RECH201 : Quantum Majorization

Maxence Jauberty

January 18, 2024

Contents

1 Introduction

The main purpose of this document is to provide the understanding of the current knowledge on the majorization, an order on vectors and matrices, in quantum information theory. We may introduce together quantum and classical information notions to enlighten the analogies used to apprehend the difficult physical ressource of quantum information. Historically, the notion of information has often been bound to an obscur notion called *entropy*. In 1948, when Shannon [15] introduced measures of randomness, entropy is therefore introduced as a notion of information and a physical resource. Before Shannon, Von Neumman had already introduced the so-called Von Neumman's entropy as a measure of *"how mixed a state is ?"*.

The order of majorization, independently, have been introduced by Muirhead to find answers for sum inequalities [9] . The work of Polya and Hardy has then illustrated the link between this order and the *disorder*. Therefore, one can interpret the relation $x \prec y$ (x is majorized by y) as x is more mixed than y, that is x is closer than y to the uniform distribution. This interpretation has lead to new considerations of a general framework for classical information, and for its quantum extension the Hermitian majorization. In Rioul [12], Fano-type inequalities and Pinsker-type inequalities (with respect to the uniform distribution) are derived using majorization relation.

I have studied the use of majorization and several extensions such as d-majorization in order to find entropic inequalities and inequalities on divergences in quantum information and classical information such as Pinsker-type inequalities. The main goal of this paper is then to present the framework of quantum information and then the majorization and its generalizations. I have also studied independently the so-called Quantum Fano Inequality.

2 Quantum Information

2.1 Quantum States

While classical information theory relies on random variables and probability distributions, it is well-known that quantum theory cannot be described entirely by this formalism. Thus, accordingly to the first postulate of quantum mechanics quantum states are described by unit vectors of an Hilbert space. In quantum information theory however, states are mainly described by statistical mixing of these quantum mechanics states. That is, considering $|\psi_1\rangle, ..., |\psi_n\rangle$, the coupling $(|\psi_1\rangle, p_1), ..., (|\psi_n\rangle, p_n)$ where p is a probability distribution. These can be entirely described by density operators.

Definition 1 (Density Operators)**.** *Let $\mathcal{H} = \mathbb{C}^d$ be an Hilbert space. A density operator ρ on $\mathcal{H}$ is a positive semi-definite matrix that verifies $Tr(\rho) = 1$.*

Denote $\mathcal{D}(\mathcal{H})$ the set of density operators on $\mathcal{H}$.

We can explicitely define a quantum state in information theory, using density operators.

Definition 2. *Let $\mathcal{H}_A$ be an Hilbert space associated with a quantum system A, a quantum state, or state, is a density operator ρ_A on $\mathcal{H}_A$.*

The following definitions and properties enlighten how density operators are used in quantum information theory.

Definition 3 (Pure / Mixed States)**.** *Let $\rho \in \mathcal{D}(\mathcal{H})$ a quantum state. ρ is said to be pure if its rank is equal to one. Otherwise, it is said to be a mixed state.*

The notion of purity can be interpreted as a notion of certainty. There might be randomness on measurement, according to Born's rule, however if the state is pure, we are sure of which state the quantum system is in.

Proposition 1. *Let $\mathcal{H}$ be an Hilbert space, a quantum state $\rho \in \mathcal{D}(\mathcal{H})$ is pure if and only if there exists $|\psi\rangle \in \mathcal{H}$ such that*

$$\rho = |\psi\rangle \langle\psi| \tag{1}$$

Corollary 1. *Let $\mathcal{H}$ be an Hilbert space, let $\rho \in \mathcal{D}(\mathcal{H})$ a mixed quantum state, then there exists $p_1, ..., p_r$ a probability distribution and $\rho_1, ..., \rho_r$ pure quantum states in $\mathcal{D}(\mathcal{H})$ such that*

$$\rho = \sum_{i=1}^{r} p_i \rho_i \tag{2}$$

Then ρ can also be expressed in a vector notation by spectral theorem

$$\rho = \sum_{i=1}^{r} p_i |\psi_i\rangle \langle\psi_i| \tag{3}$$

with $|\psi_1\rangle, ..., |\psi_r\rangle$ orthogonal states.

Proposition 2. *Let $\mathcal{H}$ be an Hilbert space, let $\rho \in \mathcal{D}(\mathcal{H})$, then*

$$Tr\left(\rho^2\right) \leq 1 \tag{4}$$

with equality if and only if ρ is pure.

In quantum information, and generally in information theory, it is common to consider multipartite system. For example, the coupling of two random variables (X, Y) will usually be studied for bipartite communication through channel. Here, we define such multipartite system for quantum information.

Definition 4 (Hilbert Space of multipartite system)**.** *If we consider $A_1, ..., A_n$ quantum systems with associated Hilbert spaces $\mathcal{H}_1, ..., \mathcal{H}_n$, the Hilbert space associated to the multipartite system $A_1...A_n$ is $\mathcal{H}_1 \otimes ... \otimes \mathcal{H}_n$, where $\otimes$ denotes the tensor operation.*

It is relatively easy to see that if $\rho_1 \in \mathcal{D}(\mathcal{H}_1)$ and $\rho_2 \in \mathcal{D}(\mathcal{H}_2)$, then $\rho_1 \otimes \rho_2 \in \mathcal{D}(\mathcal{H}_2)$. However, the converse is not generally true. We introduce definitions for multipartite states.

Definition 5 (Product State)**.** *Let $\mathcal{H}_1, ..., \mathcal{H}_n$ be Hilbert spaces, let $\rho \in \mathcal{D}(\mathcal{H}_1 \otimes ... \otimes \mathcal{H}_n)$ be a density operator on product space. ρ is said to be a product space if*

$$\exists \rho_1 \in \mathcal{H}_1, ..., \exists \rho_n \in \mathcal{H}_n,\ \rho = \rho_1 \otimes ... \otimes \rho_n \tag{5}$$

Definition 6 (Separable / Entangled States)**.** *Let $\rho \in \mathcal{D}(\mathcal{H}_1 \otimes ... \otimes \mathcal{H}_n)$, ρ is said to be separable if it is convex combination of product states. Otherwise, the state is said to be entangled.*

Consider that AB is a bipartite state of Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$. We would like to have a linear map such that given a state of $\mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$, we get the state of system A. This motivates the definition of partial trace

Definition 7 (Partial Trace)**.** *Let $\mathcal{H}_A$,$\mathcal{H}_B$ be Hilbert spaces, the partial trace over $\mathcal{H}_B$ Tr_B is defined as the unique linear map such that*

$$Tr_B \left(|a_1\rangle \langle a_2| \otimes |b_1\rangle \langle b_2|\right) = |a_1\rangle \langle a_2| \, Tr\left(|b_1\rangle \langle b_2|\right) \tag{6}$$

$\forall\, |a_1\rangle, |a_2\rangle \in \mathcal{H}_A, \forall\, |b_1\rangle, |b_2\rangle \in \mathcal{H}_B$

This can be seen as the operation of marginalization for classical random variables, from (X, Y) we get X. Here, from ρ^{AB}, we get ρ^A the reduced operator on A.

Definition 8 (Reduced operator)**.** *Let $\mathcal{H}_A, \mathcal{H}_B$ be two Hilbert spaces, let $\rho^{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$, we define the reduced operator ρ^A on A*

$$\rho^A = Tr_B\left(\rho^{AB}\right) \tag{7}$$

2.2 Quantum Channel

In classical information theory, a channel is usually described as a markov kernel, and therefore by a stochastic matrix when we consider finite alphabets. They can be interpreted as a way of going from a random variable to another. We introduce in this paper a notation quite unusual for classical information theory that fits quantum notation. While a channel $X \sim p$ to $Y \sim q$ would be described by a stochastic matrix Q such that $q = Qp$, we will say that a channel is the action of multiplication on the left by a stochastic matrix, *i.e* $q = Q(p)$.

Definition 9 (Quantum Channel)**.** *Let $\mathcal{H}_A$, $\mathcal{H}_B$ be Hilbert spaces, we say that Φ is a quantum channel, or a channel, if :*

1. *Φ is linear from $\mathcal{L}(\mathcal{H}_A)^H$ to $\mathcal{L}(\mathcal{H}_B)^H$.*
2. *Φ is completely positive (i.e. $\forall$ $\mathcal{H}_E$ Hilbert spaces the map $id_E \otimes \Phi$ is positive)*
3. *Φ is trace preserving (i.e. $\forall \rho$ hermitian operator $Tr\left(\Phi(\rho)\right) = Tr\left(\rho\right)$*

Denote $\mathcal{Q}(\mathcal{H}_A, \mathcal{H}_B)$ the set of quantum channels from $\mathcal{L}(\mathcal{H}_A)^H$ to $\mathcal{L}(\mathcal{H}_B)^H$.

Remark 1. *The complete positivity property is necessary if we want to consider a class of channel such that tensorization is possible. For example, a well known channel positive but not completely positive is the transposition (see [11] box 8.2.). If you define then, $f : \rho \mapsto \rho^T$ and consider $\Phi = f \otimes \mathrm{id}_{\mathcal{L}(\mathcal{H}_B)^H}$, this map is not positive, we might not have a quantum state in return.*

Proposition 3. *Let $\Phi_1, ..., \Phi_n \in \mathcal{Q}(\mathcal{H}_A, \mathcal{H}_B)$, every convex combination of $\Phi_1, ..., \Phi_n$ is also a channel of $\mathcal{Q}(\mathcal{H}_A, \mathcal{H}_B)$*

Definition 10 (D-channels)**.** *If D is an hermitian operator, a $D-$channel Φ is a channel such that $\Phi(D) = D$.*

2.3 Entropy

To define precisely the notion of entropy is not easy since it usually describes a quantity we cannot apprehend. In 1948, C.Shannon introduced the so-called Shannon entropy to answer the following question *"Can we find a measure of how much "choice" is involved in the selection of the event or of how uncertain we are of the outcome? "*[15]. In information theory, one could first consider that the entropy of a distribution would be a measure of the *amount of randomness* contained.

Definition 11 (Shannon's Entropy)**.** *Let p be a probability distribution, we define the Shannon's entropy :*

$$H(p) = -\sum_{i=1}^{n} p_i \log p_i \tag{8}$$

where $0 \times \log(0) = 0$

Many papers then present Von Neumann's entropy, or quantum entropy, as a natural quantum extension. However, historically, the latter preceded the Shannon's entropy.

Definition 12. *Let $\rho \in \mathcal{D}(\mathcal{H})$, we define the Von Neumann's entropy :*

$$S(\rho) = -Tr\left(\rho \log(\rho)\right) \tag{9}$$

Remark 2. *Since ρ is hermitian, $S(\rho)$ can be seen as a function of its eigenvalues. We have then $S(\rho) = H((\lambda(\rho))$*

Example 1. *If we consider a pure state $\rho \in \mathcal{D}(\mathcal{H})$, that is a density operator of rank one, we have $\lambda(\rho) = (1, 0, ..., 0) = \delta$, therefore $S(\rho) = H(\delta) = 0$. And if we consider a uniform state, that is $\rho = \frac{1}{n} I_n$, then $\rho = \log(n)$.*

The Von Neumann's entropy can be interpreted as a measure of randomness on the statistical mixing of a quantum state. A well known interpretation of entropy is also given by Shannon's source coding theorem. To compress a data without loss of information, we need at least $H(p)$ bits. Using Schumacher's source coding theorem analogous to Shannon's, the information of a quantum state can be compressed without loss of information if we use at least $S(\rho)$ qubits.

Among all measure of randomness, Shannon and Von Neumann entropies are the favorites since they combine nice properties, such as tensorization or data processing inequality. It is possible to consider other entropies with less properties.

Definition 13 ($H_{h,\phi}$ entropies)**.** *Let h be monotone and ϕ be convex or concave, we define the $H_{h,\phi}$ entropy*

$$H_{h,\phi}(p) = h\left(\sum_{i=1}^{n} \phi(p_i)\right) \tag{10}$$

We can also define quantum $S_{h,\phi}$ entropies as function of ρ eigenvalues.

$$S_{h,\phi}(\rho) = h\left(\mathrm{Tr}\left(\phi(\rho)\right)\right) \tag{11}$$

For example, a well-known class of entropies are the α entropies, that can be considered as h, ϕ entropies with $h = \frac{1}{1-\alpha} \log$ and $\phi = x \mapsto x^{\alpha}$ for $\alpha \in (0, 1) \cup (1, +\infty)$.

2.4 Divergence

In classical information theory, divergences are used as information distance. If one excepts the source to have a certain distribution, how much information, or surprise, or randomness, will they get out of its assumption. Naturally, quantum information theory has introduced many generalizations of these classical divergences.

Definition 14. *Let ρ and σ be density operators, we define the relative entropy, or Umegaki's divergence :*

$$S(\rho||\sigma) = Tr(\rho\log(\rho)) - Tr(\rho\log(\sigma)) \tag{12}$$

if $\mathrm{Ker}(\sigma) \cap \mathrm{supp}(\rho) \neq \emptyset$ *then* $S(\rho||\sigma) = \infty$

Proposition 4 (Klein's Inequality)**.** *Let $\rho, \sigma \in \mathcal{D}(\mathcal{H})$*

$$0 \leq S(\rho||\sigma) \tag{13}$$

With equality if and only if $\sigma = \rho$.

Proposition 5. *$S(\rho||\sigma)$ is jointly convex.*

Proposition 6. *(DPI) Let ρ,σ be density operator and Φ be a quantum channel.*

$$S(\Phi(\rho)||\Phi(\sigma)) \leq S(\rho||\sigma) \tag{14}$$

These properties are usually the basis of a *useful* quantum divergence. If positivity is not necessarily a relevant property, the definiteness is much more important to use a divergence as measure of distinguishability.

Proposition 7. *Let $\rho, \sigma \in \mathcal{D}(\mathcal{H})$ and $\tau, \omega \in \mathcal{D}(\mathcal{H}')$ with $\rho \neq 0$ and $\tau \neq 0$, we have*

$$S(\rho \otimes \tau||\sigma \otimes \omega) = S(\rho||\sigma) + S(\tau||\omega) \tag{15}$$

Definition 15 (Axiomatic Divergence)**.** *A map $\mathcal{D} : \bigcup_{\mathcal{H}} \mathcal{D}(\mathcal{H}) \times \mathcal{D}(\mathcal{H}) \to (-\infty, +\infty]$ is called a divergence if it verifies*

1. *$\mathcal{D}(\rho||\sigma)$ is jointly convex.*
2. *$\mathcal{D}$ verifies the data processing inequality.*
3. *If $\mathcal{D}(\rho||\sigma) = 0$ then $\rho = \sigma$*

The strong hypothesis, and certainly the most useful is the data processing inequality. It is therefore quite interesting to consider a class of map that only has to verify this property (see [8] for more details)

Definition 16 (Generalized Divergence)**.** *A map $\mathcal{D} : \bigcup_{\mathcal{H}} \mathcal{D}(\mathcal{H}) \times \mathcal{D}(\mathcal{H})$ is called a generalized divergence if it verifies the data processing inequality.*

Proposition 8. *Let $\mathcal{D}$ be a generalized divergence, then it is invariant under isometric operations.*

Proposition 9. *Let $\mathcal{D}$ be a generalized divergence and $\tau \in \mathcal{D}(\mathcal{H})$, we have*

$$\mathcal{D}(\rho \otimes \tau||\sigma \otimes \tau) = \mathcal{D}(\rho||\sigma) \tag{16}$$

We can also consider classical reduction of these definition. An axiomatic classical divergence verifies the three axioms for any diagonal density operators. A generalized divergence only verify the data processing inequality on the diagonal density operators.

3 Majorization

It is common in information theory, or even probability theory, to ask whether or not a probability distribution is *"more or less uniform"* or *"more or less mixed"* than another distribution. This is one motivation behind the use of Shannon entropy since $H(p) = 0$ if and only if p is a dirac distribution and $H(p) = \log n$ if and only if p is the uniform distribution. However, Shannon entropy is an information theorist notion while it feels like the notion of *dispersion* between vectors may be an universal notion. That is one remarkable statement on majorization, it gives a general framework and an universal order to answer these questions. Marshall and Olkin [9] is the most important reference book for majorization.

Notation : We first introduce common notations taken in this paper.

- $\mathcal{P}_n$ denotes the set of $n-$ary probability distributions.
- $u \in \mathcal{P}_n$ denotes the uniform distribution, that is $u = \left(\frac{1}{n}, ..., \frac{1}{n}\right)$.
- $\delta^i \in \mathcal{P}_n$ denotes a dirac distribution, that is $\delta^i = (0, ..., 0, 1, 0, ..., 0)$ where the ith component is equal to 1. δ may also denote a dirac distribution.
- $\|x\|$ denotes the usual L^1 norm on $\mathbb{R}^n$.

3.1 Definitions

Several equivalent definitions of majorization can be considered (see [9]). In this paper, we consider the *doubly stochastic* characterization that will be later generalized in quantum information.

Definition 17 (Majorization)**.** *Let $x, y \in \mathbb{R}^n$, we say that x is majorized by y if there exists a classical channel Φ such that*

$$\Phi(y) = x \quad and \quad \Phi(u) = u \tag{17}$$

Remark 3. *As a reminder, classical channel in finite information theory corresponds to the application of a stochastic matrix A.$\Phi(u) = u$ if and only if A is doubly stochastic, that is*

$$\forall i \in \{1, ..., n\} \sum_j A_{i,j} = 1 \quad and \quad \forall j \in \{1, ..., n\} \sum_i A_{i,j}$$

And the condition $\Phi(y) = x$ can be interpreted as $x = Ay$.

Proposition 10. *$\prec$ defines a pre-order on $\mathbb{R}^n$. Moreover, it verifies*

1. *If $x \prec y$ and $y \prec x$, then x is a permutation of y.*
2. *If $x \prec y$, then for any permutation $\overline{x}$, $\overline{y}$ of x and y, we have $\overline{x} \prec \overline{y}$.*

Remark 4. *Therefore, in the majorization framework, distribution are usually equal up to permutation. Naturally, we can then define a proper order by 1.*

The different approach of majorization has lead to different characterizations, all gather in an intuitive comprehension of *what is ordered* by the majorization.

Proposition 11. *Let $x, y \in \mathbb{R}^n$ in increasing order, the following are equivalent*

1. *$x \prec y$*

2. $\sum_{i=1}^{n} \phi(x_i) \leq \sum_{i=1}^{n} \phi(y_i)$ *for all continuous convex functions* ϕ

3. $\forall k \in \{1, ..., n-1\}, \sum_{i=1}^{k} x_i \leq \sum_{i=1}^{k} y_i$ *and* $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$

Remark 5. *3. is also called the Lorenz curves definition [9]. Let* $x, y \in \mathbb{R}_{+}^{n}$*, denote* $S_x(k)$ *the cumulative sum up to* k*. By plotting the Lorenz curves of* x *and* y*, that is the segments of joints* $\left(\frac{k}{n}, \frac{S_x(k)}{S_x(n)}\right)$ *and* $\left(\frac{k}{n}, \frac{S_y(k)}{S_y(n)}\right)$*, we can observe if one curve is under another. If the Lorenz curve of* x *is under* y*'s one, then* $x \prec y$*.*

Remark 6. *Many entropies can be described by*

$$H_{h,\phi}(p) = h\left(\sum_{i=1}^{n} \phi(p_i)\right)$$

where h *is an increasing function and* ϕ *is a concave function, we have by 2.*

$$p \prec q \Longrightarrow H_{h,\phi}(q) \leq H_{h,\phi}(p)$$

Therefore, $x \prec y$ *is not an easy statement to assume. In some sense, majorization indicates that the disorder of* x *can be compared with* y*'s.*

An important result of the theory of majorization is the *path result.* First, we introduce an important class of doubly stochastic matrices.

Definition 18 (T-transform)**.** *Let* τ *be a transposition on* $\{1, ..., n\}$*, and* P_τ *the associated permutation matrix. We define a T-transform* $T_\tau(\lambda)$

$$T_\tau(\lambda) = \lambda I_n + (1-\lambda) P_\tau \tag{18}$$

T-transforms can be seen from a information theorist point of view as bit flip channels of probability λ (also called BSC). A remarkable result is that every stochastic matrices can be seen as a product of T-transforms [9].

Proposition 12 (Path Result)**.** *If* T *is a T-transform and* $x = Ty$*, then* $x \prec y$*, conversely if* $x \prec y$ *there exists* $T_1, \ldots, T_r$ *T-transforms such that* $x = T_r ... T_1 y$

Proposition 12 allows us to restrict some result to the case where $n = 2$, since every doubly stochastic transformation can be considered as consecutive operations on pairs of coordinates.

3.2 Properties

The following properties mainly comfort the intuition of the majorization as an order of *"how far are we from the uniform ?"*.

Proposition 13. *Let* $x \in \mathbb{R}^n$ *such that* $\max_k x_k \leq P$ *and* $\|x\| = s$ *then*

$$x \prec (P, ..., P, r, 0, ..., 0) \tag{19}$$

where $r = s - \left\lfloor \frac{s}{P} \right\rfloor P$*.*

In particular, if we consider the trivial inequality $\max_k x_k \leq \|x\|$, we get the maximal elements for majorization relation.

Proposition 14 (Maximal element)**.**

$$\forall x \in \mathbb{R}^n,\ x \prec \|x\|\delta \tag{20}$$

where δ is Dirac distribution, that is $\delta = (1, 0, ..., 0)$

Remark 7. *As a consequence, the minimal entropy is always the entropy of a dirac distribution.*

The following result is also natural since there is, *a priori*, no distribution more disordered than the uniform distribution.

Proposition 15 (Minimal element)**.**

$$\forall x \in \mathbb{R}^n,\ \|x\|u \prec x \tag{21}$$

where u is the uniform distribution on $\{1, ..., n\}$.

3.3 Distance to Uniform

In Rioul [12] majorization is used to derive entropic inequalities such as Pinsker and reverse Pinsker for a distance to uniform.

Definition 19 ((Reverse) Pinsker Type Inequalities)**.** *Let $\mathcal{D}$ be a divergence and $p, q \in \mathcal{P}_n$. Denote Δ the statistical distance. A Pinsker type inequality has the form*

$$f\left(\Delta(p, q)\right) \leq \mathcal{D}\left(p||q\right) \tag{22}$$

A reverse Pinsker type inequality has the form

$$\mathcal{D}\left(p||q\right) \leq g\left(\Delta(p, q)\right) \tag{23}$$

These inequalities

Definition 20 (Statistical randomness)**.** *Let $p \in \mathcal{P}^n$, we define the statistical randomness*

$$R(p) = 1 - \frac{1}{n} - \Delta(p, u) \tag{24}$$

Proposition 16. *Let $p \in \mathcal{P}^n$. Denote $R = R(p)$ and $\Delta = \Delta(p, u)$, then for any k such that $\left|\{p > \frac{1}{n}\}\right| \leq k \leq \left|\{p \geq \frac{1}{n}\}\right|$, we have*

$$\left(\frac{1}{n} + \frac{\Delta}{k}, ..., \frac{1}{n} + \frac{\Delta}{k}, \frac{1}{n} - \frac{\Delta}{n-k}, ..., \frac{1}{n} - \frac{\Delta}{n-k}\right) \prec p \prec \left(\Delta + \frac{1}{n}, \frac{1}{n}, ..., R - \frac{\lfloor nR \rfloor}{n}, 0, ..., 0\right) \tag{25}$$

Using the above proposition, we can derive optimal Pinsker and reverse-Pinsker inequalities in the uniform case [12].

Quantum case :
Using the above relation of majorization, we can derive optimal Pinsker and reverse-Pinsker bound when we consider distance to uniform. Consider $\mathcal{H}$ an Hilbert space of dimension d. Denote $\upsilon = \frac{1}{d} I_d$ and $\rho \in \mathcal{D}(\mathcal{H})$. Denote p_i the eigenvalues of ρ. Since they both commute, we have

$$\mathcal{D}_{KL}\left(\rho||\upsilon\right) = \mathcal{D}_{KL}\left(p||u\right) \tag{26}$$

$$\frac{1}{2}\|\rho - \upsilon\| = \frac{1}{2}\|p - u\| \tag{27}$$

4 d-Majorization

Majorization has shown a large success in applications over questions of distance to uniform. However, what if one would like to order his vectors with respect to the disorder from a given distribution. Such an order is given by the d-majorization as presented in Inequalities: Theory of Majorization and Its Applications [9].

4.1 Definition of d-majorization

In definition 17, the uniform distribution is a fixed point of our channel. One can interpret this definition as saying vectors x, y verify $x \prec y$ if and only if x is a mixed vector of y obtained by a transformation that leaves the uniform distribution invariant. This permits to consider the uniform distribution as a reference point to measure the disorder between x and y. This motivates a definition of relative majorization by considering transformation that leaves a distribution d invariant.

Definition 21 (d-Stochastic matrix). *Let $d \in \mathbb{R}^n$, a matrix A is said to be $d-$stochastic if*

1. $\forall i, j \in \{1, ..., n\},\ A_{i,j} \geq 0$
2. $e^T A = e^T$ *(column-stochastic)*
3. $Ad = d$ *(row-d-stochastic)*

Denote $\mathcal{A}_d$ the set of $d-$stochastic matrices.

Definition 22 (d-majorization). *Let $x, y \in \mathbb{R}^n$, we say that x is d-majorized by y if there exists a classical channel Φ such that*

1. $\Phi(y) = x$
2. $\Phi(d) = d$

i.e. there exists a $d-$ stochastic matrix such that $x = Ay$.

Remark 8. *d-majorization is a generalization of classical majorization, since we find the exact characterization of majorization with $d = e$. However, one should notice that it is not an order. To be precise, classical majorization is not an order on the vectors either, it can be considered as such with equality up to permutations. However $d-$majorization must consider the order of the vector. For example, for $d = \left(\frac{1}{3}, \frac{2}{3}\right)$, $p = \left(\frac{2}{3}, \frac{1}{3}\right)$. We can show that $p \not\prec_d d$, while $d \prec_d d$.*

Remark 9. *If we consider a map $\mathcal{P}_n \times \mathcal{P}_n$ that verifies the data processing inequality, then*

$$p \prec_d q \Longrightarrow \mathcal{D}\left(p||d\right) \leq \mathcal{D}\left(q||d\right) \tag{28}$$

Therefore, intuitively $p \prec_d q$ can be interpreted as : if I expect the distribution d, I will be more surprised by the distribution q than d.

4.2 Properties

An overview of geometrical and topological properties of the d-Majorization can be found in [5]. In particular, useful characterizations are presented in this paper.

Proposition 17. *Let $d \in \mathbb{R}^n_{++}$, $x, y \in \mathbb{R}^n$, the following are equivalent*

1. $x \prec_d y$
2. $\forall t \in \mathbb{R},\ \|x - td\|_1 \leq \|y - td\|$
3. $e^T x = e^T y$ *and* $\forall i,\ \|x - \frac{y_i}{d_i} d\|_1 \leq \|y - \frac{y_i}{d_i} d\|_1$
4. *For all continuous convex functions* $f : I_f \longrightarrow \mathbb{R}$ *such that* $f\left(\frac{x_j}{d_j}\right)$ *and* $f\left(\frac{y_j}{d_j}\right)$ *are defined*

$$\sum_{j=1}^{n} d_j f\left(\frac{x_j}{d_j}\right) \leq \sum_{j=1}^{n} d_j f\left(\frac{y_j}{d_j}\right) \tag{29}$$

Remark 10. *In particular, without proving something such as the data processing inequality of* $f-$*divergence, we already have that* $x \prec_d y \Longrightarrow \mathcal{D}_f(x||d) \leq \mathcal{D}_f(x||d)$.

Proposition 18 (choice of d). *Let* $d \in \mathbb{R}^n_{++}$, *let* $\alpha \in \mathbb{R}_{++}$, $\prec_d$ *and* $\prec_{\alpha d}$ *defines the same preorder.*

Therefore, we can consider without loss of generality d to be a probability distribution.

Proposition 19 (Gibbs Inequality). *Let* $x \in \mathbb{R}^n_+$ *and* $d \in {\mathcal{P}_n}^{++}$

$$\|x\| d \prec_d x \tag{30}$$

Remark 11. *Consider a generalized divergence* $\mathcal{D}$, *then for probability distributions*

$$d \prec_d p \Longrightarrow \mathcal{D}(d||d) \leq \mathcal{D}(p||d) \tag{31}$$

Usually, a good divergence should verify $\mathcal{D}(d||d) = 0$ *to be called faithful. Note that this might be not efficient to prove Gibbs inequality. Indeed, data processing inequality is a hard property to prove. However, it is interesting intuitively since the* $d-$*majorization does not need any divergence, it is more general result.*

Proposition 20 (Maximal elements). *Let* $d \in \mathbb{R}^n_{++}$, *denote* $I = \arg\min_k d_k$, *the maximal elements with respect to* $\prec_d$ *of the set* $\{x \in \mathbb{R}^n |\ |x| = s\}$ *are* $\{s\delta_i\}_{i \in I}$

$$\forall x \in \mathbb{R}^n_{++},\ |x| = s,\ \forall i \in I,\ x \prec_d s\delta_i \tag{32}$$

Moreover, we have unicity of maximal element if and only if d *admits a unique minimal element.*

Remark 12. *Note that the uniqueness of maximal element is not necessarily verified. However, for two maximal elements* m, m', *we have*

$$m \prec_d m' \ \text{ and } \ m' \prec_d m$$

Therefore, $D(m||d) = D(m'||d)$.

A main result of the theory of majorization is the *"path result"* (see Proposition 12) since it simplifies heavily the proof involving majorization by considering only transformation on pairs. Considering a generalized T-transformation for d-majorization, called simple d-majorization in [6], could we prove a generalized *"path result"*.

Definition 23 (Simple d-majorization). *Let $d \in \mathbb{R}^n_{++}$ and $q \in \mathbb{R}^n_+$. Consider a distinct pair q_i, q_j, we define a simple d-majorization by*

$$(q_i, q_j) \mapsto \left(\lambda q_i + w\overline{\lambda} q_j, \overline{\lambda} q_i + \left(1 - w\overline{\lambda}\right) q_j\right) \tag{33}$$

where $w = \frac{d_j}{d_i}$ and $\lambda \in [0,1]$ is such that $w\overline{\lambda} \in [0,1]$.

Proposition 21. *Let $p, q \in \mathbb{R}^n_+$ such that $p \prec_d q$. If p and q are similarly ordered with respect to d, that is there exists a permutation π such that $\frac{p_{\pi(1)}}{d_{\pi(1)}} \geq \cdots \geq \frac{p_{\pi(n)}}{d_{\pi(n)}}$ and $\frac{q_{\pi(1)}}{d_{\pi(1)}} \geq \cdots \geq \frac{q_{\pi(n)}}{d_{\pi(n)}}$, then p can be derived from q by simple d-majorizations.*

Explicitely, there exists $A_1, ..., A_m$ d-stochastic matrices that performs d-majorization such that $p = A_m \dots A_1 q$

The above proposition does not hold if p, q are not similarly ordered with respect to d (see [6] example 4.4).

For a given $d \in \mathbb{R}^n_{++}$, $\mathcal{A}_d$ is a convex set of matrices, it is moreover a semi-group. Therefore, the set also admits extreme points that can be seen as relevant d-majorization transformations. In Joe [6], a generalized *"path result"* enlighten the importance of these extreme points.

Theorem 1 (Generalized Path Result). *Let $d \in \mathbb{R}^n_{++}$ and $p, q \in \mathbb{R}^n_+$. If $p \prec_d q$ then there exists a combination $A_1, ..., A_m$ of extreme points and simple $d-$majorization of $\mathcal{A}_d$ such that $p = A_m \dots A_1 q$.*

Remark 13. *Note that when d is the uniform distribution u, $\mathcal{A}_d$ is the set of doubly stochastic. Therefore, the extreme points of set are permutation matrices, which heavily simplifies the path result. Since every permutation matrices can be written as products of transposition matrices, being a special case of T-transform, then the path result of Proposition 1 is equivalent to the Proposition12.*

5 Extension of Majorization to Hermitian Matrices

5.1 Majorization on Hermitian Matrices

The "natural" way of extending majorization on matrices is to consider the "$x = Dy$" characterization, where D is doubly stochastic. For two matrices A, B of size $n \times n$, $A \prec B$ is there exists a doubly stochastic matrix X such that $A = XB$. We will use an equivalent definition for hermitian matrices in this paper. Actually, in the classical case, this can be seen as the existence of a classical channel Φ that verifies $\Phi(y) = x$ and $\Phi(u) = u$. This characterization motivates the following definition.

Definition 24 (Majorization for Hermitian Matrices). *Let A, B be two hermitian matrices, we say that A is majorized by B if there exists a quantum channel Φ such that $\Phi(I_n) = I_n$ and $\Phi(B) = A$*

The theory of majorization for Hermitian matrices can be reduced to the classical majorization using Uhlmann's theorem.

Theorem 2 (Uhlmann). *Let A, B be two hermitians matrices, the following are equivalent*

1. $A \prec B$
2. $\lambda(A) \prec \lambda(B)$

3. there exists $U_1, ..., U_r$ unitary matrices and $p_1, ..., p_r$ a probability set such that

$$B = \sum_{i=1}^{r} U_i A U_i^\dagger$$

The above theorem gives powerful characterizations of the majorization on Hermitian matrices, many properties of the Hermitian majorization can be derived directly from the properties of classical majorization on vectors.

Proposition 22. *$\prec$ defines a preorder on the set of hermitian matrices. Moreover, it verifies the following properties*

1. *If $A \prec B$ and $B \prec A$ then there exists a unitary matrix U such that $A = UBU^\dagger$*
2. *If $A \prec B$, then for any unitary matrix U, we have $UAU^\dagger \prec UBU^\dagger$*

5.2 Definition of D-majorization on Hermitian Matrices

While Hermitian majorization can be reduced to majorization on vectors, it is much harder task for the $d-$majorization. A quantum generalization have been considered in Ende [4] using quantum channels with fixed point D a positive Hermitian operator.

Definition 25 (D-majorization)**.** *Let A, B be two hermitian matrices and D positive definite, we say that A is D-majorized by B if there exists a channel Φ such that $\Phi(B) = A$ and $\Phi(D) = D$.*

Remark 14. *There is no such thing as Uhlmann's Theorem for D-majorization. The eigenvalues of a matrix are usually defined up to re-ordering. Therefore, to have $\lambda(\rho) \prec_{\lambda(D)} \lambda(\sigma)$ one should first define the ordering of each eigenvalues vectors in order to have a meaningful theorem.*

Without loss of generality, we can consider D to be density operator by the following proposition.

Proposition 23. *Let $\alpha \in \mathbb{R}_{++}$, $\prec_D$ and $\prec_{\alpha D}$ defines the same relation.*

Proposition 24. *Let $A, B \in \mathcal{H}_n$ and $D \in \mathcal{H}_n$ a positive matrix, for every unitary U we have*

$$A \prec_D B \iff UAU^\dagger \prec_{UD^\dagger U} UBU^\dagger \tag{34}$$

Remark 15. *Since relative entropy defined by Umegaki, $\mathcal{D}(\rho||\sigma)$ is invariant under unitary transformation, many results on this divergence can be derived without loss of generality by considering D diagonal. The above proposition is quite useful since invariance under unitary transformation is verified for a generalized entropy $\mathcal{D}$, that is a map verifying the data processing inequality.*

Proposition 25. *If we consider ρ, σ, τ three density operators in $\mathcal{D}(\mathcal{H})$, and a generalized divergence $\mathcal{D}$, we have*

$$\rho \prec_\tau \sigma \Longrightarrow \mathcal{D}(\rho||\tau) \leq \mathcal{D}(\sigma||\tau) \tag{35}$$

Proof. Since $\rho \prec_\tau \sigma$ there exists a channel Φ such that $\Phi(\sigma) = \rho$ and $\Phi(\tau) = \tau$. A generalized divergence $\mathcal{D}$ verifies the data processing inequality

$$\mathcal{D}(\Phi(\sigma)||\Phi(\tau)) \leq \mathcal{D}(\sigma||\tau)$$

Hence the result

$$\mathcal{D}(\rho||\tau) \leq \mathcal{D}(\sigma||\tau)$$

□

Therefore, one can consider for many derivations of inequalities to use a diagonal density operators as D. In [4], links between Hermitian and classical $d-$majorization have been found.

Proposition 26. *Let $\rho, \sigma, \tau \in \mathcal{D}(\mathcal{H})$ such that $\tau > 0$. Assume that τ is diagonal. Denote by r, s, t the vector of the diagonal elements of ρ, σ, τ, the following properties hold*

1. *If ρ is diagonal, $r \prec_t s \Longrightarrow \rho \prec_\tau \sigma$*
2. *If σ is diagonal, $\rho \prec^\tau \sigma \Longrightarrow r \prec^t s$*
3. *If ρ and σ are diagonal, then $\rho \prec^\tau \sigma \Longleftrightarrow r \prec^t s$*

Proposition 27 (Minimal Element)**.** *Let $\tau \in \mathcal{D}(\mathcal{H})$ diagonal and positive. Let A be an Hermitian operator, then we have*

$$Tr\,(A)\,\tau \prec_\tau A \tag{36}$$

Moreover, τ is the unique element such that

$$\forall \rho \in \mathcal{D}(\mathcal{H}),\ \tau \prec_\tau \rho \tag{37}$$

The $D-$majorization admits maximal and minimal elements as the classical $d-$majorization [4].

Proposition 28 (Maximal Elements)**.** *Let $\tau \in \mathcal{D}(\mathcal{H})$ diagonal and positive. Define $I = \min_k \tau_k$. Let A be a positive hermitian operator, we have for $i \in I$*

$$A \prec_\tau \ Tr\,(A)\,|i\rangle\,\langle i| \tag{38}$$

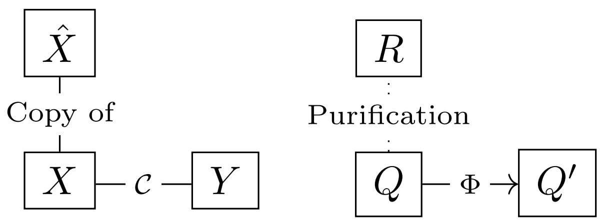
Hence, we have for density operators ρ :

$$\rho \prec_\tau |i\rangle\,\langle i| \tag{39}$$

Moreover, there is a unique maximal element if and only if τ admits a unique minimal element.

6 Quantum Fano Inequality

Consider a quantum system Q, $\mathcal{H}_Q$ a space state of Q, a given basis of pure states $\mathcal{B} = \{|\psi_i\rangle\}$ and $\rho \in \mathcal{D}(\mathcal{H})$. We would like to measure how well a channel preserve entanglement. In other words, how well does the entanglement will be preserved when noise is applied. That is in classical information the same problem as trying to figure how well the source X is preserved along a classical channel.



Therefore, we apply our channel the original source and we compare with our copied system. That is equivalent to apply to the system RQ a super channel $\mathcal{I}_R \otimes \Phi$. The system RQ is entangled before going through the channel, the measure of information on $R'Q'$ would therefore measure a difference of "entanglement information" because R would be unchanged.

6.1 Entanglement Fidelity

Definition 26 (Entanglement fidelity). *Let $\rho \in \mathcal{D}(\mathcal{H})$ a state of a quantum system Q, let R be a purification system and Φ a quantum channel. The system $R'Q'$ is defined as the transformation in the introduction, that is by applying the super channel $\mathcal{I}_R \otimes \Phi$.*

$$F(\rho, \Phi) = F(\rho^{RQ}, \rho^{R'Q'})^2 \tag{40}$$

Proposition 29 (Invariance of the reference system). *If we consider a sum operator representation of Φ $\{E_i\}$, then we have*

$$F(\rho, \Phi) = \sum_i |\, Tr\,(\rho E_i)\,|^2 \tag{41}$$

6.2 Entropy Exchange

Definition 27. *Let $\rho \in \mathcal{D}(\mathcal{H})$ a state of a quantum system Q, let R be a purification system and Φ a quantum channel. The system $R'Q'$ is defined as the transformation in the introduction, that is by applying the super channel $\mathcal{I}_R \otimes \Phi$.*

$$\mathcal{S}(\rho, \Phi) = \mathcal{S}(R', Q') \tag{42}$$

Proposition 30 (Invariance of the reference system). *If $\{E_i\}$ are a set of matrices of the sum operator representation, then we can define $W = \left[Tr\left(E_i \rho E_j^\dagger \right) \right]_{i,j}$ and we have the following equality*

$$\mathcal{S}(\rho, \Phi) = \mathcal{S}(W) \tag{43}$$

Moreover, W is a density operator.

6.3 Quantum Fano Inequality

Quantum Fano inequality is an inequality of the entropy exchange and the entanglement fidelity. In other word, we bound the amount of "energy" or simply information induced by the application of the channel on the density operator, that is $\mathcal{S}(\rho, \Phi)$, by the measure of *how well a channel preserves entanglement.* This has been proven by Schumacher in [14].

Theorem 3. *Let $\mathcal{H}$ a quantum space state of dimension d. Let $\rho \in \mathcal{D}(\mathcal{H})(\mathcal{H})$ and Φ a channel then the following inequation holds*

$$\mathcal{S}(\rho, \Phi) \leq h(1 - F(\rho, \Phi)) + (1 - F(\rho, \Phi)) \log(d^2 - 1) \tag{44}$$

Remark 16. *The above inequality illustrates the analogous role of $F(\rho, \Phi)$. If we write $\mathbb{P}_e = 1 - F(\rho, \Phi)$ and we write $\mathcal{S}(\rho, \Phi) = \mathcal{S}(\rho|\rho')$*

$$\mathcal{S}(\rho|\rho') \leq H(\mathbb{P}_e) + \mathbb{P}_e \log(d^2 - 1) \tag{45}$$

Remark 17. *Note that in the precedent equation, we could write dd_R where d_R is the dimension of the reference system introduced. Therefore, the inequality would hold for any d_R possible.*

$$\mathcal{S}(\rho, \Phi) \leq h(1 - F(\rho, \Phi)) + (1 - F(\rho, \Phi)) \log(sd - 1) \tag{46}$$

where s would be the cardinal of the support of ρ, that is the number of non null eigenvalues. In the case where ρ is defined in a precise context, we can restrict the space state such that ρ has all its eigenvalues in its support ($i.e$ $s = d$)

I have not seen yet an application of this inequality, neither I have seen a proper use of the notion of entropy exchange or entanglement fidelity used in quantum information papers. I have not seen yet a proof involving majorization. Moreover, *Quantum Fano's Inequality* is not actually a *Quantum Fano* since we don't have actual equivalent of conditional distribution, neither a proper equivalent of what is probability of error in quantum information theory. Nevertheless, one can remark that this inequality has a "Fano-form", we might investigate the analogy defined in the QFI.

7 Ideas for RECH202

As presented in this document applications of majorization in quantum information are diverse. It is mainly used as a way to characterize $LOCC$ and $ELOCC$ transformations for bipartite states. The theory for $LOCC$ transformation is well-known and may be studied further considering the lattice induced by majorization [3]. It is still an open question to determine if catalytic majorization induces a lattice structure, it has been proven true for special cases [1]. Therefore, one of my objectives could be to study catalytic majorization in order to derive useful properties and perhaps, with enough ambition, to show if the catalytic majorization may induce a lattice structure. The lattice structure has already lead for classical majorization (LOCC protocol) to optimization of approximate transformation through LOCC protocol.

In classical information theory, majorization can also be seen as a framework to determine powerful properties on randomness, in some sense to consider randomness in information theory without entropic considerations. This might be an interesting approach to have for quantum information. Quantum majorization can be easily reduced to classical majorization, hence we can translate directly results. However, if one considers what we have called *D-majorization* as an extension of *d-majorization*, the reduction is much harder. I would like to go further some characterizations of [4] or eventually to consider other extensions of $d-$majorization using equivalent definitions. An ambitious objective could be to study $d-$majorization to study Pinsker and reverse Pinsker type inequalities as it has been done in Rioul [12] for distance to uniform distribution. Moreover, work on a more general majorization called *relative* majorization in thermodynamics papers [13] have lead to results on approximate majorizations.

References

[1] Gustavo Bosyk et al. "The lattice of trumping majorization for 4D probability vectors and 2D catalysts". In: *Scientific Reports* 8 (Feb. 2018).

[2] Serge Deside. "Majorization relations and quantum entanglement". In: (2022).

[3] Serge Deside et al. "Probabilistic pure state conversion on the majorization lattice". In: (2023).

[4] Frederik vom Ende. "Strict positivity and D-majorization". In: *Linear and Multilinear Algebra* 70.19 (Dec. 2020), pp. 4023–4048.

[5] Frederik vom Ende and Gunther Dirr. "The d-Majorization Polytope". In: *Linear Algebra and its Applications* 649 (Sept. 2022), pp. 152–185.

[6] Harry Joe. "Majorization and divergence". In: *Journal of Mathematical Analysis and Applications* 148.2 (1990), pp. 287–305.

[7] Daniel Jonathan and Martin B. Plenio. "Minimal Conditions for Local Pure-State Entanglement Manipulation". In: *Phys. Rev. Lett.* 83 (7 Aug. 1999), pp. 1455–1458. DOI: 10.1103/PhysRevLett.83.1455. URL: https://link.aps.org/doi/10.1103/PhysRevLett.83.1455.

[8] Sumeet Khatri and Mark M. Wilde. *Principles of Quantum Communication Theory: A Modern Approach*. 2020.

[9] Albert W. Marshall, Ingram Olkin, and Barry C. Arnold. *Inequalities: Theory of Majorization and its Applications*. Vol. 143. 2011.

[10] M. A. Nielsen. "Conditions for a Class of Entanglement Transformations". In: *Physical Review Letters* 83.2 (July 1999), pp. 436–439. ISSN: 1079-7114. DOI: 10.1103/physrevlett.83.436. URL: http://dx.doi.org/10.1103/PhysRevLett.83.436.

[11] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2010.

[12] Olivier Rioul. "The Interplay between Error, Total Variation, Alpha-Entropy and Guessing: Fano and Pinsker Direct and Reverse Inequalities". In: *Entropy* 25.7 (2023).

[13] Takahiro Sagawa. *Entropy, Divergence, and Majorization in Classical and Quantum Thermodynamics*. 2020. arXiv: 2007.09974 [quant-ph].

[14] Benjamin Schumacher. "Sending quantum entanglement through noisy channels". In: (1996). arXiv: quant-ph/9604023 [quant-ph].

[15] C. E. Shannon. "A mathematical theory of communication". In: *The Bell System Technical Journal* 27.3 (1948), pp. 379–423.

[16] Guifré Vidal. "Entanglement of Pure States for a Single Copy". In: *Physical Review Letters* 83.5 (Aug. 1999), pp. 1046–1049. ISSN: 1079-7114. DOI: 10.1103/physrevlett.83.1046. URL: http://dx.doi.org/10.1103/PhysRevLett.83.1046.