1 Properties of the Holevo $\chi$ quantity

The Holevo $\chi$ quantity can be considered to be a generalization of the von Neumann entropy $S(\rho)$, and reduces to $S(\rho)$ for an ensemble of pure states.

The Holevo $\chi$ quantity is non-negative: $\chi(\mathcal{E}) \geq 0$. This follows easily from the concavity of the von Neumann entropy:

$$S(\rho) = S(\sum_x p_x \rho_x) \geq \sum_x p_x S(\rho_x).$$

The Holevo $\chi$ quantity can be expressed in terms of the relative entropy as follows. For an ensemble $\mathcal{E} = \{p_x, \rho_x\}$,

$$\chi(\mathcal{E}) = \sum_x p_x D(\rho_x || \rho)$$

This can be easily verified using the definition of the relative entropy. Also, \[1\] can be used to prove that the equality condition in the concavity of von Neumann entropy holds if and only if all the states $\rho_x$ are identical.

In Example Sheet 3, you proved that the relative entropy satisfies the following monotonicity property:

**Lindblad Uhlmann monotonicity:** A quantum operation (CPTP map) can never increase the relative entropy:

$$D(\Lambda(\rho_x)||\Lambda(\rho)) \leq D(\rho_x||\rho) \quad \text{for any CPTP map } \Lambda.$$ \[2\]

From \[1\] and \[2\] it follows that a quantum operation can never increase the Holevo $\chi$ quantity: If $\mathcal{E} = \{p_x, \rho_x\}$ and $\mathcal{E}' = \{p_x, \Lambda(\rho_x)\}$ then

$$\chi(\mathcal{E}') \leq \chi(\mathcal{E}).$$

The monotonicity of $\chi$ under quantum operations indicates that $\chi$ is a good measure of the amount of information encoded in a quantum system. This is because a noise process, described by a quantum operation $\Lambda$, can never increase $\chi$. In contrast, the von Neumann
entropy $S(\rho)$ is not monotonic under quantum operations. We saw for example that (i) a *depolarizing channel* transforms a pure state into a mixed state, thereby increasing the von Neumann entropy; whereas (ii) asymptotically many uses of an *amplitude damping channel* takes a mixed state to the pure state (since the excited atom decays to its ground state), thereby reducing the von Neumann entropy.

*Note:* In the case (ii), the decrease of $S(\rho)$ should *not* be looked upon as an information gain. This is because every mixed state decays to the ground state under repeated actions of the amplitude damping channel, and hence we lose the ability to distinguish between different possible preparations of the mixed state.

### 1.1 Noisy channel

In the last lecture we obtained an upper bound to the maximum amount of classical information that could be sent to Bob via a *noiseless quantum channel* by Alice, if she encoded the symbols $x$ (labelling the classical messages emitted by a source with probabilities $p_x$) into quantum states $\rho_x$, respectively. This upper bound was given by the Holevo $\chi$ quantity

$$\chi(p_x, \rho_x) = S(\sum_x p_x \rho_x) - \sum_x p_x S(\rho_x).$$

Now let us consider what would happen if the channel between Alice and Bob was *noisy*. In particular, let us consider the channel to be memoryless and noisy, and let us denote it by the CPTP map $\Lambda : \mathcal{D}(\mathcal{H}_A) \rightarrow \mathcal{D}(\mathcal{H}_B)$. A quantum channel is *memoryless* if there is no correlation in the noise acting on successive inputs to the channel. Hence if $\Lambda^{(n)} : \mathcal{D}(\mathcal{H}_A^\otimes n) \rightarrow \mathcal{D}(\mathcal{H}_B^\otimes n)$ denotes $n$ uses of the channel, then $\Lambda^{(n)} = \Lambda^\otimes n$ if the channel is memoryless.

In this case, if Alice encodes $x$ into the quantum state $\rho_x$ as before, Bob receives the state $\Lambda(\rho_x)$. Hence, the maximum amount of classical information that Bob can receive in this case (for a single use of the channel) is bounded above by the corresponding Holevo $\chi$ quantity:

$$\chi(\{p_x, \Lambda(\rho_x)\}) = S(\sum_x p_x \Lambda(\rho_x)) - \sum_x p_x S(\Lambda(\rho_x)).$$

This observation leads to the following question:

(Q) What is the *classical capacity* of a memoryless, noisy quantum channel $\Lambda$?

The *classical capacity* is defined as the maximum amount of classical information (in bits) that can be transmitted reliably, per use of the quantum channel. In other words, it is the maximum rate of reliable transmission of classical information through the quantum channel. This rate is evaluated in the limit of asymptotically uses of the channel (i.e., the limit $n \rightarrow \infty$).

Before answering this question, note that unlike a classical channel, a quantum channel has various different capacities (see following section).
1.2 Capacities of a quantum channel

Shannon’s Noisy Channel Coding Theorem gives an explicit expression for the capacity of a memoryless classical communications channel $\mathcal{N}$:

$$C(\mathcal{N}) = \max I(X : Y)$$

where $X$ and $Y$ are the random variables characterizing the input and output of the channel $\mathcal{N}$, and the maximum is over all possible input distributions.

A quantum channel, in contrast, has various distinct capacities. This is because there is a lot of flexibility in the use of a quantum channel. The particular definition of the capacity, which is applicable, depends on the following:

- whether the information transmitted is classical or quantum;
- whether the sender (Alice) is allowed to use inputs entangled over various uses of the channel (e.g. she can have an entangled state of two qubits and send the two qubits on two successive uses of the channel) or whether she is only allowed to use product inputs.
- whether the receiver (Bob) is allowed to make collective measurements over outputs of multiple uses of the channel or whether he is only allowed to measure the output of each channel use separately;
- whether the sender and the receiver have an additional resource, like shared entanglement, e.g. Alice and Bob could, to start with, each have a qubit of an EPR pair.

1.3 Classical capacity of a quantum channel

Let us consider the transmission of classical information through a quantum channel. We shall see later that any arbitrary quantum channel, $\Lambda$, can be used to transmit classical information, provided $\Lambda(\rho)$ is not identical for all input states $\rho$.

Consider the following scenario: Suppose Alice wants to send a message $M$, from a finite set of classical messages labelled by elements of the set $\mathcal{M} = \{1, 2, \ldots, |\mathcal{M}|\}$, to Bob, using a quantum system (e.g. an optical fibre) as a communications channel $\Lambda$. The messages are considered to be ... She is allowed to use the channel many times. She encodes her message $M \in \mathcal{M}$ in a quantum state, say, a state $\rho_M^{(n)} \in \mathcal{D}(\mathcal{H}^{\otimes n})$ (which is e.g. a state of $n$ qubits, if $\Lambda$ is a qubit channel). The encoding map is given by

$$\mathcal{E}^{(n)} : M \mapsto \rho_M^{(n)}.$$ 

She then sends this state to Bob through $n$ uses of the channel $\Lambda$. [Figure given in lecture]. Bob receives the output $\sigma_M^{(n)} = \Lambda^{\otimes n}(\rho_M^{(n)})$ of the channel and does a suitable measurement on it, in order to infer what the original message was. Let his measurement be described by the
POVM \( \{E_M^{(n)}\}_{M \in \mathcal{M}} \), which constitutes the decoding map which we denote by \( D^{(n)} \). Here \( E_M^{(n)} \) is the POVM element corresponding to the message \( M \). The probability of inferring the message correctly is given by

\[
\text{tr}(\sigma_M E_M^{(n)}).
\]

Hence, the probability of error corresponding to the message \( M \) is given by \( 1 - \text{tr}(\sigma_M E_M^{(n)}) \).

The maximum probability of error is therefore given by

\[
p_{\text{max}}^{(n)} = \max_{M \in \mathcal{M}} \left( 1 - \text{tr}(\sigma_M E_M^{(n)}) \right).
\]

(3)

The rate of information transmission for this encoding-decoding scheme is the number of bits of classical message that is transmitted per use of the channel. It is given by

\[
R := \frac{\log |\mathcal{M}|}{n}, \quad \text{i.e., } |\mathcal{M}| \approx 2^{Rn}.
\]

(4)

The triple \( (\mathcal{E}^{(n)}, D^{(n)}, R) \) defines a quantum error correcting code \( C^{(n)} \) of rate \( R \).

The transmission of classical information through the channel \( \Lambda \) is said to be reliable if there exists a sequence of codes \( C^{(n)} \) such that \( p_{\text{max}}^{(n)} \to 0 \) as \( n \to \infty \).

A real number \( R \geq 0 \) is said to be an achievable rate if there exists a sequence of codes of rate \( R \) for which information transmission is reliable. The capacity of the channel is defined as the maximum (more precisely, supremum) of all achievable rates.

We want to find the classical capacity of the channel that is its capacity for reliable transmission of classical information through it. Let us start with the special case in which Alice is allowed to encode her messages using only product states, i.e., states of the form \( \rho_M^{(n)} = \rho_1 \otimes \rho_2 \otimes \ldots \otimes \rho_n \), where \( \rho_1, \rho_2, \ldots \), are inputs to the channel on separate uses. Bob is allowed to decode the output of the channel by performing collective measurements over multiple uses of the channel. The capacity of the channel in this case is referred to as the product-state classical capacity and is usually denoted as \( C^{(1)}(\Lambda) \). The product state classical capacity is given by the celebrated Holevo-Schumacher-Westmoreland (HSW) Theorem \[2, 1\]:

**Theorem 1 (HSW):** The product state classical capacity of a memoryless quantum channel \( \Lambda \) is given by

\[
C^{(1)}(\Lambda) = \chi^*(\Lambda),
\]

where \( \chi^*(\Lambda) \) is the Holevo capacity (or Holevo information of the channel and is defined as follows:

\[
\chi^*(\Lambda) := \max_{\{p_x, \rho_x\}} \chi(\{p_x, \Lambda(\rho_x)\}),
\]

(5)

where

\[
\chi(\{p_x, \Lambda(\rho_x)\}) := S(\Lambda(\sum_x p_x \rho_x)) - \sum_x p_x S(\Lambda(\rho_x)).
\]

(6)

The maximum in (5) is taken over all ensembles \( \{p_x, \rho_x\} \) of possible input states \( \rho_x \) of the channel, with \( p_x \geq 0, \sum_x p_x = 1 \).
Information transmission is improved when the outputs of the noisy channel are more distinguishable. The quantity $\chi(\{p_x, \Lambda(\rho_x)\})$ is a measure of the distinguishability of the output states $\Lambda(\rho_x)$. It might seem natural that the distinguishability of the output states would be maximised by maximising the distinguishability of the input states, i.e., by using a set of mutually orthogonal input states. However, this intuition was proved to be false (see e.g. [3]). Consequently, the Holevo capacity can indeed be achieved on non-orthogonal input states.

The HSW Theorem tells us that Alice can reliably transmit classical information to Bob, through a memoryless quantum channel $\Lambda$ using product state inputs, at any rate below $\chi^*(\Lambda)$. In contrast, any rate $R > \chi^*(\Lambda)$ is not achievable.

In fact, the Holevo capacity has been shown to provide a sharp threshold for the rate of transmission of classical information through a memoryless quantum channel via product state inputs in the following sense. For any $R < \chi^*(\Lambda)$ there exists an encoding-decoding scheme of rate $R$ such that transmission of classical information via product state inputs through the channel is reliable ($p^{(n)}_{\text{max}} \to 0$ as $n \to \infty$), whereas for any encoding-decoding scheme of rate $R > \chi^*(\Lambda)$, information transmission is totally unreliable ($p^{(n)}_{\text{max}} \to 1$ as $n \to \infty$).

To any ensemble of states $\{p_x, \rho_x\}$ one can associate a classical-quantum state (or c-q state)

$$\rho_{XA} = \sum_x p_x |x\rangle \langle x| \otimes \rho_x.$$  \hfill (7)

Here $\rho_x \in D(\mathcal{H}_A)$ and $|x\rangle \in \mathcal{H}_X$. Consider the quantum channel $\Lambda : D(\mathcal{H}_A) \to D(\mathcal{H}_B)$. Then the Holevo capacity can equivalently be expressed as:

$$\chi^*(\Lambda) = \max_{\rho_{XA}} I(X : B)|\tilde{\rho}|,$$  \hfill (8)

where $\rho_{XA}$ is the state defined by (7) and

$$\tilde{\rho}_{XB} = \sum_x p_x |x\rangle \langle x| \otimes \Lambda(\rho_x).$$  \hfill (9)

It can be easily verified that

$$\chi(\{p_x, \rho_x\}) = I(X : A)_{\rho}.$$  \hfill (10)

Since, $\tilde{\rho}_{XB} = (\text{id}_X \otimes \Lambda)\rho_{XA}$, we have that

$$\chi(\{p_x, \Lambda(\rho_x)\}) = I(X : B)_{\tilde{\rho}}.$$  \hfill (11)

**Lemma 1** The maximization on the RHS of (5) (and hence on the RHS of (8)) can be restricted to pure state ensembles.

\footnote{The maximum in (6) is potentially over an unbounded set. However, it was shown in [4] that one can restrict the maximisation to pure state ensembles, containing at most $d^2$ elements, where $d$ is the dimension of the input Hilbert space of the channel.}
Proof\footnote{This closely follows the proof in the book by Mark Wilde.} This amounts to proving that

\[
\chi^*(\Lambda) = \max_{\rho_{XA}} I(X : B)_{\tilde{\rho}},
\]

where

\[
\omega_{XA} = \sum_x p_x |x\rangle\langle x| \otimes \rho_{x,y} |\phi_{x,y}\rangle\langle \phi_{x,y}|.
\]

Let the spectral decomposition of \(\rho_x\) be given by

\[
\rho_x = \sum_y \mu_{x,y} |\phi_{x,y}\rangle\langle \phi_{x,y}|.
\]

Obviously, \(\sum_y \mu_{x,y} = 1\). Define the state

\[
\sigma_{XYA} = \sum_{x,y} p(x,y) |x\rangle\langle y| \otimes |\phi_{x,y}\rangle\langle \phi_{x,y}|,
\]

Note that \(\sigma_{XA} = \text{tr}_Y \sigma_{XYA} = \rho_{XA}\). Also note that \(\sigma_{XYA}\) is of the same form as \(\omega_{XA}\). This can be seen by considering \(XY\) as the classical system. It can be written as

\[
\sigma_{XYA} = \sum_{x,y} p(x,y) |x\rangle\langle y| \otimes |\phi_{x,y}\rangle\langle \phi_{x,y}|,
\]

where \(p(x,y) := p_x \mu_{x,y}\) and \(\sum_{x,y} p(x,y) = 1\).

After the action of the quantum channel \(\Lambda\) we obtain the following:

\[
\rho_{XA} \rightarrow \tilde{\rho}_{XB} = \tilde{\sigma}_{XB} = (\text{id}_X \otimes \Lambda)(\rho_{XA})
\]

and

\[
\sigma_{XYA} \rightarrow \tilde{\sigma}_{XYB} = \sum_{x,y} |xy\rangle\langle xy| \otimes \Lambda(|\phi_{x,y}\rangle\langle \phi_{x,y}|),
\]

Hence,

\[
I(X : B)_{\tilde{\rho}} = I(X : B)_{\tilde{\sigma}} \leq I(XY : B)_{\tilde{\sigma}},
\]

where the last inequality follows from the monotonicity of the relative entropy under partial trace. Hence it follows that it suffices to restrict the maximization on the RHS of (8) to pure state ensembles.

**Exercise:** Prove that the Holevo capacity is superadditive,

\[
\chi^*(\Lambda_1 \otimes \Lambda_2) \geq \chi^*(\Lambda_1) + \chi^*(\Lambda_2),
\]

where \(\Lambda_1, \Lambda_2\) denote quantum channels.
Exercise: Use the HSW theorem to find the product state capacity of the qubit depolarizing channel, \( \Lambda \), defined by
\[
\Lambda(\rho) = p\rho + (1-p)\frac{I}{2}.
\]

An interesting consequence of the HSW theorem is the following lemma.

**Lemma 2** Any arbitrary quantum channel \( \Lambda \) can be used to transmit classical information, provided the channel is not simply a constant, i.e., \( \Lambda(\rho) \) is not identical for all \( \rho \).

**Proof**: This can be seen as follows. If \( \Lambda \) is not a constant, then there exist pure states \( |\psi\rangle \) and \( |\phi\rangle \) such that
\[
\Lambda(|\psi\rangle\langle\psi|) \neq \Lambda(|\phi\rangle\langle\phi|).
\] (18)
Consider the ensemble
\[
\{p_1 = p_2 = 1/2, \rho_1 = |\psi\rangle\langle\psi|, \rho_2 = |\phi\rangle\langle\phi|\}.
\]
From (18) and the concavity of the von Neumann entropy it follows that \( \chi(\{p_i, \Lambda(\rho_i)\}) > 0 \). Hence, \( \chi^*(\Lambda) > 0 \), which in turn implies that the quantum channel \( \Lambda \) can transmit classical information if the latter is encoded into product states which are then sent through the channel.

**References**


