ID de Contribution: **179** Type: **Invited talk**

Strongly-polynomial time and validation analysis of policy gradient methods

vendredi 1 août 2025 15:00 (30 minutes)

We propose a novel termination criterion, termed the advantage gap function, for finite state and action Markov decision processes (MDP) and reinforcement learning (RL). By incorporating this advantage gap function into the design of step size rules and deriving a new linear rate of convergence that is independent of the stationary state distribution of the optimal policy, we demonstrate that policy gradient methods can solve MDPs in strongly-polynomial time. To the best of our knowledge, this is the first time that such strong convergence properties have been established for policy gradient methods. Moreover, in the stochastic setting, where only stochastic estimates of policy gradients are available, we show that the advantage gap function provides close approximations of the optimality gap for each individual state and exhibits a sublinear rate of convergence at every state. The advantage gap function can be easily estimated in the stochastic case, and when coupled with easily computable upper bounds on policy values, they provide a convenient way to validate the solutions generated by policy gradient methods. Therefore, our developments offer a principled and computable measure of optimality for RL, whereas current practice tends to rely on algorithm-to-algorithm or baselines comparisons with no certificate of optimality.

Author: JU, Caleb (Georgia Tech)

Co-auteur: Dr LAN, Guanghui (Georgia Tech)

Orateur: JU, Caleb (Georgia Tech)

Classification de Session: ML

Classification de thématique: Machine learning