10-704: Information Processing and Learning	Spring 2012

Lecture 17: Feedback, Joint Source Channel Coding, Continuous Channel

Lecturer: Aarti Singh *Scribes: Zeyu Jin*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

17.1 DMC without Feedback

In the last lecture we know that the Channel Coding Theorem for DMC without feedback says that there exists a sequence of $(2^{nR}, n)$ codes with maximum probability of error $\lambda^{(n)} \to 0$ if

$$R < C \tag{17.1}$$

If we consider codes that are not necessarily uniquely decodable, then we can encode above the information capacity of the channel. It is important to note however that the probability of error for non-uniquely decodable codes when transmitting strings of source symbols can never approach 0. Consider the following example, we have a binary symmetric channel with cross over problem as shown in Figure 17.1 with corresponding channel code defined in Table 17.1.

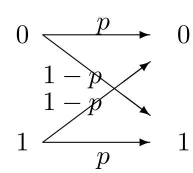

Figure 17.1: Binary Symmetric channel with crossover problem

Message	Encoding	Probability of Message
1	0	1/4
2	1	1/4
3	00	1/4
4	01	1/4

Table 17.1: Example of $R > C$ encoding, which is not uniquely decodable

The rate of the code is then:

$$R = \frac{\log_2 M}{\mathbb{E}[l]} = \frac{\log_2 4}{\frac{1}{4}(1+1+2+2)} = \frac{4}{3} \text{ bits per transmission} \tag{17.2}$$

Since $C = H(Bernoulli(p))$ we have that $R > C$.

17.2 Discrete Memoryless Feedback Channel

Consider the model shown in figure 17.3 of a channel with feedback. Instead of encoding $X_i(W)$, the encoder can use knowledge of previous channel outputs Y^{i-1} to encode $X_i\left(W, Y^{i-1}\right)$. Such a code will be called a feedback code. We will show that building feedback into memoryless channels will not increase the operational capacity of the channel; though it may help simplify the code design.

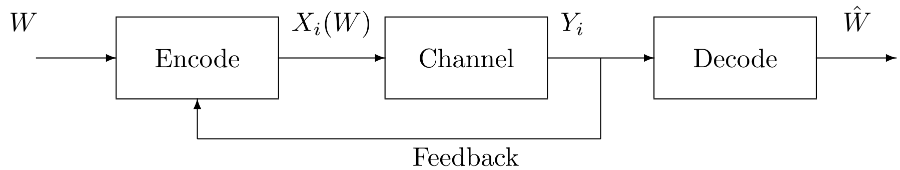

Figure 17.2: Memoryless Feedback block code model

We define C_{FB} to be the sup of all rates that is achievable using feedback codes. We have the following theorem for Feedback capacity.

Theorem 17.1 *For a discrete memoryless channel with feedback as shown in Figure 17.2. The capacity with feedback is*

$$C_{FB} = C \equiv \max_{p(x)} I(X;Y) \tag{17.3}$$

Proof: It is obvious that $C_{FB} \geq C$ since any rate achievable without feedback can be achieved with feedback; we are going to show that if the average error $p_e^{(n)} = \frac{1}{M}\sum_{i=1}^{M}\lambda_i^{(n)} \to 0$ for any $(2^{nR_{FB}}, n)$ feedback code, then $R_{FB} \leq C$.

Assume that W is uniform, then

$$nR_{FB} = H(W) = H(W|\hat{W}) + I(W,\hat{W}) \tag{17.4}$$

From Fano's inequity, we know $H(W|\hat{W}) \leq 1 + p_e^{(n)} nR_{FB}$; and from data processing inequality applied to the markov chain $W \to Y^n \to \hat{W}$, we have $I(W,\hat{W}) \leq I(W,Y^n)$. We will now focus on $I(W,Y^n)$:

$$\begin{aligned}
I(W,Y^n) &= H(Y^n) - H(Y^n|W) \\
&= H(Y^n) - \sum_{i=1}^{n} H(Y_i|Y^{i-1},W) \\
&= H(Y^n) - \sum_{i=1}^{n} H(Y_i|Y^{i-1},W,X_i) \quad (X_i \text{ is a function of } (Y^{i-1},W)) \\
&= H(Y^n) - \sum_{i=1}^{n} H(Y_i|X_i) \quad (\text{memoryless channel - conditioned on } X_i,\, Y_i \\
&\qquad\qquad \text{does not depend on } Y^{i-1} \text{ and } W)
\end{aligned}$$

$$
\begin{aligned}
&\le \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X_i) \\
&= \sum_{i=1}^{n} I(X_i, Y_i) \le nC
\end{aligned}
$$

Thus, we have

$$R_{FB} \le \frac{1}{n} + p_e^{(n)} R_{FB} + C$$

which implies that

$$\frac{R_{FB} - C}{R_{FB}} \le \frac{1}{nR_{FB}} + p_e^{(n)}$$

Hence, if $p_e^{(n)} \to 0$ as $n \to \infty$, it must be that $R_{FB} \le C$.

∎

17.3 Joint Souce-Channel Coding Theorem

So far we have considered source coding (or data compression) and channel coding (or data tranmission through a channel) independently. The source coding theorem tells us that as long as the source symbols are compressed to $R > H$ information bits/source symbol, then lossless data compression is possible. The channel coding theorem tells that as long as $R < C$ information bits are transmitted per channel use, error-free tranmission is possible. This suggests that by using a two-step procedure (compression followed by encoding), we can send a source with entropy H reliably through a channel with capacity C provided $H < C$. Can we do better if we use a single-step joint source-channel coding procedure? The answer is no, as we will see.

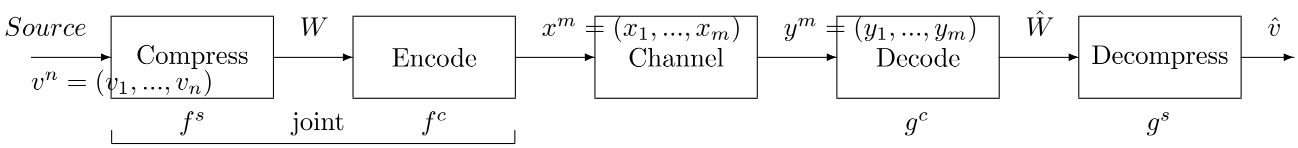

Figure 17.3: Block Diagram for Joint source-channel coding

Let n be the number of symbols in a message generated by the source and m be the number of bits in the codeword for the source string generated by the encoder.

We first define the following two quantities of the joint source-channel coding system:

- $T_s \equiv$ seconds it takes the source to generate one symbol
- $T_c \equiv$ seconds it takes the channel to transmit one bit

We claim that in any joint source-channel system, it must be that

$$nT_s \ge mT_c$$

This is because nT_s is the number of seconds used by the source to generate the message and mT_c is the number of seconds used by the channel to transmit the codeword. If the system is operating continuously, that is, there is not a growing queue of codes waiting to be transmitted through the channel, then it must be that the channel transmits the codeword at least as fast as the source is generating the messages.

Theorem 17.2 *If $v_1, v_2, ..., v_n$ is generated from a discrete stochastic ergodic process, then there exists a source-channel code with $p_e^{(n)} \to 0$ if $\frac{H(v)}{T_s} < \frac{C}{T_c}$.*

Conversely, for any stationary ergodic process $v_1, v_2, ..., v_n$, if $\frac{H(v)}{T_s} > \frac{C}{T_c}$ then $p_e^{(n)}$ cannot converge to 0.

Note that we can interprete $\frac{H(v)}{T_s}$ as number of bits of information per second generated by the source and $\frac{C}{T_c}$ as the maximum number of bits of information transmittable per second by the channel.

Proof: Achievability ($\frac{H(v)}{T_s} < \frac{C}{T_c} \Rightarrow p_e^{(n)} \to 0$):

By the source coding theorem, we know that, if $l = nR_s \geq nH(v)$, then we can compress the message to l bits with probability of error $p_e^{(n)} \leq \epsilon$.

By the channel coding theorem, we also know that if $l = mR_c \leq mC$, then we can encode l bits of messages to m bits of a codeword and transmit over the channel with probability of error $p_e^{(n)} \leq \epsilon$.

In order to chain these two results together however, we need to show that there exist a length l (of intermediate code W) that satisfies the two requirements on l. Since

$$\frac{H(v)}{T_s} \leq \frac{R_s}{T_s} = \frac{l}{nT_s}$$

and

$$\frac{C}{T_c} \geq \frac{R_c}{T_c} = \frac{l}{mT_c}$$

Combining the condition $nT_s \geq mT_c$ with the condition that $\frac{H(v)}{T_s} < \frac{C}{T_c}$, we see that there is indeed a suitable choice of l such that we can use a source code followed by a channel code, and achieve error of at most 2ϵ which goes to zero as $n \to \infty$.

Now we prove the converse ($p_e^{(n)} \to 0 \Rightarrow \frac{H(v)}{T_s} \leq \frac{C}{T_c}$):

$$\begin{aligned}
\frac{1}{n}H(v^n) &= \frac{1}{n}H(v^n|\hat{v}^n) + \frac{1}{n}I(v^n;\hat{v}^n) \\
&\leq \frac{1}{n} + p_e^{(n)}\log|\mathcal{V}| + \frac{1}{n}I(x^m;y^m) \quad \text{(Fano's and Info Processing Ineq)} \\
&\leq \frac{1}{n} + p_e^{(n)}\log|\mathcal{V}| + \frac{m}{n}C \text{ (since } I(x^m;y^m) \leq mC \text{ as in proof of converse of channel coding theorem)}
\end{aligned}$$

Since $\frac{m}{n} \leq \frac{T_s}{T_c}$ and $H(v) = \lim_{n\to\infty} \frac{H(v^n)}{n}$, we have taking limit as $n \to \infty$ on both sides:

$$H(v) - \frac{T_s}{T_c}C \leq \lim_{n\to\infty} p_e^{(n)} \log|\mathcal{V}|$$

Therefore, if $\frac{H(v)}{T_s} > \frac{C}{T_c}$, then even as $n \to \infty$, the error is always bounded away from 0.

Alternatively, if $p_e^{(n)} \to 0$, then

$$H(v) = \lim_{n\to\infty} \frac{H(v^n)}{n} \leq \frac{T_s}{T_c} C$$

i.e. $\frac{H(v)}{T_s} \leq \frac{C}{T_c}$.

■

17.4 Continuous Channel

The most common continuous alphabet channel is the Gaussian Channel, where the channel input X is real-valued and the output Y is as follows:

$$Y_i = X_i + Z_i \quad \text{where } Z_i \sim \mathcal{N}(0, \sigma^2) \ \text{ and } \mathrm{X_i} \perp \mathrm{Z_i}.$$

Without any constraints, the capacity is infinite. To see this observe that if $\sigma^2 = 0$, the channel can transmit any real-value with 0 error. Thus, the number of distinguishable input signals is infinite. Even if $\sigma^2 > 0$, we can still choose an infinite subset of input values that are arbitrarily far apart so that they are distinguishable at the output with arbitrarily small probability of error.

To avoid this, we can put a power constraint on the codeword x^n, i.e. $\frac{1}{n}\sum_{i=1}^n x_i^2 \leq P$ then we define the information capacity of a channel with power constraint P as:

$$C = \max_{p(x), E[X^2]\leq P} I(X;Y) \tag{17.5}$$

Lets evaluate the information capacity for a Gaussian channel with power constraint P is

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) = H(Y) - H(X+Z|X) = H(Y) - H(Z|X) = H(Y) - H(Z) && (17.6) \\ &\leq \frac{1}{2}\log 2\pi e(P+\sigma^2) - \frac{1}{2}\log 2\pi e\sigma^2 = \frac{1}{2}\log\frac{P+\sigma^2}{\sigma^2} && (17.7) \end{aligned}$$

The first term in the second line follows since Y is a real valued random variable with second order moment constraints $E[Y^2] \leq P + \sigma^2$, and hence its entropy is maximized if the distribution of Y is $\mathcal{N}(0, P+\sigma^2)$. Notice that this distribution of Y, and hence the upper bound on the mutual informtion, is achieved if $p(x) = \mathcal{N}(0, P)$. Thus, the capacity of a Gaussian channel with power constraint P is $\frac{1}{2}\log\frac{P+\sigma^2}{\sigma^2}$.