1. Approximating unitary transformations.

(Marker: AM)

(a) Let $|\psi\rangle$ be an arbitrary pure state. Then,

$$
\| (A - B) |\psi\rangle \| = \| (A - C + C - B) |\psi\rangle \| \\
= \| (A - C) |\psi\rangle + (C - B) |\psi\rangle \| \\
\leq \| (A - C) |\psi\rangle \| + \| (C - B) |\psi\rangle \| \\
\leq \| A - C \| + \| C - B \| 
$$

Therefore, $\| A - B \| \leq \| A - C \| + \| C - B \|$. 

To go from the second line to the third line in our equations, we apply the triangle inequality for vectors in a complex finite-dimensional inner product space. To go from the third line to the fourth line, we use the definition of spectral norm.

A common mistake in the solutions to this exercise concerned expressions of the form $\max_{|\psi\rangle} \langle \psi | X |\psi\rangle + \langle \psi | Y |\psi\rangle + f(|\psi\rangle)$, for some matrices $X$ and $Y$, and some function $f$. In some solutions, expressions of this kind were incorrectly written to be equal to $\| X \| + \| Y \| + \max_{|\psi\rangle} f(|\psi\rangle)$. This is not the case, since a value of $|\psi\rangle$ that maximizes $\langle \psi | X |\psi\rangle$ will generally not maximize $\langle \psi | Y |\psi\rangle$. Instead, one should write the inequality

$$
\max_{|\psi\rangle} \langle \psi | X |\psi\rangle + \langle \psi | Y |\psi\rangle + f(|\psi\rangle) \leq \| X \| + \| Y \| + \max_{|\psi\rangle} f(|\psi\rangle),
$$

using the definition of spectral norm.

(b) We can write any pure state $|\psi\rangle$ in $ml$ dimensions as $\sum_{i=0}^{l-1} |\psi_i\rangle |i\rangle$, where the normalization constraint of $|\psi\rangle$ corresponds to the condition $\sum_{i=0}^{l-1} \| |\psi_i\rangle \|^2 = 1$. Then, we have that
\[
\|(A \otimes I) |\psi\rangle\|^2 = \left( \sum_{i=0}^{l-1} \langle \psi_i | A^\dagger \otimes | i \rangle \right) \left( \sum_{j=0}^{l-1} A |\psi_j\rangle \otimes | j \rangle \right) = \sum_{i=0}^{l-1} \langle \psi_i | A^\dagger A |\psi_i\rangle = \sum_{i=0}^{l-1} \|A |\psi_i\rangle\|^2 \leq \sum_{i=0}^{l-1} (\|A\| \| |\psi_i\rangle\|)^2 \leq \sum_{i=0}^{l-1} (\|A\| \| |\psi_i\rangle\|)^2 = \|A\|^2 \sum_{i=0}^{l-1} \| |\psi_i\rangle\|^2 = \|A\|^2,
\]

and therefore \(\|A \otimes I\| \leq \|A\|\). To go from the first to the second line, we use that states in the computational basis are orthogonal to each other. To go from the third to the fourth line, we used that for a vector \(v\) and a scalar \(x\), \(\|xv\| = |x|\|v\|\). To go from the fourth line to the fifth line, we used the definition of the spectral norm. Note also that for any pure state \(|\psi\rangle\) in \(m\) dimensions, \((\langle\psi \otimes |0\rangle\langle\psi \otimes |0\rangle\rangle = \|A |\psi\rangle \otimes |0\rangle\| = \|A |\psi\rangle\|\), so we also have \(\|A \otimes I\| \geq \|A\|\).

A common error when trying to solve this exercise was only considering input states to \(A \otimes I\) that are tensor product states, rather than an arbitrary pure state \(|\psi\rangle\) in \(ml\) dimensions.

(c) Let \(|\psi\rangle\) be an arbitrary pure state. Then,

\[
\|U_1 U_2 |\psi\rangle\| = \sqrt{\langle\psi | U_2^\dagger A^\dagger U_1^\dagger U_1 U_2 |\psi\rangle} = \sqrt{\langle\psi | U_2^\dagger A^\dagger U_2 |\psi\rangle} = \|U_2 |\psi\rangle\|.
\]

Therefore, \(\|U_1 U_2\| = \|U_2\|\).

Let \(|\phi\rangle\) be equal to \(U_2 |\psi\rangle\). Then \(U_2 |\psi\rangle = A |\phi\rangle\). Conversely, for any state \(|\psi\rangle\), if we let \(|\phi\rangle\) be equal to \(U_2^{-1} |\psi\rangle\), then \(A |\psi\rangle = U_2 |\phi\rangle\). Therefore, the image of \(A\) is the same as the image of \(U_2\) when restricting the domain to pure states, and \(\|U_2\| = \|A\|\).
A common mistake here was simply pointing out that $AU_2|\psi\rangle = A|\phi\rangle$. This is enough to show that $\|AU_2\| \leq \|A\|$, but does not imply $\|AU_2\| = \|A\|$. To prove the equality, we also need to use that any pure state $|\psi\rangle$ can be written as $U_2|\phi\rangle$ for some pure state $|\phi\rangle$.

2. Approximate quantum Fourier transforms modulo $2^n$.

(Marker: AM)

(a) Using the definition of norm in Question 1, one observes that the value of $\|P_k - I\|$ is given by $|e^{2\pi i/2^k} - 1|$. This is the distance in the complex plane between 1 and the point in the unit circle $e^{2\pi i/2^k}$. This is less or equal than the distance traveling along the unit circle from 1 to $e^{2\pi i/2^k}$. By the definition of of an angle in radians, this distance along the unit circle is equal to $2\pi/2^k$.

One can alternatively evaluate the trigonometric expression corresponding to $|e^{2\pi i/2^k} - 1|$, and finish the proof by using either of the inequalities $\sin(x) \leq x$ and $\cos(x) \geq 1 - x^2/2$. These can be proved by noting that they hold with equality for $x = 0$, and then taking derivatives at both sides of the inequality sign.

(b) Let the total number of gates we remove be given by $R$, and fix an arbitrary ordering of the gates in the circuit. Then, after we have removed $0 \leq r \leq R$ gates, we can write $\tilde{F}_r^{2^n}$ for the resulting circuit. Note that $\tilde{F}_0^{2^n} = F_2^n$, and $\tilde{F}_R^{2^n} = \tilde{F}_2^n$. Using 1a) repeatedly, we can then write

$$\|\tilde{F}_r^{2^n} - F_2^n\| \leq \sum_{i=0}^{R-1} \|\tilde{F}_{i+1}^{2^n} - \tilde{F}_i^{2^n}\|.$$ 

Let us look now at one of the $\|\tilde{F}_{i+1}^{2^n} - \tilde{F}_i^{2^n}\|$ terms. We have that the corresponding circuits will differ in the replacement of a gate $R_j$ by $I$, for some value of $j$. Other than that, the parts left of the $R_j$ and right of the $R_j$ will be the same in both of the circuits. Let $V_1$ denote the unitary corresponding to the part of the circuit left of $R_j$, and $V_2$ the unitary for the part right of $R_j$. Then, with a suitable ordering of the qubits in the circuit, we can write $\tilde{F}_{i+1}^{2^n}$ as $V_2^iV_1$, and $\tilde{F}_i^{2^n}$ as $V_2(R_j \otimes I)V_1$. We have then that

$$\|\tilde{F}_{i+1}^{2^n} - \tilde{F}_i^{2^n}\| = \|V_2((I - R_j) \otimes I)V_1\| = \|I - R_j\| \leq 2\pi/2^j,$$

using properties 1b) and 1c), as well as the bound in 2a).

We have then that
\[
\|\tilde{F}_{2^n} - F_{2^n}\| \leq \sum_{i=0}^{R-1} 2\pi/2^i(i),
\]

where \(P_{j(i)}\) is the type of control gate removed in the \(i^{th}\) step. To evaluate this sum, we need to count the number of \(P_k\)'s that we remove for each value of \(k\). A visual inspection of the circuit for the QFT in slide 16 informs us that for each value of \(k\) such that \(n \geq k \geq t\), we will remove \(n - k + 1\) copies of \(P_k\). Therefore, we obtain the bound

\[
\|\tilde{F}_{2^n} - F_{2^n}\| \leq n \sum_{k=t}^{n} (n - k + 1) 2\pi/2^k \quad (15)
\]

\[
\leq 2\pi n \sum_{k=t}^{n} \frac{1}{2^k} \quad (16)
\]

\[
= 4\pi n \left( \frac{1}{2^t} - \frac{1}{2^n+1} \right) \quad (17)
\]

\[
\leq \frac{4\pi n}{2^t}. \quad (18)
\]

As an aside, we could have evaluated exactly the sum in the first line, and only then apply any arithmetic bounds. This process is a bit error-prone, and if doing so I would recommend using the summation formula for arithmetico-geometric progressions.

In any case, we have then that if we pick the first integer value of \(t\) greater or equal to \(\log(4\pi n/\epsilon)\), then \(\|\tilde{F}_{2^n} - F_{2^n}\| \leq \epsilon\), as we wanted. It remains then to prove that the number of gates will decrease enough when picking that value of \(t\). Indeed, a simple count gives us that the initial number of gates in the QFT circuit is \(n(n+1)/2\). Then, after our removal process the number of gates will be

\[
\frac{n(n+1)}{2} - \sum_{k=t}^{n} n - k + 1 = \sum_{i=n-t+1}^{n} i = \frac{(n-t+1)t}{2} \in O(nt).
\]

Since \(t \in O(\log(n/\epsilon))\), we have that the total number of gates is in \(O(n \log(n/\epsilon))\).


(Marker: AM)
(a) First, note that we can implement a one-qubit control $U$ by controlling each gate of $U$. Since each gate of $U$ is a 2-qubit gate, and there are $s$ gates, this will result in a circuit that uses $s$ 3-qubit gates.

Then, to implement the two-qubit controlled-$U$, we simply concatenate three copies of the one-qubit controlled-$U$. Two of them will be controlled on the qubit corresponding to the higher order bit $a$ of $ab$ (where, as in the rest of this explanation, $ab$ represents the number that we write as $ab$ in binary). The remaining copy of the one-qubit controlled-$U$ will be controlled on the qubit corresponding to the lower order bit $b$ of $ab$. This will then implement $(U^a)(U^a)(U^b) = U^{2s+a+b} = U^{ab}$.

A common mistake here was replacing each gate $U'$ of $U$ by three controlled $U'$ gates, rather than concatenating three one-qubit controlled copies of $U$. This does not work, due to the general non-commutativity of unitary operators. To see why, consider the simple case where $U$ consists of two gates $U_1$ and $U_2$, each applied to the same two qubits, and we have $ab = 11$. Then, our correct circuit will implement $U^3 = U_1U_2U_1U_2U_1U_2$. However, if we replace each gate of $U$ by three controlled copies of that gate, we would be implementing instead $U_1U_1U_1U_2U_2U_2U_2$. Since unitary operators are not commutative, this will generally not be equal to $U_1U_2U_1U_2U_2$.

(b) The circuit described in slide 27 will exactly estimate eigenvalues of the form $e^{2\pi i k/2^m}$, for $k \in \mathbb{Z}$. Since in this exercise all eigenvalues are of the form $e^{2\pi i k/4}$, if we pick $m = 2$ the circuit will estimate eigenvalues exactly. We can determine manually the circuit that implements the inverse of the QFT, by inverting each gate of the QFT individually, in reverse of their order of application. We obtain then the following circuit (credit: Morgan Mastrovich):

The $-4$ means that we are implementing a controlled phase shift gate with phase $\frac{2\pi}{4}$. As we saw before, the 2-qubit controlled-$U$ can be implemented with $3s$ 3-qubit gates. Since we also four one-qubit Hadamard gates and one two-qubit controlled-phase gate, we obtain the desired gate count.

A common mistake when trying to solve this exercise was not computing the inverse of the QFT correctly. In particular, the inverse of $H$ happens to be $H$ itself, since $H$ is both Hermitian and unitary. However, this is not generally true for unitary matrices. In particular, it is generally not the
case with controlled phase gates, because of having a non-real number in the diagonal. This means that it is not correct to use a controlled $\alpha$-phase shift gate to reverse a controlled $\alpha$-phase gate, and instead we should use a controlled $(-\alpha)$-phase shift gate.

(c) We apply an $\omega^2$-phase shift gate to the first qubit, and an $\omega$-phase shift gate to the second qubit.

This will implement then $\begin{pmatrix} 1 & 0 & 0 \\ 0 & \omega^2 & 0 \\ 0 & 0 & \omega \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 & 0 \\ 0 & \omega & 0 \\ 0 & 0 & \omega^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \omega & 0 & 0 \\ 0 & 0 & \omega^2 & 0 \\ 0 & 0 & 0 & \omega^3 \end{pmatrix}$, as desired.

4. Basic questions about density matrices.

(Marker: AM)

(a) Let $\rho$ correspond to a mixture of pure states $|\psi_i\rangle$, with probability $p_i$ each.

Then, we have:

$$\text{Tr}(\rho^2) = \text{Tr} \left( \sum_{i,j} p_ip_j |\psi_i\rangle \langle \psi_i| |\psi_j\rangle \langle \psi_j| \right) = \sum_{i,j} p_ip_j \text{Tr} (|\psi_i\rangle \langle \psi_i| |\psi_j\rangle \langle \psi_j|) = \sum_{i,j} p_ip_j |\langle \psi_i| \psi_j\rangle|^2 \leq \sum_{i,j} p_ip_j \leq \left( \sum_i p_i \right) \left( \sum_j p_j \right) = 1.$$  

We used the additive and cyclic properties of the trace to go from the first line to the second line, and from the second line to the third line, respectively. The inequality between the third and the fourth line will be an equality if and only if all terms $|\langle \psi_i| \psi_j\rangle|^2$ are equal to one. This happens if and only if there is only one $|\psi_i\rangle$, which is equivalent to $\rho$ being a pure state. This is because $|\langle \psi_i| \psi_j\rangle|^2 = 1$ holds for two pure states if and only if $|\psi_1\rangle$ and $|\psi_2\rangle$ are equal to each other (considering as the same two states written similarly up to a complex phase, as usual).
(b) Let $H$ be such an operator. Consider a spectral decomposition

$$H = \sum \lambda_i \langle \psi_i | \psi_i \rangle.$$ 

The fact that $H$ is positive semi-definite implies $\lambda_i \in \mathbb{R}$ and $\lambda_i > 0$. Since the trace is the sum of eigenvalues, $\sum \lambda_i = 1$. Therefore, the $\lambda_i$ form a valid probability distribution, and the set of all $(\lambda_i, |\psi_i\rangle)$ pairs represents a mixture of states that generates $H$.

(c) Consider a spectral decomposition

$$\rho = \lambda_1 |\phi_1\rangle \langle \phi_1 | + \lambda_2 |\phi_2\rangle \langle \phi_2 |.$$ 

Remember that in such a decomposition, $|\phi_1\rangle$ will be orthogonal to $|\phi_2\rangle$. Now, consider two arbitrary pure states $|\psi_1\rangle$ and $|\psi_2\rangle$, written in the $\{|\phi_1\rangle, |\phi_2\rangle\}$ basis as $a_1 |\phi_1\rangle + a_2 |\phi_2\rangle$ and $b_1 |\phi_1\rangle + b_2 |\phi_2\rangle$, respectively. Imposing the constraint $\rho = \frac{1}{2} |\psi_1\rangle \langle \psi_1 | + \frac{1}{2} |\psi_2\rangle \langle \psi_2 |$ will give us a system of six equations. Four of these correspond to the matrix entries of $\rho$, and the two others impose that $|\psi_1\rangle$ and $|\psi_2\rangle$ have norm equal to one. One can then observe that a solution of this system is given by $a_1 = \sqrt{\lambda_1}$, $a_2 = \sqrt{\lambda_2}$, $b_1 = \sqrt{\lambda_1}$, and $b_2 = -\sqrt{\lambda_2}$.

Alternatively, one can visualize $\rho$ as a point in the Bloch sphere, and situate it as the middle point of two points in the edge of the sphere, corresponding to two pure states $|\psi_1\rangle$ and $|\psi_2\rangle$, respectively. Then, one can use (for example) the correspondence between coordinates in the Bloch sphere and decompositions in the Pauli basis to show that $\rho = \frac{1}{2} |\psi_1\rangle \langle \psi_1 | + \frac{1}{2} |\psi_2\rangle \langle \psi_2 |$.

A common mistake here was just showing that all operators of the form $\frac{1}{2} |\psi_1\rangle \langle \psi_1 | + \frac{1}{2} |\psi_2\rangle \langle \psi_2 |$ are positive-semidefinite and have trace one. This reverses the direction of the fact that one needs to prove in this part of the exercise. That is, it proves that all matrices of that form are density matrices, instead of proving that all density matrices can be written in that form.

5. The density matrix is in the eye of the beholder.

(Marker: AM)

(a) The density matrix from Bob’s point of view will be $\cos^2(\pi/8) |0\rangle \langle 0 | + \sin^2(\pi/8) |1\rangle \langle 1 | = \begin{pmatrix} \cos^2(\pi/8) & 0 \\ 0 & \sin^2(\pi/8) \end{pmatrix}$. 

7
(b) When Alice prepares $|0\rangle \langle 0|$, Bob’s measurement will always result in the state $|0\rangle$, and his density matrix will then be $|0\rangle \langle 0|$ as well. Similarly, when Alice prepares $|1\rangle \langle 1|$, Bob’s measurement will always result in the state $|1\rangle$, and his density matrix will then be $|1\rangle \langle 1|$ as well.

We can see that in both cases, Bob’s density matrix agrees with Alice’s.

(c) The density matrix from Bob’s point of view will be $\cos^2(\pi/8) |\psi_0\rangle \langle \psi_0| + \sin^2(\pi/8) |\psi_1\rangle \langle \psi_1| = \begin{pmatrix} \cos^2(\pi/8) & 0 \\ 0 & \sin^2(\pi/8) \end{pmatrix}$.

We can see that this is the same matrix as in part a), so from the point of view of Bob the two mixtures of states are indistinguishable.

(d) Since from Bob’s point of view he is dealing with the same mixture $\rho$ as in 4b), and his measurement statistics only depend on $\rho$, Bob will again obtain the density matrix $|0\rangle \langle 0|$ with probability $\cos^2(\pi/8)$, and the density matrix $|1\rangle \langle 1|$ with probability $\sin^2(\pi/8)$.

We can see then that after Bob’s measurement, his density matrix is never the same as either of those $|\psi_0\rangle \langle \psi_0|$ and $|\psi_1\rangle \langle \psi_1|$ that could have been generated from Alice’s point of view. Note that it doesn’t agree with the density matrices corresponding to Alice’s expectation of Bob’s output measurement output, either. This is because no matter whether she generates $|\psi_0\rangle$ or $|\psi_1\rangle$, she expects Bob’s measurement to result in $|0\rangle$ with probability $\cos^2(\pi/8)$, and $|1\rangle$ with probability $\sin^2(\pi/8)$, which corresponds to the density matrix $\begin{pmatrix} \cos^2(\pi/8) & 0 \\ 0 & \sin^2(\pi/8) \end{pmatrix}$. We can see then how the difference between all of these density matrices represents Bob and Alice’s different knowledge regarding the outcome of the coin flip in the preparation procedure and the outcome of Bob’s measurement.