Quantum Information Part III review

Shelby Kimmel

February 16, 2015

Part I
States, Operators, and Measurements

1 States

- **Density Matrix:** $\rho = \sum_k p_k |\psi_k\rangle\langle\psi_k|$ with $\sum_k p_k = 1$ and positive. Any density matrix has an orthonormal decomposition $\rho = \sum_i p_i |i\rangle\langle i|$ ($|i\rangle$ are eigenstates, p_i are eigenvalues.

- **Bloch Sphere:** Only for qubits. $\rho = \frac{I + \vec{r}\vec{\sigma}}{2}$, where r is a vector with length less than 1 and σ are the Pauli matrices. Pure states on surface, mixed in interior

- **Subsystems:** Let $\rho_{12} = \sum_{i,i',j,j'} p_{i,i',j,j'} |i\rangle\langle i'||j\rangle\langle j'|$ be a bipartite pure state. Then $\rho_1 = Tr_2(\rho_{12}) = \sum_{i,i',j,j'} p_{i,i',j,j'} |i\rangle\langle i'|\langle j'|j\rangle$

- **Entanglement of pure states:** The **entropy** of a state ρ is $S(\rho) = -\sum_k \lambda_k \log_2 \lambda_k$ where λ_k are the eigenvalues of ρ. Then $|\psi_{12}\rangle$ is entangled iff $S(Tr_2(|\psi_{12}\rangle\langle\psi_{12}|)) > 0$.

- **Schmidt Decomposition:** If $|\psi\rangle_{AB}$ is a pure bipartite state, it can be written as

$$|\psi\rangle \sum_i p_i |i_A\rangle |i_B\rangle \tag{1}$$

where $\{|i_A\rangle\}$ are orthogonal states for system A, and $\{|i_B\rangle\}$ are orthogonal states for system B.

Proof. We can always write $|\psi\rangle = \sum_{ij} a_{ij} |_A\rangle |j_B\rangle$. Then any square matrix a can be decomposed using the singular value decomposition into udv where d is diagnoal and u and v are unitary. So we have $\sum_{ikj} u_{ik} d_{kk} v_{kj} |i_A\rangle |j_B\rangle$. Let $|k_A\rangle = \sum_i u_{ik} |i_A\rangle$ and $|k_B\rangle = \sum_j v_{kj} |j_B\rangle$ and $p_k = d_{kk}$. (Note this satisfies the orthogonality constraint since u and v just perform a change of basis. The **Schmidt Order** is the number of terms in the sum. More than 1 means entangled. □

- **Purification** Suppose have a density matrix to purify. Can always write in orthonormal decomposition $\rho = \sum_i p_i |i\rangle\langle i|$. Let $|\psi\rangle = \sum_i \sqrt{p_i}|i\rangle|i\rangle$. Then the partial trace is just ρ. **Two different purifications are related to each other by a unitary acting only on ρ.**

- For any states ρ and σ, can find Q and S such that $\rho - \sigma = Q - S$, where Q and S are positive operators with support on orthogonal subspaces.

2 Measurements

- **Measurement:** Is a set of any matrices $\{M_i\}$, such that $\sum M_i^\dagger M_i = \mathbb{1}$.
  - Probability of outcome i for a state $|\psi\rangle$ is $p_i = \langle\psi|M_i^\dagger M_i|\psi\rangle$.
  - State after outcome i is $\frac{M_i|\psi\rangle}{\text{normalization}}$

- **POVM Measurement** Set of positive operators $\{E_i\}$ such that $\sum_i E_i = \mathbb{I}$
  - Probability of outcome is $p_i = \langle\psi|E_i|\psi\rangle$.
  - Normally use when don't care about outcome after measurement, but can get M_i as above by taking $\sqrt{E_i}$.

- **Projective Measurement** Given a Hermitian operator (i.e. with measureable observables) A. We can write $A = \sum_i e_i P_i$ where e_i is an eigenvalue of A with eigenvector $|\psi_i\rangle$ and $P_i = |\psi_i\rangle\langle\psi_i|$.
  - Probability of outcome $p_i = \langle\psi|P_i|\psi\rangle$ and measure e_i.
  - After measurement, state is in state $\dfrac{P_i|\psi\rangle}{normalization}$.
  - If have a state ρ, then the expectation value of the measurement A is $\text{Tr}(\rho A)$.

3 Operations

Definition 3.1. *Most generally, a quantum operation is a set of operators $\mathbb{O} = \{E_i\}$ such that $\sum_i E_i^\dagger E_i = \mathbb{I}$, and the operation acts as*

$$\rho \to \sum_i E_i \rho E_i^\dagger$$

This is called the **Kraus Operator Sum Notation**. *It is not unique! $\{E_i\}\tilde{\{F_i\}}$ iff there is a unitary u such that $F_i = \sum_j u_{ji} E_i$.*

Some examples:

- Unitary map: $\mathbb{O} = U$

- Depolarizing Channel: $\mathbb{O} = \{\sqrt{1-3p/4}\mathbb{I}, \sqrt{p/4}\sigma_i\}$ for each i. This sends $\rho \to (1-p)\rho + p\mathbb{I}/d$.

Part II
Algorithms

1 Tools

- Rotation by θ about the j-axis: $e^{-i\theta/2\sigma_j} = \cos\theta/2\mathbb{I} - i\sin(\theta/2)\sigma_j$
- Can do above with any operator whose square is the identity:
  $e^{-i\theta\sigma_j\otimes\sigma_k} = \cos\theta\mathbb{I} - i\sin(\theta)\sigma_j \otimes \sigma_k$
- $H^{\otimes n}|y\rangle = \sum_x (-1)^{x\cdot y}|x\rangle$
- Phase Kickback. If $U|x\rangle|y\rangle = |x\rangle|f(x)\oplus y\rangle$ then $U|x\rangle(|0\rangle - |1\rangle) = (-1)^{f(x)}|x\rangle(|0\rangle - |1\rangle)$
- $\sum_{a=1}^{N} w^{ab} = N\delta_{b,Nm}$ where $w = e^{-2\pi i/N}$ and Nm is any multiple of N, for $m \in \mathbb{Z}$.
- Any hermitian matrix h can be diagonalized by a unitary matrix u, i.e. $d = u^\dagger h u$ where d is diagonal.
- POVMs, tr(Aρ), etc.

2 Grover's Search

Grover Operator is $G = [2|\psi\rangle\langle\psi| - \mathbb{I}]O_f$, where $|\psi\rangle = 1/\sqrt{N}\sum_x |x\rangle$. Consider how G acts on the states:

- $|\alpha\rangle = \dfrac{1}{\sqrt{N-m}}\sum_{x\notin sol}|x\rangle$
- $|\beta\rangle = \dfrac{1}{\sqrt{m}}\sum_{x\in sol}|x\rangle$

Notice $|\psi\rangle = \cos(\theta/2)|\alpha\rangle + \sin(\theta/2)|\beta\rangle$ where $\cos(\theta/2) = \sqrt{(N-m)/N}$ and $\sin(\theta/2) = \sqrt{(m)/N}$. Then

- $O_f|\alpha\rangle = |\alpha\rangle$
- $O_f|\beta\rangle = -|\beta\rangle$
- $2[|\psi\rangle\langle\psi| - \mathbb{I}]|\psi\rangle = |\psi\rangle$
- $2[|\psi\rangle\langle\psi| - \mathbb{I}]|\psi'\rangle = -|\psi'\rangle$ where $|\psi'\rangle = \cos(\theta/2)|\beta\rangle - \sin(\theta/2)|\alpha\rangle$ since $\langle\psi|\psi'\rangle = 0$.

So we see that G only moves the state in the 2-dimensional state spanned by $|\alpha\rangle$ and $|\beta\rangle$. O_f causes a reflection over the state $|\alpha\rangle$ and the rest causes a reflection over the state $|\psi\rangle$. See Figure **??**. Since we start at $\theta/2$ and we can move by θ each time, and we want to end up at angle $\pi/4$, we have Q(SEARCH)$= (\pi/4-\theta/2)/\theta \approx (\pi/4-\sqrt{m/N})/(2\sqrt{m/N}) = O(\sqrt{N/m})$.

3 Amplitude Amplification

Suppose have a quantum or classical process that succeeds with probability p, then a quantum walk over the two subspaces (succeed and not succeed) can end up in the succeed case by using $O(1/\sqrt{p})$ repitions of the process. (Since $\sqrt{p} > p$.)

4 Quantum Fourier Transform

4.1 Basics

Classical Discrete Fourier Transform. For $f : \{0,1\}^n \to D$:

$$\mathbb{F}(f(x)) \to \tilde{f}(k)$$

where

$$\tilde{f}(k) = \frac{1}{\sqrt{2^n}} \sum_x f(x) w^{xk/N}$$

where $w = e^{-2\pi i}$ and $N = 2^n$. Note if f' is f shifted by l, then only phase shifts:

$$\tilde{f}'(k) = \frac{1}{\sqrt{2^n}} \sum_x f(x+l) w^{xk/N} = \frac{1}{\sqrt{2^n}} \sum_x f(x) w^{(x-l)k/N} = w^{-lk/N} \tilde{f}(k)$$

Quantum Fourier Transform.

$$|x\rangle \to \frac{1}{\sqrt{2^n}} \sum_k w^{xk/N} |k\rangle$$

Useful fact. RHS equals

$$\otimes_{l=1}^n (|0\rangle + w^{x2^{-l}}|1\rangle) = (|0\rangle + w^{0.x_n}|1\rangle) \otimes (|0\rangle + w^{0.x_{n-1}x_n}|1\rangle$$

where $0.x_n$ is binary decimal with x_n the n^{th} binary bit of x. LHS=RHS because when divide by 2, say, all of the first $n-1$ digits remain as integers and when w is raised to integers it is just 1.

Proof. (for later) □

Algorithm (for later)

4.2 Phase Estimation

Need an eigenstate of U, and t qubits to record the phase to the t^{th} binary decimal. Put recorder qubits into equal superposition of all states. Then controlled on the last qubit, act on the eigenstate with U. This will give a phase $e^{i0.\phi_1 \dots \phi_t}$ to the $|1\rangle$ state of the last qubit. Controlled on the second to last qubit, act on eigenstate with U^2. This will give a phase $e^{i0.\phi_2 \dots \phi_t}$ to the $|1\rangle$ state of the second to last qubit. Continue working up until controlled on the first qubit, act with $U^{2^{t-1}}$ on the eigenstate. But now, the record qubits are in a fourier transformed state, and can just do an inverse Fourier transform to calculate the eigenstate phase.

4.3 Hidden Abelian Subgroup

Function: $f : G \to S$ with G a group and S is a set.
Promise: $f(x) = f(y) \leftrightarrow x = yh$ where $h \in H$ and H is a subgroup of G and multiplication is the group operation. Notice if x and y are in the same coset, then they have the same value: $y = gh_1$ and $x = gh_2 \to yh_1^{-1} = xh_2^{-1} \to y = xh_2^{-1}h$. If they have the same value then $x = yh$ and x is in the coset started by y.
Background: Any finite abelian group can be decomposed (efficiently using a quantum computer) into finite cyclic groups $\to g = {g_1}^{n_1}{g_2}^{n_2} \dots g_k^{n_k}$ where g_i is the generator of a cyclic group of size N_i. Write $g \in G$ as $(n_1, \dots, n_k)$. Then the character (homomorphism to complex numbers) of a representation of g is $\xi_h(g) = \prod_i w^{n_i h_i / N_i}$ for some set $\{h_i\}$, since the character of g_i can be w^{h_i/N_i} for any $h_i \leq N_i \in \mathbb{N}$ (characters turn group operations into multiplication, trace of identity is 1). Note $h\tilde{\{}h_i\}$ also represents an element of G, so the fourier transform is

$$|g\rangle \to \sum_{h \in G} \prod_i w^{h_i n_i / N_i} |h\rangle.$$

Can write this as

$$|g\rangle \to \sum_{h \in G} \xi_h(g) |h\rangle.$$

Algorithm:

$$\begin{aligned} |0,0\rangle &\to \sum_g |g,0\rangle && (2)\\ &\to \sum_g |g,f(g)\rangle && (3)\\ &= \sum_{\text{cosets:cgenerates}} \sum_{h \in H} |ch, f(c)\rangle && (4) \end{aligned}$$

Now throw out the last register and QFT

$$\to \sum_{k \in G} \sum_{h \in H} \xi_k(ch) |k\rangle \quad (5)$$

But characters turn group operations into normal multiplication, so this is

$$\rightarrow \sum_{k\in G} \xi_k(c) \sum_{h\in H} \xi_k(h)|k\rangle \quad (6)$$

$$= \sum_{k\in G} \xi_k(c) \sum_{h\in H} \prod_i w^{k_i h_i/N_i}|k\rangle \quad (7)$$

Will only measure k such that k is in the kernel of H. (That is $kh = 1$ for $h \in H$.) Do this a few times, and get a bunch of linear equations. (max i is log in the size of the group, so at most log generators of H, each with at most log terms to figure out.) Can solve to find h_i's since know N_i and the k_i's.

5 Simon's Algorithm

Function: $f : \{0,1\}^n \rightarrow \{0,1\}^n$
Promise: $f(x) = f(y) \leftrightarrow x = y \oplus s$ where $s \in \{0,1\}^n$.
Algorithm: (H - Oracle - H -Measure Second Register)

1. $H^{\otimes n}|0\rangle \rightarrow \sum_x |x\rangle \rightarrow O_f \sum_x |x\rangle \rightarrow \sum_x |x\rangle|f(x)\rangle$
2. $H^{\otimes n} \sum_x |x\rangle|f(x)\rangle \rightarrow \sum_{x,y}(-1)^{x\cdot y}|y\rangle|f(x)\rangle$
3. Measure 2^{nd} register
   - If f is 1-to-1 $\rightarrow \sum_y(-1)^{x\cdot y}|y\rangle$. Even amplitude on all.
   - If f is 2-to-1 $\rightarrow \sum_y(-1)^{x\cdot y} + (-1)^{(x\oplus s)\cdot y}|y\rangle$. Zero amplitude for $s \cdot y = 1$. So whatever measure learn a y such that $y \cdot s = 0$.

Repeat $n-1$ times, get $n-1$ linear equations. As long as equations are linearly independent can solve. Prob that the k^{th} is linear independent is probability it is in the free subspace, which has size $1 - \frac{1}{2^{n-k+1}}$. So the total probability is more than $\prod_i(1 - \frac{1}{2^i}) = .288$.

Classically suppose you choose to measure $2^{n/4}$ bits. Then can compare less than $2^{n/2}$ pairs. The probability that any individual pair satisfies $x = y \oplus s$ is 2^{-n} (since there are 2^n options for s). Thus have $2^{-n/2}$ probability of getting a good pair.

5.1 Factoring

Back Story
Let $N = pq$, with p and q prime. Choose $x \leq N-1$, coprime to N. Let r be the minimum number s.t. $x^r \mod N = 1$. Then if r is even and $x^{r/2} \neq -1 \mod N$ then $\gcd(x^{r/2} \pm 1 \mod N, N)$ is p or q.

Algorithm

$$|0,0\rangle \rightarrow \sum_a |a,0\rangle \tag{8}$$

$$\rightarrow \sum_a |a, x^a \mod N\rangle \tag{9}$$

$$= \sum_{l=0}^{r-1} \sum_{j=0} |jr+l, x^l \mod N\rangle \tag{10}$$

Now throw out the last register and QFT

$$\rightarrow \sum_j \sum_k w^{k(jr+l)/N} |k\rangle \tag{11}$$

$$= \sum_k w^{kl/N} \sum_j w^{kjr/N} |k\rangle \tag{12}$$

$$\tag{13}$$

$|k\rangle$ has 0 probability unless $k = 0$ or $kjr = mN$. Get k s.t. k is a multiple of mN/r. After a constant number of measurements, obtain r.

6 Element Distinctness and Collision

6.1 Collision

Problem: Check if 2-to-1 or 1-to-1.

Quantum algorithm is just grover: pick k elements at random. Check no collision. If 2-1, must be those k elements again in the $n-k$ remaining. Mark all k elements and do Grover, so there are k marked items out of n. Takes $O(\sqrt{(n-k)/k})$. Total is $O(k + \sqrt{n/k})$, and when make $k = n^{1/3}$ minimizes.

Polynomial method for tight lower bound

Classically, any randomly chosen pair has probability $1/n$ of having the same value. If randomly choose m, can look at $\binom{m}{2}\tilde{m}^2$ pairs. So to get high likelihood of choosing a pair, need $m = \sqrt{n}$.

6.2 Element Distinctness

Problem: Check if all elements are distinct.

Less good quantum algorithm: (like Collision) pick k elements at random. Check to see if two are the same. Mark all of those items and do Grover on the rest. At best there is only item that actually repeats and so gets marked, so there is 1 marked item out of n. Probability that the paired item was in the random k elements to begin with is k/n, so have k/n probability of succeeding. Use amplitude amplification for total complexity of $O((k + \sqrt{n})\sqrt{n/k})$. When $k = \sqrt{n}$, get $O(n^{3/4})$.

Better algorithm: use walk on a graph. Each node of the graph is associated with a set of $\{x_1, \ldots, x_k\}$. When are on a node, might also query some of the inputs on that node. Let

δ be the spectral gap of the adjacency matrix of the graph (distance between largest and second largest eigenvalue), ϵ be the fraction of marked nodes (in our case, nodes that contain x_i and x_j s.t. $f(x_i) = f(x_j)$, let S be the number of queries necessary to set up the initial state. U be the number of queries to update from one position in the walk to the next, and C be the checking cost to check if the node you are on is marked (if already have queried all elements at node, this might be 0). Then the quantum algorithm to find a marked item requires:

$$S + \frac{1}{\sqrt{\epsilon}}(\frac{1}{\sqrt{\delta}}U + C).$$

For E.D. use Hamming graph, and nodes with $k = n^{2/3}$.

Lower bound: Collision reduces to element distinctness; take $\sqrt{n}$ elements from collision problem, and by birthday principle, should have all elements distinct except 2 (if 2-to-1), so run element distinct algorithm. If could solve element distinctness in less than $\sqrt{n}^{2/3}$ queries, would beat the lower bound of $n^{1/3}$ for collision. So has to have a lower bound of $n^{2/3}$.

Classically probability that a randomly chosen pair is the matched pair is $1/n^2$. From collision, if randomly choose m, can look at $\binom{m}{2}\tilde{m}^2$ pairs. So to get high likelihood of choosing a pair, need $m = n$.

Part III
Lower Bounds

1 Adversary Bound

1.1 Set-Up

We have an oracle for a string x, O_x, such that $O_x|i, b\rangle = |i, b \oplus x_i\rangle$, and our goal is to determine $f(x)$. We act on an initial state $|\psi^0\rangle$, alternating between the oracle and unitary operators, to get a final state:

$$|\psi_x^t\rangle = U^t O_x U^{t-1} O_x \dots O_x U^1 O_x |\psi^0\rangle$$

We say $|\psi_x^k\rangle = U^k O_x |\psi_x^{k-1}\rangle$.

To distinguish two states with probability ϵ, require $|\langle\psi_x^t|\psi_y^t\rangle| \leq 2\sqrt{\epsilon(1-\epsilon)}$. So, for the algorithm to always succeed, need that

$$\mathbb{E}_{x,y:f(x)\neq f(y)}|\langle\psi_x^t|\psi_y^t\rangle| \leq 2\sqrt{\epsilon(1-\epsilon)}.$$

(Can interpret as letting an adversary choose a superposition of the most difficult set of oracles to use. Then add our own weighting that give greater weight to things that are difficult to tell apart).

Then we want to consider the following weighting function:

$$W^t = \sum_{x,y} a_x^* \Gamma_{xy} a_y \langle\psi_x^t|\psi_y^t\rangle$$

where Γ is our weighting function $\Gamma_{xy} = 0$ if $f(x) = f(y)$ because we want to weight the difficult things, not the easy, $\sum_x a_x^2 = 1$, (the oracle weighting). and $\Gamma_{xy} \geq 0$ and $\Gamma_{xy} = \Gamma_{xy}$. So throughout, while we sum over x and y, all terms with x and y such that $f(x) = f(y)$ are zero.

The the initial function is

$$W^0 = \sum_{xy} a_x^* \Gamma_{xy} a_y.$$

The final function is

$$W^t \leq \sum_{xy} a_x^* \Gamma_{xy} a_y 2\sqrt{\epsilon(1-\epsilon)}.$$

We will take the difference and divide by the maximum change W^t can undergo, so we need to consider the maximum value of $\sum_{xy} a_x^* \Gamma_{xy} a_y = \|\Gamma\|$, when a is a maximum valued eigenstate of Γ.

Now we just need to determine how much W^t can change with each step:

$$|W^t - W^{t-1}| = |\sum_{xy} a_x^* \Gamma_{xy} a_y (\langle \psi_x^t | \psi_y^t \rangle - \langle \psi_x^{t-1} | \psi_y^{t-1} \rangle)| \tag{14}$$

$$= |\sum_{xy} a_x^* \Gamma_{xy} a_y \langle \psi_x^{t-1} | O_x^\dagger U^{t\dagger} U^t O_y | \psi_y^{t-1} \rangle - \langle \psi_x^{t-1} | \psi_y^{t-1} \rangle)| \tag{15}$$

$$= |\sum_{xy} a_x^* \Gamma_{xy} a_y \langle \psi_x^{t-1} | (O_x^\dagger O_y - \mathbb{I}) | \psi_y^{t-1} \rangle|. \tag{16}$$

If $O_x O_y |i, b\rangle = (-1)^{b(x_i \oplus y_i)} |i, b\rangle$, $O_x O_y - \mathbb{I} |i, b\rangle$ is 0 if $b = 0$ or if $x_i = y_i$ and $-2|i, b\rangle$ otherwise. So $O_x O_y - \mathbb{I} = -2 \sum_{i: x_i \neq y_i} P_i$ where P_i is the projector onto $|i, 1\rangle\langle i, 1|$.

$$|W^t - W^{t-1}| = 2|\sum_{xy} \sum_{i: x_i \neq y_i} a_x^* \Gamma_{xy} a_y \langle \psi_x^{t-1} | P_i | \psi_y^{t-1} \rangle| \tag{17}$$

$$\leq 2 \sum_{xy} \sum_{i: x_i \neq y_i} a_x^* \Gamma_{xy} a_y |\langle \psi_x^{t-1} | P_i | \psi_y^{t-1} \rangle| \tag{18}$$

$$\leq 2 \sum_{xy} \sum_{i: x_i \neq y_i} \Gamma_{xy} \|a_x P_i | \psi_x^{t-1} \rangle\| \cdot \|a_y P_i | \psi_y^{t-1} \rangle\| \tag{19}$$

Let Γ^i be a matrix that has 0 values for Γ^i_{xy} if $x_i \neq y_i$ and the value of Γ otherwise. Let $|v_i\rangle$ have components $||a_y P_i | \psi_y^{t-1} \rangle||$.

$$|W^t - W^{t-1}| \leq 2 \sum_i \langle v_i | \Gamma_i | v_i \rangle \tag{20}$$

$$\leq 2 \sum_i \|\Gamma_i\| \| |v_i\rangle \|^2 \tag{21}$$

$$\tag{22}$$

But since P_i is a projector and $\sum_y a_y^2 = 1$, $\sum_i \| |v_i\rangle \|^2 \leq 1$,

$$|W^t - W^{t-1}| \leq 2 \max_i \|\Gamma_i\|.$$

So we have that the number of oracle calls is at least

$$\frac{\|\Gamma\|\sqrt{\epsilon(1-\epsilon)}}{\max_i \|\Gamma_i\|}.$$

2 Polynomial

For a function $f : \{0,1\} \to \{0,1\}$, where $f(i) = x_i$, there exists a polyvariate function (function of all the variables $x_1, x_2, \dots$) $p(x)$ with lowest degree possible, such that $p(x) = f(x)$ for all x.

Lemma 2.1. *The acceptance probability of a q query quantum algorithm corresponds to a polyvariate polynomial of degree less than or equal to $2q$*

Proof. Each call to the oracle adds a phase $(-1)^{x_i \oplus y} = (1 - 2(x_i \oplus y))$. So every time you call the oracle, degree of polynomial that is the amplitude of any state increases by at most 1. Unitaries are linear, so can't change degree. When you measure, take the square of the amplitude. □

So if the smallest degree polynomial $p(x)$ that corresponds to $f(x)$ has degree $deg(f)$, then you need at least $deg(f)/2$ queries.

But because we only need to approximate the polynomial, say there is another polynomial $g(x)$ such that $|g(x) - f(x)| < 1/3$. Then with bounded error, this is still OK. We call the degree of such an approximating polynomial $\widetilde{deg}(f)$. But the quantum algorithm must have degree at least of $\widetilde{deg}(f)$, which requires $\widetilde{deg}(f)/2$ queries.

Since the x_i's are either 0 or 1, we don't need to consider higher powers (since $x_i^2 = x_i$), so we can consider multi*linear* polynomials:

$$g = \sum_{S \in \{0,1,\dots,n\}} c_S \prod_{i \in S} x_i.$$

We want to reparametrize over a variable k, where $k = |x|$ = the number of x_i that have value 1. So average g over x with $|x| = k$ to get $G(k)$:

$$G(k) = \sum_{S \in \{0,1,\dots,n\}} c_S \mathbb{E}_{|x|=k} \prod_{i \in S} x_i.$$

But the product is only 1 if all are 1, and since average is $0 \times p(0) + 1 \times p(1)$ we just need the probability that all the x_i in for $i \in S$ are 1.

$$\mathbb{E}_{|x|=k} \prod_{i \in S} x_i \quad = \quad \frac{\binom{n-S}{n-k}}{\binom{n}{k}} \tag{23}$$

Here the top fraction is the number of ways of choosing the elements in S to have value 1 (which equals the ways of arranging the remaining $n-k$ 0-valued elements in $n-|S|$ spaces. The bottom fraction is the ways to have k of n bits have value 1. The RHS is

$$= (n-|S|)!/n! \times \frac{k!(n-k)!}{(n-k)!(k-|S|)!} \quad (24)$$

$$= (n-|S|)!/n! \times k(k-1)\dots(k-|S|+1) \quad (25)$$

We see that $G(k)$ is a monomial with degree less than or equal to that of $g(x)$ (since here we get a $k^{|S|}$ and before we got up to $|S|$ x_i's multiplied together.

So for functions that depend on k, we can look at $\tilde{deg}(G)$.

Examples:

Parity: For example, for parity, we can plot G (including the 1/3 error,) see Fig **??**.

Because of the separation between 2/3 and 1/3, any polynomial that satisfies must have n wiggles, and so must have degree n, so $Q(Parity) = n/2$.

Search

For search, we have $k = 0$ between $\pm 1/3$ and all the other k's between 2/3 and 4/3.

Lemma 2.2. *Let $f(x)$ be a polynomial $f : \mathbb{R} \to \mathbb{R}$, then*

$$\max_{x\in[0,n]} \frac{df(x)}{dx} =\leq \frac{deg(f)^2}{n}(\max_{x\in[0,n]} f(x) - \min_{x\in[0,n]} f(x))$$

So this means $deg(f) \geq \sqrt{nd/h}$ where d is the derivative, and h the difference between max and min. We know that d must be at least 1/3 because the function must increase by 1/3 between $k = 0$ and $k = 1$. However, the function might do crazy things in between integer values. Suppose it's max value is $m > 4/3$. Then it must have a d of at least $(m - 4/3)/2$. Similarly for the min value. But either the max or min is basically at least half of h, so we have $d \geq (h/2 - 4/3)/2$. So

$$deg(f) \geq \sqrt{\frac{n \max\{1/3, (h/2 - 4/3)/2\}}{h}}$$

So we see that the h parts cancel out, and $deg(f) \geq O(\sqrt{n})$.

Part IV
Models of Quantum Computation

1 Adiabatic

1 Cluster Computing

Part V
Building Blocks of Quantum Computers

1 Universal Gate Sets

Part VI
Error Correcting

An $[n, k, d]$ code uses n qubits to encode k qubits, and can correct errors on up to $d - 1$ qubits.

1 Shor Code 9 qubit code

The coode and correction is described in Figure **??**. Let P be the projector onto the states $|000\rangle\langle 000| + |111\rangle\langle 111|$, $|100\rangle\langle 100| + |011\rangle\langle 011|$, $|010\rangle\langle 010| + |101\rangle\langle 101|$, $|001\rangle\langle 001| + |110\rangle\langle 110|$. The syndrome is which outcome of the projective measurement is obtained. Depending on syndrome, either do nothing, or apply $\sigma_x = X$ to the corresponding qubit.

To correct for phase errors, after control gates in circuit shown in Figure **??**, apply H to each qubit. This turns $|0\rangle \to |+\rangle$ and $|1\rangle \to |-\rangle$. Then the projector is the same except with the same substitution, and the correction is to apply a $\sigma_z = Z$ operator to the offending qubit.

Then, just concatenate the two codes to be able to correct one phase and one bit-flip errors. But, since any error can be written as a sum of phase, bit-flip, and combination of phase and bit-flip, the projection will collapse to a single type of error (out of the continuum of errors), which can then be corrected.

How does it do? Use **Fidelity**

Definition 1.1. *For a pure state $|\psi\rangle$ and a mixed state ρ, the Fidelity of ρ is*

$$F = \sqrt{\langle\psi|\rho|\psi\rangle} \tag{26}$$

Without correction (since only 1 qubit):

$$F = \sqrt{(1-p) + p|\langle \psi X \psi \rangle|^2} > \sqrt{1-p}$$

With correction (more places for error, since now have 3 qubits):

$$F = \sqrt{(1-p)^3 + 3p(1-p)^2 + ...|\langle \psi X \psi \rangle|^2} > \sqrt{1-3p^2}$$

2 Criteria for Error Correcting Codes

For a code to be reasonable, the code words much be orthogonal. Otherwise you wouldn't know what word you have.

Theorem 2.1. *Let $\{E_i\}$ be the Krauss operators for the error operation. Let P be the projector onto the code space (sum of projectors of code words). Then*

$$PE_j^\dagger E_i P = \alpha_{ij} P \tag{27}$$

with α_{ij} a constant, is sufficient and necessary for a code to correct errors. Note that this is the same as

$$\langle \psi | E_j^\dagger E_i | \psi \rangle = \alpha_{ij}.$$

Proof. **Sufficient** We will show that there is a way to correct the errors.

1. From the form of α it is hermitian. So we can write as $\alpha_{ij} = u_{ik}^* d_{kk} u_{kj}$ (see useful math at beginning)

2. Using same u matrices, we can re-express the error operationlet $F_i = \sum_j u_{ji} E_j$ (by useful math, $\{F_i\}$ are just another representation of the same error operation.

3. So
$$PF_j^\dagger F_i P = \sum_{kl} u_{jk}^\dagger u_{li} P E_k E_l P = \sum_{kl} u_{jk}^\dagger u_{li} \alpha_{kl} = \delta_{ij} d_{jj} P$$

4. Using Polar Decomposition $F_j P = U_j \sqrt{P F_j^\dagger F_j P} = \sqrt{d_{jj}} U_j P$

Thus the effect of F_j is equivalent to a unitary U_j. So can measure the syndrome using $U_k^\dagger P U_k$, (which in essence looks at the space spanned by the code space after it is affected by that unitary) and fix using $U_k^\dagger$.

Necessary For any quantum state ρ, we want the recovery scheme to work on $P\rho P$ (the projection onto the code space). By work, we mean $R(E(P\rho P)) = cP\rho P$ where R is the recovery operation E is the error operation and c is a constant, because error might not be trace preserving. Written out, this means

$$\sum_{i,j} R_j E_i P \rho P E_i^\dagger R_j^\dagger = cP\rho P$$

. But this means that $\{R_j E_i P\}$ is the same operation as $\sqrt{c}P$. So by by the equivalence of Krauss operators, $R_j E_i P = u_{ji} P$. Thus (taking the adjoint) $PE_i^\dagger R_j^\dagger R_j E_k P = u_{ij}^\dagger u_{jk} P$. Then if we sum over j, we get the desired result, since $\sum_j R_j^\dagger R_j = \mathbb{I}$ since R is trace preserving.

3 Quantum Hamming Code

Assume we have a non degenerate code, so for each code word, unique errors go to a unique state (Shor code is not - phase errors on different qubits go to the same state). Encode k qubits in n qubits. Counting argument:

- $\binom{n}{j}$ places for j errors to take place
- 3^j possible combinations of j errors (X, Y, and Z in each positin)
- Each error must take each of the 2^k possible encoded states to a unique state to be non degenerate

Since the total number of states needed to deal with all errors much be less than the total available number of states, we have:

$$\sum_{j=1}^{t} \binom{n}{j} 3^j 2^k \leq 2^n$$

.

□

4 CSS Codes

4.1 Classical Linear Codes

CSS codes are based on classical codes, so we need some background. Let the code space be defined as the span of a set of k linearly independent vectors $\{v_i\}$ with length n and only 0 and 1 entries (so the coefficients in the span are only 0 and 1). If G is a matrix whose *rows* are the $\{v_i\}$, then to encode a length k word w, do

$$\tilde{w} = G^T w = wG$$

Now let H be a matrix whose *rows* are $n-k$ length n linearly independent vectors that are orthogonal to $\{v_i\}$. This is the parity check matrix. So $H(\tilde{w}+e) = He$ is the syndrome.

Notice $HG^T = 0$. However, this means $GH^T = 0$. So the row of H can be thought of as a code, (to encode length $n-k$ strings into length n strings),

Note

$$\sum_{u \in C} (-1)^{u \cdot v} = \begin{cases} |C| \text{ if } v \in C^\perp \\ 0 \text{ if } v \notin C^\perp \end{cases}$$

The top case is clear from the orthogonality of elements in C and $C^\perp$, and the bottom, we can rewrite u as wG. where w runs over all $\{0,1\}^k$. Since G is the parity check for $C^\perp$ and $v \notin C^\perp$, $Gv \neq 0$, so we have

$$\sum_{u \in C} (-1)^{u \cdot v} = \sum_{w \in \{0,1\}^k} (-1)^{wG \cdot \nu} = \sum_{w \in \{0,1\}^k} (-1)^{w \cdot v}.$$

From our prev arguments, this is 0.

Let C_1 be a code and C_2 be a subcode within C_1. Codewords are

$$|\bar{w}\rangle = \sum_{x \in C_2} |w + x\rangle$$

for $w \in C_1$. So there are as many codewords as there are cosets of C_2 in C_1.

Now let there be a phase and bit flip error:

$$|\bar{w}\rangle \to \sum_{x \in C_2} (-1)^{(w+x)\cdot e_2} |w + x + e_1\rangle$$

Attach an ancilla and reversibly apply parity check matrix for C_1:

$$\to \sum_{x \in C_2} (-1)^{(w+x)\cdot e_2} |w + x + e_1\rangle |He\rangle.$$

Measure ancilla to find syndrome, and then can correct! Next apply Hadamard to get

$$\to \sum_{y \in C_1, x \in C_2} (-1)^{(w+x)\cdot(e_2+y)} |y\rangle$$

Rewrite as $y' = e_2 + y$, and then from our above identity, summing over $x \in C^\perp$, terms only survive when $y' \in C_2^\perp$.So we have

$$\sum_{y' \in C_2^\perp} (-1)^{w\cdot y'} |y' + e_2\rangle.$$

Next apply parity check for $C_2^\perp$ to ancilla as before to correct error. Now the state is the same as the state after applying the Hadamard, except if $e_2 = 0$. So when apply Hadamard, get back to state before that, except with $e_2 = 0$, which is our original state!!

5 Stabilizer Codes

Stabilizer codes are made up of states that are "stabilized," i.e. plus one eigenstates of a subgroup S of the Pauli Group: $G_n = \{\pm I, \pm X, \pm Y, \pm Z\}^n$ where $Y = i\sigma_y$. Let $\{M_1, M_2, \ldots, M_k\}$ be the generators of S. Then the generators have the following properties:

- All M_i commute. Otherwise $M_i M_j = -M_j M_i \to |\psi\rangle = M_i M_j |\psi\rangle = -M_j M_i |\psi\rangle = -|\psi\rangle$. (Note $M_i M_j = \pm M_j M_i$ since each of the individual pauli matrices have this properties.)
- $M_i^2 \neq -\mathbb{I}$. (Note $Y^2 = -1$, which is where this constraint comes from.)

Lemma 5.1. *If you have k independent generators, then the dimension of the code space is 2^{n-k}*

Proof. One generator M has 2^n eigenstates, which all have eigenstates ± 1 (build up from individual Paulis). But there exists $N \in G^n$ that anti-commutes with M. So for every 1-eigenvector of M, N give us a -1-egenvector of M: $M|\psi\rangle = |\psi\rangle \to MN|\psi\rangle = -NM|\psi\rangle = -N|\psi\rangle$. So the eigenvectors of M come in pairs, with exactly half having eigenvector 1. So the dimension of 1-eigenvectors of M is 2^{n-1}.

Now one can find a matrix N that commutes with M from before but not M'. Consider an eigenvector that is a 1 eigenvector of M and also a 1-eigenvector of M' (might be different set than before). Then $MN|\psi\rangle = N|\psi\rangle$ so $N|\psi\rangle$ is in the set of 1-eigenvectors of M. But $M'N|\psi\rangle = -N|\psi\rangle$, so the 1-valued eigenvectors of M come in pairs, where one in the pair is a 1-eigenvector of M' and one is not. So the dimension of 1-eigenvectors of M and M' together is 2^{n-2}.

Continuing in this way, we get the stabilized space has dimension 2^{n-k}. □

Normalizer is $N(S) = \{h : h^\dagger g h \in S \forall g \in S, h \notin S\}$. Let d be the hamming weight of the largest element in $N(S)$ (hamming weight being number of pauli matrices in element h that are not the identity). Then code can correct $(d-1)/2$ errors.

Proof. Note if set of errors $\{E_i\}$ satisfies $E_j^\dagger E_i$ is not in the normalizer, then is correctable. Use $\langle\psi|E_j^\dagger E_i|\psi\rangle = \alpha_{ij}$

- If $E_j^\dagger E_i \in S$, then get $\alpha_{ij} = 1$.
- If $E_j^\dagger E_i \notin N(S)$, then $\langle\psi|E_j^\dagger E_i|\psi\rangle = \langle\psi|E_j^\dagger E_i g|\psi\rangle = -\langle\psi|gE_j^\dagger E_i|\psi\rangle = 0$

So if all errors have weight less than half of the weight of the normalizer, then any two of them together can not get to a weight equal to the normalizer, so will not be in the normalizer, and can be corrected. (Basically, things in normalizer are bad because can't be detected by syndrome, since $gh|\psi\rangle = hg_2|\psi\rangle = h|\psi\rangle$, so the syndrome will be all $+1$ (see below). □

Correct errors by acting on state with each of the generators - get ± 1 outcomes. Use sequence of ± 1 as syndrome to detect which error occured.

Can write generators in matrix form: $M = \{IXX, ZIZ, IYX\}$ can be written as

$$\left(\begin{array}{ccc|ccc} 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \end{array}\right) \tag{28}$$

Can test if generators are independent by checking if rows are independent, since adding rows is equivalent to multiplication of group elements.

5.1 Gottesman-Knill Theorem

Theorem 5.2. *Any computaion involving only $\{X, Y, Z, H, CNOT, S\}$ can be efficiently simulated using a classical computer.*

Proof. Idea is one just needs to keep track of the stabilizing generators. Initially always start in $|0\rangle^{\otimes}$ which is stabilized by $\{Z_1, \ldots, Z_n\}$. Form there, see that, e.g. acting on one bit with H is $HZH^\dagger = X$, so just changes one of the generator. Likewise, $UXU^\dagger = X_1X_2$ (where U is CNOT). Because need n generators to specify 1 state, and each generator has n Paulis in it, takes can keep track of an n bit state using polynomial resources. □

However, not universal, can prove that $TXT^\dagger$ where $T = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{pmatrix}$, does not give another product of Paulis - nor does Toffoli gate.

6 Topological Codes

7 Other Shit

Part VII
Quantum Information

1 Distance Measures

Before we can talk about information measures, we need a way to tell if quantum states are similar to each other. There are two distance measures, trace distance and fidelity.

1.1 Trace Distance

Definition 1.1. *Given states ρ and σ, the trace distance between them is*

$$D(\rho, \sigma) = \frac{1}{2} tr(|\rho - \sigma|)$$

where $|A| = \sqrt{AA^\dagger}$

Things of note about trace distance:

- Is a metric (symmetric, triangle inequality).
- Doesn't change under unitary operations on states.
- If ρ and σ commute, then they are diagonalizable in the same basis with ρ corresponding to probability distribution $\{p_i\}$ and σ to $\{s_i\}$ in that basis, then $D(\rho, \sigma) = D(p_i, s_i) = \frac{1}{2}\sum_i |p_i - s_i|$.
- Let P be any element of a POVM. Then $D(\rho, \sigma) = \max_P \mathrm{tr}(P(\rho - \sigma))$. (Proof comes from showing that $\rho - \sigma = Q - S$, where Q and S are positive operators with support on orthogonal subspaces.) This is like saying that the distance between them is outcome of the measurement that best separates them.

- Trace preserving operations can not increase trace distance:

$$D(\epsilon(\rho), \epsilon(\sigma)) \leq D(\rho, \sigma).$$

- Looking at reduced parts of the state can not increase trace distance:

$$D(\rho^A, \sigma^A) \leq D(\rho^{AB}, \sigma^{AB}),$$

1.2 Fidelity

Definition 1.2. *Given states ρ and σ, the fidelity is*

$$F(\rho, \sigma) = Tr(\sqrt{\rho^{1/2}\sigma\rho^{1/2}})$$

Things of note about Fidelity:

- Fidelity is not a distance measure (triangle property doesn't hold.
- Doesn't change under unitaries (obv)
- When one is a pure state, get

$$F(\sigma, |\psi\rangle) = \mathrm{Tr}(\sqrt{|\psi\rangle\langle\psi|\sigma|\psi\rangle\langle\psi|}) = \sqrt{\langle\psi|\sigma|\psi\rangle}\mathrm{Tr}(|\psi\rangle\langle\psi|) = \sqrt{\langle\psi|\sigma|\psi\rangle}$$

  since $(|\psi\rangle\langle\psi|)^{1/2} = |\psi\rangle\langle\psi|$,

- When ρ and σ commute then they are diagonalizable in the same basis with ρ corresponding to probability distribution $\{p_i\}$ and σ to $\{s_i\}$ in that basis. We get

$$F(\rho, \sigma) = \mathrm{Tr}(\sqrt{\sum_i p_i s_i |i\rangle\langle i|}) = \mathrm{Tr}(\sum_i \sqrt{p_i s_i}|i\rangle\langle i|) = \sum_i \sqrt{p_i s_i}$$

- Let $\{E_i\}$ be a POVM that has outcomes $\{p_m\}$ on ρ and $\{s_i\}$ on σ. Then

$$F(\rho, \sigma) = \min_{\{E_i\}} \sum_i \sqrt{p_i s_i}.$$

- To actually use Fidelity, can use Uhlmann's Theorem.

  Theorem 1.3. *Suppose ρ and σ are states of a system Q. Look at two copies of Q to get the system Q'. Let $|\psi\rangle$ and $|\phi\rangle$ be purifications of our two states in this enlarged system. Then*

$$F(\rho, \sigma) = min_{|\psi\rangle, |\phi\rangle} |\langle\psi|\phi\rangle|.$$

F and D are closely related:

$$1 - F \leq D \leq \sqrt{1 - F^2}.$$

1.3 Some Uses for Distance Measures

1. For a channel: $F_{min} = \min_{|\psi\rangle} F(|\psi\rangle, E(|\psi\rangle\langle\psi|))$.
2. For a gate U that when actually computed acts as an operation E: $F(U, E) = \min_{|\psi\rangle} F(U|\psi\rangle, E(|\psi\rangle\langle\psi|$
3. For a channel with input distribution $\{p_i, \rho_i\}$, $\bar{F} = \sum_i p_i F(\rho_i, E(\rho_i))^2$. **Note Squared!**
4. The entanglement fidelity measures how well a channel preserves entanglement. For a state ρ, purify to a state R, send the ρ part through the channel to get purified state R', and measure the fidelity:

$$F(R, R')^2 = \sum_i \langle R|E_i|R\rangle\langle R|E_i|R\rangle = \sum_i |\langle R|E_i|R\rangle|^2 \sum_i |\mathrm{Tr}(E_i\rho)$$

Since $|R\rangle = \sum_j \sqrt{p_j}|j\rangle|j\rangle$, so

$$\langle R|E_i|R\rangle = \sum_{jk} \sqrt{p_j p_k}\langle j|k\rangle\langle j|E_i|k\rangle = \sum_j p_i \rangle j|E_i|j\rangle = \mathrm{Tr}(E_i\rho)$$

The entanglement fidelity has the property that is is always smaller than the square of the normal fidelity between ρ and $E(\rho)$, since looking at a part of the subsystem can only increase fidelity. (Harder to preserve entanglement than just the state.)

The entanglement fidelity of a state ρ which describes the whole input ($\rho = \sum_i p_i\rho_i$), is less than the average fidelity $\bar{F}$ of the same system. (If preserve entanglement, preserve the state well.)

2 Entropy Measures

Von Neumann Entropy is non-negative (0 if pure), max $\log d$ in d dimensional space, two parts of pure state have same entropy,

$$S(\rho) = \mathrm{Tr}(\rho \log \rho) = \sum_i \lambda_i \log \lambda_i.$$

Quantum Relative Entropy Is always positive, and zero only for $\rho = \sigma$

$$S(\rho||\sigma) = \mathrm{Tr}(\rho \log \rho) - \mathrm{Tr}(\rho \log \sigma).$$

Joint Entropy

$$S(A, B) = \mathrm{Tr}(\rho^{A,B} \log \rho^{A,B}).$$

Conditional Entropy

$$S(A|B) = S(A, B) - S(B).$$

Mutual Information—

$$I(A : B) = S(A) + S(B) - S(A, B) = S(A) - S(A|B).$$

Some other properties of the Von Neumann Entropy:

- Max is $\log d$
- 0 if pure
- Unitaries don't change (only depends on eigenvalues)
- Given a bipartite pure state, the two parts have the same entropy.
- Concave: $S(\sum_x p_x \rho_x) \geq \sum_x p_x S(\rho_s)$.
- **Entropy of Measurement** Measurement always increases entropy (if one were to measure and then not actually look at the outcome, state would have higher entropy, unless measure in the eigenbasis of the mixed state, in which case, stays the same.)
- **Entropy of Preparation** If have a distribution of pure states to create a mixed state, there is more uncertainty in the initial distribution than in the final state (since many non-orthogonal states in the initial distribution combine to create a mixed state who's entropy depends on only d eigenvalues).
- **Subadditivity** The entropy of the whole (ρ^{AB}) is less than the sum of the parts (ρ^A, ρ^B) (think Bell state). (More uncertainty because lose the relationship between the two parts.
- **Strong Subadditivity** $S(\rho_{ABC})+S(\rho_A) \leq S(\rho_{AB})+S(\rho_{AC})$. Two overlapping systems contain more uncertainty that the whole system and the single overlapped part.
- **Triangle Inequality** $S(\rho_{AB}) \geq |S(\rho_A) - S(\rho_B)|$. For Shannon, entropy of whole exceeds either part. For von Neumann, only exceeds the difference.
- If have a state $\rho = \sum_x p_x \rho_x$ where all of the ρ's have support on orthogonal subspaces, then $S(\rho) = H(X) + \sum_x p_x S(\rho_x)$

3 Data Compression

R is the rate of compression = (# of compressed bits)/(# of original bits).

3.1 Classical Shannon Compression

If have a distribution of symbols accordion to $\{p_1, \ldots, p_k\}$, and look at a sequence of n symbols, the probability of a typical sequence is

$$p(x_1, \ldots, x_n) = p(x_1) \ldots p(x_n) \approx p_1^{np_1} \ldots p_k^{np_k},$$

since the expected number of symbol X_j is np_j, and the probability of something with probability p occurring g times is p^g. Taking the log of both sides:

$$\log p(x_1, \ldots, x_n) = n \sum_j p_j \log p_j = -nH(X).$$

So the probability of a typical sequence is $2^{-nH(X)}$. The number of typical sequences is then approximately $2^{nH(X)}$, so can encode all typical sequences using $nH(x) + \epsilon$ bits. (Another way to see this is to say the number of typical sequences is n CHOOSE $(np_1, np_2, \ldots, np_i) = \frac{n!}{\sum_j (np_j)!}$. The probability of a non- typical sequence appearing is negligible for large n, so can encode at a rate of $H(x)$ per bit.

3.2 Schumacher Compression

Suppose have a source represented by a density matrix $\rho = \sum_j p_i |i\rangle$ where $|i\rangle$ are the orthonormal decomposition. Then the probability of sending a typical state

$$p(|x_1, \ldots, x_n\rangle) = p(|x_1\rangle) \ldots p(|x_n\rangle) \approx p_1^{np_1} \ldots p_k^{np_k}.$$

So as above, there are approximately $2^{nH(X)}$ typical states. Notice that the typical states are all orthogonal, so can be represented by a $nH(X) + \epsilon$ qubits. (Note $H(X) = S(\rho)$.) First project onto space of typical states. Output error if not typical. Then create a unitary that maps the typical states onto a smaller subspace of size $nH(X) + \epsilon$ qubits, and then can reversibly recover the original state by inverting the unitary.

If $R < S(\rho)$ is used, then that means we are trying to project onto a space of dimension 2^{nR}. But as n gets large, by the theorem of typical suspaces, there is no way that most states sent (which will mostly be typical states, and thus fall into a subspace of size 2^{nS}) will end up on this subspace because it is too small. If $R > S(\rho)$, then we are projecting to a subspace of dimension 2^{nR}, and so we can include all of the typical subspace, and w.h.p all states sent will be in this subspace.

4 Channel Coding

R is the rate of encoding = (# of original bits)/(# of bits used to encode). (2^{nR} messages sent using n bits.)

4.1 Shannon Channel Coding

Setup: Have a source that outputs symbols $\{x_j\}$ with probability $\{p_j\}$. Have a channel that outputs symbols $\{y_j\}$ with known probabilities $p(y_j|x_i)$.

Idea: randomly pick 2^{nR} codewords that each consist of n symbols. Choose them to be typical (throw out if not typical). Then an output Y, assuming errors are typical, could correspond to a certain "sphere" of typical inputs. If choose few enough codewords, only one codeword in each sphere, so map to that codeword.

Details: Given an output Y, the entropy in the value of the input is $H(X|Y)$. But we saw that having an entropy of a certain value corresponds to $2^{nH(X|Y)}$ typical states. This means that there are $2^{nH(X|Y)}$ typical states that the output could correspond to (in a given sphere for an output y. Since there are $2^{nH(X)}$ total typical states, the number of possible distinct spheres is $2^{n(H(X)-H(X|Y))}$. We want there to only be one codeword per sphere, so the number of codewords must be 2^{nR}, where $R < H(X) - H(X|Y) = I(X : Y) = H(X, Y) - H(X) - H(Y)$. So any $R < I(X : Y)$ is achievable.

4.2 Holevo Bound and Quantum Channel Coding

Holevo Information

$$\xi(E) = S(\rho) - \sum_x p_x S(\rho_x).$$

The Holevo Information tells you how much the uncertainty decreases if you know how a state was prepared. Is > 0 because of concavity.

Accesible information is the amount of classical information that can be retrieved from a state. So if Alice tries to send Bob some classical states $\{x\}$ with probabilities $\{p_x\}$ by encoding them into mixed states $\{\rho_x\}$, and Bob decodes to some variables $\{y\}$ using the POVM $\{E_y\}$, then the accessible information is $I(X:Y)$.

Theorem 4.1 (Holevo Bound)**.** *Given the set-up above,*

$$I(X:Y) \leq \xi(E).$$

Proof. Use 3 systems: A is Alice's system with her classical states and probabilities, Q is her quantum encoding, and B is Bob's outcome. Initially we have a state

$$\rho^{AQB} = \sum_x p_x |x\rangle\langle x| \otimes \rho_x \otimes |0\rangle\langle 0|.$$

After Bob performs a measurement, the system becomes

$$\sum_{x,y} p_x |x\rangle\langle x| \otimes \sqrt{E_y^\dagger}\rho_x\sqrt{E_y} \otimes |y\rangle\langle y|.$$

Let primed states denote the system after the measurement. The mutual information between Alice's state and the measurement outcome must be less than between Alice's states and the original quantum encoding. (Because operations can't increase mutual information, and adding uncorrelated states and subtracting states can't increase mutual information.)

$$S(A':B') \leq S(A:Q).$$

But $S(A) = H(X)$, and $S(Q) = S(\rho)$ (taking trace), and since the sum over x is of mixed states that are mutually orthogonal, there is a theorem (see above, properties of entropy) that $H(A,Q) = H(X) + \sum_x p_x S(\rho_x) \to S(A:Q) = S(\rho) - \sum_x p_x S(\rho_x)$. For $S(A':B')$, since $\mathrm{Tr}(E_y\rho_x) = p(y|x)$, tracing out Q', we get a state of A' and B' that can be represented completely classically with $p(x,y)$ and x and y. Thus $S(A':B') = I(X:Y)$.

□

Sending classical information over a quantum channel can be thought of as trying to achieve the Holevo bound. (Can take away channel and just have Alice send Bob states for most direct connection.) For sending classical information over a quantum channel, Alice creates codewords of length n from some distribution $\{\rho_x\}$ each with distribution $\{p_x\}$. But what Bob receives are $\{\sigma_x = \varepsilon(\rho_x)\}$ with probability $\{p_x\}$. Bob is looking for codewords among the typical spaces of $\sigma^{\otimes^n}$, where $\sigma = \sum_x p_x\sigma_x$, which has size $2^{nS(\sigma)}$. But for Alice, who creates the codewords, the entropy of a codeword is $S(\sigma_{x1}\otimes\cdots\otimes\sigma_{xn})$, which on average

is $\sum_{x1,\ldots,xn} p_{x1\ldots,xn} S(\sigma_{x1} \otimes \cdots \otimes \sigma_{xn}) = \sum_x p_x S(p_x)$. This means that the states that Alice sends are only on a subspace of size $2^{n\langle S(\rho)\rangle}$. So the probability of error is

$$\frac{2^{n\langle S(\rho)\rangle}}{2^{nS(\rho)}}.$$

5 Additivity Conjectures

Above we discuss the capacity of channels where we always send unentangled states. It was conjectured that even if you send entangled states, don't get an advantage. The following conjectures were shown by Shor to be equivalent:

6 Mother of All Lemmas

7 Choi-Jamiokoslky

Part VIII
Definitions and Basic Information

8 Definitions

- **Hermitian:** $A^\dagger = A$
- **Unitary:** $UU^\dagger = \mathbb{I}$
- **Positive Operator** A such that $\langle v|A|v\rangle > 0 \rightarrow$ eigenvalues > 0 (for definite) and similarly but with $\geq$ for semi-definite.
- $S = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}$
- $H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$
- Toffoli gate first controls 3 base on 1, and then controls 3 base on 2.

9 Useful Math Theorems

- Any hermitian matrix h can be diagonalized by a unitary matrix u, i.e. $d = u^\dagger h u$ where d is diagonal.
- Any square matrix m can be diagonalized by two unitaries u, v. i.e. $d = umv$ where d is diagonal (singular value decomposition)

- Any linear operator$a$ can be broken into a unitary u and a positive opertor j, k: $a = ju = uk$. Furthermore $j = \sqrt{aa^\dagger}$ and $k = \sqrt{a^\dagger a}$.
- $\mathbb{I}/2 = (\rho + \sum_i \sigma_i \rho \sigma_i)/4$ for any ρ.
- Suppose you have two operations, defined by the operators $\{E_i\}$ and $\{F_i\}$. Then make them have the same number of operators by padding one with zeros. Then the two operations are the same iff $\exists$ a unitary matrix u such that $E_i = \sum_j u_{ij} F_j$
- If two matrices commute, they are diagonalizable in the same basis.
- $\mathrm{Tr}(A) = \sum_i \langle i|A|i\rangle$ where i is an orthonormal basis.
- $\log x \leq (1-x)/\ln 2$ for $x \leq 1$

10 Basic Quantum

- $U(t) = e^{\frac{-iHt}{\hbar}}$