Towards Robust Multi-Agent Reinforcement Learning

Aritra Mitra

North Carolina State University
amitra2@ncsu.edu

Abstract

Stochastic gradient descent (SGD) is at the heart of large-scale distributed machine learning paradigms such as federated learning (FL). In these applications, the task of training high-dimensional weight vectors is distributed among several workers that exchange information over networks of limited bandwidth. While parallelization at such an immense scale helps to reduce the computational burden, it creates several other challenges: delays, asynchrony, and most importantly, a significant communication bottleneck. The popularity and success of SGD can be attributed in no small part to the fact that it is extremely *robust* to such deviations from ideal operating conditions. Inspired by these findings, we ask: *Are common reinforcement learning (RL) algorithms also robust to similar structured perturbations?* Perhaps surprisingly, despite the recent surge of interest in multi-agent/federated RL, almost nothing is known about the above question.

To begin to fill this void, we study one of the simplest tasks in reinforcement learning, that of policy evaluation using the classical temporal difference (TD) learning algorithm with linear function approximation. In particular, we study variants of TD learning with perturbed update directions, where the perturbation is caused due to (i) a general compression operator; and (ii) arbitrary (but bounded) time-varying delays. For these settings, our main results can be summarized as follows.

- **Result 1.** We prove that compressed TD algorithms, coupled with an error-feedback mechanism used widely in optimization, exhibit the same non-asymptotic theoretical guarantees as their SGD counterparts (Mitra, Pappas, and Hassani 2023).
- **Result 2.** We prove that for multi-agent TD learning, one can achieve linear convergence speedups with respect to the number of agents while communicating just $\tilde{O}(1)$ bits per iteration (Mitra, Pappas, and Hassani 2023).
- **Result 3.** In (Dal Fabbro, Mitra, and Pappas 2023), we further extend our above analyses to account for the presence of lossy, packet-dropping channels in the context of federated TD learning.
- **Result 4.** Finally, we provide a comprehensive analysis of the effect of delays on the finite-time performance of TD learning algorithms, and propose delay-adaptive variants that provably improve performance relative to the vanilla delayed algorithms (Adibi et al. 2024).

Arriving at the above results is non-trivial, since unlike the key data-independence assumption prevalent in supervised learning, the data in RL exhibits time correlations. Nonetheless, we show that our techniques are not just limited to TD learning, but rather extend seamlessly to a much broader class of stochastic approximation algorithms driven by Markovian noise (including variants of Q-learning). The overarching message conveyed by our work is the following: *iterative RL algorithms can be just as robust to structured perturbations as their optimization counterparts.*

References

Adibi, A.; Dal Fabbro, N.; Schenato, L.; Kulkarni, S.; Poor, H. V.; Pappas, G. J.; Hassani, H.; and Mitra, A. 2024. Stochastic Approximation with Delayed Updates: Finite-Time Rates under Markovian Sampling. In *International Conference on Artificial Intelligence and Statistics*. PMLR.

Dal Fabbro, N.; Mitra, A.; and Pappas, G. J. 2023. Federated TD Learning over Finite-Rate Erasure Channels: Linear Speedup under Markovian Sampling. *IEEE Control Systems Letters*.

Mitra, A.; Pappas, G. J.; and Hassani, H. 2023. Temporal Difference Learning with Compressed Updates: Error-Feedback meets Reinforcement Learning. *arXiv preprint arXiv:2301.00944*.