Sequential dynamic teams: State of the art and future directions

Aditya Mahajan
GERAD and McGill University

Un chercheur du GERAD vous parle!
27 March 2018

- email: aditya.mahajan@mcgill.ca
- homepage: http://cim.mcgill.ca/~adityam

Brief Introduction

Education

- B.Tech. Electrical Engineering, IIT Kanpur, 2003.
- MS & PhD EE: Systems, University of Michigan, 2006 & 2008.
- Post-doc, Yale University, 2008–2010.

Current postition

- Associate Professor, Electrical and Computer Engineering, McGill University.
- Member of GERAD since September 2012.

Research interests

- Multi-agent decision making
  Dynamic programming: MDPs and POMDPs
  Structure of optimal strategies
- Networked control systems
  Overlap of control theory and communication theory
- Resource allocation and scheduling
  Multi-armed bandits, network utility maximization
- . . . reinforcement learning

Acknowledgement

Collaborators

- Ashutosh Nayyar, University of Southern California
- Demos Teneketzis, University of Michigan
- Sekhar Tatikonda, Yale University
- Serdar Yüksel, Queen's University
- Nuno Martins, University of Maryland
- Ashish Khisti, University of Toronto
- Aditya Parajape, Imperial College London

Former and current PhD students

- Jalal Arabneydi, currently post-doc at Concordia
- Jhelum Chakravorty, currently post-doc at McGill

- Mohammad Afshari
- Jayakumar Subramanian
- Nima Akbarzadeh

Funding agencies

- NSERC, FRQNT, MITACS, and CFI.

What is team theory?

A brief overview of decision making

Decision making by a single agent

Static optimization

$$\min_{u \in \mathcal{U}} c(u)$$

- Linear programming
- Convex optimization
- Non-convex optimization
- . . .

Decision making by a single agent

Static optimization

$$\min_{u \in \mathcal{U}} c(u)$$

- Linear programming
- Convex optimization
- Non-convex optimization
- ...

Bayesian optimization

$$\min_{g} \mathbb{E}[c(\omega, g(Y(\omega)))]$$

- Stochastic programming
- Stochastic approximation
- Markov Chain Monte Carlo
- ...

Decision making by a single agent

Static optimization

$$\min_{u \in \mathcal{U}} c(u)$$

- Linear programming
- Convex optimization
- Non-convex optimization
- ...

Bayesian optimization

$$\min_{g} \mathbb{E}[c(\omega, g(Y(\omega)))]$$

- Stochastic programming
- Stochastic approximation
- Markov Chain Monte Carlo
- ...

Dynamic optimization/Stochastic control

$$\min_{(g_1,\dots,g_T)} \mathbb{E}\left[\sum_{t=1}^{T} c_t(x_t, u_t)\right]$$

where

$$x_{t+1} = f_t(x_t, u_t, W_t),$$

$$y_t = h_t(x_t, N_t),$$

$$u_t = g_t(y_{1:t}, u_{1:t-1})$$

- Dynamic programming
- Pontryagin maximum principle
- Multi-stage stochastic programming
- ...

Decision making by multiple agents

Game theory — Each agent has an individual objective. Agents compete to minimize individual costs.

- One stage games: Static games, Bayesian games, . . .
- Multi-stage games: Games with perfect information, imperfect information, asymmetric information, . . .

Decision making by multiple agents

Game theory

Each agent has an individual objective. Agents compete to minimize individual costs.

- One stage games: Static games, Bayesian games, . . .
- Multi-stage games: Games with perfect information, imperfect information, asymmetric information, . . .

Team theory/Decentralized stochastic control

All agents have a common objective. Agents cooperate to minimize team costs.

- Static (Bayesian) teams
- Dynamic teams or decentralized stochastic control

Research in team theory started in Economics in mid 50's in the context of organizational behaviour. It has been studied in Systems and Control since the late 60's and in Artificial Intelligence since late 90's.

Comparison with Game Theory

Teams may be thought of as games with aligned preferences

- In teams, all players have a common utlity function

Teams are simpler than non-cooperative games

- Due to aligned preferences, all "pre-game" agreements are enforceable.

Teams are simpler than cooperative games

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

Teams and games have different solution concepts

- Global optima vs Nash equilibrium (and it's refinements).
- In teams, we are typically interested in globally optimal strategies for multi-stage problems with non-classical information structure (equivalent to dynamic games with asymmetric information).

Comparison with Centralized Stochastic Control

Not same as distributed implementation of a centralized solution

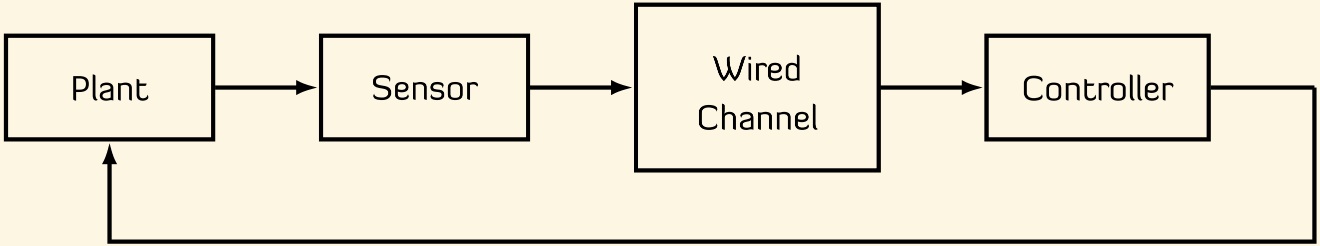
- In most applications, the information structure is given apriori and cannot be changed.
- One seeks a decentralized solution not because it is easy but because it is necessary.

Identifying team optimal strategies is significantly more complicated

- Finding optimal policies in PODMPs is NP-complete.
- Finding optimal policies in dynamics teams is NEXP

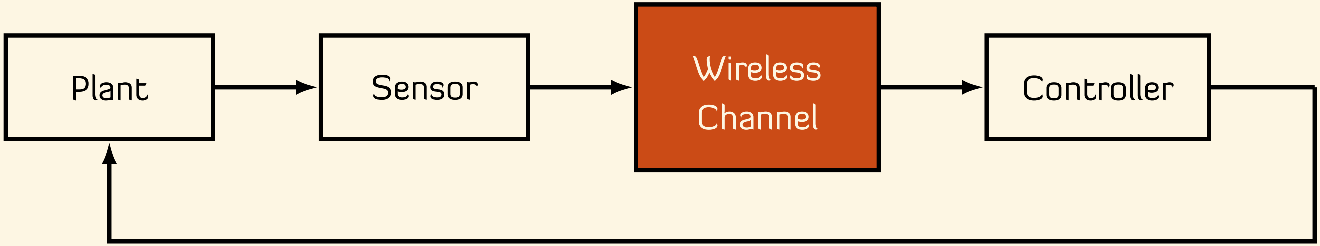

Common theme: multi-stage multi-agent
decision making under uncertainty

A traditional stochastic control system



Examples ⊳ Almost all modern applications . . .

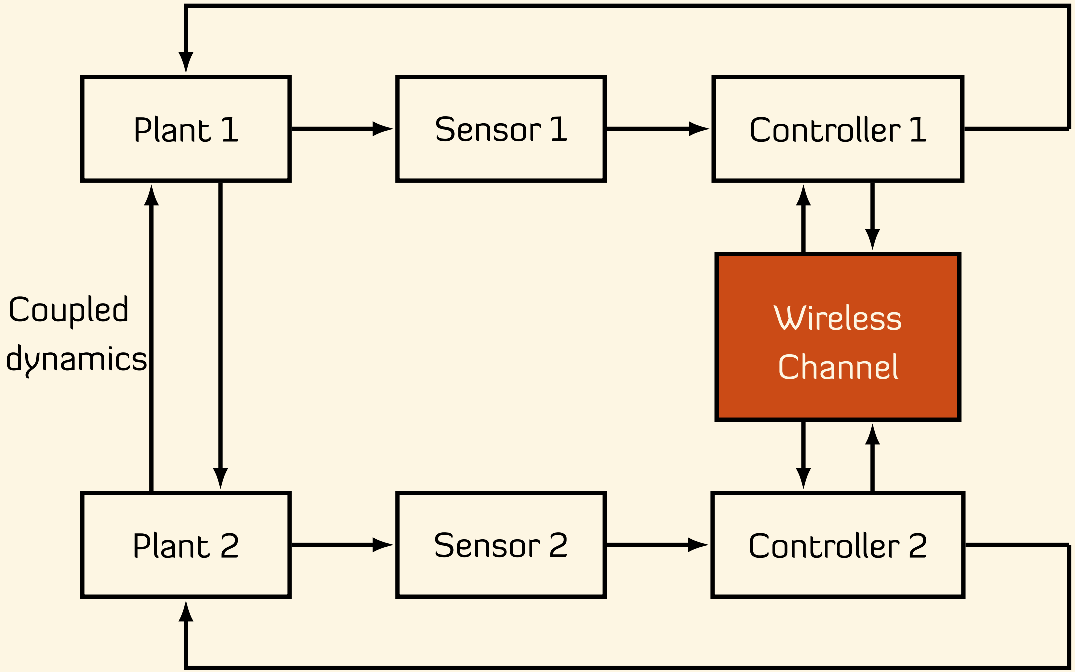
A networked stochastic control system



Examples
- Most cyber-physical systems

Design questions
- What should the controller do if it does not receive a packet?
- When (and what) should the sensor communicate if communication is expensive?

A multi-agent stochastic control system



Examples

- Truck platooning

Design question

- Controller 1 can signal info. to controller 2 through the plant

Why are team problems hard? An example.

LQG model: One of the strongest results in centralized stochastic control

Linear Model ▷ Dynamics: $x_{t+1} = Ax_t + Bu_t + w_t$ ▷ Observations: $y_t = Cx_t + v_t$

Objective Choose $u_t = g_t(y_{1:t}, u_{1:t-1})$ to minimize $\mathbb{E}\left[\sum_{t=1}^{T}\left[x_t^\top Q x_t + u_t R^\top u_t\right]\right]$

Assumption The noise processes $\{w_t\}_{t\geqslant 1}$ and $\{v_t\}_{t\geqslant 1}$ are i.i.d. Gaussian processes.

LQG model: One of the strongest results in centralized stochastic control

Linear Model ▷ Dynamics: $x_{t+1} = Ax_t + Bu_t + w_t$ ▷ Observations: $y_t = Cx_t + v_t$

Objective Choose $u_t = g_t(y_{1:t}, u_{1:t-1})$ to minimize $\mathbb{E}\left[\sum_{t=1}^{T}\left[x_t^\top Q x_t + u_t R^\top u_t\right]\right]$

Assumption The noise processes $\{w_t\}_{t\geqslant 1}$ and $\{v_t\}_{t\geqslant 1}$ are i.i.d. Gaussian processes.

Main result Define $\hat{x}_t = \mathbb{E}[x_t \mid y_{1:t}, u_{1:t-1}]$. Then, the optimal controller may be written as

$$u_t = -K_t x_t$$

where

- ▷ The gains $\{K_t\}_{t\geqslant 1}$ are same as when $y_t = x_t$ and $w_t = 0$ (certainty equivalence).
- ▷ The update of $\hat{x}_t$ is same as when $u_t = 0$ (Kalman filtering).

- ▷ Simon, “Dynamic programming under uncertainty with a quadratic criterion function”, Econometrica, 1956.
- ▷ Theil, “A note on certainty equivalence in dynamic planning”, Econometrica, 1957.
- ▷ Wonham, “On the separation theorem of stochastic control”, SICON 1968.

The Witsenhausen Counterexample

Dynamics $x_1 \sim \mathcal{N}(0, \sigma^2), \quad x_2 = x_1 + u_1, \quad x_3 = x_2 - u_2.$

Observations $y_1 = x_1, \quad y_2 = x_2 + v_2, \quad v_2 \sim \mathcal{N}(0, 1).$

Objective Choose $u_1 = g_1(y_1)$ and $u_2 = g_2(y_2)$ to minimize $\mathbb{E}[k^2 u_1^2 + x_3^2]$.

Remark Linear dynamics, quadratic cost, and Gaussian noise. But $I_1 \not\subseteq I_2$.

▷ Witsenhausen, "A counterexample in stochastic optimum control", SICON, 1968.

The Witsenhausen Counterexample

Dynamics $x_1 \sim \mathcal{N}(0, \sigma^2), \quad x_2 = x_1 + u_1, \quad x_3 = x_2 - u_2.$

Observations $y_1 = x_1, \quad y_2 = x_2 + v_2, \quad v_2 \sim \mathcal{N}(0, 1).$

Objective Choose $u_1 = g_1(y_1)$ and $u_2 = g_2(y_2)$ to minimize $\mathbb{E}[k^2 u_1^2 + x_3^2]$.

Remark Linear dynamics, quadratic cost, and Gaussian noise. But $I_1 \not\subseteq I_2$.

Best linear controller $g_1(y_1) = (\lambda - 1) y_1, \quad g_2(y_2) = \dfrac{\sigma^2\lambda^2}{1 + \sigma^2\lambda^2} y_2, \quad J(\lambda) = k^2\sigma^2(1-\lambda)^2 + \dfrac{\sigma^2\lambda^2}{1+\sigma^2\lambda^2}$

If $k\sigma = 1$ and $k^2 < 1$, then $J_a = 1 - k^2$.

▷ Witsenhausen, "A counterexample in stochastic optimum control", SICON, 1968.

The Witsenhausen Counterexample

Dynamics $x_1 \sim \mathcal{N}(0, \sigma^2), \quad x_2 = x_1 + u_1, \quad x_3 = x_2 - u_2.$

Observations $y_1 = x_1, \quad y_2 = x_2 + v_2, \quad v_2 \sim \mathcal{N}(0, 1).$

Objective Choose $u_1 = g_1(y_1)$ and $u_2 = g_2(y_2)$ to minimize $\mathbb{E}[k^2 u_1^2 + x_3^2]$.

Remark Linear dynamics, quadratic cost, and Gaussian noise. But $I_1 \not\subseteq I_2$.

Best linear controller $g_1(y_1) = (\lambda - 1)y_1, \quad g_2(y_2) = \frac{\sigma^2\lambda^2}{1 + \sigma^2\lambda^2} y_2, \quad J(\lambda) = k^2\sigma^2(1-\lambda)^2 + \frac{\sigma^2\lambda^2}{1+\sigma^2\lambda^2}$

If $k\sigma = 1$ and $k^2 < 1$, then $J_a = 1 - k^2$.

A non-linear strategy $g_1(y_1) = \sigma \operatorname{sgn}(y_1) - y_1, \quad g_2(y_2) = \sigma \operatorname{sgn}(y_2), \quad J = 2k^2\sigma^2(1 - \sqrt{2/\pi}) + 4\sigma^2 \operatorname{erfc}(\sigma).$

If $k\sigma = 1$ and $k \to 0$, then $J_n = 2(1 - \sqrt{2/\pi}) \approx 0.404.$

▷ Witsenhausen, "A counterexample in stochastic optimum control", SICON, 1968.
▷ Mitter and Sahai. "Information and control: Witsenhausen revisited", Learning, control and hybrid systems. Springer, 1999.

The Witsenhausen Counterexample

Dynamics $x_1 \sim \mathcal{N}(0, \sigma^2), \quad x_2 = x_1 + u_1, \quad x_3 = x_2 - u_2.$

If $I_1 \not\subseteq I_2$, non-linear strategies may outperform linear strategies even in LQG systems.

Best linear controller $g_1(y_1) = (\lambda - 1)y_1, \quad g_2(y_2) = \dfrac{\sigma^2\lambda^2}{1+\sigma^2\lambda^2} y_2, \quad J(\lambda) = k^2\sigma^2(1-\lambda)^2 + \dfrac{\sigma^2\lambda^2}{1+\sigma^2\lambda^2}$

If $k\sigma = 1$ and $k^2 < 1$, then $J_a = 1 - k^2$.

A non-linear strategy $g_1(y_1) = \sigma\,\mathrm{sgn}(y_1) - y_1, \quad g_2(y_2) = \sigma\,\mathrm{sgn}(y_2), \quad J = 2k^2\sigma^2(1 - \sqrt{2/\pi}) + 4\sigma^2\mathrm{erfc}(\sigma).$

If $k\sigma = 1$ and $k \to 0$, then $J_n = 2(1 - \sqrt{2/\pi}) \approx 0.404$.

▷ Witsenhausen, "A counterexample in stochastic optimum control", SICON, 1968.
▷ Mitter and Sahai. "Information and control: Witsenhausen revisited", Learning, control and hybrid systems. Springer, 1999.

Research directions since Witsenhausen's counterexample

Numerical methods to find the best non-linear strategy for the counterexample

- Baglietto, Parisini, and Zoppoli. TAC 2001, using neural networks,
- Lee, Lau, and Ho. TAC 2001, using hierarchical search
- . . .
- Ho, CDC 2008.

Identify conditions under which linear strategies are optimal in LQG systems

- Radnar, Annals of Math. Stats, 1962 showed that linear strategies are optimal for static teams.
- Ho and Chu, TAC 1972 showed that linear strategies are optimal for partially nested teams.
- A few specific examples: one-step delay sharing, two-player problem, . . .

Identify conditions under which dynamic programming works for multi-agent systems

- A few specific examples in the literature: Yoshikawa, TAC 1975; Aicardi, Davoli, Minciardi, TAC 1987; Walrand and Varaiya, TIT 1982; . . .
- Nayyar, Mahajan, Teneketzis, TAC 2013: general method called the common information approach.

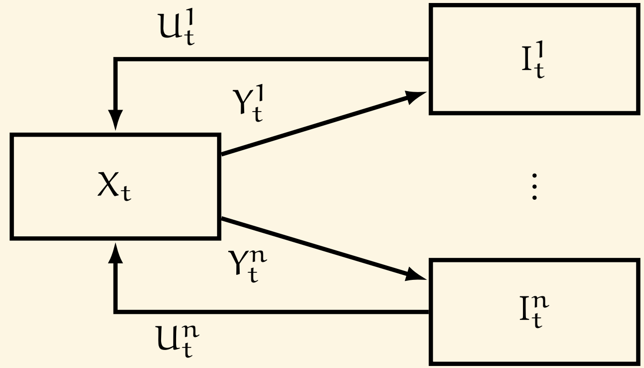
Why are team problems hard?

Conceptual difficulties in obtaining a dynamic program

Simplest general model of a decentralized control system

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

Conceptual difficulties

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

The domain I_t^i of control law g_t^i increases with time.

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

Dynamic programming for centralized stochastic control, revisited

Centralized stochastic control: Information state

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

Centralized stochastic control: Information state

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

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 $Z_t = f_t(I_t)$.

- Controlled Markov property

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

- Sufficient for performance evaluation

  $$\mathbb{E}[C_t \mid I_t = i_t, U_t = u_t] = \mathbb{E}[C_t \mid Z_t = F_t(i_t), U_t = u_t].$$

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 $g_{(t)} = (g_{t+1}, g_{t+2}, \dots)$

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

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 $g_{(t)} = (g_{t+1}, g_{t+2}, \dots)$

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

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 $g_{(t)} = (g_{t+1}, g_{t+2}, \dots)$

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

Therefore,

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

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

Centralized control: Dynamic programming

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

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

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

Relies on $I_t \subseteq I_{t+1}$

Centralized control: Dynamic programming

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

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

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

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

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

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 $g_t : Z_t \mapsto U_t$,

- Can we identify a **sufficient statistic** Z_t^i and restrict attention to $g_t^i : Z_t^i \mapsto U_t^i$?

- 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 $\{Z^i_t\}_{t=0}^{\infty}$ for agent i.

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

Write coupled dynamic programs to identify the best response strategy $\boxed{\mathbf{g}^i = \mathcal{D}^i(\mathbf{g}^{-i})}$

Remarks
- Is the best-response strategy time-homogeneous?
- Does there exist a fixed-point of the coupled dynamic program?
- 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. Arbitrarily fix the strategies g^{-i} of all other agents.

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.

- Does there exist a fixed-point of the coupled dynamic program?
- 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

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

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

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

An example: coupled subsystems with control sharing

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

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

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," TAC 2013.

An example: coupled subsystems with control sharing

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," TAC 2013.

Two approaches to dynamic programming:

The common-information approach

One dynamic program to rule them all

$$V({\color{orange}\blacksquare}) = \min_{\blacksquare} \mathbb{E}[C_t + \beta V({\color{orange}\blacksquare}_{t+1}) \mid {\color{orange}\blacksquare}_t = {\color{orange}\blacksquare}, \blacksquare_t = \blacksquare]$$

One dynamic program to rule them all

$$V(\blacksquare) = \min_{\blacksquare} \mathbb{E}[C_t + \beta V(\blacksquare_{t+1}) \mid \blacksquare_t = \blacksquare, \blacksquare_t = \blacksquare]$$

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

One dynamic program to rule them all

$$V(\blacksquare) = \min_{\blacksquare} \mathbb{E}[C_t + \beta V(\blacksquare_{t+1}) \mid \blacksquare_t = \blacksquare, \blacksquare_t = \blacksquare]$$

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

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

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

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

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

$$V(z) = \min_{\text{[redacted]}} \mathbb{E}[C_t + \beta V(Z_{t+1}) \mid Z_t = z, \text{[redacted]}_t = \text{[redacted]}]$$

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

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

- 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

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

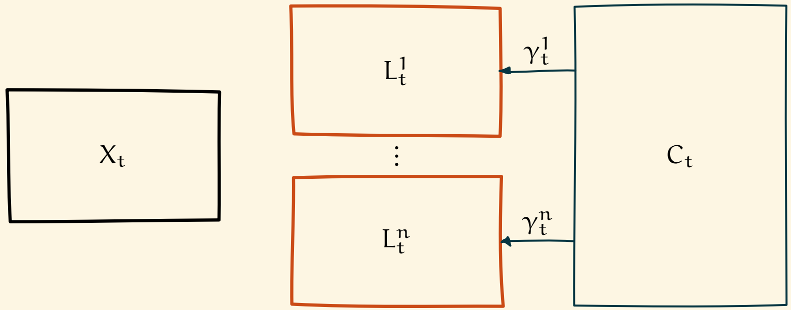
- The **information state** must be a function of the information available to **every** controller.

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

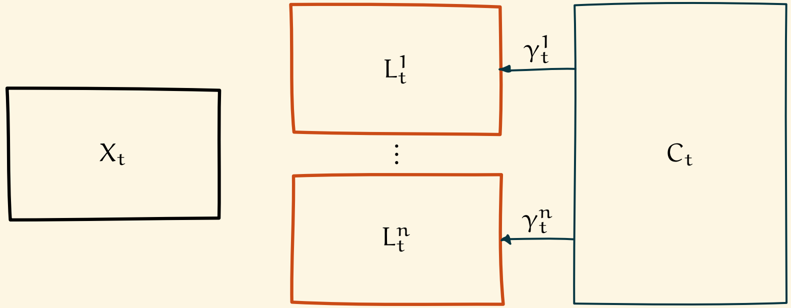
- Each step of the dynamic programming must determine a mapping from $(C_t, L^i_t) \mapsto U^i_t$.
  - The information state Z_t only depends on C_t
  - Thus, the "action" at each step must be a mapping $L^i_t \mapsto U^i_t$. Call it **prescription** and denote it by γ^i_t.

A virtual coordinator

A virtual coordinator



A virtual coordinator



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 simpler)
- "Standard" POMDP results apply, value function is piecewise linear and concave.
- Subsumes many previous results on DP for decentralized stochastic control.

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

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

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

Info 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 control strategies $g^i_t \colon (L^i_t, \Pi_t) \mapsto U^i_t$.
- Dynamic program: $V(\pi) = \min_{\gamma} \mathbb{E}[R_t + \beta V(\Pi_{t+1}) \mid \Pi_t = \pi, \boldsymbol{\Gamma}_t = \boldsymbol{\gamma}]$.

- Witsenhausen, "Separation of estimation and control," Proc IEEE, 1971.
- Nayyar, Mahajan and Teneketzis, "Optimal control strategies in delayed sharing information structures," 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,” 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 $\mathbf{\Xi}_t = (\Xi^1_t, \dots, \Xi^n_t)$ is an information state.

Results

- No loss of optimality in using control strategies $g^i_t \colon (X^i_t, \mathbf{\Xi}_t) \mapsto U^i_t$.
- Dynamic program: $V(\xi) = \min_{\gamma} \mathbb{E}[R_t + \beta V(\mathbf{\Xi}_{t+1}) \mid \mathbf{\Xi}_t = \xi, \mathbf{\Gamma}_t = \gamma]$.

Mahajan, "Optimal decentralized control of coupled subsystems with control sharing," 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 $M_t = \sum_{i=1}^{n} \delta_{X^i_t}$.

Info structure $I^i_t = \{X^i_t, M_{1:t}\}$, and assume identical control laws.

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 $M_t = \sum_{i=1}^{n} \delta_{X^i_t}$.

Info structure $I^i_t = \{X^i_t, M_{1:t}\}$, and assume identical control laws.

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 control 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: Finite memory controller

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.

An example: Finite memory controller

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

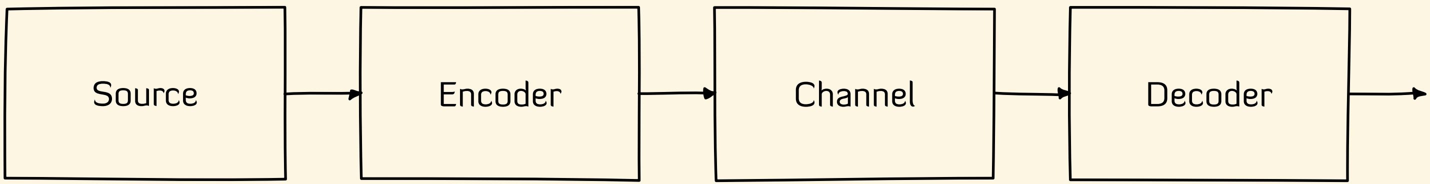
Results

- 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.

Some examples

Real-time communication with or without feedback



Variations

- Source coding, channel coding, or joint source-channel coding setup;
- Feedback from channel output to encoder;
- No feedback or noisy feedback (but either encoder or decoder has finite memory);

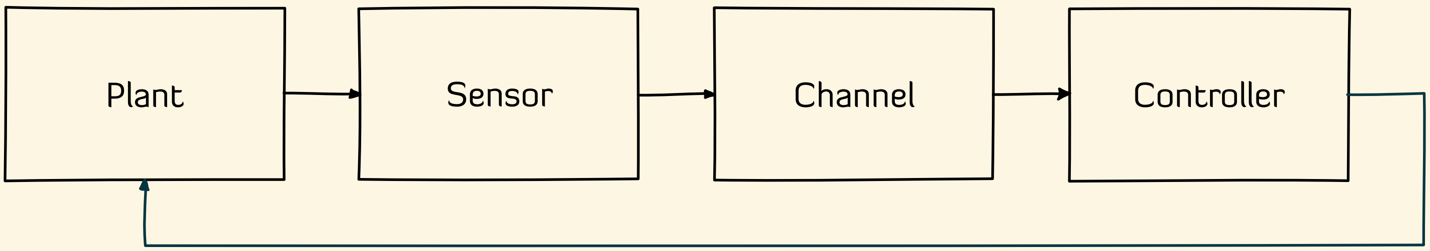
Generalization

- Multi-terminal real-time communication

- Witsenhausen, "On the structure of real-time source coders", BSTJ 1979.
- Walrand and Varaiya, "Optimal causal coding—decoding problems", TIT 1983.
- Borkar, Mitter, and Tatikonda, "Optimal sequential vector quantization of Markov sources", SICON 2001.
- Mahajan Teneketzis, "Optimal design of sequential real-time communication systems", TIT 2009

Source coding, channel coding, joint source-channel coding

Networked control systems



Variations

- Feedback from channel output to sensor;
- No feedback from channel output to sensor (but either the sensor or the controller has finite memory);

- Walrand and Varaiya, "Causal coding and control of Markov chains", System Control Lett., 1983.
- Mahajan and Teneketzis, "Optimal performance of networked control systems with non-classical information structures", SICON 2009.
- Yüksel and Başar, "Stochastic Networked Control Systems: Stabilization and Optimization under Information Constraints", Springer, 2013.

Other examples

- Paging and registration in cellular networks
  Hajek, Mitzel, Yang, TIT 2008
- Multi-access broadcast
  Hlyuchi Gallager, NTC 1983; Ooi, Wornell, CDC 1996; Mahajan, Allerton 2011
- Decentralized balancing of queues
  Ouyang, Teneketzis, Annals OR, 2015
- Remote Estimation
  Lipsa, Martins TAC 2011; Nayyar, Başar, Teneketzis, Veeravalli, TAC 2013; Chakravorty, Mahajan, TAC 2017.
- Decentralized sequential hypothesis testing
  Nayyar, Teneketzis, TIT, 2011.

Unresolved questions and research directions

Identifying optimal linear control laws for partially nested teams

- Common information approach doesn't work directly.
- Let $u^i_t = -K^{i,\text{com}}_t C_t - K^{i,\text{loc}}_t L^i_t$. The prescription $\gamma^i_t(L^i_t) = K^{i,\text{loc}}_t L^i_t$ does not depend on common information.

Mahajan and Nayyar, "Sufficient statistics for linear control strategies in decentralized systems with partial history sharing", TAC 2015.

Unresolved questions and research directions

Identifying optimal linear control laws for partially nested teams

- Common information approach doesn't work directly.
- Let $u_t^i = -K_t^{i,\text{com}} C_t - K_t^{i,\text{loc}} L_t^i$. The prescription $\gamma_t^i(L_t^i) = K_t^{i,\text{loc}} L_t^i$ does not depend on common information.

Developing good numerical algorithms

- Algorithms for POMDP with funcation-valued actions.
- Exploit some feature of the DP.

- Mahajan and Nayyar, "Sufficient statistics for linear control strategies in decentralized systems with partial history sharing", TAC 2015.

Unresolved questions and research directions

Identifying optimal linear control laws for partially nested teams

- Common information approach doesn't work directly.
- Let $u_t^i = -K_t^{i,\text{com}} C_t - K_t^{i,\text{loc}} L_t^i$. The prescription $\gamma_t^i(L_t^i) = K_t^{i,\text{loc}} L_t^i$ does not depend on common information.

Developing good numerical algorithms

- Algorithms for POMDP with funcation-valued actions.
- Exploit some feature of the DP.

Monotonicity of optimal policies

- In MDPs and POMDPs, sufficient conditions based on stochastic dominance, submodularity, and MLR dominance guarantee monotonicity of optimal policies. What is the equivalent for dynamic teams?

- Mahajan and Nayyar, "Sufficient statistics for linear control strategies in decentralized systems with partial history sharing", TAC 2015.

Unresolved questions and research directions

Identifying optimal linear control laws for partially nested teams

- Common information approach doesn't work directly.
- Let $u^i_t = -K^{i,\text{com}}_t C_t - K^{i,\text{loc}}_t L^i_t$. The prescription $\gamma^i_t(L^i_t) = K^{i,\text{loc}}_t L^i_t$ does not depend on common information.

Developing good numerical algorithms

- Algorithms for POMDP with funcation-valued actions. ▷ Exploit some feature of the DP.

Monotonicity of optimal policies

- In MDPs and POMDPs, sufficient conditions based on stochastic dominance, submodularity, and MLR dominance guarantee monotonicity of optimal policies. What is the equivalent for dynamic teams?

Reinforcement learning in dynamic teams

- In principle, we can obtain a reinforcement learning algorithm based on the dynamic program.
- Several interesting features: impact of reward coupling and reward structure.

▷ Mahajan and Nayyar, "Sufficient statistics for linear control strategies in decentralized systems with partial history sharing", TAC 2015.