Nielsen and Chuang Solutions

Jacob Watkins

December 2020

1 Outline and plan

There exist other partial solution manuals to N&C, most of them on github. It appears, taken together, they have covered chapters 2,3,4 and 9 almost completely, with scattered solutions for other chapters. Here, we wish to help fill in the gaps, and perhaps ultimately create the most comprehensive solution manual to date. The strategy here is as follows:

- Create solution manuals for chapters 5-8
- Create solutions for chapters 10-12
- (If motivated) Fill in remaining problems in chapters already covered by others (in chapters 2-4 for example.)
- (If REALLY motivated) Compile together solutions already created, and bring them into a common format, so that we may come closer to a universal solutions manual!

2 To-do

- Fix QFT circuit, recreate it in Qiskit and set barrier option to True
- Double check circuit shown in problem 5.4

Chapter 5: The Quantum Fourier Transform and its applications

5.1

Problem: 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 $|j\rangle, |k\rangle$,

$$\langle j| (QFT)^\dagger (QFT) |k\rangle = \langle j|k\rangle = \delta_{ij}. \tag{1}$$

To do this, we substitute the definition into the above equation.

$$\begin{aligned}
\langle k| (QFT)^{\dagger}(QFT) |j\rangle &= \Big(\frac{1}{\sqrt{N}}\sum_{p=0}^{N-1} e^{-2\pi ikp/N} \langle p|\Big)\Big(\frac{1}{\sqrt{N}}\sum_{q=0}^{N-1} e^{2\pi ijq/N} |q\rangle\Big) \\
&= \frac{1}{N}\sum_{p=0}^{N-1}\sum_{q=0}^{N-1} e^{2\pi i(jq-kp)/N} \langle p|q\rangle \qquad (2) \\
&= \frac{1}{N}\sum_{p=0}^{N-1} e^{2\pi i(j-k)p/N}
\end{aligned}$$

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 $e^{2\pi in/N}$.

$$e^{2\pi in/N}\sum_{p=0}^{N-1}\big(e^{2\pi in/N}\big)^{p} = \sum_{p=0}^{N-1}\big(e^{2\pi in/N}\big)^{p+1} = \sum_{p=1}^{N}\big(e^{2\pi in/N}\big)^{p} \qquad (3)$$

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.

5.2

Problem: 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,

$$\begin{aligned}
|00...0\rangle &\to \frac{1}{2^{n/2}}\sum_{k=0}^{2^n-1} e^{0} |k\rangle \qquad (4) \\
&= \frac{1}{2^{n/2}}\sum_{k=0}^{2^n-1} |k\rangle , \qquad (5)
\end{aligned}$$

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.

5.3

Problem (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 $k = k_1k_2...k_n$. Using (5.4) the amplitude a_k corresponding to the state $|k\rangle$ is given by

$$\langle k|\, QFT\, |j\rangle = \frac{1}{2^{n/2}}\left(\delta_{k_1 0} + e^{2\pi i 0.j_n}\delta_{k_1 2}\right)\left(\delta_{k_2 0} + e^{2\pi i 0.j_{n-1}j_n}\delta_{k_2 2}\right)...\left(\delta_{k_n 0} + e^{2\pi i 0.j_1...j_n}\delta_{k_n 2}\right). \tag{6}$$

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.

5.4

Problem: Give a decomposition of the controlled-R_k gate into single qubit and CNOT gates.

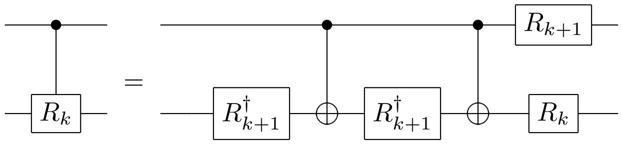
Solution: We use the ABC construction of Corollary 4.2 to make our controlled R_k according to Figure 4.6. First, note that

$$R_k = \begin{pmatrix} 1 & 0 \\ 0 & e^{2\pi i/2^k} \end{pmatrix} = e^{2\pi i/2^{k+1}} \begin{pmatrix} e^{-2\pi i/2^{k+1}} & 0 \\ 0 & e^{2\pi i/2^{k+1}} \end{pmatrix} = e^{i\alpha} R_z(\beta) \tag{7}$$

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,

$$\begin{aligned} A &= R_z(\beta) \\ B &= R_z(-\beta/2) \\ C &= R_z(-\beta/2) \end{aligned} \tag{8}$$

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.

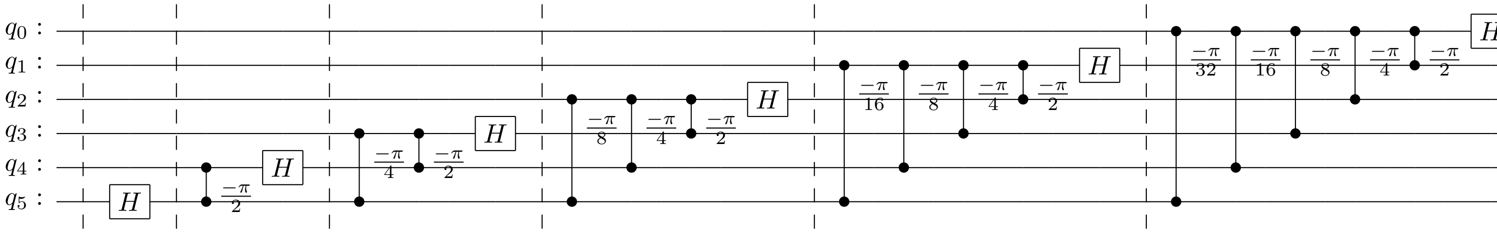


(9)

5.5

Problem: 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.

5.6

Problem (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

$$E(\mathcal{X}\mathcal{Y}, XY) \leq E(\mathcal{X}, X) + E(\mathcal{Y}, Y). \tag{10}$$

To proceed, we set things up to make use of our friend: the triangle inequality. For any state $|\psi\rangle$ we have

$$\begin{aligned} \|(\mathcal{X}\mathcal{Y} - XY)|\psi\rangle\| &= \|(\mathcal{X}\mathcal{Y} - X\mathcal{Y} + X\mathcal{Y} - XY)|\psi\rangle\| & (11)\\ &= \|(\mathcal{X} - X)\mathcal{Y}|\psi\rangle + X(\mathcal{Y} - Y)|\psi\rangle\| & (12)\\ &\leq \|(\mathcal{X} - X)\mathcal{Y}|\psi\rangle\| + \|X(\mathcal{Y} - Y)|\psi\rangle\| & (13) \end{aligned}$$

Since this inequality holds for any $|\psi\rangle$, it certainly holds if we take the max of both sides. Hence,

$$E(\mathcal{X}\mathcal{Y}, XY) \leq \max_{|\psi\rangle} \left(\|(\mathcal{X} - X)\mathcal{Y}|\psi\rangle\| + \|X(\mathcal{Y} - Y)|\psi\rangle\| \right) \tag{14}$$

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,

$$\begin{aligned} E(\mathcal{X}\mathcal{Y}, XY) &\leq \max_{|\phi\rangle} \|(\mathcal{X} - X)|\phi\rangle\| + \max_{|\psi\rangle} \|(\mathcal{Y} - Y)|\psi\rangle\| & (15)\\ &= E(\mathcal{X}, X) + E(\mathcal{Y}, Y). & (16) \end{aligned}$$

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,

$$E(U,V) \leq \frac{n(n+1)}{2}\Delta \in \Theta(n^2\Delta). \tag{17}$$

Here, the use of Θ is appropriate rather than $\mathcal{O}$ since our bound is tight.

5.7

Problem: 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 $j_{t-1}...j_1j_0$, 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

$$|j\rangle = \bigotimes_{k=0}^{t-1} |j_k\rangle \equiv |j_{t-1}\rangle \dots |j_0\rangle \, . \tag{18}$$

The action of the controlled-U^2k controlled on the kth qubitis given by

$$|j_k\rangle |u\rangle \longrightarrow |j_k\rangle U^{2^k j_k} |u\rangle \tag{19}$$

as can be readily verified. Thus, the full sequence of controlled gates acts as follows.

$$\begin{aligned} |j\rangle |u\rangle = \bigotimes_{k=0}^{t-1} |j_k\rangle |u\rangle &\longrightarrow |j\rangle U^{2^{t-1}j_{t-1}} ... U^{2^0 j_0} |u\rangle \\ &= |j\rangle U^{2^0 j_0 + ... + 2^{t-1}j_{t-1}} |u\rangle \\ &= |j\rangle U^j |u\rangle \end{aligned} \tag{20}$$

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 $|u\rangle$.

5.8

Problem: 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

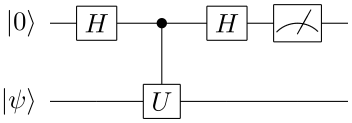
$$p_u p_{succ} \geq |c_u|^2(1-\epsilon). \tag{21}$$

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 (21) remains a lower bound.

5.9

Problem: 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.



(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.

5.10

Problem: Show that the order of $x = 5$ modulo $N = 21$ is 6.

Solution: We proceed by exhaustive calculation.

$$\begin{aligned} 5^1 \bmod 21 &= 5 \\ 5^2 \bmod 21 &= 4 \\ 5^3 \bmod 21 &= 20 \\ 5^4 \bmod 21 &= 16 \\ 5^5 \bmod 21 &= 17 \\ 5^6 \bmod 21 &= 1 \end{aligned} \tag{23}$$

Hence, 6 is the smallest positive integer r such that $5^r \bmod 21 = 1$. Thus, 6 is the order of 5 modulo 21.

5.11

Problem: 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

$$\begin{aligned} x^j - x^i &= 0 \bmod N \\ x^i(x^{j-i} - 1) &= 0 \bmod N. \end{aligned} \tag{24}$$

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.

5.12

Problem: 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,

$$\begin{aligned} xy \bmod N &= xz \bmod N \\ x(y - z) &= 0 \pmod N \end{aligned} \tag{25}$$

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

$$|y\rangle = |z\rangle \tag{26}$$

which proves that U is injective on the basis, as desired.

5.13

Problem: Prove (5.44). (*Hint*:$\sum_{s=0}^{r-1}\exp(-2\pi isk/r)=r\delta_{k0}$). In fact, prove that

$$\frac{1}{\sqrt{r}}\sum_{s=0}^{r-1} e^{2\pi isk/r}|u_s\rangle = \left|x^k \bmod N\right\rangle \tag{27}$$

Solution: Let us crank the wheel: putting in the definition of $|u_s\rangle$ to the left side of (5.44) and regrouping.

$$\begin{aligned}\frac{1}{\sqrt{r}}\sum_{s=0}^{r-1} e^{2\pi isk/r}|u_s\rangle &= \frac{1}{\sqrt{r}}\sum_{s=0}^{r-1} e^{2\pi isk/r}\frac{1}{\sqrt{r}}\sum_{t=0}^{r-1} e^{-2\pi ist/r}\left|x^t \bmod N\right\rangle && (28)\\ &= \frac{1}{r}\sum_{s=0}^{r-1}\sum_{t=0}^{r-1} e^{2\pi is(k-t)/r}\left|x^t \bmod N\right\rangle && (29)\\ &= \frac{1}{r}\sum_{t=0}^{r-1}\left|x^t \bmod N\right\rangle\left(\sum_{s=0}^{r-1} e^{2\pi is(k-t)/r}\right) && (30)\end{aligned}$$

Making gracious use of the hint, we see the rightmost sum on the last line equals $r\delta_{kt}$. Hence,

$$\begin{aligned}\frac{1}{\sqrt{r}}\sum_{s=0}^{r-1} e^{2\pi isk/r}|u_s\rangle &= \sum_{t=0}^{r-1}\delta_{tk}\left|x^t \bmod N\right\rangle && (31)\\ &= \left|x^k \bmod N\right\rangle. && (32)\end{aligned}$$

To obtain (5.44) from the text, simply set $k=0$.

5.14

Problem: The quantum state produced in the order-finding algorithm, before the inverse Fourier transform, is

$$|\psi\rangle = \sum_{j=0}^{2^t-1}|j\rangle U^j|1\rangle = \sum_{j=0}^{2^t-1}|j\rangle\left|x^j \bmod N\right\rangle \tag{33}$$

if we initialize the second register as $|1\rangle$. Show that the same state is obtained if we replace U^j with a *different* unitary transform V, which computes

$$V|j\rangle|k\rangle = |j\rangle\left|k + x^j \bmod N\right\rangle, \tag{34}$$

and start the second register in the state $|0\rangle$. 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. Finding V is going to require techniques in reversible computation.

Chapter 6: Quantum Search Algorithms

6.1

Problem: Show that the unitary operator corresponding to the phase shift in the Grover iteration is $(2\left|0\right\rangle\left\langle 0\right| - I)$.

Solution: For $\left|x\right\rangle = \left|0\right\rangle$,

$$(2\left|0\right\rangle\left\langle 0\right| - I)\left|0\right\rangle = 2\left|0\right\rangle - \left|0\right\rangle = \left|0\right\rangle \tag{35}$$

Meanwhile, for $\left|x\right\rangle \neq \left|0\right\rangle$,

$$(2\left|0\right\rangle\left\langle 0\right| - I)\left|x\right\rangle = 2\left|0\right\rangle\langle 0|x\rangle - \left|x\right\rangle = -\left|x\right\rangle \tag{36}$$

Altogether,

$$(2\left|0\right\rangle\left\langle 0\right| - I)\left|x\right\rangle = (-1)^{\delta_{0x}}\left|x\right\rangle \tag{37}$$

6.2

Problem: 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

$$\sum_k (-\alpha_k + 2\langle\alpha\rangle)\left|k\right\rangle \tag{38}$$

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,

$$\begin{aligned}(2\left|\psi\right\rangle\left\langle\psi\right| - I)\sum_k \alpha_k\left|k\right\rangle &= \sum_k 2\alpha_k\left(\left|\psi\right\rangle\left\langle\psi\right|\right)\left|k\right\rangle - \sum_k \alpha_k\left|k\right\rangle \\ &= 2\left|\psi\right\rangle\sum_k \alpha_k\langle\psi|k\rangle - \sum_k \alpha_k\left|k\right\rangle\end{aligned}$$

Because $\left|\psi\right\rangle$ is uniform superposition over the computational basis states, for all k we have $\langle\psi|k\rangle = 1/\sqrt{N}$. Hence,

$$\begin{aligned}(2\left|\psi\right\rangle\left\langle\psi\right| - I)\sum_k \alpha_k\left|k\right\rangle &= \frac{2}{\sqrt{N}}\left(\sum_k \alpha_k\right)\left|\psi\right\rangle - \sum_k \alpha_k\left|k\right\rangle & (39)\\ &= 2\sqrt{N}\langle\alpha\rangle\left|\psi\right\rangle - \sum_k \alpha_k\left|k\right\rangle & (40)\end{aligned}$$

Finally, we expand out the definition of $\left|\psi\right\rangle$ and cancel the factors of $\sqrt{N}$.This gives our result.

$$2\sqrt{N}\langle\alpha\rangle\left|\psi\right\rangle - \sum_k \alpha_k\left|k\right\rangle = \sum_k (2\langle\alpha\rangle - \alpha_k)\left|k\right\rangle \tag{41}$$

6.3

Problem: Show that in the $\left|\alpha\right\rangle, \left|\beta\right\rangle$ basis, we may write the Grover iteration as

$$G = \begin{bmatrix}\cos\theta & -\sin\theta \\ \sin\theta & \cos\theta\end{bmatrix} \tag{42}$$

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

$$\sin\theta = \frac{2\sqrt{M(N-M)}}{N} \tag{43}$$

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 $V = \mathit{span}\left(\left|\alpha\right\rangle, \left|\beta\right\rangle\right)$ 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 $V = \mathit{span}\left(\left|\alpha\right\rangle, \left|\beta\right\rangle\right)$, 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 (32) for *some* angle θ. It remains to show θ satisfies relation (33). To do this, we simply compute the matrix element $\left\langle\beta\right| G \left|\alpha\right\rangle$.

$$\begin{aligned}\left\langle\beta\right| G \left|\alpha\right\rangle &= \left\langle\beta\right| (2\left|\psi\right\rangle\left\langle\psi\right| - I)\, O \left|\alpha\right\rangle \\ &= \left\langle\beta\right| (2\left|\psi\right\rangle\left\langle\psi\right| - I) \left|\alpha\right\rangle \\ &= \left\langle\beta\right| (2\left|\psi\right\rangle\left\langle\psi|\alpha\right\rangle - \left|\alpha\right\rangle) \\ &= 2\left\langle\beta|\psi\right\rangle\left\langle\psi|\alpha\right\rangle,\end{aligned} \tag{44}$$

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.

$$\left\langle\beta\right| G \left|\alpha\right\rangle = \sin\theta = \frac{2\sqrt{M(N-M)}}{N} \tag{45}$$

Note that, in fact, we did not require the assumption that $M \leq N/2$ in our derivation.

Appendix 1: Notes on basic probability theory

A1.1

Problem: Prove Bayes' rule.

Solution: From the definition of conditional probability, we have

$$p(x,y) = p(y|x)p(x) = p(x|y)p(y) \tag{46}$$

Rearranging the last of these equations gives the desired result.

A1.2

Problem: 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.

$$p(y) = \sum_x p(x,y) \tag{47}$$

We arrive at our result by noting that, from the definition of conditional probability, $p(x,y) = p(y|x)p(x)$.

A1.3

Problem: 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,

$$\mathbf{E}(X) = \sum_{x \in X} x(x) < \mathbf{E}(X) \sum_{x \in X} p(x) = \mathbf{E}(X) \tag{48}$$

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 $p(x) > 0$.

A1.4

Problem: Prove that $\mathbf{E}(X)$ is linear in X.

Solution: The following computation gives us the result.

$$\begin{aligned}
\mathbf{E}(aX + bY) &= \sum_{(x,y) \in (X,Y)} (ax + by) p(x, y) \\
&= \sum_{x \in X} \sum_{y \in Y} axp(x, y) + byp(x, y) \\
&= \sum_{x \in X} ax \sum_{y \in Y} p(x, y) + \sum_{y \in Y} by \sum_{x \in X} p(x, y) \\
&= a \sum_{x \in X} xp(x) + b \sum_{y \in Y} yp(y) \\
&= a\mathbf{E}(X) + b\mathbf{E}(Y).
\end{aligned} \tag{49}$$

Here, a, b are constants. Along the way, we used $p(x) = \sum_y p(x, y)$ and the definition of expectation value.

A1.5

Problem: Prove that for independent random variables X and Y, $\mathbf{E}(\mathrm{XY}) = \mathbf{E}(\mathrm{X})\mathbf{E}(\mathrm{Y})$.

Solution: Recall that, for independent random variables, the joint probability distribution breaks into a product of individual probabilities. This yields the following computation.

$$\begin{aligned}
\mathbf{E}(XY) &= \sum_x \sum_y xy\, p(x, y) \\
&= \sum_x \sum_y xy\, p(x)p(y) \\
&= \sum_x p(x) \sum_y p(y) \\
&= \mathbf{E}(X)\mathbf{E}(Y)
\end{aligned} \tag{50}$$