Foundations of Multi-Agent Learning in Dynamic Environments: Where Reinforcement Learning Meets Strategic Decision-Making

Kaiqing Zhang

University of Maryland, College Park
kaiqing@umd.edu

Background. Recent years have witnessed tremendous successes of learning for *sequential* decision-making, and in particular, *Reinforcement Learning (RL)*. Prominent application examples include playing Go and video games, robotics, autonomous driving, and recently large language models. Most such success stories naturally involve ***multiple agents***. Hence, there has been surging research interest in advancing *Multi-Agent Learning in Dynamic Environments*, and particularly, ***multi-agent RL (MARL)***, to which my research has led and made significant contributions.

Sample Complexity of MARL. My work (Zhang et al. 2023) established the ***first near-optimal sample complexity*** for two-player zero-sum Stochastic Games, the most fundamental model of MARL (Shapley 1953; Littman 1994). We proved that a simple **model-based RL** approach matches the *lower-bound* **in all parameters** except a $|\mathcal{A}||\mathcal{B}|$ v.s. $|\mathcal{A}|+|\mathcal{B}|$ gap, where $\mathcal{A}, \mathcal{B}$ correspond to the *action spaces* of agents 1 and 2. More intriguingly, (Zhang et al. 2023) also showed that when the reward function is only given *after* the transition model is estimated, which we termed as the *reward-agnostic* setting, the lower bound is $\Omega(|\mathcal{A}||\mathcal{B}|)$ and exactly matches the upper bound. Note that model-based RL can *inherently* handle this more challenging reward-agnostic setting, which thus implies that it **is minimax optimal** in this regime. **This separation demonstrated the unique *power*** (it can inherently handle *multiple* rewards in hindsight) and ***limitation*** (it is *less adaptive* and has a $|\mathcal{A}| + |\mathcal{B}|$ gap even with reward information) of **model-based MARL.** My work thus raised a more fundamental question: **when will the sample complexities *not grow exponentially* in the number of agents?** Many works have since then been inspired to show that *with the guidance of reward*, some *model-free* MARL approaches can address this issue (Jin et al. 2023; Song, Mei, and Bai 2022; Li et al. 2022), including my works (Cui, Zhang, and Du 2023; Daskalakis*, Golowich*, and Zhang* 2023), on the sample complexity of MARL.

Computational Complexity of MARL. My work (Daskalakis*, Golowich*, and Zhang* 2023) also **settled that, computing *stationary Markov coarse correlated equilibrium (CCE)*** in **general-sum** Stochastic Games is **`PPAD-hard`** and thus computationally **intractable**. **The result is significant as it stands in sharp contrast with** that for **matrix/static games** *without* Markovian states, where computing CCE is known to be *tractable*; it also contrasts **single-agent RL**, where finding a *stationary Markov* optimal policy is tractable. **Hence, it illustrated another *unique* challenge in MARL**: it is the *combined effect* of both ***multi-agency*** and ***sequential*** decision-making that made MARL (computationally) hard. This is the **first computational hardness** result of finding CCE in general-sum Stochastic Games and MARL, with many followup works.

Independent Learning: An Econ Perspective of MARL. My work also reflected on the intriguing question: **Why always using *equilibrium* as the solution concept in MARL?** In fact, in Economics, *(Nash) equilibrium* is *not* necessarily the *target*, but the *natural long-run outcome* from multiple **strategic agents'** interactions via ***independent learning*** (Fudenberg and Levine 1998). Such a perspective has been extensively studied for stateless games, but remains elusive for *Stochastic Games* with *state dynamics*. My works (Sayin* et al. 2021; Sayin, Zhang, and Ozdaglar 2022; Park*, Zhang*, and Ozdaglar 2023) have led the exploration of this unique perspective of MARL, and have constituted an **invited article at International International Congress of Mathematicians (ICM) 2022** (Ozdaglar*, Sayin*, and Zhang* 2022).

Other Notable Contributions. My works have also led the exploration of distributed *networked* MARL (Zhang et al. 2018), and a *minimax* (*zero-sum-game*) view of *robust adversarial* and risk-sensitive RL (Zhang, Hu, and Başar 2020, 2021), safe RL (Ding et al. 2020), and offline RL (Ozdaglar et al. 2023).

Applications (in Robotics). Trained as a theorist, I have also actively engaged with applied researchers to broaden my research impact. For example, (Suh et al. 2022) on *Differentiable Simulators* in Robotics has won **ICML Outstanding Paper Award** in 2022. My work (Wang et al. 2024) has deployed the networked MARL framework to *Robot Fleet Learning* for robotic manipulation tasks, with application to the real-world data from *Amazon warehouses*.

Acknowledgments. K.Z. acknowledges the support from the U.S. Army Research Laboratory (ARL) Grant W911NF-

24-1-0085 and NSF CAREER Award #2443704.

References

Cui, Q.; Zhang, K.; and Du, S. 2023. Breaking the curse of multiagents in a large state space: Rl in Markov games with independent linear function approximation. In *COLT*.

Daskalakis*, C.; Golowich*, N.; and Zhang*, K. 2023. The complexity of Markov equilibrium in stochastic games. In *COLT*.

Ding, D.; Zhang, K.; Başar, T.; and Jovanovic, M. 2020. Natural policy gradient primal-dual method for constrained Markov decision processes. In *NeurIPS*.

Fudenberg, D.; and Levine, D. K. 1998. *The Theory of Learning in Games*. MIT press.

Jin, C.; Liu, Q.; Wang, Y.; and Yu, T. 2023. V-learning – A simple, efficient, decentralized algorithm for multiagent reinforcement learning. *Mathematics of Operations Research*.

Li, G.; Chi, Y.; Wei, Y.; and Chen, Y. 2022. Minimax-optimal multi-agent RL in Markov games with a generative model. In *NeurIPS*.

Littman, M. L. 1994. Markov games as a framework for multi-agent reinforcement learning. In *Machine Learning Proceedings*. Elsevier.

Ozdaglar, A.; Pattathil, S.; Zhang, J.; and Zhang, K. 2023. Revisiting the linear-programming framework for offline RL with general function approximation. In *ICML*.

Ozdaglar*, A.; Sayin*, M. O.; and Zhang*, K. 2022. Independent learning in stochastic games. *International Congress of Mathematicians (ICM)*.

Park*, C.; Zhang*, K.; and Ozdaglar, A. 2023. Multi-player zero-sum Markov games with networked separable interactions. In *NeurIPS*.

Sayin*, M. O.; Zhang*, K.; Leslie, D.; Başar, T.; and Ozdaglar, A. 2021. Decentralized Q-learning in zero-sum Markov games. In *NeurIPS*.

Sayin, M. O.; Zhang, K.; and Ozdaglar, A. 2022. Fictitious Play in Markov Games with Single Controller. In *ACM Conference on Economics and Computation*.

Shapley, L. S. 1953. Stochastic games. *PNAS*.

Song, Z.; Mei, S.; and Bai, Y. 2022. When can we learn general-sum Markov games with a large number of players sample-efficiently? In *ICLR*.

Suh, H. J.; Simchowitz, M.; Zhang, K.; and Tedrake, R. 2022. Do differentiable simulators give better policy gradients? In *ICML*.

Wang, L.; Zhang, K.; Zhou, A.; Simchowitz, M.; and Tedrake, R. 2024. Robot fleet learning via policy merging. In *ICLR*.

Zhang, K.; Hu, B.; and Başar, T. 2020. On the stability and convergence of robust adversarial reinforcement learning: A case study on linear quadratic systems. In *NeurIPS*.

Zhang, K.; Hu, B.; and Başar, T. 2021. Policy Optimization for $\mathcal{H}_2$ Linear Control with $\mathcal{H}_\infty$ Robustness Guarantee: Implicit Regularization and Global Convergence. *SIAM Journal on Control and Optimization*, 59(6): 4081–4109.

Zhang, K.; Kakade, S. M.; Başar, T.; and Yang, L. F. 2023. Model-based multi-agent RL in zero-sum Markov games with near-optimal sample complexity. *Journal of Machine Learning Research (short version Accepted at NeurIPS 2020, Spotlight)*, 24(175): 1–53.

Zhang, K.; Yang, Z.; Liu, H.; Zhang, T.; and Başar, T. 2018. Fully Decentralized Multi-Agent Reinforcement Learning with Networked Agents. In *ICML*.