Information Theory of Quantum Computing

Grant Yang, Rohan Ramkumar

May 1, 2025

Fundamentals of Quantum Computing I

1. In Quantum Mechanics, objects no longer have a single deterministic state like a classical object and are instead in a superposition.
2. For example, uncertainty in position and uncertainty in momentum are inversely proportional (Heisenberg Uncertainty).
3. So instead of using a point in phase space to describe an object, we use a **state vector** $|\psi\rangle$**,** which can be thought of as a linear combination of some basis states
4. However, for this presentation, we will be focusing on discrete states which attain discrete values, such as the states of an electron: spin up ($|\uparrow_x\rangle$) or spin down ($|\downarrow_x\rangle$).
5. Not only can we have states like $|\uparrow_x\rangle$ and $|\downarrow_x\rangle$ but also $\frac{1}{\sqrt{2}}|\uparrow_x\rangle + \frac{1}{\sqrt{2}}|\downarrow_x\rangle$, which might represent spin right ($|\uparrow_z\rangle$).

Fundamentals of Quantum Computing II

6. By the Born Rule (see later), this particle has a 0.5 probability of being measured spin up and a 0.5 probability of being measured spin down.
7. The Kronecker product $\otimes$ is used to glue two quantum states together.

$$(a_1 |\uparrow\rangle + a_2 |\downarrow\rangle) \otimes (b_1 |\uparrow\rangle + b_2 |\downarrow\rangle) =$$

$$a_1 b_1 |\uparrow\rangle \otimes |\uparrow\rangle + a_1 b_2 |\uparrow\rangle \otimes |\downarrow\rangle + a_2 b_1 |\downarrow\rangle \otimes |\uparrow\rangle + a_2 b_2 |\downarrow\rangle \otimes |\downarrow\rangle$$

This is also how we describe entangled pairs of particles: $\frac{1}{\sqrt{2}} |\uparrow\rangle \otimes |\downarrow\rangle + \frac{1}{\sqrt{2}} |\downarrow\rangle \otimes |\uparrow\rangle$.

8. For convenience, we will omit the Kronecker product symbol when writing it out, e.g.

$$|0\rangle \otimes |0\rangle \rightarrow |00\rangle .$$

But, do not forget that the Kronecker product is still there.

How Quantum is Different

1. Say you have a mixture of quantum states: 50% spin right $|\uparrow_x\rangle$, 50% spin up $|\uparrow_y\rangle$.
2. Classically, this is 1 bit of information.
3. However, you cannot make a measurement without skewing the distribution away from 50-50! The least you can disturb it is with a 45-degree axis, yielding only $H(\cos^2 \frac{\pi}{8}) \approx 0.6$ qubits.
4. Because of entanglement, we can also do funny stuff like send 2 classical bits in 1 qubit given a shared entangled bit (superdense coding).

The Born Rule and Dirac-von Neumann Axioms I

1. $|\psi\rangle = c_1 |a\rangle + c_2 |b\rangle + \ldots, \qquad c_1, c_2, \cdots \in \mathbb{C}$
2. $P(\psi \text{ is in state } a) = |\langle a|\psi\rangle|^2$
3. $\langle\psi|\psi\rangle = 1, \quad \langle a|b\rangle = 0. \qquad \langle\psi| = c_1^* \langle a| + c_2^* \langle b| + \ldots$
4. Any observable can be represented as an operator (matrix) $\hat{A}$. The expected value of $\hat{A}$ is denoted $\langle\hat{A}\rangle = \langle\psi| \hat{A} |\psi\rangle$.
    1. Let $|A_i\rangle$ be a set of eigenvectors of $\hat{A}$. Assuming that the eigenvectors are non-degenerate, we can get an **orthonormal** basis that spans the space of quantum states (spectral theorem).
    2. If we express $|\psi\rangle$ in this basis and interpret the eigenvalues a_i as the values for the observable, $\langle\psi|\hat{A}|\psi\rangle$ reduces to the usual formula for expected value!

The Born Rule and Dirac-von Neumann Axioms II

$$\hat{A} = \sum_i a_i \left|A_i\right\rangle \left\langle A_i\right|$$

$$\left|\psi\right\rangle = \sum_i c_i \left|A_i\right\rangle$$

$$\hat{A}\left|\psi\right\rangle = \sum_i a_i c_i \left|A_i\right\rangle \qquad \hat{A}\left|A_i\right\rangle = a_i \left|A_i\right\rangle$$

$$\langle\psi|\hat{A}|\psi\rangle = \sum_i a_i c_i^* c_i = \sum_i a_i P_\psi(|A_i\rangle) = \langle\hat{A}\rangle$$

Mixture States and the Density Matrix

1. The density matrix is a more generalized way to write a wavefunction that allows you to deal with mixture states.
2. $\rho = |\psi\rangle\langle\psi|$, so the probability of measuring state a is $\langle a|\rho|a\rangle$. ρ is an observable representing 'how likely is ψ?' and $\langle\hat{A}\rangle = \mathrm{tr}(\hat{A}\rho)$.

$$\mathrm{tr}(\hat{A}\rho) = \mathrm{tr}\left[\left(\sum_i a_i |A_i\rangle\langle A_i|\right)\left(\sum_{j,k} c_j c_k^* |A_j\rangle\langle A_k|\right)\right]$$

$$= \mathrm{tr}\left[\sum_{i,k} a_i c_i c_k^* |A_i\rangle\langle A_k|\right] = \sum_i a_i |c_i|^2 |A_i\rangle\langle A_i| = \langle\hat{A}\rangle.$$

3. If we want to represent a *mixture state* instead, that is a distribution of possible states, we can just do $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$. We can't use wavefunctions because those can only handle superpositions, not mixtures! Think about the expected value formulae.

Von Neumann Entropy I

1. $$S = -\mathrm{tr}(\rho \ln \rho)$$
2. A **projector** is an operator such that $\Pi^2 = \Pi$. For example, $\Pi = |x\rangle \langle x|$. If we sum the projectors for each basis vector in an orthonormal basis, we get the identity. Projectors are important because they represent the process of measuring and collapsing a possibly mixed quantum state.
3. We can show that
$$S = \min_{\{\Pi_{1,2,\ldots}\}} \left[-\sum_i \mathrm{tr}(\Pi_i \rho) \ln(\mathrm{tr}\,(\Pi_i \rho)) \right]$$
   - where the minimum is taken over all sets of projectors such that $\sum_i \Pi_i = I$ and for each projector $\mathrm{tr}(\Pi_i) \leq 1$.
4. In other words, S is the absolute minimum amount of uncertainty under the most efficient measurement we can make, which happens to be in the orthonormal eigenbasis.

Von Neumann Entropy II

Since ρ is a hermitian matrix (i.e. $\rho^\dagger = \rho$),

$$\rho = \sum_i p_i \left|\psi_i\right\rangle \left\langle\psi_i\right| = \sum_i \eta_j \left|j\right\rangle \left\langle j\right|,$$

where $\left|j\right\rangle$ are orthonormal eigenvectors with eigenvalues η_j. Furthermore, by Born's rule, the eigenbasis measurement is optimal for distinguishing states. Since the probability of measuring $\left|j\right\rangle$ is η_j, it's natural to set $S = -\sum_j \eta_j \log \eta_j$. Also, note that, for some unitary U and some real diagonal matrix D and using the matrix logarithm,

$$\begin{aligned} -\mathrm{tr}(\rho \log \rho) &= -\mathrm{tr}(UDU^\dagger U \log DU^\dagger) \\ &= -\mathrm{tr}(UD \log DU^\dagger) \\ &= -\mathrm{tr}(D \log D) \\ &= -\sum_j \eta_j \log \eta_j = S. \end{aligned}$$

Schumacher Compression

Recall Shannon's source coding theorem:

Theorem

(Source Coding Theorem) If we send n symbols drawn i.i.d. from some R.V. X with entropy $H(X)$, then we can compress our codewords so that we only need to send $nH(X)$ bits, which is lossless as $n \to \infty$.

We used Huffman coding to actually implement such a compression, which achieves this rate for large n. We have a similar theorem from Benjamin Schumacher:

Theorem

(Schumacher Compression Theorem) Given n qubits drawn from some source ρ denoted as $\rho^{\otimes n}$, with von Neumann entropy $S(\rho)$, then we can compress the source down to $nS(\rho)$ qubits, which is lossless as $n \to \infty$.

An Example I

Let's say that we want to send three qubits drawn from the distribution with $P(\psi = |\uparrow_z\rangle) = P(\psi = |\uparrow_x\rangle) = 0.5$, where

$$|\uparrow_z\rangle = |0\rangle, |\uparrow_x\rangle = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle$$

in the basis $|0\rangle, |1\rangle$. We can calculate the density matrix of ρ to be

$$\begin{aligned}\rho = \begin{bmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{bmatrix} &= \lambda_{0'}|0'\rangle\langle 0'| + \lambda_{1'}|1'\rangle\langle 1'| \\ &= \cos^2\frac{\pi}{8}\begin{bmatrix}\cos\frac{\pi}{8} \\ \sin\frac{\pi}{8}\end{bmatrix}\begin{bmatrix}\cos\frac{\pi}{8} & \sin\frac{\pi}{8}\end{bmatrix} \\ &\quad + \sin^2\frac{\pi}{8}\begin{bmatrix}\sin\frac{\pi}{8} \\ -\cos\frac{\pi}{8}\end{bmatrix}\begin{bmatrix}\sin\frac{\pi}{8} & -\cos\frac{\pi}{8}\end{bmatrix}.\end{aligned}$$

We see that $S = H_2(\cos^2\frac{\pi}{8}) \approx 0.601$, and $3S < 2$, so we can compress our 3 qubits into 2 qubits without losing too much information.

An Example II

The probability measuring $|0'\rangle$ is $\cos^2 \frac{\pi}{8} \approx 0.854$, and the probability of measuring $|1'\rangle$ is $\sin^2 \frac{\pi}{8} \approx 0.146$, so if we measure our three qubits in the $|0'\rangle, |1'\rangle$ basis, we expect for them mostly to be $|0'\rangle$. Specifically, the probability that our message is in the set spanned by $\{|0'0'0'\rangle, |1'0'0'\rangle, |0'1'0'\rangle, |0'0'1'\rangle\}$ is $\cos^6 \frac{\pi}{8} + 3\sin^2 \frac{\pi}{8} \cos^4 \frac{\pi}{8} \approx 0.942$.
This smaller set can be rotated to $\{|000\rangle, |010\rangle, |100\rangle, |110\rangle\}$, and we can measure the third qubit to collapse (project) our state onto a smaller subspace. after discarding the third qubit, we have compressed our three qubits into two. The person receiving the two qubits can append $|0\rangle$ and apply the inverse of the encoding rotation to obtain their own density matrix ρ'. We can calculate the **average fidelity** between ρ' and ρ to be around 0.923, which means that there is a 92.3% chance that ρ' and ρ would be measured identically, so this is a pretty good compression.

The General Process

1. But wait! 92.3% is not good enough. We want lossless, not lossy!
2. Turns out that as $n \to \infty$ fidelity approaches 1. In general, Schumacher compression involves projecting ρ into a "typical subspace," the subspace spanned by the most likely eigenvectors.

Typical Subspaces I

1. The law of large numbers dictates that as $n \to \infty$ and for i.i.d. X_i,

$$\begin{aligned} -\log p(X^n) = -\log \prod p(X_i) &\to H(X_1, X_2, \ldots, X_n) \\ &= nH(X), \end{aligned}$$

where X^n is a string of n i.i.d. random variables X_i. This is the **asymptotic equipartion property.**

2. A typical subspace for some arbitrary δ is the space spanned by

$$T_\delta = \left\{ |x^n\rangle : \left| -\frac{1}{n} \log p_{X^n}(x^n) - H(X) \right| < \delta \right\}$$

where $H(X) = -\sum_x p_X(x) \log p_X(x) = S$ and p_X is our eigenvalue decomposition for $\rho = \sum_x p_X(x) |x\rangle \langle x|$.

Typical Subspaces II

3. The probability that our sequence $\rho^{\otimes n}$ lies in this subspace is $\mathrm{tr}(\Pi\rho^{\otimes n})$, where Π is the projector $\sum_{x^n \in T_\delta} |x^n\rangle \langle x^n|$.
4. In the limit of infinitely long sequences, all sequences lie in the typical subspace:

$$\lim_{n\to\infty} \mathrm{tr}(\Pi\rho^{\otimes n}) = 1$$

.

5. However, the dimension of the subspace is bounded by $2^{n(S\pm\delta)}$. This is often significantly smaller than our full space, which is of size 2^n. In short, the subspace is small but the probability that $\rho^{\otimes n}$ lies in it approaches 1 asymptotically.

So What?

1. Quantum Computing can be a very very powerful tool in certain scenarios, e.g. Grover's Algorithm, Shor's Algorithm, HHL Algorithm, allowing us to solve some very specific problems faster than any classical computer could.
2. But, if we want to send qubits across distances, like from the output of these algorithms, it would probably be very expensive because we don't want to disturb our very delicate and complicated superposition state.
3. So, Schumacher compression provides a scheme that allows us to (for large n) transmit our qubits at a better rate without loss, which will save a lot of money.
4. We didn't mention this, but there are also error correction codes for qubits, which involves a "syndrome" measurement to check if the qubit is in the correct subspace. There are also theorems relating to channel theory and quantum information, but they are pretty complicated.

References

[1] John Preskill. *Quantum Information: Chapter 10, Quantum Shannon Entropy*. California Institute of Technology, 2022. URL: http://theory.caltech.edu/~preskill/ph219/chap10_6A_2022.pdf.

[2] Mark M. Wilde. *From Classical to Quantum Shannon Theory*. Cambridge University Press, July 2019. DOI: 10.1017/9781316809976.001. URL: https://arxiv.org/pdf/1106.1445.