Homogenization and Mean-field approximation(s) for systems of heterogeneous agents

Application to MultiAgent Reinforcement Learning

Rama Cont (Oxford) & Anran HU (Columbia IEOR)

Based on:
Rama Cont and Anran Hu (2025)
Homogenization and Mean-Field Approximation for Multi-Player Games

Outline

1. Heterogeneity in Mean-field games
2. Approximation results for N-Player Games of Mean-Field Type
3. Mean-Field approximations for generic Heterogeneous N-Player Games
4. Optimal partitioning of agents
5. Application to large scale MultiAgent Reinforcement Learning (MARL)
6. Example: Multiagent RL for a heterogeneous robot swarm

Mean-Field Games (MFG)

- Huang, Malhamé and Caines (2006), Lasry and Lions (2007)
- *Symmetric* game with large *homogeneous* population of (interchangeable) interacting agents whose individual impact is small.
- Exchangeability $\Rightarrow$ one may represent the game as the result of interaction between a **typical (representative) player** and an aggregate variable **"mean-field"** which only depends on the population *distribution*
  - **Large population**: a "continuum" of players
  - **Homogeneity**: players are viewed as IID instances of a common distribution
  - **Symmetric interactions**: payoffs invariant under permutation of agents
- Nash equilibria (NE) of symmetric stochastic dynamic N-player games with homogeneous players converge to MFG limit (Carmona & Delarue, 2013,...).
- Conversely: MFG may be used to construct approximate equilibria for **symmetric** N-player games with exchangeable payoff structure

MFGs as Approximations to Multiplayer Games

- Intuitively, MFGs represent the limit of a symmetric N-player dynamic games for large N. (Kolokoltsov, Li and Yang ,2011), (Carmona and Delarue, 2013) (Carmona and Lacker, 2015) (Fischer 2017) (Cardaliaguet et al 2018).
- This approximation relies on **homogeneity** of the agent population and the exchangeability of the payoffs.
- Law of large numbers is assumed to hold *exactly* in the MFG formulation: fluctuations are neglected.
- **Sample heterogeneity**: In a population of size N, even if population characteristics are indeed IID, we expect in general sample fluctuations of order $O(1/\sqrt{N})$ in mean-field variables.
- In physics problems $N \sim 10^{23}$ so $1/\sqrt{N} < 10^{-11} \ll 1$. But in socio-economic systems $N \sim 10^2 - 10^6$ so error is not negligible!
- **Population heterogeneity**: if population is not homogeneous we may expected further deviations from the MFG limit.

Incorporating Heterogeneity in Mean-Field Games

- The framework of Mean-field games can be modified to include certain forms of heterogeneity by introducing **agent types**.
- Multi-population MFG:
  - $k = 1..K$ homogeneous populations
  - subpopulation k characterized by distribution μ^k
  - agent in class k interacts with distributions $\mu^j,\ j = 1..K$
- This approach accomodates certain forms of statistical heterogeneity but still neglects effect of fluctuations in each subpopulation.
- In practice if $k-$th population has size N_k with $\sum_k N_k = N$ we have an approximation of order $1/\sqrt{N_k}$ for each mean-field variable in the best: sample fluctuations even greater!
- Deviation from statistical homogeneity in each population leads to additional fluctuations.

Motivating Example: Sequential Auction Games

Example: Bid recommendation in sequential ad auction

search query $\Longrightarrow$ (hybrid) auction[1]

Characteristics:

- A large number of (competing) advertisers (over 1MM)
- For each search query, different ads have different probabilities of being clicked (click-through rates, CTR)
- Heterogeneity in CTR values leads to heterogeneity of payoffs across advertisers: higher CTR value $\Rightarrow$ larger payoffs *in same state*
- Clustering by CTR values lead to approximate (but not exact) homogeneity within each cluster (brand reputation, product quality, advertisement quality, company market value, etc.)

[1]

Motivating Example: Competing Market Makers

- Competition among market makers in a dealer market (Cont, Guo and Xu 2021; Cont & Xiong 2021)
- Each market maker optimize tradeoff between expected profit and inventory risk by setting bid/ask prices in response to market order flow.
- Market makers affected by *aggregate* supply/demand/order flow $\Rightarrow$ "Mean-field" formulation
- Heterogeneity: Market makers may differ by size and risk aversion
- One may pool market makers into *near*-homogeneous clusters (large/small, high/low risk aversion) but heterogeneity still exists in each cluster
- Number of market makers *not extremely large*: $N \sim 10^2$. Fluctuation effects non-negligeable.

Homogenization of N-player games

We propose a systematic **homogenization** method to construct *controlled* mean-field approximations for dynamic N-player games with heterogeneity, with error estimates at each step.

Given a N-player game with heterogeneity we:

- Partition the population of players into K "near"-heterogeneous groups.
- Replace interactions between players by interactions between typical players in each group j with the **distributions** L^k of state/actions of each group $k = 1..K$
- Consider a multi-population mean-field game (MFG) whose Nash equilibria are ϵ-Nash equilibria of the N-player game
- Provide non-asymptotic estimates for ϵ in terms of population size and heterogeneity.

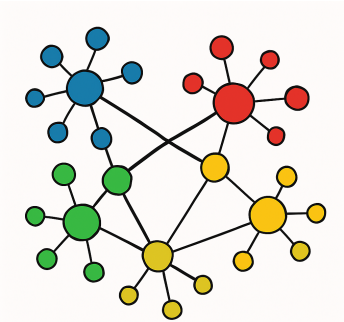
Accuracy of mean-field approximations

Two sources of approximation error:

- **Mean-field approximation error** due to the finite number of players in each group
- **Heterogeneity** : deviations from homogeneity in each group

Natural trade-off :

- Mean-field approximation error decreases with group size N_k
- Heterogeneity error increases with group size N_k, decreases with number of groups K



Problem Setup: (Heterogeneous) N-Player Games

Dynamic N-player game G

- N players, $i = 1, \dots, N$, finite state and action spaces ,
- Finite-time-horizon $\mathcal{T} = \{0, 1, \dots, T\}$
- Player i: state $s_t^i \in \mathcal{S}$, action $a_t^i \in \mathcal{A}$
- Initialization of states: $_0 \in^N$
- State $_t = (s_t^1, \dots, s_t^N) \in^N$, actions $_t = (a_t^1, \dots, a_t^N) \in^N$
- Reward (payoff) of player i in state $_t$: $R_t^i(_t, _t)$
- Transition probability for state of player i: $P_t^i(\cdot|_t, _t)$

Nash Equilibrium

- Player i's strategy: $\pi_t^i : \mathcal{S} \to \mathcal{P}(\mathcal{A})$, randomized, local
- Player i's expected total reward $V^i(\boldsymbol{\pi}; G) := \mathbb{E}_{\boldsymbol{\pi}} \left[\sum_{t \in \mathcal{T}} R_t^i(\mathbf{s}_t, \mathbf{a}_t) \right]$

Nash equilibrium (NE) of N-player game G

A strategy profile $\boldsymbol{\pi} = \{\pi_t^i\}_{t \in \mathcal{T}, i \in [N]}$ is a Nash equilibrium if each strategy is the best response to the other players strategies:

$$\forall i = 1..N, \quad V^i(\boldsymbol{\pi}; G) = \max_{\tilde{\pi}^i \in \Pi} V^i(\pi^1, \ldots, \tilde{\pi}^i, \ldots, \pi^N; G)$$

- Distance from Nash equilibrium:

$$\texttt{Expl}(\boldsymbol{\pi}; G) := \frac{1}{N} \sum_{i \in [N]} \max_{\tilde{\pi}^i \in \Pi} V^i(\pi^1, \ldots, \tilde{\pi}^i, \ldots, \pi^N; G) - V^i(\boldsymbol{\pi}; G).$$

- $\boldsymbol{\pi}$ Nash equilibrium $\iff$ $\texttt{Expl}(\boldsymbol{\pi}; G) = 0$.
- $\boldsymbol{\pi}$ is a ϵ-Nash equilibrium if $\texttt{Expl}(\boldsymbol{\pi}; G) \leq \epsilon$.

Homogenized Model: K-Population Mean-Field Game

K-population mean-field game $G^{\mathtt{MF}}$

- Split population into K (near-)homogeneous groups of players
- Represent each group $k = 1, \dots, K$ by the joint distribution $L_t^k \in \Delta(\mathcal{S} \times \mathcal{A})$ of state $s_t^k \in \mathcal{S}$ and action $a_t^k \in \mathcal{A}$ of its members
- Initialization of state of representative agent in group k: $s_0^k \sim \mu_0^k$.
- Reward/payoff of representative agent in group k:

$$\bar{R}_t^k(s_t^k, a_t^k, L_t^1, \dots, L_t^K)$$

- Dynamics: transition probabilities for state s_t^k of representative agent in group k

$$\bar{P}_t^k(s_{t+1}^k | s_t^k, a_t^k, L_t^1, \dots, L_t^K)$$

Homogenized Model: K-Population Mean-Field Game

Mean-field strategy and exploitability

- Mean-field strategy profile $\bar{\boldsymbol{\pi}} = \{\bar{\pi}_t^k\}_{t\in\mathcal{T},k\in[K]}$: $\bar{\pi}_t^k :\to \mathcal{P}()$ is the policy adopted by players in population k at time step t
- Under mean-field strategy profile $\bar{}$, joint state-action distributions of the populations $\{L_t^{k,\bar{\boldsymbol{\pi}}}\}_{t\in\mathcal{T},k\in[K]}$ are induced by the population mean-field strategy profile $\bar{\boldsymbol{\pi}}$.
- Any player in group k taking strategy $\tilde{\pi}^k$ has expected reward

$$\bar{V}^k(\tilde{\pi}^k, \bar{\boldsymbol{\pi}}; G^{\mathtt{MF}}) := \mathbb{E}_{\tilde{\pi}^k}[\sum_{t\in\mathcal{T}} \bar{R}_t^k(s_t^k, a_t^k, L_t^{1,\bar{\boldsymbol{\pi}}}, \ldots, L_t^{K,\bar{\boldsymbol{\pi}}})],$$

Exploitability

For any $k \in [K]$, the k-th population exploitability is defined as

$$\mathtt{Expl}^k(\bar{\boldsymbol{\pi}}; G^{\mathtt{MF}}) := \max_{\tilde{\pi}^k\in\Pi} \bar{V}^k(\tilde{\pi}^k, \bar{\boldsymbol{\pi}}; G^{\mathtt{MF}}) - \bar{V}^k(\bar{\pi}^k, \bar{\boldsymbol{\pi}}; G^{\mathtt{MF}}).$$

Nash equilibrium of $G^{\mathtt{MF}}$

A mean-field strategy profile $\bar{\boldsymbol{\pi}}$ is an NE of $G^{\mathtt{MF}}$ if the k-th population exploitability $\mathrm{Expl}^k(\bar{\boldsymbol{\pi}}; G^{\mathtt{MF}}) = 0$ for all $k \in [K]$

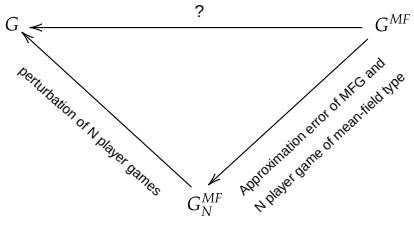
Weighted ϵ-NE of $G^{\mathtt{MF}}$

For any $w \in \mathbb{R}^k_{\geq 0}$, a $\bar{\boldsymbol{\pi}}$ with $\sum_{k\in[K]} w_k \mathtt{Expl}^k(\bar{\boldsymbol{\pi}}; G^{\mathtt{MF}}) \leq \epsilon$ is called a w-weighted ϵ-NE.

- When $K = 1$, a 1-weighted ϵ-NE is simply referred to as an ϵ-NE
- Existence of NE of $G^{\mathtt{MF}}$ is guaranteed by the continuity of rewards and transition probabilities as functions of mean-field(s) L

Decomposition of Homogenization Error

- How large is the gap between the N-player game G and the approximating MFG $G^{\mathtt{MF}}$?
- Bridge: N-player game of mean-field type $G_N^{\mathtt{MF}}$



- The two edges of the triangle correspond to the two types of error.

Bridge: N-Player Games of Mean-Field Type $G_N^{\mathtt{MF}}$

Mean-field games with subpopulations are naturally described as limits of "N-player games of mean-field type".
An N-player game is of mean-field type if

- There is a K-partition of the players into $K \ll N$ disjoint **homogeneous subpopulations**

$$\mathcal{I}_1, \dots, \mathcal{I}_K, \qquad [N] = \bigcup_{k \in [K]} \mathcal{I}_k, \quad N_k := |\mathcal{I}_k|.$$

- For $i \in \mathcal{I}_k$:
  - Rewards $\hat{R}_t^i(\_{t,t}) = \bar{R}_t^k(s_t^i, a_t^i, L_t^{1,N}, \dots, L_t^{K,N})$
  - Transitions $\hat{P}_t^i(s_{t+1}^i|\_{t,t}) = \bar{P}_t^k(s_{t+1}^i|s_t^i, a_t^i, L_t^{1,N}, \dots, L_t^{K,N})$

  for some mean-field type reward $\bar{R}_t^k$ and transition probability $\bar{P}_t^k$
- $L_t^{k,N} \in \Delta(\mathcal{S} \times \mathcal{A})$: state-action empirical distribution of group k at time step t, defined as $L_t^{k,N}(s,a) := \frac{1}{N_k} \sum_{i \in \mathcal{I}_k} \mathbf{1}\{s_t^i = s, a_t^i = a\}$

Bridge: N-Player Game of Mean-Field type $G_N^{\mathtt{MF}}$

Connection with K-population mean-field game

Consider the limiting K-population mean-field game $G^{\mathtt{MF}}$

- Representative player in group k has rewards $\bar{R}^k$ and transitions $\bar{P}^k$
- The initializations of players in group k: $\mu_0^{k,N}(s) := \frac{1}{N_k} \sum_{i \in \mathcal{I}_k} \mathbf{1}\{s_0^i = s\}$ the empirical distributions of states for each group in the N-player game
- If $\bar{\boldsymbol{\pi}}$ is a $(N_1/N, \ldots, N_K/N)$-weighted ϵ-NE of $G^{\mathtt{MF}}$, ,

$$\mathtt{Expl}(\bar{\boldsymbol{\pi}}; G^{\mathtt{MF}}) := \sum_{k \in [K]} \frac{N_k}{N} \mathtt{Expl}^k(\bar{\boldsymbol{\pi}}; G^{\mathtt{MF}}) \leq \epsilon$$

- Such a strategy profile can then be executed in the original N-player game by taking $\pi_t^i := \bar{\pi}_t^k$ for any player $i \in \mathcal{I}_k$, $k \in [K]$

Weighted Lipschitz Continuity

[Weighted Lipschitz continuity]

$$|\bar{R}_t^k(s_t, a_t, L_t^{1:K}) - \bar{R}_t^k(s_t, a_t, \tilde{L}_t^{1:K})| \leq \sum_{j=1}^K w_R^j \|L_t^j - \tilde{L}_t^j\|_1,$$

$$\|\bar{P}_t^k(\cdot|s_t, a_t, L_t^{1:K}) - \bar{P}_t^k(\cdot|s_t, a_t, \tilde{L}_t^{1:K})\|_1 \leq \sum_{j=1}^K w_P^j \|L_t^j - \tilde{L}_t^j\|_1.$$

Examples:

- ℓ_1-Lipschitz continuity in the concatenated vector of $L_t^1, \ldots, L_t^K$
  - Uniform weights: $w_R^j \equiv L_R$, $w_P^j \equiv L_P$, $j \in [K]$
- ℓ_1-Lipschitz continuity in the aggregated mean-field $\bar{L}_t := \sum_{k=1}^K \frac{N_k}{N} L_t^k$
  - Proportional weights: $w_R^j = L_R N_j/N$, $w_P^j = L_P N_j/N$, $j \in [K]$

$W_{\max}$ = maximum Lipschitz weight, $\bar{R}_{\max}$ = maximum reward.

Mean-field approximation error

Suppose $\bar{R}_t^k$ and $\bar{P}_t^k$ satisfy weighted Lipschitz continuity. Then

$$\begin{aligned}&\left|\texttt{Expl}(\boldsymbol{\pi}; G_N^{\texttt{MF}}) - \texttt{Expl}(\bar{\boldsymbol{\pi}}; G^{\texttt{MF}})\right| \\ &\quad \le C(T, W_{\max}, \bar{R}_{\max})\sqrt{SA}\left(\sum_{j\in[K]} \frac{w_P^j + w_R^j}{\sqrt{N_j}} + \sum_{j\in[K]} \frac{\sqrt{N_j}}{N}\right).\end{aligned}$$

- Uniform weights: $O\left(\sqrt{SA}\sum_{j\in[K]} 1/\sqrt{N_j}\right)$
- Proportional weights: $O(\sqrt{SA}\sum_{j\in[K]} \sqrt{N_j}/N)$

Population heterogeneity

N-player game vs N-player game with mean-field type

Theorem 2

For any strategy profile $\boldsymbol{\pi} = \{\pi_t^i\}_{t\in\mathcal{T}, i\in[N]}$, we have

$$\left|\texttt{Expl}(\boldsymbol{\pi}; G) - \texttt{Expl}(\boldsymbol{\pi}; G_N^{\texttt{MF}})\right| \leq 2TR_{\max}\sum_{t=1}^{T}\epsilon_t^P + 2\sum_{t=1}^{T}\epsilon_t^R,$$

where

$$\epsilon_t^R = \sqrt{\frac{1}{N}\sum_{i\in[N]}\max_{\in\mathcal{S}^N, \in\mathcal{A}^N}\left|R_t^i(\mathbf{s},\mathbf{a}) - \hat{R}_t^i(\mathbf{s},\mathbf{a})\right|^2},$$

$$\epsilon_t^P = \max_{\in\mathcal{S}^N, \in\mathcal{A}^N}\sum_{i\in[N]}\left\|P_t^i(\cdot|\mathbf{s},\mathbf{a}) - \hat{P}_t^i(\cdot|\mathbf{s},\mathbf{a})\right\|_1.$$

- ϵ^R measures heterogeneity in rewards/payoffs
- ϵ^P measures heterogeneity in transition probabilities

Theorem 3

Suppose $\bar{R}_t^k$ and $\bar{P}_t^k$ satisfy weighted Lipschitz continuity. Let $\bar{\boldsymbol{\pi}}$ be an $\left(\frac{N_1}{N}, \ldots, \frac{N_K}{N}\right)$-weighted ϵ-NE of $G^{\mathtt{MF}}$. Then $\boldsymbol{\pi} = \{\pi_t^i\}_{t\in\mathcal{T}, i\in[N]}$, $\pi_t^i = \bar{\pi}_t^k$, $i \in \mathcal{I}_k$, $k \in [K]$ is an $(\epsilon + \epsilon_{\mathtt{approx}} + \epsilon_{\mathtt{heter}})$-NE of G, where

$$\epsilon_{\mathtt{approx}} = C(T, W_{\max}, \bar{R}_{\max})\sqrt{SA}\left(\sum_{j\in[K]} \frac{w_P^j + w_R^j}{\sqrt{N_j}} + \sum_{j\in[K]} \frac{\sqrt{N_j}}{N}\right),$$

$$\epsilon_{\mathtt{heter}} = 2TR_{\max}\sum_{t=1}^{T}\epsilon_t^P + 2\sum_{t=1}^{T}\epsilon_t^R.$$

Effect of not accounting for population heterogeneity

Trade-off in players partitions

- Half of the agents have $R_t^1(s, a, \bar{L})$, while the other half of the agents have $R_t^2(s, a, \bar{L})$
- Consider two different cases
  - Case 1: fail to identify the two groups, approximate both R_t^1 and R_t^2 with a single R_t
  - Case 2: correctly identify the two groups of agents
- Then case 1 has larger $\epsilon_{\texttt{heter}}$ but smaller $\epsilon_{\texttt{approx}}$ than case 2
  - $\epsilon_{\texttt{heter}}$: $\sqrt{\frac{1}{2}(|R_t^1 - R_t|^2 + |R_t^2 - R_t|^2)} \Longrightarrow 0$
  - $\epsilon_{\texttt{approx}}$: $2C/\sqrt{N} \Longrightarrow 3\sqrt{2}C/\sqrt{N}$

Discussion: Players with Heterogeneous Rewards

Suppose that there exists a partition $\{\mathcal{I}_k\}_{k\in[K]}$ of N players and that for any $k \in [K]$, $\theta^i \sim \mathcal{D}^k$ i.i.d. for any $i \in \mathcal{I}_k$, with $\mathcal{D}^k$ some distribution over a compact subset $\mathbb{R}^d$.

- Parameterized N-player game with $R_t^i(,) = \bar{R}_t^k(s_t^i, a_t^i, L_t^1, \ldots, L_t^K, \theta^i)$ and $P_t^i(s_{t+1}^i|_{t,t}) = \bar{P}_t^k(s_{t+1}^i|s_i^i, a_t^i, L_t^1, \ldots, L_t^K)$

Discussion: Players with Heterogeneous Rewards

Suppose that there exists a partition $\{\mathcal{I}_k\}_{k\in[K]}$ of N players and that for any $k \in [K]$, $\theta^i \sim \mathcal{D}^k$ i.i.d. for any $i \in \mathcal{I}_k$, with $\mathcal{D}^k$ some distribution over a compact subset $\mathbb{R}^d$.

- Parameterized N-player game with $R_t^i(,) = \bar{R}_t^k(s_t^i, a_t^i, L_t^1, \ldots, L_t^K, \theta^i)$ and $P_t^i(s_{t+1}^i|_{t,t}) = \bar{P}_t^k(s_{t+1}^i|s_i^i, a_t^i, L_t^1, \ldots, L_t^K)$
- Sample mean approximation: approximate θ^i by $\hat{\theta}^k := \dfrac{1}{N_k}\sum_{i\in\mathcal{I}_k}\theta^i$,

Discussion: Players with Heterogeneous Rewards

Suppose that there exists a partition $\{\mathcal{I}_k\}_{k\in[K]}$ of N players and that for any $k \in [K]$, $\theta^i \sim \mathcal{D}^k$ i.i.d. for any $i \in \mathcal{I}_k$, with $\mathcal{D}^k$ some distribution over a compact subset $\mathbb{R}^d$.

- Parameterized N-player game with $R_t^i(,) = \bar{R}_t^k(s_t^i, a_t^i, L_t^1, \ldots, L_t^K, \theta^i)$ and $P_t^i(s_{t+1}^i|_{t,t}) = \bar{P}_t^k(s_{t+1}^i|s_i^i, a_t^i, L_t^1, \ldots, L_t^K)$
- Sample mean approximation: approximate θ^i by $\hat{\theta}^k := \dfrac{1}{N_k}\sum_{i\in\mathcal{I}_k}\theta^i$,
- Suppose that $\bar{R}_t^k$ is $w_\theta-$Lipschitz continuous in θ, then concentration inequality gives

$$\epsilon_{\texttt{heter}} \le 2T\sqrt{w_\theta^2\sum_{k\in[K]}\frac{N_k}{N}\sigma_k^2 + \frac{w_\theta^2\sum_{k\in[K]}\sqrt{N_k}}{N}\left(\sqrt{\frac{1}{C_1}\log(1/\delta)} + d\sqrt{\frac{1}{C_2}\log(d/\delta)}\right)}$$

with probability at least $1-2(1+d)K\delta$, where $\sigma_k^2 = \mathbb{E}\|\theta^i - \bar{\theta}_k\|_2^2 = \mathbb{E}\|\theta^i\|_2^2 - \|\bar{\theta}_k\|_2^2$, and $\bar{\theta}_k := \mathbb{E}_{\theta\sim\mathcal{D}^k}\theta$.

Discussion: Players with Heterogeneous Rewards

- This approximation works well when the variability in each group $\sigma_k^2 (k \in [K])$ is small.
- If for some group $k \in [K]$, the variability σ_k^2 is large, it is reasonable to further divide the group into n_k smaller groups to reduce the heterogeneity error.
- Finding the best partition for the agents that minimizes the heterogeneity can be reduced to finding the best clustering for θ^i that minimizes

$$\sum_{l=1}^{n_k} \sum_{i \in \mathcal{I}_k^l} \|\theta^i - \bar{\theta}_k^l\|_2^2,$$

  where $\bar{\theta}_k^l = \frac{1}{|\mathcal{I}_k^l|} \sum_{i \in \mathcal{I}_k^l} \theta^i$ is the mean (also called centroid) of θ^i for $i \in \mathcal{I}_k^l$.
- This coincides with the objective of k-means method.

Optimal partitioning of agents

Theorem (Optimal partition as mixed-integer optimization)

The optimal partitioning problem

$$\min_{\{\mathcal{I}_k, k\in[K]\} \text{ partition of } [N]} \epsilon_{MF}(\{\mathcal{I}_k\}_{k\in[K]}) + \epsilon_{heter}(\{\mathcal{I}_k\}_{k\in[K]}) \tag{1}$$

is equivalent to solving the following mixed-integer convex optimization (MIC) problem with variables $d_{ik}, w_{ik} \in \mathbb{R}$ *(*$i \in [N], k \in [K]$*),* $y_k \in \mathbb{R}, c_k \in \mathbb{R}^d$ *(*$k \in [K]$*) and* $x \in \mathbb{R}$*:*

$$\begin{aligned}
\min \quad & 2w_D T x + \max\{\bar{R}_{\max}, 1\} C\sqrt{SA}\left(\frac{w_P + w_R}{N} + \frac{1}{\sqrt{N}}\right) \sum_{k\in[K]} y_k \\
& x \in \mathbb{R}, \qquad d_{ik}, w_{ik} \in \mathbb{R}, \quad y_k \in \mathbb{R}, c_k \in \mathbb{R}^d, \quad i \in [N], k \in [K] \\
\text{subject to} \quad & d_{ik} \geq \|\theta^i - c_k\|_2 - 2D(1 - w_{ik}), \quad d_{ik} \geq 0, \quad i \in [N], k \in [K], \\
& w_{ik} \in \{0, 1\}, \quad \textstyle\sum_{k\in[K]} w_{ik} = 1, \quad i \in [N], \\
& \sqrt{\textstyle\sum_{i\in[N]} w_{ik}^2} \leq y_k, \quad \|c_k\|_2 \leq D, \quad k \in [K],
\end{aligned}$$

Optimal partitioning of agents

- More precisely, if the partition $\{\mathcal{I}_k\}_{k=1}^K$ is a minimizer of (1) then by setting

$$w_{ik} = \mathbf{1}_{\{i \in_k\}}, \quad y_k = |_k|, \quad c_k = 1_{y_k>0} \frac{\sum_{i \in_k} \theta^i}{|_k|},$$

$$d_{ik} = \|\theta^i - c_k\|_2 1_{i \in_k}, \quad x = \sqrt{\sum_{k\in[K]} \sum_{i\in[N]} d_{ik}^2}$$

  for $i \in [N], k \in [K]$ then $x, \{d_{ik}, w_{ik}\}_{i\in[N],k\in[K]}$, $\{y_k, c_k\}_{k\in[K]}$ solve (MIC).
- Conversely, if $x, \{d_{ik}, w_{ik}\}_{i\in[N],k\in[K]}$, $\{y_k, c_k\}_{k\in[K]}$ solve (MIC), then $\mathcal{I}_k = \{i \in [N] : w_{ik} = 1\}$ yields an optimal partition of $[N]$ which minimizes (MIC).
- The complexity of this mixed-integer problem is similar to a K-means clustering problem for which efficient algorithms exist.

Summary

- Quantitative **Homogenization estimates** for heterogenous N-player games.
- Given a dynamic N-player game with heterogeneity, we:
- Construct a multi-population mean-field game (MFG) whose Nash equilibria are ϵ-Nash equilibria of the N-player game
- Provide non-asymptotic estimates for ϵ in terms of population size and heterogeneity.
- Explicit nature of approximation error estimates allows to compare different partitioning schemes and assess the impact of misspecified homogeneity assumptions.
- Computation of optimal population partition is formulated as mixed-integer optimization, simlar to a K-means clustering problem.

Application to large-scale MultiAgent Reinforcement Learning (MARL)

Large-scale decentralized MARL faces two distinct scalability issues.

Observation scalability

Each agent observes a variable-size set of neighbors:

$$O_i(t) = \{o_t^{i,j} : j \in \mathcal{N}_i(t)\}.$$

A neural policy needs a fixed-dimensional input invariant to neighbor ordering and swarm size.

Population heterogeneity

Agents may differ in dynamics, costs, reward parameters, sensing, or objectives:

$$\theta_i = (A_i, B_i, Q_i, R_i, M_i, \ldots).$$

A single shared policy may be biased when the swarm is not homogeneous.

Clustered mean-embedding MARL

We combine two ingredients:

Mean embeddings for swarms

Hüttenrauch et al (2019):
encode a variable-size local neighborhood

$$\widehat{\mu}_{O_i} = \frac{1}{|O_i|} \sum_{o^{i,j} \in O_i} \phi_\eta(o^{i,j}).$$

This is permutation-invariant and independent of the number of observed neighbors.

Homogenization for heterogeneity

Partition agents into near-homogeneous groups:

$$[N] = I_1 \cup \cdots \cup I_K.$$

Then train one representative policy per group and control the error by

$$\epsilon_{\mathrm{RL}} + \epsilon_{\mathrm{MF}} + \epsilon_{\mathrm{heter}}.$$

Clustered mean-embedding MARL

For each cluster k, learn a separate decentralized policy architecture

$$\pi_\theta^k(a \mid x, \text{local mean embeddings}, \text{aggregate signals}).$$

For an agent $i \in I_k$,

$$a_t^i \sim \pi_\theta^k(\cdot \mid z_t^i).$$

The input z_t^i contains:

$$z_t^i = \left(x_t^i,\ k,\ \widehat{\mu}_{i,t}^{k\leftarrow 1}, \ldots, \widehat{\mu}_{i,t}^{k\leftarrow K},\ M_t^1, \ldots, M_t^K\right).$$

Here:

- $\widehat{\mu}_{i,t}^{k\leftarrow \ell}$ encodes neighbors of type ℓ seen by agent $i \in I_k$.
- M_t^ℓ is an optional population-level aggregate for cluster ℓ.

Typed Local Mean Embeddings

Define the neighbors of agent $i \in I_k$ belonging to group ℓ:

$$\mathcal{N}_i^\ell(t) = \mathcal{N}_i(t) \cap I_\ell.$$

For each agent cluster ℓ, compute

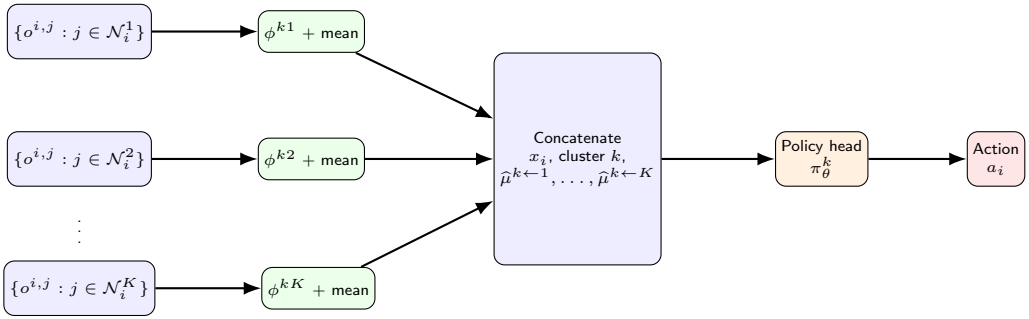
$$\boxed{\widehat{\mu}_{i,t}^{k\leftarrow\ell} = \frac{1}{|\mathcal{N}_i^\ell(t)|} \sum_{j \in \mathcal{N}_i^\ell(t)} \phi_\eta^{k\ell}\left(o_t^{i,j}\right)}$$

with a default zero vector if $|\mathcal{N}_i^\ell(t)| = 0$.

The policy for group k is

$$\boxed{a_t^i \sim \pi_\theta^k\left(\cdot \mid x_t^i, \widehat{\mu}_{i,t}^{k\leftarrow 1}, \ldots, \widehat{\mu}_{i,t}^{k\leftarrow K}\right), \qquad i \in I_k.}$$

Architecture



One policy architecture is learned for each recipient cluster k.

Why Type-Conditioned Embeddings?

A single untyped embedding

$$\frac{1}{|\mathcal{N}_i|}\sum_{j\in\mathcal{N}_i}\phi(o^{i,j})$$

mixes heterogeneous neighbors into one distribution.

Typed embeddings preserve the interaction structure:

$$\left(\widehat{\mu}^{k\leftarrow 1},\widehat{\mu}^{k\leftarrow 2},\ldots,\widehat{\mu}^{k\leftarrow K}\right).$$

Interpretation

For an agent in population k, the effect of population ℓ is summarized by a local empirical distribution of ℓ-type neighbors.

This is the local-observation analogue of a multi-population mean-field game.

Decentralized Execution

After training, each agent executes independently.

$$a_t^i = f_\theta^k\left(x_t^i, \widehat{\mu}_{i,t}^{k\leftarrow 1}, \ldots, \widehat{\mu}_{i,t}^{k\leftarrow K}\right), \qquad i \in I_k.$$

No agent needs the full joint state

$$(x_t^1, \ldots, x_t^N).$$

Each agent only needs:

- its own local state x_t^i,
- local neighbor observations $o_t^{i,j}$,
- neighbor cluster labels, or inferred types,
- optionally a low-dimensional broadcast mean field.

Centralized Training, Decentralized Execution

Training can be centralized:

$$\max_{\theta^1,\dots,\theta^K} \sum_{k=1}^{K} \frac{N_k}{N} J_k(\theta^1,\dots,\theta^K).$$

Each J_k is the return of a representative agent in population k.

In policy-gradient form:

$$\nabla_{\theta^k} J \approx \frac{1}{|\mathcal{B}_k|} \sum_{(i,t)\in\mathcal{B}_k} \nabla_{\theta^k} \log \pi^k_{\theta^k}(a^i_t \mid z^i_t) \widehat{A}^i_t.$$

Data pooling

All agents in cluster k contribute samples to the same policy update.

Algorithmic Template

1. Estimate or learn agent descriptors θ_i.
2. Partition the population:
$$[N] = I_1 \cup \cdots \cup I_K.$$
3. For each agent $i \in I_k$, compute typed local embeddings
$$\widehat{\mu}_i^{k \leftarrow \ell}, \qquad \ell = 1, \ldots, K.$$
4. Train K decentralized policies
$$\pi^1, \ldots, \pi^K$$
with any actor–critic, PPO, TRPO, or Q-learning method.
5. Deploy one policy per cluster.
6. Periodically re-cluster if agent parameters drift.

Homogenization Error Bound

If the learned K-population policy is an ϵ_{RL}-equilibrium of the approximate multi-population game, then the lifted decentralized policy satisfies

$$\boxed{\mathrm{NashConv}(\pi; G) \leq \epsilon_{\mathrm{RL}} + \epsilon_{\mathrm{MF}} + \epsilon_{\mathrm{heter}}.}$$

ϵ_{MF} finite-population / sampling error within clusters,

$\epsilon_{\mathrm{heter}}$ within-cluster heterogeneity error.

This separates learning error from modeling error.

Finite-Population Error

A typical bound has the form

$$\boxed{\epsilon_{\mathrm{MF}} = C\left[\sum_{j=1}^{K}(w_P^j + w_R^j)\left(\frac{1}{N_j} + \frac{1}{\sqrt{N_j}}\right) + \sum_{j=1}^{K}\frac{\sqrt{N_j}}{N}\right].}$$

Implications:

- Larger clusters reduce finite-population error.
- Splitting into too many clusters increases ϵ_{MF}.
- Small clusters are statistically fragile.

Heterogeneity Error

Let $\bar{\theta}_k$ be the representative parameter of cluster k. A typical bound is

$$\epsilon_{\text{heter}} \lesssim C_T \left[\frac{1}{N} \sum_{k=1}^{K} \sum_{i \in I_k} \|\theta_i - \bar{\theta}_k\|^2 \right]^{1/2}.$$

Implications:

- More homogeneous clusters reduce ϵ_{heter}.
- Increasing K usually lowers heterogeneity error.
- The partition should balance ϵ_{MF} and ϵ_{heter}.

Bias–Variance Trade-Off in K

Cluster selection solves

$$\min_{I_1,\dots,I_K} \left(\epsilon_{\mathrm{MF}}(I_1,\dots,I_K) + \epsilon_{\mathrm{heter}}(I_1,\dots,I_K)\right).$$

Choice of K	Benefit	Cost
Small K	large clusters, stable learning	high heterogeneity bias
Large K	homogeneous clusters	small-sample / mean-field error
Optimal K	balanced error	model selection required

For large N, the heterogeneity term dominates and clustering resembles K-means.

Example: MultiAgent RL for a heterogeneous swarm of LQ robots

MultiAgent RL for a heterogeneous swarm of LQ robots

Robot i has linear dynamics

$$x_{t+1}^i = A_i x_t^i + B_i u_t^i + \eta_t^i.$$

Swarm should stay not far from average position:

$$\bar{x}_t = \frac{1}{N} \sum_{j=1}^{N} x_t^j.$$

Quadratic objective:

$$J_i = \mathbb{E} \sum_{t=0}^{T} \left[(x_t^i)^\top Q_i x_t^i + (u_t^i)^\top R_i u_t^i + (x_t^i - \bar{x}_t)^\top M_i (x_t^i - \bar{x}_t) \right].$$

Agent parameters are allowed to be heterogeneous:

$$\theta_i = (A_i, B_i, Q_i, R_i, M_i).$$

Clustered Mean-Embedding Policy for LQ Robots

For $i \in I_k$,

$$u_t^i = \pi_\theta^k \left(x_t^i, \widehat{\mu}_{i,t}^{k \leftarrow 1}, \ldots, \widehat{\mu}_{i,t}^{k \leftarrow K} \right).$$

Parameterization of policies:

$$u_t^i = -K_t^k x_t^i + g_\theta^k \left(\widehat{\mu}_{i,t}^{k \leftarrow 1}, \ldots, \widehat{\mu}_{i,t}^{k \leftarrow K} \right).$$

where

- K_t^k captures local LQ feedback for type k.
- g_θ^k learns nonlinear corrections from local neighbor distributions.

Two-Type Robot Example

Suppose

$$\theta^A = (A_A, B, Q, R, M), \qquad \theta^B = (A_B, B, Q, R, M),$$

with fractions α and $1 - \alpha$.

Single population

One shared policy:

$$\epsilon^{(1)}_{\text{heter}} \lesssim C_T \sqrt{\alpha(1-\alpha)} \|A_A - A_B\|_F.$$

Two populations

Two policies, one per type:

$$\epsilon^{(2)}_{\text{heter}} = 0$$

if clusters match the true types.

When are multipopulation models better?

Two clusters improve the bound when

$$\boxed{\epsilon^{(2)}_{\mathrm{MF}} - \epsilon^{(1)}_{\mathrm{MF}} < \epsilon^{(1)}_{\mathrm{heter}}.}$$

For two LQ robot types, this is roughly

$$C\left(\frac{1}{\sqrt{\alpha N}} + \frac{1}{\sqrt{(1-\alpha)N}} - \frac{1}{\sqrt{N}}\right) < C_T\sqrt{\alpha(1-\alpha)}\|A_A - A_B\|_F.$$

Use clustering when type differences are large enough to justify smaller cluster sizes.

Computational Complexity

Let p be the embedding dimension and $\deg_i$ the local neighborhood size.

Method	Policies	Execution per agent
Fully independent MARL	N	policy-specific
Single mean-embedding swarm	1	$O(\deg_i p)$
Clustered mean-embedding MARL	K	$O(\deg_i p)$

Training parameter count scales as

$$O(KP)$$

instead of

$$O(NP),$$

where P is the number of parameters per policy.

Design Choices

Encoders

- Separate encoder $\phi^{k\ell}$ for each pair of populations.
- Shared encoder $\phi(o, \ell)$ with type embedding.
- Shared encoder plus cluster-specific policy heads.

Clustering

- Hard clustering: $i \in I_k$.
- Soft clustering: weights $w_{ik} \in [0, 1]$.
- Online clustering when agent parameters evolve.

Practical Training Objective

For cooperative tasks:

$$\max_{\theta} \ \mathbb{E}\left[\sum_{t=0}^{T} r_t^{\text{team}}\right].$$

For game-theoretic tasks:

$$\min_{\theta} \ \text{NashConv}(\pi_\theta; G) \quad \text{or} \quad \max_{\pi^k} J_k(\pi^k, \pi^{-k}).$$

Population-weighted objective:

$$J(\theta) = \sum_{k=1}^{K} \frac{N_k}{N} J_k(\theta^1, \ldots, \theta^K).$$

This matches the multi-population mean-field equilibrium objective.

Summary: scalable MARL for heterogeneous populations

Clustered mean-embedding MARL = homogenization + mean-embedding neural policies.

- Partition heterogeneous agents into K near-homogeneous populations.
- Train one decentralized mean-embedding policy per population.
- Encode local neighbors by typed empirical mean embeddings.
- Scalable decentralized execution.
- If the learned population policies are close to equilibrium, the lifted N-agent policy is close to equilibrium.
- The error separates learning, finite-population, and heterogeneity components.
- Non-asymptotic estimates+ mixed integer optimization give a principled criterion for choosing K.