

Lecture Notes

**Quantum Systems for Information
Technology — Theory**

Matthias Christandl and Lidia del Rio

October 16, 2012

Contents

1. The qubit	3
1.1. Revisiting the bit	3
1.2. Qubit	3
1.2.1. Definition	3
1.2.2. Measurements	5
1.2.3. Mixed states	7
1.2.4. Operations	10
2. Measurements generalized	12
2.1. Properties and terminology of density operators	12
2.2. Observables	12
2.3. Measurement with respect to a basis	13
2.4. POVMs	14
3. Multiple systems: partial trace and entanglement	16
3.1. Partial trace	16
A. Crash course on Dirac notation	18

1. The qubit

In this chapter we introduce the smallest unit of quantum information, the qubit. Before we get there, though, we should formalize what we already know about the classical counterpart of the qubit, the humble bit.

1.1. Revisiting the bit

The bit is an abstraction of any two-state classical system. Think of a light switch that can be turned on and off, a coin that we can see heads or tails, an emperor's hand showing thumbs up or down, or, let's be wild, an explorer, Amy Bit, who might be either on the north or the south pole. All sort of systems can be used to encode one bit of information: a variable that can take one out of two possible values, $x \in \{0, 1\}$.

Information theory would not be very interesting if it could only speak of known bits, like the switch that we remember having turned on, or the coin that we can see tails up on the table. We want to make predictions about future events, based on our current partial information: for instance, we may toss a coin that we know is biased, or we might know that the emperor had a bad night's sleep and is more likely to give thumbs down. To express this notion of uncertainty about a random bit, we represent it as a **random variable** X . This is just an object that is associated with two things: a set of possible outcomes (0 or 1 in the case of a bit), which we call the **alphabet** (or range) $\mathcal{X} = \{0, 1\}$, and a **probability distribution** P_X that gives us the probability of each outcome. In the case of a fair coin, or any other uniformly random bit, this would just be $P_X(0) = \text{Prob}[X = 0] = \frac{1}{2}$ and $P_X(1) = \frac{1}{2}$.

Finally, what operations can we perform on a single bit? When it comes to reversible operations, all we can do is flip the bit. In other words, we can apply a NOT operation (or gate), which maps 0 to 1 and vice-versa. To reverse it, we only need to apply the same gate again. In the case of random qubits, the NOT gate flips the probability distribution of outcomes. For instance, say you have a biased coin, with probability 70% of giving tails. If you close your eyes, toss the coin and then flip it, you know you have a 70% chance of facing heads.

1.2. Qubit

1.2.1. Definition

[tl;dr: bit + superposition = qubit.]

The qubit is just a generalization of the bit. We start by defining two states, which correspond to the 0 and 1 of a bit. Now, however, we say that these states correspond

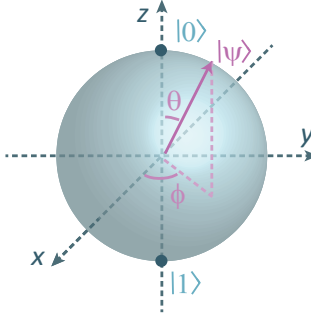
to vectors in \mathbb{C}^2 :

$$0 \rightarrow |0\rangle := \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad 1 \rightarrow |1\rangle := \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Now we postulate that a qubit can also be in any superposition of these two states,

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \quad |\alpha|^2 + |\beta|^2 = 1. \quad (1.1)$$

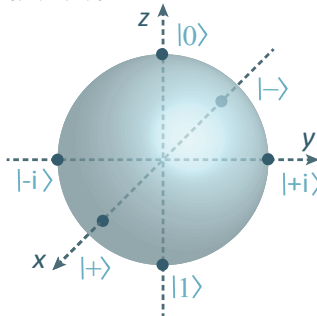
Because the coefficients are normalized, we can express them in terms of angles,

$$|\psi\rangle = e^{i\gamma} \left[\cos \frac{\theta}{2} |0\rangle + e^{i\phi} \sin \frac{\theta}{2} |1\rangle \right].$$


For reasons that we will see soon, the overall phase γ is irrelevant. Since the state of the qubit depends on two angles, $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi[$, we can represent it as a point on a sphere. We call this the Bloch sphere representation. States $|0\rangle$ and $|1\rangle$ sit on the North and South poles of the Bloch sphere, along the z -axis. Amy Bit was a polar explorer (she had to be $|0\rangle$ or $|1\rangle$), but Amy Qubit can be in any point of the globe.

Other points on the sphere come up often enough to get their own names. On the x -axis, we have states $|+\rangle$ and $|-\rangle$, and on the y -axis, states $|+i\rangle$ and $|-i\rangle$. Every pair of antipodes on the Bloch sphere forms an orthonormal basis for the vector space of a qubit. For instance, we have that $\langle +|+\rangle = \langle -|-\rangle = 1$, $\langle +|-\rangle = 0$, and we could describe the state of any qubit as $|\psi\rangle = \alpha'|+\rangle + \beta'|-\rangle$. Out of convention, we usually write down states on the so-called computational basis, $\{|0\rangle, |1\rangle\}$.

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

$$|\pm i\rangle = \frac{|0\rangle \pm i|1\rangle}{\sqrt{2}}.$$


If you have learned some quantum mechanics, this setting should sound familiar. In fact, a qubit can be implemented as a two-level quantum system, like a spin- $\frac{1}{2}$ particle or a polarized photon. The state $|0\rangle$ could be “spin down”, or “horizontal polarization”. The superposition of states of a qubit is also directly inspired by the superposition principle of quantum theory.

A few notes about the Bloch sphere: the factor of $\frac{1}{2}$ in the angle θ is there to make $\theta \in [0, \pi]$. Instead of the two angles, we can also describe any point on the sphere with a vector \mathbf{r} with origin in the centre of the sphere. We call this the Bloch vector.

So what can we do with a qubit? A bit could encode a single unit of information (0 or 1), but it sounds like we might encode infinite information on a qubit: we imposed no restrictions on the two angles, which may take any real value within that interval. We can take a long letter, write it as two real numbers, and use them to define the two angles of the qubit. Now we have, say, a photon whose polarization corresponds to our long letter. Seems a little over-powerful for an object described as the minimal unit of quantum information. As it happens, there is a catch. Although we may *prepare* any state we want, we can only *read* one out of two outcomes.

Before we proceed: if you are not familiar with Dirac notation (kets and bras) and inner products, read Appendix A. The short version is: a ket $|\psi\rangle$ is a column vector; a bra $\langle\phi|$ is the complex-conjugate of ket $|\phi\rangle$ (so a row vector); inner products are given by $\langle\phi|\psi\rangle$, and matrices by $\sum_i a_{ij}|\psi_i\rangle\langle\psi_j|$.

1.2.2. Measurements

[tl;dr: Measuring a qubit is projecting it on an axis of the Bloch sphere; measurement statistics as $\text{Prob}[\phi]_\psi = |\langle\phi|\psi\rangle|^2$; qubit collapses to state detected.]

When reading a bit we can obtain either 0 or 1. With qubits, we still only obtain one bit of information, but with a fundamental difference: we can choose the basis of measurement.

First, we have to define quantum measurements. We postulate that the probability of reading out a given state (like $|0\rangle$ or any other state $|\phi\rangle$) when measuring a qubit in state $|\psi\rangle$ is given by the overlap between the two states, measured by

$$\text{Prob}[\phi]_\psi = |\langle\phi|\psi\rangle|^2. \quad (1.2)$$

For instance, if we try to read an arbitrary state $|\psi\rangle$ in the computational basis, the probabilities come

$$\begin{aligned} |\psi\rangle &= \cos\frac{\theta}{2}|0\rangle + e^{i\phi}\sin\frac{\theta}{2}|1\rangle = \alpha|0\rangle + \beta|1\rangle, \\ \text{Prob}[0]_\psi &= \left| \langle 0 | \left(\cos\frac{\theta}{2}|0\rangle + e^{i\phi}\sin\frac{\theta}{2}|1\rangle \right) \right|^2 = \cos^2\frac{\theta}{2} = |\alpha|^2, \\ \text{Prob}[1]_\psi &= \left| \langle 1 | \left(\cos\frac{\theta}{2}|0\rangle + e^{i\phi}\sin\frac{\theta}{2}|1\rangle \right) \right|^2 = \sin^2\frac{\theta}{2} = |\beta|^2. \end{aligned}$$

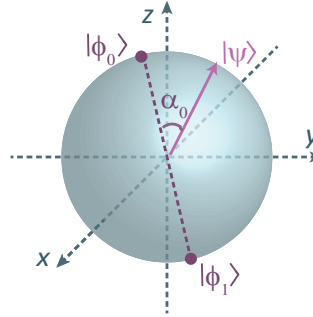
This why we normalize the qubit states as we do: we want

$$1 = \text{Prob}[0]_\psi + \text{Prob}[1]_\psi = |\alpha|^2 + |\beta|^2. \quad (1.3)$$

We could also choose any other orthonormal basis to measure a qubit. In other words, we can take two antipodes on the Bloch sphere, which form a basis $\{|\phi_0\rangle, |\phi_1\rangle\}$, and project the Bloch vector of our qubit onto them¹ to obtain the probability of each outcome,

¹using $\frac{1}{2}$ of the overlap angle in the Bloch sphere

$$\begin{aligned}\text{Prob}[\phi_0]_\psi &= |\langle\phi_0|\psi\rangle|^2 = \cos^2\left(\frac{\alpha_0}{2}\right), \\ \text{Prob}[\phi_1]_\psi &= |\langle\phi_1|\psi\rangle|^2 = \cos^2\left(\frac{\frac{\pi}{2} - \alpha_0}{2}\right), \\ \text{Prob}[\phi_0]_\psi + \text{Prob}[\phi_1]_\psi &= 1.\end{aligned}$$



Now we can see why the overall phase γ of a qubit does not matter: it does not change the measurement statistics, as

$$|\langle\phi|e^{i\gamma}\psi\rangle|^2 = e^{i\gamma}e^{-i\gamma}|\langle\phi|\psi\rangle|^2 = |\langle\phi|\psi\rangle|^2. \quad (1.4)$$

Let us look at one more example to see how important the choice of basis is. Say that you prepare the qubit $|\psi\rangle = |+\rangle$. What happens if you measure it in the basis $\{|+\rangle, |-\rangle\}$? Naturally, the probability of obtaining $|+\rangle$ is 1:

$$\begin{aligned}\text{Prob}[+]_+ &= |\langle+|+\rangle|^2 = 1, \\ \text{Prob}[-]_+ &= |\langle-|+\rangle|^2 = 0.\end{aligned}$$

However, if we measure the same qubit in the computational basis, we are equally likely to obtain $|0\rangle$ or $|1\rangle$,

$$\begin{aligned}\text{Prob}[0]_+ &= |\langle 0|+\rangle|^2 = \left|\langle 0|\left(\frac{|0\rangle + |1\rangle}{\sqrt{2}}\right)\right|^2 = \frac{1}{2}, \\ \text{Prob}[1]_+ &= |\langle 1|+\rangle|^2 = \left|\langle 1|\left(\frac{|0\rangle + |1\rangle}{\sqrt{2}}\right)\right|^2 = \frac{1}{2}.\end{aligned}$$

What happens to a qubit after it was measured? If we read a bit and see a 0, then close our eyes and look again, it remains a 0. The same happens with qubits: if we measure a qubit $|\psi\rangle$ in a basis $\{|\phi_0\rangle, |\phi_1\rangle\}$ and obtain $|\phi_0\rangle$, then measure it again in the same basis, we will observe the same state $|\phi_0\rangle$ with probability 1. For all purposes, we can say that the state of the qubit is now $|\phi_0\rangle$. The infamous collapse of the wave function is as simple as that: the post-measurement state of a qubit corresponds to what was measured. We will talk about this again later in the lecture.

We can rewrite the measurement statistics with the help of the matrix trace,

$$\text{Prob}[\phi]_\psi = \text{Tr}(|\phi\rangle\langle\phi| |\psi\rangle\langle\psi|). \quad (1.5)$$

To see this, recall that the trace is the sum of the diagonal elements of a matrix, with

respect to any basis $\{|i\rangle\}_i$, $\text{Tr}(A) = \sum_i \langle i|A|i\rangle$.² We have

$$\begin{aligned} \text{Tr}(|\phi\rangle\langle\phi| |\psi\rangle\langle\psi|) &= \sum_i \langle i|\phi\rangle\langle\phi|\psi\rangle\langle\psi|i\rangle \\ &= \langle\phi|\psi\rangle \sum_i \langle\psi|i\rangle\langle i|\phi\rangle \\ &= \langle\phi|\psi\rangle\langle\psi| \underbrace{\left(\sum_i |i\rangle\langle i|\right)}_{\mathbb{1}} |\phi\rangle \\ &= \langle\phi|\psi\rangle\langle\psi|\phi\rangle = |\langle\phi|\psi\rangle|^2. \end{aligned}$$

Expressing the measurement statistics as a trace will be useful in a minute, when we introduce random qubits.

1.2.3. Mixed states

Like in the case of random bits, sometimes we don't have complete information about the state of a qubit. For instance, you may have a flawed source of qubits, which was supposed to produce state $|\psi\rangle$, but may have a malfunction with probability p , producing state $|\tau\rangle$ instead. What is the probability of obtaining $|x\rangle$ when you measure your unknown state in a basis $\{|x\rangle\}_x$? Well, it has to be [probability of having $|\psi\rangle$] * [probability of measuring $|x\rangle$ on $|\psi\rangle$] + [probability of having $|\tau\rangle$] * [probability of measuring $|x\rangle$ on $|\tau\rangle$],

$$\begin{aligned} \text{Pr}(x) &= (1-p)\text{Pr}(x)_\psi + p\text{Pr}(x)_\tau \\ &= (1-p)\text{Tr}(|x\rangle\langle x| |\psi\rangle\langle\psi|) + p\text{Tr}(|x\rangle\langle x| |\tau\rangle\langle\tau|) \\ &= \text{Tr}\left((1-p)|x\rangle\langle x| |\psi\rangle\langle\psi| + p|x\rangle\langle x| |\tau\rangle\langle\tau|\right) \\ &= \text{Tr}\left(|x\rangle\langle x| \underbrace{\left[(1-p)|\psi\rangle\langle\psi| + p|\tau\rangle\langle\tau|\right]}_{=: \rho}\right) \\ &= \text{Tr}(|x\rangle\langle x| \rho). \end{aligned} \tag{1.6}$$

Here, we defined the density operator $\rho = (1-p)|\psi\rangle\langle\psi| + p|\tau\rangle\langle\tau|$. This is a matrix that reflects our ignorance about the actual state of the system. For instance, if $|\psi\rangle = |1\rangle$ and $|\tau\rangle = |-\rangle$, we have

$$\begin{aligned} \rho &= (1-p)|1\rangle\langle 1| + p|-\rangle\langle -| \\ &= (1-p) \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} + p \frac{1}{2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} \frac{p}{2} & -\frac{p}{2} \\ -\frac{p}{2} & 1 - \frac{p}{2} \end{pmatrix}. \end{aligned}$$

²The trace is **linear**, $\alpha\text{Tr}(A) + \beta\text{Tr}(B) = \text{Tr}(\alpha A + \beta B)$; **cyclic**, $\text{Tr}(ABC) = \text{Tr}(CAB)$; and **basis-independent**, i.e., invariant under unitary transformations, $\text{Tr}(A) = \text{Tr}(UAU^\dagger)$, for a unitary matrix U .

The probability of obtaining a “1” when measuring this state in the computational basis is

$$\begin{aligned}\Pr(1)_\rho &= \text{Tr}[|1\rangle\langle 1| \rho] \\ &= \text{Tr} \left[\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{p}{2} & -\frac{p}{2} \\ -\frac{p}{2} & 1 - \frac{p}{2} \end{pmatrix} \right] \\ &= \text{Tr} \left[\begin{pmatrix} 0 & 0 \\ -\frac{p}{2} & 1 - \frac{p}{2} \end{pmatrix} \right] = 1 - \frac{p}{2}.\end{aligned}$$

Note that we could also have computed this quantity the old way,

$$\begin{aligned}\Pr(1)_\rho &= (1 - p) \Pr(1)_{|1\rangle} + p \Pr(1)_{|-\rangle} \\ &= (1 - p) 1 + p \frac{1}{2} = 1 - \frac{p}{2}.\end{aligned}$$

In practice, however, it is much more convenient to use density matrices. Density operators are the quantum generalization of probability distributions over bits.

The Bloch ball

In the Bloch representation, pure states lie on the surface of a ball. Mixed states are inside the ball. In fact, if two pure states $|\tau\rangle, |\psi\rangle$ have the Bloch vectors \mathbf{t}, \mathbf{p} , respectively, then the Bloch vector of mixed state $\rho = p|\tau\rangle\langle\tau| + (1 - p)|\psi\rangle\langle\psi|$ is the weighted average of \mathbf{t} and \mathbf{p} : $\mathbf{r} = p\mathbf{r} + (1 - p)\mathbf{t}$.

[insert pic]

In general, any qubit can be written as

$$\rho = \frac{1}{2}(\mathbb{1} + \mathbf{r} \cdot \boldsymbol{\sigma}), \quad (1.7)$$

where $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ and $\mathbf{r} = (r_x, r_y, r_z)$, $|\mathbf{r}| \leq 1$ is the state’s Bloch vector, that gives us the position of a point in the unit ball. The matrices $\sigma_x, \sigma_y, \sigma_z$ are called the Pauli matrices, and are given by

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (1.8)$$

The length of the Bloch vector tells us how mixed the state is: pure states have $|\mathbf{r}| = 1$, and mixtures have shorter vectors, all the way to the fully mixed state that lies in the centre of the ball, $\rho = \frac{\mathbb{1}}{2} = \frac{1}{2}(|\phi_0\rangle\langle\phi_0| + |\psi_1\rangle\langle\psi_1|)$ for any orthonormal basis $\{|\phi_0\rangle, |\phi_1\rangle\}$.

Superposition vs mixture

It is easy to confuse mixtures of quantum states with quantum superpositions in the beginning. For instance, you may be wondering about the difference between the states

$$|+\rangle = \frac{|0\rangle + |1\rangle}{\sqrt{2}} \quad \text{and} \quad \rho = \frac{|0\rangle\langle 0| + |1\rangle\langle 1|}{2}.$$

In a nutshell, the difference is that the former is a pure state (you know that the system is in that exact state), while the latter is a probabilistic mixture of two possible states. To clarify things, we can compute the density operators corresponding to both states, in two different bases,

$$\begin{aligned}
\sigma &= |+\rangle\langle +| \\
&= \frac{|0\rangle + |1\rangle}{\sqrt{2}} \frac{\langle 0| + \langle 1|}{\sqrt{2}} \\
&= \frac{|0\rangle\langle 0| + |0\rangle\langle 1| + |1\rangle\langle 0| + |1\rangle\langle 1|}{2}, \\
\rho &= \frac{|0\rangle\langle 0| + |1\rangle\langle 1|}{2} \\
&= \frac{1}{2} \left(\frac{|+\rangle + |-\rangle}{\sqrt{2}} \frac{\langle +| + \langle -|}{\sqrt{2}} + \frac{|+\rangle - |-\rangle}{\sqrt{2}} \frac{\langle +| - \langle -|}{\sqrt{2}} \right) \\
&= \frac{|+\rangle\langle +| + |-\rangle\langle -|}{2},
\end{aligned}$$

basis	$\{ 0\rangle, 1\rangle\}$	$\{ +\rangle, -\rangle\}$
σ	$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$
ρ	$\begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$	$\begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$.

So these two states have different density matrices. We can also see a difference in the Bloch representation of the two states: $|+\rangle$, as a pure state, sits on the surface of the ball ($|\mathbf{r}| = 1$), while ρ is precisely in the centre ($|\mathbf{r}| = 0$). But does that have any operational meaning? For instance, would we get different results if we measured the two states? Let's check. If you measure each state in basis $\{|0\rangle, |1\rangle\}$, the probability of obtaining $|0\rangle$ would be

$$\begin{aligned}
\Pr(0)_\sigma &= \text{Tr}[|0\rangle\langle 0| \sigma] = \text{Tr} \left[\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \right] = \frac{1}{2}, \\
\Pr(0)_\rho &= \text{Tr}[|0\rangle\langle 0| \rho] = \text{Tr} \left[\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \right] = \frac{1}{2}.
\end{aligned}$$

This does not look very promising: we could not distinguish the two states solely from the statistics of this measurement. But what happens if instead we measure them in

basis $\{|+\rangle, |-\rangle\}$? The probability of obtaining $|+\rangle$ is, for each of them,

$$\begin{aligned}\Pr(+)_\sigma &= \text{Tr}[|+\rangle\langle+| \sigma] = \text{Tr} \left[\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \right] = 1, \\ \Pr(+)_\rho &= \text{Tr}[|+\rangle\langle+| \rho] = \text{Tr} \left[\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \right] = \frac{1}{2}.\end{aligned}$$

The states have different measurement statistics in this basis. If you have to find out whether you have fifty copies of $|+\rangle$ or fifty copies of ρ , you can just measure all your systems in basis $\{|+\rangle, |-\rangle\}$. If you get at least one outcome “–”, then you know that you had state ρ . Question for bonus points: what if you are only given one copy of the state?

Here is another difference between the two states: $|+\rangle$ is a pure state, which means that you can in principle transform it into any other pure state via unitary evolution, while $\rho = \frac{1}{2}$ is fully mixed, and invariant under unitary transformations. See, for instance, what happens to each state under the following change of basis,

$$\begin{aligned}U &= |+\rangle\langle 0| + |0\rangle\langle +| + |1\rangle\langle -| + |-\rangle\langle 1| = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \\ U|+\rangle &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = |0\rangle, \\ U\rho U^\dagger &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} = \frac{1}{2}.\end{aligned}$$

1.2.4. Operations

With one classical bit, there is only one reversible operation available: the NOT gate. In the Bloch representation, we can implement that gate by rotating the ball clockwise around the x axis by π , so that $|0\rangle$ is mapped to $|1\rangle$ and vice-versa. To reverse the operation, we just need to rotate the ball anticlockwise by another π . More generally, all reversible operations on qubits are represented by rotations of the Bloch ball, which correspond to unitary matrices acting on the vector space of a qubit: $U|\psi\rangle$. Note that we have much more freedom than in the classical case: we can implement even tiny rotations.

Operations on mixed qubits

Now imagine that you are again given a state $|\psi\rangle$ or $|\tau\rangle$ at random, with probabilities $1-p$ and p respectively. We saw that you can represent your knowledge of the qubit via a mixed state $\rho = (1-p)|\psi\rangle\langle\psi| + p|\tau\rangle\langle\tau|$. But what happens to the density operator when you apply an operation U on the qubit? Well, we know that after applying U , the qubit is either in state $U_t|\psi\rangle$, with probability $1-p$, or in state $U_t|\tau\rangle$, with probability p . Imagine that now you want to measure it in some basis $\{|x\rangle\}_x$. The probability of

reading $|x\rangle$ after the measurement is given by

$$\begin{aligned}
\text{Prob}[x] &= (1-p)\text{Prob}[x]_{U|\psi\rangle} + p\text{Prob}[x]_{U|\tau\rangle} \\
&= (1-p)\text{Tr}[|x\rangle\langle x| U|\psi\rangle\langle\psi|U^\dagger] + p\text{Tr}[|x\rangle\langle x| U|\tau\rangle\langle\tau|U^\dagger] \\
&= \text{Tr}\left[|x\rangle\langle x| U \underbrace{((1-p)|\psi\rangle\langle\psi| + p|\tau\rangle\langle\tau|)}_{\rho} U^\dagger\right] \\
&= \text{Tr}\left[|x\rangle\langle x| \underbrace{U\rho U^\dagger}_{=:\tilde{\rho}}\right].
\end{aligned}$$

We can therefore express the density operator after applying an operation U as

$$\tilde{\rho} := U \cdot \rho := U \rho(0) U^\dagger. \quad (1.9)$$

Now, this is more relevant than it might appear at first sight. Think about what you can do to qubits in a lab: you can apply unitary operations, and you can measure them.³ We have seen that both the evolution under a unitary operation and measurement statistics of a system depend only on the density operator. This implies that it is impossible to distinguish two systems with the same density operator, even if they were produced in different ways. Let me give you a somewhat trivial example: suppose that your friend Alice has two machines that produce polarized photons. The only problem is that those machines are not reliable at all: machine A produces state $|0\rangle$ with probability $1/2$ and state $|1\rangle$ with probability $1/2$, while machine B produces state $|+\rangle$ with probability $1/2$ and state $|-\rangle$ with probability $1/2$. The density operators of a state coming from each machine are

$$\begin{aligned}
\rho_A &= \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \frac{\mathbb{1}}{2}, \\
\rho_B &= \frac{1}{2}|+\rangle\langle +| + \frac{1}{2}|-\rangle\langle -| = \frac{1}{2}\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} + \frac{1}{2}\begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix} = \frac{\mathbb{1}}{2}.
\end{aligned}$$

Now imagine that Alice gives you a million photons, and tells you that she used the same machine to produce all of them. But she won't tell you which machine. You would think that it would be easy to find out, right? Here are a million photons, either half are $|0\rangle$ and half are $|1\rangle$ or half are $|+\rangle$ and half are $|-\rangle$. You can do whatever you want to your photons: rotate the states, measure them, let them evolve under convoluted Hamiltonians, make them interact with an external system, let your local soothsayer examine them. And yet you will never be able to determine which machine was used, because photons from A and from B have the same density matrix. The origins of a quantum state do not matter.⁴

³Time evolution is just a special case of a unitary operation, $U = e^{-iHt}$, where H is the Hamiltonian of the system

⁴This proves particularly handy in quantum cryptography.

2. Measurements generalized

There are a few conventions on quantum measurements. We will go through them with simple examples. Before we start, though, let us just generalize what we know about density operators.

2.1. Properties and terminology of density operators

The following applies to any Hilbert space \mathcal{H} of finite dimension N .

- Density operators are endomorphisms on a Hilbert space, $\rho : \mathcal{H} \rightarrow \mathcal{H}$.
- Density matrices can be diagonalized as

$$\rho = \sum_{i=1}^N p_i |\psi_i\rangle\langle\psi_i|, \quad (2.1)$$

where the eigenvalues $\{p_i\}$ form a probability distribution (meaning $\forall p_i \geq 0$ and $\sum_i p_i = 1$), and the eigenstates $\{|\psi_i\rangle\}$ form a basis of \mathcal{H} . This has the following physical interpretation: the system may be in quantum state $|\psi_i\rangle$ with probability p_i .

- Since the eigenvalues of ρ form a probability distribution, $\text{Tr}(\rho) = \sum_i p_i = 1$, and ρ is positive semi-definite, and Hermitian.
- A system is said to be in a *pure* state if the corresponding density operator only has one non-zero eigenvalue, $\{p_i\} = \{1, 0, 0 \dots\}$, and therefore $\rho = |\psi\rangle\langle\psi|$. Otherwise a state is considered *mixed* (meaning that there is more than one possibility for the exact state of the system).
- If $\rho = \mathbb{1}/N$, it is called *fully mixed*.

2.2. Observables

In traditional quantum mechanics courses, people talk about observables all the time. An observable is an operator that represents a measurement. These operators are of the form $O = \sum_x x P_x$, where

1. $\{x\}_x$ are the possible outcomes of the measurement (for instance if you are measuring a distance then you could have $\{x\} = \{10m, 23km, 3cm\}$);

2. $\{P_x\}_x$ are the projectors (operators) that determine the statistics of the measurement: the probability of having outcome x when performing a measurement on state ρ is given by $\text{Tr}(P_x\rho)$, and the state “collapses” to $\rho_x = (P_x\rho P_x)/\text{Tr}(P_x\rho)$;
3. sometimes the name of the observable gives us a hint of what it measures (like calling it P if it measures momentum), but many times we just stick to O ;
4. note that because $\{x\}_x$ are the outcomes (what you read in your machine when you measure a state), O is in general not normalised. In fact, the x_x don’t even need to be numbers, but more like labels to the different outcomes (see ahead with the giraffes);
5. $\sum_x P_x = \mathbb{1}$.

This is why we say we “measure an observable” or “perform a measurement represented by an observable” (maybe even “with respect to the observable”), because an observable has everything that is needed to identify and specify a measurement.

2.3. Measurement with respect to a basis

Consider one qubit in state ρ , and a scientist, Alice, who wants to measure it in the computational basis $\{|0\rangle, |1\rangle\}$. The probabilities of obtaining each outcome are given by

$$P_0 = \text{Tr}(|0\rangle\langle 0|\rho),$$

$$P_1 = \text{Tr}(|1\rangle\langle 1|\rho),$$

as we saw before. If $\rho = a|0\rangle\langle 0| + b|1\rangle\langle 1| + c|0\rangle\langle 1| + d|1\rangle\langle 0|$, then

$$P_0 = \text{Tr}\left(a|0\rangle\langle 0|0\rangle\langle 0| + b|0\rangle\langle 0|1\rangle\langle 1| + c|0\rangle\langle 0|0\rangle\langle 1| + d|0\rangle\langle 0|1\rangle\langle 0|\right)$$

$$= a$$

$$P_1 = \text{Tr}\left(a|1\rangle\langle 1|0\rangle\langle 0| + b|1\rangle\langle 1|1\rangle\langle 1| + c|1\rangle\langle 1|0\rangle\langle 1| + d|1\rangle\langle 1|1\rangle\langle 0|\right)$$

$$= b$$

For instance, if ρ is the pure state $|1\rangle\langle 1|$, then $P_0 = 0$ and $P_1 = 1$, which means that if she performs a measurement in the computational basis she will always obtain $|1\rangle$.

If we were to represent this measurement as an observable, it could be $O = x_0|0\rangle\langle 0| + x_1|1\rangle\langle 1|$, where x_0 is what Alice sees in her machine when the state is projected to $|0\rangle$ and x_1 what she sees when it goes to $|1\rangle$. Here it is clear why the $\{x\}_x$ are not “numbers”: if we used 0 and 1, then we would have $O = |1\rangle\langle 1|$, which is not very representative of our measurement!

2.4. POVMs

POVMs (positive operator valued measures) are a more general way of representing a measurement. We forget about what Alice sees in her machine (she can have the machine just saying “*the state collapsed to $|0\rangle$* ” instead of “*I got x_0* ”; she could have the machine saying “*elephant*” for 0 and “*giraffe*” for 1, for that matter) and focus only on the projectors that determine the statistics of the measurement—the important part.

In our previous example, the POVM that represents Alice’s measurement is just the set of projectors $M = \{|0\rangle\langle 0|, |1\rangle\langle 1|\}$.

Now suppose that Alice’s machine does not work very well: with a small probability p it won’t measure the state at all. Then, when she measures the state ρ , three things can happen:

1. with probability p the measurement fails (the machine says “*I failed*”) and the final state is

$$\rho_{\text{fail}} = \rho = \mathbb{1}\rho\mathbb{1};$$

2. with probability $1 - p$ the measurement works, and then

- with probability $(1 - p)\text{Tr}(|0\rangle\langle 0|\rho)$ she measures “ x_0 ” and the state goes to

$$\rho_0 = \frac{|0\rangle\langle 0|\rho|0\rangle\langle 0|}{(1 - p)\text{Tr}(|0\rangle\langle 0|\rho)} = |0\rangle\langle 0|;$$

- with probability $(1 - p)\text{Tr}(|1\rangle\langle 1|\rho)$ she measures “ x_1 ” and the state goes to

$$\rho_1 = \frac{|1\rangle\langle 1|\rho|1\rangle\langle 1|}{(1 - p)\text{Tr}(|1\rangle\langle 1|\rho)} = |1\rangle\langle 1|.$$

All in all there are three possible outcomes, which may be represented by an observable

$$\begin{aligned} O' &= p \text{ “fail”} + (1 - p) * O \\ &= \text{“fail”} \ p \ \mathbb{1} \\ &\quad + x_0 (1 - p) |0\rangle\langle 0| \\ &\quad + x_1 (1 - p) |1\rangle\langle 1|, \end{aligned}$$

or by a POVM

$$M' = \{p \ \mathbb{1}, \ (1 - p) |0\rangle\langle 0|, \ (1 - p) |1\rangle\langle 1|\}.$$

The difference is just that in the POVM we get rid of the labels. Formally, a POVM is a set of positive operators $\{M_x\}_x$ that sum up to the identity operator: $\sum_x M_x = \mathbb{1}$. Each operator M_x corresponds to a different outcome, x .

A measurement with respect to a basis is simply a special case of a POVM with a perfect machine that can distinguish all the orthogonal states in a basis perfectly. Explicitly, it is a POVM $\{|x\rangle\langle x|\}_x$, where $\{|x\rangle\}_x$ for a basis of our Hilbert space.

Collapse with POVMs

Imagine that you have a four-dimensional Hilbert space (two qubits), with basis $\{|1\rangle, |2\rangle, |3\rangle, |4\rangle\}$. You also have a measurement device, but not a very good one: it only distinguishes the first two from the last two states. It has two measurement outcomes, “<” for $|1\rangle$ and $|2\rangle$, and “>” for $|3\rangle$ and $|4\rangle$. The probability of obtaining outcome < when measuring a state ρ is given by

$$\begin{aligned} \text{Prob} [<]_\rho &= \text{Prob} [1 \vee 2]_\rho \\ &= \text{Prob} [1]_\rho + \text{Prob} [2]_\rho \\ &= \text{Tr}(|1\rangle\langle 1| \rho) + \text{Tr}(|2\rangle\langle 2| \rho) \\ &= \text{Tr}\left(\underbrace{(|1\rangle\langle 1| + |2\rangle\langle 2|)}_{P_<} \rho\right). \end{aligned}$$

We can represent this measurement with a POVM $\{P_<, P_>\}$, whose elements are projectors,

$$\begin{aligned} P_< &= |1\rangle\langle 1| + |2\rangle\langle 2|, & \text{Prob} [<]_\rho &= \text{Tr}(P_< \rho), \\ P_> &= |3\rangle\langle 3| + |4\rangle\langle 4|, & \text{Prob} [>]_\rho &= \text{Tr}(P_> \rho). \end{aligned}$$

and, not surprisingly, people call it a *projective measurement*. What happens to the system after a measurement with outcome “<”? What matters here is that, if we perform the same measurement again, we obtain the same outcome, “<”. This means that the state can go to either $|1\rangle$ or $|2\rangle$, we just don’t know which. In other words, it becomes a probabilistic mixture of $|1\rangle$ and $|2\rangle$,

$$\begin{aligned} \rho^{<} &= \frac{\text{Prob} [1]_\rho |1\rangle\langle 1| + \text{Prob} [2]_\rho |2\rangle\langle 2|}{\text{Prob} [1]_\rho + \text{Prob} [2]_\rho} \\ &= \frac{\text{Tr}(|1\rangle\langle 1| \rho) |1\rangle\langle 1| + \text{Tr}(|2\rangle\langle 2| \rho) |2\rangle\langle 2|}{\text{Tr}(P_< \rho)} \\ &= \frac{\langle 1|\rho|1\rangle |1\rangle\langle 1| + \langle 2|\rho|2\rangle |2\rangle\langle 2|}{\text{Tr}(P_< \rho)} \\ &= \frac{|1\rangle\langle 1|\rho|1\rangle\langle 1| + |2\rangle\langle 2|\rho|2\rangle\langle 2|}{\text{Tr}(P_< \rho)} \\ &= \frac{(|1\rangle\langle 1| + |2\rangle\langle 2|)\rho(|1\rangle\langle 1| + |2\rangle\langle 2|)}{\text{Tr}(P_< \rho)} \\ &= \frac{P_< \rho P_<}{\text{Tr}(P_< \rho)}. \end{aligned}$$

This generalizes to any POVM $\{M_x\}_x$,

$$\text{Prob} [x]_\rho = \text{Tr}(M_x \rho), \quad \rho^x = \frac{M_x \rho M_x^\dagger}{\text{Tr}(M_x \rho)}.$$

3. Multiple systems: partial trace and entanglement

3.1. Partial trace

Consider a composed system $\mathcal{H}_A \otimes \mathcal{H}_B$. Any state of that system can be expressed as a density matrix. For instance, in the basis $\{|a_i\rangle \otimes |b_j\rangle\}_{i,j}$, we have

$$\begin{aligned}\rho_{AB} &= \sum_{i,j} \sum_{k,l} c_{ij}^{kl} (|a_i\rangle \otimes |b_j\rangle) (\langle a_k| \otimes \langle b_l|) \\ &= \sum_{i,k} \sum_{j,l} c_{ij}^{kl} |a_i\rangle \langle a_k| \otimes |b_j\rangle \langle b_l|.\end{aligned}$$

Notation: We write composed states of the form $|x\rangle \otimes |m\rangle$ as $|x\rangle|m\rangle$ or simply $|xm\rangle$.

The (usually mixed) state of one of the subsystems can be obtained by *tracing out* the other. In practice, this means we ignore the other system (for instance, Alice may not have access to Bob's system). This is done by means of a *partial trace*,

$$\rho_A = \text{Tr}_B(\rho_{AB}). \quad (3.1)$$

The resulting *reduced density matrix* of subsystem A is

$$\rho_A = \sum_{i,k} \sum_j c_{ij}^{kj} |a_i\rangle \langle a_k|. \quad (3.2)$$

If systems A and B are independent, i.e., if ρ_{AB} is a product state, $\rho_{AB} = \rho_A \otimes \rho_B$, we have $\text{Tr}_B(\rho_A \otimes \rho_B) = \rho_A \text{Tr}(\rho_B) = \rho_A$, since density matrices are normalized.

Let us look at an example. Consider a system with two qubits A and B , prepared in the global pure state

$$|\phi\rangle = \frac{1}{\sqrt{3}} (|00\rangle + |10\rangle + |11\rangle).$$

The density matrix of the composed system is given by

$$\begin{aligned}\rho_{AB} &= |\Phi_{AB}\rangle \langle \Phi_{AB}| \\ &= \frac{1}{3} \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix}.\end{aligned}$$

This matrix has eigenvalues $\{1, 0, 0, 0\}$; the only non-zero eigenvalue corresponds to the eigenstate $|\phi\rangle$.

To obtain the reduced density matrix of the subsystem A , one has to evaluate the sum of Eq. 3.2. For instance, the coefficient α_0^1 that corresponds to the mixture $|0_A\rangle\langle 1_A|$ (in blue ahead) is given by the sum of the coefficients of the terms corresponding to $|0_A0_B\rangle\langle 1_A0_B|$ and $|0_A1_B\rangle\langle 1_A1_B|$ of the original density matrix. In the basis $\{|0_A\rangle, |1_A\rangle\}$, the reduced state is represented by the matrix

$$\rho_A = \frac{1}{3} \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}.$$

The colours indicate the elements of the ρ_{AB} that were summed up to each of the entries of ρ_A . Since the basis of the global system was nicely ordered, each element of the new one was calculated by the trace of the 2×2 “submatrix” that is in the corner of the original matrix indicated by the position of the desired element in the new one. The resulting state is mixed, as the reduced density matrix has two non-zero eigenvalues.

A. Crash course on Dirac notation

A Hilbert space \mathcal{H} is a vector space with a well-behaved inner product: for any $\phi, \psi \in \mathcal{H}$, the inner product $\langle \phi, \psi \rangle = \langle \psi, \phi \rangle^* < \infty$. In this course, we restrict ourselves to finite spaces, like a finite number of qubits.

Formally, a ket is a function

$$\begin{aligned} |\psi\rangle &: \mathbb{C} \rightarrow \mathcal{H} \\ \alpha &\mapsto \alpha \psi, \quad \psi \in \mathcal{H}. \end{aligned} \tag{A.1}$$

In practice, we usually identify the ket with the element of the Hilbert space. Let $\{\phi_i\}_i$ be an orthonormal basis of the Hilbert space, with corresponding kets $\{|i\rangle\}_i$ (we drop the ϕ in kets for simplicity). Just like we can expand any element $\psi \in \mathcal{H}$ in basis $\{\phi_i\}_i$, we can do it for any ket $|\psi\rangle$,

$$|\psi\rangle = \sum_i \langle \phi_i, \psi \rangle |i\rangle. \tag{A.2}$$

We represent this as a vertical vector in basis $\{|i\rangle\}_i$,

$$|\psi\rangle = \begin{pmatrix} \langle \phi_0, \psi \rangle \\ \langle \phi_1, \psi \rangle \\ \vdots \\ \langle \phi_N, \psi \rangle \end{pmatrix}, \tag{A.3}$$

such that row i corresponds to ket $|i\rangle$.

A bra is also defined as a function

$$\begin{aligned} \langle \phi| &: \mathcal{H} \rightarrow \mathbb{C} \\ \psi &\mapsto \langle \phi, \psi \rangle. \end{aligned} \tag{A.4}$$

However, we can simply represent bras as horizontal vectors,

$$\langle \tau| = (\langle \tau, \phi_0 \rangle, \langle \tau, \phi_1 \rangle, \dots, \langle \tau, \phi_N \rangle) \tag{A.5}$$

Indeed, concatenating a bra and a ket gives us an inner product,

$$\begin{aligned} \langle \tau| |\psi\rangle &= (\langle \tau, \phi_0 \rangle, \langle \tau, \phi_1 \rangle, \dots, \langle \tau, \phi_N \rangle) \begin{pmatrix} \langle \phi_0, \psi \rangle \\ \langle \phi_1, \psi \rangle \\ \vdots \\ \langle \phi_N, \psi \rangle \end{pmatrix} \\ &= \sum_i \langle \tau, \phi_i \rangle \langle \phi_i, \psi \rangle \\ &= \langle \tau, \psi \rangle, \end{aligned} \tag{A.6}$$

hence we can write inner products as contractions between bras and kets, $\langle \tau, \psi \rangle = \langle \tau | \psi \rangle =: \langle \tau | \psi \rangle$.

Similarly, linear operators that act on the Hilbert space (endomorphisms on \mathcal{H}) can be represented as matrices,

$$\begin{aligned}
A &= \underbrace{\left(\sum_i |i\rangle\langle i| \right)}_{\mathbb{1}} A \underbrace{\left(\sum_j |j\rangle\langle j| \right)}_{\mathbb{1}} \\
&= \sum_{i,j} \underbrace{\langle i|A|j\rangle}_{\langle \phi_i, A\phi_j \rangle} |i\rangle\langle j| \\
&= \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & & \\ \vdots & & \ddots & \\ a_{N1} & & & a_{NN} \end{pmatrix}, \quad a_{ij} = \langle i|A|j\rangle. \tag{A.7}
\end{aligned}$$

Examples. Let \mathcal{H} be the Hilbert space of a qubit. Here go a few quick examples of kets, bras, and operators, written in the computational basis, $\{|0\rangle, |1\rangle\}$.

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad |1\rangle\langle 1| = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad |0\rangle\langle 1| = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix},$$

$$|-\rangle = \frac{|0\rangle - |1\rangle}{\sqrt{2}} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad |-\rangle\langle -| = \left(\frac{|0\rangle - |1\rangle}{\sqrt{2}} \right) \left(\frac{\langle 0| - \langle 1|}{\sqrt{2}} \right) = \frac{1}{2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix},$$

$$|+i\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix}, \quad \langle +i| = \frac{1}{\sqrt{2}}(1, -i), \quad |+i\rangle\langle +i| = \frac{1}{2} \begin{pmatrix} 1 & -i \\ i & 1 \end{pmatrix},$$

$$\Pr[1]_{|-\rangle} = \text{Tr}[|1\rangle\langle 1| |-\rangle\langle -|] = \text{Tr}\left[\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \frac{1}{2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \right] = \text{Tr}\left[\frac{1}{2} \begin{pmatrix} 0 & 0 \\ -1 & 1 \end{pmatrix} \right] = \frac{1}{2}.$$