Task 2.1 (2019-20)

IoBT REIGN

Reinforcement Learning in Linear-Quadratic Mean-Field Games

Muhammad Aneeq Uz Zaman, Kaiqing Zhang, Erik Miehling, Tamer Başar
University of Illinois at Urbana-Champaign

Motivation

Scaling game-theoretic models to multiple agents quickly runs into scalability issues.

The main idea behind mean-field games is that a game with N agents can be reasonably approximated by an infinite population game:

- Impose that every agent in the game play the same strategy as computed in the infinite-population game
- Obtain bounds on the quality of the computed strategies in the original finite-population game

Goal: We wish to learn the mean-field equilibrium (the equilibrium arising from assuming an infinite population) and evaluate how well the associated strategies perform in the finite population game as a function of both N and the given stopping tolerance (error) associated with the learning algorithm.

Model

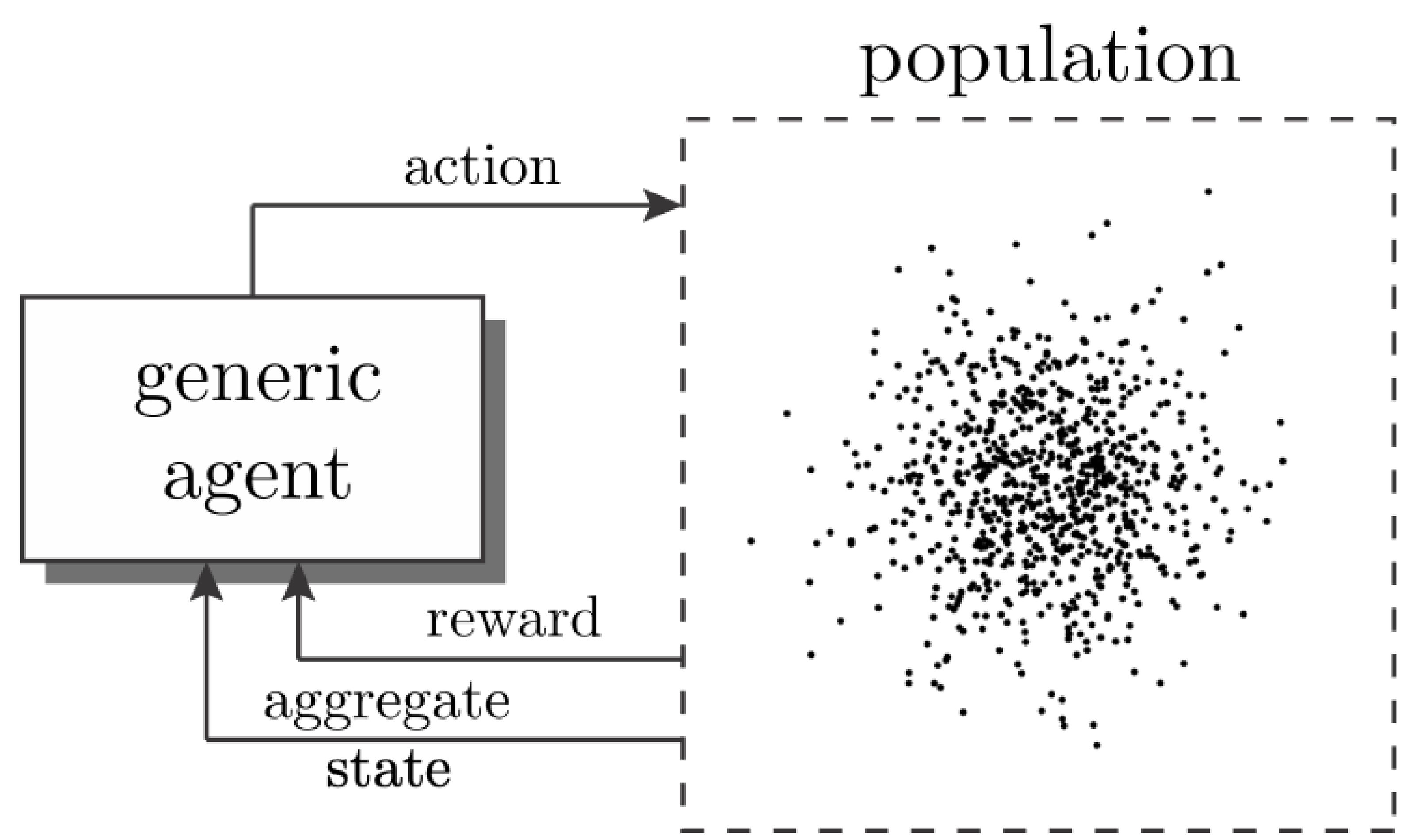
Agents follow uncoupled linear dynamics:

$$Z^n_{t+1} = AZ^n_t + BU^n_t + W^n_t$$

Each agent's cost is quadratic and coupled to other agents' costs by a state average term:

$$J^n_{\nu_0}(\eta) = \lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T} \mathbb{E}_{\eta,\nu_0}\left[\left\| Z^n_t - \frac{1}{N-1}\sum_{n'\neq n} Z^{n'}_t \right\|^2_{C_Z} + \|U^n_t\|^2_{C_U}\right]$$

The game is solved by taking the infinite population limit and considering the interaction between a **representative (generic) agent** and the **mean-field**.



Contribution

We prove that under the LQ structure, the mean-field equilibrium trajectory obeys linear dynamics, allowing us to cast the equilibrium learning problem as learning the **parameterized** dynamics of the mean-field state trajectory (a fixed matrix). This is done via actor-critic.

Derived bounds quantifying the quality of the resulting policy that take into account the following errors:

i) Error due to the infinite-population approximation of the finite-population game

ii) Error introduced due to the stopping condition of the learning (actor-critic) algorithm

Main Result

If the number of iterations of the actor-critic algorithm are on the order of $\Omega(\log(1/\varepsilon))$ and the number of agents are on the order of $\Omega(1/\varepsilon^2)$, then the cost degradation is bounded by ε with probability of at least $1-\varepsilon^5$.

Main Takeaway

Learning in games is challenging due to poor scalability in the number of agents. Instead of trying to scale up, we **scale-down** from the infinite (mean-field) population setting — much more tractable without sacrificing theoretical guarantees.

Future Directions

1) **Mean-field games on networks (current work!)** Oftentimes in battlefield scenarios, there are multiple subpopulations of agents with strong coupling within a subpopulation and weak coupling between subpopulations (e.g., multiple troops on a single mission, or multiple concurrent missions) — we are currently extending above results to the networked setting.

2) **Mean-field games with major and minor agents:** Agents are typically assumed to all have the same level of influence over the population. Often, especially in the case of adversarial interactions, one agent may have much more pronounced influence over the evolution of the population (e.g., drone swarms: evading a projectile, or tracking a target). We plan to investigate RL algorithms in this heterogenous (major-minor) agent environment.

Publications

Zaman, M. A.U., Zhang, K., Miehling, E., & Başar, T. (2020, July). **Approximate equilibrium computation for discrete-time linear-quadratic mean-field games.** In 2020 American Control Conference (ACC) (pp. 333-339).

Zaman, M. A.U., Zhang, K., Miehling, E., & Başar, T. (2020, December). **Reinforcement learning in non-stationary discrete-time linear-quadratic mean-field games**. In 2020 59th IEEE Conference on Decision and Control (CDC) (pp. 2278-2284).