# 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

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