ECOLE
POLYTECHNIQUE
DE BRUXELLES

ULB

UNIVERSITÉ LIBRE DE BRUXELLES

Majorization lattice in the theory of quantum entanglement

Master thesis presented in order to be awarded the
Master's Degree in Physics Engineering

Alexander Stévins

Supervisor
Prof. Nicolas Cerf

Advisor
Ir. Serge Deside

Research department
Centre for Quantum Information and Communication (QuIC)

Academic year
2024 - 2025

Extended abstract

Keywords: Majorization lattice, entanglement, Quantum Resource Theory, incomparability, entropic volume, LOCC

Entanglement has long been one of the most puzzling features of quantum mechanics. Some physicists originally believed that entangled states were an artifact of a yet incomplete quantum theory as they deemed some of their properties too absurd for them to truly exist. Notably, it was shown that such states could violate so-called Bell's inequalities, which must be satisfied by any local hidden-variable model, thereby implying that quantum mechanics – and thus our world – is nonlocal if such states were ever observed experimentally. Such states and the accompanying violation of Bell's inequalities were finally confirmed experimentally in the last decades of the 20th century, sparking considerable interest in the theory of entanglement. This led to both deeper theoretical results in the foundations of quantum mechanics and powerful prospects for an emerging quantum technology, which is slowly coming to fruition with, for instance, quantum cryptography, quantum teleportation, or quantum computing.

On the other hand, the mathematical theory of majorization has been used in a variety of fields to compare probability distributions in terms of intrinsic disorder or uncertainty. It provides a (partial) order on probability vectors and is fundamentally tied to information theory via a property of Shannon entropy called Schur-concavity. If a probability vector majorizes another one, it means that is more ordered, implying that its Shannon entropy (as well as every Schur-concave function) must be lower. Majorization theory has found deep applications in quantum information theory, especially in connection with entanglement. It provides a strong separability criterion for detecting whether a state is entangled, which is, in general, a hard problem, and most notably it gives a condition for the interconversion between entangled states via Local Operations and Classical Communications (LOCC), which is a practically relevant class of protocols for manipulating entangled states shared between two quantum computers. The necessary and sufficient condition for a pure bipartite entangled state to be LOCC-convertible into another one can be expressed as a majorization relation implying the so-called Schmidt vectors of the two states. In recent years, the lattice structure induced by the majorization relation – the so-called majorization lattice – has seen growing interest in the field. By drawing on the majorizing supremum (so-called "join") and infimum (so-called "meet") of two probability vectors, the majorization lattice enables a more sophisticated approach to majorization relations with many probability vectors and it is at the heart of majorization-based Quantum Resource Theories (QRT). For example, in a recent work at QuIC, the best conversion protocol to produce the so-called optimal common resource state was proven by using the majorization lattice.

In spite of the close connection between majorization theory and entanglement theory (and some other QRTs), surprisingly little research has been devoted to the majorization lattice since its introduction twenty years ago. The primary objective of this MSc thesis was to address this gap and explore new applications of the majorization lattice to the theory of quantum entanglement. Our starting point was the separability criterion based on a pair of entropy inequalities (comparing the von Neumann entropy of the state to the one of each of its two reduced states). We improved this criterion by leveraging the "meet" of the two reduced states and could prove that the meet-based entropic separability criterion may, in some cases, detect entangled states which remain undetected by the best of the two entropy inequalities. This brought us to the notion of incomparability between two probability vectors (or two bipartite pure entangled states), which arises when a majorization relation linking the two does not exist either way. Incomparability is a key feature in this context because the meet (and join) of two vectors are trivial if the two vectors are comparable. It is therefore tempting to view incomparability as a desirable resource, which is the topic of the core of our MSc thesis.

By taking inspiration from QRTs, we have defined new entropic quantities that measure the amount of incomparability of a probability vector (or bipartite pure entangled state), called

probe, with respect to another one, called reference. We introduced a future-incomparability function, which compares the probe with the future majorization cone emerging from the reference (i.e., the set of vectors that majorize the reference). Symmetrically, we considered a past-incomparability function, which compares the probe with the past majorization cone (i.e., the cone of vectors that are majorized by the reference). Importantly, we could prove that these incomparability functions increase or decrease monotonically under a bistochastic degradation of the probe or reference (with one caveat), promoting them to the status of incomparability monotones upon which a proper QRT could hopefully be built. As a relevant quantum application of these notions, we note that the future-incomparability monotone of an entangled state with respect to another one quantifies how much the former state is far from all the states that are LOCC-reachable from the latter state. This hints to the fact that the geometry of the majorization lattice is crucial, though it is still badly understood as of today.

As an attempt to better understand such a geometry, the last part of our MSc thesis was then centered on a notion of entropic volume in the majorization lattice, which enabled us to extend our resource-theoretical approach of incomparability to a scenario involving multiple probability vectors (or multiple bipartite pure entangled states). The incomparability of a vector with respect to a collection of vectors – what we call a bank – was quantified by using the inclusion-exclusion principle from set theory. The behavior of the proposed generalization seems to indicate that the Shannon entropy of the tip of a majorization cone behaves precisely like the volume of a specific majorization-theoretic set consisting of all LOCC-reachable states from the tip. This intuition was used to quantify the volume of the set of LOCC-reachable states brought by some entangled probe state given a bank of pre-shared entangled states. We call this quantity the uniqueness entropy because it measures what can uniquely be obtained from the probe state in terms of LOCC and is not redundant with what can already be obtained from the bank. While we could not entirely prove that the entropy of the tip of the majorization cone, viewed as a set function, is, indeed, a valid measure, we succeeded in proving at least several of the conditions that follow from measure theory. Furthermore, a convincing argument of it being a measure is that it directly implies the supermodularity of the Shannon entropy on the majorization lattice, which is a proven nontrivial inequality.

Incidentally, as a mathematical side result, our work allowed us to find alternative proofs – sometimes even generalizations – of fundamental inequalities in the majorization lattice. We could prove the subadditivity of Shannon entropy on the lattice and even generalize it to a larger subclass of Schur-concave functions. With some caveat, we could also prove the supermodularity of Shannon entropy on the majorization lattice and, more interestingly, extend it along the same lines.

Finally, the applicability of our resource-theoretical approach to incomparability based on the notion of uniqueness entropy was illustrated by constructing some decision algorithms for Resource-State Selection Strategies (RSSS). The goal of the game consists in producing a target entangled state, given a bank of pre-shared entangled states and considering that LOCC protocols are free. The question is to select, among all states of the bank, the one that can produce the target state while consuming the least amount of entanglement, thereby preserving it as much as possible for later use. By carrying our numerical simulations, we were able to achieve improvements over a simple strategy by using the uniqueness entropy function. While the improvements are not very large, we believe that further research into similar lattice-based quantities might yield even better strategies. More generally, we believe that characterizing entropic distances and volumes in the majorization lattice by exploiting set theory is a fruitful avenue, which will most probably have interesting quantum applications.

Alexander Stévins
Physics Engineering
Majorization lattice in the theory of quantum entanglement
2024-2025

Acknowledgements

Avant tout, j'aimerais sincèrement remercier le professeur Nicolas Cerf de m'avoir accordé l'occasion de travailler avec lui sur la théorie de l'intrication, et sur les subtilités des liens avec la théorie de la majorisation. De même, j'aimerais également remercier mon superviseur Serge Deside pour sa disponibilité tout au long de l'année, et tous les conseils qu'il a pu me donner concernant le treillis de majorisation et la recherche de façon générale. Un grand merci à tous les deux pour cette belle introduction au monde de la recherche.

Je remercie Corentin Vienne pour des discussions utiles concernant la théorie des treillis, la théorie de la mesure et les algèbres de Heyting. De même, je remercie Nathan Vandervelpen pour des discussions utiles concernant la théorie des jeux, les équilibres de Nash, et la théorie de l'optimisation dynamique (et beaucoup de séances d'escalade).

Un grand merci à tous les professeurs et assistants que j'ai eu durant mon parcours scolaire et universitaire, pour m'avoir permis de développer ma culture et mes compétences, et de m'avoir introduit à tant de beaux domaines. Parmi eux, je remercie tout particulièrement Philippe Fernandez de m'avoir transmis sa passion pour les mathématiques. J'espère un jour la communiquer aussi efficacement que lui à mon tour.

J'aimerais ensuite remercier mes parents pour leur soutien continu, et de m'avoir inculqué l'amour de la science dès le plus jeune âge. Je remercie également ma soeur de m'avoir supporté toutes ces années. Og tak til Henry, fordi han gad at lytte på mine teorier selvom han ikke forstod det hele.

Je remercie le Cercle Polytechnique et la Guilde Mandarine dans leur ensemble, que ce soit pour toutes les belles rencontres que j'y ai fait ou pour l'opportunité d'y prendre des postes à responsabilités, à travers lesquels j'ai beaucoup appris.

A Alice, Antoine et Tom, merci d'avoir été la fine équipe de physiciens avec qui partager et décompresser (et souvent ragoter). A Vivien, Rémi et Elitsa, pour avoir été à mes côtés à travers vents et marées depuis des années, un énorme merci. Ces dernières années auraient été fort différentes sans votre amitié, et toutes les opportunités auxquelles vous m'avez introduit.

Enfin, *last but not least*, j'aimerais surtout remercier Mathias d'avoir été mon inséparable compagnon de (més)aventure, et sans qui ces 5 années d'études n'auraient pas été ne serait-ce qu'à moitié aussi amusantes qu'elles ne l'ont été.

Contents

Acronyms

LHS	Left-hand side
RHS	Right-hand side
SPLC	Separable, Piecewise-Linear, Concave
LHVM	Local Hidden-Variable Model
QRT	Quantum Resource Theory
LO	Local Operations
CC	Classical Communications
LOCC	Local Operations and Classical Communications
QC	Quantum Communications
OCR	Optimal Common Resource
CPTP	Completely Positive and Trace-Preserving
RSSS	Resource-State Selection Strategy

Introduction

Entanglement has long been one of the most puzzling features of quantum mechanics. It was first described in 1935 by Einstein, Podolsky and Rosen, who noticed that the Hilbert space nature of quantum states allowed some particles to be in some special superpositions that would connect two particles even at a distance, and that measuring one of the particles could determine the state of the other particle. They conjectured in their seminal paper that the properties of such states were so absurd that their theoretical existence had to imply that quantum mechanics had to be an incomplete theory [1]. Investigating the question, Bell published in 1964 a set of correlation inequalities that any local, hidden-variable model had to satisfy, and then showed that entangled states were capable of violating them [2]. The philosophical ramifications of such a result, which implies that quantum mechanics (and thus our world) must be inherently non-local if the existence of entangled states was ever demonstrated experimentally, are profound.

The experimental confirmation of entanglement only came much later, when experiments showed that photon pairs could indeed violate Bell's inequalities, showing stronger correlations than classically possible [3]. This verification prompted new enthusiasm for the theory of entanglement and its ramifications in many fields, looking for deeper insight into the nature of the universe and advanced technological applications. One such field is quantum information, where entanglement has been the source of powerful technological prospects such as quantum cryptography, quantum key distribution or distributed quantum computing [4].

In 1999, Nielsen proved the conditions under which an entangled pure state shared between two quantum computers can be deterministically transformed into another if the computers can only communicate through a classical channel, a practical paradigm known as Local Operations and Classical Communications (LOCC) [5]. This result showed that manipulating entangled states is theoretically feasible, and a few months later Vidal proved a generalization of his theorem, giving the maximal success probability of non-deterministic entanglement transformations [6].

Interestingly, both of these impactful theorems are written in the language of majorization theory, which is a mathematical field that studies how to compare probability distributions based on how certain they are [7]. Such concepts can be connected to information theory, thanks to a property of the Shannon entropy called Schur-concavity, which is not surprising as it also attempts to capture how uncertain a random variable is [7, 8]. The majorization relation is not a total order, and so probability distributions do not have a *more/less ordered than* relationship with every other probability distribution: we say that the majorization relation induces a *preorder* on probability distributions. Such preorders sometimes admit an infimum (the meet) and supremum (the join), generating mathematical structures called lattices, which are well-studied in order theory [9]. In 2002, Cicalese and Vaccaro showed that the majorization relation does admit such an infimum and supremum, introduced the majorization lattice, and showed some new properties of the Shannon entropy on these meet and join distributions known as *supermodularity* and *subadditivity* [10]. In 2013, along with Gargano, they also used the majorization lattice to define an entropic distance for probability distributions [11].

The majorization lattice is a more sophisticated tool to study majorization relations and entropic inequalities. Taking inspiration from economics, recent years have seen the development of Quantum Resource Theories (QRT) such as the theories of quantum entanglement, quantum

thermodynamics, quantum coherence, and many others [12]. Several of those are majorization-based, and the majorization lattice was used to define special states, called the Optimal Common Resource (OCR) of a set of targets which is the state with the least amount of (entanglement) resource capable of reaching (by LOCC) all of the targets of the set, ensuring minimal waste of resource [13]. A recent work at QuIC also found the best conversion protocol to produce such OCRs, and the optimality of the protocol was proven using the lattice [14].

The goal of this work was to explore the applications of the majorization lattice to the theory of quantum entanglement. Several avenues were explored, but the main idea that was explored was to study and quantify the incomparability between probability distributions. This was done because in the LOCC context, incomparability seems to enable more diversity in the set of reachable states, and is thus a desirable property for two states. This idea can be generalized to sets of states, enabling a new volumic intuiton for the Shannon entropy, which can be used to select states in LOCC protocols based on their redundancy relative to a set of states.

This MSc thesis is separated in two parts. The first part provides an overview of the litterature. First, in Chapter 1, a theoretical explanation of majorization theory and the majorization lattice is given. Then, Chapter 2 provides a review of quantum information, going from classical information and the Shannon entropy to Nielsen's LOCC reachability theorems, which are the main connection point to majorization theory.

The second part of this manuscript goes over new results found during this MSc thesis. Chapter 3 proves some of the most important mathematical properties of entropy we use in the following chapters. We go over a more general proof of subadditivity for a broader class of functions than only the Shannon entropy, which were achieved by using a concatenation-based technique to find an underlying majorization relation (which we call a *majorization precursor*). A similar majorization precursor, which seems to hold numerically, is also conjectured for supermodularity. Then, Chapter 4 quickly goes over an improvement to entropic separability criteria based on the meet of two distributions (which might be of interest in some experimental setups). However, the improvement over the previous criterion directly depends on how incomparable the two vectors are, motivating an attempt to quantify the incomparability between pairs of distributions. In an attempt to do this, we go over definitions of new entropic quantities which we call incomparability monotones, quantifying how incomparable two probability distributions are by taking inspiration from QRTs, along with theorems showing their behavior under LOCC. Finally, the next logical step was to generalize the incomparability monotones to sets of states, which is done in Chapter 5. This was done by exploiting the inclusion-exclusion principle and a relation between the intersection of majorization cones and the cone of their join, taking inspiration from geometrical properties of majorization cones that have recently been studied in the field of quantum thermodynamics [15]. With this new entropic inclusion-exclusion formula, we prove several properties which seem to indicate that the Shannon entropy behaves like the volume of majorization cones, and use this idea to quantify the uniqueness of a state relative to a bank of entangled states. Using these uniqueness measures, Resource-State Selection Strategies (RSSS) were proposed which, according to our numerical simulations, run out of entangled states capable of reaching successive entangled targets through LOCC slightly slower *on average* than naive strategies based only on the entropy of individual states.

Part I

State of the art

Chapter 1

Majorization theory

Definitions for majorization theory mostly follow Ref. [7]. The notations for majorization and the majorization lattice, as well as the reordering convention $p^{\downarrow}$ (cf. Section 1.1.1) mostly come from Ref. [10].

1.1 The majorization relation

The theory of majorization gives a framework for what it means for a probability distribution (or vectors) to be more disordered, more random, than another one. A rigorous notion of disorder such as the one that majorization suggests is interesting because it is linkable to other objects that attempt to quantify randomness in information theory, and most notably the Shannon entropy. In order to see that this is the case, we first need to rigorously define the majorization relation.

1.1.1 Definition using cumulative sums

This section is based on Ref. [7, pp. 4–10]. Let p be a probability vector, *i.e.* a vector such that $p_i \geq 0$ and $\sum_i^d p_i = 1$ for a vector of dimension d. Let $p^{\downarrow}$ be the non-increasing reordering of p, meaning that the entries of $p^{\downarrow}$ are the same as those of p, but sorted such that $p_1^{\downarrow} \geq p_2^{\downarrow} \geq ... \geq p_d^{\downarrow}$. We will always be considering reordered vectors in this master thesis, and we will therefore denote the set of d-dimensional probability vectors sorted in non-increasing order $\mathcal{P}^d$. Formally, $\mathcal{P}^d = \{p \in \mathbb{R}^d | \sum_i^d p_i = 1, p_i \geq p_{i+1}\}$.

Definition 1.1 (Cumulative sum)**.** Let p be a vector in $\mathcal{P}^d$. The k^{th} non-increasing[1] cumulative sum of p is the sum of the j biggest components of p, which can be written as

$$S_k^{\downarrow}(p) \coloneqq \sum_{i=1}^{k} p_i^{\downarrow}, \tag{1.1}$$

where $p_i^{\downarrow}$ is the i^{th} largest entry of p.

Definition 1.2 (Majorization relation)**.** Let $p, q \in \mathbb{R}^d$ be two vectors of dimension d. We say that p is majorized by q, written $p \prec q$, if and only if

$$\begin{cases} S_k^{\downarrow}(q) \geq S_k^{\downarrow}(p) & \forall k = 1, ..., d-1, \\ S_d^{\downarrow}(q) = S_d^{\downarrow}(p). \end{cases} \tag{1.2}$$

If the dimensions of the vectors do not match, one can always append zeroes to the one of lower dimension and apply the definition on the enlarged vector.

[1] We could also have defined $\mathcal{P}^d$ to be non-decreasing vectors and use non-decreasing cumulative sums $S_k^{\uparrow}(p)$ instead. We would get an equivalent description of majorization, the only effect being reversed inequalities in (1.2).

Remark 1.1. If p and q are probability vectors, then the last equality is automatically verified because $S_d^{\downarrow}(p) = S_d^{\downarrow}(q) = 1$.

This definition encapsulates the idea of disorder: if $p \prec q$, then the largest probability of q ($q_1^{\downarrow}$) is larger than the largest probability of p ($p_1^{\downarrow}$), and the rest of the distribution in q is lowered for the sum of probabilities to give 1, whereas p has a distribution that is more spread out. As an example, consider $q = (0.9, 0.1)$ and $p = (0.6, 0.4)$, and applying Definition 1.2 we see that $p \prec q$. This feels fairly intuitive : q is not as uncertain as p, because a process described by probability distribution q sees one of the 2 events happen most of the time, whereas the outcome of a process with probability distribution p is almost equally likely to be one or the other. It is clear from this example that the *majorizer* is more ordered than the *majorized*.

Remark 1.2. For any $p \in \mathcal{P}^d$, we have

$$\left(\frac{1}{d}, ..., \frac{1}{d}\right) \prec p \prec (1, 0, ..., 0) \tag{1.3}$$

which is fairly intuitive, considering that the flat distribution $\overline{1}_d \coloneqq \left(\frac{1}{d}, ..., \frac{1}{d}\right)$ is the most uncertain distribution and the peaked distribution $\overline{\delta}_d \coloneqq (1, 0, ..., 0)$ is the most certain distribution.

The majorization relation creates a partial order on $\mathcal{P}^d$, meaning that the ordering is both reflexive and transitive on the set, and that[2] $p \prec q$ and $p \succ q \implies p = q$ for any $p, q \in \mathcal{P}^d$. When either $p \prec q$ or $p \succ q$ is verified, we will say that p and q are *comparable* (written $p \sim q$). However, there also exist cases where both $p \not\prec q$ and $p \not\succ q$ are true, and we will then say that p and q are *incomparable* (written $p \nsim q$). For example, $(0.5, 0.25, 0.25) \nsim (0.4, 0.4, 0.2)$.

Remark 1.3. Two vectors of dimension 2 are always comparable. Incomparability is only possible from dimension 3 and above.

1.1.2 Definition using bistochastic matrices

This section is based on Ref. [7, pp. 29–33]. A second equivalent definition for majorization can be reached by a different intuition involving bistochastic matrices. The idea is the following: if the probability vector q can be obtained by randomly mixing the components of p together, then p is more ordered than q. It turns out that this simple idea is equivalent to the definition using cumulative sums. Let us express this rigorously.

Definition 1.3 (Bistochastic matrix)**.** A bistochastic matrix D is a matrix in $\mathbb{R}^{d\times d}$ such that

$$\begin{cases} D_{i,j} \geq 0 \quad \forall i, j = 1, ..., d \\ \sum_{i=1}^{d} D_{i,j} = 1 \quad \forall j = 1, ..., d \\ \sum_{j=1}^{d} D_{i,j} = 1 \quad \forall i = 1, ..., d \end{cases} \tag{1.4}$$

A bistochastic matrix (also called doubly stochastic matrix) is basically a matrix with non-negative entries such that the columns and rows each sum to one. The following theorem formalizes the way that bistochastic matrices mix entries of vectors together.

Theorem 1.1 (Birkhoff's theorem)**.** *If D is a d-dimensional bistochastic matrix, then there exists a probability distribution $\{p_j\}$ and a set of d-dimensional permutation matrices P_j such that*

$$D = \sum_j p_j P_j \tag{1.5}$$

[2] This is not the case if we choose the subset of $\mathbb{R}^d$ of vectors with non-negative entries, where having both $p \succ q$ and $p \prec q$ only means that p and q are the same up to a reordering. On non-ordered vectors, the majorization relation therefore only generates a preorder instead of a proper partial order.

From this theorem, we can see that the effect of a bistochastic matrix on a vector is essentially to perform a convex mixture of the vector's entries, producing a vector that is more smoothed out, and thus more disordered.

Theorem 1.2 (Hardy, Littlewood and Pólya)**.** *Let $p, q \in \mathcal{P}^d$. Then, $p \prec q$ if and only if there exists a d-dimensional bistochastic matrix D such that*

$$p = Dq. \tag{1.6}$$

Because a distribution loses certainty after being applied a bistochastic matrix, we will often say that the vector has gone through a *bistochastic degradation*. One can take Theorem 1.2 as a definition of the majorization relation, as the notion of a vector being more disordered is perhaps clearer in the bistochastic matrix picture. Then, the nature of the convex combination of entries would have given us the set of inequalities (1.2). Let's take the same example as earlier, $p = (0.6, 0.4)$ and $q = (0.9, 0.1)$. Then we can find that $D = \begin{pmatrix} 5/8 & 3/8 \\ 3/8 & 5/8 \end{pmatrix}$ gives us $p = Dq$, and therefore $p \prec q$.

Whichever picture one prefers intuitively, the fact that there is a full equivalence between the two is quite powerful when working on proofs, as sometimes one picture is easier to work with than the other depending on the context.

Remark 1.4. The bistochastic matrix $\begin{pmatrix} \frac{1}{d} & \cdots & \frac{1}{d} \\ \vdots & & \vdots \\ \frac{1}{d} & \cdots & \frac{1}{d} \end{pmatrix}$ produces the probability vector $(\frac{1}{d}, \ldots, \frac{1}{d})$ regardless of the probability vector on which it is applied, hence $(\frac{1}{d}, \ldots, \frac{1}{d})$ is majorized by every probability vector in $\mathcal{P}^d$.

1.1.3 Visualization using Lorenz curves

This section is based on Ref. [7, pp. 5–6]. Lorenz curves were first introduced in 1905 by economist Max Lorenz as a way to represent income inequality. By plotting the cumulative sums of two distributions, we can say that if one curve is above the other at all times, then it is more unequal (more disordered) than the second. By normalizing the distributions for total income, this graphical tool could then be used to quickly compare income inequality between different regions or countries.

More formally, the Lorenz curve of a vector $p \in \mathbb{R}^d$ is obtained by linear interpolation between the points of the set $\{(i, S_i^{\downarrow}(p)) | i \in \mathbb{N}, i \leq d\}$. We will be using the convention that $S_0^{\downarrow}(v) = 0$ for any vector v. A couple of examples are given in Figures 1.1 and 1.2.

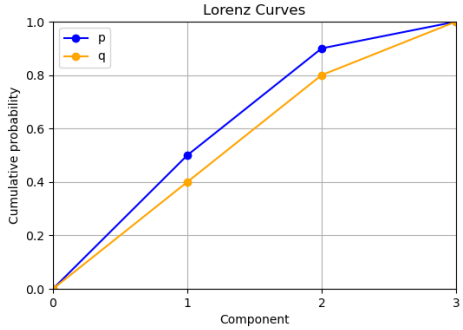


Figure 1.1: Lorenz curves for comparable vectors $p = (0.5, 0.4, 0.1)$, $q = (0.4, 0.4, 0.2)$.

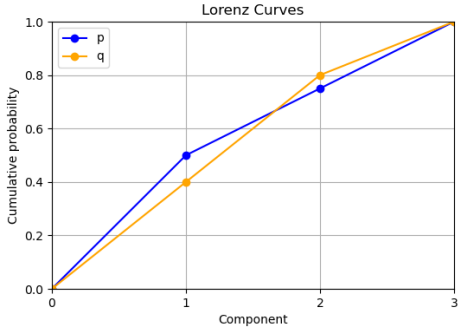


Figure 1.2: Lorenz curves for incomparable vectors $p = (0.5, 0.25, 0.25)$, $q = (0.4, 0.4, 0.2)$.

At integer abscissa i, we see the values of $S_i^{\downarrow}(v)$ for each vector v. At each point, we can thus compare the value of the cumulative sums of each vector, which is precisely the same

as an inequality of the majorization relation (1.2). From these figures, we quite quickly get the intuition that if one vector majorizes another (i.e. they are comparable), then all of the inequalities must be satisfied and the majorizer's curve must be above the majorized's curve at all points. If the two vectors are incomparable, then at least one inequality must fail, and so there must be a reversal in the largest cumulative sum, and the Lorenz curves must cross at least once.

1.2 Schur-convex and Schur-concave functions

Some functions preserve the ordering defined by the majorization relation and are therefore interesting to study, as a majorization relation directly implies an inequality on these functions.

1.2.1 Definition and Karamata's inequality

This section is based on Ref. [7, pp. 79–92]. The following few definitions will be useful.

Definition 1.4 (Convex function)**.** A function $f \in C(\mathcal{I} \to \mathbb{R})$ is convex on an interval $\mathcal{I}$ iff for all $x_1, x_2 \in \mathcal{I}$

$$f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha) f(x_2) \quad \forall \alpha \in [0,1]. \tag{1.7}$$

Definition 1.5 (Schur-convex function)**.** A function $f \in C(\mathbb{R}^d \to \mathbb{R})$ is Schur-convex iff for all $p, q \in \mathcal{P}^d$

$$p \prec q \implies f(p) \leq f(q). \tag{1.8}$$

Conversely, a function f is (Schur-)concave if $-f$ is (Schur-)convex. Additional properties linking convex and concave functions to majorization exist.

Lemma 1.1 (Karamata's inequality)**.** *If $f \in C(\mathcal{I} \to \mathbb{R})$ is convex on $\mathcal{I}$, and if $p \prec q$, then*

$$\sum_{i=1}^{d} f(p_i) \leq \sum_{i=1}^{d} f(q_i). \tag{1.9}$$

This can be restated as saying that the function $\sum_i f$ is Schur-convex.

It is interesting to note that a statement very similar to Lemma 1.1 can also be seen as a definition of majorization. The statement is the following: if $\sum_{i=1}^{d} f(p_i) \leq \sum_{i=1}^{d} f(q_i)$ for *any convex function* f, then $p \prec q$.

1.2.2 Shannon entropy

This section is based on Refs. [7, p. 101] and [8, p. 88]. The Shannon entropy $H(p) \coloneqq -\sum_{i=1}^{d} p_i \log p_i$ is a foundational quantity in classical information theory (cf. Section 2.1). $H(p)$ essentially captures the uncertainty content of a random process distributed over a probability distribution p. We therefore expect it to be linked to majorization theory which gives a partial order on the uncertainty of different probability distributions. The following lemma establishes this link through Schur-concavity.

Lemma 1.2 (Schur-concavity of the Shannon entropy)**.** *For any $p, q \in \mathcal{P}^d$,*

$$p \prec q \implies H(p) \geq H(q). \tag{1.10}$$

This property of the Shannon entropy is fairly intuitive: the Shannon entropy captures the uncertainty content of a probability distribution, and $p \prec q$ means that q is unequivocally[3] more ordered (so more certain) than p, so we would expect Schur-concavity to hold. However, a majorization relation is stronger than an entropic inequality, and so we should study majorization relations alongside entropy to gain a better understanding of information theory.

[3] As opposed to incomparable distributions being more ordered than the other each in their own way.

1.2.3 Rényi entropies

This section is based on [16]. A generalization of the Shannon entropy is the concept of Rényi entropies, which are a class of functions that can all be used to quantify the uncertainty of a random variable.

Definition 1.6 (Rényi entropy)**.** Let p be a probability distribution of dimension d. Then, its Rényi entropy of order α is

$$H_\alpha(p) = \frac{1}{1-\alpha} \log \left(\sum_i p_i^\alpha \right) \tag{1.11}$$

The Shannon entropy is a special case of the Rényi entropy and is obtained as the limit of H_α as α goes to 1. Other values also hold specific interpretations, but we will not go over them here. A classical result is the following, proven using the Schur-Ostrowski criterion [7, p. 84].

Lemma 1.3 (Schur-concavity of the Rényi entropy)**.** *All Rényi entropies of order $\alpha \geq 0$ are Schur-concave.*

1.3 The majorization lattice

1.3.1 Lattice structures

This section is based on [10]. The majorization relation creates a proper partial ordering on the set of *ordered* probability vectors $\mathcal{P}^d$. With the ordering induced by the majorization relation, the set of ordered vectors $\{z | z_1 \geq z_2 \geq ... \geq z_d\}$ becomes a lattice.

Definition 1.7 (Lattice)**.** A lattice is a quadruple $\langle \mathcal{L}, \sqsubseteq, \wedge, \vee \rangle$ where $\mathcal{L}$ is a set, $\sqsubseteq$ is a partial ordering $\mathcal{L}$, and for all $a, b \in \mathcal{L}$ there is a unique greatest lower bound (glb) $a \wedge b$ and a unique least upper bound (lub) $a \vee b$. That is, $a \wedge b$ precedes ($\sqsubseteq$) both a and b and $a \vee b$ succedes ($\sqsupseteq$) both a and b, and for any c that precedes both a and b, and for any d that succedes both a and b, we have

$$c \sqsubseteq a \wedge b \quad \text{and} \quad a \vee b \sqsubseteq d. \tag{1.12}$$

A simple example of a lattice is the set $\mathbb{N}$ of natural numbers ordered by the relation "divides". Then for $a, b \in \mathbb{N}$ the lub is the least common multiple of (a, b) and the glb is the greatest common divisor of (a, b).

It was shown that the set $\mathcal{P}^d$ endowed with the partial ordering $\prec$ is a lattice. As the lattice is a central part of this master thesis, we will go through the main ideas behind the proof, which hinges on defining the correct greatest lower bound (glb) $p \wedge q$, which is called the *meet* of p and q, and the correct least upper bound (lub) $p \vee q$, which is called the *join*.

Let us first go through what it means for a vector to be the glb, the meet, of two vectors $p, q \in \mathcal{P}^d$. Intuitively, the glb of p and q in terms of majorization is the most ordered vector that is majorized by both p and q. While this may not seem very interesting at a first glance, recall that two vectors may be incomparable, meaning that both $p \not\prec q$ and $p \not\succ q$ are true. Even if they are incomparable, the subset of vectors that are majorized by both p and q is nonempty[4] and has a supremum, $p \wedge q$, *which majorizes every other vector in the subset.* Note that an equivalent definition of a lattice is to require that every partially ordered subset has a supremum and infimum [9, p. 34]. Since the two definitions are equivalent, we can consider this to be a property of the lattice, guaranteeing the existence of the supremum $p \wedge q$.

In a similar fashion, the lub, the join, of two vectors $p, q \in \mathcal{P}^d$ is the least ordered vector that majorizes both p and q. The subset of vectors that majorize both p and q is nonempty[5]

[4] It is guaranteed to at least contain the flat distribution $\overline{1}_d = (\frac{1}{d}, \ldots, \frac{1}{d})$ since it is majorized by every other distribution.

[5] It is guaranteed to at least contain the certain distribution $\overline{\delta}_d = (1, 0, \ldots, 0)$ since it majorizes every other distribution.

and has an infimum, $p \vee q$, *which is majorized by every other vector in the subset.* Moreover, the same property as before guarantees the existence of the infimum.

Corollary 1.2.1. *If two vectors are comparable, say $p \prec q$ (the $p \succ q$ case being symmetric), then the question becomes trivial as the least ordered vector that majorizes both p and q is q and the least disordered vector that is majorized by both p and q is p, i.e. $p \wedge q = p$ and $p \vee q = q$.*

1.3.2 Majorization cones and geometric intuition

Before going further with the proof sketch, let us first give a 2D visual representation of the lattice, which is very helpful in building intuition for using the lattice, first proposed in the field of thermomajorization [17]. More accurate representations exist but are much more complicated, living in the canonical Weyl chamber of a Δ_{d-1} simplex [15]. Nevertheless, the simple 2D depiction is very useful to think about, but one should not draw hasty conclusions from it.

Definition 1.8 (Majorization cones)**.** The set of states that majorize p is called the *future cone*[6] $\mathcal{T}_+(p)$. The set of states that are majorized by p is called the *past cone* $\mathcal{T}_-(p)$. The set of states that are neither in the past nor in the future cone of p is called the *incomparable region* $\mathcal{T}_\emptyset(p)$.

Figures 1.3 and 1.4 show examples of the 2D depiction of vectors $p, d \in \mathcal{P}^d$ on the lattice and their majorization cones, as well as their meet and join.

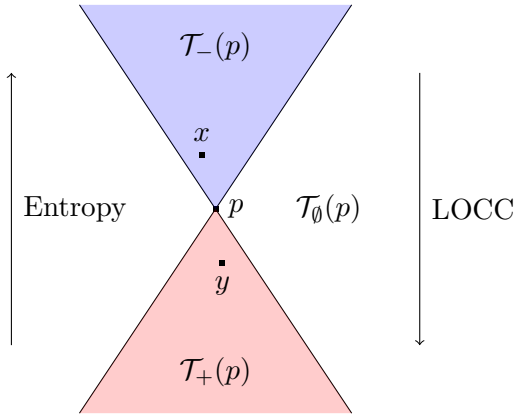


Figure 1.3: Depiction of the majorization cones of a nondescript vector $p \in \mathcal{P}^d$. The future cone $\mathcal{T}_+(p)$ and past cone $\mathcal{T}_-(p)$ of p have been shaded red and blue, respectively. The incomparable region to p, $\mathcal{T}_\emptyset(p)$ is shaded white. The arrow shows the direction in which $H(p)$ increases. We also see that $x \prec p \prec y$.

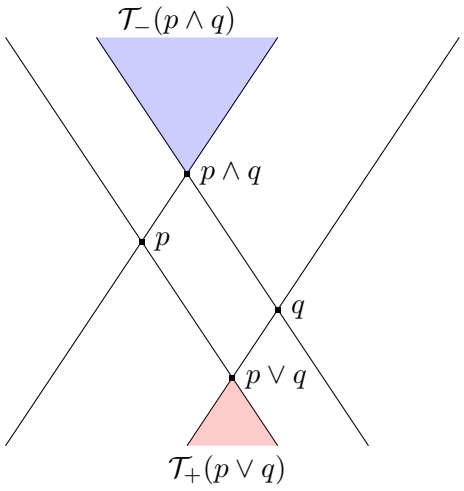


Figure 1.4: Depiction of the intersection of the majorization cones of incomparable vectors $p, q \in \mathcal{P}^d$. Note that by definition of the meet and join, $\mathcal{T}_+(p) \cap \mathcal{T}_+(q) = \mathcal{T}_+(p \vee q)$ and $\mathcal{T}_-(p) \cap \mathcal{T}_-(q) = \mathcal{T}_-(p \wedge q)$.

In classical and quantum information, the lattice is usually drawn with more disordered vectors at the top and more ordered vectors at the bottom because the entropy of a state is considered a resource (cf. Section 2.3.1). This means that the uniform distribution sits at the very top of the diagram, and the certain distribution at the very bottom. One should therefore be careful, because the majorization relation runs from top to bottom, which is the opposite of what one would picture at first (a greatest lower bound looks like a least upper bound and vice-versa). While this is inconvenient for this chapter, this convention will make more sense later down the road.

[6] Our naming conventions and notations are for the most part entirely opposite to the thermomajorization conventions for reasons that will be made clear in Section 2.3.3. Most notably, our future cone corresponds to their past cone and vice versa. Moreover, entropy goes down from past to future because it will be linked to entanglement going down under LOCC.

1.3.3 Constructing the meet

In order to show that the quadruple $\langle \mathcal{P}^d, \prec, \wedge, \vee \rangle$ is indeed a lattice, we need to find an appropriate definition for the glb $p \wedge q$ and lub $p \vee q$ of two arbitrary elements p and q in $\mathcal{P}^d$. Let us start with the meet $p \wedge q$. Thinking in terms of Lorenz curves is actually very helpful here and can lead us very intuitively to the correct algorithm for constructing the meet. Let us choose the example $p = (0.6, 0.2, 0.2)$ and $q = (0.45, 0.45, 0.1)$, both in $\mathcal{P}^3$ (below dimension 3 all vectors are comparable and the lattice is trivial). Intuitively, the most ordered state that is majorized by both p and q should have a Lorenz curve that hugs the curves of p and q from below, let us write it $\alpha(p, q)$. Figure 1.5 shows their Lorenz curves.

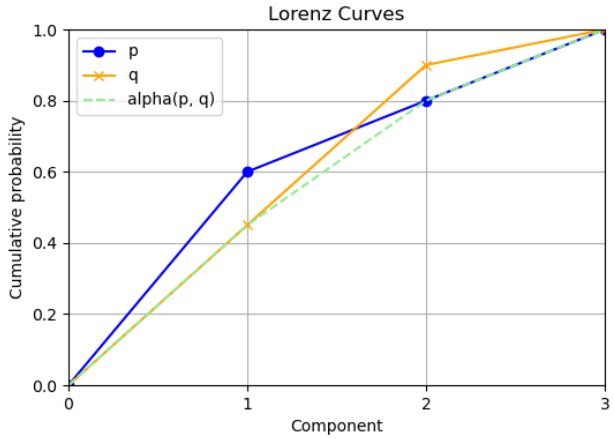


Figure 1.5: Representation of the $p = (0.6, 0.2, 0.2)$ and $q = (0.45, 0.45, 0.1)$ example, along with the proposed $\alpha(p, q)$ glb vector.

More formally, we can define the vector $\alpha(p, q) = (a_1, \ldots, a_d)$ as

$$a_i = \min\Big\{\sum_{j=1}^{i} p_j, \sum_{j=1}^{i} q_j\Big\} - \sum_{j=1}^{i-1} a_j \tag{1.13}$$

$$= \min\Big\{\sum_{j=1}^{i} p_j, \sum_{j=1}^{i} q_j\Big\} - \min\Big\{\sum_{j=1}^{i-1} p_j, \sum_{j=1}^{i-1} q_j\Big\}. \tag{1.14}$$

Lemma 1.4 (Cicalese and Vaccaro, 2002 [10]). *For all $p, q \in \mathcal{P}^d$ we have $\alpha(p, q) = p \wedge q$.*

1.3.4 Constructing the join

We can proceed in a similar way for the join of two vectors $p \vee q$, and thinking in terms of Lorenz curves will once again give us the right idea to move forward. As a first attempt, let us take $p = (0.6, 0.2, 0.2)$ and $q = (0.45, 0.45, 0.1)$ again, both in $\mathcal{P}^3$. This time, the least ordered vector that still majorizes both p and q should have a Lorenz curve that hugs their curves from above, let us write it $\beta(p, q)$. Figure 1.6 shows their Lorenz curves. Formally, we can define the vector $\beta(p, q) = (b_1, ..., b_d)$ as

$$b_i = \max\Big\{\sum_{j=1}^{i} p_j, \sum_{j=1}^{i} q_j\Big\} - \sum_{j=1}^{i-1} b_j \tag{1.15}$$

$$= \max\Big\{\sum_{j=1}^{i} p_j, \sum_{j=1}^{i} q_j\Big\} - \max\Big\{\sum_{j=1}^{i-1} p_j, \sum_{j=1}^{i-1} q_j\Big\}. \tag{1.16}$$

While this is a step in the right direction, things are unfortunately more complicated than for the meet, as this vector is not guaranteed to be sorted and is thus not necessarily in $\mathcal{P}^d$. In

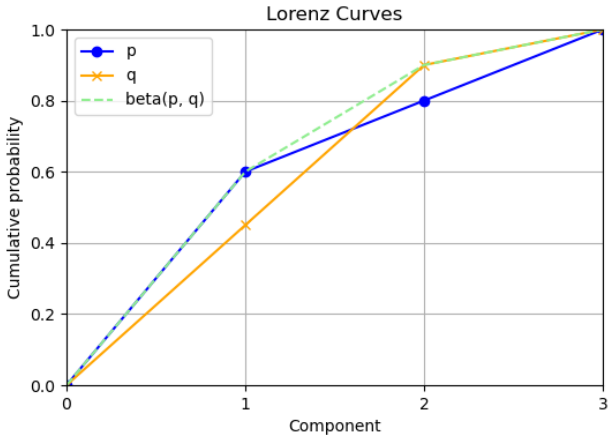


Figure 1.6: Representation of the $p = (0.6, 0.2, 0.2)$ and $q = (0.45, 0.45, 0.1)$ example, along with the prototype $\beta(p, q)$ lub vector.

order to see why, let us take a slightly more complicated example with $p = (0.6, 0.15, 0.15, 0.1)$ and $q = (0.5, 0.25, 0.20, 0.05)$ and let us define $\beta'(p, q) = (\beta(p, q))^{\downarrow} \in \mathcal{P}^d$. We have

$$\beta(p, q) = (0.6, 0.15, 0.2, 0.05) \notin \mathcal{P}^4 \tag{1.17}$$

$$\beta'(p, q) = (0.6, 0.2, 0.15, 0.05) \in \mathcal{P}^4. \tag{1.18}$$

Figure 1.7 shows the Lorenz curves of $p, q, \beta(p, q)$ and $\beta'(p, q)$. It is immediate to see that $p, q \prec \beta'(p, q)$, but while $\beta'(p, q)$ is indeed an upper bound on p and q it is not necessarily majorized by all of the vectors in $\mathcal{T}_+(p) \cap \mathcal{T}_+(q)$. Intuitively, we can still find a Lorenz curve below $\beta'(p, q)$ by choosing the shortest chord (a straight line) between b'_1 and b'_3, i.e. by smoothing over the concave dent in $\beta(p, q)$. This idea yields the vector $\beta''(p, q) = (0.6, 0.175, 0.175, 0.05)$ which is clearly majorized by $\beta'(p, q)$, as shown in Figure 1.8.

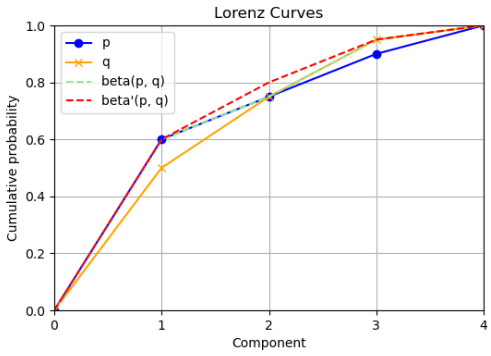


Figure 1.7: Lorenz curves for the vector $p = (0.6, 0.15, 0.15, 0.1)$ and the vector $q = (0.5, 0.25, 0.20, 0.05)$, along with the join prototypes $\beta(p, q)$ and $\beta'(p, q)$.

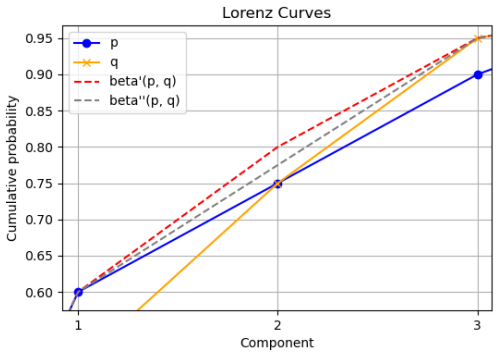


Figure 1.8: Zoom on the region that visually shows that the $\beta'(p, q)$ vector is not the join, as there exist vectors such as $\beta''(p, q)$ that are majorized by it but still majorize both p and q.

We have the following statement:

$$p, q \prec (0.6, 0.175, 0.175, 0.05) \prec \beta'(p, q). \tag{1.19}$$

The following definition provides the smoothing algorithm to apply to $\beta(p, q)$ to obtain the join of p and q.

Definition 1.9 (Concave dent smoothing algorithm)**.** Let $b = (b_1, ..., b_d)$, and let j be the smallest integer in $\{2, ..., d\}$ such that $b_j > b_{j-1}$. Moreover, let i be the greatest integer in $\{1, 2, ..., j-1\}$ such that

$$b_{i-1} \geq \frac{\sum_{r=1}^{j} b_r}{j-i+1} =: a. \tag{1.20}$$

Let the probability distribution $c = (c_1, ..., c_d)$ be defined as

$$c_r = \begin{cases} a, & \text{for } r = i, i+1, ..., j \\ b_r & \text{otherwise.} \end{cases} \tag{1.21}$$

This vector is the same as the original vector, but where the *first* concave dent found in the Lorenz curve has been smoothed out.

Lemma 1.5 (Cicalese and Vaccaro, 2002 [10])**.** *For any $p, q \in \mathcal{P}^d$ we obtain $p \vee q$ in at most $d-1$ iterations[7] of the algorithm described in Definition 1.9 applied to the vector $\beta(p, q)$.*

Lemmas 1.4 and 1.5 conclude the proof sketch for the set $\mathcal{P}^d$ endowed with the majorization relation being a lattice, as we have found the lub and glb of any two vectors in $\mathcal{P}^d$. Working implementations of the join and meet algorithms in Python are available on the GitHub for the project[8], which consists of a small library for majorization-related tasks such as plotting Lorenz curves or applying bistochastic matrices on probability vectors. The library can be used with QuTip which is a Python quantum information library for generating quantum state representations [18].

1.3.5 Algebraic properties of the meet and join

This section is based on Ref. [9, pp. 39–41]. In any lattice, the meet and the join can also be seen as two operations[9] acting on elements of a set $\mathcal{L}$, which enjoy some useful properties.

- Commutativity: $a \wedge b = b \wedge a,\ a \vee b = b \vee a \quad \forall a, b \in \mathcal{L}$.
- Associativity: $a \wedge (b \wedge c) = (a \wedge b) \wedge c,\ a \vee (b \vee c) = (a \vee b) \vee c \quad \forall a, b, c \in \mathcal{L}$.
- Idempotency: $a \wedge a = a,\ a \vee a = a \quad \forall a \in \mathcal{L}$.
- Absorption: $a \wedge (a \vee b) = a,\ a \vee (a \wedge b) = a \quad \forall a, b \in \mathcal{L}$.

While it is fairly clear how the algorithms from Sections 1.3.3 and 1.3.4 lead to commutativity, associativity, and idempotency, it might not be as clear why they lead to the absorption rule. Recall Definition 1.7 and the interpretation of glb and lub of the meet and join respectively. Then it becomes clearer that the first absorption rule for the majorization lattice states that *the most ordered vector that is majorized by both a and some vector that majorizes a (and b) is none other than a*. Similarly, the second absorption rules states that *the least ordered vector that majorizes both a and some vector that is majorized by a (and b) is a*. Yet another way to understand it is that this boils down to Corollary 1.2.1: a is necessarily comparable to $a \vee b$ (resp. $a \wedge b$) since by definition $a \prec a \vee b$ (resp. $a \succ a \wedge b$), and when two vectors are comparable their meet (resp. join) is simply the least ordered (resp. most ordered) of the two vectors, being a in this case.

These properties are extremely helpful when working on proofs on the lattice, and will appear often in Chapters 4 and 5.

[7] Smoothing a concave dent can lead to a concave dent with the following point, so there is at most $d-1$ concave dents to smoothe.

[8] https://github.com/traaldbjerg/MajoLat

[9] This yields an equivalent definition of a lattice, which can be defined as a triple $\langle \mathcal{L}, \wedge, \vee \rangle$ with operations $\wedge, \vee$ acting on elements of $\mathcal{L}$. If they satisfy the properties listed above, an ordering $\sqsubseteq$ can then be derived from the triple by defining $a \sqsubseteq b \iff a \wedge b = a$.

1.4 Properties of the Shannon entropy on the lattice

This section is based on Refs. [10] and [11]. The Schur-concavity of the Shannon entropy, which preserves the ordering of the majorization relation, has already been discussed in Section 1.2.2. However, the Shannon entropy enjoys additional properties on the majorization lattice: *supermodularity* and *subadditivity*.

1.4.1 Supermodularity

Definition 1.10. A real-valued function f defined on a lattice $\langle \mathcal{L}, \sqsubseteq, \wedge, \vee \rangle$ is *supermodular* iff for any $a, b \in \mathcal{L}$

$$f(a \wedge b) + f(a \vee b) \geq f(a) + f(b). \tag{1.22}$$

Theorem 1.3 (Supermodularity of the Shannon entropy)**.** *The entropy function H is supermodular on the majorization lattice, and so for all $p, q \in \mathcal{P}^d$,*

$$H(p \wedge q) + H(p \vee q) \geq H(p) + H(q). \tag{1.23}$$

Note that if $p \sim q$, there is trivially equality.

This property is quite useful when working on the lattice. The original proof is long and hinges on unintuitive index tricks, however one of the first results of this master thesis is to prove a generalization of Theorem 1.3 which is applicable to a broader class of functions than only the Shannon entropy. The original proof is available in Appendix A, but we will leave our proof for Section 3.2. We will also encounter the reverse concept of *submodularity* later on.

Definition 1.11. A function f defined on a lattice $\langle \mathcal{L}, \sqsubseteq, \wedge, \vee \rangle$ is *submodular* iff for any $a, b \in \mathcal{L}$, $-f$ is supermodular. Alternatively,

$$f(a \wedge b) + f(a \vee b) \leq f(a) + f(b). \tag{1.24}$$

1.4.2 Subadditivity

Definition 1.12 (Subadditivity)**.** A real-valued function f defined on a lattice $\langle \mathcal{L}, \sqsubseteq, \wedge, \vee \rangle$ is *subadditive* iff for any $a, b \in \mathcal{L}$

$$f(a \wedge b) \leq f(a) + f(b). \tag{1.25}$$

Theorem 1.4 (Subadditivity of the Shannon entropy)**.** *The entropy function H is subadditive on the majorization lattice, so for all $p, q \in \mathcal{P}^d$*

$$H(p \wedge q) \leq H(p) + H(q). \tag{1.26}$$

One of the first results of this master thesis is to prove a stronger version of Theorem 1.4 which is applicable to a broader class of functions than only the Shannon entropy. As such, Theorem 1.4 becomes a corollary of our theorem. We will leave the proof for Section 3.3. We will also encounter the reverse concept of *superadditivity* later on.

Definition 1.13 (Superadditivity)**.** A real-valued function f defined on a lattice $\langle \mathcal{L}, \sqsubseteq, \wedge, \vee \rangle$ is *superadditive* iff for any $a, b \in \mathcal{L}$, $-f$ is subadditive. Alternatively,

$$f(a \wedge b) \geq f(a) + f(b). \tag{1.27}$$

1.4.3 Entropic distance on the lattice

Entropy can be used to define an entropic distance on the lattice, which allows us to quantify how far away two probability vectors are from each other on the lattice. We will use it quite a lot in Chapter 4 to quantify the amount of incomparability between 2 probability vectors.

Definition 1.14 (Entropic distance)**.** For all $p, q \in \mathcal{P}^d$, we define

$$d(p, q) = H(p) + H(q) - 2H(p \vee q). \tag{1.28}$$

Theorem 1.5 (Cicalese, Gargano and Vaccaro, 2013)**.** *d is a distance on $\mathcal{P}^d$, meaning that it satisfies the following properties:*

- *Symmetry: $d(p, q) = d(q, p)$,*
- *Positivity: $d(p, q) \geq 0$ with equality iff $p = q$,*
- *Triangular inequality: $d(p, q) \leq d(p, t) + d(t, q)$.*

This distance has a few non-intuitive properties, but remains a rigorous notion of distance nonetheless. We will also use a quasidistance going through the meet instead of the join.

Definition 1.15 (Meet-based entropic quasidistance)**.** For all $p, q \in \mathcal{P}^d$, we define

$$d'(p, q) = 2H(p \wedge q) - H(p) - H(q). \tag{1.29}$$

This notion of distance[10] is unfortunately not as foolproof as the first one, which goes through the join. In particular, while it does satisfy symmetry and positivity, it fails the triangular inequality in some cases. For example, consider $p, q \in \mathcal{P}^d$. One can show that in general, $d'(p, q) \nleq d'(p, p \vee q) + d'(p \vee q, q)$. We have

$$\begin{aligned} d'(p, p \vee q) + d'(p \vee q, q) &= 2H(p \wedge (p \vee q)) - H(p) - H(p \vee q) \\ &\quad + 2H((p \vee q) \wedge q) - H(p \vee q) - H(q) && (1.30) \\ &= 2H(p) - H(p) - H(p \vee q) \\ &\quad + 2H(q) - H(p \vee q) - H(q) && (1.31) \\ &= H(p) + H(q) - 2H(p \vee q). && (1.32) \end{aligned}$$

This quantity can be compared to $d'(p, q) = 2H(p \wedge q) - H(p) - H(q)$, and we have $d'(p, q) \geq d'(p, p \vee q) + d'(p \vee q, q)$ iff

$$\begin{aligned} & 2H(p \wedge q) - H(p) - H(q) \geq H(p) + H(q) - 2H(p \vee q) && (1.33) \\ \iff & 2H(p \wedge q) + 2H(p \vee q) \geq 2H(p) + 2H(q), && (1.34) \end{aligned}$$

which is always true by supermodularity, and so the triangular inequality always fails for incomparable p and q and is saturated otherwise. As such, one can not use it as a rigorous notion of distance, but it still has a few applications nonetheless.

An interpretation of these notions of distance is given in Figure 1.9. Notably, the distance d seems to induce some form of hyperbolic metric in the majorization lattice, in the sense that two segments that look identical in our geometrical representation (due to parallel lines) are not according to this notion of distance. In general, segments become larger the higher they are on the lattice. This intuition is helpful to understand some results of Chapter 4.

[10] Note that if $p \sim q$, then $d'(p, q) = d(p, q)$.

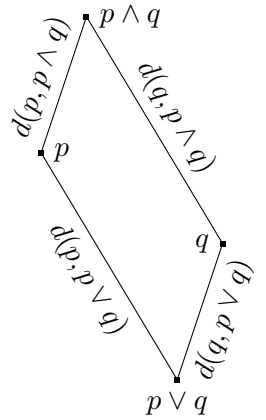


Figure 1.9: Depiction of the region (nicknamed *incomparability diamond*) $p, q, p \wedge q, p \vee q$ generated by two incomparable vectors $p, q \in \mathcal{P}^d$. Supermodularity implies that $d(p, p \vee q) = H(p) - H(p \vee q) \leq d(q, q \wedge p) = H(p \wedge q) - H(q)$, which seems to indicate some form of hyperbolic metric on the lattice where segments higher up are larger despite the parallel lines.

Finally, the following lemma will help us see another unusual property of this entropic distance, represented on Figure 1.10.

Lemma 1.6. *(Cicalese, Gargano and Vaccaro, 2013 [11]) The join $p \vee q$ of two probability vectors $p, q \in \mathcal{P}^d$ saturates the triangular inequality for the entropic distance $d(p, q) = H(p) + H(q) - 2H(p \vee q)$.*

Proof. Let $p, q \in \mathcal{P}^d$. We have

$$\begin{aligned} d(p, p \vee q) + d(p \vee q, q) &= H(p) + H(p \vee q) - 2H(p \vee (p \vee q)) + H(p \vee q) \\ &\quad + H(q) - 2H((p \vee q) \vee q) && (1.35) \\ &= H(p) + H(p \vee q) - 2H(p \vee q) + H(p \vee q) + H(q) \\ &\quad - 2H(p \vee q) && (1.36) \\ &= H(p) + H(q) - 2H(p \vee q) && (1.37) \\ &= d(p, q), && (1.38) \end{aligned}$$

where we have used associativity of successive joins to simplify the expressions. □

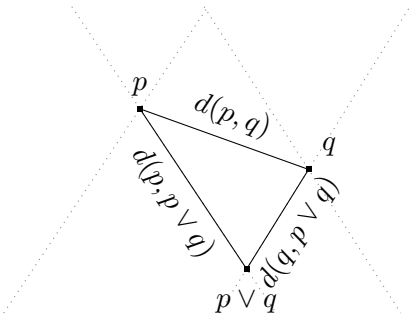


Figure 1.10: Depiction of the saturated triangular inequality $d(p, q) = d(p, p \vee q) + d(p \vee q, q)$ from Lemma 1.6 for two incomparable vectors $p, q \in \mathcal{P}^d$.

A way to interpret this result is that the distance always *goes through the join* when two vectors are incomparable, instead of measuring the "segment joining the two vectors" directly. Again, this idea is compatible with the hyperbolic intuition given above: in any euclidean metric, saying that the middle distance is as long as the sum of the two bottom ones would be absurd.

Chapter 2

Quantum information

2.1 From classical to quantum information

In this section, we will only deal with discrete random variables and probability mass functions, however these notions can be generalized to continous variables and probability density functions. Moreover, we will also assume some familiarity with quantum mechanics, such as the elementary postulates, Hilbert spaces, Dirac notation, and superpositions.

2.1.1 Information and entropy

This section is based on Ref. [8, pp. 1–56] and Ref. [19, pp. 500–506]. Classical information theory is concerned with the transmission and compression of information. For example, it provides bounds for the best possible lossless compression of information, as well as for the most amount of information that can be passed through a noisy channel without losses (asymptotically). Information, being a fairly abstract concept, can be hard to define. A generally accepted vision of information is to see it as the resolution of *uncertainty* contained in a random variable X *before* measurement.

Let us illustrate this notion of uncertainty with a simple example. Say you have a friend who just finished taking a statistics exam with a very simple marking grid: you can either get 0, 10 or 20 out of 20. Having already passed statistics, you use data from the previous years to determine that the distribution for these marks is 70%, 20%, and 10% respectively for teacher 1 (who isn't very nice). The question is the following: what is the most efficient way on average to ask your friend what mark they got? An efficient first question would be whether or not they got 0/20. There is then a 70% chance that the answer is yes, and you would not need to ask a second question, which is quite efficient. If the answer is no, a second efficient question would be whether they got 10/20, etc.

Now, imagine that they took the exam with teacher 2 who is much nicer, and actually gives those same marks with an equal probability of 1/3 instead. In that case, you could once again start by asking whether they got 0, but this time the answer would be no 2/3 of the time, and you would be much more likely to have to ask follow-up questions. In this sense, the mark of your friend is more uncertain *from the start* under teacher 2 than under teacher 1.

Similarly, this intuition of the number of binary questions can be though of as the *average* number of bits of the *shortest* representation of the outcome of a random process, i.e. the average number of binary answers. The idea this time is that it is efficient to assign a short description to an outcome that happens often. This notion of uncertainty can be formalized with the notion of Shannon entropy.

Definition 2.1 (Shannon entropy)**.** Let X be a random variable over an alphabet $\mathcal{X}$, with outcome $x \in \mathcal{X}$ having probability $p(x)$. Then, the entropy $H(X)$ (also written $H(p)$) of X is

defined as

$$H(X) = -\sum_{x\in\mathcal{X}} p(x)\log(p(x)) \tag{2.1}$$

$$= -E[\log p(x)], \tag{2.2}$$

with the convention that $0\log 0 = 0$ (an event that never happens should not contribute).

Usually, the logarithm is taken in base 2, and the entropy is said to be measured in bits. Less commonly, the logarithm can be taken in base e, and the entropy is then in units of nats. It is interesting to note that this definition is very consistent, as the vision of entropy being the expected value of the random variable $\log\frac{1}{p(x)}$ holds even for compositions of random variables, which Appendix B quickly goes over. Going back to our statistics exam example, the marks from teacher 1 and 2 are distributed along $p = (0.7, 0.2, 0.1)$ and $q = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$ respectively. We have $H(p) = -3\cdot\frac{1}{3}\log\frac{1}{3} = \log 3 = 1.585$ bits, and $H(q) = -\frac{7}{10}\log\frac{7}{10} - \frac{2}{10}\log\frac{2}{10} - \frac{1}{10}\log\frac{1}{10} = 1.157$ bits, and so a less certain process has higher entropy.

Remark 2.1. This is consistent with the intuition that we built in Chapter 1: the first distribution is more uncertain than the second because $p \prec q$. By Schur-concavity of H we directly get $H(p) \geq H(q)$.

Remark 2.2. Let X be a random variable with d possible outcomes. Then, the flat distribution has maximal entropy: $H(X) \leq H\left(\overline{1}_d\right) = \log d$. Again, this well-known fact from information theory ties in nicely with the fact that the flat distribution is majorized by every other distribution.

Remark 2.3. The entropy of a random variable X only depends on its probability distribution, not on the actual values associated with specific outcomes.

Recall the question that we asked above: what is the average length of the shortest description of a random process distributed along p? It turns out that the answer to this question is $H(p)$. A naive idea would be to assign the code 0 if our friend got 0/20, 1 if they got 10/20, and 01 if they got 20/20. Then, in the case of teacher 1, on average, we expect to use $7/10+2/10+2\cdot 1/10 = 1.1$ bits $< H(p)$. Why did we beat the Shannon entropy with this simple code ? The answer is that our code is not uniquely decodable (which is a hypothesis of the theory), and is thus not a very good code. Imagine we receive the string 0100 for the results of several students. Are the scores 20/20, 0/20, 0/20, or are they 0/20, 10/20, 0/20, 0/20?

A better idea would be to use codes such that no code is a prefix for another code, such as 0 for 0/20, 10 for 10/20, and 11 for 20/20. This is called a Huffman code, which can be proven to be the optimal family of codes [8, p. 123]. In this case, we expect to use $7/10+2\cdot 2/10+2\cdot 1/10 = 1.3$ bits on average, which is close to $H(p)$ with a small overhead due to the constraint of an integer representation. However, we can get arbitrarily close to $H(p)$ per symbol on average if we code larger strings [8, p. 114]. This is perhaps the most intuitive interpretation of entropy. As such, one should be careful with the fact that entropy is exact only in the asymptotic limit. Moreover, underlying hypotheses in the theory can sometimes be broken in practice[1].

2.1.2 Density matrices

This section is based on Ref. [19, pp. 98–108]. Extending the notion of information to a quantum setting, where states can be in a superposition of several states depending on the basis choice, is not a trivial question. In order to do this, we first need to go through a few mathematical preliminaries.

[1] A famous example is the 56 kbit/s modem which broke the theoretical channel capacity for analog modems by cleverly using an asymmetrical digital/analog channel and thus changing the signal/noise ratio [8].

For the rest of this master thesis, we will be considering computational states of a qubit, $|0\rangle$ and $|1\rangle$. The actual implementation of those states matters little for our purposes, be it an optical implementation with vertically and horizontally polarized photons, the spin of an ion vacancy in a crystalline lattice, phonon vibrations of ions in an ion trap, ... It is useful to define the *dual basis* of the computational states of a qubit. The basis vectors read

$$|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \tag{2.3}$$

$$|-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle). \tag{2.4}$$

The most general notion of quantum state is the notion of density operator, which captures both superpositions and classical uncertainty. Imagine that we have a device that internally flips a fair coin and prepares the state $|0\rangle$ if the result is heads, and the state $|1\rangle$ otherwise. However, the device does not tell you the result of the coin flip, and so you do not know whether it prepared one state or the other. Let us call this state ρ. It is tempting to assign $|\psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) = |+\rangle$ (the relative phase is set to 0 for simplicity's sake) to our unknown state, however it does not capture the physics of our state correctly: if we sent ρ through an interferometer (assuming an optical implementation), the state would not interfere with itself, whereas a state in a superposition would (like $|\psi\rangle$). A description in terms of state vectors is therefore not sufficient, as we need to be capable of capturing the classical uncertainty introduced by the coin flip[2].

Definition 2.2 (Density operator)**.** Suppose that a quantum system is in a state $|\psi_i\rangle$ with probability p_i. The set $\{p_i, |\psi_i\rangle\}$ is called an *ensemble of pure states*. The density operator ρ (also called density matrix) for the system is defined as

$$\rho = \sum_i p_i \, |\psi_i\rangle \, \langle\psi_i| \, . \tag{2.5}$$

If the sum contains only one term, the state is said to be *pure*, i.e. one can describe it with a single state vector. Otherwise, the state is said to be *mixed*, because it can be seen as a *classical mixture* of pure states. We will also use the convention that when we speak of nondescript mixed states, we explicitly mean states that are not pure (for some authors mixed states include pure states). The two states mentioned in our coin-flip example can be written

$$\rho = 1/2 \, |0\rangle \, \langle 0| + 1/2 \, |1\rangle \, \langle 1| = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}, \tag{2.6}$$

$$\rho' = |\psi\rangle \, \langle\psi| = |+\rangle \, \langle +| = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)\frac{1}{\sqrt{2}}(\langle 0| + \langle 1|) = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}. \tag{2.7}$$

Here, ρ is a mixed state, but ρ' is a pure state. The fact that their density matrices are not the same is thus a first sign that we have captured the difference in their behavior.

It is possible to fully forego the notion of state vectors in the definition of density operators. In doing so, it turns out that for an operator O to be a valid density operator, the only requirements are to be a positive operator, and to have $\operatorname{tr} O = 1$. In order to use this new description for quantum states and to understand that it can correctly capture this new type of behavior, the postulates of quantum mechanics must be reformulated in the language of density operators, in order to define how measurements and unitaries act on such states.

Postulate 1. The state of an isolated physical system is completely described by its density operator, acting on the state space of the system. If the system is in state ρ_i with probability p_i, then the density operator for the system is $\sum_i p_i \rho_i$ (a convex combination of density operators is still a density operator).

[2]It is actually still possible to capture this phenomenon correctly with state vectors by entangling the output state vector with a virtual environment, in a process called *purification* (of the mixed state) [19, p. 110].

Postulate 2. The time evolution of a closed quantum system is described by unitary transformation. The state of the system at time t_2, $\rho(t_2)$, is related to the state of the system at time t_1, $\rho(t_1)$, according to the following expression

$$\rho(t_2) = U\rho(t_1)U^\dagger, \tag{2.8}$$

where U is a unitary operator which in general depends on t_1 and t_2.

Postulate 3. Quantum measurements are described by a set of measurement operators $\{M_m\}$ which act on the state space of the system, where the index m labels the possible measurement ouctomes. If the system is in the state ρ at time of measurement, then the probability of result m is[3]

$$p(m) = \operatorname{tr}(M_m^\dagger M_m \rho), \tag{2.9}$$

and the system is left in the state

$$\rho' = \frac{M_m \rho M_m^\dagger}{\operatorname{tr}(M_m^\dagger M_m \rho)}. \tag{2.10}$$

Moreover, the measurement operators must satisfy

$$\sum_m M_m^\dagger M_m = I, \tag{2.11}$$

which is called the completeness relation.

Postulate 3 is illustrated with our coin-flip example in Appendix C. The most striking difference between a pure state and a mixed state is that for a pure state, there always exists a choice of measurement operators under which the outcome can be guaranteed, i.e. one of the outcomes has probability 1. However, for a mixed state, no matter the choice of measurement operators, the outcome cannot be guaranteed. In this sense, mixed states are fundamentally uncertain.

2.1.3 Von Neumann entropy

This section is based on Ref. [19, pp. 510–527]. With the notion of density operator, quantum analogues for information-theoretic quantities can be defined. In classical information theory, the Shannon entropy describes the uncertainty of a probability distribution. Instead of a probability distribution, however, we are working with a density operator, which contains the probabilities to be in each state. It is thus quite natural to attempt to use the same definition for these density operators.

Definition 2.3 (Von Neumann entropy)**.** Let ρ be a density operator. The von Neumann entropy $S(\rho)$ of this quantum state is defined as

$$S(\rho) = -\operatorname{tr}(\rho \log \rho), \tag{2.12}$$

where the logarithm of an operator O is another operator $\log O$ such that its exponential $e^{\log O} = O$, and the exponential of an operator is defined with the usual power series.

This definition is a direct generalization of the Shannon entropy to density matrices. We will not prove them, but it turns out that it has all of the properties one would expect from a measure of uncertainty. Two interesting properties are

- $S(\rho) \geq 0$, with equality iff ρ is a pure state.
- In a d-dimensional Hilbert space, $S(\rho) \leq \log d$, with equality iff $\rho = \frac{I}{d}$ (where I is the identity operator), which is called the *fully mixed state*.

These properties show an interpretation of von Neumann entropy: $S(\rho)$ quantifies how mixed a state is. If the outcome of measuring ρ can be made certain by choosing a specific measurement basis (and so ρ is pure), then ρ contains no uncertainty, and $S(\rho) = 0$.

[3]Note that the trace is cyclic, and so $\operatorname{tr}(M_m^\dagger M_m \rho) = \operatorname{tr}(M_m \rho M_m^\dagger)$, but the former expression is used because it sometimes simplifies calculations thanks to matching bras and kets.

2.2 Theory of entanglement

2.2.1 Composite systems

This section is based on Ref. [19, pp. 71–75, 93–96]. In quantum mechanics, state vectors describe the state of a system. However, it is sometimes useful to describe composite systems made of smaller subsystems, each in their own specific states. For example, say that we have two systems in states $|\psi\rangle_A$ and $|\phi\rangle_B$, where the indices A and B denote each subsystem, usually nicknamed Alice's system and Bob's system. Note that the 2 systems need not span the same Hilbert space, i.e. $\mathcal{H}^A = \mathcal{H}^B$ is not necessarily true, which means that the systems can be of different natures and different dimensions (e.g. Alice could hold a qubit and Bob a qutrit). The joint system AB, called a *bipartite system* because it is made up of two subsystems[4], is a valid quantum system as well, and so one would expect that it can be fully described by the states of the individual subsystems $|\psi\rangle_A$ and $|\phi\rangle_B$. This description is done using tensor products, which are a way of building larger vector spaces from smaller vector spaces in a way that still preserves the structure of the individual smaller vector spaces.

Definition 2.4 (Tensor product of vector spaces)**.** Let V and W be two vector spaces, of dimension m and n respectively. The tensor product of V and W, written $V \otimes W$, is a mn-dimensional vector space, whose vectors are tensor products $|v\rangle \otimes |w\rangle$ (often abbreviated $|v\rangle |w\rangle, |v, w\rangle$ or $|vw\rangle$) of elements $|v\rangle \in V$ and $|w\rangle \in W$. If $|i\rangle$ and $|j\rangle$ are orthonormal bases of V and W, then $|i\rangle \otimes |j\rangle$ forms an orthonormal basis for $V \otimes W$ too.

This notion of tensor product is used for an additional postulate on composite systems.

Postulate 4. Let $\mathcal{H}^i$ be the Hilbert space of subsystem i, numbered 1 through n. Then, the Hilbert space of the composite system $\mathcal{H}^c$ is the tensor product of the Hilbert spaces of the individual subsystems, i.e. $\mathcal{H}^c = \mathcal{H}^1 \otimes \mathcal{H}^2 \otimes \cdots \otimes \mathcal{H}^n$. Moreover, if each subsystem i is prepared in state ρ_i, then the joint system is in state $\rho_1 \otimes \rho_2 \otimes \cdots \otimes \rho_n$.

This formal definition and this postulate are perhaps better understood with a simple example. Imagine Alice holds a qubit in the state $|+\rangle_A = \frac{1}{\sqrt{2}}(|0\rangle_A + |1\rangle_A)$, and Bob holds a qubit in the state $|-\rangle_B = \frac{1}{\sqrt{2}}(|0\rangle_B - |1\rangle_B)$, where the system indices are added for now to keep track of the systems. $\mathcal{H}^A$ and $\mathcal{H}^B$ can both be described in a computational basis $|0\rangle_i, |1\rangle_i$ with $i = A, B$. Then, the Hilbert space of the joint system $\mathcal{H}^{AB}$ has a basis $|0\rangle_A \otimes |0\rangle_B, |0\rangle_A \otimes |1\rangle_B, |1\rangle_A \otimes |0\rangle_B, |1\rangle_A \otimes |1\rangle_B$, or in shorthand notation $|00\rangle_{AB}, |01\rangle_{AB}, |10\rangle_{AB}, |11\rangle_{AB}$. The composite description is thus quite intuitive: the joint system is simply described by the state of each subsystem, e.g. a joint state $|01\rangle_{AB}$ simply means that system A is in state $|0\rangle_A$ and system B is in state $|1\rangle_B$. In our slightly more complicated example with the $|+\rangle$ and $|-\rangle$ states, the joint state becomes

$$|+-\rangle_{AB} = \frac{1}{2}(|0\rangle_A + |1\rangle_A) \otimes (|0\rangle_B - |1\rangle_B) = \frac{1}{2}(|00\rangle_{AB} - |01\rangle_{AB} + |10\rangle_{AB} - |11\rangle_{AB}). \quad (2.13)$$

2.2.2 Partial trace

This section is based on Ref. [19, pp. 105–109]. From now on, we will omit subsystem indices, as they should be clear from context. Imagine that we have the AB system in the state $|\Phi^+\rangle := \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. What is an accurate description of Alice's qubit from her point of view, if she has no knowledge of Bob's qubit, i.e. what measurement statistics will she observe if she

[4] We will mostly concern ourselves with bipartite systems, but of course one could also decide to consider that a system is made of more subsystems than 2. We can sometimes still use a bipartite description because an ABC tripartite system can also be considered a bipartite system with subsystems A and BC (which is itself a bipartite system), etc.

measures her qubit ? Will she have a coherent superposition $|+\rangle$, or will she have a mixed state $\frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|$, which cannot interfere with itself? In order to answer this question, we will have to go back to a density operator representation to make sure that we can capture both superpositions and classical mixtures. We are essentially looking to build a reduced density operator ρ^A from the composite density operator ρ^{AB}, by discarding system B which Alice has no knowledge over. This is done with an operation called the *partial trace*.

Definition 2.5 (Partial trace)**.** The partial trace over a subsystem B is a map of operators on a joint system AB, defined as

$$\mathrm{tr}_B(|a_1\rangle\langle a_2| \otimes |b_1\rangle\langle b_2|) \coloneqq |a_1\rangle\langle a_2| \,\mathrm{tr}(|b_1\rangle\langle b_2|), \tag{2.14}$$

where $|a_1\rangle$ and $|a_2\rangle$ are any two elements of the state space of subsystem A, and $|b_1\rangle$ and $|b_2\rangle$ are any two elements of the state space of subsystem B, and where the trace operation on the right-hand side (RHS) is the usual trace for system B. The system B is said to be *traced out*.

Essentially, the partial trace maps the joint operator $|a_1\rangle\langle a_2| \otimes |b_1\rangle\langle b_2|$ acting on $\mathcal{H}^{AB}$ to a new operator $|a_1\rangle\langle a_2| \,\mathrm{tr}(|b_1\rangle\langle b_2|)$, which only acts on the Hilbert space of the subsystem A, $\mathcal{H}^A$. Using this partial trace, the reduced density operators can be defined.

Definition 2.6 (Reduced density operator)**.** Let AB be a joint system in state ρ^{AB}. The reduced density operator of subsystem A is defined as

$$\rho^A = \mathrm{tr}_B(\rho^{AB}). \tag{2.15}$$

While it may not be obvious from the definition, the reduced density operator does lead to the correct measurement statistics on each subsystem, signalling that it is an accurate description of their state. We will not do it here, but this can be proven by studying the effects of measurement operators on the joint and reduced density operators. However, we can convince ourselves with a simple example that the reduced density operators work as expected. Let $\rho^{AB} = \rho^A \otimes \rho^B$, then we have

$$\begin{aligned} \mathrm{tr}_B \rho^{AB} &= \mathrm{tr}_B(\rho^A \otimes \rho^B) && (2.16)\\ &= \rho^A \,\mathrm{tr}(\rho^B) && (2.17)\\ &= \rho^A, && (2.18) \end{aligned}$$

which is the expected result. Similarly, tracing out A gives ρ^B. While this may seem trivial, these notions become much more interesting when studying states that can not be described as the tensor product of subsystems. In fact, the previous example $|\Phi^+\rangle \coloneqq \frac{1}{2}(|00\rangle + |11\rangle)$ was not chosen innocently: it is such a state. Even though it lives in the $\mathcal{H}^A \otimes \mathcal{H}^B$ Hilbert space, there is no basis choice for $\mathcal{H}^A$ and $\mathcal{H}^B$ which turns this state into a product state, i.e. a state such that $\rho^{AB} = \rho^A \otimes \rho^B$. In particular, tracing out system B, we get

$$\begin{aligned} \mathrm{tr}_B |\Phi^+\rangle\langle\Phi^+| &= \frac{1}{2}\,\mathrm{tr}_B\left((|00\rangle + |11\rangle)(\langle 00| + \langle 11|)\right) && (2.19)\\ &= \frac{1}{2}\,\mathrm{tr}_B\Big[\,|00\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 00| + |11\rangle\langle 11|\,\Big] && (2.20)\\ &= \frac{1}{2}\,\mathrm{tr}_B\Big[\,|0\rangle\langle 0| \otimes |0\rangle\langle 0| + |0\rangle\langle 1| \otimes |0\rangle\langle 1| + |1\rangle\langle 0| \otimes |1\rangle\langle 0| \\ &\qquad + |1\rangle\langle 1| \otimes |1\rangle\langle 1|\,\Big] && (2.21)\\ &= \frac{1}{2}\Big[\,|0\rangle\langle 0| \,\mathrm{tr}(|0\rangle\langle 0|) + |0\rangle\langle 1| \,\mathrm{tr}(|0\rangle\langle 1|) + |1\rangle\langle 0| \,\mathrm{tr}(|1\rangle\langle 0|) \\ &\qquad + |1\rangle\langle 1| \,\mathrm{tr}(|1\rangle\langle 1|)\Big] && (2.22) \end{aligned}$$

$$= \frac{1}{2}\Big[\, |0\rangle\langle 0| + |1\rangle\langle 1| \,\Big] \tag{2.23}$$

$$= \frac{I}{2}, \tag{2.24}$$

which is the maximally mixed state. Similarly, tracing out system A also gives $I/2$. This result is quite counterintuitive: the joint system is in a perfectly known state (a projective measurement of AB on the state $|\Phi^+\rangle$ succeeds with probability 1), yet Alice's qubit is in an uncertain state from her point of view! This surprising property is a sign of *entanglement*.

2.2.3 Entangled states

This section is based on Refs. [4] and [20]. An entangled state is a composite state that is not a product state, i.e. it can not be written $\rho^{AB} = \rho^A \otimes \rho^B$. The name stems from the fact that only a global description captures the state without additional uncertainty, and so the state of each subsystem is linked to the state of the other subsystem even if they are spatially separated. Analogously, product states are also called *separable* states, because studying separated subsystems (tracing one out) does not introduce uncertainty into the reduced states. It is useful to define the *Bell basis* for the joint Hilbert space of two qubits. The basis vectors read

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \tag{2.25}$$

$$|\Phi^-\rangle = \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle) \tag{2.26}$$

$$|\Psi^+\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle) \tag{2.27}$$

$$|\Psi^-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle), \tag{2.28}$$

which are all entangled states.

Any one of these states is often called an Einstein-Podolsky-Rosen (EPR) pair, because those authors famously were the first to use the predicted properties of entangled[5] states to argue that they were so absurd that they deemed them impossible, indicating that quantum mechanics must then be an incomplete theory because it theoretically allows entangled states [1]. Since then such states have been realized experimentally, disproving the EPR argument, but the name stuck, perhaps to honor their thought experiment which sparked so much debate and discoveries. Two fundamental properties are the following.

1. **Quantum nonlocality:** Bell's inequalities cannot be violated without entanglement. Note that all entangled states do not necessarily violate Bell's inequalities, e.g. Werner states (mixed states of the form $p\,|\Psi^-\rangle\langle\Psi^-| + (1-p)\frac{I}{p}$) showcase entanglement but can admit a Local Hidden-Variable Model (LHVM) or not, depending on the value of p. However, the entangled states that do violate the inequalities do not admit a LHVM, and are inherently nonlocal.

2. **Monogamy:** if systems A and B are maximally[6] entangled, then A and B cannot be entangled at all with a third party C. This effect is strictly quantum: if two systems A and B are classically correlated, nothing prevents them from having the same correlations with system C.

These properties are counterintuitive. Let us explore go through an example. Let us assume that Alice and Bob share an EPR pair $|\Phi^+\rangle$. Imagine that Bob measures his half of the entangled

[5] Although the term *entangled* was only coined by Schrödinger a few months later.
[6] This notion is formalized in Section 2.3.3.

pair in the computational basis $|0\rangle, |1\rangle$, and measures his state to be $|1\rangle$. Then the joint state has been reduced to $|11\rangle$. Now, if Alice measures her qubit after Bob's measurement, she will also measure her qubit to be in the state $|1\rangle$, showcasing the inherent nonlocality of such a state: Bob influenced Alice's qubit by exploiting the joint quantum state. What is even more striking is that quantum mechanics does not predict a delay for the reduction of the joint state: the effect is *instantaneous*.

With such processes, can Alice immediately know the result of Bob's measurement, enabling faster-than-light communication ? The key that saves causality is that until Bob tells Alice that he has measured his qubit, if Alice measures her half of the pair she cannot know whether she was the one responsible for breaking the superposition with her measurement, or whether Bob was. Moreover, since the (classical) message from Bob is bound by the speed of light, no information can reach Alice faster than light, and causality is preserved.

In the last decades, the properties of entangled states have been of great interest as they enable powerful applications, such as post-quantum cryptography, secure key distribution, quantum teleportation, dense coding, ... This master thesis does not focus on these specific applications, and mostly focuses on entangled state manipulation in general. It is however interesting to note that manipulating entangled states in clever ways is of very high interest for all of these applications.

The surprising property of entanglement that a joint state can be more certain globally than locally can be formalized into a theorem.

Theorem 2.1 (Entropic separability criterion [21, 22])**.** *Let ρ^{AB} be the state of a bipartite system AB, and ρ^A and ρ^B the reduced states of subsystems A and B. Then, ρ^{AB} is an entangled state if*

$$S(\rho^{AB}) < S(\rho^A) \quad or \quad S(\rho^{AB}) < S(\rho^B). \tag{2.29}$$

This theorem can be understood fairly intuitively from our previous remark. It simply states that a state is entangled if the joint state is more certain than the reduced states, i.e. not having knowledge over one subsystem prevents you from having certainty on the other subsystem. It is important to note however that the implication only goes one way, and so some lightly entangled states do not satisfy the entropic inequality.

2.2.4 Schmidt decomposition

This section is based on Ref. [19, pp. 109–111]. A powerful tool to describe entangled states is the *Schmidt decomposition*.

Theorem 2.2 (Schmidt decomposition)**.** *Let $|\psi\rangle$ be the state of a composite system AB. Then, there exist orthonormal bases $|i_A\rangle$ and $|i_B\rangle$ for subsystems A and B respectively such that*

$$|\psi\rangle = \sum_i \sqrt{\lambda_i}\, |i_A\rangle\, |i_B\rangle\,, \tag{2.30}$$

where the λ_i are real numbers called Schmidt coefficients which satisfy $\sum_i \lambda_i = 1$, and where the RHS is called the Schmidt decomposition of $|\psi\rangle$.

A related quantity is the *Schmidt number* (or *Schmidt rank*) of the decomposition of a joint state $|\psi\rangle$, defined as the number of non-zero Schmidt coefficients of the decomposition. We will also often put the non-zero Schmidt coefficients of a joint state $|\psi\rangle$ in a vector $\lambda = (\lambda_1, \ldots, \lambda_n) \in \mathcal{P}^n$ for a state of Schmidt rank n. We will call this vector the *Schmidt vector* of $|\psi\rangle$.

Corollary 2.2.1. *If $|\psi\rangle$ is a product state, then its Schmidt rank is 1.*

This corollary directly implies that entangled states all have non-trivial Schmidt vectors. Moreover, the Schmidt vector has an interesting interpretation in terms of information theory, which will be made clearer with the following corollary.

Corollary 2.2.2. *Let $|\psi\rangle$ have Schmidt vector λ. Then $\{\lambda_i\}$ is the set of eigenvalues of the reduced density matrices of both Alice and Bob.*

Proof.

$$|\psi\rangle = \sum_i \sqrt{\lambda_i}\, |i_A\rangle \otimes |i_B\rangle \tag{2.31}$$

$$\implies \rho_\psi = \sum_{i,j} \sqrt{\lambda_i \lambda_j}\, |i_A\rangle \langle j_A| \otimes |i_B\rangle \langle j_B| \tag{2.32}$$

$$\implies \mathrm{tr}_B(\rho_\psi) = \sum_{i,j} \sqrt{\lambda_i \lambda_j}\, |i_A\rangle \langle j_A| \,\mathrm{tr}(|i_B\rangle \langle j_B|) \tag{2.33}$$

$$= \sum_{i,j} \sqrt{\lambda_i \lambda_j}\, |i_A\rangle \langle j_A|\, \delta_{ij} \tag{2.34}$$

$$= \sum_i \lambda_i\, |i_A\rangle \langle i_A|\,, \tag{2.35}$$

the tr_A case being symmetric. □

As we have seen before, the reduced state of a perfectly known joint state can be mixed, meaning that the result of measuring it is uncertain. Let Alice and Bob share an entangled state $|\psi\rangle$ with Schmidt vector λ. How much uncertainty on Bob's system does Alice resolve by assigning a local description to her half of the state, i.e. by measuring it ? Using our information-theoretic intuition, we can interpret the uncertainty to be $H(\lambda)$ (in the asymptotic limit), with λ the probability distribution for each local state. This means that Alice needs on average $H(\lambda)$ yes/no measurements in her half of the Schmidt basis to determine the state Bob holds (and her own), with an additional overhead due to an integer constraint (which vanishes as the number of measured systems increase).

2.2.5 Majorization separability criterion

As we have seen in Section 1.2.2, majorization relations directly imply entropic inequalities. As such, an entropic inequality can sometimes be the sign of an underlying majorization relation, which we will call a *majorization precursor*[7]. Finding majorization precursors to existing inequalities is an active research domain. In the case of entanglement, the entropic criterion in Theorem 2.1 can be strengthened to a majorization criterion.

Theorem 2.3 (Majorization separability criterion [23])**.** *Let ρ^{AB} be the state of a bipartite system AB, ρ^A and ρ^B the reduced states of subsystems A and B, and λ^{AB}, λ^A and λ^B are their vectors of eigenvalues, sorted in decreasing order. Then, ρ^{AB} is an entangled state if*

$$\lambda^{AB} \not\prec \lambda^A \quad \textit{or} \quad \lambda^{AB} \not\prec \lambda^B. \tag{2.36}$$

The interpretation of this theorem is identical to the entropic version, but stated with more sophisticated tools to characterize disorder. Once again, the implication only goes one way and so this criterion cannot tell some lightly entangled states apart from separable states. However, this theorem is still a strengthening of the entropic separability criterion, which covers cases that the entropic criterion could not conclude on.

This theorem is a first glimpse of the power of majorization as a tool in quantum information. Other powerful applications exist, notably in the field of entanglement transformations, which are of high interest in distributed quantum computation and quantum communications.

[7] This is not a standard denomination.

2.3 Entanglement transformations

2.3.1 Quantum Resource Theories

This section is based on Ref. [12]. Resource theories are originally economic theories, stemming from the simple idea that an object is more valuable if it is difficult to acquire. Resource theories, then, are a theory of *interconversions* between objects, e.g. if an object is worth more money than another, you can sell it to acquire the second. Essentially, a resource theory consists of 2 concepts from which resources can be defined.

- Free operations: set of permissible operations in the theory, each mapping a set of states to another set of states, e.g. sell an object for its monetary value and buy back another object of lower monetary value (the reverse operation is not necessarily achievable and is thus not a free operation).
- Free states: set of states that can be accessed from any state in the theory using free operations, e.g. any object that can bought by selling any other object in the theory (which would be an object of value €0).

A state is said to contain at least as much resource as a second state if it can be converted into the second state using the free operations of the theory, i.e. anything that can be done with the second state can be done with the first. We will often say that the second state is *reachable* from the first. In the monetary example, the resource of an object could simply be quantified by its monetary value in €.

Resource theories have been applied in many fields, including quantum information. In a quantum setting, a resource theory is called a *Quantum Resource Theory* (QRT), involving objects and processes at atomic and subatomic levels. Most notably, the theory of entanglement is a QRT, but other examples exist, such as the theories of quantum coherence, quantum thermodynamics or non-Gaussianity in bosonic systems. Just like in the monetary example, an entanglement QRT allows us to say that one state is more entangled than a second state if it can be transformed into the second using the free operations of the theory.

Let us define the notions of *free operations* and *free states* more formally, which are defined on Hilbert spaces for a QRT.

Definition 2.7 (Quantum Resource Theory)**.** Let $\mathcal{O}$ be a mapping that assigns to any two input/output physical systems A and B, with corresponding Hilbert spaces $\mathcal{H}^A$ and $\mathcal{H}^B$, a unique set of completely positive and trace-preserving (CPTP) operations $\mathcal{O}(A \to B) \coloneqq \mathcal{O}(\mathcal{H}^A \to \mathcal{H}^B) \subset \mathcal{Q}(A \to B)$. Let $\mathcal{F}$ be the induced mapping $\mathcal{F} \coloneqq \mathcal{O}(\mathbb{C} \to \mathcal{H})$, where $\mathcal{H}$ is an arbitrary Hilbert space. Then the tuple $\mathcal{R} = (\mathcal{F}, \mathcal{O})$ is called a *Quantum Resource Theory* if the following two conditions hold:

1. For any physical system A the set $\mathcal{O}(A \to A)$ contains the identity map $\mathtt{id}^A$.
2. For any three physical systems A, B and C, if $\Phi \in \mathcal{O}(A \to B)$ and $\Lambda \in \mathcal{O}(B \to C)$, then $\Lambda \circ \Phi \in \mathcal{O}(A \to C)$.

In a QRT, the set $\mathcal{F}(\mathcal{H}) \subset \mathcal{S}(\mathcal{H})$ defines the set of *free states* acting on $\mathcal{H}$, and the elements belonging to $\mathcal{S}(\mathcal{H}) \backslash \mathcal{F}(\mathcal{H})$ are called *resource states*. Likewise, the CPTP maps in $\mathcal{O}(A \to B)$ are called *free operations*, and the CPTP maps that are not in $\mathcal{O}(A \to B)$ are called *dynamical resources*.

Property 2 makes preorders arise quite naturally in QRTs. Let us define $\alpha \xrightarrow{\mathcal{O}} \beta$ as meaning "β can be reached from α using free operations from $\mathcal{O}$". Then, for any $\rho, \sigma, \gamma \in \mathcal{S}(A)$, $\rho \xrightarrow{\mathcal{O}} \sigma$ and $\sigma \xrightarrow{\mathcal{O}} \gamma \implies \rho \xrightarrow{\mathcal{O}} \gamma$. However, for any $\rho, \sigma \in \mathcal{S}(A)$, $\rho \xrightarrow{\mathcal{O}} \sigma$ and $\sigma \xrightarrow{\mathcal{O}} \rho$ does not

necessarily mean $\rho = \sigma$, so in general we only have a preorder and not a full partial order, e.g. if an apple and an orange were worth the same amount of money (and so could be converted back and forth by selling and rebuying), it wouldn't mean that an apple is an orange. Another useful concept tied to reachability is the notion of *resource monotone*.

Definition 2.8 (Resource monotone)**.** A resource monotone is a function $F : \mathcal{H} \to \mathbb{R}$ such that

$$\rho \xrightarrow{\mathcal{O}} \sigma \implies F(\rho) \geq F(\sigma). \tag{2.37}$$

Essentially, a resource monotone is a function of the states of the theory that can only decrease under free operations. It is thus a way of quantifying the amount of resource contained in different states. Notice however that the implication only goes one way, because some states may be incomparable under the preorder $\xrightarrow{\mathcal{O}}$.

Remark 2.4. While the monetary example is helpful in understanding the basic idea of resource theory, it does not fully capture the richness of such theories. This is essentially because the value of an object is directly a resource monotone which can always be compared between states, and so the preorder $\xrightarrow{\text{sell/buy}}$ is promoted to a *total order*. In general, additional richness arises in theories where $\xrightarrow{\mathcal{O}}$ remains a preorder, because some states can then be incomparable, holding a similar amount of resource, each in a different way. If the preorder supports a meet and a join, a lattice can then be defined.

The definition of a QRT may be quite formal, but the game here is essentially to find the set of free operations $\mathcal{O}(A \to B)$ that correctly captures the possible physical processes we can apply to the states in our possession. In general, starting from the free operations is the more natural way to proceed, as they give a single set of free states associated to the operations. The opposite approach is sometimes applied, for example in the theory of quantum coherence, but has the disadvantage that several different choices of free operations can lead to the same free states, and so a choice for the set of free operations must be made.

The definition of QRT is perhaps best understood with an example. In the theory of entanglement such an example arises when we confine ourselves to a practically relevant set of operations, called *Local Operations and Classical Communications* (LOCC), which is at the center of this master thesis, along with the majorization lattice.

2.3.2 Local Operations and Classical Communications

This section is based on Ref. [12, pp. 19–21]. For the rest of this master thesis, we assume a bipartite setting, and that Alice and Bob both possess a quantum computer capable of applying arbitrary[8] *Local Operations* (LO) to the quantum states in their possession, i.e. of applying a unitary only on their reduced state but not on the full joint state, which may be entangled[9]. The actual implementation of the computer and of those states matters is not particularly important to us. Moreover, we will also assume that Alice and Bob can use a Classical Communications (CC) channel to send messages to each other (which they can use to decide collaboratively on which local unitaries to apply on each end). Hence, the paradigm we are working with is called *Local Operations and Classical Communications* (LOCC). We will also limit ourselves to pure states, because mixed states are by definition noisy states that are a liability in quantum computations, i.e. any usable quantum computer should have low enough decoherence rates that the qubit states can be considered pure.

[8] To be more rigorous, one should say capable of approaching to arbitrary precision any given unitary.

[9] How Alice and Bob came into possession of shared entangled states matters little for our purposes. They could either have met physically, made some of their qubits interact in a way that entangled them, then separated their computers physically again. Alternatively, entangled photons coming from a nondescript source (see generation techniques in Section II.B.7 of Ref. [24]) could be sent separately to the two parties using a QC channel, and stored. While more convenient logistically, the second approach requires a QC channel. Both approaches require a good quantum memory (i.e. capability of storing the qubit in its state without losing it due to noise).

That this physical setting is relevant is not surprising, as classical communication channels are readily available with modern technological standards, whereas Quantum Communication (QC) channels are still in very early development stages. As such, it is reasonable to limit ourselves to the information-related tasks achievable using LOCC. LO and Quantum Communications[10] is a more powerful paradigm, but technological relevance is not expected to be achieved soon. Moreover, given the technological challenge, QC channels are expected to be very expensive to communicate on, further hindering the usage of LO and Quantum Communications. Finally, LO and Quantum Communications operations are a less natural set of operations to study the properties of entangled states, because nonlocal properties can be caused by the operations (sending quantum systems allows to create entanglement).

Conversely, by definition LOCC operations are only local in nature, and thus cannot increase nonlocal properties on average (cf. Sections 2.3.3 and D.2). As such, LOCC is the natural framework with which to study entanglement as a resource, considering that any nonlocal properties showcased will be due to the entangled states themselves, and not due to any manipulation.

Given the experimental conditions given above, the free operations of LOCC can be defined as being made of successive rounds of local measurements followed by a global broadcast of the measurement result, which can be used to determine the next local measurements to perform. Concatenating the successive operators for the measurements on each party into a single measurement operator M, we get that every multipartite LOCC map Λ will have the form

$$\Lambda(\cdot) = \sum_k \left(\otimes_{i=1}^N M_{k,i}^{A_i}\right)(\cdot)\left(\otimes_{i=1}^N M_{k,i}^{A_i}\right)^\dagger, \tag{2.38}$$

where $M_{k,i}^{A_i}$ acts on the Hilbert space of party A_i, i.e. A_1 denotes Alice, A_2 denotes Bob, etc. The bipartite case arises for $N = 2$. Eq. (2.38) hides some of the complexity induced by the number of communication rounds by concatenating the successive measurement operators into a single set of operators, however it is interesting to note that the properties and the form of the set LOCC are still an active area of research, notably concerning the question of round complexity.

Without surprise, from these free operations, the free states are simply the set of separable states, denoted $\mathrm{SEP}(\mathcal{H})$. A simple way to see this is that one can generate any separable state from any joint state by discarding the joint state entirely, and generating each piece of the reduced states directly. Such an operation can be expressed in the form of Eq. (2.38), and so $\mathrm{SEP}(\mathcal{H}) \subseteq \mathcal{F}(\mathcal{H})$. Moreover, it can be shown by an invariance argument that $\mathrm{SEP}(\mathcal{H}) = \mathcal{F}(\mathcal{H})$. Essentially, the idea is that it can be shown that $\mathrm{SEP}(\mathcal{H}) \xrightarrow{\mathrm{LOCC}} \mathrm{SEP}(\mathcal{H})$, but since by definition every free state is reachable from any other state (including free states) by LOCC, then $\mathrm{SEP}(\mathcal{H})$ being the image of itself under the mapping LOCC means that it must contain all free states.

2.3.3 Majorization criterion for deterministic transformations

This section is based on Ref. [5]. We now know that the theory of entanglement with the mapping LOCC yields a QRT, and so entanglement can be seen as a resource. Thankfully additional results exist giving the conditions under which a state is reachable from another (without having to try every possible LOCC map until one works).

Theorem 2.4 (Nielsen's theorem)**.** *Let $|\phi\rangle$ and $|\psi\rangle$ be two pure states of a bipartite system AB. Then,*

$$|\psi\rangle \xrightarrow{LOCC} |\phi\rangle \iff \lambda_\psi \prec \lambda_\phi, \tag{2.39}$$

where λ_ψ and λ_ϕ are the Schmidt vectors of $|\psi\rangle$ and $|\phi\rangle$, respectively.

[10]The LOQC abreviation usually stands for Linear Optical Quantum Computing, so we will avoid using it to denote something else here.

This theorem is the second time we encounter majorization in the theory of entanglement. This theorem is very powerful, and confirms some of our suspicions. From our QRT discussion we expect $\mathrm{SEP}(\mathcal{H})$ to only be able to reach itself. Theorem 2.4 is clearly compatible with this result, because separable states have Schmidt vector $\lambda = \overline{\delta}_d = (1, 0, \ldots, 0)$, which majorizes every distribution. As such, LOCC only allows you to convert a state $|\psi\rangle$ with Schmidt vector $\lambda_\psi = \overline{\delta}_d$ into another target state $|\phi\rangle$ with Schmidt vector $\lambda_\phi = \overline{\delta}_d$, i.e. separable states can only reach other separable states. As such, the states with the least amount of entanglement resource are the free states, as expected. In QRT terminology, we say that such states are *minimally entangled* (which is consistent with the fact that they are not entangled).

Remark 2.5. Two states with identical Schmidt coefficients can always be converted back and forth without loss of entanglement. This is consistent with the idea that they are equivalent up to a local change of basis which is a local operation.

The same logic can be applied to find the state with the maximal amount of entanglement resource, i.e. the state that can be transformed into any other entangled state using LOCC: the Schmidt vector $\lambda_\psi = \overline{1}_d = \left(\frac{1}{d}, \ldots, \frac{1}{d}\right)$ is majorized by every other Schmidt vector, and so the condition of Theorem 2.4 is fulfilled no matter the target Schmidt vector λ_ϕ. In QRT terminology, we say that states with $\lambda = \overline{1}_d$ are *maximally entangled*.

Just like other resource theories, free operations can only make a state less resourceful. In our case, LOCC can transform a strongly entangled state (in the sense that its Schmidt vector is close to the flat distribution and can thus reach many other states) into a lightly entangled state, but not vice versa. We will say that after transformation, the state has been *degraded*. Keep in mind however that a bistochastic degradation (cf. Section 1.1.2) of the Schmidt vector would make the state more resourceful, and not less. To avoid any confusion, we will always attempt to precise which type of degradation we speak of if not clear from context. Finally, we can particularise the idea of resource monotone to an entanglement monotone in the context of LOCC.

Definition 2.9 (Entanglement monotone)**.** An entanglement monotone is a function $F : \mathcal{H} \to \mathbb{R}$ such that

$$|\psi\rangle \xrightarrow{\text{LOCC}} |\phi\rangle \implies F(|\psi\rangle) \geq F(|\phi\rangle). \tag{2.40}$$

Remark 2.6. Considering that Theorem 2.4 gives us a LOCC reachability condition in terms of a majorization relation on Schmidt vectors, it is easy to see that any Schur-concave function applied on the Schmidt vector of a state is a valid entanglement monotone. In particular, $H(\lambda_\psi)$ is a valid entanglement monotone, which is quite satisfying considering the information-theoretic interpretation of the Schmidt vector (cf. Section 2.2.4).

Figure 2.1 shows the geometric illustration of entanglement cones (cf. Section 1.3.2). With Theorem 2.4 in mind, it is clearer why we decided to call $\mathcal{T}_+(\lambda_\psi)$ the future cone of λ_ψ, as it is exactly the set containing all of the LOCC reachable Schmidt vectors from λ_ψ. Likewise, $\mathcal{T}_-(\lambda_\psi)$ is the set of states from which λ_ψ could have been constructed through LOCC, hence the past cone denomination. Finally, the incomparable region $\mathcal{T}_\emptyset(\lambda_\psi)$ is the set of states that could not have reached λ_ψ, and can not be reached from λ_ψ. As such, holding additional states in $\mathcal{T}_-(\lambda_\psi) \cup \mathcal{T}_\emptyset(\lambda_\psi)$ is interesting, because it allows you to reach states not directly reachable from λ_ψ. This will be touched on further in Section 5.2.

Moreover, the convention of representing the lattice with increasing entropy towards the top makes more sense now that the notion of entanglement monotone has been defined, as more resourceful states are therefore drawn at the top. Separable states are at the very bottom of the lattice and can always be reached from any state of the theory.

As a final remark, it is important to note that from now on, we will mostly work with probability distributions and their entropic properties on the lattice. We will not explicitly remind it each time, however to bring ourselves back to the quantum case, it suffices to plug a

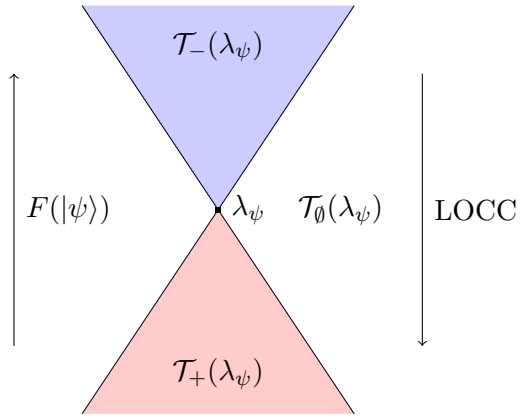


Figure 2.1: Depiction of the entanglement cones of a pure state $|\psi\rangle$ with Schmidt vector $\lambda_\psi \in \mathcal{P}^d$. F is any entanglement monotone. In the LOCC context, the denomination of future cone $\mathcal{T}_+$ and past cone $\mathcal{T}_-$ is clearer, as they are the sets that a state λ can become through LOCC or from which it could have been constructed by LOCC, respectively.

Schmidt vector λ into distributions p and q. This works because a Schmidt vector has a sum of one and can simply be seen as a probability distribution (cf. Remark 2.6), and so entropic properties of the Schmidt vector are of interest to quantify the amount of resource of an entangled state.

We will not use probabilistic transformations very much in this master thesis, however they are also another appearance of (weak) majorization in the theory of entanglement. A quick overview is available in Appendix D.

Part II

Results

Chapter 3

Supermodularity and subadditivity on the majorization lattice

This chapter proves some of the mathematical properties of the Shannon entropy we will use extensively in the following chapters, and is thus of interest for the study of quantum applications of the majorization lattice. In particular, we show alternative proofs of supermodularity and subadditivity, which often come up when working with entropic distances and thus appear a lot in Chapter 4.

3.1 Concatenations, sum-convex and sum-concave functions

We will work with concatenated probability vectors. We will need the following definitions for the main theorems of the following sections.

Definition 3.1 (Probability vector concatenation)**.** Let $p, q \in \mathcal{P}^d$, then the ordered concatenation of p and q, $\text{concat}(p, q)$, is defined as

$$\text{concat}(p, q) = (p_1, p_2, ..., p_d, q_1, q_2, ..., q_d)^{\downarrow} \in \mathbb{R}^{2d}, \tag{3.1}$$

and so $S^{\downarrow}_{2d}(\text{concat}(p, q)) = 2$.

Definition 3.2 (Sum-convex function)**.** Let $F : \mathcal{P}^n \to \mathbb{R}$. F is sum-convex if there exists a convex function $\phi : [0, 1] \to \mathbb{R}$ such that

$$F(p) = \sum_i \phi(p_i) \tag{3.2}$$

Definition 3.3 (Sum-concave function)**.** Let $F : \mathcal{P}^n \to \mathbb{R}$. F is sum-concave if there exists a concave function $\varphi : [0, 1] \to \mathbb{R}$ such that

$$F(p) = \sum_i \varphi(p_i) \tag{3.3}$$

Alternatively, F is sum-concave if $-F$ is sum-convex. Moreover, note that Lemma 1.1 implies that all sum-convex functions (resp. sum-concave) are Schur-convex (resp. Schur-concave).

3.2 Supermodularity of all sum-concave functions

3.2.1 Main statement

Theorem 3.1 (Supermodularity of all sum-concave functions)**.** *All sum-concave functions F are supermodular on the majorization lattice*[1] *provided*[2] *$p, q \in \mathcal{P}^d$ are such that $\beta(p,q) = p \vee q$ (as defined in (1.16)). For those vectors, we have*

$$F(p \wedge q) + F(p \vee q) \geq F(p) + F(q). \tag{3.4}$$

Note that if $p \sim q$, there is trivially equality. Moreover, in dimension 3, $\beta(p,q) = p \vee q$ is always true.

Proof. We will prove the slightly stronger result on $\beta(p,q)$ as defined in (1.16):

$$F(p \wedge q) + F(\beta(p,q)) \geq F(p) + F(q). \tag{3.5}$$

Let us define[3]

$$r = p \wedge q, \;\; s = \beta(p,q), \tag{3.6}$$
$$A = \mathrm{concat}(p,q) \in \mathbb{R}^{2d}, \tag{3.7}$$
$$B = \mathrm{concat}(r,s) \in \mathbb{R}^{2d}, \tag{3.8}$$

where $\mathrm{concat}(\cdot,\cdot)$ is the ordered concatenation of two vectors. We have $\sum_{i=1}^{2d} A_i = \sum_{i=1}^{2d} B_i = 2$. We will now show that $A \succ B$. To do so, we need to show that the majorization inequalities (cf. Definition 1.2) are satisfied for each $k \leq 2d$. Clearly, the k^{th} cumulative sum of A is made up of the first entries of p and q, and the same goes for cumulative sums of B being made of the first entries of r and s. To avoid ill-defined terms, we additionally introduce the convention that $S_k^{\downarrow}(v) = 1 \; \forall k > d$ for any d-dimensional probability vector v. For each $k \leq 2d$, we have

$$S_k^{\downarrow}(A) = \max_{0 \leq l \leq k} \left(S_l^{\downarrow}(p) + S_{k-l}^{\downarrow}(q) \right), \tag{3.9}$$
$$S_k^{\downarrow}(B) = \max_{0 \leq m \leq k} \left(S_{k-m}^{\downarrow}(r) + S_m^{\downarrow}(s) \right). \tag{3.10}$$

Recall Eqs. (1.13) and (1.15), from which we immediately deduce[4]

$$S_i^{\downarrow}(r) = \min \left\{ S_i^{\downarrow}(p), S_i^{\downarrow}(q) \right\}, \tag{3.11}$$
$$S_i^{\downarrow}(s) = \max \left\{ S_i^{\downarrow}(p), S_i^{\downarrow}(q) \right\}. \tag{3.12}$$

Let us call the value of m that realizes the maximum of Eq. (3.10) m'. There are two possible cases, either $S_{m'}^{\downarrow}(p) \geq S_{m'}^{\downarrow}(q)$ or $S_{m'}^{\downarrow}(p) \leq S_{m'}^{\downarrow}(q)$. Let us consider the $S_{m'}^{\downarrow}(p) \geq S_{m'}^{\downarrow}(q)$ case, which gives us

$$S_k^{\downarrow}(B) = \min \left\{ S_{k-m'}^{\downarrow}(p), S_{k-m'}^{\downarrow}(q) \right\} + S_{m'}^{\downarrow}(p) \tag{3.13}$$

[1] It is important to specify which lattice, because a different lattice means a different ordering and different meet and join operations, leading to different properties. For example, on the boolean lattice, the Shannon entropy is *submodular* instead [10].

[2] This proof has since been extended to remove this condition, see Appendix I.

[3] Credit must be given to OpenAI's o4-mini LLM for the idea of constructing such concatenated vectors. After digging, it turns out that similar techniques are used in Ref. [7, pp. 133–136], where non-trivial vector constructions are also used to prove convexity results.

[4] This is the step where the $\beta(p,q) = p \vee q$ limitation comes into play: the expression in terms of maxima of cumulative sums of $\beta(p,q)$ was defined with the unordered vector, but majorization relations require the ordered vector. As such, we cannot use the maximum relation directly if $\beta(p,q) \neq \beta^{\downarrow}(p,q)$, on which this proof hinges.

$$\leq S^{\downarrow}_{m'}(p) + S^{\downarrow}_{k-m'}(q) \tag{3.14}$$

$$\leq \max_{0\leq l\leq k} \left(S^{\downarrow}_{l}(p) + S^{\downarrow}_{k-l}(q) \right) = S^{\downarrow}_{k}(A), \tag{3.15}$$

which is precisely one inequality of the majorization relation for a specific k. The $S^{\downarrow}_{m'}(p) \leq S^{\downarrow}_{m'}(q)$ case being symmetric, and this being true for all $k \leq 2d$, we have shown that $A \succ B$. Using Lemma 1.1 on φ ($-\varphi$ being convex on the same interval), we get

$$\sum_{i=1}^{2d} \varphi(A_i) \leq \sum_{i=1}^{2d} \varphi(B_i). \tag{3.16}$$

By the sum nature of F, the LHS of (3.16) is precisely $F(p) + F(q)$, and the RHS is precisely $F(p \wedge q) + F(\beta(p, q))$, so we have proven (3.5) which implies (3.4). □

3.2.2 Corollaries and main conjecture

Of course, the corollary we care the most about is the supermodularity of the Shannon entropy. We will not write it here however, because Theorem 1.3 is stronger than a corollary we can build from Theorem 3.1 (because of the limiting $\beta(p, q) = p \vee q$ condition).

However, the above proof can be reused for a sum-convex function f (because $-f$ is sum-concave) by simply reversing the final Karamata inequality, giving the following corollary.

Corollary 3.1.1 (Submodularity of all sum-convex functions)**.** *All sum-convex functions F are submodular on the majorization lattice for all $p, q \in \mathcal{P}^d$ such that $\beta(p, q) = p \vee q$ (as defined in (1.16)). For those vectors, we have*

$$F(p \wedge q) + F(p \vee q) \leq F(p) + F(q). \tag{3.17}$$

Note that if $p \sim q$, there is trivially equality. Moreover, in dimension 3, $\beta(p, q) = p \vee q$ is always true.

Another interesting corollary is that the exponential of Rényi entropies 2^{H_α} can be expressed as sum-concave or sum-convex functions too (depending on the parameter α). This is simply because $2^{H_\alpha} = 2^{\frac{1}{1-\alpha} \log \sum_i p_i^\alpha} = 2^{\frac{1}{1-\alpha}} \sum_i p_i^\alpha$, which is clearly sum-concave if $\alpha < 1$ and sum-convex if $\alpha > 1$.

Corollary 3.1.2 (Supermodularity of the exponential of $\alpha < 1$ Rényi entropies)**.** *The exponential of Rényi entropy functions 2^{H_α} of order $\alpha < 1$ are supermodular on the majorization lattice for all $p, q \in \mathcal{P}^d$ such that $\beta(p, q) = p \vee q$, for which we have*

$$2^{H_\alpha(p\wedge q)} + 2^{H_\alpha(p\vee q)} \geq 2^{H_\alpha(p)} + 2^{H_\alpha(q)}. \tag{3.18}$$

Note that if $p \sim q$, there is trivially equality.

Corollary 3.1.3 (Submodularity of the exponential of $\alpha > 1$ Rényi entropies)**.** *The exponential of Rényi entropy functions 2^{H_α} of order $\alpha > 1$ are submodular on the majorization lattice for all $p, q \in \mathcal{P}^d$ such that $\beta(p, q) = p \vee q$, for which we have*

$$2^{H_\alpha(p\wedge q)} + 2^{H_\alpha(p\vee q)} \leq 2^{H_\alpha(p)} + 2^{H_\alpha(q)}. \tag{3.19}$$

Note that if $p \sim q$, there is trivially equality.

Theorem 3.1 is a generalization of supermodularity to a broader class of functions. However, the condition $\beta(p, q) = p \vee q$ is limiting. Numerical testing seems to indicate that the relation $A \succ B$ holds even in cases where $\beta(p, q) \neq \beta^{\downarrow}(p, q)$ (and so $\beta^{\downarrow}$ must be smoothed to get the join), which yields the first conjecture of this master thesis[5].

[5]This conjecture has since been proven, see Appendix I.

Conjecture 3.1. *All sum-concave functions F are supermodular on the majorization lattice, and so for all $p, q \in \mathcal{P}^d$,*

$$F(p \wedge q) + F(p \vee q) \geq F(p) + F(q). \tag{3.20}$$

Note that if $p \sim q$, there is trivially equality.

3.3 Subadditivity of all sum-concave functions

3.3.1 Main statement

Thankfully, in the case of subadditivity we are not bound by a $\beta(p, q) = p \vee q$ condition, because we will not make use of the join at all during the proof. We therefore have a form of subadditivity that is applicable to more functions than Cicalese and Vaccaro's original theorem. Moreover, the d-dimensional certain distribution $\overline{\delta}_d = (1, 0, \ldots, 0)$ also appears in the proof, and so our theorem is actually a slightly stronger form of subadditivity.

Theorem 3.2 (Subadditivity of all sum-concave functions)**.** *All sum-concave functions F are subadditive on the majorization lattice for all $p, q \in \mathcal{P}^d$. Formally,*

$$F(p \wedge q) \leq F(p) + F(q) - F(\overline{\delta}_d). \tag{3.21}$$

Proof. Let $p, q \in \mathcal{P}^d$. Define

$$r = p \wedge q, \tag{3.22}$$
$$A = \mathrm{concat}(p, q), \tag{3.23}$$
$$B = \mathrm{concat}(\overline{\delta}_d, r), \tag{3.24}$$

where $\mathrm{concat}(\cdot, \cdot)$ is the ordered concatenation of two vectors. We have $\sum_{i=1}^{2d} A_i = \sum_{i=1}^{2d} B_i = 2$. We will now show that the majorization precursor $A \prec B$ is true. Clearly, the k^{th} cumulative sum of A is made up of the first entries of p and q, and the same goes for cumulative sums of B being made of the first entries of r and $\overline{\delta}_d$. To avoid ill-defined terms, we additionally introduce the convention that $S_k^{\downarrow}(v) = 1 \; \forall k > d$ and $S_0^{\downarrow}(v) = 0$ for any d-dimensional probability vector v. For all $k \leq 2d$, we have

$$S_k^{\downarrow}(A) = \max_{0 \leq l \leq k} \left(S_l^{\downarrow}(p) + S_{k-l}^{\downarrow}(q) \right), \tag{3.25}$$
$$S_k^{\downarrow}(B) = 1 + S_{k-1}^{\downarrow}(r). \tag{3.26}$$

Recall Eq. (1.13), from which we immediately deduce

$$S_i^{\downarrow}(r) = \min \left\{ S_i^{\downarrow}(p), S_i^{\downarrow}(q) \right\}. \tag{3.27}$$

We need to show that the majorization inequalities (cf. Definition 1.2) are satisfied for each $k \leq 2d$. First, we directly see that for $k = 1$, we have $S_1^{\downarrow}(B) = 1 \geq S_1^{\downarrow}(A) = \max\{p_1, q_1\}$.

Now, for $k \geq 2$, there are several cases depending on the value of l, which ranges from 0 to k. Having no knowledge over p and q, we must consider several cases separately.

1. $\underline{l = 0}$: we have
$$S_k^{\downarrow}(B) \geq 1 \geq S_k^{\downarrow}(q). \tag{3.28}$$

2. $\underline{l = k}$: symmetric to the $l = 0$ case, because
$$S_k^{\downarrow}(B) \geq 1 \geq S_k^{\downarrow}(p). \tag{3.29}$$

For the remaining cases, we first do the following manipulation

$$S_k^{\downarrow}(B) = 1 + \min\left\{S_{k-1}^{\downarrow}(p), S_{k-1}^{\downarrow}(q)\right\} \tag{3.30}$$

$$\geq \max\left\{S_{k-1}^{\downarrow}(p), S_{k-1}^{\downarrow}(q)\right\} + \min\left\{S_{k-1}^{\downarrow}(p), S_{k-1}^{\downarrow}(q)\right\} \tag{3.31}$$

$$= S_{k-1}^{\downarrow}(p) + S_{k-1}^{\downarrow}(q), \tag{3.32}$$

where we have used the fact that the cumulative sum of any d-dimensional vector is always less or equal to 1 to write Eq. (3.31). Therefore, for the majorization relation $S_k^{\downarrow}(B) \geq S_k^{\downarrow}(A)$ to be verified, it is enough to show that for any choice of l that we haven't treated yet,

$$S_{k-1}^{\downarrow}(p) + S_{k-1}^{\downarrow}(q) \geq \max_{0<l<k}\left(S_l^{\downarrow}(p) + S_{k-l}^{\downarrow}(q)\right) \tag{3.33}$$

$$\iff 0 \geq \max_{0<l<k}\left(S_l^{\downarrow}(p) - S_{k-1}^{\downarrow}(p) + S_{k-l}^{\downarrow}(q) - S_{k-1}^{\downarrow}(q)\right), \tag{3.34}$$

and so the majorization inequality is satisfied if we show that the RHS of Eq. (3.34) is less than or equal to 0.

3. $\underline{l = 1}$: Eq. (3.34) becomes
$$S_1^{\downarrow}(p) - S_{k-1}^{\downarrow}(p) \leq 0, \tag{3.35}$$
which is verified $\forall k \geq 2$.

4. $\underline{l = k-1}$: symmetric to the $l = 1$ case, because Eq. (3.34) becomes
$$S_1^{\downarrow}(q) - S_{k-1}^{\downarrow}(q) \leq 0, \tag{3.36}$$
which is verified $\forall k \geq 2$.

5. $\underline{2 \leq l \leq k-2}$: this case can only arise for $k \geq 4$, otherwise the inequality allows no l value[6]. We have both
$$S_l^{\downarrow}(p) \leq S_{k-1}^{\downarrow}(p) \quad \text{and} \quad S_{k-l}^{\downarrow}(q) \leq S_{k-1}^{\downarrow}(q), \tag{3.37}$$
which are both true $\forall k \geq 4$, and so Eq. (3.34) is verified.

Therefore, no matter the value of l, the majorization inequalities are satisfied $\forall k \leq 2d$. We have thus shown that $B \succ A$. Using Lemma 1.1 on φ ($-\varphi$ being convex on the same interval), we get

$$\sum_{i=1}^{2d} \varphi(A_i) \geq \sum_{i=1}^{2d} \varphi(B_i). \tag{3.38}$$

By the sum nature of F, the LHS of (3.16) is precisely $F(p) + F(q)$, and the RHS is precisely $F(p \wedge q) + F(\overline{\delta}_d)$, which we can rearrange into

$$F(p \wedge q) \leq F(p) + F(q) - F(\overline{\delta}_d). \tag{3.39}$$

□

It is interesting to note that the appearance of the additional term $F(\overline{\delta}_d)$ makes this theorem stronger than a subadditivity property. In the case of the Shannon entropy, the additional term simply vanishes, because the entropy of the certain distribution $\overline{\delta}_d$ is 0, however for other sum-concave functions that are not 0 for the certain distribution, this is a better upper bound than simple subadditivity.

[6]This is not an issue because if $k = 2$, $l \in \{0, 1, 2\}$, and if $k = 3$, $l \in \{0, 1, 2, 3\}$. All of these possibilities fall under cases 1, 2, 3 or 4, thus all possible cases are taken into account.

3.3.2 Corollaries

Theorem 1.4 concerning the subadditivity of H becomes a corollary of Theorem 3.2, because H is sum-concave from its definition.

Just like with supermodularity, the above proof can be reused for a sum-convex function f (because $-f$ is sum-concave) by simply reversing the final Karamata inequality, giving the following corollary.

Corollary 3.2.1 (Superadditivity of all sum-convex functions)**.** *All sum-convex functions F are superadditive on the majorization lattice for all $p, q \in \mathcal{P}^d$. Formally,*

$$F(p \wedge q) \geq F(p) + F(q) - F(\overline{\delta}_d). \tag{3.40}$$

Theorem 3.2 also implies corollaries analoguous to 3.1.2 and 3.1.3 for subadditivity and superadditivity of the exponential of Rényi entropies of order $\alpha < 1$ and $\alpha > 1$, respectively.

3.4 Discussion

We have proven that properties akin to supermodularity and subadditivity hold for a broader class of functions than only the Shannon entropy. Some sum-concave functions is used in recent litterature, particularly the subset of *separable, piecewise-linear, concave* (SPLC) utilities (which is a subset of the class of sum-concave functions) in the theory of market equilibria in economics and algorithmic game theory [25, 26]. Conjecture 3.1 and Theorem 3.2 might then be of interest if researchers decide to study some of their probability distributions on the majorization lattice. Some game equilibria have already been studied on lattice structures [27], so such a prospect might be interesting.

The exponential of Rényi entropies has also long been studied, and inequalities and concavity results have already been shown for differential entropy [28, 29]. Corollaries 3.1.2 and 3.1.3 and their subadditivity/superadditivity counterparts might be of interest in the discrete case. Moreover, in the quantum context, Rényi entropies are of interest (cf. Section 4.1.3), so these new properties on the lattice for derived quantities of Rényi entropies might be of interest for future research.

We also believe that the concatenation technique used in this chapter might be of use to find new majorization precursors for existing entropic inequalities where other techniques have failed. For example, there is currently some interest in finding a majorization precursor for the entropic version of Bell inequalities from Ref. [30], because such a majorization relation could enable deeper insight into Bell nonlocality.

Chapter 4

Quantification of incomparability for pairs of states

This chapter first shows an improvement over traditional entropic separability criteria that can be achieved using the lattice. As this improvement depends on the incomparability between the two reduced states'eigenvalues, this chapter then tries to quantify the amount of incomparability between probability vectors by defining new objects . It is interesting to note that some unpublished work attempting to quantify the incomparability of quantum states (without the lattice) does exist in Ref. [31], however the proposed measures do not seem very natural to us.

4.1 Motivation

4.1.1 Improved separability criterion based on meet

The second result of this master thesis is to improve some entropic criteria tied to majorization relations. As stated in Theorem 2.3, if a bipartite state ρ^{AB} is separable, then

$$\lambda^{AB} \prec \lambda^A \quad \text{and} \quad \lambda^{AB} \prec \lambda^B, \tag{4.1}$$

where λ^X is the ordered vector of eigenvalues of the density matrix of system X. By Schur-concavity of H, this directly implies

$$H(\lambda^{AB}) \geq H(\lambda^A) \quad \text{and} \quad H(\lambda^{AB}) \geq H(\lambda^B). \tag{4.2}$$

These entropic inequalities[1] are of course not as strong as the majorization relations we deduced them from, as illustrated in the example from Section 2.2.5. Using the lattice, however, the entropic condition can be strengthened. The relations $\lambda^{AB} \prec \lambda^A$ and $\lambda^{AB} \prec \lambda^B$ mean that λ^{AB} is majorized by both λ^A and λ^B. By definition, the meet of λ^A and λ^B is the vector $\lambda^A \wedge \lambda^B$ that majorizes all the vectors that are majorized by both λ^A and λ^B. This fact can also be restated as

$$\mathcal{T}_-(\lambda^A) \cap \mathcal{T}_-(\lambda^B) = \mathcal{T}_-(\lambda^A \wedge \lambda^B), \tag{4.3}$$

using Definition 1.8 of majorization cones. Applying this to the separability criterion, we get that if ρ^{AB} is a separable state, then

$$\lambda^{AB} \prec \lambda^A \text{ and } \lambda^{AB} \prec \lambda^B \iff \lambda^{AB} \prec \lambda^A \wedge \lambda^B \implies H(\lambda^{AB}) \geq H(\lambda^A \wedge \lambda^B). \tag{4.4}$$

This new formulation in terms of entropies is a strengthening of the initial entropic criterion (4.2), providing a better lower bound on the entropy of a separable state. However, it is still

[1] Historically, these entropic inequalities (expressed equivalently in von Neumann entropies) were known before the strengthening to a majorization relation [21, 23].

not as strong as the majorization criterion itself. Moreover, this new entropic criterion based on the meet is not very interesting for pure bipartite states because of Corollary 2.2.2, which essentially states that $\lambda^A = \lambda^B$ if ρ^{AB} is pure. In that case $\lambda^A \wedge \lambda^B = \lambda^A = \lambda^B$, and the new entropic criterion reduces to the original one. While pure bipartite states are the more interesting states for quantum computation, as decoherence (i.e. losing purity by interacting with the environment) is never a desired feature for qubits as mixed states introduce uncertainty into computations, in practice noise always makes states at least slightly mixed. Figure 4.1 illustrates the new criterion.

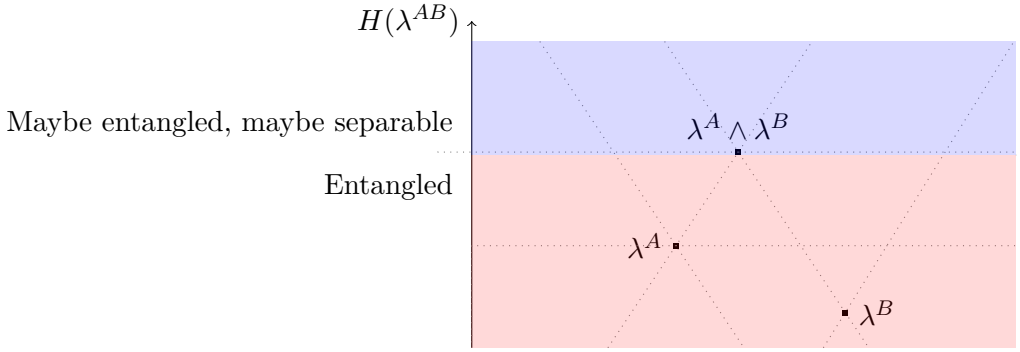


Figure 4.1: Depiction of the improved entropic criterion for entanglement. The previous criterion could not decide whether states with λ^{AB} with an entropy between the two horizontal doted lines going through λ^A and $\lambda^A \wedge \lambda^B$ were entangled. The new criterion shows that they are.

Another caveat to this criterion is that if $\lambda^A \sim \lambda^B$, then $\lambda^A \wedge \lambda^B = \lambda^A$ or λ^B (whichever is majorized by the other), and the criterion reduces to the original one. As such, it seems that something interesting is happening when two vectors are incomparable, as incomparability enables us to improve entropic criteria. This very notion, of quantifying the amount of incomparability between two probability vectors and considering it a resource, is the idea that led to Chapter 4.

However, it could still be useful in experiments trying to assert whether a state is entangled or not. In such cases, the majorization relations can be hard to use because the vector of eigenvalues of the density operator can be hard to determine experimentally. However, given many copies, it is easier to use the measurement statistics to compute the Shannon entropy of the distributions. As such, if one could experimentally construct a state with the same eigenvalues as $\lambda^A \wedge \lambda^B$, one could measure its statistics and use the better lower bound to determine whether the joint state is separable. Unfortunately, no obvious way to construct this state, which we will note $\rho_{A \wedge B}$, was found. For example, constructing this state with an interferometer in quantum optics would require the construction of the meet (cf. Section 1.3.3) to have a known formulation in terms of unitary operators. Outputting both the join and the meet would be even more interesting (and one could restrict themselves to measuring the statistics of the meet state for the separability criterion).

We suspect that such an interferometer might be possible, but no clear lead on designing the interferometer has been found because it is not clear how the meet and join constructions might be possible with unitaries. We have

$$a_i = \min\Big\{\sum_{j=1}^{i} p_j, \sum_{j=1}^{i} q_j\Big\} - \sum_{j=1}^{i-1} a_j \tag{4.5}$$

$$= \min\Big\{\sum_{j=1}^{i} p_j, \sum_{j=1}^{i} q_j\Big\} - \min\Big\{\sum_{j=1}^{i-1} p_j, \sum_{j=1}^{i-1} q_j\Big\} \tag{4.6}$$

$$b_i = \max\Big\{\sum_{j=1}^{i} p_j, \sum_{j=1}^{i} q_j\Big\} - \sum_{j=1}^{i-1} b_j \tag{4.7}$$

$$= \max\Big\{\sum_{j=1}^{i} p_j, \sum_{j=1}^{i} q_j\Big\} - \max\Big\{\sum_{j=1}^{i-1} p_j, \sum_{j=1}^{i-1} q_j\Big\}, \tag{4.8}$$

where we use $\beta(p, q)$ instead of the true join because the smoothing algorithm is an additional layer of complexity. The minima and maxima of those expressions seem difficult to realize with unitaries.

4.1.2 Other criteria

The previous discussion is not unique to quantum information and separability. Any other branch of physics working with coupled majorization relations but struggling to use them in practice might benefit from constructing the equivalent of a state $\rho_{A\wedge B}$ and computing the entropy of the resulting distribution. For example, another field that deals with majorization relations is the field of quantum thermodynamics, where allowed state transformations under thermal operations are bound to thermomajorization cones (called thermal cones), analogously to entanglement cones in LOCC [15, 17]. On the flip-side, this kind of entropic criterion based on the meet could also be used as an experimental sign of an underlying majorization relation, in cases where one suspects that there is something deeper going on than only the entropic inequalities.

4.1.3 Simulations for Rényi entropies

The improved criterion is not unique to the Shannon entropy. All Rényi entropies H_α are Schur-concave, and so

$$\lambda^{AB} \prec \lambda^A \text{ and } \lambda^{AB} \prec \lambda^B \iff \lambda^{AB} \prec \lambda^A \wedge \lambda^B \tag{4.9}$$

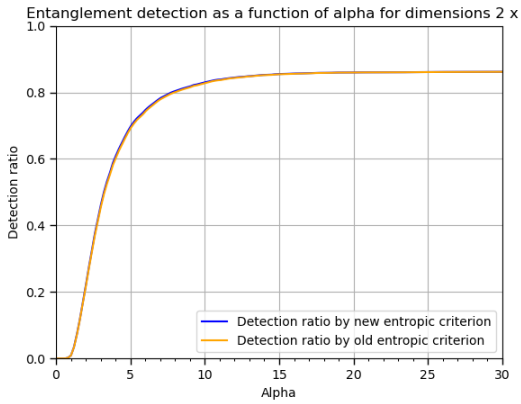
$$\implies H_\alpha(\lambda^{AB}) \geq H_\alpha(\lambda^A \wedge \lambda^B) \quad \forall \alpha \in \mathbb{R}^+. \tag{4.10}$$

A numerical analysis was carried out to figure out whether there exists an optimal α for which the improved criterion detects entanglement the most amount of times, relative to the original criterion. This was simulated in the context of entanglement by generating random density matrices using QuTip with a specified tensor structure until the majorization criterion $\lambda^A \wedge \lambda^B \succ \lambda^{AB}$ is failed (and so we are guaranteed to not have a separable state). This way, 10,000 strongly entangled states are generated. Then, the Rényi-entropic criterion was used on each of these strongly entangled states by iterating over the value of α, with a step of 0.2, and each entanglement detection was recorded. Finally, this process is repeated for several dimensions of interest. An implementation of this test is available on the GitHub for the project[2].

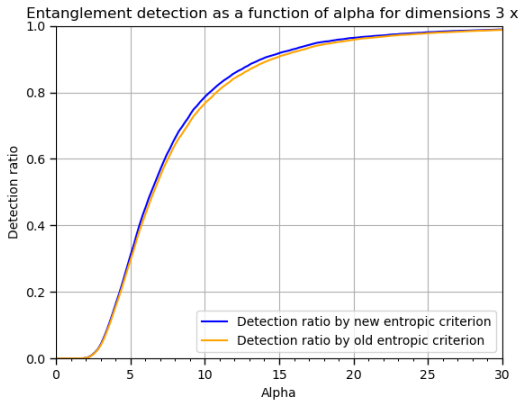
Computationally, the eigenvalue calculations of the operators and reduced operators become very heavy as the dimensions increase, so the largest dimensions attempted were 4×4 before the computation time became prohibitive. The 2×2 case is not shown because incomparability is not possible in two dimensions and so the meet criterion reduces to the old one. The final results are shown in Figures 4.2.

Interestingly, the Rényi simulation shows that there does seem to be an optimum α where the meet criterion seems to perform better than the old criterion (the difference then vanishes again at large values of α), which seems to be dimension-dependent. Perhaps this hints at some deeper connection between Rényi entropy and majorization, but this research avenue was not explored further. It is also interesting to note that the improved criterion does not seem to show significant improvements according to the simulation. This might be because of a non-uniform

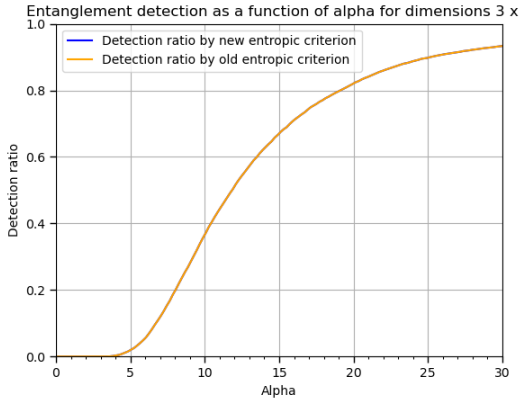
[2] https://github.com/traaldbjerg/MajoLat



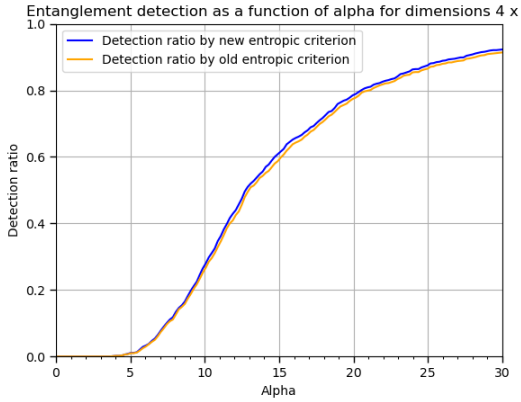
(a) Dimensions 2×3.



(b) Dimensions 3×3.



(c) Dimensions 3×6.



(d) Dimensions 4×4.

Figure 4.2: Comparison of the entanglement detection ratios for the old and new entropic criteria for various subsystem dimensions, depending on the value of the Rényi order parameter α. 10000 strongly entangled states were generated, and the Rényi criterion was applied to each of them for each value of α between 0 and 30, with a step of 0.2. The only exception is the 4×4 case, where only 1000 states were sampled due to the very long computation times (which explains the larger statistical fluctuations for curve 4.2d).

density of quantum states in the Δ_{d-1} simplex, which might lead to a significant number of random density matrices being drawn close to the bottom of the lattice where both criteria can tell entanglement. Perhaps, instead, this might be due to reduced states of a density matrix not tending to be too imcomparable from each other, and thus being close to their meet. This final idea leads nicely into the second part of this chapter, where we define and study new objects quantifying the incomparability between probability vectors.

4.2 Entropic distance approach

4.2.1 Resource-theoretic intuition

We took inspiration from resource theories to define incomparability monotones. What is often done in resource theories is to define a monotone as being the minimum distance from the studied state to the set of free states. For instance, the entanglement monotone $H(\lambda_\psi)$ is such a monotone. Consider the entropic distance (cf. Section 1.4.3) $d(\lambda_\psi, \lambda_{\text{free}})$. We know that all minimally entangled states have $\lambda_{\text{free}} = \overline{\delta}_d = (1, 0, \ldots, 0)$ (they are separable), and so the

entropic distance reduces to

$$d(\lambda_\psi, \lambda_{\text{free}}) = H(\lambda_\psi) + H(\lambda_{\text{free}}) - 2H(\lambda_\psi \vee \lambda_{\text{free}}) \tag{4.11}$$
$$= H(\lambda_\psi), \tag{4.12}$$

where $H(\lambda_{\text{free}}) = 0$ (the certain distribution has no entropy) and $\lambda_\psi \vee \lambda_{\text{free}} = \lambda_{\text{free}}$ because the certain distribution is the bottom of the lattice. In this particular case, there is no need to take the minimum over all free states because they all have the same Schmidt vector, but in other resource theories or using other notions of distance between quantum states, they might not all be at the same distance from the state λ_ψ.

4.2.2 Expected properties

There are some key differences to usual resource monotones when quantifying incomparability. The first major difference is that a vector can only be incomparable to another vector (a vector is not incomparable in and of itself). As such, treating incomparability as a resource is a bit strange because it would characterize some form of collective property of a pair of states, rather than an intrinsic property of individual states which most resource theories do. In the following, we will call p the *probe state*, and q the *reference state* that p is compared to. Of course, nothing prevents us from reversing the roles of p and q to study the other half of the collective property.

Let E be some incomparability function. To emphasize the asymmetrical role of p and q, we propose the notation $E(p \parallel q)$, which is inspired by the notation for the *relative entropy* (also known as Kullback-Leibler divergence, cf. Appendix B) between 2 probability distributions, which is also asymmetric [8, p. 19]. A good measure of incomparability E between p and q should be 0 when $p \sim q$ and greater than 0 only when $p \nsim q$. That is,

$$E(p \parallel q) \neq 0 \iff p \in \mathcal{T}_\emptyset(q). \tag{4.13}$$

Unfortunately, the second difference with usual resource monotones is that the set of comparable vectors to q is not convex. Figure 4.3 illustrates why this is the case: a convex mixture of a vector from $\mathcal{T}_+(q)$ and $\mathcal{T}_-(q)$ could land in $\mathcal{T}_\emptyset(q)$.

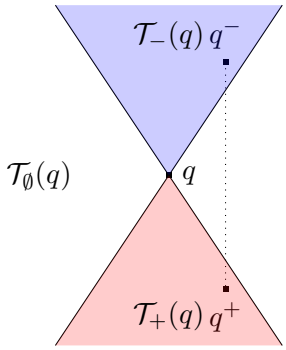


Figure 4.3: Depiction of the set $\mathcal{T}_+(q) \cup \mathcal{T}_-(q)$ not being convex, as there exists elements q^+, q^- such that the chord joining them is not entirely contained in the set.

This problem unfortunately prevents us from achieving a monotone obeying Eq. (4.13), at least directly. However, we can define two separate notions of incomparability, one as the minimal distance to the future cone of q and the other as the minimal distance to the past cone of q.

4.2.3 Future incomparability function

Considering the previous discussion, there are 2 notions of incomparability we can define. The first will be called the notion of *future incomparability*. Essentially, it is simply going to charac-

terize how far p is from the future cone of q. We propose the following definition for a future incomaparability function $E^+(p \parallel q)$, illustrated by Figure 4.4 directly on the lattice.

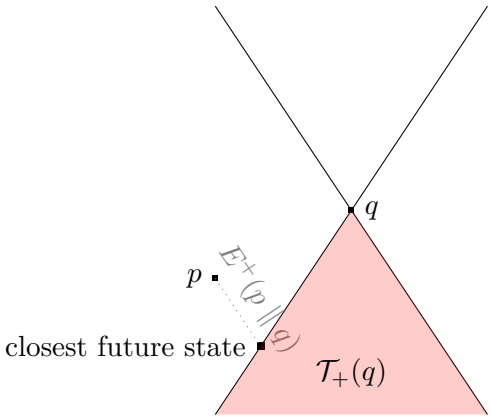


Figure 4.4: Depiction of the definition of the future incomparability function, $E^+(p \parallel q)$, which is the entropic distance from p to the closest point of the future cone of q. Visually, we expect that point to be the join.

Definition 4.1 (Future incomparability function)**.** Let $p, q \in \mathcal{P}^d$. The future incomparability $E^+(p \parallel q)$ of p to q is defined as

$$E^+(p \parallel q) = \min_{s \mid s \succ q} d(p, s). \tag{4.14}$$

Remark 4.1. From the definition it is easy to see that $p \in \mathcal{T}_+(q) \iff E^+(p \parallel q) = 0$, because then the closest state to p that also majorizes q is simply p.

Essentially, we are looking for the vector s in $\mathcal{T}_+(q)$ which is closest to p in terms of entropic distance. As hinted by Figure 4.4 expect this vector to be $p \vee q$. Lemma 1.6 gives us the tools to prove the first major result of this chapter, which essentially states that our geometric intuition is correct.

Theorem 4.1. *The future incomparability of a vector p to another vector q is the entropic distance from p to their join. Formally,*

$$E^+(p \parallel q) = d(p, p \vee q). \tag{4.15}$$

Essentially, this theorem simply states that our geometric intution concerning majorization cones is correct (in this case). Let us now prove this result.

Proof. Let us find the minimal value of $d(p, s)$ (with $s \succ q$), and let us show that it is realized for $s = p \vee q$.

$$E^+(p \parallel q) = \min_{s \mid s \succ q} d(p, s) \tag{4.16}$$

$$\overset{\text{Lemma 1.6}}{=} \min_{s \succ q} d(p, p \vee s) + d(p \vee s, s) \tag{4.17}$$

Let $s'[s] = p \vee s$. The absorption law $a \vee a = a$, along with the associativity of the join operation, implies that choosing $s = s'[s]$ (i.e. requiring that s majorizes p) makes the second term of the sum vanish, while the first remains invariant.

$$\implies E^+(p \parallel q) = \min_{s \succ q} d(p, p \vee s'[s]) + d(p \vee s'[s], s'[s]) \tag{4.18}$$

$$= \min_{s \mid s \succ q} d(p, p \vee (p \vee s)) + d(p \vee (p \vee s), p \vee s) \tag{4.19}$$

$$= \min_{s \mid s \succ q} d(p, (p \vee p) \vee s) + d((p \vee p) \vee s, p \vee s) \tag{4.20}$$

$$= \min_{s \mid s \succ q} d(p, p \vee s) + d(p \vee s, p \vee s) \tag{4.21}$$

$$= \min_{s \mid s \succ q} d(p, p \vee s) \tag{4.22}$$

and so for any choice of s, taking $p \vee s$ instead will always give a smaller distance. Because s majorizes q, $s'[s]$ majorizes q too. Moreover, to be the join of p with another vector, $s'[s]$ must majorize p as well, and so we can write

$$E^+(p \parallel q) = \min_{s' \mid s' \succ q,p} d(p, s') \tag{4.23}$$

$$= \min_{s' \mid s' \succ q,p} H(p) + H(s') - 2H(p \vee s') \tag{4.24}$$

$$= \min_{s' \mid s' \succ q,p} H(p) + H(s') - 2H(s') \tag{4.25}$$

$$= \min_{s' \mid s' \succ q,p} H(p) - H(s'). \tag{4.26}$$

The Schur-concavity of the Shannon entropy implies that the maximum value of $H(s')$ for $s' \in \{v \mid v \succ p, q\}$[3]$\subseteq \mathcal{P}^d$ is reached for a vector s' that is majorized by all other vectors in the subset. By definition, that vector is the join $p \vee q$. Thus, $E^+(p \parallel q) = H(p) - H(p \vee q)$. □

4.2.4 Past incomparability function

One could attempt the same definition for a *past incomparability* function of p to q as being the minimal distance from p to the past cone of q, $\mathcal{T}_-(q)$. However, using the same entropic distance does not yield properties as nice as we had with the join. A simple interpretation of this would be the hyperbolic geometry induced by our entropic distance, which complicates any attempts at working with the meet. Instead, we worked with the entropic quasidistance $d'(p, q) = 2H(p \wedge q) - H(p) - H(q)$ as introduced in Definition 1.15. We propose the following definition for a past incomparability function $E^-(p \parallel q)$, illustrated by Figure 4.5.

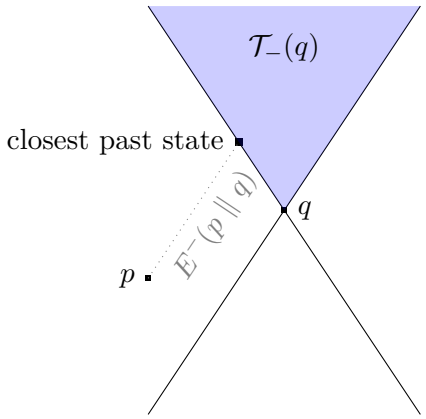


Figure 4.5: Depiction of the definition of the past incomparability function, $E^-(p \parallel q)$, which is the entropic quasidistance from p to the closest point of the past cone of q. Visually, we expect that point to be the meet.

Definition 4.2 (Past incomparability function)**.** Let $p, q \in \mathcal{P}^d$. The past incomparability $E^-(p \parallel q)$ of p to q is defined as

$$E^-(p \parallel q) = \min_{s \mid s \prec q} d'(p, s). \tag{4.27}$$

[3]This is equivalent to saying $s' \in \mathcal{T}_+(p) \cap \mathcal{T}_+(q)$.

The same interpretation can be given as the future incomparability function, though one should be careful that we are working with a quasidistance this time around, and so geometric intuitions can be deceiving.

Remark 4.2. From the definition it is easy to see that $p \in \mathcal{T}_-(q) \iff E^-(p \parallel q) = 0$, because then the closest state to p that is also majorized by q is simply p.

With this quasidistance, we fall on similar properties as we did with the future incomparability monotone. Namely, the minimal quasidistance to the past cone is reached for the meet $p \wedge q$. Before proving this, we will need the following lemma, very similar to Lemma 1.6 except that this time around we cannot speak of triangular inequality because of the quasidistance.

Lemma 4.1. *Let $p, q \in \mathcal{P}^d$. For the entropic quasidistance $d'(p,q) = 2H(p \wedge q) - H(p) - H(q)$, we have*

$$d'(p,q) = d'(p, p \wedge q) + d'(p \wedge q, q). \tag{4.28}$$

Proof. Let $p, q \in \mathcal{P}^d$. Then,

$$\begin{aligned} d'(p, p \wedge q) + d'(p \wedge q, q) &= 2H(p \wedge (p \wedge q)) - H(p) - H(p \wedge q) + 2H((p \wedge q) \wedge q) \\ &\quad - H(p \wedge q) - H(q) && (4.29) \\ &= 2H(p \wedge q) - H(p \wedge q) - H(p) + 2H(p \wedge q) - H(p \wedge q) \\ &\quad - H(q) && (4.30) \\ &= 2H(p \wedge q) - H(p) - H(q) && (4.31) \\ &= d'(p,q). && \square \end{aligned}$$

We are now ready to prove the main result of this section, which again states that our geometric intuition is correct (though this time the interpretation is murkier given the quasidistance d').

Theorem 4.2. *The past incomparability of a vector p to another vector q is the entropic distance*[4] *from p to their meet. Formally,*

$$E^-(p \parallel q) = d(p, p \wedge q). \tag{4.32}$$

Proof. Let us find the minimal value of $d'(p,s)$ (with $s \prec q$), and let us show it is realized for $s = p \wedge q$.

$$\begin{aligned} E^-(p \parallel q) &= \min_{s \mid s \prec q} d'(p,s) && (4.33) \\ &\overset{\text{Lemma 4.1}}{=} \min_{s \mid s \prec q} d'(p, p \wedge s) + d'(p \wedge s, s) && (4.34) \\ &\implies \min_{s \mid s \prec q} d'(p,s) \geq \min_{s \mid s \prec q} d'(p, p \wedge s) && (4.35) \end{aligned}$$

Let $s'[s] = p \wedge s$. Because s is majorized by q, $s'[s]$ is majorized by q as well. Moreover, to be the meet of p with another vector, $s'[s]$ must be majorized by p as well.

$$\begin{aligned} \implies \min_{s \mid s \prec q} d'(p, p \wedge s) &= \min_{s' \mid s' \prec q,p} d'(p, s') && (4.36) \\ &= \min_{s' \mid s' \prec q,p} 2H(p \wedge s') - H(p) - H(s') && (4.37) \\ &= \min_{s' \mid s' \prec q,p} H(s') - H(p). && (4.38) \end{aligned}$$

[4]While the definition made no mention of the rigorous entropic distance d, this theorem brings us back to the actual distance.

The Schur-concavity of the Shannon entropy implies that the minimum value of $H(s')$ for $s' \in \{v \mid v \prec p, q\}$[5]$\subseteq \mathcal{P}^d$ is reached for a vector s' that majorizes all other vectors in the subset. By definition, that vector is the meet $p \wedge q$, and so $\min_{s' \prec q,p} H(s') - H(p) = H(p \wedge q) - H(p)$. The vector $p \wedge q$ is also part of the original subset $\{v \mid v \prec q\}$, and so plugging $s = p \wedge q$ into (4.35) shows that the LHS realizes the value $H(p \wedge q) - H(p)$ as well, and the 2 minima must therefore be equal. Moreover, at the final line we have $H(s') - H(p) = d'(s', p) = d(s', p)$, because the two distance notions are equal when vectors are comparable. □

4.3 Properties

We are now interested in the behavior of our incomparability functions under LOCC transformations of Schmidt vectors. For this section, we will denote bistochastic matrices of dimension $d \times d$ by the letter D. Recall that an equivalent definition of majorization is $p \succ q \iff \exists D \mid q = Dp$ (cf. Section 1.1.2). Now that we have defined our two incomparability functions and shown that they are equal to the (quasi)distance to the meet or join, let us show that they are monotones under a bistochastic degradation. An important consequence of such a property is that in the quantum picture, if they are increasing (resp. decreasing) monotones under a bistochastic degradation, they are a decreasing (resp. increasing) monotone under a LOCC degradation. However, we are working with 2 states, and so we can study degradations of p and of q separately.

4.3.1 Monotonicity under bistochastic degradation of the probe

Let us start with studying the future incomparability function. If we can show that $E^+(Dp \parallel q)$ is greater than $E^+(p \parallel q)$ for any bistochastic matrix D (implying $Dp \in \mathcal{T}_-(p)$), then we can promote our future incomparability function E^+ to an increasing monotone under bistochastic degradation of p, or (more importantly) to a decreasing monotone under LOCC degradation. Figure 4.6 shows why we geometrically expect this to be the case. The representation seems to indicate that the states with the lowest distance to $\mathcal{T}_+(q)$ are on the edge facing towards q. All of the states on the edge visually have the same distance to the cone, like p'. But remembering the hyperbolic metric induced by the entropic distance (cf. Section 1.4.3), we can deduce that $d(p', p' \vee q) \geq d(p, p \vee q)$ for any p' on that edge.

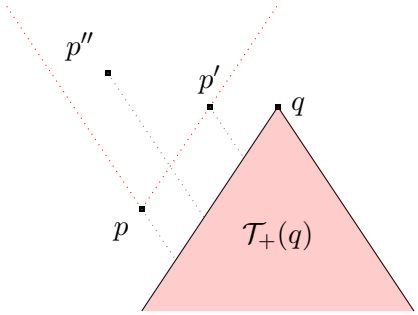


Figure 4.6: Depiction of the expected monotonicity in p of E^+. The dotted red half-cone contains all possible bistochastic degradations of p. Visually, points on the edge closest to q, like p', have the smallest distance to the future cone of q. Other choices of degradation, like p'', clearly increase the distance to the cone.

In order to rigorously prove monotonicity, we will first need the following lemma (cf. Figure 1.9).

Lemma 4.2. *In any incomparable diamond $p, q, p \wedge q, p \vee q$ we have $d(p, p \vee q) \leq d(q, p \wedge q)$.*

[5]This is equivalent to saying $s' \in \mathcal{T}_-(p) \cap \mathcal{T}_-(q)$.

Proof. Let $p, q \in \mathcal{P}^d$. We have

$$\begin{aligned} d(p, p \vee q) &= H(p) + H(p \vee q) - 2H(p \vee (p \vee q)) && (4.39)\\ &= H(p) - H(p \vee q) && (4.40)\\ &= H(p) + H(p \wedge q) - H(p \wedge q) - H(p \vee q) && (4.41)\\ &\overset{\text{supermod}}{\leq} H(p \wedge q) - H(q) && (4.42)\\ &= d(q, p \wedge q). \end{aligned}$$ □

This lemma is essentially just an equivalent way of stating supermodularity. We are now ready to prove the main theorem of this section.

Theorem 4.3. *For all bistochastic matrices D, $E^+(Dp \parallel q) \geq E^+(p \parallel q)$.*

This theorem essentially states that E^+ is an increasing monotone under bistochastic degradation of p. This proof is perhaps the trickiest of the chapter, so Figure 4.7 illustrates the different vectors of the construction on the lattice to help with comprehension.

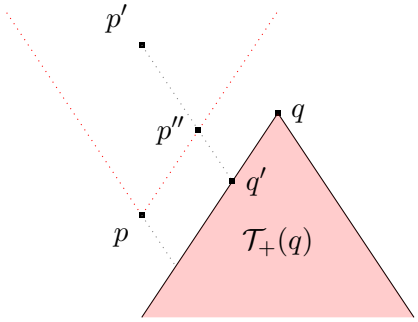


Figure 4.7: Depiction of the lattice construction for the proof of Theorem 4.3. Essentially, the purpose of q' is only to define p'' from q', which is guaranteed to be on the edge closest to q where the (visually) closest points to $\mathcal{T}_+(q)$ lie no matter the choice of p'.

Proof. Let us show that the minimal value of $E^+(p' \parallel q)$ (with $p' \prec p$) is realized for $p' = p$. Let p' be a vector majorized by p, i.e. there exists a bistochastic matrix D such that $p' = Dp$. Let $q'[p'] = p' \vee q$, and let $p''[p'] = p \wedge q'[p']$ which are both functions of the variable to minimize over, p'. By hypothesis and by definition of q', p' is majorized by both p and q', which is equivalent to p' being majorized by $p \wedge q'$. We have

$$d(p'', q') = H(p \wedge q') - H(q') \leq H(p') - H(q') = d(p', q'), \tag{4.43}$$

which can be used in the following development

$$\begin{aligned} \min_{p' \,|\, p' \prec p} E^+(p' \parallel q) &\overset{\text{Theorem 4.1}}{=} \min_{p' \,|\, p' \prec p} d(p', p' \vee q) && (4.44)\\ &\overset{\text{Def. of } q'}{=} \min_{p' \,|\, p' \prec p} d(p', q'[p']) && (4.45)\\ &\overset{(4.43)}{=} \min_{p' \,|\, p' \prec p} d(p''[p'], q'[p']) && (4.46)\\ &\overset{\text{Def. of } p''}{=} \min_{p' \,|\, p' \prec p} d(p \wedge q'[p'], q'[p']) && (4.47)\\ &\overset{\text{Lemma 4.2}}{\geq} \min_{p' \,|\, p' \prec p} d(p, p \vee q'[p']) && (4.48) \end{aligned}$$

$$\overset{\text{Def. of } q'}{=} \min_{p' \mid p' \prec p} d(p, p \vee (p' \vee q)) \tag{4.49}$$

$$= \min_{p' \mid p' \prec p} d(p, (p \vee p') \vee q) \tag{4.50}$$

$$= d(p, p \vee q) \tag{4.51}$$

$$= E^+(p \parallel q), \tag{4.52}$$

and so the minimal value of $E^+(Dp \parallel q)$ is reached for the identity degradation which leaves p invariant. □

This theorem is quite satisfying in the sense that it seems to indicate that our geometric intuitions on the lattice are valid. Moreover, we can now promote the future incomparability function to a future incomparability *monotone*. Theorem 4.3 has a few corollaries in the quantum regime.

Corollary 4.3.1. *Let $|\psi\rangle$ and $|\phi\rangle$ be two pure quantum states, and let λ_ψ and λ_ϕ be the associated Schmidt vectors. If $|\psi\rangle \overset{LOCC}{\longrightarrow} |\phi\rangle$ with probability 1, then $E^+(\lambda_\psi \parallel \lambda_\alpha) \geq E^+(\lambda_\phi \parallel \lambda_\alpha)$ with λ_α some Schmidt vector.*

Corollary 4.3.2. *Let ρ_{AB} be a bipartite quantum state, and let ρ_A and ρ_B be the reduced states. Let λ_{AB}, λ_A and λ_B be the vectors of eigenvalues of their density matrices. Then, if ρ_{AB} is separable, $E^+(\lambda_{AB} \parallel \lambda_B) \geq E^+(\lambda_A \parallel \lambda_B)$ and $E^+(\lambda_{AB} \parallel \lambda_A) \geq E^+(\lambda_B \parallel \lambda_A)$.*

These corollaries, while not very useful by themselves, still hold some interpretational value, because they show that in the sense of our future incomparability monotone, some states must be more incomparable than others relatively to some set states. Corollary 4.3.1, while involving some nondescript Schmidt vector λ_α, essentially states that it is not possible to create a state that is more future-incomparable to λ_α through LOCC than the original state λ_ψ. Corollary 4.3.2 is perhaps slightly less obvious, but states that if a joint state is separable, then the joint state is more future-incomparable to each of the reduced states than the reduced states are to each other.

One would hope that the analogue of Theorem 4.3 would hold for the measure of past incomparability E^-, however a bistochastic matrix D such that $E^-(Dp \parallel q) < E^-(p \parallel q)$ can exist, but a bistochastic matrix D' such that $E^-(D'p \parallel q) > E^-(p \parallel q)$ can exist too. Consider the following counterexample. Let $q = (0.6, 0.4)$, $p = (0.7, 0.29, 0.01)$, $p' = p \wedge q$ and $p'' = (0.7, 0.15, 0.15)$. One can verify that p majorizes both p' and p'', yet $E^-(p \parallel q) = 0.0939$ bits, $E^-(p' \parallel q) = 0$ bits and $E^-(p'' \parallel q) = 0.171$ bits. Figure 4.8 gives a geometric intuition.

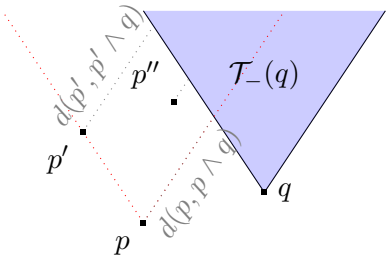


Figure 4.8: Depiction of why monotonicity in p of E^- fails. The dotted red half-cone contains all possible bistochastic degradations of p. We can of course make E^- decrease by going straight towards the past cone of q, but the issue arises when sticking to the left edge, as the hyperbolic geometry makes E^- slightly increase. This explains why there is no monotonicity in p.

4.3.2 Monotonicity under bistochastic degradation of the reference

Let us now turn our attention to bistochastic degradations of q, and attempt to promote our incomparability functions to monotones. Figures 4.9 and 4.10 show the monotonicity relations we expect. The proofs of these properties are simpler than for probe degradation, and can be found in Appendix E. Starting with the future incomparability function, if we can show that $E^+(p \parallel Dq) \leq E^+(p \parallel q)$ for any bistochastic matrix D, then E^+ would be a decreasing monotone, like we expect.

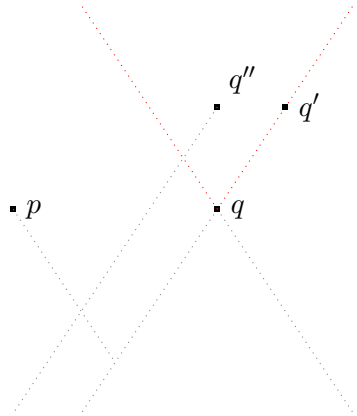


Figure 4.9: Depiction of the expected monotonicity in q of E^+. The dotted red half-cone contains all possible bistochastic degradations of q. Visually, p has highest E^+ relative to states on the furthest edge from p, like q', and all other choices decrease E^+, like q''. The cones of q, q' and q'' have not been fully drawn and shaded to avoid visual cluttering.

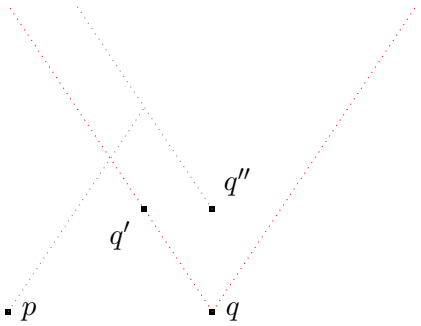


Figure 4.10: Depiction of the expected monotonicity in q of E^-. The dotted red half-cone contains all possible bistochastic degradations of q. Visually, p has lowest E^- relative to states on the closest edge to p, like q', and all other choices increase E^-, like q''. The cones of q, q' and q'' have not been fully drawn and shaded to avoid visual cluttering.

Theorem 4.4. *For all bistochastic matrices D, $E^+(p \parallel Dq) \leq E^+(p \parallel q)$.*

Let us turn our attention to E^-. This time around, the expected property for E^- does hold, and we do have monotonicity in q. If we can show that $E^-(p \parallel Dq) \geq E^-(p \parallel q)$ for any bistochastic matrix D, then E^- would be an increasing monotone, like we expect.

Theorem 4.5. *For all bistochastic matrices D, $E^-(p \parallel Dq) \geq E^-(p \parallel q)$.*

To conclude this section, we will simply note that Theorems 4.4 and 4.5 have similar corollaries to those of Theorem 4.3.

Remark 4.3. While Theorems 4.4 and 4.5 imply that the maximal value of E^+ and the minimal value of E^- are reached for the identity degradation of q, other degradations can also lead to the same value. This is made clearer by comparing Figure 4.6 for p monotonicity and Figures 4.9, 4.10 for q monotonicity, and we can see that the situation in p and q is a bit different. In p, the minimal value is reached *only* for $Dp = p$, whereas in q there are other states in $\mathcal{T}_-(q)$ that lead to the same value of the incomparability monotone.

4.3.3 Compositions

Now that we have these monotones and have proved some monotonicity properties, we could attempt to build compositions of the two that would have some nice properties. We propose the following definitions.

Definition 4.3 (Distance-like incomparability function)**.** Let $p, q \in \mathcal{P}^d$. The distance-like incomparability function F is defined as

$$F(p \parallel q) = E^+(p \parallel q) - E^-(p \parallel q). \tag{4.53}$$

Definition 4.4 (Area-like incomparability function)**.** Let $p, q \in \mathcal{P}^d$. The area-like incomparability function G is defined as

$$G(p \parallel q) = E^+(p \parallel q)E^-(p \parallel q). \tag{4.54}$$

Both of these definitions have interesting properties that might be desirable depending on different use cases. For F, a nice property is that Theorems 4.4 and 4.5 guarantee that F is a decreasing monotone under bistochastic degradation of q, and numerical testing seems to indicate that it is monotone under bistochastic degradation of p (which seems intuitive in the geometrical representation). However, F is not necessarily positive.

For G, the advantage is that it is positive, and it does satisfy Eq. (4.13), because $E^+(p \parallel q) = 0 \iff p \in \mathcal{T}_+(q)$ and $E^-(p \parallel q) = 0 \iff p \in \mathcal{T}_-(q)$. The only region where $G(p \parallel q) \neq 0$ is $\mathcal{T}_\emptyset(q)$, which is a very nice property for a measure of incomparability. However, G loses the monotonicity under bistochastic degradation of p and q.

4.4 Discussion

We will voluntarily keep the discussion in this section short because a better interpretation of E^+ and E^- will be made possible with the properties of the Shannon entropy we uncover in Chapter 5. We also postpone discussing applications of those 2 monotones to Chapter 5. While the incomparability monotones did not explain the behavior of the Rényi criteria from Section 4.1.3, the quantities E^+ and E^- are still interesting in their own right. The properties of the distance-like and area-like functions F and G could also be interesting to study, but another direction was chosen. Of the two, G seems like the more promising candidate for a measure of incomparability between two vectors.

While we did take inspiration from a resource-theoretic approach, we have not defined a full resource theory. Indeed, although we have identified free states as being the future-comparable or past-comparable states of q, it is not clear what the free operations of the theory might be. Perhaps some more general notion of operation should be defined which would act on both p and q, and where p ending in a reachable region for q would make it free because it is already reachable from q. This (potential) new type of relative resource theory might be interesting in considering the intrinsic value of having a diverse set of states: in the quantum picture, if $\lambda \nsim \lambda'$, then λ can reach some states that λ' can't (e.g. in a LOCC context). In this sense, incomparability seems like it enables *diversity* in reachable states.

Therefore, incomparability seems like a desirable property, and it is this idea that was explored further. More precisely, now that we have quantified the incomparability between a probe state and a reference state, the question that ended up being interesting to ask was how can we generalize this notion of incomparability to more than one single reference state?

Chapter 5

Quantification of incomparability for sets of states

This chapter is a direct continuation of Chapter 4. The question that this chapter tries to answer is the following: how can we generalize the notion of incomparability function of a probe state, relative to a *set of reference states* ? For the rest of this master thesis, we will call this set of reference states the *bank*. Because we are interested in quantum applications and LOCC transformations, which are confined to the $\mathcal{T}_+$ sets, we will only discuss the generalization of E^+. We did not attempt it, but we believe that reversing the whole discussion of this chapter is also possible for other theories such as quantum thermodynamics, which work with the $\mathcal{T}_-$ sets instead.

5.1 Entropic volume of majorization cones

5.1.1 Venn diagrams and the inclusion-exclusion principle

Let us remain in the 2 state context for now. Incomparability seems to enable *diversity* in the set of LOCC-reachable states from p and q. In some sense, we would expect E^+ to characterize how many states are only accessible from p, but not from q. The set-theoretic intuition is helpful here: we are looking to compute the volume (which is directly linked to the amount of reachable states) $V(\mathcal{T}_+(p)\backslash\mathcal{T}_+(q))$, where V is a measure that sends the cone of a state to a real number. For sets A and B, set theory gives us the following expression

$$V(A\backslash B) = V(A) - V(A \cap B), \tag{5.1}$$

which is illustrated with Venn diagrams in Figure 5.1.

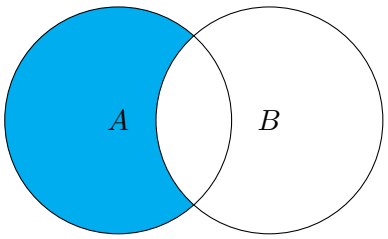


Figure 5.1: Venn diagram of two sets A, B with non-empty intersection. $A\backslash B$ is colored in cyan.

The key insight which we will use extensively for this chapter, which to our knowledge has not been used in the QRT field before, is that $\mathcal{T}_+(p) \cap \mathcal{T}_+(q) = \mathcal{T}_+(p \vee q)$ (we had briefly mentioned this relationship in Figure 1.4). It is very important to note at this stage that while visually similar, the notation for the lattice-theoretic meet and join $\wedge$ and $\vee$ should not be confused with

the notations for the set-theoretic union and intersection $\cup$ and $\cap$. With this insight, we can rewrite Eq. (5.1) with $A = \mathcal{T}_+(p)$ and $B = \mathcal{T}_+(q)$ as

$$V(\mathcal{T}_+(p)\backslash\mathcal{T}_+(q)) = V(\mathcal{T}_+(p)) - V(\mathcal{T}_+(p \vee q)). \tag{5.2}$$

Thankfully, it is possible to generalize this idea further. For example, for three sets A, B, C we have the statement

$$V(A\backslash(B \cup C)) = V(A) - V(A \cap B) - V(A \cap C) + V(A \cap B \cap C), \tag{5.3}$$

which is illustrated with Venn diagrams in Figure 5.2.

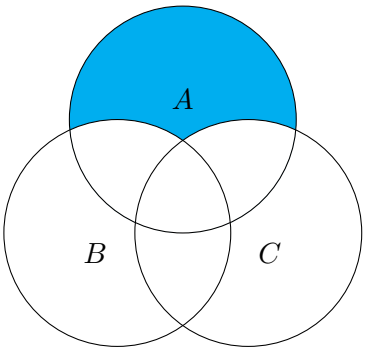


Figure 5.2: Venn diagram of three sets A, B, C with non-empty intersections. $A\backslash(B \cup C)$ is colored in cyan.

The idea is essentially that when removing both $A \cap B$ and $A \cap C$ from A, the region $A \cap B \cap C$ gets removed twice, and we must add it back to get the correct set size. This formula can then be rewritten with $A = \mathcal{T}_+(p)$, $B = \mathcal{T}_+(q_1)$ and $C = \mathcal{T}_+(q_2)$ to compute the volume of the region that is *only* LOCC-reachable from p, and not from the bank $\{q_1, q_2\}$. The fully general set-theoretic formula is known as the *inclusion-exclusion principle*, which gives the following statement.

Theorem 5.1 (Inclusion-exclusion principle [32, p. 99])**.** *Let $A_i \subseteq \Omega, 0 \leq i \leq k$, and let V be a measure over Ω. We have*

$$V\left(A_0\backslash\left(\bigcup_{i=1}^{k} A_i\right)\right) = V(A_0) + \sum_{\emptyset \neq J \subseteq \{1,\ldots,k\}} (-1)^{|J|} V\left(A_0 \bigcap_{i \in J} A_i\right), \tag{5.4}$$

where the sum $\sum_{\emptyset \neq J \subseteq \{1,\ldots,k\}}$ denotes the sum over all possible combinations J of elements from $\{1, \ldots, k\}$ (excluding the combination with no elements) and where $|J|$ denotes the cardinality of the set J.

Several equivalent versions of Theorem 5.1 exist, but this version is the relevant one in our context. We propose the following corollary.

Corollary 5.1.1 (LOCC inclusion-exclusion)**.** *Let $p, q_1, \ldots, q_k \in \mathcal{P}^d$, and let V be a measure over $\mathcal{P}^d$. Then*

$$V\left(\mathcal{T}_+(p)\backslash\left(\bigcup_{i=1}^{k} \mathcal{T}_+(q_i)\right)\right) = V\left(\mathcal{T}_+(p)\right) + \sum_{\emptyset \neq J \subseteq \{1,\ldots,k\}} (-1)^{|J|} V\left(\mathcal{T}_+\left(p \bigvee_{i \in J} q_i\right)\right), \tag{5.5}$$

where $p \bigvee_{i \in J} q_i$ is defined as $p \vee q_{j_1} \vee \cdots \vee q_{j_n}$ (with $\{j_1, \ldots, j_n\} = J$), is the volume of the set of states that is only LOCC-reachable from p but not from the bank of states $\{q_1, \ldots, q_k\}$.

5.1.2 Generalization of incomparability monotones to a bank of states

The volume of convex polytopes in Weyl chambers (which our majorization cones are) under a Euclidean measure can be computed explicitly. Unfortunately, no closed-form solution exists past dimension 3, although algorithms do exist to compute them in any dimension [15]. However, compute time and memory requirements can quickly become limiting as the dimensions increase, rendering exact methods difficult to use [33].

There is an additional difficulty in the field of LOCC transformations: quantum states are not distributed evenly on the canonical Weyl chamber of a Δ_{d-1} simplex. This is because the Schmidt decomposition induces a non-euclidean measure, called a Haar measure (which essentially captures the density of quantum states in the Schmidt vector space). Volume formulae taking this Haar measure into account do exist, and are useful in quantifying things like the average entanglement in a region of density matrices [34].

However, we will not study them further in this master thesis. Indeed, in terms of an entanglement QRT, it is not entirely clear why a Schmidt vector shared by many quantum states should be weighted more than a Schmidt vector that is shared by few quantum states. This is because states with the same Schmidt vector are equivalent up to a local change of basis, and can thus be interchanged without losing any resource. Of course, we are not trying to argue that assigning a non-uniform weight to Schmidt vectors is necessarily useless, but rather that it is not clear why there couldn't exist other relevant choices of weighting than the density of LOCC-equivalent quantum states.

With this in mind, we postulate that the Shannon entropy is itself a measure of the volume of majorization cones under some unknown measure. This makes our future incomparability monotone $E^+(p \parallel q) = H(p) - H(p \vee q)$ from Chapter 4 much easier to interpret: according to Eq. (5.6) it measures the volume of states that are reachable from p, but not from q. For this reason, it is perhaps more natural to call E^+ the *uniqueness entropy*. We propose the following generalization of E^+.

Definition 5.1 (Uniqueness entropy)**.** Let $p, q_1, \dots, q_k \in \mathcal{P}^d$. The uniqueness entropy of p relative to the bank $\{q_1, \dots, q_k\}$, which we note $E^+(p \parallel q_1, \dots, q_k)$, is defined as

$$E^+(p \parallel q_1, \dots, q_k) = H(p) + \sum_{\emptyset \neq J \subseteq \{1,\dots,k\}} (-1)^{|J|} H\left(p \bigvee_{i \in J} q_i\right). \tag{5.6}$$

A simple example with three states gives $E^+(p \parallel q_1, q_2) = H(p) - H(p \vee q_1) - H(p \vee q_2) + H(p \vee q_1 \vee q_2)$. This is better understood with the cone diagram in Figure 5.3 (which now has a new Venn-like interpretation).

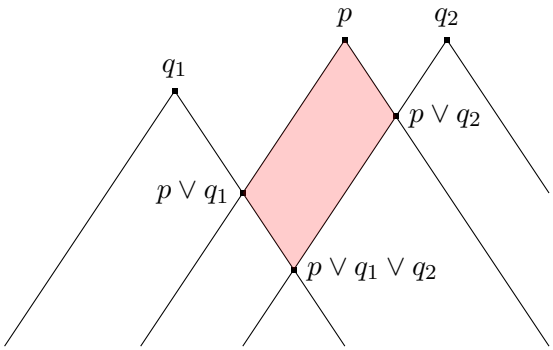


Figure 5.3: Geometric representation of the uniqueness entropy $E^+(p \parallel q_1, q_2)$ (shaded in red) for 3 states $p, q_1, q_2 \in \mathcal{P}^d$. The analogy with Figure 5.2 is quite clear: we add back $H(p \vee q_1 \vee q_2)$ because it is already removed twice in the formula.

A small overview of the essentials of measure theory is available in Appendix F. We will also denote the set of all future majorization cones by $\mathbb{T}_+$. While this definition and postulate may not seem too far-fetched considering the previous discussion, this essentially means that if E^+ behaves like a volume, then a set function μ_0 which sends cones onto the entropy of their tip p is a valid measure of the set $\mathcal{P}^d$. In other words, the volume of a cone would be entirely determined by its tip (meaning that μ_0 would be an unusual set function). That this might be possible is not too surprising however, because existing formulae for computing the volume of convex polytopes already only use the vertices of the polytope [35]. Moreover, algorithms to compute the location of the vertices of the majorization cone of any vector p do exist, and Ref. [15] even provides a Mathematica implementation[1] (in the context of thermomajorization[2] which is similar). Therefore, the volume of a majorization cone is entirely determined by its tip.

Some work did go into studying the basics of measure theory and defining a rigorous set function which would send cones onto the entropy of their tip, which would then be a valid measure for majorization cones if it could verify the properties of countable additivity and positivity on the σ-algebra (cf. Def. F.2) generated by the majorization cones. Such a result would be very interesting, and would give new insight into the nature of entropy on the majorization lattice.

For this to be possible we need our set function to work on a given collection of subsets of $\mathcal{P}^d$ of interest. We are only really interested in the measure of the cones, however a rigorous measure must at least be defined on a σ-algebra. We can generate the σ-algebra from the true collection of interest: the σ-algebra of our majorization cones contains all the possible unions, complements and intersections of cones. Intersections are easy because $\mathcal{T}_+(p) \cap \mathcal{T}_+(q) = \mathcal{T}_+(p \vee q)$ (and so they form a π-system), but unions and complements of cones are not necessarily cones and so the naive set function which sends cones on the entropy of their tip is not directly defined for those. Moreover, for μ_0 to be a measure, it needs to be countably additive. However, countable additivity can only be shown for disjoint sets, but the set of future cones does not contain a single pairwise disjoint set[3] (but the σ-algebra generated from the majorization cones $\sigma(\mathbb{T}_+)$ does contain disjoint sets). It is unfortunately not clear if and how one could extend the naive set function to $\sigma(\mathbb{T}_+)$. We believe that a recursive definition for an extended set function μ_{ext} sending a set A to $\mu_0(A)$ if A is a cone, and to $\mu_{\text{ext}}(A') + \mu_{\text{ext}}(A'') - \mu_{\text{ext}}(A' \cap A'')$ if A is the union of 2 sets A' and A'' might be enough, however proving countable additivity for such a recursive function seemed out of scope for this master thesis and was not studied further.

5.1.3 Expected properties

Unfortunately, without a clear set function we cannot work on showing finite additivity, and so we will not show that an entropy-based set function is rigorously a volume. However, we will still attempt to show that our uniqueness entropy E^+ holds properties compatible with the geometrical intuition one could have for such a volume. For any $p, q_1, \dots, q_k, q_{k+1} \in \mathcal{P}^d$, we have

1. **Commutativity:** $E^+(p \parallel q_1, \dots, q_i, \dots, q_j, \dots, q_k) = E^+(p \parallel q_1, \dots, q_j, \dots, q_i, \dots, q_k)$ for any $i \neq j$.
2. **Empty volume:** $E^+(\overline{\delta}_d \parallel q_1, \dots, q_k) = 0$.
3. **Absorption in p:** $E^+(p \parallel q_1, \dots, q_k) = 0$ if $\exists i \leq k \mid p \succ q_i$.
4. **Absorption in q:** $E^+(p \parallel q_1, \dots, q_k, q_{k+1}) = E^+(p \parallel q_1, \dots, q_k)$ if $\exists i \leq k \mid q_{k+1} \succ q_i$.

[1] https://github.com/AdeOliveiraJunior/Thermal-Cones

[2] It is interesting to note at this point that the volume of thermal cones has already been studied in Ref. [15], and that they also give an interpretation of the volume being akin to a resource. Moreover, they also propose a euclidean and Haar volume for entanglement cones. However, they studied the volume of individual cones, and thus did not attempt to characterize the volume uniquely accessible from a state.

[3] No future cone is disjoint because they are all guaranteed to at least contain $\overline{\delta}_d \coloneqq (1, 0, \dots, 0)$.

5. **Positivity:** $E^+(p \parallel q_1, \ldots, q_k) \geq 0$.

6. **Increasing monotonicity in p:** $E^+(Dp \parallel q_1, \ldots, q_k) \geq E^+(p \parallel q_1, \ldots, q_k)$ for any bistochastic matrix D.

7. **Decreasing monotonicity in q:** $E^+(p \parallel q_1, \ldots, Dq_i, \ldots, q_k) \leq E^+(p \parallel q_1, \ldots, q_i, \ldots, q_k) \forall i \leq k$, for any bistochastic matrix D.

The key insight needed to prove these properties is given by the following lemma (proven in Appendix G), which lends itself very well to induction proofs.

Lemma 5.1. *Let $p, q_1, \ldots, q_k, q_{k+1} \in \mathcal{P}^d$. Then,*

$$E^+(p \parallel q_1, \ldots, q_k, q_{k+1}) = E^+(p \parallel q_1, \ldots, q_k) - E^+(p \vee q_{k+1} \parallel q_1, \ldots, q_k). \tag{5.7}$$

It turns out that only Props. 5, 6 and 7 are difficult to prove. The other proofs come fairly naturally from Definition 5.1, see Appendix G. Note that Prop. 1 implies that the ordering of the bank is arbitrary, and so we can always make any vector the last vector of the bank, which is useful for proving Props. 3, 4 and 7 as showing the property for $i = k$ is sufficient. In order to prove Props. 5, 6 and 7, we will need an additional small lemma.

Lemma 5.2. *Let $p, q_1, \ldots, q_k, \in \mathcal{P}^d$. Then,*

$$E^+(p \parallel q_1, \ldots, q_{k-1}, q_k) = E^+(p \parallel q_1, \ldots, q_{k-1}, p \vee q_k). \tag{5.8}$$

Proof. Immediate from commutativity, associativity and idempotency of the join (cf. Section 1.3.5). □

Theorem 5.2 (Monotonicity in p and q of the uniqueness entropy)**.** *Let $p, q_1, \ldots, q_k \in \mathcal{P}^d$. Then, for any bistochastic matrix D, E^+ is an increasing monotone in p and a decreasing monotone in q. Formally,*

$$\begin{cases} E^+(p \parallel q_1, \ldots, q_k) \leq E^+(Dp \parallel q_1, \ldots, q_k) \\ E^+(p \parallel q_1, \ldots, q_k) \geq E^+(p \parallel q_1, \ldots, Dq_k). \end{cases} \tag{5.9}$$

Proof. We will do an induction proof. Let us first assume that $E^+(p \parallel q_1, \ldots, q_i)$ is monotonically decreasing in q for some value of i, and let us show that this property then propagates to $i+1$, meaning that $E^+(p \parallel q_1, \ldots, q_i, q_{i+1}) - E^+(p \parallel q_1, \ldots, q_i, Dq_{i+1}) \geq 0$. First, by using Lemma 5.1, for any bistochastic matrix D we get

$$\begin{aligned} &E^+(p \parallel q_1, \ldots, q_i, q_{i+1}) - E^+(p \parallel q_1, \ldots, q_i, Dq_{i+1}) \\ &= E^+(p \parallel q_1, \ldots, q_i) - E^+(p \vee q_{i+1} \parallel q_1, \ldots, q_i) - E^+(p \parallel q_1, \ldots, q_i) \\ &\quad + E^+(p \vee Dq_{i+1} \parallel q_1, \ldots, q_i) \qquad (5.10) \\ &= E^+(p \vee Dq_{i+1} \parallel q_1, \ldots, q_i) - E^+(p \vee q_{i+1} \parallel q_1, \ldots, q_i), \qquad (5.11) \end{aligned}$$

which is greater or equal to 0 if $E^+(p \parallel q_1, \ldots, q_i)$ is monotonically increasing in p for i (because $p \vee q_{i+1} \succ p \vee Dq_{i+1}$ is true for any D [9, p. 35]). Let us now show that $E^+(p \parallel q_1, \ldots, q_i)$ being monotonically decreasing in q for i implies that it is also monotonically increasing in p for i. Let us first assume that $p \prec q_i$, i.e. $\exists D' \mid D'q_i = p$, and consider the following statement, which is true $\forall\, D$ because of Prop. 4.

$$E^+(Dp \parallel q_1, \ldots, q_i, p) = E^+(Dp \parallel q_1, \ldots, D'q_i, p) \tag{5.12}$$

$$\begin{aligned} \implies E^+(Dp \parallel q_1, \ldots, q_i) - E^+(Dp \vee p \parallel q_1, \ldots, q_i) &= E^+(Dp \parallel q_1, \ldots, q_{i-1}, p) \\ &\quad - E^+(\underbrace{Dp \vee D'q_i}_{=p} \parallel q_1, \ldots, q_{i-1}, p) \qquad (5.13) \end{aligned}$$

$$\implies E^+(Dp \parallel q_1, \ldots, q_i) - E^+(p \parallel q_1, \ldots, q_i) = E^+(Dp \parallel q_1, \ldots, q_{i-1}, p), \tag{5.14}$$

where the term $E^+(p \parallel q_1, \ldots, q_{i-1}, p)$ vanishes by Prop. 3. To show increasing monotonicity in p, we only need to show that the RHS of (5.14) is greater or equal to 0. By hypothesis, we know that $E^+(Dp \parallel q_1, \ldots, q_{i-1}, p)$ is monotonically decreasing in q. In particular, applied on the vector p in the bank,

$$E^+(Dp \parallel q_1, \ldots, q_{i-1}, p) \geq E^+(Dp \parallel q_1, \ldots, q_{i-1}, Dp) = 0 \tag{5.15}$$

by Prop. 3, and so the RHS of (5.14) is greater or equal to 0. Moreover, Lemma 5.2 guarantees that if q_i does not majorize p, we can simply replace it with $p \vee q_i$, which does majorize p. Therefore, we can lift the $p \prec q_i$ restriction. We have thus shown that if $E^+(p \parallel q_1, \ldots, q_i)$ is monotone in q for some value of i, then it is too in p for i, which implies monotonicity in q for $i+1$. Finally, Theorem 4.4 provides the base case for $i = 1$, and so it is monotonically increasing and decreasing in p and q (respectively) for all $i \in \mathbb{N}$. □

With this new result, the last missing property is easy to show.

Theorem 5.3 (Positivity of the uniqueness entropy)**.** *Let $p, q_1, \ldots, q_k \in \mathcal{P}^d$. Then,*

$$E^+(p \parallel q_1, \ldots, q_k) \geq 0. \tag{5.16}$$

Proof. Using Lemma 5.1, we have

$$E^+(p \parallel q_1, \ldots, q_i) = E^+(p \parallel q_1, \ldots, q_{i-1}) - E^+(\underbrace{p \vee q_i}_{\succ p} \parallel q_1, \ldots, q_{i-1}), \tag{5.17}$$

which is greater or equal to 0 because of the increasing monotonicity in p (cf. Theorem 5.2) applied on the second term of the RHS, for any value of i. □

In the quantum picture, corollaries analoguous to those of Theorems 4.3 and 4.4 are also applicable here, e.g. if the volumic intuition is rigorously proven, an entangled state can never reach more unique states (relative to a bank) after any LOCC transformation than before.

5.1.4 New intuition for supermodularity from entropic volumes

If one could rigorously prove that an entropy-based set function would give a true measure on the σ-algebra of majorization cones, then one would immediately have a new proof of supermodularity of the Shannon entropy. This is because any measure μ is monotone (cf. Theorem F.1), in the sense that for any 2 sets A, B, if $A \subseteq B$, then $\mu(A) \leq \mu(B)$. Consider now Figure 5.4.

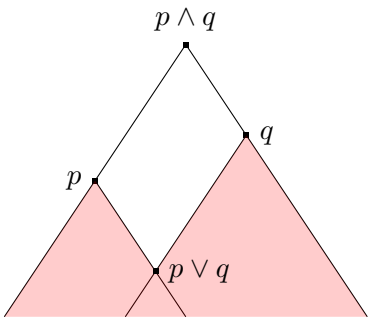


Figure 5.4: Depiction of the implication of supermodularity by monotonicity of μ. The region $\mathcal{T}_+(p) \cup \mathcal{T}_+(q)$ is shaded in red. Clearly, $\mathcal{T}_+(p) \cup \mathcal{T}_+(q) \subseteq \mathcal{T}_+(p \wedge q)$, with equality iff $p \sim q$.

It is easy to see that we have

$$\mathcal{T}_+(p) \cup \mathcal{T}_+(q) \subseteq \mathcal{T}_+(p \wedge q) \implies \mu\left(\mathcal{T}_+(p) \cup \mathcal{T}_+(q)\right) \leq \mu\left(\mathcal{T}_+(p \wedge q)\right). \tag{5.18}$$

However, by the inclusion-exclusion principle, we have

$$\mu\left(\mathcal{T}_+(p) \cup \mathcal{T}_+(q)\right) = \mu\left(\mathcal{T}_+(p)\right) + \mu\left(\mathcal{T}_+(q)\right) - \mu\underbrace{\left(\mathcal{T}_+(p) \cap \mathcal{T}_+(q)\right)}_{=\mathcal{T}_+(p \vee q)}, \tag{5.19}$$

where all the sets on the RHS are valid majorization cones. If μ is such that it sends every cone on the entropy of its tip, we would get

$$\mu\left(\mathcal{T}_+(p)\right) + \mu\left(\mathcal{T}_+(q)\right) - \mu\left(\mathcal{T}_+(p \vee q)\right) \leq \mu\left(\mathcal{T}_+(p \wedge q)\right) \tag{5.20}$$

$$\implies H(p) + H(q) - H(p \vee q) \leq H(p \wedge q), \tag{5.21}$$

which is precisely the supermodularity property (cf. Theorem 1.3). While we have not shown rigorously that such a set function exists, we believe that the properties of E^+ are a good sign that such a measure is possible.

5.2 Resource-State Selection Strategies

While most of this master thesis has been very mathematical, this section goes over the main quantum application we have found for the quantities studied in Chapters 4 and 5.

5.2.1 Definition

Let us assume Alice and Bob are have a pre-shared bank of entangled states $\{q_1, \ldots, q_k\} = Q$, and let us assume they are required to use their bank of states for a LOCC protocol, in which they decide through a CC channel on successive target states that they need to construct, e.g. for some form of distributed quantum computing. In general, the states in their possession can be different, i.e. they could have received some of their entangled states from one provider, and the rest from another provider with a different preparation standard. The problem we are trying to solve is the following. Suppose they decide that they need to construct a target with Schmidt vector t, and after looking in their bank, they realize that two states in their bank can reach t through LOCC, e.g. $q_1 \prec t$ and $q_2 \prec t$ (cf. Theorem 2.4). Should they use q_1 or q_2 to construct the target t ?

As far as we know, while research has gone into finding and constructing a state which is the Optimal Common Resource (OCR) of a given set of possible targets $\{t_1, \ldots, t_n\}$ in the sense that it can reach all of the possible targets (the OCR of the set is $\wedge_{i=1}^n t_i$) [13, 14], the slightly different question of choosing between states of a bank that does not exclusively contain OCRs has not been studied much. If the bank contains states less entangled than the OCR, but that can also reach a given target t, then it would be a waste to use up an OCR when a less entangled state could do the job. If several non-OCR states can reach t, how do Alice and Bob choose which to use ? We propose the following definition.

Definition 5.2 (Resource-State Selection Strategy)**.** A Resource-State Selection Strategy (RSSS) is any decision algorithm that chooses which state from a pre-shared entanglement bank Q to use to construct a target entangled state t through LOCC, e.g. by minimising some loss function.

The goal of a good RSSS is then to run out of states capable of reaching successive targets (on which we might have no knowledge over) as slow as possible *on average*. We will assume that after transformation, the LOCC state is consumed (for a quantum computing task the state needs to be measured at the end of the quantum circuit to get a result)[4].

[4]In practice, the measurement might not need to fully determine the state. For instance, a valid measurement

5.2.2 Individual strategy

The simplest form of RSSS could be to use the least entangled state that can reach t. This simple idea yields the following algorithm, which we use as a benchmark.

Definition 5.3 (Individual strategy)**.** Let Q be a bank of pre-shared entangled states, and let t be a target state. The following algorithm decides which state $q \in Q$ to use to construct t.

1. Initialize the set $Q' = \emptyset$. For each $q_i \in Q$, if $q_i \prec t$, add q_i to Q', which contains all the states that can reach the target.
2. For each $q_i \in Q'$, compute $a_i = H(q_i)$.
3. Finally, construct t using the state $q_i \in Q'$ with the lowest value of a_i.

This simple strategy essentially uses up the least resourceful state each time. While simple, we believe that more sophisticated strategies might yield better results on average.

5.2.3 Uniqueness strategy

Let us now turn to the main application of E^+. A state q_i with a high uniqueness entropy relative to the rest of the bank $Q \backslash q_i$ is inherently valuable because it can reach many states that the rest of the bank can't reach. If we have no knowledge over the future targets to construct, we would want to avoid using up a state with a high E^+ as long as possible, because if we use it up instead of a state that has a low uniqueness entropy, we have a higher risk that at a later step in the protocol a target would fall in the region $\mathcal{T}_+(q_i) \backslash \left(\cup_{q_j \in Q \backslash q_i} \mathcal{T}_+(q_j) \right)$, which would now be unreachable. However, directly computing the uniqueness entropy of all the states that can reach t and deciding based on that alone is not sufficient. Consider a bank of 5 entangled states with which to construct a target t, represented in Figure 5.5.

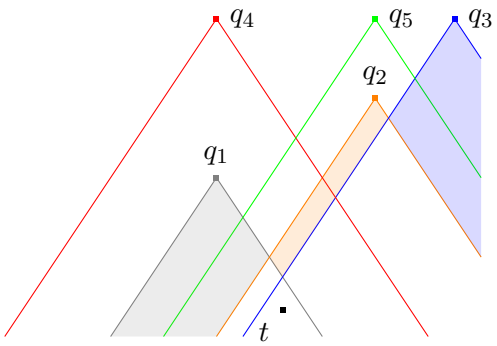


Figure 5.5: Geometric representation of a bank of 5 states, and the target t to construct. The shaded regions show some uniqueness entropies we can use to choose the state to consume (hinting that q_2 is the least valuable state even if $H(q_1) < H(q_2)$). The value of the uniqueness entropy depends on which states are kept in the reference set, which we call a *state filtering*.

Clearly, it would be a waste to use up q_4 or q_5, because they are majorized by q_1 and q_2 (respectively), and can therefore reach all of the states that q_1 and q_2 can reach, respectively. However, $E^+(q_1 \parallel Q \backslash q_1) = E^+(q_2 \parallel Q \backslash q_2) = 0$ because their cones are contained in the cones of q_4 and q_5, and so without additional restriction on the bank a uniqueness entropy criterion is not sufficient. The approach also fails if we have several copies of the same state, because then all the copies have zero uniqueness entropy. Therefore, all states should not be considered in

on a ququart could be to measure whether it is in the subspace generated by $\{|0\rangle, |1\rangle\}$ or in the subspace generated by $\{|2\rangle, |3\rangle\}$. Such a measurement would not break a superposition (and thus reduce entanglement of a joint state) if the ququart is in a $\alpha|0\rangle + \beta|1\rangle$ superposition. We will not treat such cases.

the uniqueness entropy calculation, and we call the selection algorithm to choose which states to keep in the calculation a *state filtering*. Most of the subtilities in our strategies lie in the filtering. With this in mind, we propose the following strategy.

Definition 5.4 (Uniqueness strategy)**.** Let Q be a bank of pre-shared entangled states[5], and let t be a target state. The following algorithm decides which state $q \in Q$ to use to construct t.

1. Initialize the set $Q' = \emptyset$. For each $q_i \in Q$, if $q_i \prec t$, add q_i to Q', which contains all the states that can reach the target.

2. Initialize the set $Q'' = \emptyset$. For each $q_i \in Q'$, if $\nexists q_j \in Q' \mid q_i \prec q_j, q_i \neq q_j$ and if $q_i \notin Q''$[6], add q_i to Q'', which contains only the least resourceful states that can reach the target.

3. For each $q_i \in Q''$, compute $a_i = E^+(q_i \parallel Q'' \backslash q_i)$, and initialize $b_i = 1$; for each $q_j \in Q' \backslash q_i$, if $q_j \prec q_i$, increment b_i by 1.

4. Finally, construct t using the state $q_i \in Q''$ with the lowest value of the ratio $c_i = \frac{a_i}{b_i}$.

It is interesting to note at this stage that this decision algorithm is computationaly more demanding than the entropic strategy.

Essentially, the parameter b, which we will call the *redundancy factor*, is used here to take into account the number of times one can reach a region. For example, if one of the high E^+ states left in Q'' has many copies (or majorized states), then it makes sense to use it to construct the target instead of a lower E^+ state with few copies, because using it up makes that region of the bank significantly weaker. Other strategies might choose a different a, which we will call the *loss function*. Moreover, the decision criterion is comparing the values of c, which we will call the *weighted loss function*, and could also be changed in other strategies[7].This algorithm mostly consists of state filters, where the sets Q, Q', Q'' are only defined to compare the states with those that we want. It is important to note however that the choices made here are fairly arbitrary and might be improved upon.

5.2.4 Mixed strategies

We also define an additional variation, which mixes both proposed loss functions with a specific ratio that can be adjusted.

Definition 5.5 (Mixed strategy)**.** Choose $\alpha \in [0, 1]$. The mixed strategy is the same as the uniqueness strategy (cf. Definition 5.4), except that the loss function is replaced with $a_i = \alpha H(q_i) + (1 - \alpha) E^+(q_i \parallel Q'' \backslash q_i)$ in step 3.

5.2.5 Comparison and statistical sampling

A simulation of the RSSS and successive attemps (nicknamed "games") to construct as many successive targets as possible is available on the GitHub for the project[8]. To compare the strategies depending on the mixed strategy parameter α, a bank of 9 states (of dimension 4) sampled uniformly on the Δ_{d-1} simplex and a maximally entangled state, and a sequence of 10 targets (of dimension 4), are generated. A skew parameter allows to control the Dirichlet distribution from which the targets are sampled: if the sampling is uniform then all strategies perform very poorly because strongly entangled targets are likely to be sampled and are quickly not constructible depending on how many states have already been used. The same bank and

[5] A bank may contain several copies of the same state.

[6] This ensures that if there are several copies of the same state, only one of them enters Q'', ensuring that the E^+ calculation at the next step does not yield 0 due to the copies.

[7] For example, one could choose $c = \frac{a}{b^2}$, which might yield a different behavior of the strategy in some cases.

[8] https://github.com/traaldbjerg/MajoLat

same sequence of targets are given to Algorithm 5.5 for each value of α (with a step of 0.01). This way we simulated 600000 games, and the average number of succesful targets for each value of α was recorded, both for the standard version of Def. 5.5, and a variation where we set $b_i = 1 \, \forall \, i$ in step 3 to test whether the redundancy factor is useful or not. The results of this process are shown in Figure 5.6.

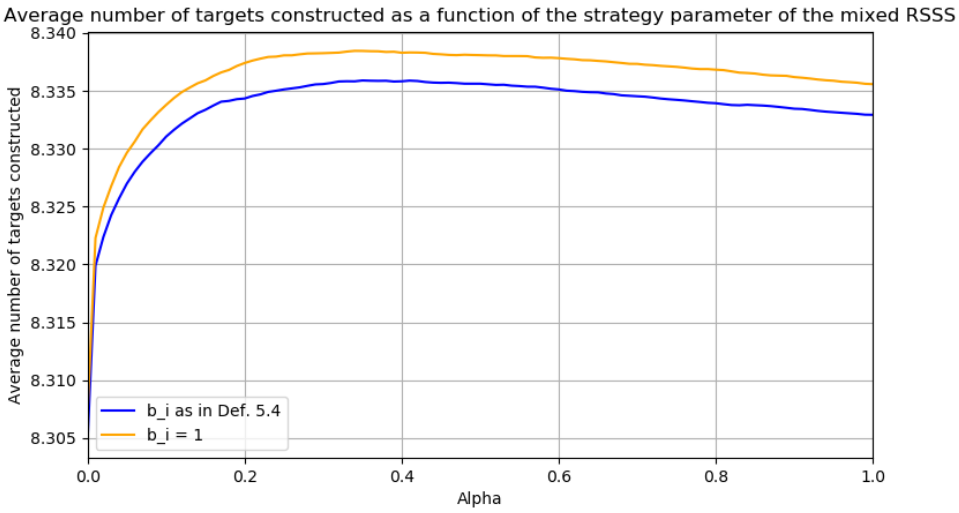


Figure 5.6: Average number of succesful target constructions depending on the value of the strategy parameter α for a bank of 9 uniformly sampled states (dimension 4) and 1 maximally entangled state, and 10 targets sampled with Dirichlet parameters $\left(2, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$. The blue curve corresponds to the algorithm of Definition 5.4, and the orange curve was initially a test to see if the redundancy parameter is helpful or not, and it seems not to be, which was a bit surprising. Note that on the blue curve the $\alpha = 0$ extremity is the uniqueness strategy, and $\alpha = 1$ is similar to the individual strategy, but also uses a redundancy factor b and a weighted loss function c. For both curves, 600000 games were simulated. The entropic strategy as in Def. 5.3 is found on the orange curve, at $\alpha = 1$, and so the curve having a maximum towards $\alpha = 0.3$ shows that we outperform the naive strategy.

5.3 Discussion

Eq. (5.6) is computationally very demanding (which is why we didn't go over 10 states for the games, as otherwise the number of combinations is sometimes very large). Most of the time, Prop. 4 allows most states to vanish from the computation and the number of combinations is low enough that the computation time is manageable. However, sometimes a target such that many states fall into Q'' is generated, and the uniqueness entropy computation is costly.

Although the variations are not very large (less than 1%), we were surprised that the uniqueness strategy performed worse than the entropic strategy. However, it is very interesting to see that there exists a nontrivial maximum at around $\alpha = 0.3$, where combining both the uniqueness entropy loss function and the entropic loss function performs better than a simple individual strategy. This seems to validate our lattice-based approach and the utility of E^+, and the mathematical theory of incomparability that we have discussed in this manuscript. Moreover, drastically different results can be reached by changing some of the state filtering choices in step 3, and we believe that further research might lead to better filterings or weigthed loss functions which could further improve the performance of the different strategies. The results for another filtering (which is closer to our original vision) is shown in Appendix H.

Chapter 6

Conclusion

In this MSc thesis, we have mainly defined the foundations of a mathematical theory of incomparability by taking inspiration from resource-theoretic approaches.

In Chapter 3, we have proven some more general forms of the properties of supermodularity and subadditivity – which Shannon entropy was known to enjoy on the majorization lattice – to all sum-concave functions. The concatenation technique used to prove the underlying majorization precursor seems fairly powerful, and might yield interesting majorization precursors in other fields that deal with entropic inequalities.

In Chapter 4, we have gone over a small improvement to coupled entropic criteria for which a majorization relation already exists, notably the majorization separability criterion. This criterion is only an improvement for incomparable Schmidt vectors, and so we defined new quantities to measure the incomparability of two distributions, and studied some of their properties under entanglement transformations.

Finally, in Chapter 5, the ideas from Chapter 4 came to fruitition as we succesfully generalized the future entropy monotone into the uniqueness entropy, yielding new geometrical intuitions explaining some of the properties of the Shannon entropy on the majorization lattice. We didn't prove a rigorous measure, but we were recently made aware of the relevance of Heyting lattices, on which many measures have already been defined, from which inspiration might be taken to define our own measure. We have also found an application to the uniqueness entropy, and shown that some more general forms of resource monotones which take into account all of the states in our possession and not only each individual state separately yields better results in some situations, giving the idea that on some level it really is a valuable notion of resource in the quantum picture. As such, we have shown that there is some merit to such approaches, and that the lattice is a useful tool to quantify collective properties. Moreover, while the improvements over a simple entropic strategy are not very large, one has to keep in mind that some of the choices made to define the different RSSS were fairly arbitrary, so we believe that further study of lattice-based quantities and state filters might yield better results. We were recently made aware of the relevance of the field of dynamic optimization theory, which is precisely concerned with such resource-selection decision algorithms.

We believe that further research in defining other RSSS, either strictly algorithmically or by defining other lattice-based quantities might yield better results. The example with the successive target constructions might have been a little artificial, but it serves as a proof-of-concept for this new approach, and we believe that reversing the discussion to define a RSSS for OCR construction might be a more practically relevant case (although more complicated because of probabilistic transformations). In another direction, we also believe that further research on an entropic measure could be interesting, as such a result could lead to the discovery of new properties of the Shannon entropy. Finally, the foundations of the mathematical theory of incomparability that we have defined in this manuscript could be applied to other QRTs than entanglement theory, which might be another fruitful research avenue.

Declaration of Generative AI and AI-assisted Technologies in the Writing Process

During the preparation of this work, the author used OpenAI's o4-mini and GPT4.5 in order to find references, discuss results and find alternative techniques. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

Bibliography

[1] A. Einstein, B. Podolsky, and N. Rosen. Can Quantum-Mechanical Description of Physical Reality Be Considered Complete? *Phys. Rev.* 47(10), 777–780 (1935).

[2] J. S. Bell. On the Einstein Podolsky Rosen Paradox. *Phys. Phys. Fiz.* 1(3), 195–200 (1964).

[3] A. Aspect, P. Grangier, and G. Roger. Experimental Realization of Einstein-Podolsky-Rosen-Bohm Gedankenexperiment: A New Violation of Bell's Inequalities. *Phys. Rev. Lett.* 49(2), 91–94 (1982).

[4] R. Horodecki et al. Quantum Entanglement. *Rev. Mod. Phys.* 81(2), 865–942 (2009).

[5] M. A. Nielsen. Conditions for a Class of Entanglement Transformations. *Phys. Rev. Lett.* 83(2), 436–439 (1999).

[6] G. Vidal. Entanglement of Pure States for a Single Copy. *Phys. Rev. Lett.* 83(5), 1046–1049 (1999).

[7] A. W. Marshall, I. Olkin, and B. C. Arnold. *Inequalities: Theory of Majorization and Its Applications*. 2nd ed. (Springer, New York, NY, 2011).

[8] T. M. Cover. *Elements of Information Theory*. 2nd ed. (Wiley-Interscience, Hoboken, NJ, 2006).

[9] B. A. Davey and H. A. Priestley. *Introduction to Lattices and Order*. 2nd ed. (Cambridge University Press, Cambridge, 2002).

[10] F. Cicalese and U. Vaccaro. Supermodularity and subadditivity properties of the entropy on the majorization lattice. *IEEE Trans. Inf. Theory* 48(4), 933–938 (2002).

[11] F. Cicalese, L. Gargano, and U. Vaccaro. Information theoretic measures of distances and their econometric applications. In *Proceedings of the 2013 IEEE International Symposium on Information Theory*, 409–413 (2013).

[12] E. Chitambar and G. Gour. Quantum resource theories. *Rev. Mod. Phys.* 91(2), 025001 (2019).

[13] G. M. Bosyk et al. Optimal Common Resource in Majorization-Based Resource Theories. *New J. Phys.* 21(8), 083028 (2019).

[14] S. Deside et al. Probabilistic pure state conversion on the majorization lattice. *Phys. Rev. Res.* 6(2), 023156 (2024).

[15] A. de Oliveira Junior et al. Geometric structure of thermal cones. *Phys. Rev. E* 106(6), 064109 (2022).

[16] A. Rényi. On Measures of Entropy and Information. In *Proceedings of the Fourth Berkeley Symposium on Mathematics, Statistics and Probability*, 547–561 (1961).

[17] K. Korzekwa. Structure of the thermodynamic arrow of time in classical and quantum theories. *Phys. Rev. A* 95(5), 052318 (2017).

[18] J. R. Johansson, P. D. Nation, and F. Nori. QuTiP: An Open-Source Python Framework for the Dynamics of Open Quantum Systems. *Comput. Phys. Commun.* 183(8), 1760–1772 (2012).

[19] M. A. Nielsen and I. L. Chuang. *Quantum computation and quantum information: 10th anniversary edition* (Cambridge University Press, Cambridge, 2010).

[20] N. Brunner et al. Bell Nonlocality. *Rev. Mod. Phys.* 86(2), 419–478 (2014).

[21] N. J. Cerf and C. Adami. Negative Entropy and Information in Quantum Mechanics. *Phys. Rev. Lett.* 79(26), 5194–5197 (1997).

[22] R. Horodecki and M. Horodecki. Information-Theoretic Aspects of Inseparability of Mixed States. *Phys. Rev. A* 54(3), 1838–1843 (1996).

[23] M. A. Nielsen and J. Kempe. Separable States Are More Disordered Globally than Locally. *Phys. Rev. Lett.* 86(22), 5184–5187 (2001).

[24] C. Weedbrook et al. Gaussian Quantum Information. *Rev. Mod. Phys.* 84, 621–669 (2012).

[25] V. Vazirani and M. Yannakakis. Market Equilibrium under Separable, Piecewise-Linear, Concave Utilities. *J. ACM* 58(3), 156–165 (2011).

[26] N. Anari et al. Nash Social Welfare for Indivisible Items under Separable, Piecewise-Linear Concave Utilities. In *Proceedings of the 29th Annual ACM-SIAM Symposium on Discrete Algorithms*, 2274–2290 (2017).

[27] E. Agliardi. A Generalization of Supermodularity. *Econ. Lett.* 68(3), 251–254 (2000).

[28] O. Rioul. Information Theoretic Proofs of Entropy Power Inequalities. *IEEE Trans. Inf. Theory* 57(1), 33–55 (2011).

[29] L. Guo, C.-M. Yuan, and X.-S. Gao. A Generalization of the Concavity of Rényi Entropy Power. *Entropy* 23(12), 1593 (2021).

[30] N. J. Cerf and C. Adami. Entropic Bell Inequalities. *Phys. Rev. A* 55(5), 3371–3374 (1997).

[31] L. Hu. Characterizing Incomparability in Quantum Information. arXiv:1803.10408. (2018).

[32] T. Tao. *An Introduction to Measure Theory* (American Mathematical Society, Providence, RI, 2011).

[33] B. Büeler, A. Enge, and K. Fukuda. Exact Volume Computation for Polytopes: A Practical Study. *Polytopes — Combinatorics and Computation.* Ed. by G. Kalai and G. M. Ziegler, 131–154 (2000).

[34] K. Zyczkowski and H.-J. Sommers. Induced measures in the space of mixed quantum states. *J. Phys. A: Math. Gen.* 34(35), 7111 (2001).

[35] B. Braden. The Surveyor's Area Formula. *Coll. Math. J.* 17(4), 326–337 (1986).

[36] M. A. Nielsen and G. Vidal. Majorization and the interconversion of bipartite states. *Quant. Inf. Comput.* 1(1), 76–93 (2001).

[37] W. Rudin. *Real and Complex Analysis.* 3rd ed. (McGraw-Hill, New York, NY, 1987).

Appendices

A Cicalese and Vaccaro's original proof of supermodularity

This section goes over the original proof of Theorem 1.3 for the Shannon entropy, taken from Ref. [10]. The proof is mainly here for reader convenience, as the original paper is not easily accessible. The proof is structured in two lemmas. Note that the convention $0 \log \frac{1}{0} = 0$ is used (an event that never happens does not contribute to entropy). We will first need the following definition.

Definition A.1 (Inversion point)**.** Let $p, q \in \mathcal{P}^d$. We say that the index $i \in \{2, \dots, n\}$ is an *inversion point* for p and q if either

$$\sum_{l=1}^{i} p_l < \sum_{l=1}^{i} q_l \quad \text{and} \quad \sum_{l=1}^{i-1} p_l > \sum_{l=1}^{i-1} q_l \tag{A.1}$$

or, vice versa, if

$$\sum_{l=1}^{i} p_l > \sum_{l=1}^{i} q_l \quad \text{and} \quad \sum_{l=1}^{i-1} p_l < \sum_{l=1}^{i-1} q_l. \tag{A.2}$$

Let $2 \leq i_1 < i_2 < \cdots < i_k \leq n$ be all and the only inversion points for $p, q \in \mathcal{P}^d$, and let

$$t = (t_1, \dots, t_n) = p \wedge q \tag{A.3}$$

and

$$s = (s_1, \dots, s_n) = \beta(p, q). \tag{A.4}$$

The first lemma is the following, where we assume $i_0 = 0$ and $i_{k+1} = n + 1$ for the sake of definiteness.

Lemma A.1. For each inversion point i_r, $r = 0, \dots, k$, we have

$$\sum_{l=i_r+1}^{i_{r+1}-1} p_l \log \frac{1}{p_l} + q_l \log \frac{1}{q_l} = \sum_{l=i_r+1}^{i_{r+1}-1} t_l \log \frac{1}{t_l}. \tag{A.5}$$

Proof. Let us assume, without loss of generality, that

$$\sum_{l=1}^{i_r} p_l > \sum_{l=1}^{i_r} q_l. \tag{A.6}$$

Then

$$\sum_{l=1}^{i_r} s_l = \sum_{l=1}^{i_r} p_l, \tag{A.7}$$

and for all $s = i_r + 1, \ldots, i_{r+1} - 1$, we also have

$$s_s = \sum_{l=1}^{s} p_l - \sum_{l=1}^{s} q_l = p_s. \tag{A.8}$$

Accordingly, we have

$$\sum_{l=1}^{i_r} t_l = \sum_{l=1}^{i_r} q_l, \tag{A.9}$$

and for all $s = i_r + 1, \ldots, i_{r+1} - 1$, we have

$$t_s = \sum_{l=0}^{s} q_l - \sum_{l=0}^{s} t_l = q_s. \tag{A.10}$$

□

As an immediate consequence of Lemma A.1, we have

$$\sum_{l=i_r+1}^{i_{r+1}-1} \left(p_l \log \frac{1}{p_l} + q_l \log \frac{1}{q_l} - t_l \log \frac{1}{t_l} \right) = 0. \tag{A.11}$$

Lemma A.2. For each inversion point i_r, $r = 1, 2, \ldots, k$, we have

$$s_{i_r} \log \frac{1}{s_{i_r}} + t_{i_r} \log \frac{1}{t_{i_r}} > p_{i_r} \log \frac{1}{p_{i_r}} + q_{i_r} \log \frac{1}{q_{i_r}}, \tag{A.12}$$

where $p_{i_r}, q_{i_r}, s_{i_r}, t_{i_r} > 0$.

Proof. Let us write p, q, s, t instead of $p_{i_r}, q_{i_r}, s_{i_r}, t_{i_r}$, respectively. Without loss of generality let us assume that for the inversion point i_r it holds that

$$\sum_{l=1}^{i_r} p_l > \sum_{l=1}^{i_r} q_l \quad \text{and} \quad \sum_{l=1}^{i_r-1} p_l < \sum_{l=1}^{i_r-1} q_l. \tag{A.13}$$

It follows that $s > p$ and $q > t$, so $s > t$. Moreover, it is not hard to see that $s + t = p + q$. Let $s = q + \Delta$, then $t = p - \Delta$ and $s + t = p + q$. Then we have

$$\begin{aligned}
s \log \frac{1}{s} + t \log \frac{1}{t} - p \log \frac{1}{p} - q \log \frac{1}{q} &= -(s \log s + t \log t) + p \log p + q \log q && \text{(A.14)} \\
&= -(q + \Delta) \log(q + \Delta) - (p - \Delta) \log(p - \Delta) \\
&\quad + p \log p + q \log q && \text{(A.15)} \\
&= (q \log q + p \log p) - ((q + \Delta) \log(q + \Delta) \\
&\quad + (p - \Delta) \log(p - \Delta)) && \text{(A.16)} \\
&\overset{\text{Jensen}}{\geq} (q + p + \Delta) \log \left(\frac{(q + p + \Delta)^2}{(q + \Delta)(p - \Delta)} \right) && \text{(A.17)} \\
&= (q + p + \Delta) \log \left(\frac{s + t + \Delta}{s + \Delta} \cdot \frac{s + t + \Delta}{t - \Delta} \right) && \text{(A.18)} \\
&= (q + p + \Delta) \log \left(\frac{(s + t + \Delta)^2}{(s + \Delta)(t - \Delta)} \right) > 0, && \text{(A.19)}
\end{aligned}$$

since $s > t$ and $\Delta < t$. The remaining case $s \leq t$ is completely symmetric, provided that we set $s = p + \Delta$ and $t = q - \Delta$. □

Note that Eq. (A.17) invokes Jensen's inequality. We have not gone over it in this master thesis, but it is sufficient to know that in the discrete case it is essentially a (less general) form of Karamata's inequality (Lemma 1.1). We are now ready to prove Theorem 1.3.

Proof of Theorem 1.3. We have

$$
\begin{aligned}
H(s)+H(t)-H(p)-H(q) &= \sum_{l=1}^{k}\left(s_l\log\frac{1}{s_l}+t_l\log\frac{1}{t_l}-p_l\log\frac{1}{p_l}-q_l\log\frac{1}{q_l}\right) && \text{(A.20)}\\
&= \sum_{r=1}^{k}\Bigg(\sum_{l=i_r+1}^{i_{r+1}-1}\left(s_l\log\frac{1}{s_l}+t_l\log\frac{1}{t_l}\right)+s_{i_r}\log\frac{1}{s_{i_r}}+t_{i_r}\log\frac{1}{t_{i_r}}\\
&\quad -p_{i_r}\log\frac{1}{p_{i_r}}-q_{i_r}\log\frac{1}{q_{i_r}}-\sum_{l=i_r+1}^{i_{r+1}-1}\left(p_l\log\frac{1}{p_l}+q_l\log\frac{1}{q_l}\right)\Bigg)\\
&\quad +\sum_{l=1}^{i_1-1}\left(s_l\log\frac{1}{s_l}+t_l\log\frac{1}{t_l}-p_l\log\frac{1}{p_l}-q_l\log\frac{1}{q_l}\right)\\
&\quad +\sum_{l=i_k+1}^{n}\left(s_l\log\frac{1}{s_l}+t_l\log\frac{1}{t_l}-p_l\log\frac{1}{p_l}-q_l\log\frac{1}{q_l}\right) && \text{(A.21)}\\
&\overset{\text{Lemma A.1}}{=} \sum_{r=1}^{k}\left(s_{i_r}\log\frac{1}{s_{i_r}}+t_{i_r}\log\frac{1}{t_{i_r}}-p_{i_r}\log\frac{1}{p_{i_r}}\right.\\
&\qquad\qquad \left.-q_{i_r}\log\frac{1}{q_{i_r}}\right) && \text{(A.22)}\\
&\overset{\text{Lemma A.2}}{>} 0. && \square
\end{aligned}
$$

B Entropy of compositions of random variables

This section is based on Ref. [8, pp. 16–22]. In practice, we are often concerned with more than one process at a given time, and so defining the uncertainty of two variables X, Y is interesting. Moreover, knowing how much measuring one variable reduces the uncertainty on the second one (and so how much information on the second process we gain) on average is valuable as well. These concepts are captured by *joint entropy*, *conditional entropy* and *mutual information*.

Definition B.1 (Joint entropy)**.** Let X and Y be random variables over alphabets $\mathcal{X}$ and $\mathcal{Y}$ with joint probability distribution $p(x, y)$. Then the joint entropy $H(X, Y)$ of X and Y is defined as

$$H(X,Y) = -\sum_{x\in\mathcal{X},y\in\mathcal{Y}} p(x,y)\log p(x,y) \tag{B.1}$$

$$= -E[\log p(x,y)]. \tag{B.2}$$

Definition B.2 (Conditional entropy)**.** Let X and Y be random variables over alphabets $\mathcal{X}$ and $\mathcal{Y}$ with marginal probability distributions $p(x)$ and $p(y)$. Then the conditional entropy $H(X|Y)$ of X knowing Y is defined as

$$H(X|Y) = \sum_{y\in\mathcal{Y}} p(y)H(X|Y=y) \tag{B.3}$$

$$= -\sum_{y\in\mathcal{Y}} p(y)\sum_{x\in\mathcal{X}} p(x|y)\log p(x|y) \tag{B.4}$$

$$= -\sum_{y\in\mathcal{Y}}\sum_{x\in\mathcal{X}} p(x,y)\log p(x|y) \tag{B.5}$$

$$= -E[\log p(x|y)]. \tag{B.6}$$

It is interesting to note that these definitions are all very consistent with each other, as the vision of entropy being the expected value of the random variable $\log\frac{1}{p(x)}$ holds even for compositions of random variables. Before defining mutual information, we introduce another mathematical tool, the Kullback-Leibler divergence $D(p \parallel q)$ of two probability mass functions p and q, sometimes called *relative entropy*.

Definition B.3 (Kullback-Leibler divergence)**.** Let p and q be two probability mass functions over an alphabet $\mathcal{X}$. Then the Kullback-Leibler divergence $D(p \parallel q)$ of p and q is defined as

$$D(p \parallel q) = \sum_{x\in\mathcal{X}} p(x)\frac{p(x)}{q(x)} \tag{B.7}$$

$$= E_p\left[\log\frac{p(x)}{q(x)}\right], \tag{B.8}$$

with conventions $0\log\frac{0}{0} = 0$, $p\log\frac{p}{0} = \infty$ and $0\log\frac{0}{q} = 0$ by continuity arguments.

This quantity can be understood as measuring how far from each other the two distributions p and q are. If they are identical, i.e. $p(x) = q(x)\ \forall x \in \mathcal{X}$, then all of the terms in the sum are $\log\frac{p(x)}{p(x)} = \log 1 = 0$ and the relative entropy is simply 0. It is important to note however that this quantity is not a proper notion of distance, as it is not symmetric in its arguments, and fails the triangular inequality. Nonetheless, this notion of distance is used to define mutual information between two random variables as the Kullback-Leibler divergence between the joint probability distribution and the product of the marginal probability distributions.

Definition B.4 (Mutual information)**.** Let X and Y be random variables over alphabets $\mathcal{X}$ and $\mathcal{Y}$ with marginal probability distributions $p(x)$ and $p(y)$ and joint distribution $p(x, y)$. Then the mutual information $I(X, Y)$ of X and Y is defined as

$$
\begin{aligned}
I(X,Y) &= D(p(x,y)||p(x)p(y)) && \text{(B.9)}\\
&= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} && \text{(B.10)}\\
&= E_{p(x,y)}\left[\log\frac{p(x,y)}{p(x)p(y)}\right] && \text{(B.11)}\\
&= E_{p(x,y)}[-\log p(x)] + E_{p(x,y)}[-\log p(y)] - E_{p(x,y)}[-\log p(x,y)] && \text{(B.12)}\\
&= E_{p(x)}[-\log p(x)] + E_{p(y)}[-\log p(y)] - E_{p(x,y)}[-\log p(x,y)] && \text{(B.13)}\\
&= H(X) + H(Y) - H(X,Y). && \text{(B.14)}
\end{aligned}
$$

Again, this definition agrees with what we would expect intuitively from a notion of the information that two random variables contain about each other: if the variables are independent, and thus contain no information about each other whatsoever (as the outcome of one does not affect the other), then their joint distribution reduces to the product of the marginal distributions, and the relative entropy in the definition becomes 0. Then, the less independent the joint distribution becomes, the farther it becomes from the product distribution, and the more measuring the outcome of one variable can tell you about the outcome of the second. It is interesting to note that the mutual information is symmetric in its arguments.

To finish this small incursion into classical information theory, it turns out that the Shannon entropy is actually the only function that satisfies all the properties that one would expect from a measure of information, such as (non-exhaustive list):

- $H(X) \geq 0$.
- $H(X, Y) = H(Y, X)$.
- $H(X, Y) = H(X) + H(Y|X)$.
- **Subadditivity**[1]**:** $H(X, Y) \leq H(X) + H(Y)$ with equality iff X and Y are independent.
- $H(X|Y) \leq H(X)$, which directly implies that $I(X, Y) \geq 0$, with equality iff X and Y are independent.

[1]This subadditivity is not related to the subadditivity on the lattice.

C Mixed state example

Let us take a qubit, and let us try to find the difference between the states ρ and ρ' from the coin-flip example in Section 2.1.2. In the computational basis, they are written

$$\rho = 1/2\,|0\rangle\langle 0| + 1/2\,|1\rangle\langle 1| = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}, \tag{C.1}$$

$$\rho' = |\psi\rangle\langle\psi| = |+\rangle\langle +| = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)\frac{1}{\sqrt{2}}(\langle 0| + \langle 1|) = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}. \tag{C.2}$$

In both cases, let us measure if the qubit is in state $|0\rangle$ or $|1\rangle$. Those measurements are described by the operators $M_0 = |0\rangle\langle 0|$ and $M_1 = |1\rangle\langle 1|$, which are simple projective measurements. For ρ, we have

$$p(0) = \mathrm{tr}\left(\,|0\rangle\langle 0|\,|0\rangle\langle 0|\,(1/2\,|0\rangle\langle 0| + 1/2\,|1\rangle\langle 1|\,)\right) \tag{C.3}$$

$$= \mathrm{tr}\left(1/2\,|0\rangle\langle 0|\,|0\rangle\langle 0| + 1/2\,|0\rangle\langle 0|\,|1\rangle\langle 1|\,\right) \tag{C.4}$$

$$= \mathrm{tr}(1/2\,|0\rangle\langle 0|\,|0\rangle\langle 0|) \tag{C.5}$$

$$= 1/2, \tag{C.6}$$

$$p(1) = \mathrm{tr}\left(\,|1\rangle\langle 1|\,|1\rangle\langle 1|\,(1/2\,|0\rangle\langle 0| + 1/2\,|1\rangle\langle 1|\,)\right) \tag{C.7}$$

$$= \mathrm{tr}\left(1/2\,|1\rangle\langle 1|\,|0\rangle\langle 0| + 1/2\,|1\rangle\langle 1|\,|1\rangle\langle 1|\,\right) \tag{C.8}$$

$$= \mathrm{tr}(1/2\,|1\rangle\langle 1|\,|1\rangle\langle 1|) \tag{C.9}$$

$$= 1/2. \tag{C.10}$$

For ρ', we have

$$p(0) = \mathrm{tr}\left(\,|0\rangle\langle 0|\,|0\rangle\langle 0|\,(1/2\,|0\rangle\langle 0| + 1/2\,|0\rangle\langle 1| + 1/2\,|1\rangle\langle 0|\,1/2\,|1\rangle\langle 1|\,)\right) \tag{C.11}$$

$$= \mathrm{tr}\left(1/2\,|0\rangle\langle 0|\,|0\rangle\langle 0| + 1/2\,|0\rangle\langle 0|\,|0\rangle\langle 1| + 1/2\,|0\rangle\langle 0|\,|1\rangle\langle 0| + 1/2\,|0\rangle\langle 0|\,|1\rangle\langle 1|\,\right) \tag{C.12}$$

$$= \mathrm{tr}(1/2\,|0\rangle\langle 0| + 1/2\,|0\rangle\langle 1|) \tag{C.13}$$

$$= 1/2, \tag{C.14}$$

$$p(1) = \mathrm{tr}\left(\,|1\rangle\langle 1|\,|1\rangle\langle 1|\,(1/2\,|0\rangle\langle 0| + 1/2\,|0\rangle\langle 1| + 1/2\,|1\rangle\langle 0|\,1/2\,|1\rangle\langle 1|\,)\right) \tag{C.15}$$

$$= \mathrm{tr}\left(1/2\,|1\rangle\langle 1|\,|0\rangle\langle 0| + 1/2\,|1\rangle\langle 1|\,|0\rangle\langle 1| + 1/2\,|1\rangle\langle 1|\,|1\rangle\langle 0| + 1/2\,|1\rangle\langle 1|\,|1\rangle\langle 1|\,\right) \tag{C.16}$$

$$= \mathrm{tr}(1/2\,|1\rangle\langle 0| + 1/2\,|1\rangle\langle 1|) \tag{C.17}$$

$$= 1/2. \tag{C.18}$$

These are the expected results, and the difference between the pure and the mixed state are not so clear in this case. However, something interesting happens if we now try to measure the qubit in the dual basis, i.e. measuring along operators $M_+ = |+\rangle\langle +|$ and $M_- = |-\rangle\langle -|$. The calculation is easier in the dual basis, in which $\rho' = |+\rangle\langle +|$, and so we directly get $p(+) = 1$. In this basis, using $|0\rangle = \frac{1}{\sqrt{2}}(|+\rangle + |-\rangle)$ and $|1\rangle = \frac{1}{\sqrt{2}}(|+\rangle - |-\rangle)$, ρ becomes

$$\rho = 1/2\,|0\rangle\langle 0| + 1/2\,|1\rangle\langle 1| \tag{C.19}$$

$$= 1/2(\frac{1}{\sqrt{2}}(|+\rangle + |-\rangle))(\frac{1}{\sqrt{2}}(\langle +| + \langle -|)) + 1/2(\frac{1}{\sqrt{2}}(|+\rangle - |-\rangle))(\frac{1}{\sqrt{2}}(\langle +| - \langle -|)) \tag{C.20}$$

$$= 1/2\,|+\rangle\langle+| + 1/2\,|-\rangle\langle-|\,. \tag{C.21}$$

And so the probabilities associated to outcomes + and - are

$$p(+) = \mathrm{tr}\left(\,|+\rangle\langle+|\,|+\rangle\langle+|\,(1/2(|+\rangle\langle+|) + 1/2(|-\rangle\langle-|))\right) \tag{C.22}$$

$$= \mathrm{tr}\left(1/2\,|+\rangle\langle+|\,|+\rangle\langle+| + 1/2\,|+\rangle\langle+|\,|-\rangle\langle-|\,\right) \tag{C.23}$$

$$= \mathrm{tr}(1/2\,|+\rangle\langle+|\,|+\rangle\langle+|) \tag{C.24}$$

$$= 1/2, \tag{C.25}$$

$$p(-) = \mathrm{tr}\left(\,|-\rangle\langle-|\,|-\rangle\langle-|\,(1/2\,|+\rangle\langle+| + 1/2\,|-\rangle\langle-|\,)\right) \tag{C.26}$$

$$= \mathrm{tr}\left(1/2\,|-\rangle\langle-|\,|-\rangle\langle-| + 1/2\,|-\rangle\langle-|\,|-\rangle\langle-|\,\right) \tag{C.27}$$

$$= \mathrm{tr}(1/2\,|-\rangle\langle-|\,|-\rangle\langle-|) \tag{C.28}$$

$$= 1/2. \tag{C.29}$$

This is interesting, because this example showcases a fundamental difference between a superposition and a classical mixture. An appropriate choice of basis can make the outcome of measuring a state in a superposition certain, whereas the outcome of a measurement on a mixed state is *always* uncertain, no matter the choice of measurement operator.

D Probabilistic entanglement transformations

D.1 Weak majorization

This section is based on Ref. [7, pp. 10–15]. It is sometimes useful to relax a bit the conditions on the majorization relation, in order to compare vectors that do not have the same sum. Two complementary notions can be defined, *submajorization* and *supermajorization*.

Definition D.1 (Submajorization)**.** Let $p, q \in \mathbb{R}^d$ be two vectors of dimension d. We say that p is submajorized by q, written $p \prec_w q$, if and only if

$$S_k^{\downarrow}(q) \geq S_k^{\downarrow}(p) \quad \forall k = 1, ..., d. \tag{D.1}$$

If the dimensions of the vectors do not match, one can always append zeroes to the one of lower dimension and apply the definition on the enlarged vector.

Definition D.2 (Supermajorization)**.** Let $p, q \in \mathbb{R}^d$ be two vectors of dimension d. We say that p is supermajorized by q, written $p \prec^w q$, if and only if

$$S_k^{\uparrow}(q) \leq S_k^{\uparrow}(p) \quad \forall k = 1, ..., d, \tag{D.2}$$

where $S_k^{\uparrow}(p)$ and $S_k^{\uparrow}(q)$ denote the k^{th} non-decreasing cumulative sums[2] of p and q. If the dimensions of the vectors do not match, one can always append zeroes to the one of lower dimension and apply the definition on the enlarged vector.

Essentially, these definitions are the same as majorization, but they remove the condition that $S_d(p) = S_d(q)$. This allows to compare objects that do not have the same 1-norm.

D.2 Vidal's theorem

This section is based on Refs. [6] and [36]. Until now, only deterministic entanglement transformations have been discussed, i.e. transformations that succeed with probability one. However, a more general notion of entanglement transformations is the class of *probabilistic* entanglement transformations. In the context of LOCC, we will denote the class of LOCC transformations that succeed with probability p by LOCC_p. As such, the class of deterministic LOCC transformations we have discussed until now is simply LOCC_1. Probabilistic transformations are powerful, because they can break the majorization condition of Theorem 2.4, meaning that such transformations can sometimes gain entanglement, even though the operations in themselves are not nonlocal in nature. While this may seem strange, the key lies in the word *sometimes*: such a protocol has probability p of succeeding, but when it fails with probability $1-p$, the resulting state is less entangled, and *on average* the amount of entanglement is (at most) preserved. The following theorem gives the maximal success probability with which a state can be transformed into another.

Theorem D.1 (Vidal's theorem)**.** Let $|\phi\rangle$ and $|\psi\rangle$ be two pure states of a bipartite system AB. Then,

$$|\psi\rangle \xrightarrow{\text{LOCC}_p} |\phi\rangle \iff \lambda_\psi \prec^w p\lambda_\phi, \tag{D.3}$$

where λ_ψ and λ_ϕ are the Schmidt vectors of $|\psi\rangle$ and $|\phi\rangle$, respectively.

This time, a supermajorization relation appears. The maximal success probability can thus be acquired from the theorem by finding the limiting probability p.

Corollary D.1. Probabilistic LOCC transformations cannot increase the Schmidt rank of a state, because if λ_ψ has Schmidt rank k and λ_ϕ has Schmidt rank $k' > k$, then supermajorization relation $k+1$ is only satisfied for $p = 0$.

[2] Notice that using the reverse ordering means that the first components of $p^{\uparrow}$ are the smallest components of p, and so the smallest components of p should be greater than those of q in order to be more uncertain. This is why the inequality in Eq. (D.2) is reversed.

E Proofs of monotonicity of the incomparability functions under bistochastic degradation of the reference

We prove here Theorems 4.4 and 4.5 from Section 4.3.2. Starting with the future incomparability function, if we can show that $E^+(p \parallel Dq) < E^+(p \parallel q)$ for any bistochastic matrix D, then E^+ would be a decreasing monotone, like we expect. We first need a preliminar lemma.

Lemma E.1. For any $p, q \in \mathcal{P}^d$, we have

$$p \vee q \succ p \vee q' \quad \forall\, q' \prec q. \tag{E.1}$$

Proof. Let $p, q \in \mathcal{P}^d$ and let q' be any probability vector majorized by q.

$$p \vee q = p \vee (q \vee q') = p \vee (q' \vee q) = (p \vee q') \vee q \succ p \vee q'. \tag{E.2}$$

□

This lemma is all we need to prove Theorem E.1 concerning the monotonicity of E^+ under bistochastic degradation of q.

Proof of Theorem 4.4. Let us show that the maximum value of $E^+(p \parallel q')$ (with $q' \prec q$) is realized for $q' = q$.

$$\max_{q' \mid q' \prec q} E^+(p \parallel q') = \max_{q' \prec q} d(p, p \vee q') \tag{E.3}$$

$$= \max_{q' \mid q' \prec q} H(p) - H(p \vee q') \tag{E.4}$$

$$\overset{\text{Lemma E.1}}{=} H(p) - H(p \vee q) \tag{E.5}$$

$$= d(p, p \vee q) \tag{E.6}$$

$$= E^+(p \parallel q), \tag{E.7}$$

and so the maximal value of $E^+(p \parallel Dq)$ is reached for the identity degradation. □

Let us turn our attention to E^-. This time around, the expected property for E^- does hold, and we do have monotonicity in q. If we can show that $E^-(p \parallel Dq) > E^+(p \parallel q)$ for any bistochastic matrix D, then E^- would be an increasing monotone, like we expect. To prove it, we first need the following lemma.

Lemma E.2. For any $p, q \in \mathcal{P}^d$, we have

$$p \wedge q \succ p \wedge q' \quad \forall\, q' \prec q. \tag{E.8}$$

Proof. Let $p, q \in \mathcal{P}^d$, and let q' be any probability vector majorized by q. We have

$$p \wedge q' = p \wedge (q \wedge q') = (p \wedge q) \wedge q' \prec p \wedge q. \tag{E.9}$$

□

This lemma is all we need to prove Theorem 4.5 concerning the monotonicity of E^- under bistochastic degradation of q.

Proof of Theorem 4.5. Let q' be a vector majorized by q, i.e. there exists a bistochastic matrix D such that $q' = Dq$.

$$\min_{q' \mid q' \prec q} E^-(p \parallel q') = \min_{q' \mid q' \prec q} d(p, p \wedge q') \tag{E.10}$$

$$= \min_{q' \mid q' \prec q} H(p \wedge q') - H(p) \tag{E.11}$$

$$\overset{\text{Lemma E.2}}{=} H(p \wedge q) - H(p) \tag{E.12}$$

$$= d(p, p \wedge q) \tag{E.13}$$

$$= E^-(p \parallel q), \tag{E.14}$$

and so the minimal value of $E^-(p \parallel Dq)$ is reached for the identity degradation. □

F Notions of measure theory

This section is based on Ref. [37, pp. 8–16]. Measure theory is the mathematical field that studies the size of sets. It has applications in many fields, most notably in probability theory where probability measures are integral to proper definitions, but also in integration theory. A measure is a set function defined on a σ-algebra satisfying special properties. Let us define those notions.

Definition F.1 (Set function)**.** Let Ω be a set, and let $P(\Omega)$ be the power set of Ω, i.e. the set containing all subsets of Ω. Finally, let $\tilde{\Omega} \subseteq P(\Omega)$. Then

$$f : \tilde{\Omega} \to \mathbb{R}, A \mapsto f(A) \tag{F.1}$$

is a set function. Essentially, it is a function that sends subsets of Ω to a number (we will limit ourselves to real numbers here).

A measure, then, is a special set function that satisfies volume-like properties (which we will define soon). Thankfully, a measure does not necessarily need to be defined on absolutely all subsets of the universal set Ω. Instead, we can choose to limit ourselves to a collection of subsets of interest, as long as they form a σ-algebra.

Definition F.2 (σ-algebra)**.** A collection Σ of subsets of a set Ω is said to be a σ-algebra in Ω if Σ has the following properties.

1. Contains the universal set: $\Omega \in \Sigma$.
2. Closed under complements: If $A \in \Sigma$, then $A^c \in \Sigma$, where A^c is the complement of A relative to Ω.
3. Closed under countable unions: If $A = \bigcup_{i=1}^{\infty} A_i$ and if $A_i \in \Sigma \, \forall \, i \in \mathbb{N}$, then $A \in \Sigma$.

In other words, a collection Σ of subsets of Ω is a σ-algebra if it is closed under countable unions and complements[3].

If Σ is a σ-algebra in Ω, then Ω is called a *measurable space* and the elements of Σ are said to be the *measurable sets* of Ω. Moreover, given a collection S of subsets of interest of Ω, a σ-algebra can be generated from S by adding all of the countable unions and complements of the members of S to the generated set, which we will write $\sigma(S)$. Note that if we require the collection of subsets to be closed under finite unions instead of countable unions, the collection is called an algebra over Ω instead of a σ-algebra. A related (simpler) notion is the notion of π-system.

Definition F.3 (π-system)**.** A collection Π of subsets of a set Ω is said to be a π-system in Ω if Π has the following properties.

1. Nonempty: $\Pi \neq \emptyset$.
2. Closed under finite intersections: If $A = \bigcap_{i=1}^{n} A_i$ and if $A_i \in \Pi \, \forall \, i \leq n$, then $A \in \Pi$.

Remark F.1. π-systems are fairly interesting to us, because it is easy to see that the collection of subsets of $\mathcal{P}^d$ $\{\mathcal{T}_+(p) \mid p \in \mathcal{P}^d\}$ is a π-system. This is because by definition of the join, the intersection of two future cones is the future cone of the join (the same discussion is possible with past cones and the meet), and so all possible intersections of future cones is already contained in the set of future cones, and so the collection is closed under finite intersections, satisfying Def. F.3.

[3] And implicitly under countable intersections, as they can be constructed by using $\bigcap_{i=1}^{\infty} A_i = \left(\bigcup_{i=1}^{\infty} A_i^c \right)^c$ (finite intersections can be obtained from this formula by setting $A_i = A_1 \, \forall \, i > n$ for a chosen n).

The final notion of interest we will discuss here is the notion of measure, which we have already mentioned before.

Definition F.4 (Measure)**.** A set function μ, defined on a σ-algebra Σ, is a measure on Σ iff it is countably additive, i.e. if $\{A_i\}$ is a collection of *disjoint* members of Σ, then

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i) \tag{F.2}$$

We say that a *measure space* is a measurable space which has a positive measure defined on the σ-algebra of its measurable sets. Alternatively, one can also define a measure on a semi-ring, but this did not seem as relevant in our context. Measures also enjoy some useful properties.

Theorem F.1. Let μ be a positive measure over Σ. Then, μ has the following properties.

1. Emtpy volume: $\mu(\emptyset) = 0$.
2. Finite additivity: $\mu(A_1 \cup \cdots \cup A_n) = \mu(A_1) + \cdots + \mu(A_n)$ if $\{A_i\}$ is a set of pairwise disjoint elements of Σ.
3. Monotonicity: $A \subseteq B \implies \mu(A) \leq \mu(B)$.

An interesting notion is that measures automatically give a well-defined notion of the amount of elements in a set, i.e. a volume. Finally, the notion that we are perhaps most interested in is that a measure automatically verifies the inclusion-exclusion principle (cf. Theorem 5.1).

G Proofs of the basic properties of the uniqueness entropy

We prove here Props. 1, 2, 3 and 4 from Section 5.1.3. The commutativity Prop. 1 is immediate by commutativity of the join, and so the ordering of the bank does not matter. This is what we would expect for an inclusion-exclusion formula, as the order in which we remove intersections of sets does not matter as long as we perform the full sum. Because the ordering of the bank is arbitrary, we can always make any vector the last vector of the bank, which is useful for proving Props. 3, 4 and 7 as showing the property for $i = k$ is sufficient.

Prop. 2 is immediate as well, because the join of the certain distribution $\overline{\delta}_d$ with any other vector remains the certain distribution. As such, all of the terms in Eq. (5.6) are the entropy of $\overline{\delta}_d$, which is 0. The key insight needed to prove the other properties is given by Lemma 5.1, which we prove here.

Proof of Lemma 5.1. Let $p, q_1, \dots, q_k, q_{k+1} \in \mathcal{P}^d$. Developing all of the terms in Eq. (5.6), and regrouping the combinations containing q_{k+1}, we get

$$\begin{aligned}
E^+(p \parallel q_1, \dots, q_k, q_{k+1}) = H(p) &- \Big[H(p \vee q_1) + \dots + H(p \vee q_k) + H(p \vee q_{k+1})\Big] \\
&+ \Big[H(p \vee q_1 \vee q_2) + \dots + H(p \vee q_{k-1} \vee q_k) + H(p \vee q_1 \vee q_{k+1}) \\
&+ \dots + H(p \vee q_k \vee q_{k+1})\Big] - \dots \\
&+ (-1)^{k+1} H(p \vee \dots \vee q_k \vee q_{k+1}) \qquad \text{(G.1)} \\
= \Big(H(p) &- [H(p \vee q_1) + \dots + H(p \vee q_k)] + [H(p \vee q_1 \vee q_2) \\
&+ \dots + H(p \vee q_{k-1} \vee q_k)] - \dots + (-1)^k H(p \vee \dots \vee q_{k-1} \vee q_k)\Big) \\
&- \Big(H(p \vee q_{k+1}) - [H(p \vee q_1 \vee q_{k+1}) + \dots + H(p \vee q_k \vee q_{k+1})] \\
&+ [H(p \vee q_1 \vee q_2 \vee q_{k+1}) + \dots + H(p \vee q_{k-1} \vee q_k \vee q_{k+1})] \\
&- \dots + (-1)^k H(p \vee \dots \vee q_k \vee q_{k+1})\Big). \qquad \text{(G.2)}
\end{aligned}$$

By commutativity of the join, we can push q_{k+1} next to p, and so Eq. (G.2) becomes

$$\begin{aligned}
E^+(p \parallel q_1, \dots, q_k, q_{k+1}) = \Big(H(p) &- [H(p \vee q_1) + \dots + H(p \vee q_k)] + [H(p \vee q_1 \vee q_2) \\
&+ \dots + H(p \vee q_{k-1} \vee q_k)] - \dots + (-1)^k H(p \vee \dots \vee q_{k-1} \vee q_k)\Big) \\
&- \Big(H(p \vee q_{k+1}) - [H(p \vee q_{k+1} \vee q_1) + \dots + H(p \vee q_{k+1} \vee q_k)] \\
&+ [H(p \vee q_{k+1} \vee q_1 \vee q_2) + \dots + H(p \vee q_{k+1} \vee q_{k-1} \vee q_k)] \\
&- \dots + (-1)^k H(p \vee q_{k+1} \vee \dots \vee q_k)\Big). \qquad \text{(G.3)}
\end{aligned}$$

Comparing terms one by one, we can see that the second half of the expression is exactly the same as the first half (which is equal to $E^+(p \parallel q_1, \dots, q_k)$) if p is replaced by $p \vee q_{k+1}$, and so Eq. (G.3) becomes

$$E^+(p \parallel q_1, \dots, q_k, q_{k+1}) = E^+(p \parallel q_1, \dots, q_k) - E^+(p \vee q_{k+1} \parallel q_1, \dots, q_k). \qquad \text{(G.4)}$$

□

Such an expression lends itself very well to induction proofs. Let us now prove the remaining properties. For Prop. 3, we have the following.

Lemma G.1 (Absorption in p of the uniqueness entropy)**.** Let $p, q_1, \dots, q_k \in \mathcal{P}^d$. If $\exists i \leq k \,|\, p \succ q_i$, then

$$E^+(p \parallel q_1, \dots, q_k) = 0. \qquad \text{(G.5)}$$

Proof. The proof is immediate from Lemma 5.1. Because the ordering of the bank is arbitrary, let us first choose $i = k$, and so $p \succ q_k$. We have

$$E^+(p \parallel q_1, \dots, q_k) = E^+(p \parallel q_1, \dots, q_{k-1}) - E^+(\underbrace{p \vee q_k}_{=p} \parallel q_1, \dots, q_{k-1}) \tag{G.6}$$

$$= 0. \tag{G.7}$$

The same result can be reached by comparing term by term the possible combinations of states in the bank[4]. If $i \neq k$, commuting q_i until it becomes the last vector of the bank and applying Lemma 5.1 yields the same result. □

This property is again quite intuitive, because if p majorizes one of the vectors of the bank q_i, then $\mathcal{T}_+(p) \subseteq \mathcal{T}_+(q_i)$, and so p holds no unique volume relative to the bank. The geometric interpretation of Prop. 4 is very similar, and we have the following.

Lemma G.2 (Absorption in q of the uniqueness entropy)**.** Let $p, q_1, \dots, q_k, q_{k+1} \in \mathcal{P}^d$. If $\exists i \leq k \mid q_{k+1} \succ q_i$, then

$$E^+(p \parallel q_1, \dots, q_k, q_{k+1}) = E^+(p \parallel q_1, \dots, q_k). \tag{G.8}$$

Proof. Let $p, q_1, \dots, q_k, q_{k+1} \in \mathcal{P}^d$. Because the ordering of the bank is arbitrary, let us first choose $i = k$, and so $q_{k+1} \succ q_k$. Using Lemma 5.1 twice in a row, we get

$$E^+(p \parallel q_1, \dots, q_k, q_{k+1}) = E^+(p \parallel q_1, \dots, q_k) - E^+(p \vee q_{k+1} \parallel q_1, \dots, q_k) \tag{G.9}$$

$$= E^+(p \parallel q_1, \dots, q_k) - E^+(p \vee q_{k+1} \parallel q_1, \dots, q_{k-1}) + E^+(p \vee \underbrace{q_{k+1} \vee q_k}_{=q_{k+1}} \parallel q_1, \dots, q_{k-1}) \tag{G.10}$$

$$= E^+(p \parallel q_1, \dots, q_k). \tag{G.11}$$

The same result can be reached by comparing term by term the possible combinations of states in the bank[5]. If $i \neq k$, commuting q_i until it becomes the k^{th} vector of the bank and applying Lemma 5.1 twice yields the same result. □

[4]Exactly half of the terms in the sum contain q_k. Because each of those can be paired up with the same combination excluding q_k, we have terms of same absolute value but opposite sign, and so all of the terms cancel out.

[5]Exactly half of the combinations containing q_{k+1} also contain q_k. Because each of those can be paired up with the same combination excluding q_k, we have terms of same absolute value but opposite sign, and so all of the terms containing q_{k+1} cancel out.

H Simulation for alternative filterings for the mixed strategy

For the sake of brevity, we will also define special subsets of the bank. If $\tilde{Q}$ is a subset of the bank, then $\tilde{Q}_i$ is the subset of states $\{q \in \tilde{Q} \,|\, q \not\prec q_i\} \subseteq \tilde{Q}$. This implies that $\tilde{Q}_i$ does not contain q_i either, since $q_i \prec q_i$ is always true.

As mentioned in Section 5.3, the performance of the mixed strategy greatly varies if we choose other state filterings in step 3 of Def. 5.4. We propose the following variation.

Definition H.1. Choose $\alpha \in [0, 1]$. Let Q be a bank of pre-shared entangled states, and let t be a target state. The following algorithm decides which state $q \in Q$ to use to construct t.

1. Initialize the set $Q' = \emptyset$. For each $q_i \in Q$, if $q_i \prec t$, add q_i to Q', which contains all the states that can reach the target.
2. Initialize the set $Q'' = \emptyset$. For each $q_i \in Q'$, if $\nexists q_j \in Q' \,|\, q_i \prec q_j, q_i \neq q_j$ and if $q_i \notin Q''$, add q_i to Q'', which contains only the least resourceful states that can reach the target.
3. For each $q_i \in Q''$, compute $a_i = \alpha H(q_i) + (1 - \alpha) E^+(q_i \parallel Q'_i)$, and initialize $b_i = 1$; for each $q_j \in Q' \backslash q_i$, if $q_j \prec q_i$, increment b_i by 1.
4. Finally, construct t using the state $q_i \in Q''$ with the lowest value of the ratio $c_i = \frac{a_i}{b_i}$.

Note that in step 3, the uniqueness entropy is computed relative to Q'_i relative to Q''. This behavior is actually closer to what we originally envisioned for the uniqueness strategy, however Figure H.1 shows that it actually performs strictly worse than the entropic strategy (with the additional redundancy factor), and the curve showcases no interesting maximum.

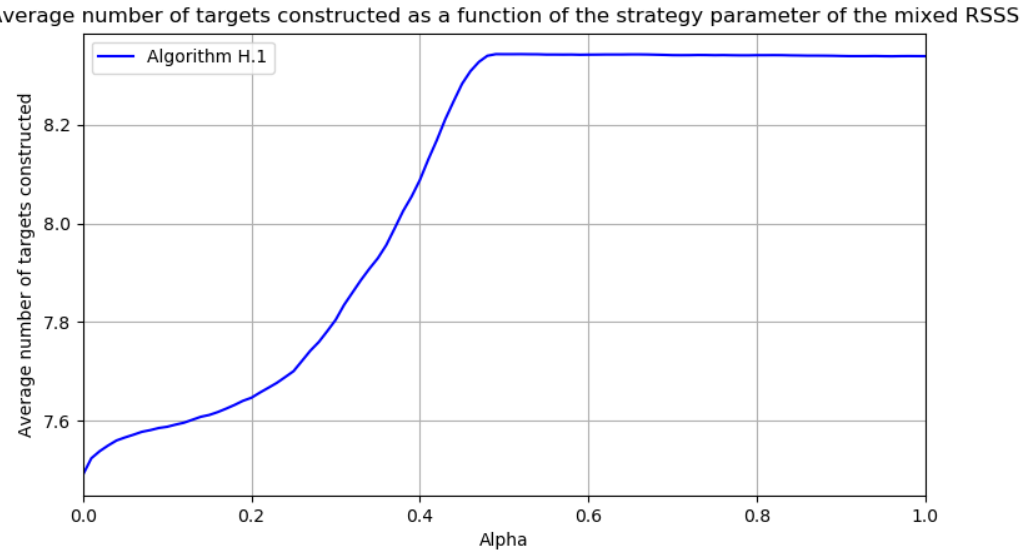


Figure H.1: Average number of succesful target constructions depending on the value of the strategy parameter α for a bank of 9 uniformly sampled states and 1 maximally entangled state, and 10 targets sampled with Dirichlet parameters $\left(2, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$, sampled over 10000 games. Clearly, the uniqueness entropy is only detrimental for this variant. It has little impact until $\alpha \approx 0.45$, and then the sharp drop seems to indicate that uniqueness entropy becomes significant enough in the loss function under this filtering to kick in and change the selection of states. This can be explained by the fact that more states make it into Q'_i than into Q'' (from the original algorithm), and so the value of the uniqueness entropy is generally lower this time.

In Section 5.3 we showed that improvements over the naive individual strategy are possible using the uniqueness entropy, and perhaps more generally with other lattice-based quantities.

However, considering how much of a difference there is between Figures 5.6 and H.1 for such a small change in the filtering process, we hope that other filtering choices could lead to even better improvements over the naive entropic strategy. We were recently made aware of the relevance of the field of dynamic optimization in this context, and hope that it might hold answers to this problem.

I Full proof of supermodularity of sum-concave functions.

This section provides the full proof of supermodularity for all sum-concave functions, showing that Conjecture 3.1 is true. This proof was not included in the original master thesis and was only added in this second version, hence its place in the appendix.

Theorem .1 (Supermodularity of all sum-concave functions)**.** *All sum-concave functions F are supermodular on the majorization lattice. Formally, for any $p, q \in \mathcal{P}^d$, we have*

$$F(p \wedge q) + F(p \vee q) \geq F(p) + F(q). \tag{I.1}$$

Note that if $p \sim q$, there is trivially equality.

Proof. The proof is in two steps. Let us define[6]

$$m = p \wedge q,\ j = p \vee q, j' = \beta(p, q), \tag{I.2}$$
$$A = \text{concat}(p, q) \in \mathbb{R}^{2d}, \tag{I.3}$$
$$B = \text{concat}(m, j) \in \mathbb{R}^{2d}, \tag{I.4}$$
$$B' = \text{concat}(m, j') \in \mathbb{R}^{2d}, \tag{I.5}$$

where $\text{concat}(\cdot, \cdot)$ is the ordered concatenation of two vectors. The first step of the proof is to show that $A \succ B'$. The second is to show that since $\beta(p, q)^{\downarrow} \succ p \vee q$, then $B' \succ B$ too. This implies that $A \succ B$ by transitivity, on which we can then use a Karamata inequality to retrieve Eq. (3.4). We have $\sum_{i=1}^{2d} A_i = \sum_{i=1}^{2d} B_i = \sum_{i=1}^{2d} B'_i = 2$. We will now show that $A \succ B'$, by showing that all of the terms in B' can be obtained by successive Robin Hood transfers of A. Clearly, the k^{th} cumulative sum of A is made up of the largest entries of p and q, and the same goes for cumulative sums of B' being made of the largest entries of m and j'. To avoid ill-defined terms, we additionally introduce the convention that $S_k^{\downarrow}(v) = 1\ \forall k > d$ for any d-dimensional probability vector v. Let us assume that for a given cumulative sum index $i \leq d - 1$, we have $S_i^{\downarrow}(p) \geq S_i^{\downarrow}(q)$ (the other case being perfectly symmetric). We have 2 possible cases for the index $i + 1$.

1. $\underline{S_{i+1}^{\downarrow}(p) \geq S_{i+1}^{\downarrow}(q)}$: in this case, the expressions for m_{i+1} and j'_{i+1} are simple. We have

$$m_{i+1} = S_{i+1}^{\downarrow}(q) - S_i^{\downarrow}(q) = q_{i+1} \tag{I.6}$$
$$j'_{i+1} = S_{i+1}^{\downarrow}(p) - S_i^{\downarrow}(p) = p_{i+1}, \tag{I.7}$$

   and so for the index $i + 1$, m and j' are simply components of p and q. Therefore, for all of the indices falling under this case, the components of B' are simply the same as those of A.

2. $\underline{S_{i+1}^{\downarrow}(p) < S_{i+1}^{\downarrow}(q)}$: we have

$$m_{i+1} = S_{i+1}^{\downarrow}(p) - S_i^{\downarrow}(q) = p_{i+1} + S_i^{\downarrow}(p) - S_i^{\downarrow}(q) \tag{I.8}$$
$$j'_{i+1} = S_{i+1}^{\downarrow}(q) - S_i^{\downarrow}(p) = q_{i+1} + S_i^{\downarrow}(q) - S_i^{\downarrow}(p). \tag{I.9}$$

   Let us define $\Delta = S_i^{\downarrow}(p) - S_i^{\downarrow}(q)$, which is greater or equal to 0 by hypothesis. Eqs. (I.8) and (I.9) can thus be rewritten as

$$m_{i+1} = S_{i+1}^{\downarrow}(p) - S_i^{\downarrow}(q) = p_{i+1} + \Delta \tag{I.10}$$

[6] Credit must be given to OpenAI's o4-mini LLM for the idea of constructing such concatenated vectors. After digging, it turns out that similar techniques are used in Ref. [7, pp. 133–136], where non-trivial vector constructions are also used to prove convexity results.

$$j'_{i+1} = S^{\downarrow}_{i+1}(q) - S^{\downarrow}_{i}(p) = q_{i+1} - \Delta. \tag{I.11}$$

Moreover, we know that $p_{i+1} < q_{i+1}$, but also that $p_{i+1} + \Delta < q_{i+1}$ and $q_{i+1} - \Delta > p_{i+1}$, because

$$\begin{aligned} S^{\downarrow}_{i+1}(q) - S^{\downarrow}_{i+1}(p) &> 0 && \text{(I.12)}\\ &= q_{i+1} - p_{i+1} + S^{\downarrow}_{i}(q) - S^{\downarrow}_{i}(p) && \text{(I.13)}\\ \iff q_{i+1} - p_{i+1} &> \underbrace{S^{\downarrow}_{i}(p) - S^{\downarrow}_{i}(q)}_{=\Delta} \geq 0. && \text{(I.14)} \end{aligned}$$

Therefore, since all components of B' can be constructed from cases 1 and 2, and that both cases can be expressed as Robin Hood transfers from A to B', we have that $A \succ B'$.

Now, in order to finalize the proof, we need to show that $B' \succ B$. We already know from Ref. [10] that $\beta(p,q)^{\downarrow} \succ p \vee q$, so what we now need to show is essentially that $j' \succ j \implies \text{concat}(j', m) \succ \text{concat}(j, m)$. We have

$$S^{\downarrow}_{k}(B) = \max_{0 \leq l \leq k} \left(S^{\downarrow}_{l}(j) + S^{\downarrow}_{k-l}(m) \right), \tag{I.15}$$

$$S^{\downarrow}_{k}(B') = \max_{0 \leq l' \leq k} \left(S^{\downarrow}_{l'}(j') + S^{\downarrow}_{k-l'}(m) \right). \tag{I.16}$$

Let us call the value of l that realizes the maximum of Eq. (I.15) L, and the value of l' that realizes the maximum of Eq. (I.16) L'. Since $j' \succ j$, we know that $S^{\downarrow}_{i}(j') \geq S^{\downarrow}_{i}(j)$ is true for all $i \leq d$. In order to show majorization, we need to show that for all $k \leq 2d$, $S^{\downarrow}_{k}(B) \leq S^{\downarrow}_{k}(B')$ is true. For any $k \leq 2d$, there are 2 possible cases.

1. $\underline{L' = L}$: we have

$$S^{\downarrow}_{k}(B) = S^{\downarrow}_{L}(j) + S^{\downarrow}_{k-L}(m) \leq S^{\downarrow}_{L}(j') + S^{\downarrow}_{k-L}(m) = S^{\downarrow}_{k}(B'). \tag{I.17}$$

2. $\underline{L' > L}$: we have

$$\begin{aligned} S^{\downarrow}_{k}(B') &= \max_{0 \leq l' \leq k} \left(S^{\downarrow}_{l'}(j') + S^{\downarrow}_{k-l'}(m) \right) && \text{(I.18)}\\ &\geq S^{\downarrow}_{L}(j') + S^{\downarrow}_{k-L}(m) && \text{(I.19)}\\ &\geq S^{\downarrow}_{L}(j) + S^{\downarrow}_{k-L}(m) && \text{(I.20)}\\ &= S^{\downarrow}_{k}(B). && \text{(I.21)} \end{aligned}$$

And so in both cases, $S^{\downarrow}_{k}(B) \leq S^{\downarrow}_{k}(B')$ is true, and so majorization is verified. Therefore, $B' \succ B$, and so $A \succ B$ as well. Using Lemma 1.1 on φ ($-\varphi$ being convex on the same interval), we get

$$\sum_{i=1}^{2d} \varphi(A_i) \leq \sum_{i=1}^{2d} \varphi(B_i). \tag{I.22}$$

By the sum nature of F, the LHS of Eq. (I.22) is precisely $F(p) + F(q)$, and the RHS is precisely $F(p \wedge q) + F(p \vee q)$, so we have proven Eq. (I.1). □