



HAL
open science

Von Neumann's trace inequality for tensors

Stephane Chretien, Tianwen Wei

► **To cite this version:**

Stephane Chretien, Tianwen Wei. Von Neumann's trace inequality for tensors. *Linear Algebra and its Applications*, 2015, 482, pp.149 - 157. 10.1016/j.laa.2015.05.015 . hal-02515996

HAL Id: hal-02515996

<https://hal.science/hal-02515996>

Submitted on 25 Mar 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

VON NEUMANN'S INEQUALITY FOR TENSORS

STÉPHANE CHRÉTIEN AND TIANWEN WEI

ABSTRACT. For two matrices in $\mathbb{R}^{n_1 \times n_2}$, the von Neumann inequality says that their scalar product is less than or equal to the scalar product of their singular spectrum. In this short note, we extend this result to real tensors and provide a complete study of the equality case.

1. INTRODUCTION

The goal of this paper is to generalize von Neumann's inequality from matrices to tensors. Consider two matrices X and Y in $\mathbb{R}^{n_1 \times n_2}$. Denote their singular spectrum, i.e. the vector of their singular values, by $\sigma(X)$ (resp. $\sigma(Y)$). The classical matrix von Neumann's inequality [6] says that

$$\langle X, Y \rangle \leq \langle \sigma(X), \sigma(Y) \rangle,$$

and equality is achieved if and only if X and Y have the same singular subspaces. Von Neumann's inequality, and the characterization of the equality case in this inequality, are important in many aspects of mathematics.

For tensors, the task of generalizing Von Neumann's inequality is rendered harder because of the necessity to appropriately define the singular values and the Singular Value Decomposition (SVD). In this paper, we will use the SVD defined in [1], which is based on the Tucker decomposition.

Our main result is given in Theorem 3.1 below and gives a characterization of the equality case. We expect this result to be useful for the description of the subdifferential of some tensor functions as the matrix counterpart has proved for matrix functions [4]. Such functions occur naturally in computational statistics, machine learning and numerical analysis [2, 3] due to the recent interest of sparsity promoting norms as a convex surrogate to rank penalization.

2. MAIN FACTS ABOUT TENSORS

Let D and n_1, \dots, n_D be positive integers. Let $\mathcal{X} \in \mathbb{R}^{n_1 \times \dots \times n_D}$ denote a D -dimensional array of real numbers. We will also denote such arrays as tensors.

2.1. Basic notations and operations. A subtensor of \mathcal{X} is a tensor obtained by fixing some of its coordinates. As an example, fixing one coordinate $i_d = k$ in \mathcal{X} for some $k \in \{1, \dots, n_d\}$ yields a tensor in $\mathbb{R}^{n_1 \times \dots \times n_{d-1} \times n_{d+1} \times \dots \times n_D}$. In the sequel, we will denote this subtensor of \mathcal{X} by $\mathcal{X}_{i_d=k}$.

The fibers of a tensor are subtensors that have only one mode, i.e. obtained by fixing every coordinate except one. The mode- d fibers are the vectors

$$\left(\mathcal{X}_{i_1, \dots, i_{d-1}, i_d, i_{d+1}, \dots, i_D} \right)_{i_d=1, \dots, n_d}.$$

They extend the notion of columns and rows from the matrix to the tensor framework. For a matrix, the mode-1 fibers are the columns and the mode-2 fibers are the rows.

The mode- d matricization $\mathcal{X}_{(d)}$ of \mathcal{X} is obtained by forming the matrix whose columns are the mode- d fibers of the tensor, arranged in the lexicographic ordering [7]. Clearly, the k th column of $\mathcal{X}_{(d)}$ consists of the entries of $\mathcal{X}_{i_d=k}$.

The mode- d multiplication of a tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times \dots \times n_D}$ by a matrix $U \in \mathbb{R}^{n'_d \times n_d}$, denoted by $\mathcal{X} \times_d U$, gives a tensor in $\mathbb{R}^{n_1 \times \dots \times n'_d \times \dots \times n_D}$. It is defined as

$$(\mathcal{X} \times_d U)_{i_1, \dots, i_{d-1}, i'_d, i_{d+1}, \dots, i_D} = \sum_{i_d=1}^{n_d} \mathcal{X}_{i_1, \dots, i_{d-1}, i_d, i_{d+1}, \dots, i_D} U_{i'_d, i_d}.$$

2.2. Higher Order Singular Value Decomposition (HOSVD). The Tucker decomposition of a tensor is a very useful decomposition, which can be chosen so that with appropriate orthogonal transformations, one can reveal a tensor \mathcal{S} hidden inside \mathcal{X} with interesting rank and orthogonality properties. More precisely, we have

$$(2.1) \quad \mathcal{X} = \mathcal{S}(\mathcal{X}) \times_1 U^{(1)} \times_2 U^{(2)} \dots \times_D U^{(D)},$$

where each $U^{(d)} \in \mathbb{R}^{n_d \times n_d}$ is orthogonal and $\mathcal{S}(\mathcal{X})$ is a tensor of the same size as \mathcal{X} defined as follows. Moreover, subtensors $\mathcal{S}(\mathcal{X})_{i_d=k}$ for $k = 1, \dots, n_d$ are all orthogonal to each other for each $d = 1, \dots, D$.

2.2.1. Relationship with matricization. A tensor can be matricized along each of its modes. Let \otimes denote the standard Kronecker product for matrices. Then, the mode- d matricization of a tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times \dots \times n_D}$ is given by

$$(2.2)_{(d)} = U^{(d)} \cdot \mathcal{S}(\mathcal{X})_{(d)} \cdot \left(U^{(d+1)} \otimes \dots \otimes U^{(D)} \otimes U^{(1)} \otimes \dots \otimes U^{(d-1)} \right)^t.$$

Take the (usual) SVD of the matrix $\mathcal{X}_{(d)}$

$$\mathcal{X}_{(d)} = U^{(d)} \Sigma^{(d)} V^{(d)t}$$

and based on (2.2), we can set

$$\mathcal{S}(\mathcal{X})_{(d)} = \Sigma^{(d)} V^{(d)t} \left(U^{(d+1)} \otimes \dots \otimes U^{(D)} \otimes U^{(1)} \otimes \dots \otimes U^{(d-1)} \right),$$

where $\mathcal{S}(\mathcal{X})_{(d)}$ is the mode- d matricization of $\mathcal{S}(\mathcal{X})$. One proceeds similarly for all $d = 1, \dots, D$ and one recovers the orthogonal matrices $U^{(1)}, \dots, U^{(D)}$ which allow us to decompose \mathcal{X} as in (2.1).

2.2.2. The spectrum. The mode- d spectrum is defined as the vector of singular values of $\mathcal{X}_{(d)}$ and we will denote it by $\sigma^{(d)}(\mathcal{X})$. Notice that this construction implies that $\mathcal{S}(\mathcal{X})$ has orthonormal fibers for every modes. With a slight abuse of notation, we will denote by σ the mapping which to each tensor \mathcal{X} assigns the vector $1/\sqrt{D} (\sigma^{(1)}, \dots, \sigma^{(D)})$ of all mode- d singular spectra.

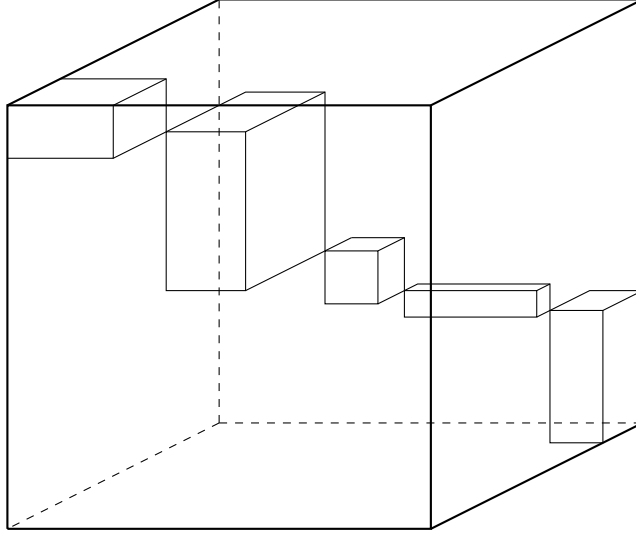


FIGURE 1. A block-wise diagonal tensor.

3. MAIN RESULT

3.1. The main theorem. The main result of this paper is the following theorem.

Theorem 3.1. *Let $\mathcal{X}, \mathcal{Y} \in \mathbb{R}^{n_1 \times \dots \times n_D}$ be tensors. Then for all $d = 1, \dots, D$, we have*

$$(3.3) \quad \langle \mathcal{X}, \mathcal{Y} \rangle \leq \langle \sigma^{(d)}(\mathcal{X}), \sigma^{(d)}(\mathcal{Y}) \rangle.$$

The equality in (3.3) holds simultaneously for all $d = 1, \dots, D$ if and only if there exist orthogonal matrices $W^{(d)} \in \mathbb{R}^{n_d \times n_d}$ for $d = 1, \dots, D$ and tensors $\mathcal{D}(\mathcal{X}), \mathcal{D}(\mathcal{Y}) \in \mathbb{R}^{n_1 \times \dots \times n_D}$ such that

$$\begin{aligned} \mathcal{X} &= \mathcal{D}(\mathcal{X}) \times_1 W^{(1)} \dots \times_D W^{(D)}, \\ \mathcal{Y} &= \mathcal{D}(\mathcal{Y}) \times_1 W^{(1)} \dots \times_D W^{(D)}, \end{aligned}$$

where $\mathcal{D}(\mathcal{X})$ and $\mathcal{D}(\mathcal{Y})$ satisfy the following properties:

- $\mathcal{D}(\mathcal{X})$ and $\mathcal{D}(\mathcal{Y})$ are block-wise diagonal with the same number and size of blocks.
- Let L be the number of blocks and $\{\mathcal{D}_l(\mathcal{X})\}_{l=1, \dots, L}$ (resp. $\{\mathcal{D}_l(\mathcal{Y})\}_{l=1, \dots, L}$) be the blocks on the diagonal of $\mathcal{D}(\mathcal{X})$ (resp. $\mathcal{D}(\mathcal{Y})$). Then for each $l = 1, \dots, L$, the two blocks $\mathcal{D}_l(\mathcal{X})$ and $\mathcal{D}_l(\mathcal{Y})$ are proportional.

3.2. Proof of the main theorem. In this section, we prove Theorem 3.1. If \mathcal{X} or \mathcal{Y} is a zero tensor, then the result is trivial. In the sequel, we assume that both \mathcal{X} and \mathcal{Y} are non-zero tensors.

3.2.1. The "if" part. The "if" part of the result is straightforward. Notice that $\langle \mathcal{X}, \mathcal{Y} \rangle = \langle \mathcal{D}(\mathcal{X}), \mathcal{D}(\mathcal{Y}) \rangle$ and the singular vectors of \mathcal{X} (resp. \mathcal{Y}) are equal to those of $\mathcal{D}(\mathcal{X})$ (resp. $\mathcal{D}(\mathcal{Y})$). Therefore, it remains to prove that

$$(3.4) \quad \langle \mathcal{D}(\mathcal{X}), \mathcal{D}(\mathcal{Y}) \rangle = \langle \sigma^{(d)}(\mathcal{D}(\mathcal{X})), \sigma^{(d)}(\mathcal{D}(\mathcal{Y})) \rangle, \quad d = 1, \dots, D.$$

The conditions that $\mathcal{D}(\mathcal{X})$ and $\mathcal{D}(\mathcal{Y})$ are block-wise diagonal and that $\mathcal{D}_i(\mathcal{X})$ and $\mathcal{D}_i(\mathcal{Y})$ are proportional implies that each row of $\mathcal{D}_{(d)}(\mathcal{X})$ and that of $\mathcal{D}_{(d)}(\mathcal{Y})$ are parallel. It then follows that $\mathcal{D}_{(d)}(\mathcal{X})$ and $\mathcal{D}_{(d)}(\mathcal{Y})$ have the same left and right singular vectors. Then, applying the matrix von Neumann's result immediately gives (3.4).

3.2.2. *The "only if" part: first step.* Assume that

$$\langle \mathcal{X}, \mathcal{Y} \rangle = \langle \sigma^{(d)}(\mathcal{X}), \sigma^{(d)}(\mathcal{Y}) \rangle, \quad d = 1, \dots, D.$$

By the classical results of matrix von Neumann's inequality, we know that the equality holds if and only if there exist orthogonal matrices $U^{(d)}$ and $V^{(d)}$ such that

$$(3.5) \quad \begin{aligned} \mathcal{X}_{(d)} &= U^{(d)} \text{Diag} \left(\sigma^{(d)}(\mathcal{X}) \right) V^{(d)t} \quad \text{and} \\ \mathcal{Y}_{(d)} &= U^{(d)} \text{Diag} \left(\sigma^{(d)}(\mathcal{Y}) \right) V^{(d)t} \end{aligned}$$

for all $d = 1, \dots, D$. From this remark, we obtain the following HOSVD of \mathcal{X} and \mathcal{Y} :

$$\begin{aligned} \mathcal{X} &= \mathcal{S}(\mathcal{X}) \times_1 U^{(1)} \dots \times_D U^{(D)}, \\ \mathcal{Y} &= \mathcal{S}(\mathcal{Y}) \times_1 U^{(1)} \dots \times_D U^{(D)}. \end{aligned}$$

3.2.3. *The "only if" part: second step.* We now show that subtensors $\mathcal{S}(\mathcal{X})_{i_d=k}$ and $\mathcal{S}(\mathcal{Y})_{i_d=k}$ must be parallel for all $k = 1, \dots, n_d$ and $d = 1, \dots, D$.

Comparing (3.5) with (2.2), we deduce that

$$(3.6) \quad \mathcal{S}_{(d)}(\mathcal{X}) = \begin{pmatrix} \sigma_1^{(d)}(\mathcal{X}) \cdot p_1^t \\ \vdots \\ \sigma_{n_d}^{(d)}(\mathcal{X}) \cdot p_{n_d}^t \end{pmatrix},$$

where p_i^t denotes the i th row of matrix $V^{(d)t} (U^{(d+1)} \otimes \dots \otimes U^{(D)} \otimes U^{(1)} \otimes \dots \otimes U^{(d-1)})$. Similarly, we have

$$(3.7) \quad \mathcal{S}_{(d)}(\mathcal{Y}) = \begin{pmatrix} \sigma_1^{(d)}(\mathcal{Y}) \cdot p_1^t \\ \vdots \\ \sigma_{n_d}^{(d)}(\mathcal{Y}) \cdot p_{n_d}^t \end{pmatrix}.$$

Comparing now (3.6) and (3.7) reveals that the i th row of $\mathcal{S}_{(d)}(\mathcal{X})$ and the i th row of $\mathcal{S}_{(d)}(\mathcal{Y})$ must be proportional, for all $i = 1, \dots, n_d$. Formally, this means

$$(3.8) \quad \sigma_{i_d}^{(d)}(\mathcal{Y}) \cdot \mathcal{S}(\mathcal{X})_{i_1 \dots i_d \dots i_D} = \sigma_{i_d}^{(d)}(\mathcal{X}) \cdot \mathcal{S}(\mathcal{Y})_{i_1 \dots i_d \dots i_D}.$$

for all possible values of i_1, \dots, i_D .

3.2.4. *The "only if" part: third step.* For $d = 1, \dots, D$, let $r_x^{(d)}$ (resp. $r_y^{(d)}$) be the rank of $\mathcal{S}_{(d)}(\mathcal{X})$ (resp. $\mathcal{S}_{(d)}(\mathcal{Y})$). Let (i_1, \dots, i_D) be such that

$$\begin{aligned} (i_1, \dots, i_D) &\not\leq (r_y^{(1)}, \dots, r_y^{(D)}) \\ (i_1, \dots, i_D) &\not\geq (r_y^{(1)}, \dots, r_y^{(D)}). \end{aligned}$$

Then, $\mathcal{S}(\mathcal{Y})_{i_1, \dots, i_D} = 0$ but there exists $d \in \{1, \dots, D\}$ such that $\sigma_{i_d}^{(d)}(\mathcal{Y}) > 0$. Using (3.8), we obtain that $\mathcal{S}(\mathcal{X})_{i_1, \dots, i_D} = 0$. Thus, if $r_x^{(d)} > r_y^{(d)}$ for some $d = 1, \dots, D$, then there exists some

$$(i_1, \dots, i_D) > (r_y^{(1)}, \dots, r_y^{(D)})$$

such that $\mathcal{S}(\mathcal{X})_{i_1, \dots, i_D} \neq 0$ and thus, $r_x^{(d)} > r_y^{(d)}$ for all $d = 1, \dots, D$. By symmetry, we deduce that either $r_x^{(d)} > r_y^{(d)}$ for all $d = 1, \dots, D$, or $r_x^{(d)} < r_y^{(d)}$ for all $d = 1, \dots, D$ or else $r_x^{(d)} = r_y^{(d)}$ for all $d = 1, \dots, D$.

3.2.5. *The "only if" part: fourth step.* Assume that $(r_x^{(1)}, \dots, r_x^{(D)}) \leq (r_y^{(1)}, \dots, r_y^{(D)})$. The other case may be treated in the same way (with an overlap in the equality case) by interchanging the role of \mathcal{X} and \mathcal{Y} . For all $(i_1, \dots, i_D) \leq (r_y^{(1)}, \dots, r_y^{(D)})$, we have $\sigma_{i_d}^{(d)}(\mathcal{Y}) > 0$ for all $d = 1, \dots, D$. Thus, (3.8) gives

$$\mathcal{S}(\mathcal{X})_{i_1 \dots i_d \dots i_D} = \frac{\sigma_{i_d}^{(d)}(\mathcal{X})}{\sigma_{i_d}^{(d)}(\mathcal{Y})} \cdot \mathcal{S}(\mathcal{Y})_{i_1 \dots i_d \dots i_D}.$$

We deduce from this equation that for two indices (i_1, \dots, i_D) and (i'_1, \dots, i'_D) , if

- (i) there exists some d in $\{1, \dots, D\}$ such that $i_d = i'_d$,
- (ii) $\mathcal{S}(\mathcal{X})_{i_1, \dots, i_d, \dots, i_D}$ and $\mathcal{S}(\mathcal{X})_{i'_1, \dots, i'_d, \dots, i'_D}$ are different from zero,

then

$$(\mathcal{S}(\mathcal{X})_{i_1 \dots i_d \dots i_D}, \mathcal{S}(\mathcal{X})_{i'_1, \dots, i'_d, \dots, i'_D}) = \rho \cdot (\mathcal{S}(\mathcal{Y})_{i_1 \dots i_d \dots i_D}, \mathcal{S}(\mathcal{Y})_{i'_1, \dots, i'_d, \dots, i'_D}),$$

where

$$\rho = \frac{\sigma_{i_1}^{(1)}(\mathcal{X})}{\sigma_{i_1}^{(1)}(\mathcal{Y})} = \dots = \frac{\sigma_{i_D}^{(D)}(\mathcal{X})}{\sigma_{i_D}^{(D)}(\mathcal{Y})} > 0.$$

3.2.6. *The "only if" part: fifth step.* Let $\rho_1 > \dots > \rho_L$ denote the possible values of the ratio $\sigma_{i_d}^{(d)}(\mathcal{X})/\sigma_{i_d}^{(d)}(\mathcal{Y})$, for all $(i_1, \dots, i_D) \leq (r_y^{(1)}, \dots, r_y^{(D)})$. Let $I_{d,l}$, $d = 1, \dots, D$, $l = 1, \dots, L$ denote the possibly empty set of indices in $\{1, \dots, r_x^{(d)}\}$ such that

$$\frac{\sigma_{i_d}^{(d)}(\mathcal{X})}{\sigma_{i_d}^{(d)}(\mathcal{Y})} = \rho_l$$

and let $m_{d,l}$ denote the cardinality of $I_{d,l}$. Then, for each $d = 1, \dots, D$, we can find a permutation π_d on $\{1, \dots, n_d\}$ such that $\pi_d(I_{d,1}) = \{1, \dots, m_{d,1}\}$, $\pi_d(I_{d,2}) = \{m_{d,1} + 1, \dots, m_{d,1} + m_{d,2}\}$, and so on and so forth.

Thus, for each mode $d = 1, \dots, D$, there exists a permutation matrix Π_d such that the matrices

$$\mathcal{D}(\mathcal{X}) = \mathcal{S}(\mathcal{X}) \times_1 \Pi_1 \cdots \times_D \Pi_D,$$

and

$$\mathcal{D}(\mathcal{Y}) = \mathcal{S}(\mathcal{Y}) \times_1 \Pi_1 \cdots \times_D \Pi_D.$$

contain L blocks and each block in $\mathcal{D}(\mathcal{X})$ is proportionnal to the corresponding block in $\mathcal{D}(\mathcal{Y})$. Moreover, any entry $\mathcal{D}(\mathcal{X})_{i_1, \dots, i_D}$ with $(i_1, \dots, i_D) \leq (r_x^{(1)}, \dots, r_x^{(D)})$ and lying outside the union of these L blocks is null since if it were not, by (3.8) combined with $(r_x^{(1)}, \dots, r_x^{(D)}) \leq (r_y^{(1)}, \dots, r_y^{(D)})$, it would be proportionnal to a nonzero component $\mathcal{D}(\mathcal{Y})_{i_1, \dots, i_D}$ with two different ratios, thus a contradiction.

Finally, setting $W^{(d)} = \Pi_d U^{(d)}$ for $d = 1, \dots, D$ achieves the proof.

REFERENCES

1. De Lathauwer, L., De Moor, B. and Vandewalle, J. (2000). A multilinear singular value decomposition. *SIAM journal on Matrix Analysis and Applications*, 21(4), 1253-1278.
2. Gandy, S., Recht, B. and Yamada, I. (2011). Tensor completion and low- n -rank tensor recovery via convex optimization. *Inverse Problems*, 27(2).
3. Kressner, D., Steinlechner, M. and Vandereycken, B. (2013). Low-rank tensor completion by Riemannian optimization. *BIT Numerical Mathematics*, 1-22.
4. Lewis, A. S., The convex analysis of unitarily invariant matrix functions. *J. Convex Anal.* 2 (1995), no. 1-2, 173–183.
5. B. Romera-Paredes and M. Pontil, A New Convex Relaxation for Tensor Completion. *Neural Information Processing Systems, NIPS 2013*. Lake Tahoe, USA.
6. Von Neumann, J. (1937). Some matrix inequalities and metrization of matrix space. *Tomsk Univ. Rev.* 1(11), 286-300.
7. Kolda, T. and Bader, B. (2009), *Tensor Decompositions and Applications*, *SIAM Review*, 455-500.

LABORATOIRE DE MATHÉMATIQUES, UMR 6623, UNIVERSITÉ DE FRANCHE-COMTÉ,
16 ROUTE DE GRAY, 25030 BESANCON, FRANCE
E-mail address: `stephane.chretien@univ-fcomte.fr`

LABORATOIRE DE MATHÉMATIQUES, UMR 6623, UNIVERSITÉ DE FRANCHE-COMTÉ,
16 ROUTE DE GRAY, 25030 BESANCON, FRANCE
E-mail address: `tianwen.wei@univ-fcomte.fr`