Separation result for delayed sharing information structures

Aditya Mahajan

McGill University

Joint work with: Ashutosh Nayyar and Demos Teneketzis, UMichigan

Queen's University, July 12, 2011

Common theme: multi-stage multi-agent decision making under uncertainty

Outline of this talk

What is the conceptual difficulty with
multi-agent decision making? How to resolve it?

- Modeling decision making under uncertainty
- Overview of single-agent decision making
- Delayed sharing information structure:
  A “simple” model for multi-agent decision making
  - History of the problem
  - Our approach
  - Main results
- Conclusion

Modeling multi-stage decision making under uncertainty

Model of uncertainty

Model of information

Model of objective

Modeling multi-stage decision making under uncertainty

Model of uncertainty

- Stochastic dynamics

$$X_{t+1} = f\left(X_t, U_t, W_t\right)$$

- Noisy observations

$$Y_t = h\left(X_t, N_t\right)$$

- State disturbance and noise are i.i.d. stochastic processes with known distribution.

- System dynamics f and observation function h are known.

Model of information

Model of objective

Modeling multi-stage decision making under uncertainty

Model of uncertainty

- Stochastic dynamics

$$X_{t+1} = f(X_t, U_t, W_t)$$

- Noisy observations

$$Y_t = h(X_t, N_t)$$

- State disturbance and noise are i.i.d. stochastic processes with known distribution.

- System dynamics f and observation function h are known.

Model of information

- Single DM with perfect recall

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

Model of objective

Modeling multi-stage decision making under uncertainty

Model of uncertainty

- Stochastic dynamics

$$X_{t+1} = f(X_t, U_t, W_t)$$

- Noisy observations

$$Y_t = h(X_t, N_t)$$

- State disturbance and noise are i.i.d. stochastic processes with known distribution.

- System dynamics f and observation function h are known.

Model of information

- Single DM with perfect recall

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

Model of objective

- Cost at time t = $c(X_t, U_t)$.

- Objective: minimize expected total cost

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

More general setups

Model of uncertainty

- Non-i.i.d. dynamics (Markov, ergodic, etc.)
- Unknown distribution
- Unknown model, unknown cost, etc.

Model of information

- Fixed memory/complexity at decision maker
- More than one decision maker

Model of objective

- Worse-case performance (instead of expected performance)
- Minimize regret (instead of minimizing total cost)
- Remain in a desirable set (rather than minimize total cost)

More general setups

- Model of uncertainty
  - Non-i.i.d. dynamics (Markov, ergodic, etc.)
  - Unknown distribution
  - Unknown model, unknown cost, etc.
- Model of information
  - Fixed memory/complexity at decision maker
  - More than one decision maker
- Model of objective
  - Worse-case performance (instead of expected performance)
  - Minimize regret (instead of minimizing total cost)
  - Remain in a desirable set (rather than minimize total cost)

Outline of this talk

What is the conceptual difficulty with
multi-agent decision making? How to resolve it?

- Modeling decision making under uncertainty
- Overview of single-agent decision making
- Delayed sharing information structure:
  A "simple" model for multi-agent decision making
  - History of the problem
  - Our approach
  - Main results
- Conclusion

Single-agent decision making

Design difficulties

- $U_t = g_t(Y_{1:t}, U_{1:t-1})$
  Domain of control laws increases with time

- $\min_{(g_1,\dots,g_T)} \mathbb{E}\Big[\sum_{t=1}^{T} c(X_t, U_t)\Big]$
  Search of optimal control policy is a functional optimization problem

Single-agent decision making

Design difficulties

- $U_t = g_t(Y_{1:t}, U_{1:t-1})$
  Domain of control laws increases with time

- $\min_{(g_1,\dots,g_T)} \mathbb{E}\Big[\sum_{t=1}^{T} c(X_t, U_t)\Big]$
  Search of optimal control policy is a functional optimization problem

Structural results

Define, information state:

$$\pi_t = \mathbb{P}(X_t \mid Y_{1:t}, U_{1:t-1})$$

Then, there is no loss of optimality in restricting attention to control laws of the form

$$U_t = g_t(\pi_t)$$

Single-agent decision making

Design difficulties

- $U_t = g_t(Y_{1:t}, U_{1:t-1})$
  Domain of control laws increases with time

- $\min_{(g_1,\dots,g_T)} \mathbb{E}\Big[\sum_{t=1}^{T} c(X_t, U_t)\Big]$
  Search of optimal control policy is a functional optimization problem

Structural results

Define, information state:

$$\pi_t = \mathbb{P}(X_t \mid Y_{1:t}, U_{1:t-1})$$

Then, there is no loss of optimality in restricting attention to control laws of the form

$$U_t = g_t(\pi_t)$$

Dynamic Programming

The following recursive equations provide an optimal control policy

$$V_t(\pi_t) = \min_{U_t} \mathbb{E}\Big[c(X_t, U_t) + V_{t+1}(\pi_{t+1}) \,\Big|\, \pi_t, U_t\Big]$$

Single-agent decision making

Estimation

Structural results

Define, information state:

$$\pi_t = \mathbb{P}(X_t \mid Y_{1:t}, U_{1:t-1})$$

Then, there is no loss of optimality in restricting attention to control laws of the form

$$U_t = g_t(\pi_t)$$

Control

Dynamic Programming

The following recursive equations provide an optimal control policy

$$V_t(\pi_t) = \min_{U_t} \mathbb{E}\Big[c(X_t, U_t) + V_{t+1}(\pi_{t+1}) \,\Big|\, \pi_t, U_t \Big]$$

Single-agent decision making

- π_t is policy independent

Estimation

Control

Structural results

Define, information state:

$$\pi_t = \mathbb{P}(X_t \mid Y_{1:t}, U_{1:t-1})$$

Then, there is no loss of optimality in restricting attention to control laws of the form

$$U_t = g_t(\pi_t)$$

Dynamic Programming

The following recursive equations provide an optimal control policy

$$V_t(\pi_t) = \min_{U_t} \mathbb{E}\Big[c(X_t, U_t) + V_{t+1}(\pi_{t+1}) \,\Big|\, \pi_t, U_t\Big]$$

Single-agent decision making

- π_t is policy independent

Estimation

Structural results

Define, information state:

$$\pi_t = \mathbb{P}(X_t \mid Y_{1:t}, U_{1:t-1})$$

Then, there is no loss of optimality in restricting attention to control laws of the form

$$U_t = g_t(\pi_t)$$

- Each step of DP is a parameter optimization.

Control

Dynamic Programming

The following recursive equations provide an optimal control policy

$$V_t(\pi_t) = \min_{U_t} \mathbb{E}\Big[c(X_t, U_t) + V_{t+1}(\pi_{t+1}) \,\Big|\, \pi_t, U_t\Big]$$

(One-way) separation between estimation and control

In single-agent decision making, estimation is separated from control. This separation is critical for decomposing the search of optmial control policy into a sequence of parameter optimization problems.

Does this separation extend to multi-agent decision making?

Outline of this talk

What is the conceptual difficulty with
multi-agent decision making? How to resolve it?

- Modeling decision making under uncertainty
- Overview of single-agent decision making
- Delayed sharing information structure:
  A "simple" model for multi-agent decision making
  - History of the problem
  - Our approach
  - Main results
- Conclusion

Delayed-sharing information structure

Model of uncertainty

Model of objective

Model of information

Delayed-sharing information structure

Model of uncertainty

- Stochastic dynamics

$$X_{t+1} = f\left(X_t, U_t^{1:2}, W_t\right)$$

- Noisy observations

$$Y_t^k = h^k\left(X_t, N_t^k\right)$$

Model of information

Model of objective

Delayed-sharing information structure

Model of uncertainty

- Stochastic dynamics

$$X_{t+1} = f\left(X_t, U_t^{1:2}, W_t\right)$$

- Noisy observations

$$Y_t^k = h^k\left(X_t, N_t^k\right)$$

Model of information

- Each DM has perfect recall
- Each DM observes n-step delayed information (observation and actions) of other DMs

Model of objective

Delayed-sharing information structure

Model of uncertainty

- Stochastic dynamics

$$X_{t+1} = f\left(X_t, U_t^{1:2}, W_t\right)$$

- Noisy observations

$$Y_t^k = h^k\left(X_t, N_t^k\right)$$

Model of information

- Each DM has perfect recall
- Each DM observes n-step delayed information (observation and actions) of other DMs

$$U_t^1 = g_t^1\begin{pmatrix} Y_{1:t}^1, & U_{1:t-1}^1 \\ Y_{1:t-n}^2, & U_{1:t-n}^2 \end{pmatrix}$$

Model of objective

Delayed-sharing information structure

Model of uncertainty

- Stochastic dynamics

$$X_{t+1} = f\left(X_t, U_t^{1:2}, W_t\right)$$

- Noisy observations

$$Y_t^k = h^k\left(X_t, N_t^k\right)$$

Model of information

- Each DM has perfect recall
- Each DM observes n-step delayed information (observation and actions) of other DMs

Model of objective

- Cost at time t = $c\left(X_t, U_t^{1:2}\right)$.
- Objective: minimize

$$\mathbb{E}\Big[\sum_{t=1}^{T} c_t(X_t, U_t^{1:2})\Big]$$

$$U_t^1 = g_t^1\begin{pmatrix} Y_{1:t}^1, & U_{1:t-1}^1 \\ Y_{1:t-n}^2, & U_{1:t-n}^2 \end{pmatrix}$$

Delayed-sharing information structure

Some Notation

$$U_t^1 = g_t^1\begin{pmatrix} Y^1_{1:t}, & U^1_{1:t-1} \\ Y^2_{1:t-n}, & U^2_{1:t-n} \end{pmatrix} \qquad U_t^2 = g_t^1\begin{pmatrix} Y^1_{1:t-n}, & U^1_{1:t-n} \\ Y^2_{1:t}, & U^2_{1:t-1} \end{pmatrix}$$

Thus,

$$U_t^k = g_t^k\left(C_t, L_t^k\right)$$

where

- Common info $C_t = (Y^{1:2}_{1:t-n}, U^{1:2}_{1:t-n})$
- Local info $L^k_t = (Y^k_{t-n+1:t}, U^k_{t-n+1:t-n})$

Delayed-sharing information structure

Some Notation

$$U_t^1 = g_t^1\begin{pmatrix} Y^1_{1:t}, & U^1_{1:t-1} \\ Y^2_{1:t-n}, & U^2_{1:t-n} \end{pmatrix} \qquad U_t^2 = g_t^1\begin{pmatrix} Y^1_{1:t-n}, & U^1_{1:t-n} \\ Y^2_{1:t}, & U^2_{1:t-1} \end{pmatrix}$$

Thus,

$$U_t^k = g_t^k\left(C_t, L_t^k \right)$$

where

- Common info $C_t = (Y^{1:2}_{1:t-n}, U^{1:2}_{1:t-n})$
- Local info $L_t^k = (Y^k_{t-n+1:t}, U^k_{t-n+1:t-n})$

Same design difficulties as single-agent case

Literature overview

- (Witsenhausen, 1971): Proposed delayed-sharing information structure. Asserted a structure of optimal control law (without proof).

Literature overview

- (Witsenhausen, 1971): Proposed delayed-sharing information structure. Asserted a structure of optimal control law (without proof).
- (Varaiya and Walrand, 1978): Proved Witsenhausen's assertion for $n = 1$. Showed via a counter-example that the assertion is false for delay $n > 2$.

Literature overview

- (Witsenhausen, 1971): Proposed delayed-sharing information structure. Asserted a structure of optimal control law (without proof).
- (Varaiya and Walrand, 1978): Proved Witsenhausen's assertion for $n = 1$. Showed via a counter-example that the assertion is false for delay $n > 2$.
- (Nayyar, Mahajan, and Teneketzis, 2011): Prove two alternative structures of optimal control law.

Literature overview

- (Witsenhausen, 1971): Proposed delayed-sharing information structure. Asserted a structure of optimal control law (without proof).
- (Varaiya and Walrand, 1978): Proved Witsenhausen's assertion for $n = 1$. Showed via a counter-example that the assertion is false for delay $n > 2$.
- (Nayyar, Mahajan, and Teneketzis, 2011): Prove two alternative structures of optimal control law.
- NMT 2011 also obtain a recursive algorithm to find optimal control laws. At each step, we need to solve a functional optimization problem.

Structure of optimal control law

Original setup

$$U_t^k = g_t^k(C_t, L_t^k)$$

Structure of optimal control law

Original setup

$$U_t^k = g_t^k(C_t, L_t^k)$$

W71 Assertion

$$U_t^k = g_t^k(\mathbb{P}(X_{t-n+1} \mid C_t), L_t^k)$$

Structure of optimal control law

Original setup

$$U_t^k = g_t^k(C_t, L_t^k)$$

W71 Assertion

$$U_t^k = g_t^k(\mathbb{P}(X_{t-n+1} \mid C_t), L_t^k)$$

NMT11 First result

$$U_t^k = g_t^k(\mathbb{P}^g(X_t, L_t^{1:2} \mid C_t), L_t^k)$$

Structure of optimal control law

Original setup

$$U_t^k = g_t^k(C_t, L_t^k)$$

W71 Assertion

$$U_t^k = g_t^k(\mathbb{P}(X_{t-n+1} \mid C_t), L_t^k)$$

NMT11 First result

$$U_t^k = g_t^k(\mathbb{P}^g(X_t, L_t^{1:2} \mid C_t), L_t^k)$$

NMT11 Second result

$$U_t^k = g_t^k(\mathbb{P}(X_{t-n+1} \mid C_t), L_t^k, q_{t-n+1:t-1}^{1:2})$$

Structure of optimal control law

Original setup

$$U_t^k = g_t^k(C_t, L_t^k)$$

W71 Assertion

$$U_t^k = g_t^k(\mathbb{P}(X_{t-n+1} \mid C_t), L_t^k)$$

NMT11 First result

$$U_t^k = g_t^k(\mathbb{P}^g(X_t, L_t^{1:2} \mid C_t), L_t^k)$$

NMT11 Second result

$$U_t^k = g_t^k(\mathbb{P}(X_{t-n+1} \mid C_t), L_t^k, q_{t-n+1:t-1}^{1:2})$$

Contrast dependence on policy for the different results.

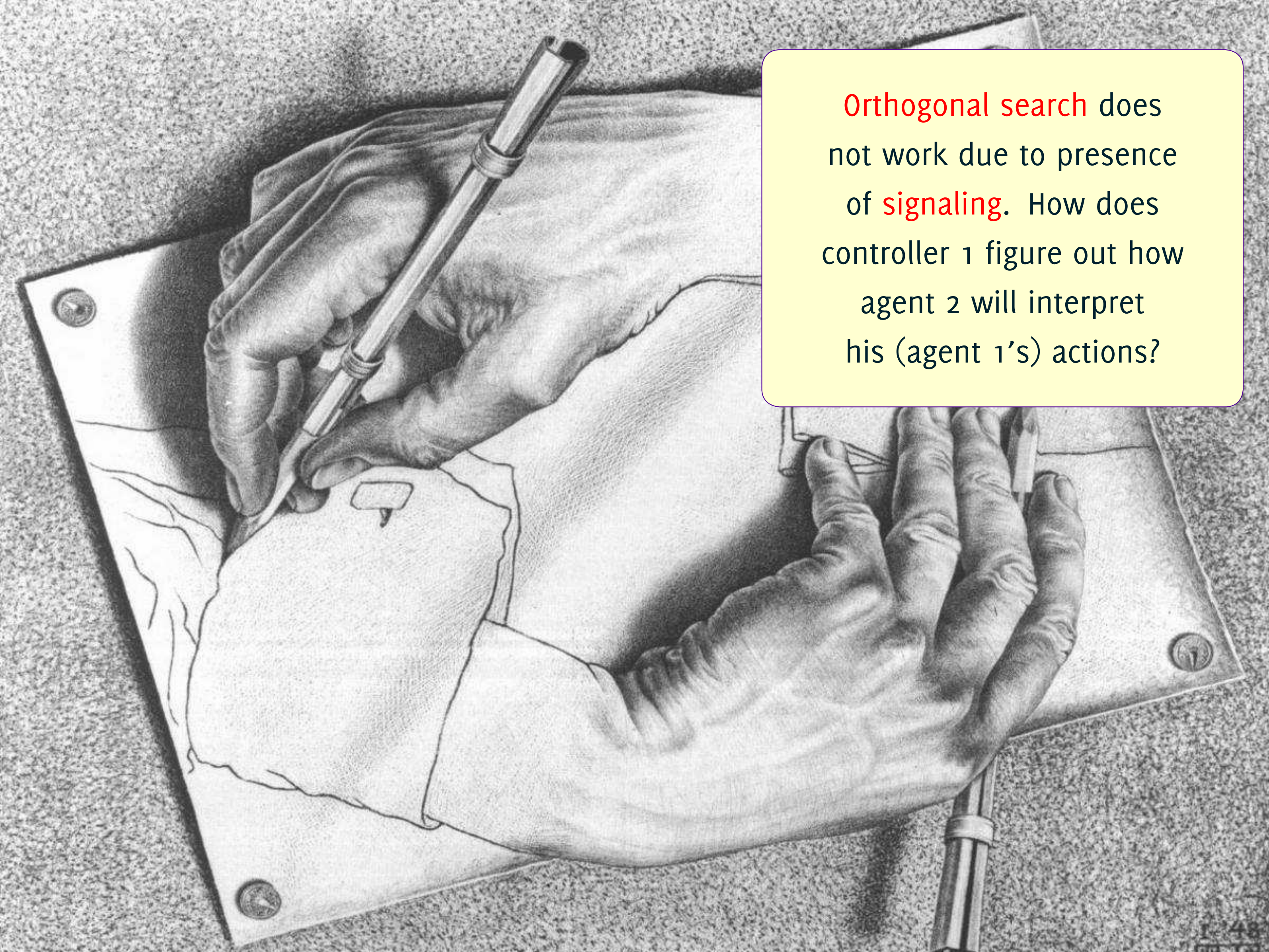
Importance of the problem

- Applications (of one step delay sharing)
  - Power systems: Altman *et. al*, 2009
  - Queueing theory: Kuri and Kumar, 1995
  - Communication networks: Grizzle *et. al*, 1982
  - Stochastic games: Papavassilopoulos, 1982; Chang and Cruz, 1983
  - Economics: Li and Wu, 1991

Importance of the problem

- Applications (of one step delay sharing)
  - Power systems: Altman *et. al*, 2009
  - Queueing theory: Kuri and Kumar, 1995
  - Communication networks: Grizzle *et. al*, 1982
  - Stochastic games: Papavassilopoulos, 1982; Chang and Cruz, 1983
  - Economics: Li and Wu, 1991
- Conceptual significance
  - Understanding the design of networked control systems
  - Bridge between centralized and decentralized systems
  - Insights for the design of general decentralized systems

Orthogonal search does not work due to presence of signaling. How does controller 1 figure out how agent 2 will interpret his (agent 1's) actions?



Orthogonal search does not work due to presence of signaling. How does controller 1 figure out how agent 2 will interpret his (agent 1's) actions?

Proof outline

1. Construct a coordinated system
2. Show that any policy of the coordinated system is implementable in the original system and vice-versa. Hence, the two systems are equivalent.
3. Optimal design of the coordinated system is a single-agent multi-stage decision problem. Find a solution for the coordinated system.
4. Translate this solution back to the original system.

Step 1: The coordinated system

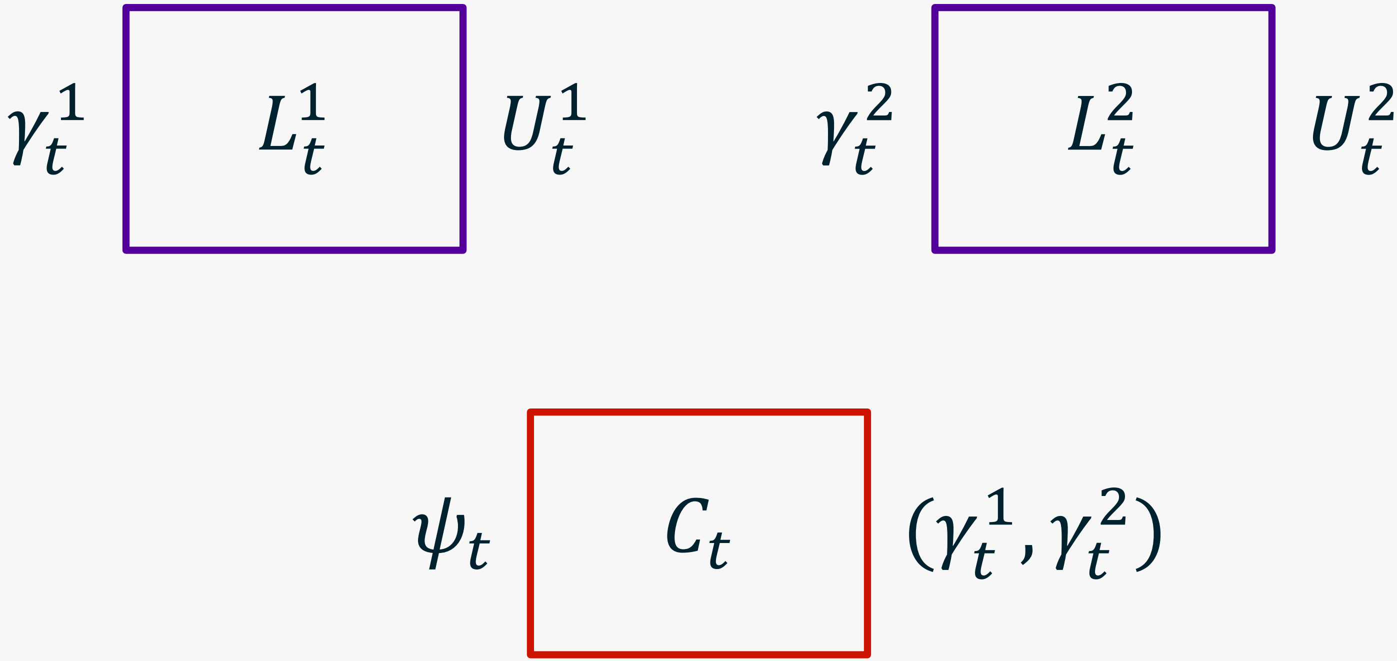
g_t^1 $\boxed{C_t, L_t^1}$ U_t^1

g_t^2 $\boxed{C_t, L_t^2}$ U_t^2

Step 1: The coordinated system

Define partially evaluated control law: $\gamma_t^i(\cdot) = g_t^i(C_t, \cdot)$

Step 1: The coordinated system



Define partially evaluated control law: $\gamma_t^i(\cdot) = g_t^i(C_t, \cdot)$

Coordinator prescribes (γ_t^1, γ_t^2) to the controllers as

$$(\gamma_t^1, \gamma_t^2) = \psi_t(C_t, \gamma_{1:t-1}^1, \gamma_{1:t-1}^2)$$

Step 2: Equivalence

- For any policy $(g_1, \dots, g_T)$ of the original system, we can construct a policy $(\psi_1, \dots, \psi_T)$ of the coordinated system such that the system variables $\{(X_t, Y_t^{1:2}, U_t^{1:2}),\ t = 1, \dots, T\}$ have the same realization along all sample paths in both cases.

Step 2: Equivalence

- For any policy $(g_1, \dots, g_T)$ of the original system, we can construct a policy $(\psi_1, \dots, \psi_T)$ of the coordinated system such that the system variables $\{(X_t, Y_t^{1:2}, U_t^{1:2}), t = 1, \dots, T\}$ have the same realization along all sample paths in both cases.

$$\psi_t(C_t) = (\gamma_t^1, \gamma_t^2) = \big(g_t^1(C_t, \cdot), g_t^2(C_t, \cdot)\big)$$

Step 2: Equivalence

- For any policy $(g_1, \dots, g_T)$ of the original system, we can construct a policy $(\psi_1, \dots, \psi_T)$ of the coordinated system such that the system variables $\{(X_t, Y_t^{1:2}, U_t^{1:2}),\ t = 1, \dots, T\}$ have the same realization along all sample paths in both cases.

$$\psi_t(C_t) = (\gamma_t^1, \gamma_t^2) = \big(g_t^1(C_t, \cdot), g_t^2(C_t, \cdot)\big)$$

- For any policy $(\psi_1, \dots, \psi_T)$ of the coordinated system, we can construct a policy $(g_1, \dots, g_T)$ of the original system such that the system variables $\{(X_t, Y_t^{1:2}, U_t^{1:2}),\ t = 1, \dots, T\}$ have the same realization along all sample paths in both cases.

Step 2: Equivalence

- For any policy $(g_1, \dots, g_T)$ of the original system, we can construct a policy $(\psi_1, \dots, \psi_T)$ of the coordinated system such that the system variables $\{(X_t, Y_t^{1:2}, U_t^{1:2}), t = 1, \dots, T\}$ have the same realization along all sample paths in both cases.

$$\psi_t(C_t) = (\gamma_t^1, \gamma_t^2) = \big(g_t^1(C_t, \cdot), g_t^2(C_t, \cdot)\big)$$

- For any policy $(\psi_1, \dots, \psi_T)$ of the coordinated system, we can construct a policy $(g_1, \dots, g_T)$ of the original system such that the system variables $\{(X_t, Y_t^{1:2}, U_t^{1:2}), t = 1, \dots, T\}$ have the same realization along all sample paths in both cases.

  - At time 1, both controllers know C_1. Choose

$$g_1^k(C_1, L_1^k) = \psi_1^k(C_1)(L_1^k).$$

  - At time 2, both controllers knows C_2, γ_1^1, and γ_1^2. Choose

$$g_2^k(C_2, L_2^k) = \psi_2^k(C_2, \gamma_1^1, \gamma_1^2)(L_2^k).$$

Step 3: Solve the coordinated system

- By construction, the coordinated system has a single decision maker with perfect recall.

- Use result for single-agent decision making:
  Define:

$$\pi_t = \mathbb{P}(\text{“Current state”} \mid \text{past history})$$

  Then, there is no loss of optimality in restricting attention to control laws of the form:

$$\text{control action} = \text{Fn}(\pi_t)$$

Step 3: Solve the coordinated system

- By construction, the coordinated system has a single decision maker with perfect recall.
- Use result for single-agent decision making:
  Define:

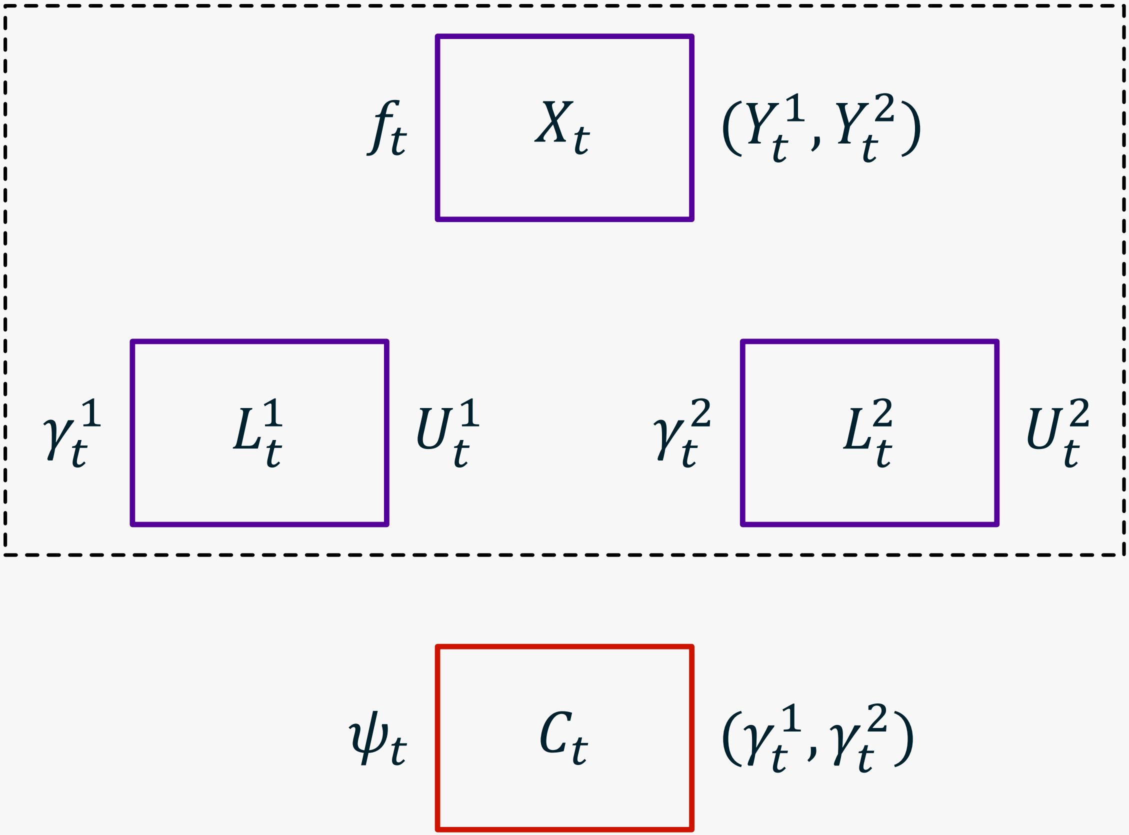
$$\pi_t = \mathbb{P}(\text{"Current state"} \mid \text{past history})$$

  Then, there is no loss of optimality in restricting attention to control laws of the form:

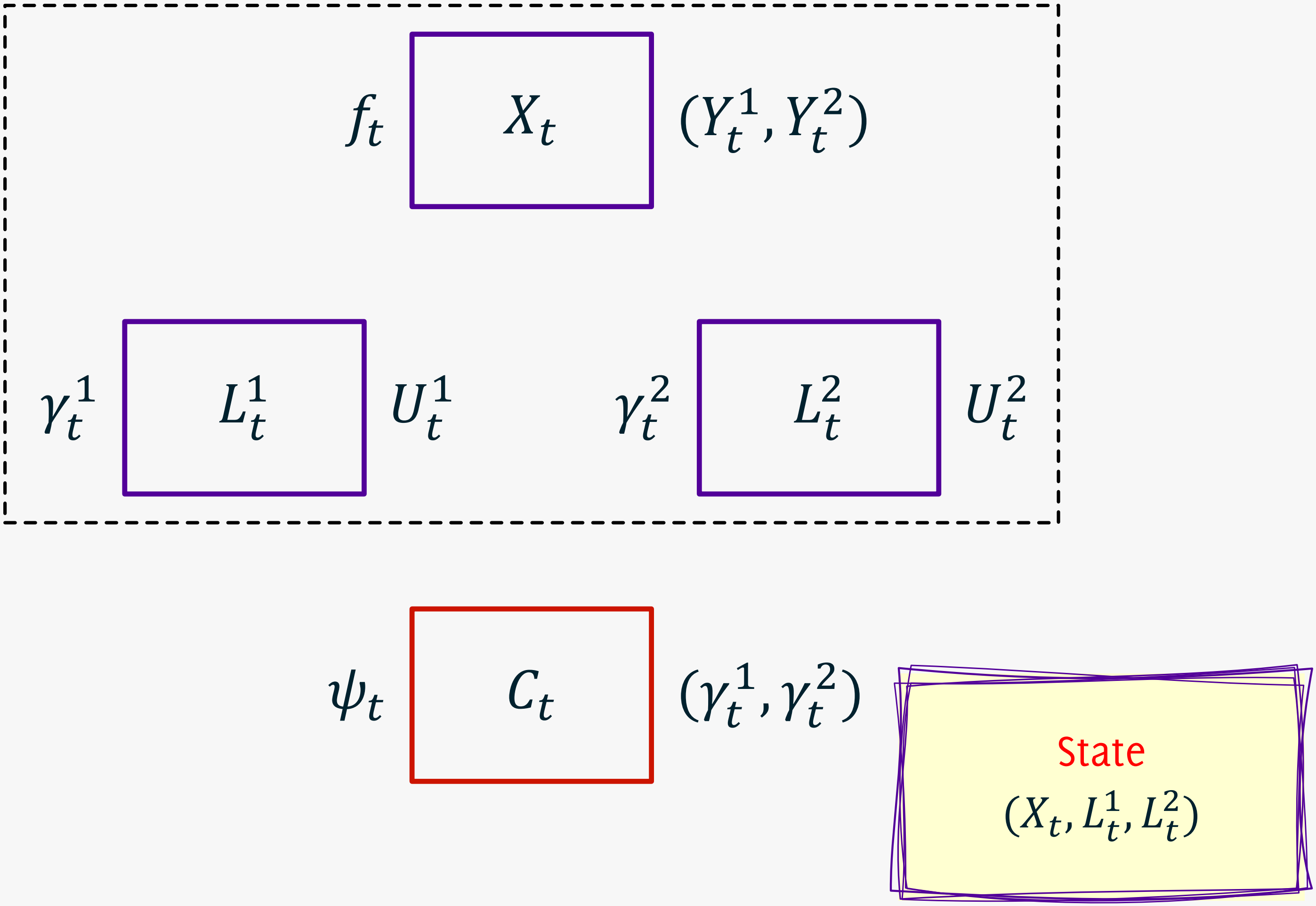
$$\text{control action} = \text{Fn}(\pi_t)$$

What is the state (for I/O mapping) for the system.

State for the coordinated system



State for the coordinated system



Structure of optimal control law

- Define

$$\pi_t = \mathbb{P}(X_t, L^1_t, L^2_t \mid C_t, \gamma^1_{1:t-1}, \gamma^2_{1:t-1})$$

Then, there is no loss of optimality in restricting attention to coordination laws of the form

$$(\gamma^1_t, `\gamma^2_t) = \psi_t(\pi_t)$$

Structure of optimal control law

- Define

$$\pi_t = \mathbb{P}(X_t, L^1_t, L^2_t \mid C_t, \gamma^1_{1:t-1}, \gamma^2_{1:t-1})$$

Then, there is no loss of optimality in restricting attention to coordination laws of the form

$$(\gamma^1_t, `\gamma^2_t) = \psi_t(\pi_t)$$

- The following recursive equations provide an optimal coordination policy

$$V_t(\pi_t) = \min_{(\gamma^1_t, \gamma^2_t)} \mathbb{E}\Big[c(X_t, U_t) + V_{t+1}(\pi_{t+1}) \,\Big|\, \pi_t, \gamma^1_t, \gamma^2_t \Big]$$

Step 4: Translate the solution

For a system with delayed-sharing information structure, there is no loss of optimality in restricting attention to control laws of the form

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

Optimal control laws can be obtained by the solution of the following recursive equations

$$V_t(\pi_t) = \min_{(g^1(\pi_t,\cdot), g^2(\pi_t,\cdot))} \mathbb{E}\Big[c(X_t, U_t) + V_{t+1}(\pi_{t+1}) \,\Big|\, \pi_t, g_t^1(\pi_t,\cdot), g_t^2(\pi_t,\cdot)\Big]$$

Features of the solution

- The space of realizations of $\pi_t = \mathbb{P}(X_t, L_t^1, L_t^2 \mid C_t, \gamma_{1:t-1}^1, \gamma_{1:t-1}^2)$ is time-invariant. Thus, the domain of the control laws $g_t^k(\pi_t, L_t^k)$ is time-invariant.
- π_t is not policy independent! Estimation is not separated from control. This is always the case when signaling is present.
- In each step of the dynamic program, we choose the partially evaluated control laws $g_t^1(\pi_t, \cdot)$, $g_t^2(\pi_t, \cdot)$. Choosing partially evaluated functions (instead of values) allows us to write a dynamic program even in the presence of signaling.

Outline of this talk

What is the conceptual difficulty with
multi-agent decision making? How to resolve it?

- Modeling decision making under uncertainty
- Overview of single-agent decision making
- Delayed sharing information structure:
  A “simple” model for multi-agent decision making
  - History of the problem
  - Our approach
  - Main results
- Conclusion

Summary

- Simple methodology to resolve a 40 year old open question:
  - Find common information at each time
  - Look at the problem for the point of view of a coordinator that observes this common information and chooses partiallly evaluated functions
  - Find an information state for the problem at the coordinator
    - $\mathbb{P}$(state for input-output mapping | common information)
    - ($\mathbb{P}$(past state | common information), past partial control laws)
- This methodology is also applicable to systems with more general information structures (Mahajan, Nayyar, Teneketzis, 2008).

Salient Features

- The size of the information state is time-invariant

> The methodology is also applicable to infinite horizon problems

- Each step of DP is a functional optimization problem
  - Form of the DP is similar to that of POMDP
  - Can borrow from the POMDP literature for numerical approaches