8.371 J

4/04/06

Capacities of Quantum Channels

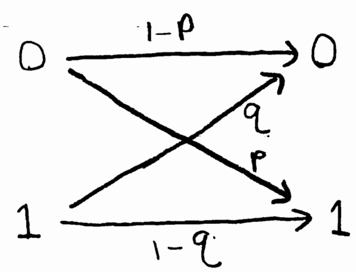
1) What is a classical channel?

- phone line, radio waves, etc.

As a simplified abstraction we'll look at discrete memoryless channels.

Alice sends 0 or 1 $\xrightarrow{\text{channel}}$ Bob recieves 0 or 1

Memoryless means each use is independent. Any discrete memoryless channel for bits can be fully described by two probabilities p and q:



If $p = q$ then this is the binary symmetric channel.

2) What is a quantum channel?

Discrete: finite-dimensional Hilbert space

Memoryless: each use is independent

= quantum operation or "superoperator"

$$\rho \to \sum_K A_K \rho A_K^+ \quad \text{where} \quad \sum_K A_K^+ A_K = I$$

Examples

1) Depolarizing channel (this is the quantum analogue to the binary symmetric channel)

$$\rho \to (1-\lambda)\rho + \lambda I/d \qquad \left(\text{for } d\text{-dimensional qudits}\right)$$

For 2-dimensions (qubits) we can use the following identity to put this into operator sum form:

$$\frac{I}{2} = \frac{\rho + \sigma_x \rho \sigma_x + \sigma_y \rho \sigma_y + \sigma_z \rho \sigma_z}{4}$$

For all ρ with trace 1.

So for qubits the depolarizing channel can be written as:

$$\rho \to (1-\gamma)\rho + \frac{\gamma}{3}\left[\sigma_x \rho \sigma_x + \sigma_y \rho \sigma_y + \sigma_z \rho \sigma_z\right]$$

(which doesn't look obviously symmetric anymore, but of course it still is)

2) Dephasing channel

$$\rho \to (1-\gamma)\rho + \gamma \sigma_z \rho \sigma_z$$

3) Amplitude damping channel

$|1\rangle$ is likely to go to $|0\rangle$ but not the other way around, e.g. photon loss in an optical fiber

8.371J
4/04/06

The operator sum representation of an amplitude damping channel is:

$$A_1 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{pmatrix} \qquad A_2 = \begin{pmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{pmatrix}$$

$$\rho \to \sum_i A_i \rho A_i^\dagger$$

For all the examples shown so far, the channels have been representable as a mixture of unitaries:

$$\rho \to \sum_i \underbrace{p_i}_{\text{probability}} \underbrace{U_i}_{\text{unitary}} \rho U_i^\dagger$$

Question: can all channels Φ with $\Phi(I/d) = I/d$ be represented as mixtures of unitaries?

For $d = 2$ the answer is yes.
For $d \geq 3$ the answer is no.

Now we'll look at a counterexample for $d=3$.

1) Project onto the x-y, x-z, or y-z planes.

This is a POVM with elements

$$\frac{1}{2}\begin{pmatrix} 1 & & \\ & 1 & \\ & & 0 \end{pmatrix} \qquad \frac{1}{2}\begin{pmatrix} 1 & & \\ & 0 & \\ & & 1 \end{pmatrix} \qquad \frac{1}{2}\begin{pmatrix} 0 & & \\ & 1 & \\ & & 1 \end{pmatrix}$$

2) Flip coordinates in planes. The 3 corresponding Kraus operators are

$$A_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad A_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad A_3 = \frac{1}{\sqrt{2}}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

This quantum channel cannot be represented as a mixture of unitaries.

Now let's look at what type of argument can be used to prove this.

If we input $|0\rangle$ to this channel we get

$$\frac{1}{2}\left(|1\rangle\langle 1| + |2\rangle\langle 2|\right)$$

as output.

If we input the state $(|0\rangle + |1\rangle)/\sqrt{2}$ then with probability 1/2 the state gets projected onto the 0-1 plane. This state is already in the 0-1 plane, so in this case the output is just $(|0\rangle + |1\rangle)/\sqrt{2}$.

With probability 1/4 the state will get projected onto the 1-2 plane, in which case the output is $|1\rangle$.

With probability 1/4 the state gets projected onto the 0-2 plane, in which case the output is $|0\rangle$.

Thus we see that $\mathcal{E}(|0\rangle)$ has zero amplitude for $|0\rangle$ and $\mathcal{E}((|0\rangle + |1\rangle)/\sqrt{2})$ has zero amplitude for $|2\rangle$.

If $\mathcal{E}$ is a mixture of unitaries then each of the unitaries must satisfy these constraints. By considering the action of $\mathcal{E}$ on a few more input states it is possible to compile enough constraints so that no unitary can satisfy all of them.

As a final example, we'll look at the erasure channel, which takes d-dimensional inputs to a d+1 dimensional space of outputs.

$$|\psi\rangle \xrightarrow{1-p} |\psi\rangle$$

$$|\psi\rangle \xrightarrow{p} |e\rangle \leftarrow \text{not in input space}$$

For $d=2$:

$$A_1 = \begin{bmatrix} \sqrt{1-p} & 0 \\ 0 & \sqrt{1-p} \\ 0 & 0 \end{bmatrix} \qquad A_2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ \sqrt{p} & 0 \end{bmatrix} \qquad A_3 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & \sqrt{p} \end{bmatrix}$$

Classical Shannon theory:

2 big theorems in Shannon's famous paper:

1) noiseless coding thm ("source coding thm")
2) noisy coding thm ("channel coding thm")

A discrete probability distribution $p_1 \ldots p_K$ has entropy

$$H(p) = -\sum_{j=1}^{K} p_j \log p_j$$

($H(p)$ is the information theory notation. The physics notation for entropy is $S(p)$.)

A source can be coded so that n symbols are sent using

$$(H(s) + \epsilon)\,n \quad \text{bits}$$

and be recovered with high probability. (The source S is modelled as producing independent identically distributed random variables from some probability distribution.)

Def The capacity of a channel is:

$$\max_A I(A:B)$$

where $I(A:B)$ is the mutual information between the input A and the output B.

Def Mutual information is defined by

$$I(A:B) \equiv H(A) + H(B) - H(A,B)$$

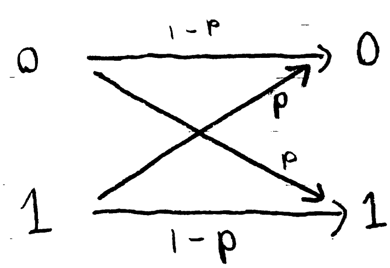
where $H(A,B)$ denotes the entropy of the joint probability distribution of A and B.

It is also true that

$$I(A:B) = H(B) - H(B|A)$$

where $B|A$ denotes the conditional probability distribution for B given A.

Example: Binary Symmetric Channel



$$H(B|A) = -p\log p - (1-p)\log(1-p)$$

regardless of what A is.

And $$H(A,B) = p_0\, p\,(0\to 1) + p_0\,(1-p)\,(0\to 0) + p_1\, p\,(1\to 0) + p_1\,(1-p)\,(1\to 1)$$

To find I we choose $p_0 = p_1 = 1/2$ since this maximizes $H(B)$, whereas $H(B|A)$ is not affected by p_0 and p_1.

$n(C-\epsilon)$ bits of information can be transmitted over a channel and recovered with high probability by using the channel n times, where C is the channel capacity.

More precisely, for any ϵ, there exists an n and a coding scheme which takes $n(C-\epsilon)$ bits and encodes them so that they can be transmitted by using the channel n times

Shannons theorems are proved using typical sequences.

Typical Sequences

Suppose you have a probability distribution

probability p_1 for symbol α_1
probability p_2 for symbol α_2
$\vdots$

A length n string is ϵ-typical if X_i, the number of occurrences of symbol α_i satisfies

$$n(p_i - \epsilon) \le X_i \le n(p_i + \epsilon)$$

for all i.

Theorem: with high probability a length n string produced by a given source is ϵ-typical.

(More precisely, the probability of a string not being ϵ-typical goes to zero exponentially as $n \to \infty$.)

The proof of this theorem is relatively simple and works by applying Stirling's formula to the multinomial distribution.

Encoding

A source S outputs a typical string with high probability.

It is only necessary to encode typical strings, since by simply throwing out non-typical strings one only fails with exponentially small probability.

The number of bits needed to encode a typical string is $\log_2$ (# of typical strings)

$$\simeq \log_2 \binom{n}{p_1 n \quad p_2 n \quad \dots \quad p_k n}$$

$$\simeq n H(p_1, p_2, \dots, p_n)$$

$$= n H(\text{source})$$

Quantum coding

Alice gets a source which outputs an unknown pure state $|V_i\rangle$ with probability p_i.

Alice's goal is to send these states to Bob using as few qubits as possible.

Alice can use quantum data compression to successfully transmit the states to Bob with high probability using a reduced number of qubits.

Criteria for success

Alice gets n symbols from the source

$$|V\rangle = |V_{i_1}\rangle \otimes |V_{i_2}\rangle \otimes \ldots \otimes |V_{i_n}\rangle$$

She wants to compress this into some number of qubits, send them through a noiseless quantum channel to Bob, Bob then decompresses and obtains ρ.

require: $$\underset{\text{expectation}}{E}\ \underbrace{\langle V|\rho|V\rangle}_{\text{fidelity}} \geq 1-\epsilon$$

To be concrete, we could imagine that there is a referee who, unlike Alice and Bob, knows $|V_i\rangle$. The referee does the best measurement allowed by quantum mechanics and can't distinguish Bob's decompressed state from $|V\rangle$ except with probability ϵ.

Def The entropy of a source is $H(\rho)$ where ρ is the density matrix defined by

$$\rho = \sum_i p_i |V_i\rangle\langle V_i|$$

and $H(\rho) = -\operatorname{tr}(\rho \log \rho)$.

Note that $H(\rho)$ is equal to the Shannon entropy of the eigenvalues of ρ. Also note that $\log(\rho)$ is uniquely defined since density matrices are always positive operators.

Now let's see how Alice performs the compression.

Alice projects onto a typical subspace. This is the quantum analogue of typical sequences and next we'll see what typical subspaces are.

Suppose the source produces states $|V_1\rangle, |V_2\rangle, \ldots, |V_K\rangle$ with probabilities $p_1, p_2, \ldots, p_K$. Then we say that the source is

$$\rho = \sum_{i=1}^{K} p_i |V_i\rangle\langle V_i|$$

Call the eigenvectors of ρ $|\tilde{V}_i\rangle$ and the eigenvalues λ_i. The source which produces states $|V_1\rangle, |V_2\rangle, \ldots, |V_K\rangle$ with probabilities $\lambda_1, \lambda_2, \ldots, \lambda_K$ has the same density matrix ρ.

These states are orthogonal so this source is essentially classical.

<u>Thm</u> Any two sources with identical density matrices behave the same in any experiment (i.e. are indistinguishable).

The typical subspace is the subspace spanned by typical sequences of $|\tilde{v}_i\rangle$'s.

Alice's compression procedure:

Perform a projective measurement to see whether the string of n states produced by the source lies in the typical subspace or not. If it does then send the resulting projection onto the typical subspace through the quantum channel. This requires $n(H(\rho)+\epsilon)$ qubits, since this is just the Shannon entropy of the λ_i's.

The probability that the state will project into the typical subspace is nearly 1.

Let's let T be the projector onto the typical subspace and let

$$|\psi\rangle = |V_{i_1}\rangle \otimes |V_{i_2}\rangle \otimes \ldots \otimes |V_{i_n}\rangle$$

be the output of the source. Then

$$\underset{\substack{\uparrow \\ \text{expectation over all outputs of source}}}{E_\psi} \langle\psi| T |\psi\rangle = 1 - \underset{\substack{\uparrow \\ \text{small}}}{\epsilon}$$

If Alice projects successfully she sends

$$\frac{T|\psi\rangle\langle\psi|T}{(1-\epsilon)} \leftarrow \text{normalization factor}$$

Recall that $\langle\psi|T|\psi\rangle = 1-\epsilon$. Thus the fidelity is:

$$\langle\psi| \left(\frac{T|\psi\rangle\langle\psi|T}{1-\epsilon} \right) |\psi\rangle = \frac{(1-\epsilon)^2}{1-\epsilon} = 1-\epsilon$$

Capacity of quantum channel

Example:

Alice is given 2 non-orthogonal quantum states with which to encode. We'll denote these as ↕ and ⤢.

One thing she could do is:

0 1 1 0 1 0 1 1 0 → ↕ ⤢ ⤢ ↕ ⤢ ↕ ⤢ ⤢ ↕

Bob measures each ↕ or ⤢, distinguishes them as well as possible, and decodes.

More specifically, suppose ↕ and ⤢ represent the following states of a qubit:

$$\updownarrow = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad ⤢ = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$$

The optimal measurement to distinguish these is a Von Neumann measurement symmetric about these two states:

project onto these perpendicular lines denoted by dotted lines.

The accessible information is:

$$I_{acc} = 1 - H_2\left(\frac{1}{2} + \frac{\sin\theta}{2}\right)$$

where $H_2(p)$ denotes $-p\log p - (1-p)\log(1-p)$.

Suppose Alice is instead given 3 quantum states

$$\updownarrow \quad \nearrow\!\!\!\!\swarrow \quad \nwarrow\!\!\!\!\searrow \qquad (60^\circ \text{ apart})$$

Suppose Alice encodes her bits just using two of these states. In this case she can only transmit 0.6454 bits per channel use.

There is another strategy Alice can use which is much better. Use two-state blocks and send either

$$|{\updownarrow}\rangle|{\updownarrow}\rangle, \quad |{\nearrow\!\!\!\!\swarrow}\rangle|{\nearrow\!\!\!\!\swarrow}\rangle, \quad \text{or} \quad |{\nwarrow\!\!\!\!\searrow}\rangle|{\nwarrow\!\!\!\!\searrow}\rangle$$

If Bob uses the Von Neumann measurement that best distinguishes these then he gets 1.369 bits per channel usage.