Information Theory — Foundations and Applications

Introduction to quantum information

Lucas Chibebe Céleri

Institute of Physics
Federal University of Goiás

**2024 — Basque Center for Applied Mathematics
University of Basque Country**

The qubit

The simplest quantum system is the qubit, which is a two level quantum system, like the polarization of light, for instance. Qubits are the basic units in quantum information.

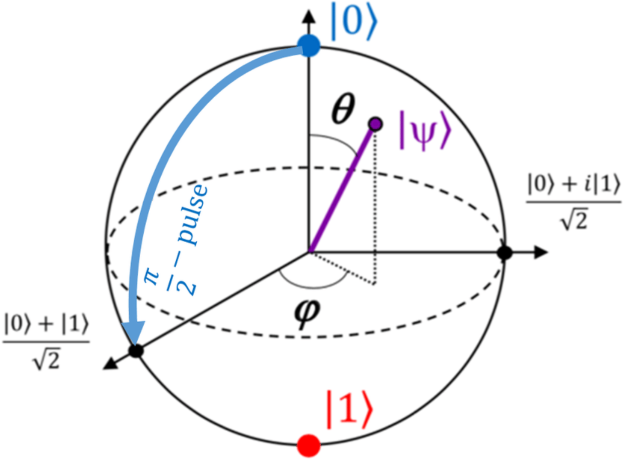
In the *computational basis* the two states of the qubit are represented by the kets

$$\{|0\rangle, |1\rangle\}$$

A fundamental difference between the classical and the quantum bit is that the last one can be in a superposition. Therefore, we cannot use Boolean algebra to describe the qubit. We need linear algebra.

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \qquad |\alpha|^2 + |\beta|^2 = 1$$

The Bloch sphere



$|0\rangle$
θ
$|\psi\rangle$
$\frac{|0\rangle + i|1\rangle}{\sqrt{2}}$
$\frac{\pi}{2}$ – pulse
φ
$\frac{|0\rangle + |1\rangle}{\sqrt{2}}$
$|1\rangle$

$$|\psi\rangle = cos\frac{\theta}{2}|0\rangle + e^{i\varphi}sin\frac{\theta}{2}|1\rangle$$

Schmidt decomposition

One of the most important resources in quantum information is entanglement. And one of the main tools to study pure state quantum entanglement is the Schmidt decomposition.

Let us consider a bipartite system AB. A pure state is said to be entangled if it cannot be written in the separable form

$$|\psi\rangle_{AB} = |\chi\rangle_{A} \otimes |\eta\rangle_{B}$$

Schmidt decomposition states that it is always possible to write

$$|\psi\rangle_{AB} = \sum_{i=1}^{d} \lambda_i \, |i\rangle_{A} \otimes |i\rangle_{B}$$

Schmidt decomposition

$$|\psi\rangle_{AB} = \sum_{i=1}^{d} \lambda_i \, |i\rangle_A \otimes |i\rangle_B$$

$\{|i\rangle_A\}$ ($\{|i\rangle_B\}$) is an orthonormal basis in $\mathcal{H}_A$ ($\mathcal{H}_B$), the Hilbert space of system A (B). The set of positive λ_i, with $\sum_i \lambda_i^2 = 1$, are called Schmidt coefficients, whose number respect

$$d \leq \text{min}\,\{\text{dim}\mathcal{H}_A, \text{dim}\mathcal{H}_B\}$$

Entanglement

$|\psi\rangle_{AB}$ is not entangled if and only if its Schmidt rank (the number of Schmidt coefficients) is 1

Distances measures

Every physical process is noisy. This means that the output state is different from expected. How can we know how well a given protocol is running? A set of measures called distances measures provides several tools for achieving this goal. Here we describe a few of them.

Trace norm

Or Schatten-1 norm of the operator O

$$||O||_1 = \text{Tr}\left[|O|\right] \qquad |O| = \sqrt{O^\dagger O}$$

Trace norm

It can be proved that the trace norm of an operator is equal to the sum of its singular values

$$\text{Tr}\,[||O||] = \sum_i \sigma_i$$

Also, it satisfies the following properties

- Non-negativity: Tr $[||O||] \geq 0$, with equality only for $O = 0$
- Homogeneity: $||cO||_1 = |c|||O||_1$
- Triangle inequality

$$||O_1 + O_2||_1 \leq ||O_1||_1 + ||O_2||_1$$

These properties shows that the trace norm is indeed a distance.

Trace norm

Some other important properties of this norm are

- Isometric invariance:

$$||O||_1 = ||UOV^\dagger||_1$$

- Convexity

$$||\lambda O_1 + (1-\lambda)O_2||_1 \leq \lambda||O_1||_1 + (1-\lambda)||O_2||_1$$

- Variational characterization

$$||O||_1 = \max_U \mathrm{Tr}\,[OU]$$

The trace distance

From the trace norm we can define a distance between two density operators

Trace distance

$$||\rho - \sigma||_1 \quad \text{which satisfies} \quad 0 \leq ||\rho - \sigma||_1 \leq 2$$

It it is maximum, there is a measurement that can distinguish ρ from σ perfectly.
It if is zero, them no measurement can distinguish the states.

There is a very interesting interpretation of the trace distance in the context of Hypothesis-testing.

Hypothesis-Testing

- Alice has to distinguish between two quantum states ρ_0 and ρ_b that are prepared with equal probabilities $p_X(0) = p_X(1) = 1/2$.
- Alice can perform a binary POVM $\Lambda = \{\Lambda_0, \Lambda_1\}$.
- Alice guesses ρ_0 (ρ_1) if the outcome of the outcome of the measurement is 0 (1). This outcome is denoted by the random variable Y.

The success probability is them defined as

$$\begin{aligned} p_{\text{succ}}(\Lambda) &= p_{Y|X}(0|0)p_X(0) + p_{Y|X}(1|1)p_X(1) = \frac{1}{2}\text{Tr}\left[\Lambda_0\rho_0\right] + \frac{1}{2}\text{Tr}\left[\Lambda_1\rho_1\right] \\ &= \frac{1}{2}\left[1 + \text{Tr}\left(\Lambda_0(\rho_0 - \rho_1)\right)\right] \end{aligned}$$

The last equality follows from the fact that $\Lambda_1 = \mathbb{1} - \Lambda_0$.

Hypothesis-Testing

The success probability depends explicitly on the quantum measurement. Alice wants to maximize it. Therefore, we have

$$p_{\rm succ} = \max_{\Lambda} \left[p_{\rm succ}(\Lambda)\right] = \frac{1}{2}\left[1 + \frac{1}{2}||\rho_0 - \rho_1||_1\right]$$

If the trace distance is zero, Alice cannot have any information regarding distinguishability and $p_{\rm succ} = 1/2$. If the trace distance is maximum, them $p_{\rm succ} = 1$ and Alice can find a measurement that let her to perfectly guess the correct state.

Fidelity

Another interesting measure of distinguishability is the quantum fidelity

Fidelity

Let us consider two quantum pure state $|\psi\rangle$ and $|\phi\rangle$. The quantum fidelity is defined as

$$F(\psi,\phi) = |\,\langle\psi|\phi\rangle\,|^2 \qquad 0 \leq F(\psi,\phi) \leq 1$$

Uhlmann Fidelity

The most general fidelity, which is a dintinguishability measure between to density operators, is the Uhlmann fidelity

$$F(\rho,\sigma) = ||\sqrt{\rho}\sqrt{\sigma}||_1^2$$

Quantum entropies

We now start our journey into the quantum Shannon theory. The first fundamental measure we introduce is the von Neumann entropy, which gives meaning to the notion of information qubit, which is a fundamental quantum information unit measure. It is defined for any quantum state ρ as

$$S(\rho) = -\mathrm{Tr}\left[\rho \log \rho\right]$$

Let us consider the spectral decomposition of the density operator

$$\rho = \sum_x p_X(x) \left|x\right\rangle\!\left\langle x\right|$$

Them

$$S(\rho) = H(X)$$

Properties of von Neumann entropy

- Non-negativity:

$$S(\rho) \geq 0$$

  $S(\rho) = 0$ for pure states

- Upper bound:

$$S(\rho) \leq \log d$$

  with $d =$ dim$\mathcal{H}$

- Concavity:

$$S(\rho) \geq \sum_x p_X(x) S(\rho_x) \qquad \rho = \sum_x p_X(x) \rho_x$$

- Since isometries do not change eigenvalues, we must have

$$S(\rho) = S(U\rho U^\dagger)$$

Joint entropy

We can extend this definition to more than one system. Let us consider the bipartite system AB.

$$S(\rho_{AB}) = -\text{Tr}\left[\rho_{AB}\log\rho_{AB}\right]$$

Let us consider $\rho_{AB} = |\phi\rangle\langle\phi|$. We can employ Schmidt decomposition to write

$$|\phi\rangle = \sum_i \sqrt{\lambda_i}\,|i_A\rangle \otimes |i_B\rangle$$

The respective marginals are

$$\rho_A = \sum_i \lambda_i\,|i_A\rangle\langle i_A| \quad \text{and} \quad \rho_A = \sum_i \lambda_i\,|i_B\rangle\langle i_B|$$

Joint entropy

Therefore, both marginals have exactly the same eigenvalues. Thus

$$S(\rho_A) = S(\rho_B)$$

while $\rho_{AB} = 0$. Remembering that, in the classical case

$$H(X, Y) \geq H(X) + H(Y)$$

This is the first radical difference between the classical and the quantum cases. We can know everything about the global system, but completely ignore the states of the individual systems.

This is what Schödinger called the fundamental characteristic of quantum mechanics.

The conditional quantum entropy

Let us write

$$S(A|B)_\rho = S(\rho_{AB}) - S(\rho_B)$$

Such definition is natural and it obeys many of the relations that the classical conditional entropy obeys., like that conditioning does not increase entropy

$$S(A)_\rho \geq S(A|B)_\rho$$

As an example, let us consider the maximum entangled state of two qubits

$$\left|\Phi^+\right\rangle = \frac{|00\rangle_{AB} + |11\rangle_{AB}}{\sqrt{2}}$$

It is not difficult to show that $\rho_A = \rho_B = \mathbb{1}/2$. Therefore

$$S(A|B)_{|\Phi^+\rangle} = -1$$

Negative conditional quantum entropy

This property is so important in quantum information that it has its own name

Coherent information
is a measure of the quantum correlations shared by subsystems A and B

$$I(A;B) = -S(A|B) = S(\rho_B) - S(\rho_{AB})$$

Coherent information is positive for entangled states. In such cases, it is not possible to assign a state vector to the subsystems.

Total correlations

How much correlations, classical and quantum, are shared by two subsystems? The answer is the quantum mutual information, which is defined analogously to the classical case

$$I(A:B)_\rho = S(A)_\rho + S(B)_\rho - S(A,B)_\rho$$

It is easy to prove that

$$I(A:B)_\rho \geq 0 \quad \forall \rho$$

For the state $|\Phi^+\rangle$ we have

$$I(A:B)_{|\Phi^+\rangle} = 2$$

We have 1 bit of entanglement (the coherent information) and 1 bit of classical correlations.

The quantum relative entropy

One of the most important quantities in quantum Shannon theory is the quantum relative entropy

$$D(\rho||\sigma) = \mathrm{Tr}\left[\rho\left(\log\rho - \log\sigma\right)\right]$$

Which respect one of the fundamental inequalities in the theory

> **Monotonicity**
> The quantum relative entropy is monotonic under quantum operation $\mathcal{N}$
>
> $$D(\rho||\sigma) \geq D(\mathcal{N}(\rho)||\mathcal{N}(\sigma))$$
>
> which implies non-negativity of the relative entropy, $D(\rho||\sigma) \geq 0$.

Finally, one interesting inequality is the quantum Pinsker inequality

$$D(\rho||\sigma) \geq \frac{1}{2\ln 2}||\rho - \sigma||_1^2$$

Data processing inequality

Processing quantum data reduces correlations. This is the statement of fundamental inequalities known as quantum data processing inequalities

Coherent information

Let $\mathcal{N}_{B\to B'}$ be a quantum channel and $\sigma_{AB'} = \mathcal{N}_{B\to B'}(\rho_{AB})$. Then

$$I(A;B)_\rho \geq I(A;B')_\sigma$$

Mutual information

Let $\mathcal{N}_{A\to A'}$ and $\mathcal{M}_{B\to B'}$ be quantum channels on subsystems A and B, respectively, and $\sigma_{A'B'} = \mathcal{N}_{A\to A'} \otimes \mathcal{M}_{B\to B'}(\rho_{AB})$. Then

$$I(A:B)_\rho \geq I(A':B')_\sigma$$

The Holevo bound

Suppose that Alice prepares the ensemble

$$\mathcal{E} = \{p_X(x), \rho_B^x\}$$

and then send it to Bob without any information regarding the value of x. Let us thing about it as a quantum channel whose input random variable is X. Since Bob does not know the value of this random variable, its expected density operator is

$$\rho_B = \sum_x p_X(x)\rho_B^x$$

Bob then has to determine x by performing some measurement on this state.

The Holevo bound

The information accessible to Bob is defined as

$$I_{\text{acc}}(\mathcal{E}) = \max_{\Lambda} I(X : Y)$$

with Y being the random variable associated withe Bob's measurement. This is a very intuitive definition that characterizes the maximum information about x Bob can obtain. Although it is highly difficult to compute it, an upper bound was determined by Holevo

$$I_{\text{acc}}(\mathcal{E}) \leq \chi(\mathcal{E}) = S(\rho_B) - \sum_x p_X(x) S(\rho_B^x)$$

χ is called the Holevo quantity and it is straightforward to compute.

Meaning of von Neumann entropy

We now can give a meaning of the von Neumann entropy with the Schumacher's compression theorem, which is the quantum data compression theorem.

The protocol is characterized by three parameters n, R, and ϵ, corresponding to the length of the original quantum data sequence, the rate, and the error, respectively.

The protocol consists of four steps: preparation, encoding, transmission and decoding.

Meaning of von Neumann entropy

- **Preparation**. The quantum information source outputs a sequence $|\psi_{x^n}\rangle_{A^n}$ of quantum states according to the ensemble $\{p_X(x), |\psi_x\rangle\}$ where

  $$|\psi_{x^n}\rangle_{A^n} = |\psi_{x_1}\rangle_{A_1} \otimes \cdots \otimes |\psi_{x_n}\rangle_{A_n}$$

  For someone ignorant of the classical sequence x^n, the density operator is $\rho^{\otimes n}$ where

  $$\rho = \sum_x p_X(x)\, |\psi_x\rangle\langle\psi_x|$$

- **Encoding**. Alice encodes the systems A^n according to some compression channel $\mathcal{E}_{A^n \to W}$, where W is a quantum system of size 2^{nR}, with R being the rate of compression

  $$R = \frac{1}{n} \log \mathrm{dim} \mathcal{H}_W$$

Meaning of von Neumann entropy

- **Transmission**. Alice transmits the system to Bob using nR quantum qubit channels.
- **Decompression**. Bob sends the system to the decompression channel $\mathcal{D}_{W\to A^n}$. The error is characterized by the trace distance

$$\frac{1}{2}||\rho^{\otimes n} - (\mathcal{E}_{A^n\to W} \circ \mathcal{D}_{W\to A^n}(\rho^{\otimes n}))||_1 \leq \epsilon$$

The compression rate R is achievable if there exists an (n, $R+\delta$, ϵ) quantum compression code for all $\delta > 0$, $\epsilon \in (0,1)$, and sufficiently large n. The quantum data compression limit of ρ is equal to the infimum of all achievable quantum compression rates.

Meaning of von Neumann entropy

In this context, the quantum data compression theorem states that

Schumacher's compression theorem

Suppose that ρ_A is the density operator corresponding to a quantum information source. Then the quantum entropy $S(A)_\rho$ is equal to the quantum data compression limit of ρ.

Such a theorem gives and operational meaning to the von Neumann entropy in the same sense that classical data compression (first Shannon coding theorem) gives a meaning to the Shannon entropy.

Thank you for your attention

lucas@qpequi.com

www.qpequi.com