ID de Contribution: **267** Type: **Non spécifié**

A Two-Timescale Primal-Dual Framework for Reinforcement Learning via Online Dual Variable Guidance

mercredi 30 juillet 2025 12:15 (30 minutes)

We study reinforcement learning by combining recent advances in regularized linear programming formulations with the classical theory of stochastic approximation.
Motivated by the challenge of designing algorithms that leverage off-policy data while maintaining on-policy exploration, we propose PGDA-RL, a novel primal-dual Projected Gradient Descent-Ascent algorithm for solving regularized Markov Decision Processes (MDPs). PGDA-RL integrates experience replay-based gradient estimation with a two-timescale decomposition of the underlying nested optimization problem.
The algorithm operates asynchronously, interacts with the environment through a single trajectory of correlated data, and updates its policy online in response to the dual variable associated with the occupation measure of the underlying MDP. We prove that PGDA-RL converges almost surely to the optimal value function and policy of the regularized MDP. Our convergence analysis relies on tools from stochastic approximation theory and holds under weaker assumptions than those required by existing primal-dual RL approaches, notably removing the need for a simulator or a fixed behavioral policy.

Authors: WOLTER, Axel Friedrich (University of Konstanz); Prof. SUTTER, Tobias (University of Konstanz)

Orateur: WOLTER, Axel Friedrich (University of Konstanz)

Classification de Session: Sequential decision-making under uncertainty

Classification de thématique: Sequential decision making under uncertainty