Fixed-finite delay decoding of i.i.d. sequences

Aditya Mahajan
McGill University

Information Theory and Applications (ITA) Workshop

Motivation

Block Markov coding schemes

Used to derive achievable schemes for multi-user info theory

- MAC with feedback
- Broadcast with feedback
- Relay

Share common features

- i.i.d. messages at each block
- decoding done with a fixed (block) delay

Motivation

- Proof techniques to derive structural results in teams

Usual way to prove structural properties:

- Person-by-person optimality
- Coordinator with common information

Both proceed by **backward induction**. Are there situations where backward induction does not work and one has to use a **forward induction** argument?

Motivation

- Proof techniques to derive structural results in teams

  Our proof technique

  - Genie-aided nodes that know additional information.
  - Forward induction to **discard irrelevant variables**

What is the optimal structure (at block level) of block Markov coding schemes?

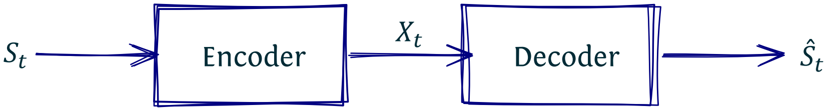
A new proof technique to prove structural results in teams.

Outline

- Point-to-point communication
  - Source coding
  - Channel coding
  - Channel coding with feedback
- Two-transmitters one-receiver setup
  - ...
- One-transmitter two-receivers
  - ...

Point-to-point communication

Source Coding Model



i.i.d. source	$\{S_1, S_2, \ldots\}$
Sequential encoder	$X_t = f_t(S_{1:t}, X_{1:t\text{-}1})$
Sequential decoder	$\hat{S}_t = g_t(X_{1:t})$
Fixed-finite delay	$\rho_t(S_{t\text{-}d}, \hat{S}_t)$

Objective: Choose $f = (f_1, \ldots, f_T)$ and $g = (g_1, \ldots, g_T)$ to minimize

$$J(f, g) = \mathbb{E}\Big[\sum_{t=d+1}^{T} \rho_t(S_{t\text{-}d+1}, \hat{S}_t)\Big]$$

Source Coding Model

Literature Overview (delay = 0)

- i.i.d. source, delay = 0 (Ericson '79)

$$X_t = f_t(S_t)$$

- Markov source, delay = 0 (Witsenhausen '79, Walrand-Varaiya '82)

$$X_t = f_t(S_t, M_{t-1}) \quad \text{and} \quad X_t = f_t(S_t, \mathbb{P}(S_t \mid X_{1:t-1}))$$

Source Coding Model

Literature Overview (delay > 0)

- Markov source, delay > 0 (Witsenhausen '79)

$$X_t = f_t(S_{t\text{-}d:t}, M_{t\text{-}1})$$

- i.i.d. source, delay > 0 (Asnani Wiessman '11)

$$X_t = f_t(S_{t\text{-}d:t}, \mathbb{P}(S_{t\text{-}d:t} \mid X_{1:t\text{-}1}))$$

Source Coding Model

Literature Overview (delay > 0)

- Markov source, delay > 0 (Witsenhausen '79)

$$X_t = f_t(S_{t\text{-}d:t}, M_{t\text{-}1})$$

- i.i.d. source, delay > 0 (Asnani Wiessman '11)

$$X_t = f_t(S_{t\text{-}d:t}, \mathbb{P}(S_{t\text{-}d:t} \mid X_{1:t\text{-}1}))$$

Our result

$$X_t = f_t(S_{t\text{-}d:t}, X_{t\text{-}d:t\text{-}1}) \quad \text{and} \quad \hat{S}_t = g_t(X_{t\text{-}d:t})$$

A detour to MDPs

Markov decision processes

$$X_{t+1} = f_t(X_t, U_t, W_t) \quad \text{and} \quad U_t = g_t(X_{1:t}, U_{1:t\text{-}1})$$

$$\min_g \mathbb{E}^g \Big[\sum_{t=1}^{T} c_t(X_t, U_t) \Big]$$

A detour to MDPs

- Markov decision processes

$$X_{t+1} = f_t(X_t, U_t, W_t) \quad \text{and} \quad U_t = g_t(X_{1:t}, U_{1:t\text{-}1})$$

$$\min_g \mathbb{E}^g\Big[\sum_{t=1}^{T} c_t(X_t, U_t)\Big]$$

- Usual proof

A detour to MDPs

Markov decision processes

$$X_{t+1} = f_t(X_t, U_t, W_t) \quad \text{and} \quad U_t = g_t(X_{1:t}, U_{1:t-1})$$

$$\min_g \mathbb{E}^g \Big[\sum_{t=1}^{T} c_t(X_t, U_t) \Big]$$

Usual proof

Sequentially show that

$$U_T = g_T(X_T)$$

A detour to MDPs

Markov decision processes

$$X_{t+1} = f_t(X_t, U_t, W_t) \quad \text{and} \quad U_t = g_t(X_{1:t}, U_{1:t\text{-}1})$$

$$\min_g \mathbb{E}^g \Big[\sum_{t=1}^{T} c_t(X_t, U_t) \Big]$$

Usual proof

Sequentially show that

$$U_T = g_T(X_T)$$

$$U_{T\text{-}1} = g_{T\text{-}1}(X_{T\text{-}1})$$

A detour to MDPs

Markov decision processes

$$X_{t+1} = f_t(X_t, U_t, W_t) \quad \text{and} \quad U_t = g_t(X_{1:t}, U_{1:t-1})$$

$$\min_g \mathbb{E}^g \Big[\sum_{t=1}^{T} c_t(X_t, U_t) \Big]$$

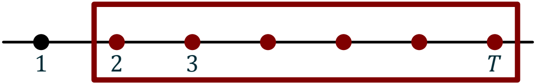
Usual proof

Sequentially show that

$$\begin{aligned} U_T &= g_T(X_T) \\ U_{T-1} &= g_{T-1}(X_{T-1}) \\ \cdots &= \cdots \\ U_1 &= g_1(X_1) \end{aligned}$$

An alternative proof idea

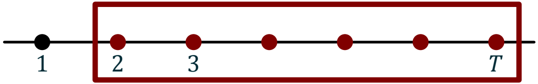
At time 2, consider a **coordinator** for all controllers from 2 up to T.



A detour to MDPs

An alternative proof idea

At time 2, consider a **coordinator** for all controllers from 2 up to T.



Common info (X_1, U_1, X_2)

Control action $\gamma^{(2)} = (\gamma_2^{(2)}, \gamma_3^{(2)}, \dots, \gamma_T^{(2)}) = h^{(2)}(X_1, U_1, X_2)$

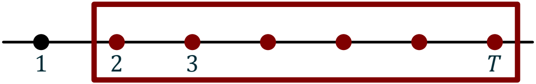
where $\gamma_t^{(2)}(\cdot) = g_t(X_{1:2}, U_1, \cdot) : X_{3:t} \mapsto U_t$

Cost at coordinator $\tilde{c}(X_2, \gamma^{(2)})$

A detour to MDPs

An alternative proof idea

At time 2, consider a **coordinator** for all controllers from 2 up to T.

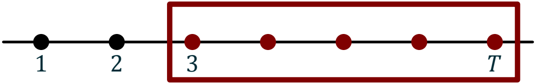


Without loss of optimality $\gamma^{(2)} = h^{(2)}(X_2)$, $\implies$

$$U_t = g_t(X_{2:t}, U_{2:t\text{-}1}), \quad \forall t > 2$$

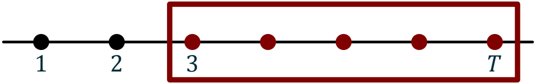
An alternative proof idea

Repeat the argument for $t = 3$



An alternative proof idea

Repeat the argument for $t = 3$



Common info (X_2, U_2, X_3)

Control action $\gamma^{(3)} = (\gamma_3^{(3)}, \gamma_4^{(3)}, \dots, \gamma_T^{(3)}) = h^{(3)}(X_1, U_1, X_2)$

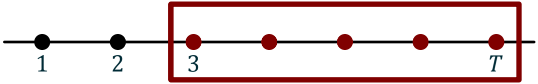
where $\gamma_t^{(3)}(\cdot) = g_t(X_{2:3}, U_2, \cdot) : X_{3:t}, U_{3:t\text{-}1} \mapsto U_3$

Cost at coordinator $\tilde{c}(X_3, \gamma^{(3)})$

A detour to MDPs

An alternative proof idea

Repeat the argument for $t = 3$



Without loss of optimality $\gamma^{(3)} = h^{(3)}(X_3)$, $\implies$

$$U_t = g_t(X_{3:t}, U_{3:t\text{-}1}), \quad \forall t > 3$$

A detour to MDPs

An alternative proof idea

Proceeding this way,

$$U_t = g_t(X_t), \quad \forall t$$

A detour to MDPs

An alternative proof idea

Proceeding this way,

$$U_t = g_t(X_t), \quad \forall t$$

> Use this **forward induction** argument to prove the structural result for the source coding setup

Structure of Optimal Encoder and Decoder

- Information structure of the system

S_t $\boxed{S_{1:t}, X_{1:t-1}}$ X_t

Encoder

X_t $\boxed{X_{1:t}}$ $\hat{S}_t$

Decoder

Structure of Optimal Encoder and Decoder

- Information structure of the system

$$S_t \rightarrow \boxed{S_{1:t}, X_{1:t\text{-}1}} \rightarrow X_t \qquad X_t \rightarrow \boxed{X_{1:t}} \rightarrow \hat{S}_t$$

Encoder Decoder

- Genie aided decoder

$$S_t \rightarrow \boxed{S_{1:t}, X_{1:t\text{-}1}} \rightarrow X_t \qquad X_t \rightarrow \boxed{S_{1:t\text{-}d\text{-}1}, X_{1:t}} \rightarrow \hat{S}_t$$

Encoder Decoder

Structure of Optimal Encoder and Decoder

- Coordinator for $[d + 2, T]$

$$S_t \;\rightarrow\; \boxed{S_{1:t}, X_{1:t\text{-}1}} \;\rightarrow\; X_t \qquad\qquad X_t \;\rightarrow\; \boxed{S_{1:t\text{-}d\text{-}1}, X_{1:t}} \;\rightarrow\; \hat{S}_t$$

Encoder — Decoder

Structure of Optimal Encoder and Decoder

Coordinator for $[d + 2, T]$

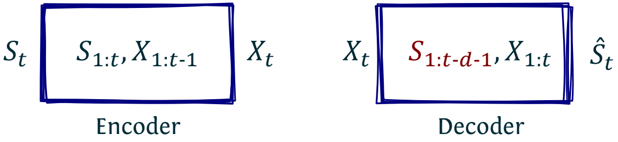
S_t $\boxed{S_{1:t}, X_{1:t\text{-}1}}$ X_t X_t $\boxed{S_{1:t\text{-}d\text{-}1}, X_{1:t}}$ $\hat{S}_t$

Encoder Decoder

- Common info $(S_1, X_{1:d+1})$
- Control action: appropriate **partial functions** $\gamma^{(d+2)}$
- Cost at coordinator (depends on $\mathbb{P}(S_{2:T}, \hat{S}_{d+2:T})$)

$$\tilde{c}(X_{2:d+2}, \gamma^{(d+2)})$$

Coordinator for $[d+2, T]$



Without loss of optimality, $\gamma^{(d+2)} = h^{(d+2)}(X_{2:d+2})$.
All nodes in the coordinator may discard (S_1, X_1).

- Coordinator for $[d + 2, T]$

S_t $\boxed{S_{2:t}, X_{2:t\text{-}1}}$ X_t

Encoder

X_t $\boxed{S_{2:t\text{-}d\text{-}1}, X_{2:t}}$ $\hat{S}_t$

Decoder

Structure of Optimal Encoder and Decoder

- Coordinator for $[d + 3, T]$

$$S_t \rightarrow \boxed{S_{3:t}, X_{3:t-1}} \rightarrow X_t$$

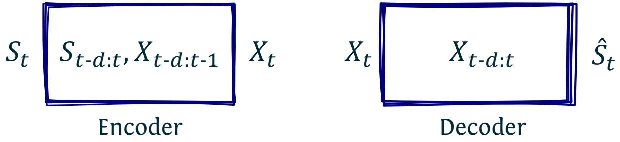
Encoder

$$X_t \rightarrow \boxed{S_{3:t-d-1}, X_{3:t}} \rightarrow \hat{S}_t$$

Decoder

Structure of Optimal Encoder and Decoder

- Coordinator for T



Structure of Optimal Encoder and Decoder

Coordinator for T

S_t $\boxed{S_{t\text{-}d:t}, X_{t\text{-}d:t\text{-}1}}$ X_t

Encoder

X_t $\boxed{X_{t\text{-}d:t}}$ $\hat{S}_t$

Decoder

The decoder does not use the additional information $S_{1:t\text{-}d\text{-}1}$.

Channel coding model

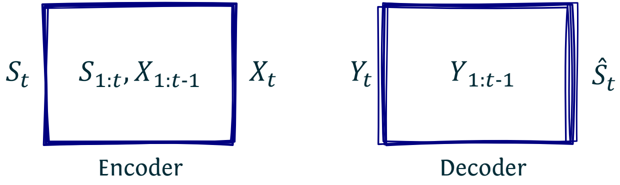
i.i.d. source	$\{S_1, S_2, \dots\}$
Sequential encoder	$X_t = f_t(S_{1:t}, X_{1:t\text{-}1})$
Sequential decoder	$\hat{S}_t = g_t(Y_{1:t})$
Fixed-finite delay	$\rho_t(S_{t\text{-}d}, \hat{S}_t)$

Objective: Choose $f = (f_1, \dots, f_T)$ and $g = (g_1, \dots, g_T)$ to minimize

$$J(f, g) = \mathbb{E}\Big[\sum_{t=d+1}^{T} \rho_t(S_{t\text{-}d+1}, \hat{S}_t) \Big]$$

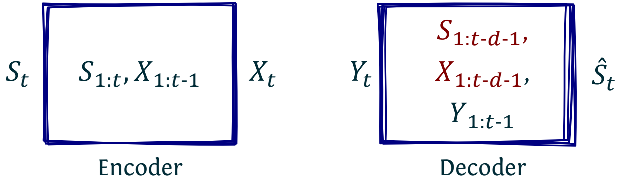
Structure of optimal encoder and decoder

- Information structure of the system



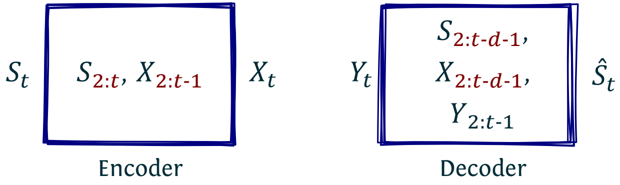
Structure of optimal encoder and decoder

- Genie aided decoder



Structure of optimal encoder and decoder

- Coordinator for $[d+2, T]$



Structure of optimal encoder and decoder

- Coordinator for $[d+3, T]$

S_t $\boxed{S_{3:t}, X_{3:t\text{-}1}}$ X_t

Encoder

Y_t $\boxed{\begin{matrix} S_{3:t\text{-}d\text{-}1}, \\ X_{3:t\text{-}d\text{-}1}, \\ Y_{3:t\text{-}1} \end{matrix}}$ $\hat{S}_t$

Decoder

Structure of optimal encoder and decoder

- Coordinator for T

S_t $\quad \boxed{S_{t\text{-}d:t}, X_{t\text{-}d:t}} \quad$ X_t

Encoder

Y_t $\quad \boxed{Y_{t\text{-}d:t\text{-}1}} \quad$ $\hat{S}_t$

Decoder

Channel coding with feedback

i.i.d. source	$\{S_1, S_2, \dots\}$
Sequential encoder	$X_t = f_t(S_{1:t}, X_{1:t\text{-}1}, Y_{1:t\text{-}1})$
Sequential decoder	$\hat{S}_t = g_t(Y_{1:t})$
Fixed-finite delay	$\rho_t(S_{t\text{-}d}, \hat{S}_t)$

$$\min J(f, g) = \min \mathbb{E}\Big[\sum_{t=d+1}^{T} \rho_t(S_{t\text{-}d+1}, \hat{S}_t) \Big]$$

Structure of optimal encoder and decoder

- Information structure of the system

S_t $\boxed{\begin{array}{c} S_{1:t}, X_{1:t\text{-}1}, \\ Y_{1:t\text{-}1} \end{array}}$ X_t

Encoder

Y_t $\boxed{Y_{1:t\text{-}1}}$ $\hat{S}_t$

Decoder

Structure of optimal encoder and decoder

- Genie aided decoder

$$S_t \;\; \boxed{\begin{array}{c} S_{1:t}, X_{1:t-1}, \\ Y_{1:t-1} \end{array}} \;\; X_t \qquad\qquad Y_t \;\; \boxed{\begin{array}{c} {\color{darkred} S_{1:t-d-1}}, \\ Y_{1:t-1} \end{array}} \;\; \hat{S}_t$$

Encoder — Decoder

Structure of optimal encoder and decoder

- Coordinator for $[d + 2, T]$

S_t $\boxed{\begin{matrix} S_{2:t}, X_{2:t}, \\ Y_{2:t-1} \end{matrix}}$ X_t

Encoder

Y_t $\boxed{\begin{matrix} S_{2:t-d-1}, \\ Y_{2:t-1} \end{matrix}}$ $\hat{S}_t$

Decoder

Structure of optimal encoder and decoder

- Coordinator for $[d+3, T]$

$$S_t \quad \boxed{\begin{matrix} S_{3:t}, X_{3:t}, \\ Y_{3:t\text{-}1} \end{matrix}} \quad X_t \qquad\qquad Y_t \quad \boxed{\begin{matrix} S_{3:t\text{-}d\text{-}1}, \\ Y_{3:t\text{-}1} \end{matrix}} \quad \hat{S}_t$$

Encoder　　　　Decoder

Structure of optimal encoder and decoder

- Coordinator for T

S_t $\boxed{S_{t\text{-}d:t}, X_{t\text{-}d:t}, Y_{t\text{-}d:t\text{-}1}}$ X_t

Encoder

Y_t $\boxed{Y_{t\text{-}d:t\text{-}1}}$ $\hat{S}_t$

Decoder

Structure of optimal encoder and decoder

- Fix decoder, consider optimal encoder

$$S_t \; \boxed{S_{t\text{-}d:t},\, Y_{t\text{-}d:t\text{-}1}} \; X_t \qquad\qquad Y_t \; \boxed{\quad Y_{t\text{-}d:t\text{-}1} \quad} \; \hat{S}_t$$

Encoder　　　　　　　Decoder

Given channel outputs and corresponding source outputs, channel inputs are irrelevant.

Summary of results for point-to-point setup

- Source coding

$$X_t = f_t(S_{t\text{-}d:t}, X_{t\text{-}d:t\text{-}1}) \quad \text{and} \quad \hat{S}_t = g_t(X_{t\text{-}d:t})$$

- Channel coding

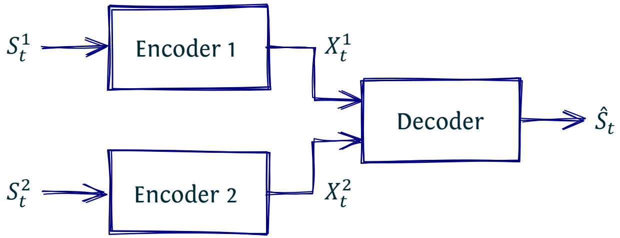
$$X_t = f_t(S_{t\text{-}d:t}, X_{t\text{-}d:t\text{-}1}) \quad \text{and} \quad \hat{S}_t = g_t(Y_{t\text{-}d:t})$$

- Channel coding with feedback

$$X_t = f_t(S_{t\text{-}d:t}, Y_{t\text{-}d:t\text{-}1}) \quad \text{and} \quad \hat{S}_t = g_t(Y_{t\text{-}d:t})$$

Multi-terminal communication
Two transmitters, one receiver

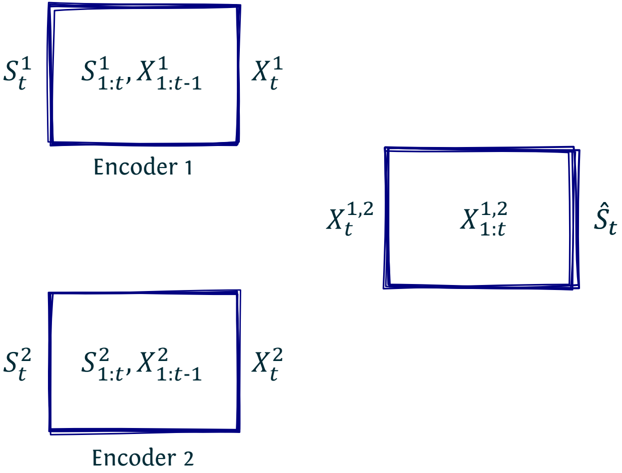
Distributed source coding



i.i.d. source	$\{(S_1^1, S_1^2), (S_2^1, S_2^2), \dots\}$
Sequential encoder	$X_t^k = f_t(S_{1:t}^k, X_{1:t-1}^k)$
Sequential decoder	$\hat{S}_t = g_t(X_{1:t}^1, X_{1:t}^2)$
Fixed-finite delay	$\rho_t(S_{t-d}^1, S_{t-d}^2, \hat{S}_t)$

Structue of optimal encoders and decoder

- Original information structure



Structue of optimal encoders and decoder

Genie aided nodes

S_t^1 → $S_{1:t}^1$, $S_{1:t\text{-}d\text{-}1}^2$, $X_{1:t\text{-}1}^1$, $X_{1:t\text{-}d\text{-}1}^2$ → X_t^1

Encoder 1

$X_t^{1,2}$ → $S_{1:t\text{-}d\text{-}1}^{1,2}$, $X_{1:t}^{1,2}$ → $\hat{S}_t$

S_t^2 → $S_{1:t}^2$, $S_{1:t\text{-}d\text{-}1}^1$, $X_{1:t\text{-}d\text{-}1}^1$, $X_{1:t\text{-}1}^2$ → X_t^2

Encoder 2

Structue of optimal encoders and decoder

- Coordinator for $[d+2, T]$

S^1_t → $\boxed{S^1_{2:t}, S^2_{2:t\text{-}d\text{-}1}, X^1_{2:t\text{-}1}, X^2_{2:t\text{-}d\text{-}1}}$ → X^1_t

Encoder 1

$X^{1,2}_t$ → $\boxed{S^{1,2}_{2:t\text{-}d\text{-}1}, X^{1,2}_{2:t}}$ → $\hat{S}_t$

S^2_t → $\boxed{S^2_{2:t}, S^1_{2:t\text{-}d\text{-}1}, X^1_{2:t\text{-}d\text{-}1}, X^2_{2:t\text{-}1}}$ → X^2_t

Encoder 2

Structue of optimal encoders and decoder

Coordinator for $[d + 3, T]$

S_t^1 → $S_{3:t}^1, S_{3:t\text{-}d\text{-}1}^2, X_{3:t\text{-}1}^1, X_{3:t\text{-}d\text{-}1}^2$ → X_t^1

Encoder 1

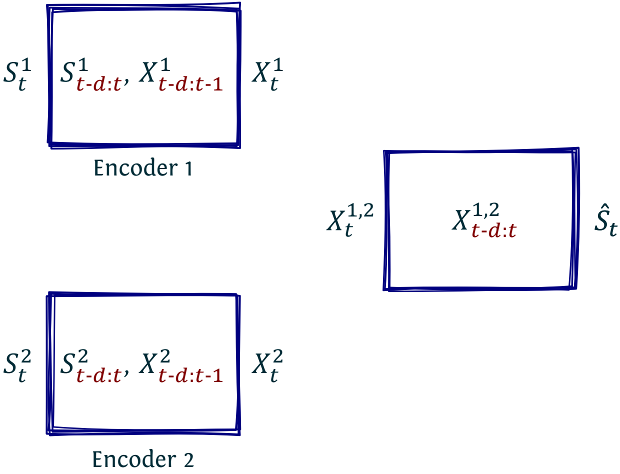
$X_t^{1,2}$ → $S_{3:t\text{-}d\text{-}1}^{1,2}, X_{3:t}^{1,2}$ → $\hat{S}_t$

S_t^2 → $S_{3:t}^2, S_{3:t\text{-}d\text{-}1}^1, X_{3:t\text{-}d\text{-}1}^1, X_{3:t\text{-}1}^2$ → X_t^2

Encoder 2

Structue of optimal encoders and decoder

Coordinator for T



Summary of results for 2-Tx 1-Rx

- Multi-terminal source coding

$$X_t^k = f_t^k(S^k_{t\text{-}d:t}, X^k_{t\text{-}d:t\text{-}1}) \quad \text{and} \quad \hat{S}_t = g_t(X^{1,2}_{t\text{-}d:t})$$

- Multiple access channel (MAC)

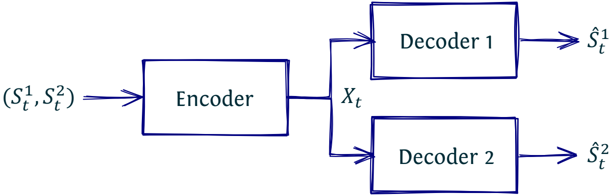
$$X_t^k = f_t^k(S^k_{t\text{-}d:t}, X^k_{t\text{-}d:t\text{-}1}) \quad \text{and} \quad \hat{S}_t = g_t(Y_{t\text{-}d:t})$$

- MAC with with feedback

$$X_t^k = f_t^k(S^k_{t\text{-}d:t}, X^k_{t\text{-}d:t\text{-}1}, Y_{t\text{-}d:t\text{-}1}) \quad \text{and} \quad \hat{S}_t = g_t(Y_{t\text{-}d:t})$$

Multi-terminal communication
One transmitter, two receivers

Distributed reconstruction source coding



i.i.d. source	$\{(S_1^1, S_1^2), (S_2^1, S_2^2), \ldots\}$
Sequential encoder	$X_t = f_t(S_{1:t}, X_{1:t\text{-}1})$
Sequential decoder	$\hat{S}_t^k = g_t(X_{1:t})$
Fixed-finite delay	$\rho_t(S_{t\text{-}d}^1, S_{t\text{-}d}^2, \hat{S}_t^1, \hat{S}_t^2)$

Summary of results for 1-Tx 2-Rx

- Distributed reconstruction

$$X_t = f_t(S^{1,2}_{t\text{-}d:t}, X_{t\text{-}d:t\text{-}1}) \quad \text{and} \quad \hat{S}^k_t = g^k_t(X_{t\text{-}d:t})$$

- Broadcast channel

$$X_t = f_t(S^{1,2}_{t\text{-}d:t}, X_{t\text{-}d:t\text{-}1}) \quad \text{and} \quad \hat{S}^k_t = g^k_t(Y^k_{t\text{-}d:t})$$

- Broadcast channel with feedback

$$X_t = f_t(S^{1,2}_{t\text{-}d:t}, Y_{t\text{-}d:t\text{-}1}) \quad \text{and} \quad \hat{S}^k_t = g^k_t(Y^k_{t\text{-}d:t})$$

Concluding remarks

- Optimal structure for fixed-finite delay reconstruction of i.i.d. sequences
  - Might be useful for block Markov coding schemes.
- A (new?) proof technique to prove structural results
  - Genie-aided nodes with additional variables
  - Forward induction argument to remove irrelevant variables

Concluding remarks

- Dynamic programming decomposition

  - All nodes have fixed memory. So we can use the approach of Sandell '74, M '08.

    $$\text{info state} = \mathbb{P}(\text{Memory of all nodes at } t)$$

    Time-varying optimal control laws!

  - "State" observed after a delay. So we can also use the approach of Aicardi *et al.* '87, Nayyar M Teneketzis '12

    $$\text{info state} = (\text{local memory}, (partially evaluated functions))$$

    Time-invariant optimal control laws!

Thank you