Quantum Information Processing

Quantum Information Theory (I)

Richard Cleve

Institute for Quantum Computing & Cheriton School of Computer Science

University of Waterloo

November 29, 2021

Abstract

The goal of these notes is to explain the basics of quantum information processing, with intuition and technical definitions, in a manner that is accessible to anyone with a solid understanding of linear algebra and probability theory.

These are lecture notes for the third part of a course entitled "Quantum Information Processing" (with numberings QIC 710, CS 768, PHYS 767, CO 681, AM 871, PM 871 at the University of Waterloo). The other parts of the course are: a primer for beginners, quantum algorithms, and quantum cryptography. The course web site `http://cleve.iqc.uwaterloo.ca/qic710` contains other course materials, including video lectures.

I welcome feedback about errors or any other comments. This can be sent to `cleve@uwaterloo.ca` (with "Lecture notes" in subject, if at all possible).

Contents

1 Quantum states as density matrices

Let's begin by considering a couple of situations where Alice has an apparatus for creating quantum states for Bob, who has no apparatus. When Bob needs a specific state, he asks Alice to create it and send it to him.

Figure 1: Alice uses her apparatus to prepare states for Bob.

The story of the fake-plus state

Suppose that Bob asks Alice to create a qubit in state $|+\rangle = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle$ and send it to him. But suppose that Alice's state preparation device is broken in that it can *only* prepare qubits in state $|0\rangle$ or $|1\rangle$.

What is Alice to do? She cannot create the state $|+\rangle$ literally. Suppose she tries to fake it by flipping a fair coin and then creating either $|0\rangle$ or $|1\rangle$, depending on the coin's outcome—while keeping the coin's outcome secret from Bob. The state that Alice creates can be described as

$$\begin{cases} |0\rangle & \text{with probability } \frac{1}{2} \\ |1\rangle & \text{with probability } \frac{1}{2}. \end{cases} \tag{1}$$

Let's call this the *fake-plus state*.

How good is this fake-plus state as a substitution for the real plus state $|+\rangle$? Is there any way that Bob can tell the difference? If, for any measurement that Bob can perform, the outcome probabilities are exactly the same then the substitution is a good one. Note that if Bob measures the fake plus state in the computational basis then the outcome probabilities are the same as measuring $|+\rangle$ in the computational basis. So far so good.

But there are other measurements for which the outcome probabilities are different. Can you think of one?

Exercise 1.1. *Give a measurement which has different outcome probabilities for the fake-plus state* (1) *than for the true plus state* $|+\rangle$.

So the fake-plus state is *not* a good substitute for the plus state.

The story of the fake-fake-plus state

Now, let's consider a different scenario. Suppose that Bob doesn't want a $|+\rangle$ state; instead, he wants Alice to prepare for him a fake-plus state, the state in Eq. (1). But this time, let's suppose that Alice's apparatus is broken in a different way: it can *only* prepare $|+\rangle$ and $|-\rangle$ states.

What can Alice do in this case? It won't do to send Bob a $|+\rangle$ state, because we already know that the plus state and the fake-plus state do not behave the same for all measurements. What if Alice tries to fake it this time by flipping a fair coin and then sending $|+\rangle$ or $|-\rangle$, depending on the outcome (again keeping the coin outcome secret). Such a state can be described as

$$\begin{cases} |+\rangle & \text{with probability } \frac{1}{2} \\ |-\rangle & \text{with probability } \frac{1}{2}. \end{cases} \tag{2}$$

Let's call this the *fake-fake-plus state*. Is the fake-fake-plus state (2) a good substitute for the fake-plus state (1)?

The two states certainly don't look the same. So your first guess might be that there's a measurement for which the outcome probabilities are different. But it turns out that, for *every* measurement that Bob can make, the outcome probabilities for state (1) are exactly the same as they are for state (2). Even though the two states don't look the same, the fake-fake-plus state is a good substitute for the fake-plus state.

1.1 Probabilistic mixtures of states

We are now in the realm of *probabilistic mixtures* of states. These are states where a random process is used to decide which pure state to prepare. Let $(p_1, p_2, ..., p_m)$ be a probability vector and let $|\psi_1\rangle, |\psi_1\rangle, \ldots, |\psi_m\rangle$ be d-dimensional quantum states (they need not be orthogonal). Imagine that $k \in \{1, 2, \ldots, m\}$ is sampled according the the probability distribution $(p_1, p_2, ..., p_m)$, and then state $|\psi_k\rangle$ is produced (but k is not revealed). Such a state can be described as

$$\begin{cases} |\psi_1\rangle & \text{with probability } p_1 \\ |\psi_2\rangle & \text{with probability } p_2 \\ \vdots & \vdots \qquad\qquad \vdots \\ |\psi_m\rangle & \text{with probability } p_m. \end{cases} \tag{3}$$

These states are called *mixed states*. The "ordinary" states, describable by a single normalized vector $|\psi\rangle$, are called *pure states*. In Eq. (3), if one of the probabilities is 1 and the others are 0 then the state is a pure state.

Let's look at some examples of probabilistic mixtures of states. We have already seen these two mixed states:

$$\begin{cases} |0\rangle & \text{with probability } \frac{1}{2} \\ |1\rangle & \text{with probability } \frac{1}{2} \end{cases} \quad \text{and} \quad \begin{cases} |+\rangle & \text{with probability } \frac{1}{2} \\ |-\rangle & \text{with probability } \frac{1}{2}, \end{cases} \tag{4}$$

and I claimed that they are indistinguishable—but I did not explain why. How do these two states compare with

$$\begin{cases} |0\rangle & \text{with probability } \frac{1}{2} \\ |-\rangle & \text{with probability } \frac{1}{2}? \end{cases} \tag{5}$$

Are they also indistinguishable from this state?

Mixed states for qubits can be probability distributions on any number of vectors, for example

$$\begin{cases} |0\rangle & \text{with prob. } \frac{1}{4} \\ |1\rangle & \text{with prob. } \frac{1}{4} \\ |+\rangle & \text{with prob. } \frac{1}{4} \\ |-\rangle & \text{with prob. } \frac{1}{4} \end{cases} \quad \text{and} \quad \begin{cases} |0\rangle & \text{with prob. } \frac{1}{3} \\ -\frac{1}{2}|0\rangle + \frac{\sqrt{3}}{2}|1\rangle & \text{with prob. } \frac{1}{3} \\ -\frac{1}{2}|0\rangle - \frac{\sqrt{3}}{2}|1\rangle & \text{with prob. } \frac{1}{3}. \end{cases} \tag{6}$$

And the probability distribution need not be uniform, as shown in this example

$$\begin{cases} \cos(\frac{\pi}{8})|0\rangle - \sin(\frac{\pi}{8})|1\rangle & \text{with prob. } \cos^2(\frac{\pi}{8}) \\ \sin(\frac{\pi}{8})|0\rangle + \cos(\frac{\pi}{8})|1\rangle & \text{with prob. } \sin^2(\frac{\pi}{8}). \end{cases} \tag{7}$$

Definition 1.1 (indistinguishable states)**.** *Two probabilistic mixtures of states are* indistinguishable *if, for all possible measurements, the outcome probabilities are the same for the two states.*

Now, consider the six mixed states appearing in in (4)(5)(6)(7) above. Which pairs are indistinguishable? To address such questions, we need to understand these kinds of states better. A very useful approach is to express these states in terms of their *density matrices*.

1.2 Density matrices

Given a probability distribution on a set of state vectors, one might be tempted to consider the *weighted average* of the state vectors $p_1 |\psi_1\rangle + p_2 |\psi_2\rangle + \cdots + p_m |\psi_m\rangle$ as a useful object. However, this kind of average turns out to be of little use. One indicator that it's not worth much is that the weighted average can change dramatically depending on the global phases associated with the vectors—which shouldn't matter. Also, notice that, for the second mixed state in Eq. (6), the weighted average is the zero vector.

Mixed states can be nicely characterized by a different kind of averaging, which occurs in the definition of the density matrix.

Definition 1.2 (density matrix)**.** *For a mixed state of the form of Eq.* (3)*, where* $|\psi_1\rangle, |\psi_1\rangle, \ldots, |\psi_m\rangle$ *are d-dimensional, its* density matrix *is the* $d \times d$ *matrix*

$$\rho = p_1 |\psi_1\rangle \langle\psi_1| + p_2 |\psi_2\rangle \langle\psi_2| + \cdots + p_m |\psi_m\rangle \langle\psi_m| . \tag{8}$$

Note that the density matrix of any pure state $|\psi\rangle$ is $|\psi\rangle \langle\psi|$. For example, the state $|\psi\rangle = \alpha_0 |0\rangle + \alpha_1 |1\rangle$ has density matrix

$$\rho = \big(\alpha_0 |0\rangle + \alpha_1 |1\rangle\big)\big(\bar{\alpha}_0 \langle 0| + \bar{\alpha}_1 \langle 1|\big) \tag{9}$$

$$= \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} \begin{bmatrix} \bar{\alpha}_0 & \bar{\alpha}_1 \end{bmatrix} \tag{10}$$

$$= \begin{bmatrix} |\alpha_0|^2 & \alpha_0 \bar{\alpha}_1 \\ \alpha_1 \bar{\alpha}_0 & |\alpha_1|^2 \end{bmatrix} . \tag{11}$$

The entries along the diagonal are the absolute values squared of the amplitudes and the off-diagonal entries are cross-terms involving the amplitudes.

Also note from Definition 1.2 that the density matrix of a probabilistic mixture of pure states is the weighted average of the density matrices of the pure states. For example, the density matrices of $|0\rangle$ and $|1\rangle$ are

$$|0\rangle \langle 0| = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \qquad \text{and} \qquad |1\rangle \langle 1| = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \tag{12}$$

and the density matrix of the first mixed state in Eq. (4) is

$$\tfrac{1}{2} |0\rangle\langle 0| + \tfrac{1}{2} |1\rangle\langle 1| = \tfrac{1}{2} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \tfrac{1}{2} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} . \tag{13}$$

Regarding the second mixed state in Eq. (4), the density matrices of $|+\rangle$ and $|-\rangle$ are

$$|+\rangle\langle+| = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \qquad \text{and} \qquad |-\rangle\langle-| = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix} \tag{14}$$

and the density matrix of their mixture is

$$\tfrac{1}{2}|+\rangle\langle+| \;+\; \tfrac{1}{2}|-\rangle\langle-| = \tfrac{1}{2}\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} + \tfrac{1}{2}\begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}. \tag{15}$$

Notice that the density matrices for the two mixed stated in Eq. (4) are the same. This is related to the fact that the states are indistinguishable—which will be explained shortly.

Exercise 1.2 (a straightforward calculation)**.** *Work out the density matrices of the mixed states appearing in* (5)(6)(7)*.*

Let me make a comment about global phases in vector states. When we represent (pure) states as vectors, there's this issue that if multiply the vector by a unit complex number (of the form $e^{i\theta}$, for $\theta \in \mathbb{R}$) then it's essentially the same state; it's indistinguishable from the original state. The density matrix of $e^{i\theta}|\psi\rangle$ is

$$e^{i\theta}|\psi\rangle\langle\psi|e^{-i\theta} = |\psi\rangle\langle\psi|. \tag{16}$$

So global phases don't even show up in density matrices, which is nice. It means that, using density matrices, we don't need to define an equivalence relation to account for global phases.

Now, let's look at some examples of density matrices for higher dimensional systems than qubits. The density matrix of $|00\rangle$ is

$$|00\rangle\langle 00| = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \tag{17}$$

and the density matrices of the other computational basis states are also diagonal matrices with a 1 in one position.

The density matrix of the Bell state $\frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|11\rangle$ is

$$\begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & 0 & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{bmatrix} \tag{18}$$

and the density matrix of the state

$$\begin{cases} |00\rangle & \text{with probability } \frac{1}{2} \\ |11\rangle & \text{with probability } \frac{1}{2} \end{cases} \tag{19}$$

is

$$\tfrac{1}{2}\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} + \tfrac{1}{2}\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} \end{bmatrix}. \tag{20}$$

If we were adopt the terminology used for state (1), we might call state (19) a *fake-Bell state*. Notice that the difference between the density matrices of the Bell state $\frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|11\rangle$ and the fake-Bell state (19) is in the two off-diagonal entries of their density matrices.

1.2.1 Effect of unitaries on mixed states

Every probabilistic mixture of states has a density matrix associated with it; however, different probabilistic mixtures can result in the same density matrix (for example the two states in Eq. (4)). Suppose that we apply a unitary operation U on a probabilistic mixture of states. How does this affect the density matrix?

If we begin with a mixed state of the form

$$\begin{cases} |\psi_1\rangle & \text{with prob. } p_1 \\ |\psi_2\rangle & \text{with prob. } p_2 \\ \vdots & \vdots & \vdots \\ |\psi_m\rangle & \text{with prob. } p_m \end{cases} \tag{21}$$

then applying U changes the state to

$$\begin{cases} U|\psi_1\rangle & \text{with prob. } p_1 \\ U|\psi_2\rangle & \text{with prob. } p_2 \\ \vdots & \vdots & \vdots \\ U|\psi_m\rangle & \text{with prob. } p_m. \end{cases} \tag{22}$$

This is because, whatever $|\psi_k\rangle$ is randomly selected, it gets converted to $U|\psi_k\rangle$.

Let ρ denote the density matrix of the original state. That is,

$$\rho = \sum_{k=1}^{m} p_k \left|\psi_k\right\rangle \left\langle\psi_k\right| . \tag{23}$$

Then the density matrix after U is applied is

$$\sum_{k=1}^{m} p_k \big(U \left|\psi_k\right\rangle\big) \big(U \left|\psi_k\right\rangle\big)^* = \sum_{k=1}^{m} p_k U \left|\psi_k\right\rangle \left\langle\psi_k\right| U^* \tag{24}$$

$$= U \left(\sum_{k=1}^{m} p_k \left|\psi_k\right\rangle \left\langle\psi_k\right| \right) U^* \tag{25}$$

$$= U \rho U^* . \tag{26}$$

What is remarkable is that the density matrix of the modified state depends *only* on the density matrix of the original state. It does not depend on what specific probabilistic mixture is used to create the original state.

1.2.2 Effect of measurement on mixed states

Now, let's consider the effect of a measurement of a d-dimensional mixed state. As usual, the computational basis is denoted as $\left|0\right\rangle, \left|1\right\rangle, \ldots, \left|d-1\right\rangle$. Let the mixture be

$$\begin{cases} \left|\psi_1\right\rangle & \text{with prob. } p_1 \\ \left|\psi_2\right\rangle & \text{with prob. } p_2 \\ \quad\vdots & \quad\vdots \qquad\quad \vdots \\ \left|\psi_m\right\rangle & \text{with prob. } p_m . \end{cases} \tag{27}$$

There are two different ways that randomness arises in such a measurement: the randomness that was used to select one of the pure states; and, the randomness that arises in the measurement process for the selected state.

If the selected state is $\left|\psi_j\right\rangle$ then the probability of measurement outcome k is

$$\left| \left\langle k|\psi_j\right\rangle \right|^2 = \left\langle k|\psi_j\right\rangle \left\langle\psi_j|k\right\rangle = \left\langle k\right| \Big(\left|\psi_j\right\rangle \left\langle\psi_j\right| \Big) \left|k\right\rangle \tag{28}$$

(where we are using the fact that the expressions are all products of row matrices and column matrices and that matrix multiplication is associative).

If we average this over all possibilities of $|\psi_j\rangle$ then the probability of outcome k is

$$\sum_{j=1}^{m} p_j \langle k| \left(|\psi_j\rangle \langle\psi_j| \right) |k\rangle = \langle k| \left(\sum_{j=1}^{m} p_j |\psi_j\rangle \langle\psi_j| \right) |k\rangle \tag{29}$$

$$= \langle k| \rho |k\rangle , \tag{30}$$

where ρ is the density matrix of the original state. Also, when the measurement outcome is k, the residual state is $|k\rangle$.

Once again, the result of the operation depends *only* on the density matrix of the original state. It does not depend on what specific probabilistic mixture is used to create the original state.

1.2.3 Information processing solely in terms of density matrices

In sections 1.2.1 and 1.2.2, we saw that, for a mixed state with density matrix ρ:

- Applying a unitary operation U to the state changes it to one with density operator $U\rho U^*$.
- Applying a measurement in the computational basis to the state produces classical and quantum outcomes

$$\begin{cases} \big(0,\, |0\rangle\big) & \text{with prob. } \langle 0| \rho |0\rangle \\ \big(1,\, |1\rangle\big) & \text{with prob. } \langle 1| \rho |1\rangle \\ \quad\vdots & \quad\vdots \qquad\qquad \vdots \\ \big(d{-}1,\, |d{-}1\rangle\big) & \text{with prob. } \langle d{-}1| \rho |d{-}1\rangle. \end{cases} \tag{31}$$

In both cases, the result of the operation depends only on the density matrix of the state (not on the specific probabilistic mixture that is used to generate the state).

From this, we can deduce[1] the following theorem.

Theorem 1.1. *Whenever two mixed states have the same density matrix, the states are equivalent.*

Theorem 1.1 implies that the fake-plus state (1) and fake-fake-plus state (2) are indistinguishable.

[1] Actually, we have not yet shown that the outcomes of exotic measurements depend only on the density matrix. Although this is indeed true, it is more convenient to address this later. (Exotic measurements are those where the state being measured is isometrically embedded into a larger space, followed by a unitary and measurement in the larger space.)

1.3 Some properties of matrices

Prior to our use of the density matrix framework, our matrices have mostly been unitary, representing unitary operations on quantum states. Now, we also have density matrices that describe states (and which are not unitary). Henceforth, more types of matrices will arise as we develop our models of quantum information theory. In this section, we review some useful definitions and properties of matrices that will be used.

Definition 1.3 (normal matrix)**.** *A matrix $M \in \mathbb{C}^{d\times d}$ is* normal *if $M^*M = MM^*$.*

An important property of normal matrices is that they are diagonalizable in some orthonormal basis.

Theorem 1.2 (spectral theorem)**.** *A matrix $M \in \mathbb{C}^{d\times d}$ is* normal *if and only if there exists a unitary $U \in \mathbb{C}^{d\times d}$ such that*

$$M = U^* \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_d \end{bmatrix} U. \tag{32}$$

Although Definition 1.3 is the common textbook definition of *normal*, the statement of Theorem 1.2 can be taken as an alternative definition.

To help understand normal matrices, it's useful to see examples of *ab*normal matrices. Consider these two:

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \qquad\qquad \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}. \tag{33}$$

The first matrix is not normal because it is not even diagonalizable. The second matrix is diagonalizable but not unitarily (it has eigenvectors $\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$ and $\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]$, which are not orthogonal).

For any normal matrix, by Theorem 1.2, we can imagine it to be diagonal in the coordinate system of some orthonormal basis. Note that a square matrix M is unitary (defined as $M^*M = I$) if and only if all its eigenvalues have absolute value 1 (i.e., they are points on the unit circle in $\mathbb{C}$. And a matrix M is Hermitian (defined as $M = M^*$) if and only if all its eigenvalues are in $\mathbb{R}$.

Definition 1.4 (positive). *A matrix $M \in \mathbb{C}^{d\times d}$ is* positive[2] *if and only if, M is normal and, for all states $|\psi\rangle \in \mathbb{C}^d$, it holds that $\langle\psi| M |\psi\rangle \geq 0$.*

A normal matrix is positive if and only if all of its eigenvalues are in $\mathbb{R}$ and greater than or equal to 0.

Definition 1.5 (trace). *The* trace *of a matrix $M \in \mathbb{C}^{d\times d}$ (denoted as $\mathrm{Tr}(M)$) is defined as the sum of its diagonal entries*

$$\mathrm{Tr}(M) = \sum_{k=1}^{d} M_{k,k}. \tag{34}$$

As simple as the definition of the trace is, it has some interesting properties. An obvious property is that it is linear. That is, for all $A, B \in \mathbb{C}^{d\times d}$ and all $\alpha, \beta \in \mathbb{C}$,

$$\mathrm{Tr}(\alpha A + \beta B) = \alpha \,\mathrm{Tr}(A) + \beta \,\mathrm{Tr}(B). \tag{35}$$

Also, for all $A \in \mathbb{C}^{d_1\times d_2}$ and $B \in \mathbb{C}^{d_2\times d_1}$,

$$\mathrm{Tr}(AB) = \mathrm{Tr}(BA). \tag{36}$$

Equation (36) implies that the trace is coordinate system independent, in the sense that, for all $S, A \in \mathbb{C}^{d\times d}$ where S is invertible,

$$\mathrm{Tr}(S^{-1}AS) = \mathrm{Tr}(A). \tag{37}$$

Also, notice that, in Eq. (36), A and B need not be square matrices. For example,

$$\mathrm{Tr}(|\psi\rangle\langle\phi|) = \mathrm{Tr}(\langle\phi|\psi\rangle) = \langle\phi|\psi\rangle\,. \tag{38}$$

A word of caution: here are some properties that, in general, the trace does *not* have:

⚠ In general, the trace is not multiplicative (in the sense that the determinant is). In general, $\mathrm{Tr}(AB) = \mathrm{Tr}(A)\,\mathrm{Tr}(B)$ does *not* hold.

⚠ Moreover, Eq. (36) does not mean that you can arbitrarily reorder any product in the argument of the trace. For example, $\mathrm{Tr}(ABC) = \mathrm{Tr}(BAC)$ does *not* hold in general. But, for the trace of a product, the product can always be *cyclically* permuted as $\mathrm{Tr}(A_1A_2\ldots A_{m-1}A_m) = \mathrm{Tr}(A_mA_1A_2\ldots A_{m-1})$.

[2] In some communities, the terminology *positive semidefinite* is used instead of *positive*.

1.3.1 Characterizing density matrices

Not all matrices arise as the density matrix of some probabilistic mixture of states. The following theorem precisely characterizes which matrices are density matrices.

Theorem 1.3 (characterization of valid density matrix)**.** *A matrix $\rho \in \mathbb{C}^{d\times d}$ is the density matrix of some probabilistic mixture of pure states if and only if ρ is positive and $\mathrm{Tr}(\rho) = 1$.*

Exercise 1.3. *Prove Theorem 1.3.*

Theorem 1.4 (characterization of pure states)**.** *If $\rho \in \mathbb{C}^{d\times d}$ is a density matrix then ρ is a pure state if and only if $\mathrm{Tr}(\rho^2) = 1$.*

Exercise 1.4. *Prove Theorem 1.4.*

1.4 Bloch sphere for qubits

The set of density matrices for qubits has a nice representation as points in the *Bloch sphere*. In this section, I explain this correspondence. Consider the Pauli matrices

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \qquad Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \tag{39}$$

(where in this context I is an honorary Pauli matrix). Every 2×2 matrix can be expressed as a linear combination of I, X, Y, Z. Since $\mathrm{Tr}(I) = 2$ and $\mathrm{Tr}(X) = \mathrm{Tr}(Y) = \mathrm{Tr}(Z) = 0$, we can express any density matrix ρ as

$$\rho = \frac{I + c_x X + c_y Y + c_z Z}{2}. \tag{40}$$

Let's develop a geometric picture for the set of all possible triples (c_x, c_y, c_z) that correspond to valid 2×2 density matrices. Let's start with pure states. Any pure state of a qubit can be written as

$$|\psi\rangle = \cos(\tfrac{\theta}{2})\,|0\rangle + e^{i\phi}\sin(\tfrac{\theta}{2})\,|1\rangle\,, \tag{41}$$

for some $\theta, \phi \in [0, 2\pi]$.

The density matrix of this state is

$$|\psi\rangle\langle\psi| = \begin{bmatrix} \cos^2(\frac{\theta}{2}) & e^{-i\phi}\cos(\frac{\theta}{2})\sin(\frac{\theta}{2}) \\ e^{i\phi}\cos(\frac{\theta}{2})\sin(\frac{\theta}{2}) & \sin^2(\frac{\theta}{2}) \end{bmatrix} \tag{42}$$

$$= \frac{1}{2}\begin{bmatrix} 1+\cos(\theta) & e^{-i\phi}\sin(\theta) \\ e^{i\phi}\sin(\theta) & 1-\cos(\theta) \end{bmatrix} \tag{43}$$

$$= \frac{1}{2}\begin{bmatrix} 1+\cos(\theta) & \big(\cos(\phi)-i\sin(\phi)\big)\sin(\theta) \\ \big(\cos(\phi)+i\sin(\phi)\big)\sin(\theta) & 1-\cos(\theta) \end{bmatrix} \tag{44}$$

$$= \frac{I+\cos(\phi)\sin(\theta)X+\sin(\phi)\sin(\theta)Y+\cos(\theta)Z}{2}. \tag{45}$$

Therefore, the coefficients in Eq. (40) are

$$(c_x, c_y, c_z) = (\cos(\phi)\sin(\theta),\ \sin(\phi)\sin(\theta),\ \cos(\theta)). \tag{46}$$

These triples are the *polar coordinates* of points on the surface of a sphere, as shown in figure 2.

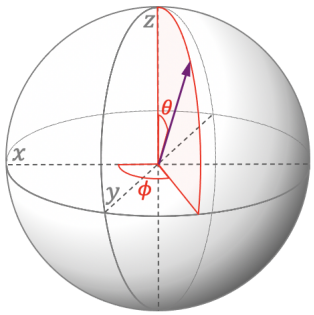

Figure 2: The coordinates (c_x, c_y, c_z) of a pure state are a point on the surface of a sphere.

Think of this sphere as the Earth with the North Pole at the top. Points on the surface can be expressed in terms of their latitude and longitude. The *latitude* θ is the angular distance away from the North Pole. The *longitude* ϕ is an angle representing the East-West distance from some arbitrary[3] starting point. So all the pure states correspond to points[4] on the surface of this sphere, called the *Bloch sphere*.

[3]In geography, the convention is to set 0° at the *Prime Meridian* in Greenwich, UK.

[4]To avoid redundancy, it is natural to restrict the range of θ to $[0, \pi]$.

Where are states $|0\rangle$ and $|1\rangle$ situated on the Bloch sphere? State $|0\rangle$ lies at the North Pole and $|1\rangle$ is at the South Pole, as illustrated in figure 3.

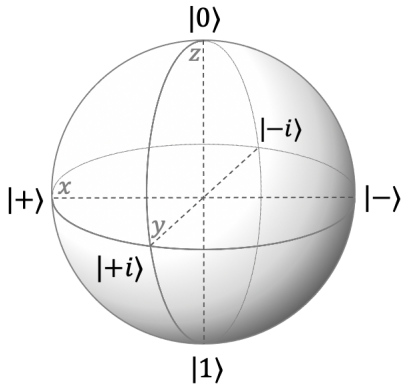

Figure 3: Position of states $|0\rangle$, $|1\rangle$, $|+\rangle$, $|-\rangle$, $|+i\rangle$, and $|-i\rangle$ on the Bloch sphere.

Notice that $|0\rangle$ and $|1\rangle$ are orthogonal as vectors; however, their positions on the Bloch sphere are 180° apart. Any two orthogonal vectors map to antipodal points on the sphere (180° apart).

Where are $|+\rangle$ and $|-\rangle$? They lie on the equator. State $|+\rangle$ has longitude $\phi = 0$ (it lies at the intersection of the equator and the Prime Meridian) and $|-\rangle$ is at the antipodal point.

Notice the angle-doubling again. The angle between $|0\rangle$ and $|+\rangle$ is 45°, but the angle between their points on the Bloch sphere is 90°. In general, for any two state vectors, if we map them to the sphere, the angular distance between them doubles.

There are two other points on the sphere that are in natural positions relative to the points we have considered so far: those that are 90° from $|0\rangle$, $|1\rangle$, $|+\rangle$, and $|-\rangle$. They correspond to the states

$$|+i\rangle = \tfrac{1}{\sqrt{2}}|0\rangle + \tfrac{i}{\sqrt{2}}|1\rangle \tag{47}$$

$$|-i\rangle = \tfrac{1}{\sqrt{2}}|0\rangle - \tfrac{i}{\sqrt{2}}|1\rangle\,. \tag{48}$$

There is some nice symmetry among the six states $|0\rangle$, $|1\rangle$, $|+\rangle$, $|-\rangle$, $|+i\rangle$, and $|-i\rangle$.

The surface of the Bloch sphere consists of all the pure states, and it turns out the *mixed* states are all the points inside the sphere. For any probabilistic mixture of pure states, its position in the Bloch sphere is the weighted average of the positions of the pure states. For example, an equally weighted mixture of $|0\rangle$ and $|1\rangle$ (the so-called fake-plus state from Eq. (1)) is the point right at the centre of the sphere. And an

equally weighted mixture of $|0\rangle$ and $|+\rangle$ is the midpoint of the line connecting their positions on the sphere.

For qubit systems, it's often very useful to think of states—and the operations acting on them—on the Bloch sphere.

A word of caution:

⚠ Do not conflate state vectors with points on the Bloch sphere!

⚠ The Bloch sphere is for one-qubit states. For higher dimensional systems, there's an analogous geometric shape—but it's not a hypersphere. It's shape is somewhat complicated and it doesn't satisfy all the properties that one might expect to hold based on the case of qubits. For example, not all points on the surface of this shape are pure states (some are mixed states). For qutrits, the shape is 7-dimensional.

2 State transitions in the Kraus form

So far, we have seen various kinds of operations that can be performed on quantum systems. We have seen unitary operations and measurements. There are also operations that are described in words (or annotated quantum circuits), such as "add an ancilla qubit", "measure the second qubit", and "take the first qubit as the output." The Kraus form is a unified framework for describing all of these, as well as some other kinds of quantum operations. We begin with this definition.

Definition 2.1 (Kraus operators)**.** *A sequence of $d_1 \times d_2$ matrices $A_0, A_2, \ldots, A_{m-1}$ is a sequence of* Kraus operators *if*

$$\sum_{k=0}^{m-1} A_k^* A_k = I, \tag{49}$$

where I denotes the $d_2 \times d_2$ identity matrix.

Note that the matrices in the above definition need not be square: if A_k is $d_1 \times d_2$ then $A_k^* A_k$ is $d_2 \times d_2$.

At first glance, Definition 2.1 may look mysterious. In this section, we'll see that several quantum state transformations, including measurements, unitary operations (and other natural transformations) are expressible in terms of Kraus operators.

2.1 Measurements via Kraus operators

For any Kraus operators $A_0, A_2, \ldots, A_{m-1} \in \mathbb{C}^{d_1 \times d_2}$, define the the following measurement operation, whose classical output is $k \in \{0, 1, \ldots, m-1\}$.

Input to the measurement: is a d_2-dimensional quantum system, whose state can be described by a $d_2 \times d_2$ density matrix ρ.

Output of the measurement: can be described as as the probabilistic mixture

$$\begin{cases} \left(\quad 0\,, \quad \dfrac{A_0 \rho A_0^*}{\operatorname{Tr}(A_0 \rho A_0^*)} \quad \right) & \text{with prob. } \operatorname{Tr}(A_0 \rho A_0^*) \\ \left(\quad 1\,, \quad \dfrac{A_1 \rho A_1^*}{\operatorname{Tr}(A_1 \rho A_1^*)} \quad \right) & \text{with prob. } \operatorname{Tr}(A_1 \rho A_1^*) \\ \quad \vdots \qquad \vdots & \qquad \vdots \qquad \vdots \\ \left(m-1\,, \ \dfrac{A_{m-1} \rho A_{m-1}^*}{\operatorname{Tr}(A_{m-1} \rho A_{m-1}^*)} \right) & \text{with prob. } \operatorname{Tr}(A_{m-1} \rho A_{m-1}^*), \end{cases} \tag{50}$$

where the first component is the classical outcome $k \in \{0, 1, \ldots, m-1\}$ and the second component is the residual state, which is a $d_1 \times d_1$ density matrix (if $d_1 \neq d_2$ then the dimensions of the input and output systems are different).

The first question is whether the above makes sense. Are the probabilities non-negative real numbers that sum to 1? Are the residual states valid density matrices? To get an idea why this measurement makes sense, consider the case of pure states. If $\rho = |\psi\rangle\langle\psi|$ then, for all k,

$$\operatorname{Tr}(A_k \rho A_k^*) = \operatorname{Tr}(A_k |\psi\rangle\langle\psi| A_k^*) = \operatorname{Tr}(\langle\psi| A_k^* A_k |\psi\rangle) = \langle\psi| A_k^* A_k |\psi\rangle\,. \tag{51}$$

Clearly, $\langle\psi| A_k^* A_k |\psi\rangle \geq 0$, since this is the inner product of $A_k |\psi\rangle$ with itself. Also,

$$\sum_{k=0}^{m-1} \operatorname{Tr}(A_k \rho A_k^*) = \sum_{k=0}^{m-1} \langle\psi| A_k^* A_k |\psi\rangle = \langle\psi| \left(\sum_{k=0}^{m-1} A_k^* A_k \right) |\psi\rangle = \langle\psi|\psi\rangle = 1. \tag{52}$$

The more general case where ρ is a mixed state can by analyzed by averaging over pure states.

Exercise 2.1 (straightforward)**.** *Show that, for an arbitrary $d_2 \times d_2$ density matrix ρ, it holds that $\operatorname{Tr}(A_k \rho A_k^*) \geq 0$ (for all k) and $\sum_{k=0}^{m-1} \operatorname{Tr}(A_k \rho A_k^*) = 1$. Also show that $(A_k \rho A_k)/\operatorname{Tr}(A_k \rho A_k^*)$ is a valid density matrix (for all k).*

Next, we'll see some measurements expressed in terms of Kraus operators.

2.1.1 Computational basis measurements

Let us begin with the basic measurement with respect to the computational basis. For a d-dimensional system, the computational basis is $|0\rangle, |1\rangle, \ldots, |d-1\rangle$. To express this measurement in the Kraus form, set

$$A_k = |k\rangle\langle k| \tag{53}$$

for each $k \in \{0, 1, \ldots, d-1\}$. It's easy to check that these are valid Kraus operators, in the sense of Definition 2.1.

Exercise 2.2. *Show that $A_0, A_1, \ldots, A_{d-1}$, as defined as in Eq. (53), are valid Kraus operators.*

For A_k, as defined as in Eq. (53), it holds that

$$\mathrm{Tr}(A_k \rho A_k^*) = \mathrm{Tr}(|k\rangle\langle k| \rho |k\rangle\langle k|) = \mathrm{Tr}(\langle k|k\rangle \langle k| \rho |k\rangle) = \langle k| \rho |k\rangle \tag{54}$$

and

$$\frac{A_k \rho A_k^*}{\mathrm{Tr}(A_k \rho A_k^*)} = \frac{|k\rangle \langle k| \rho |k\rangle \langle k|}{\langle k| \rho |k\rangle} = |k\rangle \langle k| \tag{55}$$

(where we have used the fact that $\langle k| \rho |k\rangle$ is a scalar). This is consistent with our definition of the measurement in the computational basis in section 1.2.2.

2.1.2 Projective measurements

A *projective measurement* is a measurement with respect to orthogonal subspaces. In the case of pure states, the effect of such a measurement is for the state to project to one of the subspaces, where the probabilities are the projection lengths squared.

These measurements were discussed in the notes [*Part 1: A Primer for Beginners*, Section 8.1]. What follows is a description of these measurements in the Kraus form using projectors.

Definition 2.2 (projector)**.** *A matrix Π is a* projector *if Π is normal and $\Pi^2 = \Pi$.*

Note that the eigenvalues of a projector are 0 or 1. Geometrically, if a projector is applied to a vector then the result is its component in the 1-eigenspace.

Definition 2.3 (orthogonal and complete projectors)**.** *Let $\Pi_0, \Pi_1, \ldots, \Pi_{m-1} \in \mathbb{C}^d$ be a sequence of projectors. The projectors are* orthogonal *if $\Pi_j \Pi_k = 0$ (the zero matrix) for all $j \neq k$. The projectors are* complete *if $\Pi_0 + \Pi_1 + \cdots + \Pi_{m-1} = I$.*

Here's a simple example of orthogonal and complete projectors in $\mathbb{C}^3$.

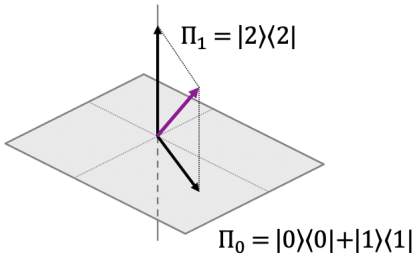

Figure 4: $\Pi_0 = |0\rangle\langle 0| + |1\rangle\langle 1|$ and $\Pi_1 = |2\rangle\langle 2|$ are orthogonal and complete projectors in $\mathbb{C}^3$.

It's easy to see that any sequence of orthogonal and complete projectors are Kraus operators.

Exercise 2.3. *Show that if* $\Pi_0, \Pi_1, \ldots, \Pi_{m-1}$ *are orthogonal and complete projectors then they are Kraus operators, in that they satisfy Eq.* (49).

Therefore, a sequence of orthogonal and complete projectors defines a measurement in the Kraus form. The probability of outcome k is $\mathrm{Tr}(\Pi_k \rho \Pi_k) = \mathrm{Tr}(\rho \Pi_k)$.

Let's look at what these measurements do for pure states. If $\rho = |\psi\rangle\langle\psi|$ then

$$\begin{aligned} \mathrm{Tr}(\rho \Pi_k) &= \mathrm{Tr}\big(|\psi\rangle \langle\psi| \Pi_k\big) && (56)\\ &= \langle\psi| \Pi_k |\psi\rangle && (57)\\ &= \big|\Pi_k |\psi\rangle\big|^2, && (58) \end{aligned}$$

which is the projection length squared of $|\psi\rangle$ to the 1-eigenspace of Π_k. And the corresponding residual state can be shown to be $\Pi_k |\psi\rangle$ normalized.

2.1.3 Measuring the first of two registers

Suppose that we have a system consisting of two registers, with respective dimensions d_1 and d_2.

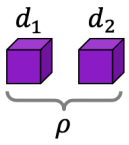

Figure 5: A system consisting of a d_1-dimensional register and a d_2-dimensional register.

All the pure states on this combined system are $d_1 d_2$-dimensional vectors and the density matrices are in $d_1 d_2 \times d_1 d_2$ matrices. A measurement of the *first* register in the computational basis can be defined along the lines of the notes [*Part 1: A Primer for Beginners*, Section 8.2]. In the language of Kraus operators, we can define this measurement as follows. Let $|0\rangle, |1\rangle, \ldots, |d_1 - 1\rangle$ be the computational basis for the first register and set the Kraus operators to be

$$A_k = \big(|k\rangle\langle k|\big) \otimes I, \tag{59}$$

for $k \in \{0, 1, \ldots, d_1 - 1\}$ (where I denotes the $d_2 \times d_2$ identity matrix). Following Eq. (50), for each $k \in \{0, 1, \ldots, d_1\}$, the output can be shown to be[5]

$$\left(k\,, \quad |k\rangle\langle k| \otimes \frac{\big(\langle k| \otimes I\big)\rho\big(|k\rangle \otimes I\big)}{\mathrm{Tr}\Big(\big(\langle k| \otimes I\big)\rho\big(|k\rangle \otimes I\big)\Big)} \right) \tag{60}$$

with probability $\mathrm{Tr}\Big(\big(\langle k| \otimes I\big)\rho\big(|k\rangle \otimes I\big)\Big)$.

2.1.4 Trine state measurement

So far, all the Kraus measurements that we've seen are projective measurements. However, Kraus measurements need not be projective, as the next example shows.

Let's begin by considering the problem of distinguishing between the trine states

$$|\phi_0\rangle = |0\rangle \tag{61}$$

$$|\phi_1\rangle = -\tfrac{1}{2}|0\rangle + \tfrac{\sqrt{3}}{2}|1\rangle \tag{62}$$

$$|\phi_2\rangle = -\tfrac{1}{2}|0\rangle - \tfrac{\sqrt{3}}{2}|1\rangle\,, \tag{63}$$

which are three vectors in $\mathbb{C}^2$, with angle 120° between each pair.

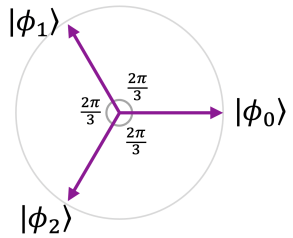

Figure 6: The trine states in $\mathbb{C}^2$.

[5]The key step is that $\big(|\psi\rangle\langle\psi| \otimes I\big)\rho\big(|\psi\rangle\langle\psi| \otimes I\big) = \big(|\psi\rangle \otimes I\big)\Big(\big(\langle\psi| \otimes I\big)\rho\big(|\psi\rangle \otimes I\big)\Big)\big(\langle\psi| \otimes I\big)$
$= |\psi\rangle\langle\psi| \otimes \Big(\big(\langle\psi| \otimes I\big)\rho\big(|\psi\rangle \otimes I\big)\Big)$.

Suppose that we're given one of these states (we're not told which one) and our goal is to perform a measurement that guesses the state correctly with as high a probability as possible. It turns out the optimal performance (for a worst-case input state) cannot attained by any projective measurement in $\mathbb{C}^2$.

Define these three Kraus operators

$$A_0 = \sqrt{\tfrac{2}{3}}\,|\phi_0\rangle\langle\phi_0| \;=\; \begin{bmatrix} \sqrt{2/3} & 0 \\ 0 & 0 \end{bmatrix} \tag{64}$$

$$A_1 = \sqrt{\tfrac{2}{3}}\,|\phi_1\rangle\langle\phi_1| \;=\; \frac{1}{4}\begin{bmatrix} \sqrt{2/3} & -\sqrt{2} \\ -\sqrt{2} & \sqrt{6} \end{bmatrix} \tag{65}$$

$$A_2 = \sqrt{\tfrac{2}{3}}\,|\phi_2\rangle\langle\phi_2| \;=\; \frac{1}{4}\begin{bmatrix} \sqrt{2/3} & \sqrt{2} \\ \sqrt{2} & \sqrt{6} \end{bmatrix}. \tag{66}$$

Notice that these are *not* projectors (because of the factor $\sqrt{2/3}$), and they are not orthogonal. Nevertheless, since $A_0^*A_0 + A_1^*A_1 + A_2^*A_2 = I$, these are valid Kraus operators. Using this measurement for the trine state distinguishing problem results in success probability $\frac{2}{3}$. This is not achievable using projective measurements (however, it is achievable using one of the so-called exotic measurements, where the system is embedded into a larger dimensional space before a projective measurement).

2.2 Quantum channels via Kraus operators

For a sequence of Kraus operators, $A_0, A_2, \ldots, A_{m-1} \in \mathbb{C}^{d_1 \times d_2}$ define the following state transformation, called a *quantum channel*, which maps quantum states to quantum states with no classsical side information.

Input to the channel: is a d_2-dimensional quantum system, whose state can be described by a $d_2 \times d_2$ density matrix ρ.

Output of the channel: is a d_1-dimensional quantum system, whose state is

$$A_0\rho A_0^* + A_1\rho A_1^* + \cdots + A_{m-1}\rho A_{m-1}^*. \tag{67}$$

2.2.1 Unitary operations

Any $d \times d$ unitary operation U corresponds to a quantum channel with one single Kraus operator U. The channel maps each $d \times d$ density matrix ρ maps ρ to $U\rho U^*$. We can think of quantum channels as generalizations of unitary operations.

2.2.2 Decoherence of a qubit

I will first explain what this channel does, and then show you two different ways of "implementing" the channel in terms of Kraus operators. The decoherence channel changes the state of its input qubit from

$$\rho = \begin{bmatrix} \rho_{00} & \rho_{01} \\ \rho_{10} & \rho_{11} \end{bmatrix} \qquad \text{to} \qquad \begin{bmatrix} \rho_{00} & 0 \\ 0 & \rho_{11} \end{bmatrix}. \tag{68}$$

The diagonal density matrix can be viewed as a probabilistic mixture of $|0\rangle\langle 0|$ and $|1\rangle\langle 1|$. On the Bloch sphere, the diagonal density matrices are on the axis connecting $|0\rangle\langle 0|$ and $|1\rangle\langle 1|$. The effect of the channel is to move the state "horizontally" (i.e., parallel to the equatorial plane) to the vertical axis connecting $|0\rangle\langle 0|$ and $|1\rangle\langle 1|$.

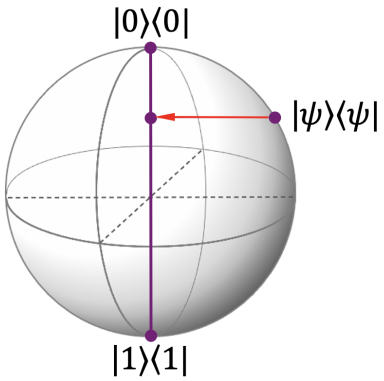

Figure 7: Effect of the decoherence channel on pure state $|\psi\rangle\langle\psi|$.

I will show you two operationally different ways of implementing this channel.

Measuring without looking at the outcome

Our first way of implementing the decoherence channel can be intuitively thought of as measuring the qubit in the computational basis—but without looking at the classical outcome. We might imagine that Bob performs the measurement, but covers his eyes so that he doesn't see the classical outcome. But let's think about it this way: Bob sends the qubit to Alice, who performs the measurement (and sees the outcome) and then Alice sends the qubit back to Bob, but she does not send him the classical output of the measurement.

The quantum part of the outcome of Alice's measurement is

$$\begin{cases} |0\rangle & \text{with prob. } \langle 0|\,\rho\,|0\rangle \\ |1\rangle & \text{with prob. } \langle 1|\,\rho\,|1\rangle. \end{cases} \tag{69}$$

Since Alice obtains the classical outcome, from *her* perspective, the quantum outcome is always either $|0\rangle\langle 0|$ or $|1\rangle\langle 1|$. But Bob does not receive the classical outcome so, from *his* perspective, the quantum outcome is the density matrix

$$\langle 0|\rho|0\rangle\,|1\rangle\langle 1| + \langle 1|\rho|1\rangle\,|1\rangle\langle 1|\,. \tag{70}$$

This can be expressed in the Kraus form by setting the Kraus operators of a quantum channel to $|0\rangle\langle 0|$ and $|1\rangle\langle 1|$. Then a density matrix $\rho = \left[\begin{smallmatrix}\rho_{00} & \rho_{01}\\ \rho_{10} & \rho_{11}\end{smallmatrix}\right]$ maps to

$$\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}\rho_{00} & \rho_{01}\\ \rho_{10} & \rho_{11}\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix} + \begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}\rho_{00} & \rho_{01}\\ \rho_{10} & \rho_{11}\end{bmatrix}\begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix} = \begin{bmatrix}\rho_{00} & 0\\ 0 & \rho_{11}\end{bmatrix}. \tag{71}$$

Probabilistic mixture of I and Z

Another way of implementing the decoherence channel is intuitively based on applying a randomly selected unitary to the state. Bob sends the qubit to Alice, who does the following. She flips a fair coin, and then either applies I or Z to the qubit, depending on the outcome of the coin flip. Then she sends the qubit back to Bob, but she does not reveal the coin flip.

Since Alice knows outcome of the coin flip, from *her* perspective, the state is either ρ or $Z\rho Z$. But Bob does not know the coin flip so, from *his* perspective, the state is

$$\begin{cases} \rho & \text{with prob. } \frac{1}{2} \\ Z\rho Z & \text{with prob. } \frac{1}{2}. \end{cases} \tag{72}$$

and the density matrix of this mixture is

$$\tfrac{1}{2}\rho + \tfrac{1}{2}Z\rho Z. \tag{73}$$

This can be expressed in the Kraus form by setting the Kraus operators of a quantum channel to $\frac{1}{\sqrt{2}}I$ and $\frac{1}{\sqrt{2}}Z$. Then a density matrix $\rho = \left[\begin{smallmatrix}\rho_{00} & \rho_{01}\\ \rho_{10} & \rho_{11}\end{smallmatrix}\right]$ maps to

$$\left(\tfrac{1}{\sqrt{2}}I\right)\rho\left(\tfrac{1}{\sqrt{2}}I\right)^* + \left(\tfrac{1}{\sqrt{2}}Z\right)\rho\left(\tfrac{1}{\sqrt{2}}Z\right)^* = \tfrac{1}{2}\rho + \tfrac{1}{2}Z\rho Z \tag{74}$$

$$= \tfrac{1}{2}\begin{bmatrix}\rho_{00} & \rho_{01}\\ \rho_{10} & \rho_{11}\end{bmatrix} + \tfrac{1}{2}\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}\begin{bmatrix}\rho_{00} & \rho_{01}\\ \rho_{10} & \rho_{11}\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix} \tag{75}$$

$$= \begin{bmatrix}\rho_{00} & 0\\ 0 & \rho_{11}\end{bmatrix}. \tag{76}$$

Comparison of the two implementations of the decoherence channel

From Bob's perspective, who doesn't receive any classical measurement outcomes or coin flip outcomes, the two implementations of the decoherence channel are identical. However, they are not literally the same. In the first implementation, the state is actually measured and it cannot be recovered at a later time, even with the classical information that Alice has. In the second implementation, Alice performs no measurement. If, at some later time, she reveals the coin flip to Bob then he can recover the initial state (by applying either I or Z to the state).

So there's an advantage to the second implementation. But there is also a disadvantage: if Bob asks Alice later on "what was the measurement outcome?", she cannot answer that question. There is no classical bit $b \in \{0, 1\}$ that Alice can produce and send to Bob such that, if Bob then measures his decohered state in the computational basis, the outcome is guaranteed to be b.

Exercise 2.4 (conceptual)**.** *Suppose that Bob believes that he has figured out a new way of measuring a qubit that is reversible. His idea is to first implement the random unitary method to create the decohered state, which can serve as the quantum outcome (remembering what the coin flip is, so that the can undo the unitary later on). Now all that's lacking is that classical outcome. Bob's idea is to measure the decohered qubit in the computational basis to obtain a bit that can serve as the the classical outcome of the measurement. Will doing all this result in a faithful simulation of the measurement operation? And, after all these operations have been performed, is there a way for Bob to recover the original state?*

2.2.3 General measurement without seeing the outcome

For any sequence of Kraus operators $A_0, A_2, \ldots, A_{m-1}$, we have defined an associated measurement in section 2.1 and an associated channel in section 2.2. The associated channel can *always* be interpreted as performing the associated measurement without looking at the classical outcome.

2.2.4 General mixed unitary channels

For any sequence of unitary operations $U_0, U_1, \ldots, U_{m-1}$ with associated probabilities $p_0, p_1, \ldots, p_{m-1}$, consider the operation where $k \in \{0, 1 \ldots, m-1\}$ is randomly chosen according to probabilities $p_0, p_1, \ldots, p_{m-1}$ and then U_k is applied. If the selected k is

not revealed then this procedure maps any input state ρ to the output state

$$p_0\, U_0\, \rho\, U_0^* \;+\; p_1\, U_1\, \rho\, U_1^* \;+\; \cdots \;+\; p_{m-1}\, U_{m-1}\, \rho\, U_{m-1}^*. \tag{77}$$

This is easy to express in the Kraus form, by setting the Kraus operators to

$$A_k = \sqrt{p_k}\, U_k \tag{78}$$

for $k \in \{0, 1, \ldots, m-1\}$.

Exercise 2.5 (may be challenging)**.** *Can* every *quantum channel, as defined in Eq.* (67)*, be expressed as a probability distribution on a set of unitary operations? Either prove this to be the case or give a counterexample.*

2.2.5 Adding an ancilla

A natural quantum operation is to append an ancilla in state $|\psi\rangle$ after a register. Let $|\psi\rangle$ be d_2-dimensional. The input to this operation is a d_1-dimensional system, whose state is described by a $d_1 \times d_1$ density matrix ρ. The output is a $d_1 d_2$-dimensional system, whose state is $\rho \otimes \big(|\psi\rangle\langle\psi|\big)$.

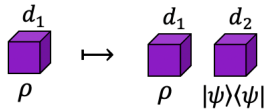

Figure 8: The operation of appending an ancilla in state $|\psi\rangle$ after a register.

This can be expressed as a channel in the Kraus form with one Kraus operator

$$A_0 = I \otimes |\psi\rangle\,. \tag{79}$$

Applying the channel to state $\rho \in \mathbb{C}^{d_1 \times d_1}$ produces the state

$$\begin{aligned}
A_0 \rho A_0^* &= \big(I \otimes |\psi\rangle\big)\rho\big(I \otimes \langle\psi|\big) && (80)\\
&= \big(I \otimes |\psi\rangle\big)\big(\rho \otimes [\,1\,]\big)\big(I \otimes \langle\psi|\big) \qquad \text{where } [\,1\,] \text{ is a } 1\times 1 \text{ matrix} && (81)\\
&= \big(I\rho I\big) \otimes \big(|\psi\rangle\,[\,1\,]\,\langle\psi|\big) && (82)\\
&= \rho \otimes \big(|\psi\rangle\langle\psi|\big). && (83)
\end{aligned}$$

(The insertion of the 1×1 matrix $[\,1\,]$ above is an optional step to make the product easier to parse in a form where the identity $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$ can be applied [*Part 1: A Primer for Beginners*, Section 6.6, Lemma 6.1].)

An explicit example is the addition of an ancilla in state $|0\rangle$ after a qubit. This is accomplished by the Kraus operator

$$A_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}. \tag{84}$$

Note that we have addressed the case of adding an ancilla in a pure state. What if we want to add an ancilla in a mixed state? I leave this as an exercise.

Exercise 2.6. *Suppose that we want to add an ancilla in the mixed state $\sigma \in \mathbb{C}^{d_2 \times d_2}$ to the end of a d_1-dimensional system. Show how to express this as a quantum channel in the Kraus form.*

2.2.6 Partial trace

Suppose that Bob is in possession of a system consisting of two registers. Let his first register be d_1-dimensional and his second register be d_2-dimensional. Suppose that Bob wants to discard his first register. What does this mean? Intuitively, we can imagine that Bob sends his first register to a faraway place where he will never access it again. Another way of thinking about this is that the first register doesn't move, but Bob decides to henceforth completely ignore it. He ghosts his first register. What's the state of Bob's remaining register?

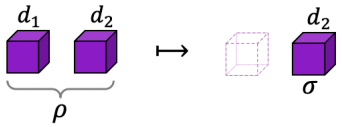

Figure 9: Tracing out the first of two registers.

This question arose the context of pure states in the notes [*Part 1: A Primer for Beginners*, Section 6.3]. If one restricts to pure states (representable as unit vectors) then subsystems might not have states of their own. For example, there is no pure state that captures the state of the second qubit of the Bell state $\frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|11\rangle$.

However, we are now working in a broader context that includes mixed states (representable as density matrices). In this broader context, subsystems always have well-defined states of their own. The states of subsystems are captured by a quantum channel called the *partial trace*.

One way of deriving the definition of the partial trace is to use the fact that what happens to the discarded system is inconsequential to the remaining system. In particular, there is no harm in measuring the discarded system in some orthonormal basis, say $|0\rangle, |1\rangle, \ldots, |d_1 - 1\rangle$ (and not looking at the classical or quantum outcomes of the measurement). Following Eq. (59), the quantum channel corresponding to this measurement (without looking at the classical outcome) has Kraus operators

$$|k\rangle\langle k| \otimes I, \tag{85}$$

for $k \in \{0, 1, \ldots, d_1 - 1\}$. But the output of this channel includes the residual quantum state of the first register (see Eq. (60)). To eradicate this residual state, we modify the Kraus operators to

$$\langle k| \otimes I. \tag{86}$$

It's easy to check that $\langle 0| \otimes I$, $\langle 1| \otimes I$, $\ldots$, $\langle d_1 - 1| \otimes I$ are valid Kraus operators and the quantum channel that they define is the partial trace.

Definition 2.4 (partial trace)**.** *This definition is in the context of a system with a d_1-dimensional register and a d_2-dimensional register. There are two partial traces. The* partial trace $\mathrm{Tr}_1 : \mathbb{C}^{d_1 d_2 \times d_1 d_2} \to \mathbb{C}^{d_2 \times d_2}$ *is defined as, for all $\rho \in \mathbb{C}^{d_1 d_2 \times d_1 d_2}$,*

$$\mathrm{Tr}_1(\rho) = \sum_{k=0}^{d_1 - 1} \big(\langle k| \otimes I\big)\rho\big(|k\rangle \otimes I\big). \tag{87}$$

And the partial trace $\mathrm{Tr}_2 : \mathbb{C}^{d_1 d_2 \times d_1 d_2} \to \mathbb{C}^{d_1 \times d_1}$ *is defined as, for all $\rho \in \mathbb{C}^{d_1 d_2 \times d_1 d_2}$,*

$$\mathrm{Tr}_2(\rho) = \sum_{k=0}^{d_2 - 1} \big(I \otimes \langle k|\big)\rho\big(I \otimes |k\rangle\big). \tag{88}$$

The subscript of Tr denotes which system is being traced out. In the above definition, the measurement is with respect to the computational basis, but the channel is the same if a different orthonormal basis is used.

Recall that the *trace* of a square matrix is the sum of its diagonal entries. We can also call this the *full trace* and its definition can be written as $\mathrm{Tr}(\rho) = \sum_{k=0}^{d-1} \langle k| \rho |k\rangle$. And the entries of the partial trace of ρ are sums of the matrix entries of ρ.

For the case of two 1-qubit registers,

$$\mathrm{Tr}_1 \begin{bmatrix} \rho_{00,00} & \rho_{00,01} & \rho_{00,10} & \rho_{00,11} \\ \rho_{01,00} & \rho_{01,01} & \rho_{01,10} & \rho_{01,11} \\ \rho_{10,00} & \rho_{10,01} & \rho_{10,10} & \rho_{10,11} \\ \rho_{11,00} & \rho_{11,01} & \rho_{11,10} & \rho_{11,11} \end{bmatrix} = \begin{bmatrix} \rho_{00,00} + \rho_{10,10} & \rho_{00,01} + \rho_{10,11} \\ \rho_{01,00} + \rho_{11,10} & \rho_{01,01} + \rho_{11,11} \end{bmatrix} \tag{89}$$

$$\mathrm{Tr}_2 \begin{bmatrix} \rho_{00,00} & \rho_{00,01} & \rho_{00,10} & \rho_{00,11} \\ \rho_{01,00} & \rho_{01,01} & \rho_{01,10} & \rho_{01,11} \\ \rho_{10,00} & \rho_{10,01} & \rho_{10,10} & \rho_{10,11} \\ \rho_{11,00} & \rho_{11,01} & \rho_{11,10} & \rho_{11,11} \end{bmatrix} = \begin{bmatrix} \rho_{00,00} + \rho_{01,01} & \rho_{00,10} + \rho_{01,11} \\ \rho_{10,00} + \rho_{11,01} & \rho_{10,10} + \rho_{11,11} \end{bmatrix}. \tag{90}$$

It should be noted that, although a measurement was introduced to derive[6] the formulas for the partial trace, the measurement does not have to occur. If one register is discarded then the state of the other register is given by the formula for the partial trace whether or not the discarded register is measured.

Now, let's calculate the state of the second qubit of the Bell state $\frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|11\rangle$. Applying the formula in Eq. (89) to the density matrix of the state, we obtain

$$\mathrm{Tr}_1\Big(\big(\tfrac{1}{\sqrt{2}}|00\rangle + \tfrac{1}{\sqrt{2}}|11\rangle\big)\big(\tfrac{1}{\sqrt{2}}\langle 00| + \tfrac{1}{\sqrt{2}}\langle 11|\big)\Big) = \mathrm{Tr}_1 \begin{bmatrix} \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{bmatrix} \tag{91}$$

$$= \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}. \tag{92}$$

There's something remarkable about this. Until now, all our mixed states have been expressed as probabilistic mixtures of pure states. However, a mixed state can arise from a process without any explicit occurrence of randomness or measurement. For the pure state $\frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|11\rangle$, the state of each of its individual qubits is $\begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$.

[6] An alternative way of deriving the formula for $\mathrm{Tr}_1 : \mathbb{C}^{d_1 d_2 \times d_1 d_2} \to \mathbb{C}^{d_2 \times d_2}$ is to define Tr_1 as the unique linear operator with the property that, for all $\rho \in \mathbb{C}^{d_1 \times d_1}$ and $\sigma \in \mathbb{C}^{d_2 \times d_2}$, $\mathrm{Tr}_1(\rho \otimes \sigma) = \mathrm{Tr}(\rho)\sigma$.

3 State transitions in the Stinespring form

In the previous section, I showed you how to express quantum measurements and quantum channels in terms of Kraus operators. In this section, I'm going to show you another form for expressing state transitions, called the *Stinespring form*.

3.1 Measurements in the Stinespring form

Imagine that the input state is a d-dimensional register. First, we append an m-dimensional ancilla register in some computational basis state, say $|0\rangle$. The combined system is md-dimensional. Next, we apply some $md \times md$ unitary operation U to the combined system. Finally, we measure one register in the computational basis, yielding a classical outcome k and a residual quantum state in the other register.

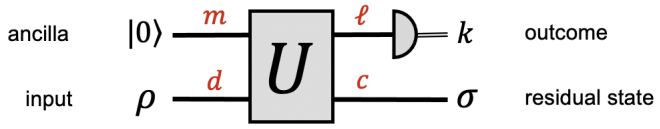

Figure 10: Quantum circuit for a measurement in the Stinespring form.

It's natural for the dimensions of the registers coming out of U to be the same as those of the registers going in ($\ell = m$ and $c = d$). But we allow for the dimensions of the outgoing registers to be different, as long as the total dimension is the same ($md = \ell c$). To get a feeling for this, consider the case where all the dimensions are powers of 2. In that case, we can assume that each register is a bunch of qubits.

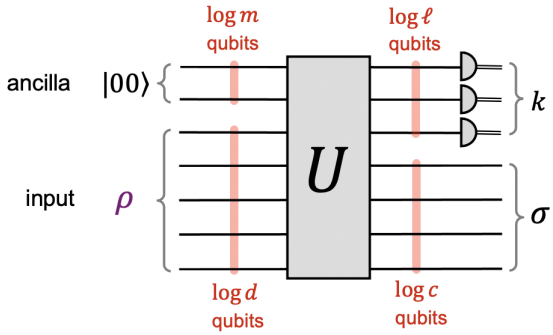

Figure 11: Quantum circuit on qubits for a measurement in the Stinespring form.

In this example, the input state is 5 qubits, whereas the residual outgoing state is 4 qubits. Also, the ancilla is 2 qubits, whereas the number of qubits that are measured

is 3. As long as the total number of qubits going into U and coming out of U is preserved, this makes perfect sense.

Also, note that some of the dimensions can be 1. A 1-dimensional register is essentially the same as no register. A 1×1 density matrix is $\begin{bmatrix} 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \end{bmatrix} \otimes \rho = \rho$. For example, for a measurement in an orthonormal basis specified by U, a very strict translation into the form of figure 11 is obtained by setting $m = c = 1$ and $\ell = d$ (where $U = I$ in the case of the computational basis).

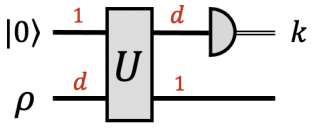

Figure 12: Measurement with respect to an orthonormal basis specified by U.

But figure 12 is pedantic, and we can freely omit the wires of dimension 1 (and omit any I gates). With this relaxation, we can denote a measurement with respect to an orthonormal basis in the Stinespring form as follows.

Figure 13: Measurement with respect to the computational basis and a basis specified by U.

Remember the "exotic measurements" in the notes [*Part 1: A Primer for Beginners*, Section 9]? It should be clear that those measurements are subsumed by these Stinespring measurements.

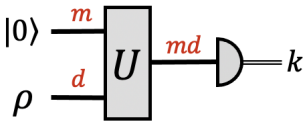

Figure 14: Exotic measurement in the Stinespring form.

3.2 Channels in the Stinespring form

As we noted earlier, one way of thinking about a channel is as a measurement where we don't look at the classical part of the outcome. So we could define Stinespring channels that way. We'll do that, but we'll simplify things by noting that, if we're

not going to see the classical outcome then, instead of performing the measurement, we can trace out that register, like this.

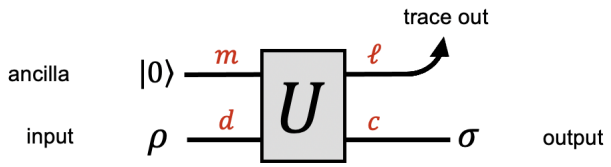

Figure 15: Quantum circuit for a channel in the Stinespring form.

Notice the circuit notation that I'm using here for tracing out a register: the imagery is supposed evoke that the register is tossed away.

Here are some of our most basic channels in the Stinespring form (loosely in the form of figure 15).

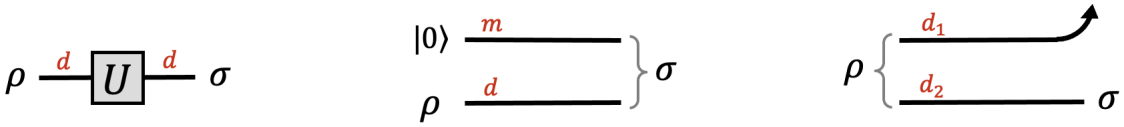

Figure 16: Unitary channel, add ancilla $|0\rangle\langle 0|$ channel, and partial trace Tr_1 channel.

All these channels are rather trivial examples. In the next subsections, we review some more interesting examples.

3.2.1 Decoherence of a qubit

The qubit decoherence channel was defined in the Kraus form in section 2.2.2. The output of this channel corresponds to the residual state when a qubit is measured in the computational basis (but where we don't see the classical part of the outcome). Here's a Stinespring circuit for the decoherence channel.

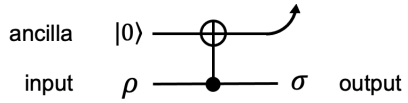

Figure 17: Decoherence of a qubit channel.

How does this work? Consider the case of a pure state $\alpha_0 |0\rangle + \alpha_1 |1\rangle$. The CNOT gate causes the state of the two qubits to become $\alpha_0 |00\rangle + \alpha_1 |11\rangle$, and tracing out

the first qubit yields the mixed state

$$\begin{bmatrix} |\alpha_0|^2 & 0 \\ 0 & |\alpha_1|^2 \end{bmatrix} = |\alpha_0|^2 |0\rangle\langle 0| + |\alpha_1|^2 |1\rangle\langle 1| , \tag{93}$$

which is consistent with the definition of this channel. So at least this works for the special case of pure states.

Exercise 3.1 (easy)**.** *Show that the circuit in figure 17 implements the decoherence channel, as defined in section 2.2.2.*

3.2.2 Reset channel

Here's a very simple channel that I haven't mentioned before, that I'll call the *reset channel*. The input is a qubit and the output is a qubit in state $|0\rangle$ (regardless of what the input state is). Here's a very simple Stinespring circuit for this.

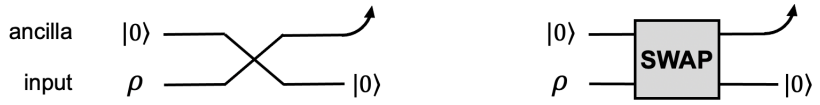

Figure 18: Circuit for the reset channel (with two different notations for the SWAP gate).

It's obvious that this circuit works: it traces out the input qubit and produces a qubit in state $|0\rangle$ as output. The following question about the reset channel is non-trivial.

Exercise 3.2. *Express the reset circuit in the Kraus form (in terms of Kraus operators). (Hint: two Kraus operators suffice.)*

Later in this section, we will see recipes for converting between the Stinespring form and the Kraus form, but there is a simple solution to the above which you might try to discover directly.

3.2.3 Depolarizing channel

The *depolarizing channel* is fundamental, and used as a natural model of noise. We'll be seeing more of this channel when we get to the subject of quantum error-correcting codes. The channel is parameterized by $p \in [0, 1]$, and it maps an input qubit in state $\rho \in \mathbb{C}^{2\times 2}$ to an output qubit in state

$$p\rho + (1-p) \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} . \tag{94}$$

In other words, with probability p, the state is left alone and with probability $1-p$ the state is changed to the maximally mixed state $\left[\begin{smallmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{smallmatrix}\right]$. Here's what the effect of this channel looks like on the Bloch sphere.

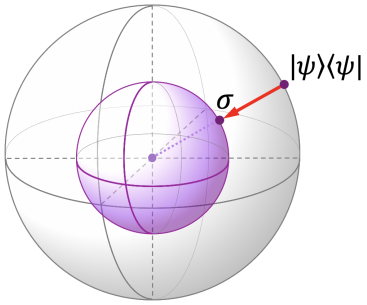

Figure 19: Effect of depolarizing channel on pure state $|\psi\rangle\langle\psi|$.

The maximally mixed state is at the centre of the Bloch sphere. The channel moves states towards the centre. In fact, the channel shrinks the entire Bloch sphere by a factor of p towards the centre.

Can we represent this channel in Stinespring form? Here's one Stinespring circuit for this channel, where $R = \left[\begin{smallmatrix} \sqrt{1-p} & -\sqrt{p} \\ \sqrt{p} & \sqrt{1-p} \end{smallmatrix}\right]$.

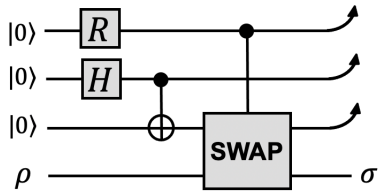

Figure 20: The depolarizing channel in the Stinespring form.

At first glance, this circuit may look complicated. But we can understand it by first looking at the two middle qubits. The H and CNOT gate are manufacturing a Bell state $\frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|11\rangle$. Note that each qubit of the Bell state is in state $\left[\begin{smallmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{smallmatrix}\right]$. Then the rest of the circuit applies

$$\begin{cases} \mathsf{SWAP} & \text{with prob. } p \\ I & \text{with prob. } 1-p. \end{cases} \tag{95}$$

This can be seen by noting that the circuit is equivalent to this.

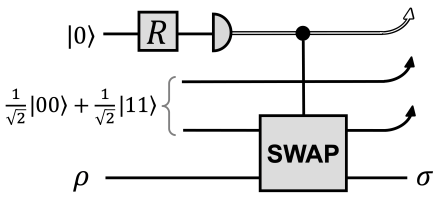

Figure 21: An equivalent circuit for the depolarizing channel.

Notice that this Stinespring form uses three ancilla qubits. Can this channel be constructed with fewer ancilla qubits?

A very easy way to reduce the ancilla to two qubits is to skip the Bell state and just initialize one ancilla qubit to the mixed state $\frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|$. Technically, this isn't in the Stinespring form of figure 15, since that requires all ancilla qubits to be initialized to a pure state. But a construction allowing an ancilla to be initialized to a mixed state might nevertheless be useful in some contexts.

However, it turns out that there is a different Stinespring circuit for the depolarizing channel that uses only two ancilla qubits initialized to state $|00\rangle$. The construction is rather elegant, and I leave it as an exercise.

Exercise 3.3. *Give a Stinespring form for the depolarizing channel that uses only two ancilla qubits in initial state* $|00\rangle$*.*

Is two qubits the optimal size of the ancilla? It turns out that one ancilla qubit is not enough for the depolarizing channel.

Exercise 3.4. *Prove that there is no Stinespring form for the depolarizing channel that uses only one ancilla qubit.*

And we can make more a fine-grained distinction regarding the size of the ancilla: what if the ancilla is allowed to be a qutrit?

Exercise 3.5. *Is there a Stinespring form for the depolarizing channel that uses one qutrit as ancilla? Justify your answer.*

3.3 Equivalence of Kraus and Stinespring channels

Recall from Definition 2.1 that $A_0, A_1, \ldots, A_{\ell-1}$ is a sequence of Kraus operators if

$$\sum_{k=0}^{\ell-1} A_k^* A_k = I. \tag{96}$$

Associated with any sequence of Kraus operators, we have two transformations: a Kraus measurement and a Kraus channel. We're now going to prove two theorems.

Theorem 3.1 (Kraus to Stinespring)**.** *Any transformation in the Kraus form can be simulated in the Stinespring form.*

Theorem 3.2 (Stinespring to Kraus)**.** *Any transformation in the Stinespring form can be simulated in the Kraus form.*

3.3.1 Kraus to Stinespring

In this section we prove Theorem 3.1. For Kraus operators $A_0, A_1, \ldots, A_{\ell-1} \in \mathbb{C}^{c\times d}$, consider the block matrix

$$\begin{bmatrix} A_0 \\ A_1 \\ \vdots \\ A_{\ell-1} \end{bmatrix}, \tag{97}$$

which is an $\ell c \times d$ matrix. The columns of this matrix are orthonormal because

$$\begin{bmatrix} A_0^* & A_1^* & \cdots & A_{\ell-1}^* \end{bmatrix} \begin{bmatrix} A_0 \\ A_1 \\ \vdots \\ A_{\ell-1} \end{bmatrix} = \sum_{k=0}^{\ell-1} A_k^* A_k = I. \tag{98}$$

One consequence of this is that $d \leq \ell c$. Otherwise, the number of orthonormal vectors would have to exceed the dimension of the space in which they exist, which is impossible.

So we have ℓ Kraus operators that are $c \times d$ matrices and $d \leq \ell c$. To make the dimensions work out nicely, I'd like to assume that d divides ℓc. For this to hold, we might have to increase ℓ. It's straightforward to show that this can be done, where the new value of ℓ is less than double the original value of ℓ. If ℓ is increased we can add more Kraus operators that are zero matrices. Note that the larger set of matrices are still Kraus operators. So we can assume that d divides ℓc and set $m = \frac{\ell c}{d}$. Then we have $\ell c = md$.

Now, consider to the block matrix in Eq. (97) of Kraus operators again. Since its columns are orthonormal, we can extend this set of ℓc-dimensional column vectors to be an orthonormal basis of size ℓc. If we add these column vectors to the block

matrix, we end up with a square unitary matrix. Call this $\ell c \times \ell c$ matrix U, which is of the form

$$U = \left[\begin{array}{c|c} A_0 & \\ A_1 & \\ \vdots & W \\ A_{\ell-1} & \end{array}\right]. \tag{99}$$

Now consider this circuit.

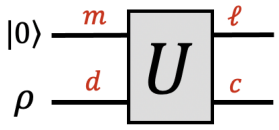

Figure 22: Stinespring circuit (omitting the final measurement/trace-out stage).

The input to the circuit consists of two registers: an m-dimensional ancilla and our d-dimensional input state. The circuit applies U to this. We can calculate the density matrix of the output state as

$$U\big(|0\rangle\langle 0| \otimes \rho\big)U^* = \left[\begin{array}{c|c} A_0 & \\ A_1 & \\ \vdots & W \\ A_{\ell-1} & \end{array}\right] \begin{bmatrix} \rho & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \left[\begin{array}{c|ccc} A_0^* & A_1^* & \cdots & A_{\ell-1}^* \\ \hline & & W^* & \end{array}\right] \tag{100}$$

$$= \left[\begin{array}{c|c} A_0\rho & \\ A_1\rho & \\ \vdots & 0 \\ A_{\ell-1}\rho & \end{array}\right] \left[\begin{array}{c|ccc} A_0^* & A_1^* & \cdots & A_{\ell-1}^* \\ \hline & & W^* & \end{array}\right] \tag{101}$$

$$= \begin{bmatrix} A_0^*\rho A_0 & A_0^*\rho A_1 & \cdots & A_0^*\rho A_{\ell-1} \\ A_1^*\rho A_0 & A_1^*\rho A_1 & \cdots & A_1^*\rho A_{\ell-1} \\ \vdots & \vdots & \ddots & \vdots \\ A_{\ell-1}^*\rho A_0 & A_{\ell-1}^*\rho A_1 & \cdots & A_{\ell-1}^*\rho A_{\ell-1} \end{bmatrix}. \tag{102}$$

For the Stinespring channel, the final step is to trace out the first register. This partial trace is the sum of the $c \times c$ blocks along the diagonal, which is

$$\sum_{k=0}^{\ell-1} A_k \rho A_k^*. \tag{103}$$

For the Stinesrping measurement, the final step is to measure the first register in the computational basis. This measurement is defined in the Kraus form in section 2.1.3, and it is straightforward to deduce that the probability that the outcome of this measurement is k is $\mathrm{Tr}(A_k \rho A_k^*)$.

3.3.2 Stinepring to Kraus

In this section I give a brief overview of the proof of Theorem 3.2. The Stinespring form consists of three stages. The first two are: adding an ancilla in state $|0\rangle$; and then applying a unitary operation U. The third stage is the partial trace for the Stinespring channel, and the measurement of the first register for the Stinespring measurement. Note that, in section 2, we have Kraus forms for each of these individual operations. We can compose these Kraus forms to obtain a Kraus channel from a Stinespring channel. And we can compose them to obtain a Kraus measurement from a Stinespring measurement.

3.4 Unifying measurements and channels

I have been describing state transitions as if there's a clear dichotomy between measurements and channels. You either measure and get a classical outcome and a residual state or you apply a channel and get just a quantum state as outcome. In fact, there's a general notion that unifies these.

Let $f : \{0, 1, \ldots, \ell - 1\} \to T$ be some function. Suppose that we apply the Kraus measurement and then apply f to the classical outcome. So the classical outcome is $f(k)$, rather than k.

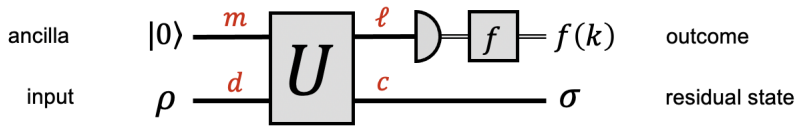

Figure 23: Generalized quantum transformation.

If f is a constant function, then seeing $f(k)$ provides us with no information about k. So that corresponds to the case of a channel. The other extreme case is where f is a bijection, for which knowing $f(k)$ provides full information about k. And there are in-between cases where f is not constant nor a bijection. In those cases, we receive *partial* information about k. The classical outcome $f(k)$ might narrow down the possible values of k, but without uniquely determining k.

3.5 POVM measurements

A final topic concerns *POVM measurements* (POVM stands for *positive operator valued measure*[7]). This is a simplified way of describing a Kraus measurement, that works if we *only* care about the classical outcome (so we do *not* care about the residual quantum state).

Recall that, for Kraus operators $A_0, A_1, \ldots, A_{\ell-1}$, the associated measurement of a state ρ produces outcome k with probability

$$\mathrm{Tr}(A_k \rho A_k^*) = \mathrm{Tr}(\rho\, A_k^* A_k). \tag{104}$$

For each Kraus operator A_k, define $E_k = A_k^* A_k$. All we need to know is the sequence $E_0, E_1, \ldots, E_{\ell-1}$ to define the classical part of the measurement outcome. And we can characterize such sequences $E_0, E_1, \ldots, E_{\ell-1}$ in a simple way.

Definition 3.1 (POVM elements)**.** *A sequence $E_0, E_1, \ldots, E_{\ell-1}$ is a* sequence of POVM elements *if, for all k, it holds that E_k is positive,*[8] *and*

$$E_0 + E_1 + \cdots + E_{\ell-1} = I. \tag{105}$$

For a sequence of POVM elements $E_0, E_1, \ldots, E_{\ell-1}$ and a quantum state ρ, applying the associated *POVM measurement* produces outcome $k \in \{0, 1, \ldots, \ell - 1\}$ with probability $\mathrm{Tr}(\rho E_k)$.

⚠ A word of caution: for a POVM measurement, there is no way to define a residual quantum state. This is because we cannot uniquely deduce a set of underlying Kraus operators from POVM elements. We can find an A_k such that $E_k = A_k^* A_k$, but this A_k is not unique, and for a different choices of A_k the residual state is different. So we should use Kraus operators if we want to be able to refer to the quantum state after the measurement.

[7] Regarding terminology, a *positive operator valued measure* is a generalization of the notion of a *measure*, that you may have seen in probability theory or functional analysis. The word "measure" is distinct from "measurement". So it makes sense to say "POVM measurement".

[8] Meaning that E_k is normal and all its eigenvalues are nonnegative reals.