Dynamic programming in infinite horizon dynamic teams with non-classical information structure

Aditya Mahajan

McGill University

Fifth Workshop on Dynamic Games in Management Science, GERAD, Montreal

28 Nov, 2014

Joint work with: Ashutosh Nayyar and Demosthenis Teneketzis

What are dynamic teams

Dynamic games where all players have identical payoff function

What are dynamic teams

Dynamic games where all players have identical payoff function

Literature overview

- Economics Literature
  - Radner, "Team decision problems," Ann Math Stat, 1962.
  - Marschak and Radner, "Economics Theory of Teams," 1972.
  - . . .
- Systems & Control Literature
  - Witsenhausen, "Separation of estimation and control," Proc IEEE, 1971.
  - Witsenhausen, "On information structures, feedback and causality," SICON 1971.
  - Ho and Chu, "Team decision theory and information structures," IEEE TAC 1972.
  - . . .

What are dynamic teams

Dynamic games where all players have identical payoff function

Literature overview

- Economics Literature
  - Radner, "Team decision problems," Ann Math Stat, 1962.
  - Marschak and Radner, "Economics Theory of Teams," 1972.
  - . . .
- Systems & Control Literature
  - Witsenhausen, "Separation of estimation and control," Proc IEEE, 1971.
  - Witsenhausen, "On information structures, feedback and causality," SICON 1971.
  - Ho and Chu, "Team decision theory and information structures," IEEE TAC 1972.
  - . . .

Simpler than non-cooperative game theory.

All "pre-game" agreements are enforceable.

Simpler than cooperative game theory.

The value of the game does not need to be split between the players.

What are dynamic teams

Dynamic games where all players have identical payoff function

Main difficulties:

- Information decentralization

  Non-classical information structure

- Seeking global optimality

1.

ICON 1971.

TAC 1972.

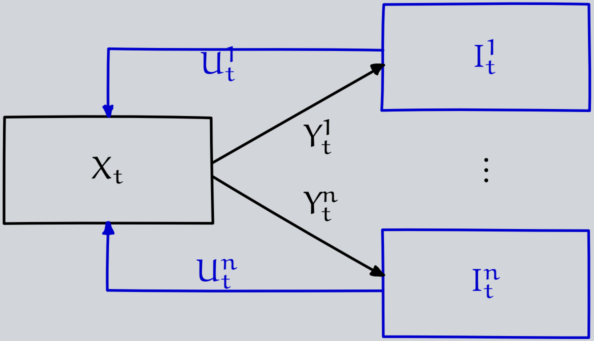
Simpl

All “pre-game” agreements are enforceable.

Simpler than cooperative game theory.

The value of the game does not need to be split between the players.

Simplest general model of a dynamic team



Dynamics $X_{t+1} = f_t(X_t, \mathbf{U}_t, W_t^0)$, where $\mathbf{U}_t = (U_t^1, \dots, U_t^n)$.

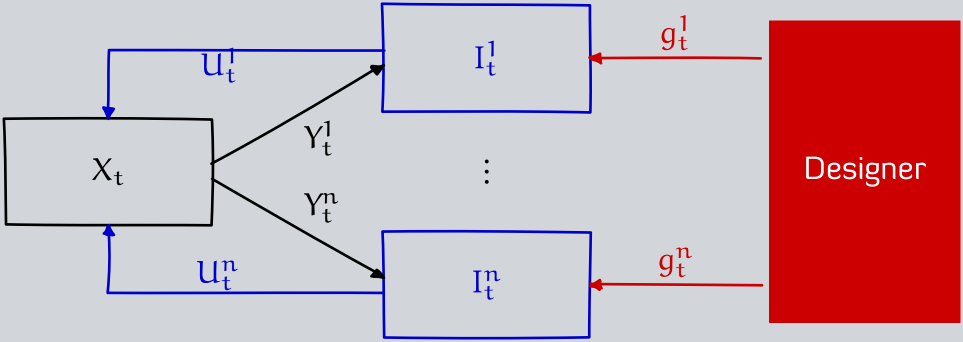
Observation $Y_t^i = h_t^i(X_t, W_t^i)$.

Information structure $\boxed{\{Y_{1:t}^i, U_{1:t-1}^i\} \subseteq I_t^i \subseteq \{\mathbf{Y}_{1:t}, \mathbf{U}_{1:t-1}\}, \quad U_t^i = g_t^i(I_t^i).}$

Control Strategy $\mathbf{g} = (\mathbf{g}^1, \dots, \mathbf{g}^n)$, where $\mathbf{g}^i = (g_1^i, g_2^i, \dots)$.

Performance
- Per-step reward $R_t = \rho(X_t, \mathbf{U}_t)$.
- $\boxed{J(\mathbf{g}) = \mathbb{E}^{\mathbf{g}}\left[\sum_{t=0}^{\infty} \beta^t R_t\right]}$

Simplest general model of a dynamic team



Dynamics $X_{t+1} = f_t(X_t, \mathbf{U}_t, W_t^0)$, where $\mathbf{U}_t = (U_t^1, \dots, U_t^n)$.

Observation $Y_t^i = h_t^i(X_t, W_t^i)$.

Information structure

$$\boxed{\{Y_{1:t}^i, U_{1:t-1}^i\} \subseteq I_t^i \subseteq \{\mathbf{Y}_{1:t}, \mathbf{U}_{1:t-1}\}, \quad U_t^i = g_t^i(I_t^i).}$$

Control Strategy $\mathbf{g} = (\mathbf{g}^1, \dots, \mathbf{g}^n)$, where $\mathbf{g}^i = (g_1^i, g_2^i, \dots)$.

Performance
- Per-step reward $R_t = \rho(X_t, \mathbf{U}_t)$.
- $$\boxed{J(\mathbf{g}) = \mathbb{E}^{\mathbf{g}}\left[\sum_{t=0}^{\infty} \beta^t R_t\right]}$$

Conceptual difficulties

The optimal control problem is a functional optimization problem where we have to choose an infinite sequence of control laws $\mathbf{g}$ to maximize the expected total reward.

The domain I^i_t of control law g^i_t increases with time.

- Can the optimization problem be solved?
- Can we implement the optimal solution?

Agent based methods lead to infinite regress.

Signaling (or the communication aspect of control)

Centralized stochastic control: Information state

$$\boxed{I_t \subseteq I_{t+1}}$$

Centralized stochastic control: Information state

$$\boxed{I_t \subseteq I_{t+1}}$$

A process $\{Z_t\}_{t=0}^{\infty}$ is called an information state if

- **Function of available information**

  There exists a series of functions $\{F_t\}_{t=0}^{\infty}$ such that $Z_t = f_t(I_t)$.

- **Absorbs the effect of available information on current rewards**

$$\mathbb{P}(R_t \in \mathcal{B} \mid I_t = i_t, U_t = u_t) = \mathbb{P}(R_t \in \mathcal{B} \mid Z_t = F_t(i_t), U_t = u_t).$$

- **Controlled Markov property**

$$\mathbb{P}(Z_{t+1} \in \mathcal{A} \mid I_t = i_t, U_t = u_t) = \mathbb{P}(Z_{t+1} \in \mathcal{A} \mid Z_t = F_t(i_t), U_t = u_t).$$

Examples: ▸ System state in MDPs ▸ Belief state in POMDPs

Centralized control: Structure of optimal strategies

The information state absorbs the effect of available information on expected future cost, i.e., for any choice of future strategy $g_{(t)} = (g_{t+1}, g_{t+2}, \dots)$

$$\mathbb{E}^{g_{(t)}}\left[\sum_{\tau=t}^{\infty} \beta^{\tau} R_{\tau} \,\middle|\, I_t = i_t, U_t = u_t\right] = \mathbb{E}^{g_{(t)}}\left[\sum_{\tau=t}^{\infty} \beta^{\tau} R_{\tau} \,\middle|\, Z_t = F_t(i_t), U_t = u_t\right].$$

Centralized control: Structure of optimal strategies

The information state absorbs the effect of available information on expected future cost, i.e., for any choice of future strategy $g_{(t)} = (g_{t+1}, g_{t+2}, \dots)$

$$\mathbb{E}^{g_{(t)}}\left[\sum_{\tau=t}^{\infty}\beta^{\tau}R_{\tau}\,\middle|\,I_t = i_t, U_t = u_t\right] = \mathbb{E}^{g_{(t)}}\left[\sum_{\tau=t}^{\infty}\beta^{\tau}R_{\tau}\,\middle|\,Z_t = F_t(i_t), U_t = u_t\right].$$

Therefore,

- Z_t is a sufficient statistic for performance evaluation,
- there is no loss of optimality is using control laws of the form $g_t\colon Z_t \mapsto U_t$

Centralized control: Structure of optimal strategies

The information state absorbs the effect of available information on expected future cost, i.e., for any choice of future strategy $g_{(t)} = (g_{t+1}, g_{t+2}, \dots)$

$$\mathbb{E}^{g_{(t)}}\left[\sum_{\tau=t}^{\infty} \beta^\tau R_\tau \,\middle|\, I_t = i_t, U_t = u_t\right] = \mathbb{E}^{g_{(t)}}\left[\sum_{\tau=t}^{\infty} \beta^\tau R_\tau \,\middle|\, Z_t = F_t(i_t), U_t = u_t\right].$$

Therefore,

- Z_t is a sufficient statistic for performance evaluation,
- there is no loss of optimality is using control laws of the form $g_t \colon Z_t \mapsto U_t$

Examples

- In MDPs, $g_t \colon X_t \mapsto U_t$.
- In POMDPs, $g_t \colon B_t \mapsto U_t$, where B_t is the belief state.

Centralized control: Dynamic programming

For any strategy $\mathbf{g}$ of the form $g_t \colon Z_t \mapsto U_t$,

$$\mathbb{E}^{\mathbf{g}(t)}\left[\mathbb{E}^{\mathbf{g}(t+1)}\left[\sum_{\tau=t+1}^{\infty}\beta^{\tau}R_{\tau}\,\middle|\, Z_{t+1}, U_{t+1} = g_{t+1}(Z_{t+1})\right]\,\middle|\, Z_t = z_t, U_t = u_t\right]$$

$$= \mathbb{E}^{\mathbf{g}(t)}\left[\sum_{\tau=t+1}^{\infty}\beta^{\tau}R_{\tau}\,\middle|\, Z_t = z_t, U_t = u_t\right]$$

Relies on $\mathbf{I}_t \subseteq \mathbf{I}_{t+1}$

Centralized control: Dynamic programming

For any strategy $\mathbf{g}$ of the form $g_t \colon Z_t \mapsto U_t$,

$$\mathbb{E}^{\mathbf{g}(t)}\left[\mathbb{E}^{\mathbf{g}(t+1)}\left[\sum_{\tau=t+1}^{\infty}\beta^{\tau}R_{\tau}\,\middle|\, Z_{t+1}, U_{t+1} = g_{t+1}(Z_{t+1})\right]\,\middle|\, Z_t = z_t, U_t = u_t\right]$$

$$= \mathbb{E}^{\mathbf{g}(t)}\left[\sum_{\tau=t+1}^{\infty}\beta^{\tau}R_{\tau}\,\middle|\, Z_t = z_t, U_t = u_t\right] \qquad \text{Relies on } \mathbf{I}_t \subseteq \mathbf{I}_{t+1}$$

There exists a time-homogeneous optimal strategy $\mathbf{g}^* = (g^*, g^*, \dots)$ that is given by the fixed point of the following dynamic program

$$V(z) = \min_{u \in \mathcal{U}} \mathbb{E}[R_t + \beta V(Z_{t+1}) \mid Z_t = z, U_t = u]$$

Centralized control: Dynamic programming

For any strategy g of the form $g_t : Z_t \mapsto U_t$,

Both these results rely on an appropriate choice of **information state**.

Note that **information state for DP** is also a **sufficient statistic for control**.

Centralized control: Dynamic programming

For any strategy g of the form $g_t \colon Z_t \mapsto U_t$,

- Can we identify a sufficient statistic Z^i_t and restrict attention to $g^i_t \colon Z^i_t \mapsto U^i_t$?
- Can we show that there exist time-homogeneous optimal control strategies?
- Can we identify appropriate information states to determine a dynamic program that computes such optimal strategies?

Two approaches to dynamic programming:
The person-by-person approach

The person-by-person approach

Pick an agent, say i.

Arbitrarily fix the strategies $\mathbf{g}^{-i}$ of all other agents.

Identify an information-state process $\{Z_t^i\}_{t=0}^{\infty}$ for agent i.

Structure of optimal strategies

If $\mathcal{Z}_t^i$, the space of realization of Z_t^i, does not depend on $\mathbf{g}^{-i}$, then there is no loss of optimality in using $g_t^i \colon Z_t^i \mapsto U_t^i$.

- Radner, "Team decision problems," Ann Math Stat, 1962.
- Marschak and Radner, "Economics Theory of Teams," 1972.

The person-by-person approach

Pick an agent, say i.

Arbitrarily fix the strategies $\mathbf{g}^{-i}$ of all other agents.

Identify an information-state process $\{Z_t^i\}_{t=0}^{\infty}$ for agent i.

Structure of optimal strategies: If $\mathcal{Z}_t^i$, the space of realization of Z_t^i, does not depend on $\mathbf{g}^{-i}$, then there is no loss of optimality in using $g_t^i \colon Z_t^i \mapsto U_t^i$.

Write coupled dynamic programs to identify the best response strategy

$$\mathbf{g}^i = \mathcal{D}^i(\mathbf{g}^{-i})$$

Remarks

- Is the best-response strategy time-homogeneous?
- The coupled dynamic program always has a fixed-point.
- Is the fixed point unique?

- Radner, "Team decision problems," Ann Math Stat, 1962.
- Marschak and Radner, "Economics Theory of Teams," 1972.

The person-by-person approach

Pick an agent, say i.

The person-by-person approach:

- May identify the structure of globally optimal control strategies.
- Provides coupled dynamic programs, which, at best, may determine person-by-person optimal control strategies. Such strategies can be arbitrarily bad compared to globally optimal strategies.

Remarks
- Is the best-response strategy time-homogeneous?
- The coupled dynamic program always has a fixed-point.
- Is the fixed point unique?

- Radner, "Team decision problems," Ann Math Stat, 1962.
- Marschak and Radner, "Economics Theory of Teams," 1972.

An example: coupled subsystems with control sharing

Global state $\mathbf{X}_t = (X^1_t, \dots, X^n_t)$

Dynamics $X^i_{t+1} = f^i(X^i_t, \mathbf{U}_t, W^i_t), \quad$ where $\mathbf{U}_t = (U^1_t, \dots, U^n_t)$.

Information structure $\boxed{I^i_t = \{X^i_{1:t}, \mathbf{U}_{1:t-1}\}}$

- Mahajan, "Optimal decentralized control of coupled subsystems with control sharing," IEEE TAC 2013.

An example: coupled subsystems with control sharing

Global state $\mathbf{X}_t = (X^1_t, \dots, X^n_t)$

Dynamics $X^i_{t+1} = f^i(X^i_t, \mathbf{U}_t, W^i_t), \quad$ where $\mathbf{U}_t = (U^1_t, \dots, U^n_t)$.

Information structure

$$\boxed{I^i_t = \{X^i_{1:t}, \mathbf{U}_{1:t-1}\}}$$

Conditional independence

For any arbitrary choice of control strategies $\mathbf{g}$:

$$\mathbb{P}(\mathbf{X}_{1:t} \mid \mathbf{U}_{1:t-1} = \mathbf{u}_{1:t-1}) = \prod_{i=1}^{n} \mathbb{P}(X^i_{1:t} \mid \mathbf{U}_{1:t-1} = \mathbf{u}_{1:t-1})$$

- Mahajan, "Optimal decentralized control of coupled subsystems with control sharing," IEEE TAC 2013.

An example: coupled subsystems with control sharing

Global state $\mathbf{X}_t = (X^1_t, \dots, X^n_t)$

Dynamics $X^i_{t+1} = f^i(X^i_t, \mathbf{U}_t, W^i_t), \quad \text{where } \mathbf{U}_t = (U^1_t, \dots, U^n_t).$

Information structure

$$\boxed{I^i_t = \{X^i_{1:t}, \mathbf{U}_{1:t-1}\}}$$

Conditional independence For any arbitrary choice of control strategies $\mathbf{g}$:

$$\mathbb{P}(\mathbf{X}_{1:t} \mid \mathbf{U}_{1:t-1} = \mathbf{u}_{1:t-1}) = \prod_{i=1}^{n} \mathbb{P}(X^i_{1:t} \mid \mathbf{U}_{1:t-1} = \mathbf{u}_{1:t-1})$$

Structure of optimal strategies

- Arbitrarily fix strategies $\mathbf{g}^{-i}$, and consider the “best-response” strategy at agent i.
- $\{X^i_t, \mathbf{U}_{1:t-1}\}$ is an information-state at agent i.

- Mahajan, “Optimal decentralized control of coupled subsystems with control sharing,” IEEE TAC 2013.

Two approaches to dynamic programming:
The common-information approach

Illustrative example: Two agent static Bayesian team

Observations $I^1 = (C, Y^1)$ and $I^2 = (C, Y^2)$. Joint probability on (X, Y^1, Y^2, C).

Actions $U^1 = g^1(C, Y^1)$ and $U^2 = g^2(C, Y^2)$.

Utility Maximize $J(g^1, g^2) = \mathbb{E}^{(g^1, g^2)}[\rho(X, U^1, U^2)]$.

Illustrative example: Two agent static Bayesian team

Observations $I^1 = (C, Y^1)$ and $I^2 = (C, Y^2)$. Joint probability on (X, Y^1, Y^2, C).

Actions $U^1 = g^1(C, Y^1)$ and $U^2 = g^2(C, Y^2)$.

Utility Maximize $J(g^1, g^2) = \mathbb{E}^{(g^1, g^2)}[\rho(X, U^1, U^2)]$.

Complexity Possibilities for $g^i = |\mathcal{U}^i|^{|\mathcal{C}| \cdot |\mathcal{Y}^i|}$

Illustrative example: Two agent static Bayesian team

Observations $I^1 = (C, Y^1)$ and $I^2 = (C, Y^2)$. Joint probability on (X, Y^1, Y^2, C).

Actions $U^1 = g^1(C, Y^1)$ and $U^2 = g^2(C, Y^2)$.

Utility Maximize $J(g^1, g^2) = \mathbb{E}^{(g^1, g^2)}[\rho(X, U^1, U^2)]$.

Complexity Possibilities for $g^i = |\mathcal{U}^i|^{|\mathcal{C}| \cdot |\mathcal{Y}^i|}$

Partial strategies
- Define a functional-valued function ψ^i such that $g^i(C, Y^i) = \psi^i(C)(Y^i)$.
- We call $\gamma^i_C = \psi^i(C)$ a partial strategy (or prescriptions)

Illustrative example: Two agent static Bayesian team

Observations $I^1 = (C, Y^1)$ and $I^2 = (C, Y^2)$. Joint probability on (X, Y^1, Y^2, C).

Actions $U^1 = g^1(C, Y^1)$ and $U^2 = g^2(C, Y^2)$.

Utility Maximize $J(g^1, g^2) = \mathbb{E}^{(g^1, g^2)}[\rho(X, U^1, U^2)]$.

Complexity Possibilities for $g^i = |\mathcal{U}^i|^{|\mathcal{C}|\cdot|\mathcal{Y}^i|}$

Partial strategies

- Define a functional-valued function ψ^i such that $g^i(C, Y^i) = \psi^i(C)(Y^i)$.
- We call $\gamma^i_C = \psi^i(C)$ a partial strategy (or prescriptions)

The idea of a coordinator

Consider a virtual coordinator that:

- Observes the common information C
- And prescribes the partial strategy (γ^1_C, γ^2_C)
- The decision rule at the coordinator is $\psi \colon C \mapsto (\gamma^1_C, \gamma^2_C)$.

$$\tilde{J}(\psi) = \mathbb{E}^{\psi}[\rho(X, \gamma^1_C(Y^1), \gamma^2_C(Y^2))]$$

Illustrative example: Two agent static Bayesian team

- Centralized optimization problem
  Can be solved by solving looking at the conditional reward
  $\mathbb{E}[\rho(X, \gamma^1_C(Y^1), \gamma^2_C(Y^2) \mid C = c]$.

- Complexity: $|\mathcal{C}| \cdot |\mathcal{U}^1|^{|\mathcal{Y}^1|} \cdot |\mathcal{U}^2|^{|\mathcal{Y}^2|}$

- Contrast from: $|\mathcal{U}^1|^{|\mathcal{Y}^1| \cdot |\mathcal{C}|} \cdot |\mathcal{U}^2|^{|\mathcal{Y}^2| \cdot |\mathcal{C}|}$

(Y^i).

- The decision rule at the coordinator is $\psi \colon C \mapsto (\gamma^1_C, \gamma^2_C)$.

$$\tilde{J}(\psi) = \mathbb{E}^{\psi}[\rho(X, \gamma^1_C(Y^1), \gamma^2_C(Y^2))]$$

One dynamic program to rule them all

$$V(\textcolor{blue}{\blacksquare}) = \min_{\textcolor{red}{\blacksquare}} \mathbb{E}[R_t + \beta V(\textcolor{blue}{\blacksquare}_{t+1}) \mid \textcolor{blue}{\blacksquare}_t = \textcolor{blue}{\blacksquare}, \textcolor{red}{\blacksquare}_t = \textcolor{red}{\blacksquare}]$$

One dynamic program to rule them all

$$V({\color{blue}\blacksquare}) = \min_{{\color{red}\blacksquare}} \mathbb{E}[R_t + \beta V({\color{blue}\blacksquare}_{t+1}) \mid {\color{blue}\blacksquare}_t = {\color{blue}\blacksquare}, {\color{red}\blacksquare}_t = {\color{red}\blacksquare}]$$

- The information state must be a function of the information available to every player.

One dynamic program to rule them all

$$V({\color{blue}\blacksquare}) = \min_{{\color{red}\blacksquare}} \mathbb{E}[R_t + \beta V({\color{blue}\blacksquare}_{t+1}) \mid {\color{blue}\blacksquare}_t = {\color{blue}\blacksquare}, {\color{red}\blacksquare}_t = {\color{red}\blacksquare}]$$

- The information state must be a function of the information available to every player.

Common information: $C_t = \bigcap_{\tau \geqslant t} \bigcap_{i=1}^{n} I^i_\tau,$ Local information: $L^i_t = I^i_t \setminus C_t$

One dynamic program to rule them all

$$V(z) = \min_{\blacksquare} \mathbb{E}[R_t + \beta V(Z_{t+1}) \mid Z_t = z, \blacksquare_t = \blacksquare]$$

- The information state must be a function of the information available to every player.

Common information: $C_t = \bigcap_{\tau \geqslant t} \bigcap_{i=1}^{n} I^i_\tau,$ Local information: $L^i_t = I^i_t \setminus C_t$

One dynamic program to rule them all

$$V(z) = \min_{\blacksquare} \mathbb{E}[R_t + \beta V(Z_{t+1}) \mid Z_t = z, \blacksquare_t = \blacksquare]$$

- The information state must be a function of the information available to every player.

$$\text{Common information: } C_t = \bigcap_{\tau \geqslant t} \bigcap_{i=1}^{n} I^i_\tau, \qquad \text{Local information: } L^i_t = I^i_t \setminus C_t$$

- Each step of the dynamic programming must determine a mapping from $(C_t, L^i_t) \mapsto U^i_t$.
  - The information state Z_t only depends on C_t
  - Thus, the "action" at each step must be a mapping $L^i_t \mapsto U^i_t$. Call it prescription and denote it by γ^i_t.

One dynamic program to rule them all

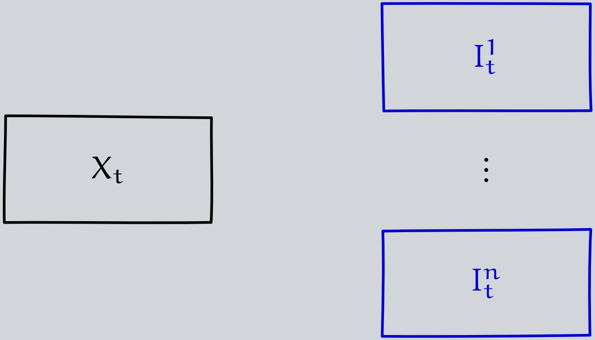
$$V(z) = \min_{\gamma} \mathbb{E}[R_t + \beta V(Z_{t+1}) \mid Z_t = z, \Gamma_t = \gamma]$$

- The information state must be a function of the information available to every player.

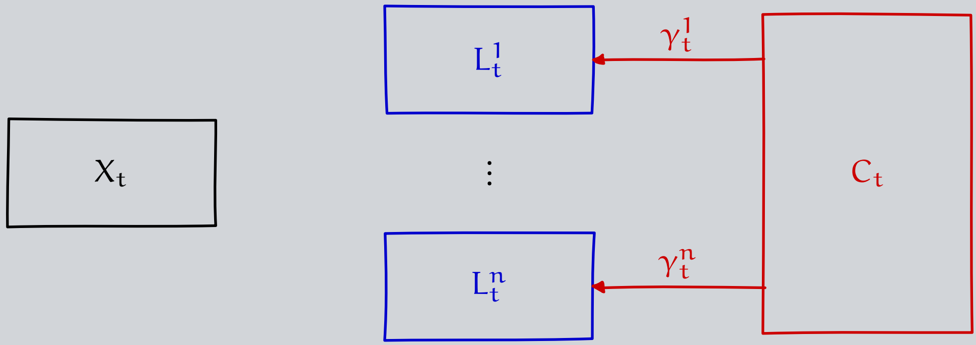
Common information: $C_t = \bigcap_{\tau \geqslant t} \bigcap_{i=1}^{n} I^i_\tau$, Local information: $L^i_t = I^i_t \setminus C_t$

- Each step of the dynamic programming must determine a mapping from $(C_t, L^i_t) \mapsto U^i_t$.
  - The information state Z_t only depends on C_t
  - Thus, the "action" at each step must be a mapping $L^i_t \mapsto U^i_t$. Call it prescription and denote it by γ^i_t.

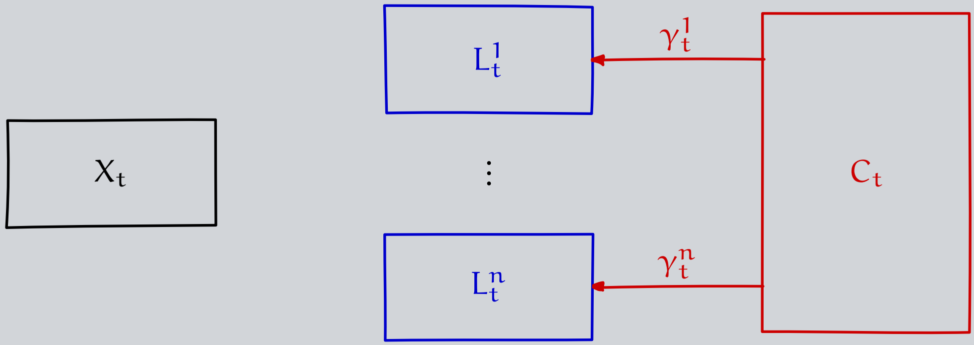
A virtual coordinator



A virtual coordinator



A virtual coordinator



When does this work: Partial history sharing

- $|\mathcal{L}_t^i|$ is uniformly bounded (over i and t) and

$$\mathbb{P}(L_{t+1}^i \in \mathcal{A} \mid C_t, L_t^i, U_t^i, Y_{t+1}^i) = \mathbb{P}(L_{t+1}^i \in \mathcal{A} \mid L_t^i, U_t^i, Y_{t+1}^i)$$

Centralized POMDP

- Information state: $\mathbb{P}(X_t, \mathbf{L}_t \mid C_t = c)$ (or something else)
- "Standard" POMDP results apply, value function is PWLC.
- Subsumes many previous results on DP for decentralized stochastic control.

Example 1: Delayed sharing information structure

Dynamics $X_{t+1} = f_t(X_t, \mathbf{U}_t, W^0_t)$, where $\mathbf{U}_t = (U^1_t, \dots, U^n_t)$.

Observations $Y^i_t = h^i_t(X_t, W^i_t)$.

Information structure $I^i_t = \{Y^i_{1:t}, U^i_{1:t-1}, \mathbf{Y}_{1:t-k}, \mathbf{U}_{1:t-k}\}$. k is the sharing delay.

- Witsenhausen, "Separation of estimation and control," Proc IEEE, 1971.
- Nayyar, Mahajan and Teneketzis, "Optimal control strategies in delayed sharing information structures," IEEE TAC 2011.

Example 1: Delayed sharing information structure

Dynamics $X_{t+1} = f_t(X_t, \mathbf{U}_t, W^0_t)$, where $\mathbf{U}_t = (U^1_t, \dots, U^n_t)$.

Observations $Y^i_t = h^i_t(X_t, W^i_t)$.

Information structure $I^i_t = \{Y^i_{1:t}, U^i_{1:t-1}, \mathbf{Y}_{1:t-k}, \mathbf{U}_{1:t-k}\}$. k is the sharing delay.

Common info.: $C_t = \{\mathbf{Y}_{1:t-k}, \mathbf{U}_{1:t-k}\}$, Local Info.: $L^i_t = I^i_t \setminus C_t$, Pres.: $\Gamma^i_t \colon L^i_t \mapsto U^i_t$

Information State $\Pi_t = \mathbb{P}(X_t, \mathbf{L}_t \mid C_t)$

Results

- No loss of optimality in using decision strategies $g^i_t \colon (L^i_t, \Pi_t) \mapsto U^i_t$.
- Dynamic program: $V(\pi) = \min_{\gamma} \mathbb{E}[R_t + \beta V(\Pi_{t+1}) \mid \Pi_t = \pi, \mathbf{\Gamma}_t = \boldsymbol{\gamma}]$.

- Witsenhausen, "Separation of estimation and control," Proc IEEE, 1971.
- Nayyar, Mahajan and Teneketzis, "Optimal control strategies in delayed sharing information structures," IEEE TAC 2011.

Example 2: Control sharing information structure

Dynamics $X^i_{t+1} = f^i(X^i_t, \mathbf{U}_t, W^i_t)$, where $\mathbf{U}_t = (U^1_t, \dots, U^n_t)$.

Information structure

Original : $I^i_t = \{X^i_{1:t}, \mathbf{U}_{1:t-1}\}$

Using p-by-p approach: $\tilde{I}^i_t = \{X^i_t, \mathbf{U}_{1:t-1}\}$.

- Mahajan, "Optimal decentralized control of coupled subsystems with control sharing," IEEE TAC 2013.

Example 2: Control sharing information structure

Dynamics $X^i_{t+1} = f^i(X^i_t, \mathbf{U}_t, W^i_t)$, where $\mathbf{U}_t = (U^1_t, \dots, U^n_t)$.

Information structure
- Original : $I^i_t = \{X^i_{1:t}, \mathbf{U}_{1:t-1}\}$
- Using p-by-p approach: $\tilde{I}^i_t = \{X^i_t, \mathbf{U}_{1:t-1}\}$.

Common info.: $C_t = \mathbf{U}_{1:t-1}$, Local Info.: $L^i_t = X^i_t$, Prescriptions: $\Gamma^i_t \colon X^i_t \mapsto U^i_t$

Information State
Define $\Xi^i_t(x) = \mathbb{P}(X^i_t = x \mid \mathbf{U}_{1:t-1})$.
Then $\Xi_t = (\Xi^1_t, \dots, \Xi^n_t)$ is an information state.

Results
- No loss of optimality in using decision strategies $g^i_t \colon (X^i_t, \Xi_t) \mapsto U^i_t$.
- Dynamic program: $V(\xi) = \min_\gamma \mathbb{E}[R_t + \beta V(\Xi_{t+1}) \mid \Xi_t = \xi, \Gamma_t = \gamma]$.

- Mahajan, "Optimal decentralized control of coupled subsystems with control sharing," IEEE TAC 2013.

Example 3: Mean-field sharing information structure

Dynamics $X^i_{t+1} = f_t(X^i_t, U^i_t, M_t, W^i_t)$, where $M_t = \sum_{i=1}^{n} \delta_{X^i_t}$.

Information structure $I^i_t = \{X^i_t, M_{1:t}\}$, and assume identical decision rules.

- Arabneydi, Mahajan "Team optimal control of coupled subsystems with mean field sharing," CDC 2014.

Example 3: Mean-field sharing information structure

Dynamics $X^i_{t+1} = f_t(X^i_t, U^i_t, M_t, W^i_t)$, where $M_t = \sum_{i=1}^{n} \delta_{X^i_t}$.

Information structure $I^i_t = \{X^i_t, M_{1:t}\}$, and assume identical decision rules.

Common info.: $C_t = M_{1:t}$, Local info.: $L^i_t = X^i_t$, Prescriptions: $\Gamma_t \colon X^i_t \mapsto U^i_t$.

Information state Due to the symmetry of the system, M_t is an information-state.

Results

- No loss of optimality in using decision strategies: $g^i_t(X^i_t, M_t)$.
- Dynamic program: $V(m) = \min_{\gamma} \mathbb{E}[R_t + \beta V(M_{t+1}) \mid M_t = m, \Gamma_t = \gamma]$
- Size of state space = poly(n); Size of action space $\mathcal{U}^{\mathcal{X}}$.

- Arabneydi, Mahajan "Team optimal control of coupled subsystems with mean field sharing," CDC 2014.

What if the shared information is empty?
The designer's approach

An example: Players with finite memory

Dynamics $X_{t+1} = f_t(X_t, U_t, W_t), \quad Y_t = h_t(X_t, N_t).$

Information structure $I_t = \{Y_t, M_t\}$ Simplest non-classical information structure

$[U_t, M_{t+1}] = g_t(Y_t, M_t)$

- Witsenhausen, "A standard form for sequential stochastic control," Math. Sys. Theory, 1973.
- Mahajan, "Sequential decomposition of sequential teams," PhD Thesis, 2008.

An example: Players with finite memory

Dynamics $X_{t+1} = f_t(X_t, U_t, W_t), \quad Y_t = h_t(X_t, N_t).$

Information structure $I_t = \{Y_t, M_t\}$ Simplest non-classical information structure

$[U_t, M_{t+1}] = g_t(Y_t, M_t)$

Common info.: $C_t = \emptyset$, Local info.: $L_t = (Y_t, M_t)$, Prescriptions: $g_t \colon (Y_t, M_t) \mapsto U_t$.

Information state $\Pi_t = \mathbb{P}(X_t, M_t \mid g_{1:t-1})$

Results
- Dynamic program: $V(\pi) = \min_{g} \mathbb{E}[R_t + \beta V(\Pi_{t+1}) \mid \Pi_t = \pi, g_t = g]$
- Cannot show that time-homogeneous strategies are optimal!

- Witsenhausen, "A standard form for sequential stochastic control," Math. Sys. Theory, 1973.
- Mahajan, "Sequential decomposition of sequential teams," PhD Thesis, 2008.

Final Thoughts

Relation to game-theory

- Generalizes to Markov perfect equilibrium in stochastic games with asymmetric information (Nayyar, Gupta, Langbort, Başar, 2014).
- Implications for dynamic mechanism design.

Is common information (or PHS) a realistic assumption?

- Arises naturally in certain applications.
- Use (a faster time-scale) consensus dynamics to generate common information
- Provide upper and lower bounds

Are there good numerical algorithms?

- Are there POMDP algorithms for large action spaces?
- Is there some structure in the DP that can be exploited?

Interesting variations

- ε common-information
- Approximation techniques
- Reinforcement learning
- Other information structures (sparse structures)?

References

Nayyar, Mahajan and Teneketzis, "Decentralized stochastic control with partial history sharing: A common information approach," IEEE TAC 2013.

Mahajan, Nayyar, and Teneketzis, "Identifying tractable decentralized problems on the basis of information structures", Allerton 2008.

Nayyar, "Sequential Decision-Making in Decentralized systems," PhD Thesis, Univ of Michigan, 2011.

Nayyar, Mahajan and Teneketzis, "Optimal control strategies in delayed sharing information structures," IEEE TAC 2011.

Mahajan, "Optimal decentralized control of coupled subsystems with control sharing," IEEE TAC 2013.

Arabneydi and Mahajan, "Team optimal control of coupled subsystems with mean field sharing," CDC 2014.

Mahajan and Mannan, "Decentralized Stochastic Control," Annals of OR, (in print).