Contribution ID: **27** Type: **not specified**

Convergence for Natural Policy Gradient on Infinite-State Average-Reward Markov Decision Processes

Monday, June 17, 2024 2:00 PM (30 minutes)

Infinite-state Markov Decision Processes (MDPs) are essential in modeling and optimizing a wide variety of engineering problems. In the reinforcement learning (RL) context, a variety of algorithms have been developed to learn and optimize these MDPs. At the heart of many popular policy-gradient based learning algorithms, such as natural actor-critic, TRPO, and PPO, lies the Natural Policy Gradient (NPG) algorithm. Convergence results for these RL algorithms rest on convergence results for the NPG algorithm. However, all existing results on the convergence of the NPG algorithm are limited to finite-state settings.

We prove the first convergence rate bound for the NPG algorithm for infinite-state average-reward MDPs, proving a $O(1/\sqrt{T})$ convergence rate, if the NPG algorithm is initialized with a good initial policy. Moreover, we show that in the context of a large class of queueing MDPs, the MaxWeight policy suffices to satisfy our initial-policy requirement and achieve a $O(1/\sqrt{T})$ convergence rate. Key to our result are state-dependent bounds on the relative value function achieved by the iterate policies of the NPG algorithm.

Primary author: GROSOF, Isaac (University of Illinois, Urbana-Champaign; Northwestern University)

Co-authors: Prof. SRIKANT, R (University of Illinois, Urbana-Champaign); Prof. MAGULURI, Siva Theja (Georgia Institute of Technology)

Presenter: GROSOF, Isaac (University of Illinois, Urbana-Champaign; Northwestern University)

Session Classification: Parallel session: Policy gradient methods: optimization and convergence