5.4 Accesible information and Holevo's bound

A source emits state ρ_x with probability p_x. We wish to extract as much information about which state ρ_x was emitted by performing POVM measurements $\{E_y\}$ on the source output. That is, we wish to maximize the mutual information $I(X;Y)$ where X is the index of the emitted state and Y is the result of the measurement performed. The accessible information is defined as

$$I_{acc} = \sup_{\{E_y\}} I(X;Y)$$

where the mutual information is computed using the joint distribution $p(x,y) = p_x \mathrm{tr}(E_y \rho_x)$. The problem we consider in this section is to determine I_{acc} and the POVM $\{E_y\}$ that achieves it. The exact solution of this problem is not known. We will present an upper bound known as Holevo's bound on I_{acc}.

Let us begin with an example. Suppose the source emits the states $|v_1\rangle = |0\rangle$ and $|v_2\rangle = \cos(\theta)|0\rangle + \sin(\theta)|1\rangle$ with probability 1/2 each. It turns out the optimal measurements for this case are projective measurements with projectors onto the directions

$$|w_1\rangle = \cos(\pi/4+\theta/2)|0\rangle + \sin(\pi/4+\theta/2)|1\rangle, \quad |w_2\rangle = \cos(-\pi/4+\theta/2)|0\rangle + \sin(-\pi/4+\theta/2)|1\rangle.$$

More precisely, the measurement operators are $E_1 = |w_1\rangle\langle w_1|$ and $E_2 = |w_2\rangle\langle w_2|$. Since $\langle w_1|w_2\rangle = 0$, $E_1 + E_2 = I$, and $\{E_1, E_2\}$ forms a POVM. Let X denote the index of the source output and Y the index of the measurement result. The conditional probability that $Y = y$, given that $X = x$, for $x, y \in \{1,2\}$, is calculated as

$$p(y|x) = \mathrm{tr}(E_y |v_x\rangle\langle v_x|) = |\langle w_y|v_x\rangle|^2 = \begin{cases} \frac{1+\sin(\theta)}{2} & \text{if } x = y, \\ \frac{1-\sin(\theta)}{2} & \text{if } x \neq y. \end{cases}$$

Thus, it is as if X is the input, Y the output of a binary symmetric channel (BSC) with error probability $\epsilon = (1-\sin(\theta))/2$. Since X is uniformly distributed on the input of the BSC, the output Y is also uniformly distributed, so $H(Y) = 1$. The mutual information is calculated easily as

$$I(X;Y) = H(Y) - H(Y|X) = 1 - H(\frac{1-\sin(\theta)}{2}),$$

where $H(x) = -x\log_2(x) - (1-x)\log_2(1-x)$. The density matrix of the source output is given by

$$\rho = \frac{1}{2}|v_1\rangle\langle v_1| + \frac{1}{2}|v_2\rangle\langle v_2| = \frac{1}{2}\begin{pmatrix} 1+\cos^2(\theta) & \sin(\theta)\cos(\theta) \\ \sin(\theta)\cos(\theta) & 1-\cos^2(\theta) \end{pmatrix}$$

which has eigenvalues $\frac{1}{2} \pm \cos(\theta)$, so the von Neumann entropy is

$$S(\rho) = H(\frac{1-\cos(\theta)}{2}).$$

The accessible information I_{acc} and $S(\rho)$ are sketched in Fig. 5.1.

We observe that $I_{acc} \le S(\rho)$. This is true in general. In fact, a stronger upper bound holds, as given next.

Theorem 1 *(Holevo, 1973) Consider a source emitting state ρ_x with probability p_x. Then, the accessible information for the source is bounded by*

$$I_{acc} \le \chi \triangleq S(\sum_x p_x \rho_x) - \sum_x p_x S(\rho_x).$$

Equality holds iff all the ρ_x commute.

For a proof see Preskill 5.4 or Nielsen and Chuang 12.1.

5.4.1 Peres-Wootters Method

Let's consider a second example. Suppose Alice wishes to send Bob information over a quantum channel. The channel is such that Alice can send one of three designated states

$$|\psi_1\rangle = (1,0)^T, \;\; |\psi_2\rangle = (-1/2, \sqrt{3}/2)^T, \;\; |\psi_3\rangle = (-1/2, -\sqrt{3}/2)^T$$

in each use of the channel. Suppose Alice sends each state with probability 1/3. Notice that the states are separated by $2\pi/3$ radians, with

$$\langle\psi_i|\psi_j\rangle = -1/2, \;\; i \ne j.$$

From Bob's point of view, the density matrix of the state prepared by Alice is given by

$$\rho = \frac{1}{3}\left(|\psi_1\rangle\langle\psi_1| + |\psi_2\rangle\langle\psi_2| + |\psi_3\rangle\langle\psi_3|\right) = \frac{1}{2}I;$$

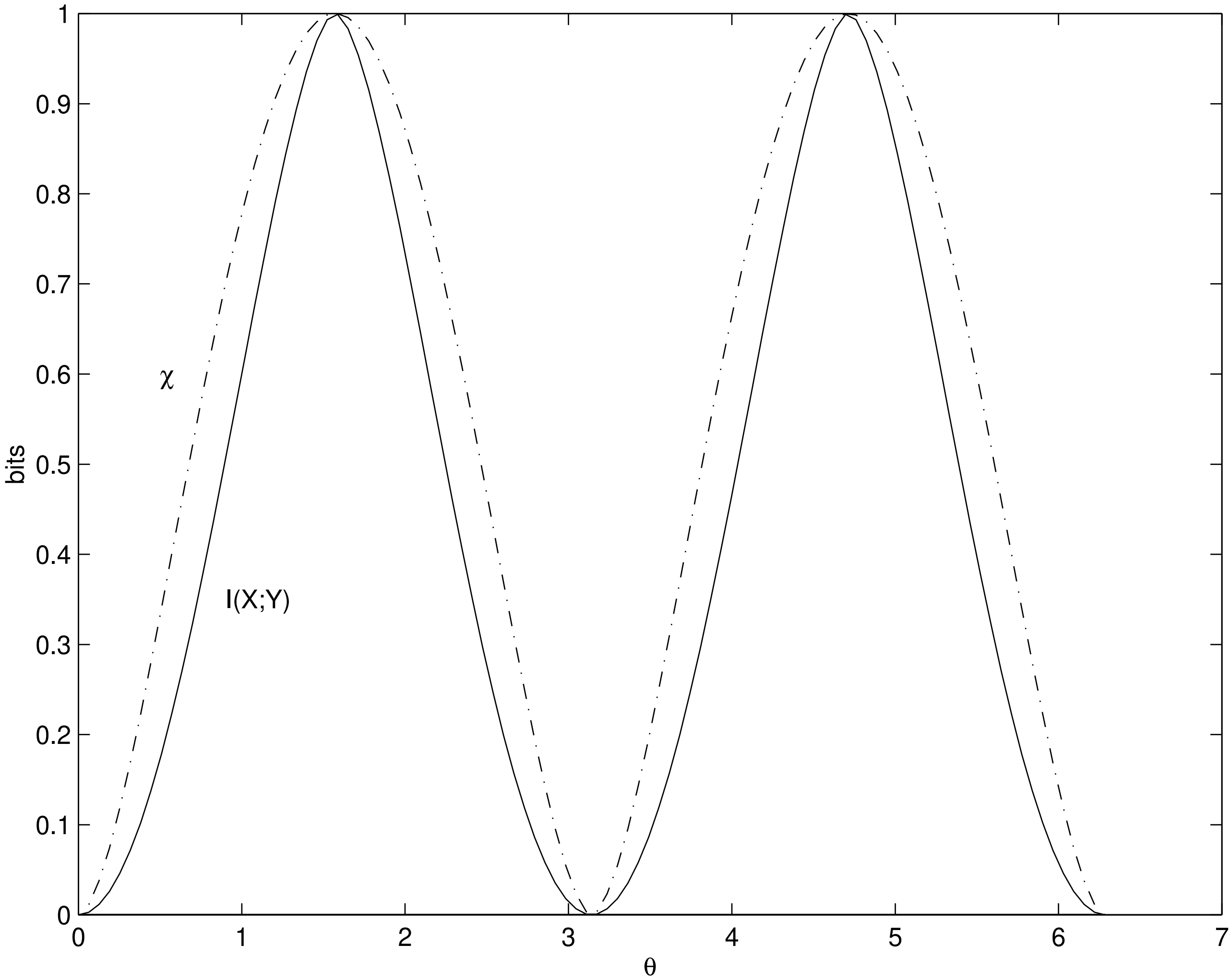

Figure 5.1: Accessible information and Holevo bound.

so, $S(\rho) = 1$. The Holevo function χ equals 1 as well, since the channel input states are pure states.

The Holevo bound tells us that the accessible information I_{acc} is bounded by 1 bit. In fact, the Holevo bound is pretty loose here. The optimal measurement turns out to be the POVM with operators

$$E_y = \frac{2}{3}\left(I - |\psi_y\rangle\langle\psi_y|\right), \quad y = 1, 2, 3.$$

The transition probabilities are now given as

$$p(y|x) = \mathrm{tr}(E_y|\psi_x\rangle\langle\psi_x|) = \langle\psi_x|E_y|\psi_x\rangle = \begin{cases} 0 & x = y, \\ \frac{1}{2} & x \neq y. \end{cases}$$

The mutual information obtained by this measurement equals

$$I(X;Y) = H(Y) - H(Y|X) = \log(3) - 1 = .58496\text{bits}.$$

It is known that this measurement is optimal, so $I_{acc} = .58496$ bits, which is considerably less than the Holevo bound $\chi = 1$ bit.

In an effort to transmit more information, Alice and Bob come up with a block coding strategy. Instead of using the channel on a one-shot basis, with each use independent of each other, they agree on a code consisting of three codewords,

$$|\Psi_x\rangle = |\psi_x\rangle|\psi_x\rangle, \quad x = 1, 2, 3.$$

It takes two uses of the channel to transmit one of three messages now. Assuming that the messages are equally likely, the channel input density is now given by

$$\rho = \sum_x \frac{1}{3}|\Psi_x\rangle\langle\Psi_x|,$$

which has nonzero eigenvalues are $1/2, 1/4, 1/4$, and entropy $S(\rho) = 3/2$ bits. On a per channel-use basis the entropy of channel input density is 0.75 bits, which is larger than the entropy achievable by using the channel on a one-shot basis. This promises that the accessible information may be improved by using coding. The underlying reason for this is that by using coding we choose a smaller subset of channel inputs that are more nearly orthogonal to each other; e.g., here,

$$\langle\Psi_y|\Psi_x\rangle = \frac{1}{4}, \quad x \neq y,$$

instead of the inner-product value of $-1/2$ for the one-shot case.

Peres and Wootters suggested using the following measurement operators for this type of coded communication over a quantum channel

$$E_y = G^{-1/2}|\Psi_y\rangle\langle\Psi_y|G^{-1/2}$$

where

$$G = \sum_x |\Psi_x\rangle\langle\Psi_x|.$$

To verify that this is a POVM, note that

$$\sum_y E_y = G^{-1/2}|\sum_y \Psi_y\rangle\langle\Psi_y|G^{-1/2} = G^{-1/2}GG^{-1/2} = I$$

on the span of $\{|\Psi_x\rangle\}$.[1] This measurement is variously called *square-root measurement* and *pretty good measurement* (PGM). It is not optimal but close to being so to yield tight results asymptotically as the block length of the codes go to infinity. One special case in which PGM is optimal is when the codewords are orthogonal; then, PGM reduces to the optimal projective mesaurement.

The PGM for the code here corresponds to a projective measurement along the three axes obtained by pulling the three states $\Psi_x\rangle$ apart until they are orthogonal. The channel transition probabilities are obtained as follows

$$p(y|x) = \begin{cases} .971405 & y = x, \\ .0142977 & y \neq x. \end{cases}$$

This results in an information transfer of $I(X;Y) = 1.36907$ bits. On a per-channel use, we transfer .684535 bits, which is larger than the information rate for the one-shot case considered earlier.[2]

[1] In case this span is not the full space, it is necessary to augment the set of E_y with one more operator that equals the identity in the orthogonal complement of the span.

[2] We should note that for the one-shot case, the maximum information transfer rate is actually achieved not by using the channel input states equiprobably, but by using only two of the states with probability 1/2 each. That gives a rate of .6454, which is still less than the rate achieved by the length-2 code here.

5.4.2 Classical capacity of a noiseless quantum channel

Since length-2 codes perform better than length-1 codes, it is natural that one should consider codes with arbitrary block lengths to maximize the information transfer rate. The solution to this problem is as follows.

Theorem 2 *(Holevo, Schumacher-Westmoreland) The maximum accessible information by using codewords composed of signal states ρ_x is given by*

$$\chi_{max} = \max_{p_x} \chi(p_x, \rho_x) = \max_{p_x} \left[S(\sum_x p_x \rho_x) - \sum_x p_x S(\rho_x) \right].$$

This result shows that the Holevo bound is saturated by using block codes. The proof of this theorem can be found in Preskill 5.4. The proof employs random coding, PGM, and several inequalities about von Neumann entropy, including strong subadditivity. It is noteworthy that PGM is asymptotically optimal.

To summarize, the above theorem gives us the best attainable information transfer rate by using a noiseless quantum channel that, in each use, can transfer one of a designated set of states ρ_x from its input to its output. We are permitted use block codes of any length that we may choose, but the codewords are constrained to product-form states,

$$\rho = \rho_{x_1} \otimes \cdots \otimes \rho_{x_n},$$

where n is the block-length of the code and ρ_{x_i} is the state transmitted in the ith use of the channel. In other words, entanglement over successive channel input states is not permitted in this version of the problem.