NUCLEAR NORM OF HIGHER ORDER TENSORS

SHMUEL FRIEDLAND AND LEK-HENG LIM

1. Introduction

The nuclear norm of a 2-tensor (or, in coordinate form, a matrix) has recently found widespread use as a convex surrogate for relaxing various intractable non-convex problems into tractable convex problems. More generally, the nuclear $p$-norm of a $d$-tensor $A \in V_1 \otimes \cdots \otimes V_d$ is defined by

$$ \|A\|_{*,p} := \min \left\{ \sum_{k=1}^{r} \|x_{1,k}\|_p \cdots \|x_{d,k}\|_p : A = \sum_{k=1}^{r} x_{1,k} \otimes \cdots \otimes x_{d,k}, r \in \mathbb{N} \right\}. $$

where $\| \cdot \|_p$ is the $l^p$-norm and $x_{j,k} \in V_j$ for $j = 1, \ldots, d, k = 1, \ldots, r$. The usual nuclear norm of a matrix is then the case when $d = p = 2$. This is known to be polynomial-time computable to arbitrary accuracy. Obviously, the computational tractability of the matrix nuclear norm is critical to its recent widespread use.

An immediate application of our results are the answers to the computational complexity of the nuclear norm in cases when $p \neq 1, 2, \infty$. Henceforth we will follow standard convention and write ‘nuclear norm’ for ‘nuclear 2-norm’ (regardless of the order $d$). For matrices this is standard and is equivalent to the usual definition as a sum of singular values, often known as the Schatten 1-norm (c.f. [2, 18] for example). For higher-order tensors it was defined explicitly in [21, 22] (see also [8, 5]) although the original idea dates back to Grothendieck [11] and Schatten [27].

Using the aforementioned hardness results, we immediately have (i) and (ii): the NP-hardness of the nuclear $p$-norm of 2-tensors may be deduced from that of the matrix $p$-norm for $p \neq 1, 2, \infty$ [14]; the NP-hardness of the nuclear norm of real 3-tensor may be deduced from that of the spectral norm of real 3-tensors [15].

The NP-hardness of the nuclear norm of complex 4-tensors in (iii) however requires that we first establish the corresponding hardness of the spectral norm of a 4-tensor, which we achieve by characterizing the clique number of a graph (well-known to be NP-hard) as the spectral norm of a 4-tensor. In fact the 4-tensor that we construct will bihermitian, positive semidefinite, and nonnegative-valued, allowing us to deduce a side result: deciding weak membership in the set of bipartite separable quantum states is NP-hard.

Equivalently, these hardness results may also be stated in an alternative form: the nuclear $p$-norm of 2-tensors, the nuclear norm of 3-tensors over $\mathbb{R}$, and the nuclear norm of 4-tensors over $\mathbb{R}$ and $\mathbb{C}$, are all not polynomial-time approximable to arbitrary accuracy.

1.1. Outline. In Section ?? we show that weak membership in the unit ball of a given norm can be decided in polynomial time if and only if weak membership in the unit ball of its dual norm can be decided in polynomial time. In Section ?? we show that weak membership in the unit ball of a given norm can be decided in polynomial if and only if $\epsilon$-approximation of the given norm is polynomial time. In Section 3 we deduce that the nuclear $p$-norm for matrices for any $p \neq 1, 2, \infty$, is NP-hard. In Section 2 we define the nuclear and spectral norms for tensors of arbitrary orders over $\mathbb{C}$ and $\mathbb{R}$. In Section 5 we relate Motzkin–Strauss’s characterization of the clique number of a
graph as the spectral norm of a 4-tensor defined by the graph. It then follows that ϵ-approximation of the nuclear norm for 4-tensors over \( \mathbb{C} \) and \( \mathbb{R} \), or equivalently, of bipartite density matrices, is NP-hard. In Section 6 we apply our results to show that deciding membership in the tensor nuclear norm unit ball is NP-hard.

2. Tensor nuclear and spectral norms

We will identify \( \mathbb{F}^{n_1 \times \cdots \times n_d} \) with \( \mathbb{F}^n \), where \( n = \prod_{i=1}^d n_i \). For any \( A \in \mathbb{F}^{n_1 \times \cdots \times n_d} \), the Hilbert–Schmidt norm, which is the Euclidean norm on \( \mathbb{F}^n \), is denoted \( \|A\| \), consistent with the notation in the previous paragraph.

Lemma 2.1. Let \( \| \cdot \|_{\sigma, \mathbb{F}} \) be the spectral norm in \( \mathbb{F}^{n_1 \times \cdots \times n_d} \). Then

\[
\frac{1}{\sqrt{n_1 n_2 \cdots n_d}} \|A\| \leq \|A\|_{\sigma, \mathbb{F}} \leq \|A\| \quad \text{for all } A \in \mathbb{F}^{n_1 \times \cdots \times n_d}. \tag{1}
\]

Proof. Clearly, \( \|A\|_{\sigma, \mathbb{F}} \leq \|A\| \). Assume that \( A = (a_{i_1 \ldots i_d}) \). Let \( \|A\|_{\text{max}} = \max \{|a_{i_1 \ldots i_d}| : i_1, \ldots, i_d \} \). Clearly, \( \|A\| \leq \sqrt{n_1 n_2 \cdots n_d} \|A\|_{\text{max}} \). Choose in characterization (4) each \( x_j \in \mathbb{F}^{n_j} \) to be a standard unit vector. It then follows that \( \|A\|_{\text{max}} \leq \|A\|_{\sigma, \mathbb{F}} \). Hence the left-hand side of (1) holds. \( \square \)

Let \( \mathbb{F} = \mathbb{R} \) or \( \mathbb{C} \). Let \( \mathbb{F}^{n_1 \times \cdots \times n_d} := \mathbb{F}^{n_1} \otimes \cdots \otimes \mathbb{F}^{n_d} \) be the space of \( d \)-tensors of dimensions \( n_1, \ldots, n_d \in \mathbb{N} \). If desired, these may be viewed as \( d \)-dimensional hypermatrices \( A = (a_{i_1 \ldots i_d}) \) with entries \( a_{i_1 \ldots i_d} \in \mathbb{F} \).

The hermitian inner product of two \( d \)-tensors \( A, B \in \mathbb{C}^{n_1 \times \cdots \times n_d} \) is given by

\[
\langle A, B \rangle = \sum_{i_1, \ldots, i_d=1}^{n_1, \ldots, n_d} \bar{a}_{i_1 \ldots i_d} b_{i_1 \ldots i_d}.
\]

This induces the Hilbert–Schmidt norm, denoted by

\[
\|A\| = \sqrt{\langle A, A \rangle} = \left( \sum_{i_1, \ldots, i_d=1}^{n_1, \ldots, n_d} |a_{i_1 \ldots i_d}|^2 \right)^{\frac{1}{2}}.
\]

We adopt the convention that an unlabeled \( \| \cdot \| \) will always denote the Hilbert–Schmidt norm. Note that when \( d = 1 \), this is just the regular hermitian or \( l^2 \)-norm of a vector in \( \mathbb{C}^n \) and when \( d = 2 \), this is the Frobenius norm of a matrix in \( \mathbb{C}^{n \times n} \).

The norms of greatest interest to us in this article are the spectral norm and nuclear norm of a \( d \)-tensor \( A \in \mathbb{C}^{n_1 \times \cdots \times n_d} \). These are denoted and defined respectively by

\[
\|A\|_{\sigma} := \max_{x_1, \ldots, x_d \neq 0} \frac{\langle A, x_1 \otimes \cdots \otimes x_d \rangle}{\|x_1\| \cdots \|x_d\|} = \max_{x_1, \ldots, x_d \neq 0} \frac{\text{Re}(\langle A, x_1 \otimes \cdots \otimes x_d \rangle)}{\|x_1\| \cdots \|x_d\|}, \tag{2}
\]

\[
\|A\|_* := \min \left\{ \sum_{p=1}^r \|x_{1,p} \cdots x_{d,p}\| : A = \sum_{p=1}^r x_{1,p} \otimes \cdots \otimes x_{d,p}, \ r \in \mathbb{N} \right\}. \tag{3}
\]

It is easy to see that \( \| \cdot \|_{\sigma} \) and \( \| \cdot \|_* \) are dual norms [22, Lemma 21] and that

\[
\|x_1 \otimes \cdots \otimes x_d\|_{\sigma} = \|x_1 \otimes \cdots \otimes x_d\|_* = \|x_1\| \cdots \|x_d\|.
\]

We would like to point out that (3) is the definition of tensor nuclear norm as originally defined by Grothendieck [11] and Schatten [27]. An alternate definition of ‘tensor nuclear norm’ as the average of nuclear norms of matrices obtained from flattenings of a tensor has gained recent popularity. While this alternate definition may be useful for various purposes, it is nevertheless not the definition commonly accepted in mathematics [4, 26, 23, 29] (see also [8, 22]). The nuclear norm defined in (3) is precisely the dual norm of the spectral norm in (2) and is naturally related to the notion of tensor rank (cf. [20]). Moreover, nuclear norm as defined in (3) has physical meaning — equivalent to bipartite separability of quantum states in an appropriate sense [6]. As such, a tensor nuclear norm in this article will always be the one in (3).
Tensor rank is known to depend on the choice of base field [20]. We do not know if it might be the same for nuclear and spectral norms. As such, we need a more careful definition. Let

$$B(\mathbb{F}^n) := \{ x \in \mathbb{F}^n : \|x\| \leq 1 \}, \quad S(\mathbb{F}^n) := \{ x \in \mathbb{F}^n : \|x\| = 1 \},$$

be the Euclidean unit ball and sphere in $\mathbb{F}^n$ (recall that $\| \cdot \|$ denotes the Euclidean norm on $\mathbb{F}^n$). Then for $A \in \mathbb{F}^{n_1 \times \cdots \times n_d}$, we define

$$\|A\|_{\sigma,\mathbb{F}} := \max \left\{ \langle A, x_1 \otimes \cdots \otimes x_d \rangle : x_i \in B(\mathbb{F}^{n_i}) \right\},$$

$$\|A\|_{*,\mathbb{F}} := \min \left\{ \sum_{p=1}^r \prod_{i=1}^d \|x_{i,p}\| : A = \sum_{p=1}^r x_{1,p} \otimes \cdots \otimes x_{d,p}, \ x_{i,p} \in \mathbb{F}^{n_i}, \ r \in \mathbb{N} \right\}. \quad (4)$$

The indices $i$ and $p$ are always assumed to run over $i = 1, \ldots, d$ and $p = 1, \ldots, r$ respectively. Clearly, for any $A \in \mathbb{F}^{n_1 \times \cdots \times n_d}$,

$$\|A\|_{\sigma,\mathbb{R}} \leq \|A\|_{\sigma,\mathbb{C}}, \quad \|A\|_{*,\mathbb{R}} \geq \|A\|_{*,\mathbb{C}}. \quad (5)$$

It is well known that for $d = 2$ we have equality signs in the above inequalities. We suspect that this is not the case for $d \geq 3$.

Let $x = (x_1, \ldots, x_n)^T \in \mathbb{C}^n$. Denote by $|x| := (|x_1|, \ldots, |x_n|)^T$. Then $x$ is called a nonnegative vector, denoted as $x \geq 0$, if $x = |x|$. We will also use this notation for tensors in $\mathbb{C}^{n_1 \times \cdots \times n_d}$.

**Lemma 2.2.** Let $A \in \mathbb{C}^{n_1 \times \cdots \times n_d}$. Then

$$\|A\|_{\sigma,\mathbb{C}} \leq \|A\|_{\sigma,\mathbb{C}}, \quad \|A\|_{\sigma,\mathbb{C}} = \|A\|_{\sigma,\mathbb{R}}. \quad (6)$$

**Proof.** The triangle inequality yields

$$\langle A, x_1 \otimes \cdots \otimes x_d \rangle \leq \langle |A|, x_1 \otimes \cdots \otimes x_d \rangle.$$ 

Recall that the Euclidean norm on $\mathbb{C}^n$ is an absolute norm, i.e., $\|x\| = ||x||$. The definitions of $\| \cdot \|_{\sigma,\mathbb{C}}$ and $\| \cdot \|_{\sigma,\mathbb{R}}$ and the above inequality yields the result. \qed

The equality in the above lemma fails for nuclear norm even for matrices. Indeed, let

$$Q = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}. $$

Then $\|Q\|_* = 2 > \|Q\|_{\sigma,\mathbb{R}} = \sqrt{2}$. A recent example of Ryota Tomioka shows that the equality in Lemma 2.2 fails for $4 \times 4$ hermitian matrices. Henceforth, by spectral and nuclear norm we will always mean over $\mathbb{C}$, unless stated otherwise. We will denote

$$\| \cdot \|_{\sigma} := \| \cdot \|_{\sigma,\mathbb{C}}, \quad \| \cdot \|_* := \| \cdot \|_{*,\mathbb{C}}.$$ 

In what follows we show that $\|A\|_{\sigma,\mathbb{F}}$ and $\|A\|_{*,\mathbb{F}}$ are NP-hard to compute for $d \geq 3$ by using appropriate results. If we show that $\|A\|_{\sigma,\mathbb{R}}$ is NP-hard to compute for $A \geq 0$, then Lemma 2.2 yields that $\|A\|_{\sigma,\mathbb{C}}$ is also NP-hard to compute.

### 3. Nuclear $(p,q)$-norm of a matrix

We show that a matrix norm induced by two norm spaces is the dual to a corresponding nuclear norm on matrices. Using the known results on the complexity of matrix norms we analyze exactly the complexity of nuclear norms of a matrix for two Hölder norms $p,q$.

For simplicity of the exposition we assume here that all normed space are over $\mathbb{R}$. Let $V$ be a finite dimensional vector space of dimension $\dim V = n$. Denote by $V^*$ the dual space of linear functionals $f : V \to \mathbb{R}$. Let $\| \cdot \|$ be a norm on $V$. Then the dual norm $\| \cdot \|^{\ast}$ is given by $\|f\|^{\ast} = \max_{\|x\|=1} f(x)$. Since $V$ is finite dimensional then $(V^*)^* = V$ and $(\| \cdot \|^{\ast})^* = \| \cdot \|$. 


We will identify $V$ and $V^*$ with $\mathbb{R}^n$, when necessary, with $f(x) = f^T x$, and no ambiguity will arise. Recall that for $p \in [1, \infty]$, the $\ell_p$ norm on $\mathbb{R}^n$ is given by

$$\| (x_1, \ldots, x_n)^T \|_p = \left( \sum_{i=1}^n |x_i|^p \right)^{1/p}.$$  

Moreover, $\| \cdot \|_p^* = \| \cdot \|_{p^*}$, where $p^* = \frac{p}{p-1}$, i.e. $\frac{1}{p} + \frac{1}{p^*} = 1$.

Let $V_1, V_2$ be two finite dimensional vector spaces. Denote by $L(V_1, V_2)$ the linear space of all linear transformations from $V_1$ to $V_2$. Assume now that $V_1$ is a normed space with the norms $\| \cdot \|_i$ for $i \in [2]$. The these norms induce the following operator norm on $L(V_1, V_2)$:

$$\|T\|_{1,2} = \max_{\|x\|_1 = 1} \| Tx \|_2, \quad T \in L(V_1, V_2).$$

Let $T \in L(V_1, V_2)$ and denote by $T^*$ the induced linear operator in $L(V_2^*, V_1^*)$ by the equality: $(T^* g)(x) = g(Tx)$ for all $x \in V_1$. The following equality is well known:

**Lemma 3.1.** $\|T^*\|_{2^*,1^*} = \|T\|_{1,2}$.  

**Proof.** Recall that $\| Tx \|_2 = \max_{\|g\|_{2^*} = 1} g(Tx)$. As $g(Tx) = (T^*g)(x) = x^*(T^*g)$ we deduce the lemma. \hfill \Box

We now recall the nuclear norm on $L(V_1, V_2)$. We first identify $L(V_1, V_2)$ with $V_2 \otimes V_1^*$. So $y \otimes f$ is the linear transformation $x \to f(x)y$ for each $x \in V_1$. We now define $\| y \otimes f \|_{\nu_2} := \| y \|_2 \| f \|_1$. Then the convex hull spanned by all ranks one matrices $y \otimes f$, where $\| y \|_2 \| f \|_1 \leq 1$ is the unit ball of the nuclear norm $\| \cdot \|_{\nu_2}$ in $L(V_1, V_2)$. Equivalently,

$$\| T \|_{\nu_2} := \min \left\{ \sum_{i=1}^r \| y_i \|_2 \| f_i \|_1^* : T = \sum_{i=1}^r y_i \otimes f_i \right\}.$$  

By interchanging the factors $y$ and $f$ we deduce the analog of Lemma 3.1:

$$\| T^* \|_{\nu_{1^*,2^*}} = \| T \|_{\nu_{2,1}}. \quad (7)$$

The main result of this note is:

**Theorem 3.2.** Let $V_i$ be a finite dimensional norm space with norm $\| \cdot \|_i$ over $\mathbb{R}$ for $i \in [2]$. Then the operator norm on $L(V_1, V_2)$ is the dual norm to the nuclear norm on $L(V_2, V_1)$.  

**Proof.** Let $T \in L(V_1, V_2)$ and $x \otimes g \in V_1 \otimes V_2^* \sim L(V_2, V_1)$. Then $T(x \otimes g) := g(Tx)$. So $T : L(V_2, V_1) \to \mathbb{R}$. As the unit ball of the nuclear norm $\nu_{1,2}$ is a convex combination of the matrices of rank one it follows

$$\| T \|_{1,2}^* = \max_{\|x\|_1 = 1, \|g\|_2 = 1} T(x \otimes g) = \max_{\|x\|_1 = 1, \|g\|_2 = 1} g(Tx) = \max_{\|x\|_1 = 1} \| Tx \|_2 = \| T \|_{1,2}. \quad \Box$$

We now specify ourselves to $\ell_p, \ell_q$ norms. So $V_1 = \mathbb{R}^n, V_2 = \mathbb{R}^m$ and $L(V_1, V_2)$ is represented by a matrix $A \in \mathbb{R}^{m \times n}$. Then $\| A \|_{p,q}$ is the above operator norm. Then nuclear norm is $\| A \|_{q,p}^*$. So $\| \cdot \|_{p,q}$ is dual to $\| \cdot \|_{q,p}^*$. The results of [28] yields that for $1 \leq q < p \leq \infty$ the computation of $\| A \|_{p,q}$ in NP-hard. For $p = q$ it is also NP-hard if $p \neq 1, 2, \infty$ [14]. It is well known that $\| A \|_{p,q}$ is polynomial if $T = q \in \{ 1, 2, \infty \}$. It is shown that in [28] that that $\| A \|_{p,q}$ is polynomial if $p = 1$ and $q > 1$ is rational, and the conjugate case given by Lemma 3.1.

So now we can use this results for the nuclear norms of matrices $\| A \|_{p,q}$ and our results n the complexity of the dual norms.

**Proposition 3.3.** Let $e_{j,n} := (\delta_{j,1}, \ldots, \delta_{jn})^T, i = 1, \ldots, n$ be the standard basis in $\mathbb{R}^n$. For a positive integer $m$ denote $[m] := \{1, \ldots, m\}$. Then for $A \in \mathbb{R}^{m \times n}$ the following equalities hold:

$$\| A \|_{1,p} = \max_{j \in [m]} \| Ae_{j,n} \|_p, \quad \| A \|_{p,\infty} = \max_{i \in [m]} \| A^T e_{i,m} \|_{p^*}, \quad p \in [1, \infty], \quad (8)$$

$$\| A \|_{1,p}^* = \sum_{j=1}^n \| Ae_{j,n} \|_{p^*}, \quad \| A \|_{p,\infty}^* = \sum_{i=1}^m \| A^T e_{i,m} \|_p, \quad p \in [1, \infty]. \quad (9)$$
Proof. Recall that unit ball of the $\ell_1$ norm, denoted as $B_{1,n}$ is the convex hull spanned by $\pm e_{j,n}, j \in [n]$. As $\|Ax\|_p$ is a convex function on $B_{1,n}$, we deduce that $\max_{x \in B_{1,n}} \|Ax\|_p = \max_{j \in [n]} \| \pm Ae_{j,n} \|$. Hence the first equality of (8) holds. Since $\|A\|_{p,\infty} = \|A^T\|_{1,p^*}$ we deduce the second equality of (8) from the first one.

Recall that $f \in (\mathbb{R}^{m \times n})^*$ is of the form $f_B$ for some $B \in \mathbb{R}^{m \times n}$, where $f_B(A) = \text{tr}(B^T A)$. Hence

$$\|A\|_{1,p}^* = \max_{\|B\|_{1,p} \leq 1} \text{trace } B^T A = \max_{\|Be_{k,n}\|_{1,p} \leq 1, k \in [n]} \left| \sum_{j=1}^n (Be_{j,n})^T (Ae_{j,n}) \right| = \sum_{j=1}^n \|Ae_{j,n}\|_{p^*}.$$ 

This proves the first equality of (9). The second equality of (9) follows similarly from the second equality of (8).

4. Nuclear norm for higher order tensors is special

One might think that it is also possible to define Schatten $p$-norm for any $p \in (0, \infty]$ as follows

$$\|A\|_{*,p} := \inf \left\{ \left[ \sum_{i=1}^r \lambda_i \right]^{1/p} : A = \sum_{i=1}^r \lambda_i u_i \otimes v_i \otimes w_i, \|u_i\| = \|v_i\| = \|w_i\| = 1, \, r \in \mathbb{N} \right\}.$$ 

Proposition 4.1. $\| \cdot \|_{*,p}$ is identically zero: $\|A\|_{*,p} = 0$ for all $p \neq 1$ and all $A \in \mathbb{C}^{m \times n \times p}$.

Proof. Write

$$u \otimes v \otimes w = \frac{1}{2^p} u \otimes v \otimes w + \cdots + \frac{1}{2^p} u \otimes v \otimes w$$

and observe that

$$\inf_{n \in \mathbb{N}} \left[ \sum_{i=1}^r 2^{-np} \right]^{1/p} = 0.$$ 

Why does this work for matrices then? The reasons is that we may impose orthogonality in the factors for the case $d = 2$ but this is not possible when $d > 2$.

5. Clique number and spectral norm

Let $G = (V, E)$ be a graph with $V := \{1, \ldots, n\}$ and the set of undirected edges $E = \{(i_k, j_k) : i_k = 1, \ldots, j_k - 1, k = 1, \ldots, m\}$. Let $\kappa(G)$ be the clique number of $G$, i.e., the size of the maximal clique in $G$. Denote by $A_G$ the adjacency matrix of $G$. Let $\Delta^n$ be the simplex of probability vectors on $\mathbb{R}^n$. Motzkin and Straus [24] showed that

$$\frac{\kappa(G) - 1}{\kappa(G)} = \max_{x \in \Delta^n} x^T A_G^* x.$$ 

Equality is attained when $x$ is uniformly distributed on the largest clique.

We now transform (10) to one involving $4$-tensors. Let $x = (y^2) \in \mathbb{C}^2$, i.e., $x = (y_1^2, \ldots, y_n^2)^T$. Then

$$(y^2)^T A_G y^2 = \sum_{(i,j) \in E(G)} y_i^2 y_j^2, \quad y^T y = 1.$$ 

We will see how a term of the form $y_i^2 y_j^2$ may be regarded as a $4$-tensor.

Definition 5.1. Let $A = (a_{ijpq})_{i,j,p,q=1}^{m,n,m,n} \in \mathbb{C}^{m \times n \times m \times n}$ be a $4$-tensor. We call it bisymmetric if

$$a_{ijpq} = a_{pqij} \quad \text{for all } i, p = 1, \ldots, m, j, q = 1, \ldots, n,$$

and bihermitian if

$$a_{ijpq} = \bar{a}_{pqij} \quad \text{for all } i, p = 1, \ldots, m, j, q = 1, \ldots, n.$$ 

A bihermitian tensor $A = (a_{ijpq})$ is said to be positive semidefinite if

$$\sum_{i,j,p,q=1}^{m,n,m,n} a_{ijpq} x_{ij} \bar{x}_{pq} \geq 0$$

for all $X = [x_{ij}] \in \mathbb{C}^{m \times n}$. 

Proof. It is easy to see that the tensor $A = (a_{i,j,p,q})^{m,n,m,n} \in \mathbb{C}^{m \times n \times m \times n}$ as a matrix $C(A) = [c_{(i,j),(p,q)}] \in \mathbb{C}^{m \times m \times n \times n}$, where $c_{(i,j),(p,q)} = a_{i,j,p,q}$. Evidently, $A$ is bisymmetric (resp. bihermitian) if and only if $C(A)$ is symmetric (resp. hermitian).

For integers $1 \leq s < t \leq n$, we consider $A_{st} = (a_{i,j,p,q}^{(s,t)})^{m,n,m,n}_{i,j,p,q=1} \in \mathbb{C}^{n \times n \times n \times n}$ defined by

$$
a_{i,j,p,q}^{(s,t)} = \begin{cases} 
1/2 & i = s, j = t, p = s, q = t, \\
1/2 & i = t, j = s, p = t, q = s, \\
1/2 & i = s, j = t, p = t, q = s, \\
1/2 & i = t, j = s, p = s, q = t, \\
0 & \text{otherwise.}
\end{cases}
$$

(14)

Note that $m = n$ for us and the quartic form $\langle A_{st}, y \otimes y \otimes y \otimes y \rangle = 2y_x y_t^2$.

**Lemma 5.2.** The tensor $A_{st}$ is bihermitian, positive semidefinite, and has all entries nonnegative.

**Proof.** It is easy to see that $A_{st}$ is bihermitian and nonnegative. It is positive semidefinite because

$$
\sum_{i,j,s,t=1}^{n,n,m,n} a_{i,j,p,q}^{(s,t)} x_i x_j x_p x_q = \frac{1}{2}(x_{st} + x_{ts})(\bar{x}_{st} + \bar{x}_{ts}) \geq 0
$$

for all $X = [x_{ij}] \in \mathbb{C}^{n \times n}$. \qed

$C(A_{st})$ is evidently a nonnegative definite, rank-one matrix with trace one. For those familiar with quantum information theory, this means that $C(A_{st})$ represents a bipartite density matrix [6]. For any graph $G$, we will let

$$
A_G := \frac{1}{m} \sum_{(s,t) \in E(G)} A_{st} \in \mathbb{C}^{n \times n \times n \times n},
$$

(15)

and $C(A_G) \in \mathbb{C}^{n^2 \times n^2}$ be its corresponding matrix. Note that $A_G$ and $C(A_G)$ have real-valued entries.

**Theorem 5.3.** Let $G$ be a simple undirected graph on $n$ vertices with $m$ edges. Let $A_G$ be as above. Then

$$
\|A_G\|_\sigma := \max_{0 \neq x,y,u,v \in \mathbb{C}^n} \frac{|\langle A_G, x \otimes y \otimes u \otimes v \rangle|}{\|x\| \|y\| \|u\| \|v\|} = \max_{0 \neq x,y,u,v \in \mathbb{R}^n_+} \frac{\langle A_G, x \otimes y \otimes u \otimes v \rangle}{\|x\| \|y\| \|u\| \|v\|} \geq 0
$$

(16)

$$
= \max_{0 \neq y \in \mathbb{R}^n_+} \frac{\langle A_G, y \otimes y \otimes y \otimes y \rangle}{\|y\|^4}.
$$

(17)

Since all entries of $A_G$ are nonnegative we immediately deduce the equality (16). If $A_G$ is a symmetric 4-tensor as opposed to merely bisymmetric, then we may apply a classical result of Banach [1, 7] to deduce that the maximum is attained at $x = y = u = v$ and thus (17). Unfortunately for us, $A_G$ is not symmetric and we need to prove Theorem 5.3 from scratch. We start with the following lemma which may be of independent interest.

**Lemma 5.4.** Let $H = [h_{(i,j),(p,q)}] \in \mathbb{C}^{m \times m \times m \times m}$ be a hermitian nonnegative definite matrix. Define $A = (a_{i,j,p,q}) \in \mathbb{C}^{m \times n \times n \times n}$ by $a_{i,j,p,q} = h_{(i,j),(p,q)}$. Then

$$
\|A\|_\sigma = \max_{0 \neq x \in \mathbb{C}^n, 0 \neq y \in \mathbb{C}^n} \frac{\langle A, x \otimes y \otimes \bar{x} \otimes \bar{y} \rangle}{\|x\|^2 \|y\|^2}
$$

(18)
Proof. Recall that we may always write $H = R^2$ for some hermitian $R \in \mathbb{C}^{mn \times mn}$. Consider the sesquilinear form $\bar{w}^T H z = (\bar{w} w)^T (R z)$. Use Cauchy–Schwarz inequality to see that

$$|\bar{z}^T H w| \leq \sqrt{\bar{z}^T H \bar{z}} \sqrt{\bar{w}^T H \bar{w}} \leq \max(\bar{z}^T H z, \bar{w}^T H w).$$

Now let $z := x \otimes y$ and $w := \bar{u} \otimes \bar{v}$ to deduce the lemma. \qed

Proof of Theorem 5.3. We first apply Lemma 5.4 to (16). We can assume that $u, v \geq 0$ and $y = v \geq 0$. Hence $2(A_{st}, x \otimes y \otimes x \otimes y) = (x_s y_t + x_t y_s)^2$. Use Cauchy–Schwarz inequality to see that

$$(x_s y_t + x_t y_s)^2 \leq 4 \frac{(x_s^2 + y_s^2)}{2} \times \frac{(x_t^2 + y_t^2)}{2}.$$ 

So by introducing two new variables $a_s = \sqrt{(x_s^2 + y_s^2)/2}$ and $a_t = \sqrt{(x_t^2 + y_t^2)/2}$, we reduced our problem to the problem of degree 4 in one vector variable $a = (a_1, \ldots, a_n)$, where $\|a\| = 1$. This is exactly the characterization (17) that we require. \qed

It was shown in [15] that a computation and approximation of spectral norm of $d$-tensor, for $d \geq 3$, is NP-hard over $\mathbb{R}$. We will now prove the NP-hardness of tensor spectral norm for bisymmetric 4-tensors over $\mathbb{R}$ and bihermitian 4-tensor over $\mathbb{C}$.

**Theorem 5.5.** It is NP-hard to approximate the spectral norm of bihermitian 4-tensors over $\mathbb{C}$ and of bisymmetric real 4-tensor over $\mathbb{R}$.

**Proof.** Let $A_G$ be defined by (15). Then $A_G$ is a nonnegative bisymmetric tensor, as we have pointed out. Motzkin–Strauss theorem yields that

$$\|A_G\|_\sigma = \|A_G\|_{\sigma,\mathbb{R}} = \frac{\kappa(G) - 1}{m \kappa(G)}. \quad (19)$$

Since the computation of clique number of a graph is NP-hard, the above identity implies that the computation of the spectral norm corresponding to bipartite density matrices is NP-hard over $\mathbb{C}$ and $\mathbb{R}$. Since the clique number of a graph is an integer, it follows that it is NP-hard to approximate the spectral norms of the corresponding 4-tensors over complex or real numbers. \qed

Observe that the matrix $C(A_G)$ is positive semidefinite and since $m$ is the number of edges in $E(G)$ it follows that the trace of $C(A_G)$ is 1. Hence $C(A_G)$ represents a real bipartite density matrix. It follows that the problem remains NP-hard even when we restrict ourselves to the set of 4-tensors corresponding to bipartite density matrices. See [6] for further discussions.

6. Weak membership in tensor nuclear norm unit ball is NP-hard

We show that the NP-hardness of the weak membership problem for the nuclear norm unit ball of 4-tensor over $\mathbb{C}$ and $\mathbb{R}$. In the following, we write $\mathbb{Q}[i] := \{a + bi : a, b \in \mathbb{Q}\}$ for the Gaussian rationals.

**Theorem 6.1.** Given $A \in \mathbb{Q}[i]^{n \times n \times n \times n}$ or $A \in \mathbb{Q}^{n \times n \times n \times n}$ and $0 < \delta \in \mathbb{Q}$, deciding whether $A \in S(B_{\| \cdot \|}, \delta)$ or $A \notin S(B_{\| \cdot \|}, -\delta)$ is an NP-hard problem.

**Proof.** By Theorems ?? and 5.5, we deduce that $\text{wmem}$ for the tensor spectral norm unit ball is NP-hard over $\mathbb{C}$ and $\mathbb{R}$. Since the tensor spectral and nuclear norms are dual to each other over $\mathbb{C}$ and $\mathbb{R}$, by Theorem ??, the $\text{wmem}$ for the tensor nuclear norm unit ball is also NP-hard over $\mathbb{C}$ and $\mathbb{R}$. \qed
7. Analogues of Comon’s conjecture and Banach’s theorem for nuclear norm

Recall that Comon’s conjecture asserts that rank = symmetric rank: For \( A \in S^d(\mathbb{F}^n) \),
\[
\min \left\{ r : A = \sum_{i=1}^r \lambda_i u_i \otimes v_i \otimes w_i \right\} \leq \min \left\{ r : A = \sum_{i=1}^r \lambda_i v_i \otimes u_i \otimes v_i \right\}
\]
Banach’s theorem on the other hand shows that spectral norm = symmetric spectral norm: For \( A \in S^d(\mathbb{F}^n) \),
\[
\sup_{x,y,z \neq 0} \frac{|A(x, y, z)|}{\|x\|\|y\|\|z\|} = \sup_{x \neq 0} \frac{|A(x, x, x)|}{\|x\|^3}
\]
In this section, we establish that nuclear norm = symmetric nuclear norm: For \( A \in S^d(\mathbb{F}^n) \),
\[
\inf \left\{ \sum_{i=1}^r |\lambda_i| : A = \sum_{i=1}^r \lambda_i u_i \otimes v_i \otimes w_i, \ r \in \mathbb{N} \right\} = \inf \left\{ \sum_{i=1}^r |\lambda_i| : A = \sum_{i=1}^r \lambda_i v_i \otimes u_i \otimes v_i, \ r \in \mathbb{N} \right\}
\]
Let \( \mathbb{F} = \mathbb{R}, \mathbb{C} \). Let \( d, n_1, \ldots, n_d \in \mathbb{N} \). Denote \( \mathbb{F}^{n_1 \times \cdots \times n_d} := \mathbb{F}^{n_1} \otimes \cdots \otimes \mathbb{F}^{n_d} \), the space of corresponding mode \( d \) tensors. Denote by \( \langle B, A \rangle \) the \( \mathbb{F} \)-inner product on \( \mathbb{F}^{n_1 \times \cdots \times n_d} \). Then \( \|A\| := \sqrt{\langle A, A \rangle} \) is the Hilbert–Schmidt norm on \( \mathbb{F}^{n_1 \times \cdots \times n_d} \). Recall the notion of the spectral and nuclear norm of \( A \in \mathbb{F}^{n_1 \times \cdots \times n_d} \):
\[
\|A\|_{\sigma, \mathbb{F}} = \max_{\|x_i\|=1} \text{Re} \langle A, x_1 \otimes \cdots \otimes x_d \rangle = \max_{\|x_i\|=1} |\langle A, x_1 \otimes \cdots \otimes x_d \rangle|,
\]
\[
\|A\|_{\ast, \mathbb{F}} = \min \left\{ \sum_{j=1}^r \prod_{i=1}^d \|x_{j,i}\| : A = \sum_{j=1}^r x_{j,1} \otimes \cdots \otimes x_{j,d}, \ r \in \mathbb{N} \right\}
\]
It is well known that the spectral and the nuclear norms are dual norms.

Clearly,
\[
\|A\|_{\sigma, \mathbb{R}} \leq \|A\|_{\sigma, \mathbb{C}}, \quad \|A\|_{\ast, \mathbb{R}} \geq \|A\|_{\ast, \mathbb{C}} \quad \text{for } A \in \mathbb{F}^{n_1 \times \cdots \times n_d}.
\]
It is well known that for \( d = 2 \) equalities hold in the above inequalities. It is known that for \( d \geq 3 \) one can have the following strict inequalities in the above inequalities.

Let \( \mathbb{F}^n \otimes \mathbb{R} \supset S^d(\mathbb{F}^n) \) be space of order \( d \) tensors of dimensions \( n \) and the subspace of symmetric tensors. Banach’s theorem for symmetric tensors claims [1]: For every \( B \in S^d(\mathbb{F}^n) \),
\[
\|B\|_{\sigma, \mathbb{F}} = \max_{\|x_i\|=1} |\langle B, x_1 \otimes \cdots \otimes x_d \rangle|.
\]
The aim of this note is to show that Banach’s theorem extends to the nuclear norm of symmetric tensors: For every \( B \in S^d(\mathbb{F}^n) \),
\[
\|B\|_{\ast, \mathbb{F}} = \min \left\{ \sum_{j=1}^r \|x_j\|^d : B = \sum_{j=1}^r \varepsilon_j x_j \otimes \cdots \otimes x_j, \ \varepsilon_j \in \{-1, 1\}, \ j = 1, \ldots, r \right\}
\]
Note that we can choose \( \varepsilon_j = 1, \ j = 1, \ldots, r \) unless \( \mathbb{F} = \mathbb{R} \) and \( d \) is an even integer.

As in [7] we can extend Banach’s theorem to partially symmetric tensors. Assume that \( d = \sum_{k=1}^l d_k \) where we assume that \( (d_1, \ldots, d_l) \in \mathbb{N}^l \) is a partition of \( d \), \( d_1 \geq \cdots \geq d_l \geq 1 \) and \( d_1 \geq 2 \). We now assume that
\[
n_{d_k+1} \cdots = n_{d_{k+1}}, \quad k = 0, \ldots, l, \quad d_0 = 0, \quad d_k = \sum_{i=1}^k d_i, \quad k = 1, \ldots, l.
\]
A tensor \( B = [b_{i_1 \cdots i_d}] \in \mathbb{F}^{n_1 \times \cdots \times n_d} \) is called \( (d_1, \ldots, d_l) \)-symmetric if the entries of \( B \) are not changed by a permutation of indices \( i_{d_k+1}, \ldots, i_{d_{k+1}} \) for \( k \in \{0, \ldots, l-1\} \). Denote by \( S^{(d_1, \ldots, d_l)}(\mathbb{F}^{n_1 \times \cdots \times n_d}) \) the subspace of symmetric tensors in \( \mathbb{F}^{n_1 \times \cdots \times n_d} \). It is straightforward to show as in [7] that: For all \( B \in S^{(d_1, \ldots, d_l)}(\mathbb{F}^{n_1 \times \cdots \times n_d}) \),
\[
\|B\|_{\sigma, \mathbb{F}} = \max_{\|x_i\|=1} |\langle B, x_1 \otimes \cdots \otimes x_i^d \rangle|.
\]
The analog of (24) is: For all \( B \in S^{d_1 \cdots d_l}(\mathbb{R}^{n_1 \times \cdots \times n_d}) \),
\[
\|B\|*F = \min \left\{ \sum_{j=1}^{r} \prod_{i=1}^{l} \|x_{j,i}\|^{d_i} : B = \sum_{j=1}^{r} \varepsilon_j x_{j,1}^{\otimes d_1} \otimes \cdots \otimes x_{j,l}^{\otimes d_l}, \varepsilon_j \in \{-1,1\}, j = 1, \ldots, r \right\}. \tag{27}
\]

Note that we can choose \( \varepsilon_j = 1, j = 1, \ldots, r \) unless \( F = \mathbb{R} \) and \( d_1, \ldots, d_l \) are even integers.

7.1. **Real symmetric tensors.** We first prove (27) for \( F = \mathbb{R} \). Let \( C := \operatorname{conv}(E) \subset \mathbb{R}^{n_1 \times \cdots \times n_d} \) be the convex hull of all vectors of the form
\[
E := \{ \pm x^{\otimes d} : x \in \mathbb{R}^n, \|x\| = 1 \}.
\]
As \( x^{\otimes d} + (-1)x^{\otimes d} = 0 \) it follows that \( C \) is a symmetric set in \( S^d(\mathbb{R}^n) \). Since any symmetric tensor is a linear combinations of vectors of the form \( x^{\otimes d} \) it follows that \( C \subset S^d(\mathbb{R}^n) \) with an interior. Hence \( C \) defines a unit ball with respect to the norm \( \nu : S^d(\mathbb{R}^n) \rightarrow [0, \infty) \). Note that \( \nu(x^{\otimes d}) \leq 1 \) for \( \|x\| = 1 \). We claim that each \( \pm x^{\otimes d} \), \( \|x\| = 1 \) is an extreme point of \( C \). Indeed, consider the unit ball of the Hilbert–Schmidt norm \( \|\cdot\|_{B} \). We deduce that \( \nabla \) is given by the right-hand of (24) we deduce that \( \nabla \) does not hold. Let \( C \) be the unit ball of the Hilbert–Schmidt norm \( \|\cdot\|_{B} \). Hence \( \nabla \) defines a unit ball with respect to the norm \( \nu(x^{\otimes d}) = \max \|x\|_{B} \) yields \( 1 = \max \|x\|_{B} \). Assume to the contrary that this equality holds for \( x^{\otimes d} \). Assume that \( \nu(B) \leq \sum_{i=1}^{r} \|t_i\|\|x_i\| \). (Without loss of generality we may assume that \( t_i \in \{-1,1\} \) for \( i = 1, \ldots, r \) Hence
\[
\nu(B) \leq \inf \left\{ \sum_{j=1}^{r} \prod_{i=1}^{l} \|x_{j,i}\|^{d_i} : B = \sum_{j=1}^{r} \varepsilon_j x_{j,1}^{\otimes d_1} \otimes \cdots \otimes x_{j,l}^{\otimes d_l}, \varepsilon_j \in \{-1,1\}, j = 1, \ldots, r \right\}.
\]
We claim that the infimum is achieved. It is enough to consider the case \( \nu(B) = 1 \). Hence \( B \in C \). Therefore \( B \) is a convex combination of the extreme points of \( C \). Hence
\[
B = \sum_{i=1}^{r} t_i \varepsilon_i x_i^{\otimes d}, \quad t_i > 0, \|x_i\| = 1, \varepsilon_i = \pm 1, i = 1, \ldots, r, \sum_{i=1}^{r} t_i = 1. \tag{28}
\]
Caratheodory’s theorem yields that \( r \leq 1 + \prod_{i=1}^{d} n_j \). The triangle inequality yields \( 1 = \nu(S) \leq \sum_{i=1}^{r} t_i \nu(x_i^{\otimes d}) = \sum_{i=1}^{r} t_i = 1 \). Combine that with (28) to deduce that \( \nu(B) \) is given by the right-hand side of (24).

Let \( \nu^* \) be the dual norm of \( \nu \) on \( S^d(\mathbb{R}^n) \). By the definition of the dual norm it follows that
\[
\nu^*(B) = \max_{A \in S^d(\mathbb{R}^n), \|A\| \leq 1} \langle B, A \rangle = \max_{A \in E} \langle B, A \rangle = \max_{\|x\| = 1} \langle B, x^{\otimes d} \rangle.
\]
In view of Banach’s theorem \( \nu^*(B) = \|B\|_{\sigma, \mathbb{R}} \). From the definition \( \|B\|_{\sigma, \mathbb{R}} \) and the fact that \( \nu(B) \) is given by the right-hand of (24) we deduce that \( \|B\|_{\sigma, \mathbb{R}} \leq \nu(B) \) for each \( B \in S^d(\mathbb{R}^n) \). Let \( \nu_1(\cdot) : S^d(\mathbb{R}^n) \rightarrow [0, \infty) \) be the induced norm on \( S^d(\mathbb{R}^n) \) by the nuclear norm on \( (\mathbb{R}^n)^{\otimes d} \). So \( \nu_1(B) = \|B\|_{\sigma, \mathbb{R}} \) for \( B \in S^d(\mathbb{R}^n) \). We claim that \( \nu = \nu_1 \). Assume to the contrary that this equality does not hold. Let \( C := \{ B : \nu_1(B) \leq 1 \} \) be the unit ball of \( \nu_1 \). Our assumption yields that \( C \subset C_1 \). Let \( \nu_1^* \) be the dual norm of \( \nu_1 \). Let \( C^* \) and \( C_1^* \) be the unit balls of the norms \( \nu^* \) and \( \nu_1^* \) respectively. It is well known that the assumption \( C \subset C_1 \) yields that \( C_1^* \subset C^* \). Hence there exists \( B \) such that \( \nu_1^*(B) > \nu(B) \) for some \( B \in S^d(\mathbb{R}^n) \). Now
\[
\nu_1^*(B) = \max_{A \in S^d(\mathbb{R}^n), \|A\| \leq 1} \langle B, A \rangle \leq \max_{A \in (\mathbb{R}^n)^{\otimes d}, \|A\| \leq 1} \langle B, A \rangle = \|B\|_{\sigma, \mathbb{R}}.
\]
As \( \nu^*(B) = \|B\|_{\sigma, \mathbb{R}} \) we obtain a contradiction.
7.2. **Complex symmetric tensor.** To deduce similar results for complex valued tensors we have to identify \((\mathbb{C}^n)^{\otimes d}\) with \((\mathbb{R}^n)^{\otimes d} \times (\mathbb{R}^n)^{\otimes d} = \mathbb{R}^{2n^d}\). That is let \(A \in (\mathbb{C}^n)^{\otimes d}\). View \(A = X + iY\), where \(X, Y \in (\mathbb{R}^n)^{\otimes d}\). Identify \(A\) with \((X, Y)\). On \((\mathbb{R}^n)^{\otimes d} \times (\mathbb{R}^n)^{\otimes d}\) define a real inner product
\[
\langle (X,Y),(W,Z) \rangle = \langle X,W \rangle + \langle Y,Z \rangle = \text{Re}(X+iY,W+iZ).
\]

Note that the Hilbert–Schmidt norm on \((\mathbb{C}^n)^{\otimes d}\) is the same as the Hilbert–Schmidt norm on \((\mathbb{R}^n)^{\otimes d} \times (\mathbb{R}^n)^{\otimes d}\). Hence the spectral norm on \((\mathbb{C}^n)^{\otimes d}\), given by the first equality of (20), is translated straightforward to the spectral norm on the real space \((\mathbb{R}^n)^{\otimes d} \times (\mathbb{R}^n)^{\otimes d}\). Hence its dual norm on \((\mathbb{R}^n)^{\otimes d} \times (\mathbb{R}^n)^{\otimes d}\) is given also by (21). (This follows from the observation that the extreme points of the unit ball in the nuclear norm is exactly \(E = \{x_1 \otimes \cdots \otimes x_d, x_j \in \mathbb{C}^n, \|x_j\| = 1, j \in [d]\}\). So \(S^d(\mathbb{C}^n)\) is a viewed as a real subspace \(S^d(\mathbb{R}^n) \times S^d(\mathbb{R}^n)\). Now we can repeat the arguments as in the real case using Banach’s theorem for complex valued symmetric tensors.

It can be deduced from the Banach theorem for the spectral norm of partial symmetric tensors the same way as for symmetric tensors.

8. **Nuclear decomposition of symmetric tensors for nuclear norm**

We say that
\[
A = \sum_{i=1}^{r} x_{j,i} \otimes \cdots \otimes x_{j,d}
\]
is a nuclear decomposition over \(\mathbb{F}\) if and only if
\[
\|A\|_{*,\mathbb{F}} = \sum_{i=1}^{r} \prod_{j=1}^{d} \|x_{j,i}\| \quad \text{and} \quad \prod_{j=1}^{d} \|x_{j,i}\| > 0 \text{ for each } i = 1, \ldots, r.
\]

**Lemma 8.1.** Let \(A \in F_{\mathbb{F}}^{n_1 \times \cdots \times n_d}\). Then (29) is a nuclear decomposition over \(\mathbb{F}\) if and only if there exists \(X \in F_{\mathbb{F}}^{n_1 \times \cdots \times n_d}\) satisfying
\[
\|X\|_{*,\mathbb{F}} = 1 \quad \text{and} \quad \langle x_{j,1} \otimes \cdots \otimes x_{j,d}, X \rangle = \prod_{j=1}^{d} \|x_{j,i}\| \text{ for all } i = 1, \ldots, r.
\]

**Proof.** Recall that \(\text{Re}\langle A, X \rangle \leq \|A\|_{*,\mathbb{F}}\|X\|_{*,\mathbb{F}}\). Suppose furthermore \(\|X\|_{*,\mathbb{F}} = 1\) and \(A \neq 0\). Then \(\text{Re}\langle A, X \rangle = \|A\|_{*,\mathbb{F}}\|X\|_{*,\mathbb{F}}\) if and only if the real functional \(\text{Re}(C, X)\) is a supporting hyperplane of the ball \(\{C \in F_{\mathbb{F}}^{n_1 \times \cdots \times n_d}, \|C\|_{*,\mathbb{F}} \leq \|A\|_{*,\mathbb{F}}\}\) at the point \(C = A\).

Assume that (30) holds. Without loss of generality we assume that \(\prod_{j=1}^{d} \|x_{j,i}\| > 0 \text{ for each } i = 1, \ldots, r\). Suppose first that (31) holds. Then
\[
\|A\|_{*,\mathbb{F}} = \|A\|_{*,\mathbb{F}}\|X\|_{*,\mathbb{F}} \geq \text{Re}\langle A, X \rangle = \sum_{i=1}^{r} \langle x_{j,1} \otimes \cdots \otimes x_{j,d}, B \rangle = \sum_{i=1}^{r} \prod_{j=1}^{d} \|x_{j,i}\|.
\]
The minimal characterization of \(\|A\|_{*,\mathbb{F}}\) yields that (30) is a nuclear decomposition of \(A\) over \(\mathbb{F}\).

Vice versa, suppose that (30) is a nuclear decomposition of \(A\) over \(\mathbb{F}\). Then there exists \(X \in F_{\mathbb{F}}^{n_1 \times \cdots \times n_d}, \|X\|_{*,\mathbb{F}} = 1\) such that \(\text{Re}\langle A, X \rangle = \|A\|_{*,\mathbb{F}}\). Hence
\[
\|A\|_{*,\mathbb{F}} = \text{Re}\langle A, X \rangle = \sum_{i=1}^{r} \text{Re}\langle x_{j,1} \otimes \cdots \otimes x_{j,d}, B \rangle \leq \sum_{i=1}^{r} \prod_{j=1}^{d} \|x_{j,i}\| = \|A\|_{*,\mathbb{F}}.
\]
Therefore (31) holds. \(\square\)

9. **Base field dependence**

Recall that rank is dependent on base fields: \(x, y \in \mathbb{R}^n\) linearly independent and \(z = x + iy \in \mathbb{C}^n\),
\[
x \otimes x \otimes x - x \otimes y \otimes y + y \otimes x \otimes y + y \otimes y \otimes x = \frac{1}{2}(z \otimes \bar{z} \otimes \bar{z} + \bar{z} \otimes z \otimes z),
\]
then \(\text{rank}_\mathbb{C}(A) = 2 < 3 = \text{rank}_\mathbb{R}(A)\).

Here we show that spectral norm is dependent on base field
\[
A = e_1 \otimes e_1 \otimes e_2 + e_1 \otimes e_2 \otimes e_1 + e_2 \otimes e_1 \otimes e_1 - e_2 \otimes e_2 \otimes e_2,
\]
then \( \|A\|_{\sigma,C} = 2 < 2\sqrt{2} = \|A\|_{\sigma,R} \). In fact, so is nuclear norm

\[
B = e_1 \otimes e_1 \otimes e_2 + e_1 \otimes e_2 \otimes e_1 + e_2 \otimes e_1 \otimes e_1,
\]

then \( \|B\|_{*,C} = \frac{3}{2} \sqrt{3} < 3 = \|B\|_{*,R} \).

We now provide the details.

**Lemma 9.1.** Let

\[
B_1 = \frac{1}{\sqrt{3}}(e_1 \otimes e_1 \otimes e_2 + e_1 \otimes e_2 \otimes e_1 + e_2 \otimes e_1 \otimes e_1),
\]

\[
B_2 = \frac{1}{2}(e_1 \otimes e_1 \otimes e_2 + e_1 \otimes e_2 \otimes e_1 + e_2 \otimes e_1 \otimes e_1 - e_2 \otimes e_2 \otimes e_2),
\]

where \( e_i = (\delta_{i1}, \delta_{i2})^T, i = 1, 2 \). Then

\[
\|B_1\|_{\sigma,R} = \|B_1\|_{\sigma,C} = \frac{2}{3}, \quad \|B_1\|_{*,R} = \sqrt{3}, \quad \|B_1\|_{*,C} = \frac{3}{2},
\]

\[
\|B_2\|_{\sigma,R} = \frac{1}{2}, \quad \|B_2\|_{\sigma,C} = \frac{1}{\sqrt{2}}, \quad \|B_2\|_{*,R} = 2.
\]

**Proof.** As \( B_1 \) and \( B_2 \) symmetric to compute the spectral norms of these tensors over reals and complex numbers we will use Banach theorem in most of the cases. For computations we will use \( X = \sqrt{3}B_1, Y = 2B_2 \). Let \( x = (x, y)^T \). Then \( 3f(x, y) := \langle X, x^{\otimes 3} \rangle = 3x^2y \). As \( X \) has nonnegative entries, clearly \( \|X\|_{\sigma,R} = \|X\|_{*,C} \). So it is enough to consider the case \( x, y > 0, x^2 + y^2 = 1 \). So \( f(x, y) = x^2\sqrt{1 - x^2}, x \in [0, 1] \). A straightforward calculation yields the maximum is achieved at \( x^2 = 2/3, y = 1/\sqrt{3} \). This establishes the first equality in (34).

To show that \( \|Y\|_{\sigma,R} = 1 \) we consider the maximum of the trilinear form \( \langle Y, x \otimes y \otimes z \rangle \), where \( \|x\| = \|y\| = \|z\| = 1 \). Note the the two slices of \( Y = [b_{i,j,k}] \) are

\[
B_1 = [b_{i,j,1}] = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad B_2 = [b_{i,j,2}] = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}
\]

Let \( z = (\cos \theta, \sin \theta)^T \). Then \( Q(\theta) = \langle Y, z \rangle = \cos \theta B_1 + \sin \theta B_2 \) is an orthogonal matrix for each real \( \theta \). As the singular value of \( Q(\theta) \) are equal to 1 we deduce that \( \|Y\|_{\sigma,R} = 1 \). Clearly

\[
\langle Y, e_1 \otimes e_1 \otimes e_2 \rangle = \langle Y, e_1 \otimes e_2 \otimes e_1 \rangle = \langle Y, e_2 \otimes e_1 \otimes e_1 \rangle = 1.
\]

Lemma 8.1 yields that the decomposition (32) is the nuclear decomposition over \( R \). Hence \( \|B_1\|_{*,R} = \sqrt{3} \). Banach’s theorem yields that \( B_1 \) has a nuclear decomposition of the form \( B_1 = \sum_{i=1}^r x_i^{\otimes 3} \).

Here is such a decomposition:

\[
B_1 = \frac{4}{3\sqrt{3}} \left[ \sqrt{\frac{3}{2}} e_1 + \frac{1}{2} e_2 \right]^{\otimes 3} + \left[ -\sqrt{\frac{3}{2}} e_1 + \frac{1}{2} e_2 \right]^{\otimes 3} + \frac{1}{4} (-e_2)^{\otimes 3}.
\]

The minimality of this decomposition follows from the identities

\[
\langle Y, (-e_2)^{\otimes 3} \rangle = \langle Y, \left[ \sqrt{\frac{3}{2}} e_1 + \frac{1}{2} e_2 \right]^{\otimes 3} \rangle = \langle Y, \left[ -\sqrt{\frac{3}{2}} e_1 + \frac{1}{2} e_2 \right]^{\otimes 3} \rangle = 1.
\]

To prove the equality \( \|B_1\|_{*,C} = 1.5 \) we first observe the following symmetric decomposition of \( B_1 \) over \( C \):

\[
B_1 = \frac{3}{8} \left( \left[ \frac{\sqrt{2}}{\sqrt{3}} e_1 + \frac{1}{\sqrt{3}} e_2 \right]^{\otimes 3} + \left[ -\frac{\sqrt{2}}{\sqrt{3}} e_1 + \frac{1}{\sqrt{3}} e_2 \right]^{\otimes 3} + \left[ i\frac{\sqrt{2}}{\sqrt{3}} e_1 - \frac{1}{\sqrt{3}} e_2 \right]^{\otimes 3} + \left[ -i\frac{\sqrt{2}}{\sqrt{3}} e_1 + \frac{1}{\sqrt{3}} e_2 \right]^{\otimes 3} \right)
\]
We next observe that
\[ \left\langle S_1, \left[ \frac{\sqrt{2}}{\sqrt{3}} e_1 + \frac{1}{\sqrt{3}} e_2 \right]^{\otimes 3} \right\rangle = \left\langle S_1, \left[ -\frac{\sqrt{2}}{\sqrt{3}} e_1 + \frac{1}{\sqrt{3}} e_2 \right]^{\otimes 3} \right\rangle = \left\langle S_1, \left[ \frac{\sqrt{2}}{\sqrt{3}} e_1 - \frac{1}{\sqrt{3}} e_2 \right]^{\otimes 3} \right\rangle = \| B_1 \|_{\sigma, C}. \]

Lemma 8.1 yields that the decomposition (39) is the nuclear decomposition over \( \mathbb{C} \). Hence \( \| B_1 \| = 3/2 \). Note that this nuclear decomposition over \( \mathbb{C} \) has four terms while the rank of \( B_1 \) is three.

We now discuss the spectral norms of \( B_2 \) over \( \mathbb{R} \) and \( \mathbb{C} \). Let \( x = (x, y)^T, |x|^2 + |y|^2 = 1 \). Then \( g(x, y) := \langle Y, x^{\otimes 3} \rangle = 3x^2y - y^3 = y(3x^2 - y^2) \). Suppose first that \( x, y \in \mathbb{R} \). Then \( x^2 = (1 - y^2) \) and \( g(x, y) = y(3 - 4y^2), y \in [0, 1] \). The maximum of this polynomial on \([0, 1]\) is achieved for \( y = 1/2 \). So \( g(x, y) \leq 1 = g(\sqrt{3}/2, 1/2) \). Hence \( \| B_2 \| = 1/2 \). Assume now that \( x, y \in \mathbb{C} \). Clearly, \( |g(x, y)| \leq |y|(3|x|^2 + |y|^2) \). Choose \( y = -t, x = is \) where \( s, t \geq 0, s^2 + t^2 = 1 \). Then \( g(x, y) = h(s, t) = t(3s^2 + t^2) = t(3 - 2t^2) \) for \( t \in [0, 1] \). The maximum of this polynomial is \( \sqrt{2} \) achieved for \( t = 1/\sqrt{2} \). This shows that \( \| B_2 \|_{\sigma, \mathbb{C}} = 1/\sqrt{2} \).

It is left to show that \( \| B_2 \|_{\sigma, \mathbb{R}} = 2 \). We claim that the decomposition (33) is a nuclear decomposition over \( \mathbb{R} \). This follows from Lemma 8.1 combined with the identities (36) and the first identity in (38). Consider the following \( \mathbb{R} \)-symmetric decomposition of \( B_2 \):
\[
B_2 = \frac{1}{12} ((\sqrt{3}e_1 + e_2)^{\otimes 3} + (-\sqrt{3}e_1 + e_2)^{\otimes 3} + (-e_2)^{\otimes 3}).
\]

In view of Lemma 8.1 and the equalities (38) we deduce that the above decomposition is a nuclear decomposition over \( \mathbb{R} \). Note that the \( \mathbb{R} \)-rank of \( B_2 \) is three. So a nuclear decomposition of \( B_2 \) over \( \mathbb{R} \) (38) is also a rank decomposition of \( B_2 \) over \( \mathbb{R} \). However a nuclear decomposition of \( B_2 \) over \( \mathbb{R} \) given by (33) is not a rank decomposition over \( \mathbb{R} \).

10. Multilinear rank and nuclear norm

For matrix \( A = (a_{ij}) \in \mathbb{C}^{m \times n} \), we have a nice equality between three numerical invariants associated with \( A \):
\[
\text{rank}(A) = \dim \text{span}_\mathbb{C} \{ A_{1}, \ldots, A_{n} \} = \dim \text{span}_\mathbb{C} \{ A_{1}, \ldots, A_{m} \} = \min \{ r : A = xy^T + \cdots + x_r y_r^T \}. \tag{43}
\]
Here we let \( A_{i} = (a_{i1}, \ldots, a_{in}) \in \mathbb{C}^n \) and \( A_{j} = (a_{ij}, \ldots, a_{nj}) \in \mathbb{C}^n \) denote the \( i \)th row and \( j \)th column vectors of \( A \). The numbers in (42) and (41) are the row and column ranks of \( A \). Their equality is a standard fact in linear algebra and the common value is called the rank of \( A \). The number in (43) is also easily seen to be equal to \( \text{rank}(A) \), a fact that follows either from a direct proof or can be deduced from any rank revealing factorization of \( A \).

For a tensor \( A = (a_{ijk}) \in \mathbb{C}^{l \times m \times n} \), one may also define analogous numbers (which, like their matrix counterparts, are also invariant under the action of \( \text{GL}_{l,m,n}(\mathbb{C}) \)):
\[
r_1 = \dim \text{span}_\mathbb{C} \{ A_{1}, \ldots, A_{m} \}, \tag{44}
\]
\[
r_2 = \dim \text{span}_\mathbb{C} \{ A_{1}, \ldots, A_{n} \}, \tag{45}
\]
\[
r_3 = \dim \text{span}_\mathbb{C} \{ A_{1}, \ldots, A_{n} \}, \tag{46}
\]
\[
r = \min \{ r : A = x_1 \otimes y_1 \otimes z_1 + \cdots + x_r \otimes y_r \otimes z_r \}. \tag{43}
\]
Here \( A_{i} = (a_{ijk})_{j,k=1}^{m,n} \in \mathbb{C}^{m \times n} \), \( A_{j} = (a_{ijk})_{i,k=1}^{l,m} \in \mathbb{C}^{l \times n} \), \( A_{k} = (a_{ijk})_{i,j=1}^{l,m} \in \mathbb{C}^{l \times m} \) are ‘matrix slices’ of the 3-tensor — the analogues of the row and column vectors of a matrix. The last number is of course the tensor rank that have discussed in the last few section. In this case, however, we
generally have that $r_1, r_2, r_3, r$ are all distinct numbers, although some simple inequalities hold, the most obvious one being:

$$\max(r_1, r_2, r_3) \leq r \leq \min(r_1r_2, r_1r_3, r_2r_3).$$

We refer the reader to [3] for several other such relations among these numbers. The notion of multilinear rank was due also to Hitchcock [17], as a special case (2-plex rank) of his *multiplex rank*.

Since in general $r_1 \neq r_2 \neq r_3$, for a 3-tensor $A \in \mathbb{C}^{l \times m \times n}$, all three numbers would have to be recorded, which leads us to the following notion.

**Definition 10.1.** The multilinear rank of $A \in \mathbb{C}^{l \times m \times n}$ is $\mu\text{rank}(A) := (r_1, r_2, r_3)$ as defined in (44), (45), (46).

The multilinear rank is essentially matrix rank and so inherits many of the latter’s properties. So we do not see the sort of anomalies like border rank. To qualify this claim, we define the *flattening maps*

$$b_1 : \mathbb{C}^{l \times m \times n} \to \mathbb{C}^{l \times mn}, \ b_2 : \mathbb{C}^{l \times m \times n} \to \mathbb{C}^{m \times ln}, \ b_3 : \mathbb{C}^{l \times m \times n} \to \mathbb{C}^{n \times lm}$$

that takes a 3-tensor $A \in \mathbb{C}^{l \times m \times n}$, viewed as a cubic array of numbers, and ‘flatten’ it in three different ways to yield square array of numbers. Instead of giving explicit formulae, we will illustrate these maps with an example: Let

$$A = \begin{bmatrix} a_{111} & a_{121} & a_{131} & a_{112} & a_{122} & a_{132} \\ a_{211} & a_{221} & a_{231} & a_{212} & a_{222} & a_{232} \\ a_{311} & a_{321} & a_{331} & a_{312} & a_{322} & a_{332} \\ a_{411} & a_{421} & a_{431} & a_{412} & a_{422} & a_{432} \end{bmatrix} \in \mathbb{C}^{4 \times 3 \times 2},$$

then

$$b_1(A) = \begin{bmatrix} a_{111} & a_{112} & a_{121} & a_{122} & a_{131} & a_{132} \\ a_{211} & a_{212} & a_{221} & a_{222} & a_{231} & a_{232} \\ a_{311} & a_{312} & a_{321} & a_{322} & a_{331} & a_{332} \\ a_{411} & a_{412} & a_{421} & a_{422} & a_{431} & a_{432} \end{bmatrix} \in \mathbb{C}^{4 \times 6},$$

$$b_2(A) = \begin{bmatrix} a_{111} & a_{112} & a_{121} & a_{122} & a_{131} & a_{132} \\ a_{211} & a_{212} & a_{221} & a_{222} & a_{231} & a_{232} \\ a_{311} & a_{312} & a_{321} & a_{322} & a_{331} & a_{332} \\ a_{411} & a_{412} & a_{421} & a_{422} & a_{431} & a_{432} \end{bmatrix} \in \mathbb{C}^{3 \times 8},$$

$$b_3(A) = \begin{bmatrix} a_{111} & a_{121} & a_{131} & a_{112} & a_{122} & a_{132} \\ a_{211} & a_{221} & a_{231} & a_{212} & a_{222} & a_{232} \\ a_{311} & a_{312} & a_{321} & a_{312} & a_{322} & a_{332} \\ a_{411} & a_{412} & a_{421} & a_{422} & a_{431} & a_{432} \end{bmatrix} \in \mathbb{C}^{2 \times 12}.$$

It follows immediately from Definition 10.1 that the multilinear rank $\mu\text{rank}(A) := (r_1, r_2, r_3)$ is given by

$$r_1 = \text{rank}(b_1(A)), \quad r_2 = \text{rank}(b_2(A)), \quad r_3 = \text{rank}(b_3(A)),$$

where rank here is of course the usual matrix rank of the matrices $b_1(A), b_2(A), b_3(A)$.

One recently popular definition of tensor nuclear norm is as the arithmetic mean of the (matrix) nuclear norm of the flattenings:

$$\|A\|_b = \frac{1}{3}(\|b_1(A)\|_\star + \|b_2(A)\|_\star + \|b_3(A)\|_\star).$$

### 11. Approximability of Tensor Nuclear Norm and Spectral Norm

Let $\|\cdot\|_k$ be a sequence of norms in $\mathbb{R}^n$ for $n \in \mathbb{N}$. Denote by $\|\cdot\|_n^\vee$ the dual norm in $\mathbb{R}^n$:

$$\|x\|_n^\vee := \max_{\|y\|_n \leq 1} y^T x = \max \langle x, y \rangle.$$

We claim that $\|x\|_n$ is polynomially computed if and only if $\|x\|_n^\vee$ polynomially computed.
We first try to approximate the spectral norm. Let $b_k(A)$ be the flattening of $b_k(A)$ in the kth index. Let $\|b_k(A)\|_\sigma = \sigma_1(b_k(A))$ be the spectral of the flattened matrix. Clearly, $\|A\|_\sigma \leq \sigma_1(b_k(A))$. Hence

$$\|A\|_\sigma \leq \min\{\sigma_1(b_1(A)), \sigma_1(b_2(A)), \sigma_1(b_3(A))\}. \quad (47)$$

We next observe the following characterization of $\|A\|_\star$:

$$\|A\|_\star = \min\left\{\sum_{i=1}^N \|x_i\|_A : A = \sum_{i=1}^N x_i \otimes A_i, x_i \in \mathbb{R}^l, A_i \in \mathbb{R}^{m \times n}, N \in \mathbb{N}\right\}. \quad (48)$$

Indeed, if we assume that $A_i = y_i \otimes z_i$ for $i = 1, \ldots, N$, then the minimal characterization (48) reduces to the minimal characterization (3). So the minimum of (3) is not more than the minimum given by (48), which is $\|A\|_\star$. On the other hand, write each $A_i$ as the sum of rank one matrices given by the singular value decomposition to deduce that (48) reduces (3).

Recall that

$$\|A_i\|_F \leq \|A_i\|_\star \leq \sqrt{\text{rank} A_i} \|A_i\|_F.$$ 

Clearly, $\text{rank} A_i \leq \min\{r_2(A), r_3(A)\}$. Since

$$\|b_1(A)\|_\star = \min\left\{\sum_{i=1}^N \|x_i\|_F : A = \sum_{i=1}^N x_i \otimes A_i, x_i \in \mathbb{R}^l, A_i \in \mathbb{R}^{m \times n}, N \in \mathbb{N}\right\},$$

we obtain

$$\|b_1(A)\|_\star \leq \|A\|_\star \leq \sqrt{\min\{r_2(A), r_3(A)\}} \|b_1(A)\|_\star.$$ 

Hence, we have the following relation for the dual norms.

$$\|b_1(A)\|_\sigma \geq \|A\|_\sigma \geq \frac{1}{\sqrt{\min\{r_2(A), r_3(A)\}}} \|b_1(A)\|_\sigma.$$ 

Similar inequalities can be deduced for $A_2$ and $A_3$.

As a corollary we deduce that we have computable bounds on $\|A\|_\sigma$ and $\|A\|_\star$.

ACKNOWLEDGMENT

We thank Harm Derksen and Jiawang Nie for their enormously helpful discussions. SF’s work is partially supported by NSF DMS-1216393. LH’s work is partially supported by AFOSR FA9550-13-1-0133, DARPA D15AP00109, NSF IIS 1546413, DMS 1209136, DMS 1057064.

REFERENCES


Department of Mathematics, Statistics and Computer Science, University of Illinois, Chicago
E-mail address: friedlan@uic.edu

Computational and Applied Mathematics Initiative, Department of Statistics, University of Chicago
E-mail address: lekheng@galton.uchicago.edu