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

$$I(X^n \to Y^n) \triangleq \sum_{i=1}^{n} I(X^i; Y_i | Y^{i-1})$$

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]

$$
\begin{aligned}
I(X^n \to Y^n) &\triangleq H(Y^n) - H(Y^n||X^n) \\
I(X^n; Y^n) &\triangleq H(Y^n) - H(Y^n|X^n)
\end{aligned}
$$

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

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

Definitions

Directed Information [Massey90] inspired by [Marko 73]

$$\begin{aligned} I(X^n \to Y^n) &\triangleq H(Y^n) - H(Y^n||X^n) \\ I(X^n;Y^n) &\triangleq H(Y^n) - H(Y^n|X^n) \end{aligned}$$

Causal Conditioning [Kramer98]

$$\begin{aligned} H(Y^n||X^n) &\triangleq E[-\log P(Y^n||X^n)] \\ H(Y^n|X^n) &\triangleq E[-\log P(Y^n|X^n)] \end{aligned}$$

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

Definitions

Directed Information [Massey90] inspired by [Marko 73]

$$
\begin{aligned}
I(X^n \to Y^n) &\triangleq H(Y^n) - H(Y^n||X^n) \\
I(X^n; Y^n) &\triangleq H(Y^n) - H(Y^n|X^n)
\end{aligned}
$$

Causal Conditioning [Kramer98]

$$
\begin{aligned}
H(Y^n||X^n) &\triangleq E[-\log P(Y^n||X^n)] \\
H(Y^n|X^n) &\triangleq E[-\log P(Y^n|X^n)]
\end{aligned}
$$

$$
\begin{aligned}
P(y^n||x^n) &\triangleq \prod_{i=1}^{n} P(y_i|x^i, y^{i-1}) \\
P(y^n|x^n) &= \prod_{i=1}^{n} P(y_i|x^n, y^{i-1})
\end{aligned}
$$

Definitions

Directed Information [Massey90] inspired by [Marko 73]

$$I(X^n \to Y^n) \triangleq H(Y^n) - H(Y^n||X^n)$$
$$I(X^n; Y^n) \triangleq H(Y^n) - H(Y^n|X^n)$$

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 lossless 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”,

Directed information optimization

How to find

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

Recall

$$\begin{aligned} I(X^n \to Y^n) &= \sum_{i=1}^{n} I(X^i; Y_i | Y^{i-1}) \\ &= H(Y^n) - H(Y^n||X^n) \\ &= \sum_{y^n, x^n} p(x^n, y^n) \log \frac{p(y^n||x^n)}{p(y^n)} \end{aligned}$$

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

$$p(x^n, y^n) = p(x^n||y^{n-1}) p(y^n||x^n)$$

Convexity of directed information causal conditioning

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:

$$\begin{array}{ll} p(x^n||y^{n-1}) \geq 0, & \forall x^n, y^{n-1}, \\ \sum_{x_{i+1}^n} p(x^n||y^{n-1}) = \gamma_{x^i,y^{i-1}}, & \forall x^i, y^{n-1}, i \geq 1, \\ \sum_{x_1^n} p(x^n||y^{n-1}) = 1, & \forall y^{n-1}. \end{array}$$

Convexity of directed information causal conditioning

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:

$$\begin{aligned} &p(x^n||y^{n-1}) \geq 0, && \forall x^n, y^{n-1}, \\ &\textstyle\sum_{x_{i+1}^n} p(x^n||y^{n-1}) = \gamma_{x^i,y^{i-1}}, && \forall x^i, y^{n-1}, i \geq 1, \\ &\textstyle\sum_{x_1^n} p(x^n||y^{n-1}) = 1, && \forall y^{n-1}. \end{aligned}$$

Lemma: concavity of directed information

For a fixed channel $p(y^n||x^n)$, the directed information $I(X^n \to Y^n)$ 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

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

Directed information as a functional

$$
\begin{aligned}
I(X^n;Y^n) &\triangleq H(Y^n) - H(Y^n|X^n) \\
&= \sum_{y^n,x^n} Q(x^n)P(y^n|x^n) \ln \frac{P(y^n|x^n)}{\sum_{x^n} Q(x^n)P(y^n|x^n)} \\
&\triangleq \mathcal{I}(Q(x^n), P(y^n|x^n))
\end{aligned}
$$

Directed information as a functional

$$
\begin{aligned}
I(X^n;Y^n) &\triangleq H(Y^n) - H(Y^n|X^n) \\
&= \sum_{y^n,x^n} Q(x^n)P(y^n|x^n) \ln \frac{P(y^n|x^n)}{\sum_{x^n} Q(x^n)P(y^n|x^n)} \\
&\triangleq \mathcal{I}(Q(x^n), P(y^n|x^n))
\end{aligned}
$$

Q - input distribution, P- channel

Directed information as a functional

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

Q - input distribution, P- channel

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

Directed information as a functional

$$
\begin{aligned}
I(X^n;Y^n) &\triangleq H(Y^n) - H(Y^n|X^n) \\
&= \sum_{y^n,x^n} Q(x^n)P(y^n|x^n)\ln\frac{P(y^n|x^n)}{\sum_{x^n} Q(x^n)P(y^n|x^n)} \\
&\triangleq \mathcal{I}(Q(x^n), P(y^n|x^n))
\end{aligned}
$$

Q - input distribution, P- channel

$$
\begin{aligned}
&\mathcal{I}(Q(x^n||y^{n-1}), P(y^n||x^n)) \\
&= \sum_{x^n,y^n} Q(x^n||y^{n-1})P(y^n||x^n)\ln\frac{P(y^n||x^n)}{\sum_{x^n} Q(x^n||y^{n-1})P(y^n||x^n)}
\end{aligned}
$$

Chain rule $P(x^n, y^n) = Q(x^n||y^{n-1})P(y^n||x^n)$

Directed information as a functional

$$
\begin{aligned}
I(X^n;Y^n) &\triangleq H(Y^n) - H(Y^n|X^n) \\
&= \sum_{y^n,x^n} Q(x^n)P(y^n|x^n) \ln \frac{P(y^n|x^n)}{\sum_{x^n} Q(x^n)P(y^n|x^n)} \\
&\triangleq \mathcal{I}(Q(x^n), P(y^n|x^n))
\end{aligned}
$$

Q - input distribution, P- channel

$$
\begin{aligned}
&\mathcal{I}(Q(x^n||y^{n-1}), P(y^n||x^n)) \\
&= \sum_{x^n,y^n} P(x^n,y^n) \ln \frac{P(y^n||x^n)}{\sum_{x^n} Q(x^n||y^{n-1})P(y^n||x^n)}
\end{aligned}
$$

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

$$
\begin{aligned}
I(X^n;Y^n) &\triangleq H(Y^n) - H(Y^n|X^n) \\
&= \sum_{y^n,x^n} Q(x^n)P(y^n|x^n)\ln\frac{P(y^n|x^n)}{\sum_{x^n} Q(x^n)P(y^n|x^n)} \\
&\triangleq \mathcal{I}(Q(x^n),P(y^n|x^n))
\end{aligned}
$$

Q - input distribution, P- channel

$$
\begin{aligned}
&\mathcal{I}(Q(x^n||y^{n-1}),P(y^n||x^n)) \\
&= \sum_{x^n,y^n} P(x^n,y^n)\ln\frac{P(y^n||x^n)}{P(y^n)}
\end{aligned}
$$

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

$$
\begin{aligned}
I(X^n;Y^n) &\triangleq H(Y^n) - H(Y^n|X^n) \\
&= \sum_{y^n,x^n} Q(x^n)P(y^n|x^n) \ln \frac{P(y^n|x^n)}{\sum_{x^n} Q(x^n)P(y^n|x^n)} \\
&\triangleq \mathcal{I}(Q(x^n), P(y^n|x^n))
\end{aligned}
$$

Q - input distribution, P- channel

$$
\begin{aligned}
&\mathcal{I}(Q(x^n||y^{n-1}), P(y^n||x^n)) \\
&= \sum_{x^n,y^n} P(x^n,y^n) \ln \frac{P(y^n||x^n)}{P(y^n)} \\
&= H(Y^n) - H(Y^n||X^n) \\
&= I(X^n \to Y^n)
\end{aligned}
$$

Chain rule $P(x^n,y^n) = Q(x^n||y^{n-1})P(y^n||x^n)$

Property of the optimization problem

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

Good news

- $I(X^n \to Y^n)$ 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

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

Good news

- $I(X^n \to Y^n)$ 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

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

Good news

- $I(X^n \to Y^n)$ 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 carefull

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

The Alternating maximization procedure

Lemma (Double maximization)

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

The Alternating maximization procedure

Lemma (Double maximization)

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

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

$$\max_{u_1\in\mathcal{A}_1,u_2\in\mathcal{A}_2} f(u_1, u_2).$$

The procedure is

$$u_1^{(k+1)} = \arg\max_{u_1\in\mathcal{A}_1} f(u_1^{(k)}, u_2^{(k)}),\ u_2^{(k+1)} = \arg\max_{u_2\in\mathcal{A}_2} f(u_1^{(k+1)}, u_2^{(k)}).$$

$$f^{(k)} = f(u_1^{(k)}, u_2^{(k)}).$$

Theorem (The Alternating maximization procedure)

$$\lim_{k\to\infty} f^{(k)} = \max_{u_1\in\mathcal{A}_1,u_2\in\mathcal{A}_2} f(u_1, u_2).$$

BA for directed information

Compute by the alternating maximization procedure

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

BA for directed information

Compute by the alternating maximization procedure

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

1st step

Lemma ($\max_{q(x^n|y^n)} I(X^n \to Y^n)$)

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

$$q^*(x^n|y^n) = \frac{p(x^n\|y^{n-1})p(y^n\|x^n)}{\sum_{x^n} p(x^n\|y^{n-1})p(y^n\|x^n)}.$$

2nd Step

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

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

Main ideas of 2nd Step

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

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

- $I(X^n \to Y^n)$ 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 $p_i = p(x_i | x^{i-1}, y^{i-1})$

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

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

$$q(x^n | y^n), p_{i+1}, p_{i+2}, ..., p_n$$

Main ideas of 2nd Step

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

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

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

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

- Hence we can find

$$\max_{p_1} \ldots \left(\max_{p_{n-1}} \left(\max_{p_n} I(X^n \to Y^n) \right) \right)$$

despite being nonconvex.

How to terminate the algorithm?

- Using steps 1 and 2 we can compute

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

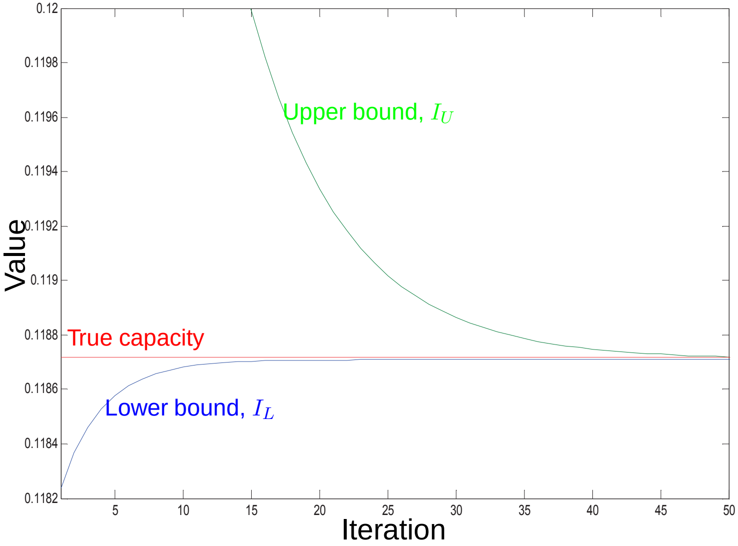
  which converges from below to $\max_{p(x^n \| y^{n-1})} I(X^n \to Y^n)$
- We also have an upper bound

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

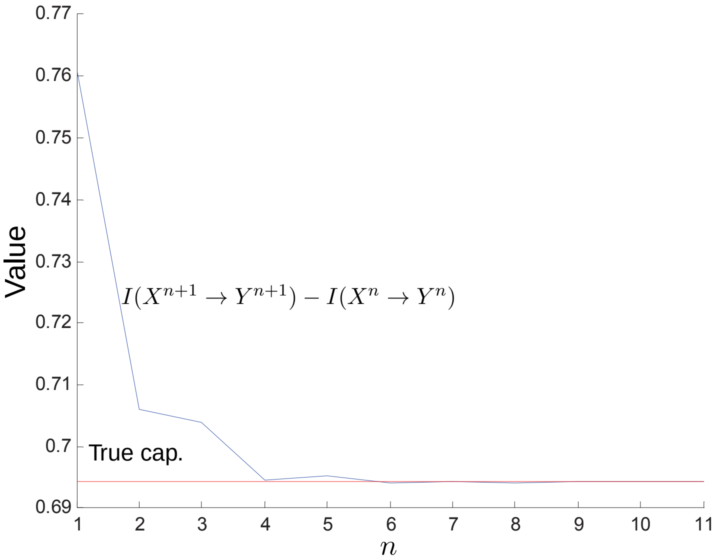
- The algorithm terminates when

$$|I_U - I_L| \leq \epsilon$$

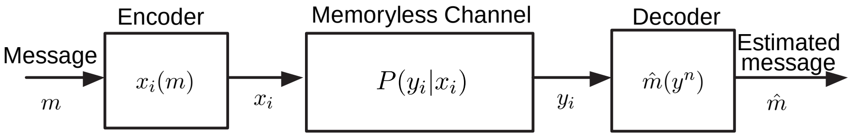
maximizing the directed information for BSC(0.3)



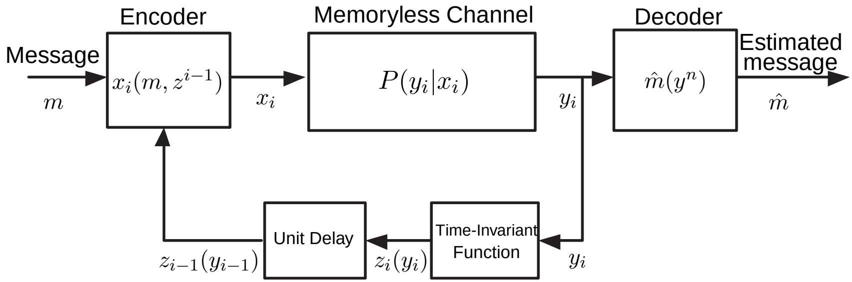
Directed information rate



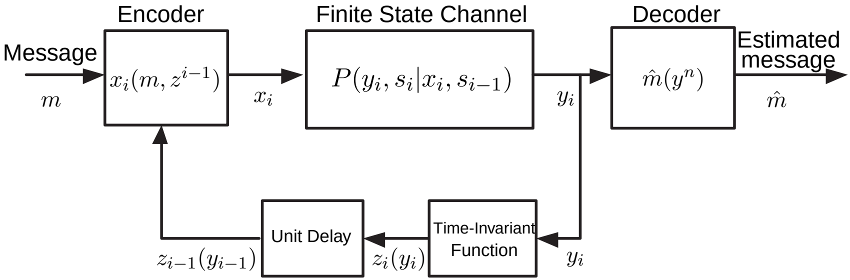
Channels without feedback



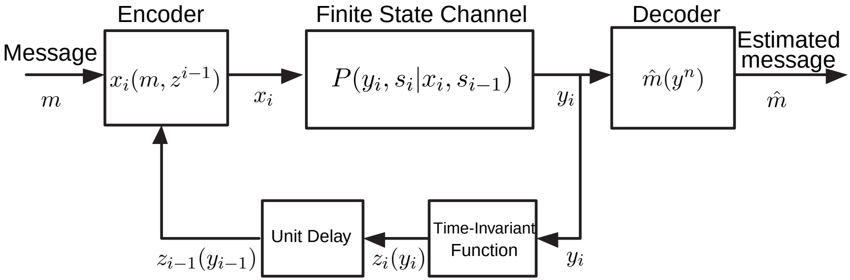
Channels with feedback



Channels with feedback



Channels with feedback



Finite State Channel(FSC) property:

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

Exact capacity computations

- 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.

$$C = 1 - \lim_{n\to\infty} H(Z_i|Z^{i-1})$$

[with feedback, by Alajaji95]

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]

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]

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]

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]

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]

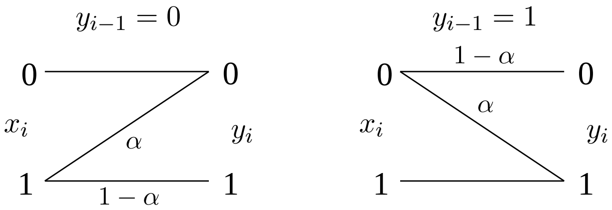
  - Additive Gaussian channel $Y_i = X_i + Z_i$. [Shannon49]
  - Additive Gaussian ARMA channel with feedback [Kim10]
  - State known at En and De [Goldsmith/Varaiya97, Chen/Berger05]
  - Trapdoor channel with feedback [P./Cuff/Weissman/Van-Roy10]
  - Ising channel with feedback [Elischo/P.12]

POST

Previous Output is the STate

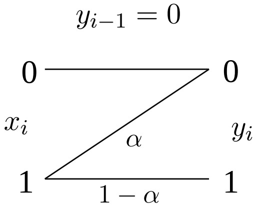
POST(α) channel

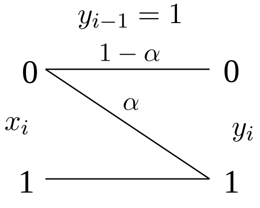
- If $y_{i-1} = 0$ then the channel behaves as an Z channel with parameter α
- If $y_{i-1} = 1$ then it behaves an S channel with parameter α.



POST(α) channel

- If $y_{i-1} = 0$ then the channel behaves as an Z channel with parameter α
- If $y_{i-1} = 1$ then it behaves an S channel with parameter α.





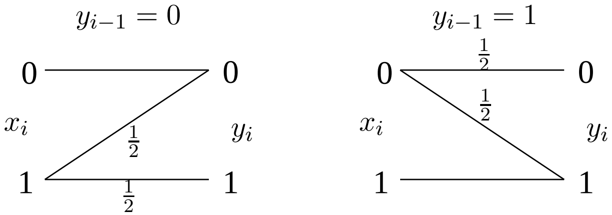
Alternatively,

$$\text{if } X_i = Y_{i-1}, \quad Y_i = X_i$$
$$\text{otherwise,} \quad Y_i = X_i \oplus Z_i, \text{ where } Z_i \sim \text{Bernnouli}\,(\alpha)$$

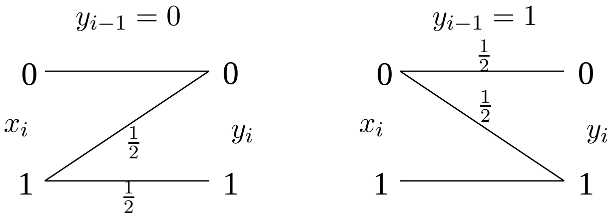
Simple POST channel or POST($\alpha = \frac{1}{2}$)

if $X_i = Y_{i-1}$, $\quad Y_i = X_i$

otherwise, $\quad Y_i \sim Bernouli(\frac{1}{2})$



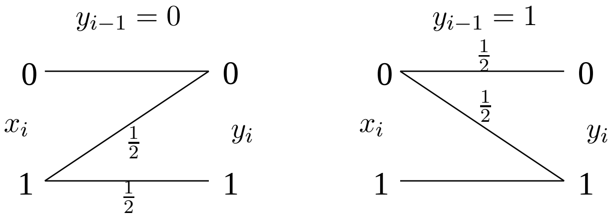
Goals and motivation



Questions

- What is the capacity with feedback?

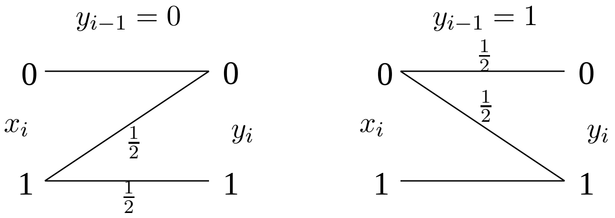
Goals and motivation



Questions

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

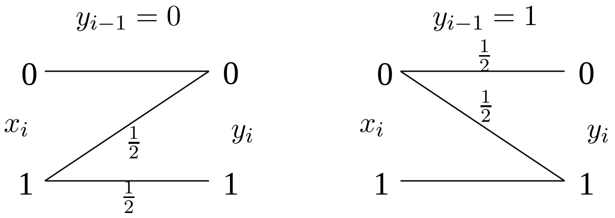
Goals and motivation



Questions

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

Goals and motivation



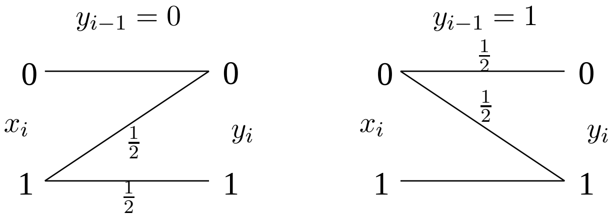
Questions

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

Motivation

- Simple channel with memory

Goals and motivation



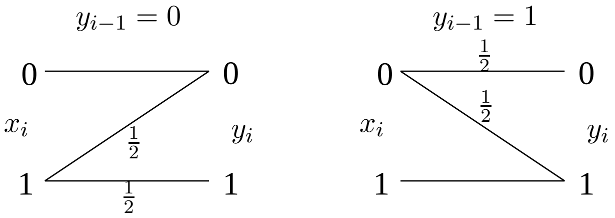
Questions

- 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

Goals and motivation



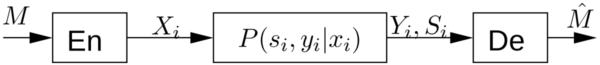
Questions

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

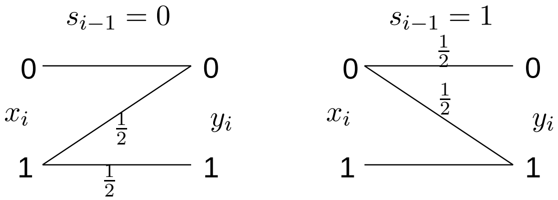
Motivation

- 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 i.i.d. $\sim$Bernoulli($\frac{1}{2}$)



- Regular capacity

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

- 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]*

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

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

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

Main result and idea

Theorem

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

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

$$\max_{P(x^n)} I(X^n \to Y^n)$$

A convex optimization problem

Definition

A *convex optimization problem* is of the form

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

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

A convex optimization problem

Definition

A *convex optimization problem* is of the form

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \le b_i \quad i = 1, \cdots, k \\ & g_j(x) = 0 \quad j = 1, \cdots, l \end{aligned}$$

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

Necessary and sufficient for $\max I(X^n \to Y^n)$

Theorem

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

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

Main corollary we use to prove equality of the optimization problems

Corollary

Let $P^*(x^n||y^{n-1})$ achieve the maximum of $\max_{P(x^n||y^{n-1})} I(X^n \to Y^n)$ and let $P^*(y^n)$ be the induced output pmf. If there exists an input probability distribution $P(x^n)$ such that

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

for any n then the feedback capacity and the nonfeedback capacity are the same.

Simple POST channel

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

$$P_0(y^n) = \left[\begin{array}{c} 0.8P_0(y^{n-1}) \\ 0.2P_1(y^{n-1}) \end{array}\right] \qquad P_1(y^n) = \left[\begin{array}{c} 0.2P_0(y^{n-1}) \\ 0.8P_1(y^{n-1}) \end{array}\right],$$

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

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

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

Conditional probabilities:

$P(Y_1|X_1, s_0 = 0)$

Y_1 \ X_1	0	1
0	1	$\frac{1}{2}$
1	0	$\frac{1}{2}$

$P(Y_1|X_1, s_0 = 1)$

Y_1 \ X_1	0	1
0	$\frac{1}{2}$	0
1	$\frac{1}{2}$	1

Simple POST channel

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

$$P_{n,0} = \begin{bmatrix} 1 \cdot P_{n-1,0} & \frac{1}{2} \cdot P_{n-1,0} \\ 0 \cdot P_{n-1,1} & \frac{1}{2} \cdot P_{n-1,1} \end{bmatrix} \; P_{n,1} = \begin{bmatrix} \frac{1}{2} \cdot P_{n-1,0} & 0 \cdot P_{n-1,0} \\ \frac{1}{2} \cdot P_{n-1,1} & 1 \cdot P_{n-1,1} \end{bmatrix}$$

Using

$$P_0(x^n) = P_{n,0}^{-1} P_0(y^n), \;\; P_1(x^n) = P_{n,1}^{-1} P_1(y^n)$$

we obtained

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

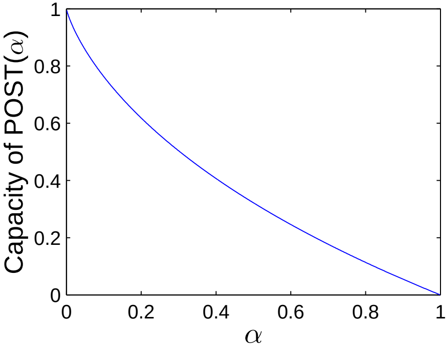
$$P_1(x^n) = \begin{bmatrix} 0.4P_0(x^{n-1}) \\ 0.8P_1(x^{n-1}) - 0.2P_0(x^{n-1}) \end{bmatrix}.$$

Main result

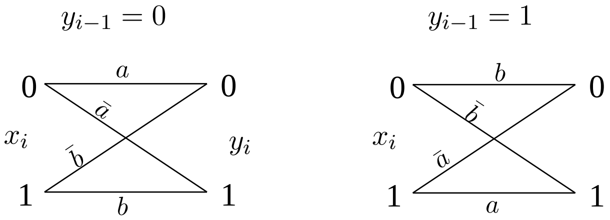
Feedback does not increase capacity of POST(α)

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

$$c = (1 + \bar{\alpha}\alpha^{\frac{\alpha}{\bar{\alpha}}})^{-1}$$



POST(a, b) channel



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.

Difficulty

We where able to obtain an input distribution that attains $P^*(y^n)$,

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

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

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

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

where

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

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

where

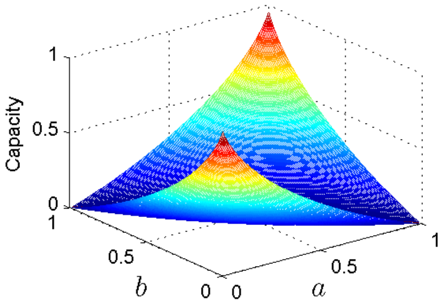
$$\gamma = 2^{\frac{H(b)-H(a)}{a+b-1}}.$$

Main result

Feedback does not increase capacity of a POST(a, b) channel

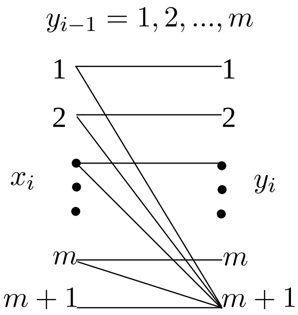
The feedback and the non-feedback capacity of POST(a, b) channel is the same as of a binary DMC channel with parameters (a, b), which is given by

$$C = \log\left[2^{\frac{\bar{a}H_b(b)-bH_b(a)}{a+b-1}} + 2^{\frac{\bar{b}H_b(a)-aH_b(b)}{a+b-1}}\right].$$

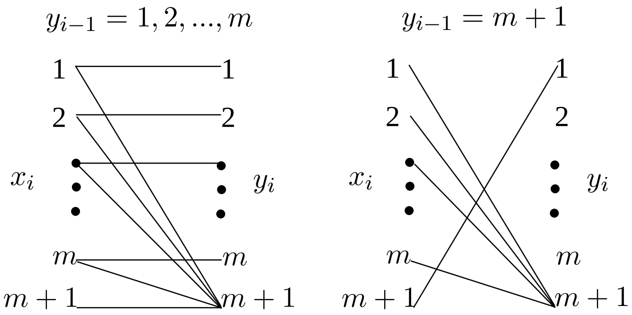


Is there a POST channel where feedback increases capacity?

Is there a POST channel where feedback increases capacity?



Is there a POST channel where feedback increases capacity?



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

Summary

Directed Information

- Directed information is a multi letter expression

Summary

Directed Information

- Directed information is a multi letter expression
- Directed information has similar properties as mutual information

Summary

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

Summary

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!

Convex Optimization vs Dynamic Programming

Comparing two approaches to compute

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

	Convex Optimization	Dynamic Programming
Channel	any FSC	unifilar FSC
Length	$n \leq 15$	unlimited
Solution	exact for $n < \infty$	approximate
Suff. cond	kkt conditions for $n < \infty$	Bellman Eq. for $n = \infty$

Convex Optimization vs Dynamic Programming

Comparing two approaches to compute

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

	Convex Optimization	Dynamic Programming
Channel	any FSC	unifilar FSC
Length	$n \leq 15$	unlimited
Solution	exact for $n < \infty$	approximate
Suff. cond	kkt conditions for $n < \infty$	Bellman Eq. for $n = \infty$