A First Course in Quantum Information Theory

Designed for junior students of Tsinghua University

BY

GIULIO CHIRIBELLA

Associate Professor and Young 1000 Talents Program Member
IIIS, Tsinghua University

Beijing,
Academic year 2014-2015

Contents

Part I

The basic rules of quantum theory

Chapter 1

Pure states, basic measurements, and reversible gates

With this chapter you are entering in the first part of the course, where you will learn about the basic rules of quantum mechanics. In the following you will see what are the possible states of a quantum system, what are the possible measurements that you can use to extract information, and what are the possible transformations that you can perform on the system.

1.1 Bits versus qubits

The basic unit of information in classical information theory is the *bit*, a classical system with two perfectly distinguishable states, denoted by 0 and 1, respectively. Practically, the bit can be realized by different physical systems: for example, it can be realized by an electric circuit, where the state 0 means that the current in the circuit is below a certain threshold, and the state 1 means that the current is above the threshold. Alternatively, the bit can be realized by a magnetic memory, where the state 0 means that the magnetic moment is pointing upwards and the state 1 means that the magnetic moment is pointing downwards.

In quantum information, the bit is replaced by a *quantum bit* or (*qubit*), a quantum system with two perfectly distinguishable states, denoted by $|0\rangle$ and $|1\rangle$, respectively. Also in this case the qubit can be realized by different physical systems. For example, it can be realized but the polarization of a single photon, where the state $|0\rangle$ means that the polarization of the light is vertical, and the state $|1\rangle$ means that the polarization is horizontal. Another physical system that realizes a qubit is a nuclear magnet of spin 1/2, where the state $|0\rangle$ means that the spin is pointing upwards and the state $|1\rangle$ means that it is pointing

downwards.

But besides the fact that we wrote $|0\rangle$ and $|1\rangle$ instead of 0 and 1, what is the difference between bit and qubit?

The key difference is this: The bit has *only* two possible states, 0 and 1. Instead, the quantum bit has many possible states in addition to $|0\rangle$ and $|1\rangle$—as a matter of fact, a qubit has an *infinite number of possible states*! Precisely, the states of a qubit belong to a *two-dimensional complex vector space* and any linear combination of the form

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \qquad \alpha, \beta \in \mathbb{C} \quad |\alpha|^2 + |\beta|^2 = 1 \tag{1.1}$$

represents a possible state. For example, a photon can have not only vertical and horizontal polarization, but also 45 degrees polarization, -45 degrees polarization, and every other possible polarization. In particular, the 45 and -45 degrees polarizations correspond to the states

$$|+\rangle := \frac{|0\rangle + |1\rangle}{\sqrt{2}} \qquad \text{and} \qquad |-\rangle := \frac{|0\rangle + |1\rangle}{\sqrt{2}}\,, \tag{1.2}$$

respectively.

In order to be really precise, I need to clarify a little annoying thing. Every vector of the form Eq. (1.1) is a possible state of a qubit, but not all different vectors represent different states: if two vectors $|\psi\rangle$ and $|\psi'\rangle$ are such that

$$|\psi'\rangle = e^{i\gamma}|\psi\rangle \tag{1.3}$$

for some global phase $\gamma \in [0, 2\pi]$, then they represent the *same state*. For example, the vectors $|0\rangle, -|0\rangle, i|0\rangle$ and $-i|0\rangle$ all represent the same state—e.g. for a photon, the state of vertical polarization.

In order to get rid of this ambiguity, we can parametrize the vectors in Eq. (1.1) as

$$|\psi\rangle = e^{i\gamma}\left[\cos\frac{\theta}{2}\,|0\rangle + \sin\frac{\theta}{2}\,e^{i\varphi}\,|1\rangle\right] \qquad \gamma \in [0, 2\pi]\,, \theta \in [0, \pi]\,, \varphi \in [0, 2\pi]$$

and remove the phase γ. In this way, the states of a qubit are in one-to-one correspondence with the vectors of the form

$$|\psi\rangle = \cos\frac{\theta}{2}\,|0\rangle + \sin\frac{\theta}{2}\,e^{i\varphi}\,|1\rangle \qquad \theta \in [0, \pi]\,, \varphi \in [0, 2\pi] \tag{1.4}$$

Hence, we can visualize the states of a qubit as points on a sphere, called the *Bloch sphere*, as in Figure 1.1. In the Bloch sphere, the state $|0\rangle$ is the North pole and the state $|1\rangle$ is the South pole.

In summary, we can think of the difference between bit and qubit in this way: a bit can be in only two states, corresponding to the North and South pole of the sphere, while a qubit can be in infinitely many states, one state for every point on the sphere. The fact that a qubit has many more states

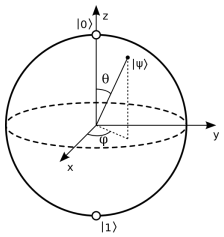

Figure 1.1: **The Bloch sphere** (image from Wikipedia). The possible states of a qubit are in one-to-one correspondence with points on the sphere. In this representation, the states $|0\rangle$ and $|1\rangle$ are the North pole and the South pole, respectively. We can think of a classical bit as a qubit where we use only these two states.

than a bit suggests that in quantum mechanics we have more opportunities for information processing. But does this mean that we extract infinite information from a qubit?

To answer this question, we need to learn more about the rules of quantum mechanics. And in order see the rules, you have first to learn a useful notation, invented by Dirac.

1.2 The Dirac notation

We saw that the states of a qubit are represented by vectors in a two-dimensional complex vector space. This means that we can think of the vector in Eq. (1.1) as a column vector $\begin{pmatrix}\alpha\\ \beta\end{pmatrix} \in \mathbb{C}^2$.

The Dirac notation is a convenient way to represent column vectors, row vectors, and their products. Here is how it works: in the Dirac notation, column vectors are called "ket" and are denoted as $|\psi\rangle$. The mapping between the usual vector notation and the Dirac notation is as follows:

$$\begin{array}{ccccccc} \psi & = \begin{pmatrix}\alpha\\ \beta\end{pmatrix} = & \alpha\begin{pmatrix}1\\ 0\end{pmatrix} & + & \beta\begin{pmatrix}0\\ 1\end{pmatrix} \\ \updownarrow & & \updownarrow & & \updownarrow \\ |\psi\rangle & = & \alpha|0\rangle & + & \beta|1\rangle \end{array}$$

Hence, a generic column vector in $\mathbb{C}^2$ can be written as

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$$

Row vectors are called "bra" and denoted by $\langle\psi|$. The mapping between the

usual vector notation and the Dirac notation is as follows:

$$\begin{array}{ccccccc} \psi^\dagger & = \begin{pmatrix} \overline{\alpha} & \overline{\beta} \end{pmatrix} = & \overline{\alpha} \begin{pmatrix} 1 & 0 \end{pmatrix} & + & \overline{\beta} \begin{pmatrix} 0 & 1 \end{pmatrix} & \\ \updownarrow & & \updownarrow & & \updownarrow & & \\ \langle\psi| & = & \bar{\alpha}\langle 0| & + & \bar{\beta}\langle 1| & \text{(Bra)} \end{array}$$

The scalar product between a row vector $\langle\psi|$ and a column vector $|\psi'\rangle$ is then called "braket" and is denoted $\langle\psi|\psi'\rangle$. The mapping between the usual vector notation and the Dirac notation is as follows:

$$\begin{array}{ll} \psi^\dagger\psi' & = \begin{pmatrix} \overline{\alpha} & \overline{\beta} \end{pmatrix} \begin{pmatrix} \alpha' \\ \beta' \end{pmatrix} = \overline{\alpha}\alpha' + \overline{\beta}\beta' \in \mathbb{C} \\ \updownarrow & \\ \langle\psi|\psi'\rangle & \end{array}$$

The norm of a vector can then be written as $\|\psi\| = \sqrt{\psi^\dagger\psi} = \sqrt{\langle\psi|\psi\rangle}$.

The Dirac notation can be defined not only for vectors in $\mathbb{C}^2$, but also for vectors in $\mathbb{C}^d$ with arbitrary $d < \infty$. The definition is the obvious one: a column vector $\psi \in \mathcal{H}$ is denoted as $|\psi\rangle$ according to the mapping

$$\psi = \begin{pmatrix} \psi_1 \\ \vdots \\ \psi_d \end{pmatrix} \qquad \leftrightarrow \qquad |\psi\rangle = \textstyle\sum_{n=1}^d \psi_n |n\rangle,$$

where $|n\rangle$ denotes the column vector with a 1 in the n-th entry and zeros elsewhere.

Similarly, a row vector $\psi^\dagger$ is denoted as $\langle\psi|$ according to the mapping

$\psi^\dagger = (\bar{\psi}_1, \cdots, \bar{\psi}_d) \qquad \leftrightarrow \qquad \langle\psi| = \sum_{n=1}^d \bar{\psi}_n \langle n|,$

where $\langle n|$ denotes the row vector with 1 in the n-th entry and zeros elsewhere.

The scalar product of two vectors $|\psi\rangle$ and $|\psi\rangle$ is then given by

$$\langle\psi|\psi\rangle = \sum_{n=1}^d \bar{\psi}_n \psi_n \,,$$

and, again, the norm of a vector is given by $\|\psi\| := \sqrt{\langle\psi|\psi\rangle}$.

The Dirac notation is not just a fancy way to write vectors. Throughout this course you will have many occasions to see how much using this notation can simplify your life. For the moment, if you are not already familiar with the Dirac notation, try the following exercise:

Exercise 1 Prove that:

1. $\langle\psi|\psi\rangle = \overline{\langle\psi|\psi\rangle} \qquad \forall|\psi\rangle, |\psi\rangle \in \mathcal{H}$
2. $\langle\psi|\psi\rangle \geqslant 0 \qquad \forall|\psi\rangle \in \mathcal{H} \quad \text{and} \quad \langle\psi|\psi\rangle = 0 \Leftrightarrow |\psi\rangle = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

3. $\langle m|n\rangle = \delta_{mn}$.

Remark. In the following, we will call the complex vector vector space $\mathbb{C}^d$ a *Hilbert space*. Since we are in finite dimension ($d < \infty$), for us "Hilbert space" is just a short name for "complex vector space with a fixed scalar product".

1.3 Quantum systems and pure states

In physics, we say that a system is in a "pure state" is we have "maximal knowledge" about the properties of the system. For example, a pure state of the polarization of a photon is a state where we know exactly *how the photon is polarized* (e.g. for linear polarization, we know the direction of polarization). The first rule of quantum mechanics tells us how to describe a quantum system and its pure states:

Rule 1 (the pure states of a quantum system):
A quantum system is associated to a Hilbert space $\mathcal{H} = \mathbb{C}^d$.
The pure states of the system are represented by unit vectors $|\psi\rangle \in \mathcal{H} \quad \|\psi\| = 1$, up to a global phase—i. e. two unit vectors $|\psi\rangle$ and $|\psi'\rangle$ such that

$$|\psi'\rangle = e^{i\gamma}|\psi\rangle \qquad \gamma \in [0, 2\pi]$$

represent the same pure state.

We already encountered this rule in our introductory discussion about qubits. Now you know the full story for quantum systems of arbitrary dimension d. These systems are usually called *qudits*.

Remark. In physics, it is often useful to consider Hilbert spaces of infinite dimension, describing quantum systems that have infinitely many distinguishable states. However, in this course we focus on finite dimensions.

1.4 Basic measurements

We saw that every quantum system has an infinite set of possible states. Does it mean that we can extract an infinite amount of information from it? Well, unfortunately no. Because in order to extract information we need to perform a *measurement*. What is a measurement in quantum mechanics?

The answer to this question is provided by the second rule of quantum mechanics. But before I tell you the rule, it is useful to recall the definition of orthonormal basis:

Definition 1 (Orthonormal basis) *A set of vectors* $\{|\psi_n\rangle\}_{n=1}^d$ *is an* orthonormal basis (ONB) *for* $\mathcal{H} = \mathbb{C}^d$ *iff*

$$\langle\psi_m|\psi_n\rangle = \delta_{mn} \qquad \forall m, n \in \{1, \ldots, d\}\,. \tag{1.5}$$

For example, the vectors $\{|n\rangle\}_{n=1}^{d}$ are an ONB in dimension d. This basis is called the *computational basis*.

Another example of ONB is given by the vectors $\{|f_n\rangle\}_{n=1}^{d}$ defined by

$$|f_n\rangle := \frac{1}{\sqrt{d}} \sum_{n=1}^{d} \exp\left[\frac{2\pi i m n}{d}\right] |m\rangle \,.$$

This basis is called the *Fourier basis*.

Exercise 2 Verify that the vectors $\{|f_n\rangle\}_{n=1}^{d}$ satisfy the ONB condition of Eq. (1.5).

The computational basis and the Fourier basis are very important in quantum information and you will meet them many times in this course. Note that in dimension $d = 2$ we have

$$|f_1\rangle = \frac{|1\rangle + |2\rangle}{\sqrt{2}} \equiv |+\rangle$$
$$|f_2\rangle = \frac{-|1\rangle + |2\rangle}{\sqrt{2}} \equiv -|-\rangle$$

Hence, up to an irrelevant minus sign, the Fourier basis is the basis $\{|+\rangle, |-\rangle\}$.

We are now ready to give the second rule of quantum mechanics:

Rule 2 (the basic measurements): The basic measurements that can be performed on a quantum system are represented by ONBs on the corresponding Hilbert space.

Each basic measurement has d possible outcomes, corresponding to the vectors of the ONB.

If a system is in the state $|\psi\rangle$ and we measure it on the ONB $\{|\psi_n\rangle\}_{n=1}^{d}$, the probability that we obtain the outcome n is given by the *Born rule*

$$p(n|\,|\psi\rangle) = |\langle\psi_n|\psi\rangle|^2 \,. \tag{1.6}$$

Here we talk about "basic measurements" because in the next chapters we will see other types of measurements. However, all the measurements that we will see can be realized in terms of the basic measurements of Rule 2.

Remark 1 (number of outcomes of a basic measurement). Note that every basic measurement has the same number of outcomes: the number of outcomes is equal to d, the dimension of the Hilbert space. This means that, even if we have an infinite number of pure states, when we make a basic measurement we can only extract a finite amount of information.

Remark 2 (normalization of the probabilities). Rule 1 and Rule 2 guarantee that, if we sum over all possible outcomes of the measurements, the

total probability is equal to 1. Indeed, we have

$$\begin{aligned}\sum_{n=1}^{d} p(n\,||\psi\rangle) &= \sum_{n=1}^{d} |\langle\psi_n|\psi\rangle|^2 && \text{by Rule 1}\\ &= \langle\psi|\psi\rangle \\ &= 1 && \text{by Rule 2}\,.\end{aligned}$$

Remark 3 (global phases do not change the probabilities). Rule 1 tells us that pure states are described by unit vectors, *up to a global phase*. Rule 2 is consistent with this fact: if $|\psi'\rangle = e^{i\gamma}|\psi\rangle$, then $|\psi\rangle$ and $|\psi'\rangle$ give the same probabilities for every possible measurement, i.e. $p(n|\,|\psi'\rangle) = p(n|\,|\psi\rangle)$ for every possible outcome n and for every possible ONB $\{|\psi_n\rangle\}_{n=1}^{d}$.

Remark 4 (a quantum state is identified by the probabilities of all possible measurements). Note that, if two states $|\psi\rangle$ and $|\psi'\rangle$ give the same probabilities for every possible basic measurement, that is, if

$$|\langle\psi_n|\psi\rangle|^2 = |\langle\psi_n|\psi\rangle|^2 \qquad \forall\text{ONB } \{|\psi_n\rangle\}_{n=1}^{d}\,, \tag{1.7}$$

then $|\psi\rangle$ and $|\psi'\rangle$ must be the same quantum state, that is, we must have $|\psi'\rangle = e^{i\gamma}|\psi\rangle$ for some global phase $\gamma \in [0, 2\pi)$.

Exercise 3 Prove that Eq. (1.7) implies that the vectors $|\psi\rangle$ and $|\psi'\rangle$ differ by a global phase.

In other words, if two quantum states are different, then they give different probabilities for at least one basic measurement.

1.5 A physical example

Basic measurements are described by ONBs. Good, but how can one perform these measurements in the lab? The answer depends on the specific quantum system that we are trying to measure, but just to give you an idea, here I tell you how to measure the polarization of a photon on the computational basis $\{|0\rangle, |1\rangle\}$.

Suppose that we have a qubit, encoded in the polarization of a photon. Then, measuring on the computational basis $\{|0\rangle, |1\rangle\}$ means making a measurement of polarization that tests the vertical polarization against the horizontal polarization. Practically, this measurement can be implemented in the following way:

1. take a polarizing filter and align it in the vertical direction
2. put a photodetector after the filter
3. send the photon through the filter

4. if the detector clicks, then report the outcome 0, otherwise report the outcome 1.

In this physical setup, if at the beginning the photon is polarized vertically (i.e. if it is in the state $|0\rangle$), then it will pass the filter. Hence, the detector will click and the measurement will give outcome 0. Instead, if the photon is polarized horizontally (i.e. if it is in the state $|1\rangle$), then the filter will block it, the detector will not click, and the outcome will be 1. This is in agreement with the Born rule, which gives

$$p(0|\,|0\rangle) = p(0|\,|1\rangle) = 1$$
$$p(1|\,|0\rangle) = p(0|\,|1\rangle) = 0\,.$$

The situation is described in Figure 1.5.

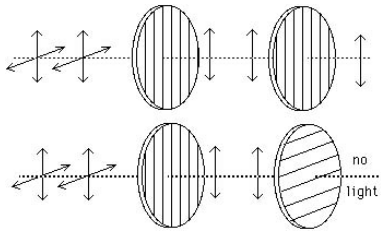

Figure 1.2: **Polarization measurement with a linear polarization filter in the vertical direction** (image from http://www.eiu.edu/ ddavis/chapter_16/ch16_6.htm). (Top) A first filter prepares photons with vertical polarization. Then, a vertically polarized photon passes through the filter and hits a photon detector. The detector clicks, thus triggering the outcome 0. (Bottom) Vertically polarized photons are blocked by the a filter in the horizontal direction. In the same way, a horizontally polarized photon is blocked by the vertical filter in the top: in this case, the photon detector does not click, thus triggering the outcome 1.

We have seen that, if the polarization of the photon is vertical or horizontal, then the outcome of the measurement in the basis $\{|0\rangle, |1\rangle\}$ can be predicted with certainty. But what happens if the photon is polarized in another direction, like 45 degrees? For a photon polarized at 45 degrees, the Born rule gives

$$p(0|\,|+\rangle) = p(1|\,|+\rangle) = \frac{1}{2}.$$

This means that the outcome of our measurement is completely random. Quantum mechanics does not tell us what the outcome will be, it only tells us that, if we repeat the experiment a large number of times, we will get the outcome 0 approximately half of the times (and the outcome 1 in the remaining cases).

The same situation takes place if we send a vertically polarized photon through a filter that is oriented at some angle θ, as in figure 1.5. In this case,

the Born rule tells us that the probability of a detector click after the filter is given by $\cos^2(\theta/2)$. By varying the angle θ we can then measure on all the ONBs corresponding to linear polarizations.

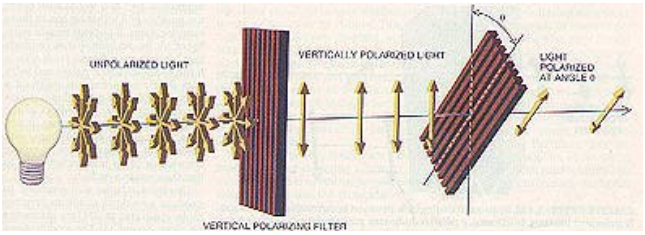

Figure 1.3: A vertically polarized photon passes through filter oriented at an angle θ (image source: http://www.csa.com/discoveryguides/crypt/overview.php)

1.6 What we have learnt so far

With Rules 1 and 2, you already learnt many important things about quantum mechanics. Let us highlight them:

1. **The intrinsic randomness in quantum measurements.** If we prepare a qubit in the state $|+\rangle$ and we measure it in the computational basis $\{|0\rangle, |1\rangle\}$, then we find that the result is **random**! Even if the system was in a pure state! This means that even if you have maximal knowledge of the system, you will not be able to predict the outcome of the measurement. See Figure 1.5 for the illustration in the example of a polarization of a single photon. This idea is the working principle of **quantum random-number generators**: random numbers can be generated by preparing the system in state $|+\rangle$ and measuring it in the computational basis.

2. **The maximum number of perfectly distinguishable states.** By definition, the states in an ONB can be distinguished perfectly using the corresponding measurement: *if* the system is in one of the states $\{|\psi_n\rangle\}_{n=1}^d$, then we can find out exactly what the state is, by measuring on the ONB $\{|\psi_n\rangle\}_{n=1}^d$. In other words, a ONB is a **maximal set** of **perfectly distinguishable** pure states. "Maximal" because in a d-dimensional Hilbert space, you cannot distinguish perfectly more than d pure states in one shot: your basic measurements have only d outcomes!

3. **Bohr's complementarity.** Different ONBs are complementary ways to look at a quantum system: if you measure in one basis, you lose the opportunity to find out what the outcome would have been *if* you had measured in another basis. For example, suppose that we know that our qubit is either in the state $|+\rangle$ or in the state $|-\rangle$. By measuring on the Fourier basis $\{|+\rangle, |-\rangle\}$ we can find out if the state is $|+\rangle$ or $|-\rangle$. But if

we measure in the computational basis $\{|0\rangle, |1\rangle\}$ we will never be able to find out if the state was $|+\rangle$ or $|-\rangle$.

4. **Mutually unbiased bases.** The two ONBs $\{|0\rangle, |1\rangle\}$ and $\{|+\rangle, |-\rangle\}$ have a special property:

   - If the state is $|0\rangle$(or $|1\rangle$) the outcome of the measurement on $\{|+\rangle, |-\rangle\}$ is completely random.
   - If the state is $|+\rangle$(or $|-\rangle$) the outcome of the measurement on $\{|0\rangle, |1\rangle\}$ is completely random.

   i.e. if we know the polarization in one direction, then we must be completely ignorant about the outcome of a polarization measurement in another direction.

   In general, two ONBs $\{|\psi_m\rangle\}_{m=1}^d$ and $\{\psi_m\}_{n=1}^d$ are called *mutually unbiased* iff one has

   $$|\langle\psi_m|\psi_n\rangle| = \frac{1}{d} \qquad \forall m, n \in \{1, \ldots, d\}\,.$$

   Exercise 4 Prove that the computational basis and the Fourier basis are mutually unbiased in every dimension.

1.7 Reversible transformations

Until now, we talked about the possible states of a quantum system, and about the possible measurements that we can perform. It remans to say what are the possible transformations that we can perform on the state of a quantum system.

In physics, a time evolution T is *reversible* if there exists another time evolution T^{-1} such that $T^{-1}T = I$ and $TT^{-1} = I$, where I is the identity transformation, that leaves every state unchanged.

$$\begin{array}{ccc} \text{Input state} & \longrightarrow & \text{Output state} \\ |\psi\rangle \in \mathcal{H} & \text{Time evolution } T & |\psi\rangle = T|\psi\rangle \end{array}$$

Now, by Rule 1 we know that the pure states of a system are unit vectors in a Hilbert space: hence, a reversible transformation must be represented by an invertible operator that sends unit vectors to unit vectors. The only linear operators with this property are are the *unitary operators*:

Definition 2 (Unitary operator) *A linear operator $U : \mathcal{H} \to \mathcal{H}$ is* **unitary** *iff*

$$U^\dagger U = UU^\dagger = I$$

with I the identity operator ($I|\psi\rangle = |\psi\rangle \quad \forall|\psi\rangle \in \mathcal{H}$).

Recall the definition of the adjoint $A^\dagger$ for an operator:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots \\ a_{21} & a_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \qquad \overline{A} = \begin{pmatrix} \overline{a_{11}} & \overline{a_{12}} & \cdots \\ \overline{a_{21}} & \overline{a_{22}} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

$$A^T = \begin{pmatrix} a_{11} & a_{21} & \cdots \\ a_{12} & a_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \qquad A^\dagger = \left(\overline{A}\right)^T = \overline{(A^T)}$$

The condition of unitarity can be schematically represented as follows: for every state $|\psi\rangle \in \mathcal{H}$ one has

$$|\psi\rangle \xrightarrow{U} U|\psi\rangle \xrightarrow{U^\dagger} |\psi\rangle$$

$$|\psi\rangle \xrightarrow{U^\dagger} U^\dagger|\psi\rangle \xrightarrow{U} |\psi\rangle$$

We then have the following rule:

Rule 3 (the reversible evolutions): The reversible evolutions of a quantum system are described by **unitary operators** on the corresponding Hilbert space.

Reversible evolutions are also called *reversible gates* in quantum information.

Examples
Reversible qubit gates:

- Bit flip (a.k.a. NOT gate): $X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ $X|0\rangle = |1\rangle$ $X|1\rangle = |0\rangle$.
- Phase flip: $Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ $Z|0\rangle = |0\rangle$ $Z|1\rangle = -|1\rangle$.
- Hadamard gate (a.k.a. Fourier transform): $H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ $H|0\rangle = |+\rangle$ $H|1\rangle = |-\rangle$.

Exercise 5 Prove that:

1. X, Z and H are unitary.
2. $X^\dagger = X$, $Z^\dagger = Z$, $H^\dagger = H$.
3. $XZ = -ZX$.
4. $Z = HXH$.

1.8 Dirac notation for operators

In order to write down unitary operators more efficiently, it is good to make a little parenthesis on the Dirac notation for operators. This is actually the situation where the Dirac notation is most powerful.

For two ket vectors $|\alpha\rangle$ and $|\beta\rangle$ we can define an operator as follows:

$$|\alpha\rangle = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_d \end{pmatrix} \in \mathcal{H}_A \quad |\beta\rangle = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_d \end{pmatrix} \in \mathcal{H}_B$$

$$|\beta\rangle\langle\alpha| := \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_d \end{pmatrix} \begin{pmatrix} \overline{\alpha_1} & \cdots & \overline{\alpha_d} \end{pmatrix} = \begin{pmatrix} \beta_1\overline{\alpha_1} & \beta_1\overline{\alpha_2} & \cdots \\ \beta_2\overline{\alpha_1} & \beta_2\overline{\alpha_2} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}.$$

$|\beta\rangle\langle\alpha|$ is an operator from $\mathcal{H}_A$ to $\mathcal{H}_B$.

To become more proficient with the Dirac notation, we suggest you this exercise:

Exercise 6 If $A = |\beta\rangle\langle\alpha|$, prove that

1. $A|\psi\rangle = |\beta\rangle\langle\alpha|\psi\rangle \quad \forall|\psi\rangle \in \mathcal{H}_A$
2. $\overline{A} = |\overline{\beta}\rangle\langle\overline{\alpha}| \quad |\overline{\alpha}\rangle = \begin{pmatrix} \overline{\alpha_1} \\ \vdots \\ \overline{\alpha_d} \end{pmatrix} \quad |\overline{\beta}\rangle = \begin{pmatrix} \overline{\beta_1} \\ \vdots \\ \overline{\beta_d} \end{pmatrix}$
3. $A^T = |\overline{\alpha}\rangle\langle\overline{\beta}|$
4. $A^\dagger = |\alpha\rangle\langle\beta|$
5. $\sum_{n=1}^d |n\rangle\langle n| = I$
6. for every orthonormal basis $\{|\psi_n\rangle\}_{n=1}^d$, one has $\sum_{n=1}^d |\psi_n\rangle\langle\psi_n| = I$

1.9 General form of a unitary

Using the Dirac notation for operators, we can now see what is the general form of a unitary gate in every dimension.

Property 1 A linear operator $U : \mathcal{H} \to \mathcal{H}$ is unitary iff

$$\exists \text{ ONBs } \{|\psi_n\rangle\}_{n=1}^d \text{ and } \{|\psi_n\rangle\}_{n=1}^d \quad \text{such that} \quad U = \sum_{n=1}^d |\psi_n\rangle\langle\psi_n|.$$

The physical meaning of this result is that we can always transform a maximal set of distinguishable pure states in another one with a reversible gate.

Examples of unitary gates in higher dimensions:

1. Shift operator: $S = \sum_{n=1}^{d} |(n+1) \mod d\rangle\langle n| \quad S|1\rangle = |2\rangle \cdots S|d\rangle = |1\rangle$. For qubits ($d = 2$), it is just the bit flip X.

2. Multiply operator: $M = \sum_{n=1}^{d} \exp\left[\frac{2\pi i}{d} n\right] |n\rangle\langle n|$. For qubits, it is just the phase flip Z

3. Fourier transform: $F = \sum_{n=1}^{d} |f_n\rangle\langle n| \quad |f_n\rangle = \frac{1}{\sqrt{d}} \sum_{m=1}^{d} \exp\left[\frac{2\pi i}{d} mn\right] |m\rangle$. For qubits, it is just the Hadamard gate.

4. Reduced oracles for Boolean functions. A Boolean functions is just a function that outputs a bit, $\quad f : \{1, 2, \ldots, d\} \to \{0, 1\}$. To every Boolean function f, one can associate a unitary gate V_f, in the following way:

$$V_f = \sum_{n=1}^{d} (-1)^{f(n)} |n\rangle\langle n|.$$

   You can easily check that this is a unitary operator.

   The gate V_f is usually called the *reduced oracle* for the function f. In many quantum algorithms, the evaluation of the function f is replaced by one use of the gate V_f.

1.10 A little bit of fun

Using the ideas of this chapter, we can already enjoy the basic part of one of the first quantum algorithms: the Deutsch-Jozsa algorithm.

Let us see which problem the algorithm tries to solve:

> Problem: You are given a black box that computes for you a Boolean function $f : \{1, \ldots, d\} \to \{0, 1\}$, with d even. You don't know the function, but you know that f can be either *constant* or *balanced*.
>
> ["constant" means $f(n) = b \quad \forall n = \{1, \ldots, d\}$ with $b = 0$ or $b = 1$.
>
> "balanced" means that the number of inputs such that $f(n) = 0$ is equal to the number of inputs such that $f(n) = 1$.]
>
> You want to find out which of the two properties holds: Is f constant, or is f balanced?

The best classical algorithm for this problem requires to evaluate the function $\frac{d}{2}$ times (in the worst case situation): until you see two different values on two different inputs, the function could always be constant. But if someone constructs V_f for you, the Deutsch-Jozsa algorithm can give you the answer with only 1 **use** of V_f.

How does this work?

Deutsch-Jozsa algorithm

1. Prepare the Fourier state $|f_d\rangle = \frac{1}{\sqrt{d}} \sum_{n=1}^{d} |n\rangle$.

2. Apply V_f:
   If f is constant($f(n) = b \quad \forall n$), then $V_f|f_d\rangle$ is **proportional** to $|f_d\rangle$:

   $$V_f|f_d\rangle = (-1)^b|f_d\rangle.$$

   If f is balanced, then $V_f|f_d\rangle$ is **orthogonal** to $|f_d\rangle$:

   $$\langle f_d|V_f|f_d\rangle = \frac{1}{d}\sum_{n=1}^{d}(-1)^{f(n)} = 0.$$

3. Measure the Fourier basis $\{|f_n\rangle\}_{n=1}^{d}$:
   If the outcome is "d", then f is constant.
   If the outcome is not "d", then f is balanced.

Of course, you may ask what V_f has to do with f: *after all, using the gate V_f is not the same thing as using the function f!* To know in what sense V_f is the "quantum version of f" you will have to wait for the next chapter.

1.11 Chapter summary

In this chapter you learnt the three basic rules to describe a single quantum system. We can now summarize these rules in a sort of *"Rosetta stone"* where one the left you have physical notions and on the right you have their translation in the mathematical language of quantum theory:

A quantum Rosetta stone	
Physics	**Mathematics**
Quantum system	Hilbert space
Pure state	Unit vector, up to a global phase
Basic measurement	Orthonormal basis
Probability of outcomes for a given measurement	Born rule with the corresponding basis
Reversible transformation	Unitary operator

Did you think that quantum theory was hard to learn? If you did, it is time to take heart: with this chapter, you learnt already almost all of it! *Almost,* because there is still a fundamental ingredient missing: how to compose two systems together. This will be the subject of the next chapter.

Chapter 2

Composite Systems

2.1 How to describe a composite system?

Suppose that two physicists, Alice and Bob, have two laboratories, and in their laboratories they are making experiments on two quantum systems, say A for Alice's system and B for Bob's system. For example, A and B could be the polarizations of two photons. We want to consider the two systems together as a *composite system* and see what Alice and Bob can do with it. But how to describe a composite quantum system? From the previous chapter, we know that quantum systems are associated to Hilbert spaces. Hence, the question is: What is the Hilbert space of the composite system AB? And how does it relate to the Hilbert spaces of the individual systems A and B?

You will see the precise mathematical answer in the next two sections. Before looking at the answer, however, try to ask yourself which properties should be satisfied by the Hilbert space of a composite system. Think of what Alice and Bob can do in their laboratories:

1. First of all, Alice and Bob can always prepare their systems independently: Alice can prepare system A is in the state $|\alpha\rangle \in \mathcal{H}_A$ and Bob can prepare system B is in the state $|\beta\rangle \in \mathcal{H}_B$. Hence, the Hilbert space $\mathcal{H}_{AB}$ should contain a state that describes this situation. To give a name to this state, let us call it $|\alpha\rangle|\beta\rangle$.

2. Second, Alice and Bob can always measure their systems independently, performing a basic measurement on A and a basic measurement on B. Representing the two basic measurements with the corresponding ONBs $\{|\alpha_m\rangle\}_{m=1}^{d_A}$ and $\{|\beta_n\rangle_{n=1}^{d_B}\}$, we have that the vectors $\{|\alpha_m\rangle|\beta_n\rangle \mid m = 1, \dots d_A,\, n = 1, \dots, d_B\}$ must represent a valid measurement of the composite system AB. Going a little further, we can ask that these vectors represent a *basic* measurement, that is, that they are an ONB for the Hilbert space $\mathcal{H}_{AB}$.

3. Alice and Bob can always prepare *and* measure their systems independently. If they do that, we expect that the outcomes of their measurements will be uncorrelated. This means that, if Alice and Bob prepare the states of point 1 and perform the basic measurements of point 2, then the probability that Alice gets outcome m and Bob gets outcome n must be of the product form

$$p_{AB}(m,n) = p_A(m)\, p_B(n)\,,$$

where $p_A(m)$ and $p_B(n)$ are given by the Born rule $p_A(m) = |\langle \alpha_m|\alpha\rangle|^2$ and $p_B(n) = |\langle \beta_n|\beta\rangle|^2$.

Note that the second requirement implies that the dimension of $\mathcal{H}_{AB}$ is the product of the dimensions of $\mathcal{H}_A$ and $\mathcal{H}_B$. In other words, if $\mathcal{H}_A \simeq \mathbb{C}^{d_A}$ and $\mathcal{H}_B \simeq \mathbb{C}^{d_B}$, then $\mathcal{H}_{AB} \simeq \mathbb{C}^{d_A d_B}$. We will now see a concrete way to construct the Hilbert space $\mathcal{H}_{AB}$ from $\mathcal{H}_A$ and $\mathcal{H}_B$.

2.2 The tensor product Hilbert space

Here we introduce the *tensor product* of two Hilbert spaces. For many of you, this will be a new piece of math.

Let us start from the case of qubits: $\mathcal{H}_A \simeq \mathcal{H}_B \simeq \mathbb{C}^2$. This is the case, for example, when Alice and Bob have one photon each. The tensor product Hilbert space, denoted by $\mathcal{H}_A \otimes \mathcal{H}_B$ is the Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B \simeq \mathbb{C}^4$, equipped with a rule that associates every pair of column vectors $|\alpha\rangle \in \mathcal{H}_A$ and $|\beta\rangle \in \mathcal{H}_B$ to a column vector $|\alpha\rangle \otimes |\beta\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$: precisely, the rule associates the vectors

$$|\alpha\rangle = \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} \qquad\qquad |\beta\rangle = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

to their tensor product

$$|\alpha\rangle \otimes |\beta\rangle := \begin{pmatrix} \alpha_0\beta_0 \\ \alpha_0\beta_1 \\ \alpha_1\beta_0 \\ \alpha_1\beta_1 \end{pmatrix} . \tag{2.1}$$

For example, for the computational basis vectors $|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $|1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ one has

$$|0\rangle \otimes |0\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \qquad |0\rangle \otimes |1\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \qquad |1\rangle \otimes |0\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \qquad |1\rangle \otimes |1\rangle = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} .$$

Clearly, these vectors form an ONB. Precisely, they form the computational basis for $\mathcal{H}_A \otimes \mathcal{H}_B \simeq \mathbb{C}^4$. Hence, every vector in $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ can be

expanded as

$$|\Psi\rangle = \sum_{m,n=0,1} \Psi_{mn} \, |m\rangle \otimes |n\rangle \, .$$

Let us now define the tensor product of two Hilbert spaces of general dimension $\mathcal{H}_A \simeq \mathbb{C}^{d_A}$ and $\mathcal{H}_B \simeq \mathbb{C}^{d_B}$. The definition is exactly the same as for qubits: the tensor product Hilbert space is the Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B \simeq \mathbb{C}^{d_A \cdot d_B}$, equipped with the rule that associates the pair of vectors

$$|\alpha\rangle = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_{d_A} \end{pmatrix} \qquad\qquad |\beta\rangle = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_{d_B} \end{pmatrix}$$

to their tensor product

$$|\alpha\rangle \otimes |\beta\rangle := \begin{pmatrix} \alpha_1\beta_1 \\ \vdots \\ \alpha_1\beta_{d_B} \\ \vdots \\ \alpha_{d_A}\beta_1 \\ \vdots \\ \alpha_{d_A}\beta_{d_B} \end{pmatrix} .$$

Like in the case of qubits, one can construct the computational basis for $\mathcal{H}_A \otimes \mathcal{H}_B$ from the computational bases of $\mathcal{H}_A$ and $\mathcal{H}_B$ and expand an arbitrary vector $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ as

$$|\Psi\rangle = \sum_{m=1}^{d_A} \sum_{n=1}^{d_B} \Psi_{mn} \, |m\rangle \otimes |n\rangle \, .$$

It is also useful to introduce a notation for the tensor product of two row vectors, and, more generally, for the tensor product of two operators. Given two vectors $|\alpha\rangle \in \mathcal{H}_A$ and $|\beta\rangle \in \mathcal{H}_B$, we define $\langle\alpha| \otimes \langle\beta|$ to be the row vector

$$\langle\alpha| \otimes \langle\beta| := \left(\, |\alpha\rangle \otimes |\beta\rangle \, \right)^\dagger$$

For example, for the vectors $|\alpha\rangle = \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix}$ and $|\beta\rangle = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$ one has

$$\langle\alpha| \otimes \langle\beta| = \begin{pmatrix} \alpha_0\beta_0 \\ \alpha_0\beta_1 \\ \alpha_1\beta_0 \\ \alpha_1\beta_1 \end{pmatrix}^\dagger = \begin{pmatrix} \bar{\alpha}_0\bar{\beta}_0 & \bar{\alpha}_0\bar{\beta}_1 & \bar{\alpha}_1\bar{\beta}_0 & \bar{\alpha}_1\bar{\beta}_1 \end{pmatrix}$$

Note that by definition, we have the property

$$(\,\langle\alpha|\otimes\langle\beta|\,)\,(\,|\alpha'\rangle\otimes|\beta'\rangle\,)=\langle\alpha|\alpha'\rangle\,\langle\beta|\beta'\rangle \qquad\qquad \forall|\alpha\rangle,|\alpha'\rangle\in\mathcal{H}_A\,\forall|\beta\rangle,|\beta'\rangle\in\mathcal{H}_B\,.$$

Of course, when you do a calculation it is much better to use this property, rather than doing brute-force manipulations of the matrix elements!

Let us define now the tensor product of two operators. Suppose that A is a linear operator transforming vectors in $\mathcal{H}_A$ into vectors in $\mathcal{H}_{A'}$ and B is an operator transforming vectors in $\mathcal{H}_B$ into vectors in $\mathcal{H}_{B'}$. Then, $A\otimes B$ is a linear operator transforming vectors in $\mathcal{H}_A\otimes\mathcal{H}_B$ into vectors in $\mathcal{H}_{A'}\otimes\mathcal{H}_{B'}$ and is defined by the relation:

$$(A\otimes B)(|\alpha\rangle\otimes|\beta\rangle):=A|\alpha\rangle\otimes B|\beta\rangle \qquad \forall|\alpha\rangle\in\mathcal{H}_A \quad \forall|\beta\rangle\in\mathcal{H}_B.$$

This relation defines $A\otimes B$ on *every* vector: indeed, for a general vector $|\Psi\rangle=\sum_{m,n}\Psi_{mn}|m\rangle\otimes|n\rangle$ we have

$$\begin{array}{rcl}(A\otimes B)|\Psi\rangle & = & \sum_{m,n}\Psi_{mn}(A\otimes B)(|m\rangle\otimes|n\rangle)\\ & \uparrow & \\ & \text{by} & \text{by linearity}\\ & & \\ & = & \sum_{m,n}\Psi_{mn}\,A|m\rangle\otimes B|n\rangle\,.\\ & \uparrow & \\ & \text{by} & \text{by definition of } A\otimes B \text{ on product vectors.}\end{array}$$

For example, the tensor product of the matrices $A=\begin{pmatrix}A_{00} & A_{01}\\ A_{10} & A_{11}\end{pmatrix}$ and $B=\begin{pmatrix}B_{00} & B_{01}\\ B_{10} & B_{11}\end{pmatrix}$ is the matrix

$$A\otimes B=\begin{pmatrix}A_{00}B_{00} & A_{00}B_{01} & A_{01}B_{00} & A_{01}B_{01}\\ A_{00}B_{10} & A_{00}B_{11} & A_{01}B_{10} & A_{01}B_{11}\\ A_{10}B_{00} & A_{10}B_{01} & A_{11}B_{00} & A_{11}B_{01}\\ A_{10}B_{10} & A_{10}B_{11} & A_{11}B_{10} & A_{11}B_{11}\end{pmatrix}.$$

To become more familiar with the basic properties of tensor products, try the following exercises. By knowing the following properties you will be able to avoid doing boring calculations with matrix elements.

Exercise 7 (Tensor product of vectors) Check the following properties:

1. For every vectors $|\alpha\rangle,|\alpha'\rangle\in\mathcal{H}_A$, $|\beta\rangle\in\mathcal{H}_B$, one has $(|\alpha\rangle+|\alpha'\rangle)\otimes|\beta\rangle=|\alpha\rangle\otimes|\beta\rangle+|\alpha'\rangle\otimes|\beta\rangle$.

2. For every vectors $|\alpha\rangle\in\mathcal{H}_A$, $|\beta\rangle,|\beta'\rangle\in\mathcal{H}_B$, one has $|\alpha\rangle\otimes(|\beta\rangle+|\beta'\rangle)=|\alpha\rangle\otimes|\beta\rangle+|\alpha\rangle\otimes|\beta'\rangle$.

3. For every vectors $|\alpha\rangle\in\mathcal{H}_A$, $|\beta\rangle\in\mathcal{H}_B$ and for every number $\lambda\in\mathbb{C}$, one has $(\lambda|\alpha\rangle)\otimes|\beta\rangle=|\alpha\rangle\otimes(\lambda|\beta\rangle)=\lambda\,|\alpha\rangle\otimes|\beta\rangle$.

4. If $\{|\alpha_m\rangle\}_{m=1}^{d_A}$ and $\{|\beta_n\rangle\}_{n=1}^{d_B}$ are two ONBs for $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively, then $\{|\alpha_m\rangle \otimes |\beta_n\rangle \mid m = 1, \dots d_A,\, n = 1, \dots, d_B\}$ is a ONB for $\mathcal{H}_A \otimes \mathcal{H}_B$.

Exercise 8 (Tensor product of operators) Let A be an operator from $\mathcal{H}_A$ to $\mathcal{H}_{A'}$ and B be an operator from $\mathcal{H}_B$ to $\mathcal{H}_{B'}$. Check the following properties:

1. $(\langle\alpha| \otimes \langle\beta|)(A \otimes B)(|\alpha'\rangle \otimes |\beta'\rangle) = \langle\alpha|A|\alpha'\rangle\langle\beta|B|\beta'\rangle$ for every vectors $|\alpha\rangle \in \mathcal{H}_A, |\alpha'\rangle \in \mathcal{H}_{A'}$, $|\beta\rangle \in \mathcal{H}_B\,, |\beta'\rangle \in \mathcal{H}_{B'}$
2. $(A \otimes B)^\dagger = A^\dagger \otimes B^\dagger$
3. For $\mathcal{H}_A = \mathcal{H}_{A'}$ and $\mathcal{H}_B = \mathcal{H}_{B'}$, one has $(A \otimes B)(A' \otimes B') = AA' \otimes BB'$. In particular, $(A \otimes I_B)(I_A \otimes B) = A \otimes B = (I \otimes B)(A \otimes I)$(operators on different Hilbert spaces commute).
4. If U_A and U_B are unitaries, then $U_A \otimes U_B$ is unitary.

2.3 The rule for describing composite systems

We have now all the ingredients we need to describe a composite quantum system:

Rule 4 (Composite systems): Let $\mathcal{H}_A$ and $\mathcal{H}_B$ be the Hilbert spaces of two quantum systems A and B. Then, the Hilbert space of the composite system AB is $\mathcal{H}_{AB} := \mathcal{H}_A \otimes \mathcal{H}_B$.

If system A is in the state $|\alpha\rangle \in \mathcal{H}_A$ and system B is in the state $|\beta\rangle \in \mathcal{H}_B$, then the composite system AB is in the state $|\alpha\rangle|\beta\rangle := |\alpha\rangle \otimes |\beta\rangle$.

If system A undergoes the basic measurement $\{|\alpha_m\rangle\}_{m=1}^{d_A}$ and system B undergoes the basic measurement $\{|\beta_n\rangle\}_{n=1}^{d_B}$, then system AB undergoes the basic measurement $\{|\alpha_m\rangle|\beta_n\rangle \mid m = 1, \dots d_A,\, n = 1, \dots, d_B\}$.

If system A undergoes the reversible gate U_A and system B undergoes the reversible gate U_B, then system AB undergoes the reversible gate $U_A \otimes U_B$

This rule satisfies the all intuitive requirements that we introduced at the beginning of the chapter. In particular, if Alice and Bob prepare their systems in the states $|\alpha\rangle$ and $|\beta\rangle$, and perform the basic measurements $\{|\alpha_m\rangle\}_{m=1}^{d_A}$ and $\{|\beta_n\rangle\}_{n=1}^{d_B}$, then they get the outcomes (m, n) with probability $p_{AB}(m, n) = |\langle\alpha_m|\alpha\rangle|^2 \, |\langle\beta_n|\beta\rangle|^2$. When Alice and Bob perform independent experiments in their labs, the outcomes are uncorrelated, pretty much like the outcomes that you get when you roll two dices.

Is everything clear about Rule 4 and Rules 1-3 from the previous chapter? If yes, congratulations: now you know everything about the basics of quantum theory! What we will do in the rest of the course will be only to understand more and more deeply the consequences of these four rules.

Remark (tensor product of more than two systems). We have seen how to describe the tensor product of two systems. The extension of Rule 4 to more than two systems is immediate: if you have k systems A_1, A_2, ..., A_k, the composite system is described by the Hilbert space

$$\mathcal{H}_{A_1 A_2 \ldots A_k} := \mathcal{H}_{A_1} \otimes \mathcal{H}_{A_2} \otimes \cdots \otimes \mathcal{H}_{A_k} \,. \tag{2.2}$$

A pure state of k systems can be written as a linear combination

$$|\Psi\rangle = \sum_{n_1=1}^{d_{A_1}} \sum_{n_2=1}^{d_{A_2}} \cdots \sum_{n_k=1}^{d_{A_k}} \Psi_{n_1 n_2 \ldots n_k} \, |n_1\rangle \otimes |n_2\rangle \otimes \cdots \otimes |n_k\rangle \,.$$

Note that in order to specify a pure state one needs $d_{A_1} d_{A_2} \cdots d_{A_k}$ complex numbers—for example, if all the systems are qubits one needs 2^k complex numbers, a number which grows exponentially with the number of qubits.

2.4 Product states vs entangled states

Until now, we met the states $|\alpha\rangle|\beta\rangle$, which describe the situation "system A is in the state $|\alpha\rangle$ and system B is in the state $|\beta\rangle$". These states are called *product states*. By definition, when the system AB is in a product state, we can associate a pure state to the subsystems A and B. For example, suppose that the states $|0\rangle$ and $|1\rangle$ represent the vertical and horizontal polarization of a photon, respectively. If two photons are in the product state $|0\rangle|0\rangle$, then we can say that each photon has vertical polarization. Likewise, if the photons are in the state $|1\rangle|1\rangle$, then we can say that each photon has horizontal polarization.

However, the Hilbert space contains also many other states: for example, it contains the state

$$|\Phi^+\rangle = \frac{|0\rangle|0\rangle + |1\rangle|1\rangle}{\sqrt{2}} \,.$$

What does *this* represent? It is an easy exercise to check that $|\Phi^+\rangle$ cannot be seen as the product of two states: it is impossible to find two states $|\alpha\rangle$ and $|\beta\rangle$ such that $|\Phi^+\rangle = |\alpha\rangle|\beta\rangle$. In other words, when two photons are in the state $|\Phi^+\rangle$ there is no way to assign a pure state to each photon. This is very surprising! On the one hand, the state $|\Phi^+\rangle$ is *pure*: we know everything we could possibly know about the composite system AB. On the other hand, the system A alone is not in a pure state. And the system B alone is not in a pure state either.

Many pioneers of quantum mechanics have been fascinated by the fact that the theory predicts the existence of states like $|\Phi^+\rangle$. For example, in a famous 1935 paper Schrödinger wrote:

"Maximal knowledge of a total system does not necessarily include total knowledge of all its parts, not even when these are fully separated from each other and at the moment are not influencing each other at all."

In the same paper, Schrödinger introduced the term *entangled*, to indicate the states like $|\Phi^+\rangle$. Precisely, we say that a pure state $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ is *entangled* if there is no way to write it as $|\Psi\rangle = |\alpha\rangle|\beta\rangle$.

The existence of pure entangled states is one of the most profound features of quantum mechanics. In another part of the paper, Schrödinger wrote, speaking about entanglement: *"I would not call that* one *but rather* the *characteristic trait of quantum mechanics, the one that enforces its entire departure from classical lines of thought."*

As we will see in this course, entangled states have many magical properties, and are the key ingredient for a huge number of applications, such as secure quantum cryptography, quantum games, quantum teleportation, and many others.

Entangled states are so surprising that you may wonder whether they really exist in nature. Can we generate entangled states in the lab? The answer is *yes*: the generation of entangled states has been demonstrated for all the implementations of quantum information (photons, NMR, ions, superconductors). For example, one way to generate entanglement in the quantum optics lab is to use the phenomenon of *spontaneous parametric down conversion (SPDC)*. In this phenomenon, a photon of higher energy passing through a non-linear crystal splits into two photons of lower energy. When the splitting takes place, the polarizations of the two photons are in the entangled state $|\Phi^+\rangle$.

Let us see a curious property of the state $|\Phi^+\rangle$. Suppose that Alice and Bob have two photons in the state $|\Phi^+\rangle$ and they both make a measurement of polarization, putting their polarizing filters in the vertical direction. Mathematically, this means that they are both measuring on the computational basis $\{|0\rangle, |1\rangle\}$. Using the Born rule, we can compute the probabilities of their outcomes:

$$\begin{aligned} p_{AB}(0,0) &= |\langle 0|\langle 0||\Phi^+\rangle|^2 = \frac{1}{2} \\ p_{AB}(1,1) &= |\langle 1|\langle 1||\Phi^+\rangle|^2 = \frac{1}{2} \\ p_{AB}(0,1) &= |\langle 0|\langle 1||\Phi^+\rangle|^2 = 0 \\ p_{AB}(1,0) &= |\langle 1|\langle 0||\Phi^+\rangle|^2 = 0\,. \end{aligned}$$

The outcome of each measurement is completely random: both Alice and Bob can find vertical or horizontal polarization with probability 1/2. However, Alice's and Bob's outcomes are perfectly correlated. Whenever Alice's photon passes through the filter, also Bob's photon will pass. Whenever Alice's photon is absorbed by the filter, Bob's photon will be absorbed too. This is like having two coins that are perfectly correlated: when one coin gives "heads" also the the other coin gives "heads", when one gives "tails" also the other gives "tails".

If you are not surprised enough, try to see what happens if Alice and Bob put their polarizers in the 45 degrees direction. Mathematically, this means that they are both measuring on the Fourier basis $\{|+\rangle, |-\rangle\}$. The probabilities of

their outcomes are:

$$p_{AB}(+,+) = |\langle +|\langle +||\Phi^+\rangle|^2 = \frac{1}{2}$$
$$p_{AB}(-,-) = |\langle -|\langle -||\Phi^+\rangle|^2 = \frac{1}{2}$$
$$p_{AB}(+,-) = |\langle +|\langle -||\Phi^+\rangle|^2 = 0$$
$$p_{AB}(-,+) = |\langle -|\langle +||\Phi^+\rangle|^2 = 0\,.$$

Again, the outcome of each measurement is completely random, but there is a perfect correlation between the outcome on Alice's side and the outcome on Bob's side!

Finally, suppose that Alice and Bob put their polarizers at an arbitrary angle θ, measuring on the basis $\{|0,\theta\rangle, |1,\theta\rangle\}$, defined by

$$|0,\theta\rangle = \cos(\theta/2)|0\rangle + \sin(\theta/2)|1\rangle$$
$$|1,\theta\rangle = -\sin(\theta/2)|0\rangle + \cos(\theta/2)|1\rangle\,.$$

Guess what happens to the outcomes of their measurement? Again, they are random, but perfectly correlated!

This fact is quite puzzling. When Alice and Bob make their measurements, the two photons can be very far from each other. We can even imagine that Alice and Bob make their measurement at the same time. And yet—almost magically—for every direction of their polarizers, Alice and Bob will find the same random outcome. Einstein was very disturbed by this fact, and called it *"spooky action at distance"*. We will come back to this point in the next chapter.

2.5 Reversible gates on a composite system

Until now we discussed the states of a composite system. It is time to discuss the gates.

1. **Product gates.** Suppose that Alice and Bob have two quantum systems A and B and that Alice applies to her system the unitary gate U_A , while Bob applies to his system the unitary gate U_B. In this case, the two systems A and B evolve independently, and it is easy to see that resulting transformation of the composite system AB is given by the product gate $U_A \otimes U_B$. This is clear when Alice's and Bob's systems are in product state $|\alpha\rangle|\beta\rangle$: since $|\alpha\rangle$ is transformed into $U_A|\alpha\rangle$ and $|\beta\rangle$ is transformed into $U_B|\beta\rangle$, the product state $|\alpha\rangle|\beta\rangle$ is transformed into $(U_A \otimes U_B)|\alpha\rangle|\beta\rangle$. In general, product gates transform product states into product states.

2. **Non-product gates.** Not all unitary gates on $\mathcal{H}_A \otimes \mathcal{H}_B$ are product gates. There exist unitary gates U that cannot be written as $U = U_A \otimes U_B$. These gates represent an *interaction* between Alice's and Bob's system and

often they can transform product states into entangled states [1]. If a gate can transform a product state into an entangled state, then the gate is called an *entangling gate*. We will now give some examples.

The CNOT gate. A very important example is the *controlled-NOT (CNOT)* gate. This is the two qubit gate

$$\text{CNOT} = |0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes \sigma_x \,,$$

where σ_x is the Pauli matrix $\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

The reason for the name CNOT is clear from the action of the gate on the states of the computational basis

$$\begin{aligned} \text{CNOT}|0\rangle|0\rangle &= |0\rangle|0\rangle \\ \text{CNOT}|0\rangle|1\rangle &= |0\rangle|1\rangle \\ \text{CNOT}|1\rangle|0\rangle &= |1\rangle|1\rangle \\ \text{CNOT}|1\rangle|1\rangle &= |1\rangle|0\rangle \end{aligned}$$

From this table it is easy to see that the first qubit acts as a control system:

- if the state of the first qubit is $|0\rangle$, then nothing is done on the second qubit
- if the state of the first qubit is $|1\rangle$, then the second qubit undergoes the NOT gate.

Now, the interesting thing happens when the first qubit is in a state of the Fourier basis $\{|+\rangle, |-\rangle\}$. In this case, the CNOT gate generates entanglement:

$$\begin{aligned} \text{CNOT}|+\rangle|0\rangle &= \frac{|0\rangle|0\rangle + |1\rangle|1\rangle}{\sqrt{2}} := |\Phi^+\rangle \\ \text{CNOT}|-\rangle|0\rangle &= \frac{|0\rangle|0\rangle - |1\rangle|1\rangle}{\sqrt{2}} := |\Phi^-\rangle \\ \text{CNOT}|+\rangle|1\rangle &= \frac{|0\rangle|1\rangle + |1\rangle|0\rangle}{\sqrt{2}} := |\Psi^+\rangle \\ \text{CNOT}|-\rangle|1\rangle &= \frac{|0\rangle|1\rangle - |1\rangle|0\rangle}{\sqrt{2}} := |\Psi^-\rangle \end{aligned}$$

The four states $\{|\Phi^+\rangle, |\Phi^-\rangle, |\Psi^+\rangle, |\Psi^-\rangle\}$ are all entangled: after the interaction, the two qubits do not have anymore an individual quantum state.

[1] Note that there exist also some non-product gates that do not generate entanglement: for example, the SWAP gate, defined by $\text{SWAP}|\alpha\rangle|\beta\rangle = |\beta\rangle|\alpha\rangle \,, \forall |\alpha\rangle \in \mathcal{H}_A, |\beta\rangle \in \mathcal{H}_B$, transforms all product states into product states. However, it is not a product gate: in order to exchange the state of two systems, you have to make them interact.

The four entangled states $\{|\Phi^+\rangle, |\Phi^-\rangle, |\Psi^+\rangle, |\Psi^-\rangle\}$ form an orthonormal basis, known as the *Bell basis*.

If we have two qubits in the state $|\Psi\rangle$ and we want to measure them in the Bell basis, we only need to apply first a CNOT gate and then to measure on the basis $\{|+\rangle|0\rangle, |+\rangle|1\rangle, |-\rangle|0\rangle, |-\rangle|1\rangle\}$. For example, we have

$$p_{AB}(+,0) = |\langle +|\langle 0|\mathrm{CNOT}|\Psi\rangle|^2 = |\langle \Phi^+|\Psi\rangle|^2 .$$

The probability of finding the outcomes $(+,0)$ after applying the CNOT gate is the same of the probability of the outcome "Φ^+" for the Bell basis.

A key step in every experimental implementation of quantum information is to realize the CNOT gate. For example, with photons there is an easy way to implement the CNOT, using a suitable crystal known as *polarizing beam splitter(PBS)*. The idea is that photons that have vertical polarization are transmitted by the crystal, while photons with horizontal polarization are reflected in another direction. In this way, the polarization of the photon becomes a control qubit, that controls the path followed by the photon. Regarding the path as a second qubit, this provides a realization of the CNOT gate.

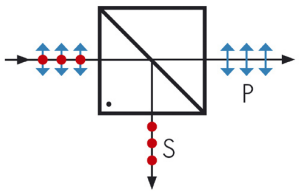

Figure 2.1: Polarizing beam splitter. The input photon is transmitted if the polarization is vertical (state $|0\rangle$, blue arrows in the figure), and reflected if the polarization is horizontal (state $|1\rangle$, red dots in the figure).

"System A" = polarization of the photon, $|0\rangle_A$ = |vertical polarization⟩ and $|1\rangle_A$ = |horizontal polarization⟩.

"System B"= path of the photon, $|0\rangle_B$ = |horizontal path⟩ and $|1\rangle_B$ = |vertical path⟩.

$$|+\rangle_A|0\rangle_B \xrightarrow{PBS} \frac{|\text{vertical}\rangle|\text{horizontal}\rangle + |\text{horizontal}\rangle|\text{vertical}\rangle}{\sqrt{2}}$$

Control-unitary gates. The definition of the CNOT can be generalized to systems of arbitrary dimensions in the following way. Let d_A be the dimension of Alice's system, and let $\{U_m\}_{m=1}^{d_A}$ be a collection of unitary gates acting on Bob's system. Then, we can define the *control-unitary gate*

$$U = \sum_{m=1}^{d_A} |m\rangle\langle m| \otimes U_m .$$

It is indeed easy to see that the operator U defined in this way is unitary.

Reversible quantum computation and quantum oracles. Control unitary gates play a fundamental role in quantum computation. Suppose that you have a function $f : \{1, \dots, d_A\} \to \{1, \dots, d_B\}$ and that you want to compute the function f using elementary quantum particles. To do that, you need to find a unitary gate U_f that does the computation for you. At first sight, this seems difficult: the function f may not be invertible, while quantum gates are invertible. However, the problem can be solved with the following trick: consider *two* quantum systems with Hilbert spaces $\mathcal{H}_A = \mathbb{C}^{d_A}$ and $\mathcal{H}_B = \mathbb{C}^{d_B}$ and define the gate U_f by its action on the computational basis:

$$U_f|m\rangle|n\rangle := |m\rangle|n \oplus f(m)\rangle\,,$$

where $\oplus$ denotes the addition modulo d_B. It is easy to see that the gate U_f is unitary. Moreover, if you apply the gate U_f to the input state $|m\rangle|0\rangle$, you get the output state $|m\rangle|f(m)\rangle$: the outcome of the computation is now written on system B! For this reason, this unitary U_f is called the *quantum oracle for the function* f. The two quantum systems A and B are usually called the *input register* and the *output register*, respectively.

It is easy to see that U_f is a control-unitary gate: it can be written as

$$U_f = \sum_{m=1}^{d_A} |m\rangle\langle m| \otimes S^{f(m)}\,,$$

where S is the shift operator $S = \sum_{n=1}^{d_B} |n \oplus 1\rangle\langle n|$. Depending on the state of the first system, we shift the second system by $f(m)$.

The gate U_f is the quantum analogue of the function f: whenever we have a classical problem with access to a black box computing f, we can translate it into a quantum problem with access to a black box implementing the gate U_f. Of course, the interesting thing about quantum computation is that we can apply the gate U_f to states that are not in the computational basis. For example, we can apply U_f to the state $|f_{d_A}\rangle|0\rangle$, where the system A is in the vector

$$|f_{d_A}\rangle = \frac{1}{\sqrt{d_A}} \sum_{m=1}^{d_A} |m\rangle$$

of the Fourier basis. After the application of the gate U_f system A and B end up in the state

$$U_f|f_{d_A}\rangle|0\rangle = \frac{1}{\sqrt{d_A}} \sum_{m=1}^{d_A} |m\rangle|f(m)\rangle\,.$$

Some authors like to think that, applying the gate U_f in this way, we have evaluated the function f on "all possible inputs at the same time". This way of putting it is a little bit overdone, though. In fact, the information about the values of $f(m)$ is not accessible from the above state: to find $f(m)$ we would have to measure in the computational basis, and the measurement will only give us one pair $(m, f(m))$ where m is chosen at random.

Remark on quantum oracles and reduced quantum oracles. In the previous chapter we have seen a baby version of the Deutsch-Jozsa algorithm, which used the reduced oracle

$$V_f = \sum_{n=1}^{d} (-1)^{f(n)} \, |n\rangle\langle n|$$

for the Boolean function $f : \{1, \ldots, d\} \to \{0, 1\}$. We can now clarify what the reduced oracle V_f has to do with the function f. As we have just seen, the reversible computation of the function f is described by the gate U_f. If we prepare the output register in the state $|-\rangle$, then we obtain

$$\begin{aligned} U_f|n\rangle \otimes |-\rangle &= |n\rangle \otimes \frac{|f(n)\rangle - |f(n) \oplus 1\rangle}{\sqrt{2}} \\ &= (-1)^{f(n)} \, |n\rangle \otimes |-\rangle \\ &= V_f|n\rangle \otimes |-\rangle \, . \end{aligned}$$

By linearity, this also implies that, for every quantum state $|\psi\rangle$, we have

$$U_f|\psi\rangle \otimes |-\rangle = V_f|\psi\rangle \otimes |-\rangle \, .$$

In other words, if we have the oracle U_f, we can always implement the reduced oracle V_f, by preparing the output register in the state $|-\rangle$.

2.6 The Pavia notation

Before presenting other applications of entangled states, it is useful to introduce a notation that can simplify calculations a lot. To do that, recall that a generic vector $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ can be expanded on the computational basis as

$$|\Psi\rangle = \sum_{m=1}^{d_A} \sum_{n=1}^{d_B} \Psi_{mn} \, |m\rangle|n\rangle \, .$$

Now, the coefficients Ψ_{mn} can be used to define an operator Ψ in the following way:

$$\Psi := \sum_{m=1}^{d_A} \sum_{n=1}^{d_B} \Psi_{mn} \, |m\rangle\langle n| \, .$$

This is an operator transforming vectors in $\mathcal{H}_B$ into vectors in $\mathcal{H}_A$. It is easy to see that the correspondence between the state $|\Psi\rangle$ and the operator Ψ is one-to-one.

The inverse of this correspondence is very useful in quantum information and we will adopt a special notation for it. For a given operator Ψ from $\mathcal{H}_B$ to $\mathcal{H}_A$, we will define the state

$$|\Psi\rangle\!\rangle := \sum_{m=1}^{d_A} \sum_{n=1}^{d_B} \langle m|\Psi|n\rangle \, |m\rangle|n\rangle \, . \tag{2.3}$$

For example, we will write the entangled states in the Bell basis as

$$\begin{aligned} |\Phi^+\rangle &= \frac{1}{\sqrt{2}}|I\rangle\!\rangle \\ |\Phi^-\rangle &= \frac{1}{\sqrt{2}}|\sigma_z\rangle\!\rangle \\ |\Psi^+\rangle &= \frac{1}{\sqrt{2}}|\sigma_x\rangle\!\rangle \\ |\Psi^-\rangle &= \frac{i}{\sqrt{2}}|\sigma_y\rangle\!\rangle \end{aligned}$$

We will call this special notation "double-ket notation" or "Pavia notation" (from the name of the group of Prof. D'Ariano at Pavia University, which made extensive use of it). Like the Dirac notation, the double-ket notation is useful because it avoids boring calculations with matrix elements. The main advantages of this notation are summarized in the following exercise:

Exercise 9 (Properties of the double-ket notation) Check the following properties:

1. $\langle\!\langle \Phi|\Psi\rangle\!\rangle = \mathrm{Tr}[\Phi^\dagger \Psi]$.

2. $|\Psi\rangle\!\rangle = (\Psi \otimes I_B)|I_B\rangle\!\rangle = (I_A \otimes \Psi^T)|I_A\rangle\!\rangle$
   I_A and I_B denoting the identity operator on $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively.

3. $(A \otimes B)|\Psi\rangle\!\rangle = |A\Psi B^T\rangle\!\rangle \qquad \forall A : \mathcal{H}_A \to \mathcal{H}_A, \forall B : \mathcal{H}_B \to \mathcal{H}_B, \forall \Psi : \mathcal{H}_B \to \mathcal{H}_A$.

4. If $\Psi = |\alpha\rangle\langle\beta|$, then $|\Psi\rangle\!\rangle = |\alpha\rangle|\overline{\beta}\rangle$.

5. $\langle\alpha|\langle\beta||\Psi\rangle\!\rangle = \langle\alpha|\Psi|\overline{\beta}\rangle$.

Of course, to check the properties you have to go through the definition of Eq. (2.3), which is in terms of matrix elements. However, the point of this exercise is that *after you have done it* you can use the above properties to do calculations without matrix elements.

The double-ket notation tells us something very useful about the Bell basis: indeed, using the second property in the above exercise, we have

$$
\begin{aligned}
|\Phi^+\rangle &= \frac{1}{\sqrt{2}}|I\rangle\rangle \\
|\Phi^-\rangle &= \frac{1}{\sqrt{2}}|\sigma_z\rangle\rangle = (\sigma_z \otimes I_B)|\Phi^+\rangle \\
|\Psi^+\rangle &= \frac{1}{\sqrt{2}}|\sigma_x\rangle\rangle = (\sigma_x \otimes I_B)|\Phi^+\rangle \\
|\Psi^-\rangle &= \frac{i}{\sqrt{2}}|\sigma_y\rangle\rangle = (i\sigma_y \otimes I_B)|\Phi^+\rangle\,.
\end{aligned}
$$

In other words, if Alice and Bob have two photons in the state $|\Phi^+\rangle$, then Alice can generate any state in the Bell basis by just rotating the polarization of her photon! This fact is the core of a surprising communication protocol known as *dense coding*.

2.7 Application: the dense coding protocol

Dense coding is a protocol that allows Alice and Bob to communicate classical data (bits) by using entanglement and transmitting quantum systems (qubits). It is a nice illustration of the basic facts about composite systems that we saw in this chapter.

The protocol works as follows:

- Alice and Bob are far apart, but each of them has a qubit in her/his lab, and the two qubits are in the Bell state $|\Phi^+\rangle$.
- Alice wants to send a two bits to Bob. To do that, she encodes the value of the two bits $\{00, 01, 10, 11\}$ by applying on her qubit one of the four unitary gates $\{I, \sigma_x, \sigma_y, \sigma_z\}$, respectively.
- Alice sends her qubit to Bob
- To decode Alice's message, Bob measures the two qubits on the Bell basis.

The trick of the protocol is that Alice's operation on her qubit generates one of the four Bell states. Since the Bell states are orthogonal, by performing a measurement on the Bell basis, Bob can identify perfectly the state of the two qubits and decode Alice's message. Thanks to the use of entanglement, Alice communicated **2 bits** to Bob, sending only **one qubit**!

We can summarize the dense coding protocol as a transformation of resources: 1 Bell state + 1 qubit of quantum communication $\longrightarrow$ 2 bits of classical communication.

Again, dense coding seems like magic. How can a single qubit carry two bits of information? You could complain that, in fact, in order to establish the entangled state $|\Phi^+\rangle\rangle$, Alice has to send the qubit B to Bob before the beginning

of the protocol: after all, she is sending two qubits. True, but even if you put it in this way, when Alice sends the first qubit B, she does not need to know the classical message that she will send later to Bob. She may not have any idea of the message yet! The first qubit does not carry any information about the 2 bits that she will transmit later.

2.8 Chapter summary

In this chapter we saw how to describe composite systems in quantum theory: if systems A and B have Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively, then the composite system has Hilbert space $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$. The tensor product rule is very natural, but has one revolutionary consequence: two system A and B can be in a *pure entangled state*, that is, a state that is not of the product form $|\alpha\rangle \otimes |\beta\rangle$. When A and B are entangled, we cannot associate a pure state to them separately. We have seen that entanglement gives rise to a curious phenomenon, known as dense coding, and to strange correlations between the outcomes of the measurements of two distant parties. In the next chapter, we will explore more closely the properties of these strange new correlations.

Chapter 3

Non-locality, no-signalling, and the density matrix

3.1 Hidden variables?

In the previous chapter we saw a curious feature of the Bell state $|\Phi^+\rangle = \frac{|I\rangle\rangle}{\sqrt{2}}$: if both Alice and Bob hold two qubits in this state and each of them measures her/his qubit in the same ONB $\{|0,\theta\rangle, |1,\theta\rangle\}$, with

$$|0,\theta\rangle := \cos\frac{\theta}{2}|0\rangle + \sin\frac{\theta}{2}|1\rangle \qquad \theta \in [0, 2\pi)$$
$$|1,\theta\rangle := -\sin\frac{\theta}{2}|0\rangle + \cos\frac{\theta}{2}|1\rangle\,,$$

then their outcomes will be completely random ($p_A(0) = p_A(1) = p_B(0) = p_B(1) = 1/2$) but perfectly correlated ($p_{AB}(0,0) = p_{AB}(1,1) = 1/2$).

At first sight, this seems surprising: how does Bob's qubit know the result of the measurement on Alice's qubit? For Einstein, all this sounded fishy and he referred to it as "spooky action at distance". On second thought, however, the perfect correlations in the Bell state are less surprising. Indeed, we can simulate these correlations by just tossing a coin, writing the result of the toss inside two boxes, giving the boxes to Alice and Bob, and declaring that, for every given θ, the outcome of Alice's and Bob's measurement will be the generated by opening the box and declaring 0 if the coin gave heads and 1 if the coin gave tails. So, what is the big fuss about quantum entanglement? What is difference from measuring a Bell state and having a bunch of correlated random variables?

Maybe Quantum Mechanics is just like classical probability theory, in this case, the values of Alice's and Bob's measurements are **pre-determined** by some classical random variable λ.

Maybe Quantum Theory is an incomplete theory, and has to be replaced by a more fundamental theory where the outcomes of the measurements are not random, but instead they are function of some **hidden variable** λ.

This is what Einstein, and many with him, were hoping. However, in 1964 John Bell showed that this is not possible: the correlations arising from hidden variables must obey to a set of inequalities, now known as **Bell inequalities**. However, in many cases, quantum theory **violates** these inequalities.

The simplest example is a Bell inequality invented by Clauser, Home, Shimony and Holt and known as the CHSH inequality. The best way understand this inequality is through a game.

3.2 The CHSH game

The CHSH game is a game played by two cooperating players, Alice and Bob, and a referee, who challenges Alice and Bob to provide the right answers to a set of questions. The rules of the game are the following

1. Alice and Bob cannot communicate after the game starts.

2. There are two possible questions for Alice: $a \in \{0,1\}$ and two possible questions for Bob: $b \in \{0,1\}$.
   All questions have the same probability $p(a,b) = \frac{1}{4} \quad \forall a, \forall b$.

3. Alice and Bob have to answer one bit, $x \in \{0,1\}$ and $y \in \{0,1\}$, respectively.

4. If $x \oplus y = a \cdot b$ they win one coin, otherwise they lose one coin. In short, they win $(-1)^{x+y+a\cdot b}$ coins.

(a,b) \ (x,y)	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$	Anwsers
$(0,0)$	$+1$	-1	-1	$+1$	
$(0,1)$	$+1$	-1	-1	$+1$	
$(1,0)$	$+1$	-1	-1	$+1$	
$(1,1)$	-1	$+1$	$+1$	-1	
Questions					

Table 3.1: Payoff table for the CHSH game.

In the classical world, the simplest strategies are the *pure strategies*, where $x = f_A(a)$ and $y = f_B(b)$. In this case, the expected payoff is

$$\omega = \frac{1}{4} \sum_{a,b \in \{0,1\}} (-1)^{f_A(a)+f_B(b)+ab}.$$

It is easy to see that the maximum payoff is achieved for

$$f_A(a) = f_B(b) = 0 \quad \forall a, b$$

In this way, Alice and Bob win $\omega_c = \frac{1}{4}(1+1+1-1) = \frac{1}{2}$ coins. It is also easy to see that mixed strategies do not help. Indeed, the most general mixed strategy is:

- Alice and Bob meet before the start of the game and choose λ at random with probability $p(\lambda)$.
- To the question a, Alice answers at random with probability $p_A(x|a,\lambda)$.
- To the question b, Bob answers at random with probability $p_B(y|b,\lambda)$.

The expected payoff for the classical strategy of CHSH game will be:

$$
\begin{aligned}
\omega &= \sum_{\lambda} p(\lambda) \left\{ \frac{1}{4} \sum_{a,b\in\{0,1\}} \sum_{x,y\in\{0,1\}} (-1)^{x+y+ab}\, p_A(x|a,\lambda)\, p_B(y|b,\lambda) \right\} \\
&\leqslant \max_{\lambda} \left\{ \frac{1}{4} \sum_{a,b} \sum_{x,y} (-1)^{x+y+ab}\, p_A(x|a,\lambda)\, p_B(y|b,\lambda) \right\} \\
&= \frac{1}{4} \sum_{a} \sum_{x} p_A(x|a,\lambda_{\max}) \left\{ \sum_{b} \sum_{y} (-1)^{x+y+ab}\, p_B(y|b,\lambda_{\max}) \right\} \\
&\leqslant \frac{1}{4} \sum_{a} \max_{x} \left\{ \sum_{b} \sum_{y} (-1)^{x+y+ab}\, p_B(y|b,\lambda_{\max}) \right\} \\
&= \frac{1}{4} \sum_{a,b} \sum_{y} (-1)^{f_A(a)+y+ab}\, p_B(y|b,\lambda_{\max}) \qquad f_A(a) := \operatorname*{argmax}_{x} \left\{ \sum_{b,y} (-1)^{x+y+ab}\, p_B(y|b,\lambda_{\max}) \right\} \\
&= \frac{1}{4} \sum_{b} \sum_{y} p_B(y|b,\lambda_{\max}) \left\{ \sum_{a} (-1)^{f_A(a)+y+ab} \right\} \\
&\leqslant \frac{1}{4} \sum_{b} \max_{y} \left\{ \sum_{a} (-1)^{f_A(a)+y+ab} \right\} \\
&= \frac{1}{4} \sum_{a,b} (-1)^{f_A(a)+f_B(b)+ab} \qquad f_B(b) := \operatorname*{argmax}_{y} \left\{ \sum_{a} (-1)^{x+y+ab} \right\} \\
&\leqslant \omega_c = \frac{1}{2}
\end{aligned}
$$

In summary, in the classical world Alice and Bob cannot win more than $\frac{1}{2}$ coins (in average): no matter if they use a pure or a mixed strategy, in the classical world they will always have an expected payoff $\omega \leq 1/2$. This is the *CHSH inequality* and it is the simplest example of a Bell inequality.

Surprisingly, in the quantum world Alice and Bob can win more! How?

- Before the game starts, they prepare the Bell state $|\Phi^+\rangle = \frac{|I\rangle\rangle}{\sqrt{2}}$. Then they separate and each one keeps one qubit.

- If the question is a, Alice measures her qubit on the basis $\{|0,\theta_a\rangle, |1,\theta_a\rangle\}$ with $\theta_0 = 0$ and $\theta_1 = \frac{\pi}{2}$. The outcome of the measurement will be her answer x.

- If the question is b, Bob measures his qubit on the basis $\{|0,\tau_b\rangle, |1,\tau_b\rangle\}$ with $\tau_0 = \frac{\pi}{4}$ and $\tau_1 = -\frac{\pi}{4}$. The outcome of the measurement will be his answer y.

How much do they win in this way? Let us compute the payoff. Using the Born rule for the outcome probabilities, we obtain

$$\begin{aligned}\omega_Q &= \frac{1}{4}\sum_{a,b}\sum_{x,y}(-1)^{x+y+ab}\left|\frac{\langle x,\theta_a|\langle y,\tau_b|\,|I\rangle\rangle}{\sqrt{2}}\right|^2 \\ &= \frac{1}{8}\sum_{a,b}\sum_{x,y}(-1)^{x+y+ab}\,|\langle x,\theta_a|y,\tau_b\rangle|^2\,,\end{aligned}$$

the second property coming from the relation $\langle\alpha|\langle\beta||\Psi\rangle\rangle = \langle\alpha|\Psi|\bar{\beta}\rangle$ (cf. Exercise 3 of chapter 2). Continuing the chain of equalities, we have

$$\begin{aligned}\omega_Q &= \frac{1}{8}\sum_{a,b}\sum_{x,y}(-1)^{x+y+ab}\,\mathrm{Tr}\left[|x,\theta_a\rangle\langle x,\theta_a|y,\tau_b\rangle\langle y,\tau_b|\right] \\ &= \frac{1}{8}\sum_{a,b}(-1)^{ab}\,\mathrm{Tr}\left\{\left[\sum_x(-1)^x\,|x,\theta_a\rangle\langle x,\theta_a|\right]\,\left[\sum_y(-1)^y\,|y,\tau_b\rangle\langle y,\tau_b|\right]\right\} \\ &= \frac{1}{8}\sum_{a,b}(-1)^{ab}\,\mathrm{Tr}\left[(\mathbf{m}_a\cdot\boldsymbol{\sigma})\,(\mathbf{n}_b\cdot\boldsymbol{\sigma})\right]\,.\end{aligned}$$

In the last equality we used the relations

$$\sum_{x=0,1}(-1)^x\,|x,\theta_a\rangle\langle x,\theta_a| = \mathbf{m}_a\cdot\boldsymbol{\sigma}\,, \qquad \mathbf{m}_a := \begin{pmatrix}\sin\theta_a\\0\\\cos\theta_a\end{pmatrix} \tag{3.1}$$

$$\sum_{y=0,1}(-1)^y\,|y,\tau_b\rangle\langle y,\tau_b| = \mathbf{n}_b\cdot\boldsymbol{\sigma}\,, \qquad \mathbf{n}_b := \begin{pmatrix}\sin\tau_b\\0\\\cos\tau_b\end{pmatrix}\,, \tag{3.2}$$

and the notation

$$\mathbf{n}\cdot\boldsymbol{\sigma} := n_x\sigma_x + n_y\sigma_y + n_z\sigma_z\,,$$

where σ_x, σ_y and σ_z are the three Pauli matrices, and $\mathbf{n} = (n_x.n_y, n_z)$. The proof of Eqs. (3.1) and (3.2) is easy and is left to you as an exercise.

Recalling that $\mathrm{Tr}[\sigma_k \sigma_l] = 2\delta_{kl}$, for every $k, l \in \{x, y, z\}$, we then obtain

$$\begin{aligned}
\omega_Q &= \frac{1}{4} \sum_{a,b} (-1)^{ab}\, \mathbf{m}_a \cdot \mathbf{n}_b \\
&= \frac{1}{4} \sum_{a,b} (-1)^{ab} \cos\left(\theta_a - \tau_b\right) \\
&= \frac{1}{4} \left[\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}} - \left(-\frac{1}{\sqrt{2}} \right) \right] \\
&= \frac{1}{\sqrt{2}} \approx 0.707\,.
\end{aligned}$$

Hence, quantum theory allows us to violate the CHSH inequality and to win more than classical probability theory!

Remark 1 (the Bell state as a one-time resource). Every time that Alice and Bob play the game with the above strategy, they consume one pair of qubits in the Bell state $|\Phi^+\rangle$. This means that if they want to play the game k times, they need first to prepare k pairs of qubits, in such a way that the Alice's first qubit is entangled with Bob's first qubit, Alice's second qubit is entangled with Bob's second qubit, and so on.

Remark 2 (optimality of the value $\omega_Q = 1/\sqrt{2}$). Excited by their success in the game, Alice and Bob have become greedy and hope to win more. "Maybe", they think, "using two entangled *qudits* with $d > 2$ we can win more than $1/\sqrt{2}$". Unfortunately for them, this is not the case: $\omega_Q = 1/\sqrt{2}$ is the very best that one can do in quantum mechanics. In a quantum world, there is no way to win more than $1/\sqrt{2}$ at the CHSH game, in the same way as in the classical world there is no way to win more than $1/2$.

Remark 3 (entanglement is necessary to violate the CHSH inequality). We have seen that using the Bell state $|\Phi^+\rangle$ one can win more than $\omega_c = 1/2$, i. e. one can violate the CHSH inequality. Is it possible to do the same without entanglement? If is easy to see that the answer is "no": if Alice and Bob have two qubits in a product state $|\alpha\rangle \otimes |\beta\rangle$, then the probability of their answers given the questions will be

$$p(x, y|a, b) = p(x|a)\, p(y|b) \qquad \text{with} \quad p(x|a) = |\langle x, \theta_a | \alpha \rangle|^2, \quad p(y|b) = |\langle y, \tau_b | \beta \rangle|^2\,.$$

Since the probability distribution is of the product form, we can apply to it the same proof that that we used in the classical case and obtain the bound $\omega \leq 1/2$. In other words, all product states satisfy the CHSH inequality.

3.3 Less certainty, more correlations

What are the implications of the fact that quantum theory violates the CHSH inequality?

- $\omega_Q > \omega_C$ implies that the outcomes of Alice's and Bob's measurements, x and y, cannot be computed by Alice and Bob as functions of the questions a and b, respectively, and of some classical random variable λ.

- $\omega_Q > \omega_C$ implies that the outcomes of Alice's and Bob's measurements **cannot exist before the measurement**! Otherwise we would have $\omega_Q = \omega_C$.

- Having intrinsic randomness offers an advantage in the game! The problem of classical probability theory is that in the classical world Alice and Bob are *forced* to assign the value of the answers to all possible questions *before playing the game*. Compared to classical probability theory, quantum theory offers less certainty (the outcomes of the Alice's and Bob's measurements are not determined before the measurements are performed), but thanks to this fact, it also offers stronger correlations.

From a fundamental point of view, the violation of the CHSH inequality tells us that the following picture of the world is **wrong**:

1. Alice's qubit has a property described by a variable λ_A;
   Bob's qubit has a property described by a variable λ_B.

2. The outcome x of Alice's measurement depends only on Alice's choice of measurement a and on the variable λ_A.
   The outcome y of Bob's measurement depends only on Bob's choice of measurement b and on the variable λ_B.

Quantum Theory is not compatible with 1) and 2): if it were compatible, you could call $\lambda := (\lambda_A, \lambda_B)$ and you would get $\omega_Q = \omega_C$. The fact that quantum mechanics can violate the CHSH inequality (as well as other Bell inequalities) is known as **quantum non-locality** and has been confirmed by many experiments with different physical systems.

3.4 No-signalling

As we observed more than one time, when Alice and Bob share two qubits in the Bell state $|\Phi^+\rangle = \frac{|I\rangle\rangle}{\sqrt{2}}$ and each of them measures her/his qubit on the basis $\{|0,\theta\rangle, |1,\theta\rangle\}$, the outcomes of their measurements are random and perfectly correlated. Somehow, it seems that Alice's measurement can "affect Bob's qubit remotely". Can we use this fact to communicate at unbounded speed? After all, in the previous chapter we saw that Alice can use entanglement to communicate 2 bits to Bob using only 1 qubit... Can she can use entanglement to communicate 1 bit using 0 qubits?

The answer is: *NO WAY!*

If Alice does not send any physical system to Bob, then she cannot communicate any bit to him, even if they share an entangled state. Let us prove it in general: suppose that Alice and Bob have two quantum systems in the state

$|\Psi\rangle\!\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ and that they measure their systems on the ONBs $\{|\alpha_m\rangle\}_{m=1}^{d_A}$ and $\{|\beta_n\rangle\}_{n=1}^{d_B}$, respectively. The joint probability distribution of their outcomes is

$$\begin{aligned} p_{AB}(m,n) &= |\langle\alpha_m|\langle\beta_n||\Psi\rangle\!\rangle|^2 \\ &= \left|\langle\alpha_m|\Psi|\overline{\beta_n}\rangle\right|^2 , \end{aligned}$$

having used one of the properties for the double-ket notation (cf. Exercise 3 of the previous chapter). Now, it is easy to see that the choice of Alice's measurement does not influence the probabilities of Bob's outcomes: indeed, we have

$$\begin{aligned} p_B(n) :&= \sum_{m=1}^{d_A} p_{AB}(m,n) \\ &= \sum_{m=1}^{d_A} |\langle\alpha_m|\Psi|\overline{\beta_n}\rangle|^2 \\ &= \sum_{m=1}^{d_A} \langle\overline{\beta_n}|\Psi^\dagger|\alpha_m\rangle\langle\alpha_m|\Psi|\overline{\beta_n}\rangle \\ &= \langle\overline{\beta_n}|\Psi^\dagger \underbrace{\left(\sum_{m=1}^{d_A} |\alpha_m\rangle\langle\alpha_m|\right)}_{=I_A} \Psi|\overline{\beta_n}\rangle \\ &= \langle\overline{\beta_n}|\Psi^\dagger\Psi|\overline{\beta_n}\rangle \\ &= \overline{\langle\overline{\beta_n}|\Psi^\dagger\Psi|\overline{\beta_n}\rangle} \\ &= \langle\beta_n|\Psi^T\overline{\Psi}|\beta_n\rangle \\ &= \langle\beta_n|\rho_B|\beta_n\rangle \qquad\qquad \rho_B := \Psi^T\overline{\Psi}\,. \end{aligned}$$

Every dependence on Alice's measurement has disappeared: the probability of Bob's outcome depends only on the operator ρ_B. Note that the situation does not change if Alice performs a unitary gate U_A before doing the measurement: performing U_A and then measuring on the ONB $\{|\alpha_m\rangle\}_{m=1}^{d_A}$ is the same as measuring on the ONB $\{|\alpha'_m\rangle\}_{m=1}^{d_A}$ with $|\alpha'_m\rangle := U^\dagger|\alpha_m\rangle$. Hence, the probability of Bob's outcomes is independent on which gate Alice performs on her system.

Of course, here the role of Alice and Bob is symmetric. Also the probabilities

of Alice's outcomes are independent of Bob's choice of basis $\{|\beta_n\rangle\}$:

$$\begin{aligned} p_A(m) &:= \sum_{n=1}^{d_B} p_{AB}(m,n) \\ &= \sum_{n=1}^{d_B} |\langle \alpha_m | \Psi | \overline{\beta_n} \rangle|^2 \\ &= \sum_{n=1}^{d_B} \langle \alpha_m | \Psi | \overline{\beta_n} \rangle \langle \overline{\beta_n} | \Psi^\dagger | \alpha_m \rangle \\ &= \langle \alpha_m | \Psi \underbrace{\left(\sum_{n=1}^{d_B} |\overline{\beta_n}\rangle\langle\overline{\beta_n}| \right)}_{=I_B} \Psi^\dagger | \alpha_m \rangle \\ &= \langle \alpha_m | \Psi \Psi^\dagger | \alpha_m \rangle \\ &= \langle \alpha_m | \rho_A | \alpha_m \rangle \qquad\qquad \rho_A := \Psi\Psi^\dagger. \end{aligned}$$

As expected, every dependence on Bob's measurement has been washed out: the probability of Alice's outcome depends only on the operator ρ_A.

Summarizing, we have the following

No-signalling property of Quantum Theory: sharing an entangled state is not enough for Alice and Bob to communicate. If Alice wants to communicate something to Bob, she has to send a physical system to him.

Note that the no-signalling property forbids instantaneous communication: since physical systems cannot travel at infinite speed, when Alice messages will always take a finite time before reaching Bob.

3.5 Marginal states

As we have just seen, the joint probability distribution $p_{AB}(m,n) = |\langle \alpha_m | \langle \beta_n | | \Psi \rangle\rangle|^2$ has marginals given by

$$p_A(m) = \langle \alpha_m | \rho_A | \alpha_m \rangle \qquad\qquad p_B(n) = \langle \beta_n | \rho_B | \beta_n \rangle \,, \tag{3.3}$$

where $\rho_A = \Psi\Psi^\dagger$ and $\rho_B = \Psi^T\overline{\Psi}$. Here the operator ρ_A plays the role of the *state* of system A: it allows us to compute the probability of the outcomes for every measurement performed on A. Same for the operator ρ_B, which plays the role of the state of system B. For this reason, we call the operators ρ_A and ρ_B the **marginal states** of $|\Psi\rangle\rangle$ on A and B, respectively.

Note that here "state" does not mean "unit vector": ρ_A and ρ_B are *operators*, not vectors. Unit vectors represent *pure* states, but when the composite system

AB is in an entangled state $|\Psi\rangle\rangle$ it is impossible to associate a pure states to A and B!

Marginal states have two important mathematical properties. The first is **positivity**. Recall that an operator ρ on $\mathcal{H}$ is called *positive*, denoted $\rho \geq 0$, iff $\langle\varphi|\rho|\varphi\rangle \geq 0$ for every vector $|\varphi\rangle$. It is easy to see that both ρ_A and ρ_B are positive operators: indeed, we have

$$\begin{aligned}\langle\alpha|\rho_A|\alpha\rangle &= \langle\alpha|\Psi\Psi^\dagger|\alpha\rangle \\ &= \left\|\Psi^\dagger|\alpha\rangle\right\|^2 \\ &\geqslant 0 \qquad \forall|\alpha\rangle \in \mathcal{H}_A\end{aligned}$$

and similarly

$$\begin{aligned}\langle\beta|\rho_B|\beta\rangle &= \langle\beta|\Psi^T\overline{\Psi}|\beta\rangle \\ &= \left\|\overline{\Psi}|\beta\rangle\right\|^2 \\ &\geqslant 0 \qquad \forall|\alpha\rangle \in \mathcal{H}_A\end{aligned}$$

In summary, the marginal states are positive operators and the physical meaning of this fact is that the outcome probabilities must be positive for every possible measurement that Alice and Bob can do.

The second important property is **normalization**: it is easy to see that the trace of a marginal state is equal to 1:

$$\begin{aligned}\mathrm{Tr}[\rho_A] &= \mathrm{Tr}[\Psi\Psi^\dagger] \\ &= \mathrm{Tr}[\Psi^\dagger\Psi] \\ &= \langle\langle\Psi|\Psi\rangle\rangle = 1\end{aligned}$$

and similarly

$$\begin{aligned}\mathrm{Tr}[\rho_B] &= \mathrm{Tr}[\Psi^T\overline{\Psi}] \\ &= \overline{\mathrm{Tr}[\Psi^\dagger\Psi]} \\ &= \overline{\langle\langle\Psi|\Psi\rangle\rangle} = 1.\end{aligned}$$

The physical meaning of normalization is that the sum of the probabilities for all possible outcomes of Alice's and Bob's measurements must be 1.

3.6 The density matrix

The analysis of marginal states suggested a **more general notion of "quantum state"** instead of vectors, we can use operators. Moreover, the properties of marginal states suggest the following definition:

Definition 3 (Density matrix) *We say that an operator ρ on the Hilbert space $\mathcal{H}_A$ is a* density matrix *iff*

$$\rho \geqslant 0 \qquad \textit{and} \qquad \mathrm{Tr}[\rho] = 1.$$

It is easy to see that *every density matrix ρ is the marginal state of some pure entangled state* $|\Psi\rangle\!\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$, with $\mathcal{H}_B \simeq \mathcal{H}_A$. Indeed, the positivity condition $\rho \geq 0$ implies that ρ is self-adjoint ($\rho^\dagger = \rho$) and can be diagonalized as

$$\rho = \sum_{m=1}^{d_A} p_m \, |\alpha_m\rangle\langle\alpha_m| \,,$$

where $\{|\alpha_m\rangle\}_{m=1}^{d_A}$ is an ONB and the eigenvalues $p_m \geq 0$ for every m. The normalization condition requires that the eigenvalues are probabilities: $\sum_{m=1}^{d_A} p_m = \mathrm{Tr}[\rho] = 1$. Hence, defining the unit vector

$$|\Psi\rangle\!\rangle = \sum_{m=1}^{d_A} \sqrt{p_m} \, |\alpha_m\rangle|\bar{\alpha}_m\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B \qquad \mathcal{H}_B \simeq \mathcal{H}_A$$

we have that ρ is the marginal state of $|\Psi\rangle\!\rangle$ on system A: indeed, by definition of the double-ket notation we have $\Psi = \sum_m \sqrt{p_m} \, |\alpha_m\rangle\langle\alpha_m|$ and, therefore $\rho_A := \Psi\Psi^\dagger = \rho$.

3.7 All the states of a quantum system

Since every unit vector in $\mathcal{H}_A \otimes \mathcal{H}_B$ is a pure state of the composite system AB, every density matrix on $\mathcal{H}_A$ *must* be a valid state of the individual system A. In other words, every density matrix has the right to be called **quantum state**. In light of this discussion, we have to update the rule for the description of quantum states:

Upgraded Rule 1 (all the states of a quantum system): The states of a quantum system are represented by the density matrices on the system's Hilbert space. When a system with density matrix ρ undergoes a basic measurement with ONB $\{|\varphi_m\rangle\}_{m=1}^{d}$, the probability of the outcome m is given by

$$p(m) = \langle\varphi_m|\rho|\varphi_m\rangle \,. \tag{3.4}$$

From now on, we will denote by $\mathsf{St}(\mathcal{H})$ the set of density matrices (a.k.a. quantum states) on the Hilbert space $\mathcal{H}$.

What is the relation between the pure states, described by unit vectors, and the general states, described by density matrices? What is the density matrix

of a quantum system in a pure state $|\varphi\rangle$? The answer can be easily found by applying the Born rule: when the system is in the pure state $|\varphi\rangle$ and we measure on a generic ONB $\{|\varphi_m\rangle\}_{m=1}^d$, we know that the probability is given by

$$p(m) = |\langle\varphi_m|\varphi\rangle|^2 = \langle\varphi_m|\varphi\rangle\langle\varphi|\varphi_m\rangle\,. \tag{3.5}$$

Comparing Eqs. (3.5) and (3.4)—note how helpful is the Dirac notation for this comparison!—we then obtain

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

that is, the density matrix of a pure state is a *rank-one projector*. Thanks to the density matrix, we can finally get rid of an annoying fact, namely that two unit vectors that differ for a global phase give the same outcome probabilities for every possible measurement: what is the point of having two different mathematical descriptions of the same physical situation and then to identify them by hand? This inelegant feature is removed by the density matrix: if $|\varphi'\rangle = e^{i\theta}|\varphi\rangle$, it is immediate to see that $|\varphi'\rangle\langle\varphi'| = |\varphi\rangle\langle\varphi|$, that is, two unit vectors that differ only for a global phase correspond to the *same quantum state*.

Exercise 10 Show that the following are equivalent:

1. $\rho = |\varphi\rangle\langle\varphi|$ for some unit vector $|\varphi\rangle \in \mathcal{H}$
2. ρ is a density matrix and it is a projector ($\rho^2 = \rho$)
3. ρ is a density matrix and its rank is equal to 1.

3.8 Pure states vs mixed states

In the first chapter, we stated that unit vectors represent *pure* states, i.e. states in which the information about how the system is maximal. What does this mean, exactly? What is the difference between a density matrix associated to a pure state and a generic density matrix?

Suppose that we do not know exactly the state of a quantum system, but we have some partial information about it. For example, suppose that we only know that the system is in the state ρ_0 with probability p, while with probability $1-p$ it is in the state ρ_1 (with ρ_1 distinct from ρ_0). In this situation, if we perform a measurement with ONB $\{|\varphi_m\rangle\}_{m=1}^d$, we expect to obtain the outcome m with probability

$$p(m) = p\,\langle\varphi_m|\rho_0|\varphi_m\rangle + (1-p)\,\langle\varphi_m|\rho_1|\varphi_m\rangle\,.$$

This can be re-written as $p(m) = \langle\varphi_m|\,\rho_p\,|\varphi_m\rangle$, where ρ_p is the density matrix defined by

$$\rho_p := p\,\rho_0 + (1-p)\,\rho_1 \tag{3.6}$$

(you can easily check that $\rho_p \geq 0$ and $\mathrm{Tr}[\rho_p] = 1$). The density matrix ρ_p represents a *mixed state*, i. e. a state in which we have partial information about the system's preparation.

Mathematically, the matrix ρ_p is called a *convex combination* of ρ_0 and ρ_1 with weight p. The fact that every convex combination of two density matrices is a density matrix tells us that the set of all density matrices is a *convex set* (i.e. a set that is closed under convex combinations). This is an important property, because it guarantees that every mixed state that we can imagine in quantum mechanics will be described by a density matrix. Note that we can also consider mixtures of more than two density matrices: for a set of probabilities $\{p_m\}_{m=1}^M$ and a set of density matrices $\{\rho_m\}_{m=1}^M$ we have that the operator $\rho_{\mathbf{p}} = \sum_{m=1}^M p_m \rho_m$ is a density matrix, representing the partial information that we have about the system when we know only that the system is prepared in the state ρ_m with probability p_m.

Now, we know that the set of all density matrices is a convex set. When we have a convex set, it is natural to ask what are the *extreme points*, that is, the points that *cannot* be obtained as nontrivial convex combinations (i.e. convex combination as in Eq. (3.6) where $\rho_0 \neq \rho_1$ and $p \neq 0, 1$). In other words, an extreme point of $\mathsf{St}(\mathcal{H})$ is a quantum state cannot be prepared by choosing at random between two other distinct states: the notion of extreme point is the right mathematical notion to express the fact that we have maximal knowledge about the preparation of the system.

It is easy to see that the extreme points of $\mathsf{St}(\mathcal{H})$ are exactly the rank-one density matrices, corresponding to the pure states. Indeed, every density matrix can be diagonalized as $\rho = \sum_{m=1}^d p_m |\varphi_m\rangle\langle\varphi_m|$. We can interpret this as a mixture of the density matrices $\rho_m = |\varphi_m\rangle\langle\varphi_m|$ with probabilities p_m. Now, the only way to avoid that ρ is a mixed state is to have all probabilities equal to zero, except one—that is, to have $\rho = |\varphi\rangle\langle\varphi|$ for some unit vector $|\varphi\rangle$. Hence, only rank-one density matrices can be extreme points of $\mathsf{St}(\mathcal{H})$. Conversely, it is an easy exercise to check that every rank-one density matrix is an extreme point:

Exercise 11 For every unit vector $|\varphi\rangle \in \mathcal{H}$, show that the condition $|\varphi\rangle\langle\varphi| = p\,\rho_0 + (1-p)\,\rho_1$ with $p \neq 0, 1$ and $\rho_0, \rho_1 \in \mathsf{St}(\mathcal{H})$, implies $\rho_0 = \rho_1 = |\varphi\rangle\langle\varphi|$.

Summarizing, we have proven the following fact: *a density matrix ρ is an extreme point of $\mathsf{St}(\mathcal{H})$ if and only if $\rho = |\varphi\rangle\langle\varphi|$ for some unit vector $|\varphi\rangle$*, that is, if and only if ρ corresponds to a pure state.

3.9 The partial trace

Starting from the pure states of a composite system AB, we discovered that we have to extend our notion of quantum states from unit vectors to density matrices. What happens now if we have a composite system AB in a mixed state? Are we going to discover something even more general than the density matrices?

Luckily, the answer is *no*: we don't need to revise again our notion of quantum state. To see that, suppose that Alice and Bob have two quantum systems in the mixed state $\rho_{AB} \in \mathsf{St}(\mathcal{H}_A \otimes \mathcal{H}_B)$ and that they measure their systems

on the ONBs $\{|\alpha_m\rangle\}_{m=1}^{d_A}$ and $\{|\beta_n\rangle\}_{n=1}^{d_B}$, respectively. Using the Born rule for density matrices, we know that the joint probabilities of their outcomes is given by

$$p_{AB}(m,n) = \langle\alpha_m|\langle\beta_n|\rho_{AB}|\alpha_m\rangle|\beta_n\rangle\,.$$

It is then easy to see that the Alice's marginal probability distribution $p_A(m)$ is independent on Bob's choice of measurement and is given by

$$\begin{aligned}
p_A(m) &= \sum_{n=1}^{d_B}\langle\alpha_m|\langle\beta_n|\rho_{AB}|\alpha_m\rangle|\beta_n\rangle \\
&= \sum_{n=1}^{d_B}\mathrm{Tr}[\rho_{AB}\left(|\alpha_m\rangle\langle\alpha_m|\otimes|\beta_n\rangle\langle\beta_n|\right)] \\
&= \mathrm{Tr}\left[\rho_{AB}\left(\sum_{n=1}^{d_B}|\alpha_m\rangle\langle\alpha_m|\otimes|\beta_n\rangle\langle\beta_n|\right)\right] \\
&= \mathrm{Tr}\left[\rho_{AB}\left(\sum_{n=1}^{d_B}|\alpha_m\rangle\langle\alpha_m|\otimes I_B\right)\right] \\
&= \mathrm{Tr}\left[\rho_{AB}\left(\sum_{n=1}^{d_B}|\alpha_m\rangle\langle\alpha_m|\otimes|n\rangle\langle n|\right)\right] \\
&= \sum_{n=1}^{d_B}\mathrm{Tr}[\rho_{AB}\left(|\alpha_m\rangle\langle\alpha_m|\otimes|n\rangle\langle n|\right)] \\
&= \sum_{n=1}^{d_B}\langle\alpha_m|\langle n|\rho_{AB}|\alpha_m\rangle|n\rangle \\
&= \langle\alpha_m|\rho_A|\alpha_m\rangle\,,
\end{aligned}$$

where ρ_A is the operator with matrix elements

$$\langle m|\rho_A|m'\rangle := \sum_{n=1}^{d_B}\langle m|\langle n|\rho_{AB}|m'\rangle|n\rangle\,. \tag{3.7}$$

Similarly, Bob's marginal probability distribution $p_B(n)$ is independent on Alice's choice of measurement and is given by

$$p_B(n) = \langle\beta_n|\rho_B|\beta_n\rangle\,,$$

where ρ_B is the operator with matrix elements

$$\langle n|\rho_A|n'\rangle := \sum_{m=1}^{d_A}\langle m|\langle n|\rho_{AB}|m\rangle|n'\rangle\,. \tag{3.8}$$

As you can easily check, ρ_A and ρ_B are both density matrices:

Exercise 12 If ρ_{AB} is a density matrix on $\mathcal{H}_A \otimes \mathcal{H}_B$, then the operators ρ_A and ρ_B defined in Eqs. (3.7) and (3.8) are density matrices on $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively.

We will call ρ_A (respectively, ρ_B) the **marginal state** of ρ_{AB} on system A (respectively, B). When the state ρ_{AB} is pure, this definition is consistent with the definition that we gave earlier in this chapter:

Exercise 13 For a rank-one density matrix $\rho_{AB} = |\Psi\rangle\!\rangle\langle\!\langle\Psi|$, show that the operators ρ_A and ρ_B defined in Eqs. (3.7) and (3.8) are given by $\rho_A = \Psi\Psi^\dagger$ and $\rho_B = \Psi^T\overline{\Psi}$, respectively.

For a generic operator ρ_{AB} (not necessarily a density matrix), the map from ρ_{AB} to ρ_A defined by Eq. (3.7) is linear and is called *partial trace on* $\mathcal{H}_B$. We will denote it by Tr_B and we will write $\rho_A = \mathrm{Tr}_B[\rho_{AB}]$. Similarly, the map from ρ_{AB} to ρ_B is called *partial trace on* $\mathcal{H}_A$ and is denoted as Tr_A. Probably, the partial trace is a new mathematical notion for most of you. To become more familiar with it, try the following exercises:

Exercise 14 Prove the following properties:

1. if ρ_A and ρ_B are generic operators on $\mathcal{H}_A$ and $\mathcal{H}_B$, then $\mathrm{Tr}_A[\rho_A \otimes \rho_B] = \mathrm{Tr}[\rho_A]\,\rho_B$ and $\mathrm{Tr}_B[\rho_A \otimes \rho_B] = \mathrm{Tr}[\rho_B]\,\rho_A$
2. for every operator ρ on $\mathcal{H}_A \otimes \mathcal{H}_B$, one has $\mathrm{Tr}[\rho] = \mathrm{Tr}\left\{\mathrm{Tr}_A[\rho]\right\} = \mathrm{Tr}\left\{\mathrm{Tr}_B[\rho]\right\}$
3. for every operator ρ on $\mathcal{H}_A \otimes \mathcal{H}_B$ and for every pair of operators S, T on $\mathcal{H}_A$, one has
$$\mathrm{Tr}_B[(S \otimes I_B)\rho(T \otimes I_B)] = S\ \mathrm{Tr}_B[\rho]\,T\,.$$

Exercise 15 For a positive operator ρ acting on $\mathcal{H}_A \otimes \mathcal{H}_B$, prove that the following are equivalent

1. $\rho = 0$
2. $\mathrm{Tr}[\rho] = 0$
3. $\mathrm{Tr}_A[\rho] = 0$
4. $\mathrm{Tr}_B[\rho] = 0$.

3.10 Chapter summary

Quantum theory is a **non-local** (entanglement can be used to violate the CHSH inequality), but it is also **non-signalling** (entanglement cannot be used to communicate without exchanging physical systems).

The reason for no-signalling is that the probabilities of Alice's and Bob's measurements depend only on their **marginal states**. This fact forced us to introduce a more general notion of quantum state: the **density matrix**.

Pure states, previously described by unit vectors $|\varphi\rangle$, are now described by rank-one projectors $|\varphi\rangle\langle\varphi|$. In addition to pure states, the notion of density matrix allows us to consider *mixed states*, namely states where the information about the system's preparation is not maximal. The random choice between two different quantum states ρ_0 and ρ_1 is still described by a density matrix, given by a convex combination of ρ_0 and ρ_1. The density matrices that cannot be obtained as convex combinations (except in the trivial way) are exactly the rank-one projectors corresponding to pure states.

Chapter 4

Quantum channels and POVMs

In the previous lecture we considered two quantum systems together and asked what are the outcome probabilities for a measurement performed on one subsystem alone. This forced us to extend our notion of quantum state from unit vectors to density matrices. In physical terms, this means considering not only *pure states* but also *mixed states*, i.e. states for which we have only incomplete information about the preparation of the system.

In this lecture we will follow the same path for evolutions and measurements.

For evolutions, we will have to include in the picture not only *reversible evolutions*, described by unitary matrices, but also *irreversible evolutions*, described by *quantum channels*.

For measurements, we will have to include not only *basic measurements*, described by ONBs, but also *general quantum measurements*, described by *POVMs*.

The concepts of quantum channels and POVMs will be introduced and presented in detail in the following sections.

4.1 Reversible evolutions of the density matrix

Let us start from reversible evolutions. For pure states, we know that reversible evolutions are described by unitary matrices: a unitary gate U will transform a unit vector $|\varphi\rangle$ into another unit vector $U|\varphi\rangle$. But what if the initial state of the system is not pure? How can we describe the evolution of the density matrix?

The answer is easy. We know that every density matrix ρ can be written as a convex combination $\rho = \sum_i p_i \, |\varphi_i\rangle\langle\varphi_i|$, where $\{p_i\}$ are probabilities and each $|\varphi_i\rangle$ is a unit vector. Hence we can think of ρ as the mixed state describing a system that is in the pure state $|\varphi_i\rangle$ with probability p_i. Now, the gate U transforms each pure state $|\varphi_i\rangle$ into the pure state $U|\varphi_i\rangle$. This means that after

the evolution the system will be in the state $U|\varphi_i\rangle$ with probability p_i and the density matrix of the system will be

$$\begin{aligned}\rho' &= \sum_i p_i U|\varphi_i\rangle\langle\varphi_i|U^\dagger \\ &= U\rho U^\dagger.\end{aligned}$$

In summary, the evolution of a density matrix under the unitary gate U is given by the map

$$\mathscr{U}(\rho) := U\rho U^\dagger \tag{4.1}$$

where ρ can be an arbitrary $d \times d$ matrix.

Mathematically, the map $\mathscr{U}$ has three important properties. First of all, $\mathscr{U}$ is *linear*, namely

$$\mathscr{U}\left(\sum_k c_k M_k\right) = \sum_k c_k \mathscr{U}(M_k)\,,$$

for every set of complex coefficients $\{c_k\}$ and for every set of $d \times d$ matrices $\{M_k\}$. Linearity guarantees that mixed states evolve in the correct way: if before the evolution the system is in the state ρ_i with probability p_i, then after the evolution the system will be in the state $\mathscr{U}(\rho_i)$ with probability p_i.

Moreover, the map $\mathscr{U}$ is *trace-preserving*: for every $d \times d$ matrix M we have

$$\begin{aligned}\mathrm{Tr}[\mathscr{U}(M)] &= \mathrm{Tr}[UMU^\dagger] \\ &= \mathrm{Tr}[MU^\dagger U] \\ &= \mathrm{Tr}[M]\,.\end{aligned}$$

Finally, the map $\mathscr{U}$ is *positive*: if P is a positive matrix, then also $\mathscr{U}(P)$ is a positive matrix.

Trace-preservation and positivity are very important properties: they guarantee that $\mathscr{U}(\rho)$ is a density matrix whenever ρ is a density matrix.

4.2 The seed of irreversibility: discarding information

Suppose that we have two physical systems, A and B, and that we throw away system B. This is a very common situation: for example, if you have two sheets of paper on our desk, you may want to throw one of them in the waste basket, or you have a candy wrapped in some paper, you may want to eat the candy and throw away the paper. How can we describe operations like this in the language of quantum theory? Again, the answer is simple: if the joint state of the candy and the paper is $\rho \in \mathsf{St}(\mathcal{H}_A \otimes \mathcal{H}_B)$, then the state of the candy after throwing away the paper is given by the partial trace $\rho' = \mathrm{Tr}_B[\rho] \in \mathsf{St}(\mathcal{H}_A)$.

Of course, throwing things away is an *irreversible* process: for example, if we start off with systems A and B in the product state $\rho = \rho_A \otimes \rho_B$ and we discard

system B, then we end up with system A in the state $\rho' = \rho_A$. At that point, the information that was originally contained in the state ρ_B is completely lost.

Discarding information is often an undesired process, due to the fact that there are physical systems that are not under our control. For example, an excited atom (system A) could emit a photon (system B) that flies away from our laboratory. The physical systems that are not under our control are sometimes called the *environment*.

Mathematically, the map Tr_B transforms density matrices on $\mathcal{H}_A \otimes \mathcal{H}_B$ into density matrices on $\mathcal{H}_A$. More generally, it transforms $(d_A d_B) \times (d_A d_B)$ matrices into $d_A \times d_A$ matrices. It is easy to verify that the map Tr_B is linear, trace-preserving, and positive. As we noted in the previous paragraph, these three properties are necessary in order to have the correct evolution of mixed states.

4.3 General quantum evolutions: quantum channels

Up to now, we considered two types of evolution: the reversible evolution associated to a unitary gate and the irreversible evolution described by the partial trace. Combining these two ingredients together, we will now generate *all possible quantum evolutions*.

Suppose that you are an engineer, who wants to design a machine that transforms the state of a quantum system A. In order to transform the state of A we can

1. prepare another system B in a given fixed state σ
2. evolve A and B together using a unitary gate U_{AB}
3. discard B.

The three steps above are described by the maps

$$\begin{aligned} \mathscr{C}_1(\rho) &:= \rho \otimes \sigma \\ \mathscr{C}_2(\rho_{AB}) &:= U_{AB}\, \rho_{AB}\, U_{AB}^\dagger \\ \mathscr{C}_3(\rho_{AB}) &:= \mathrm{Tr}_B[\rho_{AB}] \end{aligned}$$

and the composition of them gives the map

$$\begin{aligned} \mathscr{C}(\rho) &:= \mathscr{C}_3 \mathscr{C}_2 \mathscr{C}_1(\rho) \\ &= \mathrm{Tr}_B \left[U_{AB}\, (\rho \otimes \sigma)\, U_{AB}^\dagger \right] . \end{aligned} \tag{4.2}$$

Note that, since the maps $\mathscr{C}_1, \mathscr{C}_2$ and $\mathscr{C}_3$ are linear, trace-preserving, and positive, also $\mathscr{C}$ has these properties. Again, this is a fundamental fact, because these three properties guarantee that $\mathscr{C}$ can represent a valid evolution of the mixed states of system A.

An evolution that can be realized as in Eq. (4.2) is called *quantum channel* (here the channel transforms system A into itself). In general, quantum channels are irreversible evolutions, because some information is lost when we throw away system B.

Let us see some examples of quantum channels transforming a system into itself:

1. *Erasure channels.* Let us choose B of the same dimension of A ($\mathcal{H}_B \simeq \mathcal{H}_A \sim \mathbb{C}^d$) and let U_{AB} be the SWAP gate, defined by

$$\mathrm{SWAP}|\varphi\rangle \otimes |\psi\rangle = |\psi\rangle \otimes |\varphi\rangle \qquad \forall |\varphi\rangle, |\psi\rangle \in \mathbb{C}^d \,.$$

   With this choice we have

$$\begin{aligned} \mathscr{C}[\rho] &= \mathrm{Tr}_B \left[\mathrm{SWAP}(\rho \otimes \sigma)\mathrm{SWAP}\right] \\ &= \mathrm{Tr}_B[\sigma \otimes \rho] \\ &= \sigma \,. \end{aligned}$$

   The channel $\mathscr{C}$ produces an output state that is completely independent of the input. This is what happens, for example, when you erase the memory of a hard disk and reset it to its initial state.

2. *Random-unitary channels.* Suppose that U_{AB} is a control-unitary gate controlled by B:

$$U = \sum_{n=1}^{d_B} U_n \otimes |n\rangle\langle n| \,.$$

   Expanding σ as $\sigma = \sum_{n,n'} \sigma_{nn'} \, |n\rangle\langle n'|$, we obtain

$$\begin{aligned} \mathscr{C}(\rho) &= \mathrm{Tr}_B \left[U_{AB}(\rho \otimes \sigma)U_{AB}^\dagger\right] \\ &= \sum_{n,n'} \sigma_{nn'} \; \mathrm{Tr}_B \left[U_n \rho U_{n'}^\dagger \otimes |n\rangle\langle n'|\right] \\ &= \sum_{n,n'} \sigma_{nn'} \, U_n \rho U_{n'}^\dagger \; \underbrace{\mathrm{Tr}[|n\rangle\langle n'|]}_{\delta_{nn'}} \\ &= \sum_n p_n \, U_n \rho U_n^\dagger \qquad p_n := \sigma_{nn} \\ &= \sum_n p_n \, \mathscr{U}_n(\rho) \,. \end{aligned}$$

   Hence, an alternative way to realize the channel $\mathscr{C}$ is to

   - choose n at random with probability p_n
   - apply the unitary channel $\mathscr{U}_n$ to system A.

In this way, the input state ρ_A will evolve to $\mathscr{U}_n(\rho)$ with probability p_n and, on average, the density matrix of the final state will be

$$\begin{aligned}\rho' &= \sum_n p_n \mathscr{U}_n(\rho) \\ &\equiv \mathscr{C}(\rho).\end{aligned}$$

In summary, applying a joint control-unitary gate and throwing away the control system is the same as applying a unitary gate chosen at random.

3. *Dephasing channels.* Suppose that U_{AB} is a control-unitary gate, now controlled by A:

$$U_{AB} = \sum_{n=1}^{d_A} |n\rangle\langle n| \otimes U_n \,.$$

Then, we have

$$\begin{aligned}\mathscr{C}(\rho) &= \mathrm{Tr}_B \left[U_{AB}(\rho \otimes \sigma) U_{AB}^\dagger \right] \\ &= \sum_{n,n'} [\rho]_{nn'} \ \mathrm{Tr}_B \left[|n\rangle\langle n'| \otimes U_n \sigma U_{n'}^\dagger \right] \\ &= \sum_{n,n'} [\rho]_{nn'} \ \mathrm{Tr} \left[U_n \sigma U_{n'}^\dagger \right] \ |n\rangle\langle n'| \,.\end{aligned}$$

Note that the channel does not change the outcome probabilities for a measurement in the computational basis: for every n one has $\langle n|\mathscr{C}(\rho)|n\rangle = \langle n|\rho|n\rangle$. In general, it is easy to see that $\mathscr{C}$ can only reduce the off-diagonal coefficients of ρ: for every n, n' one has

$$|\langle n|\mathscr{C}(\rho)|n'\rangle| \leq |\langle n|\rho|n'\rangle| \,.$$

Channels of this type are called *dephasing channels*. The extreme example of dephasing channel arises if we choose $U_n := S^n$, where S is the shift on the computational basis and $\rho_B = |0\rangle\langle 0|$. With this choice we have $\mathrm{Tr}[U_n \sigma U_{n'}^\dagger] = \langle n'|n\rangle = \delta_{nn'}$ and, therefore, $\mathscr{C}(\rho) = \sum_n \rho_{nn} |n\rangle\langle n|$. All the off-diagonal matrix elements $\langle n|\rho|n'\rangle$ are killed by the channel $\mathscr{C}$. In other words, the state ρ_S has been transformed into a classical mixture of states in the computational basis.

4.4 Channels with different input and output systems

Until now, we considered evolutions that transform a system A into itself. However, nothing forbids to consider evolutions that transform an input system A into a different system A'. For example, we can consider an evolution that encodes the polarization state of one photon into the joint state of two photons:

using the CNOT gate, we can transform the single-photon state $|\varphi\rangle = \alpha|0\rangle+\beta|1\rangle$ into the two-photon state $|\varphi'\rangle = \alpha|0\rangle|0\rangle + \beta|1\rangle|1\rangle$. In general, to transform an input system A into an output system A' it is enough to find two systems B and B' such that $\mathcal{H}_A \otimes \mathcal{H}_B \simeq \mathcal{H}_{A'} \otimes \mathcal{H}_{B'}$. Then, one can

1. prepare system B in a given state σ
2. transform the state of the joint system AB into a state of the joint system $A'B'$ using a unitary gate $U_{AB\to A'B'}$
3. discard B'.

The result of these three steps is the map

$$\mathscr{C}(\rho) = \mathrm{Tr}_{B'}\left[U_{AB\to A'B'}\,(\rho\otimes\sigma)\,U^\dagger_{AB\to A'B'}\right]. \tag{4.3}$$

Like in the case $A = A'$, it is easy to see that the map $\mathscr{C}$ is linear, trace-preserving, and positive. Once more, recall that these properties are fundamental requirements for a valid transformation of mixed states of A into mixed states of A'. We will call an evolution described by Eq. (4.3) a *quantum channel* (transforming A into A').

Quantum channels with different input and output states are actually very common: for example, your classical computer implements such channels all the time:

1. *Classical computations.* Consider a function $f : \{1,\dots,M\} \to \{1,\dots,N\}$ and the control-unitary gate that computes f reversibly

$$U_f = \sum_{x=1}^{M} |x\rangle\langle x| \otimes S^{f(x)}\,.$$

The gate U_f acts on the composite system AB, where $\mathcal{H}_A = \mathbb{C}^M$ and $\mathcal{H}_B = \mathbb{C}^N$. Suppose that we set system B to the initial state $\sigma = |0\rangle\langle 0|$ and throw away system A after the application of U_f. In this way, we obtain a channel from A to B, given by Eq. (4.3) with $A' = B$, $B' = A$ and $U_{AB\to A'B'} = U_f$:

$$\begin{aligned}\mathscr{C}_f(\rho) &= \mathrm{Tr}_A\left[U_f(\rho\otimes\sigma)U_f^\dagger\right]\\ &= \sum_{m,m'} \rho_{mm'}\ \mathrm{Tr}_A\left[|m\rangle\langle m'|\otimes|f(m)\rangle\langle f(m)|\right]\\ &= \sum_m \rho_{mm}\,|f(m)\rangle\langle f(m)|\,.\end{aligned}$$

An equivalent way to implement this channel is to

(a) measure the input system A on the computational basis $\{|m\rangle\}$

(b) if the outcome is m, compute $f(m)$

(c) write down the outcome of the computation on system B.

Since the probability of the outcome m is $p_m = \langle m|\rho|m\rangle = \rho_{mm}$, the state that one obtain in this way is, on average,

$$\begin{aligned}\rho' &= \sum_m p_m \, |f(m)\rangle\langle f(m)| \\ &\equiv \mathscr{C}_f(\rho)\,.\end{aligned}$$

2. *Isometric encodings.* In many applications, including communication, cryptography, and error correction, it is useful to encode the information carried by a system A into a larger system A'. Mathematically, one way to do it is to pick an isometry $V : \mathcal{H}_A \to \mathcal{H}'_A$, that is, a rectangular $d_{A'} \times d_A$ matrix satisfying $V^\dagger V = I_A$, and to encode the density matrix $\rho \in \mathsf{St}(\mathcal{H}_A)$ into the new density matrix

$$\mathscr{V}(\rho) := V\rho V^\dagger\,.$$

It is easy to see that this mathematical operation of encoding can be actually implemented by a physical process that one can engineer in the laboratory. In other words, the isometric encodings are quantum channels:

Property 2 (Physical realization of isometric encodings) For every isometry $V : \mathcal{H}_A \to \mathcal{H}_{A'}$, there exist two quantum systems B and B', a pure state $|\beta\rangle \in \mathcal{H}_B$ and a unitary gate $U_{AB\to A'B'}$ such that, for every density matrix $\rho \in \mathsf{St}(\mathcal{H}_A)$, one has

$$\mathscr{V}(\rho) = \mathrm{Tr}_{B'}\left[U_{AB\to A'B'}\left(\rho \otimes |\beta\rangle\langle\beta|\right) U^\dagger_{AB\to A'B'}\right] \tag{4.4}$$

Proof. Choose two quantum systems B and B' such that $d_A d_B = d_{A'} d_{B'}$. Then, for every $m = 1, \ldots, d_A$, define the vector $|\Phi_{m,1}\rangle := V|m\rangle \otimes |1\rangle \in \mathcal{H}_{A'} \otimes \mathcal{H}_{B'}$. Since V is an isometry, the vectors $\{|\Phi_{m,1}\rangle\}$ are orthonormal:

$$\begin{aligned}\langle\Phi_{m,1}|\Phi_{m',1}\rangle &= \langle m|V^\dagger V|m'\rangle \\ &= \langle m|m'\rangle \\ &= \delta_{mm'}.\end{aligned}$$

Hence, we can extend them to an ONB for $\mathcal{H}_{A'} \otimes \mathcal{H}_{B'}$. Denote by $\{|\Phi_{m,n}\rangle\ ,\ m = 1, \ldots, d_{A'}\ ,\ n = 1, \ldots, d_{B'}\}$ the vectors of this ONB. Defining the unitary

$$U_{AB\to A'B'} := \sum_{m,n} |\Phi_{m,n}\rangle\langle m|\langle n|\,,$$

we have that U and V satisfy the relation

$$(V|\varphi\rangle) \otimes |1\rangle = U(|\varphi\rangle \otimes |1\rangle) \qquad \forall |\varphi\rangle \in \mathcal{H}_A\,,$$

or, equivalently,

$$V\rho V^\dagger \otimes |1\rangle\langle 1| = U_{AB\to A'B'}(\rho \otimes |1\rangle\langle 1|)U^\dagger_{AB\to A'B'} \qquad \forall \rho \in \mathsf{St}(\mathcal{H}_A)\,.$$

In conclusion, we obtain $\mathscr{V}(\rho) = \mathrm{Tr}_{B'}\left[U_{AB\to A'B'}\ (\rho \otimes |1\rangle\langle 1|)\ U^\dagger_{AB\to A'B'}\right]$, which is of the desired form of Eq. (4.4), with $|\beta\rangle \equiv |1\rangle$. ■

Note that in the physical realization of Eq. (4.4) is more specific than the one of Eq. (4.3), because the state of system B is pure.

4.5 The Kraus representation

The definition of quantum channel in Eqs. (4.2) and (4.3) has an intuitive physical meaning, but in general is not very convenient, because it requires us to specify all the details of how the channel is implemented.

A more convenient representation is the *Kraus representation*, given by the following

Theorem 1 (Kraus theorem) *The following are equivalent:*

1. *$\mathscr{C}$ is a quantum channel with input space A and output space A'*
2. *$\mathscr{C}$ can be written as $\mathscr{C}(\rho) = \sum_{k=1}^K C_k \rho C_k^\dagger$ where each C_k is a linear operator from $\mathcal{H}_A$ to $\mathcal{H}'_A$ and $\sum_{k=1}^K C_k^\dagger C_k = I_A$.*

Proof. Suppose that $\mathscr{C}$ is a quantum channel. Then, there exists two quantum systems B and B', a state $\sigma \in \mathsf{St}(\mathcal{H}_B)$ and a unitary gate U such that

$$\mathscr{C}(\rho) = \mathrm{Tr}_{B'}\left[U(\rho \otimes \sigma)U^\dagger\right]\,.$$

Writing σ as $\sigma = \sum_i p_i |\beta_i\rangle\langle\beta_i|$, we then have

$$\begin{aligned}\mathscr{C}(\rho) &= \sum_i p_i\ \mathrm{Tr}_{B'}\left[U\ (\rho \otimes |\beta_i\rangle\langle\beta_i|)\right] \\ &= \sum_{i,j} p_i\ C_{ij}\rho C_{ij}^\dagger\end{aligned}$$

where C_{ij} is the operator with matrix elements $\langle m|C_{ij}|m'\rangle := \langle m|\langle j|U|m'\rangle|\beta_i\rangle$. Note that we have

$$\begin{aligned}\langle m|\left(\sum_{i,j} p_i C_{ij}^\dagger C_{ij}\right)|n\rangle &= \sum_{m'}\left(\sum_{i,j}\langle m|C_{ij}^\dagger|m'\rangle\,\langle m'|C_{ij}|n\rangle\right) \\ &= \sum_i p_i\left(\sum_{m',j}\langle m|\langle\beta_i|U^\dagger|m'\rangle|j\rangle\,\langle m'|\langle j|U|n\rangle|\beta_i\rangle\right) \\ &= \sum_i p_i\,\langle m|\langle\beta_i|\,U^\dagger U\,|n\rangle|\beta_i\rangle \\ &= \delta_{m,n}\,.\end{aligned}$$

In other words, we obtained the relation

$$\sum_{i,j} p_i \, C_{ij}^\dagger C_{ij} = I_A \, .$$

Defining $k := (i,j)$ and $C_k = \sqrt{p_i} C_{ij}$ we then obtain the desired result.

Conversely, suppose that $\mathscr{C}$ can be written as $\mathscr{C}(\rho) = \sum_{k=1}^K C_k \rho C_k^\dagger$ with $\sum_{k=1} C_k^\dagger C_k = I_A$. Then, pick a quantum system $\tilde{A}$ of dimension $d_{\tilde{A}} = K$ and define the operator $V : \mathcal{H}_A \to \mathcal{H}_{A'} \otimes \mathcal{H}_{\tilde{A}}$ as

$$V := \sum_{k=1}^K C_k \otimes |k\rangle \tag{4.5}$$

[this means $V|\varphi\rangle = \sum_{k=1}^K C_k|\varphi\rangle \otimes |k\rangle \quad \forall |\varphi\rangle \in \mathcal{H}_A$]. With this definition, we have $\mathscr{C}(\rho) = \mathrm{Tr}_{B'}[V\rho V^\dagger]$: indeed, one has

$$\begin{aligned}
\mathrm{Tr}_{B'}[V\rho V^\dagger] &= \sum_{k,k'} \mathrm{Tr}_{\tilde{A}} \left[C_k \rho C_{k'} \otimes |k\rangle\langle k'| \right] \\
&= \sum_{k,k'} C_k \rho C_{k'} \ \ \mathrm{Tr}[|k\rangle\langle k'|] \\
&= \sum_k C_k \rho C_k^\dagger \\
&= \mathscr{C}(\rho) \, .
\end{aligned}$$

Moreover, V is an *isometry*:

$$\begin{aligned}
V^\dagger V &= \sum_{k,k'} (C_k^\dagger \otimes \langle k|)(C_{k'} \otimes |k'\rangle) \\
&= \sum_{k,k'} C_k^\dagger C_{k'} \ \langle k|k'\rangle \\
&= \sum_k C_k^\dagger C_k \\
&= I_A .
\end{aligned}$$

The isometry V is called a *Stinespring isometry* for channel $\mathscr{C}$. To conclude, it is enough to apply Proposition 2 to the isometric encoding $\mathscr{V}(\rho) = V\rho V^\dagger$: Proposition 2 ensures that there exists two systems B and B', a pure state of $|\beta\rangle \in \mathcal{H}_B$, and a unitary gate $U_{AB\to A'\tilde{A}B'}$ such that

$$\mathscr{V}(\rho) = \mathrm{Tr}_{B'} \left[U_{AB\to A'\tilde{A}B'} \left(\rho \otimes |\beta\rangle\langle\beta| \right) U_{AB\to A'\tilde{A}B'}^\dagger \right] ,$$

which implies

$$\begin{aligned}
\mathscr{C}(\rho) &= \mathrm{Tr}_{\tilde{A}}[\mathscr{V}(\rho)] \\
&= \mathrm{Tr}_{\tilde{A}} \left\{ \mathrm{Tr}_{B'} \left[U_{AB\to A'\tilde{A}B'} \left(\rho \otimes |\beta\rangle\langle\beta| \right) U_{AB\to A'\tilde{A}B'}^\dagger \right] \right\} \\
&= \mathrm{Tr}_{\tilde{A}B'} \left[U_{AB\to A'\tilde{A}B'} \left(\rho \otimes |\beta\rangle\langle\beta| \right) U_{AB\to A'\tilde{A}B'}^\dagger \right] .
\end{aligned} \tag{4.6}$$

Hence, we conclude that the map $\mathscr{C}$ is a quantum channel. ■

Kraus' theorem has many important consequences. First of all, it gives a powerful way to construct quantum channels: once we have a set of operators $\{C_k\}$ satisfying the normalization condition $\sum_k C_k^\dagger C_k = I_A$, the theorem guarantees that the map $\mathscr{C}(\rho) := \sum_k C_k \rho C_k^\dagger$ is automatically a quantum channel—no need to worry about finding the systems B, B', the state ρ_B, and the unitary gate U!

Moreover, the proof of the Kraus theorem reveals a very important point: every quantum channel can be implemented as an *isometric encoding*, using the Stinespring isometry of Eq. (4.5). As a consequence of this, Eq. (4.6) shows that it is always possible to realize a quantum channel by choosing the initial state ρ_B to be **pure**. Even if the channel describes a random process, such as a random-unitary channel $\mathscr{C} = \sum_k p_k \mathscr{U}_k$, can be realized with the auxiliary system B initialized in a pure state. This feature is specific of Quantum Mechanics: in the classical world, if we want to realize a random process through a reversible gate with we need a source of randomness (think of the importance of having a good random number generator for Montecarlo simulations).

4.6 Product channels: the importance of being completely positive

Suppose that Alice and Bob have two quantum systems A and B, respectively, prepared in a joint state $\rho \in \mathsf{St}(AB)$. Alice and Bob decide to transform their systems independently, using two quantum channels $\mathscr{A}$ and $\mathscr{B}$. After the evolution, the state input systems A and B will be transformed into two output systems A' and B', in a quantum state ρ'. But what is this state?

Let us call $\mathscr{A} \otimes \mathscr{B}$ the map from ρ to ρ'. In order to represent a valid transformation of mixed states of AB, the map $\mathscr{A} \otimes \mathscr{B}$ must be linear. Moreover, if ρ is a product state $\rho = \rho_A \otimes \rho_B$, then the output state must be

$$(\mathscr{A} \otimes \mathscr{B})(\rho_A \otimes \rho_B) := \mathscr{A}(\rho_A) \otimes \mathscr{B}(\rho_B)\,. \tag{4.7}$$

Combined with linearity, this condition determines the action of the map $\mathscr{A} \otimes \mathscr{B}$ on *every state*, and, more generally, on every operator acting on $\mathcal{H}_A \otimes \mathcal{H}_B$. This point is discussed in detail in the Appendix, where we show that every operator C acting on $\mathcal{H}_A \otimes \mathcal{H}_B$ can be written as a linear combination

$$C = \sum_k \gamma_k\, \rho_{A,k} \otimes \rho_{B,k}\,, \tag{4.8}$$

where $\{\gamma_k\}$ are complex coefficients and $\{\rho_{A,k}\}$ are density matrices on $\mathcal{H}_A$ and $\{\rho_{B,k}\}$ are density matrices on $\mathcal{H}_B$.

In order to become more familiar with the tensor product of maps, you can try the following exercise

Exercise 16 Let $\{A_n\}$ be operators on A and $\{B_n\}$ be operators on B. Show that

$$(\mathscr{A} \otimes \mathscr{B}) \left(\sum_n A_n \otimes B_n \right) = \sum_n \mathscr{A}(A_n) \otimes \mathscr{B}(B_n) \, .$$

Using the Kraus form, it is easy to show that, if $\mathscr{A}$ and $\mathscr{B}$ are quantum channels, then the map $\mathscr{A} \otimes \mathscr{B}$ is a quantum channel:

Exercise 17 Let $\mathscr{A}$ and $\mathscr{B}$ be two quantum channels and let $\mathscr{A}(\rho_A) = \sum_k A_k \rho_A A_k^\dagger$ and $\mathscr{B}(\rho_B) = \sum_l B_l \rho_B B_l^\dagger$ be their Kraus representations. Show that the product map $\mathscr{A} \otimes \mathscr{B}$ has Kraus representation

$$(\mathscr{A} \otimes \mathscr{B})(\rho) = \sum_{k,l} (A_k \otimes B_l) \rho (A_k \otimes B_l)^\dagger \, .$$

Use Kraus theorem to prove that $\mathscr{A} \otimes \mathscr{B}$ is a quantum channel.

Since $\mathscr{A} \otimes \mathscr{B}$ is a quantum channel, in particular it is a positive map, transforming positive $(d_A d_B) \times (d_A d_B)$ matrices into positive $(d_{A'} d_{B'}) \times (d_{A'} d_{B'})$ matrices. As a special case, we can consider the case where Bob does not do anything on his system: here the channel acting on B is the *identity channel*, given by

$$\mathscr{I}_B(\rho_B) = \rho_B \qquad \forall \rho_B \in \mathsf{St}(\rho_B)$$

and the map $\mathscr{A} \otimes \mathscr{I}_B$ is positive.

In mathematics, a linear map $\mathscr{A}$ with the property that $\mathscr{A} \otimes \mathscr{I}_B$ is positive for every B is called *completely positive (CP)*. This is a non-trivial property: indeed, there exist maps that are positive, but not completely positive. The most famous example is the *transpose* Θ_A defined by

$$\Theta_A(\rho) := \rho^T$$

Clearly, Θ_A is positive. However, it is also easy to see that $\Theta_A \otimes \mathscr{I}_B$ is not positive for every system B of dimension $d_B \geq 2$:

Exercise 18 Let P be the positive matrix $P = |\Phi\rangle\langle\Phi|$ with $|\Phi\rangle = \sum_{m=1}^{\min\{d_A, d_B\}} |m\rangle|m\rangle$. Show that the matrix $(\Theta_A \otimes \mathscr{I}_B)(P)$ has the eigenvalue -1, corresponding to the eigenvector $|1\rangle|2\rangle - |2\rangle|1\rangle$.

Hence, the transpose cannot be a quantum channel! A device that realize the transpose in the lab will lead to absurd consequences, like negative probabilities.

4.7 All the evolutions of a quantum system

Summarizing what we have seen until now, all quantum channels are linear, completely positive, trace-preserving maps. Interestingly, also the converse is true: every linear, completely positive, trace-preserving map is a quantum channel.

This is not only a beautiful mathematical characterization of quantum channels, but also a deep fact about Quantum Theory. Since linearity, complete positivity, and trace-preservation are the minimum requirements that a map should satisfy in order to be a valid evolution of mixed states, this means that the most general evolutions of quantum states that one can ever imagine can be actually realized as quantum channels.

Using this fact, that will be proved in the next lectures, we can now update the rule to describe quantum evolutions:

Upgraded Rule 3: The evolutions transforming an input quantum system A into an output quantum system A' are linear, completely positive, trace-preserving maps transforming operators on $\mathcal{H}_A$ into operators on $\mathcal{H}_{A'}$.

4.8 All the measurements on a quantum system

The idea of restricting the attention from a composite system AB to one of its components (say A) has been very fruitful up to now: it has stimulated us to extend our notion of states and evolutions from unit vectors and unitary gates to density matrices and quantum channels. It is natural to expect that applying the same idea will lead us to extend our notion of measurement from ONBs to something more general. But what is this "something"?

After our discussion on quantum channels, the answer is easy. Suppose that we have a quantum system A, prepared in some state ρ. In order to measure the system A, we can

1. apply a quantum channel $\mathscr{C}$ that transforms system A into another system B

2. perform a basic measurement on B, described by an ONB $\{|\beta_n\rangle\}_{n=1}^{d_B}$.

Note that the dimension of system B can be larger (or smaller) than the dimension of system A: the number of outcomes of the measurement here is arbitrary.

If we follow the two steps above, the probability of the outcome n is $p(n) = \langle\beta_n|\mathscr{C}(\rho)|\beta_n\rangle$. Using the Kraus representation $\mathscr{C}(\rho) = \sum_k C_k \rho C_k^\dagger$, we can reduce this expression to:

$$\begin{aligned} p(n) &= \sum_k \langle\beta_n|C_k\rho C_k^\dagger|\beta_n\rangle \\ &= \sum_k \mathrm{Tr}\left[C_k^\dagger|\beta_n\rangle\langle\beta_n|C_k\rho\right] \\ &= \mathrm{Tr}[P_n\rho] \qquad P_n := \sum_k C_k^\dagger|\beta_n\rangle\langle\beta_n|C_k\,. \end{aligned}$$

By definition, each operator P_n is positive and $\sum_{n=1}^{d_B} P_B = I_A$. Indeed, we have

$$\begin{aligned}\sum_{n=1}^{d_B} P_n &= \sum_{n=1}^{d_B}\sum_k C_k^\dagger|\beta_n\rangle\langle\beta_n|C_k \\ &= \sum_k C_k^\dagger \underbrace{\left(\sum_{n=1}^{d_B}|\beta_n\rangle\langle\beta_n|\right)}_{I_B} C_k \\ &= \sum_k C_k^\dagger C_k \\ &= I_A\,.\end{aligned}$$

In general, a collection of positive operators $\{P_n\}$ on a Hilbert space $\mathcal{H}$, satisfying the condition $\sum_n P_n = I$, is called *positive operator-valued measure (POVM)*. A POVM is the operator generalization of a probability distribution: indeed, when the Hilbert space $\mathcal{H}$ is one-dimensional, the POVM is just a collection of positive numbers that sum up to 1, i.e. it is a probability distribution.

Given a quantum state ρ and a POVM $\{P_n\}$ we can abstractly define a set of probabilities as

$$p(n) := \mathrm{Tr}[P_n\rho]\,. \tag{4.9}$$

Is it possible to generate the probability distribution $\{p(n)\}$ by transforming the state ρ with a channel and performing a basic measurement on the output? The answer is *yes*: indeed, consider the linear map defined by

$$\mathscr{C}(\rho) = \sum_n \mathrm{Tr}[P_n\rho]\ |n\rangle\langle n|\,.$$

Using the Kraus theorem it is easy to see that $\mathscr{C}$ is a quantum channel. On the other hand, one clearly has

$$\langle n|\mathscr{C}(\rho)|n\rangle = \mathrm{Tr}[P_n\rho]$$

for every n.

Based on these observations, we can now update the rule for describing measurements in Quantum Theory:

Upgraded Rule 2: The measurements on a quantum system A are described by POVMs. When a system is prepared in the state ρ and measured with the POVM $\{P_n\}$, the probability of the outcome n is $p(n) = \mathrm{Tr}[P_n\rho]$.

Let us see a few examples of POVMs:

1. *Measuring on a random basis.* Suppose that you want to measure the polarization of a photon. You have a linear polarizer oriented at some angle θ, which measures on the ONB

$$|0,\theta\rangle := \cos\frac{\theta}{2}|0\rangle + \sin\frac{\theta}{2}|1\rangle \qquad \theta \in [0, 2\pi)$$
$$|1,\theta\rangle := -\sin\frac{\theta}{2}|1\rangle + \cos\frac{\theta}{2}|0\rangle\,.$$

Now, in practice one cannot know perfectly the angle θ. Instead, one can have some partial information, e.g. one can know the probability distribution of θ, say $p(\theta)$. Averaging over the possible values of θ, the probability of the outcomes 0 and 1 will be given by the POVM $\{P_0, P_1\}$ with outcomes

$$P_0 = \int \mathrm{d}\theta\ p(\theta)\ |0,\theta\rangle\langle 0,\theta|$$
$$P_1 = \int \mathrm{d}\theta\ p(\theta)\ |1,\theta\rangle\langle 1,\theta|\,.$$

POVM of this form arise from the uncertainty about some parameter of the measuring device (like the orientation of the polarizer). In our example, note that if the angle θ is completely unknown, i.e. $p(\theta) = 1/(2\pi)$ we have $P_0 = P_1 = I/2$. Hence, the outcome probabilities $p(0)$ and $p(1)$ are equal to $1/2$, independently of the state of the system.

2. *Projective measurement.* A projective POVM is a POVM $\{P_n\}$ where each operator P_n is a projector, that is $P_n^2 = 1$. Projective POVMs generalize the old basic measurements on ONBs: a measurement on the ONB $\{|\alpha_m\rangle\}_{m=1}^{d_A}$ is equivalent to a projective POVM with rank-one projectors

$$P_m = |\alpha_m\rangle\langle\alpha_m|\,.$$

In fact, it is easy to see that projective POVMs can be obtained from ONB measurements by coarse-graining:

Exercise 19 Let $\{P_n\}_{n=1}^N$ be a projective POVM on a Hilbert space $\mathcal{H} = \mathbb{C}^d$. Then,

(a) show that $\sum_{n\neq n'} P_n P_n' = 0$

(b) show that $P_n P_n' = 0$ for $n' \neq n$

(c) show that there exists an ONB $\{|\alpha_k\rangle\}_{k=1}^d$ and a partition of the set $\{1,\ldots,d\}$ into disjoint subsets $\{\mathbf{S}_n\}$ such that $P_n = \sum_{k\in\mathbf{S}_n} |\alpha_n\rangle\langle\alpha_n|$.

Also, projective POVMs play a special role in quantum theory due to theorem by Naimark, which reduces *every* quantum measurement to the measurement of a suitable projective POVM on a composite system:

Theorem 2 (Naimark's theorem) *For every POVM $\{P_k\}$ on $\mathcal{H}_A$ there exists a system B, a pure state $|\beta\rangle \in \mathcal{H}_B$ and a projective POVM $\{E_n\}$ on $\mathcal{H}_A \otimes \mathcal{H}_B$ such that*

$$\mathrm{Tr}[\rho P_n] = \mathrm{Tr}[(\rho \otimes |\beta\rangle\langle\beta|)\ E_n]$$

for every outcome n.

Proof. Consider the channel $\mathscr{C}$ from A to A' defined by $\mathscr{C}(\rho) = \sum_n \mathrm{Tr}[\rho P_n]\ |n\rangle\langle n|$. Since $\mathscr{C}$ is a channel, one can realize it as

$$\mathscr{C}(\rho) = \mathrm{Tr}_{B'}\left[U_{AB\to A'B'}(\rho \otimes |\beta\rangle\langle\beta|)U^\dagger_{AB\to A'B'}\right] .$$

Hence, one has

$$\begin{aligned}\mathrm{Tr}[\rho P_n] &= \langle n|\mathscr{C}(\rho)|n\rangle \\ &= \mathrm{Tr}_{A'B'}\left[U_{AB\to A'B'}(\rho \otimes |\beta\rangle\langle\beta|)U^\dagger_{AB\to A'B'}(|n\rangle\langle n| \otimes I_{B'})\right] \\ &= \mathrm{Tr}_{A'B'}\left[(\rho \otimes |\beta\rangle\langle\beta|)E_n\right] ,\end{aligned}$$

where E_n is the projector $E_n := U^\dagger_{AB\to A'B'}(|n\rangle\langle n| \otimes I_{B'})\ U_{AB\to A'B'}$. ■

4.9 Chapter summary

In this chapter we considered the most general evolution of a quantum system in interaction with another quantum system. In general, such evolution will be *irreversible*, due to the fact that some output system is discarded, and the information carried by it is lost. To describe these general evolutions we introduced the notion of *quantum channel*. Similarly, we applied the same philosophy to measurements: we considered all possible ways to measure a quantum system by putting it in interaction with another system, and then performing a measurement. Equivalently, this means applying a quantum channel to the system and measuring the output. To describe these general measurements, we introduced the notion of *POVM*.

Density matrices, quantum channels, and POVMs provide the most general description of states, evolutions, and measurements in quantum mechanics. Using these notions, we can now expand the "quantum Rosetta stone" drawn in the end of Chapter 1.

The extended quantum Rosetta stone for quantum systems in interaction	
Physics	**Mathematics**
Quantum system	Hilbert space
Composite system	Tensor product space
State (pure state)	Density matrix (rank-one matrix)
Measurement (basic measurement)	POVM (rank-one projective POVM)
Probability of outcomes for a given measurement	Born rule with the corresponding POVM
Physical transformation (reversible transformation)	Quantum channel (unitary channel)

Now you can really say that you know quantum theory. Everything was already contained in the basic rules explained in Chapters 1 and 2, but we had to do some work to pull out the full-blown formalism of density matrices, channels, and POVMs. In the next chapter we will see two fundamental properties of the quantum formalism that will lead us straight into the core of quantum information.

Appendix: spanning sets of operators

Here we show how generic operators can be written as linear combinations of density matrices.

Operators on a single Hilbert space

Let A be an operator acting on the Hilbert space $\mathcal{H}$. The operator A can be written as a linear combination of two self-adjoint operators using the *Cartesian decomposition*:

$$A = \mathrm{Re}(A) + i\,\mathrm{Im}(A)\,, \qquad \mathrm{Re}(A) := \frac{A + A^\dagger}{2} \qquad \mathrm{Im}(A) := \frac{A - A^\dagger}{2i}\,. \tag{4.10}$$

Since the operators $\mathrm{Re}(A)$ and $\mathrm{Im}(A)$ are self-adjoint, they can be diagonalized as

$$\mathrm{Re}(A) = \sum_i x_i\,|\varphi_i\rangle\langle\varphi_i| \qquad \mathrm{Im}(A) = \sum_j y_j\,|\psi_j\rangle\langle\psi_i|\,, \tag{4.11}$$

where $\{x_i\}$ ($\{y_j\}$) are real coefficients and $\{|\varphi_i\rangle\}$ ($\{|\psi_j\rangle\}$) is a set of orthonormal vectors. Putting together Eqs. (4.10) and (4.11) we obtain

$$A = \sum_k c_k\,\rho_k\,, \tag{4.12}$$

where $\{c_k\}$ is a set of complex coefficients and $\{\rho_k\}$ is a set of (rank-one) density matrices. In conclusion: every operator acting on $\mathcal{H}$ is a (complex) linear combination of density matrices.

Operators on two Hilbert spaces

Let C be an operator acting on $\mathcal{H}_A \otimes \mathcal{H}_B$. Then, it is easy to write C as a linear combination

$$C = \sum_i c_i\,A_i \otimes B_i\,, \tag{4.13}$$

where $\{c_i\}$ are complex coefficients, $\{A_i\}$ are operators acting on $\mathcal{H}_A$, and $\{B_i\}$ are operators acting on $\mathcal{H}_B$. For example, one can expand C as

$$C = \sum_{m,n,k,l} C_{mn,kl}\,|\alpha_m\rangle\langle\alpha_n| \otimes |\beta_k\rangle\langle\beta_l|\,,$$

where $\{|\alpha_m\rangle\}$ and $\{|\beta_k\rangle\}$ are two ONBs for $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively. Defining $i := (m,n,k,l)$ one obtains the decomposition of Eq. (4.13), with $c_i := C_{mn,kl}$, $A_i := |\alpha_m\rangle\langle\alpha_n|$ and $B_i = |\beta_k\rangle\langle\beta_l|$.

Now, we know from the previous paragraph that every operator A_i (B_i) can be written as a linear combination of (rank-one) density matrices on $\mathcal{H}_A$ ($\mathcal{H}_B$). In conclusion, we obtain that every operator C can be written as a linear combination

$$C = \sum_k \gamma_k \, \rho_{A,k} \otimes \rho_{B,k} \,, \tag{4.14}$$

where $\{\gamma_k\}$ are complex coefficients and $\{\rho_{A,k}\}$ ($\{\rho_{B,k}\}$) are rank-one density matrices on $\mathcal{H}_A$ ($\mathcal{H}_B$).

Chapter 5

Steering and purification

In the last chapter we completed our presentation of the basic rules of quantum mechanics: we know that states are described by density matrices, evolutions by channels, and measurements by POVMs. In this chapter we will apply these rules to analyze three fundamental topics. First, we explore the idea of *indirect measurements*, i. e. measurements that extract information without destroying it. Second, we will discuss the phenomenon of *steering*, which allows on to prepare a statistical mixture of quantum states by making measurements on one part of a composite system. Third, we will explore a fundamental property of quantum theory, known as *purification*. Purification is a powerful tool and will allow us to unveil many deep facts about quantum states, channels and POVMs. As an example of its power, we will demonstrate the No Information Without Disturbance principle, a result that provides the stepping stone to the security of quantum cryptography.

5.1 Indirect measurements

Until now, when we talked about quantum measurements, we only cared about the probability of the outcomes. For example, in the previous chapter we said that the most general quantum measurement is described by a POVM $\{P_n\}$ and that when a system in the state ρ is measured with the POVM $\{P_n\}$, the probability of the outcome n is given by the Born rule $p(n) = \mathrm{Tr}[P_n\rho]$. For measurements like the detection of a photon, predicting the probability of the outcomes is the best that we can do, because the photon is absorbed by the measurement device and the only thing that remains after the measurement is the information about the outcome. However, in other cases the system that we measured is still there after the measurement. Consider for example what happens when we observe the colour of an apple. In this case, our eye detects the photons that are reflected by the apple and uses this information to establish the colour. The photons are absorbed by our eye, but the apple is still there. The measurement here is an *indirect measurement*, where we measure one system in

order to extract information about another.

Let us see a simple example of indirect measurement in quantum mechanics. Suppose that we have a photon A in some state $|\varphi\rangle = \alpha|0\rangle + \beta|1\rangle$ and that we want to measure the polarization of a photon A without destroying it. One way to do it is to let interact our photon A with another photon B, initially prepared in the state $|0\rangle$, and then to perform a usual measurement of polarization on B, thus destroying photon B. As an interaction, we can apply a CNOT gate to the two photons, implementing the transformation

$$\texttt{CNOT}(\alpha|0\rangle + \beta|1\rangle)|0\rangle = \alpha|0\rangle|0\rangle + \beta|1\rangle|1\rangle\,. \tag{5.1}$$

Now, the marginal state of the photon B after the interaction is $\rho_B = |\alpha|^2|0\rangle\langle 0| + |\beta|^2|1\rangle\langle 1|$. This means that if we measure the photon B on the computational basis, we find the outcome 0 with probability

$$\begin{aligned} p_B(0) &= \mathrm{Tr}[|0\rangle\langle 0|\rho_B] \\ &= |\alpha|^2 \end{aligned}$$

and the outcome 1 with probability

$$\begin{aligned} p_B(1) &= \mathrm{Tr}[|1\rangle\langle 1|\rho_B] \\ &= |\beta|^2 \end{aligned}$$

These are exactly the outcome probabilities that we would have obtained if we had measured directly the photon A in the computational basis. However, the difference is that using the indirect measurement, the photon A *is still there*. Hence the question: what is the state of the photon A after the measurement? In order to answer the question, we need a rule that tells us to how to update the state of a system A after we discover the outcome of a measurement on system B. This is done in the next paragraph.

5.2 The quantum Bayes rule

In classical probability theory, the rule to update the probability distribution of a random variable is the Bayes rule. Suppose that we have two random variables A and B, with values m and n, respectively, and with joint probability distribution $p_{AB}(m,n)$. When we discover the value n of the random variable B, the probability distribution of the random variable A is updated to

$$p_A(m|n) = \frac{p_{AB}(m,n)}{p_B(n)}\ ,$$

where $p_B(n) := \sum_m p_{AB}(m,n)$ is the marginal probability distribution for system B. Of course, all this holds provided that the probability $p_B(n)$ is non-zero, otherwise, there is no point talking about the probability of A *given that* B *had value* n!

What we want now is a generalization of the Bayes rule for quantum states. Suppose that two quantum systems A and B are in the joint state σ_{AB}. We measure the POVM $\{P_m\}$ on system A and the POVM $\{Q_n\}$ on system B. Then, the joint probability distribution for the outcomes of the two measurements is

$$p_{AB}(m,n) = \mathrm{Tr}[(P_m \otimes Q_n)\sigma_{AB}]\,.$$

On the other hand, the marginal probability distribution for the outcomes of the B measurement is

$$p_B(n) = \mathrm{Tr}[Q_n\,\sigma_B]$$

where $\sigma_B := \mathrm{Tr}_A[\sigma_{AB}]$ is the marginal state of system B. When we discover the outcome of the B measurement, the Bayes rule requires us to update the probability distribution for the outcomes of the A measurement as follows:

$$\begin{aligned} p_A(m|n) &= \frac{p_{AB}(m,n)}{p_B(n)} \\ &= \frac{\mathrm{Tr}[(P_m \otimes Q_n)\sigma_{AB}]}{\mathrm{Tr}[Q_n\,\sigma_B]} \\ &= \frac{\mathrm{Tr}[(P_m \otimes Q_n)\sigma_{AB}]}{\mathrm{Tr}[(I_A \otimes Q_n)\sigma_{AB}]} \\ &= \mathrm{Tr}[P_m\ \sigma_{A|n}] \end{aligned}$$

where $\sigma_{A|n}$ is the density matrix defined by

$$\sigma_{A|n} := \frac{\mathrm{Tr}_B[(I_A \otimes Q_n)\sigma_{AB}]}{\mathrm{Tr}[(I_A \otimes Q_n)\sigma_{AB}]}$$

Again, for this to make sense the probability of the outcome n must be non-zero. In conclusion, we obtained the relation

$$p_A(m|n) = \mathrm{Tr}[P_m\,\sigma_{A|n}]\,.$$

Since the POVM $\{P_m\}$ is arbitrary, this relation tells us that $\sigma_{A|n}$ is the state of system A, conditional to the outcome n of the measurement on system B.

Summarizing, we have proven the following

Property 3 (Quantum Bayes rule) If two systems A and B are in the joint state σ_{AB} and a measurement with POVM $\{Q_n\}$ is performed on system B, the state of system A conditional to the outcome n is

$$\sigma_{A|n} = \frac{\mathrm{Tr}_B[(I_A \otimes Q_n)\sigma_{AB}]}{\mathrm{Tr}[(I_A \otimes Q_n)\sigma_{AB}]}\,. \tag{5.2}$$

For example, consider the case $\sigma_{AB} = |\Psi\rangle\langle\Psi|$, with $|\Psi\rangle = \alpha|0\rangle|0\rangle + \beta|1\rangle|1\rangle$, which appeared in the example of indirect measurement considered in Eq. (5.1).

If system B is measured in the computational basis, the state of system A conditional on the outcome 0 will be

$$\begin{aligned}\sigma_{A|0} &= \frac{\mathrm{Tr}_B[(I_A \otimes |0\rangle\langle 0|)|\Psi\rangle\langle\Psi|]}{\mathrm{Tr}[(I_A \otimes |0\rangle\langle 0|)|\Psi\rangle\langle\Psi|]} \\ &= \frac{|\alpha|^2|0\rangle\langle 0|}{|\alpha|^2} \\ &= |0\rangle\langle 0|\,.\end{aligned}$$

Similarly, the state of system A conditional on the outcome 1 will be $\sigma_{A|1} = |1\rangle\langle 1|$.

5.3 Quantum Instruments

Suppose that we want to perform an indirect measurement on system A, initially prepared in some state ρ. In general, an indirect measurement can be realized as follows:

1. transform system A into a composite system AB, using a quantum channel $\mathscr{C}$
2. measure system B with a POVM $\{Q_n\}_{n=1}^N$.

After step 1, we will have systems A and B in the state $\sigma_{AB} = \mathscr{C}(\rho)$. At step 2, when the outcome n occurs, the state of system A will be transformed into

$$\begin{aligned}\sigma_{A|n} &= \frac{\mathrm{Tr}_B[(I_A \otimes Q_n)\mathscr{C}(\rho)]}{\mathrm{Tr}[(I_A \otimes Q_n)\mathscr{C}(\rho)]} \\ &:= \frac{\mathscr{Q}_n(\rho)}{\mathrm{Tr}[\mathscr{Q}_n(\rho)]} \qquad\qquad \mathscr{Q}_n(\rho) := \mathrm{Tr}_B[(I_A \otimes Q_n)\mathscr{C}(\rho)]\,.\end{aligned}$$

Note that the map $\mathscr{Q}_n$, defined above, has three important properties:

1. it is *linear*
2. it is *trace non-increasing*: $\mathrm{Tr}[\mathscr{Q}_n(\rho)] \leq \mathrm{Tr}[\rho]$ for every ρ
3. it is *completely positive*[1].

A map with these three properties is called *quantum operation.*

Moreover, the collection of maps $\{\mathscr{Q}_n\}$ associated to our indirect measurement has the property that, for every matrix ρ,

$$\begin{aligned}\sum_{n=1}^N \mathrm{Tr}[\mathscr{Q}_n(\rho)] &= \mathrm{Tr}[(I_A \otimes Q_n)\mathscr{C}(\rho)] \\ &= \mathrm{Tr}[\mathscr{C}(\rho)] \\ &= \mathrm{Tr}[\rho]\,.\end{aligned}$$

[1] In order to check complete positivity, you can use the Kraus representation for the channel $\mathscr{C}$ and show that $\mathscr{Q}_n$ can also be written in a Kraus representation. As we saw in the previous chapter, every map that can be written in the Kraus representation is completely positive.

In other words, the linear map $\mathscr{Q} := \sum_{n=1}^{N} \mathscr{Q}_n$ is *trace-preserving*. Note that $\mathscr{Q}$ is also completely positive, because is a sum of completely positive maps. In conclusion, $\mathscr{Q}$ is a quantum channel.

Let us summarize what we have discovered until now. We have discovered that

- an indirect quantum measurement is described by a set of quantum operations $\{\mathscr{Q}_n\}_{n=1}^{N}$ with the property that $\sum_n \mathscr{Q}$ is a quantum channel.
- when the measurement is performed, the probability of the outcome n is $p_n = \mathrm{Tr}[\mathscr{Q}_n(\rho)]$, where ρ is the state of the system before the measurement
- if the outcome n occurs, the state of the system after the measurement is $\mathscr{Q}_n(\rho)/\,\mathrm{Tr}[\mathscr{Q}_n(\rho)]$.

A set of quantum operations $\{\mathscr{Q}_n\}_{n=1}^{N}$ with the property that $\sum_{n=1}^{N} \mathscr{Q}_n$ is a quantum channel is called a *quantum instrument*.

As usual, you may ask if all possible quantum instruments that can be defined mathematically can also be realized with a physical scheme of indirect measurement. And as usual, the answer is *yes*. The easiest way to see it is to choose a system B with N-dimensional Hilbert space and, for the transformation from A to AB, to use the quantum channel $\mathscr{C}$ defined by

$$\mathscr{C}(\rho) := \sum_{n=1}^{N} \mathscr{Q}_n(\rho) \otimes |n\rangle\langle n| \,.$$

Clearly, measuring system B on the computational basis we retrieve the quantum operations $\{\mathscr{Q}_n\}$: indeed, for every n and for every ρ, we have $\mathrm{Tr}_B[(I_A \otimes |n\rangle\langle n|)\mathscr{C}(\rho)] \equiv \mathscr{Q}_n(\rho)$.

Let us see some examples of quantum instruments:

1. **The von Neumann instrument.** Let A be a d-dimensional quantum system, with an ONB $\{|\alpha_m\rangle\}_{m=1}^{d}$. Consider the indirect measurement described the instrument $\{\mathscr{Q}_m\}_{m=1}^{d}$ where the quantum operations are defined as $\mathscr{Q}_m(\rho) := \langle\alpha_m|\rho|\alpha_m\rangle\, |\alpha_m\rangle\langle\alpha_m|$. Clearly, if the state of the system before the measurement is ρ, then the probability of the outcome m is $p_m = \langle\alpha_m|\rho|\alpha_m\rangle$. Conditional on the outcome m, the state of the system after the measurement is $|\alpha_m\rangle\langle\alpha_m|$. Following von Neumann, some old-fashioned textbooks present this as *the* rule for the evolution of the state after a basic measurement. This particular evolution is often called "collapse of the wavefunction", to stress the fact that a quantum superposition (where all possible outcomes are virtually possible) is destroyed by the measurement, which forces the system to choose one particular state of an ONB and to jump on it. It is instructive to see a concrete way to implement the von Neumann instrument as an indirect measurement. Consider another d-dimensional system B, initially prepared in the state

$|0\rangle$, and imagine that A and B interact with the control-shift gate

$$U = \sum_n |\alpha_m\rangle\langle\alpha_m| \otimes S^m \,.$$

From previous examples (cf. chapter 4), we know that the state of A and B after the gate will be

$$U\rho \otimes |0\rangle\langle 0|U^\dagger = \sum_{m,m'} \langle\alpha_m|\rho|\alpha_{m'}\rangle \; |\alpha_m\rangle\langle\alpha_{m'}| \otimes |m\rangle\langle m'| \,.$$

Now, if we measure system B on the computational basis, we obtain the quantum operations

$$\begin{aligned} \mathscr{Q}_m(\rho) &= \mathrm{Tr}_B[(I_A \otimes |m\rangle\langle m|)U\rho \otimes |0\rangle\langle 0|U^\dagger] \\ &= \langle\alpha_m|\rho|\alpha_m\rangle \; |\alpha_m\rangle\langle\alpha_m| \,. \end{aligned}$$

In other words, measuring on the computational basis after a control shift is a physical procedure that implements the von Neumann instrument.

2. **Other instruments associated to a basic measurement.** It is good to remember that the von Neumann instrument is just *one* of the possible ways to update the state of the system in an indirect measurement on an ONB $\{|\alpha_m\rangle\}$. For example, for a quantum state ρ_m we can define the quantum operation $\mathscr{Q}_m$ as $\mathscr{Q}_m(\rho) := \langle\alpha_m|\rho|\alpha_m\rangle \, \rho_m$. Also in this case the probability of the outcome m is $p_m = \langle\alpha_m|\rho|\alpha_m\rangle$. However, the state of the system after the measurement, conditional to outcome m, is now ρ_m, instead of $|\alpha_m\rangle\langle\alpha_m|$. Intuitively, the important thing about a basic measurement is that we extract classical information about the outcome m. After we know m, we can reset the state of the system A to any desired state that depends on m. For example, one way to realize the instrument as an indirect measurement is to choose a system B of dimension d and consider the channel from A to AB defined by

$$\mathscr{C}(\rho) = \sum_{m=1}^{d} \langle m|\rho|m\rangle \; \rho_m \otimes |m\rangle\langle m| \,.$$

Clearly, if we measure B on the computational basis, we obtain the desired quantum instrument:

$$\begin{aligned} \mathscr{Q}_m(\rho) &= \mathrm{Tr}_B[(I_A \otimes |m\rangle\langle m|)\mathscr{C}(\rho)] \\ &= \langle m|\rho|m\rangle \; \rho_m \,. \end{aligned}$$

3. **Lüders instruments.** This is a generalization of the von Neumann instruments: let $\{P_m\}$ be a collection of projectors such that $\sum_m P_m = I_A$ and define the quantum instrument $\{\mathscr{Q}_m\}$ with quantum operations defined by $\mathscr{Q}_m(\rho) := P_m \rho P_m$. The big difference with the previous examples

is that, when the projectors are of rank larger than 1, the state after the measurement does not depend only on the outcome m, but depends only on the initial state ρ. Let us see this in an example: for a system of dimension 3, consider the projectors $P_1 = |1\rangle\langle 1|$ and $P_{23} = |2\rangle\langle 2| + |3\rangle\langle 3|$, which projects on the subspace $\mathcal{H}_{23} = \mathsf{Span}\{|2\rangle, |3\rangle\}$. Now, for every state of the form $|\varphi\rangle = \alpha|2\rangle + \beta|3\rangle$ we will have that $P_{23}|\varphi\rangle = |\varphi\rangle$, and, therefore $\mathscr{Q}_{23}(|\varphi\rangle\langle\varphi|) = |\varphi\rangle\langle\varphi|$. In other words, if the system is initially in the state $|\varphi\rangle$, the probability of the outcome 23 is equal to 1 and the state after the measurement will be still $|\varphi\rangle$. All the quantum superposition inside the subspace $\mathcal{H}_{23}$ are preserved! In a few chapters, we will see that Lüders measurement play an important role in the theory of quantum error correction.

4. **Nielsen-Chuang's instrument.** In the Nielsen-Chuang book, it is said that a measurement process is described by a set of operators $\{M_n\}_{n=1}^N$, which are required to have the property $\sum_n M_n^\dagger M_n = I_A$. The outcome probabilities and the state after the measurement are given by the following rules:

   (a) If the measured system A is in the state ρ, then the probability of the outcome n is given by $p_n = \mathrm{Tr}[M_n^\dagger M_n \rho]$

   (b) conditionally to the outcome n, the state of system A' is $\rho'_n = \frac{M_n \rho M_n^\dagger}{\mathrm{Tr}[M_n^\dagger M_n \rho]}$.

   How can we fit this in the framework of these chapter notes? The answer is immediate: Nielsen and Chuang are considering a special kind of quantum instrument, with quantum operations $\{\mathscr{Q}_n\}$ defined as $\mathscr{Q}_n(\rho) = M_n \rho M_n^\dagger$. It is interesting to see one way to realize the Nielsen-Chuang instrument as an indirect measurement. To this purpose, take a system B of dimension N and consider the isometry $V : \mathcal{H}_A \to \mathcal{H}_A \otimes \mathcal{H}_B$ given by

$$V := \sum_{n=1}^{N} M_n \otimes |n\rangle \tag{5.3}$$

   If we transform system A into the composite system AB using this isometry, and we measure system B on the computational basis we obtain the quantum operation

$$\begin{aligned} \mathscr{Q}_n(\rho) &= \mathrm{Tr}_B[(I_A \otimes |n\rangle\langle n|) V \rho V^\dagger] \\ &= M_n \rho M_n^\dagger \,. \end{aligned}$$

Exercise 20 Check that the operator V defined in Eq. (5.3) is indeed an isometry.

Remark (instruments with different input and output spaces). For completeness, note that we can consider also quantum instruments where the output system A' is different from the output system A. For example, we let a photon (input system A) interact with an atom (system B) and then measure the photon, thus remaining only with the atom (output system A').

5.4 Quantum steering

At the beginning of this chapter, we saw what happens when we have two systems A and B in a state ρ_{AB}, and we measure B with a POVM $\{Q_n\}$. We saw that this measurement will produce the outcome n with probability $p_n = \mathrm{Tr}[(I_A \otimes Q_n)\rho_{AB}]$ and that, conditional on outcome n, the state of system A will be

$$\rho_{A|n} = \frac{\mathrm{Tr}_B[(I_A \otimes Q_n)\rho_{AB}]}{\mathrm{Tr}[(I_A \otimes Q_n)\rho_{AB}]} .$$

Now, on average the state will be

$$\begin{aligned} \sum_n p_n \ \rho_{A|n} &= \sum_n \mathrm{Tr}_B[(I_A \otimes Q_n)\rho_{AB}] \\ &= \mathrm{Tr}_B[\rho_{AB}] \\ &= \rho_A . \end{aligned}$$

This means that the marginal state ρ_A is a mixture of the states $\rho_{A|n}$ with probabilities p_n. In other words, by measuring system B we realized a source of quantum states for system A, which emits the state ρ_n with probability p_n. The average state of the source is the state ρ_A. Note that, changing the measurement on system B, we can obtain a new source of states $\{\rho'_{A|n}\}$ with probabilities $\{p'_n\}$, still with the property that the average state is ρ_A.

For example, if A and B are two photons in the Bell state $|\Phi^+\rangle$, by measuring B on the basis

$$\begin{aligned} |0,\theta\rangle &= \cos(\theta/2)|0\rangle + \sin(\theta/2)|1\rangle \\ |1,\theta\rangle &= -\sin(\theta/2)|0\rangle + \cos(\theta/2)|1\rangle \end{aligned}$$

we can realize a random source of states for system A, which produces the state $|0,\theta\rangle$ with probability $p_0 = 1/2$ and the state $|1,\theta\rangle$ with probability $p_1 = 1/2$. For every value of θ, the average state is the state $\rho_A = I/2$.

This simple example shows one of the striking differences of quantum theory with classical probability theory: in quantum theory there are infinitely many ways to decompose a mixed state as a mixture of pure states! And if this were not surprising enough, the decomposition the state of Alice's photon can depend on a measurement done by Bob on another photon far away... In his famous 1935 paper, Schrödinger described the situation with these words *"It is rather discomforting that the theory should allow a system to be steered or piloted into one or the other type of state at the experimenter's mercy in spite of his having no access to it."*

Let us go deeper into this topic. We have seen that, when two systems are in a joint state ρ_{AB}, a measurement on system B will induce a decomposition of the state ρ_A as a random mixture of quantum states. A natural question is: can we find a *single* state ρ_{AB} which allows us to prepare *every desired decomposition of* ρ_A via a measurement on B?

To answer this question, we need to meet one of the most important features of quantum theory: purification.

5.5 Purification

From chapter 3 we know that every density matrix ρ_A is the marginal state of some pure entangled state $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$, for some system B, that is

$$\rho_A = \Psi\Psi^\dagger .$$

When this relation is satisfied, we say that $|\Psi\rangle$ is a *purification* of ρ_A and we call system B the *purifying system*. Can we find **all** the purifications of ρ?

The answer to this question is provided by the Schmidt decomposition of the state $|\Psi\rangle$:

Theorem 3 (Schmidt decomposition) *Let $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ be a purification of ρ_A and let*

$$\rho_A = \sum_{m=1}^{r} p_m |\alpha_m\rangle\langle\alpha_m| \qquad \langle\alpha_m|\alpha_n\rangle = \delta_{mn} \qquad \forall m, n \in \{1, \ldots, r\}$$
$$p_m > 0 \qquad \forall m \in \{1, \ldots, r\}$$

be a diagonalization of ρ_A. Then, there exist r orthonormal vectors $\{|\beta_m\rangle\}_{m=1}^r \subseteq \mathcal{H}_B$ such that

$$|\Psi\rangle = \sum_{m=1}^{r} \sqrt{p_m} |\alpha_m\rangle|\beta_m\rangle.$$

The proof is an immediate consequence of a fact in linear algebra known as the *singular value decomposition (SVD)*. You can find both the proof and the full information about the SVD in the Appendix of this chapter.

The Schmidt decomposition has many important consequences:

1. **Minimal dimension of the purifying system.** The purifying system B must have Hilbert space of dimension $d_B \geqslant r$, the rank of ρ_A. Incidentally, note that the converse is also true: if you have a Hilbert space of dimension $d_B \geq r$, you can always take the state

   $$|\Psi\rangle := \sum_{m=1}^{r} \sqrt{p_m}\ |\alpha_m\rangle|m\rangle$$

   as your purification of ρ_A.

2. **The rank of ρ_A as an indicator of entanglement.** From the Schmidt decomposition, it is easy to see that if ρ_A is pure, then $|\Psi\rangle$ is a product state. In general, we can consider the rank of ρ_A as a rough measure of "how mixed" is ρ_A: for a full rank state, every outcome of every basic measurement has some non-zero probability to occur. Based on the intuition that a pure state of AB is "more entangled" the more ρ_A is mixed, we can consider r as a measure of "how entangled" is $|\Psi\rangle$. This particular measure is called the *Schmidt rank*.

3. **Perfect correlations of measurement outcomes.** The Schmidt decomposition tells us that for every pure state of the composite system AB there exists two basic measurements on A and B, respectively, such that the outcomes of the two measurements are perfectly correlated. Indeed, we can always extend the two sets of vectors $\{|\alpha_m\rangle\}_{m=1}^r$ and $\{|\beta_m\rangle\}_{m=1}^r$ to two ONBs for A and B, respectively. The joint probability distribution for these two ONBs will be

$$\begin{aligned} p_{AB}(m,n) &= |\langle\alpha_m|\langle\beta_n||\Psi\rangle\rangle|^2 \\ &= p_m \ \delta_{m,n} \, . \end{aligned}$$

   The outcomes of the two measurements are two perfectly correlated random variables. Of course, correlated random variables exist also in classical probability theory, but what is surprising here is that we can obtain perfect correlations by making measurements on a *pure* state.

5.6 The uniqueness of purification

Thanks to the Schmidt decomposition, we can now find all the purifications of a given state ρ_A and show that, in a certain sense, they are all equivalent.

Since it is easier, we will first explain the mathematical meaning of the equivalence of purifications, and then we will elaborate in more detail on the physical meaning of this equivalence.

Property 4 Let $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ and $|\Psi'\rangle \in \mathcal{H}_A \otimes \mathcal{H}_{B'}$ be two purifications of the same state ρ_A. Then, there exists a partial isometry $S : \mathcal{H}_B \to \mathcal{H}_{B'}$ such that

$$|\Psi'\rangle = (I_A \otimes S)|\Psi\rangle \, .$$

Recall that, by definition, we say that an operator S is a *partial isometry* iff $S^\dagger S$ and $SS^\dagger$ are two projectors. The proof of Proposition 5 is immediate: once you have the Schmidt decompositions

$$\begin{aligned} |\Psi\rangle &= \sum_{m=1}^r \sqrt{p_m}|\alpha_m\rangle|\beta_m\rangle \\ |\Psi'\rangle &= \sum_{m=1}^r \sqrt{p_m}|\alpha_m\rangle|\beta'_m\rangle \, , \end{aligned}$$

it is enough to define the operator S as $S := \sum_{m=1}^{r} |\beta'_m\rangle\langle\beta_m|$. By construction, it is obvious that S is a partial isometry.

Mathematically, Proposition 5 tells us the purification of a density matrix is unique *up to partial isometries* on the purifying system. What does this mean *physically*?

It is easy to see that the physical meaning of the uniqueness of the purification is that we can transform any purification into any other by applying a suitable quantum channel on the purifying system. In the following you can find some more explicit comments on this:

1. **Equal purifying systems.** When $B = B'$, it is easy to extend the partial isometry S to a unitary gate U_B on B, thus obtaining the following little variation of Proposition 5:

   Property 5 Let $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ and $|\Psi'\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ be two purifications of the same state ρ_A, with the same purifying system B. Then, there exists a unitary gate U_B such that

   $$|\Psi'\rangle = (I_A \otimes U_B)|\Psi\rangle\,.$$

   You have already seen an example of this fact when we discussed the Bell basis. The four Bell states $|\Phi^+\rangle, |\Phi^-\rangle, |\Psi^+\rangle$, and $|\Psi^-\rangle$ are purifications of the mixed state $\rho_A = I/d$. As we saw already, we can transform any Bell state into any other Bell state by applying a suitable Pauli matrix on system B.

2. **Different purifying systems.** When the two systems B and B' have different dimensions, we cannot connect the two purifications with a unitary gate, but still we can do it with a suitable quantum channel. Precisely, we can transform the system B into the system B' using the channel $\mathscr{B}$ defined by

   $$\mathscr{B}(\rho) := S\rho S^\dagger + \mathrm{Tr}[P_\perp \rho]\ \rho_0\,,$$

   where S is the partial isometry that we constructed in Proposition 5, $P_\perp$ is the projector defined by $P_\perp := I_B - S^\dagger S$, and ρ_0 is a fixed state of system B'. Note that, since S was defined as $S = \sum_{m=1}^{r} |\beta'_m\rangle\langle\beta_m|$, we have

   $$P_\perp = \sum_{m>r} |\beta_n\rangle\langle\beta_n|$$

   and, therefore

   $$\mathrm{Tr}[(I_A \otimes P_\perp)|\Psi\rangle\langle\Psi|] = 0\,.$$

   Hence, when we apply the channel $\mathscr{C}$ on system B we obtain

   $$\begin{aligned}(\mathscr{I}_A \otimes \mathscr{B})(|\Psi\rangle\langle\Psi|) &= (I_A \otimes S)|\Psi\rangle\langle\Psi|(I_A \otimes S^\dagger) \\ &= |\Psi'\rangle\langle\Psi'|\,.\end{aligned}$$

   In other word, we have been able to transform the purification $|\Psi\rangle$ into the purification $|\Psi'\rangle$ by applying a channel on the purifying system.

Among the properties of quantum theory, the uniqueness of the purification is one of the most important properties for quantum information (perhaps *the* most important). We will have many occasions to encounter it again in the next chapters. For the moment, let us use it to answer the question about steering that we asked here.

5.7 Universal steering

We will now see that every decomposition of a mixed state ρ_A can be generated from an arbitrary purification $|\Psi\rangle_{AB}$ by performing a measurement on the purifying system B. This is a strong version of steering, which we refer to as *universal steering*:

Theorem 4 (Universal steering) *Let $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ be a purification of ρ_A. For every decomposition of ρ_A as a mixture $\rho_A = \sum_{m=1}^{M} p_m \rho_m$, there exists a measurement on B, with POVM $\{Q_m\}_{m=1}^{M}$, such that*

$$\mathrm{Tr}_B[(I_A \otimes Q_m)|\Psi\rangle\langle\Psi|] = p_m \rho_m \qquad \forall m = 1, \ldots, M\,.$$

Proof. Since the states ρ_m can always be decomposed as mixtures of pure states, there is no loss of generality in proving the theorem for decompositions into pure states. Hence, let us assume that each ρ_m is pure, i.e. $\rho_m = |\alpha_m\rangle\langle\alpha_m|$ for some unit vector $|\alpha_m\rangle \in \mathcal{H}_A$. Then, take a system B' of dimension M and consider the purification

$$|\Psi'\rangle = \sum_m \sqrt{p_m}|\alpha_m\rangle|m\rangle\,.$$

Clearly, this is a purification of the state ρ_A with purifying system ρ_B. Hence, we know that there is a channel $\mathscr{B}$, sending states of B to states of B', such that

$$|\Psi'\rangle\langle\Psi'| = (\mathscr{I}_A \otimes \mathscr{B})(|\Psi\rangle\langle\Psi|)\,.$$

Now, by definition of $|\Psi'\rangle$ we have

$$\mathrm{Tr}_{B'}[(I_A \otimes |m\rangle\langle m|)|\Psi'\rangle\langle\Psi'|] = p_m\ |\alpha_m\rangle\langle\alpha_m|\,,$$

that is, measuring B' on the computational basis we can steer the desired states. Now, measuring on B' on the computational basis after having transformed B into B' with the channel $\mathscr{B}$ can be viewed as a measurement on B. As such, it will be described by a POVM $\{Q_m\}$ (the explicit expression of the POVM can be obtained from a Kraus representation of the channel $\mathscr{B}$, in the same way showed in chapter 4). ■

An important example is the one where the state of system A is maximally mixed, namely $\rho_A = I/d$. In this case, we can decompose ρ_A in as

$$\rho_A = \frac{1}{d}\ \rho + \left(1 - \frac{1}{d}\right)\ \sigma \qquad\qquad \sigma := \frac{I - \rho}{d-1}\,,$$

where ρ is an arbitrary state of system A. Now, consider the purification $|\Phi\rangle = \frac{|I\rangle\!\rangle}{\sqrt{d}} \in \mathcal{H}_A \otimes \mathcal{H}_B$, with $\mathcal{H}_A \simeq \mathcal{H}_B \simeq \mathbb{C}^d$. The steering theorem ensures us that there is a two-outcome measurement on system B that generates the state ρ with probability $p_\rho = 1/d$ and the state σ with probability $p_\sigma = 1 - 1/d$. In this case, it is easy to find the measurement: it is described by the POVM $\{\rho^T, I - \rho^T\}$. Indeed, for the outcome with POVM operator ρ, we have

$$\begin{aligned}
\mathrm{Tr}_B\left[(I_A \otimes \rho^T)|\Phi\rangle\langle\Phi|\right] &= \frac{1}{d}\ \mathrm{Tr}_B\left[(I_A \otimes \rho^T)|I\rangle\!\rangle\langle\!\langle I|\right] \\
&= \frac{1}{d}\ \mathrm{Tr}_B\left[|\rho\rangle\!\rangle\langle\!\langle I|\right] \\
&= \frac{1}{d}\ \mathrm{Tr}_B\left[(\rho \otimes I_B)|I\rangle\!\rangle\langle\!\langle I|\right] \\
&= \frac{\rho}{d}\ \mathrm{Tr}_B[|I\rangle\!\rangle\langle\!\langle I|] \\
&= \frac{\rho}{d}\,.
\end{aligned}$$

This means that using a measurement on B we can generate, with probability $1/d$, every desired state of system A. Summarizing, we have proven the equation

$$\mathrm{Tr}_B\left[(I_A \otimes \rho^T)|\Phi\rangle\langle\Phi|\right] = \frac{\rho}{d}\,, \tag{5.4}$$

valid for every density matrix ρ (and, more generally, for every matrix). This fact has rather deep consequences, one of which will be shown in the next paragraph.

5.8 Encoding a quantum operation in a quantum state

For a general map $\mathscr{M}$, consider the matrix $C_{\mathscr{M}}$ defined by

$$\Phi_{\mathscr{M}} := (\mathscr{M} \otimes \mathscr{I}_B)(|\Phi\rangle\langle\Phi|)\,,$$

where $|\Phi\rangle$ is the Bell state $|\Phi\rangle = \frac{|I\rangle\!\rangle}{\sqrt{d}} \in \mathcal{H}_A \otimes \mathcal{H}_B\,, \mathcal{H}_A \simeq \mathcal{H}_B \simeq \mathbb{C}^d$. The matrix $\Phi_{\mathscr{M}}$ is called *Choi matrix*.

The first important property of the Choi matrix is that it is in one-to-one correspondence with the map $\mathscr{M}$. Indeed, if we have the Choi matrix we can always compute the map $\mathscr{M}$ through the formula

$$\mathscr{M}(\rho) = d\ \mathrm{Tr}_B\left[\left(I_A \otimes \rho^T\right)\Phi_{\mathscr{M}}\right] \qquad\qquad \forall \rho\,. \tag{5.5}$$

The proof of the formula is simple: thanks to Eq. (5.4) we have

$$\begin{aligned}
\mathscr{M}(\rho) &= d\ \mathscr{M}\left(\mathrm{Tr}_B[\left(I_A \otimes \rho^T\right)|\Phi\rangle\langle\Phi|]\right) \\
&= d\ \mathrm{Tr}_B[\left(I_A \otimes \rho^T\right)\Phi_{\mathscr{M}}]\,,
\end{aligned}$$

where the second equation can be checked easily using the matrix elements. Eq. (5.5) implies that two linear maps with the same Choi matrix are actually the same linear map. The physical meaning of this fact is that a quantum channel (or a quantum operation) is completely identified by its action on the Bell state $|\Phi\rangle$.

The Choi matrix is an easy way to represent linear maps. Using the Choi matrix, we can prove a fact that we left without proof in the previous chapter: the fact that every completely positive trace-preserving (CPTP) map can be realized as quantum channel. The way to obtain the desired result is to show that every CPTP map $\mathscr{C}$ from A to A' can be written in the Kraus representation

$$\mathscr{C}(\rho) = \sum_m C_m \rho C_m^\dagger \,, \qquad\qquad \sum_m C_m^\dagger C_m = I_A \,.$$

Once we have this, we know by Kraus' theorem that the map $\mathscr{C}$ can be realized through a unitary interaction of system A with some other system B that is discarded after the interaction.

Theorem 5 *Let $\mathscr{C}$ be a completely positive map sending operators on $\mathcal{H}_A$ to operators on $\mathcal{H}_{A'}$. Then $\mathscr{C}$ can be written in the Kraus form*

$$\mathscr{C}(\rho) = \sum_m C_m \rho C_m^\dagger \,.$$

Proof. Since $\mathscr{C}$ is completely positive, $\Phi_{\mathscr{C}}$ must be a positive matrix. Then, we can diagonalize it as

$$\Phi_{\mathscr{C}} = \frac{1}{d} \sum_m |C_m\rangle\!\rangle \langle\!\langle C_m| \,,$$

where $|C_m\rangle\!\rangle$ are unnormalized eigenvectors. With this choice, the eigenvalues of $\Phi_{\mathscr{C}}$ are given by $\mathrm{Tr}[C_m^\dagger C_m]/d$: the reason why we choose such a strange way of writing them is just for later convenience. Using Eq. (5.5) we then have

$$\begin{aligned}
\mathscr{C}(\rho) &= d\,\mathrm{Tr}_B\left[\left(I_A \otimes \rho^T\right)\Phi_{\mathscr{C}}\right] \\
&= \mathrm{Tr}_B\left[\sum_m \left(I_A \otimes \rho^T\right)|C_m\rangle\!\rangle\langle\!\langle C_m|\right] \\
&= \mathrm{Tr}_B\left[\sum_m |C_m\rho\rangle\!\rangle\langle\!\langle C_m|\right] \\
&= \mathrm{Tr}_B\left[\sum_m (C_m\rho\otimes I_B)|I\rangle\!\rangle\langle\!\langle I|(C_m^\dagger\otimes I_B)\right] \\
&= \sum_m C_m\rho\ \left\{\mathrm{Tr}_B\left[|I\rangle\!\rangle\langle\!\langle I|\right]\right\} C_m^\dagger \\
&= \sum_m C_m \rho C_m^\dagger \,.
\end{aligned}$$

This is the desired Kraus representation of the map $\mathscr{C}$. Clearly, if $\mathscr{C}$ is trace-preserving, then we must have $\sum_m C_m^\dagger C_m = I_A$. ■

5.9 No Information Without Disturbance

We conclude this chapter by showing an important property of quantum mechanics, known as the *No Information Without Disturbance*. Suppose that we have a quantum system A in some unknown state ρ and that we want to extract some information about ρ without disturbing it. In order to do that, we can try to use a quantum instrument $\{\mathscr{Q}_n\}$, which gives us some outcome n with probability $p_n = \mathrm{Tr}[\mathscr{Q}_n(\rho)]$. As we learnt earlier in this chapter, the state of system A after the measurement, conditional to outcome n will be $\rho_n = \frac{\mathscr{Q}_n(\rho)}{\mathrm{Tr}[\mathscr{Q}_n(\rho)]}$. The condition that the measurement does not disturb the state is then

$$\sum_n p_n \, \rho_n = \rho \, .$$

Substituting the expressions for p_n and ρ_n, we obtain $\sum_n \mathscr{Q}_n(\rho) = \rho$ for every state ρ. In turn, this is equivalent to the condition

$$\sum_n \mathscr{Q}_n = \mathscr{I}_A \, , \tag{5.6}$$

where $\mathscr{I}_A$ is the identity channel on system A. The No Information Without Disturbance principle is contained in the following:

Property 6 (No Information Without Disturbance) If an instrument $\{\mathscr{Q}_n\}$ is does not disturb quantum states [Eq. (5.6)], then the probabilities of the outcome n is independent on the state of the measured system (that is, the measurement does not extract any information about ρ).

Proof. Since $\sum_n \mathscr{Q}_n = \mathscr{I}_A$, in terms of the Choi matrix we have $\sum_n \Phi_{\mathscr{Q}_n} = \Phi_{\mathscr{I}_A} \equiv |\Phi\rangle\langle\Phi|$. Now, this is a convex decomposition of the *pure* state $|\Phi\rangle\langle\Phi|$. Hence the only possibility is that each term in the decomposition is proportional to $|\Phi\rangle\langle\Phi|$, that is $\Phi_{\mathscr{Q}_n} = p_n |\Phi\rangle\langle\Phi|$ for some probability p_n. Since we have $\Phi_{\mathscr{Q}_n} = \Phi_{p_n \mathscr{I}_A}$, we conclude that $\mathscr{Q}_n = p_n \mathscr{I}_A$. This implies that the probabilities of the measurement are independent of ρ: $\mathrm{Tr}[\mathscr{Q}_n(\rho)] = \mathrm{Tr}[p_n \rho] = p_n$ for every ρ. ■

Note that the No Information Without Disturbance is a specific feature of quantum information: in the classical world, one can always read the value of random bit without changing it. Instead, in the quantum world, one cannot extract any information about a generic state of a qubit without affecting it. This non-classical feature is the working principle of quantum cryptography: when Alice and Bob communicate using quantum systems, they have always a chance to discover if between them there is an eavesdropper trying to extract

some information from the systems that they are exchanging! Strictly speaking, if the states received by Bob are *exactly* the states sent by Alice (for a sufficiently large number of states ρ), then Alice and Bob can be sure that, by the laws of nature, nobody else can have extracted any information from their states.

5.10 Chapter summary

In this chapter we started from a basic question: how to measure a quantum system without destroying it? In order to answer the question, we had to explore what happens when to the state of a composite system AB when a measurement is performed on system B. This led us to discover a Bayes rule for quantum states, which was later used to construct indirect measurements. The Bayes rule for quantum states also led us to the phenomenon of steering, which allows to prepare a random mixture of states of system A via a measurement on system B. We then asked the question: can we find a joint state of system AB that allows us prepare *every* random mixture that decomposes the state of A? In order to answer this question, we had to encounter one key feature of quantum theory: purification. We saw that every two purifications of the same quantum state are equivalent, modulo quantum channels on the purifying system. Using this fact, we learnt three important notions: universal steering, the Choi matrix, and the No Information Without Disturbance principle.

This chapter concludes the first part of the course, which focussed on on the basic rules of quantum mechanics. With it, we concluded the first and most substantial part of the book, which introduced you to the basic language of quantum mechanics. In this part, you became familiar with the mathematical description of quantum states (density matrices), quantum evolutions (quantum channels), and quantum measurements (POVMs and instruments)—congratulations on the great job done so far! Equipped with what you have learnt, in the next chapters we will enter into the core of quantum information theory.

Appendix: the singular value decomposition.

One of the most useful results about matrices is the singular value decomposition (SVD), which can be applied to *every* matrix.

Let $\Psi : \mathcal{H}_B \to \mathcal{H}_A$ be an arbitrary matrix. By definition, it is immediate to see that the matrix $\Psi\Psi^\dagger$ is positive (and hermitian). As such, it can be diagonalized as

$$\Psi\Psi^\dagger = \sum_{m=1}^{r} p_m |\alpha_m\rangle\langle\alpha_m| \qquad \langle\alpha_m|\alpha_n\rangle = \delta_{mn} \qquad \forall m,n \in \{1,\dots,r\} \quad (5.7)$$
$$p_m > 0 \qquad \forall m \in \{1,\dots,r\}\,,$$

where r is the rank of $\Psi\Psi^\dagger$. Note that in general $\Psi\Psi^\dagger$ is not a density matrix, so the eigenvalues p_m do not need to sum up to 1.

Theorem 6 (Singular value decomposition) *Let $\Psi : \mathcal{H}_B \to \mathcal{H}_A$ be a linear operator and $\Psi\Psi^\dagger = \sum_{m=1}^{r} p_m |\alpha_m\rangle\langle\alpha_m|$ be a diagonalization of $\Psi\Psi^\dagger$ as in Eq. (5.7). Then, there exists a set of orthonormal vectors $\{|\alpha_m\rangle\}_{m=1}^r \subseteq \mathcal{H}_B$ such that*

$$\Psi = \sum_{m=1}^{r} \sqrt{p_m} |\alpha_m\rangle\langle\beta_m|\,.$$

Proof. The proof is easy: it is enough to define $|\beta_m\rangle := \frac{\Psi^\dagger|\alpha_m\rangle}{\sqrt{p_m}}$ and to check that the set of vectors $\{|\beta_m\rangle\}_{m=1}^r$ has all the desired properties:

1. The vectors $\{|\beta_m\rangle\}_{m=1}^r$ are orthonormal: indeed, we have

$$\begin{aligned}\langle\beta_m|\beta_n\rangle &= \frac{\langle\alpha_m|\Psi\Psi^\dagger|\alpha_n\rangle}{\sqrt{p_m p_n}} \\ &= \frac{p_m\ \langle\alpha_m|\alpha_n\rangle}{\sqrt{p_m p_n}} \\ &= \delta_{mn}.\end{aligned}$$

2. When the vectors $\{|\alpha_m\rangle\}_{m=1}^r$ are extended to an ONB for $\mathcal{H}_A$, we have $\Psi^\dagger|\alpha_m\rangle = 0$ for every $m > r$: indeed,

$$\begin{aligned}\left\|\Psi^\dagger|\alpha_m\rangle\right\|^2 &= \langle\alpha_m|\Psi\Psi^\dagger|\alpha_m\rangle \\ &= \sum_{n=1}^{r} p_n\ |\langle\alpha_m|\alpha_n\rangle|^2 \\ &= 0 \qquad \forall m > r\,.\end{aligned}$$

Hence, we have

$$\begin{aligned}\Psi^\dagger &= \Psi^\dagger \left(\sum_{m=1}^{d_A} |\alpha_m\rangle\langle\alpha_m| \right) \\ &= \Psi^\dagger \left(\sum_{m=1}^{r} |\alpha_m\rangle\langle\alpha_m| \right) \\ &= \sum_{m=1}^{r} \sqrt{p_m} |\beta_m\rangle\langle\alpha_m| \,.\end{aligned}$$

Taking the adjoint, this proves the SVD: $\Psi = \sum_{m=1}^{r} \sqrt{p_m} |\alpha_m\rangle\langle\beta_m|$. ■

Easy as it may be, the SVD is extremely useful. For example, it allows us to see immediately that the matrices Ψ, $\Psi^\dagger$, $\Psi\Psi^\dagger$ and $\Psi^\dagger\Psi$ have all the same rank. In quantum information, it allows us to prove the Schmidt decomposition:

Proof of the Schmidt decomposition. Let us write the purification of ρ_A with the double-ket notation as $|\Psi\rangle\!\rangle$, where Ψ is an operator from $\mathcal{H}_B$ to $\mathcal{H}_A$. Using the SVD for the operator Ψ and the properties of the double-ket notation, we obtain

$$\begin{aligned}|\Psi\rangle\!\rangle &= \left| \sum_{m=1}^{r} \sqrt{p_m} |\alpha_m\rangle\langle\beta_m| \right\rangle\!\right\rangle \\ &= \sum_{m=1}^{r} \sqrt{p_m}\ |\alpha_m\rangle|\overline{\beta_n}\rangle \,.\end{aligned}$$

This is the Schmidt decomposition, up to re-naming the orthonormal vectors $|\overline{\beta_m}\rangle$ as $|\beta_m\rangle$. ■

Part II

The core of Quantum Information

Chapter 6

No-cloning and Teleportation

Every surprise and every mystery about quantum mechanics is already contained in the rules that you learnt in the first five chapters of this course, and everything that will follow from now on will be just a consequence of these rules. In your first exploration, you already discovered that the language of quantum mechanics describes a world that is radically different from the familiar world of classical physics: a world where the outcomes of the measurement do not exist before the measurement is performed (recall the discussion of the CHSH game in chapter 3!) and where two systems together can be in a pure state, even though each of them is in a mixed state. In the group of chapters that starts here we will delve more deeply into the features of the quantum world, exploring the new ways in which information can be processed using quantum systems. These new ways often start from impossibility results, that the creativity of quantum information theorists has been able to turn into surprising new advantages. An example of this situation is the no-cloning theorem, which is the starting point for new cryptographic protocols.

6.1 Copy machines in the quantum world

In our everyday world we are used to the fact that data can be copied. Newspapers, CDs, DVDs, and even our own DNA can be copied, in principle, without any error. This is because in the classical world there are no fundamental limits to our ability to copy data, but only *practical limits*. For example, copying a DVD was hard some time ago, but now it has become easy and cheap, because the technology advanced. We can even imagine that, one day, copying a strand of DNA will be as easy as copying a DVD.

Things are very different in the quantum world, where there is a **fundamental limit** to our ability to copy data: the no-cloning theorem. The no-cloning theorem tells us that no machine can copy perfectly an unknown quantum state.

This limit was first discovered by Wootters and Zurek (Nature, 1982) and independently Dieks (PLA, 1982), stimulated by a strange proposal by Nick Herbert entitled *"FLASH: A superluminal communicator based upon a new kind of quantum measurement"*. In his proposal, Herbert stated that he found a way to communicate faster than light (actually, to communicate at infinite speed—that is, faster than everything). If this fact were correct, it would be shocking, because it would mean that Einstein's theory of relativity is actually wrong! Of course, Herbert's proposal is **not correct**, because, as you know since the third chapter of our course, quantum mechanics satisfies the no-signalling property. However, this fact was not so well known in 1982 and although Herbert proposal was wrong, it was wrong in an interesting way that stimulated the discovery of the no-cloning theorem by Wootters-Zurek and Dieks.

Let us have a quick look into Herbert's proposal. The idea was very simple: suppose that Alice and Bob have two photons in the Bell state $|\Phi^+\rangle$ and that Alice measures the polarization of her photon in the direction θ. As we know very well, Alice's measurement will reduce the state of Bob's photon to be either $|0,\theta\rangle$ or $|1,\theta\rangle$. Now, suppose that Bob has a device that can produce many copies of the state of his photon, so that, instead of having only one photon in the state $|b,\theta\rangle, b = 0, 1$ he can have N photons in the state $|b,\theta\rangle^{\otimes N}$. Of course, in this way he could discover the value of θ! But if this were correct, Alice and Bob could use this trick to communicate faster than light: Alice could encode a message in the value of θ, and Bob could read out the message immediately after Alice's measurement is performed.

So... where is the mistake? The point here is that there is no way to discover θ, due to the no-signalling property of quantum mechanics. Hence, also the copy machine that Herbert suggested must be forbidden by the laws of quantum mechanics. The general result that forbids Herbert's copy machine, as well as many other machines, is the no-cloning theorem.

6.2 The no-cloning theorem

The theorem can be formulated as follows:

Theorem 7 (No-cloning) *It is impossible to construct a copy machine that takes as input a quantum system A in one of two distinct non-orthogonal states $|\varphi_0\rangle$ and $|\varphi_1\rangle$ and deterministically returns as output two systems, A and B, each of them in the same state as the original one. In other words, there is no quantum channel with input A and output AB such that*

$$\mathscr{C}(|\varphi_i\rangle\langle\varphi_i|_A) = |\varphi_i\rangle\langle\varphi_i|_A \otimes |\varphi_i\rangle\langle\varphi_i|_B \qquad \forall i \in \{0,1\}\,.$$

Proof. To prove the theorem, we proceed by contrapositive: we start from the assumption that there exist a channel that copies the two states $|\varphi_0\rangle$ and $|\varphi_1\rangle$ and we conclude that the two states must be either two orthogonal states (namely, $\langle\varphi_0|\varphi_1\rangle = 0$) or they correspond to the same density matrix (namely, $|\varphi_0\rangle\langle\varphi_0| = |\varphi_1\rangle\langle\varphi_1|$).

There are many ways to prove the no-cloning theorem. The nicest one is probably to think of the channel $\mathscr{C}$ as the result of a unitary gate U that describes the interaction between the input system A, one blank copy B, and the copy machine M.

To start with, since system A is initially in a pure state, the state of the composite system ABM must be of the form $|\varphi_i\rangle_A|\Psi_0\rangle_{BM}$, where $|\Psi_0\rangle_{BM}$ the state of the blank copy and of the machine before the cloning process. Similarly, in order to have systems A and B in the desired state $|\varphi_i\rangle_A|\varphi_i\rangle_B$, the state of ABM after the cloning process must be $|\varphi_i\rangle_A|\varphi_i\rangle_B|\mu_i\rangle_M$, where $|\mu_i\rangle_M$ is some state of the machine. Note that, in principle, the state of the machine can possibly depend on i.

Hence, the unitary gate U that describes the cloning process must satisfy the conditions

$$\begin{aligned} U|\varphi_0\rangle_A|\Psi_0\rangle_{BM} &= |\varphi_0\rangle_A|\varphi_0\rangle_B|\mu_0\rangle_M \\ U|\varphi_1\rangle_A|\Psi_0\rangle_{BM} &= |\varphi_1\rangle_A|\varphi_1\rangle_B|\mu_1\rangle_M \,. \end{aligned}$$

Now, since U is a unitary gate, the scalar product between the two states before the action of U must be equal to the scalar product of the states after the action of U: this means that we have

$$\langle\varphi_0|\varphi_1\rangle = \langle\varphi_0|\varphi_1\rangle\langle\varphi_0|\varphi_1\rangle\langle\mu_0|\mu_1\rangle \,.$$

and, taking the moduli, $|\langle\varphi_0|\varphi_1\rangle| = |\langle\varphi_0|\varphi_1\rangle|^2|\langle\mu_0|\mu_1\rangle|$. Now, since $|\mu_0\rangle$ and $|\mu_1\rangle$ are unit vectors, we have $|\langle\mu_0|\mu_1\rangle| \leq 1$. Hence, the scalar product between $|\varphi_0\rangle$ and $|\varphi_1\rangle$ must satisfy the inequality

$$|\langle\varphi_0|\varphi_1\rangle| \leq |\langle\varphi_0|\varphi_1\rangle|^2 \,.$$

The only values satisfying the inequality $x \leqslant x^2$ with $x \in [0,1]$ are $x = 0$ and $x = 1$. Here, $x = 0$ means that $|\varphi_0\rangle$ and $|\varphi_1\rangle$ are orthogonal, while $x = 1$ implies that $|\varphi_0\rangle$ and $|\varphi_1\rangle$ are equal up to a global phase ($|\varphi_1\rangle = e^{i\theta}|\varphi_0\rangle$) and hence they correspond to the same quantum state.

■

In summary, the no-cloning theorem says that there is no quantum copy machine that can copy two non-orthogonal states perfectly and with probability 1. Any physical process that attempts to copying the information encoded in two non-orthogonal states will necessarily incur in an error, or, if succeeds, cannot succeed with probability equal to 1.

Of course, the fact that no machine can copy two distinct non-orthogonal states also tells us that there is no *perfect universal cloning machine*, that is, no machine that can copy perfectly all possible pure states of a quantum system. This is a big difference between quantum theory and classical theory: in classical theory, a bit has only two pure states $\{0,1\}$ and these states can be copied perfectly, while in quantum theory, a qubit has infinite pure states, and there is no machine that can copy all of them.

6.3 What is left open by the no-cloning theorem

The no-cloning theorem leaves a few doors open about copying information encoded into quantum systems:

1. **Cloning orthogonal states.** Of course, nothing forbids that we copy orthogonal states. Indeed, this is what our copy machines to everyday in the classical word: they copy information encoded into the computational basis! The way a classical copy machine works is to first read the input data and then to produce a copy depending on the information that has been read. In the language of quantum mechanics, we can model a classical copy machine as a quantum channel, where the system is measured first on the computational basis $\{|m\rangle\}_{m=1}^{d}$ and, depending on the outcome, one re-prepares two copies of the state $|m\rangle$:

$$\mathscr{C}(\rho_A) = \sum_{m=1}^{d} \langle m|\rho|m\rangle \; |m\rangle\langle m|_A \otimes |m\rangle\langle m|_B \, .$$

   As you can easily check, this channel copies perfectly (and with probability 1) the states in the computational basis. Another, more sophisticate way to copy states in the computational basis is with the control-shift gate

$$U = \sum_{m}^{d} |m\rangle\langle m| \otimes S^m \, ,$$

   which satisfies $U|m\rangle_A|0\rangle_B = |m\rangle_A|m\rangle_B$ for every m. Note that, in the model of reversible quantum computation, the unitary U is exactly the gate that computes the identity function: since in reversible computation we always keep a quantum register with the input of the computation, computing the identity requires copying the data encoded in the computational basis from the input register to the output register.

2. **Approximate cloning.** The no-cloning theorem prevents us from making perfect copies, but how about trying to make some imperfect copies? And how good can be these approximate copies? These questions stimulated a lot of research over the past 15 years. It turns out that, indeed, there exists quantum copy machines that produce approximate copies of non orthogonal states. In the assignments you will find an example of approximate cloning machine known as the *universal cloning machine*, because it clones equally well all pure states. For qubits, the universal cloning machine is described by the quantum channel

$$\mathscr{C}(\rho) := \frac{1}{6} \; (I_A \otimes I_B + \texttt{SWAP}) \, (\rho \otimes I_B) \, (I_A \otimes I_B + \texttt{SWAP}) \; ,$$

   where SWAP is the SWAP gate, defined by $\texttt{SWAP}|\alpha\rangle_A|\beta\rangle_B = |\beta\rangle_A|\alpha\rangle_B \, , \forall |\alpha\rangle, |\beta\rangle \in \mathbb{C}^2$. You will see in the assignments that the marginal states of qubits A

and B are equal, and, precisely, they are equal to

$$\rho' = p\ \rho + (1-p)\ \frac{I}{2} \qquad\qquad p = \frac{2}{3}\,.$$

The cloning machine introduces an error in the state of the two copies. In this case, the error is equivalent to mixing the original state with the maximally mixed state $I/2$.

3. **Probabilistic cloning.** The no-cloning theorem prevents us from constructing machines that copy non-orthogonal quantum states *deterministically.* How about machines that produces copies in a probabilistic way? A probabilistic copy machine is described by a quantum instrument $\{\mathscr{D}_{yes}, \mathscr{D}_{no}\}$, where the outcome *yes* indicates that the machine succeeded in producing two copies, while the outcome *no* indicates that the machine failed. It turns out that, conditionally to the outcome *yes*, it is actually possible to copy *some* sets of non-orthogonal quantum states without error: precisely, for every set of linearly-independent states one can invent a probabilistic copy machine that copies these states perfectly (Duan and Guo, PRL 1998). However, no probabilistic machine can copy perfectly a set of linearly independent states, like $\{|0\rangle, |1\rangle, |+\rangle, |-\rangle\}$.

6.4 Consequences of the no-cloning theorem

The no-cloning theorem has many important consequences and is the starting point of many other applications in quantum information, such as quantum cryptography.

1. **A no-distinguishability theorem.** Using the no-cloning theorem it is easy to prove something that probably you suspected already from the beginning of the course: if two distinct states $|\varphi_0\rangle$ and $|\varphi_1\rangle$ are not orthogonal, then there is no way to distinguish them without error. Let us put this statement in a precise form:

   Corollary 1 (No-distinguishability theorem) *It is impossible to construct a machine that distinguishes perfectly between two non-orthogonal states $|\varphi_0\rangle$ and $|\varphi_1\rangle$.*

   Proof. Again, we prove this result by contrapositive: we can prove that if there is a way to distinguish between two states perfectly, then they *must* be orthogonal. Indeed, if there is a way to distinguish between two states $|\varphi_0\rangle$ and $|\varphi_1\rangle$, then there must be a quantum measurement, described by some POVM $\{P_0, P_1\}$, such that

   $$p(i|j) = \langle\varphi_j|P_i|\varphi_j\rangle = \delta_{ij}.$$

   If this condition is true, we can construct a machine that copies these states: indeed, our machine can simply perform the measurement with

POVM $\{P_0, P_1\}$, in order to identify the state, and, after the state has been identified, can prepare two perfect copies of it. Mathematically, this machine is described by the channel $\mathscr{C}$ defined as

$$\mathscr{C}(\rho) := \mathrm{Tr}[P_0\rho]\ |\varphi_0\rangle\langle\varphi_0| \otimes |\varphi_0\rangle\langle\varphi_0| + \mathrm{Tr}[P_1\rho]\ |\varphi_1\rangle\langle\varphi_1| \otimes |\varphi_1\rangle\langle\varphi_1|.$$

Clearly, this channel satisfies the condition $\mathscr{C}(|\varphi_i\rangle\langle\varphi_i|) = |\varphi_i\rangle\langle\varphi_i| \otimes |\varphi_i\rangle\langle\varphi_i|$. By the no-cloning theorem, we then get that $|\varphi_0\rangle$ and $|\varphi_1\rangle$ must be orthogonal. ■

As we have just seen, two pure states are perfectly distinguishable if and only if they are orthogonal. An obvious consequence of this fact is that, if you have a single quantum system in some unknown pure state, there you have no way to establish exactly what is the state of your system. For example, if in your lab you have a single photon with unknown polarization, there is no way to establish what is the polarization. To discover an unknown direction of polarization, you need a large (ideally infinite) number of photons that are equally polarized along that direction.

2. **Quantum money (Wiesner, 1970).** The no-cloning theorem is not only an impossibility result: it is also a result that opens the way to new opportunities. One is to solve a problem that is impossible to solve in the classical world: making money that cannot be faked.

   Since in the classical world there are no limits on the ability to copy data, in principle every banknote that we use in our everyday life can be copied. Maybe it is hard to do it, because the banknote has some complex pattern that is difficult to copy (such as a complex image with many layers of colour), but in principle, with a more powerful technology, one can always fake this banknote. Now, thanks to the no-cloning theorem, the situation is different in the quantum case. Imagine the following situation: a bank can produce a banknote containing a particle in one of the four states $\{|0\rangle, |1\rangle, |+\rangle, |-\rangle\}$. The bank can check if the banknote is valid by performing a measurement in the right basis(either $\{|0\rangle, |1\rangle\}$ or $\{|+\rangle, |-\rangle\}$) and checking that the outcome is correct. A counterfeiter who wants to copy the banknote would have to copy the state of the particle. But making perfect copies is forbidden by the no-cloning theorem! And even if we consider imperfect copies, copying the banknote becomes harder and harder if we put more particles on it. For a banknote containing N particles in a product state like $|0\rangle|+\rangle|0\rangle|0\rangle|-\rangle|1\rangle \cdots |1\rangle|1\rangle$ the probability that a fake banknote passes the test set up by the bank goes to 0 exponentially with N. This means that there is a security level guaranteed in principle by the laws of quantum mechanics.

3. **Secure key distribution: the BB84 protocol** Another striking example where the no-cloning theorem opens the way to a new opportunity is the case of quantum key distribution. Key distribution is the task where

two parties, Alice and Bob, try to establish a string of bits (the *secret key*) that is known only to them. The quantum protocol that was invented to achieve this goal is the BB84 protocol, named after the name if its inventors, Bennett and Brassard, and after the date of its publication, 1984. In this protocol, Alice encodes a bit of secret key in the state of a qubit. With probability $\frac{1}{2}$ she encodes the bit in the basis $\{|0\rangle, |1\rangle\}$(i.e. she prepares the state $|0\rangle$ when the bit value is "0", and she prepares the state $|1\rangle$ when the bit value is "1"), with probability $\frac{1}{2}$ she encodes the bit in the basis $\{|+\rangle, |-\rangle\}$(preparing $|+\rangle$ for value "0" and $|-\rangle$ for value "1"). Then she sends the qubit to Bob over an insecure transmission line controlled by an eavesdropper Eve. In order to be in the same situation of Bob, Eve should be able to make two copies of the state and to keep one copy for herself. In other words, she should be able to copy an unknown state in the set $\{|0\rangle, |1\rangle, |+\rangle, |-\rangle\}$. But this is forbidden by the no-cloning theorem!

Intuitively, the best that Eve can do is to copy either the states in the computational basis, or the states in the Fourier basis. However, in this way there is a high probability that Alice and Bob will discover that she is trying to steal the key. If Alice and Bob conclude that their communication is not secure, they can just throw away the key that they generated. Of course, one can consider a more detailed analysis that takes into account the possibility that Eve makes some *approximate* copies. In this case, in order to guarantee security, one has to consider a quantitative tradeoff between the probability that Alice and Bob discover the presence of Eve, and the amount of information that Eve can extract. I will not take you through the detailed analysis, but instead I will give you the conclusion: the BB84 protocol allows Alice and Bob to generate a string of bits $\mathbf{k} = (k_1, \ldots, k_N)$, with the guarantee that these bits are, with high probability, unknown to Eve. Once Alice and Bob have such string of bits, they can use it to communicate secretly: if Alice wants to communicate the secret message $\mathbf{x} = (x_1, \ldots, x_N)$, she can encrypt the message as $\mathbf{x}' = \mathbf{x} \oplus \mathbf{k} := (x_1 \oplus k_1, \ldots, x_N \oplus k_N)$ and send it to Bob over an insecure communication channel. No matter if Eve reads the encrypted message $\mathbf{x}'$, she cannot figure out anything about the original message $\mathbf{x}$, because she does not know the key $\mathbf{k}$. For Bob, instead, everything is easy: to decode the message he only has to sum each bit of the message with each bit of the secret key: indeed, one has $\mathbf{x}' \oplus \mathbf{k} = (x_1 + 2k_1, \ldots, x_N + 2k_N) = (x_1, \ldots, x_N) = \mathbf{x}$.

6.5 Quantum teleportation

We saw that it is impossible to build up a quantum copy machine that can copy the data written in an unknown quantum state. Now we will ask whether we can construct the quantum version of a fax machine, that sends quantum data from one place to another, by only sending some classical message over the

telephone line.

Suppose that Alice and Bob are far apart. Alice has one qubit A in some state $|\varphi\rangle$ and Bob has another qubit B in some other state. Can Alice transfer the state $|\varphi\rangle$ from her qubit to Bob's qubit?

For example, Alice could have a photon that is polarized in direction θ and may want to transfer the polarization of her photon A to the photon B.

The boring way to transfer the state of Alice's photon to Bob's photon is to send qubit A to Bob, who will apply a SWAP gate to it and to photon B. But is it possible to transmit *just the state*, without sending the photon? This is what we mean be a "*quantum fax*", where the information contained in Alice's qubit is first "scanned" and then transmitted to Bob, who reconstructs the state of Alice's qubit on his side.

6.6 Obstacles to constructing a quantum fax machine

Considering what you have learnt about quantum mechanics, constructing fax machine for quantum states seems to be an impossible task. At least, there are three big obstacles that tell us that a quantum fax machine cannot work in the same way of the fax machines that we know in our everyday life:

1. **The information on Alice's side must be destroyed.** The fax machines that we use nowadays scan a document and transmit the content to the receiver, where the fax machine produces an identical copy of the original document. However, by the no-cloning theorem, it is impossible to produce a second copy of $|\varphi\rangle$ on Bob's side. In order for Alice to transmit her quantum document to Bob, she must destroy her copy of the document!

2. **Alice cannot convert a quantum state into classical data.** An ordinary fax machine scans the document and converts it into some classical message that is used by the receiver to reconstruct the document. This approach cannot work with a quantum document: indeed, we know that there is no measure-and-prepare channel $\mathscr{C}(\rho) = \sum_{n=1}^{N} \mathrm{Tr}[P_n\rho]\ \rho_n$ such that $\mathscr{C}(\rho) = \rho$ for every density matrix ρ [cf. Exercises from chapter 4]. This means that a protocol where

   - Alice measures her qubit with a POVM $\{P_n\}_{n=1}^{N}$;
   - Alice communicates the measurement outcome "n" to Bob;
   - Bob prepares the state ρ_n when the outcome is "n".

   cannot work.

   In summary, *if you want to fax a quantum document, you are not allowed to read it!*

3. **The amount of classical bits needed to describe a quantum state is infinite.** *Even if Alice knows the state of her qubit*, in order to describe the state of her qubit to Bob she has to specify two complex numbers, namely the coefficients α and β in the linear combination $|\varphi\rangle = \alpha|0\rangle + \beta|1\rangle$. This means that she has to send to Bob an **infinite amount of classical bits**!

6.7 Quantum teleportation

Quantum teleportation seems to be an impossible task... However, using entanglement this task can be accomplished, and surprisingly, *Alice has only to send to Bob 2 classical bits!* The way to achieve this goal is a spectacular quantum protocol, known as *quantum teleportation* and invented by Bennett, Brassard, Crépeau, Jozsa, Peres, Wotters [Phys. Rev. Lett. **70**, 1895-1899 (1993)].

The resource used by the quantum teleportation protocol is an extra qubit, A', which is entangled with Bob's qubit B, so that the two qubits together are in the Bell state $|\Phi^+\rangle_{A'B}$.

The protocol consists in three simple steps:

1. Alice measures qubits A and A' together on the Bell basis. To denote the four states of the Bell basis, here we will use the notation

$$|\Phi_0\rangle := \frac{|I\rangle\rangle}{\sqrt{2}} \qquad |\Phi_1\rangle := \frac{|\sigma_x\rangle\rangle}{\sqrt{2}}$$

$$|\Phi_2\rangle := \frac{|\sigma_y\rangle\rangle}{\sqrt{2}} \qquad |\Phi_3\rangle := \frac{|\sigma_z\rangle\rangle}{\sqrt{2}}.$$

2. Alice sends the outcome of her measurement to Bob.

3. Bob applies on qubit B a unitary gate U_n that depends on the outcome $n = 0, 1, 2, 3$ in the following way:

$$U_0 = I, U_1 = \sigma_x, U_2 = \sigma_y, U_3 = \sigma_z.$$

Since there are 4 possible outcomes of the Bell measurement, the amount of communication required by the protocol is just of 2 bits. The magic of teleportation is that, after the three simple steps above, qubit B will be in the state $|\varphi\rangle$! In the next sections we will see how this is possible.

6.8 Describing quantum teleportation as a quantum instrument

If you thought that quantum instruments, quantum operations, and the Choi matrix are just mathematical complications, this will be a good chance to change your mind: these notions are essential if we want to understand quantum teleportation in depth.

First of all, note that the teleportation protocol is nothing but a particular way to implement a quantum instrument with input system A and output system B. For the moment, let us ignore the unitary gates applied by Bob. Everything that happens up to this point can be simulated by the following recipe:

1. prepare qubits A' and B in the state $|\Phi_0\rangle_{BA'}$

2. put qubit A together qubits A' and B, and finally,

3. measure A' and A on the Bell basis.

The measurement in the final step induces a random transformation of the state of system A into a state of system B. This transformation is described by a quantum instrument $\{\mathscr{Q}_n\}_{n=0}^3$, where each quantum operation $\mathscr{Q}_n$ corresponds to the evolution of the system conditional to a particular outcome of the Bell measurement. Precisely, for outcome n the state of qubit A will be transformed into the state of qubit B given by

$$\mathscr{Q}_n(\rho_A) := \mathrm{Tr}_{A'A}[(I_B \otimes P_n)(|\Phi_0\rangle\langle\Phi_0|_{BA'} \otimes \rho_A)]\,, \tag{6.1}$$

where $P_n = |\Phi_n\rangle\langle\Phi_n|$ is the POVM operator corresponding to outcome n.

Now, the magic of quantum teleportation protocol is that the quantum instrument $\{\mathscr{Q}_n\}$ can can be realized by another, equivalent recipe, namely

1. Alice sends qubit A to Bob and Bob transfers the state of system A to system B

2. apply a unitary gate U_n to system B, with the value of n chosen at random with probability 1/4.

Here the first step is described by a quantum channel[1] from A to B that, in term of matrices, is just the identity map: it only transforms the qubit A with density matrix ρ into the qubit B with density matrix ρ. Let us call this channel $\mathscr{T}_{A\to B}$ and write

$$\mathscr{T}_{A\to B}(\rho_A) = \rho_B\,,$$

for every 2×2 density matrix ρ. The combination of the two steps above we obtain a quantum instrument, with quantum operations $\{\mathscr{Q}'_n\}$ defined by

$$\mathscr{Q}'_n(\rho_A) = \frac{1}{4}\,\mathscr{U}_n\left[\mathscr{T}_{A\to B}(\rho_A)\right]\,,$$

[1] Physically, one way to realize the channel $\mathscr{T}_{A\to B}$ through a unitary evolution is to

1. prepare qubit B in a fixed state $|0\rangle$

2. apply a SWAP gate to qubits A and B

3. throw away qubit A.

where $\mathscr{U}_n$ is the unitary channel $\mathscr{U}_n(\rho_B) = U_n \rho_B U_n^\dagger$, which transforms states of system B. Now, the surprising thing is that the quantum operation $\mathscr{Q}'_n$ is exactly the same as $\mathscr{Q}_n$: in other words

$$\mathrm{Tr}_{A'A}[(I_B \otimes P_n)(|\Phi_0\rangle\langle\Phi_0|_{BA'} \otimes \rho_A)]\,(\rho) = \frac{1}{4}\,\mathscr{U}_n\left[\mathscr{T}_{A\to B}(\rho_A)\right]\,, \tag{6.2}$$

for every possible outcome n. The proof of Eq. (6.2) is provided in the Appendix, where we analyze the teleportation protocol in general dimension $d \geq 2$. Note the power of the math: on the two sides of the equation we have two abstract transformations of the quantum state that arise in two completely different ways—the teleportation protocol on the left-hand-side and the physical transfer of qubit A from Alice to Bob in the right-hand-side!

Of course, Eq. (6.2) means that, if Alice tells to Bob the outcome of her measurement, Bob can cancel the unitary gate U_n, and in this way he gets exactly the state ρ. Hence, the heart of teleportation is just Eq. (6.2).

6.9 Constructing the quantum teleportation protocol from quantum steering

To understand Eq. (6.2), one approach is just to do a brute-force calculation and to verify that it holds. However, there is something more interesting and more deep that can be said about Eq. (6.2): in fact, this equation can be derived from the property of quantum steering, which we saw in the previous chapter.

Let us see how. First, we can start from something that you know very well, namely that the marginal of the Bell state is maximally mixed: in particular, one has

$$\mathrm{Tr}_{A'}[|\Phi_0\rangle\langle\Phi_0|_{BA'}] = \frac{I_B}{2}\,.$$

Now, let us introduce another qubit R, called the *reference*, and let us consider a Bell state of A and R. Of course, we have

$$\mathrm{Tr}_{A}[|\Phi_0\rangle\langle\Phi_0|_{AR}] = \frac{I_R}{2}\,.$$

As a consequence, the marginal state of $|\Phi^+\rangle_{BA'}|\Phi^+\rangle_{AR}$ on systems BR is maximally mixed:

$$\mathrm{Tr}_{A'A}[|\Phi_0\rangle\langle\Phi_0|_{BA'} \otimes |\Phi_0\rangle\langle\Phi_0|_{AR}] = \frac{I_B \otimes I_R}{4}\,.$$

Now, the maximally mixed state can be seen as a random mixture of Bell states:

$$\frac{I_B \otimes I_R}{4} = \frac{1}{4}\sum_{n=0}^{3} |\Phi_n\rangle\langle\Phi_n|_{BR}\,.$$

The steering property of quantum mechanics then tells us that there exists a measurement on the system AA' that generates the Bell states on system BR

with probability 1/4: denoting the POVM of this measurement by $\{P_n\}_{n=0}^3$, we have

$$\mathrm{Tr}_{A'A}\left[(I_B \otimes P_n \otimes I_R)\left(|\Phi_0\rangle\langle\Phi_0|_{BA'} \otimes |\Phi^+\rangle\langle\Phi^+|_{AR}\right)\right] = \frac{1}{4}\,|\Phi_n\rangle\langle\Phi_n|_{BR}\,, \quad (6.3)$$

for every possible $n \in \{0,1,2,3\}$. Note that here we did not specify the POVM $\{P_n\}$. In fact, it is not really so important to know what is $\{P_n\}$: to construct our teleportation protocol, we only need to know that, thanks to the steering theorem, such a POVM exists [2].

Let us compare carefully the two sides of Eq. (6.3):

1. On the right hand side of Eq. (6.3), we have the Bell state $|\Phi_n\rangle_{BR}$. As we know very well, every Bell state can be generated from the Bell state $|\Phi_0\rangle$ by acting with a unitary: for every n, we have $|\Phi_n\rangle_{BR} = (U_n \otimes I_R)|\Phi_0\rangle_{BR}$. Equivalently, we have

   $$|\Phi_n\rangle\langle\Phi_n|_{BR} = (\mathscr{U}_n \otimes \mathscr{I}_R)\left(|\Phi_0\rangle\langle\Phi_0|_{BR}\right)\,,$$

   where $\mathscr{U}_n$ is the unitary channel $\mathscr{U}_n(\rho) = U_n \rho U_n^\dagger$. Moreover, the Bell state of qubits BR can be obtained from the Bell state of qubits AR, by transforming system A into system B with the channel $\mathscr{T}_{A\to B}$. Hence, we have

   $$|\Phi_n\rangle\langle\Phi_n|_{BR} = (\mathscr{U}_n\mathscr{T}_{A\to B} \otimes \mathscr{I}_R)\left(|\Phi_0\rangle\langle\Phi_0|_{AR}\right)\,,$$

   Recalling the definition of the Choi matrix

   $$\Phi_{\mathscr{C}} := (\mathscr{C} \otimes \mathscr{I})(|\Phi_0\rangle\langle\Phi_0|)\,.$$

   we have that, by this definition, $|\Phi_n\rangle\langle\Phi_n|_{BR}$ is the Choi matrix of the unitary channel $\mathscr{U}_n\mathscr{T}_{A\to B}$, namely

   $$|\Phi_n\rangle\langle\Phi_n|_{BR} = \Phi_{\mathscr{U}_n\mathscr{T}_{A\to B}}\,.$$

2. The left-hand-side of Eq. (6.3) can be written as

   $$\mathrm{Tr}_{A'A}\left[(I_B \otimes P_n \otimes I_R)\left(|\Phi_0\rangle\langle\Phi_0|_{BA'} \otimes |\Phi_0\rangle\langle\Phi_0|_{AR}\right)\right] = (\mathscr{Q}_n \otimes \mathscr{I}_R)(|\Phi_0\rangle\langle\Phi_0|_{AR})]$$

   where $\mathscr{Q}_n$ is the quantum operation defined by

   $$\mathscr{Q}_n(\rho_A) := \mathrm{Tr}_{A'A}[(I_B \otimes P_n)(|\Phi_0\rangle\langle\Phi_0|_{BA'} \otimes \rho_A)]\,, \quad (6.4)$$

   In other words, the left-hand-side of Eq. (6.3) is the Choi matrix of the quantum operation $\mathscr{Q}_n$.

[2] However, if you are curious to know what POVM is $\{P_n\}$, the answer is: the measurement on the Bell basis, of course!

Putting the above observations together, Eq. (6.3) can be re-written as an equality between two Choi matrices:

$$\begin{aligned}\Phi_{\mathscr{Q}_n} &= \frac{1}{4}\,\Phi_{\mathscr{U}_n \mathscr{T}_{A\to B}} \\ &= \Phi_{\frac{1}{4}\mathscr{U}_n \mathscr{T}_{A\to B}}\,.\end{aligned}$$

But we know from the previous chapter that every map is in one-to-one correspondence with its Choi matrix: therefore, we conclude that

$$\mathscr{Q}_n = \frac{1}{4}\,\mathscr{U}_n \mathscr{T}_{A\to B}\,,$$

which is the equation that we wanted to prove.

6.10 Some comments on quantum teleportation

The quantum teleportation protocol is so different from everything we know from our everyday classical world that some comments are in order.

1. **Quantum teleportation and no information without disturbance.** One of the problems that we mentioned, was that Alice should not try to read her quantum document in order to sent it to Bob. But in the teleportation protocol she performs a *measurement* on A and A'! Isn't this in contradiction with that we said before? The answer is *no*, and the reason is that the measurement on the Bell basis *does not extract any information about Alice's state!* The probability of the outcome n is always 1/4, independent of the state of Alice's system.

   In other words, from the beginning to the end of the protocol, Alice and Bob do not need to know the state $|\varphi\rangle$ at all! The quantum teleportation protocol is a completely blind transfer of information from one place to another and can be implemented by two parties, Alice and Bob, that have no idea of what this information is.

2. **Quantum teleportation and the no-signalling theorem.** One intriguing question is: what happens to the information that was in Alice's qubit at the moment of the Bell measurement? In a sense, this "information" has been transferred instantaneously to Bob's lab. However, this "information" is not information for Bob: because, before he receives the classical communication from Alice, he only knows that he has, with probability 1/4, one of the possible states

   $$|\varphi\rangle \qquad \sigma_x|\varphi\rangle \qquad \sigma_y|\varphi\rangle \qquad \sigma_z|\varphi\rangle\,.$$

   Averaging over this possibilities, the state of Bob's qubit alone, before

known the outcome of Alice's measurement, is

$$\begin{aligned}\rho_B &= \frac{1}{4}\sum_{n=0}^{3} U_n|\varphi\rangle\langle\varphi|U_n^\dagger \\ &= \frac{I}{2} \qquad \forall|\varphi\rangle\,.\end{aligned}$$

In other words, Bob has no clue of what the state of his system is, before receiving Alice's measurement outcome: the information is virtually there, but in order to unlock it, he needs to know the value of the classical message from Alice. This tells us that it is not possible to use teleportation to send a signal faster than the speed of light: in order to reconstruct the state $|\varphi\rangle$, Bob needs to wait until he receives Alice's message. But the the message has to be carried by some physical system, and therefore, its speed is bounded by the speed of the light in the vacuum.

3. **The resource inequality of quantum teleportation.** In an abstract sense, we can think of quantum teleportation as a kind of one way to transform some input resources into some other resource: specifically, we can summarize the protocol in a sort of "chemical reaction":

   "1 Bell state + 2 bits of classical communication $\Rightarrow$ 1 qubit of quantum communication."

   Later in this course, we will see that the Bell state of two qubits can be considered as a *standard unit of entanglement*, called *ebit*. Calling the bits of classical communication *cbits* and the bits of quantum communication *qubits*, we can then put the balance in the more compact form:

   $$\boxed{1 \text{ ebit} + 2 \text{ cbits} \Rightarrow 1 \text{ qubit}}$$

   Note that this is an *inequality* of resources: one qubit of quantum communication is not enough to generate one ebit *and* two cbits. What we can do with one qubit of quantum communication is just to generate an ebit: Alice can prepare two qubits AB in the Bell state $|\Phi^+\rangle$ and send qubit B to Bob. This will generate an ebit of entanglement between Alice and Bob, but no additional classical bits:

   $$\boxed{1 \text{ qubit} \Rightarrow 1 \text{ ebit}}$$

   Now, *if we assume that classical communication is a free resource*, then we get a resource equality:

   $$\boxed{1 \text{ ebit} = 1 \text{ qubit} \qquad \text{modulo classical communication}}$$

   Assuming that classical communication is cheap is often a well-motivated assumption: at least communication over the phone is practically much easier than transmitting quantum states!

4. **Quantum teleportation and dense coding.** Recall the dense coding protocol, that we saw in chapter 2: using as a resource two qubits in a Bell state, dense coding allows Alice to communicate 2 classical bits by sending only one qubit. We can summarize this fact in the resource inequality:

$$\boxed{1 \text{ ebit} + 1 \text{ qubit} \Rightarrow 2 \text{ cbits}.}$$

It is interesting to compare this inequality with the one of teleportation: in this case, *if we assume that entanglement is a free resource*, we get the resource equality

$$\boxed{1 \text{ qubit} = 2 \text{ cbits} \qquad \text{modulo entanglement}}$$

However, assuming that entanglement is a free resource is not truly realistic: maintaining an entangled state, especially across a big distance, is a very hard task!

6.11 Quantum teleportation for d-dimensional systems

Let us see briefly how to generalize the quantum teleportation protocol to d-dimensional quantum systems with $d \geq 2$. In order to transfer the state of a d-dimensional system A, the resource is a Bell state of two d-dimensional systems A' and B. For d-dimensional systems, the Bell basis is defined as follows:

$$|\Phi_{p,q}\rangle := (S^p M^q \otimes I)|\Phi\rangle$$

where S and M are the shift and multiply operators that we already encountered often in the course

$$S = \sum_{n=1}^{d} |(n+1) \mod d\rangle\langle n| \qquad M = \sum_{n=1}^{d} e^{\frac{2\pi i n}{d}} |n\rangle\langle n|.$$

Exercise 21 Check that the states $\{|\Phi_{p,q}\rangle\}_{p,q=0}^{d-1}$ are an orthonormal basis.

In the d-dimensional version of the teleportation protocol, Alice measures on the Bell basis and sends to Bob the outcome, now described by the two numbers p and q. When Bob receives the outcome (p, q), he performs on his qubit the unitary gate $S^p M^q$. The proof that this protocol works is provided in the Appendix.

6.12 Chapter summary

In this chapter we encountered our first two examples of quantum machines: machines that try to copy the state of a quantum system and machines that try to transfer the state from one system to another. The key points are that non-orthogonal quantum states cannot be copied (no-cloning theorem) and cannot be perfectly distinguished. This opens new possibilities for secure protocols (cf. quantum money, BB84 protocol). Despite the impossibility to copy and distinguish non-orthogonal states, quantum states can be teleported. Quantum teleportation leads to the resource inequality 1 ebit + 2 cbits $\Rightarrow$ 1 qubit.

Appendix: explicit calculation of the quantum teleportation protocol

Here we prove that the teleportation protocol works for quantum systems of arbitrary dimension $d \geq 2$. Mathematically, this means proving the equation

$$\mathrm{Tr}_{AA'}\left[(|\Phi_{pq}\rangle\langle\Phi_{pq}|_{AA'} \otimes I_B)(\rho_A \otimes |\Phi_{00}\rangle\langle\Phi_{00}|_{A'B})\right] = \frac{1}{d^2}\, U_{pq}^{\dagger}\, \rho_B U_{pq} \tag{6.5}$$

for every p and q in $\{0, \ldots, d-1\}$ and for every density matrix ρ. Here the notation ρ_A (ρ_B) means that the operator ρ acts on the Hilbert space $\mathcal{H}_A$ ($\mathcal{H}_B$).

Since every density matrix is a mixture of the rank-one density matrices corresponding to pure states, it is enough to prove the equation for pure states, that is, to prove the equation

$$\mathrm{Tr}_{AA'}\left[(|\Phi_{pq}\rangle\langle\Phi_{pq}|_{AA'} \otimes I_B)(|\varphi\rangle\langle\varphi|_A \otimes |\Phi_{00}\rangle\langle\Phi_{00}|_{A'B})\right] = \frac{1}{d^2} U_{pq}^{\dagger}|\varphi\rangle\langle\varphi|_B U_{pq} \tag{6.6}$$

for every vector $|\varphi\rangle \in \mathbb{C}^d$.

The proof of Eq. (6.6) uses a number of intermediate results, provided in the following.

Property 7 For $\mathcal{H}_A \simeq \mathcal{H}'_A \simeq \mathcal{H}_B \simeq \mathbb{C}^d$ and for every vector $|\varphi\rangle \in \mathbb{C}^d$ we have the relations

$$\left(\langle\!\langle I|_{AA'} \otimes I_B\right)\, \left(|\varphi\rangle_A \otimes |I\rangle\!\rangle_{A'B}\right) = |\varphi\rangle_B \tag{6.7}$$

and

$$\left(\langle\varphi|_A \otimes \langle\!\langle I|_{A'B}\right)\, \left(|I\rangle\!\rangle_{AA'} \otimes I_B\right) = \langle\varphi|_B\,. \tag{6.8}$$

Proof. Eq. (6.8) can be obtained from Eq. (6.7) by taking the adjoint on both sides of the equality. Hence, it is enough to prove Eq. (6.7). This can be done as follows:

$$\begin{aligned}
\left(\langle\!\langle I|_{AA'} \otimes I_B\right)\left(|\varphi\rangle_A \otimes |I\rangle\!\rangle_{A'B}\right) &= \sum_{m,n=1}^{d} \left(\langle m|_A \otimes \langle m|_{A'} \otimes I_B\right)\left(|\varphi\rangle_A \otimes |n\rangle_{A'} \otimes |n\rangle_B\right) \\
&= \sum_{m,n=1}^{d} \langle m|\varphi\rangle \underbrace{\langle m|n\rangle}_{\delta_{mn}} |n\rangle_B \\
&= \sum_{m=1}^{d} \langle m|\varphi\rangle |m\rangle_B \\
&= |\varphi\rangle_B
\end{aligned}$$

■

As a consequence of the above result, we obtain the following

Property 8 (Probabilistic teleportation) For every vector $|\varphi\rangle \in \mathbb{C}^d$ one has

$$\mathrm{Tr}_{AA'}\left[(|\Phi_{00}\rangle\langle\Phi_{00}| \otimes I_B)(|\varphi\rangle\langle\varphi|_A \otimes |\Phi_{00}\rangle\langle\Phi_{00}|_{A'B})\right] = \frac{1}{d^2}\ |\varphi\rangle\langle\varphi|_B \tag{6.9}$$

Proof.
We have

$$\begin{aligned}
&\mathrm{Tr}_{AA'}\left[(|\Phi_{00}\rangle\langle\Phi_{00}|_{AA'} \otimes I_B)(|\varphi\rangle\langle\varphi|_A \otimes |\Phi_{00}\rangle\langle\Phi_{00}|_{A'B})\right] \\
&\qquad = (\langle\Phi_{00}|_{AA'} \otimes I_B)(|\varphi\rangle\langle\varphi|_A \otimes |\Phi_{00}\rangle\langle\Phi_{00}|_{A'B})(|\Phi_{00}\rangle_{AA'} \otimes I_B) \\
&\qquad = \left[(\underbrace{\langle\Phi_{00}|_{AA'}}_{\frac{\langle\langle I|_{AA'}}{\sqrt{d}}} \otimes I_B)(|\varphi\rangle_A \otimes \underbrace{|\Phi_{00}\rangle_{A'B}}_{\frac{|I\rangle\rangle_{A'B}}{\sqrt{d}}})\right]\left[(\langle\varphi|_A \otimes \underbrace{\langle\Phi_{00}|_{A'B}}_{\frac{\langle\langle I|_{A'B}}{\sqrt{d}}})(\underbrace{|\Phi_{00}\rangle_{AA'}}_{\frac{|I\rangle\rangle_{AA'}}{\sqrt{d}}} \otimes I_B)\right] \\
&\qquad = \frac{1}{d^2}\left[\ (\langle\langle I|_{AA'} \otimes I_B)\ (\,|\varphi\rangle_A \otimes |I\rangle\rangle_{A'B})\right]\left[(\langle\varphi|_A \otimes \langle\langle I|_{A'B})\ (|I\rangle\rangle_{AA'} \otimes I_B)\ \right] \\
&\qquad = \frac{1}{d^2}\,|\varphi\rangle\langle\varphi|_B
\end{aligned}$$

having used Eqs. (6.7) and (6.8) for the last equality. ■

The meaning of Eq. (6.9) is the following: if Alice measures A and A' together in the Bell basis, she will get outcome 00 with probability $p_{00} = \frac{1}{d^2}$ and the conditional state of Bob's qubit will be $|\varphi\rangle\langle\varphi|_B$. Note that *this is already teleportation, with probability* $\frac{1}{d^2}$*:* if Alice and Bob are lucky and the Bell measurement gives outcome 00, then Bob does not need to do anything.

Teleportation with probability $p_0 = \frac{1}{d^2}$ is already surprising, but the quantum teleportation protocol does even better. We are almost ready to show that the teleportation protocols works as promised. The last step is the following

Property 9 For every vector $\varphi \in \mathbb{C}^d$ and for every operator $U : \mathbb{C}^d \to \mathbb{C}^d$ one has

$$(\langle\langle U|_{AA'} \otimes I_B)\ (|\varphi\rangle_A \otimes |I\rangle\rangle_{A'B}) = U^\dagger|\varphi\rangle_B\,. \tag{6.10}$$

and

$$(\langle\varphi|_A \otimes \langle\langle I|_{A'B})\ (|U\rangle\rangle_{AA'} \otimes I_B) = \langle\varphi|_B U\,. \tag{6.11}$$

Proof. Eq. (6.11) can be obtained from Eq. (6.10) by taking the adjoint on both sides. Hence, we only need to prove Eq. (6.10). This can be done as follows:

$$
\begin{aligned}
(\langle\!\langle U|_{AA'} \otimes I_B)(|\varphi\rangle_A \otimes |I\rangle\!\rangle_{A'B}) &= (\langle\!\langle I|_{AA'} \otimes I_B)U^\dagger|\varphi\rangle_A \otimes |I\rangle\!\rangle_{A'B}) \\
&= (\langle\!\langle I|_{AA'} \otimes I_B)(|\psi\rangle_A \otimes |I\rangle\!\rangle_{A'B}) \qquad |\psi\rangle := U^\dagger|\varphi\rangle_A \\
&= |\psi\rangle_B \\
&= U^\dagger|\varphi\rangle_B \,,
\end{aligned}
$$

having used Eq. (6.7) for the third equality. ■

We are ready to prove that teleportation works as we promised, that is, to prove Eq. (6.6). The proof is as follows

$$
\begin{aligned}
&\mathrm{Tr}_{AA'}\left[(|\Phi_{pq}\rangle\langle\Phi_{pq}|_{AA'} \otimes I_B)(|\varphi\rangle\langle\varphi|_A \otimes |\Phi_{00}\rangle\langle\Phi_{00}|_{A'B})\right] \\
&\qquad = (\langle\Phi_{pq}|_{AA'} \otimes I_B)(|\varphi\rangle\langle\varphi|_A \otimes |\Phi_{00}\rangle\langle\Phi_{00}|_{A'B})(|\Phi_{pq}\rangle_{AA'} \otimes I_B) \\
&\qquad = \left[(\underbrace{\langle\Phi_{pq}|_{AA'}}_{\frac{\langle\langle U_{pq}|_{AA'}}{\sqrt{d}}} \otimes I_B)\ (|\varphi\rangle_A \otimes \underbrace{|\Phi_{00}\rangle_{A'B}}_{\frac{|I\rangle\!\rangle_{A'B}}{\sqrt{d}}})\right]\ \left[(\langle\varphi|_A \otimes \underbrace{\langle\Phi_{00}|_{AB}}_{\frac{\langle\langle I|_{A'B}}{\sqrt{d}}})\ (\underbrace{|\Phi_{pq}\rangle_{AA'}}_{\frac{|U_{pq}\rangle\!\rangle_{AA'}}{\sqrt{d}}} \otimes I_B)\right] \\
&\qquad = \frac{1}{d^2}\left[\ (\langle\!\langle U_{pq}|_{AA'} \otimes I_B)\ (|\varphi\rangle_A \otimes |I\rangle\!\rangle_{A'B})\ \right]\ \left[\ (\langle\varphi|_A \otimes \langle\!\langle I|_{A'B})\ (|U_{pq}\rangle\!\rangle_{AA'} \otimes I_B)\ \right] \\
&\qquad = \frac{1}{d^2} U_{pq}^\dagger|\varphi\rangle\langle\varphi|_B U_{pq} \,,
\end{aligned}
$$

having used Eqs. (6.10) and (6.11) for the last equality. ■

Chapter 7

Quantum State Discrimination

In this chapter we continue our journey in the world of elementary quantum machines. After the impossible copy machine and the quantum fax, it is time to move to machines that decode a classical message x encoded into a quantum state ρ_x. For simplicity, we will start from the case where the message is a single bit, encoded into two possible states ρ_0 and ρ_1.

7.1 The minimum error state discriminator

In the previous chapter we observed that it is impossible to distinguish perfectly between two non-orthogonal quantum states, even if we use arbitrary POVMs: when we try to distinguish between non-orthogonal states there will be always an error.

However, in some cases the error can be very small. For example, suppose that we have a photon and that we know that the photon is polarized either horizontally (corresponding to the state $|\varphi_0\rangle = |0\rangle$), or with a polarization angle θ (corresponding to the state $|\varphi_1\rangle = \cos\theta|0\rangle + \sin\theta|1\rangle$). In order to identify the unknown polarization, we can measure on the ONB $\{|0\rangle, |1\rangle\}$ and, if the outcome is "x", declare that the state was $|\varphi_x\rangle$. Clearly, sometimes this strategy will give a wrong answer: if the state is $|\varphi_1\rangle$, then there is a non-zero probability that we declare $|\varphi_0\rangle$. Precisely, the Born rule tells us that the probability is given by

$$p(0|1) = |\langle 0|\varphi_1\rangle|^2 = \cos^2\theta\,.$$

However, when θ is close to 90°, the probability of the error will be close to zero. Practically, this means that we can distinguish *almost perfectly* between the two states $|\varphi_0\rangle$ and $|\varphi_1\rangle$.

Motivated by this example, we can ask the following questions:

- What is the **minimum error** in distinguishing between two quantum states?

- What is the **best measurement** that we have to perform if we want to reduce the error to the minimum?

7.2 A state discrimination game

Let us put the problem of distinguishing between two states in the form of a game, involving one referee, Alice, and one player, Bob. The game is the following

- Alice chooses the value of a bit $x \in \{0,1\}$ at random with probability π_x.
- She encodes the value x in the state $\rho_x \in \mathsf{St}(\mathcal{H})$ of a quantum system.
- She sends the system to Bob and asks him to guess the value of the bit.
- When Bob guesses correctly, he wins one coin. When Bob guesses incorrectly, he loses one coin.

In this game, Bob's strategy will be described by a quantum machine that decodes the classical bit x encoded in the quantum state ρ_x. Of course, Bob's machine is just a measuring device with two possible outcomes $\{0,1\}$, so that, if the measurement outcome is y, then Bob will guess that the value of the bit is y. Mathematically, Bob's measurement is described by a POVM $\{P_y\}_{y\in\{0,1\}}$ and that the probability of guessing y when the true value is x is given by the Born rule

$$p(y|x) = \mathrm{Tr}[P_y \rho_x]\,.$$

Note that the amount of coins that Bob wins when he guesses y and the bit value is x is given by $\omega(x,y) = (-1)^{x+y}$. Hence, his expected payoff is

$$\begin{aligned}\omega &= \sum_{x,y\in\{0,1\}} (-1)^{x+y}\; p(x,y) \\ &= \sum_{x,y\in\{0,1\}} (-1)^{x+y}\;\; \mathrm{Tr}[P_y\rho_x]\,\pi_x \\ &= \mathrm{Tr}\left[(P_0 - P_1)(\pi_0\rho_0 - \pi_1\rho_1)\right]. \end{aligned} \tag{7.1}$$

7.3 Helstrom's decoder

In order to win the maximum amount of coins, Bob has to find the POVM that maximizes his payoff. Luckily for Bob, Helstrom solved this problem many years ago, proving a theorem that tells exactly how much is the maximum payoff and which POVM should be used to achieve it:

Theorem 8 (Helstrom's minimum error decoder) *The maximum of the payoff ω in Eq. (7.1) is*

$$\omega_{\max} = \sum_{n=1}^{d} |\delta_n|\,, \tag{7.2}$$

where δ_n are the eigenvalues of the operator $\Delta = \pi_0\rho_0 - \pi_1\rho_1$. The maximum payoff is achieved by measuring on the basis of the eigenvectors of Δ and by making a guess according to the following rule:

- *if the measurement outcome n corresponds to a positive eigenvalue $\delta_n > 0$, then guess that the state is ρ_0*
- *if the measurement outcome corresponds to a non-positive eigenvalue $\delta_n \leq 0$, then guess that the state is ρ_1.*

Proof. Let us diagonalize Δ as $\Delta = \sum_{n=1}^{d} \delta_n \, |\varphi_n\rangle\langle\varphi_n|$ and write it as $\Delta = \Delta_+ - \Delta_-$, where Δ_+ and Δ_- are the two positive operators defined by

$$\begin{aligned} \Delta_+ &:= \sum_{n:\delta_n>0} \delta_n |\varphi_n\rangle\langle\varphi_n| \\ \Delta_- &:= \sum_{n:\delta_n\leq 0} -\delta_n |\varphi_n\rangle\langle\varphi_n| \, . \end{aligned}$$

With this definition, Bob's payoff is

$$\begin{aligned} \omega &= \mathrm{Tr}\left[(P_0 - P_1)\Delta\right] \\ &= \mathrm{Tr}\left[P_0\Delta_+ - P_1\Delta_+ - P_0\Delta_- + P_1\Delta_-\right] \\ &\leqslant \mathrm{Tr}[P_0\Delta_+ + P_1\Delta_+ + P_0\Delta_- + P_1\Delta_-] \\ &\leqslant \mathrm{Tr}[\Delta_+ + \Delta_-] \\ &= \sum_{n=1}^{d} |\delta_n| \, , \end{aligned}$$

where the inequality in the third line comes from the fact that $\mathrm{Tr}[P_1\Delta_+]$ and $\mathrm{Tr}[P_0\Delta_-]$ are non-negative numbers, while the equality in the last line comes from the normalization of the POVM ($P_0 + P_1 = I$).

We have just proven that, no matter which POVM Bob chooses, his payoff will be upper bounded as $\omega \leq \sum_n |\delta_n|$. On the other hand, it is easy to see that the strategy described in Helstrom's theorem achieves the bound: Helstrom strategy is described by the projective POVM $\{P_0, P_1\}$ with

$$\begin{aligned} P_0 &:= \sum_{n:\delta_n>0} |\varphi_n\rangle\langle\varphi_n| \\ P_1 &:= \sum_{n:\delta_n\leq 0} |\varphi_n\rangle\langle\varphi_n|. \end{aligned}$$

By definition, this POVM gives

$$\begin{aligned} \omega &= \mathrm{Tr}[(P_0 - P_1)\Delta] \\ &= \left(\sum_{n:\delta_n>0} \langle\varphi_n|\Delta|\varphi_n\rangle\right) - \left(\sum_{n:\delta_n\leq 0} \langle\varphi_n|\Delta|\varphi_n\rangle\right) \\ &= \sum_n |\delta_n| \, . \end{aligned}$$

■

7.4 Minimum error discrimination

It is easy to see that Helstrom's POVM is the POVM that maximizes the probability that the Bob's guess y coincides with the true value x. Indeed, Bob's expected payoff can be expressed as

$$\begin{aligned}\omega &= p(0,0) + p(1,1) - p(0,1) - p(1,0) \\ &= p_{succ} - p_{err}\,,\end{aligned}$$

where $p_{succ} := p(0,0) + p(1,1)$ and $p_{err} := p(0,1) + p(1,0)$ are the (average) success and error probabilities, respectively. Since $p_{succ} + p_{err} = 1$, we have

$$\begin{aligned}\omega &= 2p_{succ} - 1 \\ &= 1 - 2p_{err}\,.\end{aligned}$$

Therefore, maximizing ω is equivalent to maximizing p_{succ}, which is in turn equivalent to minimizing p_{err}. The maximum success probability and the minimum error probability will be given by

$$\begin{aligned}p_{succ}^{\max} &= \frac{1}{2}(1 + \omega_{\max}) \\ p_{err}^{\min} &= \frac{1}{2}(1 - \omega_{\max})\,,\end{aligned}$$

where $\omega_{\max}$ is given by Eq. (7.2).

In order to understand better the intuitive meaning of Helstrom's theorem, it is good to apply it to the case of two states that are diagonal in the same basis, like the states

$$\rho_0 := \sum_{n=1}^{d} p_n \,|n\rangle\langle n| \qquad\qquad \rho_1 = \sum_{n=1}^{d} q_n \,|n\rangle\langle n|\,,$$

where $\{p_n\}_{n=1}^d$ and $\{q_n\}_{n=1}^d$ are two probability distributions. In this case, the operator $\Delta = \pi_0\rho_0 - \pi_1\rho_1$ is already diagonalized and its eigenvalues are $\delta_n = \pi_0 p_n - \pi_1 q_n$. Then, Helstrom's theorem tells us that the maximum payoff that Bob can get is

$$\omega_{\max} = \sum_{n=1}^{d} |\pi_0 p_n - \pi_1 q_n|\,,$$

and the optimal strategy is simply to measure on the computational basis $\{|n\rangle\}$ and declare that the state is

- ρ_0 for the outcomes n such that $\pi_0 p_n > \pi_1 q_n$

- ρ_1 for the outcomes n such that $\pi_0 p_n \leq \pi_1 q_n$.

This strategy is very reasonable: Suppose that you perform the measurement and you get outcome n. Then, which probability would you assign to the state ρ_x? The answer is immediate from Bayes' theorem:

$$p(\rho_0|n) = \frac{p_n \pi_0}{p_n \pi_0 + q_n \pi_1}$$
$$p(\rho_1|n) = \frac{q_n \pi_1}{p_n \pi_0 + q_n \pi_1} .$$

Helstrom's strategy just tells you to guess the state that has the largest probability, given the information that you have: ρ_0 if $p(\rho_0|n) > p(\rho_1|n)$ and ρ_1 otherwise. Note that, in general, Helstrom's strategy is different from the maximum likelihood approach, where one chooses the state ρ_x that maximizes the probability of obtain the outcome n: here we have

$$p(n|\rho_0) = p_n$$
$$p(n|\rho_1) = q_n ,$$

which means that Helstrom's strategy and the maximum likelihood approach give different results unless $\pi_0 = \pi_1 = 1/2$.

7.5 Distinguishing between two pure states

Suppose that ρ_0 and ρ_1 are two pure states, given by $\rho_0 = |\varphi_0\rangle\langle\varphi_0|$ and $\rho_1 = |\varphi_1\rangle\langle\varphi_1|$, respectively. In this case, Helstrom's theorem yields the following

Corollary 2 (Helstrom's decoder for two pure states) *The maximum payoff in distinguishing between two pure states $|\varphi_0\rangle$ and $|\varphi_1\rangle$, given with prior probabilities π_0 and π_1, respectively, is given by*

$$\omega_{\max} = \sqrt{1 - 4\pi_0\pi_1 F} \qquad F := |\langle\varphi_0|\varphi_1\rangle|^2 . \tag{7.3}$$

The quantity F is called *fidelity* and plays a very important role in quantum information theory. We will come back to it several times in this course.

Proof. Since the correspondence between the density matrix $\rho_1 = |\varphi_1\rangle\langle\varphi_1|$ and the vector $|\varphi_1\rangle$ is up to a global phase, we can choose without loss of generality a vector $|\varphi_1\rangle$ such that $\langle\varphi_0|\varphi_1\rangle \geq 0$. With this choice, we can write

$$|\varphi_1\rangle = \sqrt{F}|\varphi_0\rangle + \sqrt{1-F}|\varphi_0^\perp\rangle$$

where $|\varphi_0^\perp\rangle$ is a unit vector orthogonal to $|\varphi_0\rangle$. Now, the two vectors $|\varphi_0\rangle$ and $|\varphi_0^\perp\rangle$ are a basis for the two-dimensional subspace containing $|\varphi_0\rangle$ and $|\varphi_1\rangle$. Using the basis $\{|\varphi_0\rangle, |\varphi_0^\perp\rangle\}$, the two quantum states ρ_0 and ρ_1 can be written as the following matrices

$$\rho_0 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \qquad \rho_1 = \begin{pmatrix} F & \sqrt{F(1-F)} \\ \sqrt{F(1-F)} & 1-F \end{pmatrix} ,$$

and we have

$$\begin{aligned}\Delta &= \pi_0\rho_0 - \pi_1\rho_1 \\ &= \pi_0 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} - \pi_1 \begin{pmatrix} F & \sqrt{F(1-F)} \\ \sqrt{F(1-F)} & (1-F) \end{pmatrix} \\ &= \begin{pmatrix} \pi_0 - \pi_1 F & -\pi_1\sqrt{F(1-F)} \\ -\pi_1\sqrt{F(1-F)} & -\pi_1(1-F) \end{pmatrix}.\end{aligned}$$

The eigenvalues of Δ, denoted by δ_+ and δ_- have sum $s = \mathrm{Tr}[\Delta] = \pi_0 - \pi_1$ and product $p = \det(\Delta) = -\pi_0\pi_1(1-F)$, which is always negative (or zero, if $F = 1$). Since the product is negative, the two eigenvalues must have opposite signs, say $\delta_+ \geq 0$ and $\delta_- \leq 0$. Hence, we have

$$\begin{aligned}\omega_{\max} &= |\delta_+| + |\delta_-| \\ &= |\delta_+ - \delta_-| \\ &= \sqrt{(\pi_0 - \pi_1)^2 + 4\pi_0\pi_1(1-F)} \\ &= \sqrt{1 - 4\pi_0\pi_1 F}\end{aligned} \tag{7.4}$$

having used the expression $\delta_\pm = \left(s \pm \sqrt{s^2 - 4p}\right)/2$.

■

Let us apply Eq. (7.3) in a couple of interesting cases:

1. **Distinguishing between two linear polarization states.** Let us go back to the example considered in the introduction, where the quantum system used to encode the bit is the polarization of a photon and the polarization states used for the encoding are $|\varphi_0\rangle = |0\rangle$ and $|\varphi_1\rangle = \cos\theta|0\rangle + \sin\theta|1\rangle$. For simplicity, let us consider the case of uniform prior probability $\pi_0 = \pi_1 = 1/2$. Using Eq. (7.3), we have

   $$\omega_{\max} = |\sin\theta|\,.$$

   Note that Helstrom's strategy allows Bob to win more coins than the naive strategy presented in the introduction: indeed, measuring on the computational basis $\{|0\rangle, |1\rangle\}$ gives Bob the payoff

   $$\begin{aligned}\omega &= \frac{1}{2}(1 + \sin^2\theta - \cos^2\theta) \\ &= \sin^2\theta \\ &= \omega_{\max}^2\,.\end{aligned}$$

2. **Distinguishing between two pure states using multiple copies.** Suppose that Alice decides to help Bob and, instead of preparing just one quantum system, she prepares N identical quantum systems, each of them

in the same state $|\varphi_x\rangle$, where x is the value of the bit that she wants to communicate. Now, Bob has to distinguish between the states

$$|\psi_0\rangle := \underbrace{|\varphi_0\rangle|\varphi_0\rangle\cdots|\varphi_0\rangle}_{N \text{ times}} = |\varphi_0\rangle^{\otimes N}$$

and

$$|\psi_1\rangle := \underbrace{|\varphi_1\rangle|\varphi_1\rangle\cdots|\varphi_1\rangle}_{N \text{ times}} = |\varphi_1\rangle^{\otimes N}.$$

Since we have $|\langle\psi_0|\psi_1\rangle|^2 = |\langle\varphi_0|\varphi_1\rangle|^{2N} = F^N$, where F is the fidelity between $|\varphi_0\rangle$ and $|\varphi_1\rangle$ the probability of error will go to zero exponentially fast with N:

$$\begin{aligned} p_{err} &= \frac{1}{2}\left(1 - \sqrt{1 - 4\pi_0\pi_1 F^N}\right) \\ &\approx \pi_0\pi_1 F^N \qquad\qquad N \gg 1. \end{aligned}$$

If two pure states are distinct (that is if $F < 1$), then the probability of error in distinguishing between them with N copies goes to zero exponentially fast in N!

This is one more reason why we cannot make perfect copies of non-orthogonal quantum states: if perfect cloning were possible, we could use the cloning machine to make the two states $|\varphi_0\rangle$ and $|\varphi_1\rangle$ more distinguishable, making the probability of error as small as we wanted. This would violate Helstrom's theorem!

7.6 The trace norm

Mathematically, the maximum payoff that Bob can achieve in the state discrimination game can be interpreted as the *trace norm* of the operator $\Delta = \pi_0\rho_0 - \pi_1\rho_1$. We will now see the definition of the trace norm and some elementary properties.

Definition and operational meaning. The trace norm of a generic operator is defined as follows:

Definition 4 (Trace-norm) *Let Ψ be an operator from $\mathcal{H}_B$ to $\mathcal{H}_A$ and let $\Psi = \sum_n \lambda_n\ |\alpha_n\rangle\langle\beta_n|$, $\lambda_n > 0$, be its singular value decomposition (SVD). Then, the* trace-norm *of Ψ is given by*

$$\|\Psi\|_1 := \sum_n \lambda_n\,.$$

For a self-adjoint operator $\Delta = \sum_n \delta_n\ |\varphi_n\rangle\langle\varphi_n|$, the bases in the SVD can be chosen to be

$$\begin{aligned} |\beta_n\rangle &:= |\varphi_n\rangle \\ |\alpha_n\rangle &:= \begin{cases} |\varphi_n\rangle & \forall n : \delta_n > 0 \\ -|\varphi_n\rangle & \forall n : \delta_n \leq 0\,. \end{cases} \end{aligned},$$

which implies that the trace-norm is given by $\|\Delta\|_1 = \sum_n |\delta_n|$. Hence, Bob's maximum payoff can be expressed as

$$\omega_{\max} = \|\pi_0\rho_0 - \pi_1\rho_1\|_1 \,.$$

It is easy to check that, like the name suggests, the trace norm satisfies all the properties of a norm:

Exercise 22 Show that the following properties hold:

1. $\|c\Psi\|_1 = |c| \, \|\Psi\|_1$ for every complex number $c \in \mathbb{C}$
2. $\|\Psi\|_1 = 0$ only if $\Psi = 0$
3. $\|\Psi + \Psi'\|_1 \leq \|\Psi\|_1 + \|\Psi'\|_1$.

If you are curious to know why the trace norm is called *trace* norm, this is because $\|\Psi\|_1$ is the trace of an operator. Precisely, it is the trace of the operator $|\Psi|$ defined as

$$|\Psi| := \sum_n \lambda_n |\beta_n\rangle\langle\beta_n| \,.$$

If you are curious to know why the trace norm is denoted by $\|\cdot\|_1$, this is because it belongs to a larger families of norms, called *p-norms*, defined as follows:

Definition 5 (*p*-norm) *For $p > 0$, the* p-norm *of Ψ is given by*

$$\|\Psi\|_p := \left(\sum_n \lambda_n^p\right)^{\frac{1}{p}} \,.$$

p-norms have a useful property, shown in the following

Exercise 23 Show that one has $\|M \otimes N\|_p = \|M\|_p \, \|N\|_p$ for every pair of matrices M and N.

We will have some other chance to meet the p-norms later in the course. For the moment, our interest is in the case $p = 1$. An alternative way to compute the trace-norm is given by the following

Proposition 1 (Alternative characterization of the trace norm) *The trace norm of an operator $\Psi : \mathcal{H}_B \to \mathcal{H}_A$ is given by*

$$\|\Psi\|_1 = \max_{\substack{V \,:\, \mathcal{H}_A \to \mathcal{H}_B \\ V^\dagger V = I_A}} \mathrm{Tr}[\Psi V] \,. \tag{7.5}$$

Proof. Writing Ψ in the singular value decomposition

$$\Psi = \sum_{n=1}^{r} \lambda_n |\alpha_n\rangle\langle\beta_n|$$

we get

$$\begin{aligned}\mathrm{Tr}[\Psi V] &= \sum_{n=1}^{r} \lambda_n \langle \beta_n | V | \alpha_n \rangle \\ &\leqslant \sum_{n=1}^{r} \lambda_n \underbrace{|\langle \beta_n | V | \alpha_n \rangle|}_{\leqslant 1} \\ &\leqslant \sum_{n=1}^{r} \lambda_n \\ &\equiv \|\Psi\|_1 \,.\end{aligned}$$

The bound is achieved when the isometry V satisfies $|\beta_n\rangle = V|\alpha_n\rangle$ for every $n \in \{1, \ldots, r\}$. ■

Properties of the trace norm. We saw that the trace norm is a measure of the distinguishability between the two quantum states ρ_0 and ρ_1. Since the explicit calculation of the trace norm can be hard sometimes, it is useful to know some properties.

Property 10 (Quantum channels cannot increase the distinguishability) For every pair of states ρ_0, $\rho_1 \in \mathsf{St}(\mathcal{H}_A)$, for every pair or prior probabilities π_0, π_1, and for every quantum channel $\mathscr{C} : \mathsf{St}(\mathcal{H}_A) \to \mathsf{St}(\mathcal{H}_B)$ we have

$$\|\pi_0 \mathscr{C}(\rho_0) - \pi_1 \mathscr{C}(\rho_1)\|_1 \leqslant \|\pi_0 \rho_0 - \pi_1 \rho_1\|_1 \,. \tag{7.6}$$

Here, the intuitive meaning of the property is simple: no deterministic physical process can make two states more distinguishable.

Proof. Consider a state discrimination game with the states ρ_0 and ρ_1, given with prior probabilities π_0 and π_1 respectively. By Helstrom's theorem, the maximum payoff in this game is given by

$$\omega = \|\pi_0 \rho_0 - \pi_1 \rho_1\|_1 \,.$$

On the other hand, one strategy to play the game is the following:

1. apply the quantum channel $\mathscr{C}$ to the state ρ_x, $x \in \{0, 1\}$, transforming it into $\rho'_x = \mathscr{C}(\rho_x)$
2. use the minimum error POVM to distinguish between the two states ρ'_0 and ρ'_1.

By Helstrom's theorem, this strategy gives the payoff

$$\begin{aligned}\omega'_{\max} &= \|\pi_0 \rho'_x - \pi_1 \rho'_1\|_1 \\ &= \|\pi \mathscr{C}(\rho_0) - \pi_1 \mathscr{C}(\rho_1)\|_1 \,.\end{aligned}$$

By definition, performing a measurement on system B after the action of the channel $\mathscr{C}$ is just one way to perform a measurement on system A [1]. Since this measurement cannot give a larger payoff than the best possible measurement, we conclude $\omega'_{\max} \leq \omega_{\max}$. ■

Examples illustrating property 1

- Erasure channel $\mathscr{C}(\rho) = \frac{I}{d} \quad \forall \rho$,

$$\|\mathscr{C}(\rho_0) - \mathscr{C}(\rho_1)\|_1 = 0.$$

- Depolarizing channel $\mathscr{C}(\rho) = p\rho + (1-p)\frac{I}{d} \quad p \in (0,1)$,

$$\begin{aligned} \|\mathscr{C}(\rho_0) - \mathscr{C}(\rho_1)\|_1 &= p\|\rho_0 - \rho_1\|_1 \\ &< \|\rho_0 - \rho_1\|_1. \end{aligned}$$

- Quantum-to-classical channel. Let $\{P_n\}_{n=1}^N$ be a POVM on system A and let $\mathscr{C}$ the quantum channel from system A to an N-dimensional system B defined by

$$\mathscr{C}(\rho) = \sum_n \mathrm{Tr}[P_n \rho] \, |n\rangle\langle n| \, .$$

  Essentially, the channel $\mathscr{C}$ transform the quantum state ρ into the classical probability distribution $p(n|\rho) = \mathrm{Tr}[P_n\rho]$. Now, by Property 1 one has

$$\begin{aligned} \|\pi_0\rho_0 - \pi_1\rho_1\|_1 \geq \|\pi_0\mathscr{C}(\rho_0) - \pi_1\mathscr{C}(\rho_1)\|_1 \\ \sum_n |\pi_0 p(n|\rho_0) - \pi_1 p(n|\rho_1)| \, . \end{aligned}$$

  This means that the classical probability distributions $\{p(n|\rho_0)\}$ and $\{p(n|\rho_1)\}$, obtained by measuring a quantum system in the states ρ_0 and ρ_1, respectively, cannot be more distinguishable than the quantum states ρ_0 and ρ_1.

If the channel $\mathscr{C}$ is a unitary gate, then it is easy to see that it does not change the trace norm. This property holds not only for unitary gates, but also to channels that are *correctable*, in the following sense:

Definition 6 *A channel $\mathscr{C}$ from A to B is* correctable *if there exists a channel $\mathscr{R}$ from B to A such that $\mathscr{R}\mathscr{C}$ is the identity channel on system A.*

[1] Precisely, let $\{P'_0, P'_1\}$ be a POVM on B. Writing $\mathscr{C}$ in the Kraus form $\mathscr{C}(\rho) = \sum_i C_i \rho C_i^\dagger$, we can define a POVM on A in the following way

$$\begin{aligned} P_0 &:= \sum_i C_i^\dagger P'_0 C_i \\ P_1 &:= \sum_i C_i^\dagger P'_1 C_i \, . \end{aligned}$$

With this definition, we have $\mathrm{Tr}[P_y\rho] = \mathrm{Tr}[P'_y\mathscr{C}(\rho)]$ for every state ρ and for every outcome y.

Property 11 (Correctable channels preserve the distinguishability) If a channel $\mathscr{C}$ is correctable, then

$$\|\pi_0\mathscr{C}(\rho_0) - \pi_1\mathscr{C}(\rho_1)\|_1 = \|\pi_0\rho_0 - \pi_1\rho_1\|_1 \,,$$

for every pair of states ρ_0 and ρ_1 and for every prior probabilities π_0 and π_1.

Proof. Applying Property 1 to the channel $\mathscr{R}$ and to the states $\mathscr{C}(\rho_x)$, $x = 0, 1$ we obtain

$$\begin{aligned}\|\pi_0\rho_0 - \pi_1\rho_1\|_1 &= \|\pi_0\mathscr{R}\left[\mathscr{C}(\rho_0)\right] - \pi_1\mathscr{R}\left[\mathscr{C}(\rho_1)\right]\|_1 \\ &\leq \|\pi_0\mathscr{C}(\rho_0) - \pi_1\mathscr{C}(\rho_1)\|_1 \,.\end{aligned}$$

Combined with the inequality of Eq. (7.6), this gives the desired equality. ■

For example, if $V : \mathcal{H}_A \to \mathcal{H}_B$ is an isometry ($V^\dagger V = I_A$), then the channel $\mathscr{C}(\rho) = V\rho V^\dagger$ does not change the trace norm. Interestingly, one can also show that if a channel preserves the trace norm for all possible states and prior probabilities, then the channel must be correctable.

7.7 The fidelity between two mixed states

We saw that the trace distance between two pure states $|\varphi_0\rangle$ and $|\varphi_1\rangle$ is a simple function of the fidelity $F = |\langle\varphi_0|\varphi_1\rangle|^2$. This is very useful. Can we have something similar for **mixed states**? **Idea:** Take purifications!

Definition of fidelity for mixed states. Suppose that ρ_0, ρ_1 are two states of system B and that $|\Psi_0\rangle, |\Psi_1\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ are two purifications of them, respectively, i. e.

$$\begin{aligned}\rho_0 &= \mathrm{Tr}_A[|\Psi_0\rangle\langle\Psi_0|] \\ \rho_1 &= \mathrm{Tr}_A[|\Psi_1\rangle\langle\Psi_1|].\end{aligned}$$

Now, the partial trace over A is a quantum channel. Hence, using Property 1 (quantum channels cannot increase the trace-norm) we obtain the bound

$$\begin{aligned}\|\pi_0\rho_0 - \pi_1\rho_1\|_1 &\leqslant \|\pi_0|\Psi_0\rangle\langle\Psi_0| - \pi_1|\Psi_1\rangle\langle\Psi_1|\|_1 \\ &= \sqrt{1 - 4\pi_0\pi_1|\langle\Psi_0|\Psi_1\rangle|^2}.\end{aligned}$$

This upper bound is valid for **every choice of purifications**!

Choosing the best bound, we can define the *fidelity* between two mixed states:

Definition 7 (Fidelity) *The fidelity between two mixed states ρ_0 and ρ_1 is defined as the maximum over all possible purifications $|\Psi_0\rangle, |\Psi_1\rangle$:*

$$F(\rho_0, \rho_1) := \sup_{\mathcal{H}_A} \max_{\substack{|\Psi_0\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B \\ \mathrm{Tr}_A[|\Psi_0\rangle\langle\Psi_0|] = \rho_0}} \max_{\substack{|\Psi_1\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B \\ \mathrm{Tr}_A[|\Psi_1\rangle\langle\Psi_1|] = \rho_1}} |\langle\Psi_0|\Psi_1\rangle|^2.$$

With this definition, we have the bound:

$$\|\pi_0\rho_0 - \pi_1\rho_1\|_1 \leqslant \sqrt{1 - 4\pi_0\pi_1 F(\rho_0, \rho_1)} \tag{7.7}$$

which limits the payoff that Bob can win in a state discrimination game with states $\rho_0 = \rho$ and $\rho_1 = \sigma$.

Remark 1 (State discrimination games with an evil Alice) Right hand side of Eq. (7.7) can be also interpreted as a payoff in a state discrimination game, where now Alice plays *against Bob*. Suppose that Alice is forced to obey the following rule: she has to give Bob a pure state with fixed marginals ρ_x on system B. If her goal is to reduce Bob's payoff, she will choose the purifications $|\Psi_0\rangle$ and $|\Psi_1\rangle$ that are *less distinguishable*. Hence, she will maximize the fidelity $F = |\langle\Psi_0|\Psi_1\rangle|^2$ over all possible purifications.

The definition of fidelity in terms of purifications is conceptually fundamental. However, sometimes it is useful to have an expression that involves less optimizations. An expression that is simpler to calculate is given by Uhlmann's theorem:

Theorem 9 (Uhlmann's theorem) *The fidelity between two states ρ_0 and ρ_1 is given by*

$$F(\rho_0, \rho_1) = \|\sqrt{\rho_0}\sqrt{\rho_1}\|_1^2$$

Proof. Diagonalizing ρ_0 as $\rho = \sum_{m=1}^{r} p_m \, |\beta_m\rangle\langle\beta_m|$ we know that the purification $|\Psi_0\rangle$ must be of the Schmidt form

$$|\Psi_0\rangle = \sum_{m=1}^{r} \sqrt{p_m}|\alpha_m\rangle|\beta_m\rangle$$

for some orthonormal vectors $\{|\alpha_m\rangle\}_{m=1}^{r} \subset \mathcal{H}_A$. In the same way, diagonalizing ρ_1 as $\rho_1 = \sum_{n=1}^{r'} p'_n \, |\beta'_n\rangle\langle\beta'_n|$ we know that the purification $|\Psi_1\rangle$ must be of the Schmidt form

$$|\Psi_1\rangle = \sum_{n=1}^{r'} \sqrt{p'_n}|\alpha'_n\rangle|\beta'_n\rangle$$

for some orthonormal vectors $\{|\alpha'_n\rangle\}_{n=1}^{n'} \subset \mathcal{H}_A$. Hence, we have

$$\langle\Psi_0|\Psi_1\rangle = \sum_{m=1}^{r}\sum_{n=1}^{r'} \sqrt{p_m p'_n} \underbrace{\langle\alpha_m|\alpha'_n\rangle}_{U_{mn}} \langle\beta_m|\beta'_n\rangle$$

matrix elements of a partial isometry of rank $t = \min\{r, r'\} \leq \min\{d_A, d_B\}$.

Extending U_{mn} to a $d_B \times d_B$ matrix and defining the unitary gate $U := \sum_{m,n=1}^{d_B} U_{mn} |\beta'_n\rangle\langle\beta_m|$ we then have $U_{mn} = \langle\beta'_n|U|\beta_m\rangle$. Substituting in the expression of the fidelity, we obtain

$$\begin{aligned}\langle\Psi_0|\Psi_1\rangle &= \sum_{m=1}^{r}\sum_{n=1}^{r'} \sqrt{p_m p'_n}\, \langle\beta'_n|U|\beta_m\rangle\langle\beta_m|\beta'_n\rangle \\ &= \mathrm{Tr}[U\sqrt{\rho_0}\sqrt{\rho_1}]\,.\end{aligned}$$

Taking the modulus and maximizing over all possible unitaries we obtain

$$\begin{aligned}|\langle\Psi_0|\Psi_1\rangle| &= |\,\mathrm{Tr}[U\sqrt{\rho_0}\sqrt{\rho_1}]| \\ &\leq \max_{U:U^\dagger U=I} \mathrm{Tr}[U\sqrt{\rho_0}\sqrt{\rho_1}] \\ &= \|\sqrt{\rho_0}\sqrt{\rho_1}\|_1\ ,\end{aligned}$$

where in the last line we used Proposition 1. The bound can be achieved by choosing the system A to be of the same dimension as system B and by choosing the purifications $|\Phi_0^{\max}\rangle$ and $|\Phi_1^{\max}\rangle$, defined as

$$\begin{aligned}|\Psi_0^{\max}\rangle &= |\,(\sqrt{\rho_0})^T\,\rangle\!\rangle \\ |\Psi_1^{\max}\rangle &= |\,(\sqrt{\rho_1}U_{\max})^T\,\rangle\!\rangle\,,\end{aligned}$$

where $U_{\max}$ is the unitary such that $\mathrm{Tr}[\sqrt{\rho_0}\sqrt{\rho_1}U_{\max}] = \|\sqrt{\rho_0}\sqrt{\rho_1}\|_1$. ■

For example, Uhlmann's theorem can be used to calculate the fidelity between two states that are diagonal in the same basis, such as $\rho_0 = \sum_n p_n|n\rangle\langle n|$ and $\rho_1 = \sum_n q_n|n\rangle\langle n|$. In this case, we have

$$\begin{aligned}F(\rho_0,\rho_1) &= \max_U \sum_n \sqrt{p_n q_n}\langle n|U|n\rangle \\ &= \sum_n \sqrt{p_n q_n}\,.\end{aligned}$$

In classical probability theory, the quantity $B := \sum_n \sqrt{p_n q_n}$ is known as *Bhattacharya coefficient* and is a measure of how close two probability distributions are.

Another easy application of Uhlmann's theorem is to show that the fidelity between a pure state $\rho_0 = |\varphi\rangle\langle\varphi|$ and a general state ρ_1 is given by

$$F(|\varphi\rangle\langle\varphi|,\rho_1) = \langle\varphi|\rho_1|\varphi\rangle\,.$$

Proving this relation is an easy exercise [2].

[2] This is a subtle invitation for you to try.

7.8 Lower bound on the trace norm in terms of the fidelity

By definition, we know that the trace norm can be upper bounded in terms of the fidelity as

$$\|\pi_0\rho_0 - \pi_1\rho_1\|_1 \leq \sqrt{1 - 4\pi_0\pi_1 F(\rho_0, \rho_1)}\,.$$

Can we find a lower bound?

A positive answer is given by the following

Proposition 2 *For every pair of states ρ_0, ρ_1 and for every pair of probabilities π_0, π_1 one has*

$$\|\pi_0\rho_0 - \pi_1\rho_1\|_1 \geqslant 1 - \sqrt{4\pi_0\pi_1 F(\rho_0, \rho_1)}\,.$$

The proof is provided in the Appendix.

Using this lower bound we can prove that the probability of error in distinguishing between ρ_0 and ρ_1 using N identical copies goes to zero exponentially fast in N. Indeed, in this case we have

$$\begin{aligned}\tau_0 &= \underbrace{\rho_0 \otimes \rho_0 \otimes \cdots \otimes \rho_0}_{N \text{ times}} =: \rho_0^{\otimes N} \\ \tau_1 &= \underbrace{\rho_1 \otimes \rho_1 \otimes \cdots \otimes \rho_1}_{N \text{ times}} =: \rho_1^{\otimes N}\,.\end{aligned}$$

The fidelity can be computed with Uhlmann's theorem, which gives

$$\begin{aligned}F(\tau_0, \tau_1) &= \left\| \underbrace{\sqrt{\rho_0}\sqrt{\rho_1} \otimes \sqrt{\rho_0}\sqrt{\rho_1} \otimes \cdots \otimes \sqrt{\rho_0}\sqrt{\rho_1}}_{N \text{ times}} \right\|_1 \\ &= \|\sqrt{\rho_0}\sqrt{\rho_1}\|_1^N\,,\end{aligned}$$

the second equality coming from exercise 23 with $p = 1$. Using Helstrom's theorem, the minimum probability of error is equal to

$$p_{err} = \frac{1}{2}\left(1 - \|\pi_0\tau_0 - \pi_1\tau_1\|_1\right) \leqslant \sqrt{\pi_0\pi_1}\,[\,F(\rho_0, \rho_1)\,]^{\frac{N}{2}}\,.$$

This bound, however, is not tight. Using more refined techniques it is possible to show that the error probability goes to zero at rate C^N, where $C < \sqrt{F(\rho_0, \rho_1)}$. Precisely, one has

$$C = \min_{p:0\leqslant p\leqslant 1} \mathrm{Tr}[\rho_0^p \rho_1^{1-p}].$$

The scaling $p_{err} = O(C^N)$ is called *quantum Chernoff bound*.

7.9 The unambiguous state discriminator

Until now a talked about measurements that distinguish quantum states with the minimum error compatible with the rules of quantum mechanics. But can we invent some measurements that distinguish non-orthogonal states *without error*?

At first sight, this seems to be a silly question: haven't we already proved that it is impossible to distinguish perfectly between non-orthogonal quantum states? However, there is a subtle difference between distinguishing states *perfectly* (i.e. with probability of success equal to 1) and distinguishing them *without errors*:

Consider the following measurement, designed distinguish between two non-orthogonal states $|\varphi_0\rangle = |0\rangle$ and $|\varphi_1\rangle = \cos\theta|0\rangle + \sin\theta|1\rangle$: the measurement that has three outcomes $\{0, 1, ?\}$ where the outcomes 0 and 1 tells us that the state was $|\varphi_0\rangle$ and $|\varphi_1\rangle$, respectively, while the outcome ? tells us "I don't know, I don't want to make a guess for the value of the bit x". The three outcomes correspond to the POVM $\{P_0, P_1, P_?\}$ with

$$\begin{aligned} P_0 &= \frac{1}{1+\cos\theta}|\varphi_1^\perp\rangle\langle\varphi_1^\perp| \qquad & |\varphi_1^\perp\rangle &= \sin\theta|0\rangle - \cos\theta|1\rangle \\ P_1 &= \frac{1}{1+\cos\theta}|\varphi_0^\perp\rangle\langle\varphi_0^\perp| \qquad & |\varphi_0^\perp\rangle &= |1\rangle \\ P_? &= I - P_0 - P_1 \,. \end{aligned}$$

This POVM has a very interesting features: *when it gives an answer*, it never makes mistakes. Indeed, we have $p(0|\rho_1) = p(1|\rho_0) = 0$. Discrimination strategies with this feature are called *unambiguous discrimination strategies*. The price to pay, of course, is that sometimes the measurement gives no answer: the outcome ? occurs with probability

$$p(?|\rho_0) = p(?|\rho_1) = 1 - \frac{\sin\theta^2}{1+\cos\theta} = \cos\theta \,.$$

The good thing about it is that, when the states $|\varphi_0\rangle$ and $|\varphi_1\rangle$ are almost orthogonal $\theta \approx 90°$, the probability that the machine gives no answer is small.

Can we generalize this example to the discrimination of N pure states?

Suppose that Alice has N possible messages for Bob and that she encodes the n-th message in the quantum state $|\varphi_n\rangle$. In order to decode the message without errors Bob should to use a POVM with operators $\{P_n\}_{n=1}^N \cup \{P_?\}$, with the property

$$\langle\varphi_n|P_m|\varphi_n\rangle = p_n\,\delta_{mn} \qquad \forall m, n \in \{1, \ldots, N\}\,, \tag{7.8}$$

where $p_n > 0$ is the probability that the POVM gives outcome n when the state is n.

Eq. (7.8) sets a pretty strict constraint on the states $\{|\varphi_n\rangle\}$: indeed, it requires them to be linearly independent.

Proposition 3 *Unambiguous discrimination is possible only if the states* $\{|\varphi_n\rangle\}_{n=1}^N$ *are linearly independent.*

Proof. For $m \neq n$, note that Eq. (7.8) gives

$$\begin{aligned} 0 &= \langle\varphi_n|P_m|\varphi_n\rangle \\ &= \|\sqrt{P_m}|\varphi_n\rangle\|^2 , \end{aligned}$$

which implies that $\sqrt{P_m}|\varphi_n\rangle = 0$, and, therefore $P_m|\varphi_n\rangle = 0$. Now, if we have $\sum_n c_n|\varphi_n\rangle = 0$ for some coefficients c_n, then, by multiplying by $\langle\varphi_m|P_m$ on both sides of the equality we get

$$c_m\langle\varphi_m|P_m|\varphi_m\rangle = 0 .$$

Since $\langle\varphi_m|P_m|\varphi_m\rangle = p_m > 0$, this implies $c_m = 0$. Hence, the states $\{|\varphi_n\rangle\}$ must be linearly independent. ■

Conversely, we can show that if the states $\{|\varphi_m\rangle\}$ are linearly independent, then unambiguous discrimination is possible. In addition, linear independence is enough to construct a special kind of unambiguous discrimination strategy, where the probability $\langle\varphi_n|P_n|\varphi_n\rangle$ is the same for all possible values of n.

Proposition 4 (The equal-probability unambiguous decoder) *If the states* $\{|\varphi_n\rangle\}$ *are linearly independent, then there exists a POVM* $\{P_n\}_{n=1}^N \cup \{P_?\}$ *that achieves unambiguous discrimination and satisfies*

$$\langle\varphi_n|P_n|\varphi_n\rangle = p \qquad\qquad \forall n \in \{1, \ldots, N\} .$$

The maximum value that the probability p *can achieve is equal to the minimum eigenvector of the operator* $\Phi := \sum_n |\varphi_n\rangle\langle\varphi_n|$*, called the* frame operator.

Proof. For simplicity, let us suppose that the linearly states $\{|\varphi_n\rangle\}$ are a basis for the Hilbert space (if they are not, we can always restrict without loss of generality to the subspace spanned by them). Since the states $\{|\varphi_m\rangle\}$ are linearly independent, the frame operator $\Phi = \sum_m |\varphi_m\rangle\langle\varphi_m|$ is invertible. Let us define the reciprocal vectors

$$|\psi_m\rangle := \Phi^{-1}|\varphi_m\rangle .$$

By definition, we have $\sum_m |\varphi_m\rangle\langle\psi_m| = I$ and, therefore

$$|\varphi_n\rangle = \sum_m \langle\psi_m|\varphi_n\rangle|\varphi_m\rangle .$$

Now, since the states $\{|\varphi_m\rangle\}$ are linearly independent, the last equation implies the relation

$$\langle\psi_m|\varphi_n\rangle = \delta_{mn} .$$

Hence, the reciprocal vectors $\{|\psi_m\rangle\}$ are exactly the vectors that we need to define a POVM for unambiguous discrimination. We can now define the POVM operators $P_n = p|\varphi_n\rangle\langle\varphi_n|$ and find the maximum probability p that is compatible with the normalization of the POVM. Precisely, the normalization of the POVM reads

$$\sum_{n=1}^{N} P_n + P_? = I\,,$$

or, equivalently,

$$\sum_{n=1}^{N} P_n \leq I\,.$$

Substituting the definition, we have

$$\begin{aligned}\sum_n P_n &= \sum_n p\,\Phi^{-1}|\varphi_n\rangle\langle\varphi_n|\Phi^{-1} \\ &= p\,\Phi^{-1}\Phi\Phi^{-1} \\ &= p\,\Phi^{-1}\,.\end{aligned}$$

Hence, the normalization of the POVM gives the constraint $p\,\Phi^{-1} \leq I$, or, equivalently, $p\,I \leq \Phi$. The maximum probability p compatible with this condition is the minimum eigenvalue of Φ. ■

It is interesting to see what happens in the particular case where $N = 2$. In this case, it is easy to compute the eigenvalues of the frame operator. Indeed, we can write without loss of generality

$$|\varphi_1\rangle = \sqrt{F}|\varphi_0\rangle + \sqrt{1-F}|\varphi_0^{\perp}\rangle$$

and we can use the basis $\{|\varphi_0\rangle, |\varphi_0^{\perp}\rangle\}$ to write the density matrices of the states $|\varphi_0\rangle$ and $|\varphi_1\rangle$ as

$$|\varphi_0\rangle\langle\varphi_0| = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \qquad |\varphi_1\rangle\langle\varphi_1| = \begin{pmatrix} F & \sqrt{F(1-F)} \\ \sqrt{F(1-F)} & 1-F \end{pmatrix}\,,$$

so that

$$\Phi = \begin{pmatrix} 1+F & \sqrt{F(1-F)} \\ \sqrt{F(1-F)} & 1-F \end{pmatrix}\,.$$

Now, the sum of the eigenvalues of Φ is $\mathrm{Tr}[\Phi] = 1$ and the product is $\det[\Phi] = 1 - F$. Clearly, this means that the eigenvalues are $1 - \sqrt{F}$ and $1 + \sqrt{F}$. Hence, we have that the probability of the outcome ? is

$$p_? = \sqrt{F}\,.$$

It is interesting to compare the probability of the inconclusive outcome ? with the probability of error in the minimum error discrimination. In that case, the probability of error was given by

$$p_{err}^{\min} = \frac{1 - \sqrt{1 - 4\pi_0\pi_1 F}}{2}\,,$$

which, for small F, scaled like

$$p_{err}^{\min} \approx \pi_0 \pi_1 F \, .$$

When N identical copies are available, this means that the probability of the inconclusive outcome in the unambiguous decoder goes to zero much slower than the probability of error in the minimum error decoder. Somehow, allowing for a non-zero probability of error enables a faster improvement of discrimination performances.

7.10 Chapter summary

In this chapter we explored the task of distinguishing states. Starting from a simple game of discrimination between two states, we discovered that the measurement that minimizes the probability of error is given by Helstrom's minimum error decoder. The minimum error probability is a simple function of the *trace distance*, an important quantity that measures the distinguishability of quantum states. We then saw that the trace distance between two pure states can be computed in terms of the *fidelity* and we extended the relation between fidelity and trace norm to mixed states. Finally, we analyzed another type of measurement for distinguishing quantum states: the unambiguous state discrimination measurement, which never makes error, but sometimes refuses to give an answer. For pure states, we showed that unambiguous state discrimination is possible if and only if the states are linearly independent. In this case, we found out exactly what is the measurement that achieves unambiguous state discrimination with equal probabilities. Finally, we compared minimum error discrimination and unambiguous state discrimination for two pure states, showing that the minimum error probability goes to zero quadratically faster than the probability of the inconclusive result ? in the limit of many copies.

Appendix: Proof of the lower bound on the trace norm

The proof of the lower bound on the trace norm uses an expression for the fidelity that is proven in Nielsen-Chuang's book. This expression is actually very interesting, because it says that the fidelity between two quantum states is equal to the minimum of the Bhattacharya coefficient over all possible probability distributions resulting from measurements on these two states: precisely, one has

$$F(\rho_0,\rho_1) = \min_N \min_{\substack{\text{all POVMs} \\ \{P_n\}_{n=1}^N}} \left(\sum_n \sqrt{\mathrm{Tr}[P_n\rho_0]\,\mathrm{Tr}[P_n\rho_1]} \right)^2 . \tag{7.9}$$

The proof of this fact can be found at page 412 of Nielsen-Chuang's book. Let us use this fact to prove the desired result:

Proof of the lower bound on the trace norm.

Let $\{P_n\}_{n=1}^N$ be an arbitrary POVM and let $\{p_n\}$ and $\{q_n\}$ be the probability distributions

$$\begin{aligned} p_n &:= \mathrm{Tr}[P_n\rho_0] \\ q_n &:= \mathrm{Tr}[P_n\rho_1] \,. \end{aligned}$$

By the monotonicity of the trace norm, we have

$$\|\pi_0\rho_0 - \pi_1\rho_1\|_1 \geq \sum_{n=1}^N |\pi_0 p_n - \pi_1 q_n| \,.$$

On the other hand,

$$\begin{aligned} \sum_{n=1}^N |\pi_0 p_n - \pi_1 q_n| &= \sum_{n=1}^N |\sqrt{\pi_0 p_n} - \sqrt{\pi_1 q_n}|\ |\sqrt{\pi_0 p_n} + \sqrt{\pi_1 q_n}| \\ &\geqslant \sum_{n=1}^N |\sqrt{\pi_0 p_n} - \sqrt{\pi_1 q_n}|^2 \\ &= \left(1 - 2 \sum_{n=1}^N \sqrt{\pi_0\pi_1 p_n q_n} \right) \\ &\geq \max_{\substack{\text{all POVMs} \\ \{P_n\}_{n=1}^N}} \left(1 - 2 \sum_{n=1}^N \sqrt{\pi_0\pi_1 p_n q_n} \right) \\ &= 1 - 2\sqrt{\pi_0\pi_1 F(\rho_0,\rho_1)} \,, \end{aligned}$$

where the last equality uses Eq. (7.9). Summarizing, we have proven the bound $\|\pi_0\rho - \pi_1\sigma\|_1 \geqslant 1 - \sqrt{4\pi_0\pi_1 F(\rho_0,\rho_1)}$. ■

Chapter 8

Quantum Channel Discrimination and Programming

In the last chapter we saw how to distinguish between two quantum states. Precisely, we considered the situation where Alice prepares a quantum system in a state ρ_x with probability π_x and asks Bob to identify the value of x, giving him a score $+1$ if he succeeds and -1 if he fails. When x can have only two possible values $x = 0, 1$, the average payoff that Bob can get in this game is

$$\omega_{\max} = \|\pi_0\rho_0 - \pi_1\rho_1\|_1 ,$$

where $\|A\|_1$ is the trace norm of an operator A. The maximum payoff is achieved if Bob uses the Helstrom's measurement.

These results can be immediately adapted to the discrimination of two quantum channels, instead of two quantum states. In this chapter we will see how to do this and we will discover some surprising consequences. The discrimination of quantum channels will lead us to prove an important result called the no-programming theorem, which restricts our ability to program a quantum computer. Finally, we will see how an arbitrary computation can be performed with arbitrary accuracy without violating the no-programming theorem.

8.1 Distinguishing between two quantum channels: the advantage of entanglement

Consider the following game: Alice gives Bob a **black box** that transforms an input system A into an output system A'. She promises to Bob that the black box implements a channel $\mathscr{C}_x$, where the index x can assume the values $x = 0, 1$ with probabilities π_0, π_1, respectively. Then, Bob is asked to guess if the channel

is $\mathscr{C}_0$ or $\mathscr{C}_1$, using the black box only one time. If Bob guesses correctly, Alice gives him one coin, otherwise she takes one coin from him. For example, Alice's black box could be a piece of glass that rotates the polarization of a photon by an unknown angle θ_0 or θ_1, and Bob's task could be to discover the angle.

How can Bob identify the unknown channel $\mathscr{C}_x$? The simplest strategy for him is to apply $\mathscr{C}_x$ to some input state ρ and to decode the value of x from the output state $\rho_x := \mathscr{C}_x(\rho)$. By Helstrom's theorem, the best quantum measurement that Bob can use will give him the average payoff $\|\pi_0\mathscr{C}_0(\rho) - \pi_1\mathscr{C}_1(\rho)\|_1$. By the triangular inequality, it is also clear that Bob can choose without loss of generality a pure state $\rho = |\alpha\rangle\langle\alpha|$. Hence, choosing the best input state, Bob will obtain the payoff

$$\omega_{\max} = \max_{|\alpha\rangle\in\mathcal{H}_A} \|\pi_0\mathscr{C}_0(|\alpha\rangle\langle\alpha|) - \pi_1\mathscr{C}_1(|\alpha\rangle\langle\alpha|)\|_1 \ .$$

For example, consider the case where the input system A is the polarization of a photon and the two channels $\mathscr{C}_0$ and $\mathscr{C}_1$ are given by

$$\begin{aligned}\mathscr{C}_0(\rho) &= \rho \\ \mathscr{C}_1(\rho) &= \frac{\sigma_x\rho\sigma_x + \sigma_y\rho\sigma_y + \sigma_z\rho\sigma_z}{3} \ .\end{aligned}$$

In this case, Bob has to discover whether the polarization vector is left untouched or whether instead it is rotated by π around one axis chosen at random among the three axes x, y, z. In the case where $\pi_0 = \pi_1 = 1/2$, it is easy to see that the maximum payoff that Bob can get by sending a photon through the black box is equal to $2/3$. If you don't believe this, try the following

Exercise 24 Prove the relation

$$\rho + \sigma_x\rho\sigma_x + \sigma_y\rho\sigma_y + \sigma_z\rho\sigma_z = 2I \ \mathrm{Tr}[\rho] \, ,$$

valid for every two-by-two matrix ρ. Use this relation to show that, in the above example, $\|\mathscr{C}_0(|\alpha\rangle\langle\alpha|) - \mathscr{C}_1(|\alpha\rangle\langle\alpha|)\|_1 = \frac{4}{3}$ for every pure state $|\alpha\rangle$.

However, applying the unknown channel $\mathscr{C}_x$ to an input state of system A is not the most clever strategy that Bob can adopt. As you know very well by now, using two quantum systems A and B in an entangled state is often helpful. What if Bob prepares two systems A and B, instead of just preparing system A?

In our example of the polarization, this strategy gives a striking result: Bob can distinguish between $\mathscr{C}_0$ and $\mathscr{C}_1$ *with certainty*. Indeed, suppose that Bob prepares two photons in the Bell state $|\Phi^+\rangle = \frac{|I\rangle\rangle}{\sqrt{2}}$ and applies the unknown channel $\mathscr{C}_x$ on system A. In this way, he obtains either the state

$$(\mathscr{C}_0 \otimes \mathscr{I}_B)(\rho) = \frac{|I\rangle\rangle\langle\langle I|}{2}$$

or the state

$$(\mathscr{C}_1 \otimes \mathscr{I}_B)(\rho) = \frac{1}{3}\left(\frac{|\sigma_x\rangle\!\rangle\langle\!\langle\sigma_x|}{2} + \frac{|\sigma_y\rangle\!\rangle\langle\!\langle\sigma_y|}{2} + \frac{|\sigma_z\rangle\!\rangle\langle\!\langle\sigma_z|}{2}\right).$$

Using a measurement on the Bell basis, one can distinguish between ρ_0 and ρ_1 without any error: with the assistance of entanglement, his payoff can be guaranteed to be 1!

This easy example demonstrates that, once again, entanglement allows us to increase our payoff in a game. In general, the best strategy to distinguish between two quantum channels is to

1. choose an auxiliary system B
2. prepare a pure state $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$
3. apply the channel $\mathscr{C}_x$ on system A
4. perform the Helstrom measurement on systems A and B together.

Optimizing over the choice of the auxiliary system and over all possible input states, Bob can achieve the payoff

$$\omega_{\max}^{ent} = \max_{\mathcal{H}_B} \max_{\substack{|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B \\ \||\Psi\rangle\| = 1}} \|\pi_0(\mathscr{C}_0 \otimes \mathscr{I}_B)(|\Psi\rangle\langle\Psi|) - \pi_1(\mathscr{C}_1 \otimes \mathscr{I}_B)(|\Psi\rangle\langle\Psi|)\|_1 \tag{8.1}$$

This expression gives the maximum payoff that can be obtained using the most general kind of quantum strategy.

Note that one can always choose $\mathcal{H}_B = \mathcal{H}_A$ without loss of generality:

Exercise 25 Show that for every Hilbert space $\mathcal{H}_B$ and for every unit vector $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ there exists a unit vector $|\Psi'\rangle \in \mathcal{H}_A \otimes \mathcal{H}_A$ and an isometry $V : \mathcal{H}_A \to \mathcal{H}_B$ such that $|\Psi\rangle = (I_A \otimes V)|\Psi'\rangle$. Use this fact to prove that

$$\omega_{\max}^{ent} = \max_{\substack{|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_A \\ \||\Psi\rangle\| = 1}} \|\pi_0(\mathscr{C}_0 \otimes \mathscr{I}_A)(|\Psi\rangle\langle\Psi|) - \pi_1(\mathscr{C}_1 \otimes \mathscr{I}_A)(|\Psi\rangle\langle\Psi|)\|_1$$

Remark 2 (The diamond norm) The payoff of Eq. (8.1) suggest us to define a norm in the real vector space generated by quantum channels: for a real linear combination of quantum channels Δ, we can define the *diamond norm* of Δ as

$$\|\Delta\|_\diamond := \max_{\substack{|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_A \\ \||\Psi\rangle\| = 1}} \|(\Delta \otimes \mathscr{I}_A)(|\Psi\rangle\langle\Psi|)\|_1 .$$

It is easy to see that $\|\cdot\|_\diamond$ satisfies all the properties of a norm. Using the definition of diamond norm, we can rewrite the maximum payoff in the channel discrimination game as

$$\omega_{max}^{ent} = \|\pi_0 \mathscr{C}_0 - \pi_1 \mathscr{C}_1\|_\diamond \, .$$

In other words, the diamond norm quantifies how well we can distinguish between two quantum channels using the best entangled state for the input and the best joint measurement on the output.

8.2 Distinguishing between two unitary gates

Let us consider a special case of channel discrimination, where the two channels are two unitary gates U_0 and U_1. In this case the output states are pure and we can evaluate the trace norm using the expression

$$\|\pi_0 |\varphi_0\rangle\langle\varphi_0| - \pi_1 |\varphi_1\rangle\langle\varphi_1|\|_1 = \sqrt{1 - 4\pi_0\pi_1 F} \qquad F = |\langle\varphi_0|\varphi_1\rangle|^2 \, .$$

For simplicity, let us consider first the case where Bob does not use an auxiliary system, and he just apply the unknown gate U_x to some input state $|\alpha\rangle \in \mathcal{H}_A$, thus obtaining the output state $|\alpha_x\rangle := U_x|\alpha\rangle$. In this case, his payoff will be

$$\begin{aligned} \omega &= \|\pi_0 \, |\alpha_0\rangle\langle\alpha_0| - \pi_1 \, |\alpha_1\rangle\langle\alpha_1|\|_1 \\ &= \sqrt{1 - 4\pi_0\pi_1 F} \qquad\qquad F = \Big|\langle\alpha|U_0^\dagger U_1|\alpha\rangle\Big|^2 \, . \end{aligned}$$

Can we find the maximum payoff? By definition, the best input state $|\alpha\rangle$ is the one that minimizes the fidelity. To find the minimum, we can expand $|\alpha\rangle$ on a basis of eigenvectors of $U_0^\dagger U_1$ as

$$|\alpha\rangle = \sum_{m=1}^{d_A} \sqrt{p_m} |\alpha_m\rangle \, ,$$

where $\{p_m\}$ are probabilities and each $|\alpha_m\rangle$ is an eigenvector of $U_0^\dagger U_1$. Recall that the eigenvalues of a unitary matrix are complex numbers with unit modulus and can be written as $e^{i\theta_m}$ where each $\theta_m \in [0, 2\pi)$ is an angle. Using this fact, we can write the fidelity as

$$F = \left| \sum_m p_m \, e^{i\theta_m} \right|^2 \, . \tag{8.2}$$

In the complex plane, the number $z = \sum_m p_m \, e^{i\theta_m}$ is just an arbitrary point inside the polygon with vertices $\{e^{i\theta_m}\}$. Hence, the minimum fidelity is nothing but the square of the distance between this polygon and the origin of the complex plane! In other words, we proved that the maximum payoff that Bob can win preparing an input state $|\alpha\rangle \in \mathcal{H}_A$ is

$$\omega_{\max} = \sqrt{1 - 4\pi_0\pi_1 [d(\mathbf{P})]^2} \, , \tag{8.3}$$

where $d(\mathbf{P})$ is the distance between the origin of the complex plane and the polygon $\mathbf{P}$ whose vertices are the eigenvalues of $U_0^\dagger U_1$. In particular, when the polygon contains the origin, the minimum fidelity is zero: this means that the output states $|\alpha_0\rangle$ and $|\alpha_1\rangle$ are orthogonal, and Bob can identify the black box without errors, winning the game with probability 1.

Eq. (8.3) has some surprising consequences:

1. **Entanglement is useless in the discrimination of two unitary gates.** Suppose that, instead of preparing the state $|\alpha\rangle \in \mathcal{H}_A$, Bob prepares an entangled state $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ and applies the unknown unitary U_x on system A, thus generating the output state

$$|\Psi_x\rangle := (U_x \otimes I_B)|\Psi\rangle\,.$$

   Clearly, the best input state $|\Psi\rangle$ is the state that minimizes the fidelity $F = |\langle\Psi|U_0^\dagger U_1 \otimes I_B|\Psi\rangle|^2$ and the minimum of the fidelity is equal to the distance between the origin and the polygon $\mathbf{P}$ whose vertices are the eigenvalues of $U_0^\dagger U_1 \otimes I_B$. But the eigenvalues of $U_0^\dagger U_1 \otimes I_B$ coincide with the eigenvalues of $U_0^\dagger U_1$. Hence, the payoff remains the same even if we introduce the auxiliary system! In mathematical terms, we just showed that

$$\|\mathscr{U}_0 - \mathscr{U}_1\|_\diamond = 2\sqrt{1 - d^2(\mathbf{P})}\,,$$

   where $\mathscr{U}_x$ is the unitary channel defined by $\mathscr{U}_x(\rho) = U_x \rho U_x^\dagger$.

2. **Bob can always win the game using the black box a finite number of times.** Suppose that Alice allows Bob to use the black box N times. In this case, he can prepare N input systems in a joint entangled state $|\Psi\rangle \in \mathcal{H}_A^{\otimes N}$, apply the unknown gate U_x to each system, and then perform a measurement on the output state $|\Psi_x\rangle = U_x^{\otimes N}|\Psi\rangle$. Now, the eigenvalues of $(U_0^\dagger U_1)^{\otimes N}$ are of the form

$$e^{i(\theta_{m_1} + \theta_{m_2} + \cdots + \theta_{m_N})}\,,$$

   where $\{e^{i\theta_m}\}$ are the eigenvalues of $U_0^\dagger U_1$. Hence, the polygon of the eigenvalues becomes larger as N increases. (except in the trivial case where $U_0^\dagger U_1$ is proportional to the identity, where the two unitaries U_0 and U_1 correspond to the same quantum channel). This means that there exists a finite number N_* such that the polygon contains the origin. Using the black box N_* times, Bob can find an input state $|\Psi\rangle$ such that the output states $|\Psi_0\rangle$ and $|\Psi_1\rangle$ are orthogonal.

   Remark 3 (Distinguishing gates vs distinguishing states) Note the difference between the discrimination of unitary gates and the discrimination of pure states. Two gates U_0 and U_1 may not be distinguishable with one use of the black box, but become perfectly distinguishable when N_* uses are allowed. Instead, two pure states $|\varphi_0\rangle$ and $|\varphi_1\rangle$ that are not orthogonal, will never become orthogonal if we take N copies of them: they will become "more and more orthogonal" (fidelity going to zero exponentially fast) but never *exactly* orthogonal.

8.3 Distinguishing between more than two unitary gates

We just saw that an unknown gate U_x with $x = 0, 1$ can be identified without error using the black box a finite number of times. This result can be extended to the case where there are more than two possibilities: suppose that the black box performs one of the gates $\{U_x\}_{x=1}^N$. In this case, if Bob is allowed to use the black box a large (but finite) number of times, he can identify the unknown gate without errors in the following way:

1. Set up a test to distinguish without error between U_1 and U_2 (this can be done using the black box a finite number of times)

2. if the the outcome of the first test is $x_1 \in \{1, 2\}$, then set up a test to distinguish without error between U_{x_1} and U_3

3. if the outcome of the second test is $x_2 \in \{x_1, 3\}$, then set up a test to distinguish without error between U_{x_2} and U_4

4. iterate the procedure $N - 1$ times

5. if the outcome of the $(N-1)$-th test is x_{N-1}, declare that the gate is U_x, with $x = x_{N-1}$

The above protocol eliminates one wrong alternative at each step, so that in the end only the correct alternative remains.

Since the protocols consists in a finite number of steps, and each step uses the black box only a finite number of times, we proved the following:

Property 12 For every finite set of gates $\{U_x\}_{x=1}^N$ there is a finite number K such that an unknown gate in the set can be identified without error using the black box K times.

In the next section we will use this result to prove a surprising result, which is somehow similar to the no-cloning theorem.

8.4 Programming quantum gates

Suppose that you want to construct a small quantum computer, which performs unitary gates on a quantum system A. You want your computer to act as a programmable machine, which can perform one of the gates $\{U_n\}_{n=1}^N$ depending on your instructions. Without loss of generality, we can assume that your instructions are encoded in a pure state $|\beta_n\rangle$ of some quantum system B, called *the program*. The machine should be able to read the instructions contained in the state $|\beta_n\rangle$ and to perform the desired gate U_n.

One example of programmable machine is described by a control-unitary gate

$$V = \sum_{n=1}^{N} U_n \otimes |n\rangle\langle n| \,. \tag{8.4}$$

If you want this machine to perform the gate U_n you have only to encode your instructions in the state $|n\rangle$: for every state $|\alpha\rangle \in \mathcal{H}_A$, the machine will implement the desired gate U_n

$$V|\alpha\rangle|n\rangle = U_n|\alpha\rangle|n\rangle \,.$$

Now, the machine here needs a program B of dimension N equal to the number of gates. But can we use a smaller program of dimension $d_B < N$? And can we construct a machine that can perform *every* unitary gate on system A using only a finite dimensional program B?

Surprisingly, the answer is *no*. The proof of this fact is known as the *no-programming theorem* (Nielsen-Chuang, PRL 1997):

Theorem 10 (No-programming) *In order to program N distinct unitary gates, one needs N orthogonal program states.*

The no-programming theorem tells us that, in terms of dimension of the program, the control-unitary gate V in Eq. (8.4) is already the best possible programmable machine that executes the gates $\{U_n\}_{n=1}^N$.

Let us see the proof of the theorem:

Proof. Suppose that there is a programmable machine using the program states $\{|\beta_n\rangle\}$. Let us represent the joint evolution of the input data A, the program B, and the machine M as a unitary gate W and let us denote by $|0\rangle \in \mathcal{H}_M$ the initial state of the machine. By definition, we must have

$$W|\alpha\rangle|\beta_n\rangle|0\rangle = U_n|\alpha\rangle|\Gamma_n\rangle \qquad \forall|\alpha\rangle \in \mathcal{H}_A \,, \tag{8.5}$$

where $|\Gamma_n\rangle \in \mathcal{H}_B \otimes \mathcal{H}_M$ is the final state of the program and the machine [1].

On the other hand, applying $W^\dagger$ on both sides of the Eq. (8.5), we obtain

$$|\alpha\rangle|\beta_n\rangle|0\rangle = W^\dagger U_n|\alpha\rangle|\Gamma_n\rangle \qquad \forall|\alpha\rangle \in \mathcal{H}_A \,,$$

and, defining $|\alpha'\rangle := U_n|\alpha\rangle$,

$$U_n^\dagger|\alpha'\rangle|\beta_n\rangle|0\rangle = W^\dagger|\alpha'\rangle|\Gamma_n\rangle \qquad \forall|\alpha'\rangle \in \mathcal{H}_A \,. \tag{8.6}$$

[1] A priori, one may ask if $|\Gamma_n\rangle$ can depend also on $|\alpha\rangle$. In fact, it is easy to see that this is not the case: the condition

$$W|\alpha\rangle|\beta_n\rangle|0\rangle = U_n|\alpha\rangle|\Gamma_{\alpha,n}\rangle \qquad \forall|\alpha\rangle \in \mathcal{H}_A$$

implies

$$\langle\alpha|\alpha'\rangle = \langle\alpha|\alpha'\rangle\langle\Gamma_{\alpha,n}|\Gamma_{\alpha',n}\rangle \qquad \forall|\alpha\rangle,|\alpha'\rangle \in \mathcal{H}_A \,.$$

For $\langle\alpha|\alpha'\rangle \neq 0$, this requires $|\Gamma_{\alpha,n}\rangle = |\Gamma_{\alpha',n}\rangle$. Since the original condition is required to hold *for every* $|\alpha\rangle$, the state $|\Gamma_{\alpha,n}\rangle$ must be independent of $|\alpha\rangle$.

This means that the states $\{|\Gamma_n\rangle\}$ can be used to program the gates $\{U_n^\dagger\}$ using the inverse evolution $W^\dagger$ and that $W^\dagger$ resets the program and the machine to the state $|\beta_n\rangle|0\rangle$.

Using Eqs. (8.5) and (8.6) we can program the gate $U_n \otimes U_n^\dagger$ and apply it as many times as we want. Indeed, suppose that we want to apply the gate $(U_n \otimes U_n^\dagger)^{\otimes K}$ to the state $|\Psi\rangle$ of $2K$ copies of A. We can do it in the following way:

1. start with BM in the state $|\beta_n\rangle|0\rangle$

2. apply the gate W to A_1BM, where A_1 is the first copy (the result of this step is to apply U_n on system A_1 and to leave BM in the state $|\Gamma_n\rangle$)

3. apply the gate $W^\dagger$ to A_2BM (the result of this step is to apply $U_n^\dagger$ on system A_2 and to leave BM in the state $|\beta_n\rangle|0\rangle$)

4. apply the gate W to A_3BM, the gate $W^\dagger$ to A_4BM and continue until all the $2K$ systems have been used.

As a result of this protocol, the systems $A_1A_2\dots A_{2K}$ are in the state $|\Psi_n\rangle = (U_n \otimes U_n^\dagger)|\Psi\rangle$. Now, choosing K sufficiently large, we know that there is an input state $|\Psi\rangle$ such that the states $\{|\Psi_n\rangle\}_{n=1}^N$ are orthogonal. Since the states $\{|\Psi_n\rangle\}$ are generated from the program states $\{|\beta_n\rangle\}$, also the states $\{|\beta_n\rangle\}$ must be orthogonal. ■

The no-programming theorem highlights a big difference between classical and quantum information theory. In classical information, the only reversible gates that we can apply to a classical system with d_A distinguishable states are permutations, which transform the input state m into the output state $\pi(m)$. For example, the only reversible gates for a single bit are the identity and the bit flip, implementing the permutation $\pi(0) = 1$ and $\pi(1) = 0$. Since there is a finite number of permutations, a classical gate can be programmed perfectly with a finite program of size $d_B = d_A!$, the number of permutations on a set of d_A elements. Instead, in quantum information even a two-level system has an infinite set of reversible gates. This means that, if we want to program an arbitrary qubit gate without error we have to use an infinite program!

Of course, this cannot be done in a computer: the computations performed by a computer must be performed using only a finite set of basic gates. Otherwise, how could we write down a program for this computation?

8.5 Universal sets of quantum gates

The no-programming theorem seems to put quantum computation into trouble. A finite quantum system has an infinite set of gates, which cannot be controlled by a finite program. How can we hope to perform arbitrary quantum computations?

Luckily, one can show that *every* quantum gate U can be approximated with arbitrary accuracy using only a finite set of gates. To see that, let us consider first a simple example for qubits: suppose that you constructed a device that implements the gate

$$U_{\varphi,z} = \exp \frac{i\varphi\sigma_z}{2} \,,$$

where $\varphi \in [0, 2\pi)$ is some angle such that $\varphi/(2\pi)$ is irrational. Clearly, every rotation around the z axis can be approximated with arbitrary precision by applying $U_{\varphi,z}$ many times. Indeed, since φ is not rational to 2π, there is no number N such that $N\varphi = 0 \mod 2\pi$. This means that the multiples of φ are dense in the circle: for every angle θ and for every desired $\epsilon > 0$ we can find some number N such that $|\theta - N\varphi| \mod 2\pi \leq \epsilon$. Clearly, this means that every rotation of the form $U_{\theta,z} = \exp \frac{i\theta\sigma_z}{2}$ can be approximated arbitrarily well be by $U^N_{\varphi,z}$, provided that we choose the right value of N. Measuring the quality of the approximation in terms of the diamond norm, we have the following statement:

Proposition 5 *For every angle $\theta \in [0, 2\pi)$ and for every $\epsilon > 0$ there is an integer N such that*

$$\left\| \mathscr{U}_{\theta,z} - \mathscr{U}^N_{\varphi,z} \right\|_\diamond \leq \epsilon \,,$$

where $\mathscr{U}_\theta$ is the quantum channel defined by $\mathscr{U}_\theta(\rho) = U_{\theta,z}\rho U^\dagger_{\theta,z}$.

We have just seen how to approximate all rotations around the z axis by repeatedly applying a single rotation. Clearly, the same idea can be used to approximate rotations around arbitrary axes. Indeed, if you are able to construct a gate implementing the rotation

$$U_{\psi,x} = \exp \frac{i\psi\sigma_x}{2} \,,$$

where $\psi/(2\pi)$ is irrational, then you can approximate arbitrary rotations around the x axis. Then, combining rotations around x with rotations around z you can generate every rotation. In conclusion, we have proved the following

Theorem 11 *Every qubit gate can be approximated with arbitrary precision with a circuit consisting only of 2 elementary gates.*

Of course, there are many different pairs of qubit gates that can be used to approximate arbitrary qubit gates. The most popular pair used by quantum computer scientists is the pair $\{H, T\}$, where H is the Hadamard gate

$$\begin{aligned} H &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \\ &= i \exp \frac{-i\pi \mathbf{n} \cdot \boldsymbol{\sigma}}{2} \qquad \mathbf{n} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \end{aligned}$$

and T is a rotation of $\pi/4$ around z [2]:

$$T = \exp \frac{-i\pi\sigma_z}{8} .$$

If you are surprised because the two gates H and T are both rotations of angles that are rational to π, try to compute HT: you will note that it is a rotation of an angle φ such that $\varphi/(2\pi)$ is not rational.

For a quantum system of dimension $d \geq 2$, one can obtain a similar result: there is always a finite set of gates $\mathbf{U}$ such that every unitary gate U can be approximated arbitrarily well by a sequence $U' = U_1 U_2 \cdots U_N$ where each gate belongs to $\mathbf{U}$. A finite set of gates with this property is called a *universal set of gates in dimension d*.

You may wonder how many gates are needed to approximate every unitary gate in dimension d. Is it order of d, order of d^2, or anything else? Probably the answer will surprise you: if is enough to use order of $(\log d)^2$ elementary gates!

The way to see it is a powerful result about quantum systems consisting of N qubits:

Theorem 12 *Let* $\mathbf{S}$ *be a universal set of single-qubit gates and let* W *be a two-qubit entangling gate, that is, a gate that transforms at least one product state* $|\alpha\rangle|\beta\rangle$ *into an entangled state* $W|\alpha\rangle|\beta\rangle$*. Then, the set* $\mathbf{U}$ *containing all single qubit gates in* $\mathbf{S}$ *(applied to all possible qubits) and all possible two qubit gates* W_{ij}*, where* W_{ij} *denotes the gate* W *acting on the qubits* i *and* j*, is a universal in dimension* 2^N*.*

For example, one can choose the universal set of single-qubit gates to be the pair $\{H, T\}$ and the entangling gate W to be the control-NOT gate

$$\texttt{CNOT} = |0\rangle\langle 0| \otimes I + |1\rangle\langle 1| \otimes \sigma_x .$$

With this choice, we need only $2N$ single qubit gates and $N(N-1)/2$ two qubit gates: using $O(N^2)$ elementary gates is enough to approximate an arbitrary gate on N qubits. The Nielsen-Chuang book provides a (rather long) proof that this fact.

Clearly, we can encode the states of a general quantum system A into the states of $N = \lceil \log_2 d_A \rceil$ qubits. Using this encoding we can reduce the problem of performing a unitary gate on A to the problem of performing a unitary gate on N_A qubits. Using the previous result, we know that since $O(\log_2^2 d_A)$ elementary gates are enough.

Remark 4 Programming quantum channels. Note that, since every quantum channel can be realized as a unitary gate followed by partial trace, we have that programming all possible unitary gates is enough to program all possible quantum channels.

[2] Historically, the T gate was defined as $T = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{pmatrix}$, which coincides with our definition up to an irrelevant phase factor $e^{-i\pi/8}$.

We just saw that every gate in dimension d_A can be approximated using a finite set of $O(\log^2 d_A)$ gates. But how long will be the sequence of gates needed for a good approximation? This is a very important question: the fact that we can approximate every gate would not be very useful if the length of the circuit needed for the approximation were too long!

Luckily, the following theorem gives us a favourable scaling:

Theorem 13 (Solovay-Kitaev) *Let* $\mathbf{U}$ *be a universal set of unitary gates in dimension* d *with the property that for every gate* $U \in \mathbf{U}$ *one has* $U^\dagger \in \mathbf{U}$*. Then, every unitary gate in dimension* d *can be approximated within an error* ϵ *as a product of* N *gates in* $\mathbf{U}$*, where* N *is of order* $O(\log^c \epsilon^{-1})$*, where* c *is a suitable constant between 0 and 2.*

In other words, the Solovay-Kitaev theorem tells us that, whenever a set of gates can approximate all gates, it also does it "quickly", in the sense that taking sequences of gates in the set we quickly get close to every desired gate. Importantly, there are constructive algorithms to find how to decompose a desired gate into a sequence of $O(\log^c \epsilon^{-1})$ elementary gates.

The existence of universal sets of gates and the Solovay-Kitaev theorem ensure that we can *efficiently* program a quantum computation using a finite set of instructions that can be encoded into a set of orthogonal program states. This is great news, but it is important not to confuse great news with miracles: the Solovay-Kitaev theorem guarantees that the number of gates scales well with the approximation parameter ϵ, but it does not guarantees the same for the scaling with the dimension. There is no theorem stating that all unitary gates in dimension d can be approximated using $O(\log^\alpha d)$ for some exponent α. Such a theorem would be a miracle, because it would imply that every unitary gate on N qubits can be implemented with a number of elementary gates scaling polynomially with N. Instead, we know that there are counterexamples of unitary gates that require a number of elementary gates that is exponential in N.

8.6 Chapter summary

Equipped with the results of the previous chapter, today we studied the task of distinguishing quantum channels, highlighting a number of quantum effects. First, we have seen that preparing the input of an unknown channel in an entangled state with an additional system, one can increase the probability of correct identification. Then, we considered the special case of unitary gates, showing that an unknown unitary gate in a finite set can be identified perfectly using the gate a finite number of times. This result lead us to the No Programming Theorem, which tells us that it is impossible to construct a machine that can be programmed to perform an arbitrary gate using as program a finite dimensional system. Finally, we showed that the No Programming theorem does not threaten quantum computation: every desired gate can be *approximated* with

arbitrary accuracy with a finite sequence of elementary gates, which can be programmed using a finite dimensional program. Moreover, the approximation is efficient, thanks to the Solovay-Kitaev theorem, which guarantees that a unitary gate in dimension d can be implemented with accuracy ϵ using a number of elementary gates that is polynomial in $\log \epsilon^{-1}$.

Chapter 9

Quantum Error Correction

In the previous chapters we saw various types of quantum machines: machines that try to copy data, machines that transfer data from one place to another, machine that decode a classical message encoded into quantum states and channels, machines that can be programmed in order to perform a desired computation. All these machines are doing operations on quantum systems. But in general quantum systems are subject to noise, which corrupts their state. How can we protect quantum information from these imperfections? Protecting quantum information is the goal of quantum error correction, which will be studied in this chapter.

9.1 Why quantum error correction is challenging

In real implementations, quantum systems are subject to noise and errors. Doing quantum computation seems an impossible task, because these errors can easily destroy quantum superpositions and entanglement, and eliminate the advantages of quantum information.

Errors and noise also exist in the classical world, of course, but in the classical world correcting errors seems to be a lot easier. Let us see an example:

Suppose that you want to store a bit x in the memory of your computer and to retrieve it at a later time. However, after this time there is a small probability p that the value of the bit is flipped, i.e. that x becomes $x \oplus 1$. In order to protect the information from this error, there is one simple strategy: make many copies of your data, so that it is likely that the majority of them will remain correct, and, if there is an error, it will affect at most one copy. For example, you can encode the value x into three classical bits in the memory, encoding 0 into the string 000 and 1 into the string 111. Since p is small, with high probability the error will affect only one bit of the string: for example, the

string 000 will become

$$\begin{aligned} 000 \quad &\xrightarrow{(1-p)^3 \simeq 1-3p} \quad 000 \\ &\xrightarrow{\quad p \quad} \quad 100 \\ &\xrightarrow{\quad p \quad} \quad 010 \\ &\xrightarrow{\quad p \quad} \quad 001. \end{aligned}$$

To retrieve the value of the bit, you just have to read the string xyz and check if the majority of digits is "0" or "1". If the majority is "0", you bring back the string xyz to 000; if the majority is "1", you bring back the string to 111. Assuming that p^2 is negligible, using this strategy you will always be able to maintain the memory of the computer in the correct state. And of course, you can always retrieve the value of your initial bit x by measuring the three bits in the memory.

Let us now consider the same problem in the quantum world: you have a qubit in some pure state $|\varphi\rangle = \alpha|0\rangle + \beta|1\rangle$ and you want to store it in the memory of a quantum computer. However, after some time there is a small probability p that a qubit in the memory is rotated by the gate X (Pauli matrix σ_x), which is the quantum analogue of the bit flip. When this happens, the state $|\varphi\rangle = \alpha|0\rangle + \beta|1\rangle$ is transformed into $|\varphi'\rangle = \alpha|1\rangle + \beta|0\rangle$. How can you store your quantum data in a way that is protected by this error?

Of course, by the no-cloning theorem you cannot make three copies of your initial data: this means that it is impossible to encode $|\varphi\rangle$ into $|\varphi\rangle|\varphi\rangle|\varphi\rangle$. On top of that, if you do not know $|\varphi\rangle$ it is not even clear how to check that the majority of the qubits is in the state $|\varphi\rangle$!

Since the obvious approach does not work, you have to invent some other trick. One possibility is to apply classical error correction in the computational basis: you can encode the state $|\varphi\rangle = \alpha|0\rangle + \beta|1\rangle$ into the three-qubit state

$$|\varphi_0\rangle = \alpha|0\rangle|0\rangle|0\rangle + \beta|1\rangle|1\rangle|1\rangle$$

applying the isometry $V := |0\rangle|0\rangle|0\rangle\langle 0| + |1\rangle|1\rangle|1\rangle\langle 1|$, which transforms one-qubit states into three-qubit states. Then, you can store the state $|\varphi_0\rangle$ in the memory of the computer. As a result of the noise in the memory, the state $|\varphi_0\rangle$ will become

$$\begin{aligned} |\varphi_0\rangle \quad &\xrightarrow{(1-p)^3 \simeq 1-3p} \quad |\varphi_0\rangle := \alpha|000\rangle + \beta|111\rangle \\ &\xrightarrow{\quad p \quad} \quad |\varphi_1\rangle := (X \otimes I \otimes I)|\varphi_0\rangle = \alpha|100\rangle + \beta|011\rangle \\ &\xrightarrow{\quad p \quad} \quad |\varphi_2\rangle := (I \otimes X \otimes I)|\varphi_0\rangle = \alpha|010\rangle + \beta|101\rangle \\ &\xrightarrow{\quad p \quad} \quad |\varphi_3\rangle := (I \otimes I \otimes X)|\varphi_0\rangle = \alpha|001\rangle + \beta|110\rangle\,. \end{aligned}$$

This means that, on average, the state of the memory will be the mixed state

$$\rho' = (1-3p)|\varphi_0\rangle\langle\varphi_0| + p\sum_{i=1}^{3}|\varphi_i\rangle\langle\varphi_i|\,.$$

Now, suppose that you would like to take back the state of the memory to the initial state $|\varphi_0\rangle$. Here there is a problem: you cannot measure the three qubits in the computational basis to see if the majority of the qubits is in the state $|0\rangle$ or if it is in the state $|1\rangle$. Indeed, measuring on the computational basis $\{|x\rangle|y\rangle|z\rangle\}$ destroys all quantum superpositions!

However, the good thing is that the four states $|\varphi_0\rangle$, $|\varphi_1\rangle$, $|\varphi_2\rangle$, $|\varphi_3\rangle$ belong to four **orthogonal** subspaces

$$\begin{aligned}
\mathcal{S}_0 &:= \mathsf{Span}\{|0\rangle|0\rangle|0\rangle, |1\rangle|1\rangle|1\rangle\} \\
\mathcal{S}_1 &:= \mathsf{Span}\{|1\rangle|0\rangle|0\rangle, |0\rangle|1\rangle|1\rangle\} \\
\mathcal{S}_2 &:= \mathsf{Span}\{|0\rangle|1\rangle|0\rangle, |1\rangle|0\rangle|1\rangle\} \\
\mathcal{S}_3 &:= \mathsf{Span}\{|0\rangle|0\rangle|1\rangle, |1\rangle|1\rangle|0\rangle\},
\end{aligned}$$

no matter what the coefficients α and β are. This means that you can apply a quantum instrument that reveals in which subspace is the state of the three qubits. Consider the instrument $\{\mathscr{Q}_i\}_{i=0}^3$, with quantum operations defined as $\mathscr{Q}_i(\rho) = P_i \rho P_i$, where P_i is the projector on $\mathcal{S}_i$. Since we have $P_i|\varphi_j\rangle = \delta_{ij}|\varphi_j\rangle$, if you obtain the outcome "i", then the state of the three qubits after the measurement will be

$$\rho_i' = \frac{P_i \rho' P_i}{\mathrm{Tr}[P_i \rho']} \equiv |\varphi_i\rangle\langle\varphi_i|\,.$$

At this point, to correct the error you just need to apply the unitary gate U_i given by

$$\begin{aligned}
U_0 &= I \otimes I \otimes I \\
U_1 &= X \otimes I \otimes I \\
U_2 &= I \otimes X \otimes I \\
U_3 &= I \otimes I \otimes X.
\end{aligned}$$

In this way, the state of the memory has been brought back to $|\varphi_0\rangle = \alpha|0\rangle|0\rangle|0\rangle + \beta|1\rangle|1\rangle|1\rangle$. Of course, if you want to retrieve the state $|\varphi\rangle = \alpha|0\rangle + \beta|1\rangle$, you can now do it by applying the decoding channel

$$\mathscr{D}(\rho) = V^\dagger \rho V + \mathrm{Tr}[(I - VV^\dagger)\rho]\,|0\rangle\langle 0|,$$

where $V = |0\rangle|0\rangle|0\rangle\langle 0| + |1\rangle|1\rangle|1\rangle\langle 1|$ is the isometry that was used to encode the state of the initial qubit in the state of three qubits.

This is the simplest example of quantum error correction. Summarizing, our example consisted of five steps:

1. Encoding: the state of the input system A (one qubit, in our example) is encoded into the state of another system A' (three qubits, in our example) using the channel $\mathscr{V}(\rho) = V\rho V^\dagger$ for a suitable isometry V

2. Error: a quantum channel $\mathscr{E}$, representing the noise, transforms the state of system A' [in our example, this was the channel $\mathscr{E}(\rho) = (1-3p)\rho + p\sum_{i=1}^3 \mathscr{U}_i(\rho)$, where $\mathscr{U}_i(\rho) = U_i \rho U_i^\dagger$]

3. Measurement: system A' is measured with a quantum instrument $\{\mathscr{Q}_i\}$

4. Recovery: if the measurement gives outcome i, then one applies a unitary gate U_i, which brings A' back to the initial state

5. Decoding: system A' is transformed into system A, which is brought back to the initial state.

Putting the last three steps together, we have the *recovery channel* $\mathscr{R}$, defined by

$$\mathscr{R}(\rho) := \sum_i \mathscr{D}\left\{U_i\left[\mathscr{Q}_i(\rho)\right]U_i^\dagger\right\} .$$

The encoding channel $\mathscr{V}(\rho) = V\rho V^\dagger$, the error $\mathscr{E}$ and the recovery channel $\mathscr{R}$ satisfy the relation

$$\mathscr{R}\mathscr{E}\mathscr{V} = \mathscr{I}_A ,$$

which means that after applying the three channels the state of the input system A remains the same.

Our example was a very specific one:

1. bit flips are not the only types of error that can happen in the quantum world. For single qubit errors, in addition to X we have the other Pauli matrices Y, Z, and all possible rotations.

2. in the example, the errors where guaranteed to happen only on one qubit at one time. But in general one could have errors like $X \otimes X \otimes X$, acting on all qubits together.

In the next paragraph, we will consider errors described by arbitrary quantum channels. However, the structure of error correction will be exactly the same that we saw in this simple example.

9.2 When there is hope to correct an error

Suppose that the state of a quantum system A has been encoded into the state of another system A', using some encoding operation $\mathscr{V}$. Suppose that, after the encoding, system A' undergoes an error, described by a quantum channel $\mathscr{E}$. How can we know if this error can be corrected or not?

We will now answer this question. First of all, we can consider the encoding and the error together, defining the channel $\mathscr{C} := \mathscr{E}\mathscr{V}$. Using this definition, the question "Can I correct the error $\mathscr{E}$?" is equivalent to the question "Is it possible to invert the channel $\mathscr{C}$ with another quantum channel $\mathscr{R}$?"

This motivates the following definition:

Definition 8 (Correctability) *A quantum channel from A to A' is* **correctable** *if and only if there exists a* **recovery channel** *$\mathscr{R}$ from A' to A such that $\mathscr{R}\mathscr{C} = \mathscr{I}_A$.*

Of course, not every channel is correctable. For example, the erasure channel, that transforms every state ρ into a fixed state (e.g. $|0\rangle\langle 0|$) cannot be corrected, because after this channel has been applied every information about the input state is gone.

A mathematical condition that can be used to establish whether or not a quantum channel $\mathscr{C}$ is correctable is the following

Theorem 14 (Knill-Laflamme (KL) condition) *Let $\mathscr{C}$ be a channel with Kraus representation $\mathscr{C}(\rho) = \sum_{i=1}^{K} C_i \rho C_i^\dagger$. If channel $\mathscr{C}$ is correctable, then*

$$C_j^\dagger C_i = \sigma_{ij} I_A$$

where σ is the density matrix of a quantum system B of dimension $d_B = K$.

Proof. If $\mathscr{C}$ is correctable, then there exists a recovery channel $\mathscr{R}$ such that $\mathscr{R}\mathscr{C} = \mathscr{I}_A$. Taking Kraus representations for $\mathscr{C}$ and $\mathscr{R}$, we then have

$$\sum_m \sum_i R_m C_i \rho C_i^\dagger R_m^\dagger = \rho \,,$$

for every state ρ. By the no-information without disturbance theorem, we have that each quantum operation $\mathscr{Q}_{mi}$ defined by $\mathscr{Q}_{mi}(\rho) = R_m C_i \rho C_i^\dagger R_m^\dagger$ must be proportional to the identity. Hence, we have

$$R_m C_i = \lambda_{mi} I_A$$

for some coefficient $\lambda_{mi} \in \mathbb{C}$ with the property $\sum_{m,i} |\lambda_{mi}|^2 = 1$. Taking the adjoint, we get

$$C_i^\dagger R_m^\dagger = \bar{\lambda}_{mi} I_A .$$

Hence, using the normalization condition $\sum_m R_m^\dagger R_m = I_{A'}$ we obtain

$$\begin{aligned} C_j^\dagger C_i &= \sum_m C_j^\dagger (R_m^\dagger R_m) C_i \\ &= \left(\sum_m \bar{\lambda}_{mj} \lambda_{mi} \right) I_A \\ &= \sigma_{ij} \, I_A \qquad \sigma_{ij} := \left(\sum_m \bar{\lambda}_{mj} \lambda_{mi} \right) \end{aligned}$$

Clearly, the matrix σ is positive and $\mathrm{Tr}[\sigma] = \sum_{m,i} |\lambda_{mi}|^2 = 1$. Hence σ is a quantum state. ■

In the case where the input and output systems are the same ($A = A'$), the KL condition gives an immediate characterization of the correctable channels:

Corollary 3 *A channel $\mathscr{C}$ transforming system A into itself is correctable if and only if it is unitary (i.e. $\mathscr{C}(\rho) = U \rho U^\dagger$ for some unitary U).*

Proof. For $A = A'$, the condition $C_i^\dagger C_i = \sigma_{ii}\, I_A$ implies that $U_i := C_i/\sqrt{\sigma_{ii}}$ is a unitary gate. On the other hand, the condition $C_j^\dagger C_i \propto I_A$ implies that $U_i \propto \omega_{ij}\, U_j$ for every i, j, with $\omega_{ij} \in \mathbb{C}$ and $|\omega_{ij}| = 1$. Denoting $U_i := \omega_{i1} U$, we then have

$$\begin{aligned}\mathscr{C}(\rho) &= \sum_{i=1}^{k} (\sqrt{\sigma_{ii}} U_i) \rho k (\sqrt{\sigma_{ii}} U_i) \\ &= \mathrm{Tr}[\sigma]\, U \rho U^\dagger \\ &= U \rho U^\dagger \,.\end{aligned}$$

The converse is trivial: if a channel is unitary it is also correctable, with recovery channel $\mathscr{R}(\rho) = U^\dagger \rho U$. ■

9.3 How to correct a channel

We saw that the KL condition is a necessary condition for the correctability of a channel. Is it also sufficient? Luckily, the answer is *yes* and we will now see explicitly how to correct a channel that satisfies the Knill-Laflamme condition. Interestingly, the general correction procedure has the same structure of the correction procedure that we saw in the example at the beginning of the chapter.

In order to prove these facts, it is useful to derive a simple consequence of the KL condition:

Lemma 1 *If a channel $\mathscr{C}$ satisfies the KL condition, then it can be written as*

$$\mathscr{C}(\rho) = \sum_{m=1}^{R} p_m\; V_m \rho V_m^\dagger \,,$$

where $\{V_m\}_{m=1}^R$ are orthogonal isometries*, that is, $V_m^\dagger V_n = \delta_{mn}\, I_A$ for every $m, n \in \{1, \ldots, R\}$.*

Proof. Let us start from the KL condition $C_j^\dagger C_i = \sigma_{ij}\, I_A$ and diagonalize the density matrix σ as

$$\sigma = \sum_{l=1}^{R} p_l\, U|l\rangle\langle l|U^\dagger \,,$$

where U is a suitable $K \times K$ unitary. Then, we have $\sigma_{ij} = \sum_{l=1}^R p_l\, U_{il} U_{lj}^\dagger$. Now, define the operators $\{V_m\}_{m=1}^R$ as

$$V_m := \frac{1}{\sqrt{p_m}} \sum_j U_{mj}^\dagger C_j \,.$$

These operators are orthogonal isometries: indeed, using the KL condition we obtain

$$\begin{aligned} V_m^\dagger V_n &= \frac{1}{\sqrt{p_m p_n}} \sum_{i,j} \overline{U}^\dagger_{mj} \left(C_j^\dagger C_i \right) U^\dagger_{ni} \\ &= \frac{1}{\sqrt{p_m p_n}} \sum_{i,j} U_{jm} \left(\sum_{l=1}^{r} p_l \, U_{il} U^\dagger_{lj} \, I_A \right) U^\dagger_{ni} \\ &= \frac{1}{\sqrt{p_m p_n}} \left[\sum_{i,j,l} p_l \left(U^\dagger_{ni} U_{il} \right) \left(U^\dagger_{lj} U_{jm} \right) \right] \, I_A \\ &= \frac{1}{\sqrt{p_m p_n}} \left[\sum_{l} p_l \, \delta_{nl} \, \delta_{lm} \right] \, I_A \\ &= \delta_{mn} \, I_A \, . \end{aligned}$$

On the other hand, one has

$$\begin{aligned} \sum_{m=1}^{R} p_m \, V_m \rho V_m^\dagger &= \sum_{m,i,j} \left(U^\dagger_{mj} C_j \right) \rho \left(\overline{U}^\dagger_{mi} C_i^\dagger \right) \\ &= \sum_{m,i,j} \left(U_{im} U^\dagger_{mj} \right) \, C_j \rho C_i^\dagger \\ &= \sum_{i} C_i \rho C_i^\dagger \\ &= \mathscr{C}(\rho) \, . \end{aligned}$$

This concludes the proof. ■

The meaning of the above lemma is that every correctable channel can be realized in the following way:

1. with probability p_m, system A is transformed with an isometry V_m, chosen at random in a set of orthogonal isometries
2. average over m.

In other words, every correctable channel is a *random-isometry channel*, with orthogonal isometries.

Now, suppose that system A is originally in a pure state $|\varphi\rangle$. Then, with probability p_m, the state will be transformed into $|\varphi_m\rangle = V_m|\varphi\rangle$, which belongs to the subspace $\mathcal{S}_m := V_m \mathcal{H}_A$. The subspaces $\{\mathcal{S}_m\}_{m=1}^R$ are **orthogonal**, thanks to the the fact that the isometries $\{V_m\}_{m=1}^R$ are orthogonal: indeed, every two vectors $|\varphi_m\rangle$ and $|\psi_n\rangle$ of the form

$$\begin{aligned} |\varphi_m\rangle &= V_m|\varphi\rangle \qquad |\varphi\rangle \in \mathcal{H}_A \\ |\psi_n\rangle &= V_n|\psi\rangle \qquad |\psi\rangle \in \mathcal{H}_A \end{aligned}$$

will be orthogonal for $m \neq n$.

Using this fact, it is immediate to see that the KL condition is also sufficient for correctability. Indeed, to correct the errors we can use the same two-step procedure used in the example at the beginning of the chapter:

1. **Measurement:** Measure system A' with the quantum instrument $\{\mathscr{Q}_m\}_{m=1}^{R+1}$, where the quantum operations are defined as

$$\mathscr{Q}_m(\rho) := P_m \rho P_m$$

   where $P_m = V_m V_m^\dagger$ $(m = 1, \ldots, R)$ is the projector on $\mathscr{S}_m$ and $P_{R+1} = I_{A'} - \sum_{m=1}^{R} P_m$.

2. **Recovery:** if the outcome is "m", then apply the recovery channel $\mathscr{R}_m$ that corrects the isometry V_m, namely

$$\mathscr{R}_m(\rho) = V_m^\dagger \rho V_m + \mathrm{Tr}[(I_{A'} - V_m V_m^\dagger)\rho]\, |0\rangle\langle 0| \,.$$

The functionality of the protocol is clear: after the measurement is performed, if the outcome is m the state will be

$$\begin{aligned} \rho'_m &:= \frac{\mathscr{Q}_m(\mathscr{C}(\rho))}{\mathrm{Tr}[\mathscr{Q}_m(\mathscr{C}(\rho))]} \\ &= V_m \rho V_m^\dagger \,. \end{aligned}$$

Then, the recovery channel $\mathscr{R}_m$ will undo the isometry V_m, transforming the state ρ'_m into the original state ρ. Summing over all possible outcomes, our protocol defines a recovery channel

$$\mathscr{R} := \sum_m \mathscr{R}_m \mathscr{Q}_m \,,$$

which satisfies the desired condition

$$\mathscr{R}\mathscr{C} = \mathscr{I}_A \,.$$

In summary, we have discovered the following facts: the only channels that can be corrected are those that encode randomly the state of system A in different **orthogonal subspaces**. The correction consists in discovering in which subspace the information has been encoded and in bringing it back to the original system. Collecting together these results, we have the following

Theorem 15 *Let $\mathscr{C}$ be a channel from A to A', with Kraus representation $\mathscr{C}(\rho) = \sum_{i=1}^{K} C_i \rho C_i^\dagger$. Then, the following are equivalent:*

1. *$\mathscr{C}$ is correctable*
2. *$C_j^\dagger C_i = \sigma_{ij}\, I_A$ for some state σ of a K-dimensional quantum system*
3. *there exist a set of orthogonal isometries $\{V_m\}_{m=1}^{R}$ and a set of probabilities $\{p_m\}_{m=1}^{R}$ such that $\mathscr{C}(\rho) = \sum_{m=1}^{R} p_m\, V_m \rho V_m^\dagger$ for every state ρ.*

9.4 Quantum packing bounds

A correctable channel has to encode the input state into orthogonal subspaces. This implies that the dimension of the output space A' must be larger than a multiple of the dimension of the input space A: only in this way it is possible to pack the orthogonal d_A-dimensional subspaces $\mathcal{S}_m = V_m \mathcal{H}_A$ into $\mathcal{H}_{A'}$. Precisely, we have the following *packing bound*

$$d_{A'} \geq d_A \, R \qquad\qquad R := \mathsf{rank}(\sigma) \, . \tag{9.1}$$

Among other things, this tells us an obvious thing: if the dimension of A' is smaller than the dimension of A, then there is no way to correct.

Let us elaborate more on this idea. Note that for every channel $\mathscr{C}$ it is always possible to find a Kraus representation where the Kraus operators are orthogonal, that is

$$\mathrm{Tr}[C_j^\dagger C_i] = 0 \qquad \forall i \neq j \, . \tag{9.2}$$

If you don't believe this, try the following

Exercise 26 For a quantum channel $\mathscr{C}$, consider the Choi matrix $\Phi_{\mathscr{C}} = (\mathscr{C} \otimes \mathscr{I}_A)(|\Phi\rangle\langle\Phi|)$ and diagonalize it as

$$\Phi_{\mathscr{C}} = \frac{1}{d_A} \sum_{i=1}^{K_{\min}} |C_i\rangle\!\rangle \langle\!\langle C_i| \, ,$$

where $\{|C_i\rangle\!\rangle\}_{i=1}^{K_{\min}}$ are orthogonal vectors in $\mathcal{H}_{A'} \otimes \mathcal{H}_A$. Show that $\mathscr{C}(\rho) = \sum_i C_i \rho C_i^\dagger$ is a Kraus representation of $\mathscr{C}$ with orthogonal Kraus operators.

Now, if the channel is correctable, we can combine the KL condition $C_j^\dagger C_i = \sigma_{ij} \, I_A$ with Eq. (9.2), getting

$$\sigma_{ij} = p_i \, \delta_{ij} \qquad p_i = \mathrm{Tr}[C_i^\dagger C_i]/d_A \, .$$

This means that the rank of σ is equal to the number of orthogonal Kraus operators $K_{\min}$, which in turn is equal to the rank of the Choi matrix $\Phi_{\mathscr{C}}$. Hence the packing bound can be equivalently written as

$$d_{A'} \geq d_A \, \mathsf{rank}(\Phi_{\mathscr{C}}) \, .$$

9.5 The physical meaning of the Knill-Laflamme condition

The KL condition is a mathematical condition for correctability. At first sight, it may look mysterious: Why do we have a quantum state σ appearing in the condition for the correctability of a channel?

The answer is a deep fact about quantum error correction, and also about quantum mechanics itself. In order to find the answer, it is important to recall that every quantum channel can be realized by an isometry that encodes the state of the input system A into the state of a composite system $A'B$, followed by a partial trace that discards system B. In the context of error correction, system B is usually called the *environment*, because it represents the physical system that interacts with system A and causes noise on it.

For our channel $\mathscr{C}$, let us choose a system B and an isometry $W : \mathcal{H}_A \to \mathcal{H}_{A'} \otimes \mathcal{H}_B$ such that

$$\mathscr{C}(\rho) = \mathrm{Tr}_B[W\rho W^\dagger]\,.$$

The channel $\mathscr{C}$ represents the flow of information from A to A'. But what about the flow of information from A to B? This flow of information is described by the *complementary channel* $\widetilde{\mathscr{C}}$, which is the channel from A to B defined by

$$\widetilde{\mathscr{C}}(\rho) = \mathrm{Tr}_{A'}[W\rho W^\dagger] \qquad \forall \rho \in \mathsf{St}(\mathcal{H}_A)\,.$$

Now, the intuition tells us that, if the channel from A to A' cannot be corrected, this is because some information has been transferred to B and therefore became unavailable. Conversely, one expects that if no information goes to B, then all the information about the initial state should be contained in system A'.

Mathematically, the condition that no information goes to system B is that the channel $\widetilde{\mathscr{C}}$ is an *erasure channel*, that is, there exists a fixed state σ of system B such that

$$\widetilde{\mathscr{C}}(\rho) = \sigma \qquad \forall \rho \in \mathsf{St}(\mathcal{H}_A)\,.$$

Now, it is not by chance that we used the letter σ to denote the state of system B: σ is exactly the density matrix that appears in the KL condition! The proof is contained in the following:

Theorem 16 (Alternative form of the Knill-Laflamme condition) *A channel $\mathscr{C}$ is correctable if and only if the complementary channel $\widetilde{\mathscr{C}}$ is an erasure channel.*

Proof. Let K be the dimension of system B and let us write the isometry W as

$$\begin{aligned} W &= \sum_{i=1}^{K} (I_{A'} \otimes |i\rangle\langle i|) W \\ &= \sum_{i=1}^{K} C_i \otimes |i\rangle \qquad C_i := (I_{A'} \otimes \langle i|)\, W\,. \end{aligned}$$

By definition, $\mathscr{C}(\rho) = \sum_{i=1}^{K} C_i \rho C_i^\dagger$ is a Kraus representation of channel $\mathscr{C}$. On

the other hand, the complementary channel $\widetilde{\mathscr{C}}$ can be written as

$$\begin{aligned}\widetilde{\mathscr{C}}(\rho) &= \sum_{i,j} \mathrm{Tr}_{A'}\left[(C_i \otimes |i\rangle)\rho(C_j^\dagger \otimes \langle j|)\right] \\ &= \sum_{i,j} \mathrm{Tr}[C_i \rho C_j^\dagger]\, |i\rangle\langle j| \\ &= \sum_{i,j} \mathrm{Tr}[C_j^\dagger C_i \rho]\, |i\rangle\langle j|\,.\end{aligned}$$

Now, the channel $\mathscr{C}$ is correctable if and only if $C_j^\dagger C_i = \sigma_{ij}\, I_A$. In turn, this is true if and only if

$$\mathrm{Tr}[C_j^\dagger C_i \rho] = \sigma_{ij} \qquad \forall \rho \in \mathsf{St}(\mathcal{H}_A)\,,$$

that is, if and only if the complementary channel $\widetilde{\mathscr{C}}$ satisfies $\widetilde{\mathscr{C}}(\rho) = \sigma$ for every state ρ. ■

Note that the information balance between system and environment is a specific feature of Quantum Information. In the classical world, it is still possible to have an interaction that copies information in the environment. Instead, in the quantum world, this is not possible due to the no-cloning theorem. Even more strongly, if we can retrieve the quantum state of the system, then no information *at all* can flow to the environment.

This fact is fundamental not only for quantum error correction, but also for quantum cryptography. Suppose that $\mathscr{C}$ is a quantum communication channel used by Alice (system A) to send quantum information to Alice' (system A'). Now, if Alice and Alice' can make sure that the communication channel is correctable, then they are automatically sure that no one else can extract information about the state that Alice is sending. After all, checking whether the channel is correctable is not too hard: Alice can send states either in the computational basis or in the Fourier basis, and Alice' can try to retrieve them after the channel. If she succeeds, this means that the channel is correctable, and, therefore, also cryptographically secure.

9.6 How to find good encodings

Until now we asked how to invert the action of a quantum channel $\mathscr{C}$ with some other channel $\mathscr{R}$. However, quantum error correction is more than just inverting a channel: in quantum error correction one is given a channel $\mathscr{E}$, which represents the noise, and one has to find a way to encode information into the input system of the channel. Typically, $\mathscr{E}$ is a channel from a system A' to itself, and one has to find a way to encode the states of the original system A into states of system A' in a way that can be corrected. A good encoding is a quantum channel from A to A' with the property that one can find a recovery channel $\mathscr{R}$ such that

$$\mathscr{R}\mathscr{E}\mathscr{V} = \mathscr{I}_A\,. \tag{9.3}$$

How can we find good encodings? First of all, note that, in order to be a good encoding, $\mathscr{V}$ must be a correctable channel:

Lemma 2 *If the channel $\mathscr{E}\mathscr{V}$ is correctable, then also $\mathscr{V}$ must be correctable.*

Proof. If $\mathscr{E}\mathscr{V}$ is correctable, then there is a recovery channel $\mathscr{R}$ such that $\mathscr{R}\mathscr{E}\mathscr{V} = \mathscr{I}_A$. Defining $\mathscr{R}' := \mathscr{R}\mathscr{E}$ we have that $\mathscr{R}'\mathscr{V} = \mathscr{I}_A$, that is, $\mathscr{V}$ is correctable. ■

We have proved that every encoding channel $\mathscr{V}$ must be correctable. Equivalently, this means that $\mathscr{V}$ is a random mixture of orthogonal isometries, that is $\mathscr{V} = \sum_m p_m \mathscr{V}_m$, where $\{p_m\}$ are probabilities and $\mathscr{V}_m(\rho) = V_m \rho V_m^\dagger$ for some isometry V_m. Clearly, if $\mathscr{V}$ is a good encoding, then each isometry $\mathscr{V}_m$ is also a good encoding: indeed, the condition of Eq. (9.3), together with the no-information without disturbance theorem, implies that we must have

$$\mathscr{R}\mathscr{E}\mathscr{V}_m = \mathscr{I}_A \qquad \forall m\,.$$

Hence, from now on we will restrict our attention to isometric encodings $\mathscr{V}$, of the form $\mathscr{V}(\rho) = V\rho V^\dagger$ for some isometry V. Such an encoding encodes the state of system A into the subspace $\mathcal{S} := V\mathcal{H}_A$, which is called the *code*. Note that the correspondence between the isometry V and the subspace $\mathcal{S}$ is not one-to-one: there are many isometries such that $V\mathcal{H}_A = \mathcal{S}$.

The problem now is to find good isometric encodings. To to this, we need to find which subspaces are *good quantum codes*: a subspace $\mathscr{S}$ is a good quantum code if there is some isometry V such that

1. $\mathscr{E}\mathscr{V}$ is correctable
2. $\mathcal{S} = V\mathcal{H}_A$

The KL condition gives a criterion to establish whether or not a subspace $\mathcal{S}$ is a good code for a channel $\mathscr{C}$:

Theorem 17 (Knill-Laflamme condition for good codes) *Let $\mathscr{E}$ be a quantum channel acting on A', with Kraus representation $\mathscr{E}(\rho) = \sum_{i=1}^k E_i \rho E_i^\dagger$, and let $\mathcal{S}$ be a subspace of $\mathcal{H}_{A'}$. The subspace $\mathcal{S}$ is a good code for $\mathscr{E}$ if and only if*

$$PE_j^\dagger E_i P = \sigma_{ij}\, P \tag{9.4}$$

where P is the projector on $\mathcal{S}$ and σ is a quantum state.

Proof. Suppose that $\mathcal{S}$ is a good code for $\mathscr{E}$. By definition, this means that there exists an isometry V such that the channel $\mathscr{E}\mathscr{V}$ is correctable and $V\mathcal{H}_A = \mathcal{S}$ (equivalently, $VV^\dagger = P$). Applying the KL condition, we obtain the condition

$$V^\dagger E_j^\dagger E_i V = \sigma_{ij}\, I_A\,,$$

which implies $VV^\dagger E_j^\dagger E_i VV^\dagger = \sigma_{ij}\, VV^\dagger$. Recalling that $VV^\dagger = P$ we then obtain the desired KL condition $PE_j^\dagger E_i P = \sigma_{ij}\, P$. Conversely, suppose that

the code $\mathcal{S}$ satisfies the KL condition $PE_j^\dagger E_i P = \sigma_{ij}\, P$. Take any isometry V such that $V\mathcal{H}_A = \mathcal{S}$ (equivalently, $VV^\dagger = P$). Then one has

$$\begin{aligned}
\sigma_{ij}\, I_A &= \sigma_{ij}\, (V^\dagger V)(V^\dagger V) \\
&= \sigma_{ij} V^\dagger P V \\
&= V^\dagger P E_j^\dagger E_i P V \\
&= V^\dagger (VV^\dagger) E_j^\dagger E_i (VV^\dagger) V \\
&= V^\dagger E_j E_i V\,.
\end{aligned}$$

By KL, this means that the channel $\mathscr{E}\mathscr{V}$ is correctable. Hence, $\mathcal{S}$ is a good quantum code. ■

The above theorem tells us whether or not a subspace $\mathcal{S}$ is a good quantum code. Moreover, it tells us that, once we found a good code $\mathcal{S}$, *every* isometry V such that $V\mathcal{H}_A = \mathcal{S}$ will be a good encoding of system A into system A'.

9.7 Quantum packing bound for non-degenerate codes

How big should be the dimension of A' in order to have a good quantum code for a channel $\mathscr{E}$? As we saw before, an answer is given by the quantum packing bound:

$$d_{A'} \geq d_A\, R \qquad\qquad R = \mathsf{rank}(\sigma)\,,$$

where σ is the quantum state appearing in the KL condition $PE_j^\dagger E_i P = \sigma_{ij}\, P$. Essentially, the packing bound says that $\mathcal{H}_{A'}$ should be big enough to contain R orthogonal subspaces of dimension d_A.

Now, it would be nice if we could compute $\mathsf{rank}(\sigma)$ only in terms of properties of the channel $\mathscr{E}$, as we did before. Unfortunately, this is not possible: the matrix σ depends not only on the channel $\mathscr{E}$, but also on the code $\mathcal{S}$ that we are using (equivalently, it depends on the projector P).

To make things simpler, it is useful to restrict our attention to a particular class of codes:

Definition 9 *A good code $\mathcal{S}$ is* non-degenerate *if, given an orthogonal Kraus representation $\mathscr{E}(\rho) = \sum_i^{K_{\min}} E_i \rho E_i^\dagger$, the matrix σ_{ij} appearing in the KL condition $PE_j^\dagger E_i P = \sigma_{ij} P$ is non-degenerate.*

For a non-degenerate code, we have $\mathsf{rank}(\sigma) = K_{\min}$. Hence, the packing bound for non-degenerate codes is

$$d_{A'} \geq d_A\, K_{\min}\,.$$

Recall that $K_{\min}$, the number of Kraus operators in an orthogonal Kraus representation is equal to the rank of the Choi matrix $\Phi_{\mathscr{E}}$.

In the following, we give a few examples of application of the quantum packing bound for non-degenerate codes:

1. **Single-qubit bit flips.** Consider the quantum channel $\mathscr{E}$, acting on N qubits, defined by

$$\mathscr{E}(\rho) := (1-p)\rho + \frac{p}{N}\sum_{n=1}^{N}\mathscr{X}_n(\rho)\,,$$

where $\mathscr{X}_n(\rho)$ is the unitary channel that consists in applying the Pauli matrix X on the n-th qubit. How many qubits do we need in order to encode one qubit with a non-degenerate code? Here, we have $d_A = 2$ and $d_{A'} = 2^N$ and $K_{\min} = 1+N$. Hence, the packing bound for non-degenerate codes gives:

$$2^{N-1} \geq N + 1\,.$$

Clearly, the minimum number of qubits needed to satisfy this condition is $N_{\min} = 3$. This bound is saturated by the example that we showed at the beginning of the chapter, where we saw that 1 qubit can be encoded into 3 qubits in a way that allows to correct random bit flips on single qubits.

2. **Arbitrary Pauli errors on a single qubit.** Consider a more general type of error, where one of the qubits that we use undergoes an arbitrary Pauli gate $\{X, Y, Z\}$. For example, consider the following channel acting on N qubits

$$\mathscr{E}(\rho) := (1-p)\rho + \frac{p}{3N}\sum_{n=1}^{N}\sum_{k=1}^{3}\mathscr{U}_{n,k}(\rho)\,,$$

where $\mathscr{U}_{n,k}$ is the unitary channel that consists in applying the Pauli matrix U_k ($U_1 = X, U_2 = Y, U_3 = Z$) to the n-th qubit. Now, suppose that we want to encode one qubit into N qubits in a way that can correct the channel $\mathscr{E}$. Here we have $d_A = 2, d_{A'} = 2^N$ and $K_{\min} = 1 + 3N$, and, therefore, the packing bound for non-degenerate codes gives:

$$2^{N-1} \geq 1 + 3N\,.$$

The smallest N satisfying this bound is $N = 5$. An example of quantum code that achieves this bound is the 5 qubit code by DiVincenzo and Shor, which encodes a qubit into 5 qubits in a way that can correct arbitrary single-qubit Pauli errors.

3. **Arbitrary Pauli errors on at most t qubits: the quantum Hamming bound.** Consider the case of a channel $\mathscr{E}$ acting on N qubits, which consists in applying at random some some Pauli gates to a group of $m \leq t$ qubits

$$\mathscr{E}(\rho) = (1-p)\rho + \frac{p}{t}\sum_{m=0}^{t}\frac{1}{\binom{N}{m}}\sum_{\mathbf{n}}\frac{1}{3^m}\sum_{\mathbf{k}}\mathscr{U}_{\mathbf{n},\mathbf{k}}(\rho)\,,$$

where $\mathbf{n} = (n_1, \ldots, n_m)$ labels a group of m qubits that are affected by the errors, $\mathbf{k} = (k_1, \ldots, k_m)$ tells which Pauli error acts on the m qubits, and $\mathscr{U}_{\mathbf{n},\mathbf{k}}$ is the unitary channel which corresponds to having the Pauli error U_{k_j} on the qubit n_j, for $j = 1, \ldots, m$. For this channel, the number of orthogonal Kraus operators is

$$K_{\min} = \sum_{m=0}^{t} \binom{N}{m} 3^m \, .$$

Suppose now that we want to encode K qubits into the N qubits that undergo the channel. The quantum packing bound for non-degenerate codes then gives

$$2^{N-K} \geq \sum_{m=0}^{t} \binom{N}{m} 3^m \, .$$

This particular bound is known with the name of *quantum Hamming bound*. Note, that the quantum Hamming bound, is a bound for *non-degenerate codes*. In principle, *degenerate codes* could do better: it is still possible that one can encode K qubits into N qubits, correcting errors on t qubits, in such a way that N is smaller than the value given by the Hamming bound. However, over the past years nobody has ever found a degenerate code that beats the Hamming bound. Finding a degenerate code that violates the Hamming bound, or proving that the Hamming bound holds for *every* code, is one of the open problems in the theory of quantum error correction.

9.8 Correct one to correct them all

In this chapter we saw how to correct a quantum channel $\mathscr{C}$. Precisely, we saw that every correctable channel $\mathscr{C}$ can be written as

$$\mathscr{C} = \sum_m p_m \, V_m \rho V_m^\dagger \, ,$$

where $\{V_m\}$ is a set of orthogonal isometries. The way to correct was simply to apply a quantum instrument, that selects one value of m, and, for the value m, to correct the isometry V_m.

However, this protocol seems to depend on the channel $\mathscr{C}$. It seems that for every channel we have to design a specific, taylor-made, error correction protocol. If this were true, error correction would be very hard: we could correct only channels that we know perfectly. Luckily, this is not the case. Instead, it is possible to see that the same protocol that corrects a given channel $\mathscr{C}$ can be used for many other channels.

Proposition 6 *Let $\mathscr{C}$ be a quantum channel with Kraus representation $\mathscr{C}(\rho) = \sum_i C_i \rho C_i^\dagger$ and let $\mathscr{D}$ be a channel with Kraus representation $\mathscr{D}(\rho) = \sum_j D_j \rho D_j^\dagger$,*

with $D_j \in \mathsf{Span}\{C_i\}$. If $\mathscr{C}$ is correctable with recovery channel $\mathscr{R}$, then also $\mathscr{D}$ is correctable with recovery channel $\mathscr{R}$.

Proof. Let $\mathscr{R}(\rho) = \sum_m R_m \rho R_m^\dagger$ be an Kraus representation of the channel $\mathscr{R}$. As we saw in the proof of the KL condition, the requirement $\mathscr{R}\mathscr{C} = \mathscr{I}_A$ implies $R_m C_i = \lambda_{mi} I_A$, for some coefficients $\{\lambda_{mi}\}$. Now, since the Kraus operators of $\mathscr{D}$ are in the linear span of the Kraus operators of $\mathscr{C}$, we can write $D_j = \sum_i d_{ji} C_i$. Hence, we have

$$\begin{aligned}
\mathscr{R}\mathscr{D}(\rho) &= \sum_m R_m \left(\sum_i d_{ji} C_i \right) \rho \left(\sum_k \bar{d}_{jk} C_k^\dagger \right) R_m^\dagger \\
&= \sum_{m,i,k} d_{ji}\, \bar{d}_{jk}\, (R_m C_i)\, \rho \left(C_k^\dagger R_m^\dagger \right) \\
&= \left(\sum_{m,i,k} d_{ji}\, \bar{d}_{jk}\, \lambda_{mi}\, \bar{\lambda}_{mk} \right) \rho \\
&= c\, \rho
\end{aligned}$$

By definition, since $\mathscr{R}\mathscr{D}$ is trace-preserving, one must have $c = 1$, that is, $\mathscr{R}\mathscr{D}(\rho) = \rho$ for every ρ. This proves that $\mathscr{D}$ is correctable with recovery channel $\mathscr{R}$. ■

The same result holds for good quantum codes:

Corollary 4 *Let $\mathscr{C}$ be a quantum channel with Kraus representation $\mathscr{C}(\rho) = \sum_i C_i \rho C_i^\dagger$ and let $\mathscr{D}$ be a channel with Kraus representation $\mathscr{D}(\rho) = \sum_j D_j \rho D_j^\dagger$, with $D_j \in \mathsf{Span}\{C_i\}$. If $\mathcal{S}$ is a good quantum code for $\mathscr{C}$, then $\mathcal{S}$ is a good quantum code for $\mathscr{D}$ and the same error correction protocol can be used to correct both $\mathscr{C}$ and $\mathscr{D}$.*

For example, this result can be applied to the channel $\mathscr{E}$ that corresponds to arbitrary Pauli errors acting on N qubits:

$$\mathscr{E}(\rho) = (1-p)\rho + \frac{p}{3N} \sum_{n=1}^{N} \sum_{k=1}^{3} \mathscr{U}_{n,k}(\rho)\,,$$

where $\mathscr{U}_{n,k}$ is the unitary channel that consists in applying the Pauli matrix U_k ($U_1 = X, U_2 = Y, U_3 = Z$) to the n-th qubit. It is easy to see that *every* operator C acting on a single qubit can be written as

$$C = c_0 I + \sum_{k=1}^{3} c_k U_k$$

where $\{c_k\}_{k=0}^3$ are arbitrary complex numbers. This means that a good code for channel $\mathscr{E}$ is a good code for *every quantum channel acting on a single qubit.*

For example, the five qubits code by Shor and DiVincenzo allows one to correct not only channel $\mathscr{E}$, but also the channel $\mathscr{D}$ given by

$$\mathscr{D}(\rho) = \frac{1}{5} \sum_{m=1}^{5} \mathscr{C}_{0,m}(\rho) \,,$$

where $\mathscr{C}_{0,m}$ is an erasure channel acting on qubit m. Even if we loose one qubit completely, the state of the remaining four qubits is enough to decode the initial state! This error correcting code can be used also for secret sharing: the state of one qubit is encoded in the state of 5 qubits in such a way that every group of 4 qubits is enough to recover the initial state.

9.9 Chapter summary

In this chapter we discussed how to protect quantum information from noise, by choosing a suitable encoding. We first noted that the no-cloning theorem poses a challenge to the design of error correcting codes. We then developed a general theory of error correction, discussing the Knill-Laflamme condition, which can be used to establish whether or not a given channel can be corrected. Physically, the Knill-Laflamme condition implies that a channel is correctable if and only if it does not allow any information to flow to the environment. When a channel can be corrected, the correction protocol has a general form: it consist in a quantum instrument—which projects the state of the system inside a set of orthogonal subspaces—followed by a suitable recovery operation. We then provide packing bounds, which are lower bounds on the total dimension of the Hilbert space when the channel is correctable. A special case of packing bound is the Hamming bound, which quantifies the number of physical qubits needed to encode k qubits, with errors on at most t qubits—when the code is non-degenerate. It is an open question whether or not there exist degenerate codes that require less qubits. Finally, we showed a fundamental fact about quantum error correction: if a channel $\mathscr{C}$ is correctable, then every other channel with Kraus operators that are linear combinations of the Kraus operation $\mathscr{C}$ can be corrected by the same correction protocol that corrects $\mathscr{C}$.

Part III

Quantum entanglement and communication

Chapter 10

Entanglement of pure bipartite states

With this chapter we start the more advanced part of our journey in the quantum world. In the first part, you learnt the basic rules of quantum theory and saw some first examples of the advantages of quantum information, like the Deutsch-Josza's algorithm, the dense coding protocol, and the CHSH game. In the second part, you encountered a number of machines that process information in the quantum world: machines that try to copy information and machines that transfer information from one place to another, machines that try to distinguish quantum states and machines that can perform a set of quantum gates depending on a set of instructions. Finally, you saw how these machines can correct errors using a clever choice of encoding and decoding operations. Now it is time to enter more deeply in the realm of quantum information. In the last four chapters we will find out how to compare different quantum resources and how to use quantum systems to make faster computations.

10.1 When is a quantum state more entangled than another?

Entanglement is a precious resource. If Alice and Bob have two systems in an entangled state, they can obtain many advantages: for example, we know that when Alice and Bob have two photons in the Bell state $|\Phi^+\rangle$ they can

- increase their probability to win in the CHSH game
- communicate 2 classical bits by sending only 1 qubit (dense coding)
- transmit the state of a qubit by sending only 2 classical bits (teleportation).

Since entanglement is a resource, it is important to understand when a state "**more entangled**" than another. But how can we compare two quantum states?

In this chapter we will answer this question in the case where the two states are pure.

10.2 Transforming a pure state into another by a LOCC protocol

Suppose that Alice and Bob are far apart and have two quantum systems in the state $|\Psi\rangle_{AB}$. They would like to transform this state into another state $|\Psi'\rangle_{AB}$, but, for some reason, they are not allowed to send quantum systems to each other. Instead, they can only send classical data and use these data to coordinate the operations that they perform in their labs. For example, they can have a cellphone and communicate to each other the outcomes of their measurements.

The set of protocols that Alice and Bob can realize by exchanging classical data and by performing local operations in their labs depending on these data is called *LOCC (Local Operations and Classical Communication)*. The most general LOCC protocol has the following form

1. Alice performs a measurement on her system, using a quantum instrument $\{\mathscr{A}_{x_1}\}$, obtains the outcome x_1, and communicates it to Bob

2. Bob performs a measurement on his system, using a quantum instrument $\left\{\mathscr{B}_{x_2}^{(x_1)}\right\}$, that depends on Alice's outcome. He obtains the outcome y_1 and communicates it to Alice.

3. Alice performs a measurement on her system, using a quantum instrument $\left\{\mathscr{A}_{x_3}^{(x_1x_2)}\right\}$, that depends on all the outcomes obtained so far. She obtains the outcome x_3 and communicates it to Bob.

4. Bob performs a measurement on his systems, using a quantum instrument $\left\{\mathscr{B}_{x_4}^{(x_1x_2x_3)}\right\}$,

5. They continue for N rounds, and at each round Alice or Bob performs a measurement that depends on all the classical outcomes obtained so far.

Note that

- the quantum instruments used by Alice and Bob can have different input systems and different output systems. For example, Alice's first instrument $\{\mathscr{A}_{x_1}\}$ could transform the input state A into another quantum system A' in Alice's lab.

- the quantum instruments used by Alice and Bob can have as many outcomes as we want. For example, Alice's first instrument $\{\mathscr{A}_{x_1}\}$ can have a single outcome, corresponding to a quantum channel $\mathscr{A}$: $\{\mathscr{A}_{x_1}\} \equiv \{\mathscr{A}\}$. In this case there is no classical communication from Alice to Bob.

- it does not matter who starts the protocol: the case where Bob starts can be also put in the above form by choosing that Alice's instrument has a single outcome, and that the corresponding operation is the identity channel: $\{\mathscr{A}_{x_1}\} \equiv \{\mathscr{I}_A\}$. As a matter of fact, this means that Alice does not do anything on her system and does not send any classical information.

Summing over all the outcomes obtained in a LOCC protocol, we obtain a quantum channel $\mathscr{L}$, that transforms Alice's and Bob's system. We call such a channel a *LOCC channel*.

Let us see some examples of LOCC protocols and LOCC channels:

1. *Local protocols.* Suppose that Alice and Bob do not communicate at all. In this case, their operations are described by two quantum channels $\mathscr{A}$ and $\mathscr{B}$, respectively. Hence, the protocol is described by the quantum channel
$$\mathscr{L} = \mathscr{A} \otimes \mathscr{B}\,,$$
which transforms Alice's and Bob's systems independently.

2. *One-way LOCC protocols.* Suppose that the protocol has only one round of classical communication—for example, from Alice to Bob. This means that Alice performs a measurement with instrument $\{\mathscr{A}_{x_1}\}$ and communicates the outcome to Bob, who performs a quantum channel $\mathscr{B}^{(x_1)}$. This is the situation in the quantum teleportation protocol, where Alice makes a measurement, transmits the outcome to Bob, and then Bob performs a unitary gate that depends on the outcome. The protocol is described by the channel
$$\mathscr{L} = \sum_{x_1} \mathscr{A}_{x_1} \otimes \mathscr{B}^{(x_1)}\,.$$
Of course, we can also have protocols where the information goes from Bob to Alice. This means that Bob performs a measurement with instrument $\{\mathscr{B}_{x_2}\}$ and communicates the outcome to Alice, who performs a quantum channel $\mathscr{A}^{(x_2)}$. In this case, the protocol is described by the channel
$$\mathscr{L} = \sum_{x_2} \mathscr{A}^{(x_2)} \otimes \mathscr{B}_{x_2}\,.$$

In general, it is hard to classify all possible LOCC channels. However, we will not need to do it here: the following theorem will be enough.

Theorem 18 (Lo-Popescu theorem) *The following are equivalent:*

1. $|\Psi\rangle$ *can be converted into* $|\Psi'\rangle$ *by a LOCC protocol*

2. $|\Psi\rangle$ *can be converted into* $|\Psi'\rangle$ *by a one-way LOCC protocol involving communication from Alice to Bob, where Bob's channel* $\mathscr{B}^{(x_1)}$ *is unitary*

3. $|\Psi\rangle$ *can be converted into* $|\Psi'\rangle$ *by a one-way LOCC protocol involving communication from Bob to Alice, where Alice's channel* $\mathscr{A}^{(x_2)}$ *is unitary.*

Proof. See Proposition 12.14 p. 575 of Nielsen-Chuang or the Appendix of these chapter notes. ■

The Lo-Popescu theorem tells us that in order to find out whether or not $|\Psi\rangle$ can be transformed into $|\Psi'\rangle$ is by an LOCC protocol it is enough to consider a very special class of protocols: the 1-way protocols using unitary gates. Clearly, this simplifies thee problem. In the following we will look for an event simpler condition.

10.3 Definition: when a quantum state is more entangled than another

Definition 10 *We say that* $|\Psi\rangle$ *is* **more entangled** *than* $|\Psi'\rangle$ *if there exists a LOCC channel that transforms* $|\Psi\rangle$ *into* $|\Psi'\rangle$.

Let us see that this definition makes sense, by observing a few examples:

1. **Product states.** Product states are the less entangled of all states. Indeed, every state $|\Psi\rangle$ is more entangled than a product state $|\Psi'\rangle = |\alpha\rangle|\beta\rangle$. In order to transform $|\Psi\rangle$ into $|\alpha\rangle|\beta\rangle$, Alice and Bob do not need to communicate at all: it is enough that Alice applies the erasure channel

$$\mathscr{A}(\rho) := |\alpha\rangle\langle\alpha| \qquad \forall \rho \in \mathsf{St}(A)\,,$$

and Bob applies the erasure channel

$$\mathscr{B}(\rho) := |\beta\rangle\langle\beta| \qquad \forall \rho \in \mathsf{St}(B)\,.$$

Clearly, in this way they obtain

$$(\mathscr{A} \otimes \mathscr{B})(|\Psi\rangle\langle\Psi|) = |\alpha\rangle\langle\alpha| \otimes |\beta\rangle\langle\beta|\,.$$

In other words, a product state is less entangled than any other bipartite state.

2. **Bell states.** Let A and B be two d-dimensional quantum systems. Then, every state $|\Psi\rangle$ less entangled than the Bell state $|\Phi^+\rangle = \frac{|I\rangle\!\rangle}{\sqrt{d}}$. An LOCC protocol transforming $|\Phi^+\rangle$ into $|\Psi\rangle$ is the following:

   - In her lab, Alice prepares two d-dimensional quantum systems A_1 and A_2 in the state $|\Psi\rangle_{A_1A_2}$.
   - Alice and Bob use the Bell state $|\Phi^+\rangle_{AB}$ to teleport the state of system A_2 to Bob, thus obtaining the state $|\Psi\rangle_{A_1B}$.
   - Since A_1 and A have the same dimension, having the state $|\Psi\rangle_{A_1B}$ is equivalent to having the state $|\Psi\rangle_{AB}$.

10.4 Definition: when a quantum state is more mixed than another

We know that a pure state $|\Psi\rangle_{AB}$ is entangled if and only if its marginal states on A and B are mixed. This is clear from the Schmidt decomposition $|\Psi\rangle_{AB} = \sum_i \sqrt{p_i}|\alpha_i\rangle|\beta_i\rangle$, which tells us that the marginal states on A and B are given by

$$\rho_A = \sum_i p_i \, |\alpha_i\rangle\langle\alpha_i| \qquad\qquad \rho_B = \sum_i p_i \, |\beta_i\rangle\langle\beta_i| \,, \tag{10.1}$$

respectively. Following this idea, it is natural to expect that a pure state $|\Psi\rangle_{AB}$ is *more entangled* than another pure state $|\Psi'\rangle_{AB}$ if and only if the marginal states of $|\Psi\rangle$ are *more mixed* than the marginal state of $|\Psi'\rangle$.

In order to prove this result, we have first to define what we mean when we say that a state is "more mixed" than another. A reasonable choice is the following:

Definition 11 *$\rho \in \mathsf{St}(\mathcal{H})$ is* **more mixed** *than $\rho' \in \mathsf{St}(\mathcal{H})$ if ρ can be obtained from ρ' by applying a random-unitary (RU) channel:*

$$\rho = \sum_{i=1}^{k} p_i \, \mathscr{U}_i(\rho')$$

where $\{p_i\}_{i=1}^k$ are probabilities and each $\mathscr{U}_i$ is a unitary channel, i.e. a channel of the form $\mathscr{U}_i(\rho) = U_i \rho U_i^\dagger$ for some unitary matrix U_i.

In words, the definition says that ρ is more mixed than ρ' iff ρ can be obtained from ρ' applying a unitary gate chosen at random.

In order to become familiar with the definition, it is good to see a couple of examples:

1. **Pure states.** Every state ρ is more mixed than a pure state $\rho' = |\varphi\rangle\langle\varphi|$: indeed, we can diagonalize ρ as $\rho = \sum_i p_i \, |\varphi_i\rangle\langle\varphi_i|$ and find a set of unitaries $\{U_i\}$ such that
$$U_i|\varphi\rangle = |\varphi_i\rangle \qquad \forall i \,.$$
In this way, we have $\rho = \sum_i p_i \, \mathscr{U}_i(|\varphi\rangle\langle\varphi|)$.

2. **Maximally mixed states.** No state ρ is more mixed than the state $\frac{I}{d}$. Indeed, if $\rho = \sum_i p_i \, \mathscr{U}_i(I/d)$, then we must have $\rho = \sum_i p_i \frac{U_i U_i^\dagger}{d} = \frac{I}{d}$. For this reason, we say that $\frac{I}{d}$ is the *maximally mixed state*. It is also easy to see that I/d is more mixed than any other state ρ': indeed, we can diagonalize ρ' as $\rho' = \sum_{n=1}^{d} p_n \, |\varphi_n\rangle\langle\varphi_n|$ and define the unitary gates
$$U_i := \sum_{n=1}^{d} |\varphi_{n\oplus i}\rangle\langle\varphi_n| \,,$$

where $\oplus$ denotes the sum modulo d. With this definition we have

$$\begin{aligned}\frac{1}{d}\sum_{i=1}^{d}\mathscr{U}_i(\rho) &= \frac{1}{d}\sum_{i,n} p_n|\varphi_{n\oplus i}\rangle\langle\varphi_{n\oplus i}| \\ &= \frac{1}{d}\sum_{j,n} p_n\,|\varphi_j\rangle\langle\varphi_j| \\ &= \frac{I}{d}\,.\end{aligned}$$

It is immediate to see that the property of being "more mixed" is invariant under the action of unitary gates:

Proposition 7 *If ρ is more mixed than ρ', then $U\rho U^\dagger$ is more mixed than $V\rho V^\dagger$ for every unitary gates U and V.*

Proof. By definition, if ρ is more mixed than ρ' there is a random-unitary channel such that $\rho = \sum_i p_i\,\mathscr{U}_i(\rho')$. Now, for every unitaries U and V, also $\sum_i p_i\,\mathscr{U}\mathscr{U}_i\mathscr{V}^\dagger$ is a ranom-unitary channel, and we have

$$\begin{aligned}\sum_i p_i\,\mathscr{U}\mathscr{U}_i\mathscr{V}^\dagger\left[\mathscr{V}(\rho)\right] &= \sum_i p_i\,\mathscr{U}\mathscr{U}_i(\rho) \\ &= \mathscr{U}(\rho')\,.\end{aligned}$$

■

As a consequence, the property of being "more mixed" must depend **only on the eigenvalues**. Let $\mathbf{p} := (p_1, p_2, \ldots, p_d)$ be the vector of eigenvalues of ρ and $\mathbf{p}' := (p'_1, \ldots, p'_d)$ be the vector of eigenvalues of ρ'. Since $\mathbf{p}$ and $\mathbf{p}'$ are classical probability distributions, we can think of them as density matrices that are diagonal in the computational basis. In short, we will say that $\mathbf{p}$ is *more mixed than* $\mathbf{p}'$ iff the diagonal density matrix $\mathsf{diag}(\mathbf{p})$ is more mixed than $\mathsf{diag}(\mathbf{p}')$.

Corollary 5 *Let ρ and ρ' be two quantum states, diagonalized as $\rho = \sum_i p_i\,|\varphi_i\rangle\langle\varphi_i|$ and $\rho' = \sum_i p'_i|\varphi'_i\rangle\langle\varphi'_i|$. Then, the following are equivalent*

1. *ρ is more mixed than ρ'*
2. *$\mathbf{p}$ is more mixed than $\mathbf{p}'$.*

Proof. Choose U and V to be the unitaries that diagonalize ρ and ρ', i.e. $\mathscr{U}(\rho) = \mathsf{diag}(\mathbf{p})$ and $\mathscr{V}(\rho') = \mathsf{diag}(\mathbf{p}')$. By the previous result we know that $\mathsf{diag}(\mathbf{p})$ must be more mixed than $\mathsf{diag}(\mathbf{p}')$, i.e. $\mathbf{p}$ must be more mixed than $\mathbf{p}'$. With the same reasoning it is immediate to prove the reverse implication: if $\mathbf{p}$ is more mixed than $\mathbf{p}'$, then ρ is more mixed than ρ'. ■

10.5 The relation between entanglement and mixedness

We are now ready to prove the desired theorem on the relation between mixedness and entanglement. Without loss of generality we consider two bipartite states $|\Psi\rangle$ and $|\Psi'\rangle$ on the **same** Hilbert spaces $\mathcal{H}_A \otimes \mathcal{H}_B$.

First of all, recall Eq. (10.1): for a pure state $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$, the marginal states ρ_A and ρ_B have the same eigenvalues. Now, if $|\Psi'\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ is another pure state, with marginal states ρ'_A and ρ'_B, it is clear that the following conditions are equivalent:

- ρ_A is more mixed than ρ'_A
- ρ_B is more mixed than ρ'_B.

In other words, it does not matter if we consider the marginal on Alice's side, or the marginal on Bob's side. For this reason, we will say that *the marginal of* $|\Psi\rangle$ *is more mixed than the marginal of* $|\Psi'\rangle$ without the need of specifying if we are talking about the marginal on A or the marginal on B. The goal of this paragraph is to prove the theorem: $|\Psi\rangle$ *is more entangled than* $|\Psi'\rangle$ *if and only if the marginal of* $|\Psi\rangle$ *is more mixed than the marginal of* $|\Psi'\rangle$.

Let us prove the first implication:

Lemma 3 *If the marginal of* $|\Psi\rangle$ *is more mixed than the marginal of* $|\Psi'\rangle$*, then* $|\Psi\rangle$ *is more entangled than* $|\Psi'\rangle$*.*

Proof. Without loss of generality, let us consider the marginals on system A, denoted by ρ_A and ρ'_A. By hypothesis $\rho_A = \sum_{i=1}^k p_i \, \mathscr{U}_i(\rho'_A)$. Let us define the mixed state ρ_{AB} as

$$\rho_{AB} := \sum_i p_i \, (\mathscr{U}_i \otimes \mathscr{I}_B)(|\Psi'\rangle\langle\Psi'|) \, .$$

By definition, the marginal of ρ_{AB} on system A is ρ_A:

$$\begin{aligned} \mathrm{Tr}_B[\rho_{AB}] &= \sum_i p_i \, \mathrm{Tr}_B \left[(\mathscr{U}_i \otimes \mathscr{I}_B)(|\Psi'\rangle\langle\Psi'|) \right] \, . \\ &= \sum_i p_i \, \mathscr{U}_i \left(\mathrm{Tr}_B[|\Psi'\rangle\langle\Psi'] \right) \\ &= \sum_i p_i \, \mathscr{U}_i(\rho'_A) \\ &= \rho_A \, . \end{aligned}$$

Now, consider a purification of ρ_{AB}, given by a pure state $|\Gamma\rangle_{ABC}$, where C is a purifying system. By definition, $|\Gamma\rangle$ is also a purification of ρ_A, with purifying system BC. Since, the state $|\Psi\rangle_{AB}|0\rangle_C$ is also a purification of ρ_A, the uniqueness of purification implies there must be a unitary gate U, acting on BC, such that

$$|\Gamma\rangle_{ABC} = (I_A \otimes U)|\Psi\rangle_{AB}|0\rangle_C \, .$$

In other words, if Alice and Bob have the state $|\Psi\rangle$, Bob can transform it in the state $|\Gamma\rangle$ by using only local operations in his lab: precisely, he just has to

- prepare system C in the state $|0\rangle$
- apply the gate U to systems B and C.

Once Alice and Bob have the state $|\Gamma\rangle$, it is clear how to transform it into the state $|\Psi'\rangle$ by a LOCC protocol. Indeed, by the steering property of quantum mechanics (cf. chapter 5), we know that there exists a measurement on C, represented by a POVM $\{P_i\}$, such that

$$\mathrm{Tr}_C\left[(I_{AB}\otimes P_i)|\Gamma\rangle\langle\Gamma|\right] = p_i\,(\mathscr{U}_i\otimes\mathscr{I}_B)(|\Psi'\rangle\langle\Psi'|)\,.$$

This means that, if Bob measures system C and finds outcome i, then he is sure that the remaining state of systems A and B is $(\mathscr{U}_i\otimes\mathscr{I}_B)(|\Psi'\rangle\langle\Psi'|)$. To obtain the state $|\Psi'\rangle$, he has only to communicate the outcome i to Alice, so that she can apply the unitary channel $\mathscr{U}_i^\dagger$ on system A. Since $|\Psi\rangle$ can be transformed into $|\Psi'\rangle$ by a LOCC channel, we proved that $|\Psi\rangle$ is more entangled than $|\Psi'\rangle$. ■

Note that the LOCC protocol in the proof of Theorem 3 is actually a *one-way* LOCC protocol, where Bob communicates an outcome to Alice and Alice performs a unitary channel depending on the outcome. This form is in agreement with the Lo-Popescu theorem, which states that every LOCC protocol that transforms a pure state into another is equivalent to a one-way protocol.

The Lo-Popescu theorem becomes very useful when we want to prove the other implication:

Lemma 4 *If $|\Psi\rangle$ is more entangled than $|\Psi'\rangle$, then the marginal of $|\Psi\rangle$ is more mixed than the marginal of $|\Psi'\rangle$.*

Proof. By the Lo-Popescu theorem, there exists a quantum instrument on B, say $\{\mathscr{B}_i\}$ and, for every outcome i, a unitary channel $\mathscr{U}_i$ acting on A, such that

$$\sum_i(\mathscr{U}_i\otimes\mathscr{B}_i)(|\Psi\rangle\langle\Psi|) = |\Psi'\rangle\langle\Psi'|\,.$$

Since $|\Psi'\rangle$ is pure, this means that there is a set of probabilities $\{p_i\}$ such that

$$(\mathscr{U}_i\otimes\mathscr{B}_i)(|\Psi\rangle\langle\Psi|) = p_i\,|\Psi'\rangle\langle\Psi'|$$

for every i. Equivalently, we have $(\mathscr{I}_A\otimes\mathscr{B}_i)(|\Psi\rangle\langle\Psi|) = p_i\,(\mathscr{U}_i^\dagger\otimes\mathscr{I}_B)(|\Psi'\rangle\langle\Psi'|)$, and, summing over i

$$(\mathscr{I}_A\otimes\mathscr{B})(|\Psi\rangle\langle\Psi|) = (\mathscr{A}\otimes\mathscr{I}_B)(|\Psi'\rangle\langle\Psi'|)\,,$$

where $\mathscr{B} := \sum_i\mathscr{B}_i$ is the channel corresponding to the instrument $\{\mathscr{B}_i\}$ and $\mathscr{A}$ is the random-unitary channel $\mathscr{A} := \sum_i p_i\,\mathscr{U}_i$. Discarding system B, we then obtain

$$\begin{aligned}\mathrm{Tr}_B[(\mathscr{I}_A\otimes\mathscr{B})(|\Psi\rangle\langle\Psi|)] &= \mathrm{Tr}_B[|\Psi\rangle\langle\Psi|]\\ &= \rho_A\end{aligned}$$

and

$$\begin{aligned}\mathrm{Tr}_B[(\mathscr{A}\otimes\mathscr{I}_B)(|\Psi'\rangle\langle\Psi|')] &= \mathscr{A}\left(\mathrm{Tr}_B[|\Psi'\rangle\langle\Psi'|]\right)\\ &= \mathscr{A}(\rho'_A)\\ &= \sum_i p_i\,\mathscr{U}_i(\rho'_A)\,.\end{aligned}$$

Therefore, we conclude that $\rho_A = \sum_i p_i\,\mathscr{U}_i(\rho'_A)$, i.e. ρ_A is more mixed than ρ'_A. ■

In conclusion, we established that the entanglement of pure bipartite states is equivalent to the mixedness of their marginals:

Theorem 19 *Let $|\Psi\rangle, |\Psi'\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ be two pure bipartite states. Then, the following are equivalent*

1. *$|\Psi\rangle$ is more entangled than $|\Psi'\rangle$*

2. *the marginal of $|\Psi\rangle$ is more mixed than the marginal of $|\Psi'\rangle$*

Two remarks are in order here:

1. **Entangled states of different systems.** It is immediate to generalize the results of this chapter to the case where $|\Psi\rangle$ and $|\Psi'\rangle$ belong to different Hilbert spaces, say $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ and $|\Psi'\rangle \in \mathcal{H}_{A'} \otimes \mathcal{H}_{B'}$. This situation is interesting in several applications: for example, Alice and Bob could have two qubit each ($\mathcal{H}_A \simeq \mathcal{H}_B \simeq \mathbb{C}^2 \otimes \mathbb{C}^2$) and each of them could transform her/his two qubits into a single qubit ($\mathcal{H}_{A'} \simeq \mathcal{H}_{B'} \simeq \mathbb{C}^2$). In order to extend our theorem to this case, we need to define what we mean when we say that the marginal of $|\Psi\rangle$ is more mixed than the marginal of $|\Psi'\rangle$. This can be done as follows:

   Definition 12 *Let ρ and ρ' be two quantum states of A and A', respectively. We say that ρ is more mixed than ρ' iff $\rho \otimes |\alpha'_0\rangle\langle\alpha'_0|$ is more mixed than $|\alpha_0\rangle\langle\alpha_0| \otimes \rho'$, where $|\alpha_0\rangle$ and $|\alpha'_0\rangle$ are two fixed pure states of A and A', respectively.*

   Note that the definition is independent of the choice of $|\alpha_0\rangle$ and $|\alpha'_0\rangle$, thanks to proposition 7.

2. The relation between entanglement and mixedness holds for pure bipartite states, but *not* for mixed bipartite states. For example, consider the product state $\rho_A \otimes \rho_B$, where $\rho_A = I/d$. In this case, the marginal on Alice's system is very mixed (in fact, maximally mixed), but the state is a product, and therefore it is not entangled at all!

10.6 The majorization criterion

The relation between entanglement of pure bipartite states and mixedness of their marginals is a very deep and fundamental fact. But how can we establish if a state is more mixed than another? Should we try all possible random-unitary channels? Of course, this does not seem very practical. Luckily, there is a simple necessary and sufficient condition that simplifies our life: this condition is called *majorization criterion* and is expressed only in terms of the eigenvalues of the density matrix.

For a density matrix ρ, let us denote by $\mathbf{p} = (p_1, \ldots, p_d)$ the vector of the eigenvalues, arranged in decreasing order:

$$p_1 \geq p_2 \cdots \geq p_d \geq 0\,.$$

Similarly, denote by $\mathbf{p}'$ the eigenvalues of ρ', arranged in decreasing order. With this notation, we have the following:

Theorem 20 (Majorization criterion) *ρ is more mixed than ρ' if and only if, for every*

$$\sum_{i=1}^{k} p_i \leq \sum_{i=1}^{k} p'_i \qquad\qquad \forall k \in \{1, \ldots, d-1\}\,. \tag{10.2}$$

When the conditions of Eq. (10.2) hold, we say that $\mathbf{p}$ *is majorized by* $\mathbf{p}'$, and denote this relation by $\mathbf{p} \preceq \mathbf{p}'$. Intuitively, $\mathbf{p} \preceq \mathbf{p}'$ means that $\mathbf{p}$ is "more flat" than $\mathbf{p}'$. The majorization criterion tells us a very simple and intuitive thing: a state is more mixed than another if and only if the probability distribution of its eigenvalues is more flat than the probability distribution of the eigenvalues of the other.

Let us apply the majorization criterion in the simple case of qubit states, where $\mathbf{p} = (p_1, 1-p_1)$ and $\mathbf{p}' = (p'_1, 1-p'_1)$. In this case we have that $\mathbf{p}$ is more mixed than $\mathbf{p}'$ if and only if $p_1 \leq p'_1$. This means that every two qubit states can be compared by just checking their maximum eigenvalues.

Note, however, that this simple result does not hold in dimension $d > 2$: for higher dimensional systems, not every two states can be compared. There are many examples of states ρ and ρ' such that neither ρ is more mixed than ρ' nor ρ' is more mixed than ρ. For instance, consider the probability distributions and $\mathbf{p} = \left(\frac{2}{5}, \frac{2}{5}, \frac{1}{5}\right)$ and $\mathbf{p}' = \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right)$:

- $\mathbf{p}'$ cannot be more mixed than $\mathbf{p}$, because $p'_1 = 1/2 > p_1 = 2/5$
- $\mathbf{p}$ cannot be more mixed than $\mathbf{p}'$, because $p_1 + p_2 = 4/5 > p'_1 + p'_2 = 3/4$.

10.7 Measuring mixedness, measuring entanglement

The majorization criterion is a necessary and sufficient condition for a state to be more mixed than another. However, checking the majorization criterion

means checking $d-1$ inequalities, where d is the dimension of the Hilbert space. Can we have a single quantity that tells us "how mixed" is a quantum state?

This quantity would be a *measure of mixedness*: a measure of mixedness should be a function M from the set of the density matrices to the set of real numbers, with the property that M that is *monotone under mixing*, that is,

$$M(\rho) \geqslant M(\sigma)$$

whenever ρ is more mixed than σ. Since we know that the relation of being "more mixed" depends only on the eigenvalues, every measure of mixedness must be of the form

$$M(\rho) = f(\mathbf{p})\,,$$

where $f(\mathbf{p})$ is a function $f : \mathbb{R}^d \to R$ with the property

$$f(\mathbf{p}) \geqslant f(\mathbf{p}') \qquad\qquad \mathbf{p} \leq \mathbf{p}'\,.$$

In mathematics, a function f with the above property is called *Schur-concave*, and, in fact, there are a lot of Schur-concave functions.

Now, every Schur-concave function defines a **measure of mixedness** and equivalently, a **measure of entanglement for pure bipartite states**. An important class of Schur-concave functions are the *Rényi entropies*

$$H_\alpha(\mathbf{p}) = \frac{1}{1-\alpha}\log\left[\sum_{i=1}^{d} p_i^\alpha\right] \qquad \alpha \geqslant 0\,,$$

where, for $\alpha = 1$, the Renyi entropy is defined as

$$H_1(\mathbf{p}) = \lim_{\alpha\to 1} H_\alpha(\mathbf{p})\,.$$

Using the Rényi entropies, we can define the *quantum Renyi entropies* as

$$\begin{aligned} S_\alpha(\rho) &:= \frac{1}{1-\alpha}\log \mathrm{Tr}[\rho^\alpha] \qquad \alpha \geqslant 0 \\ &= \left(\frac{\alpha}{1-\alpha}\right)\log \|\rho\|_\alpha \end{aligned}$$

where $\|\rho\|_\alpha$ is the "α-norm" $\|\rho\|_\alpha := (\mathrm{Tr}[\rho^\alpha])^{\frac{1}{\alpha}}$.

The quantum Rényi entropies have the following nice properties:

1. for every α, $S_\alpha(\rho) = 0$ if and only if ρ is pure.
2. for $\alpha > 0$, $S_\alpha(\rho) = \log d$ if and only if ρ is maximally mixed ($\rho = \frac{I}{d}$).
3. for every α, $0 \leqslant S_\alpha(\rho) \leqslant \log d$ for every ρ (this follows from Schur-concavity)
4. for every pair of states, ρ and σ, one has $S_\alpha(\rho\otimes\sigma) = S_\alpha(\rho) + S_\alpha(\sigma)$ (additivity property)

5. for every state ρ, $S_\alpha(\rho) \leq S_\beta(\rho)$ when $\beta \leq \alpha$.

Let us see a few examples of quanrum Rényi entropies for some different values of α:

1. $\alpha = 0$

$$S_0(\rho) = \log\left[\mathsf{rank}(\rho)\right]$$

The corresponding measurement of entanglement is the (logarithm of) the Schmidt rank:

if $|\Psi\rangle = \sum_{m=1}^{r} \sqrt{p_m}|\alpha_m\rangle|\beta_m\rangle$ is more entangled than $|\Psi'\rangle = \sum_{n=1}^{r'} \sqrt{p'_n}|\alpha'_n\rangle|\beta'_n\rangle$, then $r \geqslant r'$.

Since $S_0(\rho) \geqslant S_\alpha(\rho) \quad \forall\alpha$, S_0 is called the *max-entropy*.

2. $\alpha \to \infty$

$$S_\infty(\rho) = -\log\|\rho\|_\infty = -\log p_1$$

(logarithm of the maximum eigenvalue)

It is called *min-entropy*, because $S_\infty(\rho) \leqslant S_\alpha(\rho) \quad \forall\alpha$.

3. $\alpha = 2$

$$S_2(\rho) = -\log \mathrm{Tr}[\rho^2]$$

The quantity $\mathrm{Tr}[\rho^2]$ is called *purity*: if ρ is more mixed than σ, then $\mathrm{Tr}[\rho^2] \leqslant \mathrm{Tr}[\sigma^2]$.

4. $\alpha \to 1$

In classical information theory, $\lim_{\alpha\to 1} H_\alpha(\mathbf{p}) = \sum_{m=1}^{d} -p_m \log p_m =: H(\mathbf{p})$ is the *Shannon entropy*. In the quantum case:

$$\lim_{\alpha\to 1} S_\alpha(\rho) = -\mathrm{Tr}[\rho\log\rho] =: S(\rho)$$

is the *von Neumann entropy*.

The measure of mixedness/entanglement provide a quick way to see that some transformations are **impossible**: if $S_\alpha(\rho) < S_\alpha(\rho')$ for some α, then we are sure that ρ cannot be more mixed than ρ'.

But can we use a measure of mixedness/entanglement to argue that a transformation is **possible**? Unfortunately, the answer is no. For example, if $S(\rho) > S(\rho')$ we cannot conclude that ρ is "more mixed" than ρ'. We will see in the next paragraph that this problem disappears if we have many copies of ρ and ρ'.

10.8 Asymptotic transformations of pure bipartite states

Suppose that Alice and Bob have N copies of a state $|\Psi\rangle$ and want to convert them into M copies of a state $|\Psi'\rangle$. We know from the previous paragraph that a necessary condition is for this transformation is that

$$S\left(\rho'^{\otimes M}\right) \leq S\left(\rho^{\otimes N}\right) ,$$

where ρ and ρ' are the marginal states of $|\Psi\rangle$ and $|\Psi'\rangle$, respectively. Using the additivity property of the entropy, the above condition is equivalent to $MS(\rho') \leq NS(\rho)$, which implies the upper bound

$$\frac{M}{N} \leq \frac{S(\rho)}{S(\rho')} . \tag{10.3}$$

Now, suppose that Alice and Bob can tolerate a small error in their protocol: denoting by $\mathscr{L}_N$ the LOCC channel implemented by the protocol, Alice and Bob can accept that the output state $\mathscr{L}_N\left[(|\Psi\rangle\langle\Psi'|)^{\otimes N}\right]$ is not exactly the desired state $(|\Psi'\rangle\langle\Psi'|)^{\otimes M}$, but some state close to it. Here "close" means that

$$\left\|\mathscr{L}_N\left[(|\Psi\rangle\langle\Psi'|)^{\otimes N}\right] - (|\Psi'\rangle\langle\Psi'|)^{\otimes M}\right\|_1 < \epsilon,$$

for some positive ϵ. Recall that the trace norm $\|\cdot\|_1$ quantifies the distinguishability of two states: saying that the trace norm is small means saying that the two states are almost indistinguishable, even if we use the best possible measurement.

In addition, suppose that Alice and Bob can have as many copies of $|\Psi\rangle$ as they want. Then the question is: how many copies of $|\Psi'\rangle$ can they obtain from N copies of $|\Psi\rangle$ under the requirement that the error goes to zero in the limit $N \to \infty$?

Let us formulate the question in a mathematically precise way. Suppose that for every N, Alice and Bob have a LOCC protocol that produces $M = M(N)$ approximate copies of $|\Psi'\rangle$. Let us denote by $\{\mathscr{L}_N\}_{N\in\mathbb{N}}$ the sequence of LOCC channels that correspond to Alice's and Bob's protocols. Naturally, we want to know how big is the ratio $M(N)/N$ when N becomes large, that is, we want to know how big is the *rate*

$$R := \liminf_{N\to\infty} \frac{M(N)}{N} .$$

Note that the rate is a property of the sequence of protocols that Alice and Bob use. Equivalently, it is a property of the sequence of LOCC channels $\{\mathscr{L}_N\}_{N\in\mathbb{N}}$. Of course, Alice and Bob also want the error to be small: ideally, when N is infinite, they want the error to be zero. This requirement motivates the definition of *achievable rate*:

Definition 13 *A rate R is achievable if for every N there exists a sequence of LOCC channels $\{\mathscr{L}_N\}_{N\in\mathbb{N}}$ that has rate R and satisfies the condition*

$$\lim_{N\to\infty} \left\|\mathscr{L}_N\left((|\Psi\rangle\langle\Psi'|)^{\otimes N}\right) - (|\Psi'\rangle\langle\Psi'|)^{\otimes RN}\right\|_1 = 0.$$

In other words, when a rate R is achievable, Alice and Bob can transform N copies of $|\Psi\rangle$ into RN copies of $|\Psi'\rangle$ with an error that is arbitrarily small, provided that N is sufficiently large.

If we want to know how many copies Alice and Bob can produce, we have to answer the question: what is the largest achievable rate? The answer to this question is beautiful, and it is given by the following theorem:

Theorem 21 *The supremum over all achievable rates is given by*

$$\sup\{R \mid R \ \ is \quad achievable\} = \frac{S(\rho)}{S(\rho')}.$$

We will see the idea behind the proof of this result in the next chapter. For the moment, let us appreciate the meaning of this result: the result tells us that, when Alice and Bob have an arbitrarily large number of copies of $|\Psi\rangle$, the condition of Eq. (10.3) is not only necessary, but also sufficient to produce copies of $|\Psi'\rangle$, up to an error that vanishes in the limit.

This means that, in the asymptotic setting, the von Neumann entropy is the unique measure of entanglement of pure bipartite states! This fact has profound implications. To get a feeling of them, let us consider two special cases

1. **Entanglement distillation.** Suppose that $|\Psi'\rangle = \frac{|I\rangle\rangle}{\sqrt{2}}$. This means that we want to know how many qubit Bell states can be produced from N copies of $|\Psi\rangle$. Since $S(\frac{I}{2}) = 1$ (taking the logarithm in base two), for large N the anwer is $N_{dist} \approx N\ S(\rho)$. The quantity N_{dist} is the amount of Bell states that can be extracted from $|\Psi\rangle^{\otimes N}$.

2. **Entanglement dilution.** Suppose that $|\Psi\rangle = \frac{|I\rangle\rangle}{\sqrt{2}}$. This means that we want to know how many qubit Bell states are needed to produce M copies of $|\Psi'\rangle$. For large M, the answer is $M_{dil} \approx M\ S(\rho')$. The quantity M_{dil} is the "cost" of preparing $|\Psi'\rangle^{\otimes M}$ using qubit Bell states.

Putting together 1. and 2. we have that, in the asymptotic scenario, every state $|\Psi\rangle$ can be **reversibly converted** into the Bell state $\frac{|I\rangle\rangle}{\sqrt{2}}$ at a rate given by the von Neumann entropy $S(\rho)$. In summary, the Bell state can be used as the **standard unit of entanglement**, like the kilogram for the mass or the meter for the length. This is why the Bell state $|\Phi^+\rangle = \frac{|I\rangle\rangle}{\sqrt{2}}$ is called **ebit**, in analogy to the notion of *bit* in communication theory. Having one Bell state means having one unit of entanglement.

10.9 Chapter summary

In this chapter we considered the set of transformations that can be implemented by two parties by using only local operations and classical communication (LOCC transformations). We saw that LOCC transformations cannot generate entanglement. When two parties are restricted to use LOCC transformations, entanglement becomes a resource. We then used this idea to compare

the *degree* of entanglement of pure states: if a state can be obtained from the other via a LOCC transformation, then we say that it is *less entangled*. We showed that the degree of entanglement of a pure state depends only on the spectrum of the reduced density matrix. The exact condition is given by the theory of *majorization*, which links entanglement with mixedness. Using this fact we defined many quantitative measures of entanglement and showed the surprising effect of *entanglement cathalysis*. Finally, we observed that all the measures of entanglement for pure bipartite states become equivalent when many copies of the states are available. This is a deep result, which we will be able to prove in the next chapter.

Appendix

This Appendix contains a sketch of proof of the Lo-Popescu theorem which is somehow different from the one given by Nielsen-Chuang's book. The proof is based on a simple fact: if Alice and Bob have two systems in a pure bipartite state, they can swap the two systems by local operations. Precisely, consider a generic state $|\Psi\rangle = \sum_{i=1}^{r} \sqrt{p_i}\,|\alpha_i\rangle|\beta_i\rangle$ and define the two operators

$$C = \sum_i |\beta_i\rangle\langle\alpha_i|$$
$$D = \sum_i |\alpha_i\rangle\langle\beta_i|\,.$$

Clearly one has

$$(C \otimes D)|\Psi\rangle = \sum_i \sqrt{p_i}\,|\beta_i\rangle|\alpha_i\rangle = \texttt{SWAP}|\Psi\rangle\,. \tag{10.4}$$

From the operators C and D one can construct two quantum channels (i.e. trace-preserving maps) $\mathscr{C}(\rho) = C\rho C^\dagger + \mathrm{Tr}[(I_A - C^\dagger C)\rho]\,|\alpha\rangle\langle\alpha|$ and $\mathscr{D}(\rho) = D\rho D^\dagger + \mathrm{Tr}[(I_B - D^\dagger D)\rho]\,|\beta\rangle\langle\beta|$ for two fixed states $|\alpha\rangle$ and $|\beta\rangle$. From Eq. (10.4) it follows that

$$(\mathscr{C} \otimes \mathscr{D})(\,|\Psi\rangle\langle\Psi|\,) = \texttt{SWAP}|\Psi\rangle\langle\Psi|\texttt{SWAP} \tag{10.5}$$

Note that the channels $\mathscr{C}$ and $\mathscr{D}$ depend on the state $|\Psi\rangle$: in order to swap their systems by local operations, Alice and Bob need to know which pure state they have.

The possibility of swapping pure states by local operations implies the following

Lemma 5 *The action of an N-way LOCC protocol on a pure state $|\Psi\rangle$ can be simulated by the action of a one-way LOCC protocol.*

Proof. Let us consider first the case $N = 2$. A 2-way LOCC protocol can be described as channel of the form

$$\mathscr{L}(\rho) = \sum_{i_1,i_2} \left(\mathscr{A}^{(i_1,i_2)} \otimes \mathscr{I}_B\right) \left(\mathscr{I}_A \otimes \mathscr{B}_{i_2}^{(i_1)}\right) (\mathscr{A}_{i_1} \otimes \mathscr{I}_B)(\rho)\,,$$

where $\{\mathscr{A}_{i_1}\}$ is a quantum instrument, $\{\mathscr{B}_{i_2}^{(i_1)}\}$ is a quantum instrument for every i_1, and $\mathscr{A}^{(i+1,i_2)}$ is a quantum channel for every i_1 and i_2.

Note that, without loss of generality we can choose the quantum operations $\mathscr{A}_{i_1}$ and $\mathscr{B}_{i_2}^{(i_1)}$ to have a single Kraus operator, that is

$$\mathscr{A}_{i_1}(\rho) = A_{i_1}\rho A_{i_1}^\dagger$$
$$\mathscr{B}_{i_2}^{(i_1)}(\rho) = B_{i_2}^{(i_1)}\rho B_{i_2}^{(i_1)\dagger}\,,$$

for some operators A_{i_1} and $B_{i_2}^{(i_1)}$. Now, if we apply the LOCC channel $\mathscr{L}$ to the pure state $|\Psi\rangle$ we obtain

$$\begin{aligned}\mathscr{L}(|\Psi\rangle\langle\Psi|) &= \sum_i \left(\mathscr{A}^{(i_1,i_2)} \otimes \mathscr{I}_B\right) \left(\mathscr{A}_{i_1} \otimes \mathscr{B}_{i_2}^{(i_1)}\right) (|\Psi\rangle\langle\Psi|) \\ &= \sum_i \left(\mathscr{A}^{(i_1,i_2)} \otimes \mathscr{I}_B\right) \left(\mathscr{I}_A \otimes \mathscr{B}_{i_2}^{(i_1)}\right) (|\Psi_{i_1}\rangle\langle\Psi_{i_1}|)\,,\end{aligned}$$

where $|\Psi_{i_1}\rangle$ is the unnormalized vector $|\Psi_{i_1}\rangle := (A_{i_1} \otimes I_B)|\Psi\rangle$.

Since $|\Psi_{i_1}\rangle$ is proportional to a pure state, by Eq. (10.5) there must be two channels $\mathscr{C}^{(i_1)}$ and $\mathscr{D}^{(i_1)}$ such that

$$\left(\mathscr{C}^{(i_1)} \otimes \mathscr{D}^{(i_1)}\right) (\,|\Psi_{i_1}\rangle\langle\Psi_{i_1}|\,) = \texttt{SWAP}|\Psi_{i_1}\rangle\langle\Psi_{i_1}|\texttt{SWAP}\,. \tag{10.6}$$

Similarly, since the vector $|\Psi_{i_1,i_2}\rangle := \left(I_A \otimes B_{i_2}^{(i_1)}\right)|\Psi_{i_1}\rangle = \left(A_{i_1} \otimes B_{i_2}^{(i_1)}\right)|\Psi\rangle$ is proportional to a pure state, there must be two channels $\mathscr{C}^{(i_1,i_2)}$ and $\mathscr{D}^{(i_1,i_2)}$ such that

$$\left(\mathscr{C}^{(i_1,i_2)} \otimes \mathscr{D}^{(i_1,i_2)}\right) (\,|\Psi_{i_1,i_2}\rangle\langle\Psi_{i_1,i_2}|\,) = \texttt{SWAP}|\Psi_{i_1,i_2}\rangle\langle\Psi_{i_1,i_2}|\texttt{SWAP}\,. \tag{10.7}$$

Combining Eqs. (10.6) and (10.7) we obtain

$$\begin{aligned}\left(\mathscr{I}_A \otimes \mathscr{B}_{i_2}^{(i_1)}\right) (|\Psi_{i_1}\rangle\langle\Psi_{i_1}|) &= \left[\texttt{SWAP}\left(B_{i_2}^{(i_1)} \otimes I_A\right)\texttt{SWAP}\right] |\Psi_i\rangle\langle\Psi_i| \left[\texttt{SWAP}\left(I_A \otimes B_{i_2}^{(i_1)}\right)^\dagger \texttt{SWAP}\right] \\ &= \left(\mathscr{D}^{(i_1,i_2)} \otimes \mathscr{C}^{(i_1,i_2)}\right) \left(\mathscr{B}_{i_2}^{(i_1)} \otimes \mathscr{I}_A\right) \left(\mathscr{C}^{i_1} \otimes \mathscr{D}^{i_1}\right) (|\Psi_{i_1}\rangle\langle\Psi_{i_1}|) \\ &= (\mathscr{D}_{i_1,i_2}\mathscr{B}_{i_2}^{(i_1)}\mathscr{C}_{i_1} \otimes \mathscr{C}_{i_1,i_2}\mathscr{D}_{i_1})(|\Psi_i\rangle\langle\Psi_i|) \\ &= \left(\mathscr{D}^{(i_1,i_2)}\mathscr{B}_{i_2}^{(i_1)}\mathscr{C}^{(i_1)}\mathscr{A}_{i_1} \otimes \mathscr{C}^{(i_1,i_2)}\mathscr{D}^{(i_1)}\right) (|\Psi\rangle\langle\Psi|)\end{aligned}$$

and therefore

$$\begin{aligned}\left(\mathscr{A}^{(i_1,i_2)} \otimes \mathscr{I}_B\right) \left(\mathscr{I}_A \otimes \mathscr{B}_{i_2}^{(i_1)}\right) (|\Psi_{i_1}\rangle\langle\Psi_{i_1}|) &= \left(\mathscr{A}^{(i_1,i_2)}\mathscr{D}^{(i_1,i_2)}\mathscr{B}_{i_2}^{(i_1)}\mathscr{C}^{(i_1)}\mathscr{A}_{i_1} \otimes \mathscr{C}^{(i_1,i_2)}\mathscr{D}^{(i_1)}\right) (|\Psi\rangle\langle\Psi|) \\ &= \left(\mathscr{A}_i \otimes \mathscr{B}^{(i)}\right) (|\Psi\rangle\langle\Psi|)\,, \end{aligned} \tag{10.8}$$

having set $i := (i_1, i_2)$, defined the quantum instrument [1]

$$\mathscr{A}_i := \mathscr{A}^{(i_1,i_2)}\mathscr{D}^{(i_1,i_2)}\mathscr{B}_{i_2}^{(i_1)}\mathscr{C}^{(i_1)}\mathscr{A}_{i_1}$$

and the quantum channel

$$\mathscr{B}_i := \mathscr{C}^{(i_1,i_2)}\mathscr{D}^{(i_1)}$$

[1] It is easy to check that the quantum instrument is normalized, i.e. that $\sum_i \mathscr{A}_i$ is trace-

Hence, we reduced the action of the 2-way LOCC protocol on Ψ to the action of a 1-way LOCC protocol: indeed, we have

$$\begin{aligned}\mathscr{L}(|\Psi\rangle\langle\Psi|) &= \sum_{i_1,i_2} \left(\mathscr{A}^{(i_1,i_2)} \otimes \mathscr{I}_B\right) \left(\mathscr{I}_A \otimes \mathscr{B}_{i_2}^{(i_1)}\right) (\mathscr{A}_{i_1} \otimes \mathscr{I}_B)(|\Psi\rangle\langle\Psi|) \\ &= \sum_{i_1,i_2} \left(\mathscr{A}^{(i_1,i_2)} \otimes \mathscr{I}_B\right) \left(\mathscr{I}_A \otimes \mathscr{B}_{i_2}^{(i_1)}\right) (|\Psi_{i_1}\rangle\langle\Psi_{i_1}|) \\ &= \left(\mathscr{A}_i \otimes \mathscr{B}^{(i)}\right) (|\Psi\rangle\langle\Psi|)\,,\end{aligned}$$

having used Eq. (10.8) in the last equality. The generalization to $N > 2$ is quite straightforward: we can fix the classical communication up to the step $N-2$ and consider the (unnormalized) pure state $|\Psi_{i_1,i_2,\ldots,i_{N-2}}\rangle$ that results from the first $N-2$ steps. Applying the same argument as above, we can reduce the last two rounds of the LOCC protocol to only 1 round. In this way, we reduced an N-way protocol to and $(N-1)$-way protocol. Iterating this procedure, we remain with a 1-way LOCC protocol. ■

To conclude the proof of the Lo-Popescu theorem, it only remains to show that the channel $\mathscr{B}_i$ in the proof of the previous lemma can be chosen to be unitary. But this is easy:

Lemma 6 *Every 1-way LOCC protocol that transforms a pure state $|\Psi\rangle$ into another pure state $|\Psi'\rangle$ can be reduced to a 1-way LOCC protocol where the channels performed by Bob are unitary.*

Proof. By hypothesis we have

$$\sum_i \left(\mathscr{A}_i \otimes \mathscr{B}^{(i)}\right) (\,|\Psi\rangle\langle\Psi|\,) = |\Psi'\rangle\langle\Psi'|\,,$$

---

preserving. Indeed, for every state $\rho \in \mathsf{St}(\mathcal{H}_A)$ one has

$$\begin{aligned}\mathrm{Tr}\left[\sum_i \mathscr{A}(\rho)\right] &= \sum_{i_1,i_2} \mathrm{Tr}\left[\mathscr{A}^{(i_1,i_2)}\mathscr{D}^{(i_1,i_2)}\mathscr{B}_{i_2}^{(i_1)}\mathscr{C}^{(i_1)}\mathscr{A}_{i_1}(\rho)\right] \\ &= \sum_{i_1,i_2} \mathrm{Tr}\left[\mathscr{B}_{i_2}^{(i_1)}\mathscr{C}^{(i_1)}\mathscr{A}_{i_1}(\rho)\right] \\ &= \sum_{i_1} \mathrm{Tr}\left[\left(\sum_{i_2} \mathscr{B}_{i_2}^{(i_1)}\right)\mathscr{C}^{(i_1)}\mathscr{A}_{i_1}(\rho)\right] \\ &= \sum_{i_1} \mathrm{Tr}\left[\mathscr{C}^{(i_1)}\mathscr{A}_{i_1}(\rho)\right] \\ &= \sum_{i_1} \mathrm{Tr}\left[\mathscr{A}_{i_1}(\rho)\right] \\ &= \mathrm{Tr}\left[\left(\sum_{i_1} \mathscr{A}_{i_1}\right)(\rho)\right] \\ &= \mathrm{Tr}[\rho]\,.\end{aligned}$$

or, equivalently

$$\left(\mathscr{A}_i \otimes \mathscr{B}^{(i)}\right)(\,|\Psi\rangle\langle\Psi|\,) = p_i\,|\Psi'\rangle\langle\Psi'|$$

for some probabilities $\{p_i\}$. Defining the (unnormalized) pure state $|\Psi_i\rangle := (A_i \otimes I_B)|\Psi\rangle$ we then have

$$\begin{aligned}\mathrm{Tr}_B[|\Psi_i\rangle\langle\Psi_i|] &= \mathrm{Tr}_B[(\mathscr{A}_i \otimes \mathscr{I}_B)|\Psi\rangle\langle\Psi|] \\ &= \mathrm{Tr}_B[\left(\mathscr{A}_i \otimes \mathscr{B}^{(i)}\right)|\Psi\rangle\langle\Psi|] \\ &= p_i\,\mathrm{Tr}_B[|\Psi'\rangle\langle\Psi'|]\,.\end{aligned}$$

In other words, the vector $|\Psi_i\rangle$ is proportional to a purification of $|\Psi'\rangle$. Hence, by the uniqueness of the purification, there must be a unitary gate $U^{(i)}$ such that $\left(I_A \otimes U^{(i)}\right)|\Psi_i\rangle \propto |\Psi'\rangle$. We conclude that

$$\begin{aligned}\left(\mathscr{A}_i \otimes \mathscr{B}^{(i)}\right)(\,|\Psi\rangle\langle\Psi|\,) &= p_i\,|\Psi'\rangle\langle\Psi'| \\ &= \left(\mathscr{I}_A \otimes \mathscr{U}^{(i)}\right)(\,|\Psi_i\rangle\langle\Psi_i|\,) \qquad \mathscr{U}^{(i)}(\rho) := U^{(i)}\rho U^{(i)\dagger} \\ &= \left(\mathscr{A}_i \otimes \mathscr{U}^{(i)}\right)(\,|\Psi\rangle\langle\Psi|\,)\end{aligned}$$

In other words, the quantum channel $\mathscr{B}^{(i)}$ can be replaced by a unitary channel $\mathscr{U}^{(i)}$. Summing over i we obtain

$$\begin{aligned}|\Psi'\rangle\langle\Psi'| &= \sum_i \left(\mathscr{A}_i \otimes \mathscr{B}^{(i)}\right)(\,|\Psi\rangle\langle\Psi|\,) \\ &= \sum_i \left(\mathscr{A}_i \otimes \mathscr{U}^{(i)}\right)(\,|\Psi\rangle\langle\Psi|\,)\,.\end{aligned}$$

In conclusion, the state $|\Psi\rangle$ can be transformed into the state $|\Psi'\rangle$ by a 1-way LOCC protocol with a unitary channel on Bob's side. ■

Chapter 11

Quantum Data Compression

This chapter contains a short introduction to the topic of "Quantum Shannon Theory", the quantum generalization of the mathematical theory of information developed by Claude Shannon starting from 1948.

What is special about the quantum theory of information is the relation between the transmission of quantum data and the transmission of entanglement. In this chapter we will discuss this relation in the simplest possible case: the compression of the information contained in a random source of quantum states. The techniques developed for quantum data compression will be then used to study the tasks of entanglement dilution and entanglement distillation, which we introduced in the previous chapter.

11.1 A quantum delivery service

Imagine that, in the future, a shipping company offers a delivery service for quantum systems. By paying a fee to the company, Alice can ship a quantum system to Bob, with the guarantee that the state of the system will be unchanged during the transmission.

For example, the company can provide the desired service in the following way: at Alice's location, they encode the state of Alice's system in the polarization state of N photons, send the photons to Bob's location, and then decode the state of the N photons, transferring it back to a system of the same type of Alice's system. Of course, the company has to pay a lot of care in making sure that the state of the photons is not altered during the transmission: to protect the state of the photons from noise, they will have to use all sorts of error correction tricks, which will require a quite sophisticated quantum technology. Clearly, sending quantum systems must be an expensive service. We can imagine that the company will charge a certain amount of money to Alice for every photon that they have to send: if encoding a state of Alice's system

requires N photons, then Alice will have to pay N coins. In other words, this means that if she wants to send to Bob a quantum system of dimension d, she will have to pay $\lceil \log d \rceil$ coins, where log is the logarithm in base 2.

Clearly, if Alice and Bob want to send quantum data to one another, they will try to pay the minimum amount of money possible. In other words, they will try to encode their data in a quantum system of the smallest possible dimension. In the next paragraph, we will see how can they achieve this goal.

11.2 Compressing data vs compressing entanglement

In order to pay less money to the shipping company, Alice and Bob can use a protocol of this form:

1. Alice encodes the state of her system A into the state of quantum system B of dimension $d_B < d_A$, using a suitable encoding channel $\mathscr{E}$.

2. Alice sends system B to Bob using the delivery service, at the expense of $\lceil \log d_B \rceil$ coins.

3. Bob decodes the original state of Alice's system using a decoding channel $\mathscr{D}$, which transforms states of system B back into states of system A.

A protocol of this form is called a *quantum data compression* protocol, because it compresses the states of system A into a smaller system B.

Now the question is, how good is the protocol? How much of the original quantum data will reach Bob? The answer to these questions depends on what are the possible states that carry the original data. Suppose that Alice's system can be in a set of possible states $\{|\varphi_x\rangle\}_{x\in\mathbf{X}}$, and that p_x is the probability that the system is in the state $|\varphi_x\rangle$. In the compression protocol, Alice's encoding transforms the initial state $\rho_x = |\varphi_x\rangle\langle\varphi_x|$ into the state $\rho'_x := \mathscr{E}(|\varphi_x\rangle\langle\varphi_x|)$, while Bob's decoding transforms this state into $\rho''_x := \mathscr{D}\mathscr{E}(|\varphi_x\rangle\langle\varphi_x|)$. In order to measure the quality of the transmission, we can use the fidelity between ρ_x and ρ''_x, which is given by

$$F(\rho_x, \rho''_x) = \langle\varphi_x| \mathscr{D}\mathscr{E}(|\varphi_x\rangle\langle\varphi_x|) |\varphi_x\rangle .$$

Averaging over all possible states, the fidelity becomes

$$F = \sum_x p_x \langle\varphi_x| \mathscr{D}\mathscr{E}(|\varphi_x\rangle\langle\varphi_x|) |\varphi_x\rangle . \tag{11.1}$$

A fidelity close to 1 indicates that the quality of the transmission is very good, while a fidelity close to 0 indicates that the quality of the transmission is very bad. In general, Alice and Bob will consider themselves happy if the fidelity is lower bounded as $F \geq 1 - \epsilon$, where $\epsilon > 0$ is a small error. Intuitively, if some input states have a very small probability to occur, then Alice and Bob do not

need to worry about these states, because anyways they do not contribute much to the fidelity. This is the reason why Alice and Bob can expect to save money in the transmission and still have a fidelity close to 1.

OK, now Alice and Bob know what they have to do: for a fixed dimension d_B, they have to find the best encoding and decoding channels in order to maximize the fidelity of Eq. (11.1). However, solving this problem is not so easy. Instead of solving the problem directly, Alice and Bob can take advantage of a fundamental connection between the task of compressing quantum data and the task of *compressing quantum entanglement*.

The problem of compressing entanglement is the following: Alice has two systems A and R in some entangled state $|\Psi\rangle_{AR}$ and wants to send system A to Bob. Again we can imagine that sending quantum systems has a cost, and, in order to reduce the cost, Alice and Bob can use the following protocol:

1. Alice encodes system A into a smaller system B, using an encoding channel $\mathscr{E}$.

2. She sends system B to Bob using the delivery service.

3. Bob applies a decoding channel $\mathscr{D}$ to system B, transforming it back into system A.

In the encoding step, Alice transforms the state $\rho_{AR} := |\Psi\rangle\langle\Psi|$ into the state $\rho'_{BR} := (\mathscr{E} \otimes \mathscr{I}_R)(|\Psi\rangle\langle\Psi|)$, while in the decoding step, Bob transforms the state ρ'_{BR} into the state

$$\rho''_{AR} := (\mathscr{D}\mathscr{E} \otimes \mathscr{I}_R)(|\Psi\rangle\langle\Psi|)\,.$$

The quality of the transmission can be measured by the fidelity between ρ_{AR} and ρ''_{AR}, given by

$$F_{ent} := \langle\Psi|\,(\mathscr{D}\mathscr{E} \otimes \mathscr{I}_R)(|\Psi\rangle\langle\Psi|)\,|\Psi\rangle \tag{11.2}$$

As you can see, the tasks of data compression and entanglement compression are essentially the same. The only difference is that in the case of data compression Alice has *a set of pure states* of system A, while in the case of entanglement compression she has just *one bipartite state* of the composite system AR. We will now see a fundamental connection between the two tasks.

Consider the problem of data compression with a set of states $\{|\varphi_x\rangle\}_{x\in\mathbf{X}}$, given with probabilities $\{p_x\}$ and consider the purification

$$|\Psi\rangle := \sum_x \sqrt{p_x}\,|\varphi_x\rangle|x\rangle\,.$$

With this setting we have the following lower bound:

Proposition 8 *The data compression fidelity is lower bounded by the entanglement fidelity, i.e.*

$$F \geq F_{ent}\,.$$

Proof. By definition, the entanglement fidelity is given by

$$\begin{aligned} F_{ent} &= \sum_{x,y} p_x p_y \, \langle\varphi_x| \, \mathscr{D}\mathscr{E}(|\varphi_x\rangle\langle\varphi_y|) \, |\varphi_y\rangle \\ &\leq \sum_{x,y} p_x p_y \, | \, \langle\varphi_x| \, \mathscr{D}\mathscr{E}(|\varphi_x\rangle\langle\varphi_y|) \, |\varphi_y\rangle \, | \, . \end{aligned}$$

Now, using the Kraus representation and the Schwarz inequality, it is easy to show that, for every quantum channel $\mathscr{C}$ and for every pair of vectors $|\alpha\rangle$ and $|\beta\rangle$, one has

$$|\langle\alpha|\mathscr{C}(|\alpha\rangle\langle\beta|)|\beta\rangle| \leq \sqrt{\langle\alpha|\mathscr{C}(|\alpha\rangle\langle\alpha|)|\alpha\rangle \, \langle\beta|\mathscr{C}(|\beta\rangle\langle\beta|)|\beta\rangle} \, .$$

Inserting this inequality in the bound for F_{ent}, we obtain

$$F_{ent} \leq \left(\sum_x p_x \, \sqrt{\langle\varphi_x| \, \mathscr{D}\mathscr{E}(|\varphi_x\rangle\langle\varphi_x|) \, |\varphi_x\rangle} \right)^2 \, ,$$

and, using the concavity of the function $f(t) = \sqrt{t}$,

$$\begin{aligned} F_{ent} &\leq \left(\sqrt{\sum_x p_x \, \langle\varphi_x| \, \mathscr{D}\mathscr{E}(|\varphi_x\rangle\langle\varphi_x|) \, |\varphi_x\rangle} \right)^2 \\ &\leq \sum_x p_x \, \langle\varphi_x| \, \mathscr{D}\mathscr{E}(|\varphi_x\rangle\langle\varphi_x|) \, |\varphi_x\rangle \\ &\equiv F \, . \end{aligned}$$

■

In other words, if Alice and Bob have a good protocol for entanglement compression, then they also have a good protocol for data compression. In the next paragraph we will use this fact to design some good data compression protocols.

11.3 Subspace encodings

The easiest example of quantum data compression protocol is the following: Alice chooses a proper subspace of $\mathcal{H}_A$, say $\mathcal{S} \subset \mathcal{H}_A$, and decides to encode perfectly only the density matrices that have support inside this subspace. For the states that are orthogonal to the subspace $\mathcal{S}$, she will send to Bob a failure message, informing him that in this case he should not expect any useful data. Mathematically, this procedure is described by the encoding channel

$$\mathscr{E}(\rho) = P\rho P + \mathrm{Tr}[(I - P)\rho] \, |\varphi_0\rangle\langle\varphi_0| \, , \tag{11.3}$$

where P is the projector on $\mathcal{S}$ and $|\varphi_0\rangle$ is a state in the orthogonal complement of $\mathcal{S}$. In this way, the density matrix ρ is squeezed inside a subspace of dimension $d_B = \dim \mathcal{S} + 1$ [1].

In general, the dimension d_B will be smaller than the dimension of A. Let us see more closely what is the effect of the encoding channel on the quantum states of system A:

- for a density matrix ρ with support inside the subspace $\mathcal{S}$, we get $\mathscr{E}(\rho) = \rho$: the state is encoded without any error
- for a density matrix ρ with support in the orthogonal complement of $\mathcal{S}$, we get $\mathscr{E}(\rho) = |\varphi_0\rangle\langle\varphi_0|$: the information contained in the state has been completely erased.
- in general, for a density matrix that has neither support in $\mathcal{S}$, nor in its orthogonal complement, the encoding will introduce some error, which will be as bigger as $\mathrm{Tr}[(I-P)\rho]$ is larger.

For the encoding of Eq. (11.3), Bob can just use the trivial decoding operation $\mathscr{D} = \mathscr{I}_A$ [2]. After all, this would ensure that the state is decoded correctly at least for the states with support inside $\mathcal{S}$.

11.4 Finding good subspaces

How much money can Alice and Bob save if they use a subspace encoding like the one in Eq. (11.3)? And how can they find a good subspace for their encoding?

The answer is easy: essentially, how good is a subspace depends only on how large is the probability that the input states are found in that subspace. Precisely, consider the POVM let $\{P_{yes}, P_{no}\}$, defined by $P_{yes} := P$ and $P_{no} := I - P$. This POVM can be used to distinguish between states that have support inside the subspace $\mathcal{S}$ (outcome *yes*) and states that have support in the orthogonal complement of $\mathcal{S}$ (outcome *no*). Now, the probability of the outcome *yes* is

$$\begin{aligned} p_{yes} &:= \sum_x p_x \langle\varphi_x|P|\varphi_x\rangle \\ &= \mathrm{Tr}[P\rho]\,, \end{aligned}$$

where $\rho := \sum_x p_x |\varphi_x\rangle\langle\varphi_x|$ is the average state of Alice's system.

The following result tells us that $\mathcal{S}$ is a good subspace if the probability p_{yes} is close to 1:

Proposition 9 *Consider a compression protocol with encoding as in Eq. (11.3) and with trivial decoding. Then, the entanglement fidelity of the protocol is lower bounded as* $F_{ent} \geq p_{yes}^2$.

[1] With a little abuse of notation, we can think of channel $\mathscr{E}$ as a channel from system A to the quantum system B with Hilbert space $\mathcal{H}_B = \mathsf{Span}\{\mathcal{S}, |\varphi_0\rangle\}$.

[2] More precisely, Bob's decoding consists in encode back the states of system B into the subspace $\mathcal{H}_B \subset \mathcal{H}_A$.

Proof. By definition, the entanglement fidelity is equal to

$$\begin{aligned} F_{ent} &= \langle\Psi| \left(\mathscr{D}\mathscr{E} \otimes \mathscr{I}_R\right)(|\Psi\rangle\langle\Psi|) |\Psi\rangle \\ &\geq |\langle\Psi|(P \otimes I_R)|\Psi\rangle|^2 \\ &= \left|\sum_x p_x \langle\varphi_x|P|\varphi_x\rangle\right|^2 \\ &\equiv p_{yes}^2 \,. \end{aligned}$$

■

In other words, every high-probability subspace is a good subspace for entanglement compression, and, therefore, for data compression. In particular, if the subspace satisfies $p_{yes} \geq 1 - \epsilon$, then we have

$$F \geq F_{ent} \geq p_{yes}^2 \geq 1 - 2\epsilon \,. \tag{11.4}$$

This inequality has a very important consequence: for a high-probability subspace, the corresponding compression protocol works equally well for all sets of states $\{|\varphi_x\rangle\}$ and probabilities $\{p_x\}$ such that $\sum_x p_x |\varphi_x\rangle\langle\varphi_x| = \rho$. Here is the magic of quantum compression: Alice does not even need to know which states she is sending to Bob, she only needs to know what is the average state! For this reason, we will talk about *compressing the state* ρ, meaning that we will consider compression protocols that work for arbitrary sets of states that average to ρ.

This observation is very useful to construct compression protocols. Let $\rho = \sum_n q_n |\psi_n\rangle\langle\psi_n|$ be a diagonalization of ρ, with the eigenvalues arranged in decreasing order

$$q_1 \geq q_2 \geq \cdots \geq q_{d_A} \,.$$

Clearly, if Alice and Bob want to compress the data in a quantum system of dimension d_B, they can maximize the probability $p_{yes} = \mathrm{Tr}[P\rho]$ by choosing the subspace spanned by the first d_B eigenvalues.

For example, the state

$$\rho = (1-\epsilon)\frac{|0\rangle\langle 0| + |1\rangle\langle 1|}{2} + \frac{\epsilon}{d_A - 2}\sum_{n=2}^{d_A} |n\rangle\langle n|$$

can be compressed into a two-dimensional quantum system, with a fidelity lower bounded by $F \leq 1 - 2\epsilon$.

11.5 Quantum compression in the asymptotic scenario

Suppose that Alice and Bob use a subspace encoding to compress the state $\rho^{\otimes N}$, where N is some integer number. To those purpose, they can use a high-probability subspace $\mathscr{S}_N \subset \mathcal{H}_A^{\otimes N}$, corresponding to a projector P_N. Now the

question is: how much can they compress the state $\rho^{\otimes N}$ in the limit of N going to infinity?

For every N, denote by d_N the dimension of the subspace $\mathcal{S}_N$. Then, the cost of the protocol will be $\log d_N$. We define the *compression rate* to be the unit cost

$$R := \limsup_{N\to\infty} \frac{\log d_N}{N} .$$

Note that the rate is the property of the sequence of subspaces $\{\mathcal{S}_N\}_{N\in\mathbb{N}}$ used by Alice and Bob.

Of course, Alice and Bob also want the error to be small: ideally, when N is infinite, they want the error to be zero, or equivalently, the fidelity to be 1. This requirement motivates the definition of *achievable rate*:

Definition 14 *A rate R is* achievable *if there exists a sequence of coding protocols with rate R such that*

$$\liminf_{N\to\infty} F_N = 1 ,$$

where F_N is the fidelity of the N-th compression protocol.

Now, we know that the good subspaces for compression are the maximum probability subspaces for the average state. In other words, we need to find a sequence of subspaces such that

$$\liminf_{N\to\infty} \mathrm{Tr}\left[P_N \rho^{\otimes N}\right] = 1 .$$

Once this condition is satisfied, we can have entanglement compression and data compression for every set of quantum states that average to $\rho^{\otimes N}$.

In order to find a good sequence of subspaces, it is enough to diagonalize ρ, writing

$$\rho = \sum_{m=1}^{d_A} q(m)\, |\psi_m\rangle\langle\psi_m| .$$

Clearly, a diagonalization of $\rho^{\otimes N}$ is given by

$$\rho^{\otimes N} = \sum_{\mathbf{m}} q_N(\mathbf{m})\, |\psi_{\mathbf{m}}\rangle\langle\psi_{\mathbf{m}}| ,$$

where

- $\mathbf{m}$ is the sequence $\mathbf{m} = (m_1, \ldots, m_N) \in \{1, \ldots, d_A\}^{\times N}$
- $q(\mathbf{m})$ is the probability $q_N(\mathbf{m}) = q(m_1)\, q(m_2) \cdots q(m_N)$
- $|\psi_{\mathbf{m}}\rangle$ is the product vector $|\psi_{\mathbf{m}}\rangle := |\psi_{m_1}\rangle|\psi_{m_2}\rangle \cdots |\psi_{m_N}\rangle$.

For large N, the high-probability eigenspaces of $\rho^{\otimes N}$ are determined by the theory of types. A little review of the main results on this topic is presented in the next paragraph.

11.6 A quick summary about types and probabilities

For a given sequence $\mathbf{m} = (m_1, \ldots, m_N)$, consider the number of times that the value m appears in the sequence, namely

$$N_m := |\{k \in \{1, \ldots, N\} \mid m_k = m\}| \, .$$

Definition 15 (Type of a sequence) *The* type *of the sequence* $\mathbf{m} = (m_1, \ldots, m_N)$ *is the probability distribution* $t_{\mathbf{m}}$ *defined by* $t_{\mathbf{m}} := (N_1/N, \ldots, N_{d_A}/N)$.

With this notation, the probability of the sequence can be rewritten as

$$q_N(\mathbf{m}) = \prod_{m=1}^{d_A} [q(m)]^{N_m}$$

Clearly, the probability of a sequence $\mathbf{m}$ depends only on its type $t_{\mathbf{m}}$. The property that all sequences of the same type have the same probability is called *equipartition*.

In the following, we summarize the most important facts about types and about their probabilities. For two functions of N, we will use the notation $f(N) \approx g(N)$ to mean that f and g are equivalent up to polynomials in N: that is, if there exists two polynomials $poly_1(N)$ and $poly_2(N)$ such that

$$f(N) \leq poly_1(N)\, g(N) \, ,$$

(denoted as $f(N) \lesssim g(N)$) and

$$g(N) \leq poly_2(N)\, f(N) \, .$$

Theorem 22 (Facts about types) *The main facts about types are summarized in the following:*

1. *the total number of types, denoted by* T_N*, is a polynomial in* N*, namely* $T_N \approx 1$ [3]

2. *the number of sequences of type* t*, denoted by* $S_{N,t}$*, satisfies*

$$S_{N,t} \approx \exp[NH(t)] \, ,$$

where $H(t) := -\sum_{m=1}^{d_A} t(m) \log t(m)$ *is the Shannon entropy*

[3] Precisely, the number of types is the number of partitions of N into d_A non-negative integers, and is given by $T_N = \binom{N + d_A - 1}{d_A - 1}$

3. *the probability that a sequence is of type t, denoted by $Q_{N,t}$, satisfies*

$$Q_{N,t} \approx \exp[-ND(t||q)]\,,$$

where $D(t||p)$ is the Kullback-Leibler divergence $D(t||q) := \sum_{m=1}^{d_A} t(m) \log \frac{t(m)}{q(m)}$.

The facts summarized in the above theorem come from a few elementary arguments of combinatorics and probability theory.

Another important piece of information, that will be used in the next section, are the properties of the Kullback-Leibler divergence:

Theorem 23 *The Kulback-Leibler divergence satisfies the following properties*

1. *for every two probability distributions t and p, $D(t||q) \geq 0$, and the equality holds iff $t = q$*

2. *if $\lim_{N\to\infty} D(t_N||q) = 0$, then $\lim_{N\to\infty} H(t_N) = H(q)$*

Putting together the two results above, we have that the probability that a sequence is of a type t close to p tends to one in the limit $N \to \infty$:

Corollary 6 *Let $\{\epsilon_N\}$ be a sequence such that $\lim_{N\to\infty} N\epsilon_N = \infty$. The probability that a sequence is of type t with $D(t||q) < \epsilon_N$, denoted by Q_N, satisfies*

$$\lim_{N\to\infty} Q_N = 1\,.$$

Proof. By definition, we have

$$\begin{aligned} 1 - Q_N &= \sum_{t:D(t||p)\geq\epsilon_N} P_{N,t} \\ &\lesssim \exp[-N\epsilon_N] \left(\sum_{t:D(t||q)\geq\epsilon_N} 1 \right) \\ &\leq \exp[-N\epsilon_N]\; T_N \\ &\approx \exp[-N\epsilon_N] \\ &\to 0\,, \end{aligned}$$

where in the fourth line we used the fact that the total number of types is a polynomial ($T_N \approx 1$). ■

In other words, this is the mathematical proof that, if we toss an unbiased coin a large number of times, the number of times the coin gives head and the number of times it gives tails will be nearly the same with very high probability.

Another easy corollary is that the number of sequences of type t close to q, scales like $\exp[NH(q)]$. Precisely, one has the following

Corollary 7 *Let $\{\epsilon_N\}$ be a sequence such that $\lim_{N\to\infty} \epsilon_N = 0$. Then, the number of sequences of type t satisfying $D(t||q) < \epsilon_N$, denoted by S_N, satisfies*

$$\lim_{N\to\infty} \frac{\log S_N}{N} = H(q)\,.$$

Proof. Among the types t such that $D(t||q) < \epsilon_N$, pick the one that maximizes the number $S_{N,t}$ and call it t_N^*. By definition, we have the bounds

$$S_N = \sum_{t:D(t||q)<\epsilon_N} S_{N,t} \geq S_{N,t_N^*}$$

and

$$S_N = \sum_{t:D(t||q)<\epsilon_N} S_{N,t} \leq S_{N,t_N^*} \left(\sum_{t:D(t||q)<\epsilon_N} 1 \right) \leq S_{N,t_N^*}\ T_N \approx S_{N,t_N^*}\,.$$

The two bounds imply that we have

$$S_N \approx S_{N,t_N^*} \approx \exp[NH(t_N^*)]\,.$$

Taking the logarithm and the limit for $N \to \infty$, we obtain

$$\begin{aligned} \lim_{N\to\infty} \frac{\log S_N}{N} &= \lim_{N\to\infty} \frac{\log\{\exp[NH(t_N^*)]\}}{N} \\ &= \lim_{N\to\infty} H(t_N^*) \\ &= H(q)\,. \end{aligned}$$

■

In summary, we proved that in the limit of large N, a sequence is of type close to q with high probability and the number of sequences of type close to q scales like $\exp[NH(q)]$.

11.7 Schumacher's compression theorem

We will now see a quantum protocol, developed by Benjamin Schumacher, that is able to compress quantum information at rate $S(\rho)$, where $S(\rho)$ is the von Neumann entropy

$$S(\rho) = -\operatorname{Tr}[\rho \log \rho]\,.$$

The special feature of quantum compression is that *Alice and Bob do not need to know the states $\{|\varphi_x\rangle \mid x \in \mathsf{X}\}$* but only the density matrix ρ: there is a coding protocol that works *for every set of states $\{|\varphi_x\rangle\}$ and for every set of probabilities $\{p_x\}$ such that $\sum_{x\in\mathbf{X}} p_x\ |\varphi_x\rangle\langle\varphi_x| = \rho$.*

Theorem 24 (Schumacher's theorem: direct part) *Let $\rho \in \mathsf{St}(\mathcal{H})$ be a density matrix. Then, every compression rate $R \geq S(\rho)$ is achievable.*

Proof. To achieve rate equal to the entropy, one can choose subspaces spanned by the eigenvectors of $\rho^{\otimes N}$ corresponding to types that are close to the the probability distribution of the eigenvalues of ρ. Precisely, let us define

$$\mathscr{S}_N := \mathsf{Span}\left\{|\psi_{\mathbf{m}}\rangle \mid D(t_{\mathbf{m}}||p) < \epsilon_N\right\} ,$$

for some $\epsilon_N > 0$ and let P_N be the projector on $\mathcal{S}_N$. By definition we have that

$$\begin{aligned} p_{yes,N} &:= \mathrm{Tr}[P_N \rho] \\ &= \sum_{\mathbf{m}:\ D(t_{\mathbf{m}}||p)<\epsilon_N} q(\mathbf{m}) \\ &\equiv Q_N \end{aligned}$$

As we know from the previous paragraph, choosing ϵ_N such that $\lim_{N\to\infty} N\epsilon_N = \infty$ we obtain that $\lim_{N\to\infty} p_{yes,N} = 1$. The relation $F_N \geq p^2_{yes,N}$, then implies that the compression is reliable, namely

$$\lim_{N\to\infty} F_N = 1\,.$$

On the other hand, the dimension of the subspace $\mathcal{S}_N$ is equal to the number of sequences of type close to the probability distribution:

$$\begin{aligned} d_N &\equiv \mathrm{Tr}[P_N] \\ &= \sum_{\mathbf{m}:D(t_{\mathbf{m}}||q)<\epsilon} 1 \\ &\equiv S_N . \end{aligned}$$

If in addition we choose ϵ_N such that $\lim_{N\to\infty} \epsilon_N = 0$ [4], then we obtain the rate

$$\begin{aligned} R &:= \lim_{N\to\infty} \frac{\log d_N}{N} \\ &= \lim_{N\to\infty} \frac{\log S_N}{N} \\ &= H(q) \\ &= S(\rho)\,. \end{aligned}$$

The last equality uses the fact that the Shannon entropy of the probability distribution of the eigenvalues is equal to the von Neumann entropy of the quantum state ρ. In summary, we constructed a reliable compression protocol with rate $S(\rho)$. ■

The compression protocol given in the proof of Schumacher's theorem tells us that asymptotically we can compress the state $\rho^{\otimes N}$ using NR qubits, where

[4] One possible choice satisfying both conditions is $\epsilon_N = N^{-\alpha}$ for some $0 < \alpha < 1$.

R is any number satisfying $R > S(\rho)$ [5]. Now you may ask the question: Is it possible to find some compression protocol that is reliable but has a rate smaller than $S(\rho)$?

The answer is *no*: even if we consider general encoding and decoding channels there is no way to go below the value $R = S(\rho)$:

Theorem 25 (Schumacher's theorem: strong converse) *Let*

$$\rho = \sum_m q(m) \, |\psi_m\rangle\langle\psi_m|$$

be a diagonalization of ρ and let F_N be the fidelity of data compression for the states $\{|\psi_m\rangle\}$ with probabilities $\{q(m)\}$. Then, for every compression rate $R < S(\rho)$ one has $\lim_{N\to 1} F_N = 0$.

The proof is provided in the Appendix.

Schumacher's theorem can be reformulated in terms of number of qubits needed to encode reliably the quantum information contained in a source:

Corollary 8 (Schumacher's theorem, in terms of qubits) *The information contained in N uses of a quantum source with average density matrix ρ can be transmitted reliably (i.e. without errors in the limit $N \to \infty$) using a NR qubits, where R is every rate larger than $S(\rho)$. Every attempt to transmit the information at a rate less than $S(\rho)$ will lead to failure.*

Thanks to Schumacher's theorem, we can *measure* the quantum information content of a source in qubits, in the same way we measure the mass of an object in kilograms, or the volume of a liquid in litres. Precisely, the number of qubits that are necessary for the transmission as a measure of the amount of quantum information contained in ρ.

11.8 Entanglement dilution and distillation

In this paragraph we will see an application of the techniques used for Schumacher's compression theorem to the problems of entanglement dilution and entanglement distillation, which were mentioned in the previous chapter. Entanglement dilution is the easiest application: essentially, to construct a good entanglement dilution protocol you need only to use Schumacher compression and teleportation.

Entanglement dilution. Alice and Bob have at disposal N identical pairs of qubits, each pair in the Bell state $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|0\rangle|0\rangle + |1\rangle|1\rangle)$. Using a LOCC protocol, they want to transform their pairs into M_N identical pairs of d-dimensional quantum systems, with each pair in the pure state $|\Psi\rangle$.

[5] Using only $NS(\rho)$ qubits may not be enough, because the dimension of the subspace $\mathscr{S}_N$ is not *exactly* equal to $\exp[NS(\rho)]$, but only polynomially equivalent to it.

The rate of their protocol is defined as

$$R_{\rm dil} = \liminf_{N\to\infty} \frac{M_N}{N}.$$

A rate $R_{\rm dil}$ is *achievable*, if there exists a LOCC protocol that produces an output state ρ'_N that is close to $R_{\rm dil}N$ copies of $|\Psi\rangle$, that is

$$\lim_{N\to\infty} F_N = 1$$

where $F_N := \langle\Psi|^{\otimes N}\rho'_N|\Psi\rangle^{\otimes N}$.

From quantum teleportation we know that having N Bell states is equivalent to having at disposal an perfect quantum channel sending N qubits. Combining this fact with Schumacher compression we can show that every rate

$$R_{\rm dil} < 1/S(\rho)\,, \qquad\qquad \rho = {\rm Tr}_A[|\Psi\rangle\langle\Psi|]$$

is achievable. The protocol is the following:

1. Alice prepares in her lab $N_{\rm dil}R$ pairs in the state $|\Psi\rangle$
2. She encodes half of the systems into N qubits (N qubits are sufficient because, by Schumacher's compression, the minimum number of qubits needed to compress the state $\rho^{\otimes NR_{\rm dil}}$ is $(NR_{\rm dil})S(\rho) < N$).
3. Alice uses the N Bell states to transfer the N qubits from her lab to Bob's lab via quantum teleportation
4. Bob decodes the state of the systems he received.

Using the result about compression, one can also show that rates $R_{\rm dil} > 1/S(\rho)$ cannot be achieved.

Entanglement distillation. Suppose that Alice and Bob have at disposal N identical pairs of quantum systems, each pair being in the pure bipartite state $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$. They want to convert their states into M_N Bell states $|\Phi^+\rangle \in \mathbb{C}^2 \otimes \mathbb{C}^2$. How big can M_N be?

Again, we can define the rate of a protocol as

$$R_{\rm dist} = \liminf_{N\to\infty} \frac{M_N}{N}$$

and we can look for the best achievable rate. And again, the answer is simple: any rate $R < S(\rho)$ can be achieved, where $S(\rho)$ is the von Neumann entropy of the marginal state of $|\Psi\rangle$, say, on system A [6].

One important thing that we should say here is that while entanglement dilution was essentially a corollary of Schumacher compression, entanglement distillation is a different story. We will not give the full detail of the proof,

[6] Remember that the marginal states of a bipartite state on systems A and B have the same entropy.

but it is nice to see the main idea. Essentially, the idea is to use the basic properties of types to construct an LOCC protocol that transforms $|\Psi\rangle^{\otimes N}$ into approximately $NS(\rho)$ Bell states.

Writing $|\Psi\rangle$ in the Schmidt form $|\Psi\rangle = \sum_m \sqrt{q(m)}|\alpha_m\rangle|\beta_m\rangle$, the initial state in Alice's and Bob's hands can be written as

$$|\Psi\rangle^{\otimes N} = \sum_{\mathbf{m}} \sqrt{q(\mathbf{m})}|\alpha_{\mathbf{m}}\rangle|\beta_{\mathbf{m}}\rangle\,,$$

where the states $|\alpha_{\mathbf{m}}\rangle$ and $|\beta_{\mathbf{m}}\rangle$ are defined in the same way as the states $|\psi_{\mathbf{m}}\rangle$ in Schumacher's compression. Now, define Alice's *subspace of type t* as

$$\mathcal{H}_{N,t} := \mathsf{Span}\left\{|\alpha_{\mathbf{m}}\rangle \mid t_{\mathbf{m}} = t\right\}.$$

By definition, the dimension of $\mathcal{H}_{N,t}$ is the number of sequences of type t, given by $S_{N,t} \approx \exp[NH(t)]$.

Suppose that Alice measures her systems with the instrument $\{\mathscr{A}_t\}$, there the outcomes are labelled by types and the quantum operations are given by

$$\mathscr{A}_{N,t}(\rho) = P_{N,t}\,\rho\,P_{N,t}$$

where $P_{N,t}$ is the projector on the subspace of type t. By definition, the probability of the outcome t is equal to the probability that a sequence is of type t:

$$\begin{aligned}\mathrm{Tr}\left[(\mathscr{A}_{N,t}\otimes\mathscr{I}_B)(|\Psi\rangle\langle\Psi|)\right] &= \langle\Psi|(P_{N,t}\otimes I_B)|\Psi\rangle \\ &= \sum_{\mathbf{m}:t_{\mathbf{m}}=t} q(\mathbf{m}) \\ &\equiv Q_{N,t} \\ &\approx \exp[-ND(t||p)]\,.\end{aligned}$$

Conditionally to the outcome t, the state of Alice's and Bob's systems will be

$$|\Psi_t\rangle = \frac{1}{\sqrt{S_{N,t}}} \sum_{\mathbf{m}\in S_{N,t}} |\alpha_{\mathbf{m}}\rangle|\alpha_{\mathbf{m}}\rangle.$$

Now, using the majorization criterion it is easy to see that this state is more entangled than the state $|\Phi^+\rangle^{\otimes M_t}$, where $M_t = \lfloor \log S_{N,t} \rfloor$.

The idea now is that with large probability we will obtain a type with entropy $H(t)$ close to the Shannon entropy $H(q)$, which is equal to the von Neumann entropy $S(\rho)$. Hence, at least heuristically, we can see that we can produce approximately $NS(\rho)$ Bell pairs with high probability.

Asymptotic transformations of pure entangled states. Putting together entanglement dilution with entanglement distillation, we obtain that, in the limit of large N, N copies of a pure bipartite state $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ can be reversibly transformed by LOCC into $NS(\rho)/S(\rho')$ copies of $|\Psi'\rangle$, where

$$\begin{aligned}\rho &:= \mathrm{Tr}_A[|\Psi\rangle\langle\Psi|] \\ \rho' &:= \mathrm{Tr}_A[|\Psi'\rangle\langle\Psi'|].\end{aligned}$$

In summary, all pure bipartite states are asymptotically equivalent under LOCC transformations, and we can transform any bipartite state in any other at a rate given by the ratio of their von Neumann entropies.

11.9 Chapter summary

In this chapter we studied the task of quantum data compression, where Alice has a state $|\varphi_x\rangle$ of system A, chosen at random with probability p_x, and wants to encode it into a smaller system B, in such a way that the encoded information can be decoded by Bob. We established a connection between data compression with the task of entanglement compression and we showed how to construct good compression protocols. The special feature of quantum compression is that Alice does not need to know which states she is trying to encode: in order to construct a good compression protocol, she only needs to know the average state $\rho = \sum_x p_x |\varphi_x\rangle\langle\varphi_x|$. In the rest of the chapter we studied the asymptotic version of quantum compression. The key result here is Schumacher's theorem, which tells us how well we can compress N uses of a quantum source with average state ρ. Precisely: *i)* every compression rate $R \geq S(\rho)$ can be achieved reliably, while *ii)* if we try to compress at a rate $R < S(\rho)$, the fidelity will drop to zero for large N. Finally, we applied the techniques developed in this chapter to study the two tasks of entanglement dilution and entanglement distillation. Using these results we established that in the asymptotic scenario all entangled pure states can be transformed into one another, at a rate equal to the ratio of the von Neumann entropies of their marginal states.

Appendix

Proof of Theorem 8.3: Let $\{\mathscr{E}_N, \mathscr{D}_N\}$ be a generic sequence of encoding and decoding channels (not necessarily subspace encodings). Diagonalizing $\rho^{\otimes N}$ as $\rho^{\otimes N} = \sum_{\mathbf{m}} q(\mathbf{m}) |\psi_{\mathbf{m}}\rangle\langle\psi_{\mathbf{m}}|$ and defining $\rho'_{\mathbf{m}} = \mathscr{E}_N(|\psi_{\mathbf{m}}\rangle\langle\psi_{\mathbf{m}}|)$, we can write the fidelity as

$$\begin{aligned} F_N &= \sum_{\mathbf{m}} q(\mathbf{m}) \langle\psi_{\mathbf{m}}|\mathscr{D}_N(\rho'_{\mathbf{m}})|\psi_{\mathbf{m}}\rangle \\ &= \sum_{\mathbf{m}} q(\mathbf{m}) \operatorname{Tr}[P_{\mathbf{m}}\rho'_{\mathbf{m}}], \end{aligned}$$

where the operators $\{P_{\mathbf{m}}\}$ is the POVM that corresponds to measuring on the basis $\{|\psi_{\mathbf{m}}\rangle\}$ after the decoding channel $\mathscr{D}_N$. In other words, the average fidelity is the average probability that we correctly identify the states $\rho'_{\mathbf{m}}$.

Now, we can regroup the sequences $\mathbf{m}$ according to their types and write

$$F_N = \sum_t Q_{N,t} \left(\sum_{\mathbf{m}: t_{\mathbf{m}} = t} \frac{\operatorname{Tr}[P_{\mathbf{m}}\rho'_{\mathbf{m}}]}{S_{N,t}} \right).$$

Each term in the round bracket is the probability that we distinguish correctly between the states $\{\rho'_{\mathbf{m}} \mid t_{\mathbf{m}} = t\}$, given with uniform prior $1/S_{N,t}$. Since for large N the sum is dominated by the types close to q, we can consider only the types such that $D(t||q) < \epsilon$

$$\begin{aligned}\lim_{N\to\infty} F_N &= \lim_{N\to\infty} \sum_{t:D(t||q<\epsilon)} Q_{N,t} \left(\sum_{\mathbf{m}:t_{\mathbf{m}}=t} \frac{\mathrm{Tr}[P_{\mathbf{m}}\rho'_{\mathbf{m}}]}{S_{N,t}} \right) \\ &\leq \lim_{N\to\infty} \sum_{t:D(t||q<\epsilon)} \left(\sum_{\mathbf{m}:t_{\mathbf{m}}=t} \frac{\mathrm{Tr}[P_{\mathbf{m}}\rho'_{\mathbf{m}}]}{S_{N,t}} \right) .\end{aligned}$$

Moreover, we know from a previous exercise of this course that if we try to distinguish between N states in dimension d, the probability of success is upper bounded by d/N. Hence, we obtain

$$\lim_{N\to\infty} F_N \leq \lim_{N\to\infty} \sum_{t:D(t||q)<\epsilon} \frac{d_N}{S_{N,t}} ,$$

where d_N is the dimension of the system used for the encoding and $S_{N,t}$ is the number of sequences of type t. We know that $S_{N,t} \approx \exp[NH(t)]$ and we know that the entropy $H(t)$ is as close to $H(q)$ as we want, provided that we choose ϵ small enough. Hence, whenever the rate satisfies $R < H(q)$, the ratio $d_N/S_{N,t}$ will go to zero exponentially fast, which implies

$$\begin{aligned}\lim_{N\to\infty} F_N &\leq \lim_{N\to\infty} T_N \max_{t:D(t||q)<\epsilon_N} \frac{d_N}{S_{N,t}} \\ &\to 0 .\end{aligned}$$

■

Part IV

Quantum computation

Chapter 12

Quantum search

In the last two chapters of the course we will see how quantum mechanics can be used to speed up calculations. This is the subject of quantum computing and, more specifically, of quantum algorithms.

We will see two quantum algorithms that allow one to speed up two important tasks:

1. Finding a marked element in a list
2. Finding the prime factors of a number.

The two algorithms are Grover's quantum search algorithm and Shor's algorithm for factoring numbers, a problem that can be reduced to the problem of finding the period of a function. Grover's and Shor's algorithm are not only important because they are "famous", but also because they can be used to illustrate the two most important tricks of quantum computation:

1. Amplitude amplification (in Grover's algorithm).
2. Phase estimation (in Shor's algorithm).

The origin of both tricks is the same: in quantum mechanics we are not forced to make computations in the computational basis, but we can encode data in other, more efficient, ways. For example, we can use the Fourier basis

$$|e_n\rangle = \frac{1}{\sqrt{d}} \sum_{k=1}^{d} \exp\left[\frac{2\pi i n k}{d}\right] |k\rangle \,.$$

As we saw at the beginning of the course, when a system is prepared in a state of the Fourier basis, the outcomes of a measurement on the computational basis are completely random:

$$p(k) = |\langle k|e_n\rangle|^2 = \frac{1}{d} \,.$$

In general, we will see that using bases that are different from the computational basis allows us to make faster computations. Essentially, quantum computations

are faster because we do not need to know the state of the computer during the computation, but instead, we only need to measure the state of the computer in the end.

12.1 A quantum game of boxes and prizes

You are given N boxes. Most boxes are empty, but a few of them, say $S \ll N$, contain a very valuable prize, like an airticket to Hawai, or the solution of your Quantum Information homeworks. You would like to find a prize, but every time you open a box you have to pay a fee. What is the minimum number of boxes that you have to open before you find a prize?

In the classical world, it is easy to see that you need to open $\Theta(N)$ boxes in order to find prize. In the worst case, this is obvious: if you are really unlucky, you have to open $N-S$ boxes before you find a prize. On average, the situation is similar: if you want to find a prize with high probability you still need to open a number of boxes that grows like N. If this is not immediately clear to you, try the following exercise:

Exercise 27 Suppose that you open k boxes. Show that

1. the probability that you find at least one box containing a prize is equal to $p_k = 1 - \dfrac{\binom{N-s}{k}}{\binom{N}{k}}$
2. if S is fixed and $k = k(N)$ is a function satisfying $\lim_{N\to\infty} k(N)/N = 0$, then $\lim_{N\to\infty} p_{k(N)} = 0$.

Let us make a mathematical model of our game of boxes. First of all, the boxes can be labelled with numbers from 1 to N and we can imagine that each box can have two states 0 and 1, where 0 describes an empty box and 1 describes a box containing the prize. When we open a box, we are actually evaluating a function:

$$f : \{1, \ldots, N\} \mapsto \{0, 1\}.$$

Precisely, the value of the function $f(n)$ tells us whether the n-th box is empty $[f(n) = 0]$ or whether there is a prize $[f(n) = 1]$.

What about the quantum version of this game? We can model the boxes as N qubits, so that an empty box is in the state $|0\rangle$ and a box with the prize is in the state $|1\rangle$. Let is denote by A the system of N qubits and by $|\alpha\rangle \in \left(\mathbb{C}^2\right)^{\otimes N}$ the state of the system. By definition we have

$$|\alpha\rangle = |f(1)\rangle\, |f(2)\rangle \cdots |f(N)\rangle\,.$$

Now, "opening the n-th box" means measuring the n-th qubit in the computational basis. However, in the quantum world we can do many other things on

a box, other than "opening" it. For example, we can apply unitary gates, such as the Pauli gate Z, whose action in the computational basis is given by

$$Z|0\rangle = |0\rangle$$
$$Z|1\rangle = -|1\rangle\,.$$

Let us denote by Z_n the Pauli gate Z acting on the n-th qubit. Then we have

$$Z_n|\alpha\rangle = (-1)^{f(n)}\,|\alpha\rangle\,.$$

This does not seem very useful, because the density matrix of the qubits after the gate is still $|\alpha\rangle\langle\alpha|$. However, we can introduce an extra system B, of dimension N, and consider the control-unitary gate

$$U = \sum_{n=1}^{N} Z_n \otimes |n\rangle\langle n|\,.$$

The gate U represents the action of a programmable machine that, depending on the state of system B, applies the gate Z to one of the N qubits:

$$U|\alpha\rangle|n\rangle = (-1)^{f(n)}|\alpha\rangle|n\rangle\,.$$

Intuitively, one would say that the machine is only "acting one box": if we think of acting on a box as applying a gate on the corresponding qubit, then the machine is only acting one box, which depends on the state of the control system B. Surprisingly, in the next paragraphs we will see an algorithm (*Grover's quantum search algorithm*) that allows us to find the prize using the gate U only $O(\sqrt{N})$ times!

How is this possible? The key point is that the control system B does not have to be in a state of the computational basis. If we prepare B in a pure state $|\beta\rangle = \sum_n \beta_n\,|n\rangle$ and apply the gate U, then we obtain the output state

$$\begin{aligned} U|\alpha\rangle|\beta\rangle &= \sum_n \beta_n\,(-1)^{f(n)}\,|\alpha\rangle|n\rangle \\ &= |\alpha\rangle|\beta_f\rangle \qquad |\beta_f\rangle := \sum_{n=1}^{N}(-1)^{f(n)}\,\beta_n\,|n\rangle\,. \end{aligned}$$

The information about the function f has been encoded in the state $|\beta_f\rangle$. Note also that we can express the state $|\beta_f\rangle$ as $|\beta_f\rangle = V_f|\beta\rangle$, where V_f is the unitary gate

$$V_f = \sum_{n=1}^{N}(-1)^{f(n)}\,|n\rangle\langle n|\,. \tag{12.1}$$

In other words, if we have a programmable machine that performs the Pauli gate Z on the n-th qubit depending on the state of a control system B, then

we can also implement the gate V_f. In the following we will call the gate V_f *Grover's gate*. Mathematically, the problem of quantum search is the problem of finding a solution of the equation $f(n) = 1$ using the Grover's gate V_f the minimum number of times.

It is important to note that the problem of quantum search is not exactly the same as the problem of classical search: in the quantum problem we are giving ourselves some extra-power that we did not have in the classical problem, namely the ability to operate on the boxes in a way that is controlled by the state of a quantum system—a system that can be in a superposition of different computational basis states.

12.2 Alternative formulation of the quantum search problem

Before presenting Grover's algorithm, it is good to see another way to look at the search problem. Forgetting about the game of boxes and prizes, we can model a classical search problem as follows:

- we have a black box that evaluates a function $f : \{1, \ldots, N\} \to \{0, 1\}$
- we want to find a value n such that $f(n) = 1$ using the black box the minimum number of times [1].

Since classical probability distribution can be represented as density matrices that are diagonal in the computational basis, in the classical case the black box can be described as a channel $\mathscr{C}_f$ that transforms an N-dimensional quantum system A into a 2-dimensional quantum system B [2]:

$$\mathscr{C}_f(\rho) := \sum_n \langle n|\rho|n\rangle \, |f(n)\rangle\langle f(n)| \, .$$

In the quantum world, the channel $\mathscr{C}_f$ can be realized by

1. preparing system B in the state $|0\rangle \equiv |N \mod N\rangle$
2. applying to systems A and B the gate $U_f = \sum_n |n\rangle\langle n| \otimes X^{f(n)}$
3. discarding system A.

Indeed, it is easy to check that

$$\mathscr{C}_f(\rho) = \mathrm{Tr}_A \left[U_f \ (\rho \otimes |0\rangle\langle 0|) \ U_f^\dagger \right] \, .$$

[1] Note that here we do not care about the *complexity* of computing f, but only on the *number of times* we use it. The number of uses of f needed to find a solution to the equation $f(n) = 1$ is called the *query complexity*, to distinguish it from the *gate complexity*, which is the number of elementary gates needed to find a solution and includes the number of elementary gates needed to compute f.

[2] Note that the quantum systems A and B here have nothing to do with the quantum systems A and B in the previous section.

In other words, the black box that implements the unitary gate U_f is a more powerful black box than the one that implements the channel $\mathscr{C}_f$: if someone gives you the black box U_f, you can always simulate $\mathscr{C}_f$, but if somebody gives you $\mathscr{C}_f$, there is no way to simulate U_f.

Now, suppose that suppose that, instead of the channel $\mathscr{C}_f$, we can use the unitary gate U_f The question is: how many times do we need to use the gate U_f in order to find a solution of the equation $f(n) = 1$?

Formally, searching for the desired n is equivalent to the quantum game of boxes and prizes that we discussed in the previous section. In order to see this, suppose that we apply U_f to an input of the product form $|\alpha\rangle|-\rangle$, with

$$|\alpha\rangle = \sum_n \alpha_n |n\rangle \qquad\qquad |-\rangle = \frac{|0\rangle - |1\rangle}{\sqrt{2}} .$$

Since $X|-\rangle = -|-\rangle$, we have

$$\begin{aligned} U_f|\alpha\rangle|-\rangle &= \sum_{n=1}^{N} \alpha_n (-1)^{f(n)} |n\rangle|-\rangle \\ &= |\alpha_f\rangle|-\rangle \qquad |\alpha_f\rangle = \sum_n \alpha_n \, (-1)^{f(n)} |n\rangle . \end{aligned}$$

Again, the function f has been encoded in a quantum state, in this case $|\alpha_f\rangle$. Note that the state $|\alpha_f\rangle$ can be expressed as $|\alpha_f\rangle = V_f|\alpha\rangle$, where V_f is the Grover's gate defined in Eq. (12.1).

In summary, if we have access to one use of the gate U_f, then we have also access to one use of the Grover's gate V_f (because we are free to prepare the second system in the state $|-\rangle$). Also in this formulation, the problem of quantum search is to find a solution of the equation $f(n) = 1$ using the Grover's gate V_f the minimum number of times.

12.3 The quantum search algorithm

We will now present an algorithm, invented by Lov Grover, which allows one to find a winning outcome n with high probability using the gate V_f only $O(\sqrt{N})$ times. For simplicity of presentation, we will consider the case where the number of solutions to the equation $f(n) = 1$ is small compared to N, i.e. $S \ll N$. The input system for the Grover's gate V_f will be denoted by A. With this notation, the quantum search algorithm works as follows:

1. Prepare system A in the Fourier basis state

$$|e_N\rangle = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} |n\rangle.$$

2. Apply Grover's unitary V_f (cf. "open one box", in the classical case)

3. Apply the unitary
$$W = 2|e_N\rangle\langle e_N| - I.$$

4. Repeat steps 2. and 3. for a number of times k that is the closest integer to $\frac{\pi}{4}\sqrt{\frac{N}{S}}$.

5. Measure the system on the computational basis. With probability $p_{succ} \geq 1 - \frac{S}{N}$ the outcome of the measurement is a solution of the equation $f(n) = 1$.

12.4 How the algorithm works: amplitude amplification

At first sight, the quantum search algorithm seems a little bit mysterious. Why does the algorithm work?

Here we will show a simple geometrical explanation. Note that the first step of the algorithm is pretty intuitive: since we do not know anything about the solution, we prepare an input state $|e_N\rangle$ that is an equal superposition of all solutions. Let us write the input state as

$$|e_N\rangle = \sqrt{1 - \frac{S}{N}}\,|\varphi_0\rangle + \sqrt{\frac{S}{N}}\,|\varphi_1\rangle$$

where

$$|\varphi_0\rangle := \frac{1}{\sqrt{N-S}} \sum_{n:f(n)=0} |n\rangle$$
$$|\varphi_1\rangle := \frac{1}{\sqrt{S}} \sum_{n:f(n)=1} |n\rangle$$

Hence, we can consider $|e_N\rangle$ as a unit vector in the **real plane** $\mathbb{R}^2$,

$$|e_N\rangle = \cos\theta|\varphi_0\rangle + \sin\theta|\varphi_1\rangle$$

with $\cos\theta = \sqrt{1 - \frac{S}{N}}$ and $\sin\theta = \sqrt{\frac{S}{N}} \approx \theta$.

Now, in the real plane V_f and W are two **reflections**:

- V_f is a reflection around $|\varphi_0\rangle$: indeed, one has
$$\begin{cases} V_f|\varphi_0\rangle = \ \ |\varphi_0\rangle \\ V_f|\varphi_1\rangle = -|\varphi_1\rangle \end{cases}$$

- W is a reflection around $|e_N\rangle$: indeed, one has
$$\begin{cases} W|e_N\rangle \ \ = \ \ |e_N\rangle \\ W|e_N^\perp\rangle \ = -|e_N^\perp\rangle\,, \end{cases}$$

where $|e_N^{\perp}\rangle$ is the orthogonal vector defined by

$$|e_N^{\perp}\rangle := -\sin\theta|\varphi_0\rangle + \cos\theta|\varphi_1\rangle\,.$$

Since V_f and W are **reflections**, WV_f is a **rotation**. How much is the rotation angle? The easiest way to answer the question is to follow the evolution of the state $|e_N\rangle$: choosing the x and y axes to point in the directions of the vectors $|\varphi_0\rangle$ and $|\varphi_1\rangle$, respectively, we have that

- initially $|e_n\rangle$ is at an angle θ with the x-axis
- since V_f is a reflection around the x-axis, the vector $V_f|e_N\rangle$ is at an angle $-\theta$ with the x-axis , and therefore at angle -2θ with $|e_N\rangle$
- since W is a reflection around $|e_N\rangle$, the vector $WV_f|e_N\rangle$ is at an angle 2θ with $|e_N\rangle$.

Since WV_f rotated the vector $|e_N\rangle$ by 2θ and we know that WV_f is a rotation, we conclude that WV_f is rotates every vector by 2θ.

Since WV_f is a rotation of a positive angle, the first times we apply it, we bring the state of the system closer to $|\varphi_1\rangle$: indeed, we have

$$(WV_f)^k|e_N\rangle = \cos[(2k+1)\theta]\,|\varphi_0\rangle + \sin[(2k+1)\theta]\,|\varphi_1\rangle.$$

This phenomenon is known as **amplitude amplification**: initially, at each step the coefficient in front of the state $|\varphi_1\rangle$ will be increased. Note that $[\sin(k\tau+\theta)]^2$ is the probability that, if we apply the unitary WV_f for k times and we measure the output in the computational basis, the outcome of the measurement is a solution of the equation $f(n)=1$. Indeed, denoting the probability by $p_{succ}^{(k)}$ we have

$$\begin{aligned} p_{succ}^{(k)} &= \sum_{n:f(n)=1} |\langle n|\,(WV_f)^k\,|e_N\rangle|^2 \\ &= \sum_{n:f(n)=1} [\sin(2k+1)\theta]^2\;\; |\langle n|\varphi_1\rangle|^2 \\ &= [\sin(2k+1)\theta]^2\;. \end{aligned}$$

The probability will be close to 1 when

$$(2k+1)\theta \approx \frac{\pi}{2}.$$

Using the fact that θ is small with respect to $\pi/2$, we obtain that the optimal value of k is given by

$$k_{\rm opt} \approx \frac{\pi}{4\theta} \approx \frac{\pi}{4}\sqrt{\frac{N}{S}}.$$

Without approximation, the best choice of k is to choose $k_{\rm opt}$ to be the integer that is closest to $\frac{\pi}{4\theta} - \frac{1}{2}$, with $\theta = \arcsin\sqrt{\frac{S}{N}}$. With this choice, we have

$$\begin{aligned} p_{succ}^{(k_{\rm opt})} &= [\sin(2k_{\rm opt}+1)\theta]^2 \\ &= \left[\sin\left(\frac{\pi}{2}+\epsilon\right)\right]^2 \qquad \epsilon \leq \theta \\ &\geq [\cos\theta]^2 \\ &= 1 - \frac{S}{N}. \end{aligned}$$

In summary, using the gate V_f for $O(\sqrt{N})$ times we can find a solution with a probability lower bounded by $1 - S/N$, where S is the number of solutions (in the assumption $S \ll N$).

Remark 5 (Dependence of the algorithm on the number of solutions.) In Grover's algorithm it is important to know **how many solutions** we have, or, more precisely, it is important to know S/N. This is because we need to apply the unitary WV_f for

$$k_{\rm opt} \approx \frac{\pi}{4}\sqrt{\frac{N}{S}}$$

times (if $S \ll N$). If we apply the gate more times, the probability to find a solution will start to **decrease**. This is pretty bad, because in many interesting cases, we do not know the number of solutions of the equation $f(n) = 1$. Actually, we may not even know if the equation *has* any solution! Luckily, quantum mechanics offers us one way to solve this problem. Here is the idea: we know that the gate WV_f is a rotation of an angle τ that depends on S/N [precisely $\tau = 2\arcsin\sqrt{S/N}$]. Hence, if we know the rotation angle τ we can find the number of solutions. Quantum mechanics offers a great algorithm that allows us to estimate the rotation angle, known as the *quantum phase estimation* algorithm. This algorithm, which is the core of Shor's factorization algorithm, will be explained in depth in the next chapter. For the moment, it is enough to know that the quantum phase estimation algorithm allows us to find out the rotation angle τ, with a precision that grows quickly with the number of times we use the gate V_f. More precisely, the phase estimation algorithm assumes that we are able to control the number of times we apply V_f, implementing the control-unitary unitary gate

$$T = \sum_{m=1}^{M} V_f^m \otimes |m\rangle\langle m| \,,$$

where the control is an M-dimensional quantum system. Using this gate, the angle τ can be estimated within an interval of size $1/M$. In conclusion, when the number of solutions is not know we can use first the phase estimation algorithm to estimate the number of solutions, and then Grover's algorithm to find a solution in $O(\sqrt{N})$ steps. Since we only need to know the identify the angle

with a precision $1/\sqrt{N}$, the phase estimation step does not spoil the scaling $O(\sqrt{N})$ of Grover's algorithm.

12.5 Optimality of Grover's $O(\sqrt{N})$ scaling

Grover's algorithm is just one possible algorithm to find a solution in $O(\sqrt{N})$ steps. Can we find a better algorithm, which is even faster?

In particular, can we find a solution in $O(\log N)$ steps? This would be an astonishing improvement: you would be able to solve many hard problems in polynomial time, just by searching blindly among all possible solutions.

Unfortunately, this is not the case. We can rigorously prove that $O(\sqrt{N})$ is the **best scaling allowed by quantum mechanics**. For simplicity, we will give the proof for the case where there is **only one solution** to the equation $f(n) = 1$.

Let us denote the solution by by x. In this case we have

$$\begin{aligned} V_f &= -|x\rangle\langle x| + \sum_{n \neq x} |n\rangle\langle n| \\ &= -2|x\rangle\langle x| + I \\ &=: V_x \,. \end{aligned}$$

Now, let us consider the **most general** quantum algorithm that uses the gate V_x a finite number of times k. To this purpose, we can introduce an auxiliary system with Hilbert space $\mathcal{H}_B$ of some dimension d_B (the value of d_B is not important for the proof: system B can be as big as we want). The most general quantum circuit that uses V_x for k times will be of the following form:

1. prepare a state $|\Psi_0\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$
2. apply the gate V_x on system A
3. apply a gate U_1 to systems A and B
4. at the $2n$-th step, $n = 2, \dots, k$, apply the gate V_x on system A
5. at the $(2n+1)$-step, $n = 2, \dots, k$, apply a gate U_n on systems A and B.

In formula, the output of the circuit will be

$$|\Psi_{k,x}\rangle = U_k(V_x \otimes I_B)U_{k-1} \cdots U_1(V_x \otimes I_B)|\Psi_0\rangle.$$

The goal of the circuit is to produce states $|\Psi_{k,x}\rangle$ that are almost orthogonal. Without loss of generality, this means that we want $|\Psi_{k,x}\rangle$ to be close to a state $|\Phi_{k,x}\rangle$ of the form

$$|\Phi_{k,x}\rangle = |x\rangle|\beta_{k,x}\rangle$$

where $\{|\beta_{k,x}\rangle\}$ are some pure states of system B.

For a fixed value of x, we can quantify the error as

$$\eta_{k,x} := \| |\Psi_{k,x}\rangle - |\Phi_{k,x}\rangle \|^2 .$$

This quantity tells us how far is the state that we actually get—i.e. $|\Psi_{k,x}\rangle$—from the ideal state that we would like to get—i.e. $|\Phi_{k,x}\rangle$. Since x is unknown and can take N possible values, it is good to consider the *average error*

$$\eta_k := \frac{1}{N} \sum_{s=1}^{N} \| |\Psi_{k,x}\rangle - |\Phi_{k,x}\rangle \|^2.$$

Of course, we want the error to be small. We will now see that this requires k to be of the order of $\sqrt{N}$.

Proof. The proof uses a clever trick: consider the state that we obtain by removing the black box U_s from the circuit:

$$|\Psi_k\rangle := U_k U_{k-1} \cdots U_2 U_1 |\Psi\rangle.$$

This state is "bad", because carries no information about the value of the solution.

Now, we have

$$\begin{aligned} \eta_k &= \frac{1}{N} \sum_s \| |\Psi_{k,x}\rangle - |\Psi_k\rangle + |\Psi_k\rangle - |\Phi_{k,x}\rangle \|^2 \\ &\geqslant \frac{1}{N} \sum_s \left(\underbrace{\| |\Psi_{k,x}\rangle - |\Psi_k\rangle \|}_{=a_{k,x}} - \underbrace{\| |\Psi_k\rangle - |\Phi_{k,x}\rangle \|}_{=b_{k,x}} \right)^2 . \end{aligned}$$

Defining the real vectors

$$\begin{aligned} \mathbf{a}_k &:= (a_{k,1}, \ldots, a_{k,N}) \\ \mathbf{b}_k &:= (b_{k,1}, \ldots, b_{k,N}) \end{aligned}$$

the chain of inequalities can be continued as

$$\begin{aligned} \eta_k &= \frac{1}{N} \|\mathbf{a}_k - \mathbf{b}_k\|^2 \\ &\geq \left(\frac{\|\mathbf{a}_k\| - \|\mathbf{b}_k\|}{\sqrt{N}} \right)^2 \end{aligned}$$

Now, in order for the error to be small, $\|\mathbf{a}_k\|$ should be close to $\|\mathbf{b}_k\|$, in the sense that

$$\left| \frac{\|\mathbf{a}_k\| - \|\mathbf{b}_k\|}{\sqrt{N}} \right| \ll 1.$$

Let us see what $\|\mathbf{a}_k\|$ and $\|\mathbf{b}_k\|$ are. Intuitively, the length of $\mathbf{b}_k$ tells us how far is the "bad state" $|\Psi_k\rangle$ from the "good" state $|\Phi_{k,x}\rangle$, summed over all possible

values of x. Since the "bad" state is "bad" and the "good" state is "good", it is natural to expect that the distance between them will be large. This intuition is correct: indeed, we have

$$\begin{aligned}\|\mathbf{b}_k\|^2 &= \sum_x \| |\Psi_k\rangle - |\Phi_{k,x}\rangle \|^2 \\ &= \sum_x 2\left(1 - \mathsf{Re}\langle\Psi_k|\Phi_{k,x}\rangle\right) \\ &\geqslant \sum_x 2\left(1 - |\langle\Psi_k|\Phi_{k,x}\rangle|\right) \\ &\geqslant 2\left(N - \sqrt{N\sum_x |\langle\Psi_k|\Phi_{k,x}\rangle|^2}\right) ,\end{aligned}$$

having used the Schwartz inequality for the bound $\sum_x |\langle\Psi_k|\Psi_{k,x}\rangle| \leq \sqrt{N\sum_x |\langle\Psi_k|\Phi_{k,x}\rangle|^2}$. Now, the states $|\Phi_{k,x}\rangle$ are **orthogonal** and they can be part of an orthonormal basis. Hence, we have $\sum_{x=1}^N |\langle\Psi_k|\Phi_{k,x}\rangle|^2 \leqslant 1$, which implies

$$\|\mathbf{b}_k\|^2 \geqslant 2(N - \sqrt{N}) .$$

In other words, the distance between the bad states and good states grows like N.

Now, the only way that the error can be small is that also $\|\mathbf{a}_k\|^2$ is of order N, so that $|\|\mathbf{a}_k\| - \|\mathbf{b}_k\||$ is small. We now show that this is possible only if the number of times we use V_x, given by k, is of order $\sqrt{N}$. Indeed, we can prove that $\|\mathbf{a}_k\|$ grows only linearly with k, and, precisely

$$\|\mathbf{a}_k\| \leqslant 2k .$$

Intuitively, the idea is that $\|\mathbf{a}_k\|$ quantifies the distance between the bad state and the states that we obtain from the circuit. In order to produce states that are far from the bad state we need to apply the gate V_x many times.

Let us prove the bound $\|\mathbf{a}_k\| \leq 2k$. The proof is by induction on k:

- For $k = 1$, we have $\|\mathbf{a}_1\| = \| |\Psi_{1,x}\rangle - |\Psi_1\rangle \| \leq 2$ due to the triangular inequality.

- Suppose that $\|\mathbf{a}_k\|^2 \leqslant 4k^2$. Then,

$$\begin{aligned}
\|\mathbf{a}_{k+1}\|^2 &= \sum_x \|U_{k+1}(V_x \otimes I_B)|\Psi_{x,k}\rangle - U_{k+1}|\Psi_k\rangle\|^2 \\
&= \sum_x \|(V_x \otimes I_B)|\Psi_{k,x}\rangle - |\Psi_k\rangle\|^2 \\
&= \sum_x \|(V_x \otimes I_B)|\Psi_{k,x}\rangle - (V_x \otimes I_B)|\Psi_k\rangle + (V_x \otimes I_B)|\Psi_k\rangle - |\Psi_k\rangle\|^2 \\
&\leqslant \sum_x \left(\| \underbrace{|\Psi_{k,x}\rangle - |\Psi_k\rangle}_{a_{k,x}} \| + \| \underbrace{(V_x \otimes I_B - I_A \otimes I_B)|\Psi_k\rangle}_{c_{k,x}} \| \right)^2 \\
&= \|\mathbf{a}_k - \mathbf{c}_k\|^2 \\
&\leq \left(\|\mathbf{a}_k\| + \|\mathbf{c}_k\|\right)^2 .
\end{aligned}$$

It only remains to upper bound $\|\mathbf{c}_k\|$. Since $V_x = -2|x\rangle\langle x| + I_A$, we have

$$\begin{aligned}
(V_x \otimes I_B) - (I_A \otimes I_B) &= (V_x - I_A) \otimes I_B \\
&= -2|x\rangle\langle x| \otimes I_B .
\end{aligned}$$

Using this fact, we obtain

$$\begin{aligned}
\|\mathbf{c}_k\|^2 &= 4 \sum_x \|(|x\rangle\langle x| \otimes I_B)|\Psi_k\rangle\|^2 \\
&= 4 \sum_x \langle\Psi_k|(|x\rangle\langle x| \otimes I_B)|\Psi_k\rangle \\
&= 4 . \qquad (12.2)
\end{aligned}$$

In conclusion,

$$\begin{aligned}
\|\mathbf{a}_{k+1}\| &\leq \|\mathbf{a}_k\| + \|\mathbf{c}_k\| \\
&\leq 2k + 2 \\
&= 2(k+1).
\end{aligned}$$

This proves that $\|\mathbf{a}_k\| \leqslant 2k$ for every k. In summary, since $\|b_k\|$ is of order $\sqrt{N}$, the error $\eta_k \geq |\|\mathbf{a}_k\| - \|\mathbf{b}_k\||^2/N$ can go to zero only if k is of order $\sqrt{N}$.

■

The above proof shows that finding a solution with high probability requires $\Theta(\sqrt{N})$ uses of the unitary gate V_x. In the game of boxes and prizes, this means that, in order to find a box with the prize, we need to operate on $\Theta(\sqrt{N})$ boxes. In other words, the quadratic speed-up of Grover's algorithm is the maximum improvement that is allowed by quantum mechanics.

Remark 6 Note that the above result does not change if we replace V_x with the gate $U_x = \sum_{n=1}^{N} |n\rangle\langle n| \otimes X^{f(n)}$, which implements the quantum computation of the function f such that $f(n) = \delta_{n,x}$. To see why the result does not change, it is enough to write U_x as

$$U_x = I_A \otimes |+\rangle\langle +| + V_x \otimes |-\rangle\langle -|$$

and use this expression to obtain the same result as in Eq. (12.2). In other words, using the gate U_x, the length of the vector $\mathbf{a}_k$ still grows only as $\|\mathbf{a}_k\| \leq 2k$.

Remark 7 There is also another interesting observation about the optimality of Grover's algorithm: not only we need to operate on $\Theta(\sqrt{N})$ boxes in order to find a solution, but also we have to do it in $\Theta(\sqrt{N})$ *time steps*. Let us clarify this point. In principle, there is a difference between using the gate V_x for k times and using k time steps: for example, we could apply the gate V_x in parallel on k systems at the same time, producing the state

$$|\Psi_x\rangle = (V_x^{\otimes k} \otimes I_B)|\Psi\rangle\,.$$

In this case, the gate V_x has been used k times in a single time step. In the case of Grover's algorithm, the gate V_x is used k times in different k time steps. So the question is, can we find a solution to the equation $f(n) = 1$ using the gate V_f for $\Theta(N)$ times *in a single time step*? Or, more generally, can we find a solution using less than $\Theta(\sqrt{N})$ time steps? The answer is that $\Theta(\sqrt{N})$ time steps are absolutely necessary: the time resource used by Grover's algorithm is optimal, in the sense that no quantum algorithm can find a solution of the equation $f(n) = 1$ in less than $\Theta(\sqrt{N})$ steps.

12.6 Chapter summary

In this chapter we considered the problem of quantum search. This problem is slightly different from the problem of classical search, because in the model of quantum search we allow ourselves some extra-power: either *i)* the ability to operate on a particular qubit n determined by the quantum state of a control system, or *ii)* the ability to compute a function $f : \{1, \ldots, N\} \to \{0, 1\}$ using a reversible unitary gate U_f.

In both scenarios *i)* and *ii)* the problem of quantum search is reduced to finding a solution of the equation $f(n) = 1$ using the gate V_f in a quantum circuit.

Grover's algorithm shows one efficient way to do quantum search: using the gate V_f for $O(\sqrt{N})$ times one can obtain a solution with high probability $p_{succ} \geq 1 - S/N$, where $S \ll N$ is the total number of solutions. The Grover's algorithm gives a quadratic speed-up over classical search. Although this does not change complexity classes, this result is still interesting and can have useful applications.

The scaling $\Theta(\sqrt{N})$ of Grover's algorithm is optimal: no other algorithm can find a solution with less than $\Theta(\sqrt{N})$ uses of V_f. In addition, Grover's algorithm is also optimal in terms of time steps: no other algorithm can find a solution in less than $\Theta(\sqrt{N})$ time steps. These results are quite fundamental as they represent the ultimate limits allowed by quantum mechanics to the speed of search.

Finally, one important feature of Grover's algorithm is that it requires you to know in advance the number of solutions. When such a number is unknown, one can first estimate the number of solutions using the phase estimation algorithm, which will be presented in the next chapter.

Chapter 13

Breaking the RSA code with quantum mechanics

With this lecture we conclude the course, exploring what is probably the most famous result in Quantum Information: Shor's algorithm for factoring numbers in polynomial time. Our exploration will be a rapid trip, touching topics in cryptography, complexity theory, number theory, and, of course, highlighting some deep facts about quantum mechanics.

13.1 The RSA code

Suppose that Alice and Bob want to communicate some secret message over an insecure communication channel. Unfortunately, they don't share a secret key that can be used to encrypt the message. Even more unfortunately, they cannot send quantum systems to each other and therefore they cannot use the BB84 protocol to generate a secret key.

This sounds a pretty bad situation. Actually, this our everyday situation when we send an email or when we input our credit card number in an online purchase. So, how do we protect our emails from being read and our credit cards from being stolen?

Here is where complexity theory comes to the rescue. Under the assumption that solving some mathematical problem is hard, there exist cryptographic protocols that allow Alice and Bob to communicate securely even in the presence of an eavesdropper, Eve. The protocols are designed in such a way that encoding and decoding the message is computationally easy for Alice and Bob, but computationally hard for Eve.

The most famous such protocol is the RSA code, named after the initials of its inventors Rivest, Shamir, and Adleman. The purpose of the protocol is to generate a binary string a that is known only to Alice and Bob and can be used later to encrypt a message. The steps of the protocol are as follows

1. Bob chooses at random two large prime numbers p and q and a number c that is coprime with $M = (p-1)(q-1)$
2. Bob communicates to Alice the value of $N = pq$ and the value of c
3. Alice chooses at random a number $a < N$ that is coprime with N. The binary expansion of a will be the secret key.
4. Alice encodes a in the number $b = a^c \mod N$ and communicates b to Bob.
5. Bob computes the inverse of c modulo M, denoted by d, and decodes $a' = b^d \mod N$.

The promise of the protocol is that *i)* Bob's decoding is correct (i.e. $a' = a$) and that *ii)* Eve cannot discover a. In this way, a can be used later as a key to encrypt an L bit message, with $L = \lfloor \log a \rfloor$.

At first sight, the protocol looks quite mysterious. Why does it work? And why it is secure? You will find the answer to these questions in the next two paragraphs.

13.2 Why the RSA code works

In order to prove that the RSA code works we have to prove that *i)* Bob decodes correctly Alice's message (that is, $a' = a$) and that *ii)* all the steps involved in the protocol are easy (that is, Alice and Bob have to perform only a number of elementary operations that is polynomial in L).

Correctness of Bob's decoding. Let us check first that $a' = a$. In order to decode, Bob first computes d, the inverse of c modulo M. The inverse exists because the integers in the interval $\{1, M-1\}$ that are coprime with M form a group under multiplication modulo M. If this is not immediately clear to you, try the following exercise:

Exercise 28 Let $\mathbf{G}_M \subset \{1, M-1\}$ be the set of integers that are coprime with M. Using the notation $[x] := x \mod M$ and $x \circ y := [xy]$, show that the following properties hold

1. for every $x, y \in \mathbf{G}_M$, one has $x \circ y \in \mathbf{G}_M$.
2. for every $x, y, z \in \mathbf{G}_M$, one has $(x \circ y) \circ z = x \circ (y \circ z)$
3. for every $x, y, z \in \mathbf{G}_M$, $x \circ y = x \circ z$ implies $y = z$
4. for every $x \in \mathbf{G}_M$ there exist two distinct integers $m > n$ such that $[x^m] = [x^n]$
5. for every $x \in \mathbf{G}_M$ there exist an integer $k > 0$ such that $[x^k] = 1$
6. for every $x \in \mathbf{G}_M$ there exists $y \in \mathbf{G}_M$ such that $x \circ y = 1$
7. $(\mathbf{G}_M, \circ)$ is a group

After computing d, Bob will evaluate $a' = b^d \mod N$. To prove that the decoding is correct it remains to show that $a' = a$. Note that we have

$$\begin{aligned} a' &= b^d \mod N \\ &= [a^c \mod N]^d \mod N \\ &= a^{cd} \mod N \\ &= a^{1+s(p-1)(q-1)} \mod N \end{aligned}$$

for some positive integer s. In the last equality we used the fact that, by definition, $cd = 1 \mod M$ and $M = (p-1)(q-1)$.

To prove that the decoding is correct it remains to show that

$$a^{s(p-1)(q-1)} \mod N = 1\,, \tag{13.1}$$

for $N = pq$.

The proof is based on a cute result known as *Fermat's little theorem*:

Theorem 26 *Let p be a prime and x be an integer that is coprime with p. Then, $x^{p-1} \mod p = 1$.*

Unlike the proof of the more famous "Fermat's last theorem", the proof of Fermat's little theorem is quite easy [1]. Equipped with this result, it is immediate to prove Eq. (13.1):

Proof of Eq. (13.1). Recall that, by construction, a is comprime with $N = pq$. Since p and q are primes, this means that a is coprime both with p and with q. Hence, a^{p-1} is coprime with q and a^{q-1} is coprime with p. Applying Fermat's little theorem, we then get

$$\begin{aligned} \left(a^{p-1}\right)^{q-1} &\mod q = 1 \\ \left(a^{q-1}\right)^{p-1} &\mod p = 1\,. \end{aligned}$$

This means that $a^{(p-1)(q-1)} - 1$ is multiple both of p and q. Since p and q are primes, this also means that $a^{(p-1)(q-1)} - 1$ is a multiple of pq, namely

$$a^{(p-1)(q-1)} \mod pq = 1\,.$$

[1] For those of you who know a little bit of group theory, here is an easy proof. An elementary fact in group theory is that for every group $\mathbf{G}$ and for every element $g \in \mathbf{G}$ one has

$$g^{|\mathbf{G}|} = e\,, \tag{13.2}$$

where $|\mathbf{G}|$ is the number of elements of the group and $e \in \mathbf{G}$ is the identity element. Now, consider the group $\mathbf{G}_p$ as defined in Exercise 28. Since p is prime, every positive integer smaller than p is coprime with p. Hence, $|\mathbf{G}_p| = p - 1$. Moreover, since x is coprime with p, one has that $[x] \equiv x \mod p \neq 0$, that is, $[x]$ is an element of $\mathbf{G}_p$. As a consequence, one obtains $x^{p-1} \mod p = [x^{p-1}] = [[x]^{p-1}] = [1] = 1$, having used the notation of Exercise 28. ■

Raising both sides of the equality to the power s we obtain the desired result. ■

In conclusion, we have verified that the Bob's decoding works correctly: $a' = a$.

Complexity of Alice's and Bob's operations. Here the size of the problem is the number of bits needed to write down the large primes p and q. This number is of the order of a few hundreds in the typical applications, and its order is also the order of the number of bits needed to write the product pq. Let us denote by $L = \lfloor \log(pq) \rfloor$.

The key reasons why the RSA protocol is easy for Alice and Bob are the following

1. the probability that two numbers chosen at random are coprime is larger than $1/2$
2. finding the greater common divisor (*gcd*) of two numbers is easy: it can be done using Euclid's algorithm, which requires $O(L^2)$ operations if the two numbers are of size L
3. if x and y are coprime and $x < y$, then finding the inverse of x modulo y is easy and can be done in $O(L^2)$ steps as a byproduct of Euclid's algorithm

Using these three facts, it is easy to see that the RSA protocol is easy to implement for Alice and Bob. First, Bob has no difficulty in finding a random number c that is coprime with M. The probability that he gets such a number is quite high. Moreover, to check if c is really coprime with M, Bob can compute the *gcd* between c and M and check if it is 1 or not. In the unlucky case when the *gcd* is not 1, he can try with another value of c.

Similarly, for Alice it is easy to find a random a that is coprime with N.

Finally, computing the function

$$f(x) = a^x \mod N$$

is also easy and can be done in $O(L^3)$ steps. Indeed, writing x in binary, as

$$x = \sum_{j=1}^{L} 2^{L-j}\, x_j\,,$$

we have

$$\begin{aligned} a^x &= a^{\left[\sum_{j=1}^{L} 2^{L-j}\, x_j\right]} \\ &= \prod_{j=1}^{L} \left(a^{2^{L-j}}\right)^{x_j} \end{aligned}$$

so that

$$a^x \mod N = \left(\prod_{j=1}^{L} A_j^{x_j} \right) \mod N \,,$$

where $A_j := a^{2^{L-j}} \mod N = (A_{j+1})^2 \mod N$. More explicitly, we have

$$\begin{aligned} A_L &= a \\ A_{L-1} &= A_L^2 \mod N \\ A_{L-2} &= A_{L-1}^2 \mod N \\ &\vdots \\ A_1 &= A_2^2 \mod N \,. \end{aligned}$$

In other words, to compute $a^x \mod N$ it is enough to compute L numbers $\{A_j\}_{j=1}^{L}$. Computing each of them requires us to compute the square of an L-bit number modulo another L-bit number (this takes $O(L^2)$ operations in total). Hence, in total we need $O(L^3)$ operations. This observation will also be important in the discussion of Shor's algorithm, because the same argument we showed here can be used to prove that the *unitary gate* U_f can be realized from $O(L^3)$ elementary gates: after all, what is easy on a classical computer should also be easy on a quantum computer.

We have just proved that encoding a into $b = a^c \mod N$ is easy for Alice. Finally, at the decoding step, we know that finding the inverse of c is easy [$O(L^2)$ operations] and, again, that computing $b^d \mod N$ is easy [$O(L^3)$ operations].

13.3 Security of the RSA code

Suppose that you are a spy and want to crack the RSA code. Which strategy would you use?

The most natural strategy is to try to factor the number N. Once you have its prime factors, p and q, you can compute $M = (p-1)(q-1)$, compute the inverse of c modulo M, and then decode b in the same way as Bob does. The reason why the RSA is considered secure is that nobody knows how to factor large numbers in an efficient way [2]. The best classical algorithm known nowadays is the *field number sieve*, which takes $\exp[\Theta(L^{1/3} \log^{2/3} L)]$ operations before finding the prime factors. In practice, with L of the order of a few hundreds, this superpolynomial scaling means that it will take you forever to compute the prime factors p and q.

[2] In fact, the general opinion in the computer science community is that factoring numbers in polynomial time should be impossible. Although no one has found a rigorous proof yet, there are pretty strong arguments that support this opinion.

Another, less obvious way to crack the RSA code is to find the period of the function

$$f(x) = a^x \mod N\,.$$

Of course, you do not know a, but it turns out that the period of the function $f(x)$ is the same as the period of the function $g(x) = b^x \mod N$. And you know b. The reason why the period of g is equal to the period of f is that *i)* by definition, b is a power of a—which implies that the period of g is no longer than the period of f, and *ii)* by the correctness of the decoding, a is a power of b—which implies that the period of f is no longer than the period of g.

Now, once you find the period of f, call it r, you can find the inverse of c modulo r, call it $\tilde{d}$ [3]. Once you have the inverse of c, using the relation

$$c\tilde{d} = 1 + rs \qquad s \in \mathbb{N}$$

you can decode

$$\begin{aligned} b^{\tilde{d}} \mod N &= a^{c\tilde{d}} \mod N \\ &= a^{1+rs} \mod N \\ &= f(1+rs) \\ &= f(1) \\ &= a\,. \end{aligned}$$

However, also in this case, there is no known classical algorithm that can find efficiently the period of g. This is because finding a period is a primitive for factoring: if you could find periods efficiently, then you could also factor efficiently. We will see this in more detail in the next paragraph.

In conclusion, although there is no formal proof that the RSA code is secure, the code is believed to be secure under the assumption that some mathematical problems, such as factoring numbers or finding the period of the function $g(x) = b^x \mod N$, are hard.

13.4 From period finding to factoring

Here we establish that finding the period of the function $f(x) = a^x \mod N$ is a more powerful task than factoring N: if a friend borrows you a magical machine that can find the period efficiently, then you can use this machine to factor numbers efficiently.

For simplicity, suppose that N is the product of two primes p and q, like in the RSA code. Nevertheless, the same argument can be easily extended to the case of generic N.

The algorithm works as follows:

[3] The inverse exists because c and r must have no factor in common: if $c = c'f$ and $r = r'f$ for some common factor f then the period of $g(x) = a^{cx} = a^{c'fx}$ would be r' instead of r. This is not possible, because the period r is the *smallest* number such that $g(x + r) = g(x)$.

1. Take a random integer $a < N$ and check if a divides N.

2. If a divides N, you are done: this means that either $a = p$ or $a = q$. If not, then proceed to the next step.

3. Use the period-finding machine to find the period of the function $f(x) = a^x \mod N$. Call the period r.

4. If r is odd, then go back to Step 1. If r is even, then proceed to the next step.

5. Compute $x_+ = a^{r/2} + 1$ and $x_- = a^{r/2} - 1$.

6. If $x_+ = 0 \mod N$, then go back to Step 1. Otherwise, proceed to the next step.

7. Output the solution $\{p, q\} = \{gcd(N, x_+)\,,\, gcd(N, x_-)\}$.

When the algorithm terminates, it is easy to see that it gives the correct answer. Indeed, we have that

- x_- is not a multiple of N. Otherwise, we would have $a^{r/2} = 1 \mod N$, which would lead to the contradiction that the period of the function $f(x)$ is $r/2$, instead of r.

- since x_+ and x_- are not multiples of N, the only way that their product x_+x_- can be a multiple of N is that

  1. either x_+ is a multiple of p (but not of q) and x_- is a multiple of q (but not of p)
  2. either x_+ is a multiple of q (but not of p) and x_- is a multiple of p (but not of q)

  In the first case we conclude that $p = gcd(x_+, N)$ and $q = gcd(x_-, N)$. In the second case we conclude $q = gcd(x_+, N)$ and $p = gcd(x_-, N)$.

In addition to being correct, the algorithm terminates after a few steps with high probability. The reason is that the probability that r is even and $x_+ \neq 0 \mod N$ is pretty high. In practice, this means that a few repetitions of the algorithm are sufficient to find the solution [4].

In summary, if you have a period-finding machine, you can factor numbers efficiently. And of course, you can crack the RSA code. Since in the classical world nobody has found an algorithm that can find factor numbers efficiently, this also means that nobody has found a classical algorithm that can find the period of the function $f(x) = a^x \mod N$ efficiently.

[4] The actual value of the probability of the favourable event "(r is even) $\wedge$ (x_+ is not multiple of N)" does not matter so much, as long as such a probability is lower bounded by a constant ϵ independent of N. After $O(\log 1/\epsilon)$ repetitions, the algorithm will be likely to output the solution.

However, the classical world is a small place, where all computations have to happen in the computational basis. Finding the period may be hard on a classical computer, but can still be easy on a quantum computer. This is indeed the case: in the rest of the lecture we will see that quantum mechanics allows one to construct the "magical period-finding machine" that we imagined earlier in this paragraph.

13.5 A quantum period-finding machine

Let $f : \{1, \dots, N\} \to \{1, \dots, N'\}$ be a function with the property that, for some $r < N$, one has

$$f(x) = f(y) \qquad \Longleftrightarrow \qquad x = y + kr, \quad k \in \mathbb{Z}\,. \tag{13.3}$$

In other words, *i)* the function has period r and *ii)* two different inputs give the same output only if they differ by an integer multiple of the period.

One example of function with this property is the function that we used for factoring:

Exercise 29 Let $a < N$ be coprime with N. Then the function $f(x) = a^x \mod N$ satisfies Eq. (13.3).

Now, we want to find the period of f. In general, this is a hard problem if $f(x)$ is an arbitrary function: in the worst case, finding the period may require $O(N)$ evaluations of f. We will now see a quantum algorithm that can be used to find the period exponentially faster. Strictly speaking, the quantum algorithm does not find directly the period, but instead if finds a *divisor of the period*. However, once you have a fast way to find divisors of the period, a little bit of classical analysis shows that you can find the period with high probability. The quantum algorithm uses the unitary gate U_f that computes the function f reversibly. Recall that the gate U_f acts on two system A and B of dimensions N and M, respectively, and its action is given by

$$U_f = \sum_{x=1}^{d_A} |x\rangle\langle x| \otimes S^{f(x)}\,,$$

where S is the shift gate. For reasons that will become clear very soon, it is a good idea to choose system A to be of dimension $\widetilde{N} > N$, for a suitable integer $\widetilde{N}$ that will be specified later. The value of $f(x)$ for the inputs $x \in \{N+1, \dots, \widetilde{N}\}$ is completely determined by the fact that f is periodic and that the period is smaller than N.

With the above settings, we are ready to describe the period-finding algorithm, which is the subroutine at the core of Shor's algorithm:

1. Prepare system A in the Fourier state $|e_0\rangle = \frac{1}{\sqrt{\widetilde{N}}} \sum_x |x\rangle$ and system B in the computational state $|0\rangle$.

2. Apply U_f

3. Measure B in the computational basis.

4. Measure A in the Fourier basis.

5. From the outcome of the Fourier basis measurement, call it n one can extract a divisor by classical methods. In the easy case where $\widetilde{N}$ is a multiple of the period, it is enough to compute the fraction $n/\widetilde{N}$ and reduce it to minimal terms $n/\widetilde{N} = k_0/r_0$—in this case, r_0 is a divisor of the period.

Note that the order of the Steps 3 and 4 does not matter, because the two measurements are performed on independent systems. The reason why we list the measurement on B before the measurement on A is that this makes it easier to see why the algorithm works.

Let us follow what happens through the steps of the protocol: After the application of U_f, the state of system AB is

$$\begin{aligned} U_f|e_0\rangle|0\rangle &= \frac{1}{\sqrt{\widetilde{N}}}\sum_{x=1}^{\widetilde{N}}|x\rangle|f(x)\rangle \\ &= \frac{1}{\sqrt{\widetilde{N}}}\sum_{x=1}^{r}\left(\sum_{m=0}^{M_x}|x+mr\rangle\right)|f(x)\rangle. \end{aligned}$$

In the second equality we used the strong periodicity of the function f and we denoted by $M_x + 1$ the number of inputs that are equal to x modulo r.

Now, if we measure system B on the computational basis, the outcome of the measurement on the second system will be equal to $f(x)$, with probability $(M_x + 1)/\widetilde{N}$. Conditionally on this outcome, the state of system A will be

$$|\varphi_x\rangle := \frac{1}{\sqrt{M_x+1}}\sum_{m=0}^{M_x}|x+mr\rangle.$$

Finally, when we measure system A in the Fourier basis we obtain the outcome n with probability

$$\begin{aligned} p(n|x) &= |\langle e_n|\varphi_x\rangle|^2 \\ &= \frac{1}{M_x+1}\left|\sum_{m=0}^{M_x}\langle e_n|x+mr\rangle\right|^2 \\ &= \frac{1}{(M_x+1)\widetilde{N}}\left|\sum_{m=0}^{M_x}\exp\left[\frac{-2\pi i n(x+mr)}{\widetilde{N}}\right]\right|^2 \\ &= \frac{1}{(M_x+1)\widetilde{N}}\left|\sum_{m=0}^{M_x}\exp\left[\frac{-2\pi i\, n\, m\, r}{\widetilde{N}}\right]\right|^2 . \end{aligned} \tag{13.4}$$

Now, consider the easy case when $\widetilde{N}$ is a multiple of the period r, namely $\widetilde{N} = rM$ for some integer M. In this case, since $\widetilde{N}$ contains M full periods, we have $M_x = M - 1$ for every x. Then, the probability becomes

$$\begin{aligned} p_n &= \frac{1}{r} \left| \frac{1}{M} \sum_{m=0}^{M-1} \exp\left[\frac{-2\pi i\, n\, m}{M}\right] \right|^2 \\ &= \frac{1}{r} \sum_{k=1}^{r} \delta_{n,kM} \end{aligned}$$

In other words, the outcome of the Fourier measurement is guaranteed to be a multiple of M, chosen at random in the set $\{M, 2M, \ldots, rM\}$. Now if the outcome of the measurement is $n = kM$, recalling that $M = \widetilde{N}/r$, we have that $n/\widetilde{N} = k/r$. Reducing the fraction to minimal terms, we obtain $n/\widetilde{N} = k_0/r_0$, where r_0 is a divisor of the period.

Things are less trivial in the case where $\widetilde{N}$ is not a multiple of the period. However, when M_x is large, it is not difficult to see that the probability p_n in Eq. (13.4) is peaked around the values n that are close to integer multiples of $\widetilde{N}/r$. The larger M_x becomes, the more peaked the probability becomes. One way to guarantee that each M_x is large is to choose $\widetilde{N}$ to be large enough. A good choice is, for example, $\widetilde{N} = N^2$. The change from N to $\widetilde{N} = N^2$ does not change the performance of the algorithm in a significant way: in terms of bits, this just means that we are changing from L bits to $2L$. In this case, the classical analysis of the algorithm is much more complicated and will be omitted here. If you are interested in it, you can find more details in Mermin's book "Quantum Computer Science" and in Nielsen-Chuang book.

In summary, we have a quantum algorithm that, with high probability produces numbers that are divisors of the period. How can we get the period from this?

A simple recipe is this:

1. Check if the output of the quantum algorithm, r_0, is the period (e.g. by evaluating $f(1)$ and $f(1 + r_0)$ and checking if they are the same).

2. If r_0 is the period, you are done. If r_0 is not the period, try a few low multiples of r_0, say $2r_0, 3r_0, \ldots, t\, r_0$ with $t = O(1)$.

3. If one of them is the period, you are done. If none of them is the period, run the quantum algorithm again, obtaining a new value r_0', and go back to Step 1.

A little bit of classical analysis shows that this procedure terminates in few steps with high probability. Intuitively, since the measurement outcome is a random multiple of M, say kM, it is quite unlikely that k has a big factor in common with r. Hence, there is a good chance that the period is one of the low multiples $r_0, 2r_0, \ldots, tr_0$. Moreover, if we run the quantum algorithm a second time, we will get another random multiple of M, say $k'M$. Since k and k' are random,

it is quite likely that they are coprime. If this is the case, then r is the least common multiple of r_0 and r_0'. In all the steps, the probability to succeed is reasonably high (larger than 1/2) probability, and checking whether or not we found the period is cheap (requires one evaluation of f).

In summary, we have seen a quantum algorithm that finds a period of a function f using the gate U_f for $O(1)$ times and making elementary classical computations (each of them requiring at most $O(L^3)$ operations). In particular, choosing the function $f(x) = a^x \mod N$, the period finding algorithm can be used to factor numbers.

Note that the ingredients of the algorithm are very simple:

- preparation of the Fourier basis state $|e_0\rangle$
- access to the reversible gate U_f
- measurement on the Fourier basis.

13.6 The complexity of finding the period

Shor's period-finding algorithm gives us a super-fast way to compute the period of a function and to factor large numbers, *provided that we can*

- prepare of the Fourier basis state $|e_0\rangle$
- implement the reversible gate U_f
- measure on the Fourier basis.

The implementation of U_f is the not a big problem if f can be computed classically in an efficient way. After all, everything that is easy in the classical world, is also easy in the quantum world, where we can replace elementary reversible operations on classical bits with elementary quantum gates on qubits. For example, the function $f(x) = a^x \mod N$ can be computed in $O(L^3)$ steps using classical gates, and the corresponding gate U_f can be computed in $O(L^3)$ steps using quantum gates.

But what guarantees that we can do prepare and measure states on the Fourier basis quickly? Suppose that measuring in the Fourier basis required $\Theta(N)$ computational steps. If this were the case, the exponential advantage of the quantum algorithm would be lost!

This is a really scary observation. To make it even scarier, note that computing the Fourier transform on a classical computer *does require* $\Omega(N)$ *steps.* If you have a vector of coefficients $(c_1, \ldots, c_N)$, computing the Fourier components

$$\tilde{c}_k = \sum_{n=1}^{N} e^{2\pi i k n} c_n$$

requires $\Theta(N \log N)$ steps using the fastest classical algorithm known nowadays (the classical algorithm that achieves this scaling is known as *fast Fourier transform*).

Luckily, realizing a measurement in the Fourier basis requires only $O\left[(\log N)^2\right]$ steps. Precisely, imagine that we have a quantum computer that processes information encoded into qubits. In addition, imagine that our quantum computer can do the following things:

1. prepare qubits in the state $|0\rangle$
2. perform an universal set of one qubit gates, say $\{H, T\}$
3. perform an entangling gate, say the `CNOT` gate
4. perform measurements on the computational basis.

We want to program the quantum computer to realize the Fourier measurement. How many elementary operations do we need to perform?

By the Solovay-Kitaev theorem we know that performing single-qubit and two-qubit gates is easy: if we want to approximate a single-qubit gate with precision ϵ, we need to to apply a sequence of only $O\left(\log \epsilon^{-1}\right)$ elementary gates. Note that the number of elementary gates depends only on the desired precision ϵ, although we may need a precision that depends on N. As long as the required precision scales like $1/N$, however, this fact does not represent a problem [5].

We will now see an easy way to prepare states and to measure on the Fourier basis using only $\Theta\left[(\log N)^2\right]$ two qubit gates.

13.7 The complexity of preparing states in the Fourier basis

We start from the observation that the states of the Fourier basis are generated by the "multiply operator" M defined as

$$M = \sum_{n=1}^{N} \exp\left[\frac{2\pi i n}{N}\right] |n\rangle\langle n| .$$

Indeed, we have

$$|e_k\rangle = M^K |e_0\rangle \qquad \forall k = 1, \ldots, N \tag{13.5}$$

where

$$|e_0\rangle \equiv |e_N\rangle = \frac{1}{\sqrt{N}} \sum_n |n\rangle .$$

From now on, we suppose that N is a power of 2 and we define $L := \log N$. Since our quantum computer operates on qubits, we must write the numbers n and k in binary, and encode them into the state of L qubits. Precisely, we write a number $x \in \{1, \ldots, N\}$ as a vector

$$x = (x_1, x_2, \ldots, x_L) \in \{0, 1\}^{\times L} ,$$

[5] It is an important part of the analysis of Shor's algorithm to show that, indeed, we do not need to implement the gates with precision better than $1/N$.

meaning that $x = \sum_{j=1}^{L} 2^{L-j}\, x_j$. Using this fact, the number x can be encoded in the computational basis state

$$|x\rangle = |x_1\rangle|x_2\rangle \cdots |x_L\rangle$$

The binary representation has two beautiful properties. First, the Fourier vector $|e_0\rangle$ can be written as

$$|e_0\rangle = |+\rangle^{\otimes L}\,.$$

Since $|+\rangle = H|0\rangle$, this means that we can generate $|+\rangle^{\otimes L}$ from the computational basis vector $|0\rangle^{\otimes L}$ by applying the Hadamard gate H on each qubit. This is good, because this requires only L single-qubit gates.

The second beautiful property is that the multiply operator is a product of single qubit gates. Indeed, we have

$$\begin{aligned} M|n\rangle &= \exp\left(\frac{2\pi i n}{N}\right)|n\rangle \\ &= \exp\left[\frac{2\pi i\left(\sum_{j=1}^{L} 2^{L-j}\, n_j\right)}{2^L}\right]|n_1\rangle|n_2\rangle\ldots|n_L\rangle \\ &= \bigotimes_{j=1}^{L}\left[\exp\left(\frac{2\pi i\, n_l}{2^j}\right)|n_j\rangle\right] \\ &= R_1|n_1\rangle \otimes R_2|n_2\rangle \otimes \cdots \otimes R_L|n_L\rangle\,, \end{aligned} \tag{13.6}$$

where R_j is the single-qubit gate defined by

$$R_j := \begin{pmatrix} 1 & 0 \\ 0 & \exp\left[\frac{2\pi i}{2^j}\right] \end{pmatrix}.$$

Since Eq. (13.6) holds for every vector of the computational basis, we proved the relation

$$M = R_1 \otimes R_2 \otimes \cdots \otimes R_L\,.$$

In other words, the multiply operator is easy to implement: it requires only L single-qubit gates.

Now, we would like to generate the Fourier basis from the fixed state $|e_0\rangle$, using Eq. (13.5). However, there is a problem: if we take the relation $|e_k\rangle = M^k|e_0\rangle$ too literally and prepare the state $|e_k\rangle$ by applying M for k times, then we may have to apply $O(N)$ gates. This is not good at all!

Luckily, also here there is a trick: by definition, we have

$$(R_j)^{2^m} = \begin{cases} R_{j-m} & m < j \\ I & m \geq j \end{cases}.$$

Using this relation, we obtain

$$\begin{aligned} R_j^k &= R_j^{\left(\sum_{l=1}^{L} 2^{L-l} k_l\right)} \\ &= \prod_{l=L-j+1}^{L} R_{j-L+l}^{k_l} \,, \end{aligned}$$

or, explicitly,

$$\begin{aligned} R_L^k &= R_L^{k_L} \, \ldots \, R_2^{k_2} \, R_1^{k_1} \\ R_{L-1}^k &= R_{L-1}^{k_L} \, \ldots \, R_2^{k_3} \, R_1^{k_2} \\ &\vdots \\ R_1^k &= R_1^{k_L} \,. \end{aligned} \tag{13.7}$$

This means that R_j^k can be realized by using at most $j \leq L$ single-qubit gates. This is exponentially better than repeating R_l for k times, because, as we already observed, k can be of order $N = 2^L$.

Summarizing what the results that we obtained up to now:

- the multiply operator M can be written as a product of L single-qubit gates as $M = R_1 \otimes R_2 \otimes \cdots \otimes R_L$
- the exponential R_j^k can be realized using a sequence of at most L single qubit gates.

Putting the two results together, we have proved that the gate M^k can be realized using a sequence of at most L^2 single qubit gates. More precisely, it can be realized using

$$\sum_{j=1}^{L} j = \frac{L(L+1)}{2}$$

single qubit gates.

Good. We now know a cheap way to prepare every desired *state* in the Fourier basis using a few single-qubit gates. But can we find a cheap way to implement the unitary Fourier *gate* F?

The key to measuring on the Fourier basis is to realize the *Fourier gate* F, which we met in the first lecture of this course. For a N-dimensional quantum system, the Fourier gate is the unitary operator

$$F = \sum_{n=1}^{N} |e_n\rangle\langle n| \,,$$

which maps states in the computational basis into states in the Fourier basis.

It turns out that also realizing F is easy and can be done using $O(L^2)$ two-qubit gates. Indeed, we can define the control-unitary gates.

$$C_j := I \otimes |0\rangle\langle 0| + R_j \otimes |1\rangle\langle 1| \,.$$

Here, the control qubit (right) decides whether or not we apply the gate R_j on a given targed qubit (left). Let us use the notation $C_j^{(mn)}$ to mean that we apply the gate C_j to the qubits m and n and that m is the target, while n is the control. For convenience of notation, we also define $C_1^{(mm)}$ to be the Hadamard gate H acting on qubit m: note that we have

$$C_1^{(mm)}|k_m\rangle = R_1^{k_m}|+\rangle \,.$$

With this definition, we have

$$\begin{aligned}
C_L^{(1L)} \cdots C_3^{(13)} C_2^{(12)} C_1^{(11)} |k_1\rangle|k_2\rangle \cdots |k_L\rangle &= R_L^k|+\rangle|k_2\rangle \cdots |k_L\rangle \\
C_{L-1}^{(2L)} \cdots C_3^{(24)} C_2^{(23)} C_1^{(22)} |k_2\rangle|k_3\rangle \ldots |k_L\rangle &= R_{L-1}^k|+\rangle|k_3\rangle \ldots |k_L\rangle \\
&\vdots \\
C_1^{(LL)}|k_L\rangle &= R_1^k|+\rangle \,,
\end{aligned}$$

Hence, defining

$$\begin{aligned}
U_1 &:= C_L^{(1L)} \cdots C_3^{(13)} C_2^{(12)} C_1^{(11)} \\
U_2 &:= C_{L-1}^{(2L)} \cdots C_3^{(24)} C_2^{(23)} C_1^{(22)} \\
&\vdots \\
U_L &:= C_1^{LL} \,,
\end{aligned}$$

we obtain

$$\begin{aligned}
(I_1 \otimes \cdots \otimes I_{L-1} \otimes U_L) \cdots (I_1 \otimes I_2 \otimes U_3)(I_1 \otimes U_2) U_1 |k\rangle &= R_L^k|+\rangle \otimes R_{L-1}^k|+\rangle \otimes \cdots \otimes R_1^k|+\rangle \\
&\simeq M^k|e_0\rangle \\
&= |e_k\rangle \\
&= F|k\rangle \,, \qquad (13.8)
\end{aligned}$$

where the symbol $\simeq$ in the third line denotes the equality up to reshuffling of the L qubits.

Since the Eq. (13.8) holds for every k, we proved the decomposition

$$(I_1 \otimes \cdots \otimes I_{L-1} \otimes U_L) \cdots (I_1 \otimes I_2 \otimes U_3)(I_1 \otimes U_2) U_1 \simeq F \,.$$

Note that the sequence of gates $U_1, U_2, \ldots, U_L$ can be realized using $O(L^2)$ two-qubit gates, and the reshuffling of the qubits also requires $O(L)$ two-qubit gates. In total, the Fourier gate F can be constructed from a circuit that uses $\Theta(L^2)$

two-qubit gates [6]. If you think about it, this is a pretty spectacular result, which also tells you a deep thing: *computing* the Fourier transform of a vector is hard on a classical computer, but *realizing* a physical evolution that implements the Fourier gate is easy on a quantum computer. Note that the Fourier gate is really the quantum version of the Fourier transform: if you apply F to a quantum state

$$|\varphi\rangle = \sum_n \varphi_n |n\rangle$$

you will obtain

$$F|\varphi\rangle = \sum_k \tilde{\varphi}_k |k\rangle \qquad \tilde{\varphi}_k := \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \exp\left[\frac{2\pi i k n}{N}\right] \varphi_n \,.$$

However, the big difference with the classical Fourier transform is that the Fourier transform *computes* the values of all the Fourier coefficients, while the Fourier gates only *processes* the quantum state. There is no way to compute the value of the Fourier coefficients from a single copy of the quantum state $F|\varphi\rangle$.

13.8 Realizing a measurement on the Fourier basis

Once we know how to realize the unitary gate F, it is easy to find how to measure on the Fourier basis. The trick is just to apply first the Fourier gate F, and then measure on the computational basis. Indeed, we have

$$\begin{aligned} \langle k|F &= \sum_{n=1}^{N} \langle k|e_n\rangle\langle n| \\ &= \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \exp\left[\frac{2\pi i k n}{N}\right] \langle n| \\ &= \langle e_{-k}| \,. \end{aligned}$$

In other words, finding the outcome k for a measurement in the computational basis after the Fourier gate is equivalent to finding the outcome $-k$ for a measurement in the Fourier basis.

In summary, the number of elementary operations needed for the quantum part of Shor's algorithm is

- $O(L)$ to prepare the Fourier basis state $|e_0\rangle$
- $O(L^2)$ to implement the Fourier gate.
- $O(L^3)$ to implement the unitary gate U_f, with $f(x) = a^x \mod N$

[6] More precisely, $O(L^2 \log \epsilon^{-1})$, where ϵ is the precision required in the realization of each two-qubit gate.

After all, the bottleneck is just the calculation of f! Putting everything together, the complexity of factoring numbers on a quantum computer is $O(L^3)$.

Now, if factoring numbers in polynomial time is not possible on a classical computer, as most complexity theorist believe, then we must conclude that quantum mechanics can change the complexity class of a problem. Problems that are hard in the classical world, like factoring, can become easy in the quantum world.

13.9 Chapter summary

In this chapter we first saw how our emails and credit cards are secured: typically, their security is based on the RSA code, which is believed to be very hard to break. Then, we saw that the RSA could be broken by someone how had a magical machines that can efficiently find the period of a desired function. Here is where the surprising quantum algorithm by Shor comes into play: the algorithm shows that quantum mechanics allows us to build the magical "period-finding machine", which allows to break the RSA code and, therefore, the security of emails and credit cards. In order to see this possibility, we showed that the problem of factoring numbers can be reduced to a special problem of period-finding, and saw that for such problem, the magical quantum machine can compute the answer in a time that is only polynomial in the size of the input. In short, when somebody will be able to build a quantum computer, we will have to use new quantum ways to protect our emails and credit cards—for example, using the quantum cryptographic protocols that we saw earlier in this book.

Chapter 14

Epilogue

Im this book we went together from the very basics in quantum mechanics to pretty sophisticate results like quantum data compression, entanglement transformations, and quantum algorithms. Through this journey, you discovered many surprising features of quantum information, like quantum non-locality in the CHSH game, you encountered many elementary quantum machines—machines that try to copy data, machines that transmit data, machines that can be programmed, and machines that correct errors—, and you learnt about banknotes that cannot be copied, passwords that cannot be stolen, and algorithms that can search quickly into a database or crack the RSA code. Looking back to all this, I hope you will feel that learning this chapter of contemporary physics and computer science was worth the effort. And I hope that you had fun, too. But most of all, I hope you have also a feeling of the deep unity that keeps together all the curious features of quantum information, opening new

possibilities that were unthinkable in our old, familiar, classical world.

$\langle\!\langle$ THE | END $\rangle\!\rangle$