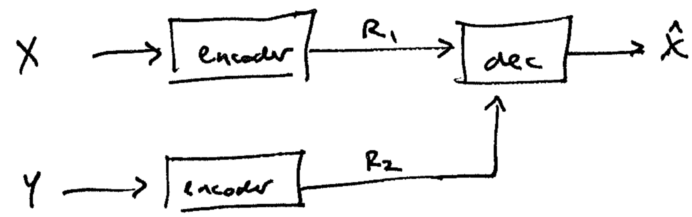
Source Coding with Coded Side Information

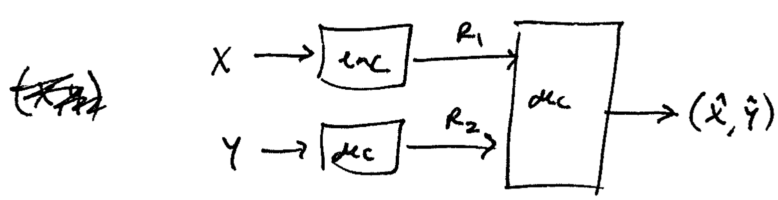
- Helper Problem
- Wyner-Ahlswede-Korner (WAK) problem



X and Y are correlated random variables, they are encoded separately, jointly decoded, but only X is recovered.

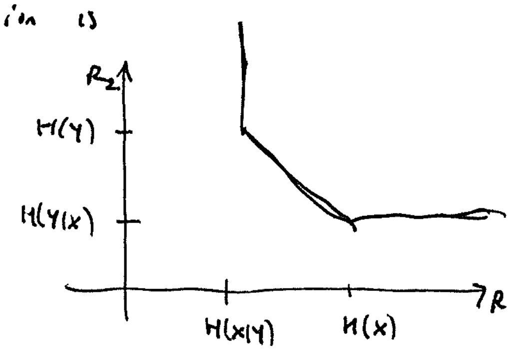
If we allow R_2 bits to describe Y, which will help us in describing X, how many bits R_1 are needed to describe X?

We can use the standard source coding thm and the Slepian-Wolf thm to get ~~the~~ "corner points":



<u>Slepian-Wolf thm</u> For the distributed source problem of source (X,Y) drawn iid $\sim p(x,y)$, achievable rate region is

$$R_1 \geq H(X|Y)$$
$$R_2 \geq H(Y|X)$$
$$R_1 + R_2 \geq H(X,Y)$$



If $R_2 > H(Y)$ then Y can be described perfectly, and by Slepian-Wolf, $R_1 = H(X|Y)$ are enough to describe X.

If $R_2 = 0$, then describe X without help, and $R_1 = H(X)$.

In between, our intuition is we will use a lossy code for Y, some sort of rate-distortion code s.t. $R_2 = I(Y;\hat{Y})$.

This will allow us to describe X using $H(X|\hat{Y})$ bits in presence of the coded side informatn.

Thm: Let $(X,Y) \sim p(x,y)$. ~~If~~ If Y is encoded at rate R_2 and X is encoded at rate R_1, we can recover X with arbitrarily small error probability if and only if

$$R_1 \geq H(X|U)$$
$$R_2 \geq I(Y;U)$$

for some joint probability mass function $p(x,y)\,p(u|y)$, where $|\mathcal{U}| \leq |\mathcal{Y}|+2$.

Here U is an auxilliary r.v. that represents the coded side information and to find closed-form expression for the rate region, we must optimize over the auxilliary r.v. that satisfies the appropriate Markov condition.

Let us prove converse first.

A key way to prove converses is to connect operational notions like error probability to informational quantities using Fano's ineq.

<u>Fano's ineq</u>: Let X and $\hat{X}$ be random variables taking values in the same alphabet $\mathcal{X}$. Then letting $P_e = \Pr[X \neq \hat{X}]$

$$H(X|\hat{X}) \leq h_2(P_e) + P_e \log(|\mathcal{X}|-1)$$

<u>Converse proof</u>: A source code for the setting consists of mappings $f_n(X^n)$ and $g_n(Y^n)$ such that rates of f_n and g_n are less than R_1 and R_2, respectively. Also decoder h_n such that

$$P_e^{(n)} = \Pr\left[h_n\left(f_n(X^n), g_n(Y^n)\right) \neq X^n \right] < \varepsilon.$$

So we assert our losslessness and rate requirements.

Let $S = f_n(X^n)$ and $T = g_n(Y^n)$ as new random variables to represent the encodings.

Since we can recover X^n from S and T with small error probability, by Fano's ineq

$$H(X^n | S,T) \leq n\varepsilon_n.$$

$nR_2 \geq H(T)$ since g_n maps to $\{1, 2, \ldots, 2^{nR_2}\}$

$\geq I(Y^n; T)$ by mutual information property.

$= \sum_{i=1}^{n} I(Y_i; T \mid Y_1, \ldots, Y_{i-1})$

$= \sum_{i=1}^{n} I(Y_i; T, Y_1, \ldots, Y_{i-1})$ by chain rule and since Y_i independent of $Y_1, \ldots, Y_{i-1}$ so $I(Y_i; Y_1, \ldots, Y_{i-1}) = 0$.

$= \sum_{i=1}^{n} I(Y_i; U_i)$

if we define $U_i = (T, Y_1, \ldots, Y_{i-1})$.

Similarly for R_1

$nR_1 \geq H(S)$

$\geq H(S \mid T)$ since conditioning only reduces entropy.

$= H(S \mid T) + H(X^n \mid S, T) - H(X^n \mid S, T)$.

$\geq H(X^n, S \mid T) - n\varepsilon_n$ by Fano's ineq.

$= H(X^n \mid T) - n\varepsilon_n$ since S is function of X^n and chain rule

$= \sum_{i=1}^{n} H(X_i \mid T, X^{i-1}, Y^{i-1}) - n\varepsilon_n$ by chain rule.

$\geq \sum_{i=1}^{n} H(X_i \mid T, X^{i-1}, Y^{i-1}) - n\varepsilon_n$ since conditioning only reduces entropy

$= \sum_{i=1}^{n} H(X_i \mid T, Y^{i-1}) - n\varepsilon_n$ which is from the Markov structure of how the system is constructed: $X_i \to (T, Y^{i-1}) \to X^{i-1}$ forms Markov chain since X_i doesn't have info. about X^{i-1} that is not in Y^{i-1} and T.

$= \sum_{i=1}^{n} H(X_i \mid U_i) - n\varepsilon_n$ from our previous definition of U.

Also since X_i contains no more information about U_i than is in Y_i,

we have $X_i - Y_i - U_i$ as a Markov chain. Now we get:

$$R_1 \geq \frac{1}{n} \sum_{i=1}^{n} H(X_i | U_i)$$

$$R_2 \geq \frac{1}{n} \sum_{i=1}^{n} I(Y_i ; U_i)$$

Now introduce a timesharing random variable Q, that allows us to essentially interpolate (over time), we get

$$R_1 \geq \frac{1}{n} \sum_{i=1}^{n} H(X_i | U_i, Q=i) = H(X_Q | U_Q, Q)$$

$$R_2 \geq \frac{1}{n} \sum_{i=1}^{n} I(Y_i ; U_i | Q=i) = I(Y_Q ; U_Q | Q).$$

since Q is independent of Y_Q (the distribution of Y_i does not depend on i)

we have

$$I(Y_Q ; U_Q | Q) = I(Y_Q ; U_Q, Q) - I(Y_Q ; Q) = I(Y_Q ; U_Q, Q).$$

Now notice X_Q and Y_Q have joint distribution $p(x,y)$ from theorem.

Define auxilliary r.v. $U = (U_Q, Q)$, $X = X_Q$, and $Y = Y_Q$, we have

existence of r.v. U s.t.

$$R_1 \geq H(X|U)$$
$$R_2 \geq I(Y;U).$$

Note that the alphabet size restriction on U will come from Caratheodory's thm in convex analysis.

Achievability side

random coding argument like for $R(D)$, but specifically make use of the idea of random binning.

It is a way to connect codewords for X representation and Y representation.

For analysis, we will use strong typicality in proof.

If we have a triple of r.v.s X, Y, Z, a triplet of sequences x^n, y^n, z^n is ε-strongly typical if

$$\left| \frac{1}{n} N(a,b,c \mid x^n, y^n, z^n) - p(a,b,c) \right| < \frac{\varepsilon}{|\mathcal{X}||\mathcal{Y}||\mathcal{Z}|}$$

has the same concentration properties we've seen before.