Quantum Computing: Foundations to Frontier Fall 2018

Lecture 5

Lecturer: Henry Yuen *Scribes: Jared Barron, Sita Gakkhar*
Nils Wenzler, Robin Gaudreau

1 Admin

This is the last of the foundation material for the class. From here we will transition to the Quantum Computing "frontier."

Agenda

1. Grover Search
2. Hamiltonians
3. Open quantum systems: noise and decoherence

2 Grover's Search Algorithm

So far we have considered quantum algorithms for structured problems, for example, Simon's algorithm for identifying the hidden shift (and others with structure like the Quantum Fourier Transform and factoring using Shor's algorithm) getting exponential speedup over classical approaches. Next we consider a much more unstructured, general search problem. The Grover's search algorithm will give us a quadratic speed up here compared to exponential before but the search problem is very general as most problems can be cast this way.

The Grover search algorithm was devised by Lov Grover in 1997 a few years after Shor's factoring algorithm.

We start with an unknown function $f : \{0,1\}^n \rightarrow \{0,1\}$ with the goal of finding x such that $f(x) = 1$ if it exists.

The classical query complexity is $\Omega(N)$ where $N = 2^n$ because we have to look at all inputs in the worst case. The quantum query complexity will turn out to be $\mathcal{O}(\sqrt{N})$.

2.1 The Algorithm

Let $f : \{0,1\}^n \rightarrow \{0,1\}$, $N = 2^n$. We will use a different query model. Before we had the XOR oracle:

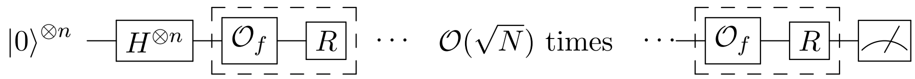
$$U_f|x, b\rangle = |x, b \oplus f(x)\rangle$$

acting on $n+1$ qubits. Now we have the phase oracle:

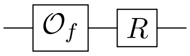
$$O_f|x\rangle = -1^{f(x)}|x\rangle$$

acting on n qubits. It can be shown that these orcales are equivalent.

Grover's algorithm is defined by the circuit:



where we have $\sqrt{N}$ copies of the following block (called the *Grover Iterate*)



After applying $H^{\otimes n}$ on $|0\rangle^{\otimes n}$, $|\psi_1\rangle$ is the uniform superposition of all states:

$$|\psi_1\rangle = \frac{1}{\sqrt{N}} \sum_{\{0,1\}^n} |x\rangle$$

And let R denote the n-qubit unitary,

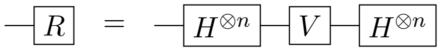
$$R = 2|\psi_1\rangle\langle\psi_1| - I$$

which has components R_{xy} for $x, y \in \{0,1\}^n$ given by,

$$R_{xy} = \begin{cases} \dfrac{2}{N}, & \text{if } x \neq y \\ \\ \dfrac{2}{N} - \mathbb{I}, & \text{if } x = y \end{cases}$$

Geometrically R reflects about ψ_1. As there are $\sqrt{N}$ Grover Iterate blocks, each querying f once, the query complexity is $\mathcal{O}(\sqrt{N})$.

Note that R can be implemented as the following circuit.



where

$$V|x\rangle = \begin{cases} V|0\rangle^{\otimes n} = |0\rangle^{\otimes n} \\ V|x\rangle = -|x\rangle \qquad \text{if } |x\rangle \neq |0\rangle^{\otimes n} \end{cases}$$

V can be implemented with $\mathcal{O}(n)$ gates, while H also has a small circuit *(exercise)*. Here, though, we are only interested in query complexity.

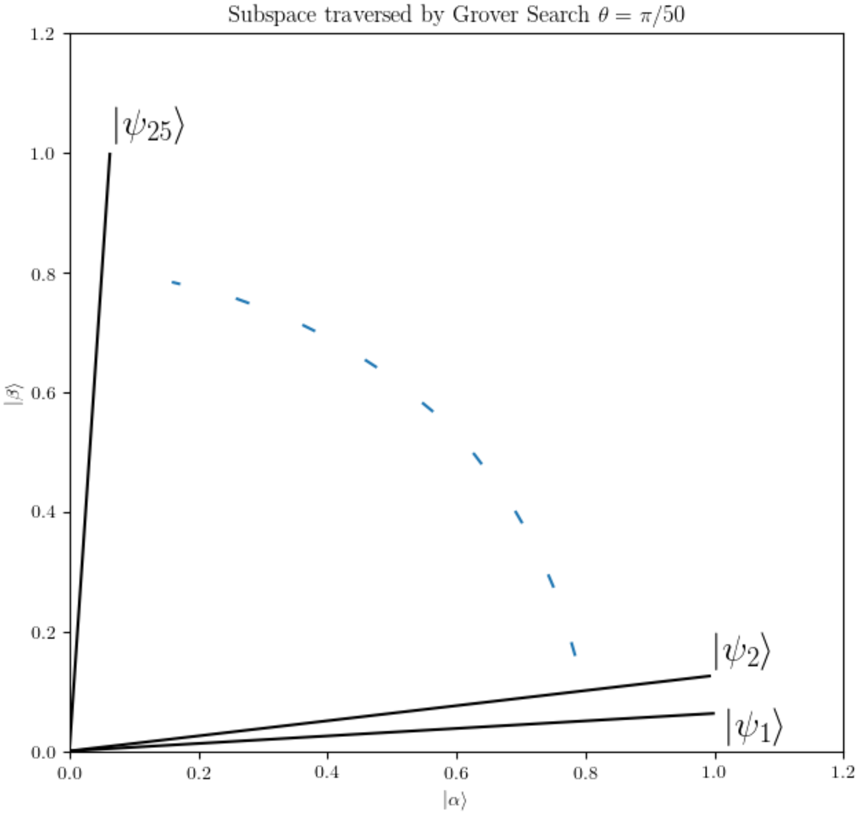
2.2 Analysis

Let $M = |\{x : f(x) = 1\}|$. Assume $M = 1$, so $\exists! x$ such that $f(x) = 1$. Let this x be x^*.

A way to visualize Grover's algorithm is to note that after each application of the Grover Iterate, the state of the algorithm lives in a two dimensional subspace spanned by

$$|\alpha\rangle = \frac{1}{\sqrt{N-1}} \sum_{x \neq x^*} |x\rangle \text{ and } |\beta\rangle = |x^*\rangle$$

.



That is, $|\alpha\rangle$ is uniform superposition over all non-solutions, and $|\beta\rangle$ over solutions.

We want that after $k \approx \sqrt{N}$ iterations the state of the algorithm has high overlap with β. Thus, on measurement we will get $|x^*\rangle$ with high probability.

To analyze the algorithm we'll work exclusively with basis consisting of $|\alpha\rangle$, $|\beta\rangle$.

Let $|\psi_1\rangle$ be the state after the initial $H^{\otimes n}$ applied to $|0\rangle^{\otimes n}$,

$$|\psi_1\rangle = \sqrt{\frac{N-1}{N}}|\alpha\rangle + \sqrt{\frac{1}{N}}|\beta\rangle$$

This gives

$$|\psi_2\rangle = O_f|\psi_1\rangle = \sqrt{\frac{N-1}{N}} O_f|\alpha\rangle + \sqrt{\frac{1}{N}} O_f|\beta\rangle \implies |\psi_2\rangle = \sqrt{\frac{N-1}{N}}|\alpha\rangle - \sqrt{\frac{1}{N}}|\beta\rangle$$

Since $f(x) = 1$ iff $x = x^*$ meaning

$$O_f|\alpha\rangle = \frac{1}{\sqrt{N-1}} \sum_{x \neq x^*} O_f|x\rangle = \frac{1}{\sqrt{N-1}} \sum_{x \neq x^*} |x\rangle$$

and

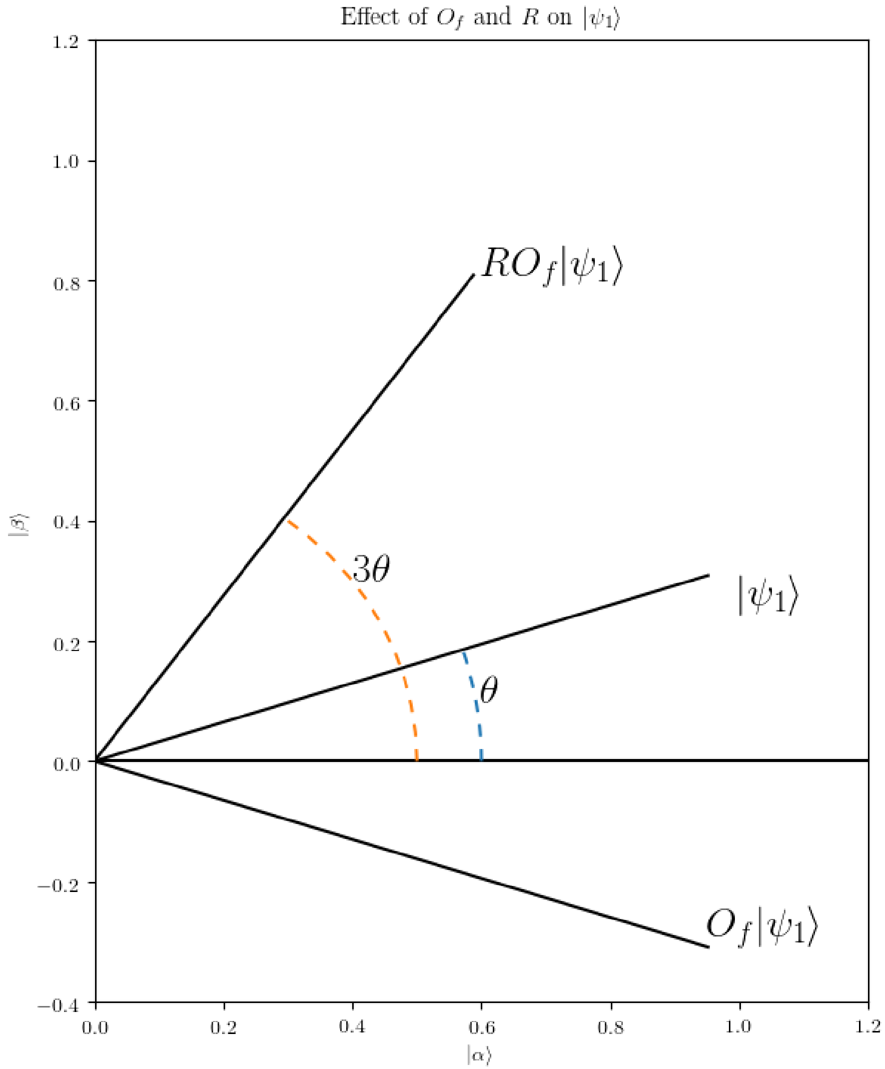
$$O_f|x^*\rangle = -|x^*\rangle$$

Computing $R|\psi_2\rangle$ gives

$$\begin{aligned}
|\psi_3\rangle = R|\psi_2\rangle &= (2|\psi_1\rangle\langle\psi_1| - \mathbb{I})\,|\psi_2\rangle \\
&= 2|\psi_1\rangle\langle\psi_1|\psi_2\rangle - |\psi_2\rangle \\
&= 2|\psi_1\rangle\left(\frac{N-1}{N} - \frac{1}{N}\right) - |\psi_2\rangle \\
&= \frac{N-4}{N}\sqrt{\frac{N-1}{N}}|\alpha\rangle + \frac{3N-4}{N}\frac{1}{\sqrt{N}}|\beta\rangle
\end{aligned}$$

Now let θ be such that $\cos(\theta) = \sqrt{(N-1)/N}$. Then from trigonometry it follows that:

$$|\psi_3\rangle = \cos(3\theta)|\alpha\rangle + \sin(3\theta)|\beta\rangle$$



Note that O_f reflects about $|\alpha\rangle$ and R reflects about $|\psi_1\rangle$. So each iteration gets us towards $|\beta\rangle$ as the state vector sweeps towards it.

After applying k Grover Iterates, we get to state:

$$|\psi_{2k+1}\rangle = \cos((2k+1)\theta)|\alpha\rangle + \sin((2k+1)\theta)|\beta\rangle$$

.

We want k such that $\sin((2k+1)\theta) \approx 1$ meaning $(2k+1)\theta \approx \pi/2$, implying

$$k \approx \frac{\pi}{4\theta} = \frac{\pi}{4\cos^{-1}((N-1)/N)}$$

This gives $\theta \approx \sin(\theta) = 1/\sqrt{N}$ using small angle approximation. Thus, for $k \approx \sqrt{N}\pi/4$, $|\psi_{2k+1}\rangle \approx |\beta\rangle$, more precisely the angle between them is $\mathcal{O}(\theta) \approx \mathcal{O}(1/\sqrt{N})$.

$$|\psi_{2k+1}\rangle \approx \sqrt{\frac{N-1}{N}}|x^*\rangle + \dots$$

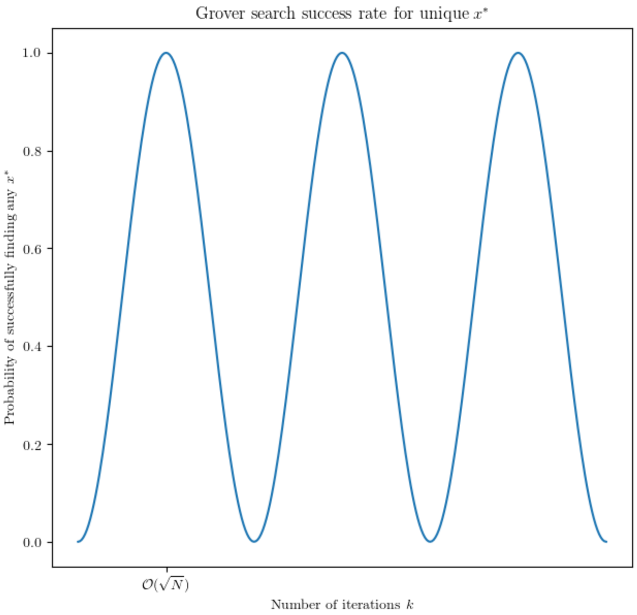
So with probability $(N-1)/N \approx 1$ for large N, we measure to obtain x^*.

And as number of Grover iterates is the number of oracle queries, this is the number of queries.

2.3 Success Probability

A few things to notice:

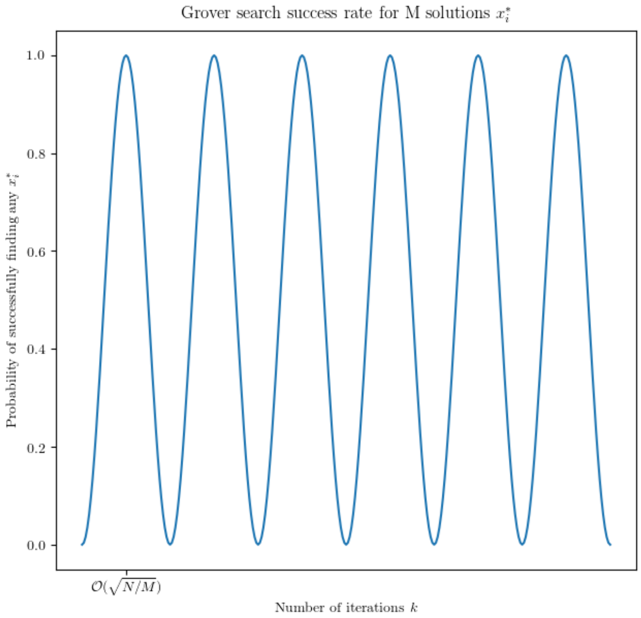
1. Success probability oscillates as we vary the number of queries k.



   So overshooting the optimal may decrease the probability. This is different from classical case where success probability can be made to increases monotonically with k.

2. The probability curve is dependent on M, the number of solution. This curves is valid only for $M = 1$.

When $M > 1$, we do the same analysis but with $\cos(\theta) = \sqrt{(N-M)/N}$, and the success probability now oscillates faster:



and the number of queries needed for high probability of success turns out to be $\mathcal{O}(\sqrt{N/M})$.

Since the probability of success is not monotone increasing, it seems like knowing M is important to being able to estimate k. But what if M is unknown?

The easy solution is to guess k uniformly randomly between 0 and $2\sqrt{N}$.

Picking a random k, we will end up somewhere with success probability at least $1/2$ with probability $1/4$, just by how area under the sine curve varies *(exercise:work out exact numbers)*. So if we measure we will get a solution with constant probability. Which means that trying this a constant number of times will get a solution with high probability with $\mathcal{O}(\sqrt{N})$ total queries.

Exercise: We can solve this search problem w.h.p. even is M is unknown.

Now the question is if this is the best possible performance for search. We will prove that this is indeed the best. You cannot do better than $\mathcal{O}(\sqrt{N/M})$ queries.

2.4 Postscript

We have so far seen two paradigms for quantum computing: Fourier transform-based and Grover-based. In fact, many quantum algorithms use ideas from one or the other.

Another paradigm that is important and distinct from these is quantum simulation. In this we simulate the behavior of a physical quantum system - which was the original motivation of Richard Feynman for coming up with the notion of quantum computers: simulating quantum physics. We will discuss simulation algorithms later in the course, but next we will establish the background material needed; namely, we will need to discuss Hamiltonians.

3 Hamiltonians

For understanding quantum simulation and topics in quantum machine learning, it is necessary to discuss Hamiltonians.

In the very first lecture, we learned of the postulate that isolated quantum systems (meaning no measurements occur) evolve over time through unitary evolution.

Starting with a state $|\psi\rangle$, it changes into a state $|\psi'\rangle = U|\psi\rangle$ for some unitary U.

While this notion is correct, it is not the most natural formulation of dynamics (change of a system over time) from a physics perspective.

The reasons for that are:

- *discrete perspective on time* – The unitary evolution does not really incorporate a notion of time. It treats time as being discrete (i.e. looking at our analysis of evolution through single states in a quantum circuit). In physics time is typically treated as continuous.
- *physics like forces* – Physical processes are rarely described in terms of unitaries. In high school physics, the laws of nature are usually described as forces like gravity or EM, and even in non high school physics you learn about things like weak and strong nuclear force. The continuous action of these forces on matter is what causes things to change continuously.

That is why it is much more natural for physicists to describe physics using Hamiltionians.

Hamiltonians are a convenient way to describe how states change in a continous manner.

Formally, a Hamiltonian H acting on $\mathbb{C}^d$ is a $d \times d$ complex Hermitian matrix.

Definition 1. *H is Hamiltonian H if* $H^\dagger = H$
i.e.,

$$H_{ij} = H^*_{ji}$$

In quantum mechanics the central equation is the Schrödinger Equation:

$$i\frac{\mathrm{d}|\psi\rangle}{\mathrm{d}t} = H|\psi\rangle$$

It is a diffferential equation that describes how quantum states evolve over time.

$|\psi(t)\rangle$ is a function of time. Solving the differential equation yields

$$|\psi(t)\rangle = e^{-iHt}|\psi(0)\rangle$$

where $|\psi(0)\rangle$ is the initial state at time 0 and e^{-iHt} is matrix exponential. It does not mean exponentiating every entry of H.

Theorem 2 (Spectral Theorem)**.** *Every Hermitian matrix* $M \in \mathbb{C}^{d\times d}$ *can be diagonalized: i.e., it exists an orthonormal basis* $\{|u_1\rangle, ..., |u_d\rangle\}$ *for* $\mathbb{C}^d$ *and real numbers* $\lambda_1, ..., \lambda_d$ *so that*

$$M = \sum_{j=1}^{d} \lambda_j |u_j\rangle\langle u_j|$$

The λ_j's are eigenvalues for the correspoding eigenvectors $|u_j\rangle$,

Functions of Hermitian matrices

Let $f : \mathbb{R} \to \mathbb{R}$ be a real function. Then for a Hermitian matrix $M = \sum_j \lambda_j |u_j\rangle\langle u_j|$,

$$f(M) = \sum_j f(\lambda_j)|u_j\rangle\langle u_j|$$

In other words, $f(M)$ is another Hermitian matrix such that its eigenvectors are the same as the ones of M, but its eigenvalues change.

Thus,

$$e^{iHt} = \sum_j e^{-i\lambda_j t}|u_j\rangle\langle u_j|$$

Notice that e^{iHt} is a unitary matrix, because all of its eigenvalues are on the unit circle.

$$|\psi(t)\rangle = U(t)|\psi(0)\rangle, \quad U(t) = e^{-iHt}$$

Given a Hamiltonian H describing a physical system, the eigenvalues and eigenvectors hold special significance.

The eigenvalues are called energies.

The eigenvectors are called energy eigenstates.

Since the eigenvalues of H are real, we can order them from smallest to biggest:

$$\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_d$$

λ_1 is called the ground energy.

$|u_1\rangle$ is called a ground state.

Why does this make sense? Let's say have a physical system in the state $|\psi\rangle$, and there is also an associated Hamiltonian H that describes the dynamics. The energy of $|\psi\rangle$ with regard to H is

$$\langle\psi|H|\psi\rangle \in \mathbb{R}$$

$$\begin{aligned} &= \sum_j \lambda_j \langle\psi|u_j\rangle\langle u_j|\psi\rangle \\ &= \sum_j \lambda_j |\langle\psi|u_j\rangle|^2 \end{aligned}$$

This measure of ernergy coincides with how energy is measured in many physical systems. For example, in classical mechanics, the energy measures sum of kinetic and potential energy in the system.

In physics, we are usually interested in the ground states, or more generally, the low-energy states of a Hamiltonian describing a physical system.

Why? This is because, as a principle of physics systems they tend to eveolve into lower energy configurations.

Why? Well, this has to do with the second law of thermodynamics, which states that this is the general trend of the universe.

So a very, very high level picture of what a lot of physics is, is:

1. write down a candidate Hamiltonian H describing your physical system
2. diagonalize H
3. try to analysze the low energy states of H; come up with qualitative and quantitative features of those states
4. experimentally test. Rinse, repeat.

Of course, this is a very simplistic cartoon picture of what physics is, and it hopefully did not offend anyone.

Doing each of these steps is generally incredibly hard!

Adding Hamiltonians

So why dow we work with Hamiltonians, rather just unitaries?

One of these reasons is because one can add Hamiltonians.

If H_1 and H_2 are Hamiltonians acting on $\mathbb{C}^d$, then $H = H_1 + H_2$ is also a Hamiltonianacting on $\mathbb{C}^d$.

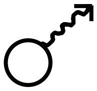
It means that two different "laws" or "constraints" are active simultaneously. For example, these could describe seperate forces.

We can easily build up a description of complex physical systems by adding up Hamiltonians.

For example, consider a bunch of particles in a 2-dimensional grid.

Let's model them as qubits. Let's furthermore say that the system has these rules:

- each qubit experiences a force described by a 2×2 Hamiltonian H_{single}



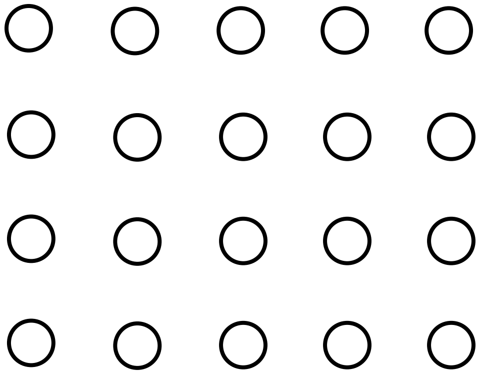


Figure 1: 20 qubit system

- each neighbouring pair of qubits experience an interaction described by a 4 × 4 Hamiltonian H_{pair}

These Hamiltonians are called local terms because they only act on a few qubits at a time.

The overall Hamiltonian can be written as

$$H = \sum_{i} H_{\text{single}}(i) + \sum_{i\ j} H_{\text{pair}}(i,j)$$

This Hamiltonian acts on $(\mathbb{C}^2)^n$. The qubits are labelled $1, ..., n$.

$H_{\text{single}}(i)$ is Hamiltonian H_{single} acting on the i^{th} qubit. It's like a quantum gate in the sense that

$$H_{\text{single}}(i) = H_{\text{single}} \otimes I$$

where H_{single} acts on $\mathbb{C}^2$ and the i^{th} qubit and I acts on $(\mathbb{C}^2)^{\otimes(n-1)}$.

Similiarry, $H_{\text{pair}}(i,j)$ acts on the neighbouring qubits i and j, and trivially on all other qubits.

H is called a local Hamiltonian, because it is written as a sum of local terms.

In general, diagonalizing a local Hamiltonian like H - or even just estimating the ground energy of it - is extraordinarily difficult. Later in the course, we will see that trying to diagonalize or compute ground energies of local hamiltonians is harder than solving NP complete problems. In fact, this is complete for the quantum analogue of NP, called *Quantum Merlin-Arthur* (or *QMA*).

Why is it so hard, though? We'll get into this in more depth later on, but here's the intuition. Think of the $H_{\text{pair}}(i,j)$ and $H_{\text{single}}(i)$ as constraints, and think of the qubits as variables.

You're given a bunch of constraints, and your task is to find an assignment of values to the variables such that is minimizes the number of violated, unsatisfied constraints.

This should tell you that this problem is at least NP-hard, because you can encode your favorite NP-complete problem as finding the ground stat of a local Hamiltonian.

This perspective will come up in quantum machine learning. Many optimization problems can be cast in terms of finding low energy states of local Hamitonians.

4 Mixed States and Open Quantum Systems

Let's revisit another basic setup of quantum information theory. Up to now, quantum states are complex unit vectors. However, such a $|\psi\rangle \in \mathbb{C}^d$ only describes a *pure state*. This means that the state of the system is deterministic, and there is no uncertainty. Of course, if we measure the state, there could be uncertainty, but if we leave the system alone, it is fixed.

This is not the best way to describe the state in all situations. If we want to talk about probabilistic mixtures of pure quantum states, we use *mixed states*. For example, a qubit can be in the state $|\psi_0\rangle$ with probability p and in the state $|\psi_1\rangle$ with probability $1-p$. This is a perfectly valid thing to talk about, and it cannot be described as a pure state. This is what happened to cause the quantum weirdness in the first lecture, where we saw a demonstration of interference in Experiment B but not in Experiment A.

The unwieldy way to denote a mixed state is to write $\{(|\psi_i\rangle, p_i)\}$, meaning it is in the state $|\psi_i\rangle$ with probability p_i. This is called an *ensemble*. It can better be described by a density matrix

$$\rho = \sum_i |\psi_i\rangle\langle\psi_i| \in \mathbb{C}^{d\times d}.$$

Density matrices have the following properties:

1. They are Hermitian.
2. They are positive definite (all the eigenvalues are non-negative).
3. Their trace is 1.

Proof. Since $\rho^\dagger = \sum_i (p_i|\psi_i\rangle\langle\psi_i|)^\dagger = \sum_i p_i|\psi_i\rangle\langle\psi_i|$, the density matrix obtained from an ensemble is clearly Hermitian.

If the pure states chosen were orthonormal, it is also automatic that ρ is positive semi-definite.

Finally, $\mathrm{Tr}(\rho) = \mathrm{Tr}(\sum_i p_i|\psi_i\rangle\langle\psi_i|) = \sum_i p_i \mathrm{Tr}(|\psi_i\rangle\langle\psi_i|) = \sum_i p_i = 1$. □

Conversely, one can obtain an ensemble, non-uniquely, from a density matrix by applying the spectral theorem. There are multiple ensembles with the same density matrix, but this is fine as ensembles with equal density matrices are correspond to states which are undistinguishable.

Examples:

1. For a pure state $|\psi\rangle$, the density matrix if $\rho = |\psi\rangle\langle\psi|$.
2. The density matrix $\frac{1}{2}(|0\rangle\langle 0| + |+\rangle\langle +|)$ is not written in an orthonormal decomposition.
3. Given a density matrix $\rho = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|$, the ensemble is $\{(|0\rangle, \frac{1}{2}), (|1\rangle, \frac{1}{2})\}$.

Evolution A unitary U acts on a density matrix ρ by conjugation $\rho \mapsto U\rho U^\dagger$.

Measurement Let $B = \{|v_1\rangle, \ldots, |v_d\rangle\}$ be an orthonormal basis. The probability of measuring outcome $|v_j\rangle$ is $\langle v_j|\rho|v_j\rangle \geq 0$. The post-measurement state of the density matrix is $|v_j\rangle\langle v_j|$, which is a pure state.

Mixed states are necessary for describing parts of an entangled state, such as $|\Phi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. There is no way of describing any subsystem of $|\Psi\rangle$ as a pure state. The state of the left qubit is $\rho = \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|) = \frac{1}{2}(|+\rangle\langle +| + |-\rangle\langle -|)$. This is the maximally mixed state. It is different from the pure state corresponding to the superposition $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$.

5 Noise and Decoherence

Why don't we see quantum effects all the time? Because when quantum states are measured, the information disappears. It is very difficult to avoid doing measurement. But, who is doing the measurement? Hopefully, it is the experimenter, but a particle from the environment hitting the qubit can be considered a measurement. Quantum states are fragile because of such spurious (unintended) measurements and unitaries. Quantum states are rarely isolated. While the system evolves according to a Hamiltonian, it also belongs to a bigger system, the environment and inevitably interact with it. Environment influencing qubit in the system is called *noise*. This interaction is often detrimental to quantum information, as the environment essentially steals away the information.

Simple example The desired state of the first qubit at the end of the circuit is $|+\rangle\langle +|$. This density matrix called *coherent* since it has strong off-diagonal terms. This is the signature of quantum-ness, showing a superposition of classical states. However, because of the interaction with the environment, the observed state is $\frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|)$ it is called a *decohered* state.

Battling noise and decoherence is the biggest issue facing the construction and scaling of quantum computers. We will talk more about ways to deal with errors in quantum systems towards the end of the course.