QI Lecture 14

The Grover Search Algorithm

The Grover Search Algorithm is the first arguably useful quantum algorithm we'll be looking at. The problem is to find one item out of N from an unsorted database. A better way to picture this is trying to find a bean that's hidden under one of many shells.

Figure 1: The classical solution for searching an unsorted database.

Grover

We describe the Grover algorithm using another blackbox function

$$f(x) = \begin{cases} 1, & x = x_0 \\ 0, & x \neq x_0 \end{cases}$$

where $0 \leq x \leq N-1$ and our goal is to find x_0. We'll define two states

$$|w_0\rangle = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} |j\rangle \qquad\qquad |x_0\rangle$$

where $|w_0\rangle$ is the simplest superposition of all the inputs and $|x_0\rangle$ is the target of the search. Because we'll need it in a minute, the inner product of these states is

$$\boxed{\langle x_0|w_0\rangle = \frac{1}{\sqrt{N}}\sum_{j=0}^{N-1}\langle x_0|j\rangle = \frac{1}{\sqrt{N}}\sum_{j=0}^{N-1}\delta_{x_0 j} = \frac{1}{\sqrt{N}}}$$

With these two states we can define two operations

$$\boxed{U_0 = I - 2|w_0\rangle\langle w_0| \qquad\qquad U_f = I - 2|x_0\rangle\langle x_0|}$$

U_f can also be written $U_f|x\rangle = (-1)^{f(x)}|x\rangle$. U_0 is called the "diffusive operator" and U_f is the "function call" or "oracle".

Grover's algorithm is the repeated application of $Q = -U_0U_f$, causing the initial state $|w_0\rangle$ to evolve into $|x_0\rangle$. Notice what Q does to $|\psi\rangle = \alpha|w_0\rangle + \beta|x_0\rangle \in Span\{|w_0\rangle, |x_0\rangle\}$

$$\begin{aligned}
Q\left[\alpha|w_0\rangle + \beta|x_0\rangle\right] &= -\left(I - 2|w_0\rangle\langle w_0|\right)\left(I - 2|x_0\rangle\langle x_0|\right)\left[\alpha|w_0\rangle + \beta|x_0\rangle\right] \\
&= -\left(I - 2|w_0\rangle\langle w_0|\right)\left[\alpha|w_0\rangle - \alpha\tfrac{2}{\sqrt{N}}|x_0\rangle - \beta|x_0\rangle\right] \\
&= -\left(I - 2|w_0\rangle\langle w_0|\right)\left[\alpha|w_0\rangle - \left(\alpha\tfrac{2}{\sqrt{N}} + \beta\right)|x_0\rangle\right] \\
&= -\left[-\alpha|w_0\rangle - \left(\alpha\tfrac{2}{\sqrt{N}} + \beta\right)|x_0\rangle + \left(\alpha\tfrac{2}{\sqrt{N}} + \beta\right)\tfrac{2}{\sqrt{N}}|w_0\rangle\right] \\
&= \left(\alpha\left(1 - \tfrac{2}{\sqrt{N}}\right) - \beta\right)\tfrac{2}{\sqrt{N}}|w_0\rangle + \left(\alpha\tfrac{2}{\sqrt{N}} + \beta\right)|x_0\rangle
\end{aligned}$$

Details aside, $Q\left[\alpha|w_0\rangle + \beta|x_0\rangle\right] \in Span\{|w_0\rangle, |x_0\rangle\}$ which means that we don't have to worry about an N dimensional space, we can just focus on the two dimensional space $S \equiv Span\{|w_0\rangle, |x_0\rangle\}$. In fact, if $\alpha, \beta \in \mathbb{R}$, then we can treat S as a *real* two dimensional space, meaning that we can talk about angles and draw pictures!

In particular, since the space is two dimensional, we can define $|x_0^\perp\rangle$ as the component of $|w_0\rangle$ perpendicular to $|x_0\rangle$.[1] The form $|x_0^\perp\rangle$ takes turns out to be simple enough that it should seem obvious after the calculation is done.[2]

[1] Since $S = Span\{|x_0\rangle, |w_0\rangle\}$, using $|w_0\rangle$ to define $|x_0^\perp\rangle$ is the only option. There are no other directions to work with.

[2] But it's still better not to guess.

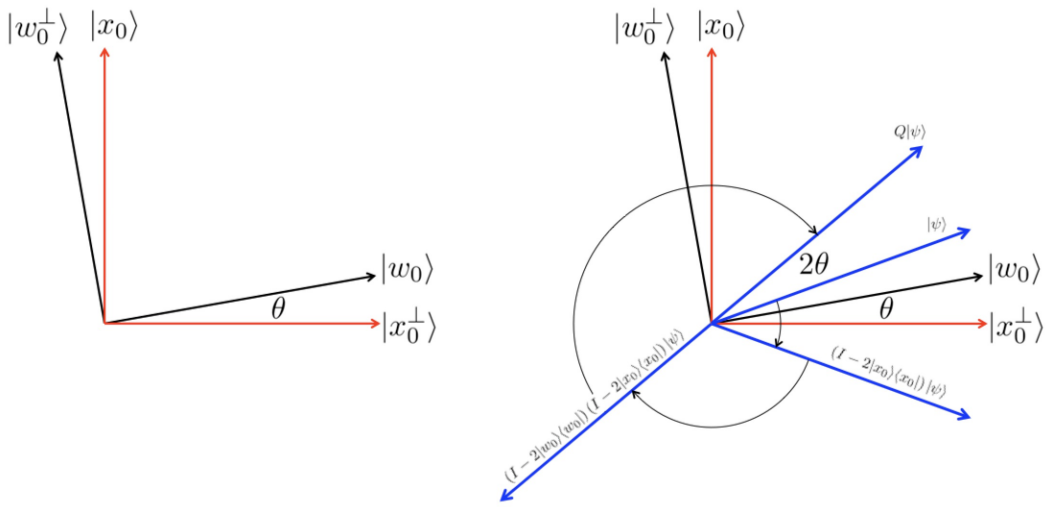

Figure 2: Left: The anatomy of $S = Span\{|w_0\rangle, |x_0\rangle\}$, where θ is the angle between $|w_0\rangle$ and $|x_0^\perp\rangle$. Right: Restricted to S, $Q|\psi\rangle$ is a reflection over $|x_0^\perp\rangle$ followed by a reflection over $|w_0\rangle$, amounting to a rotation by 2θ.

$$
\begin{aligned}
|x_0^\perp\rangle &= \frac{|w_0\rangle - |x_0\rangle\langle x_0|w_0\rangle}{\||w_0\rangle - |x_0\rangle\langle x_0|w_0\rangle\|} \\
&= \frac{|w_0\rangle - |x_0\rangle\langle x_0|w_0\rangle}{\sqrt{\langle w_0|w_0\rangle - 2\langle w_0|x_0\rangle\langle x_0|w_0\rangle + \langle w_0|x_0\rangle\langle x_0|w_0\rangle}} \\
&= \frac{|w_0\rangle - \frac{1}{\sqrt{N}}|x_0\rangle}{\sqrt{1-\frac{1}{N}}} \\
&= \sqrt{\frac{N}{N-1}}\left(\frac{1}{\sqrt{N}}\sum_{x=0}^{N-1}|x\rangle - \frac{1}{\sqrt{N}}|x_0\rangle\right) \\
&= \frac{1}{\sqrt{N-1}}\sum_{x\neq x_0}|x\rangle
\end{aligned}
$$

U_f flips the sign on the component of a vector in the $|x_0\rangle$ direction, so it's a reflection over $|x_0^\perp\rangle$. Similarly, $-U_0$ (note the negative) flips the sign on the component of a vector in the $|w_0^\perp\rangle$ direction, so it's a reflection over $|w_0\rangle$. It turns out that two sequential reflections over intersecting lines is a rotation about their point of intersection by twice the angle between them in the direction pointing from the first line to the second. This is an old geometric fact and proving it is exactly as hard as just looking at what $Q = -U_0U_f$ does to an arbitrary state. Define

$$|\phi\rangle \equiv \cos(\phi)|x_0^\perp\rangle + \sin(\phi)|x_0\rangle$$

In this notation $|\frac{\pi}{2}\rangle = |x_0\rangle$ and we define the angle θ with

$$|w_0\rangle \equiv |\theta\rangle = \cos(\theta)|x_0^\perp\rangle + \sin(\theta)|x_0\rangle$$

We're now equipped to look at the effect of Q on the arbitrary state $|\phi\rangle = \cos(\phi)|x_0^\perp\rangle + \sin(\phi)|x_0\rangle$. First applying U_f:

$$\begin{aligned}
U_f|\phi\rangle &= (I - 2|x_0\rangle\langle x_0|)\left[\cos(\phi)|x_0^\perp\rangle + \sin(\phi)|x_0\rangle\right] \\
&= \cos(\phi)|x_0^\perp\rangle + \sin(\phi)|x_0\rangle - 2\cos(\phi)|x_0\rangle\langle x_0|x_0^\perp\rangle - 2\sin(\phi)|x_0\rangle\langle x_0|x_0\rangle \\
&= \cos(\phi)|x_0^\perp\rangle - \sin(\phi)|x_0\rangle \\
&= \cos(-\phi)|x_0^\perp\rangle + \sin(-\phi)|x_0\rangle
\end{aligned}$$

This is a reflection over $|x_0^\perp\rangle$. Now applying U_0:[3]

$$\begin{aligned}
U_0U_f|\phi\rangle &= (I - 2|w_0\rangle\langle w_0|)\left[\cos(\phi)|x_0^\perp\rangle - \sin(\phi)|x_0\rangle\right] \\
&= \cos(\phi)|x_0^\perp\rangle - \sin(\phi)|x_0\rangle - 2\cos(\phi)|w_0\rangle\langle w_0|x_0^\perp\rangle + 2\sin(\phi)|w_0\rangle\langle w_0|x_0\rangle \\
&= \cos(\phi)|x_0^\perp\rangle - \sin(\phi)|x_0\rangle - 2\cos(\theta)\cos(\phi)|w_0\rangle + 2\sin(\theta)\sin(\phi)|w_0\rangle \\
&= \cos(\phi)|x_0^\perp\rangle - \sin(\phi)|x_0\rangle + \left[-2\cos(\theta)\cos(\phi) + 2\sin(\theta)\sin(\phi)\right]|w_0\rangle \\
&= \cos(\phi)|x_0^\perp\rangle - \sin(\phi)|x_0\rangle + \left[-2\cos(\theta)\cos(\phi) + 2\sin(\theta)\sin(\phi)\right]\left[\cos(\theta)|x_0^\perp\rangle + \sin(\theta)|x_0\rangle\right] \\
&= \begin{cases} & \left[\cos(\phi) - 2\cos^2(\theta)\cos(\phi) + 2\sin(\theta)\cos(\theta)\sin(\phi)\right]|x_0^\perp\rangle \\ + & \left[-\sin(\phi) - 2\sin(\theta)\cos(\theta)\cos(\phi) + 2\sin^2(\theta)\sin(\phi)\right]|x_0\rangle \end{cases} \\
&= \begin{cases} & \left[-\cos(\phi)\cos(2\theta) + \sin(2\theta)\sin(\phi)\right]|x_0^\perp\rangle \\ + & \left[-\sin(\phi)\cos(2\theta) - \sin(2\theta)\cos(\phi)\right]|x_0\rangle \end{cases} \\
&= -\cos(\phi + 2\theta)|x_0^\perp\rangle - \sin(\phi + 2\theta)|x_0\rangle
\end{aligned}$$

Finally,

$$Q|\phi\rangle = -U_0U_f|\phi\rangle = \cos(\phi + 2\theta)|x_0^\perp\rangle + \sin(\phi + 2\theta)|x_0\rangle$$

which is a reflection of $\cos(-\phi)|x_0^\perp\rangle + \sin(-\phi)|x_0\rangle$ over $|w_0\rangle$. This means that,

$$Q^n|w_0\rangle = \cos(\theta + 2n\theta)|x_0^\perp\rangle + \sin(\theta + 2n\theta)|x_0\rangle$$

We want the state of the system to be as nearly equal to $|x_0\rangle$ as possible, so that a measurement is almost guaranteed to produce that target state. This happens when

[3] By the way: $\cos(A + B) = \cos(A)\cos(B) - \sin(A)\sin(B)$, $\sin(2A) = 2\sin(A)\cos(A)$, and $\cos(2A) = \cos^2(A) - 1 = 1 - \sin^2(A)$.

$\frac{\pi}{2} = (2n+1)\theta$. Presumably N, the number of "shells to look under", is large and that means we can use the small-angle-approximation:

$$\frac{1}{\sqrt{N}} = \langle x_0 | w_0 \rangle = \sin(\theta) \approx \theta$$

Therefore, for large values of N, the number of times, n, that we apply Q before measuring the state of the system is

$$\boxed{n \approx \frac{\pi}{4}\sqrt{N}}$$

Quantum Walk Approach

In the usual approach to the Grover algorithm we switch back and forth between the diffusive operator and the function call, but we can gain some insight by describing them as a single operation on a more complicated space. In this case a "star graph". Normally, a classical random walk is described on the vertices of a graph, but we find that formalism is insufficient for quantum walks[4], so we find that if we define our walk on the edges and put an operations on the vertices that a lot of problems clear up. We've actually seen something very much like this before, in lecture 1 when we looked at light traveling down paths (edges) and interacting at beam splitters (vertices). The state on an edge going from vertex a to b is $|a, b\rangle$ and the state going from b to a is $|b, a\rangle$.

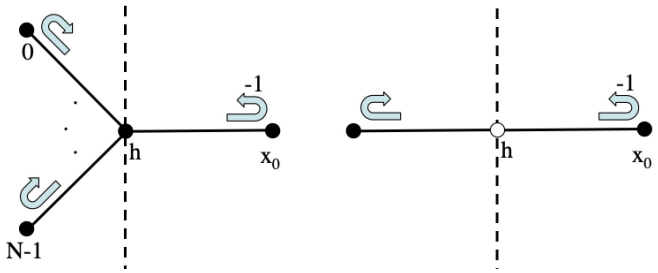

Figure 3: Left: A graph representing the Grover algorithm. States are defined on the directed edges and operations are defined on the vertices. Right: A simplification of the graph where all of the unmarked edges are collected together.

Rather than worry about two different operations, we'll consider a single time-step operation U. In figure 3 the "spokes" are the states that we need to search through and when the vertices at the ends reflect edge states back they either leave the phase alone or,

[4]For example, if the state is on a particular vertex, then where was it one time-step ago? Generally an ancillary "coin space" is added to each vertex to keep track of where a state was at the last time step (or possibly quite a bit more).

for the marked vertex, multiply it by -1. In this way they execute the oracle operation on outgoing states.

$$U|h,j\rangle = |j,h\rangle \qquad\qquad U|h,x_0\rangle = -|x_0,h\rangle$$

The h vertex at the center is the "hub vertex" and it performs the diffusive operation. For any ingoing state

$$U|j,h\rangle = r|h,j\rangle + t\sum_{k\neq j}|h,k\rangle$$

Unitarity at the hub requires that $|r|^2 + (N-1)|t|^2 = 1$ and $2Re(r^*t) + (N-2)|t|^2 = 0$ and the simplest solution is that the hub is a "standard diffusive vertex" meaning that

$$r = -1 + \frac{2}{N} \qquad\qquad t = \frac{2}{N}$$

For ease of notation define the following:

$$|in\rangle \equiv \frac{1}{\sqrt{N-1}}\sum_{j\neq x_0}|j,h\rangle \qquad\qquad |out\rangle \equiv \frac{1}{\sqrt{N-1}}\sum_{j\neq x_0}|h,j\rangle$$

We can now restrict our attention to four states, $|in\rangle, |out\rangle$ on the left and $|h,x_0\rangle, |x_0,h\rangle$ on the right. Clearly,

$$U|out\rangle = |in\rangle \qquad\qquad U|h,x_0\rangle = -|x_0,h\rangle$$

but the hub vertex has become slightly more complicated.

$$\begin{aligned}
U|x_0,h\rangle &= \left(-1+\tfrac{2}{N}\right)|h,x_0\rangle + \tfrac{2}{N}\textstyle\sum_{k\neq x_0}|h,k\rangle \\
&= \left(-1+\tfrac{2}{N}\right)|h,x_0\rangle + \tfrac{2\sqrt{N-1}}{N}\tfrac{1}{\sqrt{N-1}}\textstyle\sum_{k\neq x_0}|h,k\rangle \\
&= \left(-1+\tfrac{2}{N}\right)|h,x_0\rangle + 2\sqrt{\tfrac{1}{N}-\tfrac{1}{N^2}}|out\rangle
\end{aligned}$$

$$\begin{aligned}
U|in\rangle &= \tfrac{1}{\sqrt{N-1}}\textstyle\sum_{j\neq x_0}U|j,h\rangle \\
&= \tfrac{1}{\sqrt{N-1}}\textstyle\sum_{j\neq x_0}\left[\left(-1+\tfrac{2}{N}\right)|h,j\rangle + \tfrac{2}{N}\sum_{k\neq j}|h,k\rangle\right] \\
&= \left(-1+\tfrac{2}{N}\right)|out\rangle + \tfrac{2}{N\sqrt{N-1}}\textstyle\sum_{j\neq x_0}\sum_{k\neq j}|h,k\rangle \\
&= \left(-1+\tfrac{2}{N}\right)|out\rangle + \tfrac{2}{N\sqrt{N-1}}\textstyle\sum_{j\neq x_0}\sum_{k\neq j,x_0}|h,k\rangle + \tfrac{2}{N\sqrt{N-1}}\sum_{j\neq x_0}|h,x_0\rangle \\
&= \left(-1+\tfrac{2}{N}\right)|out\rangle + \tfrac{2(N-2)}{N\sqrt{N-1}}\textstyle\sum_{k\neq x_0}|h,k\rangle + \tfrac{2(N-1)}{N\sqrt{N-1}}|h,x_0\rangle \\
&= \left(-1+\tfrac{2}{N}\right)|out\rangle + 2\left(1-\tfrac{2}{N}\right)|out\rangle + 2\sqrt{\tfrac{1}{N}-\tfrac{1}{N^2}}|h,x_0\rangle \\
&= \left(1-\tfrac{2}{N}\right)|out\rangle + 2\sqrt{\tfrac{1}{N}-\tfrac{1}{N^2}}|h,x_0\rangle
\end{aligned}$$

In the $\{|out\rangle, |in\rangle, |h, x_0\rangle, |x_0, h\rangle\}$ basis we can write U and it's first order approximation

$$U = \begin{bmatrix} 0 & 1-\frac{2}{N} & 0 & 2\sqrt{\frac{1}{N}-\frac{1}{N^2}} \\ 1 & 0 & 0 & 0 \\ 0 & 2\sqrt{\frac{1}{N}-\frac{1}{N^2}} & 0 & -1+\frac{2}{N} \\ 0 & 0 & -1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \frac{2}{\sqrt{N}} \\ 1 & 0 & 0 & 0 \\ 0 & \frac{2}{\sqrt{N}} & 0 & -1 \\ 0 & 0 & -1 & 0 \end{bmatrix} + O\left(\frac{1}{N}\right)$$

as well as the eigenvalues and states

$$\begin{aligned} \lambda_1 &= e^{i\frac{1}{\sqrt{N}}} + O\left(\tfrac{1}{N}\right) & |v_1\rangle &= \tfrac{|out\rangle+|in\rangle-i|h,x_0\rangle+i|x_0,h\rangle}{2} + O\left(\tfrac{1}{\sqrt{N}}\right) \\ \lambda_2 &= e^{-i\frac{1}{\sqrt{N}}} + O\left(\tfrac{1}{N}\right) & |v_2\rangle &= \tfrac{|out\rangle+|in\rangle+i|h,x_0\rangle-i|x_0,h\rangle}{2} + O\left(\tfrac{1}{\sqrt{N}}\right) \\ \lambda_3 &= -e^{i\frac{1}{\sqrt{N}}} + O\left(\tfrac{1}{N}\right) & |v_3\rangle &= \tfrac{|out\rangle-|in\rangle-i|h,x_0\rangle-i|x_0,h\rangle}{2} + O\left(\tfrac{1}{\sqrt{N}}\right) \\ \lambda_4 &= -e^{-i\frac{1}{\sqrt{N}}} + O\left(\tfrac{1}{N}\right) & |v_4\rangle &= \tfrac{|out\rangle-|in\rangle+i|h,x_0\rangle+i|x_0,h\rangle}{2} + O\left(\tfrac{1}{\sqrt{N}}\right) \end{aligned}$$

We'll now restrict our attention to the $\lambda_1 \approx \lambda_2 \approx 1$ eigenstates. In the limit as $N \to \infty$ something important happens:

$$U_\infty = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \end{bmatrix}$$

The weak coupling, $\frac{2}{\sqrt{N}}$, between the left and right sides of figure 3 disappears, the eigenspaces become degenerate, $\lambda_1 = \lambda_2 = 1$, and initial conditions (such as this one) become concentrated on the left side:

$$|\psi\rangle \equiv \frac{1}{\sqrt{2N}} \sum_{j=0}^{N-1} [|j,h\rangle + |h,j\rangle] = \sqrt{\frac{N-1}{N}} \frac{|in\rangle + |out\rangle}{\sqrt{2}} + \frac{1}{\sqrt{N}} \frac{|x_0,h\rangle + |h,x_0\rangle}{\sqrt{2}} \longrightarrow \frac{|in\rangle + |out\rangle}{\sqrt{2}} = \frac{|v_1\rangle + |v_2\rangle}{\sqrt{2}}$$

In the $N = \infty$ scenario, initial conditions remain on the left side because $U_\infty\left(\frac{|v_1\rangle+|v_2\rangle}{\sqrt{2}}\right) = \frac{|v_1\rangle+|v_2\rangle}{\sqrt{2}}$. The 1-eigenspace is two dimensional, with one on both the left and right. As we perturbate the variable $\frac{1}{N}$ from zero to small, non-zero values the eigenspaces "break apart" and the new non-degenerate eigenspaces, $Span\{|v_1\rangle\}$ and $Span\{|v_2\rangle\}$ have support on both the left and right.

The initial condition is still approximately equal to a linear combination of $|v_1\rangle$ and $|v_2\rangle$,

$$|\psi\rangle = \frac{1}{\sqrt{2N}} \sum_{j=0}^{N-1} [|j,h\rangle + |h,j\rangle] = \frac{|v_1\rangle + |v_2\rangle}{\sqrt{2}} + O\left(\frac{1}{\sqrt{N}}\right)$$

The big difference is that, now that the eigenvalues don't match, these two eigenvectors slowly migrate from adding together to subtracting. Applying U a total of $n = \frac{\pi}{2}\sqrt{N}$ times:

$$\begin{aligned} U^n|\psi\rangle &= \frac{U^n|v_1\rangle + U^n|v_2\rangle}{\sqrt{2}} \\ &= \frac{e^{i\frac{n}{\sqrt{N}}}|v_1\rangle + e^{-i\frac{n}{\sqrt{N}}}|v_2\rangle}{\sqrt{2}} \\ &= \frac{e^{i\frac{\pi}{2}}|v_1\rangle + e^{-i\frac{\pi}{2}}|v_2\rangle}{\sqrt{2}} \\ &= \frac{i|v_1\rangle - i|v_2\rangle}{\sqrt{2}} \\ &= \frac{|h,x_0\rangle - |x_0,h\rangle}{\sqrt{2}} \end{aligned}$$

Approximations here are accurate to within $O\left(\frac{1}{\sqrt{N}}\right)$.

So after $n = \frac{\pi}{2}\sqrt{N}$ steps an initial condition spread across all of the "spokes" becomes concentrated on the marked spoke and a measurement will reveal it with probability approximately 1. Properly read, this is the same number of steps that the first treatment of Grover's algorithm took, since it takes two steps for states to reflect off the ends of the spokes (oracle operation) and to scatter off the hub (diffusive operation). In the first iteration, these two operations are considered together, one after the other.

Optimality

The important thing to notice here is that $n = O\left(\sqrt{N}\right)$!! Classically, if there were a $N = 10^6$ shells to look under, you'd need to look under half, 500,000, on average before finding the bean. "$\frac{N}{2}$" scales proportionately to N, meaning that it's $O(N)$. Grovers algorithm would only require about $\frac{\pi}{4}\sqrt{10^6} \approx 785$ steps to find the target shell. Because it scales proportional to the square root of the size of the unsorted database, Grover's algorithm becomes more useful the larger the database.

I'm including this section because it's rare to see a proof of the optimality of Grover, although this is a well known fact (before the end of this section you'll see why it's so rare). This also gives you an example of one of the more difficult problems in computation: putting bounds on computational difficulty.

However, it would be nice to know if it's possible to do any better. Define $U_x = I - 2|x\rangle\langle x|$, which is the blackbox function targeting the state $|x\rangle$. We'd like a search algorithm to work after a fixed number of steps for *any* x, since we don't know what x is.[5] Each application

[5] That's the whole point of a search.

of U_x is a function call, and in between them we'll assume that we can apply any unitary operation, U_j, that may change between calls. This is about as generalized as you can get. With initial state $|\psi_0\rangle$, define

$$|\psi_j^x\rangle = U_j U_x U_{j-1} U_x \dots U_1 U_x |\psi_0\rangle \qquad |\psi_k\rangle = U_j U_{j-1} \dots U_1 |\psi_0\rangle$$

when the requisite number of function calls, j, has been met, we'd like the probability of success to be at least $\frac{1}{2}$, so

$$\left|\langle x|\psi_j^x\rangle\right|^2 > \frac{1}{2}$$

This is a fairly standard requirement for probabilistic algorithms which aren't guaranteed to return a correct result. With a high enough probability, repeating the algorithm several times makes it effectively non-probabilistic.

We'll consider the quantity

$$D_j \equiv \sum_x \left\| |\psi_j^x\rangle - |\psi_j\rangle \right\|^2$$

and by finding both upper and lower bounds for D_j we'll establish the optimality of Grover's $O(\sqrt{N})$ speed. D_j doesn't really mean anything useful; it's just something to use for this proof. This proof involves a lot of inequalities and identities, so they'll be noted as they're used. As a reminder:

Triangle inequality:	$\|\|\eta\rangle \pm |\phi\rangle\| \le \||\eta\rangle\| + \||\phi\rangle\|$
Cauchy-Schwarz:	$\left\|\sum_k A_k B_k\right| \le \left(\sum_j A_j^2\right)^{\frac{1}{2}} \left(\sum_k B_k^2\right)^{\frac{1}{2}}$
Unitarity:	$\|U|\phi\rangle\| = \||\phi\rangle\|$
Pythagoras:	$\||\phi\rangle\|^2 = \|P_s|\phi\rangle + P_{s^\perp}|\phi\rangle\|^2 = \|P_s|\phi\rangle\|^2 + \|P_{s^\perp}|\phi\rangle\|^2$
Normality:	$\sum_k \left|\langle k|\phi\rangle\right|^2 = 1$

where P_s and $P_{s^\perp}$ are projections onto orthogonal spaces.

$$
\begin{array}{rcll}
D_{j+1} &=& \sum_x \left\| |\psi_{j+1}^x\rangle - |\psi_{j+1}\rangle \right\|^2 & \\
&=& \sum_x \left\| U_{j+1}U_x|\psi_j^x\rangle - U_{j+1}|\psi_j\rangle \right\|^2 & \\
&=& \sum_x \left\| U_{j+1}\left(U_x|\psi_j^x\rangle - |\psi_j\rangle\right) \right\|^2 & \\
&=& \sum_x \left\| U_x|\psi_j^x\rangle - |\psi_j\rangle \right\|^2 & Unitarity \\
&=& \sum_x \left\| U_x\left(|\psi_j^x\rangle - |\psi_j\rangle\right) + \left(U_x - I\right)|\psi_j\rangle \right\|^2 & \\
&\leq& \sum_x \left[\left\| U_x\left(|\psi_j^x\rangle - |\psi_j\rangle\right) \right\| + \left\| \left(U_x - I\right)|\psi_j\rangle \right\| \right]^2 & Triangle \\
&=& \sum_x \left[\left\| |\psi_j^x\rangle - |\psi_j\rangle \right\| + \left\| \left(U_x - I\right)|\psi_j\rangle \right\| \right]^2 & Unitarity \\
&=& \sum_x \left[\left\| |\psi_j^x\rangle - |\psi_j\rangle \right\| + \left\| -2|x\rangle\langle x|\psi_j\rangle \right\| \right]^2 & U_x = I - 2|x\rangle\langle x| \\
&=& \sum_x \left[\left\| |\psi_j^x\rangle - |\psi_j\rangle \right\| + 2\left|\langle x|\psi_j\rangle\right| \, \||x\rangle\| \right]^2 & \\
&=& \sum_x \left[\left\| |\psi_j^x\rangle - |\psi_j\rangle \right\| + 2\left|\langle x|\psi_j\rangle\right| \right]^2 & \\
&=& \sum_x \left[\left\| |\psi_j^x\rangle - |\psi_j\rangle \right\|^2 + 4\left\| |\psi_j^x\rangle - |\psi_j\rangle \right\| \left|\langle x|\psi_j\rangle\right| + 4\left|\langle x|\psi_j\rangle\right|^2 \right] & \\
&=& D_j + \sum_x \left[4\left\| |\psi_j^x\rangle - |\psi_j\rangle \right\| \left|\langle x|\psi_j\rangle\right| + 4\left|\langle x|\psi_j\rangle\right|^2 \right] & \\
&=& D_j + 4 + 4\sum_x \left\| |\psi_j^x\rangle - |\psi_j\rangle \right\| \left|\langle x|\psi_j\rangle\right| & Normality \\
&\leq& D_j + 4 + 4\left(\sum_x \left\| |\psi_j^x\rangle - |\psi_j\rangle \right\|^2 \right)^{\frac{1}{2}} \left(\sum_x \left|\langle x|\psi_j\rangle\right|^2 \right)^{\frac{1}{2}} & Cauchy - Schwarz \\
&\leq& D_j + 4 + 4\left(\sum_x \left\| |\psi_j^x\rangle - |\psi_j\rangle \right\|^2 \right)^{\frac{1}{2}} & Normality \\
&=& D_j + 4\sqrt{D_j} + 4 &
\end{array}
$$

We can use this to inductively show that

$$\boxed{D_j \leq 4j^2}$$

The base case is $D_1 = \sum_x \|U_1U_x|\psi_0\rangle - U_1|\psi_0\rangle\|^2 = \sum_x \|U_x|\psi_0\rangle - |\psi_0\rangle\|^2 = \sum_x \|(U_x - I)|\psi_0\rangle\|^2 = \sum_x \|-2|x\rangle\langle x|\psi_0\rangle\|^2 = \sum_x 4\left|\langle x|\psi_0\rangle\right|^2 = 4$ and the inductive case is

$$D_{j+1} \leq D_j + 4\sqrt{D_j} + 4 \leq 4j^2 + 8j + 4 = 4(j+1)^2$$

So that's the upper bound. Here's the lower bound.

We'll use two perpendicular projections: $P_x = |x\rangle\langle x|$ and $P_{x^\perp} = I - |x\rangle\langle x|$.

$$
\begin{array}{lcll}
\left\||\psi_j^x\rangle - |\psi_j\rangle\right\|^2 & = & \left\|P_x\left(|\psi_j^x\rangle - |\psi_j\rangle\right) + P_{x^\perp}\left(|\psi_j^x\rangle - |\psi_j\rangle\right)\right\|^2 & \\
& = & \left\|P_x\left(|\psi_j^x\rangle - |\psi_j\rangle\right)\right\|^2 + \left\|P_{x^\perp}\left(|\psi_j^x\rangle - |\psi_j\rangle\right)\right\|^2 & Pythagoras \\
& = & \left\||x\rangle\left(\langle x|\psi_j^x\rangle - \langle x|\psi_j\rangle\right)\right\|^2 + \left\|P_{x^\perp}\left(|\psi_j^x\rangle - |\psi_j\rangle\right)\right\|^2 & \\
& = & \left|\langle x|\psi_j^x\rangle - \langle x|\psi_j\rangle\right|^2 + \left\|P_{x^\perp}\left(|\psi_j^x\rangle - |\psi_j\rangle\right)\right\|^2 & \\
& = & \left\{\begin{array}{ll} & \left(\langle x|\psi_j^x\rangle - \langle x|\psi_j\rangle\right)^*\left(\langle x|\psi_j^x\rangle - \langle x|\psi_j\rangle\right) \\ + & \left(\langle\psi_j^x| - \langle\psi_j|\right)P_{x^\perp}P_{x^\perp}\left(|\psi_j^x\rangle - |\psi_j\rangle\right)\end{array}\right. & \\
& = & \left\{\begin{array}{ll} & \left(\langle x|\psi_j^x\rangle - \langle x|\psi_j\rangle\right)^*\left(\langle x|\psi_j^x\rangle - \langle x|\psi_j\rangle\right) \\ + & \left(\langle\psi_j^x| - \langle\psi_j|\right)P_{x^\perp}\left(|\psi_j^x\rangle - |\psi_j\rangle\right)\end{array}\right. & P_{x^\perp}^\dagger P_{x^\perp} = P_{x^\perp}^2 = P_{x^\perp} \\
& = & \left\{\begin{array}{ll} & \left|\langle x|\psi_j^x\rangle\right|^2 + \left|\langle x|\psi_j\rangle\right|^2 - \langle x|\psi_j^x\rangle\langle x|\psi_j\rangle^* - \langle x|\psi_j^x\rangle^*\langle x|\psi_j\rangle \\ + & \left\|P_{x^\perp}|\psi_j^x\rangle\right\|^2 + \left\|P_{x^\perp}|\psi_j\rangle\right\|^2 - \langle\psi_j|P_{x^\perp}|\psi_j^x\rangle - \langle\psi_j^x|P_{x^\perp}|\psi_j\rangle\end{array}\right. & \\
& = & \left\{\begin{array}{ll} & \left|\langle x|\psi_j^x\rangle\right|^2 + \left|\langle x|\psi_j\rangle\right|^2 - 2Re\left[\langle x|\psi_j^x\rangle\langle x|\psi_j\rangle^*\right] \\ + & \left\|P_{x^\perp}|\psi_j^x\rangle\right\|^2 + \left\|P_{x^\perp}|\psi_j\rangle\right\|^2 - 2Re\left[\langle\psi_j|P_{x^\perp}|\psi_j^x\rangle\right]\end{array}\right. & \\
& \geq & \left\{\begin{array}{ll} & \left|\langle x|\psi_j^x\rangle\right|^2 + \left|\langle x|\psi_j\rangle\right|^2 - 2\left|\langle x|\psi_j^x\rangle\right|\left|\langle x|\psi_j\rangle\right| \\ + & \left\|P_{x^\perp}|\psi_j^x\rangle\right\|^2 + \left\|P_{x^\perp}|\psi_j\rangle\right\|^2 - 2\left|\langle\psi_j|P_{x^\perp}|\psi_j^x\rangle\right|\end{array}\right. & Re[A] \leq |A| \\
& = & \left\{\begin{array}{ll} & \left(\left|\langle x|\psi_j^x\rangle\right|^2 + \left\|P_{x^\perp}|\psi_j^x\rangle\right\|^2\right) + \left(\left|\langle x|\psi_j\rangle\right|^2 + \left\|P_{x^\perp}|\psi_j\rangle\right\|^2\right) \\ - & 2\left|\langle x|\psi_j^x\rangle\right|\left|\langle x|\psi_j\rangle\right| - 2\left|\langle\psi_j|P_{x^\perp}|\psi_j^x\rangle\right|\end{array}\right. & \\
& = & \left\||\psi_j^x\rangle\right\|^2 + \left\||\psi_j\rangle\right\|^2 - 2\left|\langle x|\psi_j^x\rangle\right|\left|\langle x|\psi_j\rangle\right| - 2\left|\langle\psi_j|P_{x^\perp}|\psi_j^x\rangle\right| & Pythagoras \\
& = & 2 - 2\left|\langle x|\psi_j^x\rangle\right|\left|\langle x|\psi_j\rangle\right| - 2\left|\langle\psi_j|P_{x^\perp}|\psi_j^x\rangle\right| & \\
& \geq & 2 - 2\left|\langle x|\psi_j^x\rangle\right|\left|\langle x|\psi_j\rangle\right| - 2\left\||\psi_j\rangle\right\|\left\|P_{x^\perp}|\psi_j^x\rangle\right\| & Cauchy - Schwarz \\
& = & 2 - 2\left|\langle x|\psi_j^x\rangle\right|\left|\langle x|\psi_j\rangle\right| - 2\left\|P_{x^\perp}|\psi_j^x\rangle\right\| & \\
& \geq & 2 - 2\left|\langle x|\psi_j\rangle\right| - 2\left\|P_{x^\perp}|\psi_j^x\rangle\right\| &
\end{array}
$$

Because (by Pythagoras) $\left|\langle x|\psi_j^x\rangle\right|^2 + \left\|P_{x^\perp}|\psi_j^x\rangle\right\|^2 = 1$ and $\left|\langle x|\psi_j^x\rangle\right|^2 > \frac{1}{2}$ we must have that $\left\|P_{x^\perp}|\psi_j^x\rangle\right\| < \frac{1}{\sqrt{2}}$ and therefore

$$\left\||\psi_j^x\rangle - |\psi_j\rangle\right\|^2 \geq 2 - 2\left|\langle x|\psi_j\rangle\right| - \sqrt{2}$$

Keeping in mind that there are N different values of x,

$$
\begin{array}{rcll}
D_j & = & \sum_x \left\| |\psi_j^x\rangle - |\psi_j\rangle \right\|^2 & \\
& \geq & \sum_x \left[2 - \sqrt{2} - 2\,|\langle x|\psi_j\rangle| \right] & \\
& = & \left(2-\sqrt{2}\right) N - 2 \sum_x |\langle x|\psi_j\rangle| & \\
& \geq & \left(2-\sqrt{2}\right) N - 2 \left(\sum_x 1^2\right)^{\frac{1}{2}} \left(\sum_x |\langle x|\psi_j\rangle|^2\right)^{\frac{1}{2}} & Cauchy - Schwarz \\
& = & \left(2-\sqrt{2}\right) N - 2 \left(\sum_x 1^2\right)^{\frac{1}{2}} & Normality \\
& = & \left(2-\sqrt{2}\right) N - 2\sqrt{N} &
\end{array}
$$

So at long last

$$4j^2 \geq D_j \geq \left(2-\sqrt{2}\right) N - 2\sqrt{N}$$

and therefore

$$j \geq \sqrt{N} \sqrt{\frac{2-\sqrt{2}}{4} - \frac{1}{2\sqrt{N}}} = O\left(\sqrt{N}\right)$$

This means that quantum search algorithms for unsorted sets of size N require *at least* $O\left(\sqrt{N}\right)$ iterations, and since Grover's algorithm does the job in $j \approx \frac{\pi}{4}\sqrt{N}$ steps no algorithm will ever be faster by more than a scale factor; twice as fast maybe, but not $O\left(N^{\frac{1}{3}}\right)$.

Exercises

#1) One and Done.

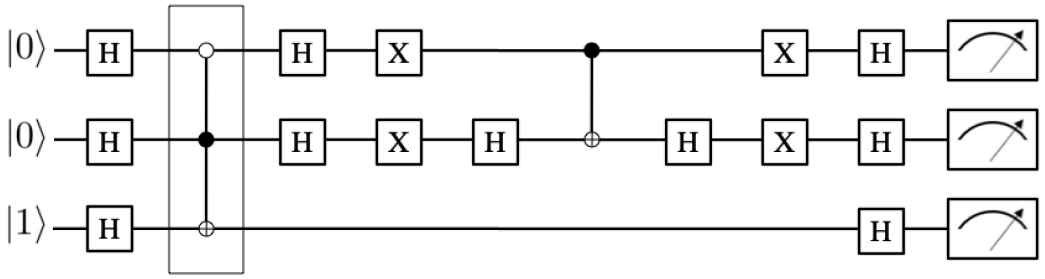

Figure 4: The circuit for a search over $N = 4$ states.

This circuit performs a Grover search in one run.[6] The boxed section is the part of the circuit that would normally be a "black box". It's a multiply-controlled not gate: if the top qubit[7] is 0 and the middle qubit is 1, then the bottom qubit is flipped.

The output of the algorithm is given in binary by a measurement of the top two qubits. The bottom qubit will always be found in the $|1\rangle$ state at the end of this algorithm.

a) Follow the state of the system all the way through the circuit.

b) What is the target state, $|x_0\rangle$?

c) Indicate which parts of the circuit are:

1) Preparing the initial state, $|w_0\rangle$.
2) Performing the oracle operation, $U_f = I - 2|x_0\rangle\langle x_0|$.
3) Performing the diffusive operation, $U_0 = I - 2|w_0\rangle\langle w_0|$.

d) What would the circuit look like if it applied Q^2, instead of just Q as it does now?

e) What would be the output state just before measurement?

[6] Which is possible because for $N = 2^n = 4$ the angle between $\frac{|00\rangle+|01\rangle+|10\rangle+|11\rangle}{2}$ and the target state, $|00\rangle$, $|01\rangle$, $|10\rangle$, or $|11\rangle$, is 30^o and therefore a single run will rotate the state 60^o to stop at 90^o, ending the algorithm.

[7] The white dot indicates that you're using the opposite of a regular CNOT gate; the target qubit is flipped when the control is zero and left alone when the control is 1.

#2) You Never Lose Just One Key.

The Grover algorithm can be expanded to finding any one of many marked states. Instead of one marked state, assume (for simplicity) that the marked states are all the states in the set $S = \{|1\rangle, |2\rangle, \ldots, |s\rangle\}$. Define

$$|S\rangle = \frac{1}{\sqrt{s}} \sum_{x=1}^{s} |x\rangle \qquad\qquad |w_0\rangle = \frac{1}{\sqrt{N}} \sum_{x=1}^{N} |x\rangle$$

The diffusive operator is the same, $U_0 = I - 2|w_0\rangle\langle w_0|$, but the oracle operation is now $U_f = I - 2|S\rangle\langle S|$.

a) Find $|S^\perp\rangle$, the normalized state in $Span\{|S\rangle, |w_0\rangle\}$ such that $\langle S|S^\perp\rangle = 0$.

b) Write the initial state, $|w_0\rangle$, as a linear combination of $|S\rangle$ and $|S^\perp\rangle$.

c) How many iterations of $Q = -U_0 U_f$ does it take to transform $|w_0\rangle \to |S\rangle$?

#3) When "Good Enough" is Good Enough.

If the Grover algorithm takes a thousand iterations, then the last few don't really seem necessary. If N is the number of items, then as $n \to \frac{\pi}{4}\sqrt{N}$ the state of the system is already really close to $|x_0\rangle$. So why not stop early? Assume that it's easy and fast to verify a correct answer, so we only need the Grover algorithm to produce a single correct result.

a) Assume there are N items. What is the probability of a success after n iterations?

b) For a given number of iterations, n, what is the expected number of times, T, you will need to run the Grover algorithm before you see a correct result?

c) The total processing time is approximately nT. For a given N, what is the optimal number of iterations to use in each run of the algorithm? On average, will this be faster or slower than $\frac{\pi}{4}\sqrt{N}$?