MATHEMATICS OF COMPUTATION
Volume 87, Number 311, May 2018, Pages 1255–1281
http://dx.doi.org/10.1090/mcom/3239
Article electronically published on September 19, 2017

NUCLEAR NORM OF HIGHER-ORDER TENSORS

SHMUEL FRIEDLAND AND LEK-HENG LIM

ABSTRACT. We establish several mathematical and computational properties of the nuclear norm for higher-order tensors. We show that like tensor rank, tensor nuclear norm is dependent on the choice of base field; the value of the nuclear norm of a real 3-tensor depends on whether we regard it as a real 3-tensor or a complex 3-tensor with real entries. We show that every tensor has a nuclear norm attaining decomposition and every symmetric tensor has a symmetric nuclear norm attaining decomposition. There is a corresponding notion of nuclear rank that, unlike tensor rank, is lower semicontinuous. We establish an analogue of Banach's theorem for tensor spectral norm and Comon's conjecture for tensor rank; for a symmetric tensor, its symmetric nuclear norm always equals its nuclear norm. We show that computing tensor nuclear norm is NP-hard in several ways. Deciding weak membership in the nuclear norm unit ball of 3-tensors is NP-hard, as is finding an ε-approximation of nuclear norm for 3-tensors. In addition, the problem of computing spectral or nuclear norm of a 4-tensor is NP-hard, even if we restrict the 4-tensor to be bi-Hermitian, bisymmetric, positive semidefinite, nonnegative valued, or all of the above. We discuss some simple polynomial-time approximation bounds. As an aside, we show that computing the nuclear (p, q)-norm of a matrix is NP-hard in general but polynomial-time if $p = 1$, $q = 1$, or $p = q = 2$, with closed-form expressions for the nuclear $(1, q)$- and $(p, 1)$-norms.

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 rank, allowing one to relax various intractable rank minimization problems into tractable convex optimization problems. More generally, for $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, the nuclear norm of a d-tensor $A \in \mathbb{F}^{n_1} \otimes \cdots \otimes \mathbb{F}^{n_d} = \mathbb{F}^{n_1 \times \cdots \times n_d}$ is defined by

$$(1.1)\quad \|A\|_{*,\mathbb{F}} = \inf\Bigl\{\sum\nolimits_{i=1}^{r}|\lambda_i| : A = \sum\nolimits_{i=1}^{r} \lambda_i u_{1,i} \otimes \cdots \otimes u_{d,i},\ \|u_{k,i}\| = 1,\ r \in \mathbb{N}\Bigr\},$$

where $\|\cdot\|$ is the l^2-norm and $u_{k,i} \in \mathbb{F}^{n_k}$ for $k = 1, \dots, d$, $i = 1, \dots, r$. The nuclear norm of a matrix is then the case when $d = 2$ and is equivalent to the usual definition as a sum of singular values, also known as the Schatten 1-norm [9]. For higher-order tensors it was defined explicitly in [18, 19] (see also [6, 10]) although the original idea dates back to Grothendieck [12] and Schatten [24]. In Section 2 we will discuss the definitions and basic properties of Hilbert–Schmidt, spectral, and

Received by the editor May 28, 2016, and, in revised form, November 2, 2016.
2010 *Mathematics Subject Classification.* Primary 15A69, 90C60; Secondary 47B10, 68Q25.
Key words and phrases. Tensor nuclear norm, tensor spectral norm, tensor rank, nuclear rank, nuclear decomposition, NP-hard, dual norm.
The first author was partially supported by NSF DMS-1216393.
The second author was partially supported by AFOSR FA9550-13-1-0133, DARPA D15AP00109, NSF IIS 1546413, DMS 1209136, DMS 1057064.

nuclear norms for tensors of arbitrary orders over $\mathbb{C}$ and $\mathbb{R}$ as well as their relations with the projective and injective norms in operator theory.

1.1. **Mathematical properties of tensor nuclear norm.** We start by showing in Section 3 that the expression in (1.1) defines a norm and that the infimum is always attained, i.e., there is a finite r and a decomposition into a linear combination of r norm-one rank-one terms such that the l^1-norm of the r coefficients gives the nuclear norm. We call this a *nuclear decomposition*. Such a decomposition gives a corresponding notion of *nuclear rank* that, unlike the usual tensor rank, is lower semicontinuous and thus avoids the ill-posedness issues in the best rank-r approximation problem for tensor rank [4]. As an aside, we show that one cannot get a Schatten p-norm for tensors in this manner: If the l^1-norm of the coefficients is replaced by an l^p-norm for any $p > 1$, the infimum is identically zero. In Section 4, we give a necessary and sufficient condition for checking whether a given decomposition of a tensor into rank-one terms is a nuclear decomposition of that tensor. We also show that every norm on a real finite-dimensional vector space may be regarded as a nuclear norm in an appropriate sense.

For notational simplicity let $d = 3$ but the following conjecture and results may be stated for any $d \geq 3$. Let $A \in \mathsf{S}^3(\mathbb{F}^n)$ be a symmetric tensor. Comon's conjecture [3] asserts that the rank and symmetric rank of A are always equal, i.e.,

$$\min\Big\{r : A = \sum\nolimits_{i=1}^{r} \lambda_i u_i \otimes v_i \otimes w_i\Big\} \stackrel{?}{=} \min\Big\{r : A = \sum\nolimits_{i=1}^{r} \lambda_i v_i \otimes v_i \otimes v_i\Big\}. \tag{1.2}$$

Banach's theorem [1,8] on the other hand shows that the analogous statement holds for the spectral norm in place of rank, i.e.,

$$\sup_{x,y,z\neq 0} \frac{|\langle A, x \otimes y \otimes z\rangle|}{\|x\|\|y\|\|z\|} = \sup_{x\neq 0} \frac{|\langle A, x \otimes x \otimes x\rangle|}{\|x\|^3}.$$

We prove the analogous statement for nuclear norm (for arbitrary d) in Section 5:

$$\begin{aligned}\inf\Big\{\sum\nolimits_{i=1}^{r} |\lambda_i| : A = \sum\nolimits_{i=1}^{r} \lambda_i u_i \otimes v_i \otimes w_i\Big\} \\ = \inf\Big\{\sum\nolimits_{i=1}^{r} |\lambda_i| : A = \sum\nolimits_{i=1}^{r} \lambda_i v_i \otimes v_i \otimes v_i\Big\},\end{aligned} \tag{1.3}$$

where the infimum is taken over all $r \in \mathbb{N}$ and $\|u_i\| = \|v_i\| = \|w_i\| = 1$, $i = 1, \dots, r$. This may be viewed as a dual version of Banach's theorem or, if we regard tensor nuclear norm as a continuous proxy for tensor rank, then this shows that the continuous analogue of Comon's conjecture is true. In addition, we show that every symmetric tensor over $\mathbb{F}$ has a *symmetric nuclear decomposition* over $\mathbb{F}$, i.e., a decomposition that attains the right-hand side of (1.3).

Tensor rank is known to depend on the choice of base field [2, 4]. We show in Section 6 that the same is true for nuclear and spectral norms. If we define $B, C \in \mathbb{R}^{2\times 2\times 2} \subseteq \mathbb{C}^{2\times 2\times 2}$ by

$$\begin{aligned} B &= \frac{1}{2}(e_1 \otimes e_1 \otimes e_2 + e_1 \otimes e_2 \otimes e_1 + e_2 \otimes e_1 \otimes e_1 - e_2 \otimes e_2 \otimes e_2), \\ C &= \frac{1}{\sqrt{3}}(e_1 \otimes e_1 \otimes e_2 + e_1 \otimes e_2 \otimes e_1 + e_2 \otimes e_1 \otimes e_1), \end{aligned}$$

where $e_1, e_2 \in \mathbb{R}^2$ are the standard basis vectors, then

$$\|B\|_{\sigma,\mathbb{R}} = 1/2 < 1/\sqrt{2} = \|B\|_{\sigma,\mathbb{C}}, \qquad \|C\|_{*,\mathbb{C}} = 3/2 < \sqrt{3} = \|C\|_{*,\mathbb{R}}.$$

$\|\cdot\|_{\sigma,\mathbb{F}}$ denotes the spectral norm over the field $\mathbb{F}$ and is defined in (2.2). We give explicit nuclear decompositions and symmetric nuclear decompositions of B and C over $\mathbb{R}$ and $\mathbb{C}$.

As our title indicates, most of this article is about nuclear norms of d-tensors where $d \geq 3$. Section 7 is an exception in that it is about the nuclear (p,q)-norm for matrices,

$$\|A\|_{*,p,q} = \inf\Big\{\sum\nolimits_{i=1}^{r}|\lambda_i| : A = \sum\nolimits_{i=1}^{r} \lambda_i u_i v_i^{\mathsf{T}},\ \|u_i\|_p = \|v_i\|_q = 1,\ r \in \mathbb{N}\Big\}.$$

We discuss its computational complexity—polynomial-time if $p = 1$ or $q = 1$ or $p = q = 2$, but NP-hard otherwise—and show that the nuclear $(1,q)$- and $(p,1)$-norms have nice closed-form expressions.

1.2. **Computational properties of tensor nuclear norm.** More generally, we may also define the nuclear p-norm of a d-tensor $A \in \mathbb{F}^{n_1\times\cdots\times n_d}$ by

$$\|A\|_{*,p} = \inf\Big\{\sum\nolimits_{i=1}^{r}|\lambda_i| : A = \sum\nolimits_{i=1}^{r} \lambda_i u_{1,i}\otimes\cdots\otimes u_{d,i},\ \|u_{k,i}\|_p = 1,\ r \in \mathbb{N}\Big\},$$

where $\|\cdot\|_p$ is the l^p-norm and $u_{k,i} \in \mathbb{F}^{n_k}$ for $k = 1,\dots,d$, $i = 1,\dots,r$. When $p = 2$, the nuclear 2-norm is just the nuclear norm in (1.1).

For the special case $d = p = 2$, the matrix nuclear norm is polynomial-time computable to arbitrary accuracy, as we had pointed out above. Obviously, the computational tractability of the matrix nuclear norm is critical to its recent widespread use. In Sections 7 and 8, we discuss the computational complexity of the nuclear norm in cases when $p \neq 2$ and $d \neq 2$. We will show that the following norms are all NP-hard to compute:

(i) nuclear p-norm of 2-tensors if $p \neq 1, 2, \infty$,
(ii) nuclear 2-norm of d-tensors over $\mathbb{R}$ for all $d \geq 3$,
(iii) nuclear 2-norm of d-tensors over $\mathbb{C}$ for all $d \geq 4$.

We rely on our earlier work [11] for (i) and (ii): The NP-hardness of the nuclear p-norm of 2-tensors follows from that of the operator p-norm for $p \neq 1, 2, \infty$ [13]; the NP-hardness of the nuclear norm of real 3-tensors follows from that of the spectral norm of real 3-tensors [14].

For (iii), we establish a stronger result: we show that even if we require our 4-tensor to be bi-Hermitian, bisymmetric, positive semidefinite, nonnegative-valued, or all of the above, the problem of deciding its weak membership in either the spectral or nuclear norm unit ball in $\mathbb{C}^{n\times n\times n\times n}$ remains NP-hard. We provide a direct proof by showing that the clique number of a graph (well-known to be NP-hard) is the spectral norm of a 4-tensor satisfying these properties, and applying [11] to deduce the corresponding result for nuclear norm. Since we do not regard d-tensors as special cases of $(d+1)$-tensors, we provide a simple argument for extending such hardness results to higher order, giving us the required NP-hardness when $d \geq 3$ (for real tensors) and $d \geq 4$ (for complex tensors).

These hardness results may be stated in an alternative form, namely, the nuclear p-norm of 2-tensors for $p \neq 1, 2, \infty$, 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. We provide some simple polynomial-time computable approximation bounds for the spectral and nuclear norms in Section 9.

2. Hilbert–Schmidt, spectral, and nuclear norms of tensors

We let $\mathbb{F}$ denote either $\mathbb{R}$ or $\mathbb{C}$ throughout this article. A result stated for $\mathbb{F}$ holds true for both $\mathbb{R}$ and $\mathbb{C}$. 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 $n_1,\dots,n_d\in\mathbb{N}$. If desired, these may be viewed as d-dimensional hypermatrices $A=(a_{i_1\cdots i_d})$ with entries $a_{i_1\cdots i_d}\in\mathbb{F}$.

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

$$\langle A,B\rangle = \sum\nolimits_{i_1,\dots,i_d=1}^{n_1,\dots,n_d} a_{i_1\cdots i_d}\overline{b_{i_1\cdots i_d}}. \tag{2.1}$$

When restricted to $\mathbb{R}^{n_1\times\cdots\times n_d}$, (2.1) becomes the Euclidean inner product. This induces the *Hilbert–Schmidt norm* on $\mathbb{F}^{n_1\times\cdots\times n_d}$, denoted by

$$\|A\| = \sqrt{\langle A,A\rangle} = \Bigl(\sum\nolimits_{i_1,\dots,i_d=1}^{n_1,\dots,n_d}|a_{i_1\cdots i_d}|^2\Bigr)^{\frac{1}{2}}.$$

We adopt the convention that an unlabeled $\|\cdot\|$ will always denote the Hilbert–Schmidt norm. When $d=1$, this is the l^2-norm of a vector in $\mathbb{C}^n$ and when $d=2$, this is the Frobenius norm of a matrix in $\mathbb{C}^{m\times n}$. As an $\mathbb{F}$-vector space, $\mathbb{F}^{n_1\times\cdots\times n_d}\simeq\mathbb{F}^n$ where $n=\prod_{k=1}^d n_k$, and the Hilbert–Schmidt norm on $\mathbb{F}^{n_1\times\cdots\times n_d}$ equals the Euclidean norm on $\mathbb{F}^n$.

Let $A\in\mathbb{F}^{n_1\times\cdots\times n_d}$. We define its *spectral norm* by

$$\|A\|_{\sigma,\mathbb{F}} := \sup\biggl\{\frac{|\langle A,x_1\otimes\cdots\otimes x_d\rangle|}{\|x_1\|\cdots\|x_d\|} : 0\neq x_k\in\mathbb{F}^{n_k}\biggr\}, \tag{2.2}$$

and its *nuclear norm* by

$$\|A\|_{*,\mathbb{F}} := \inf\Bigl\{\sum\nolimits_{i=1}^r\|x_{1,i}\|\cdots\|x_{d,i}\| : A=\sum\nolimits_{i=1}^r x_{1,i}\otimes\cdots\otimes x_{d,i},\ x_{k,i}\in\mathbb{F}^{n_k},\ r\in\mathbb{N}\Bigr\}. \tag{2.3}$$

It is straightforward to show that these may also be expressed, respectively, as

$$\|A\|_{\sigma,\mathbb{F}} = \sup\bigl\{|\langle A,u_1\otimes\cdots\otimes u_d\rangle| : \|u_k\|=1\bigr\}, \tag{2.4}$$

$$\|A\|_{*,\mathbb{F}} = \inf\Bigl\{\sum\nolimits_{i=1}^r|\lambda_i| : A=\sum\nolimits_{i=1}^r\lambda_i u_{1,i}\otimes\cdots\otimes u_{d,i},\ \|u_{k,i}\|=1\Bigr\}. \tag{2.5}$$

The Hilbert–Schmidt norm is clearly independent of the choice of base field, i.e., $A\in\mathbb{R}^{n_1\times\cdots\times n_d}\subseteq\mathbb{C}^{n_1\times\cdots\times n_d}$ has the same Hilbert–Schmidt norm whether it is regarded as a real tensor, $A\in\mathbb{R}^{n_1\times\cdots\times n_d}$, or a complex tensor, $A\in\mathbb{C}^{n_1\times\cdots\times n_d}$. As we will see, this is not the case for spectral and nuclear norms when $d>2$, which is why there is a subscript $\mathbb{F}$ in their notations. When $\mathbb{F}=\mathbb{C}$, the absolute value in (2.2) and (2.4) may be replaced by the real part, giving

$$\|A\|_{\sigma,\mathbb{C}} = \sup_{x_k\neq 0}\frac{\operatorname{Re}(\langle A,x_1\otimes\cdots\otimes x_d\rangle)}{\|x_1\|\cdots\|x_d\|} = \sup_{\|u_k\|=1}\operatorname{Re}(\langle A,u_1\otimes\cdots\otimes u_d\rangle).$$

Henceforth we will adopt the convention that whenever the discussion holds for both $\mathbb{F}=\mathbb{R}$ and $\mathbb{C}$, we will drop the subscript $\mathbb{F}$ and write

$$\|\cdot\|_\sigma = \|\cdot\|_{\sigma,\mathbb{F}} \qquad\text{and}\qquad \|\cdot\|_* = \|\cdot\|_{*,\mathbb{F}}.$$

By (2.2) and (2.3), we have

$$|\langle A,B\rangle| \le \|A\|_\sigma\|B\|_*.$$

In fact they are dual norms (see [10, Corollary 5.18] or [19, Lemma 21]) since

$$\|A\|_*^* = \sup_{\|B\|_* \le 1} |\langle A, B\rangle| \le \sup_{\|B\|_* \le 1} \|A\|_\sigma \|B\|_* = \|A\|_\sigma$$

and, on the other hand, it follows from $|\langle A, B\rangle| \le \|A\|_*^* \|B\|_*$ that

$$\|A\|_\sigma = \sup_{\|x_k\|=1} |\langle A, x_1 \otimes \cdots \otimes x_d\rangle| \le \sup_{\|x_k\|=1} \|A\|_*^* \|x_1 \otimes \cdots \otimes x_d\|_* = \|A\|_*^*.$$

It is also easy to see that

$$\|x_1 \otimes \cdots \otimes x_d\| = \|x_1 \otimes \cdots \otimes x_d\|_\sigma = \|x_1 \otimes \cdots \otimes x_d\|_* = \|x_1\| \cdots \|x_d\|.$$

In fact, the following generalization is clear from the definitions (2.2) and (2.3).

Proposition 2.1. *Let* $A \in \mathbb{F}^{n_1 \times \cdots \times n_d}$ *and* $x_1 \in \mathbb{F}^{m_1}, \ldots, x_e \in \mathbb{F}^{m_e}$. *Then*

$$\|A \otimes x_1 \otimes \cdots \otimes x_e\|_{\sigma,\mathbb{F}} = \|A\|_{\sigma,\mathbb{F}} \|x_1\| \cdots \|x_e\|,$$
$$\|A \otimes x_1 \otimes \cdots \otimes x_e\|_{*,\mathbb{F}} = \|A\|_{*,\mathbb{F}} \|x_1\| \cdots \|x_e\|.$$

In this article, we undertake a coordinate dependent point-of-view for broader appeal; a d-tensor is synonymous with a d-dimensional hypermatrix. Nevertheless, we could also have taken a coordinate-free approach. A d-tensor is an element of a tensor product of d vector spaces $V_1, \ldots, V_d$ and choosing a basis on each of these vector spaces allows us to represent the d-tensor $\mathbf{A} \in V_1 \otimes \cdots \otimes V_d$ as a d-hypermatrix $A \in \mathbb{F}^{n_1 \times \cdots \times n_d}$. Strictly speaking, the d-hypermatrix A is a coordinate representation of the d-tensor $\mathbf{A}$ with respect to our choice of bases; a different choice of bases would yield a different hypermatrix for the same tensor [17].

This can be extended to the tensor product of d norm spaces $(V_1, \|\cdot\|_1), \ldots, (V_d, \|\cdot\|_d)$ or d inner product spaces $(V_1, \langle\cdot,\cdot\rangle_1), \ldots, (V_d, \langle\cdot,\cdot\rangle_d)$. For inner product spaces, defining an inner product on rank-one tensors by

$$\langle u_1 \otimes \cdots \otimes u_d, v_1 \otimes \cdots \otimes v_d\rangle := \langle u_1, v_1\rangle_1 \cdots \langle u_d, v_d\rangle_d,$$

and extending bilinearly to the whole of $V_1 \otimes \cdots \otimes V_d$ defines an inner product on $V_1 \otimes \cdots \otimes V_d$. For norm spaces, there are two natural ways of defining a norm on $V_1 \otimes \cdots \otimes V_d$. Let $V_1^*, \ldots, V_d^*$ be the dual spaces[1] of $V_1, \ldots, V_d$. Then

$$\|\mathbf{A}\|_\sigma := \sup\left\{ \frac{|\varphi_1 \otimes \cdots \otimes \varphi_d(\mathbf{A})|}{\|\varphi_1\|_1^* \cdots \|\varphi_d\|_d^*} : 0 \ne \varphi_k \in V_k^* \right\}, \tag{2.6}$$

$$\|\mathbf{A}\|_* := \inf\left\{ \sum_{i=1}^r \|v_{1,i}\|_1 \cdots \|v_{d,i}\|_d : \mathbf{A} = \sum_{i=1}^r v_{1,i} \otimes \cdots \otimes v_{d,i},\ v_{k,i} \in V_k,\ r \in \mathbb{N} \right\}, \tag{2.7}$$

i.e., essentially the spectral and nuclear norm that we defined in (2.2) and (2.3).

For the special case $d = 2$, (2.6) and (2.7) are the well-known *injective* and *projective norms* [5, 12, 20, 23, 24, 27]. In operator theory, $V_1, \ldots, V_d$ are usually infinite-dimensional Banach or Hilbert spaces and so one must allow $r = \infty$ in (2.7). Also, the tensor product $\otimes$ has to be more carefully defined (differently for (2.6) and (2.7)) so that these norms are finite-valued on $V_1 \otimes V_2$.

We are primarily interested in the higher-order case $d \ge 3$ in this article and all our spaces will be finite-dimensional to avoid such complications.

[1] For norm space $(V, \|\cdot\|)$, dual space $V^* := \{\varphi : V \to \mathbb{F} \text{ linear functional}\}$ has *dual norm* $\|\varphi\|^* := \sup_{\|v\|=1} |\varphi(v)|$.

3. Tensor nuclear norm is special

We would like to highlight that (2.3) is the definition of tensor nuclear norm as originally defined by Grothendieck [12] and Schatten [24]. 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 [5,20,23,27] (see also [10,19]). In particular, the nuclear norm defined in (2.3) is precisely the dual norm of the spectral norm in (2.2), is naturally related to the notion of tensor rank [17], and has physical meaning: for a d-Hermitian tensor $A \in (\mathbb{C}^{n_1\times\cdots\times n_d})^2$ representing a density matrix, $\|A\|_{*,\mathbb{C}} = 1$ if and only if A is d-partite separable[2] [7]. Furthermore, it has been shown in [28, Section 3.2] that tensor nuclear norm as defined in (2.3) quantifies optimal numerical stability much like tensor rank quantifies optimal complexity of bilinear operations like matrix-matrix or matrix-vector products. As such, a tensor nuclear norm in this article will always be the one in (2.3) or its equivalent expression (2.5).

One might think that it is possible to extend (2.5) to get a definition of 'Schatten p-norm' for any $p > 1$. Let us take $d = 3$ for illustration. Suppose we define

$$\nu_p(A) \coloneqq \inf\Bigl\{ \Bigl[\sum\nolimits_{i=1}^r |\lambda_i|^p\Bigr]^{1/p} : A = \sum\nolimits_{i=1}^r \lambda_i u_i \otimes v_i \otimes w_i, \\ \|u_i\| = \|v_i\| = \|w_i\| = 1,\ r \in \mathbb{N}\Bigr\}. \tag{3.1}$$

Then $\nu_1 = \|\cdot\|_*$ but in fact ν_p is identically zero for all $p > 1$. To see this, write $u \otimes v \otimes w$ as a sum of 2^n identical terms

$$u \otimes v \otimes w = \tfrac{1}{2^n} u \otimes v \otimes w + \cdots + \tfrac{1}{2^n} u \otimes v \otimes w$$

and observe that if $p > 1$, then

$$\inf_{n\in\mathbb{N}} \Bigl[\sum\nolimits_{i=1}^{2^n} 2^{-np}\Bigr]^{1/p} = \lim_{n\to\infty} 2^{-n(p-1)/p} = 0.$$

This of course also applies to the case $d = 2$ but note that in this case we may impose orthonormality on the factors, i.e.,

$$\nu_p(A) \coloneqq \inf\Bigl\{ \Bigl[\sum\nolimits_{i=1}^r |\lambda_i|^p\Bigr]^{1/p} : A = \sum\nolimits_{i=1}^r \lambda_i u_i \otimes v_i,\ \langle u_i, u_j\rangle = \delta_{ij} = \langle v_i, v_j\rangle \Bigr\},$$

and the result gives us precisely the matrix Schatten p-norm. This is not possible when $d > 2$. A d-tensor $A \in \mathbb{F}^{n_1\times\cdots\times n_d}$ is said to be *orthogonally decomposable* [30] if it has an *orthogonal decomposition* given by

$$A = \sum\nolimits_{i=1}^r \lambda_i u_{1,i} \otimes \cdots \otimes u_{d,i}, \quad \langle u_{k,i}, u_{k,j}\rangle = \delta_{ij}, \quad i,j = 1,\dots,n_k,\ k = 1,\dots,d.$$

There is no loss of generality if we further assume that $\lambda_1 \ge \cdots \ge \lambda_r > 0$. An orthogonal decomposition does not exist when $d \ge 3$, as a simple dimension count would show. Nonetheless we would like to point out that this notion has been vastly generalized in [6].

The case $p = 1$ is also special. In this case (3.1) reduces to (2.5) (for $d = 3$), which indeed defines a norm for any d-tensors.

---

[2]This result appeared in an earlier preprint version of this article (see `https://arxiv.org/abs/1410.6072v1`), but has been moved to a more specialized article [7] focusing on quantum information theory.

Proposition 3.1 (Tensor nuclear norm). *The expression in* (2.3) *or, equivalently,* (2.5) *defines a norm on* $\mathbb{F}^{n_1\times\cdots\times n_d}$. *Furthermore, the infimum is attained and* inf *may be replaced by* min *in* (2.3).

Proof. Consider the set of all norm-one rank-one tensors,

$$\mathcal{E} := \{u_1\otimes\cdots\otimes u_d \in \mathbb{F}^{n_1\times\cdots\times n_d} : \|u_1\| = \cdots = \|u_d\| = 1\}.$$

The Hilbert–Schmidt norm is strictly convex, i.e., for $A, B \in \mathbb{F}^{n_1\times\cdots\times n_d}$, $\|A+B\| < 2$ whenever $A \neq B$, $\|A\| = \|B\| = 1$. Hence in $\mathbb{F}^{n_1\times\cdots\times n_d}$ the extreme points of the unit ball are precisely the points on the unit sphere. It follows that any rank-one tensor $A \in \mathcal{E}$ is not a convex combination of any finite number of points in $\mathcal{E}\setminus\{A\}$. Let $\mathcal{C}$ be the convex hull of $\mathcal{E}$. By Carathéodory's theorem, each tensor $A \in \mathcal{C}$ is a convex combination of $n+1$ rank-one tensors in $\mathcal{E}$. Since $\mathcal{E}$ is compact it follows that $\mathcal{C}$ is also compact. As $\mathcal{C}$ is a compact balanced convex set with 0 as an interior point, it must be a unit ball of some norm ν on $\mathbb{F}^{n_1\times\cdots\times n_d}$. Clearly $\nu(A) = 1$ for all $A \in \mathcal{E}$. So if

$$A = \sum\nolimits_{i=1}^{r} \lambda_i u_{1,i}\otimes\cdots\otimes u_{d,i}, \quad \|u_{1,i}\otimes\cdots\otimes u_{d,i}\| = 1,$$

then

$$\sum\nolimits_{i=1}^{r} |\lambda_i| \geq \nu(A).$$

Hence $\|A\|_* \geq \nu(A)$. We claim that $\|A\|_* = \nu(A)$. Assume first that $\nu(A) = 1$. Then $A \in \{B \in \mathbb{F}^{n_1\times\cdots\times n_d} : \nu(B) \leq 1\} = \mathcal{C}$. So A is a convex combination of a finite number of points in $\mathcal{E}$, i.e.,

$$A = \sum\nolimits_{i=1}^{r} \lambda_i u_{1,i}\otimes\cdots\otimes u_{d,i}, \quad \|u_{1,i}\otimes\cdots\otimes u_{d,i}\| = 1,\ \lambda_1,\dots,\lambda_r > 0,\ \sum\nolimits_{i=1}^{r}\lambda_i = 1.$$

By the definition of nuclear norm (2.5), $\|A\|_* \leq 1 = \nu(A)$. So $\|A\|_* = 1$ and the above decomposition of A attains its nuclear norm. Thus if $\nu(A) = 1$, the infimum in (2.5) is attained. For general $A \neq 0$, we consider $B = \frac{1}{\nu(A)}A$. As $\|B\|_* = \nu(B) = 1$, we have $\nu(A) = \|A\|_*$ and the infimum in (2.5) is likewise attained. ☐

4. Nuclear decompositions of tensors

We will call the nuclear norm attaining decomposition in Proposition 3.1 a nuclear decomposition for short, i.e., for $A \in \mathbb{F}^{n_1\times\cdots\times n_d}$,

$$A = \sum\nolimits_{i=1}^{r} x_{1,i}\otimes\cdots\otimes x_{d,i} \tag{4.1}$$

is a *nuclear decomposition* over $\mathbb{F}$ if and only if

$$\|A\|_{*,\mathbb{F}} = \sum\nolimits_{i=1}^{r} \|x_{1,i}\|\cdots\|x_{d,i}\|, \tag{4.2}$$

where $x_{k,i} \in \mathbb{F}^{n_k}$, $k = 1,\dots,d$, $i = 1,\dots,r$. We define the *nuclear rank* of $A \in \mathbb{F}^{n_1\times\cdots\times n_d}$ by

$$\operatorname{rank}_*(A) := \min\Big\{r \in \mathbb{N} : A = \sum\nolimits_{i=1}^{r} x_{1,i}\otimes\cdots\otimes x_{d,i}, \quad \|A\|_{*,\mathbb{F}} = \sum\nolimits_{i=1}^{r} \|x_{1,i}\|\cdots\|x_{d,i}\|\Big\}, \tag{4.3}$$

and we will call (4.1) a *nuclear rank decomposition* if $r = \operatorname{rank}_*(A)$. Alternatively, we may write the decomposition in a form that resembles the matrix SVD, i.e.,

$$A = \sum\nolimits_{i=1}^{r} \lambda_i u_{1,i}\otimes\cdots\otimes u_{d,i} \tag{4.4}$$

is a nuclear decomposition over $\mathbb{F}$ if and only if

$$\|A\|_{*,\mathbb{F}} = \sum_{i=1}^{r} \lambda_i \qquad \text{and} \qquad \lambda_1 \geq \cdots \geq \lambda_r > 0, \quad \|u_{k,i}\| = 1,$$

where $u_{k,i} \in \mathbb{F}^{n_k}$, $k = 1, \dots, d$, $i = 1, \dots, r$. Unlike the matrix SVD, $\{u_{k,1}, \dots, u_{k,r}\}$ does not need to be orthonormal.

The following lemma provides a way that allows us to check, in principle, when a given decomposition is a nuclear decomposition.

Lemma 4.1. *Let $A \in \mathbb{F}^{n_1 \times \cdots \times n_d}$. Then (4.1) is a nuclear decomposition over $\mathbb{F}$ if and only if there exists $0 \neq B \in \mathbb{F}^{n_1 \times \cdots \times n_d}$ with*

$$\langle B, x_{1,i} \otimes \cdots \otimes x_{d,i} \rangle = \|B\|_{\sigma,\mathbb{F}} \|x_{1,i}\| \cdots \|x_{d,i}\|, \qquad i = 1, \dots, r. \tag{4.5}$$

Alternatively, (4.4) is a nuclear decomposition over $\mathbb{F}$ if and only if there exists $0 \neq B \in \mathbb{F}^{n_1 \times \cdots \times n_d}$ with

$$\langle B, u_{1,i} \otimes \cdots \otimes u_{d,i} \rangle = \|B\|_{\sigma,\mathbb{F}}, \qquad i = 1, \dots, r.$$

Proof. Since the nuclear and spectral norms are dual norms,

$$\operatorname{Re}\langle A, B \rangle \leq \|A\|_{*,\mathbb{F}} \|B\|_{\sigma,\mathbb{F}}.$$

Suppose $\|B\|_{\sigma,\mathbb{F}} = 1$ and $A \neq 0$. Then $\operatorname{Re}\langle A, B \rangle = \|A\|_{*,\mathbb{F}} \|B\|_{\sigma,\mathbb{F}}$ if and only if the real functional $X \mapsto \operatorname{Re}\langle X, B \rangle$ is a supporting hyperplane of the ball $\{X \in \mathbb{F}^{n_1 \times \cdots \times n_d} : \|X\|_{*,\mathbb{F}} \leq \|A\|_{*,\mathbb{F}}\}$ at the point $X = A$. So $\operatorname{Re}\langle A, B \rangle = \|A\|_{*,\mathbb{F}}$ is always attained for some B with $\|B\|_{\sigma,\mathbb{F}} = 1$.

Suppose (4.1) is a nuclear decomposition, i.e., (4.2) holds. Let $B \in \mathbb{F}^{n_1 \times \cdots \times n_d}$, $\|B\|_{\sigma,\mathbb{F}} = 1$ be such that $\operatorname{Re}\langle A, B \rangle = \|A\|_{*,\mathbb{F}}$. Then

$$\|A\|_{*,\mathbb{F}} = \operatorname{Re}\langle A, B \rangle = \sum_{i=1}^{r} \operatorname{Re}\langle x_{1,i} \otimes \cdots \otimes x_{d,i}, B \rangle \leq \sum_{i=1}^{r} \prod_{k=1}^{d} \|x_{k,i}\| = \|A\|_{*,\mathbb{F}}.$$

Therefore equality holds and we have (4.5).

Suppose (4.5) holds. We may assume without loss of generality that $\|B\|_{\sigma,\mathbb{F}} = 1$ and $\prod_{k=1}^{d} \|x_{k,i}\| > 0$ for each $i = 1, \dots, r$. Then

$$\begin{aligned} \|A\|_{*,\mathbb{F}} = \|A\|_{*,\mathbb{F}} \|B\|_{\sigma,\mathbb{F}} &\geq \operatorname{Re}\langle A, B \rangle \\ &= \sum_{i=1}^{r} \langle x_{1,i} \otimes \cdots \otimes x_{d,i}, B \rangle = \sum_{i=1}^{r} \prod_{k=1}^{d} \|x_{k,i}\|. \end{aligned}$$

It follows from the minimality in (2.3) that (4.1) is a nuclear decomposition of A. $\square$

In Proposition 4.3(iii), we will generalize Lemma 4.1 to any norm. As an illustration of Lemma 4.1, we prove that for an orthogonally decomposable tensor, every orthogonal decomposition is a nuclear decomposition, a special case of [6, Theorem 1.11].

Corollary 4.2. *Let $A \in \mathbb{F}^{n_1 \times \cdots \times n_d}$ be orthogonally decomposable and let*

$$A = \sum_{i=1}^{r} \lambda_i u_{1,i} \otimes \cdots \otimes u_{d,i}, \quad \langle u_{k,i}, u_{k,j} \rangle = \delta_{ij}, \tag{4.6}$$

be an orthogonal decomposition. Then

$$\|A\| = \Big(\sum_{i=1}^{r} |\lambda_i|^2 \Big)^{1/2}, \quad \|A\|_{\sigma,\mathbb{F}} = \max_{i=1,\dots,r} |\lambda_i|, \quad \|A\|_{*,\mathbb{F}} = |\lambda_1| + \cdots + |\lambda_r|.$$

Proof. The expression for the Hilbert–Schmidt norm is immediate from the Pythagorean theorem since $\{u_{1,i}\otimes\cdots\otimes u_{d,i} : i = 1,\ldots,r\}$ is orthonormal. We may assume that $\lambda_1 \geq \cdots \geq \lambda_r > 0$. Let $v_k \in \mathbb{F}^{n_k}$, $k = 1,\ldots,d$, be unit vectors. Clearly, $|\langle u_{k,i}, v_k\rangle| \leq 1$ for all i and k. By Bessel's inequality, $\sum_{i=1}^r |\langle u_{k,i}, v_k\rangle|^2 \leq |v_k|^2 = 1$ for $k = 1, 2$. Hence

$$\begin{aligned}|\langle A, v_1\otimes\cdots\otimes v_d\rangle| &\leq \sum\nolimits_{i=1}^r \lambda_i |\langle u_{1,i}\otimes\cdots\otimes u_{d,i}, v_1\otimes\cdots\otimes v_d\rangle| \\ &= \sum\nolimits_{i=1}^r \lambda_i \prod\nolimits_{k=1}^d |\langle u_{k,i}, v_k\rangle| \leq \lambda_1 \sum\nolimits_{i=1}^r |\langle u_{1,i}, v_1\rangle||\langle u_{2,i}, v_2\rangle| \\ &\leq \lambda_1\Big(\sum\nolimits_{i=1}^r |\langle u_{1,i}, v_1\rangle|^2\Big)^{1/2}\Big(\sum\nolimits_{i=1}^r |\langle u_{2,i}, v_2\rangle|^2\Big)^{1/2} \leq \lambda_1.\end{aligned}$$

Choose $v_k = u_{k,i}$ for $k = 1,\ldots,d$ to deduce that $\|A\|_{\sigma,\mathbb{F}} = \lambda_1 = \max_{i=1,\ldots,r}\lambda_i$. Now take $B := \sum_{i=1}^r u_{1,i}\otimes\cdots\otimes u_{d,i}$ and observe that $\|B\|_{\sigma,\mathbb{F}} = 1$ and that $\langle B, u_{1,i}\otimes\cdots\otimes u_{d,i}\rangle = 1$ for all $i = 1,\ldots,r$. Hence by Lemma 4.1, (4.6) is a nuclear decomposition and $\|A\|_* = \sum_{i=1}^r \lambda_i$. ☐

For $\mathbb{F} = \mathbb{R}$, we establish a generalization of nuclear decomposition that holds true for any finite-dimensional norm space V. The next result essentially says that 'every norm is a nuclear norm' in an appropriate sense.

Proposition 4.3. *Let V be a real vector space of dimension n and let $\nu : V \to [0,\infty)$ be a norm and $\nu^* : V^* \to [0,\infty)$ its dual norm. Let $\mathcal{E}$ be the set of the extreme points of the unit ball $B_\nu := \{x \in V : \nu(x) \leq 1\}$.*

(i) *If $\nu(x) = 1$, then there exists a decomposition*

$$x = \sum\nolimits_{i=1}^r \lambda_i x_i, \tag{4.7}$$

where $\lambda_1,\ldots,\lambda_r > 0$, $\lambda_1 + \cdots + \lambda_r = 1$, and $x_1,\ldots,x_r \in \mathcal{E}$ are linearly independent.

(ii) *For any $x \in V$,*

$$\nu(x) = \min\Big\{\sum\nolimits_{i=1}^n |\lambda_i| : x = \sum\nolimits_{i=1}^n \lambda_i x_i,\ x_1,\ldots,x_n \in \mathcal{E} \text{ lin. indep.}\Big\}. \tag{4.8}$$

(iii) *If (4.7) holds, then equality holds in*

$$\nu(x) \leq \sum\nolimits_{i=1}^r \lambda_i \tag{4.9}$$

if and only if there exists $\varphi \in V^*$ such that $\nu^*(\varphi) = 1$ and $\varphi(x_1) = \cdots = \varphi(x_r) = 1$.

Proof. We start with (i). Let $\nu(x) = 1$. Then by Krein–Milman, x is a convex combination of the extreme points of B_ν,

$$x = \sum\nolimits_{i=1}^r \lambda_i x_i, \quad x_1,\ldots,x_r \in \mathcal{E}, \quad \lambda_1,\ldots,\lambda_r > 0, \quad \sum\nolimits_{i=1}^r \lambda_i = 1.$$

Let r be minimum. We claim that for such a minimum decomposition $x_1,\ldots,x_r$ must be linearly independent. Suppose not, then there is a nontrivial linear combination

$$\sum\nolimits_{i=1}^r \beta_i x_i = 0. \tag{4.10}$$

We claim that $\sum_{i=1}^r \beta_i = 0$. Suppose not. Then we may assume that $\sum_{i=1}^r \beta_i > 0$ (if not, we replace β_i by $-\beta_i$ in (4.10)). Choose $t > 0$ such that $\lambda_i - t\beta_i \geq 0$ for $i = 1, \dots, r$. Then

$$1 = \nu(x) = \nu\Bigl(\sum\nolimits_{i=1}^r (\lambda_i - t\beta_i)x_i\Bigr) \leq \sum\nolimits_{i=1}^r (\lambda_i - t\beta_i)\nu(x_i)$$
$$= \sum\nolimits_{i=1}^r \lambda_i - t\beta_i = 1 - t\sum\nolimits_{i=1}^r \beta_i < 1,$$

a contradiction. Hence $\sum_{i=1}^r \beta_i = 0$. By our earlier assumption that the linear combination in (4.10) is nontrivial, not all β_i's are zero; so we may choose $t > 0$ such that $\lambda_i - t\beta_i \geq 0$ for all $i = 1, \dots, r$ and $\lambda_i - t\beta_i = 0$ for at least one i. In which case the decomposition $x = \sum_{i=1}^r (\lambda_i - t\beta_i)x_i$ contains fewer than r terms, contradicting the minimality of r. Hence $x_1, \dots, x_r$ are linearly independent. Clearly $r \leq n$.

Next we prove (ii). Since $-B_\nu = B_\nu$, it follows that $-\mathcal{E} = \mathcal{E}$. Since B_ν has nonempty interior, $\operatorname{span}_{\mathbb{R}}(\mathcal{E}) = V$. So any $x \in V$ may be written as a linear combination

$$x = \sum\nolimits_{i=1}^n \lambda_i x_i, \qquad x_1, \dots, x_n \in \mathcal{E} \text{ linearly independent.} \tag{4.11}$$

Since $\nu(x_i) = 1$ for $i = 1, \dots, n$, $\nu(x) \leq \sum_{i=1}^n |\lambda_i|$, and thus the right-hand side of (4.8) is not less than $\nu(x)$. It remains to show that there exist linearly independent $x_1, \dots, x_n \in \mathcal{E}$ such that the decomposition (4.11) attains $\nu(x) = \sum_{i=1}^n |\lambda_i|$. This is trivial for $x = 0$ and we may assume that $x \neq 0$. Upon normalizing, we may further assume that $\nu(x) = 1$. By the earlier part, we have a convex decomposition $x = \sum_{i=1}^r \lambda_i x_i$ where $x_1, \dots, x_r \in \mathcal{E}$ and $\sum_{i=1}^r \lambda_i = 1$. If $r = n$, we are done. If $r < n$, we extend $x_1, \dots, x_r$ to $x_1, \dots, x_n \in \mathcal{E}$, a basis of V; note that this is always possible since $\mathcal{E}$ is a spanning set. Then $x = \sum_{i=1}^n \lambda_i x_i$ by setting $\lambda_i := 0$ for $i = r+1, \dots, n$. Hence $1 = \nu(x) = \sum_{i=1}^n |\lambda_i|$.

For (iii), suppose (4.7) holds for some $\lambda_1, \dots, \lambda_r > 0$ and $x_1, \dots, x_r \in \mathcal{E}$. The inequality $\nu(x) \leq \sum_{i=1}^r \nu(\lambda_i x_i) \leq \sum_{i=1}^r \lambda_i \nu(x_i) = \sum_{i=1}^r \lambda_i$ follows trivially from the properties of a norm. If $\nu(x) = \sum_{i=1}^r \lambda_i$, then as $x \neq 0$, there exists $\varphi \in V^*$ such that $\varphi^*(x) = \nu(x)$ and $\nu^*(\varphi) = 1$. Hence $\nu(x) = \sum_{i=1}^n \lambda_i \varphi(x_i)$. As $\nu(x_i) = 1$, it follows that $\varphi(x_i) \leq 1$, and so $\nu(x) = \sum_{i=1}^r \lambda_i$ implies $\varphi(x_i) = 1$ for $i = 1, \dots, r$ as required. Conversely, if $\varphi \in V^*$ is such that $\nu^*(\varphi) = 1$ and $\varphi^*(x_i) = 1$ for $i = 1, \dots, r$, then $\varphi(x) = \sum_{i=1}^r \lambda_i$. Since

$$\nu(x) = \max\{\psi(x) : \psi \in V^*,\ \nu^*(\psi) = 1\},$$

it follows that $\nu(x) \geq \sum_{i=1}^r \lambda_i$, which yields the required equality as (4.9) always hold. $\square$

For any $0 \neq x \in V$, we may apply Proposition 4.3 to the unit vector $x/\nu(x)$ to obtain a *nuclear decomposition* for x with respect to ν,

$$x = \lambda_1 x_1 + \cdots + \lambda_r x_r, \qquad \nu(x) = \lambda_1 + \cdots + \lambda_r, \qquad \lambda_1 \geq \cdots \geq \lambda_r > 0, \tag{4.12}$$

where $x_1, \dots, x_r$ are extreme points of B_ν. We define the *nuclear rank* of $x \in V$ with respect to ν, denoted by $\operatorname{rank}_\nu(x)$, to be the minimum $r \in \mathbb{N}$ such that (4.12) holds. We set $\operatorname{rank}_\nu(x) = 0$ iff $x = 0$. A nuclear decomposition (4.12) where $r = \operatorname{rank}_\nu(x)$ is called a *nuclear rank decomposition* with respect to ν. Note that the linear independence of $x_1, \dots, x_r$ in (4.12) is automatic if it is a nuclear rank decomposition with respect to any norm.

Proposition 4.4. *Let V be a real vector space of dimension n and let $\nu : V \to [0, \infty)$ be a norm. Suppose $\mathcal{E}$, the set of the extreme points of the unit ball B_ν, is compact. Then the nuclear rank* $\operatorname{rank}_\nu : V \to \mathbb{R}$ *is a lower semicontinuous function, i.e., if $(x_m)_{m=1}^\infty$ is a convergent sequence in V with* $\operatorname{rank}_\nu(x_m) \le r$ *for all $m \in \mathbb{N}$, then $x = \lim_{m\to\infty} x_m$ must have* $\operatorname{rank}_\nu(x) \le r$.

Proof. For each $m \in \mathbb{N}$, since $\operatorname{rank}_\nu(x_m) \le r$, x_m has a nuclear decomposition $x_m = \sum_{i=1}^r \lambda_{m,i} x_{m,i}$ with $\sum_{i=1}^r \lambda_{m,i} = \nu(x_m)$, $\lambda_{m,1}, \dots, \lambda_{m,r} \ge 0$, and $x_{m,1}, \dots, x_{m,r} \in \mathcal{E}$. Since $\mathcal{E}$ is compact, by passing through subsequences r times, we obtain a nuclear decomposition $x = \sum_{i=1}^r \lambda_i x_i$ with $\sum_{i=1}^r \lambda_i = \nu(x)$, $\lambda_1, \dots, \lambda_r \ge 0$, and $x_1, \dots, x_r \in \mathcal{E}$. Hence $\operatorname{rank}_\nu(x) \le r$. $\square$

Readers familiar with *X-rank* [29] (see [21] for a generalization) may observe that for $X = B_\nu$, the X-rank and nuclear rank with respect to ν are related in the following sense:

$$\operatorname{rank}_X(x) = \min\Bigl\{ r \in \mathbb{N} : x = \sum\nolimits_{i=1}^r \lambda_i x_i,\ x_i \in X \Bigr\},$$
$$\operatorname{rank}_\nu(x) = \min\Bigl\{ r \in \mathbb{N} : x = \sum\nolimits_{i=1}^r \lambda_i x_i,\ \|x\|_* = \sum\nolimits_{i=1}^r \|\lambda_i x_i\|,\ x_i \in X \Bigr\}.$$

If $V = \mathbb{R}^{n_1 \times \cdots \times n_d}$ and $\nu = \|\cdot\|_{*,\mathbb{R}}$, then $\mathcal{E} = \{u_1 \otimes \cdots \otimes u_d : \|u_1\| = \cdots = \|u_d\| = 1\}$ and (4.12) gives a nuclear decomposition in the sense it was defined in (4.3). A remarkable fact is that tensor nuclear rank is lower semicontinuous since $\mathcal{E}$ is compact. The lack of lower semicontinuity in tensor rank has been a source of many problems [4], particularly the best rank-r approximation problem for d-tensors does not have a solution when $r \ge 2$ and $d \ge 3$. We note that the use of nuclear rank would alleviate this problem.

Corollary 4.5. *For any $A \in \mathbb{R}^{n_1 \times \cdots \times n_d}$, the best nuclear rank-$r$ approximation problem*

$$\operatorname{argmin}\{\|A - X\| : \operatorname{rank}_*(X) \le r\}$$

always has a solution.

Proof. By Proposition 4.4, $\mathcal{S} = \{X \in \mathbb{R}^{n_1 \times \cdots \times n_d} : \operatorname{rank}_*(X) \le r\}$ is a closed set and the result follows from the fact that in any metric space the distance between a point A and a closed set $\mathcal{S}$ must be attained by some $X \in \mathcal{S}$. $\square$

5. Comon's conjecture and Banach's theorem for nuclear norm

We write $\mathsf{T}^d(\mathbb{F}^n) := (\mathbb{F}^n)^{\otimes d} = \mathbb{F}^{n \times \cdots \times n}$ for the space of cubical d-tensors and $\mathsf{S}^d(\mathbb{F}^n)$ for the subspace of symmetric d-tensors in $\mathsf{T}^d(\mathbb{F}^n)$. See [3] for definition and basic properties of symmetric tensors. Let $A \in \mathsf{S}^d(\mathbb{F}^n)$. Comon's conjecture [3] asserts that the rank and symmetric rank of a symmetric tensor are always equal,

$$\min\left\{ r : A = \sum\nolimits_{i=1}^r \lambda_i u_{1,i} \otimes \cdots \otimes u_{d,i} \right\} \stackrel{?}{=} \min\left\{ r : A = \sum\nolimits_{i=1}^r \lambda_i u_i^{\otimes d} \right\}.$$

Banach's theorem [1], [8] on the other hand shows that the analogous assertion for spectral norm is true over both $\mathbb{R}$ and $\mathbb{C}$,

$$\sup_{x_1, \dots, x_d \neq 0} \frac{|\langle A, x_1 \otimes \cdots \otimes x_d \rangle|}{\|x_1\| \cdots \|x_d\|} = \sup_{x \neq 0} \frac{|\langle A, x^{\otimes d} \rangle|}{\|x\|^d}. \tag{5.1}$$

Here we show that the analogous assertion for nuclear norm is also true over both $\mathbb{R}$ and $\mathbb{C}$,

$$\inf\left\{\sum_{i=1}^{r}|\lambda_i| : A=\sum_{i=1}^{r}\lambda_i u_{1,i}\otimes\cdots\otimes u_{d,i}\right\} = \inf\left\{\sum_{i=1}^{r}|\lambda_i| : A=\sum_{i=1}^{r}\lambda_i u_i^{\otimes d}\right\}. \tag{5.2}$$

We will first prove a slight variation of (5.2) over $\mathbb{R}$ below. Note that (5.2) follows from (5.3). If d is odd in (5.2), we may drop the ε_i's.

Theorem 5.1. *Let $A\in\mathsf{S}^d(\mathbb{R}^n)$. Then*

$$\|A\|_{*,\mathbb{R}} = \min\left\{\sum_{i=1}^{r}\|x_i\|^d : A=\sum_{i=1}^{r}\varepsilon_i x_i^{\otimes d},\ \varepsilon_i\in\{-1,1\}\right\}. \tag{5.3}$$

The infimum is taken over all possible symmetric rank-one decompositions of A with $r\in\mathbb{N}$ and is attained (therefore denoted by minimum).

Proof. Let $\mathcal{C}\coloneqq\operatorname{conv}(\mathcal{E})\subseteq\mathsf{T}^d(\mathbb{R}^n)$ be the convex hull of all vectors of the form

$$\mathcal{E}\coloneqq\{\pm x^{\otimes d} : x\in\mathbb{R}^n,\ \|x\|=1\}.$$

As $x^{\otimes d}+(-x^{\otimes d})=0$, $\mathcal{C}$ is a symmetric set in $\mathsf{S}^d(\mathbb{R}^n)$. Since any symmetric tensor is a linear combination of symmetric rank-one terms $x^{\otimes d}$, $\mathcal{C}$ has nonempty interior in $\mathsf{S}^d(\mathbb{R}^n)$. Hence $\mathcal{C}$ is the unit ball of some norm $\nu:\mathsf{S}^d(\mathbb{R}^n)\to[0,\infty)$. Note that $\nu(x^{\otimes d})\le 1$ for $\|x\|=1$. We claim that each point of $\mathcal{E}$ is an extreme point of $\mathcal{C}$. Indeed, consider the unit ball of the Hilbert–Schmidt norm $\{A\in\mathsf{S}^d(\mathbb{R}^n) : \|A\|\le 1\}$. Note that $\|\pm x^{\otimes d}\|=1$ for $\|x\|=1$, and as $\|\cdot\|$ is a strictly convex function, no point on $\mathcal{E}$ is a convex combination of other points of $\mathcal{E}$. Hence $\nu(\pm x^{\otimes d})=\|x\|^d$ for $\|x\|=1$. The homogeneity of ν implies that $\nu(\pm x^{\otimes d})=\|x\|^d$.

Suppose $A=\sum_{i=1}^r\alpha_i x_i^{\otimes d}$. Then the triangle inequality for ν and the above equality yields $\nu(A)\le\sum_{i=1}^r|\alpha_i|\|x_i\|^d$. By scaling the norm of x_i appropriately, we may assume without loss of generality that $\alpha_i\in\{-1,1\}$ for $i=1,\dots,r$. Hence

$$\nu(A)\le\inf\left\{\sum_{i=1}^{r}\|x_i\|^d : A=\sum_{i=1}^{r}\varepsilon_i x_i^{\otimes d},\ \varepsilon_i\in\{-1,1\}\right\}.$$

We claim that the infimum is attained. It is enough to consider the case $\nu(A)=1$. So $A\in\mathcal{C}$ and A is a convex combination of the extreme points of $\mathcal{C}$, i.e.,

$$A=\sum_{i=1}^{r}t_i\varepsilon_i x_i^{\otimes d},\qquad \sum_{i=1}^{r}t_i=1, \tag{5.4}$$

where $t_i\ge 0$, $\|x_i\|=1$, $\varepsilon_i=\pm1$, for all $i=1,\dots,r$. Since $\dim_{\mathbb{R}}\mathsf{S}^d(\mathbb{R}^n)=\binom{n+d-1}{d}$, Caratheodory's theorem implies that $r\le 1+\binom{n+d-1}{d}$. The triangle inequality gives

$$1=\nu(A)\le\sum_{i=1}^{r}t_i\nu(x_i^{\otimes d})=\sum_{i=1}^{r}t_i=1. \tag{5.5}$$

We deduce from (5.4) and (5.5) that $\nu(A)$ is given by the right-hand side of (5.3).

Let ν^* be the dual norm of ν in $\mathsf{S}^d(\mathbb{R}^n)$. By definition,

$$\nu^*(A)=\max_{B\in\mathsf{S}^d(\mathbb{R}^n),\ \nu(B)\le1}\langle A,B\rangle=\max_{B\in\mathcal{E}}\langle A,B\rangle=\max_{\|x\|=1}|\langle A,x^{\otimes d}\rangle|.$$

Since Banach's theorem (5.1) may be written in the form

$$\|A\|_{\sigma,\mathbb{R}}=\max_{\|x\|=1}|\langle A,x^{\otimes d}\rangle|,$$

we get

$$\nu^*(A)=\|A\|_{\sigma,\mathbb{R}}. \tag{5.6}$$

From the definition of nuclear norm (2.3) and the fact that $\nu(A)$ is given by the right-hand side of (5.3) we deduce that $\|A\|_{*,\mathbb{R}} \le \nu(A)$ for all $A \in \mathsf{S}^d(\mathbb{R}^n)$.

Let $\nu_1 : \mathsf{S}^d(\mathbb{R}^n) \to [0,\infty)$ be the nuclear norm $\|\cdot\|_{*,\mathbb{R}}$ on $\mathsf{T}^d(\mathbb{R}^n)$ restricted to $\mathsf{S}^d(\mathbb{R}^n)$. So $\nu_1(A) = \|A\|_{*,\mathbb{R}}$ for $A \in \mathsf{S}^d(\mathbb{R}^n)$. We claim that $\nu = \nu_1$. Suppose not. Then the ν_1 unit ball $\mathcal{C}_1 \coloneqq \{A : \nu_1(A) \le 1\}$ must strictly contain the ν unit ball, i.e., $\mathcal{C} \subsetneq \mathcal{C}_1$. Let $\nu_1^* : \mathsf{S}^d(\mathbb{R}^n) \to [0,\infty)$ be the dual norm of ν_1. Let $\mathcal{C}^*$ and $\mathcal{C}_1^*$ be the unit balls of ν^* and ν_1^*, respectively. Then $\mathcal{C} \subsetneq \mathcal{C}_1$ implies that $\mathcal{C}_1^* \subsetneq \mathcal{C}^*$. So there exists $A \in \mathsf{S}^d(\mathbb{R}^n)$ such that $\nu_1^*(A) > \nu^*(A)$. Hence

$$\nu^*(A) < \nu_1^*(A) = \max_{B\in\mathsf{S}^d(\mathbb{R}^n),\ \|B\|_{*,\mathbb{R}}\le 1} \langle A, B\rangle \le \max_{B\in\mathsf{T}^d(\mathbb{R}^n),\ \|B\|_{*,\mathbb{R}}\le 1} \langle A, B\rangle = \|A\|_{\sigma,\mathbb{R}},$$

which contradicts (5.6). $\square$

The complex case may be deduced from the real case as follows. Note that the ε_i's in (5.3) are unnecessary regardless of the order d since $\mathbb{C}$ contains all dth roots of unity.

Corollary 5.2. *Let $A \in \mathsf{S}^d(\mathbb{C}^n)$. Then*

$$\|A\|_{*,\mathbb{C}} = \min\left\{\sum\nolimits_{j=1}^{r} \|x_j\|^d : A = \sum\nolimits_{j=1}^{r} x_j^{\otimes d}\right\}.$$

The infimum is taken over all possible symmetric rank-one decompositions of A with $r \in \mathbb{N}$ and is attained (therefore denoted by minimum).

Proof. We identify $\mathsf{T}^d(\mathbb{C}^n)$ with $\mathsf{T}^d(\mathbb{R}^n) \times \mathsf{T}^d(\mathbb{R}^n)$, i.e., we write $B \in \mathsf{T}^d(\mathbb{C}^n)$ as $B = X + iY$ where $X, Y \in \mathsf{T}^d(\mathbb{R}^n)$ and identify B with (X, Y). On $\mathsf{T}^d(\mathbb{R}^n)\times\mathsf{T}^d(\mathbb{R}^n)$, we define a real inner product

$$\langle (X, Y), (W, Z)\rangle = \langle X, W\rangle + \langle Y, Z\rangle = \operatorname{Re}\langle X + iY, W + iZ\rangle,$$

under which the Hilbert–Schmidt norm on $\mathsf{T}^d(\mathbb{C}^n)$ is the same as the Hilbert–Schmidt norm on $\mathsf{T}^d(\mathbb{R}^n) \times \mathsf{T}^d(\mathbb{R}^n)$. The spectral norm on $\mathsf{T}^d(\mathbb{C}^n)$ defined in (2.2) translates to a spectral norm on the real space $\mathsf{T}^d(\mathbb{R}^n) \times \mathsf{T}^d(\mathbb{R}^n)$. Furthermore its dual norm on $\mathsf{T}^d(\mathbb{R}^n) \times \mathsf{T}^d(\mathbb{R}^n)$ is precisely the nuclear norm on $\mathsf{T}^d(\mathbb{C}^n)$ as defined in (2.3). This follows from the observation that the extreme points of the nuclear norm unit ball in $\mathsf{T}^d(\mathbb{C}^n)$ is exactly

$$\mathcal{E} = \{x_1 \otimes \cdots \otimes x_d : x_1, \dots, x_d \in \mathbb{C}^n,\ \|x_1\| = \cdots = \|x_d\| = 1\}.$$

So $\mathsf{S}^d(\mathbb{C}^n)$ may be viewed as a real subspace of $\mathsf{S}^d(\mathbb{R}^n)\times\mathsf{S}^d(\mathbb{R}^n)$. We may repeat the arguments as in the real case and use Banach's theorem (5.1) for complex-valued symmetric tensors. $\square$

An immediate consequence of Theorem 5.1 and Corollary 5.2 is the existence of a *symmetric nuclear decomposition* for symmetric d-tensors.

Corollary 5.3 (Symmetric nuclear decomposition)**.** *Let $A \in \mathsf{S}^d(\mathbb{F}^n)$. Then there exists a decomposition*

$$A = \sum\nolimits_{i=1}^{r} \lambda_i u_i^{\otimes d}$$

with finite $r \in \mathbb{N}, r \le 1 + \binom{n+d-1}{d}$, and $\|u_1\| = \cdots = \|u_r\| = 1$ such that

$$\|A\|_{*,\mathbb{F}} = |\lambda_1| + \cdots + |\lambda_r|.$$

As in [8] we may extend Theorem 5.1 and Corollary 5.2 to partially symmetric tensors. Let $d_1, \dots, d_m \in \mathbb{N}$ and $d = d_1 + \cdots + d_m$. A d-tensor $A \in \mathsf{S}^{d_1}(\mathbb{F}^{n_1}) \otimes \cdots \otimes \mathsf{S}^{d_m}(\mathbb{F}^{n_m})$ is called a $(d_1, \dots, d_m)$-symmetric tensor. The following analogue of Banach's theorem (5.1) for such tensors was established in [8]:

$$\|A\|_{\sigma,\mathbb{F}} = \max_{\|x_i\|=1} |\langle A, x_1^{\otimes d_1} \otimes \cdots \otimes x_m^{\otimes d_m}\rangle|$$

for all $A \in \mathsf{S}^{d_1}(\mathbb{F}^{n_1}) \otimes \cdots \otimes \mathsf{S}^{d_m}(\mathbb{F}^{n_m})$. Using this and the same arguments used to establish Theorem 5.1 and Corollary 5.2, we may obtain the following. Note that the ε_i's in (5.7) may be dropped in all cases except when $\mathbb{F} = \mathbb{R}$ and $d_1, \dots, d_m$ are all even integers.

Corollary 5.4. *Let $A \in \mathsf{S}^{d_1}(\mathbb{F}^{n_1}) \otimes \cdots \otimes \mathsf{S}^{d_m}(\mathbb{F}^{n_m})$. Then*

$$\|A\|_{*,\mathbb{F}} = \min\Big\{\sum_{i=1}^{r} \|x_{1,i}\|^{d_1} \cdots \|x_{m,i}\|^{d_m} : \\ A = \sum_{i=1}^{r} \varepsilon_j x_{1,i}^{\otimes d_1} \otimes \cdots \otimes x_{m,i}^{\otimes d_m},\ \varepsilon_i \in \{-1, 1\}\Big\}. \tag{5.7}$$

6. Base field dependence

It is well-known [2], [4] that tensor rank is dependent on the choice of base fields when the order of the tensor $d \geq 3$. Take any linearly independent $x, y \in \mathbb{R}^n$ and let $z = x + iy \in \mathbb{C}^n$. If we define

$$A \coloneqq x \otimes x \otimes x - x \otimes y \otimes y + y \otimes x \otimes y + y \otimes y \otimes x = \frac{1}{2}(z \otimes \bar{z} \otimes \bar{z} + \bar{z} \otimes z \otimes z),$$

then $\operatorname{rank}_{\mathbb{C}}(A) = 2 < 3 = \operatorname{rank}_{\mathbb{R}}(A)$. We show that the same is true for spectral and nuclear norms of d tensors when $d \geq 3$.

Lemma 6.1. *Let $e_1, e_2 \in \mathbb{R}^2$ be the standard basis vectors. Define $B \in \mathbb{R}^{2\times2\times2} \subseteq \mathbb{C}^{2\times2\times2}$ by*

$$B = \frac{1}{2}(e_1 \otimes e_1 \otimes e_2 + e_1 \otimes e_2 \otimes e_1 + e_2 \otimes e_1 \otimes e_1 - e_2 \otimes e_2 \otimes e_2). \tag{6.1}$$

Then (6.1) is a nuclear decomposition over $\mathbb{R}$, and

$$\|B\|_{\sigma,\mathbb{R}} = \frac{1}{2}, \qquad \|B\|_{\sigma,\mathbb{C}} = \frac{1}{\sqrt{2}}, \qquad \|B\|_{*,\mathbb{R}} = 2, \qquad \|B\|_{*,\mathbb{C}} = \sqrt{2}.$$

Furthermore, $B \in \mathsf{S}^3(\mathbb{R}^2) \subseteq \mathsf{S}^3(\mathbb{C}^2)$ has a symmetric nuclear decomposition over $\mathbb{R}$ given by

$$B = \frac{2}{3}\left(\left[\frac{\sqrt{3}}{2}e_1 + \frac{1}{2}e_2\right]^{\otimes 3} + \left[-\frac{\sqrt{3}}{2}e_1 + \frac{1}{2}e_2\right]^{\otimes 3} + (-e_2)^{\otimes 3}\right), \tag{6.2}$$

and a symmetric nuclear decomposition over $\mathbb{C}$ given by

$$B = \frac{1}{\sqrt{2}}\left(\left[-\frac{1}{\sqrt{2}}e_2 + \frac{i}{\sqrt{2}}e_1\right]^{\otimes 3} + \left[-\frac{1}{\sqrt{2}}e_2 - \frac{i}{\sqrt{2}}e_1\right]^{\otimes 3}\right). \tag{6.3}$$

Proof. Since $B \in \mathsf{S}^3(\mathbb{R}^2)$, we may rely on (5.1) and (5.3) in Section 5 to calculate its spectral and nuclear norms over $\mathbb{R}$ and $\mathbb{C}$. Set $Y = 2B$ for convenience.

Let $x = (x_1, x_2)^{\mathsf{T}}$ with $|x_1|^2 + |x_2|^2 = 1$. Then $g(x_1, x_2) \coloneqq \langle Y, x^{\otimes 3}\rangle = 3x_1^2x_2 - x_2^3 = x_2(3x_1^2 - x_2^2)$. Suppose first that $x_1, x_2 \in \mathbb{R}$. Then $x_1^2 = 1 - x_2^2$ and the maximum of $g(x_1, x_2) = x_2(3 - 4x_2^2)$ over $x_2 \in [0, 1]$ is attained at $x_2 = 1/2$, $x_1 = \sqrt{3}/2$. Hence $\|Y\|_{\sigma,\mathbb{R}} = 1$ and $\|B\|_{\sigma,\mathbb{R}} = 1/2$.

Suppose now that $x_1, x_2 \in \mathbb{C}$. Clearly, $|g(x_1, x_2)| \le |x_2|(3|x_1|^2 + |x_2|^2)$. Choose $x_2 = -t$, $x_1 = is$ where $s, t \ge 0$ and $s^2 + t^2 = 1$. Then the maximum of $g(x_1, x_2) = h(s, t) = t(3s^2 + t^2) = t(3 - 2t^2)$ over $t \in [0, 1]$ is $\sqrt{2}$, attained at $t = 1/\sqrt{2} = s$. Hence $\|B\|_{\sigma,\mathbb{C}} = 1/\sqrt{2}$ and $\|Y\|_{\sigma,\mathbb{C}} = \sqrt{2}$.

That (6.1) is a nuclear decomposition over $\mathbb{R}$ and $\|B\|_{*,\mathbb{R}} = 2$ follows from Lemma 4.1 and the observation

$$\begin{aligned} \langle Y, e_1 \otimes e_1 \otimes e_2 \rangle &= \langle Y, e_1 \otimes e_2 \otimes e_1 \rangle \\ &= \langle Y, e_2 \otimes e_1 \otimes e_1 \rangle = \langle Y, (-e_2)^{\otimes 3} \rangle = 1 = \|Y\|_{\sigma,\mathbb{R}}. \end{aligned} \tag{6.4}$$

That (6.3) is a symmetric nuclear decomposition over $\mathbb{C}$ follows from Lemma 4.1 and the observation

$$\left\langle Y, \left[\frac{1}{\sqrt{2}}(-e_2 + ie_1)\right]^{\otimes 3} \right\rangle = \left\langle Y, \left[\frac{1}{\sqrt{2}}(-e_2 - ie_1)\right]^{\otimes 3} \right\rangle = \sqrt{2} = \|Y\|_{\sigma,\mathbb{C}}.$$

This also shows that $\|B\|_{*,\mathbb{C}} = \sqrt{2}$. $\square$

Lemma 6.2. *Let $e_1, e_2 \in \mathbb{R}^2$ be the standard basis vectors. Define $C \in \mathbb{R}^{2\times 2\times 2} \subseteq \mathbb{C}^{2\times 2\times 2}$ by*

$$C = \frac{1}{\sqrt{3}}(e_1 \otimes e_1 \otimes e_2 + e_1 \otimes e_2 \otimes e_1 + e_2 \otimes e_1 \otimes e_1). \tag{6.5}$$

Then (6.5) is a nuclear decomposition over $\mathbb{R}$, and

$$\|C\|_{\sigma,\mathbb{R}} = \|C\|_{\sigma,\mathbb{C}} = \frac{2}{3}, \qquad \|C\|_{*,\mathbb{R}} = \sqrt{3}, \qquad \|C\|_{*,\mathbb{C}} = \frac{3}{2}. \tag{6.6}$$

Furthermore, $C \in \mathsf{S}^3(\mathbb{R}^2) \subseteq \mathsf{S}^3(\mathbb{C}^2)$ has a symmetric nuclear decomposition over $\mathbb{R}$ given by

$$C = \frac{4}{3\sqrt{3}}\left(\left[\frac{\sqrt{3}}{2}e_1 + \frac{1}{2}e_2\right]^{\otimes 3} + \left[-\frac{\sqrt{3}}{2}e_1 + \frac{1}{2}e_2\right]^{\otimes 3} + \frac{1}{4}(-e_2)^{\otimes 3}\right), \tag{6.7}$$

and a symmetric nuclear decomposition over $\mathbb{C}$ given by

$$\begin{aligned} C = \frac{3}{8}\Bigg(&\left[\sqrt{\frac{2}{3}}e_1 + \frac{1}{\sqrt{3}}e_2\right]^{\otimes 3} + \left[-\sqrt{\frac{2}{3}}e_1 + \frac{1}{\sqrt{3}}e_2\right]^{\otimes 3} \\ &+ \left[i\sqrt{\frac{2}{3}}e_1 - \frac{1}{\sqrt{3}}e_2\right]^{\otimes 3} + \left[-i\sqrt{\frac{2}{3}}e_1 - \frac{1}{\sqrt{3}}e_2\right]^{\otimes 3}\Bigg). \end{aligned} \tag{6.8}$$

Proof. Since C is a symmetric tensor, we may rely on (5.1) and (5.3) in Section 5 to calculate its spectral and nuclear norms over $\mathbb{R}$ and $\mathbb{C}$. Set $X = \sqrt{3}C$ for convenience.

Let $x = (x_1, x_2)^\mathsf{T}$. Then $f(x_1, x_2) := \frac{1}{3}\langle X, x^{\otimes 3}\rangle = x_1^2 x_2$. Clearly $\|X\|_{\sigma,\mathbb{R}} = \|X\|_{\sigma,\mathbb{C}}$ since all entries of X are nonnegative. For the maximum of $|f(x)|$ when $\|x\| = 1$, we may restrict to $x_1, x_2 \ge 0$, $x_1^2 + x_2^2 = 1$. Since the maximum of $f(x_1, x_2) = x_1^2\sqrt{1 - x_1^2}$ over $x_1 \in [0, 1]$ occurs at $x_1^2 = 2/3$, $x_2 = 1/\sqrt{3}$, we get the first two equalities in (6.6).

By Lemma 4.1 and (6.4) in the proof of Lemma 6.1, (6.5) is a nuclear decomposition over $\mathbb{R}$. Hence $\|C\|_{*,\mathbb{R}} = \sqrt{3}$.

By Corollary 5.3, C has symmetric nuclear decompositions over both $\mathbb{R}$ and $\mathbb{C}$. That (6.7) is a symmetric nuclear decomposition over $\mathbb{R}$ follows from Lemma 4.1 and the observation that

$$\langle Y, (-e_2)^{\otimes 3}\rangle = \left\langle Y, \left[\frac{\sqrt{3}}{2}e_1 + \frac{1}{2}e_2\right]^{\otimes 3}\right\rangle = \left\langle Y, \left[-\frac{\sqrt{3}}{2}e_1 + \frac{1}{2}e_2\right]^{\otimes 3}\right\rangle = 1 = \|Y\|_{\sigma,\mathbb{R}},$$

where Y is as defined in the proof of Lemma 6.1. Likewise, (6.8) is a symmetric nuclear decomposition over $\mathbb{C}$ by Lemma 4.1 and the observation that

$$\begin{aligned}\left\langle C, \left[\sqrt{\frac{2}{3}}e_1 + \frac{1}{\sqrt{3}}e_2\right]^{\otimes 3}\right\rangle &= \left\langle C, \left[-\sqrt{\frac{2}{3}}e_1 + \frac{1}{\sqrt{3}}e_2\right]^{\otimes 3}\right\rangle \\ &= \left\langle C, \left[i\sqrt{\frac{2}{3}}e_1 - \frac{1}{\sqrt{3}}e_2\right]^{\otimes 3}\right\rangle = \left\langle C, \left[-i\sqrt{\frac{2}{3}}e_1 - \frac{1}{\sqrt{3}}e_2\right]^{\otimes 3}\right\rangle = \|C\|_{\sigma,\mathbb{C}}.\end{aligned}$$

Since (6.8) is a symmetric nuclear decomposition over $\mathbb{C}$, we obtain $\|C\|_{*,\mathbb{C}} = 3/2$. ☐

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

Lemma 6.3. *Let $A \in \mathbb{C}^{n_1\times\cdots\times n_d}$. Then*

$$\|A\|_{\sigma,\mathbb{C}} \le \|\,|A|\,\|_{\sigma,\mathbb{C}}, \qquad \|\,|A|\,\|_{\sigma,\mathbb{C}} = \|\,|A|\,\|_{\sigma,\mathbb{R}}.$$

Proof. The triangle inequality yields

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

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. ☐

A plausible nuclear norm analogue of the inequality $\|A\|_{\sigma,\mathbb{C}} \le \|\,|A|\,\|_{\sigma,\mathbb{C}}$ is $\|A\|_{*,\mathbb{C}} \le \|\,|A|\,\|_{*,\mathbb{C}}$. It is easy to show that this inequality holds in special cases (e.g. if A is a Hermitian positive semidefinite matrix) but it is false in general. For example, let

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

Then $\|A\|_* = 2 > \sqrt{2} = \|\,|A|\,\|_*$.

7. Nuclear (p,q)-norm of a matrix

In this section, we study the special case where $d = 2$. Let $\|\cdot\|_p$ denote the l^p-norm on $\mathbb{R}^n$, i.e.,

$$\|x\|_p = \left(\sum\nolimits_{i=1}^n |x_i|^p\right)^{1/p}, \qquad \|x\|_\infty = \max\{|x_1|, \dots, |x_n|\}.$$

Recall that the dual norm $\|\cdot\|_p^* = \|\cdot\|_{p^*}$ where $p^* := p/(p-1)$, i.e., $1/p + 1/p^* = 1$.

The nuclear (p,q)-norm of a matrix $A \in \mathbb{R}^{m\times n}$ is

$$\|A\|_{*,p,q} = \inf\left\{\sum\nolimits_{i=1}^r |\lambda_i| : A = \sum\nolimits_{i=1}^r \lambda_i u_i \otimes v_i,\ \|u_i\|_p = \|v_i\|_q = 1,\ r \in \mathbb{N}\right\} \tag{7.1}$$

for any $p, q \in [1, \infty]$. The spectral (p,q)-norm on $\mathbb{R}^{m\times n}$ is

$$\|A\|_{\sigma,p,q} = \max_{x,y\ne 0} \frac{y^{\mathsf{T}}Ax}{\|x\|_p\|y\|_q} = \max_{\|x\|_p=\|y\|_q=1} y^{\mathsf{T}}Ax$$

for any $p, q \in [1, \infty]$. The operator (p, q)-norm on $\mathbb{R}^{m\times n}$ is

$$\|A\|_{p,q} = \max_{x\neq 0} \frac{\|Ax\|_q}{\|x\|_p} = \max_{\|x\|_p=1} \|Ax\|_q$$

for any $p, q \in [1, \infty]$. When $p = q$, we write

$$\|\cdot\|_{p,p} = \|\cdot\|_p, \qquad \|\cdot\|_{\sigma,p,p} = \|\cdot\|_{\sigma,p}, \qquad \|\cdot\|_{*,p,p} = \|\cdot\|_{*,p},$$

and call them the operator, spectral, nuclear p-norm, respectively. The case $p = 2$ gives the usual spectral and nuclear norms.

It is well-known that the operator (p, q)-norm and the spectral (p, q^*)-norm are related via

$$\|A\|_{\sigma,p,q} = \|A\|_{p,q^*} \qquad \text{for all } A \in \mathbb{R}^{m\times n}. \tag{7.2}$$

It follows from $\|Ax\|_q = \max_{\|y\|_{q^*}=1} y^\mathsf{T} Ax$ and $y^\mathsf{T} Ax = x^\mathsf{T} A^\mathsf{T} y$ that

$$\|A^\mathsf{T}\|_{p,q} = \|A\|_{q^*,p^*}, \qquad \|A^\mathsf{T}\|_{\sigma,p,q} = \|A\|_{\sigma,q,p}. \tag{7.3}$$

Equivalently, (7.1) may be written as

$$\|A\|_{*,p,q} := \min\Big\{ \sum\nolimits_{i=1}^{r} \|x_i\|_p \|y_i\|_q : A = \sum\nolimits_{i=1}^{r} x_i \otimes y_i,\ r \in \mathbb{N}\Big\}, \tag{7.4}$$

or as the norm whose unit ball is the convex hull of all rank-one matrices $x \otimes y$, where $\|x\|_p\|y\|_q \le 1$. It is trivial to deduce from (7.4) an analogue of (7.3),

$$\|A^\mathsf{T}\|_{*,q,p} = \|A\|_{*,p,q}. \tag{7.5}$$

Theorem 7.1. *The dual norm of the spectral (p, q)-norm is the nuclear (q, p)-norm. The dual norm of the operator (p, q)-norm is the nuclear (q^*, p)-norm.*

$$\|A\|^*_{\sigma,p,q} = \|A\|_{*,q,p}, \qquad \|A\|^*_{p,q} = \|A\|_{*,q^*,p}$$

for all $A \in \mathbb{R}^{m\times n}$ and all $p, q \in [1, \infty]$.

Proof. We prove the equality on the right and deduce the other from (7.2). As in the proof of Corollary 5.2, the unit ball of the (q^*, p)-nuclear norm $\|\cdot\|_{*,q^*,p}$ on $\mathbb{R}^{m\times n}$ is the convex hull of $\mathcal{E} = \{xy^\mathsf{T} : \|x\|_{q^*} = \|y\|_p = 1\}$. Hence

$$\begin{aligned}\|A\|^*_{*,q^*,p} &= \max_{\|B\|_{*,q^*,p}\le 1} \operatorname{tr}(B^\mathsf{T} A) = \max_{xy^\mathsf{T}\in\mathcal{E}} \operatorname{tr}(yx^\mathsf{T} A)\\ &= \max_{\|x\|_{q^*}=\|y\|_p=1} x^\mathsf{T} Ay = \max_{\|y\|_p=1} \|Ay\|_q = \|A\|_{p,q}. \qquad \square\end{aligned}$$

It is well-known that the operator (p, q)-norm is NP-hard in many instances [13], [25] notably:

(i) $\|\cdot\|_{p,q}$ is NP-hard if $1 \le q < p \le \infty$.
(ii) $\|\cdot\|_p$ is NP-hard if $p \ne 1, 2, \infty$.

The exceptional cases [25] are also well-known:

(iii) $\|\cdot\|_p$ is polynomial-time computable if $p = 1, 2, \infty$.
(iv) $\|\cdot\|_{p,q}$ is polynomial-time computable if $p = 1$ and $1 \le q \le \infty$, or if $q = \infty$ and $1 \le p \le \infty$.

By [11], the computational complexity of norms and their dual norms are polynomial-time interreducible. So we obtain the following from Theorem 7.1:

(v) $\|\cdot\|_{*,p,q}$ is NP-hard if $1 \le p^* < q \le \infty$.
(vi) $\|\cdot\|_{*,p^*,p}$ is NP-hard if $p \ne 1, 2, \infty$.
(vii) $\|\cdot\|_{*,p^*,p}$ is polynomial-time computable if $p = 1, 2, \infty$.

(viii) $\|\cdot\|_{*,p,q}$ is polynomial-time computable if $p=1$ and $1\le q\le\infty$, or if $q=1$ and $1\le p\le\infty$.

In (iv) and (viii), we assume that the values of p and q are rational.

In fact, as further special cases of (viii), the nuclear $(1,p)$-norms and $(p,1)$-norms have closed-form expressions, a consequence of the well-known closed-form expressions for the operator $(1,p)$-norms and (p,∞)-norms.

Proposition 7.2. *Let $e_1,\dots,e_n$ be the standard basis vectors in $\mathbb{R}^n$. Let $A\in\mathbb{R}^{m\times n}$ and write*

$$A=[A_{\bullet 1},\dots,A_{\bullet n}]=\begin{bmatrix}A_{1\bullet}^{\mathsf{T}}\\ \vdots\\ A_{m\bullet}^{\mathsf{T}}\end{bmatrix},$$

$A_{\bullet 1},\dots,A_{\bullet n}\in\mathbb{R}^m$ are the column vectors and $A_{1\bullet},\dots,A_{m\bullet}\in\mathbb{R}^n$ are the row vectors of A. Then

$$\|A\|_{1,p}=\max_{j=1,\dots,n}\|Ae_j\|_p=\max\{\|A_{\bullet 1}\|_p,\dots,\|A_{\bullet n}\|_p\},\tag{7.6}$$

$$\|A\|_{p,\infty}=\max_{i=1,\dots,m}\|A^{\mathsf{T}}e_i\|_{p^*}=\max\{\|A_{1\bullet}\|_{p^*},\dots,\|A_{m\bullet}\|_{p^*}\},\tag{7.7}$$

$$\|A\|_{*,1,p}=\sum\nolimits_{i=1}^{m}\|A^{\mathsf{T}}e_i\|_p=\|A_{1\bullet}\|_p+\dots+\|A_{m\bullet}\|_p,\tag{7.8}$$

$$\|A\|_{*,p,1}=\sum\nolimits_{j=1}^{n}\|Ae_j\|_p=\|A_{\bullet 1}\|_p+\dots+\|A_{\bullet n}\|_p\tag{7.9}$$

for all $p\in[1,\infty]$.

Proof. Note that $\mathcal{C}=\{x\in\mathbb{R}^n:\|x\|_1\le 1\}$ is the convex hull of $\{\pm e_j: j=1,\dots,n\}$. As $x\mapsto\|Ax\|_p$ is a convex function on $\mathcal{C}$, we deduce that $\|A\|_{1,p}=\max_{x\in\mathcal{C}}\|Ax\|_p=\max_{j=1,\dots,n}\|\pm Ae_j\|$. Hence (7.6) holds. Equation (7.7) then follows from (7.3) and (7.6). Now observe that

$$\|A\|_{1,p^*}^*=\max_{\|B\|_{1,p^*}\le 1}\operatorname{tr}(B^{\mathsf{T}}A)=\max_{\|Be_j\|_{p^*}\le 1}\Bigl|\sum\nolimits_{j=1}^{n}(Be_j)^{\mathsf{T}}(Ae_j)\Bigr|=\sum\nolimits_{j=1}^{n}\|Ae_j\|_p.$$

Using Theorem 7.1, we obtain (7.9). Equation (7.8) then follows from (7.5) and (7.9). $\square$

The operator $(\infty,1)$-norm is NP-hard to compute by (i) but it has a well-known expression (7.11) that arises in many applications. We will describe its dual norm, the nuclear ∞-norm. In the following, we let

$$\begin{aligned}\mathbb{E}^n&:=\{\varepsilon=(\varepsilon_1,\dots,\varepsilon_n)^{\mathsf{T}}\in\mathbb{R}^n:\varepsilon_i=\pm1,\ i=1,\dots,n\},\\ \mathbb{E}^m\otimes\mathbb{E}^n&:=\{E=(\varepsilon_{ij})\in\mathbb{R}^{m\times n}:\varepsilon_{ij}=\pm1,\ i=1,\dots,m,\\ &\qquad\qquad j=1,\dots,n,\ \operatorname{rank}(E)=1\}.\end{aligned}$$

Note that $\#\mathbb{E}^n=2^n$ and $\#\mathbb{E}^m\otimes\mathbb{E}^n=2^{m+n-1}$.

Lemma 7.3. *Let $A\in\mathbb{R}^{m\times n}$. Then*

$$\|A\|_{\infty,p}=\max_{\varepsilon\in\mathbb{E}^n}\|A\varepsilon\|_p.\tag{7.10}$$

In particular,

$$\|A\|_{\infty,1}=\max_{\varepsilon_1,\dots,\varepsilon_m,\delta_1,\dots,\delta_n\in\{-1,+1\}}\sum\nolimits_{i=1}^{m}\sum\nolimits_{j=1}^{n}a_{ij}\varepsilon_i\delta_j,\tag{7.11}$$

and its dual norm is

$$\|A\|_{*,\infty} = \min\Bigl\{\sum\nolimits_{i=1}^{mn} |\lambda_i| : A = \sum\nolimits_{i=1}^{mn} \lambda_i E_i, \quad E_1, \dots, E_{mn} \in \mathbb{E}^m \otimes \mathbb{E}^n \text{ linearly independent}\Bigr\}. \tag{7.12}$$

Proof. Observe that the convex hull of $\mathbb{E}^n$ is precisely the unit cube, i.e.,

$$\operatorname{conv}(\mathbb{E}^n) = \{x \in \mathbb{R}^n : \|x\|_\infty \le 1\},$$

giving us (7.10). For $x \in \mathbb{R}^m$, note that $\|x\|_1 = \max_{\varepsilon \in \mathbb{E}^m} \varepsilon^\mathsf{T} x$ and thus

$$\|A\|_{\infty,1} = \max_{\delta \in \mathbb{E}^n} \|A\delta\|_1 = \max_{\varepsilon \in \mathbb{E}^m,\ \delta \in \mathbb{E}^n,} \varepsilon^\mathsf{T} A \delta,$$

giving us (7.11). It follows from Theorem 7.1 that $\|\cdot\|_{\infty,1}^* = \|\cdot\|_{*,\infty,\infty} = \|\cdot\|_{*,\infty}$ and (7.12) follows from Proposition 4.3. $\square$

We have thus far restricted our discussions over $\mathbb{R}$. We may use similar arguments to show that (7.3), (7.5), Theorem 7.1, and Proposition 7.2 all remain true over $\mathbb{C}$. In addition, (7.3) and (7.5) also hold if we have A^* in place of A^T.

Nevertheless for $A \in \mathbb{R}^{m \times n}$, the values of its operator (p, q)-norm over $\mathbb{R}$ and over $\mathbb{C}$ may be different; likewise for its nuclear (p, q)-norm. In fact, a classical result [26] states that $\|A\|_{p,q,\mathbb{C}} = \|A\|_{p,q,\mathbb{R}}$ for all $A \in \mathbb{R}^{m \times n}$ if and only if $p \le q$. We deduce the following analogue for nuclear (p, q)-norm using Theorem 7.1.

Corollary 7.4. *$\|A\|_{*,p,q,\mathbb{C}} = \|A\|_{*,p,q,\mathbb{R}}$ for all $A \in \mathbb{R}^{m \times n}$ if and only if $q \le p^*$.*

8. Tensor nuclear norm is NP-hard

The computational complexity of a norm and that of its dual norm are polynomial-time interreducible [11]. If a norm is polynomial-time computable, then so is its dual; if a norm is NP-hard to compute, then so is its dual. Consequently, computing the nuclear norm of a 3-tensor over $\mathbb{R}$ is NP-hard since computing the spectral norm of a 3-tensor over $\mathbb{R}$ is NP-hard [14]. In fact, it is easy to extend to higher orders by simply invoking Proposition 2.1.

Theorem 8.1. *The spectral and nuclear norms of d-tensors over $\mathbb{R}$ are NP-hard for any $d \ge 3$.*

In this section, we will extend the NP-hardness of tensor spectral and nuclear norms to $\mathbb{C}$. In addition, we will show that even the *weak* membership problem is NP-hard, a *stronger* claim than the membership problem being NP-hard (Theorem 8.1 refers to the membership problem). In the study of various tensor problems, it is sometimes the case that imposing certain special properties on the tensors makes the problems more tractable. Examples of such properties include: (i) even order, (ii) symmetric or Hermitian, (iii) positive semidefinite, (iv) nonnegative valued (we will define these formally later). We will show that computing the spectral or nuclear norm for tensors having all of the aforementioned properties remains an NP-hard problem.

Let $G = (V, E)$ be an undirected graph with vertex set $V := \{1, \dots, n\}$ and edge set $E := \bigl\{\{i_k, j_k\} : k = 1, \dots, m\bigr\}$. Let $\kappa(G)$ be the clique number of G, i.e., the size of the largest clique in G, well-known to be NP-hard to compute [16]. Let

M_G be the adjacency matrix of G, i.e., $m_{ij} = 1 = m_{ji}$ if $\{i, j\} \in E$ and is zero otherwise. Motzkin and Straus [22] showed that

$$\frac{\kappa(G) - 1}{\kappa(G)} = \max_{x \in \Delta^n} x^{\mathsf{T}} M_G x, \tag{8.1}$$

where $\Delta^n := \{x \in \mathbb{R}^n : x \geq 0, \ \|x\|_1 = 1\}$ is the probability simplex. Equality is attained in (8.1) when x is uniformly distributed on the largest clique.

We transform (8.1) into a problem involving 4-tensors. Let $x = y^{\circ 2}$, i.e., $x = (y_1^2, \dots, y_n^2)^{\mathsf{T}}$. Then[3]

$$x^{\mathsf{T}} M_G x = 2 \sum\nolimits_{\{i,j\} \in E} y_i^2 y_j^2. \tag{8.2}$$

For integers $1 \leq s < t \leq n$, let $\mathcal{A}_{st} = \big(a_{ijkl}^{(s,t)}\big)_{i,j,k,l=1}^{n} \in \mathbb{C}^{n \times n \times n \times n}$ be defined by

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

Observe that $\mathcal{A}_{st}$ is *not* a symmetric tensor but we have

$$\langle \mathcal{A}_{st}, y \otimes y \otimes y \otimes y \rangle = 2 y_s^2 y_t^2. \tag{8.3}$$

Definition 8.2. Let $\mathcal{A} = (a_{ijkl})_{i,j,k,l=1}^{m,n,m,n} \in \mathbb{C}^{m \times n \times m \times n}$ be a 4-tensor. We call it *bisymmetric* if

$$a_{ijkl} = a_{klij} \quad \text{for all } i, k = 1, \dots, m, \ j, l = 1, \dots, n,$$

and *bi-Hermitian* if

$$a_{ijkl} = \bar{a}_{klij} \quad \text{for all } i, k = 1, \dots, m, \ j, l = 1, \dots, n.$$

A bi-Hermitian tensor is said to be *bipositive semidefinite* if

$$\sum\nolimits_{i,j,k,l=1}^{m,n,m,n} a_{ijkl} x_{ij} \bar{x}_{kl} \geq 0 \quad \text{for all } X = (x_{ij}) \in \mathbb{C}^{m \times n}.$$

We may regard a 4-tensor $\mathcal{A} = (a_{ijkl})_{i,j,k,l=1}^{m,n,m,n} \in \mathbb{C}^{m \times n \times m \times n}$ as a matrix $M(\mathcal{A}) := [a_{(i,j),(k,l)}] \in \mathbb{C}^{mn \times mn}$, where $a_{(i,j),(k,l)} := a_{ijkl}$. Then $\mathcal{A}$ is bisymmetric, bi-Hermitian, or bipositive semidefinite if and only if $M(\mathcal{A})$ is symmetric, Hermitian, or positive semidefinite.

Clearly bi-Hermitian and bisymmetric are the same notion over $\mathbb{R}$. If $m = n$, a bisymmetric 4-tensor is not necessarily a symmetric 4-tensor although the converse is trivially true. However, if $m = n$, a real bipositive semidefinite tensor $\mathcal{A} \in \mathbb{R}^{n \times n \times n \times n}$ is clearly a positive semidefinite tensor in the usual sense, i.e.,

$$\sum\nolimits_{i,j,k,l=1}^{n,n,n,n} a_{ijkl} x_i x_j x_k x_l \geq 0 \quad \text{for all } x \in \mathbb{R}^n.$$

Lemma 8.3. *The tensor $\mathcal{A}_{st} \in \mathbb{C}^{n \times n \times n \times n}$ is bi-Hermitian, bisymmetric, bipositive semidefinite, and has all entries nonnegative.*

---

[3]By convention, we sum once over each edge; e.g., if $E = \{\{1, 2\}\}$, then $\sum_{\{i,j\} \in E} a_{ij} = a_{12}$, not $a_{12} + a_{21}$.

Proof. It follows from the way it is defined that A_{st} is bi-Hermitian, bisymmetric, and nonnegative valued. It is positive semidefinite because

$$\sum\nolimits_{i,j,s,t=1}^{n,n,n,n} a_{ijkl}^{(s,t)} x_{ij}\bar{x}_{kl} = \frac{1}{2}(x_{st}+x_{ts})(\bar{x}_{st}+\bar{x}_{ts}) \geq 0$$

for all $X = (x_{ij}) \in \mathbb{C}^{n\times n}$. $\square$

$M(A_{st})$ is evidently a nonnegative definite, rank-one matrix with trace one. Those familiar with quantum information theory may note that $M(A_{st})$ represents a bipartite density matrix [7]. For any graph $G = (V, E)$, we define

$$A_G := \sum\nolimits_{\{s,t\}\in E} A_{st} \in \mathbb{C}^{n\times n\times n\times n}. \tag{8.4}$$

Then A_G is bi-Hermitian, bisymmetric, bipositive semidefinite, and has all entries nonnegative. Summing (8.3) over $\{s,t\} \in E$ gives

$$\langle A_G, y\otimes y\otimes y\otimes y\rangle = x^{\mathsf{T}} M_G x, \tag{8.5}$$

where $x = y^{\circ 2}$. Hence

$$\max_{\|y\|=1} \langle A_G, y\otimes y\otimes y\otimes y\rangle = \max_{x\in\Delta^n} x^{\mathsf{T}} M_G x = \frac{\kappa(G)-1}{\kappa(G)}. \tag{8.6}$$

Theorem 8.4. *Let G be a simple undirected graph on n vertices with m edges. Let A_G be defined as in (8.4). Then*

$$\|A_G\|_{\sigma,\mathbb{C}} := \max_{0\neq x,y,u,v\in\mathbb{C}^n} \frac{|\langle A_G, x\otimes y\otimes u\otimes v\rangle|}{\|x\|\|y\|\|u\|\|v\|} = \max_{0\neq y\in\mathbb{R}^n_+} \frac{\langle A_G, y\otimes y\otimes y\otimes y\rangle}{\|y\|^4}. \tag{8.7}$$

Furthermore, we have

$$\frac{\kappa(G)-1}{\kappa(G)} = \|A_G\|_{\sigma,\mathbb{C}} = \|A_G\|_{\sigma,\mathbb{R}}. \tag{8.8}$$

If A_G were a symmetric 4-tensor as opposed to merely bisymmetric, then we may apply Banach's theorem (5.1) to deduce that the maximum is attained at $x = y = u = v$ and thus (8.7) would follow. However A_G is not symmetric and we may not invoke Banach's theorem. Instead we will rely on the following lemma, which may be of independent interest.

Lemma 8.5. *Let $A = (a_{ijkl}) \in \mathbb{C}^{m\times n\times m\times n}$. If $M(A) \in \mathbb{C}^{mn\times mn}$ is Hermitian positive semidefinite, then*

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

Proof. Let $M = M(A)$. Then M is a Hermitian positive semidefinite matrix. Cauchy–Schwarz applied to the sesquilinear form $\bar{w}^{\mathsf{T}} M z$ gives

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

Let $z = \operatorname{vec}(x\otimes y)$ and $w = \operatorname{vec}(\bar{u}\otimes\bar{v}) \in \mathbb{C}^{mn}$ and observe that

$$|\langle A, x\otimes y\otimes u\otimes v\rangle| = |\bar{w}^{\mathsf{T}} M z| \leq \max(\langle A, x\otimes y\otimes\bar{x}\otimes\bar{y}\rangle, \langle A, \bar{u}\otimes\bar{v}\otimes u\otimes v\rangle),$$

from which the required equality follows upon taking max over unit vectors. $\square$

Proof of Theorem 8.4. We apply Lemma 8.5 to A_G and note that we may take our maximum over $\mathbb{R}_+^n$ since A_G is nonnegative valued.

$$\|A_G\|_{\sigma,\mathbb{C}} = \max_{0\neq x,y,u,v\in\mathbb{R}_+^n} \frac{\langle A_G, x\otimes y\otimes u\otimes v\rangle}{\|x\|\|y\|\|u\|\|v\|} = \max_{0\neq x,y\in\mathbb{R}_+^n} \frac{\langle A_G, x\otimes y\otimes x\otimes y\rangle}{\|x\|^2\|y\|^2}.$$

Since $2\langle A_{st}, x\otimes y\otimes x\otimes y\rangle = (x_s y_t + x_t y_s)^2$, we may use Cauchy–Schwarz to see that

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

If we do a change-of-variables $a_s = \sqrt{(x_s^2+y_s^2)/2}$ for $s = 1,\dots,n$, we obtain

$$\langle A_{st}, x\otimes y\otimes x\otimes y\rangle \le 2a_s^2 a_t^2 = \langle A_{st}, a\otimes a\otimes a\otimes a\rangle.$$

Upon summing over $\{s,t\} \in E$, we get

$$\langle A_G, x\otimes y\otimes x\otimes y\rangle \le \langle A_G, a\otimes a\otimes a\otimes a\rangle,$$

where the left-hand side follows from (8.4) and the right-hand side follows from (8.2) and (8.5). The last inequality gives us (8.7) easily. We then get (8.8) from (8.6) and (8.7). $\square$

In the following, we let $\mathbb{Q}_\mathbb{F}$ be the field of rational numbers $\mathbb{Q}$ if $\mathbb{F} = \mathbb{R}$ and the field of Gaussian rational numbers $\mathbb{Q}[i] := \{a + bi : a, b \in \mathbb{Q}\}$ if $\mathbb{F} = \mathbb{C}$. As is customary, we will restrict our problem inputs to $\mathbb{Q}_\mathbb{F}$ to ensure that they may be specified in finitely many bits. We refer the reader to [11, Definitions 2.1 and 4.1] for the formal definitions of the weak membership problem and the approximation problem.

Computing the clique number of a graph is an NP-hard problem [16] and so the identity (8.8) implies that the computing the spectral norm of A_G is NP-hard over both $\mathbb{R}$ and $\mathbb{C}$. Since the clique number is an integer, it is also NP-hard to approximate the spectral norm to arbitrary accuracy.

Theorem 8.6. *Let $\delta > 0$ be rational and $A \in \mathbb{Q}_\mathbb{F}^{n\times n\times n\times n}$ be bi-Hermitian, bipositive semidefinite, and nonnegative-valued. Computing an approximation $\omega(A) \in \mathbb{Q}$ such that*

$$\|A\|_{\sigma,\mathbb{F}} - \delta < \omega(A) < \|A\|_{\sigma,\mathbb{F}} + \delta$$

is an NP-hard problem for both $\mathbb{F} = \mathbb{R}$ and $\mathbb{C}$.

For any $\delta > 0$ and any convex set with nonempty interior $K \subseteq \mathbb{F}^n$, we define

$$S(K,\delta) := \bigcup\nolimits_{x\in K} B(x,\delta) \quad \text{and} \quad S(K,-\delta) := \{x \in K : B(x,\delta) \subseteq K\},$$

where $B(x,\delta)$ is the δ-ball centered at x with respect to the Hilbert–Schmidt norm in $\mathbb{F}^n$. Using [11, Theorem 4.2], we deduce the NP-hardness of the weak membership problem from Theorem 8.6.

Corollary 8.7. *Let K be the spectral norm unit ball in $\mathbb{F}^{n\times n\times n\times n}$ and $0 < \delta \in \mathbb{Q}$. Given $A \in \mathbb{Q}_\mathbb{F}^{n\times n\times n\times n}$ that is bi-Hermitian, bipositive semidefinite, and nonnegative-valued, deciding whether $A \in S(K,\delta)$ or $x \notin S(K,-\delta)$ is an NP-hard problem for both $\mathbb{F} = \mathbb{R}$ and $\mathbb{C}$.*

It then follows from [11, Theorem 3.1] and the duality of spectral and nuclear norms that Corollary 8.7 also holds true for the nuclear norm of 4-tensors.

Corollary 8.8. *Let K be the nuclear norm unit ball in $\mathbb{F}^{n\times n\times n\times n}$ and $0<\delta\in\mathbb{Q}$. Given $A\in\mathbb{Q}_{\mathbb{F}}^{n\times n\times n\times n}$ that is bi-Hermitian, bipositive semidefinite, and nonnegative-valued, deciding whether $A\in S(K,\delta)$ or $x\notin S(K,-\delta)$ is an NP-hard problem for both $\mathbb{F}=\mathbb{R}$ and $\mathbb{C}$.*

Using [11, Theorem 4.2] a second time, we may deduce the nuclear norm analogue of Theorem 8.6.

Corollary 8.9. *Let $\delta>0$ be rational and $A\in\mathbb{Q}_{\mathbb{F}}^{n\times n\times n\times n}$ be bi-Hermitian, bipositive semidefinite, and nonnegative-valued. Computing an approximation $\omega(A)\in\mathbb{Q}$ such that*

$$\|A\|_{\sigma,\mathbb{F}}-\delta<\omega(A)<\|A\|_{\sigma,\mathbb{F}}+\delta$$

is an NP-hard problem for both $\mathbb{F}=\mathbb{R}$ and $\mathbb{C}$.

As we did for Theorem 8.1, we may use Corollaries 8.7 and 8.8 along with Proposition 2.1 to deduce a complex analogue of Theorem 8.1.

Theorem 8.10. *The spectral and nuclear norms of d-tensors over $\mathbb{C}$ are NP-hard for any $d\geq 4$.*

9. Polynomial-time approximation bounds

Assuming that $P\neq NP$, then by Corollaries 8.7 and 8.8, one cannot approximate the spectral and nuclear norms of d-tensors to arbitrary accuracy in polynomial time. In this section, we will discuss some approximation bounds for spectral and nuclear norms that are computable in polynomial time.

The simplest polynomial-time computable bounds for the spectral and nuclear norms are those that come from the equivalence of norms in finite-dimensional spaces. The following lemma uses the Hilbert–Schmidt norm but any other Hölder p-norms [17],

$$\|A\|_{H,p}\coloneqq\Bigl(\sum\nolimits_{i_1,\dots,i_d=1}^{n_1,\dots,n_d}|a_{i_1\cdots i_d}|^p\Bigr)^{1/p},$$

where $p\in[1,\infty]$, which are all polynomial-time computable, may also serve the role.

Lemma 9.1. *Let $A\in\mathbb{F}^{n_1\times\cdots\times n_d}$. Then*

$$\frac{1}{\sqrt{n_1\cdots n_d}}\|A\|\leq\|A\|_\sigma\leq\|A\|\qquad\textit{and}\qquad\|A\|\leq\|A\|_*\leq\sqrt{n_1\cdots n_d}\|A\|.$$

Proof. We start with the bounds for the spectral norm. Clearly $\|A\|_\sigma\leq\|A\|$. Let $A=(a_{i_1\cdots i_d})$ and set $\|A\|_{H,\infty}=\max\{|a_{i_1\cdots i_d}| : i_k=1,\dots,n_k,\ k=1,\dots,d\}$. Clearly, $\|A\|\leq\sqrt{n_1\cdots n_d}\,\|A\|_{H,\infty}$. Note that $a_{i_1\cdots i_d}=\langle A,e_{i_1}\otimes\cdots\otimes e_{i_d}\rangle$ where e_{i_k} are standard basis vectors in $\mathbb{F}^{n_k}$. In particular $\|A\|_{H,\infty}=|\langle A,u_1\otimes\cdots\otimes u_d\rangle|$ for some unit vectors $u_1,\dots,u_d$ and thus $\|A\|_{H,\infty}\leq\|A\|_\sigma$ by (2.4). The corresponding inequalities for the nuclear norm follows from it being a dual norm. ☐

One downside of universal bounds like those in Lemma 9.1 is that they necessarily depend on the dimension of the ambient space. We will now construct tighter polynomial-time computable bounds for the spectral and nuclear norms of 3-tensors that depend only on the 'intrinsic dimension' of the specific tensor we are approximating. The *multilinear rank* [4] of a 3-tensor $A \in \mathbb{F}^{m\times n\times p}$ is the 3-tuple $\mu\operatorname{rank}(A) \coloneqq (r_1, r_2, r_3)$ where

$$\begin{aligned} r_1 &= \dim \operatorname{span}_{\mathbb{F}}\{A_{1\bullet\bullet}, \dots, A_{m\bullet\bullet}\},\\ r_2 &= \dim \operatorname{span}_{\mathbb{F}}\{A_{\bullet 1\bullet}, \dots, A_{\bullet n\bullet}\},\\ r_3 &= \dim \operatorname{span}_{\mathbb{F}}\{A_{\bullet\bullet 1}, \dots, A_{\bullet\bullet p}\}. \end{aligned}$$

Here $A_{i\bullet\bullet} = (a_{ijk})_{j,k=1}^{n,p} \in \mathbb{F}^{n\times p}$, $A_{\bullet j\bullet} = (a_{ijk})_{i,k=1}^{m,p} \in \mathbb{F}^{m\times p}$, $A_{\bullet\bullet k} = (a_{ijk})_{i,j=1}^{m,p} \in \mathbb{F}^{m\times n}$ are 'matrix slices' of the 3-tensor — the analogues of the row and column vectors of a matrix. This was due originally to Hitchcock [15], a special case (2-plex rank) of his *multiplex rank*.

We define the *flattening maps* along the 1st, 2nd, and 3rd index by

$$\flat_1 : \mathbb{F}^{m\times n\times p} \to \mathbb{F}^{m\times np}, \quad \flat_2 : \mathbb{F}^{m\times n\times p} \to \mathbb{F}^{n\times mp}, \quad \flat_3 : \mathbb{F}^{m\times n\times p} \to \mathbb{F}^{p\times mn},$$

respectively. Intuitively, these take a 3-tensor $A \in \mathbb{F}^{m\times n\times p}$ and 'flatten' it in three different ways to yield three matrices. Instead of giving precise but cumbersome formulae, it suffices to illustrate these simple maps with an example: Let

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

then

$$\flat_1(A) = \begin{bmatrix} a_{111} & a_{112} & a_{121} & a_{122} & a_{131} & a_{132}\\ a_{211} & a_{212} & a_{221} & a_{222} & a_{231} & a_{232}\\ a_{311} & a_{312} & a_{321} & a_{322} & a_{331} & a_{332}\\ a_{411} & a_{412} & a_{421} & a_{422} & a_{431} & a_{432} \end{bmatrix} \in \mathbb{F}^{4\times 6},$$

$$\flat_2(A) = \begin{bmatrix} a_{111} & a_{112} & a_{211} & a_{212} & a_{311} & a_{312} & a_{411} & a_{412}\\ a_{121} & a_{122} & a_{221} & a_{222} & a_{321} & a_{322} & a_{421} & a_{422}\\ a_{131} & a_{132} & a_{231} & a_{232} & a_{331} & a_{332} & a_{431} & a_{432} \end{bmatrix} \in \mathbb{F}^{3\times 8},$$

$$\flat_3(A) = \begin{bmatrix} a_{111} & a_{121} & a_{131} & a_{211} & a_{221} & a_{231} & a_{311} & a_{321} & a_{331} & a_{411} & a_{421} & a_{431}\\ a_{112} & a_{122} & a_{132} & a_{212} & a_{222} & a_{232} & a_{312} & a_{322} & a_{332} & a_{412} & a_{422} & a_{432} \end{bmatrix} \in \mathbb{F}^{2\times 12}.$$

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

$$r_1 = \operatorname{rank}(\flat_1(A)), \quad r_2 = \operatorname{rank}(\flat_2(A)), \quad r_3 = \operatorname{rank}(\flat_3(A)),$$

where rank here is the usual matrix rank of the matrices $\flat_1(A), \flat_2(A), \flat_3(A)$. Although we will have no use for it, a recently popular definition of tensor nuclear norm is as the arithmetic mean of the (matrix) nuclear norm of the flattenings:

$$\|A\|_\flat = \frac{1}{3}(\|\flat_1(A)\|_* + \|\flat_2(A)\|_* + \|\flat_3(A)\|_*).$$

We first provide alternative characterizations for the spectral and nuclear norms of a 3-tensor.

Lemma 9.2. *Let $A \in \mathbb{F}^{m\times n\times p}$. Then*

$$\|A\|_\sigma = \max\left\{\frac{|\langle A, x\otimes M\rangle|}{\|x\|\|M\|_*} : 0 \neq x \in \mathbb{F}^m,\ 0 \neq M \in \mathbb{F}^{n\times p}\right\}, \tag{9.1}$$

$$\begin{aligned} \|A\|_* = \min\Bigl\{\sum\nolimits_{i=1}^r \|x_i\|\|M_i\|_* : A = \sum\nolimits_{i=1}^r x_i \otimes M_i,& \\ x_i \in \mathbb{F}^m,\ M_i \in \mathbb{F}^{n\times p},\ r \in \mathbb{N}\Bigr\}.& \end{aligned} \tag{9.2}$$

Furthermore there is a decomposition of A that attains the minimum in (9.2) where $x_1 \otimes M_1, \dots, x_r \otimes M_r$ are linearly independent.

Proof. If we set $M_i = y_i \otimes z_i$, $i = 1, \dots, r$, then (9.2) reduces to (2.3). So the minimum in (2.3) is not more than the minimum in (9.2). On the other hand, we may write each M_i as a sum of rank-one matrices, in which case (9.2) reduces to (2.3). The existence of a decomposition that attains (9.2) follows from the same argument that we used in the proof of Proposition 3.1. The linear independence of $x_1 \otimes M_1, \dots, x_r \otimes M_r$ follows from Proposition 4.3. Equation (9.1) then follows from (9.2) by duality. □

Lemma 9.3. *Let $A \in \mathbb{F}^{m\times n\times p}$ with $\mu\,\mathrm{rank}(A) = (r_1, r_2, r_3)$. If the decomposition*

$$A = \sum\nolimits_{i=1}^{r} x_i \otimes M_i$$

attains (9.2), then for all $i = 1, \dots, r$,

$$\operatorname{rank} M_i \le \min(r_2, r_3). \tag{9.3}$$

Proof. Suppose $\mathbb{F} = \mathbb{R}$; the proof for $\mathbb{C}$ is similar except that we have unitary transformations in place of orthogonal ones. Using any one of the multilinear rank decompositions [17], we may reduce $A \in \mathbb{R}^{m\times n\times p}$ to a tensor $U \in \mathrm{O}(m)$, $V \in \mathrm{O}(n)$, $W \in \mathrm{O}(p)$ such that

$$A = (U, V, W) \cdot C,$$

where $C \in \mathbb{R}^{m\times n\times p}$ is such that $c_{ijk} = 0$ if $i > r_1$, $j > r_2$, or $k > r_3$. So we have

$$(U, V, W) \cdot C = \sum\nolimits_{\ell=1}^{r} x_\ell \otimes M_\ell,$$

and applying the multilinear transform $(U^\mathsf{T}, V^\mathsf{T}, W^\mathsf{T})$ to both sides, we get

$$C = \sum\nolimits_{\ell=1}^{r} (U^\mathsf{T} x_\ell) \otimes (V M_\ell W^\mathsf{T}).$$

Let $\ell = 1, \dots, r$. Let us partition $\widetilde{x}_\ell = U^\mathsf{T} x_\ell \in \mathbb{R}^m$ and $\widetilde{M}_\ell = V M_\ell W^\mathsf{T} \in \mathbb{R}^{n\times p}$ into

$$\widetilde{x}_\ell = \begin{bmatrix} y_\ell \\ z_\ell \end{bmatrix} \qquad \text{and} \qquad \widetilde{M}_\ell = \begin{bmatrix} J_\ell & K_\ell \\ L_\ell & N_\ell \end{bmatrix},$$

where $y_\ell \in \mathbb{R}^{r_1}$, $z_\ell \in \mathbb{R}^{m-r_1}$ and $J_\ell \in \mathbb{R}^{r_2\times r_3}$, $K_\ell \in \mathbb{R}^{r_2\times(p-r_3)}$, $L_\ell \in \mathbb{R}^{(n-r_2)\times r_3}$, $N_\ell \in \mathbb{R}^{(n-r_2)\times(p-r_3)}$. Now set

$$x'_\ell = \begin{bmatrix} y_\ell \\ 0 \end{bmatrix}, \qquad M'_\ell = \begin{bmatrix} J_\ell & 0 \\ 0 & 0 \end{bmatrix}.$$

As $c_{ijk} = 0$ if $i > r_1$, $j > r_2$, or $k > r_3$, it follows that

$$C = \sum\nolimits_{\ell=1}^{r} x'_\ell \otimes M'_\ell.$$

Since orthogonal matrices preserve Hilbert–Schimdt and nuclear norms, $\|x_i\| = \|\widetilde{x}_i\| \ge \|x'_i\|$ and $\|M_\ell\|_* = \|\widetilde{M}_\ell\|_* \ge \|M'_\ell\|_*$ and so

$$\sum\nolimits_{\ell=1}^{r} \|x_\ell\| \|M_\ell\|_* \ge \sum\nolimits_{\ell=1}^{r} \|x'_\ell\| \|M'_\ell\|_*.$$

Clearly $\operatorname{rank} M'_\ell \le \min(r_2, r_3)$. □

By Definition 2.3,

$$\|\flat_1(A)\|_* = \min\Big\{\sum\nolimits_{i=1}^r \|x_i\|\|M_i\| : A = \sum\nolimits_{i=1}^r x_i \otimes M_i,\ x_i \in \mathbb{F}^m,\ M_i \in \mathbb{F}^{n\times p}\Big\},$$

and since any matrix satisfies

$$\|M_i\| \le \|M_i\|_* \le \sqrt{\operatorname{rank} M_i}\|M_i\|_\sigma,$$

using (9.2) and (9.3), we obtain

$$\|\flat_1(A)\|_* \le \|A\|_* \le \sqrt{\min(r_2(A), r_3(A))}\|\flat_1(A)\|_*. \tag{9.4}$$

From (9.4), we deduce the corresponding bounds for its dual norm,

$$\|\flat_1(A)\|_\sigma \ge \|A\|_\sigma \ge \frac{1}{\sqrt{\min(r_2(A), r_3(A))}}\|\flat_1(A)\|_\sigma.$$

Moreover, we may deduce analogous inequalities in terms of $\flat_2(A)$ and $\flat_3(A)$. We assemble these to get the bounds in the following theorem.

Theorem 9.4. *Let $A \in \mathbb{F}^{m\times n\times p}$ with $\mu\operatorname{rank}(A) = (r_1, r_2, r_3)$. Then*

$$\max\left\{\frac{\|\flat_1(A)\|_\sigma}{\sqrt{\min(r_2, r_3)}}, \frac{\|\flat_2(A)\|_\sigma}{\sqrt{\min(r_1, r_3)}}, \frac{\|\flat_3(A)\|_\sigma}{\sqrt{\min(r_1, r_2)}}\right\} \le \|A\|_\sigma$$
$$\le \min\{\|\flat_1(A)\|_\sigma, \|\flat_2(A)\|_\sigma, \|\flat_3(A)\|_\sigma\}$$

and

$$\max\{\|\flat_1(A)\|_*, \|\flat_2(A)\|_*, \|\flat_3(A)\|_*\} \le \|A\|_*$$
$$\le \min\Big\{\sqrt{\min(r_2, r_3)}\|\flat_1(A)\|_*, \sqrt{\min(r_1, r_3)}\|\flat_2(A)\|_*, \sqrt{\min(r_1, r_2)}\|\flat_3(A)\|_*\Big\}.$$

Note that both upper and lower bounds are computable in polynomial time. Clearly, we may extend Theorem 9.4 to any $d > 3$ simply by flattening along d indices.

Acknowledgment

The authors thank Harm Derksen, Shenglong Hu, Jiawang Nie, and the two anonymous referees for their enormously helpful comments. The authors also thank Li Wang for help with numerical experiments that suggested that $\|C\|_{*,\mathbb{C}} = \sqrt{2}$ in Lemma 6.1.

References

[1] S. Banach, *Uber homogene Polynome in* (L^2), Studia Math. **7** (1938), 36–44.

[2] J.-L. Brylinski, *Algebraic measures of entanglement*, Mathematics of quantum computation, Comput. Math. Ser., Chapman & Hall/CRC, Boca Raton, FL, 2002, pp. 3–23. MR2007941

[3] P. Comon, G. Golub, L.-H. Lim, and B. Mourrain, *Symmetric tensors and symmetric tensor rank*, SIAM J. Matrix Anal. Appl. **30** (2008), no. 3, 1254–1279, DOI 10.1137/060661569. MR2447451

[4] 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), no. 3, 1084–1127, DOI 10.1137/06066518X. MR2447444

[5] A. Defant and K. Floret, *Tensor norms and operator ideals*, North-Holland Mathematics Studies, vol. 176, North-Holland Publishing Co., Amsterdam, 1993. MR1209438

[6] H. Derksen, *On the nuclear norm and the singular value decomposition of tensors*, Found. Comput. Math. **16** (2016), no. 3, 779–811, DOI 10.1007/s10208-015-9264-x. MR3494510

[7] H. Derksen, S. Friedland, L.-H. Lim, and L. Wang, *Theoretical and computational aspects of entanglement*, preprint (2017), arXiv:1707.02569.

[8] S. Friedland, *Best rank one approximation of real symmetric tensors can be chosen symmetric*, Front. Math. China **8** (2013), no. 1, 19–40, DOI 10.1007/s11464-012-0262-x. MR3010470
[9] S. Friedland, *Matrices—algebra, analysis and applications*, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2016. MR3467205
[10] S. Friedland, *Variation of tensor powers and spectra*, Linear and Multilinear Algebra **12** (1982/83), no. 2, 81–98, DOI 10.1080/03081088208817475. MR670715
[11] S. Friedland and L.-H. Lim, *The computational complexity of duality*, SIAM J. Optim. **26** (2016), no. 4, 2378–2393, DOI 10.1137/16M105887X. MR3566920
[12] A. Grothendieck, *Produits tensoriels topologiques et espaces nucléaires*, Mem. Amer. Math. Soc. **No. 16** (1955), 140. MR0075539
[13] J. M. Hendrickx and A. Olshevsky, *Matrix p-norms are NP-hard to approximate if $p \neq 1, 2, \infty$*, SIAM J. Matrix Anal. Appl. **31** (2010), no. 5, 2802–2812, DOI 10.1137/09076773X. MR2740634
[14] C. J. Hillar and L.-H. Lim, *Most tensor problems are NP-hard*, J. ACM **60** (2013), no. 6, Art. 45, 39, DOI 10.1145/2512329. MR3144915
[15] F. L. Hitchcock, *The expression of a tensor or a polyadic as a sum of products*, J. Math. Phys. **6** (1927), no. 1, 164–189.
[16] R. M. Karp, *Reducibility among combinatorial problems*, Complexity of computer computations (Proc. Sympos., IBM Thomas J. Watson Res. Center, Yorktown Heights, NY, 1972), Plenum, New York, 1972, pp. 85–103. MR0378476
[17] L.-H. Lim, *Tensors and hypermatrices*, Handbook of Linear Algebra, 2nd Ed., CRC Press, Boca Raton, FL, 2013.
[18] L.-H. Lim and P. Comon, *Multiarray signal processing: Tensor decomposition meets compressed sensing*, C. R. Acad. Sci. Paris, Series IIB – Mechanics **338** (2010), no. 6, 311–320.
[19] L.-H. Lim and P. Comon, *Blind multilinear identification*, IEEE Trans. Inform. Theory **60** (2014), no. 2, 1260–1280, DOI 10.1109/TIT.2013.2291876. MR3164974
[20] A. Pappas, Y. Sarantopoulos, and A. Tonge, *Norm attaining polynomials*, Bull. Lond. Math. Soc. **39** (2007), no. 2, 255–264, DOI 10.1112/blms/bdl033. MR2323457
[21] Y. Qi, P. Comon, and L.-H. Lim, *Semialgebraic geometry of nonnegative tensor rank*, SIAM J. Matrix Anal. Appl. **37** (2016), no. 4, 1556–1580, DOI 10.1137/16M1063708. MR3565551
[22] T. S. Motzkin and E. G. Straus, *Maxima for graphs and a new proof of a theorem of Turán*, Canad. J. Math. **17** (1965), 533–540. MR0175813
[23] R. A. Ryan, *Introduction to Tensor Products of Banach Spaces*, Springer Monographs in Mathematics, Springer-Verlag London, Ltd., London, 2002. MR1888309
[24] R. Schatten, *A Theory of Cross-Spaces*, Annals of Mathematics Studies, no. 26, Princeton University Press, Princeton, NJ, 1950. MR0036935
[25] D. Steinberg, *Computation of Matrix Norms with Applications to Robust Optimization*, M.Sc. thesis, Technion Israel Institute of Technology, Haifa, Israel, 2005.
[26] A. E. Taylor, *The norm of a real linear transformation in Minkowski space*, Enseignement Math. (2) **4** (1958), 101–107. MR0096687
[27] Y. C. Wong, *Schwartz Spaces, Nuclear Spaces and Tensor Products*, Lecture Notes in Mathematics, vol. 726, Springer, Berlin, 1979. MR541034
[28] K. Ye and L.-H. Lim, *Fast structured matrix computations: Tensor rank and Cohn–Umans method*, Found. Comput. Math. (2016), DOI 10.1007/s1020, to appear.
[29] F. L. Zak, *Determinants of projective varieties and their degrees*, Algebraic Transformation Groups and Algebraic Varieties, Encyclopaedia Math. Sci., vol. 132, Springer, Berlin, 2004, pp. 207–238, DOI 10.1007/978-3-662-05652-3_11. MR2090676
[30] T. Zhang and G. H. Golub, *Rank-one approximation to high order tensors*, SIAM J. Matrix Anal. Appl. **23** (2001), no. 2, 534–550, DOI 10.1137/S0895479899352045. MR1871328

Department of Mathematics, Statistics and Computer Science, University of Illinois at Chicago, 851 S. Morgan Street, Chicago, Illinois 60607
E-mail address: friedlan@uic.edu

Computational and Applied Mathematics Initiative, Department of Statistics, University of Chicago, 5747 S. Ellis Avenue, Chicago, Illinois 60637
E-mail address: lekheng@galton.uchicago.edu