Refined Sample Complexity for Markov Games with Independent Linear Function Approximation

Yan Dai YAN-DAI20@MAILS.TSINGHUA.EDU.CN
Institute for Interdisciplinary Information Sciences, Tsinghua University

Qiwen Cui QWCUI@CS.WASHINGTON.EDU
Simon S. Du SSDU@CS.WASHINGTON.EDU
Paul G. Allen School of Computer Science & Engineering, University of Washington

Editors: Shipra Agrawal and Aaron Roth

Abstract

Markov Games (MG) is an important model for Multi-Agent Reinforcement Learning, and various algorithms have been proposed to tackle different types of MGs. In this paper, we focus on the special type of MGs called *independent linear MGs* where the number of states and agents are both large, and independent linear function approximations are employed for each agent. This function approximation scheme is recently proposed by Cui et al. (2023) and Wang et al. (2023), which is shown to be capable of avoiding the "curse of multi-agents" (*i.e.*, the number of samples needed for finding an equilibrium depends polynomially on the number of agents – instead of exponentially).

However, while the recent algorithms proposed by Cui et al. (2023) and Wang et al. (2023) successfully avoided the curse of multi-agents in independent linear MGs, they either *i)* had a sub-optimal convergence rate of $\mathcal{O}(T^{-1/4})$ (Cui et al., 2023), or *ii)* had a polynomial dependency on the number of actions $A_{\max}$ (Wang et al., 2023) (which is avoidable in single-agent cases).

In this paper, we give a single algorithm for finding Markov Coarse Correlated Equilibria (CCE) for independent linear MGs. It simultaneously *i)* resolves the "curse of multi-agents", *ii)* attains the optimal $\mathcal{O}(T^{-1/2})$ convergence rate, and *iii)* avoids $\mathrm{poly}(A_{\max})$ dependencies.

Our approach exploits the following two technical innovations:

1. When refining the `AVLPR` framework by Wang et al. (2023), we propose that designing *data-dependent (i.e., stochastic) pessimistic estimations* of the sub-optimality gap can allow a broader choice of plug-in algorithms. Specifically, instead of the original requirement that the gap estimator must be deterministic, we allow it to be stochastic but with a bounded expectation. This avoids the $\mathrm{poly}(A_{\max})$ factors.
2. To ensure $\mathcal{O}(T^{-1/2})$ convergence rate, we not only borrowed state-of-the-art techniques from the single-agent RL literature (Zimmert and Lattimore, 2022; Dai et al., 2023; Liu et al., 2023), but also proposes a novel technique called *action-dependent bonuses*. It can be used to cancel estimation errors that occasionally has very extreme magnitudes (so classical techniques fail), but only those "rarely-visited" actions are related to enormous errors. We expect this technique to be of independent interest.

Our final algorithm attained the three aforementioned properties at the same time: It finds an ϵ-CCE in independent linear MGs with only $\widetilde{\mathcal{O}}(m^4d^5H^6\epsilon^{-2})$ samples, which polynomially depends on m, attains the optimal $\mathcal{O}(T^{-1/2})$ convergence rate, and avoids $\mathrm{poly}(A_{\max})$ factors.

Concurrent to this work, Fan et al. (2024) considered the same problem of finding CCEs in independent linear MGs but under the much stronger assumption of *local access models*. Under this stronger assumption, they proposed an algorithm with $\widetilde{\mathcal{O}}(m^2d^2H^6\epsilon^{-2})$ sample complexity, also enjoying the three nice properties. A detailed comparison can be found in our arXiv version.[1]

Keywords: Game Theory, Reinforcement Learning Theory, Multi-Agent Reinforcement Learning

1. Extended abstract. Full version available at `https://arxiv.org/abs/2402.07082v2`.

References

Qiwen Cui, Kaiqing Zhang, and Simon Du. Breaking the curse of multiagents in a large state space: Rl in markov games with independent linear function approximation. In *Proceedings of Thirty Sixth Conference on Learning Theory*, volume 195, pages 2651–2652. PMLR, 2023.

Yan Dai, Haipeng Luo, Chen-Yu Wei, and Julian Zimmert. Refined regret for adversarial mdps with linear function approximation. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202, pages 6726–6759. PMLR, 2023.

Junyi Fan, Yuxuan Han, Jialin Zeng, Jian-Feng Cai, Yang Wang, Yang Xiang, and Jiheng Zhang. Rl in markov games with independent function approximation: Improved sample complexity bound under the local access model. In *International Conference on Artificial Intelligence and Statistics*, pages 2035–2043. PMLR, 2024.

Haolin Liu, Chen-Yu Wei, and Julian Zimmert. Bypassing the simulator: Near-optimal adversarial linear contextual bandits. *arXiv preprint arXiv:2309.00814*, 2023.

Yuanhao Wang, Qinghua Liu, Yu Bai, and Chi Jin. Breaking the curse of multiagency: Provably efficient decentralized multi-agent rl with function approximation. In *Proceedings of Thirty Sixth Conference on Learning Theory*, volume 195, pages 2793–2848. PMLR, 2023.

Julian Zimmert and Tor Lattimore. Return of the bias: Almost minimax optimal high probability bounds for adversarial linear bandits. In *Conference on Learning Theory*, pages 3285–3312. PMLR, 2022.