Chapter 7

Accessible information

In this chapter we prove our first important result in quantum information theory, namely Holevo's bound. This is a general bound on the *information on the preparation of a mixed state, that can be extracted from the mixed state by a measurement process.* We will see in a later chapter that it has an important applications in channel coding theory.

7.1 Notion of accessible information

We argued in chapter 2 that non-orthogonal quantum states are not perfectly distinguishable. The Holevo bound quantifies this statement. Suppose a system with Hilbert space $\mathcal{H}$ is prepared in a mixed state $\{p_x, \rho_x\}$ where ρ_x are density matrices (hence the preparation of the system is a mixture of mixed states). The total density matrix of the system is

$$\rho = \sum_x p_x \rho_x$$

We imagine that Alice has prepared the mixture $\{p_x, \rho_x\}$ but gives only ρ to Bob who wants to extract information about the preparation by performing measurements on ρ. Let us formalize the problem.

- The preparation of Alice is described by a classical random variable X taking value x with $\text{Prob}(X = x) = p_x$.

  For example Alice flip a coin: if Face is obtained with $p_F = \frac{1}{2}$ she prepares a photon in state $\rho_F = |0\rangle\langle 0|$, while if Tail is obtained with $p_T = \frac{1}{2}$ she prepares a photon in state $\rho_T = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)(\frac{1}{\sqrt{2}}(\langle 0| + \langle 1|)$.

- Bob is given $\rho = \sum_x p_x\rho_x$ but does not know the preparation $\{p_x, \rho_x\}$. He has full access to ρ in the sense that he can manipulate and measure the state.

  In the example

$$\begin{aligned}\rho &= \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)(\frac{1}{\sqrt{2}}(\langle 0| + \langle 1|)\\ &= \frac{3}{4}|0\rangle\langle 0| + \frac{1}{4}|0\rangle\langle 1| + \frac{1}{4}|1\rangle\langle 0| + \frac{1}{4}|1\rangle\langle 1|\\ &= \begin{bmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{bmatrix}\end{aligned}$$

- Bob makes measurements with an apparatus corresponding to a measurement basis $\{P_y\}$ where $P_y^2 = P_y$ are projectors and $\sum_y P_y = 1$. The outcome of the measurement is a random variable Y such that

$$\text{Prob}(Y = y|X = x) = Tr\rho_x P_y \qquad (= p_{y|x})$$

  This follows by the measurement postulate. We can now define a joint probability distribution

$$\text{Prob}(X = x, Y = y) = p_x p_{y|x} = p_x Tr\rho_x P_y \qquad (= p_{x,y})$$

  and the marginal

$$\text{Prob}(Y = y) = \sum_x p_x Tr\rho_x P_y = Tr\rho P_y \qquad (= p_y)$$

  Note that the last equation also follows directly from the measurement postulate applied to ρ.

  In the example, suppose that Bob uses the canonical basis $\{|0\rangle\langle 0|; |1\rangle\langle 1|\}$. For the conditional distribution one obtains

$$p_{y|x} = \begin{bmatrix} 0 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{bmatrix}, \tag{7.1}$$

  for the joint distribution

$$p_{x,y} = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{4} \end{bmatrix}, \tag{7.2}$$

  and for the marginal

$$p_y = \begin{bmatrix} \frac{3}{4} \\ \frac{1}{4} \end{bmatrix}. \tag{7.3}$$

- The mutual information $I(X;Y)$ defined from $p_{x,y}$ is the information about X that Bob can extract from ρ by his measurements outcomes Y. We define the *accessible information* as the maximum possible mutual information obtained by the best possible measurement

  $$\mathrm{Acc}(\{p_x,\rho_x\}) = \sup_{\{P_y\}} I(X;Y)$$

  In the example we have $H(X) = \ln 2$, $H(Y) = \ln 4 - \frac{3}{4}\ln 3$ and $H(X,Y) = \frac{3}{2}\ln 2$. Thus for the particular measurement in the canonical basis $I(X;Y) = \frac{3}{2}\ln 2 - \frac{3}{4}\ln 3 = 0.215$. This equals $0.31 \ln 2$ so Bob retrieves 0.31 bits from this type of measurement. He can do better by choosing a more clever basis but, since the states ρ_F and ρ_T are not perfectly distinguishable, his accessible information will always be strictly smaller than 1 bit (the entropy of X). An interesting question partly answered in the next paragraph is : how much smaller is it ?

  Note that if ρ_x are pure orthogonal states they form a subset of a basis of the Hilbert space. Thus by choosing *this basis* as a measurement basis Bob gets $Y = X$ so that $I(X;Y) = H(X)$. This means that a mixture of orthogonal states behaves as a classical probability distribution and can be perfectly known by suitable measurements.

7.2 The Holevo bound

In general it is very difficult to compute the supremum over all possible measurement basis, involved in the definition of the accessible information. Holevo (following pioneering works of Gordon and levitin) gave a bound which gives us an estimate that is independent measurement basis. In general this bound is loose and is not achievable by a measurement basis. The achievability holds for special mixtures, as briefly discussed in the next paragraph, and plays an important role in channel coding theorems.

Theorem [Holevo bound]. Let X be a classical random variable $\{p_x = \mathrm{Prob}(X = x)\}$ and $\{p_x, \rho_x\}$ a mixture of mixed quantum states. Let Y be the random variable describing outcomes of measurements on the state $\rho = \sum_x p_x \rho_x$ in the basis $\{P_y\}$ (these can be measuerements of any observable that has the spectral decomposition $A = \sum_y a_y P_y$). Then

$$I(X;Y) \leq \chi(\{p_x,\rho_x\}), \qquad \text{so also} \qquad \mathrm{Acc}(\{p_x,\rho_x\}) \leq \chi(\{p_x,\rho_x\})$$

where

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

In the example ρ_F and ρ_T are pure so their individual entropies vanish, and $\chi(\{p_x, \rho_x\}) = S(\rho)$. The eigenvalues of ρ are $\rho_\pm = \frac{1}{2} \pm \frac{\sqrt{2}}{4}$. So the von Neuman entropy is

$$S(\rho) = -(\frac{1}{2} + \frac{\sqrt{2}}{4}) \ln(\frac{1}{2} + \frac{\sqrt{2}}{4}) - (\frac{1}{2} - \frac{\sqrt{2}}{4}) \ln(\frac{1}{2} - \frac{\sqrt{2}}{4}) = 0.41 = 0.59 \ln 2$$

We can conclude that there are no measurements that would retrieve more than 0.59 bits of information from X. *Exercise*: compute $Acc(\{p_x, \rho_x\})$.

Proof of the Holevo Bound. Bob is given the mixed state $\rho_Q = \sum_x p_x \rho_x$ which we view as a state belonging to $\mathcal{H}_Q$. We introduce a larger Hilbert space $\mathcal{H}_X \otimes \mathcal{H}_Q \otimes \mathcal{H}_Y$ and a state

$$\rho_{XQY} = \sum_x p_x |x\rangle\langle x| \otimes \rho_x \otimes |0\rangle\langle 0|$$

The interpretation of this state is as follows: $|x\rangle\langle x|$ are mutual orthogonal states describing Alice's preparation (or r.v X) and $|0\rangle\langle 0|$ is a blank state where Bob will record his measurement outcomes. Note that $dim\mathcal{H}_X =$ number of values of x, $dim\mathcal{H}_Q$ is the dimension of the Hilbert space in which Bob's state lives (e.g 2 if this is a single Qbit) and $dim\mathcal{H}_Y = dim\mathcal{H}_Q$ since $\mathcal{H}_Y$ records the measurement outcomes. For the measurement basis of Bob we take $\{P_y = |y\rangle\langle y|\}$.

We introduce the unitary operation

$$U_{XQY} = Id \otimes U_{QY}$$

where

$$U_{QY}|\phi\rangle_Q \otimes |a\rangle_Y = \sum_y P_y|\phi\rangle_Q \otimes |a \oplus y\rangle_Y$$

Here $a \oplus y$ is computed modulo $dim\mathcal{H}_Y$. Let us check that this is a unitary operation. We have

$$\begin{aligned}
\langle\psi| \otimes \langle b|U_{QY}^\dagger U_{QY}|\phi\rangle \otimes |a\rangle &= \sum_{y,y'} \langle\psi|P_{y'} \otimes \langle b \oplus y'|P_y\phi\rangle \otimes |a \oplus y\rangle \\
&= \sum_{y,y'} \langle\psi P_{y'} P_y|\phi\rangle\langle b \oplus y'|a \oplus y\rangle \\
&= \sum_{y,y'} \delta_{y,y'}\langle\psi P_y|\phi\rangle\langle b \oplus y|a \oplus y\rangle \\
&= \langle\psi|\phi\rangle\langle b|a\rangle
\end{aligned}$$

Thus $Id \otimes U_{QY}$ preserves the inner product and is unitary.

Now we define

$$\rho'_{XQY} = U_{XQY}\rho_{XQY}U^\dagger_{XQY} = \sum_{x,y,y'} p_x|x\rangle\langle x| \otimes P_y\rho_x P_{y'} \otimes |y\rangle\langle y'|$$

The two density matrices ρ_{XQY} and ρ'_{XQY} have the same eigenvalues (since they are unitarily related) therefore their von Neumann entropies are the same

$$S(\rho_{XQY}) = S(\rho'_{XQY})$$

The two partial density matrices

$$\rho_{QY} = Tr_X\rho_{XQY} = \sum_x p_x\rho_x \otimes |0\rangle\langle 0| = \rho \otimes |0\rangle\langle 0|$$

and

$$\rho'_{QY} = Tr_X\rho'_{XQY} = \sum_x p_x P_y\rho_x P_y \otimes |y\rangle\langle y'|$$

are also unitarily related because of the tensor product form of $U_{XQY} = Id \otimes U_{QY}$. Thus we also have

$$S(\rho_{QY}) = S(\rho'_{QY})$$

From the strong subadditivity

$$S(\rho'_{XQY}) - S(\rho'_{QY}) \leq S(\rho'_{XY}) - S(\rho'_Y),$$

thus we get

$$S(\rho_{XQY}) - S(\rho_{QY}) \leq S(\rho'_{XY}) - S(\rho'_Y).$$

The rest of the proof is a computation of all the entropies appearing in this last inequality. For the first one we have

$$S(\rho_{XQY}) = S\big(\sum_x p_x|x\rangle\langle x| \otimes \rho_x\big)$$

To compute this entropy we use the spectral decomposition $\rho_x = \sum_{a_x}\lambda_{a_x}|a_x\rangle\langle a_x|$. Then

$$\sum_x p_x|x\rangle\langle x| \otimes \rho_x = \sum_{x,a_x} p_x\lambda_{a_x}|x\rangle\langle x| \otimes |a_x\rangle\langle a_x|$$

Since this is a convex combination of mutualy orthogonal states we have that its entropy is

$$-\sum_{x,a_x} p_x\lambda_{a_x}\ln p_x\lambda_{a_x} = H(X) - \sum_x p_x\sum_{a_x}\lambda_{a_x}\ln\lambda_{a_x} \tag{7.4}$$

$$= H(X) + \sum_x p_x S(\rho_x)$$

Thus

$$S(\rho_{XQY}) = H(X) + \sum_x p_x S(\rho_x)$$

For the second entropy since $\rho_{QY} = \rho \otimes |0\rangle\langle 0|$ we simply have

$$S(\rho_{QY}) = S(\rho).$$

For the third one, we first compute the reduced density matrix

$$\rho'_{XY} = Tr\rho'_{XQY} = \sum_{x,y,y'} p_x |x\rangle\langle x| \otimes |y\rangle\langle y'| Tr_Q P_y \rho_x P_{y'}$$

By the cyclicity of the trace

$$Tr_Q P_y \rho_x P_{y'} = Tr_Q P_{y'} P_y \rho_x = \delta_{yy'} Tr_Q P_y \rho_x = \delta_{yy'} p_{y|x}$$

Thus we find

$$\rho'_{XY} = \sum_{x,y} p_{x,y} |x\rangle\langle x| \otimes |y\rangle\langle y|$$

The states $|x\rangle\langle x| \otimes |y\rangle\langle y|$ are mutualy orthogonal (so they can be perfectly distinguished, fo example by performing measurements in a basis containing them). Thus this density matrix is just another representation for the random variable (X, Y). The von Neumann entropy is

$$S(\rho'_{XY}) = H(X, Y)$$

Now it remains to compute the last entropy $S(\rho'_Y)$. We have

$$\rho'_Y = Tr\rho'_{XY} = \sum_{x,y} p_{x,y} |y\rangle\langle y| = \sum_y p_y |y\rangle\langle y|$$

therefore

$$S(\rho'_Y) = H(Y)$$

Collecting all these entropies and replacing them in the strong subadditivity inequality we obtain

$$H(X) + \sum_x p_x S(\rho_x) - S(\rho) \le H(X, Y) - H(Y)$$

which is the same as

$$I(X; Y) \le S(\rho) - \sum_x p_x S(\rho_x) = \chi(X; \rho)$$

This ends the proof of Holevo's bound.

7.3 Remarks on the achievability of Holevo's bound

Given a measurement basis $\{P_y\}$ we have

$$\begin{aligned} I(X;Y) =& H(Y) - H(Y|X) \\ =& -\sum_y (TrP_y\rho)\ln(TrP_y\rho) + \sum_x p_x \sum_y (TrP_y\rho_x)\ln(TrP_y\rho_x) \end{aligned}$$

The Holevo bound states that for any $\{P_y\}$ this expression is less than

$$S(\rho) - \sum_x p_x S(\rho_x)$$

In general, given a mixture $\{p_x, \rho_x\}$ it is difficult to assess if there exists a measurement basis $\{P_y\}$ such that the bound is achieved. A positive answer can be given in special important cases.

For a mixture of pure states $\{p_x, |\phi_x\rangle\langle\phi_x|\}$ and a measurement basis $P_y = |y\rangle\langle y|$ we have

$$\begin{aligned} I(X;Y) = & -\sum_y \big(\sum_x p_x |\langle y|x\rangle|^2\big)\ln\big(\sum_x p_x |\langle y|x\rangle|^2\big) \\ & + \sum_x p_x \sum_y |\langle y|x\rangle|^2 \ln |\langle y|x\rangle|^2 \end{aligned}$$

If we have a mixture of orthonormal states and we choose a measurement basis containing all these states we find

$$I(X;Y) = H(X)$$

But since $S(\rho) \leq H(X) + \sum_x p_x S(\rho_x)$ (a general bound proved in chapter 6) we always have

$$\chi(p_x;\rho_x) \leq H(X)$$

Therefore we see that the equality is achieved for mixtures of orthonormal states, by a measurement basis containing these states. This result is an expression of the fact that orthonormal states can be perffectly distinguished: we gain the maximum possible amount of mutual information by doing the right measurements.

These arguments can be generalized to the case of a mixture such that the density matrices ρ_x are mutually orthonormal in the sense that

$$Tr\rho_x\rho_{x'} = 0, \qquad x \neq x'$$

This means that for no zero eigenvalues the eigenprojectors of ρ_x and $\rho_{x'}$ are mutually orthogonal. If we set $\rho_x = \sum_j \lambda_j P_{j,x}$ for the spectral decomposition, we have (for non zero λ_j's)

$$Tr P_{j,x} P_{j',x'} = \delta_{j,j'} \delta_{x,x'}$$

This can be checked by replacing the spectral decompositions of the density matrices in the trace and noting that all terms in the sum are non-negative. We leave it as an exercise for the reader to check that if the measurement basis $\{P_y\}$ contains $\{P_{j,x}\}$ one gets

$$I(X;Y) = S(\rho) - \sum_x p_x S(\rho_x)$$

Summarizing, when the density matrices are mutually orthogonal, the Holevo bound can again be attained by an appropriate measurement basis.

This last fact has an important application in channel coding as we will see in chapter 8. There one shows that in a high dimensional Hilbert space $\mathcal{H}^{\otimes}$, $n \to \infty$ it is possible to show the existence of mixed tensor product states that are mutualy orthogonal,

$$Tr(\rho_{x_1} \otimes \cdots \otimes \rho_{x_n})(\rho_{x'_1} \otimes \cdots \otimes \rho_{x'_n}) = 0$$

The proof uses the probababilistic method, in the spirit of Shannon's theory.