Capacity of a POST Channel with and without Feedback

Workshop On Coding and Information Theory

The University of Hong Kong, December 2013

Tsachy Weissman
based on joint work with:
Haim Permuter
Himanshu Asnani
directed information

\[ I(X^n \rightarrow Y^n) \triangleq \sum_{i=1}^{n} I(X^i; Y_i | Y^{i-1}) \]
directed information

\[ I(X^n \rightarrow Y^n) \triangleq \sum_{i=1}^{n} I(X^i; Y_i | Y^{i-1}) \]

compare to:

\[ I(X^n; Y^n) = \sum_{i=1}^{n} I(X^n; Y_i | Y^{i-1}) \]
**Directed Information**  
[Massey90] inspired by [Marko 73]

\[ I(X^n \rightarrow Y^n) \triangleq H(Y^n) - H(Y^n \mid X^n) \]
\[ I(X^n; Y^n) \triangleq H(Y^n) - H(Y^n \mid X^n) \]

**Causal Conditioning**  
[Kramer98]

\[ H(Y^n \mid X^n) \triangleq E[- \log P(Y^n \mid X^n)] \]
\[ H(Y^n \mid X^n) \triangleq E[- \log P(Y^n \mid X^n)] \]

\[ P(y^n \mid x^n) \triangleq \prod_{i=1}^{n} P(y_i \mid x^i, y^{i-1}) \]
\[ P(y^n \mid x^{n-1}) \triangleq \prod_{i=1}^{n} P(y_i \mid x^{i-1}, y^{i-1}) \]
causal conditioning

\[
P(y^n|x^n) \triangleq \prod_{i=1}^{n} P(y_i|x^i, y^{i-1})
\]

\[
P(y^n|x^{n-1}) \triangleq \prod_{i=1}^{n} P(y_i|x^{i-1}, y^{i-1})
\]

\[
p(x^n, y^n) = p(x^n|y^{n-1})p(y^n|x^n)
\]
why directed information?

consider, e.g.:

\[ \text{BSC}(1/2) \]

\[ X_i = Y_{i-1} \]

\[ I(X^n; Y^n) =? \]

\[ I(X^n \rightarrow Y^n) =? \]
optimization

\[ \max_{p(x^n||y^{n-1})} I(X^n \rightarrow Y^n) \]

note:

- concavity of the function
- convexity of the set
on capacity

(under conditions)

\[ C = \lim_{n \to \infty} \max_{X^n \to Y^n} \frac{1}{n} I(X^n \to Y^n) \]

[Massey 1990], [Kramer 1998], [Chen and Berger 2005], [Tatikonda and Mitter 2010], [Kim 2010], [Permuter, Goldsmith and W. 2010]
Finite State Channels

$$P(y_i, s_i | x^i, s^{i-1}, y^{i-1}) = P(y_i, s_i | x_i, s_{i-1})$$
explicit computations

- memoryless channels
- mod-additive channels
- Gaussian with and without FB
- trapdoor with FB
- Ising with FB
- some more
POST channel

Previous Output is the STate

\[ S_i = Y_{i-1} \]

\[ p(y_i|x^i, y^{i-1}) = p(y_i|x_i, y_{i-1}) \]

T. Berger, 2002 Shannon lecture
“living information theory”
motivation

• simple
• good model
• to feed or not
"To Feed or Not to Feed Back"
questions for the POST channel

• $C_{FB}$
• $C_{NFB}$
• $C_{FB} > C_{NFB}$?
feedback capacity of the POST channel

[Chen and Berger 2005]:

\[ C_{FB} = \max_{p(x_1|y_0)} I(X_1; Y_1|Y_0) \]

(under benign conditions)
If $y_{i-1} = 0$, then the channel behaves as a $Z$ channel with parameter $\alpha$. If $y_{i-1} = 1$, then it behaves as an $S$ channel with parameter $\alpha$. 

![POST(\alpha) channel diagram]
(simple) \[ \text{POST}(\alpha = \frac{1}{2}) \]

\[ y_{i-1} = 0 \]

\[ x_i \quad | \quad \frac{1}{2} \quad | \quad y_i \]

\[ y_{i-1} = 1 \]

\[ x_i \quad | \quad \frac{1}{2} \quad | \quad y_i \]

alternatively:

if \( X_i = Y_{i-1} \),

otherwise,

\[ Y_i = X_i \]

\[ Y_i \sim Bernoulli(\frac{1}{2}) \]
intuition?

Regular capacity

$$C = \max_{P(x)} I(X; Y, S) = H_b\left(\frac{1}{4}\right) - \frac{1}{2} = 0.3111$$

Feedback capacity is the capacity of the Z channel

$$C_{fb} = -\log_2 0.8 = 0.3219$$
channel probing ([Asnani, Permuter and W. 2011])
POST(\(\alpha\)) channel

If \(y_{i-1} = 0\) then the channel behaves as an \(Z\) channel with parameter \(\alpha\).

If \(y_{i-1} = 1\) then it behaves as an \(S\) channel with parameter \(\alpha\).

- \(C_{FB}\)
- \(C_{NFB}\)
- \(C_{FB} > C_{NFB}\)?

\(P(X) \times P(Y)_{\text{naive}}\) small if 
\[ E_d(\hat{X}, Y) > D_X(R) + D_{\text{naive}}(D) \]

(assuming \(d\) is a metric)

\[ P_{X,Y} = P_X \times P_Y \]

\[ \mathcal{E}_Q(R, 0) \]

\[ \mathcal{E}_Q(R, n) \]
Theorem

Feedback does not increase the capacity of the POST($\alpha$) channel.
main idea

show:

\[
\max_{P(x^n || y^{n-1})} I(X^n \rightarrow Y^n) = \max_{P(x^n)} I(X^n \rightarrow Y^n)
\]
Necessary and sufficient for $\max I(X^n \rightarrow Y^n)$

**Theorem**

A set of necessary and sufficient conditions for an input probability $P(x^n||y^{n-1})$ to maximize $I(X^n \rightarrow Y^n)$ is that for some numbers $\beta_{y^{n-1}}$

$$\sum_{y_n} p(y^n|x^n) \log \frac{p(y^n|x^n)}{\text{ep}(y^n)} = \beta_{y^{n-1}}, \ \forall x^n, y^{n-1}, \text{ if } p(x^n||y^{n-1}) > 0$$

$$\sum_{y_n} p(y^n|x^n) \log \frac{p(y^n|x^n)}{\text{ep}(y^n)} \leq \beta_{y^{n-1}}, \ \forall x^n, y^{n-1}, \text{ if } p(x^n||y^{n-1}) = 0$$

where $p(y^n) = \sum_{x^n} p(y^n|x^n)p(x^n||y^{n-1})$. The solution of the optimization is

$$\max_{P(x^n||y^{n-1})} I(X^n \rightarrow Y^n) = \sum_{y^{n-1}} \beta_{y^{n-1}} + 1.$$
the case $n=1$

[Gallager 1968]

4.5 Finding Channel Capacity for a Discrete Memoryless Channel

Theorem 4.5.1. A set of necessary and sufficient conditions on an input probability vector $Q = [Q(0), \ldots, Q(K-1)]$ to achieve capacity on a discrete memoryless channel with transition probabilities $P(j \mid k)$ is that for some number $C$,

$$I(x = k; Y) = C; \quad \text{all } k \text{ with } Q(k) > 0 \quad (4.5.1)$$

$$I(x = k; Y) \leq C; \quad \text{all } k \text{ with } Q(k) = 0 \quad (4.5.2)$$

in which $I(x = k; Y)$ is the mutual information for input $k$ averaged over the outputs,

$$I(x = k; Y) = \sum_{j} P(j \mid k) \log \frac{P(j \mid k)}{\sum_{i} Q(i)P(j \mid i)} \quad (4.5.3)$$
Necessary and sufficient for \( \max I(X^n \rightarrow Y^n) \)

**Theorem**

A set of necessary and sufficient conditions for an input probability \( P(x^n||y^{n-1}) \) to maximize \( I(X^n \rightarrow Y^n) \) is that for some numbers \( \beta_{y_{n-1}} \)

\[
\sum_{y_n} p(y^n|x^n) \log \frac{p(y^n|x^n)}{ep(y^n)} = \beta_{y_{n-1}}, \quad \forall x^n, y^{n-1}, \quad \text{if} \quad p(x^n||y^{n-1}) > 0
\]

\[
\sum_{y_n} p(y^n|x^n) \log \frac{p(y^n|x^n)}{ep(y^n)} \leq \beta_{y_{n-1}}, \quad \forall x^n, y^{n-1}, \quad \text{if} \quad p(x^n||y^{n-1}) = 0
\]

where \( p(y^n) = \sum_{x^n} p(y^n|x^n)p(x^n||y^{n-1}) \). The solution of the optimization is

\[
\max_{P(x^n||y^{n-1})} I(X^n \rightarrow Y^n) = \sum_{y^{n-1}} \beta_{y_{n-1}} + 1.
\]
The main tool we use to prove the equality of the optimization problems is a corollary.

**Corollary**

Let $P^*(x^n|y^{n-1})$ achieve the maximum of

$$\max_{P(x^n|y^{n-1})} I(X^n \to Y^n)$$

and let $P^*(y^n)$ be the induced output pmf. If there exists an input probability distribution $P(x^n)$ such that

$$p^*(y^n) = \sum_{x^n} p(y^n|x^n)p(x^n),$$

then

$$\max_{P(x^n|y^{n-1})} I(X^n \to Y^n) = \max_{P(x^n)} I(X^n \to Y^n)$$
route for showing $C_{NFB} = C_{FB}$

Find:

1. $p^*(y^n)$
2. $P_n^{-1}$ (where $P_n(y^n, x^n) = P(y^n|x^n)$)
3. $P_n^{-1} \cdot p^*(y^n)$

is $P_n^{-1} \cdot p^*(y^n) \geq 0$?

if yes $\forall n \Rightarrow C_{NFB} = C_{FB}$
specifically, for POST

• simple structure and evolution of:
  • optimal output distribution
  • channel matrix
  • its inverse
Binary symmetric Markov $\{Y\}_{i \geq 1}$ with transition probability 0.2 can be described recursively

$$P_0(y^n) = \begin{bmatrix} 0.8P_0(y^{n-1}) \\ 0.2P_1(y^{n-1}) \end{bmatrix} \quad P_1(y^n) = \begin{bmatrix} 0.2P_0(y^{n-1}) \\ 0.8P_1(y^{n-1}) \end{bmatrix},$$

where $P_0(y^0) = P_1(y^0) = 1.$
Simple POST channel

Using P we obtained n,

Conditional probabilities:

\[
P(Y_1 | X_1, s_0 = 0)
\]

<table>
<thead>
<tr>
<th></th>
<th>X_1</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y_1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1/2</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1/2</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>X_1</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y_1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1/2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1/2</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

\[
P(Y_1 | X_1, s_0 = 1)
\]

\[
P_n,0 = \begin{bmatrix}
1 \cdot P_{n-1,0} & \frac{1}{2} \cdot P_{n-1,0} \\
0 \cdot P_{n-1,1} & \frac{1}{2} \cdot P_{n-1,1}
\end{bmatrix}
\]

\[
P_n,1 = \begin{bmatrix}
\frac{1}{2} \cdot P_{n-1,0} & 0 \cdot P_{n-1,0} \\
\frac{1}{2} \cdot P_{n-1,1} & 1 \cdot P_{n-1,1}
\end{bmatrix}
\]

\[
P_n^{-1} = \begin{bmatrix}
1 \cdot P_{n-1,0} & \alpha \cdot P_{n-1,0} \\
0 \cdot P_{n-1,1} & \bar{\alpha} \cdot P_{n-1,1}
\end{bmatrix}^{-1} = \begin{bmatrix}
P_{n-1,0} & -\frac{\alpha}{\bar{\alpha}} P_{n-1,1,1} \\
0 & \frac{1}{\bar{\alpha}} P_{n-1,1,1}
\end{bmatrix}
\]

\[
P_n^{-1} = \begin{bmatrix}
\bar{\alpha} \cdot P_{n-1,0} & 0 \cdot P_{n-1,0} \\
\alpha \cdot P_{n-1,1} & 1 \cdot P_{n-1,1}
\end{bmatrix}^{-1} = \begin{bmatrix}
\frac{1}{\bar{\alpha}} P_{n-1,0} & 0 \\
-\frac{\alpha}{\bar{\alpha}} P_{n-1,0,1} & P_{n-1,1,1}
\end{bmatrix}
\]
using:

\[ P_0(x^n) = P_{n,0}^{-1}P_0(y^n), \quad P_1(x^n) = P_{n,1}^{-1}P_1(y^n) \]

we obtain:

\[
\begin{align*}
P_0(x^n) &= \begin{bmatrix} 0.8P_0(x^{n-1}) - 0.2P_1(x^{n-1}) \\ 0.4P_1(x^{n-1}) \end{bmatrix} \\
P_1(x^n) &= \begin{bmatrix} 0.4P_0(x^{n-1}) \\ 0.8P_1(x^{n-1}) - 0.2P_0(x^{n-1}) \end{bmatrix}
\end{align*}
\]
Feedback does not increase capacity of POST(\(\alpha\))

The feedback and the non-feedback capacity of POST(\(\alpha\)) channel is the same as of the memoryless \(\mathcal{Z}\) channel with parameter \(\alpha\), which is \(C = -\log_2 c\) where

\[
c = \left(1 + \bar{\alpha} \alpha \frac{\alpha}{\bar{\alpha}}\right)^{-1}
\]
If \( y_{i-1} = 0 \) then the channel behaves as DMC with parameters \((a, b)\) and if \( y_{i-1} = 1 \) then the channel behaves as DMC with parameters \((b, a)\).

We are able to show numerically on a grid of resolution \(10^{-5}\times10^{-5}\) on \((a, b)\in[0,1] \times [0,1]\) that feedback does not increase the capacity.
the input distribution

\[ P_0(x^n) = \frac{1}{(a + b - 1)(\gamma + 1)} \left[ b\gamma P_0(x^{n-1}) - \bar{b} P_1(x^{n-1}) \right. \]
\[ \left. - \bar{a}\gamma P_0(x^{n-1}) + a P_1(x^{n-1}) \right] \]

\[ P_1(x^n) = \frac{1}{(a + b - 1)(\gamma + 1)} \left[ a P_0(x^{n-1}) - \bar{a}\gamma P_1(x^{n-1}) \right. \]
\[ \left. - \bar{b} P_0(x^{n-1}) + b\gamma P_1(x^{n-1}) \right] \]

\[ \gamma = 2^\frac{H(b) - H(a)}{a + b - 1}. \]
In order to prove that $P(x^n)$ is valid we needed:

- $\gamma \geq \frac{\bar{b}}{b}$
- $\gamma \leq \frac{a}{\bar{a}}$
- $\gamma \geq \frac{a}{b}$ for $a \geq \bar{b}$
- $\gamma^2 \leq \frac{a^2}{b\bar{a}}$ for $a \geq \bar{b}$
- $\frac{\gamma(\bar{a}+b)}{2b} \geq 1$ for $a \geq \bar{b}$ and $a\bar{a} \leq b\bar{b}$
- $\gamma^2(\bar{a} + b)^2 - 4a\bar{b} \geq 0$
- $\gamma(\bar{a} + b) - \sqrt{\gamma^2(\bar{a} + b)^2 - 4ab} \leq 2\bar{b}$, for $a \geq \bar{b}$ and $a\bar{a} \leq b\bar{b}$

where

$$\gamma = 2^\frac{H(b) - H(a)}{a + b - 1}.$$
Feedback does not increase capacity of a POST\((a, b)\) channel

The feedback and the non-feedback capacity of POST\((a, b)\) channel is the same as of a binary DMC channel with parameters \((a, b)\), which is given by

\[
C = \log \left[ 2 \frac{\bar{a}H_b(b) - bH_b(a)}{a+b-1} + 2 \frac{\bar{b}H_b(a) - aH_b(b)}{a+b-1} \right].
\]
can feedback help for a general POST channel?
large alphabet

\[ y_{i-1} = 1, 2, \ldots, m \]

\[ y_{i-1} = m + 1 \]

\[
\begin{align*}
x_i & \quad \bullet \quad \bullet \quad \bullet \\
m & \quad \text{---} \quad \text{---} \\
m + 1 & \quad \text{---} \quad \text{---}
\end{align*}
\]

\[
\begin{align*}
1 & \quad \text{---} \\
2 & \quad \text{---} \\
y_i & \quad \bullet \quad \bullet \\
m & \quad \text{---} \\
m + 1 & \quad \text{---}
\end{align*}
\]

\[
\begin{align*}
x_i & \quad \bullet \quad \bullet \quad \bullet \\
m & \quad \text{---} \quad \text{---} \\
m + 1 & \quad \text{---} \quad \text{---}
\end{align*}
\]

\[
\begin{align*}
1 & \quad \text{---} \\
2 & \quad \text{---} \\
y_i & \quad \bullet \quad \bullet \\
m & \quad \text{---} \\
m + 1 & \quad \text{---}
\end{align*}
\]
$$y_{i-1} = 1, 2, \ldots, m$$

$$y_{i-1} = m + 1$$

**Table:**

| $m$  | upper bound on capacity $\frac{1}{6} \max_{s_0} \max_{P(x^6)} I(X^6; Y^6 | s_0)$ | lower bound on $C_{fb}$ $R = \frac{\log_2 m}{3}$ |
|------|-------------------------------------------------|---------------------------------|
| $2^9$ | 2.5376                                         | 3.0000                          |
is the sufficient condition necessary?

- we didn’t know
- numerics in binary case were inconclusive
condition is necessary and sufficient

Theorem:

Let $C_{FB} = \max_{p(x_1|y_0)} I(X_1; Y_1|Y_0)$ and let $p^*(y_0)$ be induced by $p^*(x_1|y_0)$. $C_{FB} = C_{NFB}$ if and only if $\forall n$:

Can induce $p^*_{fb}(y^n)$ without feedback and without knowledge of $Y_0 \sim p^*(y_0)$.
implication

\[ \exists \text{ binary POST channels with } C_{FB} > C_{NFB} \]
conclusions

- $C_{FB} = C_{NFB} \iff$ can induce $p^*_{fb}(y^n)$ without feedback $\forall n$
- $C_{FB} = C_{NFB}$ for binary POST$(a, b)$ channels
- $\exists$ POST channels with $C_{FB} > C_{NFB}$
questions

• crisper necessary and sufficient condition for $C_{FB} = C_{NFB}$

• capacity achieving schemes