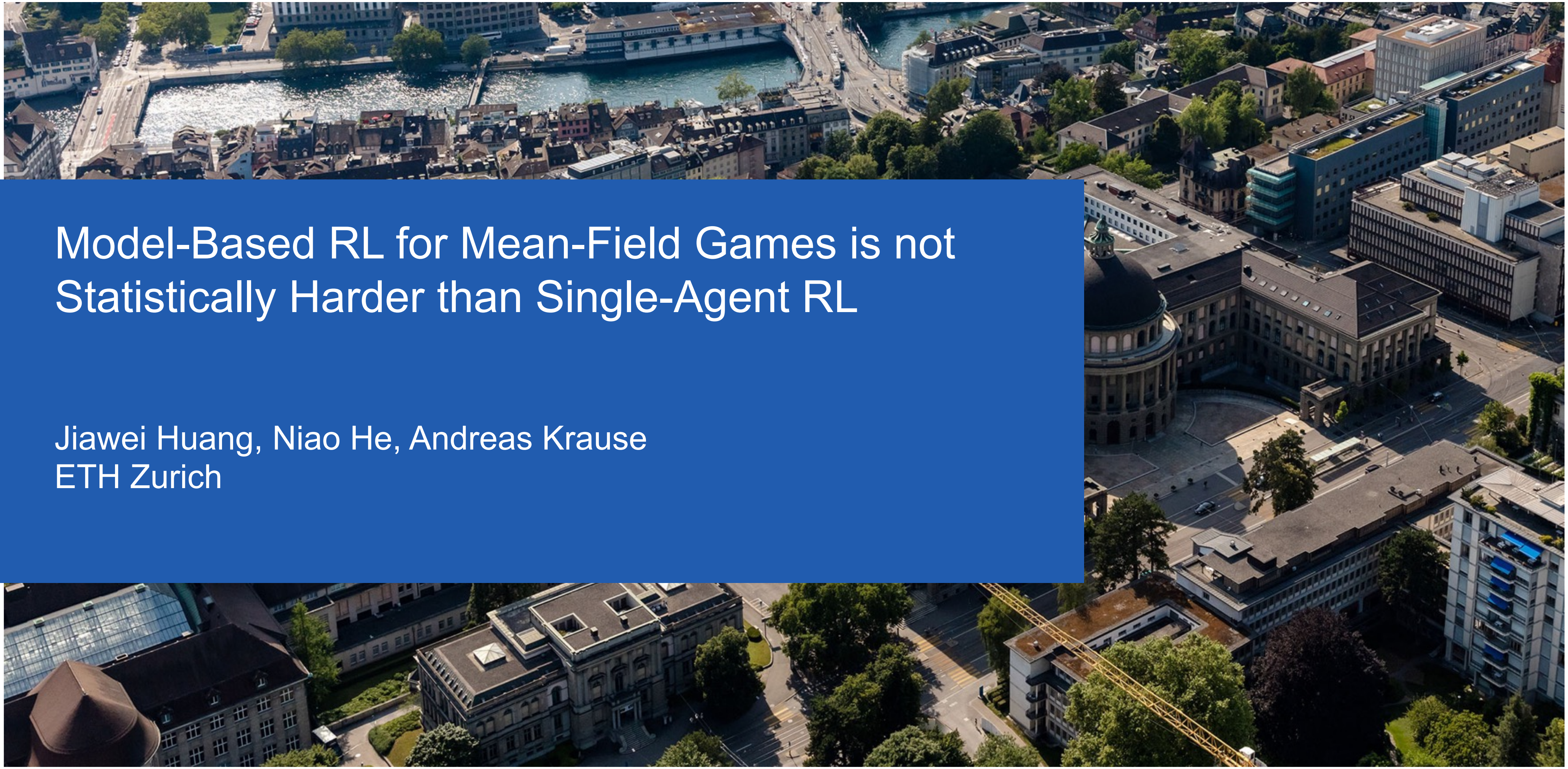
ETH zürich



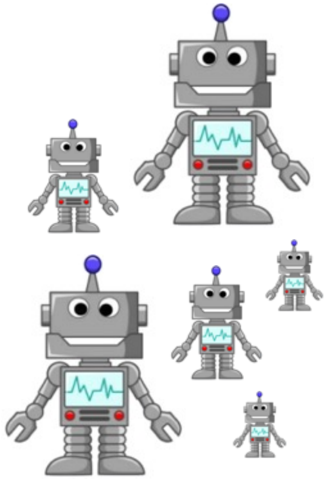
Model-Based RL for Mean-Field Games is not Statistically Harder than Single-Agent RL

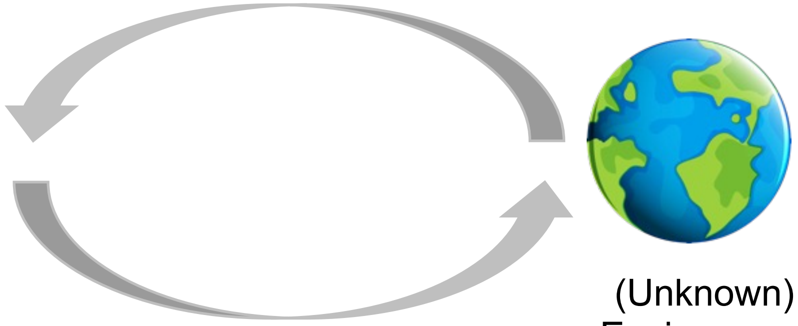
Jiawei Huang, Niao He, Andreas Krause
ETH Zurich

D INFK

Many Real-World Scenarios are Multi-Agent Systems

For each agent $n = 1, \dots, N$
- Rewards $r^n \sim r^n(s^1, \dots, s^N, a^1, \dots, a^n)$,
- Transition $s^{n\prime} \sim \mathbb{P}^n(\cdot \mid s^n, \dots, s^N, a^1, \dots, a^n)$

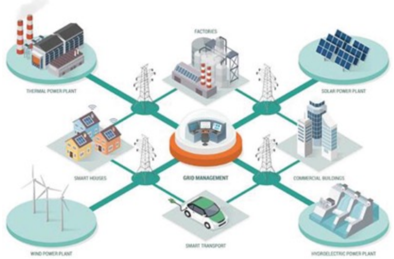




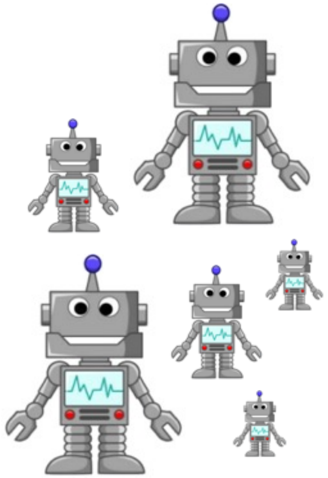
(Unknown)
Environment

N Agent

For each agent $n = 1, \dots, N$
- Actions $a^n \sim \pi^n(\cdot \mid s^n)$

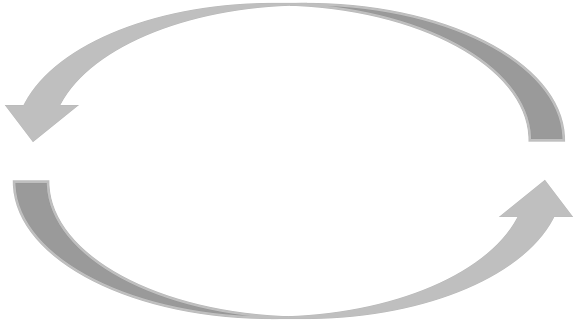


Many Real-World Scenarios are Multi-Agent Systems



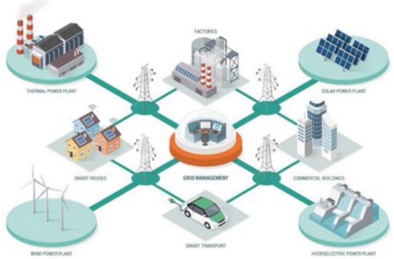
N Agent

For each agent $n = 1, \dots, N$
- Rewards $r^n \sim r^n(s^1, \dots, s^N, a^1, \dots, a^n)$,
- Transition $s^{n\prime} \sim \mathbb{P}^n(\cdot \mid s^n, \dots, s^N, a^1, \dots, a^n)$



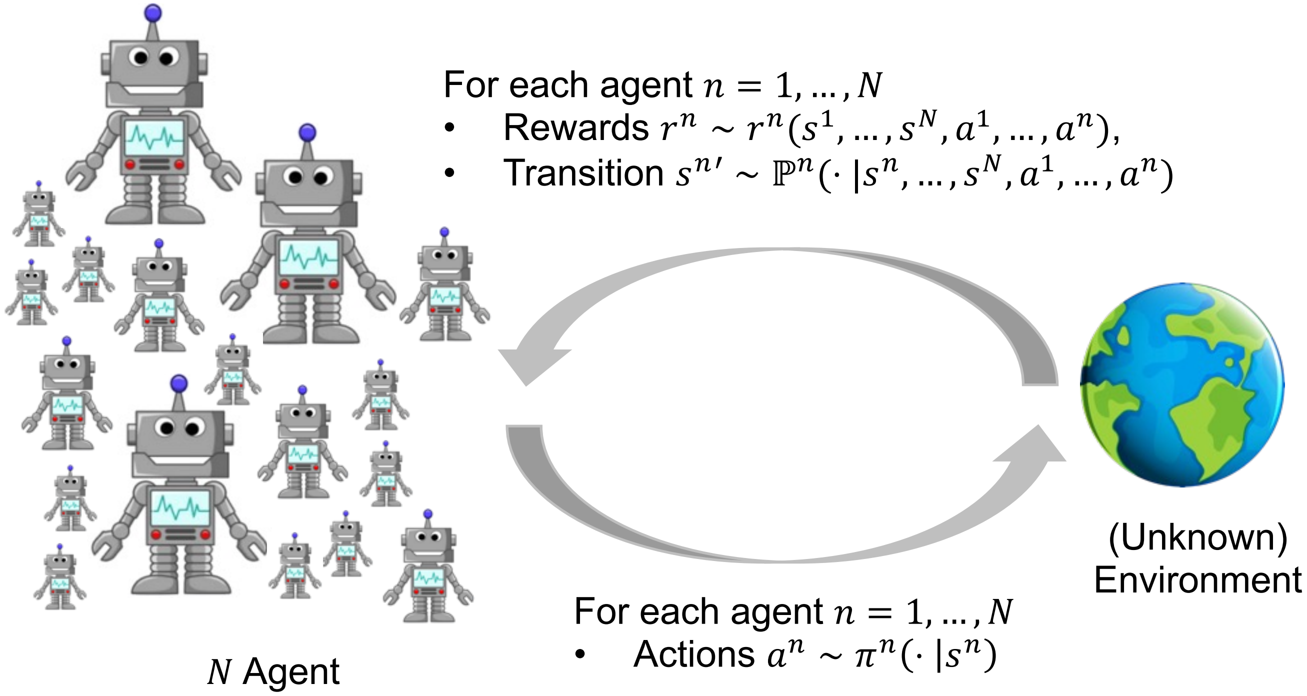
For each agent $n = 1, \dots, N$
- Actions $a^n \sim \pi^n(\cdot \mid s^n)$

(Unknown)
Environment



Main Objective: Learn equilibrium policies $\pi^1, \dots, \pi^N$, *s.t.* no agent can increase its return by deviation.

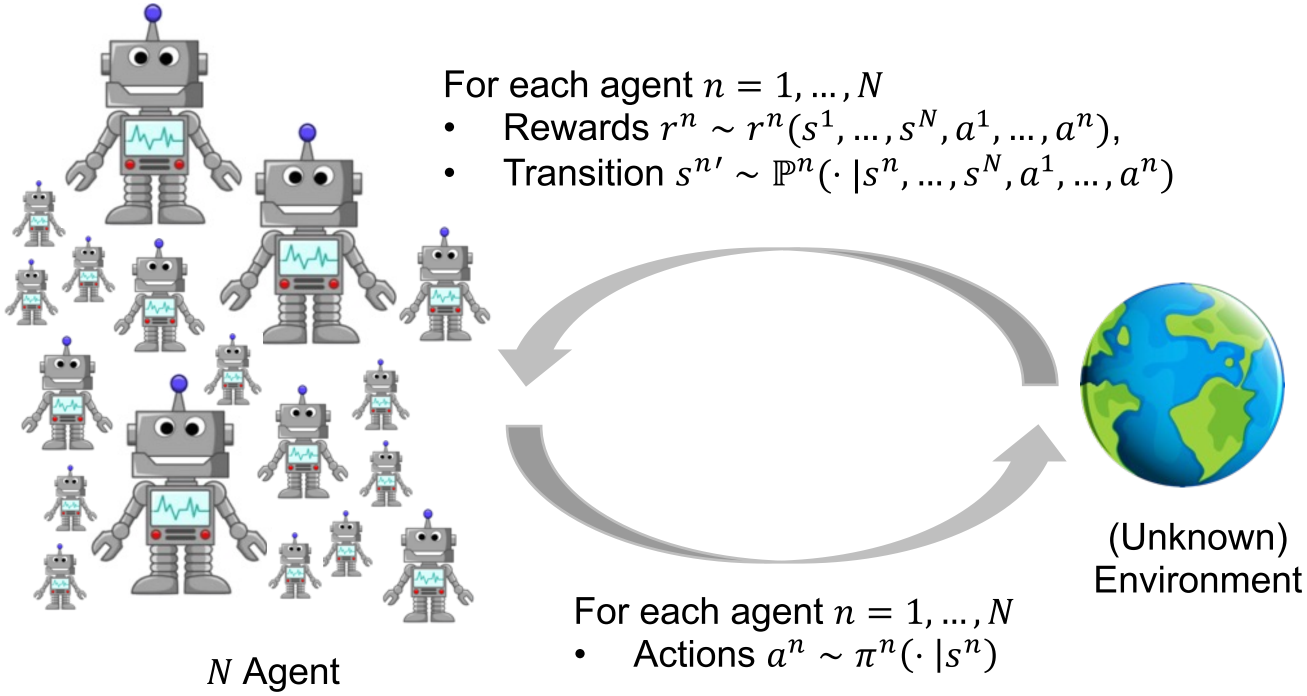
Challenges for Large N



Main Challenge: Curse of “Multi-Agency”

State space grows as $\exp(N)$, and exploration over entire state space is intractable.

Challenges for Large N

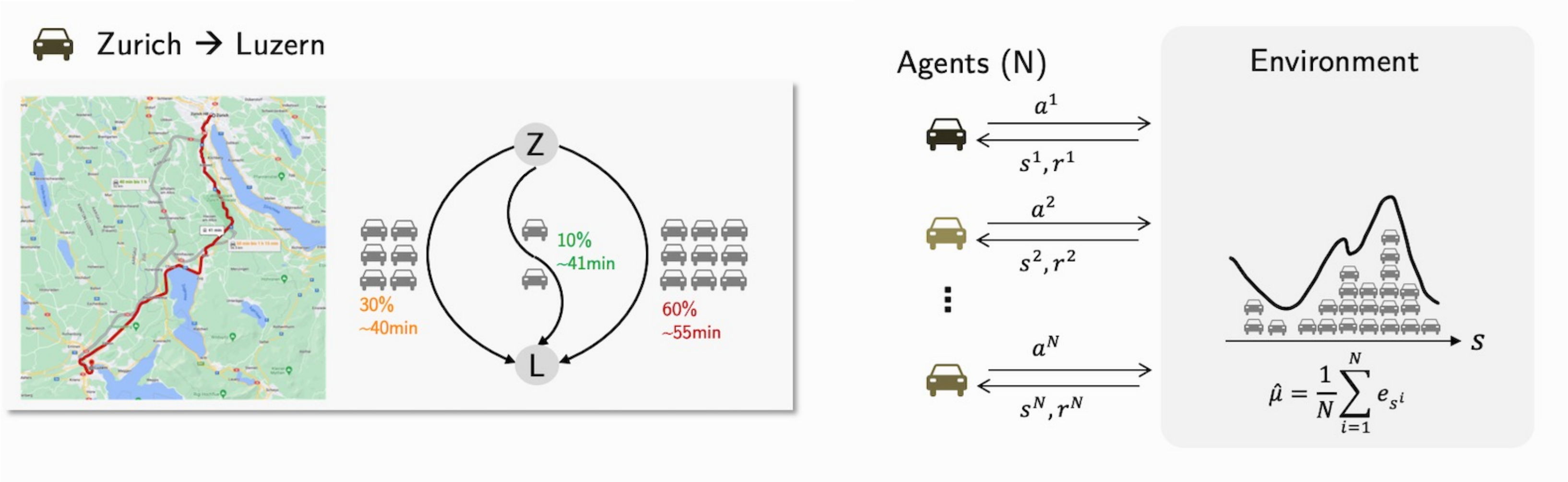


Main Challenge: Curse of “Multi-Agency”

State space grows as $\exp(N)$, and exploration over entire state space is intractable.

Leveraging additional problem structures can be helpful!

N-Player Games with Homogeneous Agents

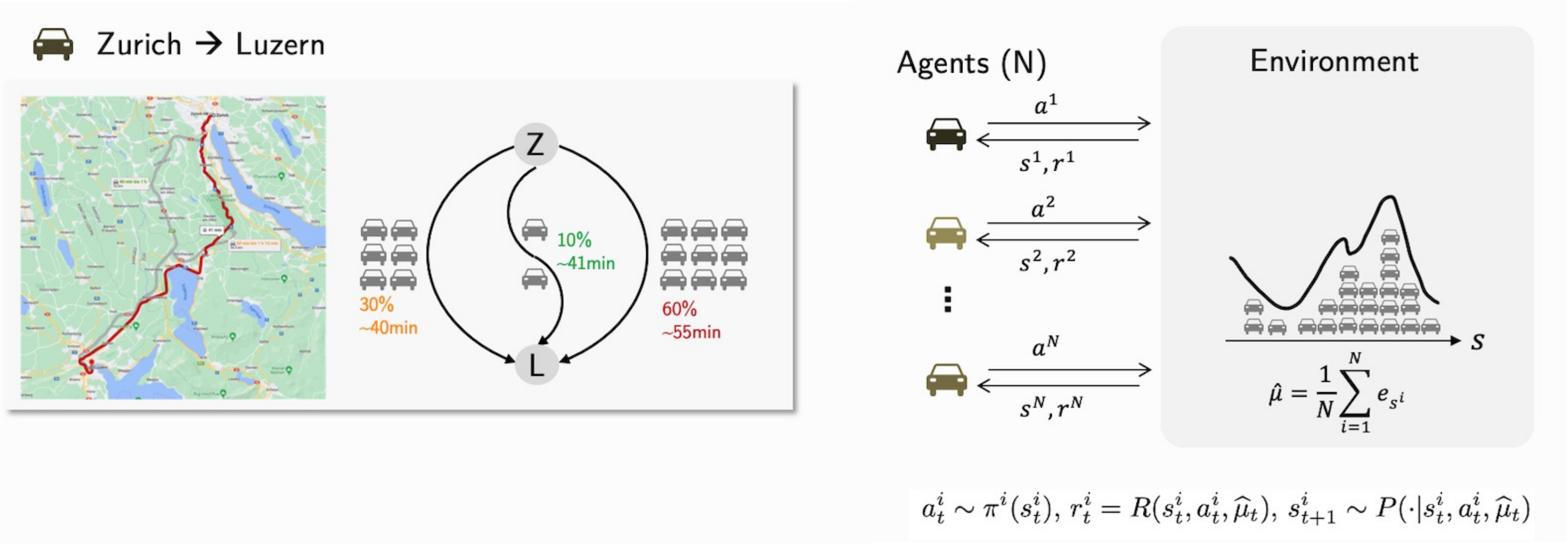


$$a_t^i \sim \pi^i(s_t^i),\ r_t^i = R(s_t^i, a_t^i, \widehat{\mu}_t),\ s_{t+1}^i \sim P(\cdot|s_t^i, a_t^i, \widehat{\mu}_t)$$

Special Structures: Symmetricity

- All the agents share the state/action space and dynamics
- transitions & rewards only affected through empirical state (action) distributions

N-Player Games with Homogeneous Agents

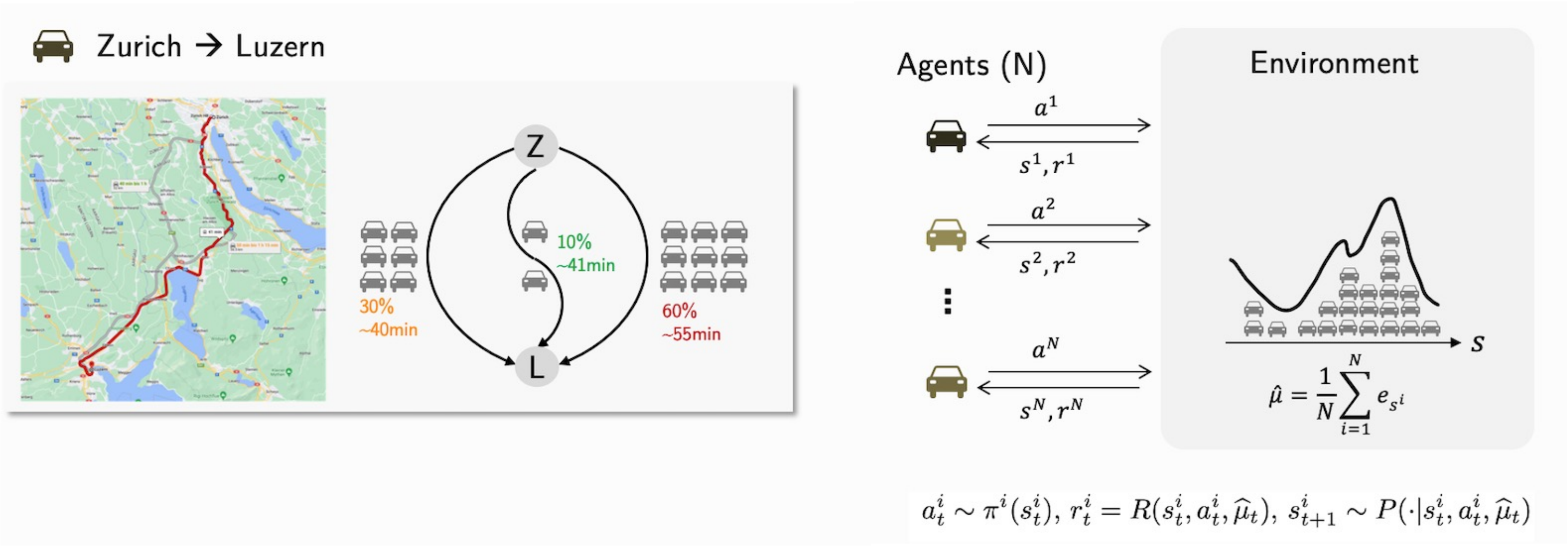


Special Structures: Symmetricity

- All the agents share the state/action space and dynamics
- transitions & rewards only affected through empirical state (action) distributions

When N is large, known as **Mean-Field Games** [Huang et al., 2006], [Lasry & Lions, 2007]

N-Player Games with Homogeneous Agents



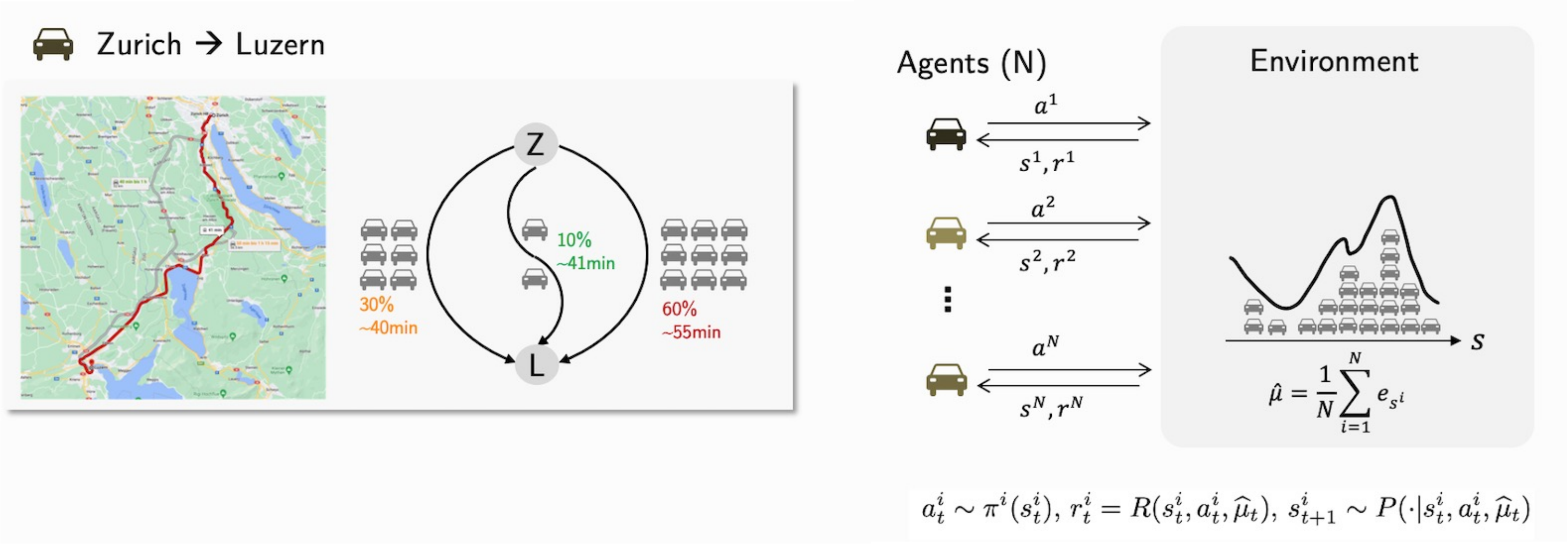
Special Structures: Symmetricity

- All the agents share the state/action space and dynamics
- transitions & rewards only affected through empirical state (action) distributions

When N is large, known as **Mean-Field Games** [Huang et al., 2006], [Lasry & Lions, 2007]

Various applications in **finance** [Carmona, 2020], **economics** [Gomes et al., 2017], **industrial engineering** [Paola et al., 2019]

N-Player Games with Homogeneous Agents



Special Structures: Symmetricity

- All the agents share the state/action space and dynamics
- transitions & rewards only affected through empirical state (action) distributions

When N is large, known as **Mean-Field Games** [Huang et al., 2006], [Lasry & Lions, 2007]

Various applications in **finance** [Carmona, 2020], **economics** [Gomes et al., 2017], **industrial engineering** [Paola et al., 2019]

Our main focus!

Mean-Field Games

(Episodic) Mean-Field Games

- $M \coloneqq (\mu_1, \mathcal{S}, \mathcal{A}, H, \mathbb{P}, r)$
- If all agents take the same $\pi \coloneqq \{\pi_1, \dots, \pi_H\}$
- For a representative agent,
  - start with $s_1 \sim \mu_1$
    - for $h = 1, \dots, H$:
      - Take action $a_h \sim \pi_h(\cdot \mid s_h)$
      - Observe next state $s_{h+1} \sim \mathbb{P}_h(\cdot \mid s_h, a_h, \mu_h^\pi)$, and reward $r_h \sim r(s_h, a_h, \mu_h^\pi)$
      - Density evolves $\mu_{h+1}^\pi(\cdot) = \sum_{s_h, a_h} \mu_h^\pi(s_h) \pi(a_h \mid s_h) \mathbb{P}_h(\cdot \mid s_h, a_h, \mu_h^\pi)$
- No curse of multi-agency issue
  - Only the density matters

Mean-Field Games

(Episodic) Mean-Field Games

- Nash Equilibrium (NE)
  - Given a policy π and a possible deviation $\tilde{\pi}$, define

$$J_M(\tilde{\pi}, \pi) \coloneqq \mathbb{E}_{\tilde{\pi}, M(\pi)}[\sum_{h \in [H]} r(s_h, a_h, \mu_h^\pi)]$$

Mean-Field Games

(Episodic) Mean-Field Games

- Nash Equilibrium (NE)
  - Given a policy π and a possible deviation $\tilde{\pi}$, define

$$J_M(\tilde{\pi}, \pi) := \mathbb{E}_{\tilde{\pi}, M(\pi)}[\sum_{h \in [H]} r(s_h, a_h, \mu_h^\pi)]$$

Short note of $\mathbb{E}[\cdot | \forall h, a_h \sim \tilde{\pi}, s_h \sim \mathbb{P}(\cdot\, | s_h, a_h, \mu_h^\pi), r_h \sim r(s_h, a_h, \mu_h^\pi)]$

Mean-Field Games

(Episodic) Mean-Field Games

- Nash Equilibrium (NE)
  - Given a policy π and a possible deviation $\tilde{\pi}$, define

$$J_M(\tilde{\pi}, \pi) \coloneqq \mathbb{E}_{\tilde{\pi}, M(\pi)}[\sum_{h \in [H]} r(s_h, a_h, \mu_h^\pi)]$$

  - NE is defined to be π_M^{NE}

$$\forall \pi, \qquad J_M\left(\pi_M^{\mathrm{NE}}, \pi_M^{\mathrm{NE}}\right) \geq J_M(\pi, \pi_M^{\mathrm{NE}})$$

  - ϵ- NE is defined to be $\hat{\pi}_M^{\mathrm{NE}}$

$$\forall \pi, \qquad J_M\left(\hat{\pi}_M^{\mathrm{NE}}, \hat{\pi}_M^{\mathrm{NE}}\right) \geq J_M\left(\pi, \hat{\pi}_M^{\mathrm{NE}}\right) - \epsilon$$

Key Question and Motivation

Key Question:
What is the fundamental sample complexity of learn ϵ-NE in MFGs with ***general function approximation***?

Key Question and Motivation

Key Question:
What is the fundamental sample complexity of learn ϵ-NE in MFGs with ***general function approximation***?

1. Usually, the learner does not have full knowledge about the Mean-Field model M
   - Efficient exploration is needed

Key Question and Motivation

Key Question:
What is the fundamental sample complexity of learn ϵ-NE in MFGs with ***general function approximation***?

1. Usually, the learner does not have full knowledge about the Mean-Field model M
   - Efficient exploration is needed

2. Real-world applications has rich observation and action spaces
   - Function approximation (e.g. neural networks) is necessary

Key Question and Motivation

Key Question:
What is the fundamental sample complexity of learn ϵ-NE in MFGs with ***general function approximation***?

1. Usually, the learner does not have full knowledge about the Mean-Field model M
   - Efficient exploration is needed

2. Real-world applications has rich observation and action spaces
   - Function approximation (e.g. neural networks) is necessary

3. Understanding of sample efficiency of learning NE in MFGs is limited
   - MFGs has special structure. Results in single-agent RL or Markov Games usually cannot be generalized here.

Basic Setting

- Model-Based Function Approximation
  - A model class $\mathcal{M}$ available, $|\mathcal{M}| < +\infty$
  - Only consider unknown transition in this paper
    - Each $M \in \mathcal{M}$ associates the same known reward r and different transition functions $\mathbb{P}_M$
    - Can be extended to unknown reward setting

Basic Setting

- Model-Based Function Approximation
  - A model class $\mathcal{M}$ available, $|\mathcal{M}| < +\infty$
  - Only consider unknown transition in this paper
    - Each $M \in \mathcal{M}$ associates the same known reward r and different transition functions $\mathbb{P}_M$
    - Can be extended to unknown reward setting
- Assumptions
  - **Realizability**: true unknown model $M^* \in \mathcal{M}$
  - **Lipschitz in density**: $\forall M \in \mathcal{M}, \forall h \in [H], \forall \pi, \pi'$
    - $||\mathbb{P}_M(\cdot \mid s_h, a_h, \mu^{\pi}_{M,h}) - \mathbb{P}_M(\cdot \mid s_h, a_h, \mu^{\pi'}_{M,h})||_1 \leq L_T ||\mu^{\pi}_{M,h} - \mu^{\pi'}_{M,h}||_1$
    - $|r(s_h, a_h, \mu^{\pi}_{M,h}) - r(s_h, a_h, \mu^{\pi'}_{M,h})| \leq L_r ||\mu^{\pi}_{M,h} - \mu^{\pi'}_{M,h}||_1$
    - NE may not exist if not Lipshictz

Basic Setting

- Model-Based Function Approximation
  - A model class $\mathcal{M}$ available, $|\mathcal{M}| < +\infty$
  - Only consider unknown transition in this paper
    - Each $M \in \mathcal{M}$ associates the same known reward r and different transition functions $\mathbb{P}_M$
    - Can be extended to unknown reward setting
- Assumptions
  - **Realizability**: true unknown model $M^* \in \mathcal{M}$
  - **Lipschitz in density**: $\forall M \in \mathcal{M}, \forall h \in [H], \forall \pi, \pi'$
    - $||\mathbb{P}_M(\cdot|s_h, a_h, \mu^{\pi}_{M,h}) - \mathbb{P}_M(\cdot|s_h, a_h, \mu^{\pi'}_{M,h})||_1 \le L_T||\mu^{\pi}_{M,h} - \mu^{\pi'}_{M,h}||_1$
    - $|r(s_h, a_h, \mu^{\pi}_{M,h}) - r(s_h, a_h, \mu^{\pi'}_{M,h})| \le L_r||\mu^{\pi}_{M,h} - \mu^{\pi'}_{M,h}||_1$
    - NE may not exist if not Lipshictz
- Comparing with assumptions in MFGs literature:

Contractivity [Guo et al. 2019]	Monotonicity [Perolat et al. 2021]
L_T, L_r are small and others	$L_T = 0$ and others

Main Results

Main Theorem (Informal)

Learning ϵ-NE in MFG is as sample-efficient as solving $\log|\mathcal{M}|$ single-agent RL problems

Sample complexity = $\mathrm{Poly}(1 + L_r, 1 + L_P, \frac{1}{\epsilon}, H, \log\frac{|\mathcal{M}|}{\delta}, \mathrm{dimPE}(\mathcal{M}),)$

Main Results

Main Theorem (Informal)

Learning ϵ-NE in MFG is as sample-efficient as solving $\log|\mathcal{M}|$ single-agent RL problems

Sample complexity = $\mathrm{Poly}(1 + L_r, 1 + L_P, \frac{1}{\epsilon}, H, \log\frac{|\mathcal{M}|}{\delta}, \mathrm{dimPE}(\mathcal{M}),)$

Partial Model-Based Eluder Dimension: Can be regarded as the complexity of the ***single-agent*** model class converted from $\mathcal{M}$

Main Results

Main Theorem (Informal)

Learning ϵ-NE in MFG is as sample-efficient as solving $\log|\mathcal{M}|$ single-agent RL problems

Sample complexity = $\mathrm{Poly}(1 + L_r, 1 + L_P, \frac{1}{\epsilon}, H, \log\frac{|\mathcal{M}|}{\delta}, \mathrm{dimPE}(\mathcal{M}),)$

Partial Model-Based Eluder Dimension: Can be regarded as the complexity of the ***single-agent*** model class converted from $\mathcal{M}$

- Concrete Examples
  - Tabular MFG: $\mathrm{dimPE}(\mathcal{M}) \leq SA$
    - Tabular MFG is sample-efficient in general
  - Linear MFG: $\mathrm{dimPE}(\mathcal{M}) \leq d$
    - $\mathbb{P}(s'|s,a,\mu) = \phi(s,a)^\top U(\mu)\psi(s')$ with known $\phi(s,a) \in \mathbb{R}^d$

Algorithm Details

For $\mathrm{k} = 1,2 \ldots, O(\log|\mathcal{M}|)$, (start with $\mathcal{M}^1 = \mathcal{M}$)

- Find a desired policy π^k
- Construct $\mathcal{M}^k_{|\pi^k} \coloneqq \{M_{|\pi^k}, M \in \mathcal{M}^k\}$
  - $M_{|\pi^k}$ is the single-agent model with $\left\{r\left(\cdot,\cdot,\mu^{\pi^k}_{M,h}\right), \mathbb{P}_M\left(\cdot \,\middle|\, \cdot,\cdot,\mu^{\pi^k}_{M,h}\right)\right\}_{h\in[H]}$
- Collect samples and $\mathcal{M}^{k+1} \leftarrow \{M \in \mathcal{M}^k \big| M_{|\pi^k} \text{ agrees with } M^*_{|\pi^k}\}$

Algorithm Details

For $k = 1,2 \ldots, O(\log|\mathcal{M}|)$, (start with $\mathcal{M}^1 = \mathcal{M}$)

- Find a desired policy π^k
- Construct $\mathcal{M}^k_{|\pi^k} := \{M_{|\pi^k}, M \in \mathcal{M}^k\}$
  - $M_{|\pi^k}$ is the single-agent model with $\left\{r\left(\cdot,\cdot,\mu^{\pi^k}_{M,h}\right), \mathbb{P}_M\left(\cdot \,\middle|\, \cdot,\cdot,\mu^{\pi^k}_{M,h}\right)\right\}_{h\in[H]}$
- Collect samples and $\mathcal{M}^{k+1} \leftarrow \{M \in \mathcal{M}^k | M_{|\pi^k} \text{ agrees with } M^*_{|\pi^k}\}$

- mode elimination in single-agent function class
- the only step require samples, origin of dependence on $\text{dimPE}(\mathcal{M})$
- any single-agent model learning algorithm.

Algorithm Details

For $\mathrm{k} = 1,2 \ldots, O(\log|\mathcal{M}|)$, (start with $\mathcal{M}^1 = \mathcal{M}$)

- Find a **desired policy** π^k
- Construct $\mathcal{M}^k_{|\pi^k} := \{M_{|\pi^k}, M \in \mathcal{M}^k\}$
  - $M_{|\pi^k}$ is the single-agent model with $\left\{r\left(\cdot,\cdot,\mu^{\pi^k}_{M,h}\right), \mathbb{P}_M\left(\cdot \,\middle|\, \cdot,\cdot,\mu^{\pi^k}_{M,h}\right)\right\}_{h\in[H]}$
- Collect samples and $\mathcal{M}^{k+1} \leftarrow \{M \in \mathcal{M}^k \,|\, M_{|\pi^k} \text{ agrees with } M^*_{|\pi^k}\}$

How to find the "desired policy" for fast learning?

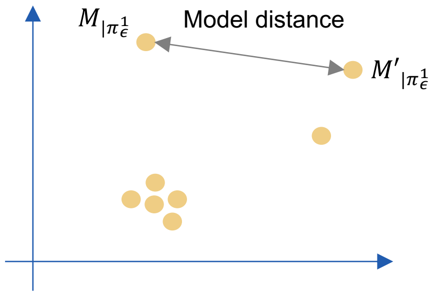
Algorithm Details

For $k = 1,2 \ldots, O(\log|\mathcal{M}|)$, (start with $\mathcal{M}^1 = \mathcal{M}$)

- Find a **desired policy** π^k
- Construct $\mathcal{M}^k_{|\pi^k} := \{M_{|\pi^k}, M \in \mathcal{M}^k\}$
  - $M_{|\pi^k}$ is the single-agent model with $\left\{r\left(\cdot,\cdot,\mu^{\pi^k}_{M,h}\right), \mathbb{P}_M\left(\cdot \,\middle|\, \cdot,\cdot,\mu^{\pi^k}_{M,h}\right)\right\}_{h\in[H]}$
- Collect samples and $\mathcal{M}^{k+1} \leftarrow \{M \in \mathcal{M}^k \,|\, M_{|\pi^k} \text{ agrees with } M^*_{|\pi^k}\}$

How to find the "desired policy" for fast learning?

Consider ϵ-cover for policy space $\Pi_\epsilon := \{\pi^1_\epsilon, \pi^2_\epsilon \ldots\}$



Visualization of $\mathcal{M}^k_{|\pi^1_\epsilon}$

(dots represent models in $\mathcal{M}^k_{|\pi^1_\epsilon}$)

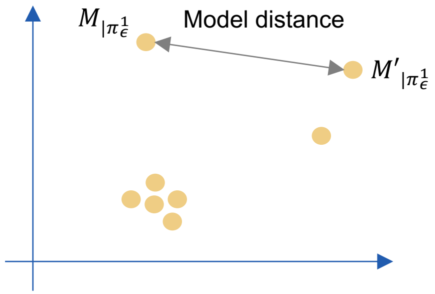
Algorithm Details

For $\mathrm{k} = 1,2 \ldots, O(\log|\mathcal{M}|)$, (start with $\mathcal{M}^1 = \mathcal{M}$)

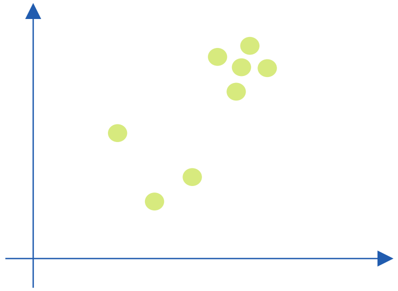
- Find a **desired policy** π^k
- Construct $\mathcal{M}^k_{|\pi^k} \coloneqq \{M_{|\pi^k}, M \in \mathcal{M}^k\}$
  - $M_{|\pi^k}$ is the single-agent model with $\left\{r\left(\cdot,\cdot, \mu^{\pi^k}_{M,h}\right), \mathbb{P}_M\left(\cdot \,\middle|\, \cdot,\cdot, \mu^{\pi^k}_{M,h}\right)\right\}_{h\in[H]}$
- Collect samples and $\mathcal{M}^{k+1} \leftarrow \{M \in \mathcal{M}^k \,|\, M_{|\pi^k} \text{ agrees with } M^*_{|\pi^k}\}$

How to find the “desired policy” for fast learning?

Consider ϵ-cover for policy space $\Pi_\epsilon \coloneqq \{\pi^1_\epsilon, \pi^2_\epsilon \ldots\}$

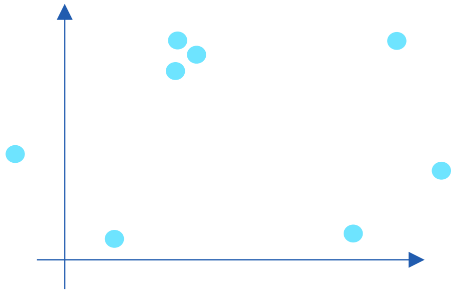


Visualization of $\mathcal{M}^k_{|\pi^1_\epsilon}$
(dots represent models in $\mathcal{M}^k_{|\pi^1_\epsilon}$)



Visualization of $\mathcal{M}^k_{|\pi^2_\epsilon}$
(dots represent models in $\mathcal{M}^k_{|\pi^2_\epsilon}$)

…



Visualization of $\mathcal{M}^k_{|\pi^i_\epsilon}$
(dots represent models in $\mathcal{M}^k_{|\pi^i_\epsilon}$)

…

Algorithm Details

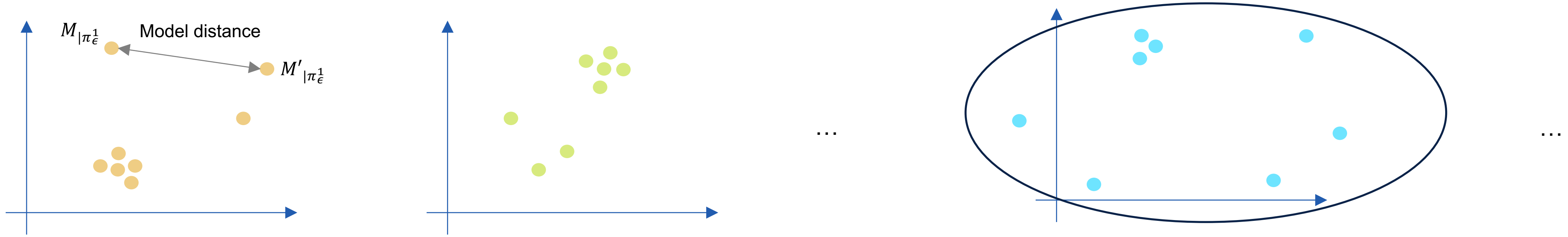
For $\mathrm{k} = 1,2 \dots, O(\log|\mathcal{M}|)$, (start with $\mathcal{M}^1 = \mathcal{M}$)

- Find a **desired policy** π^k
- Construct $\mathcal{M}^k_{|\pi^k} := \{M_{|\pi^k}, M \in \mathcal{M}^k\}$
  - $M_{|\pi^k}$ is the single-agent model with $\left\{r\left(\cdot,\cdot,\mu^{\pi^k}_{M,h}\right), \mathbb{P}_M\left(\cdot \middle| \cdot,\cdot,\mu^{\pi^k}_{M,h}\right)\right\}_{h\in[H]}$
- Collect samples and $\mathcal{M}^{k+1} \leftarrow \{M \in \mathcal{M}^k | M_{|\pi^k} \text{ agrees with } M^*_{|\pi^k}\}$

How to find the “desired policy” for fast learning?

- Case 1 (Non-Concentrated): Existence of π^i_ϵ, s.t. no model cluster contains half of models

Consider ϵ-cover for policy space $\Pi_\epsilon := \{\pi^1_\epsilon, \pi^2_\epsilon \dots\}$



Visualization of $\mathcal{M}^k_{|\pi^1_\epsilon}$ (dots represent models in $\mathcal{M}^k_{|\pi^1_\epsilon}$)

Visualization of $\mathcal{M}^k_{|\pi^2_\epsilon}$ (dots represent models in $\mathcal{M}^k_{|\pi^2_\epsilon}$)

Visualization of $\mathcal{M}^k_{|\pi^i_\epsilon}$ (dots represent models in $\mathcal{M}^k_{|\pi^i_\epsilon}$)

Algorithm Details

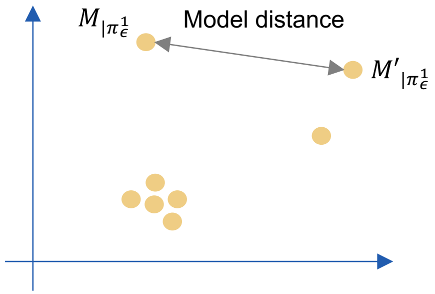
For $\mathrm{k} = 1,2 \ldots, O(\log|\mathcal{M}|)$, (start with $\mathcal{M}^1 = \mathcal{M}$)

- Find a **desired policy** π^k
- Construct $\mathcal{M}^k_{|\pi^k} := \{M_{|\pi^k}, M \in \mathcal{M}^k\}$
  - $M_{|\pi^k}$ is the single-agent model with $\left\{r\left(\cdot,\cdot, \mu^{\pi^k}_{M,h}\right), \mathbb{P}_M\left(\cdot \,\middle|\, \cdot,\cdot, \mu^{\pi^k}_{M,h}\right)\right\}_{h\in[H]}$
- Collect samples and $\mathcal{M}^{k+1} \leftarrow \{M \in \mathcal{M}^k | M_{|\pi^k} \text{ agrees with } M^*_{|\pi^k}\}$

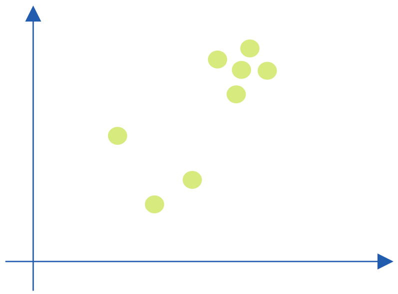
How to find the "desired policy" for fast learning?

- Case 1 (Non-Concentrated): Existence of π^i_ϵ, s.t. no model cluster contains half of models
  - Set $\pi^k = \pi^i_\epsilon$,, then $|\mathcal{M}^{k+1}| \leq \frac{|\mathcal{M}^k|}{2}$

Consider ϵ-cover for policy space $\Pi_\epsilon := \{\pi^1_\epsilon, \pi^2_\epsilon \ldots\}$

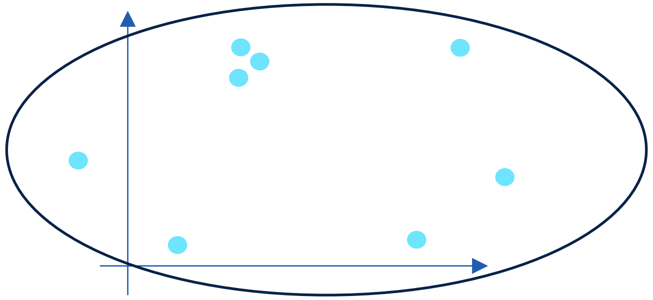


Visualization of $\mathcal{M}^k_{|\pi^1_\epsilon}$
(dots represent models in $\mathcal{M}^k_{|\pi^1_\epsilon}$)



Visualization of $\mathcal{M}^k_{|\pi^2_\epsilon}$
(dots represent models in $\mathcal{M}^k_{|\pi^2_\epsilon}$)

…



Visualization of $\mathcal{M}^k_{|\pi^i_\epsilon}$
(dots represent models in $\mathcal{M}^k_{|\pi^i_\epsilon}$)

…

Algorithm Details

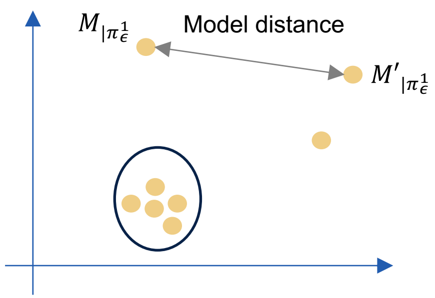
For $k = 1,2 \ldots, O(\log|\mathcal{M}|)$, (start with $\mathcal{M}^1 = \mathcal{M}$)

- Find a **desired policy** π^k
- Construct $\mathcal{M}^k_{|\pi^k} := \{M_{|\pi^k}, M \in \mathcal{M}^k\}$
  - $M_{|\pi^k}$ is the single-agent model with $\left\{r\left(\cdot,\cdot,\mu^{\pi^k}_{M,h}\right), \mathbb{P}_M\left(\cdot \,\middle|\, \cdot,\cdot,\mu^{\pi^k}_{M,h}\right)\right\}_{h\in[H]}$
- Collect samples and $\mathcal{M}^{k+1} \leftarrow \{M \in \mathcal{M}^k | M_{|\pi^k} \text{ agrees with } M^*_{|\pi^k}\}$

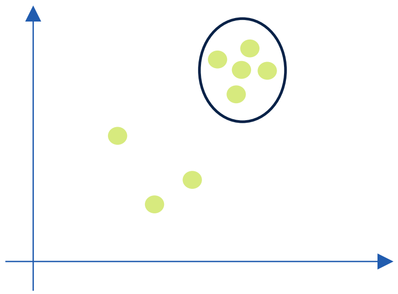
How to find the "desired policy" for fast learning?

- Case 2 (Concentrated): $\forall \pi \in \Pi_\epsilon$, there is a cluster with half of models

Consider ϵ-cover for policy space $\Pi_\epsilon := \{\pi^1_\epsilon, \pi^2_\epsilon \ldots\}$

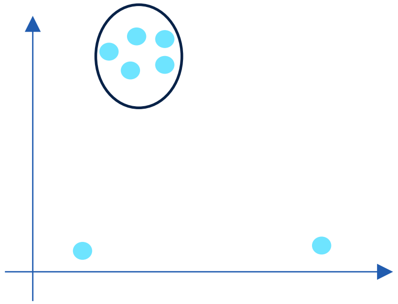


Visualization of $\mathcal{M}^k_{|\pi^1_\epsilon}$
(dots represent models in $\mathcal{M}^k_{|\pi^1_\epsilon}$)



Visualization of $\mathcal{M}^k_{|\pi^2_\epsilon}$
(dots represent models in $\mathcal{M}^k_{|\pi^2_\epsilon}$)

…



Visualization of $\mathcal{M}^k_{|\pi^i_\epsilon}$
(dots represent models in $\mathcal{M}^k_{|\pi^i_\epsilon}$)

…

Algorithm Details

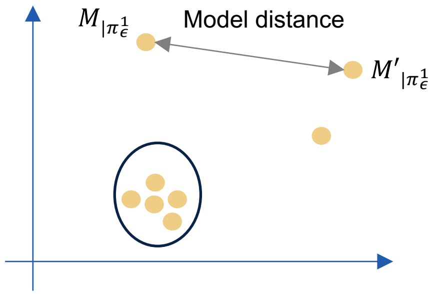
For $\mathrm{k} = 1,2 \ldots, O(\log|\mathcal{M}|)$, (start with $\mathcal{M}^1 = \mathcal{M}$)

- Find a **desired policy** π^k
- Construct $\mathcal{M}^k_{|\pi^k} := \{M_{|\pi^k}, M \in \mathcal{M}^k\}$
  - $M_{|\pi^k}$ is the single-agent model with $\left\{r\left(\cdot,\cdot,\mu^{\pi^k}_{M,h}\right), \mathbb{P}_M\left(\cdot \,\middle|\, \cdot,\cdot,\mu^{\pi^k}_{M,h}\right)\right\}_{h\in[H]}$
- Collect samples and $\mathcal{M}^{k+1} \leftarrow \{M \in \mathcal{M}^k | M_{|\pi^k} \text{ agrees with } M^*_{|\pi^k}\}$

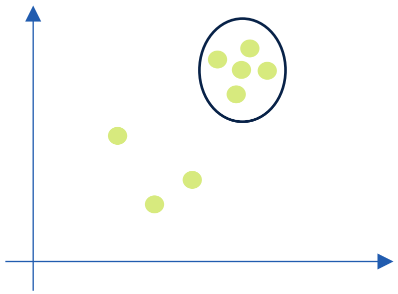
How to find the "desired policy" for fast learning?

- Case 2 (Concentrated): $\forall \pi \in \Pi_\epsilon$, there is a cluster with half of models
  - **Key Observation [Local Alignment Lemma]:** If $M_{|\pi} \approx M^*_{|\pi}$ and π is NE of M, then $\pi \approx$ NE of M^*

Consider ϵ-cover for policy space $\Pi_\epsilon := \{\pi^1_\epsilon, \pi^2_\epsilon \ldots\}$

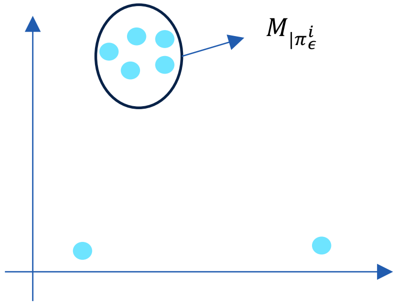


Visualization of $\mathcal{M}^k_{|\pi^1_\epsilon}$
(dots represent models in $\mathcal{M}^k_{|\pi^1_\epsilon}$)



Visualization of $\mathcal{M}^k_{|\pi^2_\epsilon}$
(dots represent models in $\mathcal{M}^k_{|\pi^2_\epsilon}$)

…



Visualization of $\mathcal{M}^k_{|\pi^i_\epsilon}$
(dots represent models in $\mathcal{M}^k_{|\pi^i_\epsilon}$)

…

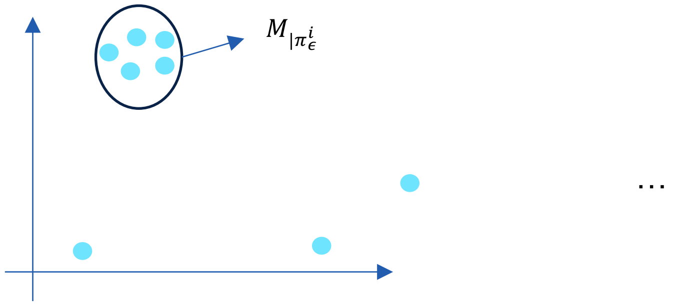
Algorithm Details

For $\mathrm{k} = 1,2 \ldots, O(\log|\mathcal{M}|)$, (start with $\mathcal{M}^1 = \mathcal{M}$)

- Find a **desired policy** π^k
- Construct $\mathcal{M}^k_{|\pi^k} \coloneqq \{M_{|\pi^k}, M \in \mathcal{M}^k\}$
  - $M_{|\pi^k}$ is the single-agent model with $\left\{r\left(\cdot,\cdot,\mu_{M,h}^{\pi^k}\right), \mathbb{P}_M\left(\cdot \,\middle|\, \cdot,\cdot,\mu_{M,h}^{\pi^k}\right)\right\}_{h\in[H]}$
- Collect samples and $\mathcal{M}^{k+1} \leftarrow \{M \in \mathcal{M}^k | M_{|\pi^k} \text{ agrees with } M^*_{|\pi^k}\}$

How to find the “desired policy” for fast learning?

- Case 2 (Concentrated): $\forall \pi \in \Pi_\epsilon$, there is a cluster with half of models
  - **Key Observation [Local Alignment Lemma]:** If $M_{|\pi} \approx M^*_{|\pi}$ and π is NE of M, then $\pi \approx$ NE of M^*
  - Under Lipschitz continuity, by construction (Algorithm 3 in paper, omitted here), there exists a π_ϵ^i, such that $\pi_\epsilon^i \approx$ NE of any $M_{|\pi_\epsilon^i} \in$ 



Visualization of $\mathcal{M}^k_{|\pi_\epsilon^i}$
(dots represent models in $\mathcal{M}^k_{|\pi_\epsilon^i}$)

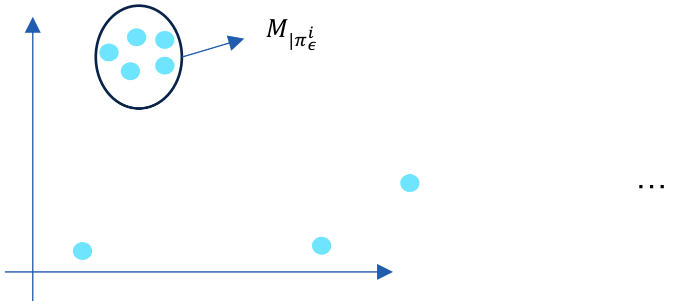
Algorithm Details

For $\mathrm{k} = 1,2 \ldots, O(\log|\mathcal{M}|)$, (start with $\mathcal{M}^1 = \mathcal{M}$)

- Find a **desired policy** π^k
- Construct $\mathcal{M}^k_{|\pi^k} \coloneqq \{M_{|\pi^k}, M \in \mathcal{M}^k\}$
  - $M_{|\pi^k}$ is the single-agent model with $\left\{r\left(\cdot,\cdot,\mu^{\pi^k}_{M,h}\right), \mathbb{P}_M\left(\cdot \,\middle|\, \cdot,\cdot,\mu^{\pi^k}_{M,h}\right)\right\}_{h\in[H]}$
- Collect samples and $\mathcal{M}^{k+1} \leftarrow \{M \in \mathcal{M}^k | M_{|\pi^k} \text{ agrees with } M^*_{|\pi^k}\}$

How to find the "desired policy" for fast learning?

- Case 2 (Concentrated): $\forall \pi \in \Pi_\epsilon$, there is a cluster with half of models
  - **Key Observation [Local Alignment Lemma]:** If $M_{|\pi} \approx M^*_{|\pi}$ and π is NE of M, then $\pi \approx$ NE of M^*
  - Under Lipschitz continuity, by construction (Algorithm 3 in paper, omitted here), there exists a π^i_ϵ, such that $\pi^i_\epsilon \approx$ NE of any $M_{|\pi^i_\epsilon} \in$ 
    - Set $\pi^k = \pi^i_\epsilon$
      - If  $\in \mathcal{M}^{k+1}$, it implies:
        - $M^*_{|\pi^i_\epsilon} \in$  and $\pi^i_\epsilon \approx$ NE of M^*
      - If  $\notin \mathcal{M}^{k+1}$, it implies:
        - $M^*_{|\pi^i_\epsilon} \notin$  and $|\mathcal{M}^{k+1}| \le \frac{|\mathcal{M}^k|}{2}$



Visualization of $\mathcal{M}^k_{|\pi^i_\epsilon}$
(dots represent models in $\mathcal{M}^k_{|\pi^i_\epsilon}$)

Algorithm Details

For $\mathrm{k} = 1,2 \ldots, O(\log|\mathcal{M}|)$, (start with $\mathcal{M}^1 = \mathcal{M}$)

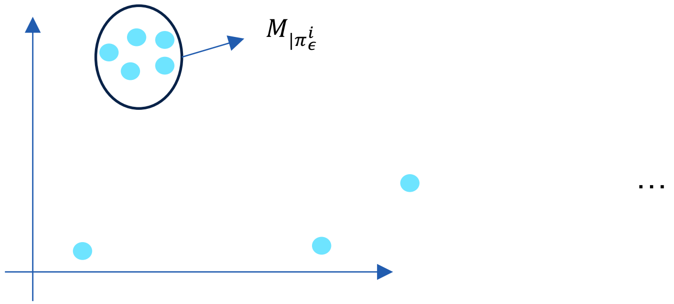
- Find a **desired policy** π^k
- Construct $\mathcal{M}^k_{|\pi^k} := \{M_{|\pi^k}, M \in \mathcal{M}^k\}$
  - $M_{|\pi^k}$ is the single-agent model with $\left\{r\left(\cdot,\cdot,\mu^{\pi^k}_{M,h}\right), \mathbb{P}_M\left(\cdot \,\middle|\, \cdot,\cdot,\mu^{\pi^k}_{M,h}\right)\right\}_{h\in[H]}$
- Collect samples and $\mathcal{M}^{k+1} \leftarrow \{M \in \mathcal{M}^k | M_{|\pi^k} \text{ agrees with } M^*_{|\pi^k}\}$

How to find the "desired policy" for fast learning?

- Case 2 (Concentrated): $\forall \pi \in \Pi_\epsilon$, there is a cluster with half of models
  - **Key Observation [Local Alignment Lemma]:** If $M_{|\pi} \approx M^*_{|\pi}$ and π is NE of M, then $\pi \approx$ NE of M^*
  - Under Lipschitz continuity, by construction (Algorithm 3 in paper, omitted here), there exists a π^i_ϵ, such that $\pi^i_\epsilon \approx$ NE of any $M_{|\pi^i_\epsilon} \in$ 
    - Set $\pi^k = \pi^i_\epsilon$
      - If  $\in \mathcal{M}^{k+1}$, it implies:
        - $M^*_{|\pi^i_\epsilon} \in$  and $\pi^i_\epsilon \approx$ NE of M^*
      - If  $\notin \mathcal{M}^{k+1}$, it implies:
        - $M^*_{|\pi^i_\epsilon} \notin$  and $|\mathcal{M}^{k+1}| \leq \frac{|\mathcal{M}^k|}{2}$

Combining with Case 1, for each k
- either $|\mathcal{M}^{k+1}| \leq |\frac{\mathcal{M}^k}{2}|$
- or $\pi^k \approx \pi^{\mathrm{NE}}$

which concludes our main results



Visualization of $\mathcal{M}^k_{|\pi^i_\epsilon}$
(dots represent models in $\mathcal{M}^k_{|\pi^i_\epsilon}$)

Summary

Take Aways

- A new complexity measure: Partial Model Based Eluder Dimension
- A novel model elimination algorithm for Mean-Field Games setting

Under realizability and Lipschitz conditions
Model-Based RL for Mean-Field Games is not Statistically Harder than Single-Agent RL

Summary

Take Aways

- A new complexity measure: Partial Model Based Eluder Dimension
- A novel model elimination algorithm for Mean-Field Games setting

Under realizability and Lipschitz conditions
Model-Based RL for Mean-Field Games is not Statistically Harder than Single-Agent RL

Future Directions

- Decentralized learning?
- Computationally efficient solutions?

Thank you!