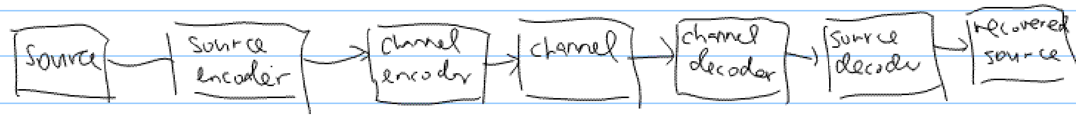
Note Title 8/13/2007

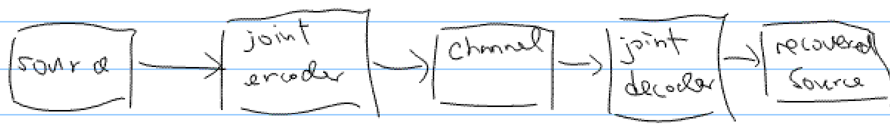
Separation Theorem

Read 7.13 (8.13 for 1st edition)

We have considered source & channel coding separately in previous lectures. In practice, if we have a source to be transmitted through a channel, a natural option is to first compress the source with a source code & then code the compressed source with channel code to withstand the noise from the channel.

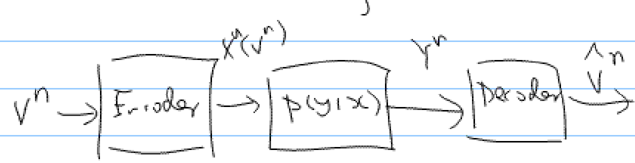


However, is this scheme optimal? Our source coding scheme doesn't take into account of the channel & our channel coding scheme ignores the source distribution. Will we get better performance if we use a "joint-source channel" approach?



It turns out, for point to point communication, we get optimum performance even we design the source & channel coding scheme separately. This is usually known as the separation Theorem. Separation Thm may seem intuitive to some of you. However, note that in many multiterminal communication systems, the separation Thm does not apply. (This is also one of the reasons why network communication systems are much harder to analyze.)

We will prove the separation Thm by showing that a joint source channel coding scheme does not perform better than separately designed source & channel coding schemes.



<u>Thm 1:</u> A source V can transmit through a channel with a capacity C losslessly iff $H(\mathcal{V}) < C$ regardless whether joint or separate design is used.

entropy rate

Pf: The converse part (if $H(\mathcal{V}) < C$, lossless transmission is possible) is easy.

Since separate design is just a special case of joint design,

We can compress source V to rate $H(\mathcal{V})$ by optimal source coder &

we can transmit the compressed bits losslessly to the decoder as long as the rate $R = H(\mathcal{V}) < C$. $\therefore$ we can recover V losslessly at the decoder.

The forward part (if transmission is lossless, $H(\mathcal{V}) < C$) can be proved very similar to the case of channel coding Thm.

Let's review the Fano's inequality here. Let

$$E = \begin{cases} 1 & \hat{V}^n \neq V^n \\ 0 & \text{o.w.} \end{cases} \quad \& \quad P_e = \Pr(E=1)$$

Then
$$H(E, V^n | \hat{V}^n) = H(E|\hat{V}^n) + H(V^n|\hat{V}^n, E)$$
$$= H(V^n|\hat{V}^n) + H(E|\cancel{V^n, \hat{V}^n})^{0}$$

$$\therefore \; H(V^n|\hat{V}^n) = H(E|\hat{V}^n) + H(V^n|\hat{V}^n, E)$$
$$\leq 1 + H(V^n|\hat{V}^n, E=1)P_e + \cancel{H(V^n|\hat{V}^n, E=0)}^{0}(1-P_e)$$
$$\leq 1 + n\log|\mathcal{V}| \cdot P_e$$

$$H(\mathcal{V}) \leq \frac{H(V^n)}{n}$$ (from the def. of entropy of stationary source $H(\mathcal{V}) = \frac{1}{n}H(V^n)$ for $n \to \infty$)

$$= \frac{1}{n}H(V^n|\hat{V}^n) + \frac{1}{n}I(V^n;\hat{V}^n)$$
$$\leq \frac{1}{n}(1 + n\log|\mathcal{V}|P_e) + \frac{1}{n}I(X^n;Y^n)$$ (data processing inequality)
$$\leq \frac{1}{n}(1 + n\log|\mathcal{V}|P_e) + I(X;Y)$$ (memoryless channel)
$$\leq \frac{1}{n}(1 + n\log|\mathcal{V}|P_e) + C$$

where the first term approach 0 as $n \to \infty$. ∎