Learning & Adaptive Systems

ETH zürich

Non-Convex Reinforcement Learning via Submodular Optimization

Project Proposal for Master Thesis

Motivation

Classic Reinforcement Learning (RL) algorithms assume the existence of a reward function that assigns a scalar reward to every state(-action) of the underlying Markov Decision Process [Puterman, 2014]. But it turns out that a wide variety of real-world problems don't admit a reward function of this type, including exploration [Hazan et al., 2019], experimental design [Mutny et al., 2023], imitation learning, diverse skill discovery and many others [Zahavy et al., 2021]. A way to face this problem has been developed within the area of Convex RL which leverages tools from convex optimization. Nonetheless, this formulation suffers some inherent limitations. Recently, another perspective on the problem has been developed using notions and algorithms from the area of submodular function maximization, giving rise to Submodular Reinforcement Learning [Prajapat et al., 2023].

Scope of the Project

While it has been shown that several objectives including maximum entropy exploration and D-optimal experimental design, can be captured via Submodular RL objectives, within this project we aim to show that Submodular RL is able to capture a much richer class of problems, including imitation learning among many others. Moreover, we aim to better understand the complementarity between Submodular RL and Convex RL. The project type would span from building theoretical results, algorithms, and practical implementation of the algorithm.

Ideal Candidate

An ideal candidate would have completed the *Foundations of RL* course, the *Probabilistic Artificial Intelligence* course, or have an equivalent understanding of RL. Background in statistics, optimization, and/or algorithmics is very appreciated, as well as strong motivation for both developing theory and playing with experimental settings.

Contact

If you are interested, please contact Riccardo De Santi (rdesanti@ethz.ch) or Manish Prajapat (manishp@ai.ethz.ch).

References

[Hazan et al., 2019] Hazan, E., Kakade, S., Singh, K., and Van Soest, A. (2019). Provably efficient maximum entropy exploration. In *International Conference on Machine Learning*.

[Mutny et al., 2023] Mutny, M., Janik, T., and Krause, A. (2023). Active exploration via experiment design in markov chains. In *International Conference on Artificial Intelligence and Statistics*, pages 7349–7374. PMLR.

[Prajapat et al., 2023] Prajapat, M., Mutnỳ, M., Zeilinger, M. N., and Krause, A. (2023). Submodular reinforcement learning. *arXiv preprint arXiv:2307.13372*.

[Puterman, 2014] Puterman, M. L. (2014). *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons.

[Zahavy et al., 2021] Zahavy, T., O'Donoghue, B., Desjardins, G., and Singh, S. (2021). Reward is enough for convex mdps. In *35th Conference on Neural Information Processing Systems (NeurIPS 2021)*.