Ranks and Nuclear Norms of Tensors

Harm Derksen

University of Michigan

August 2, 2013

Tensor Rank

$V^{(1)}, V^{(2)}, \dots, V^{(d)}$ finite dimensional Hilbert spaces and let

$$V = V^{(1)} \otimes V^{(2)} \otimes \cdots \otimes V^{(d)}.$$

A *pure tensor* is a tensor of the form

$$v = v^{(1)} \otimes v^{(2)} \otimes \cdots \otimes v^{(d)} \in V.$$

Definition

The *rank* of a tensor $T \in V$ is the smallest nonnegative integer r such that we can write T as a sum of r pure tensors.

It is difficult to determine the rank of a tensor.

Example: Matrix Multiplication

$V = \mathrm{Mat}_{n,n}(\mathbb{C}) \otimes \mathrm{Mat}_{n,n}(\mathbb{C}) \otimes \mathrm{Mat}_{n,n}(\mathbb{C})$

$$T_n = \sum_{i,j,k=1}^n e_{i,j} \otimes e_{j,k} \otimes e_{k,i}$$

$\mathrm{rank}(T_n)$ is the number of multiplications needed to multiply two $n \times n$ matrices. Clearly $\mathrm{rank}(T_n) \leq n^3$.

Example: Matrix Multiplication

$V = \mathrm{Mat}_{n,n}(\mathbb{C}) \otimes \mathrm{Mat}_{n,n}(\mathbb{C}) \otimes \mathrm{Mat}_{n,n}(\mathbb{C})$

$$T_n = \sum_{i,j,k=1}^{n} e_{i,j} \otimes e_{j,k} \otimes e_{k,i}$$

$\mathrm{rank}(T_n)$ is the number of multiplications needed to multiply two $n \times n$ matrices. Clearly $\mathrm{rank}(T_n) \leq n^3$.

Theorem

If $\mathrm{rank}(T_m) \leq k$, *then* $\mathrm{rank}(T_n) = O(n^{\log_m(k)})$.

Example:Matrix Multiplication

Theorem (Strassen 1969)

$\mathrm{rank}(T_2) \leq 7$, *so* $\mathrm{rank}(T_n) = O(n^{\log_2(7)}) = O(n^{2.8073\ldots})$.

Theorem (Williams 2012)

$\mathrm{rank}(T_n) = O(n^{2.3727})$

Theorem (Landsberg 2012)

$$\mathrm{rank}(T_n) \geq 3n^2 - 4n^{3/2} + n.$$

Motivation/Digression: Convex Relaxation

Suppose $A \in \mathrm{Mat}_{m,n}(\mathbb{C})$, $b \in \mathbb{C}^n$ and you want to solve a linear system $Ax = b$ where $x \in \mathbb{C}^m$ is a sparse vector.
Want to minimize

$$\|x\|_0 = \#\{i \mid x_i \neq 0\}.$$

But, $\|\cdot\|_0$ is not convex and this optimization problem is difficult.

Instead, minimize the convex function $\|x\|_1$. Often this will also give the optimal solution for $\|\cdot\|_0$.

Motivation/Digression: Convex Relaxation

Suppose $A \in \mathrm{Mat}_{m,n}(\mathbb{C})$, $b \in \mathbb{C}^n$ and you want to solve a linear system $Ax = b$ where $x \in \mathbb{C}^m$ is a sparse vector.
Want to minimize

$$\|x\|_0 = \#\{i \mid x_i \neq 0\}.$$

But, $\|\cdot\|_0$ is not convex and this optimization problem is difficult.

Instead, minimize the convex function $\|x\|_1$. Often this will also give the optimal solution for $\|\cdot\|_0$.

The tensor rank is not convex, so we can use convex relaxation.

Nuclear Norm

Definition

The nuclear norm $\|T\|_\star$ is the smallest value of $\sum_{i=1}^r \|v_i\|$ where $T = \sum_{i=1}^r v_i$ and $v_1, \ldots, v_r$ are pure tensors.

Nuclear Norm

Definition

The nuclear norm $\|T\|_\star$ is the smallest value of $\sum_{i=1}^r \|v_i\|$ where $T = \sum_{i=1}^r v_i$ and $v_1, \ldots, v_r$ are pure tensors.

Theorem (D.)

For the matrix multiplication tensor we have

$$\|T_n\|_\star = n^3.$$

(proof sketch later)

Spectral Norm

Definition

The spectral norm is defined by

$$[T] = \max\{|\langle T, v\rangle| \mid v \text{ pure tensor with } \|v\| = 1\}.$$

The spectral norm is *dual* to the nuclear norm, in particular

$$|\langle T, S\rangle| \leq \|T\|_\star [S]$$

for all tensors S, T.

Example: Determinant Tensor

Consider the tensor

$$D_n = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) e_{\sigma(1)} \otimes e_{\sigma(2)} \otimes \cdots \otimes e_{\sigma(n)} \in \mathbb{C}^n \otimes \cdots \otimes \mathbb{C}^n.$$

Clearly $\operatorname{rank}(D_n) \leq n!$, but actually $\operatorname{rank}(D_n) \leq (\frac{5}{6})^{\lfloor n/3 \rfloor} n!$.

Example: Determinant Tensor

Consider the tensor

$$D_n = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) e_{\sigma(1)} \otimes e_{\sigma(2)} \otimes \cdots \otimes e_{\sigma(n)} \in \mathbb{C}^n \otimes \cdots \otimes \mathbb{C}^n.$$

Clearly $\operatorname{rank}(D_n) \leq n!$, but actually $\operatorname{rank}(D_n) \leq (\frac{5}{6})^{\lfloor n/3 \rfloor} n!$.

$$[D_n] = \max\{\det(v_1 v_2 \cdots v_n) \mid \|v_1\| = \cdots = \|v_n\| = 1\} = 1$$

by Hadamard's inequality.

$$\|D_n\|_\star = \|D_n\|_\star [D_n] \geq \langle D_n, D_n \rangle = n!, \text{ so}$$

Theorem (D.)

$$\|D_n\|_\star = n!$$

Example: Permanent Tensor

$$P_n = \sum_{\sigma \in S_n} e_{\sigma(1)} \otimes e_{\sigma(2)} \otimes \cdots \otimes e_{\sigma(n)} \in \mathbb{C}^n \otimes \cdots \otimes \mathbb{C}^n.$$

$$[P_n] = \max\{\mathrm{perm}(v_1 v_2 \cdots v_n) \mid \|v_1\| = \cdots = \|v_n\| = 1\}$$

Example: Permanent Tensor

$$P_n = \sum_{\sigma \in S_n} e_{\sigma(1)} \otimes e_{\sigma(2)} \otimes \cdots \otimes e_{\sigma(n)} \in \mathbb{C}^n \otimes \cdots \otimes \mathbb{C}^n.$$

$$[P_n] = \max\{\mathrm{perm}(v_1 v_2 \cdots v_n) \mid \|v_1\| = \cdots = \|v_n\| = 1\}$$

Theorem (Carlen, Lieb and Moss, 2006)

$$\max\{\mathrm{perm}(v_1 v_2 \cdots v_n) \mid \|v_1\| = \cdots = \|v_n\| = 1\} = n!/n^{n/2}$$

$$\|P_n\|_\star = \frac{n^{n/2}}{n!}\|P_n\|_\star [P_n] \geq \frac{n^{n/2}}{n!}\langle P_n, P_n\rangle = n^{n/2}.$$

Example: Permanent Tensor

Theorem (Glynn 2010)

$$P_n = \frac{1}{2^{n-1}} \sum_{\delta} (\textstyle\sum_{i=1}^n \delta_i e_i) \otimes \cdots \otimes (\textstyle\sum_{i=1}^n \delta_i e_i)$$

where δ *runs over* $\{1\} \times \{-1, 1\}^{n-1}$.

Example: Permanent Tensor

Theorem (Glynn 2010)

$$P_n = \frac{1}{2^{n-1}} \sum_{\delta} (\textstyle\sum_{i=1}^n \delta_i e_i) \otimes \cdots \otimes (\sum_{i=1}^n \delta_i e_i)$$

where δ *runs over* $\{1\} \times \{-1, 1\}^{n-1}$.

In particular, $\operatorname{rank}(P_n) \leq 2^{n-1}$ and $\|P_n\|_\star \leq n^{n/2}$, so

Theorem (D.)

$$\|P_n\|_\star = n^{n/2}$$

t-Orthogonality

Definition

Tensors $v_1, v_2, \ldots, v_r$ are t-orthogonal if

$$\sum_{i=1}^{r} |\langle v_i, w\rangle|^{2/t} \leq 1$$

for every pure tensor w with $\|w\| = 1$.

t-Orthogonality

Definition

Tensors $v_1, v_2, \ldots, v_r$ are t-orthogonal if

$$\sum_{i=1}^{r} |\langle v_i, w\rangle|^{2/t} \leq 1$$

for every pure tensor w with $\|w\| = 1$.

1-orthogonal $\Leftrightarrow$ orthogonal

If $t > s$ then t-orthogonal $\Rightarrow$ s-orthogonal.

t-Orthogonality

Definition

Tensors $v_1, v_2, \ldots, v_r$ are t-orthogonal if

$$\sum_{i=1}^{r} |\langle v_i, w\rangle|^{2/t} \leq 1$$

for every pure tensor w with $\|w\| = 1$.

1-orthogonal $\Leftrightarrow$ orthogonal

If $t > s$ then t-orthogonal $\Rightarrow$ s-orthogonal.

Theorem (D.)

If $v_1, \ldots, v_r \in V$ are t-orthogonal, then $r \leq \dim(V)^{1/t}$.

Horizontal and Vertical Tensor Product

Theorem (“horizontal tensor product”, D.)

If $v_1, \ldots, v_r$ are t-orthogonal, and $w_1, \ldots, w_r$ are s-orthogonal, then $v_1 \otimes w_1, \ldots, v_r \otimes w_r$ are $(s+t)$-orthogonal.

Horizontal and Vertical Tensor Product

Theorem ("horizontal tensor product", D.)

If $v_1, \dots, v_r$ are t-orthogonal, and $w_1, \dots, w_r$ are s-orthogonal, then $v_1 \otimes w_1, \dots, v_r \otimes w_r$ are $(s+t)$-orthogonal.

If $V = V^{(1)} \otimes \cdots \otimes V^{(d)}$ and $W = W^{(1)} \otimes \cdots \otimes W^{(d)}$, then

$$V \boxtimes W := (V^{(1)} \otimes W^{(1)}) \otimes \cdots \otimes (V^{(d)} \otimes W^{(d)}).$$

Theorem ("vertical tensor product", D.)

If $v_1, v_2, \dots, v_r \in V$ and $w_1, \dots, w_s \in W$ are t-orthogonal, then $\{v_i \boxtimes w_j \mid 1 \leq i \leq r, 1 \leq j \leq s\}$ are t-orthogonal.

The Diagonal Singular Value Decomposition

Definition

Suppose that $(\star) : T = \sum_{i=1}^{r} \lambda_i v_i$ such that $\lambda_1 \geq \cdots \geq \lambda_r > 0$ and $v_1, \ldots, v_r$ are 2-orthogonal pure tensors of unit length, then $(\star)$ is called a *diagonal singular value decomposition* of T (DSVD).

If $d = 2$ (tensor product of 2 spaces) then the DSVD is the usual singular value decomposition. For $d > 2$, the DSVD is different from the *Higher Order Singular Value Decomposition* defined by De Lathauer, De Moor, and Vandewalle. Not every tensor has a DSVD.

The Diagonal Singular Value Decomposition

Definition

Suppose that $(\star) : T = \sum_{i=1}^{r} \lambda_i v_i$ such that $\lambda_1 \geq \cdots \geq \lambda_r > 0$ and $v_1, \ldots, v_r$ are 2-orthogonal pure tensors of unit length, then $(\star)$ is called a *diagonal singular value decomposition* of T (DSVD).

If $d = 2$ (tensor product of 2 spaces) then the DSVD is the usual singular value decomposition. For $d > 2$, the DSVD is different from the *Higher Order Singular Value Decomposition* defined by De Lathauer, De Moor, and Vandewalle. Not every tensor has a DSVD.

The Diagonal Singular Value Decomposition

Theorem (D.)

If T has a DSVD then

$$\|T\|_\star = \sum_i \lambda_i, \quad \|T\| = \sqrt{\sum_i \lambda_i^2}, \quad [T] = \lambda_1$$

The Diagonal Singular Value Decomposition

Theorem (D.)

If T has a DSVD then

$$\|T\|_\star = \sum_i \lambda_i, \quad \|T\| = \sqrt{\sum_i \lambda_i^2}, \quad [T] = \lambda_1$$

Theorem (D.)

If $\lambda_1 > \lambda_2 > \cdots > \lambda_r$ then the DSVD is unique.

Theorem (D.)

If $v_1, \ldots, v_r$ are t-orthogonal with $t > 2$, then the DSVD is unique.

Example: Matrix Multiplication Tensor

$e_1, \dots, e_n \in \mathbb{C}^n$ are orthogonal
$e_1 \otimes e_1, \dots, e_n \otimes e_n \in \mathbb{C}^n \otimes \mathbb{C}^n$ are 2-orthogonal

Example: Matrix Multiplication Tensor

$e_1, \ldots, e_n \in \mathbb{C}^n$ are orthogonal
$e_1 \otimes e_1, \ldots, e_n \otimes e_n \in \mathbb{C}^n \otimes \mathbb{C}^n$ are 2-orthogonal

$e_1 \otimes e_1 \otimes 1, \ldots, e_n \otimes e_n \otimes e_1 \in \mathbb{C}^n \otimes \mathbb{C}^n \otimes \mathbb{C}$ are 2-orthogonal
$e_1 \otimes 1 \otimes e_1, \ldots, e_n \otimes 1 \otimes e_n \in \mathbb{C}^n \otimes \mathbb{C} \otimes \mathbb{C}^n$ are 2-orthogonal
$1 \otimes e_1 \otimes e_1, \ldots, 1 \otimes e_n \otimes e_n \in \mathbb{C} \otimes \mathbb{C}^n \otimes \mathbb{C}^n$ are 2-orthogonal

Example: Matrix Multiplication Tensor

$e_1, \ldots, e_n \in \mathbb{C}^n$ are orthogonal
$e_1 \otimes e_1, \ldots, e_n \otimes e_n \in \mathbb{C}^n \otimes \mathbb{C}^n$ are 2-orthogonal

$e_1 \otimes e_1 \otimes 1, \ldots, e_n \otimes e_n \otimes e_1 \in \mathbb{C}^n \otimes \mathbb{C}^n \otimes \mathbb{C}$ are 2-orthogonal
$e_1 \otimes 1 \otimes e_1, \ldots, e_n \otimes 1 \otimes e_n \in \mathbb{C}^n \otimes \mathbb{C} \otimes \mathbb{C}^n$ are 2-orthogonal
$1 \otimes e_1 \otimes e_1, \ldots, 1 \otimes e_n \otimes e_n \in \mathbb{C} \otimes \mathbb{C}^n \otimes \mathbb{C}^n$ are 2-orthogonal

Using vertical tensor product, we get

$$\{(e_i \otimes e_j) \otimes (e_j \otimes e_k) \otimes (e_k \otimes e_i) \mid 1 \leq i, j, k \leq n\}$$

are 2-orthogonal.

Example: Matrix Multiplication Tensor

Theorem (D.)

The matrix multiplication tensor

$$T_n = \sum_{i,j,k=1}^{n} e_{i,j} \otimes e_{j,k} \otimes e_{k,i}$$

is a DSVD.

The singular values of T_n are

$$\underbrace{1, 1, \ldots, 1}_{n^3}$$

In particular,

$$\|T_n\|_\star = \sum_{i=1}^{n^3} 1 = n^3.$$

DFT

Define

$$F_n = \sum_{\substack{1 \leq i,j,k \leq n \\ i+j+k \equiv 0 \bmod n}} e_i \otimes e_j \otimes e_k$$

This tensor is related to the multiplication of univariate polynomials. Clearly $\operatorname{rank}(F_n) \leq n^2$ and $\|F_n\|_\star \leq n^2$.

DFT

Define

$$F_n = \sum_{\substack{1\leq i,j,k\leq n \\ i+j+k\equiv 0 \bmod n}} e_i \otimes e_j \otimes e_k$$

This tensor is related to the multiplication of univariate polynomials. Clearly $\operatorname{rank}(F_n) \leq n^2$ and $\|F_n\|_\star \leq n^2$.

Discrete Fourier Transform (DFT):

$$F_n = \sum_{j=1}^{n} \sqrt{n}\left(\tfrac{1}{\sqrt{n}} \textstyle\sum_{i=1}^n \zeta^{ij} e_i\right) \otimes \left(\tfrac{1}{\sqrt{n}} \textstyle\sum_{i=1}^n \zeta^{ij} e_i\right) \otimes \left(\tfrac{1}{\sqrt{n}} \textstyle\sum_{i=1}^n \zeta^{ij} e_i\right).$$

where $\zeta = e^{\pi i/n}$. This is the *unique* DSVD of F_n. So the singular values are $\sqrt{n}, \dots, \sqrt{n}$ (n times), $\operatorname{rank}(F_n) = n$ and $\|F_n\|_\star = n\sqrt{n}$.

Generalization: Group Algebra Multiplication Tensor

G is a group with n elements and $\mathbb{C}G \cong \mathbb{C}^n$ is the group algebra

$$T_G = \sum_{g,h \in G} g \otimes h \otimes h^{-1}g^{-1}.$$

DFT case corresponds to $G = \mathbb{Z}/n\mathbb{Z}$.

Generalization: Group Algebra Multiplication Tensor

G is a group with n elements and $\mathbb{C}G \cong \mathbb{C}^n$ is the group algebra

$$T_G = \sum_{g,h \in G} g \otimes h \otimes h^{-1} g^{-1}.$$

DFT case corresponds to $G = \mathbb{Z}/n\mathbb{Z}$.

Theorem (D.)

T_G has a DSVD and its singular values are

$$\underbrace{\sqrt{\frac{n}{d_1}}, \dots, \sqrt{\frac{n}{d_1}}}_{d_1^3}, \dots, \underbrace{\sqrt{\frac{n}{d_s}}, \dots, \sqrt{\frac{n}{d_s}}}_{d_s^3}$$

where $d_1, d_2, \dots, d_s$ are the dimension of the irreducible representations of G.