Mean-field games among teams

Aditya Mahajan
McGill University

CDC Workshop on Large Population Teams
15 Dec 2024

- email: aditya.mahajan@mcgill.ca
- web: https://adityam.github.io


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

Types of multi-agent decision problems

Teams

- All agents have **common objective**
- Agents **cooperate** to minimize team cost
- Agents are **not strategic**
- Solution concepts: person-by-person optimality, global optimality . . .

Types of multi-agent decision problems

Teams

- All agents have **common objective**
- Agents **cooperate** to minimize team cost
- Agents are **not strategic**
- Solution concepts: person-by-person optimality, global optimality . . .

Games

- Each agent has **individual objective**
- Agents **compete** to minimize individual cost
- Agents are **strategic**
- Solution concepts: Nash equil, Bayesian Nash, Subgame perfect, Markov perfect, Bayesian perfect, . . .

Emerging applications are neither teams nor games

Emerging applications

- Aggregators competing in demand response market
- Ride-sharing companies in a city
- DARPA spectrum sharing challenge
- StarCraft

Emerging applications are neither teams nor games

Emerging applications

- Aggregators competing in demand response market
- Ride-sharing companies in a city
- DARPA spectrum sharing challenge
- StarCraft

Salient features

- Multiple teams are competing in the same environment (non zero-sum).
- Agents belonging to the same team are willing to cooperate with one another.
- But agents have partial information about other members of their teams and agents in other teams.

Emerging applications are neither teams nor games

Emerging applications

- Aggregators competing in demand response market
- Ride-sharing companies in a city
- DARPA spectrum sharing challenge
- StarCraft

Salient features

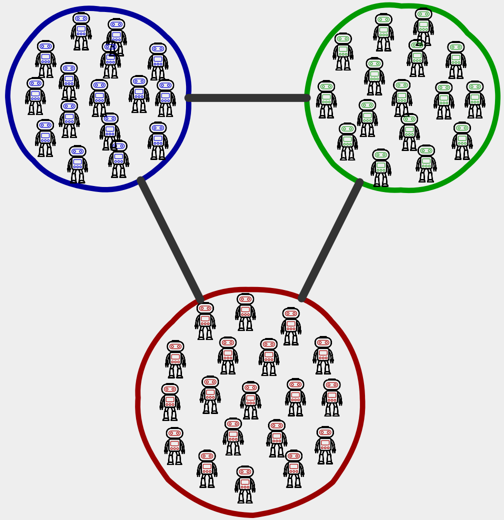
- Multiple teams are competing in the same environment (non zero-sum).
- Agents belonging to the same team are willing to cooperate with one another.
- But agents have partial information about other members of their teams and agents in other teams.

See Tang et al (2024) for general solution.
Are there tractable models for games among teams?

Model of games among teams

System Model

- K teams, indexed by $\mathcal{K} = \{1, \dots, K\}$.
- Team k has $N^{(k)}$ agents, indexed by $\mathcal{N}^{(k)}$.

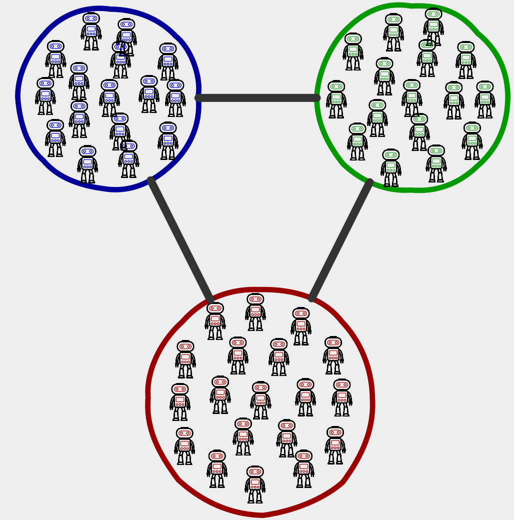


Model of games among teams

System Model

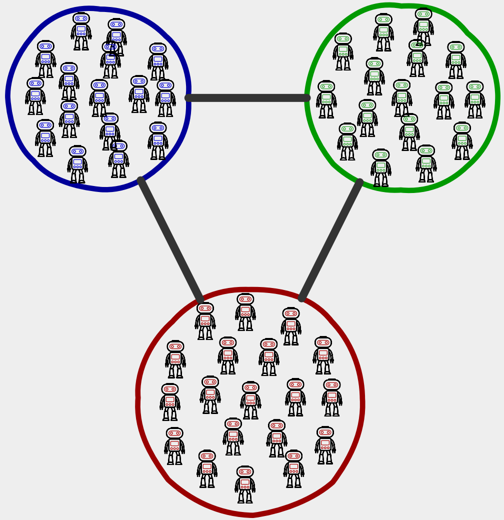
- K teams, indexed by $\mathcal{K} = \{1, \dots, K\}$.
- Team k has $N^{(k)}$ agents, indexed by $\mathcal{N}^{(k)}$.

- For agent i in team k:
  state: $S^i_t \in \mathcal{S}^{(k)}$; action: $A^i_t \in \mathcal{A}^{(k)}$.
- Team k:
  state: $S^{(k)}_t = (S^i_t)_{i \in \mathcal{N}^{(k)}}$; action: $A^{(k)}_t = (A^i_t)_{i \in \mathcal{N}^{(k)}}$.
- For the system:
  state: $S_t = (S^{(1)}_t, \dots, S^{(K)}_t)$; action: $A_t = (A^{(1)}_t, \dots, A^{(K)}_t)$.



Model of games among teams

System Model



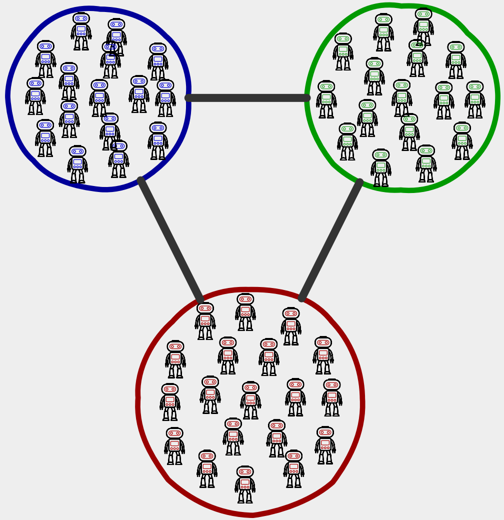
- K teams, indexed by $\mathcal{K} = \{1, \dots, K\}$.
- Team k has $N^{(k)}$ agents, indexed by $\mathcal{N}^{(k)}$.

- For agent i in team k:
  state: $S^i_t \in \mathcal{S}^{(k)}$; action: $A^i_t \in \mathcal{A}^{(k)}$.
- Team k:
  state: $S^{(k)}_t = (S^i_t)_{i \in \mathcal{N}^{(k)}}$; action: $A^{(k)}_t = (A^i_t)_{i \in \mathcal{N}^{(k)}}$.
- For the system:
  state: $S_t = (S^{(1)}_t, \dots, S^{(K)}_t)$; action: $A_t = (A^{(1)}_t, \dots, A^{(K)}_t)$.

- Dynamics: $\mathbb{P}(S_{t+1} \mid S_t, A_t)$.
- Per-step cost for team k: $C^{(k)}_t = c^{(k)}(S_t, A_t)$.

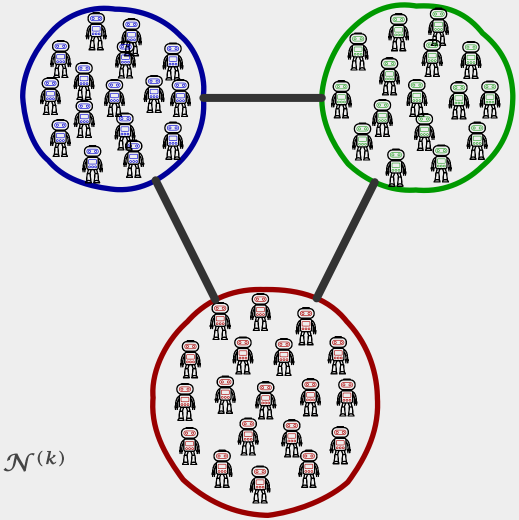
Model of games among teams

Information Structure



- $I^i_t \subseteq \{S_{1:t}, A_{1:t-1}\}$.

Model of games among teams



Information Structure

- $I_t^i \subseteq \{S_{1:t}, A_{1:t-1}\}$.

Strategy profiles

- Strategy of agent i: $A_t^i \sim \pi_t^i(I_t^i)$.
- Strategy of team k: $\pi^{(k)} = (\pi_1^{(k)}, \dots, \pi_T^{(k)})$ where $\pi_t^{(k)} = (\pi_t^i)_{i \in \mathcal{N}^{(k)}}$
- Strategy profile: $\pi = (\pi^{(1)}, \dots, \pi^{(K)})$.

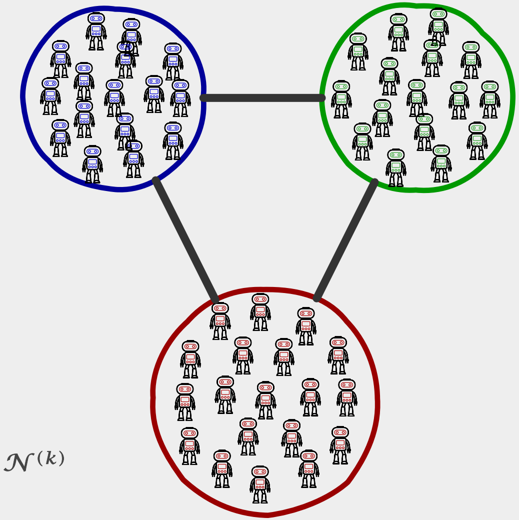
Model of games among teams

Information Structure

- $I_t^i \subseteq \{S_{1:t}, A_{1:t-1}\}$.

Strategy profiles

- Strategy of agent i: $A_t^i \sim \pi_t^i(I_t^i)$.
- Strategy of team k: $\pi^{(k)} = (\pi_1^{(k)}, \dots, \pi_T^{(k)})$ where $\pi_t^{(k)} = (\pi_t^i)_{i \in \mathcal{N}^{(k)}}$
- Strategy profile: $\pi = (\pi^{(1)}, \dots, \pi^{(K)})$.



Performance of team k

- $J^{(k)}(\pi) = \mathbb{E}^{\pi}\Big[\sum_{t=1}^{T} C_t^{(k)}\Big]$.

Solution concept

Team-Nash equilibrium (Tang et al, 2024)

A strategy profile π is a team-Nash equilibrium (TNE) if for every team k and every alternative policy $\tilde{\pi}^{(k)}$ for team k, we have

$$J^{(k)}(\pi^{(k)}, \pi^{(-k)}) \leq J^{(k)}(\tilde{\pi}^{(k)}, \pi^{(-k)})$$

Solution concept

Team-Nash equilibrium (Tang et al, 2024)

A strategy profile π is a team-Nash equilibrium (TNE) if for every team k and every alternative policy $\tilde{\pi}^{(k)}$ for team k, we have

$$J^{(k)}(\pi^{(k)}, \pi^{(-k)}) \le J^{(k)}(\tilde{\pi}^{(k)}, \pi^{(-k)})$$

Salient features

- Different from Nash equilibrium because agents of the same team can deviate together.
- Agents within a team have a non-classical information structure
- Agents in different teams have asymmetric information.

Solution concept

Team-Nash equilibrium (Tang et al, 2024)

A strategy profile π is a team-Nash equilibrium (TNE) if for every team k and every alternative policy $\tilde{\pi}^{(k)}$ for team k, we have

$$J^{(k)}(\pi^{(k)}, \pi^{(-k)}) \leq J^{(k)}(\tilde{\pi}^{(k)}, \pi^{(-k)})$$

Salient features

- Different from Nash equilibrium because agents of the same team can deviate together.
- Agents within a team have a non-classical information structure
- Agents in different teams have asymmetric information.

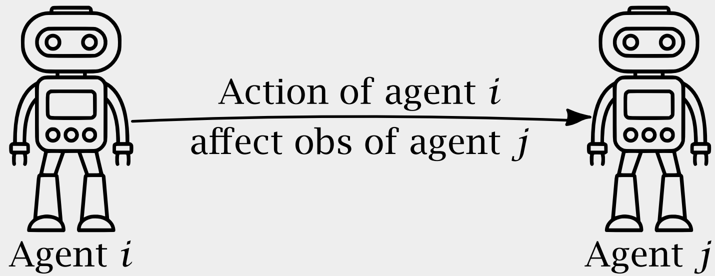
Special cases

- Dynamic team $K = 1$
- Dynamic games $\mathcal{N}^{(k)} = 1$ for all k.

Games among teams inherit the conceptual challenges of teams and games.

Are there tractable models for games among teams?

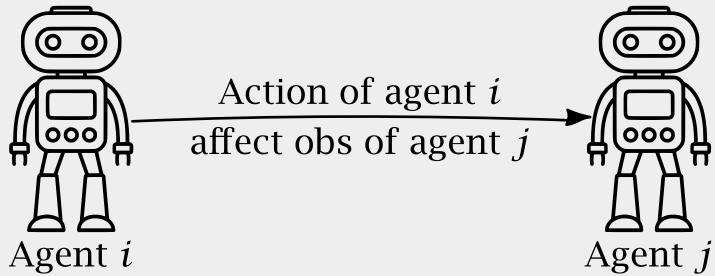
Why are dynamic multi-agent problems hard?



Signaling

- Actions of agent i can convey information to agent j.

Why are dynamic multi-agent problems hard?

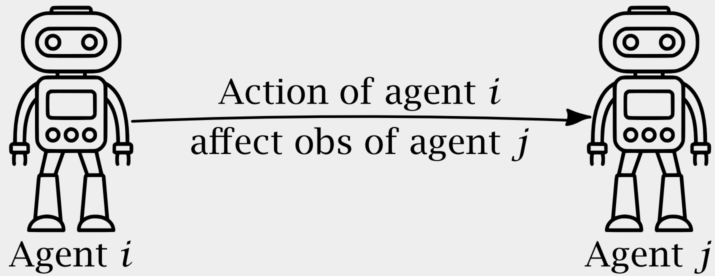


Signaling

- Actions of agent i can convey information to agent j.

- In teams, agents can exploit signaling to inform other agents.
- In games, agents can use signaling to confuse other agents.

Why are dynamic multi-agent problems hard?



Signaling

- Actions of agent i can convey information to agent j.

Usually, problems with lack of signaling are tractable:

- **Teams:** MDPs, POMDPs, partially nested teams, mean-field teams . . .
- **Games:** Markov perfect equilibrium (MPE), some common-info based refinements of MPE, mean-field games, . . .

Models with exchangeable agents

Assumption

Agents in each team are **exchangeable**. In particular, let

$$\sigma(S_t) = \big(\sigma^1(S_t^{(1)}, \dots, \sigma^{(K)}(S_t^{(K)})\big)$$

denote a permulation of agents in each team.

Then,

$$\mathbb{P}(\sigma(S_{t+1}) \mid \sigma(S_t), \sigma(A_t)) = \mathbb{P}(S_{t+1} \mid S_t, A_t)$$

and

$$c^{(k)}(\sigma(S_t), \sigma(A_t)) = c^{(k)}(S_t, A_t)$$

Models with exchangeable agents

Assumption

Agents in each team are **exchangeable**. In particular, let

$$\sigma(S_t) = (\sigma^1(S_t^{(1)}, \dots, \sigma^{(K)}(S_t^{(K)})$$

denote a permulation of agents in each team.

Then,

$$\mathbb{P}(\sigma(S_{t+1}) \mid \sigma(S_t), \sigma(A_t)) = \mathbb{P}(S_{t+1} \mid S_t, A_t)$$

and

$$c^{(k)}(\sigma(S_t), \sigma(A_t)) = c^{(k)}(S_t, A_t)$$

Rationale

- In many applications, dynamics and cost do not depend on how we index agents.
- E.g.: demand response, DARPA spectrum challenge, ...

Models with exchangeable agents

Assumption

Agents in each team are **exchangeable**. In particular, let

$$\sigma(S_t) = (\sigma^1(S_t^{(1)}, \dots, \sigma^{(K)}(S_t^{(K)})$$

denote a permulation of agents in each team. Then,

$$\mathbb{P}(\sigma(S_{t+1}) \mid \sigma(S_t), \sigma(A_t)) = \mathbb{P}(S_{t+1} \mid S_t, A_t)$$

and

$$c^{(k)}(\sigma(S_t), \sigma(A_t)) = c^{(k)}(S_t, A_t)$$

Rationale

- In many applications, dynamics and cost do not depend on how we index agents.
- E.g.: demand response, DARPA spectrum challenge, ...

Implications

- Exchangeable couplings ≡ mean-field coupling
- Mean-field couplings lead to lack of signaling

- Haung, Malhame, Caines, "Large population stochastic dynamic games ...", 2006.
- Arabneydi, Mahajan, "Team optimal control ... with mean field sharing", 2013
- Saldi, Raginsky, Başar, "Markov-Nash equil in mean-field games ...", 2018.
- Sanjari, Saldi, Yüksel, "Optimality of ... with mean-field information sharing," 2024.

Outline of the talk

Mean-field Teams ($\mathcal{K} = 1$)	▷ Common information based DP ▷ Simplification under symmetry assumptions ▷ Sampling-based algorithms for solving DP ▷ Infinite population approximation

Outline of the talk

Mean-field Teams ($\mathcal{K} = 1$)	▷ Common information based DP ▷ Simplification under symmetry assumptions ▷ Sampling-based algorithms for solving DP ▷ Infinite population approximation

Mean-field Games among Teams ($\mathcal{K} > 1$)	▷ Common information based MPE ▷ Infinite population approximation

Outline of the talk

Mean-field Teams ($\mathcal{K} = 1$)

- Common information based DP
- Simplification under symmetry assumptions
- Sampling-based algorithms for solving DP
- Infinite population approximation

Mean-field Games among Teams ($\mathcal{K} > 1$)

- Common information based MPE
- Infinite population approximation

Notational Simplification

System Model

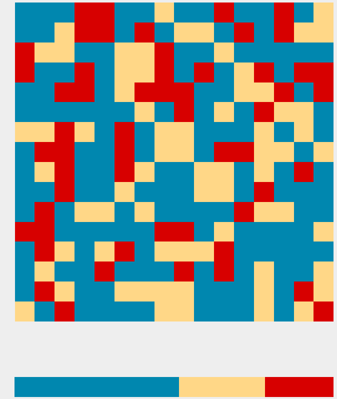
- Only one team with N agents indexed by $\mathcal{N}$.

- For agent i:
  state: $S^i_t \in \mathcal{S}$; action: $A^i_t \in \mathcal{A}$.

Arabneydi, Mahajan, “Team optimal control of coupled subsystems with mean field sharing,” 2014.

Notational Simplification

System Model



- Only one team with N agents indexed by $\mathcal{N}$.

- For agent i:

  state: $S^i_t \in \mathcal{S}$; action: $A^i_t \in \mathcal{A}$.

- Mean-field coupled dynamics:

$$\mathbb{P}(S_{t+1} \mid S_t, A_t) = \prod_{i \in \mathcal{N}} \mathbb{P}(S^i_{t+1} \mid S^i_t, A^i_t, Z_t) \text{ where } Z_t = \xi(S_t) \coloneqq \frac{1}{N} \sum_{i \in \mathcal{N}} \delta_{S^i_t}.$$

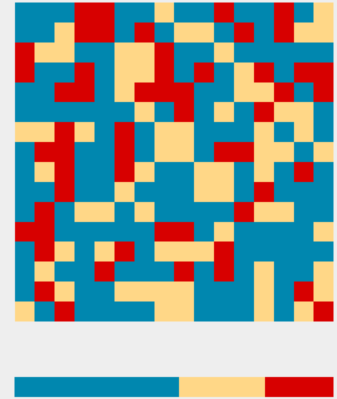
- Mean-field coupled cost:

$$C_t = \frac{1}{N} \sum_{i \in \mathcal{N}} c_t(S^i_t, A^i_t, Z_t).$$

Arabneydi, Mahajan, “Team optimal control of coupled subsystems with mean field sharing,” 2014.

Notational Simplification

System Model



- Only one team with N agents indexed by $\mathcal{N}$.
- For agent i:
  state: $S^i_t \in \mathcal{S}$; action: $A^i_t \in \mathcal{A}$.
- Mean-field coupled dynamics:

$$\mathbb{P}(S_{t+1} \mid S_t, A_t) = \prod_{i \in \mathcal{N}} \mathbb{P}(S^i_{t+1} \mid S^i_t, A^i_t, Z_t) \text{ where } Z_t = \xi(S_t) \coloneqq \frac{1}{N} \sum_{i \in \mathcal{N}} \delta_{S^i_t}.$$

- Mean-field coupled cost:

$$C_t = \frac{1}{N} \sum_{i \in \mathcal{N}} c_t(S^i_t, A^i_t, Z_t).$$

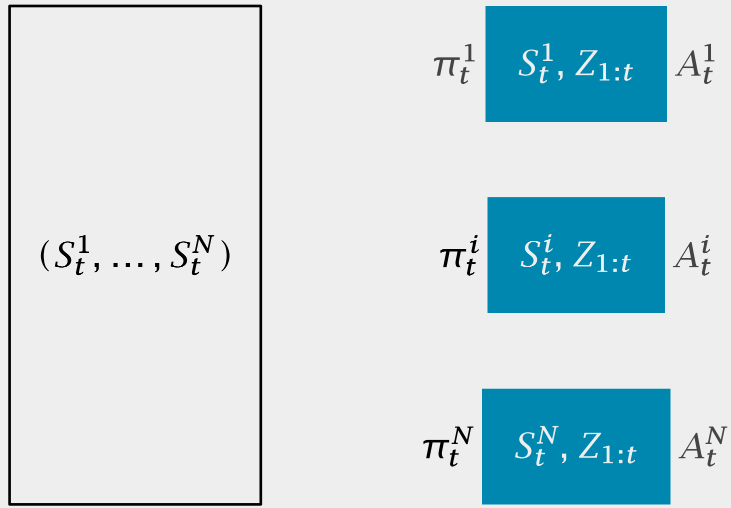
Mean-field sharing info structure	$I^i_t = \{S^i_t, Z_{1:t}\}$

Arabneydi, Mahajan, “Team optimal control of coupled subsystems with mean field sharing,” 2014.

Common-information based simplification

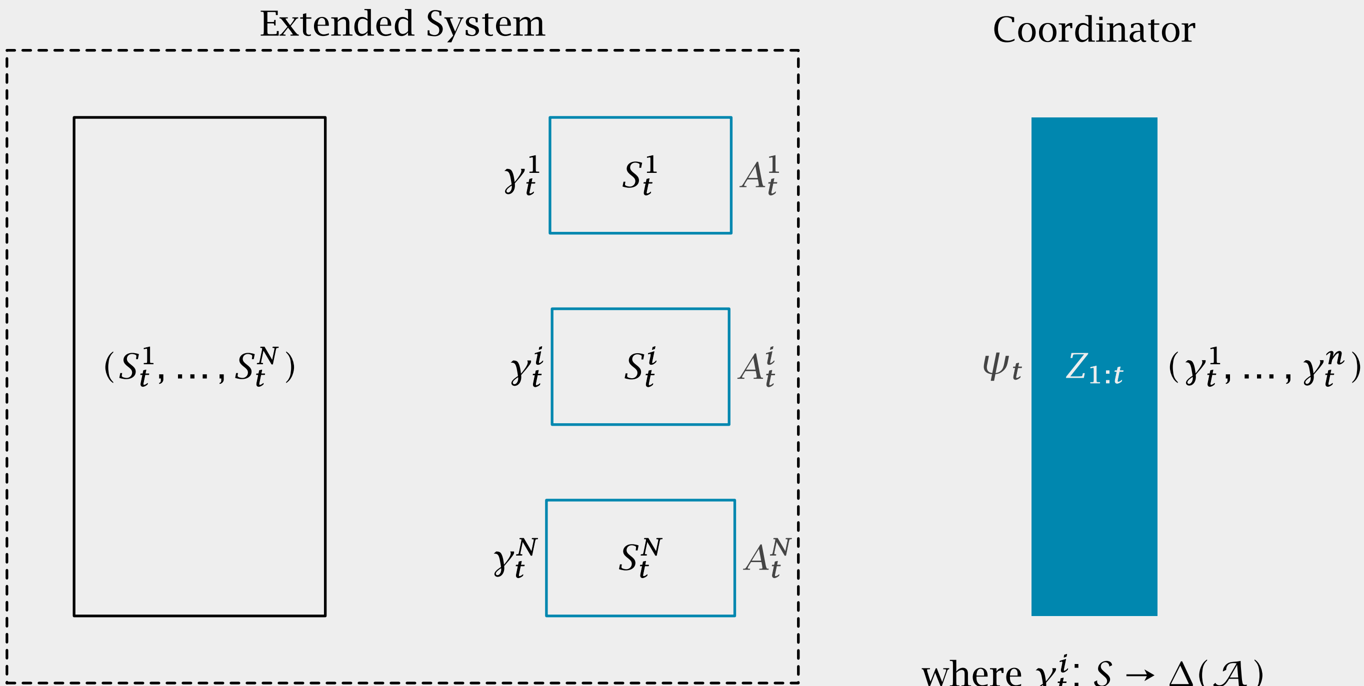
📰 Nayyar, Mahajan, Teneketzis, "Decentralized stochastic control . . .: A common info approach," IEEE TAC 2013.

Common-information based simplification



Nayyar, Mahajan, Teneketzis, "Decentralized stochastic control ...: A common info approach," IEEE TAC 2013.

Common-information based simplification



where $\gamma_t^i \colon \mathcal{S} \to \Delta(\mathcal{A})$

Nayyar, Mahajan, Teneketzis, "Decentralized stochastic control . . . : A common info approach," IEEE TAC 2013.

Equivalent centralized problem

System Model (from the p.o.v. of the coordinator)

- Unobserved state: $(S_t^1, \dots, S_t^n)$
- Observations: Z_t
- Control action: prescriptions $\gamma_t = (\gamma_t^1, \dots, \gamma_t^N)$, where $\gamma_t^i \colon \mathcal{S} \to \Delta(\mathcal{A})$.
- Control law: $(\gamma_t^1, \dots, \gamma_t^N) = \psi_t(Z_{1:t})$.

Equivalent centralized problem

System Model (from the p.o.v. of the coordinator)

- Unobserved state: $(S_t^1, \dots, S_t^n)$
- Observations: Z_t
- Control action: prescriptions $\gamma_t = (\gamma_t^1, \dots, \gamma_t^N)$, where $\gamma_t^i \colon \mathcal{S} \to \Delta(\mathcal{A})$.
- Control law: $(\gamma_t^1, \dots, \gamma_t^N) = \psi_t(Z_{1:t})$.

"Standard" centralized POMDP

Equivalent centralized problem

System Model (from the p.o.v. of the coordinator)

- Unobserved state: $(S_t^1, \dots, S_t^n)$
- Observations: Z_t
- Control action: prescriptions $\gamma_t = (\gamma_t^1, \dots, \gamma_t^N)$, where $\gamma_t^i \colon \mathcal{S} \to \Delta(\mathcal{A})$.
- Control law: $(\gamma_t^1, \dots, \gamma_t^N) = \psi_t(Z_{1:t})$.

"Standard" centralized POMDP

Info state and DP

- Information state: $B_t = \mathbb{P}(S_t^1, \dots, S_t^N \mid Z_{1:t}, \gamma_{1:t-1})$
- Dynamic programming decomposition:

$$V_t(b) = \min_{\gamma_t = (\gamma_t^1, \dots, \gamma_t^N)} \mathbb{E}[C_t + V_{t+1}(B_{t+1}) \mid B_t = b, \Gamma_t = \gamma_t]$$

Limitations and simlifying assumption

Computational challenges

- The size of the unobserved state $(S_t^1, \dots, S_t^N)$ is exponential in the number of agents.
- The size of the belief state B_t is doubly exponential in the number of agents.
- The size of the "actions" $(\gamma_t^1, \dots, \gamma_t^N)$ is exponential in the number of agents.

Limitations and simlifying assumption

Computational challenges

- The size of the unobserved state $(S_t^1, \dots, S_t^N)$ is exponential in the number of agents.
- The size of the belief state B_t is doubly exponential in the number of agents.
- The size of the "actions" $(\gamma_t^1, \dots, \gamma_t^N)$ is exponential in the number of agents.

Simplifying assumption: All agents use identical strategies

- In general, entails loss of optimality.
- Without loss of optimality in some settings (LQG teams, asymptotically large population)
- Ensures simplicity, fairness, and robustness.

Arabneydi, Mahajan, "Team optimal control of coupled subsystems with mean field sharing," 2014.
Arabneydi, Mahajan, "Linear Quadratic Mean Field Teams: ...", 2016.
Sanjari, Saldi, Yüksel, "Optimality of Decentralized Symmetric Policies for Stochastic Teams with Mean-Field Information Sharing," 2024.

Agents using identical strategies implies that the state process is an exchangeable Markov process

Exchangeable random vector

Definition

- A random vector $S \in \mathcal{S}^N$ is called exchangeable if for any permutation σ,

$$(S^1, \dots, S^N) \stackrel{d}{=} (S^{\sigma(1)}, \dots, S^{\sigma(N)})$$

Exchangeable random vector

Definition

- A random vector $S \in \mathcal{S}^N$ is called exchangeable if for any permutation σ,

$$(S^1, \dots, S^N) \stackrel{d}{=} (S^{\sigma(1)}, \dots, S^{\sigma(N)})$$

Note

- Equivalently, S is exchangeable if for any $s, s' \in \mathcal{S}^N$ such that $\xi(s) = \xi(s')$, we have

$$\mathbb{P}(S = s) = \mathbb{P}(S = s')$$

Exchangeable random vector

Definition

- A random vector $S \in \mathcal{S}^N$ is called exchangeable if for any permutation σ,

$$(S^1, \dots, S^N) \stackrel{d}{=} (S^{\sigma(1)}, \dots, S^{\sigma(N)})$$

Note

- Equivalently, S is exchangeable if for any $s, s' \in \mathcal{S}^N$ such that $\xi(s) = \xi(s')$, we have

$$\mathbb{P}(S = s) = \mathbb{P}(S = s')$$

Example

- For $\mathcal{S} = \{0, 1\}$ and $N = 3$, $S \in \mathcal{S}^N$ is exchangeable iff

$$p(001) = p(010) = p(100) \quad \text{and} \quad p(011) = p(110) = p(101).$$

where $p(ijk) = \mathbb{P}(S = (i, j, k))$.

Exchangeable Markov chain

Definition

A Markov chain $\{S_t\}_{t \ge 1}$ defined on $\mathcal{S}^N$ is called exchangeable if

- The initial state S_1 is exchangeable
- The transition matrix is invariant under permutations, i.e., for any permutation σ

$$\mathbb{P}(S_{t+1} = \sigma(s') \mid S_t = \sigma(s)) = \mathbb{P}(S_{t+1} = s' \mid S_t = s).$$

Exchangeable Markov chain

Definition

A Markov chain $\{S_t\}_{t \ge 1}$ defined on $\mathcal{S}^N$ is called exchangeable if

- The initial state S_1 is exchangeable
- The transition matrix is invariant under permutations, i.e., for any permutation σ

$$\mathbb{P}(S_{t+1} = \sigma(s') \mid S_t = \sigma(s)) = \mathbb{P}(S_{t+1} = s' \mid S_t = s).$$

Note

- If $\{S_t\}_{t \ge 1}$ is an exchangeable Markov chain, then for every t, the state S_t is an exchangeable random vector.

Exchangeable Markov chain

Definition

A Markov chain $\{S_t\}_{t \ge 1}$ defined on $\mathcal{S}^N$ is called exchangeable if

- The initial state S_1 is exchangeable
- The transition matrix is invariant under permutations, i.e., for any permutation σ

$$\mathbb{P}(S_{t+1} = \sigma(s') \mid S_t = \sigma(s)) = \mathbb{P}(S_{t+1} = s' \mid S_t = s).$$

Note

- If $\{S_t\}_{t \ge 1}$ is an exchangeable Markov chain, then for every t, the state S_t is an exchangeable random vector.

Example

- Interacting particle systems

Mean-field projection of an exchangeable Markov chain

Definition

Let $\{S_t\}_{t \ge 1}$ be an $\mathcal{S}^N$-valued exchangeable Markov chain. Its mean-field projection is the process $\{Z_t\}_{t \ge 1}$ where $Z_t = \xi(S_t)$.

Mean-field projection of an exchangeable Markov chain

Definition

Let $\{S_t\}_{t \ge 1}$ be an $\mathcal{S}^N$-valued exchangeable Markov chain. Its mean-field projection is the process $\{Z_t\}_{t \ge 1}$ where $Z_t = \xi(S_t)$.

Proposition

- The mean-field projection of an exchangeable Markov chain is a Markov chain, i.e.,

$$\mathbb{P}(Z_{t+1} = z_{t+1} \mid S_{1:t} = s_{1:t}) = \mathbb{P}(Z_{t+1} = z_{t+1} \mid Z_t = \xi(s_t))$$

Mean-field projection of an exchangeable Markov chain

Definition

Let $\{S_t\}_{t\geq 1}$ be an $\mathcal{S}^N$-valued exchangeable Markov chain. Its mean-field projection is the process $\{Z_t\}_{t\geq 1}$ where $Z_t = \xi(S_t)$.

Proposition

- The mean-field projection of an exchangeable Markov chain is a Markov chain, i.e.,

$$\mathbb{P}(Z_{t+1} = z_{t+1} \mid S_{1:t} = s_{1:t}) = \mathbb{P}(Z_{t+1} = z_{t+1} \mid Z_t = \xi(s_t))$$

Theorem

Conditioned on the mean-field, all feasible realizations are equally likely, i.e.,

$$\mathbb{P}(S_t = s_t \mid Z_{1:t} = z_{1:t}) = \mathbb{P}(S_t = s_t \mid Z_t = z_t) = \mathbb{P}(S_t = \sigma(s_t) \mid Z_t = z_t) = \frac{\mathbb{1}\{\xi(s_t) = z_t\}}{\Xi(z_t)}$$

Leveraging exchageability leads to simpler info state

Intuition

When all agents use the same prescription γ, the controlled process $\{S_t\}_{t \ge 1}$ is an exchageable Markov chain.

Leveraging exchageability leads to simpler info state

Intuition

When all agents use the same prescription γ, the controlled process $\{S_t\}_{t \ge 1}$ is an exchageable Markov chain.

Theorem

In the mean-field teams problem, the mean-field $Z_t = \xi(X_t)$ is an info state for the (centralized) coordinated system.

- Z_t is sufficient to compute belief B_t:

$$B_t = \mathbb{P}(S_t = s_t \mid Z_{1:t} = z_{1:t}, \Gamma_{1:t} = \gamma_{1:t}) = \mathbb{P}(S_t = s_t \mid Z_t = z_t) = \frac{\mathbb{1}\{\xi(s_t) = z_t\}}{\Xi(z_t)}$$

- Z_t evolves in a controlled Markov manner:

$$\mathbb{P}(Z_{t+1} = z_{t+1} \mid Z_{1:t}, \Gamma_{1:t} = \gamma_{1:t}) = \mathbb{P}(Z_{t+1} = z_{t+1} \mid Z_t = z_t, \Gamma_t = \gamma_t).$$

Leveraging exchageability leads to simpler info state

Intuition

When all agents use the same prescription γ, the controlled process $\{S_t\}_{t \ge 1}$ is an exchageable Markov chain.

Theorem

In the mean-field teams problem, the mean-field $Z_t = \xi(X_t)$ is an info state for the (centralized) coordinated system.

Dynamic program

$$V_t(z_t) = \min_{\gamma_t : \mathcal{S} \to \Delta(\mathcal{A})} \mathbb{E}\big[c(S_t, A_t) + V_{t+1}(Z_{t+1}) \bigm| Z_t = z_t, \Gamma_t = \gamma_t \big].$$

▷ Dimension of state and action space doesn't increase with number of agents!

How to efficiently solve the DP

Method 1: Sampling-based algorithms for solving DP

Key idea: Work with counts rather than mean-field

- State counts: $M_t(s) = \sum_{i \in \mathcal{N}} \mathbb{1}\{S^i_t = s\}$.
- State-action counts: $\bar{M}_t(s, a) = \sum_{i \in \mathcal{N}} \mathbb{1}\{S^i_t = s, A^i_t = a\}$.
- State-action-next-state counts: $\hat{M}_t(s, a, s') = \sum_{i \in \mathcal{N}} \mathbb{1}\{S^i_t = s, A^i_t = a, S^i_{t+1} = s'\}$.

Nguyen, Kumar, Lau, "Collective multiagent sequential decision making under uncertainty", 2017

Method 1: Sampling-based algorithms for solving DP

Key idea: Work with counts rather than mean-field

Dynamics of the count

- From state counts to state-action counts:

$$\mathbb{P}(\bar{M}_t = \bar{m}_t \mid M_t = m_t, \Gamma_t = \gamma_t) = \frac{m_t(s)!}{\prod_{a \in \mathcal{A}} \bar{m}_t(s,a)!} \prod_{a \in \mathcal{A}} \gamma_t(a \mid s)^{\bar{m}_t(s,a)}.$$

- From state-action counts to state-action-state counts:

$$\mathbb{P}(\hat{M}_t = \hat{m}_t \mid \bar{M}_t = \bar{m}_t) = \frac{\bar{m}_t(s,a)!}{\prod_{s' \in \mathcal{S}} \hat{m}_t(s,a,s')!} \prod_{s' \in \mathcal{S}} P(s'|s,a,z_t)^{\hat{m}_t(s,a,s')}.$$

- From state-action-state counts to updated state counts:

$$m_{t+1}(s') = \sum_{s \in \mathcal{S}} \sum_{a \in \mathcal{A}} \hat{m}_t(s,a,s')$$

Nguyen, Kumar, Lau, “Collective multiagent sequential decision making under uncertainty”, 2017

Method 1: Sampling-based algorithms for solving DP

Key idea: Work with counts rather than mean-field

Dynamics of the count

- From state counts to state-action counts:

$$\mathbb{P}(\bar{M}_t = \bar{m}_t \mid M_t = m_t, \Gamma_t = \gamma_t) = \frac{m_t(s)!}{\prod_{a\in\mathcal{A}} \bar{m}_t(s,a)!} \prod_{a\in\mathcal{A}} \gamma_t(a \mid s)^{\bar{m}_t(s,a)}.$$

- From state-action counts to state-action-state counts:

$$\mathbb{P}(\hat{M}_t = \hat{m}_t \mid \bar{M}_t = \bar{m}_t) = \frac{\bar{m}_t(s,a)!}{\prod \hat{m}_t(s,a,s')!} \prod_{s'\in\mathcal{S}} P(s'|s,a,z_t)^{\hat{m}_t(s,a,s')}.$$

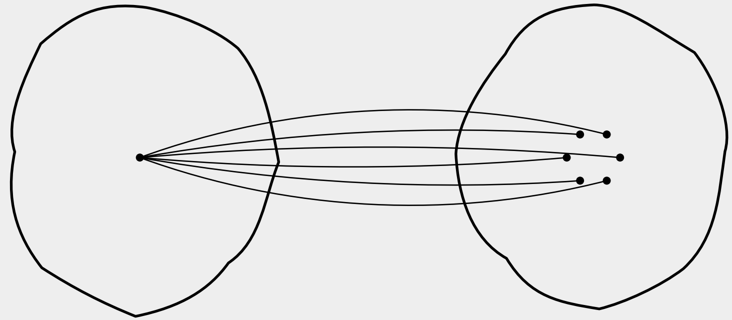
Can use sampling-based MDP algorithms for solve the DP!

Nguyen, Kumar, Lau, “Collective multiagent sequential decision making under uncertainty”, 2017

Method 2: Infinite population approximation

Main Idea

- Consider the infinite population limit
- Emperical MF is replaced by statistical MF



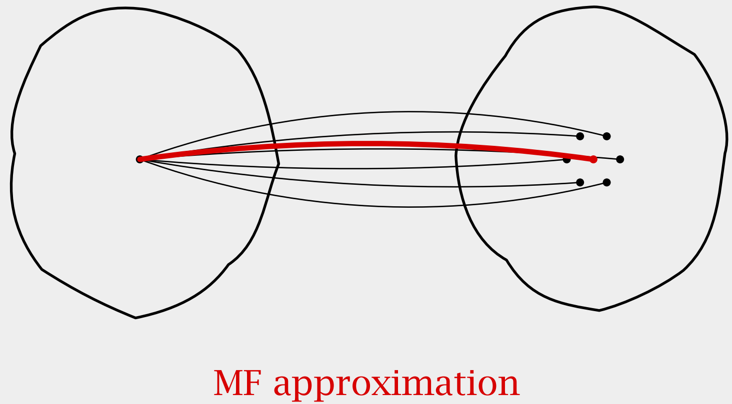
True dynamics

Subramanian, "RL in partially observed and multi-agent environments," 2020.

Method 2: Infinite population approximation

Main Idea

- Consider the infinite population limit
- Emperical MF is replaced by statistical MF
- Statistical MF evolves deterministically
- **Stochastic DP simplifies to deterministic DP**



Infinite population dynamics

$$\bar{Q}(\bar{z}' \mid \bar{z}, \gamma) = \mathbb{1}\{\bar{z}' = \bar{q}(\bar{z}, \gamma)\} \quad \text{where } \bar{q}(\bar{z}, \gamma)(s') = \sum_{s \in S} \bar{z}(s) P(s' \mid s, \gamma(s), \bar{z})$$

Subramanian, “RL in partially observed and multi-agent environments,” 2020.

Method 2: Infinite population approximation

Main Idea

- Consider the
- Emperical M
- Statistical M
- Stochastic D

Infinite pop

$\bar{Q}(\bar{z}' \mid \bar{z}, y$

Inf pop policy is approximately optimal for finite pop team

Let ψ_∞ be the optimal coordination strategy of the inf population team. Then,

$$J_N^{\psi_\infty} \le J_N^\star + 2 \sum_{t=1}^{T} \frac{\kappa \mathcal{L}_t}{\sqrt{N}}$$

where

- κ is a constant that depends on the metric on S.
- $\mathcal{L}_t$ is the Lipschitz constant of total cost $V_{N,t}^\star$.

Subramanian, “RL in partially observed and multi-agent environments,” 2020.

Outline of the talk

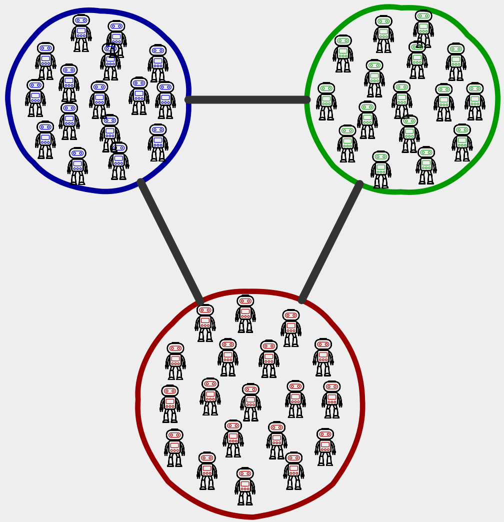
Mean-field Teams ($\mathcal{K} = 1$)

- Common information based DP
- Simplification under symmetry assumptions
- Sampling-based algorithms for solving DP
- Infinite population approximation

Mean-field Games among Teams ($\mathcal{K} > 1$)

- Common information based MPE
- Infinite population approximation

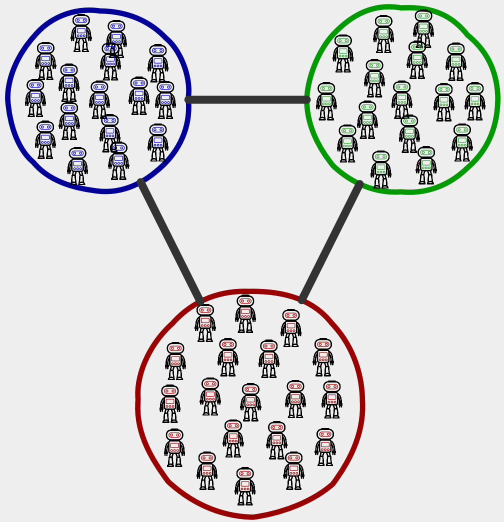
Exactly the same idea works for games among teams



Exactly the same idea works for games among teams

Assumptions

- Mean-field sharing information structure: $I^i_t = \{S^i_t, Z_{1:t}\}$, where $Z_t = (Z^{(1)}_t, \dots, Z^{(N)}_t)$.
- Agents in a team use identical strategies.



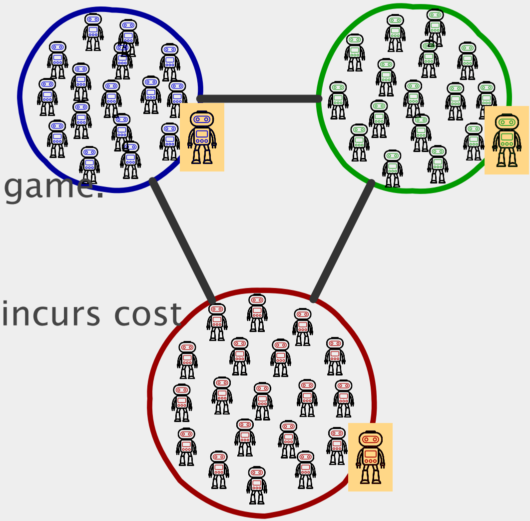
Subramanian, Kumar, Mahajan, "Meam-field Markov perfect equilibrium for mean-field games among teams," 2023.

Exactly the same idea works for games among teams

Assumptions

- Mean-field sharing information structure: $I^i_t = \{S^i_t, Z_{1:t}\}$, where $Z_t = (Z^{(1)}_t, \dots, Z^{(N)}_t)$.
- Agents in a team use identical strategies.

Common info based Markov perfect equilibrium



- Adapt the idea of Nayyar et al (2014) to construct an equivalent game.
- There is a virtual player associated with each team.
- Virtual player k observes Z_t, chooses $\gamma^{(k)}_t : \mathcal{S}^{(k)} \to \Delta(\mathcal{A}^{(k)})$, and incurs cost

$$\ell^{(k)}_t(Z_t, \gamma^{(k)}_t) = \mathbb{E}[C^{(k)}_t \mid Z_t, \gamma^{(k)}_t]$$

Subramanian, Kumar, Mahajan, "Meam-field Markov perfect equilibrium for mean-field games among teams," 2023.
Nayyar, Gupta, Langbort, Başar, "Common info based MPE for stochastic games in asymmetric info ...", 2014.

Exactly the same idea works for games among teams

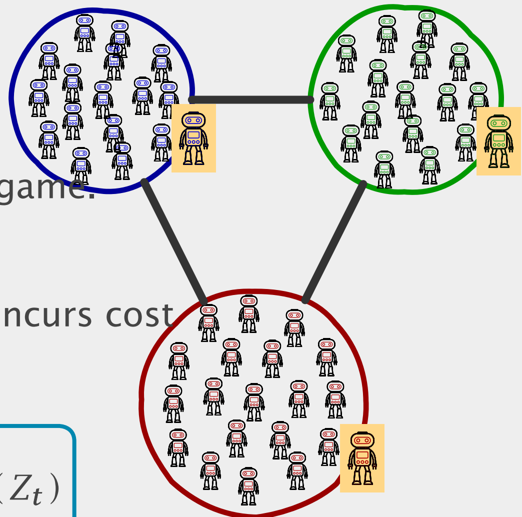
Assumptions

- Mean-field sharing information structure: $I^i_t = \{S^i_t, Z_{1:t}\}$, where $Z_t = (Z^{(1)}_t, \dots, Z^{(N)}_t)$.
- Agents in a team use identical strategies.

Common info based Markov perfect equilibrium

- Adapt the idea of Nayyar et al (2014) to construct an equivalent game.
- There is a virtual player associated with each team.
- Virtual player k observes Z_t, chooses $\gamma^{(k)}_t \colon \mathcal{S}^{(k)} \to \Delta(\mathcal{A}^{(k)})$, and incurs cost

$$\ell^{(k)}_t(Z_t, \gamma^{(k)}_t) = \mathbb{E}[C^{(k)}_t \mid Z_t, \gamma^{(k)}_t]$$



Strategy of virtual player k:	$\psi^{(k)} = (\psi^{(k)}_t, \dots, \psi^{(k)}_T)$, where $\gamma^{(k)}_t \sim \psi^{(k)}_t(Z_t)$

Subramanian, Kumar, Mahajan, "Meam-field Markov perfect equilibrium for mean-field games among teams," 2023.

Nayyar, Gupta, Langbort, Başar, "Common info based MPE for stochastic games in asymmetric info ...", 2014.

Exactly the same idea works for games among teams

Assumptions

- ▷ Mean-field s ...).
- ▷ Agents in a ...

Common in...

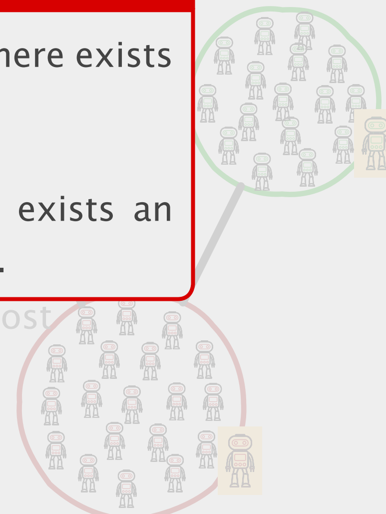
- ▷ Adapt the id...
- ▷ There is a vi...
- ▷ Virtual player k observes Z_t, chooses $\gamma_t^{(k)} \colon \mathcal{S}^{(k)} \to \Delta(\mathcal{A}^{(k)})$, and incurs cost

$$\ell_t^{(k)}(Z_t, \gamma_t^{(k)}) = \mathbb{E}[C_t^{(k)} \mid Z_t, \gamma_t^{(k)}]$$

Strategy of virtual player k:	$\psi^{(k)} = (\psi_1^{(k)}, \dots, \psi_T^{(k)})$, where $\gamma_t^{(k)} \sim \psi_t^{(k)}(Z_t)$

Equivalence of the two models

- ▷ For every Team-Nash equilibrium of the original game there exists an equivalent Nash equilibrium of the virtual game.
- ▷ For every Nash equilibrium of the virtual game there exists an equivalent Team-Nash equilibrium of the original game.



Subramanian, Kumar, Mahajan, "Meam-field Markov perfect equilibrium for mean-field games among teams," 2023.
Nayyar, Gupta, Langbort, Başar, "Common info based MPE for stochastic games in asymmetric info ...", 2014.

How to solve the virtual game

Since the virtual game is a game with perfect information, we can identify a Markov perfect equilibrium (MPE) via DP.

Maskin and Tirole, "A theory of dynamic oligopoly, i: Overview and quantity competition with large fixed costs", Econometrica, 1988.

How to solve the virtual game

Since the virtual game is a game with perfect information, we can identify a Markov perfect equilibrium (MPE) via DP.

Dynamic programming decomposition

A Markov strategy $\psi = (\psi^{(k)})_{k \in \mathcal{K}}$ is a MPE of the virtual game

if and only if

for each $k \in \mathcal{K}$, there exist a sequence of value functions $\{V_t^{(k)}\}_{t=1}^{T+1}$ such that $V_{T+1}^{(k)} \equiv 0$ and

$$V_t^{(k)}(z_t) = \min_{\gamma_t^{(k)}} \Big\{ \ell_t^{(k)}(z_t, \gamma_t^{(k)}) + \mathbb{E}\big[V_{t+1}^{(k)}(Z_{t+1}) \mid z_t, \gamma_t^{(k)}, \pi_t^{(-k)} \big] \Big\}$$

where all pure strategies in $\text{supp}(\psi_t^{(k)}(z_t))$ are minimizers of the right hand side.

Maskin and Tirole, "A theory of dynamic oligopoly, i: Overview and quantity competition with large fixed costs", Econometrica, 1988.

How to solve the virtual game

Mean-field Markov Perfect Equilibrium

- Recall that MPE is a refinement of NE (of the virtual game)
- There is an equivalent TNE of the original game among teams.
- We call it MF-MPE (Mean-Field Markov Perfect Equilibrium)

Dynamic pr

A Markov strat

if and only

for each $k \in \mathcal{K}$, there exist a sequence of value functions $\{V_t^{(k)}\}_{t=1}^{T+1}$ such that $V_{T+1}^{(k)} \equiv 0$ and

$$V_t^{(k)}(z_t) = \min_{\gamma_t^{(k)}} \Big\{ \ell_t^{(k)}(z_t, \gamma_t^{(k)}) + \mathbb{E}[V_{t+1}^{(k)}(Z_{t+1}) \mid z_t, \gamma_t^{(k)}, \pi_t^{(-k)}] \Big\}$$

where all pure strategies in $\text{supp}(\psi_t^{(k)}(z_t))$ are minimizers of the right hand side.

Maskin and Tirole, "A theory of dynamic oligopoly, i: Overview and quantity competition with large fixed costs", Econometrica, 1988.

Solving the DP

Conceptual challenges

- The state space $\mathcal{Z}^* = \prod_{k \in \mathcal{K}} \mathcal{Z}^{(k)}$ is the space of all MFs.
- $|\mathcal{Z}^*| \le \prod_{k \in \mathcal{K}} (N^{(k)} + 1)^{|\mathcal{S}^{(k)}|}$ curse of dimensionality

Solving the DP

Conceptual challenges

- The state space $\mathcal{Z}^* = \prod_{k \in \mathcal{K}} \mathcal{Z}^{(k)}$ is the space of all MFs.
- $|\mathcal{Z}^*| \le \prod_{k \in \mathcal{K}} (N^{(k)} + 1)^{|\mathcal{S}^{(k)}|}$ curse of dimensionality
- Action space is $\mathbb{R}^{|\mathcal{A}^{(k)}| \times |\mathcal{Z}^*|}$ curse of decentralization

Solving the DP

Conceptual challenges

- The state space $\mathcal{Z}^* = \prod_{k \in \mathcal{K}} \mathcal{Z}^{(k)}$ is the space of all MFs.
- $|\mathcal{Z}^*| \le \prod_{k \in \mathcal{K}} (N^{(k)} + 1)^{|\mathcal{S}^{(k)}|}$ **curse of dimensionality**
- Action space is $\mathbb{R}^{|\mathcal{A}^{(k)}| \times |\mathcal{Z}^*|}$ **curse of decentralization**

Efficiently solving the DP

- Using sampling-based algorithms with counts (instead of MF)
- Approx. finite population model with infinite population limit (similar approx bounds hold)

Generalizations and Discussions

Multiple types of agents

- Types can simply be encoded as part of state space
- Even allows for the possibility of agents changing types

Generalizations and Discussions

Multiple types of agents

- Types can simply be encoded as part of state space
- Even allows for the possibility of agents changing types

Some small and some large teams

- Simply take the MF limit of large teams.
- If number of agents in a team are not known, simply use an estimate!

Generalizations and Discussions

Multiple types of agents

- Types can simply be encoded as part of state space
- Even allows for the possibility of agents changing types

Closed loop vs open loop control

- Even for inf pop limit, it is assumed that MF is observed $\implies$ closed loop implementation
- Open-loop implementation (MF not observed) leads to an additional error which can be bounded using mixing properties of Markov chains

Some small and some large teams

- Simply take the MF limit of large teams.
- If number of agents in a team are not known, simply use an estimate!

Generalizations and Discussions

Multiple types of agents

- Types can simply be encoded as part of state space
- Even allows for the possibility of agents changing types

Some small and some large teams

- Simply take the MF limit of large teams.
- If number of agents in a team are not known, simply use an estimate!

Closed loop vs open loop control

- Even for inf pop limit, it is assumed that MF is observed $\implies$ closed loop implementation
- Open-loop implementation (MF not observed) leads to an additional error which can be bounded using mixing properties of Markov chains

Stability of approximate strategies

- Results generalize to cts compact state spaces.
- More nuanced analysis needed for non-compact state spaces with unbounded cost.

Key Messages

Exchangeability of agents (or MF coupling)
and observation of MF

$\Longrightarrow$ lack of signaling in teams,
games, and games among teams.

Key Messages

Exchangeability of agents (or MF coupling) and observation of MF

$\Longrightarrow$ lack of signaling in teams, games, and games among teams.

Exact DP population is possible (but computationally challenging) for finite population

Is well approximated by infinite population limit (but still computationally challenging)

- email: aditya.mahajan@mcgill.ca
- web: https://adityam.github.io

Thank you