– Tensors in Quantum Information Theory –

Pierre Botteron[|1|]

Supervisor. Ion Nechita, Chargé de Recherche (CNRS).
Laboratoire de Physique Théorique (LPT), Toulouse, France.

Goal. Our main motivation in this report is to examine some properties of tensors, more particularly their rank, decomposition and norm in a tensor space of the form $\mathbb{C}^{n_1} \otimes \cdots \otimes \mathbb{C}^{n_N}$. If on the one hand we see matrices of $\mathbb{C}^{m \times n}$ as two-dimensional arrays of complex numbers, then on the other hand we may view tensors as a generalization of matrices: in some sense, they are multi-dimensional arrays of complex numbers.

One may wonder whether or not some nice properties satisfied by matrices are preserved in the more general case of tensors, and we will see that it is not always the case, since for instance contrary to matrices, *stricto sensu*, there is no Singular Value Decomposition for tensors. Eventually, we will take an interest in some applications to Quantum Information Theory, showing in particular the surprising result that the so-called *geometric measure of entanglement* is non-additive.

Keywords. Tensor, Rank, Decomposition, Norm, Entanglement, Geometric Measure.

Contents

Introduction

Tensors are one of the most famous generalizations of matrices. They are basic and convenient tools that have been widely used in many fields at the nexus of Mathematics, Physics or Computer Science, and they are applied in numerous ways. Physicists use them as a concise mathematical framework to establish equations and to solve problems in various areas such as mechanics [Ari90], electrodynamics [NYLBSB97] or general relativity [Gre72] among others.

Tensor theory was introduced in 1900 by the mathematician Tullio Levi-Civita in conjonction with his doctoral supervisor Gregorio Ricci-Curbastro in *Méthodes de calcul différentiel absolu et leurs applications* [RLC70] as

[|1|]Last edition: Friday 25$^{\text{th}}$ February, 2022.

part of the *Absolute Differential Calculus*. It was then developed and popularized by Jan Arnoldus Schouten in 1924 [Sch24] at the time when general relativity was booming.

However, as a fly on the ointment, in a same way as generality grows on the one hand, complexity arrives on the other hand with all its ensuing problems. Hillar and Lim summarized all of it in the following sentence:

> *Tensor problems are almost invariably computationally hard.* [HL13]

What is behind the scenes is that simple problems for matrices, such as computing their rank or exhibiting a certain decomposition, turn out to be NP-hard when we take a look at the more general case of tensors. The purpose of the first two sections of this report will be to understand better how difficult it is to generalize tools as simple as rank, norm and decomposition, from matrices to tensors, and we will in particular introduce two tensor norms, the *spectral norm* $||\cdot||_\sigma$ and the *nuclear norm* $||\cdot||_*$, and explore a tensor decomposition approximation to the Singular Value Decomposition known as *Canonical Polyadic Decomposition*.

Finally, in a third and last section, we will try to apply those concepts to Quantum Information Theory. Finding firm roots in Classical Information Theory which was introduced by Claude E. Shannon in 1948 [Sha48], Quantum Information Theory is now a blooming subject, as shown by publication dates distributed in the last two decades for most of them. In particular, we will study the notion of *entanglement* and a tool to measure it: the *geometric measure of entanglement*. As a last result, we will provide a proof of the astonishing result that this measure is not additive.

I. Linear Algebra: Matrices

In this section, we will recall some basic facts related to matrices. In particular, our motivation is to express the notions of rank, decomposition, and norm, so that we may try to generalize them to tensors in the next section. The main reference for this section is the course [Wat11]. See also [Bha97] and [HJ91] for more details about linear algebra.

I.1. Basic Notations

Let us introduce some fundamental notations that will be used throughout this report. We denote by $\mathbb{C}^{m\times n}$ the vector space of matrices with m rows and n columns, and with coefficients in $\mathbb{C}$. Given a matrix $\mathtt{A} := (a_{i,j})_{i,j} \in \mathbb{C}^{m\times n}$, we define its *entry-wise conjugate* $\bar{\mathtt{A}}$, its *transpose* $\mathtt{A}^\intercal$ and its *adjoint* $\mathtt{A}^*$ as follows:

$$\bar{\mathtt{A}} := (\bar{a}_{i,j})_{i,j} \in \mathbb{C}^{m\times n}, \qquad \mathtt{A}^\intercal := (a_{j,i})_{i,j} \in \mathbb{C}^{n\times m} \qquad \text{and} \qquad \mathtt{A}^* := \overline{\mathtt{A}^\intercal} \in \mathbb{C}^{n\times m}.$$

Spectrum and Trace. If there is a non-zero vector $\mathtt{x} \in \mathbb{C}^n$ such that $\mathtt{A}\mathtt{x} = \lambda\mathtt{x}$ for some $\lambda \in \mathbb{C}$, then we call $\mathtt{x}$ an *eigenvector* and λ an *eigenvalue* of $\mathtt{A}$. The *spectrum* of $\mathtt{A}$, denoted $\mathrm{spec}(\mathtt{A})$, is then the tuple $(\lambda_1, \dots, \lambda_n)$ of all the eigenvalues of $\mathtt{A}$, where the λ_i's are not necessarily distinct. In the particular case of a square matrix $\mathtt{A} \in \mathbb{C}^{n\times n}$, the *trace* $\mathrm{Tr}(\mathtt{A})$ is defined as the sum of the diagonal elements, or equivalently as the sum of all the eigenvalues:

$$\mathrm{Tr}(\mathtt{A}) := \sum_{i=1}^n a_{ii} = \sum_{i=1}^n \lambda_i \in \mathbb{C}.$$

Of course, the trace is $\mathbb{C}$-linear:

$$\mathrm{Tr}(\mathtt{A} + \mu\mathtt{B}) = \sum_{i=1}^n \left(a_{ii} + \mu\, b_{ii}\right) = \sum_{i=1}^n a_{ii} + \mu \sum_{i=1}^n b_{ii} = \mathrm{Tr}(\mathtt{A}) + \mu\,\mathrm{Tr}(\mathtt{B}). \tag{1}$$

Normal and Semi-Definite Matrices. We want to introduce two important classes of matrices: normal matrices and positive semi-definite matrices. On the one hand, a *normal* matrix is a matrix $\mathtt{A} \in \mathbb{C}^{m\times n}$ such that $\mathtt{A}\mathtt{A}^* = \mathtt{A}^*\mathtt{A}$, or in other words such that $\mathtt{A}$ and $\mathtt{A}^*$ commute:

$$[\mathtt{A}, \mathtt{A}^*] = 0.$$

On the other hand, the set $\mathrm{Pos}(\mathbb{C}^{m\times m})$ is the set of *positive semi-definite* matrices, *i.e.* square matrices $\mathtt{P} \in \mathbb{C}^{m\times m}$ that can be written as $\mathtt{P} = \mathtt{B}^*\mathtt{B}$ for some $\mathtt{B} \in \mathbb{C}^{m\times m}$. Such matrices are obviously normal,

hermitian, and they only have non-negative eigenvalues. This explains why we often denotes those matrices by simply $\mathtt{P} \geq 0$, while more generally $\mathtt{A} \geq \mathtt{B}$ means that $\mathtt{A} - \mathtt{B}$ is positive semi-definite.

Inner Product. We define on $\mathbb{C}^{m\times n}$ the following *inner product*:

$$\langle \mathtt{A}, \mathtt{B} \rangle := \mathrm{Tr}(\mathtt{A}^*\mathtt{B}) \in \mathbb{C}.$$

Note that in the particular case of vectors $\mathtt{x} := (x_i)_i$ and $\mathtt{y} := (y_i)_i$ in $\mathbb{C}^m \simeq \mathbb{C}^{m\times 1}$, we have the usual inner product of $\mathbb{C}^m$:

$$\langle \mathtt{x}, \mathtt{y} \rangle := \mathrm{Tr}\big(\underbrace{\mathtt{x}^*\mathtt{y}}_{\in\mathbb{C}}\big) = \mathtt{x}^*\mathtt{y} = \begin{bmatrix}\overline{x_1} & \dots & \overline{x_m}\end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = \sum_{i=1}^{m} \overline{x_i}\, y_i.$$

Still in the case of vectors, physicists will prefer rather using the quite convenient notation $\langle \mathtt{x}|\mathtt{y}\rangle := \langle \mathtt{x}|\,|\mathtt{y}\rangle = \langle \mathtt{x}, \mathtt{y}\rangle$ for the inner product, where $\langle \mathtt{x}| := \mathtt{x}^*$ is called *bra* of $\mathtt{x}$ and $|\mathtt{y}\rangle := \mathtt{y}$ is called *ket* of $\mathtt{y}$. This notation is called *Dirac notation*. Naturally, we have the induced norm on $\mathbb{C}^m$ defined by $||\mathtt{x}|| := \sqrt{\langle \mathtt{x}, \mathtt{x}\rangle}$. We say that a family $\{\mathtt{x}_1, \dots, \mathtt{x}_k\}$ of vectors in $\mathbb{C}^m$ is *orthonormal* if:

$$\langle \mathtt{x}_i, \mathtt{x}_j \rangle = \delta_{ij},$$

for any $1 \leq i, j \leq k$. Note that in particular $||\mathtt{x}_i|| = \sqrt{\langle \mathtt{x}_i, \mathtt{x}_i\rangle} = \sqrt{\delta_{ii}} = 1$.

I.2. Matrix Rank

There are many equivalent ways to define the rank of a matrix. One of the most common ways is the following one: the *rank* of a matrix $\mathtt{A} \in \mathbb{C}^{m\times n}$, denoted $\mathrm{rank}(\mathtt{A})$, is defined as the dimension of the range subspace $\mathrm{Ran}(\mathtt{A}) := \big\{\mathtt{A}\mathtt{x} : \mathtt{x} \in \mathbb{C}^n\big\} \subseteq \mathbb{C}^m$ spanned by the columns of $\mathtt{A}$. Recall that we have the *rank-nullity theorem*:

$$\dim\big(\ker(\mathtt{A})\big) + \mathrm{rank}(\mathtt{A}) = \dim\big(\mathbb{C}^n\big) = n.$$

We find in [BFŻ19, Lemma 1] some equivalent definitions of $\mathrm{rank}(\mathtt{A})$:

- It is the minimum $k \in \mathbb{N}$ such that $\mathtt{A} = \sum_{i=1}^{k} \mathtt{x}_i\mathtt{y}_i^*$ for some $\mathtt{x}_i \in \mathbb{C}^m$ and $\mathtt{y}_i \in \mathbb{C}^n$.
- It is the dimension of the subspace spanned by rows of $\mathtt{A}$.
- It is the dimension of a maximal non-zero minor of $\mathtt{A}$.

Matrices of rank 1 are naturally called *rank-one matrices* and they are of the form $\mathtt{A} = \mathtt{x}\mathtt{y}^*$, with non-zero vectors $\mathtt{x} \in \mathbb{C}^m$ and $\mathtt{y} \in \mathbb{C}^n$. More generally, we call *rank-k matrix* a matrix $\mathtt{A}$ such that $\mathrm{rank}(\mathtt{A}) = k$, and we say that a matrix $\mathtt{A} \in \mathbb{C}^{m\times n}$ if *of maximal rank* if $\mathrm{rank}(\mathtt{A}) = \min\{m, n\}$.

Complexity. An easy way to compute the rank of a given matrix in $\mathbb{C}^{m\times n}$ consists in using *Gaussian elimination*: once the matrix has a row echelon form, it suffices to count its non-zero rows. So, using this method, the complexity of computing the rank of a matrix is $O\big(\min(m,n)^2 \max(m,n)\big)$ in exact arithmetic. We will see in the next section that it does not remain the same in the case of tensor rank.

Limit of a Matrix Sequence. We want to show that, given a positive integer k, the limit of a convergent sequence of matrices of rank at most k is at most of rank k. In other words, we want to prove that the set of matrices of rank $\leq k$ is closed. This result is interesting since we will see in the next section that it is no longer true for tensors. We still work in $\mathbb{C}^{m\times n}$ and we assume without loss of generality that $m \leq n$. So we take $k \leq m$, and we even do not have to consider the case $k = m$ since the result would be trivial. We use the trick that matrices of rank at most k are matrices whose minors of order $k+1$ are all null. We have:

$$\begin{aligned}\big\{\mathtt{A} \in \mathbb{C}^{m\times n} : \mathrm{rank}(\mathtt{A}) \leq k\big\} &= \big\{\mathtt{A} \in \mathbb{C}^{m\times n} : \forall I \subseteq [m], \forall J \subseteq [n],\, |I| = |J| = k+1,\, \det X_{I,J} = 0\big\} \\ &= \bigcap_{I,J} \big\{\mathtt{A} \in \mathbb{C}^{m\times n} : \det X_{I,J} = 0\big\}.\end{aligned}$$

The sets in the intersection are closed as being the sets of vanishing points of some polynomial equations, and therefore the set of matrices of rank at most k is closed by finite intersection of closed sets.

I.3. Matrix Decompositions

Matrix decompositions are widely used in various fields of Mathematics, Physics and theoretical Computer Science in order to simplify some computations. In this part, we want to provide two of the most important decompositions of a given matrix: its *spectral decomposition* and its *singular-value decomposition* (SVD).

The Spectral Decomposition. On the one hand, the *spectral decomposition*, also known as *eigendecomposition*, holds for normal matrices. Basically, it consists in decomposing a normal matrix $\mathtt{A}$ into a linear combinaison of self-adjoint rank-one matrices $\mathtt{x}_i\mathtt{x}_i^*$. The coefficients λ_i arise from the spectrum of $\mathtt{A}$, which explains in some sense the name of this decomposition.

Proposition 1 (Spectral Decomposition)**.** Let $\mathtt{A} \in \mathbb{C}^{m\times m}$ be a normal matrix with spectrum $\mathrm{spec}(\mathtt{A}) := \{\lambda_1(\mathtt{A}), \dots, \lambda_m(\mathtt{A})\}$. Then there exists an orthonormal basis $\{\mathtt{x}_1, \dots, \mathtt{x}_m\}$ of $\mathbb{C}^m$ such that:

$$\mathtt{A} = \sum_{i=1}^{m} \lambda_i(\mathtt{A})\, \mathtt{x}_i\mathtt{x}_i^* = \sum_{i=1}^{m} \lambda_i(\mathtt{A})\, |\mathtt{x}_i\rangle\langle \mathtt{x}_i|. \qquad \square$$

With this notation, one may recall that if we consider normal matrices $\mathtt{A}, \mathtt{B} \in \mathbb{C}^{m\times m}$ that commute, then they are diagonalizable in the same basis, *i.e.* there is an orthonormal basis $\{\mathtt{x}_1, \dots, \mathtt{x}_m\}$ of $\mathbb{C}^m$ such that:

$$\mathtt{A} = \sum_{i=1}^{m} \lambda_i(\mathtt{A})\, \mathtt{x}_i\mathtt{x}_i^* \qquad \text{and} \qquad \mathtt{B} = \sum_{i=1}^{m} \lambda_i(\mathtt{B})\, \mathtt{x}_i\mathtt{x}_i^*.$$

Moreover, this decomposition is quite useful for instance to solve a system of linear ordinary differential equations or linear difference equations. Consider for example the difference equation $\mathtt{y}_{t+1} = \mathtt{A}\,\mathtt{y}_t$ where $\mathtt{y}_0 \in \mathbb{C}^m$ is given. Its solution is:

$$\mathtt{y}_t = \mathtt{A}^t\mathtt{y}_0 = \sum_{i=1}^{m} \lambda_i(A)^t\, \mathtt{x}_i\mathtt{x}_i^*\mathtt{y}_0 = \sum_{i=1}^{m} \lambda_i(A)^t\, |\mathtt{x}_i\rangle\langle \mathtt{x}_i|\mathtt{y}_0\rangle.$$

Functional Calculus. An interesting remark is that this decomposition enables us to extend a function $f : \mathbb{C} \to \mathbb{C}$ to a function between normal matrices of $\mathbb{C}^{m\times m}$. Indeed, for any normal matrix $\mathtt{A} \in \mathbb{C}^{m\times m}$, it suffices to define:

$$f(\mathtt{A}) := \sum_{i=1}^{m} f\big(\lambda_i(\mathtt{A})\big)\, \mathtt{x}_i\mathtt{x}_i^*.$$

Naturally, the same idea works in the case of a function f only defined on $\mathrm{spec}(\mathtt{A})$. In particular, we may apply this technique to the square-root function, and we obtain for any positive semi-definite matrix $\mathtt{A} \in \mathrm{Pos}(\mathbb{C}^{m\times m})$:

$$\sqrt{\mathtt{A}} := \sum_{i=1}^{m} \sqrt{\lambda_i(\mathtt{A})}\, \mathtt{x}_i\mathtt{x}_i^*.$$

(This is well-defined since, as we already saw, eigenvalues of a positive semi-definite matrix are non-negative.) This operation will turn out to be especially helpful in the next subsection as it will be involved in the definition of the nuclear norm $||\cdot||_1$. In addition, notice that $\sqrt{\mathtt{A}} \in \mathbb{C}^{m\times m}$ is the only positive semi-definite matrix such that $\sqrt{\mathtt{A}} \times \sqrt{\mathtt{A}} = \mathtt{A}$:

$$\sqrt{\mathtt{A}} \times \sqrt{\mathtt{A}} = \left(\sum_{i=1}^{m} \sqrt{\lambda_i(\mathtt{A})}\, \mathtt{x}_i\mathtt{x}_i^*\right)\left(\sum_{j=1}^{m} \sqrt{\lambda_j(\mathtt{A})}\, \mathtt{x}_j\mathtt{x}_j^*\right) = \sum_{i,j=1}^{m} \sqrt{\lambda_i(\mathtt{A})\lambda_j(\mathtt{A})}\, \mathtt{x}_i \underbrace{\mathtt{x}_i^*\mathtt{x}_j}_{=\delta_{ij}} \mathtt{x}_j^* = \sum_{i=1}^{m} \lambda_i(\mathtt{A})\, \mathtt{x}_i\mathtt{x}_i^* = \mathtt{A}.$$

The Singular-Value Decomposition. On the other hand, the SVD is a bit more general than the above decomposition in the sense that it may be applied to an arbitrary matrix. Note that this decomposition is also called *Schmidt decomposition* by physicists. In some sense, this is the canonical way to make a minimal rank decomposition of a matrix.

Proposition 2 (SVD). For a given matrix $\mathtt{A} \in \mathbb{C}^{m\times n}$ of rank $r := \mathrm{rank}(\mathtt{A})$, there exist real numbers $\sigma_1(\mathtt{A}) \geq \cdots \geq \sigma_r(\mathtt{A}) > 0$ such that:

$$\mathtt{A} = \sum_{i=1}^{r} \sigma_i(\mathtt{A})\, \mathtt{x}_i \mathtt{y}_i^*,$$

where $\{\mathtt{x}_i\}_i$ and $\{\mathtt{y}_i\}_i$ are orthonormal families of $\mathbb{C}^m$ and $\mathbb{C}^n$ respectively (*i.e.* $\mathtt{x}_i^*\mathtt{x}_j = \mathtt{y}_i^*\mathtt{y}_j = \delta_{ij}$). Moreover, the $\sigma_i(\mathtt{A})$'s are uniquely determined (up to permutation) and are called *singular values* of $\mathtt{A}$. □

Note that there is a link between the former two decompositions. Indeed, for any $1 \leq i \leq \mathrm{rank}(\mathtt{A})$, the following equalities show that we have $\sigma_i(\mathtt{A}) = \sqrt{\lambda_i(\mathtt{A}^*\mathtt{A})} = \sqrt{\lambda_i(\mathtt{A}\mathtt{A}^*)}$, or in other words that the $\sigma_i(\mathtt{A})^2$'s are the positive eigenvalues of $\mathtt{A}^*\mathtt{A}$:

$$\mathtt{A}^*\mathtt{A}\,\mathtt{y}_i = \left(\sum_{k=1}^{r} \sigma_k(\mathtt{A})\,\mathtt{y}_k\mathtt{x}_k^*\right)\left(\sum_{j=1}^{r} \sigma_j(\mathtt{A})\,\mathtt{x}_j\mathtt{y}_j^*\right)\mathtt{y}_i = \sum_{k,j=1}^{r} \sigma_k(\mathtt{A})\sigma_j(\mathtt{A})\,\mathtt{y}_k \underbrace{\mathtt{x}_k^*\mathtt{x}_j}_{=\delta_{kj}} \underbrace{\mathtt{y}_j^*\mathtt{y}_i}_{=\delta_{ji}} = \sigma_i(\mathtt{A})^2\,\mathtt{y}_i. \tag{2}$$

I.4. Matrix Norms

There exist several matrix norms, but in this report we will mainly focus only on two of them, the *trace norm* $||\cdot||_1$ and the *operator norm* $||\cdot||_\infty$. Those two different norms are actually particular cases of a more general norm know as *Schatten p-norm* $||\cdot||_p$, so let us introduce it first.

Schatten p-norm. For any $1 \leq p < +\infty$ and $\mathtt{A} \in \mathbb{C}^{m\times n}$, define:

$$||\mathtt{A}||_p := \left[\mathrm{Tr}\left((\mathtt{A}^*\mathtt{A})^{p/2}\right)\right]^{1/p} \qquad \text{and} \qquad ||\mathtt{A}||_\infty := \max\left\{||\mathtt{A}\mathtt{x}|| : \mathtt{x} \in \mathbb{C}^n, ||\mathtt{x}|| = 1\right\},$$

where $||\cdot||_\infty$ coincides with the limit of $||\cdot||_p$ when $p \to +\infty$. Recall that the meaning of $\sqrt{\mathtt{A}\mathtt{A}^*}$ has been explained with the spectral decomposition in the previous subsection. It happens that some very interesting properties hold for this norm (see [Wat11, Subsection 2.3] for more details), such as:

$$||\mathtt{A}\mathtt{B}||_p \leq ||\mathtt{A}||_p\, ||\mathtt{B}||_p \qquad \text{and} \qquad ||\mathtt{A}||_p = ||\mathtt{A}^\intercal||_p = ||\mathtt{A}^*||_p = ||\bar{\mathtt{A}}||_p. \tag{3}$$

Trace and Operator Norms. In particular, the *trace and operator norms* $||\cdot||_1$ and $||\cdot||_\infty$ are respectively the Schatten 1- and ∞-norm, and therefore they satisfy the hereinabove properties. Notice that the operator norm $||\cdot||_\infty$ is in fact the usual operator norm induced by the Euclidean norm $||\cdot||$ on $\mathbb{C}^n$:

$$||\mathtt{A}||_\infty := \max\left\{||\mathtt{A}\mathtt{x}|| : \mathtt{x} \in \mathbb{C}^n, ||\mathtt{x}|| = 1\right\} = \sup_{\mathtt{x}\neq 0} \frac{||\mathtt{A}\mathtt{x}||}{||\mathtt{x}||}.$$

Concerning the trace norm $||\cdot||_1$, it is often referred to with such a name simply because it is of the form:

$$||\mathtt{A}||_1 := \mathrm{Tr}\left(\sqrt{\mathtt{A}^*\mathtt{A}}\right).$$

Moreover, a duality occurs between those two norms for any $\mathtt{A} \in \mathbb{C}^{m\times n}$:

$$||\mathtt{A}||_1 = \max\left\{|\langle \mathtt{B}, \mathtt{A}\rangle| : \mathtt{B} \in \mathbb{C}^{m\times n}, ||\mathtt{B}||_\infty \leq 1\right\} \quad \text{and} \quad ||\mathtt{A}||_\infty = \max\left\{|\langle \mathtt{B}, \mathtt{A}\rangle| : \mathtt{B} \in \mathbb{C}^{m\times n}, ||\mathtt{B}||_1 \leq 1\right\}. \tag{4}$$

Link with SVD. Recall from Proposition 2 that we may decompose any $\mathtt{A} \in \mathbb{C}^{m\times n}$ in SVD as follows:

$$\mathtt{A} = \sum_{i=1}^{r} \sigma_i(A)\, \mathtt{x}_i \mathtt{y}_i^*,$$

where $r := \mathrm{rank}(\mathtt{A})$, $\sigma_1(\mathtt{A}) \geq \cdots \geq \sigma_r(\mathtt{A}) > 0$ are real numbers called *singular values* of $\mathtt{A}$, and $\{\mathtt{x}_i\}_i$ and $\{\mathtt{y}_i\}_i$ are orthogonal families of $\mathbb{C}^m$ and $\mathbb{C}^n$ respectively. Then, we have the following link between the two norms and the singular values.

Proposition 3. For any matrix $\mathtt{A} \in \mathbb{C}^{m\times n}$ of rank r, the following holds:

$$||\mathtt{A}||_1 = \sum_{i=1}^{r} \sigma_i(\mathtt{A}) \qquad \text{and} \qquad ||\mathtt{A}||_\infty = \max_{1\leq i\leq r} \sigma_i(\mathtt{A}) = \sigma_1(\mathtt{A}).$$

Proof. On the one hand, if $n \leq m$, then we obtain the first equality:

$$||\mathtt{A}||_1 \;=\; \mathrm{Tr}\left(\sqrt{\mathtt{A}^*\mathtt{A}}\right) \;=\; \sum_{i=1}^{n} \lambda_i\left(\sqrt{\mathtt{A}^*\mathtt{A}}\right) \overset{\text{by (2)}}{=} \sum_{i=1}^{n} \sigma_i(\mathtt{A}) \;=\; \sum_{i=1}^{r} \sigma_i(\mathtt{A}),$$

where the last equality holds because $\sigma_{r+1}(\mathtt{A}) = \cdots = \sigma_{m\wedge n}(\mathtt{A}) = 0$. Otherwise, if $m \leq n$, then it suffices to use the property seen in (3) to obtain $||\mathtt{A}||_1 = ||\mathtt{A}^*||_1 = \sum_{i=1}^{m} \lambda_i\left(\sqrt{\mathtt{A}\mathtt{A}^*}\right)$, and we conclude similarly as above using again the equality (2).
On the other hand, as $\{\mathtt{x}_i\}_i \subseteq \mathbb{C}^m$ and $\{\mathtt{y}_i\}_i \subseteq \mathbb{C}^n$ are respectively orthogonal families, we may extend them to orthogonal bases $\{\tilde{\mathtt{x}}_i\}_i \subseteq \mathbb{C}^m$ and $\{\tilde{\mathtt{y}}_i\}_i \subseteq \mathbb{C}^n$ respectively. Now, for any $\mathtt{u} := u_1\tilde{\mathtt{y}}_1 + \cdots + u_n\tilde{\mathtt{y}}_n \in \mathbb{C}^n$ such that $||\mathtt{u}|| = 1$, we use the orthogonality property of $\{\tilde{\mathtt{y}}_i\}_i$ to get:

$$\mathtt{A}\mathtt{u} \;=\; \sum_{i=1}^{r} \sigma_i(\mathtt{A})\, \mathtt{x}_i\mathtt{y}_i^*\left(u_1\tilde{\mathtt{y}}_1 + \cdots + u_n\tilde{\mathtt{y}}_n\right) \;=\; \sum_{i=1}^{r} \sigma_i(\mathtt{A})\, u_i\, \mathtt{x}_i \;=\; \sum_{i=1}^{m} \sigma_i(\mathtt{A})\, u_i\, \tilde{\mathtt{x}}_i,$$

where we use the convenient notation $\sigma_{r+1}(\mathtt{A}) = \cdots = \sigma_m(\mathtt{A}) = 0$. Then, taking the norm, we obtain:

$$||\mathtt{A}\mathtt{u}||^2 \;=\; \sum_{i=1}^{m} \underbrace{\sigma_i(\mathtt{A})^2}_{\leq \sigma_1(\mathtt{A})^2}\, u_i^2 \;\leq\; \sigma_1(\mathtt{A})^2 \underbrace{\sum_{i=1}^{m} u_i^2}_{=||\mathtt{u}||^2=1} \;\leq\; \sigma_1(\mathtt{A})^2.$$

So $||\mathtt{A}||_\infty := \max ||\mathtt{A}\mathtt{u}|| \leq \sigma_1(\mathtt{A})$, and we even have equality since $\sigma_1(\mathtt{A})$ is reached by $||\mathtt{A}\mathtt{u}||$ for $\mathtt{u} := (1, 0, \ldots, 0)$ in the basis $\{\tilde{\mathtt{y}}_i\}_i \subseteq \mathbb{C}^n$. □

II. Multilinear Algebra: Tensors

Now, let us try to generalize to tensors what we have just seen for matrices in the former section. The notion of tensor is quite useful in Quantum Physics especially when we desire to describe a few independent subsystems at the same time with a compact notation. In some sense, tensors could be seen as multi-way arrays of data, and they have particular properties.

II.1. Link Between Tensors and Matrices

We consider a tensor space of the form $\mathcal{H} := \mathbb{C}^{n_1} \otimes \cdots \otimes \mathbb{C}^{n_N}$ for some positive integers $n_1, \ldots, n_N$, and we call its elements *N-partite states* as it is the convention in Quantum Mechanics. These so-called *states* are either of the simple form $|\Psi_N\rangle := |\mathtt{x}_1\rangle \otimes \cdots \otimes |\mathtt{x}_N\rangle$ for some $|\mathtt{x}_i\rangle \in \mathbb{C}^{n_i}$, and in that case they are called *product states*, or they are made out of a linear combinaison of product states and called *entangled states*. Recall from Subsection I.1. that $|\mathtt{x}_i\rangle$ is the Dirac notation used by physicists to describe a vector, and $\langle \mathtt{x}_i|$ denotes its dual. Physicists will say that entangled states are a *superposition* of product states (as in the Schröderger's cat experiment, where the cat is surprisingly considered both dead and alive, it is a superposition of the state "dead" and the state "alive"), and they correspond to correlated events. If we see a tensor $|\Psi_N\rangle \in \mathcal{H}$ as an N-way array of complex numbers, we may have access to each of these numbers using a coordinate notation similar to one of matrices: $(\psi_N)_{i_1\ldots i_N} \in \mathbb{C}$, where $1 \leq i_1 \leq n_1$, …, $1 \leq i_N \leq n_N$. In other words, as for matrices, we will sometimes use the coefficient notation as follows:

$$|\Psi_N\rangle = \left((\psi_N)_{i_1\ldots i_N}\right).$$

The link with matrices is the following: when $N = 2$, we are in the *bipartite* case and our space $\mathcal{H} = \mathbb{C}^{n_1} \otimes \mathbb{C}^{n_2}$ consists precisely of all matrices. Indeed, recall that we have the isomorphism $\mathbb{C}^{n_1\times n_2} \simeq \mathbb{C}^{n_1} \otimes (\mathbb{C}^{n_2})^*$. But,

as $\mathbb{C}^{n_2}$ is finite dimensional, we have $(\mathbb{C}^{n_2})^* \simeq \mathbb{C}^{n_2}$ using $\langle \tilde{\mathbf{x}}_i| \mapsto |\tilde{\mathbf{x}}_i\rangle$, where the $|\tilde{\mathbf{x}}_i\rangle$'s form a basis of $\mathbb{C}^{n_2}$. Therefore, we obtain that the space of $n_1 \times n_2$ matrices is well isomorphic to $\mathcal{H}$ when $N = 2$:

$$\mathbb{C}^{n_1 \times n_2} \simeq \mathbb{C}^{n_1} \otimes (\mathbb{C}^{n_2})^* \simeq \mathbb{C}^{n_1} \otimes \mathbb{C}^{n_2} = \mathcal{H}.$$

Hence, our goal in this section is to generalize the results known for $N = 2$ to the *mutlipartite* case $N \geq 3$. The number N will be designated as the *order* of the tensors of $\mathcal{H}$. Thus, scalars are called *zeroth-order* tensors, vectors are called *first-order* tensors, matrices are called *second-order* tensors, and eventually all other tensors are called *higher-order* tensors (*i.e.* when we are in the multipartite case $N \geq 3$).

II.2. Tensor Rank

The notion of rank is quite helpful since it could be seen as a simple way to measure entanglement of a state — recall that we defined *entangled states* in the former subsection as non-trivial linear combinaisons of product states. After defining the notion of tensor rank, we will try to compare some of its aspects with the notion of matrix rank defined in the previous section. The main reference for this subsection will be [BFŻ19, Sections 3 and 4].

Definition. Given an N-partite state $\Psi_N \in \mathcal{H}$, the *rank* $\operatorname{rank}(\Psi_N)$ of Ψ_N may be defined as the minimal integer r such that Ψ_N could be represent as a superposition of r product states:

$$\Psi_N = \sum_{i=1}^{r} |\mathbf{x}_1^{(i)}\rangle \otimes \cdots \otimes |\mathbf{x}_N^{(i)}\rangle,$$

for some $|\mathbf{x}_j^{(i)}\rangle \in \mathbb{C}^{n_j}$, $i = 1, \ldots, r$. This definition of tensor rank is a generalization of one of the equivalent definitions of matrix rank seen in the previous section. Naturally, if $\operatorname{rank}(\Psi_N) = 1$, then it means that Ψ_N is a product state and we call it *rank-one* tensor; this notion will be useful in the next subsection in order to define the *canonical polyadic decomposition* of a tensor. One can find various different approches to tensor rank in [BFŻ19, Section 3].

Some Properties. First, in general, one can easily see that the rank of a product of some tensors is sub-multiplicative:

$$\operatorname{rank}(\Psi_N \otimes \Phi_N) \leq \operatorname{rank}(\Psi_N) \operatorname{rank}(\Phi_N),$$

where $\Psi_N, \Phi_N \in \mathcal{H}$. Indeed, if we write $\Psi_N = \sum_{i=1}^{r} |\mathbf{x}_1^{(i)}\rangle \otimes \cdots \otimes |\mathbf{x}_N^{(i)}\rangle$ and $\Phi_N = \sum_{j=1}^{s} |\mathbf{y}_1^{(j)}\rangle \otimes \cdots \otimes |\mathbf{y}_N^{(j)}\rangle$, then:

$$\Psi_N \otimes \Phi_N = \sum_{i,j} |\mathbf{x}_1^{(i)}\rangle \otimes \cdots \otimes |\mathbf{x}_N^{(i)}\rangle \otimes |\mathbf{y}_1^{(j)}\rangle \otimes \cdots \otimes |\mathbf{y}_N^{(j)}\rangle,$$

where the latter sum consists of $r \cdot s$ summands, and we get the wanted inequality by minimality of the rank. Moreover, note that the rank $\operatorname{rank}(\Psi_N)$ of any tensor is a quantity that can be majored by the product $n_1 \ldots n_N$ since this last is the dimension of $\mathcal{H}$, but we can find a more precise and interesting majoration in [BFŻ19, Theorem 14]:

$$\operatorname{rank}(\Psi_N) \leq \frac{n_1 \ldots n_N}{\max\{n_1, \ldots, n_N\}}.$$

Additionally, we may define the direct sum $\Psi_N \oplus \Phi_N \in \mathcal{H}^2$ of $\Psi_N, \Phi_N \in \mathcal{H}$ as follows:

$$(\psi_N \oplus \phi_N)_{i_1,\ldots,i_N} := (\psi_N)_{i_1,\ldots,i_N} \quad \text{and} \quad (\psi_N \oplus \phi_N)_{n_1+i_1,\ldots,n_N+i_N} := (\phi_N)_{i_1,\ldots,i_N},$$

for $1 \leq i_k \leq n_k$, and we set all other entries to 0. In some sense, the "first" coefficients of $\Psi_N \oplus \Phi_N$ are those of Ψ_N, and the "last" coefficients are those of Φ_N. We obviously have the following inequality:

$$\operatorname{rank}(\Psi_N \oplus \Phi_N) \leq \operatorname{rank}(\Psi_N) + \operatorname{rank}(\Phi_N),$$

and we have equality in the case of matrices. Strassen wondered in 1973 whether equality holds or not for more general tensors in the so-called *Strassen's direct sum conjecture* (see [Str73]). Actually, the answer is that equality is satisfied in some particular cases, but Shitov showed in 2017 that it is not true in general (see [Shi17]) contrary to the case of matrices.

NP-Hard to Compute. We have seen in the former section that, in the case of matrices where $N = 2$, the complexity of computing the rank of a matrix is $O\big(\min(m,n)^2 \max(m,n)\big)$ in exact arithmetic. Nevertheless, in the multipartite case where $N \geq 3$, the calculation of the rank of a tensor is much harder. It is actually even NP-hard in most cases (see [Hå90, HL13] for more details). In particular, find in [HL13, Table I] a list of NP-hard problems related to tensors. This last reference also provides an idea why it is NP-hard: "3-tensor problems form a boundary separating classes of tractable linear/convex problems from intractable non-linear/non-convex ones". Thereby, contrary to matrices, there is no general and trivial algorithm that computes tensor rank.

A Limit Case. Now, let us show a fascinating result: the set of rank-2 tensors is not closed, in the sense that we can exhibit tensor $|\Phi\rangle$ of rank 3 which can be written as the limit of some tensors of rank 2. Consider the tripartite case $N = 3$, and fix all the dimensions n_1, n_2 and n_3 of the parties to 2. In other words, we are in the case $\mathcal{H} = \mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2 = (\mathbb{C}^2)^{\otimes 3}$. Define the following state of $\mathcal{H}$:

$$|\Phi\rangle := |121\rangle + |211\rangle + |112\rangle,$$

where $|\mathtt{xyz}\rangle$ is a convenient shortcut for $|\mathtt{x}\rangle \otimes |\mathtt{y}\rangle \otimes |\mathtt{z}\rangle \in \mathcal{H}$ for some $|\mathtt{x}\rangle$, $|\mathtt{y}\rangle$ and $|\mathtt{z}\rangle$ in $\mathbb{C}^2$, and where $|1\rangle := |\tilde{\mathtt{x}}_1\rangle$ and $|2\rangle := |\tilde{\mathtt{x}}_2\rangle$ form an orthonormal basis of $\mathbb{C}^2$. Consider any tensor $|\Psi_3\rangle$ in $\mathcal{H}$. We want to introduce the notion of *slice*. In the same way as we can view a matrix either as an array of numbers or as a list of vectors, we may view a tensor either as a multi-way array or as a list of slices. In our case, the tensor $|\Psi_3\rangle$ of $\mathcal{H}$ could be seen not only as a $2 \times 2 \times 2$ "cube" of data, but also as a list of two slices. Indeed, fixing the last coordinate of $|\Psi_3\rangle$ to $|1\rangle$, we get the "top" slice $\mathtt{A}_1$, and fixing its last coordinate to $|2\rangle$, we obtain the "bottom" slice $\mathtt{A}_2$. Note that those two slices are actually 2×2 matrices, and they are called *frontal slices*. Now, we can write $|\Psi_3\rangle$ as a sum of those two slices:

$$|\Psi_3\rangle = \mathtt{A}_1 \otimes |1\rangle + \mathtt{A}_2 \otimes |2\rangle$$

This definition of slice is used in the following proposition, which will tell us that $|\Phi\rangle$ is of rank 3.

Proposition 4. [BFŻ19, Lemma 6]. Let $|\Psi_3\rangle \in \mathcal{H}$ a tensor where $\mathcal{H} = \mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2$, and denote $\mathtt{A}_1$ and $\mathtt{A}_2$ its two frontal slices. Then $\operatorname{rank}(|\Psi_3\rangle) = 3$ if, and only if, the matrices $\mathtt{A}_1$ and $\mathtt{A}_2$ are linearly independent and $\operatorname{span}(\mathtt{A}_1, \mathtt{A}_2)$ contains two matrices $\mathtt{A}$ and $\mathtt{B}$ such that $\mathtt{A}$ is invertible and $\mathtt{A}^{-1}\mathtt{B}$ is not diagonalizable. □

In our case, we notice that:

$$|\Phi\rangle = \Big(|12\rangle + |21\rangle\Big) \otimes |1\rangle + |11\rangle \otimes |2\rangle = \mathtt{A}_1 \otimes |1\rangle + \mathtt{A}_2 \otimes |2\rangle, \quad \text{where:} \quad \mathtt{A}_1 := \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \mathtt{A}_2 := \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

Now, we see that $\mathtt{A}_1$ and $\mathtt{A}_2$ are well linearly independent, that $\mathtt{A}_1$ is invertible with itself as inverse, and that $\mathtt{A}_1^{-1}\mathtt{A}_2 = \mathtt{A}_1\mathtt{A}_2 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$ is not diagonalizable. Therefore, by Proposition 4, we deduce that:

$$\operatorname{rank}\big(|\Phi\rangle\big) = 3.$$

However, the state $|\Phi\rangle$ can be written as a limit of tensors of rank 2:

$$|\Phi\rangle = \lim_{t \to 0} \frac{1}{t}\Big(\big(|1\rangle + t|2\rangle\big)^{\otimes 3} - |111\rangle\Big),$$

because:

$$\begin{aligned} \big(|1\rangle + t|2\rangle\big)^{\otimes 3} &= \big(|1\rangle + t|2\rangle\big) \otimes \big(|1\rangle + t|2\rangle\big) \otimes \big(|1\rangle + t|2\rangle\big) \\ &= |111\rangle + t\Big(\underbrace{|211\rangle + |121\rangle + |112\rangle}_{=|\Phi\rangle}\Big) + t^2\Big(|122\rangle + |212\rangle + |221\rangle\Big) + t^3|222\rangle. \end{aligned}$$

Hence, we showed that the set of rank-2 tensors is not "closed" contrary to the case of matrices (see the former section).

II.3. Tensor Norms

We would like to introduce two tensor norms known as *spectral norm* $||\cdot||_\sigma$ and *nuclear norm* $||\cdot||_*$. Note that, in the literature, these are also known in Banach theory as respectively *injective* and *projective* tensor norms. They have many applications to quantum information theory, *e.g.* the *geometric measure of entanglement* that we will study in the next section, and they also induce the notion of *nuclear rank* that, unlike the tensor rank, is lower semi-continuous. After defining these two tensor norms, we will see a link with the two matrice norms $||\cdot||_1$ and $||\cdot||_\infty$ that we defined in the previous section. Finally, even though it is NP-hard to compute these norms, we will briefly mention some alternating approaches in order to compute them. Our main references for this subsection will be [FL16, DFLW17, Nie17].

Spectral and Nuclear Norms. A norm $||\cdot||$ on $\mathcal{H} := \mathbb{C}^{n_1} \otimes \cdots \otimes \mathbb{C}^{n_N}$ is a *tensor norm* if for any product state $\mathtt{x} = \mathtt{x}_1 \otimes \cdots \otimes \mathtt{x}_N \in \mathcal{H}$ we have $||\mathtt{x}|| = ||\mathtt{x}_1||_{\mathbb{C}^{n_1}} \dots ||\mathtt{x}_N||_{\mathbb{C}^{n_N}}$. For instance, given a tensor $\Psi_N \in \mathcal{H}$, we respectively define the *spectral norm* $||\cdot||_\sigma$ and the *nuclear norm* $||\cdot||_*$ as follows:

$$||\Psi_N||_\sigma := \max\Big\{\big|\langle \Psi_N,\, \mathtt{x}_1 \otimes \cdots \otimes \mathtt{x}_N\rangle\big| : ||\mathtt{x}_j|| = 1,\, \mathtt{x}_j \in \mathbb{C}^{n_j}\Big\},$$
$$||\Psi_N||_* := \min\Big\{\textstyle\sum_{i=1}^r |\lambda_i| : \Psi_N = \sum_{i=1}^r \lambda_i\, \mathtt{x}_1^{(i)} \otimes \cdots \otimes \mathtt{x}_N^{(i)},\, ||\mathtt{x}_j^{(i)}|| = 1,\, \mathtt{x}_j^{(i)} \in \mathbb{C}^{n_j},\, \lambda_i \in \mathbb{C},\, r \in \mathbb{N}\Big\},$$

where $\langle\cdot,\cdot\rangle$ is the usual scalar product of tensors, defined as follows given two tensors $\Psi_N, \Phi_N \in \mathcal{H}$ of same order:

$$\langle \Psi_N, \Phi_N\rangle := \sum_{i_1=1}^{n_1}\sum_{i_2=1}^{n_2}\cdots\sum_{i_N=1}^{n_N} (\psi_N)_{i_1\dots i_N}\overline{(\phi_N)}_{i_1\dots i_N} \in \mathbb{C}. \tag{5}$$

Proposition 5. In the bipartite case $N = 2$, we have the following equalities for any matrix $\mathtt{A} \in \mathbb{C}^{m\times n}$ of maximal rank:

$$||\mathtt{A}||_\sigma = \max_{1\le i\le \mathrm{rank}(\mathtt{A})} \big(\sigma_i(\mathtt{A})\big) = \sigma_1(\mathtt{A}). \tag{6}$$

$$||\mathtt{A}||_* = \sum_{i=1}^{\mathrm{rank}(\mathtt{A})} \sigma_i(\mathtt{A}), \tag{7}$$

where the $\sigma_i(\mathtt{A})$'s are the singular values of $\mathtt{A}$ and $\mathrm{rank}(\mathtt{A}) = \min\{n, m\}$ is the rank of $\mathtt{A}$.

Proof. Recall from Proposition 2 that the matrix $\mathtt{A} \in \mathbb{C}^{m\times n}$ may be written under the following SVD form:

$$\mathtt{A} = \sum_{i=1}^{\mathrm{rank}(\mathtt{A})} \sigma_i(A)\, \mathtt{u}_i \mathtt{v}_i^*, \qquad \text{where:} \qquad \langle \mathtt{u}_i, \mathtt{u}_j\rangle = \langle \mathtt{v}_i, \mathtt{v}_j\rangle = \delta_{ij} \text{ for any } 1 \le i, j \le r.$$

(6) First, notice that, in our bipartite case $N = 2$, the spectral norm $||\cdot||_\sigma$ is reduced to:

$$||\mathtt{A}||_\sigma := \max\Big\{\big|\langle \mathtt{A}, \mathtt{x}\mathtt{y}^*\rangle\big| : ||\mathtt{x}|| = ||\mathtt{y}|| = 1, \mathtt{x} \in \mathbb{C}^m, \mathtt{y} \in \mathbb{C}^n\Big\} = \max_{||\mathtt{x}||=||\mathtt{y}||=1} \big|\langle \mathtt{A}\mathtt{y}, \mathtt{x}\rangle\big|.$$

If we take $\mathtt{x} := \mathtt{u}_1 \in \mathbb{C}^m$ and $\mathtt{y} := \mathtt{v}_1 \in \mathbb{C}^n$ from the SVD of $\mathtt{A}$, then we well have $||\mathtt{x}|| = ||\mathtt{y}|| = 1$ and:

$$\big|\langle \mathtt{A}\mathtt{y}, \mathtt{x}\rangle\big| = \big|\mathtt{x}^*\mathtt{A}\mathtt{y}\big| = \Big|\mathtt{u}_1^*\Big(\textstyle\sum_{i=1}^{\mathrm{rank}(\mathtt{A})} \sigma_i(\mathtt{A})\, \mathtt{u}_i\, \mathtt{v}_i^*\Big)\mathtt{v}_1\Big| = \Big|\sum_{i=1}^{\mathrm{rank}(\mathtt{A})} \sigma_i(\mathtt{A})\, \underbrace{\langle \mathtt{u}_1, \mathtt{u}_i\rangle}_{=\delta_{1i}}\, \underbrace{\langle \mathtt{v}_i, \mathtt{v}_1\rangle}_{=\delta_{1i}}\Big| = |\sigma_1(\mathtt{A})| = \sigma_1(\mathtt{A}).$$

By maximality of $||\cdot||_\sigma$, we deduce that $\sigma_1(\mathtt{A}) \le ||\mathtt{A}||_\sigma$.
Conversely, consider any $\mathtt{x} \in \mathbb{C}^m$ and $\mathtt{y} \in \mathbb{C}^n$ satisfying $||\mathtt{x}|| = ||\mathtt{y}|| = 1$, and let us show that $\big|\langle \mathtt{x}, \mathtt{A}\mathtt{y}\rangle\big| \le \sigma_1(\mathtt{A})$. We have:

$$|\langle \mathtt{x}, \mathtt{A}\mathtt{y}\rangle|^2 = \Big|\Big\langle \mathtt{x},\, \Big(\textstyle\sum_{i=1}^{\mathrm{rank}(\mathtt{A})} \sigma_i(\mathtt{A})\, \mathtt{u}_i\, \mathtt{v}_i^*\Big)\mathtt{y}\Big\rangle\Big|^2 \le \Big|\sum_{i=1}^{\mathrm{rank}(\mathtt{A})} \underbrace{\sigma_i(\mathtt{A})}_{\le \sigma_1(\mathtt{A})} \langle \mathtt{x}, \mathtt{u}_i\rangle \langle \mathtt{v}_i, \mathtt{y}\rangle\Big|^2$$
$$\le \sigma_1(\mathtt{A})^2 \textstyle\sum_{i=1}^{\mathrm{rank}(\mathtt{A})} |\langle \mathtt{x}, \mathtt{u}_i\rangle|^2\, |\langle \mathtt{v}_i, \mathtt{y}\rangle|^2 \le \sigma_1(\mathtt{A})^2 \underbrace{\Big(\textstyle\sum_{i=1}^{\mathrm{rank}(\mathtt{A})} |\langle \mathtt{x}, \mathtt{u}_i\rangle|^2\Big)}_{\overset{(*)}{\le} ||\mathtt{x}||^2=1} \underbrace{\Big(\textstyle\sum_{j=1}^{\mathrm{rank}(\mathtt{A})} |\langle \mathtt{v}_j, \mathtt{y}\rangle|^2\Big)}_{\overset{(*)}{\le} ||\mathtt{y}||^2=1} \le \sigma_1(\mathtt{A})^2,$$

where the inequalities $(*)$ hold because, if we denote $(\mathtt{x}_1, \ldots, \mathtt{x}_m)$ the coordinates of x in a basis of $\mathbb{C}^m$ made out of $\mathtt{u}_1, \ldots, \mathtt{u}_r$ and some other vectors, then we have $\sum_{i=1}^r |\langle \mathtt{x}, \mathtt{u}_i\rangle|^2 = \sum_{i=1}^{\mathrm{rank}(\mathtt{A})} |\mathtt{x}_i|^2 \leq \sum_{i=1}^m |\mathtt{x}_i|^2 = ||\mathtt{x}||^2$. Hence, passing to the maximum, we get $||\mathtt{A}||_\sigma \leq \sigma_1(\mathtt{A})$ and we obtain equation (6).

(7) Next, still in the bipartite case $N = 2$, the nuclear norm $||\cdot||_*$ is reduced to:

$$||\mathtt{A}||_* := \min\Big\{ \textstyle\sum_{i=1}^r |\lambda_i| : \mathtt{A} = \sum_{i=1}^r \lambda_i\, \mathtt{x}_i\mathtt{y}_i^*,\ ||\mathtt{x}_i|| = ||\mathtt{y}_i|| = 1,\ \mathtt{x}_i \in \mathbb{C}^m,\ \mathtt{y}_i \in \mathbb{C}^n,\ \lambda_i \in \mathbb{C},\ r \in \mathbb{N}\Big\}.$$

With the notations of the above SVD, we set:

$$\mathtt{x}_i := \mathtt{u}_i \in \mathbb{C}^m, \quad \mathtt{y}_i := \mathtt{v}_i \in \mathbb{C}^n, \quad \lambda_i := \sigma_i(\mathtt{A}) \in \mathbb{C}, \quad r := \mathrm{rank}(\mathtt{A}) \in \mathbb{N},$$

and we well have $||\mathtt{x}_i|| = ||\mathtt{y}_i|| = 1$ and $\mathtt{A} = \sum_{i=1}^r \lambda_i\, \mathtt{x}_i\mathtt{y}_i^*$. So, by minimality of $||\cdot||_*$, we get:

$$||\mathtt{A}||_* \ \leq\ \textstyle\sum_{i=1}^{\mathrm{rank}(\mathtt{A})} |\lambda_i(\mathtt{A})| \ =\ \sum_{i=1}^{\mathrm{rank}(\mathtt{A})} \sigma_i(\mathtt{A}).$$

Reciprocally, write any decomposition $\mathtt{A} = \sum_{i=1}^r \lambda_i\, \mathtt{x}_i\mathtt{y}_i^*$ as in the above definition of $||\cdot||_*$, and show that $\sum_{i=1}^{\mathrm{rank}(\mathtt{A})} \sigma_i(\mathtt{A}) \leq \sum_{i=1}^r |\lambda_i|$. Without loss of generality, we may assume that $m \leq n$, which means that $\mathrm{rank}(\mathtt{A}) = m$. As stated in [HJ91, Theorem 3.4.1], we have:

$$\sum_{i=1}^{\mathrm{rank}(\mathtt{A})} \sigma_i(\mathtt{A}) = \max\Big\{ |\operatorname{Tr}(\mathtt{AC})| : \mathtt{C} \in \mathbb{C}^{n\times m} \text{ is a partial isometry of rank } \mathrm{rank}(\mathtt{A})\Big\}.$$

Recall that a matrix $\mathtt{C} \in \mathbb{C}^{n\times m}$ is said to be a partial isometry if $||\mathtt{Cx}|| \leq ||\mathtt{x}||$ for all $\mathtt{x} \in \mathbb{C}^m$. Hence, we may restate our goal as follows:

$$\sum_{i=1}^{\mathrm{rank}(\mathtt{A})} \sigma_i(\mathtt{A}) = \max_{\mathtt{C}} |\operatorname{Tr}(\mathtt{AC})| \overset{\text{(goal)}}{\leq} \sum_{i=1}^r |\lambda_i|.$$

For any partial isometry $\mathtt{C} \in \mathbb{C}^{m\times n}$, notice that $\operatorname{Tr}(\mathtt{x}_i\mathtt{y}_i^*\,\mathtt{C}) = \operatorname{Tr}(\mathtt{y}_i^*\mathtt{Cx}_i) = \operatorname{Tr}(\langle \mathtt{y}_i, \mathtt{Cx}_i\rangle) = \langle \mathtt{y}_i, \mathtt{Cx}_i\rangle$ because it is a scalar, so we have:

$$\begin{aligned} |\operatorname{Tr}(\mathtt{AC})| \ &=\ \Big|\operatorname{Tr}\Big(\textstyle\sum_{i=1}^r \lambda_i\, \mathtt{x}_i\mathtt{y}_i^*\,\mathtt{C}\Big)\Big| \ \overset{(1)}{=}\ \Big|\sum_{i=1}^r \lambda_i\, \operatorname{Tr}(\mathtt{x}_i\mathtt{y}_i^*\,\mathtt{C})\Big| \\ &\leq\ \sum_{i=1}^r |\lambda_i|\, \big|\langle \mathtt{y}_i, \mathtt{Cx}_i\rangle\big| \ \overset{\text{C.-S.}}{\leq}\ \sum_{i=1}^r |\lambda_i|\, \underbrace{||\mathtt{y}_i||}_{=1}\, \underbrace{||\mathtt{Cx}_i||}_{\leq ||\mathtt{x}_i|| = 1} \ \leq\ \sum_{i=1}^r |\lambda_i|. \end{aligned}$$

Thus, taking the maximum of $|\operatorname{Tr}(\mathtt{AC})|$ over $\mathtt{C}$ and the minimum of $\sum_{i=1}^r |\lambda_i|$, we obtain the other inequality $\sum_{i=1}^{\mathrm{rank}(\mathtt{A})} \sigma_i(\mathtt{A}) \leq ||\mathtt{A}||_*$, and equation (7) holds. □

Link with Matrix Norms. Now, recall that we saw in the previous section the definition of the *trace norm* $||\cdot||_1$ and the *operator norm* $||\cdot||_\infty$, introduced as follows for any matrix $\mathtt{A} \in \mathbb{C}^{m\times n}$:

$$||\mathtt{A}||_1 := \operatorname{Tr}\big(\sqrt{\mathtt{A}^*\mathtt{A}}\big) \qquad \text{and} \qquad ||\mathtt{A}||_\infty := \max\Big\{ ||\mathtt{Ax}|| : \mathtt{x} \in \mathbb{C}^n, ||\mathtt{x}|| = 1\Big\}.$$

Using Proposition 3 and Proposition 5, we get the following corollary which states that these two matrix norms are particular cases of tensor norms.

Corollary 6. When $N = 2$, the spectral and nuclear norms coincide respectively with the operator and trace norms for any matrix $\mathtt{A} \in \mathbb{C}^{m\times n}$ of maximal rank:

$$||\mathtt{A}||_\sigma = ||\mathtt{A}||_\infty \qquad \text{and} \qquad ||\mathtt{A}||_* = ||\mathtt{A}||_1. \qquad \square$$

Moreover, it is known that we also have a duality between $||\cdot||_\sigma$ and $||\cdot||_*$, as for $||\cdot||_1$ and $||\cdot||_\infty$ in (4), meaning that for any tensor $\Psi \in \mathcal{H}$:

$$||\Psi||_\sigma = \max\Big\{ \big|\langle \Psi, \Phi\rangle\big| : \Phi \in \mathcal{H},\ ||\Phi||_* = 1\Big\} \quad \text{and} \quad ||\Psi||_* = \max\Big\{ \big|\langle \Psi, \Phi\rangle\big| : \Phi \in \mathcal{H},\ ||\Phi||_\sigma = 1\Big\}.$$

NP-Hard to Compute. As explained in [FL16], even if we consider a case as simple as $N = 4$, the computation of the spectral norm $||\cdot||_\sigma$ and the nuclear norm $||\cdot||_*$ is NP-hard. Actually, it is even the case when we restrict tensors to be bi-Hermitian, bi-symmetric, positive semi-definite, non-negative valued, or all the above. However, we find in [DFLW17, Sections 9 and 10] an alternating method for computing the nuclear norm of non-symmetric tensors in a first time, and then of symmetric tensors, where a tensor is said *symmetric* if it remains unchanged after permuting two of its indices. In addition, one may take a look at [Nie17] for a reverse approach, focusing first on symmetric tensors and based on the fact that the nuclear norm of symmetric tensors can be achieved with a symmetric decomposition, and then extending results to non-symmetric tensors in its last section.

II.4. Tensor Decomposition

We know from Proposition 1 and Proposition 2 that there are two very useful decompositions for matrices, the so-called *spectral decomposition* and *singular value decomposition* (SVD), and we may wonder if there exists any generalization of them to tensors. While the first one would not make sense for tensors since it requires normal matrices and we do not have a canonical notion of tensor transposition so *a fortiori* neither a notion of normal tensor, the latter would worth the try for tensors. However, as we will see, strictly speaking, such an SVD for tensors cannot be defined. Nevertheless, we will then provide a tensor decomposition that try to keep some aspects of the SVD, namely the *canonical polyadic decomposition* (CPD). Our main references for this subsection are [KB09] and [RSG17].

No SVD for Tensors. As a generalization of the SVD to general tensors, the *higher-order singular value decomposition* (HOSVD) would decompose a tensor $\Psi_N \in \mathcal{H}$ as follows:

$$\Psi_N = \sum_{i=1}^{\operatorname{rank}(\Psi_N)} \sigma_i \, \mathtt{x}_1^{(i)} \otimes \cdots \otimes \mathtt{x}_N^{(i)}, \tag{8}$$

where each of the families $\{\mathtt{x}_1^{(i)}\}_i, \ldots, \{\mathtt{x}_N^{(i)}\}_i$ would be respectively orthonormal, and where the σ_i's would be (positive) real numbers. Our goal is to show that such an HOSVD cannot stand.

The Case of Matrices. In order to understand why such an HOSVD cannot hold, we will begin by seeing a reason why SVD is possible in the case $N = 2$ of matrices. Pick a matrix $\mathtt{A} \in \mathbb{C}^{m\times n}$ and assume without loss of generality that $m \leq n$. It is quite natural to consider that the rank of $\mathtt{A}$ is m, *i.e.* that $\mathtt{A}$ is of maximal rank, since if $\mathtt{A}$ is randomly chosen in $\mathbb{C}^{m\times n}$ — say with a Gaussian distribution applied on each coordinate, then $\operatorname{rank}(\mathtt{A}) = \min\{m, n\} = m$ almost surely (see [BFŻ19, Subsection 3.1]). In addition, this assumption will deeply simplify our calculations. We want to compare the number of real parameter carried by $\mathtt{A}$ with those carried by its SVD.

Proposition 7. If $N = 2$, with the settings as above, we have:

$$\#\text{ real parameters carried by } \mathtt{A} \;=\; \#\text{ real parameters carried by SVD.}$$

Proof. On the one hand, as $\mathtt{A}$ lies in $\mathbb{C}^{m\times n}$, it carries the following number of real parameters:

$$\#\text{ real parameters carried by } \mathtt{A} = 2nm,$$

because $\mathtt{A}$ consists of nm complex coefficients, and each of them is composed with a real part and an imaginary part. On the other hand, its SVD is of the form:

$$\sum_{i=1}^{m} \sigma_i \, \mathtt{x}_i \mathtt{y}_i^*,$$

where the σ_i's are positive real numbers, and where $\{\mathtt{x}_i\}_i$ and $\{\mathtt{y}_i\}_i$ are orthogonal families of $\mathbb{C}^m$ and $\mathbb{C}^n$ respectively. First, the σ_i's provide m real parameters. Secondly, let us show that the family $\{\mathtt{y}_i\}_i \subseteq \mathbb{C}^n$

carries $2nm - m^2$ real parameters. These vectors are normalized, so $\mathtt{y}_1 \in \mathbb{C}^n$ carries $2n-1$ real parameters. Then, as $\mathtt{y}_2$ is orthogonal to $\mathtt{y}_1$, we have following two conditions:

$$\Re\big(\langle \mathtt{y}_1, \mathtt{y}_2\rangle\big) = 0 \qquad \text{and} \qquad \Im\big(\langle \mathtt{y}_1, \mathtt{y}_2\rangle\big) = 0,$$

and it yields that $\mathtt{y}_2$ carries $2n-1-2$ parameters. Similarly, for any $1 \leq k \leq m$, using the facts that $\mathtt{y}_k$ is normalized and that it is orthogonal to each of $\mathtt{y}_1, \dots, \mathtt{y}_{k-1}$, we get that $\mathtt{y}_k$ carries $2n-(2k-1)$ real parameters. In result, the family $\{\mathtt{y}_i\}_i$ carries the following number of real parameters:

$$\#\text{ real parameters in } \{\mathtt{y}_i\}_i = \sum_{k=1}^{m} \big(2n-(2k-1)\big) = 2nm - \underbrace{\sum_{k=1}^{m}(2k-1)}_{=m^2 \text{ by induction}} = 2nm - m^2.$$

Thirdly, one may similarly show that the family $\{\mathtt{x}_i\}_i \subseteq \mathbb{C}^m$ carries m^2 real parameters. Finally, there is a last condition, we have not to take into consideration the phase of the $\mathtt{x}_i$'s since we can reach any wanted phase for $\mathtt{x}_i\mathtt{y}_i^*$ by simply changing the phase of $\mathtt{y}_i$. So we need to remove m real parameters. At the end of the day, the number of real parameters for the SVD is the following:

$$\#\text{ real parameters carried by SVD} = m + \big(2nm - m^2\big) + m^2 - m = 2nm. \qquad \square$$

Hence, we have the same number of real parameters contained in $\mathtt{A}$ and the SVD, which explains why it could work in the bipartite case $N = 2$.

The General case. Now, let us consider the case of higher-order tensors, *i.e.* when $N \geq 3$. Let a tensor $\Psi_N \in \mathcal{H} := \mathbb{C}^{n_1} \otimes \cdots \otimes \mathbb{C}^{n_N}$ and assume without loss of generality that $n_1 = \min\big\{n_k : 1 \leq k \leq N\big\}$. For the sake of convenience, we will suppose as in the above case of matrices that the rank $\operatorname{rank}(\Psi_N)$ of Ψ_N is equal to n_1. Again, the strategy consists in comparing the number of real parameters carried by Ψ_N and by the HOSVD.

Proposition 8. If $N \geq 3$, with the settings as above, we have:

$$\#\text{ real parameters carried by } \Psi_N \;>>\; \#\text{ real parameters carried by HOSVD.} \tag{9}$$

Proof. In order to prove it, our main arguments will be similar to the ones in the proof of Proposition 7, so all the details will not be provided again. On the one hand, we observe that $\Psi_N \in \mathbb{C}^{n_1} \otimes \cdots \otimes \mathbb{C}^{n_N}$ carries the following number of real parameters:

$$\#\text{ real parameters carried by } \Psi_N \;=\; 2\,n_1 \times \cdots \times n_N.$$

On the other hand, recall from (8) that HOSVD is of the following form:

$$\Psi_N = \sum_{i=1}^{n_1} \sigma_i\, \mathtt{x}_1^{(i)} \otimes \cdots \otimes \mathtt{x}_N^{(i)}.$$

So, without considering the relations of phase cancelation between parties, the number of real parameters carried by HOSVD can be majored by:

$$\begin{aligned}
\#\text{ real param. carried by HOSVD} \;&\leq\; \underbrace{n_1}_{\text{due to the } \sigma_i\text{'s}} + \underbrace{n_1^2}_{\text{due to } \{\mathtt{x}_1^{(i)}\}_i} + \underbrace{\big(2\,n_1n_2 - n_1^2\big)}_{\text{due to } \{\mathtt{x}_2^{(i)}\}_i} + \cdots + \underbrace{\big(2\,n_1n_N - n_1^2\big)}_{\text{due to } \{\mathtt{x}_N^{(i)}\}_i} \\
&\leq\; n_1^2 + n_1^2 + \big(2n_2^2 - n_1^2\big) + \cdots + \big(2n_N^2 - n_1^2\big) \\
&=\; 2\sum_{k=1}^{N} n_k^2 - (N-1)\,n_1^2.
\end{aligned}$$

Thereby, we obtain the wanted inequality (9) when the dimensions n_i are big enough — otherwise, use the relations of phase cancelation between parties as in the case of matrices and get the same inequality. $\square$

Thus, HOSVD would have been a great tool to simplify some computations with tensors, but we fall upon the disappointing fact that it cannot exist in the multipartite case $N \geq 3$. However, one may try to approach

it in order to keep some of the interesting properties of the SVD, and this is the motivation of the following decomposition.

The Canonical Polyadic Decomposition. In principle, the *canonical polyadic decomposition* (CPD) is meant to approximate a tensor Ψ_N with a sum of finitely many rank-one tensors. This notion was first proposed by Hitchcock in 1927 [Hit27], who called this decomposition the *polyadic form* of a tensor; but it needed to wait until the 70's to get known, under the name CANDECOMP (canonical decomposition) [CC70] and PARAFAC (parallel factors) [Har70]. All these decompositions will be designated here by *canonical polyadic decomposition* (CPD) as in our main references [KB09, RSG17].

Definition. Given a tensor $\Psi_N \in \mathcal{H}$, the CPD consists in finding a tensor $\hat{\Psi}_N \in \mathcal{H}$ that minimizes the quantity:

$$||\Psi_N - \hat{\Psi}_N||, \qquad \text{where } \hat{\Psi}_N \text{ is of the form:} \qquad \hat{\Psi}_N := \sum_{i=1}^{R} \lambda_r \, \mathtt{x}_1^{(i)} \otimes \cdots \otimes \mathtt{x}_N^{(i)}, \tag{10}$$

for some positive integer R, some real numbers λ_j and some normalized vectors $\mathtt{x}_j^{(i)} \in \mathbb{C}^{n_j}$, for $j = 1, \dots, N$ and $i = 1, \dots, R$. The norm $||\cdot||$ used here is the norm induced by the scalar product $\langle \cdot, \cdot \rangle$ defined in (5). Such a decomposition always exists, but we may wonder if it is unique. In fact, even though rank decompositions are generally not unique in the case of matrices, it is often the case for higher-order tensors, as we can see in [RSG17, 4.1.4] and in [KB09, 3.2] which provide us with a sufficient and necessary condition for uniqueness.

Algorithms. In order to compute such a decomposition, one may use an *alternating least square* algorithm, which is a method proposed in the two papers of the 70's [CC70, Har70] that we cited above. The main idea of this algorithm results in minimizing the the norm $||\Psi_N - \hat{\Psi}_N||$ while fixing all the vectors $\mathtt{x}_2^{(i)}, \dots, \mathtt{x}_N^{(i)}$ except the $\mathtt{x}_1^{(i)}$'s, for all $1 \leq i \leq R$, and then repeating this step for the following vectors until we reach some stopping criterion. An issue with this algorithm is that it might be slow to converge, and that it might not even converge toward a global minimum depending on initial conditions. This is why we have other algorithms, *e.g.* the *Jenrich's algorithm* or the *tensor power method*, where both are more efficient than the alternating least square algorithm, but they are only available on particular cases of tensors. Actually, there is no perfect procedure for fitting the CPD for a given number of components. The difficulty comes from the fact that finding the rank of a tensor is NP-hard, as seen in the former subsection, so it is difficult to chose a relevant parameter R, and data is usually corrupted by some kind of noise. Find more details in [RSG17, 4.1] and [KB09, 3.4].

Link with HOSVD. One may notice that the form of $\hat{\Psi}_N$ in (10) is very similar to the HOSVD in (8). In the case of a "perfect HOSVD", we would have $||\Psi_N - \hat{\Psi}_N|| = 0$ and $R = \text{rank}(\Psi_N)$, but it is not always the case since we previously saw that a general HOSVD cannot hold. In that sense, the CPD is an approximation of the HOSVD, and it allows us to apply all the good properties of the HOSVD to the approximation $\hat{\Psi}_N$ instead of applying them directly to Ψ_N.

Applications. The canonical polyadic decomposition finds many applications, first in psychometrics in the 70's [CC70, Har70], but then also as far as numerical analysis, data mining, neuroscience, and beyond. Refer to [KB09, 3.5] for an exhaustive list of applications of the CPD. Find also in [RSG17, Sections 5 and 6] some detailed applications of tensor decomposition to machine learning and estimation of mixture models.

III. Application to Quantum Information: the Entanglement

Our goal here is to provide an example of how tensor products can apply to Quantum Information Theory. More precisely, we will study the *geometric measure*, which is, in some sense, the *spectral norm* $||\cdot||_\sigma$ that we studied in the former section, and we will show the surprising fact that it is not additive. But first, let us introduce some of the basic notions of quantum information theory.

III.1. Quantum Information Formalism

Classical Information Theory was first introduced in 1948 by Shannon [Sha48] and it established the mathematical basis of communication. Here is a quotation:

> *The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point.*

It then gave birth to Quantum Information Theory, where the transmission and processing of information are made by quantum systems, and which is nowadays in wide expansion. It appeals to both Mathematics and Mathematical Physics, with for instance Group Theory, Probability Theory or also Quantum Statistical Physics, and finds many applications as in Cryptography. The monographs [Wat11, Tim04, Pre98] contain more details.

Basic Notions. As before, we work in a tensor space of the form $\mathcal{H} := \mathbb{C}^{n_1} \otimes \cdots \otimes \mathbb{C}^{n_N}$, for some positive integers $n_1, \dots, n_N$, whose elements are called *(N-partite) states*. We will indifferently employ the Dirac notation $|\Psi\rangle$ or simply Ψ to designate any state in $\mathcal{H}$, and we will refer to operators from $\mathcal{H}$ to $\mathcal{H}$ by greek letters as ρ or σ. We will often put an N as indice to Ψ_N or ρ_N as to recall that we are in the N-partite case. Recall that the bipartite case $N = 2$ corresponds to the case of matrices, and that we are in the *multipartite* case when $N \geq 3$. We denote by $\langle\Psi| \in \mathcal{H}^*$ the dual of $|\Psi\rangle$ defined by the scalar product over $\mathcal{H}$ in (5). This enables us to define $\langle\Psi|\rho|\Psi\rangle$, which is simply the composition of the linear form $\langle\Psi|$ with the operator ρ and with the state $|\Psi\rangle$. It will be helpful for computations to notice the following fact:

$$\langle\Psi \otimes \Phi|\rho \otimes \sigma|\Psi \otimes \Phi\rangle \;=\; \langle\Psi|\rho|\Psi\rangle \times \langle\Phi|\sigma|\Phi\rangle.$$

Pure States. A vector $|\Psi\rangle \in \mathcal{H}$ is said to be a *pure quantum state* when it has norm 1. More generally, one considers in quantum physics *mixed states* ρ, which are positive semi-definite operators of unit trace acting on $\mathcal{H}$. As a particular case of mixed states, we have *rank-one projections*, which are of the form $\rho = |\Psi\rangle\langle\Psi|$, for Ψ a unit vector from $\mathcal{H}$. Note that these states Ψ are also called *pure states*. Without any precision, we adopt the convention that a state is thought to be pure, as it is the case for physicists. Notice that when $\rho = |\Psi\rangle\langle\Psi|$, we have:

$$\langle\Phi|\rho|\Phi\rangle \;=\; \langle\Phi|\Psi\rangle\langle\Psi|\Phi\rangle \;=\; \big|\langle\Phi, \Psi\rangle\big|^2.$$

In addition, notice that we have a bijective correspondance between pure operators and states:

$$\rho = |\Psi\rangle\langle\Psi| : \mathcal{H} \to \mathcal{H} \qquad \leftrightarrow \qquad |\Psi\rangle \in \mathcal{H}.$$

Product States. A state $\Psi \in \mathcal{H}$ is called *product state* or *separable state* if it is of rank 1, *i.e.* if it is of the form $\Psi = \mathtt{x}_1 \otimes \cdots \otimes \mathtt{x}_N$ for some vectors $\mathtt{x}_i \in \mathbb{C}^{n_i}$, $i = 1, \dots, N$. Otherwise, Ψ is a linear combinaison of product states, and is called *entangled state*. We define in a similar way a *product* operator or a *separable* operator.

Entanglement. As defined hereinabove, a state $\Psi \in \mathcal{H}$ is entangled if it is not a product state, or in other words if $\mathrm{rank}(\Psi) \geq 2$. In that sense, the rank could be seen a tool that measures entanglement. But we will study in the next subsection a more commonly used tool to do it, namely the *geometric measure*. For physicists, entangled states corresponds to correlated events, whereas quantum product states can be compared to probability vectors of the form $p(a, b) = p_1(a)p_2(b)$, with independent variables. In other words, in the case of a product state, a joint physical system AB could be divided into two disjoint subsystems A and B.

On the contrary, quantum entanglement occurs when two particles share some properties concerning their position, momentum or polarization that can not have happened by chance. In fact, what is surprising is that if we know something about one of these characteristics for one particle, then we are able to deduce something about the same characteristic for the other particle. A common comparaison is maid with a pair of gloves. Imagine we only find the left glove in our drawer, then we know that the missing glove is the right glove. Even more than that, one might say that the two gloves are entangled since knowing something about the first one tells us something about the second one. This topic is one the main ones that separates Classical Physics from Quantum Physics since entanglement is a basic feature of Quantum Mechanics missing in Classical Mechanics.

III.2. Entanglement Measure

In this last subsection, we will introduce an entanglement measure known as *geometric measure* (GM). Recently, it has been applied to various parts of Quantum Computation, as for instance in order to show that most entangled states are too entangled to be useful as computational resources [GFE09]. Recall that a state $|\Psi\rangle$ is a *product state* if it is of the form $|\Psi\rangle = \mathtt{x}_1 \otimes \dots \mathtt{x}_N$ for some $\mathtt{x}_i \in \mathbb{C}^{n_i}$, otherwise $|\Psi\rangle$ is a sum of such states and is said to be *entangled*. In some sense, the geometric measure is meant to describe "how much" a state $|\Psi\rangle$ is entangled, or in other words how much $|\Psi\rangle$ is "far" from a product state. After defining the geometric measure, we will see a link between this measure and the spectral norm $||\cdot||_\sigma$, and eventually we will land on the surprising result that the GM is non-additive. Our main references here will be [ZCH10, WG03].

The Geometric Measure (GM). Pick an N-partite operator ρ acting on the Hilbert space $\mathcal{H} = \mathbb{C}^{n_1} \otimes \dots \otimes \mathbb{C}^{n_N}$. The *geometric measure of entanglement* (GM) measures the closest distance between ρ and the set of separable operators SEP, or equivalently the set of pure product states PRO:

$$G(\rho) := -2\log\big(\Lambda(\rho)\big) \qquad \text{where} \qquad \Lambda^2(\rho) := \max_{\sigma \in SEP} \mathrm{Tr}(\rho\sigma) = \max_{|\Psi\rangle \in PRO} \langle\Psi|\rho|\Psi\rangle,$$

where the logarithm log has base 2. Notice that if $\rho = |\Phi\rangle\langle\Phi|$ is a pure state, then:

$$\Lambda^2\big(|\Phi\rangle\langle\Phi|\big) = \max_{|\Psi\rangle \in PRO} |\langle\Psi, \Phi\rangle|^2.$$

Any pure product state $|\Psi\rangle$ maximizing $\Lambda^2(\rho)$ is a *closest product state* to ρ. There exist alternative ways of defining GM, for instance through a convex roof construction [WG03], which could explain why we give such a name to this measure.

GM of a Tensor Product. Given two N-partite operators ρ and σ from $\mathcal{H} = \mathbb{C}^{n_1} \otimes \dots \otimes \mathbb{C}^{n_N}$ to itself, the tensor product $\rho \otimes \sigma$ is a $2N$-partite operator that goes from $\mathcal{H} \otimes \mathcal{H}$ to itself. However, here, for the sake of simplification, we will assume that the first partite of ρ and the first one σ are located in the same lab, in other words we identify them together so that they may be viewed as a single partite. We apply the same trick for the second partite of ρ and the second one of σ, and similarly so on and so forth for all of their parties. At the end of the day, the tensor product $\rho \otimes \sigma$ may be seen as an N-partite operator, which means that we are allowed to compute its geometric measure $G(\rho \otimes \sigma)$. This will be useful in order to define the additivity property of the geometric measure.

Additivity. Given an operator $\rho : \mathcal{H} \to \mathcal{H}$, the *asymptotic geometric measure* $G^\infty(\rho)$ (AGM) is the geometric measure applied to the tensor product $\rho^{\otimes n}$ of a large number of copies of ρ:

$$G^\infty(\rho) := \lim_{n\to\infty} \frac{1}{n} G\big(\rho^{\otimes n}\big).$$

Based on this definition, we say that the geometric measure G is *additive* if $G^\infty(\rho) = G(\rho)$, and it is *strong additive* if $G(\rho \otimes \sigma) = G(\rho) + G(\sigma)$ for any $\rho, \sigma : \mathcal{H} \to \mathcal{H}$. Obviously, strong additivity implies additivity, hence its name. We will see in Proposition 10 that GM is additive for bipartite pure states, but we will see in Theorem 13 that it is no longer the case in the multipartite antisymmetric case. However, we need to keep in mind that it is quite difficult to show that the geometric measure G is symmetric or not given a state, or to compute the asymptotic geometric measure G^∞.

Proposition 9. For any operators ρ_N and ρ'_N, we have that GM is sub-additive:

$$G(\rho_N \otimes \rho'_N) \leq G(\rho_N) + G(\rho'_N).$$

Proof. We have:

$$\begin{aligned}
\Lambda^2(\rho_N \otimes \rho'_N) &= \max_{|\Phi\rangle} \langle\Phi|\rho_N \otimes \rho'_N|\Phi\rangle \\
&\geq \max_{|\Phi\rangle \text{ of the form } |\varphi\rangle\otimes|\varphi'\rangle} \langle\Phi|\rho_N \otimes \rho'_N|\Phi\rangle \\
&= \max_{|\varphi\rangle} \langle\varphi|\rho_N|\varphi\rangle \times \max_{|\varphi'\rangle} \langle\varphi'|\rho'_N|\varphi'\rangle \\
&= \Lambda^2(\rho_N) \times \Lambda^2(\rho'_N).
\end{aligned}$$

Hence the result using the equality $G(\rho) := -2\log\big(\Lambda(\rho)\big)$. □

Remarks. There are two other famous entanglement measures, known as *relative entropy of entanglement* (REE) and *logarithmic global robustness* (LGR). But under some theoretical conditions, Hayashi *et al* showed in [ZCH10, Proposition 3] that there is an equality relation between GM, REE and LGR which enables us to treat the additivity problem of REE and LGR through the additivity of GM.

Link between GM and $||\cdot||_\sigma$. Recall from Subsection II.3. that we introduced the *spectral norm* $||\cdot||_\sigma$ as a particular case of Schatten norm. Now, in the pure case where ρ_N is of the form $\rho_N = |\Psi\rangle\langle\Psi|$ for some state $|\Psi\rangle \in \mathcal{H}$, there is a link between the geometric measure G and the spectral norm $||\cdot||_\sigma$. Indeed, we have:

$$\Lambda^2(\rho_N) \;=\; \max_{||\mathtt{x}_i||=1} \big\langle \mathtt{x}_1\otimes\cdots\otimes\mathtt{x}_N|\rho_N|\mathtt{x}_1\otimes\cdots\otimes\mathtt{x}_N\big\rangle \;=\; \max_{||\mathtt{x}_i||=1} \big|\big\langle \mathtt{x}_1\otimes\cdots\otimes\mathtt{x}_N|\Psi\big\rangle\big|^2 \;=\; ||\Psi||_\sigma^2.$$

So $G(\rho_N) = -2\log\big(||\Psi||_\sigma\big)$. Note that there is also a link between GM and the *nuclear norm* $||\cdot||_*$: they attain their maximal for the same states [DFLW17]. From the link between GM and $||\cdot||_\sigma$, one can deduce that GM is strong additive for pure bipartite states as stated in the following proposition.

Proposition 10. For $N = 2$, we may associate any two pure operators $\rho = |\Psi\rangle\langle\Psi|$ and $\sigma = |\Phi\rangle\langle\Phi|$ to matrices $\mathtt{A}$ and $\mathtt{B}$ respectively. Then GM is strong additive, *i.e.* :

$$G(\mathtt{A}\otimes\mathtt{B}) = G(\mathtt{A}) + G(\mathtt{B}).$$

Proof. Recall from Proposition 5 that, when $N = 2$, the spectral norm $||\cdot||_\sigma$ is simply of the form $||\mathtt{A}||_\sigma = \sigma_1(\mathtt{A})$, *i.e.* it is the biggest singular value (note that the quoted proposition requires that $\mathtt{A}$ has maximal rank, but this part of the proposition does not need this assumption). One may see that σ_1 is multiplicative in the sense that $\sigma_1(\mathtt{A}\otimes\mathtt{B}) = \sigma_1(\mathtt{A})\times\sigma_1(\mathtt{B})$ because the singular values of a tensor product are defined as the singular values of the factors. Therefore, we have:

$$\Lambda^2(\mathtt{A}\otimes\mathtt{B}) \;=\; ||\mathtt{A}\otimes\mathtt{B}||_\sigma^2 \;=\; \sigma_1(\mathtt{A}\otimes\mathtt{B})^2 \;=\; \sigma_1(\mathtt{A})^2\times\sigma_1(\mathtt{B})^2 \;=\; ||\mathtt{A}||_\sigma^2\times||\mathtt{B}||_\sigma^2 \;=\; \Lambda^2(\mathtt{A})\times\Lambda^2(\mathtt{B}),$$

which implies the wanted result using $G(\rho) := -2\log\big(\Lambda(\rho)\big)$. □

Non-additivity of GM. The non-additivity of GM is a quite astonishing fact since GM is additive in many cases of both practical and theoretical interest; see for instance [ZCH10, Theorem 5] that states that GM is additive for non-negative multipartite states. In other words, before wondering about the non-additivity of GM, no one would intuitively call it into question. Indeed, historically, since GM was defined in 1995 [SHI95], it has been conjectured to be additive, until the surprising discovery of Werner and Holevo in 2002 [WH02]. They discovered that GM is non-additive in the case of antisymmetric multipartite states, which is what we will aim to show hereinbelow. See [ZCH10, Section 5] for a statistical approach that shows that almost all multipartite pure states have non-additive GM.

Symmetric and Antisymmetric States. A pure product state $|\Psi_N\rangle = \mathtt{x}_1\otimes\cdots\otimes\mathtt{x}_N \in \mathcal{H}$ is said to be *symmetric* if it remains unchanged after permuting its parties with an odd permutation σ:

$$|\Psi_N\rangle = \mathtt{x}_1\otimes\cdots\otimes\mathtt{x}_N = \mathtt{x}_{\sigma(1)}\otimes\cdots\otimes\mathtt{x}_{\sigma(N)}.$$

Similarly, we say that a pure product state $|\Psi_N\rangle = \mathtt{x}_1\otimes\cdots\otimes\mathtt{x}_N \in \mathcal{H}$ is *antisymmetric* if an odd permutation of its parties induces a sign change. Now, these notions of symmetric and antisymmetric states could be generalized to any pure state by linear combinaison of pure product states, and then more generally to any state ρ_N by convex combinaison of pure states. We define in a similar way *symmetric* and *antisymmetric* operators.

Proposition 11. [ZCH10, Proposition 4]. Given a multipartite symmetric operator ρ, any closest product state $|\Psi\rangle$ to ρ (*i.e.* a pure product state $|\Psi\rangle$ maximizing $\Lambda^2(\rho)$) is necessarily also symmetric. Similarly for antisymmetric states. □

Proposition 12. [ZCH10, Proposition 16]. Given an N-partite antisymmetric state $\rho_N : \mathcal{H} \to \mathcal{H}$, and given a closest product state $|\Psi_N\rangle = \mathtt{x}_1 \otimes \cdots \otimes \mathtt{x}_N$ to ρ, the set $\{\mathtt{x}_1, \ldots, \mathtt{x}_N\}$ of is orthogonal. □

Theorem 13 (Non-additivity of GM)**.** [ZCH10, Theorem 17]. When $N \geq 3$, GM is non-additive for antisymmetric operators. More precisely, given any two N-partite antisymmetric operators ρ_N and ρ'_N from $\mathcal{H}$ to itself, we have:

$$G(\rho_N \otimes \rho'_N) < G(\rho_N) + G(\rho'_N).$$

Proof. Recall from the paragraph *GM of a Tensor Product* at the beginning of the current subsection that the parties of ρ_N and ρ'_N are identified so that we may view the tensor product $\rho_N \otimes \rho'_N$ not only as a $2N$-partite operator, but also as an N-partite operator. Now, as ρ_N and ρ'_N are antisymmetric, the tensor product $\rho_N \otimes \rho'_N$ is necessarily symmetric because an odd permutation of the parties changes the signs of both ρ_N and ρ'_N, thus $\rho_N \otimes \rho'_N$ remains unchanged after permuting the parties. Therefore, using Proposition 11, as $N \geq 3$, any closest product state $|\Psi\rangle$ to $\rho_N \otimes \rho'_N$ needs also to be symmetric.
Now, let us assume that we may write $|\Psi\rangle = |\Phi_N\rangle \otimes |\Phi'_N\rangle$ for some closest product states $|\Phi_N\rangle$ and $|\Phi'_N\rangle$ to respectively ρ_N and ρ'_N, and let us show that this leads to a contradiction. Write $|\Phi_N\rangle = \mathtt{y}_1 \otimes \cdots \otimes \mathtt{y}_N$ and $|\Phi'_N\rangle = \mathtt{y}'_1 \otimes \cdots \otimes \mathtt{y}'_N$. Since ρ_N and ρ'_N are antisymmetric, we have by Proposition 12 that the sets $\{\mathtt{y}_1, \ldots, \mathtt{y}_N\}$ and $\{\mathtt{y}'_1, \ldots, \mathtt{y}'_N\}$ are respectively orthogonal. We have:

$$\begin{aligned}
|\Psi\rangle &= |\Phi_N\rangle \otimes |\Phi'_N\rangle = (\mathtt{y}_1 \otimes \cdots \otimes \mathtt{y}_N) \otimes (\mathtt{y}'_1 \otimes \cdots \otimes \mathtt{y}'_N) \\
&= (\mathtt{y}_1 \otimes \mathtt{y}'_1) \otimes (\mathtt{y}_2 \otimes \mathtt{y}'_2) \otimes (\mathtt{y}_3 \otimes \mathtt{y}'_3) \otimes \cdots \otimes (\mathtt{y}_N \otimes \mathtt{y}'_N) \\
&= (\mathtt{y}_2 \otimes \mathtt{y}'_2) \otimes (\mathtt{y}_1 \otimes \mathtt{y}'_1) \otimes (\mathtt{y}_3 \otimes \mathtt{y}'_3) \otimes \cdots \otimes (\mathtt{y}_N \otimes \mathtt{y}'_N) \qquad \text{because } |\Psi\rangle \text{ is symmetric}
\end{aligned}$$

It follows that $\mathtt{y}_1 \otimes \mathtt{y}'_1 = \mathtt{y}_2 \otimes \mathtt{y}'_2$ up to a phase, and therefore $\mathtt{y}_1 = \mathtt{y}_2$ up to a phase, which contradicts the fact that $\{\mathtt{y}_1, \ldots, \mathtt{y}_N\}$ is orthogonal.
Hence, there is no closest product state $|\Psi\rangle$ to $\rho_N \otimes \rho'_N$ that could be written as a tensor product of closest product states to ρ_N and ρ'_N respectively, which implies:

$$\begin{aligned}
\Lambda^2(\rho_N \otimes \rho'_N) &= \langle\Psi|\rho_N \otimes \rho'_N|\Psi\rangle = \max_{|\varphi\rangle} \langle\varphi|\rho_N \otimes \rho'_N|\varphi\rangle \\
&> \max_{|\varphi\rangle \text{ of the form } |\varphi_N\rangle \otimes |\varphi'_N\rangle} \langle\varphi|\rho_N \otimes \rho'_N|\varphi\rangle \\
&= \max_{|\varphi_N\rangle} \langle\varphi_N|\rho_N|\varphi_N\rangle \times \max_{|\varphi'_N\rangle} \langle\varphi'_N|\rho'_N|\varphi'_N\rangle \\
&= \langle\Phi_N|\rho_N|\Phi_N\rangle \times \langle\Phi'_N|\rho'_N|\Phi'_N\rangle = \Lambda^2(\rho_N) \times \Lambda^2(\rho'_N).
\end{aligned}$$

Whence, we obtain the wanted inequality $G(\rho_N \otimes \rho'_N) < G(\rho_N) + G(\rho'_N)$. □

Conclusions

As a short conclusion, let us recap all that we have seen. First of all, we introduced the rank $\operatorname{rank}(\mathtt{A})$ of a matrix $\mathtt{A} \in \mathbb{C}^{m \times n}$ as the smallest integer r such that $\mathtt{A} = \sum_{i=1}^{r} \mathtt{x}_i \mathtt{y}_i^*$ for some vectors $\mathtt{x}_i$ and $\mathtt{y}_i$, and we used a similar definition when we generalized it to tensors. As opposed to tensors whose rank is NP-hard to compute, we saw that the complexity of computing $\operatorname{rank}(\mathtt{A})$ is $O\big(\min(m,n)^2 \max(m,n)\big)$ using Gaussian elimination, and we also proved that the set of matrices of rank at most k is closed, contrary to tensors for which we exhibited a counter-example.

In addition, we analyzed two matrix decompositions, the spectral decomposition and the Singular Value Decomposition. The first decomposition was in particular helpful in order to define objects as $\sqrt{\mathtt{A}}$. We tried to generalize them to tensors, but both of these generalizations failed; that is the reason why we thought about the Canonical Polyadic Decomposition as a convenient approximation.

Moreover, we studied two particular cases of the Schatten norms: the trace norm $||\cdot||_1$ and the operator norm $||\cdot||_\infty$. We generalized these norms so that we may get tensor norms, and we obtained the nuclear

norm $||\cdot||_*$ and the spectral norm $||\cdot||_\sigma$. We also viewed some links with the SVD and, even though those norms are NP-hard to compute, we proposed a few alternating methods of computations.

Last but not least, we focused on Quantum Information Theory, trying to apply some results about tensors in it. After introducing the basis of this new formalism, we mainly worked on the geometric measure of entanglement. In particular, we saw that this measure is always sub-additive, that it is additive in the bipartite pure case, and eventually that it is surprisingly non-additive in the multi-partite antisymmetric case.

References

[**Ari90**] R. Aris. *Vectors, Tensors and the Basic Equations of Fluid Mechanics*. Dover Books on Mathematics. Dover Publications, 1990.

[**BFŻ19**] Wojciech Bruzda, Shmuel Friedland, and Karol Życzkowski. Tensor rank and entanglement of pure quantum states. *arXiv e-prints*, December 2019. arXiv: 1912.06854.

[**Bha97**] Rajendra Bhatia. *Matrix Analysis*. Graduate Texts in Mathematics. Springer-Verlag New York, 1997. DOI: 10.1007/978-1-4612-0653-8.

[**CC70**] J. Douglas Carroll and Jih-Jie Chang. Analysis of individual differences in multidimensional scaling via an n-way generalization of "eckart-young" decomposition. *Psychometrica*, 35(3):283–319, September 1970. http://cda.psych.uiuc.edu/psychometrika_highly_cited_articles/carroll_chang_1970.pdf.

[**DFLW17**] Harm Derksen, Shmuel Friedland, Lek-Heng Lim, and Li Wang. Theoretical and computational aspects of entanglement. *arXiv e-prints*, May 2017. arXiv: 1705.07160.

[**FL16**] Shmuel Friedland and Lek-Heng Lim. Nuclear Norm of Higher-Order Tensors. *arXiv e-prints*, May 2016. arXiv: 1410.6072.

[**GFE09**] David Gross, Steven T. Flammia, and Jens Eisert. Most quantum states are too entangled to be useful as computational resources. *Phys. Rev. Lett.*, 102, May 2009. DOI: 10.1103/PhysRevLett.102.190501.

[**Gre72**] Phillip J. Greenberg. The algebra of the riemann curvature tensor in general relativity: Preliminaries. *Studies in Applied Mathematics*, 51(3):277–308, 1972. DOI: 10.1002/sapm1972513277.

[**Har70**] R. A. Harshman. Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis. *UCLA Working Papers in Phonetics*, 16:1–84, 1970.

[**Hit27**] Frank L. Hitchcock. The expression of a tensor or a polyadic as a sum of products. *Journal of Mathematics and Physics*, 6(1-4):164–189, 1927. DOI: 10.1002/sapm192761164.

[**HJ91**] Roger A. Horn and Charlie R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, USA, 1991. DOI: 10.1017/CBO9780511840371.

[**HL13**] Christopher Hillar and Lek-Heng Lim. Most tensor problems are NP-hard. *arXiv e-prints*, July 2013. arXiv: 0911.1393.

[**Hå90**] Johan Håstad. Tensor rank is NP-complete. *Journal of Algorithms*, pages 644–654, 1990. DOI: 10.1016/0196-6774(90)90014-6.

[**JHK+08**] Eylee Jung, Mi-Ra Hwang, Hungsoo Kim, Min-Soo Kim, Daekil Park, Jin-Woo Son, and Sayatnova Tamaryan. Reduced state uniquely defines the Groverian measure of the original pure state. *Phys. Rev*, 77(6):062317, June 2008. arXiv: 0709.4292.

[**KB09**] Tamara G. Kolda and Brett W. Bader. Tensor decompositions and applications. *SIAM*, pages 455–500, August 2009. DOI: 10.1137/07070111X.

[**Nie17**] Jiawang Nie. Symmetric tensor nuclear norms. *SIAM*, pages 599–625, August 2017. DOI: 10.1137/16M1083384.

[**NYLBSB97**] H. N. Núñez-Yépez, J. L. López-Bonilla, and A. L. Salas-Brito. On the radiative part of the maxwell tensor for a liénard - wiechert field. *Journal of Physics A: Mathematical and General*, 30(10):3663–3669, may 1997. DOI: 10.1088/0305-4470/30/10/036.

[**Pre98**] John Preskill. *Quantum Information Theory*, volume Chap 5. California Institute of Technologie, 1998. Lecture notes, http://www.theory.caltech.edu/people/preskill/ph229/notes/chap5.pdf.

[**RLC70**] M. M. G. Ricci and T. Levi Civita. Méthodes de calcul différentiel absolu et leurs applications, January 1970. DOI: 10.1007/bf01454201.

[**RSG17**] Stephan Rabanser, Oleksandr Shchur, and Stephan Günnemann. Introduction to Tensor Decompositions and their Applications in Machine Learning. *arXiv e-prints*, November 2017. arXiv: 1711.10781.

[**Sch24**] Jan Arnoldus Schouten. *Der Ricci-Kalkül – Eine Einführung in die neueren Methoden und Probleme der mehrdimensionalen Differentialgeometrie (Ricci Calculus – An introduction in the latest methods and problems in multi-dimensional differential geometry)*. Springer, Berlin, 1924. 10.1007/978-3-662-06545-7.

[**Sha48**] Claude E. Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.*, 27(3):379–423, October 1948. http://people.math.harvard.edu/~ctm/home/text/others/shannon/entropy/entropy.pdf.

[**SHI95**] ABNER SHIMONY. Degree of entanglement. *Annals of the New York Academy of Sciences*, 755(1):675–679, 1995. DOI: 10.1111/j.1749-6632.1995.tb39008.x.

[**Shi17**] Yaroslav Shitov. A counterexample to Strassen's direct sum conjecture. *arXiv e-prints*, December 2017. arXiv: 1712.08660.

[**Str73**] Volker Strassen. Vermeidung von divisionen. *Journal für die reine und angewandte Mathematik*, 264:184–202, 1973. http://eudml.org/doc/151394.

[**Tim04**] Christopher Gordon Timpson. Quantum Information Theory and the Foundations of Quantum Mechanics. *arXiv e-prints*, December 2004. arXiv: quant-ph/0412063.

[**Wat11**] John Watrous. *Theory of Quantum Information*, volume CS 766/QIC 820. Institute for Quantum Computing, University of Waterloo, 2011. Lecture notes, https://cs.uwaterloo.ca/~watrous/LectureNotes/CS766.Fall2011/all.pdf.

[**WG03**] Tzu-Chieh Wei and Paul M. Goldbart. Geometric measure of entanglement and applications to bipartite and multipartite quantum states. *Phys. Rev. A*, 68:042307, Oct 2003. DOI: 10.1103/PhysRevA.68.042307.

[**WH02**] R. F. Werner and A. S. Holevo. Counterexample to an additivity conjecture for output purity of quantum channels. *Journal of Mathematical Physics*, 43(9):4353–4357, September 2002. arXiv: quant-ph/0203003.

[**ZCH10**] Huangjun Zhu, Lin Chen, and Masahito Hayashi. Additivity and non-additivity of multipartite entanglement measures. *New Journal of Physics*, 12(8):083002, August 2010. arXiv: 1002.2511.