# 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