Directed Information Optimization and Capacity of the POST Channel with and without Feedback

Haim Permuter

Ben-Gurion University, Israel

Information Systems Laboratory Colloquium
Sep 2013
Directed Information

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

POST Channel
Previous Output is the STate

Convex Optimization
Definitions

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

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

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

**Directed Information**

[Massey90] inspired by [Marko 73]

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

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

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

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

**Directed Information**  
[Massey90] inspired by [Marko 73]

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

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

**Causal Conditioning**  
[Kramer98]

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

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

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

**Directed Information**

[Massey90] inspired by [Marko 73]

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

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

**Causal Conditioning**

[Kramer98]

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

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

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

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

**Directed Information**  
[Massey90] inspired by [Marko 73]

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

**Causal Conditioning**  
[Kramer98]

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

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

\[ P(y^n | x^{n-1}) \triangleq \prod_{i=1}^{n} P(y_i | x^{i-1}, y^{i-1}) \]
Directed information and causal conditioning characterizes:

1. rate reduction in losless compression due to causal side information at the decoder,
2. the gain in growth rate in horse-race gambling due to causal side information
3. channel capacity with feedback,
4. multi user capacity with feedback: broadcast, MAC, compound, memory-in-block networks
5. rate distortion with feedforward,
6. causal MMSE for additive Gaussian noise,
7. stock investment with causal side information,
8. measure of causal relevance between processes,
9. actions with causal constraint such as “to feed or not to feed back”,

H. Permuter  Directed Information and the POST Channel
Directed information optimization

How to find

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

Recall

\[
I(X^n \rightarrow Y^n) = \sum_{i=1}^{n} I(X^i; Y_i|Y^{i-1})
\]
\[
= H(Y^n) - H(Y^n||X^n)
\]
\[
= \sum_{y^n,x^n} p(x^n, y^n) \log \frac{p(y^n|x^n)}{p(y^n)}
\]

\(P(x^n, y^n)\) can be expressed by the chain-rule

\[
p(x^n, y^n) = p(x^n||y^{n-1})p(y^n||x^n)
\]
Lemma: causal conditioning is a polyhedron

The set of all causal conditioning distributions of the form $P(x^n||y^{n-1})$ is a polyhedron in $\mathbb{R}^{|\mathcal{X}|^n|\mathcal{Y}|^{n-1}}$ and is given by the following linear equalities and inequalities:

\[
\begin{align*}
\ p(x^n||y^{n-1}) & \geq 0, & & \forall x^n, y^{n-1}, \\
\sum_{x_{i+1}} x^n p(x^n||y^{n-1}) & = \gamma_{x^i, y^{i-1}}, & & \forall x^i, y^{n-1}, i \geq 1, \\
\sum_{x_1} x^n p(x^n||y^{n-1}) & = 1, & & \forall y^{n-1}.
\end{align*}
\]
Lemma: causal conditioning is a polyhedron

The set of all causal conditioning distributions of the form
\( P(x^n \| y^{n-1}) \) is a polyhedron in \( \mathbb{R}^{|X|^n|Y|^{n-1}} \) and is given by the following linear equalities and inequalities:

\[
\begin{align*}
p(x^n \| y^{n-1}) & \geq 0, \quad \forall x^n, y^{n-1}, \\
\sum x^i_{i+1} p(x^n \| y^{n-1}) &= \gamma_{x^i, y^{i-1}}, \quad \forall x^i, y^{n-1}, i \geq 1, \\
\sum x^1 p(x^n \| y^{n-1}) &= 1, \quad \forall y^{n-1}.
\end{align*}
\]

Lemma: concavity of directed information

For a fixed channel \( p(y^n \| x^n) \), the directed information \( I(X^n \rightarrow Y^n) \) is concave in \( p(x^n \| y^{n-1}) \).
Directed information as a functional

\[ I(X^n; Y^n) \triangleq H(Y^n) - H(Y^n | X^n) \]
Directed information as a functional

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

\[ = \sum_{y^n,x^n} Q(x^n) P(y^n | x^n) \ln \frac{P(y^n | x^n)}{\sum_{x^n} Q(x^n) P(y^n | x^n)} \]
$I(X^n; Y^n) \triangleq H(Y^n) - H(Y^n | X^n)$

$= \sum_{y^n, x^n} Q(x^n) P(y^n | x^n) \ln \frac{P(y^n | x^n)}{\sum_{x^n} Q(x^n) P(y^n | x^n)}$

$\triangleq \mathcal{I}(Q(x^n), P(y^n | x^n))$
Directed information as a functional

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

\[ = \sum_{y^n, x^n} Q(x^n)P(y^n | x^n) \ln \frac{P(y^n | x^n)}{\sum_{x^n} Q(x^n)P(y^n | x^n)} \]

\[ \triangleq \mathcal{I}(Q(x^n), P(y^n | x^n)) \]

\( Q \) - input distribution, \( P \) - channel
Directed information as a functional

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

\[
= \sum_{y^n, x^n} Q(x^n) P(y^n | x^n) \ln \frac{P(y^n | x^n)}{\sum_{x^n} Q(x^n) P(y^n | x^n)}
\]

\[
\triangleq I(Q(x^n), P(y^n | x^n))
\]

\(Q\) - input distribution, \(P\) - channel

\[
I(Q(x^n | y^{n-1}), P(y^n | x^n))
\]

\[
= \sum_{x^n, y^n} Q(x^n | y^{n-1}) P(y^n | x^n) \ln \frac{P(y^n | x^n)}{\sum_{x^n} Q(x^n | y^{n-1}) P(y^n | x^n)}
\]

H. Permuter  Directed Information and the POST Channel
Directed information as a functional

\[ I(X^n; Y^n) \triangleq H(Y^n) - H(Y^n | X^n) \]
\[ = \sum_{x^n, y^n} Q(x^n) P(y^n | x^n) \ln \frac{P(y^n | x^n)}{\sum_{x^n} Q(x^n) P(y^n | x^n)} \]
\[ \triangleq \mathcal{I}(Q(x^n), P(y^n | x^n)) \]

\( Q \) - input distribution, \( P \) - channel

\[ \mathcal{I}(Q(x^n || y^{n-1}), P(y^n || x^n)) \]
\[ = \sum_{x^n, y^n} Q(x^n || y^{n-1}) P(y^n || x^n) \ln \frac{P(y^n || x^n)}{\sum_{x^n} Q(x^n || y^{n-1}) P(y^n || x^n)} \]

Chain rule \( P(x^n, y^n) = Q(x^n || y^{n-1}) P(y^n || x^n) \)
Directed information as a functional

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

\[ = \sum_{y^n, x^n} Q(x^n) P(y^n | x^n) \ln \frac{P(y^n | x^n)}{\sum_{x^n} Q(x^n) P(y^n | x^n)} \]

\[ \triangleq \mathcal{I}(Q(x^n), P(y^n | x^n)) \]

\( Q \) - input distribution, \( P \) - channel

\[ \mathcal{I}(Q(x^n \| y^{n-1}), P(y^n \| x^n)) \]

\[ = \sum_{x^n, y^n} P(x^n, y^n) \ln \frac{P(y^n \| x^n)}{\sum_{x^n} Q(x^n \| y^{n-1}) P(y^n \| x^n)} \]

Chain rule \( P(x^n, y^n) = Q(x^n \| y^{n-1}) P(y^n \| x^n) \)
Directed information as a functional

\[ I(X^n; Y^n) \triangleq H(Y^n) - H(Y^n | X^n) \]
\[ = \sum_{y^n, x^n} Q(x^n) P(y^n | x^n) \ln \frac{P(y^n | x^n)}{\sum_{x^n} Q(x^n) P(y^n | x^n)} \]
\[ \triangleq \mathcal{I}(Q(x^n), P(y^n | x^n)) \]

Q - input distribution, P - channel

\[ \mathcal{I}(Q(x^n | y^{n-1}), P(y^n | x^n)) \]
\[ = \sum_{x^n, y^n} P(x^n, y^n) \ln \frac{P(y^n | x^n)}{\sum_{x^n} P(x^n, y^n)} \]

Chain rule \[ P(x^n, y^n) = Q(x^n | y^{n-1}) P(y^n | x^n) \]
Directed information as a functional

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

\[ = \sum_{y^n, x^n} Q(x^n)P(y^n|x^n) \ln \frac{P(y^n|x^n)}{\sum_{x^n} Q(x^n)P(y^n|x^n)} \]

\[ \triangleq \mathcal{I}(Q(x^n), P(y^n|x^n)) \]

\( Q \) - input distribution, \( P \) - channel

\[ \mathcal{I}(Q(x^n || y^{n-1}), P(y^n || x^n)) \]

\[ = \sum_{x^n, y^n} P(x^n, y^n) \ln \frac{P(y^n || x^n)}{P(y^n)} \]

Chain rule \( P(x^n, y^n) = Q(x^n || y^{n-1}) P(y^n || x^n) \)
Directed information as a functional

\[ I(X^n; Y^n) \triangleq H(Y^n) - H(Y^n|X^n) \]
\[ = \sum_{y^n, x^n} Q(x^n) P(y^n|x^n) \ln \frac{P(y^n|x^n)}{\sum_{x^n} Q(x^n) P(y^n|x^n)} \]
\[ \triangleq \mathcal{I}(Q(x^n), P(y^n|x^n)) \]

- \( Q \) - input distribution, \( P \) - channel

\[ \mathcal{I}(Q(x^n||y^{n-1}), P(y^n||x^n)) \]
\[ = \sum_{x^n, y^n} P(x^n, y^n) \ln \frac{P(y^n||x^n)}{P(y^n)} \]
\[ = H(Y^n) - H(Y^n||X^n) \]

Chain rule \( P(x^n, y^n) = Q(x^n||y^{n-1}) P(y^n||x^n) \)
Directed information as a functional

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

\[ = \sum_{y^n, x^n} Q(x^n) P(y^n | x^n) \ln \frac{P(y^n | x^n)}{\sum_{x^n} Q(x^n) P(y^n | x^n)} \]

\[ \triangleq \mathcal{I}(Q(x^n), P(y^n | x^n)) \]

\(Q\) - input distribution, \(P\) - channel

\[ \mathcal{I}(Q(x^n \| y^{n-1}), P(y^n \| x^n)) \]

\[ = \sum_{x^n, y^n} P(x^n, y^n) \ln \frac{P(y^n \| x^n)}{P(y^n)} \]

\[ = H(Y^n) - H(Y^n \| X^n) \]

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

Chain rule \( P(x^n, y^n) = Q(x^n \| y^{n-1}) P(y^n \| x^n) \)
Property of the optimization problem

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

**Good news**

- \(I(X^n \rightarrow Y^n)\) is convex in \(p(x^n||y^{n-1})\) for a fixed \(p(y^n||x^n)\).
- \(p(x^n||y^{n-1})\) is a convex set.
Property of the optimization problem

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

**Good news**

- \(I(X^n \rightarrow Y^n)\) is convex in \(p(x^n||y^{n-1})\) for a fixed \(p(y^n||x^n)\).
- \(p(x^n||y^{n-1})\) is a convex set.

**Benefits:**

- Efficient algorithm for finding the maximum.
- Necessary and sufficient conditions (KKT conditions) for having the optimum.
Property of the optimization problem

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

Good news

- \( I(X^n \rightarrow Y^n) \) is convex in \( p(x^n||y^{n-1}) \) for a fixed \( p(y^n||x^n) \).
- \( p(x^n||y^{n-1}) \) is a convex set.

Benefits:

- Efficient algorithm for finding the maximum.
- Necessary and sufficient conditions (KKT conditions) for having the optimum.

To be careful

- \( I(X^n \rightarrow Y^n) \) non-convex in \( p(x_1), \ldots, p(x_n||x^{n-1}, y^{n-1}) \)
- Cannot optimize each term in \( \sum_i I(X^i; Y_i|Y^{i-1}) \) separately.
The Alternating maximization procedure

Lemma (Double maximization)

\[
\max_{p(x^n \mid y^{n-1})} I(X^n \rightarrow Y^n) = \max_{p(x^n \mid y^{n-1}), q(x^n \mid y^n)} I(X^n \rightarrow Y^n).
\]
The Alternating maximization procedure

**Lemma (Double maximization)**

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

Let \( f(u_1, u_2) \), be a convex fun and we want to find

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

The procedure is

\[
 u_1^{(k+1)} = \arg \max_{u_1 \in A_1} f(u_1^{(k)}, u_2^{(k)}), \quad u_2^{(k+1)} = \arg \max_{u_2 \in A_2} f(u_1^{(k+1)}, u_2^{(k)}).
\]

\[
f^{(k)} = f(u_1^{(k)}, u_2^{(k)}).
\]

**Theorem (The Alternating maximization procedure)**

\[
\lim_{k \rightarrow \infty} f^{(k)} = \max_{u_1 \in A_1, u_2 \in A_2} f(u_1, u_2).
\]
Compute by the alternating maximization procedure

$$\max_{p(x^n|y^{n-1})} \max_{q(x^n|y^n)} I(X^n \rightarrow Y^n).$$
Compute by the alternating maximization procedure

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

1st step

**Lemma (max \(q(x^n|y^n) \ I(X^n \rightarrow Y^n))\)**

*For fixed \(p(x^n|y^{n-1})\), \(q^*(x^n|y^n)\) that achieves max \(q(x^n|y^n) \ I(X^n \rightarrow Y^n)\), is*

\[
q^*(x^n|y^n) = \frac{p(x^n|y^{n-1})p(y^n|x^n)}{\sum_{x^n} p(x^n|y^{n-1})p(y^n|x^n)}.
\]
Lemma \((\max_p(x^n\|y^{n-1}) \ I(X^n \to Y^n))\)

For fixed \(q(x^n|y^n), p^*(x^n\|y^{n-1})\) that achieves
\[\max_p(x^n\|y^{n-1}) \ I(X^n \to Y^n),\]
is:

starting from \(i = n\), compute \(p(x_i|x^{i-1}, y^{i-1})\)

\[
p_i = p^*(x_i|x^{i-1}, y^{i-1}) = \frac{p'(x^i, y^{i-1})}{\sum x_i p'(x^i, y^{i-1})},
\]

where

\[
p'(x^i, y^{i-1}) = \prod_{x_{i+1}^n, y_i^n} \left[ \frac{q(x^n|y^n)}{\prod_{j=i+1}^n p_j} \right] \prod_{j=i}^n p(y_j|x^j, y^{j-1}) \prod_{j=i+1}^n p_j,
\]

and do so backwards until \(i = 1\).
Main ideas of 2nd Step

- Exchange $p(x^n \parallel y^{n-1})$ by the set $\{p_i\}_{i=1}^n$ where
  
  $p_i = p(x_i \mid x^{i-1}, y^{i-1})$

  $$\max_{p(x^n \parallel y^{n-1})} I(X^n \rightarrow Y^n) = \max_{p_1} \max_{p_2} \ldots \max_{p_n} I(X^n \rightarrow Y^n)$$
Main ideas of 2nd Step

- Exchange $p(x^n \parallel y^{n-1})$ by the set $\{p_i\}_{i=1}^n$ where
  
  $$p_i = p(x_i | x^{i-1}, y^{i-1})$$

  $$\max_{p(x^n \parallel y^{n-1})} I(X^n \rightarrow Y^n) = \max_{p_1} \max_{p_2} \ldots \max_{p_n} I(X^n \rightarrow Y^n)$$

- $I(X^n \rightarrow Y^n)$ is concave in each $p_i$. 
Main ideas of 2nd Step

- Exchange $p(x^n||y^{n-1})$ by the set $\{p_i\}_{i=1}^n$ where $p_i = p(x_i|x^{i-1}, y^{i-1})$

$$\max_{p(x^n||y^{n-1})} I(X^n \rightarrow Y^n) = \max_{p_1} \max_{p_2} \ldots \max_{p_n} I(X^n \rightarrow Y^n)$$

- $I(X^n \rightarrow Y^n)$ is concave in each $p_i$.
- For fixed $q(x^n|y^n)$, $p_i^*$ that achieves $\max_{p_i} I(X^n \rightarrow Y^n)$, depends only on $q(x^n|y^n), p_i+1, p_i+2, \ldots, p_n$
Main ideas of 2nd Step

- Exchange $p(x^n\|y^{n-1})$ by the set $\{p_i\}_{i=1}^n$ where $p_i = p(x_i|x_{i-1}, y_{i-1})$

\[
\max_{p(x^n\|y^{n-1})} I(X^n \rightarrow Y^n) = \max_{p_1} \max_{p_2} \cdots \max_{p_n} I(X^n \rightarrow Y^n)
\]

- $I(X^n \rightarrow Y^n)$ is concave in each $p_i$.
- For fixed $q(x^n|y^n)$, $p^*_i$ that achieves $\max_{p_i} I(X^n \rightarrow Y^n)$, depends only on

\[
q(x^n|y^n), p_{i+1}, p_{i+2}, \ldots, p_n
\]

- Hence we can find

\[
\max_{p_1} \cdots \left( \max_{p_{n-1}} \left( \max_{p_n} I(X^n \rightarrow Y^n) \right) \right)
\]

despite being nonconvex.
How to terminate the algorithm?

- Using steps 1 and 2 we can compute

\[ I_L = \sum_{y^n, x^n} p(y^n\|x^n)r(x^n\|y^{n-1}) \log \frac{q(x^n|y^n)}{p(x^n\|y^{n-1})} \]

which converges from below to \( \max_{p(x^n\|y^{n-1})} I(X^n \rightarrow Y^n) \)

- We also have an upper bound

\[ I_U = \max_{x_1} \sum_{y_1} \max_{x_2} \cdots \sum_{y_{n-1}} \max_{x_n} \sum_{y_n} p(y^n\|x^n) \log \frac{p(y^n\|x^n)}{\sum_{x'n} p(y^n\|x'n)p(x'n\|y^{n-1})} \]

- The algorithm terminates when

\[ |I_U - I_L| \leq \epsilon \]
maximizing the directed information for BSC(0.3)

Upper bound, $I_U$

Lower bound, $I_L$

True capacity

Value

Iteration
$I(X^{n+1} \rightarrow Y^{n+1}) - I(X^n \rightarrow Y^n)$
Channels without feedback

Message $m$ is encoded as $x_i(m)$, transmitted through a memoryless channel with probability $P(y_i|x_i)$, and decoded into an estimated message $\hat{m}(y^n)$. The time-invariant function $z_i = y_{i-1}$ represents the unit delay.
Channels with feedback

Message $x_i(m, z^{i-1})$ is encoded and sent through a Memoryless Channel $P(y_i|x_i)$, and the output $y_i$ is decoded into an estimated message $\hat{m}(y^n)$. The feedback path is denoted by $z_i(y_i)$ and $z_{i-1}(y_{i-1})$. The unit delay function maps the previous output $y_{i-1}$ to $z_{i-1}(y_{i-1})$.
Channels with feedback

Message $x_i(m, z_{i-1})$

Finite State Channel

Encoder

$P(y_i, s_i | x_i, s_{i-1})$

Finite State Channel

Decoder

Estimated message $\hat{m}(y^n)$

Unit Delay

$z_{i-1}(y_{i-1})$

Time-Invariant Function

$z_i(y_i)$

$y_i$

$y_i$

$\hat{m}$
Channels with feedback

**Finite State Channel (FSC) property:**

\[
P(y_i, s_i | x_i, s_{i-1}, y_{i-1}) = P(y_i, s_i | x_i, s_{i-1})
\]
For memoryless channels we know the exact capacity:

- Binary Symmetric channel (BSC)
- Erasure channel
- Z-channel
- Arbitrary memoryless binary channel
Exact capacity computations

- For memoryless channels we know the exact capacity:
  - Binary Symmetric channel (BSC)
  - Erasure channel
  - Z-channel
  - Arbitrary memoryless binary channel
  - Numerical solutions [Blahut72, Arimoto72]
Exact capacity computations

For memoryless channels we know the exact capacity:

- Binary Symmetric channel (BSC)
- Erasure channel
- Z- channel
- Arbitrary memoryless binary channel
- Numerical solutions [Blahut72, Arimoto72]
- Feedback does not increase capacity [Shannon56]
Exact capacity computations

- For memoryless channels we know the exact capacity:
  - Binary Symmetric channel (BSC)
  - Erasure channel
  - Z-channel
  - Arbitrary memoryless binary channel
  - Numerical solutions [Blahut72, Arimoto72]
  - Feedback does not increase capacity [Shannon56]

- What about channels with memory?
Exact capacity computations

- For memoryless channels we know the exact capacity:
  - Binary Symmetric channel (BSC)
  - Erasure channel
  - Z- channel
  - Arbitrary memoryless binary channel
  - Numerical solutions [Blahut72, Arimoto72]
  - Feedback does not increase capacity [Shannon56]

- What about channels with memory?
  - Mod-2 addition channel $Y_i = X_i \oplus Z_i$, where $Z_i$ stationary.
    \[
    C = 1 - \lim_{n \to \infty} H(Z_i | Z_{i-1}) \quad [\text{with feedback, by Alajaji95}]
    \]
For memoryless channels we know the exact capacity:

- Binary Symmetric channel (BSC)
- Erasure channel
- Z- channel
- Arbitrary memoryless binary channel
- Numerical solutions [Blahut72, Arimoto72]
- Feedback does not increase capacity [Shannon56]

What about channels with memory?

- Mod-2 addition channel $Y_i = X_i \oplus Z_i$, where $Z_i$ stationary.
  \[ C = 1 - \lim_{n \to \infty} H(Z_i | Z_{i-1}) \]  
  [with feedback, by Alajaji95]

- Additive Gaussian channel $Y_i = X_i + Z_i$. [Shannon49]
Exact capacity computations

- For memoryless channels we know the exact capacity:
  - Binary Symmetric channel (BSC)
  - Erasure channel
  - Z-channel
  - Arbitrary memoryless binary channel
  - Numerical solutions [Blahut72, Arimoto72]
  - Feedback does not increase capacity [Shannon56]

- What about channels with memory?
  - Mod-2 addition channel \( Y_i = X_i \oplus Z_i \), where \( Z_i \) stationary.
    \[
    C = 1 - \lim_{n \to \infty} H(Z_i | Z_i^{i-1})
    \]
    [with feedback, by Alajaji95]
  - Additive Gaussian channel \( Y_i = X_i + Z_i \). [Shannon49]
  - Additive Gaussian ARMA channel with feedback [Kim10]
For memoryless channels we know the exact capacity:

- Binary Symmetric channel (BSC)
- Erasure channel
- Z-channel
- Arbitrary memoryless binary channel
- Numerical solutions [Blahut72, Arimoto72]
- Feedback does not increase capacity [Shannon56]

What about channels with memory?

- Mod-2 addition channel \( Y_i = X_i \oplus Z_i \), where \( Z_i \) stationary.

\[
C = 1 - \lim_{n \to \infty} H(Z_i | Z_{i-1}) \\
\text{[with feedback, by Alajaji95]}
\]

- Additive Gaussian channel \( Y_i = X_i + Z_i \). [Shannon49]
- Additive Gaussian ARMA channel with feedback [Kim10]
- State known at En and De [Goldsmith/Varaiya97, Chen/Berger05]
Exact capacity computations

- For memoryless channels we know the exact capacity:
  - Binary Symmetric channel (BSC)
  - Erasure channel
  - Z- channel
  - Arbitrary memoryless binary channel
  - Numerical solutions [Blahut72, Arimoto72]
  - Feedback does not increase capacity [Shannon56]

- What about channels with memory?
  - Mod-2 addition channel \( Y_i = X_i \oplus Z_i \), where \( Z_i \) stationary.
    \[
    C = 1 - \lim_{n \to \infty} H(Z_i | Z_{i-1})
    \]
    [with feedback, by Alajaji95]

- Additive Gaussian channel \( Y_i = X_i + Z_i \). [Shannon49]
- Additive Gaussian ARMA channel with feedback [Kim10]
- State known at En and De [Goldsmith/Varaiya97, Chen/Berger05]
- Trapdoor channel with feedback [P./Cuff/Weissman/Van-Roy10]
For memoryless channels we know the exact capacity:
- Binary Symmetric channel (BSC)
- Erasure channel
- Z-channel
- Arbitrary memoryless binary channel
- Numerical solutions [Blahut72, Arimoto72]
- Feedback does not increase capacity [Shannon56]

What about channels with memory?
- Mod-2 addition channel $Y_i = X_i \oplus Z_i$, where $Z_i$ stationary.
  \[
  C = 1 - \lim_{n \to \infty} H(Z_i | Z_i^{i-1})
  \]
  [with feedback, by Alajaji95]

- Additive Gaussian channel $Y_i = X_i + Z_i$. [Shannon49]
- Additive Gaussian ARMA channel with feedback [Kim10]
- State known at En and De [Goldsmith/Varaiya97, Chen/Berger05]
- Trapdoor channel with feedback [P./Cuff/Weissman/Van-Roy10]
- Ising channel with feedback [Elischo/P.12]
POST
Previous Output is the STate
If $y_{i-1} = 0$ then the channel behaves as an $Z$ channel with parameter $\alpha$.

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

\begin{align*}
\text{if } y_{i-1} = 0 & \quad \text{then the channel behaves as an } Z \text{ channel with parameter } \alpha \\
\text{if } y_{i-1} = 1 & \quad \text{then it behaves an } S \text{ channel with parameter } \alpha.
\end{align*}
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 an \(S\) channel with parameter \(\alpha\).

Alternatively,

if \(X_i = Y_{i-1}\), \(Y_i = X_i\)
otherwise, \(Y_i = X_i \oplus Z_i\), where \(Z_i \sim \text{Bernoulli}(\alpha)\)
Simple POST channel or \( \text{POST}(\alpha = \frac{1}{2}) \)

\[
\begin{align*}
\text{if } X_i &= Y_{i-1}, \\
\text{otherwise,} \\
Y_i &= X_i \\
Y_i &\sim \text{Bernoulli}(\frac{1}{2})
\end{align*}
\]

![Diagram of the POST channel](image)
Goals and motivation

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

\[ 0 \hspace{2cm} 0 \]

\[ \frac{1}{2} \]

\[ x_i \]

\[ 1 \hspace{2cm} 1 \]

\[ \frac{1}{2} \]

\[ y_i \]

\[ \frac{1}{2} \]

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

\[ 0 \hspace{2cm} 0 \]

\[ \frac{1}{2} \]

\[ x_i \]

\[ 1 \hspace{2cm} 1 \]

\[ \frac{1}{2} \]

\[ y_i \]

Questions

\[ \text{What is the capacity with feedback?} \]
Goals and motivation

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

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

Questions

- What is the capacity with feedback?
- What is the capacity without feedback?
Goals and motivation

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

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

\[ 1 \quad \frac{1}{2} \quad 1 \]

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

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

\[ 1 \quad \frac{1}{2} \quad 1 \]

Questions

- What is the capacity with feedback?
- What is the capacity without feedback?
- Does feedback increase capacity?
Goals and motivation

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

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

Questions
- What is the capacity with feedback?
- What is the capacity without feedback?
- Does feedback increase capacity?

Motivation
- Simple channel with memory
Goals and motivation

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

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

Questions

- What is the capacity with feedback?
- What is the capacity without feedback?
- Does feedback increase capacity?

Motivation

- Simple channel with memory
- Models writing on memory with cell interference
Questions

- What is the capacity with feedback?
- What is the capacity without feedback?
- Does feedback increase capacity?

Motivation

- Simple channel with memory
- Models writing on memory with cell interference
- “To feed or not to feed back”
Gaining intuition via a similar example

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

\[
\begin{array}{c}
0 \\
\frac{1}{2} \\
1
\end{array}
\ 
\begin{array}{c}
0 \\
\frac{1}{2} \\
1
\end{array}
\ 
\begin{array}{c}
0 \\
\frac{1}{2} \\
1
\end{array}
\ 
\begin{array}{c}
0 \\
\frac{1}{2} \\
1
\end{array}
\]

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

\[
\begin{array}{c}
0 \\
\frac{1}{2} \\
1
\end{array}
\ 
\begin{array}{c}
0 \\
\frac{1}{2} \\
1
\end{array}
\ 
\begin{array}{c}
\frac{1}{2} \\
0 \\
\frac{1}{2}
\end{array}
\ 
\begin{array}{c}
\frac{1}{2} \\
0 \\
\frac{1}{2}
\end{array}
\]

- 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

\[
C_{fb} = -\log_2 0.8 = 0.3219
\]
Capacity of FSC with feedback

**Theorem**

*For any FSC with feedback* \([P.\ &\ Weissman\ &\ Goldsmith09]\)

\[
C_{FB} \geq \frac{1}{n} \max_{x^n|z^{n-1}} P(x^n||z^{n-1}) \min_{s_0} I(X^n \rightarrow Y^n|s_0) - \frac{\log |S|}{n}
\]

\[
C_{FB} \leq \frac{1}{n} \max_{x^n|z^{n-1}} P(x^n||z^{n-1}) \max_{s_0} I(X^n \rightarrow Y^n|s_0) + \frac{\log |S|}{n}
\]

- \(I(X^n \rightarrow Y^n)\) is the *directed information*.
- \(P(x^n||z^{n-1})\) is a *causally conditioned distribution*.
- \(|S|\) is the number of states.
Theorem

Feedback does not increase the capacity of the POST(\(\alpha\)) channel.
Main result and idea

**Theorem**

Feedback does not increase the capacity of the POST(α) channel.

**Main Idea:** show that for any $n$ the two optimization problems have the same value.

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

$$\max_{P(x^n)} I(X^n \rightarrow Y^n)$$
A convex optimization problem

**Definition**

A *convex optimization problem* is of the form

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_i(x) \leq b_i \quad i = 1, \cdots, k \\
& \quad g_j(x) = 0 \quad j = 1, \cdots, l
\end{align*}
\]

where \( f_0(x) \) and \( \{ f_i(x) \}_{i=1}^k \) are convex functions, and \( \{ g_j(x) \}_{j=1}^l \) are affine.

- The problem \( \max_{P(x^n|y^{n-1})} I(X^n \rightarrow Y^n) \) is a convex optimization problem.
A convex optimization problem

Definition

A convex optimization problem is of the form

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_i(x) \leq b_i \quad i = 1, \ldots, k \\
& \quad g_j(x) = 0 \quad j = 1, \ldots, l
\end{align*}
\]

where \( f_0(x) \) and \( \{f_i(x)\}_{i=1}^k \) are convex functions, and \( \{g_j(x)\}_{j=1}^l \) are affine.

- The problem \( \max_{P(x^n|y^{n-1})} I(X^n \rightarrow Y^n) \) is a convex optimization problem.
- **Tool:** KKT conditions are sufficient and necessary conditions for a solution to be optimal.
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 \( \beta_{y^{n-1}} \)

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

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

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

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

Corollary

Let \( P^*(x^n || y^{n-1}) \) achieve the maximum of
\[
\max_{P(x^n || y^{n-1})} I(X^n \rightarrow Y^n)
\]
and let \( P^*(y^n) \) be the induced output pmf. If there exists an input probability distribution \( P(x^n) \) such that
\[
p^*(y^n) = \sum_{x^n} p(y^n || x^n) p(x^n),
\]
for any \( n \) then the feedback capacity and the nonfeedback capacity are the same.
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

Binary symmetric Markov $\{Y\}_{i \geq 1}$ with transition probability 0.2 can be described recursively

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

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

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

Conditional probabilities:

<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>$\frac{1}{2}$</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>$\frac{1}{2}$</td>
<td></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>$\frac{1}{2}$</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>$\frac{1}{2}$</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
Simple POST channel

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

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

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 obtained

\[
P_0(x^n) = \begin{bmatrix} 0.8P_0(x^{n-1}) - 0.2P_1(x^{n-1}) \\ 0.4P_1(x^{n-1}) \end{bmatrix},
\]

\[
P_1(x^n) = \begin{bmatrix} 0.4P_0(x^{n-1}) \\ 0.8P_1(x^{n-1}) - 0.2P_0(x^{n-1}) \end{bmatrix}.
\]
Main result

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 \(Z\) channel with parameter \(\alpha\), which is \(C = -\log_2 c\) where

\[c = (1 + \bar{\alpha}\alpha)^{-1}\]
POST \((a, b)\) channel

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

\[
\begin{array}{ccc}
0 & a & 0 \\
\tilde{a} & \tilde{b} & \tilde{b} \\
x_i & 1 & 1 \\
y_i & 0 & 0 \\
\end{array}
\]

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

\[
\begin{array}{ccc}
0 & b & 0 \\
\tilde{b} & \tilde{a} & \tilde{a} \\
x_i & 1 & 1 \\
y_i & 0 & 0 \\
\end{array}
\]

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 \(10^{-5} \times 10^{-5}\) on \((a, b) \in [0, 1] \times [0, 1]\) that feedback does not increase the capacity.
We were able to obtain an input distribution that attains $P^*(y^n)$,

$$P_0(x^n) = \frac{1}{(a + b - 1)(\gamma + 1)} \left[ b\gamma P_0(x^{n-1}) - \bar{b}P_1(x^{n-1}) \\
- \bar{a}\gamma P_0(x^{n-1}) + aP_1(x^{n-1}) \right],$$

$$P_1(x^n) = \frac{1}{(a + b - 1)(\gamma + 1)} \left[ aP_0(x^{n-1}) - \bar{a}\gamma P_1(x^{n-1}) \\
- \bar{b}P_0(x^{n-1}) + b\gamma P_1(x^{n-1}) \right],$$

$$\gamma = 2 \frac{H(b) - H(a)}{a + b - 1}.$$

but how to show analytically that $P_0(x^n)$ and $P_1(x^n)$ are valid.
Inequalities that we needed.

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

where

$$\gamma = 2 \frac{H(b) - H(a)}{a + b - 1}.$$
Inequalities that we needed.

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

where

$$\gamma = 2 \frac{H(b) - H(a)}{a+b-1}.$$
Main result

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

$$C = \log \left[ 2 \frac{aH_b(b) - bH_b(a)}{a+b-1} + 2 \frac{bH_b(a) - aH_b(b)}{a+b-1} \right].$$
Is there a POST channel where feedback increases capacity?
Is there a POST channel where feedback increases capacity?

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

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

H. Permuter

Directed Information and the POST Channel
Is there a POST channel where feedback increases capacity?

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

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

<table>
<thead>
<tr>
<th>( m )</th>
<th>upper bound on capacity</th>
<th>lower bound on ( C_{fb} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 2^9 )</td>
<td>( \frac{1}{6} \max_{s_0} \max_{P(x^6)} I(X^6; Y^6</td>
<td>s_0) )</td>
</tr>
<tr>
<td></td>
<td>( 2.5376 )</td>
<td>( 3.0000 )</td>
</tr>
</tbody>
</table>
Directed Information

- Directed information is a multi-letter expression
Directed Information

- Directed information is a multi letter expression
- Directed information has similar properties as mutual information
Directed Information is a multi-letter expression. Directed information has similar properties as mutual information. For any finite $n$ directed information is a computable measure due to convexity properties.
Directed Information

- Directed information is a multi-letter expression.
- Directed information has similar properties as mutual information.
- For any finite $n$, directed information is a computable measure due to convexity properties.

Channels with memory

- If we can generate $P_{fb}^*(y^n)$ using non-feedback input, then feedback does not increase capacity.
- Feedback does not increase capacity of POST$(a, b)$.

Thank you very much!
Comparing two approaches to compute

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

<table>
<thead>
<tr>
<th>Channel</th>
<th>Convex Optimization</th>
<th>Dynamic Programming</th>
</tr>
</thead>
<tbody>
<tr>
<td>Channel</td>
<td>any FSC</td>
<td>unifilar FSC</td>
</tr>
<tr>
<td>Length</td>
<td>(n \leq 15)</td>
<td>unlimited</td>
</tr>
<tr>
<td>Solution</td>
<td>exact for (n &lt; \infty)</td>
<td>approximate</td>
</tr>
<tr>
<td>Suff. cond</td>
<td>kkt conditions for (n &lt; \infty)</td>
<td>Bellman Eq. for (n = \infty)</td>
</tr>
</tbody>
</table>
Comparing two approaches to compute

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

<table>
<thead>
<tr>
<th></th>
<th>Convex Optimization</th>
<th>Dynamic Programming</th>
</tr>
</thead>
<tbody>
<tr>
<td>Channel</td>
<td>any FSC</td>
<td>unifilar FSC</td>
</tr>
<tr>
<td>Length</td>
<td>( n \leq 15 )</td>
<td>unlimited</td>
</tr>
<tr>
<td>Solution</td>
<td>exact for ( n &lt; \infty )</td>
<td>approximate</td>
</tr>
<tr>
<td>Suff. cond</td>
<td>kkt conditions for ( n &lt; \infty )</td>
<td>Bellman Eq. for ( n = \infty )</td>
</tr>
</tbody>
</table>