Quantum Shannon Theory and Beyond.

A course by JProf. Dr Angela Capel-Cuevas,

with the script, in parts, written by Paul Gondolf.*

Institute of Mathematics
University of Tübingen

May 31, 2022

*In case you find mistakes or typos, let me know: paul.gondolf@web.de

Contents

Chapter 1

Introduction and Fundamental Concepts

1.1 Scope of the course and Bibliography

In this course, we will study the transmission of information over a noisy quantum communication channel. In particular, students will learn about quantum mechanics, entanglement, teleportation, tomography, quantum estimation, hypothesis testing, and various capacity theorems involving classical bits, qubits, and entangled bits. There will be a strong focus on entropy measures and their application to numerous quantum tasks.

This course is intended for students of the Master in Mathematical Physics of the University of Tübingen, but is open to anyone with some basic knowledge on mathematical analysis, linear algebra, probability theory, as well as some interest on learning about the exciting world of quantum information theory.

1.1.1 Bibliography

There is a plethora of references in the literature that concern the topics that we will discuss in this course, such as Quantum Information Theory, Quantum Entropies and Applications, Shannon Theory, Quantum Channels, etc. Below we include a short list with some of the main texts which are frequently used in this community in courses of similar scope. Furthermore, we list as a reference the current Lecture Notes, since they will contain a summary of the contents presented in the lectures, as well as some complementary material/references in some parts. By no means these Lecture Notes intend to replace any of the other texts, written by some of the main authors in the quantum information community, and the students are encouraged to consult these texts to complement their knowledge on the subject.

The short selection of manuscripts has been done accordingly to the contents intended for this course. The lectures, and therefore these notes, have been prepared consulting these texts, as well as some others, and they are properly referenced in this respect. More specifically, most of the contents of Chapter 1 have been extracted from the well-known books of Nielsen and Chuang, as well as Wilde; Chapter 3 is inspired in the Lecture Notes by Wolf in quantum channels; Chapters 4 and **??** use as base materials the texts of Wilde and Carlen; and Chapter **??** is inspired in various parts of all the texts mentioned below. Some other books/notes used for the construction of these notes will be properly referenced in the main text.

1. Lectures Notes

2. Nielsen-Chuang, "Quantum Computation and Quantum Information" [11]

3. Wilde, "From Classical to Quantum Shannon Theory" [23]

4. Watrous, "The Theory of Quantum Information" [22]

5. Carlen, "Trace Inequalities and Quantum Entropy" [5]

6. Wolf, "Quantum Channels and Operations. Guided Tour" [24]

1.2 What is Quantum Information?

The scientific field of Quantum Information has a lot of different facets and encompasses the field of mathematics, physics and computer science. Its main questions concern the control of quantum systems. I.e. *can we construct and manipulate complex quantum systems? And if so, what are the scientific and technological applications?* It is important to remark that the field of Quantum Information Science does not study the frontier of short (subnuclear) distances or long (cosmological) distances, but rather the frontier of highly complex quantum systems, what is usually known as *the entanglement frontier*.

Compared to the classical world that we know, as it is the world we experience every day, the quantum world exhibit behaviours that are counter intuitive to our classical understanding of the world. These additional properties, which will be discussed in detail throughout the course, provide quantum systems with, in a sense, more complex and richer behaviour, meaning that we can expect to simulate a classical system using a quantum system, and it is generally believed that this is not possible the other way around (although it still remains an unproven conjecture). However, quantum systems present the phenomenon of *decoherence*, as opposed to classical systems, and this effect tends to destroy information very fast and, thus, quantum systems end up losing their special properties after some time and behaving like classical ones. It is a major problem to determine how hard solving the problem arising from decoherence is and whether we will be able to overcome it with the current techniques of science. Nevertheless, the special properties of quantum systems torched a large research effort whose main goal is to control the quantum behaviour of scalable quantum systems and achieve the "quantum advantage", which will allow us to prepare and control complex quantum systems that behave in ways that cannot be predicted using digital computers. For that, we will need to find what quantum tasks are feasible and which quantum problems are hard to simulate classically.

1.2.0.1 Shor's factoring algorithm

To give an example of the theoretically expected improvement in behaviour of a quantum computer with respect to a classical one, let us present some basic calculations on the required to factorize a certain number with both devices. For that, we make use of one of the first breakthroughs in the first steps of the field of Quantum Computing in the past century, namely the algorithm devised by Shor to factor numbers in their prime components. We are not going to discuss such an algorithm in detail in teh current text, but we strongly recommend the avid reader to consult the original article [16] as well as the references [25, 20].

Assume that we want to factor a number into its two prime factors $n = p_1 \cdot p_2$. Some theoretical computations show that we have the following comparison of computational time using Shor's algorithm on a quantum computer and a classical algorithm on a classical computer:

Numbers	Classical computer	Quantum computer
193 digits	30 CPU years	0.1 seconds
500 digits	10^{12} CPU years	2 seconds

Moreover, as a hint of the meaning of the previous table in an impactful case, the energy consumption to crack RSA-encryption would demand 10^6 terawatt hours for the classical and 10 megawatt hours for a quantum computer.[1]

1.2.1 Emergence of Quantum Information Science

There were several coetaneous facts that could be considered as the seed for the creation of the new field of Quantum Information Science. Some of the most remarkable facts which gave rise to this field are:

- A genuine concern regarding the true value of Moore's law in the coming years. This concern was based on the physical limit of computer chips, i.e. the space per bit cannot be shrunk indefinitely but is limited by the physical properties of the chip material (diameter of atoms, etc.).
- At a similar time, it was the first moment in history in which researchers in labs managed to control "single quantum systems", isolating them from systems with many quantum systems.
- Moreover, there was also an increase in the recognition of the computational power generated by quantum mechanics, which might allow for the design of computational devices based on the laws of such a field.
- Finally, another motivating aspect was the relevance of certain implications of quantum mechanics in practical aspects for society, such as to the security of public key cryptography.

1.2.2 Quantum Information vs. Classical Information

To conclude this short introduction to Quantum Information Theory, let us briefly mention the main differences with respect to the realm of Classical Information Theory. The three key properties of a quantum system compared to a classical system are the following ones:

- **(True) randomness.** Note that, even though we sometimes discuss some processes in classical mechanics as random ones, they are frequently just "pseudo-random", in the sense that their outcome might be predetermined, even if we do not know it in advance (and that is why we take it to be random). However, clicks in a Geiger counter, for instance, are intrinsically random, not pseudo-random, as, at every instant of time, there is always a certain probability of having more clicks in the next second or not having them, but the outcome is not predetermined in any way.
- **Uncertainty.** If we consider two operators A and B which do not commute, this means that measuring A influences the outcome of a subsequent measurement of B and vice versa.
- **Entanglement.** This property can be summarized as "the whole is more definite than the parts". This means that even knowing a joint system AB (pure), the (mixed) state of A may be highly uncertain.

All these terms will be further defined and formalized as we proceed with the course.

[1] This estimate stems from about 10 years ago.

1.3 Classical Information Theory: A Brief Overview

1.3.0.1 Starting point

In the following we are trying to understand Shannon's approach to classical information. He started with the idea formalised by the theorem of Bayes (1763). This theorem phrased the idea that probabilities depend on what one knows, meaning that acquiring additional information modifies the probabilities. Almost 200 years later, Claude Shannon (1916 - 2001) took this as basis and framed the term *information*. He did this in a series of papers and works from the year 1948 onward [14, 15], starting with one of the most influential papers ever written, called "The mathematical approach to communication". In this paper he for example introduced the *Shannon Entropy* over a probability distribution, which nowadays appears various different contexts

$$S(p) = -\sum_{i \in I} p_i \log(p_i). \tag{1.3.1}$$

It is noteworthy that this quantity actually coincides with Boltzmann's formula of entropy.

Note that Bayes constructed his theory based on the idea that probabilities are not absolute, but rather depend on the available information, whereas Shannon framed the concept of information as a function that is precisely defined using a probability distribution. Hence, any set of probabilities can be associated with a quantity of information. Reversing this we get that every probabilistic phenomenon has an associated information theory and since quantum theory is a probabilistic theory the existence of a quantum information theoretic field is only natural.

1.3.0.2 Probabilities and conditional probabilities

Shannon frames the notion of "content of information" using probability theory and the concept of a probabilistic ensemble. This limits the use of such a notion to systems that can be described by random variables. Let us collect below some basic notions and properties concerning random variables, which will be of use in the next pages of these notes.

Definition. 1.3.1 (Ensemble)
An ensemble is given by a tuple $X = (x, A_X, P_X)$ with

- x: the value of the random variable,
- A_X: the set of possible values that the random variable can take (sample space),
- P_X: the probability distribution.

The probability distribution is a function from $\mathcal{F}$ a σ-algebra over A_X, called event space, to the reals, i.e.

$$P_X : \mathcal{F} \to \mathbb{R} \tag{1.3.2}$$

satisfying the following properties:

1. $\forall A \in \mathcal{F}$: $1 \geq P(A) \geq 0$.
2. For a disjoint family of sets $\{A_i\}_i$ it holds that
$$P(\bigcup_i A_i) = \sum_i P(A_i). \tag{1.3.3}$$
3. $P(A_X) = 1$

In the case that A_X is finite we usually have that $\mathcal{F} = \mathcal{P}(A_X)$, the power set. We then also leave $\mathcal{F}$ implicit and just assume it to be the power set of A_X. Abusing notation we then often write for $a \in A_x$, $P_X(a)$ or $P(X = a)$ and mean $P_X(\{a\})$, or if the elements of A_X are enumerated we might write $p_i = p(a_i) = P_X(a_i) = P_X(\{a_i\})$.

Example. Modelling the tossing of a coin with the outcomes heads (h) and tails (t), results in $A_X = \{h, t\}$, $\mathcal{F} = \{\{h\}, \{t\}\}$ and $P(\{h\}) = 1/2 = P(\{t\})$.

Definition. 1.3.2 (Joint probability distribution)
These distributions describe the joint outcome of two events. We now have that $A_X = A_A \times A_B$ is a set of tuples. Apart from that the situation is completely analogous. We denote (again in the situation that A_A and A_B are finite and enumerated)

$$\begin{aligned} P_X(\{a_i, b_j\}) &= P(X = (A, B) = (a_i, b_j)) \\ &= P(A = a_i, B = b_j) \\ &= P(a_i, b_j). \end{aligned} \tag{1.3.4}$$

We have that the joint distribution fulfills the following properties:

- In the case that they are independent, we have that $P(a_i, b_j) = P(a_i)P(b_j)$ (this is actually the defining property of independence).
- It holds that the marginal distributions can be recovered from the joint distribution by

$$P(a_i) = \sum_j P(a_i, b_j) \qquad P(b_j) = \sum_i P(a_i, b_j) \tag{1.3.5}$$

- We now frame the concept of conditional probability using *Bayes rule*. We namely have

$$P(a_i, b_j) = P(b_j|a_i)P(a_i) = P(a_i|b_j)P(b_j). \tag{1.3.6}$$

- From the last point we immediately get the *Bayes theorem*:

$$P(a_i|b_j) = \frac{P(b_j|a_i)P(a_i)}{P(b_j)}. \tag{1.3.7}$$

1.3.0.3 Entropy and information

The link between entropy and information is rather subtle and beautiful. In their book from 1949 [15], Shannon and Weaver gave a description of this notion using axioms that should be satisfied by it. As reproducing this description would require some effort that goes beyond the scope of this course, we will instead roughly sketch the ideas behind their formalization in the following: Let A be a single event with the possible outcomes $\{a_i\}_i$

1. Intuitively, if the outcome of the event is almost certain, e.g. $A = a_0$ there is no information present when it happens. If, on the other hand, $A = a_0$ is unlikely, then the information content of this event happening is very high. Hence, our entropy function h should have the property that $h(P(a_i))$ increases as $P(a_i)$ decreases.
2. In the case that the $P_X = P_{(A,B)}$ is a joint distribution, with, however, the events A and B unrelated, i.e. $P(a_i, b_j) = P(a_i)P(b_j)$, the information provided by each of the events should add up, namely the following should hold:

$$h[P(a_i, b_j)] = h[P(a_i)P(b_j)] = h[P(a_i)] + h[P(b_j)] \tag{1.3.8}$$

3. The function should be positive.

The natural candidate for a function representing the information that satisfies the previous conditions for a specific outcome is the minus logarithm, i.e.

$$h(P(a_i)) := -K \log P(a_i)\,,$$

where K is a positive constant to be determined later. However, note that we want to define "information" as the weighted average of the previous h for every possible outcome, namely:

$$\begin{aligned} H &:= -K \sum_i P(a_i) \log P(a_i) \\ &= -\sum_i P(a_i) \log_2 P(a_i). \end{aligned} \tag{1.3.9}$$

in the base two logarithm. This is the formula Shannon gave and we call the unit of this entropy *bits*. If instead we consider the natural logarithm, i.e.

$$H_e := -\sum_i P(a_i) \log P(a_i) \tag{1.3.10}$$

the unit is called nats. The transformation formula is just $H_e = \log(2)H$, giving us that 1 nat = $\log(2)$ bits.

Example. One example would be the binary entropy (see Figure 1.1). In this case $A_X = \{0, 1\}$ and $P_X = \{p, 1-p\}$. The entropy then turns out to be

$$H_{(2)} = p \log_2 \frac{1}{p} + (1-p) \log_2 \frac{1}{1-p}. \tag{1.3.11}$$

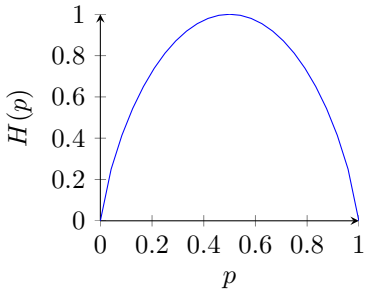


fig. 1.1: The binary entropy.

The binary entropy obviously attains its maximum if $p = 1-p$, i.e. the probability distribution is uniform.

It is noteworthy that the discovery we made in the example with the binary entropy, actually also applies to every other random variable. This means that the maximum of the entropy is reached if the probability distribution is uniform.

Our goal is now to discuss the more general case in which we have joint variables, i.e. a system composed of two subsystems with two variables (x, A_X, P_X) and (y, A_Y, P_Y) respectively. We then define

Definition. 1.3.3 (Joint entropy, Mutual information, Conditional entropy)
Let two random variables X, Y be given and A_X and A_Y finite and enumerated and P the joint distribution of X and Y. Then

- The *joint entropy* is given by

$$H(X,Y) := -\sum_{i,j} P(x_i, y_j) \log P(x_i, y_j); \tag{1.3.12}$$

- The *marginal entropies* are given by

$$\begin{aligned} H(X) &= -\sum_{i,j} P(x_i, y_j) \log \sum_k P(x_i, y_k), \\ H(Y) &= -\sum_{i,j} P(x_i, y_j) \log \sum_k P(x_k, y_j); \end{aligned} \tag{1.3.13}$$

- The *mutual information* is defined as

$$I(X:Y) = H(X) + H(Y) - H(X,Y). \tag{1.3.14}$$

  It measures how much two random variables are codependent and has the following properties

  - $I(X:Y) \geq 0$
  - $I(X:Y) = 0 \Leftrightarrow P(x_i, x_j) = P(x_i)P(x_j)\ \forall i, j.$

  Further it provides a bound on the rate at which we can communicate.

1.3.0.4 Shannon entropy and Boltzmann entropy

We have already stated that the Shannon entropy coincides with Boltzmann's concept of entropy and want to elaborate on that further.

Note that we obtain the Boltzmann distribution by maximizing the information subject to a constraint on the average energy. Considering a system which is in thermal equilibrium with its environment we, in the classical case, have an continuous energy spectrum, while in the discrete case the energy levels are discrete. As we are more interested in the quantum case, we limit our calculations to that case.

What we now want to do is to determine the probability p_i that the system actually has energy E_i and are doing so by maximizing information subject to the mean energy being fixed. Using Lagrange's method:

$$\max H_e = \max\Big(-\sum_i p_i \log(p_i)\Big) \quad \text{subject to} \quad \sum_i p_i E_i = \tilde{E}_i, \quad \sum_i p_i = 1 \tag{1.3.15}$$

This gives us

$$\tilde{H} = H_e + \lambda\Big(1 - \sum_i p_i\Big) + \beta\Big(\tilde{E} - \sum_i p_i E_i\Big) \tag{1.3.16}$$

and

$$d\tilde{H} = \sum_i (-\log(p_i) - 1 - \lambda - \beta E_i) dp_i. \tag{1.3.17}$$

After some simple manipulations we obtain

$$p_i = e^{-(1+\lambda)} e^{-\beta E_i} \tag{1.3.18}$$

and imposing the constraints:

$$p_i = \frac{\exp\left(-E_i/(k_B T)\right)}{\sum_j \exp\left(-E_j/(k_B T)\right)}. \tag{1.3.19}$$

1.3.0.5 Communication theory

As mentioned earlier in this section, Shannon introduced his information theory as the Mathematical Theory of Communication. Once we have set some preliminary concepts and properties in the field of Classical Information Theory, we are ready to describe how classical communication takes place. For that, we can sketch a scheme with the different elements of a classical communication channel, as done in figure 1.2:

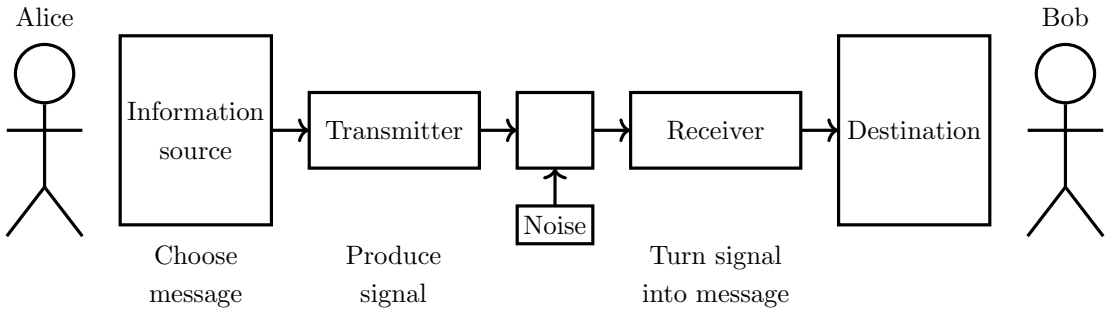


fig. 1.2: A scheme of classical communication through a noisy channel.

This simple scheme raises two fundamental questions, which will be addressed in the next pages, namely:

1. How much can a message be "compressed"?
2. Which is the best rate of reliable communication through a noisy channel?

The answers to those questions were given by Shannon through proving two very important theorems. First the "Noiseless Shannon theorem" and second the "Noisy Shannon theorem". In the following we will visit both of them, however, only prove the first one. For the second one we give some intuition but leave the proof to the several sources that exist on this topic. In any case, in the quantum part we will present a complete proof of the analogous result.

Before we dive into the math, however, we want to make the first question about the necessity for compressing a message more illustrative on an example.

Example. Let $\{a, b, c, d\}$ be an alphabet that we want to use to send a message which we will subsequently try to compress. The appearance probabilities of every letter of the alphabet in a message are:

$$P(\{a\}) = 1/2,\ P(\{b\}) = 1/8,\ P(\{c\}) = 1/4,\ P(\{d\}) = 1/8. \tag{1.3.20}$$

Our goal is now to encode this input using bits, as only bits can be transmitted in our communication channel.

- Attempt 1. Assume we encode as follows

$$\begin{aligned} a &\to 00, \quad b \to 01, \\ c &\to 10, \quad d \to 11. \end{aligned} \tag{1.3.21}$$

  Then the expected length of a message is 2N bits. We have, however, not taken the probabilities into account!

- Attempt 2. Since we expect a to appear more frequently than any other letter, we can associate to it a "shorter" code in terms of number of bits. Assume now that we encode

$$\begin{aligned} a &\to 0, \quad b \to 110, \\ c &\to 10, \quad d \to 111, \end{aligned} \tag{1.3.22}$$

then the expected length of a message is

$$N(\frac{1}{2}\cdot 1+\frac{1}{8}\cdot 3+\frac{1}{4}\cdot 2+\frac{1}{8}\cdot 3)=\frac{7}{4}N \tag{1.3.23}$$

hence only 1.75 bits.

One could wonder whether there is another way to further compress the message so that its expected length is even smaller. However, we will see below that this is not the case, as Shannon's noiseless theorem states that the optimal rate of compression is given by the Shannon entropy of the distribution, and in the example above its value is exactly 1.75.

1.3.0.6 Source coding theorems, informal version

In the context of the theorem, let $(X_1, X_2, \ldots, X_n) = X^n$ be a sequence of n binary i.i.d. random variables. We denote a single outcome by

$$(x_1, \ldots, x_n) = x^n, \tag{1.3.24}$$

and the set of all possible outcomes by

$$A_{X^n} = \{(x_1, \ldots, x_n) \ : \ x_i \in \{0,1\}\ \forall i = 1, \ldots n\}. \tag{1.3.25}$$

Then, as we will show in the next subsections, the source coding theorem proves that there exists a subset $S_{X^n} \subseteq A_{X^n}$, called a "typical set", such that almost all information in A_{X^n} is contained in S_{X^n}. The typical set has a total of 2^{nH} elements. Informally, this theorem is stated as follows.

Theorem. 1.3.4 (Source coding, informal version) *Consider n binary i.i.d. random variables, each with entropy H. Then, they can be compressed into more than 2^{nH} bits with negligible risk of losing information.*

Conversely, if they are compressed into fewer than nH bit, virtually certain some information will be lost.

1.3.1 Shannon's noiseless coding theorem

Let a sequence of n i.i.d. random variables $(X_1, \ldots, X_n)$ be given, which take on letters in a finite alphabet of symbols $\{a_1, \ldots, a_k\}$. Let further the probability distribution of X be given by $(p_x = p(a_x))_{x=1}^k$.

Example. Let X be a binary random variable with an alphabet $\{0,1\}$, $p(0) = 1-p$ and $p(1) = p$ with $0 \leq p \leq 1$. Then the Shannon entropy of X is given by

$$H(X) = -\sum p_x \log_2(p_x) \tag{1.3.26}$$

Exercise. Show that $0 \leq H(X) \leq \log_2(K)$.

We now consider a long message $x^n = x_1 \ldots x_n$ and ask ourselves the question: Can we compress this message to a shorter string of letters with essentially the same information? The answer to this question will be given by Shannon's first theorem.

But before we dive into that, we have to establish some foundations. First, the protocol we are going to use can be schematically reproduced as in Figure 1.3. In such a scheme, we appreciate that Alice encodes n bits as a block with an encoder $\mathcal{E}$, from which the output is a codeword with less than n bits. Next, this codeword is transmitted over noiseless bit channels and Bob receives them. He further decodes it using a decoder $\mathcal{D}$, (hopefully) obtaining the original sequence that Alice sent.

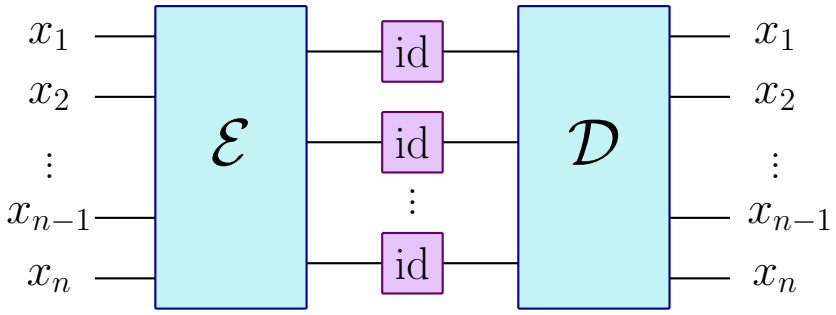


fig. 1.3: Schematic representation of a classical communication channel between Alice and Bob without any noise.

Once the protocol is introduced, we need some further notions and technical results before proceeding to the main result of this section. First, we notice that, by independence,

$$p(x_1 \dots x_n) = p(x_1) \dots p(x_n) \tag{1.3.27}$$

Next, we need to introduce the notion of *typical strings*. As the name suggests, such strings will be the ones that most typically appear and, therefore, the ones to be considered in the asymptotic limit with n.

Definition. 1.3.5 (ε-typical string)
For every $\varepsilon > 0$, the string of source symbols $x_1 \dots x_n$ is called *ε-typical*, if, and only if,

$$2^{-n(H(X)+\varepsilon)} \leq p(x_1 \dots x_n) \leq 2^{-n(H(X)-\varepsilon)}\,, \tag{1.3.28}$$

which can be equivalently written as

$$\left| \frac{1}{n} \log_2 \left(\frac{1}{p(x_1 \dots x_n)} \right) - H(X) \right| \leq \varepsilon\,. \tag{1.3.29}$$

We introduce the notation $T(n, \varepsilon)$ for the set of ε-typical sequences of length n.

Moreover, the proof of the main result of this section, we will need to previously introduce and prove three lemmas. The first one of them essentially states that, in the asymptotic limit, a sequence is typical with high probability.

Lemma. 1.3.6 *Given $\varepsilon > 0$, for every $\delta > 0$ and for large enough n, the probability that a sequence is ε-typical is, at least, $1 - \delta$.*

Proof. Let $\{X_1, \dots, X_n\}$ be i.i.d. random variables. Then we find that

$$\{-\log p(X_1), \dots, -\log p(X_n)\} \tag{1.3.30}$$

are also i.i.d. random variables. We then find that

$$-\frac{1}{n} \sum_{l=1}^{n} \log(p(X_l)) \tag{1.3.31}$$

is a random variable. By the law of large numbers we find that, for every $\varepsilon > 0$ and $\delta > 0$, the following holds for n large enough:

$$P\Big(\Big| -\frac{1}{n} \sum_{l=1}^{n} \log_2 p(X_l) + \sum_{x=1}^{k} p_x \log_2 p_x \Big| \leq \varepsilon \Big) \geq 1 - \delta \tag{1.3.32}$$

If we use that $-\sum_{l=1}^{n} \log_2 p(X_l) = -\log_2 \prod_{l=1}^{n} p(X_l) = -\log_2 p(X_1, \ldots, X_n)$ and $-\sum_{x=1}^{k} p_x \log_2 p_x = H(X)$, we obtain

$$P\Big(\Big|\frac{1}{n}\log_2 \frac{1}{p(X_1,\ldots,X_n)} - H(X)\Big| \leq \varepsilon\Big) \geq 1-\delta\,, \tag{1.3.33}$$

concluding thus the proof. □

Next, we give an approximation that yields the size of the set of typical sequences in the asymptotic limit with n.

Lemma. 1.3.7 *Given $\varepsilon > 0$ and $\delta > 0$, for large enough n the following inequality holds:*

$$(1-\delta)2^{n(H(X)-\varepsilon)} \leq |T(n,\varepsilon)| \leq 2^{n(H(x)+\varepsilon)}\,. \tag{1.3.34}$$

Proof. a) For the second inequality, we have that

$$1 \geq \sum_{x^n \in T(n,\varepsilon)} p(x^n) \geq \sum_{x^n \in T(n,\varepsilon)} 2^{-n(H(X)+\varepsilon)} = |T(n,\varepsilon)|2^{-n(H(X)+\varepsilon)}\,, \tag{1.3.35}$$

where the first inequality comes from the fact that we are adding up probabilities and the second one the definition of ε-typical sets. This gives us immediately that

$$|T(n,\varepsilon)| \leq 2^{n(H(X)+\varepsilon)}. \tag{1.3.36}$$

b) For the lower bound of $|T(n,\varepsilon)|$, we use that

$$1-\delta \leq \sum_{x^n \in T(n,\varepsilon)} p(x^n) \leq \sum_{x^n \in T(n,\varepsilon)} 2^{-n(H(X)-\varepsilon)} = |T(n,\varepsilon)|2^{-n(H(X)-\varepsilon)}, \tag{1.3.37}$$

where the first inequality follows from Lemma 1.3.6 and the second one from the definition of ε-typical sets again. Hence, we can conclude that

$$|T(n,\varepsilon)| \geq (1-\delta)2^{n(H(X)-\varepsilon)}. \tag{1.3.38}$$

□

Note that the content of the previous lemma can be graphically see in Figure 1.4. Finally, we prove that the probability that sequences belong to a set of size smaller than $2^{nH(X)}$ is negligible in the asymptotic limit with n.

Lemma. 1.3.8 *Let $R < H(X)$. We consider a set $S(n)$ of size smaller or equal to 2^{nR} composed of length n sequences from the source. Then for all $\delta > 0$, for n large enough the following holds:*

$$\sum_{x^n \in S(n)} p(x^n) \leq \delta. \tag{1.3.39}$$

Proof. First, we fix $\varepsilon > 0$ such that $R < H(X) - \delta$ and $0 < \varepsilon < \frac{\delta}{2}$. Then, we can split $S(n))$ into its typical and atypical sequences, namely

$$S(n) \begin{cases} S^{\text{typ}}(n) \\ S^{\text{atyp}}(n) \end{cases} \tag{1.3.40}$$

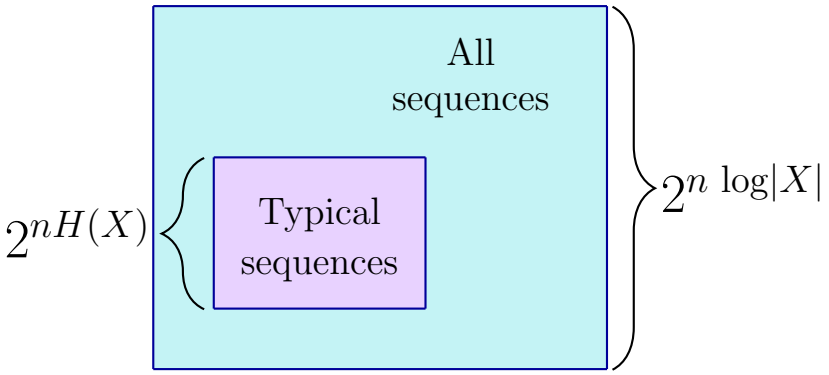


fig. 1.4: A schematic representation of the fact that the set of typical sequences is much smaller than the set of all sequences in general. They have approximately the same size only when the entropy $H(X)$ is equal to $\log|X|$, and in such a case the random variable X is uniformly distributed.

For the atypical sequences, we know that $P(x^n \in S^{\text{atyp}}(n)) \leq \frac{\delta}{2}$ by Lemma 1.3.6. It is only remaining to bound $S^{\text{typ}}(n)$. For that, first note that $|S^{\text{typ}}(n)| \leq 2^{nR}$, and each sequence in this set has probability smaller or equal to $2^{-n(H(X)-\varepsilon)}$ by assumption. This gives us

$$\sum_{x^n \in S^{\text{typ}}(n)} p(x^n) \leq 2^{nR} 2^{-n(H(X)-\varepsilon)} = 2^{-n(H(X)-R-\varepsilon)} \overset{n\to\infty}{\longrightarrow} 0. \tag{1.3.41}$$

We can, hence, conclude that

$$\sum_{x^n \in S(n)} p(x^n) = \sum_{x^n \in S^{\text{typ}}(n)} p(x^n) + \sum_{x^n \in S^{\text{atyp}}(n)} p(x^n) \leq 2^{-n(H(X)-R-\varepsilon)} + \frac{\delta}{2} \leq \delta, \tag{1.3.42}$$

if one chooses n large enough. □

Before introducing the main theorem of this section, we need to introduce some last concepts, which have to do with the compression and decompression of information throughout the communication channel (see Figure 1.3).

Definition. 1.3.9
A *compression scheme* of rate R is a map

$$x^n = (x_1, \ldots, x_n) \mapsto C^n(x^n) \equiv C^n(x_1 \ldots x_n) \tag{1.3.43}$$

producing a bit string of length $\lceil nR \rceil$.
A *decompression scheme* of rate R maps the bit string back to n letters

$$D^n(C^n(x^n)) = \tilde{x}^n \tag{1.3.44}$$

A compression-decompression scheme is *reliable* if

$$P(D^n(C^n(x^n)) = x^n) \overset{n\to\infty}{\longrightarrow} 1 \tag{1.3.45}$$

Now we are in position to state and prove the main result of this section, namely the best rate of compression for a message sent by a communication channel without noise.

Theorem. 1.3.10 (Shannon's noiseless coding theorem) *Let $\{X_1, \ldots, X_n\}$ be i.i.d. random variables with entropy rate $H(X)$. Then*

1. *If $R > H(X)$, there exists a reliable compression-decompression scheme of rate R for the source.*

2. *If $R < H(X)$, any compression scheme of rate R will not be reliable.*

Proof. 1. Let $R > H(X)$. We fix $\varepsilon > 0$ such that $R > H(X) + \varepsilon$. Now we consider $T(n, \varepsilon)$. By Lemma. 1.3.7, we have

$$|T(n,\varepsilon)| \leq 2^{n(H(X)+\varepsilon)} < 2^{nR} \tag{1.3.46}$$

and $P(x^n \in T(n,\varepsilon)) \geq 1 - \delta$ by Lemma. 1.3.6. Then we have that $\{y_1, \ldots, y_k\} \subseteq T(n,\varepsilon)$, $k < 2^{nR}$ where $y_1, \ldots, y_k$ is just enumeration by nR bit strings. I.e.

$$\begin{aligned} y_1 &= (0, \ldots, 0) \\ y_2 &= (1, 0, \ldots, 0) \\ &\vdots \end{aligned} \tag{1.3.47}$$

all with nR entries and a 1 in one entry and zeros in all others. Note that, in general, $k < 2^{nR}$, and thus $y_k \neq (1, \ldots, 1, 1)$ in general. In the rest of the proof of this statement, we are going to construct the compression-decompression scheme using the previous enumeration of $T(n,\varepsilon)$. First, the coding scheme is constructed as

$$x^n \mapsto C^n(x^n) = \begin{cases} y_j & \text{if } x^n \text{ is } \varepsilon\text{-typical} \\ & (\text{corresponding } y_j \text{ representation}) \\ y_1 & \text{if } x^n \text{ is } \varepsilon\text{-atypical} \end{cases} \tag{1.3.48}$$

Note that the choice of the bit string to which we map the atypical sequences (above y_1) is not important; we just map them all to the same string for simplicity. Moreover, the decoding scheme is defined in the following way:

$$y_j \mapsto D^n(y_j) = \begin{cases} x^n & \text{if } j \leq K \\ & (y_j \text{ and } x^n \text{ corresponding message}) \\ x_1 \ldots x_1 & \text{if } j > K \end{cases} \tag{1.3.49}$$

In this way, it is clear that we have defined a bijection over the typical sequences. More specifically, for any $x^n \in T(n,\varepsilon)$, $D^n(C^n(x^n)) = x^n$, and therefore, since

$$P(x^n \in T(x,\varepsilon)) \geq 1 - \delta\,, \tag{1.3.50}$$

we obtain

$$P(D^n(C^n(x^n)) = x^n) \geq 1 - \delta \overset{n\to\infty}{\longrightarrow} 1 \tag{1.3.51}$$

2. Let $R < H(X)$. As seen above, the compression-decompression scheme has at most 2^{nR} possible outputs. Then, at most 2^{nR} sequences can be compressed (and decompressed) without an error with $S(n)$ the set of such sequences. By Lemma. 1.3.8, for large enough n, $P(x^n \in S(n)) \leq \delta \overset{n\to\infty}{\longrightarrow} 0$. We can, hence, conclude that the compression-decompression scheme is not reliable.

□

1.3.2 Shannon's noisy channel coding theorem

Let X, Y be random variables in the following. We revisit some already known quantities and define a new one:

- The joint entropy
$$H(X,Y) = -\sum_{x,y} p_{x,y} \log p_{x,y} \tag{1.3.52}$$

- Conditional entropy
$$H(X|Y) = \sum_{y} p_y H(X|Y=y) = H(X,Y) - H(Y) \tag{1.3.53}$$

- Mutual information
$$I(X:Y) = H(X) + H(Y) - H(X,Y) \tag{1.3.54}$$

For these notions, we can prove a series of fundamental properties, whose proof we do not include in the main text as they were left as exercises for the students attending the course.

Proposition. 1.3.11 *Given two random variables X and Y, the following properties hold for their joint entropy, conditional entropy and mutual information:*

1. *Show that $H(X|Y) \geq 0$ and further*
   (a) *$H(X,Y) \geq H(X)$ with equality if and only if $X = f(Y)$*
   (b) *$I(X:Y) \leq H(Y)$ with equality if and only if $X = f(Y)$*
2. *Show the subadditivity of the joint entropy, i.e.*
$$H(X,Y) \leq H(X) + H(Y) \tag{1.3.55}$$
   with equality if and only if X, Y are independent.
3. *$H(X|Y) \leq H(X)$ and thus $I(X:Y) \geq 0$ with equality if and only if X, Y are independent.*
4. *Show the chain rule for $X_1, \ldots, X_n, Y$, i.e.*
$$H(X_1, \ldots, X_n|Y) = \sum_{i=1}^{n} H(X_i|Y, X_1, \ldots, X_{i-1}) \tag{1.3.56}$$

The following theorem is a well-known inequality which is required for the sketch of the proof of the main result of this section, namely Shannon's noisy theorem.

Theorem. 1.3.12 (Fano's inequality) *Let X and Y be two random variables and $\tilde{X} = f(Y)$ a function of Y with which we intend to guess the value of X. Let $p_e = p(X \neq \tilde{X})$ be the error made by guessing X with $\tilde{X}$.*

$$H(p_e) + p_e \log(|X| - 1) \geq H(X|Y) \tag{1.3.57}$$

Proof. We define the random variable

$$E = \begin{cases} 1 & \text{if } X \neq \tilde{X} \\ 0 & \text{if } X = \tilde{X} \end{cases} \tag{1.3.58}$$

then the following equations hold

- $H(E) = H(p_e)$
- $H(E|X,Y) = 0$
- $H(E|Y) \leq H(E) = H(p_e)$

By the chain rule and using the above equations we immediately get

$$H(E, X|Y) = H(X|Y) + H(E|X, Y) = H(X|Y) \tag{1.3.59}$$

and

$$H(E, X|Y) = H(E|Y) + H(X|E, Y) \leq H(p_e) + H(X|E, Y). \tag{1.3.60}$$

Thus we find that

$$H(X|Y) \leq H(p_e) + H(X|E, Y). \tag{1.3.61}$$

What remains is to upper bound $H(X|E, Y)$ which can be done by noticing that

$$\begin{aligned} H(X|E, Y) &= p(E = 0)H(X|E = 0, Y) + p(E = 1)H(X|E = 1, Y) \\ &\leq p(E = 0)0 + p_e \log(|X| - 1) \\ &= p_e \log(|X| - 1) \end{aligned} \tag{1.3.62}$$

□

Shannon's noisy coding theorem deals with a channel that in addition to encoding and decoding is perturbed by a source of noise. An schematic representation of the communication channel in such a case is provided in Figure 1.5. Note that the situation is similar to that of the case without noise, with the main difference that, now, the codeword that outputs the encoder is sent to Bob through some channels which model the noise present in the environment. To understand mathematically how to describe such a noise, we start by introducing some new notions.

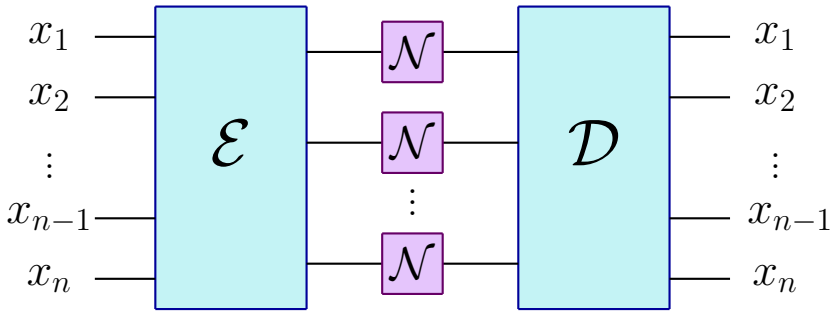


fig. 1.5: Schematic representation of a classical communication channel between Alice and Bob in the presence of noise.

Definition. 1.3.13 (Classical channel)
A (classical) channel is a positive linear map

$$\mathcal{N} : \mathbb{R}_A^n \to \mathbb{R}_B^n \tag{1.3.63}$$

verifying

$$\sum_{i=1}^{m} \mathcal{N}(p_i) = 1 \quad \forall (p_i)_{i=1}^n \text{ s.t. } \sum_{i=1}^{n} p_i = 1 \text{ and } p_i \geq 0\,. \tag{1.3.64}$$

Notation. In the following we will denote

$$\ell_1^k = (\mathbb{R}^k, \|\cdot\|_1) \tag{1.3.65}$$

and write

$$\mathcal{N} : \ell_1^n \to \ell_1^m. \tag{1.3.66}$$

Definition. 1.3.14 (Capacity of a channel)
Intuitively, the capacity of a channel $\mathcal{N}$ is an asymptotic limit of

$$\frac{\#\text{ transmitted bits with error } \varepsilon \to 0}{\#\text{ required uses of the channel in parallel}}. \tag{1.3.67}$$

More precisely $\mathcal{N} : \ell_1^n \to \ell_1^n$, the *(classical) capacity* is defined as

$$C_c(\mathcal{N}) := \lim_{\varepsilon \to 0} \limsup_{k \to \infty} \left\{ \frac{m}{k} \ : \ \exists \mathcal{E}, \exists \mathcal{D} \text{ s.t. } \left\| \mathrm{id}_{\ell_1^{2m}} - \mathcal{D} \circ \mathcal{N}^{\otimes k} \circ \mathcal{E} \right\| < \varepsilon \right\}. \tag{1.3.68}$$

Here we used the following notation

$$\begin{aligned} &\mathcal{E} \text{ encoder}, \varepsilon : \ell_1^{2m} \to \otimes^k \ell_1^n, \\ &\mathcal{D} \text{ decoder}, \ \mathcal{D} : \otimes^k \ell_1^n \to \ell_1^{2m}, \end{aligned} \tag{1.3.69}$$

and the k parallel uses of the channel $\mathcal{N}$ are denoted by

$$\mathcal{N}^{\otimes k} : \otimes^k \ell_1^n \to \otimes^k \ell_1^n. \tag{1.3.70}$$

Now we are in position to state the main result of this section, namely Shannon's noisy channel coding theorem. A complete proof of this theorem would require many more preliminaries and is far beyond the scope of this course. Therefore, below we only sketch some ideas behind the proof of this result. However, we will provide a complete proof of its quantum analogue later in the course, and since the quantum setting will be shown to constitute an extension of the classical one, in particular such a proof will be valid for the current result.

Theorem. 1.3.15 (Noisy channel coding theorem) *For a noisy channel $\mathcal{N} : \ell_1^n \to \ell_1^n$, its classical capacity can be recovered from:*

$$C_c(\mathcal{N}) = \max_{P=(p(x))_{x=1}^n} I(X:Y) \tag{1.3.71}$$

where the maximum is taken over the input distributions $(p(x))_{x=1}^n$ for X, for one use of the channel, and Y the corresponding random variable at the output of $(\mathcal{N}(p))_{y=1}^n$.

Proof (sketch). Our first goal is to show that

$$C_c(\mathcal{N}) \geq \max_{P=((p(x))_{x=1}^n} I(X:Y). \tag{1.3.72}$$

Let $X = \{x, p_x\}$ be the probability distribution for the input letters. Using X and $\mathcal{N} = (P(x|y))_{x,y}$ we determine $Y = \{y, p_y\}$.

We further have that codewords are chosen with a prior probability distribution governed by X^n and that they are chosen from a typical set of $2^{nH(X)}$ elements with high probability. On the other hand for a received message Y^n about $2^{nH(X|Y)}$ messages could have been sent. This means, to make our decoding reliable, we need our input codewords to be chosen so that the error spheres of two different codewords do not overlap (with high probability).

Decoding is now done by associating to Y^n a sphere of $2^{n(H(X|Y)+\delta)}$ possible inputs such that only one codeword is contained in this sphere. However, as every sphere contains a fraction of

$$\frac{2^{n(H(X|Y)+\delta)}}{2^{nH(X)}} = 2^{-n(H(X)-H(X|Y)-\delta} = 2^{-n(I(X:Y)-\delta} \tag{1.3.73}$$

codewords it can happen that more than one codeword is contained in a sphere. If we set the number of codewords to 2^{nR}, with R the channel capacity (or rate), then the probability that all of them lie in the decoding sphere is

$$2^{nR} 2^{-n(I(X:Y)-\delta)} = 2^{-n(I(X:Y)-R-\delta)} \xrightarrow{\delta \to 0} R \to I(X:Y). \tag{1.3.74}$$

Hence the average probability of error is small over all codewords. □

Example. If we consider a binary random variable with

$$\begin{array}{ll} P(0|0) = 1-p & P(1|1) = 1-p \\ P(0|1) = p & P(1|0) = p \end{array} \tag{1.3.75}$$

$x^n = x_1 \dots x_n$ is manipulated by the error channel and eventually np bits are flipped. The goal then is to find so called Hamming spheres with center x^n and radius np. The assure this condition one chooses from a typical set of $2^{nH(x)}$ elements.

1.4 Quantum Information Theory: Preliminaries

In this section, we are going to provide a brief introduction to quantum mechanics and its formalism, from a mathematical perspective. The concepts that we will introduce below are essential for the postulates that we will present in the following subsections.

From now on, we will be working with n-dimensional (complex) Hilbert spaces $\mathcal{H}$, which can be identified with $\mathbb{C}^n$. If the dimension is irrelevant or clear from the context we will just write $\mathcal{H}$.

Notation. We further introduce the following (*bra-ket*) notation originally used by Paul Dirac (1904 - 1984). He set

$$\begin{array}{ll} |\psi\rangle \in \mathbb{C}^n & \text{to be a vector,} \\ \langle\psi| \in (\mathbb{C}^n)^* & \text{to be a dual vector.} \end{array} \tag{1.4.1}$$

This notation originated from the notation for the inner product on $\mathcal{H}$:

$$\langle\psi|\psi\rangle \in \mathbb{R}\,. \tag{1.4.2}$$

In this notation one can write the rank one operators onto the space spanned by $|\psi\rangle$ as a ket-bra

$$|\psi\rangle\langle\psi| : \mathbb{C}^n \to \mathbb{C}^n \tag{1.4.3}$$

allowing for the convenient use of these objects, for every $|\xi\rangle \in \mathbb{C}^n$,

$$|\psi\rangle\langle\psi|\xi\rangle = \langle\psi|\xi\rangle\,|\psi\rangle \in \mathbb{C}^n\,. \tag{1.4.4}$$

The content of the following subsections has been inspired by the courses some basic texts of quantum information theory, such as the courses [12], [25] and [20], as well as the books [23], although one of the most fundamental texts in this field is [11]. We refer the reader to any of those texts for further knowledge on the topic.

1.4.1 Qubits and basic operations

Another essential concept for quantum information theory is the one of a qubit. A *qubit* is the simplest quantum mechanical system and plays the same role in quantum information theory as the *bit* in classical information theory, which can be 0 or 1. Hence, it is the basic unit of information and extends the the concept of a classical bit, which is just 0 or 1 to a *superposition* of those two. Formally, it is the system described by a two-dimensional Hilbert space. We will denote the canonical basis of this vector space $\mathbb{C}^n$ as $\{|0\rangle, |1\rangle\}$, i.e.

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{1.4.5}$$

This basis is usually called the *computational basis*. Then, while a classical bit can be in the state 0 or in the state 1, an arbitrary state for a qubit is a vector

$$|\psi\rangle = a_0\,|0\rangle + a_1\,|1\rangle \in \mathbb{C}^2 \tag{1.4.6}$$

with $a_0, a_1 \in \mathbb{C}$ and $|a_0|^2 + |a_1|^2 = 1$. If $a_0 \neq 0, a_1 \neq 0$, we say that the state is in superposition of the situations $|0\rangle$ and $|1\rangle$. Notice that this fact leads to the essential difference between the possible states of a bit, which are just two, 0 or 1, and the possible states of a qubit, which, in principle, are infinite. This new situation allows us to perform new protocols for quantum information processing. Indeed, this principle constitutes the basis of the theoretical quantum computer, for which we can briefly mention the main idea behind it in a nutshell:

- If we consider one bit, we can perform 1 operation at a time, whereas we can perform 2 operations simultaneously with one qubit. This is due to the superposition phenomenon mentioned above, since now an arbitrary state is of the form

$$|\psi\rangle = a_0 |0\rangle + a_1 |1\rangle \in \mathbb{C}^2 ,$$

  where one can see two basic bits (thus, operations) being performed at the same time.

- If we now have two bits, we can perform 2 operations at a time, while, if we consider two qubits, 4 operations can be performed at the same time (we will see that when we consider a superposition of the four elements of the Bell basis).

- In general, with n qubits, one can perform n operations simultaneously (one per each bit), whereas with n qubits one can perform $2n$ operations at the same time.

Hence, theoretically, a quantum computer should be able give an exponential improvement to the amount of operations performed in parallel compared to a classical computer (i.e. an exponential speed-up).

Let us go back now to the definition and basic properties of qubits. It is important to remark that, even though a given qubit can be in any superposition state $a_0 |0\rangle + a_1 |1\rangle$, if we *measure* the state of such a qubit, we will obtain either the value $|0\rangle$ or $|1\rangle$ for the state of the qubit (these states can be seen as classical ones), with certain probabilities. Hence, we cannot "observe" the superposition phenomenon, although we are able to use it, as we will see below.

In the following, another basis will be also rather important and we want to introduce it here. It is given as

$$|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \qquad |-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) \tag{1.4.7}$$

To extract classical information from a quantum system one performs a measurement. Performing such a measurement on such an arbitrary state the systems turns out to be in the state $|0\rangle$ with probability $|a_0|^2$ and in the state $|1\rangle$ with $|a_1|^2$.

A single qubit lives in $\mathbb{C}^2$. However, one can consider systems of more qubit to have richer spaces. For example, if we consider 2 qubits, the 2-qubit system that we get has four elements in a possible basis:

$$\{|0\rangle \otimes |0\rangle , |0\rangle \otimes |1\rangle , |1\rangle \otimes |0\rangle , |1\rangle \otimes |1\rangle\},$$

where, in each case, the qubit in the left part denotes the first qubit (and associated to the first system), and the right one denotes the second qubit. If one considers $|0\rangle \otimes |1\rangle$, for instance, this element can be also expressed by $|0\rangle |1\rangle$ or $|01\rangle$, and the structure of tensor product implies that in $\mathbb{C}^4$ can be written as:

$$\begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}.$$

More generally, as mentioned previously, if one considers a system of n qubits, a basis of such system has 2^n elements (it is equivalent to saying that, with n qubits, one can perform 2^n operations simultaneously). In particular, one can always consider for such elements of the basis the elements $|a_1\rangle \otimes |a_2\rangle \otimes \ldots \otimes |a_n\rangle$, with $a_i \in \{0,1\}$ for all $i = 1, \ldots, n$. Since there are 2^n elements in this basis, we can change this previous notation to $|0\rangle, |1\rangle, \ldots |2^{n-1}\rangle$, to simplify it.

Therefore, a quantum state on n qubits, because of superposition, is given by

$$\alpha_0 |0\rangle + \alpha_1 |1\rangle + \ldots + \alpha_{2^n-1} |2^n - 1\rangle, \quad \sum_{i=0}^{2^n-1} |\alpha_i|^2 = 1.$$

Moreover, as in the case of a single qubit, if one measures this in the computational basis, one just gets a "classical" n-bit state, $|i\rangle$, with probability $|\alpha_i|^2$.

In a more general setting, consider a physical system that can be in N different, mutually exclusive classical states (in the case of the qubit, $N = 2$, and for n qubits, $N = 2^n$). A *pure quantum state* $|\varphi\rangle$ is a superposition of classical states in the following form:

$$|\varphi\rangle = \alpha_1 |1\rangle + \alpha_2 |2\rangle + \ldots + \alpha_N |N\rangle.$$

The elements α_i in the previous expression are complex numbers that are called *amplitudes*, and, in this expression, it is easy to read the superposition phenomenon as the possibility of a quantum system to be in N classical states states at the same time (or perform N operations simultaneously, as mentioned above).

1.4.1.1 Measurement

In general, given a quantum state, we can consider two different scenarios: Either we measure it, or we let it evolve under a unitary without measuring it. In this subsection, we explain the first case.

Let us recall some basic notions about Hilbert spaces and their scalar products. Let $\mathcal{H}$ be a Hilbert space (in general, we will just consider finite-dimensional spaces) and let $T : \mathcal{H} \to \mathcal{H}$ be a linear operator on it. Since $\mathcal{H}$ is a Hilbert space, in particular it is a normed space with an associated norm $\|\cdot\|_{\mathcal{H}}$, which comes from a scalar product $\langle \cdot, \cdot \rangle$.

We say that T is a bounded operator if

$$\|T\|_{\mathcal{H}\to\mathcal{H}} := \sup_{x\in\mathcal{H}} \frac{\|T(x)\|_{\mathcal{H}}}{\|x\|_{\mathcal{H}}} < \infty,$$

and denote by $\mathcal{B}(\mathcal{H})$ the space of bounded linear operators on $\mathcal{H}$. Moreover, if $T : \mathcal{H} \to \mathcal{H}$, we can define its dual operator, and denote it by T^*, as the operator that satisfies

$$\langle y, T(x)\rangle = \langle T^*(y), x\rangle \qquad \text{for every } x, y \in \mathcal{H}.$$

Now we are in position to formally define a measurement in the following form:

Definition. 1.4.1 (Measurement)
Let $\{M_n\}_n \subset \mathcal{B}(\mathcal{H})$ be a collection of operators verifying

$$\sum_n M_n^* M_n = \mathbb{1} \tag{1.4.8}$$

where $\mathbb{1}$ denotes the identity operator (we drop the subindex with the dimension when there is no possible confusion) and M_n^* denotes the dual of the operator M_n. This collection of operators is called **quantum measurements** when the following holds: Given a state of a quantum system

$|\varphi\rangle$ before performing this operation to measure it, the probability that result $|n\rangle$ occurs is given by

$$p(n) = \langle \varphi, M_n^* M_n \varphi \rangle \tag{1.4.9}$$

and the state of the system after this operation is given by

$$\frac{M_n |\varphi\rangle}{\sqrt{p(n)}}.$$

Consider again the state

$$|\varphi\rangle = \alpha_1 |1\rangle + \alpha_2 |2\rangle + \ldots + \alpha_N |N\rangle.$$

and assume that we measure it. As we have already mentioned, we will obtain the classical state $|i\rangle$, with probability $|\alpha_i|^2$, thus we cannot "see" the superposition itself. Among some other things, this means that the probability to get specifically the state $|i\rangle$ when we measure, and not another one, is $|\alpha_i|^2$. Hence, since the quantum state induces a probability distribution on the classical states, this implies

$$\sum_{i=1}^{N} |\alpha_i|^2 = 1.$$

Notice that when we measure $|\varphi\rangle$ and get a classical state, $|\varphi\rangle$ disappears, and all that is left is the classical state itself. We say then that $|\varphi\rangle$ has *collapsed* to the classical state that we got, and the information encoded in the amplitudes α_i is now gone.

In general, we will measure in the computational basis. However, there are several ways to perform these measurements, that we will present throughout this section. Let us begin with the easiest one, the measurement of a qubit in its computational basis. It is defined by the measurement operators:

$$M_0 = |0\rangle \langle 0| \qquad \text{and} \qquad M_1 = |1\rangle \langle 1|.$$

Notice that both operators are *selfadjoint*, i.e., they coincide with their dual operators (actually, they are projections), and they verify $M_i^* M_i = M_i^2 = M_i$, for $i = 1, 2$, where M_i^* denotes the dual of the operator M_i, and $M_0 + M_1 = \mathbb{1}$. Also, when we measure

$$|\varphi\rangle = \alpha_0 |0\rangle + \alpha_1 |1\rangle,$$

the probability to obtain the outcome $|i\rangle$ is $|\alpha_i|^2$, and the state after measurement in that case is $\frac{\alpha_i}{|\alpha_i|} |i\rangle$. Actually, we can see that this state is equivalent to $|i\rangle$ (since it is just a rotation of the latter).

Indeed, consider $|\varphi\rangle$ and $e^{i\theta} |\varphi\rangle$ (which is a more general expression for the element mentioned above) and assume that we measure both states with a measurement $\{M_n\}_n$. Then, the probability of getting outcome n for the second element is

$$\langle \varphi e^{-i\theta} | M_n^* M_n | e^{i\theta} \varphi \rangle = \langle \varphi | M_n^* M_n | \varphi \rangle,$$

the same that for the first element. Hence, both states are operationally identical.

1.4.1.2 Distinguishability of quantum states

One typical problem in quantum information is to distinguish between two (or more) quantum states. Namely, among several possible states, we have a particle in one of them and we want to find out in which one the particle actually is. We study this problem now in the simplest case: to distinguish between two possible states.

First, let us assume that the states that we want to distinguish, $|\varphi_1\rangle$ and $|\varphi_2\rangle$ are orthogonal. Then, if we choose the measurement operators $M_i = |\varphi_i\rangle\langle\varphi_i|$ for $i = 1, 2$ and define $M_0 = \mathbb{1} - \sum_i M_i$, it is easy to see that

$$M_0 + M_1 + M_2 = \mathbb{1}.$$

Therefore, given $i \in \{1, 2\}$, if $|\varphi\rangle$ is prepared in the state $|\varphi_i\rangle$, we have

$$p(i) = \langle\varphi|M_i|\varphi\rangle = 1, \qquad \text{and} \qquad p(j) = 0 \text{ for } j \neq i,$$

thus both states can be unambiguously distinguished.

Now, suppose that we want to distinguish two non-orthogonal states $|\varphi_1\rangle$ and $|\varphi_2\rangle$. We can prove that there is no way to do that in general.

Proposition. 1.4.2 *If $|\varphi_1\rangle$ and $|\varphi_2\rangle$ are not orthogonal, then we cannot distinguish between them.*

Proof. By reduction to the absurd, let as assume that there is a measurement $\{M_n\}_{n\in I}$ capable of doing that, i.e., distinguishing both states. Then, we can consider a partition of the set I ($I_1, I_2 \subset I$ such that $I_1 \neq \emptyset \neq I_2$, $I_1 \cup I_2 = I$ and $I_1 \cap I_2 = \emptyset$) and this allows us to decide that if the result of the measurement is $|m\rangle$, with the index $m \in I_i$, then the state is $|\varphi_i\rangle$.

Consider, for $i = 1, 2$, the operations $\mathbb{E}_i = \sum_{j\in I_i} M_j^* M_j$, which satisfy by construction $\mathbb{1} = \mathbb{E}_1 + \mathbb{E}_2$. It is clear that we have

$$\langle\varphi_i|\,\mathbb{E}_i\,|\varphi_i\rangle = 1 \quad \text{for } i = 1, 2,$$

and

$$\langle\varphi_1|\,\mathbb{E}_2\,|\varphi_1\rangle = \langle\varphi_2|\,\mathbb{E}_1\,|\varphi_2\rangle = 0.$$

Consider the first term in the previous equality. Since $\mathbb{E}_2$ is positive, one can write:

$$0 = \langle\varphi_1|\,\mathbb{E}_2\,|\varphi_1\rangle = \Big\langle\varphi_1|\sqrt{\mathbb{E}_2}\sqrt{\mathbb{E}_2}|\varphi_1\Big\rangle,$$

so we get $\sqrt{\mathbb{E}_2}\,|\varphi_1\rangle = 0$. By assumption, $|\varphi_1\rangle$ and $|\varphi_2\rangle$ are not orthogonal, thus there exist $\alpha_1 \neq 0 \neq \alpha_2$ and another state $|\phi\rangle$, orthogonal to $|\varphi_1\rangle$, so that

$$|\varphi_2\rangle = \alpha_1\,|\varphi_1\rangle + \alpha_2\,|\phi\rangle.$$

Applying $\sqrt{\mathbb{E}_2}$ to this expression, we get:

$$\sqrt{\mathbb{E}_2}\,|\varphi_2\rangle = \alpha_1\sqrt{\mathbb{E}_2}\,|\varphi_1\rangle + \alpha_2\sqrt{\mathbb{E}_2}\,|\phi\rangle = \alpha_2\sqrt{\mathbb{E}_2}\,|\phi\rangle.$$

However, this is a contradiction, since this implies $|\alpha_2| = 1$ and, by assumption, we have $|\alpha_2| < 1$ (since $\alpha_1 \neq 0$). □

1.4.1.3 Projective Measurements

In this subsubsection, we are going to introduce *projective measurements*, which play an special role in Postulate III of quantum mechanics (as we will see in the following section). They can be defined in the following form.

Definition. 1.4.3 (Projective measurement)
Consider a collection $\{M_n\}_n \subset \mathcal{B}(\mathcal{H})$ of measurements, as described in Definition 1.4.1. Assume that they have the additional property that the M_n are orthogonal projections, i.e., they are self-adjoint and verify

$$M_n M_m = \delta_{mn} M_n,$$

where $\delta_{mn} = 1$ iff $m = n$ and 0 otherwise. These measurements are called **projective measurements**.

It is clear that each one of these operators M_n projects on a subspace $\mathcal{H}_n \subset \mathcal{H}$ of the global Hilbert space. Hence, an *observable* M can be defined as the Hermitian operator

$$M = \sum_n \lambda_n M_n,$$

where the term in the right hand-side is, in fact, the spectral decomposition of M. Moreover, the possible outcomes of the measurement correspond to the eigenvalues λ_n of the observable, and when we measure the state $|\varphi\rangle$, the probability of getting state $|n\rangle$ is:

$$p(n) = \langle\varphi|M_n|\varphi\rangle.$$

With this notation, the average value of the measurement, with respect to the state $|\varphi\rangle$, is

$$\sum_n np(n) = \sum_n n \langle\varphi|M_n|\varphi\rangle = \langle\varphi|M|\varphi\rangle.$$

1.4.1.4 POVM Measurements

In many situations, we will not be as interested in the post measurement state of our particle itself as in the probabilities of the different possible measurement outcomes. In this case, we can reduce to the formalism of the so called *Positive Operator Valued Measurements (POVM's)*.

Let us recall that an operator $T \in \mathcal{B}(\mathcal{H})$ is said to be *positive* (shortened form of *positive semidefinite*) if

$$\langle x, T(x)\rangle \geq 0 \qquad \forall x \in \mathcal{H}.$$

As we will see below, the operators mentioned in the definition of POVM are clearly positive, since

$$\langle x, E_n(x)\rangle = \langle M_n(x), M_n(x)\rangle = \|M_n(x)\| \geq 0 \qquad \forall x \in \mathcal{H}.$$

Definition. 1.4.4 (Positive operator valued measure)
Consider a measurement $\{M_n\}_n \in \mathcal{B}(\mathcal{H})$ as in the Definition 1.4.1. Then, we can define the *positive* operators

$$E_n = M_n^* M_n.$$

This family of operators $\{E_n\}_n$ is called a **POVM**.

The operators presented in the definition of POVM clearly satisfy

$$\sum_n E_n = \mathbb{1}$$

and their probability of obtaining outcome m is

$$p(m) = \langle\varphi|E_m|\varphi\rangle.$$

Conversely, if we have a collection of positive operators $\{E_n\}_n$ verifying $\sum_n E_n = \mathbb{1}$, we can define a measurement $\{M_n\}_n$ from them just by considering $M_n = \sqrt{E_n}$.

1.4.1.5 Unitary Evolution

As opposed to the previous subsection, now we let our quantum state evolve without measuring it. Assume we have a system of the form

$$|\varphi\rangle = \alpha_1 |1\rangle + \ldots + \alpha_N |N\rangle \tag{1.4.10}$$

and want to transform this to the system

$$|\psi\rangle = \beta_1 |1\rangle + \ldots + \beta_N |N\rangle . \tag{1.4.11}$$

Quantum mechanics only allows linear operations to be applied to quantum states. This means that, after a change of notation (identifying $|\varphi\rangle$ with an n-dimensional vector), applying an operation that changes $|\varphi\rangle$ to $|\psi\rangle$ corresponds just to a multiplication by an $N \times N$ complex-valued matrix. With the previous expressions for $|\varphi\rangle$ and $|\psi\rangle$, one has

$$U \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_N \end{pmatrix} \tag{1.4.12}$$

and adding the condition that

$$\sum_{i=1}^{N} |\beta_i|^2 = 1, \tag{1.4.13}$$

we immediately get that U has to be a unitary. This means

$$UU^* = U^*U = \mathbb{1} \tag{1.4.14}$$

Since it is unitary, then, in particular, $U^{-1} = U^*$, and this inverse always exists, what can be translated in the quantum setting to the fact that every non-measuring operation on a quantum state must be reversible (in contrast with measurements, which were clearly non-reversible).

We present now a prominent example of unitaries, the *Pauli matrices*.

Example.

$$\begin{aligned} \sigma_0 = \mathbb{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad & \sigma_x = X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \\ \sigma_y = Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad & \sigma_z = Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \end{aligned} \tag{1.4.15}$$

From a quantum computational point of view we can think of unitary matrices as quantum logical gates. We will deepen in this connection in the first section of the following chapter.

1.4.1.6 Density operators

In this subsection, we introduce the density operators formalism that will be necessary to present the Postulates of the Quantum Mechanics in the Schrödinger picture. Before moving to the definition of density operators, let us start by recalling some basic concepts. We start by recalling the notion of trace of an operator.

Definition. 1.4.5 (Trace)
Let $T : \mathcal{H} \to \mathcal{H}$ a linear map represented by a matrix M in a certain basis. We then define

$$\mathrm{Tr}[T] = \mathrm{Tr}[M] = \sum_i M_{ii} \in \mathbb{C} \tag{1.4.16}$$

as the sum of the diagonal elements of the matrix M. The trace of T is well defined as it is cyclic and linear. This means it it is invariant under basis change.

It is easy to see that the trace is linear and cyclic, i.e., for A and B matrices,

$$\mathrm{Tr}(AB) = \mathrm{Tr}(BA).$$

From this last property, one also gets unitary invariance: For every unitary operator U,

$$\mathrm{Tr}(UAU^*) = \mathrm{Tr}(U^*UA) = \mathrm{Tr}(A).$$

Finally, another useful and interesting property concerning the trace is the following. Let $|\varphi\rangle \in \mathcal{H}$ be a state (or unit vector), and consider the rank-one operator $|\varphi\rangle\langle\varphi| : \mathcal{H} \to \mathcal{H}$, which projects in the direction of $|\varphi\rangle$. Consider now an arbitrary operator $T \in \mathcal{B}(\mathcal{H})$ and suppose that we want to compute $\mathrm{Tr}(A\,|\varphi\rangle\langle\varphi|)$. To do that, before we express $|\varphi\rangle$ in a basis $\{|i\rangle\}$ of $\mathcal{H}$ where the first element is exactly $|\varphi\rangle$, i.e., $|\varphi\rangle = |1\rangle$. Then, we get:

$$\mathrm{Tr}(A\,|\varphi\rangle\langle\varphi|) = \sum_i \langle i|A|\varphi\rangle\langle\varphi|i\rangle = \langle\varphi|A|\varphi\rangle$$

Now, let us move to the formalism of density operators. In the previous subsections, we have described the state of a physical system identifying it with a unit vector in the Hilbert space $\mathcal{H}$. However, there is an equivalent description with trace-class operators on the Hilbert space. One of the main advantages of this description with respect to certain problems appears, for example, when dealing with real experimental systems where noise is present.

A motivation for this formalism comes from the following situation: Sometimes, we do not know whether our system is in a specific state $|\varphi\rangle$, but rather that it is in each one of the states $|\varphi_i\rangle$ with probability p_i, respectively. Hence, we would like to be able to consider the element

$$\sum_i p_i\,|\varphi_i\rangle,$$

with the constants p_i verifying

$$\sum_i p_i = 1,$$

and work with it as a state. However, it is not a state anymore, since it is not a unit vector. To avoid this difficulty, one can associate each state $|\varphi_i\rangle$ to the rank-one projector $|\varphi_i\rangle\langle\varphi_i|$. Hence, the state in the previous scenario can be described, instead, in the following form

$$\rho = \sum_i p_i\,|\varphi_i\rangle\langle\varphi_i|,$$

where ρ is a Hermitian, positive semidefinite, and trace one operator. Indeed, it is clear from its description that ρ is Hermitian and has trace one (because of the linearity of the trace and the fact that $\mathrm{Tr}(|\varphi_i\rangle\langle\varphi_i|) = 1$). To see that it is positive semidefinite, notice that for any $|\phi\rangle \in \mathcal{H}$,

$$\langle\phi|\rho|\phi\rangle = \sum_i p_i\,\langle\phi|\varphi_i\rangle\langle\varphi_i|\phi\rangle = \sum_i p_i\,|\langle\phi|\varphi_i\rangle|^2 \geq 0.$$

These operators are called *density operators* or *density matrices* and the set of such elements is usually denoted by $\mathcal{S}(\mathcal{H})$.

Definition. 1.4.6 (Quantum state/Density operators)
A quantum state or density operator is an linear continues operator $\rho \in \mathcal{B}(\mathcal{H})$, which is positive semi-definite, i.e.

$$\langle\psi|\rho|\psi\rangle \geq 0 \qquad \forall\, |\psi\rangle \in \mathcal{H}, \tag{1.4.17}$$

and has trace one, i.e. $\mathrm{Tr}[\rho] = 1$.

1.4.2 Postulates of quantum mechanics

The postulates of quantum mechanics were derived after a long process of trial and error, which involved a considerable amount of guessing and fumbling by the originators of the theory. The motivation for them is not always clear; even to experts the basic postulates of quantum mechanics appear surprising.

In this section, we mostly focus on the mathematical formulation for the postulates of quantum mechanics in two different (and dual) settings, Heisenberg and Schrödinger picture. These two descriptions will help us to understand the topics presented above.

1.4.2.1 Heisenberg picture

The postulates in the Heisenberg picture can be stated as follows.

Postulate. 1
Given an isolated physical system, there is a complex Hilbert space $\mathcal{H}$ associated to it, called **state space**. Moreover, the physical system is described by a **state vector**, a normalised vector in this space.

In general, the state space $\mathcal{H}$ of the system under study will depend on the specific physical system, but we know that it is a *separable* Hilbert space. Frequently, one restricts to finite-dimensional Hilbert spaces for simplicity.

Postulate. 2
Given an isolated physical system, its evolution is described by a **unitary**. If the system is in the state $|\varphi_1\rangle$ at time $t = t_1$ and in the state $|\varphi_2\rangle$ at time $t = t_2$, then there exists a unitary $U(t_1, t_2) = U_{t_1,t_2}$ such that

$$|\varphi_2\rangle = U_{t_1,t_2}\,|\varphi_1\rangle\,. \tag{1.4.18}$$

This can be generalised using the Schrödinger equation: Given a closed quantum system (with no interaction with an environment), the time evolution of a state on such system is described by

$$i\hbar\frac{d}{dt}\,|\varphi_t\rangle = H\,|\varphi_t\rangle\,. \tag{1.4.19}$$

where $\hbar$ is the Planck's constant. The linear self-adjoint operator H (generally time dependent) is called *Hamiltonian* and describes the dynamics of the system. Let us consider the spectral decomposition of the Hamiltonian (since it is a Hermitian operator):

$$H = \sum_{E_i} E_i\,|E_i\rangle\,\langle E_i|,$$

where we denote by E_i the eigenvalues and by $|E_i\rangle$ the corresponding normalized eigenvectors, to emphasize the fact that these eigenvalues represent some energies of the physical system. Indeed, the states $|E_i\rangle$ are usually called *energy eigenstates* or *stationary states*, with associated energy E_i.

The lowest energy is known as *ground state energy*, and its associated eigenstate is known as the *ground state*, a fundamental element in the theory of quantum systems. Moreover, when the difference between the two smallest eigenvalues is strictly positive, this difference is called *spectral gap*, and we say in that case that the system is *gapped*. Determining whether a physical system has or not a spectral gap is a really important problem in Quantum Physics.

The states $|E_i\rangle$ mentioned above are called stationary because their only change in time is of the form

$$|E_i\rangle \mapsto \exp(-iE_i t/\hbar)\,|E_i\rangle.$$

Let us see now the connection between the two formulations for this postulate. If we consider the Schrödinger equation, we can see:

$$|\varphi(t_2)\rangle = \exp\left[\frac{-iH(t_2-t_1)}{\hbar}\right]|\varphi(t_1)\rangle = U(t_1,t_2)\,|\varphi(t_1)\rangle,$$

where we are defining:

$$U(t_1,t_2) := \exp\left[\frac{-iH(t_2-t_1)}{\hbar}\right].$$

This operation is easily seen to be unitary, and, furthermore, one can see that any unitary operator U can be written in the form

$$U = \exp(iK),$$

for some Hermitian operator K.

Postulate. 3
Given a physical system, with associated Hilbert space $\mathcal{H}$, the quantum measurements over such system are described by a collection $\{M_n\}_n \subset \mathcal{B}(\mathcal{H})$ of measurements as defined in Definition 1.4.1.

More specificaly, the index n refers to the measurement outcomes that may occur in the experiment, and given a state of a quantum system $|\varphi\rangle$ before a measurement, the probability that result $|n\rangle$ occurs is given by

$$p(n) = \langle\varphi|M_n^* M_n|\varphi\rangle$$

and the state of the system after the measurement is given by

$$\frac{M_n\,|\varphi\rangle}{\sqrt{p(n)}}.$$

Finally, measurement operators satisfy:

$$\sum_n M_n^* M_n = \mathbb{1}.$$

Finally, the fourth postulate can be stated as follows.

Postulate. 4
Given a composite physical system, its state space is also composite, and corresponds to the tensor product of the state spaces of the component physical systems. Moreover, if each system i is prepared in the state $|\varphi_i\rangle$, then the composite system is in the state $|\varphi_1\rangle \otimes \ldots \otimes |\varphi_n\rangle$.

After introducing the fourth postulate, it is necessary to make the following remark, which leads to introducing the concept of *entanglement*. Consider two Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$. Since these two Hilbert spaces have inner products (resp. $\langle\cdot,\cdot\rangle_1$ and $\langle\cdot,\cdot\rangle_2$), it is a natural question whether one can introduce an inner product, and therefore a topology, on the tensor product that arise naturally from those of the factors. This can be done by defining the inner product as:

$$\langle \varphi_1 \otimes \varphi_2, \psi_1 \otimes \psi_2 \rangle = \langle \varphi_1, \psi_1 \rangle_1 \langle \varphi_2, \psi_2 \rangle_2$$

for every $\varphi_1, \psi_1 \in \mathcal{H}_1$ and $\varphi_2, \psi_2 \in \mathcal{H}_2$, and extending by linearity. Finally, we take the completion under this inner product, and we get as the resulting Hilbert space the tensor product of $\mathcal{H}_1$ and $\mathcal{H}_2$. This can be generalized to the tensor product of n Hilbert spaces.

Now, a composite Hilbert space, i.e., a Hilbert space of the form $\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \ldots \otimes \mathcal{H}_n$, contains elements which are not tensor products of elements of each one of the components. In other words, if $|\varphi\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \ldots \otimes \mathcal{H}_n$, there not exist, in general, $|\varphi_i\rangle \in \mathcal{H}_i$ for all i so that

$$|\varphi\rangle = |\varphi_1\rangle \otimes |\varphi_2\rangle \otimes \ldots \otimes |\varphi_n\rangle.$$

A standard example of a non trivial two qubit state is the *EPR pair*[7], or *Bell state* is the following state:

$$|\phi^+\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}.$$

This structure of tensor products leads to the definition of *quantum entanglement*, a behavior that seems to be at the root of many of the most surprising phenomena in quantum mechanics.

Definition. 1.4.7 (Entanglement)
Given a state $|\varphi\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \ldots \otimes \mathcal{H}_n$, we say that $|\varphi\rangle$ is **entangled** if it cannot be written as an elementary tensor product of the form

$$|\varphi_i\rangle \otimes \ldots \otimes |\varphi_n\rangle \tag{1.4.20}$$

Notice that, in particular, one needs to have more than one system to talk about entangled states.

1.4.2.2 Schrödinger Picture

In the Schrödinger picture, we consider density matrices instead of states.

Postulate. 1
Given an isolated physical system, there is a complex Hilbert space $\mathcal{H}$ which is known as the state space of the system. This system is completely described by its **density operator**, which is a Hermitian, positive semidefinite and trace one operator $\rho \in \mathcal{S}(\mathcal{H})$.

Moreover, if we know the probability of the system in every state (for each state ρ_i, the probability that the system is in that state is p_i), then the state ρ can be written as

$$\sum_i p_i \rho_i.$$

Since the Heisenberg and Schrödinger picture are duals, there is an identification between observables in the Heisenberg picture and density matrices in the Schrödinger one. This leads to directly calling by *states* the density matrices, in a slight abuse of notation. With this notation, we denote by *pure states* the density matrices of the form

$$\rho = |\varphi\rangle \langle\varphi|$$

and by *mixed states* the ones of the form

$$\rho = \sum_i p_i |\varphi_i\rangle \langle\varphi_i|.$$

For the second postulate, concerning evolution of systems, we have the following formulation.

Postulate. 2
Given an isolated physical system, with associated Hilbert space $\mathcal{H}$, its evolution is described by a **unitary transformation**. More specifically, if the state of the system t_1 is described by the density matrix ρ_1 and the state of the system at instant $t_2 > t_1$ is described by ρ_2, then there exist a unitary operator U, which depends only on t_1 and t_2, such that

$$\rho_2 = U\rho_1 U^*.$$

As in the Heisenberg picture, the evolution of a density matrix is given by a unitary. To explain the form of the statement of the second postulate, consider

$$\rho = \sum_i p_i \left|\varphi_i\right\rangle \left\langle\varphi_i\right|.$$

Notice that, since the system initially is in the state $|\varphi_i\rangle$ with probability p_i, then after the evolution given by a unitary U it will be in state $U|\varphi_i\rangle$ with probability p_i. Therefore, the associated density operator will be given by

$$\sum_i p_i U \left|\varphi_i\right\rangle \left\langle\varphi_i\right| U^* = U\left(\sum_i p_i \left|\varphi_i\right\rangle \left\langle\varphi_i\right|\right) U^* = U\rho U^*.$$

Moving to the third postulate and the relation with quantum measurements, we have the following formulation for it.

Postulate. 3
Given an isolated physical system, with associated Hilbert space $\mathcal{H}$, any quantum measurements on it are described by a collection of measurement operators $\{M_n\}_n$ as the ones described in Definition 1.4.1. As in the case of the Heisenberg picture, each index n refers to the different outcomes that may occur when measuring. Indeed, If the state of the quantum system is ρ before the measurement, the probability that we get result n is given by

$$p(n) = \mathrm{Tr}(M_n^* M_n \rho),$$

and the state that we get after the measurement is given by:

$$\frac{M_n \rho M_n^*}{p(n)}.$$

Moreover, since probabilities need to sum one, these operators have to satisfy

$$\sum_n M_n^* M_n = \mathbb{1}.$$

Suppose that we measure with the measurement $\{M_n\}_n$ a mixed state of the form

$$\rho = \sum_i p_i \left|\varphi_i\right\rangle \left\langle\varphi_i\right|.$$

Then, if the initial state is $|\varphi_i\rangle$, for instance, the probability of having outcome n is

$$p(n|i) = \langle\varphi_i|M_n^* M_n|\varphi_i\rangle = \mathrm{Tr}(M_n^* M_n \left|\varphi_i\right\rangle \left\langle\varphi_i\right|).$$

Hence, the total probability of this outcome is

$$p(n) = \sum_i p(n|i)p_i = \sum_i p_i \,\mathrm{Tr}(M_n^* M_n \left|\varphi_i\right\rangle \left\langle\varphi_i\right|) = \mathrm{Tr}(\rho M_n^* M_n),$$

because of the definition of ρ. And analogously, one can see that the post-measurement state is given by:

$$\frac{M_n \rho M_n^*}{p(n)}.$$

Finally, concerning the state space of a composite physical system, we get the following postulate, due to the linearity of tensor products.

Postulate. 4
Given a composite physical system, its state space is the tensor product of the state spaces of the component physical systems.

Moreover, if each system i is initially prepared in state ρ_i, then the state in which the composite system is prepared is given as the tensor product of the ρ_i, i.e., $\rho_1 \otimes \rho_2 \otimes \ldots \otimes \rho_n$.

These reformulations of the postulates of quantum mechanics in terms of the density operator are, clearly, mathematically equivalent to the description in terms of the state vector. However, as a way of thinking about quantum mechanics, the density operator approach has advantages with respect to two main facts: the description of quantum systems whose state is not known, and the description of subsystems of a composite quantum system.

1.4.3 Quantum circuits

In this subsection, first, we present a brief survey on classical Boolean circuits, and, then, we introduce some notions of quantum circuits, by outlining the difference with respect to the latter ones.

1.4.3.1 Classical circuits

A classical circuit is used to represent functions from $\{0,1\}^n$ to $\{0,1\}$. It is a computational model that consists of decomposing each function in some elemental operations, so that this procedure allows us to represent all the possible functions in the domain. This model has good properties in general and is fundamental in computational theory.

In classical complexity theory, we can define a Boolean circuit more formally as follows.

Definition. 1.4.8 (Boolean circuit)
A Boolean circuit is a finite directed acyclic graph composed of AND, OR an NOT gates (see Figure 1.6).

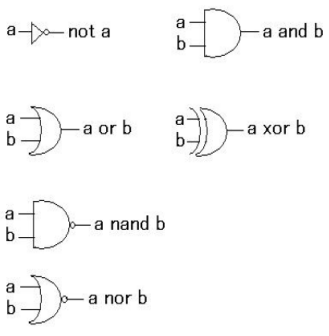


fig. 1.6: Some classical gates for two qubits. AND, OR and NOT are used to construct the rest.

An important theoretical result is that any function on bits can be computed from the composition of NAND gates alone, which is thus known as a *universal gate*. By contrast, the XOR alone or even with NOT is not universal (one can notice that just by taking a look at the parity).

The idea of classical circuits lays on the following facts:

- Every circuit has n input nodes, which contain n input bits.
- The circuit is made of those three gates (AND, OR and NOT), and combinations of then, as well as some output nodes.
- The initial input bits are fed into combinations of the previous gates, so that eventually the output nodes assume some value.

Let $f : \{0,1\}^n \to \{0,1\}^m$ be a Boolean function. Then, we say that a circuit *computes* it if the output nodes get the right value $f(x)$ for every $x \in \{0,1\}^n$.

Now we can introduce some concepts related to the complexity of some circuits. Let us denote a *circuit family* by a set $\mathcal{C} = \{C_n\}$, each one of them of *input size* n (which means that the number of input nodes, and hence bits, is exactly n). We assume that each one of these circuits has one output bit. Then, we say that this family *recognizes* a certain *language* $L \subseteq \bigcup_{n\geq 0} \{0,1\}^n$ (which we denote hereafter by $\{0,1\}^*$) if, for every $x \in \{0,1\}^n$, the circuit C_n outputs:

- 1 if $x \in L$.
- 0 if $x \notin L$.

1.4.3.2 Quantum gates

Let us move now to quantum circuits, which generalize the idea of classical circuit families. In this case, we replace the AND, OR and NOT gates by elementary *quantum gates*. We define a quantum gate as a unitary transformation in a small number of qubits, usually 1, 2 or 3. The following are the most important gates 1-qubit gates:

1. **Bitflip gate**: It negates the bit, i.e., swaps $|0\rangle$ and $|1\rangle$. It can be represented by:

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{1.4.21}$$

2. **Phaseflip gate**: It puts a - in front of $|1\rangle$. It can be represented by:

$$Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{1.4.22}$$

3. **Phase gate**: It rotates the phase of the $|1\rangle$-state by an angle θ:

$$R_\theta = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\theta} \end{pmatrix} \tag{1.4.23}$$

4. **Hadamard gate**: It is specified by:

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \tag{1.4.24}$$

The last one, the Hadamard gate, is possibly the most important 1-qubit gate. If we apply H to an initial state $|0\rangle$ and then measure, we have the same probability of observing $|0\rangle$ or $|1\rangle$, and analogously if we apply it to initial $|1\rangle$. However, when applied to the superposition state

$$\frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle,$$

the Hadamard gate provides the value $|0\rangle$. The effect that we get in this case (both positive and negative amplitudes for $|1\rangle$ cancelling out) is called *interference*. It is completely analogous to the interference patterns that one can notice in light or sound waves.

We can further define gates that act on 2 qubits:

5. **CNOT (Controlled not)**: Given two input bits, this gate is used to negate the second bit if the first one is 1, and to leave it invariant if the first bit is 0. It can be represented by

$$\text{CNOT} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{1.4.25}$$

In this scenario, the first qubit is called the *control* qubit, since it is the one that determines the effect of the gate, and the second one is called the *target* qubit, as it is the one that receives the effect.

In general, if U is a 1-qubit gate (as the ones that we have defined above), then the we can define the 2-qubit controlled-U gate analogously to the previous one, i.e., if the first bit is 0 it does nothing, and, if it is 1, the gate applies the unitary to the second bit. We can represent it in the following matrix form:

6. **Controlled-U gate**: If the first bit is 0 it does nothing, and, if it is 1, the gate applies the unitary to the second bit. It is given by

$$C_U = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & U_{11} & U_{12} \\ 0 & 0 & U_{21} & U_{22} \end{pmatrix} \tag{1.4.26}$$

Another way to understand the quantum CNOT gate is as a generalization of the classical XOR gate. Note, however, that there are some classical gates, like NAND or XOR, which cannot be understood as unitary gates in a sense similar to the way the quantum NOT gate represents the classical NOT gate. The reason is because these two gates are essentially irreversible.

We can see that in the following example: Given an output $A \oplus B$ of a XOR gate, it is not possible to determine what the inputs A and B were. This can be also stated by saying that there is a loss of information associated with the irreversible action of the XOR gate. On the other hand, since quantum gates are described by unitary matrices, it is important to remark that they can always be inverted by another quantum gate.

Further we name the following 3-qubit gate, which is particularly interesting as it is classically universal. This means every classical computation can be implemented by a sequence of Toffoli gates.

7. **Toffoli gate or CCNOT (Controlled-Controlled-NOT gate)**: It negates the third bit of its input if both the first two bits are 1.

All those gates mentioned above can be composed into bigger unitary operations in the following ways:

- By taking *tensor products*, if the gates are applied *in parallel*.
- By taking *matrix products*, if the gates are applied *sequentially*.

We show now an example of these operations. If we apply a Hadamard gate H to each bit in a register of n zeros, we get

$$\frac{1}{\sqrt{2^n}} \sum_{j \in \{0,1\}^n} |j\rangle,$$

a superposition of all n-bit strings, whereas applying $H^{\otimes n}$ to an initial state $|i\rangle$, with $i \in \{0,1\}^n$ gives us

$$H^{\otimes n} |i\rangle = \frac{1}{\sqrt{2^n}} \sum_{j \in \{0,1\}^n} (-1)^{i \cdot j} |j\rangle,$$

with $i \cdot j = \sum_{k=1}^{n} i_k j_k$ the inner product of the n-bit strings $i, j \in \{0,1\}^n$. In this case, one can also notice that the Hadamard is its own inverse. Thus, if we apply it again on the right-hand side of the previous expression, we get the initial $|i\rangle$. This makes the Hadamard gate quite useful for the development of algorithms, as we will see in the following section.

To sum up, as in the case for classical circuits, one can define a quantum circuit in the following form.

Definition. 1.4.9 (Quantum circuit)
A *quantum circuit* is a finite directed acyclic graph composed by:

- **Input nodes.** Some of these nodes (n nodes) contain the input, and some more nodes are initially $|0\rangle$ (they are called the *workspace*).
- **Quantum gates.** Each of them operates on, at most, two or three qubits of the state.
- **Output nodes.** The previous gates transform the initial state vector into a final state, which will generally be a superposition.

Let us see now how one can draw these circuits. We usually consider that time progresses from left to right. As briefly mentioned above, each qubit is represented as a wire, and the circuit prescribes which gates are applied to each wire.

With this notation of wires, it is clear that 1-qubit gates act on just one wire, whereas 2-qubit and 3-qubit gates act, respectively, on 2 or 3 wires. Moreover, when a gate acts on more than 1 qubit, and one of them is the *control* one, its wire is drawn with a dot linked vertically to the *target* qubits, i.e., the qubits where this effect is applied.

We show an example of this notation in the following figure.

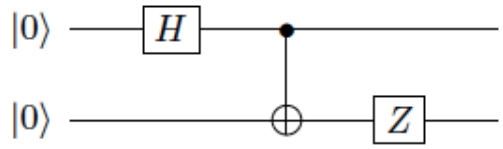


fig. 1.7: Circuit used to turn $|00\rangle$ into $\frac{1}{\sqrt{2}}(|00\rangle - |11\rangle)$.

In this example, and in general, we denote the quantum CNOT by $\oplus$. If we study every step separately, and taking into account the definition for every gate mentioned above, we can see that, after each step, the resulting state is:

- **Step 0.** We start with $|\varphi_0\rangle = |00\rangle$.

- **Step 1.** After the Hadarmard gate, we have $|\varphi_1\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |10\rangle)$.
- **Step 2.** When we apply the CNOT gate, we get $|\varphi_2\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$.
- **Step 3.** Finally, after the Z gate, we have $|\varphi_3\rangle = \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle)$.

1.4.4 No-Cloning theorem

In this subsection, the question that we want to pose is: "Can we clone a classical/quantum bit?".

In the classical setting the answer is clearly yes. The cloning can be done just by the circuit shown in Figure 1.8. It is a trivial application of the classical CNOT gate.

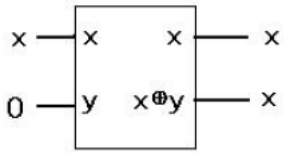


fig. 1.8: Classical way to clone a bit.

The interpretation of this circuit is the following. Let us assume that we start with the bit x that we want to clone, and we take it as a *control* bit, as well as the 0 bit, which we take as the *target* bit. Then, applying a classical CNOT gate, one automatically gets an output given by two copies of x.

In the quantum setting the answer is no! In principle, one could be tempted to think that a similar argument as in the classical case can follow using instead a quantum CNOT gate. However, it does not work [26], as we show below.

To frame our question more mathematically, we phrase it in the following terms: Consider a quantum machine with two input qubits labeled by A and B. The first one is the *control* qubit, denoted by $|psi\rangle$, and the second one the *target* qubit, initially $|\phi\rangle$. Hence, the initial state is given by

$$|\psi\rangle \otimes |\phi\rangle.$$

Assume now that our quantum machine facilitates cloning. Then this machine fulfills the properties: There exists a unitary U such that

- For a given $|\psi\rangle \otimes |\phi\rangle$, we have

$$|\psi\rangle \otimes |\phi\rangle \xrightarrow{U} U(|\psi\rangle \otimes |\phi\rangle) = |\psi\rangle \otimes |\psi\rangle \tag{1.4.27}$$

- For another state $|\varphi\rangle \otimes |\phi\rangle$, we find again

$$|\varphi\rangle \otimes |\phi\rangle \xrightarrow{U} U(|\varphi\rangle \otimes |\phi\rangle) = |\varphi\rangle \otimes |\varphi\rangle \tag{1.4.28}$$

From this it immediately follows that

$$\langle\varphi|\psi\rangle = \lambda = \lambda^2 = \langle\varphi|\psi\rangle^2. \tag{1.4.29}$$

This implies that

$$\lambda = \begin{cases} 0 & \text{States are orthogonal} \\ 1 & \text{States are the same} \end{cases}. \tag{1.4.30}$$

Hence, we can clone only classical information embedded into a quantum system, making no quantum cloning device possible in general. Even if we consider U not to be a unitary (as we implicitly did), there is no general cloning device.

1.4.5 Quantum teleportation

We present now another example of how the gates of the previous subsections can be used to get results of relevance. In this particular case, we will explain *quantum teleportation* [3] as a combination of some elementary gates from the ones above.

Quantum teleportation is one of the most representative communication protocols of quantum information theory. The protocol is schematically described in Figure 1.9. Suppose there are two parties, Alice and Bob, which are spatially separated, and Alice wants to send a qubit of information to Bob by just sending two classical bits via a classical channel. For that, they first met and shared an EPR (Einstein-Podolsky-Rosen) state, given by:

$$\text{EPR} = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) =: |\varphi\rangle\,. \tag{1.4.31}$$

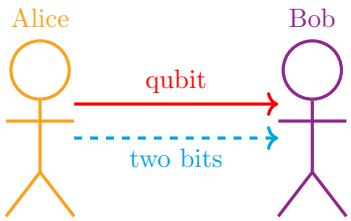


fig. 1.9: Schematic representation of the protocol of Quantum teleportation

In a sense, this protocol means that to send one qubit, one needs one EPR state and two classical bits that have to be exchanged. They jointly carry more information then one qubit. The fact that a qubit can be sent by means of this procedure is usually expressed by writing:

$$\boxed{1\text{ EPR} + 2\text{ bits} \geq 1\text{ qubit}}$$

It is important to remark that Alice does not need to know her own qubit in order to send it to Bob. Let us see how this works. Alice has a quantum state of the form

$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle \tag{1.4.32}$$

that she wants to transfer. We can assume that Alice does not know the values of α and β. We can then split the procedure in the following steps:

1. The combined initial system is

$$\begin{aligned} |\varphi_0\rangle_{AA'B'} &= |\psi\rangle_A \otimes |\varphi\rangle_{A'B'} = (\alpha\,|0\rangle_A + \beta\,|1\rangle_A) \otimes \left(\frac{1}{\sqrt{2}}(|00\rangle_{A'B'} + |11\rangle_{A'B'})\right) \\ &= \frac{1}{\sqrt{2}}\big(\alpha(|000\rangle_{AA'B'} + |011\rangle_{AA'B'}) + \beta(|100\rangle_{AA'B'} + |111\rangle_{AA'B'})\big), \end{aligned} \tag{1.4.33}$$

   where we are writing subindices in the last line to outline whom each bit belongs to.

2. Alice now applies now a CNOT gate to this state to her part, i.e. the first two qubits:

$$\begin{aligned} |\varphi_1\rangle &= \text{CNOT}_{AA'}\,|\varphi_0\rangle \\ &= \frac{1}{\sqrt{2}}\big[\alpha(|000\rangle_{AA'B'} + |011\rangle_{AA'B'}) + \beta(|110\rangle_{AA'B'} + |101\rangle_{AA'B'}\big] \end{aligned} \tag{1.4.34}$$

3. Alice applies a Hadamard gate on her first qubit, obtaining

$$\begin{aligned}|\varphi_2\rangle &= H_A \otimes \mathbb{1}_{A'B'} |\varphi_1\rangle \\ &= \frac{1}{\sqrt{2}} \left[\alpha \frac{|0\rangle_A + |1\rangle_A}{\sqrt{2}} \otimes (|00\rangle_{A'B'} + |11\rangle_{A'B'}) + \beta \frac{|0\rangle_A - |1\rangle_A}{\sqrt{2}} (|00\rangle_{A'B'} + |11\rangle_{A'B'} \right] \\ &= \frac{1}{2} \left[|00\rangle_{AA'} \otimes (\alpha |0\rangle + \beta |1\rangle) + |10\rangle_{AA'} \otimes (\alpha |0\rangle_{B'} - \beta |1\rangle_{B'}) \right. \\ &\qquad \left. + |01\rangle_{AA'} \otimes (\alpha |1\rangle_{B'} + \beta |0\rangle_{B'}) + |11\rangle_{AA'} \otimes (\alpha |1\rangle_{B'} - \beta |0\rangle_{B'} \right]\end{aligned} \tag{1.4.35}$$

with

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \tag{1.4.36}$$

We have written the state in this way, because of the considerations in the next step.

4. This last expression has four different terms, and each one of them can be seen as the product one of the elements of the 2-qubit computational basis for Alice and another qubit (in four different forms) for Bob. Thus, if Alice measures her pair in the computational basis, i.e.,

$$\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\},$$

she gets one of the previous four elements and, thus, she can read off Bob's postmeasurement from this value, given her measurement, in the following way:

$$\begin{cases} |00\rangle_A \mapsto |\varphi_3(00)\rangle_B := \alpha |0\rangle_B + \beta |1\rangle_B \\ |01\rangle_A \mapsto |\varphi_3(01)\rangle_B := \alpha |1\rangle_B + \beta |0\rangle_B \\ |10\rangle_A \mapsto |\varphi_3(10)\rangle_B := \alpha |0\rangle_B - \beta |1\rangle_B \\ |11\rangle_A \mapsto |\varphi_3(11)\rangle_B := \alpha |1\rangle_B - \beta |0\rangle_B . \end{cases}$$

Hence, the way of proceeding is the following: Alice measures her two bits in the computational basis, sends the result to Bob over a classical channel, and Bob knows which transformation he must do on his qubit to regain the desired qubit $|\varphi\rangle$. In each case, Bob has to do the following:

Alice sends	**Bob receives**	**Bob does (to obtain $\|\psi\rangle$)**
$\|00\rangle \to \{0,0\}$	$\|\psi\rangle$	Nothing
$\|10\rangle \to \{1,0\}$	$\alpha \|0\rangle - \beta \|1\rangle$	Applies a Z gate
$\|01\rangle \to \{0,1\}$	$\alpha \|1\rangle + \beta \|0\rangle$	Applies a X gate (NOT gate)
$\|11\rangle \to \{1,1\}$	$\alpha \|1\rangle - \beta \|0\rangle$	Applies a X and Z gate

It is noteworthy that Alice has to communicate through a classical channel to send her measurement, thus there is no instant transfer of information. More specifically, quantum teleportation does not allow for faster than light communication, as, to complete the protocol, in the third step above, Alice must transmit her measurement's result to Bob over a classical communication channel.

We also need to remark that we are not creating a copy of $|\psi\rangle$ being teleported. Alice destroys her state in the process of measuring, so this protocol is also no contradiction to the no-cloning theorem.

1.4.6 Superdense coding

Now we present a third example of use of the previous quantum gates for a certain communication protocol. More specifically, in this subsection we discuss superdense coding [2], which is a communication protocol that allows to transmit 2 classical bits of information by just sending 1 qubit, assuming that Alice and Bob share an EPR pair. Analogously to the previous example, we can express the fact that two bits can be sent by means of this procedure by writing:

$$\boxed{1 \text{ EPR} + 1 \text{ qubit} \geq 2 \text{ bits}}$$

The protocol in this case can be schematically represented as in Figure 1.10:

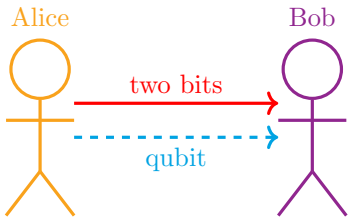


fig. 1.10: Superdense coding.

1. First, Alice and Bob share an EPR state:

$$|\varphi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle). \tag{1.4.37}$$

2. Next, let us denote by $\{x, y\}$ the classical bits Alice wants to send. Alice performs the following operations to the first qubit of $|\varphi\rangle$ (which is in her possession) and then sends the state she obtains:

Alice wants to send	**Alice does to $\|\varphi\rangle$**	**She gets**
$\{0,0\}$	Nothing	$\|\varphi_1\rangle = \frac{1}{\sqrt{2}}(\|00\rangle + \|11\rangle)$
$\{1,0\}$	Applies a Z gate	$\|\varphi_1\rangle = \frac{1}{\sqrt{2}}(\|00\rangle - \|11\rangle)$
$\{0,1\}$	Applies a X gate	$\|\varphi_1\rangle = \frac{1}{\sqrt{2}}(\|10\rangle + \|01\rangle)$
$\{1,1\}$	Applies a iY gate (or X and Z)	$\|\varphi_1\rangle = \frac{1}{\sqrt{2}}(\|01\rangle - \|10\rangle)$

   Since these four last elements for the *Bell basis* of $\mathbb{C}^2 \otimes \mathbb{C}^2$, each state can be perfectly distinguished by an appropriate quantum measurement.

3. Alice sends her part of the state to Bob, so that he is now in possession of the whole state. By measurement the whole state in the Bell basis, Bob can then determine which of the four possible bit strings Alice sent him.

1.4.6.1 Universality of quantum gates

In this section we want to discuss the universality of certain sets of elementary quantum gates. In the classical setting, we have that AND and NOT are universal, in the sense that any classical Boolean circuit can be implemented just using AND and NOT gates. We can also show that the Toffoli gates are universal for classical computation by reducing the Toffoli gate to an AND and NOT gate respectively:

$$\text{Tofolli} = \begin{cases} \text{AND} & \text{Fix the third input to 0,} \\ \text{NOT} & \text{Fix the first and second input to 1.} \end{cases} \tag{1.4.38}$$

Hence, if we apply Toffoli gates, we can implement any classical computation in a reversible manner.

Now, if we move to the quantum case, there are also several possibilities for universal sets of elementary gates. Let us mention here some examples:

- **All 1-qubit operations + 2-qubit CNOT.** This set is universal, in the sense that any other unitary transformation can be built from these gates.

  However, it is difficult to consider this set, as 'all' possible 1-qubit gates are difficult to be described (there are continuously many of them). Also, we cannot expect that experimentalists can implement these gates with infinite precision. Hence, the practical model that is usually considered allows just a small finite set of 1-qubit gates from which the rest can be efficiently approximated.

- **CNOT, Hadamard and $R_{\pi/4}$.** It is universal concerning approximation. It means that any other unitary can be arbitrarily well approximated using circuits, and it is a consequence of the *Solovay-Kitaev theorem* [19, 10].

  If we restrict to real numbers, then we also have the following set:

- **Hadamard and Toffoli.** This set is universal for all unitaries with real entries, again in the sense of approximation.

1.4.6.2 Some further basic definitions

Before concluding this chapter, we need to introduce some further notions and basic results which will be of use for the exercises of the next lectures.

Let us start by introducing the partial trace of a density operator. For this discussion, let $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$ be a composite system and let

$$\mathcal{S}(\mathcal{H}) := \{\rho \in \mathcal{B}(\mathcal{H}) \ : \ \rho = \rho^*, \ \rho \geq 0, \ \mathrm{Tr}[\rho] = 1\} \tag{1.4.39}$$

be the set of quantum states (or density matrices) on $\mathcal{H}_{AB}$.

Definition. 1.4.10 (Partial trace)
The partial trace Tr_B is a linear, trace preserving completely positive map which is defined over basis product states of the composite space by

$$\mathrm{Tr}_B : \mathcal{S}(\mathcal{H}_{AB}) \to \mathcal{S}(\mathcal{H}_A), \qquad |a_1\rangle\langle a_2| \otimes |b_1\rangle\langle b_2| \mapsto \mathrm{Tr}[|b_1\rangle\langle b_2|]|a_1\rangle\langle a_2|. \tag{1.4.40}$$

This notion extends to other density matrix by linearity. In particular, for a product state $\rho = \rho_A \otimes \rho_B$, we get that

$$\mathrm{Tr}_B[\rho] = \mathrm{Tr}_B[\rho_A \otimes \rho_B] = \mathrm{Tr}[\rho_B]\rho_A = \rho_A. \tag{1.4.41}$$

We want to highlight the following interesting property. Let $M_A \in \mathcal{B}(\mathcal{H}_A)$ and $M_A \otimes \mathbb{1}_B$, then

$$\mathrm{Tr}[M_A \rho_A] = \mathrm{Tr}[M_A \otimes \mathbb{1}_B \, \rho_{AB}] \tag{1.4.42}$$

in the case that $\rho_{AB} \in \mathcal{S}(\mathcal{H}_{AB})$.

We will also need for the exercises to be able to decompose pure states into basis products of a composite Hilbert space. We use the Schmidt decomposition for that purpose.

Theorem. 1.4.11 (Schmidt decomposition) *Assume that $|\psi\rangle$ is a pure state in a composite system $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$. Then, there exist orthonormal states $\{|i_A\rangle\}$ in $\mathcal{H}_A$ and $\{|i_B\rangle\}$ in $\mathcal{H}_B$ such that*

$$|\psi\rangle = \sum_i \lambda_i \, |i_A\rangle \otimes |i_B\rangle \, , \qquad \textit{with} \quad \lambda_i \geq 0, \, \sum_i \lambda_i^2 = 1. \tag{1.4.43}$$

The λ_i are called Schmidt coefficients and the number of λ_i in the decomposition, with multiplicity, is the Schmidt rank of the state.

Consequence. 1.4.11.1 (Purification) *Consider a state $\rho_A \in \mathcal{S}(\mathcal{H}_A)$. We can introduce an auxiliary system $\mathcal{H}_R$, and construct a pure state $|\psi_{AR}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_R$ such that*

$$\rho_A = \mathrm{Tr}_R[|\psi_{AR}\rangle\langle\psi_{AR}|]. \tag{1.4.44}$$

Hence the theorem allows us to make a connection between mixed states and pure states in a larger Hilbert space.

Chapter 2

Quantum Nonlocality

2.1 Quantum Nonlocality

In the past section we introduced the formalism of quantum mechanics. This formulation, however, was widely questioned since its origins in the physics as well as in the mathematical worlds. The only thing the scientific community could agree on was that the theory is a useful description of physical laws and allows to predict them in a very precise way. However, some important scientists showed some skepticism about the nondeterministic nature of the theory.

Throughout the years there has been a major discussion on this topic, in particular about the uncertainty part of quantum mechanics and its mathematical formalisation. Alternative theories have been proposed which incorporated the uncertainty principle that is generally believed to be intrinsic to nature. The most relevant models under this framework are called "Local Hidden Variable Models". The idea behind these models is that there exists a hidden probability over all possible states in the world that we cannot know. However, once one of these states is fixed, we are in a completely deterministic situation.

More specifically, in the first Chapter, when we discussed the Postulates of Quantum Mechanics, we said that the vector state contains all the information that we can obtain about the system. This, in particular, implies that the impossibility to obtain more accurate information about a physical system does not depend on our precision, but is something intrinsic of Nature. On the other hand, the main idea behind the Hidden Variable Models is that such an ignorance about Nature is due to our own restrictions. According to these theories, there exists a hidden probability over all the possible states of the world that we cannot know. However, once one of these states is fixed, we are in a completely deterministic situation.

This can be rephrased and summarised as follows:

> Uncertainty in Nature can be understood as a classical average over deterministic states.

These discussions about completeness of quantum mechanics and alternative theories started in 1935 with a paper by Einstein, Podolsky and Rosen, who proposed and experiment to "prove" the incompleteness of quantum mechanics as a model of Nature. It took almost 30 years until Bell understood that the EPR paradox could be reformulated in terms of certain assumptions which naturally lead to a refutable prediction. In particular, Bell showed that the assumption of a local hidden variable model implies some inequalities on the set of correlations obtained in the scenario of a certain measurement (called *Bell inequalities*). Bell inequalities are violated by certain quantum correlations produced with an entangled state.

In summary, *quantum nonlocality* can be identified with the violations of Bell inequalities. These considerations have strong implications for quantum information, as they are key for some of its branches, such as the security of quantum cryptography protocols as well as the quantum advantage in communication complexity and information theoretical protocols. In the next few pages, we will discuss the notion of quantum nonlocality, associated to the violations of Bell inequalities. Subsequently, we will provide a brief review on nonlocal games and sets of correlations. We will conclude by mentioning some very recent results on the topic. The content of this chapter has been mainly extracted from [12] and [22].

2.1.1 Correlation in EPR Bell's result

Let us consider the following experiment: We have Alice, Bob and Charlie and the latter send one particle to each of the former. Alice can measure two properties of this particle, named A_1 and A_2, which output as a value $+1$ or -1. Analogously, Bob can measure two properties B_1 and B_2 on his particle, outputting also $+1$ or -1. This setup is schematically represented in Figure 2.1.

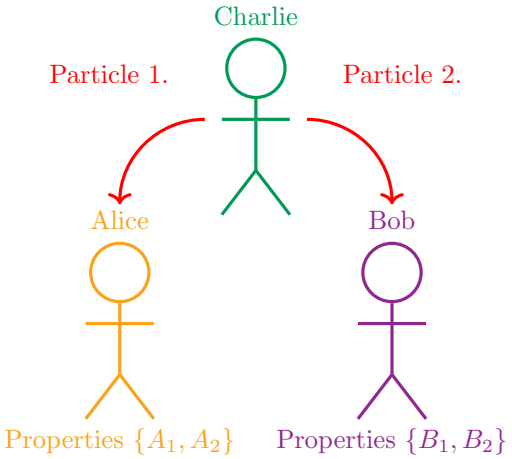


fig. 2.1: EPR setup

We now want to better understand the the EPR paradox and the reformulation of it by Bell. For that, this experiment has the following conditions:

- We consider measurements in a disconnected manner, i.e. simultaneous measurements that therefore cannot influence each other.
- We repeat the experiment as many times as possible.

We consider the following combination of the outcomes

$$A_1B_1 + A_1B_2 + A_2B_1 - A_2B_2 = (A_1 + A_2)B_1 + (A_1 - A_2)B_2\,. \tag{2.1.1}$$

It is clear that, in this expression, either $A_1 + A_2$ or $A_1 - A_2$ takes the value 0. Therefore,

$$A_1B_1 + A_1B_2 + A_2B_1 - A_2B_2 = \pm 2 \tag{2.1.2}$$

- Let us assume that we are in the setting of a **local hidden variable model**. Note that the locality condition allows us to perform measurements in a disconnected manner, whereas the hidden variable model provides a hidden probability on the space of all possible deterministic states of the world. Let us denote by

$$P(a_1, a_2, b_1, b_2) = P(A_1 = a_1, A_2 = a_2, B_1 = b_1, B_2 = b_2) \tag{2.1.3}$$

the hidden probability. Then the CHSH (Clauser-Horn-Shimony-Holt) inequality states:

$$\begin{aligned} &|\,\mathbb{E}[A_1B_1 + A_1B_2 + A_2B_1 - A_2B_2]| \\ &\qquad = \Big| \sum_{a_1,a_2,b_1,b_2} p(a_1,a_2,b_1,b_2)(a_1b_1 + a_1b_2 + a_2b_1 - a_2b_2) \Big| \leq 2 \end{aligned} \tag{2.1.4}$$

- Let us assume that we are in the setting of **Quantum Mechanics**. Consider the following state formed by the two particles sent by Charlie

$$|\varphi\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}}\,, \tag{2.1.5}$$

with the first qubit going to Alice, who measures it with $A_1 = X$ or with $A_2 = Z$, and the second one to Bob, who measures it with $B_1 = \frac{-Z-X}{\sqrt{2}}$ or $B_2 = \frac{Z-X}{\sqrt{2}}$. Then, however,

$$\langle\varphi|A_1B_1|\varphi\rangle = \langle\varphi|A_2B_1|\varphi\rangle = \langle\varphi|A_1B_2|\varphi\rangle = \frac{1}{\sqrt{2}} \tag{2.1.6}$$

and

$$\langle\varphi|A_2B_2|\varphi\rangle = -\frac{1}{\sqrt{2}} \tag{2.1.7}$$

giving us

$$\langle\varphi|A_1B_1 + A_1B_2 + A_2B_1 - A_2B_2|\varphi\rangle = 2\sqrt{2} > 2. \tag{2.1.8}$$

From this it immediately follows, that local hidden variable models cannot describe quantum mechanics. Or, in other words, certain correlations in the previous experiment cannot be explained by a local hidden variable model.

2.1.2 Tsirelson's Theorem

We can generalize the previous scenario to N measurements. For that, we define the correlation matrix as

$$\gamma_{ij} = \mathbb{E}[A_iB_j] \quad \forall i,j = 1,\ldots,N. \tag{2.1.9}$$

In the local hidden variable model this evaluates to

$$\gamma_{ij} = \int_\Omega A_i(\omega)B_j(\omega)\,d\mathbb{P}(\omega) \tag{2.1.10}$$

with $(\Omega,\mathbb{P})$ the hidden probability distribution. For each $\omega \in \Omega$ we have that $A_i(\omega) = \pm 1$, $B_j(\omega) = \pm 1$. In our finite context we find

$$\gamma_{ij} = \sum_k p(k)A_i(k)B_j(k). \tag{2.1.11}$$

Hence $\gamma = (\gamma_{i,j})_{i,j=1}^N$ can be written as a *classical correlation matrix*. We denote the set of classical correlation matrices of dimension $N \times N$ as $\mathcal{C}_{cl}(N)$.

In the quantum mechanical setting we describe a bipartite system with a quantum state $\rho \in \mathcal{S}(\mathbb{C}^{n\times n} \otimes \mathbb{C}^{n\times n})$. In this case, it holds

Output of Alice's measurement	A_i: POVM $\{\mathbb{E}_i, \mathbb{1} - \mathbb{E}_i\}$
Output of Bob's measurement	B_j: POVM $\{\mathbb{F}_j, \mathbb{1} - \mathbb{F}_j\}$

and we obtain

$$p(i',j') = P(A_i = i', B_j = j') = \begin{cases} \mathrm{Tr}[(\mathbb{E}_i \otimes \mathbb{F}_j)\rho] & \text{if } (i',j') = (1,1) \\ \mathrm{Tr}[((\mathbb{1} - \mathbb{E}_i) \otimes \mathbb{F}_j)\rho] & \text{if } (i',j') = (-1,1) \\ \mathrm{Tr}[(\mathbb{E}_i \otimes (\mathbb{1} - \mathbb{F}_j))\rho] & \text{if } (i',j') = (1,-1) \\ \mathrm{Tr}[((\mathbb{1} - \mathbb{E}_i) \otimes (\mathbb{1} - \mathbb{F}_j))\rho] & \text{if } (i',j') = (-1,-1) \end{cases} \tag{2.1.12}$$

Hence we get

$$\begin{aligned} \gamma_{ij} &= \mathbb{E}[A_i B_j] = [p(1,1) + p(-1,-1)] - [p(1,-1) + p(-1,1)] \\ &= \mathrm{Tr}[(\mathbb{E}_i \otimes \mathbb{F}_j + (\mathbb{1} - \mathbb{E}_i) \otimes (\mathbb{1} - \mathbb{F}_j) - \mathbb{E}_i \otimes (\mathbb{1} - \mathbb{F}_j) - (\mathbb{1} - \mathbb{E}_i) \otimes \mathbb{F}_j)\rho] \\ &= \mathrm{Tr}[((\mathbb{1} - 2\,\mathbb{E}_i) \otimes (\mathbb{1} - 2\,\mathbb{F}_j))\rho] \end{aligned} \tag{2.1.13}$$

Note that $\mathbb{1} - 2\,\mathbb{E}_i$ and $\mathbb{1} - 2\,\mathbb{F}_j$ are self-adjoint and we have $\|\mathbb{1} - 2\,\mathbb{E}_i\| \leq 1$ and $\|\mathbb{1} - 2\,\mathbb{F}_j\| \leq 1$. At this point it is noteworthy that every operator O with $\|O\| \leq 1$ can actually be written as $\mathbb{1} - 2\,\mathbb{E}$. This gives rise to the following definition:

Definition. 2.1.1 (Quantum correlation matrix)
We have that

$$\gamma = (\gamma_{i,j})_{i,j=1}^{N} \tag{2.1.14}$$

is a *quantum correlation matrix* if there exist self-adjoint operators $A_1, \ldots, A_N$, $B_1, \ldots, B_N$ acting on $\mathbb{C}^n$ such that $\max_{i,j=1,\ldots,N} \{\|A_i\|, \|B_j\|\} \leq 1$ and a state ρ acting on $\mathbb{C}^n \times \mathbb{C}^n$ such that

$$\gamma_{i,j} = \mathrm{Tr}[A_i \otimes B_j\, \rho]\,, \qquad \forall i,j = 1, \ldots, N\,.$$

We denote the set of quantum correlation matrices of size N with $\mathcal{C}_q(N)$.

Proposition. 2.1.2 *We have that $\mathcal{C}_{cl}(N) \subseteq \mathcal{C}_q(N)$.*

Proof. Consider $\gamma \in \mathcal{C}_{cl}(N)$. Then, for a fixed size N, we can always assume that the $\gamma_{i,j}$ are defined as:

$$\gamma_{i,j} = \sum_{k=1}^{K} p(k) A_i(k) B_j(k) \tag{2.1.15}$$

Consider now the matrices given by

$$A_i = \begin{pmatrix} A_i(1) & 0 & \ldots & & 0 \\ 0 & A_i(2) & 0 & \ldots & 0 \\ \vdots & 0 & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ 0 & 0 & \ldots & 0 & A_i(k) \end{pmatrix} \tag{2.1.16}$$

and

$$B_j = \begin{pmatrix} B_j(1) & 0 & \ldots & & 0 \\ 0 & B_j(2) & 0 & \ldots & 0 \\ \vdots & 0 & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ 0 & 0 & \ldots & 0 & B_j(k) \end{pmatrix}. \tag{2.1.17}$$

Take $\rho = \sum_{k=1}^{K} p(k)\, |kk\rangle\langle kk|$. Then, it clearly holds that

$$\mathrm{Tr}[(A_i \otimes B_j)] = \sum_{k=1}^{K} p(k) A_i(k) B_j(k) = \gamma_{i,j}\,, \quad \forall i,j\,. \tag{2.1.18}$$

Moreover, since A_i and B_j are clearly self-adjoint, we conclude that $\gamma \in \mathcal{C}_q(N)$. □

We notice the following properties of $\mathcal{C}_{cl}(N)$ and $\mathcal{C}_q(N)$:

- Both are convex sets.
- $\mathcal{C}_{cl}(N)$ is a polytope, and thus it has a finite number of extreme points. Its facets are usually called "correlation Bell inequalities".

Definition. 2.1.3 (Bell inequalities)
In the above framework the Bell inequalities are given by

$$\sum_{i,j=1}^{N} M_{ij}\gamma_{ij} \leq C\,, \qquad .\forall \gamma = (\gamma_{i,j})_{i,j=1}^{N} \in \mathcal{C}_{cl}(N)\,. \tag{2.1.19}$$

with $M = (M_{i,j})_{i,j=1}^{N}$ the coefficients of the corresponding inequality and C an independent term.

Example. (CHSH) In the case of the CHSH inequality,

$$M = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \tag{2.1.20}$$

and $C = 2$. Moreover, there exist some correlations $\hat{\gamma}_i j$ such that

$$\sum_{i,j=1}^{N} M_{i,j}\hat{\gamma}_{i,j} = 2\sqrt{2}. \tag{2.1.21}$$

This means $\hat{\gamma}$ violates the corresponding Bell inequality, meaning $\mathcal{C}_{cl}(N) \subsetneq \mathcal{C}_q(N)$.

Definition. 2.1.4 (Classical value)
Given $M = (M_{ij})_{i,j=1}^{N}$ with real entries, we can associate

$$\Big| \sum_{i,j=1}^{N} M_{i,j}\gamma_{i,j} \Big| \leq \omega(M) \tag{2.1.22}$$

with

$$\begin{aligned} \omega(M) &:= \sup\Big\{ \Big| \sum_{i,j=1}^{N} M_{i,j}\gamma_{i,j} \Big| \;:\; \gamma = (\gamma_{i,j}) \in \mathcal{C}_{cl}(N) \Big\} \\ &= \sup\Big\{ \Big| \sum_{i,j=1}^{N} M_{i,j}x_i y_j \Big| \;:\; x_i = \pm 1,\; y_j = \pm 1 \;\forall i,j = 1,\dots,N \Big\}. \end{aligned} \tag{2.1.23}$$

Note that the last equality follows by convexity. We then call $\omega(M)$ the *classical value* of M.

Remark. By abuse of notation, we will sometimes call M above just a *Bell inequality*.

Definition. 2.1.5 (Quantum value)
Analogously we can define for $M = (M_{ij})_{i,j=1}^{N}$ with real entries, the *quantum value* as

$$\omega^*(M) := \sup\Big\{ \Big| \sum_{i,j=1}^{N} M_{i,j}\gamma_{i,j} \Big| \;:\; \gamma = (\gamma_{i,j})_{i,j=1}^{N} \in \mathcal{C}_q(N) \Big\} \tag{2.1.24}$$

Definition. 2.1.6 (Largest violation)
For $M = (M_{i,j})_{i,j=1}^{N}$ with real entries, we define the largest violation as

$$\mathrm{LV}(M) := \frac{\omega^*(M)}{\omega(M)} \tag{2.1.25}$$

It is clear that $C_{cl}(N) \subseteq \mathcal{C}_q(N)$ is equivalent to $\omega^*(M) \geq \omega(M)$, which is equivalent to $\mathrm{LV}(M) \geq 1$ for all M.

Example. (CHSH) For the CHSH example we get

$$\omega(M) \leq 2,\ \omega^*(M) = 2\sqrt{2} \tag{2.1.26}$$

hence $\mathrm{LV}(M) \geq \sqrt{2}$. As an interesting note, this value is not far from being optimal, as we will see in a few pages.

Before stating and proving the main result of this section, namely Tsirelson's theorem, we need to introduce a previous notion which we will use in its proof.

Definition. 2.1.7 (CAR-algebra)
Given $N \geq 2$, a set of operators $X_1, \ldots, X_N$ is said to satisfy the *Classical Anticommutation Relations* if:

- $X_i^* = X_i\ \forall i = 1, \ldots, N$.
- $X_i X_j + X_j X_i = \{X_i, X_j\} = 2\delta_{ij}\,\mathbb{1}\ \forall i, j = 1, \ldots, N$.

An idea to construct such a set of operators is via the Pauli matrices. For even N (i.e. $N = 2m$), we have

$$\begin{aligned} X_1 &= \underbrace{X \otimes \mathbb{1} \otimes \ldots \otimes \mathbb{1}}_{m} & X_2 &= Y \otimes \mathbb{1} \otimes \ldots \otimes \mathbb{1} \\ X_3 &= Z \otimes X \otimes \mathbb{1} \otimes \ldots \otimes \mathbb{1} & X_4 &= Z \otimes Y \otimes \mathbb{1} \otimes \ldots \otimes \mathbb{1} \\ &\vdots & &\vdots \\ X_{2m-1} &= Z \otimes Z \otimes \ldots \otimes Z \otimes X & X_{2m} &= Z \otimes Z \otimes \ldots \otimes Z \otimes Y \end{aligned} \tag{2.1.27}$$

If N is odd, we further add the element $X_{2m+1} = Z \otimes Z \otimes \ldots \otimes Z \otimes Z$.

Now we are in position to state and prove the following formulation of Tsirelson's theorem (we will also introduce another reformulation for it later in the text).

Theorem. 2.1.8 (Tsirelson's theorem) *Let $\gamma = (\gamma_{i,j})_{i,j=1}^N$ be a correlation matrix with real entries. Then, the following are equivalent:*

1. $\gamma \in \mathcal{C}_q(N)$.
2. $\exists$ *normalised* $x_1, \ldots, x_N, y_1, \ldots, y_N$ *in a real Hilbert space such that*

$$\gamma_{ij} = \langle x_i, y_j \rangle\ , \quad \forall i, j = 1, \ldots, N\,.$$

In particular,

$$\omega^*(M) = \sup_{1 = \|x_i\| = \|y_j\|} \Big\{ \Big| \sum_{i,j=1}^{N} M_{i,j} \langle x_i, y_j \rangle \Big| \Big\} \tag{2.1.28}$$

Proof. 1. $\Rightarrow$ 2. Consider the real vector space $\mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_2)_{sa}$. We define the real Hilbert space as $\mathcal{H} := (\mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_2)_{sa}, \langle \cdot, \cdot \rangle)$. The inner product is given by

$$\langle A, B \rangle = \mathrm{Re}(\mathrm{Tr}[AB\rho]) \tag{2.1.29}$$

for every $A, B \in \mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_2)_{sa}$, where we get from γ the POVMs $\{A_i\}$ in $\mathcal{H}_1$ and $\{B_j\}$ in $\mathcal{H}_2$, as well as ρ. We further define

$$\widetilde{\mathcal{H}} := \mathrm{span}\{x_i = A_i \otimes \mathbb{1}\ :\ i = 1, \ldots, N\}\,, \tag{2.1.30}$$

and set $\mathbb{P} : \mathcal{H} \to \widetilde{\mathcal{H}}$ the orthogonal projection from $\mathcal{H}$ to $\widetilde{\mathcal{H}}$. We further set

$$y_j = \mathbb{P}(\mathbb{1} \otimes B_j) \qquad \forall j = 1, \ldots, N. \tag{2.1.31}$$

We then get a collection of $x_1, \dots, x_N, y_1, \dots, y_N$ in a real Hilbert space with $\dim(\widetilde{H}) = k \leq N$ verifying

$$\|x_i\| \leq 1,\ \|y_j\| \leq 1,\ \gamma_{ij} = \langle x_i, y_j \rangle = \mathrm{Tr}[A_i \otimes B_j \rho] \qquad \forall i,j = 1, \dots, N\,. \tag{2.1.32}$$

Since the A_is and the B_js are self-adjoint, we drop the Re in the trace above. Finally, since their norms should be 1, we normalize them using the following procedure:

$$\tilde{x}_i = x_i \oplus \sqrt{1 - \|x_i\|^2} \oplus 0, \quad \tilde{y}_j := y_j \oplus 0 \oplus \sqrt{1 - \|y\|_j^2} \tag{2.1.33}$$

2. $\Rightarrow$ 1. Consider $(\mathbb{R}^M, \langle \cdot, \cdot \rangle)$, the real Hilbert space where $(x_i)_{i=1}^N$, $(y_j)_{j=1}^N$ live. Assume w.l.o.g. that M is even. Then we can construct the products of $\frac{M}{2}$ Pauli matrices as introduced above and define

$$T : \mathbb{R}^m \to \mathrm{span}\{X_1, \dots, X_m\}, \quad e_k \mapsto X_k \tag{2.1.34}$$

with the X_i constructed as in eq. (2.1.27). Then

$$\left\| T : \ell_2^m \to (\mathbb{C}^{2\times 2})^{\otimes m} \right\| \leq 1 \tag{2.1.35}$$

directly. In particular, for every $x \in \mathbb{R}^m$ with $\|x\| \leq 1$, we have $\|T(x)\| \leq 1$. Moreover, we find

$$\frac{1}{2^{\frac{m}{2}}} \mathrm{Tr}[(Tx)(Ty)] = \langle x, y \rangle \qquad \forall x, y \in \mathbb{R}^M\,. \tag{2.1.36}$$

Consider further

$$|\psi\rangle = \frac{1}{2^{\frac{M}{4}}} \sum_{i,j=1}^{\frac{M}{2}} |ij\rangle \in (\mathbb{C}^2 \times \mathbb{C}^2)^{\otimes M} \tag{2.1.37}$$

which gives us that for $A, B \in (\mathbb{C}^2 \times \mathbb{C}^2)^{\otimes m}$,

$$\frac{1}{2^{\frac{M}{2}}} \mathrm{Tr}[AB] = \mathrm{Tr}[A \otimes B |\psi\rangle\langle\psi|] = \langle \psi | A \otimes B | \psi \rangle\,. \tag{2.1.38}$$

We define

$$A_i := T(x_i),\ B_j := T(y_j) \qquad \forall i,j = 1, \dots, N\,. \tag{2.1.39}$$

Then, we obtain a family of self-adjoint operators with $\|\cdot\| \leq 1$ such that

$$\langle \psi | A_i \otimes B_j | \psi \rangle = \frac{1}{2^{\frac{M}{2}}} \mathrm{Tr}[A_i B_j] = \langle x_i, y_j \rangle \quad \forall i,j = 1, \dots, N \tag{2.1.40}$$

□

2.1.3 Grothendieck's Theorem

The main result of this section is Grothendieck's theorem.

Theorem. 2.1.9 (Grothendieck's Theorem) *There exists a universal constant K_G such that $\forall N \in \mathbb{N}$, $\forall (M_{i,j})_{i,j=1}^N \in \mathbb{R}^{N\times N}$,*

$$\begin{aligned} &\sup\left\{ \left| \sum_{i,j=1}^N M_{i,j} \langle x_i, y_j \rangle \right| \ :\ \|x_i\| = \|y_j\| = 1\ \forall i,j = 1, \dots N \right\} \\ &\qquad \leq K_G \sup\left\{ \left| \sum_{i,j=1}^N M_{i,j} t_i s_j \right| \ :\ t_i = \pm 1\,, s_j = \pm 1\ \forall i,j = 1, \dots N \right\}, \end{aligned} \tag{2.1.41}$$

with K_G the (real) Grothendieck's constant:

$$1.67696 \leq K_G < \frac{\pi}{2\log(1+\sqrt{2})}. \tag{2.1.42}$$

We can rephrase that to

$$\omega^*(M) \leq K_G\, \omega(M) \tag{2.1.43}$$

or, equivalently,

$$LV(M) \leq K_G. \tag{2.1.44}$$

Example. CHSH. Let us recall that, for the example of the CHSH, we had

$$\mathrm{LV}(M_{\mathrm{CHSH}}) \geq \sqrt{2}\,.$$

Now we can confirm the previously mentioned statement that the bound obtained for the largest violation of the CHSH inequality is relatively close to the optimal one, given by K_G.

Before finishing this section and starting with nonlocal games, let us summarize what we have discussed in the last pages. The idea for the use of quantum nonlocality in various fields of quantum information appears with high frequency in the past few years. In certain fields such as quantum communication or quantum cryptography, some quantum correlations which are not classical, so that they violate a Bell inequality, can be used to define "certain protocols".

In general, Bell inequalities allow us to realize advantages of quantum mechanics with respect to the classical theory. Therefore, in some sense, $LV(M)$, for a certain Bell inequality M, can be understood as a measure of "how better is quantum mechanics than classical mechanics". The previous theorems, however, provide some limitations of quantum mechanics, as the violations for Bell inequalities are bounded.

A natural question that arises is whether one can get larger violations of Bells inequality in a broader context? This question was answered by Tsirelson with a yes by designing a three player experiment, completely analogous to the two-player one.

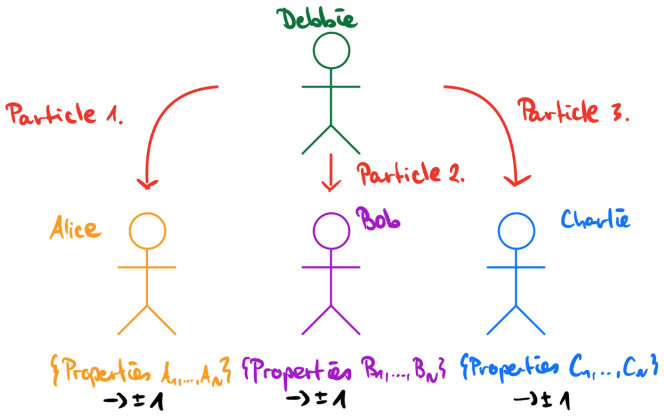


fig. 2.2: Three player game.

In this case, we can follow a similar analysis as for the case of two players. First, the classical correlations become

$$\gamma_{i,j,k} = \int_\Omega A_i(\omega) B_j(\omega) C_k(\omega)\, d\mathbb{P}(\omega)\,, \tag{2.1.45}$$

with $(\Omega, \mathbb{P})$ the hidden probabilty. The quantum correlations are given by using that for $\gamma = (\gamma_{ijk})_{i,j,k=1}^N$, there exists $A_1, \ldots, A_N, B_1, \ldots, B_n, C_1, \ldots, C_N$ self adjoint and completely positive acting on $\mathbb{C}^n$ with

$$\max_{i,j,k=1,\ldots,N} \{\|A_i\|\,, \|B_j\|\,, \|C_k\|\} \leq 1 \tag{2.1.46}$$

and ρ a density operator acting on $\mathbb{C}^n \times \mathbb{C}^n \times \mathbb{C}^n$ with

$$\gamma_{i,j,k} = \mathrm{Tr}[A_i B_j C_k \rho] \qquad \forall i, j, k = 1, \ldots, N. \tag{2.1.47}$$

Finally, the classical and entangled value of a Bell inequality, as well as its largest violation, are defined analogously to the case of the two-player scenario:

$$LV(M) = \frac{\omega^*(M)}{\omega(M)}\,.$$

Theorem. 2.1.10 (Tsirelson) *For every $D > 0$, there exist a large enough $N \in \mathbb{N}$ and a Bell inequality $M = (M_{ijk})_{i,j,k=1}^{N}$ such that*

$$LV(M) \geq D\,.$$

The most direct implication of this result is that, as soon as we consider three players, we get an unlimited amount of violation of Bell inequalities. Moreover, the best estimate for D in terms of N up to date is $D \backsimeq N^4$.

Consequently, the tripartite scenario allows for unlimited advantages by using quantum mechanics rather than with classical mechanics.

2.2 Non local games

In this section, we introduce the notion of non-local games and translate the results of non-locality introduced in the previous pages to the context of games. Unless we explicitly say it, we are only going to consider games with two players, Alice and Bob, and a referee, Charlie, as in the following picture. This section is largely inspired in [21, 12].

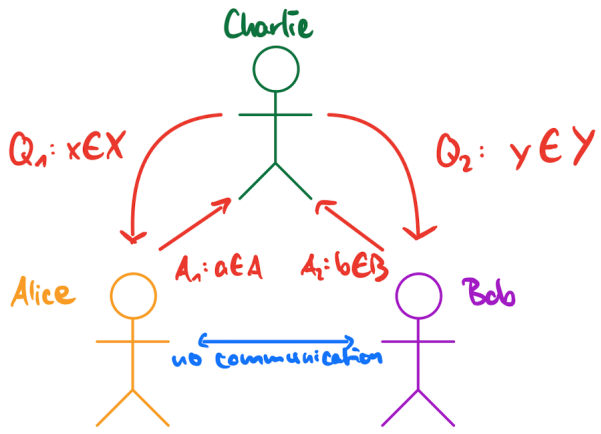


fig. 2.3: Sketch of non-local games.

In Figure 2.3, we see schematically the construction of a non-local game with two players and one referee. The latter one sends a question $x \in X$ to the first player, Alice, and another question $y \in Y$ to the other player, Bob. This is done simultaneously, but Alice and Bob cannot share any information, so they do not know about the question received by the other. Then, after receiving $x \in X$, Alice answers with $a \in A$, and Bob acts analogously replying with $b \in B$. Both answers are sent to Charlie, who checks the answers to the given questions following a protocol named *verifier*, and decides whether Alice and Bob won or lost the game. The goal is, of course, to win the game as often as possible, and for that, Alice and Bob need to agree prior to the game on a common strategy.

Let us study all these notions formally, by mathematically introducing the concept of *non-local game*.

Definition. 2.2.1 (Non local game)
A *non-local game* is a 6-tuple G, with $G = (X, Y, A, B, \Pi, V)$ such that

1. X, Y are sets of *questions* and A, B are corresponding sets of *answers*. Both are finite (and non-empty) sets.
2. $\Pi \in \mathcal{P}(X, Y)$ is a probability vector (over the questions).
3. $V : A \times B \times X \times Y \to \{0, 1\}$ is a *predicate*, which is basically the referee/verifier, defined as

$$V(a, b|x, y) \equiv V(a, b, x, y) = \begin{cases} 1 & \text{if answering } (a, b) \text{ to } (x, y) \text{ WINS} \\ 0 & \text{if answering } (a, b) \text{ to } (x, y) \text{ LOSES} \end{cases} \tag{2.2.1}$$

Let us show now some examples of basic non-local games and their sets of questions, answers, probability vectors and predicates.

Example. CHSH game. We set $X = Y = A = B = \{0, 1\}$ and the probability vector is given by:

$$\Pi(0, 0) = \Pi(1, 0) = \Pi(0, 1) = \Pi(1, 1) = \frac{1}{4}\,.$$

The predicate is

$$V(a, b|x, y) = \begin{cases} 1 & a \oplus b = x \wedge y \\ 0 & a \oplus b \neq x \wedge y \end{cases}, \tag{2.2.2}$$

where $a \oplus b$ denotes a XOR b and $x \wedge y$ is x AND y.

Example. FFL game. The name stands for Fortnau, Feige and Lorasz, who first came up with the example. In this case, we set $X = Y = A = B = \{0, 1\}$, and the probability vector is given by:

$$\Pi(1, 1) = 0, \Pi(0, 1) = \mathbb{P}(1, 0) = \Pi(0, 0) = \frac{1}{3}\,.$$

Moreover, the predicate is now given by

$$V(a, b|x, y) = \begin{cases} 1 & a \vee x \neq b \vee y \\ 0 & a \vee x = b \vee y \end{cases}, \tag{2.2.3}$$

where $a \vee x$ denotes a OR x.

Example. Graph coloring game. We set $H = (V, E)$ to be an undirected graph, with $n = |V|$, $m = |E|$, $m \geq 1$, and we take $k \in \mathbb{N}$. Let us consider the questions $X = Y = \{1, \ldots, n\}$ and the answers $A = B = \{1, \ldots, k\}$, which are called *colors*. We further set the probability vector

$$\Pi(x, y) = \begin{cases} \frac{1}{2n} & x = y \\ \frac{1}{4m} & x \neq y \,(x \text{ adjacent to } y) \\ 0 & \text{otherwise} \end{cases} \tag{2.2.4}$$

and the predicate

$$V(a, b|x, y) = \begin{cases} 1 & x = y, a = b \\ 1 & x \neq y, a \neq b \\ 0 & \text{otherwise} \end{cases}. \tag{2.2.5}$$

Note that this game is used to model the problem of coloring of a graph. Namely, given a set of colors, which is the minimal number of colors we need so that we can associate a color to each

vertex in such a way that two adjacent vertices have different colors. It is not difficult to realize that winning the previous game with certainty 1 for a certain number of colors k implies that the graph of such a problem can be completely colored according to this rule. We will go in further detail on this later in the text.

As mentioned above, the main purpose of Alice and Bob is winning their non-local game as often as possible, and for that they need to devise a previous strategy that they can use when playing the game. We show below a list of different forms of strategies, depending on the tools that Alice and Bob are allowed for such a game (i.e., the type of measurements that can be used to determine their answers, given their questions).

Definition. 2.2.2 (Strategies for non local games)

1. **Deterministic strategies:** This is the simplest possible case. In this form of strategy, Alice and Bob consider some deterministic functions and associate to each question a certain answer prior to the game. In detail, they consider:

$$f : X \to A,\ g : Y \to B \quad (x, y) \mapsto (f(x), g(y)) \,, \tag{2.2.6}$$

   and given any question pair (a, b), they output the answer $(f(x), g(y))$. Note that, in this whole procedure, they do not commute during the came. Moreover, note that this strategy is completely classical, as no quantum information is employed whatsoever.

2. **Randomized strategy:** This is slightly more involved that the previous case, but the probability for winning is the same as before, and it does not use any quantum information either. The only difference with respect to the case above is that, now, answers are not predetermined, given the questions, but they are drawn from the sets of answers randomly, using some probability distributions for that. As these strategies are just a random selection of deterministic strategies, the benefit of this case with respect to the previous one in our problem is none.

3. **Entangled strategy:** In a similar fashion as in the non-locality scenarios presented in the previous section, Alice and Bob share a (hopefully entangled) quantum state $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ in a finite-dimensional bipartite Hilbert space. The strategy is then

   - Given $x \in X$, Alice performs a POVM $\{P_a^x\}_{a\in A}$ only on $\mathcal{H}_A$ and sends the output observed as an answer.
   - Given $y \in Y$, Bob performs a POVM $\{Q_b^y\}_{b\in B}$ only on $\mathcal{H}_B$ and sends the output observed as an answer.

   Then, the probability of answering (a, b) to (x, y) is given by

$$P(x, y, a, b) = \langle P_a^x \otimes Q_b^y | \rho \rangle \ ,$$

   with $\rho = |\psi\rangle \langle\psi|$.

4. **Commuting strategy:** This strategy is similar to the previous one. Now, Alice and Bob also share a quantum state $|\psi\rangle \in \mathcal{H}$, but now in a possibly infinite-dimensional Hilbert space. The strategy is then defined as follows:

   - Given $x \in X$, Alice performs a POVM $\{P_a^x\}_{a\in A}$ on the whole $\mathcal{H}$ and sends the output observed as an answer.

- Given $y \in Y$, Bob performs a POVM $\{Q_b^y\}_{b\in B}$ on the whole $\mathcal{H}$ and sends the output observed as an answer.

Then, the probability of answering (a, b) to (x, y) is given by

$$P(x, y, a, b) = \mathrm{Tr}[P_a^x Q_b^y \rho]\,,$$

with $\rho = |\psi\rangle\langle\psi|$. For tractability of the previous quantity, we have to assume in this case that $[P_a^x, Q_b^y] = 0\ \forall a, b, x, y$.

2.2.1 Values of a non-local game

Once we have introduced the different strategies that Alice and Bob can consider for their non-local game, we can associate to each strategy the notion of the *value* of the non-local game. In general, this is just the maximal success probability for Alice and Bob in the game.

Definition. 2.2.3 (Value of a game)
Consider a non-local game $G = (X, Y, A, B, \Pi, V)$. The maximal success probability for Alice and Bob is given by

1. **Classical value.** This is the value associated to a non-local game, assuming that the strategy followed by Alice and Bob was deterministic or randomized (in any case, no quantum information considered). The maximal success probability in this case is:

$$\omega_c(G) \equiv \omega(G) := \max_{\substack{f:X\to A,\\ g:Y\to B}} \sum_{(x,y)\in X\times Y} \Pi(x, y)V(f(x), g(y)|x, y) \tag{2.2.7}$$

2. **Entangled value.** In this case, the strategy considered is an entangled one. The quantum (or entangled) value is computed taking the supremum over all possible entangled strategies of the following quantity:

$$\omega_q(G) \equiv \omega^*(G) = \sup_{\substack{\psi\in\mathcal{H}_A\otimes\mathcal{H}_B\\ \{P_a^x\}_{a\in A}\\ \{Q_b^x\}_{b\in B}}} \sum_{(x,y)\in X\times Y} \Pi(x, y) \sum_{(a,b)\in A\times B} \langle\psi|P_a^x \otimes Q_b^y|\psi\rangle\, V(a, b|x, y) \tag{2.2.8}$$

with $\{P_a^x\}_{a\in A}$ and $\{Q_b^x\}_{b\in B}$ POVMs on $\mathcal{H}_A$ and $\mathcal{H}_B$ respectively.

3. **Commuting operator value.** In an analogous way to the previous value, we introduce the commuting operator value by taking supremum now over all possible commuting strategies:

$$\omega_{co}(G) := \sup_{\substack{\psi\in\mathcal{H}_A\otimes\mathcal{H}_B\\ \text{commuting}\\ \text{strategies}}} \sum_{(x,y)\in X\times Y} \Pi(x, y) \sum_{(a,b)\in A\times B} \langle\psi|Q_b^y P_a^x|\psi\rangle\, V(a, b|x, y) \tag{2.2.9}$$

Given all these notions for values of a game, we can compute at least the previous two ones (classical and entangled values) for the examples of non-local games introduced above.

Example. CHSH game. We can compute the classical and entangled values for the CHSH game, respectively. We leave as an exercise to show that that

$$\omega_c(G_{\mathrm{CHSH}}) = \frac{3}{4}\,, \quad \text{and} \quad \omega_q(G_{\mathrm{CHSH}}) = \cos^2\left(\frac{\pi}{8}\right)\,.$$

In this case, using a quantum strategy, the average win probability is strictly better than what is possible using a classical one.

Example. FFL game. For this game, we can also compute the classical and entangled values for the CHSH game, which we also leave as an exercise. In this case, we have:

$$\omega_c(G_{\text{CHSH}}) = \frac{2}{3} = \omega_q(G_{\text{CHSH}})\,.$$

In this case, quantum strategies and classical ones can perform equally well. Note that, in both examples, computing the classical values is just done by testing all deterministic strategies.

Example. Graph coloring game. Given $k \in \mathbb{N}$, the fact that $\omega(G) = 1$ is equivalent to the chromatic number of H being at most k. Moreover, there are known H, $k \in N$, for which $\omega(G) < 1$ and $\omega^*(G) = 1$

Let us conclude this subsection by collecting some of the information we already have about values of games.

1. Derived from the increasing order in the restrictivity for the strategies presented above, we find a hierarchy in values for games, namely $\omega_c(G) \leq \omega_q(G) \leq \omega_{co}(G)$.

2. For the CHSH game in particular, we find that $\omega_c(G) < \omega_q(G)$.

2.2.2 Correlations

In this subsection, we aim at introducing the notion of correlations, derived from the strategies for non-local games presented above. Therefore, we will be able to associate sets of correlations to values of non-local games.

Definition. 2.2.4 (Correlations)
Let us fix the sizes of the sets of questions and answers to $|X| = m$, $|Y| = n$, $|A| = k$, $|B| = l$. In general, we will will write the sizes of the four involved sets as a superscript in the set of correlations, but we will drop them when they are clear from the context. Then, we define the following various sets of correlations:

- The set of **classical correlations** (which we will not study in detail in this text) is denoted by

$$\mathcal{C}_{cl} \equiv \mathcal{C}_{cl}^{m,n,k,l}\,. \tag{2.2.10}$$

- The set of **quantum correlations** is given by

$$\mathcal{C}_q^{k,l,m,n} := \left\{ p(a,b|x,y) = \langle\psi|P_a^x \otimes Q_b^y|\psi\rangle \;:\; \begin{array}{c} |\psi\rangle \in \mathcal{H} \text{ normalised }, \{P_a^x\}_{a\in A},\ \{Q_b^y\}_{b\in B} \\ \text{an entangled strategy,} \\ \text{on } \mathcal{H}_A \text{ and } \mathcal{H}_B \text{ finite dimensional} \end{array} \right\} \tag{2.2.11}$$

- The set of **quantum spatial correlations** is defined in an analogous way to the set of quantum correlations, but now we allow for infinite-dimensional Hilbert spaces

$$\mathcal{C}_{qs}^{k,l,m,n} := \left\{ p(a,b|x,y) = \langle\psi|P_a^x \otimes Q_b^y|\psi\rangle \;:\; \begin{array}{c} |\psi\rangle \in \mathcal{H} \text{ normalised }, \{P_a^x\}_{a\in A},\ \{Q_b^y\}_{b\in B} \\ \text{an entangled strategy,} \\ \text{on } \mathcal{H}_A \text{ and } \mathcal{H}_B \text{ possibly } \infty\text{-dimensional} \end{array} \right\} \tag{2.2.12}$$

- The set of **quantum correlations well-approximated by tensor products** (finite-dimensional) is by definition

$$\mathcal{C}_{qa}^{k,l,m,n} := \overline{\mathcal{C}}_q^{k,l,m,n} \tag{2.2.13}$$

- The set of **quantum commuting correlations** is denoted by

$$\mathcal{C}_{qc}^{k,l,m,n} \equiv \mathcal{C}_{co}^{k,l,m,n} = \left\{ p(a,b|x,y) = \langle\psi|Q_b^y P_a^x|\psi\rangle \ : \begin{array}{c} |\psi\rangle \in \mathcal{H} \text{ normalised }, \{P_a^x\}_{a\in A},\ \{Q_b^y\}_{b\in B} \\ \text{a commuting strategy} \\ \text{on } \mathcal{H} \text{ (possibly infinite dimensional)} \end{array} \right\} \tag{2.2.14}$$

If we fix m, n, k, l we find the following chain of (strict) inclusions

$$\boxed{\mathcal{C}_{cl} \subsetneq \mathcal{C}_q \subsetneq \mathcal{C}_{qs} \subsetneq \mathcal{C}_{qa} \subsetneq \mathcal{C}_{co}} \tag{2.2.15}$$

The inclusions of the previous chain are relatively straightforward, just by considering the definition for each of the sets of correlations involved. The fact that all of them are strict, though, is much more involved. First, note that the fact that there are quantum correlations which are not classical is due to Bell's theorem in 1964. Moreover, the problem of proving that there are (quantum) correlations in any of the other sets which do not belong to the previous one has been a very active field of research in the past few years, giving rise to some seminal works in the last decade. The timeline of the discoveries is the following:

- **(Bell '64)** As a starting point for the previous chain of inequalities, in 1964 Bell proved in [1] that there are quantum correlations which are not classical:

$$\mathcal{C}_{cl} = \mathcal{C}_q \,.$$

- **(Tsirelson, '06)** For the set of commuting quantum correlations, we alllow for infinte-dimensional Hilbert spaces in general. In 2006, Tsirelson showed that, if we restrict to finite-dimensional Hilbert spaces, then

$$\mathcal{C}_q = \mathcal{C}_{co}^{\text{finite}} \,.$$

  The question whether the same situation could hold for infinite-dimensional Hilbert spaces was then named after him **Tsirelson's problem**. With the appearance of the next results mentioned below, this problem was reduced to the question whether the last two sets of correlations presented in the previous chain of inclusions coincide or not.

- **(Scholz & Werner, '08)** Two years after Tsirelson's result for finite-dimensional Hilbert spaces, Scholz and Werner extended it in [13] to the so-called "effectively finite-dimensional", i.e. to finite-dimensional von Neumann algebras.

- **(Slofstra, '16 and '17)** Almost a decade later, Slofstra showed in 2016 in [18] that that there are quantum commuting correlations which are not quantum spatial ones,

$$\mathcal{C}_{qs} \neq \mathcal{C}_{co} \,,$$

  and a year later he showed in [17] that the set of quantum spatial correlations is not closed, yielding then

$$\mathcal{C}_{qs} \subsetneq \mathcal{C}_{qa} \,.$$

- **(Coladangelo-Stark, '18)** In 2018, Coladangelo and Stark found in [6] a particular set of values of m, n, k, l for which there are quantum spatial correlations (in infinite-dimensional Hilbert spaces) which are not quantum correlations (in finite-dimensional Hilbert spaces), i.e.

$$\mathcal{C}_q \subsetneq \mathcal{C}_{qs} \,.$$

  More specifically, they showed:

$$C_q^{4,5,3,3} \neq \mathcal{C}_{qs}^{4,5,3,3} \,.$$

- **(MIP* = RE, '20)** Finally, in the major breakthrough [9], Ji et al. showed that, in general,

$$\mathcal{C}_{qa} \subsetneq \mathcal{C}_{co}\,.$$

This solved in the negative the aforementioned Tsirelson's problem, as well as the well-known **Connes embedding problem**, which had been previously shown to be equivalent to Tsirelson's problem.

2.2.3 Non-local games as hyperplanes

Before moving to the next section, in which we will provide an approach to finding quantum values for non-local games using semidefinite programs, here we give a reinterpretation of non-local games as hyperplanes. Previously, we need the following technical lemma.

Lemma. 2.2.5 *The sets of correlations $\mathcal{C}_{qs}, \mathcal{C}_{qa}$ and $\mathcal{C}_{co}$ are all convex.*

Proof.

- Let us show that $\mathcal{C}_{qs}$ is convex. For that, we need to show that, given two correlations $P_1, P_2 \in C_{qs}$, and $\lambda \in [0,1]$, we then have $\lambda P_1 + (1-\lambda)P_2 \in C_{qs}$.

  Since $P_1, P_2 \in C_{qs}$, we set

$$P_i := p_i(a,b|x,y) = \langle \psi_i | P_a^{(i)x} \otimes Q_b^{(i)y} | \psi_i \rangle \qquad i = 1,2 \tag{2.2.16}$$

  on some Hilbert spaces $\mathcal{H}_A^{(i)}$ and $\mathcal{H}_B^{(i)}$, with $|\psi_i\rangle \in \mathcal{H}_A^{(i)} \otimes \mathcal{H}_B^{(i)}$. We then construct a correlation from them. For that, we need to properly define the Hilbert spaces, the POVMs and the state:

  - We construct the new Hilbert space combining the previous ones in the following form:

$$\begin{aligned}(\mathcal{H}_A^{(1)} \oplus \mathcal{H}_A^{(2)}) \otimes (\mathcal{H}_B^{(1)} \oplus \mathcal{H}_B^{(2)}) \\ \cong (\mathcal{H}_A^{(1)} \otimes \mathcal{H}_B^{(1)}) \oplus (\mathcal{H}_A^{(1)} \otimes \mathcal{H}_B^{(2)}) \oplus (\mathcal{H}_A^{(2)} \otimes \mathcal{H}_B^{(1)}) \oplus (\mathcal{H}_A^{(2)} \otimes \mathcal{H}_B^{(2)})\,.\end{aligned} \tag{2.2.17}$$

  - Each of the POVMs is constructed as a direct sum of the ones associated to P_1 and P_2, namely:

$$\begin{aligned}P_a^x &= P_a^{(1)x} \oplus P_a^{(2)x}\,, \\ Q_b^y &= Q_b^{(1)y} \oplus Q_b^{(2)y}\,.\end{aligned} \tag{2.2.18}$$

  - Finally, the state is defined combining the previous ones and normalizing it:

$$|\psi\rangle = \sqrt{\lambda}\,|\psi_1\rangle \oplus 0 \oplus 0 \oplus \sqrt{1-\lambda}\,|\psi_2\rangle\,. \tag{2.2.19}$$

  It is clear that these three elements define a quantum spatial correlation and, moreover, that this correlation coincides with

$$\lambda P_1 + (1-\lambda)P_2\,.$$

- To prove that $\mathcal{C}_{qa}$ is convex, note that $C_{qa} = \overline{C}_q s$ and the closure of a convex set is convex.

- Finally, the proof for C_{co} is completely analogous to that of C_{qs}.

□

With this idea on mind, we can compare correlations to separating hyperplanes.

Definition. 2.2.6 (Separating hyperplane)
Given an element $H = (H_{a,b,x,y})_{a,b,x,y} \in \mathbb{R}^{m,n,k,l}$, H can be regarded as a linear functional acting on the correlation $(P(a,b|x,y))_{a,b,x,y}$ as follows:

$$\langle H, P\rangle = \sum_{a,b,x,y} H_{a,b,x,y} P(a,b|x,y). \tag{2.2.20}$$

Therefore, it is reasonable to define a maximal value of a given hyperplane H with respect to a set $\mathcal{C}$ of correlations

$$\max_{\mathcal{C}}(H) = \sup_{P\in\mathcal{C}} |\langle H, P\rangle| \tag{2.2.21}$$

This allows us to establish an identification between correlations and values of non-local games; e.g. $\mathcal{C}_{co} \leftrightarrow \omega^{co}(G)$.

Remark. We conclude from the previous identification that non-local games are just hyperplanes with positive coefficients.

2.3 Semi-definite programs for the entangled bias of a XOR game

In this section, we will introduce semi-definite programs (SDPs) which will allow us to compute the entangled bias of a XOR game (and, thus, the entangled value). Before introduce the SDPs, let us recall the notions of bias of a game and XOR games.

Definition. 2.3.1 (XOR games)
XOR games are a restricted type of non-local games $G = (X, Y, A, B, \Pi, V)$ in which there are only two answers for each player, namely $A = B = \{0, 1\}$. Moreover, in this case the predicate is given by

$$V(a,b|x,y) = \begin{cases} 1 & a \oplus b = f(x,y) \\ 0 & a \oplus b \neq f(x,y) \end{cases}, \tag{2.3.1}$$

for a function $f : X \times Y \to \{0, 1\}$. Then, we can just identify the game with $G = (X, Y, \Pi, f)$.

Example. CHSH game. The CHSH game is a particular case of a XOR game, since it is of the form $(\{0,1\}, \{0,1\}, \Pi, f)$ for $f(x,y) = x \wedge y$.

Definition. 2.3.2 (Bias of a game)
The bias of a strategy is the difference between the probability of winning it and the probability of losing it using that strategy. Moreover, the bias of a XOR non-local game is the supremum of the biases over all strategies considered for that game.

Since the probability of winning plus the probability of losing a game is 1, the classical and entangled biases for a XOR non-local game G are given by

$$\mathcal{E}(G) = 2\omega(G) - 1 \quad , \qquad \mathcal{E}^*(G) = 2\omega^*(G) - 1\,,$$

or, equivalently, we have

$$\omega(G) = \frac{1 + \mathcal{E}(G)}{2} \quad , \qquad \omega^*(G) = \frac{1 + \mathcal{E}^*(G)}{2}\,.$$

Let us discuss now how XOR game strategies can be described by observables. Let us consider a XOR game $G = (X, Y, \Pi, f)$ and consider a entangled strategy for the game, represented by a state $\rho \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ and measurement operators

$$\{P_0^x, P_1^x\} \text{ POVMs in } \mathcal{H}_A \quad \text{and} \{Q_0^y, Q_1^y\} \text{ POVMs in } \mathcal{H}_B\,.$$

Consider the following quantity:

$$\sum_{(x,y)\in X\times Y} \Pi(x,y)(-1)^{f(x,y)} \operatorname{Tr}[(P_0^x - P_1^x)\otimes(Q_0^y - Q_1^y)\rho]\,.$$

It is clear that the supremum over this quantity coincides with the bias of a game described above. Moreover, if we define $A_x := P_0^x - P_1^x$ and $B_y := Q_0^y - Q_1^y$, these two operators are observable. Therefore, the bias of a game can be computed as the supremum of the quantity

$$\sum_{(x,y)\in X\times Y} \Pi(x,y)(-1)^{f(x,y)} \operatorname{Tr}[A_x\otimes B_y\rho]\,,$$

over observables $\{A_x\}_{x\in X}$ and $\{B_y\}_{y\in Y}$ with $\|A_x\|, \|B_y\| \leq 1$ for every $x\in X$ and $y\in Y$.

Let us move now to the definition of the SDPs for computing the entangled value of certain XOR games. Before introducing the formal definition, we need the following reformulation of Tsirelson's theorem, which we will not prove here, as the proof completely resembles that of the original version for the theorem presented above.

Proposition. 2.3.3 *For $X, Y \neq \emptyset$ and $M \in \mathcal{B}(\mathbb{R}^{|Y|}, \mathbb{R}^{|X|})$, the following are equivalent:*

1. *There exist complex Hilbert spaces $\mathcal{H}_A, \mathcal{H}_B$, as well as $\rho \in \mathcal{S}(\mathcal{H}_A\otimes\mathcal{H}_B)$ a density operator and $\{A_x : x\in X\} \subseteq \mathcal{B}(\mathcal{H}_A)_{sa}$, $\{B_y : y\in Y\} \subseteq \mathcal{B}(\mathcal{H}_B)_{sa}$ such that*

$$\|A_x\| \leq 1 \quad , \quad \|B_y\| \leq 1 \quad , \quad M(x,y) = \langle Ax\otimes By, \rho\rangle \quad \forall x\in X \quad , \quad \forall y\in Y.$$

2. *There exists positive semi-definite operators $R \in \mathcal{B}(\mathbb{C}^{|X|})_+$, $S\in\mathcal{B}(\mathbb{C}^{|Y|})_+$, $R(x,x) = 1$, $S(y,y) = 1$ $\forall x\in X$, $\forall y\in Y$, such that*

$$\begin{pmatrix} R & M \\ M^* & S \end{pmatrix} \geq 0\,. \tag{2.3.2}$$

Let us recall that the entangled bias introduced above is defined by

$$\mathcal{E}^*(G) = 2\omega^*(G) - 1 \tag{2.3.3}$$

and can be equivalently obtained by means of the following formulation:

$$\mathcal{E}^*(G) := \sup_{\substack{\text{POVMs } \{P_a^x\}_{a\in A}, \{Q_b^y\}_{b\in B} \\ \rho\in\mathcal{S}(\mathcal{H}_A\otimes\mathcal{H}_B)}} \sum_{(x,y)\in X\times Y} \Pi(x,y)(-1)^{f(x,y)} \operatorname{Tr}[P_a^x\otimes Q_b^y\rho]. \tag{2.3.4}$$

By the reformulation of Tsirelson's theorem presented in this section, namely Proposition 2.3.3, this is equivalent to

$$\sup \sum_{(x,y)\in X\times Y} \Pi(x,y)(-1)^{f(x,y)} M(x,y)\,, \tag{2.3.5}$$

over $M\in\mathcal{B}(\mathbb{R}^{|X|}, \mathbb{R}^{|Y|})$ such that there exist $R\in\mathcal{B}(\mathbb{C}^{|X|})_+$, and $S\in\mathcal{B}(\mathbb{C}^{|Y|})_+$, with $R(x,x) = 1$, $S(y,y) = 1$, $\forall x\in X$, $\forall y\in Y$, and

$$\begin{pmatrix} R & M \\ M^* & S \end{pmatrix} \geq 0$$

This gives us an idea of how to construct an SPD for estimating the entangled bias of a XOR game by defining

$$D(x,y) = \Pi(x,y)(-1)^{f(x,y)}\,. \tag{2.3.6}$$

Moreover, we need to define $\Delta \in \mathcal{B}(\mathcal{B}(\mathbb{C}^{|X|} \oplus \mathbb{C}^{|Y|}))$ as the complete dephasing channel (zeroing out the non-diagonal entries of a density matrix and leaving the diagonal entries unmodified; we will provide more details on this channel in the next chapter). Further, let us denote

$$H := \frac{1}{2}\begin{pmatrix} 0 & D \\ D^* & 0 \end{pmatrix} \qquad H \in \mathcal{B}(\mathbb{C}^{|X|} \oplus \mathbb{C}^{|Y|})\,. \tag{2.3.7}$$

We are now in position of introducing the desired SDP.

Definition. 2.3.4 (SDP for XOR non-local games)
A semi-definite program (SDP) to compute the entangled bias of a XOR game is denoted by $(\Delta, H, \mathbb{1}_{\mathbb{C}^{|X|} \oplus \mathbb{C}^{|Y|}})$ and is given by the following two equivalent formulations:

Primal Problem		**Dual Problem**	
maximize	$\langle H, Z\rangle$	minimize	$\mathrm{Tr}[\omega]$
subject to	$\Delta(Z) = \mathbb{1}_{\mathbb{C}^{\|X\|} \oplus \mathbb{C}^{\|Y\|}}$	subject to	$\Delta(\omega) \geq H$
	$Z \in \mathcal{B}(\mathbb{C}^{\|X\|} \oplus \mathbb{C}^{\|Y\|})_+$		$\omega \in \mathcal{B}(\mathbb{C}^{\|X\|} \oplus \mathbb{C}^{\|Y\|})_{sa}$

In the dual problem, we have used that the completely dephasing channel is self-adjoint, i.e. $\Delta = \Delta^*$.

Strong duality and achievability of optimal values in both the primal and the dual problems follow from Slater's theorem. A feasible solution Z for the primal problem is $Z = \mathbb{1}_{\mathbb{C}^{|X|} \oplus \mathbb{C}^{|Y|}}$, while for the dual problem a feasible solution ω is $\omega = \lambda\, \mathbb{1}_{\mathbb{C}^{|X|} \oplus \mathbb{C}^{|Y|}}$ for λ large enough.

2.3.1 Primal problem

Let us verify that the optimal value in the primal problem indeed agrees with the entangled bias of the XOR game G.

Assume that $M \in \mathcal{B}(\mathbb{R}^{|X|}, \mathbb{R}^{|Y|})$ is such that there exist $R \in \mathcal{B}(\mathbb{C}^{|X|})_+$, and $S \in \mathcal{B}(\mathbb{C}^{|Y|})_+$, with $R(x,x) = 1$, $S(y,y) = 1$, $\forall x \in X$, $\forall y \in Y$,

$$Z := \begin{pmatrix} R & M \\ M^* & S \end{pmatrix} \in \mathcal{B}(\mathbb{C}^{|X|} \oplus \mathbb{C}^{|Y|})_+\,. \tag{2.3.8}$$

We need to check that the conditions of the primal problem are satisfied:

- Z is primal feasible, since $\Delta(Z) = \mathbb{1}_{\mathbb{C}^{|X|} \oplus \mathbb{C}^{|Y|}}$ is equivalent to R and S having diagonal entries equal to 1, which is something we are assuming for our R and S.

- The objective value is

$$\langle H, Z\rangle = \frac{1}{2}\langle D, M\rangle + \frac{1}{2}\langle D^*, M^*\rangle = \langle D, M\rangle\ . \tag{2.3.9}$$

  Since

$$\langle D, M\rangle = \sum_{(x,y)\in X\times Y} \Pi(x,y)(-1)^{f(x,y)} M(x,y)\,, \tag{2.3.10}$$

  then the optimal value of the (primal) SDP is at least $\mathcal{E}^*(M)$.

Let us check now that they actually coincide, for which we consider $Z \in \mathcal{B}(\mathbb{C}^{|X|} \oplus \mathbb{C}^{|Y|})_+$ expressed as

$$Z = \begin{pmatrix} R & K \\ K^* & S \end{pmatrix}$$

with $R \in \mathcal{B}(\mathbb{C}^{|X|})_+$, $S \in \mathcal{B}(\mathcal{C}^{|Y|})_+$, $K \in \mathcal{B}(\mathbb{C}^{|X|}, \mathbb{C}^{|Y|})$. Moreover, note that

$$\Delta(Z) = \mathbb{1}_{\mathbb{C}^{|X|} \oplus \mathbb{C}^{|Y|}} \quad \leftrightarrow \quad R, S \text{ only having 1s in diagonal.} \tag{2.3.11}$$

Since K might not have real number entries, we write

$$\langle H, Z \rangle = \frac{1}{2} \langle D, K \rangle + \frac{1}{2} \langle D^*, K^* \rangle = \langle D, M \rangle \tag{2.3.12}$$

with $M = \frac{K+K^*}{2}$, as D has real entries. From this it follows that

$$\frac{1}{2} \begin{pmatrix} R & K \\ K^* & S \end{pmatrix} + \frac{1}{2} \begin{pmatrix} R & K \\ K^* & S \end{pmatrix}^T = \begin{pmatrix} \frac{R+R^T}{2} & K \\ K^* & \frac{S+S^T}{2} \end{pmatrix} \geq 0 \,, \tag{2.3.13}$$

and the diagonal entries of both $\frac{R+R^T}{2}$ and $\frac{S+S^T}{2}$ are 1. Then, $\langle D, M \rangle$ is no larger than $\mathcal{E}^*(G)$.

2.3.2 Dual problem

Now, let us check what the outcome of the dual problem is. Without loss of generality, let us assume that

$$\omega = \frac{1}{2} \begin{pmatrix} \text{Diag}(u) & 0 \\ 0 & \text{Diag}(v) \end{pmatrix} \geq 0 \,,$$

for $u \in \mathbb{R}^{|X|}$ and $v \in \mathbb{R}^{|Y|}$. The objective function is given by

$$\text{Tr}[\omega] = \frac{1}{2} \sum_{x \in X} u(x) + \frac{1}{2} \sum_{y \in Y} v(y) \,,$$

and then the constraint in such an objective function is equivalent to

$$\omega = \begin{pmatrix} \text{Diag}(u) & -D \\ -D & \text{Diag}(v) \end{pmatrix} \geq 0 \,.$$

2.3.2.1 SDP for XOR games, simplified

Considering the information presented just above for both formulations of the SDPs for the XOR games, we can write the following simplified reformulations:

Primal Problem		**Dual Problem**	
max	$\langle D, M \rangle$	min	$\frac{1}{2} \sum_{x \in X} u(x) + \sum_{y \in Y} v(y)$
subject to	$\begin{pmatrix} R & M \\ M^* & S \end{pmatrix} \geq 0$	subject to	$\begin{pmatrix} \text{Diag}(u) & -D \\ -D^* & \text{Diag}(v) \end{pmatrix} \geq 0$
	$R(x,x) = 1 \ \forall x$		$u \in \mathbb{R}^{|X|}$, $v \in \mathbb{R}^{|Y|}$
	$S(y,y) = 1 \ \forall y$		
	$R \geq 0$		
	$S \geq 0$		
	$M \in \mathcal{B}(\mathbb{R}^{|Y|}, \mathbb{R}^{|X|})$		

Example. **(CHSH game)** For the primal problem, in this case,

$$D = \frac{1}{4} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} .$$

We consider the following candidate for M:

$$M = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} .$$

For $R = S = \mathbb{1}$, it is clear that

$$Z = \begin{pmatrix} R & M \\ M^* & S \end{pmatrix} \geq 0\,.$$

Moreover, since the diagonal of R and S is clearly composed only of 1s, then Z is primal feasible and $\langle D, M\rangle = 1/\sqrt{2}$, which coincides with the known entangled bias for the CHSH game.

Now, for the dual problem, we can consider

$$u = \left(\frac{1}{2\sqrt{2}}, \frac{1}{2\sqrt{2}}\right) \quad , \quad v = \left(\frac{1}{2\sqrt{2}}, \frac{1}{2\sqrt{2}}\right)\,,$$

which is dual feasible, and then, the objective value is

$$\frac{1}{2}\sum_{x\in X} u(x) + \sum_{y\in Y} v(y) = \frac{1}{\sqrt{2}}\,.$$

Therefore, in this way we also obtain that the entangled bias for the CHSH game is $\frac{1}{\sqrt{2}}$.

Chapter 3

Quantum Channels

In this chapter, we will introduce the main elements for the transmission of quantum information, namely quantum channels. For an overview on this topic, we recommend the lecture notes [24], which provide a wide collection of properties and results concerning completely positive and trace-preserving maps, a.k.a. quantum channels.

However, before starting with the first definitions and properties of quantum channels, we need to recall some notions which might be of use during this chapter. Since they are relatively external to the main topic of this chapter, we collect all of them in a section of preliminaries.

3.1 Preliminaries

3.1.1 Bloch Sphere

In this subsection, we introduce the Bloch sphere, which is frequently used in many applications in quantum information theory and will help us in the next pages to describe some aspects in an easier way. For that, we need to recall the Pauli-Matrices

$$\begin{aligned} \mathbb{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \\ Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \end{aligned}. \tag{3.1.1}$$

They form a basis of $M_{2\times 2}(\mathbb{C})$. We find that for $M \in M_{2\times 2}(\mathbb{C})$

$$M = \frac{1}{2}(x_0\,\mathbb{1} + \vec{x}\cdot\vec{\gamma}) \tag{3.1.2}$$

with $x_0 = \mathrm{Tr}[M]$, $\vec{x} \in \mathbb{C}^3$, and $\vec{x}\cdot\vec{\sigma} = \sum_{i=1}^{3} x_i\sigma_i$. Note that

1. M is Hermitian if and only if x_0 and $\vec{x}$ are real.
2. $M \geq 0$ if and only if $\|\vec{x}\|_2 \leq x_0$

In particular for $\rho \in \mathcal{S}(\mathbb{C}^{2\times 2})$ we find

$$\rho = \frac{1}{2}(\mathbb{1} + \vec{x}\cdot\vec{\gamma}) \tag{3.1.3}$$

with $\|\vec{x}\|_2 \leq 1$. We call this the *Bloch ball* and further for $\|\vec{x}\|_2 = 1$ the *Bloch sphere*. The closer we get to the center, the more mixed ρ becomes.

We further define

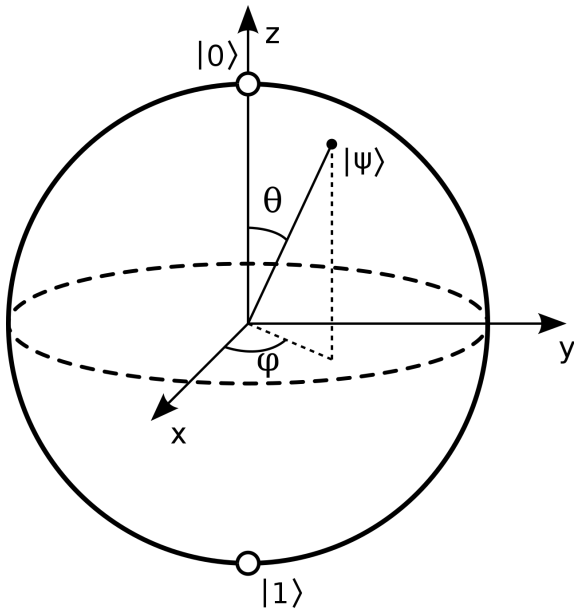


fig. 3.1: Bloch sphere.

- The *purity* as

$$\mathrm{Tr}[\rho^2] = \frac{1}{2}(1 + \|\vec{x}\|^2) \tag{3.1.4}$$

- For $\vec{x} = 0$, we have the maximally mixed state.

Then, we have the following connection between two different classes of objects:

$$\begin{array}{ccc} \text{Orthogonal rotations} & \longleftrightarrow & \text{Unitaries at the level} \\ \text{on the Bloch ball} & & \text{of density matrices} \end{array}$$

using the identification map

$$\rho \mapsto U_{\vec{x},\theta}\, \rho\, U^*_{\vec{x},\theta} \tag{3.1.5}$$

with

$$U_{\vec{x},\theta} = e^{-i\theta\frac{\vec{x}\cdot\vec{\gamma}}{2}} = \mathbb{1}\cos(\theta/2) - i\vec{x}\cdot\vec{\gamma}\sin(\theta/2)\,. \tag{3.1.6}$$

Since every unitary is of this form up to a phase, this in particular shows the following well-known relation between $SO(3)$ and $SU(2)$:

$$SO(3) \cong SU(2)/\{+1, -1\}\,. \tag{3.1.7}$$

Moreover, in general, we have for qubits

$$|\psi\rangle = \cos(\theta/2)\,|0\rangle + e^{i\varphi}\sin(\theta/2)\,|1\rangle \tag{3.1.8}$$

for $0 \le \theta \le \pi$ and $0 \le \varphi \le 2\pi$.

3.1.2 Born's Rule

Any experiment can be described following Born's rule as in the following Figure:

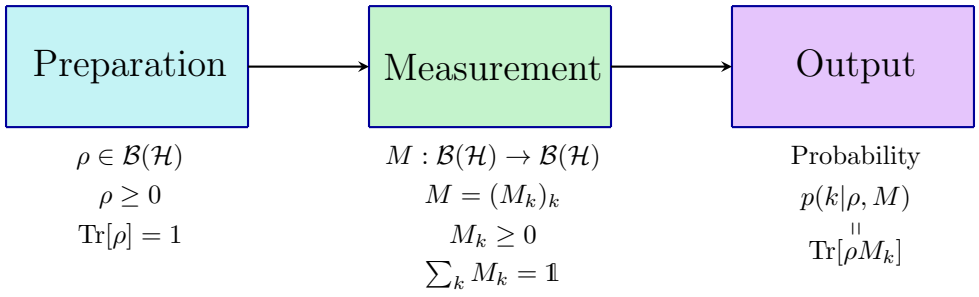


fig. 3.2: Schematic representation of an experiment.

3.1.3 Composite systems

Let $\mathcal{H}_1, \mathcal{H}_2$ finite dimensional Hilbert spaces. We consider the composite Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2$.

Theorem. 3.1.1 (Schmidt decomposition) *Let $|\phi\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2$, then there exists a set of $\{|e_i\rangle\} \subset \mathcal{H}_1$, $\{|f_i\rangle\} \subset \mathcal{H}_2$ and $\lambda_i \geq 0$ $\forall i$, such that*

$$|\phi\rangle = \sum_i \sqrt{\lambda_i}\, |e_i\rangle \otimes |f_i\rangle \tag{3.1.9}$$

Proof. We find in general

$$|\varphi\rangle = \sum_{k,l} \beta_{kl}\, |\psi_k\rangle \otimes |\varphi_l\rangle \tag{3.1.10}$$

for orthonormal basis $\{|\psi_k\rangle\} \subset \mathcal{H}_1$, $\{|\varphi_l\rangle\} \subset \mathcal{H}_2$. Through the singular value decomposition, we can decomposte the matrix

$$M = (\beta_{kl})_{kl} = U \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V \tag{3.1.11}$$

with Σ an $m \times m$ matrix and 0 a $(n-m) \times m$ matrix and $\Sigma \geq 0$ and hence

$$\beta_{kl} = \sum_i U_{ki} s_i V_{il} \tag{3.1.12}$$

which gives immediately

$$|\phi\rangle = \sum_i s_i \underbrace{\Big(\sum_k U_{ki} |\varphi_k\rangle \Big)}_{|e_i\rangle} \otimes \underbrace{\Big(\sum_l V_{il} |\varphi_l\rangle \Big)}_{|f_i\rangle} \tag{3.1.13}$$

□

Definition. 3.1.2 (Schmidt coefficients...)

...

Definition. 3.1.3 (Partial trace)
The partial trace is a linear map

$$\mathrm{Tr}_B : \mathcal{S}(\mathcal{H}_{AB}) \to \mathcal{S}(\mathcal{H}_A), \quad \rho_{AB} \mapsto \rho_A \tag{3.1.14}$$

defined by

$$\mathrm{Tr}[\rho_{AB}(M_A \otimes \mathbb{1}_B)] = \mathrm{Tr}[\rho_A M_A] \qquad \forall M_A \in \mathcal{B}(\mathcal{H}_A). \tag{3.1.15}$$

In the case that

- ρ_{AB} is a density matrix, ρ_A is called "reduced density matrix" of ρ_{AB} in A.

- $\rho_{AB} = \rho_A \otimes \rho_B$ we call ρ_{AB} a product state.

Proposition. 3.1.4 *Consider $\rho_{AB} \in \mathcal{S}(\mathcal{H}_{AB})$, $\rho_A := \mathrm{Tr}_B[\rho_{AB}]$, then*

1. $\mathrm{Tr}[\rho_A] = \mathrm{Tr}[\rho_{AB}]$
2. *$\rho_{AB} \geq 0$, then $\rho_A \geq 0$*
3. *From the first two we immediately get that ρ_{AB} density matrix, then ρ_A is a density matrix.*
4. $$\langle \varphi_i, \rho_A \varphi_i \rangle = \sum_k \langle \varphi_i \otimes \psi_k, \rho_{AB} \varphi_i \otimes \psi_k \rangle \tag{3.1.16}$$
5. *If $\rho_{AB} = |\phi\rangle\langle\phi|$ with Schmidt decomposition:*
$$|\phi\rangle = \sum_i \sqrt{\lambda_i}\, |e_i\rangle \otimes |f_i\rangle \tag{3.1.17}$$
then
$$\rho_A = \sum_i \lambda_i |e_i\rangle\langle e_i| \tag{3.1.18}$$
6. *If $\rho_{AB} = \tilde{\rho}_A \otimes \tilde{\rho}_B$ with $\mathrm{Tr}[\tilde{\rho}_B] = 1$, then $\tilde{\rho}_A = \rho_A$*

Proof. 1. $\mathrm{Tr}[\rho_A] = \mathrm{Tr}[\rho_A \mathbb{1}] = \mathrm{Tr}[\rho_{AB}(\mathbb{1} \otimes \mathbb{1})] = \mathrm{Tr}[\rho_{AB}]$

2. $\langle \psi, \rho_A \psi \rangle = \mathrm{Tr}[\rho_A |\psi\rangle\langle\psi|] = \mathrm{Tr}[\rho_{AB}(|\psi\rangle\langle\psi|_A \otimes \mathbb{1}_B)] \geq 0$

3. The result is immediate.

4. $$\begin{aligned} \langle \varphi_i, \rho_A \varphi_i \rangle &= \mathrm{Tr}[\rho_A |\varphi_i\rangle\langle\varphi_i|] = \mathrm{Tr}[\rho_{AB}(|\varphi_i\rangle\langle\varphi_i|] \\ &= \sum_{k,l} \langle \varphi_l \otimes \psi_k, \rho_{AB}(|\varphi_i\rangle\langle\varphi_i| \otimes \mathbb{1}) \varphi_l \otimes \psi_k \rangle \\ &= \sum_k \langle \varphi_i \otimes \psi_k, \rho_{AB} \varphi_i \otimes \psi_k \rangle \end{aligned} \tag{3.1.19}$$

5. $\langle e_i, \rho_A e_i \rangle = \sum_k \langle e_i \otimes f_k, \phi \rangle \langle \phi, e_i \otimes f_k \rangle = \delta_{ij} \sqrt{\lambda_i} \sqrt{\lambda_j}$

6. For all $X_A \in \mathcal{B}(\mathcal{H}_A)$
$$\begin{aligned} \mathrm{Tr}[\rho_A X_A] &= \mathrm{Tr}[\rho_{AB}(X_A \otimes \mathbb{1}_B)] = \mathrm{Tr}[\tilde{\rho}_A \otimes \tilde{\rho}_B (X_A \otimes \mathbb{1})] \\ &= \mathrm{Tr}[\tilde{\rho}_A X_A]\, \mathrm{Tr}[\tilde{\rho}_B] = \mathrm{Tr}[\tilde{\rho}_A X_A] \end{aligned} \tag{3.1.20}$$

□

3.1.4 Measurement on Subsystem

Let $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$. If

$$M_A : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(H_{AB}) \qquad \text{POVM } \{M_A\} \mapsto \text{POVM } \{M_A \otimes \mathbb{1}_B\} \tag{3.1.21}$$

Theorem. 3.1.5 (Naimark) *For every POVM $(M_x)_{x=1}^m \subseteq \mathcal{B}(\mathcal{H})$, there is a $|\psi\rangle \in \mathbb{C}^m$ and a probability measure $(P_x)_{x=1}^m \subseteq \mathcal{B}(\mathcal{H} \otimes \mathbb{C}^m)$ such that*

$$\mathrm{Tr}[(\rho \otimes |\psi\rangle\langle\psi| P_x] \qquad \forall x = 1, \ldots, m \tag{3.1.22}$$

Proof. Let $V : \mathcal{H} \to \mathcal{H} \otimes \mathbb{C}^m$, $V = \sum_{x=1}^{m} \sqrt{M_x} \otimes |x\rangle$ for an ONB $\{|x\rangle\}$ in $\mathbb{C}^m$. Then

$$V^*V = \sum_{x=1}^{m} M_x = \mathbb{1} \tag{3.1.23}$$

hence V is an isometry and

$$V = U(\mathbb{1} \otimes |\psi\rangle) \tag{3.1.24}$$

for some $U \in \mathcal{B}(\mathcal{H} \otimes \mathbb{C}^m)$ and $|\psi\rangle \in \mathbb{C}^m$. Then,

$$\mathrm{Tr}[\rho M_x] = \mathrm{Tr}[V\rho V^*(\mathbb{1} \otimes \langle x\rangle\, x)] = \mathrm{Tr}[(\rho \otimes |\psi\rangle\langle\psi|) \underbrace{U^*(\mathbb{1} \otimes |x\rangle\langle x|)U}_{P_X}] \tag{3.1.25}$$

□

3.2 Quantum channels

They are used to describe the evolution or process in quantum systems (QIT). Schematically we can represent them by

Example.

- Closed system evolution:

$$\rho \mapsto U\rho U^* \tag{3.2.1}$$

  with U a unitary.

- Open system evolution

$$\rho \mapsto \mathrm{Tr}_E[U(\rho \otimes \rho_E)U^*] \tag{3.2.2}$$

Definition. 3.2.1 (Quantum channel)
A quantum channel is a linear map $T : \mathcal{S}(\mathcal{H}_{in}) \to \mathcal{S}(\mathcal{H}_{out})$ such that

1. Trace preserving, i.e. $\mathrm{Tr}[T(\rho)] = \mathrm{Tr}[\rho]\ \forall \rho \in \mathcal{S}(\mathcal{H}_{in})$
2. Positive: $\rho \geq 0$, then $T(\rho) \geq 0$.
3. Completely positive: For all $n \in \mathbb{N}_0$ $T \otimes \mathbb{1}_n$ is positive, with $\mathbb{1}_n$ the identity map on $\mathcal{B}(\mathbb{C}^n)$

i.e. a completely positive trace preserving (CPTP) map.

Example.

- Unitary evolution $\rho \mapsto U\rho U^*$.
- Adding an ancilla $\rho \mapsto \rho \otimes \rho_E$.
- Partial trace.

Definition. 3.2.2 (Maximally entangled state)
The maximally entangled state is given by

$$|\phi\rangle = \frac{1}{\sqrt{d}} \sum_{k=1}^{d} |kk\rangle \tag{3.2.3}$$

Definition. 3.2.3 (Choi-Jamiolkowski matrix)
Let $T : \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^d)$ linear, then the Choi-Jamiolkowski matrix of T is given by

$$C := (T \otimes \mathbb{1}_d)(|\phi\rangle\langle\phi|) \tag{3.2.4}$$

Remark. C determines T by

$$\langle ij, Ckl\rangle = \frac{1}{d}\sum_{m,n=1}^{d} \langle i, T(|n\rangle\langle m|)k\rangle \langle j, n\rangle \langle m, k\rangle = \frac{1}{d}\langle i, T(|j\rangle\langle l|)k\rangle \tag{3.2.5}$$

Theorem. 3.2.4 (Characterisation of quantum channels) *Let $T : \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^{d'})$ linear. Then the following are equivalent.*

1. *T is a quantum channel.*

2. *$C \geq 0$ and $\mathrm{Tr}_1[C] = \frac{1}{d}$, with C the Choi-Jambilowski matrix of T:*

$$C := (T \otimes \mathbb{1}_d)(|\phi\rangle\langle\phi|) \tag{3.2.6}$$

and

$$|\phi\rangle = \frac{1}{\sqrt{d}}\sum_{k=1}^{d} |kk\rangle \tag{3.2.7}$$

the maximally entangled state.

3. *Kraus decomposition:*

$$T(\rho) = \sum_{k=1}^{dd'} A_k \rho A_k^* \tag{3.2.8}$$

with

$$\sum_{k=1}^{dd'} A_k^* A_k = \mathbb{1} \tag{3.2.9}$$

4. *Stinespring dilation:*

$$T(\rho) = \mathrm{Tr}_2[U(\rho \otimes |\psi\rangle\langle\psi|)U^*] \tag{3.2.10}$$

with U a unitary on $\mathbb{C}^d \otimes \mathbb{C}^{dd'}$ and $|\psi\rangle$ a state.

Proof. 1. $\Rightarrow$ 2. $C \geq 0$ follows from T being completely positive

$$\begin{aligned}
\mathrm{Tr}_1[C] &= \frac{1}{d}\sum_{n,m=1}^{d} \mathrm{Tr}[T(|n\rangle\langle m|)]|n\rangle\langle m| \\
&= \frac{1}{d}\sum_{n,m=1}^{d} \underbrace{\mathrm{Tr}[|n\rangle\langle m|]}_{\delta_{nm}} |n\rangle\langle m| \\
&= \frac{1}{d}\sum_{n,m} \delta_{nm}|n\rangle\langle m| \\
&= \frac{1}{d}\sum_{n=1}^{d} |n\rangle\langle n| \\
&= \mathbb{1}
\end{aligned} \tag{3.2.11}$$

2. $\Rightarrow$ 3. We use

- $(A \otimes \mathbb{1})|\phi\rangle = (\mathbb{1} \otimes A^T)|\phi\rangle \ \forall A \in \mathcal{B}(\mathbb{C}^d)$
- $\forall |\psi\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$, $\exists A$ s.t. $|\psi\rangle = (A \otimes \mathbb{1})|\phi\rangle$

Then since $C \geq 0$,

$$\begin{aligned} C &= \sum_{k=1}^{dd'} |\psi_k\rangle\langle\psi_k| \\ &= \sum_{k=1}^{dd'} \underbrace{(A_k \otimes \mathbb{1})}_{\mathbb{1}\otimes A^T} |\phi\rangle\langle\phi| \underbrace{(A_k^* \otimes \mathbb{1})}_{\mathbb{1}\otimes \overline{A}} \\ &= (T \otimes \mathbb{1})(|\phi\rangle\langle\phi|) \end{aligned} \tag{3.2.12}$$

and

$$\begin{aligned} \frac{\mathbb{1}}{d} = \mathrm{Tr}_1[C] &= \sum_{n=1}^{d} \langle n, Cn\rangle \\ &= \frac{1}{d} \sum_{n=1}^{d} \sum_{k=1}^{d'} A_k^T |n\rangle\langle n| \overline{A}_k \\ &= \frac{1}{d} \sum_{k=1}^{d'} A_k^T \overline{A}_k \\ &= \frac{1}{d} \sum_{k=1}^{d'} \tilde{A}_k^* \tilde{A}_k \end{aligned} \tag{3.2.13}$$

To conclude, we just take $\tilde{A}_k := \overline{A}_k$. This concludes this step.

3. $\Rightarrow$ 4. Define $V = \sum\limits_{k=1} A_k \otimes |k\rangle$, it is an isometry ($V^*V = \mathbb{1}$). $\{|k\rangle\}$ is an ONB of $\mathbb{C}^{dd'}$.

$$\begin{aligned} \mathrm{Tr}_E[V\rho V^*] &= \sum_{kl} A_k \rho A_l^* \underbrace{\mathrm{Tr}[|k\rangle\langle l|]}_{\delta_{kl}} \\ &= \sum_{k} A_k \rho A_k^* \\ &= T(\rho) \end{aligned} \tag{3.2.14}$$

Hence, $T(\rho) = \mathrm{Tr}_E[V\rho V^*]$. We choose $V = U(\mathbb{1} \otimes |\psi\rangle)$ for some $|\psi\rangle$ pure state and some unitary U.

4. $\Rightarrow$ 1. Now it remains to show

$$T(\rho) = \mathrm{Tr}_E[U(\rho \otimes |\psi\rangle\langle\psi|)U^*] \tag{3.2.15}$$

implies T being a quantum channel. We set

$$\rho \mapsto \rho \otimes |\psi\rangle\langle\psi| \mapsto U(\rho \otimes |\psi\rangle\langle\psi|)U^* \mapsto \mathrm{Tr}_E[U(\rho \otimes |\psi\rangle\langle\psi|)U^*] \tag{3.2.16}$$

The above mappings are all quantum channels and, hence, their composition is a quantum channel as well.

□

Remark.

- The number k in the Kraus decomposition is called Kraus rank of T (it coincides with the Choi rank). It is not to be confused with the rank of T as a map.
- T is a completely positive linear map. Hence, there is always a representation for T with $r = \mathrm{rank}(\tau)$ orthogonal Kraus operators (HS product).

$$\tau := (T \otimes \mathbb{1}_d)(|\phi\rangle\langle\phi|) \tag{3.2.17}$$

the Choi-Jamilowski state.

- Two sets of Kraus operators $\{K_j\}_j^n$ and $\{\tilde{K}_l\}_l^m$ represent the same map T, if and only if $\exists$ a unitary map s.t.
$$K_j = \sum_l U_{kl}\tilde{K}_l \tag{3.2.18}$$
(the smallest set is complemented with zeros).

Proposition. 3.2.5 (Equivalence of ensembles) *Two ensembles of vectors $\{|\psi_j\rangle\}$ and $\{|\psi_l\rangle\}$ (not necessarily normalised) satisfy*
$$\sum_j |\psi_j\rangle\langle\psi_j| = \sum_l |\varphi_l\rangle\langle\varphi_l| \tag{3.2.19}$$
if and only if $\exists U$ a unitary such that $|\psi_j\rangle = \sum_l U_{jl}|\varphi_l\rangle$

Proof. $\Leftarrow$ Trivial.

$\Rightarrow$ Without loss of generality we assume that $\sum_j |\psi_j\rangle\langle\psi_j|$ represents a density matrix ρ. Through purification, one obtains
$$\rho = \mathrm{Tr}_B[|\Psi\rangle\langle\Psi|] \tag{3.2.20}$$
for $|\Psi\rangle = \sum_j |\psi_j\rangle \otimes |j\rangle$ and $|\Phi\rangle = \sum_l |\varphi_l\rangle \otimes |l\rangle$. $|\Psi\rangle$ and $|\Phi\rangle$ (from the Schmidt decomposition) differ only in a unitary (an isometry):
$$|\Psi\rangle = (\mathbb{1} \otimes U)|\Phi\rangle . \tag{3.2.21}$$
Take $\langle j|$:
$$|\psi_j\rangle = \sum_l U_{jl}|\varphi_l\rangle \tag{3.2.22}$$
□

3.2.1 Examples of quantum channels

3.2.1.1 Depolarizing channel

In three dimensions we can get 3 kinds of errors, if we restrict to the two dimensional case. Those are

1. Bit flip error, which can be modeled by the X Pauli matrix $\begin{array}{ccc} |0\rangle & \mapsto & |1\rangle \\ |1\rangle & \mapsto & |0\rangle \end{array}$.
2. Phase flip error, modeled by Z $\begin{array}{ccc} |0\rangle & \mapsto & |0\rangle \\ |1\rangle & \mapsto & -|1\rangle \end{array}$.
3. Combination of both: Y

As a unitary representation (from $\mathcal{H}_A \to \mathcal{H}_{AE}$ with the environment E of dimension four) we get
$$U_{A\to AE} : |\psi\rangle_A \mapsto \sqrt{1-p}\,|\psi\rangle_A \otimes |0\rangle_E + \sqrt{\frac{p}{3}}\left(X|\psi\rangle_A \otimes |1\rangle_E + Y|\psi\rangle_A \otimes |2\rangle_E + Z|\psi\rangle_A \otimes |3\rangle_E\right) \tag{3.2.23}$$
Or in the operator representation
$$M_a := {}_E\langle a|\, U_{A\to AE} \tag{3.2.24}$$
with ${}_E\langle a| \in \{{}_E\langle 0|, {}_E\langle 1|, {}_E\langle 2|, {}_E\langle 3|\}$ and
$$M_0 = \sqrt{1-p}\,\mathbb{1},\ M_1 = \sqrt{\frac{p}{3}}X,\ M_2 = \sqrt{\frac{p}{3}}Y,\ M_3 = \sqrt{\frac{p}{3}}Z \tag{3.2.25}$$

It is straight forward to see that

$$\sum M_a^* M_a = ((1-p) + \frac{p}{3} + \frac{p}{3} + \frac{p}{3})\,\mathbb{1} = \mathbb{1}\,. \tag{3.2.26}$$

The depolarisation channel is given by

$$\rho \mapsto \rho' = (1-p)\rho + \frac{p}{3}(X\rho X + Y\rho Y + Z\rho Y) \tag{3.2.27}$$

In general for an arbitrary dimension $D \in \mathbb{N}$, the decoding channel becomes

$$\rho \mapsto (1-p)\rho + p\sigma \tag{3.2.28}$$

with σ usually taken to be $\frac{\mathbb{1}}{d}$.

3.2.1.2 Phase damping channel

The phase damping channel in operator representation is given by

$$\rho \in \mathbb{C}^{2\times 2} \qquad \rho = \begin{pmatrix} \rho_{00} & \rho_{01} \\ \rho_{10} & \rho_{11} \end{pmatrix} \mapsto \tag{3.2.29}$$

In state representation we get for the ONB of the environment $\{|0\rangle_E, |1\rangle_E, |2\rangle_E\}$

$$\begin{array}{rcl} |0\rangle_A & \mapsto & \sqrt{1-p}\,|0\rangle_A \otimes |0\rangle_E + \sqrt{p}\,|0\rangle_A \otimes |1\rangle_E \\ |1\rangle_A & \mapsto & \sqrt{1-p}\,|1\rangle_A \otimes |0\rangle_E + \sqrt{p}\,|1\rangle_A \otimes |2\rangle_E \end{array}\,. \tag{3.2.30}$$

In Kraus operators, we evaluate

$$\begin{gathered} A_0 = \sqrt{1-p}\,\mathbb{1} = \begin{pmatrix} \sqrt{1-p} & 0 \\ 0 & \sqrt{1-p} \end{pmatrix}, \quad A_1 = \sqrt{p}|0\rangle\langle 0| = \begin{pmatrix} \sqrt{p} & 0 \\ 0 & 0 \end{pmatrix} \\ A_2 = \sqrt{p}|1\rangle\langle 1| = \begin{pmatrix} 0 & 0 \\ 0 & \sqrt{p} \end{pmatrix} \end{gathered} \tag{3.2.31}$$

With

$$T(\rho) = \sum_{k=0}^{2} A_k \rho A_k^* \Rightarrow (\tag{3.2.32}$$

$$T(\rho) = \left(1 - \frac{1}{2}p)\right)\rho + \frac{1}{2}pZ\rho Z \tag{3.2.33}$$

3.2.2 Entanglement breaking channels

Definition. 3.2.6 (Separability)
Let $\rho \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$ is separable, if and only if it is a convex combination of products of the form

$$\rho = \sum_i \lambda_i \rho_i^A \otimes \rho_i^B \tag{3.2.34}$$

otherwise it is entangled.

Definition. 3.2.7 (Breaking of entanglement)
A quantum channel T is entanglement breaking if its Choi matrix is separable. This is equivalent to the existence of a POVM $\{M_x\}$ and a set of density matrices $\{\rho_x\}$ such that

$$T(\rho) = \sum_x \mathrm{Tr}[M_x \rho]\rho_x \tag{3.2.35}$$

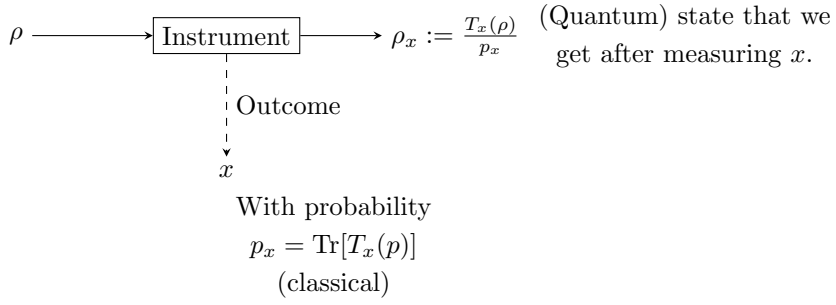


fig. 3.3: Scheme of an instrument.

3.2.3 Instruments

Definition. 3.2.8 (Instrument)
An instrument is a set of CPTP maps (quantum channels) $\{T_x\}$ whose sum is $\sum_x T_x$ is trace preserving. x can be interpreted as the outcome of a measurement with probability

$$p_x = \mathrm{Tr}[T_x(\rho)], \quad \rho \mapsto \frac{T_x(\rho)}{p_x}. \tag{3.2.36}$$

In that sense it encompasses the notion of quantum channel and POVMs in the following way

$$\begin{cases} \text{Quantum channel:} & \text{Ignore the measurement outcome} \\ & \rho \mapsto \sum_x p_x \rho_x = \sum_x T_x(\rho) =: T \\ \text{POVM:} & \text{Ignore the quantum system} \\ & p_x = \mathrm{Tr}[T_x(\rho)] = \mathrm{Tr}[T_x(\rho)\,\mathbb{1}] = \mathrm{Tr}[\rho T_x^*(\mathbb{1})] =: \mathrm{Tr}[\rho M_x] \\ & \{M_x\}_x \text{ is a POVM} \end{cases} \tag{3.2.37}$$

Remark. Instruments can be viewed as special case of quantum channels by assigning to them

$$\rho \mapsto \sum_{x \in X} T_x(\rho) \otimes |x\rangle\langle x| \tag{3.2.38}$$

with $\{|x\rangle\}$ an orthonormal basis.

Theorem. 3.2.9 (No information without disturbance) *Consider an instrument $\{T_x\}_{x\in X}$ such that $\forall \rho$*

$$\rho \mapsto \sum_x p_x \rho_x = \rho \tag{3.2.39}$$

Then, ρ_x is independent of ρ, i.e.

$$\mathrm{Tr}[T_x(\rho)] = \mathrm{Tr}[T_x(\rho')] \tag{3.2.40}$$

for all ρ, ρ' density operators.

Proof. Based on the Choi-Jamiolkowski representation for channels. Since for all density operators ρ

$$\rho = \sum_x p_x \rho_x = \sum_x T_x(\rho). \tag{3.2.41}$$

This means that $\sum_x T_x = \mathbb{1}$. This further gives us that also the Choi Jamiolkowski matrices coincide.

$$\begin{aligned} T_x &\mapsto \tau_x := (T_x \otimes \mathbb{1})(|\phi\rangle\langle\phi|) \\ \sum_x T_x &\mapsto \sum_x \tau_x = (\mathbb{1} \otimes \mathbb{1})(|\phi\rangle\langle\phi|) = |\phi\rangle\langle\phi| \end{aligned} \tag{3.2.42}$$

Moreover, since $\tau_x \geq 0 \ \forall x$, we find that there exists $q_x \geq 0$ such that

$$\tau_x = q_x |\phi\rangle\langle\phi|, \tag{3.2.43}$$

with $\sum_x q_x = 1$ □

3.3 Open system representation

3.3.0.1 Partial order of CP maps

We write $T_2 \geq T_1$, if and only if $T_2 - T_1$ is completely positive. By Choi-Jamiolkowski representation, this is equivalent to $\tau_2 \geq \tau_1$, i.e. $\tau_2 - \tau_1$ positive semi-definite.

Theorem. 3.3.1 (Relation CP maps) *Let for $i = 1, 2$*

$$T_i : \mathbb{C}^{d' \times d'} \to \mathbb{C}^{d \times d} \tag{3.3.1}$$

CP linear maps be given. Assume that $T_2 \geq T_1$. If for $i = 1, 2$

$$V_i : \mathbb{C}^d \to \mathbb{C}^{d'} \otimes \mathbb{C}^{r_i} \tag{3.3.2}$$

provide Stinespring dilations for T_i, then there is a contraction $[T_i(A) = V_i^*(A \otimes \mathbb{1}_{r_i})V_i]$,

$$C : \mathbb{C}^{r_2} \to \mathbb{C}^{r_1} \quad \text{such that} \quad V_1 = (\mathbb{1}_{d'} \otimes C)V_2. \tag{3.3.3}$$

If V_2 belongs to a minimal dilation then C is unique.

Proof. We use the equivalence $T_2 \geq T_1 \Leftrightarrow \tau_2 \geq \tau_1$. Starting with defining

$$W_i := (\mathbb{1}_{r_i} \otimes \langle\phi|)(V_i \otimes \mathbb{1}_{d'}) \in \mathcal{B}(\mathbb{C}^d \otimes \mathbb{C}^{d'} \otimes \mathbb{C}^{r_i}), \tag{3.3.4}$$

we find that $\forall\, |\psi\rangle \in \mathbb{C}^d \otimes \mathbb{C}^{d'}$

$$\|W_2 |\psi\rangle\|^2 = \langle\psi, \tau_2\psi\rangle \geq \langle\varphi, \tau_1\varphi\rangle = \|W_1 |\varphi\rangle\|^2 . \tag{3.3.5}$$

This gives us the existence of $C : \mathbb{C}^{r_2} \to \mathbb{C}^{r_1}$ a contraction ($C^*C \leq \mathbb{1}$) such that

$$W_1 = CW_2. \tag{3.3.6}$$

Since $V_i \to W_i$ is one to one, $V_1 = (\mathbb{1}_{d'} \otimes C)V_2$. If $r_2 = \mathrm{rank}(\tau_2)$, then W_2 is surjective and hence C is uniquely determined. □

Theorem. 3.3.2 (Radon-Nikodym) *Let $\{T_i\}$ a set of CPTP maps such that $\sum_i T_i = T \in \mathcal{B}(\mathbb{C}^{d' \times d'}, \mathbb{C}^{d \times d})$ with Stinespring representation*

$$T(A) = V^*(A \otimes \mathbb{1}_r)V. \tag{3.3.7}$$

Then there exists a set of non negative operators $P_i \in \mathbb{C}^{r \times r}$, $\sum_i P_i = \mathbb{1}_r$, such that

$$T_i(A) = V^*(A \otimes P_i)V. \tag{3.3.8}$$

Remark. Since $T = \sum_i T_i$ we find that

$$T(A) = \sum_i V^*(A \otimes P_i)V \tag{3.3.9}$$

with $\{P_i\}$ a POVM. To say it with words: We have found a possibility to represent a quantum channel using a POVM.

Proposition. 3.3.3 (Quantum steering) *Let $\rho \in \mathcal{B}(\mathcal{H}_A)$ density operator with purification*

$$|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B \tag{3.3.10}$$

(i.e. $\mathrm{Tr}_B[|\psi\rangle\langle\psi|] = \rho$). Then for every convex combination $\rho = \sum_i \lambda_i \rho_i$, then there is an instrument $\{T_i\}_i$

$$T_i : \mathcal{B}(\mathcal{H}_B) \to \mathcal{B}(\mathcal{H}_B) \tag{3.3.11}$$

such that

$$\lambda_i \rho_i = \mathrm{Tr}_B[(\mathbb{1} \otimes T_i)(|\psi\rangle\langle\varphi|] \tag{3.3.12}$$

Proof. (sketch). The idea of the proof is just to form Schmidt decompositions of $|\psi\rangle$ and applying the transposition map. □

3.3.0.2 Open system representation

Theorem. 3.3.4

- *Let $T : \mathbb{C}^{d\times d} \to \mathbb{C}^{d'\times d'}$ be a CPTP amp. Then there exists $U \in \mathbb{C}^{dd'\times dd'}$ and a normalised vector $|\varphi\rangle \in \mathbb{C}^{d'} \otimes \mathbb{C}^d$ such that $\forall \rho$*

  $$T(\rho) = \mathrm{Tr}_E[U(\rho \otimes |\varphi\rangle\langle\varphi|)U^*], \tag{3.3.13}$$

  with Tr_E the partial trace over the first two tensor factors of $\mathbb{C}^d \otimes \mathbb{C}^{d'} \otimes \mathbb{C}^{d'}$.

- *Equivalently, there exists an isometry*

  $$V : \mathbb{C}^d \to \mathbb{C}^{d'} \otimes \mathbb{C}^r \tag{3.3.14}$$

  with $r \geq rank(\tau)$, $\forall A \in \mathbb{C}^{d'\times d'}$,

  $$T^*(A) = V^*(A \otimes \mathbb{1}_r)V \tag{3.3.15}$$

- *If $T = \sum_i T_i$ can be decomposed into CPTP maps, there exists a POVM $\{P_i\}$ such that*

  $$T_i(\rho) = \mathrm{Tr}_E[(P_i \otimes \mathbb{1}_{d'})U(\rho \otimes |\phi\rangle\langle\phi|)U^*] \tag{3.3.16}$$

  If k_i is the Kraus rank associated to T_i then $k_i \leq rank(P_i)$ for all i.

Proof. We only give a sketch of a proof for the second statement as the other ones are straight forward with the considerations above. Hence, let τ the Choi matrix of T, consider its purification:

$$|\psi\rangle := (\mathbb{1}_d \otimes U)(|\Omega\rangle \otimes |\varphi\rangle) \tag{3.3.17}$$

$\tau = \mathrm{Tr}_E[|\psi\rangle\langle\varphi|]$. The decomposition of $T = \sum_i T_i$ gives us $\tau = \sum_i \tau_i$. We conclude by applying Quantum Steering to

$$\ldots \tag{3.3.18}$$

□

Remark. V is an isometry ($V^*V = \mathbb{1}_d$) if and only if T is trace preserving.

3.4 Quantum hypothesis testing

Lets assume the following setting: We are given as set $\rho_1, \dots, \rho_n \in \mathcal{S}(\mathcal{H})$ of density operators with corresponding probabilities $p_1, \dots, p_n$ that satisfy $p_x \geq 0 \ \forall x = 1, \dots, n$ and $\sum_{x=1}^{n} p_x = 1$. This can be interpreted as a set of n hypothesis with corresponding a priori probability p_x. The goal is to discriminate among the hypothesis with a measurement described by a POVM $M = (M_x)_{x=1}^n \subset \mathcal{B}(\mathcal{H})$. Hence, we want to maximize

$$\mathcal{P}(M) := \sum_{x=1}^{n} \mathrm{Tr}[M_x \underbrace{p_x \rho_x}_{=:\sigma_x}] \tag{3.4.1}$$

over POVMs $M = (M_x)_{x=1}^n \in \mathcal{M}$. We set

$$\mathcal{P}(\mathcal{M}) := \sup_{M \in \mathcal{M}} P(M). \tag{3.4.2}$$

Definition. 3.4.1 (Maximum likelihood measurement)
The maximum likelihood measurement is defined as

$$L := \sum_x M_x \sigma_x = \sum_x M_x p_x \rho_x \tag{3.4.3}$$

With this definition, we can write $\forall M$

$$\mathcal{P}(M) = \mathrm{Tr}[L]. \tag{3.4.4}$$

Lemma. 3.4.2 (Existence of optimal measurement) *The supremum in $\mathcal{P}$ is always attained, i.e. there exists a measurement $\widehat{M}$ such that*

$$\mathcal{P}(\mathcal{M}) = \mathcal{P}(\widehat{M}) \tag{3.4.5}$$

Proof. We define

$$\mathcal{M} := \mathrm{span}\{(M_1, \dots, M_n) \cong M \ : \ M = (M_x)_{x=1}^n \text{ measures}\} \tag{3.4.6}$$

with $\mathcal{M} \subseteq \mathcal{B}(\mathcal{H})^n$. I.e. $\mathcal{M}$ is the space of n-outcome POVMs on the Hilbert space $\mathcal{H}$ equipped with ... □

Theorem. 3.4.3 *Let $(p_x)_{x=1}^n$ and $(p_x)_{x=1}^n$ as above. Then, for every $M = (M_x)_{x=1}^n$, $L = \sum_{x=1}^{n} M_x p_x \rho_x$ the following are equivalent*

1. *M is an optimal measurement, i.e*

$$\max_{M' = (M'_x)_{x=1}^n} \mathcal{P}(M') = \mathcal{P}(M) \tag{3.4.7}$$

2. *$\forall x = 1, \dots, n$, $\frac{1}{2}(L + L^*) \geq p_x \rho_x$*
3. *$\forall x = 1, \dots, n$, $L \geq p_x \rho_x$*
4. *$\exists K \in \mathcal{B}(\mathcal{H})$ such that $\forall x = 1, \dots, n$, $K \geq p_x \rho_x$ and $(K - p_x \rho_x) M_x = 0$*
5. *$\mathcal{P}(M) = \min\{\mathrm{Tr}[A] \ : \ A \in \mathcal{A}\}$, $\mathcal{A} := \{A \in \mathcal{S}(\mathcal{H}) \ : \ \forall x = 1, \dots, n, \ A \geq p_x \rho_x\}$*

Example. 1. **Commuting states** $\rho_1, \dots, \rho_n$ commuting states (mutually commute). This means that there exists an orthonormal basis $\{|i\rangle\}_{i=1}^n$ such that

$$\max_M P(M) = \sum_i \max_x \underbrace{\langle i, \rho_x i \rangle}_{\lambda_i^x} \tag{3.4.8}$$

2. **Uniformly distributed pure states** We assume that $\rho_1, \dots, \rho_n$ are pure states and that the associated a priory probability is $\frac{1}{n}$. We further assume that

$$\sum_{x=1}^{n} p_x \rho_x = \frac{\mathbb{1}}{d} \tag{3.4.9}$$

(i.e. in particular $d \leq n$). Lets consider $M_x = \frac{d}{n}\rho_x$ which clearly constitute $M = (M_x)_{x=1}^n$ a POVM which has an optimal measurement.

- $\rho_x^2 = \rho_x$, $L = \sum_{x=1}^{n} M_x p_x \rho_x$. We find for all

$$\begin{aligned} L &= \sum_{x=1}^{n} M_x p_x \rho_x = \frac{d}{n} \sum_{x=1}^{n} \underbrace{\frac{1}{n}}_{p_x} \rho_x^2 \\ &= \frac{d}{n^2} \sum_{x=1}^{n} \rho_x = \frac{d}{n} \underbrace{\sum_{x=1}^{n} p_x \rho_x}_{\mathbb{1}/d} = \frac{\mathbb{1}}{n} \\ &\geq \frac{1}{n}\rho_x = p_x \rho_x \quad \forall x = 1, \dots, n \end{aligned} \tag{3.4.10}$$

i.e. $\mathcal{P}(M) = \mathrm{Tr}[L] = \frac{d}{n}$.

3.4.1 Binary hypothesis testing

Let ρ_1, ρ_2 be density matrices with a priory probability p and $(1-p)$. Further $M = (M_1, M_2) \cong (\mathbb{1}, \mathbb{1} - P)$ a POVM (i.e. $M_1 + M_2 = \mathbb{1}$) with P an orthogonal projection. Assigning P to ρ_1 and $(\mathbb{1} - P) \to \rho_2$ the error becomes

$$\mathcal{E}(M) := p \,\mathrm{Tr}[\rho_1(\mathbb{1} - P)] + (1-p)\,\mathrm{Tr}[\rho_2 P] \tag{3.4.11}$$

Remark. It is rather obvious that for

$$\mathcal{P}(M) = p \,\mathrm{Tr}[\rho_1 P] + (1-p)\,\mathrm{Tr}[\rho_2(\mathbb{1} - P)] \tag{3.4.12}$$

we find

$$\mathcal{P}(M) + \mathcal{E}(M) = 1 \tag{3.4.13}$$

Theorem. 3.4.4 (Quantum Neyman-Pearson) *We find that in the above setting we have the inequality*

$$\mathcal{E}(M) \geq \frac{1}{2}(1 - \|p\rho_1 - (1-p)\rho_2\|_1) \tag{3.4.14}$$

with equality, if and only if P is a projection onto $(p_1\rho_1 - (1-p)\rho_2)_+$.

Proof. For every Hermitian A, we can write $A = A_+ + A_-$ and find

- $\mathrm{Tr}[A_+] = \frac{\|A\|_1 + \mathrm{Tr}[A]}{2}$, since $\|A\|_1 = \mathrm{Tr}[|A|] = \mathrm{Tr}[A_+ - A_-]$ and $\mathrm{Tr}[A] = \mathrm{Tr}[A_+ + A_-]$

This consideration allows us to write

$$\begin{aligned} \min_M \mathcal{E}(M) &= \min_M \{p \,\mathrm{Tr}[\rho_1(\mathbb{1} - P)] + (1-p)\,\mathrm{Tr}[\rho_2 P]\} \\ &= \min\{p - \mathrm{Tr}[P(p\rho_1 - (1-p)\rho_2)]\} \\ &= p - \max_M\{\mathrm{Tr}[P \underbrace{(p\rho_1 - (1-p)\rho_2)}_{A = A_+ + A_-}] \end{aligned} \tag{3.4.15}$$

Hence the maximum is attained if $PA_+ = A_+$ and $PA_- = 0$, i.e. P is an orthonormal projection onto $A_+ = (p\rho_1 + (1-p)\rho_2)_+$. This gives

$$\begin{aligned} &= p - \{\mathrm{Tr}[(p\rho_1 - (1-p)\rho_2)_+]\} \\ &= p - \frac{\|p\rho_1 - (1-p)\rho_2\|_1 + \mathrm{Tr}[p\rho_1] - \mathrm{Tr}[(1-p)\rho_2]}{2} \\ &= \frac{1}{2}(1 - \|p\rho_1 - (1-p)\rho_2\|_1) \end{aligned} \tag{3.4.16}$$

which concludes the proof. Alternatively we could just argue that for $\mathcal{P}(M)$, we can prove that $P = (p\rho_1 - (1-p)\rho_2)_+$ provides an optimal measurement as

$$L = Pp\rho_1 + (\mathbb{1} - P)(1-p)\rho_2 \geq \begin{cases} p\rho_1 \\ (1-p)\rho_2 \end{cases} . \tag{3.4.17}$$

□

We are now interested in sending $m \in \mathbb{N}$ copies of ρ_1 and ρ_2 respectively, i.e. $\rho_1^{\otimes m}$ and $\rho_2^{\otimes m}$. It turns out that for the optimal measurement we find the error rate

$$\mathcal{E}_m^{opt} = \frac{1}{2}(1 - \left\|p\rho_1^{\otimes m} - (1-p)\rho_2^{\otimes m}\right\|_1) \tag{3.4.18}$$

and $\mathcal{E}_m^{opt}$ decays exponentially with $-\xi_m$, with ξ a rate given as

$$\mathcal{E}_m^{opt} \leq Ke^{-\xi m} \tag{3.4.19}$$

Theorem. 3.4.5 *If $p \neq 0, 1$, it holds that*

$$\xi := \lim_{m\to\infty}\left(-\frac{1}{m}\log(\mathcal{E}_m^{opt})\right) = -\log(\inf_{s\in[0,1]} \mathrm{Tr}[\rho_1^{1-s}\rho_2^s]) \tag{3.4.20}$$

Proof. For $A, B \in \mathcal{B}(\mathcal{H})$ positive, $\forall s \in [0,1]$

$$\underbrace{\mathrm{Tr}[(A^s - B^s)A^{1-s}]}_{\mathrm{Tr}[A] - \mathrm{Tr}[B^s A^{1-s}]} \leq \mathrm{Tr}[(A-B)_+] \tag{3.4.21}$$

which is a consequence of $z \mapsto z^s$ being operator monotone. Then

$$\begin{aligned} \frac{1}{2}(\mathrm{Tr}[A+B] - \|A-B\|_1) &= \frac{1}{2}(2\,\mathrm{Tr}[A] - \mathrm{Tr}[A-B] - \mathrm{Tr}[(A-B)_+] + \mathrm{Tr}[(A-B)_-]) \\ &= \mathrm{Tr}[A] - \mathrm{Tr}[(A-B)_+] \leq \mathrm{Tr}[B^s A^{1-s}] \end{aligned} \tag{3.4.22}$$

If we choose $A = p\rho_1^{\otimes m}$ and $B = (1-p)\rho_2^{\otimes m}$, then

$$\begin{aligned} \frac{1}{2}(1 - \left\|p\rho_1^{\otimes m} - (1-p)\rho_2^{\otimes m}\right\|_1) &\leq p^{1-s}(1-p)^s\,\mathrm{Tr}[(\rho_1^{\otimes m\,1-s}\rho_2^{\otimes m\,s} \\ &= p^{1-s}(1-p)^s\,\mathrm{Tr}[(\rho_1^{1-s}\rho_2^s)^{\otimes m}] = p^{1-s}(1-p)^s\,\mathrm{Tr}[\rho_1^{1-s}\rho_2^s]^m \end{aligned} \tag{3.4.23}$$

This gives us that

$$\mathcal{E}_m^{opt} \leq \inf_{s\in[0,1]} p^{1-s}(1-p)^{1-s}\,\mathrm{Tr}[\rho_1^{1-s}\rho_2^s]^m \leq \inf_{s\in[0,1]} \mathrm{Tr}[\rho_1^{1-s}\rho_2^s]^m \tag{3.4.24}$$

and hence for all m

$$-\frac{1}{m}\log\mathcal{E}_m^{opt} \geq -\log\inf_{s\in[0,1]} \mathrm{Tr}[\rho_1^{1-s}\rho_2^s] \tag{3.4.25}$$

which in the limit gives us

$$\lim_{m\to\infty} -\frac{1}{m}\log \mathcal{E}_m^{opt} \geq -\log(\inf_{s\in[0,1]} \mathrm{Tr}[\rho_1^{1-s}\rho_2^s]). \tag{3.4.26}$$

Equality is achieved when ρ_1, ρ_2 are given by $\hat{\rho}_1, \hat{\rho}_2$ such that $[\hat{\rho}_1, \hat{\rho}_2] = 0$. This allows us to write for $x = 1, 2$

$$\rho_x = \sum_i \lambda_i^x |\psi_i^x\rangle\langle\psi_i^x| \tag{3.4.27}$$

and hence

$$\begin{aligned} \hat{\rho}_1 &= \sum_{i,j} \lambda_i^1 |\langle\psi_i^1, \psi_j^2\rangle||ij\rangle\langle ij| \\ \hat{\rho}_2 &= \sum_{i,j} \lambda_i^2 |\langle\psi_i^1, \psi_j^2\rangle||ij\rangle\langle ij| \end{aligned} \tag{3.4.28}$$

with $\{|ij\rangle\}$ a ONB of $\mathcal{H}\otimes\mathcal{H}$. □

3.4.2 The pretty good measurement

Pretty good measurement	**Square measurement**
$R = \sum_{i=1}^n p_x\rho_x$ and then	$S = \sum_{x=1}^n p_x^2\rho_x^2$ and then
$M_x^p = R^{-1/2}p_x\rho_x + \frac{1}{n}\underbrace{(\mathbb{1} - R^{-1/2}RR^{-1/2})}_{\mathbb{1}_{\ker(R)}}$	$M_x^s := S^{-1/2}p_x^2\rho_x^2S^{-1/2} + \frac{1}{n}(\mathbb{1} - S^{-1/2}SS^{-1/2})$
$M^P = (M_x^p)_{x=1}^n$	$M^S = (M_x^s)_{x=1}^n$

with R^{-1} and S^{-1} the Moore-Penrose pseudo inverse. We need the following relations (which we won't proof here) in the following:

Definition. 3.4.6 (Schatten p-norms)
Let $\mathcal{H}$ be a finite dimensional Hilbert space. Then for $p \in [1, \infty)$

$$\|\cdot\|_p : \mathcal{B}(\mathcal{H}) \to [0, \infty), \quad A \mapsto \|A\|_p = \mathrm{Tr}[|A|^p]^{1/p} \tag{3.4.29}$$

is a norm on $\mathcal{B}(\mathcal{H})$.

Theorem. 3.4.7 (Hoelder's inequality) *For $p, q \in [1, \infty]$ and $\frac{1}{p} + \frac{1}{q} = 1$ we find that*

$$\|AB\|_1 = \mathrm{Tr}[|AB|] \leq \|A\|_p \|B\|_q \tag{3.4.30}$$

Theorem. 3.4.8 (Jensen's inequality) *Let f be a continuous function on an interval I. Then the following are equivalent*

1. *f is operator convex in I.*
2. *For each $n \in \mathbb{N}$*
$$f\Big(\sum_{i=1}^n A_i^* X_i A_i\Big) \leq \sum_{i=1}^n A_i^* f(X_i) A_i \tag{3.4.31}$$
with $(X_1, \ldots, X_n)$ a n-tuple of bounded self-adjoint operators with spectra contained in I and $A_1, \ldots, A_n$ operators on $\mathcal{H}$ with $\sum_{i=1}^n A_i^* A_i = \mathbb{1}$.
3. *$f(V^*XV) \leq V^* f(X) V$, with X Hermitian with spectrum in I and V an isometry.*

Now we will come to the results we obtain from this very basic relations.

Proposition. 3.4.9 *We find that in the setting above, we find*

$$(\mathrm{Tr}[S^{1/2}])^2 \leq \mathcal{P}(M^S) \leq \mathcal{P}^{opt} \leq \mathrm{Tr}[S^{1/2}] \tag{3.4.32}$$

Proof. 1.

$$\begin{aligned}(\mathrm{Tr}[S^{1/2}])^2 = (\mathrm{Tr}[SS^{-1/2}])^2 &= \Big(\sum_x \mathrm{Tr}[\underbrace{p_x^2\rho_x^2}_{\sigma_x^2} S^{-1/2}]\Big)^2 \\ &= \Big(\sum_x \mathrm{Tr}[\sigma_x(\sigma_x^{1/2}S^{-1/2}\sigma_x^{1/2})]\Big)^2 \\ &\overset{\text{Jensen}}{\leq} \sum_x \mathrm{Tr}[\sigma_x(\sigma_x^{1/2}S^{-1/2}\sigma_x^{1/2})^2] \\ &= \sum_x \mathrm{Tr}[\sigma_x^2 S^{-1/2}\sigma_x S^{-1/2}] \\ &= \sum_x \mathrm{Tr}[\sigma_x \underbrace{S^{-1/2}\sigma_x^2 S^{-1/2}}_{M_x^s}] = \mathcal{P}(M^S)\end{aligned} \tag{3.4.33}$$

2. Using that $z \mapsto z^{1/2}$ is operator monotone we find that

$$\sigma_x^2 \leq \sum_x \sigma_x^2 = S \tag{3.4.34}$$

giving us that

$$\sigma_x \leq S^{1/2} \quad \forall x = 1, \ldots, n \tag{3.4.35}$$

As a consequence we obtain

$$\sum_x \mathrm{Tr}[M_x\sigma_x] \leq \sum_x \mathrm{Tr}[M_x S^{1/2}] = \mathrm{Tr}[(\underbrace{\sum_x M_x}_{=\mathbb{1}} S^{1/2}] = \mathrm{Tr}[S^{1/2}] \tag{3.4.36}$$

□

Proposition. 3.4.10 *We find that*

$$(\mathcal{P}^{opt})^2 \leq \mathcal{P}(M^p) \leq \mathcal{P}^{opt} \tag{3.4.37}$$

Proof. Let $M = (M_x)_{x=1}^n$ be a POVM. We then findt that

$$\begin{aligned}\Big(\sum_x \mathrm{Tr}[M_x\sigma_x]\Big)^2 &= \Big(\sum_x \mathrm{Tr}[(R^{1/4}M_xR^{1/4})(R^{-1/4}\sigma_xR^{-1/4})]\Big)^2 \\ &\overset{Hoelder}{\leq} \Big(\sum_x \left\|R^{1/4}M_xR^{1/4}\right\|_2 \left\|R^{-1/4}\sigma_xR^{-1/4}\right\|_2\Big)^2 \\ &\leq \underbrace{\sum_x \left\|R^{1/4}M_xR^{1/4}\right\|_2^2}_{(1)} \underbrace{\sum_x \left\|R^{-1/4}\sigma_xR^{-1/4}\right\|_2^2}_{(2)} \leq \mathcal{P}(M^P)\end{aligned} \tag{3.4.38}$$

We find that

$$\begin{aligned}(1) = \sum_x \left\|R^{1/4}M_xR^{1/4}\right\|_2^2 &= \sum_x \mathrm{Tr}[(R^{1/4}M_xR^{1/4})^2] = \sum_x \mathrm{Tr}[R^{1/2}M_xR^{1/2}M_x] \\ &\leq \sum_x \mathrm{Tr}[R^{1/2}M_xR^{1/2}] = \mathrm{Tr}[R] = 1 \\ (2) = \sum_x \left\|R^{-1/4}\sigma_xR^{-1/4}\right\|_2^2 &= \sum_x \mathrm{Tr}[\underbrace{R^{-1/2}\sigma_xR^{-1/2}}_{M_x^p}\sigma_x] = \mathcal{P}(M^p)\end{aligned} \tag{3.4.39}$$

□

In summary with the relation $\mathcal{E}(M) = 1 - \mathcal{P}(M)$, $\mathcal{E}^{opt} = 1 - \mathcal{P}^{opt}$ we find

$$(\mathcal{P}^{opt})^2 \leq \begin{Bmatrix} \mathcal{P}(M^P) \\ \mathcal{P}(M^S) \end{Bmatrix} \leq \mathcal{P}^{opt}, \qquad (\mathcal{E}^{opt}) \leq \begin{Bmatrix} \mathcal{E}(M^P) \\ \mathcal{P}(M^S) \end{Bmatrix} \leq 2\,\mathcal{E}^{opt} \tag{3.4.40}$$

3.5 Separability criteria

Definition. 3.5.1 (Separable states)
Let ρ be a density matrix in $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$, then ρ is separable if it can be written as

$$\rho = \sum_j p_j \rho_j^A \otimes \rho_j^B, \tag{3.5.1}$$

with $0 \leq p_j \leq 1$, $\sum\limits_j p_j = 1$ and ρ_j^A and ρ_j^B states on their respective Hilbert space. If ρ is not separable, then we call ρ entangled.

Remark. A standard way to measure correlations:

$$\mathrm{Cor}_\rho(A:B) := \sup_{\substack{\|O_A\|_1 \leq 1 \\ \|O_B\| \leq 1}} |\operatorname{Tr}[\rho O_A O_B] - \operatorname{Tr}[\rho O_A]\operatorname{Tr}[\rho O_B]| \tag{3.5.2}$$

For a separable state

$$\mathrm{Cor}_{\sum p_j \rho_j^A \otimes \rho_j^B}(A:B) \tag{3.5.3}$$

actually measures classical correlation. In that sense separable states are classically correlated state.

3.5.0.1 Entanglement entropy

Let $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ and $d = \min\{\dim \mathcal{H}_A, \dim \mathcal{H}_B\}$. In Schmidt decomposition we find

$$|\psi\rangle = \sum_{i=1}^{d} \sqrt{\lambda_i}\, |e_i\rangle \otimes |f_i\rangle \tag{3.5.4}$$

with $\lambda_i \geq 0$, $\{|e_i\rangle\}$ an ONB of $\mathcal{H}_A$ and $\{|f_i\rangle\}$ an ONB of $\mathcal{H}_B$ respectively. With $\rho = |\psi\rangle\langle\psi|$ a pure state, we define the entanglement entropy of ρ as the von Neumann entropy of $\{\lambda_i\}_{i=1}^d$, i.e.

$$S_{ENT}(\rho) := -\sum_{i=1} \lambda_i \log(\lambda_i) \tag{3.5.5}$$

We then find that

- Separable: $S_{ENT}(\rho) = 0 \Leftrightarrow$ Schmidt rank of $|\psi\rangle$ is 1.
- Maximally entangled: $\lambda_i = \frac{1}{d}\ \forall i = 1, \ldots, d$.

In the above discussion we have taken ρ to be a pure state, we would, however, like to answer the question if a matrix is entangled or separable in all generality. It turns out that measuring separability in a broader framework is rather difficult. We cannot dive into the discussion directly but first have to develop some tools.

Definition. 3.5.2 (Partial transpose)
The partial transpose is a positive linear map, which is not completely positive. We first introduce the transposition map

$$\Theta : A \mapsto A^t, \qquad \langle i, A^T j\rangle = \langle j, Ai\rangle \qquad \forall i, j. \tag{3.5.6}$$

Using this map we define the partial transpose through its action on the maximally entangled state $|\Omega\rangle = \frac{1}{d}\sum_{i=1}^{d}|ii\rangle$,

$$(\Theta \otimes \mathbb{1})(|\Omega\rangle\langle\Omega|) = \frac{1}{d}\,\mathbb{F} \qquad \mathbb{F} := \sum_{i,j=1}^{n} |ij\rangle\langle ji| \tag{3.5.7}$$

The partial trace can be used to detect entanglement, as follows:

Proposition. 3.5.3 *$\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$. Consider $\rho \in \mathcal{S}(\mathcal{H}_{AB})$. If ρ^{T_A} has a negative eigenvalue, then ρ is entangled.*

Proof. ρ separable $\Rightarrow \rho = \sum_j p_j \rho_j^A \otimes \rho_j^A \overset{T_A}{\Rightarrow} \rho^{T_A} = \sum_j p_j (\rho_j^A)^T \otimes \rho_j^B \geq 0$ □

We again need to introduce some nomenclature to proof the next proposition.

Definition. 3.5.4 (Entanglement witness)
We first set

$$\delta := \{\text{separable density matrices}\} \tag{3.5.8}$$

is convex and compact set. By the Hahn-Banach Theorem $\rho \notin \delta$, then there exists a hyperplane ω such that

$$\mathrm{Tr}[\rho\omega] < 0 \quad \text{and} \quad \mathrm{Tr}[\sigma\omega] \geq 0 \quad \sigma \in \delta. \tag{3.5.9}$$

We then call ω a entanglement witness. The Choi-Jamiolkowski matrix of this state is given through its action on the maximally entangled state

$$\omega = (\Lambda^* \otimes \mathbb{1})(|\Omega\rangle\langle\Omega|) \tag{3.5.10}$$

for Λ a quantum channel.

Proposition. 3.5.5 *Let $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$ with $\rho \in \mathcal{S}(\mathcal{H}_{AB})$ is separable if and only if $(\Lambda \otimes \mathbb{1}_B)(\rho) \geq 0$, for every $\Lambda : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_A)$ positive map.*

Proof. $\Rightarrow$ We just use the explicit form of the entangled state. For a separable state and a positive map $\Lambda : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_A)$, we find

$$(\Lambda \otimes \mathbb{1}_B)(\rho) = (\Lambda \otimes \mathbb{1}_B)\Big(\sum_j \lambda_j \rho_j^A \otimes \rho_j^B\Big) = \sum_j \lambda_j \underbrace{\Lambda(\rho_j^A)}_{\geq 0} \otimes \rho_j^B \geq 0 \tag{3.5.11}$$

$\Leftarrow$ Given ρ entangled, we want to show that there exists A positive map such that $(\Lambda \otimes \mathbb{1}_B)(\rho)$ has a negative eigenvalue. Using Definition 3.5.4, we find

$$\begin{aligned}\mathrm{Tr}[(A \otimes B)\omega] &= \mathrm{Tr}[B^T \Lambda(A)] = \mathrm{Tr}[\mathbb{F}(\Lambda(A) \otimes B^T)] = d\,\mathrm{Tr}[(\Lambda \otimes \mathbb{1}_B)(A \otimes B)|\Omega\rangle\langle\Omega|] \\ &= d\,\langle\Omega, (\Lambda \otimes \mathbb{1}_B)(A \otimes B)\Omega\rangle\end{aligned} \tag{3.5.12}$$

In the second step we used that $\mathrm{Tr}[XY] = \mathrm{Tr}[\mathbb{F}\,X \otimes Y]$. Now this means

$$\mathrm{Tr}[\rho\omega] = d\,\langle\Omega, (\Lambda \otimes \mathbb{1}_B)(\rho)\Omega\rangle \tag{3.5.13}$$

This means, if ρ is entangled, then $\mathrm{Tr}[\rho\omega] < 0$, which gives us $\langle\Omega, (\Lambda \otimes \mathbb{1}_B)(\rho)\Omega\rangle < 0$ and finally $(\Lambda \otimes \mathbb{1}_B)(\rho)$ has a negative eigenvalue.

□

Remark. The idea to implement this map in a lab is

$$\rho \overset{T}{\mapsto} \frac{p}{d^2}\,\mathbb{1}_d \otimes \mathbb{1}_d + (1-p)(\Lambda \otimes \mathbb{1})\rho. \tag{3.5.14}$$

Then T is a completely positive map. If we apply T to a separable state, the minimal eigenvalue of $T(\rho)$ has to be larger than the threshold. If that is not the case we can conclude that ρ is entangled.

3.5.0.2 Partial Transpose

We want to take a closer look at the partial transpose. This map is clearly not unique, as one obtains a different map by just changing the basis in question. Take for example $\tilde{T}_A$ which can be written in terms of a unitary and the "original" partial transpose

$$\rho^{\tilde{T}_A} = (U \otimes \mathbb{1})[(U^* \otimes \mathbb{1})\rho(U \otimes \mathbb{1})]^{\tilde{T}_A} = [(UU^T) \otimes \mathbb{1}]\rho^{T_A}[(UU^T)^* \otimes \mathbb{1}] \neq \rho^{T_A}. \tag{3.5.15}$$

This non-uniqueness does, however, not interfere with the criteria that we developed, as those are only concerned about the eigenvalues and hence are not affected by basis changes (composition with unitaries).

Definition. 3.5.6 (Decoposable map)
We call $\Lambda : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ a decomposable map, if $\Lambda = \Lambda_1 + \Lambda_2 \otimes \Theta$, with Λ_1 and Λ_2 positive maps and Θ a partial transpose.

Remark. The above definition allows us to write the entanglement withness as

$$\omega = Q_1 + Q_2^T \tag{3.5.16}$$

with the PSD

$$Q_i = d(\Lambda_i^* \otimes \mathbb{1})(|\Omega\rangle\langle\Omega|). \tag{3.5.17}$$

In general the separability criteria of the entanglement witness is weaker than of transpositions, i.e.

$$\rho^{T_A} \geq 0 \quad \Rightarrow \quad (\Lambda \otimes \mathbb{1})(\rho) \geq 0 \tag{3.5.18}$$

Example. Let $\Lambda_{red}(A) = \mathrm{Tr}[A]\,\mathbb{1} - A$ we get the separability criteria

$$(\Lambda_{red} \otimes \mathbb{1})(\rho) \geq 0 \Leftrightarrow \begin{cases} \rho_A \otimes \mathbb{1}_B \geq \rho_{AB} \\ \mathbb{1}_A \otimes \rho_B \geq \rho_{AB} \end{cases}. \tag{3.5.19}$$

We get for the witness

$$\omega_{red} = (\mathbb{1} - \mathbb{F})^{T_A} = 2P_-^{T_A} \tag{3.5.20}$$

with P_- the projector onto the anti-symmetric space. Further the ω prop is

$$\mathrm{Tr}[\rho\omega] < 0 \quad \Leftrightarrow \quad \langle\Omega, \rho\Omega\rangle \leq \frac{1}{d} \tag{3.5.21}$$

with $|\Omega\rangle$ the maximally entangled state. In case that $\mathcal{H} = \mathbb{C}^2 \otimes \mathbb{C}^2$, $P_-^{T_A}$ is one dimensional, which gives us that the entanglement witness criterion is equivalent to the PPT criterion.

Proposition. 3.5.7 *Let $\rho \in \mathcal{S}(\mathbb{C}^2 \otimes \mathbb{C}^3)$ or $\mathcal{S}(\mathbb{C}^2 \otimes \mathbb{C}^2)$, then*

$$\rho \text{ separable} \quad \Leftrightarrow \quad \rho^{T_A} \geq 0 \tag{3.5.22}$$

which is a consequence of the complete decomposability of every positive map in $2 \otimes 2$ and $2 \otimes 3$.

Proposition. 3.5.8 *Entangled states with PPT exists if and only if there are non-decomposable maps.*

Chapter 4

Trace Distances, Fidelity and Entropy Measures

4.1 Quantum Entropies

4.1.1 Von Neumann Entropy

In the classical information theory, the setting is the one of an ensemble $X = \{x, P_X\}$. We want to prepare a message of n letters with the n letters drawn independently from X. In this context we define the Shannon Entropy as

$$H(X) := -\sum_x p_x \log p_x. \tag{4.1.1}$$

This quantity gives the value of information, meaning the number of incompressable bits carried per letter (asymptotically with $n \to \infty$).

If we have two ensembles $X = \{x, P_X\}$ and $Y = \{y, Q_Y\}$, we can compare them and compute their correlation

$$I(X : Y) = H(X) + H(Y) - H(X, Y). \tag{4.1.2}$$

This quantity is called the mutual information and has the following interpretations:

- It computes the information per letter about X that can be acquired by reading Y or vice versa.
- Or it can be understood as the amount of information sent through a (classical) channel.

4.1.1.1 Quantum Generalisation

We now translate this quantities into the quantum information context. The setting now is the one of n letters from an ensemble of $\{\rho_x\}$ states with a priori probability $\{p_x\}$, i.e.

$$\rho = \sum_x p_x \rho_x \tag{4.1.3}$$

Definition. 4.1.1 (Von Neumann Entropy)
Let $\rho \in \mathcal{S}(\mathcal{H})$ a positive semidefinite matrix $\mathcal{H}$. We define the von Neumann entropy of ρ as:

$$S(\rho) = -\operatorname{Tr}[\rho \log \rho] = -\operatorname{Tr}[UDU^{-1} \log(UDU^{-1})] = -\operatorname{Tr}[D \log(D)] = -\sum_x \lambda_x \log(\lambda_x) \tag{4.1.4}$$

with

$$\rho = UDU^{-1} \quad \text{with} \quad D = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_n \end{pmatrix} \tag{4.1.5}$$

Remark.

$$\log \rho = U \begin{pmatrix} \log \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \log \lambda_n \end{pmatrix} U^{-1} \tag{4.1.6}$$

Remark. If all the states are mutually orthogonal pure states, then the quantum source reduces to the classical one and they are perfectly distinguishable, i.e. $S(\rho) = H(X)$

The operational interpretation and the meaning of the Von Neumann entropy is versatile.

- The Von Neumann entropy quantifies the quantum information content per letter of ensemble (the minimum number of qubits per letter that are necessary to reliably encode a message).
- It quantifies the entanglement of a bipartite pure state.

Proposition. 4.1.2 (Properties of the Von Neumann entropy) *The following are the essential properties of the Von Neumann entropy which will be the basis for all that follows.*

1. ***Purity*** $\rho = |\psi\rangle\langle\psi| \Rightarrow S(\rho) = 0$
2. ***Unitary invariance*** $S(U\rho U^{-1}) = S(\rho)$
3. ***Maximum*** $S(\rho) \leq \log(D)$ *(the logarithm of the dimension of the underlying Hilbert space).*
4. ***Concavity*** *For* $\lambda_i \geq 0 \sum_i \lambda_i = 1$, $\rho_1, \ldots, \rho_n$ *states, we find*

$$S\Big(\sum_i \lambda_i \rho_i\Big) \geq \sum_i \lambda_i S(\rho_i). \tag{4.1.7}$$

5. ***Entropy of measurement***, $A = \sum_y \lambda_y |a_y\rangle\langle a_y|$. *We measure in the eigenbasis of* A *and define the ensemble*

$$Y = \{a_y, p(a_y)\} \qquad p(a_y) = \langle a_y, \rho a_y \rangle . \tag{4.1.8}$$

It is immediately clear that

$$H(Y) \geq S(\rho) \tag{4.1.9}$$

with equality, if and only if $[A, \rho] = 0$. *More loosely put,* $S(\rho)$ *increases if we replace all off-diagonal terms of* ρ *by 0. The randomness of the measurement outcome is minimized if we choose to measure an observable that commutes with* ρ. *This means if we choose a "bad observable" our measurement becomes less predictable.*

6. ***Entropy of preperation*** *Let* $\{|\varphi_x\rangle, p_x\}$ *be given and* $\rho = \sum_x \lambda_x |\varphi_x\rangle\langle\varphi_x|$. *Then*

$$H(X) \geq S(\rho) \tag{4.1.10}$$

with equality, if and only if $\{|\varphi_x\rangle\}$ *are mutually orthogonal.*

7. ***Additivity*** $\rho_{AB} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$, $\rho_A \in \mathcal{S}(\mathcal{H}_A)$, $\rho_B \in \mathcal{S}(\mathcal{H}_B)$, *then*

$$S(\rho_{AB}) = S(\rho_A) + S(\rho_B) \tag{4.1.11}$$

8. ***Subadditivity*** *It holds in general that*

$$S(\rho_{AB}) \leq S(\rho_A) + S(\rho_B) \tag{4.1.12}$$

This statement is equivalent to the quantum mutual information

$$I_\rho(A:B) = S(\rho_A) + S(\rho_B) - S(\rho_{AB}) \tag{4.1.13}$$

being greater or equal to zero.

9. ***Strong subadditivity*** *We further have that*

$$S(\rho_{ABC}) + S(\rho_B) \leq S(\rho_{AB}) + S(\rho_{BC}), \tag{4.1.14}$$

$\rho_{ABC} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_B)$.

10. ***Triangle inequality*** *(Araki-Lieb inequality)*

$$S(\rho_{AB}) \geq |S(\rho_A) - S(\rho_B)| \tag{4.1.15}$$

Exercise. Let ρ_{AB} a pure quantum state in a bipartite Hilbert space, i.e. $S(\rho_{AB}) = 0$ and $S(\rho_A) = S(\rho_B)$ but in general $S(\rho_A) = S(\rho_B) \neq 0$. Show this.

4.1.1.2 Open system evolution

Let $\rho_{SE} = \rho_S \otimes \rho_E$. We then have that $S(\rho_{SE}) = S(\rho_S) + S(\rho_E)$. If we now look at the evolution map

$$\rho_{SE} \mapsto U_{SE}\rho_{SE}U_{SE}^{-1} = \rho'_{SE} \tag{4.1.16}$$

From Proposition 4.1.2, we find

$$S(\rho_S) + S(\rho_E) = S(\rho_{SE}) = S(\rho'_{SE}) \leq S(\rho'_S) + S(\rho'_E). \tag{4.1.17}$$

Put in different terms Equation (4.1.17) is the second law of thermodynamics.

Definition. 4.1.3 (Conditional entropy)
Let $\rho_{AB} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$, then we define the conditional quantum entropy as

$$H(A|B)_\rho := S(\rho_{AB}) - S(\rho_B) \tag{4.1.18}$$

Remark. We have the following properties

- $H(A|B)_\rho \geq -\log(d_A)$ but can be negative, which we will later show.
- $S(\rho_A) \geq H(A|B)_\rho$

Definition. 4.1.4 (Coherent information)
The coherent information for $\rho_{AB} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ is given by

$$I(A\langle B)_\rho := S(\rho_B) - S(\rho_{AB}) \tag{4.1.19}$$

Definition. 4.1.5 (Mutual information)
Let $\rho_{AB} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$, then the mutual information is given by

$$I(A:B)_\rho = S(\rho_A) + S(\rho_B) - S(\rho_{AB}) = S(\rho_A \otimes \rho_B) - S(\rho_{AB}) \tag{4.1.20}$$

Remark. We have the following properties

- $I(A:B)_\rho \geq 0$
- Chain rule: $I(A:BC)_\rho = I(A:B)_\rho + I(A:C|B)_\rho$. With the last quantity the conditional mutual information defined in the following.

Definition. 4.1.6 (Conditional mutual information)
The conditional mutual information for $\rho_{ABC} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$ is given by

$$I(A:C|B)_\rho = S(\rho_{AB}) + S(\rho_{BC}) - S(\rho_B) - S(\rho_{ABC}) \geq 0 \tag{4.1.21}$$

Definition. 4.1.7 (Quantum relative entropy)
Let $\rho, \sigma \in \mathcal{S}(\mathcal{H})$, then the quantum relative entropy is given as

$$D(\rho\|\sigma) = \begin{cases} \mathrm{Tr}[\rho(\log\rho - \log\sigma)] & \ker\sigma \subseteq \ker\rho \\ +\infty & \text{otherwise} \end{cases}. \tag{4.1.22}$$

Remark. The quantum relative entropy is inspired by the Kullback-Leibler divergence. For $p = \{p_x\}$, $q = \{q_x\}$ probability distributions

$$KL(p\|q) = \sum_x p_x \log\frac{p_x}{q_x} \tag{4.1.23}$$

Definition. 4.1.8 (Belavkin-Staszewski relative entropy)
Let $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ full rank, then the Belavkin-Staszewski relative entropy is given by

$$\widehat{D}(\rho\|\sigma) = \mathrm{Tr}[\rho\log\rho^{1/2}\sigma\rho^{1/2}] \tag{4.1.24}$$

Remark. We have the relation

$$D(\rho\|\sigma) \leq \widehat{D}(\rho\|\sigma) \tag{4.1.25}$$

with equality if and only if $[\rho, \sigma] = 0$.

Proposition. 4.1.9 (Properties of the quantum relative entropy I) *The quantum relative entropy has the following properties*

- *Continuity:* $\rho \mapsto D(\rho\|\sigma)$
- *Additive:* $\rho_A, \sigma_A \in \mathcal{S}(\mathcal{H}_A)$, $\rho_B, \sigma_B \in \mathcal{S}(\mathcal{H}_B)$, *then*

$$D(\rho_A \otimes \rho_B\|\sigma_A \otimes \sigma_B) = D(\rho_A\|\sigma_A) + D(\rho_B\|\sigma_B) \tag{4.1.26}$$

- *Superadditivity:* $\rho_{AB} \in \mathcal{S}(\mathcal{H}_{AB})$, $\sigma_A \in \mathcal{S}(\mathcal{H}_A)$, $\sigma_B \in \mathcal{S}(\mathcal{H}_B)$

$$D(\rho_{AB}\|\sigma_A \otimes \sigma_B) \geq D(\rho_A\|\sigma_A) + D(\rho_B\|\sigma_B). \tag{4.1.27}$$

- *Data processing inequality:* $\rho, \sigma \in \mathcal{S}(\mathcal{H})$, T *a quantum channel, then*

$$D(\rho\|\sigma) \geq D(T(\rho)\|T(\sigma)) \tag{4.1.28}$$

Remark. There is an axiomatic characterization of the relative entropy. We find that every function

$$f : \mathcal{S}(\mathcal{H}) \times \mathcal{S}(\mathcal{H}) \to [0, +\infty) \tag{4.1.29}$$

that satisfies the properties in Proposition 4.1.9 is already the relative entropy.

Proposition. 4.1.10 (Properties of the quantum relative entropy II) *The relative entropy further has the properties:*

- *Non-negativity:* $D(\rho\|\sigma) \geq 0$
- *Unitary invariance:* $D(U\rho U^*\|U\sigma U^*) = D(\rho\|\sigma)$

4.1.2 Relative entropy

We once again revisit the relative entropy and give again the definition

Definition. 4.1.11 (Relative entropy)
Let $\rho, \sigma \in \mathcal{S}(\mathcal{H})$, with $\mathcal{H}$ finite-dimensional Hilber space, we define their relative entropy by:

$$D(\rho\|\sigma) := \begin{cases} \mathrm{Tr}[\rho(\log\rho - \log\sigma)] & \text{if } \ker\sigma \subseteq \ker\rho \\ +\infty & \text{otherwise} \end{cases} \tag{4.1.30}$$

Proposition. 4.1.12 (Properties of the relative entropy)

- ***Unitary invariance:*** $D(U^*\rho U\|U^*\sigma U) = D(\rho\|\sigma)$ $\forall U$ *unitaries*
- ***Non-negativity:*** $D(\rho\|\sigma) \geq 0$ *with equality if and only if* $\rho = \sigma$ *(direct consequence of DPI for the* $\mathrm{Tr}[\cdot]$*)*

1. ***Continuity:*** $\rho \to D(\rho\|\sigma)$ *is continuous.*
2. ***Additivity:*** $\rho_{AB}, \sigma_{AB} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$, $D(\rho_A \otimes \rho_B\|\sigma_A \otimes \sigma_B) = D(\rho_A\|\sigma_A) + D(\rho_B\|\sigma_B)$.
3. ***Superadditivity:*** $\rho_{AB} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$, $\sigma_A \in \mathcal{S}(\mathcal{H}_A)$, $\sigma_B \in \mathcal{S}(\mathcal{H}_B)$, *then*

$$D(\rho_{AB}\|\sigma_A \otimes \sigma_B) \geq D(\rho_A\|\sigma_B) + D(\rho_B\|\sigma_B) \tag{4.1.31}$$

4. ***Data processing inequality*** *(monotonicity)* $\rho, \sigma \in \mathcal{S}(\mathcal{H})$, T *CPTP map, then*

$$D(\rho\|\sigma) \geq D(T(\rho)\|T(\sigma)) \tag{4.1.32}$$

Theorem. 4.1.13 (Axiomatic characterisation of the relative entropy) *If* $f : \mathcal{S}(\mathcal{H}) \times \mathcal{S}(\mathcal{H}) \to [0, \infty)$ *satisfies 1. - 4. from Proposition 4.1.12, then* f *is the relative entropy.*

Proof. Step 1. 1. - 3. imply "lower asymptotic semicontinuity" (LAS) Definition 4.1.14. Let therefore be (ρ, σ) and $\{\rho'_n\}$ sequence of state be given, such that

$$\left\|\rho^{\otimes n} - \rho'_n\right\|_1 \overset{n\to\infty}{\longrightarrow} \tag{4.1.33}$$

Through the DPI for $\|\cdot\|_1$ we can conclude that

$$\|\rho - (\rho'_n)_i\|_1 \overset{n\to\infty}{\longrightarrow} . \tag{4.1.34}$$

Now applying superadditivity to the first and addtitivity to the second summand gives

$$\begin{aligned} \frac{1}{n}(D(\rho'_n\|\sigma^{\otimes n}) - D(\rho^{\otimes}\|\sigma^{\otimes n})) &\overset{2.+3.}{\geq} \frac{1}{n}\sum_{i=1}^{n}\left[D((\rho'_n)_i\|\sigma) - D(\rho\|\sigma)\right] \\ &\geq \min_{i=1,\dots,n}\left[D((\rho'_n)_i\|\sigma) - D(\rho\|\sigma)\right] \overset{n\to\infty}{\underset{1.}{\longrightarrow}} 0 \end{aligned} \tag{4.1.35}$$

Step 2. 2. + 4. + LAS gives us the relative entropy. To see this let w.l.o.g. $\rho_0, \sigma_0 \in \mathcal{S}(\mathcal{H})$ such that $f(\rho_0, \sigma_0) = D(\rho_0\|\sigma_0)$. For any $\rho, \sigma \in \mathcal{S}(\mathcal{H})$, there exists $l, m, l', m' \in \mathbb{N}$ such that

$$\frac{l'}{m'}D(\rho_0\|\sigma_0) \leq D(\rho\|\sigma) \leq \frac{l}{m}D(\rho_0\|\sigma_0). \tag{4.1.36}$$

Now the upper bound is equivalent to saying that

$$\begin{aligned} & mD(\rho\|\sigma) \leq lD(\rho_0\|\sigma_0) \\ \overset{2.}{\iff} \quad & D(\rho^{\otimes m}\|\sigma^{\otimes m}) \leq D(\rho_0^{\otimes l}\|\sigma_0^{\otimes l}) \\ \overset{\text{Lemma 4.1.15}}{\implies} \quad & \Psi^n(\sigma_0^{\otimes ln}) = \sigma^{\otimes nm}, \quad \lim_{n\to\infty}\left\|\Psi^n(\rho_0^{\otimes ln}) - \rho^{\otimes mn}\right\|_1 = 0 \end{aligned} \tag{4.1.37}$$

Therefore,

$$\begin{aligned} mf(\rho,\sigma) &\overset{2.}{=} f(\rho^{\otimes n},\sigma^{\otimes n}) \overset{2.}{=} \limsup_{n\to\infty}\frac{1}{n}f(\rho^{\otimes mn},\sigma^{\otimes mn}) \\ &\leq \liminf_{n\to\infty}\frac{1}{n}f(\Psi^n(\rho_0^{\otimes ln}),\underbrace{\Psi^n(\sigma_0^{\otimes ln})}_{\sigma^{\otimes nm}}) \\ &\overset{4.+\text{DPI}}{\leq} \liminf_{n\to\infty}\frac{1}{n}f(\rho_0^{\otimes ln},\sigma_0^{\otimes ln}) \\ &\overset{2.}{=} f(\rho_0^{\otimes l},\sigma_0^{\otimes l}) \overset{2.}{=} lf(\rho_0,\sigma_0) \end{aligned} \tag{4.1.38}$$

which is equivalent to

$$f(\rho,\sigma) \leq \frac{l}{m}f(\rho_0,\sigma_0) = \frac{l}{m}D(\rho_0,\sigma_0). \tag{4.1.39}$$

Analogously one obtains

$$\frac{l'}{m'}D(\rho_0\|\sigma_0) = \frac{l'}{m'}f(\rho_0,\sigma_0) \leq f(\rho,\sigma) \leq \frac{l}{m}D(\rho\|\sigma_0). \tag{4.1.40}$$

Choosing properly l, m, l', m', we can conclude that $f(\rho,\sigma) \propto D(\rho\|\sigma)$.

□

Definition. 4.1.14 (Lower asymptotic semicontinuity (LAS))
For $\rho,\sigma \in \mathcal{S}(\mathcal{H})$ a pair of states, $\mathcal{H}^{\otimes n}$, $\{\rho'_n\}$ a sequence in $\mathcal{S}(\mathcal{H}^{\otimes n})$. We say that f is LAS with respect to σ if

$$\lim_{n\to\infty}\left\|\rho^{\otimes n}-\rho'_n\right\|_1 = 0 \tag{4.1.41}$$

then, this implies that

$$\liminf_{n\to\infty}\frac{1}{n}(f(\rho'_n,\sigma^{\otimes n}) - f(\rho^{\otimes n},\sigma^{\otimes n})) \geq 0 \tag{4.1.42}$$

Lemma. 4.1.15 *If $\rho,\sigma,\rho_0,\sigma_0 \in \mathcal{S}(\mathcal{H})$ such that $D(\rho\|\sigma) \leq D(\rho_0\|\sigma_0)$, then there exists as sequence (Ψ_n) of CPTP maps such that*

$$\Psi_n(\sigma_0^n) = \sigma^{\otimes n}, \quad \lim_{n\to\infty}\left\|\Psi_n(\rho_0^{\otimes n}) - \rho^{\otimes n}\right\|_1 = 0 \tag{4.1.43}$$

4.1.3 Non-commmutative L_p norms

4.1.3.1 Schatten p-norms

Definition. 4.1.16 (Schatten p-norms)
Let $X \in \mathcal{B}(\mathcal{H})$ and $p \in [1,+\infty)$, then the Schatten p-Norm of X is given by

$$\|X\|_p := (\mathrm{Tr}[|X|^p])^{1/p} \tag{4.1.44}$$

where $|X| = \sqrt{X^*X}$. For $p=\infty$, we defin

$$\|\cdot\|_\infty = \lim_{p\to\infty}\|\cdot\|_p \tag{4.1.45}$$

the operator norm. For $p<1$, it does not satisfy the triangle inequality.

Proposition. 4.1.17 (Properties of Schatten p-norms) *1. **Monotonicity:** For $1 \leq p \leq p' \leq +\infty$,*

$$\|X\|_1 \leq \|X\|_p \leq \|X\|_{p'} \leq \|X\|_\infty . \tag{4.1.46}$$

*2. **Unitary invariance** For U a unitary, we have*

$$\|UXU^*\|_p = \|X\|_p . \tag{4.1.47}$$

3. ***Minkowski's inequality:*** $\|X+Y\|_p \leq \|X\|_p + \|Y\|_p$.

4. ***Hölder's inequality:*** $\frac{1}{p}+\frac{1}{q}=1$, $\|XY\|_1 \leq \|X\|_p \|Y\|_q$.

5. ***Duality:***
$$\|X\|_q := \sup\left\{|\langle X,Y\rangle| \ : \ \|Y\|_p = 1\right\} \tag{4.1.48}$$

6. ***Submultiplicativity:*** $\|XY\|_p \leq \|X\|_p \|Y\|_p$

7. ***Generalised Hölder's inequality:*** *For* $\frac{1}{r}=\frac{1}{p}+\frac{1}{q}$, $0<r<\infty$, $\|XY\|_r \leq \|X\|_p \|Y\|_q$

8. $\|X\| = \|X^*\|_p$, $\|X\|_{2p}^p = \|X^*X\|_p$.

9. $\|\cdot\|_1$ *satisfies DPI:* $\|\rho-\sigma\|_1 \geq \|T(\rho)-T(\sigma)\|_1$ *for* T *CPTP.*

Definition. 4.1.18 (Weighted p-norms)
Let $p\in[1,\infty)$ again, $\rho\in\mathcal{S}_+(\mathcal{H})$ full-rank, then the weighted p-norm is given by:

$$\|X\|_{L_p(\rho)} := \mathrm{Tr}[\left|\rho^{\frac{1}{2p}} X \rho^{\frac{1}{2p}}\right|^p]^{1/p} \tag{4.1.49}$$

for all $X\in\mathcal{B}(\mathcal{H})$. We further define the KMS (Kubo-Martin-Schwinger) inner product

$$\langle X,Y\rangle_{\rho,KMS} := \mathrm{Tr}[\rho^{1/2} X j \rho^{1/2} Y] \tag{4.1.50}$$

and the GNS (Gelfant-Naimark-Segal) inner product

$$\langle X,Y\rangle_{\rho,GNS} := \mathrm{Tr}[\rho X^* Y] \tag{4.1.51}$$

for all $X,Y\in\mathcal{B}(\mathcal{H})$

Proposition. 4.1.19 (Properties of the weighted p-norms)

1. ***Monotonicity:*** $\forall p,q,\in[1,\infty)$, $p\leq q$, $\|X\|_{L_p(\rho)} \leq \|X\|_{L_q(\rho)}$ $\forall X\in\mathcal{B}(\mathcal{H})$.

2. ***Duality:*** $\|X\|_{L_p(\rho)} := \sup\left\{|\langle X,Y\rangle_{\rho,GMS}| \ : \ \|Y\|_{L_q(\rho)}=1\right\}$

3. ***Operator norm:*** $\|X\|_\infty = \|X\|_{L_\infty(\rho)} := \lim_{p\to\infty}\|X\|_{L_p(\rho)}$

4.2 Divergences

In 1961 Alfred Renyi, supplemented the Kullback-Leibler divergence, given for probability distributions $\{p_x\}_{x=1}^n$, $\{q_x\}_{x=1}^n$ as

$$KL(p\|q) = \sum_{x=1}^n p_x \log\frac{p_x}{q_x}, \tag{4.2.1}$$

by starting with an axiomatic description and then arriving at all possible families of divergences that satisfy those axioms. We will build from the classical axioms to the quantum ones

Definition. 4.2.1 (Classical axiomatic definition of a divergence)
We call a function $\mathbb{D}:\mathcal{B}(\mathcal{H})_+\times\mathcal{B}(\mathcal{H})_+\to[0,+\infty)$ a divergence if for $X,Y\in\mathcal{B}(\mathcal{H})_+$ i.e. unnormalised Hermitian positive semi-definite operators that satisfy the kernel inclusion $\ker Y\subseteq\ker X$, the following hold

1. **Continuity:** $X\mapsto\mathbb{D}(X\|Y)$ is continuous (problems with continuity on Y), $Y\mapsto\mathbb{D}(X\|Y)$ is continuous if $X,Y>0$.

2. **Unitary invariance:** $\mathbb{D}(X\|Y) = \mathbb{D}(UXU^*\|UYU^*)$ for all unitaries U.

3. **Order:** If $X \geq Y$, then $\mathbb{D}(X\|Y) \geq 0$. If $X \leq Y$, then $\mathbb{D}(X\|Y) \leq 0$.

4. **Additivity:** $\mathbb{D}(X_1 \otimes X_2\|Y_1 \otimes Y_2) = \mathbb{D}(X_1\|Y_1) + \mathbb{D}(X_2\|y_2)$.

5. **General mean:** There exists a continuous, strictly monotonic function g, s.t.

$$\mathbb{Q}(\cdot\|\cdot) = g(\mathbb{D}(\cdot\|\cdot)) \tag{4.2.2}$$

and for $X_1, Y_1 \in \mathcal{B}(\mathcal{H}_1)_+$, $X_2, Y_2 \in \mathcal{B}(\mathcal{H}_2)_+$

$$\mathbb{Q}(X_1 \oplus X_2\|Y_1 \oplus Y_2) = \frac{\mathrm{Tr}[X_1]}{\mathrm{Tr}[X_1] + \mathrm{Tr}[X_2]}\,\mathbb{Q}(X_1\|Y_1) + \frac{\mathrm{Tr}[X_2]}{\mathrm{Tr}[X_1] + \mathrm{Tr}[X_2]}\,\mathbb{Q}(X_2\|Y_2) \tag{4.2.3}$$

Proposition. 4.2.2 (Classical case) *A divergence satisfying Definition 4.2.1 is either the Kullback-Leibler divergence of the Renyi divergence:*

$$D_\alpha(p\|q) = \frac{1}{\alpha - 1} \log \frac{\sum_{x=1}^n p_x^\alpha q_x^{1-\alpha}}{\sum_{x=1}^n p_x} \tag{4.2.4}$$

for $\alpha \in (0,1) \cup (1, +\infty)$. In the case of the KL-divergence $g = 1$ and in the α-divergence case $g_\alpha(t) = \exp((\alpha - 1)t)$. In the limit $\alpha \nearrow 1$, $\alpha \searrow 1$ one gets $D_\alpha \to KL$

Definition. 4.2.3 (Quantum axiomatic definition of a divergence)
We are in the setting of Definition 4.2.1 and add some additional axioms to restrict the number of families of divergences and also make them able to work with.

1. **Positive definiteness:** If $\rho, \sigma \in \mathcal{S}(\mathcal{H})$, $\mathbb{D}(\rho\|\sigma) \geq 0$ with equality if and only if $\rho = \sigma$.

2. **Data processing inequality:** For T a CPTP map $\mathbb{D}(\rho\|\sigma) \geq \mathbb{D}(T(\rho)\|T(\sigma))$.

3. (a) **Joint convexity:** $(\alpha > 1)$ $\{\rho_i\}_i$, $\{\sigma_i\}_i$, $0 \leq \lambda_i \leq 1$, then

$$\mathbb{Q}\Big(\sum_i \lambda_i \rho_i \| \sum_i \lambda_i \sigma_i\Big) \leq \sum_i \lambda_i\, \mathbb{D}(\rho_i\|\sigma_i) \tag{4.2.5}$$

   (b) **Joint concavity:** $(\alpha < 1)$ $\{\rho_i\}_i$, $\{\sigma_i\}_i$, $0 \leq \lambda_i \leq 1$, then

$$\mathbb{Q}\Big(\sum_i \lambda_i \rho_i \| \sum_i \lambda_i \sigma_i\Big) \geq \sum_i \lambda_i\, \mathbb{D}(\rho_i\|\sigma_i) \tag{4.2.6}$$

4. **Dominance:** $X, Y, Y' \in \mathcal{B}(\mathcal{H})_+$, $Y \leq Y'$, then $\mathbb{D}(X\|Y) \geq \mathbb{D}(X\|Y')$

4.2.1 Minimal Divergence

Definition. 4.2.4 (Pinching map)
We call the CPTP map $\mathcal{P} : L \mapsto \sum_{x=1}^n P_x L P_x$ with $\{P_x\}_{x=1}^n$ orthogonal projections, i.e. $P_x = P_x^*$, $\sum_{x=1}^n P_x = 1$, which can be represented by

$$\mathcal{P}(L) = \sum_{x=1}^n P_x L P_x = \sum_{y=1}^n U_y L U_y^* \tag{4.2.7}$$

with $U_y = \sum_{x=1}^n e^{\frac{2\pi i y x}{n}} P_x$. From that representation it is also quite obvious that $\mathcal{P}$ is indeed CPTP.

Remark. For H Hermitian, $H = \sum_{x=1}^n \lambda_x |e_x\rangle\langle e_x|$, we can set $P_\lambda = \sum_{x:\lambda_x = \lambda} |e_x\rangle\langle e_x|$ which means $H = \sum_x \lambda_x P_x$. We then can create the pinching map using H that we call $\mathcal{P}_H : L \mapsto \sum_x P_x L P_x$ and get the following properties

- $\mathcal{P}_H(L) \geq \frac{1}{|\mathrm{spec}H|} L$
- $[\mathcal{P}_H(L), H] = 0$

Definition. 4.2.5 (Preperation map)
We define for $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ the preperation map Λ which is a CPTP map. For that purpose we set $\Lambda = \sigma^{-1/2}\rho\sigma^{-1/2}$ and in spectral decomposition $\Delta = \sum_x \lambda_x \Pi_x$. Using this we define

$$q(x) = \mathrm{Tr}[\sigma\Pi_x], \qquad p(x) = \lambda_x q(x) \tag{4.2.8}$$

and with that

$$\Lambda(\cdot) = \sum_x \langle x, \cdot x\rangle \frac{1}{q(x)} \sigma^{1/2}\Pi_x\sigma^{1/2}. \tag{4.2.9}$$

We find that $\Lambda(p) = \rho$ and $\Lambda(q) = \sigma$.

Using the above we can define the minimal Renyi divergences

Definition. 4.2.6 (Minimal Renyi divergence (Sandwiched Renyi Divergences))
For $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ $\alpha \in (1/2, 1) \cup (1, \infty)$,

$$\begin{aligned}\mathcal{D}_\alpha(\rho\|\sigma) \lim_{n\to\infty} \frac{1}{n} \mathbb{D}_\alpha(\rho^{\otimes n}\|\sigma^{\otimes n}) &\overset{\mathrm{DPI}}{\geq} \lim_{n\to\infty} \frac{1}{n} \mathbb{D}_\alpha(\mathcal{P}_{\sigma^{\otimes n}}(\rho^{\otimes n})\|\sigma^{\otimes n}) \\ &= \frac{1}{\alpha - 1} \log \mathrm{Tr}[(\sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}})^\alpha] =: \widetilde{D}_\alpha(\rho\|\sigma)\end{aligned} \tag{4.2.10}$$

and also the maximal ones

Definition. 4.2.7 (Maximal Renyi divergences (Maximal Renyi Divergences))
For $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ $\alpha \in (1, 2)$, we find that

$$\mathcal{D}_\alpha(\rho\|\sigma) = \mathcal{D}_\alpha(\Lambda(p)\|\Lambda(q)) \overset{\mathrm{DPI}}{\leq} \mathcal{D}_\alpha(p\|q) = \frac{1}{\alpha - 1} \log \mathrm{Tr}[\sigma(\sigma^{-1/2}\rho\sigma^{-1/2})^\alpha] =: \widehat{D}_\alpha(\rho\|\sigma). \tag{4.2.11}$$

The quantity in the argument of the trace is called a geometric mean, which we also write as

$$\sigma \#_\alpha \rho = \sigma^{1/2}(\sigma^{-1/2}\rho\sigma^{-1/2})^\alpha \sigma^{1/2} \tag{4.2.12}$$

Remark.
- Clearly $\widehat{D}_\alpha(\rho\|\sigma) \geq \widetilde{D}_\alpha(\rho\|\sigma)$ with equality if and only if $[\rho, \sigma] = 0$.
- We have further that

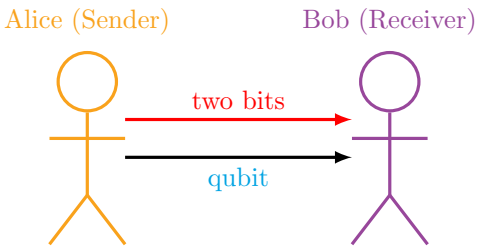
$$\lim_{\alpha\to 1} \widehat{D}_\alpha(\rho\|\sigma) = \widehat{D}(\rho\|\sigma) := \mathrm{Tr}[\rho \log(\rho^{1/2}\sigma^{-1}\rho^{1/2}] \tag{4.2.13}$$

$$\lim_{\alpha\to 1} \widetilde{D}_\alpha(\rho\|\sigma) = D(\rho\|\sigma) = \mathrm{Tr}[\rho(\log(\rho) - \log(\sigma))] \tag{4.2.14}$$

Definition. 4.2.8 (Petz Renyi Divergence)
For $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ $\alpha \in (0, 1)$ we define the Petz Renyi Divergence as

$$\overline{D}_\alpha(\rho\|\sigma) := \frac{1}{\alpha - 1} \log \mathrm{Tr}[\rho^\alpha \sigma^{1-\alpha}] \tag{4.2.15}$$

Definition. 4.2.9



4.3 Quantum Hypothesis Testing

4.3.1 Symmetric State Discrimination

$$\mathcal{P}(\mathcal{M}) = \sum \tag{4.3.1}$$

Definition. 4.3.1
States $\rho, \sigma \in \mathcal{S}(\mathcal{H})$. A system

- Null hypothesis: The state of A^n is $\rho^{\otimes n}$.
- Alternate hypothesis: The state of A^n is $\sigma^{\otimes n}$.

We get a POVM $\{\mathbb{P}, 1-\mathbb{P}\}$ with $\mathbb{P}$ an orthogonal projection. We call T_n a "hypothesis test". We can make two kind of errors

1. **First kind error:** We wrongly conclude that the alternate hypothesis is correct even if the state is $\rho^{\otimes n}$
$$\alpha_n(T_n;\rho) := \mathrm{Tr}[\rho^{\otimes n}(\mathbb{1} - T_n)]. \tag{4.3.2}$$
2. **Second kind error:** We wrongly conclude that the null hypothesis is correct even if the state is $\sigma^{\otimes n}$
$$\beta_n(T_n;\sigma) := \mathrm{Tr}[\sigma^{\otimes n} T_n] \tag{4.3.3}$$

We have the Chernoff bound as

$$\min_{\substack{T_n \\ \text{hypothesis} \\ \text{test}}} \frac{1}{2}(\alpha_n(T_n;\rho) + \beta_n(T_n;\sigma)) = \frac{1}{2}(1 - \left\| p\rho^{\otimes n} - (1-p)\sigma^{\otimes n} \right\|_1) \tag{4.3.4}$$

and the quantum Chernoff bound

$$\begin{aligned} \lim_{n\to\infty} -\frac{1}{n} \log \min_{\substack{T_n \\ \text{hypothesis} \\ \text{test}}} \frac{1}{2}(\alpha_n(T_n;\rho) + \beta_n(T_n;\sigma)) &= \max_{0\le s\le 1} -\mathrm{Tr}[\rho^s \sigma^{1-s}]. \\ &= -\min_{0\le s\le 1} \log \overline{Q}_s(\rho\|\sigma) \\ &= \max_{0\le s\le 1} (1-s)\overline{D}_s(\rho\|\sigma) \end{aligned} \tag{4.3.5}$$

We find a building block of the Petz Renyi divergence.

$$\overline{D}_\alpha(\rho\|\sigma) := \frac{1}{\alpha - 1} \log \mathrm{Tr}[\rho^\alpha \sigma^{1-\alpha}] \qquad \alpha \in (0,1) \cup (1, +\infty). \tag{4.3.6}$$

So the interpretation of the Petz Renyi divergence is that it provides optimal exponential rate for the error committed in the task of binary hypothesis testing when considering errors of kinds first and second jointly.

4.3.2 Asymmetric hypothesis testing

The goal of asymmetric quantum hypothesis testing is to minimize

$$\beta_n(T_n;\sigma) := \mathrm{Tr}[\sigma^{\otimes n} T_n] \tag{4.3.7}$$

under the constraint

$$\alpha_n(T_n;\rho) = \mathrm{Tr}[\rho^{\otimes n}(\mathbb{1} - T_n)] \leq \varepsilon \tag{4.3.8}$$

Lemma. 4.3.2 *Let $T = \mathcal{P}_{\sigma^{\otimes n}} : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ the pinching map, $\rho, \sigma \in \mathcal{S}(\mathcal{H})$*

$$\mathcal{P}_{\sigma^{\otimes n}}(X) := \sum_{i=1}^{\alpha} P_i X P_i, \tag{4.3.9}$$

from the spectral decomposition of $\sigma^{\otimes n} = \sum_{i=1}^{k} \lambda_i P_i$ (k runs over the distinct eigenvalues of $\sigma^{\otimes n}$.) We have

$$D(\rho\|\sigma) = \lim_{n\to\infty} \frac{1}{n} D(\mathcal{P}_{\sigma^{\otimes n}}(\rho^{\otimes n}), \sigma^{\otimes n}) \tag{4.3.10}$$

4.3.3 Quantum Stein Lemma

The task is to distinguish two quantum states $\rho, \sigma \in \mathcal{B}(\mathcal{H})$. For every $\varepsilon \in (0,1)$, we find

$$\lim_{n\to\infty} -\frac{1}{n} \log \beta_n = D(\rho\|\sigma) \tag{4.3.11}$$

Proof. We want to proof that $D(\rho\|\sigma)$ is a lower bound on $-\frac{1}{n}\log\beta_n$. Using Lemma 4.3.3 with $A = \rho^{\otimes n}$ and $B = e^{\lambda n}\sigma^{\otimes n}$ ($\lambda \in \mathbb{R}$ will be chosen later). I.e. for $s \in [0,1]$,

$$e^{-s\lambda n}\,\mathrm{Tr}[\rho^{1+s}\sigma^{-1}]^n \geq \mathrm{Tr}[(\rho^{\otimes n} - e^{\lambda n}\sigma^{\otimes n})T_n] \geq (1-\varepsilon) - e^{\lambda n}\beta_n(T_n;\sigma_n) \tag{4.3.12}$$

where we used in the last step that $\alpha_n \leq \varepsilon$. This gives us

$$\beta_n(T_n;\sigma) \geq e^{-n\lambda}[(1-\varepsilon) - e^{-n(\lambda n - f(s))}] \tag{4.3.13}$$

with $f(s) := \log \mathrm{Tr}[\rho^{1+s}\sigma^{-s}]$ having the properties

- $f(0) = 0$
- $f'(0) = D(\rho\|\sigma)$.

If we choose $\lambda = D(\rho\|\sigma) + \delta$ for $\delta > 0$. Hence there exists a $s \in (0,1]$ such that

$$\lambda s > f(s) \tag{4.3.14}$$

which allows us to take the limit

$$\lim_{n\to\infty} -\frac{1}{n}\log\beta_n(T_n;\sigma) \leq D(\rho\|\sigma) + \delta \tag{4.3.15}$$

Since δ was arbitrary □

Lemma. 4.3.3 *We have for self-adjoint A, B and all $s \in [0,1]$ that*

$$\|A - B\|_1 \geq \mathrm{Tr}[A+B] - 2\,\mathrm{Tr}[A^S B^{1-s}] \tag{4.3.16}$$

and further

$$\mathrm{Tr}[(A+B)_+] \leq \mathrm{Tr}[A^{1-s}B^s] \tag{4.3.17}$$

4.4 Quantum source coding

Let $\{|\varphi(x)\rangle, p(x)\}$ be an ensemble of states and probabilities. We create

$$\rho = \sum_x p(x)|\varphi(x)\rangle\langle\varphi(x)| \tag{4.4.1}$$

Alice sends n-letters to Bob, meaning she sends $\rho^{\otimes n}$. Hence there is a redundancy in the message and the question now is: How much can we compress the message (sent by Alice) so that it is perfectly understandable? We will see that the rate of compression is the von Neumann entropy as we would expect abstracting from the classical case. The proof is quite extensive and not very instructive so we will skip it here and just give an example.

Example. Let the ensemble be given by

$$\begin{aligned} |1\rangle &= \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad p = \frac{1}{2} \\ |+\rangle &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad p = \frac{1}{2} \end{aligned} \tag{4.4.2}$$

then clearly

$$\rho = \frac{1}{2}|1\rangle\langle 1| + \frac{1}{2}|+\rangle\langle +| = \begin{pmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{pmatrix} \tag{4.4.3}$$

having the eigenvalues and eigenstates

$$|0'\rangle = \begin{pmatrix} \cos\frac{\pi}{8} \\ \sin\frac{\pi}{8} \end{pmatrix}, \quad \lambda_{0'} = \cos^2\frac{\pi}{8}, \qquad |1'\rangle = \begin{pmatrix} \sin\frac{\pi}{8} \\ -\cos\frac{\pi}{8} \end{pmatrix}, \quad \lambda_{1'} = \sin^2\frac{\pi}{8}. \tag{4.4.4}$$

We further find that

$$\begin{aligned} |\langle 0', 1\rangle|^2 = |\langle 0', +\rangle|^2 = \cos^2\frac{\pi}{8} \approx 0.8535, \\ |\langle 1', 1\rangle|^2 = |\langle 1', +\rangle|^2 = \sin^2\frac{\pi}{8} \approx 0.1465 \end{aligned} \tag{4.4.5}$$

Bob, hence receives $|\varphi\rangle$ (with the sent state being either $|1\rangle$ or $|+\rangle$) and measures $|\varphi\rangle = |0'\rangle$. The fidelity is then

$$F = \sup_{|\varphi\rangle}(\frac{1}{2}|\langle 1, \varphi\rangle|^2 + \frac{1}{2}|\langle +, \varphi\rangle|^2). = 0.8545 \tag{4.4.6}$$

Assume now that Alice can only send 3 letters, i.e. The question appears if there exists a clever

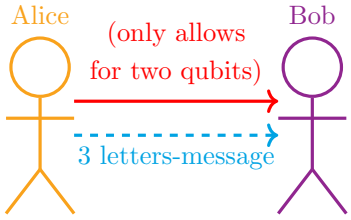


fig. 4.1: 3 letter transmission.

procedure that achieves a higher fidelity. The answer is yes! We find the overlaps

$$\begin{aligned} |\langle 0'0'0', \varphi\rangle|^2 &= \cos^6\frac{\pi}{8} \approx 0.629 \\ |\langle 0'0'1', \varphi\rangle|^2 = |\langle 0'1'0', \varphi\rangle|^2 = |\langle 1'0'0', \varphi\rangle|^2 &= \cos^4\frac{\pi}{8}\sin^2\frac{\pi}{8} \approx 0.1067 \\ |\langle 1'1'0', \varphi\rangle|^2 = |\langle 1'0'1', \varphi\rangle|^2 = |\langle 0'1'1', \varphi\rangle|^2 &= \cos^2\frac{\pi}{8}\sin^4\frac{\pi}{8} \approx 0.0183 \\ |\langle 1'1'1', \varphi\rangle|^2 &= \sin^6\frac{\pi}{8} \approx 0.0031 \end{aligned} \tag{4.4.7}$$

Hence the 'likely' subspace is

$$\Lambda = \text{span}\left\{|0'0'0'\rangle, |0'0'1\rangle, |0'1'0\rangle, |1'0'0'\rangle\right\} \tag{4.4.8}$$

and the 'unlikely' one is $\Lambda^{\perp}$. We find that the probabilities are

$$\begin{aligned} P_{\Lambda} &\approx 0.6219 + 3 \cdot 0.1067 = 0.9419 \\ P_{\Lambda^{\perp}} &\approx 3 \cdot 0.0183 + 0.0031 = 0.0581 \end{aligned} \tag{4.4.9}$$

after making an incomplete orthogonal measurement. The only way to perform this measurement is

1. Alice applies U unitary that rotates the basis of Λ to $\{|\cdot\rangle \otimes |\cdot\rangle \otimes |0\rangle\}$ and the basis of $\Lambda^{\perp}$ to $\{|\cdot\rangle \otimes |\cdot\rangle \otimes |1\rangle\}$

2. Alice measures the third qubit to perform the projection

   - If the outcome is $|0\rangle$
   - If the outcome is $|1\rangle$, Alice's input was projected onto $\Lambda^{\perp}$. She send $|\varphi_{\text{compr}}\rangle$ such that

   $$|\varphi'\rangle = U^{-1}(|\varphi_{\text{compr}}\rangle \otimes |0\rangle) = |0'0'0'\rangle \tag{4.4.10}$$

3. - Bob gets $|\varphi_{\text{compr}}\rangle$ the compressed state from Alice and then decompresses as follows

   $$|\varphi'\rangle = U^{-1}(|\varphi_{\text{compr}}\rangle \otimes |0\rangle) \tag{4.4.11}$$

   - Bob does the same thing as before, i.e.

   $$|\varphi'\rangle = U^{-1}(|\varphi_{\text{compr}}\rangle \otimes |0\rangle) \tag{4.4.12}$$

   He is always obtaining $|0'0'0'\rangle$

The procedure is hence, as follows: Alice encodes qubits $|\varphi\rangle$, she sends two qubits to Bob. Bob then decodes the message and obtains ρ':

$$|\varphi\rangle\langle\varphi| \mapsto \rho' = P_{\Lambda}|\varphi\rangle\langle\varphi|P_{\Lambda} + |0'0'0'\rangle\langle\varphi|(1 - P_{\Lambda})|\varphi\rangle\langle 0'0'0'|. \tag{4.4.13}$$

The fidelity becomes

$$F = \sup_{|\varphi\rangle \in \{|1\rangle, |+\rangle\}^{\otimes 3}} \langle \varphi, \rho'\varphi\rangle = \sup_{|\varphi\rangle \in \{|1\rangle, |+\rangle\}^{\otimes 3}} \{\langle \varphi, P_{\Lambda}\varphi\rangle^2 + (|\langle \varphi, (1 - P_{\Lambda})\varphi\rangle \langle \varphi, 0'0'0'\rangle|)^2\} \approx 0.9234 \tag{4.4.14}$$

If we consider a longer message with more letters the fidelity improves if we compress not to much the fidelity will continue to improve. Naturally we ask the question: How much can we compress and the answer is: The optimal rate of compression is given by the von Neumann entropy. For the specific example we find $S(\rho) \approx 0.6008$. We will give the following theorem which tackles exactly the above question without proof

Theorem. 4.4.1 (Schumacher's theorem) *The optimal rate of compression is given by the von Neumann entropy.*

This theorem is clearly the quantum version of the noiseless Shannon's theorem.

4.5 Entanglement

4.5.1 Entanglement concentration and dilution.

4.5.1.1 LOCC (Local operations and classical communication) maps

This maps describe a method in quantum information theory in which a local operation (product) on a part of a system is performed and the result is communicated classically (usually one needs to perform another local operation on the receiver).

For $r \geq 1$, we define the LOCC_r maps to be the set of LOCC operations that can be achieved with r rounds of classical communication. E.g. the LOCC_1 maps are the set of quantum instruments $\{E_x\}$, with E_x being a CP maps that do not increase the trace and are local for all measurements, i.e.

$$E_x = \bigotimes_j (E_x^j). \tag{4.5.1}$$

We find that

$$\mathrm{LOCC}_1 \subset \mathrm{LOCC}_r \subset \mathrm{LOCC}_{r+1} \subset \mathrm{LOCC}_{\mathbb{N}} \subset \overline{\mathrm{LOCC}}_{\mathbb{N}} \subset \mathrm{SEP} \tag{4.5.2}$$

with

$$\mathrm{SEP} := \{\mathcal{E} \ : \ \mathcal{E}(\rho) = \sum K_1^i \otimes K_2^i \otimes \ldots \otimes K_N^i \rho (K_1^i \otimes K_2^i \otimes \ldots \otimes K_N^i)^*\} \tag{4.5.3}$$

We want to give a fact that tackles the entanglement transformation under these LOCC maps. It states that LOCC maps cannot generate entangled states out of product states. In general, LOCC cannot increase entanglement.

Let now $|\varphi\rangle_{AB}$ an Alice-Bob state. We form n-copies, i.e. $((|\varphi\rangle_{AB})^{\otimes n})$. Lets further assume that Alice and Bob share a large supply of maximally entangled Bell states. $((|\phi^+\rangle_{AB})^{\otimes K})$. The question is now, is there a transformation that transform between those two with high fidelity.

4.5.1.2 Asymptotic setting

Definition. 4.5.1
We say that a rate of conversion R from $|\phi^+\rangle$ to $|\varphi\rangle$ is **asymptotically achievable** if for any $\varepsilon, \delta > 0$ there exists an LOCC protocol

$$\frac{K}{n} \leq R + \delta \tag{4.5.4}$$

which prepares a target state $|\psi^+\rangle^{\otimes n}$ with fidelity $F \geq 1 - \varepsilon$.

Definition. 4.5.2 (Entanglement cost)
We define the entanglement cost to be

$$E_C(|\varphi\rangle) := \inf\{\text{achievable rate for creating } |\varphi\rangle \text{ from Bell states}\} \tag{4.5.5}$$

Definition. 4.5.3
We say that a rate of conversion R' from $|\varphi\rangle$ to $|\phi^+\rangle$ is asymptotically achievable if for any $\varepsilon, \delta > 0$ there exists an LOCC protocol such that

$$\frac{K'}{n} \geq R' - \delta \tag{4.5.6}$$

prepares $|\phi^+\rangle^{\otimes K'}$ with fidelity $F \geq 1 - \varepsilon$.

Definition. 4.5.4 (Distillable entanglement)
We define the distillable entanglement as

$$E_D(|\varphi\rangle) := \sup\{\text{achievable rate for distilling Bell states from } |\varphi\rangle\}. \tag{4.5.7}$$

It clearly holds that

$$E_D(|\varphi\rangle) \leq E_C(|\varphi\rangle) \tag{4.5.8}$$

Example. Let

$$\rho_A = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}^{\otimes K} = \rho_B \tag{4.5.9}$$

We choose $|\varphi\rangle_{AB}$ such that $\rho_{AB} = |\varphi\rangle\langle\varphi|_{AB}$ is mapped to ρ_A and ρ_B under the respective partial trace. Then

$$E_C(|\varphi\rangle) = E_D(|\varphi\rangle) = S(\rho_A) = S(\rho_B). \tag{4.5.10}$$

One finds this in ...

Definition. 4.5.5 (Squashed entanglement)
We define the squashed entanglement as

$$E_{sq}(\rho_{AB}) = \inf\{\frac{1}{2} I_\rho(A:B|C) \ : \ \rho_{AB} = \mathrm{Tr}_C(\rho_{ABC})\} \tag{4.5.11}$$

Proposition. 4.5.6 (Properties of the squashed entanglement) *We have that for $\rho_{AB} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ we find the following*

- $E_{sq}(\rho_{AB} \otimes \sigma_{AB}) = E_{sq}(\rho_A B) + E_{sq}(\sigma_{AB})$
- $E_C \geq E_{sq} \geq E_D$

4.5.2 Entanglement "Monogamy"

We have the following

- Classical correlations are "polyamorous"
- Quantum correlations are not.
  - If ρ_B is pure then $\rho_{ABC} = \rho_{AC} \otimes \rho_B$.
  - If ρ_{AB} is maximally entangled, then $\rho_{ABC} = \rho_{AB} \otimes \rho_C$.
  - If Bob and Charlie share a pure state, then $\rho_{ABC} = \rho_A \otimes \rho_{BC}$.

  We hence have that quantum entanglement implies monogamy and that

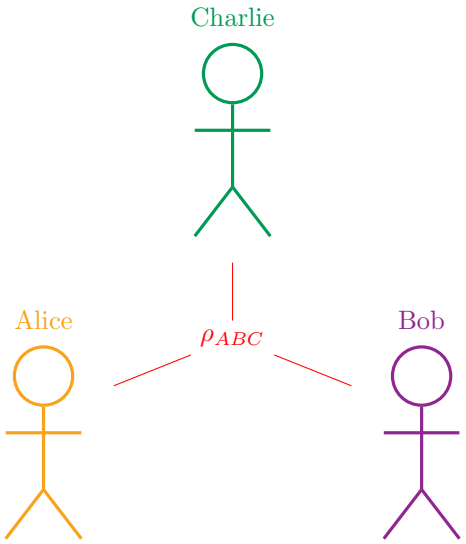
$$E_{sq}(A:B) + E_{sq}(A:C) \leq E_{sq}(A:BC) \tag{4.5.12}$$

We need to proof Equation (4.5.12).

Proof. We have

1. $I_\rho(A:BC) = I_\rho(A:C) + I_\rho(A:B|C)$
2. $I_\rho(A:BC|D) = I_\rho(A:C|D) + I_\rho(A:B|CD)$
3. $I_\rho(A:BC|D) \geq E_{sq}(A:B) + E_{sq}(A:C)$ ρ_{ABCD} is also an extension of ρ_{AB} and ρ_{AC}.
4. Taking the infimum over ρ_{ABCD} now gives the result.

□



4.5.2.1 Accessible Information

The question that leads to the concept of the accessible information is: How much can one learn from a measurement? To answer this question we make the following considerations

- Bob always guesses $\rho(x)$ (with certainty) if $\{\rho(x)\}$ is orthogonal.
- The conditional probability of Bob obtaining outcome y, if Alice sent $\rho(x)$, is $p(x|y) = \mathrm{Tr}[E(y)\rho(x)]$. The joint distribution of Alice & Bob is $p(x,y) = p(y|x)p(x)$.
- Before Bob's ignorance about Alice's state is quantified by $H(X) = S(\rho)$, where $\rho = \sum_x \rho(x)p(x)$.
- After the measurement the ignorance of Bob changes to $H(X|Y) = H(XY) - H(Y)$.

We define

Definition. 4.5.7 (Information gain and accessible information)
The information gain is given by

$$I(X:Y) = H(X) - H(X|Y) \tag{4.5.13}$$

further the accessibel information is given by

$$\mathrm{Acc}(\mathcal{E}) = \max_{E \text{ POVMs}} I(X:Y) = \max_{E \text{ POVMs}} H(X) - H(X|Y) \leq H(X) \tag{4.5.14}$$

Equality in the last inequality holds, if and only if $\{\rho(x)\}$ are orthogonal.

Proposition. 4.5.8 (Holevo Bound) *We find that*

$$Acc(\mathcal{E}) \leq S(\rho) = S\Big(\sum_x |\varphi(x)\rangle\langle\varphi(x)|p(x)\Big) \tag{4.5.15}$$

for the ensemble $\mathcal{E} = \{|\varphi(x)\rangle, p(x)\}_x$.

There are several facts that hold

- For mixed states the bound can be improved.
- Alice sends n qubits to Bob (2^n bits), then one ?? more then n bits.

- T CPTP map $T : \mathcal{S}(\mathcal{H}_B) \to \mathcal{S}(\mathcal{H}_{B'})$, then $I(A : B) \geq I(A : B')$.
- Alice records her chosen state in X (classical)
- Bob records her chosen state in Y (classical)
- Bob's information gain is $I(X : Y)$
- Alice prepares $\rho_{XE} = \sum_x p(x)|x\rangle\langle x| \otimes \rho(x)$
- Bob measures it:

$$\rho(x) \mapsto \sum_y M(y)\rho(x)M(y)^* \otimes |y\rangle\langle y| \tag{4.5.16}$$

  where $E(y) = M(y)^* M(y)$, then one finds

$$\rho'_{XAY} = \sum_{x,y} p(x)|x\rangle\langle x| \otimes \tag{4.5.17}$$

- We have: $I_{\rho'}(X : Y) \leq I_{\rho'}(X : AY) \leq I_\rho(X : Y)$
- If $I_\rho(X : A)$ is an intrinsic property of the ensemble.

We define two quantities

Definition. 4.5.9 (Holevo chi and the Holevo bound)
The Holevo chi is given by $\chi(\mathcal{E}) := I(X : Y)$. The Holevo bound states that $\text{Acc}(\mathcal{E}) \leq \chi(\mathcal{E})$.

- Ensemble of pure states, gives that $\chi(\mathcal{E}) = H(A)$.
- $\mathcal{E} = \{\rho(x), p(x)\}$, $\mathcal{E}' = \{T(\rho(x)), p(x)\}$ with T a CPTP map, then

$$\chi(\mathcal{E}') \leq \chi(\mathcal{E}). \tag{4.5.18}$$

  Hence we can set

$$\chi(T) = \max_{\mathcal{E}'} \chi(\mathcal{E}') = \max_{\mathcal{E}'} I(A : B) \tag{4.5.19}$$

  with $T : \mathcal{E} \to \mathcal{E}'$.

Compare the last point to the classical capacity of a channel

$$C_{cl}(T) \leq \lim_{n\to\infty} \frac{1}{n} \chi(T^{\otimes n}) \tag{4.5.20}$$

where it holds that $\chi(T_1 \otimes T_2) \geq \chi(T_1) + \chi(T_2)$.

4.6 Geometric Renyi divergences and its application in quantum channel capacities [8]

The two goals that we have are

- Study of geometric Renyi divergence
- Application to various channel capacity problems.

4.6.1 Introduction

We assume an imperfect communication link between a sender and receiver, hence a noisy channel. The capacity of a channel is then defined as the maximum rate at which information can be transmitted through the channel reliably. For a classical channel, the (Shannon) capacity is just the mutual information. In the case of a quantum channel, there are several different forms of capacities (classical, quantum, private, with or without assistance, two-way or one-way, etc.). For example, the entanglement-assisted quantum capacity is given by the quantum mutual information. It is hard to find exact representations, and several works, therefore, focus on finding achievable (lower) and ... (upper) bounds.

4.6.2 Desirable criteria (for bounds on capacities)

- **Single-letter:** Depends on a single use of the channel.
- **Computable:** Explicitly computed for a given quantum channel.
- **General:** Holds for arbitrary quantum channels.
- **Strong converse:** If the communication rate exceeds this bound, the fidelity of transmission goes to zero (with many uses of the channel).

4.6.3 Geoemtric Renyi divergences

- Divergence: $D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)) \leq D(\rho\|\sigma)$ with $\mathcal{N}$ a quantum channel.

$$\begin{aligned} \text{Sandwitched} \quad & \widetilde{D}_\alpha(\rho\|\sigma) := \tfrac{1}{\alpha}\log \mathrm{Tr}[(\sigma^{\frac{1-\alpha}{2\alpha}}\rho\sigma^{\frac{1-\alpha}{2\alpha}})^\alpha] \\ & \downarrow \alpha \to 1 \\ & D(\rho\|\sigma) := \mathrm{Tr}[\rho\log\rho - \rho\log\sigma] \\ & \uparrow \alpha \to 1 \\ \text{Petz} \quad & \overline{D}_\alpha(\rho\|\sigma) := \tfrac{1}{\alpha-1}\log \mathrm{Tr}[\rho^\alpha\sigma^{1-\alpha}] \end{aligned} \tag{4.6.1}$$

  We further have

$$\begin{aligned} \text{Geometric} \quad & \widehat{D}_\alpha(\rho\|\sigma) := \tfrac{1}{\alpha-1}\mathrm{Tr}[\sigma^{1/2}(\sigma^{-1/2}\rho\sigma^{-1/2})^\alpha\sigma^{1/2}] \\ & \downarrow \alpha \to 1 \\ & \hat{D}(\rho\|\sigma) := \mathrm{Tr}[\rho\log(\rho^{1/2}\sigma^{-1}\rho^{1/2})] \end{aligned} \tag{4.6.2}$$

  where $G_\alpha(\sigma\|\rho) = \sigma^{1/2}(\sigma^{-1/2}\rho\sigma^{-1/2})^\alpha\sigma^{1/2}$ is called the geometric mean.

- We find the following relations
  - $D(\rho\|\sigma) \leq \widehat{D}(\rho\|\sigma)$
  - $\widetilde{D}_\alpha(\rho\|\sigma) \leq \overline{D}_\alpha(\rho\|\sigma)$
  - $\widehat{D}(\rho\|\sigma) \leq \widehat{D}_\alpha(\rho\|\sigma)$ for $\alpha > 1$
  - ...

And further, summarise properties in the following proposition

Proposition. 4.6.1 *For $\mathcal{N}, \mathcal{M} : \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B) \to \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ CPTP maps.*

- *The geometric Renyi channel divergence is given by*

$$\widehat{D}_\alpha(\mathcal{N}\,\|\,\mathcal{M}) := \max_{\rho_A \in \mathcal{S}(\mathcal{H}_A)} \widehat{D}_\alpha(\mathcal{N}_{A'\to B}(\phi_{AA'})\|\,\mathcal{M}_{A'\to B}(\phi_{AA'}) \tag{4.6.3}$$

  for $\phi_{AA'}$ a purification of ρ_A.

- *We further have* $D(\rho\|\sigma) \leq \widehat{D}(\rho\|\sigma) \leq \widehat{D}_\alpha(\rho\|\sigma) \leq D_{\max}(\rho\|\sigma)$.
- *The closed-form expression is given by*

$$\widehat{D}_\alpha(\mathcal{N} \,\|\, \mathcal{M}) := \frac{1}{\alpha - 1} \log \left\| \mathrm{Tr}[G_{1-\alpha}(J^{\mathcal{N}}_{AB}, J^{\mathcal{M}}_{AB})] \right\|_\infty \tag{4.6.4}$$

- $\widehat{D}_\alpha(\mathcal{N}_1 \otimes \mathcal{N}_2 \,\|\, \mathcal{M}_1 \otimes \mathcal{M}_2) = \widehat{D}_\alpha(\mathcal{N}_1 \,\|\, \mathcal{M}_1) + \widehat{D}_\alpha(\mathcal{N}_2 \,\|\, \mathcal{M}_2)$
- *Chain rule:*

$$\widehat{D}_\alpha(\mathcal{N}_{A\to B}(\rho_{AR}) \| \, \mathcal{M}_{A\to B}(\sigma_{AR})) \leq \widehat{D}_\alpha(\rho_{AR}\|\sigma_{AR}) + \widehat{D}_\alpha(\mathcal{N}_{A\to B} \,\|\, \mathcal{M}_{A\to B}) \tag{4.6.5}$$

- *Semi-definite representation:* $\alpha(l) = 1 + 2^{-l}$, $l \in \mathcal{N}$. $\mathcal{V} :=$ *Set of subchannels characterized by certain semidefinite conditions. We compute*

$$\min_{\mathcal{M}\in\mathcal{V}} \widehat{D}_\alpha(\mathcal{N} \,\|\, \mathcal{M}) \tag{4.6.6}$$

using an SDP:

$$\begin{array}{ll} \textit{Compute:} & 2^l \log \min y \\ \textit{subject to} & \mathcal{M}, \{\mathcal{N}_i\}_{i=0}^l,\ J_{\mathcal{M}},\ y \textit{ are Hermitian.} \\ & \begin{pmatrix} \mathcal{M} & J_{\mathcal{N}} \\ J_{\mathcal{N}} & \mathcal{N}_l \end{pmatrix}, \left\{ \begin{pmatrix} J_{\mathcal{N}} & \mathcal{N}_i \\ \mathcal{N}_i & NN_{i-1} \end{pmatrix} \right\}_{i=1}^{l} \geq 0 \\ & y\,\mathbb{1}_A \geq \mathrm{Tr}_B[\mathcal{M}],\ \mathcal{N}_0 = J_{\mathcal{M}} \textit{ with } \mathcal{M} \in \mathcal{V} \end{array}$$

Definition. 4.6.2 (Amortized channel divergence)

$$\widehat{D}^a_\alpha := \max_{\rho_{AR},\sigma_{AR}} \left[\widehat{D}_\alpha(\mathcal{N}_{A\to B}(\rho_{AR}) \| \, \mathcal{M}_{A\to B}(\sigma_{AR})) - \widehat{D}_\alpha(\rho_{AR}\|\sigma_{AR}) \right] \tag{4.6.7}$$

From the chain rule, we immediately get

$$\widehat{D}^a_\alpha(\mathcal{N} \,\|\, \mathcal{M}) = \widehat{D}_\alpha(\mathcal{N} \,\|\, \mathcal{M}\,0 \tag{4.6.8}$$

the amortized collapse for $\alpha \in (1, 2]$

4.6.4 Quantum communication

- Quantum capacities: $\begin{cases} *) \text{ (Unassisted) quantum capacity } Q \\ *) \text{ Two-way assisted capacity } Q^{\leftrightarrow} \end{cases}$

Definition. 4.6.3 (Quantum capacity)
The quantum capacity or regularized coherent information is given by

$$Q(\mathcal{N}) := \lim_{n\to\infty} \frac{1}{n} I_c(\mathcal{N}^{\otimes n}) \tag{4.6.9}$$

where

$$I_c(\mathcal{N}) := \max_{\rho\in SS(\mathcal{H})} [S(\mathcal{N}(\rho)) - S(\mathcal{N}^c(\rho))] \tag{4.6.10}$$

and $\mathcal{N}^c$ the complementary channel.

One finds the following, so-called generalized Rains bounds:

$$\mathcal{R}(\rho_{AB}) := \min_{\sigma_{AB}\in PPT'(A:B)} D(\rho_{AB}\|\sigma_{AB}) \tag{4.6.11}$$

where

$$PPT'(A:B) = \{\sigma_{AB} \ : \ \sigma_{AB} \geq 0,\ \sigma^{T_B}_{AB} \geq 0,\ \left\|\sigma^{T_B}_{AB}\right\|_1 \leq 1\} \tag{4.6.12}$$

Definition. 4.6.4 (Generalised Rains information (induced by $\widehat{D}_\alpha$)
We define the generalised Rains information as

$$\widehat{R}(\mathcal{N}) := \max_{\rho_A \in \mathcal{S}(\mathcal{H}_A)} \min_{\sigma_{AB} \in PPT'(A:B)} \widehat{D}_\alpha(\mathcal{N}_{A' \to B}(\phi_{AA'}) \| \sigma_{AB}) \tag{4.6.13}$$

with $\phi_{AA'}$ purification of ρ_A.

Theorem. 4.6.5

$$Q(\mathcal{N}) \leq Q^{\dagger}(\mathcal{N}) \leq R(\mathcal{N}) \leq \widehat{R}_\alpha(\mathcal{N}) \leq R_{\max}(\mathcal{N}) \tag{4.6.14}$$

Chapter 5

Miscellanea

5.1 Monotonicity of the relative entropy

Theorem. 5.1.1 (Data processing inequality of the relative entropy) *Let $\rho, \sigma \in \mathcal{S}(\mathcal{H})$, $T : \mathcal{H} \to \mathcal{H}$ a CPTP map. Then we have*

$$D(\rho\|\sigma) \geq D(T(\rho)\|T(\sigma))\,. \tag{5.1.1}$$

We further get that equality holds, i.e. $D(\rho\|\sigma) = D(T(\rho)\|T(\sigma))$, if and only if there exists $\mathcal{R}_T^\sigma$ a recovery map such that $\mathcal{R}_T^\sigma(\rho) = \rho$, which is equivalent to

$$\rho = \sigma^{1/2} T^*(T(\sigma)^{-1/2} T(\rho) T(\sigma)^{1/2}) \sigma^{1/2} = \mathcal{P}_T^\sigma(\rho) \tag{5.1.2}$$

We are now interested in the distance between

$$D(\rho\|\sigma) - D(T(\rho)\|T(\sigma) \geq \text{"dist}(\rho, \mathcal{P}_T^\sigma(\rho))\text{"} \tag{5.1.3}$$

the ρ and its recovery.

Theorem. 5.1.2 (Fawzi-Renner, '1.) *If there exists a recovery map $\mathcal{R}_T^\sigma$ then*

$$D(\rho\|\sigma) - D(T(\rho)\|T(\sigma) \geq -\log F(\rho, (\mathcal{R}_T^\sigma \circ)(\rho)) \geq 0 \tag{5.1.4}$$

with the fidelity

$$F(\rho, \sigma) := \left\|\sqrt{\rho}\sqrt{\sigma}\right\|_1^2 \tag{5.1.5}$$

For the proof of this theorem we want to recall some facts from complex analysis

5.1.1 Brief overview on complex analysis

- $f : \mathbb{C} \to \mathbb{C}$, derivative at $z_0 \in \mathbb{C}$,

$$\frac{df}{dz}|_{z=z_0} = \lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0} \tag{5.1.6}$$

- $U \subseteq \mathbb{C}$ an open set, if f is differentiable at $z_0 \in U$ for every $z_0 \in U \Rightarrow f$ is holomorphic in U.

- $f(x + iy) = u(x, y) + iv(x, y)$, for $x, y \in \mathbb{R}$, and $u, v : \mathbb{R} \to \mathbb{R}$ if f is holomorphic $\Rightarrow$ Cauchy-Riemann equations.

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} \tag{5.1.7}$$

- If the first partial derivative of u and v are continuous and satisfy CR equations $\Rightarrow$ f is holomorphic.

Theorem. 5.1.3 (Liouville's theorem) *For $f : \mathbb{C} \to \mathbb{C}$ holomorphic on $U \subseteq \mathbb{C}$, open, bounded and connected. Then, if $z_0 \in U$ s.t. $|f(z_0)| \geq |f(z)|$ $\forall z$ in a neighbourhood (open set containing z_0) of $z_0 \Rightarrow$ f is constant in U.*

Corollary. 5.1.3.1 (Maximum modulus principle on a strip) *Let*

$$S = \{z \in \mathbb{C} \mid 0 < Re(z) < 1\} \tag{5.1.8}$$

$$\overline{S} = \{z \in \mathbb{C} \mid 0 \leq Re(z) \leq 1\} \tag{5.1.9}$$

$$\partial\overline{S} = \{z \in \mathcal{C} \mid Re(z) = 0 \text{ or } Re(z) = 1\} \tag{5.1.10}$$

If $f : \overline{S} \to \mathcal{C}$ is holomorphic on S, and continuous on $\partial\overline{S}$, then the supremum of $|f|$ is attained on $\partial\overline{S}$.

Theorem. 5.1.4 (Hadamard three-lines) *Let $f : \overline{S} \to \mathcal{C}$ be bounded on $\overline{S}$, holomorphic on S and continuous on $\partial\overline{S}$. Let $\theta \in (0, 1)$ and*

$$M(\theta) := \sup_{t\in\mathbb{R}} |f(\theta + it)| . \tag{5.1.11}$$

Then, $\log M(\theta)$ is convex, which gives that

$$\log M(\theta) \leq (1 - \theta) \log M(0) + \theta \log M(1) \tag{5.1.12}$$

Theorem. 5.1.5 (Hirschmann) *Let $f : \overline{S} \to \mathbb{C}$ bounded on $\overline{S}$, holomorphic on S and continous on $\partial\overline{S}$. Then $\theta \in (0, 1)$,*

$$\log |f(\theta)| \leq \int_{-\infty}^{\infty} dt\Big(\alpha_\theta(t) \log[|f(it)|^{1-\theta}] + \beta_\theta(t) \log[|f(1 + it)|^\theta]\Big) \tag{5.1.13}$$

with

$$\alpha_\theta(t) = \frac{\sin(\pi\theta}{2(1 - \theta)[\cosh(\theta t] - cos(\pi\theta)}, \qquad \beta_0(t) = \frac{\sin(\pi\theta)}{2\theta[\cosh(\pi t) + \cos(\pi\theta)]}. \tag{5.1.14}$$

Taking the limit $\theta \to 0$, we find

$$\frac{\pi}{2(\cosh(\pi t) + 1)} \int_{-\infty}^{\infty} dt\beta_\theta(t) = \int_{-\infty}^{\infty} dt\alpha_\theta(t) = 1 \qquad \alpha_\theta(t), \beta_0(t) \geq 0 \; \forall t, \forall\theta \tag{5.1.15}$$

Theorem. 5.1.6 (Stein-Hirschman) *$G : \overline{S} \to \mathcal{B}(\mathcal{H})$ operator-valued function, with G bounded on $\overline{S}$, holomorphic on S, continuous on $\partial\overline{S}$. Let $\theta \in (0, 1)$ and define*

$$\frac{1}{p_\theta} = \frac{1 - \theta}{p_0} + \frac{\theta}{p_1}, \quad p_0, p_1 \in [1, \infty] . \tag{5.1.16}$$

We then find that

$$\log \|G(\theta)\|_{p_\theta} \leq \int_{-\infty}^{\infty} dt\big(\alpha_\theta(t) \log[\|G(it)\|_{p_0}^{1-\theta}] + \beta_\theta(t) \log[\|G(1 + it)\|_{p_1}^{\theta}]\big) \tag{5.1.17}$$

Proof of Theorem 5.1.2.

- Main ingredient: Interpolation
- Isometric map:

$$\mathcal{U}_{\sigma,t}(M) := \sigma^{it} M \sigma^{-it}, \qquad \mathcal{U}_{\sigma,t}(\sigma) = \sigma \quad \forall t \in \mathbb{R} \tag{5.1.18}$$

- Rotated Petz recovery map:

$$\mathcal{R}_T^{\sigma,t})(X) := (\mathcal{U}_{\sigma,-t} \circ \mathcal{P}_T^{\sigma} \circ \mathcal{U}_{T(\sigma),t})(X) \tag{5.1.19}$$

- Renyi infromation measure

$$\begin{aligned} \widetilde{\Delta}_\alpha(\rho,\sigma,T) &:= \frac{1}{\alpha-1}\log \widetilde{Q}_\alpha(\rho,\sigma,T) \\ &= \frac{1}{\alpha-1}\log \left\| (T(\rho)^{\frac{1-\alpha}{2\alpha}} T(\sigma)^{\frac{\alpha-1}{2\alpha}} \otimes \mathbb{1}_E) U \sigma^{\frac{1-\alpha}{2\alpha}} \rho^{\frac{1}{2}} \right\|_{2\alpha}^{2\alpha} \end{aligned} \tag{5.1.20}$$

  where $U : \mathcal{H} \to \mathcal{H}' \otimes \mathcal{H}_E$ is the isometric extension of T.

- $\lim_{\alpha\to 1} \widetilde{\Delta}_\alpha(\rho,\sigma,T) = D(\rho\|\sigma) - D(T(\rho)\|T(\sigma))$

- $\widetilde{\Delta}_{\alpha=\frac{1}{2}}(\rho,\sigma,T) = -\log F(\rho, \mathcal{P}_T^{\sigma} \circ T(\rho))$.

- We find that

$$D(\rho\|\sigma) - D(T(\rho)\|T(\sigma)) \geq -\int_{-\infty}^{\infty} dt\, \beta_0(t) \log[F(\rho, (\mathcal{R}_T^{\sigma,\frac{t}{2}} \circ T)(\rho))] \tag{5.1.21}$$

  and with

$$G(z) := (T(\rho)^{z/2} T(\sigma)^{-z/2} \otimes \mathbb{1}_E) U \sigma^{z/2} \rho^{1/2} \tag{5.1.22}$$

  $p_0 = 2$, $p_1 = 1$, $\theta \in (0,1)$, $p_\theta = \frac{2}{1+\theta}$. With those we find

$$\begin{aligned} \|G(\theta)\|_{p_\theta = \frac{2}{1+\theta}} &= \left\| (T(\rho)^{\frac{\theta}{2}} T(\sigma)^{-\frac{\theta}{2}} \otimes \mathbb{1}_E) U \sigma^{\frac{\theta}{2}} \rho^{\frac{1}{2}} \right\|_{\frac{2}{1+\theta}} \\ \|G(it)\|_2 &= \left\| (T(\rho)^{\frac{it}{2}} T(\sigma)^{-\frac{it}{2}} \otimes \mathbb{1}_E) U \sigma^{\frac{it}{2}} \rho^{\frac{1}{2}} \right\|_{\frac{2}{1+\theta}} \leq \left\| \rho^{1/2} \right\|_2 = 1 \\ \|G(1+it)\|_1 &= \left\| (T(\rho)^{\frac{1+it}{2}} T(\sigma)^{-\frac{1+it}{2}} \otimes \mathbb{1}_E) U \sigma^{\frac{1+it}{2}} \rho^{\frac{1}{2}} \right\|_1 = \sqrt{F}(\rho, \mathcal{R}_T^{\sigma,\frac{t}{2}} \circ T(\rho)) \,. \end{aligned} \tag{5.1.23}$$

  Applying Stein-Hirschman gives

$$\log \left\| (T(\rho)^{\frac{\theta}{2}} T(\sigma)^{-\frac{\theta}{2}} \otimes \mathbb{1}_E) U \sigma^{\frac{\theta}{2}} \rho^{\frac{1}{2}} \right\|_{\frac{2}{1+\theta}} \leq \int dt \beta_\theta(t) \log[(F(\rho, \mathcal{R}_T^{\sigma,\frac{t}{2}} \circ T(\rho))^{\frac{\theta}{2}}] \tag{5.1.24}$$

  modifying this a little we find

$$-\frac{2}{\theta}\log \left\| (T(\rho)^{\frac{\theta}{2}} T(\sigma)^{-\frac{\theta}{2}} \otimes \mathbb{1}_E) U \sigma^{\frac{\theta}{2}} \rho^{\frac{1}{2}} \right\|_{\frac{2}{1+\theta}} \leq -\int dt \beta_\theta(t) \log[F(\rho, \mathcal{R}_T^{\sigma,\frac{t}{2}} \circ T(\rho)]. \tag{5.1.25}$$

  Now setting $\theta = \frac{1-\alpha}{\alpha}$, gives

$$\widetilde{\Delta}_\alpha(\rho,\sigma,T) \geq -\int_{-\infty}^{\infty} dt \beta_{\frac{1-\alpha}{\alpha}}(t) \log[F(\rho, \mathcal{R}_T^{\sigma,\frac{t}{2}} \circ T(\rho))] \tag{5.1.26}$$

  taking the limit $\alpha \to 1$ gives now

$$D(\rho\|\sigma) - D(T(\rho)\|T(\sigma) \geq -\int_{-\infty}^{\infty} \beta_0(t) \log[F(\rho, \mathcal{R}_T^{\sigma,\frac{t}{2}} \circ T(\rho))] \tag{5.1.27}$$

□

From this result one immediately gets that $D(\rho\|\sigma) = D(T(\rho)\|T(\sigma))$ holds if and only if $\mathcal{R}_{\sigma,T}^{\sigma,t}(\rho) = \rho$ for all $t \in \mathcal{R}$.

5.1.2 Chainrule of quantum channels [4]

Definition. 5.1.7
Let P, Q be probability densities over $X = \{x\}$. Let further $\alpha \in (1, \infty)$, further

$$D_\alpha(P\|Q) = \frac{1}{\alpha - 1} \log Q_\alpha(P\|Q) \tag{5.1.28}$$

with

$$Q_\alpha(P\|Q) = \begin{cases} \sum\limits_{x \in X} P(x) \left(\frac{P(x)}{Q(x)}\right)^{\alpha - 1} & P \ll Q \\ +\infty \end{cases} \tag{5.1.29}$$

For $\alpha \to 1$, we find that

$$D_{KL}(P\|Q) = \begin{cases} \sum\limits_{x \in X} P(x) \log \left(\frac{P(x)}{Q(x)}\right) & P \ll Q \\ +\infty \end{cases} \tag{5.1.30}$$

For $\alpha \to \infty$ we find

$$D_\alpha(P\|Q) = \max_x \log \frac{P(x)}{Q(x)} \tag{5.1.31}$$

For two systems X, Y one can apply the chain rule. Further for $\alpha \in [1, \infty]$ we find

$$D_\alpha(P_{XY}\|Q_{X,Y}) \leq D_\alpha(P_X\|Q_X) + \max_{x \in X} D(P_{Y|X=x}\|Q_{Y|X=x} \tag{5.1.32}$$

In the quantum setting for $\rho, \sigma \in \mathcal{S}(\mathcal{H})$.

Definition. 5.1.8
We define

$$\mathbb{D}_\alpha(\rho\|\sigma) := \frac{1}{\alpha - 1} \log \mathbb{Q}_\alpha(\rho\|\sigma) \tag{5.1.33}$$

which has the following properties

1. If $[\rho, \sigma] = 0$, then $\mathbb{D}_\alpha(\rho\|\sigma) = D_\alpha(\rho\|\sigma)$. Classical quantum states, for $\rho \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ with

$$\begin{aligned} \rho &= \sum_{x \in X} p_x |x\rangle\langle x| \otimes \rho^x \\ \sigma &= \sum_{x \in X} q_x |x\rangle\langle x| \otimes \sigma^x \end{aligned} \tag{5.1.34}$$

   we find

$$\mathbb{Q}_\alpha(\rho\|\sigma) = \sum_x p_x^\alpha q_x^{1-\alpha}\, \mathbb{Q}_\alpha(\rho^x\|\sigma^x) \tag{5.1.35}$$

2. Data processing inequality. We find

$$\mathbb{D}_\alpha(\rho\|\sigma) \geq \mathbb{D}_\alpha(T(\rho)\|T(\sigma)) \tag{5.1.36}$$

   with T a quantum channel.

The smallest of those is the Renyi quantum divergence: Measured Renyi divergence defined as

$$D_\alpha^M(\rho\|\sigma) := \sup_M D_\alpha(M(\rho)\|M(\sigma)) \tag{5.1.37}$$

with M a rank-one projective measurment, defined using a ONB $\{|x\rangle\}$ and

$$M(\cdot) = \sum_x \langle x, \cdot x\rangle\, |x\rangle\langle x|. \tag{5.1.38}$$

The largest are the Geometric Renyi divergence, defined via

$$\hat{D}_\alpha(\rho\|\sigma) := \frac{1}{\alpha-1}\hat{Q}_\alpha(\rho\|\sigma) \tag{5.1.39}$$

where

$$\hat{Q}_\alpha(\rho\|\sigma) := \mathrm{Tr}[\sigma^{1/2}(\sigma^{-1/2}\rho\sigma^{-1/2})^\alpha\sigma^{1/2}] \tag{5.1.40}$$

In the limit:

$$\hat{D}_\alpha(\rho\|\sigma) \overset{\alpha\to1}{\longrightarrow} \hat{D}(\rho\|\sigma) := \mathrm{Tr}[\rho\log(\rho^{1/2}\sigma^{-1}\rho^{1/2})] \tag{5.1.41}$$

Definition. 5.1.9 (Pinching map)
The pinching map is defined via $\sigma = \sum_\lambda \lambda P_\lambda$ with $\mathcal{P}_\sigma : \rho \mapsto \sum P_\lambda\rho P_\lambda$. It has the following properties:

1. $[\mathcal{P}_\sigma(\rho),\sigma] = 0\ \forall\rho$
2. $\mathrm{Tr}[\mathcal{P}_\sigma(\rho)\sigma] = \mathrm{Tr}[\rho\sigma]\ \forall\rho$
3. $\mathcal{P}_\sigma(\rho) \geq |\operatorname{spec}(\sigma)|^{-1}\rho$, which scales as $|\operatorname{spec}(\sigma^{\otimes n})| = O(\mathrm{poly}(n))$
4. $\mathcal{P}_\alpha(\cdot) = \int\mu(dt)\sigma^{it}(\cdot)\sigma^{-it}$

Lemma. 5.1.10 (Matsumoto) *For $\rho,\sigma\in\mathcal{S}(\mathcal{H})$, $\alpha\in(0,2]$, we have*

$$\hat{D}_\alpha(\rho\|\sigma) = \inf_{(P,Q,\Gamma)} D_\alpha(P\|Q) \tag{5.1.42}$$

with $\Gamma(P)=\rho$, $\Gamma(Q)=\sigma$, where Γ is a quantum channel. The infimum is attained for the following map: We first decompose

$$\sigma^{-1/2}\rho\sigma^{-1/2} = \sum_{x\in X}\lambda_x\Pi_x \tag{5.1.43}$$

and then define

$$\Gamma(\cdot) = \sum_x \frac{\langle x,\cdot x\rangle}{Q(x)}\sigma^{1/2}\Pi_x\sigma^{1/2} \tag{5.1.44}$$

with $Q(x) := \mathrm{Tr}[\sigma\Pi_x]$, $P(x) := \lambda_x Q(x)$

Definition. 5.1.11
If in addition to Items 1 and 2 we demand additivity, we obtain the Sandwiched Renyi divergences, defined as

$$\widetilde{D}_\alpha(\rho\|\sigma) := \frac{1}{\alpha-1}\log\widetilde{Q}_\alpha(\rho\|\sigma) \tag{5.1.45}$$

with

$$\widetilde{Q}_\alpha(\rho\|\sigma) := \mathrm{Tr}[(\sigma^{\frac{1-\alpha}{2\alpha}}\rho\sigma^{\frac{1-\alpha}{2\alpha}})^\alpha] \tag{5.1.46}$$

we find for $\alpha\to1$

$$D(\rho\|\sigma) := \mathrm{Tr}[\rho(\log\rho-\log\sigma)] \tag{5.1.47}$$

and for $\alpha\to\infty$

$$\widetilde{D}_\infty(\rho\|\sigma) := \inf\{\lambda\in\mathcal{R}\ :\ \rho 2^\lambda\sigma\} \tag{5.1.48}$$

Lemma. 5.1.12 *Let $\rho,\sigma\in\mathcal{S}(\mathcal{H})$, T_1,T_2 quantum channels, $\alpha\in(0,\infty)$. Then*

$$\widetilde{Q}_\alpha(T_1(\rho)\|T_2(\sigma)) \leq |\operatorname{spec}(\sigma)|^\alpha\widetilde{Q}_\alpha(T_1\circ\mathcal{P}(\rho)\|T_2(\sigma)). \tag{5.1.49}$$

Proof. We have

$$T_1(\rho) \leq |\operatorname{spec}(\sigma)| T_1 \circ \mathcal{P}_\sigma(\rho) \tag{5.1.50}$$

then

$$\operatorname{Tr}[(T_2(\sigma)^{\frac{1-\alpha}{2\alpha}} T_1(\rho) T_2(\sigma)^{\frac{1-\alpha}{2\alpha}})^\alpha] \leq |\operatorname{spec}(\sigma)|^\alpha \operatorname{Tr}[(T_2(\sigma)^{\frac{1-\alpha}{2\alpha}} T_1 \circ \mathcal{P}_\sigma(\rho) T_2(\sigma)^{\frac{1-\alpha}{2\alpha}})^\alpha] \tag{5.1.51}$$

as the trace is monotone and $t \mapsto t^\alpha$ as well. □

Definition. 5.1.13 (Renyi Divergence of quantum channels)
Let T_1, T_2 CPTP maps, then the Renyi Divergence of quantum channels is defined as

$$\mathbb{D}_\alpha(T_1 \| T_2) := \sup_{\rho \in \mathcal{S}(\mathcal{H})} \mathbb{D}_\alpha(T_1(\rho) \| T_2(\rho)) \tag{5.1.52}$$

Theorem. 5.1.14 (Chain rule for quantum channels) *Let $\rho, \sigma \in \mathcal{S}(\mathcal{H})$, T_1, T_2 CPTP maps, $\alpha \in (0, \infty)$, then*

$$\mathbb{D}_\alpha(T_1(\rho) \| T_2(\sigma)) \leq \widehat{D}_\alpha(\rho \| \sigma) + \mathbb{D}_\alpha(T_1 \| T_2) \tag{5.1.53}$$

In particular, $\hat{D}_\alpha(T_1(\rho) \| T_2(\sigma)) \leq \widehat{D}_\alpha(\rho \| \sigma) + \widehat{D}_\alpha(T_1 \| T_2)$.

Proof. Let $T_1(\rho) = \sum_x \hat{p}_x T_1(\Gamma(|x\rangle\langle x|))$ and $T_2(\sigma) = \sum_x \hat{q}_x T_2(\Gamma(|x\rangle\langle x|))$. Then

$$\begin{aligned}
\mathbb{D}_\alpha(T_1(\rho) \| T_2(\sigma)) &\leq \mathbb{D}_\alpha\Big(\sum_x \hat{p}_x T_1(\Gamma(|x\rangle\langle x|)) \| \sum_x \hat{q}_x T_2(\Gamma(|x\rangle\langle x|))\Big) \\
&= \frac{1}{\alpha - 1} \log \sum_x \hat{p}_x^\alpha \hat{q}_x^{1-\alpha} 2^{\mathbb{D}_\alpha(T_1(\Gamma(|x\rangle\langle x|)) \| T_2(\Gamma(|x\rangle\langle x|)))} \\
&\leq D_\alpha(\hat{p} \| \hat{q}) + \underbrace{\max_x \mathbb{D}_\alpha(T_1(\rho^x) \| T_2(\rho^x))}_{\leq \mathbb{D}_\alpha(T_1 \| T_2)}
\end{aligned} \tag{5.1.54}$$

where we used that $\Gamma(|x\rangle\langle x|) = \rho^x$ □

We have that

$$\widehat{D}(T_1 \| T_2) = \lim_{n \to \infty} \frac{1}{n} \hat{D}(T_1^{\otimes n} \| T_2^{\otimes n}) \tag{5.1.55}$$

further we can define

$$\hat{D}_\alpha^\infty(T_1 \| T_2) = \lim_{n \to \infty} \frac{1}{n} \mathbb{D}_\alpha(T_1^{\otimes n} \| T_2^{\otimes n}) \tag{5.1.56}$$

and the stabilised version of this

$$\mathbb{D}_\alpha^{stab}(T_1 \| T_2) = \sup_{\rho_{AR} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{S}_R)} \mathbb{D}_\alpha((T_1 \otimes I_R)(\rho_{AR}) \| (T_2 \otimes I_R)(\rho_{AR})) \tag{5.1.57}$$

and now

$$\mathbb{D}_\alpha^{stab,\infty}(T_1 \| T_2) = \lim_{n \to \infty} \frac{1}{n} \mathbb{D}_\alpha^{stab}(T_1^{\otimes n} \| T_2^{\otimes n}) \tag{5.1.58}$$

The question is now if

$$\lim_{\alpha \to 1} \widehat{D}_\alpha^{stab,\infty}(T_1 \| T_2) = \widehat{D}^{stab}(T_1 \| T_2) \tag{5.1.59}$$

Appendix A

Interlude

A.1 Quantum Many Body Systems

A.1.1 Master Equation

The master equation is an approximate version of the physical processes that are happening. It comes in the form of a differential equation which constitutes a good approximation to the evolution of a density matrix on a system S.

- Coherent case (evolution of a closed system). The system is isolated which means the dynamics of the system is described by the Schrödinger equation (infinitesimal). We obtain the global evolution by integrating.

- Decoherent case (open system). We make one fundamental assumption, namely Markovianity, i.e for

$$t \mapsto \rho(t) \tag{A.1.1}$$

$\rho(t+dt)$ only depends on $\rho(t)$ (but not on previous times). Differently put the environment holds no memory. We can now look at the system and its environment as a closed system to obtain:

$$\rho_{SE} = \rho_S \otimes \rho_E \overset{\text{QC}}{\mapsto} U_{SE}(\rho_S \otimes \rho_E)U_{SE}^* \overset{\text{QC}}{\mapsto} \text{Tr}_E[U_{SE}(\rho_S \otimes \rho_E)U_{SE}^*] \tag{A.1.2}$$

This means

$$\rho_S \mapsto \rho_S' := \text{Tr}_E[U_{SE}(\rho_S \otimes \rho_E)U_{SE}^*] \tag{A.1.3}$$

is a quantum channel (CPTP map). This is still an infinitesimal description. We further assume that there is only weak coupling between the system and the environment. Meaning we can do something like the following

$$\text{"}\underbrace{\rho_S(t) \otimes \rho_E}_{\rho_{SE}(t)} \overset{dt}{\mapsto} \rho_{SE}(t+dt) = \rho_S(t+dt) \otimes \rho_E\text{"}, \tag{A.1.4}$$

i.e. the environment does not evolve. We now frame this concepts a little more formally. The Markovian approximation allows us to describe the system using a quantum Markov

semigroup (QMS), which is a 1-parameter semigroup[1] $\{T_t\}_{t\geq 0}$ of CPTP maps

$$T_t : \mathcal{S}(\mathcal{H}) \to \mathcal{S}(\mathcal{H}) \tag{A.1.5}$$

We obtain the generator of the group by differentiation

$$\frac{d}{dt}T_t = \mathcal{L}\circ T_t = T_t \circ \mathcal{L}\,. \tag{A.1.6}$$

This is called the Liouville equation. It gives us the generator of the group

$$T_t = e^{t\,\mathcal{L}} \quad \text{with} \quad \mathcal{L} \text{ the Liouvillian} \tag{A.1.7}$$

With all the above at hand, we will now formalise our two description above

- Closed system:

$$\dot{\rho} = -i[H_S, \rho] \tag{A.1.8}$$

with the solution

$$\rho(t) = e^{-iHt}\rho(0) \tag{A.1.9}$$

- Open system: The Hamilton operator can be decomposed as follows

$$H = H_S + H_E + H_{SE}. \tag{A.1.10}$$

The evolution equation becomes

$$\dot{\rho} = \mathcal{L}[\rho] \tag{A.1.11}$$

and its solution

$$\rho(t) = e^{t\,\mathcal{L}}\rho(0). \tag{A.1.12}$$

This gives us using the Kraus decomposition for the channel T_t

$$\rho(t) = T_t(\rho(0)) = \sum_\mu M_\mu(t)\rho(0)M_\mu(t)^*, \tag{A.1.13}$$

with $M_\mu(t)$ called jump operators. We have

$$\rho(dt) = \rho(0) + O(dt). \tag{A.1.14}$$

We have

- $M_0 = \mathbb{1} + O(dt)$
- Others: Order $O(\sqrt{dt})$

meaning

$$\begin{aligned} M_\mu &= \sqrt{dt}L_\mu \\ M_0 &= \mathbb{1} + (-iH_S + K)dt. \end{aligned} \tag{A.1.15}$$

We need

$$\mathbb{1} = \sum_\mu M_\mu^*(t)M_\mu(t) = \mathbb{1} + dt\Big(2K + \sum_{\mu>0} L_\mu^* L_\mu\Big) \tag{A.1.16}$$

[1] We have the properties

- $T_0 = \mathbb{1}$
- $T_s \circ T_t = T_{t+s}$

from which we get

$$K = -\frac{1}{2}\sum_{\mu>0} L_\mu^* L_\mu \tag{A.1.17}$$

giving us the master equation

$$\dot{\rho} = \mathcal{L}[\rho] = -i[\tilde{H}_s, \rho] + \sum_{\mu>0}(L_\mu \rho L_\mu^* - \frac{1}{2}\rho L_\mu^* L_\mu \rho) \tag{A.1.18}$$

A.2 Operator monotone functions

Definition. A.2.1 (Operator montone function)
A function $f : I \subset \mathbb{R} \to \mathbb{R}$ it called operator monotone, if

$$f(A) \leq f(B) \tag{A.2.1}$$

for A, B Hermitian operators with spectrum in I and $A \leq B$.

Definition. A.2.2 (Operator convex function)
A function $f : I \subset \mathbb{R} \to \mathbb{R}$ is operator convex if

$$f(\lambda A + (1-\lambda)B) \leq \lambda f(A) + (1-\lambda) f(B) \tag{A.2.2}$$

for Hermitian A, B with spectrum in I and $\lambda \in [0, 1]$

Theorem. A.2.3 (Loewner-Heinz theorem)

- $-1 \leq P \leq 0$, $f(t) = -t^p$ *operator monotone and operator convex.*
- $0 < p \leq 1$, $f(t) = t^p$ *operator monotone and operator concave.*
- $1 < p \leq 2$, $f(t) = t^p$ *operator convex.*
- $f(t) = log(t)$ *is operator monotone and operator concave.*
- $f(t) = t\log(t)$ *is operator convex.*

Bibliography

[1] J.S. Bell. "On the Einstein Podolsky Rosen paradox". In: *Physics Physique Fizika* 1 (1964), p. 195.

[2] C. Bennet and S. Wiesner. "Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states". In: *Physical Review Letters* 69 (1992), pp. 2881–2884.

[3] C. Bennet et al. "Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels". In: *Physical Review Letters* 70 (1993), pp. 1895–1899.

[4] Mario Berta and Marco Tomamichel. *Chain rules for quantum channels*. 2022. DOI: 10.48550/ARXIV.2204.11153. URL: https://arxiv.org/abs/2204.11153.

[5] Eric Carlen. "Trace inequalities and quantum entropy: An introductory course". In: 529 (Jan. 2010). DOI: 10.1090/conm/529/10428.

[6] A. Coladangelo and J. Stark. "Unconditional separation of finite and infinite-dimensional quantum correlations". In: *Nature Communications* 11 (2020), pp. 1–6.

[7] A. Einstein, B. Podolsky, and N. Rosen. "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?" In: *Physical Review* 47 (1935), p. 777.

[8] Kun Fang and Hamza Fawzi. "Geometric Rényi Divergence and its Applications in Quantum Channel Capacities". In: *Communications in Mathematical Physics* 384.3 (May 2021), pp. 1615–1677. DOI: 10.1007/s00220-021-04064-4. URL: https://doi.org/10.1007%2Fs00220-021-04064-4.

[9] Z. Ji et al. "MIP* = RE". In: *Communications of the ACM* 64.11 (2021), pp. 131–138.

[10] A. Y. Kitaev. "Quantum computations: algorithms and error correction". In: *Russian Mathematical Surveys* 52.6 (1997), p. 1191.

[11] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010. DOI: 10.1017/CBO9780511976667. URL: http://mmrc.amss.cas.cn/tlb/201702/W020170224608149940643.pdf.

[12] C. Palazuelos. *Introduction to Quantum Information Theory*. 2013.

[13] V. B. Scholz and R. F. Werner. "Tsirelson's problem". In: *arXiv preprint:0812.4305* (2008).

[14] C.E. Shannon. "A mathematical theory of communication". In: *Bell System Technical Journal* 27 (1948), pp. 379–423.

[15] C.E. Shannon and W. Weaver. *The Mathematical Theory of Communication*. University of Illinois Press, 1949.

[16] P. W. Shor. "Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer". In: SIAM Journal on Computing.26(5) (1997), pp. 1484–1509.

[17] W. Slofstra. "The set of quantum correlations is not closed". In: *Forum of Mathematics, Pi* 7 (2019).

[18] W. Slofstra. "Tsirelson's problem and an embedding theorem for groups arising from non-local games". In: *J. Amer. Math. Soc.* 33 (2020), pp. 1–56.

[19] R. Solovay. "Lie Groups and Quantum Circuits". In: *Mathematics of Quantum Computation* 07 (2000).

[20] U. Vazirani. *Notes of the Course CS294-2 Quantum Computation*. University of Berkeley, 2004.

[21] John Watrous. *Advanced Topics in Quantum Information Theory*. Lecture Notes, 2021. URL: https://cs.uwaterloo.ca/~watrous/QIT-notes/.

[22] John Watrous. *The Theory of Quantum Information*. Cambridge University Press, 2018. DOI: 10.1017/9781316848142. URL: https://cs.uwaterloo.ca/~watrous/TQI/TQI.pdf.

[23] Mark M. Wilde. "Preface to the Second Edition". In: *Quantum Information Theory*. Cambridge University Press, pp. xi–xii. DOI: 10.1017/9781316809976.001. URL: https://arxiv.org/pdf/1106.1445.pdf.

[24] Michael M. Wolf. "Quantum Channels & Operations Guided Tour". In: (July 2012). URL: https://www-m5.ma.tum.de/foswiki/pub/M5/Allgemeines/MichaelWolf/QChannelLecture.pdf.

[25] R. de Wolf. *Quantum Computing: Lecture Notes*. 2011.

[26] W. K. Wootters and W. H. Zurek. "A single quantum cannot be cloned". In: *Nature* 299 (1982), pp. 802–803.