The capacity of the $(1, \infty)$-RLL Input-Constrained Erasure Channel with Feedback

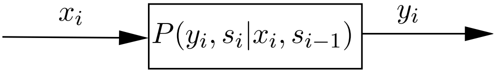
Oron Sabag, Haim Permuter and Navin Kashyap

Ben Gurion University and Indian Institute of Science

CAM16, Hong-Kong

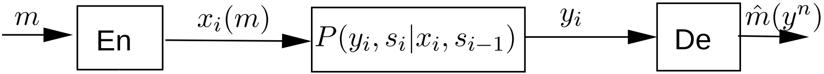
Main Ingredients

1. Channel with memory



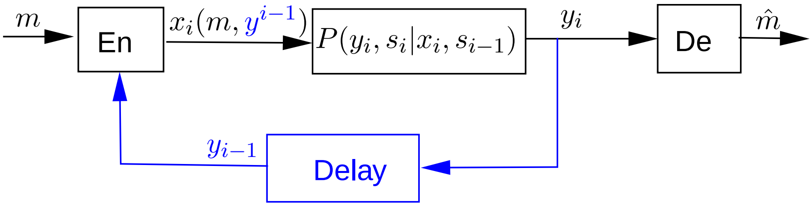
Main Ingredients

1. Channel with memory



Main Ingredients

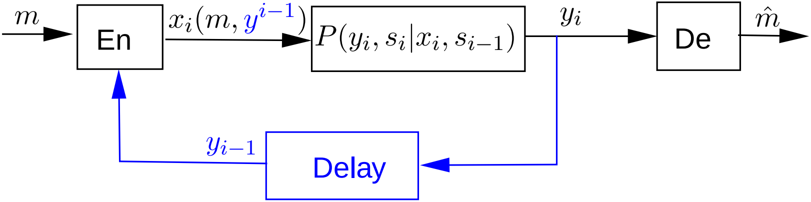
1. Channel with memory



2. Feedback

Main Ingredients

1. Channel with memory



2. Feedback
3. Directed information

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

Definition of directed information

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

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

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

$$X^n \triangleq (X_1, X_2, X_3, ..., X_n)$$

Definition of directed information

Directed Information [Massey90]

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

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

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

$$X^n \triangleq (X_1, X_2, X_3, ..., X_n)$$

Definition of directed information

Directed Information [Massey90]

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

Causal Conditioning [Kramer98]

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

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

$$
X^n \triangleq (X_1, X_2, X_3, ..., X_n)
$$

Definition of directed information

Directed Information [Massey90]

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

Causal Conditioning [Kramer98]

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

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

$$X^n \triangleq (X_1, X_2, X_3, ..., X_n)$$

Definition of directed information

Directed Information [Massey90]

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

Causal Conditioning [Kramer98]

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

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

$$X^n \triangleq (X_1, X_2, X_3, ..., X_n)$$

Feedback capacity of FSC

Theorem

For any FSC <u>with</u> feedback *[P.&Weissman&Goldsmith09]*

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

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

Feedback capacity of FSC

Theorem

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

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

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

- $\max_{P(x^n||y^{n-1})} I(X^n \to Y^n)$ can be computed using an extension of Blahut-Arimoto algorithm. [Naiss.&P.13]

Feedback capacity of FSC

Theorem

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

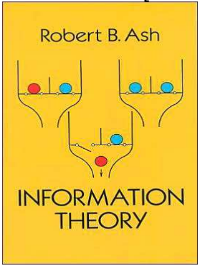
$$C_{FB} \geq \frac{1}{n} \max_{P(x^n||y^{n-1})} \min_{s_0} I(X^n \to Y^n|s_0) - \frac{\log|\mathcal{S}|}{n}$$

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

- $\max_{P(x^n||y^{n-1})} I(X^n \to Y^n)$ can be computed using an extension of Blahut-Arimoto algorithm. [Naiss.&P.13]
- Under mild conditions

$$C_{FB} = \lim_{n\to\infty} \frac{1}{n} \max_{P(x^n||y^{n-1})} I(X^n \to Y^n)$$

Trapdoor Channel [Blackwell61]



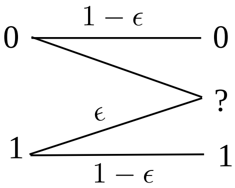
Ising Channel [Berger90]

$$y_i = \begin{cases} x_i, & \text{with prob. } \frac{1}{2} \\ x_{i-1}, & \text{with prob. } \frac{1}{2} \end{cases}$$

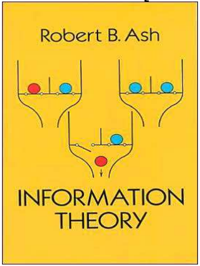
Dicode Erasure Channel [Pfister08]

$$y_i = \begin{cases} x_i - x_{i-1}, & \text{with prob. } 1-\epsilon \\ ?, & \text{with prob. } \epsilon \end{cases}$$

Erasure Channel with no repeated 1's



Trapdoor Channel [Blackwell61]



$$C_{fb} = \log \phi, \ \ \phi = \frac{\sqrt{5}+1}{2}$$

Ising Channel [Berger90]

$$y_i = \begin{cases} x_i, & \text{with prob. } \frac{1}{2} \\ x_{i-1}, & \text{with prob. } \frac{1}{2} \end{cases}$$

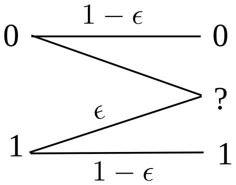
$$C_{fb} = \max_p \frac{2H_2(p)}{3+p} \approx 0.575$$

Dicode Erasure Channel [Pfister08]

$$y_i = \begin{cases} x_i - x_{i-1}, & \text{with prob. } 1-\epsilon \\ ?, & \text{with prob. } \epsilon \end{cases}$$

$$C_{fb} = \max_p (1-\epsilon) \frac{p+\epsilon H_2(p)}{\epsilon + (1-\epsilon)p}$$

Erasure Channel with no repeated 1's



$$C_{fb} = \max_p \frac{H_2(p)}{p + \frac{1}{1-\epsilon}}$$

Feedback capacity computation

For a unifilar channel we have

$$C_{FB} = \lim_{n\to\infty} \frac{1}{n} \max_{Q(x^n||y^{n-1})} I(X^n \to Y^n)$$

Feedback capacity computation

For a unifilar channel we have

$$C_{FB} = \lim_{n\to\infty} \frac{1}{n} \max_{Q(x^n||y^{n-1})} I(X^n \to Y^n)$$

$$= \sup_{\{Q(x_t|s_{t-1},y^{t-1})\}_{t\geq 1}} \liminf_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} I(X_t, S_{t-1}; Y_t|Y^{t-1})$$

Feedback capacity computation

For a unifilar channel we have

$$C_{FB} = \lim_{n\to\infty} \frac{1}{n} \max_{Q(x^n||y^{n-1})} I(X^n \to Y^n)$$
$$= \sup_{\{Q(x_t|s_{t-1},y^{t-1})\}_{t\geq 1}} \liminf_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} I(X_t, S_{t-1}; Y_t|Y^{t-1})$$

This reminds an infinite-horizon average-reward dynamic programming.

Feedback capacity computation

For a unifilar channel we have

$$C_{FB} = \lim_{n\to\infty} \frac{1}{n} \max_{Q(x^n||y^{n-1})} I(X^n \to Y^n)$$
$$= \sup_{\{Q(x_t|s_{t-1},y^{t-1})\}_{t\geq 1}} \liminf_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} I(X_t, S_{t-1}; Y_t|Y^{t-1})$$

This reminds an infinite-horizon average-reward dynamic programming.

Feedback capacity computation

For a unifilar channel we have

$$C_{FB} = \lim_{n\to\infty} \frac{1}{n} \max_{Q(x^n||y^{n-1})} I(X^n \to Y^n)$$

$$= \sup_{\{Q(x_t|s_{t-1},y^{t-1})\}_{t\geq 1}} \liminf_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} I(X_t, S_{t-1}; Y_t|Y^{t-1})$$

This reminds an infinite-horizon average-reward dynamic programming.

Feedback capacity computation

For a unifilar channel we have

$$C_{FB} = \lim_{n\to\infty} \frac{1}{n} \max_{Q(x^n||y^{n-1})} I(X^n \to Y^n)$$
$$= \sup_{\{Q(x_t|s_{t-1},y^{t-1})\}_{t\geq 1}} \liminf_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} I(X_t, S_{t-1}; Y_t | Y^{t-1})$$

This reminds an infinite-horizon average-reward dynamic programming.

Dynamic programming (DP) formulation

- state: β_{t-1}
- action: a_t
- disturbance: w_t

$$P(w_t|\beta^{t-1}, w^{t-1}, a^t) = P(w_t|\beta_{t-1}, a_t); \qquad \beta_t = F(\beta_{t-1}, a_t, w_t)$$

- reward per unit time:

$$g(\beta_{t-1}, a_t)$$

- objective:

$$\sup_{\{a_t\}_{t\geq 1}} \liminf_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \mathbb{E}\left[g(\beta_{t-1}, a_t)\right]$$

Dynamic programming (DP) formulation

- state: $\beta_{t-1} = p(s_{t-1}|y^{t-1}), \quad \beta \in [0,1]$
- action: a_t
- disturbance: w_t

$$P(w_t|\beta^{t-1}, w^{t-1}, a^t) = P(w_t|\beta_{t-1}, a_t); \qquad \beta_t = F(\beta_{t-1}, a_t, w_t)$$

- reward per unit time:

$$g(\beta_{t-1}, a_t)$$

- objective:

$$\sup_{\{a_t\}_{t\geq 1}} \lim \inf_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \mathbb{E}\left[g(\beta_{t-1}, a_t)\right]$$

Dynamic programming (DP) formulation

- state: $\beta_{t-1} = p(s_{t-1}|y^{t-1}), \quad \beta \in [0,1]$
- action: $a_t = p(x_t|s_{t-1}), \quad a_t \in [0,1]$
- disturbance: w_t

$$P(w_t|\beta^{t-1}, w^{t-1}, a^t) = P(w_t|\beta_{t-1}, a_t); \qquad \beta_t = F(\beta_{t-1}, a_t, w_t)$$

- reward per unit time:

$$g(\beta_{t-1}, a_t)$$

- objective:

$$\sup_{\{a_t\}_{t\geq 1}} \liminf_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \mathbb{E}\left[g(\beta_{t-1}, a_t)\right]$$

Dynamic programming (DP) formulation

- state: $\beta_{t-1} = p(s_{t-1}|y^{t-1}), \quad \beta \in [0,1]$
- action: $a_t = p(x_t|s_{t-1}), \quad a_t \in [0,1]$
- disturbance: $w_t = y_{t-1}, \quad w_t \in \{0,1,?\}$

$$P(w_t|\beta^{t-1}, w^{t-1}, a^t) = P(w_t|\beta_{t-1}, a_t); \qquad \beta_t = F(\beta_{t-1}, a_t, w_t)$$

- reward per unit time:

$$g(\beta_{t-1}, a_t)$$

- objective:

$$\sup_{\{a_t\}_{t\geq 1}} \lim\inf_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \mathbb{E}\left[g(\beta_{t-1}, a_t)\right]$$

Dynamic programming (DP) formulation

- state: $\beta_{t-1} = p(s_{t-1}|y^{t-1}), \quad \beta \in [0,1]$
- action: $a_t = p(x_t|s_{t-1}), \quad a_t \in [0,1]$
- disturbance: $w_t = y_{t-1}, \quad w_t \in \{0,1,?\}$

$$P(w_t|\beta^{t-1}, w^{t-1}, a^t) = P(w_t|\beta_{t-1}, a_t); \qquad \beta_t = F(\beta_{t-1}, a_t, w_t)$$

- reward per unit time:

$$g(\beta_{t-1}, a_t) = I(X_t, S_{t-1}; Y_t|y^{t-1}).$$

- objective:

$$\sup_{\{a_t\}_{t\geq 1}} \lim \inf_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \mathbb{E}\left[g(\beta_{t-1}, a_t)\right]$$

Dynamic programming (DP) formulation

- state: $\beta_{t-1} = p(s_{t-1}|y^{t-1}), \quad \beta \in [0,1]$
- action: $a_t = p(x_t|s_{t-1}), \quad a_t \in [0,1]$
- disturbance: $w_t = y_{t-1}, \quad w_t \in \{0,1,?\}$

$$P(w_t|\beta^{t-1}, w^{t-1}, a^t) = P(w_t|\beta_{t-1}, a_t); \qquad \beta_t = F(\beta_{t-1}, a_t, w_t)$$

- reward per unit time:

$$g(\beta_{t-1}, a_t) = I(X_t, S_{t-1}; Y_t|y^{t-1}).$$

- objective:

$$\sup_{\{a_t\}_{t\geq 1}} \lim \inf_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \mathbb{E}\left[g(\beta_{t-1}, a_t)\right]$$

$$= \sup_{\{p(x_t|s_{t-1}, y^{t-1})\}_{t\geq 1}} \liminf_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} I(X_t, S_{t-1}; Y_t|Y^{t-1}) = C_{FB}$$

Dynamic programming operator, T

- The dynamic programming operator T is given by

$$(T \circ J)(\beta) = \sup_{a \in \mathcal{A}} \left(g(\beta, a) + \sum_{w} P(w|\beta, a) J(F(\beta, a, w)) \right)$$

Dynamic programming operator, T

- The dynamic programming operator T is given by

$$(T \circ J)(\beta) = \sup_{a \in \mathcal{A}} \left(g(\beta, a) + \sum_{w} P(w|\beta, a) J(F(\beta, a, w)) \right)$$

- In our case: $J : [0,1] \to \mathbb{R}$ and

$$(T \circ J)(\beta) = \sup_{0 \leq \delta \leq z} \bar{\epsilon} H_b(\delta) + (1-\delta)\bar{\epsilon} J(1) + \epsilon J(1-\delta) + \delta \bar{\epsilon} J(0)$$

Dynamic programming operator, T

- The dynamic programming operator T is given by

$$(T \circ J)(\beta) = \sup_{a \in \mathcal{A}} \left(g(\beta, a) + \sum_{w} P(w|\beta, a) J(F(\beta, a, w)) \right)$$

- In our case: $J : [0,1] \to \mathbb{R}$ and

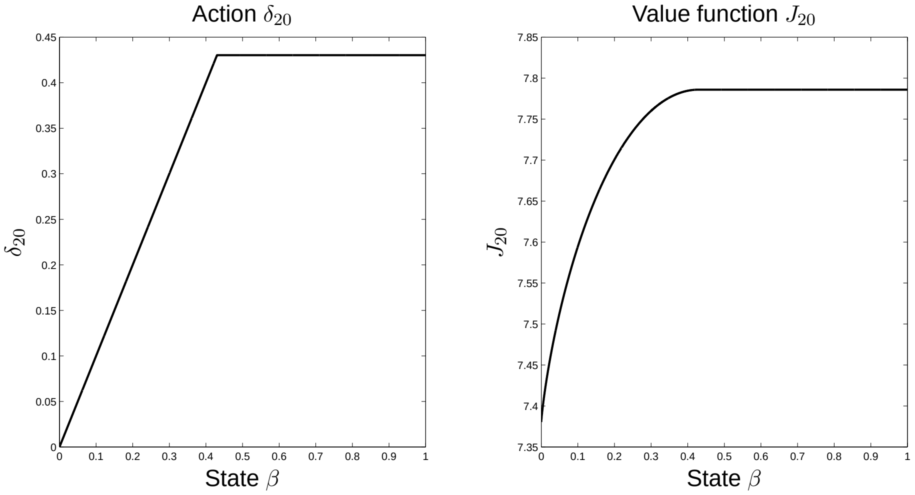
$$(T \circ J)(\beta) = \sup_{0 \le \delta \le z} \bar{\epsilon} H_b(\delta) + (1-\delta)\bar{\epsilon} J(1) + \epsilon J(1-\delta) + \delta \bar{\epsilon} J(0)$$

- Value iteration algorithms: executed 20 iterations

$$J_{k+1} = T \circ J_k$$

$$C_{FB} \approx 0.405 \text{ bits}$$

Result of value iteration



Dynamic programming- Bellman equation

Theorem (Bellman Equation)

If there exist a function $J(\beta)$ and a constant ρ that satisfy

$$T \circ J(\beta) = J(\beta) + \rho$$

then ρ is the optimal infinite horizon average reward.

Dynamic programming- Bellman equation

Theorem (Bellman Equation)

If there exist a function $J(\beta)$ and a constant ρ that satisfy

$$T \circ J(\beta) = J(\beta) + \rho$$

then ρ is the optimal infinite horizon average reward.

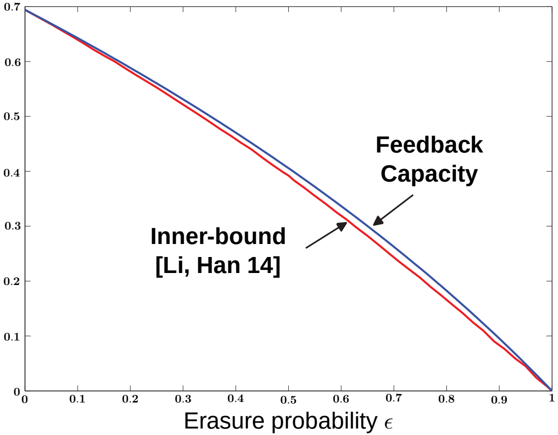
$$J^*(z) = \begin{cases} \bar{\epsilon} H_b(z) - z\bar{\epsilon}\frac{H_b(p_\epsilon)}{p_\epsilon + \frac{1}{1-\epsilon}} & \text{if } 0 \leq z \leq p_\epsilon \\ \frac{H_b(p_\epsilon)}{p_\epsilon + \frac{1}{1-\epsilon}} & \text{if } p_\epsilon \leq z \leq 1. \end{cases}$$

$$\rho^*_\epsilon = \max_{0 \leq p \leq 1} \frac{H_b(p)}{p + \frac{1}{1-\epsilon}},$$

We showed that $J^*(z)$ and ρ^*_ϵ solve the Bellman equation.

Non-Feedback capacity

Comparing to achievable rate of non-feedback [Li, Han14].



Dynamic programming- Bellman equation

Theorem (Bellman Equation)

If there exist a function $J(\beta)$ and a constant ρ that satisfy

$$T \circ J(\beta) = J(\beta) + \rho$$

then ρ is the optimal infinite horizon average reward.

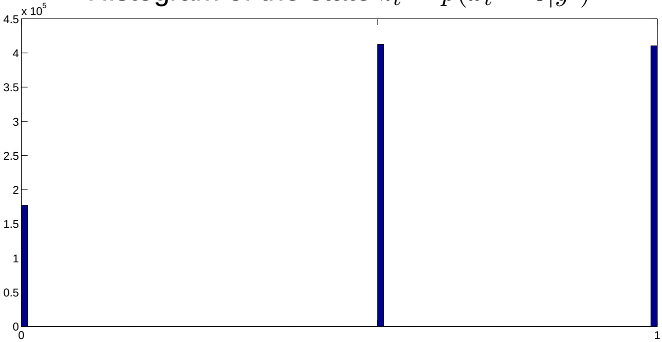
$$J^*(z) = \begin{cases} \bar{\epsilon} H_b(z) - z\bar{\epsilon}\frac{H_b(p_\epsilon)}{p_\epsilon + \frac{1}{1-\epsilon}} & \text{if } 0 \leq z \leq p_\epsilon \\ \frac{H_b(p_\epsilon)}{p_\epsilon + \frac{1}{1-\epsilon}} & \text{if } p_\epsilon \leq z \leq 1. \end{cases}$$

$$\rho^*_\epsilon = \max_{0 \leq p \leq 1} \frac{H_b(p)}{p + \frac{1}{1-\epsilon}},$$

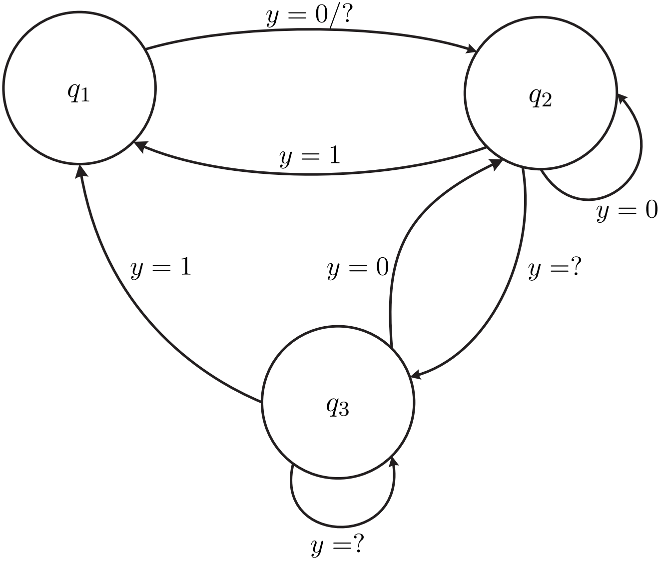
We showed that $J^*(z)$ and ρ^*_ϵ solve the Bellman equation.

Result of Value iteration

Histogram of the state $z_t \triangleq p(x_t = 0|y^t)$



The DP optimal policy [Sabag/P./Kashyap15]



Upper bound on capacity

Upper bound on capacity

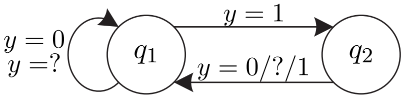
$$I(X^n \to Y^n) = \sum_i I(X_i, S_{i-1}; Y_i | Y^{i-1})$$

Upper bound on capacity

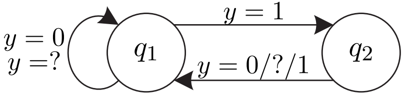
$$I(X^n \to Y^n) = \sum_i I(X_i, S_{i-1}; Y_i | Y^{i-1}) \leq \sum I(X_i, S_{i-1}; Y_i | Q_{i-1}(Y^{i-1})).$$

Upper bound on capacity

$$I(X^n \to Y^n) = \sum_i I(X_i, S_{i-1}; Y_i | Y^{i-1})$$
$$\leq \sum I(X_i, S_{i-1}; Y_i | Q_{i-1}(Y^{i-1})).$$



$$I(X^n \to Y^n) = \sum_i I(X_i, S_{i-1}; Y_i | Y^{i-1})$$
$$\leq \sum I(X_i, S_{i-1}; Y_i | Q_{i-1}(Y^{i-1})).$$



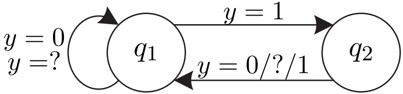
Theorem [Sabag/P./Pfiser16]

The feedback capacity of a unifilar FSC is bounded by

$$C_{fb} \leq \sup_{p(x|s,q)} I(X, S; Y|Q), \quad \forall Q\text{-graph}$$

Upper bound on capacity

$$I(X^n \to Y^n) = \sum_i I(X_i, S_{i-1}; Y_i | Y^{i-1})$$
$$\leq \sum I(X_i, S_{i-1}; Y_i | Q_{i-1}(Y^{i-1})).$$



Theorem [Sabag/P./Pfiser16]

The feedback capacity of a unifilar FSC is bounded by

$$C_{fb} \leq \sup_{p(x|s,q)} I(X, S; Y|Q), \quad \forall Q\text{-graph}$$

For BEC with no repeated 1's

$$C_{fb} \leq \max_p \frac{H_2(p)}{\frac{1}{1-\epsilon} + p}$$

Summary and future work

- Obtained the capacity of fundamental feedback problem.

Summary and future work

- Obtained the capacity of fundamental feedback problem.
- Used directed information and DP.

Summary and future work

- Obtained the capacity of fundamental feedback problem.
- Used directed information and DP.
- Obtained optimal code from the DP solution.

Summary and future work

- Obtained the capacity of fundamental feedback problem.
- Used directed information and DP.
- Obtained optimal code from the DP solution.
- Obtained single letter upper bound that is tight for all known cases.

Summary and future work

- Obtained the capacity of fundamental feedback problem.
- Used directed information and DP.
- Obtained optimal code from the DP solution.
- Obtained single letter upper bound that is tight for all known cases.
- **Future goal:** find a unified solution for all FSC with feedback.

Summary and future work

- Obtained the capacity of fundamental feedback problem.
- Used directed information and DP.
- Obtained optimal code from the DP solution.
- Obtained single letter upper bound that is tight for all known cases.
- **Future goal:** find a unified solution for all FSC with feedback.

Thank you !

Sufficient condition

- The channel state estimation (DP state):

$$p(s_t|y^t) = \frac{p(s_t, y_t|y^{t-1})}{p(y_t|y^{t-1})}$$
$$= \frac{\sum_{x_t, s_{t-1}} p(s_t, y_t, x_t, s_{t-1}|y^{t-1})}{\sum_{x_t, s_{t-1}} p(y_t, x_t, s_{t-1}|y^{t-1})}.$$

Sufficient condition

- The channel state estimation (DP state):

$$p(s_t|y^t) = \frac{p(s_t, y_t|y^{t-1})}{p(y_t|y^{t-1})}$$
$$= \frac{\sum_{x_t, s_{t-1}} p(s_t, y_t, x_t, s_{t-1}|y^{t-1})}{\sum_{x_t, s_{t-1}} p(y_t, x_t, s_{t-1}|y^{t-1})}.$$

- This mapping is denoted by $\text{BCJR} : \mathcal{Z} \times \mathcal{Y} \rightarrow \mathcal{Z}$.

Sufficient condition

- The channel state estimation (DP state):

$$p(s_t|y^t) = \frac{p(s_t, y_t|y^{t-1})}{p(y_t|y^{t-1})}$$
$$= \frac{\sum_{x_t,s_{t-1}} p(s_t, y_t, x_t, s_{t-1}|y^{t-1})}{\sum_{x_t,s_{t-1}} p(y_t, x_t, s_{t-1}|y^{t-1})}.$$

- This mapping is denoted by BCJR : $\mathcal{Z} \times \mathcal{Y} \to \mathcal{Z}$.

Sufficient condition

- The channel state estimation (DP state):

$$p(s_t|y^t) = \frac{p(s_t, y_t|y^{t-1})}{p(y_t|y^{t-1})}$$
$$= \frac{\sum_{x_t, s_{t-1}} p(s_t, y_t, x_t, s_{t-1}|y^{t-1})}{\sum_{x_t, s_{t-1}} p(y_t, x_t, s_{t-1}|y^{t-1})}.$$

- This mapping is denoted by BCJR : $\mathcal{Z} \times \mathcal{Y} \to \mathcal{Z}$.

Sufficient condition

- The channel state estimation (DP state):

$$p(s_t|y^t) = \frac{p(s_t, y_t|y^{t-1})}{p(y_t|y^{t-1})}$$
$$= \frac{\sum_{x_t, s_{t-1}} p(s_t, y_t, x_t, s_{t-1}|y^{t-1})}{\sum_{x_t, s_{t-1}} p(y_t, x_t, s_{t-1}|y^{t-1})}.$$

- This mapping is denoted by BCJR : $\mathcal{Z} \times \mathcal{Y} \rightarrow \mathcal{Z}$.

Sufficient condition

- The channel state estimation (DP state):

$$p(s_t|y^t) = \frac{p(s_t, y_t|y^{t-1})}{p(y_t|y^{t-1})}$$
$$= \frac{\sum_{x_t,s_{t-1}} p(s_t, y_t, x_t, s_{t-1}|y^{t-1})}{\sum_{x_t,s_{t-1}} p(y_t, x_t, s_{t-1}|y^{t-1})}.$$

- This mapping is denoted by $\mathrm{BCJR} : \mathcal{Z} \times \mathcal{Y} \to \mathcal{Z}$.
- If $|\mathcal{Z}| \leq \infty$ and $P_{X|S,Z}$ satisfies the BCJR mapping, we call it *BCJR-invariant*.

Sufficient condition

- The channel state estimation (DP state):

$$p(s_t|y^t) = \frac{p(s_t, y_t|y^{t-1})}{p(y_t|y^{t-1})}$$
$$= \frac{\sum_{x_t, s_{t-1}} p(s_t, y_t, x_t, s_{t-1}|y^{t-1})}{\sum_{x_t, s_{t-1}} p(y_t, x_t, s_{t-1}|y^{t-1})}.$$

- This mapping is denoted by BCJR : $\mathcal{Z} \times \mathcal{Y} \to \mathcal{Z}$.
- If $|\mathcal{Z}| \leq \infty$ and $P_{X|S,Z}$ satisfies the BCJR mapping, we call it *BCJR-invariant*.

Theorem (Lower bound)

The feedback capacity satisfies

$$C_{\text{fb}} \geq I(X, S; Y|Q),$$

for all BCJR-invariant inputs.

Upper bound with sufficient condition

Theorem [Sabag/P./Pfiser16]

The feedback capacity of a unifilar FSC is bounded by

$$C_{fb} \leq \max_{p(x|s,q)} I(X,S;Y|Q), \quad \forall Q\text{-graph}$$

and if $p^*(x|s,q)$ is BCJR-invariant input, equality holds.