LECTURE 12

Last time:

- Error Exponents
- Strong Coding Theorem
- Binary Source/ BSC

Lecture outline

- Computing the channel capacity.
- Blahut-Arimoto Algorithm.
Review

- Strong Coding Theorem

For $R < C$, the error probability decays exponentially with the codeword length $n$.

- Two different ranges of rate $C > R \geq R_{\text{crit}}$ and $R < R_{\text{crit}}$.

- Random code is optimal for high data rate, but not optimal for lower rates.
Maximizing the Mutual Information

- For a given channel $P_{Y|X}(y|x)$,

$$
C = \max_{P_X} I(X;Y) \\
= \max_{P_X} \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P_X(x) P_{Y|X}(y|x) \log \frac{P_X(x) P_{Y|X}(y|x)}{P_X(x) \sum_{x' \in \mathcal{X}} P_X(x') P_{Y|X}(y|x')} \\
= \max_{P_X} \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P_X(x) P_{Y|X}(y|x) \log \left( \frac{P_{X|Y}(x|y)}{P_X(x)} \right)
$$

- Recall for symmetric channels, the optimal input distribution is the uniform distribution.

- In general, analytical solutions of the optimal input distribution is not always available.

- The motivations of numerically computing the capacity.

- The Difficulties, high dimensional non-linear optimization.
Alternating Optimization

Consider the double supremum

\[
\max_{u_1 \in A_1} \max_{u_2 \in A_2} f(u_1, u_2)
\]

suppose \( f \) is bounded from above and is continuous and has continuous partial derivatives, \( A_1, A_2 \) are convex.

- Suppose now for all \( u_1 \in A_1 \), there exists unique \( c_2(u_1) \) that
  \[
  f(u_1, c_2(u_1)) = \max_{u'_2} f(u_1, u'_2)
  \]
  and for any \( u_2 \in A_2 \), there exists unique \( c_1(u_2) \) that
  \[
  f(c_1(u_2), u_2) = \max_{u'_1} f(u'_1, u_2)
  \]
Now we can find the maximum with an alternating algorithm.

- Initialize: pick any $u_1^{(0)}$,
- find $u_2^{(0)} = c_2(u_1^{(0)})$,
- In the $k^{th}$ step, suppose we have already computed $u_1^{(k-1)}, u_2^{(k-1)}$, now we compute

$$u_1^{(k)} = c_1(u_2^{(k-1)})$$
$$= \arg \max_{u_1'} f(u_1', u_2^{(k-1)})$$

and

$$u_2^{(k)} = c_2(u_1^{(k)})$$
$$= \arg \max_{u_2'} f(u_1^{(k)}, u_2')$$

- The function value is non-decreasing in each step, since bounded from above, the sequence converges.
- Does it necessarily converges to the maximum?
- Why this is related to maximizing the mutual information?
Lemma Fix any input distribution $P_X$, and channel $P_{Y|X}$. Suppose $P_X(x) > 0$ for any $x$. Consider the following maximization problem

$$\max_{Q_{X|Y}} \sum_x \sum_y P_X(x) P_{Y|X}(y|x) \log \frac{Q_{X|Y}(x|y)}{P_X(x)}$$

where the maximization is taken over all $Q$ such that

$$Q_{X|Y}(x|y) = 0 \quad \text{iff} \quad P_{Y|X}(y|x) = 0$$

The maximum is obtained at

$$Q_{X|Y}^*(x|y) = \frac{P_X(x) P_{Y|X}(y|x)}{\sum_{x'} P_X(x') P_{Y|X}(y|x')}$$

i.e., the maximizing $Q$ is the true transition distribution from $Y$ to $X$ given the input $P_X$ and the channel $P_{Y|X}$. 
Proof First check that $Q^*$ satisfies the constraint.

$$\sum_x \sum_y P_X(x) P_{Y|X}(y|x) \log \frac{Q^*_{X|Y}(x|y)}{P_X(x)}$$

$$- \sum_x \sum_y P_X(x) P_{Y|X}(y|x) \log \frac{Q_X|Y(x|y)}{P_X(x)}$$

$$= \sum_x \sum_y P_X(x) P_{Y|X}(y|x) \log \frac{Q^*_{X|Y}(x|y)}{Q_{X|Y}(x|y)}$$

$$= \sum_x \sum_y P_Y(y) Q^*(X|Y)(x|y) \log \frac{Q^*_{X|Y}(x|y)}{Q_{X|Y}(x|y)}$$

$$= \sum_y P_Y(y) D(Q^*_{X|Y}||Q_{X|Y})$$

$$\geq 0$$
How to Use this?

\[ C = \max_{P_X} I(X; Y) \]

\[ = \sup_{P_X > 0} \max \sum P_X(x) P_{Y|X}(y|x) \log \frac{Q_{X|Y}(x|y)}{P_X(x)} \]

- To obtain the second equality above, we need \( P_X(x) > 0 \) strictly for any \( x \), while the optimal input distribution might have for some \( x \), \( P^*_X(x) = 0 \).

- The continuity of \( I(X; Y) \) with respect to the input.

- Now we have a joint optimization of two variables: \( P_X \) and \( Q_{X|Y} \), to apply the alternative algorithm, need to check:
  
  - The space of distributions is convex.
  - The function is bounded from above.

\[ f(P, Q) \leq I_P(X; Y) \leq H(X) \leq \log |\mathcal{X}| \]
Updating Rules

- Given $P^{(k)}$

\[
Q^{(k)} = \arg \max_{Q_{X|Y}} \sum_{x,y} P^{(k)}(x) P_{Y|X}(y|x) \log \frac{Q_{X|Y}(x|y)}{P^{(k)}(x)} = \frac{P^{(k)}(x) P_{Y|X}(y|x)}{\sum_{x'} P^{(k)}(x') P_{Y|X}(y|x')} \]

- Given $Q^{(k)}$

\[
P^{(k+1)} = \arg \sup_{P>0} \sum_{x,y} P(x) P_{Y|X}(y|x) \log \frac{Q^{(k)}_{X|Y}(x|y)}{P(x)} \]

the maximization is taken over all the distributions on $\mathcal{X}$.

Constraints

- $\sum_x P(x) = 1$,
- $P(x) > 0$ for any $x$. 
• ignore the positive constraint first, define

\[ J = \sum_{x,y} P(x)P(y|x) \log \frac{Q(x|y)}{P(x)} - \lambda \sum_x P(x) \]

for each \( x \in \mathcal{X} \), set

\[
0 = \frac{\partial J}{\partial P(x)}
\]

\[
= \sum_y P(y|x) \log Q(x|y)
\]

\[
- \sum_y P(y|x)[\log P(x) + 1] - \lambda
\]

\[
= \sum_y P(y|x) \log Q(x|y) - [\log P(x) + 1] - \lambda
\]

Therefore,

\[
P(x) = e^{-(\lambda+1)} \prod_y Q(x|y)^{P(y|x)}
\]

Solve for \( \lambda \) and get

\[
P(x) = \frac{\prod_y Q(x|y)^{P(y|x)}}{\sum_{x'} \prod_y Q(x'|y)^{P(y|x')}}
\]
Convergence of the Alternating Optimization Algorithm

**Lemma** if $f$ is concave,

$$f(u_1^{(k)}, u_2^{(k)}) \to f^*$$

where $f^*$ is the desired maximum value.

- The sequence on the LHS converges, but is it possible to converge at some point with $f(u_1, u_2) < f^*$?
- For a concave function, local maximum is global maximum.

Suppose the algorithm converges at $u^{(\infty)} = (u_1^{(\infty)}, u_2^{(\infty)})$, this means

$$\left. \frac{\partial f}{\partial u_1} \right|_{u^{(\infty)}} = 0$$

$$\left. \frac{\partial f}{\partial u_2} \right|_{u^{(\infty)}} = 0$$
This means the derivative of $f$ at $u^{(\infty)}$ in any direction is 0.

Suppose there is a $u^*$ for which

$$f(u^*) > f(u^{(\infty)})$$

look at $f$ along the direction $u^* - u^{(\infty)}$ to get back to the single dimension case — Contradiction
Apply to the Channel Capacity

Now we only need to check

\[
f(P, Q) = \sum_{x,y} P(x) P_{Y|X}(y|x) \log \frac{Q(x|y)}{P(x)}
\]

is concave in \( P \) and \( Q \).

For each value of \((x, y)\), let \( p = P(x) \) and \( q = Q(x|y) \), \( r \in [0, 1] \) and \( \bar{r} = 1 - r \), it is sufficient to show that

\[
(rp_1 + \bar{r}p_2) \log \frac{rq_1 + \bar{r}q_2}{rp_1 + \bar{r}p_2} \leq rp_1 \log \frac{q_1}{p_1} + \bar{r}p_2 \log \frac{q_2}{p_2}
\]

Let \( a = \frac{rp_1}{rp_1 + \bar{r}p_2} \), \( b = \frac{rq_1}{rq_1 + \bar{r}q_2} \).

\[
LHS - RHS = (rp_1 + \bar{r}p_2) \left[ a \log \frac{b}{a} + (1 - a) \log \frac{1 - b}{1 - a} \right] = -(rp_1 + \bar{r}p_2) D(a||b) \leq 0
\]