Communication with Side Information at the Transmitter

Aaron Cohen
6.962
February 22, 2001
Outline

• Basic model of communication with side info at the transmitter
  – Causal vs. non-causal side information
  – Examples
• Relationship with watermarking and other problems
• Capacity results
• Writing on dirty paper and extensions
Basic Model

- Message $M$ uniformly distributed in $\{1, \ldots, 2^{nR}\}$.
- State vector $S^n$ generated IID according to $p(s)$.
- Channel memoryless according to $p(y|x, s)$.
- Sets $S$, $X$, and $Y$ are finite.
**Types of side information**

1. Causal side information: $x_i$ depends only on $m$ and $s^i$.
   - Denote capacity with $C_c$.

2. Non-causal side information: $x^n$ depends on $m$ and $s^n$.
   - In particular, $x_i$ depends on $m$ and $s^n$ (the entire state sequence) for all $i$.
   - Denote capacity with $C_{nc}$.

Comments:
- $C_{nc} \geq C_c$.
- Non-causal assumption relevant for watermarking.
Comparison with last week

- Diagram of “lossy” source coding with side information.
- “Lossless” would require another encoder for $Y^n$.
- Encoder has non-causal side information.
Example 1

<table>
<thead>
<tr>
<th>State</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prob</td>
<td>$1 - p$</td>
<td>$p$</td>
</tr>
</tbody>
</table>

- $S = \mathcal{X} = \mathcal{Y} = \{0, 1\}$.
- $Y_i = X_i + S_i \mod 2$.
- $C_c = C_{nc} = 1$.
- With no side information, capacity is $1 - h(p)$. 
Example 2: Memory with defects

State: $a$ $b$ $c$
Prob: $1 - p$ $p/2$ $p/2$

- $S = \{a, b, c\}$, $\mathcal{X} = \mathcal{Y} = \{0, 1\}$.
- We will see that $C_{nc} > C_c$. 
Example 3: Writing on Dirty Paper

- $S^n$ is IID $\mathcal{N}(0, Q)$.
- $Z^n$ is IID $\mathcal{N}(0, P)$.
- $X^n$ subject to power constraint of $P$.
- Will show that $C_{nc} = \frac{1}{2} \log \left(1 + \frac{P}{N}\right)$. 

$M$ \hspace{1cm} Encoder \hspace{1cm} $X^n$ \hspace{1cm} $Y^n$ \hspace{1cm} Decoder \hspace{1cm} $\hat{M}$

- Encoder \hspace{1cm} $S^n$ \hspace{1cm} $Z^n$
Relationship with watermarking

\[ M \rightarrow \text{Encoder} \rightarrow X^n \rightarrow \text{Channel} \rightarrow Y^n \rightarrow \hat{M} \]

- \( S^n \) is original data (e.g. Led Zeppelin song)
- \( M \) is information to embed (e.g. owner ID number)
- Encoder restricted in choice of \( X^n \).
- Non-causal side information reasonable assumption.
- Might want more general model for “Channel”.

Data Generator
Other related problems

Different types of side information:
- At any combination of encoder and decoder.
- Noisy or compressed versions of state sequence.

Different state generators:
- Non-memoryless.
- Non-probabilistic – the arbitrarily varying channel.
- One probabilistic choice then fixed – the compound channel.
- Current state depending on past inputs.

Applications:
- Wireless – fading channels.
- Computer memories.
Capacity results

1. Causal case:

\[ C_c = \max_{p(u), f: U \times S \rightarrow X} I(U; Y), \]

where \( U \) is an auxiliary random variable with \(|U| \leq |Y|\) and

\[ p(s, u, x, y) = \begin{cases} 
  p(s)p(u)p(y|x, s) & \text{if } x = f(u, s) \\
  0 & \text{otherwise}
\end{cases}. \]

2. Non-causal case:

\[ C_{nc} = \max_{p(u|s), f: U \times S \rightarrow X} I(U; Y) - I(U; S), \]

where \(|U| \leq |X| + |S|\) and

\[ p(s, u, x, y) = \begin{cases} 
  p(s)p(u|s)p(y|x, s) & \text{if } x = f(u, s) \\
  0 & \text{otherwise}
\end{cases}. \]
Comments on Capacity Results

- $C_c \leq C_{nc}$.
  - If not, then we are in trouble.
  - Same objective function, but different feasible regions.

- Compare $C_{nc}$ with rate distortion region for “lossy” source coding with side information. Given $p(s, y)$,

$$R(D) = \min_{p(u|s), f:U \times Y \to S, \mathbb{E}[d(S, f(U, Y))] \leq D} I(U; Y) - I(U; S),$$

where $p(u, s, y) = p(s, y)p(u|s)$, which gives the Markov condition ($Y \ni S \ni U$).
Achievability : Causal Side Information

- Larger DMC – Input $\mathcal{X}^S$ and output $\mathcal{Y}$.
- Each input letter is a function from $\mathcal{S}$ to $\mathcal{X}$.
- Only need to use $|\mathcal{Y}|$ of the $|\mathcal{X}|^{|S|}$ input letters.
- Auxiliary RV $U$ indexes the input letters.
- Example: Memory with defects
  - $t_0(s) = 0$ for all $s$, $\Pr(Y = 0) = (1 - \epsilon)(1 - p) + p/2$.
  - $t_1(s) = 1$ for all $s$, $\Pr(Y = 1) = (1 - \epsilon)(1 - p) + p/2$.
  - Any other function from $\mathcal{S}$ to $\mathcal{X}$ gives one of these distributions on $\mathcal{Y}$.
  - $C_c = 1 - h(p/2 + \epsilon(1 - p))$. 
Converse : Causal Side Information

Let $U(i) = (M, S_{i-1})$.

- $(M, Y_{i-1}) \leftrightarrow U(i) \leftrightarrow Y_i$.
- $U(i)$ and $S_i$ are independent.
- For small probability of error:

$$n(R - \delta) \leq I(M; Y^n) \leq \sum_{i=1}^{n} I(M, Y_{i-1}; Y_i) \leq \sum_{i=1}^{n} I(U(i); Y_i) \leq nC_c.$$
Achievability: Non-causal Side Information

Use dual to binning technique from last week.

- Choose distribution $p(u|s)$ and function $f : \mathcal{U} \times \mathcal{S} \mapsto \mathcal{X}$.

- Codebook generation:
  - For each $m \in \{1, \ldots, 2^{nR}\}$, generate $U(m, 1), \ldots, U(m, 2^{nR_0})$ IID according to $p(u)$.
  - A total of $2^{n(R+R_0)}$ codewords.

- Encoding:
  - Given $m$ and $s^n$, find $u(m, j)$ jointly typical with $s^n$.
  - Set $x^n = f(u(m, j), s^n)$.

- Decoding:
  - Find $(\hat{m}, \hat{j})$ such that $u(\hat{m}, \hat{j})$ jointly typical with $y^n$. 


Achievability: Non-causal Side Information

- Encoding failure small if $R_0 > I(U;S)$
- Decoding failure small if $R + R_0 < I(U;Y)$.
  - Need Markov lemma.
- Rate achievable if $R < I(U;Y) - I(U;S)$.
- Intuition:
  - Codebook bin $\approx$ quantizer for state sequence.
  - If $I(U;S) > 0$, then use non-causal feedback non-trivially.
Example : Memory with defects

- $\mathcal{U} = \{u_0, u_1\}$, $f(u_i, s) = i$.
- Joint distribution of $S, U$ and $X$:

<table>
<thead>
<tr>
<th></th>
<th>$u_0, 0$</th>
<th>$u_1, 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>$(1-p)/2$</td>
<td>$(1-p)/2$</td>
</tr>
<tr>
<td>$b$</td>
<td>$(1-\epsilon)p/2$</td>
<td>$\epsilon p/2$</td>
</tr>
<tr>
<td>$c$</td>
<td>$\epsilon p/2$</td>
<td>$(1-\epsilon)p/2$</td>
</tr>
</tbody>
</table>

- $I(U; S) = H(U) - H(U|S) = 1 - (1-p) - p h(\epsilon) = p(1 - h(\epsilon))$.
- $I(U; Y) = H(Y) - H(Y|U) = 1 - h(\epsilon)$.
- $C_{nc} = I(U; Y) - I(U; S) = (1-p)(1 - h(\epsilon)) > C_c$.
  - Also capacity when state known at decoder.
  - Mistake in summary.
Converse: Non-causal side information

- Let $U(i) = (M, Y_1, \ldots, Y_{i-1}, S_{i+1}, \ldots, S_n)$.
- For small probability of error:
  $$n(R - \delta) \leq I(M; Y^n) - I(M; S^n)$$
  $$\leq \sum_{i=1}^{n} I(U(i); Y_i) - I(U(i); S_i)$$
  $$\leq nC_{nc}$$

- Second step: mutual information manipulations.
- Markov chain in causal case not valid here.
Writing on Dirty Paper

- $S_i \sim N(0, Q)$, $Z_i \sim N(0, N)$, $\frac{1}{n} \sum X_i^2 \leq P$.
- Costa shows $C_{nc} = \frac{1}{2} \log (1 + \frac{P}{N})$.
  - Same as if $S^n$ known to decoder.
  - Dual to Gaussian lossy source coding with side info.
Capacity for Writing on Dirty Paper

- Pick joint distribution on known noise $S$, input $X$, and auxiliary random variable $U$:
  - $X \sim \mathcal{N}(0, P)$, independent of $S$.
  - $U = X + \alpha S$
- Costa: Compute $I(U; Y) - I(U; S)$ and optimize over $\alpha$.
- New proof: Choose $\alpha = \frac{P}{P + N}$ and see what happens.
- Important properties:
  1. $X - \alpha(X + Z)$ and $X + Z$ are independent.
  2. $X - \alpha(X + Z)$ and $Y = X + S + Z$ are independent.
  3. $X$ has capacity achieving distribution for AWGN channel.
- Cannot do better than $C(P, N) = \frac{1}{2} \log \left(1 + \frac{C}{N}\right)$. 
Writing on Dirty Paper, continued

• Step 1

\[ I(U; Y) - I(U; S) = (h(U) - h(Y|U)) - (h(U) - h(U|S)) \]
\[ = h(U|S) - h(U|Y) \]

• Step 2

\[ h(U|S) = h(X + \alpha S|S) \]
\[ = h(X|S) \]
\[ = h(X) \quad X \text{ and } S \text{ independent} \]
Writing on Dirty Paper, continued

• Step 3

\[ h(U|Y) = h(X + \alpha S|Y) \]
\[ = h(X + \alpha(S - Y)|Y) \]
\[ = h(X - \alpha(X + Z)|Y) \]
\[ = h(X - \alpha(X + Z)) \quad \text{Property 2} \]
\[ = h(X - \alpha(X + Z)|X + Z) \quad \text{Property 1} \]
\[ = h(X|X + Z) \]

• Step 4

\[ I(U;Y) - I(U;S) = h(X) - h(X|X + Z) \quad \text{Steps 1, 2 & 3} \]
\[ = I(X;X + Z) \]
\[ = C(P,N) \quad \text{Property 3} \]
Extension of “Writing on Dirty Paper”

For any distributions on $S$ and $Z$, similar result if there exists $X$ such that both

- $X$ is capacity achieving for channel with additive noise $Z$.
- $X - a(X + Z)$ and $X + Z$ independent for some linear $a(\cdot)$.

In particular,

- $S$ can have any (power-limited) distribution.
- $Z$ can be colored Gaussian.
  - Capacity achieving distribution also Gaussian (waterfilling).

Similar extension given by Erez, Shamai & Zamir ’00.
Writing on Dirty Tape

What about $C_c$ for this problem?

- Only definitive result (Erez et. al.):
  \[
  \lim_{N \to 0} \lim_{Q \to \infty} C_{nc} - C_c = \frac{1}{2} \log \left( \frac{\pi e}{6} \right)
  \]

- $\frac{\pi e}{6}$ = ultimate “shaping gain”
  - Asymptotic MSE difference of vector vs. scalar quantization.

- Suggested scheme: Codewords as sequences of scalar quantizers.
  - Version of Quantization Index Modulation (Brian Chen).

- Any ideas for how to find capacity non-asymptotically?