a one-step refinement of π_2. Without loss of generality, assume π_2 is obtained by merging two blocks B_1, $B_2 \in \pi_1$ into a single block $B \in \pi_2$. Let $\mathrm{Unfold}_{\pi_1}(\mathscr{A}) \in \mathbb{R}^{d_1 \times \cdots \times d_\ell}$ denote the tensor unfolding induced by $\pi_1 = \{B_1, \ldots, B_\ell\}$. Then, $\mathrm{Unfold}_{\pi_2}(\mathscr{A})$ is a $(d_1 \cdot d_2, d_3, \ldots, d_\ell)$-dimensional tensor. By definition of the spectral norm,

$$
\begin{aligned}
\left\|\mathrm{Unfold}_{\pi_1}(\mathscr{A})\right\|_\sigma &= \sup \left\{ \langle \mathrm{Unfold}_{\pi_1}(\mathscr{A}), \boldsymbol{x}_1 \otimes \cdots \otimes \boldsymbol{x}_\ell \rangle : \boldsymbol{x}_n \in S^{d_n - 1}, n \in [\ell] \right\} \\
&= \sup \left\{ \langle \mathrm{Unfold}_{\pi_2}(\mathscr{A}), \mathrm{Vec}(\boldsymbol{x}_1 \otimes \boldsymbol{x}_2) \otimes \boldsymbol{x}_3 \otimes \cdots \otimes \boldsymbol{x}_\ell \rangle : \boldsymbol{x}_n \in S^{d_n - 1}, n \in [\ell] \right\} \\
&\leq \sup \left\{ \langle \mathrm{Unfold}_{\pi_2}(\mathscr{A}), \boldsymbol{y} \otimes \boldsymbol{x}_3 \cdots \otimes \boldsymbol{x}_\ell \rangle : \boldsymbol{y} \in S^{d_1 \cdot d_2 - 1}, \boldsymbol{x}_n \in S^{d_n - 1}, n = 3, \ldots, \ell \right\} \\
&= \left\|\mathrm{Unfold}_{\pi_2}(\mathscr{A})\right\|_\sigma,
\end{aligned}
$$

where the third line comes from the fact that the set $\{\mathrm{Vec}(\boldsymbol{x}_1 \otimes \boldsymbol{x}_1) : (\boldsymbol{x}_1, \boldsymbol{x}_2) \in S^{d_1 - 1} \times S^{d_2 - 1}\}$ is contained in the set $\{\boldsymbol{y} : \boldsymbol{y} \in S^{d_1 \cdot d_2 - 1}\}$.

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

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 $\mathscr{B} \in \mathbb{R}^{d_1 \times \cdots \times d_\ell}$ be an order-ℓ tensor unfolding of an order-k tensor $\mathscr{A}$ induced by the partition $\{B_1, \ldots, B_\ell\} \in \mathscr{P}_{[k]}$, and let $\mathscr{C}$ be an order-$(\ell - 1)$ tensor unfolding of $\mathscr{B}$ induced by merging blocks B_i and B_j for some $i, j \in [\ell]$. Then,

$$\min \ (d_i, d_j)^{-1/2} \|\mathscr{C}\|_\sigma \le \|\mathscr{B}\|_\sigma \le \|\mathscr{C}\|_\sigma. \quad (10)$$

Proof—The upper bound follows readily from Proposition 4.1. To prove the lower bound, without loss of generality, assume $\mathscr{C}$ corresponds to the merging of blocks $B_{\ell-1}$ and B_ℓ so that $\mathscr{C}$ is a $(d_1, \ldots, d_{\ell-2}, d_{\ell-1} \cdot d_\ell)$-dimensional tensor. Note that $S^{d_1-1} \times \cdots \times S^{d_{\ell-2}-1} \times S^{d_{\ell-1}d_{\ell}-1}$ is a compact set, so the supremum (9) is attained in that set for $\mathscr{C}$. Then, there exists $(\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_{\ell-2}^*, \boldsymbol{y}^*) \in S^{d_1-1} \times \cdots \times S^{d_{\ell-2}-1} \times S^{d_{\ell-1}d_\ell-1}$ such that

$$\|\mathscr{C}\|_\sigma = \langle \mathscr{C}, \boldsymbol{x}_1^* \otimes \cdots \otimes \boldsymbol{x}_{\ell-2}^* \otimes \boldsymbol{y}^* \rangle. \quad (11)$$

Define $\mathscr{C}^* = \mathscr{C}(\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_{\ell-2}^*, \boldsymbol{I}_{d_{\ell-1}d_\ell}) \in \mathbb{R}^{d_{\ell-1}d_\ell}$. By the self duality of the Frobenius norm in $\mathbb{R}^{d_{\ell-1}d_\ell}$, we can rewrite (11) as

$$\|\mathscr{C}\|_\sigma = \sup \ \left\{ \langle \mathscr{C}^*, \boldsymbol{y} \rangle : \boldsymbol{y} \in S^{d_{\ell-1}d_\ell-1} \right\} = \|\mathscr{C}^*\|_F. \quad (12)$$

On the other hand, if we define $\mathrm{Mat}(\mathscr{C}^*) := \mathscr{B}(\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_{\ell-2}^*, \boldsymbol{I}_{\ell-1}, \boldsymbol{I}_\ell) \in \mathbb{R}^{d_{\ell-1} \times d_\ell}$, then $\mathscr{C}^*$ is simply the vectorization of $\mathrm{Mat}(\mathscr{C}^*)$. Hence, by Remark 3.7, we obtain

$$\|\mathscr{C}^*\|_F = \|\mathrm{Mat}(\mathscr{C}^*)\|_F. \quad (13)$$

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

$$\begin{aligned} \|\mathscr{B}\|_\sigma &\ge \sup \ \left\{ \langle \mathscr{B}, \boldsymbol{x}_1^* \otimes \boldsymbol{x}_2^* \otimes \cdots \otimes \boldsymbol{x}_{\ell-2}^* \otimes \boldsymbol{x}_{\ell-1} \otimes \boldsymbol{x}_\ell \rangle : (\boldsymbol{x}_{\ell-1}, \boldsymbol{x}_\ell) \in S^{d_{\ell-1}-1} \times S^{d_\ell-1} \right\} \\ &= \sup \ \left\{ \langle \mathrm{Mat}(\mathscr{C}^*), \boldsymbol{x}_{\ell-1} \otimes \boldsymbol{x}_\ell \rangle : (\boldsymbol{x}_{\ell-1}, \boldsymbol{x}_\ell) \in S^{d_{\ell-1}-1} \times S^{d_\ell-1} \right\} \\ &= \|\mathrm{Mat}(\mathscr{C}^*)\|_\sigma. \end{aligned}$$

$$(14)$$

Recall that $\mathrm{Mat}(\mathscr{C}^*)$ is a matrix of size $d_{\ell-1} \times d_\ell$ meaning [8]

$$\|\mathrm{Mat}(\mathscr{C}^*)\|_\sigma \ge \min \ (d_{\ell-1}, d_\ell)^{-1/2} \|\mathrm{Mat}(\mathscr{C}^*)\|_F. \quad (15)$$

Using (14) in conjunction with (15), (13), and (12), we then obtain

$$\|\mathscr{B}\|_\sigma \ge \min \ (d_{\ell-1}, d_\ell)^{-1/2} \|\mathrm{Mat}(\mathscr{C}^*)\|_F = \min \ (d_{\ell-1}, d_\ell)^{-1/2} \|\mathscr{C}^*\|_F = \min \ (d_{\ell-1}, d_\ell)^{-1/2} \|\mathscr{C}\|_\sigma,$$

which proves the lower bound.

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-ℓ tensor unfolding

$\mathscr{B} = \left(\sum_{i=1}^{d_1} \boldsymbol{e}_{1,i} \otimes \boldsymbol{e}_{2,i} \right) \otimes \mathscr{D} \in \mathbb{R}^{d_1 \times d_2 \times \cdots \times d_\ell}$, where $\{\boldsymbol{e}_{1,i} : i \in [d_1]\}$ is the standard orthonormal basis of $\mathbb{R}^{d_1}$, $\{\boldsymbol{e}_{2,i} : i \in [d_1]\}$ is a set of d_1 standard basis vectors in $\mathbb{R}^{d_2}$, and $\mathscr{D}$ is an arbitrary $(d_3, \ldots, d_\ell)$-dimensional tensor. Assume $d_1 \quad d_2$ 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,

$\mathscr{C} = \left(\sum_{i=1}^{d_1} \mathrm{Vec}(\boldsymbol{e}_{1,i} \otimes \boldsymbol{e}_{2,i}) \right) \otimes \mathscr{D} \in \mathbb{R}^{d_1 \cdot d_2 \times d_3 \times \cdots \times d_\ell}$, which is an order-$(\ell - 1)$ tensor unfolding by merging the first two modes of $\mathscr{B}$ into a single mode. It follows that $\|\mathscr{B}\|_\sigma = \|\mathscr{D}\|_\sigma$ and $\|\mathscr{C}\|_\sigma = \sqrt{d_1} \|\mathscr{D}\|_\sigma$, 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 π_1, $\pi_2 \in \mathscr{P}_{[k]}$ that are not necessarily comparable. To do so, we must first introduce the following definition.

Definition 4.4

Given an order-k tensor $\mathscr{A} \in \mathbb{R}^{d_1 \times \cdots \times d_k}$, we define the map $\dim_{\mathscr{A}} : \mathscr{P}_{[k]} \times \mathscr{P}_{[k]} \to \mathbb{N}_+$ as

$$\dim_{\mathscr{A}}(\pi_1, \pi_2) = \prod_{B \in \pi_1} \max_{B' \in \pi_2} D_{\mathscr{A}}(B, B'), \quad \text{where } D_{\mathscr{A}}(B, B') = \prod_{n \in B \cap B'} d_n,$$

where π_1, $\pi_2 \in \mathscr{P}_{[k]}$.

We label this quantity as $\dim_{\mathscr{A}}(\cdot, \cdot)$ because it involves a product of a subset of the dimensions of $\mathscr{A}$, and $\dim_{\mathscr{A}}(\pi_1, \pi_2) \quad \dim(\mathscr{A})$ with equality only when $\pi_1 = \pi_1 \wedge \pi_2$. Intuitively, $\dim_{\mathscr{A}}(\pi_1, \pi_2)$ reflects the overlap between the unfoldings induced by π_1 and $\pi_1 \wedge \pi_2$. Example 4.6 presents a concrete illustration.

Remark 4.5

We set $D_{\mathscr{A}}(B, B') = 0$ when $B \cap B' = \varnothing$. This does not affect the multiplication in Definition 4.4 because a block of π_1 cannot be disjoint from every block of π_2. Note that $D_{\mathscr{A}}(B, B') = D_{\mathscr{A}}(B', B)$, but in general $\dim_{\mathscr{A}}(\pi_1, \pi_2) \quad \dim_{\mathscr{A}}(\pi_2, \pi_1)$.

Example 4.6

To illustrate the above map, let $\mathscr{A} \in \mathbb{R}^{d_1 \times d_2 \times d_3 \times d_4}$ and consider the partitions $\pi_1 = \{\{1, 2\}, \{3, 4\}\}$ and $\pi_2 = \{\{1, 2, 3\}, \{4\}\}$, for which $\pi_1 \wedge \pi_2 = \{\{1, 2\}, \{3\}, \{4\}\}$. From Definition 4.4,

$$D_{\mathscr{A}}(\{1, 2\}, \{1, 2, 3\}) = d_1 d_2$$

$$D_{\mathscr{A}}(\{3,4\},\{1,2,3\})=d_3$$

$$D_{\mathscr{A}}(\{1,2\},\{4\})=0$$

$$D_{\mathscr{A}}(\{3,4\},\{4\})=d_4.$$

Then,

$$\begin{aligned}
\dim_{\mathscr{A}}(\pi_1,\pi_2)&=\max\ \{D_{\mathscr{A}}(\{1,2\},\{1,2,3\}),D_{\mathscr{A}}(\{1,2\},\{4\})\}\\
&\quad\times\max\ \{D_{\mathscr{A}}(\{3,4\},\{1,2,3\}),D_{\mathscr{A}}(\{3,4\},\{4\})\}\\
&=d_1d_2\max\{d_3,d_4\}.
\end{aligned}$$

Exchanging arguments, we find

$$\begin{aligned}
\dim_{\mathscr{A}}(\pi_2,\pi_1)&=\max\ \{D_{\mathscr{A}}(\{1,2,3\},\{1,2\}),D_{\mathscr{A}}(\{1,2,3\},\{3,4\})\}\\
&\quad\times\max\ \{D_{\mathscr{A}}(\{4\},\{1,2\}),D_{\mathscr{A}}(\{4\},\{3,4\})\}\\
&=d_4\ \max\{d_1d_2,d_3\}.
\end{aligned}$$

Remark 4.7

As in Lemma 4.2, if $\boldsymbol{\pi}_2$ is a one-step coarsening of $\boldsymbol{\pi}_1$ 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

$$D_{\mathscr{A}}(B_i^{\pi_1},B_1^{\pi_2})=d_i,$$

$$D_{\mathscr{A}}(B_j^{\pi_1},B_1^{\pi_2})=d_j.$$

Hence, Lemma 4.2 can be written as

$$\min\ \{D_{\mathscr{A}}(B_i^{\pi_1},B_1^{\pi_2}),D_{\mathscr{A}}(B_j^{\pi_1},B_1^{\pi_2})\}^{-1/2}\|\mathrm{Unfold}_{\pi_2}(\mathscr{A})\|_\sigma\leq\|\mathrm{Unfold}_{\pi_1}(\mathscr{A})\|_\sigma\leq\|\mathrm{Unfold}_{\pi_2}(\mathscr{A})\|_\sigma.$$

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 $\mathscr{A}\in\mathbb{R}^{d_1\times\cdots\times d_k}$ be an arbitrary order-k tensor, and $\boldsymbol{\pi}_1$, $\boldsymbol{\pi}_2$ any two partitions in $\mathscr{P}_{[k]}$. Then,

$$\left[\frac{\dim(\mathscr{A})}{\dim_{\mathscr{A}}(\pi_1, \pi_2)}\right]^{-1/2} \|\text{Unfold}_{\pi_1}(\mathscr{A})\|_\sigma \leq \|\text{Unfold}_{\pi_2}(\mathscr{A})\|_\sigma \leq \left[\frac{\dim(\mathscr{A})}{\dim_{\mathscr{A}}(\pi_2, \pi_1)}\right]^{1/2} \|\text{Unfold}_{\pi_1}(\mathscr{A})\|_\sigma.$$

Remark 4.9

a. Note that π_1 and π_2 need not be comparable.

b. If $d_n = d$ for all $n \in [k]$, then the result reduces to

$$d^{-c_1/2}\|\text{Unfold}_{\pi_1}(\mathscr{A})\|_\sigma \leq \|\text{Unfold}_{\pi_2}(\mathscr{A})\|_\sigma \leq d^{c_2/2}\|\text{Unfold}_{\pi_2}(\mathscr{A})\|_\sigma,$$

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 $\mathscr{A} = \boldsymbol{I}_d \otimes \boldsymbol{I}_d$, and partitions $\boldsymbol{\pi} = \{\{1, 2\}, \{3, 4\}\}$ and $\boldsymbol{\pi}' = \{\{1\}, \{2\}, \{3\}, \{4\}\}$, for which we have $\|\text{Unfold}_{\boldsymbol{\pi}}(\mathscr{A})\|_\sigma = d$ and $\|\text{Unfold}_{\boldsymbol{\pi}'}(\mathscr{A})\|_\sigma = 1$. If $\pi_1 = \boldsymbol{\pi}$ and $\pi_2 = \boldsymbol{\pi}'$, then $c_1 = 2$ and $d^{-c_1/2}\|\text{Unfold}_{\pi_1}(\mathscr{A})\|_\sigma = \|\text{Unfold}_{\pi_2}(\mathscr{A})\|_\sigma$. On the other hand, if $\pi_1 = \boldsymbol{\pi}'$ and $\pi_2 = \boldsymbol{\pi}$, then we have $c_2 = 2$ and $\|\text{Unfold}_{\pi_2}(\mathscr{A})\|_\sigma = d^{c_2/2}\|\text{Unfold}_{\pi_1}(\mathscr{A})\|_\sigma$. 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 π_1 and π_2 in the partition lattice. To do so, we consider $\text{Unfold}_{\pi_1 \wedge \pi_2}(\mathscr{A})$ and compare its spectral norm to that of $\text{Unfold}_{\pi_1}(\mathscr{A})$. Since $\pi_1 \wedge \pi_2 \preceq \pi_1$, from Proposition 4.1 we have

$$\|\text{Unfold}_{\pi_1 \wedge \pi_2}(\mathscr{A})\|_\sigma \leq \|\text{Unfold}_{\pi_1}(\mathscr{A})\|_\sigma. \quad (16)$$

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 π_2 such that

$$D_{\mathscr{A}}(B_1^{\pi_1}, B_{m_1}^{\pi_2}) \geq D_{\mathscr{A}}(B_1^{\pi_1}, B_{m_2}^{\pi_2}) \geq \cdots \geq D_{\mathscr{A}}(B_1^{\pi_1}, B_{m_{\ell_2}}^{\pi_2}), \quad (17)$$

and define $B_{1j} = B_1^{\pi_1} \cap B_{m_j}^{\pi_2}$. Consider a sequence $(\mathscr{T}_1, \mathscr{T}_2, \ldots, \mathscr{T}_{\ell_2-1})$ of unfoldings of $\mathscr{A}$ where $\mathscr{T}_j$, for $1 \leq j \leq \ell_2 - 1$, is obtained from the tensor $\text{Unfold}_{\pi_1 \wedge \pi_2}(\mathscr{A})$ 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

$$\|\text{Unfold}_{\pi_1 \wedge \pi_2}(\mathscr{A})\|_\sigma \geq \min\{D_{\mathscr{A}}(B_1^{\pi_1}, B_{m_1}^{\pi_2}), D_{\mathscr{A}}(B_1^{\pi_1}, B_{m_2}^{\pi_2})\}^{-1/2}\|\mathscr{T}_1\|_\sigma = [D_{\mathscr{A}}(B_1^{\pi_1}, B_{m_2}^{\pi_2})]^{-1/2}\|\mathscr{T}_1\|_\sigma.$$

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

$$\|\mathscr{T}_i\|_\sigma \geq \min\left\{D_{\mathscr{A}}(B_1^{\pi_1}, B_{m_{i+2}}^{\pi_2}), \prod_{j=1}^{i+1} D_{\mathscr{A}}(B_1^{\pi_1}, B_{m_j}^{\pi_2})\right\}^{-1/2} \|\mathscr{T}_{i+1}\|_\sigma = [D_{\mathscr{A}}(B_1^{\pi_1}, B_{m_{i+2}}^{\pi_2})]^{-1/2}\|\mathscr{T}_{i+1}\|_\sigma.$$

Combining these inequalities gives

$$\|\mathrm{Unfold}_{\pi_1 \wedge \pi_2}(\mathscr{A})\|_\sigma \geq \left[\prod_{j=2}^{\ell_2} D_{\mathscr{A}}(B_1^{\pi_1}, B_{m_j}^{\pi_2})\right]^{-1/2} \|\mathscr{T}_{\ell_2-1}\|_\sigma = \left[\frac{\prod_{j=1}^{\ell_2} D_{\mathscr{A}}(B_1^{\pi_1}, B_j^{\pi_2})}{\max_{j\in[\ell_2]}\{D_{\mathscr{A}}(B_1^{\pi_1}, B_j^{\pi_2})\}}\right]^{-1/2} \|\mathscr{T}_{\ell_2-1}\|_\sigma.$$

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

$$\|\mathrm{Unfold}_{\pi_1 \wedge \pi_2}(\mathscr{A})\|_\sigma \geq \prod_{i=1}^{\ell_1}\left[\frac{\prod_{j=1}^{\ell_2} D_{\mathscr{A}}(B_i^{\pi_1}, B_j^{\pi_2})}{\max_{j\in[\ell_2]}\{D_{\mathscr{A}}(B_i^{\pi_1}, B_j^{\pi_2})\}}\right]^{-1/2} \|\mathrm{Unfold}_{\pi_1}(\mathscr{A})\|_\sigma$$

$$= \left[\frac{\prod_{i=1}^{\ell_1}\prod_{j=1}^{\ell_2} D_{\mathscr{A}}(B_i^{\pi_1}, B_j^{\pi_2})}{\prod_{i=1}^{\ell_1}\max_{j\in[\ell_2]}\{D_{\mathscr{A}}(B_i^{\pi_1}, B_j^{\pi_2})\}}\right]^{-1/2} \|\mathrm{Unfold}_{\pi_1}(\mathscr{A})\|_\sigma$$

$$= \left[\frac{\prod_{n\in[k]} d_n}{\dim_{\mathscr{A}}(\pi_1, \pi_2)}\right]^{-1/2} \|\mathrm{Unfold}_{\pi_1}(\mathscr{A})\|_\sigma.$$

Together with (16), this last inequality means

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

By symmetry,

$$\|\mathrm{Unfold}_{\pi_1 \wedge \pi_2}(\mathscr{A})\|_\sigma \leq \|\mathrm{Unfold}_{\pi_2}(\mathscr{A})\|_\sigma \leq \left[\frac{\prod_{n\in[k]} d_n}{\dim_{\mathscr{A}}(\pi_2, \pi_1)}\right]^{1/2} \|\mathrm{Unfold}_{\pi_1 \wedge \pi_2}(\mathscr{A})\|_\sigma. \tag{19}$$

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

We may immediately establish several corollaries of Theorem 4.8.

Corollary 4.10

All order-k tensors $\mathscr{A} \in \mathbb{R}^{d_1 \times \cdots \times d_k}$ satisfy

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

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

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 $\|\mathscr{A}\|_F \quad \dim(\mathscr{A})^{1/2} \|\mathscr{A}\|_\sigma$.

We now give a set of inequalities comparing the spectral norms of unfoldings at level ℓ 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 $\mathscr{A} \in \mathbb{R}^{d \times \cdots \times d}$ be an order-k tensor with the same dimension d in all modes. For all levels $1 \quad \ell \quad k$ and partitions $\pi \in \mathscr{P}_{[k]}^\ell$,

$$d^{-(k-\ell)/2} \max_{\pi \in \mathscr{P}_{[k]}^\ell} \|\mathrm{Unfold}_\pi(\mathscr{A})\|_\sigma \leq \|\mathscr{A}\|_\sigma \leq \min_{\pi \in \mathscr{P}_{[k]}^\ell} \|\mathrm{Unfold}_\pi(\mathscr{A})\|_\sigma. \tag{20}$$

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

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 $\mathscr{A} \in \mathbb{R}^{d \times \cdots \times d}$ be an order-k tensor with the same dimension d in all modes. For all levels $1 \quad \ell \quad k$ and partitions $\pi \in \mathscr{P}_{[k]}^\ell$,

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

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

Corollary 4.14

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

$$d^{-(k-\lceil k/\ell \rceil)/2}\|\mathscr{A}\|_F \leq \min_{\pi \in \mathscr{P}^\ell_{[k]}} \|\mathrm{Unfold}_\pi(\mathscr{A})\|_\sigma.$$

Proof—Note that the minimum of the maximal block sizes across all level-ℓ partitions of

$[k]$ is $\displaystyle\min_{\pi \in \mathscr{P}^\ell_{[k]}} \max_{B \in \pi} |B| = \lceil k/\ell \rceil$, and apply Corollary 4.13.

The above corollaries bound the amount by which norms can vary over a specific level ℓ. They imply that the ratios $\|\mathrm{Unfold}_\pi(\mathscr{A})\|_\sigma/\|\mathscr{A}\|_\sigma$ and $\|\mathrm{Unfold}_\pi(\mathscr{A})\|_\sigma/\|\mathscr{A}\|_F$ fall in the intervals $[1, d^{(k-\ell)/2}]$ and $[d^{-(k-\lceil k/\ell \rceil)/2}, 1]$, respectively. Therefore, in the worst case, $\|\mathrm{Unfold}_\pi(\mathscr{A})\|_\sigma$ only recovers $\|\mathscr{A}\|_\sigma$ or $\|\mathscr{A}\|_F$ 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 ℓ, 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 $\mathscr{A} = \boldsymbol{I}_d \otimes \boldsymbol{I}_d$. Straightforward calculation shows that $\|\mathscr{A}\|_\sigma = 1$. Furthermore, by symmetry, the spectral norm of an unfolding induced by any partition in $\mathscr{P}^2_{[4]}$ or $\mathscr{P}^3_{[4]}$ must fall into one of the following five representative cases:

- For $\pi_1 = \{\{1\}, \{2\}, \{3, 4\}\} \in \mathscr{P}^3_{[4]}, \|\mathrm{Unfold}_{\pi_1}(\mathscr{A})\|_\sigma = \sqrt{d}$.

- For $\pi_2 = \{\{1\}, \{3\}, \{2, 4\}\} \in \mathscr{P}^3_{[4]}, \|\mathrm{Unfold}_{\pi_2}(\mathscr{A})\|_\sigma = 1$.

- For $\pi_3 = \{\{1, 2\}, \{3, 4\}\} \in \mathscr{P}^2_{[4]}, \|\mathrm{Unfold}_{\pi_3}(\mathscr{A})\|_\sigma = d$.

- For $\pi_4 = \{\{1\}, \{2, 3, 4\}\} \in \mathscr{P}^2_{[4]}, \|\mathrm{Unfold}_{\pi_4}(\mathscr{A})\|_\sigma = \sqrt{d}$.

- For $\pi_5 = \{\{1, 3\}, \{2, 4\}\} \in \mathscr{P}^2_{[4]}, \|\mathrm{Unfold}_{\pi_5}(\mathscr{A})\|_\sigma = 1$.

Therefore

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

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

We conclude this section by generalizing Theorem 4.8 to $\mathbb{P}$-norms.

Theorem 4.16 (l^p-norm inequalities)

Let $\mathscr{A} \in \mathbb{R}^{d_1 \times \cdots \times d_k}$ be an arbitrary order-k tensor, and π_1, π_2 any two partitions in $\mathscr{P}_{[k]}$. Then,

a. *For any* $1 \quad p \quad 2$,

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

b. *For any* $2 \quad p \quad \infty$,

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

Proof—We only prove (a) since (b) follows similarly. For any given $1 \quad p \quad 2$, 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, $\dim(\mathscr{A}) = \prod_{n \in [k]} d_n$. Because the total dimension is invariant under any unfolding operation, we have

$$\dim(\mathscr{A})^{\frac{1}{2} - \frac{1}{p}} \|\mathrm{Unfold}_{\pi}(\mathscr{A})\|_\sigma \leq \|\mathrm{Unfold}_{\pi}(\mathscr{A})\|_p \leq \|\mathrm{Unfold}_{\pi}(\mathscr{A})\|_\sigma, \quad (21)$$

for all $\pi \in \mathscr{P}_{[k]}$. Combining (21) with Theorem 4.8 gives the desired results.

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 $\mathscr{A}$. 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 (π-orthogonal decomposable)

Let $\mathscr{A} \in \mathbb{R}^{d_1 \times \cdots \times d_k}$ be an order-k tensor and consider any partition $\pi \in \mathscr{P}_{[k]}$. Then $\mathscr{A}$ is called π-*orthogonal decomposable*, or π-OD, over $\mathbb{R}$ if it admits the decomposition

$$\mathscr{A} = \sum_{n=1}^{r} \lambda_n a_1^{(n)} \otimes \cdots \otimes a_k^{(n)}, \quad (22)$$

where $\lambda_1 \quad \lambda_2 \quad \cdots \quad \lambda_r \quad 0$, 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 π-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, π-OD imposes an orthogonality constraint on every block of singular vectors.

Remark 5.2 ($\mathbf{0}_{[k]}$-OD)

When $\pi = \mathbf{0}_{[k]}$ in Definition 5.1, we obtain the special case of $\mathbf{0}_{[k]}$-OD tensors, which admit the decomposition (22) while satisfying

$$\langle \boldsymbol{a}_i^{(n)}, \boldsymbol{a}_i^{(m)} \rangle = \delta_{nm},$$

for all $i \in [k]$ and all $n, m \in [r]$.

The definition of $\mathbf{0}_{[k]}$-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 $\mathbf{0}_{[k]}$-OD tensor $\mathscr{A}$ 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 $\mathscr{A} \in \mathbb{R}^{d_1 \times \cdots \times d_k}$.

    a.   *Let $\pi_1, \pi_2 \in \mathscr{P}_{[k]}$ and $\pi_1 \preceq \pi_2$. If $\mathscr{A}$ is π_1-OD, then $\mathscr{A}$ is π_2-OD.*

    b.   *Let $\pi \in \mathscr{P}_{[k]}$ and $\pi \neq \mathbf{1}_{[k]}$. If $\mathscr{A}$ is π-OD, then*

$$\|\mathrm{Unfold}_\pi(\mathscr{A})\|_\sigma = \lambda_1. \quad (23)$$

Proof—Part (a): For any two finite sets of vectors $\{\boldsymbol{x}_i\}$ and $\{\boldsymbol{y}_i\}$ for which $\boldsymbol{x}_i, \boldsymbol{y}_i \in \mathbb{R}^{d_i}$, we have

$$\left\langle \otimes_{i \in B} \boldsymbol{x}_i, \otimes_{j \in B} \boldsymbol{y}_j \right\rangle = \prod_{i \in B} \langle \boldsymbol{x}_i, \boldsymbol{y}_i \rangle, \quad (24)$$

for all $B \subset [k]$. Suppose $B \in \pi_2$. If $\pi_1 \preceq \pi_2$, then there exist subsets $C_1, \ldots, C_m \in \pi_1$ such that $C_1 \cup \cdots \cup C_m = B$. So,

$$\langle \otimes_{i \in B} \boldsymbol{x}_i, \otimes_{j \in B} \boldsymbol{y}_j \rangle = \prod_{i \in B} \langle \boldsymbol{x}_i, \boldsymbol{y}_i \rangle = \prod_{a=1}^{m} \prod_{i \in C_a} \langle \boldsymbol{x}_i, \boldsymbol{y}_i \rangle = \prod_{a=1}^{m} \langle \otimes_{i \in C_a} \boldsymbol{x}_i, \otimes_{j \in C_a} \boldsymbol{y}_j \rangle,$$

which implies that $\mathscr{A}$ is π_2-OD if $\mathscr{A}$ is π_1-OD.

Part (b): Suppose $\pi \in \mathscr{P}^{\ell}_{[k]}$ is of the form $\pi = \{B^{\pi}_i : i \in [\ell]\}$. Note that $\boldsymbol{\pi} \neq \mathbf{1}_{[k]}$ implies $\ell \geq 2$. Letting $\tau = \{B^{\pi}_1, (B^{\pi}_1)^c\}$ where $(B^{\pi}_1)^c$ denotes the complement of B^{π}_1 with respect to $[k]$, we have $\tau \in \mathscr{P}^2_{[k]}$ and $\boldsymbol{\pi} \leq \boldsymbol{\tau}$. By Lemma 5.3(a), $\mathscr{A}$ is τ-OD, so $\mathscr{A}$ admits a decomposition of the form

$$\mathscr{A} = \sum_{n=1}^{r} \lambda_n \boldsymbol{a}^{(n)}_1 \otimes \cdots \otimes \boldsymbol{a}^{(n)}_k, \qquad (25)$$

where

$$\left\langle \otimes_{i \in B_1} \boldsymbol{a}^{(n)}_i, \otimes_{j \in B_1} \boldsymbol{a}^{(m)}_j \right\rangle = \delta_{nm} \quad \text{and} \quad \left\langle \otimes_{i \in (B^{\pi}_1)^c} \boldsymbol{a}^{(n)}_i, \otimes_{j \in (B^{\pi}_1)^c} \boldsymbol{a}^{(m)}_j \right\rangle = \delta_{nm} \qquad (26)$$

for all $n, m \in [r]$.

Now define $\boldsymbol{x}_n = \mathrm{Vec}(\otimes_{i \in B^{\pi}_1} \boldsymbol{a}^{(n)}_i)$ and $\boldsymbol{y}_n = \mathrm{Vec}(\otimes_{i \in (B^{\pi}_1)^c} \boldsymbol{a}^{(n)}_i)$ for all $n \in [r]$. By (26), both $\{\boldsymbol{x}_n\}$ and $\{\boldsymbol{y}_n\}$ are sets of orthonormal vectors. By the definition of $\mathrm{Unfold}_{\tau}(\mathscr{A})$, (25) implies

$$\mathrm{Unfold}_{\tau}(\mathscr{A}) = \sum_{n=1}^{r} \lambda_n \boldsymbol{x}_n \boldsymbol{y}^T_n,$$

which is simply the matrix SVD of $\mathrm{Unfold}_{\tau}(\mathscr{A})$. Hence $\|\mathrm{Unfold}_{\tau}(\mathscr{A})\|_{\sigma} = \lambda_1$. Using monotonicity (c.f., Proposition 4.1), we have

$$\|\mathrm{Unfold}_{\pi}(\mathscr{A})\|_{\sigma} \leq \|\mathrm{Unfold}_{\tau}(\mathscr{A})\|_{\sigma} = \lambda_1. \qquad (27)$$

Conversely, by the definition of the spectral norm, we have

$$\begin{aligned}
\|\mathrm{Unfold}_{\pi}(\mathscr{A})\|_{\sigma} &\geq \left\langle \mathrm{Unfold}_{\pi}(\mathscr{A}), \mathrm{Vec}\left(\otimes_{i \in B^{\pi}_1} \boldsymbol{a}^{(1)}_i\right) \otimes \cdots \otimes \mathrm{Vec}\left(\otimes_{i \in B^{\pi}_{\ell}} \boldsymbol{a}^{(1)}_i\right) \right\rangle \\
&= \sum_{n=1}^{r} \lambda_n \left\langle \mathrm{Vec}\left(\otimes_{i \in B^{\pi}_1} \boldsymbol{a}^{(n)}_i\right) \otimes \cdots \otimes \mathrm{Vec}\left(\otimes_{i \in B^{\pi}_{\ell}} \boldsymbol{a}^{(n)}_i\right), \right. \\
&\quad \left. \mathrm{Vec}\left(\otimes_{i \in B^{\pi}_{\ell}} \boldsymbol{a}^{(1)}_i\right) \otimes \cdots \otimes \mathrm{Vec}\left(\otimes_{i \in B^{\pi}_{\ell}} \boldsymbol{a}^{(1)}_i\right) \right\rangle \\
&= \sum_{n=1}^{r} \lambda_n \prod_{j \in [\ell]} \left\langle \mathrm{Vec}\left(\otimes_{i \in B^{\pi}_j} \boldsymbol{a}^{(n)}_i\right), \mathrm{Vec}\left(\otimes_{i \in B^{\pi}_j} \boldsymbol{a}^{(1)}_i\right) \right\rangle \\
&= \sum_{n=1}^{r} \lambda_n \delta_{n,1} = \lambda_1
\end{aligned} \qquad (28)$$

where the third line comes from (24) and the last line follows from the fact that $\mathscr{A}$ is $\boldsymbol{\pi}$-OD. Combining (27) and (28), we conclude $\|\mathrm{Unfold}_{\boldsymbol{\pi}}(\mathscr{A})\|_{\sigma} = \lambda_1$.

Remark 5.4

The condition $\boldsymbol{\pi} \neq \mathbf{1}_{[k]}$ in Lemma 5.3 is needed for (23) to hold. In fact, consider a 2×2 matrix $A = 2\boldsymbol{e}_1 \otimes \boldsymbol{e}_1 + \boldsymbol{e}_1 \otimes \boldsymbol{e}_2$, where $\{\boldsymbol{e}_i, i \in [2]\}$ is the canonical basis of $\mathbb{R}^2$. Then, A is $\mathbf{1}_{[k]}$-OD with $\lambda_1 = 2$, but $\left\|\mathrm{Unfold}_{\mathbf{1}_{[k]}}(\mathscr{A})\right\|_{\sigma} = \|\mathscr{A}\|_F = \sqrt{5} \neq 2$.

Theorem 5.5 (Norm equality on upper cones)

If $\mathscr{A}$ is $\boldsymbol{\pi}$-OD, then for any partition in the upper cone $U_{\boldsymbol{\pi}} = \{\tau \in \mathscr{P}_{[k]} : \boldsymbol{\pi} \leq \tau < \mathbf{1}_{[k]}\}$ of $\boldsymbol{\pi}$, we have

$$\|\mathrm{Unfold}_{\tau}(\mathscr{A})\|_{\sigma} = \|\mathrm{Unfold}_{\boldsymbol{\pi}}(\mathscr{A})\|_{\sigma}.$$

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

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

Corollary 5.6

If $\mathscr{A}$ is $\mathbf{0}_{[k]}$-OD, then for all partitions $\boldsymbol{\pi} \neq \mathbf{1}_{[k]}$, we have

$$\|\mathrm{Unfold}_{\boldsymbol{\pi}}(\mathscr{A})\|_{\sigma} = \|\mathscr{A}\|_{\sigma}.$$

Corollary 5.6 implies that for $\mathbf{0}_{[k]}$-OD tensors, the operator norm is invariant under any unfolding operations except vectorization. Lastly, $\boldsymbol{\pi}_1, \boldsymbol{\pi}_2 \in U_{\boldsymbol{\pi}_1 \wedge \boldsymbol{\pi}_2}$ implies the following corollary:

Corollary 5.7

Let $\boldsymbol{\pi}_1, \boldsymbol{\pi}_2 \in \mathscr{P}_{[k]}$. If $\mathscr{A}$ is $(\boldsymbol{\pi}_1 \wedge \boldsymbol{\pi}_2)$-OD, then

$$\|\mathrm{Unfold}_{\boldsymbol{\pi}_1}(\mathscr{A})\|_{\sigma} = \|\mathrm{Unfold}_{\boldsymbol{\pi}_2}(\mathscr{A})\|_{\sigma}.$$

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 $\dim(\mathscr{A})$ plays a key role in the worst-case inflation factor and is a manifestation of the curse of dimensionality. Specifically, $\dim(\mathscr{A})$ 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, π-OD tensors permit much tighter bounds in which some unfoldings, including certain matricizations, leave the operator norm relatively unaffected. In practice, π-OD tensors, or those within a small neighborhood around π-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 ℓ 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 π-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-ℓ 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.

Acknowledgments

References

1. Anandkumar A, Ge R, Hsu D, Kakade SM, Telgarsky M. Tensor decompositions for learning latent variable models. The Journal of Machine Learning Research. 2014; 15(1):2773–2832.

2. Cardoso J-F. 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). 1990:2655–2658.

3. De Lathauwer L, De Moor B, Vandewalle J. A multilinear singular value decomposition. SIAM Journal on Matrix Analysis and Applications. 2000; 21(4):1253–1278.

4. Friedland S, Lim L-H. Computational complexity of tensor nuclear norm. SIAM Journal on Optimization. 2016; 4(26):2378–2393.

5. Friedland S, Lim L-H. Nuclear norm of higher-order tensors. Mathematics of Computation. (to appear).

6. Hillar CJ, Lim L-H. Most tensor problems are NP-hard. Journal of the ACM. 2013; 60(6) Article 45.

7. Hoff PD. Multilinear tensor regression for longitudinal relational data. The Annals of Applied Statistics. 2015; 9(3):1169–1193. [PubMed: 27458495]

8. Horn, RA., Johnson, CR. Matrix Analysis. Cambridge University Press; 2012.

9. Hu S. Relations of the nuclear norm of a tensor and its matrix flattenings. Linear Algebra and its Applications. 2015; 478:188–199.

10. Kroonenberg, PM. Three-Mode Principal Component Analysis: Theory and Applications. Vol. 2. DSWO Press; 1983.

11. Kuleshov V, Chaganty A, Liang P. Tensor factorization via matrix factorization. Proceedings of the 18th International Conference on Artificial Intelligence and Statistics (AISTATS). 2015:507–516.

12. Lim L-H. Singular values and eigenvalues of tensors: a variational approach. Proceedings of the IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP). 2005:129–132.

13. Liu J, Musialski P, Wonka P, Ye J. Tensor completion for estimating missing values in visual data. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2013; 35(1):208–220. [PubMed: 22271823]

14. McCullagh P. Tensor notation and cumulants of polynomials. Biometrika. 1984; 71(3):461–476.

15. Mu C, Huang B, Wright J, Goldfarb D. Square deal: Lower bounds and improved relaxations for tensor recovery. Proceedings of the 31st International Conference on Machine Learning (ICML). 2014:73–81.

16. Nickel M, Tresp V, Kriegel H-P. A three-way model for collective learning on multi-relational data. Proceedings of the 28th International Conference on Machine Learning (ICML). 2011:809–816.

17. Omberg L, Golub GH, Alter O. A tensor higher-order singular value decomposition for integrative analysis of DNA microarray data from different studies. Proceedings of the National Academy of Sciences. 2007; 104(47):18371–18376.

18. Qi L. Eigenvalues of a real supersymmetric tensor. Journal of Symbolic Computation. 2005; 40(6): 1302–1324.

19. Ragnarsson S, Van Loan CF. Block tensor unfoldings. SIAM Journal on Matrix Analysis and Applications. 2012; 33(1):149–169.

20. Richard E, Montanari A. A statistical model for tensor PCA. Advances in Neural Information Processing Systems (NIPS). 2014; 27:2897–2905.

21. Robeva E. Orthogonal decomposition of symmetric tensors. SIAM Journal on Matrix Analysis and Applications. 2016; 37:86–102.

22. Sun J, Tao D, Faloutsos C. Beyond streams and graphs: dynamic tensor analysis. Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2006:374–383.

23. Tomioka R, Hayashi K, Kashima H. Estimation of low-rank tensors via convex optimization. 2010 arXiv:1010.0789.

24. Tomioka R, Suzuki T. Spectral norm of random tensors. 2014 arXiv:1407.1870.

25. Vasilescu MAO, Terzopoulos D. Multilinear analysis of image ensembles: Tensorfaces. Proceedings of European Conference on Computer Vision (ECCV). 2002:447–460.

26. Wang M, Song YS. Orthogonal tensor decompositions via two-mode higher-order SVD (HOSVD). 2016 arXiv:1612.03839.

27. Yu Y, Cheng H, Zhang X. Approximate low-rank tensor learning. 7th NIPS Workshop on Optimization for Machine Learning. 2014

28. Yuan M, Zhang C-H. On tensor completion via nuclear norm minimization. Foundations of Computational Mathematics. 2015:1–38.

29. Zhao L, Zaki MJ. Tricluster: an effective algorithm for mining coherent clusters in 3D microarray data. Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data. 2005:694–705.
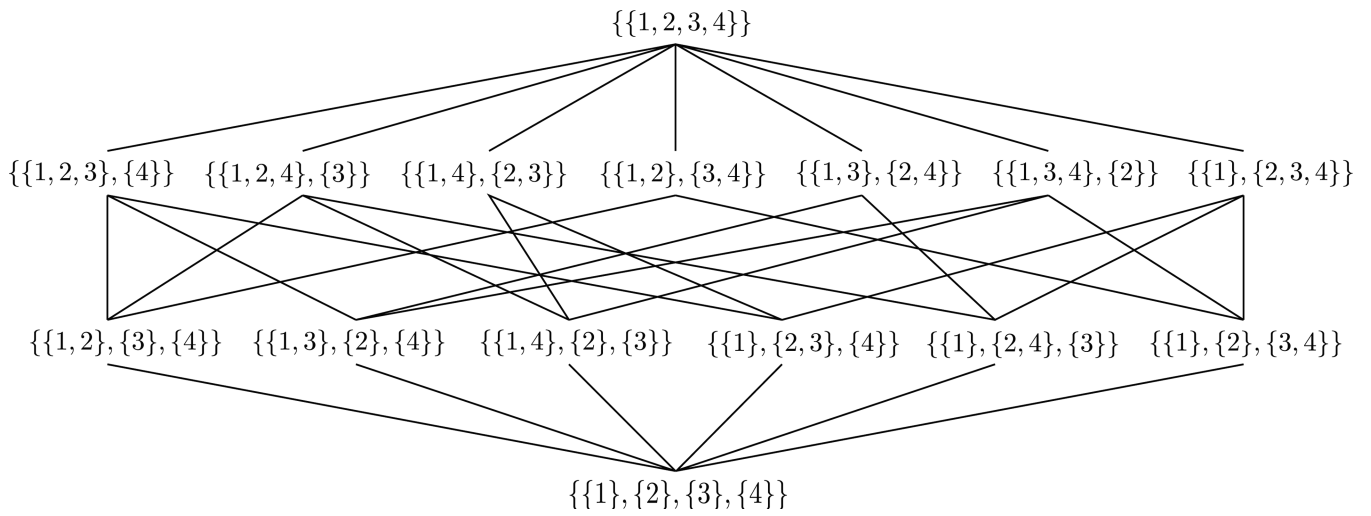
Figure 1.
The partition lattice $\mathscr{P}_{[4]}$.