Quantum Information Theory

Lecture Notes by Angela Capel and Paul Gondolf,[1]

Institute of Mathematics
University of Tübingen

July 12, 2023

[1]In case you find mistakes or typos, let us know: angela.capel@uni-tuebingen.de, paul.gondolf@uni-tuebingen.de

Contents

1 Basic Notions in Quantum Information **5**

1.1 Scope of the course and Bibliography 5

1.1.1 Bibliography 5

1.2 What is Quantum Information? 6

1.2.1 Example: Shor's factoring algorithm 6

1.2.2 Emergence of Quantum Information Science 7

1.2.3 Quantum Information vs. Classical Information 7

1.2.4 Notation 7

1.3 Qubits and basic operations 8

1.3.1 Measurement 10

1.3.2 Projective Measurements 11

1.3.3 POVM Measurements 12

1.3.4 Unitary Evolution 12

1.3.5 Density operators 13

1.4 Postulates of quantum mechanics 14

1.4.1 Heisenberg picture 14

1.4.2 Schrödinger Picture 17

1.5 Quantum circuits 18

1.5.1 Classical circuits 19

1.5.2 Quantum gates 20

2 Entanglement and Quantum Communication **23**

2.1 Entanglement and Bell inequalities 24

2.1.1 Correlation in EPR Bell's result 25

2.1.2 Tsirelson's Theorem 27

2.1.3 Grothendieck's Theorem 31

2.1.4 Non local games 33

2.1.5 Non-local games as hyperplanes 39

2.2 Quantum Communication 40

2.2.1 No-Cloning theorem 40

2.2.2 Quantum teleportation 41

2.2.3 Superdense coding 43

3 Quantum algorithms **45**

3.1 Universality of quantum gates 46

3.2 Introduction to quantum algorithms 47

3.3 Deutsch-Jozsa's algorithm 48

3.3.1 2-bit functions 49

3.3.2 n-bit functions . . . 51
3.3.3 Bernstein-Vazirani problem . . . 53
3.4 Simon's algorithm . . . 54
3.5 Fourier transform . . . 55
3.6 Shor's factoring algorithm . . . 57
3.6.1 Reduction of factoring to order-finding . . . 58
3.6.2 The order-finding problem . . . 60
3.7 Grover's algorithm . . . 66
3.7.1 Algorithm . . . 66
3.7.2 Geometrical proof of the algorithm . . . 68
3.7.3 Amplitude amplification . . . 69

4 Hamiltonian simulation **71**
4.1 Lie-Suzuki-Trotter methods . . . 72
4.2 Linear combination of unitaries (LCU) methods . . . 74
4.3 Transformation via block-encoded matrices . . . 75

5 Quantum noise and open quantum systems **77**
5.1 Quantum channels . . . 77
5.1.1 Preliminaries . . . 77
5.1.2 Equivalent formulations of quantum channels . . . 80
5.1.3 Examples of quantum channels . . . 83
5.2 Open system representation . . . 85
5.2.1 Partial order of completely positive maps . . . 85
5.2.2 Instruments . . . 86
5.2.3 Open system representation . . . 87
5.3 Quantum Many Body Systems . . . 88
5.3.1 Simplified Approach to the Master Equation . . . 88
5.3.2 Detailed Master Equation . . . 90
5.4 Quantum hypothesis testing . . . 94
5.4.1 Binary hypothesis testing . . . 96
5.4.2 The pretty good measurement . . . 98
5.5 Bonus: Separability criteria . . . 100
5.5.1 Entanglement entropy . . . 100

6 Quantum error correction **103**

7 Quantum cryptography **105**

8 Quantum Shannon Theory **107**
8.1 Quantum Entropies . . . 107
8.1.1 Von Neumann Entropy . . . 107
8.1.2 Relative entropy . . . 110
8.1.3 Non-commmutative L_p norms . . . 112
8.1.4 Quantum divergences . . . 113
8.1.5 Minimal Divergence . . . 114
8.2 Quantum Hypothesis Testing . . . 116
8.2.1 Symmetric State Discrimination . . . 116
8.2.2 Asymmetric hypothesis testing . . . 117

8.2.3 Quantum Stein Lemma 117

Chapter 1

Basic Notions in Quantum Information

1.1 Scope of the course and Bibliography

This course is primarily designed for students enrolled at the University of Tübingen in various programs such as the M.Sc. in Mathematical Physics and the M.Sc. in Advanced Quantum Physics, although students from other degrees like the B.Sc. in Physics, B.Sc. in Mathematics, and related topics, are welcome to attend it as well. In general, it is also accessible to individuals with foundational knowledge in mathematical analysis, linear algebra, probability theory, and an eagerness to explore the captivating realm of quantum information theory.

1.1.1 Bibliography

A wealth of literature exists on the topics we will explore in this course, including Quantum Information Theory, Quantum Computing, Fault Tolerance, Error Correction, and more. To assist students in their studies, we have compiled a list of the primary texts frequently used in similar courses within the community. Additionally, we offer our lecture notes as a valuable reference tool that summarizes course content and provides supplementary material and references in certain areas. It is important to note that our lecture notes are not meant to substitute any of the other texts authored by leading quantum information experts. We strongly encourage students to consult these works to augment their understanding of the subject.

The short selection of manuscripts has been made accordingly to the contents intended for this course. The lectures, and therefore these notes, have been prepared to consult these texts, as well as some others, and they are properly referenced in this respect. The main references on which our notes are based are the following:

1. Lectures Notes
2. Nielsen-Chuang, "Quantum Computation and Quantum Information" [14]
3. de Wolf, "Quantum Computing: Lecture Notes", [28]
4. Preskill, "Quantum Computation. Lecture Notes", [16]

Some other books/notes used to write these notes will be referenced in the main text.

1.2 What is Quantum Information?

The scientific field of Quantum Information has a lot of different facets and encompasses the fields 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 known as *the entanglement frontier*.

Compared to the familiar classical world we experience every day, the quantum world presents behaviours that contradict our intuition. These additional properties, which will be discussed in detail throughout the course, provide quantum systems with more complex and richer behaviour, in some sense. As a consequence, we can expect to simulate a classical system using a quantum system, while it is generally believed that the reverse is not possible (although it still remains an unproven conjecture in full generality). However, quantum systems face the challenge of *decoherence*, a phenomenon not observed in classical systems, which rapidly destroys information and eventually causes quantum systems to behave like classical ones. Overcoming decoherence, or even theoretically estimating what big of an impact decoherence has, is a significant challenge that remains to be solved. 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 appear to be intractable for digital computers. For that, we will need to identify feasible quantum tasks, meaning tasks that are hard to simulate classically.

1.2.1 Example: Shor's factoring algorithm

To give an example of a theoretically expected improvement a quantum computer would provide compared to a classical one, we look at the task of factoring. This task is notoriously difficult for classical computers and fortunately so, as some encryption algorithms (e.g. RSA) are built upon this very fact. There exists, however, a quantum algorithm devised by Peter Shor in the early days of the quantum information field, that might promise a speedup. We will discuss such an algorithm in detail in a future chapter of the current notes (see [18] as well as the references [28, 22]).

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] This estimate stems from about 10 years ago.

1.2.2 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 limitations 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 this theory
- 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.

1.2.3 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:

- **(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.

1.2.4 Notation

In the next sections, 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 subsequently.

Beforehand, we need to introduce some notation. 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 1.2.1 We further introduce the following (*bra-ket*) notation originally used by Paul Dirac (1904 - 1984). He set

$$\begin{array}{ll} |\psi\rangle \in \mathbb{C}^n & \text{to be a vector,} \\ \langle\psi| \in (\mathbb{C}^n)^* & \text{to be a dual vector.} \end{array} \tag{1.2.1}$$

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

$$\langle\psi|\psi\rangle \in \mathbb{R}\,. \tag{1.2.2}$$

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

$$|\psi\rangle\langle\psi| : \mathbb{C}^n \to \mathbb{C}^n \tag{1.2.3}$$

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

$$|\psi\rangle\langle\psi|\xi\rangle = \langle\psi|\xi\rangle\,|\psi\rangle \in \mathbb{C}^n\,. \tag{1.2.4}$$

The content of the following subsections has been inspired by some basic texts of quantum information theory, such as the courses [15], [28] and [22], as well as the books [26], although one of the most fundamental texts in this field is [14]. We refer the reader to any of those texts for further knowledge on the topic.

1.3 Qubits and basic operations

Arguably, the most 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 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 $\mathbb{C}^2$ as $\{|0\rangle, |1\rangle\}$, i.e.

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{1.3.1}$$

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

$$|\psi\rangle = a_0\,|0\rangle + a_1\,|1\rangle \in \mathbb{C}^2 \tag{1.3.2}$$

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 a superposition of the state $|0\rangle$ 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 2^n operations at the same time.

Hence, theoretically, a quantum computer should be able to give an exponential improvement to the number 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

$$|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \qquad |-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) \tag{1.3.3}$$

To extract classical information from a quantum system one performs a measurement. Performing such a measurement on such an arbitrary state the system 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 qubits 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 with 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 *mixed quantum state* $|\varphi\rangle$ is a superposition of *pure quantum state* 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 at the same time (or perform N operations simultaneously, as mentioned above).

1.3.1 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 (we will consider it to be finite-dimensional, but will still introduce all concepts in a very general framework) 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 $\|\cdot\|_{\mathcal{H}}$ induced by a scalar product $\langle \cdot \mid \cdot \rangle$.

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

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

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

Definition 1.3.1 — Measurement. Let $\{M_n\}_n \subset \mathcal{B}(\mathcal{H})$ be a collection of operators verifying

$$\sum_n M_n^* M_n = \mathbb{1} \tag{1.3.4}$$

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 $|n\rangle$ occurs is given by

$$p(n) = \langle \varphi \mid M_n^* M_n \varphi\rangle \tag{1.3.5}$$

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 this classical state, 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 satisfy $M_i^* M_i = M_i^2 = M_i$, 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

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

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

1.3.2 Projective Measurements

In this subsubsection, we are going to introduce *projective measurements*, which play a 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 1.3.2 — Projective measurement. Consider a collection $\{M_n\}_n \subset \mathcal{B}(\mathcal{H})$ of measurements, as described in Definition 1.3.1. 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 on the right-hand side is, in fact, the spectral decomposition of M. Moreover, the possible outcomes of the measurement corresponding 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

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

1.3.3 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 1.3.3 — Positive operator valued measure. Consider a measurement $\{M_n\}_n \in \mathcal{B}(\mathcal{H})$ as in the Definition 1.3.1. 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 $M_n = \sqrt{E_n}$.

1.3.4 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

$$|\varphi\rangle = \alpha_1 |1\rangle + \ldots + \alpha_N |N\rangle \tag{1.3.6}$$

and want to transform this to the system

$$|\psi\rangle = \beta_1 |1\rangle + \ldots + \beta_N |N\rangle . \tag{1.3.7}$$

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

$$U \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_N \end{pmatrix} \tag{1.3.8}$$

and adding the condition that

$$\sum_{i=1}^{N} |\beta_i|^2 = 1, \tag{1.3.9}$$

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

$$UU^* = U^*U = \mathbb{1} \tag{1.3.10}$$

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 1.3.4 — Pauli matrices.**

$$\begin{aligned} \sigma_0 = \mathbb{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad & \sigma_x = X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \\ \sigma_y = Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad & \sigma_z = Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \end{aligned} \tag{1.3.11}$$

■

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.

1.3.5 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 a trace of an operator.

Definition 1.3.5 — Trace. Let $T : \mathcal{H} \to \mathcal{H}$ be a linear map represented by a matrix M in a certain basis. We then define

$$\mathrm{Tr}[T] = \mathrm{Tr}[M] = \sum_i M_{ii} \in \mathbb{C} \tag{1.3.12}$$

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

$$\langle\psi|\rho|\psi\rangle \geq 0 \qquad \forall \, |\psi\rangle \in \mathcal{H}, \tag{1.3.13}$$

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

1.4 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's picture. These two descriptions will help us to understand the topics presented above.

1.4.1 Heisenberg picture

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

Postulate I Given an isolated physical system, there is a complex Hilbert space $\mathcal{H}$ associated with 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 finite-dimensional Hilbert spaces for simplicity.

Postulate II 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

$$|\varphi_2\rangle = U_{t_1,t_2} |\varphi_1\rangle . \tag{1.4.1}$$

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 a system is described by

$$i\hbar \frac{d}{dt} |\varphi_t\rangle = H |\varphi_t\rangle . \tag{1.4.2}$$

where $\hbar$ is 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 III Given a physical system, with associated Hilbert space $\mathcal{H}$, quantum measurements over the system are described by a collection $\{M_n\}_n \subset \mathcal{B}(\mathcal{H})$ of measurements as defined in Definition 1.3.1.

More specifically, 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 measurement, the probability that $|n\rangle$ 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 IV 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. $\langle\cdot\mid\cdot\rangle_1$ and $\langle\cdot\mid\cdot\rangle_2$), it is a natural question whether one can introduce an inner product, and therefore a topology, on the tensor product that arises naturally from those of the factors. This can be done by defining the inner product as:

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

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 do 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*[9], 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 behaviour that seems to be at the root of many of the most surprising phenomena in quantum mechanics.

Definition 1.4.1 — 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

$$|\varphi_i\rangle \otimes \ldots \otimes |\varphi_n\rangle \tag{1.4.3}$$

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

1.4.2 Schrödinger Picture

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

Postulate I 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 pictures are dual, 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 the evolution of systems, we have the following formulation.

Postulate II 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

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

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

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

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

Postulate III 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 1.3.1. 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 observe 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

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

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

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

Hence, the total probability of this outcome is

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

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 IV 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.

1.5 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.

1.5.1 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 1.5.1 — Boolean circuit. A Boolean circuit is a finite directed acyclic graph composed of AND, OR and NOT gates (see Figure 1.1).

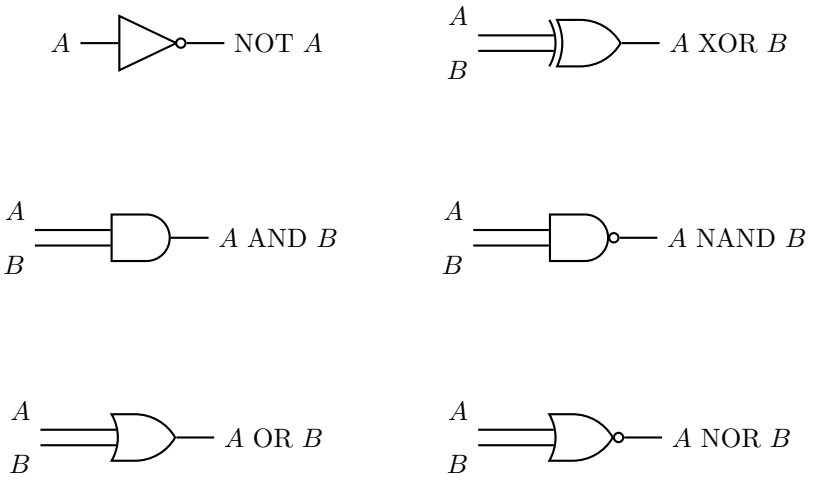

Fig. 1.1: Some classical gates for two bits. NOT, AND and OR can be used to construct the rest.

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 lies 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 them, 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* f 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 $x \notin L$.

1.5.2 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 with 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:

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{1.5.1}$$

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

$$Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{1.5.2}$$

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

$$R_\theta = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\theta} \end{pmatrix} \tag{1.5.3}$$

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

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \tag{1.5.4}$$

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

$$\text{CNOT} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{1.5.5}$$

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 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

$$C_U = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & U_{11} & U_{12} \\ 0 & 0 & U_{21} & U_{22} \end{pmatrix} \tag{1.5.6}$$

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 that these two gates are essentially irreversible.

We can see that in the following example: Given an output $A \oplus B$ of an 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 of classical circuits, one can define a quantum circuit in the following form.

Definition 1.5.2 — 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.

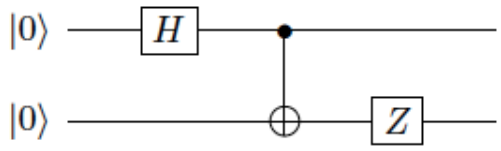

Fig. 1.2: Circuit used to turn $|00\rangle$ into $\frac{1}{\sqrt{2}}(|00\rangle - |11\rangle)$.

In this example, and in general, we denote the quantum CNOT by $\oplus$. If we study every step separately and take 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 $|\varphi_1\rangle = \dfrac{1}{\sqrt{2}}(|00\rangle + |10\rangle)$.
- **Step 2.** When we apply the CNOT gate, we get $|\varphi_2\rangle = \dfrac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$.
- **Step 3.** Finally, after the Z gate, we have $|\varphi_3\rangle = \dfrac{1}{\sqrt{2}}(|00\rangle - |11\rangle)$.

Chapter 2

Entanglement and Quantum Communication

In the past section, we introduced the formalism of quantum mechanics. This formulation, however, was widely questioned since its origins in physics as well as in the mathematical world. The only thing the scientific community could agree on was that the theory is a useful description of physical laws and allows us to predict them in a very precise way. However, some important scientists showed some scepticism 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 to Nature. On the other hand, the main idea behind the Hidden Variable Model is that such 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 the completeness of quantum mechanics and alternative theories started in 1935 with a paper by Einstein, Podolsky and Rosen, who proposed an 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 with the violations of Bell inequalities. Subsequently, we will provide a brief review of 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 [15] and [25].

2.1 Entanglement and Bell inequalities

In order to be able to provide a formal definition of entanglement, we need to provide a few preliminary results. The first one, namely the Schmidt decomposition, allows us to decompose pure states into basis products of a composite Hilbert space.

Theorem 2.1.1 — 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 $\{|j_B\rangle\}$ in $\mathcal{H}_B$ such that

$$|\psi\rangle = \sum_{i,j} \lambda_i |i_A\rangle \otimes |j_B\rangle , \qquad \text{with} \quad \lambda_{ij} \geq 0, \sum_{i,j} \lambda_{ij}^2 = 1. \tag{2.1.1}$$

The λ_{ij} are called *Schmidt coefficients* and the number of λ_{ij} in the decomposition, with multiplicity, is the *Schmidt rank* of the state.

We also need to introduce the partial trace of a density operator. For this, let $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$ be a composite system and let

$$\mathcal{S}(\mathcal{H}) := \{\rho \in \mathcal{B}(\mathcal{H}) \ : \ \rho = \rho^*, \ \rho \geq 0, \ \mathrm{Tr}[\rho] = 1\} \tag{2.1.2}$$

be the set of quantum states (or density matrices) on $\mathcal{H}_{AB}$. We define the partial trace as follows.

Definition 2.1.2 — 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

$$\mathrm{Tr}_B : \mathcal{S}(\mathcal{H}_{AB}) \to \mathcal{S}(\mathcal{H}_A), \qquad |a_1\rangle\langle a_2| \otimes |b_1\rangle\langle b_2| \mapsto \mathrm{Tr}[|b_1\rangle\langle b_2|]|a_1\rangle\langle a_2|. \tag{2.1.3}$$

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

$$\mathrm{Tr}_B[\rho] = \mathrm{Tr}_B[\rho_A \otimes \rho_B] = \mathrm{Tr}[\rho_B]\rho_A = \rho_A. \tag{2.1.4}$$

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

$$\mathrm{Tr}[M_A \rho_A] = \mathrm{Tr}[M_A \otimes \mathbb{1}_B \, \rho_{AB}] \tag{2.1.5}$$

in the case that $\rho_{AB} \in \mathcal{S}(\mathcal{H}_{AB})$. Finally, we introduce the notion of purification and find an interesting connection between mixed states and pure states as a consequence of the Schmidt decomposition.

Consequence 2.1.3 — 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

$$\rho_A = \mathrm{Tr}_R[|\psi_{AR}\rangle\langle\psi_{AR}|]. \tag{2.1.6}$$

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

Back to the discussion about entanglement, and as a reformulation of the notion of entanglement presented in the previous chapter, we say that a bipartite pure state $|\psi\rangle$ in $\mathcal{H}_A \otimes \mathcal{H}_B$ is *entangled* if its Schmidt rank is greater than one. Recall the definition of EPR from the previous chapter:

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

This state is considered a "maximally entangled state", which means that when we trace over qubit B to find the density operator ρ_A of qubit A, we obtain a multiple of the identity operator, i.e.

$$\rho_A = \mathrm{Tr}_B[|\phi^+\rangle\,\langle\phi^+|] = \frac{1}{2}\mathbb{1}_A\,.$$

and the same for ρ_B. If A and B represent two spins, this means that if we measure spin A along any axis, the result is completely random. Indeed, we find the spin up with probability $1/2$, and down with probability $1/2$. Therefore, if we perform any local measurement of A or B, we acquire no information about the preparation of the state, instead, we merely generate a random bit. With two qubits, we should be able to store two bits, but in the state $|\phi^+\rangle$ this information is *hidden*; at least, we cannot acquire it by measuring A or B.

2.1.1 Correlation in EPR Bell's result

Let us consider the following experiment: We have Alice, Bob and Charlie playing a game in which the latter sends 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.

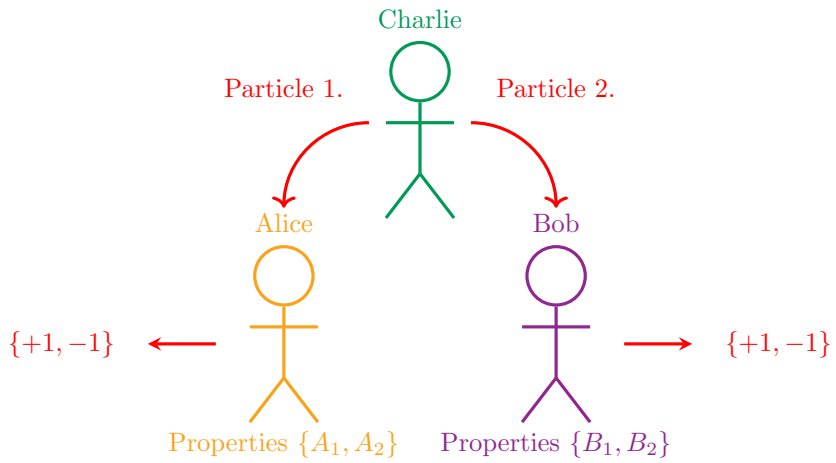

Fig. 2.1: EPR setup

We now want to better understand the 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

$$A_1B_1 + A_1B_2 + A_2B_1 - A_2B_2 = (A_1 + A_2)B_1 + (A_1 - A_2)B_2\,. \tag{2.1.7}$$

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

$$A_1B_1 + A_1B_2 + A_2B_1 - A_2B_2 = \pm 2 \tag{2.1.8}$$

- 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

$$P(a_1, a_2, b_1, b_2) = P(A_1 = a_1, A_2 = a_2, B_1 = b_1, B_2 = b_2) \tag{2.1.9}$$

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

$$\begin{aligned} &|\,\mathbb{E}[A_1B_1 + A_1B_2 + A_2B_1 - A_2B_2]| \\ &\qquad = \Big|\sum_{a_1,a_2,b_1,b_2} p(a_1, a_2, b_1, b_2)(a_1b_1 + a_1b_2 + a_2b_1 - a_2b_2)\Big| \leq 2 \end{aligned} \tag{2.1.10}$$

  Here is where the local hidden-variable assumption sneaks in since we have imagined that values in $\{\pm 1\}$ can be assigned simultaneously to all four observables, even though it is impossible to measure both of A_1 and A_2 or both of B_1 and B_2.

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

$$|\varphi\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}}\,, \tag{2.1.11}$$

  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, we have

$$\langle\varphi|A_1B_1|\varphi\rangle = \langle\varphi|A_2B_1|\varphi\rangle = \langle\varphi|A_1B_2|\varphi\rangle = \frac{1}{\sqrt{2}} \tag{2.1.12}$$

  and

$$\langle\varphi|A_2B_2|\varphi\rangle = -\frac{1}{\sqrt{2}} \tag{2.1.13}$$

  giving us

$$\langle\varphi|A_1B_1 + A_1B_2 + A_2B_1 - A_2B_2|\varphi\rangle = 2\sqrt{2} > 2. \tag{2.1.14}$$

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.

In this course, we will not discuss in further detail Bell inequalities and non-local games. However, for the avid reader, in the next subsections, we include some notes about that topic.

2.1.2 Tsirelson's Theorem

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

$$\gamma_{ij} = \mathbb{E}[A_i B_j] \quad \forall i,j = 1,\dots,N. \tag{2.1.15}$$

In the local hidden variable model, this evaluates to

$$\gamma_{ij} = \int_\Omega A_i(\omega) B_j(\omega)\, d\mathbb{P}(\omega) \tag{2.1.16}$$

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

$$\gamma_{ij} = \sum_k p(k) A_i(k) B_j(k). \tag{2.1.17}$$

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

$$\begin{array}{ll} \text{Output of Alice's measurement} & A_i\text{: POVM } \{\mathbb{E}_i, \mathbb{1} - \mathbb{E}_i\} \\ \text{Output of Bob's measurement} & B_j\text{: POVM } \{\mathbb{F}_j, \mathbb{1} - \mathbb{F}_j\} \end{array}$$

and we obtain

$$\begin{aligned} p(i',j') &= P(A_i = i', B_j = j') \\ &= \begin{cases} \mathrm{Tr}[(\mathbb{E}_i \otimes \mathbb{F}_j)\rho] & \text{if } (i',j') = (1,1) \\ \mathrm{Tr}[((\mathbb{1} - \mathbb{E}_i) \otimes \mathbb{F}_j)\rho] & \text{if } (i',j') = (-1,1) \\ \mathrm{Tr}[(\mathbb{E}_i \otimes (\mathbb{1} - \mathbb{F}_j))\rho] & \text{if } (i',j') = (1,-1) \\ \mathrm{Tr}[((\mathbb{1} - \mathbb{E}_i) \otimes (\mathbb{1} - \mathbb{F}_j))\rho] & \text{if } (i',j') = (-1,-1) \end{cases} \end{aligned} \tag{2.1.18}$$

Hence we get

$$\begin{aligned} \gamma_{ij} &= \mathbb{E}[A_i B_j] = [p(1,1) + p(-1,-1)] - [p(1,-1) + p(-1,1)] \\ &= \mathrm{Tr}[(\mathbb{E}_i \otimes \mathbb{F}_j + (\mathbb{1} - \mathbb{E}_i) \otimes (\mathbb{1} - \mathbb{F}_j) - \mathbb{E}_i \otimes (\mathbb{1} - \mathbb{F}_j) - (\mathbb{1} - \mathbb{E}_i) \otimes \mathbb{F}_j)\rho] \\ &= \mathrm{Tr}[((\mathbb{1} - 2\,\mathbb{E}_i) \otimes (\mathbb{1} - 2\,\mathbb{F}_j))\rho] \end{aligned} \tag{2.1.19}$$

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 2.1.4 — Quantum correlation matrix. We have that

$$\gamma = (\gamma_{i,j})_{i,j=1}^N \tag{2.1.20}$$

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

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

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

Proposition 2.1.5 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:

$$\gamma_{i,j} = \sum_{k=1}^{K} p(k) A_i(k) B_j(k) \tag{2.1.21}$$

Consider now the matrices given by

$$A_i = \begin{pmatrix} A_i(1) & 0 & \dots & & 0 \\ 0 & A_i(2) & 0 & \dots & 0 \\ \vdots & 0 & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ 0 & 0 & \dots & 0 & A_i(k) \end{pmatrix} \tag{2.1.22}$$

and

$$B_j = \begin{pmatrix} B_j(1) & 0 & \dots & & 0 \\ 0 & B_j(2) & 0 & \dots & 0 \\ \vdots & 0 & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ 0 & 0 & \dots & 0 & B_j(k) \end{pmatrix}. \tag{2.1.23}$$

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

$$\mathrm{Tr}[(A_i \otimes B_j)] = \sum_{k=1}^{K} p(k) A_i(k) B_j(k) = \gamma_{i,j}\,, \quad \forall i, j\,. \tag{2.1.24}$$

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 2.1.6 — Bell inequalities. In the above framework the Bell inequalities are given by

$$\sum_{i,j=1}^{N} M_{ij}\gamma_{ij} \leq C\,, \qquad .\forall \gamma = (\gamma_{i,j})_{i,j=1}^{N} \in \mathcal{C}_{cl}(N)\,. \tag{2.1.25}$$

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

■ **Example 2.1.7 — CHSH.** (CHSH) In the case of the CHSH inequality,

$$M = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \tag{2.1.26}$$

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

$$\sum_{i,j=1}^{N} M_{i,j}\hat{\gamma}_{i,j} = 2\sqrt{2}. \tag{2.1.27}$$

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

Definition 2.1.8 — Classical value. Given $M = (M_{ij})_{i,j=1}^N$ with real entries, we can associate

$$\Big| \sum_{i,j=1}^N M_{i,j}\gamma_{i,j} \Big| \leq \omega(M) \tag{2.1.28}$$

with

$$\begin{aligned} \omega(M) &:= \sup \Big\{ \Big| \sum_{i,j=1}^N M_{i,j}\gamma_{i,j} \Big| \ : \ \gamma = (\gamma_{i,j}) \in \mathcal{C}_{cl}(N) \Big\} \\ &= \sup \Big\{ \Big| \sum_{i,j=1}^N M_{i,j}x_i y_j \Big| \ : \ x_i = \pm 1, \ y_j = \pm 1 \ \forall i,j = 1,\dots,N \Big\}. \end{aligned} \tag{2.1.29}$$

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

Remark 2.1.9 By abuse of notation, we will sometimes call M above just a *Bell inequality*.

Definition 2.1.10 — Quantum value. Analogously we can define for $M = (M_{ij})_{i,j=1}^N$ with real entries, the *quantum value* as

$$\omega^*(M) := \sup \Big\{ \Big| \sum_{i,j=1}^N M_{i,j}\gamma_{i,j} \Big| \ : \ \gamma = (\gamma_{i,j})_{i,j=1}^N \in \mathcal{C}_q(N) \Big\} \tag{2.1.30}$$

Definition 2.1.11 — Largest violation. For $M = (M_{i,j})_{i,j=1}^N$ with real entries, we define the largest violation as

$$\mathrm{LV}(M) := \frac{\omega^*(M)}{\omega(M)} \tag{2.1.31}$$

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 2.1.12 — CHSH.** **(CHSH)** For the CHSH example we get

$$\omega(M) \leq 2, \ \omega^*(M) = 2\sqrt{2} \tag{2.1.32}$$

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 that we will use in its proof.

Definition 2.1.13 — CAR-algebra. Given $N \geq 2$, a set of operators $X_1, \dots, X_N$ is said to satisfy the *Classical Anticommutation Relations* if:

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

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

$$\begin{aligned} X_1 &= \underbrace{X \otimes \mathbb{1} \otimes \dots \otimes \mathbb{1}}_{m} & X_2 &= Y \otimes \mathbb{1} \otimes \dots \otimes \mathbb{1} \\ X_3 &= Z \otimes X \otimes \mathbb{1} \otimes \dots \otimes \mathbb{1} & X_4 &= Z \otimes Y \otimes \mathbb{1} \otimes \dots \otimes \mathbb{1} \\ &\vdots & &\vdots \\ X_{2m-1} &= Z \otimes Z \otimes \dots \otimes Z \otimes X & X_{2m} &= Z \otimes Z \otimes \dots \otimes Z \otimes Y \end{aligned} \tag{2.1.33}$$

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

Now we are in a 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 2.1.14 — 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 $x_1, \dots, x_N, y_1, \dots, y_N$ in a real Hilbert space such that

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

In particular,

$$\omega^*(M) = \sup_{1=\|x_i\|=\|y_j\|} \left\{ \Big| \sum_{i,j=1}^N M_{i,j} \langle x_i, y_j \rangle \Big| \right\} \tag{2.1.34}$$

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

$$\langle A, B \rangle = \mathrm{Re}(\mathrm{Tr}[AB\rho]) \tag{2.1.35}$$

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

$$\widetilde{\mathcal{H}} := \mathrm{span}\{x_i = A_i \otimes \mathbb{1} \ : \ i = 1, \dots, N\} \, , \tag{2.1.36}$$

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

$$y_j = \mathbb{P}(\mathbb{1} \otimes B_j) \qquad \forall j = 1, \dots, N. \tag{2.1.37}$$

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

$$\|x_i\| \leq 1, \ \|y_j\| \leq 1, \ \gamma_{ij} = \langle x_i, y_j \rangle = \mathrm{Tr}[A_i \otimes B_j \rho] \qquad \forall i, j = 1, \dots, N \, . \tag{2.1.38}$$

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:

$$\tilde{x}_i = x_i \oplus \sqrt{1 - \|x_i\|^2} \oplus 0, \quad \tilde{y}_j := y_j \oplus 0 \oplus \sqrt{1 - \|y\|_j^2} \tag{2.1.39}$$

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

$$T : \mathbb{R}^m \to \mathrm{span}\{X_1, \dots, X_m\}, \quad e_k \mapsto X_k \tag{2.1.40}$$

with the X_i constructed as in eq. (2.1.33). Then

$$\left\| T : \ell_2^m \to (\mathbb{C}^{2\times 2})^{\otimes m} \right\| \leq 1 \tag{2.1.41}$$

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

$$\frac{1}{2^{\frac{m}{2}}} \mathrm{Tr}[(Tx)(Ty)] = \langle x, y \rangle \qquad \forall x, y \in \mathbb{R}^M \, . \tag{2.1.42}$$

Consider further

$$|\psi\rangle = \frac{1}{2^{\frac{M}{4}}} \sum_{i,j=1}^{\frac{M}{2}} |ij\rangle \in (\mathbb{C}^2 \times \mathbb{C}^2)^{\otimes M} \tag{2.1.43}$$

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

$$\frac{1}{2^{\frac{M}{2}}} \operatorname{Tr}[AB] = \operatorname{Tr}[A \otimes B|\psi\rangle\langle\psi|] = \langle\psi|A \otimes B|\psi\rangle\,. \tag{2.1.44}$$

We define

$$A_i := T(x_i),\ B_j := T(y_j) \qquad \forall i,j = 1,\dots,N\,. \tag{2.1.45}$$

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

$$\langle\psi|A_i \otimes B_j|\psi\rangle = \frac{1}{2^{\frac{M}{2}}} \operatorname{Tr}[A_i B_j] = \langle x_i, y_j\rangle \quad \forall i,j = 1,\dots,N \tag{2.1.46}$$

■

2.1.3 Grothendieck's Theorem

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

Theorem 2.1.15 — Grothendieck's Theorem. There exists a universal constant K_G such that $\forall N \in \mathbb{N}$, $\forall (M_{i,j})_{i,j=1}^N \in \mathbb{R}^{N\times N}$,

$$\begin{aligned} &\sup\Big\{\Big|\sum_{i,j=1}^N M_{i,j}\,\langle x_i, y_j\rangle\Big| \ :\ \|x_i\| = \|y_j\| = 1\ \forall i,j = 1,\dots N\Big\} \\ &\qquad \leq K_G \sup\Big\{\Big|\sum_{i,j=1}^N M_{i,j} t_i s_j\Big| \ :\ t_i = \pm 1\,, s_j = \pm 1\ \forall i,j = 1,\dots N\Big\}\,, \end{aligned} \tag{2.1.47}$$

with K_G the (real) Grothendieck's constant:

$$1.67696 \leq K_G < \frac{\pi}{2\log(1+\sqrt{2})}. \tag{2.1.48}$$

We can rephrase that to

$$\omega^*(M) \leq K_G\,\omega(M) \tag{2.1.49}$$

or, equivalently,

$$\mathrm{LV}(M) \leq K_G. \tag{2.1.50}$$

■ **Example 2.1.16 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 the advantages of quantum mechanics with respect to 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 Bell's 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.

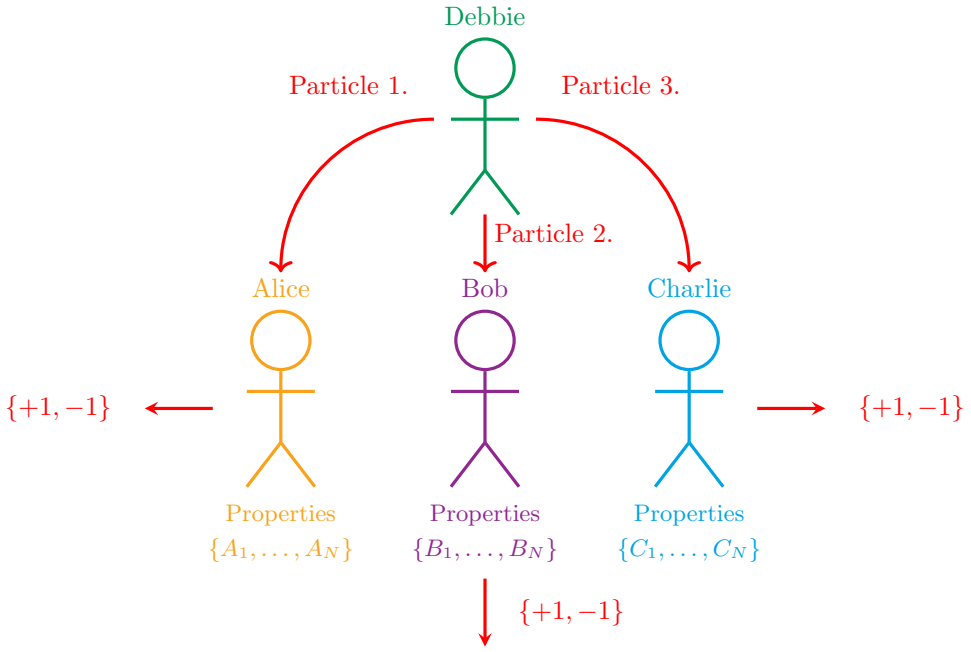

Fig. 2.2: Three-player game.

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

$$\gamma_{i,j,k} = \int_\Omega A_i(\omega)B_j(\omega)C_k(\omega)\, d\mathbb{P}(\omega)\,, \tag{2.1.51}$$

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

$$\max_{i,j,k=1,\ldots,N} \{\|A_i\|\,, \|B_j\|\,, \|C_k\|\} \leq 1 \tag{2.1.52}$$

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

$$\gamma_{i,j,k} = \mathrm{Tr}[A_i B_j C_k \rho] \qquad \forall i, j, k = 1, \ldots, N. \tag{2.1.53}$$

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 2.1.17 — 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

$$\mathrm{LV}(M) \geq D\,.$$

The most direct implication of this result is that, as soon as we consider three players, we get an unlimited amount of violations 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 classical mechanics.

2.1.4 Non local games

In this subsection, 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 [23, 15].

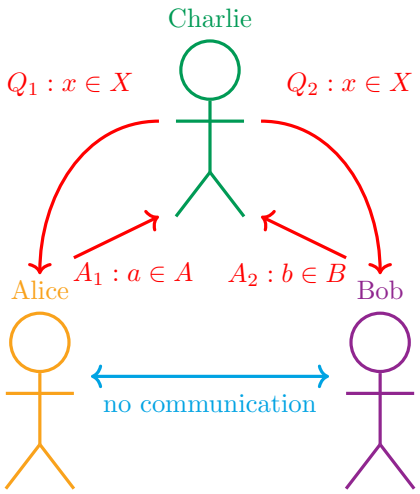

Fig. 2.3: Sketch of non-local games.

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 2.1.18 — 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

$$V(a,b|x,y) \equiv V(a,b,x,y) = \begin{cases} 1 & \text{if answering } (a,b) \text{ to } (x,y) \text{ WINS} \\ 0 & \text{if answering } (a,b) \text{ to } (x,y) \text{ LOSES} \end{cases} \tag{2.1.54}$$

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

■ **Example 2.1.19 — CHSH game. 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

$$V(a,b|x,y) = \begin{cases} 1 & a \oplus b = x \wedge y \\ 0 & a \oplus b \neq x \wedge y \end{cases}, \tag{2.1.55}$$

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

■ **Example 2.1.20 — FFL game, Fortnau, Feige and Lorasz. 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

$$V(a,b|x,y) = \begin{cases} 1 & a \vee x \neq b \vee y \\ 0 & a \vee x = b \vee y \end{cases}, \tag{2.1.56}$$

where $a \vee x$ denotes a OR x. ■

■ **Example 2.1.21 — Graph coloring game. 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

$$\Pi(x,y) = \begin{cases} \frac{1}{2n} & x = y \\ \frac{1}{4m} & x \neq y \,(x \text{ adjacent to } y) \\ 0 & \text{otherwise} \end{cases} \tag{2.1.57}$$

and the predicate

$$V(a,b|x,y) = \begin{cases} 1 & x = y, a = b \\ 1 & x \neq y, a \neq b \\ 0 & \text{otherwise} \end{cases}. \tag{2.1.58}$$

Note that this game is used to model the problem of colouring a graph. Namely, given a set of colours, we ask the question of what the minimum number of colours is, to colour each vertex in such a way that two adjacent vertices have different colours. It is not difficult to realize that winning the previous game with certainty 1 for a certain number of colours k implies that the graph of such a problem can be completely coloured according to this rule. We will go into further detail on this later in the text. ■

As mentioned above, the main purpose of Alice and Bob is to win 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 2.1.22 — 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 each question with a certain answer prior to the game. In detail, they consider:

$$f : X \to A,\ g : Y \to B \quad (x, y) \mapsto (f(x), g(y)) \,, \tag{2.1.59}$$

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 than the previous case, but the probability of 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 $[P_a^x, Q_b^y] = 0\ \forall a, b, x, y$.

2.1.4.1 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 with 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 2.1.23 — 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 with a non-local game, assuming that the strategy followed by Alice and Bob was deterministic or randomized (in any case, no quantum information was considered). The maximal success probability in this case is:

$$\omega_c(G) \equiv \omega(G) := \max_{\substack{f:X\to A,\\ g:Y\to B}} \sum_{(x,y)\in X\times Y} \Pi(x,y) V(f(x), g(y)|x, y) \tag{2.1.60}$$

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:

$$\omega_q(G) \equiv \omega^*(G) = \sup_{\substack{\psi\in\mathcal{H}_A\otimes\mathcal{H}_B\\ \{P_a^x\}_{a\in A}\\ \{Q_b^x\}_{b\in B}}} \sum_{(x,y)\in X\times Y} \Pi(x,y) \sum_{(a,b)\in A\times B} \langle\psi|P_a^x \otimes Q_b^y|\psi\rangle\, V(a, b|x, y) \tag{2.1.61}$$

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:

$$\omega_{co}(G) := \sup_{\substack{\psi\in\mathcal{H}_A\otimes\mathcal{H}_B\\ \text{commuting}\\ \text{strategies}}} \sum_{(x,y)\in X\times Y} \Pi(x,y) \sum_{(a,b)\in A\times B} \langle\psi|Q_b^y P_a^x|\psi\rangle\, V(a, b|x, y) \tag{2.1.62}$$

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 2.1.24 CHSH game.** We can compute the classical and entangled values for the CHSH game, respectively. We leave as an exercise to show 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 2.1.25 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:

$$\omega_c(G_{\mathrm{FFL}}) = \frac{2}{3} = \omega_q(G_{\mathrm{FFL}})\,.$$

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 2.1.26 Graph colouring 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 the values of games.

1. Derived from the increasing order in the restriction 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 $\omega_c(G) < \omega_q(G)$.

2.1.4.2 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 2.1.27 — 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 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
$$\mathcal{C}_{cl} \equiv \mathcal{C}_{cl}^{m,n,k,l} . \tag{2.1.63}$$

- The set of **quantum correlations** is given by
$$\mathcal{C}_q^{k,l,m,n} := \left\{ p(a,b|x,y) = \langle\psi|P_a^x \otimes Q_b^y|\psi\rangle \ : \begin{array}{c} |\psi\rangle \in \mathcal{H} \text{ normalised }, \{P_a^x\}_{a\in A}, \{Q_b^y\}_{b\in B} \\ \text{an entangled strategy,} \\ \text{on } \mathcal{H}_A \text{ and } \mathcal{H}_B \text{ finite dimensional} \end{array} \right\} \tag{2.1.64}$$

- 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
$$\mathcal{C}_{qs}^{k,l,m,n} := \left\{ p(a,b|x,y) = \langle\psi|P_a^x \otimes Q_b^y|\psi\rangle \ : \begin{array}{c} |\psi\rangle \in \mathcal{H} \text{ normalised }, \{P_a^x\}_{a\in A}, \{Q_b^y\}_{b\in B} \\ \text{an entangled strategy,} \\ \text{on } \mathcal{H}_A \text{ and } \mathcal{H}_B \text{ possibly } \infty\text{-dimensional} \end{array} \right\} \tag{2.1.65}$$

- The set of **quantum correlations well-approximated by tensor products** (finite-dimensional) is by definition
$$\mathcal{C}_{qa}^{k,l,m,n} := \overline{\mathcal{C}}_q^{k,l,m,n} \tag{2.1.66}$$

- The set of **quantum commuting correlations** is denoted by
$$\mathcal{C}_{qc}^{k,l,m,n} \equiv \mathcal{C}_{co}^{k,l,m,n} = \left\{ p(a,b|x,y) = \langle\psi|Q_b^y P_a^x|\psi\rangle \ : \begin{array}{c} |\psi\rangle \in \mathcal{H} \text{ normalised }, \{P_a^x\}_{a\in A}, \{Q_b^y\}_{b\in B} \\ \text{a commuting strategy} \\ \text{on } \mathcal{H} \text{ (possibly infinite-dimensional)} \end{array} \right\} \tag{2.1.67}$$

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

$$\boxed{\mathcal{C}_{cl} \subsetneq \mathcal{C}_q \subsetneq \mathcal{C}_{qs} \subsetneq \mathcal{C}_{qa} \subsetneq \mathcal{C}_{co}} \tag{2.1.68}$$

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 allow for infinite-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 of 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 of 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 [17] 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 [20] 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 [19] 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 [5] 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 [11], 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.

2.1.5 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 2.1.28 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

  $$P_i := p_i(a,b|x,y) = \langle \psi_i | P_a^{(i)x} \otimes Q_b^{(i)y} | \psi_i \rangle \qquad i = 1,2 \tag{2.1.69}$$

  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:

  $$\begin{aligned}(\mathcal{H}_A^{(1)} \oplus \mathcal{H}_A^{(2)}) \otimes (\mathcal{H}_B^{(1)} \oplus \mathcal{H}_B^{(2)}) \\ \cong (\mathcal{H}_A^{(1)} \otimes \mathcal{H}_B^{(1)}) \oplus (\mathcal{H}_A^{(1)} \otimes \mathcal{H}_B^{(2)}) \oplus (\mathcal{H}_A^{(2)} \otimes \mathcal{H}_B^{(1)}) \oplus (\mathcal{H}_A^{(2)} \otimes \mathcal{H}_B^{(2)})\,.\end{aligned} \tag{2.1.70}$$

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

  $$\begin{aligned}P_a^x &= P_a^{(1)x} \oplus P_a^{(2)x}\,, \\ Q_b^y &= Q_b^{(1)y} \oplus Q_b^{(2)y}\,.\end{aligned} \tag{2.1.71}$$

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

  $$|\psi\rangle = \sqrt{\lambda}\,|\psi_1\rangle \oplus 0 \oplus 0 \oplus \sqrt{1-\lambda}\,|\psi_2\rangle\,. \tag{2.1.72}$$

  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 in mind, we can compare correlations to separating hyperplanes.

Definition 2.1.29 — 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:

$$\langle H, P \rangle = \sum_{a,b,x,y} H_{a,b,x,y} P(a,b|x,y). \tag{2.1.73}$$

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

$$\max_{\mathcal{C}}(H) = \sup_{P \in \mathcal{C}} |\langle H, P \rangle| \tag{2.1.74}$$

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

2.2 Quantum Communication

2.2.1 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 2.4. It is a trivial application of the classical CNOT gate.

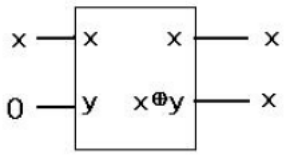

Fig. 2.4: Classical way to clone a bit.

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 [29], 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 labelled by A and B. The first one is the *control* qubit, denoted by $|psi\rangle$, and the second one is 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 fulfils the properties: There exists a unitary U such that

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

$$|\psi\rangle \otimes |\phi\rangle \xrightarrow{U} U(|\psi\rangle \otimes |\phi\rangle) = |\psi\rangle \otimes |\psi\rangle \tag{2.2.1}$$

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

$$|\varphi\rangle \otimes |\phi\rangle \xrightarrow{U} U(|\varphi\rangle \otimes |\phi\rangle) = |\varphi\rangle \otimes |\varphi\rangle \tag{2.2.2}$$

From this, it immediately follows that

$$\langle\varphi|\psi\rangle = \lambda = \lambda^2 = \langle\varphi|\psi\rangle^2 . \tag{2.2.3}$$

This implies that

$$\lambda = \begin{cases} 0 & \text{States are orthogonal} \\ 1 & \text{States are the same} \end{cases} . \tag{2.2.4}$$

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.

2.2.2 Quantum teleportation

We present now another example of how the gates of the previous chapter can be used to get results of relevance. In this particular case, we will explain *quantum teleportation* [4] 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 2.5. 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:

$$\mathrm{EPR} = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) =: |\varphi\rangle\,. \tag{2.2.5}$$

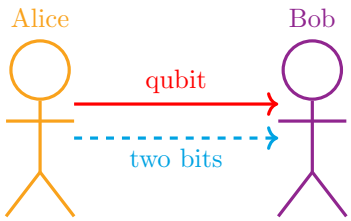

Fig. 2.5: Schematic representation of the protocol of Quantum teleportation

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 than one qubit. The fact that a qubit can be sent by means of this procedure is usually expressed by writing:

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

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

$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle \tag{2.2.6}$$

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 into the following steps:

1. The combined initial system is

$$\begin{aligned} |\varphi_0\rangle_{AA'B'} &= |\psi\rangle_A \otimes |\varphi\rangle_{A'B'} = (\alpha\,|0\rangle_A + \beta\,|1\rangle_A) \otimes \left(\frac{1}{\sqrt{2}}(|00\rangle_{A'B'} + |11\rangle_{A'B'})\right) \\ &= \frac{1}{\sqrt{2}}\big(\alpha(|000\rangle_{AA'B'} + |011\rangle_{AA'B'}) + \beta(|100\rangle_{AA'B'} + |111\rangle_{AA'B'})\big), \end{aligned} \tag{2.2.7}$$

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

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

$$\begin{aligned} |\varphi_1\rangle &= \mathrm{CNOT}_{AA'}\,|\varphi_0\rangle \\ &= \frac{1}{\sqrt{2}}\big[\alpha(|000\rangle_{AA'B'} + |011\rangle_{AA'B'}) + \beta(|110\rangle_{AA'B'} + |101\rangle_{AA'B'}\big] \end{aligned} \tag{2.2.8}$$

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

$$\begin{aligned}|\varphi_2\rangle &= H_A \otimes \mathbb{1}_{A'B'} |\varphi_1\rangle \\ &= \frac{1}{\sqrt{2}} \left[\alpha \frac{|0\rangle_A + |1\rangle_A}{\sqrt{2}} \otimes (|00\rangle_{A'B'} + |11\rangle_{A'B'}) + \beta \frac{|0\rangle_A - |1\rangle_A}{\sqrt{2}} (|00\rangle_{A'B'} + |11\rangle_{A'B'} \right] \\ &= \frac{1}{2} \left[|00\rangle_{AA'} \otimes (\alpha |0\rangle_{B'} + \beta |1\rangle_{B'}) + |10\rangle_{AA'} \otimes (\alpha |0\rangle_{B'} - \beta |1\rangle_{B'}) \right. \\ &\qquad \left. + |01\rangle_{AA'} \otimes (\alpha |1\rangle_{B'} + \beta |0\rangle_{B'}) + |11\rangle_{AA'} \otimes (\alpha |1\rangle_{B'} - \beta |0\rangle_{B'} \right]\end{aligned} \tag{2.2.9}$$

with

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \tag{2.2.10}$$

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 of 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 post measurement 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

This whole procedure can be represented as the following quantum circuit:

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.

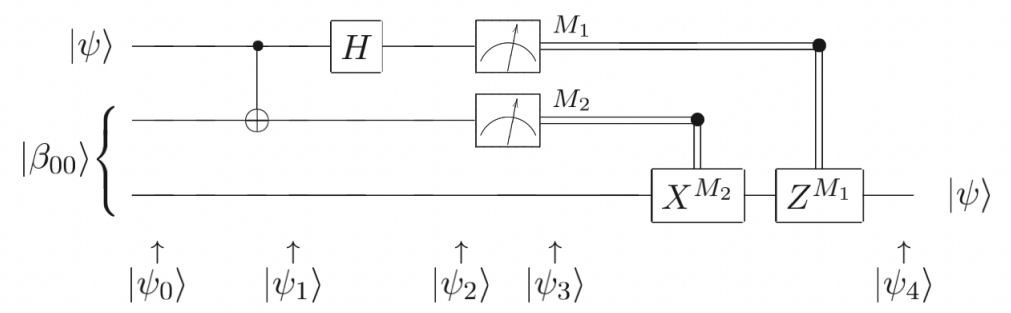

Fig. 2.6: Quantum circuit for the teleportation of a qubit, extracted from [14].

2.2.3 Superdense coding

Now we present a third example of the 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 transmitting 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 2.7:

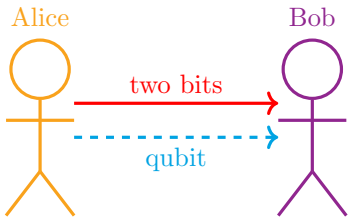

Fig. 2.7: Superdense coding.

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

$$|\varphi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle). \tag{2.2.11}$$

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 measuring the whole state in the Bell basis, Bob can then determine which of the four possible bit strings Alice sent him.

Chapter 3

Quantum algorithms

As we already discussed in the first chapter of these notes, in the future, quantum computers will have the potential to solve problems exponentially faster than classical computers. However, which problems can be solved exponentially faster? And which problems would do just as well as being solved by a classical computer? These are some of the main questions for quantum algorithm researchers.

To approach an answer to this question, it makes sense to develop the connection between quantum algorithms and computer hardware. A *computer* is a programmable machine which is based on some fundamental concepts of Physics, whereas an *algorithm* is a sequence of steps that leads to the solution of a problem. Their connection comes then into the stage when one notices that for implementing an algorithm in a computer it is necessary to check that this machine can actually run it.

Quantum computers can run algorithms that we cannot run on classical computers. Those algorithms make use of quantum effects, such as non-locality, entanglement, superposition, etc. When one designs a quantum algorithm, one of the main aims is to exploit these concepts to make the algorithm get the solution to the problem faster. In general, we can say that a quantum algorithm provides opportunities to solve problems that are difficult in the classical context, i.e., problems that we cannot solve efficiently with the known classical algorithms.

A quantum algorithm might work as follows: It starts with a classical input and turns it into a quantum state by obtaining the superposition of an exponential number of classical states. Then, it transforms the quantum state that encodes the problem into a quantum state that encodes the solution. From that quantum state, one can measure and obtain the solution.

Most of these quantum algorithms are designed for future quantum computers, which will have many qubits. Nowadays, only some small prototypes of these devices already exist, counting on, a bit more than 100 qubits. It is then an interesting research area to study what can one do with these devices before we actually have the ones with a great number of qubits.

The key to the security of public key cryptosystems is that it should be difficult to invert the encryption stage if only the public key is available. For example, inverting the encryption stage of RSA is in fact a problem closely related to factoring numbers. Hence, much of the presumed security of RSA comes from the belief that factoring is a problem hard to solve on a classical computer. However, as we will see in this chapter, Shor's fast algorithm for factoring on a quantum computer could be used to break RSA. In an analogous way, there are other public key cryptosystems that can be broken using a fast algorithm for solving the discrete logarithm problem, like Shor's quantum algorithm for discrete logarithms. This practical application of quantum computers to the break codes has excited much of the interest in quantum computation and quantum information.

The two main quantum algorithms that we will study in this chapter are *Shor's quantum factoring algorithm* and *Grover's algorithm*, which will be explained, respectively, in Section 3.6 and Section 3.7. First, we describe some of the earlier quantum algorithms that preceded them.

All quantum algorithms work with queries in some form or another, which we will explain below. The query complexity model differs from the standard model described in the first chapter, as the input, in this case, is given as a *black-box*. However, before that, we need to introduce some basic notions about *quantum circuits* [6], [30], since they constitute one of the two models that are more commonly used to explain how a quantum computer can apply computational steps to its qubits. The other model, the *quantum Turing machine*, will be further discussed in a future chapter of these notes.

Many of the topics discussed in this chapter are based in some of the following books or notes: [15], [28]. Let us begin the chapter by recalling some concepts discussed in the first chapter for quantum circuits, as well as the universality of certain sets of quantum gates.

3.1 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:

$$\text{Tofolli} = \begin{cases} \text{AND} & \text{Fix the third input to 0,} \\ \text{NOT} & \text{Fix the first and second input to 1.} \end{cases} \tag{3.1.1}$$

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 $R_{\pi/4}$.** This gate set is approximately universal, which means that any other unitary can be arbitrarily well approximated using its circuits, a consequence of the *Solovay-Kitaev theorem* [21, 12]. More specifically, this theorem states that we can approximate any gate on 1 or 2 qubits up to a certain error ε using a number of gates that only scales as $\text{polylog}(1/\varepsilon)$. This can be interpreted as the fact that simulating arbitrary gates up to exponentially small errors only costs a polynomial overhead.

  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.

3.2 Introduction to quantum algorithms

As we have mentioned in the introduction of this chapter, the two main quantum algorithms that we will study here are *Shor's quantum factoring algorithm* and *Grover's algorithm*, which will be explained, respectively, in Section 3.6 and Section 3.7. However, beforehand, we will introduce and describe some of the earlier quantum algorithms that appeared previously.

All quantum algorithms work with queries in some form or another. The query complexity model differs from the standard model described in the previous chapter, as the input, in this case, is given as a *black-box* (or an *oracle*), which implies that the exponential separation between quantum and classical that we describe in the following algorithms does not lead to an exponential separation between quantum and classical in the standard model.

The phenomenon that appears in the query setting can be further explained as follows. Consider an N-bit input $x \in \{0,1\}^N$, for $N = 2^n$, where we can write $x = (x_0, \ldots, x_{N-1})$. The restriction for N is just made to simplify the notations, as now we can address a bit x_i using an n-bit index i. The input can be seen as an N-bit memory which we can access at any point of our choice (a Random Access Memory), and this access is done via a black box, which is designed to output the bit x_i on input i.

As a quantum operation, it must always be a unitary mapping on $n+1$ qubits, given by

$$O_x : |i, 0\rangle \to |i, x_i\rangle,$$

and, in general,

$$O_x : |i, b\rangle \to |i, b \oplus x_i\rangle,$$

where we recall that $i \in \{0,1\}^n$ and $b \in \{0,1\}$. Also, by $\oplus$ we denote *exclusive-or*, which is nothing but addition modulo 2. The first n qubits of the state are called the *address bits* (in previous sections have been called *control bits*), while the qubit in position $(n+1)$ is called the *target bit*.

In the setting of matrix representation, O_x is a permutation matrix, and, as we have mentioned above, it is unitary. Here we can notice a difference between a quantum computer and a classical one since the first one can apply O_x on a superposition of several i, something that the latter cannot do (we will elaborate on this in some paragraphs below).

One application of this black box is called a *query*, and counting the required number of queries to compute a certain function of x is something essential in quantum complexity. Assume now that we make a query of the type mentioned above. Thus, we can also make a query of the form

$$|i\rangle \mapsto (-1)^{x_i} |i\rangle$$

by setting the target bit to the state

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

Hence, we obtain:

$$O_x (|i\rangle |-\rangle) = |i\rangle \frac{1}{\sqrt{2}} (|x_i\rangle - |1 - x_i\rangle) = (-1)^{x_i} |i\rangle |-\rangle.$$

The aim of term $(-1)^{x_i}$ is to put the output variable in the phase of the state:

1. If x_i is 1, then we get a -1 in the phase of basis state.
2. If x_i is 0, then the phase of the basis state remains invariant.

This "phase oracle" is sometimes more convenient than the standard type of query. We sometimes denote the corresponding n-qubit unitary transformation by $O_{x,\pm}$.

Before introducing the first proper quantum algorithm in the next section, we need to present a quantum-mechanical effect that appears exclusively in quantum mechanics and that we can (and will) use for building quantum algorithms, and which is called *quantum parallelism*. Assume that we have a Boolean function $f : \{0,1\}^n \to \{0,1\}^m$ and a classical algorithm that computes it. It is clear then that we can construct a quantum circuit of Toffoli gates (because of the universality mentioned above) that maps:

$$|x\rangle |0\rangle \mapsto |x\rangle |f(x)\rangle \qquad \text{for every } x \in \{0,1\}^n.$$

Let us denote this quantum circuit by U. Then, if we apply U to a superposition of all possible inputs, we have:

$$U\left(\frac{1}{\sqrt{2^n}} \sum_{x\in\{0,1\}^n} |x\rangle |0\rangle\right) = \frac{1}{\sqrt{2^n}} \sum_{x\in\{0,1\}^n} |x\rangle |f(x)\rangle.$$

Even though we have applied U just once, the final superposition state that we get contains all possible inputs. However, to observe the final superposition just gives one random $|x\rangle |f(x)\rangle$, so, by itself, this effect is not very useful, since it does not provide any improvement to classical randomization. It has to be combined with some other effects of interference or entanglement to improve the classical case.

In the following sections, we will use the concepts that we have just introduced to present several quantum algorithms. In general, one can classify quantum algorithms into three different classes which provide an advantage over known classical algorithms. The first one is the class of algorithms *based upon quantum versions of the Fourier transform*, a tool which is also widely used in classical algorithms. The *Deutsch–Jozsa* algorithm is an example of this type of algorithm, although the use of such transformation in the algorithm is a bit vague. The main exponent of this class of algorithms is *Shor's algorithm* for factoring and discrete logarithms.

The second class of algorithms is *quantum search algorithms*. The main algorithm from this class that we will present is *Grover's algorithm*. Finally, the third class of algorithms is *quantum simulation*, in which a quantum computer is used to simulate a quantum system.

In this chapter, we just focus on a brief description of the first two classes of algorithms mentioned above. This description will be presented in the following sections. The algorithms devoted to quantum simulation will be explored in further detail later in the course.

3.3 Deutsch-Jozsa's algorithm

In the early 1980s, Richard Feynman, in the US, and Yuri Manin, in the Soviet Union, suggested using quantum computers to simulate quantum mechanics. Some years later, in 1985, David Deutsch tried to actually formulate this. For that, he defined a quantum Turing machine and explained how to solve what later was going to turn out to be the 2-bit Deutsch-Jozsa problem. As we will see below, solving the 2-bit Deutsch-Jozsa problem did not really impress anyone. However, in 1992, David Deutsch and Richard Jozsa generalized his construction to solve the n-bit Deutsch-Jozsa problem [8], which constituted a much more interesting and relevant result.

Problem 3.3.1 For $N = 2^n$, we are given an element $x \in \{0,1\}^N$ such that one of the following holds:

1. Every x_i has the same value.

2. $N/2$ of the x_i are 0 and $N/2$ are 1.

In the first case, we call x *constant*, whereas in the second case, we call it *balanced*. The **goal** is to find out whether x is constant or balanced.

This is what computer scientists call a *promise problem* since we get the promise that we are always going to receive an n-bit in one of the previous two situations.

We shall now put this problem into words. Suppose that Alice, in Amsterdam, selects a number x from 0 to $2^n - 1$, and mails it in a letter to Bob, in Boston. Bob then calculates some function $f(x)$ and replies with the result he gets, which is either 0 or 1. The function that Bob uses is of one of two kinds: either $f(x)$ is constant for all values of x, or else $f(x)$ is balanced. Alice's goal is to determine with certainty whether Bob has chosen a constant or a balanced function, exchanging with him as little information as possible. How fast can she succeed?

More formally, we define a function $f : \{0,1\}^N \to \{0,1\}$, i.e., it takes N-bits as input and produces either a 0 or a 1 as output for each such value. We are promised that the function is either constant (0 on all outputs or 1 on all outputs) or balanced (returns 1 for half of the input domain and 0 for the other half).

The most natural question one can think of is: How is f given? In this sense, f is given as an oracle. Hence, we are given a black box, we input an N-bit x and it outputs $f(x)$, as we show in Figure 3.1.

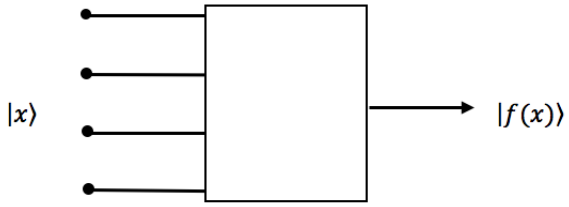

Fig. 3.1: f is given as an oracle.

It is clear that this black box is not really a quantum circuit, since it has an N-bit as input and one bit as output, and we have mentioned in previous sections that a quantum oracle has to be reversible, which means that the number of inputs should be the number of outputs. Hence, it can only be a classical oracle.

A way to get an actual quantum circuit is by recalling that any classical function can be made reversible as long as you keep the input around. In this situation, this reads as we can see in Figure 3.2 below. Moreover, one can notice that it is its own inverse.

3.3.1 2-bit functions

Let us show how we can solve the Deutsch-Jozsa problem for the 2-bit qubit case. For that, consider the quantum oracle of Figure 3.3, so that it changes the phase of the input if $f(x) = 1$, and leaves it untouched if $f(x) = 0$.

Since we are studying the 2-bit case, it is clear that $|x\rangle \in \{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$. Let us consider the input

$$\frac{1}{2}(|00\rangle + |01\rangle + |10\rangle + |11\rangle).$$

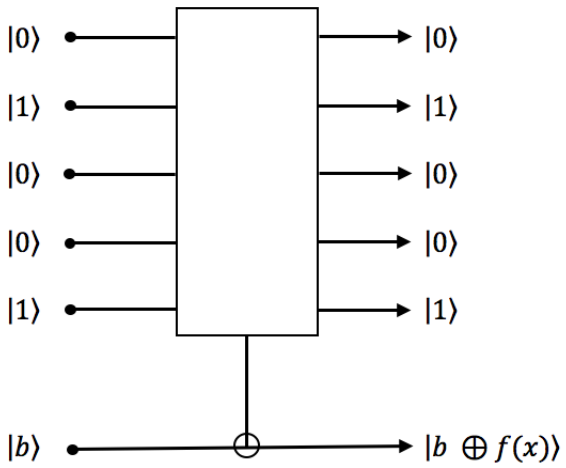

Fig. 3.2: f is given as an oracle.

Fig. 3.3: Quantum oracle.

For that input, we can have, if f is constant and equal to 0, the output

$$\frac{1}{2}\left(|00\rangle + |01\rangle + |10\rangle + |11\rangle\right)$$

and for f equal to 1

$$-\frac{1}{2}\left(|00\rangle + |01\rangle + |10\rangle + |11\rangle\right).$$

On the opposite, if f is balanced, the output would be

$$\frac{1}{2}\left(|00\rangle - |01\rangle + |10\rangle - |11\rangle\right),$$

or something like that.

If we recall that solving the Deutsch-Jozsa problem consists of distinguishing between these two possibilities, we need to develop a tool to do that. In this situation, we can just reduce to making a measurement that obtains a *yes* if the function is constant and a *no* if it is balanced. We can construct this measurement by applying two Hadamard gates and testing for 0, as in Figure 3.4.

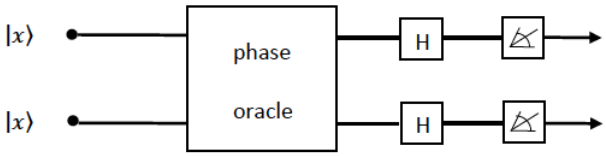

Fig. 3.4: 2-bit Deutsch-Jozsa model.

We will present in more detail this phase oracle in the next subsection. If the output equals $|00\rangle$, then f is constant. On the contrary, if the output is different from $|00\rangle$, f is balanced. However, it is important to remark that there are many cases in which f is neither constant nor balanced. For example, consider the case in which, for input $|{+}{+}\rangle$, we have the output

$$\frac{1}{2}\left(|00\rangle + |01\rangle + |10\rangle - |11\rangle\right).$$

Then, the probability of seeing $|{+}{+}\rangle$ is

$$\text{Prob}(|{+}{+}\rangle) = \frac{1}{2},$$

which is neither 0 nor 1.

3.3.2 n-bit functions

We can generalize the solution of the previous subsection to the n-bit case. Although the solution is almost exactly the same, it took seven years for Deutsch to generalize it. It is due to the fact that in Deutsch's original paper [7], it looked nothing like this circuit, and, thus, it was hard to generalize.

The generalization of the algorithm presented in the previous subsection is the following:

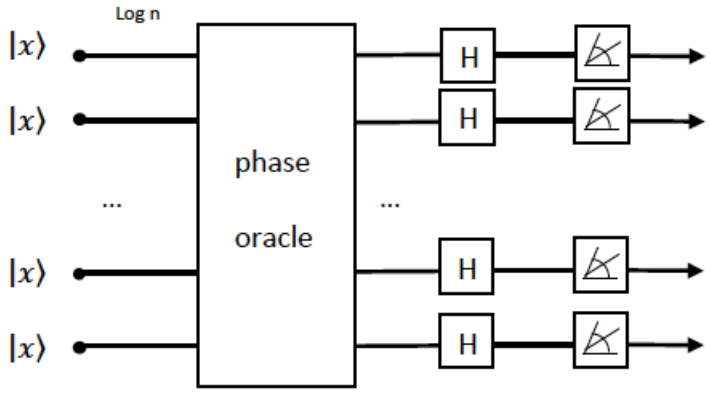

Fig. 3.5: n-bit Deutsch-Jozsa model.

Now, analogously to the previous case, if output equals $|00\ldots0\rangle$, then f is constant. On the contrary, if the output is different from $|00\ldots0\rangle$, f is balanced.

Let us explain this algorithm in some detail. We start with the zero states for n-qubits, $|00\ldots0\rangle = |0^n\rangle$, we apply a Hadamard gate to each qubit (in the previous picture encoded in the phase oracle), apply a query to everything (the phase oracle itself), then apply another Hadamard gate to each qubit, and, finally, measure the final state. Hence, this algorithm could be summarized as something like:

$$H^{\otimes n} O_{x,\pm} H^{\otimes n}.$$

Now we go step by step, seeing what we get in each one of them. Initially, we have the state $|0^n\rangle$. After the first Hadarmard gate is applied to each qubit, we get,

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

which is a uniform superposition of all possible i.

When we apply the query, the last term becomes:

$$\frac{1}{\sqrt{2^n}} \sum_{i \in \{0,1\}^n} (-1)^{x_i} |i\rangle,$$

and after the second Hadarmard gates, we have the final superposition

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

where we are denoting $i \cdot j = \sum_{k=1}^{n} i_k j_k$.

Finally, notice that $i \cdot 0^n = 0$ for all $i \in \{0,1\}^n$. Thus, we can rewrite the first part of the latter term as:

$$\frac{1}{2^n} \sum_{i \in \{0,1\}^n} (-1)^{x_i} = \begin{cases} 1 & \text{if } x_i = 0 \text{ for all } i, \\ -1 & \text{if } x_i = 1 \text{ for all } i, \\ 0 & \text{if } x \text{ is balanced .} \end{cases}$$

This is in particular the amplitude of the $|0^n\rangle$ state in the last combination. Then, the final term will yield $|0^n\rangle$ if x is constant and some other state if x is balanced. Therefore, the Deutsch-Jozsa problem can be solved using only 1 query and in $O(n)$ other operations, but this solution is not the original one of Deutsch and Jozsa, in which they used 2 queries instead of one.

Another equivalent way to present the Deutsch-Jozsa algorithm, extracted from [14] is Figure 3.6.

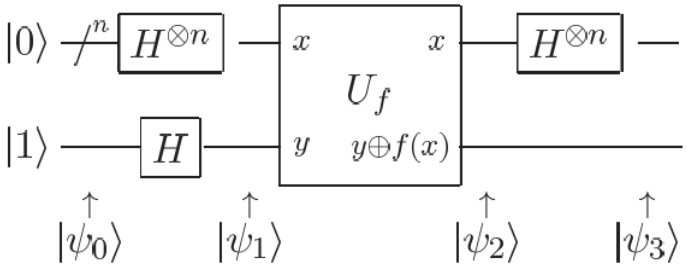

Fig. 3.6: Another scheme for the n-bit Deutsch-Jozsa model.

In this case, the initial state is $|\psi_0\rangle = |0^n\rangle |1\rangle$. With this notation, the black box U_f performs the transformation $|x\rangle |y\rangle \to |x\rangle |y \oplus f(x)\rangle$ for $x \in \{0,1\}^n$ and $f(x) \in \{0,1\}$.

After the application of the first Hadarmard gates, one gets:

$$|\psi_1\rangle = \frac{1}{\sqrt{2^n}} \sum_{i \in \{0,1\}^n} |i\rangle \left[\frac{|0\rangle - |1\rangle}{\sqrt{2}}\right],$$

where the only difference with respect to the analogous term in the previous approach is the last term.

Now, after the application of the query, which is given by $U_f : |x, y\rangle \to |x, y \oplus f(x)\rangle$, one gets

$$|\psi_2\rangle = \frac{1}{\sqrt{2^n}} \sum_{i \in \{0,1\}^n} (-1)^{f(i)} |i\rangle \left[\frac{|0\rangle - |1\rangle}{\sqrt{2}}\right].$$

Then, if we apply the second Hadamard gates, we get:

$$|\psi_3\rangle = \frac{1}{2^n} \sum_{j\in\{0,1\}^n} \sum_{i\in\{0,1\}^n} (-1)^{i\cdot j+f(i)} |i\rangle \left[\frac{|0\rangle - |1\rangle}{\sqrt{2}}\right].$$

Finally, performing the same measurement as in the previous case, one gets the final output as required.

To see the improvement with respect to classical algorithms, notice that any classical deterministic algorithm needs, at least, $N/2+1$ queries. On the other hand, a classical algorithm can solve this problem efficiently if we allow a small error probability: Query x at two random positions, and output *constant* if they are the same and *balanced* otherwise. Then, the algorithm gets the correct answer with probability 1 if x is constant, and with probability $1/2$ when x is balanced.

Therefore, the difference between quantum and classical for this problem only appears if we consider algorithms without error probability.

3.3.3 Bernstein-Vazirani problem

In this subsection, we introduce the Bernstein-Vazirani problem.

Problem 3.3.2 For $N = 2^n$, we are given an element $x \in \{0,1\}^N$ such that there is some $a \in \{0,1\}^n$ verifying:

$$x_i \equiv i \cdot a \pmod 2 \qquad \text{for every } i,$$

where $i \cdot a$ denotes the usual scalar product between vectors, i.e.,

$$i \cdot a = \sum_{k=1}^{n} i_k a_k.$$

The **goal** is to find a.

Compared to the algorithms that we have just presented, it is easy to notice that, essentially, this algorithm is the same as the Deutsch-Jozsa algorithm. The main difference yields in the fact that, in this case, the final observation yields a.

By assumption,

$$x_i \equiv i \cdot a \pmod 2 \qquad \text{for every } i,$$

so

$$(-1)^{x_i} = (-1)^{(i\cdot a)}$$

and we can write the state obtained after the query in the following form:

$$\frac{1}{\sqrt{2^n}} \sum_{i\in\{0,1\}^n} (-1)^{x_i} |i\rangle = \frac{1}{\sqrt{2^n}} \sum_{i\in\{0,1\}^n} (-1)^{i\cdot a} |i\rangle.$$

We recall now that the Hadamard gate is its own inverse since this implies that applying a Hadamard to each qubit of the above expression will turn it into $|a\rangle$. Hence, this solves the Bernstein-Vazirani problem with 1 query and $O(n)$ other operations.

Differently from the Deutsch-Jozsa algorithm, every classical algorithm needs n queries for information purposes, and the final answer consists of n bits and a classical query which contains, at most, 1 bit of information.

Remark 3.3.3 Bernstein and Vazirani also presented a recursive version of the problem stated above, which can be solved by a quantum algorithm in a polynomial number of steps. However, for this version, every classical randomized algorithm needs $n^{\Omega(\log n)}$ steps.

To finish this section, notice that the Deutsch-Jozsa problem has shown an exponential quantum improvement over the best deterministic classical algorithms, whereas the Bernstein-Vazirani problem has shown a polynomial improvement over the best randomized classical algorithms that have probability error $\leq 1/3$. In the following section, we will combine these two features, to obtain a problem where quantum computers are exponentially more efficient than bounded-error randomized algorithms.

3.4 Simon's algorithm

In this section, we change slightly the previous notation for convenience. Consider again $N = 2^n$ and denote by $[N] = \{1, \dots, N\}$, which can be identified with $\{0,1\}^n$. Remember also that by $\oplus$ we are denoting the addition modulo 2 (or addition in $\mathbb{F}_{2^n}$). Then, Simon's problem can be stated as follows:

Problem 3.4.1 For $N = 2^n$, we are given an element $x = (x_1, \dots, x_N)$ with $x_i \in \{0,1\}^n$ for every i, such that there is some $s \in \{0,1\}^n$, non null and unknown, verifying:

$$x_i = x_j \quad \text{iff } i = j \text{ or } i = j \oplus s \qquad \text{for every } i.$$

The **goal** is to find s.

Notice that x can be seen as a function $x : [N] \to [N]$ 2-to-1 (two values of the domain are projected onto the same value in the codomain), where this fact is determined by s. Differently from the previous examples, the queries now are of the same form: The input $x = (x_1, \dots, x_N)$ has variables that are not scalars, but strings of n numbers themselves, and one query produces completely such string $|i, 0^n\rangle \mapsto |i, x_i\rangle$.

This problem can also be seen in the following form: We can consider that we have $n2^n$ binary variables that we can query individually. Moreover, we can simulate one x_i-query using only n binary queries. This alternative view does not affect the number of queries too much.

The algorithm is quite similar to Deutsch-Jozsa's one. Now, we start with $2n$ zero qubits of the following form:

$$|0^n\rangle\,|0^n\rangle$$

and apply Hadamard transforms just to the first n qubits, i.e., to $|0^n\rangle$, getting:

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

The second n qubits are still zero, but when we apply a query to the whole term, we get

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

At this point, the measurement algorithm is only applied to the second n collection of qubits. This measurement is not really necessary but simplifies notably the posterior analysis. The outcome will be some value x_i, which implies that the first part also collapses to the superposition of the two indices having the x_i-value:

$$\frac{1}{\sqrt{2}} \left(|i\rangle + |i \oplus s\rangle\right) |x_i\rangle.$$

Once we have this term, we apply Hadamard gates just to the first n qubits again.

$$\frac{1}{\sqrt{2^{n+1}}}\left(\sum_{j\in\{0,1\}^n}(-1)^{i\cdot j}\,|j\rangle+\sum_{j\in\{0,1\}^n}(-1)^{(i\oplus s)\cdot j}\,|j\rangle\right),$$

and taking into account that the following property holds for the exclusive or operation

$$(i\oplus s)\cdot j=(i\cdot j)\oplus(s\cdot j),$$

we can write the resulting state as:

$$\frac{1}{\sqrt{2^{n+1}}}\left(\sum_{j\in\{0,1\}^n}(-1)^{i\cdot j}(1+(-1)^{s\cdot j})\,|j\rangle\right).$$

For the final measurement, notice that $|j\rangle$ has non-zero amplitude iff $s\cdot j=0 \mod 2$, so measuring the state gives an element from the set $\{j|s\cdot j=0 \mod 2\}$. Hence, we are getting a linear equation that gives some information about s.

We can repeat now this algorithm until we get $n-1$ independent linear equations with information on s. The solutions to these equations will be 0^n and the right s (which can be computed by a classical algorithm).

All of this can be done by means of a classical circuit of size $O(n^3)$. However, Simon's algorithm finds s using $O(n)$ operations in the form of x_i queries and a polynomial amount of many other operations.

3.5 Fourier transform

In this section, we are going to present and develop the quantum Fourier transform, which is the key ingredient for quantum factoring and many other interesting quantum algorithms, as we will see in the following sections. The quantum Fourier transform is essentially an efficient quantum algorithm for performing a Fourier transform of quantum mechanical amplitudes, and it appears in many different versions throughout classical computing, in areas ranging from signal-processing to data compression in complexity theory.

In its usual mathematical notation, the *discrete Fourier transform* takes as input a vector of complex numbers $x_0,\dots x_{N-1}$, where the length of this vector, N, is a fixed parameter, and outputs another vector of complex numbers $y_0,\dots,y_{N-1}$ defined by

$$y_k:=\frac{1}{\sqrt{N}}\sum_{j=0}^{N-1}x_j e^{2\pi ijk/N}.$$

Notice that, in the previous expression, i stands for the imaginary unit.

The quantum Fourier transform is essentially the same transformation, although the conventional notation for the quantum Fourier transform is a bit different. Consider an orthonormal basis $\{|0\rangle,\dots,|N-1\rangle\}$. Then, the *quantum Fourier transform* is defined as the linear operator with the following action on the basis states:

$$|k\rangle\mapsto\frac{1}{\sqrt{N}}\sum_{j=0}^{N-1}e^{2\pi ijk/N}\,|j\rangle.$$

Equivalently, by considering the action on an arbitrary state, we can write:

$$\sum_{j=0}^{N-1}x_j\,|j\rangle\mapsto\sum_{k=0}^{N-1}y_k\,|k\rangle,$$

where the *amplitudes* y_k are the discrete Fourier transform of the amplitudes x_j. Moreover, this definition is a unitary transformation, which implies that can be implemented as the dynamics for a quantum computer.

We can derive an equivalent expression for the quantum Fourier transform. For that, let $N = 2^n$ for some integer n and consider the basis $\{|0\rangle, \ldots, |2^n - 1\rangle\}$, which is, as mentioned previously, the computational basis for an n qubit quantum computer. For purposes of simplification of notation in the derivation of the aforementioned expression, let us write the state $|j\rangle$ using the binary representation $j = j_1 j_2 \ldots j_n$[1].

Then, after some calculations that we omit in this text for the sake of simplicity in the reading (we refer the reader to [14] for the specific algebra), the quantum Fourier transform can be given the following useful product representation:

$$|j_1, \ldots, j_n\rangle \mapsto \frac{\left(|0\rangle + e^{2\pi i 0, j_n} |1\rangle\right)\left(|0\rangle + e^{2\pi i 0, j_{n-1} j_n} |1\rangle\right)\left(|0\rangle + e^{2\pi i 0, j_1 \ldots j_n} |1\rangle\right)}{2^{n/2}}.$$

This product representation allows to construct an efficient quantum circuit to compute the Fourier transform (see Figure 3.7, extracted from [14]) and to prove that the quantum Fourier transform is unitary. It also provides insight into algorithms based upon the quantum Fourier transform.

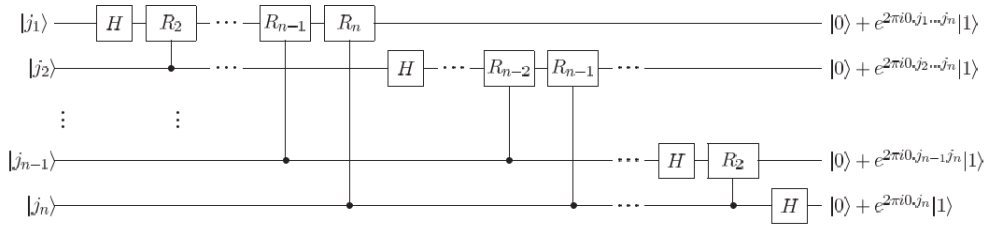

Fig. 3.7: Quantum circuit to compute the quantum Fourier transform.

An important fact to compare the quantum algorithm to compute the quantum Fourier transform with their classical analogues is the number of gates that the circuit of Figure 3.7 uses. The different steps of this circuit have the following numbers of gates, respectively:

1. One Hadamard gate and $n - 1$ phase gates on the first qubit $\mapsto n$ gates.
2. One Hadamard gate and $n - 2$ phase gates on the second qubit $\mapsto n - 1$ gates.

[1] Given a state $|j\rangle$, we say that its binary representation is given by

$$j = j_1 j_2 \ldots j_n,$$

or, more formally,

$$j = j_1 2^{n-1} + j_2 2^{n-2} + \ldots + j_n 2^0$$

if $|j\rangle$ is of the form

$$|j\rangle = |j_1\rangle \otimes |j_2\rangle \otimes \ldots |j_n\rangle.$$

Following this convention, it is also convenient to adopt the notation $0, j_1 \ldots j_n$ to represent the binary fraction

$$\frac{j_1}{2} + \ldots \frac{j_n}{2^n}.$$

3. ...

Hence, we see that

$$n + (n-1) + \ldots + 1 = \tfrac{n(n+1)}{2}$$

gates are required, plus the swap ones. Indeed, at most $n/2$ swaps are required, and each swap is accomplished using three CNOT gates.

Therefore, this circuit provides a $O(n^2)$ algorithm to perform the quantum Fourier transform. In contrast, the best classical algorithm for computing the discrete Fourier transform on 2^n elements are elements such as the *Fast Fourier Transform (FFT)*. This algorithm computes the discrete Fourier transform using $O(n2^n)$ gates. This implies that it is exponentially "worse" than its quantum analogue, meaning that it requires exponentially more operations to compute the Fourier transform on a classical computer than to implement the quantum Fourier transform on a quantum computer.

To finish this section, let us remark that, even though the previous comparison between classical and quantum algorithms to compute Fourier transforms, and the exponential improvement of the latter with respect to the first one, might allow us to think that this could have huge applications in real-world data processing applications (for example, in computer speech recognition, the first step in the recognition of phonemes is to perform the Fourier transform to the digitalized sound), in general one cannot use the quantum Fourier transform to speed up the computation of these Fourier transforms. This is due to the fact that amplitudes in a quantum computer cannot be directly accesed by a measurement and, thus, it is not possible to determine the Fourier transformed amplitudes of the original state. However, the quantum Fourier transform is still quite useful for some purposes related with some specific algorithms, as we will see in the following two sections.

3.6 Shor's factoring algorithm

The most important quantum algorithm so far is probably Shor's algorithm. If we could use a quantum computer with a sufficiently large number of qubits and without succumbing to noise and other quantum decoherence phenomena, Shor's algorithm would break public-key cryptography schemes such as the widely used RSA scheme, which is based on the assumption that factoring large numbers is computationally intractable. Currently it is known that this assumption is valid for classical (non-quantum) computers, since no classical algorithm is known that can factor in polynomial time. However, Shor's algorithm shows that factoring is efficient on an ideal quantum computer, so it may be feasible to defeat RSA by constructing a large quantum computer. Hence, it constitutes a powerful motivation for the design and construction of quantum computers and for the study of new quantum computer algorithms. Moreover, a variation of this algorithm can also be used to attack the schemes based on the discrete logarithm problem (in particular, the ones associated to elliptic curves).

We will devote this section to explain this algorithm and to introduce the necessary tools to understand the soluiton to the problem that it solves. Most of the images that appear on this section have been extracted from [22], a text that we will also follow in the proofs that appears here.

Roughly speaking, the statement of the problem of factoring integer, in which Shor's algorithm can be applied, is as follows.

Problem 3.6.1 Let N be a natural number. The goal is to find $e_1, \dots e_n$ natural numbers and $p_1, \dots, p_n$ prime numbers so that

$$N = p_1^{e_1} \cdot \dots \cdot p_n^{e_n}.$$

Notice that this decomposition of N in prime factors is unique.

We can do two initial simplifications to the problem without loss of generality. The first one is that, given N, it is enough to split it into two other integers N_1 and N_2 such that $N = N_1 \cdot N_2$. Hence, after a linear number in size of the input of such steps, we are guaranteed to reach prime factors, thus the second simplification is that we can assume that N is a product of two primes, $N = p \cdot q$.

Classically, there exist some naive algorithms for the factoring problem that work in time $O(\sqrt{N})$. The fastest known algorithm for this problem is called *Field Sieve algorithm*, and it works in time $2^{O(\sqrt[3]{\log N})}$.

Shor's result yields the fact that one can do better with a quantum computer.

Theorem 3.6.2 There exists quantum algorithm that solves the factoring problem with bounded error probability in polynomial time.

We will devote most of this section to prove this result, following several steps for that. First, we will show that the factoring problem is equivalent to the *order-finding problem*, since a fast algorithm for order-finding problem implies a fast algorithm for factoring problem. Later, we will present a fast quantum algorithm for order-finding.

3.6.1 Reduction of factoring to order-finding

First, notice that the set of modulo N coprime numbers with N, i.e.,

$$\{1 \leq x \leq N \,:\, \gcd(x, N) = 1\}$$

where $\gcd(x, N)$ denotes the *greatest common divisor* of x and N, forms a group under multiplication modulo N. Moreover, given x and N such that $\gcd(x, N) = 1$, we define the order of x by the minimum positive r such that $x^r \equiv 1 (\text{mod } N)$, and denote it by $\text{ord}(x)$.

We can now state the order-finding problem.

Problem 3.6.3 Given x and N such that $\gcd(x, N) = 1$, the **order-finding problem** consists of finding $ord(\text{x})$.

We will devote the rest of the subsection to prove that Problem 3.6.1 can be reduced to Problem 3.6.3.

For that, let us first introduce and prove three technical lemmata. The first and last one will be essential to obtain the desired result, whereas the second one will be used in the proof of the third one.

Lemma 3.6.4 Given a composite number N and a number x that is a nontrivial square root of 1 modulo N, i.e., verifying $x^2 \equiv 1 (\text{mod } N)$ with neither $x \equiv 1 (\text{mod } N)$ nor $x \equiv -1 (\text{mod } N)$ holding, we can efficiently compute a nontrivial factor of N.

Proof. Since, by assumption, $x^2 \equiv 1 (\text{mod } N)$, we have:

$$x^2 - 1 \equiv (x - 1)(x + 1) \equiv 0 (\text{mod } N).$$

Now, as none of

$$x \equiv 1(\text{mod } N)\,, \qquad x \equiv -1(\text{mod } N)$$

hold, we know that $1 < x < N-1$, and, hence, both $\gcd(x-1, N)$ and $\gcd(x+1, N)$ are nontrivial factors of N.

Finally, since there exists a fast algorithm for computing the gcd of two numbers, *Euclid's algorithm*, the efficiency follows from there.

■

Now, the next lemma reads as follows:

Lemma 3.6.5 Let p be an odd prime and let x be a uniformly random element verifying $0 \leq x < p$. Then, ord(x) is even with probability, at least, $1/2$.

Proof. In virtue of *Fermat's little theorem*, it is clear that, for every x, the following holds:

$$x^{p-1} \equiv 1(\text{mod } p).$$

Another well-known result in mathematics is that the multiplicative group modulo a prime number is a cyclic group, i.e., there is an element g (from '*generator*') that generates all the elements of the group, in the sense that any element can be written by

$$x \equiv g^k(\text{mod } p)$$

for some k. Since x is chosen uniformly at random, k is odd with probability $1/2$.

Assume now that k is odd. Then, since $x \equiv g^k(\text{mod } p)$, we have

$$x^{\text{ord}(x)} \equiv g^{k\ \text{ord}(x)} \equiv 1(\text{mod } p).$$

From here, we can deduce

$$p-1 | k\ \text{ord}(x),$$

and, thus, since both p and k are odd (and $p-1$ is even), $\text{ord}(x)$ has to be even.

To sum up, if k is odd, something that happens with probability $1/2$, $\text{ord}(x)$ is even. If k is even, however, we cannot say anything on $\text{ord}(x)$ in general. Putting both facts together, $\text{ord}(x)$ is even with probability, at least, $1/2$.

■

The last lemma that we present in this section, and in whose proof we use the previous one, is the following:

Lemma 3.6.6 Let $N = p \cdot q$, with p and q prime numbers, and assume that N is an odd number. Let us further assume that x is taken uniformly at random from $0, \ldots, N-1$.

If $\gcd(x, N) = 1$, then with probability at least $3/8$, we have that $\text{ord}(x) = r$ is even and $x^{r/2} \neq \pm 1(\text{mod } N)$.

Proof. In virtue of the *Chinese remainder theorem*, choosing x uniformly at random from $0, \ldots, N-1$ is the same that choosing x_1 uniformly at random from $0, \ldots p-1$ and, independently, x_2 uniformly at random from $0, \ldots q-1$. Moreover, if we denote $r_1 = \text{ord}(x_1)$ and $r_2 = \text{ord}(x_2)$, we can see that $r_1 | r$ and $r_2 | r$.

Now, we can first see that the probability that r is even is, at least $3/4$. Since we are assuming that N is an odd number, it is clear that both p and q also have to be odd. If x_1 is odd, then r_1 has to be even, and the same happens for x_2 and r_2. Hence, since r_1 even or r_2 even imply that

r is even, and x_1 and x_2 are chosen uniformly at random, applying Lemma 3.6.5 we get that the probability that r is even is at least $3/4$.

Finally, we can prove that the probability of $x^{r/2} \equiv \pm 1 (\text{mod } N)$ is, at most, $1/2$ when r is even. Indeed, notice that, under this assumption, $x^r \equiv 1 (\text{mod } p)$ and there are only two square roots of 1 modulo a prime number, namely ± 1. Again in virtue of the Chinese remainder theorem, it follows that there are only four roots of 1 modulo N. Only two of them make $x^{r/2} \neq \pm 1 (\text{mod } N)$. ■

In the last part of the subsection, we can see that the reduction from Problem 3.6.1 to Problem 3.6.3 follows directly from Lemma 3.6.4 and 3.6.6. Indeed, if someone computes the function $\text{ord}(\cdot)$ for us, the prime factors of N can be found classically. By checking the answer, something that can be done efficiently easily, and repeating the procedure several times, we can increase the probability of success.

3.6.2 The order-finding problem

Now that we have seen that the problem of efficiently factoring a number can be reduced to the problem of efficiently finding $\text{ord}(x) = r$, we are going to find an example for the second problem. This example is Shor's algorithm, and we devote this subsection to analyze it.

First, we will start with a simplified case, before going to the more general case.

3.6.2.1 Simplified case

In this case, we use an auxiliary number Q, which is sufficiently large, and which verifies $Q >> N^2$, and assume that $r|Q$. The case when $r \nmid Q$ does not differ drastically to the one we are studying, so we will restrict for the moment to that for the simplicity of notation, and we will proceed to the more difficult case in the next section.

The algorithm that we are going to present uses two registers:

- *Register 1* stores a number $(\text{mod } Q = 2^q)$.
- *Register 2* stores a number $(\text{mod } N)$.

It also has several steps, that we are going to present individually:

1. The registers are initially in the state $|0\rangle \otimes |0\rangle$.
2. We apply the Fourier Transform modulo Q to the first qubit, to get the state

$$\frac{1}{\sqrt{Q}} \sum_{a=0}^{Q-1} |a\rangle \otimes |0\rangle.$$

3. Consider now the function $f(a) = x^a (\text{mod } N)$, a function that we can easily compute classically, and has r as its smallest order (or *period*). Notice that this function can be computed in $\log a$ multiplications, and also that f is different in $[0, r-1]$ (otherwise, it would have a smaller period). Applying then f to the state obtained for the first qubit in the previous step, we get

$$\frac{1}{\sqrt{Q}} \sum_{a=0}^{Q-1} |a\rangle \, |f(a)\rangle.$$

4. At this point, we measure the second qubit, and when we perform this measurement, the second register collapses to some value, $f(k)$, for k uniformly random over $0, \dots r-1$. Then, all superposed states which are inconsistent with the measured value must disappear. Hence, the state of the two registers must be given by

$$\frac{1}{\sqrt{\frac{Q}{r}}} \sum_{a=0}^{\frac{Q}{r}-1} |ar+k\rangle\,|f(k)\rangle.$$

5. We have set up a periodic superposition of period r in the first register (the value that we wanted to compute), so we can drop the second register. At this moment, Shor's algorithm comes into play, by Fourier sampling modulo Q.

   As we are going to perform Fourier sampling, we can drop the shift value k by the properties of Fourier Transforms seen in Section 3.5. This allows us to move k to phase. Then, applying the Fourier sample to the state

$$\frac{1}{\sqrt{\frac{Q}{r}}} \sum_{a=0}^{\frac{Q}{r}-1} |ar+k\rangle,$$

   we get

$$\frac{1}{\sqrt{r}} \sum_{a=0}^{r-1} \omega^{ak} \left| a\frac{Q}{r} \right\rangle,$$

   where ω is a primitive qth root of unity, i.e.,

$$\omega = e^{\frac{2\pi i}{Q}}.$$

6. Now, we measure this register. The measurement provides $\left|a\frac{Q}{r}\right\rangle$, where a is a random variable uniformly from $0, \dots, r-1$. It is easy to see then that with great probability we have $\gcd\left(a, \frac{Q}{r}\right) = 1$. If that is the case, then by computing $\gcd\left(a\frac{Q}{r}, Q\right)$ we shoul get $\frac{Q}{r}$. Since we already know Q, it is clear that, from $\frac{Q}{r}$, computing r is straightforward.

3.6.2.2 General case

In the previous subsubsection we have assumed $r|Q$. As we said in its introduction, we will devote this subsubsection to the case $r \nmid Q$, which is a bit more difficult in notation, but the same in spirit.

Let us perform the same algorithm than before up to Step 3. From Step 4 on, the situation is a bit different. After applying the first measurement, we get the state:

$$\frac{1}{\sqrt{\lfloor \frac{Q}{r} \rfloor}} \sum_{a=0}^{\lfloor \frac{Q}{r} \rfloor - 1} |ar+k\rangle.$$

In this case, this is not a coset of a subgroup, so we cannot proceed as in the previous case. However, we can anyway take the Fourier transform, and we will show that we get a constructive interference primarily at the points which are close to multiples of $\dfrac{Q}{r}$, so close that they can be 'rounded' to the nearest multiple. This is, in summary, the fact that will allow us to calculate r with reasonable probability.

Let us explain that step by step. First, if we apply a Fourier transform to the previous expression, we get

$$\sum_{k=0}^{Q-1} \alpha_k |k\rangle,$$

where the coeficient α_k is given by:

$$\alpha_k = \frac{1}{\sqrt{Q}} \cdot \frac{1}{\sqrt{\lfloor \frac{Q}{r} \rfloor}} \sum_{a=0}^{\lfloor \frac{Q}{r} \rfloor - 1} \left(\omega^{rk}\right)^a.$$

One can notice that if Q is small, then terms in the sum cover only a small angle of the complex plane, and hence the magnitude of the sum is almost the sum of the magnitudes. We will see this fact explicit in a couple of lemmas that appear at the end of this section, where we will prove that, with probability more than $1/16$, we can sample a k such that

$$-\frac{r}{2} \leq kr (\text{mod } Q) \leq \frac{r}{2}.$$

Meanwhile, let us just assume that this fact is true and continue with the algorithm. First, notice that the previous expression is equivalent to

$$|kr - lQ| \leq \frac{r}{2}$$

for a certain integer l, or what is the same,

$$\left| \frac{k}{Q} - \frac{l}{r} \right| \leq \frac{1}{2Q}.$$

This condition implies that $\frac{k}{Q}$ is a $\frac{1}{2Q}$-approximation of $\frac{l}{r}$. Moreover, if we measure k, we get to know Q, so we know the quotient, which constitutes the good approximation of $\frac{l}{r}$.

Furthermore, notice that l is randomly chosen from $[0, r-1]$, which implies that, with probability at least $1/\log l$, l and r are coprimes, which allows us to compute r from l/r. Then, in principle, if we are able to choose Q much larger than N, this way of reasoning provides a good approximation. To see how much larger than N needs to be Q we use *continued fractions*[2]. Indeed, we just have to compute continued fractions until we get precision of at least $\frac{1}{2Q}$. Since l/r is rational, we know that, for a certain integer n, the continued fractions verify $CF_m(l/r) = l/r$ for every $m \geq n$. Hence, if we assume that the approximation is some rational number l'/r', clearly $r = r'$. Otherwise, we would have

[2]The idea for this part of the proof lies in the use of continued fractions to approximate real numbers using finite numbers of integers.

A *continued fraction* is defined in the following way: A real number α can be approximated by a set of positive integers $a_0, a_1 \ldots a_n$ by

$$CF_n(\alpha) = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ldots + \cfrac{1}{a_n}}}} = \frac{A_n}{B_n},$$

where the numbers a_i are chosen in a specific way, and A_n and B_n are always integers. To explain the procedure to chose these numbers, let us consider π and suppose that we want to approximate it with four decimals. Then, we get:

$$\pi \approx 3,1415 = 3 + \frac{1415}{10000} = 3 + \frac{1}{7 + \frac{95}{1415}} = 3 + \frac{1}{7 + \frac{1}{14 + \frac{85}{95}}}.$$

Moreover, continued fractions satisfy the following two important properties, which we are essential for the last step of the algorithm:

1. $CF_n(\alpha)$ is the best rational approximation of α with denominator $\leq B_n$.
2. If α is rational, then it coincides, from some n on, with the approximations $CF_n(\alpha)$.

Finally, it is easy to notice that continued fractions are easily computable for any rational number.

$$\left|\frac{l}{r}-\frac{l'}{r'}\right| \geq \frac{1}{rr'} \geq \frac{1}{N^2},$$

and this would be a contradiction, since both $\frac{l}{r}$ and $\frac{l'}{r'}$ are $\frac{1}{2Q} \leq \frac{1}{2N^2}$ close to the same rational fraction (the one that approximates $\frac{k}{Q}$ from the beginning).

Therefore, $r = r'$, so by using these continued fractions we have finished the algorithm.

In the last part of this section, we will state and prove a couple of lemmas that have been used in the development of the algorithm.

Lemma 3.6.7 If $-\frac{r}{2} \leq kr(\text{mod } Q) \leq \frac{r}{2}$ for some kr, then

$$|\alpha_k| \geq \frac{1}{2^{2/3}\sqrt{r}}.$$

Proof. Let us denote

$$\beta = e^{\frac{2\pi i r k}{Q} a} = \omega^{rk}.$$

It is clear that this expression corresponds to a vector of the complex plane. Also, the sum that appears in the term α_k, i.e.,

$$\sum_{a=0}^{\lfloor \frac{Q}{r} \rfloor - 1} \beta^a$$

is a geometric series of ratio β.

Because of the assumption of the statement of the lemma $\left(-\frac{r}{2} \leq kr(\text{mod } Q) \leq \frac{r}{2}\right)$, the terms of the series cover less than or equal to an angle π on the complex plane (since β makes a small angle with the real line). Then, as we can see in Figure 3.8,

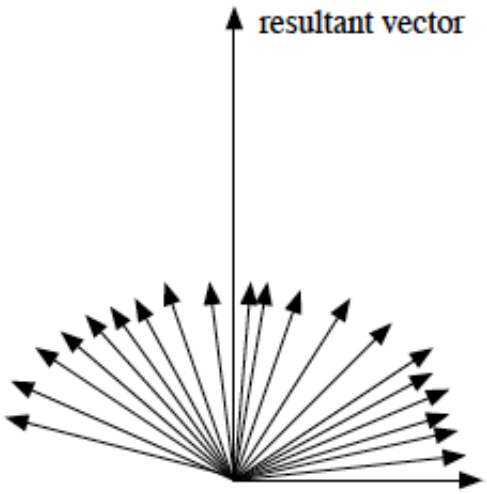

Fig. 3.8: β makes a small angle with the real line.

half of the terms on the previous series make an angle which is less than or equal to $\frac{\pi}{4}$ with the resultant vector of the addition of the terms in the series. Then, since the cosinus is a decreasing function from 0 to $\pi/4$, each term contributes with a fraction that is at least:

$$\cos\left(\frac{\pi}{4}\right) = \frac{1}{\sqrt{2}}$$

of its length to the resultant vector. Hence, the magnitude of the resultant is, at least:

$$\frac{1}{2} \cdot \frac{1}{\sqrt{2}} \cdot \frac{Q}{r} \cdot \frac{1}{\sqrt{Q}} \frac{1}{\sqrt{\lfloor \frac{Q}{r} \rfloor}} = \frac{1}{2^{3/2}} \cdot \frac{1}{\sqrt{r}}.$$

■

Lemma 3.6.8 The fact of the statement of the previous lemma, i.e.,

$$-\frac{r}{2} \leq kr(\text{mod } Q) \leq \frac{r}{2}$$

happens with probability $P(X \leq 1)$, where X is a random variable that follows a distribution $\mathcal{N}(0,1)$.

Proof. If $\gcd(r, Q) = 1$, then the element $r^{-1}(\text{mod } Q)$ exists. Thus, as the variable k varies in the range $[0, Q-1]$, $k \cdot r$ must take values that constitute a permutation of $\{0, 1, \ldots, Q-1\}$. As we can see in the following picture,

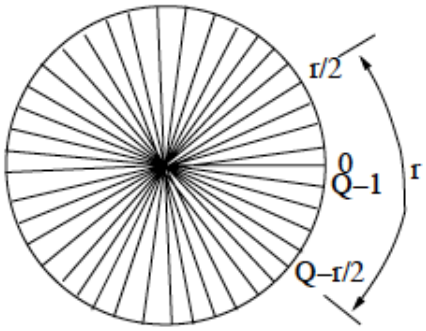

Fig. 3.9: At least r values of kr lie in the range $[Q - r/2, r/2]$.

at least r values of kr lie in the range $[Q - r/2, r/2]$.

Now, assume that $\gcd(r, Q) \neq 1$. In this case, the distribution of $k \cdot r \pmod{Q}$ is a bit different, and can be shown in the following picture:

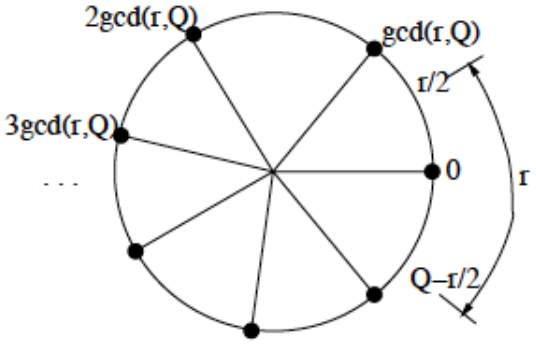

Fig. 3.10: At least $r/2$ values of kr lie in the range $[Q - r/2, r/2]$.

In this case, as we can see, at least $r/2$ values of $k \cdot r$ lie in the desired range $[Q - r/2, r/2]$.

Hence, if we consider both cases together, we have that, in any case, at least $r/2$ values of kr lie in the range $[Q - r/2, r/2]$, thus satisfy the condition:

$$-\frac{r}{2} \leq kr(\text{mod } Q) \leq \frac{r}{2}.$$

In virtue of Lemma 3.6.7, since each one of them has an amplitude which is at least

$$\frac{1}{2^{3/2}r^{1/2}},$$

we get that the probability to get such a k when sampling is, at least,

$$\frac{r}{2}\left(\frac{1}{2^{3/2}r^{1/2}}\right)^2$$

and this is greater than $1/16$. Therefore, the condition on the statement of the lemma happens with probability greater than $P(X \leq 1)$ for a normal random variable.

■

To conclude this section, let us just recall that, to factor and integer N, Shor's algorithm runs in polynomial time. More specifically, it takes quantum gates of order $O((\log N)^2(\log\log N)(\log\log\log N))$ using fast multiplication, and, thus, the integer factorization problem can be efficiently solved on a quantum computer (and belongs to the complexity class **BQP**). This proves Theorem 3.6.2.

For the sake of simplicity, we can briefly summarize the whole factorization of N in the following steps:

1. A reduction of the factoring problem to the problem of order-finding (this can be done on a classical computer).

2. A quantum algorithm to solve the order-finding problem.

   Let r be the period that we want to find.

   (a) Initialize the registers in the state $|0\rangle \otimes |0\rangle$.

   (b) Apply the Fourier Transform modulo Q to the first qubit, to get the state

   $$\frac{1}{\sqrt{Q}}\sum_{a=0}^{Q-1}|a\rangle \otimes |0\rangle.$$

   (c) Consider $f(a) = x^a(\text{mod } N)$ and apply it to the state:

   $$\frac{1}{\sqrt{Q}}\sum_{a=0}^{Q-1}|a\rangle\,|f(a)\rangle.$$

   (d) Measure the second qubit:

   $$\frac{1}{\sqrt{\lfloor \frac{Q}{r} \rfloor}}\sum_{a=0}^{\lfloor \frac{Q}{r} \rfloor - 1}|ar+k\rangle.$$

   (e) Apply Fourier transform:

   $$\sum_{k=0}^{Q-1}\alpha_k\,|k\rangle, \text{ where } \alpha_k = \frac{1}{\sqrt{Q}}\cdot\frac{1}{\sqrt{\lfloor \frac{Q}{r} \rfloor}}\sum_{a=0}^{\lfloor \frac{Q}{r} \rfloor - 1}\left(\omega^{rk}\right)^a.$$

   (f) With probability more than $1/16$, we can sample a k such that

   $$-\frac{r}{2} \leq kr(\text{mod } Q) \leq \frac{r}{2}.$$

(g) The previous fact implies that $\frac{k}{Q}$ is a $\frac{1}{2Q}$-approximation of $\frac{l}{r}$. Using continued fractions, we get r from l/r.

This is much faster than the most efficient known classical factoring algorithm, the *Field Sieve algorithm*, that works in time $2^{O(\sqrt[3]{\log N})}$. The efficiency of Shor's algorithm, as seen before, is due to the quantum Fourier transform and the modular exponentiation.

3.7 Grover's algorithm

The second most important quantum algorithm, after Shor's, is Grover's quantum search problem [10]. Attacks based on this algorithm can be used to break cryptographic schemes in symmetric key, such as AES, as well as schemes based on hash functions, such as SHA2 or SHA3.

In general, as mentioned in previous sections, much of the excitement concerning quantum computation comes from the fact that quantum algorithms can provide improvements over classical algorithms. In this particular case, Grover's algorithm exhibits a quadratic speedup with respect to the classical case. Although it does not provide exponential speedup, as Shor's did, it is much more applicable.

Problem 3.7.1 For $N = 2^n$, consider an arbitrary $x \in \{0,1\}^N$. The goal is to find $i \leq N$ such that $x_i = 1$, and to output 'no solutions' if such i does not exist.

This problem is called *the unstructured search* problem.

It can be formulated equivalently as a database search problem, in which we consider a database and we want to find an item in it which fulfils certain specifications. One example of this can be a database of N names and we want to find the position of a specific name in it.

We call this search "unstructured" because we are not given any information about how the database is ordered. If, for example, we knew in advance that the database was sorted, then we could perform a binary search and find any element in logarithmic time. However, since that is not the case, with classical circuits one cannot do better than performing a linear number of queries to find the target element.

However, in the quantum setting, Grover's algorithm leads to the following theorem:

Theorem 3.7.2 The unstructured search problem can be solved in $O(\sqrt{N})$ queries using quantum computation (and $O(\sqrt{N}\log N)$ other gates).

In fact, it was shown in [3] that, up to a constant factor, this is the best that one can do for this problem under the quantum computational model, since the query complexity is $\Theta(\sqrt{N})$. In particular, since the unstructured search problem cannot be solved in logarithmic time, this problem cannot be used as a way to solve NP problems in polynomial time. However, there is still a quadratic improvement with respect to the classical case.

3.7.1 Algorithm

Let us consider the following notation introduced in previous sections:

$$O_{x,\pm}|i\rangle = (-1)^{x_i}|i\rangle$$

to denote the $\pm$-type oracle for the input x. Let us also denote by R the unitary transformation that puts a phase -1 in front of all basis states which are different from $|0\rangle$:

$$R\,|i\rangle = \begin{cases} -\,|i\rangle & \text{if } |i\rangle \neq |0\rangle \\ |i\rangle & \text{if } |i\rangle = |0\rangle \end{cases}$$

This transformation R is clearly independent of the input x for the algorithm, and can be implemented using $O(n)$ elementary gates.

Now, let us define by *Grover iterate* the following quantity:

$$\mathcal{G} = H^{\otimes n} R H^{\otimes n} O_{x,\pm},$$

where we recall that H denotes the Hadamard gate. Notice that 1 Grover iterate provides 1 query.

Let us develop now each one of the steps of the algorithm:

1. Grover's algorithm starts with the n-bit state $|0^n\rangle$.
2. Later, it applies a Hadamard transformation to all the qubits, and gets the uniform superposition

$$|U\rangle = \frac{1}{\sqrt{N}} \sum_i |i\rangle$$

of all the indices N.

3. Now, it applies the Grover iterate $\mathcal{G}$ to this state k times, where this k will be conveniently chosen later.
4. Finally, it measures the final state.

We can see this algorithm represented graphically in Figure 3.11.

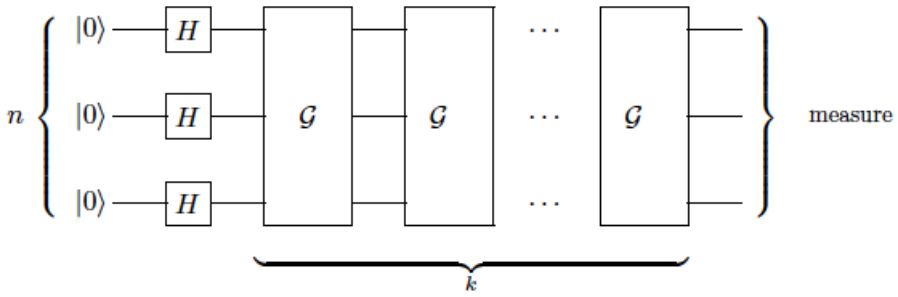

Fig. 3.11: Graphical representation of Grover's algorithm.

Intuitively, the idea behind this algorithm is that in each iteration some amplitude is moved from the indices of the 0-bits to the ones of the 1-bits. Hence, the algorithm stops when nearly all the amplitude is on the indices of the 1-bits, and in this case the measurement of the final state will probably give the index of a 1-bit.

Let us now analyze this explicitly. For that, we define the following states:

$$|G\rangle := \frac{1}{\sqrt{t}} \sum_{i\,:\,x_i=1} |i\rangle\,, \qquad |B\rangle := \frac{1}{\sqrt{N-t}} \sum_{i\,:\,x_i=0} |i\rangle,$$

where $t = \#\{i\,:\,x_i = 1\}$. The first one provides a superposition of the basis states for whose index i we have $x_i = 1$, and the second one the same when $x_i = 0$. Then, it is clear that we can write the uniform superposition of the second state of the algorithm as:

$$|U\rangle = \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} |i\rangle = \sin(\theta)\,|G\rangle + \cos(\theta)\,|B\rangle\,,$$

for a certain θ that is given by

$$\theta = \arcsin\left(\sqrt{\frac{t}{N}}\right).$$

To determine the value of k, notice that Grover iterate $\mathcal{G}$ is actually the product of two reflections in the 2-dimensional space spanned by $|G\rangle$ and $|B\rangle$. Indeed, since a *reflection* in a subspace $W \subset V$ is a unitary A such that $Av = v$ for every $v \in V$, and $Aw = -w$ for every w orthogonal to V, we can easily see that

- $O_{x,\pm}$ is a reflection through $|B\rangle$.
- $H^{\otimes n} R H^{\otimes n}$ is a reflection through $|U\rangle$.

  Indeed, the following is satisfied:

$$H^{\otimes n} R H^{\otimes n} = H^{\otimes n}(2\,|0^n\rangle\,\langle 0^n| - \mathbb{1})H^{\otimes n} = 2\,|U\rangle\,\langle U| - \mathbb{1}.$$

Hence, we can restate Grover's algorithm as follows, assuming that we know that the fraction of solutions is $\varepsilon = t/N$:

1. Grover's algorithm starts with the n-bit state $|0^n\rangle$.
2. Later, it applies a Hadamard transformation to all the qubits, and gets the uniform superposition

$$|U\rangle = \frac{1}{\sqrt{N}} \sum_{i} |i\rangle$$

   of all the indices N.
3. Now, it repeats the following k times, where $k = O(1/\sqrt{\varepsilon})$:
   (a) It reflects through $|B\rangle$, i.e., it applies $O_{x,\pm}$.
   (b) It reflects through $|U\rangle$, i.e., it applies $H^{\otimes n} R H^{\otimes n}$.
4. Finally, it measures the first register and checks that the resulting i is a solution.

3.7.2 Geometrical proof of the algorithm

In this subsection, we are going to show that the algorithm described above indeed works. This geometrical argument has been extracted from [28].

Let us consider the 2-dimensioanl real plane spanned by $|G\rangle$ and $|B\rangle$. Consider

$$|U\rangle = \sin(\theta)\,|G\rangle + \cos(\theta)\,|B\rangle.$$

Then, if we consider the two reflections mentioned on the third step of the algorithm above, we can see that the initial angle increases from θ to 3θ, moving us towards $|G\rangle$, as we can see in Figure 3.12.

For the next two reflections, the angle increases with another factor 2θ. Hence, after k applications of the two reflections, the initial state has been transformed to

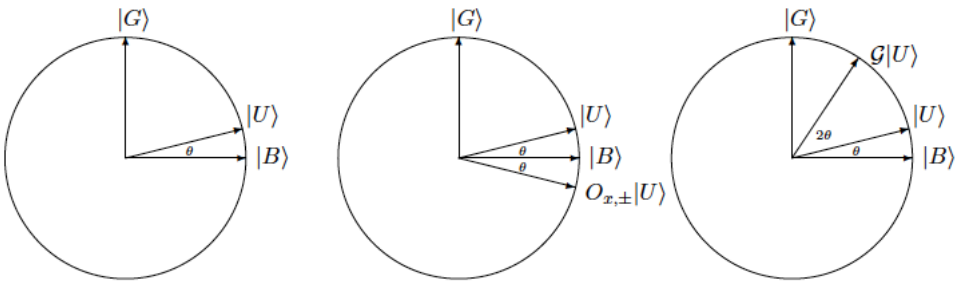

Fig. 3.12: First iteration of the third step of Grover's algorithm. In the first picture, it starts with $|U\rangle$; in the second one, it reflects through $|B\rangle$ to get $O_{x,\pm}|U\rangle$; and in the last one, it reflects through $|U\rangle$ to get $\mathcal{G}|U\rangle$.

$$\sin((2k+1)\theta)\,|G\rangle + \cos((2k+1)\theta)\,|B\rangle.$$

If we now measure this state, it is clear that the probability of seeing a solution is

$$P_k = \sin((2k+1)\theta)^2,$$

and we are interested on this P_k being as close to 1 as possible. Notice that, for example, if we choose $\tilde{k} = \pi/4\theta - 1/2$, so $(2\tilde{k}+1)\theta = \pi/2$ and, thus, $P_{\tilde{k}} = \sin(\pi/2)^2 = 1$.

One example for these quantities can be:

$$t = N/4, \qquad \theta = \pi/6, \qquad \tilde{k} = 1.$$

Unfortunately, in general, $\tilde{k} = \pi/4\theta - 1/2$ is not an integer, and we can only consider an integer number of applications of the Grover iterate. Therefore, our aim will be to choose k as the closest possible to $\tilde{k}$ integer, and in this scenario our final state will still be close to $|G\rangle$.

The failure probability is still small, as we can see below (assuming $t << N$):

$$\begin{aligned} 1 - P_k &= \cos((2k+1)\theta)^2 \\ &= \cos((2\tilde{k}+1)\theta + 2(k-\tilde{k}))^2 \\ &= \cos(\pi/2\theta + 2(k-\tilde{k}))^2 \\ &= \sin(2(k-\tilde{k}))^2 \\ &\leq \sin(\theta)^2 \\ &= \frac{t}{N}, \end{aligned}$$

where we have used $\left|k - \tilde{k}\right| \leq 1/2$ in the inequality. Finally, since $\arcsin(\theta) \geq \theta$, the number of queries is

$$k \leq \frac{\pi}{4\theta} \leq \frac{\pi}{4}\sqrt{\frac{N}{t}}.$$

3.7.3 Amplitude amplification

The analysis worked above can be extended to a much more general setting. To see that, consider $\xi : \mathbb{Z} \to \{0,1\}$ any Boolean function. We will call *solutions* the inputs $z \in \mathbb{Z}$ satisfying $\xi(z) = 1$.

Suppose now that we have an algorithm that checks whether z is a solution or not. Let us denote this by:

$$O_\xi(|z\rangle) = (-1)^{\xi(z)}\,|z\rangle.$$

We further assume that we have some quantum or classical algorithm, denoted by $\mathcal{A}$, that uses no intermediate measurements and, when applied to $|0\rangle$, has probability p of finding a solution. Hence, the *amplitude amplification* algorithm only needs to run the algorithm $\mathcal{A}$ $O(1/\sqrt{p})$ times. This algorithm is implemented as:

1. Consider the starting state $|U\rangle = \mathcal{A}\,|0\rangle$.
2. It repeats the following $O(1/\sqrt{p})$ times:
   (a) It reflects through $|B\rangle$, i.e., it applies O_ξ.
   (b) It reflects through $|U\rangle$, i.e., it applies $\mathcal{A}R\mathcal{A}^{-1}$.
3. It measures the first register and checks that the resulting element x is marked.

We can now define $\theta = \arcsin(\sqrt{p})$ and $|G\rangle$ and $|B\rangle$ as we did before, within the geometric argument for Grover's algorithm. Hence, an analogous reasoning will show that the amplitude amplification finds a solution with probability close to 1.

Therefore, we can speedup many heuristic classical algorithms by this procedure. In general, any algorithm that has non-trivial probability of finding a solution can have this probability amplified to nearly 1.

Finally, notice that Grover's algorithm is a special case of amplitude amplification, where $O_\xi = O_x$ and $\mathcal{A} = H^{\otimes n}$.

Chapter 4

Hamiltonian simulation

In the previous chapters, we have only viewed the dynamics of quantum systems from the perspective of unitary transformations. Indeed, apart from measurement, we have considered that the only way a quantum state can change is by multiplication with a unitary matrix. Examples of this are the 1-qubit, 2-qubits and 3-qubits gates tensored with identities on the other qubits described in the previous algorithms. However, the specific unitary that will actually appear in a given physical system is that which stems from a *Hamiltonian* of the system, i.e. an observable that we frequently denote by H and which corresponds to the total energy of the system.

In this case, the expectation value $\langle \psi, H|\psi\rangle$ is called the energy of the state $|\psi\rangle$, and this energy will typically be the sum of several different terms, corresponding to kinetic energy, potential energy, etc. We will also frequently work in the case in which it is the sum of many local terms such that each acts on only a few of the particles (qubits) of the system, for example, if all interactions are between pairs of particles (in which case it is called nearest neighbour interactions).

Overall, the Hamiltonian of a system describes its physical characteristics. These do not determine the initial state of the system, which we denote by $|\psi(0)\rangle$, but they do determine the evolution of the state in time. More specifically, they provide information on the state $|\psi(t)\rangle$ as a function of the t, given initial state $|\psi(0)\rangle$. This evolution is governed by the Schrödinger equation, given by

$$i\frac{h}{2\pi}\frac{d\,|\psi(t)\rangle}{dt} = H\,|\psi(t)\rangle\ .$$

From now on, $\frac{h}{2\pi}$ will be set to 1 choosing the appropriate units so that we can consider a simplified version of the Schrödinger equation. In general, we can consider the case of time-dependent Hamiltonians, i.e. when $H(t)$ changes with time. However, in this course, we will restrict to the time-independent case for simplicity. Then, we can solve the previous equation considering an initial state $|\psi(0)\rangle$, in which case the solution to this differential equation is the following unitary evolution of the state:

$$|\psi(t)\rangle = U\,|\psi(0)\rangle\ , \qquad \text{for } U = \mathrm{e}^{-iHt}\ .$$

Therefore, if we consider an evolution in n time steps of time t, this is the same as applying the unitary matrix e^{-iHt} n times. This provides a continuous-time evolution, where we find the first fundamental difference with the previous chapter, in which the circuit models with elementary gates were discrete in time.

In this chapter, we will focus on the study of the so-called *quantum Hamiltonian simulation*. This is of great importance in areas like quantum chemistry, e.g. to study of properties of molecules and their interaction, as well as in material sciences. In these fields, it is often of great importance to determine how a quantum system will evolve with time, given some initial state. Generally, this

is hard to do on classical computers, since the number of parameters we require for these studies is exponential in the number of particles. However, a quantum computer, as we have seen in the previous chapters, is supposed to provide an exponential speed-up in the number of operations, and thus it should be able to efficiently simulate every efficient quantum process, in the same way, that a classical universal Turing machine can efficiently simulate other (classical) physical processes. As a side comment, let us mention that it is actually possible to classically simulate quantum computers with a polynomial amount of space, but the best methods known to date still use an exponential amount of time. Actually, if there is a way to show that factoring a large integer into its prime components is hard classically, a fact that is widely believed, then this would constitute proof that it is impossible to simulate a quantum computer in polynomial time on a classical computer. As already discussed in previous chapters, this is the main motivation behind the introduction and invention of quantum computers.

From a practical point of view, if we want to show that quantum computers can perform tasks that classical computers are unable to do in polynomial time, we need in particular some methods to efficiently implement the unitary evolution that is induced by a given Hamiltonian. More specifically, we need methods to implement $U = \mathrm{e}^{-iHt}$ as a quantum circuit of gates and to apply this to a given initial state $|\psi(0)\rangle$. We do not ever need to obtain precisely the same value, but to approximate it well enough in norm. This is known as the problem of *Hamiltonian simulation*, and this is going to be the main content of this chapter.

In the next few pages, we will cover several methods for Hamiltonian simulation. In order to simplify the calculations, we will ignore the negative sign in front of the Hamiltonian, in the exponential, and we will instead implement $U = \mathrm{e}^{iHt}$. We will assume throughout the whole chapter that our quantum system consists of n qubits, which is a reasonable assumption, as some physical systems of relevance, like electron spins, naturally correspond to qubits. More complicated Hilbert spaces can be encoded (approximately) in binary form (as in the previous chapter we did with most of the algorithms) to reduce them to the case of qubits. The contents of this chapter will be largely based on the notes [28]. These notes are only of internal use, and for more information on the topic of this chapter, we refer the reader to [28], where all these approaches were originally explained.

4.1 Lie-Suzuki-Trotter methods

Consider an n-qubit Hamiltonian. Then, it corresponds to a $2^n \times 2^n$ matrix. Note that in the field of Hamiltonian simulation, we are dealing with very structured Hamiltonians that have a much shorter classical description. More specifically, all examples studied from now on will be of the form $H = \sum_{j=1}^{m} H_j$, where m is not excessively large (usually, we will take it to be, at most, of order $O(\log n)$) and each H_j acts only on a few of the n qubits. For the time being, let us restrict ourselves to the setting of nearest-neighbour interactions, i.e.

$$H = \sum_{j=1}^{m-1} h_{j,j+1}\,,$$

where the interactions are now written as $h_{j,j+1}$ to denote that they only act on two consecutive qubits, and thus the Hamiltonian is 2-local. That means, in particular, that our huge Hamiltonian H can be seen as the sum of many 4×4 matrices (tensored with identities in all other qubits). In particular, note that, if we fix t, for any spin j, the unitary $\mathrm{e}^{ih_{j,j+1}t}$ is nothing but a 2-qubit gate, which acts as an identity on the other $n-2$ qubits. Because of the universality results presented in this course, the idea is that now this gate can be constructed from CNOTs and single-qubit gates. Let us denote $H_j := h_{j,j+1}$ hereafter for simplicity in the notation.

The main goal of this section is to implement $U = \mathrm{e}^{iHt} = \mathrm{e}^{i\sum_j H_j t}$. It is clear that it would be much easier for us if we could write the latter term as $\prod_{j=1}^{m-1} \mathrm{e}^{iH_j t}$, since this would just be a product of $m-1$ 2-qubit unitaries. However, this is not going to be possible in general. In some particular cases, like when all H_j are diagonal, we can do that, but, in more generality, the matrix exponential does not work this way: For two given matrices A and B, $\mathrm{e}^{A+B} \neq \mathrm{e}^A\,\mathrm{e}^B$. What we can write, due to the *Baker-Campbell-Hausdorff formula* is that

$$\mathrm{e}^A\,\mathrm{e}^B = \mathrm{e}^{A+B+\frac{1}{2}[A,B]+\frac{1}{12}[A,[A,B]]-\frac{1}{12}[B,[A,B]]+\cdots}\,,$$

followed by infinitely many more commutators in A and B. Thus, from here it is easy to conclude that $\mathrm{e}^{A+B} = \mathrm{e}^A\,\mathrm{e}^B$ if, and only if, A and B commute. Nevertheless, the Lie-Suzuki-Trotter decomposition provides us with a way to deal with the exponential of the sum of many terms, in an approximate way. The idea behind it is that, if both A and B have small operator norms, then e^{A+B} and $\mathrm{e}^A\,\mathrm{e}^B$ can be seen to be approximately equal, in the sense that $\mathrm{e}^{A+B} = \mathrm{e}^A\,\mathrm{e}^B + \xi$, where the error term ξ is of order $O(\|A\|\,\|B\|)$. A non-rigorous way to see why this works, but convincing enough, is by doing the first-order Taylor expansion of the exponentials:

$$\mathrm{e}^A\,\mathrm{e}^B - \mathrm{e}^{A+B} \approx (I+A)(I+B) - (I+A+B) = AB\,.$$

Let us formalize the whole approximation process now. The first thing we need to do is construct a systematic way of approximating a unitary U by a circuit $\widetilde{U}$ of 2-qubit gates. Beforehand, let us assume that each of the terms H_j has an operator norm upper bounded by 1, and consider for simplicity the case $H = H_1 + H_2$. Then, since H_1 and H_2 do not commute in general, we can force them to infinitesimally commute, in the following form: First, we perform a small piece of H_1, then a small piece of H_2, then again a small piece of H_1, etc. More specifically, for any integer $r \geq 1$, we have:

$$U = \mathrm{e}^{iHt} = (\mathrm{e}^{iHt/r})^r = (\mathrm{e}^{iH_1 t/r + iH_2 t/r})^r = (\mathrm{e}^{iH_1 t/r}\,\mathrm{e}^{iH_2 t/r} + \xi)^r\,.$$

Note that the error term has norm

$$\|\xi\| = O(\|iH_1 t/r\|\,\|iH_2 t/r\|) = O(\|H_1\|\,\|H_2\|\,t^2/r^2) = O(t^2/r^2)\,.$$

Then, the approximating circuit will be

$$\widetilde{U} = (\mathrm{e}^{iH_1 t/r}\,\mathrm{e}^{iH_2 t/r})^r\,,$$

Now, to compute specifically the errors in this approximation, we first need to prove that error in products of unitaries add at most linearly. Indeed, assume that we have two unitaries U_1, U_2 which are ε-approximated by $\widetilde{U}_1$ and $\widetilde{U}_2$, respectively. Then, for the product we have:

$$\left\|U_1U_2 - \widetilde{U}_1\widetilde{U}_2\right\| \leq \left\|U_1U_2 - \widetilde{U}_1U_2\right\| + \left\|\widetilde{U}_1U_2 - \widetilde{U}_1\widetilde{U}_2\right\| \leq \|U_1 - U_2\| + \left\|\widetilde{U}_1 - \widetilde{U}_2\right\| \leq 2\varepsilon\,.$$

Then, in general:

$$\begin{aligned}\left\|U - \widetilde{U}\right\| &= \left\|(\mathrm{e}^{i(H_1+H_2)t/r})^r - (\mathrm{e}^{iH_1t/r}\,\mathrm{e}^{iH_2t/r})^r\right\| \\ &\leq r\left\|\mathrm{e}^{i(H_1+H_2)t/r} - \mathrm{e}^{iH_1t/r}\,\mathrm{e}^{iH_2t/r}\right\| \\ &\leq rO(\|iH_1t/r\|\,\|iH_2t/r\|) \\ &\leq O(t^2/r)\,.\end{aligned}$$

Therefore, if we want to make our approximation up to a small error ε, it is enough to choose $r = O(t^2/\varepsilon)$.

Next, this idea can be extended to the general case of many more terms in our Hamiltonian. In general, we will have for a Hamiltonian $H = \sum_{j=1}^{m} H_j$,

$$U = \mathrm{e}^{iHt} = (\mathrm{e}^{iHt/r})^r = (\mathrm{e}^{iH_1 t/r + \ldots iH_m t/r})^r = (\mathrm{e}^{iH_1 t/r} \ldots \mathrm{e}^{iH_m t/r} + \xi)^r \,,$$

Thus, repeating the same calculations as above, it is not difficult to show that

$$\|\xi\| = O(m^2 t^2 / r^2) \,,$$

and thus

$$\left\| U - \widetilde{U} \right\| = \left\| (\mathrm{e}^{iH_1 t/r + \ldots iH_m t/r})^r - (\mathrm{e}^{iH_1 t/r} \ldots \mathrm{e}^{iH_m t/r})^r \right\| \leq r \, \|\xi\| = O(m^2 t^2 / r) \,.$$

This constitutes the so-called *first-order Lie-Suzuki-Trotter approach to Hamiltonian simulation*, and it was originally presented in [13]. The number of gates of the circuit $\widetilde{U}$ depends quadratically on the time t for which we want to simulate the evolution, and this is far from optimal. One can do more complicated higher-order Lie-Suzuki-Trotter decompositions that make this dependence on time much better, even nearly linear, but this is much more involved and out of the scope of this course. What we are going to do in the next sections is to describe instead two methods that present linear dependence with time t. Additionally, here we showed that the number of gates of the approximating circuit $\widetilde{U}$ depends polynomially on the error ε. This is another point in which we can notably improve, as we will see in the next few pages. However, even though the other two methods will be much better in their dependence on t and ε, let us mention here that Trotter methods are actually considered to be quite competitive and they are much simple than the ones we will show now.

4.2 Linear combination of unitaries (LCU) methods

In this section, we will describe a method for Hamiltonian simulation whose complexity depends linearly on the time t for which we want to evolve the state, and only logarithmically on the desired error ε. The general problem we address here is the same as in the previous section: Consider a matrix M of dimension $2^n \times 2^n$ and $|\psi\rangle$ an n-qubit state. Then, the main goal is to prepare

$$\frac{M \, |\psi\rangle}{\|M \, |\psi\rangle\|} \,.$$

For this, we do not require for M to be a unitary, but consider for the time being that M is a linear combination of unitaries:

$$M = \sum_{j=1}^{m} \alpha_j V_j \,,$$

with the α_j being nonnegative reals. Note that we can assume they are reals, since we can always absorb complex phases into the unitaries. Let us denote $\|\alpha\|_1 = \sum_j \alpha_j$, and consider the following unitary map, which acts in $\lceil \log m \rceil$ qubits (we take $m = 2^n$ for simplicity, so that this holds):

$$W : |0\rangle \mapsto \frac{1}{\sqrt{\|\alpha\|_1}} \sum_j \sqrt{\alpha_j} \, |j\rangle \,.$$

Now, assume that we are in the easier scenario in which all V_j are unitaries like the ones we are more familiar with, i.e. like 2-qubit gates tensored with identity on the other $n-2$ qubits. We further need to assume that we can implement these unitaries in a controlled way, meaning that we have access to a map $V = \sum_{j=1}^{m} |j\rangle\langle j| \otimes V_j$, which maps $|j\rangle \, |\phi\rangle \mapsto |j\rangle \, V_j \, |\phi\rangle$. Therefore, the first register can be interpreted as nothing but selecting which unitary we apply to the second register.

Now, the purpose of the first part of this section is to use V and W to implement M on a given state $|\psi\rangle$. For that, we consider the following algorithm:

1. We start with a two-register state $|0\rangle |\psi\rangle$, where the first register has $\lceil \log m \rceil$ qubits.
2. We apply W to the first register.
3. We apply V to the whole state.
4. We apply W^{-1} to the first register.

Then, we can see that, after the algorithm, the resulting state is of the following form:

$$\frac{1}{\|\alpha\|_1} |0\rangle M |\psi\rangle + \sqrt{1 - \frac{\|M|\psi\rangle\|^2}{\|\alpha\|_1^2}} |\phi\rangle \ ,$$

where the last state, $|\phi\rangle$ is a normalized state whose specific form we do not need to know for now. The only important property to know of $|\psi\rangle$ is that it has no support on basis states where the first register is $|0\rangle$, and that is has norm 1.

Assume now that we measure the first register. Then, the probability of obtaining the outcome 0 is $p = \|M|\psi\rangle\|^2 / \|\alpha\|_1^2$, and the second register collapses to $M|\psi\rangle / \|M|\psi\rangle\|$ and we are done.

4.3 Transformation via block-encoded matrices

Now, we describe another approach which of great generality. The idea behind it is, as the title suggests, to encode a given matrix M in a larger one U which incidentally is going to turn to be a unitary.

Consider M a matrix on n qubits (which means its dimension is $2^n \times 2^n$) with operator norm bounded by 1. Then, we can consider the following matrix of dimension $2^{n+1} \times 2^{n+1}$:

$$U = \begin{pmatrix} M & \cdot \\ \cdot & \cdot \end{pmatrix} .$$

All the unspecified entries in the matrix are $2^n \times 2^n$-dimensional matrices, and the only constraint we put in them is that they complement M such that U is a unitary. This procedure is called *block-encoding* and

Chapter 5

Quantum noise and open quantum systems

In the previous chapters, we have only considered the dynamics of closed quantum systems, i.e. that of quantum systems that do not suffer any unwanted interactions with the outside world. These ideal systems can yield fascinating conclusions about the information processing tasks which can be accomplished by using them, but in the real world there are no perfectly closed systems, except perhaps the universe itself as a whole. In realistic situations, real systems are going to suffer from unwanted interactions with the outside world, which show up as noise in quantum information processing systems. Therefore, we need to understand and control such noise processes in order to build useful quantum information processing systems. This will be the central topic of this chapter, in which we will describe the quantum operations (a.k.a. quantum channels) formalism, a powerful set of tools enabling us to describe quantum noise and the behaviour of open quantum systems.

5.1 Quantum channels

In this section, 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 [27], 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 subsection of preliminaries.

5.1.1 Preliminaries

We start by recalling the notion of the Schmidt decomposition of a state, which has already appeared in previous chapters, but which becomes of great relevance in the current one. Let $\mathcal{H}_A, \mathcal{H}_B$ finite dimensional Hilbert spaces. We consider the composite Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$.

Theorem 5.1.1 — Schmidt decomposition. Let $|\phi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$, then there exists a set of $\{|e_i\rangle\} \subset \mathcal{H}_A$, $\{|f_i\rangle\} \subset \mathcal{H}_B$ and $\lambda_i \geq 0 \ \forall i$, such that

$$|\phi\rangle = \sum_i \sqrt{\lambda_i} \, |e_i\rangle \otimes |f_i\rangle \tag{5.1.1}$$

Proof. We find in general

$$|\phi\rangle = \sum_{k,l} \beta_{kl} |\psi_k\rangle \otimes |\varphi_l\rangle \tag{5.1.2}$$

for orthonormal basis $\{|\psi_k\rangle\} \subset \mathcal{H}_A$, $\{|\varphi_l\rangle\} \subset \mathcal{H}_B$. Through the singular value decomposition, we can decompose the matrix

$$M = (\beta_{kl})_{kl} = U \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V \tag{5.1.3}$$

with Σ an $m \times m$ matrix and 0 a $(n-m) \times m$ matrix and $\Sigma \geq 0$ and hence

$$\beta_{kl} = \sum_i U_{ki} s_i V_{il} \tag{5.1.4}$$

which gives immediately

$$|\phi\rangle = \sum_i s_i \underbrace{\Big(\sum_k U_{ki} |\varphi_k\rangle \Big)}_{|e_i\rangle} \otimes \underbrace{\Big(\sum_l V_{il} |\varphi_l\rangle \Big)}_{|f_i\rangle} \tag{5.1.5}$$

■

The $\sqrt{\lambda_i} > 0$ in the Schmidt decomposition are the *Schmidt coefficients*, and the number of positive $\sqrt{\lambda_i}$ is the *Schmidt rank* of the state. Moreover, a pure state is called *maximally entangled* if all Schmidt coefficients are $1/\sqrt{d}$.

Let us know also recall the notion of partial trace and some fundamental properties that it satisfies.

Definition 5.1.2 — Partial trace. Given a bipartite Hilbert space $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$, the partial trace in B is a linear map

$$\mathrm{Tr}_B : \mathcal{S}(\mathcal{H}_{AB}) \to \mathcal{S}(\mathcal{H}_A), \quad \rho_{AB} \mapsto \rho_A \tag{5.1.6}$$

defined by

$$\mathrm{Tr}[\rho_{AB}(M_A \otimes \mathbb{1}_B)] = \mathrm{Tr}[\rho_A M_A] \qquad \forall M_A \in \mathcal{B}(\mathcal{H}_A). \tag{5.1.7}$$

and extended by linearity.

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.

We have the following properties for the partial trace.

Proposition 5.1.3 Consider $\rho_{AB} \in \mathcal{S}(\mathcal{H}_{AB})$, $\rho_A := \mathrm{Tr}_B[\rho_{AB}]$, then

1. It holds that $\mathrm{Tr}[\rho_A] = \mathrm{Tr}[\rho_{AB}]$.
2. If $\rho_{AB} \geq 0$, then $\rho_A \geq 0$.
3. From the first two we immediately get that if ρ_{AB} density matrix, then ρ_A is a density matrix.
4. We have

$$\langle \varphi_i \mid \rho_A \varphi_i \rangle = \sum_k \langle \varphi_i \otimes \psi_k \mid \rho_{AB} \varphi_i \otimes \psi_k \rangle \tag{5.1.8}$$

5. If $\rho_{AB} = |\phi\rangle\langle\phi|$ with Schmidt decomposition:

$$|\phi\rangle = \sum_i \sqrt{\lambda_i}\, |e_i\rangle \otimes |f_i\rangle \tag{5.1.9}$$

then

$$\rho_A = \sum_i \lambda_i |e_i\rangle\langle e_i| \tag{5.1.10}$$

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. $\langle\psi \mid \rho_A \psi\rangle = \mathrm{Tr}[\rho_A |\psi\rangle\langle\psi|] = \mathrm{Tr}[\rho_{AB}(|\psi\rangle\langle\psi|_A \otimes \mathbb{1}_B)] \geq 0$.

3. The result is immediate.

4. We have

$$\begin{aligned} \langle\varphi_i \mid \rho_A \varphi_i\rangle &= \mathrm{Tr}[\rho_A |\varphi_i\rangle\langle\varphi_i|] = \mathrm{Tr}[\rho_{AB}(|\varphi_i\rangle\langle\varphi_i|] \\ &= \sum_{k,l} \langle\varphi_l \otimes \psi_k \mid \rho_{AB}(|\varphi_i\rangle\langle\varphi_i| \otimes \mathbb{1})\varphi_l \otimes \psi_k\rangle \\ &= \sum_k \langle\varphi_i \otimes \psi_k \mid \rho_{AB}\varphi_i \otimes \psi_k\rangle \ . \end{aligned} \tag{5.1.11}$$

5. $\langle e_i \mid \rho_A e_i\rangle = \sum_k \langle e_i \otimes f_k \mid \phi\rangle \langle\phi \mid e_i \otimes f_k\rangle = \delta_{ij}\sqrt{\lambda_i}\sqrt{\lambda_j}$.

6. For all $X_A \in \mathcal{B}(\mathcal{H}_A)$

$$\begin{aligned} \mathrm{Tr}[\rho_A X_A] &= \mathrm{Tr}[\rho_{AB}(X_A \otimes \mathbb{1}_B)] = \mathrm{Tr}[\tilde{\rho}_A \otimes \tilde{\rho}_B (X_A \otimes \mathbb{1})] \\ &= \mathrm{Tr}[\tilde{\rho}_A X_A]\, \mathrm{Tr}[\tilde{\rho}_B] = \mathrm{Tr}[\tilde{\rho}_A X_A]\,. \end{aligned} \tag{5.1.12}$$

■

Remark 5.1.4 The converse of 5. is called *purification*. Additionally, any mixed state ρ can be viewed as the reduced state of a pure state on a bipartite system where the dimension of the additional system's Hilbert space is $\mathrm{rank}(\rho)$.

5.1.1.1 Measurements on subsystems

Let us describe now how we can measure in subsystems. Let $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$. Then, POVMs in $\mathcal{H}_A$ can be viewed as POVMs in $\mathcal{H}_{AB}$ just by taking the following measurement in the composite system:

$$M_A : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(H_{AB}) \qquad \text{POVM } \{M_A\} \mapsto \text{POVM } \{M_A \otimes \mathbb{1}_B\} \tag{5.1.13}$$

Note that this is consistent with the definition of the reduced state $\rho_A := \mathrm{Tr}_B[\rho_{AB}]$.

Theorem 5.1.5 — 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 projective measurement $(P_x)_{x=1}^m \subseteq \mathcal{B}(\mathcal{H} \otimes \mathbb{C}^m)$ such that

$$\mathrm{Tr}[(\rho \otimes |\psi\rangle\langle\psi|)P_x] \qquad \forall x = 1, \ldots, m \tag{5.1.14}$$

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

$$V^* V = \sum_{x=1}^m M_x = \mathbb{1} \tag{5.1.15}$$

hence V is an isometry and

$$V = U(\mathbb{1} \otimes |\psi\rangle) \tag{5.1.16}$$

for some $U \in \mathcal{B}(\mathcal{H} \otimes \mathbb{C}^m)$ and $|\psi\rangle \in \mathbb{C}^m$. Then,

$$\mathrm{Tr}[\rho M_x] = \mathrm{Tr}[V\rho V^*(\mathbb{1} \otimes \langle x\rangle\, x)] = \mathrm{Tr}[(\rho \otimes |\psi\rangle\langle\psi|)\underbrace{U^*(\mathbb{1} \otimes |x\rangle\langle x|)U}_{P_X}] \tag{5.1.17}$$

■

5.1.2 Equivalent formulations of quantum channels

They are used to describe the evolution or process in quantum systems (QIT). Schematically we can represent them by

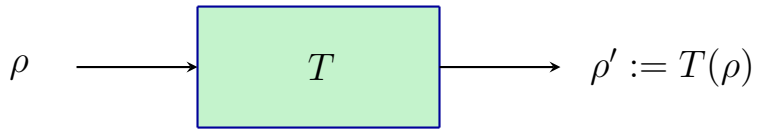

Fig. 5.1: Schematic representation of a channel.

■ **Example 5.1.6** Two prominent examples of quantum channels are the following:

- Closed system evolution:

$$\rho \mapsto U\rho U^* \tag{5.1.18}$$

  with U a unitary.

- Open system evolution

$$\rho \mapsto \mathrm{Tr}_E[U(\rho \otimes \rho_E)U^*] \tag{5.1.19}$$

■

Definition 5.1.7 — 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 5.1.8** The most important basic examples of quantum channels are the following three kinds:

- Unitary evolution $\rho \mapsto U\rho U^*$.
- Adding an ancilla $\rho \mapsto \rho \otimes \rho_E$.
- Partial trace.

Moreover, it can be shown, that any quantum channel is combination of these three. ■

Before presenting the first equivalent reformulation for a quantum channel, let us define the standard maximally entangled state that we will use in the next pages.

Definition 5.1.9 — Maximally entangled state. The maximally entangled state is given by

$$|\phi\rangle = \frac{1}{\sqrt{d}} \sum_{k=1}^{d} |kk\rangle \tag{5.1.20}$$

Now we are in position to introduce that Choi-Jamiolkowski isomorphism, which provides equivalences between quantum channels and some matrices.

Definition 5.1.10 — 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

$$C := (T \otimes \mathbb{1}_d)(|\phi\rangle\langle\phi|) \tag{5.1.21}$$

Remark 5.1.11 C determines T by

$$\langle ij \mid Ckl\rangle = \frac{1}{d} \sum_{m,n=1}^{d} \langle i \mid T(|n\rangle\langle m|)k\rangle \langle j \mid n\rangle \langle m \mid l\rangle = \frac{1}{d} \langle i \mid T(|j\rangle\langle l|)k\rangle \tag{5.1.22}$$

Theorem 5.1.12 — 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-Jamiolkowski matrix of T:

$$C := (T \otimes \mathbb{1}_d)(|\phi\rangle\langle\phi|) \tag{5.1.23}$$

and

$$|\phi\rangle = \frac{1}{\sqrt{d}} \sum_{k=1}^{d} |kk\rangle \tag{5.1.24}$$

the maximally entangled state.

3. Kraus decomposition:

$$T(\rho) = \sum_{k=1}^{dd'} A_k \rho A_k^* \tag{5.1.25}$$

with

$$\sum_{k=1}^{dd'} A_k^* A_k = \mathbb{1} \tag{5.1.26}$$

4. Stinespring dilation:

$$T(\rho) = \mathrm{Tr}_2[U(\rho \otimes |\psi\rangle\langle\psi|)U^*] \tag{5.1.27}$$

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

$$\begin{aligned}
\mathrm{Tr}_1[C] &= \frac{1}{d}\sum_{n,m=1}^{d} \mathrm{Tr}[T(|n\rangle\langle m|)]|n\rangle\langle m| \\
&= \frac{1}{d}\sum_{n,m=1}^{d} \underbrace{\mathrm{Tr}[|n\rangle\langle m|]}_{\delta_{nm}} |n\rangle\langle m| \\
&= \frac{1}{d}\sum_{n,m} \delta_{nm}|n\rangle\langle m| \\
&= \frac{1}{d}\sum_{n=1}^{d} |n\rangle\langle n| \\
&= \mathbb{1}
\end{aligned} \tag{5.1.28}$$

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 $C \geq 0$,

$$\begin{aligned}
C &= \sum_{k=1}^{dd'} |\psi_k\rangle\langle\psi_k| \\
&= \sum_{k=1}^{dd'} \underbrace{(A_k \otimes \mathbb{1})}_{\mathbb{1}\otimes A^T} |\phi\rangle\langle\phi| \underbrace{(A_k^* \otimes \mathbb{1})}_{\mathbb{1}\otimes \overline{A}} \\
&= (T \otimes \mathbb{1})(|\phi\rangle\langle\phi|)
\end{aligned} \tag{5.1.29}$$

and

$$\begin{aligned}
\frac{\mathbb{1}}{d} = \mathrm{Tr}_1[C] &= \sum_{n=1}^{d} \langle n, Cn\rangle \\
&= \frac{1}{d}\sum_{n=1}^{d}\sum_{k=1}^{d'} A_k^T |n\rangle\langle n| \overline{A}_k \\
&= \frac{1}{d}\sum_{k=1}^{d'} A_k^T \overline{A}_k \\
&= \frac{1}{d}\sum_{k=1}^{d'} \tilde{A}_k^* \tilde{A}_k
\end{aligned} \tag{5.1.30}$$

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 $\mathbb{C}^{dd'}$.

$$\begin{aligned}
\mathrm{Tr}_E[V\rho V^*] &= \sum_{kl} A_k \rho A_l^* \underbrace{\mathrm{Tr}[|k\rangle\langle l|]}_{\delta_{kl}} \\
&= \sum_{k} A_k \rho A_k^* \\
&= T(\rho)
\end{aligned} \tag{5.1.31}$$

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

$$T(\rho) = \mathrm{Tr}_E[U(\rho \otimes |\psi\rangle\langle\psi|)U^*] \tag{5.1.32}$$

implies T being a quantum channel. We set

$$\rho \mapsto \rho \otimes |\psi\rangle\langle\psi| \mapsto U(\rho \otimes |\psi\rangle\langle\psi|)U^* \mapsto \mathrm{Tr}_E[U(\rho \otimes |\psi\rangle\langle\psi|)U^*] \tag{5.1.33}$$

The above mappings are all quantum channels and, hence, their composition is a quantum channel as well.

■

Remark 5.1.13

- 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 (with the Hilbert-Schmidt product).

$$\tau := (T \otimes \mathbb{1}_d)(|\phi\rangle\langle\phi|) \tag{5.1.34}$$

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, there exists a unitary map such that

$$K_j = \sum_l U_{kl}\tilde{K}_l \tag{5.1.35}$$

(the smallest set is complemented with zeros).

5.1.3 Examples of quantum channels

5.1.3.1 Depolarizing channel

In two dimensions we can get 3 kinds of errors. 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

$$U_{A\to AE} : |\psi\rangle_A \mapsto \sqrt{1-p}\,|\psi\rangle_A \otimes |0\rangle_E + \sqrt{\frac{p}{3}}\,(X|\psi\rangle_A \otimes |1\rangle_E + Y|\psi\rangle_A \otimes |2\rangle_E + Z|\psi\rangle_A \otimes |3\rangle_E) \tag{5.1.36}$$

Or in the operator representation

$$M_a := {}_E\langle a|\, U_{A\to AE} \tag{5.1.37}$$

with ${}_E\langle a| \in \{{}_E\langle 0|, {}_E\langle 1|, {}_E\langle 2|, {}_E\langle 3|\}$ and

$$M_0 = \sqrt{1-p}\,\mathbb{1},\ M_1 = \sqrt{\frac{p}{3}}X,\ M_2 = \sqrt{\frac{p}{3}}Y,\ M_3 = \sqrt{\frac{p}{3}}Z \tag{5.1.38}$$

It is straightforward to see that

$$\sum M_a^* M_a = ((1-p) + \frac{p}{3} + \frac{p}{3} + \frac{p}{3})\,\mathbb{1} = \mathbb{1}\,. \tag{5.1.39}$$

The depolarizing channel is given by

$$\rho \mapsto \rho' = (1-p)\rho + \frac{p}{3}(X\rho X + Y\rho Y + Z\rho Y) \tag{5.1.40}$$

In general for an arbitrary dimension $D \in \mathbb{N}$, the depolarizing channel becomes

$$\rho \mapsto (1-p)\rho + p\sigma \tag{5.1.41}$$

with σ usually taken to be $\frac{\mathbb{1}}{d}$.

5.1.3.2 Phase damping channel

The phase damping channel in operator representation is given by

$$\rho \in \mathbb{C}^{2\times 2} \qquad \rho = \begin{pmatrix} \rho_{00} & \rho_{01} \\ \rho_{10} & \rho_{11} \end{pmatrix} \mapsto \begin{pmatrix} \rho_{00} & (1-p)\rho_{01} \\ (1-p)\rho_{10} & \rho_{11} \end{pmatrix} \tag{5.1.42}$$

In state representation, we get for the ONB of the environment $\{|0\rangle_E, |1\rangle_E, |2\rangle_E\}$

$$\begin{array}{rcl} |0\rangle_A & \mapsto & \sqrt{1-p}\,|0\rangle_A \otimes |0\rangle_E + \sqrt{p}\,|0\rangle_A \otimes |1\rangle_E \\ |1\rangle_A & \mapsto & \sqrt{1-p}\,|1\rangle_A \otimes |0\rangle_E + \sqrt{p}\,|1\rangle_A \otimes |2\rangle_E \end{array} . \tag{5.1.43}$$

In Kraus operators, we evaluate

$$\begin{gathered} A_0 = \sqrt{1-p}\,\mathbb{1} = \begin{pmatrix} \sqrt{1-p} & 0 \\ 0 & \sqrt{1-p} \end{pmatrix}, \quad A_1 = \sqrt{p}|0\rangle\langle 0| = \begin{pmatrix} \sqrt{p} & 0 \\ 0 & 0 \end{pmatrix} \\ A_2 = \sqrt{p}|1\rangle\langle 1| = \begin{pmatrix} 0 & 0 \\ 0 & \sqrt{p} \end{pmatrix} \end{gathered} \tag{5.1.44}$$

Note that $A_0^2 + A_1^2 + A_2^2 = \mathbb{1}$. Moreover,

$$T(\rho) = \left(1 - \frac{1}{2}p\right)\rho + \frac{1}{2}pZ\rho Z \tag{5.1.45}$$

5.1.3.3 Entanglement breaking channels

The third family of examples we present here is that of entanglement breaking channels.

Definition 5.1.14 — 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

$$\rho = \sum_i \lambda_i \rho_i^A \otimes \rho_i^B . \tag{5.1.46}$$

Otherwise it is entangled.

Definition 5.1.15 — 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

$$T(\rho) = \sum_x \mathrm{Tr}[M_x\rho]\rho_x . \tag{5.1.47}$$

5.2 Open system representation

5.2.1 Partial order of completely positive 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 $\tau_2 \geq \tau_1$ for τ_x the Choi matrix associated to the map T_x for $x = 1, 2$, i.e. $\tau_2 - \tau_1$ positive semi-definite for $\tau_x := (T_x \otimes \mathbb{1})(|\phi\rangle\langle\phi|)$.

Theorem 5.2.1 — Relation CP maps. Consider for $i = 1, 2$ the completely positive maps given by:

$$T_i : \mathbb{C}^{d' \times d'} \to \mathbb{C}^{d \times d} \,. \tag{5.2.1}$$

Assume that $T_2 \geq T_1$. If for $i = 1, 2$

$$V_i : \mathbb{C}^d \to \mathbb{C}^{d'} \otimes \mathbb{C}^{r_i} \tag{5.2.2}$$

provide Stinespring dilations for T_i $[T_i(A) = V_i^*(A \otimes \mathbb{1}_{r_i})V_i]$, then there is a contraction,

$$C : \mathbb{C}^{r_2} \to \mathbb{C}^{r_1} \quad \text{such that} \quad V_1 = (\mathbb{1}_{d'} \otimes C)V_2. \tag{5.2.3}$$

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

$$W_i := (\mathbb{1}_{r_i} \otimes \langle\phi|)(V_i \otimes \mathbb{1}_{d'}) \in \mathcal{B}(\mathbb{C}^d \otimes \mathbb{C}^{d'} \otimes \mathbb{C}^{r_i}), \tag{5.2.4}$$

we find that $\forall\, |\psi\rangle \in \mathbb{C}^d \otimes \mathbb{C}^{d'}$

$$\|W_2 |\psi\rangle\|^2 = \langle\psi, \tau_2\psi\rangle \geq \langle\varphi, \tau_1\varphi\rangle = \|W_1 |\varphi\rangle\|^2 \,. \tag{5.2.5}$$

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

$$W_1 = CW_2. \tag{5.2.6}$$

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 5.2.2 — 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

$$T(A) = V^*(A \otimes \mathbb{1}_r)V. \tag{5.2.7}$$

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

$$T_i(A) = V^*(A \otimes P_i)V. \tag{5.2.8}$$

Remark 5.2.3 Since $T = \sum_i T_i$ we find that

$$T(A) = \sum_i V^*(A \otimes P_i)V \tag{5.2.9}$$

with $\{P_i\}$ a POVM. To say it with words: We have found a possibility to represent a quantum channel using a POVM.

5.2.2 Instruments

Definition 5.2.4 — 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

$$p_x = \mathrm{Tr}[T_x(\rho)], \quad \rho \mapsto \frac{T_x(\rho)}{p_x}. \tag{5.2.10}$$

This notion is represented in the following figure

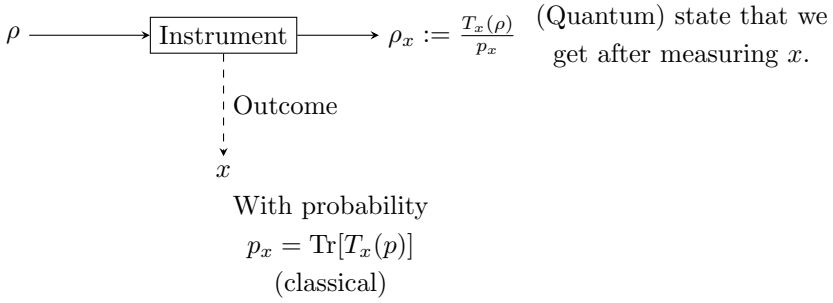

Fig. 5.2: Scheme of an instrument.

In that sense it encompasses the notion of quantum channel and POVMs in the following way

$$\begin{cases} \text{Quantum channel:} & \text{Ignore the measurement outcome} \\ & \rho \mapsto \sum_x p_x \rho_x = \sum_x T_x(\rho) =: T \\ \text{POVM:} & \text{Ignore the quantum system} \\ & p_x = \mathrm{Tr}[T_x(\rho)] = \mathrm{Tr}[T_x(\rho)\,\mathbb{1}] = \mathrm{Tr}[\rho T_x^*(\mathbb{1})] =: \mathrm{Tr}[\rho M_x] \\ & \{M_x\}_x \text{ is a POVM} \end{cases} \tag{5.2.11}$$

Remark 5.2.5 Instruments can be viewed as special case of quantum channels by assigning to them

$$\rho \mapsto \sum_{x \in X} T_x(\rho) \otimes |x\rangle\langle x| \tag{5.2.12}$$

with $\{|x\rangle\}$ an orthonormal basis.

Theorem 5.2.6 — No information without disturbance. Consider an instrument $\{T_x\}_{x\in X}$ such that $\forall \rho$

$$\rho \mapsto \sum_x p_x \rho_x = \rho \tag{5.2.13}$$

Then, ρ_x is independent of ρ, i.e.

$$\mathrm{Tr}[T_x(\rho)] = \mathrm{Tr}[T_x(\rho')] \tag{5.2.14}$$

for all ρ, ρ' density operators.

Proof. Based on the Choi-Jamiolkowski representation for channels. Since for all density operators ρ

$$\rho = \sum_x p_x \rho_x = \sum_x T_x(\rho). \tag{5.2.15}$$

This means that $\sum_x T_x = \mathbb{1}$. This further gives us that also the Choi Jamiolkowski matrices coincide.

$$\begin{aligned} T_x \mapsto \tau_x &:= (T_x \otimes \mathbb{1})(|\phi\rangle\langle\phi|) \\ \sum_x T_x \mapsto \sum_x \tau_x &= (\mathbb{1} \otimes \mathbb{1})(|\phi\rangle\langle\phi|) = |\phi\rangle\langle\phi| \end{aligned} \tag{5.2.16}$$

Moreover, since $\tau_x \geq 0 \ \forall x$, we find that there exists $q_x \geq 0$ such that

$$\tau_x = q_x |\phi\rangle\langle\phi|, \tag{5.2.17}$$

with $\sum_x q_x = 1$ ■

Proposition 5.2.7 — Quantum steering. Let $\rho \in \mathcal{B}(\mathcal{H}_A)$ density operator with purification

$$|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B \tag{5.2.18}$$

(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 $\{T_i\}_i$

$$T_i : \mathcal{B}(\mathcal{H}_B) \to \mathcal{B}(\mathcal{H}_B) \tag{5.2.19}$$

such that

$$\lambda_i \rho_i = \mathrm{Tr}_B[(\mathbb{1} \otimes T_i)(|\psi\rangle\langle\varphi|] \tag{5.2.20}$$

Proof. (sketch). The idea of the proof is just to form Schmidt decompositions of $|\psi\rangle$ and applying the transposition map. ■

5.2.3 Open system representation

In this section, we present the mathematical formulations used to model open quantum systems. Let us recall the reader that, given a quantum channel T, we denote by T^* its dual map, given by $\mathrm{Tr}[XT(Y)] = \mathrm{Tr}[T^*(X)Y]$ for every $X, Y \in \mathcal{B}(\mathcal{H})$.

Theorem 5.2.8

- Let $T : \mathbb{C}^{d\times d} \to \mathbb{C}^{d'\times d'}$ be a CPTP map. 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 $\forall \rho$

$$T(\rho) = \mathrm{Tr}_E[U(\rho \otimes |\varphi\rangle\langle\varphi|)U^*], \tag{5.2.21}$$

  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

$$V : \mathbb{C}^d \to \mathbb{C}^{d'} \otimes \mathbb{C}^r \tag{5.2.22}$$

  with $r \geq \mathrm{rank}(\tau)$, $\forall A \in \mathbb{C}^{d'\times d'}$,

$$T^*(A) = V^*(A \otimes \mathbb{1}_r)V \tag{5.2.23}$$

- If $T = \sum_i T_i$ can be decomposed into CPTP maps, there exists a POVM $\{P_i\}$ such that

$$T_i(\rho) = \mathrm{Tr}_E[(P_i \otimes \mathbb{1}_{d'})U(\rho \otimes |\phi\rangle\langle\phi|)U^*] \tag{5.2.24}$$

  If k_i is the Kraus rank associated to T_i then $k_i \leq \mathrm{rank}(P_i)$ 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:

$$|\psi\rangle := (\mathbb{1}_d \otimes U)(|\Omega\rangle \otimes |\varphi\rangle) \tag{5.2.25}$$

$\tau = \mathrm{Tr}_E[|\psi\rangle\langle\varphi|]$. The decomposition of $T = \sum_i T_i$ gives us $\tau = \sum_i \tau_i$. We conclude by applying Quantum Steering. ■

Remark 5.2.9 V is an isometry ($V^*V = \mathbb{1}_d$) if, and only if, T is trace preserving.

5.3 Quantum Many Body Systems

5.3.1 Simplified Approach to the 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

  $$t \mapsto \rho(t) \tag{5.3.1}$$

  $\rho(t+dt)$ 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:

  $$\rho_{SE} = \rho_S \otimes \rho_E \overset{\text{QC}}{\mapsto} U_{SE}(\rho_S \otimes \rho_E)U_{SE}^* \overset{\text{QC}}{\mapsto} \mathrm{Tr}_E[U_{SE}(\rho_S \otimes \rho_E)U_{SE}^*] \tag{5.3.2}$$

  This means

  $$\rho_S \mapsto \rho_S' := \mathrm{Tr}_E[U_{SE}(\rho_S \otimes \rho_E)U_{SE}^*] \tag{5.3.3}$$

  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

  $$\text{"}\underbrace{\rho_S(t) \otimes \rho_E}_{\rho_{SE}(t)} \overset{dt}{\mapsto} \rho_{SE}(t+dt) = \rho_S(t+dt) \otimes \rho_E\text{"}, \tag{5.3.4}$$

  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),

Definition 5.3.1 A quantum Markov semigroup is a 1-parameter continuous semigroup $\{T_t\}_{t\geq 0}$ of completely positive trace preserving linear maps

$$T_t : \mathcal{S}(\mathcal{H}) \to \mathcal{S}(\mathcal{H})\,. \tag{5.3.5}$$

Because of being a semigroup, it satisfies the following properties:

- $T_0 = \mathbb{1}$

- $T_s \circ T_t = T_{t+s}$

We obtain the generator of the semigroup by differentiation

$$\frac{d}{dt}T_t = \mathcal{L} \circ T_t = T_t \circ \mathcal{L}\,. \tag{5.3.6}$$

This is called the Liouville equation. It gives us the generator of the group

$$T_t = e^{t\,\mathcal{L}} \quad \text{with} \quad \mathcal{L} \text{ the Liouvillian} \tag{5.3.7}$$

With all the above at hand, we will now formalise our two description above:
Closed system: We have

$$\dot{\rho} = -i[H_S, \rho] \tag{5.3.8}$$

with the solution

$$\rho(t) = e^{-iHt}\rho(0)\,. \tag{5.3.9}$$

Open system: The Hamiltonian operator can be decomposed as follows

$$H = H_S + H_E + H_{SE}. \tag{5.3.10}$$

The evolution equation becomes

$$\dot{\rho} = \mathcal{L}[\rho] \tag{5.3.11}$$

and its solution

$$\rho(t) = e^{t\,\mathcal{L}}\rho(0). \tag{5.3.12}$$

In particular, we can infinitesimally write

$$\rho(t + dt) = T_{dt}(\rho(t))\,, \tag{5.3.13}$$

with T_{dt} a quantum channel, which comes from the Markovianity assumption. The previous representation gives us using the Kraus decomposition for the channel T_t

$$\rho(t) = T_t(\rho(0)) = \sum_\mu M_\mu(t)\rho(0)M_\mu(t)^*, \tag{5.3.14}$$

with $M_\mu(t)$ called jump operators, and in general:

$$\rho(t + dt) = T_t(\rho(0)) = \sum_\mu M_\mu(t)\rho(t)M_\mu(t)^* = \rho(t) + O(dt). \tag{5.3.15}$$

Note that we have

$$\rho(dt) = \rho(0) + O(dt) \tag{5.3.16}$$

and moreover, if we only want to retain the terms up to linear order in dt, we can assume without loss of generality:

- $M_0 = \mathbb{1} + O(dt)$
- Others: Order $O(\sqrt{dt})$

meaning

$$\begin{aligned} M_\mu &= \sqrt{dt}L_\mu \\ M_0 &= \mathbb{1} + (-iH_S + K)dt\,, \end{aligned} \tag{5.3.17}$$

with H_S and K Hermitian, and H_S, K and L having zeroth order in dt. We need

$$\mathbb{1} = \sum_\mu M_\mu^*(t)M_\mu(t) = \mathbb{1} + dt\Big(2K + \sum_{\mu>0} L_\mu^* L_\mu\Big) + \dots \tag{5.3.18}$$

from which we get

$$K = -\frac{1}{2}\sum_{\mu>0} L_\mu^* L_\mu \tag{5.3.19}$$

giving us the master equation

$$\dot{\rho} = \mathcal{L}[\rho] = -i[H_S, \rho] + \sum_{\mu>0}\left(L_\mu \rho L_\mu^* - \frac{1}{2}\rho L_\mu^* L_\mu \frac{1}{2} L_\mu^* L_\mu \rho\right) . \tag{5.3.20}$$

5.3.2 Detailed Master Equation

In this section, we present a more detailed approach to the development of the Lindblad form of the quantum master equation. Let us begin by recalling that the Liouville von Neumann equation is given by

$$\dot{\rho} = -\frac{i}{\hbar}[H_S, \rho] . \tag{5.3.21}$$

We can define a superoperator $\mathcal{L}$ such that $\mathcal{L}\rho = -\frac{i}{\hbar}[H_S, \rho]$. We assume that the Hamiltonian is time independent. Thus, we can integrate the Liouville von Neumann equation and obtain

$$\rho(t) = \mathrm{e}^{\mathcal{L}t}\,\rho(0) \equiv T_t\left(\rho(0)\right) .$$

Note that the family of maps $\{T_t\}$, which are in particular quantum channels, constitute a *dynamical quantum evolution.*

5.3.2.1 Dynamical quantum evolution

This is also found in the literature under several other names, such as *quantum dynamical semigroup* or *quantum Markov semigroup.* We are going to describe the interaction between the system of interest and its environment.

As mentioned above the combination of system and environment can be seen as a closed system, and thus it is fully coherent. Thus, it evolves unitarilly, and therefore we can write:

$$\rho_{SE}(t) = U(t)\rho_{SE}(0)U(t)^* ,$$

where ρ_{SE} represents the state of the system and the environment together. Hence, after taking the trace over the environment on both sides, we obtain:

$$\rho_S(t) = \mathrm{Tr}_E[U(t)\rho_{SE}(0)U(t)^*] .$$

Under the simplifications mentioned in the previous section, we can consider that the state between the system and the environment is a product state of the form:

$$\rho_{SE}(0) = \rho_S(0) \otimes \rho_E(0) .$$

We can therefore consider the superoperator representing the dynamical map, only for the system, given by $\{T_t\}_{t\geq 0}$. Let us use the following spectral decomposition for the state of the environment and system at instant 0:

$$\rho_E(0) = \sum_\alpha \lambda_\alpha \left|\psi_\alpha\right\rangle \left\langle\psi_\alpha\right| \; . T_t\rho_S(0) = \sum_{\alpha\beta} W_{\alpha\beta}(t)\rho_S(0)W_{\alpha\beta}^*(t) ,$$

where the operators $W_{\alpha\beta}$ act only on the Hilbert space of the system and are given by

$$W_{\alpha\beta}(t) = \sqrt{\lambda_\beta}\left\langle\psi_\alpha \mid U(t) \mid \psi_\beta\right\rangle .$$

Since the $\{W_{\alpha\beta}(t)\}$ are the Kraus operators of the Kraus decomposition of the quantum channel T_t, we have

$$\sum_{\alpha\beta} W_{\alpha\beta}(t) W^*_{\alpha\beta}(t) = \mathbb{1}_S \,,$$

In many important cases, we can make one further assumption on the dynamical map T_t, known as the Markov approximation. If the correlations in the environment decay much faster than the timescale of the evolution in the system of interest, we may neglect memory effects describing how the system has previously interacted with the environment.

To see an example of this, consider a thermal state of the environment of the form:

$$\rho_E = \sum_n \frac{\mathrm{e}^{-\beta E_n}}{Z} |n\rangle \langle n| \,,$$

where the normalization factor Z is given by $Z = \sum_n \mathrm{e}^{-\beta E_n}$ and it is called the partition function, at inverse temperature β and E_n the energy of the state $|n\rangle$. In this case, if the environment is large and its dynamics is fast enough, any energy exchanged with the system will quickly dissipate away to form a new thermal state with almost exactly the same temperature. Then, from the point of view of the system, the state of the environment will appear to be almost constant all the time. This is a further simplification we will make from now on.

Recall that, since the dynamical evolution is given by a quantum Markov semigroup, the following property is satisfied:

$$T_t \circ T_s = T_{t+s} \quad \forall t, s \geq 0 \,.$$

In particular, we can interpret the constraint on the times being positive as the fact that we can only piecewise propagate the system forward in time, or in other words the inverse of the dynamical evolution does usually not exist. This is in contrast to coherent dynamics, where the unitary evolution operator yields that there is an inverse operation corresponding to negative time arguments. Hence, in contrast to the setting of closed systems, while the dynamical maps of coherent systems form a group, the dynamical maps for open quantum systems only form a semigroup.

The generator of the semigroup is the Liouvillian $\mathcal{L}$, which is a generalization of the superoperator appearing on the right hand side of the Liouville von Neumann equation written above. One important consequence of this generalization is that the von Neumann entropy is no longer a conserved quantity. However, the Liouvillian has to fulfil the property of being the generator of a semigroup completely positive and trace-preserving maps.

5.3.2.2 Most general form of the dynamics

Because of $\mathcal{L}$ being the infinitesimal generator of the semigroup $\{T_t\}$, we can write the dynamical evolution of an open quantum system as an exponential of $\mathcal{L}$,

$$T_t = \mathrm{e}^{\mathcal{L}t} \,.$$

The generator $\mathcal{L}$ reduces to the one mentioned above for the Liouville von Neumann equation in the case of purely coherent dynamics, but in general will have additional incoherent terms in the open system case. First, similarly to the approach on the last section, expanding the exponential around t for a short time dt, we have

$$\rho(t + dt) = T_t \rho(t) = (1 + \mathcal{L}t)\rho(t) + O(t^2) \,,$$

and in the limit $dt \to 0$ we obtain the following first-order differential equation, which is known as the *quantum master equation*:

$$\frac{d}{dt}\rho(t) = \mathcal{L}\rho(t)\,.$$

The purpose of the rest of the current section is to derive an explicit expression for the master equation. For this, we are going to work on an operator basis that we are going to denote by $\{F_i\}$. For these operators, we have the following inner product:

$$\langle F_i, F_j\rangle = \mathrm{Tr}[F_i^* F_j]\,.$$

For convenience, we choose one of the operators to be proportional to the identity. i.e.

$$F_{N^2} = \frac{1}{\sqrt{N}}\,.$$

Additionally, the whole orthonormal set consists of N^2 operators, where N is the Hilbert space dimension. Because of the specific condition we imposed on F_{N^2}, all other operators are traceless. For instance, in the particular case of a two-level system, the remaining operators are proportional to the Pauli matrices.

Let us now express the action of the dynamical evolution at each instant of time using the Kraus decompositions given by these operators:

$$T_t\rho = \sum_{i,j=1}^{N^2} c_{ij}(t) F_i \rho F_j^*\,,$$

where the coefficients $c_{ij}(t)$ are given by

$$c_{ij}(t) = \sum_{\alpha\beta}\langle F_i, W_{\alpha\beta}(t)\rangle\langle F_j, W_{\alpha\beta}(t)\rangle^*\,.$$

These coefficients $c_{ij}(t)$ form a positive matrix C, as can be seen from the fact that they satisfy the following for any N^2-dimensional vector v,

$$v^* C v = \sum_{\alpha\beta}\left|\left\langle \sum_i v_i F_i, W_{\alpha\beta}(t)\right\rangle\right|^2 \geq 0\,.$$

Then, for the master equation, if we separate all terms containing $F_{N^2N^2} = \frac{1}{\sqrt{N}}$, we have

$$\begin{aligned}\mathcal{L}\rho &= \lim_{dt\to 0}\frac{T_{dt}\rho - \rho}{dt} \\ &= \lim_{dt\to 0}\left[\frac{1}{N}\frac{c_{N^2N^2}(dt) - N}{dt}\rho + \frac{1}{\sqrt{N}}\sum_{i=1}^{N^2-1}\left(\frac{c_{iN^2}(dt)}{dt}F_i\rho + \frac{c_{N^2i}(dt)}{dt}\rho F_i^*\right) + \sum_{i,j=1}^{N^2-1}\frac{c_{ij}(dt)}{dt}F_i\rho F_j^*\right]\,.\end{aligned}$$

Let us define now the following quantities:

$$\begin{aligned} a_{ij} &:= \lim_{dt\to 0}\frac{c_{ij}(dt)}{dt}\;\; i,j = 1,\ldots,N^2-1\,,\\ F &:= \frac{1}{\sqrt{N}}\sum_{i=1}^{N^2-1}\lim_{dt\to 0}\frac{c_{iN^2}(dt)}{dt}F_i\,,\\ G &:= \frac{1}{2N}\lim_{dt\to 0}\frac{c_{N^2N^2}(dt) - N}{dt} + \frac{1}{2}(F^* + F)\,,\\ H &:= \frac{1}{2i}(F^* - F)\,.\end{aligned}$$

Looking carefully at these terms, we realize that H is Hermitian and F is not, because the coefficients c_{iN^2} are complex. We can further see that the infinitesimal generator can be simplified in the following form:

$$\mathcal{L}(\rho) = -i[H,\rho] + \{G,\rho\} + \sum_{i,j=1}^{N^2-1} a_{ij} F_i \rho F_j^* \,.$$

Now we can simplify even more this expression by looking carefully at the properties of all these maps. Firstly, since the quantum dynamical evolution is trace preserving, we know that $\mathcal{L}(\rho)$ is traceless, and thus

$$\mathrm{Tr}[\mathcal{L}(\rho)] = \mathrm{Tr}\left[\left(2G + \sum_{i,j=1}^{N^2-1} a_{ij} F_i F_j^*\right)\rho\right] = 0\,,$$

from which we can conclude:

$$G = -\frac{1}{2}\sum_{i,j=1}^{N^2-1} a_{ij} F_i F_j^* \,.$$

Therefore, if we replace this back in the quantum master equation, we obtain:

$$\mathcal{L}(\rho) = -i[H,\rho] + \sum_{i,j=1}^{N^2-1} a_{ij}\left(F_i \rho F_j^* - \frac{1}{2}\{F_j^* F_i, \rho\}\right) .$$

The next condition we need to take into account is that the matrix formed by the coefficients a_{ij} is Hermitian and positive semidefinite, and thus we can diagonalize it in order to obtain the positive eigenvalues γ_i. In this way, we can conclude the most general form to the *Markovian quantum master equation*, given by

$$\mathcal{L}(\rho) = -i[H,\rho] + \sum_{i,j=1}^{N^2-1} \gamma_i\left(A_i \rho A_i^* - \frac{1}{2}\{A_i^* A_i, \rho\}\right) ,$$

where the new operators A_i are given by suitable combinations of the operators F_i, obtained in the diagonalization procedure. This form is known as the Lindblad form, as Lindblad was the first one to show that the generator of a Markovian master equation is of this form.

It is important to remark that the Hamiltonian of the system H_S is contained in the Hermitian operator H, but it doesn't completely coincide with it. On the opposite, the latter actually includes additional terms coming from the interaction with the environment. Moreover, the eigenvalues γ_i correspond to relaxation rates describing the processes of incoherent decay in the system. These will usually result in the system eventually reaching a thermal equilibrium given by a stationary state, characterized by $\mathcal{L}(\rho) = 0$ in general, and by a specific Gibbs state of a local Hamiltonian $\rho = \frac{\mathrm{e}^{-\beta H}}{\mathrm{Tr}[\mathrm{e}^{-\beta H}]}$ in the particular cases studied in this course.

5.3.2.3 Weak-coupling approximation

Now we assume that the system is coupled to a bath, and we have again the Hamiltonian given by

$$H = H_S + H_E + H_{SE}\,, \tag{5.3.22}$$

where H_S denotes the part of the Hamiltonian acting only on the system, H_E the part acting only on the environment (which in this case is a heat bath!) and H_{SE} takes care of the interactions between the two. Let us initially perform a unitary transformation for the whole state, into the interactions picture:

$$\rho = U(t)\rho' U(t)^*\,,$$

where

$$U(t) := \mathrm{e}^{-i(H_S+H_E)t} \,.$$

We can insert this now in the modified Liouville von Neumann equation, obtaining:

$$\frac{d}{dt}(U(t)\rho' U(t)^*) = -i[H, U(t)\rho' U(t)^*] \,,$$

for which we get

$$\frac{d}{dt}\rho = -i[H_{SE}(t), \rho] \,.$$

Therefore, in this particular case, the Hamiltonian modelling the interactions is time dependent, and thus we cannot solve the equation as before. Nevertheless, we can do the following:

$$H_{SE}(t) = U(t)^* H_{SE} U(t) = \mathrm{e}^{i(H_S+H_E)t} \, H_{SE} \, \mathrm{e}^{-i(H_S+H_E)t} \,,$$

and integrate the Liouville von Neumann equation to obtain:

$$\rho(t) = \rho(0) - i \int_0^t ds[H_{SE}(s), \rho(s)] \,.$$

Inserting this again into the Liouville von Neumann equation and tracing out the bath, we have:

$$\frac{d}{dt}\rho_S(t) = -\int_0^t ds \, \mathrm{Tr}_B \left[[H_{SE}(t), [H_{SE}(s), \rho(s)]] \right] ,$$

where we are assuming in particular that the initial state is such that the interaction does not generate any (first-order) dynamics in the bath, or equivalently:

$$\mathrm{Tr}_B \left[[H_{SE}(t), \rho(0)] \right] = 0 \,.$$

Since we want to obtain a closed equation for the state on the system S only, instead of the state everywhere (which is what we have so far), we need to have a further simplification, known as the Born approximation, given by

$$\rho(t) = \rho_S(t) \otimes \rho_E \,. \tag{5.3.23}$$

Thus, we can rewrite the previous integro-differential equation as:

$$\frac{d}{dt}\rho_S(t) = -\int_0^t ds \, \mathrm{Tr}_B \left[[H_{SE}(t), [H_{SE}(s), \rho_S(s) \otimes \rho_E]] \right] ,$$

which is still very difficult to handle. This equation of motion can be brought into a time-local form by replacing $\rho_S(s)$ by $\rho_S(t)$, obtaining thus

$$\frac{d}{dt}\rho_S(t) = -\int_0^t ds \, \mathrm{Tr}_B \left[[H_{SE}(t), [H_{SE}(s), \rho_S(t) \otimes \rho_E]] \right] ,$$

obtaining still a non-Markovian equation, but much more tractable. However, this does not guarantee to conserve positivity of the density matrix because of all the approximations of the latter steps.

5.4 Quantum hypothesis testing

Lets assume the following setting: We are given as set $\rho_1, \ldots, \rho_n \in \mathcal{S}(\mathcal{H})$ of density operators with corresponding probabilities $p_1, \ldots, p_n$ that satisfy $p_x \geq 0 \ \forall x = 1, \ldots, n$ 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

$$\mathcal{P}(M) := \sum_{x=1}^{n} \mathrm{Tr}[M_x \underbrace{p_x \rho_x}_{=:\sigma_x}] \tag{5.4.1}$$

over POVMs $M = (M_x)_{x=1}^n \in \mathcal{M}$. We set

$$\mathcal{P}(\mathcal{M}) := \sup_{M \in \mathcal{M}} P(M). \tag{5.4.2}$$

Definition 5.4.1 — Maximum likelihood measurement. The maximum likelihood measurement is defined as

$$L := \sum_x M_x \sigma_x = \sum_x M_x p_x \rho_x \tag{5.4.3}$$

With this definition, we can write $\forall M$

$$\mathcal{P}(M) = \mathrm{Tr}[L]. \tag{5.4.4}$$

Lemma 5.4.2 — Existence of optimal measurement. The supremum in $\mathcal{P}$ is always attained, i.e. there exists a measurement $\widehat{M}$ such that

$$\mathcal{P}(\mathcal{M}) = \mathcal{P}(\widehat{M}) \tag{5.4.5}$$

Theorem 5.4.3 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
$$\max_{M'=(M'_x)_{x=1}^n} \mathcal{P}(M') = \mathcal{P}(M) \tag{5.4.6}$$
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 5.4.4**

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
$$\max_M P(M) = \sum_i \max_x \underbrace{\langle i, \rho_x i \rangle}_{\lambda_i^x} \tag{5.4.7}$$
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
$$\sum_{x=1}^{n} p_x \rho_x = \frac{\mathbb{1}}{d} \tag{5.4.8}$$
(i.e. in particular $d \leq n$). Let us consider $M_x = \frac{d}{n} \rho_x$ which clearly constitutes $M = (M_x)_{x=1}^n$ 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

$$\begin{aligned} L &= \sum_{x=1}^{n} M_x p_x \rho_x = \frac{d}{n} \sum_{x=1}^{n} \underbrace{\frac{1}{n}}_{p_x} \rho_x^2 \\ &= \frac{d}{n^2} \sum_{x=1}^{n} \rho_x = \frac{d}{n} \underbrace{\sum_{x=1}^{n} p_x \rho_x}_{\mathbb{1}/d} = \frac{\mathbb{1}}{n} \\ &\geq \frac{1}{n} \rho_x = p_x \rho_x \quad \forall x = 1, \ldots, n \end{aligned} \tag{5.4.9}$$

  i.e. $\mathcal{P}(M) = \mathrm{Tr}[L] = \frac{d}{n}$.

■

5.4.1 Binary hypothesis testing

Let ρ_1, ρ_2 be density matrices with a priory probability p and $(1-p)$. Further $M = (M_1, M_2) \cong (P, \mathbb{1} - P)$ 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 ρ_2 the error becomes

$$\mathcal{E}(M) := p \,\mathrm{Tr}[\rho_1(\mathbb{1} - P)] + (1-p)\,\mathrm{Tr}[\rho_2 P] \tag{5.4.10}$$

Remark 5.4.5 It is rather obvious that for

$$\mathcal{P}(M) = p \,\mathrm{Tr}[\rho_1 P] + (1-p)\,\mathrm{Tr}[\rho_2(\mathbb{1} - P)] \tag{5.4.11}$$

we find

$$\mathcal{P}(M) + \mathcal{E}(M) = 1 \tag{5.4.12}$$

Theorem 5.4.6 — Quantum Neyman-Pearson. We find that in the above setting we have the inequality

$$\mathcal{E}(M) \geq \frac{1}{2}(1 - \|p\rho_1 - (1-p)\rho_2\|_1) \tag{5.4.13}$$

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

$$\begin{aligned} \min_M \mathcal{E}(M) &= \min_M \{p \,\mathrm{Tr}[\rho_1(\mathbb{1} - P)] + (1-p)\,\mathrm{Tr}[\rho_2 P]\} \\ &= \min\{p - \mathrm{Tr}[P(p\rho_1 - (1-p)\rho_2)]\} \\ &= p - \max_M \{\mathrm{Tr}[P \underbrace{(p\rho_1 - (1-p)\rho_2)}_{A = A_+ + A_-}]\} \end{aligned} \tag{5.4.14}$$

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

$$\begin{aligned} &= p - \{\mathrm{Tr}[(p\rho_1 - (1-p)\rho_2)_+]\} \\ &= p - \frac{\|p\rho_1 - (1-p)\rho_2\|_1 + \mathrm{Tr}[p\rho_1] - \mathrm{Tr}[(1-p)\rho_2]}{2} \\ &= \frac{1}{2}(1 - \|p\rho_1 - (1-p)\rho_2\|_1) \end{aligned} \tag{5.4.15}$$

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

$$L = Pp\rho_1 + (\mathbb{1} - P)(1-p)\rho_2 \geq \begin{cases} p\rho_1 \\ (1-p)\rho_2 \end{cases} . \tag{5.4.16}$$

■

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

$$\mathcal{E}_m^{opt} = \frac{1}{2}(1 - \left\| p\rho_1^{\otimes m} - (1-p)\rho_2^{\otimes m} \right\|_1) \tag{5.4.17}$$

and $\mathcal{E}_m^{opt}$ decays exponentially with $-\xi_m$, with ξ a rate given as

$$\mathcal{E}_m^{opt} \leq Ke^{-\xi m} \tag{5.4.18}$$

Theorem 5.4.7 If $p \neq 0, 1$, it holds that

$$\xi := \lim_{m\to\infty} \left(-\frac{1}{m} \log(\mathcal{E}_m^{opt}) \right) = -\log(\inf_{s\in[0,1]} \operatorname{Tr}[\rho_1^{1-s}\rho_2^s]) \tag{5.4.19}$$

Proof. For $A, B \in \mathcal{B}(\mathcal{H})$ positive, $\forall s \in [0,1]$

$$\underbrace{\operatorname{Tr}[(A^s - B^s)A^{1-s}]}_{\operatorname{Tr}[A] - \operatorname{Tr}[B^s A^{1-s}]} \leq \operatorname{Tr}[(A-B)_+] \tag{5.4.20}$$

which is a consequence of $z \mapsto z^s$ being operator monotone. Then

$$\begin{aligned} \frac{1}{2}(\operatorname{Tr}[A+B] - \|A-B\|_1) &= \frac{1}{2}(2\operatorname{Tr}[A] - \operatorname{Tr}[A-B] - \operatorname{Tr}[(A-B)_+] + \operatorname{Tr}[(A-B)_-]) \\ &= \operatorname{Tr}[A] - \operatorname{Tr}[(A-B)_+] \leq \operatorname{Tr}[B^s A^{1-s}] \end{aligned} \tag{5.4.21}$$

If we choose $A = p\rho_1^{\otimes m}$ and $B = (1-p)\rho_2^{\otimes m}$, then

$$\begin{aligned} \frac{1}{2}(1 - \left\| p\rho_1^{\otimes m} - (1-p)\rho_2^{\otimes m} \right\|_1) &\leq p^{1-s}(1-p)^s \operatorname{Tr}[(\rho_1^{\otimes m\,1-s} \rho_2^{\otimes m\,s} \\ &= p^{1-s}(1-p)^s \operatorname{Tr}[(\rho_1^{1-s}\rho_2^s)^{\otimes m}] = p^{1-s}(1-p)^s \operatorname{Tr}[\rho_1^{1-s}\rho_2^s]^m \end{aligned} \tag{5.4.22}$$

This gives us that

$$\mathcal{E}_m^{opt} \leq \inf_{s\in[0,1]} p^{1-s}(1-p)^{1-s} \operatorname{Tr}[\rho_1^{1-s}\rho_2^s]^m \leq \inf_{s\in[0,1]} \operatorname{Tr}[\rho_1^{1-s}\rho_2^s]^m \tag{5.4.23}$$

and hence for all m

$$-\frac{1}{m} \log \mathcal{E}_m^{opt} \geq -\log \inf_{s\in[0,1]} \operatorname{Tr}[\rho_1^{1-s}\rho_2^s] \tag{5.4.24}$$

which in the limit gives us

$$\lim_{m\to\infty} -\frac{1}{m} \log \mathcal{E}_m^{opt} \geq -\log(\inf_{s\in[0,1]} \operatorname{Tr}[\rho_1^{1-s}\rho_2^s]). \tag{5.4.25}$$

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 $x = 1, 2$

$$\rho_x = \sum_i \lambda_i^x |\psi_i^x\rangle\langle\psi_i^x| \tag{5.4.26}$$

and hence

$$\begin{aligned} \hat{\rho}_1 &= \sum_{i,j} \lambda_i^1 |\langle\psi_i^1, \psi_j^2\rangle| |ij\rangle\langle ij| \\ \hat{\rho}_2 &= \sum_{i,j} \lambda_i^2 |\langle\psi_i^1, \psi_j^2\rangle| |ij\rangle\langle ij| \end{aligned} \tag{5.4.27}$$

with $\{|ij\rangle\}$ a ONB of $\mathcal{H} \otimes \mathcal{H}$. ■

5.4.2 The pretty good measurement

Let us define now the two following measurements:

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
$M_x^p = R^{-1/2} p_x \rho_x + \frac{1}{n} \underbrace{(\mathbb{1} - R^{-1/2} R R^{-1/2})}_{\mathbb{1}_{\ker(R)}}$	$M_x^s := S^{-1/2} p_x^2 \rho_x^2 S^{-1/2} + \frac{1}{n}(\mathbb{1} - S^{-1/2} S S^{-1/2})$
$M^P = (M_x^p)_{x=1}^n$	$M^S = (M_x^s)_{x=1}^n$

with R^{-1} and S^{-1} the Moore-Penrose pseudo inverse. We need the following relations and definitions (which we will not prove here) in the following:

Definition 5.4.8 — Schatten p-norms. Let $\mathcal{H}$ be a finite dimensional Hilbert space. Then for $p \in [1, \infty)$

$$\|\cdot\|_p : \mathcal{B}(\mathcal{H}) \to [0, \infty), \quad A \mapsto \|A\|_p = \mathrm{Tr}[|A|^p]^{1/p} \tag{5.4.28}$$

is a norm on $\mathcal{B}(\mathcal{H})$.

Theorem 5.4.9 — Hoelder's inequality. For $p, q \in [1, \infty]$ and $\frac{1}{p} + \frac{1}{q} = 1$ we find that

$$\|AB\|_1 = \mathrm{Tr}[|AB|] \leq \|A\|_p \|B\|_q \tag{5.4.29}$$

Theorem 5.4.10 — 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 $n \in \mathbb{N}$
$$f\Big(\sum_{i=1}^{n} A_i^* X_i A_i\Big) \leq \sum_{i=1}^{n} A_i^* f(X_i) A_i \tag{5.4.30}$$
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. $f(V^* X V) \leq V^* f(X) V$, 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 5.4.11 We find that in the setting above, we find

$$(\mathrm{Tr}[S^{1/2}])^2 \leq \mathcal{P}(M^S) \leq \mathcal{P}^{opt} \leq \mathrm{Tr}[S^{1/2}] \tag{5.4.31}$$

Proof. 1.

$$\begin{aligned}(\mathrm{Tr}[S^{1/2}])^2 = (\mathrm{Tr}[SS^{-1/2}])^2 &= \Big(\sum_x \mathrm{Tr}[\underbrace{p_x^2\rho_x^2}_{\sigma_x^2} S^{-1/2}]\Big)^2 \\ &= \Big(\sum_x \mathrm{Tr}[\sigma_x(\sigma_x^{1/2}S^{-1/2}\sigma_x^{1/2})]\Big)^2 \\ &\overset{\text{Jensen}}{\leq} \sum_x \mathrm{Tr}[\sigma_x(\sigma_x^{1/2}S^{-1/2}\sigma_x^{1/2})^2] \\ &= \sum_x \mathrm{Tr}[\sigma_x^2 S^{-1/2}\sigma_x S^{-1/2}] \\ &= \sum_x \mathrm{Tr}[\sigma_x \underbrace{S^{-1/2}\sigma_x^2 S^{-1/2}}_{M_x^s}] = \mathcal{P}(M^S)\end{aligned} \tag{5.4.32}$$

2. Using that $z \mapsto z^{1/2}$ is operator monotone we find that

$$\sigma_x^2 \leq \sum_x \sigma_x^2 = S \tag{5.4.33}$$

giving us that

$$\sigma_x \leq S^{1/2} \quad \forall x = 1, \ldots, n \tag{5.4.34}$$

As a consequence we obtain

$$\sum_x \mathrm{Tr}[M_x\sigma_x] \leq \sum_x \mathrm{Tr}[M_x S^{1/2}] = \mathrm{Tr}[(\underbrace{\sum_x M_x}_{=\mathbb{1}})\, S^{1/2}] = \mathrm{Tr}[S^{1/2}] \tag{5.4.35}$$

■

Proposition 5.4.12 We find that

$$(\mathcal{P}^{opt})^2 \leq \mathcal{P}(M^p) \leq \mathcal{P}^{opt} \tag{5.4.36}$$

Proof. Let $M = (M_x)_{x=1}^n$ be a POVM. We then findt that

$$\begin{aligned}\Big(\sum_x \mathrm{Tr}[M_x\sigma_x]\Big)^2 &= \Big(\sum_x \mathrm{Tr}[(R^{1/4}M_xR^{1/4})(R^{-1/4}\sigma_xR^{-1/4})]\Big)^2 \\ &\overset{Hoelder}{\leq} \Big(\sum_x \left\|R^{1/4}M_xR^{1/4}\right\|_2 \left\|R^{-1/4}\sigma_xR^{-1/4}\right\|_2\Big)^2 \\ &\leq \underbrace{\sum_x \left\|R^{1/4}M_xR^{1/4}\right\|_2^2}_{(1)} \underbrace{\sum_x \left\|R^{-1/4}\sigma_xR^{-1/4}\right\|_2^2}_{(2)} \leq \mathcal{P}(M^P)\end{aligned} \tag{5.4.37}$$

We find that

$$\begin{aligned}(1) &= \sum_x \left\|R^{1/4}M_xR^{1/4}\right\|_2^2 = \sum_x \mathrm{Tr}[(R^{1/4}M_xR^{1/4})^2] = \sum_x \mathrm{Tr}[R^{1/2}M_xR^{1/2}M_x] \\ &\leq \sum_x \mathrm{Tr}[R^{1/2}M_xR^{1/2}] = \mathrm{Tr}[R] = 1 \\ (2) &= \sum_x \left\|R^{-1/4}\sigma_xR^{-1/4}\right\|_2^2 = \sum_x \mathrm{Tr}[\underbrace{R^{-1/2}\sigma_xR^{-1/2}}_{M_x^p}\sigma_x] = \mathcal{P}(M^p)\end{aligned} \tag{5.4.38}$$

■

In summary with the relation $\mathcal{E}(M) = 1 - \mathcal{P}(M)$, $\mathcal{E}^{opt} = 1 - \mathcal{P}^{opt}$ we find

$$(\mathcal{P}^{opt})^2 \leq \begin{Bmatrix}\mathcal{P}(M^P)\\ \mathcal{P}(M^S)\end{Bmatrix} \leq \mathcal{P}^{opt}, \qquad (\mathcal{E}^{opt}) \leq \begin{Bmatrix}\mathcal{E}(M^P)\\ \mathcal{P}(M^S)\end{Bmatrix} \leq 2\,\mathcal{E}^{opt} \tag{5.4.39}$$

5.5 Bonus: Separability criteria

5.5.1 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

$$|\psi\rangle = \sum_{i=1}^{d} \sqrt{\lambda_i}\, |e_i\rangle \otimes |f_i\rangle \tag{5.5.1}$$

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.

$$S_{ENT}(\rho) := -\sum_{i=1} \lambda_i \log(\lambda_i) \tag{5.5.2}$$

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 5.5.1 — Partial transpose. The partial transpose is a positive linear map, which is not completely positive. We first introduce the transposition map

$$\Theta : A \mapsto A^t, \qquad \langle i, A^T j\rangle = \langle j, Ai\rangle \qquad \forall i, j. \tag{5.5.3}$$

Using this map we define the partial transpose through its action on the maximally entangled state $|\Omega\rangle = \frac{1}{d}\sum_{i=1}^{d} |ii\rangle$,

$$(\Theta \otimes \mathbb{1})(|\Omega\rangle\langle\Omega|) = \frac{1}{d}\,\mathbb{F} \qquad \mathbb{F} := \sum_{i,j=1}^{n} |ij\rangle\langle ji| \tag{5.5.4}$$

The partial trace can be used to detect entanglement, as follows:

Proposition 5.5.2 $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$. Consider $\rho \in \mathcal{S}(\mathcal{H}_{AB})$. If ρ^{T_A} has a negative eigenvalue, then ρ is entangled.

Proof. ρ separable $\Rightarrow \rho = \sum_j p_j \rho_j^A \otimes \rho_j^A \overset{\cdot^{T_A}}{\Rightarrow} \rho^{T_A} = \sum_j p_j (\rho_j^A)^T \otimes \rho_j^B \geq 0$ ■

We again need to introduce some nomenclature to proof the next proposition.

Definition 5.5.3 — Entanglement witness. We first set

$$\delta := \{\text{separable density matrices}\} \tag{5.5.5}$$

is convex and compact set. By the Hahn-Banach Theorem $\rho \notin \delta$, then there exists a hyperplane ω such that

$$\mathrm{Tr}[\rho\omega] < 0 \quad \text{and} \quad \mathrm{Tr}[\sigma\omega] \geq 0 \quad \sigma \in \delta. \tag{5.5.6}$$

We then call ω a entanglement witness. The Choi-Jamiolkowski matrix of this state is given

through its action on the maximally entangled state

$$\omega = (\Lambda^* \otimes \mathbb{1})(|\Omega\rangle\langle\Omega|) \tag{5.5.7}$$

for Λ a quantum channel.

Proposition 5.5.4 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

$$(\Lambda \otimes \mathbb{1}_B)(\rho) = (\Lambda \otimes \mathbb{1}_B)\Big(\sum_j \lambda_j \rho_j^A \otimes \rho_j^B\Big) = \sum_j \lambda_j \underbrace{\Lambda(\rho_j^A)}_{\geq 0} \otimes \rho_j^B \geq 0 \tag{5.5.8}$$

$\Leftarrow$ 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 5.5.3, we find

$$\begin{aligned} \mathrm{Tr}[(A \otimes B)\omega] &= \mathrm{Tr}[B^T \Lambda(A)] = \mathrm{Tr}[\mathbb{F}(\Lambda(A) \otimes B^T)] = d\,\mathrm{Tr}[(\Lambda \otimes \mathbb{1}_B)(A \otimes B)|\Omega\rangle\langle\Omega|] \\ &= d\,\langle\Omega, (\Lambda \otimes \mathbb{1}_B)(A \otimes B)\Omega\rangle \end{aligned} \tag{5.5.9}$$

In the second step we used that $\mathrm{Tr}[XY] = \mathrm{Tr}[\mathbb{F}\, X \otimes Y]$. Now this means

$$\mathrm{Tr}[\rho\omega] = d\,\langle\Omega, (\Lambda \otimes \mathbb{1}_B)(\rho)\Omega\rangle \tag{5.5.10}$$

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 5.5.5 The idea to implement this map in a lab is

$$\rho \overset{T}{\mapsto} \frac{p}{d^2}\,\mathbb{1}_d \otimes \mathbb{1}_d + (1-p)(\Lambda \otimes \mathbb{1})\rho. \tag{5.5.11}$$

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.

5.5.1.1 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

$$\rho^{\tilde{T}_A} = (U \otimes \mathbb{1})[(U^* \otimes \mathbb{1})\rho(U \otimes \mathbb{1})]^{\tilde{T}_A} = [(UU^T) \otimes \mathbb{1}]\rho^{T_A}[(UU^T)^* \otimes \mathbb{1}] \neq \rho^{T_A}. \tag{5.5.12}$$

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 5.5.6 — 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 5.5.7 The above definition allows us to write the entanglement withness as

$$\omega = Q_1 + Q_2^T \tag{5.5.13}$$

with the PSD

$$Q_i = d(\Lambda_i^* \otimes \mathbb{1})(|\Omega\rangle\langle\Omega|). \tag{5.5.14}$$

In general the separability criteria of the entanglement witness is weaker than of transpositions, i.e.

$$\rho^{T_A} \geq 0 \quad \Rightarrow \quad (\Lambda \otimes \mathbb{1})(\rho) \geq 0 \tag{5.5.15}$$

■ **Example 5.5.8** Let $\Lambda_{red}(A) = \mathrm{Tr}[A]\,\mathbb{1} - A$ we get the separability criteria

$$(\Lambda_{red} \otimes \mathbb{1})(\rho) \geq 0 \Leftrightarrow \begin{cases} \rho_A \otimes \mathbb{1}_B \geq \rho_{AB} \\ \mathbb{1}_A \otimes \rho_B \geq \rho_{AB} \end{cases}. \tag{5.5.16}$$

We get for the witness

$$\omega_{red} = (\mathbb{1} - \mathbb{F})^{T_A} = 2P_-^{T_A} \tag{5.5.17}$$

with P_- the projector onto the anti-symmetric space. Further the ω prop is

$$\mathrm{Tr}[\rho\omega] < 0 \quad \Leftrightarrow \quad \langle\Omega, \rho\Omega\rangle \leq \frac{1}{d} \tag{5.5.18}$$

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 5.5.9 Let $\rho \in \mathcal{S}(\mathbb{C}^2 \otimes \mathbb{C}^3)$ or $\mathcal{S}(\mathbb{C}^2 \otimes \mathbb{C}^2)$, then

$$\rho \text{ separable} \quad \Leftrightarrow \quad \rho^{T_A} \geq 0 \tag{5.5.19}$$

which is a consequence of the complete decomposability of every positive map in $2 \otimes 2$ and $2 \otimes 3$.

Proposition 5.5.10 Entangled states with PPT exists if and only if there are non-decomposable maps.

Chapter 6

Quantum error correction

The notes that should be consulted for this chapter of the course are Lectures 16 and 17 of [24].

Chapter 7

Quantum cryptography

The notes that should be consulted for this chapter of the course are Lectures 18 and 19 of [24].

Chapter 8

Quantum Shannon Theory

8.1 Quantum Entropies

8.1.1 Von Neumann Entropy

In classical information theory, the setting we consider to store information is that 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

$$H(X) := -\sum_x p_x \log p_x. \tag{8.1.1}$$

This quantity gives the value of information, meaning the number of incompressible 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

$$I(X : Y) = H(X) + H(Y) - H(X, Y). \tag{8.1.2}$$

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.

8.1.1.1 Quantum Generalisation

We now translate these 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.

$$\rho = \sum_x p_x \rho_x \tag{8.1.3}$$

Definition 8.1.1 — 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:

$$S(\rho) := -\operatorname{Tr}[\rho \log \rho] = -\operatorname{Tr}[UDU^{-1} \log(UDU^{-1})] = -\operatorname{Tr}[D \log(D)] = -\sum_x \lambda_x \log(\lambda_x) \tag{8.1.4}$$

with

$$\rho = UDU^{-1} \quad \text{with} \quad D = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_n \end{pmatrix} \tag{8.1.5}$$

Remark 8.1.2

$$\log \rho = U \begin{pmatrix} \log \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \log \lambda_n \end{pmatrix} U^{-1} \tag{8.1.6}$$

Remark 8.1.3 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 8.1.4 — 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

$$S\Big(\sum_i \lambda_i \rho_i\Big) \geq \sum_i \lambda_i S(\rho_i). \tag{8.1.7}$$

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

$$Y = \{a_y, p(a_y)\} \qquad p(a_y) = \langle a_y, \rho a_y \rangle . \tag{8.1.8}$$

It is immediately clear that

$$H(Y) \geq S(\rho) \tag{8.1.9}$$

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

$$H(X) \geq S(\rho) \tag{8.1.10}$$

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

$$S(\rho_{AB}) = S(\rho_A) + S(\rho_B) \tag{8.1.11}$$

8. **Subadditivity** It holds in general that

$$S(\rho_{AB}) \leq S(\rho_A) + S(\rho_B) \tag{8.1.12}$$

This statement is equivalent to the quantum mutual information

$$I_\rho(A:B) = S(\rho_A) + S(\rho_B) - S(\rho_{AB}) \tag{8.1.13}$$

being greater or equal to zero.

9. **Strong subadditivity** We further have that

$$S(\rho_{ABC}) + S(\rho_B) \leq S(\rho_{AB}) + S(\rho_{BC}), \tag{8.1.14}$$

$\rho_{ABC} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_B)$.

10. **Triangle inequality** (Araki-Lieb inequality)

$$S(\rho_{AB}) \geq |S(\rho_A) - S(\rho_B)| \tag{8.1.15}$$

Remark 8.1.5 Let ρ_{AB} a pure quantum state in a bipartite Hilbert space, i.e. $S(\rho_{AB}) = 0$. Then, $S(\rho_A) = S(\rho_B)$ but in general $S(\rho_A) = S(\rho_B) \neq 0$.

8.1.1.2 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

$$\rho_{SE} \mapsto U_{SE}\rho_{SE}U_{SE}^{-1} = \rho'_{SE} \tag{8.1.16}$$

From Proposition 8.1.4, we find

$$S(\rho_S) + S(\rho_E) = S(\rho_{SE}) = S(\rho'_{SE}) \leq S(\rho'_S) + S(\rho'_E). \tag{8.1.17}$$

In a sense, Equation (8.1.17) can be interpreted as the second law of thermodynamics.

8.1.1.3 Conditional entropy and mutual information

Definition 8.1.6 — Conditional entropy. Let $\rho_{AB} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$, then we define the conditional quantum entropy as

$$H(A|B)_\rho := S(\rho_{AB}) - S(\rho_B)\,. \tag{8.1.18}$$

Remark 8.1.7 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 8.1.8 — Coherent information. The coherent information for $\rho_{AB} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ is given by

$$I(A\langle B)_\rho := S(\rho_B) - S(\rho_{AB}) \tag{8.1.19}$$

Definition 8.1.9 — Mutual information. Let $\rho_{AB} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$, then the mutual information is given by

$$I(A:B)_\rho = S(\rho_A) + S(\rho_B) - S(\rho_{AB}) = S(\rho_A \otimes \rho_B) - S(\rho_{AB}) \tag{8.1.20}$$

Remark 8.1.10 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 8.1.11 — 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

$$I(A:C|B)_\rho = S(\rho_{AB}) + S(\rho_{BC}) - S(\rho_B) - S(\rho_{ABC}) \geq 0 \tag{8.1.21}$$

8.1.2 Relative entropy

Definition 8.1.12 — Quantum relative entropy. Let $\rho, \sigma \in \mathcal{S}(\mathcal{H})$, then the (Umegaki) quantum relative entropy is given as

$$D(\rho\|\sigma) = \begin{cases} \mathrm{Tr}[\rho(\log\rho - \log\sigma)] & \ker\sigma \subseteq \ker\rho \\ +\infty & \text{otherwise} \end{cases}. \tag{8.1.22}$$

Remark 8.1.13 The quantum relative entropy is inspired by the Kullback-Leibler divergence. For $p = \{p_x\}$, $q = \{q_x\}$ probability distributions

$$KL(p\|q) = \sum_x p_x \log\frac{p_x}{q_x} \tag{8.1.23}$$

Definition 8.1.14 — Belavkin-Staszewski relative entropy. Let $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ full rank, then the Belavkin-Staszewski relative entropy is given by

$$\widehat{D}(\rho\|\sigma) = \mathrm{Tr}[\rho \log \rho^{1/2}\sigma^{-1}\rho^{1/2}] \tag{8.1.24}$$

Remark 8.1.15 We have the relation

$$D(\rho\|\sigma) \leq \widehat{D}(\rho\|\sigma) \tag{8.1.25}$$

with equality if and only if $[\rho, \sigma] = 0$.

Proposition 8.1.16 — 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

$$D(\rho_A \otimes \rho_B\|\sigma_A \otimes \sigma_B) = D(\rho_A\|\sigma_A) + D(\rho_B\|\sigma_B) \tag{8.1.26}$$

- Superadditivity: $\rho_{AB} \in \mathcal{S}(\mathcal{H}_{AB})$, $\sigma_A \in \mathcal{S}(\mathcal{H}_A)$, $\sigma_B \in \mathcal{S}(\mathcal{H}_B)$

$$D(\rho_{AB}\|\sigma_A \otimes \sigma_B) \geq D(\rho_A\|\sigma_A) + D(\rho_B\|\sigma_B). \tag{8.1.27}$$

- Data processing inequality: $\rho, \sigma \in \mathcal{S}(\mathcal{H})$, T a quantum channel, then

$$D(\rho\|\sigma) \geq D(T(\rho)\|T(\sigma)) \tag{8.1.28}$$

Proposition 8.1.17 — Properties of the quantum relative entropy II. The relative entropy further has the properties:

- Non-negativity: $D(\rho\|\sigma) \geq 0$
- Unitary invariance: $D(U\rho U^*\|U\sigma U^*) = D(\rho\|\sigma)$

Theorem 8.1.18 — 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 8.1.16, then f is the relative entropy.

Proof. Step 1. 1. - 3. imply "lower asymptotic semicontinuity" (LAS) Definition 8.1.19. Let therefore be (ρ, σ) and $\{\rho'_n\}$ sequence of state be given, such that

$$\left\|\rho^{\otimes n} - \rho'_n\right\|_1 \overset{n\to\infty}{\longrightarrow} \tag{8.1.29}$$

Through the DPI for $\|\cdot\|_1$ we can conclude that

$$\|\rho - (\rho'_n)_i\|_1 \overset{n\to\infty}{\longrightarrow} . \tag{8.1.30}$$

Now applying superadditivity to the first and addtitivity to the second summand gives

$$\begin{aligned} \frac{1}{n}(D(\rho'_n\|\sigma^{\otimes n}) - D(\rho^{\otimes}\|\sigma^{\otimes n})) &\overset{2.+3.}{\geq} \frac{1}{n}\sum_{i=1}^{n}\left[D((\rho'_n)_i\|\sigma) - D(\rho\|\sigma)\right] \\ &\geq \min_{i=1,\dots,n}\left[D((\rho'_n)_i\|\sigma) - D(\rho\|\sigma)\right] \overset{n\to\infty}{\underset{1.}{\longrightarrow}} 0 \end{aligned} \tag{8.1.31}$$

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

$$\frac{l'}{m'}D(\rho_0\|\sigma_0) \leq D(\rho\|\sigma) \leq \frac{l}{m}D(\rho_0\|\sigma_0). \tag{8.1.32}$$

Now the upper bound is equivalent to saying that

$$\begin{aligned} & mD(\rho\|\sigma) \leq lD(\rho_0\|\sigma_0) \\ \overset{2.}{\iff} \quad & D(\rho^{\otimes m}\|\sigma^{\otimes m}) \leq D(\rho_0^{\otimes l}\|\sigma_0^{\otimes l}) \\ \overset{\text{Lemma 8.1.20}}{\implies} \quad & \Psi^n(\sigma_0^{\otimes ln}) = \sigma^{\otimes nm}, \quad \lim_{n\to\infty}\left\|\Psi^n(\rho_0^{\otimes ln}) - \rho^{\otimes mn}\right\|_1 = 0 \end{aligned} \tag{8.1.33}$$

Therefore,

$$\begin{aligned} mf(\rho,\sigma) &\overset{2.}{=} f(\rho^{\otimes n}, \sigma^{\otimes n}) \overset{2.}{=} \limsup_{n\to\infty}\frac{1}{n}f(\rho^{\otimes mn}, \sigma^{\otimes mn}) \\ &\leq \liminf_{n\to\infty}\frac{1}{n}f(\Psi^n(\rho_0^{\otimes ln}), \underbrace{\Psi^n(\sigma_0^{\otimes ln})}_{\sigma^{\otimes nm}}) \\ &\overset{4.+\text{DPI}}{\leq} \liminf_{n\to\infty}\frac{1}{n}f(\rho_0^{\otimes ln}, \sigma_0^{\otimes ln}) \\ &\overset{2.}{=} f(\rho_0^{\otimes l}, \sigma_0^{\otimes l}) \overset{2.}{=} lf(\rho_0, \sigma_0) \end{aligned} \tag{8.1.34}$$

which is equivalent to

$$f(\rho,\sigma) \leq \frac{l}{m}f(\rho_0,\sigma_0) = \frac{l}{m}D(\rho_0,\sigma_0). \tag{8.1.35}$$

Analogously one obtains

$$\frac{l'}{m'}D(\rho_0\|\sigma_0) = \frac{l'}{m'}f(\rho_0,\sigma_0) \leq f(\rho,\sigma) \leq \frac{l}{m}D(\rho\|\sigma_0). \tag{8.1.36}$$

Choosing properly l, m, l', m', we can conclude that $f(\rho,\sigma) \propto D(\rho\|\sigma)$.

■

Definition 8.1.19 — 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

$$\lim_{n\to\infty} \left\|\rho^{\otimes n} - \rho'_n\right\|_1 = 0 \tag{8.1.37}$$

then, this implies that

$$\liminf_{n\to\infty} \frac{1}{n}(f(\rho'_n,\sigma^{\otimes n}) - f(\rho^{\otimes n},\sigma^{\otimes n})) \geq 0 \tag{8.1.38}$$

Lemma 8.1.20 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

$$\Psi_n(\sigma_0^n) = \sigma^{\otimes n}, \quad \lim_{n\to\infty} \left\|\Psi_n(\rho_0^{\otimes n}) - \rho^{\otimes n}\right\|_1 = 0 \tag{8.1.39}$$

8.1.3 Non-commmutative L_p norms

8.1.3.1 Schatten p-norms

Definition 8.1.21 — 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

$$\|X\|_p := (\mathrm{Tr}[|X|^p])^{1/p} \tag{8.1.40}$$

where $|X| = \sqrt{X^*X}$. For $p = \infty$, we define

$$\|\cdot\|_\infty = \lim_{p\to\infty} \|\cdot\|_p \tag{8.1.41}$$

the operator norm. For $p < 1$, it does not satisfy the triangle inequality.

Proposition 8.1.22 — Properties of Schatten p-norms. The Schatten p-norms have the following properties:

1. **Monotonicity:** For $1 \leq p \leq p' \leq +\infty$,

$$\|X\|_1 \leq \|X\|_p \leq \|X\|_{p'} \leq \|X\|_\infty . \tag{8.1.42}$$

2. **Unitary invariance** For U a unitary, we have

$$\|UXU^*\|_p = \|X\|_p . \tag{8.1.43}$$

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:**

$$\|X\|_q := \sup\left\{|\langle X,Y\rangle| \ : \ \|Y\|_p = 1\right\} \tag{8.1.44}$$

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 8.1.23 — 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:

$$\|X\|_{L_p(\rho)} := \mathrm{Tr}[|\rho^{\frac{1}{2p}} X \rho^{\frac{1}{2p}}|^p]^{1/p} \tag{8.1.45}$$

for all $X \in \mathcal{B}(\mathcal{H})$. We further define the KMS (Kubo-Martin-Schwinger) inner product

$$\langle X, Y \rangle_{\rho,KMS} := \mathrm{Tr}[\rho^{1/2} X j \rho^{1/2} Y] \tag{8.1.46}$$

and the GNS (Gelfant-Naimark-Segal) inner product

$$\langle X, Y \rangle_{\rho,GNS} := \mathrm{Tr}[\rho X^* Y] \tag{8.1.47}$$

for all $X, Y \in \mathcal{B}(\mathcal{H})$

Proposition 8.1.24 — 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:** $\|X\|_{L_p(\rho)} := \sup\left\{ |\langle X, Y \rangle_{\rho,GNS}| \ : \ \|Y\|_{L_q(\rho)} = 1 \right\}$
3. **Operator norm:** $\|X\|_\infty = \|X\|_{L_\infty(\rho)} := \lim_{p \to \infty} \|X\|_{L_p(\rho)}$

8.1.4 Quantum 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

$$KL(p\|q) = \sum_{x=1}^{n} p_x \log \frac{p_x}{q_x}, \tag{8.1.48}$$

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 in a similar construction.

Definition 8.1.25 — 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.

$$\mathbb{Q}(\cdot\|\cdot) = g(\mathbb{D}(\cdot\|\cdot)) \tag{8.1.49}$$

and for $X_1, Y_1 \in \mathcal{B}(\mathcal{H}_1)_+$, $X_2, Y_2 \in \mathcal{B}(\mathcal{H}_2)_+$

$$\mathbb{Q}(X_1 \oplus X_2 \| Y_1 \oplus Y_2) = \frac{\mathrm{Tr}[X_1]}{\mathrm{Tr}[X_1] + \mathrm{Tr}[X_2]} \mathbb{Q}(X_1\|Y_1) + \frac{\mathrm{Tr}[X_2]}{\mathrm{Tr}[X_1] + \mathrm{Tr}[X_2]} \mathbb{Q}(X_2\|Y_2) \tag{8.1.50}$$

Proposition 8.1.26 — Classical case. A divergence satisfying Definition 8.1.25 is either the Kullback-Leibler divergence of the Renyi divergence:

$$D_\alpha(p\|q) = \frac{1}{\alpha - 1} \log \frac{\sum_{x=1}^n p_x^\alpha q_x^{1-\alpha}}{\sum_{x=1}^n p_x} \tag{8.1.51}$$

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 8.1.27 — Quantum axiomatic definition of a divergence. We are in the setting of Definition 8.1.25 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

$$\mathbb{D}\left(\sum_i \lambda_i \rho_i \| \sum_i \lambda_i \sigma_i\right) \leq \sum_i \lambda_i \mathbb{D}(\rho_i\|\sigma_i) \tag{8.1.52}$$

   (b) **Joint concavity:** $(\alpha < 1)$ $\{\rho_i\}_i$, $\{\sigma_i\}_i$, $0 \leq \lambda_i \leq 1$, then

$$\mathbb{D}\left(\sum_i \lambda_i \rho_i \| \sum_i \lambda_i \sigma_i\right) \geq \sum_i \lambda_i \mathbb{D}(\rho_i\|\sigma_i) \tag{8.1.53}$$

4. **Dominance:** $X, Y, Y' \in \mathcal{B}(\mathcal{H})_+$, $Y \leq Y'$, then $\mathbb{D}(X\|Y) \geq \mathbb{D}(X\|Y')$

8.1.5 Minimal Divergence

Definition 8.1.28 — 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

$$\mathcal{P}(L) = \sum_{x=1}^n P_x L P_x = \sum_{y=1}^n U_y L U_y^* \tag{8.1.54}$$

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 8.1.29 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 8.1.30 — Preparation 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

$$q(x) = \mathrm{Tr}[\sigma\Pi_x], \qquad p(x) = \lambda_x q(x) \tag{8.1.55}$$

and with that

$$\Lambda(\cdot) = \sum_x \langle x, \cdot x\rangle \frac{1}{q(x)} \sigma^{1/2}\Pi_x\sigma^{1/2}. \tag{8.1.56}$$

We find that $\Lambda(p) = \rho$ and $\Lambda(q) = \sigma$.

Using the above we can define the minimal Rényi divergences

Definition 8.1.31 — Minimal Rényi divergence (Sandwiched Rényi Divergences). For $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ $\alpha \in (1/2, 1) \cup (1, \infty)$,

$$\begin{aligned} \mathcal{D}_\alpha(\rho\|\sigma) \lim_{n\to\infty} \frac{1}{n} \mathbb{D}_\alpha(\rho^{\otimes n}\|\sigma^{\otimes n}) &\overset{\text{DPI}}{\geq} \lim_{n\to\infty} \frac{1}{n} \mathbb{D}_\alpha(\mathcal{P}_{\sigma^{\otimes n}}(\rho^{\otimes n})\|\sigma^{\otimes n}) \\ &= \frac{1}{\alpha - 1} \log \mathrm{Tr}[(\sigma^{\frac{1-\alpha}{2\alpha}}\rho\sigma^{\frac{1-\alpha}{2\alpha}})^\alpha] =: \widetilde{D}_\alpha(\rho\|\sigma) \end{aligned} \tag{8.1.57}$$

and also the maximal ones

Definition 8.1.32 — Maximal Rényi divergences (Geometric Rényi Divergences). For $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ $\alpha \in (1, 2)$, we find that

$$\mathcal{D}_\alpha(\rho\|\sigma) = \mathcal{D}_\alpha(\Lambda(p)\|\Lambda(q)) \overset{\text{DPI}}{\leq} \mathcal{D}_\alpha(p\|q) = \frac{1}{\alpha - 1} \log \mathrm{Tr}[\sigma(\sigma^{-1/2}\rho\sigma^{-1/2})^\alpha] =: \widehat{D}_\alpha(\rho\|\sigma). \tag{8.1.58}$$

The quantity in the argument of the trace is called a geometric mean, which we also write as

$$\sigma \#_\alpha \rho = \sigma^{1/2}(\sigma^{-1/2}\rho\sigma^{-1/2})^\alpha\sigma^{1/2} \tag{8.1.59}$$

Remark 8.1.33

- 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

$$\lim_{\alpha\to 1} \widehat{D}_\alpha(\rho\|\sigma) = \widehat{D}(\rho\|\sigma) := \mathrm{Tr}[\rho \log(\rho^{1/2}\sigma^{-1}\rho^{1/2})] \tag{8.1.60}$$

$$\lim_{\alpha\to 1} \widetilde{D}_\alpha(\rho\|\sigma) = D(\rho\|\sigma) = \mathrm{Tr}[\rho(\log(\rho) - \log(\sigma))] \tag{8.1.61}$$

Definition 8.1.34 — Petz Rényi Divergence. For $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ $\alpha \in (0, 1)$ we define the Petz Rényi Divergence as

$$\overline{D}_\alpha(\rho\|\sigma) := \frac{1}{\alpha - 1} \log \mathrm{Tr}[\rho^\alpha\sigma^{1-\alpha}] \tag{8.1.62}$$

8.2 Quantum Hypothesis Testing

8.2.1 Symmetric State Discrimination

Let us recall the setting of binary hypothesis testing. Consider ρ_1, ρ_2 density matrices with a priory probability p and $(1-p)$. Further $M = (M_1, M_2) \cong (\mathbb{P}, \mathbb{1} - \mathbb{P})$ 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 ρ_2 the error becomes

$$\mathcal{E}(M) := p \operatorname{Tr}[\rho_1(\mathbb{1} - \mathbb{P})] + (1-p) \operatorname{Tr}[\rho_2 P] \tag{8.2.1}$$

Remark 8.2.1 It is clear that for

$$\mathcal{P}(M) = p \operatorname{Tr}[\rho_1 \mathbb{P}] + (1-p) \operatorname{Tr}[\rho_2(\mathbb{1} - \mathbb{P})] \tag{8.2.2}$$

we find

$$\mathcal{P}(M) + \mathcal{E}(M) = 1\,. \tag{8.2.3}$$

In a previous chapter, we proved the inequality

$$\mathcal{E}(M) \geq \frac{1}{2}(1 - \|p\rho_1 - (1-p)\rho_2\|_1) \tag{8.2.4}$$

with equality, if and only if P is a projection onto $(p_1\rho_1 - (1-p)\rho_2)_+$. We also discussed what happens in the case in which we send $m \in \mathbb{N}$ 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

$$\mathcal{E}_m^{opt} = \frac{1}{2}(1 - \left\|p\rho_1^{\otimes m} - (1-p)\rho_2^{\otimes m}\right\|_1) \tag{8.2.5}$$

and $\mathcal{E}_m^{opt}$ decays exponentially with $-\xi_m$, with ξ a rate given as

$$\mathcal{E}_m^{opt} \leq K e^{-\xi m}\,. \tag{8.2.6}$$

We can better describe this scenario in the following way:

Definition 8.2.2 Consider two states $\rho, \sigma \in \mathcal{S}(\mathcal{H})$. Consider A the system associated to Alice. Define the following two scenarios:

- Null hypothesis: The state of A^n is $\rho^{\otimes n}$.
- Alternate hypothesis: The state of A^n is $\sigma^{\otimes n}$.

For each of the copies, we get a POVM $\{\mathbb{P}, 1 - \mathbb{P}\}$ with $\mathbb{P}$ an orthogonal projection (for all the same). 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 $\rho^{\otimes n}$
$$\alpha_n(T_n; \rho) := \operatorname{Tr}[\rho^{\otimes n}(\mathbb{1} - T_n)]. \tag{8.2.7}$$
2. **Second kind error:** We wrongly conclude that the null hypothesis is correct even if the state is $\sigma^{\otimes n}$
$$\beta_n(T_n; \sigma) := \operatorname{Tr}[\sigma^{\otimes n} T_n] \tag{8.2.8}$$

We have the Chernoff bound as

$$\min_{\substack{T_n \\ \text{hypothesis} \\ \text{test}}} \frac{1}{2}(\alpha_n(T_n;\rho) + \beta_n(T_n;\sigma)) = \frac{1}{2}(1 - \left\|p\rho^{\otimes n} - (1-p)\sigma^{\otimes n}\right\|_1) \tag{8.2.9}$$

and the quantum Chernoff bound

$$\begin{aligned}\lim_{n\to\infty} -\frac{1}{n}\log \min_{\substack{T_n \\ \text{hypothesis} \\ \text{test}}} \frac{1}{2}(\alpha_n(T_n;\rho)+\beta_n(T_n;\sigma)) &= \max_{0\le s\le 1} -\operatorname{Tr}[\rho^s\sigma^{1-s}]. \\ &= -\min_{0\le s\le 1}\log \overline{Q}_s(\rho\|\sigma) \\ &= \max_{0\le s\le 1}(1-s)\overline{D}_s(\rho\|\sigma)\end{aligned} \tag{8.2.10}$$

We find a building block of the Petz Rényi divergence.

$$\overline{D}_\alpha(\rho\|\sigma) := \frac{1}{\alpha-1}\log \operatorname{Tr}[\rho^\alpha\sigma^{1-\alpha}] \qquad \alpha\in(0,1)\cup(1,+\infty). \tag{8.2.11}$$

So the interpretation of the Petz Rényi 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.

8.2.2 Asymmetric hypothesis testing

The goal of asymmetric quantum hypothesis testing is to minimize

$$\beta_n(T_n;\sigma) := \operatorname{Tr}[\sigma^{\otimes n}T_n] \tag{8.2.12}$$

under the constraint

$$\alpha_n(T_n;\rho) = \operatorname{Tr}[\rho^{\otimes n}(\mathbb{1}-T_n)] \le \varepsilon \tag{8.2.13}$$

Lemma 8.2.3 Let $T = \mathcal{P}_{\sigma^{\otimes n}} : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ the pinching map, $\rho,\sigma\in\mathcal{S}(\mathcal{H})$

$$\mathcal{P}_{\sigma^{\otimes n}}(X) := \sum_{i=1}^{\alpha} P_i X P_i, \tag{8.2.14}$$

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

$$D(\rho\|\sigma) = \lim_{n\to\infty}\frac{1}{n}D(\mathcal{P}_{\sigma^{\otimes n}}(\rho^{\otimes n}),\sigma^{\otimes n}) \tag{8.2.15}$$

8.2.3 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

$$\lim_{n\to\infty} -\frac{1}{n}\log\beta_n = D(\rho\|\sigma) \tag{8.2.16}$$

Proof. We want to prove that $D(\rho\|\sigma)$ is a lower bound on $-\frac{1}{n}\log\beta_n$. Using Lemma 8.2.4 with $A=\rho^{\otimes n}$ and $B=e^{\lambda n}\sigma^{\otimes n}$ ($\lambda\in\mathbb{R}$ will be chosen later). I.e. for $s\in[0,1]$,

$$e^{-s\lambda n}\operatorname{Tr}[\rho^{1+s}\sigma^{-1}]^n \ge \operatorname{Tr}[(\rho^{\otimes n}-e^{\lambda n}\sigma^{\otimes n})T_n] \ge (1-\varepsilon)-e^{\lambda n}\beta_n(T_n;\sigma_n) \tag{8.2.17}$$

where we used in the last step that $\alpha_n\le\varepsilon$. This gives us

$$\beta_n(T_n;\sigma) \ge e^{-n\lambda}[(1-\varepsilon)-e^{-n(\lambda n-f(s))}] \tag{8.2.18}$$

with $f(s) := \log\operatorname{Tr}[\rho^{1+s}\sigma^{-s}]$ 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

$$\lambda s > f(s) \tag{8.2.19}$$

which allows us to take the limit

$$\lim_{n\to\infty} -\frac{1}{n} \log \beta_n(T_n; \sigma) \leq D(\rho\|\sigma) + \delta \tag{8.2.20}$$

Since δ was arbitrary ■

Lemma 8.2.4 We have for self-adjoint A, B and all $s \in [0, 1]$ that

$$\|A - B\|_1 \geq \operatorname{Tr}[A + B] - 2\operatorname{Tr}[A^S B^{1-s}] \tag{8.2.21}$$

and further

$$\operatorname{Tr}[(A + B)_+] \leq \operatorname{Tr}[A^{1-s} B^s] \tag{8.2.22}$$

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. "Strengths and weaknesses of quantum computing". In: SIAM Journal on Computing.26(5) (1997), pp. 1510–1523.

[4] 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.

[5] A. Coladangelo and J. Stark. "Unconditional separation of finite and infinite-dimensional quantum correlations". In: *Nature Communications* 11 (2020), pp. 1–6.

[6] D. Deustch. "Quantum computational networks." In: Proceedings of the Royal Society of London.A425 (1989).

[7] D. Deustch. "Quantum theory, the Church-Turing principle, and the universal quantum Turing machine". In: Proceedings of the Royal Society of London.A400 (1985), pp. 97–117.

[8] D. Deustch and R. Josza. "Rapid solution of problems by quantum computation". In: Proceedings of the Royal Society of London.A439 (1992), pp. 553–558.

[9] A. Einstein, B. Podolsky, and N. Rosen. "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?" In: *Physical Review* 47 (1935), p. 777.

[10] L. K. Grover. "A fast quantum mechanical algorithm for database search". In: Proceedings of 28th ACM STOC.A425 (1996), pp. 212–219.

[11] Z. Ji et al. "MIP* = RE". In: *Communications of the ACM* 64.11 (2021), pp. 131–138.

[12] A. Y. Kitaev. "Quantum computations: algorithms and error correction". In: *Russian Mathematical Surveys* 52.6 (1997), p. 1191.

[13] S. Lloyd. "Universal quantum simulators". In: *Science* 273 (1996), pp. 1073–1078.

[14] 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.

[15] C. Palazuelos. *Introduction to Quantum Information Theory*. 2013.

[16] J. Preskill. *Quantum Computation. Lecture Notes*. 2015.

[17] V. B. Scholz and R. F. Werner. "Tsirelson's problem". In: *arXiv preprint:0812.4305* (2008).

[18] 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.

[19] W. Slofstra. "The set of quantum correlations is not closed". In: *Forum of Mathematics, Pi* 7 (2019).

[20] 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.

[21] R. Solovay. "Lie Groups and Quantum Circuits". In: *Mathematics of Quantum Computation* 07 (2000).

[22] U. Vazirani. *Notes of the Course CS294-2 Quantum Computation*. University of Berkeley, 2004.

[23] John Watrous. *Advanced Topics in Quantum Information Theory*. Lecture Notes, 2021. URL: https://cs.uwaterloo.ca/~watrous/QIT-notes/.

[24] John Watrous. *Introduction to Quantum Computing*. Lecture Notes, 2005. URL: https://cs.uwaterloo.ca/~watrous/QC-notes/.

[25] John Watrous. *The Theory of Quantum Information*. Cambridge University Press, 2018. DOI: 10.1017/9781316848142. URL: https://cs.uwaterloo.ca/~watrous/TQI/TQI.pdf.

[26] 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.

[27] 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.

[28] R. de Wolf. *Quantum Computing: Lecture Notes*. 2011.

[29] W. K. Wootters and W. H. Zurek. "A single quantum cannot be cloned". In: *Nature* 299 (1982), pp. 802–803.

[30] A. C-C. Yao. "Quantum circuit complexity". In: Proceedings of the 34th IEEE FOCS (1993), pp. 352–360.