Lecture 8

In this class, we review Shannon's paper on "Channels with Side Information at the Transmitter". The case considered here is different from last lecture (on Gelfand and Pinsker's paper on "Coding for Channel with Random Parameters") in that here only causal side information not non-causal side information is known to the transmitter.

A channel with side information is shown in Fig. 1. The encode has two inputs, the input message m and the side information s which can be used in the encoding.

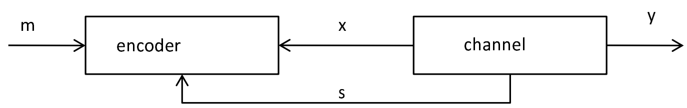

Figure 1: Channel with side information.

Problem setup:

Given a discrete memoryless channel (DMC) K with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$ and the set of channel states $\mathcal{S}$ (which is a finite set). The channel K is given by the set of conditional probabilities $p(y|x, s)$ and the set of probability $p(s)$. Suppose the channel states (side information) are known to the transmitter but not the receiver. The knowledge about s can be used in encoding in a causal way, that is, $x_i = x_i(m, s_1^i)$, where m is the information to be transmitted taking values in $\{1, 2, ...M\}$. Also we know that the channel state is independently identically distributed in different transmission instances. In short, we have:

$$\begin{aligned} x_i &= x_i(m, s_1^i) \\ p(s^n) &= \prod_{i=1}^{n} p(s_i) \\ p(y^n|x^n, s^n) &= \prod_{i=1}^{n} p(y_i|x_i, s_i) \end{aligned}$$

Find the capacity of the channel described above.

Now consider another DMC K', whose input alphabet is $\mathcal{X}^h$ where h is the cardinality of the channel state sets $\mathcal{S}$, that is, $\underline{x} = \{x_1, x_2, ...x_h\}, x_i \in \mathcal{X}$ and the output alphabet is $\mathcal{Y}$. The transition probability of channel K' is computed as:

$$p(y|x) = \sum_{s=1}^{h} p(s) \cdot p(y|[\underline{x}]_s, s) \tag{1}$$

The input $\underline{x}$ of K' can be viewed as a mapping from channel state s to the input alphabet of K, which is shown below.

$\underline{x}$:	S	$\rightarrow$	$\mathcal{X}$
	s	$\rightarrow$	$[\underline{x}]_s$
	1		x_1
	2		x_2
	...		...
	h		x_h

For example,if the channel state set $S = \{1, 2\}$ and the input alphabet for K is $X = \{0, 1\}$, then $h = 2$, and the input alphabet for K' is $\underline{x} = \mathcal{X}^2 = \{00, 01, 10, 11\}$.

Theorem 1. *The capacity of a memoryless discrete channel K with side state information, defined by $p(s)$ and $p(y|x, s)$, is equal to the capacity of the memoryless channel K' (without side information) with the same output alphabet and an input alphabet with a^h input letters $\underline{x} = (x_1, x_2, ...x_h)$ where each $x_i = 1, 2, ...a$. The transition probabilities for the channel K' are given by (1). Any code and decoding system for K' can be translated into an equivalent code and decoding system for K with the same probability of error. Any code for K has an equivocation of message (conditional entropy per letter of the message given the received sequence) at least $R - C$, where C is the capacity of K'. Any code with rate $R > C$ has a probability of error bound away from zero.*

The DMC K and K' share the same output alphabet set but K' have a larger size of input alphabet set. This theorem reduces the analysis of a given DMC channel K with side information to that for a DMC channel K' without side information.

From the description of channel K', we know that each letter $\underline{x} = (x_1, x_2, ...x_h)$ of a codeword for K' may be interpreted as a function from state s to input alphabet $\mathcal{X}$. The codes for K' are really just another way of describing certain of the codes for K. Fig. 2 gives a graphical illustration of the translation. It is clear that the statistical situations for K and K' with the translated code are identical. The probability of an input word for K' being received as a particular output word is the same as that for the corresponding operation with K. As a result, any rate achievable by K' is also achievable by K: $C(K) \geq C(K')$.

Now we prove the converse, that is, the capacity of K cannot exceed the capacity of K'. We give the notations that we use for the proof.

m :	message	m
s :	current state	s_i
S :	previous states	s_i^{i-1}
y :	current output	y_i
Y :	previous outputs	y_1^{i-1}
x :	current input	x
X :	previous inputs	x_1^{i-1}

Next we prove an equivalent statement, that is,

$$h(m|Y) - h(m|y, Y) \leq C(K')$$

which means that the uncertainty about m cannot be reduced by more than $C(K')$ upon receiving y.

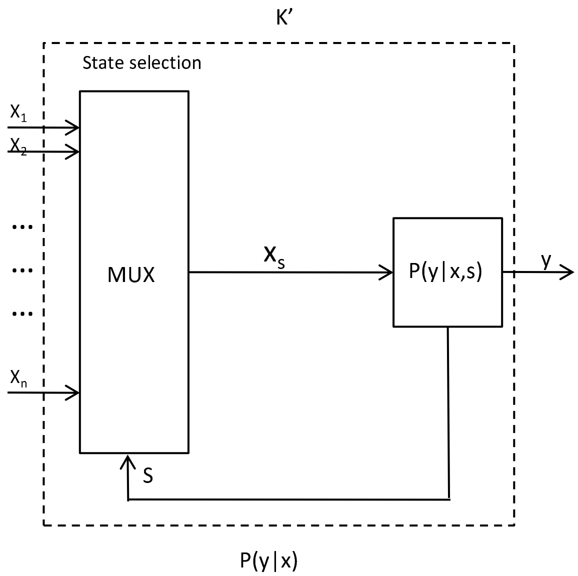

Figure 2: Graphical illustration of the translation between K and K'.

The proof goes as follows:

$$\begin{aligned} h(m|Y) - h(m|Y,y) &= I(m;y|Y) &(2)\\ &= h(y|Y) - h(y|m,Y) &(3)\\ &\leq h(y) - h(y|m,S,Y) &(4)\\ &= h(y) - h(y|m,S) &(5)\\ &= h(y) - h(y|\underline{x}) &(6)\\ &\leq C(K') &(7)\end{aligned}$$

Here from (3) to (4), we use the fact that conditioning doesn't increasing entropy. From (4) to (5), we use the Markov model shown in Fig. 3. And from (5) to (6), we follow the equivalence between (m,S) and $\underline{x}$.

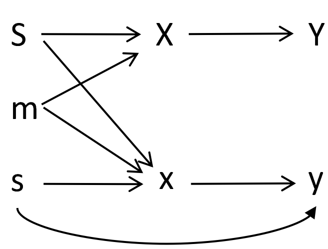

Figure 3: Markov model for m, s, x, y, S, X, Y.