LECTURE 15

Last time:

- Feedback channel: setting up the problem
- Perfect feedback
- Feedback capacity

Lecture outline

- Data compression
- Joint source and channel coding theorem
- Converse
- Robustness
- Brain teaser

Reading: Sct. 8.13.

Data compression

Consider coding several symbols together

$C : \mathcal{X}^n \mapsto \mathcal{D}^*$

expected codeword length is $\sum_{\underline{x}^n \in \mathcal{X}^n} P_{\underline{X}^n}(\underline{x}^n) l(\underline{x}^n)$

optimum satisfies

$H_D(\underline{X}^n) \leq L^* \leq H_D(\underline{X}^n) + 1$

per symbol codeword length is

$\frac{H_D(\underline{X}^n)}{n} \leq \frac{L^*}{n} \leq \frac{H_D(\underline{X}^n)}{n} + \frac{1}{n}$

Thus, we have that the rate R is lower bounded by entropy

Channel coding

Coding theorem:

For any source of messages, any rate R below C is feasible, i.e achievable with low enough probability of error

$$E_{codebooks,messages}[P_e] \leq 2^{-NE_r(R,P_X)}$$

For the P_X that yields capacity, $E_r(R, P_X)$ is positive for $R < C$

also the maximum probability of error is upper bounded

Converse:

For any source of IID messages, any transmission at rate R above C is bounded away from 0

$$R \leq \frac{1}{n} + P_e R + C$$

Joint source and channel coding

Is $H \leq C$ enough?

Should coding be done jointly or not?

Joint source channel coding

Consider a sequence of input symbols $\underline{V}^n = (V_1, \ldots V_n)$

Assume that $\underline{V}^n = (V_1, \ldots V_n)$ satisfies the AEP

We map the sequence onto a codeword $\underline{X}(\underline{V}^n)$

The receiver receives $\underline{Y}$ (a vector) and creates an estimate $\widehat{\underline{V}^n} = g(\underline{Y})$ of $\underline{V}^n$

We consider the end-to-end probability of error:

$P_e^n = P(\underline{V}^n \neq \widehat{\underline{V}^n})$

Joint source channel coding

$A_\epsilon^{(n)}$ is a typical set if it is the set of sequences in the set of all possible sequences $\underline{v}^n \in \underline{\mathcal{V}}^n$ with probability:

$$2^{-n(H(V)+\epsilon)} \leq P_{\underline{V}^n}(\underline{v}^n) \leq 2^{-n(H(V)-\epsilon)}$$

Since we assume that our V sequence satisfies the AEP, for all ϵ, there exists a typical set that has probability

$$Pr(A_\epsilon^{(n)}) \geq 1 - \epsilon$$

and has cardinality upper bounded by $2^{n(H(V)+\epsilon)}$

We encode only those elements in the typical set, the ones outside will contribute ϵ to the probability of error

Joint source channel coding

We require at most $n(H(V) + \epsilon)$ bits to describe the elements of the typical set

Let us create a set of messages from these sequences

The maximum probability of any message being decoded in error is arbitrarily close to 0 as long as $R = H(V) + \epsilon < C$ for all n LARGE ENOUGH

Note: the size of the $\underline{X}$ may not be n, but it grows in n

The receiver decodes to attempt to recover an element from the typical set

Because we have assumed that the AEP is satisfied, we have that ϵ can be chosen to be as small as we want

Joint source channel coding

The probability of error is upper bounded by the probability that we are not in the typical set, plus the probability that we are in the typical set but we have an error

Thus, we have that for any upper bound ϵ on the probability of error, there exists a large enough n such that the probability of error is upper bounded by ϵ

Note: the probability of error from not being in the typical set may be higher or lower than that from incorrect decoding

Thus, we have shown the forward part of the theorem:

for any $V_1, \ldots, V_n$ that satisfies the AEP, there exists a source code and a channel code such that $P_e^n \to 0$ as $n \to \infty$

Moreover, **the source coding and the channel coding may be done independently**

Converse to the joint source channel coding theorem

We need to show that a probability of error bounded that converges 0 as $n \rightarrow \infty$ implies that $H(V) < C$

Let us use Fano's inequality:

$H(\underline{V}^n | \widehat{\underline{V}^n}) \leq 1 + nP_e^n \log(|\mathcal{V}|)$

Let us now use the definition of entropy rate:

The entropy rate of a stochastic process V is

$$\lim_{n \rightarrow \infty} \frac{1}{n} H(\underline{V}^n)$$

if it exists

Equivalently, the entropy rate of a stochastic process is

$\frac{1}{n} H(\underline{V}^n | \widehat{\underline{V}^n}) + \frac{1}{n} I(\underline{V}^n; \widehat{\underline{V}^n})$

Converse to the joint source channel coding theorem

Hence, we have that

$$
\begin{aligned}
H(V) &\leq \frac{1}{n}(1 + nP_e^n \log(|\mathcal{V}|)) + \frac{1}{n} I(\underline{V}^n; \widehat{\underline{V}^n}) \\
&= \frac{1}{n} + P_e^n \log(|\mathcal{V}|) + \frac{1}{n} I(\underline{V}^n; \widehat{\underline{V}^n}) \\
&\leq \frac{1}{n} + P_e^n \log(|\mathcal{V}|) + \frac{1}{n} I(\underline{X}; \underline{Y}) \\
&\quad \text{from DPT} \\
&\leq \frac{1}{n} + P_e^n \log(|\mathcal{V}|) + C
\end{aligned}
$$

since $P_e^n \to 0$, it must be that $H(V) \leq C$ for the above to be true for all n

Robustness of joint source channel coding theorem and its converse

The joint channel source coding theorem and its converse hold under very general conditions:

- memory in the input

- memory in channel state

- multiple access

- but not broadcast conditions

A brain teaser

6.441 Information Theory
Spring 2010