EE 231A Information Theory
Lecture 18
Final Assessment Review, or
These are a few of my favorite things

A. Symmetry and computing channel capacity
B. Mutual information for continuous alphabets
C. Parallel Gaussian Channels
D. Computing the Rate Distortion Function
E. Multiple Access Channels
F. Slepian-Wolf

Part A: Symmetry and computing channel capacity

Weak symmetry

- $$p(y \mid x) = \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.5 & 0.3 & 0.2 \\ 0.2 & 0.5 & 0.3 \end{bmatrix}$$

	y_1	y_2	y_3
x_1	0.3	0.2	0.5
x_2	0.5	0.3	0.2
x_3	0.2	0.5	0.3

- Entry in x^{th} row and y^{th} column is $p(y|x)$.
- $p\left(y_2 \mid x_3\right) = 0.5$
- A channel is weakly symmetric if all the rows are permutations of each other and the column sums are equal.

Weak symmetry

- For a weakly symmetric channel,

$$C = \log|\mathcal{Y}| - H(r)$$

pmf of any row

which is achievable by a uniform distribution.

- $$\begin{aligned} I(X;Y) &= H(Y) - H(Y \mid X) \\ &= H(Y) - H(r) \\ &\le \log|\mathcal{Y}| - H(Y \mid X) \end{aligned}$$

Cyclic Symmetry

Matrix Conditions for Cyclic Symmetry The channel described by P has cyclic symmetry if the following two conditions are satisfied:

- All the rows of P are permutations of each other.
- The set C of columns of P can be separated into a certain collection of mutually exclusive, collectively exhaustive subsets S_i. (i.e. The equations below are satisfied:)

$$\bigcup_i S_i = C \tag{2}$$

$$\text{if } S_i \neq S_j \text{ then } S_i \bigcap S_j = \emptyset\,, \tag{3}$$

such that each subset S_i may be completely constructed from any one element of S_i as follows: S_i contains exactly one instance of each cyclic shift of that element, and nothing else. Note that the elements of both C and S_i are columns.

Cyclic Symmetry

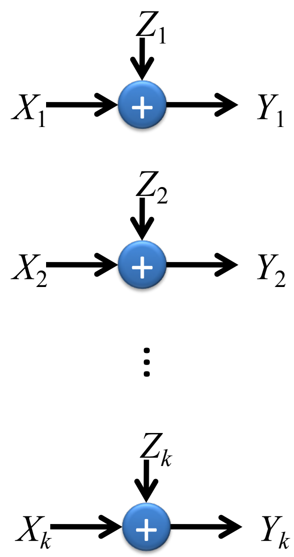
- Let's look at Problem Set 4 problem 4…

Part B: Mutual information for continuous alphabets

Let's look at problem 3 on HW 5…

Part C: Parallel Gaussian Channels

Parallel Gaussian Channels



Capacity of parallel channels

$$C = \max_{f(x_1, x_2, \ldots, x_k),\, \sum EX_i^2 \le P} I(X_1, X_2, \ldots, X_k; Y_1, Y_2, \ldots, Y_k)$$

Capacity of Parallel Gaussian Channels

$$
\begin{aligned}
I(X^k;Y^k) &= h(Y^k) - h(Y^k \mid X^k) \\
&= h(Y^k) - h(Z^k \mid X^k) \\
&= h(Y^k) - h(Z^k) \\
&= h(Y^k) - \sum_i h(Z_i) \\
&\le \sum_i \big(h(Y_i) - h(Z_i) \big) \\
&\le \sum_i \frac{1}{2} \log(1 + \frac{P_i}{N_i})
\end{aligned}
$$

Achievability

- Where $P_i = EX_i^2, \sum P_i = P$ and equality is achieved when

$$
X^k \sim \mathcal{N}\left(0, \begin{bmatrix} P_1 & 0 & \cdots & 0 \\ 0 & P_2 & 0 & 0 \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & P_k \end{bmatrix}\right)
$$

- What are the optimal P_i's?

Convert to convex optimization

$$\text{maximize} \quad \sum_i \frac{1}{2}\log\left(1+\frac{P_i}{N_i}\right) \qquad \text{(Concave function)}$$

$$\text{subject to} \quad \sum_i P_i \le P \qquad \text{(linear constraint)}$$

- Convex optimization techniques may be applied.

Lagrange multipliers (Duality)

$$J(P_1,\ldots,P_k) = \sum_i \frac{1}{2}\log\left(1+\frac{P_i}{N_i}\right) + \lambda \sum_i P_i$$

$$\frac{\partial J}{\partial P_i} = \frac{1}{2}\frac{\partial}{\partial P_i}\log_2 e\left[\ln\left(\frac{1}{N_i}\right) + \ln(N_i+P_i)\right] + \lambda \quad = \frac{\frac{1}{2}\log_2 e}{N_i+P_i} + \lambda$$

$$\boxed{\text{Set} \quad \frac{\partial J}{\partial P_i} = 0}$$

$$\frac{\frac{1}{2}\log_2 e}{N_i+P_i} + \lambda = 0 \implies \frac{1}{2}\log_2 e = -\lambda\left(N_i+P_i\right) \implies P_i = \underbrace{\frac{-1}{2\lambda}\log_2 e}_{\nu} - N_i$$

Optimal solution

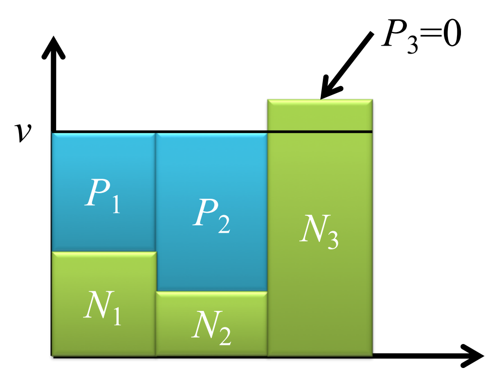
- Setting $\frac{\partial J}{\partial P_i} = 0$ is equivalent to setting $P_i = v - N_i$ for some constant v. (except where $v - N_i$ is negative.)

- Choose v to meet power constraint.

$$\sum (v - N_i)^+ = P$$

$$(x)^+ = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

Water filling



Part D: Computing the Rate-Distortion Function

Rate-Distortion function

- For X i.i.d. $\sim p(x)$ and $d(x,\hat{x})$ bounded

$$R(D) = \min_{p(\hat{x}|x), E[d] \le D} I(X;\hat{X})$$

$$E[d] = \sum_{x,\hat{x}} p(x,\hat{x}) d(x,\hat{x})$$

$$= \sum_{x,\hat{x}} p(x) \underbrace{p(\hat{x} \mid x)}_{\text{This is what we control.}} d(x,\hat{x})$$

Computing $R(D)$

- *First example:* $R(D)$ for a binary source with Hamming distortion: $P(x=1)=p\le\frac{1}{2}$
- One way to find $R(D)$ is to find a lower bound on $I(X;\hat{X})$ and then achieve it.
- We have the constraint $E[d]=P(\hat{X}\ne X)\le D$

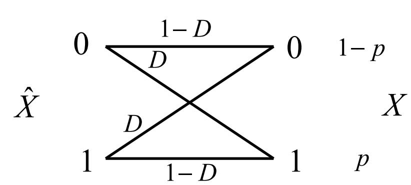
Lower bound on $I(X;\hat{X})$

$$\begin{aligned}
I(X;\hat{X}) &= H(X)-H(X\mid\hat{X})\\
&= H(p)-H(X\oplus\hat{X}\mid\hat{X})\\
&\ge H(p)-H(X\oplus\hat{X})\\
&= H(p)-H(P(X\ne\hat{X}))\\
&\ge H(p)-H(D) \qquad \text{for } D\le\frac{1}{2}
\end{aligned}$$

Achievability of the lower bound

- So $R(D) \geq H(p) - H(D)$
- Can we achieve $I(X;\hat{X}) = H(p) - H(D)$ with $E(d) \leq D$?
- We need to find a $p(x,\hat{x})$ that does that.

The Test Channel



- $H(p) - H(D)$ is the $I(X;\hat{X})$ for a BSC with transition probability D and output distribution p, 1-p.

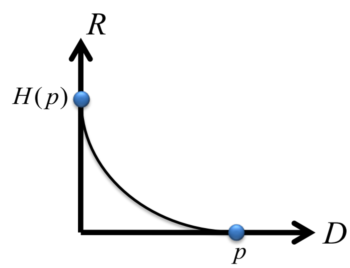
- Can we find an input distribution $p(\hat{x})$ to make it work?

Achievability of the lower bound (cont.)

$$P(X=0) = (1-D)P(\hat{X}=0) + DP(\hat{X}=1)$$
$$= 1-p$$
$$P(X=1) = (1-D)P(\hat{X}=1) + DP(\hat{X}=0)$$
$$= p$$

$$\boxed{\begin{aligned} p(\hat{x}=0) &= \frac{1-p-D}{1-2D} \\ p(\hat{x}=1) &= \frac{p-D}{1-2D} \end{aligned}}$$

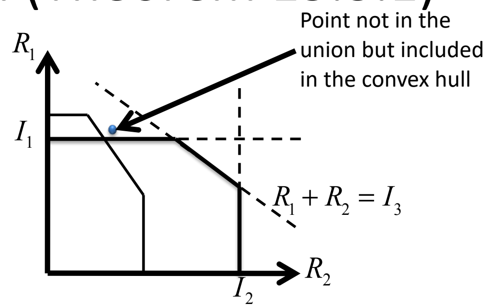
- So $R(D) = H(p) - H(D).$



Part E: Multiple Access Channels

Capacity region (Theorem 15.3.1)

$$R_1 < I(X_1;Y \mid X_2) = I_1$$
$$R_2 < I(X_2;Y \mid X_1) = I_2$$
$$R_1 + R_2 < I(X_1, X_2;Y) = I_3$$



- The capacity region is a union of many pentagons, possibly further increased by the closure of a convex hull operation.
  - Points in the convex hull are achieved by time-sharing between two points in pentagons.

Noiseless MAC

6. (5 pts) *Noiseless Multiple Access Channel.*

Consider the two-user multiple access channel with no noise so that $Y = f(X_1, X_2)$ and f is a deterministic function.

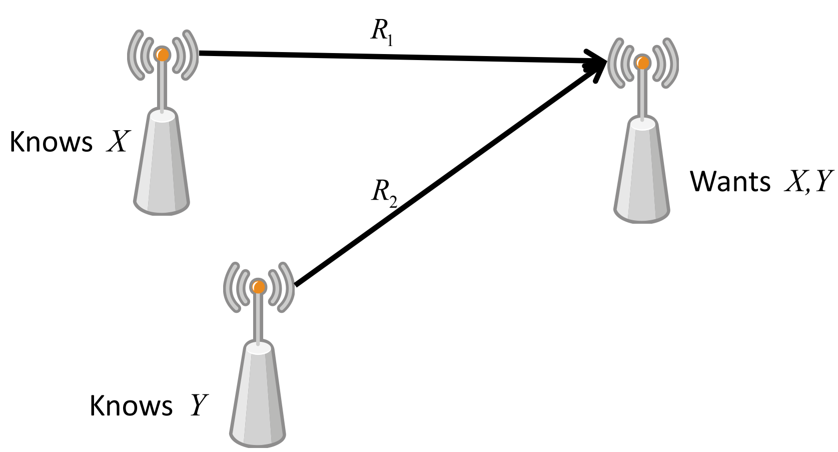
Show that for each pentagon includes the constraint $R_1 + R_2 \le H(Y)$.

Solution

$$R_1 + R_2 \le I(X_1, X_2; Y) \quad (18)$$
$$= H(Y) - \underbrace{H(Y|X_1, X_2)}_{0} \quad (19)$$
$$= H(Y) \quad (20)$$

Part F: Slepian-Wolf Encoding of Correlated Sources

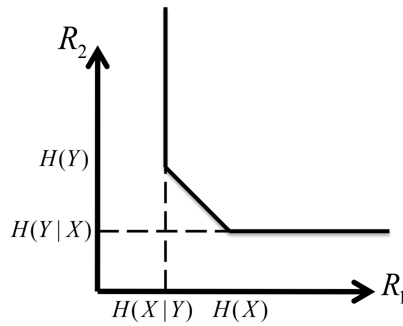
Encoding of correlated sources



Encoding of correlated sources (cont.)

- Clearly $R_1 > H(X), R_2 > H(Y)$ will work.
- Can we do better?
  - Yes!
- All we need is

$$R_1 > H(X \mid Y)$$
$$R_2 > H(Y \mid X)$$
$$R_1 + R_2 > H(X,Y)$$

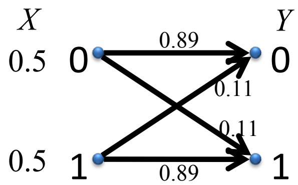


Note

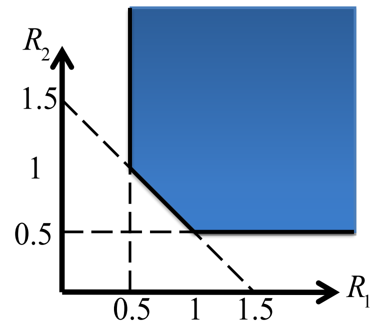
- $H(X \mid Y) + H(Y \mid X) \leq H(X \mid Y) + H(Y)$
$$= H(X,Y)$$

- The bound $R_1 + R_2 > H(X,Y)$ is always enough to satisfy the bounds on R_1 and R_2.

Example: Sending X and Y of a BSC.



$H(X) = 1$
$H(Y) = 1$
$H(Y \mid X) = H(0.89) = 0.5w$
$H(X,Y) = H(Y \mid X) + H(X) = 1.5$
$H(X \mid Y) = H(X,Y) - H(Y) = 0.5$



- So to encode n (X,Y) pairs we need $2n$ bits by separately encoding X and Y, but we only need $1.5n$ bits if we use Slepian-Wolf.

$R_1 > H(X \mid Y)$
$R_2 > H(Y \mid X)$
$R_1 + R_2 > H(X,Y)$