Regularized Anderson Acceleration for Off-Policy Deep Reinforcement Learning

Wenjie Shi, Shiji Song, Hui Wu, Ya-Chu Hsu, Cheng Wu, Gao Huang
Department of Automation, Tsinghua University, Beijing 100084, China

Tsinghua University

MOTIVATION

- Sample inefficiency remains a major limitation of current RL algorithms for problems with continuous and high dimensional state spaces.
- Sample inefficiency makes learning in real physical systems impractical and severely prohibits the applicability of RL approaches in more challenging scenarios.

Existing methods

- To learn models of the underlying system dynamics.
- Off-policy training scheme aims to reuse past experience such as DQN, DDPG.

Our method

- RL is closely linked to fixed-point iteration: the optimal policy can be found by solving a fixed-point problem of Bellman operator.
- Anderson acceleration is a method capable of speeding up the computation of fixed point iterations.

Thirty-third Conference on Neural Information Processing Systems (NeurIPS)