Quantum Information Theory

Jun-Ting Hsieh
Stanford University
junting@stanford.edu

Bingbin Liu
Stanford University
bingbin@stanford.edu

1 Introduction

In our current digital world, everything, from computing devices to communication channels, is represented as 0s and 1s. The "information" that these bits can present has been well studied in the field of classical information theory. For example, we know the capacity and limits of data compression and the rate of reliable communication over noisy channels. Recently, perhaps due to the excitement of quantum computation, these ideas in information theory have been extended to the quantum world, leading to the field of *quantum information theory*.

In this paper, we would like to introduce the quantum counterparts of some of the key ideas we studied in the course, including the quantum entropy measure, quantum source encoding, and touch briefly on quantum channels.

1.1 How is quantum different than classical?

Compared to the classical world, quantum mechanics has several "weird" properties that make things much more convoluted. Thus, we first introduce some major differences between quantum and classical information, which are the reasons that quantum information is so interesting yet harder to analyze. Refer to Section 2 for a more detailed introduction of quantum mechanics.

Qubit. The quantum version of the classical bit is the **qubit**. A classical bit is either 0 or 1, while a qubit can be *superposition* of 0 and 1. This means that it is both 0 and 1, and if we perform a measurement we will get 0 and 1 with certain probabilities. Imagine that every bit in our computer is both 0 and 1 at the same time, and if we even try to read from it, we not only get a probabilistic answer but also collapse that qubit state!

Entanglement. Suppose we have two classical bits (stored in disks) and we separate them, they become independent. However, for a system with multiple qubits, we can perform certain operations to make them *entangled*. Entanglement means that the qubits can be related in a way that is not possible in the classical world. For example, two qubits can be made entangled such that when we measure them at arbitrary locations or in arbitrary order, we know for sure that the two outcomes will be the same, which is not possible for the classical world if we think about them as, for example, two coin flips.

No-cloning theorem. In the classical world, we can make multiple copies of the same bits. This is not true in quantum mechanics. In quantum mechanics, the no-cloning theorem states that given an unknown quantum state, it is impossible to create a copy of the state. Again, imagine a computer full of qubits, and we cannot make duplicates of the same qubits.

Distinguishability. In the classical world, we can distinguish between 0 and 1 perfectly. For example, we can easily check whether a bit is flipped or whether two strings are the same. This is also not true in quantum mechanics. Given two arbitrary (or non-orthogonal in the vector space, which we will formally describe in section 2) qubits, we cannot know for sure whether they are the same. Measuring them will not give us deterministic answers, and we cannot measure them multiple times because of the no-cloning theorem.

EE376A Project Report, Winter 2019.

1.2 Why is quantum interesting?

Quantum mechanics is itself an interesting and active field of research, and it gives rise to many interesting applications. Here, we look at a simple example of quantum communication: **superdense coding**.

Suppose Alice wants to send a random 2-bit message to Bob. Classically, the best Alice can do is to send the 2 bits. Now, suppose Alice and Bob shares an entangled pair of qubits in the state $|\psi_{00}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. For a message $x \in \{0,1\}^2$, Alice first performs certain operations on her qubit such that the state becomes $|\psi_x\rangle$, which is one of the following 2-qubit states,

$$\begin{aligned} |\psi_{00}\rangle &= \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \\ |\psi_{01}\rangle &= \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle) \\ |\psi_{10}\rangle &= \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle) \\ |\psi_{11}\rangle &= \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle) \end{aligned} \tag{1}$$

Then, Alice sends her qubit to Bob. Since these are 4 orthogonal states (known as the *Bell states*), Bob can measure the qubits and with 100% chance determine which state they are in, which determines the message x.

In this example, Alice transmits 2 bits of information to Bob by sending only 1 qubit (1 use of the quantum channel), whereas in classical communication Bob needs to send 2 bits. Though one can argue that these are not comparable, superdense coding demonstrates the potential capabilities of quantum information that classical information does not have.

2 Prerequisites

Let's have a quick review of the concepts and notations needed for this paper. Feel free to skip ahead, or refer to [1] for more advanced materials.

2.1 Quantum States

The main postulate of quantum mechanics is that any isolated physical system has an associated **complex vector space** (state space), and the system is completely described as a **unit vector** (state vector) in the state space. The standard notation is the **Dirac notation**: a state vector is denoted as $|\psi\rangle$, its complex conjugate $|\psi\rangle^\dagger$ is denoted as $\langle\psi|$, and the inner product of two states $|\phi\rangle$ and $|\psi\rangle$ is $\langle\phi|\psi\rangle$. Usually, a vector $|\psi\rangle$ in a d-dimensional space is written as

$$|\psi\rangle = \sum_{i=1}^{d} a_i |e_i\rangle \tag{2}$$

where $\{|e_i\rangle\}$ is some orthonormal basis in the state space, the a_i's are complex numbers, and $\sum_i^d |a_i|^2 = 1$.

In a closed physical system, a transformation of a state $|\psi\rangle$ is described by $|\psi'\rangle = U|\psi\rangle$, where U is a **unitary matrix** ($U^\dagger U = I$). Under any unitary transformation, $\langle\psi'|\psi'\rangle = \langle\psi|U^\dagger U|\psi\rangle = \langle\psi|\psi\rangle = 1$, which ensures that the new state is still a unit vector.

2.2 Density Operator

A physical system may have an ensemble of states. For example, suppose we know the system is in state $|\psi_i\rangle$ with probability p_i, then the system can be described by the density operator,

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i| \tag{3}$$

If a system is in a single state vector $|\psi\rangle$, we say that it is a **pure state** $\rho = |\psi\rangle\langle\psi|$. Just like a state vector, a physical system is fully described by ρ, i.e. if we know ρ we know how the system behaves. This is equivalent to representing a system as an ensemble of state vectors, but the density operator is mathematically cleaner and more convenient for quantum information.

The density operator has some important properties:

- ρ is positive semi-definite and $\text{tr}(\rho) = 1$.
- Under transformation U, the state $\rho \to \sum_i p_i U |\psi_i\rangle\langle\psi_i| U^\dagger = U\rho U^\dagger$.
- $\text{tr}(\rho^2) \leq 1$. Equality holds if ρ is a pure state.

2.3 Reduced Density Operator

Suppose we have two physical systems A and B, and we know the joint state is ρ^{AB}. However, if we only care about what happens in system A, then we can look at the **reduced density operator**,

$$\rho^A = \text{tr}_B(\rho^{AB}) \tag{4}$$

where tr_B is called the **partial trace** over system B, defined as

$$\text{tr}_B(|a_1\rangle\langle a_2| \otimes |b_1\rangle\langle b_2|) = |a_1\rangle\langle a_2| \,\text{tr}(|b_1\rangle\langle b_2|) \tag{5}$$

The partial trace is useful when we want to discard a system. For example, when a system A interacts with the environment R, we get a joint state ρ^{AR}. However, we only care about the behavior of system A, which is completely represented by the reduced density operator ρ^A.

Consider two systems A and B. If ρ^{AB} is a product state $\rho^{AB} = \rho \otimes \sigma$, then $\rho^A = \rho$, directly ignoring B. Thus, the reduced density operator is most useful when ρ^{AB} is entangled. Suppose $\rho^{AB} = |\psi_{AB}\rangle\langle\psi_{AB}|$ is a pure state where $|\psi_{AB}\rangle$ is entangled. The **Schmidt decomposition** shows that there exist orthonormal states $\{|i_A\rangle\}$ for A and $\{|i_B\rangle\}$ for B such that

$$|\psi_{AB}\rangle = \sum_i \lambda_i |i_A\rangle |i_B\rangle \tag{6}$$

In particular, the reduced density operators $\rho^A = \sum_i \lambda_i^2 |i_A\rangle\langle i_A|$ and $\rho^B = \sum_i \lambda_i^2 |i_B\rangle\langle i_B|$. For example, suppose $|\psi_{AB}\rangle$ is a fully entangled Bell state $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. Then, the sub-system is $\rho^A = \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|)$, a fully mixed state.

2.4 Measurement

Given a physical system in state $|\psi\rangle$, we can make a **measurement**. Quantum measurements are described by a collection of measurement operators $\{M_m\}$. For a state $|\psi\rangle$, the probability of measuring outcome m is $p(m) = \langle\psi|M_m^\dagger M_m|\psi\rangle$. The probabilities must sum to 1, so $\sum_m M_m^\dagger M_m = I$. The most common measurement operators are projectors onto the basis vectors, $M_m = P_m$, so that the outcome probabilities correspond to the coefficients of $|\psi\rangle$.

We can further define $E_m = M_m^\dagger M_m$. The set $\{E_m\}$ is called POVM elements (for Positive Operator-Valued Measure). Any set of positive operators $\{E_m\}$ that satisfy $\sum_m E_m = I$ is a valid POVM, and the probability of measuring m is

$$p(m) = \langle\psi|E_m|\psi\rangle \tag{7}$$

In the language of density operators, given a state $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$, the probability

$$p(m) = \sum_i p_i \langle\psi_i|M_m^\dagger M_m|\psi_i\rangle = \sum_i p_i \text{tr}(|\psi_i\rangle\langle\psi_i| M_m^\dagger M_m) = \text{tr}(M_m^\dagger M_m \rho) = \text{tr}(E_m\rho) \tag{8}$$

2.5 Quantum Operations

We know that for a closed quantum system, any transformation is described by a unitary matrix U, and the state $\rho \to U\rho U^\dagger$. However, in many cases, the system may interact with another system. For

example, a state can interact with the environment in a noisy quantum channel, or we can perform operations to compress a state. Thus, we need to first define a formalism for quantum operations.

Suppose $\mathcal{N}$ is an operation on a density matrix ρ, transforming $\rho \to \rho' = \mathcal{N}(\rho)$. For a quantum operation, ρ' should be a valid density operator that satisfies the properties in Section 2.2, so $\mathcal{N}$ must satisfy 3 properties: (i) linear, (ii) trace preserving: $\text{tr}(\mathcal{N}(\rho)) = \text{tr}(\rho) = 1$, and (iii) positive: $\mathcal{N}(\rho) \geq 0$. The Choi-Kraus theorem shows that $\mathcal{N}$ has a **Choi-Kraus decomposition** as follows:

$$\mathcal{N}(\rho) = \sum_k N_k \rho N_k^\dagger \tag{9}$$

where $\{N_k\}$ are called **operation elements** that satisfy $\sum_k N_k^\dagger N_k = I$.

This allows us to represent any quantum operations as a set of operation elements $\{N_k\}$. For example, consider a quantum channel that flips a qubit (applying the Z operator) with probability ϵ. A qubit state ρ will become $Z\rho Z^\dagger$ with probability ϵ or remain the same with probability $1-\epsilon$. We can model this operation as a set of 2 operation elements $\{N_k\} = \{\sqrt{1-\epsilon}I, \sqrt{\epsilon}Z\}$ satisfying $\sum_k N_k^\dagger N_k = I$. The new state becomes

$$\rho' = (1-\epsilon)\rho + \epsilon Z\rho Z^\dagger = \sum_k N_k \rho N_k^\dagger \tag{10}$$

2.6 Distance Measures

In classical lossy compression, we need a distortion function to measure how close two messages are. In quantum mechanics, we also need a distance measure to evaluate two quantum states. Given two density matrices ρ and σ, we use **fidelity** to determine their similarity,

$$F(\rho, \sigma) = \left(\text{tr}\sqrt{\rho^{\frac{1}{2}}\sigma\rho^{\frac{1}{2}}}\right)^2 = \|\sigma^{\frac{1}{2}}\rho^{\frac{1}{2}}\|_1^2 \tag{11}$$

where $\|A\|_1$ is the trace norm $\|A\|_1 = \text{tr}\sqrt{A^\dagger A}$. The fidelity satisfies several properties:

- Symmetry. $F(\rho, \sigma) = F(\sigma, \rho)$.
- $0 \leq F(\rho, \sigma) \leq 1$. Equals 1 if and only if $\rho = \sigma$.
- If ρ is a pure state, $\rho = |\psi\rangle\langle\psi|$, then $F(\rho, \sigma) = \langle\psi|\sigma|\psi\rangle$.
- Invariance under unitary transformation. $F(\rho, \sigma) = F(U\rho U^\dagger, U\sigma U^\dagger)$.

This definition is motivated by the fidelity of two (categorical) random variables X, Y with probabilities $(p_1, \ldots, p_n)$ and $(q_1, \ldots, q_n)$: $F(X, Y) = (\sum_i \sqrt{p_i q_i})^2$. The analogy is clear when ρ and σ commute, i.e. diagonalizable by the same basis. Let the eigen-decomposition be $\rho = \sum_i p_i |i\rangle\langle i|$ and $\sigma = \sum_i q_i |i\rangle\langle i|$.

$$F(\rho, \sigma) = \left(\text{tr}\sum_i \sqrt{p_i q_i}\,|i\rangle\langle i|\right)^2 = \left(\sum_i \sqrt{p_i q_i}\right)^2 \tag{12}$$

We are interested in how much a state changes after operation $\mathcal{N}$. Suppose $\{N_k\}$ is the set of operation elements, and suppose ρ starts with a pure state $\rho = |\psi\rangle\langle\psi|$. Then, since $\mathcal{N}(\rho) = \sum_k N_k \rho N_k^\dagger$,

$$F(\rho, \mathcal{N}(\rho)) = \langle\psi|\Big(\sum_k N_k\,|\psi\rangle\langle\psi|\,N_k^\dagger\Big)|\psi\rangle = \sum_k |\,\langle\psi|N_k|\psi\rangle\,|^2 = \sum_k |\text{tr}(\rho N_k)|^2 \tag{13}$$

3 Von Neumann Entropy

We know that for a data source giving letter x with probability $p(x)$, the information per letter is given by the entropy: $H(X) = -\sum_x p(x)\log p(x)$. Now, we will extend this to a quantum source. Suppose the source outputs a quantum state ρ_x with probability p_x. The system can be described by density operator $\rho = \sum_x p_x \rho_x$. Then, we define the **Von Neumann entropy** as

$$S(\rho) = -\text{tr}(\rho \log \rho) \tag{14}$$

Let the eigen-decomposition be $\rho = \sum_i \lambda_i |i\rangle\langle i|$ where λ_i are the eigenvalues, then

$$S(\rho) = -\sum_i \lambda_i \log \lambda_i = H(\lambda_i) \tag{15}$$

same as the definition of the Shannon entropy. In fact, if the quantum source outputs orthogonal pure states, i.e. $\rho_x = |x\rangle\langle x|$ where the vectors $|x\rangle$ are orthogonal, then $S(\rho)$ is exactly $H(p_x)$.

Moreover, we can define the quantum version of relative entropy:

$$S(\rho\|\sigma) = \text{tr}(\rho \log \rho) - \text{tr}(\rho \log \sigma) \tag{16}$$

and the conditional entropy and mutual information between two systems A, B:

$$\begin{aligned} S(A|B) &= S(A,B) - S(B) \\ I(A;B) &= S(A) + S(B) - S(A,B) \end{aligned} \tag{17}$$

where $S(A,B) = S(\rho^{AB})$ and $S(A) = S(\rho^A) = S(\text{tr}_B(\rho^{AB}))$.

Some properties of $S(\rho)$ analogous to the classical entropy:

- $S(\rho) = 0$ when ρ is a pure state, i.e. we have no uncertainty about the quantum state.
- $S(\rho) \leq \log d$ in a d-dimensional space, similar to $H(X) \leq \log |\mathcal{X}|$.
- $S(\rho^{AB}) \leq S(\rho^A) + S(\rho^B)$, with equality when $\rho^{AB} = \rho^A \otimes \rho^B$ is a product state, analogous to $H(X,Y) \leq H(X) + H(Y)$ where equality holds when X, Y are independent.
- Concavity. $S(\sum_i p_i \rho_i) \geq \sum_i p_i S(\rho_i)$.
- $S(\rho\|\sigma) \geq 0$, with equality when $\rho = \sigma$.

Some properties of $S(\rho)$ that are specific to quantum,

- $S(A) \not\leq S(A,B)$, unlike classical $H(X) \leq H(X,Y)$. This is due to entanglement. If A, B is in a pure entangled state, then $S(\rho^{AB}) = 0$ but $S(\rho^A) > 0$.
- $S(\sum_i p_i \rho_i) \leq H(p_i) + \sum_i p_i S(\rho_i)$, with equality when ρ_i have orthogonal support.
- Invariance under unitary transformation. $S(U\rho U^\dagger) = S(\rho)$.

4 Quantum Source Coding

We know that in classical source coding, an n-symbol string can be compressed to $nH(X)$ symbols, and we say that this message contains nH bits of information. Now consider a quantum source that outputs $|\psi_x\rangle$ (pure state $\rho_x = |\psi_x\rangle\langle\psi_x|$) with probability p_x. The question is, how much *quantum information* does this source contain, or how many qubits can this source be compressed to?

We can see this as a communication between Alice and Bob. Suppose Alice has a message $|\psi\rangle = |\psi_{x_1}\rangle \dots |\psi_{x_n}\rangle$ obtained from the quantum source, where each $|\psi_x\rangle$ is drawn with probability p_x. She compresses $|\psi\rangle$ to nR qubits ($R < 1$), and sends them to Bob. Bob's task is to decode them and reconstruct $|\psi'\rangle$, which should be very close to $|\psi\rangle$. We will use the average fidelity (described in Section 2.6) to measure how close the reconstruction is.

If these $|\psi_x\rangle$ are mutually orthogonal, then since orthogonal states are perfectly distinguishable, we can treat each $|\psi_x\rangle$ as an independent symbol. Alice and Bob can simply use the classical compression scheme, and thus this is equivalent to the classical case.

It is more interesting when the states are not orthogonal. We can still treat them as independent classical symbols, but this will not be the optimal scheme. We can use the fact that non-orthogonal states contain redundant information. This allows us to compress the quantum source even more.

The concept of compressing a quantum state is quite different than compressing classical data. So, we first give an example.

4.1 Example of quantum state compression

Suppose the quantum source emits state $|\psi_0\rangle = |0\rangle$ and $|\psi_1\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ with probability $1/2$ each. The density operator is

$$\rho = \frac{1}{2}|\psi_0\rangle\langle\psi_0| + \frac{1}{2}|\psi_1\rangle\langle\psi_1| = \begin{bmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{bmatrix} \tag{18}$$

The eigenvalues are $\lambda_0 = \cos^2\frac{\pi}{8}$ and $\lambda_1 = \sin^2\frac{\pi}{8}$, and orthonormal eigenvectors $|e_0\rangle, |e_1\rangle$.

Now, suppose Alice has 3 qubits $|\Phi\rangle = |\phi_1\phi_2\phi_3\rangle$ from the source, where ϕ_i can be either $|\psi_0\rangle$ or $|\psi_1\rangle$, and she wants to send only 2 qubits to Bob. $|\Phi\rangle$ is in an 8-dimensional space, whereas a 2-qubit system is a 4-dimensional space. Thus, the main idea is to project $|\Phi\rangle$ onto a 4-dimensional subspace Λ that best preserves $|\Phi\rangle$.

The procedure is as follows. Let Λ be the subspace spanned by the 4 vectors with highest probability $\{|e_0e_0e_0\rangle, |e_1e_0e_0\rangle, |e_0e_1e_0\rangle, |e_0e_0e_1\rangle\}$, and $\Lambda^\perp$ be the orthogonal subspace. Alice first makes a measurement based on projectors $\{P_\Lambda, I - P_\Lambda\}$. If the outcome is Λ, then the state is projected onto Λ. Otherwise, the state is in $\Lambda^\perp$, and Alice simply transforms it to $|e_0e_0e_0\rangle$. Now that the state is in the Λ subspace, Alice can apply a transformation U such that

$$|\Phi'\rangle = U|\Phi_\Lambda\rangle = |\psi_1'\rangle|\psi_2'\rangle|0\rangle \tag{19}$$

because Λ is 4-dimensional. She sends the first two qubits $|\psi_1'\rangle|\psi_2'\rangle$ to Bob. Bob then appends a $|0\rangle$ qubit and applies $U^\dagger$ so that $U^\dagger|\Phi'\rangle = |\Phi_\Lambda\rangle$.

Now, we can calculate the fidelity, i.e. how close $|\Phi_\Lambda\rangle$ is to the original $|\Phi\rangle$. With some calculation, we can see that the fidelity is 0.92, pretty close to 1. Thus, we have compressed the 3-qubit message to 2 qubits while achieving high fidelity. The intuition behind this is that we chose Λ to be the "high-probability" subspace, i.e. with high probability, ignoring the component in $\Lambda^\perp$ does not decrease the fidelity much. This idea will be used in the next section.

4.2 Schumacher's Quantum Noiseless Coding Theorem

Let's assume that each $\rho_x = |\psi_x\rangle\langle\psi_x|$ is a qubit in a pure state, and thus ρ^n is in a 2^n-dimensional state space. To compress the state, we would like to use some *quantum codes* that transform ρ^n to some state ρ' in a 2^{nR}-dimensional subspace, which can be represented as nR qubits. nR is the number of qubits of quantum information carried by the message.

As one would expect, the optimal compression rate R as $n \to \infty$ is $S(\rho)$, and the proof is based on the **typical subspace**, analogous to the typical sequences. The main idea is that nearly all messages $|\psi_{x_1}\dots\psi_{x_n}\rangle$ will have nearly perfect overlap with the typical subspace, just like most classical messages are typical sequences.

Given a state ρ, let $\rho = \sum_i \lambda_i |i\rangle\langle i|$ be the eigen-decomposition, and $S = S(\rho)$. Take the eigenvectors $|i\rangle$ with eigenvalues

$$2^{-n(S+\epsilon)} \le \lambda_i \le 2^{-n(S-\epsilon)} \tag{20}$$

Define the typical subspace Λ as the subspace spanned by these vectors. Similar to the typical sequences, ρ is measured to be in Λ with high probability: $\text{tr}(\rho P_\Lambda) \approx 1$. Moreover, there are $\approx 2^{nS}$ such vectors, so $\dim(\Lambda) \approx 2^{nS}$.

Using Alice's procedure from the previous section, we project the state onto Λ. Since we will measure outcome Λ with probability close to 1, the error (the probability of measuring $\Lambda^\perp$) is negligible, i.e. asymptotically we preserve the state perfectly. Then, a unitary transformation will transform $|\Phi_\Lambda\rangle$ to $|\psi_{comp}\rangle|0_{rest}\rangle$, where $|\psi_{comp}\rangle$ is the compressed state in a $\approx 2^{nS}$-dimensional subspace, represented as $\approx nS$ qubits.

With some calculation, we can show that the average fidelity $F \to 1$ as $n \to \infty$. Thus, for large n, we have compressed the state to $nS(\rho)$ qubits while preserving fidelity.

4.3 Holevo Information

From the Schumacher theorem, we know how much we can compress a quantum source ensemble of pure states. However, it is possible that the quantum source outputs a mixed state ρ_x each time. Suppose $\rho = \sum_x p_x \rho_x$, where ρ_x is mixed. Is the quantum information still $S(\rho)$?

As an example, suppose a quantum source always outputs $\rho_0 = \frac{1}{2}I$, a mixed state. We first observe that this source does not encode any information: Bob can reconstruct a message without knowing anything. However, $S(\rho) = S(\rho_0) = \frac{1}{2} > 0$. This examples shows that for a source $\{p_x, \rho_x\}$, the amount of quantum information depends on $S(\rho)$ as well as each $S(\rho_x)$.

The amount of quantum information is given by the **Holevo information**:

$$\chi(\mathcal{E}) = S(\rho) - \sum_x p_x S(\rho_x) \tag{21}$$

For the example above where the source always outputs ρ_0, $\chi(\mathcal{E}) = 0$, consistent with our intuition. Moreover, if ρ_x have mutually orthogonal support, then from Section 3 we know $\chi(\mathcal{E}) = S(\rho) - \sum_x p_x S(\rho_x) = H(X)$, which is the same as classical information. This is intuitively correct because Bob can perfectly distinguish the orthogonal ρ_x, so it is equivalent to a classical source. Finally, if each ρ_x is a pure state, then $S(\rho_x) = 0$, i.e. $\chi(\mathcal{E}) = S(\rho)$, consistent with the Schumacher theorem.

5 Quantum Channels

Next, let's move on to the communication setting and look at quantum channels. This is of particular interests

Definition 5.1 *A quantum channel is any completely positive trace-preserving (CPTP) operator that takes in some quantum state ρ and outputs another quantum state ρ'.*

You may refer to section 2.5 or Lecture 19 in [2] for a refresher on CPTP operators.

Note that according to this definition, a unitary gate $U : \rho \to U\rho U^\dagger$ is an example of a valid quantum channel. Another example that adds in a bit randomness could be a channel which applies U on ρ with probability $\frac{3}{4}$, and outputs a constant state ρ_0 with probability $\frac{1}{1}$, i.e. $\rho \to \frac{3}{4}U\rho U^\dagger + \frac{1}{4}\rho_0$. More generally, a quantum channel could contain a mix of unitary transformations: $U : \rho \to \sum_x p_x U_x \rho U_x^\dagger$.

Similar to the classical case, the capacity of a quantum channel can be quantified by the mutual information between the source X and the measurement outcome Y. However we now have the option to describe this quantity in terms of either classical information or quantum information, as we describe in the following subsections.

5.1 Classical Capacity of Quantum Channels

Some may ask why would we ever want to quantify the capacity of a quantum channel with classical information; this may happen two parties are trying to communicate some classical information source, but are constraint with a quantum channel. For example, one may want to transfer classical information through optical fiber, which inevitably requires dealing with quantum states when packing photons together.

For communication, Alice would like send messages drawn from an ensemble $\mathcal{E} = \{|\phi_x\rangle, p(x)\}$, which can be considered as a quantum alphabet where the probability of drawing $|\phi_x\rangle$ is p_x. Then Bob will try to get information about x by applying a POVM of his choice on the state he receives. The **accessible information** is then defined as

$$Acc(\mathcal{E}) = \max_{\{F_y\}} I(X, Y) = \chi(\mathcal{E})$$

which is the maximum information gain over all POVMs $\{F_y\}$.

In general, each state ϕ_x can itself be a mixed state, but it can be shown that we are better off when ϕ_x are pure, so we will assume this is the case in the following part. The state that Bob receives can be mixed though given that the channel can be noisy.

5.1.1 Noisy Quantum Channel Capacity

Now, we will look at what happens when we send qubits over a noisy quantum channel $\mathcal{N}$. Suppose Alice's ensemble is $\mathcal{E} = \{\rho_x, p_x\}$, then the ensemble that Bob receives is $\mathcal{E}' = \{\mathcal{N}(\rho_x), p_x\}$. The accessible information that Bob can obtain is $\chi(\mathcal{E}')$.

Alice and Bob would like to choose an ensemble $\mathcal{E}$ that maximizes the accessible information through the channel. The maximum value is

$$\chi(\mathcal{N}) = \max_{\mathcal{E}} \chi(\mathcal{E}') = \max_{\mathcal{E}} \max_{F} I(X;Y) \tag{22}$$

maximized over Alice's ensemble and how Bob measures it. We call this the **Holevo chi** of $\mathcal{N}$.

For n channel uses, $\chi(\mathcal{N}^{\otimes n})$ is the amount of information they can send with n qubits. Alice's ensemble is $\mathcal{E} = \{\rho(x_1, \ldots, x_n), p(x_1, \ldots, x_n)\}$. Then, the classical capacity of the quantum channel $\mathcal{N}$ is

$$C(\mathcal{N}) \leq \lim_{n\to\infty} \frac{1}{n} \chi(\mathcal{N}^{\otimes n}) \tag{23}$$

This is actually much more difficult to analyze than the classical case, because the n qubits are allowed to be entangled. In general, due to entanglement, $\chi(\mathcal{N}^{\otimes n})$ can be strictly greater than $n\chi(\mathcal{N})$. Nevertheless, we can obtain a bound on the classical capacity if we restrict Alice to send only product states, i.e. $\mathcal{E} = \{\rho(x_1) \otimes \cdots \otimes \rho(x_n), p(x_1, \ldots, x_n)\}$, which might actually be a more realistic setting. In this case, it can be shown that

$$\chi(\mathcal{N}^{\otimes n}) = \max_{\mathcal{E}} I(X^n; Y^n) \leq \max_{\mathcal{E}} \sum_i I(X_i, Y_i) = n\chi(\mathcal{N}) \tag{24}$$

This is now very similar to the classical noisy channel coding theorem. As we can clearly see, $\chi(\mathcal{N})$ is analogous to the channel capacity $C^{(I)}$ over a classical channel. $C^{(I)}$ is the maximum amount of classical information per channel use in a classical channel, and $\chi(\mathcal{N})$ is the maximum amount of classical information per channel use in a quantum channel.

5.2 Quantum Capacity of Quantum Channels

As mentioned before, we can also define the quantum capacity of a quantum channel $Q(N)$, for which a single channel use gives a lower bound, and repetitive uses of the channel can asymptotically achieve the upper bound. We will not go into the details of the derivation at this point but we would recommend to check out chapter 10.7 in [3]. The proof involves a quantity called the *coherent information*, which leads to the quantum counterpart of data-processing inequality.

The proof also relaxes the previous product-state-only constraint to allow entanglement across channel inputs. In general, entanglement-assisted quantum communication, together with quantum state transfer, are considered as the "father and mother" of many interesting corollaries on the achievable rates.

6 Conclusion

Quantum information theory is a field of active and interesting research given special properties of quantum mechanics such as superposition and entanglement. There are quantities and laws that have interpretations analogous to those in the classical setting, such as von Neumann entropy and Shannon entropy, Holevo information and classical mutual information, typical subspaces and typical sequences. Similar to the classical world, these concepts are closely related to many of the theoretical bounds of compression and communication, and idea of random coding has also been widely used in proving the attainability of these bounds. Very differently from the classical setting though, many of the fundamental limits in quantum information theory still remain to be explored, and we hope this survey can be help attract more brilliant people to join the research. Thanks for reading! :)

References

[1] M. A. Nielsen and I. Chuang. Quantum computation and quantum information, 2002.

[2] R. O'Donnell. Quantum computation and information 2015. *https://www.cs.cmu.edu/ odonnell/quantum15/*, 2015.
[3] J. Preskill. Physics 219/computer science 219 quantum computation. *http://www.theory.caltech.edu/people/preskill/ph229/*, 2018.