# Notes on Proximal Policy Optimization Algorithms

D. Gueorguiev 12/13/23

## A bit of theory on Policy Gradient Reinforcement Learning Methods

Assumptions:

The environment can be represented by a finite MDP

This is equivalent of saying that its state , action and reward sets are finite, and its dynamics is given by a set of probabilities , for all and ( is plus a terminal state if the problem is episodic).

We would like to compute value functions to organize and search for good policies.

The optimal value

# Bibliography

Gueorguiev, D. (2023, 11 26). *Note on Q functions and V functions in Reinforcement Learning.* Retrieved from Reinforcement Learning and Game Theory Repository: https://github.com/dimitarpg13/reinforcement\_learning\_and\_game\_theory/blob/main/docs/Note\_on\_Q\_functions\_in\_Reinforcement\_Learning.pdf

Heeswijk, W. v. (2022, Apr 9). *Policy Gradients In Reinforcement Learning Explained.* Retrieved from Towards Data Science: https://towardsdatascience.com/policy-gradients-in-reinforcement-learning-explained-ecec7df94245

Heeswijk, W. v. (2022, 11 29). *Proximal Policy Optimization (PPO) Explained.* Retrieved from Towards Data Science: https://towardsdatascience.com/proximal-policy-optimization-ppo-explained-abed1952457b

John Schulman, e. a. (2018, Oct 20). *High-Dimensional Continuous Control Using Generalized Advantage Estimation.* Retrieved from arxiv.org: https://github.com/dimitarpg13/reinforcement\_learning\_and\_game\_theory/blob/main/articles/ReinforcementLearning/High-Dimensional\_Continuous\_Control\_Using\_Generalized\_Advantage\_Estimation\_Schulman\_2018.pdf

John Schulman, F. W. (2017, 08 28). *Proximal Policy Optimization Algorithms.* Retrieved from arxiv.org: https://github.com/dimitarpg13/reinforcement\_learning\_and\_game\_theory/blob/main/articles/ReinforcementLearning/Proximal\_Policy\_Optimization\_Algorithms\_Shulman\_2017.pdf

John Schulman, S. L. (2017, Apr 20). *Trust Region Policy Optimization.* Retrieved from arxiv.org: https://github.com/dimitarpg13/reinforcement\_learning\_and\_game\_theory/blob/main/articles/ReinforcementLearning/TrustRegionPolicyOptimization\_Schulman\_2015.pdf

OpenAI. (2017, July 20). *openai baselines proximal policy optimization.* Retrieved from openai.com: https://openai.com/research/openai-baselines-ppo

Richard D. Sutton, A. G. (2020, 10 12). *Reinforcement Learning: An Introduction, Second Edition.* Retrieved from https://github.com/dimitarpg13/reinforcement\_learning\_and\_game\_theory/blob/main/books/ReinforcementLearningSuttonSecondEdition2020.pdf

Sham Kakade, J. L. (2002, July 8). *Approximately Optimal Approximate Reinforcement Learning.* Retrieved from ICML.cc: https://github.com/dimitarpg13/reinforcement\_learning\_and\_game\_theory/blob/main/articles/ReinforcementLearning/Approximate\_Optimal\_Approximate\_Reinforcement\_Learning\_KakadeLangford-icml2002.pdf

Volodymir Mnih, e. a. (2016, June 16). *Asynchronour Methods for Deep Reinforcement Learning.* Retrieved from https://github.com/dimitarpg13/reinforcement\_learning\_and\_game\_theory/blob/main/articles/ReinforcementLearning/Asynchronous\_Methods\_for\_Deep\_Reinforcement\_Learning\_Mnih\_2016.pdf

Volodymyr Mnih, e. a. (2015, Feb 26). *Human-level control through deep reinforcement learning.* Retrieved from Nature.com: https://github.com/dimitarpg13/reinforcement\_learning\_and\_game\_theory/blob/main/articles/ReinforcementLearning/Human-level\_control\_through\_deep\_reinforcement\_learning\_Mnih\_2015.pdf