Deep Reinforcement Learning HW5 – Soft Actor Critic

In this homework I did not manage to get the algorithm to optimize the policy. I used the following implementations as reference:

https://github.com/haarnoja/sac
https://github.com/higgsfield/RL-Adventure-2/blob/master/7.soft%20actor-critic.ipynb
https://github.com/hill-a/stable-baselines/tree/master/stable_baselines/sac

In the following pages are the results of my experiments. Each experiment has 3 runs. There is strangely little variation, suggesting that there might be a bug in random initializations or sampling.

Because I was not able to use mujoco, I used the Roboschool variants of all the environments.

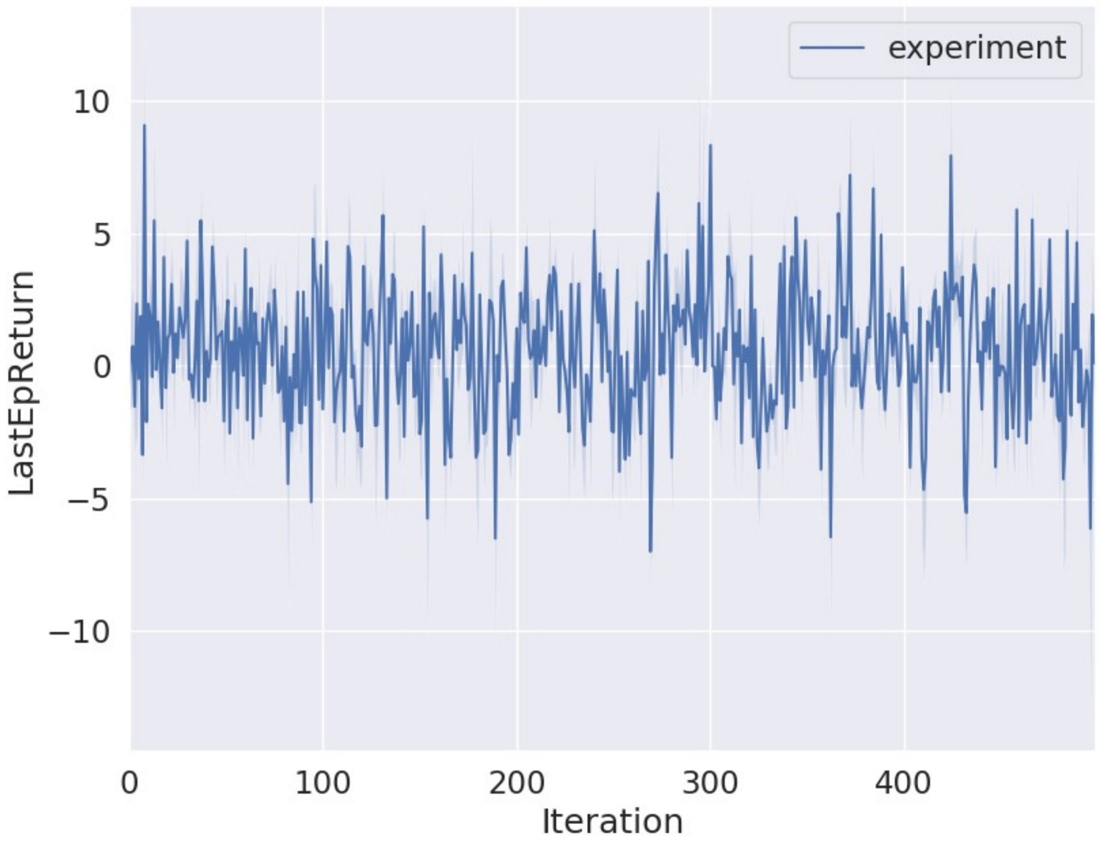

Figure 1: Half cheetah with PG and 1 Q-function

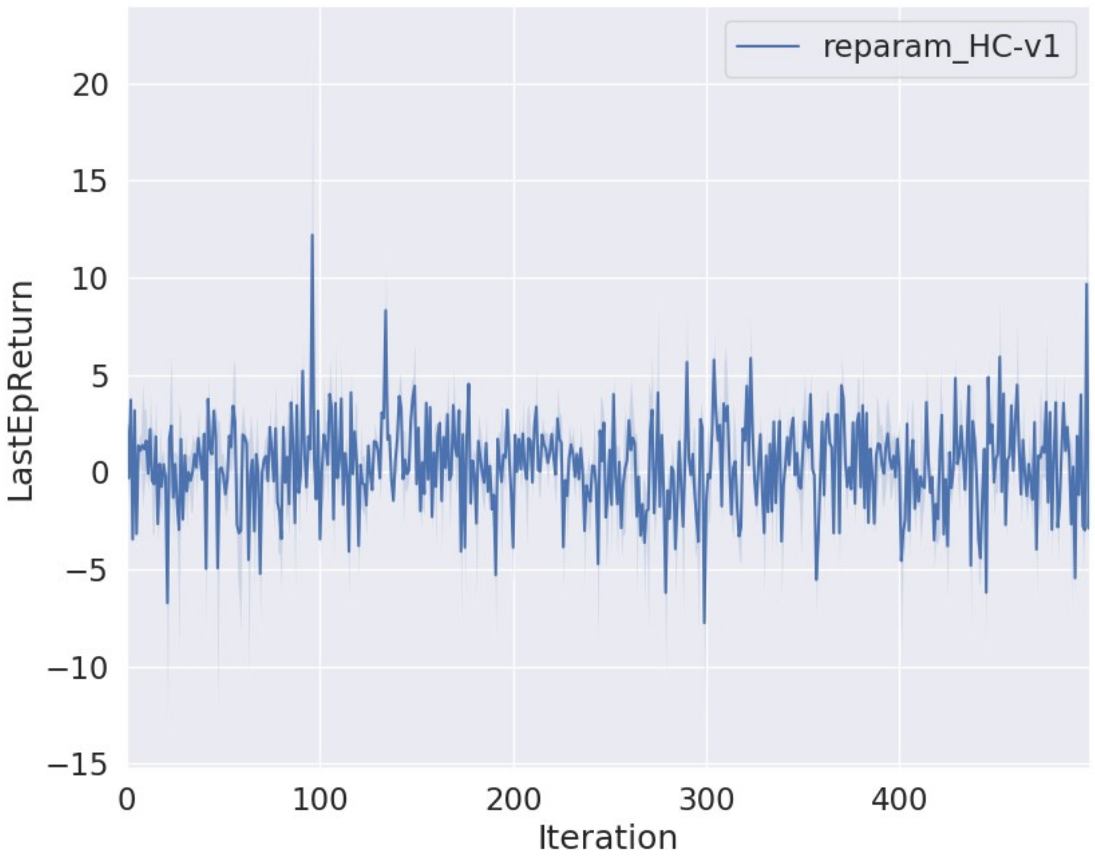

Figure 2: Half cheetah with reparameterization trick and 1 Q-function

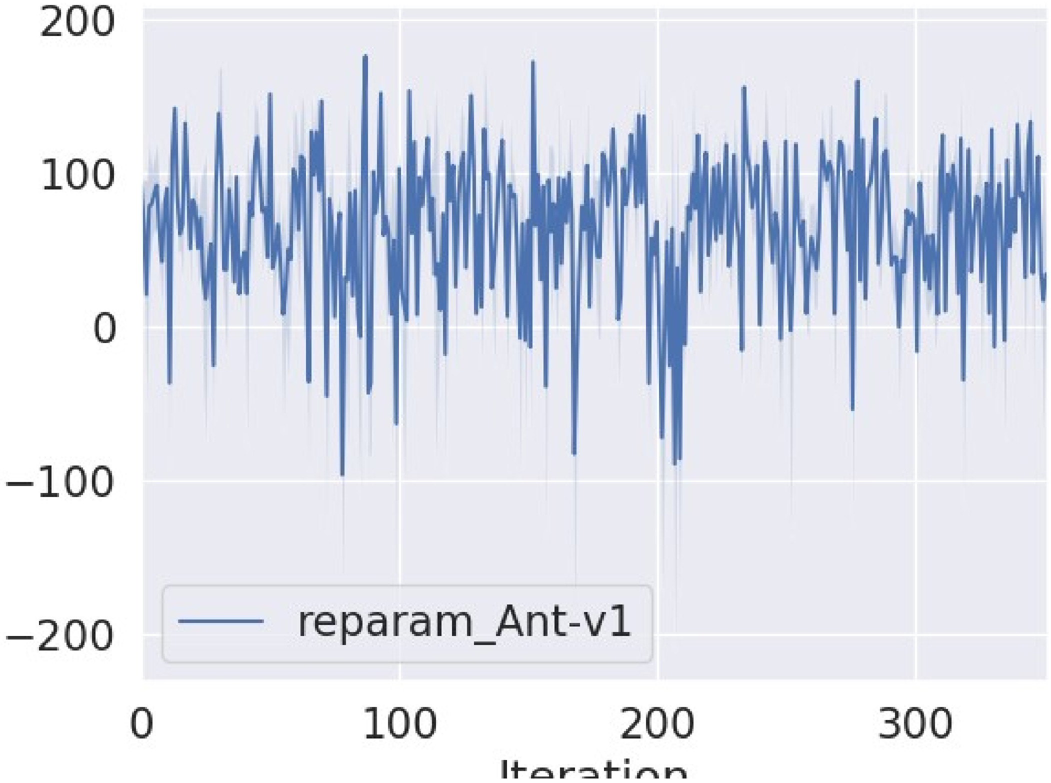

Figure 3: Ant with reparam and 1 Q-function

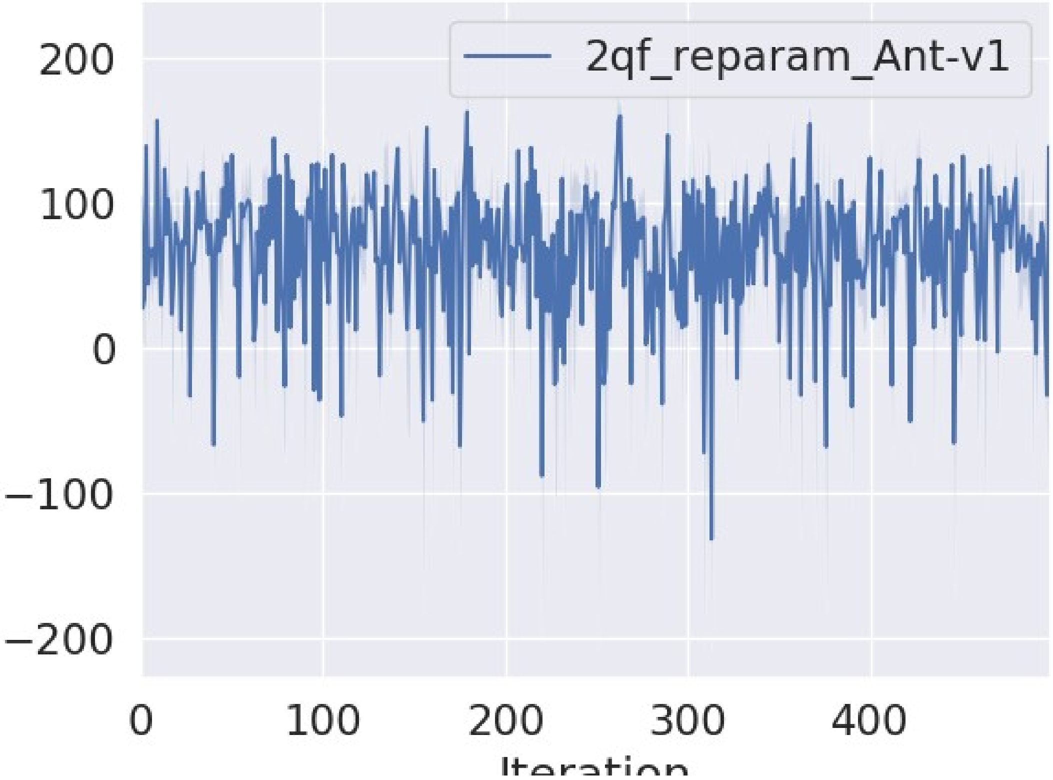

Figure 3: Ant with reparam and 2 Q-functions