On the subdifferential of symmetric convex functions of the spectrum for symmetric and orthogonally decomposable tensors

Stephane Chretien, Tianwen Wei

On the subdifferential of symmetric convex functions of the spectrum for symmetric and orthogonally decomposable tensors

Stéphane Chrétien[a], Tianwen Wei[b,*]

[a] *National Physical Laboratory, Mathematics and Modelling, Hampton Road, Teddington, TW11 OLW, UK*
[b] *Department of Mathematics and Statistics, Zhongnan University of Economics and Law, Nanhu Avenue 182, 430073, Wuhan, China*

Abstract

The subdifferential of convex functions of the singular spectrum of real matrices has been widely studied in matrix analysis, optimization and automatic control theory. Convex optimization over spaces of tensors is now gaining much interest due to its potential applications in signal processing, statistics and engineering. The goal of this paper is to present an extension of the approach by Lewis [16] for the analysis of the subdifferential of certain convex functions of the spectrum of symmetric tensors. We give a complete characterization of the subdifferential of Schatten-type tensor norms for symmetric tensors. Some partial results in this direction are also given for Orthogonally Decomposable tensors.

Keywords: Tensor, Schatten norm, spectrum, subdifferential

1. Introduction

1.1. Background

Multidimensional arrays, also known as tensors are higher-order generalizations of vectors and matrices. In recent years, they have been the subject of extensive interest in various extremely active fields such as e.g. statistics, signal processing, automatic control, etc ... where a lot of problems involve quantities that are intrinsically multidimensional such as higher order moment tensors [2]. Many natural and useful quantities in linear algebra such as the rank or the Singular Value Decomposition turn out to be very difficult to compute or generalize in the tensor setting [12, 13, 9]. Fortunately, efficient approaches exist in the case of symmetric tensors which lie at the heart of the moment approach

*Corresponding author.
Email addresses: stephane.chretien@npl.co.uk (Stéphane Chrétien), tianwen.wei.2014@ieee.org (Tianwen Wei)

which recently proved very efficient for addressing essential problems in Statistics/Machine Learning such as Clustering, estimation in Hidden Markov Chains, etc ... See the very influencial paper [2] for more details. In many statistical models such as the ones presented in [2], the rank of the involved is low and one expects that the theory of sparse recovery can be applied to recover them from just a few observations just as in the case of Matrix Completion [4], [5] Robust PCA [3] and Matrix Compressed Sensing [18]. In such approaches to Machine Learning, one usually have to solve a penalized least squares problem of the type

$$\min_{X \in \mathbb{R}^{n_1 \times n_2}} \quad \|y - \mathcal{A}(X)\| + \lambda \; p(X),$$

where the penalization p is rank-sparsity promoting such as the nuclear norm and $\mathcal{A}$ is a linear operator taking values in $\mathbb{R}^n$. In the tensor setting, we look for solutions of problems of the type

$$\min_{X \in \mathbb{R}^{n_1 \times \cdots \times n_D}} \quad \|y - \mathcal{A}(X)\| + \lambda \; p(X),$$

for $D > 2$ and p is a generalization of the nuclear norm or some Schatten-type norm for tensors. The extention of Schatten norms to the tensor setting has to be carefully defined. In particular, several nuclear norms can be naturally defined [21], [8], [17]. Moreover, the study of the efficiency of sparsity promoting penalization relies crucially on the knowledge of the subdifferential of the norm involved as achieved in [1] or [15], or at least a good approximation of this subdifferential [21] [17]. In the matrix setting, the works of [20, 16] are famous for providing a neat characterization of the subdifferential of matrix norms or more generaly functions of the matrix enjoying enough symmetries. In the 3D or higher dimensional setting, however, the case is much less understood. The relationship between the tensor norms and the norms of the flattenings are intricate although some good bounds relating one to the other can be obtained as in [11]. Notice that many recent works use the nuclear norms of the flattenings of the tensors to be optimized, especially in the field of compressed sensing; see e.g. [17, 8]. One noticeable exception is the recent work [21] where a study of the subdifferential of a purely tensorial nuclear norm is proposed. However, in [21], only a subset of the subdifferential is given but the subdifferential itself could not be fully characterized.

Our goal in the present paper is to extend previous results on matrix norms to the tensor setting. The focus will be on two special type of tensors, namely symmetric tensors and orthogonally decomposable tensors (abbreviated as *odeco* tensor hereafter). Symmetric tensors are invariant under any permutation of its indices [7]. They play an important role in many applications, e.g. Gaussian Mixture Models (GMM), Independent Component Analysis (ICA) and Hidden Markov Models (HMM), see [2] for a survey. *odeco* tensors have a diagonal core in their Higher Order Singular Value Decomposition (HOSVD) [14]. They are special structured tensors that inherit many nice properties of their matrix counterpart. In this contribution, we propose a general study the subdifferential

of certain convex functions of the spectrum of these tensors and apply our results to the computation of the subdifferential of useful and natural tensor norms. The convex conjugate approach used in this paper stems from the elegant work of [16]. One key ingredient for the understanding of tensor norms is the tensor version Von Neumann's trace inequality and the precise description of the equality case. We suspect that the lack of results on the subdifferential of tensor norms in the literature is due to the fact that an extension of the Von Neumann inequality for tensors did not exist until recently; see [6].

The plan of the paper is as follows. In Section 2, we provide a general overview of tensors and their spectral factorizations. In Section 4, we provide a general formula for the subdifferential for symmetric tensors and *odeco* tensors. Finally, in Section 5, we provide formulas for the subdifferential of Schatten norms for symmetric tensors and the subset of *odeco* tensors in the subdifferential of Schatten norms for *odeco* tensors.

1.2. Notations

1.2.1. Convex functions

For any convex function $f : \mathbb{R}^n \mapsto \mathbb{R} \cup \{+\infty\}$, the conjugate function f^* associated to f is defined by

$$f^*(g) \stackrel{\text{def}}{=} \sup_{x \in \mathbb{R}^n} \quad \langle g, x \rangle - f(x).$$

The subdifferential of f at $x \in \mathbb{R}^n$ is defined by

$$\partial f(x) \quad \stackrel{\text{def}}{=} \quad \{g \in \mathbb{R}^n \mid \forall y \in \mathbb{R}^n, \quad f(y) \geq f(x) + \langle g, y - x \rangle\}.$$

It is well known (see e.g. [10]) that $g \in \partial f(x)$ if and only if

$$f(x) + f^*(g) \quad = \quad \langle g, x \rangle.$$

1.2.2. Tensors

Let D and $n_1, \ldots, n_D$ be positive integers. In the present paper, a multi-dimensional array $\mathcal{X}$ in $\mathbb{R}^{n_1 \times \cdots \times n_D}$ is called a D-mode tensor. If $n_1 = \cdots = n_D$, then we will say that tensor $\mathcal{X}$ is cubic. The set of D-mode cubic tensors will be denoted by $\mathbb{R}^{n \times \cdots \times n}$ or $\mathbb{R}^{n^D}$ with a slight abuse of notation.

The mode-d fibers of a tensor $\mathcal{X}$ are the vectors obtained by varying the index i_d while keeping the other indices fixed.

It is often convenient to rearrange the elements of a tensor so that they form a matrix. This operation is referred to as matricization and can be defined in different ways. In this work, $\mathcal{X}_{(d)}$ stands for a matrix in $\mathbb{R}^{n_d \times \prod_{i=1; i \neq d}^{D} n_i}$ obtained by stacking the mode-d fibers of $\mathcal{X}$ one after another with forward cyclic ordering [19]. Inversely, we define the tensorization operator $\mathscr{T}^{(d)}$ as the adjoint of the mode-d matricization operator, i.e. it is such that

$$\langle \mathcal{X}_{(d)}, M \rangle \quad = \quad \langle \mathcal{X}, \mathscr{T}^{(d)}(M) \rangle$$

for all $\mathcal{X} \in \mathbb{R}^{n_1 \times \cdots \times n_D}$ and all $M \in \mathbb{R}^{n_d \times \prod_{i=1; i \neq d}^{D} n_i}$, where $\langle \cdot, \cdot \rangle$ denotes the scalar product defined in Section 2.1.3.

The mode-d multiplication of a tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times \cdots \times n_D}$ by a matrix $M \in \mathbb{R}^{n'_d \times n_d}$, denoted by $\mathcal{X} \times_d M$, yields a tensor in $\mathbb{R}^{n_1 \times \cdots \times n_{d-1} \times n'_d \times n_{d+1} \times \cdots n_d}$. It is defined by

$$(\mathcal{X} \times_d M)_{i_1,\ldots,i_D} = \sum_{i_d} \mathcal{X}_{i_1,\ldots,i_{d-1},i_d,i_{d+1},\ldots,i_D} M_{i_d,i'_d}.$$

Last, we denote by $\otimes$ the tensor product, i.e. for any $v^{(1)}, \ldots, v^{(D)}$ with $v^{(d)} \in \mathbb{R}^{n_d}$, $v^{(1)} \otimes \cdots \otimes v^{(D)}$ is a tensor in $\mathbb{R}^{n_1 \times \cdots \times n_D}$ whose entries are given by

$$\left(v^{(1)} \otimes \cdots \otimes v^{(D)}\right)_{i_1,\ldots,i_D} = v^{(1)}_{i_1} \cdots v^{(D)}_{i_D}.$$

2. Basics on tensors

2.1. General tensors

2.1.1. Tensor rank

If a tensor $\mathcal{X}$ can be written as

$$\mathcal{X} = v^{(1)} \otimes \cdots \otimes v^{(D)},$$

then we say $\mathcal{X}$ is a rank one tensor. Any tensor $\mathcal{X}$ can easily be written as a sum of rank one tensors. Indeed, if $(e^{(n)}_i)_{i=1,\ldots,n}$ denotes the canonical basis of $\mathbb{R}^n$, we have

$$\mathcal{X} = \sum_{i_1=1,\ldots,n_1,\ldots,i_D=1,\ldots,n_D} x_{i_1,\ldots,i_D} \cdot e^{(n_1)}_{i_1} \otimes \ldots \otimes e^{(n_D)}_{i_D}.$$

Among all possible decomposition as a sum of rank one tensors, one may look for the one involving the least possible number of summands, i.e.

$$\mathcal{X} = \sum_{j=1,\ldots,r} v^{(1)}_j \otimes \ldots \otimes v^{(D)}_j, \tag{2.1}$$

for some vectors $v^{(d)}_j$, $j = 1, \ldots, r$ and $d = 1, \ldots, D$. The number r is called the rank of $\mathcal{X}$. It is already known that the rank of a tensor is NP-hard to compute [13].

2.1.2. The Higher Order SVD

One of the main problems with the "sum-of-rank-one" decomposition (2.1) is that the vectors $v^{(d)}_j$, $j = 1, \ldots, r$ may not form an orthonormal family of vectors. The Tucker decomposition of a tensor is another decomposition which reveal a possibly smaller tensor $\mathcal{S}$ hidden inside $\mathcal{X}$ under orthogonal transformations. More precisely, we have

$$\mathcal{X} = \mathcal{S} \times_1 U^{(1)} \times_2 U^{(2)} \cdots \times_D U^{(D)}, \tag{2.2}$$

where each $U^{(d)}$ is orthogonal and $\mathcal{S}$ is a tensor of the same size as $\mathcal{X}$ defined as follows. Take the (usual) SVD of the matrix $\mathcal{X}_{(d)}$

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

and based on [14], we can set

$$\mathcal{S}_{(d)} \quad = \quad \Sigma^{(d)}{V^{(d)}}^t \left(U^{(d+1)} \otimes \cdots \otimes U^{(D)} \otimes U^{(1)} \otimes \cdots \otimes U^{(d-1)}\right).$$

Then, $\mathcal{S}_{(d)}$ is the mode-d matricization of $\mathcal{S}$. One proceeds similarly for all $d = 1, \ldots, D$ and one recovers the orthogonal matrices $U^{(1)}, \ldots, U^{(D)}$ which allow us to decompose $\mathcal{X}$ as in (2.2). Notice that this construction implies that $\mathcal{S}$ has orthonormal fibers for every modes.

2.1.3. Norms of tensors

Several tensor norms can be defined on the tensor space $\mathbb{R}^{n_1 \times \cdots \times n_D}$. The first one is a natural extension of the Frobenius norm or Hilbert-Schmidt norm from matrices. We start by defining the scalar product on $\mathbb{R}^{n_1 \times \cdots \times n_D}$ as

$$\langle \mathcal{X}, \mathcal{Y} \rangle \quad \stackrel{\text{def}}{=} \quad \sum_{i_1=1}^{n_1} \cdots \sum_{i_D=1}^{n_D} x_{i_1,\ldots,i_D} y_{i_1,\ldots,i_D}.$$

Using this scalar product, we can define the Frobenius norm of tensors as

$$\|\mathcal{X}\|_F \quad \stackrel{\text{def}}{=} \quad \sqrt{\langle \mathcal{X}, \mathcal{X} \rangle}.$$

In this work, we shall focus on a family of tensor norms called Schatten-(p, q) norms. The Schatten-(p, q) norm of $\mathcal{X}$ is defined by

$$\|\mathcal{X}\|_{p,q} \stackrel{\text{def}}{=} \lambda \Big(\sum_{d=1}^{D} \|\sigma^{(d)}(\mathcal{X})\|_p^q \Big)^{1/q}, \tag{2.3}$$

where $\sigma^{(d)}(\mathcal{X})$ is the vector of singular values of $\mathcal{X}_{(d)}$, called the mode-d spectrum of $\mathcal{X}$, and λ is a positive constant. In the particular case that $p = q = 1$ and $\lambda = 1/D$, the Schatten-$(1, 1)$ norm will be referred to as the *nuclear norm*, and will be denoted by $\|\cdot\|_*$ instead, i.e.

$$\|\mathcal{X}\|_* \stackrel{\text{def}}{=} \frac{1}{D} \sum_{d=1}^{D} \|\sigma^{(d)}(\mathcal{X})\|_1.$$

2.2. Orthogonally decomposable tensors

Definition 2.1. *Let $\mathcal{X}$ be a tensor in $\mathbb{R}^{n_1 \times \cdots \times n_D}$. If*

$$\mathcal{X} \quad = \quad \sum_{i=1}^{r} \alpha_i \cdot u_i^{(1)} \otimes \cdots \otimes u_i^{(D)}, \tag{2.4}$$

*where $r \leqslant n_1 \wedge \cdots \wedge n_D$, $\alpha_1 \geqslant \cdots \geqslant \alpha_r > 0$ and $\{u_1^{(d)}, \ldots, u_r^{(d)}\}$ is a family of orthonormal vectors for each $d = 1, \ldots, D$, then we say (2.4) is an orthogonal decomposition of $\mathcal{X}$ and $\mathcal{X}$ is an orthogonally decomposable (*odeco*) tensor.*

Denote $\alpha = (\alpha_1, \ldots, \alpha_r, 0, \ldots, 0)$ in $\mathbb{R}^{n_1 \wedge \cdots \wedge n_D}$. For each $d \in \{1, \ldots, D\}$, we may complete $\{u_1^{(d)}, \ldots, u_r^{(d)}\}$ with $\{u_{r+1}^{(d)}, \ldots, w_{n_d}^{(d)}\}$ so that matrix $U^{(d)} = (u_1^{(d)}, \ldots, u_{n_d}^{(d)}) \in \mathbb{R}^{n_d \times n_d}$ is orthogonal. Using $U^{(1)}, \ldots, U^{(D)}$, we may write (2.4) as

$$\mathcal{X} \quad = \quad \mathrm{diag}(\alpha) \times_1 U^{(1)} \times_2 U^{(2)} \cdots \times_D U^{(D)}, \tag{2.5}$$

where $\mathrm{diag}(\alpha)$ denotes the cubic tensor whose diagonal entries are given by vector α.

2.3. Symmetric tensors

Let $\mathfrak{S}_D$ be the set of permutations over $\{1, \ldots, D\}$. A D-mode cubic tensor $\mathcal{X} \in \mathbb{R}^{nD}$ will be said symmetric if for all $\tau \in \mathfrak{S}_D$,

$$\mathcal{X}_{i_1, \ldots, i_D} \quad = \quad \mathcal{X}_{\tau(i_1), \ldots, \tau(i_D)}$$

The set of all symmetric tensors of order n will be denoted by $\mathbb{S}_n$. An immediate result is the following useful proposition whose proof is straightforward.

2.4. Spectrum of tensors

Let $\mathbb{E}$ denote a subspace of the space of all tensors. Let us define the spectrum as the mapping which to any tensor $\mathcal{X} \in \mathbb{E}$ associates the vector $\sigma_{\mathbb{E}}(\mathcal{X})$ given by

$$\sigma_{\mathbb{E}}(\mathcal{X}) \quad \stackrel{\text{def}}{=} \quad \frac{1}{\sqrt{D}} \left(\sigma^{(1)}(\mathcal{X}), \ldots, \sigma^{(D)}(\mathcal{X})\right).$$

Here we stress that the underlying tensor subspace $\mathbb{E}$ does make a difference. For instance, although $\sigma_{\mathbb{R}^{nD}}(\mathcal{X}) = \sigma_{\mathbb{S}_n}(\mathcal{X})$ for all $\mathcal{X} \in \mathbb{S}_n$, the two different tensor space $\mathbb{R}^{nD}$ and $\mathbb{S}$ may result in different subdifferential of the same tensor norm.

The following result is straight forward.

Proposition 2.2. *If $\mathcal{X}$ is either* odeco *or symmetric, then $\sigma^{(1)}(\mathcal{X}) = \cdots = \sigma^{(D)}(\mathcal{X})$.*

3. Further results on the spectrum

In this section, we will present some further results on the spectrum such as the question of characterizing the image of the spectrum and the subdifferential of a function of the spectrum.

3.1. The Von Neumann inequality for tensors

Von Neumann's inequality says that for any two matrices X and Y in $\mathbb{R}^{n_1 \times n_2}$, we have

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

with equality when the singular vectors of X and Y are equal, up to permutations when the singular values have multiplicity greater than one. This result has proved useful for the study of the subdifferential of unitarily invariant convex functions of the spectrum in the matrix case in [16]. In order to study the subdifferential of the norms of symmetric tensors, we will need a generalization of this result to higher orders. This was worked out in [6].

Definition 3.1. *We say that a tensor $\mathcal{S}$ is blockwise decomposable if there exists an integer B and if, for all $d = 1, \ldots, D$, there exists a partition $I_1^{(d)} \cup \ldots \cup I_B^{(d)}$ into disjoint index subsets of $\{1, \ldots, n_d\}$, such that $\mathcal{X}_{i_1,\ldots,i_D} = 0$ if for all $b = 1, \ldots, B$, $(i_1, \ldots, i_D) \notin I_b^{(1)} \times \ldots \times I_b^{(D)}$.*

An illustration of this block decomposition can be found in Figure 1. The following result is a generalization of von Neumann's inequality from matrices to tensors. It is proved in [6].

Theorem 3.2. *Let $\mathcal{X}, \mathcal{Y} \in \mathbb{R}^{n_1 \times \cdots \times n_D}$ be tensors. Then for all $d = 1, \ldots, D$, we have*

$$\langle \mathcal{X}, \mathcal{Y} \rangle \leqslant \langle \sigma^{(d)}(\mathcal{X}), \sigma^{(d)}(\mathcal{Y}) \rangle. \tag{3.6}$$

Equality in (3.6) holds simultaneously for all $d = 1, \ldots, D$ if and only there exist orthogonal matrices $W^{(d)} \in \mathbb{R}^{n_d \times n_d}$ for $d = 1, \ldots, D$ and tensors $\mathcal{D}(\mathcal{X}), \mathcal{D}(\mathcal{Y}) \in \mathbb{R}^{n_1 \times \cdots \times n_D}$ such that

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

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

(i) $\mathcal{D}(\mathcal{X})$ and $\mathcal{D}(\mathcal{Y})$ are block-wise decomposable with the same number of blocks, which we will denote by B,

(ii) the blocks $\{\mathcal{D}_b(\mathcal{X})\}_{b=1,\ldots,B}$ (resp. $\{\mathcal{D}_b(\mathcal{Y})\}_{b=1,\ldots,B}$) on the diagonal of $\mathcal{D}(\mathcal{X})$ (resp. $\mathcal{D}(\mathcal{Y})$) have the same sizes,

(iii) for each $b = 1, \ldots, B$ the two blocks $\mathcal{D}_b(\mathcal{X})$ and $\mathcal{D}_b(\mathcal{Y})$ are proportional, i.e. there exist $c_1, c_2 \in \mathbb{R}$ with $c_1 c_2 \neq 0$, such that $c_1 \mathcal{D}_b(\mathcal{X}) = c_2 \mathcal{D}_b(\mathcal{Y})$.

3.2. Surjectivity of the spectrum

Note that any diagonal tensor is both *odeco* and symmetric. A diagonal tensor $\mathcal{X}$ with non-negative diagonal entries $(\lambda_1, \ldots, \lambda_n)$ satisfies $\sigma^{(1)}(\mathcal{X}) = \cdots = \sigma^{(D)}(\mathcal{X}) = (\lambda_1, \ldots, \lambda_n)^\mathsf{T}$. Therefore, for any non-negative vector s, there exist a symmetric and *odeco* tensor $\mathcal{X}$ such that $\sigma(\mathcal{X}) = (s, \ldots, s)$.

Notice that general tensors have different spectra along all the different modes and the question of analysing the surjectivity is much more subtle.

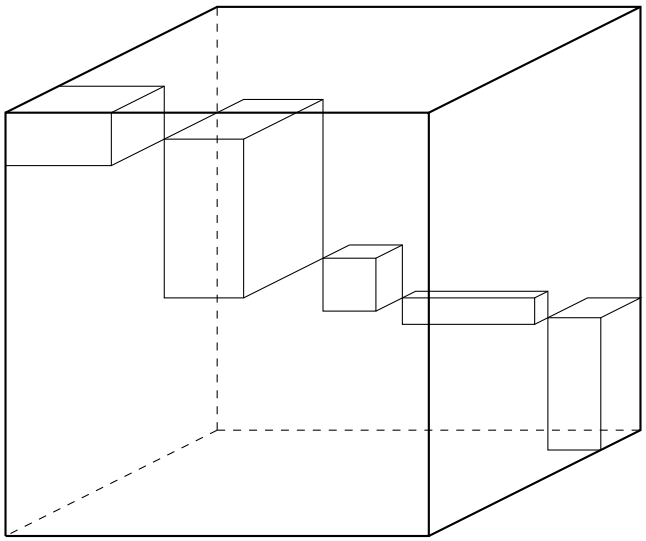

Figure 1: A block-wise diagonal tensor.

4. The subdifferential of functions of the spectrum

Hereafter, we shall always assume that $f : \mathbb{R}^n \times \cdots \times \mathbb{R}^n \mapsto \mathbb{R}$ is a proper closed convex function that satisfies the following property:

$$f(s_1, \ldots, s_D) = f(|s_{\tau(1)}|, \ldots, |s_{\tau(D)}|), \quad \forall s_1, \ldots, s_D \in \mathbb{R}, \quad \forall \tau \in \mathfrak{S}_D. \tag{4.7}$$

See Section 5 for a concrete example of f.

The purpose of this section is to present a characterization of the subdifferential of f of the spectrum for symmetric and *odeco* tensors.

4.1. Lewis' characterization of the subdifferential

Let us recall that the spectrum is defined on a subspace $\mathcal{E}$. In this section, we recall the result of Lewis in the setting of tensor spectra, which characterizes the subdifferential if the formula

$$(f \circ \sigma_{\mathbb{E}})^* = f^* \circ \sigma_{\mathbb{E}} \tag{4.8}$$

holds on the domain of definition of $\sigma_{\mathbb{E}}$. The proof is exactly the same as in [16] thanks to the tensor version of Von Neumann's inequality. We recall it here for the sake of completeness.

Theorem 4.1. *Let $\mathcal{X}$ and $\mathcal{Y}$ be two tensors in $\mathbb{R}^{nD}$. If (4.8) holds, then $\mathcal{Y} \in \partial(f \circ \sigma_{\mathbb{E}})(\mathcal{X})$ if and only if the following conditions are satisfied:*

(i) $\sigma_{\mathbb{E}}(\mathcal{Y}) \in \partial f(\sigma_{\mathbb{E}}(\mathcal{X}))$,

(ii) equality holds in the Von Neumann inequality, i.e. $\langle \mathcal{X}, \mathcal{Y} \rangle = \langle \sigma_{\mathbb{E}}(\mathcal{X}), \sigma_{\mathbb{E}}(\mathcal{Y}) \rangle$.

PROOF. As is well known, $\mathcal{Y} \in \partial(f \circ \sigma_{\mathbb{E}})(\mathcal{X})$ if and only if

$$(f \circ \sigma_{\mathbb{E}})(\mathcal{X}) + (f \circ \sigma_{\mathbb{E}})^*(\mathcal{Y}) = \langle \mathcal{X}, \mathcal{Y} \rangle.$$

Recall that we assumed the following equality to hold

$$(f \circ \sigma_{\mathbb{E}})(\mathcal{X}) + (f \circ \sigma_{\mathbb{E}})^*(\mathcal{Y}) = f(\sigma_{\mathbb{E}}(\mathcal{X})) + f^*(\sigma_{\mathbb{E}}(\mathcal{Y})).$$

On the other hand,

$$f(\sigma_{\mathbb{E}}(\mathcal{X})) + f^*(\sigma_{\mathbb{E}}(\mathcal{Y})) \geqslant \langle \sigma_{\mathbb{E}}(\mathcal{X}), \sigma_{\mathbb{E}}(\mathcal{Y}) \rangle,$$

where equality takes place if and only if

$$\sigma_{\mathbb{E}}(\mathcal{Y}) \in \partial f(\sigma_{\mathbb{E}}(\mathcal{X})).$$

Finally, by the von Neumann inequality, we have

$$\langle \sigma_{\mathbb{E}}(\mathcal{X}), \sigma_{\mathbb{E}}(\mathcal{Y}) \rangle \geq \langle \mathcal{X}, \mathcal{Y} \rangle,$$

where the equality takes place if and only if the equality condition is satisfied.

4.2. The symmetric case

Throughout this section $\mathbb{E}$ will be the set $\mathbb{S}_n$ of all symmetric tensors in $\mathbb{R}^{nD}$.

4.2.1. Proving (4.8) *for symmetric tensors*

There exists a simple formula for the conjugate of the composition of the spectrum with a convex function. This formula will be helpful for gaining useful information on the subdifferential of convex functions of the spectrum.

Theorem 4.2. *Let* $\mathcal{X}$ *be a* symmetric *tensor in* $\mathbb{R}^{nD}$. *Then,*

$$(f \circ \sigma_{\mathbb{S}_n})^*(\mathcal{X}) \quad = \quad f^* \circ \sigma_{\mathbb{S}_n}(\mathcal{X})$$

PROOF. This proof mimics the proof of [16, Theorem 2.4]. Let

$$\mathcal{X} = \mathcal{S}_{\mathcal{X}} \times_1 U \cdots \times_D U$$

denote the Higher Order singular value decomposition of $\mathcal{X}$. By definition of conjugacy, we have

$$(f \circ \sigma_{\mathbb{S}_n})^*(\mathcal{X}) \quad = \quad \sup_{\mathcal{Y} \in \mathbb{R}^{n \times \cdots \times n}} \quad \langle \mathcal{X}, \mathcal{Y} \rangle - f(\sigma_{\mathbb{S}_n}(\mathcal{Y})).$$

By the tensor von Neumann inequality we have

$$\begin{aligned} &\sup_{\mathcal{Y} \in \mathbb{R}^{n \times \cdots \times n}} \quad \langle \mathcal{X}, \mathcal{Y} \rangle - f(\sigma_{\mathbb{S}_n}(\mathcal{Y})) \\ &\qquad \leq \sup_{\mathcal{Y} \in \mathbb{R}^{n \times \cdots \times n}} \quad \frac{1}{D} \sum_{d=1}^{D} \langle \sigma^{(d)}(\mathcal{X}), \sigma^{(d)}(\mathcal{Y}) \rangle - f(\sigma_{\mathbb{S}_n}(\mathcal{Y})) \qquad (4.9) \\ &\qquad \leq \sup_{s_1, \ldots, s_D \in \mathbb{R}^n} \quad \frac{1}{\sqrt{D}} \sum_{d=1}^{D} \langle \sigma^{(d)}(\mathcal{S}_{\mathcal{X}}), s_d \rangle - f\left(\frac{1}{\sqrt{D}} (s_1, \ldots, s_D) \right) \qquad (4.10) \end{aligned}$$

By (4.7) and the symmetry of $\mathcal{X}$, there exists a maximizer s^* of the right hand side term of this last equation that satisfies $s^* \geq 0$ and $s_1^* = \ldots = s_D^*$. Now, based on Section 3.2, there exists a tensor $\mathcal{S}^*$ such that $s^* = \sigma(\mathcal{S}^*)$. Thus, we obtain that

$$\begin{aligned}\sup_{s_1,\ldots,s_D \in \mathbb{R}^n} \quad & \frac{1}{\sqrt{D}} \sum_{d=1}^{D} \langle \sigma^{(d)}(\mathcal{S}_\mathcal{X}), s_d \rangle - f\left(\frac{1}{\sqrt{D}}(s_1, \ldots, s_D) \right) \\ & = \frac{1}{D} \sum_{d=1}^{D} \langle \sigma^{(d)}(\mathcal{S}_\mathcal{X}), \sigma^{(d)}(\mathcal{S}^*) \rangle - f(\sigma_{\mathbb{S}_n}(\mathcal{S}^*)).\end{aligned}$$

On the one hand, using that $\mathcal{S}^*$ has support included in S and Theorem 3.2, we obtain that

$$\begin{aligned}\sup_{s_1,\ldots,s_D \in \mathbb{R}^n} \quad & \frac{1}{\sqrt{D}} \sum_{d=1}^{D} \langle \sigma^{(d)}(\mathcal{S}_\mathcal{X}), s_d \rangle - f\left(\frac{1}{\sqrt{D}}(s_1, \ldots, s_D) \right) \\ & = \langle \mathcal{X}, \mathcal{X}^* \rangle - f(\sigma_{\mathbb{S}_n}(\mathcal{X}^*))\end{aligned}$$

where

$$\mathcal{X}^* = \mathcal{S}^* \times_1 U \cdots \times_D U.$$

From this, we deduce that

$$\begin{aligned}\sup_{s_1,\ldots,s_D \in \mathbb{R}^n} \quad & \frac{1}{\sqrt{D}} \sum_{d=1}^{D} \langle \sigma^{(d)}(\mathcal{S}_\mathcal{X}), s_d \rangle - f\left(\frac{1}{\sqrt{D}}(s_1, \ldots, s_D) \right) \\ & = \sup_{\mathcal{Y} \in \mathbb{R}^{n \times \cdots \times n}} \quad \langle \mathcal{X}, \mathcal{Y} \rangle - f(\sigma_{\mathbb{S}_n}(\mathcal{Y})) \\ & = (f \circ \sigma_{\mathbb{S}_n})^*(\mathcal{X}).\end{aligned} \tag{4.11}$$

On the other hand, using the fact that $\sigma_{\mathbb{S}_n}(\mathcal{S}_\mathcal{X}) = \sigma_{\mathbb{S}_n}(\mathcal{X})$,

$$\begin{aligned}\sup_{s_1,\ldots,s_D \in \mathbb{R}^n} \quad & \frac{1}{\sqrt{D}} \sum_{d=1}^{D} \langle \sigma^{(d)}(\mathcal{S}_\mathcal{X}), s_d \rangle - f\left(\frac{1}{\sqrt{D}}(s_1, \ldots, s_D) \right) \\ & = \sup_{s_1,\ldots,s_D \in \mathbb{R}^n} \quad \frac{1}{\sqrt{D}} \sum_{d=1}^{D} \langle \sigma^{(d)}(\mathcal{S}_\mathcal{X}), s_d \rangle - f\left(\frac{1}{\sqrt{D}}(s_1, \ldots, s_D) \right) \\ & = f^* \circ \sigma_{\mathbb{S}_n}(\mathcal{X}).\end{aligned}$$

Therefore,

$$(f \circ \sigma_{\mathbb{S}_n})^*(\mathcal{X}) \quad = \quad f^* \circ \sigma_{\mathbb{S}_n}(\mathcal{X})$$

as announced.

4.2.2. A closed form formula for the subdifferential

We now present a closed form formula for the subdifferential of a symmetric function of the spectrum of a symmetric tensor.

Corollary 4.3. *The necessary and sufficient conditions for an symmetric tensor $\mathcal{Y}$ to belong to $\partial(f \circ \sigma_{\mathbb{R}^{nD}})(\mathcal{X})$ are*

1. *$\mathcal{Y}$ has the same mode-d singular spaces as $\mathcal{X}$ for all $d = 1, \ldots, D$*
2. *$\sigma_{\mathbb{R}^{nD}}(\mathcal{Y}) \in \partial f(\sigma_{\mathbb{R}^{nD}}(\mathcal{X}))$.*

Proof. Combine Theorem 4.2 with Theorem 4.1.

4.3. The odeco *case*

Throughout this subsection $\mathbb{E} = \mathbb{R}^{nD}$.

4.3.1. Proving (4.8) *for* odeco *tensors*

As for the symmetric case, we start with a result in the spirit of (4.8).

Theorem 4.4. *For all* odeco *tensors $\mathcal{X}$, we have*

$$(f \circ \sigma_{\mathbb{R}^{nD}})^*(\mathcal{X}) = f^*(\sigma_{\mathbb{R}^{nD}}(\mathcal{X})) \tag{4.12}$$

Proof. By definition, equality (4.12) is equivalent to

$$\sup_{\mathcal{Y}}\{\langle \mathcal{X}, \mathcal{Y}\rangle - f(\sigma_{\mathbb{R}^{nD}}(\mathcal{Y}))\} = \sup_{s_1,\ldots,s_D \in \mathbb{R}^n} \left\{ \frac{1}{D}\sum_{d=1}^{D} \langle \sigma^{(d)}(\mathcal{X}), s_d\rangle - f(s_1, \ldots, s_D) \right\}. \tag{4.13}$$

Consider the optimization problem

$$\sup_{\mathcal{Y}} \Big\{ \langle \mathcal{X}, \mathcal{Y}\rangle, \quad f(\sigma_{\mathbb{R}^{nD}}(\mathcal{Y})) \leqslant C \Big\} \tag{4.14}$$

and

$$\sup_{s_1,\ldots,s_D \in \mathbb{R}^n} \left\{ \frac{1}{D}\sum_{d=1}^{D} \langle \sigma^{(d)}(\mathcal{X}), s_d\rangle, \quad f(s_1, \ldots, s_D) \leqslant C \right\}. \tag{4.15}$$

Clearly, we have

$$\begin{aligned} & \sup_{\mathcal{Y}}\{\langle \mathcal{X}, \mathcal{Y}\rangle, \quad f(\sigma_{\mathbb{R}^{nD}}(\mathcal{Y})) \leqslant C\} \\ \leqslant \; & \sup_{\mathcal{Y}} \left\{ \frac{1}{D}\sum_{d=1}^{D} \langle \sigma^{(d)}(\mathcal{X}), \sigma^{(d)}(\mathcal{Y})\rangle, \quad f(\sigma_{\mathbb{R}^{nD}}(\mathcal{Y})) \leqslant C \right\} \\ \leqslant \; & \sup_{s_1,\ldots,s_D \in \mathbb{R}^n} \left\{ \frac{1}{D}\sum_{d=1}^{D} \langle \sigma^{(d)}(\mathcal{X}), s_d\rangle, \quad f(s_1, \ldots, s_D) \leqslant C \right\}. \end{aligned} \tag{4.16}$$

Assume that the supremum (4.15) is achieved at $(s_1^*, \ldots, s_D^*)$. Denote

$$s^\dagger = \frac{1}{D!} \sum_\tau s_{\tau(k)}.$$

Clearly, $s^\dagger$ is independent of k. Moreover, we have

$$\begin{aligned} \frac{1}{D} \sum_{d=1}^D \langle \sigma^{(d)}(\mathcal{X}), s_d \rangle &= \frac{1}{D} \sum_{d=1}^D \langle \sigma^{(d)}(\mathcal{X}), s^\dagger \rangle \\ f(s^\dagger, \ldots, s^\dagger) &\leqslant \frac{1}{D!} \sum_\tau f(s_{\tau(1)}, \ldots, s_{\tau(D)}). \end{aligned}$$

Using (4.7), we have

$$\begin{aligned} \frac{1}{D} \sum_{d=1}^D \langle \sigma^{(d)}(\mathcal{X}), s_d \rangle &= \frac{1}{D} \sum_{d=1}^D \langle \sigma^{(d)}(\mathcal{X}), s^\dagger \rangle \\ f(s^\dagger, \ldots, s^\dagger) &\leqslant C. \end{aligned}$$

This means that the supremum of (4.15) can also be achieved at $(s^\dagger, \ldots, s^\dagger)$. Now take an *odeco* tensor $\mathcal{Y}^\dagger$ such that $\sigma^{(d)}(\mathcal{Y}^\dagger) = s^\dagger$ and $\mathcal{Y}^\dagger$ has the same singular matrices as $\mathcal{X}$. For this particular $\mathcal{Y}^\dagger$, we have by the equality condition of the generalized von Neumann's Theorem

$$\langle \mathcal{X}, \mathcal{Y}^\dagger \rangle = \langle \sigma_{\mathbb{R}^{nD}}(\mathcal{X}), \sigma_{\mathbb{R}^{nD}}(\mathcal{Y}^\dagger) \rangle = \frac{1}{D} \sum_{d=1}^D \langle \sigma^{(d)}(\mathcal{X}), s^\dagger \rangle$$

and

$$f(\sigma_{\mathbb{R}^{nD}}(\mathcal{Y}^\dagger)) = f(s^\dagger, \ldots, s^\dagger) \leqslant C.$$

We then deduce that

$$\begin{aligned} &\sup_{\mathcal{Y}} \{ \langle \mathcal{X}, \mathcal{Y} \rangle, \quad f(\sigma_{\mathbb{R}^{nD}}(\mathcal{Y})) \leqslant C \} \geqslant \\ &\qquad \langle \mathcal{X}, \mathcal{Y}^\dagger \rangle = \sup_{s_1, \ldots, s_D \in \mathbb{R}^n} \left\{ \frac{1}{D} \sum_{d=1}^D \langle \sigma^{(d)}(\mathcal{X}), s_d \rangle, \quad f(s_1, \ldots, s_D) \leqslant C \right\}. \end{aligned} \tag{4.17}$$

Combining (4.16) and (4.17) gives

$$\begin{aligned} &\sup_{\mathcal{Y}} \{ \langle \mathcal{X}, \mathcal{Y} \rangle, \quad f(\sigma_{\mathbb{R}^{nD}}(\mathcal{Y})) \leqslant C \} = \\ &\qquad \sup_{s_1, \ldots, s_D \in \mathbb{R}^n} \left\{ \frac{1}{D} \sum_{d=1}^D \langle \sigma^{(d)}(\mathcal{X}), s_d \rangle, \quad f(s_1, \ldots, s_D) \leqslant C \right\}. \end{aligned}$$

Then (4.13) follows.

4.3.2. A closed form formula for a subset of the subdifferential

The following Corollary is a direct consequence of Theorem 4.1 and Theorem 4.4.

Corollary 4.5. *Let $\mathcal{X}$ be an* odeco *tensor. Then necessary and sufficient conditions for an* odeco *tensor $\mathcal{Y}$ to belong to $\partial(f \circ \sigma_{\mathbb{R}^{nD}})(\mathcal{X})$ are*

1. *$\mathcal{Y}$ has the same mode-d singular spaces as $\mathcal{X}$ for all $d = 1, \ldots, D$*
2. *$\sigma_{\mathbb{R}^{nD}}(\mathcal{Y}) \in \partial f(\sigma_{\mathbb{R}^{nD}}(\mathcal{X}))$.*

Moreover, the closure of the convex combination of these odeco *tensors belongs to $\partial(f \circ \sigma_{\mathbb{R}^{nD}})(\mathcal{X})$.*

5. The subdifferential of tensor Schatten norms for symmetric and *odeco* tensors

In this section, we compute the subdifferential of the Schatten norm (2.3) for symmetric and *odeco* tensors. Consider $f\colon \mathbb{R}^n \times \cdots \times \mathbb{R}^n \mapsto \mathbb{R}$ defined by

$$f(s_1, \ldots, s_D) = \lambda \Big(\sum_{d=1}^{D} \|s_d\|_p^q \Big)^{1/q} \tag{5.18}$$

for some integers $p, q \geqslant 1$ and constant $\lambda > 0$. Clearly, f is a convex function and

$$\|\cdot\|_{p,q} = f\Big(\sigma^{(1)}(\cdot), \ldots, \sigma^{(D)}(\cdot)\Big).$$

5.1. The case $p, q > 1$

In this case we can write (5.18) as

$$f(s_1, \ldots, s_D) = \lambda \sup_{\substack{\|w\|_{q^*}=1 \\ \|v_d\|_{p^*}=1, d=1,\ldots,D}} \left\{ \sum_{d=1}^{D} w_d \, \langle v_d, s_d \rangle \right\}. \tag{5.19}$$

Notice that since the supremum in (5.19) is taken over a compact set and the function to be maximized is continuous, then this supremum is attained. Let $\mathcal{VW}^*$ denote the set of maximizers in the variational formulation of f (5.19). Then, by [10] the subdifferential of f is given by

$$\partial f(v_1, \ldots, v_D) = \lambda \, \overline{\text{conv}}\Big\{ (w_1^* v_1^*, \ldots, w_D^* v_D^*) \mid (v_1^*, \ldots, v_D^*, w^*) \in \mathcal{V}W^* \Big\}.$$

Notice that $\mathcal{VW}^*$ is fully characterized by

$$v_d^* = \begin{cases} s_d / \|s_d\|_{p^*} \text{ if } s_d \neq 0 \\ \\ B_{p^*} \text{ otherwise} \end{cases}$$

$$w^* = \begin{cases} \omega^* / \|\omega^*|_{q^*} \text{ if } \omega^* \neq 0 \\ \\ B_{q^*} \text{ otherwise.} \end{cases}$$

with

$$\omega^* = (\langle v_d^*, s_d \rangle)_{d=1}^D$$

and B_p denotes the unit ball in the ℓ_p norm.

Using these computations, we obtain the following result.

Theorem 5.1. *We have*

1. *the subdifferential of the nuclear norm for symmetric tensors is the set of tensors $\mathcal{Y}$ satisfying*
   (a) *$\mathcal{Y}$ has the same mode-d singular spaces as $\mathcal{X}$ for all $d = 1, \ldots, D$*
   (b) *$\sigma_{\mathbb{R}^{n_D}}(\mathcal{Y})_d = w_d^* v_d^*$ if $\sigma^{(d)}(\mathcal{X}) \neq 0$,*
   (c) *$\sigma_{\mathbb{R}^{n_D}}(\mathcal{Y})_d = 0$ if $\sigma^{(d)}(\mathcal{X}) = 0$ and if $\sigma^{(d')}(\mathcal{X}) \neq 0$ for some d'.*
   (d) *$\sigma_{\mathbb{R}^{n_D}}(\mathcal{Y})_d \in w_d^* B_{p^*}$, $w^* \in B_{q^*}$ if $\sigma^{(d')}(\mathcal{X}) = 0$ for all $d' = 1, \ldots, D$.*
2. *the subdifferential of the nuclear norm for* odeco *tensors contains the closure of the convex hull of* odeco *tensors $\mathcal{Y}$ satisfying*
   (a) *$\mathcal{Y}$ has the same mode-d singular spaces as $\mathcal{X}$ for all $d = 1, \ldots, D$*
   (b) *$\sigma_{\mathbb{R}^{n_D}}(\mathcal{Y})_d = w_d^* v_d^*$ if $\sigma^{(d)}(\mathcal{X}) \neq 0$,*
   (c) *$\sigma_{\mathbb{R}^{n_D}}(\mathcal{Y})_d = 0$ if $\sigma^{(d)}(\mathcal{X}) = 0$ and if $\sigma^{(d')}(\mathcal{X}) \neq 0$ for some d'.*
   (d) *$\sigma_{\mathbb{R}^{n_D}}(\mathcal{Y})_d \in w_d^* B_{p^*}$, $w^* \in B_{q^*}$ if $\sigma^{(d')}(\mathcal{X}) = 0$ for all $d' = 1, \ldots, D$.*

5.2. The nuclear norm

Consider $f(\cdot) : \mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \mathbb{R}$ defined by

$$f(s_1, \ldots, s_D) = \frac{1}{D} \sum_{i=1}^{D} \|s_i\|_1.$$

Then for any $(s_1, \ldots, s_D) \in \mathbb{R}^n \times \cdots \times \mathbb{R}^n$, we have

$$\partial f(s_1, \ldots, s_D) = \frac{1}{D} \{(c_1, \ldots, c_D)\},$$

where $c_d = (c_{d1}, \ldots, c_{dn})^t$ for $d = 1, \ldots, D$ satisfies

$$c_{dj} = \begin{cases} 1 & s_{dj} > 0 \\ -1 & s_{dj} < 0 \\ \omega_{dj} & s_{dj} = 0 \end{cases}$$

with ω_{ij} being any real number in the interval $[-1, 1]$.

Thus, we obtain the following result.

Theorem 5.2. *We have that*

1. *the subdifferential of the nuclear norm for symmetric tensors is the set of tensors $\mathcal{Y}$ satisfying*

   (a) *$\mathcal{Y}$ has the same mode-d singular spaces as $\mathcal{X}$ for all $d = 1, \ldots, D$*

   (b) *$\sigma_{\mathbb{R}^{nD}}(\mathcal{Y})_{dj} = 1$ if $\sigma_j^{(d)}(\mathcal{X}) > 0$,*

   (c) *$\sigma_{\mathbb{R}^{nD}}(\mathcal{Y})_{dj} \in [-1, 1]$ if $\sigma_j^{(d)}(\mathcal{X}) = 0$.*

2. *the subdifferential of the nuclear norm for* odeco *tensors contains the closure of the convex hull of* odeco *tensors $\mathcal{Y}$ satisfying*

   (a) *$\mathcal{Y}$ has the same mode-d singular spaces as $\mathcal{X}$ for all $d = 1, \ldots, D$*

   (b) *$\sigma_{\mathbb{R}^{nD}}(\mathcal{Y})_{dj} = 1$ if $\sigma_j^{(d)}(\mathcal{X}) > 0$,*

   (c) *$\sigma_{\mathbb{R}^{nD}}(\mathcal{Y})_{dj} \in [-1, 1]$ if $\sigma_j^{(d)}(\mathcal{X}) = 0$.*

5.3. Remark on the remaining cases

We leave to the reader the easy task of deriving the general formulas for the cases $p = 1$ and $q > 1$, $p > 1$ and $q = 1$.

6. Conclusion and perspectives

In this paper, we studied the subdifferential of some tensor norms for symmetric tensors and *odeco* tensors. We provided a complete characterization of for the symmetric case and described a subset of the subdifferential for *odeco* tensors. The main tool in our analysis is an extension of the Von Neumann's trace inequality to the tensor setting recently proved in [6]. Such results may find applications in the field of Compressed Sensing. A lot of work remains in order to extend our results to non-symmetric settings. We plan to investigate this question in a future research project.

Acknowledgements

Both authors contributed equally to this work. The work of the first author was funded by The National Measurement Office of the UK's Department for Business, Energy and Industrial Strategy supported this work as part of its Materials and Modelling programme.

[1] Dennis Amelunxen, Martin Lotz, Michael B McCoy, and Joel A Tropp, *Living on the edge: Phase transitions in convex programs with random data*, Information and Inference (2014), iau005.

[2] Animashree Anandkumar, Rong Ge, Daniel Hsu, Sham M Kakade, and Matus Telgarsky, *Tensor decompositions for learning latent variable models*, The Journal of Machine Learning Research **15** (2014), no. 1, 2773–2832.

[3] Emmanuel J. Candès, Xiaodong Li, Yi Ma, and John Wright, *Robust principal component analysis?*, J. ACM **58** (2011), no. 3, 11:1–11:37.

[4] Emmanuel J Candès and Benjamin Recht, *Exact matrix completion via convex optimization*, Foundations of Computational mathematics **9** (2009), no. 6, 717–772.

[5] Emmanuel J. Candès and Terence Tao, *The power of convex relaxation: Near-optimal matrix completion*, IEEE Trans. Inf. Theor. **56** (2010), no. 5, 2053–2080.

[6] Stéphane Chrétien and Tianwen Wei, *Von neumann's inequality for tensors*, arXiv preprint arXiv:1502.01616 (2015).

[7] Pierre Comon, Gene Golub, Lek-Heng Lim, and Bernard Mourrain, *Symmetric tensors and symmetric tensor rank*, SIAM Journal on Matrix Analysis and Applications **30** (2008), no. 3, 1254–1279.

[8] Silvia Gandy, Benjamin Recht, and Isao Yamada, *Tensor completion and low-n-rank tensor recovery via convex optimization*, Inverse Problems **27** (2011), no. 2, 025010.

[9] Wolfgang Hackbusch, *Tensor spaces and numerical tensor calculus*, vol. 42, Springer Science & Business Media, 2012.

[10] Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal, *Convex analysis and minimization algorithms i: Fundamentals*, vol. 305, Springer Science & Business Media, 2013.

[11] Shenglong Hu, *Relations of the nuclear norm of a tensor and its matrix flattenings*, Linear Algebra and its Applications **478** (2015), 188–199.

[12] Tamara G. Kolda and Brett W. Bader, *Tensor decompositions and applications*, SIAM Review **51** (2009), no. 3, 455–500.

[13] Joseph M Landsberg, *Tensors: geometry and applications*, vol. 128, American Mathematical Society Providence, RI, USA, 2012.

[14] Lieven De Lathauwer, Bart De Moor, and Joos Vandewalle, *A multilinear singular value decomposition*, SIAM J. Matrix Anal. Appl **21** (2000), 1253–1278.

[15] Guillaume Lecué and Shahar Mendelson, *Regularization and the small-ball method i: sparse recovery*, arXiv preprint arXiv:1601.05584 (2016).

[16] Adrian S Lewis, *The convex analysis of unitarily invariant matrix functions*, Journal of Convex Analysis **2** (1995), no. 1, 173–183.

[17] Cun Mu, Bo Huang, John Wright, and Donald Goldfarb, *Square deal: Lower bounds and improved convex relaxations for tensor recovery*, (2014).

[18] Benjamin Recht, Maryam Fazel, and Pablo A Parrilo, *Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization*, SIAM review **52** (2010), no. 3, 471–501.

[19] Marco Signoretto, Lieven De Lathauwer, and Johan A. K. Suykens, *Nuclear norms for tensors and their use for convex multilinear estimation*, Linear Algebra and Its Applications (2010).

[20] G Alistair Watson, *Characterization of the subdifferential of some matrix norms*, Linear algebra and its applications **170** (1992), 33–45.

[21] Ming Yuan and Cun-Hui Zhang, *On tensor completion via nuclear norm minimization*, Foundations of Computational Mathematics (2014), 1–38.