# Decentralized stochastic control

*The person-by-person and the common information approaches*

Aditya Mahajan

McGill University

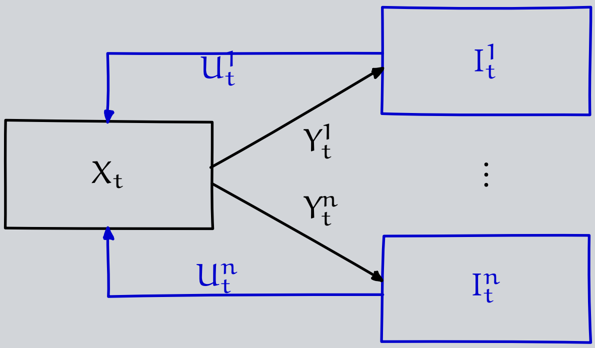
Banff Workshop on Optimal Cooperation, Communication, and Learning in Decentralized Systems, 14 Oct 2014







# Simplest general model of a decentralized control system




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

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

**Information structure**

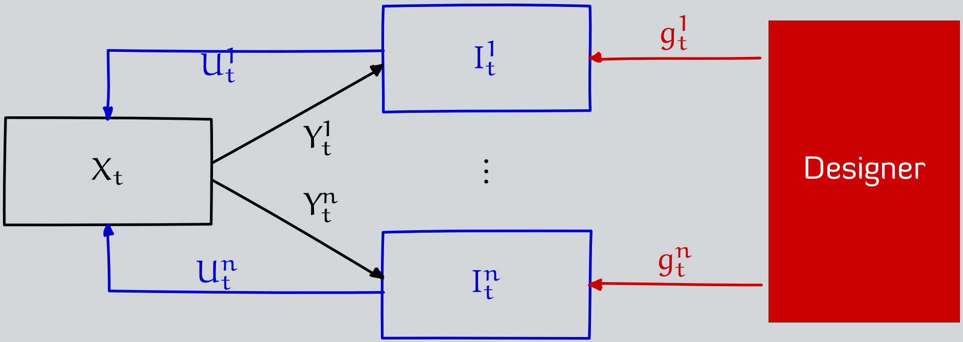
$$\{Y^i_{1:t}, U^i_{1:t-1}\} \subseteq I^i_t \subseteq \{\mathbf{Y}_{1:t}, \mathbf{U}_{1:t-1}\}, \quad U^i_t = g^i_t(I^i_t).$$

**Control Strategy** $\mathbf{g} = (\mathbf{g}^1, \dots, \mathbf{g}^n)$, where $\mathbf{g}^i = (g^i_1, g^i_2, \dots)$.

**Performance**

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

# 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**

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

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

## 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.
  - ...
- AI Literature
  - ...

## 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.
  - . . .
- AI Literature
  - . . .

### 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.

## Literature overview

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

ON 1971.

C 1972.

**Main difficulty**: Seeking global optimality

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.

# 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

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

# 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 $Z_t = f_t(I_t)$.

- 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 $g_t \colon Z_t \mapsto U_t$,

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

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

# Centralized control: Dynamic programming

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

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

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

There exists a time-homogeneous optimal strategy $g^* = (g^*, g^*, \dots)$ 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 $g_t \colon Z_t \mapsto U_t$,

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 $\{Z^i_t\}_{t=0}^{\infty}$ 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

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

**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 $i$.

Arbitr

Identi

$g^{-i}$, then

optim

Write

g

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

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

# An example: coupled subsystems with control sharing

**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," IEEE 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 $\mathbf{g}$:

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

**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

# One dynamic program to rule them all

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

# 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 controller.

# 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 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

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

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

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

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

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

- 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 $\gamma^i_t$.

# One dynamic program to rule them all

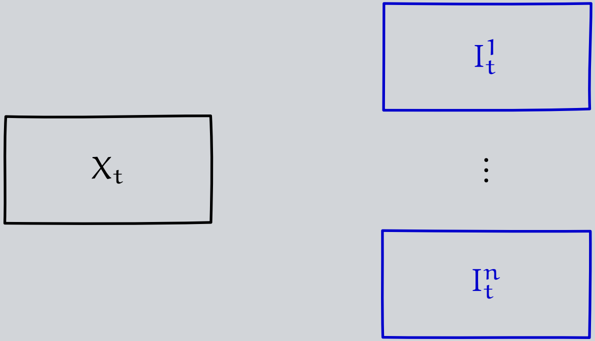
$$V(z) = \min_{\gamma} \mathbb{E}[R_t + \beta V(Z_{t+1}) \mid Z_t = z, \Gamma_t = \gamma]$$

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

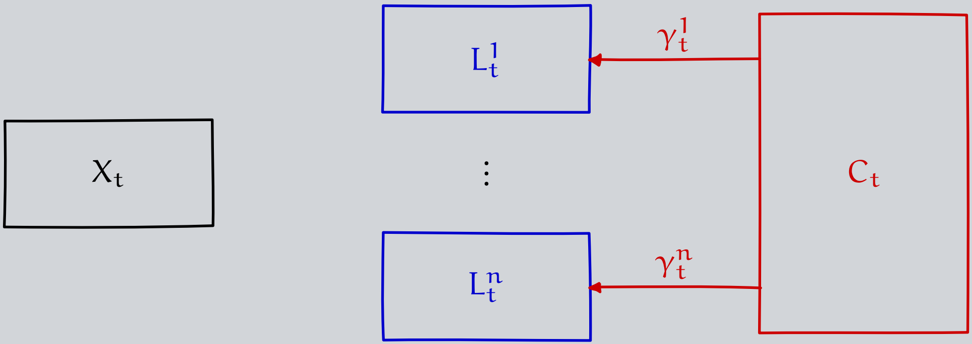
Common information: $C_t = \bigcap_{\tau \geqslant t} \bigcap_{i=1}^{n} I^i_\tau$, Local information: $L^i_t = I^i_t \setminus C_t$

- 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 $\gamma^i_t$.

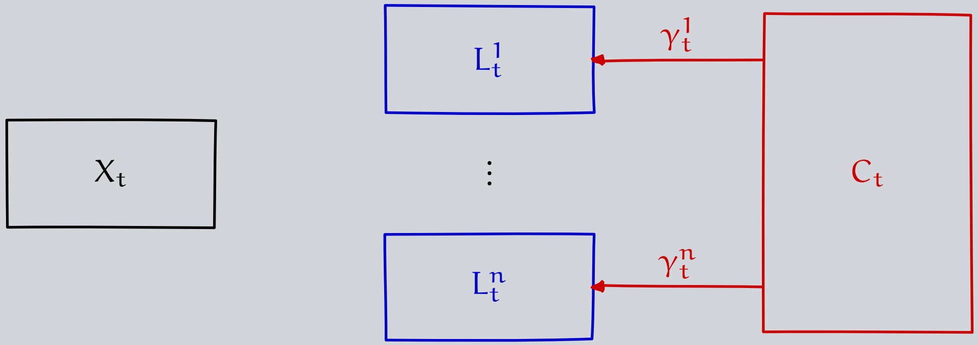
# A virtual coordinator



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

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

Information structure $I_t^i = \{Y_{1:t}^i, U_{1:t-1}^i, \mathbf{Y}_{1:t-k}, \mathbf{U}_{1:t-k}\}$. $k$ is the sharing delay.

Common info.: $C_t = \{\mathbf{Y}_{1:t-k}, \mathbf{U}_{1:t-k}\}$, Local Info.: $L_t^i = I_t^i \setminus C_t$, Pres.: $\Gamma_t^i \colon L_t^i \mapsto U_t^i$

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

Results
- No loss of optimality in using control strategies $g_t^i \colon (L_t^i, \Pi_t) \mapsto U_t^i$.
- Dynamic program: $V(\pi) = \min_{\gamma} \mathbb{E}[R_t + \beta V(\Pi_{t+1}) \mid \Pi_t = \pi, \mathbf{\Gamma}_t = \gamma]$.

---

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

**Information 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

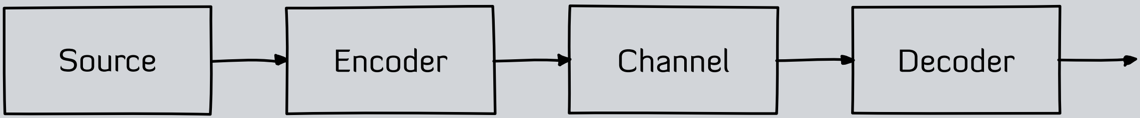
- Dynamic program: $V(\pi) = \min_g \mathbb{E}[R_t + \beta V(\Pi_{t+1}) \mid \Pi_t = \pi, g_t = g]$
- Cannot show that time-homogeneous strategies are optimal!

---

- Witsenhausen, "A standard form for sequential stochastic control," Math. Sys. Theory, 1973.

# Some applications

# Real-time communication with feedback



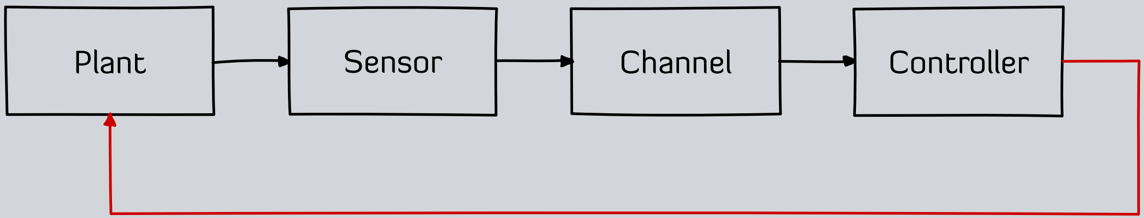

## Variations

- 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
  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);
- Connections to posterior matching

# Other examples

## Paging and registration in cellular networks

Hajek, Mitzel, Yang, IEEE TIT 2008

## Multi-access broadcast

Hlyuchi Gallager, NTC 1983; Ooi, Wornell, CDC 1996; Mahajan, Allerton 2011

## Decentralized balancing of queues

Ouyang, Teneketzis, arxiv 2014.

## Remote Estimation

Lipsa, Martins IEEE TAC 2011; Nayyar, Başar, Teneketzis, Veeravalli, IEEE TAC 2013.

## Decentralized sequential hypothesis testing

Nayyar, Teneketzis, IEEE TIT, 2011. Related to social learning.

# Further Reading

## Existence results for arbitrary spaces

- Gupta, Yüksel, Başar, Langbort, "On the Existence of Optimal Policies for a Class of Static and Sequential Dynamic Teams," arxiv preprint 2014.

## Application to Linear Quadratic Gaussian (LQG) system

- Mahajan, Nayyar, "Sufficient statistics for linear control strategies in decentralized systems with partial history sharing," IEEE TAC 2015 (in print)
- Nayyar, Lassard, "Optimal Control for LQG Systems on Graphs—Part I: Structural Results," arxiv preprint, 2014.

## Generalization to Games

- Nayyar, Gupta, Langbort, Başar, "Common Information Based Markov Perfect Equilibria for Stochastic Games With Asymmetric Information: Finite Games," IEEE TAC 2014.
- Nayyar, Gupta, Langbort, Başar, "Common Information based Markov Perfect Equilibria for Linear-Gaussian Games with Asymmetric Information," arxiv preprint 2014.

# Final Thoughts

Simple solution to a complex class of problems

## Is common information (or PHS) a realistic assumption?

- Arises naturally in certain applications.
- Use (a faster time-scale) consensus dynamics to generate common information (e.g., in mean-field sharing)
- 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

- $\varepsilon$ common-information
- Approximation techniques
- Reinforcement learning
- Other information structures (sparse structures)?

# References


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

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

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

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

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).