Guided solutions to Nielsen and Chuang: Quantum Computation and Quantum Information

Jacob Watkins

Last updated: Jan 2023

Chapter 0

Plan and Progress

0.1   Purpose of this solution manual

The textbook *Quantum Computation and Quantum Information* by Nielsen and Chuang (NC) is often considered the authoritative textbook in the subject by those in the field. Since the book was published, the field has grown and evolved enormously, and rich subfields have branched off to be of interest in their own right. In parallel with the science, more resources have become available for those interested in learning quantum computing. These include textbooks, tutorials, blog posts, software packages, and small quantum devices available to the public for playful exploration. Despite these developments, NC retains much of its original value and relevance as a comprehensive introduction to this subject.

Surprisingly, there seems to be a lack of complete, easily available, and guided solutions to NC problems. Partial solutions exist in several places on the internet, including GitHub. It appears, taken together, they have covered chapters 2,3,4 and 9 almost completely, with scattered solutions for other chapters. In many cases, these solutions are not worked through pedagogically. The goal of this document is to provide a comprehensive set of worked-through solutions to as many (if not all) NC exercises as possible, as a supplement to the textbook for those who need assistance in getting unstuck on a problem.

Under what conditions can a student best learn a technical subject such as quantum computing? For those students who learn in a formal class setting at a university, or under mentorship from an expert, the student, ideally, has access not only to the relevant information, but also *feedback* that allows them to refine their skills. In today's age of the internet, almost everyone has some access to the same raw information on quantum computing. However, this may not be enough to truly equalize the playing field for those outside academic institutions, or for those who attend a university with little to no expertise in the subject. A lack of feedback and guidance for independent learners could hamper their progress, since they are unaware of conceptual traps or useful mental shortcuts that are often discussed in informal settings such as office hours.

It has sometimes been suggested to me, by my previous professors or certain textbook authors, that a bright, motivated student can learn on their own directly from a textbook, without recurse to much else. While this might be true in some sense, a more important question is how efficiently this learning occurs, compared to those who have guided practice with feedback. I hope that these worked-through solutions will provide a different kind of resource to independent quantum learners, a kind which is sorely needed.

I would greatly appreciate any involvement from the community on this project. If anyone wants to help contribute, please send me a message!

0.2   Progress to completion

This progress was last updated May 19th, 2022.

- Chapter 1: All in-text exercises done. Two open ended problems left alone.

- Chapter 2: Exercises 1-75 complete. Many exercises remain.

- Chapter 3: 1 "complete".

- Chapter 4: 1-9 complete.

- Chapter 5: 1-14 complete, 15-29 remain, along with 6 practice problems.

- Chapter 6: 1-8 complete.

- Chapter 7: 1 exercise complete.

- Chapter 8: 1-6 complete.

- Chapter 9: All exercises complete (end-of-chapter problems remain).

- Chapter 10: 1-11 and 29-40 complete.

- Chapter 11: Only first exercise.

- Chapter 12: 1-4.

- Appendix 1: All exercises complete!

- Appendix 2: Some.

- Appendix 3: Only the first.

- Appendix 4: All exercises complete except 17 and part 2 of problem 1.

- Appendix 5: Almost done. Problem 1 needs to be cleaned up.

- Appendix 6: All exercises complete!

Chapter 1

Introduction and overview

Exercise 1.1 (Probabilistic classical algorithm)

Suppose that the problem is not to distinguish between the constant and balanced functions with certainty, but rather, with some probability of error $\epsilon < 1/2$. What is the performance of the best classical algorithm for this problem?

Solution: Intuitively, we should be able to check whether a function f as constant or balanced by looking at the output of a few randomly chosen inputs. If f were balanced, it would be very unlucky if all of the inputs we tried gave the same answer, just as flipping a coin ten times rarely gives all heads (or tails).

Lets make these ideas more precise. Suppose I sample s inputs at random from the 2^n possible bitstrings of length n. If any two of these inputs have a different output under f, I can be certain that f is balanced. If all of my inputs give the same answer, it makes sense to guess that f is constant. Of course, there is a chance that f is balanced, and I merely got unlucky. The probability that all my outputs are the same, assuming f is in fact balanced, is

$$\frac{1}{2^s} \tag{1.1}$$

where we've assumed each input is sampled (uniformly) randomly and independently. We can make this "failure probability" as small as we like. By choosing $s > \log 1/\epsilon$, the failure probability is less than ϵ. Therefore, we will be able to distinguish f constant vs balanced with an arbitrarily small chance of failure.

Notice that, not only can we make ϵ arbitrarily small, we can do so "exponentially quickly". Each new sample shrinks our failure probability by $1/2$. This suggests that, in practice, we should have no trouble solving this problem with regular computers with a high probability of success.

Exercise 1.2

Explain how a device which, upon input of one of two non-orthogonal quantum states $|\psi\rangle$ or $|\varphi\rangle$ correctly identified the state, could be used to build a device which cloned the states $|\psi\rangle$ and $|\varphi\rangle$, in violation of the no-cloning theorem. Conversely, explain how a device for cloning could be used to distinguish non-orthogonal quantum states.

Solution: Suppose we could distinguish any two quantum states $|\psi\rangle$ and $|\varphi\rangle$ perfectly, given only a

single copy. Such a device could store the result in a single (classical or quantum) bit, with, say, zero for $|\psi\rangle$ and one for $|\varphi\rangle$. Then we can define a unitary operation which will reconstruct the state from the result. Specifically, let CU be the controlled unitary which acts on a "blank" quantum system S by preparing the state $|\psi\rangle$ or $|\varphi\rangle$ depending on whether the result bit was zero or one. Applying this operation after the measurement would prepare another copy of the original state, effectively cloning it.

Conversely, cloning could be used to distinguish any two quantum states. By creating many copies of the state of interest, and performing suitable measurements on the copies, we can fully determine the quantum state to arbitrary precision. Or at least enough to tell them apart.

Together, this shows that the ability to clone and the ability to distinguish arbitrary states are equivalent in power. Since we cannot do one of them, we cannot do either!

Chapter 2

Introduction to quantum mechanics

Exercise 2.1: (Linear dependence: example)

Show that $(1, -1)$, $(1, 2)$ and $(2, 1)$ are linearly dependent.

Solution: By inspection (i.e. staring long enough), one might be able to see that $(1, -1) + (1, 2) = (2, 1)$, hence,

$$(1, -1) + (1, 2) - (2, 1) = 0. \tag{2.1}$$

This demonstrates linear dependence, the last rearrangement being more of a formality. Essentially, one of the vectors can be reached by "moving along" the directions of the other two.

A more systematic approach would be to cast this problem as a system of linear equations. We want to know if one of the three vectors is in the span of the other two. It turns out that this is equivalent to asking whether the matrix equation

$$\begin{pmatrix} 1 & 1 & 2 \\ -2 & 2 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{2.2}$$

has any "nontrivial" solutions (nontrivial meaning at least one of a, b, c is nonzero). This can be solved using your favorite method for linear equations (row reduction, substitution, computers). In fact, there are an infinite number of possible answers: give it a try!

Exercise 2.2: (Matrix representations: example)

Suppose V is a vector space with basis vectors $|0\rangle$ and $|1\rangle$, and A is a linear operator from V to V such that $A|0\rangle = |1\rangle$ and $A|1\rangle = |0\rangle$. Give a matrix representation for A, with respect to the input basis $|0\rangle, |1\rangle$, and the output basis $|0\rangle, |1\rangle$. Find input and output bases which give rise to a different matrix representation of A.

Solution: Making the identification

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{2.3}$$

it is easy to see that the matrix representation M of A must be

$$M = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{2.4}$$

On the other hand, say we choose the output basis to be $|1\rangle, |0\rangle$ instead of $|0\rangle, |1\rangle$ (note that order matters). In this case, M would take the form of the identity matrix.

Exercise 2.3: (Matrix representation for operator products)

Suppose A is a linear operator from vector space V to vector space W, and B is a linear operator from vector space W to vector space X. Let $|v_i\rangle, |w_j\rangle$, and $|x_k\rangle$ be bases for the vector spaces V, W, and X respectively. Show that the matrix representation for the linear transformation BA is the matrix product of the matrix representations for B and A, with respect to the appropriate bases.

   Solution: Let us compute the action of BA on a basis vector $|v_i\rangle$.

$$\begin{aligned} BA\,|v_i\rangle &= B(A\,|v_i\rangle) \\ &= B(\sum_j A_{ji}\,|w_j\rangle) \\ &= \sum_j A_{ji} B\,|w_j\rangle \\ &= \sum_j \sum_k A_{ji} B_{kj}\,|x_k\rangle \\ &= \sum_k \left(\sum_j B_{kj} A_{ji} \right) |x_k\rangle \\ &= \sum_k (BA)_{ki}\,|x_k\rangle\,. \end{aligned} \tag{2.5}$$

In the last step, we identified the coefficient as the product of the two matrix representations of A and B. This gives the result.

Exercise 2.4: (Matrix representation for the identity)

Show that the identity operator on a vector space V has a matrix representation which is one along the diagonal and zero everywhere else, if the matrix representation is taken with respect to the same input and output bases. This matrix is known as the *identity matrix*.

   Solution: Let M be the matrix representation of the identity operator I with respect to some basis $\{|v_i\rangle\}$. Then,

$$I\,|v_i\rangle = |v_i\rangle = \sum_j \delta_{ji}\,|v_j\rangle, \tag{2.6}$$

where δ_{ji} is the Kronecker delta. The coefficients δ_{ji} are exactly those of the identity matrix $\mathbb{1}$, hence $M = \mathbb{1}$.

Exercise 2.5

Verify that $(\cdot, \cdot)$ just defined is an inner product on $\mathbb{C}^n$.

Solution: Let's use letters x, y, z to denote elements of $\mathbb{C}^n$ (lists of n complex entries), and let $\lambda \in \mathbb{C}$ be a complex number. To check linearity, let's first consider $(x, y + z)$. Using the definition,

$$
\begin{aligned}
(x, y + z) &= \sum_{i=1}^{n} x_i^*(y_i + z_i) \\
&= \sum_{i=1}^{n} (x_i^* y_i + x_i^* z_i) \\
&= \sum_{i=1}^{n} x_i^* y_i + \sum_{i=1}^{n} x_i^* z_i \\
&= (x, y) + (x, z).
\end{aligned}
\tag{2.7}
$$

This shows that $(\cdot, \cdot)$ satisfies the additive portion of linearity. To show the multiplication part, consider $(x, \lambda y)$.

$$
\begin{aligned}
(x, \lambda y) &= \sum_{i=1}^{n} x_i^* \lambda y_i \\
&= \lambda \sum_{i=1}^{n} x_i^* y_i \\
&= \lambda (x, y).
\end{aligned}
\tag{2.8}
$$

The two properties just shown, used together, can be used to demonstrate linearity as given in the text. Can you see why? As a starting point, can we generalize to have sums of more than two vectors?

For the "skew-symmetry" property (2), we perform the following calculation.

$$
\begin{aligned}
(x, y) &= \sum_{i=1}^{n} x_i^* y_i \\
&= \left(\sum_{i=1}^{n} x_i y_i^* \right)^* \\
&= (y, x)^*
\end{aligned}
\tag{2.9}
$$

We used a nice property of complex conjugates: they distribute over sums and products of complex numbers. Finally, we demonstrate property (3) with another calculation from the definition of $(\cdot, \cdot)$.

$$
\begin{aligned}
(x, x) &= \sum_{i=1}^{n} x_i^* x_i \\
&= \sum_{i=1}^{n} |x_i|^2
\end{aligned}
\tag{2.10}
$$

We see that (x, x) is a sum of n nonnegative entries $|x_i|^2$, and is therefore also nonnegative. Moreover, (x, x) will be positive unless every term in the sum is zero, that is, if $x_i = 0$ for every i. In this case, we simply have $x = 0$. Therefore $(x, x) = 0$ if and only if $x = 0$. We've now shown that $(\cdot, \cdot)$ defines a valid inner product.

Exercise 2.6

Show that any inner product $(\cdot, \cdot)$ is conjugate-linear in the first argument,

$$\left(\sum_i \lambda_i \left| w_i \right\rangle, \left| v \right\rangle \right) = \sum_i \lambda_i^* (\left| w_i \right\rangle, \left| v \right\rangle) \tag{2.11}$$

Solution: For simplicity, I will drop the ket notation, using w_i instead of $\left| w_i \right\rangle$, for example. The conjugate-linearity follows from two of the three properties in the definition of an inner product. The following calculation shows this.

$$
\begin{aligned}
\left(\sum_i \lambda_i w_i, v \right) &= \left(v, \sum_i \lambda_i w_i \right)^* \\
&= \sum_i \lambda_i^* (v, w_i)^* \\
&= \sum_i \lambda_i^* (w_i, v).
\end{aligned}
\tag{2.12}
$$

In particular, we only needed to use properties (1) and (2)! Some properties of complex conjugation were invoked such as distribution over multiplication: $(ab)^* = a^* b^*$.

Exercise 2.7

Verify that $\left| w \right\rangle \equiv (1, 1)$ and $\left| v \right\rangle \equiv (1, -1)$ are orthogonal. What are the normalized forms of these vectors?

Solution: To see they are orthogonal, we compute their inner product as defined in the text.

$$(w, v) := \sum_{i=1}^{2} w_i^* v_i = 1^2 - 1^2 = 0 \tag{2.13}$$

Note that the complex conjugate in the definition becomes unimportant when both vectors are real. To normalize the vectors, let's first compute their norms.

$$\|w\| = \sqrt{1^2 + 1^2} = \sqrt{2} \tag{2.14}$$
$$\|v\| = \sqrt{1^2 + (-1)^2} = \sqrt{2} \tag{2.15}$$

Hence, the normalized versions $\hat{w}$, $\hat{v}$ are $(1, 1)/\sqrt{2}$ and $(1, -1)/\sqrt{2}$ respectively. As a check, can you verify that each of these really has unit norm?

Exercise 2.8

Prove that the Gram-Schmidt procedure produces an orthonormal basis for V.

Solution: There are several things we need to check that are packed into this sentence. We need to check that the $\left| v_k \right\rangle$ are orthonormal, meaning they are orthogonal and normalized. Then we need to check that they are a basis for V.

These facts are not necessarily separate. If $d = \dim V$ is the dimension of V, a set of d orthogonal, nonzero vectors will always form a basis. If it helps, this can be illustrated from the xy axes of the plane, or

an xyz grid in three dimensions. We will take this statement for granted to solve the problem, but if you'd like to learn how to prove this, consult a linear algebra textbook with chapters on linear dependence and orthogonality.

The Gram-Schmidt procedure works by subtracting off the component of each $|w_{k+1}\rangle$ that is parallel to the preceding k vectors, then normalizing the result. The "normalization" step means dividing by the norm, which we see is done in equation 2.17 of the text. We just need to make sure that the norm is not actually zero. If it was zero, then from the numerator of 2.17 (in text) we would have

$$|w_{k+1}\rangle = \sum_{i=1}^{k} \langle v_i|w_{k+1}\rangle |v_i\rangle. \tag{2.16}$$

Since each $|v_i\rangle$ is, by construction, in the subspace spanned by the $|w_j\rangle$ for $j \leq i$, this would imply $|w_{k+1}\rangle$ is a superposition of other $|w_j\rangle$, and therefore they are not linearly dependent as we assumed! Since this can't happen, it must be that the norm is not zero, so $|v_i\rangle$ as written is well defined and normalized.

Now let's check that the vectors are orthogonal. By the way we've constructed each $|v_i\rangle$ based on the previous ones, it's easiest to proceed by checking that each $|v_{k+1}\rangle$ is orthogonal to all $|v_j\rangle$ for $j \leq k$. Since, normalization does not matter, we can also just consider the numerator of equation 2.17, which we'll call $|x_{k+1}\rangle$. For any $j \leq k$, we have

$$\langle v_j|x_{k+1}\rangle = \langle v_j|w_{k+1}\rangle - \sum_{i=1}^{k} \langle v_i|w_{k+1}\rangle \langle v_j|v_i\rangle. \tag{2.17}$$

Let's assume we've already successfully shown that all pairs $|v_i\rangle, |v_j\rangle$ are orthogonal for $i, j \leq k$, $i \neq j$. Then $\langle v_j|v_i\rangle = \delta_{ij}$ in equation (2.17), where δ_{ij} is the Kronecker delta. This captures the orthonormality. We'll now show that $|v_{k+1}\rangle$ should be included in the set as well, extending the set by one. Using the orthonormality,

$$\langle v_j|x_{k+1}\rangle = \langle v_j|w_{k+1}\rangle - \langle v_j|w_{k+1}\rangle$$
$$= 0 \tag{2.18}$$

This shows $|v_1\rangle, |v_2\rangle, \ldots, |v_{k+1}\rangle$ form an orthonormal set. Repeating for $k+2, k+3$ and so on up until the dimension d, we can show that all the $|v_k\rangle$ form an orthonormal basis. This shows the Gram-Schmidt procedure works.

For those interested, one can formalize the above argument into a proof by mathematical induction. For those who found the argument confusion, try drawing some pictures to better understand the first step of the Gram-Schmidt procedure. What is going on and why does it work?

Exercise 2.9 (Pauli operators and the outer product)

The Pauli matrices (Figure 2.2 on page 65) can be considered as operators with respect to an orthonormal basis $|0\rangle, |1\rangle$ for a two-dimensional Hilbert space. Express each of the Pauli operators in the outer product notation.

Solution: The identity matrix, as was shown in the text, is given by the completeness relation for a single qubit.

$$I = \sigma_0 = |0\rangle\langle 0| + |1\rangle\langle 1| \tag{2.19}$$

To construct the other ones, we'll use the following trick. Any matrix can be broken into a linear combination of matrices, each with only one nonzero entry that is equal to one. For two dimensions, it looks like this.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = a \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + b \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + c \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} + d \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \tag{2.20}$$

What is the outer product representation of each such matrix? Let $E^{(ij)}$ represent the operator in which only the ijth matrix element is one, the rest being zero. This implies that almost all basis vectors, when multiplied by $E^{(ij)}$, are sent to zero. The only one which isn't is $|j\rangle$, and this is sent to $|i\rangle$. This leads to the outer product formula

$$E^{(ij)} = |i\rangle \langle j|, \tag{2.21}$$

which can be checked. Thus using the decomposition (2.20), we can convert any matrix into an outer product representation. We do that now for each pauli matrix.

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \rightarrow |0\rangle \langle 1| + |1\rangle \langle 0|$$

$$\sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \rightarrow -i |0\rangle \langle 1| + i |1\rangle \langle 0| \tag{2.22}$$

$$\sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \rightarrow |0\rangle \langle 0| - |1\rangle \langle 1|$$

This suggests a general method: notice that the ijth element gets mapped to a coefficient for the term with operator $|i\rangle \langle j|$. Can you write the outer product form of the matrix given in equation (2.20)?

Exercise 2.10

Suppose $|v_i\rangle$ is an orthonormal basis for an inner product space V. What is the matrix representation for the operator $|v_j\rangle \langle v_k|$ with respect to the $|v_i\rangle$ basis?

Solution: The operator $|v_j\rangle \langle v_k|$ sends every $|v_i\rangle$ to zero except when $i = k$, in which case

$$(|v_j\rangle \langle v_k|) |v_k\rangle = |v_j\rangle . \tag{2.23}$$

This means the matrix representation M has columns of all zeros except in the kth column. And even then, the only nonzero entry is the jth row, because that will make sure $|v_k\rangle \rightarrow |v_j\rangle$ as column vectors. Hence, M only has a single nonzero entry, the jkth entry, with value one. This agrees with our discussion from the previous exercise.

Exercise 2.11 (Eigendecomposition of the Pauli matrices)

Find the eigenvectors, eigenvalues, and diagonal representations of the Pauli matrices X, Y, and Z.

Solution: Let's start with Z since this is the simplest case. Why is it simple? Because Z is already a diagonal matrix! We can read off the eigenvalues as the diagonal entries: -1 and 1. Moreover, as for any diagonal matrix, the eigenvectors are just the standard basis vectors.

$$Z \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad Z \begin{pmatrix} 0 \\ 1 \end{pmatrix} = - \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{2.24}$$

The diagonal representation is given in the text, equation 2.29.

I will give a complete solution for X, and I suggest you try to solve for Y in the same manner. Let's proceed with the standard method of the characteristic equation. Suppose λ is an eigenvalue of X, with eigenvector $|v\rangle$, such that

$$X |v\rangle = \lambda |v\rangle$$
$$(X - \lambda I) |v\rangle = 0. \tag{2.25}$$

Take a minute to think about why the second line follows from the first. What rearranging did I do?

The fact that there is a nonzero $|v\rangle$ such the 2nd line of equation(2.25) holds means that $X - \lambda I$ is a "singular matrix". This means many equivalent things, one of which is that the determinant is zero.

$$\det(X - \lambda I) = 0 \tag{2.26}$$

As a reminder, all of this applies more generally to determining the eigenvalues and eigenvectors of any square matrix. Let's calculate the determinant to solve for λ.

$$\begin{aligned} \det(X - \lambda I) = \det \begin{pmatrix} -\lambda & 1 \\ 1 & -\lambda \end{pmatrix} &= 0 \\ \lambda^2 - 1 &= 0 \\ \lambda &= \pm 1. \end{aligned} \tag{2.27}$$

Thus, the eigenvalues of X are ± 1. Since X is two-dimensional, each of these eigenvalues will have exactly one eigenvector. How do we find them? Let $|+\rangle$ and $|-\rangle$ correspond to the eigenvectors for $+1$ and -1 respectively. Going back to equation (2.25), this means

$$(X - I)|+\rangle = 0 \tag{2.28}$$
$$(X + I)|-\rangle = 0. \tag{2.29}$$

These are linear equations we can solve, using our favorite method! Gaussian elimination is a great one in general. Solving each of these we will find the answer to be

$$\begin{aligned} |+\rangle &\propto \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ |-\rangle &\propto \begin{pmatrix} 1 \\ -1 \end{pmatrix} \end{aligned} \tag{2.30}$$

Here, $\propto$ means "proportional to" by some constant c which is undetermined. Any c will solve (2.28), so there are an infinite number of solutions. However, we want *normalized* solutions, so that $\langle +|+\rangle = \langle -|-\rangle = 1$. Since the norm of $|+\rangle$ and $|-\rangle$ is $\sqrt{2}$ (check it!) a good choice for c then is $1/\sqrt{2}$. But $-1/\sqrt{2}$ would also work, wouldn't it? Anyways, with this normalization, our eigenvectors are determined, and our diagonal decomposition is given as

$$X = |+\rangle \langle +| - |-\rangle \langle -|. \tag{2.31}$$

(Sorry if the notation strains your eyes a bit, it does mine as well!)

This concludes our analysis for X. I encourage you to try solving for Y, referencing the above steps as needed. I will simply list the solution. The eigenvalues are still $+1$ and -1, and the corresponding eigenvectors, which we'll call $|+i\rangle$ and $|-i\rangle$, are given by

$$\begin{aligned} |+i\rangle &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix} \\ |-i\rangle &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix}. \end{aligned} \tag{2.32}$$

Therefore, the diagonal decomposition is given by

$$Y = |+i\rangle \langle +i| - |-i\rangle \langle -i|. \tag{2.33}$$

Having solved the problem stated, let me just leave some parting comments. It is no coincidence that all the Paulis have the same eigenvalues, because they are related by a change of basis. Intuitively, changing the basis ("point of view") should not affect the stretching/compressing properties of the matrix. Once the change of basis is known, it can be used to relate all the eigenvectors together as well. These concepts will become more transparent in Chapter 4!

Exercise 2.12

Prove that the matrix

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \tag{2.34}$$

is not diagonalizable.

Solution: To investigate this claim, let's find the eigenvalues and eigenvectors. Let A be this matrix. The determinant of $A - \lambda I$ is then given by

$$\det \begin{pmatrix} 1 - \lambda & 0 \\ 1 & 1 - \lambda \end{pmatrix} = (1 - \lambda)^2. \tag{2.35}$$

This is equal to zero only when $\lambda = 1$, so there is only one eigenvalue. It may be tempting to conclude here that A is not diagonalizable: don't we only have one eigenvector? Not so: there could be *two* independent eigenvectors corresponding to this single eigenvalue. This is sometimes called "degeneracy", especially within physics. Let's determine which case is true by calculating the eigenvectors. If $|v\rangle$ is an eigenvector, it must satisfy

$$(A - I)|v\rangle = 0$$
$$\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}|v\rangle = 0. \tag{2.36}$$

We can solve this system any way we like. There is only one (linearly independent) solution.[1]

$$|v\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{2.37}$$

This is the unique eigenvector of A. But A cannot be written in terms of $|v\rangle\langle v|$ alone (why?). This shows that A is not diagonalizable.

Exercise 2.13

If $|w\rangle$ and $|v\rangle$ are any two vectors, show that $(|w\rangle\langle v|)^\dagger = (|v\rangle\langle w|)$.

Solution: We could solve this from the definition, but it also falls out from a more interesting consideration. We can think of $\langle w|$ as a linear operator from the Hilbert space H to $\mathbb{C}$, and $|v\rangle$ as another linear operator from $\mathbb{C}$ back into H. Then $|w\rangle\langle v|$ is just the composition of the two maps, and we have $(|w\rangle\langle v|)^\dagger = \langle v|^\dagger |w\rangle^\dagger = |v\rangle\langle w|$.

But for sake of clarity, let's also work from the definition. For any two vectors ϕ, ψ, we have

$$\begin{aligned}
((|w\rangle\langle v|)^\dagger \phi, \psi) &= (\phi, |w\rangle\langle v|\psi) \\
&= \langle\phi|w\rangle\langle v|\psi\rangle \\
&= (\langle\psi|v\rangle\langle w|\phi\rangle)^* \\
&= (\psi, |v\rangle\langle w|\phi)^* \\
&= (|v\rangle\langle w|\phi, \psi).
\end{aligned} \tag{2.38}$$

Comparing, the first and last line, since this holds for arbitrary ϕ, ψ, we have

$$(|w\rangle\langle v|)^\dagger = |v\rangle\langle w|. \tag{2.39}$$

---

[1] By this I mean any two solutions are related by a constant.

Exercise 2.14 (Anti-linearity of the adjoint)

Show that the adjoint operation is anti-linear,

$$\left(\sum_i a_i A_i\right)^\dagger = \sum_i a_i^* A_i^\dagger. \tag{2.40}$$

Solution: Like before, we will work directly with the definition from the inner product, making use of conjugate-linearity in the first argument (see Exercise 2.6). For any two ϕ, ψ,

$$\begin{aligned}
((\sum_i a_i A_i)^\dagger \phi, \psi) &= (\phi, \sum_i a_i A_i \psi) \\
&= \sum_i a_i (\phi, A_i \psi) \\
&= \sum_i a_i (A_i^\dagger \phi, \psi) \\
&= ((\sum_i a_i^* A_i^\dagger)\phi, \psi)
\end{aligned} \tag{2.41}$$

Comparing the beginning to end, since ϕ and ψ were completely arbitrary, we must have the operators are equal. This gives our result.

Exercise 2.15

Show that $(A^\dagger)^\dagger = A$.

Solution: Once again, from the definition! Choosing arbitrary ϕ, ψ,

$$(\phi, A\psi) = (A^\dagger \phi, \psi) = (\psi, A^\dagger \phi)^* = ((A^\dagger)^\dagger \psi, \phi)^* = (\phi, (A^\dagger)^\dagger \psi). \tag{2.42}$$

Thus, comparing the first and last of this string of equations, we see that A and $(A^\dagger)^\dagger$ are necessarily equal.

Exercise 2.16

Show that any projector P satisfies $P^2 = P$.

Solution: In some treatments, this is given the definition of a projector! Let's prove it from the textbook's definition, using equation (2.35) from the text.

$$\begin{aligned}
P^2 = PP &= \left(\sum_i |i\rangle \langle i|\right)\left(\sum_j |j\rangle \langle j|\right) \\
&= \sum_{i,j} |i\rangle \langle i|j\rangle \langle j| \\
&= \sum_{i,j} |i\rangle \langle j| \, \delta_{i,j} \\
&= \sum_i |i\rangle \langle i| = P.
\end{aligned} \tag{2.43}$$

Note we took care to use different indices for each copy of P, and exploited the orthonormality of the $|i\rangle$ via $\langle i|j\rangle = \delta_{i,j}$.

Exercise 2.17

Show that a normal matrix is Hermitian if and only if it has real eigenvalues.

Solution: ($\implies$) Suppose H is Hermitian, and let v be an eigenvector of H with eigenvalue λ. For simplicity, take v to be normalized. Then,

$$(v, Hv) = (v, \lambda v) = \lambda(v, v) = \lambda. \tag{2.44}$$

On the other hand, we can take the adjoint of H, which is just H itself, to compute the same quantity.

$$(v, Hv) = (H^\dagger v, v) = (Hv, v) = (\lambda v, v) = \lambda^*(v, v) = \lambda^* \tag{2.45}$$

where we've used conjugate-linearity in the first argument. Taken together these equations imply $\lambda = \lambda^*$. Hence, λ is real.

($\impliedby$) Suppose H is normal with real eigenvalues. By the spectral theorem, H has a diagonal representation $H = \sum_i \lambda_i |v_i\rangle \langle v_i|$ where $|v_i\rangle$ are orthonormal and λ_i real. Taking the Hermitian conjugate,

$$H^\dagger = \left(\sum_i \lambda_i |v_i\rangle \langle v_i| \right)^\dagger = \sum_i \lambda_i^*(|v_i\rangle \langle v_i|)^\dagger = \sum_i \lambda_i |v_i\rangle \langle v_i| = H. \tag{2.46}$$

Hence, H is equal to its own adjoint, hence it is Hermitian.

Exercise 2.18

Show that all eigenvalues of a unitary matrix have modulus 1, that is, can be written in the form $e^{i\theta}$ for some real θ.

Solution: Let v be an (normalized) eigenvector of unitary U with eigenvalue λ. Since unitary operators preserve the norm, we must have

$$(Uv, Uv) = (\lambda v, \lambda v) = \lambda^* \lambda = |\lambda|^2 = 1. \tag{2.47}$$

Every complex number λ has a polar representation $re^{i\theta}$ for $r \geq 0$ and real θ. Moreover, $|\lambda|^2 = r^2 = 1$, so $r = 1$. This proves our result: $\lambda = e^{i\theta}$ for some real θ.

Exercise 2.19 (Pauli matrices: Hermitian and unitary)

Show that the Pauli matrices are Hermitian and unitary.

Solution: There are many ways to show this. In previous exercises we've shown that the Pauli matrices have eigenvalues ± 1 with orthogonal eigenvectors. By Exercise 2.17, this means the Paulis are hermitian. Moreover, each Pauli squares to unity: $\sigma_i^2 = I$. This means the Paulis are their own inverse. Thus

$$\sigma_i^\dagger \sigma_i = \sigma_i \sigma_i = \sigma_i^2 = I. \tag{2.48}$$

Hence, the Paulis are also unitary.

As an aside operators which are both Hermitian and unitary are *reflections*: they necessarily have eigenvalues ± 1 about orthogonal axes.

Exercise 2.20 (Basis change)

Suppose A' and A'' are matrix representations of an operator A on a vector space V with respect to two different orthonormal bases, $|v_i\rangle$ and $|w_i\rangle$. Then the elements of A' and A'' are $A'_{ij} = \langle v_i| A |v_j\rangle$ and $A''_{ij} = \langle w_i| A'' |w_j\rangle$. Characterize the relationship between A' and A''.

Solution: Consider the matrix elements of A'. We would like to express them in terms of the elements of A''. One way to do this is to suitably insert two copies of the identity matrix into the matrix element expression, with the copies represented in the basis of A''.

$$
\begin{aligned}
A'_{ij} = \langle v_i| A |v_j\rangle &= \langle v_i| \left(\sum_k |w_k\rangle \langle w_k| \right) A \left(\sum_l |w_l\rangle \langle w_l| \right) |v_j\rangle \\
&= \sum_{k,l} \langle v_i|w_k\rangle \langle w_k| A |w_l\rangle \langle w_l|v_j\rangle \\
&= \sum_{k,l} A''_{kl} \langle v_i|w_k\rangle \langle w_l|v_j\rangle
\end{aligned}
\tag{2.49}
$$

This gives a direct relationship between the two sets of matrix elements, but might not be so illuminating. However, the double sum along with the format of the inner products suggests some kind of matrix multiplication is occurring. To facilitate this interpretation, let's define the inner products as a matrix M.

$$
M_{ij} := \langle v_i|w_j\rangle
\tag{2.50}
$$

Then we have

$$
\begin{aligned}
A'_{ij} &= \sum_{k,l} A''_{k,l} M_{ik} M^*_{jl} \\
&= \sum_{k,l} M_{ik} A''_{k,l} (M^\dagger)_{lj} \\
&= (M A'' M^\dagger)_{ij}
\end{aligned}
\tag{2.51}
$$

Hence, $A' = M A'' M^\dagger$, where M is a matrix of inner products. Can we deduce any further properties of M? Well, if $A = I$, then all matrix representations must be the identity matrix. This implies $M I M^\dagger = M M^\dagger = I$, so M is unitary.

We can summarize these results by saying A' and A'' are related via a *unitary similarity transformation*. Since A' and A'' were arbitrary matrix representations in terms of orthonormal bases, all such representations are related this way.

Exercise 2.21

Repeat the proof of the spectral decomposition in Box 2.2 for the case when M is Hermitian, simplifying the proof wherever possible.

Solution: I will just point out the places where simplifications can occur assuming M is Hermitian. Once $QMP = 0$ is known, it follows immediately that $PMQ = 0$ by taking the adjoint. It is also immediate

that QMQ is normal, since it is Hermitian! Then the inductive step allows for diagonalizing PMP and QMQ, and the rest of the proof carries through.

Exercise 2.22

Prove that two eigenvectors of a Hermitian operator with different eigenvalues are necessarily orthogonal.

Solution: Let u, v be normalized eigenvectors of a Hermitian H with eigenvalues λ, μ respectively. We've already proven in Exercise 2.17 that λ, μ are real. Hence

$$(u, Hv) = (u, \mu, v) = \mu(u, v) \tag{2.52}$$

and also

$$(u, Hv) = (Hu, v) = (\lambda u, v) = \lambda(u, v). \tag{2.53}$$

Setting these two equal, we see that if $\lambda \neq \mu$, $(u, v) = 0$. Hence, u and v are orthogonal.

Exercise 23

Show that the eigenvalues of a projector P are all either 0 or 1.

Solution: Let v be an eigenvector of P with eigenvalue λ. By Exercise 2.16, $P^2 = P$, so

$$P^2 v = P(Pv) = \lambda^2 v = Pv = \lambda v. \tag{2.54}$$

Hence $\lambda^2 = \lambda$. There are exactly two solutions to this quadratic equation: zero and one.

Exercise 2.24 (Hermiticity of positive operators)

Show that a positive operator is necessarily Hermitian. (Hint: Show that an arbitrary operator A can be written $A = B + iC$ where B and C are Hermitian.)

Solution: There is a clever trick to do the decomposition given in the hint, and it indeed works for any square operator A. Taking inspiration from Euler's formula for complex numbers, we define

$$B := \frac{A + A^\dagger}{2}, \quad C := \frac{A - A^\dagger}{2i}. \tag{2.55}$$

Indeed, we find that $A = B + iC$, and moreover both B and C are Hermitian!

Now let's assume A is a positive operator. Then for any vector ψ, we have $(\psi, A\psi) = (\psi, B\psi) + i(\psi, C\psi) > 0$. In particular, it must be real valued. This must mean that $(\psi, C\psi) = 0$ for every choice of vector ψ. The only way this condition can be met is if $C = 0$. Thus, $A = B$, and since B is Hermitian, so is A.

Exercise 2.25

Show that for any operator A, $A^\dagger A$ is positive.

Solution: We need to show that $(v, A^\dagger A v) \geq 0$ for any vector v. This follows from the definition of the adjoint.

$$(v, A^\dagger A v) = (A v, A v). \tag{2.56}$$

Since the inner product of a vector with itself is always nonnegative, $(A v, A v) \geq 0$ and we've proven our result.

Exercise 2.26

Let $|\psi\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$. Write out $|\psi\rangle^{\otimes 2}$ and $|\psi\rangle^{\otimes 3}$ explicitly, both in terms of tensor products like $|0\rangle |1\rangle$, and using the Kronecker product.

Solution: I will not do all steps because it is a bit tedious. I will just outline some things. Consider $|\psi\rangle^{\otimes n}$. Since the dimension of the Hilbert space is the product of the individual dimensions, the full space is of dimension 2^n (hence dimension four and eight for $n = 2, 3$). Moreover, the overall normalization factor will just be a product of all the $1/\sqrt{2}$, so it is $1/2^{n/2}$. Now we can just focus on the unnormalized $|0\rangle + |1\rangle$ and their tensor product.

I find it easier to work with the tensor product form instead of the Kronecker product, since we don't need to worry about "stacking" everything properly. We have a basis in terms of all possible bitstrings of length n. When we take the n-fold tensor product of $|0\rangle + |1\rangle$, we will find that we obtain a uniform superposition of all length n bitstrings. Try to reason this out yourself ($n = 2$ makes a good example case). Hence,

$$|\psi\rangle = \frac{1}{2^{n/2}} \sum_j |j\rangle \tag{2.57}$$

where the sum is taken over all bitstrings j of length n. This means, in the Kronecker product column vector, every single entry will be 1, with an overall normalization factor.

Exercise 2.27

Calculate the matrix representation of the tensor products of the Pauli operators (a) X and Z; (b) I and X; (c) X and I. Is the tensor product commutative?

Solution: I will work out part (a) explicitly and just give the answer for the other cases. It's important to remember that the basis we are using is $|00\rangle, |01\rangle, |10\rangle, |11\rangle$, in that order. For larger numbers of qubits, by convention we will always count up in binary to order the basis vectors. As for any matrix representation, the ordering of the basis vectors matters!

With this convention in mind, the usual procedure is to to "insert" the second matrix into each element of the first. So for $X \otimes Z$ we have

$$X \otimes Z = \begin{pmatrix} 0 & Z \\ Z & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}. \tag{2.58}$$

We can (and should) check that this representation matches what we expect from the simpler tensor product representation. For example,

$$\langle 1| \langle 1| X \otimes Z |0\rangle |1\rangle = \langle 1| X |0\rangle \langle 1| Z |1\rangle = -1. \tag{2.59}$$

Thus, a -1 should be located in the 2nd row and 4th column, and indeed there is. Note that 00 corresponds to the 1st row/column. This check is helpful to make sure the right matrix was inserted in the other.

Let's now work out the matrix reps for (b) and (c).

$$
\begin{aligned}
I \otimes X &= \begin{pmatrix} X & 0 \\ 0 & X \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \\
X \otimes I &= \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.
\end{aligned}
\tag{2.60}
$$

From these matrix representations it is evident that the tensor product is not commutative $I \otimes X \neq X \otimes I$. However, this is perhaps more transparent from the tensor product directly. $A \otimes B$ is different from $B \otimes A$ because we care whether A acts on the first or second space. They are two different spaces!

2.1 Exercise 2.28

Show that the transpose, complex conjugation, and adjoint operations distribute over the tensor product,

$$
(A \otimes B)^* = A^* \otimes B^*; (A \otimes B)^T = A^T \otimes B^T; (A \otimes B)^\dagger = A^\dagger \otimes B^\dagger.
\tag{2.61}
$$

Solution: As an initial caution, I'll mention that complex conjugation and transposing are only defined with respect to a *specified basis*. Only the adjoint † is basis independent, which we can see from its definition. So we will assume these operations are with respect to bases $|v_j\rangle$ and $|w_k\rangle$ on the first and second spaces respectively.

Let's consider matrix elements in (the tensor product of) the bases above. We have

$$
\begin{aligned}
\langle v_j | \langle w_k | (A \otimes B)^* |v_{j'}\rangle |w_{k'}\rangle &:= (\langle v_j | \langle w_k | A \otimes B |v_{j'}\rangle |w_{k'}\rangle)^* \\
&= (\langle v_j | A |v_j'\rangle \langle w_k | B |w_k'\rangle)^* \\
&= (\langle v_j | A |v_j'\rangle)^* (\langle w_k | B |w_k'\rangle)^* \\
&:= \langle v_j | A^* |v_j'\rangle \langle w_k | B^* |w_k'\rangle \\
&= \langle v_j | \langle w_k | A^* \otimes B^* |v_{j'}\rangle |w_{k'}\rangle.
\end{aligned}
\tag{2.62}
$$

Looking at the first and last line, we see that the matrix elements of $(A \otimes B)^*$ and $A^* \otimes B^*$ are the same in the basis of interest. Hence, they are also equal as operators in general, by linearity. Let's carry out a similar argument for the transpose.

$$
\begin{aligned}
\langle v_j | \langle w_k | (A \otimes B)^T |v_{j'}\rangle |w_{k'}\rangle &:= \langle v_j' | \langle w_k' | A \otimes B |v_j\rangle |w_k\rangle \\
&= \langle v_j' | A |v_j\rangle \langle w_k' | B |w_k\rangle \\
&:= \langle v_j | A^T |v_j'\rangle \langle w_k | B^T |w_k'\rangle \\
&= \langle v_j | \langle w_k | A^T \otimes B^T |v_{j'}\rangle |w_{k'}\rangle.
\end{aligned}
\tag{2.63}
$$

Again, this must imply that $(A \otimes B)^T = A^T \otimes B^T$ by linearity.

The adjoint operator is just complex conjugation and transposition in sequence: $A^\dagger = A^{T*}$. We've shown both of these operations distribute over the tensor product. Therefore, so must the adjoint.

Exercise 2.29

Show that the tensor product of two unitary operators is unitary.

Solution: Let U and V be operators on the first and second space, respectively, and consider the tensor product $U \otimes V$. We've shown in the previous exercise that $(U \otimes V)^\dagger = U^\dagger \otimes V^\dagger$. Hence,

$$(U \otimes V)(U \otimes V)^\dagger = (U \otimes V)(U^\dagger \otimes V^\dagger) = UU^\dagger \otimes VV^\dagger = I. \tag{2.64}$$

Here we've used the fact that U and V are unitary, and that $I \otimes I = I$ (and being somewhat sloppy by not labelling the identity differently on each space). This shows that $U \otimes V$ is unitary, and therefore the tensor product of any unitaries remains unitary.

Exercise 2.30

Show that the tensor product of two Hermitian operators is Hermitian.

Solution: If H and G are Hermitian on the first and second spaces, we have

$$(H \otimes G)^\dagger = H^\dagger \otimes G^\dagger = H \otimes G. \tag{2.65}$$

Hence, $H \otimes G$ is also Hermitian.

Exercise 2.31

Show that the tensor product of two positive operators is positive.

Solution: Let P and Q be positive operators on the first and second space, respectively. As shown in the previous exercise, $P \otimes Q$ is Hermitian since P and Q are. Moreover, the eigenvalues of $P \otimes Q$ are just the product of the eigenvalues of P and Q (why is this true?). Since the eigenvalues of P and Q are nonnegative, same for $P \otimes Q$. This means $P \otimes Q$ is a positive operator.

Exercise 2.32

Show that the tensor product of two projectors is a projector.

Solution: Let P and Q be projectors and consider $P \otimes Q$. Since P and Q are Hermitian, so is $P \otimes Q$. Moreover,

$$(P \otimes Q)^2 = P^2 \otimes Q^2 = P \otimes Q. \tag{2.66}$$

Altogether, this implies $P \otimes Q$ is a projection operator.

Exercise 2.33

The Hadamard operator on one qubit may be written as

$$H = \frac{1}{\sqrt{2}} \left[(|0\rangle + |1\rangle) \langle 0| + (|0\rangle - |1\rangle) \langle 1| \right]. \tag{2.67}$$

Show explicitly that the Hadamard transform on n qubits, $H^{\otimes n}$, may be written as

$$H^{\otimes n} = \frac{1}{\sqrt{2^n}} \sum_{x,y} (-1)^{x \cdot y} |x\rangle \langle y| . \tag{2.68}$$

Write out an explicit matrix representation for $H^{\otimes 2}$.

Solution: First, let's express H as

$$H = \frac{1}{\sqrt{2}} \sum_{x,y=0}^{1} (-1)^{x \cdot y} |x\rangle \langle y| . \tag{2.69}$$

we see then that the statement holds for $n = 1$. The $x \cdot y$ here is just the multiplication of two bits x, y, which ensures that only the last matrix element is negative. Moving now to the tensor product,

$$H^{\otimes n} = \frac{1}{\sqrt{2^n}} \bigotimes_{j=1}^{n} \sum_{x_j, y_j = 0}^{1} (-1)^{x_j \cdot y_j} |x_j\rangle \langle y_j| \tag{2.70}$$

$$= \frac{1}{\sqrt{2^n}} \sum_{x_1, y_1 = 0}^{1} \sum_{x_2, y_2 = 0}^{1} \cdots \sum_{x_n, y_n = 0}^{1} (-1)^{x_1 y_1 + x_2 y_2 + \cdots + x_n y_n} |x_1 x_2 \ldots x_n\rangle \langle y_1 y_2 \ldots y_n| , \tag{2.71}$$

where we used the distributive property of the tensor product to expand the sums. Identifying x with the length n binary string $x_1 x_2 \ldots x_n$, and similar for y, we see that the many sums in the above equation represent a sum over all possible bitstrings x and y. Moreover, the exponent of -1 is just the bitwise dot product $x \cdot y$. Making these identifications in the above equation,

$$H^{\otimes n} = \frac{1}{\sqrt{2^n}} \sum_{x,y \in \text{bitstrings}} (-1)^{x \cdot y} |x\rangle \langle y| . \tag{2.72}$$

This is the result we sought.

Exercise 2.34

Find the square root and logarithm of the matrix

$$\begin{bmatrix} 4 & 3 \\ 3 & 4 \end{bmatrix} . \tag{2.73}$$

Solution: As described in the text, these functions can be seen as acting on the eigenvalues of the matrix. Let's begin by diagonalizing this matrix. We can take some shortcuts because this matrix has a very simple form: it is Hermitian and real, so it is a real linear combination of I, Z, and X. In fact, we can readily see it is just $4I + 3X$. This has the same eigenstates as X, with eigenvalues $4 \pm 3 = 7, 1$. If $|+\rangle$ and $|-\rangle$ are the eigenstates of X, we can write this matrix as

$$\begin{bmatrix} 4 & 3 \\ 3 & 4 \end{bmatrix} = 7 |+\rangle \langle +| + 1 |-\rangle \langle -| . \tag{2.74}$$

Where the projections are given by

$$|+\rangle \langle +| = \frac{1}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$
$$|-\rangle \langle -| = \frac{1}{2} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \begin{bmatrix} 1 & -1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} . \tag{2.75}$$

For any function f on a single variable, we can evaluate $f(M)$ for this particular matrix as

$$f(7)\,|+\rangle\,\langle+|+f(3)\,|-\rangle\,\langle-| = \frac{1}{2}\begin{bmatrix} f(7)+f(3) & f(7)-f(3) \\ f(7)-f(3) & f(7)+f(3) \end{bmatrix}. \tag{2.76}$$

Using square root and logarithm functions, we obtain

$$\frac{1}{2}\begin{bmatrix} \sqrt{7}+\sqrt{3} & \sqrt{7}-\sqrt{3} \\ \sqrt{7}-\sqrt{3} & \sqrt{7}+\sqrt{3} \end{bmatrix} \tag{2.77}$$

and

$$\frac{1}{2}\begin{bmatrix} \log(21) & \log(7/3) \\ \log(7/3) & \log(21) \end{bmatrix} \tag{2.78}$$

respectively.

Exercise 2.35 (Exponential of the Pauli matrices)

Let $\vec{v}$ be any real, three-dimensional unit vector and θ a real number. Prove that

$$\exp(i\theta\vec{v}\cdot\vec{\sigma}) = \cos\theta I + i\sin\theta\vec{v}\cdot\vec{\sigma}, \tag{2.79}$$

where $\vec{v}\cdot\vec{\sigma} \equiv \sum_{i=1}^{3} v_i\sigma_i$. This exercise is generalized in Problem 2.1 on page 117.

Solution: To understand this, it will be helpful to use the fact that $(\vec{v}\cdot\vec{\sigma})^2 = I$. The essential reason for this is that, just as the individual Pauli matrices square to the identity, any other choice of axis $\vec{v}$ should give this property as well, being essentially equivalent. We can also verify this from the

Exercise 2.36

Show that the Pauli matrices except for I have trace zero.

Solution: We can simply see that the diagonal elements of each matrix add to zero.

$$\begin{aligned} \mathrm{tr}(X) &= 0+0 = 0 \\ \mathrm{tr}(Y) &= 0+0 = 0 \\ \mathrm{tr}(Z) &= 1-1 = 0. \end{aligned} \tag{2.80}$$

There are other ways to understand why the trace is zero, which are insightful. First, suppose we know at least one of the Paulis has trace zero. Then, because all the Paulis are equivalent up to a change of basis, the other two Pauli's have trace zero as well. This can be seen as a consequence of the cyclicity of the trace. For example,

$$\mathrm{tr}(X) = \mathrm{tr}\big(HZH^{-1}\big) = \mathrm{tr}\big(H^{-1}HZ\big) = \mathrm{tr}(Z). \tag{2.81}$$

From the perspective of the Lie algebra (commutation relations), each Pauli is the commutator of the other two. Since the commutator is always trace zero, so is each Pauli.

Exercise 2.37 (Cyclic property of trace)

If A and B are two linear operators show that

$$\mathrm{tr}(AB) = \mathrm{tr}(BA). \tag{2.82}$$

Solution: Let's compute the trace with respect to some basis indexed by i. We have, starting from the definition,

$$
\begin{aligned}
\mathrm{tr}(AB) &= \sum_i (AB)_{ii} \\
&= \sum_i \sum_j A_{ij} B_{ji} \\
&= \sum_j \sum_i B_{ji} A_{ij} \\
&= \sum_j (BA)_{jj} \\
&:= \mathrm{tr}(BA).
\end{aligned}
\tag{2.83}
$$

In the above, we expanded the diagonal elements of AB in terms of elements of A and B, and rearranged the sums.

It's important to realize that, despite how it appears, we can't just *arbitrarily* rearrange terms in the trace, only cyclically. So $\mathrm{tr}(ABC) = \mathrm{tr}(CAB) = \mathrm{tr}(BCA)$, but it is not necessary that $\mathrm{tr}(ABC) = \mathrm{tr}(BAC)$. The result we proved only implies invariance under *cyclic* permutations.

Exercise 2.38 (Linearity of the trace)

If A and B are two linear operators, show that

$$\mathrm{tr}(A + B) = \mathrm{tr}(A) + \mathrm{tr}(B) \tag{2.84}$$

and if z is an arbitrary complex number show that

$$\mathrm{tr}(zA) = z\,\mathrm{tr}(A). \tag{2.85}$$

Solution: Since $(A + B)_{ii} = A_{ii} + B_{ii}$, $\mathrm{tr}(A + B) = \sum_i (A + B)_{ii} = \sum_i A_{ii} + B_{ii} = \mathrm{tr}(A) + \mathrm{tr}(B)$. Moreover,

$$\mathrm{tr}(zA) = \sum_i z A_{ii} = z \sum_i A_{ii} = z\,\mathrm{tr}(A). \tag{2.86}$$

Together, these facts show that the trace is linear.

Exercise 2.39 (The Hilbert-Schmidt inner product on operators)

The set L_V of linear operators on a Hilbert space V is obviously a vector space – the sum of two linear operators is a linear operator, zA is a linear operator if A is a linear operator and z is a complex number, and there is a zero element 0. An important additional result is that the vector space L_V can be given a natural inner product structure, turning it into a Hilbert space.

1. Show that the function $(\cdot, \cdot)$ on $L_V \times L_V$ defined by

$$(A, B) \equiv \operatorname{tr}\left(A^\dagger B\right) \tag{2.87}$$

   is an inner product function. This inner product is known as the *Hilbert-Schmidt* or *trace* inner product

2. If V has dimension d, show that L_V has dimension d^2.

3. Find an orthonormal basis of Hermitian matrices for the Hilbert space L_V.

Solution: We need to show that $(\cdot, \cdot)$ is linear in the second argument. This follows directly from linearity of the trace.

$$(A, c_1 B_1 + c_2 B_2) = \operatorname{tr}\left(A^\dagger (c_1 B_1 + c_2 B_2)\right) = c_1 \operatorname{tr}\left(A^\dagger B_1\right) + c_2 \operatorname{tr}\left(A^\dagger B_2\right) = c_1(A, B_1) + c_2(A, B_2) \tag{2.88}$$

Next we need to show that (A, B) is conjugate-symmetric: $(A, B) = (B, A)^*$. To see this, note that for any operator M, $\operatorname{tr}\left(M^\dagger\right) = \operatorname{tr}(M)^*$. This follows because in a given matrix representation, the trace is just transposition follows by complex conjugation, so each diagonal entry gets a complex conjugate before being added together. Using this fact, we have

$$(A, B) = \operatorname{tr}\left(A^\dagger B\right) = \operatorname{tr}\left((B^\dagger A)^\dagger\right) = \operatorname{tr}\left(B^\dagger A\right)^* = (B, A)^*, \tag{2.89}$$

which shows conjugate symmetry. Finally, we need to show $(A, A) \geq 0$ for all $A \in L_V$, with equality iff $A = 0$. We can see this holds because for any A, $A^\dagger A$ is a positive operator. Hence, $\operatorname{tr}\left(A^\dagger A\right)$, when viewed in the eigenbasis, is the sum over all eigenvalues, all of which are nonnegative. Thus, the trace is also nonnegative. The only way it equals zero is if every eigenvalue of $A^\dagger A$ is zero, which implies $A^\dagger A = 0$. In fact, this directly implies $A = 0$, because if $A^\dagger A = 0$, then for every choice of vectors v, we have $(Av, Av) = 0$, so that $Av = 0$. This completes part 1: we've shown (A, B) as defined is truly an inner product.

To show that L_V is a Hilbert space of dimension d^2, we can think about expanding an operator A into the d^2 entries. Specifically, any operator A can be written as

$$A = \sum_{ij} A_{ij} |i\rangle \langle j| \tag{2.90}$$

for some orthonormal basis $|i\rangle$, where $A_{ij} = \langle i| A |j\rangle$ are the matrix elements. The operators $|i\rangle \langle j|$ form an orthonormal basis. Indeed,

$$(|i\rangle \langle j|, |k\rangle \langle l|) = \operatorname{tr}(|j\rangle \langle i|k\rangle \langle l|) = \langle i|k\rangle \langle j|l\rangle = \delta_{ik}\delta_{jl}. \tag{2.91}$$

So that the inner product is zero unless the two operators are equal, in which case it is one. Hence, we've found a set of d^2 orthonormal operators which span L_V. Hence, L_V is dimension d^2.

Although we have found a basis, we have not found a Hermitian basis, as asked in part (3). The operators of the form $|i\rangle \langle i|$ are Hermitian automatically, but those of the form $|i\rangle \langle j|$ for $i \neq j$ are not. However, as discussed in Exercise 2.24, we can still write $|i\rangle \langle j|$ as $B + iC$ for Hermitian B and C. In this case, we have

$$B^{(ij)} := \frac{|i\rangle \langle j| + |j\rangle \langle i|}{2} \qquad C^{(ij)} := \frac{|i\rangle \langle j| - |j\rangle \langle i|}{2i}. \tag{2.92}$$

Thus, using these operators, we can express the original basis of $|i\rangle \langle j|$ in terms of the $B^{(ij)}$ and $C^{(ij)}$. Since $B^{(ij)}$ and $B^{(ji)}$ are not independent (and similar for $C^{(ij)}$, we should only take $i < j$. Hence, our candidate basis consists of the d operators $|i\rangle \langle i|$, and the $d^2 - d$ operators $B^{(ij)}$ and $C^{(ij)}$ for $i < j$. However, these are not properly normalized: we have $(B^{(ij)}, B^{(ij)}) = 1/2$, and similar for $C^{(ij)}$. Multiplying these operators by $\sqrt{2}$ fixes this.

We need to check that these operators are all orthogonal. This can be checked using the orthonormality of the $|i\rangle \langle j|$ under the Hilbert-Schmidt inner product.

Exercise 2.40 (Commutation relations for the Pauli matrices)

Verify the commutation relations

$$[X, Y] = 2iZ; \quad [Y, Z] = 2iX; \quad [Z, X] = 2iY. \tag{2.93}$$

There is an elegant way of writing this using ϵ_{jkl}, the antisymmetric tensor on three indices, for which $\epsilon_{jkl} = 0$ except for $\epsilon_{123} = \epsilon_{231} = \epsilon_{312} = 1$, and $\epsilon_{321} = \epsilon_{213} = \epsilon_{132} = -1$:

$$[\sigma_j, \sigma_k] = 2i \sum_{l=1}^{3} \epsilon_{jkl} \sigma_l. \tag{2.94}$$

Solution: We could check this with direct matrix multiplication, of course. We won't perform that calculation here. A deeper explanation comes from the algebraic properties of the Pauli operators, which are independent of a specific matrix representation. To start, the Pauli operators *anticommute*, which implies, for instance,

$$[X, Y] := XY - YX = 2XY. \tag{2.95}$$

The product of two (different) Pauli operators is i times the third Pauli, with a plus or minus depending on the cyclic ordering according to XYZ. Thus, for example, $XY = iZ$. These properties show the commutator results, and also show why the antisymmetric tensor ϵ_{jkl} captures these relationships.

Exercise 2.41 (Anti-commutation relations for the Pauli matrices)

Verify the anti-commutation relations

$$\{\sigma_i, \sigma_j\} \tag{2.96}$$

where $i \neq j$ are both chosen from the set $1, 2, 3$. Also verify that $(i = 0, 1, 2, 3)$

$$\sigma_i^2 = I. \tag{2.97}$$

Solution: We won't go through the verification of these properties in this solution set: just compute, say, $XY + YX$ and verify that it equals zero.

As a general comment, at some level the statements we are being asked to show here should be considered as *definitions* for the set of Pauli operators. Any set of Pauli operators must satisfy the anti-commutation relations and square to the identity. Any such set gives a valid set of operators to be used for describing qubit states. This abstraction may seem peculiar, but one benefit is it takes away some of the arbitrariness in the form of the Pauli matrices (e.g. making Z diagonal).

Exercise 2.42

Verify that

$$AB = \frac{[A, B] + \{A, B\}}{2}. \tag{2.98}$$

Solution: In essence, we are breaking AB into its "symmetric and antisymmetric pieces". By this, I mean that AB does not relate nicely to BA, the operator obtained by swapping $A \Leftrightarrow B$. However, it can be written in terms of operators which do have nice swapping properties (commutators and anticommutators). These sorts of decompositions occur in other contexts as well, such as when we decomposed an operator into the sum of a Hermitian part B and an antihermitian part iC.

Having said this, the verification of the equation above is straightforward.

$$
\begin{aligned}
AB &= \frac{AB}{2} + \frac{BA}{2} + \frac{AB}{2} - \frac{BA}{2} \\
&= \frac{\{A, B\}}{2} + \frac{[A, B]}{2} \\
&= \frac{[A, B] + \{A, B\}}{2}.
\end{aligned} \tag{2.99}
$$

Exercise 2.43

Show that for $j, k = 1, 2, 3$,

$$
\sigma_j \sigma_k = \delta_{jk} I + i \sum_{l=1}^{3} \epsilon_{jkl} \sigma_l. \tag{2.100}
$$

Solution: Like several of the previous exercises, this could be computed directly from the matrices. However, we can also reason more from the algebraic properties. Since $\sigma_j^2 = I$, we see that the equation is satisfied since the antisymmetric tensor $\epsilon_{jkl} = 0$ if $j = k$. If $j \neq k$, then $\delta_{jk} = 0$, and we just have to check that the second term equals the product. But this follows from antisymmetry along with the commutation relations of Exercise 2.40.

Exercise 2.44

Suppose $[A, B] = 0$, $\{A, B\} = 0$, and A is invertible. Show that B must be 0.

Solution: From Exercise 2.42, we necessarily have that $AB = 0$. Since A is invertible, we also have $A^{-1}(AB) = IB = B = 0$.

Exercise 2.45

Show that $[A, B]^\dagger = [B^\dagger, A^\dagger]$.

Solution: We use the linearity of $\dagger$ and the way it reverses multiplication order.

$$
[A, B]^\dagger = (AB - BA)^\dagger = (AB)^\dagger - (BA)^\dagger = B^\dagger A^\dagger - A^\dagger B^\dagger = [B^\dagger, A^\dagger]. \tag{2.101}
$$

Exercise 2.46

Show that $[A, B] = -[B, A]$.

Solution: This is the statement that the commutator is antisymmetric, and important property. Showing this is straightforward.

$$[A, B] := AB - BA = -(BA - AB) = -[B, A]. \tag{2.102}$$

Exercise 2.47

Suppose A and B are Hermitian. Show that $i[A, B]$ is Hermitian.

Solution: We need to check that $(i[A, B])^\dagger = i[A, B]$. Using the results of Exercise 2.45,

$$(i[A, B])^\dagger = -i[B^\dagger, A^\dagger]. \tag{2.103}$$

Since B and A are Hermitian, $[B^\dagger, A^\dagger] = [B, A]$. Finally, $[B, A] = -[A, B]$ by Exercise 2.46. This shows our result.

Exercise 2.48

What is the polar decomposition of a positive matrix P? Of a unitary matrix U? Of a Hermitian matrix, H?

Solution: A positive operator P can be written as PI where I is the identity operator. Since I is unitary, this gives the correct decomposition. For unitary operators, we note that I is also positive, and therefore $UI = IU$ give the correct polar decomposition for any unitary U.

Finally, let H be a Hermitian operator. The positive operator part of H is given by $\sqrt{HH^\dagger} = \sqrt{H^2} = |H|$, where $|H|$ has the same eigenvectors as H, but with all eigenvalues taken in absolute value. We know $|H|$ is quite close to H, but we need to flip the sign of some of the eigenvalues as a correction. Thankfully, this can be done by a unitary operation! Let R act on the eigenbasis of H by flipping the sign of vectors with negative eigenvalues, leaving other eigenvectors alone. More succinctly, for eigenvector $|i\rangle$ of H,

$$R|i\rangle = \text{sgn}(\lambda_i)|i\rangle \tag{2.104}$$

where λ_i is the associated eigenvalue. This defines the action of R on the whole space by linearity. Moreover, the eigenvectors of $H, |H|$ and R are the same, so they all commute. Finally, we have

$$H = |H|R = R|H| \tag{2.105}$$

since, for any eigenvector $|i\rangle$,

$$|H|R|i\rangle = |\lambda_i|\,\text{sgn}(\lambda_i)|i\rangle = \lambda_i|i\rangle = H|i\rangle. \tag{2.106}$$

Since the action on the eigenbasis is the same, the operators are the same for all input vectors.

Exercise 2.49

Express the polar decomposition of a normal matrix in the outer product representation.

Solution: Let N be a normal operator, so that by the spectral theorem it may be diagonalized:

$$N = \sum_i \lambda_i \, |i\rangle \, \langle i| \tag{2.107}$$

for orthonormal eigenbasis $\{|i\rangle\}$ and eigenvalues λ_i. The positive operator part $J = K = \sqrt{NN^\dagger}$ is given by

$$\sum_i |\lambda_i| \, |i\rangle \, \langle i| \,. \tag{2.108}$$

Every complex number λ_i has a polar representation $\lambda_i = |\lambda_i|e^{i\theta_i}$ for some phase θ_i. Let U provide these phases

$$U = \sum_i e^{i\theta_i} \, |i\rangle \, \langle i| \,. \tag{2.109}$$

We see that $UJ = JU = N$, completing the polar decomposition. We see that left and right decompositions are the same for normal operators. In hindsight this may make sense: can you see why from the definition of J and K?

Exercise 2.50

Find the left and right polar decompositions of the matrix

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}. \tag{2.110}$$

Solution: Calling this matrix M, we have

$$J^2 = M^\dagger M = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \tag{2.111}$$

and

$$K^2 = MM^\dagger = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}. \tag{2.112}$$

We can take the square root through diagonalization, either by hand or by computer. Once you get the hang of matrix square roots, using a computer becomes quite convenient. The results will be

$$\begin{aligned} J &= \frac{1}{\sqrt{5}} \begin{bmatrix} 3 & -1 \\ -1 & 2 \end{bmatrix} \\ K &= \frac{1}{\sqrt{5}} \begin{bmatrix} 2 & -1 \\ -1 & 3 \end{bmatrix}. \end{aligned} \tag{2.113}$$

Since M is invertible, our expression for the unitary part U will be unique, and given by $U = MJ^{-1} = K^{-1}M$. We can compute U using either formula, and a good check is to see if we get the same answer each way. The result should be

$$U = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}. \tag{2.114}$$

Indeed, this matrix is unitary, and we have $M = UJ = KU$. This completes the polar decomposition.

Exercise 2.51

Verify that the Hadamard gate H is unitary.

Solution: The two columns of H are normalized and orthogonal to each other, can be checked quickly. For example, we see that the columns are orthogonal, since

$$\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \frac{1}{2}(1 \cdot 1 - 1 \cdot 1) = 0. \tag{2.115}$$

Exercise 2.52

Verify that $H^2 = I$.

Solution: Here's an interesting fact: any operator that is Hermitian and unitary satisfies this property. Such operators are sometimes called *reflections* (can you see why?). In the case of H, we showed it was unitary in Exercise 2.51, and it is clearly Hermitian (since it is real and symmetric). Hence,

$$H^2 = HH^\dagger = I. \tag{2.116}$$

Exercise 2.53

What are the eigenvalues and eigenvectors of H?

Solution: Here's a clever trick to find the eigenvalues: since $H^2 = I$ from Exercise 2.52, the eigenvalues are either $+1$ or -1 (why?). There are two eigenvalues, and since $\text{tr}(H) = 0$, their sum is zero. This means the two eigenvalues are $\lambda_\pm = \pm 1$. To find the eigenvalues, which we call v_+ and v_-, we can solve

$$\begin{aligned} Hv_+ &= v_+ \\ Hv_- &= -v_-. \end{aligned} \tag{2.117}$$

and then normalize. This becomes an exercise in solving systems of linear equations. The unnormalized eigenvectors are

$$v_\pm = \begin{bmatrix} 1 \pm \sqrt{2} \\ 1 \end{bmatrix}. \tag{2.118}$$

Normalizing these expressions is a bit messy, though straghtforward.

Exercise 2.54

Suppose A and B are commuting Hermitian operators. Prove that $\exp(A)\exp(B) = \exp(A + B)$. (Hint: Use the results of Section 2.1.9.)

Solution: Showing this can be done using power series, but we take a different approach by using the spectral theorem. If A and B commute, then A, B, and $A + B$ all commute and therefore have a shared eigenbasis $|i\rangle$. By definition of functions of operators, the exponentials $\exp(A)$, $\exp(B)$ and $\exp(A + B)$ all commute as well and have the same shared eigenbasis $|i\rangle$. Let λ_i and η_i represent the ith eigenvalue of A

and B respectively. Then for every eigenvector $|i\rangle$ in the basis, we have that $\exp(A)\exp(B)$ and $\exp(A+B)$ act the same way.

$$\exp(A)\exp(B)|i\rangle = \exp(A)e^{\eta_i}|i\rangle = e^{\lambda_i+\eta_i}|i\rangle = \exp(A+B)|i\rangle. \tag{2.119}$$

By linearity, the two operators act on the entire space the same way, meaning they are equal. This proves our result.

Exercise 2.55

Prove that $U(t_1,t)2)$ defined in Equation (2.91) is unitary.

Solution: The unitary $U(t_1,t_2)$ is the exponential of A, where $A = -iH(t_2-t_1)\hbar$. Since H is Hermitian, it is diagonalizable with real eigenvalues. Let

$$\sum_j E_j|j\rangle\langle j| \tag{2.120}$$

be the diagonal form of H with eigenvalues E_j and eigenstate $|j\rangle$. Then by definition of the exponential, we have

$$U(t_1,t_2) = \sum_j e^{-iE_j(t_2-t_1)}|j\rangle\langle j|. \tag{2.121}$$

This operator is unitary, since all eigenvalues have unit norm and it is diagonalizable. This proves the result.

Exercise 2.56

Use the spectral decomposition to show that $K \equiv -i\log(U)$ is Hermitian for any unitary U, and thus $U = \exp(iK)$ for some Hermitian K.

Solution: In many ways this is the reverse of the previous exercise. Since U is unitary, it may be expressed in diagonal form as

$$U = \sum_j e^{i\theta_j}|j\rangle\langle j| \tag{2.122}$$

for some orthonormal basis $\{|j\rangle\}$ and real θ_j. In fact, we may take $\theta_j \in [0,2\pi)$ by cyclicity of the phase. This will help us uniquely define the logarithm in $K = -i\log(U)$, which is given by

$$\begin{aligned} K &= \sum_j -i\log\left(e^{i\theta_j}\right)|j\rangle\langle j| \\ &= \sum_j \theta_j|j\rangle\langle j|. \end{aligned} \tag{2.123}$$

Observing this expression, we see that $U\exp(iK)$, which makes sense because the logarithm should give the inverse of the exponential. Moreover, K is Hermitian since it has a diagonal decomposition with real eigenvalues. This gives our result.

Notice that K is not unique: what other K' exist such that $U = \exp(iK')$? Can you characterize all of them? (Hint: it's easiest to work first with one-dimensional matrices, i.e. numbers.)

Exercise 2.57 (Cascading measurements are single measurements)

Suppose $\{L_l\}$ and $\{M_m\}$ are two sets of measurement operators. Show that a measurement defined by the measurement operators $\{L_l\}$ followed by a measurement defined by the measurement operators $\{M_m\}$ is physically equivalent to a single measurement defined by measurement operators $\{N_{lm}\}$ with the representation $N_{lm} \equiv M_m L_l$.

Solution: Suppose the sequence of measurements yields first the result l, then result m. After measuring l, the state of the system is given by

$$|\psi'\rangle = \frac{L_l |\psi\rangle}{\sqrt{\langle\psi| L_l^\dagger L_l |\psi\rangle}}, \tag{2.124}$$

where $|\psi\rangle$ is the initial state. Next we measure the result m for state $|\psi'\rangle$. The resulting state is given by

$$
\begin{aligned}
|\psi''\rangle &= \frac{M_m |\psi'\rangle}{\sqrt{\langle\psi'| M_m^\dagger M_m |\psi'\rangle}} \\[2mm]
&= \frac{M_m L_l |\psi\rangle}{\sqrt{\langle\psi| L_l^\dagger M_m^\dagger M_m L_l |\psi\rangle}} \frac{\sqrt{\langle\psi| L_l^\dagger L_l |\psi\rangle}}{\sqrt{\langle\psi| L_l^\dagger L_l |\psi\rangle}} \\[2mm]
&= \frac{N_{lm} |\psi\rangle}{\sqrt{\langle\psi| N_{lm}^\dagger N_{lm} |\psi\rangle}}
\end{aligned}
\tag{2.125}
$$

where $N_{lm} \equiv M_m L_l$. Hence, we see that the final state is the same as if the single measurement represented by N_{lm} was applied.

Let's check that the operators $\{N_{lm}\}$ gives the same probability as the cascading measurements. The probability of measuring both l and m is given by

$$p(m,l) = p(m|l)p(l), \tag{2.126}$$

where $p(m|l)$ is the probability of measuring m conditioned on having measured l previously. Each probability on the right side can be given in terms of the individual meaasurements.

$$p(l) = \langle\psi| L_l^\dagger L_l |\psi\rangle$$

$$p(m|l) = \langle\psi'| M_m^\dagger M_m |\psi'\rangle = \frac{\langle\psi| L_l^\dagger M_m^\dagger M_m L_l |\psi\rangle}{\langle\psi| L_l^\dagger L_l |\psi\rangle} \tag{2.127}$$

Multiplying the above expressions yields a nice cancellation, giving

$$p(m,l) = \langle\psi| N_{lm}^\dagger N_{lm} |\psi\rangle \tag{2.128}$$

as desired. Finally, let's check the completeness relation for the $\{N_{lm}\}$.

$$\sum_{l,m} N_{lm}^\dagger N_{lm} = \sum_{l,m} L_l^\dagger M_m^\dagger M_m L_l = \sum_l L_l^\dagger \Big(\sum_m M_m^\dagger M_m\Big) L_l = I \sum_l L_l^\dagger L_l = I. \tag{2.129}$$

This shows that the sequence of measurements gives rise to a description in terms of an effective single measurement described by operators N_{lm}.

Exercise 2.58

Suppose we prepare a quantum system in an eigenstate $|\psi\rangle$ of some observable M , with corresponding eigenvalue m. What is the average observed value of M, and the standard deviation?

Solution: The average observed value, or expectation value, is given by

$$\langle M \rangle = \langle \psi | M | \psi \rangle = \langle \psi | m | \psi \rangle = m. \tag{2.130}$$

Similarly,

$$\langle M^2 \rangle = \langle \psi | M^2 | \psi \rangle = m^2 \tag{2.131}$$

hence,

$$\Delta(M)^2 = \langle M^2 \rangle - \langle M \rangle^2 = m^2 - m^2 = 0. \tag{2.132}$$

Since the standard deviation is zero, every single measurement of M yields value m. This is consistent with the interpretation of $|\psi\rangle$ "having value m." Thus, when an eigenstate of an observable is measured with respect to that observable, the answer is certain, and the usual probabilistic aspects of quantum mechanics aren't present.

Exercise 2.59

Suppose we have qubit in the state $|0\rangle$, and we measure the observable X. What is the average value of X? What is the standard deviation of X?

Solution: The expectation value $\langle X \rangle$ is given by

$$\langle X \rangle = \langle 0 | X | 0 \rangle = \langle 0 | 1 \rangle = 0. \tag{2.133}$$

while the standard deviation $\Delta(X)$ is given by

$$\begin{aligned} \Delta(X)^2 := \langle X^2 \rangle - \langle X \rangle^2 &= \langle X^2 \rangle \\ &= \langle I \rangle \\ &= 1. \end{aligned} \tag{2.134}$$

In the above we used the nice fact that $X^2 = I$.

Exercise 2.60

Show that $\vec{v} \cdot \vec{\sigma}$ has eigenvalues ± 1, and that the projectors onto the corresponding eigenspaces are given by $P_\pm = (I \pm \vec{v} \cdot \vec{\sigma})/2$.

Solution: Rather than diagonalizing $\vec{v} \cdot \vec{\sigma}$ directly, let's take a slightly different approach. Let λ_1 and λ_2 denote the eigenvalues (we know there are at most two distinct values since $\vec{v} \cdot \vec{\sigma}$ is two-dimensional). Since $\text{tr}(\vec{v} \cdot \vec{\sigma}) = 0$ we must have $\lambda_1 + \lambda_2 = 0$, or $\lambda_1 = -\lambda_2$. Moreover, using the algebraic properties of the Pauli operators, which will be explored in greater detail in chapter 4 (see Exercise 4.5), we can show that $(\vec{v} \cdot \vec{\sigma})^2 = I$. This implies $\lambda_1^2 = \lambda_2^2 = 1$. Hence the eigenvalues must be ± 1.

Now let's move on to check that P_+ and P_- are indeed orthogonal projectors. The $P_\pm$ are Hermitian (can you see how?). Next, we need to show that $P_+^2 = P_+$, $P_-^2 = P_-$, and $P_+ P_- = 0$. Indeed, computing the square directly gives

$$P_\pm^2 = \left(\frac{I \pm \vec{v} \cdot \vec{\sigma}}{2} \right)^2 = \frac{I^2 + (\vec{v} \cdot \vec{\sigma})^2 \pm 2\vec{v} \cdot \vec{\sigma}}{4} = \frac{2I \pm 2\vec{v} \cdot \vec{\sigma}}{4} = P_\pm, \tag{2.135}$$

where we again used the fact that $(\vec{v} \cdot \vec{\sigma})^2 = I$. As for orthogonality, we have

$$P_+ P_- = \left(\frac{I + \vec{v} \cdot \vec{\sigma}}{2}\right)\left(\frac{I - \vec{v} \cdot \vec{\sigma}}{2}\right) = \frac{I^2 - I^2 + \vec{v} \cdot \vec{\sigma} - \vec{v} \cdot \vec{\sigma}}{4} = 0. \tag{2.136}$$

Why don't we need to check $P_- P_+ = 0$ independently? This shows that P_+ and P_- are orthogonal projectors. Moreover, $\vec{v} \cdot \vec{\sigma} = P_+ - P_-$, which gives the proper relation between the observable, the eigenvalues, and the projectors.

Exercise 2.61

Calculate the probability of obtaining the result $+1$ for a measurement of $\vec{v} \cdot \vec{\sigma}$, given that the state prior to measurement is $|0\rangle$. What is the state of the system after the measurement if $+1$ is obtained?

Solution: To calculate the probability we can use equation (2.103) from the text, which is derived from the more general rule (2.92) for obtaining measurement probabilities. In our case, we are looking for

$$p(+1) = \langle 0| P_+ |0\rangle = \frac{1}{2} + \langle 0| \vec{v} \cdot \vec{\sigma} |0\rangle /2. \tag{2.137}$$

We showed in Exercise 2.59 that $\langle X \rangle = 0$ for the $|0\rangle$ state. Similarly, $\langle Y \rangle = 0$ for much the same reason. On the other hand, since $|0\rangle$ is an eigenstate of Z, $\langle Z \rangle = +1$. Putting these altogether, $\langle 0| \vec{v} \cdot \vec{\sigma} |0\rangle = v_z$, so that

$$p(+1) = \frac{1 + v_z}{2}. \tag{2.138}$$

For the special cases $v_z = \pm 1$, this answer corresponds with the expected probabilities (can you check this?).

After the measurement, the state is projected onto the P_+ subspace, and normalized.

$$P_+ |0\rangle = |0\rangle /2 + \frac{1}{2}\left(v_x |1\rangle + v_y i |1\rangle + v_z |0\rangle\right)$$
$$= \frac{1}{2}\left((1 + v_z) |0\rangle + (v_x + iv_y) |1\rangle\right). \tag{2.139}$$

The norm of this projection is given by

$$\|P_+ |0\rangle\|^2 = \frac{1}{4}((1 + v_z)^2 + v_x^2 + v_y^2) = \frac{1}{2}(1 + v_z^2) \tag{2.140}$$

where we used the fact that $v_x^2 + v_y^2 = v_z^2 = 1$. Hence, the state $|\psi\rangle$ after measurement is given by

$$|\psi\rangle = \frac{P_+ |0\rangle}{\|P_+ |0\rangle\|} = \frac{1}{\sqrt{2(1 + v_z^2)}}((1 + v_z) |0\rangle + (v_x + iv_y) |1\rangle). \tag{2.141}$$

As a check, we see that when $v_z = 1$ (so that $v_x = v_y = 0$) this corresponds to the expected projection along the $|0\rangle$ state. Other projections along each of the pauli axes give the expected answer as well.

Exercise 2.62

Show that any measurement where the measurement operators and the POVM elements coincide is a projective measurement.

Solution: Suppose the POVM element E equals the corresponding measurement operator M. Combining this with the usual relationship, we have

$$M = M^\dagger M. \tag{2.142}$$

Since E is a positive operator, so is M, hence $M^\dagger = M$. We have that M is a positive operator such that $M^2 = M$. This is one way of *defining* what a projector is. We've shown that each E in the POVM is a projector, hence this is a projective measurement.

Exercise 2.63

Suppose a measurement is described by measurement operators M_m. Show that there exist unitary operators U_m such that $M_m = U_m\sqrt{E_m}$ where E_m is the POVM associated to the measurement.

Solution: The key is to realize that, by the polar decomposition, we can express any operator M_m as $U_m P_m$ for some unitary U_m and positive operator P_m. We can further constrain P_m by using the fact that

$$E_m = M_m^\dagger M_m = P_m^\dagger U_m^\dagger U_m P_m = P_m^2. \tag{2.143}$$

Since $P_m^2 = E_m$ we have $P_m = \sqrt{E_m}$ (careful, we were only able to define the square root nicely since the operators involved are positive). This gives us the desired formula.

Exercise 2.64

Suppose Bob is given a quantum state chosen from a set $|\psi_1\rangle, \ldots, |\psi_m\rangle$ of linearly independent states. Construct a POVM $\{E_1, E_2, \ldots, E_{m+1}\}$ such that if outcome E_i occurs, $1 \leq i \leq m$, then Bob knows with certainty that he was given the state $|\psi_i\rangle$. (The POVM must be such that $\langle\psi_i| E_i |\psi_i\rangle > 0$ for each i.

Solution: In order to ensure that outcome E_i guarantees state $|\psi_i\rangle$, we need to make the probability of every other possibility p_j zero for $j \neq i$. That is, the operators E_j must satisfy

$$\langle\psi_i| E_j |\psi_i\rangle = 0 \quad \text{for each } i \neq j. \tag{2.144}$$

In other words, E_j must be "orthogonal" to every vector $|\psi_i\rangle$ except when $i = j$. We can satisfy this by making E_j a projector onto a subspace perpendicular to all but the vector $|\psi_i\rangle$.

Let's build up this idea mathematically. Since the collection $\{|\psi_i\rangle\}_{i=1}^m$ are linearly independent, the underlying Hilbert space is dimension at least m, and for simplicity we can consider the m-dimensional subspace $\mathcal{V}$ spanned by the $|\psi_i\rangle$. Let $\mathcal{V}_i$ be the space spanned by all the $|\psi_j\rangle$ except for one special $j = i$. This subspace is $(m-1)$-dimensional, so the orthogonal complement is one-dimensional. Let $|\phi_i\rangle\langle\phi_i|$ be the unique projector onto this orthogonal subspace. By construction, we have

$$\langle\psi_j|\phi_i\rangle\langle\phi_i|\psi_j\rangle = 0 \quad (i \neq j). \tag{2.145}$$

This shows that the projector satisfies the requirements set by equation (2.144). In fact, it is almost the unique positive operator that does so. Hence, our only good candidate for E_i is of the form

$$E_i = \gamma_i |\phi_i\rangle\langle\phi_i| \quad (1 \leq i \leq m) \tag{2.146}$$

for some $\gamma_i > 0$. Note that $|\phi_i\rangle$ must have nonzero overlap with $|\psi_i\rangle$ (why?) so $p_i = \langle\psi_i| E_i |\psi_i\rangle > 0$.

Our positive operators $\{E_i\}$ are defined for $1 \leq i \leq m$, but we haven't defined E_{m+1}. By the POVM condition, we actually have no freedom in this matter.

$$E_{m+1} = I - \sum_{i=1}^{m} \gamma_i \, |\phi_i\rangle \, \langle\phi_i| \tag{2.147}$$

This defines E_{m+1}, however, we need to make sure $E_{m+1} \geq 0$ (i.e. it's a positive operator). This is satisfied if the smallest eigenvalue of E_{m+1} is nonnegative, or equivalently, if the *largest* eigenvalue of $\sum_{i=1}^{m} E_i$ is no greater than one. The largest eigenvalue of an operator A (in absolute value, a distinction which doesn't matter here since all eigenvalues are positive) is given by the *spectral norm* $\|A\|$. All norms satisfy the triangle inequality, hence

$$\left\| \sum_{i=1}^{m} \gamma_i \, |\phi_i\rangle \, \langle\phi_i| \right\| \leq \sum_{i=1}^{m} |\gamma_i| \, \| |\phi_i\rangle \, \langle\phi_i| \| = \sum_{i=1}^{m} \gamma_i. \tag{2.148}$$

We want the left hand side of the above to be bounded by 1. Thankfully, we can choose γ_i appropriately to do this. One such choice is

$$\gamma_i = 1/m \tag{2.149}$$

for every i. With this choice $E_{m+1} \geq 0$, and all operators involved are positive and sum to the identity. We have constructed the POVM with the desired properties.

How does this POVM look when the states $|\psi_i\rangle$ are orthogonal? When they are nearly parallel? Could any improvement be made?

Exercise 2.65

Express the states $(|0\rangle + |1\rangle)/\sqrt{2}$ and $(|0\rangle - |1\rangle)/\sqrt{2}$ in a basis in which they are *not* the same up to a relative phase shift.

Solution: As a first step, let's just notice these two states are orthonormal! Check this if you haven't before; you may even recognize these states already. Taken together, and labelling them $|+\rangle$ and $|-\rangle$ respectvely, they form a basis. Moreover, they aren't related by a relative phase, which becomes more obvious when we express things as

$$\begin{aligned} |+\rangle &= 1 \, |+\rangle + 0 \, |-\rangle \\ |-\rangle &= 0 \, |+\rangle + 1 \, |-\rangle . \end{aligned} \tag{2.150}$$

The coefficients in this basis do not even have the same magnitude. This example shows that two states which "differ by a relative phase" can really be quite different: even orthogonal!

Exercise 2.66

Show that the average value of the observable $X_1 Z_2$ for a two qubit system measured in the state $(|00\rangle + |11\rangle)/\sqrt{2}$ is zero.

Solution: As usual, we compute averages via the usual "sandwich" formula. Letting $|\psi\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$, we have

$$\langle X_1 Z_2 \rangle = \langle\psi| \, X_1 Z_2 \, |\psi\rangle = \frac{1}{2} \big(\langle 00| X_1 Z_2 |00\rangle + \langle 00| X_1 Z_2 |11\rangle + \langle 11| X_1 Z_2 |00\rangle + \langle 11| X_1 Z_2 |11\rangle \big). \tag{2.151}$$

We can show that each of these four terms in the right side of (2.151) is zero. We'll show this for the first two terms. Using the standard rules for tensor products, we have

$$\langle 00| X_1 Z_2 |00\rangle = (\langle 0| X |0\rangle) \times (\langle 0| Z |0\rangle) = 0 \times 1 = 0. \tag{2.152}$$

Essentially, the X_1 flips the first qubit, so $|00\rangle$ has no overlap with the result of the flip. Similarly, the second term can be computed as

$$\langle 00| X_1 Z_2 |11\rangle = (\langle 0| X |1\rangle) \times (\langle 0| Z |1\rangle) = 1 \times 0 = 0. \tag{2.153}$$

Hopefully you can see the trend: we'll find that the other terms vanish too. Hence, the expectation value is indeed zero.

Exercise 2.67

Suppose V is a Hilbert space with a subspace W. Suppose $U : W \to V$ is a linear operator which preserves inner products, that is, for any $|w_1\rangle, |w_2\rangle$ in W,

$$\langle w_1| U^\dagger U |w_2\rangle = \langle w_1|w_2\rangle. \tag{2.154}$$

Prove that there exists a unitary operator $U' : V \to V$ which *extends* U. That is $U' |w\rangle = U |w\rangle$ for all $|w\rangle$ in W, but U' is defined on the entire space V. Usually we omit the prime symbol and just write U to denote the extension.

Solution: Let $U(W) \subset V$ be the image of W under U. By restricting the codomain of U from V to its image $U(W)$, U becomes a unitary map. Try to convince yourself of this.

The picture we have is that U maps a subspace ("hyperplane") to another hyperplane in a way that preserves the lengths and angles between vectors. To extend the map to U', we need to define U on the "other piece" of V which is not in W. This notion is captured by the *orthogonal complement* $W^\perp$ of W, which is the set of vectors orthogonal to every vector in W. We will map these to the orthogonal complement $U(W)^\perp$ of the output space. There are many ways to do this, so the extension is by no means unique.

One way to go about this is to define U' on the orthogonal complement via some basis. We will take a slightly more general approach. We may write V as a *direct sum* $W \oplus W^\perp$, meaning that every vector $v \in V$ can be uniquely written as $w + w^\perp$, where $w \in W$ and $w^\perp \in W^\perp$. Let $U^\perp : W^\perp \to U(W)^\perp$ be a unitary map (i.e. it preserves inner product and relates two spaces of the same dimension). We define U' on a general element of V in a natural way.

$$U'v = U'(w + w^\perp) = Uw + U^\perp w^\perp \tag{2.155}$$

That is, U' maps each piece of V independently. Let's check that it satisfies the properties we need. First off, if $v \in W$, then $w^\perp = 0$, so we can see that $U'v = Uv$ as desired. We also need to check that U' is unitary. I'll leave that to you (remember, $U(W)$ and $U(W)^\perp$ are orthogonal by construction!)

One of the benefits of our abstract approach is that we've also classified *all* such extensions of U. The freedom to choose extension is completely determined by the choice of map $U^\perp$.

Exercise 2.68

Prove that $|\psi\rangle \neq |a\rangle |b\rangle$ for all single qubit states $|a\rangle$ and $|b\rangle$.

Solution: One way to see that $|\psi\rangle$ cannot factorize as a simple tensor product is that the measurement statistics on each qubit would be uncorrelated. Yet clearly, the result of a $0, 1$ measurement on one qubit affects the result on the other. Our proof will make these ideas mathematically precise through the use of projections onto $0, 1$ states.

Suppose, for sake of contradiction, that $|\psi\rangle = |a\rangle |b\rangle$. Then any projections (measurements) onto a single qubit must also be equal. In particular, a projection of $|\psi\rangle$ onto the zero state of the first qubit must yield

$$(\langle 0| \otimes I) |\psi\rangle = \frac{1}{\sqrt{2}} |0\rangle \tag{2.156}$$

from the definition of $|\psi\rangle$ (it's implicitly understood that the remaining state is on the second qubit; there is no ambiguity here). On the other hand, performing the same projection on $|a\rangle |b\rangle$ gives $\langle 0|a\rangle |b\rangle$. Equating the two gives two insights.

1. $\langle 0|a\rangle \neq 0$ (otherwise we could not match (2.156))

2. $|b\rangle \propto |0\rangle$

The second bullet is most important. Since $|b\rangle$ is normalized, the proportionality constant must be a complex phase, which we can readily absorb into the definition of $|a\rangle$. Hence, we see that we can write $|\psi\rangle$ as $|a\rangle |0\rangle$.

We can perform an identical argument for the other qubit, projecting onto $I \otimes \langle 0|$, and we would reach an analogous conclusion: $|\psi\rangle = |0\rangle |b\rangle$ for some b. Putting these two together, we have that $|\psi\rangle = e^{i\varphi} |00\rangle$. But we've clearly reached a contradiction.

$$\frac{|00\rangle + |11\rangle}{\sqrt{2}} \neq e^{i\varphi} |00\rangle \tag{2.157}$$

We conclude that $|\psi\rangle$ does not have a factorization $|a\rangle |b\rangle$. We say that $|\psi\rangle$ is not *separable*, hence it is *entangled*.

Exercise 2.69

Verify that the Bell basis forms an orthonormal basis for the two qubit state space.

Solution: There are four Bell states, which is the same number as the dimension of the space. Hence, if they are linearly independent, we automatically know they span the space and therefore form a basis. Moreover, nonzero orthogonal vectors are automatically linearly independent. Clearly, all four Bell states are nonzero, so we just need to check they are orthogonal. (You may have found several of the sentences above to be more information than necessary. If you thought to go ahead and start computing inner products, that's essentially the conclusion I reached above. I just wanted to give more detail for those who are less familiar with these notions)

We can almost get away without computing an inner product. The first two Bell states are automatically orthogonal to the second two since they have no computational basis states in common. Moreover, if the first two Bell states are orthogonal, then clearly so are the second two, since the mathematical forms are the same. Therefore we will just check the first two are orthogonal.

We have

$$\left(\frac{\langle 00| + \langle 11|}{\sqrt{2}}\right) \left(\frac{|00\rangle - |11\rangle}{\sqrt{2}}\right) = \frac{1}{2} \left(\langle 00|00\rangle + \langle 11|00\rangle - \langle 00|11\rangle - \langle 11|11\rangle\right)$$
$$= \frac{1}{2}(1 - 1) = 0 \tag{2.158}$$

Note that we could have left off the $1/\sqrt{2}$ normalization, as it wouldn't have changed the answer. Coupled with the arguments given above, this is enough to conclude the Bell states are orthogonal. If you're not comfortable with some of the shortcuts, go ahead and compute more inner products yourself!

Exercise 2.70

Suppose E is any positive operator acting on Alice's qubit. Show that $\langle\psi|\,E\otimes I\,|\psi\rangle$ takes the same value when $|\psi\rangle$ is any of the four Bell states. Suppose some malevolent third party ('Eve') intercepts Alice's qubit on the way to Bob in the superdense coding protocol. Can Eve infer anything about which of the four possible bit strings 00, 01, 10, 11 Alice is trying to send? If so, how, or if not, why not?

Solution: As before, let's try to reduce the amount of computation as much as possible before we dive in. As the text describes, we know that any of the Bell states can be achieved by taking the "initial" Bell state $(|00\rangle + |11\rangle)/\sqrt{2}$ and acting on Alice's qubit with X, Z or both ($iY = ZX$). Let $|\psi_{ij}\rangle$ be the ij Bell state for $i, j = 0, 1$ (counting in binary as per the text). We have

$$|\psi_{ij}\rangle = (X^i Z^j \otimes I)\,|\psi_{00}\rangle, \tag{2.159}$$

hence, the expectation value is given for any Bell state by

$$\langle\psi_{ij}|\,E\otimes I\,|\psi_{ij}\rangle = \langle\psi_{00}|\,E_{ij}\otimes I\,|\psi_{00}\rangle \tag{2.160}$$

where

$$E_{ij} := (X^i Z^j)^\dagger E (X^i Z^j) \tag{2.161}$$

is *also* and arbitrary positive operator on Alice's qubit. How do we know it's positive? Having done this pre-conditioning, let's compute the expectation value.

$$\begin{aligned}
\langle\psi_{00}|\,E_{ij}\otimes I\,|\psi_{00}\rangle &= \left(\frac{\langle00| + \langle11|}{\sqrt{2}}\right) E_{ij}\otimes I \left(\frac{|00\rangle + |11\rangle}{\sqrt{2}}\right) \\
&= \frac{1}{2}(\langle0|\,E_{ij}\,|0\rangle + \langle1|\,E_{ij}\,|1\rangle) \\
&= \frac{1}{2}\operatorname{tr}(E_{ij})
\end{aligned} \tag{2.162}$$

We've skipped some steps here, but the important aspect is that we can simplify many of these inner products by exploiting the tensor product, e.g.

$$\langle00|\,E_{ij}\otimes I\,|00\rangle = \langle0|\,E_{ij}\,|0\rangle \times \langle0|\,I\,|0\rangle = \langle0|\,E_{ij}\,|0\rangle. \tag{2.163}$$

In the last line, we just observed that the result took the form of a trace over the computational basis. Observe that, since E_{ij} is conjugated by a unitary, and the trace is cyclic,

$$\operatorname{tr}(E_{ij}) = \operatorname{tr}(E), \tag{2.164}$$

hence independent of ij. This is exactly what we wanted to show: the result is independent of choice of Bell state.

A consequence of this fact is that any local measurement on Alice's qubit is independent of the choice of Bell state. Indeed, E could represent a POVM element for a measurement scheme. Therefore, if a third party (Eve) obtains access to Alice's qubit, she cannot learn which Bell state Alice had chose for encoding her message.

Exercise 2.71 (Criterion to decide if a state is mixed or pure)

Let ρ be a density operator. Show that $\operatorname{tr}(\rho^2) \leq 1$, with equality if and only if ρ is a pure state.

Solution:

Let ρ be a state with diagonal decomposition

$$\rho = \sum_{j=1}^{n} p_j \, |j\rangle \, \langle j| . \tag{2.165}$$

Since ρ is positive with unit trace, the p_j satisfy $0 \leq p_j \leq 1$ and $\sum_j p_j = 1$. On the other hand, we have

$$\operatorname{tr}(\rho^2) = \sum_{j=1}^{n} p_j^2 . \tag{2.166}$$

Since $p_j^2 \leq p_j$ for each j, we have $\operatorname{tr}(\rho^2) \leq \operatorname{tr}(\rho) = 1$. Hence, we have $\operatorname{tr}(\rho^2) \leq 1$.

Under what conditions does $\operatorname{tr}(\rho^2) = 1$? This is certainly true for a pure state, since $\rho^2 = \rho$ for any pure state, hence the trace remains the same. Suppose, conversely, that ρ is a quantum state such that $\operatorname{tr}(\rho^2) = \operatorname{tr}(\rho)$. Using the diagonal decomposition of equation (2.166), we have

$$\operatorname{tr}(\rho^2) = \operatorname{tr}(\rho) \implies \sum_j p_j^2 = \sum_j p_j$$
$$0 = \sum_j p_j(1 - p_j). \tag{2.167}$$

Since $p_j(1 - p_j) \geq 0$ for each j, the equality holds only if $p_j(1 - p_j) = 0$ for each j, so that either $p_j = 0$ or $1 - p_j = 0$ (i.e. $p_j = 1$). By normalization, exactly one of the p_j is one, the rest zero. Going back to the expression for ρ in (2.165), this is equivalent to

$$\rho = |j\rangle \, \langle j| \tag{2.168}$$

for some j. We conclude that $\operatorname{tr}(\rho^2) = \operatorname{tr}(\rho)$ corresponds exactly to ρ being pure.

This result suggests that $\operatorname{tr}(\rho^2)$ is a measure of the *purity* of the state. It ranges from zero to one, taking on the maximum value of one when ρ is pure.

Exercise 2.72 (Bloch sphere for mixed states)

The Bloch sphere picture for pure states of a single qubit was introduced in Section 1.2. This description has an important generalization to mixed states as follows.

1. Show that an arbitrary density matrix for a mixed state qubit may be written as

$$\rho = \frac{I + \vec{r} \cdot \vec{\sigma}}{2}, \tag{2.169}$$

   where $\vec{r}$ is a real three-dimensional vector such that $\|\vec{r}\| \leq 1$. This vector is known as the *Bloch vector* for the state ρ.

2. What is the Bloch vector representation for the state $\rho = I/2$?

3. Show that a state ρ is pure if and only if $r = 1$.

4. Show that for pure states the description of the Bloch vector we have given coincides with that in Section 1.2.

Solution: (1) Let ρ be a density matrix for a qubit, so that it is a 2×2 operator. Any such operator can be written as a linear combination of the Pauli operators and the identity, since they form a basis on 2×2 matrices.

$$\rho = aI + bX + cY + dZ \tag{2.170}$$

What other conditions must be imposed on ρ? Since $\rho^\dagger = \rho$, the coefficients a, b, c, d must be real. Moreover, the trace of ρ must equal one. This places no restrictions on b, c, d since the trace of Paulis is zero, but we must have $a = 1/2$. Finally, we need $\rho \geq 0$, meaning that $\langle \psi | \rho | \psi \rangle$ for any qubit state $|\psi\rangle$. Making use of the Bloch sphere picture for pure states, we can parametrize our $|\psi\rangle$ as $\cos(\theta/2) |0\rangle + e^{i\varphi} \sin(\theta/2) |1\rangle$ for spherical angles θ, φ on the Bloch sphere. The expectation values of each Pauli then become the projection along the corresponding axis (as can be checked).

$$\langle X \rangle = \cos \varphi \sin \theta \qquad \langle Y \rangle = \sin \varphi \sin \theta \qquad \langle Z \rangle = \cos \theta \tag{2.171}$$

In particular, the vector of expectation values $\langle \vec{\sigma} \rangle = (\langle X \rangle, \langle Y \rangle, \langle Z \rangle)$ is an arbitrary unit vector. Hence,

$$\langle \rho \rangle = \langle \psi | \rho | \psi \rangle = \frac{1}{2} + \vec{b} \cdot \langle \vec{\sigma} \rangle, \tag{2.172}$$

where $\vec{b} = (b, c, d)$. Our condition for $\rho \geq 0$ therefore amounts to $\vec{b} \cdot \langle \vec{\sigma} \rangle \geq -1/2$. Since $\langle \vec{\sigma} \rangle$ is an arbitrary unit vector, and this relation must hold for every such vector, a necessary and sufficient condition becomes $\left\| \vec{b} \right\| \leq 1/2$. Reparametrizing as $\vec{r} := 2\vec{b}$, we see that $\vec{r}$ is an arbitrary unit vector. With this, we've established all criteria for ρ to be a density operator. The final parametrization of ρ is as given in the problem statement. Quantum states (mixed or pure) are therefore in exact correspondence with points in a three-dimension unit ball.

(2) Comparing with the general form of ρ from part (1), this state has Bloch vector $\vec{r} = 0$. Thus, $I/2$ lies at the very center of the ball.

(3) To test the purity of ρ, let's use the results from exercise 2.71. In fact, all we really need to check is, under what circumstances does $\rho^2 = \rho$? To do this, it will be helpful to use the fact that $(\vec{r} \cdot \vec{\sigma})^2 = r^2 I$. We have

$$\rho^2 = \frac{1}{4}(I + \vec{r} \cdot \vec{\sigma})^2 = \frac{1}{4}(I^2 + (\vec{r} \cdot \vec{\sigma})^2 + 2\vec{r} \cdot \vec{\sigma})$$

$$= \frac{1}{2} \left(I \frac{1 + r^2}{2} + \vec{r} \cdot \vec{\sigma} \right). \tag{2.173}$$

Comparing with ρ itself, we see that $\rho^2 = \rho$ exactly when $r = 1$. Thus, ρ being a pure state is equivalent to $r = 1$, meaning $\vec{r}$ is a unit vector.

(4) If ρ is a pure state, then there is some state vector $|\psi\rangle$ such that $\rho = |\psi\rangle \langle \psi|$. We've seen in Section 1.2 that such a state vector has a Bloch sphere representation $|\psi\rangle = \cos(\theta/2) |0\rangle + e^{i\varphi} \sin(\theta/2) |1\rangle$. Using this expression, we can write ρ in matrix form as

$$\rho = \begin{pmatrix} \cos^2(\theta/2) & e^{-i\varphi} \sin(\theta/2) \cos(\theta/2) \\ e^{i\varphi} \sin(\theta/2) \cos(\theta/2) & \sin^2(\theta/2). \end{pmatrix} \tag{2.174}$$

Let's expand this in the Pauli basis. Know that the diagonal parts of ρ are a mixture of I and Z, while the off-diagonals are combinations of X and Y. Let's therefore write $\rho = \rho_{\text{diag}} + \rho_{\text{off}}$, and handle each part separately. For the diagonal part, we always have

$$\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} = \frac{a + b}{2} I + \frac{a - b}{2} Z. \tag{2.175}$$

In our case,

$$\frac{a+b}{2} = \frac{\cos^2(\theta/2) + \sin^2(\theta/2)}{2} = \frac{1}{2}$$
$$\frac{a-b}{2} = \frac{\cos^2(\theta/2) - \sin^2(\theta/2)}{2} = \frac{\cos\theta}{2}. \tag{2.176}$$

Hence, $\rho_{\text{diag}} = I/2 + \cos\theta Z/2$. Moving to the off-diagonal part, let's make the simplification $\sin(\theta/2)\cos(\theta/2) = \sin(\theta)/2$. Then,

$$\rho_{\text{off}} = \frac{\sin\theta}{2} \begin{pmatrix} 0 & e^{-i\varphi} \\ e^{i\varphi} & 0 \end{pmatrix} = \frac{\sin\theta}{2}(\cos\varphi X + \sin\varphi Y), \tag{2.177}$$

where we've expanded the remaining matrix along the real and imaginary parts to extract X and Y. Altogether, we therefore have

$$\rho = \frac{I + \cos\varphi\sin\theta X + \sin\varphi\sin\theta Y + \cos\theta Z}{2}$$
$$= \frac{I + \vec{r} \cdot \vec{\sigma}}{2}, \tag{2.178}$$

where $\vec{r} = (\cos\varphi\sin\theta, \sin\varphi\sin\theta, \cos\theta)$ represents the unit Bloch vector corresponding to the pure state. Thus, we see a correspondence between the parametrization of pure state vectors $|\psi\rangle$ as given in Section 1.2, and the parametrization of Bloch vectors for density operators given presently.

Exercise 2.73

Let ρ be a density operator. A *minimal ensemble* for ρ is an ensemble $\{p_i, |\psi_i\rangle\}$ containing a number of elements equal to the rank of ρ. Let $|\psi\rangle$ be any state in the support of ρ. (The *support* of a Hermitian operator A is the vector space spanned by the eigenvectors of A with non-zero eigenvalues.) Show that there is a minimal ensemble for ρ that contains $|\psi\rangle$, and moreover that in any such ensemble $|\psi\rangle$ must appear with probability

$$p = \frac{1}{\langle\psi| \rho^{-1} |\psi\rangle} \tag{2.179}$$

where ρ^{-1} is defined to be the inverse of ρ, when ρ is considered as an operator acting only on the support of ρ. (This definition removes the problem that ρ may not have an inverse.)

Solution: The *rank* of a Hermitian operator is the dimension of the support (also known as the *range*). Let ρ have rank r, and diagonal decomposition

$$\rho = \sum_{i=1}^{r} p_i |i\rangle \langle i| \tag{2.180}$$

where each probability $p_i \in (0, 1]$ is strictly positive. The vectors $|i\rangle$ for $i = 1, \ldots r$ form an orthonormal basis for the support of ρ. From now on, we will forget the ambient vector space completely, restricting ourselves to the support. This implies, for example, that ρ is invertible. Since $|\psi\rangle$ is in the support of ρ, we decompose $|\psi\rangle$ in the eigenbasis of ρ.

$$|\psi\rangle = \sum_{i=1}^{r} c_i |i\rangle \tag{2.181}$$

We wish to know whether there exists a minimal ensemble for ρ containing $|\psi\rangle$. Suppose $\{q_i, |\psi_i\rangle\}$ were such an ensemble, with $i = 1, \ldots, L$ and $|\psi_1\rangle = |\psi\rangle$ being the state of interest. By Theorem 2.6 of the text, such a minimal ensemble exists exactly when there is an $L \times L$ unitary matrix U such that

$$\sqrt{q_i} \, |\psi_i\rangle = \sum_{j=1}^{r} U_{ij} \sqrt{p_j} \, |j\rangle. \tag{2.182}$$

In particular, we require $\sqrt{q_1} \, |\psi_1\rangle = \sum_{j=1}^{r} U_{1j} \sqrt{p_j} \, |j\rangle$. Comparing this with equation (2.181), we can relate the coefficients c_i to the other parameters as follows.

$$c_j = U_{1j} \sqrt{\frac{p_j}{q_1}} \tag{2.183}$$

By unitarity, we require $\sum_j |U_{1j}|^2 = 1$. This implies, via rearranging (2.183), that

$$q_1 \sum_{j=1}^{r} \frac{|c_j|^2}{p_j} = 1, \tag{2.184}$$

i.e. the probability $p = q_1$ of finding the state $|\psi\rangle$ in our ensemble is given by

$$q_1 = \frac{1}{\sum_{j=1}^{r} |c_j|^2 / p_j} = \frac{1}{\langle \psi | \, \rho^{-1} \, | \psi \rangle}. \tag{2.185}$$

With this choice of q_1, the coefficients U_{1j} are determined via equation (2.183).

$$U_{1j} = c_j \sqrt{\frac{q_1}{p_j}} = \frac{c_j / \sqrt{p_j}}{\sum_k |c_k|^2 / p_k} \tag{2.186}$$

The remainder of the elements of U may be chosen freely, subject to constraints of unitarity. In particular, the remaining rows will be normalized vectors, orthogonal to each other and to the first row U_{1j}. We've therefore shown the appropriate unitary can be constructed relating a minimal ensemble containing $|\psi\rangle$ to the diagonal basis of ρ. By Theorem 2.6, this ensemble is a perfectly valid one for ρ. We saw that the probability $p \equiv q_1$ was constrained by unitarity to be as stated in the problem.

Exercise 2.74

Suppose a composite system A and B is in the state $|a\rangle \, |b\rangle$, where $|a\rangle$ is a pure state of system A, and $|b\rangle$ is a pure state of system B. Show that the reduced density operator of system A alone is a pure state.

Solution: If the system is in pure state $|a\rangle \, |b\rangle$, this means the density operator representation of the state is given by

$$\rho^{AB} = |a\rangle \, |b\rangle \, \langle a| \, \langle b| = |a\rangle \, \langle a| \otimes |b\rangle \, \langle b|. \tag{2.187}$$

If this equality is unfamiliar to you, verify it by multiplying both size by an arbitrary product state $|\varphi\rangle \, |\psi\rangle$. (Recall that $|a\rangle \, |b\rangle$ is shorthand for $|a\rangle \otimes |b\rangle$.) Now let's compute the partial trace, using the definition in equation (2.178) of the text.

$$\mathrm{tr}_B(\rho^{AB}) = \mathrm{tr}_B(|a\rangle \, \langle a| \otimes |b\rangle \, \langle b|) = |a\rangle \, \langle a| \, \mathrm{tr}(|b\rangle \, \langle b|) \tag{2.188}$$

Since $|b\rangle \, \langle b|$ is a density operator, it has trace one. Hence, $\rho^A = \mathrm{tr}_B(\rho^{AB}) = |a\rangle \, \langle a|$. This state is a pure state, which we can see because it corresponds with the state vector $|a\rangle$. An explicit check reveals that $\rho^2 = \rho$.

$$(|a\rangle \, \langle a|)^2 = |a\rangle \, \langle a|a\rangle \, \langle a| = |a\rangle \, \langle a|. \tag{2.189}$$

Exercise 2.75

For each of the four Bell states, find the reduced density operator for each qubit.

Solution: This has already been calculated for the Bell state $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ in the text. Rather than performing the calculation for the other three states (which falls under very similar lines), we will discuss some ways you can get the answer using certain properties of the partial trace, as well as the answer for the first Bell state. These properties are valuable tools in general, so it will be more broadly useful to discuss them.

First, since the Bell state is completely symmetric with respect to the two qubits, we should get the same answer regardless of which qubit we trace over (convince yourself of this fact). Hence, we'll consider tracing over the second qubit, exclusively.

Intuitively, the density operator for A shouldn't change when we perform operations solely on B. The reason for this is because ρ^A is supposed to only contain information accessible from system A alone. More mathematically, for any unitary $U_B \equiv I \otimes U$ acting on system B alone, the partial trace obeys the usual cyclical relation, so that

$$\text{tr}_B(U_B \rho^{AB} U_B^\dagger) = \text{tr}_B(\rho^{AB} U_B^\dagger U_B) = \rho^A. \tag{2.190}$$

Thus, any two pure states $|\phi\rangle_{AB}$ and $|\psi\rangle_{AB}$ which are related by a local unitary on B ($|\phi\rangle_{AB} = U_B |\psi\rangle_{AB}$) will give the same density matrix on A.

Now comes the interesting part: all of the Bell states are related by a local unitary transformation on system B. Given one of the Bell states, we can obtain the others by applyin Z, X, or both to one of the qubits. Since these are local operations, the density operator will remain unchanged.

Exercise 2.76

Extend the proof of the Schmidt decomposition to the case where A and B may have state spaces of different dimensionality.

Solution: It's important to note that the singular value decomposition, which is a crucial aspect of the proof, does not depend on the matrix of coefficients a_{jk} being square. More generally, if a is $n \times m$, we have $a = udv$ where u and v are unitary of dimensions n and m, respectively. Here, d is an $n \times m$ "diagonal" matrix, in which the only possible nonzero entries are d_{ii}, which are nonnegative. Hence, just as before

$$|\psi\rangle = \sum_{ijk} u_{ji} d_{ii} v_{ik} |j\rangle |k\rangle \tag{2.191}$$

where it is understood that the indices j and k run from n and m, respectively. The index i runs from the smaller of n and m. With these changes, the proof carries through exactly as before. See if you can trace the steps yourself!

Exercise 2.77

Suppose ABC is a three component quantum system. Show by example that there are quantum states $|\psi\rangle$ of such systems which can not be written in the form

$$|\psi\rangle = \sum_i \lambda_i |i_A\rangle |i_B\rangle |i_C\rangle, \tag{2.192}$$

where λ_i are real numbers, and $|i_A\rangle$, $|i_B\rangle$, $|i_C\rangle$ are orthonormal bases of the respective systems.

Solution: The simplest counterexample we could expect to find is one in which all three systems are qubits, spanned by the orthonormal basis $|0\rangle, |1\rangle$. What could we use as inspiration for finding such an example? Notice that in equation (2.192), not only are the vectors $|i_A\rangle |i_B\rangle |i_C\rangle$ orthonormal, but every subsystem also has orthonormal vectors. Me might try to break this by having orthogonal vectors in which not all "sub vectors" are orthonormal. Here's an example.

$$|\psi\rangle = \frac{1}{\sqrt{2}} \left(|000\rangle + |011\rangle \right) \tag{2.193}$$

Notice that system A only has vector $|0\rangle$ showing, while the other two systems "follow the rule". Let's see if this will work as our counterexample.

Suppose that, somehow, we could write

$$\frac{1}{\sqrt{2}} \left(|000\rangle + |011\rangle \right) = \sqrt{p} \, |u_A u_B u_C\rangle + \sqrt{1-p} \, |v_A v_B v_C\rangle . \tag{2.194}$$

where $\{u_i, v_i\}$ form orthonormal bases for qubits A, B, C. Here, $p \in [0, 1]$. Notice that the coefficients are completely general, for the states are normalized, and if the coefficient is negative, we can absorb it into our definition of one of the basis vectors. In fact, we can say that p is *strictly* between 0 and 1 because by Exercise 2.68, the state on the left is entangled, while for $p = 0, 1$ the state on the right is separable (unentangled).

Let's take the inner product of both sides with $|1_A\rangle$. Since the left side is zero, we get

$$0 = \sqrt{p} \, \langle 1|u_A\rangle \, |u_B u_C\rangle + \sqrt{1-p} \, \langle 1|v_A\rangle \, |v_B v_C\rangle \tag{2.195}$$

The vectors remaining on the right side are orthogonal, hence linearly independent. We conclude that the coefficients are exactly zero. But p is neither 0 or 1. This means

$$\langle 1|u_A\rangle = \langle 1|v_A\rangle = 0. \tag{2.196}$$

This space is two-dimensional, so this implies both $|u_A\rangle$ and $|v_A\rangle$ are proportional to $|0\rangle$. They are certainly not orthogonal, then, contradicting our original assumption.

We conclude that the state $|\psi\rangle$ in equation (2.193) cannot be expressed in this way. This argument can be generalized to higher dimensional spaces as well, if desired.

Exercise 2.78

Prove that a state $|\psi\rangle$ of a composite system AB is a product state if and only if it has Schmidt number 1. Prove that $|\psi\rangle$ is a product state if and only if ρ^A (and thus ρ^B) are pure states.

Solution: Let $r > 0$ be the Schmidt number of $|\psi\rangle$. If $|\psi\rangle$ is a product state, such that $|\psi\rangle = |\phi\rangle \otimes |\chi\rangle$ for states $|\phi\rangle, |\chi\rangle$, then $|\psi\rangle$ is already expressed in a Schmidt decomposition, so $r = 1$. This shows product states have Schmidt number 1. Conversely, if $r = 1$, then $|\psi\rangle = \lambda_1 |1_A\rangle |1_B\rangle$, and by normalization $\lambda_1 = 1$. Hence, $|\psi\rangle$ is a product state. This proves the first statement of the problem.

Now we consider the second statement. Let $\rho^{AB} = |\psi\rangle \langle \psi|$ be the density matrix representing the pure state on the composite system. Now suppose $|\psi\rangle$ is a product state, such that $|\psi\rangle = |\phi\rangle |\chi\rangle$ for states $|\phi\rangle$ and $|\chi\rangle$ on A and B respectively. Then

$$\rho^{AB} = |\phi\rangle \langle \phi| \otimes |\chi\rangle \langle \chi| \tag{2.197}$$

and therefore

$$\rho^A = \text{tr}_B \, \rho^{AB} = |\phi\rangle \langle\phi| \, \text{tr}(|\chi\rangle \langle\chi|) = |\phi\rangle \langle\phi| \, . \tag{2.198}$$

Thus, we see that ρ^A is a pure state. A similar, symmetrical calculation would also show that $\rho^B = |\chi\rangle \langle\chi|$ is also pure. Conversely, suppose ρ^A and ρ^B are both pure states. Let's express the joint state $|\psi\rangle$ as a Schmidt decomposition.

$$|\psi\rangle = \sum_{j=1}^{r} \sqrt{p_j} \, |\phi_j\rangle \, |\chi_j\rangle \tag{2.199}$$

Here, p_j satisfy $p_j \in (0,1]$ and $\sum_j p_j = 1$. The vectors $\{|\phi_j\rangle\}$ and $\{|\chi_j\rangle\}$ are orthonormal on A and B respectively. We can compute ρ^A (and ρ^B) directly from this decomposition. When we do (try it yourself!) we find that

$$\rho^A = \text{tr}_B \, |\psi\rangle \langle\psi| = \sum_{j=1}^{r} p_j \, |\phi_j\rangle \langle\phi_j| \, . \tag{2.200}$$

For this state to be pure, as we are assuming, only one of these terms can be nonzero. Thus, there is some k such that $p_k = 1$, and all other p_j are zero. But if we return to (2.199), this means $|\psi\rangle$ is actually a product state. This is what we sought to prove.

Exercise 2.79

Consider a composite system consisting of two qubits. Find the Schmidt decompositions of the states

$$\frac{|00\rangle + |11\rangle}{\sqrt{2}}; \qquad \frac{|00\rangle + |01\rangle + |10\rangle + |11\rangle}{2}; \quad \text{and} \quad \frac{|00\rangle + |01\rangle + |10\rangle}{\sqrt{3}}. \tag{2.201}$$

Solution: We will solve the first two of these without doing an "algorithmic" Schmidt decomposition. Hopefully, this will provide some intuition that you can use going forward as well. The last of these we will do directly: via a singular value decomposition of the coefficients.

Upon inspection, the first of these states, a Bell state, is already in Schmidt form! To make this more obvious, we could distribute the factor of $1/\sqrt{2}$, and observe that $|0\rangle$ and $|1\rangle$ are orthonormal.

Next, we observe that the second state is in a uniform superposition of all qubit computational basis states. A useful fact to know is, this can be produced by applying Hadamard gates to the all zero state $|0\rangle^{\otimes n}$. The Hadamard acting on $|0\rangle$ produces $(|0\rangle + |1\rangle)/\sqrt{2}$. Hence we have that

$$\frac{|00\rangle + |01\rangle + |10\rangle + |11\rangle}{2} = \left(\frac{|0\rangle + |1\rangle}{\sqrt{2}}\right) \otimes \left(\frac{|0\rangle + |1\rangle}{\sqrt{2}}\right). \tag{2.202}$$

This is the Schmidt decomposition, with Schmidt number 1. Each state is normalized. If you aren't sure, go backwards and make sure this decomposition works. Take heart: standard methods can be used to arrive at this answer, which we will illustrate in for the final example.

Let's look at the final example, and for the moment let's ignore the normalization factor $1/\sqrt{3}$. Let's recognize the "Schmidt decomposition" is just the "singular value decomposition", with the identification

$$|i\rangle \, |j\rangle \Leftrightarrow |i\rangle \langle j| \, . \tag{2.203}$$

The math carries through exactly the same. As a matrix, this looks like

$$a = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}. \tag{2.204}$$

We can compute the singular value decomposition by diagonalizing $a^\dagger a$ and $aa^\dagger$. But our problem here is even easier than the general case, since a is diagonalizable. So let's just diagonalize it (try this yourself first; this is an essential skill!). Here is the result.

$$a = \frac{1+\sqrt{5}}{2}\,|v_1\rangle\langle v_1| + \frac{1-\sqrt{5}}{2}\,|v_2\rangle\langle v_2| \tag{2.205}$$

Here, $|v_1\rangle$ and $|v_2\rangle$ are the orthonormal eigenvectors of a given by

$$|v_1\rangle = \frac{1+\sqrt{5}}{\sqrt{2\left(5+\sqrt{5}\right)}}\,|0\rangle + \sqrt{\frac{2}{5+\sqrt{5}}}\,|1\rangle$$

$$|v_2\rangle = \frac{1-\sqrt{5}}{\sqrt{2\left(5-\sqrt{5}\right)}}\,|0\rangle + \sqrt{\frac{5+\sqrt{5}}{10}}\,|1\rangle\,. \tag{2.206}$$

Going back to the problem at hand, we then have

$$|00\rangle + |01\rangle + |10\rangle = \frac{1+\sqrt{5}}{2}\,|v_1\rangle\,|v_1\rangle + \frac{1-\sqrt{5}}{2}\,|v_2\rangle\,|v_2\rangle\,. \tag{2.207}$$

Reinserting the normalization factor, we have

$$\frac{|00\rangle + |01\rangle + |10\rangle}{\sqrt{3}} = \frac{1+\sqrt{5}}{2\sqrt{3}}\,|v_1\rangle\,|v_1\rangle + \frac{1-\sqrt{5}}{2\sqrt{3}}\,|v_2\rangle\,|v_2\rangle\,. \tag{2.208}$$

We're not done yet. For a Schmidt decomposition, we need all the coefficients to be positive. But the second coefficient is actually negative. This is related to the fact that we diagonalized a instead of taking the singular value decomposition. There is a very simple fix: group a minus sign with one of the $|v_2\rangle$.

$$\frac{|00\rangle + |01\rangle + |10\rangle}{\sqrt{3}} = \frac{1+\sqrt{5}}{2\sqrt{3}}\,|v_1\rangle\,|v_1\rangle + \frac{\sqrt{5}-1}{2\sqrt{3}}\,|v_2\rangle \otimes (-|v_2\rangle)\,. \tag{2.209}$$

This is in the appropriate form. This correction will work in general, assuming we've diagonalized a Hermitian matrix.

Exercise 2.80

Suppose $|\psi\rangle$ and $|\varphi\rangle$ are two pure states of a composite quantum system with components A and B, with identical Schmidt coefficients. Show that there are unitary transformations U on system A and V on system B such taht $|\psi\rangle = U \otimes V |\varphi\rangle$.

Solution: Let $\{\lambda_i\}$ be the shared, nonzero Schmidt coefficients of states $|\psi\rangle$ and $|\varphi\rangle$, where the index i runs from 1 to the Schmidt number r. We can express the Schmidt decomposition of each state as follows.

$$|\varphi\rangle = \sum_{i=1}^{r} \lambda_i\,|u_i\rangle \otimes |u_i'\rangle \qquad |\psi\rangle = \sum_{i=1}^{r} \lambda_i\,|v_i\rangle \otimes |v_i'\rangle \tag{2.210}$$

We stress that each of the collections $\{|u_i\rangle\}$, $\{|u_i'\rangle\}$, $\{|v_i\rangle\}$, and $\{|v_i'\rangle\}$ are orthonormal (, either on A or B. Let U and V be unitary operations on A and B respectively, such that

$$U\,|u_i\rangle = |v_i\rangle\,, \qquad V\,|u_i'\rangle = |v_i'\rangle\,. \tag{2.211}$$

If r is less than the dimension of A and/or B, then U and/or V is not uniquely defined. However, the point is they exist, because U and V map orthonormal vectors to other orthonormal vectors. With these definitions, one can check that $U \otimes V |\varphi\rangle = |\psi\rangle$.

Exercise 2.81 (Freedom in purifications)

Let $|AR_1\rangle$ and $|AR_2\rangle$ be two purifications of a state ρ^A to a composite system AR. Prove that there exists a unitary transformation U_R acting on system R such that $|AR_1\rangle = (I_A \otimes U_R)|AR_2\rangle$.

Solution: Let

$$|AR_1\rangle = \sum_i \sqrt{p_i}\, |u_i\rangle \otimes |u_i'\rangle$$
$$|AR_2\rangle = \sum_i \sqrt{q_i}\, |v_i\rangle \otimes |v_i'\rangle \tag{2.212}$$

be Schmidt decompositions of the two purifications, with $p_i, q_i > 0$. The square roots are for later convenience. The partial traces over R_1 and R_2 must give the same result ρ^A. This implies

$$\sum_i p_i |u_i\rangle \langle u_i| = \sum_i q_i |v_i\rangle \langle v_i| . \tag{2.213}$$

These are both diagonalizations of the same operator ρ^A. Thus, the set of eigenvalues and corresponding eigenspaces must be the same. Also, the rank r of each of these operators is the same, which in turn means the Schmidt rank r of $|AR_1\rangle$ and $|AR_2\rangle$ are the same.

This is all valuable information, but we will need to do some prep work to reap the full benefits. The difficulty stems from the possiblility of *degeneracy*, in which an eigenvalue has multiple independent eigenvectors. To account for this, let $\{\lambda\}$ denote the set of distinct eigenvalues of ρ^A. Then each p_j equals some λ (and similar for q_i). Let d_λ be the degeneracy of λ, meaning the dimension of the eigenspace for λ. Equivalently, d_λ is the rank of the projector P_λ onto the λ eigenspace. We also have that d_λ is the number of indices i such that $p_i = \lambda$ (and similar for q_i).

Having laid this notational groundwork, we can rewrite our two purifications in a more unified manner. First, we can write the reduced state ρ^A, through reindexing, as follows,

$$\rho^A = \sum_\lambda \lambda P_\lambda = \sum_\lambda \lambda \left(\sum_{i=1}^{d_\lambda} |u_{i,\lambda}\rangle \langle u_{i,\lambda}| \right) = \sum_\lambda \lambda \left(\sum_{i=1}^{d_\lambda} |v_{i,\lambda}\rangle \langle v_{i,\lambda}| \right) . \tag{2.214}$$

Here, the notation $|u_{i,\lambda}\rangle$ is a reindexing of the original $|u_i\rangle$, grouping them in terms of their eigenvalue λ. For each λ, let U_λ be a unitary on system A such that $U_\lambda |u_{i,\lambda}\rangle = |v_{i,\lambda}\rangle$, and acting as the identity on any orthogonal vector. Then, we have the following

$$|AR_1\rangle = \sum_\lambda \sqrt{\lambda} \sum_{i=1}^{d_\lambda} |u_{i,\lambda}\rangle \otimes |u_{i,\lambda}'\rangle$$

$$|AR_2\rangle = \sum_\lambda \sqrt{\lambda} \sum_{i=1}^{d_\lambda} |v_{i,\lambda}\rangle \otimes |v_{i,\lambda}'\rangle \tag{2.215}$$

$$= \sum_\lambda \sqrt{\lambda}\, U_\lambda \sum_{i=1}^{d_\lambda} |u_{i,\lambda}\rangle \otimes |v_{i,\lambda}'\rangle$$

In essence, each U_λ in the previous line is acting on a maximally entangled state. We need to show that such states have a spacial "ricochet property", in that acting on system A is effectively like acting on system R instead.

Lemma: Let $|\psi\rangle = \sum_i |i\rangle |i'\rangle$, where $\{|i\rangle\}$ and $\{|i'\rangle\}$ are orthonormal sets. For a unitary U, we have

$$U \otimes I\, |\psi\rangle = I \otimes U^T\, |\psi\rangle \tag{2.216}$$

where U^T is the transpose of U, with the transpose taken with respect to the orthonormal states defined by $|i\rangle$ and $|i'\rangle$.

Proof:

$$
\begin{aligned}
(U \otimes I) |\psi\rangle &= \sum_i U |i\rangle |i'\rangle \\
&= \sum_i \sum_j U_{ji} |j\rangle |i'\rangle \\
&= \sum_j |j\rangle \otimes \sum_i U_{ji} |i'\rangle \\
&= \sum_j |j\rangle \otimes \sum_i (U^T)_{ij} |i'\rangle \\
&= \sum_j |j\rangle \otimes U^T |j'\rangle \\
&= I \otimes U^T |\psi\rangle .
\end{aligned}
\tag{2.217}
$$

This is what we sought to show.

Now let's apply this result to our present case. From the last line of equation (2.215), using the lemma, we have

$$
|AR_2\rangle = \sum_\lambda \sqrt{\lambda} \sum_{i=1}^{d_\lambda} |u_{i,\lambda}\rangle \otimes U_\lambda^T |v'_{i,\lambda}\rangle .
\tag{2.218}
$$

Define U as the direct sum of all of the U_λ. Then we can write simply

$$
|AR_2\rangle = I \otimes U^T \sum_\lambda \sqrt{\lambda} \sum_{i=1}^{d_\lambda} |u_{i,\lambda}\rangle \otimes |v'_{i,\lambda}\rangle .
\tag{2.219}
$$

Finally, let V be a unitary such that $V \left|u'_{i,\lambda}\right\rangle = \left|v'_{i,\lambda}\right\rangle$. We then have simply

$$
I \otimes (U^T V) |AR_1\rangle = |AR_2\rangle .
\tag{2.220}
$$

Thus, the two purifications $|AR_1\rangle$ and $|AR_2\rangle$ are connected by a unitary on the environment.

Exercise 2.82

Suppose $\{p_i, |\psi_i\rangle\}$ is an ensemble of states generating a density matrix $\rho = \sum_i p_i |\psi_i\rangle \langle \psi_i|$ for a quantum system A. Introduce a system R with orthonormal basis $|i\rangle$.

1. Show that $\sum_i \sqrt{p_i} |\psi_i\rangle |i\rangle$ is a purification of ρ.

2. Suppose we measure R in the basis $|i\rangle$, obtaining outcome i. With what probability do we obtain the result i, and what is the corresponding state of system A?

3. Let $|AR\rangle$ be *any* purification of ρ to the sytem AR. Show that there exists an orthonormal basis $|i\rangle$ in which R can be measured such that the corresponding post-measurement state for system A is $|\psi_i\rangle$ with probability p_i.

Solution: (1) All we need to do is compute the partial trace and verify we obtain ρ. The density matrix representation of the purification $|\Psi\rangle$ is given by

$$\rho^{AR} = |\Psi\rangle\langle\Psi| = \sum_i \sum_j \sqrt{p_i p_j}\, |\psi_i\rangle\langle\psi_j| \otimes |i\rangle\langle j|, \tag{2.221}$$

so that

$$\mathrm{tr}_R\, \rho^{AR} = \sum_i \sum_j \sqrt{p_i p_j}\, |\psi_i\rangle\langle\psi_j|\, \mathrm{tr}\, |i\rangle\langle j|. \tag{2.222}$$

Since $\mathrm{tr}\, |i\rangle\langle j| = \langle j|i\rangle = \delta_{ij}$, we have

$$\mathrm{tr}_R\, \rho^{AR} = \sum_i p_i\, |\psi_i\rangle\langle\psi_i| = \rho. \tag{2.223}$$

Hence, $|\Psi\rangle$ is indeed a purification of ρ.

(2) The probability is equal to $\|\langle i|\Psi\rangle\|^2 = \mathrm{tr}\big(\rho^{AR} |i\rangle\langle i|\big)$, which in turn is given by

$$\left\|\sqrt{p_i}\, |\psi_i\rangle\right\|^2 = p_i. \tag{2.224}$$

(3) Let $|AR\rangle$ be a purification of $|\Psi\rangle$ by the results of exercise 2.81, there exists a unitary U_R such that $I \otimes |\Psi\rangle = |AR\rangle$. This means measuring R in the basis defined by $|i'\rangle = U_R |i\rangle$ will yield the same probability p_i as measuring in the $|i\rangle$ basis for state $|\Psi\rangle$.

Problem 2.1 (Functions of the Pauli matrices)

Let $f(\cdot)$ be any function from complex numbers to complex numbers. Let $\vec{n}$ be a normalized vector in three dimensions, and let θ be real. Show that

$$f(\theta\vec{n}\cdot\vec{\sigma}) = \frac{f(\theta) + f(-\theta)}{2} I + \frac{f(\theta) - f(-\theta)}{2}\vec{n}\cdot\vec{\sigma}. \tag{2.225}$$

Solution: Recall from the text that we can define such functions f on normal matrices via a spectral decomposition. In our case, since $\theta\vec{n}$ is a real vector, the matrix $\theta\vec{n}\cdot\vec{\sigma}$ is Hermitian, hence diagonalizable via the spectral theorem. Moreover, since $\theta\vec{n}\cdot\vec{\sigma}$ is traceless, and $(\vec{n}\cdot\vec{\sigma})^2 = I$, the two eigenvalues are $+\theta$ and $-\theta$. Let $|v_+\rangle$ and $|v_-\rangle$ be the corresponding (orthonormal) eigenvectors. Then

$$\vec{n}\cdot\vec{\sigma} = |v_+\rangle\langle v_+| - |v_-\rangle\langle v_-| \tag{2.226}$$

so that

$$f(\theta\vec{n}\cdot\vec{\sigma}) = f(\theta)\, |v_+\rangle\langle v_+| + f(-\theta)\, |v_-\rangle\langle v_-|. \tag{2.227}$$

This answer is true, but not in the form we need and not terribly illuminating. We need to understand how the projectors $|v_+\rangle\langle v_+|$ and $|v_-\rangle\langle v_-|$ relate to the pauli's themselves. Thankfully, we can use the fact that

$$I = |v_+\rangle\langle v_+| + |v_-\rangle\langle v_-|. \tag{2.228}$$

Using (2.226) in conjunction with (2.228), we can solve for the projectors and find that

$$|v_+\rangle\langle v_+| = \frac{I + \vec{n}\cdot\vec{\sigma}}{2}, \qquad |v_-\rangle\langle v_-| = \frac{I - \vec{n}\cdot\vec{\sigma}}{2}. \tag{2.229}$$

Plugging these into (2.227) and rearranging, we get

$$f(\theta \vec{n} \cdot \vec{\sigma}) = f(\theta) \frac{I + \vec{n} \cdot \vec{\sigma}}{2} + f(-\theta) \frac{I - \vec{n} \cdot \vec{\sigma}}{2}, \tag{2.230}$$

or, after grouping in terms of I and $\vec{n} \cdot \vec{\sigma}$,

$$f(\theta \vec{n} \cdot \vec{\sigma}) = \frac{f(\theta) + f(-\theta)}{2} I + \frac{f(\theta) - f(-\theta)}{2} \vec{n} \cdot \vec{\sigma}. \tag{2.231}$$

This is exactly what we had set out to show.

Problem 2.2 (Properties of the Schmidt number)

Suppose $|\psi\rangle$ is a pure state of a composite system with components A and B.

1. Prove that the Schmidt number of $|\psi\rangle$ is equal to the rank of the reduced density matrix $\rho_A \equiv \text{tr}_B(|\psi\rangle \langle\psi|)$. (Note that the rank of a Hermitian operator is equal to the dimension of its support.)

2. Suppose $|\psi\rangle = \sum_j |\alpha_j\rangle |\beta_j\rangle$ is a representation for $|\psi\rangle$, where $|\alpha_j\rangle$ and $|\beta_j\rangle$ are (un-normalized) states for systems A and B, respectively. Prove that the number of terms in such a decomposition is greater than or equal to the Schmidt number of $|\psi\rangle$, $\text{Sch}(\psi)$.

3. Suppose $|\psi\rangle = \alpha |\varphi\rangle + \beta |\gamma\rangle$. Prove that

$$\text{Sch}(\psi) \geq |\text{Sch}(\varphi) + \text{Sch}(\gamma)|. \tag{2.232}$$

Solution: 1. Let $|\psi\rangle = \sum_{i=1}^r \sqrt{p_i} |i\rangle |i'\rangle$ be a Schmidt decomposition of $|\psi\rangle$, with all $p_i > 0$ in the sum. The reduced density matrix ρ is given by

$$\rho = \sum_{i=1}^r p_i |i\rangle \langle i| \tag{2.233}$$

which has rank r since the rank is the span of all $|i\rangle$. Thus, $\text{rank}(\rho) = \text{Sch}(\psi) = r$.

(2) Let $r = \psi$ and and consider the state $\sum_{j=1}^n |\alpha_j\rangle |\beta_j\rangle$ for $n < r$. This is a bipartite quantum state on A and B whose reduced density matrix is given by

$$\rho_A = \sum_{j=1,k=1}^n \langle \beta_k | \beta_j \rangle |\alpha_j\rangle \langle \alpha_k|. \tag{2.234}$$

From this expression, it is clear that the range of ρ_A is contained in the span of the $|alpha_j\rangle$, so $\text{rank}(\rho_A) \leq n < r$. However, we just proved that any representation of $|\psi\rangle$ must give rise to a density matrix whose rank is r. We conclude that no such sum of $n < r$ terms can be a valid representation of $|\psi\rangle$. That is, any representation of $|\psi\rangle$ in this way must have at least r terms.

(3) It's important to recognize that this statement is not necessarily true if α or β is zero. The trouble is, there is a discontinuous jump in the Schmidt number of $\alpha |\varphi\rangle$ from $\alpha \neq 0$ to $\alpha = 0$. To avoid this annoyance, we will reformulate the problem by writing $|\psi\rangle = |\varphi\rangle + |\gamma\rangle$, where $|\varphi\rangle$ and $|\gamma\rangle$ are nonzero vectors.

We will first show a triangle inequality for the Schmidt number, namely that

$$\text{Sch}(\varphi + \gamma) \leq \text{Sch}(\varphi) + \text{Sch}(\gamma). \tag{2.235}$$

This follows directly from part (2). In particular, taking the Schmidt decomposition of both $|\varphi\rangle$ and $|\gamma\rangle$ gives us a representation of $|\psi\rangle$ that has $\text{Sch}(\varphi) + \text{Sch}(\gamma)$ terms. We already claimed this must be greater than the Schmidt number of $|\psi\rangle$. This proves our preliminary result.

We use this result twice. First, we note that $|\varphi\rangle = |\psi\rangle - |\gamma\rangle$. Applying the result, we have $\text{Sch}(\varphi) \leq \text{Sch}(\psi) + \text{Sch}(\gamma)$. By a mirror argument, interchanging φ and γ, gives $\text{Sch}(\gamma) \leq \text{Sch}(\psi) + \text{Sch}(\varphi)$. Together, these results imply

$$\begin{aligned} \text{Sch}(\varphi) - \text{Sch}(\gamma) &\leq \text{Sch}(\psi) \\ \text{Sch}(\gamma) - \text{Sch}(\varphi) &\leq \text{Sch}(\psi). \end{aligned} \tag{2.236}$$

That is,

$$\text{Sch}(\psi) \geq |\text{Sch}(\varphi) - \text{Sch}(\gamma)|. \tag{2.237}$$

This completes the proof.

Problem 2.3 (Tsirelson's inequality)

Suppose $Q = \vec{q} \cdot \vec{\sigma}, R = \vec{r} \cdot \vec{\sigma}, S = \vec{s} \cdot \vec{\sigma}, T = \vec{t} \cdot \vec{\sigma}$, where $\vec{q}, \vec{r}, \vec{s}$ and $\vec{t}$ are real unit vectors in three dimensions. Show that

$$(Q \otimes S + R \otimes S + R \otimes T - Q \otimes T)^2 = 4I + [Q, R] \otimes [S, T]. \tag{2.238}$$

Use this result to prove that

$$\langle Q \otimes S \rangle + \langle R \otimes S \rangle + \langle R \otimes T \rangle - \langle Q \otimes T \rangle \leq 2\sqrt{2}, \tag{2.239}$$

so the violation of the Bell inequiality found in Equation (2.230) is the maximum possible in quantum mechanics.

Solution: There are 16 terms when the product is expanded out. The four matching terms square to the identity, which is what we see in the answer. Let's group the other terms in an $ab + ba$ format. The result $4I$ plus the following

$$\begin{aligned} (Q \otimes S)(R \otimes S) &+ (R \otimes S)(Q \otimes S) + (Q \otimes S)(R \otimes T) + (R \otimes T)(Q \otimes S) \\ &- (Q \otimes S)(Q \otimes T) - (Q \otimes T)(Q \otimes S) + (R \otimes S)(R \otimes T) + (R \otimes T)(R \otimes S) \\ &- (R \otimes S)(Q \otimes T) - (Q \otimes T)(R \otimes S) - (R \otimes T)(Q \otimes T) - (Q \otimes T)(R \otimes T). \end{aligned} \tag{2.240}$$

After simplifying, and noting that all the operators, Q, R, S, T square to I,

$$\begin{aligned} QR \otimes I &+ RQ \otimes I + QR \otimes ST + RQ \otimes TS \\ &- I \otimes ST - I \otimes TS + I \otimes ST + I \otimes TS \\ &- RQ \otimes ST - QR \otimes TS - RQ \otimes I - QR \otimes I. \end{aligned} \tag{2.241}$$

Many of these terms cancel, leaving

$$\begin{aligned} QR \otimes ST + RQ \otimes TS - RQ \otimes ST - QR \otimes TS &= QR \otimes (ST - TS) + RQ \otimes (TS - ST) \\ &= (QR - RQ) \otimes (ST - TS) \\ &= [Q, R] \otimes [S, T]. \end{aligned} \tag{2.242}$$

Combined with the four terms which squared to I, this gives our result.

To obtain Tsirelson's inequality, we can just bound the largest (absolute value) eigenvalue of $X \equiv \langle Q \otimes S \rangle + \langle R \otimes S \rangle + \langle R \otimes T \rangle - \langle Q \otimes T \rangle$. This is given by the spectral norm $\|X\|$. First, we have

$$\|X\|^2 = \left\|X^2\right\| = \|4I + [Q, R] \otimes [S, T]\| \le 4\|I\| + \|[Q, R] \otimes [S, T]\|. \tag{2.243}$$

The spectral norm of I is just 1, while we can bound the other term as follows.

$$\|[Q, R] \otimes [S, T]\| = \|[Q, R]\|\|[S, T]\| \le 2 \times 2 = 4. \tag{2.244}$$

Hence,

$$\|X\|^2 \le 4 + 4 = 8 \tag{2.245}$$

so that $\|X\| \le \sqrt{8} = 2\sqrt{2}$.

On the other hand, the expectation value $\langle X \rangle$ is certaintly less than the magnitude of the largest eigenvalue. This completes the proof of the inequality.

Chapter 3

Introduction to computer science

Exercise 3.1 (Non-computable processes in Nature)

How might we recognize that a process in Nature computes a function not computable by a Turing machine?

Solution: This is a big and open-ended question, and I can't pretend to give a definitive answer. I encourage you to think about this on your own, then feel free to consider my ideas alongside your own.

Perhaps there is a process in nature which is not computable by a Turing machine. It seems reasonable that this process would also lie outside the capabilities of conventional computers. It would be strange if we discovered a function which our laptop can compute, but not a Turing machine, especially if our laptops essentially perform massive but elementary arithmetic. If this is the case, then we might suspect something once our computers fail to compute/model the phenomenon under consideration. Like any scientific investigation, this would never amount to a proof refuting the Church-Turing thesis, since we could never be sure we had not discovered the proper algorithm. Yet such issues would form the basis of evidence for a new type of computation.

There are problems known to be uncomputable for a Turing machine (such as the well-known Halting problem discussed later in the text). Should humans ever observe a natural process compute these functions (whatever that means) for various inputs, or should humans utilize these processes to compute the functions themselves, then this should yield strong evidence for superior computational models.

One should scrutinize what is meant when we say Nature "computes a function." Is this necessarily a function on the positive integers? By what means is the input and output stored? We typically think of intelligent agents computing a function, but in the case of Nature at large we should abandon this notion. It is also difficult to currently imagine what these processes might look like, since both classical and quantum mechanics are computable (i.e. all of physics as we know it). Hence, even quantum computers do not provide the kind of computational power we are describing.

Though perhaps "power" is not the right word here necessarily. One could imagine an exotic situation in which there is a process which computes functions intractable for a Turing machine, yet which cannot compute functions that a Turing machine can. Such a system would violate the idea that certain types of computational problems are objectively more difficult than others, as codified by the hierarchy of complexity classes.

Though this answer is necessarily incomplete, I hope it inspires your own thinking on this interesting question.

Exercise 3.2 (Turing numbers)

Show that single-tape Turing machines can each be given a number from the list $1, 2, 3, \ldots$ in such a way that the number uniquely specifies the corresponding machine. We call this number the *Turing number* of the corresponding Turing machine. (Hint: Every positive integer has a unique prime factorization $p_1^{a_1} p_2^{a_2} \ldots p_k^{a_k}$, where p_i are distinct prime numbers, and $a_1, \ldots, a_k$ are non-negative integers.)

Solution: Finish later, but essentially give a different prime number to each "attribute" of the machine (number of internal states, all possible programs, ...). Note this isn't a bijective map, but we'll get a unique integer guaranteed.

Chapter 4

Quantum circuits

Exercise 4.1

In Exercise 2.11, which you should do now if you haven't already done it, you computed the eigenvectors of the Pauli matrices. Find the points on the Bloch sphere which correspond to the normalized eigenvectors of the different Pauli matrices.

Solution: The idea is that the Pauli operators X, Y, Z represent different axes in three-dimensional space, and so their eigenvectors should point along these axes on the Bloch sphere. Let's see how this actually pans out, using the formula

$$\cos\theta/2 \, |0\rangle + e^{i\varphi}\sin\theta/2 \, |1\rangle \to (\theta, \varphi) \tag{4.1}$$

to find the appropriate angles. Let's consider Z first, with normalized eigenstates $|0\rangle, |1\rangle$. For $|0\rangle$, we need $\cos\theta/2 = 1$, which is solved by $\theta = 0$. Note $\sin\theta/2 = 0$ automatically, and φ can be set to zero for convenience. On the other hand, for $|1\rangle$ we need $\theta = \pi$. This shows what we expected: the eigenstates point along the z-axis in opposite directions.

Let's now consider the trickier case of X. The eigenvectors are $|+\rangle = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle$ and $|-\rangle = \frac{1}{\sqrt{2}}|0\rangle - \frac{1}{\sqrt{2}}|1\rangle$. Comparing the $|0\rangle$ coefficient to the formula (4.1), we see that $\theta = \pi/2$ is the appropriate choice. This sets $\sin\theta/2 = 1/\sqrt{2}$, giving the right magnitude of the $|1\rangle$ component. To get the sign correct (± 1) we can set $\phi = 0$ for $|+\rangle$ and $\phi = \pi$ for $|-\rangle$. With these choices of angles, the eigenstates are seen to align with the Bloch sphere's x-axis.

Can you perform a similar analysis to find the location of the Y eigenstates, given by $|\pm i\rangle = \frac{1}{\sqrt{2}}|0\rangle \pm i\frac{1}{\sqrt{2}}|1\rangle$? What should the answer be?

Exercise 4.2:

Let x be a real number and A a matrix such that $A^2 = I$. Show that

$$\exp(iAx) = \cos(x)I + i\sin(x)A \tag{4.2}$$

Use this result to verify Equations (4.4) through (4.6).

Solution: Let's show this in two different ways: (a) via power series and (b) via diagonalization.

(a) Recall (or accept as given by me!) that the power (Taylor) series for the exponential function of a real number x is given by

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}. \tag{4.3}$$

This actually holds for any complex number z as well. It can be taken as the *definition* for exponentiating square matrices as well.

$$e^{iAx} = \sum_{n=0}^{\infty} \frac{(iAx)^n}{n!} \tag{4.4}$$

Here, we understand that $A^0 = I$. In equation (4.4) we are taking successively larger powers of A, but we know that any power of A is either A or I (why?). This suggests we split the infinite sum into two sums: one over even and one over odd terms.

$$\begin{aligned}
e^{iAx} &= \sum_{n\text{ even}} + \sum_{n\text{ odd}} \\
&= \sum_{n=0}^{\infty} \frac{(iAx)^{2n}}{2n!} + \sum_{n=0}^{\infty} \frac{(iAx)^{2n+1}}{(2n+1)!}
\end{aligned} \tag{4.5}$$

In the last line, we reindexed so that n runs over every integer, but includes only even or odd terms in the original sum. For every n, we have $A^{2n} = (A^2)^n = I$ and $A^{2n+1} = A$. Moreover,

$$i^{2n} = (i^2)^n = (-1)^n \tag{4.6}$$

This allows us to simplify (4.5).

$$e^{iAx} = I \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{2n!} + iA \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!} + \tag{4.7}$$

Looking at the two remaining series, we may or may not recognize them as the series for $\cos x$ and $\sin x$. But if we did recognize it then we would arrive at our answer.

$$e^{iAx} = I \cos x + iA \sin x \tag{4.8}$$

(b) Eigenvalue transformations: It may not be obvious at first glance, but $A^2 = I$ automatically implies A is diagonalizable without needing to assume that A is Hermitian or unitary (or a normal operator more generally). This means we can write

$$A = S\Lambda S^{-1} \tag{4.9}$$

for some invertible matrix S and diagonal Λ. Since $A^2 = I$, the eigenvalues are ± 1, and Λ can be decomposed as

$$\Lambda = \begin{pmatrix} I_k & 0 \\ 0 & -I_{n-k} \end{pmatrix} \tag{4.10}$$

Here, n is the total dimension and I_k is the identity matrix acting on the k eigenvectors with $+1$ eigenvalue.

For a scalar valued function f, we can define $f(A)$ as acting only on the eigenvalues of A: $f(A) = Sf(\Lambda)S^{-1}$. For the particular case of the exponential function, and making use of equation (4.10),

$$\exp(i\Lambda x) = \begin{pmatrix} e^{ix} I_k & 0 \\ 0 & e^{-ix} I_{n-k} \end{pmatrix}. \tag{4.11}$$

Using Euler's formula $e^{ix} = \cos x + i \sin x$, we obtain

$$\exp(i\Lambda x) = I \cos x + i\Lambda \sin x. \tag{4.12}$$

Hence, going back to A itself,

$$\begin{aligned}
\exp(iAx) &= S(I \cos x + i\Lambda \sin x)S^{-1} \\
&= I \cos x + iS\Lambda S^{-1} \sin x \\
&= I \cos x + iA \sin x.
\end{aligned} \tag{4.13}$$

which is the result we sought.

Comparing (a) and (b), I would say method (a) is a little more straightforward computationally. Still, the idea of functions of (diagonalizable) operators acting only on the eigenvalues is a powerful one, and extends to situations where a power series may not be defined.

Exercise 4.3

Show that, up to a global phase, the $\pi/8$ gate satisfies $T = R_z(\pi/4)$.

Solution: We need only compare equation 4.3 of the text with the matrix form of R_z given in equation 4.6 of the text. We see that the two matrices align if we set $\theta/2 = \pi/8$, or $\theta = \pi/4$. The overall phase is shown in 4.3 to be $\pi/8$.

Exercise 4.4

Express the Hadamard gate H as a product of R_x and R_z rotations and $e^{i\varphi}$ for some φ.

Solution: There is value in playing around with the rotation matrices and trying to come up with the answer this way. Shortly later on in the text, you will learn tools that will let you solve this problem systematically (using, e.g. Euler angles). Instead, I will try to outline, as best as a solutions manual can, how one might come about an answer through guess and play.

Naturally, we would like to find the shortest solution possible, but it is clear that R_z or R_x alone cannot work. What about $R_x(\alpha)R_z(\beta)e^{i\gamma}$? Well, before doing the matrix multiplication, we can see that this will not send, say, the zero state to the uniform superposition $|0\rangle + |1\rangle$ as the Hadamard should. Can you see why? Moreover, the Hadamard is real symmetric, and it's not clear that this combination possesses the same symmetry. Combinations like $R_z R_x R_x$ don't add anything, since we can just combine two x rotations to produce a single one. This may lead us, therefore, to try something of the form

$$H \stackrel{?}{=} R_z(\alpha)R_x(\beta)R_z(\gamma)e^{i\delta}. \tag{4.14}$$

Before diving into messy matrix multiplication, let's again consider the action on the computational basis $|0\rangle, |1\rangle$. Thinking geometrically, we need the z axis to get mapped to the x axis. We can accomplish this if $\beta = \pi/2$ and $\alpha = \pi/2$. Think about it! We still haven't determined γ because it had no effect on the z basis. However, we also require the x axis to get mapped back to z. This happens if we set $\gamma = \pi/2$ (again thinking about the rotations geometrically).

Hence, our guess for the Hadamard becomes something like

$$H \stackrel{?}{=} R_z(\pi/2)R_x(\pi/2)R_z(\pi/2)e^{i\delta} \tag{4.15}$$

Let's carry out the matrix multiplication, simplifying from the onset given the special value $\pi/2$.

$$R_z(\pi/2) = \begin{pmatrix} e^{-i\pi/4} & 0 \\ 0 & e^{i\pi/4} \end{pmatrix} = e^{-i\pi/4}\begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}$$
$$R_x(\pi/2) = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -i \\ -i & 1 \end{pmatrix} \tag{4.16}$$

Therefore,

$$R_z(\pi/2)R_x(\pi/2)R_z(\pi/2) = \frac{e^{-i\pi/2}}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = e^{-i\pi/2}H. \tag{4.17}$$

This confirms our geometric intuition is correct. Setting $\delta = i\pi/2$ or rearranging the equation above gives an answer.

$$H = e^{i\pi/2}R_z(\pi/2)R_x(\pi/2)R_z(\pi/2) \tag{4.18}$$

Do you think this solution is unique? Can another form of H be given by sandwiching R_z between two R_x?

Exercise 4.5

Prove that $(\hat{n}\cdot\vec{\sigma})^2 = I$, and use this to verify Equation (4.8).

Solution: We will use the algebraic properties of the σ_i alone, as this gets closer to the "reason" this identity holds. First, let's summarize the relevant properties. Every pauli squares to the identity: $\sigma_i^2 = I$. Moreover, the product of two different paulis is antisymmetric: $\sigma_i\sigma_j + \sigma_j\sigma_i = 0$. With these in hand let's consider $(n\cdot\sigma)^2 = (\sum_i n_i\sigma_i)^2$, where the sum goes from $i = 1$ to 3. We have

$$\begin{aligned}
(n\cdot\sigma)^2 &= \sum_i n_i\sigma_i \sum_j n_j\sigma_j \\
&= \sum_{i,j} n_i n_j \sigma_i\sigma_j \\
&= \sum_{i=j} n_i^2\sigma_i^2 + \sum_{i<j} n_i n_j(\sigma_i\sigma_j + \sigma_j\sigma_i).
\end{aligned} \tag{4.19}$$

Along the way, we split the sum into pieces where the i, j indices were equal and pieces where they weren't. By the properties we discussed above, the sum on the right equals exactly zero, and $\sigma_i^2 = I$. Therefore,

$$(n\cdot\sigma)^2 = I\sum_i n_i^2 = I\|n\|^2 \tag{4.20}$$

When n is a unit vector, $\|n\| = 1$ and we arrive at our result.

Exercise 4.6: (Bloch sphere interpretation of rotations)

One reason why the $R_{\hat{n}}(\theta)$ operators are referred to as rotation operators is the following fact, which you are to prove. Suppose a single qubit has a state represented by the Bloch vector $\vec{\lambda}$. Then the effect of the rotation $R_{\hat{n}}(\theta)$ on the state is to rotate it by an angle θ about the $\hat{n}$ axis of the Bloch sphere. This fact explains the rather mysterious looking factor of two in the definition of the rotation matrices.

Solution: For simplicity of exposition, we will take the rotation under consideration to be about the z axis. Otherwise the matrix computations can get quite challenging. We could argue that we are just "choosing a frame" in which the rotation is about z, but this might be assuming part of the consequent.

Let $R_z(\alpha)$ represent the 2×2 matrix representing the rotation given by (4.6) of the text, while $S_z(\alpha)$ represents the actual rotation in three-dimensional space.

$$S_z(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{4.21}$$

Let $\vec{\lambda} = (\cos\varphi\sin\theta, \sin\varphi\sin\theta, \cos\theta)$ represent the unit vector on the Bloch sphere, with corresponding state

$$|\lambda\rangle = \begin{pmatrix} \cos\theta/2 \\ e^{i\varphi}\sin\theta/2. \end{pmatrix} \tag{4.22}$$

Applying the rotation of angle α to λ gives

$$R_z(\alpha)|\lambda\rangle = \begin{pmatrix} e^{-i\alpha/2}\cos\theta/2 \\ e^{i\alpha/2}e^{i\varphi}\sin\theta/2 \end{pmatrix} \cong \begin{pmatrix} \cos\theta/2 \\ e^{i(\varphi+\alpha)}\sin\theta/2. \end{pmatrix} \tag{4.23}$$

Where $\cong$ here denotes equivalence up to a phase. We see that the final state $|\lambda'\rangle$ after the rotation has a corresponding Bloch sphere representation with the same polar angle θ but azimuthal angle $\varphi' = \phi + \alpha$. Thus, we obtain the expected rotation of the state.

Exercise 4.7

Show that $XYX = -Y$ and use this to prove that $XR_y(\theta)X = Ry(-\theta)$.

Solution: Since each pauli is its own inverse (e.g. $X^2 = I$), we see that the first equation is a direct consequence of rearranging $XY = -YX$. This antisymmetry property should really be thought of as a *defining* property of the paulis, with the specific matrix form being only a representation producing the desired relations. Nevertheless, we can compute the matrix product XYX and see that the answer is indeed $-Y$.

As for the second part, we can proceed from the power series which defines the exponential for R_y, but it is easier to use the result of Exercise 4.2.

$$\begin{aligned} XR_y(\theta)X &= xe^{-iY\theta/2}X \\ &= X(I\cos\theta/2 - iY\sin\theta/2)X \\ &= XIX\cos\theta/2 - iXYX\sin\theta/2 \\ &= I\cos\theta/2 + iY\sin\theta/2 \\ &= e^{iY\theta/2} \\ &= R(-\theta). \end{aligned} \tag{4.24}$$

Exercise 4.8

An arbitrary single qubit unitary operator can be written in the form

$$U = \exp(i\alpha)R_{\hat{n}}(\theta) \tag{4.25}$$

for some real numbers α and θ, and a real three-dimensional unit vector $\hat{n}$.

1. Prove this fact.

2. Find values for α, θ, and $\hat{n}$ giving the Hadamard gate H.

3. Find values for α, θ, and $\hat{n}$ giving the phase gate

$$S = \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}. \tag{4.26}$$

Solution: 1. As our guiding idea, we will want use the spectral theorem to connect U to the notion of a rotation. The eigenvectors of U should correspond to the axis of rotation because they are unchanged by the action of U, and the eigenvalues should have to do with the angle of rotation.

Let's pursue this more formally. Any 2×2 unitary U can be written in diagonalized form as

$$U = e^{i\varphi_1} |v_1\rangle \langle v_1| + e^{i\varphi_2} |v_2\rangle \langle v_2| . \tag{4.27}$$

Since an overall phase is irrelevant, only the different in eigenvalues should matter. To capture this, define $\alpha = (\varphi_1 + \varphi_2)/2$ and $\theta = \varphi_2 - \varphi_1$. In terms of these parameters, U can be written as

$$U = e^{i\alpha}(e^{-i\theta/2} |v_1\rangle \langle v_1| + e^{i\theta/2} |v_2\rangle \langle v_2|). \tag{4.28}$$

Since v_1 and v_2 are orthogonal, they should lie on opposite ends of the Bloch sphere. Let $\hat{n}$ correspond to the unit vector pointing in the direction of $\vec{v_1}$. Then, the piece in parenthesis equation (4.28) is exactly the rotation matrix $R_{\hat{n}}(\theta)$. Thus, $U = e^{i\alpha} R_{\hat{n}}(\theta)$.

2. The rotation information will be given by the eigenvectors and eigenvalues of H. Try to calculate them yourself! The eigenvalues are $+1$ and -1 with corresponding vectors

$$|v_\pm\rangle = \frac{1 \pm \sqrt{2}}{\sqrt{2(2 \pm \sqrt{2})}} |0\rangle + \frac{1}{\sqrt{2(2 \pm \sqrt{2})}} |1\rangle . \tag{4.29}$$

Since there is no complex phase anywhere, these vectors must lie in the xz plane of the bloch sphere. We can get the appropriate polar angle by looking at the $|0\rangle$ component of one of the eigenvectors, $|v_+\rangle$ say. We need to solve

$$\cos\theta/2 = \frac{1 + \sqrt{2}}{\sqrt{2(2 + \sqrt{2})}} \tag{4.30}$$

. This may seem daunting, but the answer is actually quite simple: $\theta = \pi/4$! The rotation angle should be the difference in phase between the two eigenvalues, which is π. The overall phase is the average of the eigenvalue phases: $(\pi + 0)/2 = \pi/2$. Hence,

$$H = e^{i\pi/2} R_{\hat{n}}(\pi) \tag{4.31}$$

where $\hat{n} = (\sin\theta/4, 0, \cos\theta/4) = (1, 0, 1)/\sqrt{2}$. Can you verify that this answer is correct? Now that you know H can be thought of as a rotation, are you able to picture the trajectory of the computational basis state as they are acted on with H?

3. For the S gate, we are going to use a bit more intuition. S is diagonal, so we should be dealing with a z rotation. Factoring out an overall phase of $e^{-i\pi/4}$ (recall that $i = e^{i\pi/2}$),

$$S = e^{-i\pi/4} \begin{pmatrix} e^{-i\pi/4} & 0 \\ 0 & e^{i\pi/4} \end{pmatrix} = e^{-i\pi/4} R_z(\pi/2). \tag{4.32}$$

This gives our solution.

Exercise 4.9

Explain why any single qubit unitary operator may be written in the form (4.12).

Solution: Let's start with U having entries z_1, z_2, z_3, z_4 being complex, and understand what conditions are in place. Any complex number can be written as $z = re^{i\varphi}$ with $r \geq 0$ and $\varphi \in [0, 2\pi)$. Hence, we must have a form

$$U = \begin{pmatrix} r_1 e^{i\varphi_1} & r_3 e^{i\varphi_3} \\ r_2 e^{i\varphi_2} & r_4 e^{i\varphi_4} \end{pmatrix}. \tag{4.33}$$

What does the normalization constraint on the columns and rows impose on U? It says nothing about the phases, but does forces the amplitudes r_i to be closely related: $r_2 = r_3 = \sqrt{1 - r_1^2}$ and $r_4 = r_1$. For ease of notation, let's now call $r_1 = r$, which lies in $[0, 1]$. The phases are contrained by the orthogonality of the rows and the columns. It turns out only one of these constraints is independent, so we focus on the inner product of the two columns. It is given by

$$r\sqrt{1 - r^2} e^{i(\varphi_1 - \varphi_3)} + r\sqrt{1 - r^2} e^{i(\varphi_2 - \varphi_4)} = 0. \tag{4.34}$$

Before proceeding further with this, we need to handle edge cases where we can't divide by $r\sqrt{1 - r^2}$. Consider the edge cases $r = 0$ and $r = 1$ separately from the more general $r \in (0, 1)$. If $r = 0$, then U is off-diagonal only, and we can achieve this by setting $\gamma = \pi$. Then we just need to set α, β, δ to match φ_2 and φ_3, and some algebra shows that $\alpha = (\varphi_2 + \varphi_3 - \pi)/2$, $\beta = \varphi_2 - \varphi_3 + \pi$ and $\delta = 0$ work for this case. This handles the $r = 0$ case. For $r = 1$, the matrix U is diagonal, so we choose $\gamma = 0$. Only the phases φ_1 and φ_4 need to be expressed in terms of α, β, δ. The exact same solution given for $r = 0$ works here too.

Now going back to the main case, in equation (4.34) we can divide by $r\sqrt{1 - r^2}$ and rearrange. The result is

$$\varphi_1 + \varphi_4 - \varphi_2 - \varphi_3 = \delta_k \tag{4.35}$$

where δ_k is some odd multiple of π. Rearranging in terms of φ_3 gives

$$\varphi_3 = \varphi_1 + \varphi_4 - \varphi_2 + \delta_k. \tag{4.36}$$

Plugging this into U,

$$U = \begin{pmatrix} re^{i\varphi_i} & -\sqrt{1 - r^2} e^{i\varphi_1 + \varphi_4 - \varphi_2} \\ \sqrt{1 - r^2} e^{i\varphi_2} & re^{i\varphi_4} \end{pmatrix} \tag{4.37}$$

Let's choose $\gamma \in (0, \pi)$ such that $r = \cos \gamma/2$. This handles the amplitude part of U correctly. To parameterize the phases of U in terms of α, β, δ we will need to construct a solution to the following linear equations.

$$\begin{pmatrix} \varphi_1 \\ \varphi_2 \\ \varphi_4 \end{pmatrix} = \begin{pmatrix} 1 & -1/2 & -1/2 \\ 1 & 1/2 & -1/2 \\ 1 & 1/2 & 1/2 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix}. \tag{4.38}$$

Indeed, a unique solution can be found by inverting the matrix, meaning that the α, β, γ can be used instead of the φ to parameterize the matrix. Hence, equation (4.12) of the textbook gives a completely general characterization of 2×2 unitary matrices.

Exercise 4.10

Give a decomposition analogous to Theorem 4.1 but using Rx instead of Rz.

Chapter 5

The quantum Fourier transform and its applications

Exercise 5.1

Give a direct proof that the linear transformation defined by Equation (5.2) is unitary.

Solution: It suffices to show that, for any two computational basis states $|j\rangle, |k\rangle$,

$$\langle j| (QFT)^\dagger (QFT) |k\rangle = \langle j|k\rangle = \delta_{ij}. \tag{5.1}$$

To do this, we substitute the definition into the above equation.

$$\langle k| (QFT)^\dagger (QFT) |j\rangle = \Big(\frac{1}{\sqrt{N}} \sum_{p=0}^{N-1} e^{-2\pi ikp/N} \langle p|\Big)\Big(\frac{1}{\sqrt{N}} \sum_{q=0}^{N-1} e^{2\pi ijq/N} |q\rangle\Big)$$

$$= \frac{1}{N} \sum_{p=0}^{N-1} \sum_{q=0}^{N-1} e^{2\pi i(jq-kp)/N} \langle p|q\rangle \tag{5.2}$$

$$= \frac{1}{N} \sum_{p=0}^{N-1} e^{2\pi i(j-k)p/N}$$

where in the last step we used the orthonormality of the p, q states to eliminate one of the sums. Clearly, if $j = k$, the result is exactly one, as desired. Otherwise, $j - k$ is a nonzero integer, say n, such that $|n| < N$. We will show that in this case the sum above is zero.

The basic idea is that we are taking a sum over phases which are symmetrically distributed around the unit circle, so the result must be zero. To make this argument rigorous, multiply the sum by $e^{2\pi in/N}$.

$$e^{2\pi in/N} \sum_{p=0}^{N-1} \big(e^{2\pi in/N}\big)^p = \sum_{p=0}^{N-1} \big(e^{2\pi in/N}\big)^{p+1} = \sum_{p=1}^{N} \big(e^{2\pi in/N}\big)^p \tag{5.3}$$

In the last equation we simply reindexed. Because of the N-periodicity, $\big(e^{2\pi in/N}\big)^N = 1 = \big(e^{2\pi in/N}\big)^0$. Hence, we see that the sum is left unchanged by the multiplication. Since $e^{2\pi in/N} \neq 1$, the sum must in fact be zero. This completes the proof.

Exercise 5.2

Explicitly compute the Fourier transform of the n qubit state $|00...0\rangle$.

Solution: Suppose there are n qubits, so that $N = 2^n$. Using the definition given directly above in the textbook,

$$|00...0\rangle \rightarrow \frac{1}{2^{n/2}} \sum_{k=0}^{2^n-1} e^0 |k\rangle \tag{5.4}$$

$$= \frac{1}{2^{n/2}} \sum_{k=0}^{2^n-1} |k\rangle, \tag{5.5}$$

which is simply a uniform superposition over the computational basis states. Evidently, the QFT on the zero state simply acts the same as Hadamards on all the qubits!

We remark that this result is consistent with the interpretation that the Fourier transform decomposes a "signal" into its frequency components. Here, the signal was a sharp spike, which requires an large spread in frequency to construct. Conversely, a uniform superposition without phases is like a constant function signal, which has a frequency of zero.

Exercise 5.3 (Classical fast Fourier transform)

Suppose we wish to perform a Fourier transform of a vector containing 2^n complex numbers on a classical computer. Verify that the straightforward method for performing the Fourier transform, based upon direct evaluation of Equation (5.1) requires $\Theta(2^{2n})$ elementary arithmetic operations. Find a method for reducing this to $\Theta(n2^n)$ operations, based upon Equation (5.4). There are 2^n complex numbers we need to compute, which are the output amplitudes of the Fourier transform. If we compute each one using (5.1), each such amplitude involves a sum which contains 2^n terms. Thus, there will be $2^n \times 2^n = 2^{2n}$ summations and therefore at least as many arithmetic operations.

Solution: Let's now consider a computation based on the factored form of the QFT, Equation (5.4). As before, this involves a computation of 2^n amplitudes, one for each bitstring $k = k_1 k_2 ... k_n$. Using (5.4) the amplitude a_k corresponding to the state $|k\rangle$ is given by

$$\langle k| \, QFT \, |j\rangle = \frac{1}{2^{n/2}} \left(\delta_{k_1 0} + e^{2\pi i 0.j_n} \delta_{k_1 2}\right) \left(\delta_{k_2 0} + e^{2\pi i 0.j_{n-1}j_n} \delta_{k_2 2}\right) ... \left(\delta_{k_n 0} + e^{2\pi i 0.j_1 ... j_n} \delta_{k_n 2}\right). \tag{5.6}$$

where $|j\rangle$ is our input state. This involves a multiplication of n terms, hence there are $n \times 2^n$ total multiplications. This is a lower bound for the number of operations.

Exercise 5.4

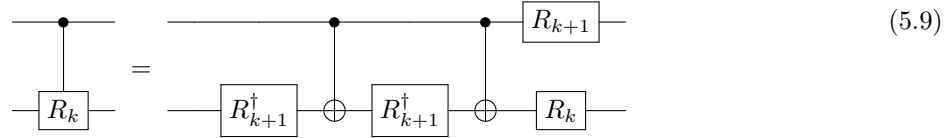Give a decomposition of the controlled-R_k gate into single qubit and CNOT gates.

Solution: We use the ABC construction of Corollary 4.2 to make our controlled R_k according to Figure 4.6. First, note that

$$R_k = \begin{pmatrix} 1 & 0 \\ 0 & e^{2\pi i/2^k} \end{pmatrix} = e^{2\pi i/2^{k+1}} \begin{pmatrix} e^{-2\pi i/2^{k+1}} & 0 \\ 0 & e^{2\pi i/2^{k+1}} \end{pmatrix} = e^{i\alpha} R_z(\beta) \tag{5.7}$$

where $\alpha = 2\pi/2^{k+1}$ and $\beta = 2\pi/2^k$. Comparing this to the Euler decomposition formula of Theorem 4.1, we set $\gamma = \delta = 0$. Following through the steps, this implies,

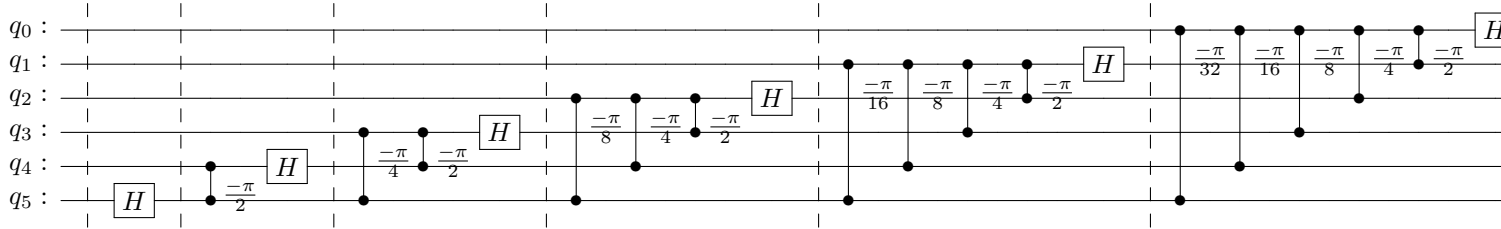$$\begin{aligned} A &= R_z(\beta) \\ B &= R_z(-\beta/2) \\ C &= R_z(-\beta/2) \end{aligned} \tag{5.8}$$

can be used in the ABC construction of R_k. As a final step, primarily one of cosmetics, we notice these gates are related to the R_k through global phases which cancel each other out. Thus, the following circuit implements the controlled-R_k, as is easy to verify.

 (5.9)

Exercise 5.5

Give a quantum circuit to perform the inverse quantum Fourier transform.

Solution: Here is the circuit for six qubits (generated using qiskit).

Here, the vertical line segments are the R_k gates, where the number alongside indicate angle of phase rotated. Like the inverse to any quantum circuit, can be obtained by reversing the order of the gates and taking the inverse of each gate. Note the Hadamard H is self-inverse.

Exercise 5.6 (Approximate quantum Fourier transform)

The quantum circuit construction of the quantum Fourier transform apparently requires gates of exponential precision in the number of qubits used. However, such precision is never required in any quantum circuit of polynomial size. For example, let U be the ideal quantum Fourier transform on n qubits, and V be the transform which results if the controlled-R_k gates are performed to a precision $\Delta = 1/p(n)$ for some polynomial $p(n)$. Show that the error $E(U,V) \equiv max_{|\psi\rangle} \|(U - V)|\psi\rangle\|$ scales as $\Theta(n^2/p(n))$, and thus polynomial precision in each gate is sufficient to guarantee polynomial accuracy in the output state.

Solution: First we will show a more general result (one which will be further generalized in part 3 of the book, when discussing quantum channels). Let $\mathcal{X}$ and $\mathcal{Y}$ be quantum gates, and let X and Y be gates which we will think of as approximating $\mathcal{X}$ and $\mathcal{Y}$ respectively. We will show that

$$E(\mathcal{X}\mathcal{Y}, XY) \le E(\mathcal{X}, X) + E(\mathcal{Y}, Y). \tag{5.10}$$

To proceed, we set things up to make use of our friend: the triangle inequality. For any state $|\psi\rangle$ we have

$$\|(\mathcal{X}\mathcal{Y} - XY)|\psi\rangle\| = \|(\mathcal{X}\mathcal{Y} - X\mathcal{Y} + X\mathcal{Y} - XY)|\psi\rangle\| \tag{5.11}$$

$$= \|(\mathcal{X} - X)\mathcal{Y}|\psi\rangle + X(\mathcal{Y} - Y)|\psi\rangle\| \tag{5.12}$$

$$\leq \|(\mathcal{X} - X)\mathcal{Y}|\psi\rangle\| + \|X(\mathcal{Y} - Y)|\psi\rangle\| \tag{5.13}$$

Since this inequality holds for any $|\psi\rangle$, it certainly holds if we take the max of both sides. Hence,

$$E(\mathcal{X}\mathcal{Y}, XY) \leq \max_{|\psi\rangle}\left(\|(\mathcal{X} - X)\mathcal{Y}|\psi\rangle\| + \|X(\mathcal{Y} - Y)|\psi\rangle\|\right) \tag{5.14}$$

Certainly, the maximum of a sum $A + B$ is less than (or equal to) maximizing each individual piece A, B, so we may distribute the max function and maintain the inequality. Let's consider each term on the left hand side. Since $\mathcal{Y}$ is unitary, it is a bijection on the space of valid states. Hence, we can maximize over $|\phi\rangle = \mathcal{Y}|\psi\rangle$ instead. Now consider the rightmost term. Since X is unitary, it preserves norm. Altogether,

$$E(\mathcal{X}\mathcal{Y}, XY) \leq \max_{|\phi\rangle}\|(\mathcal{X} - X)|\phi\rangle\| + \max_{|\psi\rangle}\|(\mathcal{Y} - Y)|\psi\rangle\| \tag{5.15}$$

$$= E(\mathcal{X}, X) + E(\mathcal{Y}, Y). \tag{5.16}$$

This places a bound on the error when composing imperfect gates. Moreover, this bound is tight, since if $X = \mathcal{X}$ and $Y = \mathcal{Y}$ we get strict equality. We can generalize this result, by simple induction, to arbitrary sequences of gates with corresponding approximations. Thus, in our present case, if each controlled R_k has precision Δ,

$$E(U, V) \leq \frac{n(n + 1)}{2}\Delta \in \Theta(n^2\Delta). \tag{5.17}$$

Here, the use of Θ is appropriate rather than $\mathcal{O}$ since our bound is tight.

Exercise 5.7

Additional insight into the circuit in Figure 5.2 may be obtained by showing, as you should now do, that the effect of the sequence of controlled-U operations like that in Figure 5.2 is to take the state $|j\rangle|u\rangle$ to $|j\rangle U^j|u\rangle$. (Note that this does not depend on $|u\rangle$ being an eigenstate of U.)

Solution: Suppose there are t qubits in the first register, so the integer j can be expressed in binary as $j_{t-1}...j_1 j_0$, with $j_k \in \{0, 1\}$ for every $0 \leq k < t$. By definition, this means $j = j_0 + 2j_1 + ... + 2^{t-1}j_{t-1}$. The state $|j\rangle$ has tensor product form

$$|j\rangle = \bigotimes_{k=0}^{t-1}|j_k\rangle \equiv |j_{t-1}\rangle...|j_0\rangle. \tag{5.18}$$

The action of the controlled-U^{2^k} controlled on the kth qubit is given by

$$|j_k\rangle|u\rangle \longrightarrow |j_k\rangle U^{2^k j_k}|u\rangle \tag{5.19}$$

since, if $j_k = 1$, the unitary is applied, otherwise ($j_k = 0$) nothing should happen. Thus, the full sequence of controlled gates acts as follows.

$$|j\rangle|u\rangle = \bigotimes_{k=0}^{t-1}|j_k\rangle|u\rangle \longrightarrow |j\rangle U^{2^{t-1}j_{t-1}}...U^{2^0 j_0}|u\rangle$$

$$= |j\rangle U^{2^0 j_0 + ... + 2^{t-1}j_{t-1}}|u\rangle \tag{5.20}$$

$$= |j\rangle U^j|u\rangle$$

In the last step we reused the definition of j being expressed in binary. This gives the desired result, which, as we see, did not rely on particular knowledge of the state $|u\rangle$.

Exercise 5.8

Suppose the phase estimation algorithm takes the state $|0\rangle\,|u\rangle$ to the state $|\tilde{\varphi}_u\rangle\,|u\rangle$, so that the input $|0\rangle\left(\sum_u c_u\,|u\rangle\right)$, the algorithm outputs $\sum_u c_u\,|\tilde{\varphi}_u\rangle\,|u\rangle$. Show that if t is chosen according to (5.35), then the probability for measuring φ_u accurate to n bits at the conclusion of the phase estimation algorithm is at least $|c_u|^2(1-\epsilon)$.

Solution: If t is chosen as such, then $\tilde{\varphi}_u$ is an n-bit approximation to φ_u with probability $p_{succ} \geq (1-\epsilon)$. Meanwhile, the probability of measuring $\tilde{\varphi}_u$ on the first register is given by the Born rule: $p_u = |c_u|^2$. These two events are independent, hence, the probability of measuring $\tilde{\varphi}_u$ *and* having it be an n-bit approximation is

$$p_u p_{succ} \geq |c_u|^2(1-\epsilon). \tag{5.21}$$

Moreover, any other eigenstate $|v\rangle$ of U such that $\varphi_v \neq \varphi_u$ might still result in an n-bit approximation to φ_u, provided they are sufficiently close (it may even be that $\tilde{\varphi}_v = \tilde{\varphi}_u$). This will only further increase the probability of success. In any case, the right side of (5.21) remains a lower bound.

Exercise 5.9

Let U be a unitary transform with eigenvalues ± 1, which acts on a state $|\psi\rangle$. Using the phase estimation procedure, construct a quantum circuit to collapse $|\psi\rangle$ into one or the other of the two eigenspaces of U, giving also a classical indicator as to which space the final state is in. Compare your result with Exercise 4.34.

Solution: If the eigenvalues of U are 1 and -1, the corresponding phases are 0.0 and 0.1 respectively. Because these phases are finite bitstrings, there is no possibility of error and, in fact, we can take $t = n$, which in our case is 1. The inverse-QFT on one qubit is simply the Hadamard, and our circuit reduces to the following.

$$\tag{5.22}$$

If a 0 (1) is measured, the final state is known to be in the plus (minus) subspace. The probability of each outcome is simply related to the initial overlap with each subspace via the Born rule.

Exercise 5.10

Show that the order of $x = 5$ modulo $N = 21$ is 6.

Solution: We proceed by exhaustive calculation.

$$5^1 \bmod 21 = 5$$
$$5^2 \bmod 21 = 4$$
$$5^3 \bmod 21 = 20$$
$$5^4 \bmod 21 = 16 \tag{5.23}$$
$$5^5 \bmod 21 = 17$$
$$5^6 \bmod 21 = 1$$

Hence, 6 is the smallest positive integer r such that $5^r \bmod 21 = 1$. Thus, 6 is the order of 5 modulo 21.

Exercise 5.11

Show that the order of x satisfies $r \leq N$.

Solution: Consider the set $\{x^n \bmod N\}_{n=1}^N$. Because we assume x and N share no common factors, it is not possible for $x^n \bmod N = 0$. Hence, $0 < x^n \bmod N < N$. By the pigeonhole principle, not all the values for $x^n \bmod N$ can be unique. There must exist some $x^i = x^j \bmod N$ for some $1 \leq i, j \leq N$ and $j \neq i$. Assume $j > i$ without loss of generality. Then, we have

$$x^j - x^i = 0 \bmod N$$
$$x^i(x^{j-i} - 1) = 0 \bmod N. \tag{5.24}$$

This implies $N | x^i(x^{j-i} - 1)$. But $N \nmid x^i$, again by assumption of no common factors. Hence, we must have $N|(x^{j-i} - 1)$, or $x^{j-i} = 1 \bmod N$. Since r is the smallest integer satisfying this condition, we must have $r \leq j - i < N$. Note this is a strict inequality, unlike what is given in the text.

Exercise 5.12

Show that U is unitary (Hint: x is co-prime to N, and therefore has an inverse modulo N).

Solution: Since U acts as a map on the computational basis states, it suffices to show this map is injective (one-to-one). If $y \geq N$, then $|y\rangle$ is mapped to itself. On the other hand, if $0 \leq y < N$, then $|y\rangle$ is certainly mapped to some $|z\rangle$ where $z < N$. Hence, U is injective on the subspace where $y \geq N$, and it suffices to focus on the case where $y < N$. To this end, suppose $U|y\rangle = U|z\rangle$ for $y, z < N$. Then,

$$xy \bmod N = xz \bmod N$$
$$x(y - z) = 0 \pmod N \tag{5.25}$$

This implies N divides $x(y - z)$. However, since N and x are coprime, they share no common factors. This implies $N|(y - z)$. But $|y - z| < N$, so it must be that $y - z = 0$. Hence

$$|y\rangle = |z\rangle \tag{5.26}$$

which proves that U is injective on the basis, as desired.

Exercise 5.13

Prove (5.44). (*Hint*:$\sum_{s=0}^{r-1} \exp(-2\pi i s k / r) = r\delta_{k0}$). In fact, prove that

$$\frac{1}{\sqrt{r}} \sum_{s=0}^{r-1} e^{2\pi i s k / r} |u_s\rangle = |x^k \bmod N\rangle \tag{5.27}$$

Solution: Let us crank the wheel: putting in the definition of $|u_s\rangle$ to the left side of (5.44) and regrouping.

$$\frac{1}{\sqrt{r}} \sum_{s=0}^{r-1} e^{2\pi i s k / r} |u_s\rangle = \frac{1}{\sqrt{r}} \sum_{s=0}^{r-1} e^{2\pi i s k / r} \frac{1}{\sqrt{r}} \sum_{t=0}^{r-1} e^{-2\pi i s t / r} |x^t \bmod N\rangle \tag{5.28}$$

$$= \frac{1}{r} \sum_{s=0}^{r-1} \sum_{t=0}^{r-1} e^{2\pi i s (k-t)/r} |x^t \bmod N\rangle \tag{5.29}$$

$$= \frac{1}{r} \sum_{t=0}^{r-1} |x^t \bmod N\rangle \left(\sum_{s=0}^{r-1} e^{2\pi i s (k-t)/r} \right) \tag{5.30}$$

Making gracious use of the hint, we see the rightmost sum on the last line equals $r\delta_{kt}$. Hence,

$$\frac{1}{\sqrt{r}} \sum_{s=0}^{r-1} e^{2\pi i s k/r} \left|u_s\right\rangle = \sum_{t=0}^{r-1} \delta_{tk} \left|x^t \bmod N\right\rangle \tag{5.31}$$

$$= \left|x^k \bmod N\right\rangle. \tag{5.32}$$

To obtain (5.44) from the text, simply set $k = 0$.

Exercise 5.14

The quantum state produced in the order-finding algorithm, before the inverse Fourier transform, is

$$\left|\psi\right\rangle = \sum_{j=0}^{2^t-1} \left|j\right\rangle U^j \left|1\right\rangle = \sum_{j=0}^{2^t-1} \left|j\right\rangle \left|x^j \bmod N\right\rangle \tag{5.33}$$

if we initialize the second register as $\left|1\right\rangle$. Show that the same state is obtained if we replace U^j with a *different* unitary transform V, which computes

$$V \left|j\right\rangle \left|k\right\rangle = \left|j\right\rangle \left|k + x^j \bmod N\right\rangle, \tag{5.34}$$

and start the second register in the state $\left|0\right\rangle$. Also show how to construct V using $O(L^3)$ gates.

Solution: It is obvious that, by setting $k = 0$, we obtain the same output using V as we would if $k = 1$ and we had acted with U. To construct V we can simply use the U^j as before and store the result in an ancillary register whose initial value was one, then *add* that result to the register initialized to 0. The addition step can be carried out bitwise, and only takes $O(L^2)$ gates naively. Hence, the algorithm remains $O(L^3)$ after this addition step.

Exercise 5.15

Show that the least common multiple of positive integers x and y is $xy/\gcd(x, y)$, and thus may be computed in $O(L^2)$ operations if x and y are L bit numbers.

Solution: Let $d = \gcd(x, y)$. Since $d|x$ and $d|y$, it divides the product, so $xy = dn$ for some integer $n > 0$. Our goal is to prove that $n = \text{lcm}(x, y)$.

By Bézout's lemma, there exists integers a, b such that $d = ax + by$. Let m be a common multiple of x and y. Then there exists integers s and t such that $m = sx = ty$. Consider the product md. Using the lemma,

$$md = axm + bym$$
$$= axty + bysx \tag{5.35}$$
$$= xy(at + bs).$$

This implies $xy|md$, or since $n = xy/d$, $n|m$. Therefore, n divides every common multiple of x and y. On the other hand, n is itself a common multiple, because

$$n = x\left(\frac{y}{d}\right) = \left(\frac{x}{d}\right)y, \tag{5.36}$$

where the terms in parenthesis are integers. Thus, n is the least common multiple, and the statement is proved.

Exercise 5.16

For all $x \geq 2$ prove that $\int_x^{x+1} 1/y^2 dy \geq 2/3x^2$. Show that

$$\sum_q \frac{1}{q^2} \leq \frac{3}{2} \int_2^\infty \frac{1}{y^2} dy = \frac{3}{4}, \tag{5.37}$$

and thus that (5.58) holds.

   Solution: Performing the integral directly yields the solution

$$\int_x^{x+1} 1/y^2 dy = \frac{1}{x^2 + x}. \tag{5.38}$$

For all $x \geq 2$,

$$x \leq \frac{x^2}{2}. \tag{5.39}$$

Thus,

$$\int_x^{x+1} 1/y^2 dy = \frac{1}{x^2 + x} \geq \frac{1}{x^2 + x^2/2} = \frac{2}{3x^2} \tag{5.40}$$

To complete the exercise, note that the sum over prime numbers q is less than the same sum over all integers starting from 2. Thus,

$$\sum_q \frac{1}{q^2} \leq \sum_{n \geq 2} 1/n^2$$

$$\leq \sum_{n \geq 2} \frac{3}{2} \int_n^{n+1} \frac{1}{y^2} dy \tag{5.41}$$

$$= \frac{3}{2} \int_2^\infty \frac{1}{y^2} dy = \frac{3}{4}$$

This proves our result.

Exercise 5.17

Suppose N is L bits long. The aim of this exercise is to find an efficient classical algorithm to determine whether $N = a^b$ for some integers $a \geq 1$, $b \geq 2$. This may be done as follows:

1. Show that b, if it exists, satisfies $b \leq L$.

2. Show that it takes at most $O(L^2)$ operations to compute $\log_2 N, x = y/b$ for $b \leq L$, and the two integers u_1 and u_2 nearest to 2^x.

3. Show that it takes at most $O(L^2)$ operations to compute u_1^b and u_2^b (use repeated squaring) and check to see if either is equal to N.

4. Combine the previous results to give an $O(L^3)$ operation algorithm to determine whether $N = a^b$ for integers a and b.

Solution: (1) Note that this statement does not actually hold in the special case $N = 1$, since $b \geq 2$ by assumption and $L = \log_2 1 = 0$. Now to the more interesting case of $N \geq 2$. Then it must be that $a \geq 2$. Assuming that b exists, it may be obtained by computing $\log_a N$. Meanwhile, $L = \lceil \log_2 N \rceil = \lceil b \log_2 a \rceil$. Since $\log_2 a \geq 1$, $L \geq b$.

(2)

Exercise 5.18 (Factoring 91)

Suppose we wish to factor $N = 91$. Confirm that steps **1** and **2** are passed. For step **3**, suppose we choose $x = 4$, which is co-prime to 91. Compute the order r of x with respect to N, and show that $x^{r/2}$ mod $91 = 64 \neq -1 \pmod{91}$, so the algorithm succeeds, giving $\gcd(64 - 1, 19) = 7$.
It is unlikely that this is the most efficient method you've seen for factoring 91. Indeed, if all computations had to be carried out on a classical computer, this reduction would not result in an efficient factoring algorithm, as no efficient method is known for solving the order-finding problem on a classical computer.

Solution: Step **1** is passed since 91 is odd.

Exercise 5.19

Show that $N = 15$ is the smallest number for which the order-finding subroutine is required, that is, it is the smallest composite number that is not even or a power of some smaller integer.

Solution: This can be checked by exhaustion. See the below list

1	Special case (N/A)
2	Prime
3	Prime
4	2^2
5	Prime
6	Even
7	Prime
8	2^3
9	3^2
10	Even
11	Prime
12	Even
13	Prime
14	Even
15	Composite, Odd, distinct prime factors ✓

Exercise 5.20

Suppose $f(x + r) = f(x)$, and $0 \leq x < N$ for N an integer multiple of r. Compute

$$\hat{f}(l) \equiv \frac{1}{\sqrt{N}} \sum_{x=0}^{N-1} e^{-2\pi i l x / N} f(x), \tag{5.42}$$

and relate the result to (5.63). You will need to use the fact that

$$\sum_{\{0,r,2r,...,N-r\}} e^{2\pi i k l/N} = \begin{cases} \sqrt{N/r} & \text{if } l \text{ is an integer multiple of } N/r \\ 0 & \text{otherwise.} \end{cases} \tag{5.43}$$

Solution:

Exercise 5.21 (Period finding and phase estimation)

Suppose you are given a unitary operator U_y which performs the transformation $U_y \left| f(x) \right\rangle = \left| f(x+y) \right\rangle$, for the periodic function described above.

1. Show that the eigenvectors of U_y are $\left| \hat{f}(l) \right\rangle$, and calculate their eigenvalues.

2. Show that given $|f(x_0)\rangle$ for some x_0, U_y can be used to realize a black box which is as useful as U in solving the period-finding problem.

Solution:

Exercise 5.22

Show that

$$\left| \hat{f}(l_1, l_2) \right\rangle = \sum_{x_1=0}^{r-1} \sum_{x_2=0}^{r-1} e^{-2\pi i (l_1 x_1 + l_2 x_2)} = \frac{1}{\sqrt{r}} \sum_{j=0}^{r-1} e^{-2\pi i l_2 j/r} \left| f(0, j) \right\rangle \tag{5.44}$$

and we are constrained to have $l_1/s - l_2$ be an integer multiple of r for this expression to be non-zero.

Solution:

Exercise 5.23

Compute

$$\frac{1}{r} \sum_{l_1=0}^{r-1} \sum_{l_2=0}^{r-1} e^{2\pi i (l_1 x_1 + l_2 x_2)/r} \left| \hat{f}(l_1, l_2) \right\rangle \tag{5.45}$$

using (5.70), and show that the result is $f(x_1, x_2)$.

Solution:

Exercise 5.24

Construct the generalized continued fractions algorithm needed in step **6** of the discrete logarithm algorithm to determine s from estimates of sl_2/r and l_2/r.

Solution:

Exercise 5.25

Construct a quantum circuit for the black box U used in the quantum discrete logarithm algorithm, which takes a and b as parameters, and performs the unitary transform $|x_1\rangle |x_2\rangle |y\rangle \rightarrow |x_1\rangle |x_2\rangle |y \oplus b^{x_1} a^{x_2}\rangle$. How many elementary operations are required?

Solution:

Exercise 5.26

Since K is a subgroup of G, when we decompose G into a product of cyclic groups of prime power order, this also decomposes K. Re-express (5.77) to show that determining l_i' allows one to sample from the corresponding cyclic subgroup K_{p_i} of K.

Solution:

Exercise 5.27

Of course, the decomposition of a general finite Abelian group G into a product of cyclic groups of prime power order is usually a difficult problem (at least as hard as factoring integers, for example). Here, quantum algorithms come to the rescue again: explain how the algorithms in this chapter can be used to efficiently decompose G as desired.

Solution:

Exercise 5.28

Write out a detailed specification of the quantum algorithm to solve the hidden subgroup problem, complete with runtime and success probability estimates, for finite Abelian groups.

Solution:

Exercise 5.29

Give quantum algorithms to solve the Deutsch and Simon problems listed in Figure 5.5, using the framework of the hidden subgroup problem.

Solution:

Problem 5.1

Construct a quantum circuit to perform the quantum Fourier transform

$$|j\rangle \rightarrow \frac{1}{\sqrt{p}} \sum_{k=0}^{p-1} e^{2\pi ijk/p} |k\rangle \tag{5.46}$$
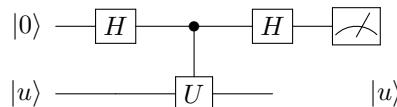
where p is prime.

Solution:

Problem 5.2 (Measured quantum Fourier transform)

Suppose the quantum Fourier transform is performed as the last step of a quantum computation, followed by a measurement in the computational basis. Show that the combination of quantum Fourier transform and measurement is equivalent to a circuit consisting entirely of *one* qubit gates and measurement, with classical control, and no two qubit gates. You may find the discussion of Section 4.4 useful.

Solution:

Problem 5.3 (Kitaev's algorithm)

Consider the quantum circuit

$$(5.47)$$

where $|u\rangle$ is an eigenstate of U with eigenvalue $e^{2\pi i\phi}$. Show that the top qubit is measured to be 0 with probability $p \equiv \cos^2(\pi\phi)$. Since the state $|u\rangle$ is unaffected by the circuit it may be reused; if U can be replaced by U^k, where k is an arbitrary integer under your control, show that by repeating this circuit and increasing k appropriately, you can efficiently obtain as many bits of p as desired, and thus, of ϕ. This is an alternative to the phase estimation algorithm.

Solution:

Problem 5.4

The runtime bound $O(L^3)$ we have given for the factoring algorithm is not tight. Show that a better upper bound of $O(L^2 \log L \log \log L)$ operations can be achieved.

Solution:

Problem 5.5 (Non-Abelian hidden subgroups- Research)

Let f be a function on a finite group G to an arbitrary finite range X, which is promised to be constant and distinct on distinct left cosets of a subgroup K. Start with the state

$$\frac{1}{\sqrt{|G|^m}} \sum_{g_1,...,g_m} |g_1,...,g_,\rangle |f(g_1),...,f(g_m)\rangle \tag{5.48}$$

and prove that picking $m = 4\log|G| + 2$ allows K to be identified with probability at least $1 - 1/|G|$. Note that G does not necessarily have to be Abelian, and being able to perform a Fourier transform over G is not required. This result shows that one can produce (using only $O(\log|G|)$ oracle calls) a final result in which the pure state outcomes corresponding to different possible hidden subgroups are nearly orthogonal. However, it is unknown whether a POVM exists or not which allows the hidden subgroup to be identified *efficiently* (i.e. using poly$(\log|G|)$ operations) from this final state.

Solution:

Problem 5.6 (Addition by Fourier transform)

Consider the task of constructing a quantum circuit to compute $|x\rangle \to |x + y \mod 2^n\rangle$, where y is a fixed constant, and $0 \le x < 2^n$. Show that one efficient way to do this, for values of y such as 1, is to first perform a quantum Fourier transform, then to apply single qubit phase shifts, then an inverse Fourier transform. What values of y can be added easily this way, and how many operations are required?

Solution:

Chapter 6

Quantum search algorithms

Exercise 6.1

Show that the unitary operator corresponding to the phase shift in the Grover iteration is $(2\left|0\right\rangle\left\langle0\right| - I)$.

Solution: For $\left|x\right\rangle = \left|0\right\rangle$,

$$(2\left|0\right\rangle\left\langle0\right| - I)\left|0\right\rangle = 2\left|0\right\rangle - \left|0\right\rangle = \left|0\right\rangle \tag{6.1}$$

Meanwhile, for $\left|x\right\rangle \neq \left|0\right\rangle$,

$$(2\left|0\right\rangle\left\langle0\right| - I)\left|x\right\rangle = 2\left|0\right\rangle\left\langle0|x\right\rangle - \left|x\right\rangle = -\left|x\right\rangle \tag{6.2}$$

Altogether,

$$(2\left|0\right\rangle\left\langle0\right| - I)\left|x\right\rangle = (-1)^{\delta_{0x}}\left|x\right\rangle \tag{6.3}$$

Exercise 6.2

Show that the operation $(2\left|\psi\right\rangle\left\langle\psi\right| - I)$ applied to a general state $\sum_k \alpha_k \left|k\right\rangle$ produces

$$\sum_k (-\alpha_k + 2\langle\alpha\rangle)\left|k\right\rangle \tag{6.4}$$

where $\langle\alpha\rangle \equiv \sum_k \alpha_k/N$ is the mean value of the α_k. For this reason, $(2\left|\psi\right\rangle\left\langle\psi\right| - I)$ is sometimes referred to as the *inversion about mean* operation.

Solution: By linearity,

$$(2\left|\psi\right\rangle\left\langle\psi\right| - I)\sum_k \alpha_k\left|k\right\rangle = \sum_k 2\alpha_k \left(\left|\psi\right\rangle\left\langle\psi\right|\right)\left|k\right\rangle - \sum_k \alpha_k\left|k\right\rangle$$
$$= 2\left|\psi\right\rangle\sum_k \alpha_k\left\langle\psi|k\right\rangle - \sum_k \alpha_k\left|k\right\rangle$$

Because $\left|\psi\right\rangle$ is uniform superposition over the computational basis states, for all k we have $\left\langle\psi|k\right\rangle = 1/\sqrt{N}$.

Hence,

$$(2 \left| \psi \right\rangle \left\langle \psi \right| - I) \sum_k \alpha_k \left| k \right\rangle = \frac{2}{\sqrt{N}} \left(\sum_k \alpha_k \right) \left| \psi \right\rangle - \sum_k \alpha_k \left| k \right\rangle \tag{6.5}$$

$$= 2\sqrt{N} \langle \alpha \rangle \left| \psi \right\rangle - \sum_k \alpha_k \left| k \right\rangle \tag{6.6}$$

Finally, we expand out the definition of $\left| \psi \right\rangle$ and cancel the factors of $\sqrt{N}$. This gives our result.

$$2\sqrt{N} \langle \alpha \rangle \left| \psi \right\rangle - \sum_k \alpha_k \left| k \right\rangle = \sum_k (2\langle \alpha \rangle - \alpha_k) \left| k \right\rangle \tag{6.7}$$

Exercise 6.3

Show that in the $\left| \alpha \right\rangle, \left| \beta \right\rangle$ basis, we may write the Grover iteration as

$$G = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \tag{6.8}$$

where θ is a real number in the range 0 to $\pi/2$ (assuming for simplicity that $M \leq N/2$; this limitation will be lifted shortly), chosen so that

$$\sin \theta = \frac{2\sqrt{M(N-M)}}{N} \tag{6.9}$$

Solution: As discussed in the text, both the oracle O and the reflection $2 \left| \psi \right\rangle \left\langle \psi \right| - I$ leave the subspace $V = span(\left| \alpha \right\rangle, \left| \beta \right\rangle)$ invariant. Hence, so does the product G. Therefore, we will from here on speak of G only in terms of its action on the 2-dimensional subspace $V = span(\left| \alpha \right\rangle, \left| \beta \right\rangle)$, and consider the matrix representation in the orthonormal basis $\{\left| \alpha \right\rangle, \left| \beta \right\rangle\}$. This representation is unitary (since G itself is), and in fact it is orthogonal, since both O and $(2 \left| \psi \right\rangle \left\langle \psi \right| - I)$ have real matrix elements. More specifically, G is *special* orthogonal, meaning it has determinant one, because it is the product of two reflections. All of this implies G is a proper rotation in the plane, and any such matrix may be parametrized as equation (6.8) for *some* angle θ. It remains to show θ satisfies relation (6.9). To do this, we simply compute the matrix element $\left\langle \beta \right| G \left| \alpha \right\rangle$.

$$\begin{aligned} \left\langle \beta \right| G \left| \alpha \right\rangle &= \left\langle \beta \right| (2 \left| \psi \right\rangle \left\langle \psi \right| - I) O \left| \alpha \right\rangle \\ &= \left\langle \beta \right| (2 \left| \psi \right\rangle \left\langle \psi \right| - I) \left| \alpha \right\rangle \\ &= \left\langle \beta \right| (2 \left| \psi \right\rangle \left\langle \psi | \alpha \right\rangle - \left| \alpha \right\rangle) \\ &= 2 \left\langle \beta | \psi \right\rangle \left\langle \psi | \alpha \right\rangle, \end{aligned} \tag{6.10}$$

where, along the way, we used the orthogonality of $\left| \alpha \right\rangle, \left| \beta \right\rangle$ and the fact that $O \left| \alpha \right\rangle = \left| \alpha \right\rangle$. Finally, using the expression given in the text for $\left| \psi \right\rangle$ expanded in the $\left| \alpha \right\rangle, \left| \beta \right\rangle$ basis, we arrive at our result.

$$\left\langle \beta \right| G \left| \alpha \right\rangle = \sin \theta = \frac{2\sqrt{M(N-M)}}{N} \tag{6.11}$$

Note that, in fact, we did not require the assumption that $M \leq N/2$ in our derivation.

Exercise 6.4

Give explicit steps for the quantum search algorithm, as above, but for the case of multiple solutions ($1 < M < N/2$).

Solution:

Algorithm: Quantum Search

Input: (1) a black box oracle O which performs the transformation $O\,|x\rangle\,|q\rangle = |x\rangle\,|q \oplus f(x)\rangle$, where f is a boolean function on n bits in which exactly M inputs are known to return 1, with the rest 0. (2) $n+1$ qubits in the state $|0\rangle$.

Output: One of the M solutions to the search problem, where a solution x is a n-bit string such that $f(x) = 1$.

Procedure:

1. $|0\rangle^{\otimes n}\,|0\rangle$ (initial state)

2. $\rightarrow \frac{1}{\sqrt{2^n}} \sum_{x=}^{2^n-1} |x\rangle \left(\frac{|0\rangle - |1\rangle}{\sqrt{2}} \right)$ (apply $H^{\otimes n}$ to the first n qubits and HX to the last).

3. $\rightarrow [(2\,|\psi\rangle\,\langle\psi| - I)O]^R \frac{1}{\sqrt{2^n}} \sum_{x=}^{2^n-1} |x\rangle \left(\frac{|0\rangle - |1\rangle}{\sqrt{2}} \right)$

   $\approx \frac{1}{\sqrt{2^M}} \sum_{s=1}^{M} |s\rangle \left[\frac{|0\rangle - |1\rangle}{\sqrt{2}} \right]$ (apply the Grover iterate $R \approx \lceil \frac{\pi}{4} \sqrt{\frac{2^n}{M}} \rceil$ times

4. $\rightarrow$ some s such that $f(s) = 1$ (measure the first n qubits)

Exercise 6.5

Show that the augmented oracle O' may be constructed using one application of O, and elementary quantum gates, using the extra qubit $|q\rangle$.

Solution: Notice that O' amounts to applying the oracle O to the first n qubits, conditioned on qubit q being in the 0 state. We can express this schematically as

 (6.12)

Thus, we are making a single access to the oracle O.

Exercise 6.6

Verify that the gates in the dotted box in the second figure of Box 6.1 perform the conditional phase shift operation $2\,|00\rangle\,\langle00| - I$, up to an unimportant global phase factor.

Solution: Because $HXH = Z$, the CNOT gate sandwiched between H gates is equivalent to a CZ gate. Moreover, $XZX = -Z$. Thus, the gates in the dotted box enact a controlled minus Z gate, where the control is taken to be the zero state. In braket form, if R is this set of gates,

$$R = |0\rangle\,\langle0| \otimes (-Z) + |1\rangle\,\langle1| \otimes I. \tag{6.13}$$

Let's consider the action of R on the computational basis. For states with a 1 on the top qubit, the state is unaffected. Otherwise, the top qubit is in the state 0, in which case a minus sign is applied only if the

bottom qubit is also zero. Thus, R applies a minus sign to $|00\rangle$ and leaves the other basis states unchanged. Hence,

$$R = I - 2 |00\rangle \langle 00| \tag{6.14}$$

which is the reflection operator up to an unimportant minus sign.

Exercise 6.7

Verify that the circuits shown in Figures 6.4 and 6.5 implement the operations $\exp(-i |x\rangle \langle x| \Delta t)$ and $\exp(-i |\psi\rangle \langle\psi| \Delta t)$, respectively, with $|\psi\rangle$ as in (6.24).

Solution:

Exercise 6.8

Suppose the simulation step is performed to an accuracy $O(\Delta t^r)$. Show that the number of oracle calls required to simulate H to reasonable accuracy is $O(N^{r/2(r-1)})$. Note that as r becomes large the exponent of N approaches $1/2$.

Solution: If the accuracy of a single step is $O(\Delta t^r)$, then the total error of the simulation is $O((t/\Delta t)\Delta t^r) = O(t\Delta t^{r-1})$, since $t/\Delta t$ is the number of steps in the simulation. For accurate results, we require this error to be $O(1)$, and from the previous section, we see that the time of simulation t is $O(\sqrt{N})$. Thus, we have $\Delta t \in O(1/N^{(r-1)/2})$.

Exercise 6.9

Verify Equation (6.25). (*Hint*: see Exercise 4.15.)

Solution:

Exercise 6.10

Show that by choosing Δt appropriately we can obtain a quantum search algorithm which uses $O(\sqrt{N})$ queries, and for which the final state is $|x\rangle$ *exactly*, that is, the algorithm works with probability 1, rather than with some smaller probability.

Solution:

Exercise 6.11 (Multiple solution continuous quantum search)

Guess a Hamiltonian with which one may solve the continuous time search problem in the case where the search problem has M solutions.

Solution:

Exercise 6.12 (Alternative Hamiltonian for quantum search)

Suppose

$$H = |x\rangle \langle \psi| + |\psi\rangle \langle x|. \tag{6.15}$$

(1) Show that it takes time $O(1)$ to rotate from the state $|\psi\rangle$ to the state $|x\rangle$, given an evolution according to the Hamiltoinian H.

(2) Explain how a quantum simulation of the Hamiltonian H may be performed, and determine the number of oracle calls your simulation technique requires to obtain the solution with high probability.

Solution:

Exercise 6.13

Consider a classical algorithm for the counting problem which samples uniformly and independently k times from the search space, and let $X_1, ..., X_k$ be the results of the oracle calls, that is, $X_j = 1$ if the jth oracle call revealed a solution to the problem, and $X_j=0$ if the jth oracle call did not reveal a solution to the problem. This algorithm returns the estimate $S = N \times \sum_j X_j/k$ for the number of solutions to the search problem. Show that the standard deviation in S is $\Delta S = \sqrt{M(N-M)/k}$. Prove that to obtain a probability at least $3/4$ of estimating M correctly to within an accuracy $\sqrt{M}$ for all values of M we must have $k = \Omega(N)$.

Solution:

Exercise 6.14

Prove that *any* classical counting algorithm with a probability at least $3/4$ for estimating M correctly to within an accuracy $c\sqrt{M}$ for some constant c and for all values of M must make $\Omega(N)$ oracle calls.

Solution:

Exercise 6.15

Use the Cauchy–Schwarz inequality to show that for any normalized state vector $|\psi\rangle$ and set of N orthonormal basis vectors $|x\rangle$,

$$\sum_x \|\psi - x\|^2 \geq 2N - 2\sqrt{N}. \tag{6.16}$$

Solution:

Exercise 6.16

Suppose we merely require that the probability of an error being made is less than $1/2$ when averaged uniformly over the possible values for x, instead of for all values of x. Show that $O(\sqrt{N})$ oracle calls are still

required to solve the search problem.

Solution:

Exercise 6.17 (Optimality for multiple solutions)

Suppose the search problem has M solutions. Show that $O(\sqrt{N/M})$ oracle applications are required to find a solution.

Solution:

Exercise 6.18

Prove that the minimum degree polynomial representing a Boolean function $F(X)$ is unique.

Solution:

Exercise 6.19

Show that $P(X) = 1 - (1 - X_0)(1 - X_1)...(1 - X_{N-1})$ represents OR.

Solution:

Exercise 6.20

Show that $Q_0(OR) \geq N$ by constructing a polynomial which represents the OR function from the output of a quantum circuit which computes OR with zero error.

Solution:

Problem 6.1 (Finding the minimum)

Suppose $x_1, ..., x_N$ is a database of numbers held in memory, as in Section 6.5. Show that only $O(\log(N)\sqrt{N})$ accesses to the memory are required on a quantum computer, in order to find the smallest element on the list, with probability at least one-half.

Solution:

Problem 6.2 (Generalized quantum searching)

Let $|\psi\rangle$ be a quantum state, and define $U_{|\psi\rangle} \equiv I - 2|\psi\rangle\langle\psi|$. That is, $U_{|\psi\rangle}$ gives the state $|\psi\rangle$ a -1 phase, and leaves the states orthogonal to $|\psi\rangle$ invariant.

(1) Suppose we have a quantum circuit implementing a unitary operator U such that $U|0\rangle^{\otimes n} = |\psi\rangle$. Explain how to implement $U_{|\psi\rangle}$.

(2) Let $|\psi_1\rangle = |1\rangle, |\psi_2\rangle = (|0\rangle - |1\rangle)/\sqrt{2}, |\psi_3\rangle = (|0\rangle - i|1\rangle)/\sqrt{2}$. Suppose an unknown oracle O is selected from the set $U_{|\psi_1\rangle}, U_{|\psi_2\rangle}, U_{|\psi_3\rangle}$. Give a quantum algorithm which identifies the oracle with just *one* application of the oracle. (*Hint*: consider superdense coding.)

(3) **Research:** More generally, given k states $|\psi_1\rangle, ..., |\psi_k\rangle$, and an unknown oracle selected from the set $U_{|\psi_1\rangle}, ..., U_{|\psi_k\rangle}$, how many oracle applications are required to identify the oracle, with high probability?

Solution:

Problem 6.3 (Database retrieval)

Given a quantum oracle which returns $|k, y \oplus X_k\rangle$ given an n qubit query (and one scratchpad qubit) $|k, y\rangle$, show that with high probability, all $N = 2^n$ bits of X can be obtained using only $N/2 + \sqrt{N}$ queries. This implies the general upper bound $Q_2(F) \le N/2 + \sqrt{N}$ for any F.

Solution:

Problem 6.4 (Quantum searching and cryptography)

Quantum searching can, potentially, be used to speed up the search for cryptographic keys. The idea is to search through the space of all possible keys for decryption, in each case trying the key, and checking to see whether the decrypted message makes 'sense'. Explain why this idea doesn't work for the Vernam cipher (Section 12.6). When might it work for cryptosystems such as DES? (For a description of DES see, for example, [MvOV96] or [Sch96a].)

Solution:

Chapter 7

Quantum computers: physical realization

Exercise 7.1

Using the fact that x and p do not commute, and that in fact $[x, p] = i\hbar$, explicitly show that $a^\dagger a = H/\hbar\omega - 1/2$.

Solution: Let us first apply the definition of a, which is given in the text.

$$
\begin{aligned}
a^\dagger a &= \frac{1}{2m\hbar\omega}(m\omega x - ip)(m\omega x + ip) \\
&= \frac{1}{2m\hbar\omega}\left((m\omega x)^2 + p^2 + im\omega(xp - px)\right)
\end{aligned}
\tag{7.1}
$$

Note that we cannot cancel the rightmost term since x and p do not commute. However, we can use the commutation relation between them. Regrouping and rearranging,

$$
\begin{aligned}
a^\dagger a &= \frac{p^2}{2m\hbar\omega} + \frac{1}{2\hbar\omega}m\omega^2 x^2 + \frac{i}{2\hbar}[x, p] \\
&= \frac{1}{\hbar\omega}\left(\frac{p^2}{2m} + \frac{1}{2}m\omega^2 x^2\right) - \frac{1}{2} \\
&= \frac{H}{\hbar\omega} - \frac{1}{2}.
\end{aligned}
\tag{7.2}
$$

In the last step, we identified the Hamiltonian of the harmonic oscillator. This completes the derivation.

Exercise 7.2

Chapter 8

Quantum noise and quantum operations

Exercise 8.1 (Unitary evolution as a quantum operation)

Pure states evolve under unitary transforms as $|\psi\rangle \to U|\psi\rangle$. Show that, equivalently, we may write $\rho \to \mathcal{E}(\rho) \equiv U\rho U^\dagger$, for $\rho = |\psi\rangle\langle\psi|$.

Solution: Consider the map f which identifies the ket representation of a pure quantum state, $|\psi\rangle$, with its representation as a rank one density operator.

$$f(|\psi\rangle) = |\psi\rangle\langle\psi| \tag{8.1}$$

Now suppose we have a unitary operation U on the quantum state, which in the ket representation acts as $|\psi\rangle \to U|\psi\rangle$. To arrive at a natural definition of the unitary transformation in the density operator formalism, we can consider what map $\mathcal{E}$ satisfies the following commutative property.

$$f(U|\psi\rangle) = \mathcal{E}(f(|\psi\rangle)) = \mathcal{E}(|\psi\rangle\langle\psi|) \tag{8.2}$$

Let $|\psi'\rangle = U|\psi\rangle$, so that $f(U|\psi\rangle) = |\psi'\rangle\langle\psi'|$. I claim that

$$|\psi'\rangle\langle\psi'| = U|\psi\rangle\langle\psi|U^\dagger. \tag{8.3}$$

To see this, we can show that the right side of this equation sends $|\psi'\rangle$ to itself and anything orthogonal to $|\psi'\rangle$ to the zero vector. Indeed,

$$U|\psi\rangle\langle\psi|U^\dagger|\psi'\rangle = U|\psi\rangle\langle\psi|\psi\rangle = \langle\psi|\psi\rangle U|\psi\rangle = |\psi'\rangle. \tag{8.4}$$

Similarly, suppose $|\psi'_\perp\rangle$ is orthogonal to $|\psi'\rangle$. Then, $|\psi'_\perp\rangle = U|\psi_\perp\rangle$ for some $|\psi_\perp\rangle$ orthogonal to $|\psi\rangle$. Hence,

$$U|\psi\rangle\langle\psi|U^\dagger|\psi'_\perp\rangle = U|\psi\rangle\langle\psi|\psi_\perp\rangle = 0. \tag{8.5}$$

Thus, we see that $|\psi'\rangle\langle\psi'| = U|\psi\rangle\langle\psi|U^\dagger$. This demonstrates that the natural definition of a unitary U acting on the density matrix representation of a quantum state is given by

$$\mathcal{E}(|\psi\rangle\langle\psi|) = U|\psi\rangle\langle\psi|U^\dagger, \tag{8.6}$$

which is precisely what we sought to show.

Exercise 8.2

Recall from Section 2.2.3 (on page 84) that a quantum measurement with outcomes labeled by m is described by a set of measurement operators M_m such that $\sum_m M_m^\dagger M_m = I$. Let the state of the system immediately before the measurement be ρ. Show that for $\mathcal{E}_m(\rho) \equiv M_m \rho M_m^\dagger$, the state of the system immediately after the measurement is

$$\frac{\mathcal{E}_m(\rho)}{\text{tr}(\mathcal{E}_m(\rho))}. \tag{8.7}$$

Also show that the probability of obtaining this measurement result is $p(m) = \text{tr}(\mathcal{E}_m(\rho))$.

Solution: For reference, the axioms for pure state measurements as provided in the text are as follows. A measurement with outcome m (an event in the probability space) has corresponding probability $p(m)$ and final state $|\psi'\rangle$ given by

$$p(m) = \langle\psi| M_m^\dagger M_m |\psi\rangle$$
$$|\psi'\rangle = \frac{M_m |\psi\rangle}{\sqrt{\langle\psi| M_m^\dagger M_m |\psi\rangle}}, \tag{8.8}$$

where $|\psi\rangle$ is the initial (pure) state of the system.

The results we are asked to "show" can be taken as axiomatic for open quantum systems. Unfortunately, because the map (8.7) is not convex linear, we cannot extend the axioms on pure states by linearity. The best we can do is show that (8.8) imply (8.7) in the case of pure states, then *assert* (i.e. define) the result to hold for any mixed state. This is what we do below.

First, consider the expression $p(m)$ in terms of ket $|\psi\rangle$. We have

$$\begin{aligned}
p(m) &= \langle\psi| M_m^\dagger M_m |\psi\rangle \\
&= \text{tr}\big(M_m^\dagger M_m |\psi\rangle \langle\psi|\big) \\
&= \text{tr}\big(M_m |\psi\rangle \langle\psi| M_m^\dagger\big) \\
&= \text{tr}(\mathcal{E}_m(\rho)).
\end{aligned} \tag{8.9}$$

where we used the property that the trace is cyclic and that, for any operator A and pure state,

$$\text{tr}(A |\psi\rangle \langle\psi|) = \langle\psi| A |\psi\rangle. \tag{8.10}$$

Now consider the pure state after measurement, which as a density operator is given by $|\psi'\rangle \langle\psi'|$. Using the transformation given by the second line of (8.8),

$$|\psi'\rangle \langle\psi'| = \frac{M_m |\psi\rangle \langle\psi| M_m^\dagger}{\langle\psi| M_m^\dagger M_m |\psi\rangle} = \frac{\mathcal{E}_m(|\psi\rangle \langle\psi|)}{\text{tr}(\mathcal{E}_m(|\psi\rangle \langle\psi|))}. \tag{8.11}$$

This shows the result for pure state density matrices that we desire. We can now extend this definition to all quantum states, mixed and pure.

Exercise 8.3

Our derivation of the operator-sum representation implicitly assumed that the input and output spaces for the operation were the same. Suppose a composite system AB initially in an unknown quantum state ρ initially in an unknown quantum state ρ is brought into contact with a composite system CD initially in some

standard state $|0\rangle$, and the two systems interact according to a unitary interaction U. After the interaction we discard systems A and D, leaving a state ρ' of the system BC. Show that the map $\mathcal{E}(\rho) = \rho'$ satisfies

$$\mathcal{E}(\rho) = \sum_k E_k \rho E_k^\dagger \tag{8.12}$$

for some set of linear operators E_k from the state space of the system AB to the state space of system BC, and such that $\sum_k E_k^\dagger E_k = I$.

Solution: The action of $\mathcal{E}$ from AB to BC is summarized as follows: append a system CD with state $|0\rangle$, apply a unitary U which acts on the combined system $ABCD$, and trace out system AD. More explicitly,

$$\mathcal{E}(\rho) = \mathrm{tr}_{AD}(U\rho \otimes |0\rangle \langle 0| U^\dagger). \tag{8.13}$$

Let $\{|u_k\rangle\}$ form an orthonormal basis for the space of system AD. Computing the trace with respect to this basis, we have

$$\begin{aligned}
\mathcal{E}(\rho) &= \sum_k \langle u_k| U\rho \otimes |0\rangle \langle 0| U^\dagger |u_k\rangle \\
&= \sum_k \langle u_k| U |0\rangle \rho \langle 0| U^\dagger |u_k\rangle
\end{aligned} \tag{8.14}$$

Note that we can freely move the ket $|0\rangle$ across ρ to match with U since ρ and $|0\rangle$ live in difference spaces. Consider the object E_k defined by

$$E_k \equiv \langle u_k| U |0\rangle. \tag{8.15}$$

This is, in fact, an operator from the space of AB to BC. From (8.14) we have

$$\mathcal{E}(\rho) = \sum_k E_k \rho E_k^\dagger \tag{8.16}$$

as desired. Moreover, we can check that the completeness relation is satisfied, as it must be for probability conservation.

$$\sum_k E_k^\dagger E_k = \sum_k \langle 0| U^\dagger |u_k\rangle |u_k\rangle U \langle 0| = \langle 0| U^\dagger \left(\sum_k |u_k\rangle \langle u_k|\right) U |0\rangle = \langle 0| U^\dagger U |0\rangle = I_{AB} \tag{8.17}$$

where the subscript emphasizes the space the identity is acting on.

Exercise 8.4 (Measurement)

Suppose we have a single qubit principal system, interacting with a single qubit environment through the transform

$$U = P_0 \otimes I + P_1 \otimes X \tag{8.18}$$

where X is the usual Pauli matrix (acting on the environment) and $P_0 \equiv |0\rangle \langle 0|$, $P_1 \equiv |1\rangle \langle 1|$ are projectors (acting on the system). Give the quantum operation for this process, in the operator-sum representation, assuming the environment starts in the state $|0\rangle$.

Solution: There will be two operation elements, E_0 and E_1, since the dimension of the environment is two. Moreover, per the preceding discussion these will be given by

$$E_j = \langle u_j| U |0\rangle \tag{8.19}$$

where the $|u_j\rangle$ form an orthonormal basis for the environment. Let's take these to be the computational basis. We then have

$$
\begin{aligned}
E_0 &= \langle 0| U |0\rangle = P_0 \otimes \langle 0| I |0\rangle + P_1 \otimes \langle 0| X |0\rangle = P_0 \\
E_1 &= \langle 1| U |0\rangle = P_0 \otimes \langle 1| I |0\rangle + P_1 \otimes \langle 1| X |0\rangle = P_1.
\end{aligned}
\tag{8.20}
$$

Thus, the quantum operation $\mathcal{E}$ can be expressed as

$$
\mathcal{E}(\rho) = P_0 \rho P_0 + P_1 \rho P_1
\tag{8.21}
$$

Note that this is equivalent to a measurement in the computational basis. It is straightforward to show that these operators satisfy the completeness relation.

Exercise 8.5 (Spin flips)

Just as in the previous exercise, but now let

$$
U = \frac{X}{\sqrt{2}} \otimes I + \frac{Y}{\sqrt{2}} \otimes X.
\tag{8.22}
$$

Give the quantum operation for this process, in the operator-sum representation.

Solution: We compute the Kraus operators as above, and get the following results.

$$
E_0 = \langle 0| U |0\rangle = \frac{X}{\sqrt{2}}
\tag{8.23}
$$

$$
E_1 = \langle 1| U |0\rangle = \frac{Y}{\sqrt{2}}
\tag{8.24}
$$

Note that the E_0 corresponds to a spin flip, while E_1 corresponds to a spin flip followed by phase flip. We can interpret each event occurring with probability $\frac{1}{2}$.

Exercise 8.6 (Composition of quantum operations)

Suppose $\mathcal{E}$ and $\mathcal{F}$ are quantum operations on the same quantum system. Show that the composition $\mathcal{F} \circ \mathcal{E}$ is a quantum operation, in the sense that it has an operator-sum representation. State and prove an extension of this result to the case where $\mathcal{E}$ and $\mathcal{F}$ do not necessarily have the same input and output spaces.

Solution: Let $\{E_i\}$ and $\{F_j\}$ be the operation elements for the channels $\mathcal{E}$ and $\mathcal{F}$ respectively. Considering the action of the composition $\mathcal{F} \circ \mathcal{E}$ on an arbitrary input state ρ.

$$
\begin{aligned}
\mathcal{F} \circ \mathcal{E}(\rho) &= \mathcal{F}(\mathcal{E}(\rho)) \\
&= \sum_j F_j \mathcal{E}(\rho) F_j^\dagger \\
&= \sum_j F_j \left(\sum_i E_i \rho E_i^\dagger \right) F_j^\dagger \\
&= \sum_{ji} F_j E_i \rho E_i^\dagger F_j^\dagger \\
&= \sum_{ji} G_{ji} \rho G_{ji}^\dagger
\end{aligned}
\tag{8.25}
$$

where $G_{ji} \equiv F_j E_i$ is the composition of the two operation elements. Note that this derivation holds as long as the output space of $\mathcal{E}$ matches the input space of $\mathcal{F}$, as is required for the composition to be well-defined anyways. One can check that the completeness relation

$$\sum_{ji} G_{ji}^\dagger G_{ji} \leq I \tag{8.26}$$

is satisfied.

Chapter 9

Distance measures for quantum information

Exercise 9.1

What is the trace distance between the probability distribution (1, 0) and the probability distribution (1/2, 1/2)? Between (1/2, 1/3, 1/6) and (3/4, 1/8, 1/8)?

Solution: Let's apply the definition in each case.

$$
\begin{aligned}
D((1,0),(1/2,1/2)) &= \frac{1}{2}(|1-1/2|+|0-1/2|) = \frac{1}{2} \\
D((1/2,1/3,1/6),(3/4,1/8,1/8)) &= \frac{1}{2}(|1/4|+|5/24|+|1/24|) = \frac{1}{4}
\end{aligned}
\tag{9.1}
$$

Exercise 9.2

Show that the trace distance between probability distributions $(p, 1-p)$ and $(q, 1-q)$ is $|p-q|$.

Solution: Again applying the definition,

$$
\begin{aligned}
D((q,1-q),(p,1-p)) &= \frac{1}{2}(|q-p|+|(1-q)-(1-p)|) \\
&= \frac{1}{2}(2|p-q|) \\
&= |p-q|
\end{aligned}
\tag{9.2}
$$

Exercise 9.3

What is the fidelity of the probability distributions (1, 0) and (1/2, 1/2)? Of (1/2, 1/3, 1/6) and (3/4, 1/8, 1/8)?

Solution: Let's once again apply the definition of fidelity to these distributions.

$$F((1,0),(1/2,1/2)) = 1 \times \sqrt{1/2} + 0 = \frac{1}{\sqrt{2}}$$

$$F((1/2,1/3,1/6),(3/4,1/8,1/8)) = \sqrt{1/2 \times 3/4} + \sqrt{1/3 \times 1/8} + \sqrt{1/6 \times 1/8} \qquad (9.3)$$

$$= \frac{1}{2}(\sqrt{3/2} + 1/\sqrt{6} + 1/(2\sqrt{3})) \qquad \approx 0.96$$

Exercise 9.4

Prove (9.3).

Solution: Let X be the index set for the outcomes x. Note that we may write

$$\frac{1}{2}\sum_{x \in X}|p_x - q_x| = \frac{1}{2}\sum_{x \in P}(p_x - q_x) + \frac{1}{2}\sum_{x \notin P}(q_x - p_x) \qquad (9.4)$$

where $P = \{x \in X | p_x \geq q_x\}$. This amounts to breaking up the absolute value depending on the sign of $p_x - q_x$. I claim that both sums on the right side of the equation are, in fact, equal. This follows from normalization of probabilities, which in this case can be cast as

$$\sum_{x \notin P} p_x = 1 - \sum_{x \in P} p_x \qquad (9.5)$$

and similar for q_x. Thus,

$$\sum_{x \notin P}(q_x - p_x) = (1 - \sum_{x \in P} q_x) - (1 - \sum_{x \in P} p_x) = \sum_{x \in P}(p_x - q_x). \qquad (9.6)$$

Applying this relationship to equation (9.4) gives

$$D(p_x, q_x) = \frac{1}{2}\sum_{x \in X}|p_x - q_x| = \sum_{x \in P} p_x - q_x. \qquad (9.7)$$

We could have just as well summed over $Q = X \setminus P$ and gotten the same result, provided we summed over $q_x - p_x$ instead.

Let's now argue that P is the set that maximizes the sum, that is,

$$\sum_{x \in P}(p_x - q_x) = \max_{S \subset X}\sum_{x \in S}(p_x - q_x). \qquad (9.8)$$

To see this, note that P contains every element x such that $p_x - q_x$ is positive (or zero, though this doesn't matter in terms of the sum), and doesn't contain elements for which it is negative. Thus, elements of P only increase the sum, never decrease. And P contains all such positive elements. This means that equation (9.8) holds. Since the left hand side equals the trace distance, we have our result. The only thing that remains is to argue that the equality remains unchanged when we include the absolute value sign; this will be done in the subsequent exercise.

Exercise 9.5

Show that the absolute value signs may be removed from Equation (9.3), that is,

$$D(p_x, q_x) = \max_s(p(S) - q(S)) = \max_S\left(\sum_{x \in S} p_x - \sum_{x \in S} q_x\right). \qquad (9.9)$$

Solution: To maximize the absolute value, we can either sum over all the x where $p_x - q_x$ is positive, or those in which it is negative, and take the maximum of these two with absolute value. However, because of the normalization of probability to one, these two are equal anyways, as I argue in the previous exercise as an intermediate step. Hence, we can always just choose the positive sum, so the absolute value sign is not necessary.

Exercise 9.6

What is the trace distrance between the density operators

$$\frac{3}{4}\,|0\rangle\,\langle 0| + \frac{1}{4}\,|1\rangle\,\langle 1|\,;\quad \frac{2}{3}\,|0\rangle\,\langle 0| + \frac{1}{3}\,|1\rangle\,\langle 1|? \tag{9.10}$$

Between:

$$\frac{3}{4}\,|0\rangle\,\langle 0| + \frac{1}{4}\,|1\rangle\,\langle 1|\,;\quad \frac{2}{3}\,|+\rangle\,\langle +| + \frac{1}{3}\,|-\rangle\,\langle -|? \tag{9.11}$$

(Recall that $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$.)

Solution: (Example 1) This example is relatively simple, since both operators are already diagonal in the computational basis. We basically just need the difference, absolute valued, of the diagonal elements. Here is the full computation.

$$\begin{aligned} &\frac{1}{2}\left(\left|\frac{3}{4} - \frac{2}{3}\right| + \left|\frac{1}{4} - \frac{1}{3}\right|\right) \\ &= \frac{1}{2}\left(\frac{1}{12} + \frac{1}{12}\right) \\ &= \frac{1}{12} \end{aligned} \tag{9.12}$$

(Example 2) This one is a bit trickier, because the two operators under consideration do not commute. To proceed, we will calculate the eigenvalues of the difference. Let ρ and σ denote the first and second density operators, respectively. We will diagonalize the difference $\rho - \sigma$, and it will be helpful to express everything as a matrix in the computational basis. These are given by

$$\rho = \begin{pmatrix} 3/4 & 0 \\ 0 & 1/4 \end{pmatrix} \quad \sigma = \begin{pmatrix} 1/2 & 1/6 \\ 1/6 & 1/2 \end{pmatrix}. \tag{9.13}$$

It's a good idea to do the computation for σ yourself. Note that the matrix forms are symmetric and have trace 1, as expected. They are also positive definite. None of this should be surprising, of course. The difference $\rho - \sigma$ is then given by

$$\rho - \sigma = \begin{pmatrix} 1/4 & -1/6 \\ -1/6 & -1/4 \end{pmatrix}. \tag{9.14}$$

Now we compute the eigenvalues. Do this using your favorite method. The two eigenvalues, λ_+, λ_- are given by

$$\lambda_\pm = \pm\frac{\sqrt{13}}{12} \tag{9.15}$$

Thus, taking the sum of the absolute values, and dividing by two, we get the trace distance: $\sqrt{13}/12$. Again, it's a good idea to check this.

Exercise 9.7

Show that for any states ρ and σ, one may write $\rho - \sigma = Q - S$, where Q and S are positive operators with support on orthogonal vector spaces. (Hint: use the spectral decomposition $\rho - \sigma = UDU^\dagger$, and split the diagonal matrix D into positive and negative parts. This fact will continue to be useful later.)

Solution: Though we are proving this for the specific case $\rho - \sigma$, with two quantum states, this is really quite more general. All we require is the difference $\rho - \sigma$ be Hermitian, which it certainly is in this case. Any Hermitian operator can be diagonalized, and one way to understand this is to say that $\rho - \sigma$ may be expressed as a combination of orthogonal projectors.

$$\rho - \sigma = \sum_j \lambda_j \, |\lambda_j\rangle \, \langle\lambda_j| \tag{9.16}$$

Here, the sum is taken over all (nonzero) eigenvalues λ_j, with corresponding eigenvector $|\lambda_j\rangle$ (in case of multiplicity, λ_j may be equal to λ_k for $j \neq k$. However, in this case we can make a choice so that $|\lambda_j\rangle$ and $|\lambda_k\rangle$ are orthogonal. The notation is tricky but useful). Let us reorganize the sum into two pieces, one in which the eigenvalues are positive, and one containing the remaining, negative terms. Call these terms Q and T, respectively. Clearly, Q is positive; its eigenvalues are either zero or λ_j for positive λ_j. Moreover, T is clearly negative, since all it's eigenvalues are negative or zero. Thus, $S = -T$ is positive. Moreover, by construction, we have

$$\rho - \sigma = Q + T = Q - S \tag{9.17}$$

The fact that Q and S act on orthogonal subspaces is also a consequence of the hermiticity of $\rho - \sigma$, which implies that all the eigenvectors in equation (9.16) are mutually orthogonal.

Exercise 9.8 (Convexity of the trace distance)

Show that the trace distance is convex in its first input,

$$D\left(\sum_i p_i \rho_i, \sigma\right) \leq \sum_i p_i D(\rho_i, \sigma). \tag{9.18}$$

By symmetry convexity in the second entry follows from convexity in the first.

Solution: We would like to somehow take advantage of the joint convexity result, which seems like it should directly imply the result we seek. To make it work in our favor, we can employ a trick: write $\sigma = \sum_i p_i \sigma_i$, where $\sigma_i = \sigma$ for all i. Now, σ is written so that we can apply joint convexity. As a final step, simply replace $\sigma_i = \sigma$ in the resulting formula. This will prove the result.

Exercise 9.9 (Existence of fixed points)

Schauder's fixed point theorem is a classic result from mathematics that implies that any continuous map on a convex, compact subset of a Hilbert space has a fixed point. Use Schauder's fixed point theorem to prove that any trace-preserving quantum operation $\mathcal{E}$ has a fixed point, that is, ρ such that $\mathcal{E}(\rho) = \rho$.

Solution: We simply need to show that a trace-preserving quantum operation $\mathcal{E}$ satisfies the criteria of the theorem. First, note that the collection of self-adjoint operators on a complex Hilbert space is *itself* a real Hilbert space, with inner product defined by $\langle A, B \rangle \equiv \text{tr}(AB)$. Quantum operations are maps on

the set of density operators, which are a subset of this Hilbert space of operators. And we know this map is continuous, since, after all, the map $\mathcal{E}$ is contractive with respect to trace distance, and therefore keeps quantum states close together (this could probably be demonstrated more rigorously). Therefore, we need to check that this set $\mathcal{D}$ of density operators is convex and compact. We already know $\mathcal{D}$ is convex: any convex combination of states ρ and σ,

$$\lambda\rho + (1 - \lambda)\sigma, \tag{9.19}$$

is itself a valid state. To prove compactness, we will take advantage of the so-called *Heine-Borel theorem*, which says that subset of a space which is the "same as" an n-dimensional real space, $\mathbb{R}^n$ is compact if and only it is closed and bounded. Bound just means the set doesn't extend indefinitely. Closed means there is some boundary that lies within the set. Both these properties can be understood, in our circumstances, as a consequence of the existence of pure states, which are *extremal points* of $\mathcal{D}$. This means pure states form the boundary for density operators, and of course this boundary is included in $\mathcal{D}$. All of this can be proved with greater rigor, but hopefully this gives a starting point for understanding the result. With these properties in hand, we can use the Schauder theorem to assert a map $\mathcal{E}$ on $\mathcal{D}$ has at least one fixed point $\rho \in \mathcal{D}$.

Exercise 9.10

Suppose E is a *strictly contractive* trace-preserving quantum operation, that is, for any ρ and σ, $D(\mathcal{E}(\rho), \mathcal{E}(\sigma)) < D(\rho, \sigma)$. Show that E has a unique fixed point.

Solution: From the previous exercise, we know that $\mathcal{E}$ has at least one fixed point. Let ρ and σ be fixed points, possibly equal. We will show that, in fact, it must be that $\rho = \sigma$. By definition of fixed point, the states are invariant under the map: $\mathcal{E}(\rho) = \rho$ and $\mathcal{E}(\sigma) = \sigma$. Therefore, $D(\mathcal{E}(\rho), \mathcal{E}(\sigma)) = D(\rho, \sigma)$. On the other hand, because D is strictly contractive, $D(\mathcal{E}(\rho), \mathcal{E}(\sigma)) \leq D(\rho, \sigma)$, with equality if and only if $\rho = \sigma$. Since we indeed have equality, it must be that $\rho = \sigma$. Thus, every fixed point of $\mathcal{E}$ is the same, and therefore unique.

Exercise 9.11

Suppose $\mathcal{E}$ is a trace-preserving quantum operation for which there exists a density operator ρ_0 and a trace-preserving quantum operation $\mathcal{E}'$ such that

$$\mathcal{E}(\rho) = p\rho_0 + (1 - p)\mathcal{E}'(\rho) \tag{9.20}$$

for some $p, 0 < p \leq 1$. Physically, this means that with probability p the input state is thrown out and replaced with the fixed state ρ_0, while with probability $1 - p$ the operation $\mathcal{E}'$ occurs. Use joint convexity to show that $\mathcal{E}$ is a strictly contractive quantum operation, and thus has a unique fixed point.

Solution: Let ρ and σ be distinct quantum states, and consider the distance between the states after application of the operation $\mathcal{E}$.

$$D(\mathcal{E}(\rho), \mathcal{E}(\sigma)) = D(p\rho_0 + (1 - p)\mathcal{E}'(\rho), p\rho_0 + (1 - p)\mathcal{E}'(\sigma)) \tag{9.21}$$

Using the joint convexity of the trace distance,

$$\begin{aligned}
D(\mathcal{E}(\rho), \mathcal{E}(\sigma)) &\leq pD(\rho_0, \rho_0) + (1 - p)D(\mathcal{E}'(\rho)\mathcal{E}'(\sigma)) \\
&= (1 - p)D(\mathcal{E}'(\rho), \mathcal{E}'(\sigma)) \\
&< D(\mathcal{E}'(\rho), \mathcal{E}'(\sigma)) \\
&\leq D(\rho, \sigma)
\end{aligned} \tag{9.22}$$

In the process we used the fact that $1 - p < 1$ and that any quantum channel, particularly $\mathcal{E}'$, is contractive. From this string of inequalities, one of which is strict, we can conclude that $D(\mathcal{E}(\rho), \mathcal{E}(\sigma)) < D(\rho, \sigma)$, hence the map is strictly contractive. By the results of the previous exercise, it also has a unique fixed point.

Exercise 9.12

Consider the depolarizing channel introduced in Section 8.3.4 on page 378, $\mathcal{E}(\rho) = pI/2 + (1 - p)\rho$. For arbitrary ρ and σ find $D(\mathcal{E}(\rho), \mathcal{E}(\sigma))$ using the Bloch representation, and prove explicitly that the map $\mathcal{E}$ is strictly contractive, that is, $D(\mathcal{E}(\rho), \mathcal{E}(\sigma)) < D(\rho, \sigma)$.

Solution: Note that the depolarizing channel meets the criteria of the map $\mathcal{E}$ of the previous exercise, and therefore we could conclude right away that the map is strictly contractive. However, as the problem suggests, we will show this explicitly in the Bloch representation, which will serve as a nice illustration. Recall that any qubit state ρ can be parametrized as

$$\rho = \frac{1}{2}(I + \vec{r} \cdot \vec{\sigma}) \tag{9.23}$$

where $\|\vec{r}\| \leq 1$. The 3-vector $\vec{r}$ gives the location of the state in the Bloch ball. The action of the depolarizing channel $\mathcal{E}$ can be expressed solely in terms of changing $\vec{r}$, and we can compute this action.

$$\begin{aligned}
\mathcal{E}(\rho) &= pI/2 + (1 - p)\frac{1}{2}(I + \vec{r} \cdot \vec{\sigma}) \\
&= I/2 + \frac{1 - p}{2}\vec{r} \cdot \vec{\sigma} \\
&= \frac{1}{2}(I + \vec{r}' \cdot \vec{\sigma})
\end{aligned} \tag{9.24}$$

where $\vec{r}' = (1 - p)\vec{r}$. Hence, the depolarizing channel shrinks any Bloch vector by the factor $1 - p < 1$. From this geometric, picture, it should be clear that this map is strictly contractive: all quantum states get closer together after application of the map. Indeed, the trace distance for qubits is directly proportional to the Euclidean distance of the corresponding Bloch vectors. This proves our result.

Exercise 9.13

Show that the bit flip channel (Section 8.3.3) is contractive but not strictly contractive. Find the set of fixed points for the bit flip channel.

Solution: We showed in Exercise 9.10 that all strictly contractive quantum operations have unique fixed points. Therefore, by the contrapositive, if an operation *does* have multiple fixed points, it cannot be strictly contractive. Since every quantum operation is contractive by Theorem 9.2, we simply need to show there are multiple fixed points. Recall that the bit flip channel $\mathcal{E}$ is defined by

$$\mathcal{E}(\rho) = pX\rho X + (1 - p)\rho. \tag{9.25}$$

Let us find the set of points ρ such that $\mathcal{E}(\rho) = \rho$. Applying the definition and simplifying,

$$pX\rho X + (1 - p)\rho = \rho \quad \Leftrightarrow \quad X\rho X = \rho. \tag{9.26}$$

In turn, this is equivalent to

$$[\rho, X] = 0. \tag{9.27}$$

Hence, any state which commutes with pauli X is unchanged by the channel. This makes sense intuitively, since these states are eigenstates of the pauli X operator. These states can be understood as the set of all Bloch vectors lying along the x-axis. Clearly there are many (in fact, infinite) such states. Therefore, $\mathcal{E}$ is not strictly contractive.

Exercise 9.14 (Invariance of fidelity under unitary transformations)

Prove (9.61) by using the fact that for any positive operator A, $\sqrt{UAU^\dagger} = U\sqrt{A}U^\dagger$.

Solution: Let $\rho' = U\rho U^\dagger$ and similarly for σ. Making use of the square root property given in the problem, we have

$$\rho'^{1/2}\sigma'\rho'^{1/2} = (U\rho^{1/2}U^\dagger)(U\sigma U^\dagger)(U\rho^{1/2}U^\dagger) = U\rho^{1/2}\sigma\rho^{1/2}U^\dagger. \tag{9.28}$$

This is the type of isomorphism property we expect matrices to respect under multiplication. Because of this, we have

$$\begin{aligned} F(\rho',\sigma') &= \text{tr}\,\sqrt{U\rho^{1/2}\sigma\rho^{1/2}U^\dagger} \\ &= \text{tr}\,U\sqrt{\rho^{1/2}\sigma\rho^{1/2}}U^\dagger \\ &= \text{tr}\,\sqrt{\rho^{1/2}\sigma\rho^{1/2}} \\ &= F(\rho,\sigma). \end{aligned} \tag{9.29}$$

Along the way, we again used the square root property for positive operators, as well as the cyclic property of the trace. This completes the proof.

Exercise 9.15

Show that

$$F(\rho,\sigma) = \max_{|\varphi\rangle} |\langle\psi|\varphi\rangle| \tag{9.30}$$

where $|\psi\rangle$ is any fixed purification of ρ, and the maximization is over all purifications of σ.

Solution: Let's return to an earlier line of the derivation for the full proof of Uhlmann's theorem. At equation (9.70) of the text, we showed that for any purifications $|\psi\rangle$ and $|\varphi\rangle$ of ρ and σ respectively,

$$|\langle\psi|\varphi\rangle| = \left|\text{tr}\big(\sqrt{\rho}\sqrt{\sigma}U\big)\right|, \tag{9.31}$$

where $U = V_Q V_R^\dagger U_R U_Q^\dagger$. As in the original derivation, let V be the unitary part of the polar decomposition of $\sqrt{\rho}\sqrt{\sigma}$.

$$\sqrt{\rho}\sqrt{\sigma} = \left|\sqrt{\rho}\sqrt{\sigma}\right|V. \tag{9.32}$$

As before, we need to set $VU = I$ in order to achieve the maximum value and match the fidelity. But this time we can only change V_Q and V_R, not U_Q and U_R (which are fixed by our choice of purification of $|\psi\rangle$). Thankfully, we can still accomplish this. Set $V_R = U_R$ and $V_Q = V^\dagger U_Q$. Then,

$$VU = VV_Q V_R^\dagger U_R U_Q^\dagger = VV^\dagger U_Q U_R^\dagger U_R U_Q^\dagger = I. \tag{9.33}$$

This completes the proof: we only need to maximize one of the two purifications.

Exercise 9.16

Suppose R and Q are two quantum systems with the same Hilbert space. Let $|i_R\rangle$ and $|i_Q\rangle$ be orthonormal basis sets for R and Q. Let A be an operator on R and B an operator on Q. Define $|m\rangle \equiv \sum_i |i_R\rangle |i_Q\rangle$. Show that

$$\operatorname{tr}(A^\dagger B) = \langle m| A \otimes B |m\rangle \tag{9.34}$$

where the multiplication on the left hand side is of matrices, and it is understood that the matrix elements of A are taken with respect to the basis $|i_R\rangle$ and those for B with respect to the basis $|i_Q\rangle$.

Solution: Let's explicitly evaluate the right side of the equation. We have

$$
\begin{aligned}
\langle m| A \otimes B |m\rangle &= \sum_i \langle i_R| \langle i_Q| A \otimes B \sum_j |j_R\rangle |j_Q\rangle \\
&= \sum_{ij} \langle i_R| A |j_R\rangle \langle i_Q| B |j_Q\rangle \\
&= \sum_{ij} A_{ij} B_{ij}
\end{aligned}
\tag{9.35}
$$

where $A_{ij} \equiv \langle i_R| A |j_R\rangle$ are the matrix elements with respect to the basis, as are B_{ij}.

This is quite an interesting result, because we can view $|m\rangle$ as a maximally entangled ("Bell") state. Apparently, the expectation value of $A \otimes B$ with respect to this entangled state computes the inner product between A and B.

Exercise 9.17

Show that $0 \le A(\rho, \sigma) \le \pi/2$, with equality in the first inequality if and only if $\rho = \sigma$.

Solution: To define the inverse of cos, which is arccos, we must choose a particular interval of angles θ on which $\cos(\theta)$ is invertible. A reasonable (and standard) choice is to take angles $\theta \in [0, \pi]$. With this choice, the arccos function has a domain $[-1, 1]$ (corresponding to the range of cosine) and a range from $[0, \pi]$.

In our present case, we are looking at arccos $F(\rho, \sigma)$, with $F(\rho, \sigma) \in [0, 1]$. The question is, for x between 0 and 1, what is $\arccos(x)$? To determine this, we note that, for the angles $\theta \in [0, \pi]$ that we are interested in, $\cos\theta$ is nonnegative only in the first quadrant (including the axes themselves). These angles are precisely the interval $[0, \pi/2]$. This implies that the image of the interval arccos$([0, 1])$ of the unit inverval under $arccos$ is $[0, \pi/2]$. Hence $A(\rho, \sigma) \in [0, \pi/2]$.

When does $A(\rho, \sigma) = 0$? Exactly when $\arccos(x) = 0$, where $x = F(\rho, \sigma)$. This occurs when $x = \cos\theta = 1$, meaning $F(\rho, \sigma) = 1$. We've already seen that this is equivalent to $\rho = \sigma$. Hence, $A(\rho, \sigma) = 0$ exactly when $\rho = \sigma$.

Exercise 9.18 (Contractivity of the angle)

Let $\mathcal{E}$ be a trace-preserving quantum operation. Show that

$$A(\mathcal{E}(\rho), \mathcal{E}(\sigma)) \le A(\rho, \sigma). \tag{9.36}$$

Solution: This follows directly from the fact that F is monotonic, while arccos is monotonically decreasing on $[0, 1]$ (with range $[0, \pi/2]$). In more detail we have that $\arccos(y) \leq \arccos(x)$ whenever $y \geq x$. In our case we have

$$F(\mathcal{E}(\rho), \mathcal{E}(\sigma)) \geq F(\rho, \sigma). \tag{9.37}$$

Hence,

$$\arccos F(\mathcal{E}(\rho), \mathcal{E}(\sigma)) \leq \arccos F(\rho, \sigma). \tag{9.38}$$

which implies the result by definition of A.

Exercise 9.19 (Joint concavity of fidelity)

Prove that the fidelity is *jointly concave*,

$$F\left(\sum_i p_i \rho_i, \sum_i, p_i \sigma_i\right) \geq \sum_i p_i F(\rho_i, \sigma_i). \tag{9.39}$$

Solution: This follows directly from strong concavity, as given in Theorem 9.7 of the text. In particular, let's set $q_i = p_i$ for all i in the index set. By the theorem, we have

$$F\left(\sum_i p_i \rho_i, \sum_i, p_i \sigma_i\right) \geq \sum_i \sqrt{p_i p_i} F(\rho_i, \sigma_i) = \sum_i p_i F(\rho_i, \sigma_i). \tag{9.40}$$

Exercise 9.20 (Concavity of fidelity)

Prove that the fidelity is concave in the first entry,

$$F\left(\sum_i p_i \rho_i, \sigma\right) \geq \sum_i p_i F(\rho_i, \sigma). \tag{9.41}$$

By symmetry the fidelity is also concavein the second entry.

Solution: This follows from joint concavity of the previous exercise (which in turn followed from strong concavity). In particular, set $\sigma_i = \sigma$ for all i in the index set. Then,

$$F\left(\sum_i p_i \rho_i, \sum_i p_i \sigma\right) = F\left(\sum_i p_i \rho_i, \sigma \sum_i p_i\right) = F\left(\sum_i p_i \rho_i, \sigma\right). \tag{9.42}$$

On the other hand, joint concavity says

$$F\left(\sum_i p_i \rho_i, \sum_i p_i \sigma_i\right) \geq \sum_i p_i F(\rho_i, \sigma_i) = \sum_i p_i F(\rho_i, \sigma). \tag{9.43}$$

Comparing lines (9.42) and (9.43), we obtain our desired inequality.

Exercise 9.21

When comparing pure states and mixed states it is possible to make a stronger statement than (9.110) about the relationship between trace distance and fidelity. Prove that

$$1 - F(|\psi\rangle, \sigma)^2 \leq D(|\psi\rangle, \sigma) \tag{9.44}$$

Solution: Recall that the fidelity takes a simpler form when comparing pure and mixed states. In particular,

$$F(|\psi\rangle, \sigma) = \sqrt{\langle\psi| \sigma |\psi\rangle}. \tag{9.45}$$

Hence, F^2 is the expectation value of σ with respect to $|\psi\rangle$. To prove our result, it remains to show that

$$D(|\psi\rangle, \sigma) \geq 1 - \langle\psi| \sigma |\psi\rangle. \tag{9.46}$$

To do this, recall that D can be expressed as a maximization over expectations with projectors P.

$$D(|\psi\rangle, \sigma) = \max_P \operatorname{tr}(P(|\psi\rangle\langle\psi| - \sigma)) \tag{9.47}$$

One particular projector is $P = |\psi\rangle\langle\psi|$. Since the maximum is greater, we have

$$
\begin{aligned}
D(|\psi\rangle, \sigma) &\geq \operatorname{tr}(|\psi\rangle\langle\psi| (|\psi\rangle\langle\psi| - \sigma)) \\
&= \operatorname{tr}(|\psi\rangle\langle\psi| - |\psi\rangle\langle\psi| \sigma) \\
&= \operatorname{tr}(|\psi\rangle\langle\psi|) - \operatorname{tr}(|\psi\rangle\langle\psi| \sigma) \\
&= 1 - \langle\psi| \sigma |\psi\rangle.
\end{aligned} \tag{9.48}
$$

This proves our result.

Exercise 9.22 (Chaining property for fidelity measures)

Suppose U, and V are unitary operators, and $\mathcal{E}$ and $\mathcal{F}$ are trace-preserving quantum operations meant to approximate U and V. Letting $d(\cdot, \cdot)$ be any metric on the space of density matrices satisfying $d(U\rho U^\dagger, U\sigma U^\dagger) = d(\rho, \sigma)$ for all density matrices ρ and σ and unitary U (such as the angle $\arccos(F(\rho, \sigma))$), define the corresponding *error* $E(U, \mathcal{E})$ by

$$E(U, \mathcal{E}) \equiv \max_\rho d(U\rho U^\dagger, \mathcal{E}(\rho)), \tag{9.49}$$

and show that $E(VU, \mathcal{F} \circ \mathcal{E} \leq E(U, \mathcal{E}) + E(V, \mathcal{F})$. Thus, to perform a quantum computation with high fidelity it suffices to complete each step of the computation with high fidelity.

Solution: We will make use of the fact that any metric satisfies the *triangle inequality*. For any points $\rho, \sigma, d(\rho, \sigma)$ satisfies

$$d(\rho, \sigma) \leq d(\rho, \eta) + d(\eta, \sigma) \tag{9.50}$$

for any other point η. In our case, the "points" are density matrices. The way we will use this property is to break up the error $E(VU, \mathcal{F} \circ \mathcal{E})$ into two pieces: the error due to V and the error due to $\mathcal{E}$ *followed by a perfect implementation of V*. Let's see how this works.

By the triangle inequality, we have

$$d(VU\rho(VU)^\dagger, \mathcal{F} \circ \mathcal{E}(\rho)) \leq d(VU\rho(VU)^\dagger, V\mathcal{E}(\rho)V^\dagger) + d(V\mathcal{E}(\rho)V^\dagger, \mathcal{F} \circ \mathcal{E}(\rho)). \tag{9.51}$$

By the "twirling" property of d with respect to unitaries as described in the problem, the first term on the right hand side above can be simplified, giving

$$d(VU\rho(VU)^\dagger, \mathcal{F} \circ \mathcal{E}(\rho)) \leq d(U\rho U^\dagger, \mathcal{E}(\rho)) + d(V\mathcal{E}(\rho)V^\dagger, \mathcal{F} \circ \mathcal{E}(\rho)). \tag{9.52}$$

Let's now take the maximum of both sides of (9.52) with respect to ρ. After doing so, we can simplify the right hand side by noting that, schematically,

$$\max(A + B) \leq \max A + \max B. \tag{9.53}$$

Hence,

$$\max_\rho d(VU\rho(VU)^\dagger, \mathcal{F} \circ \mathcal{E}(\rho)) \equiv E(VU, \mathcal{F} \circ \mathcal{E}) \leq E(U, \mathcal{E}) + \max_\rho d(V\mathcal{E}(\rho)V^\dagger, \mathcal{F} \circ \mathcal{E}(\rho)). \tag{9.54}$$

We are almost there. Let's turn our attention to the last term in the inequality above.

$$\max_\rho d(V\mathcal{E}(\rho)V^\dagger, \mathcal{F} \circ \mathcal{E}(\rho)) \tag{9.55}$$

What we want to observe is that we are maximizing over ρ, but everything is in terms of $\mathcal{E}(\rho)$. The set of all $\mathcal{E}(\rho)$ for some input density matrix ρ is a subset of all inputs σ to the quantum operation $\mathcal{F}$. If we maximize over a set, the result is larger than maximizing over a subset. This implies

$$\max_\rho d(V\mathcal{E}(\rho)V^\dagger, \mathcal{F} \circ \mathcal{E}(\rho)) \leq \max_\sigma d(V\sigma V^\dagger, \mathcal{F}(\sigma)) \equiv E(V, \mathcal{F}). \tag{9.56}$$

Combining inequality (9.56) with (9.54), we arrive at the desired result.

$$E(VU, \mathcal{F} \circ \mathcal{E}) \leq E(U, \mathcal{E}) + E(V, \mathcal{F}) \tag{9.57}$$

Exercise 9.23

Show that $\bar{F} = 1$ if and only if $\mathcal{E}(\rho_j) = \rho_j$ for all j such that $p_j > 0$.

Solution: If $\mathcal{E}(\rho_j) = \rho_j$ for all ρ_j (that actually "show up", meaning $p_j > 0$). Then the fidelity between these states is 1. Hence, it is unsurprising that the *average* (squared) fidelity is also one, as this calculation demonstrates.

$$\bar{F} = \sum_j p_j F(\rho_j, \mathcal{E}(\rho_j))^2 = \sum_j p_j = 1 \tag{9.58}$$

Conversely, suppose there exists *at least one* j' such that $\rho'_j \neq \mathcal{E}(\rho'_j)$. The the fidelity is strictly less than one for at least one state. Since the (squared) fidelity is bounded by one, the average must be strictly less than one. Making this concrete,

$$\begin{aligned} \bar{F} = \sum_j p_j F(\rho_j, \mathcal{E}(\rho_j))^2 &\leq p_{j'} F(\rho_{j'}, \mathcal{E}(\rho_{j'})) + \sum_{j \neq j'} p_j \times 1 \\ &= p_{j'} F(\rho_{j'}, \mathcal{E}(\rho_{j'})) + (1 - p_{j'}) \\ &= 1 - p_{j'}(1 - F(\rho_{j'}, \mathcal{E}(\rho_{j'}))). \end{aligned} \tag{9.59}$$

Since $1 - F(\rho_{j'}, \mathcal{E}(\rho_{j'}))$ is strictly less than 1, so is $\bar{F}$. To summarize, if the fidelity is less than one for some state ρ_j (with $p_j > 0$), then $\bar{F} < 1$.

To summarize, $\bar{F} = 1$ exactly when all of the states ρ_j are perfectly preserved under the quantum channel $\mathcal{E}$. This is what we sought to show.

Problem 9.1 (Alternate characterization of the fidelity)

Show that

$$F(\rho, \sigma) = \inf_{P} \operatorname{tr}(\rho P) \operatorname{tr}(\sigma P^{-1}) \tag{9.60}$$

where the infimum is taken over all invertible positive matrices P.

   Solution:

Problem 9.2

Let $\mathcal{E}$ be a trace-preserving quantum operation. Show that for each ρ there is a set of operation elements $\{E_i\}$ for $\mathcal{E}$ such that

$$F(\rho, \mathcal{E}) = |\operatorname{tr}(\rho E_1)|^2 \tag{9.61}$$

   Solution:

Problem 9.3

Prove fact (5) on this page.

   Solution:

Chapter 10

Theory of quantum error correction

Exercise 10.1

Verify that the encoding circuit in Figure 10.2 works as claimed.

Solution: In order for the encoding to work properly, the circuit must take $|000\rangle \rightarrow |000\rangle$, and $|100\rangle \rightarrow |111\rangle$. Why? First, note that the first qubit is the one which we are trying to encode, and the other two are initialized in the 00 state, regardless of the state of the first qubit. Thus, for example, 100 has the unencoded first qubit in the 1 state, and we need to encode it to the logical 1 state given by 111.

Consequently, whether $|\psi\rangle$ is in the 0 or 1 state, we can easily verify that the circuit does the appropriate action. Moreover, arbitrary states are taken care of by linearity.

Exercise 10.2

The action of the bit flip channel can be described by the quantum operation $\mathcal{E}(\rho) = (1-p)\rho + pX\rho X$. Show that this may be given an alternate operator-sum representation, as $\mathcal{E}(\rho) = (1-2p)\rho + 2pP_+\rho P_+ + 2pP_-\rho P_-$ where P_+ and P_- are projectors onto the $+1$ and -1 eigenstates of X, $(|0\rangle + |1\rangle)/\sqrt{2}$ and $(|0\rangle - |1\rangle)/\sqrt{2}$, respectively. This latter representation can be understood as a model in which the qubit is left alone with probability $1-2p$, and is 'measured' by the environment in the $|+\rangle, |-\rangle$ basis with probability $2p$.

Solution: The operation elements E_i are not unique for a given quantum operation. We can obtain new operation elements F_j through a 'rotation' by some unitary matrix

$$F_j = u_{ji}E_i. \tag{10.1}$$

In our case, the original operator elements are $\sqrt{1-p}I$ and $\sqrt{p}X$. The new operator elements are $\sqrt{1-2p}I$, $\sqrt{2p}P_+$, and $\sqrt{2p}P_-$. First of all, is it even possible to express the new operators as a linear combination of the old ones? Indeed it is. First we can simply rescale the identity.

$$\sqrt{1-2p}I = \sqrt{\frac{1-2p}{1-p}}\sqrt{1-p}I \tag{10.2}$$

Next, we can write the projectors $P_\pm$ as a combination of I and X: $P_\pm = (I \pm X)/2$. Starting from here,

we can insert factors of $\sqrt{p}$ and such as required to get the right operation elements.

$$\sqrt{2p}P_{\pm} = \sqrt{p}\frac{I \pm X}{\sqrt{2}}$$

$$= \frac{\sqrt{\frac{p}{1-p}}\left(\sqrt{1-p}I\right) + \left(\sqrt{p}X\right)}{\sqrt{2}} \tag{10.3}$$

Do these linear combinations constitute a unitary combination? First, we'll need to do something to make the number of quantum operations before and after the same, otherwise our 'unitary' cannot be square. A simple way to do this is to append the zero operation 0. Using the combinations above, we can express the transformation in matrix form as

$$\begin{pmatrix} \sqrt{1-2p}I \\ \sqrt{2p}P_{+} \\ \sqrt{2p}P_{-} \end{pmatrix} = \begin{pmatrix} \sqrt{\frac{1-2p}{1-p}} & 0 & ? \\ \sqrt{\frac{p}{1-p}} & \frac{1}{\sqrt{2}} & ? \\ \sqrt{\frac{p}{1-p}} & -\frac{1}{\sqrt{2}} & ? \end{pmatrix} \begin{pmatrix} \sqrt{1-p}I \\ \sqrt{p}X \\ 0 \end{pmatrix} \tag{10.4}$$

We haven't worked out the entries with ?'s, but we won't need to. The first two columns of the matrix can be checked to be orthonormal and orthogonal to each other. Since the matrix is 3x3, we can choose the remaining column to be orthogonal to the remaining two and normalized. The resulting matrix will therefore be unitary. Moreover, since the bottom entry on the right column vector is zero, the entries don't affect the truth of the equality.

Let's call this unitary matrix u. We've shown that a unitary relationship between the two sets of operation elements exists, and therefore they constitute the same quantum operation.

Exercise 10.3

Show by explicit calculation that measuring $Z_1 Z_2$ followed by $Z_2 Z_3$ is equivalent, up to labeling of the measurement outcomes, to measuring the four projectors defined by (10.5)–(10.8), in the sense that both procedures result in the same measurement statistics and post-measurement states.

Exercise 10.4

Consider the three qubit bit flip code. Suppose we had performed the error syndrome measurement by measuring the eight orthogonal projectors corresponding to projections onto the eight computational basis states.

(1) Write out the projectors corresponding to this measurement, and explain how the measurement result can be used to diagnose the error syndrome: either no bits flipped or bit number j flipped, where j is in the range one to three.

(2) Show that the recovery procedure works only for computational basis states.

(3) What is the minimum fidelity for the error-correction procedure?

Solution: (1) For a measurement of 3 qubits in the computational basis, the corresponding projectors are given by $|s\rangle \langle s|$, where s is a bit string of length 3. There are eight such bit strings, and therefore eight orthogonal projectors on this $2^3 = 8$ dimensional space. The measurement will return one of the eight bit strings, leaving the system in state $|s\rangle$. Assuming we started in one of the logical states $|000\rangle$ or $|111\rangle$ we the

procedure will be to transform $|s\rangle$ to the logical state that is closest in terms of Hamming weight (i.e. fewest number of flips). More explicitly, if there are no flips, then s is already one of the two logical states, and we do nothing. If there is one or two flips, this means one of the qubits is different from the others, and we will flip that qubit to match. Notice this produces the wrong result for two flips. Finally, if all three qubits flip, we do nothing, which also produces the incorrect result. Provided the probability for each flip is small, our correction scheme will work most of the time, bringing us back to the logical state before the error.

(2) The problem with this error correction scheme is that it destroys coherence (meaning it eliminates superpositions of computational basis states). Suppose my system, before error, is in the state (ignoring normalization)

$$|\psi\rangle = |0\rangle_L + |1\rangle_L \tag{10.5}$$
$$= |000\rangle + |111\rangle \tag{10.6}$$

Suppose an error occurs on the 2nd qubit, so our state becomes

$$|\psi\rangle = |010\rangle + |101\rangle. \tag{10.7}$$

Performing our measurement yields either the state $|010\rangle$ or $|101\rangle$, with equal probability. Applying our correction scheme gives us either a logical zero or one, but we have lost the original state $|\psi\rangle$. This demonstrates that we need to be more clever about the way we perform measurements, and in particular our measurements cannot reveal *full* information about the encoded state $|\psi\rangle$. This is why partial information such as parity are more suitable.

(3) Since, per part (2), this procedure only works for computational basis states, let's assume our initial state is prepared in either the logical zero or one. Our goal is to recover one of these states after the correction. Without loss of generality, we may just focus on logical zero, since the procedure is symmetric with respect to the states. After applying the error, the fidelity is given by

$$\mathcal{F}_{\text{error}} = \sqrt{(1-p)^3} \tag{10.8}$$

which is simply the square root of the probability that the state is unaffected by the noise. On the other hand, with correction there is a possibility of recovering the state. The full channel $\mathcal{E}$, consisting of error and correction, can be expressed as

$$\mathcal{E}(|0\rangle_L) = (1-p)^3 |0\rangle_L \langle 0|_L + (1-p)p^2 \sum_{i=1}^{3} X_i \sum_s P_s X_i |0\rangle_L \langle 0|_L X_i P_s X_i + \dots \tag{10.9}$$

where the ellipsis indicates error terms which the correction scheme does not correct properly (more than one flip). The complicated-looking second term is merely the process of an error flip, a measurement along the projectors P_s and a correction of the original flip. It can be simplified to

$$\mathcal{E}(|0\rangle_L) = \big((1-p)^3 + 3(1-p)^2 p\big) |0\rangle_L \langle 0|_L + \dots \tag{10.10}$$

which is precisely the expression obtained from the more sophisticated bit flip correction scheme shown in the text. Borrowing those results, we see again that the minimum fidelity for either computational basis state is $\sqrt{1 - 3p^2 + 2p^3}$. Thus, we come to the conclusion that, if $p < 1/2$, our scheme will improve the fidelity of computational basis states.

As a final note, observe that our procedure, even in the absence of noise, could decrease the fidelity if our state was originally in a superposition state. Hence, there is no threshold for p for which this correction scheme would always improve the fidelity. This is a serious flaw which shows why we need a more sophisticated diagnostic, such as the one from the text.

Exercise 10.5

Show that the syndrome measurement for detecting phase flip errors in the Shor code corresponds to measuring the observables $X_1X_2X_3X_4X_5X_6$ and $X_4X_5X_6X_7X_8X_9$.

Solution: To detect a phase flip error, we need to measure "X-like" operators which are capable of distinguishing $|000\rangle + |111\rangle$ and $|000\rangle - |111\rangle$ (ignoring normalization). This is just like the case of the phase flip code introduced earlier, except now we have tripled up each 0 and 1. So we need to modify the "syndrome" (error detection) procedure of the regular phase flip code to account for this repetition. We should require that $|000\rangle \mapsto |111\rangle$ and vice versa. This is accomplished by simply repeating X on each qubit. Thus, we wish to measure $X^{\otimes 3}$ on each block. We can check straightforwardly that this operator distinguishes the phases of the two states, since the eigenvalues are different.

$$X^{\otimes 3}\frac{(|000\rangle \pm |111\rangle)}{\sqrt{2}} = \frac{(|111\rangle \pm |000\rangle)}{\sqrt{2}} = \pm\frac{(|000\rangle \pm |111\rangle)}{\sqrt{2}} \tag{10.11}$$

Recall that we don't want to determine the phase in the detection protocol, but rather decide whether the phases between two neighboring blocks of three are the *same or different*. As before, to accomplish this, we can multiply two of these repeated X operators together. This gives a string of six single-qubit X operations. If we compare the phases of blocks 1 and 2, we get

$$(X_1X_2X_3)(X_4X_5X_6) \tag{10.12}$$

which is what we wanted to show. On the other hand, comparing blocks 2 and 3 gives the remaining syndrome measurement: $(X_4X_5X_6)(X_7X_8X_9)$.

Exercise 10.6

Show that recovery from a phase flip on any of the first three qubits may be accomplished by the operator $Z_1Z_2Z_3$.

Solution: The operator $Z_1Z_2Z_3$ acts as an effective Z operator on the block of three qubits. By this we simply mean that $Z_1Z_2Z_3$ acts as the Z operator for a repetition code.

$$\begin{aligned} Z^{\otimes 3}|000\rangle &= |000\rangle \\ Z^{\otimes 3}|111\rangle &= -|111\rangle \end{aligned} \tag{10.13}$$

Because of this parallel behavior, we know that $Z_1Z_2Z_3$ will flip between effective $|+\rangle$ and $|-\rangle$ states. That is, $Z_1Z_2Z_3$ will flip the phase of the first block if applied.

As you can imagine, this argument generalizes straightforwardly in the case where the phase error actually occurs in the second block (qubits 4, 5, 6) or the third block (7, 8, 9).

Exercise 10.7

Consider the three qubit bit flip code of Section 10.1.1, with corresponding projector $P = |000\rangle\langle000| + |111\rangle\langle111|$. The noise process this code protects against has the operation elements $\{\sqrt{(1-p)^3}I, \sqrt{p(1-p)^2}X_1, \sqrt{p(1-p)^2}X_2, \sqrt{p(1-p)^2}X_3\}$, where p is the probability that a bit flips. Note that this quantum operation is not trace-preserving, since we have omitted operation elements corresponding to bit flips on two and three qubits. Verify the quantum error-correction conditions for this code and noise process.

Solution: We can ignore the coefficients in front of each of the operations, such as $\sqrt{p(1-p)^2}$, since these don't affect whether the error correction conditions of Theorem 10.1 are satisfied. Another nice feature is that all of these operations are Hermitian, e.g., $X_2^\dagger = X_2$, so we can ignore the Hermitian adjoint in the conditions. We also notice that all the operation elements commute (why?), so we can happily ignore orderings of the E_i.

To reduce the number of required calculations even further, we will try to make use of the symmetry amongst all three qubits under permutations. For example, checking $E_i^\dagger E_j = X_1^\dagger X_2 = X_1 X_2$ should also give $X_2 X_3$ and $X_3 X_1$. Having made these considerations, there are only a three cases that need to be checked: $I \times I = I, I \times X_1 = X_1$ and $X_1 X_2$. If you're unconvinced by the preceding arguments, feel free to compute other combinations of the operation elements yourself.

Let's go through each one by one. The identity is the most obvious one: $PIP = P^2 = P$. So far so good. Next we have

$$PX_1P = 0 \tag{10.14}$$

because X_1 sends states in the code space, spanned by $|000\rangle$ and $|111\rangle$, to orthogonal states. Thus, $P(X_1 P)$ will equal zero. More explicitly,

$$PX_1P = (|000\rangle\langle000| + |111\rangle\langle111|)(|100\rangle\langle100| + |011\rangle\langle011|) = 0. \tag{10.15}$$

Finally, $X_1 X_2$ takes the code space to yet another orthogonal subspace, so the result is zero: $PX_1 X_2 P = 0$. You can check with a direct calculation along similar linear to equation (10.15).

Since $X_i^2 = I$, we are essentially done. Again ignoring the square root coefficients, we can express our result as a 4×4 matrix α.

$$\alpha = \operatorname{diag}(1,1,1,1) = I \tag{10.16}$$

That is, α is a diagonal matrix with real entries, hence it is Hermitian. What would the α matrix be if we included the square root factors?

Exercise 10.8

Verify that the three qubit phase flip code $|0_L\rangle = |+++\rangle$, $|1_L\rangle = |---\rangle$ satisfies the quantum error-correction conditions for the set of error operators $\{I, Z_1, Z_2, Z_3\}$.

Solution: We will not do the calculation for this problem. Instead, we will point out the fact that the mathematics for this problem is identical to the previous one (Exercise 10.7). Phase flips for superposition states are equivalent to NOT gates for computational states $0, 1$.

To be more precise, under Hadamard transformation for each qubit,

$$|+\rangle \mapsto |0\rangle \qquad |-\rangle \mapsto |1\rangle \qquad Z_i \mapsto HZ_iH = X_i, \tag{10.17}$$

all aspects of the problem transform into the corresponding bit-flip problem. For example, just as $X|0\rangle = |1\rangle$, $Z|+\rangle = |-\rangle$. Therefore, we will arrive at the same conclusion: the phase flip code is correctable.

Exercise 10.9

Again, consider the three qubit phase flip code. Let P_i and Q_i be the projectors on to the $|0\rangle$ and $|1\rangle$ states, respectively, of the ith qubit. Prove that $\{I, P_1, Q_1.P_2, Q_2, P_3, Q_3\}$.

Solution: We can make use of Theorem 10.2 We should recognize that the projectos P and Q into the computational (Z) basis are in the span of Z itself and the identity I. More precisely,

$$P = \frac{I + Z}{2} \qquad Q = \frac{I - Z}{2}. \tag{10.18}$$

Theorem 10.2, applied to our case, says we can correct any errors that are in the span of $\{I, Z_1, Z_2, Z_3\}$. Thus we can correct each of these projectors P, Q on each qubit i.

We conclude that the phase flip code also protects against "spontaneous unwanted measurement" in the computational basis.

Exercise 10.10

Explicitly verify the quantum error-correction conditions for the Shor code, for the error set containing I and the error operators X_j, Y_j, Z_j for $j = 1$ through 9.

Solution: The error operators, including the identity, total $1 + (3 \times 9) = 28$, one for each pauli (3) of each qubit (9). A brute force calculation of $P E_i^\dagger E_j P$, using all possible pairs of these 28, is clearly time consuming.

Thankfully, we can make a number of helpful simplifications. All of these operations either commute or anticommute, and all are Hermitian. Hence, we can ignore orderings up to a sign, and neglect daggers. Let's further consider the symmetries between the nine qubits. All qubits within a block of three are clearly "the same", i.e., invariant under permutation. Moreover, the blocks of three themselves are symmetric under interchange. This suggests we only need to calculate two cases which capture all possible combinations:

- Errors occuring *within* the same block (say, qubits 1 and 2)

- Errors occurring in two *distinct blocks* (say, on qubits 1 and 4).

Putting this all together, there are only a handful of classes we need to consider for $E_i E_j$:

- $E_i E_j = $ a pauli on qubit 1

- $E_i E_j = $ a pauli on qubit 1, another on qubit 2

- $E_i E_j = $ a pauli on qubit 1, another on qubit 4

We will go through these cases one at a time.

Suppose $E_i E_j = \sigma_1$, where σ_1 is some pauli (X, Y, or Z) action on the first qubit. Since the Shor code projector P is the tensor product of projectors on the three blocks, we only need to consider the first block. That is, we consider

$$P_{123} = \frac{(|000\rangle + |111\rangle)(\langle 000| + \langle 111|)}{2} \tag{10.19}$$

which is the part of the projector P on the first block. The thing to note is that, regardless of which pauli σ_1 is, it sends the state $(|000\rangle + |111\rangle)/\sqrt{2}$ to a state which is orthogonal to P_{123}, and similarly for $(|000\rangle - |111\rangle)/\sqrt{2}$. Hence, $P_{123}\sigma_1 P_{123} = 0$. We conclude that $P\sigma_1 P = 0$, so that in particular it is proportional to P. So far so good.

Next we consider $\sigma_1\sigma_2$. As before, we only focus on a single block. If either (or both) of the paulis are X or Y, then there is a bit flip involved and we will get something orthogonal to the original code space. Hence, $P\sigma_1\sigma_2 P$ will be zero. The only remaining possibility is Z_1Z_2. Interestingly we see that this leaves the states $|000\rangle \pm |111\rangle$ invariant. Thus $PZ_1Z_2P = PP = P$. In all cases, we still find something real and proportional to P.

Finally, we consider $E_iE_j = \sigma_1\sigma_4$, going across blocks. We can ignore the third block. If any bit flip occurs at all, P gets sent to something orthogonal to P. It then remains to check Z_1Z_4. But each of these sends its block to something orthogonal to the code space. In all cases here, $P\sigma_1\sigma_4 P = 0$.

Let's summarize our results. In all cases, $PE_i^\dagger E_j P$ is proportional to P and we can capture the proportionality by a 28×28 matrix of coefficients α_{ij}. How do we know it is Hermitian? The diagonals are all one, since each error operator squares to the identity. Our calculations above also show that nearly every single off-diagonal element is zero, except for Z_1Z_2, or more generally, Z_iZ_j where $i \neq j$ but i, j are in the same block. In this case, $\alpha_{ij} = 1$ (note that the indices ij means something different in α than for Z_iZ_j). Fortunately, in this case Z_i and Z_j commute.

Altogether, this implies, α is a real, symmetric matrix, hence Hermitian. We conclude that the Shor code can successfully correct arbitrary single qubit errors.

Exercise 10.11

Construct operation elements for a single qubit quantum operation $\mathcal{E}$ that upon input of any state ρ replaces it with the completely randomized state $I/2$. It is amazing that even such noise models as this may be corrected by codes such as the Shor code!

Solution: Using the material of Chapter 8 of the text, we could go through an explicit construction of the quantum operation $\mathcal{E}$ such that $\mathcal{E}(\rho) = I/2$. We will instead appeal to prior knowledge and intuition.

Suppose you know that the depolarizing channel $\mathcal{E}_p$ sends ρ to the maximally mixed state I/d with probability p, otherwise leaving it alone. That is, $\mathcal{E}_p(\rho) = (1-p)\rho + pI/d$. Suppose you also know that $\mathcal{E}_p$ can be constructed for a single qubit as a sum over Pauli sandwiches

$$\mathcal{E}_p(\rho) = (1-p)\rho + p\frac{1}{3}(X\rho X + Y\rho Y + Z\rho Z). \tag{10.20}$$

This equation is displayed on page 440, right before this problem. The intuition is, by applying a pauli with uniform randomness, we lose all information about our state.

Setting $p = 1$ gives us the desired *completely depolarizing channel* $\mathcal{E}$ on a single qubit.

$$\mathcal{E}(\rho) = \frac{1}{3}(X\rho X + Y\rho Y + Z\rho Z). \tag{10.21}$$

How do we know the Shor code corrects this? Assuming this error occured on a single qubit (and only one qubit), the operator elements match those we know to be correctable by the Shor code.

Exercise 10.12

Show that the fidelity between the state $|0\rangle$ and $|I\rangle\langle I|$ is $\sqrt{1 - 2p/3}$, and use this to argue that the minimum fidelity for the depolarizing channel is $\sqrt{1 - 2p/3}$.

Solution:

Exercise 10.13

Show that the minimum fidelity $F(|\psi\rangle, |\psi\rangle \langle\psi|)$ when $\mathcal{E}$ is the amplitude damping channel with parameter γ is $\sqrt{1-\gamma}$.

   Solution:

Exercise 10.14

Write an expression for a generator matrix encoding k bits using r repetitions for each bit. This is an $[rk, k]$ linear code, and should have an $rk \times k$ generator matrix. **Solution:**

Exercise 10.15

Show that adding one column of G to another results in a generator matrix generating the same code.

   Solution:

Exercise 10.16

Show that adding one row of the parity check matrix to another does not change the code. using Gaussian elimination and swapping of bits it is therefore possible to assume that the parity check matrix has the *standard form* $[A|I_{n-k}]$ where A is an $(n-k) \times k$ matrix.

   Solution:

Exercise 10.17

Find a parity check matrix for the $[6, 2]$ repetition code defined by the generator matrix in (10.54).

   Solution:

Exercise 10.18

Show that the parity check matrix H and generator matrix G for the same linear code satisfy $HG = 0$.

   Solution:

Exercise 10.19

Suppose an $[n, k]$ linear code C has a parity check matrix of the form $H = [A|I_{n-k}]$, for some $(n-k) \times k$ matrix A. Show that the corresponding generator matrix is

$$G = ?? \tag{10.22}$$

Solution:

Exercise 10.20

Let H be a parity check matrix such that any $d-1$ columns are linearly independent, but there exists a set of d linearly dependent columns. Show that the code defined by H has distance d.

Solution:

Exercise 10.21 (Singleton bound)

Show that an $[n, k, d]$ code must satisfy $n - k \geq d - 1$.

Solution:

Exercise 10.22

Show that all Hamming codes have distance 3, and thus can correct an error on a single bit. The Hamming codes are therefore $[2^r - 1, 2^r - r - 1, 3]$ codes.

Solution:

Exercise 10.23

Prove the Gilbert-Varshamov bound.

Solution:

Exercise 10.24

Show that a code with generator matrix G is weakly self-dual if and only if $G^T G = 0$.

Solution:

Exercise 10.25

Let C be a linear code. Show that if $x \in C^\perp$ then $\sum_{y \in C}(-1)^{x \cdot y} = |C|$, while if $x \notin C^\perp$ then $\sum_{y \in C}(-1)^{x \cdot y} = 0$.

Solution:

Exercise 10.26

Suppose H is a parity check matrix. Explain how to compute the transformation $|x\rangle |0\rangle \to |x\rangle |Hx\rangle$ using a circuit composed entirely of controlled-NOTs.

Solution:

Exercise 10.27

Show that the codes defined by

$$|x + C_2\rangle \equiv \frac{1}{\sqrt{|C_2|}} \sum_{y \in C_2}(-1)^{u \cdot y} |x + y + v\rangle \tag{10.23}$$

and parameterized by u and v are equivalent to $\mathrm{CSS}(C_1, C_2)$ in the sense that they have the same error-correcting properties. These codes, which we'll refer to ass $\mathrm{CSS}_{u,v}(C_1, C_2)$, will be useful later in our study of quantum key distribution in Section 12.6.5.

Solution:

Exercise 10.28

Verify that the transpose of the matrix in (10.77) is the generator of the $[7, 4, 3]$ Hamming code.

Solution:

Exercise 10.29

Show that an arbitrary linear combination of any two elements of V_S is also in V_S . Therefore, V_S is a subspace of the n qubit state space. Show that V_S is the intersection of the subspaces fixed by each operator in S (that is, the eigenvalue one eigenspaces of elements of S).

Solution: Let $\psi, \chi \in V_S$, and let $A \in S$. Then, A is a linear operator, and for c_1, c_2 in the underlying field $\mathbb{F}$ (in this case $\mathbb{F} = \mathbb{C}$), we have

$$\begin{aligned} A\left(c_1\psi + c_2\chi\right) &= c_1 A\psi + c_2 A\chi \\ &= c_1\psi + c_2\chi \end{aligned} \tag{10.24}$$

Hence, A stabilizes $c_1\psi + c_2\chi$. Hence, $c_1\psi + c_2\chi \in V_S$. We conclude V_S is a subspace.

Let $\cap_{s \in S} V_s$ be the intersection of all subspaces stabilized by a single element $s \in S$. We will show that this equals V_S itself by showing they are subsets of each other. Let $\psi \in V_S$, so that ψ is stabilized by s for each $s \in S$. Then, $\psi \in V_s$ for every such s, and therefore is also in the intersection. Hence, $V_S \subset \cap_{s \in S} V_s$. Conversely, suppose $\psi \in \cap_{s \in S} V_s$, so that $\psi \in V_s$ for each $s \in S$. Then, ψ is stabilized by every $s \in S$, and therefore is stabilized by S itself. Thus, $\psi \in V_S$. This shows $\cap_{s \in S} V_s \subset V_S$, which completes the proof.

Exercise 10.30

Show that $-I \notin S$ imples $\pm iI \notin S$.

Solution: To show this, we prove the (equivalent) contrapositive: $\pm iI \in S$ implies $-I \in S$. Phrased in this way, the reasoning follows simply from S being a group, and therefore closed under multiplication. In particular, if $\pm iI \in S$, so is the square

$$(\pm iI)^2 = -I \in S. \tag{10.25}$$

Exercise 10.31

Suppose S is a subgroup of G_n generated by elements $g_1, \ldots, g_l$. Show that all the elements of S commute if and only if g_i and g_j commute for each pair i, j.

Solution: If all the elements of S commute, then clearly so do the generators, since the generators are also in S. Conversely, and less trivially, if all of the generators commute, so must the whole group, since any group member is just a product of the generators.

Exercise 10.32

Verify that the generators in Figure 10.6 stabilize the codewords for the Steane code, as described in Section 10.4.2.

Solution: Here we will only verify that g_1 stabilizes the two codewords, which are given by equations (10.78) and (10.79) of the text. We have

$$g_1 |0_L\rangle = \frac{1}{\sqrt{8}}[|0001111\rangle + |1011010\rangle + |0111100\rangle + |1101001\rangle +$$
$$+ |0000000\rangle + |1010101\rangle + |0110011\rangle + |1100110\rangle] \tag{10.26}$$

In words, g_1 flips the last four bits and leaves the first three alone. Looking at this expression, we can see it is the same as $|0_L\rangle$ itself by swapping each ket on the top row with the one directly below it. This simply amounts to rearranging terms in the sum. Thus, $g_1 |0_L\rangle = |0_L\rangle$. To show that $|1_L\rangle$ by g_1, we can use a slick trick. Looking at the expression for $|1_L\rangle$, we see it is related to $|0_L\rangle$ by flipping every bit, i.e. the operation $X^{\otimes 7}$. This logical X commutes with g_1, since each X commutes with I and X. Altogether,

$$g_1 |1_L\rangle = g_1 X^{\otimes 7} |0_L\rangle = X^{\otimes 7} g_1 |0_L\rangle = X^{\otimes 7} |0_L\rangle = |1_L\rangle. \tag{10.27}$$

This completes the demonstration for g_1. We could check the remaining stabilizers in an overall similar way.

Exercise 10.33

Solution:

Exercise 10.34

Let $S = \langle g_1, \ldots, g_l \rangle$. Show that $-I$ is not an element of S if and only if $g_j^2 = I$ for all j, and $g_j \neq -I$ for all j.

Solution: Before proceeding, there is an unstated assumption in the exercise: that all the generators are independent. To see this, consider the case $S = \langle X, -X \rangle$. Clearly, $g_i^2 = I$ and $g_i \neq -I$ for both generators, yet $-I \in S$. Therefore, the theorem is false if there is no assumption of independence. We will show that, with the added assumption, the theorem is true.

($\Longrightarrow$) Suppose $-I \notin S$. Then, none of the generators are $-I$, so $g_j \neq -I$. Since any $g_j \in G_n$ is a string of pauli matrices times ± 1 or $\pm i$, we have $g_j^2 = I$ or $g_j^2 = -I$. By assumption, $g_j^2 \neq -I$, so $g_j = I$. This implies the forward direction of the equivalence.

($\Longleftarrow$) Suppose for all j, $g_j^2 = I$ and $g_j \neq -I$. Let $s \in S$, so that it can be written as a product of the generators. Since all elements of commute or anticommute, and since $g_j^2 = I$, any such element s can be expressed as

$$s = \pm g_{j_1} g_{j_2} \ldots g_{j_k} \tag{10.28}$$

where all the j_k are different integers between 1 and l. Suppose, for sake of contradiction, that $s = -I$. Then, we can express one of the generators (say, g_{j_1}) in terms of the others.

$$g_{j_1} = \mp g_{j_2} \ldots g_{j_k} \tag{10.29}$$

This violates our assumption about the independence of the generators, a contradiction. We conclude $s \neq -I$. Since s was arbitrary, $-I \notin S$. This completes the reverse implication, and therefore the proof.

Exercise 10.35

Let S be a subgroup of G_n such that $-I$ is not an element of S. Show that $g^2 = I$ for all $g \in S$, and thus $g^\dagger = g$.

Solution: This was actually shown as part of our solution to Exercise 10.34. In particular, it was argued that any element $g \in G_n$ of the pauli group squares to either I or $-I$. This follows from the fact that every pauli matrix itself squares to the identity, and the coefficient squares to 1 or -1 depending on whether it is imaginary or not. If we assume $-I \notin S$, we must have the case $g^2 = I$ since S (like any group) is closed under the group operation.

Furthermore, any element $g \in G_n$ is either Hermitian or antihermitian, depending on the coefficient being imaginary or real. If $g^2 = I$, the coefficient is real, and therefore g is Hermitian. Note that g is also a reflection, and its eigenvalues are ± 1. Going one step further, g must have the same number of plus and minus eigenvalues, since $\text{tr}(g) = 0$.

Exercise 10.36

Explicitly verify that $UX_1U^\dagger = X_1X_2$, $UX_2U^\dagger = X_2$, $UZ_1U^\dagger = Z_1$, and $UZ_2U^\dagger = Z_1Z_2$. These and other useful conjugation relations for the Hadamard, phase, and Pauli gates are summarized in Figure 10.7.

Solution: We will show each of these relationships in sequence, but rather than using the matrix representations of the operators we will use the standard operator form of the CNOT operator.

$$\text{CNOT} = |0\rangle\langle 0| I + |1\rangle\langle 1| X. \tag{10.30}$$

(We've left off the tensor product $\otimes$ for simplicity). First, we have

$$
\begin{aligned}
\text{CNOT}X_1\text{CNOT} &= (|0\rangle\langle 1| I + |1\rangle\langle 0| X)(|0\rangle\langle 0| I + |1\rangle\langle 1| X) \\
&= |1\rangle\langle 0| X + |0\rangle\langle 1| X \\
&= (|1\rangle\langle 0| + |0\rangle\langle 1|) X \\
&= X_1X_2
\end{aligned}
\tag{10.31}
$$

Next, we have

$$\text{CNOT}X_2\text{CNOT} = X_2 \tag{10.32}$$

by the simple fact that X_2 commutes with CNOT. This can be checked by computation, but it is also easy to see because X commutes with both the identity and X, which are, roughly speaking, the two possible operations depending on the control qubit.

Similarly, since Z_1 commutes with CNOT, it is invariant under conjugation.

$$\text{CNOT}Z_1\text{CNOT} = Z_1. \tag{10.33}$$

Intuitively, a Z gate does not "change" the computational basis which is being controlled on. More precisely, Z commutes with the projection operators onto the computational basis. Finally, we can compute the action on Z_2 similar to how we did X_1.

$$
\begin{aligned}
\text{CNOT}Z_2\text{CNOT} &= (|0\rangle\langle 0| Z + |1\rangle\langle 1| (XZ))(|0\rangle\langle 0| I + |1\rangle\langle 1| X) \\
&= |0\rangle\langle 0| Z + |1\rangle\langle 1| (XZX) \\
&= |0\rangle\langle 0| Z - |1\rangle\langle 1| Z \\
&= (|0\rangle\langle 0| - |1\rangle\langle 1|) Z \\
&= Z_1Z_2
\end{aligned}
\tag{10.34}
$$

Exercise 10.37

What is $UY_1U^\dagger$?

Solution: We can make use of the previous exercise by noting that $Y_1 = -iZ_1X_1$, and the conjugation by U respects multiplication. More specifically,

$$
\begin{aligned}
\text{CNOT}Y_1\text{CNOT} &= -i\text{CNOT}Z_1\text{CNOTCNOT}X_1\text{CNOT} \\
&= -iZ_1X_1X_2 \\
&= Y_1X_2.
\end{aligned}
\tag{10.35}
$$

Exercise 10.38

Suppose U and V are unitary operators on two qubits which transform Z_1, Z_2, X_1, and X_2 by conjugation in the same way. Show this implies that $U = V$.

Solution: As a first observation, we note that the statement of the exercise must be carefully understood, since $V = e^{i\delta}U$ would have the same conjugation properties as U. Of course, we know that this overall phase is irrelevant when it comes to quantum operations. Therefore, we wish to show that U and V must be the same *up to a phase*.

With some thought, it's clear that Z_i and X_i for $i = 1, 2$ generates the full Pauli group on two qubits, G_2. Therefore, U and V will also transform *any* element $P \in G_2$ the same way. In light of this observation, let's answer the more general question: if two operators U and V on n qubits transform elements of the pauli group G_n the same way, does $U = V$ (up to a phase)? This result will of course specialize to $n = 2$.

It seems hard to imagine this couldn't be the case since, thinking of operators as a linear space, the pauli group G_n forms a spanning set for all matrices of dimension 2^n. Therefore, if U and V transform G_n the same way, they transform any matrix $M \in \mathbb{C}^{2^n} \times \mathbb{C}^{2^n}$ the same way too! It seems hard to imagine U and V acting so similar and yet not being the same.

Let's make one final observation before diving into an actual proof. Suppose $UMU^\dagger = VMV^\dagger$ for all complex matrices M. Then $M = WMW^\dagger$ where $W \equiv U^\dagger V$. Since W stabilizes all matrices under conjugation, could it really be that W is not the identity, implying $U = V$ (all statements with the addendum "up to a phase")? This will be our vehicle for the proof.

Proof of exercise: Let $L(V)$ be linear operators over a complex vector space V, and suppose W is a unitary operator on V such that, for all $M \in L(V)$, $WMW^\dagger = M$. Viewing $L(V)$ as a complex vector space itself, the conjugation operation $W \cdot W^\dagger$ is a linear operator which acts trivially on each vector M. In particular, it acts trivially on rank one operators $|a\rangle \langle b|$. This necessarily implies $W|a\rangle = e^{i\delta}|a\rangle$ for all $|a\rangle$, where δ is a real phase that does not depend on $|a\rangle$. Of course, $Ie^{i\delta}$ also has the same action for every vector in V, so we must conclude $W = e^{i\delta}I$.

Now suppose W is an operator that stabilizes the pauli group G_n under conjugation. Since the pauli group forms a basis for the complex vector space of linear operators $L(\mathbb{C}^{2^n})$, then W stabilizes this space as well. By the result of the previous paragraph, $W = e^{i\delta}I$.

Suppose now that W only stabilizes a set which generates G_n. Since conjugation respects group multiplication, it must be that any product of the generators, hence the whole group itself, must be stabilized by W. By the previous paragraph, we conclude $W = e^{i\delta}I$.

As a simple corollary, suppose $UPU^\dagger = VPV^\dagger$ for every $P \in G_n$. Then $P = WPW^\dagger$, where $W \equiv U^\dagger V$. From the above results, we must have $W = e^{i\delta}I$, and therefore $V = Ue^{i\delta}$. This essentially completes the proof.

Exercise 10.39

Verify (10.91).

Solution: This can be demonstrated by direct calculation. It also just makes sense: S is a 90 degree rotation about the z axis of the Bloch sphere, thereby taking the vector X to Y and leaving Z unchanged.

Exercise 10.40

Provide an inductive proof of Theorem 10.6 as follows.

1. Prove that the Hadamard and phase gates can be used to perform any normalizer operation on a single qubit.

2. Suppose U is an $n + 1$ qubit gate in $N(G_n + 1)$ such that $UZ_1U^\dagger = X_1 \otimes g$ and $UX_1U^\dagger = Z_1 \otimes g'$ for some $g, g' \in G_n$. Define U' on n qubits by $U' |\psi\rangle \equiv \sqrt{2} \langle 0| U(|0\rangle \otimes |\psi\rangle)$. Use the inductive hypothesis to show that the construction for U in Figure 10.9 may be implemented using $O(n^2)$ Hadamard, phase and controlled-NOT gates.

3. Show that any gate $U \in N(G_n + 1)$ may be implemented using $O(n^2)$ Hadamard, phase and controlled-NOT gates.

Solution: 1. Let U be a normalizer operation for G_1, the pauli group on a single qubit. Then U is defined by how it acts on a set of generators, e.g. X and Z. Let's call these elements $u(X)$ and $u(Z)$.

$$X \mapsto UXU^\dagger \equiv u(X)$$
$$Z \mapsto UZU^\dagger \equiv u(Z) \tag{10.36}$$

From the onset, it is clear that $u(X)$ and $u(Z)$ cannot be *any* member of G_1. For example, since conjugation preserves hermiticity, the antihermitian elements such as iX are left out. Moreover, any element proportional to the identity is invariant under conjugation, and since conjugation is bijective there is no way X or Z could be mapped to them. Therefore, any normalizer U can only send X and Z to one of the set

$$\{\pm X, \pm Y, \pm Z\}. \tag{10.37}$$

These aren't the only restrictions! Since u preserves group multiplication (as an isomorphism from the group to itself, aka an automorphism), $u(X)$ and $u(Z)$ must anticommute. This implies $u(X)$ and $u(Z)$ cannot be related by a sign, and therefore the must be proportional to different pauli matrices. Altogether, this leaves us with 24 valid choices for the mapping u. Since the Hadamard interchanges X and Z, we can effectively consider only 12 of them.

The question is, can these 12 maps be represented as conjugations with a unitary U which is a product of Hadamards and phase gates? Intuitively the answer is yes, because this combination takes us to all "corners" of the Bloch sphere. Figure (WHAT) illustrates this with a graph. We work out the unitaries U for each of the 12 possibilities in the table below. This proves our result. **YOU NEED TO ADD THE TABLE**

XZ	X Z	I
	X -Z	HS^2H
	X Y	SHS
	X -Y	HSH
	-X Z	cell9
	-X -Z	cell9
	-X Y	cell9
	-X -Y	cell9
	Y Z	cell9
	Y -Z	cell9
	-Y Z	cell9
	-Y -Z	cell9

Exercise 10.41

Solution:

Chapter 11

Entropy and information

Exercise 11.1 (Simple calculations of entropy)

What is the entropy associated with the toss of a fair coin? With the roll of a fair die? How would the entropy behave if the coin or die were unfair?

Solution: For the fair coin, we have two possibilities (heads or tails) and each outcome has equal probability of $1/2$. Hence, the entropy H_{coin} is given by

$$H_{\mathrm{coin}} = -\frac{1}{2}\log 1/2 - -\frac{1}{2}\log 1/2 = \log 2. \tag{11.1}$$

Similarly, a fair die has *six* equally likely outcomes, each with probability $1/6$. There are therefore six terms in the expression for the entropy H_{die}, and each term is given by $(\log 1/6)/6$. Hence

$$H_{\mathrm{die}} = 6 \times -\frac{1}{6}\log 1/6 = \log 6. \tag{11.2}$$

Following the pattern, a set of n equally likely outcomes yields an entropy of $\log n$.

What happens if the coin is unfair? To model this, we can assign the coin a probability of heads p between 0 and 1. Necessarily, the probability of tails is $1 - p$. The entropy is now given by

$$H_{\mathrm{coin}}(p) = -p\log p - (1-p)\log(1-p). \tag{11.3}$$

Try plugging in different values for p: what is the behavior?

There are many more ways we can construct an unfair die, since there are more outcomes. What if we make one of the sides much more likely than the others? Try out some options, making sure your probabilities always add to one!

Chapter 12

Quantum information theory

Exercise 12.1

Suppose $|\psi\rangle$ and $|\varphi\rangle$ are two orthogonal quantum states of a single qubit. Design a quantum circuit with two input qubits (the 'data' and the 'target' qubits), with the data qubit in either the $|\psi\rangle$ or $|\varphi\rangle$, and the target qubit prepared in the state $|0\rangle$, which produces as output $|\psi\rangle |\psi\rangle$ or $|\varphi\rangle |\varphi\rangle$ depending on whether $|\psi\rangle$ or $|\varphi\rangle$ was input to the data qubit.

Solution: The states $|\psi\rangle, |\varphi\rangle$ define an orthonormal basis, and therefore there is a change of basis unitary U between these states and the computational basis.

$$|\psi\rangle = U |0\rangle \quad |\varphi\rangle = U |1\rangle \tag{12.1}$$

Depending on which state the data qubit is in, we want to apply U or UX to the initial state $|0\rangle$ of the target qubit. This suggests we use a CNOT gate. We will first use the inverse gate $U^\dagger$ to rotate the data qubit to the computational basis, perform the controls, then apply U to the data to bring it to the original value. Figure 12.1 depicts the full circuit.

This circuit can be readily generalized to copy any n-qubit state from a given orthonormal basis. In this context, the $U^\dagger$ will be an n-qubit unitary mapping the chosen basis to the computational basis, and the CNOT gate will become a string of n CNOTs from the ith qubit of the data register to the ith qubit of the target. These CNOTs are essentially the copy circuit for the computational basis.

$$\tag{12.2}$$

Figure 12.1: Circuit for copying one of two orthogonal states on a qubit. Here U is the change of basis from the states to the computational basis. $|data\rangle$ represents either $|\psi\rangle$ or $|\phi\rangle$.

Exercise 12.2

Define U_y to be the unitary operator acting on system M whose action on a basis is $U_y |y'\rangle \equiv |y' + y\rangle$, where the addition is done modulo $n + 1$. Show that $\{\sqrt{E_y} \otimes U_y\}$ is a set of operation elements defining a trace-preserving quantum operation $\mathcal{E}$ whose action on a state of the form $\sigma \otimes |0\rangle \langle 0|$ agrees with (12.8).

Solution: The computation is relatively straightforward. First note that we are assuming E_y are positive, hence their square root is well defined and they are Hermitian. We have

$$
\begin{aligned}
\mathcal{E}(\sigma \otimes |0\rangle \langle 0|) &= \sum_y \left(\sqrt{E_y} \otimes U_y \right) \sigma \otimes |0\rangle \langle 0| \left(\sqrt{E_y} \otimes U_y^\dagger \right) \\
&= \sum_y \left(\sqrt{E_y} \sigma \sqrt{E_y} \right) \otimes \left(U_y |0\rangle \langle 0| U_y^\dagger \right) \\
&= \sum_y \sqrt{E_y} \sigma \sqrt{E_y} \otimes |y\rangle \langle y|
\end{aligned}
\tag{12.3}
$$

which we see matches (12.8) from the text.

Exercise 12.3

Use the Holevo bound to argue that n qubits can not be used to transmit more than n bits of classical information.

Solution: Suppose that Alice wishes to communicate m bits of classical information by sending *Bob* one of an ensemble of n-qubit states. According to the Holevo bound, the mutual information $H(X : Y)$ between Alice's choice and Bob's measurement is bounded by

$$
H(X : Y) \leq S(\rho) \leq n
\tag{12.4}
$$

where ρ is some density matrix describing the ensemble of states Bob receives. We interpret this as Alice being limited to transmitting n bits of information with n-qubits.

Exercise 12.4

Suppose Alice sends Bob an equal mixture of the four pure states

$$
|X_1\rangle = |0\rangle
\tag{12.5}
$$

$$
|X_2\rangle = \sqrt{\frac{1}{3}} \left[|0\rangle + \sqrt{2} |1\rangle \right]
\tag{12.6}
$$

$$
|X_3\rangle = \sqrt{\frac{1}{3}} \left[|0\rangle + \sqrt{2} e^{2\pi i/3} |1\rangle \right]
\tag{12.7}
$$

$$
|X_4\rangle = \sqrt{\frac{1}{3}} \left[|0\rangle + \sqrt{2} e^{4\pi i/3} |1\rangle \right]
\tag{12.8}
$$

Show that the maximum mutual information between Bob's measurement and Alice's transmission is less than one bit. A POVM which achieves ≈ 0.415 bits is known. Can you construct this or, better yet, one which achieves the Holevo bound?

Appendix 1

Notes on basic probability theory

Exercise A1.1

Prove Bayes' rule.

Solution: From the definition of conditional probability, we have

$$p(x, y) = p(y|x)p(x) = p(x|y)p(y) \tag{A1.1}$$

Rearranging the last of these equations gives the desired result.

Exercise A1.2

Prove the law of total probability.

Solution: We start with the notion that, in the joint probability distribution for (X, Y), one sums over all outcomes of X to get a probability distribution on Y alone.

$$p(y) := \sum_{x} p(x, y) \tag{A1.2}$$

We arrive at our result by noting that, from the definition of conditional probability, $p(x, y) = p(y|x)p(x)$.

Exercise A1.3

Prove there exists a value of $x \geq \mathbf{E}(X)$ such that $p(x) > 0$.

Solution: Suppose, for sake of contradiction, that every value x of X with nonzero probability has the property $x < \mathbf{E}(X)$. Intuitively, we'd expect that the expectation value would have to be less then $\mathbf{E}(X)$. Indeed, using the inequality in the definition of expectation value,

$$\mathbf{E}(X) = \sum_{x \in X} x p(x) < \mathbf{E}(X) \sum_{x \in X} p(x) = \mathbf{E}(X) \tag{A1.3}$$

Hence, $\mathbf{E}(X) < \mathbf{E}(X)$, a clear contradiction. We conclude our premise was false, hence there does exist a value of $x \in X$ such that $x \geq \mathbf{E}(X)$ and $p(x) > 0$.

Exercise A1.4

Prove that $\mathbf{E}(X)$ is linear in X.

Solution: The following computation gives us the result.

$$
\begin{aligned}
\mathbf{E}(aX + bY) &= \sum_{(x,y)\in(X,Y)} (ax + by)p(x,y) \\
&= \sum_{x\in X}\sum_{y\in Y} axp(x,y) + byp(x,y) \\
&= \sum_{x\in X} ax \sum_{y\in Y} p(x,y) + \sum_{y\in Y} by \sum_{x\in X} p(x,y) \\
&= a \sum_{x\in X} xp(x) + b \sum_{y\in Y} yp(y) \\
&= a\mathbf{E}(X) + b\mathbf{E}(Y).
\end{aligned}
\tag{A1.4}
$$

Here, a, b are constants. Along the way, we used $p(x) = \sum_y p(x,y)$ and the definition of expectation value.

Exercise A1.5

Prove that for independent random variables X and Y, $\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y)$.

Solution: Recall that, for independent random variables, the joint probability distribution breaks into a product of individual probabilities. This yields the following computation.

$$
\begin{aligned}
\mathbf{E}(XY) &= \sum_x \sum_y xy\, p(x,y) \\
&= \sum_x \sum_y xy\, p(x)p(y) \\
&= \sum_x p(x) \sum_y p(y) \\
&= \mathbf{E}(X)\mathbf{E}(Y)
\end{aligned}
\tag{A1.5}
$$

Exercise A1.6

Prove Chebyshev's inequality.

Solution: Our solution is taken from Kliesch and Roth, 2021. We start by proving a more fundamental result: *Markov's inequality.* Let Y be a nonnegative random variable, and $t > 0$. Then

$$
p(Y \geq t) \leq \frac{\mathbf{E}(Y)}{t}.
\tag{A1.6}
$$

To show this, let Ω be the set of outcomes over which our random variable Y is defined. Consider the indicator function $\mathbf{1}_A$ for some $A \subset \Omega$, defined as follows.

$$
\mathbf{1}_A(\omega) = \begin{cases} 1 & y \in A \\ 0 & y \notin A \end{cases}
\tag{A1.7}
$$

Take the particular choice $A = \{\omega \in \Omega | Y(\omega) \geq t\}$, one can observe that

$$t\mathbf{1}_A(\omega') \leq Y(\omega') \tag{A1.8}$$

for any $\omega' \in \Omega$. Taking the expectation value of our results gives us Markov's inequality.

To obtain Chebyshev's inequality, let $Y = |X - \mathbf{E}(X)|^2$ for some probability distribution X, and let $\lambda^2 = t/\Delta(X)^2$. Note that $\mathbf{E}(Y) = \Delta(X)^2$. Making these substitutions for t and Y gives

$$p(|X - \mathbf{E}(X)|^2 \geq \lambda^2 \Delta(X)^2) \leq \frac{1}{\lambda^2} \tag{A1.9}$$

$$p(|X - \mathbf{E}(X)| \geq \lambda\Delta(X)) \leq \frac{1}{\lambda^2} \tag{A1.10}$$

This proves our result.

Appendix 2

Group theory

Exercise A2.1

Prove that for any element g of a finite group, there always exists a positive integer r such that $g^r = e$. That is, every element of such a group has an order.

Solution: The proof relies on the pigeonhole principle. If G is a finite group, and $g \in G$, then taking repeated powers of g must yield a repeated element. More formally, there exist distinct positive integers $i, j \in \mathbb{Z}^+$ such that $g^i = g^j$. Without loss of generality, assume $j > i$. Multiplying by g^{-i} on both sides gives $e = g^{j-i}$. Hence, there exists a positive integer $k = j - i$ such that $g^k = e$. Hence g has an order.

Exercise A2.2

Prove Lagrange's theorem.

Solution: To prove the theorem, we will introduce the notion of a coset, which is actually discussed later in the appendix. If $g \in G$, define the (left) coset $gH = \{gh | h \in H\}$. Consider the set of cosets $\{gH | g \in G\}$. I claim two things.

1. Each coset is the same size, and in particular the same size as H (which is itself a coset).

2. The collection of cosets is a *partition* of the group G. This means that

   (a) Each distinct coset is disjoint, i.e. if $g_1 H \neq g_2 H$, then $g_1 H \cap g_2 H = \emptyset$

   (b) Every element of G is in some coset.

Let us prove each of these claims one by one. Consider some coset gH, and define a bijection $g : H \to gH$ by left multiplication.

$$g(h) = gh \tag{A2.1}$$

Clearly, this is a map to and from the appropriate sets. The map is surjective (onto) since any element $x \in gH$ is equal to gh for some h, and there for the image of h under the map induced by g. To prove the map is injective (one-to-one), suppose $gh_1 = gh_2$ for some $h_1, h_2 \in H$. Taking the inverse of both sides

and using associativity, we have $h_1 = h_2$. This proves g is a bijection, and hence H and gH have the same number of elements.

Now for the second item, namely the set of cosets form a partition of G. It is easy to see that every element of $g \in G$ is in the coset gH, proving item (b). To show item (a), we prove the contrapositive: if $g_1 H \cap g_2 H \neq \emptyset$, then $g_1 H = g_2 H$. Assuming the intersection is nonempty, there exists an element g such that both $g = g_1 h_1$ and $g = g_2 h_2$ for some $h_1, h_2 \in H$. Hence, $g_1 = g_2 h_2 (h_1)^{-1} \in g_2 H$, since H is closed under multiplication. This implies $g_1 H \subset g_2 H$. But as we just showed, all cosets are the same size (for finite G). Thus, in fact, $g_1 H = g_2 H$.

Having established the claims above, it is straightforward to prove Lagrange's theorem. There are $|G|$ elements of G, partitioned into a finite number of cosets, each of size $|H|$. That is, $|G| = n|H|$, where n is the number of cosets. This proves the result.

Exercise A2.3

Show that the order of an element $g \in G$ divides $|G|$.

Solution: The order of an element g is the same as the size of the cyclic subgroup H generated by g. By Lagrange's theorem, the order must divide the size of the group.

Exercise A2.4

Show that if $y \in G_x$ then $G_y = G_x$.

Solution: Suppose $y \in G_x$. Then there exists some $g \in G$ such that $y = g^{-1}xg$. Hence, $gyg^{-1} = x$, which implies $x \in G_y$.

We will prove $G_y \subset G_x$. Let $z \in G_y$, so that $z = w^{-1}yw$ for some $w \in G$. Using the conjugacy relation between x and y,

$$z = w^{-1}(g^{-1}xg)w = u^{-1}xu \tag{A2.2}$$

where $u = gw \in G$. Thus, $z \in G_x$, so $G_y \subset G_x$.

We can prove $G_x \subset G_y$ by the argument of the previous paragraph, interchanging x with y and g with g^{-1} wherever they appear. This implies $G_x = G_y$.

Exercise A2.5

Show that if x is an element of an Abelian group G then $G_x = \{x\}$.

Solution: Since $x = exe$, x is in its own conjugacy class. Thus, G_x is nonempty. Let $y \in G_x$, so that $y = g^{-1}xg$ for some $g \in G$. Since the group is Abelian,

$$g^{-1}xg = g^{-1}gx \tag{A2.3}$$
$$= ex \tag{A2.4}$$
$$= x \tag{A2.5}$$

Hence, $y = x$. This implies any element of G_x is x itself, so $G_x = \{x\}$.

Exercise A2.6

Show that any group of prime order is cyclic.

Solution: Let G be a group of prime order. Then, $|G| > 1$, and therefore $G \setminus \{e\}$ is nonempty. Let $g \in G \setminus \{e\}$ and consider the cyclic subgroup $H = \langle g \rangle$. By Lagrange's theorem, $|H|$ divides $|G|$. However, since $|G|$ is prime, it must be that $|H| = 1$ or $|H| = |G|$. It cannot be that $|H| = 1$, since H contains e and g. Hence, $|H| = |G|$. Since $H \subset G$ we have $H = G$. Thus, G is cyclic.

Exercise A2.7

Show that every subgroup of a cyclic group is cyclic.

Solution:

Exercise A2.8

Show that if $g \in G$ has finite order r, then $g^m = g^n$ if and only if $m = n \pmod r$.

Solution: If $m = n$, the statement is true automatically. Otherwise, $m < n$ or $m > n$. Assume the former without loss of generality. Then, multiplying both sides by g^{-m},

$$e = g^{n-m} \tag{A2.6}$$

By Euclid's division lemma, there exist integers q, s such that

$$n - m = qr + s \tag{A2.7}$$

where $0 \leq s < r$. Substituting this into equation (A2.6),

$$e = g^{qr}g^s \tag{A2.8}$$
$$e = (g^r)^q g^s \tag{A2.9}$$
$$e = e^q g^s \tag{A2.10}$$
$$e = g^s. \tag{A2.11}$$

Since r is the order by assumption, and $s < r$ it cannot be that $s > 0$ or else s would be the order of g. This implies $s = 0$. Hence, $r|(n - m)$ and $m = n \pmod r$.

Exercise A2.9

Cosets define an equivalence relation between elements. Show that $g_1, g_2 \in G$ are in the same coset of H in G if and only if there exists some $h \in H$ such that $g_2 = g_1 h$.

Solution: Suppose $g_2 = g_1 h$ for some h. Then, by definition, $g_2 \in g_1 H$. Moreover, $g_1 = g_1 e \in g_1 H$. Thus, g_1 and g_2 are in the same coset.

Conversely, suppose g_1 and g_2 are in the same coset, so that there exists some $g \in G$ such that $g_1, g_2 \in gH$. This implies there exist $h_1, h_2 \in H$ such that $g_1 = gh_1$ and $g_2 = gh_2$. Inverting the second of these relations

gives $g = g_2 h_2^{-1}$, and substituting this into the relation for g_1 gives

$$g_1 = (g_2 h_2^{-1}) h_1 \tag{A2.12}$$

$$g_1 = g_2 h^{-1}, \tag{A2.13}$$

where $h = h_1^{-1} h_2 \in H$. Thus, $g_2 = g_1 h$.

Exercise A2.10

How many cosets of H are there in G?

Solution: All cosets are the same size $|H|$ and form a partition of the total group G (see Exercise A2.2). Hence, the number of cosets n satisfies

$$n = \frac{|G|}{|H|}. \tag{A2.14}$$

Exercise A2.11: (Characters)

Prove the properties of characters given above.

Solution: We will actually assume the result of the subsequent exercise: any (finite) matrix group is equivalent to a unitary matrix group. Hence, we can prove the statements for unitary matrix groups, and the result follows in general since the equivalence preserves character. Note that the proof of unitary equivalence does not rely on the results of this exercise, so there is no circular logic.

(1): The identity element of a matrix group must be the identity matrix itself. The trace of the identity matrix is just the dimension n.

$$\text{tr}(\mathbb{1}) = \sum_{i=1}^{n} \delta_{ii} = n. \tag{A2.15}$$

(2): Let $g \in G$. By unitarity, g is diagonalizable, with eigenvalues $\lambda_n = e^{i\phi_k}$ for $\phi_k \in \mathbb{R}$ and $k = 1, ..., n$. Since the trace is the sum of the eigenvalues,

$$|\chi(g)| = \left| \sum_{k=1}^{n} \lambda_k \right| \tag{A2.16}$$

$$\leq \sum_{k=1}^{n} \left| e^{i\phi_k} \right| = n. \tag{A2.17}$$

Above, we made use of the triangle inequality and the fact that the eigenvalues have unit norm. This proves this part.

(3): Using the notation of the previous part, suppose that $|\chi(g)| = n$. Then we may write $\chi(g) = ne^{i\theta}$ for some phase $\theta \in \mathbb{R}$. This implies $\chi(g') = n$, where $g' = e^{-i\theta}g$. Since g is unitary, so is g', and by the previous relationship the eigenvalues λ_k' of g' are given by $e^{i\phi_k'}$, where $\phi_k' = \phi_k - \theta$.

We will show that $\phi_k' = 0$ for all $k = 1, ..., n$, which implies that $\phi_k = \theta$. This, in turn, means that all the eigenvalues of g are the same value $e^{i\theta}$, so that $g = e^{i\theta}I$ as desired. To proceed, let's observe that, if

$\text{tr}(g') = n$, we have

$$\sum_{k=1}^{n} e^{i\phi'_k} = \sum_{k=1}^{n} 1$$

$$\implies \sum_{k=1}^{n} (1 - e^{i\phi'_k}) = 0 \tag{A2.18}$$

The only way this equality can be satisfied is if $\phi'_k = 0$, as can be seen by considering, say, the real part of the above equation. This proves (3).

(4): Suppose $g_1, g_2 \in G$ are members of the same conjugacy class, so that they are related by some similarity transform

$$g_1 = x g_2 x^{-1} \tag{A2.19}$$

for some element $x \in G$. The cyclic property of the trace implies it is invariant under a similarity transform. That is,

$$\chi(g_1) = \text{tr}(x g_2 x^{-1}) = \text{tr}(x^{-1} x g_2) = \text{tr}(e g_2) = \text{tr}(g_2) = \chi(g_2) \tag{A2.20}$$

Thus, we see that the character is the same for any two members of the same conjugacy class.

(5): We make use of our equivalence to a unitary representation, to say $g^{-1} = g^{\dagger}$. Hence,

$$\chi(g^{-1}) = \text{tr}(g^{\dagger}) = \sum_{k=1}^{n} g_{kk}^* = \chi(g)^* \tag{A2.21}$$

(6): What is an algebraic number? I don't really know how to do this one.

Exercise A2.12: (Unitary matrix groups)

A unitary matrix group is comprised solely of unitary matrices (those which satisfy $U^{\dagger}U = I$). Show that every matrix group is equivalent to a unitary matrix group. If a representation of a group consists entirely of unitary matrices, we may refer to it as being a *unitary representation*.

Solution: Let G be a finite matrix group. Given an inner product $\langle \cdot, \cdot \rangle$, we wish to construct a new inner product $\langle \cdot, \cdot \rangle'$ for which G is unitary. In general, we may relate $\langle \cdot, \cdot \rangle$ and $\langle \cdot, \cdot \rangle'$ via a positive definite matrix P.

$$\langle \cdot, \cdot \rangle' = \langle \cdot, P \cdot \rangle \tag{A2.22}$$

The property we want from P is that, for any $g \in G$,

$$g^{\dagger} P g = P. \tag{A2.23}$$

We imagine that, whatever P is, it must be expressed in terms of the group G itself. The simplest way to construct positive matrices from G is to take outer products of the form $g^{\dagger}g$. We know such constructions are strictly positive, since g is never the zero matrix. Moreover, positive linear combinations of positive matrices are themselves positive, so we might consider trying something of the form

$$P = \sum_{g \in G} c_g g^{\dagger} g, \quad c_g \geq 0. \tag{A2.24}$$

To satisfy condition (A2.23), we recall that each element $g \in G$ acts as a permutation on the group via the group multiplication. Because sums over a set are invariant under permutations of the set, we recognize that what we need is $c_g = 1$ for all g. Let's explicitly check that this works.

$$g^\dagger P g = \sum_{g' \in G} g^\dagger g'^\dagger g' g = \sum_{g' \in G} (g'g)^\dagger (g'g) = \sum_{g \in G} g^\dagger g = P. \tag{A2.25}$$

Thus, we have proven our result. As a finally note, if we wish to transform the group itself instead of the inner product of the space, we observe that we can think instead of acting on our vector space with the transformation $\sqrt{P}$, since $\langle \cdot, P \cdot \rangle = \langle \sqrt{P}^\dagger \cdot, \sqrt{P} \cdot \rangle$. The group therefore transforms as the corresponding similarity transform.

$$g' = \sqrt{P} g \sqrt{P}^{-1} \tag{A2.26}$$

This new matrix group is isomorphic to the original, it has the same characters, and is unitary with respect to the original inner product. This proves our result.

Exercise A2.13

Show that every irreducible Abelian matrix group is one dimensional.

Solution: If A is an abelian matrix group, then by definition each element commutes with one another. This implies the collection A is simulaneously diagonalizable with the same unitary U. Thus there is an equivalent representation of A consisting of diagonal matrices. Diagonal is a special case of block diagonal, and we have multiple blocks unless the dimension is 1. Thus, the group is irreducible if and only if $\dim(A) = 1$.

Exercise A2.14

Prove that if ρ is an irreducible representation of G, then $|G|/d_{rho}$ is an integer.

Solution:

Exercise A2.15

Using the Fundamental Theorem, prove that characters are orthogonal, that is:

$$\sum_{i=1}^{r} r_i (\chi_i^p)^* (\chi_i^q) = |G| \delta_{pq} \quad \text{and} \quad \sum_{p=1}^{r} (\chi_i^p)^* (\chi_j^p) = \frac{|G|}{r_i} \delta_{ij}, \tag{A2.27}$$

where p, q and δ_{pq} have the same meaning as in the theorem, and χ_i^p is the value the character of the pth irreducible representation takes on the ith conjugacy class of G, and r_i is the size of the ith conjugacy class.

Solution:

Exercise A2.16

S_3 is the group of permutations of three elements. Suppose we order these as mapping 123 to: 123; 231; 312; 213; 132, and 321, respectively. Show that there exist two one-dimensional irreducible representations

of S_3, one of which is trivial, and the other of which is 1, 1, 1, -1, -1, -1, corresponding in order to the six permutations given earlier. Also show that there exists a two-dimensional irreducible representation, with the matrices

$$
\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \frac{1}{2}\begin{bmatrix} -1 & -\sqrt{3} \\ \sqrt{3} & -1 \end{bmatrix}, \quad \frac{1}{2}\begin{bmatrix} -1 & \sqrt{3} \\ -\sqrt{3} & -1 \end{bmatrix},
$$
$$
\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \frac{1}{2}\begin{bmatrix} 1 & \sqrt{3} \\ \sqrt{3} & -1 \end{bmatrix}, \quad \frac{1}{2}\begin{bmatrix} 1 & -\sqrt{3} \\ -\sqrt{3} & -1 \end{bmatrix}. \tag{A2.28}
$$

Verify that the representations are orthogonal.

Solution:

Exercise A2.17

Prove that the regular representation is faithful.

Solution:

Exercise A2.18

Show that the character of the regular representation is zero except on the representation of the identity element, for which $\chi(I) = |G|$.

Solution:

Exercise A2.19

Use Theorem A2.5 to show that the regular representation contains d_{ρ^p} instances of each irreducible representation ρ^p. Thus, if R denotes the regular representation, and $\hat{G}$ denotes the set of all inequivalent irreducible representations, then

$$
\chi_i^R = \sum_{\rho \in \hat{G}} d_\rho \chi_i^\rho \tag{A2.29}
$$

Solution:

Exercise A2.20

The character of the regular representation is zero except for the conjugacy class i containing e, the identity element in G. Show, therefore, that

$$
\sum_{\rho \in \hat{G}} d_\rho \chi^\rho(g) = N \delta_{ge} \tag{A2.30}
$$

Solution:

Exercise A2.21

Show that $\sum_{\rho \in \hat{G}} d_\rho^2 = |G|$.

Solution:

Exercise A2.22

Substitute (A2.10) into (A2.9) and prove that $\hat{f}(\rho)$ is obtained.

Solution:

Exercise A2.23

Let us represent an Abelian group G by $g \in [0, N-1]$, with addition as the group operation, and define $\rho_h(g) = \exp[-2\pi i g h / N]$ as the h representation of g. This representation is one-dimensional, so $d_\rho = 1$. Show that the Fourier transform relations for G are

$$\hat{f}(h) = \frac{1}{\sqrt{N}} \sum_{g=0}^{N-1} f(g) e^{-2\pi i g h / N} \quad \text{and} \quad f(h) = \frac{1}{\sqrt{N}} \sum_{g=0}^{N-1} \hat{f}(g) e^{2\pi i h g / N}. \tag{A2.31}$$

Solution:

Exercise A2.24

Using the results of Exercise A2.16, construct the Fourier transform over S_3 and express it as a 6×6 unitary matrix.

Solution:

Appendix 3

The Solovay–Kitaev theorem

Exercise A3.1:

In Chapter we made use of the distance measure $E(U, V) = \max_{|\psi\rangle} \|(U - V)|\psi\rangle\|$, where the maximum is over all pure states $|\psi\rangle$. Show that when U and V are single qubit rotations, $U = R_{\hat{m}}(\theta)$, $V = R_{\hat{n}}(\varphi)$, $D(U, V) = 2E(U, V)$, and thus it does not matter whether we use the trace distance or the measure $E(\cdot, \cdot)$ for the Solovay-Kitaev theorem.

Solution: First, let's consider $E(U, V)$. We have

$$E(U, V) = \max_{|\psi\rangle} \|(U - V)|\psi\rangle\| = \max_{|\psi\rangle} \|U(\mathbb{1} - U^\dagger V)|\psi\rangle\| = \max_{|\psi\rangle} \|(\mathbb{1} - W)|\psi\rangle\| \tag{A3.1}$$

where $W = U^\dagger V \in SU(2)$. Hence, there exists a unit vector $\hat{q}$ and angle $\alpha \in [0, 4\pi)$ such that $W = R_{\hat{q}}(\alpha)$. Let us proceed to calculate the norm in E for an arbitrary state $|\psi\rangle$.

$$\begin{aligned}
\|(\mathbb{1} - W)|\psi\rangle\| &= \sqrt{\langle\psi|(\mathbb{1} - W^\dagger)(\mathbb{1} - W)|\psi\rangle} \\
&= \sqrt{1 + 1 - \langle\psi|W^\dagger|\psi\rangle - \langle\psi|W|\psi\rangle} \\
&= \sqrt{2 - 2\operatorname{Re}(\langle W\rangle)}
\end{aligned} \tag{A3.2}$$

Here, $\langle W\rangle \equiv \langle\psi|W|\psi\rangle$. Recall that, since W is a single qubit rotation,

$$W = R_{\hat{q}}(\alpha) = \cos(\alpha/2)\mathbb{1} + i\sin(\alpha/2)(\hat{q} \cdot \sigma). \tag{A3.3}$$

Hence,

$$\operatorname{Re}(\langle W\rangle) = \operatorname{Re}(\cos(\alpha/2)) + \operatorname{Re}(i\sin(\alpha/2)\langle\hat{q} \cdot \sigma\rangle) = \cos(\alpha/2), \tag{A3.4}$$

where we used the fact that $\langle\hat{q} \cdot \sigma\rangle$ is purely real. Note that this result is independent of the chosen state $|\psi\rangle$. To summarize our findings

$$E(U, V) = \sqrt{2 - 2\cos(\alpha/2)} = 2\sin(\alpha/4) \tag{A3.5}$$

Now let's consider the trace distance D(U,V). First, note that for any operator X and unitary U, $|X| = |UX|$. Hence,

$$|U - V| = |U(\mathbb{1} - U^\dagger V)| = |\mathbb{1} - W| \tag{A3.6}$$

139

with W as before. Consider the eigenvalues of $|\mathbb{1} - W|$. For any operator X, if $lambda$ is an eigenvalue, then $|\lambda|$ is an eigenvalue of $|X|$. The eigenvalues of $\mathbb{1} - W$ are given by $1 - e^{\pm i\alpha/2}$, so the eigenvalues of $|\mathbb{1} - W|$ are

$$\left| 1 - e^{\pm i\alpha/2} \right| = 2|\sin(\pm\alpha/4)| = 2\sin(\alpha/4). \tag{A3.7}$$

Note that both eigenvalues are equal. Since the trace is the sum of the eigenvalues (weighted by multiplicity)

$$D(U, V) = \mathrm{tr}\,|\mathbb{1} - W| = 4\sin(\alpha/4). \tag{A3.8}$$

Comparing with (A3.5), we see that $D(U, V) = E(U, V)$, which is the desired result. Note that the arguments used depended crucially on the fact that we are in two dimensions (one qubit).

Exercise A3.2

Appendix 4

Number theory

Exercise A4.1 (Transitivity)

Show that if $a|b$ and $b|c$, then $a|c$.

Solution: The premises imply, by definition, that there exist integers j, k such that $b = aj$ and $c = bk$. Substituting the first of these two equations into the second, we see $c = ajk = al$, where $l = jk \in \mathbb{Z}$. Hence, there exists an integer, namely l, such that $c = al$, proving $a|c$.

Exercise A4.2

Show that if $d|a$ and $d|b$ then d also divides a linear combination of a and b, $ax + by$, where x and y are integers.

Solution: From the definition of $d|a$ and $d|b$, there exist integers j, k such that $a = dj$ and $b = dk$. Hence,

$$ax + by = djx + dky = d(jx + ky). \tag{A4.1}$$

Define $m = jx + ky \in \mathbb{Z}$. Then we see $ax + by = dm$. From the definition of dividing, we have $d|(ax + by)$.

Exercise A4.3

Suppose a and b are positive integers. Show that if $a|b$ then $a \leq b$. Conclude that if $a|b$ and $b|a$ then $a = b$.

Solution: Suppose that $a|b$. Then there exists some $k \in \mathbb{Z}$ such that $b = ak$. By the hypothesis that a and b are positive, it must be that $k > 0$. Hence, $k - 1 \geq 0$. This, of course, implies

$$b(k - 1) \geq 0 \tag{A4.2}$$

since, again, b is nonnegative (positive, in fact). The result we desire comes from basic manipulations of inequalities.

$$\begin{aligned} b(k - 1) \geq 0 &\implies bk \geq b \\ &\implies a \geq b \end{aligned} \tag{A4.3}$$

As an immediate corollary, if $a|b$ and $b|a$ (both being positive integers), we have $a \geq b$ and $b \geq a$. Of course, this implies $a = b$. Note that the assumption of positivity was crucial for the proof to hold, and indeed, it is easy to see how it can be broken if negative numbers are included.

Exercise A4.4

Find the prime factorizations of 697 and 36 300.

Solution: I do not pretend to solve these in any mechanical fashion. Looking online, I notice 697 is a product of 41 and 17. Each of these numbers are themselves prime, so the prime factorization is

$$697 = 41^1 17^1 \tag{A4.4}$$

Unlike the first, the second number is easier to do in your head. We can pull out two factors of 10 and easily prime factor those. Meanwhile, 363 is divisible by 3, and then if you cared to memorize some perfect squares, $121 = 11^2$. Altogether.

$$36300 = 2^2 3^1 5^2 11^2. \tag{A4.5}$$

Exercise A4.5

For p a prime prove that all integers in the range 1 to $p - 1$ have multiplicative inverses modulo p. Which integers in the range 1 to $p^2 - 1$ do not have multiplicative inverses modulo p^2?

Solution: For any integer $a \in [1, p - 1]$, a is coprime with p. Hence, a has an inverse modulo p. In the case of p^2, the only integer which is not coprime with p^2 in the range $[0, p^2 - 1]$ is p itself. Every other integer in the range has a multiplicative inverse.

Exercise A4.6

Find the multiplicative inverse of 17 modulo 24.

Solution: We seek an positive integer $n < 24$ such that $17 * n = 1 \bmod 24$. Without yet an efficient method, we can perform an exhaustive check by hand or with a computer. It turns out the answer is $n = 17$ itself.

Exercise A4.7

Find the multiplicative inverse of $n + 1$ modulo n^2, where n is any integer greater than 1.

Solution: The answer, which might be reasonably guessed (or not). Is $n - 1$.

$$(n + 1)(n - 1) = n^2 - 1 = 1 \pmod{n}^2. \tag{A4.6}$$

Exercise A4.8 (Uniqueness of the inverse)

Suppose b and b' are multiplicative inverses of a, modulo n. Prove that $b = b' \mod n$.

Solution: If b and b' are both inverses of a, then $ab = ab' \pmod n$. This implies

$$a(b - b') = 0 \pmod n. \tag{A4.7}$$

From this, we conclude $n|a(b - b')$. But we also know, by Corollary A4.4, that n and a are coprime. Hence, $n|(b - b')$, so $b = b' \pmod n$.

Exercise A4.9

Explain how to find $\gcd(a, b)$ if the prime factorizations of a and b are known. Find the prime factorizations of 6825 and 1430, and use them to compute $\gcd(6825, 1430)$.

Solution: If the prime factorization of a and b are known, simply find the largest set (counting multiplicity) of shared prime factors.

The prime factorization of 6285 and 1430 are $3^1 5^1 15^1 419^1$ and $2^1 5^1 11^1 13^1$ respectively. The only shared prime factor is 5, hence this is also the gcd.

Exercise A4.10

What is $\varphi(187)$?

Solution: The prime factorization of 187 is 11×17. Hence,

$$\varphi(187) = \varphi(17 \times 11) = \varphi(17)\varphi(11) = 16 \times 10 = 160 \tag{A4.8}$$

Exercise A4.11

Problem: Prove that

$$n = \sum_{d|n} \varphi(d) \tag{A4.9}$$

where the sum is over all positive divisors d of n, including 1 and n. (*Hint:* Prove the result for $n = p^\alpha$ first, then use the multiplicative property (A4.22) of φ to complete the proof.

Solution: Follow the advice of the hint, suppose $n = p^\alpha$ where p is prime. The divisors of n are p^j,

where $0 \leq j \leq \alpha$. Hence, starting from the right hand side,

$$
\begin{aligned}
\sum_{d|n} \varphi(d) = \sum_{j=0}^{\alpha} \varphi(p^j) &= 1 + \sum_{j=1}^{\alpha} p^{j-1}(p-1) \\
&= 1 + (p-1)\sum_{j=1}^{\alpha} p^{j-1} \\
&= 1 + \sum_{j=0}^{\alpha} p^j - \sum_{j=0}^{\alpha} p^{j-1} \\
&= p^{\alpha},
\end{aligned}
\tag{A4.10}
$$

where, in the last step, all but p^{α} cancel from subtractions. This proves the result when n is a power of a prime.

To generalize the argument, we use the fundamental theorem of arithmetic, which says any $n \in \mathbb{Z}$ has a prime factorization.

$$
n = p_1^{\alpha_1} p_2^{\alpha_2} ... p_m^{\alpha_m}.
\tag{A4.11}
$$

Since all terms are powers of prime, they are coprime with each other, and we may use the multiplicative property of φ.

$$
\begin{aligned}
\varphi(n) &= \prod_{j=1}^{m} \varphi(p_j^{\alpha_j}) \\
&= \prod_{j=1}^{m} \sum_{k_j=0}^{\alpha_j} \varphi(p^{k_j})
\end{aligned}
\tag{A4.12}
$$

In the second step we used the first result derived above for powers of primes. By repeated use of the distributive property, the sum and product in the second line of (A4.12) can be reversed, and cast as a sum over m variables.

$$
\begin{aligned}
\prod_{j=1}^{m} \sum_{k_j=0}^{\alpha_j} \varphi(p^{k_j}) &= \sum_{k_1=0}^{\alpha_1} \sum_{k_2=0}^{\alpha_2} \cdots \sum_{k_m=0}^{\alpha_m} \prod_{j=1}^{m} \varphi(p^{k_j}) \\
&= \sum_{k_1=0}^{\alpha_1} \sum_{k_2=0}^{\alpha_2} \cdots \sum_{k_m=0}^{\alpha_m} \varphi(p^{k_1} p^{k_2} ... p^{k_m})
\end{aligned}
\tag{A4.13}
$$

A careful examination of this last equation reveals it is nothing more than a sum over all possible divisors d of n, expressed via the prime factorization. Hence,

$$
\varphi(n) = \sum_{d|n} \varphi(d)
\tag{A4.14}
$$

as desired.

Exercise A4.12

Verify that $\mathbf{Z}_n^*$ forms a group of size $\varphi(n)$ under the operation of multiplication modulo n.

Solution: That $\mathbf{Z}_n^*$ is a set of size $\varphi(n)$ follows directly from the definition of φ. Let $a, b \in \mathbf{Z}_n^*$, with inverses a^{-1}, b^{-1}. Then, the product ab has inverse $a^{-1}b^{-1}$, hence is in $\mathbf{Z}_n^*$ (note the order doesn't

matter since multiplication is commutative). Thus, the set is closed under the binary operator. Moreover, multiplication modulo n is associative. Finally, it is easy to see that $1 \in \mathbf{Z}_n^*$ (being its own inverse) and it acts as the identity operator. Of course inverses exist, by definition, therefore we have shown that $\mathbf{Z}_n^*$ satisfies the properties of a group under multiplication modulo n.

Exercise A4.13

Let a be an arbitrary element of $\mathbf{Z}_n^*$. Show that $S \equiv \{1, a, a^2, ...\}$ forms a subgroup of $\mathbf{Z}_n^*$, and that the size of S is the least value of r such that $a^r = 1 \pmod{n}$.

Solution: For any finite group G, if I take a single element $g \in G$ and generate a subset $S \subset G$ by repeatedly multiplying g by itself, the result will be a subgroup (when I include the induced binary operation). More generally, I can have multiple generators $g_1, g_2, ..., g_m$ and the result will still be a subgroup. Note this does not hold for infinite groups such as $\mathbb{Z}$, unless we allow negative exponents.

If r is the smallest positive integer satisfying $a^r = 1 \pmod{n}$, it follows that each a^i is unique for $i = 0, 1, ..., r-1$. Otherwise, $a^i = a^j$ for some $i, j < r$, which implies $a^{j-i} = 1$. This contradicts the assertion that r is the *least* such value. Hence, S has at least r values. In fact, it cannot have more than r unique values, since for any $k > r$ we have

$$k = qr + i \tag{A4.15}$$

for some $q \in \mathbb{Z}^+$ and $i < r$. But this will give the same power of a as i does.

$$a^k = a^{qr+i} = (a^r)^q a^i = 1^q a^i = a^i \tag{A4.16}$$

Here all powers are taken modulo n. Thus, S has r elements.

Exercise A4.14

Suppose g is a generator for $\mathbf{Z}_n^*$. Show that g must have order $\varphi(n)$.

Solution: If g generates $\mathbf{Z}_n^*$, then every $a \in \mathbf{Z}_n^*$ must be some power of g. Hence, $\mathbf{Z}_n^*$ is cyclic. By the results from the previous exercise, the size of $\mathbf{Z}_n^*$, which is $\varphi(n)$ must equal the order of the generator g.

Exercise A4.15

Lagrange's theorem (Theorem A2.1 on page 610) is an elementary result of group theory stating that the size of a subgroup must divide the order of the group. Use Lagrange's theorem to provide an alternative proof of Theorem A4.9, that is, show that $a^{\varphi(n)} = 1 \pmod{n}$ for any $a \in \mathbf{Z}_n^*$.

Solution: Consider the subgroup $A \subset G$ generated by a. Then the size of A is the order of a, say, r. By Lagrange's theorem, r must divide $\varphi(n)$, the size of $\mathbf{Z}_n^*$. That is, $\varphi(n) = kr$ for some $k \in \mathbb{Z}^+$. Given this,

$$a^{\varphi(n)} = a^{kr} = (a^r)^k = 1^k = 1 \tag{A4.17}$$

where all values are taken modulo n. This proves Euler's generalization of the little theorem.

Exercise A4.16

Use Theorem A4.9 to show that the order of x modulo N must divide $\varphi(N)$.

Solution: This follows directly from Lagrange's theorem (see the previous cluster of exercises). We've already shown that the size of a cyclic subgroup of $\mathbf{Z}_n^*$ is the order of a generating element. Lagrange's theorem says this order r must divide the size of the larger group $\mathbf{Z}_n^*$, which is $\varphi(n)$.

Exercise A4.17 (Reduction of order-finding to factoring)

We have seen that an efficient order-finding algorithm allows us to factor efficiently. Show that an efficient factoring algorithm would allow us to efficiently find the order modulo N of any x co-prime to N.

Solution:

Exercise A4.18

Find the continued fraction expansion for $x = 19/17$ and $x = 77/65$

Solution: In both cases we apply the repeated fraction algorithm. The case $x = 19/17$ is only a few steps.

$$\frac{19}{17} = 1 + \frac{2}{17} = 1 + \frac{1}{\frac{17}{2}} = 1 + \frac{1}{8 + \frac{1}{2}} \tag{A4.18}$$

For the case $x = 77/65$, we have to work a little harder. Here are the intermediate steps.

$$\begin{aligned}
77/65 &= 1 + 12/65 \\
65/12 &= 5 + 5/12 \\
12/5 &= 2 + 2/5 \\
5/2 &= 2 + 1/2.
\end{aligned} \tag{A4.19}$$

Hence the result is

$$\frac{77}{65} = 1 + \frac{1}{5 + \frac{1}{2 + \frac{1}{2 + \frac{1}{2}}}} \tag{A4.20}$$

Exercise A4.19

Show that $q_n p_{n-1} - p_n q_{m-1} = (-1)^n$ for $n \geq 1$. Use this fact to conclude that $\gcd(p_n, q_n) = 1$. (*Hint:* Induct on n.)

Solution: As the hint suggests, we proceed by induction on n. In the case $n = 1$, using the definitions provided in the text,

$$q_1 p_0 - p_1 q_0 = a_1 a_0 - (1 + a_0 a_1)1 = -1 \tag{A4.21}$$

as desired. By inductive hypothesis, assume the statement holds for $n = m$. Then, using the recursive definition for p and q,

$$q_{m+1}p_m - p_{m+1}q_m = (a_{m+1}q_m + q_{m-1})\,p_m - (a_{m+1}p_m + p_{m+1})\,q_m \tag{A4.22}$$

$$= \cancel{a_{m+1}q_m p_m} + q_{m-1}p_m - \cancel{a_{m+1}p_m q_m} - p_{m-1}q_m \tag{A4.23}$$

$$= -(q_m p_{m-1} - p_m q_{m-1}) \tag{A4.24}$$

$$= (-1)^m + 1, \tag{A4.25}$$

where in the last step we invoked the inductive hypothesis. Hence, the statement also holds for $n = m + 1$. By induction, the statement holds for all $n \geq 1$.

Note that the result may be reexpressed as

$$(-1)^n(q_n p_{n-1} - p_n q_{n-1}) = 1. \tag{A4.26}$$

By Theorem A4.2, we must have $\gcd(p_n, q_n) = 1$.

Problem 4.1 (Prime number estimate)

Let $\pi(n)$ be the number of prime numbers which are less than n. A difficult-to-prove result known as the *prime number theorem* asserts that $\lim_{n\to\infty} \pi(n)\log(n)/n = 1$ and thus $\pi(n) \approx n/\log(n)$. This problem gives a poor man's version of the prime number theorem which gives a pretty good lower bound on the distribution of prime numbers.

(1) Prove that $n \leq \log\binom{2n}{n}$.

Solution to (1): Note this is equivalent proving $2^n \geq \binom{2n}{n}$ (note the logarithm is base two). By definition,

$$\binom{2n}{n} = \frac{2n!}{n!n!} = \prod_{i=1}^{n} \frac{n+i}{i}. \tag{A4.27}$$

Moreover, for each i in the product, $(n+i)/i = 1 + n/i \geq 2$. Hence,

$$\prod_{i=1}^{n} \frac{n+i}{i} \geq \prod_{i=1}^{n} \frac{n+i}{i} = 2^n. \tag{A4.28}$$

This proves the result.

(2) Show that

$$\log\binom{2n}{n} \leq \sum_{p\leq 2n} \left\lfloor \frac{\log(2n)}{\log p} \right\rfloor \log p \tag{A4.29}$$

where the sum is over all primes p less than or equal to $2n$.

Solution to (2): This one is hard! I could rewrite the problem as showing

$$\binom{2n}{n} \leq \prod_{p<2n} e^{\log p \left\lfloor \frac{\log(2n)}{\log p} \right\rfloor} \tag{A4.30}$$

(3) Use the previous two results to show that

$$\pi(2n) \geq \frac{n}{\log(2n)} \tag{A4.31}$$

Solution to (3): From the previous two parts, we have

$$n \leq \sum_{p \leq 2n} \left\lfloor \frac{\log(2n)}{\log p} \right\rfloor \log p. \tag{A4.32}$$

Moreover, for any two positive real numbers x and y,

$$\lfloor x \rfloor y \leq \lfloor xy \rfloor. \tag{A4.33}$$

Hence,

$$\sum_{p \leq 2n} \left\lfloor \frac{\log(2n)}{\log p} \right\rfloor \log p \leq \sum_{p \leq 2n} \left\lfloor \frac{\log(2n)}{\log p} \log p \right\rfloor$$

$$\leq \sum_{p \leq 2n} \lfloor \log(2n) \rfloor \tag{A4.34}$$

$$\leq \log(2n)\pi(2n)$$

From this, we have $n \leq \log(2n)\pi(2n)$, and we have our result from rearranging.

Appendix 5

Public key cryptography and the RSA cryptosystem

Exercise A5.1

Written examples of the application of RSA tend to be rather opaque. It's better to work through an example yourself. Encode the word 'QUANTUM' (or at least the first few letters!), one letter at a time, using $p = 3$ and $q = 11$. Choose appropriate values for e and d, and use a representation of English text involving 5 bits per letter.

   Solution: There is choice in how we represent the letters, but a natural one is to label from 1 to 26. In this representation, we have

$$
\begin{aligned}
Q &= 17 = 10001 \\
U &= 21 = 10101 \\
A &= 01 = 00001 \\
N &= 14 = 01110 \\
T &= 20 = 10100 \\
M &= 13 = 01101.
\end{aligned}
\tag{A5.1}
$$

In our encoding, the message is 35 bits in length, and given by

$$
S = 10001101010000101110101001010101101 \tag{A5.2}
$$
$$
= 18959782573 \tag{A5.3}
$$

where we converted to decimal in the last step. Next we choose an odd number e relatively prime to $\phi(n) = (p-1)(q-1) = 20$. We will choose $e = 9$. To compute the multiplicative inverse modulo 20, d, we employ Euler's algorithm. Following the steps outlined in appendix 4,

$$
\begin{aligned}
20 &= 2 \times 9 + 2 \\
9 &= 4 \times 2 + 1 \\
2 &= 2 \times 1.
\end{aligned}
\tag{A5.4}
$$

Now we back substitute to find coefficients x, y such that $1 = 9x + 20y$.

$$
\begin{aligned}
1 &= 9 - 4 \times 2 \\
&= 9 - 4 \times (20 - 2 \times 9) \\
&= 9 - 4 \times 20 + 8 \times 9 \\
&= 9 \times 9 - 4 \times 20.
\end{aligned}
\tag{A5.5}
$$

Reading off the coefficient, we can readily see that 9 is its own inverse modulo 20. Hence $d = e = 9$.

Alas, with such a small n we can only encode in 5 bit chunks. We'll therefore simply encode each letter separately. We have

$$
\begin{aligned}
E(Q) &= 17^9 \quad (\text{mod } 33) = 02 = 00010 & \tag{A5.6} \\
E(U) &= 21^9 \quad (\text{mod } 33) = 21 = 10101 & \tag{A5.7} \\
E(A) &= 01^9 \quad (\text{mod } 33) = 01 = 00001 & \tag{A5.8} \\
E(N) &= 14^9 \quad (\text{mod } 33) = 26 = 00010 & \tag{A5.9} \\
E(T) &= 20^9 \quad (\text{mod } 33) = 05 = 00101 & \tag{A5.10} \\
E(M) &= 13^9 \quad (\text{mod } 33) = 28 = 11100 & \tag{A5.11} \\
& & \tag{A5.12}
\end{aligned}
$$

You can readily check, as expected, that taking the encoded message to the power of 9 (in 5 bit chunks) gets you back to the original message.

Exercise A5.2

Show that d is also an inverse of e modulo r, and thus $d = d'$ $(\text{mod } r)$.

Solution: We will prove a somewhat more general result, namely if $ab = 1$ $(\text{mod } n)$ and $d|n$, then $ab = 1$ $(\text{mod } d)$. Indeed, the first statement implies $ab = qn + 1$ for some $q \in \mathbb{Z}$. On the other hand, since $d|n$, there is an integer k such that $n = dk$. Using these relations, we have $ab = q(dk) + 1 = (qk)d + 1$. Thus, $ab = 1$ $(\text{mod } d)$.

This solves the exercise when we recognize that $de = 1$ $(\text{mod } \phi(n))$ and $r|\phi(n)$. A result from the previous appendix shows that the two inverses d and d' are equivalent modulo r.

Problem 5.1:

Write a computer program for performing encryption and decryption using the RSA algorithm. Find a pair of 20 bit prime numbers and use them to encrypt a 40 bit message.

Solution: I will first write pseudocode, then give an actual implementation in a common language such as python. Here is some pseudocode for the two major subroutines employed: RandomPrime and InverseMod.

Algorithm 1: RSA algorithm for public key cryptography

---

1 RSA algorithm (L, M);

    Input : An integer L specifying bit length of primes, and a $2L$-bit message M.

    Output: A public key $P = (e, n)$ and private key $M = (d, n)$.

2 $p = \text{RandomPrime}(L)$;

3 q $= \text{RandomPrime}(L)$;

4 $n = pq$;

5 $\varphi = (p - 1)(q - 1)$;

6 $d = \text{InverseMod}(e, \varphi)$;

7 $P = (e, n)$;

8 $S = (d, n)$;

9 return P, S

Algorithm 2: Algorithm for producing random prime p of given length.

---

1 RandomPrime (L);

    Input : An integer L specifying the bit length of the desired prime

    Output: A random prime p of that length

2 p $= \text{RandomInt}(L)$;

3 **do**

4 | p $= \text{RandomInt}(L)$;

5 **while** *not prime(p)*;

Appendix 6

Proof of Lieb's theorem

Exercise A6.1 ($\leq$ is preserved under conjugation)

If $A \leq B$, show that $XAX^\dagger \leq XBX^\dagger$ for all matrices X.

Solution: We will first prove that positivity is preserved under conjugation with X. That is, if P is positive, so is $XPX^\dagger$. Taking the adjoint shows that $XPX^\dagger$ is hermitian.

$$(XPX^\dagger)^\dagger = (X^\dagger)^\dagger P^\dagger X^\dagger = XPX^\dagger \tag{A6.1}$$

To prove positive-semidefiniteness, suppose λ is a (real) eigenvalus of $XPX^\dagger$, so that there is a normalized vector v such that.

$$XPX^\dagger v = \lambda v \tag{A6.2}$$

Taking the inner product of both sides of this equation with v itself,

$$\begin{aligned}
\langle v, XPX^\dagger v \rangle &= \langle v, \lambda v \rangle \\
\langle X^\dagger v, PX^\dagger v \rangle &= \lambda \langle v, v \rangle \\
\langle u, Pu \rangle &= \lambda,
\end{aligned} \tag{A6.3}$$

where $u = X^\dagger v$. Because P is positive semidefinite, we see that $\lambda \geq 0$. Hence, every eigenvalue of $XPX^\dagger$ is nonnegative. This proves our result.

Exercise A6.2

Prove that $A \geq 0$ if and only if A is a positive operator.

Solution: If $A \geq 0$, then $A - 0$ is positive semidefinite, hence so is A. Conversely, if A is positive, so is $A - 0$, and thus $A \geq 0$.

Exercise A6.3 ($\leq$ is a partial order)

Show that the relation $\leq$ is a partial order on operators – that is, it is transitive ($A \leq B$ and $B \leq C$ implies $A \leq C$), asymmetric ($A \leq B$ and $B \leq A$ implies $A = B$), and reflexive ($A \leq A$).

Solution: Let's start by proving transitivity. If $A \leq B$ and $B \leq C$, then $B - A$ and $C - B$ are positive matrices. Hence so is their sum, $C - A$. This implies $A \leq C$ by definition.

To prove asymmetry, suppose $A \leq B$ and $B \leq$. Let λ be an eigenvalue of $A - B$. It is then clear that $B - A$ must have eigenvalue $-\lambda$. By assumption of positive semidefiniteness of $A - B$ and $B - A$ we must have

$$\lambda \leq 0 \quad lambda \geq 0. \tag{A6.4}$$

Hence, $\lambda = 0$. Thus every eigenvalue of $A - B$ is zero, so $A - B = 0$. This proves asymmetry.

Finally, we note that $A - A = 0$ is positive semidefinite. Thus, $A \leq A$, proving the reflexive property.

Exercise A6.4

Suppose A has eigenvalues λ_i. Define λ to be the maximum of the set $|\lambda_i|$. Prove that

(1) $\|A\| \geq \lambda$.
(2) When A is Hermitian, $\|A\| = \lambda$.
(3) When

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \tag{A6.5}$$

$\|A\| = 3/2 > 1 = \lambda$

Solution: (1) Since A is a matrix, its set of eigenvalues is finite. Hence, there exists an eigenvalue λ_m such that $|\lambda_m| = \lambda$. Let $|u_m\rangle$ be the corresponding eigenvector, normalized. Then,

$$|\langle u_m| A |u_m\rangle| = |\lambda_m \langle u_m|u_m\rangle| = \lambda \tag{A6.6}$$

Since $\|A\|$ is the maximum over all such inner products, it is certainly at least as big as the value set by $|u\rangle = |u_m\rangle$. Hence, $\|A\| \geq \lambda$.

(2) Using part (1), it suffices to show that $\|A\| \leq \lambda$ for hermitian A. If $|u\rangle$ is a normalized state, it can be expressed as a linear combination in an orthonormal basis defined by the eigenstates of A.

$$|u\rangle = \sum_i c_i |\lambda_i\rangle \tag{A6.7}$$

Here, $|\lambda_i\rangle$ is an eigenstate of A with eigenvalue λ_i. Computing the inner product as in the defition of $\|A\|$,

$$|\langle u| A |u\rangle| = \left| \sum_i \lambda_i |c_i|^2 \right| \leq \sum_i |\lambda_i||c_i|^2 \leq \lambda \sum_i |c_i|^2 = \lambda \tag{A6.8}$$

Along the way, we used the triangle inequality, the fact that $|\lambda_i| \leq \lambda$, and the normalization of $|u\rangle$. Since λ is an upper bound for every $|u\rangle$, it is also an upper bound for the maximum, which is precisely $\|A\|$. Thus, $\|A\| \leq \lambda$, which combined with the previous result gives $\|A\| = \lambda$.

(3) A has a single eigenvalue $\lambda = 1$ with eigenvector $|\lambda\rangle = (0, 1)^T$. Hence, $\lambda = 1$. On the other hand, for some normalized $(a, b) \in \mathbb{C}^2$,

$$\begin{pmatrix} a^* & b^* \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = |a|^2 + |b|^2 + ab^* \tag{A6.9}$$

$$= 1 + ab^* \tag{A6.10}$$

If $a = b = 1/\sqrt{2}$, then this value is $3/2$. On the other hand, for more arbitrary complex values of a and b, the triangle inequality puts $3/2$ as an upper bound on the magnitude.

$$|1 + ab^*| \leq 1 + |a||b| \leq 3/2. \tag{A6.11}$$

Hence, $\|A\| = 3/2$, and we have our result.

Exercise A6.5: (AB and BA have the same eigenvalues)

Prove that AB and BA have the same eigenvalues. (*Hint*: For invertible A, show that $\det(xI - AB) = \det(xI - BA)$, and thus the eigenvalues of AB and BA are the same. By continuity this holds even when A is not invertible.

Solution: As the hint suggests, first suppose A is invertible. Because the eigenvalues of AB and BA are the zeros of the characteristic polynomial, showing these two polynomials are the same amounts to proving the statement. And these polynomials are precisely the determinants shown in the hint.

We will use the property that the determinant is indifferent to the permutation of matrices in a product.

$$\det(xI - AB) = \det(I(xI - AB)) \tag{A6.12}$$
$$= \det(A^{-1}A(xI - AB)) \tag{A6.13}$$
$$= \det(A^{-1}(xI - AB)A) \tag{A6.14}$$
$$= \det(xI - BA) \tag{A6.15}$$

This proves our result when A is invertible. If A is singular, then there exists an $\epsilon > 0$ such that

$$A' = A + \epsilon I \tag{A6.16}$$

is invertible. Then the theorem carries over as before for A', and to get the result for A we take $\epsilon \to 0$. This is valid since the determinant is only a polynomial in ϵ.

Exercise A6.6

Suppose A and B are such that AB is Hermitian. Using the previous two observations show that $\|AB\| \leq \|BA\|$.

Solution: Since AB is Hermitian, then $\|AB\| = |\lambda|$ for some eigenvalue of AB. By the previous exercise, λ is also an eigenvalue of BA, and by that same exercise we have $\lambda \leq \|BA\|$. Thus, $\|AB\| \leq \|BA\|$.

Exercise A6.7

Suppose A is positive. Show that $\|A\| \leq 1$ if and only if $A \leq I$.

Solution: ($\implies$) Suppose $\|A\| \leq 1$. Then every eigenvalue λ of A is such that $\lambda \in [0, 1]$. This implies the eigenvalues of $I - A$, which are given by $1 - \lambda$ are also in this range. In particular, $I - A$ is positive, so $A \leq I$.

($\impliedby$) Suppose $A \leq I$, so that $I - A$ is positive. As above, the eigenvalues of $I - A$ are $1 - \lambda$, where λ is an eigenvalue of A. Since both A and $I - A$ are positive, we have

$$1 - \lambda \geq 0 \quad \lambda \geq 0. \tag{A6.17}$$

This implies $\lambda \leq 1$ for each λ, so we have $\|A\| \leq 1$.

Exercise A6.8

Let A be a positive matrix. Define a superoperator (linear operator on matrices) by the equation $\mathcal{A}(X) \equiv AX$. Show that $\mathcal{A}$ is positive with respect to the Hilbert-Schmidt inner product. That is, for all X, $\mathrm{tr}\big(X^\dagger \mathcal{A}(X)\big) \geq 0$. Similarly, show that the superoperator defined by $\mathcal{A}(X) \equiv XA$ is positive with respect to the Hilbert–Schmidt inner product on matrices.

    Solution: Suppose A is positive, and let X be any complex matrix for which the multiplication AX is defined. Our goal is to show that $\mathcal{A}$ is positive, that is

$$\langle X, \mathcal{A}(X)\rangle \geq 0 \tag{A6.18}$$

with equality only when $X = 0$. The Hilbert-Schmidt inner product is defines as

$$\langle A, B\rangle = \mathrm{tr}\, A^\dagger B. \tag{A6.19}$$

Hence, we need to show

$$\mathrm{tr}\, X^\dagger A X > 0 \tag{A6.20}$$

for nonzero X. First of all, note that $X^\dagger A X$ is positive semi-definite. Why? For any vector ψ,

$$\langle \psi, X^\dagger A X \psi\rangle = \langle X\psi, AX\psi\rangle = \langle \psi', A\psi'\rangle \tag{A6.21}$$

where $\psi' := X\psi$. Since, A is positive, this expression is no less than zero. Going back to (A6.18), we simply note that the trace of any positive semi-definite operator is always at least zero. In fact, it is only zero if $X^\dagger A X = 0$, which can only happen in the very special case when $X = 0$. We've now shown that $\mathcal{A}$ is positive.