Bounds on tensor norms via tensor partitions

Zhening Li

Department of Mathematics

University of Portsmouth

June 15, 2017 LAW'17

Tensors (hypermatrices)

Consider tensor products of finite-dimensional Euclidean spaces

- A scalar (lower case letter), $x \in \mathbb{R}$, is a tensor of order zero;
- A vector (boldface lower letter), $\boldsymbol{x} = (x_i) \in \mathbb{R}^n$, is a tensor of order one;
- A matrix (capital letter), $X = (x_{ij}) \in \mathbb{R}^{n_1 \times n_2}$, is a tensor of order two;
- A tensor of order d, $\mathcal{X} = (x_{i_1 i_2 \dots i_d}) \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$.

Tensors (hypermatrices)

Consider tensor products of finite-dimensional Euclidean spaces

- A scalar (lower case letter), $x \in \mathbb{R}$, is a tensor of order zero;
- A vector (boldface lower letter), $\boldsymbol{x} = (x_i) \in \mathbb{R}^n$, is a tensor of order one;
- A matrix (capital letter), $X = (x_{ij}) \in \mathbb{R}^{n_1 \times n_2}$, is a tensor of order two;
- A tensor of order d, $\mathcal{X} = (x_{i_1 i_2 \dots i_d}) \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_d}$.

Some facts

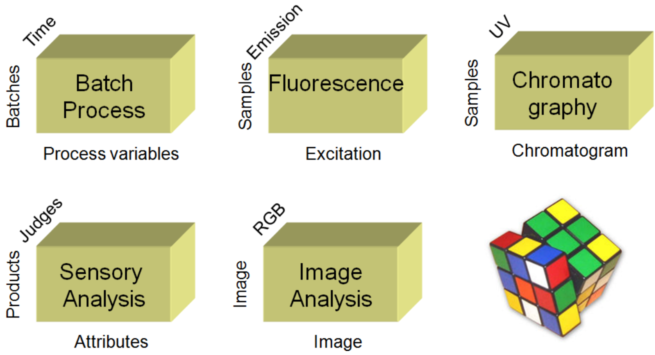
- Tensors can be one-to-one represented by multilinear forms. For instance, $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ defines a trilinear form $F : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \mathbb{R}^{n_3} \to \mathbb{R}$

$$F(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_d} a_{ijk} x_i y_j z_k.$$

- Many matrix problems are easy, while corresponding tensor problems can be very difficult (Hillar, Lim 2013).

Multiway (tensor) data problems

- Multiway empirical data analysis in psychometrics and chemometrics
- High order statistics and independent component analysis
- Algebraic properties of tensors
- ......



Some tensor basics

- $\mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ is a tensor space of dimensions $n_1 \times n_2 \times \cdots \times n_d$ and order d. Real numbers, vector spaces, and matrix spaces are tensor spaces of $d = 0$, $d = 1$, and $d = 2$, respectively.
- For $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$, the Frobenius inner product

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

- The induced Frobenius norm (Hilbert-Schmidt norm)

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

  If $d = 1$ the Frobenius norm reduces to the Euclidean norm of a vector.
- A rank-one tensor $\mathcal{T}$, also called a simple tensor, can be written as outer products of vectors

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

Tensor norms and their complexity

- Hölder p-norm $(1 \le p \le \infty)$

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

Tensor norms and their complexity

- Hölder p-norm $(1 \le p \le \infty)$

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

- Spectral norm: NP-hard to compute (He, L., Zhang 2010)

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

Tensor norms and their complexity

- Hölder p-norm ($1 \le p \le \infty$)

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

- Spectral norm: NP-hard to compute (He, L., Zhang 2010)

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

- Nuclear norm: NP-hard to compute (Friedland, Lim 2016)

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

Tensor norms and their complexity

- Hölder p-norm ($1 \le p \le \infty$)

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

- Spectral norm: NP-hard to compute (He, L., Zhang 2010)

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

- Nuclear norm: NP-hard to compute (Friedland, Lim 2016)

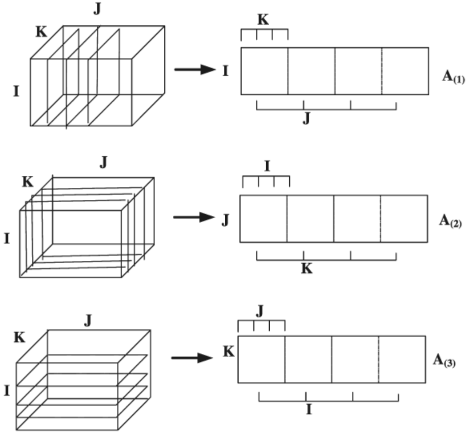
$$\|\mathcal{A}\|_* := \min \left\{ \sum_{i=1}^{r} |\lambda_i| : \mathcal{A} = \sum_{i=1}^{r} \lambda_i \boldsymbol{x}_i^1 \otimes \boldsymbol{x}_i^2 \otimes \cdots \otimes \boldsymbol{x}_i^d, \|\boldsymbol{x}_i^k\|_2 = 1, r \in \mathbb{N} \right\}$$

- The nuclear norm is the dual norm of the spectral norm:

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

A common approach to approximate

Matricization: also matricisation, matricizing, unfolding, or flattening, is the operation that turns a tensor (a multi-way array) into a matrix (a two-way array). It can be regarded as a generalization of the concept of vectorization.



A common approach to approximate

Matricization: also matricisation, matricizing, unfolding, or flattening, is the operation that turns a tensor (a multi-way array) into a matrix (a two-way array). It can be regarded as a generalization of the concept of vectorization.

- Most matrix norms are easy to compute
  - The matrix spectral norm is the largest singular value of a matrix
  - The matrix nuclear norm is the sum of all singular values of a matrix

A common approach to approximate

Matricization: also matricisation, matricizing, unfolding, or flattening, is the operation that turns a tensor (a multi-way array) into a matrix (a two-way array). It can be regarded as a generalization of the concept of vectorization.

- Most matrix norms are easy to compute
  - The matrix spectral norm is the largest singular value of a matrix
  - The matrix nuclear norm is the sum of all singular values of a matrix
- The spectral norm of a matricized tensor is an upper bound of the spectral norm of the tensor (He, L., Zhang 2010)

$$\|\mathrm{Mat}\,(\mathcal{T})\|_\sigma \geq \|\mathcal{T}\|_\sigma$$

A common approach to approximate

Matricization: also matricisation, matricizing, unfolding, or flattening, is the operation that turns a tensor (a multi-way array) into a matrix (a two-way array). It can be regarded as a generalization of the concept of vectorization.

- Most matrix norms are easy to compute
  - The matrix spectral norm is the largest singular value of a matrix
  - The matrix nuclear norm is the sum of all singular values of a matrix
- The spectral norm of a matricized tensor is an upper bound of the spectral norm of the tensor (He, L., Zhang 2010)

$$\|\mathrm{Mat}\,(\mathcal{T})\|_\sigma \geq \|\mathcal{T}\|_\sigma$$

- The nuclear norm of a matricized tensor is a lower bound of the nuclear norm of the tensor (Hu 2015)

$$\|\mathrm{Mat}\,(\mathcal{T})\|_* \leq \|\mathcal{T}\|_*$$

A new perspective from tensor partitions

Definition (Tensor partition)

A partition $\{\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_m\}$ is called a tensor partition of a tensor $\mathcal{T}$, if

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

Modal partition

Given a tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$, the indices of its mode k can be partitioned into r_k nonempty sets, i.e.,

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

Definition (Modal partition)

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

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

Regular partition

Given a tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$, a mode-$k$ tensor cut, cuts the tensor $\mathcal{T}$ at mode k into two subtensors $\mathcal{T}_1$ and $\mathcal{T}_2$, denoted by

$$\mathcal{T} = \mathcal{T}_1 \vee_k \mathcal{T}_2,$$

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

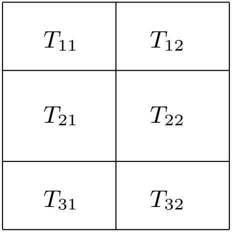
Definition (Regular partition)

$\{\mathcal{T}\}$ is called the 1-regular partition of a tensor $\mathcal{T}$. For $m \in \mathbb{N}$ with $m \geq 2$, a partition $\{\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_m\}$ is called an m-regular partition of a tensor $\mathcal{T}$, if there exist two tensors $\mathcal{A}_1, \mathcal{A}_2$ and an ℓ with $1 \leq \ell \leq m-1$, such that

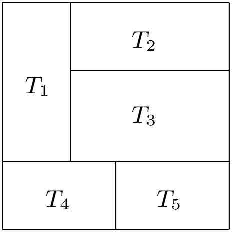
- $\mathcal{T} = \mathcal{A}_1 \vee_k \mathcal{A}_2$ for some $1 \leq k \leq d$,
- $\{\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_\ell\}$ is an ℓ-regular partition of $\mathcal{A}_1$, and
- $\{\mathcal{T}_{\ell+1}, \mathcal{T}_{\ell+2}, \ldots, \mathcal{T}_m\}$ is an $(m-\ell)$-regular partition of $\mathcal{A}_2$.

Three types of tensor partitions

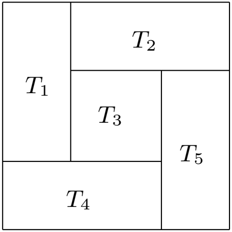
- A modal partition is a special type of regular partition, and a regular partition is a special type of tensor partition.
- For any first order tensor (vector), the three partitions are the same. This is not true for a second or higher order tensor.
- A subtensor $\mathcal{T}_j$ in a partition of a tensor $\mathcal{T} = \{\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_m\}$ may not have the same order of the original tensor $\mathcal{T}$.



A modal partition



A regular partition



An irregular partition

Bounds of tensor norms by tensor partitions

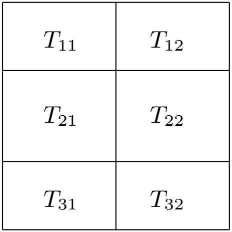
Theorem

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

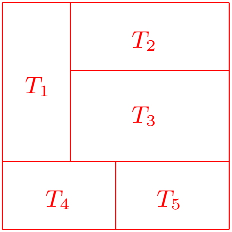
$$\|(\|\mathcal{T}_1\|_\sigma, \|\mathcal{T}_2\|_\sigma, \ldots, \|\mathcal{T}_m\|_\sigma)\|_\infty \le \|\mathcal{T}\|_\sigma \le \|(\|\mathcal{T}_1\|_\sigma, \|\mathcal{T}_2\|_\sigma, \ldots, \|\mathcal{T}_m\|_\sigma)\|_2 \,,$$

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

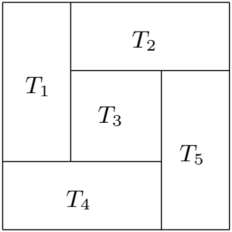
where $\|\cdot\|_p$ *is the* L_p *norm of a vector for* $1 \le p \le \infty$.



A modal partition



A regular partition



An irregular partition

Behind the main result

Some key observations:

- For any regular partition, the tensor $\mathcal{T}$ can be cut sequentially by applying a mode-k tensor cut $m-1$ times
- The L_p norm of a vector has certain additive property for $1 \le p \le \infty$, i.e., if $\boldsymbol{x} = \boldsymbol{x}^1 \vee \boldsymbol{x}^2 \in \mathbb{R}^{n_1+n_2}$ with $\boldsymbol{x}^1 \in \mathbb{R}^{n_1}$ and $\boldsymbol{x}^2 \in \mathbb{R}^{n_2}$, then

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

Behind the main result

Some key observations:

- For any regular partition, the tensor $\mathcal{T}$ can be cut sequentially by applying a mode-k tensor cut $m-1$ times
- The L_p norm of a vector has certain additive property for $1 \le p \le \infty$, i.e., if $\boldsymbol{x} = \boldsymbol{x}^1 \vee \boldsymbol{x}^2 \in \mathbb{R}^{n_1+n_2}$ with $\boldsymbol{x}^1 \in \mathbb{R}^{n_1}$ and $\boldsymbol{x}^2 \in \mathbb{R}^{n_2}$, then

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

The three-step proof:

- If $\mathcal{T} = \mathcal{A} \vee \mathcal{B}$, then $\max\{\|\mathcal{A}\|_\sigma, \|\mathcal{B}\|_\sigma\} \le \|\mathcal{T}\|_\sigma \le \sqrt{\|\mathcal{A}\|_\sigma{}^2 + \|\mathcal{B}\|_\sigma{}^2}$
- If $\mathcal{T} = \mathcal{A} \vee \mathcal{B}$, then $\sqrt{\|\mathcal{A}\|_*{}^2 + \|\mathcal{B}\|_*{}^2} \le \|\mathcal{T}\|_* \le \|\mathcal{A}\|_* + \|\mathcal{B}\|_*$ from a dual norm point of view
- Mathematical induction

Some consequence of the main result

- If $\mathcal{T}$ is partitioned entry wisely into $\prod_{k=1}^{d} n_k$ number of scalars

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

Some consequence of the main result

- If $\mathcal{T}$ is partitioned entry wisely into $\prod_{k=1}^{d} n_k$ number of scalars

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

- If $\mathcal{T}$ is partitioned into mode-k vector fibers, say $\{\boldsymbol{t}_i \in \mathbb{R}^{n_k} : i = 1, 2, \dots, m\}$ where $m = \prod_{1 \le j \le d,\, j \ne k} n_j$

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

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

Some consequence of the main result

- If $\mathcal{T}$ is partitioned entry wisely into $\prod_{k=1}^d n_k$ number of scalars

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

- If $\mathcal{T}$ is partitioned into mode-k vector fibers, say $\{\boldsymbol{t}_i \in \mathbb{R}^{n_k} : i = 1, 2, \dots, m\}$ where $m = \prod_{1\le j\le d,\, j\ne k} n_j$

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

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

- Any regular partition $\{\mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_m\}$ of a rank-one tensor $\mathcal{T}$ satisfies

$$\|(\|\mathcal{T}_1\|_\sigma, \|\mathcal{T}_2\|_\sigma, \dots, \|\mathcal{T}_m\|_\sigma)\|_2 = \|\mathcal{T}\|_\sigma = \|\mathcal{T}\|_* = \|(\|\mathcal{T}_1\|_*, \|\mathcal{T}_2\|_*, \dots, \|\mathcal{T}_m\|_*)\|_2$$

Some consequence of the main result

- If $\mathcal{T}$ is partitioned entry wisely into $\prod_{k=1}^{d} n_k$ number of scalars

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

- If $\mathcal{T}$ is partitioned into mode-k vector fibers, say $\{\boldsymbol{t}_i \in \mathbb{R}^{n_k} : i = 1, 2, \dots, m\}$ where $m = \prod_{1 \le j \le d,\, j \ne k} n_j$

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

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

- Any regular partition $\{\mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_m\}$ of a rank-one tensor $\mathcal{T}$ satisfies

$$\|(\|\mathcal{T}_1\|_\sigma, \|\mathcal{T}_2\|_\sigma, \dots, \|\mathcal{T}_m\|_\sigma)\|_2 = \|\mathcal{T}\|_\sigma = \|\mathcal{T}\|_* = \|(\|\mathcal{T}_1\|_*, \|\mathcal{T}_2\|_*, \dots, \|\mathcal{T}_m\|_*)\|_2$$

- All the bounds are sharp in general, in the sense that for any given tensor space and one of the four inequalities, there exists a tensor in that space such that the inequality becomes an equality.

Approximating tensor norms

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

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

Approximating tensor norms

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

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

An immediate fact of the main result:

Corollary

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

Approximation bounds by matrix slices

A tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ with $n_1 \le n_2 \le \cdots \le n_d$ is cut into matrix slices

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

Approximation bounds by matrix slices

A tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ with $n_1 \le n_2 \le \cdots \le n_d$ is cut into matrix slices

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

- The spectral norm of a tensor $\mathcal{T}$ can be approximated by

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

with an approximation bound $\sqrt{\prod_{k=1}^{d-2} n_k}$.

Approximation bounds by matrix slices

A tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ with $n_1 \le n_2 \le \cdots \le n_d$ is cut into matrix slices

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

- The spectral norm of a tensor $\mathcal{T}$ can be approximated by

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

  with an approximation bound $\sqrt{\prod_{k=1}^{d-2} n_k}$.

- The nuclear norm of $\mathcal{T}$ can be approximated by

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

  with an approximation bound $\sqrt{\prod_{k=1}^{d-2} n_k}$ (the best bound so far).

Approximation bounds by matrix slices

A tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ with $n_1 \le n_2 \le \cdots \le n_d$ is cut into matrix slices

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

- The spectral norm of a tensor $\mathcal{T}$ can be approximated by

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

  with an approximation bound $\sqrt{\prod_{k=1}^{d-2} n_k}$.

- The nuclear norm of $\mathcal{T}$ can be approximated by

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

  with an approximation bound $\sqrt{\prod_{k=1}^{d-2} n_k}$ (the best bound so far).

- For the case $d = 3$, both bounds are $\sqrt{n_1}$.

An algorithm to approximate the tensor spectral norm

Algorithm

Find a rank-one tensor that approximates the spectral norm of a given tensor from below, with an approximation bound $\sqrt{\prod_{k=1}^{d-2} n_k}$.

- *Input: A tensor* $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ *with* $n_1 \leq n_2 \leq \cdots \leq n_d$

1. *Compute*

$$(j_1, j_2, \ldots, j_{d-2}) = \arg \max_{1 \leq i_k \leq n_k,\, k=1,2,\ldots,d-2} \|T_{i_1 i_2 \ldots i_{d-2}}\|_\sigma$$

2. *Find the left singular vector* $\boldsymbol{x}$ *and the right singular vector* $\boldsymbol{y}$ *corresponding to the largest singular value of the matrix* $T_{j_1 j_2 \ldots j_{d-2}}$

3. *Compute* $\mathcal{X} = \boldsymbol{e}_{j_1} \otimes \boldsymbol{e}_{j_2} \otimes \cdots \otimes \boldsymbol{e}_{j_{d-2}} \otimes \boldsymbol{x} \otimes \boldsymbol{y}$ *where* $\boldsymbol{e}_j$ *is the vector whose* j*-th entry is one and other entries are zeros*

- *Output: A rank-one tensor* $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ *with* $\|\mathcal{X}\|_2 = 1$

An algorithm to approximate the tensor nuclear norm

Algorithm

Find a rank-one decomposition of a given tensor that approximates its nuclear norm from above, with an approximation bound $\sqrt{\prod_{k=1}^{d-2} n_k}$.

- *Input: A tensor* $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ *with* $n_1 \le n_2 \le \cdots \le n_d$

1. *Compute SVD for the matrix*

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

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

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

- *Output: A rank-one decomposition of* $\mathcal{T}$

Can we extend the main result?

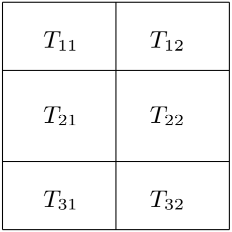
Theorem

If $\{\mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_m\}$ *is a regular partition of a tensor* $\mathcal{T}$*, then*

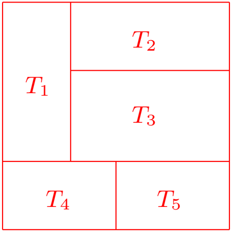
$$\left\|(\|\mathcal{T}_1\|_\sigma, \|\mathcal{T}_2\|_\sigma, \dots, \|\mathcal{T}_m\|_\sigma)\right\|_\infty \le \|\mathcal{T}\|_\sigma \le \left\|(\|\mathcal{T}_1\|_\sigma, \|\mathcal{T}_2\|_\sigma, \dots, \|\mathcal{T}_m\|_\sigma)\right\|_2 ,$$

$$\left\|(\|\mathcal{T}_1\|_*, \|\mathcal{T}_2\|_*, \dots, \|\mathcal{T}_m\|_*)\right\|_2 \le \|\mathcal{T}\|_* \le \left\|(\|\mathcal{T}_1\|_*, \|\mathcal{T}_2\|_*, \dots, \|\mathcal{T}_m\|_*)\right\|_1 ,$$

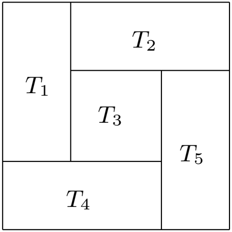
where $\|\cdot\|_p$ *is the* L_p *norm of a vector for* $1 \le p \le \infty$*.*



A modal partition



A regular partition



An irregular partition

Can we extend the main result?

Theorem

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

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

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

where $\|\cdot\|_p$ *is the* L_p *norm of a vector for* $1 \le p \le \infty$.

Questions of interest:

- How about other tensor norms?
- How about irregular (general) tensor partitions?
- How about modal (specified) partitions?

Bounds on the spectral p-norm and the nuclear p-norm

- Spectral p-norm $(1 \le p \le \infty)$

$$\|\mathcal{A}\|_{p_\sigma} := \max\left\{\left\langle \mathcal{A}, \boldsymbol{x}^1 \otimes \boldsymbol{x}^2 \otimes \cdots \otimes \boldsymbol{x}^d \right\rangle : \|\boldsymbol{x}^k\|_p = 1,\ 1 \le k \le d\right\}$$

Bounds on the spectral p-norm and the nuclear p-norm

- Spectral p-norm $(1 \le p \le \infty)$

$$\|\mathcal{A}\|_{p_\sigma} := \max\left\{\left\langle \mathcal{A}, \boldsymbol{x}^1 \otimes \boldsymbol{x}^2 \otimes \cdots \otimes \boldsymbol{x}^d \right\rangle : \|\boldsymbol{x}^k\|_p = 1,\ 1 \le k \le d\right\}$$

- Nuclear p-norm (the dual norm)

$$\|\mathcal{A}\|_{p_*} := \min\left\{\sum_{i=1}^r |\lambda_i| : \mathcal{A} = \sum_{i=1}^r \lambda_i \boldsymbol{x}_i^1 \otimes \boldsymbol{x}_i^2 \otimes \cdots \otimes \boldsymbol{x}_i^d, \|\boldsymbol{x}_i^k\|_p = 1, r \in \mathbb{N}\right\}$$

Bounds on the spectral p-norm and the nuclear p-norm

- Spectral p-norm $(1 \le p \le \infty)$

$$\|\mathcal{A}\|_{p_\sigma} := \max\left\{\left\langle \mathcal{A}, \boldsymbol{x}^1 \otimes \boldsymbol{x}^2 \otimes \cdots \otimes \boldsymbol{x}^d \right\rangle : \|\boldsymbol{x}^k\|_p = 1,\ 1 \le k \le d\right\}$$

- Nuclear p-norm (the dual norm)

$$\|\mathcal{A}\|_{p_*} := \min\left\{\sum_{i=1}^r |\lambda_i| : \mathcal{A} = \sum_{i=1}^r \lambda_i \boldsymbol{x}_i^1 \otimes \boldsymbol{x}_i^2 \otimes \cdots \otimes \boldsymbol{x}_i^d, \|\boldsymbol{x}_i^k\|_p = 1, r \in \mathbb{N}\right\}$$

- $\|\mathcal{A}\|_{1_\sigma} = \|\mathcal{A}\|_\infty$ and $\|\mathcal{A}\|_{\infty_*} = \|\mathcal{A}\|_1$

Bounds on the spectral p-norm and the nuclear p-norm

- Spectral p-norm $(1 \le p \le \infty)$

$$\|\mathcal{A}\|_{p_\sigma} := \max \left\{ \left\langle \mathcal{A}, \boldsymbol{x}^1 \otimes \boldsymbol{x}^2 \otimes \cdots \otimes \boldsymbol{x}^d \right\rangle : \|\boldsymbol{x}^k\|_p = 1,\ 1 \le k \le d \right\}$$

- Nuclear p-norm (the dual norm)

$$\|\mathcal{A}\|_{p_*} := \min \left\{ \sum_{i=1}^r |\lambda_i| : \mathcal{A} = \sum_{i=1}^r \lambda_i \boldsymbol{x}_i^1 \otimes \boldsymbol{x}_i^2 \otimes \cdots \otimes \boldsymbol{x}_i^d, \|\boldsymbol{x}_i^k\|_p = 1, r \in \mathbb{N} \right\}$$

- $\|\mathcal{A}\|_{1_\sigma} = \|\mathcal{A}\|_\infty$ and $\|\mathcal{A}\|_{\infty_*} = \|\mathcal{A}\|_1$

Theorem

If $\{\mathcal{T}_1, \mathcal{T}_2, \dots, \mathcal{T}_m\}$ is a tensor partition of a tensor $\mathcal{T}$, then

$$\|(\|\mathcal{T}_1\|_{p_\sigma}, \|\mathcal{T}_2\|_{p_\sigma}, \dots, \|\mathcal{T}_m\|_{p_\sigma})\|_\infty \le \|\mathcal{T}\|_{p_\sigma} \le \|(\|\mathcal{T}_1\|_{p_\sigma}, \|\mathcal{T}_2\|_{p_\sigma} \dots, \|\mathcal{T}_m\|_{p_\sigma})\|_q$$

$$\|(\|\mathcal{T}_1\|_{p_*}, \|\mathcal{T}_2\|_{p_*}, \dots, \|\mathcal{T}_m\|_{p_*})\|_q \le \|\mathcal{T}\|_{p_*} \le \|(\|\mathcal{T}_1\|_{p_*}, \|\mathcal{T}_2\|_{p_*}, \dots, \|\mathcal{T}_m\|_{p_*})\|_1$$

where $1 \le p \le \infty$ and $\frac{1}{p} + \frac{1}{q} = 1$.

THANK YOU!