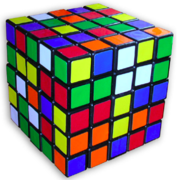
Higher-Order Tensors and Their General Unfoldings

Miaoyan Wang

Joint work with Khanh Dao Duc, Jonathan Fischer, Yun S. Song

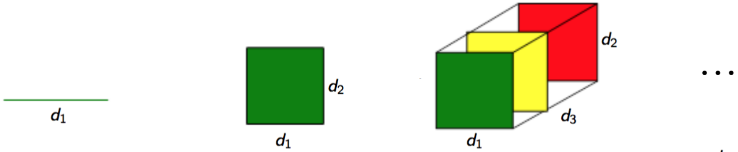
Departments of Statistics and EECS, UC Berkeley
Department of Mathematics, University of Pennsylvania

Aug 18, 2016



Higher-Order Tensors

- Tensors are generalizations of scalars, vectors and matrices:



$a \in \mathbb{R}^{d_1}$ $\quad$ $\mathbf{A} \in \mathbb{R}^{d_1 \times d_2}$ $\quad$ $\mathcal{A} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ $\quad$ $\mathcal{A} \in \mathbb{R}^{d_1 \times \cdots \times d_k}$

- Why tensors?
  - provides a natural representation for multiway data
  - allows more flexible and powerful statistical models
    $\Rightarrow$ e.g. higher-order cumulants in latent variable models

Tensor Norm

The main differences between usual matrices and higher-order tensors come from the transition from $k = 2$ to $k = 3$:

Computational complexity

Most higher-order tensor problems are NP-hard [Hillar & Lim 2013].

p-norm [Lim 2005]

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

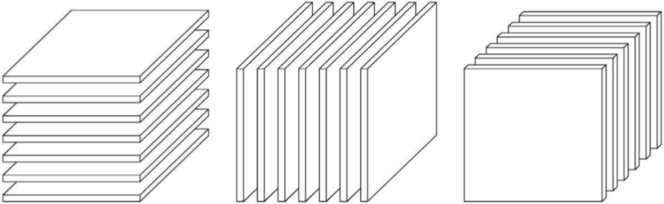
$$\|\mathcal{A}\|_p = \max_{\|\mathbf{x}_n\|_p = 1, \mathbf{x}_n \in \mathbb{R}^{d_n}, n = 1, \ldots, k} \langle \mathcal{A}, \mathbf{x}_1 \otimes \cdots \otimes \mathbf{x}_k \rangle,$$

where $\|\mathbf{x}_n\|_p$ denotes the vector l^p-norm of $\mathbf{x}_n$. The special case of $p = 2$ is called the spectral norm.

Problem: Can we give a computable bound of $\|\mathcal{A}\|_p$?

Matrixization

Tensor can be viewed as a collection of slices (matrices):

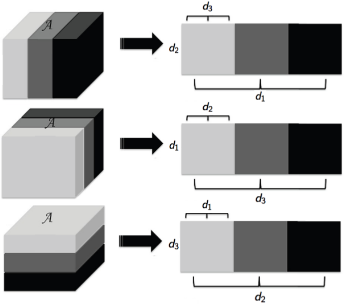


(a) Horizontal slices: $\mathbf{X}_{i::}$ (b) Lateral slices: $\mathbf{X}_{:j:}$ (c) Frontal slices: $\mathbf{X}_{::k}$ (or $\mathbf{X}_k$)

(Kolda & Bader, 2009)

General Unfoldings

- **Matricization**. Rearrange the slices of the tensor in different directions (or modes) into a matrix.



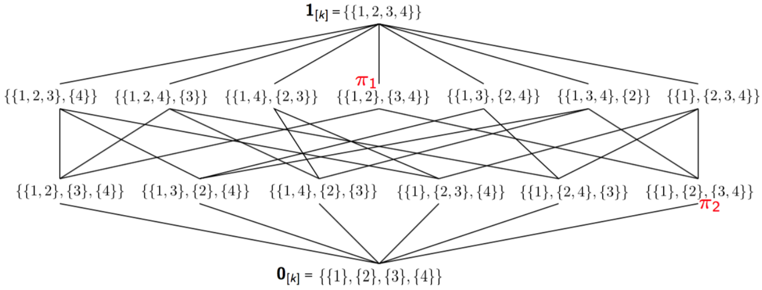
$\mathrm{Unfold}_\pi(\mathcal{A})$	Partition $\pi \in \mathcal{P}_{[3]}$
$\in \mathbb{R}^{d_2 \times d_1 d_3}$	$\pi = \{\{2\}, \{1, 3\}\}$
$\in \mathbb{R}^{d_1 \times d_2 d_3}$	$\pi = \{\{1\}, \{2, 3\}\}$
$\in \mathbb{R}^{d_3 \times d_1 d_2}$	$\pi = \{\{3\}, \{1, 2\}\}$

- **Unfolding**. We generalize this notion by considering all possible unfoldings of an order-k tensor, each of which can be viewed as being induced by a certain partition of $[k] := \{1, \ldots, k\}$.

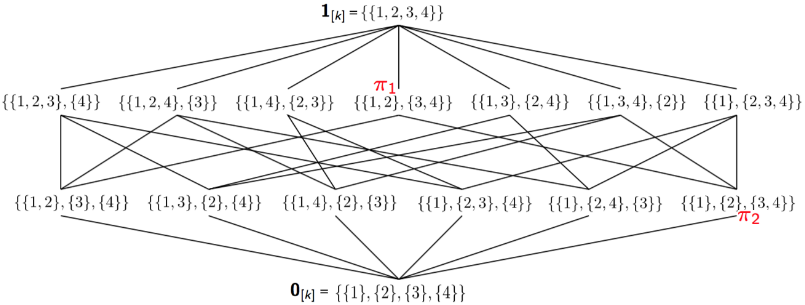
Partition Lattice $\mathcal{P}_{[k]}$

- For any $k \in \mathbb{N}_+$, a **partition** π 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 $\mathcal{P}_{[k]}$.
- A partition $\pi_1 \in \mathcal{P}_{[k]}$ is called a *refinement* of $\pi_2 \in \mathcal{P}_{[k]}$ if each block of π_1 is a subset of some block of π_2. This relationship defines a **partial order**, expressed as $\pi_1 \leq \pi_2$.

Example on $\mathcal{P}_{[4]}$:



- All possible tensor unfoldings $\overset{\text{1-to-1}}{\longleftrightarrow}$ the set of partitions of $[k]$, e.g., $\mathbf{0}_{[k]} \leftrightarrow \mathcal{A}$, $\mathbf{1}_{[k]} \leftrightarrow \text{Vec}(\mathcal{A})$.
- $\text{Unfold}_{\pi}(\mathcal{A})$ denotes the tensor unfolding induced by partition $\pi \in \mathcal{P}_{[k]}$.
- Some facts:
  - \# of possible unfoldings: B_k (Bell number).
    $B_1 = 1, B_2 = 2, B_3 = 5, B_4 = 15, B_5 = 52, \ldots$
  - \# of possible order-ℓ ($1 \le \ell \le k$) unfoldings: $S(k, \ell)$ (Stirling number of the second kind).

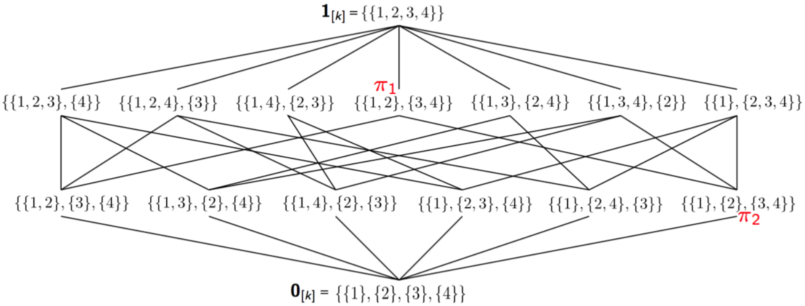


How does the spectral norm change upon unfoldings?

Example (on $\mathcal{P}_{[4]}$)

Given $\mathcal{A} \in \mathbb{R}^{d_1 \times \cdots \times d_4}$, we consider $\pi_1 = \{\{1,2\},\{3,4\}\}$ and $\pi_2 = \{\{1\},\{2\},\{3,4\}\}$.

- Spectral norm preserves the partial order on partitions:
  $\pi_2 \leq \pi_1$, $\|\mathrm{Unfold}_{\pi_2}(\mathcal{A})\|_2 \leq \|\mathrm{Unfold}_{\pi_1}(\mathcal{A})\|_2$.
- **One-step** refinement: $\|\mathrm{Unfold}_{\pi_1}(\mathcal{A})\|_2 \leq \min(\sqrt{d_1}, \sqrt{d_2})\, \|\mathrm{Unfold}_{\pi_2}(\mathcal{A})\|_2$.



Norm Inequalities Between Any Two Tensor Unfoldings

More generally, we can compare the spectral norms of tensor unfoldings induced by **any** two partitions $\pi_1, \pi_2 \in \mathcal{P}_{[k]}$.

Spectral norm inequalities

Let $\mathcal{A} \in \mathbb{R}^{d\times\cdots\times d}$ be an order-k tensor with the same dimension d in all modes. Then

$$d^{-c_1/2} \left\|\text{Unfold}_{\pi_1}(\mathcal{A})\right\|_2 \leq \left\|\text{Unfold}_{\pi_2}(\mathcal{A})\right\|_2 \leq d^{c_2/2} \left\|\text{Unfold}_{\pi_1}(\mathcal{A})\right\|_2,$$

where $c_1 = (k - \sum_{B\in\pi_1} \max_{B'\in\pi_2} |B\cap B'|)$, $c_2 = (k - \sum_{B\in\pi_2} \max_{B'\in\pi_1} |B\cap B'|)$, and $|B\cap B'|$ denotes the number of elements in the block $B\cap B'$.

- π_1 and π_2 need not be comparable.
- Proof sketch: consider the sequences of partitions $\pi_1 \geq \cdots \geq (\pi_1 \wedge \pi_2)$ and $\pi_2 \geq \cdots \geq (\pi_1 \wedge \pi_2)$, where $\pi_1 \wedge \pi_2$ is the *greatest lower bound* of π_1 and π_2, defined by $\pi_1 \wedge \pi_2 := \sup\{\pi \in \mathcal{P}_{[k]} : \pi \leq \pi_1, \pi \leq \pi_2\}$.

- See our paper for the general l^p-norm inequalities that allow unequal dimension in each mode:

  Wang, M., Dao Duc, K., Fischer, J., and Song, Y. S. Operator Norm Inequalities between Tensor Unfoldings on the Partition Lattice, Preprint. arXiv:1603.05621.

- **Application.** Recall that computing $\|\mathcal{A}\|_2$ is hard. What if we use the matrix norm to approximate the tensor norm?
  Taking $\pi_2 = \mathbf{1}_{[k]}$ in the spectral norm inequalities gives:

Bottom-up inequality

Let $\mathcal{A} \in \mathbb{R}^{d\times\cdots\times d}$ be an order-k tensor with the same dimension d in all modes, and let $\mathcal{P}_{[k]}^{\ell} \subset \mathcal{P}_{[k]}$ denote the set of partitions that have exactly ℓ blocks. Then, for all $1 \leq \ell \leq k$ and partitions $\pi \in \mathcal{P}_{[k]}^{\ell}$,

$$d^{-(k-\ell)/2} \max_{\pi\in\mathcal{P}_{[k]}^{\ell}} \|\mathrm{Unfold}_{\pi}(\mathcal{A})\|_2 \leq \|\mathcal{A}\|_2 \leq \min_{\pi\in\mathcal{P}_{[k]}^{\ell}} \|\mathrm{Unfold}_{\pi}(\mathcal{A})\|_2 .$$

Other Useful Corollaries

Top-down inequalities

1. $d^{-(k-\max_{i\in[\ell]}|B_i^\pi|)/2} \|\mathcal{A}\|_F \leq \|\mathrm{Unfold}_\pi(\mathcal{A})\|_2 \leq \|\mathcal{A}\|_F$.
2. $d^{-(k-\lceil k/\ell\rceil)/2} \|\mathcal{A}\|_F \leq \min_{\pi\in\mathcal{P}_{[k]}^\ell} \|\mathrm{Unfold}_\pi(\mathcal{A})\|_2 \leq \max_{\pi\in\mathcal{P}_{[k]}^\ell} \|\mathrm{Unfold}_\pi(\mathcal{A})\|_2 \leq \|\mathcal{A}\|_F$.

Frobenius norm vs. Spectral norm

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

$$\|\mathcal{A}\|_F \leq \left[\frac{\dim(\mathcal{A})}{\max_{n\in[k]} d_n}\right]^{1/2} \|\mathcal{A}\|_2 ,$$

where $\dim(\mathcal{A}) = \prod_{n\in[k]} d_n$ denotes the total dimension of the tensor.

This bound improves over the recent result found by Friedland and Lim [Lemma 5.1, 2016], namely $\|\mathcal{A}\|_F \leq \dim(\mathcal{A})^{1/2} \|\mathcal{A}\|_2$.

Specially-Structured Tensors

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

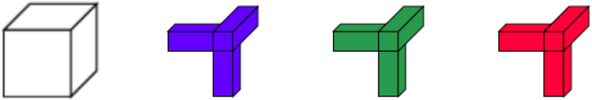
$$\mathcal{A} = \sum_{n=1}^{r} \lambda_n \mathbf{a}_1^{(n)} \otimes \cdots \otimes \mathbf{a}_k^{(n)},$$

where $\lambda_n \in \mathbb{R}_+, n \in [r]$, and the set of vectors $\{\mathbf{a}_i^{(n)} \in \mathbb{R}^{d_i} : i \in [k],\ n \in [r]\}$ satisfies

$$\langle \otimes_{i \in B} \mathbf{a}_i^{(n)},\ \otimes_{i \in B} \mathbf{a}_i^{(m)} \rangle = \delta_{nm},$$

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

Example: a symmetric tensor $\mathcal{A} \in \mathbb{R}^{d \times d \times d}$ is called $\mathbf{0}_{[k]}$-OD if it admits the following decomposition:



$$\mathcal{A} \quad = \lambda_1 (\mathbf{u}^{(1)})^{\otimes 3} + \lambda_2 (\mathbf{u}^{(2)})^{\otimes 3} + \lambda_3 (\mathbf{u}^{(3)})^{\otimes 3}$$

where $\{\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \mathbf{u}^{(3)}\}$ is a set of orthonormal vectors.

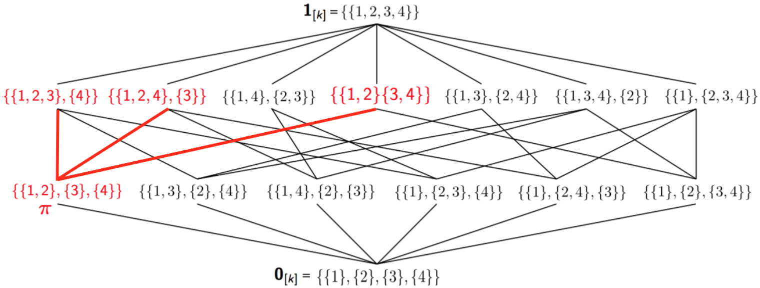
π-Orthogonal Decomposable Tensors

Norm equality on upper cones

If $\mathcal{A}$ is π-OD, then for any partition τ in the *upper cone* of π, i.e. $\tau \in U_\pi := \{\tau \in \mathcal{P}_{[k]} : \pi \leq \tau < \mathbf{1}_{[k]}\}$, we have

$$\|\mathrm{Unfold}_\tau(\mathcal{A})\|_2 = \|\mathrm{Unfold}_\pi(\mathcal{A})\|_2 .$$

Example: If $\mathcal{A}$ is π-OD tensor where $\pi = \{\{1,2\},\{3\},\{4\}\}$, then the spectral norm is invariant under the following tensor unfoldings: $\{\{1,2\},\{3\},\{4\}\}$, $\{\{1,2,3\},\{4\}\}$, $\{\{1,2,4\},\{3\}\}$, and $\{\{1,2\},\{3,4\}\}$.



References

1. C. J. Hillar and L.-H. Lim, Most tensor problems are NP-hard, Journal of the ACM 60 (2013), no. 6, 45.
2. 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.
3. T. G. Kolda and B. W. Bader, Tensor decompositions and applications, SIAM Review 51 (2009), no. 3, 455-500.
4. S. Friedland and L.-H. Lim. The computational complexity of duality, arXiv:1601.07629 (2016).
5. Wang, M., Dao Duc, K., Fischer, J., and Song, Y.S. Operator norm inequalities between tensor unfoldings on the partition lattice, arXiv:1603.05621 (2016).
6. Wang, M. and Song, Y.S. Orthogonal decomposition of symmetric tensors via two-mode higher-order SVD, manuscript.