Found Comput Math (2016) 16:779–811
DOI 10.1007/s10208-015-9264-x

FOUNDATIONS OF COMPUTATIONAL MATHEMATICS
The Journal of the Society for the Foundations of Computational Mathematics

On the Nuclear Norm and the Singular Value Decomposition of Tensors

Harm Derksen[1]

Received: 11 December 2013 / Revised: 6 January 2015 / Accepted: 13 January 2015 /
Published online: 19 May 2015

Abstract Finding the rank of a tensor is a problem that has many applications. Unfortunately, it is often very difficult to determine the rank of a given tensor. Inspired by the heuristics of convex relaxation, we consider the nuclear norm instead of the rank of a tensor. We determine the nuclear norm of various tensors of interest. Along the way, we also do a systematic study various measures of orthogonality in tensor product spaces and we give a new generalization of the singular value decomposition to higher-order tensors.

Keywords Tensor decomposition · Nuclear norm · Singular value decomposition · PARAFAC · CANDECOMP

Mathematics Subject Classification 15A69 · 15A18

1 Introduction

1.1 Tensor Decompositions

Suppose that $V = V^{(1)} \otimes \cdots \otimes V^{(d)}$ is the tensor product of finite-dimensional Hilbert spaces. For some applications, we would like to find a decomposition of a given tensor T as a sum of pure tensors:

Communicated by Peter Bürgisser.

The author was partially supported by NSF Grants DMS 0901298 and DMS 1302032.

✉ Harm Derksen
hderksen@umich.edu

[1] Department of Mathematics, University of Michigan, 530 Church Street, Ann Arbor, MI 48109-1043, USA

$$T = \sum_{i=1}^{r} v_i, \text{ where } v_i = v_i^{(1)} \otimes v_i^{(2)} \otimes \cdots \otimes v_i^{(d)} \text{ and } v_i^{(e)} \in V^{(e)}. \tag{1}$$

The smallest possible r for which a decomposition (1) exists is called the *rank* of T (see [17]). For $d = 2$, the rank of a tensor corresponds to the rank of a matrix. So, the tensor rank can be thought of as a generalization of the matrix rank to higher-dimensional arrays. For $d \geq 3$, it is difficult to determine the rank of a given tensor, or even to give good upper and lower bounds. For example, a dimension counting argument shows that a dense open subset of $\mathbb{C}^n \otimes \mathbb{C}^n \otimes \mathbb{C}^n$ consists of tensors of rank $\geq n^3/(3n-2) = O(n^2)$. So far, there are no known explicit families of examples of tensors with a proven lower bound of $\omega(n)$ (by definition, a function $f(n)$ has a lower bound $\omega(n)$ if $\lim_{n\to\infty} f(n)/n = \infty$). The problem of finding the rank of a given tensor is known to be NP-hard (see [15,16,18]). The tensor rank plays an important role in algebraic complexity theory. The complexity of matrix multiplication, for example, is closely related to the rank of a certain tensor (see Sect. 1.5).

In some applications, we just would like to find a low-rank approximation: for a small fixed value of r, we want to find pure tensors $v_1, \ldots, v_r$ such that the ℓ^2-norm

$$\left\| T - \sum_{i=1}^{r} v_i \right\|$$

is small. As pointed out in [10], there may not always be an optimal solution for which this norm is minimal. The problem of finding a low-rank approximation is known as the PARAFAC ([14]) or CANDECOMP ([8]) model. There are many applications of this model, for example fluorescence spectroscopy, statistics, psychometrics, geophysics and magnetic resonance imaging.

1.2 The Nuclear and Spectral Norms

Convex relaxation is a powerful technique that is based on the following idea: Suppose that we are trying to find the sparsest solution $x = (x_1, \ldots, x_n) \in \mathbb{C}^n$ to some problem. In other words, we are trying to find a solution x such that

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

is minimal. This is typically a very hard problem because the function $\|\cdot\|_0$ is not convex or continuous. But sometimes one can prove that minimizing the ℓ^1-norm

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|$$

yields the sparsest solution. A sparse relaxation of the rank of a matrix A is the nuclear norm $\|A\|_\star = \text{trace}(\sqrt{AA^\star})$, which is also the sum of the singular values of A. The nuclear norm is known to be dual to the spectral norm (see, for example, [5,34]). In this context, this relaxation technique has been successfully applied to matrix completion problems in [5,6,19,34].

Definition 1.1 The spectral norm $[T]$ of a tensor T is defined as the maximum value of $|\langle T, u\rangle|$ where u ranges over all pure tensors of unit length.

For a matrix, the spectral norm is just the largest singular value. Grothendieck ([13]) and Schatten ([30]) defined the nuclear norm as follows.

Definition 1.2 The nuclear norm $\|\cdot\|_\star$ is the norm that is dual to the spectral norm. This means that for a tensor T, we have

$$\|T\|_\star = \max\{|\langle T, S\rangle| \mid S \text{ is a tensor with } [S] = 1\}$$

and

$$[T] = \max\{|\langle T, S\rangle| \mid S \text{ is a tensor with} \|S\|_\star = 1\}.$$

We will discuss also an alternative definition. Let R be the compact set of all rank one tensors of unit length, and let B be its convex hull. Carathéodory's convexity theorem (see [1, Theorem 5.32]) implies that every vector in B is an affine combination of at most $n+1$ elements of R. It follows that B is compact (because it is the image of a continuous map $R^{n+1} \times \Delta_n \to B$ where Δ_n is the n-dimensional simplex, see [1, Corollary 5.33]). Also, B is invariant under multiplication with complex scalars of length 1, so we can view B as the unit ball of a norm. From Definition 1.1, it follows that this norm is dual to the spectral norm. So, we conclude that B is the unit ball of the nuclear norm.

Lemma 1.3 *The nuclear norm $\|T\|_\star$ of a tensor T is the smallest possible value of $\sum_{i=1}^r \|v_i\|$ over all possible decompositions (1).*

Proof If T is a nonzero tensor, then $T/\|T\|_\star \in B$, and we can write $T/\|T\|_\star = \sum_{i=1}^r \lambda_i w_i$, where $r \leq n+1$, $\lambda_1, \ldots, \lambda_r$ are nonnegative, $\sum_{i=1}^r \lambda_i = 1$, and $w_1, \ldots, w_r \in R$. If we set $v_i = \|T\|_\star \lambda_i w_i$ for all i, then we have $T = \sum_{i=1}^r v_i$ and $\|T\|_\star = \sum_{i=1}^r \|v_i\|$.

On the other hand, suppose that we have a decomposition $T = \sum_{i=1}^r v_i$ into pure tensors. Because the nuclear norm is dual to the spectral norm, there exists a tensor S with $[S] = 1$ and $\langle T, S\rangle = \|T\|_\star$. It follows that

$$\|T\|_\star = \langle T, S\rangle = \sum_{i=1}^r \langle S, v_i\rangle \leq \sum_{i=1}^r \|v_i\|.$$

□

See [27, Definition 17]) for a definition of the nuclear norm that is equivalent to Lemma 1.3. The nuclear norm for tensors has been used for tensor completion problems in [11].

More generally, if $\mathbf{T} = (T_1, \ldots, T_r)$ is an r-tuple of tensors, then we define $[\mathbf{T}]_\alpha$ as the maximum of

$$\left(\sum_{i=1}^{r} |\langle T_i, u\rangle|^\alpha\right)^{1/\alpha}$$

over all pure tensors u of unit length. The following theorem is useful for obtaining lower bounds for the spectral norm:

Theorem 1.4 *If T is a tensor, $\mathbf{S} = (S_1, \ldots, S_r)$ is an r-tuple of tensors and $\alpha \geq 1$, then we have*

$$\left(\sum_{i=1}^{r} |\langle T, S_i\rangle|^\alpha\right)^{1/\alpha} \leq \|T\|_\star [\mathbf{S}]_\alpha$$

The proof of Theorem 1.4 is in Sect. 5. If $\mathbf{S} = (S)$ just consists of a single tensor, then $[\mathbf{S}]_\alpha = [S]$ and we have:

Corollary 1.5 *For tensors S, T, we have*

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

Corollary 1.5 also immediately follows from the duality. If we set $S = T$, then we obtain

$$\|T\|^2 \leq \|T\|_\star [T]. \tag{2}$$

From the Cauchy–Schwartz inequality, it follows that $|\langle T, v\rangle| \leq \|T\|\|v\| = \|T\|$ for any pure tensor v of unit length, so $[T] \leq \|T\|$. Combined with (2), this gives the inequality $\|T\| \leq \|T\|_\star$. This inequality also follows from Lemma 1.3 and the triangle inequality. So, we have

$$[T] \leq \|T\| \leq \|T\|_\star.$$

Geometrically, the unit sphere circumscribes the unit ball for the spectral norm and inscribes the unit ball for the nuclear norm.

1.3 Singular Value Decomposition

The singular value decomposition (SVD) can be generalized to higher-dimensional arrays. One such generalization was given in [24]. Given a tensor T, one can choose an orthonormal bases $f_1^{(i)}, \ldots, f_{n_i}^{(i)}$ for V_i for all i and express T in these bases:

$$T = \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \cdots \sum_{i_d=1}^{n_d} \lambda_{i_1,i_2,\ldots,i_d} f_{i_1}^{(1)} \otimes f_{i_2}^{(2)} \otimes \cdots \otimes f_{i_d}^{(d)}. \tag{3}$$

Define

$$T_k^{(j)} = \sum_{i_j=k} \lambda_{i_1,i_2,\ldots,i_d} f_{i_1}^{(1)} \otimes f_{i_2}^{(2)} \otimes \cdots \otimes f_{i_d}^{(d)},$$

where the sum runs over all d-tuples $(i_1, \ldots, i_d)$ with $i_j = k$. For a proper choice of the bases, the tensors $T_1^{(j)}, \ldots, T_{n_j}^{(j)}$ are orthogonal for all j and

$$\|T_1^{(j)}\| \geq \|T_2^{(j)}\| \geq \cdots \geq \|T_{n_j}^{(j)}\|.$$

These numbers are called the *singular values in mode j*. The decomposition (3) is called the *higher-order single value decomposition* (HOSVD).

In this paper, we will give a different generalization of the SVD, which we call the *diagonal singular value decomposition* (DSVD). A given tensor may not have a diagonal singular value decomposition (see Sect. 7), but if it does, then the decomposition has many nice properties.

Definition 1.6 Suppose that $t \geq 1$ is a real number. An r-tuple $\mathbf{S} = (S_1, \ldots, S_r)$ of tensors of unit length is called t-orthogonal if $[\mathbf{S}]_{2/t} = 1$.

The term "t-orthogonal" comes from an orthogonality property described in Lemma 4.6.

Lemma 1.7 *If* $\mathbf{S}$ *is* t*-orthogonal and* $t \geq t'$*, then* $\mathbf{S}$ *is* t'*-orthogonal.*

Proof If $\mathbf{S}$ is t-orthogonal and $t \geq t'$, then we have

$$[\mathbf{S}]_{2/t'}^{2/t'} = \max_u \sum_{i=1}^r |\langle S, u\rangle|^{2/t'} \leq \max_u \sum_{i=1}^r |\langle S, u\rangle|^{2/t} = 1,$$

so $\mathbf{S}$ is t'-orthogonal. □

Lemma 1.8 *If* $\mathbf{v} = (v_1, \ldots, v_r)$ *is an* r*-tuple of pure tensors of unit length, then* t*-orthogonality is equivalent to the usual notion of orthogonality.*

Proof If $\mathbf{v}$ is orthogonal,

$$[\mathbf{v}]_2^2 = \max_u \sum_{i=1}^r |\langle v_i, u\rangle|^2 \leq \|u\|^2 = 1$$

by Pythagoras' theorem, so $\mathbf{v}$ is 2-orthogonal. The converse follows from Lemma 4.6. □

Definition 1.9 If $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ are real and $(v_1, \ldots, v_r)$ is a 2-orthogonal r-tuple of pure tensors of unit length, then a decomposition

$$T = \sum_{i=1}^r \sigma_i v_i \tag{4}$$

is called a DSVD of T and $\sigma_1, \ldots, \sigma_r$ are called the *singular values* of T.

For a tensor T that has a DSVD, we have the following results:

Theorem 1.10 *The singular values of T are uniquely determined by T (and do not depend on the choice of the diagonal singular value decomposition).*

Theorem 1.11 *Suppose that T has a singular value decomposition with singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$. Then, we have*

$$\|T\|_\star = \sum_{i=1}^{r} \sigma_i, \quad [T] = \sigma_1, \textit{ and } \|T\| = \sqrt{\textstyle\sum_{i=1}^{r} \sigma_i^2}.$$

Theorem 1.12 *If the singular values of T are distinct, then the diagonal singular value decomposition is unique.*

Theorem 1.13 *If $T = \sum_{i=1}^{r} \sigma_i v_i$ is a diagonal singular value decomposition and $(v_1, \ldots, v_r)$ is t-orthogonal for some $t > 2$, then the DSVD of T is unique.*

The proofs of Theorems 1.10–1.13 are in Sect. 6.

1.4 Tensors and Multilinear Maps

To a tensor

$$T = \sum_{i=1}^{r} v_i^{(1)} \otimes \cdots \otimes v_i^{(d)} \in V^{(1)} \otimes \cdots \otimes V^{(d)},$$

we can associate a multilinear map

$$\varphi_T : (V^{(1)})^\star \times (V^{(2)})^\star \times \cdots \times (V^{(d-1)})^\star \to V^{(d)},$$

defined by

$$\varphi_T(f^{(1)}, f^{(2)}, \ldots, f^{(d-1)}) = \sum_{i=1}^{r} \left(\textstyle\prod_{j=1}^{d-1} f^{(j)}\big(v_i^{(j)}\big)\right) v_i^{(d)}.$$

We will apply this correspondence to matrix multiplication.

1.5 Matrix Multiplication

Let $\mathbb{C}^{p\times q}$ denote the set of $p \times q$ matrices. The Hermitian form is given by $\langle A, B\rangle = \text{trace}(AB^\star)$. The matrix with a 1 in position (i, j) and zeroes everywhere else is denoted by $e_{i,j}$. Then, matrix multiplication

$$\mathbb{C}^{p\times q} \times \mathbb{C}^{q\times r} \to \mathbb{C}^{p\times r}$$

corresponds to the tensor

$$\sum_{i=1}^{p} \sum_{j=1}^{q} \sum_{k=1}^{r} e_{i,j} \otimes e_{j,k} \otimes e_{i,k} \in \mathbb{C}^{p\times q} \otimes \mathbb{C}^{q\times r} \otimes \mathbb{C}^{p\times r}.$$

If we identify $\mathbb{C}^{p\times r}$ with $\mathbb{C}^{r\times p}$, then the tensor has the following, more symmetric, form:

$$M_{p,q,r} = \sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r} e_{i,j}\otimes e_{j,k}\otimes e_{k,i} \in \mathbb{C}^{p\times q}\otimes\mathbb{C}^{q\times r}\otimes\mathbb{C}^{r\times p}. \tag{5}$$

From this formula, it is clear that $\text{rank}(M_{p,q,r}) \leq pqr$. Strassen proved that $\text{rank}(M_{2,2,2}) \leq 7$ (see [32]) by giving a decomposition

$$\begin{aligned} M_{2,2,2} = {} & (e_{1,1}+e_{2,2})\otimes(e_{1,1}+e_{2,2})\otimes(e_{1,1}+e_{2,2}) \\ & +(e_{2,1}-e_{2,2})\otimes e_{1,1}\otimes(e_{1,2}+e_{2,2}) \\ & +(e_{1,2}+e_{2,2})\otimes(e_{2,1}-e_{2,2})\otimes e_{1,1} \\ & +(e_{1,1}+e_{2,1})\otimes(e_{1,2}-e_{1,1})\otimes e_{2,2} \\ & +(e_{1,2}-e_{1,1})\otimes e_{2,2}\otimes(e_{1,1}+e_{2,1}) \\ & +e_{2,2}\otimes(e_{1,1}+e_{2,1})\otimes(e_{1,2}-e_{1,1}) \\ & +e_{1,1}\otimes(e_{1,2}+e_{2,2})\otimes(e_{2,1}-e_{2,2}) \end{aligned} \tag{6}$$

and used this to show that two $n\times n$ matrices can be multiplied by using only $O(n^{\log_2(7)})$ arithmetic where $\log_2(7)\approx 2.81 < 3$. The usual way of multiplying two matrices takes $O(n^3)$ arithmetic operations. More generally, define

$$\omega = \inf\left\{\frac{\log(\text{rank}(M_{p,q,r}))}{\log(pqr)}\middle| p,q,r\geq 2\right\}.$$

If $\varepsilon > 0$, then two $n\times n$ matrices can be multiplied using only $o(n^{\omega+\varepsilon})$ arithmetic operations (see [2] and [4]). Coppersmith and Winograd proved that $\omega < 2.376$ in [9]. Only recently, this bound was improved by Stothers ([35]) and Williams ([36]) and the current record is $\omega < 2.3728639$ by Le Gall ([25]). For most values of p, q and r, the rank of $M_{p,q,r}$ is unknown. It is easy to see that $\text{rank}(M_{n,n,n})\geq n^2$. Bläser gave a better, nontrivial lower bound in [3]. A sharper lower bound was given by Landsberg in [22], and using the same techniques, Massarenti and Raviolo (see [28]) improved this lower bound to

$$\text{rank}(M_{n,n,n}) \geq 3n^2 - 2\sqrt{2}n^{3/2} - 3n.$$

Theorem 1.14 *The decomposition (5) is a diagonal singular value decomposition. In particular, the singular values of $M_{p,q,r}$ are*

$$\underbrace{1,1,\ldots,1}_{pqr}.$$

The proof of the theorem is in Sect. 4. The following corollary follows from Theorem 1.14 and Theorem 1.11.

Corollary 1.15 *We have* $\|M_{p,q,r}\|_\star = pqr$ *and* $[M_{p,q,r}] = 1$.

Note that the sum of the lengths of the pure tensors in the decomposition (6) is $2\sqrt{2} + 12 > 8$. This shows that minimization of the rank and minimization of the nuclear norm do not always coincide.

1.6 The Discrete Fourier Transform and Group Algebras

Suppose that G is a finite group. The group algebra $\mathbb{C}G$ is the vector space with a orthonormal basis $g, g \in G$. Multiplication in the group G gives $\mathbb{C}G$ the structure of an associative algebra. The multiplication

$$\mathbb{C}G \times \mathbb{C}G \to \mathbb{C}G$$

corresponds to the tensor

$$\sum_{g\in G}\sum_{h\in G} g \otimes h \otimes gh \in \mathbb{C}G \otimes \mathbb{C}G \otimes \mathbb{C}G.$$

By permuting the basis vectors in the last factor $\mathbb{C}G$, the tensor can be written in the following symmetric form:

$$T_G := \sum_{\substack{g,h,k\in G\\ ghk=1}} g \otimes h \otimes k.$$

We have $\|T_G\| = \sqrt{n^2} = n$.

Theorem 1.16 *If G is a group of order n and $d_1 \le \cdots \le d_s$ are the dimensions of the irreducible representations of G, then the tensor T_G has a DSVD and its singular values are*

$$\underbrace{\sqrt{\frac{n}{d_1}},\ldots,\sqrt{\frac{n}{d_1}}}_{d_1^3},\underbrace{\sqrt{\frac{n}{d_2}},\ldots,\sqrt{\frac{n}{d_2}}}_{d_2^3},\ldots,\underbrace{\sqrt{\frac{n}{d_s}},\ldots,\sqrt{\frac{n}{d_s}}}_{d_s^3}.$$

The proof of Theorem 1.16 can be found in Sect. 4. The dimension of the trivial representation is $d_1 = 1$, so the largest singular value is $\sqrt{n}$. From Theorem 1.16 and Theorem 1.11, we get the following result.

Corollary 1.17 *We have* $\|T_G\|_\star = \sqrt{n}\sum_{j=1}^s d_j^{5/2}$ *and* $[T_G] = \sqrt{n}$.

Let C_n be the (multiplicative) cyclic group of order n generated by x. Then, $\mathbb{C}C_n$ is the commutative ring $\mathbb{C}[x]/(x^n - 1)$ and multiplication in $\mathbb{C}G$ corresponds to the multiplication of polynomials in one variable (modulo $x^n - 1$). We have

$$T_{C_n} = \sum_{j+k+l=0} x^j \otimes x^k \otimes x^l, \tag{7}$$

where the sum is over all $j, k, l \in \mathbb{Z}/n\mathbb{Z}$ with $j+k+l=0$. From (7), it follows that $\mathrm{rank}(T_{C_n}) \leq n^2$ and $\|T_{C_n}\|_\star \leq n^2$. Let $\zeta = e^{2\pi i/n}$ be a primitive n-th root of unity. The discrete Fourier transform is based on the following decomposition of T_{C_n}:

$$T_{C_n} = \sum_{t=0}^{n-1} \sqrt{n} \left(\tfrac{1}{\sqrt{n}} \textstyle\sum_{j=0}^{n-1} \zeta^{tj} x^j\right) \otimes \left(\tfrac{1}{\sqrt{n}} \textstyle\sum_{k=0}^{n-1} \zeta^{tk} x^k\right) \otimes \left(\tfrac{1}{\sqrt{n}} \textstyle\sum_{l=0}^{n-1} \zeta^{tl} x^l\right). \quad (8)$$

The following theorem follows from Theorems 1.16 and 1.13.

Theorem 1.18 *The decomposition (8) is the unique diagonal singular value decomposition. In particular, the singular values of T_{C_n} are*

$$\underbrace{\sqrt{n}, \sqrt{n}, \ldots, \sqrt{n}}_{n},$$

$[T_{C_n}] = \sqrt{n}$ *and* $\|T_{C_n}\|_\star = n\sqrt{n}$.

1.7 The Determinant and the Permanent

The determinant and permanent are multilinear functions

$$\underbrace{\mathbb{C}^n \times \cdots \times \mathbb{C}^n}_{n} \to \mathbb{C}.$$

Let $\mathbb{S}_n$ be the symmetric group on n letters. The sign of a permutation $\sigma \in \mathbb{S}_n$ is denoted by $\mathrm{sgn}(\sigma) \in \{1, -1\}$. The determinant and permanent correspond to the tensors

$$\det_n := \sum_{\sigma \in \mathbb{S}_n} \mathrm{sgn}(\sigma) e_{\sigma(1)} \otimes e_{\sigma(2)} \otimes \cdots \otimes e_{\sigma(n)}$$

and

$$\mathrm{per}_n := \sum_{\sigma \in \mathbb{S}_n} e_{\sigma(1)} \otimes e_{\sigma(2)} \otimes \cdots \otimes e_{\sigma(n)},$$

respectively, in

$$\underbrace{\mathbb{C}^n \otimes \cdots \otimes \mathbb{C}^n}_{n} \otimes \mathbb{C} \cong \underbrace{\mathbb{C}^n \otimes \cdots \otimes \mathbb{C}^n}_{n}.$$

From these formulas, it is clear that $\mathrm{rank}(\det_n) \leq n!$ and $\mathrm{rank}(\mathrm{per}_n) \leq n!$. The upper bound for the rank of the determinant is not sharp for $n \geq 3$ (see Sect. 8). The bound for the permanent is far from optimal. Another formula for the permanent was given by Glynn [12]:

$$\mathrm{per}_n = \frac{1}{2^{n-1}} \sum_{\delta} \left(\prod_{k=1}^{n} \delta_k\right) \left(\textstyle\sum_{j=1}^{n} \delta_j e_j\right) \otimes \left(\textstyle\sum_{j=1}^{n} \delta_j e_j\right) \otimes \cdots \otimes \left(\textstyle\sum_{j=1}^{n} \delta_j e_j\right), \quad (9)$$

where δ runs over all 2^{n-1} vectors $\delta = (\delta_1, \ldots, \delta_n) \in \{1, -1\}^n$ with $\delta_1 = 1$. From this formula, it follows that $\operatorname{rank}(\mathrm{per}_n) \leq 2^{n-1}$ and $\| \mathrm{per}_n \|_\star \leq n^{n/2}$. Some easy lower bounds for the rank of the permanent and determinant are given in Sect. 8.

Theorem 1.19 *We have* $\| \mathrm{per}_n \|_\star = n^{n/2}$.

The formula (9) minimizes the sum of the lengths of the pure tensors, but these pure tensors are not 2-orthogonal (or even orthogonal) for $n \geq 3$. In fact, for $d \geq 3$, the tensor per_n does not have a DSVD (see Example 7.1).

Theorem 1.20 *We have* $\| \det_n \|_\star = n!$.

The proofs of Theorems 1.20 and 1.19 are in Sect. 5. The determinant also does not have a diagonal singular value decomposition (see Example 7.2).

2 Orthogonality of Vectors

In this section, we will study various measures of orthogonality of vectors in a Hilbert space V. It is convenient to deal with unit vectors. Suppose that $\mathbf{v} = (v_1, \ldots, v_r)$ is an r-tuple of unit vectors.

Definition 2.1 The *coherence* is defined by

$$\mu(\mathbf{v}) = \max_{i \neq j} |\langle v_i, v_j \rangle|$$

if $r > 1$ and $\mu(\mathbf{v}) = 0$ if $r = 1$.

More generally, we will define

Definition 2.2

$$\mu_\alpha(\mathbf{v}) = \max_i \left\{ \left(\sum_{j \neq i} |\langle v_i, v_j \rangle|^\alpha \right)^{1/\alpha} \right\}.$$

Notice that $\lim_{\alpha \to \infty} \mu_\alpha(\mathbf{v}) = \mu(\mathbf{v})$. So, we may think of $\mu(\mathbf{v})$ as $\mu_\infty(\mathbf{v})$.

Suppose that $\mathbf{v} = (v_1, \ldots, v_r)$ and $\mathbf{w} = (w_1, \ldots, w_r)$ are r-tuples of unit vectors. We define the *horizontal tensor product* of $\mathbf{v}$ and $\mathbf{w}$ by

$$\mathbf{v} \otimes \mathbf{w} := (v_1 \otimes w_1, \ldots, v_r \otimes w_r) \in (V \otimes W)^r$$

If $\mathbf{v} = (v_1, \ldots, v_r) \in V^r$ and $\mathbf{w} = (w_1, \ldots, w_s) \in W^s$, then we define

$$\mathbf{v} \boxtimes \mathbf{w} = (v_1 \otimes w_1, v_1 \otimes w_2, \ldots, v_1 \otimes w_s, v_2 \otimes w_1, \ldots, v_r \otimes w_s) \in (V \otimes W)^{rs}.$$

It is easy to see that

$$\mu(v \boxtimes w) = \max\{\mu(v), \mu(w)\}.$$

Lemma 2.3 *If* $\mathbf{v} = (v_1, \ldots, v_r) \in V^r$ *and* $\mathbf{w} = (w_1, \ldots, w_s) \in W^s$ *are tuples of unit vectors, then we have*

$$\mu_\alpha(\mathbf{v} \boxtimes \mathbf{w})^\alpha + 1 = (\mu_\alpha(\mathbf{v})^\alpha + 1)(\mu_\alpha(\mathbf{w})^\alpha + 1)$$

for $\alpha > 0$ *and*

$$\mu(\mathbf{v} \otimes \mathbf{w}) = \max\{\mu(\mathbf{v}), \mu(\mathbf{w})\}.$$

Proof We have

$$\begin{aligned}
\mu_\alpha(\mathbf{v} \boxtimes \mathbf{w})^\alpha + 1 &= \max_{i,j} \left\{ \sum_{k=1}^{r} \sum_{l=1}^{s} |\langle v_i \otimes w_j, v_k \otimes w_l \rangle|^\alpha \right\} \\
&= \max_{i,j} \left\{ \sum_{k=1}^{r} \sum_{l=1}^{s} |\langle v_i, v_k \rangle \langle w_j, w_l \rangle|^\alpha \right\} \\
&= \max_{i,j} \left\{ \left(\sum_{k=1}^{r} |\langle v_i, v_k \rangle|^\alpha \right) \left(\sum_{l=1}^{s} |\langle w_j, w_l \rangle|^\alpha \right) \right\} \\
&= \max_{i} \left\{ \sum_{k=1}^{r} |\langle v_i, v_k \rangle|^\alpha \right\} \max_{j} \left\{ \sum_{l=1}^{s} |\langle w_j, w_l \rangle|^\alpha \right\} \\
&= (\mu_\alpha(\mathbf{v})^\alpha + 1)(\mu_\alpha(\mathbf{w})^\alpha + 1).
\end{aligned}$$

□

We will need a slightly more general version of the Hölder inequality.

Lemma 2.4 *(Hölder inequality) If* $a_1, \ldots, a_r, b_1, \ldots, b_r$ *are nonnegative real numbers, and* α, β, γ *are positive real numbers with* $1/\alpha + 1/\beta = 1/\gamma$*, then we have*

$$\left(\sum_{i=1}^{r} (a_i b_i)^\gamma \right)^{1/\gamma} \leq \left(\sum_{i=1}^{r} a_i^\alpha \right)^{1/\alpha} \left(\sum_{i=1}^{r} b_i^\beta \right)^{1/\beta}.$$

Proof The usual Hölder inequality states that, if $1/p + 1/q = 1$, then we have

$$\sum_{i=1}^{r} a_i b_i \leq \left(\sum_{i=1}^{r} a_i^p \right)^{1/p} \left(\sum_{i=1}^{r} b_i^q \right)^{1/q}$$

with equality if and only if the vectors $(a_1^p, \ldots, a_r^p)$ and $(b_1^q, \ldots, b_r^q)$ are dependent. Now, take $p = \alpha/\gamma$ and $q = \beta/\gamma$, and replace a_i and b_i by a_i^γ and b_i^γ, respectively,

$$\sum_{i=1}^{r} (a_i b_i)^\gamma \leq \left(\sum_{i=1}^{r} a_i^\alpha \right)^{\gamma/\alpha} \left(\sum_{i=1}^{r} b_i^\beta \right)^{\gamma/\beta}.$$

Taking the γ-th root gives the desired inequality. □

For horizontal tensor products, we have a Hölder inequality:

Lemma 2.5 *If* $1/\alpha + 1/\beta = 1/\gamma$*, then we have*

$$\mu_\gamma(\mathbf{v} \otimes \mathbf{w}) \leq \mu_\alpha(\mathbf{v})\mu_\beta(\mathbf{w}).$$

Proof By the Hölder inequality, we have

$$\begin{aligned}\left(\sum_{j\neq i} |\langle v_i \otimes w_i, v_j \otimes w_j\rangle|^\gamma\right)^{1/\gamma} &\leq \left(\sum_{j\neq i} \left(|\langle v_i, v_j\rangle| \cdot |\langle w_i, w_j\rangle|\right)^\gamma\right)^{1/\gamma} \\ &\leq \left(\sum_{j\neq i} |\langle v_i, v_j\rangle|^\alpha\right)^{1/\alpha} \left(\sum_{j\neq i} |\langle w_i, w_j\rangle|^\beta\right)^{1/\beta}\end{aligned}$$

for all i. Taking the maximum over all i on both sides gives the desired inequality. □

If we take $\beta \to \infty$, we get the inequalities

$$\mu_\alpha(\mathbf{v} \otimes \mathbf{w}) \leq \mu_\alpha(\mathbf{v})\mu(\mathbf{w})$$

and

$$\mu(\mathbf{v} \otimes \mathbf{w}) \leq \mu(\mathbf{v})\mu(\mathbf{w}).$$

Lemma 2.6 *If* $\gamma > \alpha > 0$ *and* $\mathbf{v} = (v_1, \ldots, v_r)$ *is an* r*-tuple of unit vectors, then we have*

$$\mu_\gamma(\mathbf{v})^{\gamma/\alpha} \leq \mu_\alpha(\mathbf{v}) \leq \mu_\gamma(\mathbf{v})(r-1)^{1/\alpha - 1/\gamma}.$$

Proof The inequality on the left is easy. For the inequality on the right, let $\mathbf{w} = (1, 1, \ldots, 1)$ and identify $\mathbf{v} \otimes \mathbf{w} \in (V \otimes \mathbb{C})^r \cong V^r$ with $\mathbf{v}$. Then, apply Lemma 2.5. □

If we take the limit $\gamma \to \infty$, we get

$$0 \leq \mu_\alpha(\mathbf{v}) \leq \mu(\mathbf{v})(r-1)^{1/\alpha}.$$

For an r-tuple $\mathbf{v} = (v_1, \ldots, v_r)$ of vectors, we define

$$\mathbf{v}^{\otimes d} = \underbrace{\mathbf{v} \otimes \mathbf{v} \otimes \cdots \otimes \mathbf{v}}_{d}.$$

Lemma 2.7 *For an* r*-tuple of unit vectors* $\mathbf{v}$*, we have* $\mu_\alpha(\mathbf{v}^{\otimes d}) = \mu_{d\alpha}(\mathbf{v})^d$.

Proof We have

$$\left(\sum_{j\neq i} |\langle v_i^{\otimes d}, v_j^{\otimes d}\rangle|^{\alpha}\right)^{1/\alpha} = \left(\sum_{j\neq i} |\langle v_i, v_j\rangle|^{d\alpha}\right)^{1/\alpha}.$$

Taking the maximum over all i on both sides gives the desired result. □

3 Orthogonality of Tensors

In this section, we study another measure for the orthogonality of pure tensors, which takes into account the tensor product structure of the vector space. It is important, when discussing pure tensors, to be clear which tensor product structure we are talking about. To be unambiguous, we make the following definition. An *d-th order tensor space* is a pair $\mathbf{V} = (V, (V^{(1)}, \dots, V^{(d)}))$, where $V^{(1)}, \dots, V^{(d)}$ are finite-dimensional Hilbert spaces and

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

A *pure tensor* (with respect to this tensor space) is an element in V of the form

$$v^{(1)} \otimes v^{(2)} \otimes \cdots \otimes v^{(d)}$$

with $v^{(i)} \in V^{(i)}$ for $i = 1, 2, \dots, d$. Suppose that $\mathbf{V} = (V, (V^{(1)}, \dots, V^{(d)}))$ and $\mathbf{W} = (W, (W^{(1)} \dots, W^{(e)}))$ are tensor product spaces. Then, their *horizontal tensor product* is the tensor space

$$\mathbf{V} \otimes \mathbf{W} := (V \otimes W, (V^{(1)}, \dots, V^{(d)}, W^{(1)}, \dots, W^{(e)})).$$

If $\mathbf{S} = (S_1, \dots, S_r) \in \mathbf{V}^r$ and $\mathbf{T} = (T_1, \dots, T_r) \in \mathbf{W}^r$, then we define

$$\mathbf{S} \otimes \mathbf{T} := (S_1 \otimes T_1, \dots, S_r \otimes T_r) \in (\mathbf{V} \otimes \mathbf{W})^r.$$

For the horizontal tensor product, we have a Hölder inequality.

Lemma 3.1 *If* $1/\alpha + 1/\beta = 1/\gamma$*, then we have*

$$[\mathbf{S} \otimes \mathbf{T}]_\gamma \leq [\mathbf{T}]_\alpha [\mathbf{S}]_\beta.$$

Proof Using the Hölder inequality, we get

$$\begin{aligned}\Big(\sum_{i=1}^{r} |\langle S_i \otimes T_i, x \otimes y\rangle|^{\gamma}\Big)^{1/\gamma} &= \Big(\sum_{i=1}^{r} |\langle S_i, x\rangle|^{\gamma} |\langle T_i, y\rangle|^{\gamma}\Big)^{1/\gamma} \\ &\leq \Big(\sum_{i=1}^{n} |\langle S_i, x\rangle|^{\alpha}\Big)^{1/\alpha} \Big(\sum_{i=1}^{n} |\langle T_i, y|\rangle|^{\beta}\Big)^{1/\beta}.\end{aligned}$$

Taking the supremum over all unit pure tensors x and y yields the desired inequality. □

Corollary 3.2 *We have*

$$[\mathbf{S}^{\otimes d}]_\alpha = [\mathbf{S}]^d_{d\alpha}.$$

Proof For some unit pure tensor u, we have

$$[\mathbf{S}]^d_{d\alpha} = \left(\sum_{i=1}^r |\langle S_i, u\rangle|^{d\alpha}\right)^{1/\alpha} = \left(\sum_{i=1}^r |\langle S_i^{\otimes d}, u^{\otimes d}\rangle|^{\alpha}\right)^{1/\alpha} \leq [\mathbf{S}^{\otimes d}]_\alpha.$$

The inequality in the other direction follows from Lemma 3.1. □

If $\mathbf{V} = (V, (V^{(1)}, \ldots, V^{(d)})$ and $\mathbf{W} = (W, (W^{(1)}, \ldots, W^{(d)}))$ are tensor product spaces, then their *vertical tensor product* is

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

If $S \in \mathbf{V}$ and $T \in \mathbf{W}$, we define $S \boxtimes T$ as $S \otimes T$, viewed inside the tensor product space $\mathbf{V} \boxtimes \mathbf{W}$. If $\mathbf{S} \in \mathbf{V}^r$ and $\mathbf{T} \in \mathbf{W}^s$, then we define

$$\mathbf{S} \boxtimes \mathbf{T} = (S_i \boxtimes T_j \mid 1 \leq i \leq r, 1 \leq j \leq s) \in (\mathbf{V} \boxtimes \mathbf{W})^{rs}.$$

The measure $[-]_\alpha$ also behaves multiplicatively with respect to the vertical tensor product.

Proposition 3.3 *Suppose that* $\mathbf{V} = (V, (V^{(1)}, \ldots, V^{(d)}))$ *and* $\mathbf{W} = (W, (W^{(1)}, \ldots, W^{(d)}))$ *are tensor product spaces, and* $S \in \mathbf{V}^r$ *and* $T \in \mathbf{W}^s$*. Then, we have*

$$[S \boxtimes T]_\alpha = [S]_\alpha [T]_\alpha.$$

Proof First, we will assume that $\alpha \leq 1$. For a complex vector $b = (b_1, \ldots, b_l)$, we have

$$\left|\sum_{k=1}^l b_k\right|^\alpha \leq \left(\sum_{k=1}^l |b_k|\right)^\alpha = \|b\|_1^\alpha \leq \|b\|_\alpha^\alpha = \sum_{k=1}^l |b_k|^\alpha. \tag{10}$$

Suppose that u is a pure tensor in $\mathbf{V} \boxtimes \mathbf{W}$. We can write

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

with $u^{(e)} \in Z^{(e)} = V^{(e)} \otimes W^{(e)}$ and $\|u^{(e)}\| = 1$ for $e = 1, 2, \ldots, d$. Using the singular value decomposition, we can write

$$u^{(e)} = \sum_{k=1}^{r_e} \lambda_k^{(e)} \left(x_k^{(e)} \boxtimes y_k^{(e)}\right) \in V^{(e)} \boxtimes W^{(e)}$$

for some positive integer r_e, where $x_1^{(e)}, x_2^{(e)}, \ldots, x_{r_e}^{(e)}$ and $y_1^{(e)}, y_2^{(e)}, \ldots, y_{r_e}^{(e)}$ are sequences of orthonormal vectors for $e = 1, 2, \ldots, d$, and $\lambda_k^{(e)} > 0$ if $1 \leq e \leq d$ and $1 \leq k \leq r_e$. Since $u^{(e)}$ is a unit vector, we have $\sum_{k=1}^{r_e} (\lambda_k^{(e)})^2 = 1$. We define $x^{(e)} = \sum_{k=1}^{r_e} \lambda_k^{(e)} x_k^{(e)}$, $y^{(e)} = \sum_{k=1}^{r_e} \lambda_k^{(e)} y_k^{(e)}$ for $e \leq d$, and

$$x = x^{(1)} \otimes \cdots \otimes x^{(d)}, \quad y = y^{(1)} \otimes \cdots \otimes y^{(d)}.$$

Note that $x^{(e)}$ and $y^{(e)}$ are unit vectors, and x and y are pure tensors of unit length. For a d-tuple $\underline{k} = (k_1, \ldots, k_d)$, we define

$$\lambda_{\underline{k}} = \prod_{e=1}^{d} \lambda_{k_e}^{(e)}, \quad x_{\underline{k}} = x_{k_1}^{(1)} \otimes \cdots \otimes x_{k_d}^{(d)}, \quad y_{\underline{k}} = y_{k_1}^{(1)} \otimes \cdots \otimes y_{k_d}^{(d)}.$$

So, we can write

$$u = \sum_{\underline{k}} \lambda_{\underline{k}} (x_{\underline{k}} \boxtimes y_{\underline{k}}),$$

where $\underline{k}$ runs over all d-tuples $(k_1, k_2, \ldots, k_d)$ with $1 \leq k_i \leq r_i$ for $i = 1, 2, \ldots, d$. This is a singular value decomposition of u, if u is viewed as a tensor in

$$(V \otimes W, (V^{(1)} \otimes \cdots \otimes V^{(d)}, W^{(1)} \otimes \cdots \otimes W^{(d)})).$$

Using the inequality (10) and the Cauchy–Schwarz inequality, we get

$$\begin{aligned}
\sum_{i,j} |\langle S_i \boxtimes T_j, u \rangle|^\alpha &= \sum_{i,j} \Big| \sum_{\underline{k}} \lambda_{\underline{k}} \langle S_i \boxtimes T_j, x_{\underline{k}} \boxtimes y_{\underline{k}} \rangle \Big|^\alpha \\
&\leq \sum_{i,j} \sum_{\underline{k}} \Big| \lambda_{\underline{k}} \langle S_i \boxtimes T_j, x_{\underline{k}} \boxtimes y_{\underline{k}} \rangle \Big|^\alpha \\
&= \sum_{\underline{k}} \lambda_{\underline{k}}^\alpha \sum_{i,j} |\langle S_i, x_{\underline{k}} \rangle|^\alpha |\langle T_j y_{\underline{k}} \rangle|^\alpha \\
&\leq \sum_{\underline{k}} \left(\lambda_{\underline{k}}^{\alpha/2} \sum_i |\langle S_i, x_{\underline{k}} \rangle|^\alpha \right) \left(\lambda_{\underline{k}}^{\alpha/2} \sum_j |\langle T_j, y_{\underline{k}} \rangle|^\alpha \right) \\
&\leq \sqrt{\textstyle\sum_{\underline{k}} \lambda_{\underline{k}}^\alpha \left(\sum_i |\langle S_i, x_{\underline{k}} \rangle|^\alpha \right)^2} \sqrt{\textstyle\sum_{\underline{k}} \lambda_{\underline{k}}^\alpha \left(\sum_j |\langle T_j, y_{\underline{k}} \rangle|^\alpha \right)^2} \\
&\leq \sqrt{\textstyle\sum_{\underline{k}} \lambda_{\underline{k}}^\alpha [\mathbf{S}]_\alpha^\alpha \sum_i |\langle S_i, x_{\underline{k}} \rangle|^\alpha} \sqrt{\textstyle\sum_{\underline{k}} \lambda_{\underline{k}}^\alpha [\mathbf{T}]_\alpha^\alpha \sum_j |\langle T_j, y_{\underline{k}} \rangle|^\alpha} \\
&\leq [\mathbf{S}]_\alpha^{\alpha/2} [\mathbf{T}]_\alpha^{\alpha/2} \sqrt{\textstyle\sum_i |\langle S_i, \sum_{\underline{k}} \lambda_{\underline{k}} x_{\underline{k}} \rangle|^\alpha} \sqrt{\textstyle\sum_j |\langle T_j, \sum_{\underline{k}} \lambda_{\underline{k}} y_{\underline{k}} \rangle|^\alpha} \\
&\leq [\mathbf{S}]_\alpha^{\alpha/2} [\mathbf{T}]_\alpha^{\alpha/2} \sqrt{\textstyle\sum_i |\langle S_i, x \rangle|^\alpha} \sqrt{\textstyle\sum_j |\langle T_j, y \rangle|^\alpha} \\
&\leq [\mathbf{S}]_\alpha^{\alpha/2} [\mathbf{T}]_\alpha^{\alpha/2} [\mathbf{S}]_\alpha^{\alpha/2} [\mathbf{T}]_\alpha^{\alpha/2} = [\mathbf{S}]_\alpha^\alpha [\mathbf{T}]_\alpha^\alpha
\end{aligned}$$

Since the pure tensor u was arbitrary, we have $[\mathbf{S} \boxtimes \mathbf{T}]_\alpha^\alpha \leq [\mathbf{S}]_\alpha^\alpha [\mathbf{T}]_\alpha^\alpha$ and $[\mathbf{S} \boxtimes \mathbf{T}]_\alpha \leq [\mathbf{S}]_\alpha [\mathbf{T}]_\alpha$.

If $\alpha \geq 1$, choose m such that $\alpha/m \leq 1$. By Corollary 3.2, we have

$$[(\mathbf{S} \boxtimes \mathbf{T})]_\alpha^m = [(\mathbf{S} \boxtimes \mathbf{T})^{\otimes m}]_{\alpha/m} = [\mathbf{S}^{\otimes m} \boxtimes \mathbf{T}^{\otimes m}]_{\alpha/m} \leq [\mathbf{S}^{\otimes m}]_{\alpha/m} [\mathbf{T}^{\otimes m}]_{\alpha/m} = [\mathbf{S}]_\alpha^m [\mathbf{T}]_\alpha^m,$$

so we get $[\mathbf{S} \otimes \mathbf{T}]_\alpha \leq [\mathbf{S}]_\alpha [\mathbf{T}]_\alpha$.

Suppose that $\alpha > 0$. There exist unit pure tensors $a \in \mathbf{V}$ and $b \in \mathbf{W}$ such that

$$\sum_i |\langle S_i, a\rangle|^\alpha = [\mathbf{S}]_\alpha^\alpha \text{ and } \sum_j \langle T_j, b\rangle|^\alpha = [\mathbf{T}]_\alpha^\alpha.$$

We have

$$\sum_{i,j} |\langle S_i \boxtimes T_j, a \boxtimes b\rangle|^\alpha = \sum_{i,j} |\langle S_i, a\rangle|^\alpha |\langle T_j, b\rangle|^\alpha = \sum_i |\langle S_i, a\rangle|^\alpha \sum_j |\langle T_j, b\rangle|^\alpha = [\mathbf{S}]_\alpha^\alpha [\mathbf{T}]_\alpha^\alpha.$$

So, it follows that $[\mathbf{S} \boxtimes \mathbf{T}]_\alpha^\alpha \geq [\mathbf{S}]_\alpha^\alpha [\mathbf{T}]_\alpha^\alpha$. We conclude that $[\mathbf{S} \boxtimes \mathbf{T}]_\alpha = [\mathbf{S}]_\alpha [\mathbf{T}]_\alpha$. □

The norm $[-]_\alpha$ is hard to compute in practice because we have to solve a optimization problem. But $\mu_\alpha(-)$ is easier to compute. Fortunately, $[-]_\alpha$ can be estimated in terms of μ_α:

Lemma 3.4 *For an r-tuple $\mathbf{v} = (v_1, \ldots, v_r)$ of pure tensors of unit length and $\alpha > 0$, we have*

$$\mu_\alpha(\mathbf{v})^\alpha + 1 \leq [\mathbf{v}]_\alpha^\alpha.$$

Proof Suppose that $\alpha > 0$. For every i, we have

$$1 + \sum_{j \neq i} |\langle v_i, v_j\rangle|^\alpha = \sum_{j=1}^r |\langle v_i, v_j\rangle|^\alpha \leq [\mathbf{v}]_\alpha^\alpha.$$

Taking the maximum over all i gives the desired inequality. □

Proposition 3.5 *For an r-tuple $\mathbf{v} = (v_1, \ldots, v_r)$ of pure tensors of unit length and $\alpha \geq 1$, we have*

$$[\mathbf{v}]_{2\alpha}^{2\alpha} \leq \mu_\alpha(\mathbf{v})^\alpha + 1.$$

Proof Suppose that $\alpha \geq 1$. Choose a unit pure tensor w such that

$$[\mathbf{v}]_{2\alpha} = \left(\sum_{i=1}^r |\langle v_i, w\rangle|^{2\alpha}\right)^{1/(2\alpha)}.$$

Let D be the $r \times r$ diagonal matrix with $D_{i,i} = |\langle v_i, w\rangle|^{2\alpha-2}$ and define

$$Z = \begin{pmatrix} v_1 & \cdots & v_r \end{pmatrix},$$

where $v_1, \ldots, v_r$ are viewed as column vectors with respect to some orthonormal basis. Consider the Hermitian matrix

$$A = \sum_{i=1}^{r} |\langle v_i, w\rangle|^{2\alpha-2} v_i v_i^\star = ZDZ^\star$$

Then, we have $w^\star A w = [\mathbf{v}]_{2\alpha}^{2\alpha}$. If $\lambda \in \mathbb{R}$ is the largest eigenvalue of A, then we get $[\mathbf{v}]_{2\alpha}^{2\alpha} \leq \lambda$. The largest eigenvalue of the matrix $B = Z^\star Z D$ is λ as well. Let

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_r \end{pmatrix}$$

be an eigenvector of B with eigenvalue λ. We have

$$\lambda x_i = \sum_{j=1}^{r} |\langle v_j, w\rangle|^{2\alpha-2} \langle v_j, v_i\rangle x_j,$$

for $i = 1, 2, \ldots, r$. Choose i such that $|x_i|$ is maximal. Then, we have

$$\lambda |x_i| \leq \sum_{j=1}^{r} |\langle v_j, w\rangle|^{2\alpha-2} |\langle v_j, v_i\rangle| |x_j| \leq \sum_{j=1}^{r} |\langle v_j, w\rangle|^{2\alpha-2} |\langle v_j, v_i\rangle| |x_i|.$$

If we set $\beta = \alpha/(\alpha-1)$ and $\gamma = \alpha$, then $1/\beta + 1/\alpha = 1$ and, by Hölder's inequality, we get

$$\begin{aligned}
[\mathbf{v}]_{2\alpha}^{2\alpha} &\leq \lambda \leq \sum_{j=1}^{r} |\langle v_j, w\rangle|^{2\alpha-2} |\langle v_j, v_i\rangle| \\
&\leq \left(\sum_{j=1}^{r} |\langle v_j, w\rangle|^{(2\alpha-2)\beta}\right)^{1/\beta} \left(\sum_{j=1}^{r} |\langle v_j, v_i\rangle|^{\alpha}\right)^{1/\alpha} \\
&\leq \left(\sum_{j=1}^{r} |\langle v_j, w\rangle|^{2\alpha}\right)^{(\alpha-1)/\alpha} \left(\mu_\alpha(\mathbf{v})^\alpha + 1\right)^{1/\alpha} = [\mathbf{v}]_{2\alpha}^{2\alpha-2} \left(\mu_\alpha(\mathbf{v})^\alpha + 1\right)^{1/\alpha}.
\end{aligned}$$

So, we conclude that $[\mathbf{v}]_{2\alpha}^{2\alpha} \leq \mu_\alpha(\mathbf{v})^\alpha + 1$. □

Example 3.6 The inequality in Proposition 3.5 does not hold when $\alpha < 1$. For example, if $\mathbf{e} = (e_1, e_2)$ in the (trivial) tensor product space $(\mathbb{C}^2, (\mathbb{C}^2))$, then we have $\mu_\alpha(\mathbf{e}) = 0$. If $0 < \alpha < 1$, then the maximum of

$$\left(|\langle e_1, u\rangle|^{2\alpha} + |\langle e_2, u\rangle|^{2\alpha}\right)^{1/(2\alpha)}$$

is attained for $u = (e_1 + e_2)/\sqrt{2}$. So, we have

$$[\mathbf{e}]_{2\alpha}^{2\alpha} = 2 \cdot 2^{-\alpha} = 2^{1-\alpha} > 1 = 1 + \mu_\alpha(\mathbf{v})^\alpha.$$

4 t-Orthogonality

In this section, we will discuss a notion of orthogonality for pure tensors that is stronger than the usual notion of orthogonality. Recall that an r-tuple $\mathbf{S} = (S_1, \ldots, S_r)$ of unit tensors is t-orthogonal if $[\mathbf{S}]_{2/t} = 1$.

Lemma 4.1 *Suppose that* $\mathbf{S} = (S_1, \ldots, S_r)$ *is* t*-orthogonal* r*-tuple of unit tensors and* $\mathbf{T} = (T_1, \ldots, T_r)$ *is an* u*-orthogonal* r*-tuple of unit tensors. Then,* $\mathbf{S} \otimes \mathbf{T}$ *is* $(t + u)$*-orthogonal.*

Proof From $[\mathbf{S}]_{2/t} = 1$, $[\mathbf{T}]_{2/e} = 1$ and Lemma 3.1, it follows that

$$1 \le [\mathbf{S} \otimes \mathbf{T}]_{2/(t+u)} \le [\mathbf{S}]_{2/t}[\mathbf{T}]_{2/u} \le 1$$

by Hölder's inequality. So, $[\mathbf{S} \otimes \mathbf{T}]_{2/(t+u)} = 1$ and $\mathbf{S} \otimes \mathbf{T}$ is $(t + u)$-orthogonal. □

Example 4.2 If $x_1, \ldots, x_r$ are orthogonal in $\mathbb{C}^p$, and $y_1, \ldots, y_r \in \mathbb{C}^q$ are orthogonal, then by Lemma 4.1,

$$(x_1 \otimes y_1, \ldots, x_r \otimes y_r) \in (\mathbb{C}^p \otimes \mathbb{C}^q)^r$$

is 2-orthogonal.

Orthogonality is also stable under taking vertical tensor products.

Lemma 4.3 *If* $\mathbf{S} = (S_1, \ldots, S_r)$ *is an* r*-tuple of unit tensors,* $\mathbf{T} = (T_1, \ldots, T_s)$ *is an* S*-tuple of unit tensors, and* $\mathbf{S}$ *and* $\mathbf{T}$ *are both* t*-orthogonal, then* $\mathbf{S} \boxtimes \mathbf{T}$ *is also* t*-orthogonal.*

Proof This follows from Proposition 3.3. □

Using horizontal and vertical tensor product, we can easily see that the tensor $M_{p,q,r}$ from Sect. 1.5 has a diagonal singular value decomposition.

Proposition 4.4 *We define*

$$\mathbf{e} = (e_{j,k} \otimes e_{k,i} \otimes e_{i,j} \mid 1 \le i \le p, 1 \le j \le q, 1 \le k \le r)$$
$$\times \in (\mathbb{C}^{p \times q} \otimes \mathbb{C}^{q \times r} \otimes \mathbb{C}^{r \times p})^{pqr}.$$

Then, $\mathbf{e}$ *is* 2*-orthogonal. In particular, the decomposition in (5) of* $M_{p,q,r}$ *is a diagonal singular value decomposition.*

Proof Let $e_1, \ldots, e_p$ denote the orthonormal basis of $\mathbb{C}^p$. Then, $(e_1, \ldots, e_p)$ is 1-orthogonal. So,

$$(e_i \otimes e_i \mid i = 1, \ldots, p) \in (\mathbb{C}^p \otimes \mathbb{C}^p)^p$$

is 2-orthogonal. The tuples

$$\begin{aligned} P &= (e_i \otimes 1 \otimes e_i \mid 1 \leq i \leq p) \in (\mathbb{C}^p \otimes \mathbb{C} \otimes \mathbb{C}^p)^p, \\ Q &= (e_j \otimes e_j \otimes 1 \mid 1 \leq j \leq q) \in (\mathbb{C}^q \otimes \mathbb{C}^q \otimes \mathbb{C})^q, \\ R &= (1 \otimes e_k \otimes e_k \mid 1 \leq k \leq r) \in (\mathbb{C} \otimes \mathbb{C}^r \otimes \mathbb{C}^r)^r \end{aligned}$$

are 2-orthogonal. We will write $e_{i,j}$ instead of $e_i \boxtimes e_j$. Now, the tuple

$$\begin{aligned} P \boxtimes Q \boxtimes R = (e_{i,j} \otimes e_{j,k} \otimes e_{k,i} \mid 1 \leq i \leq p, 1 \leq j \leq q, 1 \leq k \leq r) \\ \times \in (\mathbb{C}^{q \times r} \otimes \mathbb{C}^{r \times p} \otimes \mathbb{C}^{p \times q})^{pqr} \end{aligned}$$

is 2-orthogonal as well. □

Proof of Theorem 1.14 This follows immediately from the definition of the diagonal singular value decomposition and Proposition 4.4. □

Proof of Theorem 1.16 Let $Z_1, \ldots, Z_s$ be the irreducible representations of G. We have an isomorphism

$$\mathbb{C}G \cong \bigoplus_{i=1}^{s} \mathrm{Hom}(Z_i, Z_i). \tag{11}$$

For $A = \sum_{g \in G} \lambda_g g \in \mathbb{C}G$, we define $A^\star = \sum_{g \in G} \overline{\lambda_g} g^{-1}$. We may view A as an endomorphism of $\mathbb{C}G$ by left multiplication. The Hermitian form on $\mathbb{C}G$ is given by

$$\langle A, B \rangle = \tfrac{1}{n} \,\mathrm{trace}(AB^\star).$$

We can write

$$e = \sum_{i=1}^{s} \pi_i$$

where π_i is the projection onto Z_i. The decomposition (11) is orthogonal. The multiplication tensor

$$T_G \in \mathbb{C}G \otimes \mathbb{C}G \otimes \mathbb{C}G$$

decomposes

$$T_G = \sum_{i=1}^{s} T_i$$

where

$$T_i \in \mathrm{Hom}(Z_i, Z_i) \otimes \mathrm{Hom}(Z_i, Z_i) \otimes \mathrm{Hom}(Z_i, Z_i) \subseteq \mathbb{C}G \otimes \mathbb{C}G \otimes \mathbb{C}G$$

is the tensor for multiplication in $\mathrm{Hom}(Z_i, Z_i)$. We have

$$T_i = \sum_{g,h} \pi_i g \otimes h \otimes h^{-1} g^{-1}.$$

So, it follows that

$$\begin{aligned}\|T_i\|^2 &= \frac{1}{n} \sum_{g,h} \mathrm{trace}((\pi_i g)^\star (\pi_i g)) = \frac{1}{n} \sum_{g,h} \mathrm{trace}(\pi_i^\star g^{-1} g \pi_i) = \frac{1}{n} \sum_{g,h} \mathrm{trace}(\pi_i^2) \\ &= n\,\mathrm{trace}(\pi_i) = n d_i^2.\end{aligned}$$

Note that $\mathbf{T} = (T_1, T_2, \ldots, T_s)$ is 3-orthogonal. The tensor T_i corresponds to matrix multiplication in $\mathrm{Hom}(Z_i, Z_i)$. We can write

$$T_i = \sum_{j=1}^{d_i^3} \lambda_j a_j \otimes b_j \otimes c_j,$$

where $(a_j \otimes b_j \otimes c_j, 1 \leq j \leq d_i^3)$ is 2-orthogonal list of unit vectors. The norm on $\mathrm{Hom}(Z_i, Z_i)$ that is induced from the norm on $\mathbb{C}G$ may not be the same as the Euclidean norm given by $A \mapsto \mathrm{trace}(A^\star A)$, but they are the same up to a scalar. This implies that all λ_j's are the same. Since $n d_i^2 = \|T_i\|^2 = d_i^3 \lambda_j^2$, we have that $\lambda_j = \sqrt{\frac{n}{d_i}}$ for all j. So, the irreducible representation Z_i contributes the singular value $\sqrt{\frac{n}{d_i}}$ with multiplicity d_i^3. □

Proposition 4.5 *Suppose that* $\mathbf{V} = (V, (V^{(1)}, \ldots, V^{(d)}))$ *is a tensor product space,* $t \geq 1$ *and* $\mathbf{v} = (v_1, \ldots, v_r) \in V^r$ *is a* t*-orthogonal* r*-tuple of pure tensors of unit length. Then, we have* $r \leq \dim(V)^{1/t}$.

Proof Consider the $(2m-1)$-dimensional unit sphere in $\mathbb{C}^{2m}$ given by

$$|z_1|^2 + \cdots + |z_m|^2 = 1.$$

Suppose z is a random point on the sphere (with uniform distribution). We will give an estimate for the expectation $\mathbb{E}(|z|^\alpha)$. It is clear that $\mathbb{E}(|z|^\alpha) \geq \mathbb{E}(|w|^\alpha)$, where w is a random point in the $(2m)$-dimensional ball B_{2m} defined by

$$|w_1|^2 + \cdots + |w_m|^2 \leq 1.$$

Let $x = 1/\sqrt{2m}$, and let D be the body defined by

$$|w_1| \leq x \text{ and } |w_1|^2 + \cdots + |w_m|^2 \leq 1$$

and E be the body defined by

$$|w_1| \leq x \text{ and } |w_2|^2 + \cdots + |w_m|^2 \leq 1.$$

Then, $D \subseteq E$ and E is a product of an $(2m-2)$-dimensional ball with radius 1 and a disk of radius x. We have

$$\mathbb{P}(|w| \leq x) = \frac{\mathrm{vol}(D)}{\mathrm{vol}(B_{2m})} \leq \frac{\mathrm{vol}(E)}{\mathrm{vol}(B_{2m})} = \frac{\mathrm{vol}(B_{2m-2})\pi x^2}{\mathrm{vol}(B_{2m})} = \frac{x^2}{m} = \frac{1}{2},$$

where we use the formula $\mathrm{vol}(B_{2m}) = \pi^m/m!$. It follows that

$$\mathbb{E}(|z|^\alpha) \geq \mathbb{E}(|w|^\alpha) \geq x^\alpha \mathbb{P}(|w| \geq x) \geq \tfrac{1}{2}x^\alpha = 2^{-1-\alpha/2}m^{-\alpha/2}.$$

Suppose that $u = u^{(1)} \otimes \cdots \otimes u^{(d)}$ is a fixed unit pure tensor in V and $z = z^{(1)} \otimes \cdots \otimes z^{(d)}$ is a random unit pure tensor. Let $n = \dim(V)$ and $n_i = \dim(V_i)$ for all i. Then, we have

$$\mathbb{E}(|\langle z, u\rangle|^{2/t}) = \prod_{s=1}^{d} \mathbb{E}\left(|\langle z^{(i)}, u^{(i)}\rangle|^{2/t}\right) \geq 2^{-(1+1/t)d} \prod_{i=1}^{d} n_i^{-1/t} = 2^{-(1+1/t)d} n^{-1/t}.$$

It follows that

$$1 \geq \sum_{i=1}^{r} \mathbb{E}\left(|\langle v_i, z\rangle|^{2/t}\right) = 2^{-(1+1/t)d} n^{-1/t} r.$$

So, we get

$$r \leq 2^{(1+1/t)d} n^{1/t}.$$

For a positive integer q,

$$v^{\boxtimes q} = \underbrace{v \boxtimes \cdots \boxtimes v}_{q}$$

is also t-orthogonal and it has r^q vectors in an n^q-dimensional vector space. So, we have

$$r^q \leq 2^{(1+1/t)d} n^{q/t}.$$

Taking the q-th root gives

$$r \leq 2^{(1+1/t)d/q} n^{1/t}.$$

Taking the limit $q \to \infty$ yields

$$r \leq n^{1/t}.$$

□

The following lemma justifies the term t-orthogonality.

Lemma 4.6 *If (v, w) is t-orthogonal, where $v = v^{(1)} \otimes \cdots \otimes v^{(d)}$ and $w = w^{(1)} \otimes \cdots \otimes w^{(d)}$, then we have $\langle v^{(i)}, w^{(i)} \rangle = 0$ for at least $\lceil t \rceil$ values of i.*

Proof Suppose that $u = u^{(1)} \otimes \cdots \otimes u^{(d)}$ is a unit pure tensor. Then, we have

$$1 \geq \prod_i |\langle v^{(i)}, u^{(i)} \rangle|^{2/t} + \prod_i |\langle w^{(i)}, u^{(i)} \rangle|^{2/t}.$$

Choose ε with $0 < \varepsilon < \sqrt{2}$ and $u^{(i)}$ such that $|\langle v^{(i)}, u^{(i)} \rangle| = 1 - \frac{1}{2}\varepsilon^2$ and $u^{(i)}, v^{(i)}, w^{(i)}$ are dependent. If $w^{(i)}$ and $v^{(i)}$ are orthogonal, then $|\langle w^{(i)}, u^{(i)} \rangle| = \varepsilon + o(\varepsilon)$, because $|\langle v^{(i)}, u^{(i)} \rangle|^2 + |\langle w^{(i)}, u^{(i)} \rangle|^2 = 1$. If $w^{(i)}$ and $v^{(i)}$ are not orthogonal, then $|\langle w^{(i)}, u^{(i)} \rangle| = |\langle w^{(i)}, v^{(i)} \rangle| + o(\varepsilon)$. If s is the number of i for which $v^{(i)}$ and $w^{(i)}$ are orthogonal, then we have

$$1 \geq \prod_i |\langle v^{(i)}, u^{(i)} \rangle|^{2/t} + \prod_i |\langle w^{(i)}, u^{(i)} \rangle|^{2/t} = \left(1 - \tfrac{1}{2}\varepsilon^2\right)^{2/t} + |C\varepsilon^s + o(\varepsilon^s)|^{2/t}$$
$$= 1 - \tfrac{d}{t}\varepsilon^2 + o(\varepsilon^2) + C^{2/t}\varepsilon^{2s/t} + o(\varepsilon^{2s/t}),$$

for some constant C. We must have $s/t \geq 1$; otherwise, the inequality is not satisfied for small ε. □

Example 4.7 Consider the following triple of pure tensors

$$\mathbf{e} = (e_1 \otimes e_1 \otimes e_1, e_1 \otimes e_2 \otimes \otimes e_2, e_2 \otimes e_1 \otimes e_2) \in (\mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2)^3.$$

Then, every pair of vectors of $\mathbf{e}$ is 2-orthogonal. However, $\mathbf{e}$ itself is not 2-orthogonal, because it violates Proposition 4.5:

$$n = 3 > 8^{1/2} = \dim(V)^{1/2}.$$

Example 4.8 Let G be a group of order n, $\mathbb{C}G$ be its group algebra and consider the tensor product space $\mathbb{C}G \otimes \mathbb{C}G \otimes \mathbb{C}G$. Let

$$\mathbf{v} = (g \otimes h \otimes k \mid g, h, k \in G;\ ghk = 1)$$

be a list of n^2 vectors. We claim that $\mathbf{v}$ is $\frac{3}{2}$-orthogonal. Suppose that

$$w = \left(\sum_{g\in G} a_g g\right) \otimes \left(\sum_{g\in G} b_g g\right) \otimes \left(\sum_{g\in G} c_g g\right)$$

is a pure tensor with $\sum_{g\in G} |a_g|^2 = \sum_{g\in G} |b_g|^2 = \sum_{g\in G} |c_g|^2 = 1$. Using the inequality $pqr \leq \frac{1}{3}(p^3 + q^3 + r^3)$, we get

$$\begin{aligned}\sum_{\substack{g,h,k\in G\\ gkh=1}} |\langle g\otimes h\otimes k, w\rangle|^{4/3} &= \sum_{\substack{g,h,k\in G\\ ghk=1}} |a_g b_h c_k|^{4/3} = \sum_{\substack{g,h,k\in G\\ ghk=1}} |a_g b_h|^{2/3} |b_h c_k|^{2/3} |c_k a_g|^{2/3}\\ &\leq \frac{1}{3} \sum_{\substack{g,h,k\in G\\ ghk=1}} \left(|a_g b_h|^2 + |b_h c_k|^2 + |c_k a_g|^2\right) = 1.\end{aligned}$$

To see the last equality, note that

$$\sum_{\substack{g,h,k\in G\\ ghk=1}} |a_g b_h|^2 = \sum_{g\in G} |a_g|^2 \sum_{h\in G} |b_h|^2 = 1\cdot 1 = 1.$$

This proves that $\mathbf{v}$ is $\frac{3}{2}$-orthogonal. For $t > \frac{3}{2}$, $\mathbf{v}$ cannot be t-orthogonal because otherwise this would violate Proposition 4.5.

5 Lower Bounds for the Nuclear Norm

Proof of Theorem 1.4 Suppose that $\alpha \geq 1$, T is a tensor and $\mathbf{S} = (S_1, \dots, S_r)$ is an r-tuple of tensors. We can write

$$T = \sum_{i=1}^{s} \mu_j w_j,$$

where $\mu_1, \dots, \mu_s$ are positive real numbers such that $\sum_{j=1}^{s} \mu_j = \|T\|_\star$ and $w_1, \dots, w_s$ are pure unit tensors. Define

$$z_j := \begin{pmatrix} |\langle w_j, S_1\rangle| \\ \vdots \\ |\langle w_j, S_r\rangle| \end{pmatrix} \in \mathbb{C}^r$$

for $j = 1, 2, \ldots, s$. We have

$$\|z_j\|_\alpha = \left(\sum_{i=1}^{r} |\langle w_j, S_i\rangle|^\alpha\right)^{1/\alpha} \leq [\mathbf{S}]_\alpha.$$

for all j. It follows that

$$\left(\sum_{i=1}^{r} |\langle T, S_i\rangle|^\alpha\right)^{1/\alpha} \leq \left(\sum_{i=1}^{r}\left(\sum_{j=1}^{s} \mu_j |\langle w_j, S_i\rangle|\right)^\alpha\right)^{1/\alpha} = \left\|\sum_{j=1}^{s} \mu_j z_j\right\|_\alpha$$
$$\leq \sum_{j=1}^{s} |\mu_j| \|z_j\|_\alpha \leq \sum_{j=1}^{s} |\mu_j| [\mathbf{S}]_\alpha = \|T\|_\star [\mathbf{S}]_\alpha.$$

□

For a permutation $\sigma \in \mathbb{S}_n$, define

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

and

$$\mathbf{e} = (e_\sigma, \sigma \in \mathbb{S}_n).$$

We now study the determinant tensor $\sum_\sigma \operatorname{sgn}(\sigma) e_\sigma$ and the permanent tensor $\sum_\sigma e_\sigma$.

Proof of Theorem 1.20 If $a^{(1)}, \cdots, a^{(n)}$ are vectors of unit length, then Hadamard's inequality yields

$$|\langle \det_n, a^{(1)} \otimes a^{(2)} \otimes \cdots \otimes a^{(n)}\rangle| = |\det(a^{(1)}, \ldots, a^{(n)})| \leq \|a^{(1)}\| \cdots \|a^{(n)}\| = 1.$$

Therefore, we have $[\det_n] \leq 1$. It follows from Corollary 1.5 that

$$\|\det_n\|_\star \geq \|\det_n\|_\star [\det_n] \geq \|\det_n\|^2 = n!\,.$$

□

The following theorem proven in [7] is the permanent analog of Hadamard's inequality.

Theorem 5.1 *For vectors $a^{(1)}, \ldots, a^{(n)} \in \mathbb{C}^n$, we have*

$$|\operatorname{per}(A)| \leq \frac{n!}{n^{n/2}} \|a^{(1)}\| \|a^{(2)}\| \cdots \|a^{(n)}\|.$$

Proof of Theorem 1.19 For vectors $a^{(1)}, \cdots, a^{(n)}$ of unit length, we get

$$|\langle \mathrm{per}_n, a^{(1)} \otimes a^{(2)} \otimes \cdots \otimes a^{(n)} \rangle| = |\,\mathrm{per}(a^{(1)}, \ldots, a^{(n)})| \leq \frac{n!}{n^{n/2}} \|a^{(1)}\| \cdots \|a^{(n)}\| = \frac{n!}{n^{n/2}}.$$

So, we have $[\mathrm{per}_n] \leq \frac{n!}{n^{n/2}}$. From Corollary 1.5, it follows that

$$\frac{n!}{n^{n/2}} \|\mathrm{per}_n\|_\star \geq \|\mathrm{per}_n\|_\star [\mathrm{per}_n] \geq \|\mathrm{per}_n\|^2 = n!.$$

We conclude that $\|\,\mathrm{per}_n\,\|_\star \geq n^{n/2}$. □

6 The Diagonal Singular Value Decomposition

For an r-tuple $\mathbf{v} = (v_1, \ldots, v_r)$ and $k < r$, we write $\mathbf{v}^{[k]}$ for $(v_1, \ldots, v_k)$. We start with the most general, main theorem.

Theorem 6.1 *Suppose that V is a tensor product space, $\mathbf{v} = (v_1, \ldots, v_r)$ and $\mathbf{w} = (w_1, \ldots, w_s)$ consists of pure tensors in V of unit length, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_s > 0$, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ and*

$$\sum_{i=1}^{s} \lambda_i w_i = \sum_{j=1}^{r} \sigma_j v_j.$$

Also, suppose that $k \leq s$, $l \leq r$ such that $0 \leq \delta \leq [\mathbf{w}^{[k]}]_1$, where $\delta := k[\mathbf{v}]_1 - l[\mathbf{w}^{[k]}]_1$. Then, we have

$$[\mathbf{w}^{[k]}]_1(\sigma_1 + \sigma_2 + \cdots + \sigma_l) + \delta\sigma_{l+1} \geq (1 - \mu_1(\mathbf{w}))(\lambda_1 + \lambda_2 + \cdots + \lambda_k).$$

Here, we use the conventions that $0 = \lambda_{s+1} = \lambda_{s+2} = \cdots$ and $0 = \sigma_{r+1} = \sigma_{r+2} = \cdots$.

Proof Let

$$T = \sum_{i=1}^{s} \lambda_i w_i = \sum_{j=1}^{r} \sigma_j v_j.$$

We have

$$\sum_{i=1}^{k}\sum_{j=1}^{s} \lambda_i |\langle w_i, w_j\rangle| = \sum_{i=1}^{k}\sum_{j=1}^{k} \lambda_i |\langle w_i, w_j\rangle| + \sum_{i=1}^{k}\sum_{j=k+1}^{s} \lambda_i |\langle w_i, w_j\rangle|$$

$$= \sum_{i=1}^{k}\sum_{j=1}^{k}\lambda_j|\langle w_i, w_j\rangle| + \sum_{i=1}^{k}\sum_{j=k+1}^{s}\lambda_i|\langle w_i, w_j\rangle|$$
$$\geq \sum_{i=1}^{k}\sum_{j=1}^{k}\lambda_j|\langle w_i, w_j\rangle| + \sum_{i=1}^{k}\sum_{j=k+1}^{s}\lambda_j|\langle w_i, w_j\rangle|$$
$$= \sum_{i=1}^{k}\sum_{j=1}^{s}\lambda_j|\langle w_i, w_j\rangle|.$$

because $\lambda_i \geq \lambda_j$ whenever $i \geq j$. Using this, we get

$$\sum_{i=1}^{k}|\langle w_i, T\rangle| = \sum_{i=1}^{k}\left|\sum_{j=1}^{s}\lambda_j\langle w_i, w_j\rangle\right| \geq \sum_{i=1}^{k}\lambda_i - \sum_{i=1}^{k}\sum_{j\neq i}\lambda_j|\langle w_i, w_j\rangle|$$
$$= 2\sum_{i=1}^{k}\lambda_i - \sum_{i=1}^{k}\sum_{j=1}^{s}\lambda_j|\langle w_i, w_j\rangle| \geq 2\sum_{i=1}^{k}\lambda_i - \sum_{i=1}^{k}\sum_{j=1}^{s}\lambda_i|\langle w_i, w_j\rangle|$$
$$= \sum_{i=1}^{k}\lambda_i\left(2 - \sum_{j=1}^{s}|\langle w_i, w_j\rangle|\right) \geq (1-\mu_1(\mathbf{w}))\sum_{i=1}^{k}\lambda_i.$$

Let $y_{i,j} = |\langle w_i, v_j\rangle|$ if $1 \leq i \leq s$ and $1 \leq j \leq r$. We have

$$(1-\mu_1(\mathbf{w}))\sum_{i=1}^{k}\lambda_i \leq \sum_{i=1}^{k}|\langle w_i, T\rangle| \leq \sum_{i=1}^{k}\sum_{j=1}^{r}\sigma_j y_{i,j} = \sum_{j=1}^{r}\sigma_j\sum_{i=1}^{k}y_{i,j} = \sum_{j=1}^{r}\sigma_j x_j,$$

where $x_j = \sum_{i=1}^{k} y_{i,j} \leq [\mathbf{w}^{[k]}]_1$. We also have

$$x_1 + \cdots + x_r = \sum_{j=1}^{r}\sum_{i=1}^{k}y_{i,j} = \sum_{i=1}^{k}\sum_{j=1}^{r}y_{i,j} \leq \sum_{i=1}^{k}[\mathbf{v}]_1 = k[\mathbf{v}]_1.$$

If we maximize the functional $\sum_{j=1}^{r}\sigma_j x_j$ under the constraints $0 \leq x_i \leq [\mathbf{w}^{[k]}]_1$ for $i = 1, 2, \ldots, l$ and $x_1 + \cdots + x_r \leq k[\mathbf{v}]_1$, then an optimal solution is $x_1 = x_2 = \cdots = x_l = [\mathbf{w}^{[k]}]_1$, $x_{l+1} = k[\mathbf{v}]_1 - l[\mathbf{w}^{[k]}]_1 = \delta$ and $x_{l+2} = \cdots = x_r = 0$, and the optimal value is

$$[\mathbf{w}^{[k]}]_1(\sigma_1 + \cdots + \sigma_l) + \delta\sigma_{l+1}.$$

□

The following result gives a lower bound for the nuclear norm:

Theorem 6.2 *If* $\mathbf{w} = (w_1, \ldots, w_s)$ *is an orthogonal* r*-tuple of pure tensors of unit length,* $\lambda_1 \geq \cdots \geq \lambda_s > 0$ *and* $T = \sum_{i=1}^{s}\lambda_i w_i$*, then we have*

$$\|T\|_\star \geq \frac{\sum_{i=1}^k \lambda_i}{\left[\mathbf{w}^{[k]}\right]_1}.$$

Proof We can write $T = \sum_{i=1}^r \sigma_i v_i$, where v_i is a pure tensor of unit length for all i, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ and $\|T\|_\star = \sum_{i=1}^r \sigma_i$. We have

$$\sum_{i=1}^s \lambda_i w_i = \sum_{j=1}^r \sigma_j v_j$$

and $\mu_1(\mathbf{w}) = 0$ because $\mathbf{w}$ is orthogonal. From Theorem 6.1, it follows that

$$[\mathbf{w}^{[k]}]_1(\sigma_1 + \cdots + \sigma_r) \geq \lambda_1 + \cdots + \lambda_s.$$

□

Proof of Theorem 1.10 Suppose that $\mathbf{v} = (v_1, \ldots, v_r)$ and $\mathbf{w} = (w_1, \ldots, w_s)$ are 2-orthogonal tuples of pure tensors of unit length, and

$$\sum_{i=1}^s \lambda_i w_i = \sum_{j=1}^r \sigma_j v_j,$$

such that $\lambda_1 \geq \cdots \geq \lambda_s > 0$ and $\sigma_1 \geq \cdots \geq \sigma_r > 0$. We apply Theorem 6.1 with $[\mathbf{v}]_1 = [\mathbf{w}^{[k]}]_1 = 1$, $l = k$ and get

$$\sum_{j=1}^k \sigma_j \geq \sum_{i=1}^k \lambda_i$$

for all k. If we switch the roles of the v's and w's, we also get inequalities in the other directions as well. We conclude that $r = s$ and $\lambda_i = \mu_i$ for all i. □

Proof of Theorem 1.11 Suppose that the DSVD of T is

$$T = \sum_{i=1}^r \sigma_i v_i.$$

Then, we have $\|T\|_\star \leq \sum_{i=1}^r \sigma_i$. If we take $k = r$ and $\lambda_i = \sigma_i$ in Theorem 6.2, then we get

$$\|T\|_\star \geq \frac{\sum_{i=1}^r \sigma_i}{[\mathbf{v}]_1} = \sum_{i=1}^r \sigma_i,$$

so we conclude that $\|T\|_\star = \sum_{i=1}^r \sigma_i$.

Since $\mathbf{v} = (v_1, \ldots, v_r)$ is 2-orthogonal, we have

$$\|T\|^2 = \sum_{i=1}^{r} \sigma_i^2$$

and $[\mathbf{v}]_1 = 1$.

If u is a pure tensor of unit length, then

$$|\langle T, u\rangle| = \left|\sum_{i=1}^{r} \sigma_i \langle v_i, u\rangle\right| \leq \sigma_1 \sum_{i=1}^{r} |\langle v_i, u\rangle| \leq \sigma_1[\mathbf{v}]_1 = \sigma_1.$$

Clearly $\langle T, v_1\rangle = \sigma_1$. So, we conclude that $[T] = \sigma_1$. □

Proof of Theorem 1.12 Suppose that T has a DSVD with singular values $\sigma_1 > \cdots > \sigma_r > 0$. We can write $T = \sum_{j=1}^{r} \sigma_j v_j$. Suppose that we have another singular value decomposition $T = \sum_{i=1}^{r} \sigma_i w_i$ (note that the singular values are determined by T because of Theorem 1.10).

Let $y_{i,j} = |\langle w_i, v_i\rangle|$. Then, $\sum_{j=1}^{r} y_{i,j} \leq [\mathbf{v}]_1 = 1$ and $\sum_{i=1}^{r} y_{i,j} \leq [\mathbf{w}]_1 = 1$. Fix $k \leq r$ and let $x_j = \sum_{i=1}^{k} y_{i,j} \leq 1$. From the proof of Theorem 6.1, it follows that

$$x_1 + \cdots + x_r \leq k$$

and

$$\sum_{i=1}^{k} \sigma_i \leq \sum_{j=1}^{r} \sigma_j x_j.$$

Since $\sigma_1, \ldots, \sigma_r$ are distinct, we must have $x_1 = x_2 = \cdots = x_k = 1$ and $x_{k+1} = \cdots = x_r = 0$. This implies that $y_{i,j} = 0$ if $i \leq k$ and $j \geq k+1$. So, $y_{i,j} = 0$ for $i < j$ and by symmetry, $y_{i,j} = 0$ for $i > j$. This proves that $|\langle v_i, w_i\rangle| = y_{i,i} = 1$ for all i. So, w_i is equal to v_i up to a unit scalar, say $w_i = \gamma_i v_i$. It follows that

$$\sum_{i=1}^{r} \sigma_i v_i = \sum_{i=1}^{r} \sigma_i w_i = \sum_{i=1}^{r} \sigma_i \gamma_i v_i$$

and because $v_1, \ldots, v_r$ are linearly independent, it follows that $\gamma_i = 1$ and $w_i = v_i$ for all i. □

Proof of Theorem 1.13 Suppose that T is a tensor with two diagonal singular value decompositions

$$T = \sum_{i=1}^{r} \sigma_i w_i = \sum_{j=1}^{r} \sigma_j v_j$$

with $\sigma_1 \geq \cdots \geq \sigma_r > 0$, and that $\mathbf{w} = (w_1, \ldots, w_r)$ is t-orthogonal with $t > 2$. Let $y_{i,j} = |\langle w_i, v_j\rangle|$.

From the proof of Theorem 6.1, it follows that

$$\sum_{i=1}^{r} y_{i,j} = 1$$

for all j. We also have

$$\sum_{i=1}^{r} y_{i,j}^{\alpha} \leq 1,$$

where $\alpha = 2/t < 1$, because $\mathbf{w}$ is t-orthogonal. Subtracting gives

$$\sum_{i=1}^{r} y_{i,j}^{\alpha}(1 - y_{i,j}^{1-\alpha}) \leq 0.$$

It follows that $y_{i,j} \in \{0, 1\}$ for all i, j. The column sums of $Y = (y_{i,j})$ are 1. So every column has exactly one 1. So, the matrix has exactly r 1's. Since the row sums are also 1, it follows that every row has exactly one 1 as well. So, Y is a permutation matrix. There exists a permutation ϕ of $\{1, 2, \ldots, r\}$ such that

$$v_i = \gamma_i w_{\phi(i)},$$

where γ_i is a unit for all i. We have

$$\sum_{i=1}^{r} \sigma_i v_i = \sum_{i=1}^{r} \gamma_i \sigma_i w_{\phi(i)} = \sum_{i=1}^{r} \sigma_{\phi(i)} w_{\phi(i)}.$$

Since $\mathbf{w}$ is linearly independent, it follows that $\gamma_i \sigma_i = \sigma_{\phi(i)}$ for all i. So $\gamma_i = 1$ and $\sigma_i = \sigma_{\phi(i)}$ for all i. This shows that

$$\sigma_1 v_1, \ldots, \sigma_r v_r$$

is a permutation of

$$\sigma_1 w_1, \ldots, \sigma_r w_r.$$

So, the diagonal DSVD is unique. □

7 Tensors Without a Diagonal Singular Value Decomposition

Example 7.1 Consider the permanent per_n. Suppose that it has a DSVD and that its singular values are $\sigma_1, \ldots, \sigma_r$. Then, we have

$$[\mathrm{per}_n] = \sigma_1 = \frac{n!}{n^{n/2}}, \quad \| \mathrm{per}_n \|^2 = n!, \quad \| \mathrm{per}_n \|_\star = n^{n/2}.$$

We have

$$\sigma_1 \sum_{i=1}^{r} \sigma_i = [\mathrm{per}_n] \| \mathrm{per}_n \|_\star = n! = \| \mathrm{per}_n \|^2 = \sum_{i=1}^{r} \sigma_i^2,$$

so it follows that $\sigma_1 = \sigma_2 = \cdots = \sigma_r$. So

$$\frac{n^n}{n!} = \frac{\| \mathrm{per}_n \|_\star}{[\mathrm{per}_n]} = \frac{r\sigma_1}{\sigma_1} = r.$$

For $n \geq 3$, $n^n/n!$ is not an integer (the denominator is divisible by $n-1$), so per_n cannot have a diagonal singular value decomposition.

Example 7.2 Consider the determinant $\det_n$. Suppose that $\det_n$ has a DSVD. A similar argument as in the previous example shows that $\det_n$ has a singular value σ with multiplicity r, where

$$r = \frac{\| \det_n \|_\star}{[\det_n]} = n!.$$

So, there exists a 2-orthogonal r-tuple of pure tensors of unit length. This implies that $r \leq n^{n/2}$ by Proposition 4.5. For $n \geq 3$, we have $n! > n^{n/2}$, so $\det_n$ cannot have a diagonal singular value decomposition.

Acknowledgments The author thanks Zach Teitler for useful comments and a correction.

8 Appendix: The Tensor Rank of the Determinant and the Permanent

For a subset $I = \{i_1, i_2, \ldots, i_r\} \subseteq \{1, 2, \ldots, n\}$ with $i_1 < \cdots < i_r$, define

$$\det_r(I) = \sum \mathrm{sgn}(\sigma) e_{i_{\sigma(1)}} \otimes \cdots \otimes e_{i_{\sigma(r)}},$$
$$\mathrm{sgn}(I) = (-1)^{i_1 + \cdots + i_r - \binom{r+1}{2}},$$

and

$$I^c = \{1, 2, \ldots, n\} \setminus I.$$

We have the following generalized Laplace expansion

$$\det_n = \sum_I \mathrm{sgn}(I) \det_r(I) \otimes \det_{n-r}(I^c),$$

where I runs over all $\binom{n}{r}$ subsets of $\{1, 2, \ldots, n\}$ with cardinality r.

By flattening, we may view the tensor $\det_n$ as a tensor in $\mathbb{C}^{n^r} \otimes \mathbb{C}^{n^{n-r}}$. The tensors $\det_r(I)$ where I is a subset of $\{1, 2, \ldots, n\}$ with r elements are linearly independent.

The tensors $\det_r(I^c)$ are linearly independent as well. This shows that the flattened tensor has rank at least $\binom{n}{r}$. So, we have

$$\operatorname{rank}(\det_n) \geq \binom{n}{r}.$$

We get the best lower bound if $r = \lfloor n/2 \rfloor$:

$$\operatorname{rank}(\det_n) \geq \binom{n}{\lfloor n/2 \rfloor}.$$

We have a similar Laplace expansion for the permanent, so we also get

$$\operatorname{rank}(\operatorname{per}_n) \geq \binom{n}{\lfloor n/2 \rfloor}.$$

So the ranks of the determinant and permanent grow at least exponentially. We also have an exponential lower bound for the permanent. An exponential upper bound for the rank of the determinant seems not to be known. However, the obvious bound $\operatorname{rank}(\det_n) \leq n!$ is not sharp for $n \geq 3$.

For $n = 3$, we have

$$\begin{aligned}\det_3 = \tfrac{1}{2}\Big(&(e_3 + e_2) \otimes (e_1 - e_2) \otimes (e_1 + e_2) + (e_1 + e_2) \otimes (e_2 - e_3) \otimes (e_2 + e_3)\\ &+2e_2 \otimes (e_3 - e_1) \otimes (e_3 + e_1)\\ &+(e_3 - e_2) \otimes (e_2 + e_1) \otimes (e_2 - e_1) + (e_1 - e_2) \otimes (e_3 + e_2) \otimes (e_3 - e_2)\Big).\end{aligned}$$

So $\operatorname{rank}(\det_3) \leq 5$. Zach Teitler pointed out that this implies that the Waring rank of a 3×3 matrix is at most 20. He also pointed out that one can show that $\operatorname{rank}(\det_3) \geq 4$. If $n > 3$, then we can again use the generalized Laplace expansion

$$\det_n = \sum_I \operatorname{sgn}(I) \det_3(I) \otimes \det_{n-3}(I^c),$$

where I runs over all subsets of $\{1, 2, \ldots, n\}$ with three elements. This proves that

$$\operatorname{rank}(\det_n) \leq \binom{n}{3} \operatorname{rank}(\det_{n-3}) \operatorname{rank}(\det_3) \leq \frac{5 \cdot n!}{6 \cdot (n-3)!}$$

We can rewrite this as

$$\frac{\operatorname{rank}(\det_n)}{n!} \leq \frac{5}{6} \cdot \frac{\operatorname{rank}(\det_{n-3})}{(n-3)!}.$$

By induction, we get

$$\operatorname{rank}(\det_n) \leq \left(\frac{5}{6}\right)^{\lfloor \frac{n}{3} \rfloor} \cdot n!.$$

Remark 8.1 Homogeneous polynomials can be thought of as symmetric tensors. For symmetric tensors, there is also a notion of rank, the so-called *symmetric rank*. The symmetric rank is different from, but closely related to the tensor rank. The determinant and permanent can be thought of as homogeneous polynomials. Lower bounds for the symmetric tensor rank of the determinant and permanent can be found in [23] and [31].

References

1. C. D. Aliprantis and K. C. Border, *Infinite Dimensional Analysis – A Hitchhiker's Guide*, third edition, Springer, 2006.
2. D. Bini, M. Capovani, G. Lotti and F. Romani, *$O(n^{2.7799})$ complexity for matrix multiplication*, Inf. Proc. Letters **8** (1979), 234–235.
3. M. Bläser, *Beyond the Alder-Strassen bound*, Theor. Comp. Science **331** (2005), 3–21.
4. P. Bürgisser, M. Clausen and M. A. Shokrollahi, *Algebraic Complexity Theory*, Grundlehren der Mathematischen WIssenschaften **315**, Springer-Verlag, Berlin, 1997.
5. E.J. Candès and B. Recht, Exact matrix completion via convex optimization, Foundations of Computational Mathematics 9, 2009, 717–772.
6. E. J. Candès and T. Tao, The power of convex relaxation: Near-optimal matrix completion, IEEE Trans. on Information Theory 56 (2010), no. 5, 2053–2080.
7. E. Carlen, E. H. Lieb and M. Loss (2006) An inequality of Hadamard type for permanents, Methods Appl. Anal. 13(1) 1–17.
8. J. D. Carroll and J. Chang, Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition, Psychometrika 35 (1970), 218–319.
9. D. Coppersmith and S. Winograd, Matrix Multiplication via Arithmetic Progessions, J. Symbolic Computation, 9 (3) (1990), 251–280.
10. V. De Silva and L.-H. Lim, Tensor rank and the ill-posedness of the best low-rank approximation problem, SIAM J. Matrix Anal. Appl. 30 (2008), 1254–1279.
11. S. Gandy, B. Recht and I. Yamada, Tensor completion and low n-rank tensor recovery via convex optimization, Inverse Problems 27 (2011), no. 2.
12. D. Glynn, The permanent of a square matrix, European J. Combin. 31 (2010), no. 7, 1887–1891.
13. A. Grothendieck, Produits tensoriels topologiques et espaces nucléares, Mem. Amer. Math. Soc. 1955 (1955), no. 16.
14. R. A. Harshman, Foundations of the PARAFAC procedure: models and conditions for an "explanatory" multimodal factor analysis, UCLA working papers in Phonetics 16 (1970), 1–84.
15. J. Håstad, Tensor rank is NP-complete, J. Algorithms 11 (1990), no. 4, 644–654.
16. J. Håstad, *Tensor rank is NP-complete*, Automata, languages and programming (Stresa, 1989), Lecture Notes in Comput. Sci. **372**, Springer, Berlin, 1989, 451–460.
17. F. L. Hitchcock, Multiple invariants and generalized rank of a p-way matrix or tensor, J. Math. and Phys. 7 (1927), no. 1, 40–79.
18. C. J. Hillar and L.-H. Lim, Most tensor problems are NP-hard, Journal of the ACM 60 (2013), no. 6, Art. 45.
19. R. H. Keshavan, A. Montanari and S. Oh, Matrix completion from a few entries, IEEE Trans. on Information Theory 56 (2010), 2980–2998.
20. J. B. Kruskal, Three-way arrays: rank and uniqueness for trilinear decompositions, with application to arithmetic complexity and statistics, Linear Algebra Appl. 18 (2) (1977), 95–138.
21. J. M. Landsberg, *Tensors: Geometry and Applications*, Graduate Studies in mathematics **128**, American Mathematical Society, RI, 2012.

22. J. M. Landsberg, New lower bounds for the rank of matrix multiplication, SIAM J. Comput. 53 (2014), no. 1, 144–149.
23. J. M. Landsberg and Z. Teitler, On the ranks and border ranks of symmetric tensors, Foundations of Computational Mathematics 10.3 (2010), 339–366.
24. L. de Lathauwer, B. de Moor and J. Vandewalle, A multilinear singular value decomposition, Siam J. Matrix Anal. Appl. 21, no. 4, 1253–1278.
25. F. Le Gall, Powers of tensors and fast matrix multiplication, Proceedings of the 39th International Symposium on Symbolic and Algebraic Computation (ISSAC 2014), 2014, 296–303.
26. L.-H. Lim and P. Comon, Multiarray signal processing: tensor decomposition meets compressed sensing, Comptes Rendus Mecanique 338 (2010), 311–320.
27. L. H. Lim and P. Comon, Blind multilinear identification, IEEE Trans. Inf. Theory 60 (2014), 1260–1280.
28. A. Massarenti and E. Raviolo, The rank of $n \times n$ matrix multiplication is at least $3n^2 - 2\sqrt{2}n^{3/2} - 3n$, Linear Algebra and its Applications 438 (2013), 4500–4509.
29. B. Recht, M. Fazel and P. A. Parrilo, Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimzation, SIAM review 52 (2010), no. 3, 471–501.
30. R. Schatten, A Theory of Cross-Spaces, Princeton University Press, Princeton, NJ, 1950.
31. M. Shafiei, *Apolarity for determinants and permanents of generic matrices*, preprint, arXiv:1212.0515.
32. V. Strassen, Gaussian elimination is not optimal, Numer. Math. 13 (1969), 354–356.
33. N. D. Sidiropoulos and R. Bro, On the uniqueness of multilinear decomposition of N-way arrays, J. Chemometrics 14 (2000), no. 3, 229–239.
34. B. Recht, M. Fazel and P. Parillo, Guaranteed minimum rank solutions of matrix equations via nuclear norm minimization, SIAM Review 52 (2010), no. 3, 471–501.
35. A. J. Stothers, On the Complexity of Matrix Multiplication, Ph.D. Thesis, Univ. Edinburgh, 2010.
36. V. V. Williams, *Multiplying matrices faster than Coppersmith-Winograd*, STOC 2012, Proceedings of the forty-fourth annual ACM symposium on Theory of computing, ACM, 2012, 887–898.