Introduction to Sequential Teams

ADITYA MAHAJAN

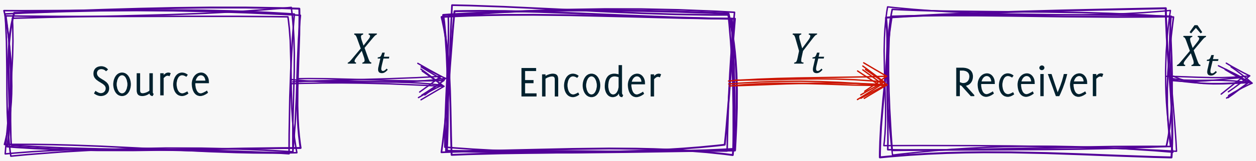
MCGILL UNIVERSITY

Joint work with: Ashutosh Nayyar and Demos Teneketzis, UMichigan

MITACS Workshop on Fusion and Inference in Networks, 2011

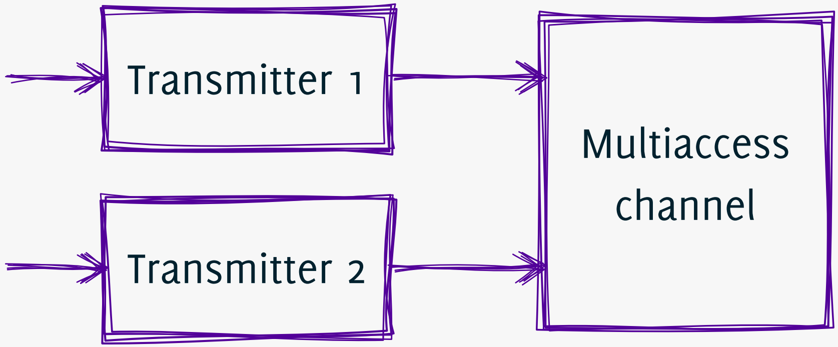
Decentralized systems are ubiquitous

Real-time quantization



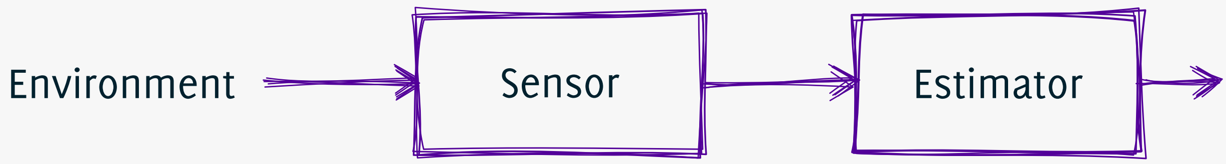
- Objective Choose transmission and estimation policy to minimize expected total distortion (over a finite or infinite horizon)

Multiaccess broadcast



- Objective Choose transmission policy to maximize throughput (over a finite or infinite horizon)

Estimating with active sensing



- Objective Choose transmission and estimation policy to minimize a weighted average of expected transmission cost and expected total distortion (over a finite or infinite horizon)

Systematic design of decentralized systems

- **Salient Features**
  - Multi-stage decision problems
  - Multiple decision makers (or agents) with decentralized information
- **Structure of optimal policy**

  Can an agent, or a group of agents

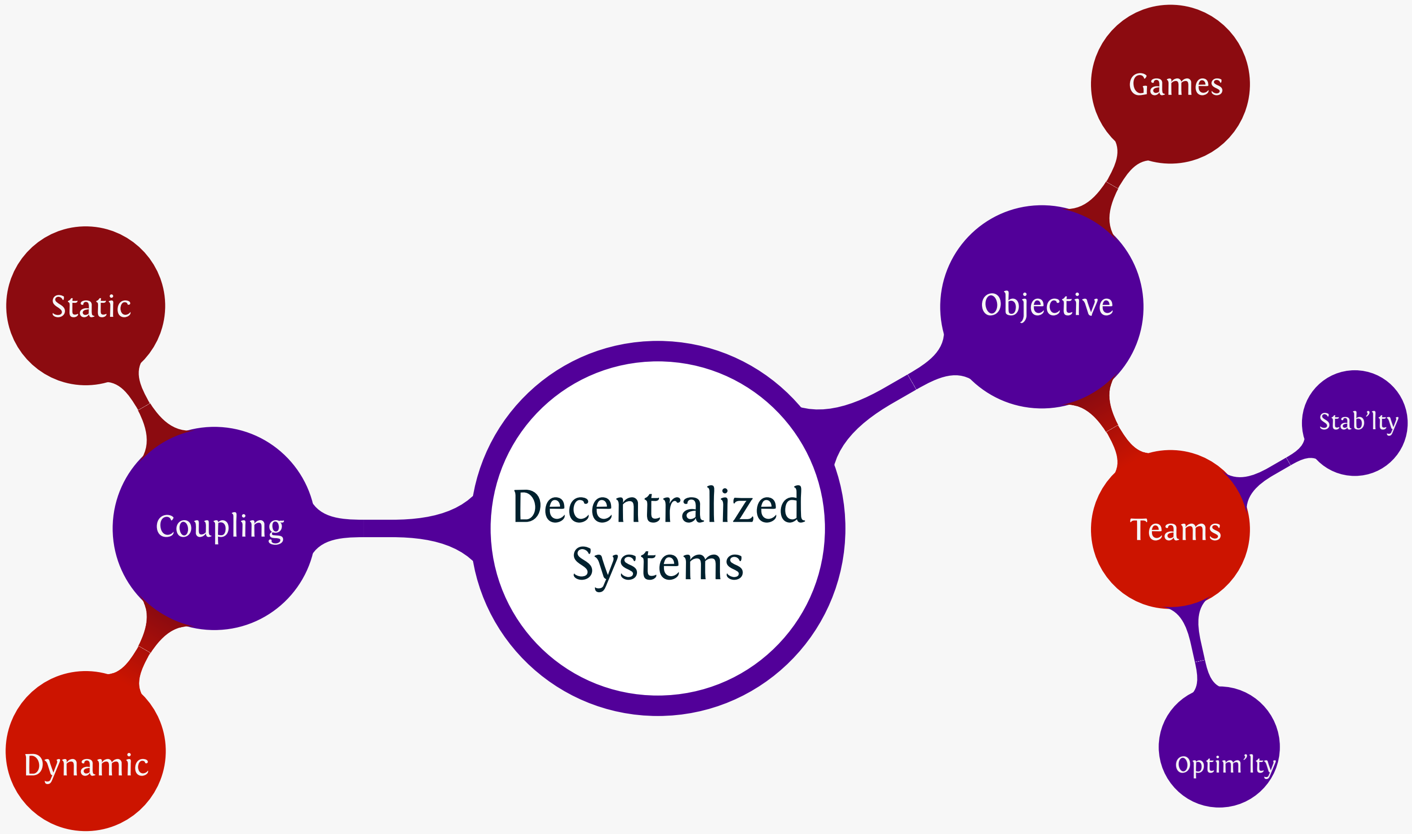
  - Shed available data
  - Compress available data

  without loss of optimality?
- **Search for optimal policies**
  - Brute force search of an optimal policy has doubly exponential complexity with time-horizon.
  - How can we search for an optimal policy efficiently?

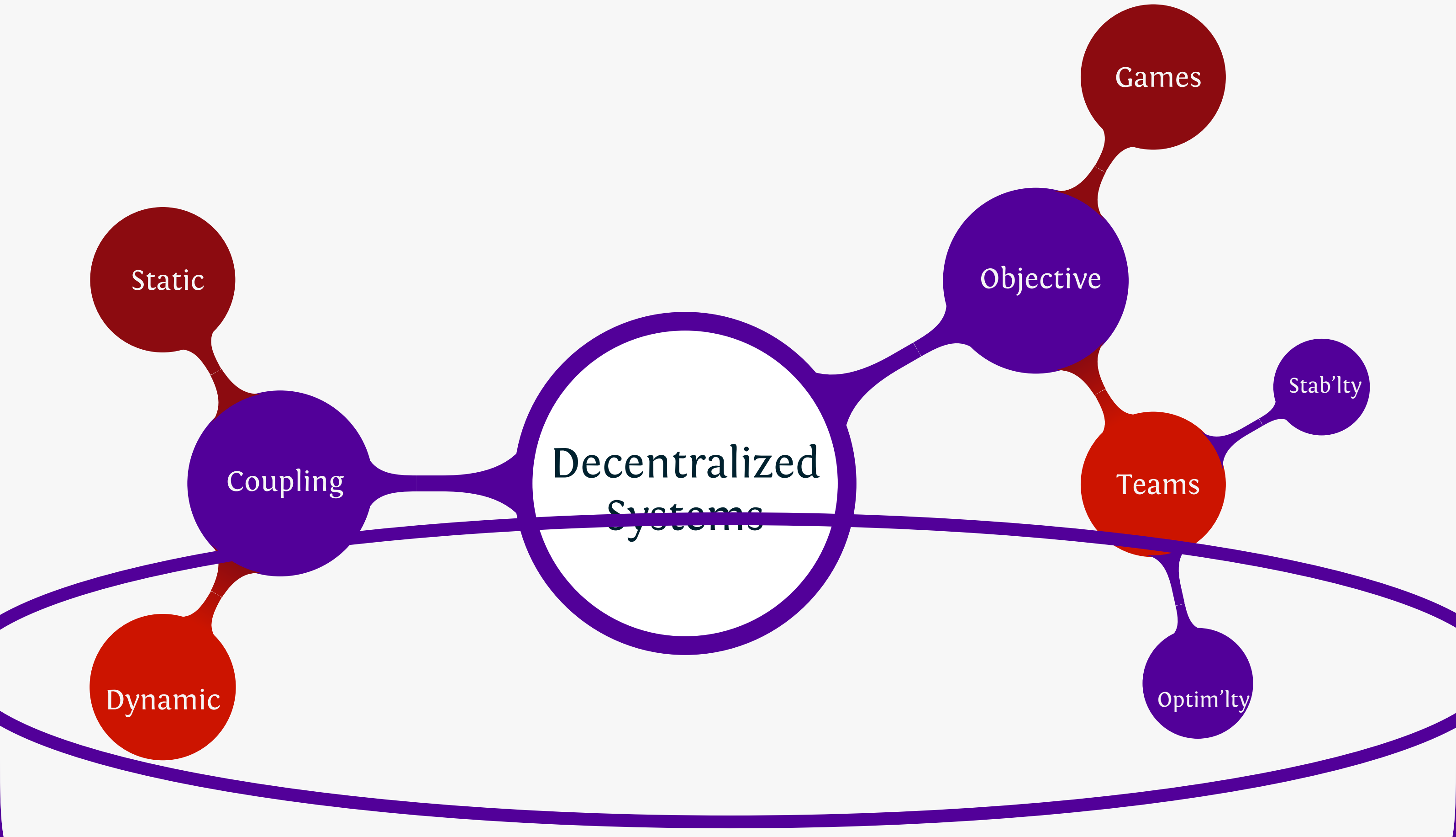
Outline

- A taxonomy of decentralized systems
- Overview of centralized stochastic control
  - Markov decision processes (MDP)
  - Partially observable Markov decision processes (POMDP)
  - Delayed state observation
- Design principle for sequential teams.
  - Delayed state observation

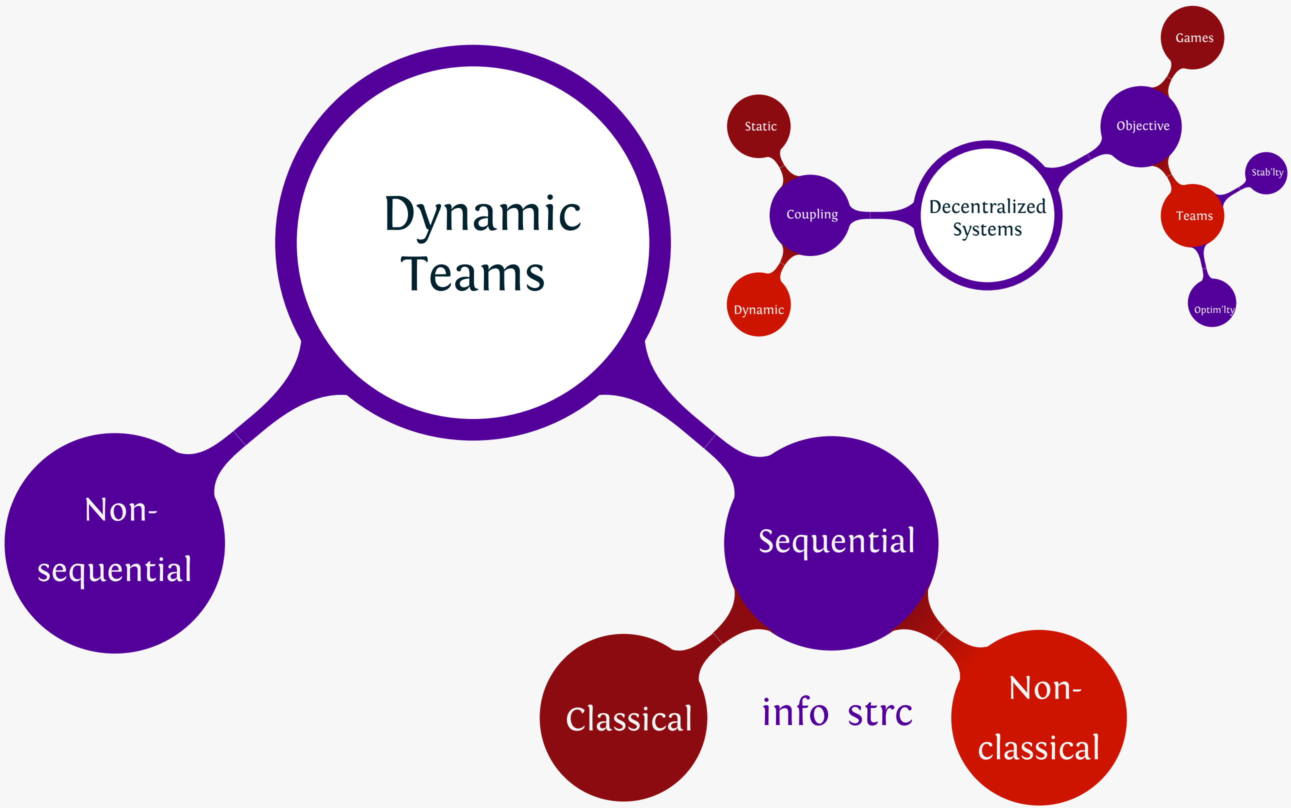
Classification of decentralized systems



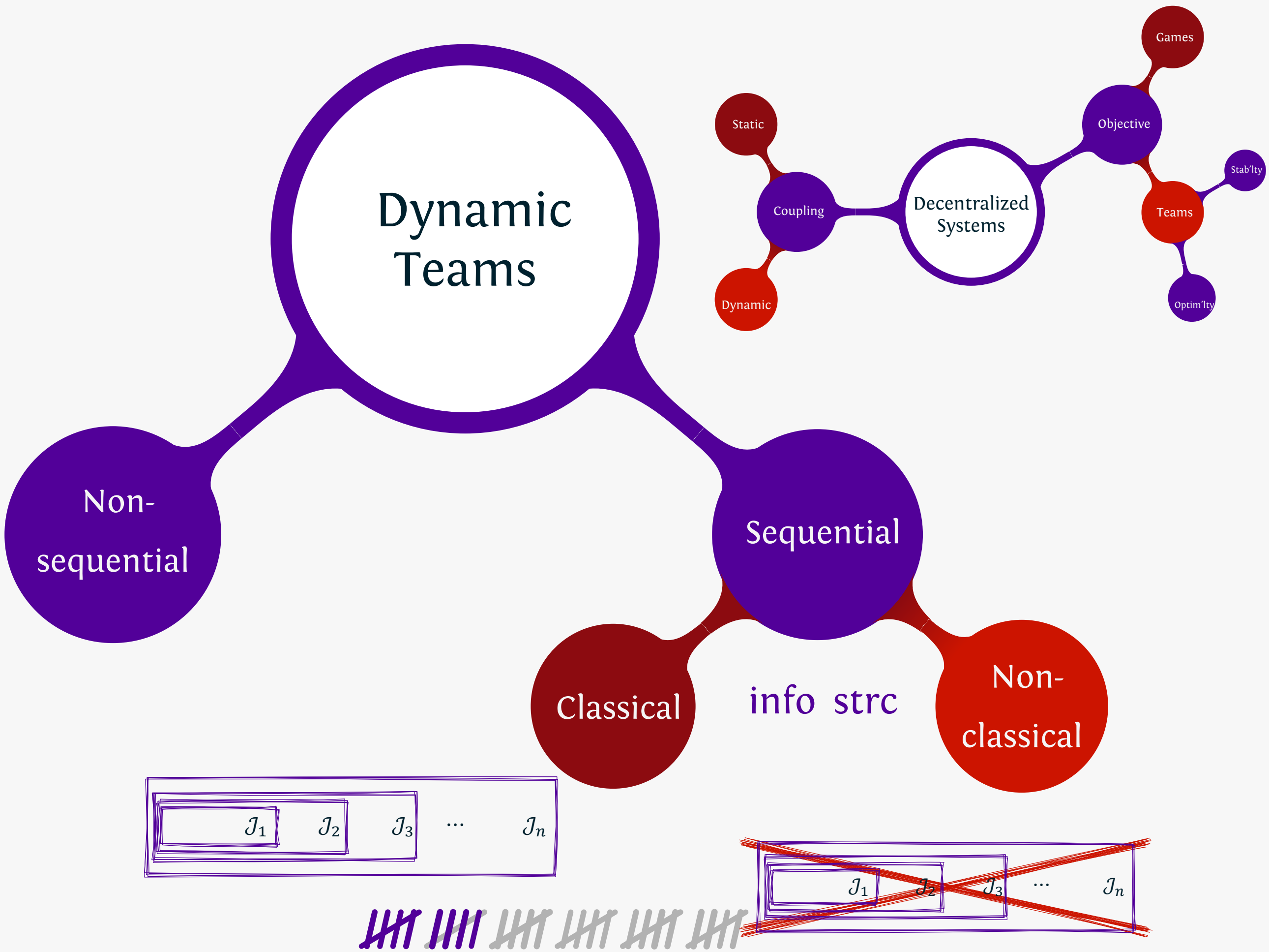
Classification of decentralized systems



Classification of decentralized systems

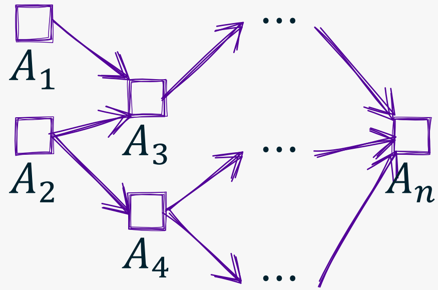


Classification of decentralized systems

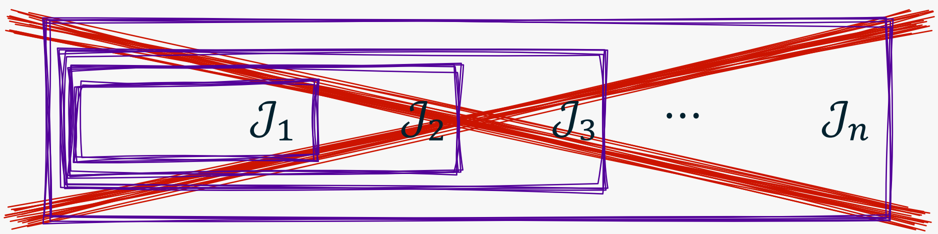


We are interested in

Sequential dynamic teams



with non-classical information structures



A bit of history . . .

TEAM DECISION PROBLEMS[1]

By R. Radner

University of California, Berkeley

1. Introduction. In a *team decision problem* there are two or more decision variables, and these different decisions can be made to depend upon different aspects of the environment, i.e., upon different information variables. For ex-

ECONOMIC THEORY OF TEAMS

by
JACOB MARSCHAK and ROY RADNER

Yale University Press, New Haven and London
1972

A bit of history . . .

SIAM J. Control
Vol. 9, No. 2, May 1971

ON INFORMATION STRUCTURES, FEEDBACK AND CAUSALITY*

H. S. WITSENHAUSEN†

Abstract. A finite number of decisions, indexed by $\alpha \in A$, are to be taken. Each decision amounts to selecting a point in a measurable space $(U_\alpha, \mathscr{F}_\alpha)$. Each decision is based on some information fed back from the system and characterized by a subfield $\mathscr{I}_\alpha$ of the product space $(\prod_\alpha U_\alpha, \prod_\alpha \mathscr{F}_\alpha)$. The decision function for each α can be any function γ_α measurable from $\mathscr{I}_\alpha$ to $\mathscr{F}_\alpha$.

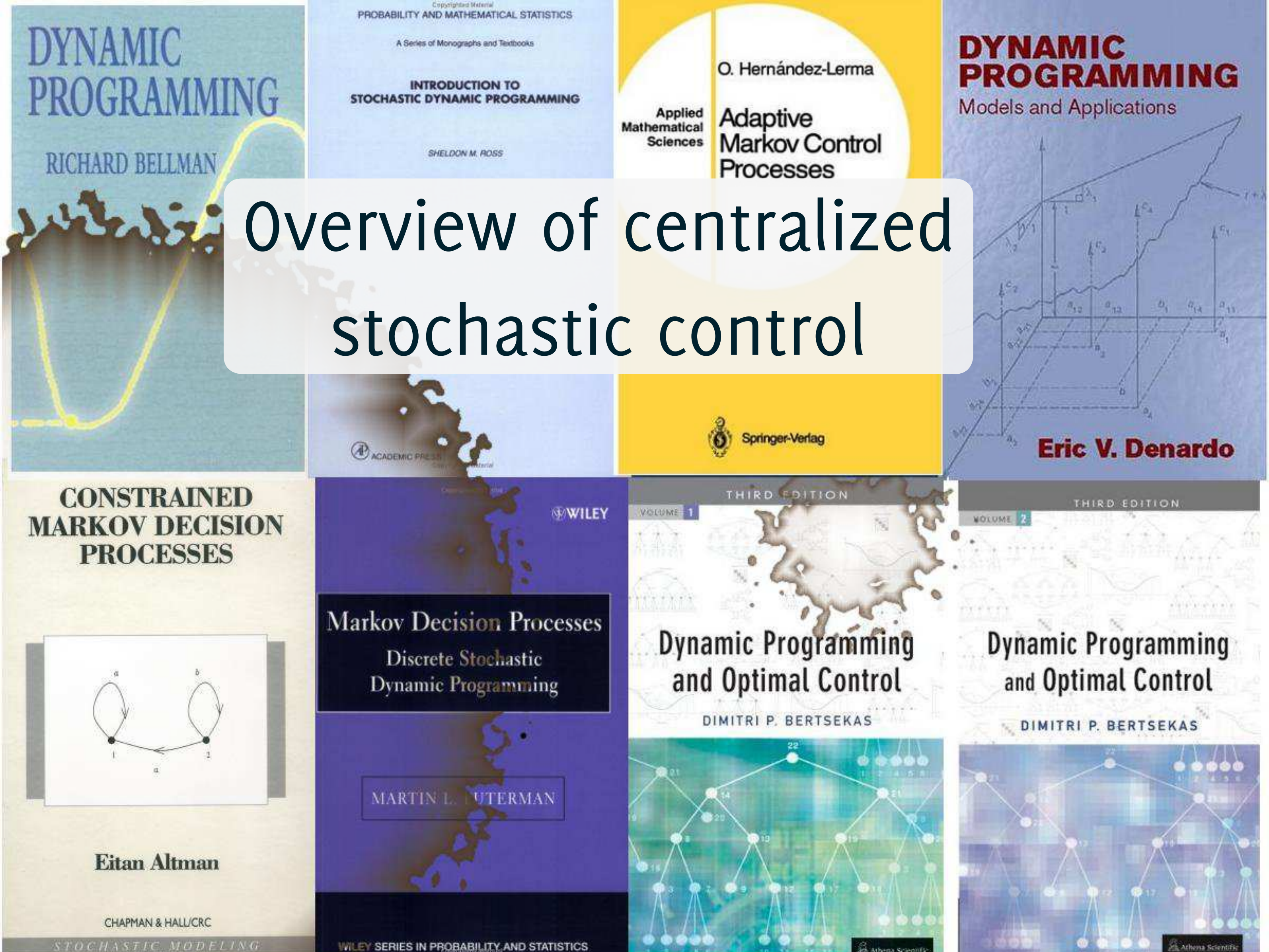
PROCEEDINGS OF THE IEEE, VOL. 59, NO. 11, NOVEMBER 1971 1557

Separation of Estimation and Control for Discrete Time Systems

HANS S. WITSENHAUSEN, MEMBER, IEEE

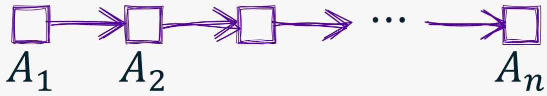
Invited Paper

Overview of centralized stochastic control

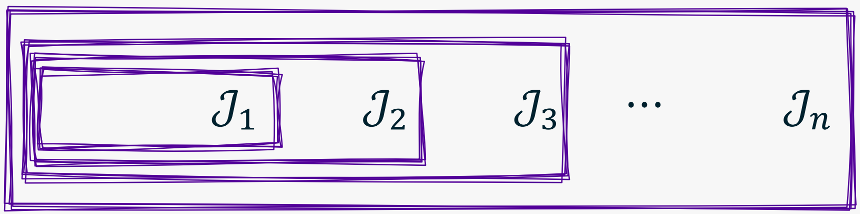


Centralized stochastic control

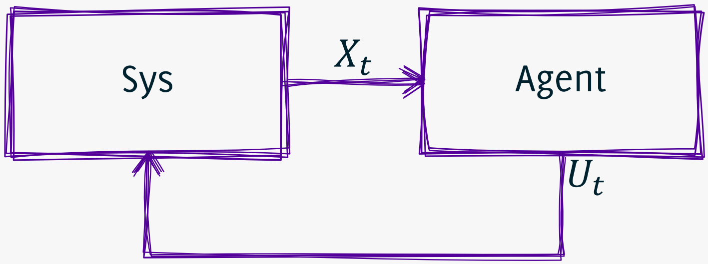
Single decision maker



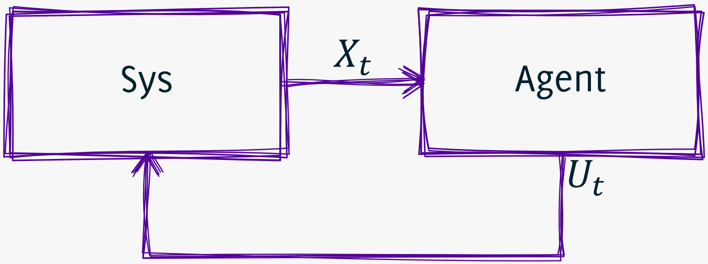
with classical information structures



MDP: Structural properties



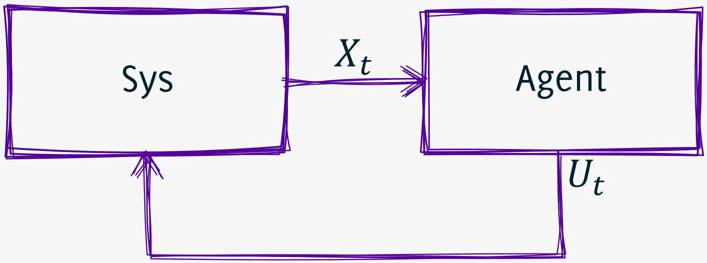
MDP: Structural properties

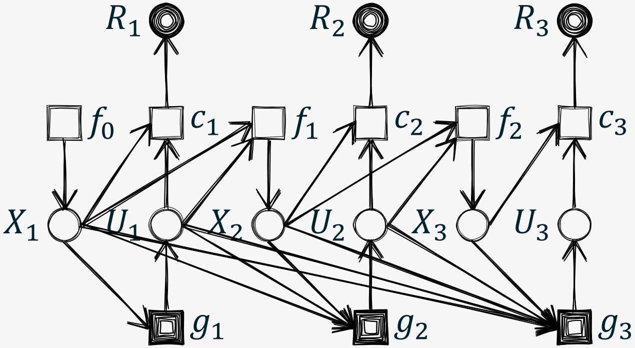


Structure of optimal policy

Choose current action
based on current state $\mathbf{X_t}$

MDP: Structural properties

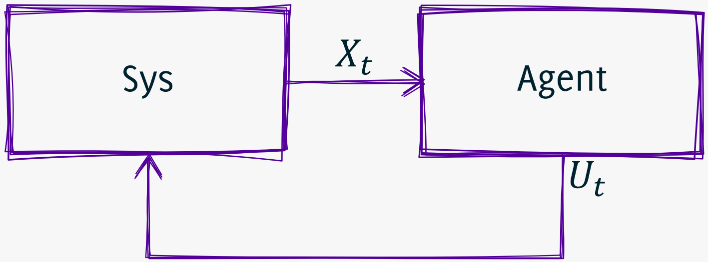


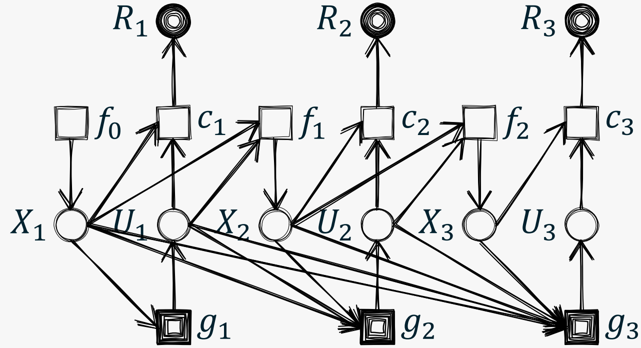


Structure of optimal policy

Choose current action based on current state $\mathbf{X_t}$

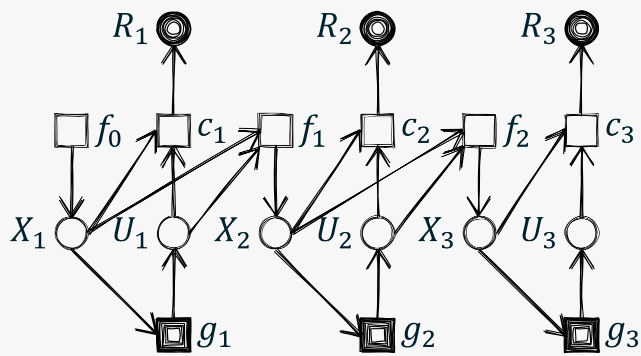
MDP: Structural properties



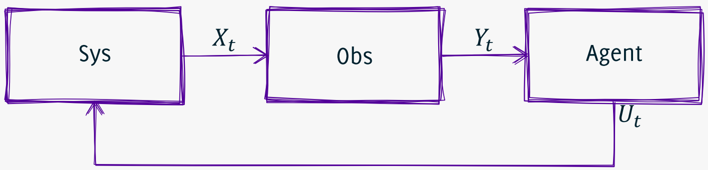


Structure of optimal policy

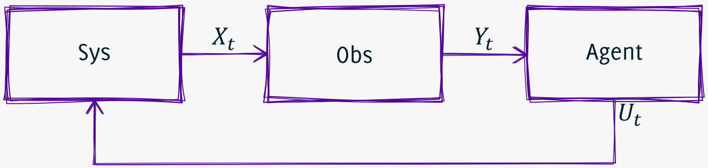
Choose current action
based on current state $\mathbf{X_t}$



POMDP: Structural properties



POMDP: Structural properties

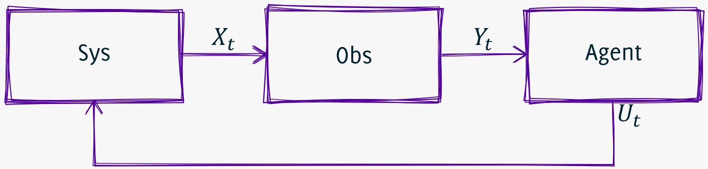


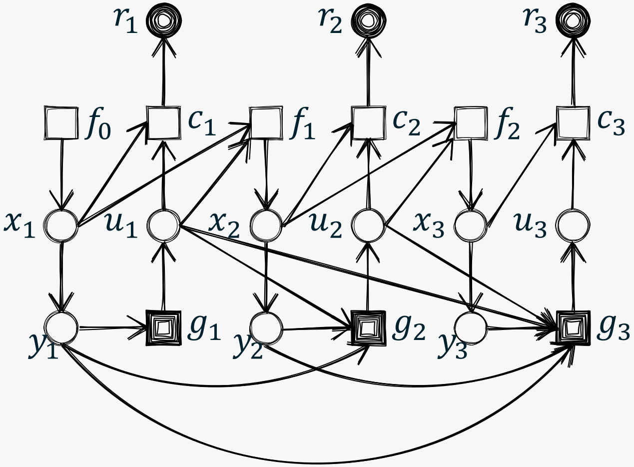
Structure of optimal policies

Choose current action
based on current info state

Pr(state of system | all data at agent)

POMDP: Structural properties



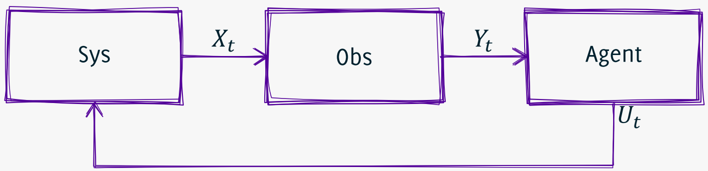


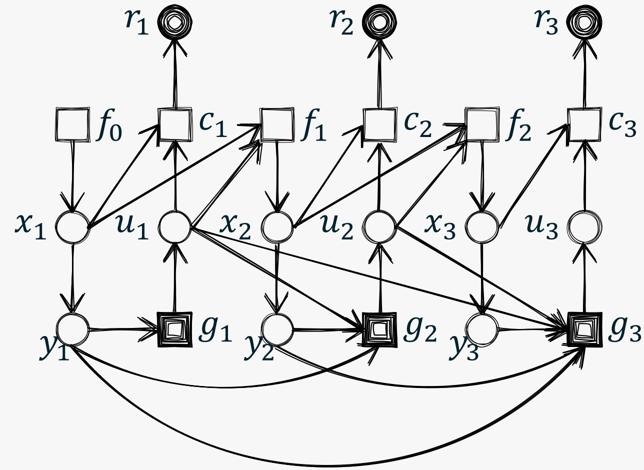
Structure of optimal policies

Choose current action
based on current info state

Pr(state of system | all data at agent)

POMDP: Structural properties

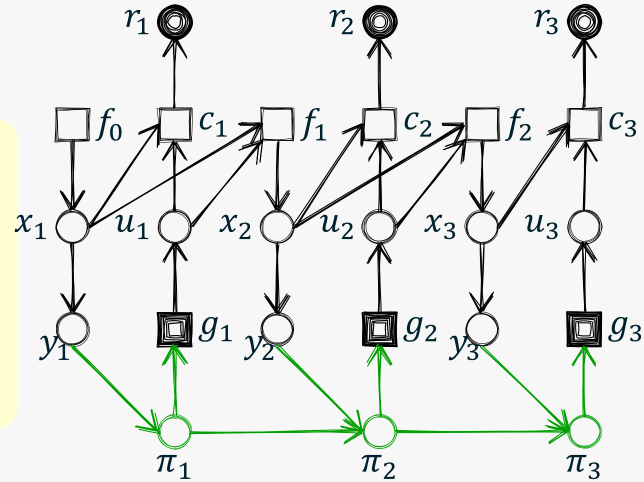




Structure of optimal policies

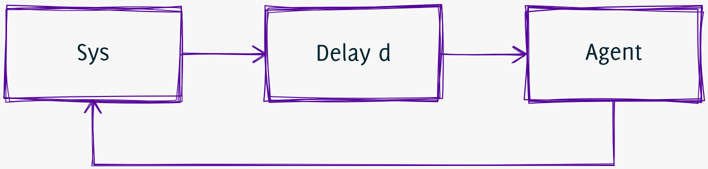
Choose current action
based on current info state

Pr(state of system | all data at agent)



An example: Delayed observation

An example: Delayed observation

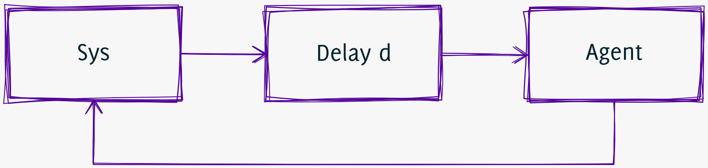


Structure of optimal policies

Choose control action based on:

$$\pi_t = \Pr(X_t | X_{1:t-d}, U_{1:t-1})$$
$$\equiv (X_{t-d}, U_{t-d:t-1})$$

An example: Delayed observation



Original form of control laws

$$U_t = g_t(X_{1:t-d}, U_{1:t-1})$$

Structure of optimal policies

Choose control action based on:

$$\pi_t = \Pr(X_t | X_{1:t-d}, U_{1:t-1})$$
$$\equiv (X_{t-d}, U_{t-d:t-1})$$

An example: Delayed observation

Original form of control laws

$U_t = g_t(X_{1:t-d}, U_{1:t-1})$

Structure of optimal policies

Choose control action based on:

$\pi_t = \Pr(X_t | X_{1:t-d}, U_{1:t-1})$

$\equiv (X_{t-d}, U_{t-d:t-1})$

Simplified form of control laws

$U_t = g_t(X_{t-d}, U_{t-d:t-1})$

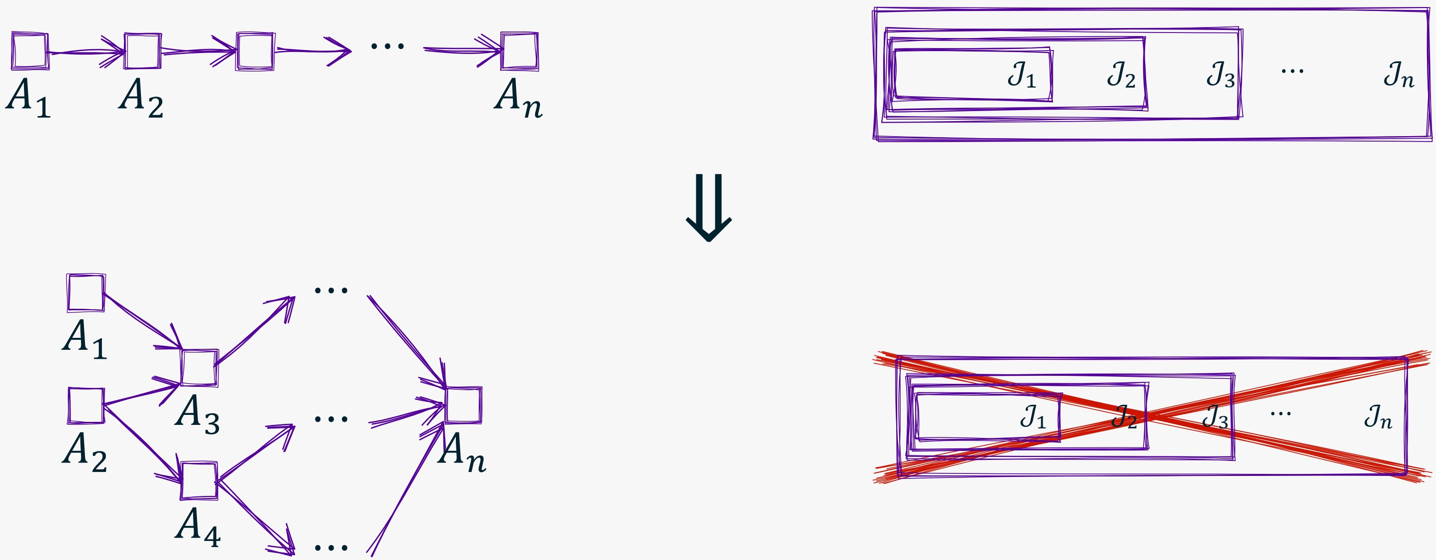
Structural policies in stochastic control

- Structure of optimal policies
  - Shed irrelevant information
  - Compress relevant information to a compact statistic
  - Hopefully, the data at the agent is not increasing with time

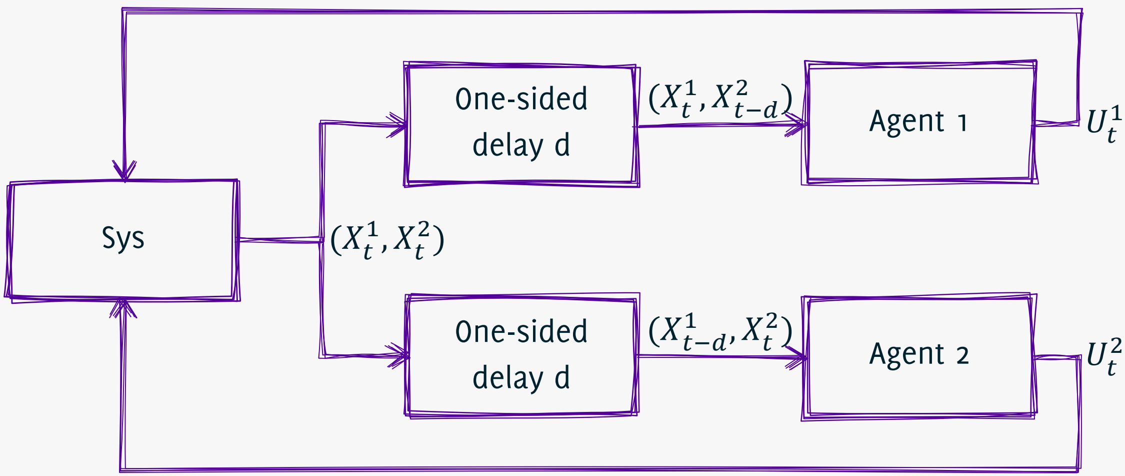
Structural policies in stochastic control

- Structure of optimal policies
  - Shed irrelevant information
  - Compress relevant information to a compact statistic
  - Hopefully, the data at the agent is not increasing with time
- Implication of the results
  - Simplify the functional form of the decision rules
  - Simplify search for optimal decision rules
  - A prerequisite for deriving dynamic programming decomposition.

Extending ideas to decentralized control



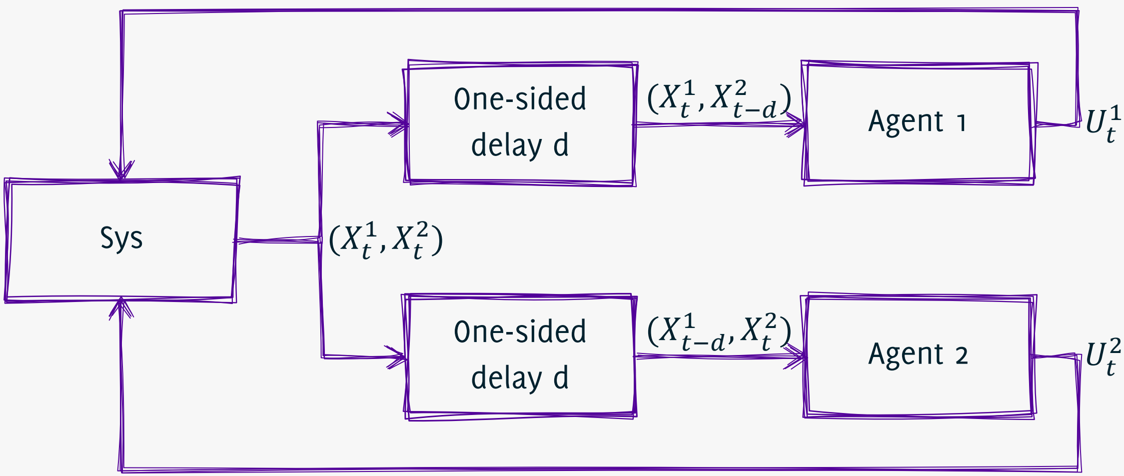
Delayed observation of state



Original form of control laws

$$U_t^i = g_t^i\left(\begin{bmatrix} X_{1:t}^i \\ U_{1:t-1}^i \end{bmatrix}, \begin{bmatrix} X_{1:t-d}^j \\ U_{1:t-d}^j \end{bmatrix}\right)$$

Delayed observation of state



Original form of control laws

$$U_t^i = g_t^i\left(\begin{bmatrix} X_{1:t}^i \\ U_{1:t-1}^i \end{bmatrix}, \begin{bmatrix} X_{1:t-d}^j \\ U_{1:t-d}^j \end{bmatrix}\right)$$

Is this structure correct?

$$U_t^i = g_t^i\left(\begin{bmatrix} X_{t-d:t}^i \\ U_{t-d:t-1}^i \end{bmatrix}, \begin{bmatrix} X_{t-d}^j \\ U_{t-d}^j \end{bmatrix}\right)$$

Lets consider delay d = 2

- At Agent 1

$$U_1^1 = g_1^1(X_1^1)$$

$$U_2^1 = g_2^1(X_{1:2}^1, U_1^1)$$

$$U_3^1 = g_3^1(X_{1:3}^1, U_{1:2}^1, X_1^2, U_1^2)$$

$$U_4^1 = g_4^1(X_{1:4}^1, U_{1:3}^1, X_{1:2}^2, U_{1:2}^2)$$

Lets consider delay d = 2

- At Agent 1

$$U_1^1 = g_1^1(X_1^1)$$

$$U_2^1 = g_2^1(X_{1:2}^1, U_1^1)$$

$$U_3^1 = g_3^1(X_{1:3}^1, U_{1:2}^1, X_1^2, U_1^2)$$

$$U_4^1 = g_4^1(X_{1:4}^1, U_{1:3}^1, X_{1:2}^2, U_{1:2}^2)$$

- At Agent 2

$$U_1^2 = g_1^2(X_1^2)$$

$$U_2^2 = g_2^2(X_{1:2}^2, U_1^2)$$

$$U_3^2 = g_3^2(X_{1:3}^2, U_{1:2}^2, X_1^1, U_1^1)$$

$$U_4^2 = g_4^2(X_{1:4}^2, U_{1:3}^2, X_{1:2}^1, U_{1:2}^1)$$

Lets consider delay d = 2

- At Agent 1

At time 4, agent 1 can't remove X_1^1 because X_1^1 gives some information about U_4^2.

$$U_1^1 = g_1^1(X_1^1)$$

$$U_2^1 = g_2^1(X_{1:2}^1, U_1^1)$$

$$U_3^1 = g_3^1(X_{1:3}^1, U_{1:2}^1, X_1^2, U_1^2)$$

$$U_4^1 = g_4^1(X_{1:4}^1, U_{1:3}^1, X_{1:2}^2, U_{1:2}^2)$$

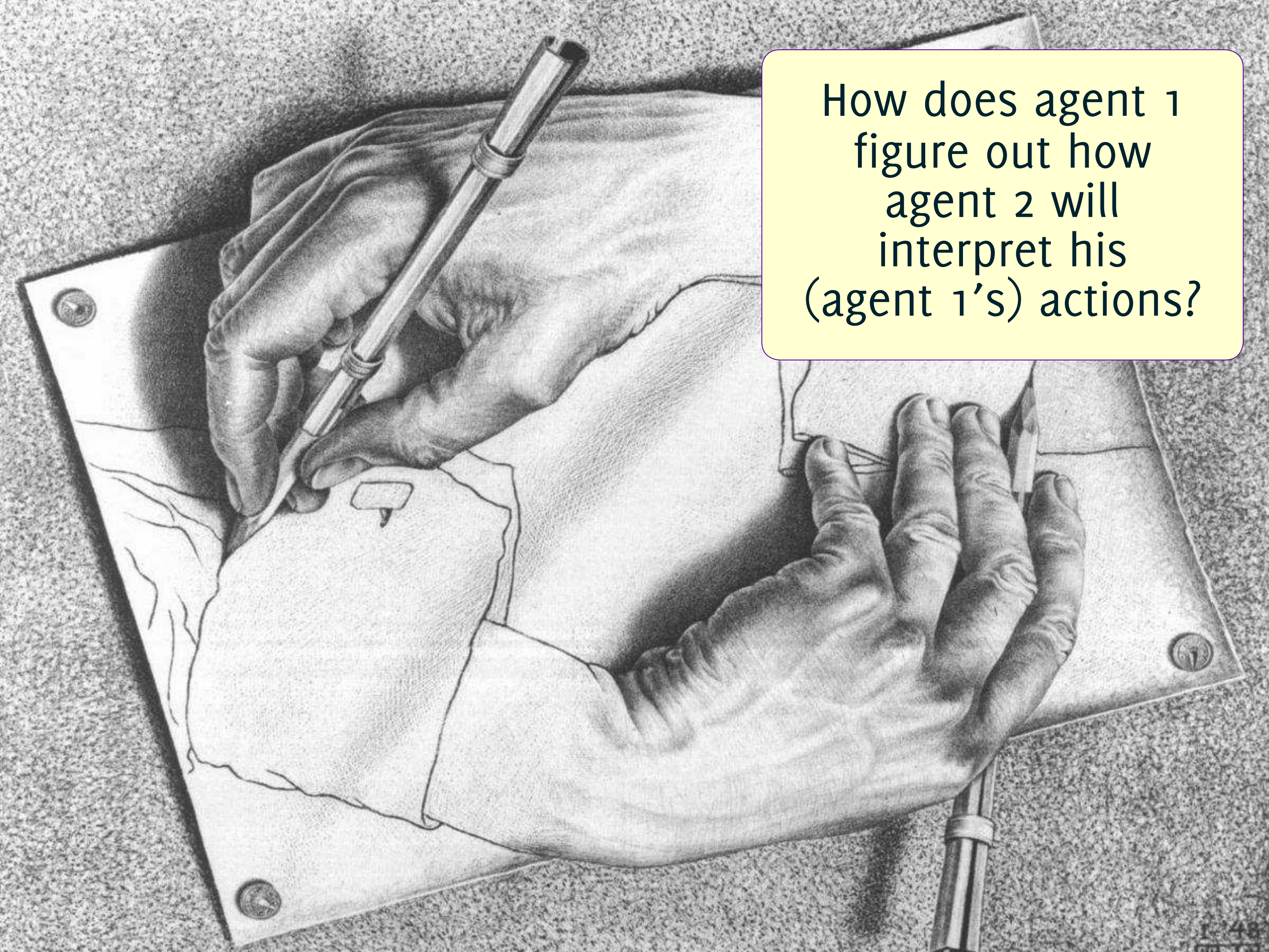
$$U_1^2 = g_1^2(X_1^2)$$

$$U_2^2 = g_2^2(X_{1:2}^2, U_1^2)$$

$$U_3^2 = g_3^2(X_{1:3}^2, U_{1:2}^2, X_1^1, U_1^1)$$

$$U_4^2 = g_4^2(X_{1:4}^2, U_{1:3}^2, X_{1:2}^1, U_{1:2}^1)$$

How does agent 1 figure out how agent 2 will interpret his (agent 1's) actions?



How does agent 1 figure out how agent 2 will interpret his (agent 1's) actions?

Solution Approach

[Mahajan 2008, 2009; Nayyar Mahajan Teneketzis 2008, 2011]

- Adapt based on common knowledge

Solution Approach

[Mahajan 2008, 2009; Nayyar Mahajan Teneketzis 2008, 2011]

- Adapt based on common knowledge
- Split observations into two parts:
  - Common data: $C_t = (X^1_{1:t-2}, X^2_{1:t-2}, U^1_{1:t-2}, U^2_{1:t-2})$
  - Local data: $L^i_t = (X^i_{t-1}, X^i_t, U^i_{t-1})$.

Solution Approach

[Mahajan 2008, 2009; Nayyar Mahajan Teneketzis 2008, 2011]

- Adapt based on common knowledge
- Split observations into two parts:
  - Common data: $C_t = (X^1_{1:t-2}, X^2_{1:t-2}, U^1_{1:t-2}, U^2_{1:t-2})$
  - Local data: $L^i_t = (X^i_{t-1}, X^i_t, U^i_{t-1})$.
- A three step approach:
  1. Consider a coordinated system
  2. Show that the coordinated system is equivalent to the original system
  3. Simplify the coordinated system

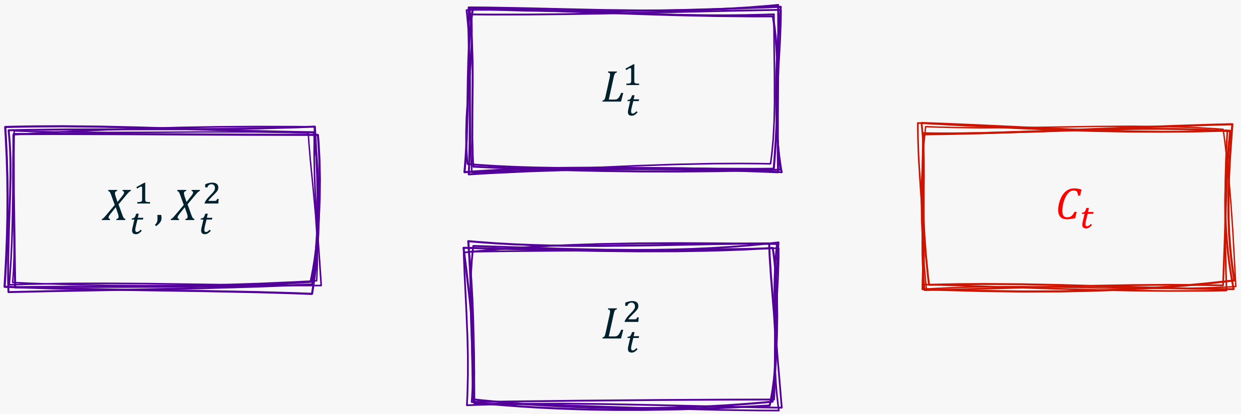
Original System

X_t^1, X_t^2

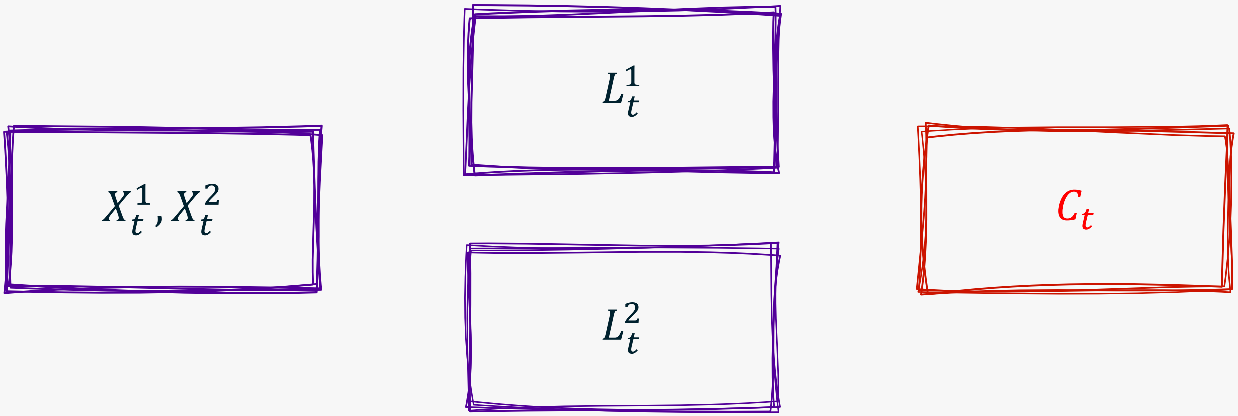
C_t, L_t^1

C_t, L_t^2

Coordinated System



Coordinated System



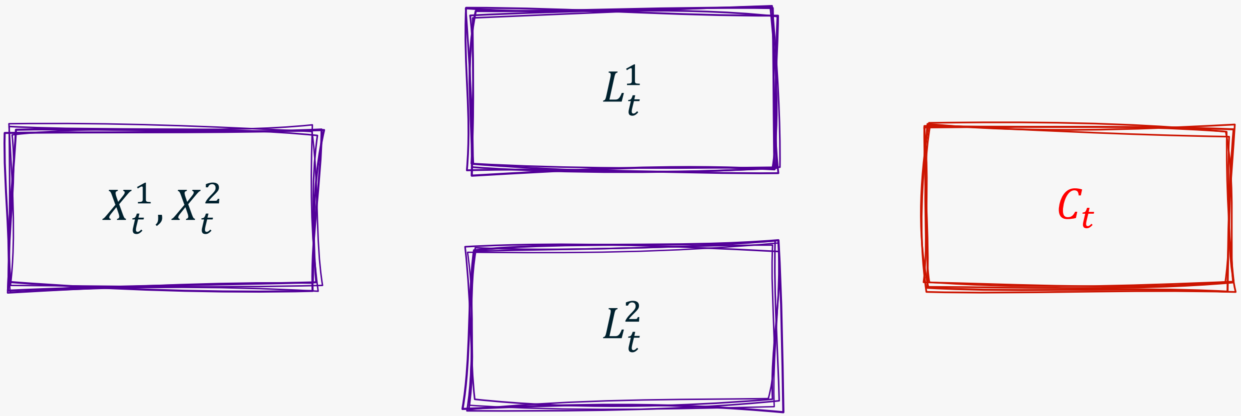
- Observations: $C_t = (X^1_{1:t-2}, X^2_{1:t-2}, U^1_{1:t-2}, U^2_{1:t-2})$
- Control "actions": Function sections γ_t^1, γ_t^2

$$\gamma_t^i(\cdot) = g_t^i(\cdot, C_t)$$

- Agents are dumb and simply follow the prescription

$$U_t^i = \gamma_t^i(L_t^i) = \gamma_t^i(X_{t-1}^i, X_t^i, U_{t-1}^i)$$

Coordinated System



- Observations: $C_t = (X^1_{1:t-2}, X^2_{1:t-2}, U^1_{1:t-2}, U^2_{1:t-2})$
- Control “actions”: Function sections γ_t^1, γ_t^2

$$\gamma_t^i(\cdot) = g_t^i(\cdot, C_t)$$

- Agents are dumb and simply follow the prescription

$$U_t^i = \gamma_t^i(L_t^i) = \gamma_t^i(X^i_{t-1}, X^i_t, U^i_{t-1})$$

- The two systems are equivalent

Structure of the coordination policy

$$(\gamma_t^1, \gamma_t^2) = \psi_t(Z_{1:t-2}, \gamma_{1:t-1}^1, \gamma_{1:t-1}^2), \quad \text{where } Z_t = (X_t^1, X_t^2, U_t^1, U_t^2)$$

Structure of the coordination policy

$$(\gamma_t^1, \gamma_t^2) = \psi_t(Z_{1:t-2}, \gamma_{1:t-1}^1, \gamma_{1:t-1}^2), \quad \text{where } Z_t = (X_t^1, X_t^2, U_t^1, U_t^2)$$

- Sufficient statistic

  $\pi_t = \Pr(\text{state} | \text{all past data})$

Structure of the coordination policy

$$(\gamma_t^1, \gamma_t^2) = \psi_t(Z_{1:t-2}, \gamma^1_{1:t-1}, \gamma^2_{1:t-1}), \quad \text{where } Z_t = (X_t^1, X_t^2, U_t^1, U_t^2)$$

- Sufficient statistic

$$\pi_t = \Pr(\text{state} | \text{all past data})$$
$$= \Pr(X_t^1, X_t^2, L_t^1, L_t^2 | Z_{1:t-2}, \gamma^1_{1:t-1}, \gamma^2_{1:t-1})$$

Structure of the coordination policy

$$(\gamma_t^1, \gamma_t^2) = \psi_t(Z_{1:t-2}, \gamma_{1:t-1}^1, \gamma_{1:t-1}^2), \quad \text{where } Z_t = (X_t^1, X_t^2, U_t^1, U_t^2)$$

- Sufficient statistic

$$\pi_t = \Pr(\text{state} | \text{all past data})$$
$$= \Pr(X_t^1, X_t^2, L_t^1, L_t^2 | Z_{1:t-2}, \gamma_{1:t-1}^1, \gamma_{1:t-1}^2)$$
$$= \Pr(X_t^1, X_t^2, Z_{t-1} | Z_{1:t-2}, \gamma_{1:t-1}^1, \gamma_{1:t-1}^2)$$

Structure of the coordination policy

$$(\gamma_t^1, \gamma_t^2) = \psi_t(Z_{1:t-2}, \gamma_{1:t-1}^1, \gamma_{1:t-1}^2), \quad \text{where } Z_t = (X_t^1, X_t^2, U_t^1, U_t^2)$$

- Sufficient statistic

$$\begin{aligned}\pi_t &= \Pr(\text{state}|\text{all past data}) \\ &= \Pr(X_t^1, X_t^2, L_t^1, L_t^2 | Z_{1:t-2}, \gamma_{1:t-1}^1, \gamma_{1:t-1}^2) \\ &= \Pr(X_t^1, X_t^2, Z_{t-1} | Z_{1:t-2}, \gamma_{1:t-1}^1, \gamma_{1:t-1}^2)\end{aligned}$$

- Structural result:

$$(\gamma_t^1, \gamma_t^2) = \psi_t(\pi_t)$$

Structure of the coordination policy

$$(\gamma_t^1, \gamma_t^2) = \psi_t(Z_{1:t-2}, \gamma_{1:t-1}^1, \gamma_{1:t-1}^2), \quad \text{where } Z_t = (X_t^1, X_t^2, U_t^1, U_t^2)$$

- Sufficient statistic

$$\begin{aligned} \pi_t &= \Pr(\text{state}|\text{all past data}) \\ &= \Pr(X_t^1, X_t^2, L_t^1, L_t^2 | Z_{1:t-2}, \gamma_{1:t-1}^1, \gamma_{1:t-1}^2) \\ &= \Pr(X_t^1, X_t^2, Z_{t-1} | Z_{1:t-2}, \gamma_{1:t-1}^1, \gamma_{1:t-1}^2) \end{aligned}$$

- Structural result:

$$(\gamma_t^1, \gamma_t^2) = \psi_t(\pi_t)$$

Or equivalently,

$$U_t^i = g_t^i(\pi_t, L_t^i)$$

Further Simplification

- Recall

$$U_t^i = \gamma_t^i(X_{t-1}^i, X_t^i, U_{t-1}^i)$$

- Define

$$\hat{\gamma}_t^i(\cdot) = \gamma_t^i(X_{t-1}^i, U_{t-1}^i, \cdot)$$

Further Simplification

- Recall

$$U_t^i = \gamma_t^i(X_{t-1}^i, X_t^i, U_{t-1}^i)$$

- Define

$$\hat{\gamma}_t^i(\cdot) = \gamma_t^i(X_{t-1}^i, U_{t-1}^i, \cdot)$$

- We can show that

$$\pi_t = \Pr(X_t^1, X_t^2, Z_{t-1} | Z_{1:t-2}, \gamma_{1:t-1}^1, \gamma_{1:t-1}^2) \equiv (Z_{t-2}, \hat{\gamma}_{t-1}^1, \hat{\gamma}_{t-1}^2)$$

Further Simplification

- Recall

$$U_t^i = \gamma_t^i(X_{t-1}^i, X_t^i, U_{t-1}^i)$$

- Define

$$\hat{\gamma}_t^i(\cdot) = \gamma_t^i(X_{t-1}^i, U_{t-1}^i, \cdot)$$

- We can show that

$$\pi_t = \Pr(X_t^1, X_t^2, Z_{t-1} | Z_{1:t-2}, \gamma_{1:t-1}^1, \gamma_{1:t-1}^2) \equiv (Z_{t-2}, \hat{\gamma}_{t-1}^1, \hat{\gamma}_{t-1}^2)$$

- Equivalent structural result

$$U_t^i = g_t^i\left(\begin{bmatrix} X_{t-2:t}^i \\ U_{t-2:t-1}^i \end{bmatrix}, \begin{bmatrix} X_{t-2}^j \\ U_{t-2}^j \end{bmatrix}, \hat{\gamma}_{t-1}^1, \hat{\gamma}_{t-1}^2\right)$$

Recap: Solution approach

[Mahajan 2008, 2009; Nayyar Mahajan Teneketzis 2008, 2011]

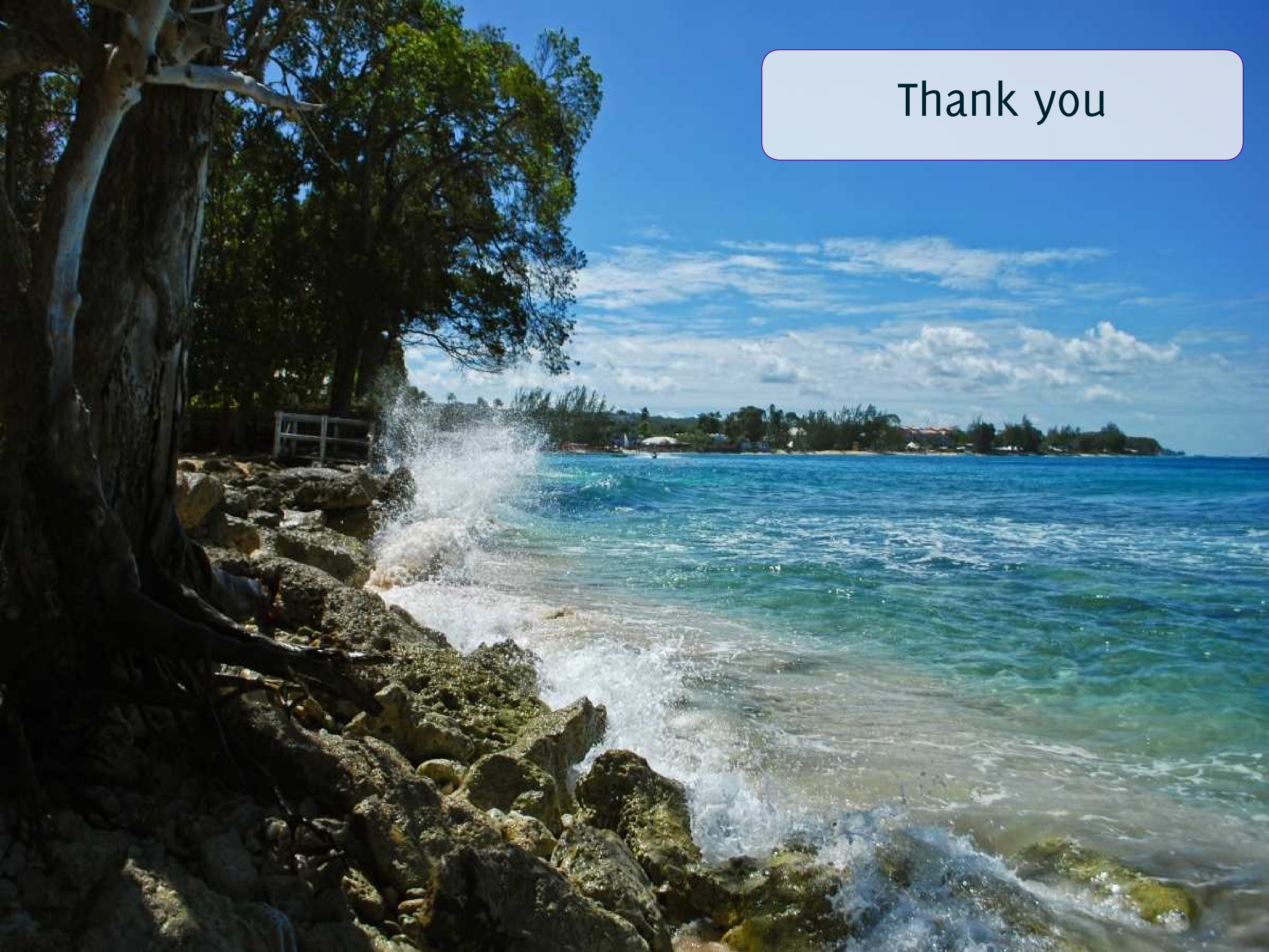
- Adapt based on common knowledge
- Split observations into two parts:
  - Common data: $C_t = (X^1_{1:t-2}, X^2_{1:t-2}, U^1_{1:t-2}, U^2_{1:t-2})$
  - Local data: $L^i_t = (X^i_{t-1}, X^i_t, U^i_{t-1})$.
- A three step approach:
  1. Consider a coordinated system
  2. Show that the coordinated system is equivalent to the original system
  3. Simplify the coordinated system

Applications

- Delayed sharing info structure (Open problem for 40 years) [Nayyar Mahajan Teneketzis 2011]
- real-time communication, feedback communication, multi-user communication, decentralized sequential hypothesis testing, multiaccess broadcast, active sensing, . . .

Future directions

- Randomized decision rules
- Unknown model


Thank you