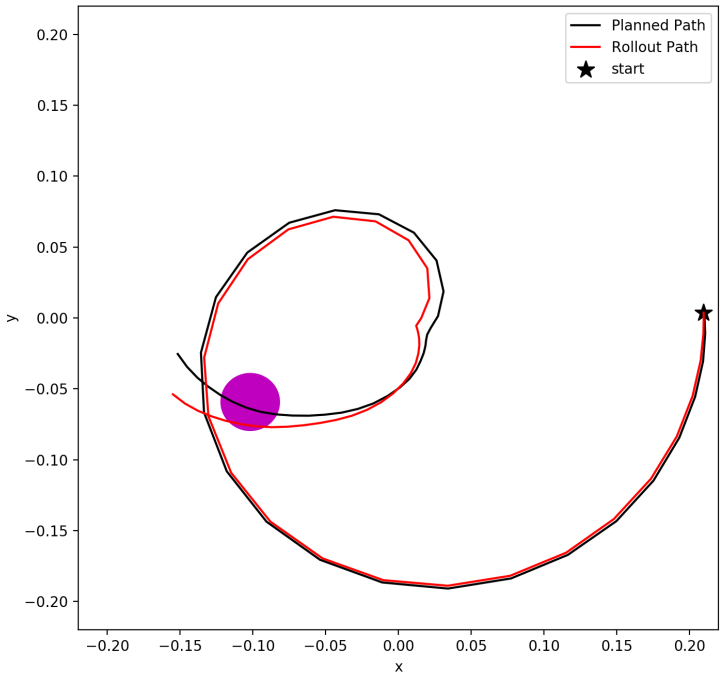
Meeting
06/23/2020

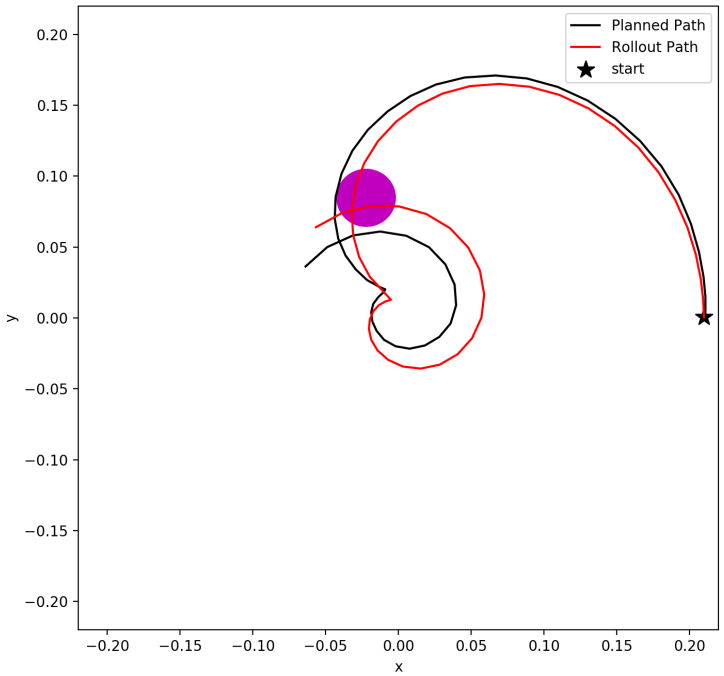
Shuo Zhang

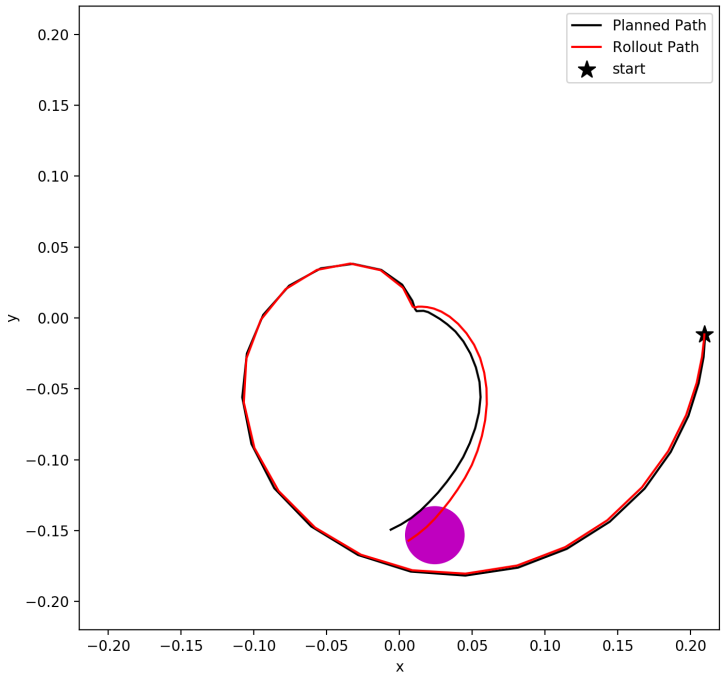
Progress

- PPO Policy Evaluation + Rollout: Reacher-v2
- PPO Policy Evaluation + Rollout: Acrobot-v1
- Many experiments for PPO on adaptive hand with hyperparameter search (Both locally and on iLab server)

Policy Evaluation + Rollout: Reacher-v2

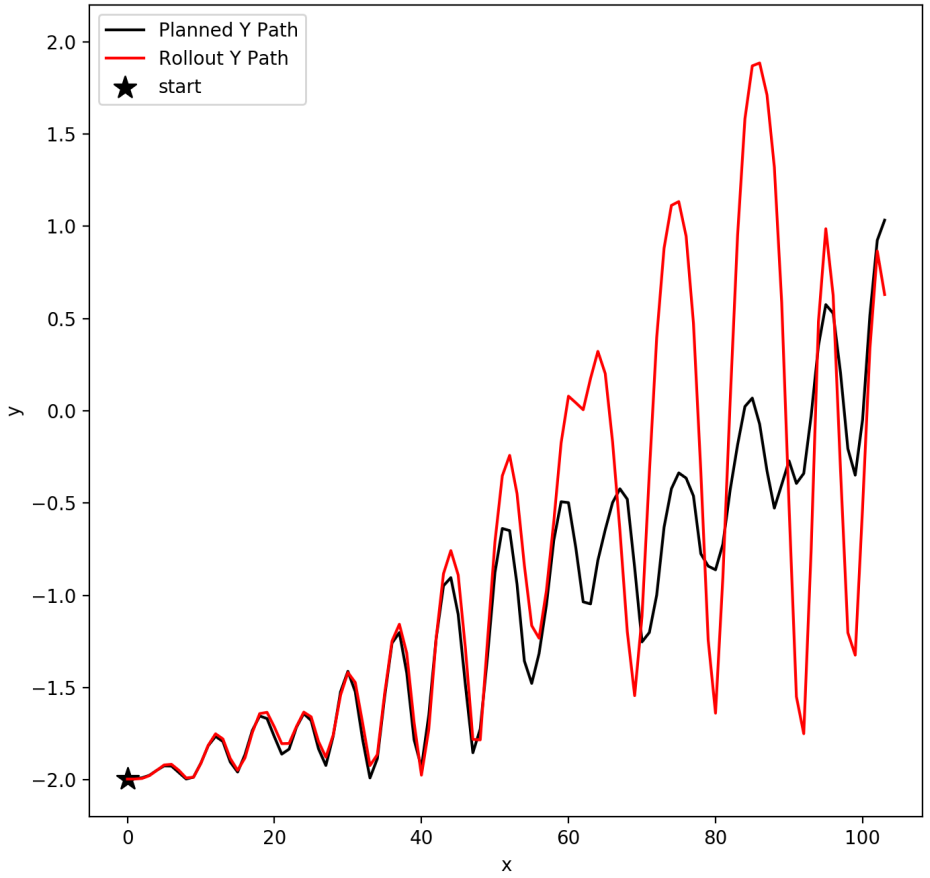






PPO is trained with learning rate of 3e-4 and 1 million timesteps

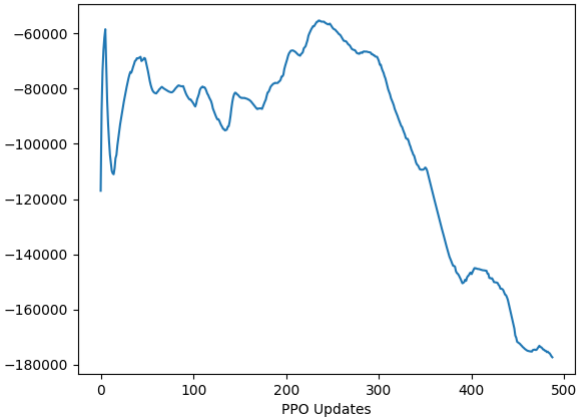
Policy Evaluation + Rollout: Acrobot-v1

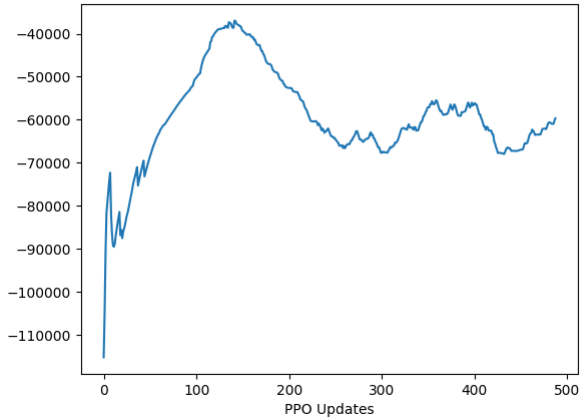


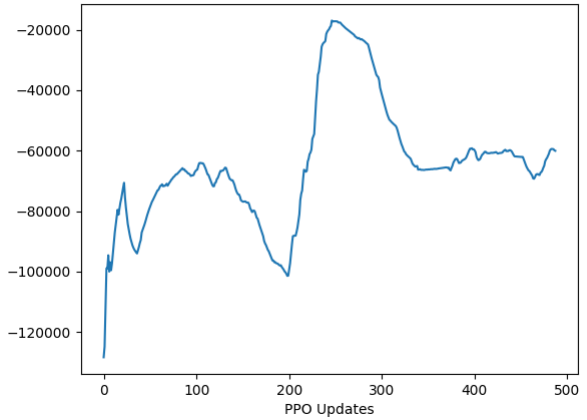
PPO is trained with learning rate of 3e-4 and 1 million timesteps

Adaptive Hand PPO: Goal Loc 8
learning rate: 3e-4; 1 million timesteps

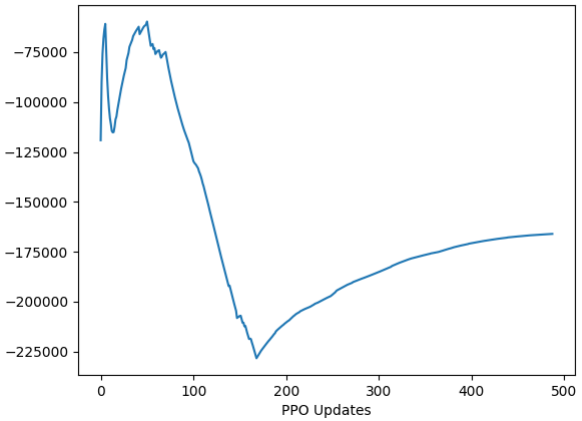
Without control reward

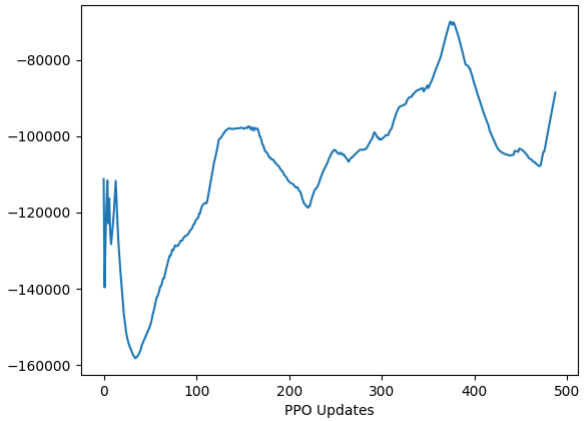


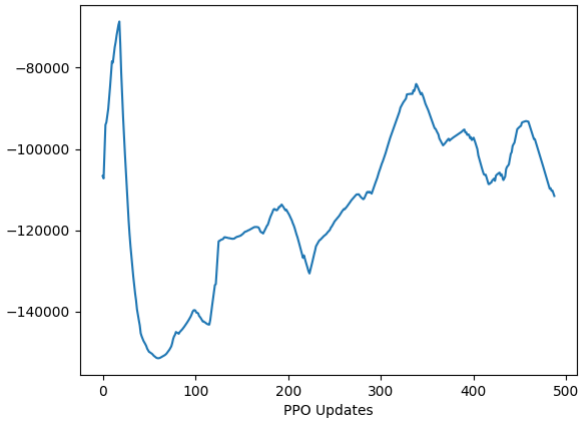




With control reward



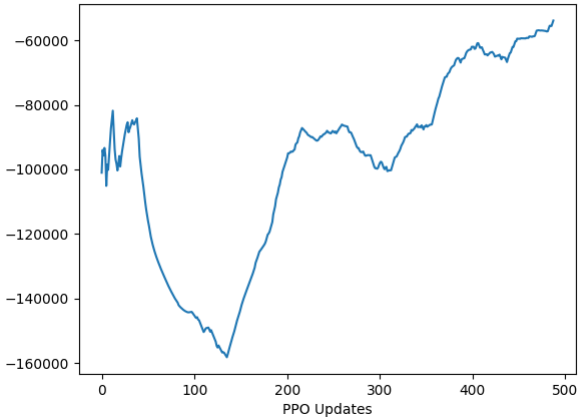


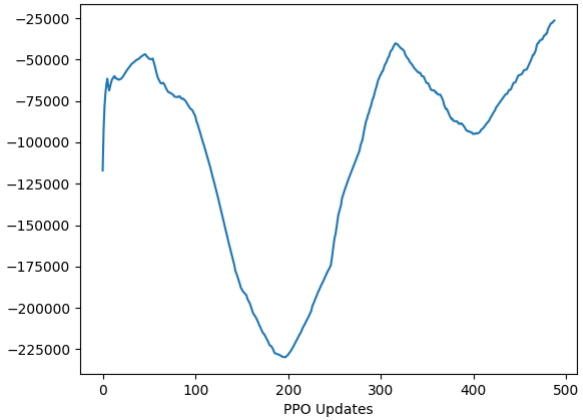


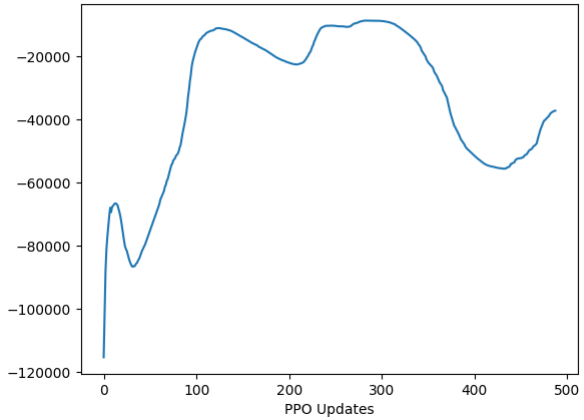
Seed 0 Seed 1 Seed 2

Adaptive Hand PPO: Goal Loc 8
learning rate: 1e-4; 1 million timesteps

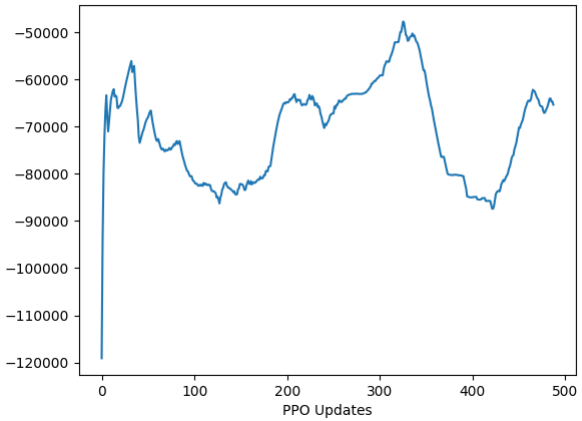
Without control reward

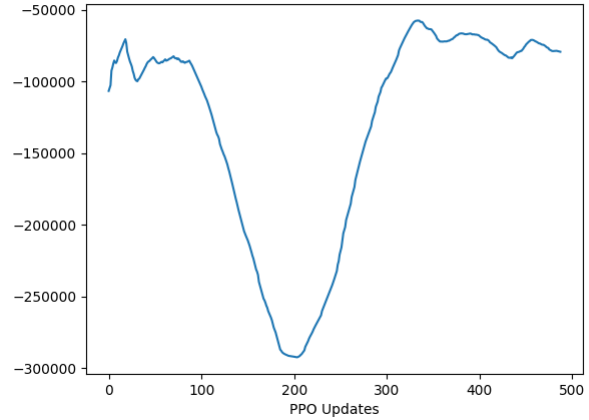


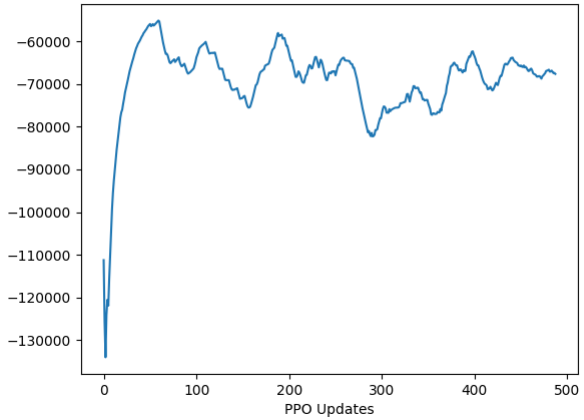




With control reward



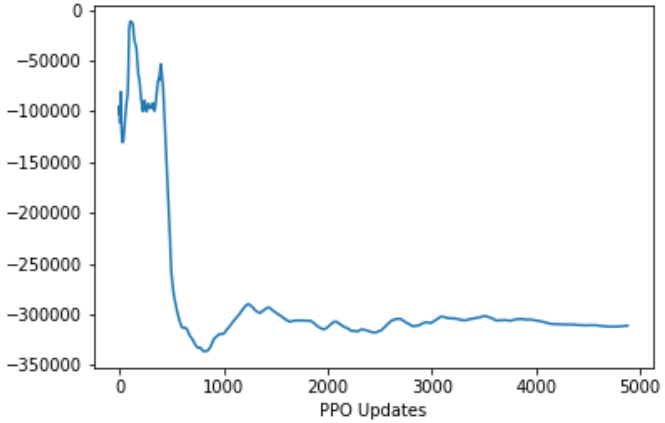


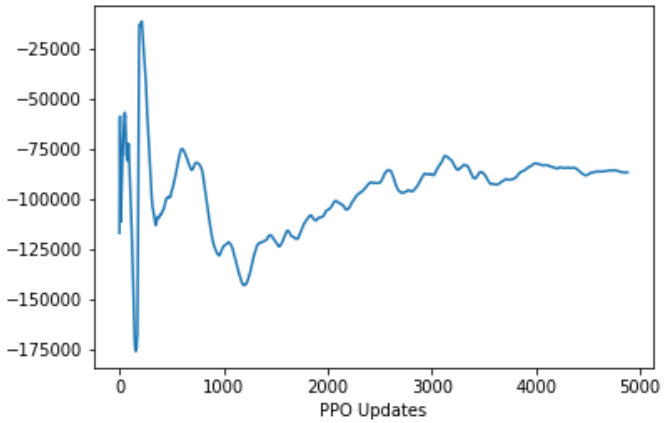


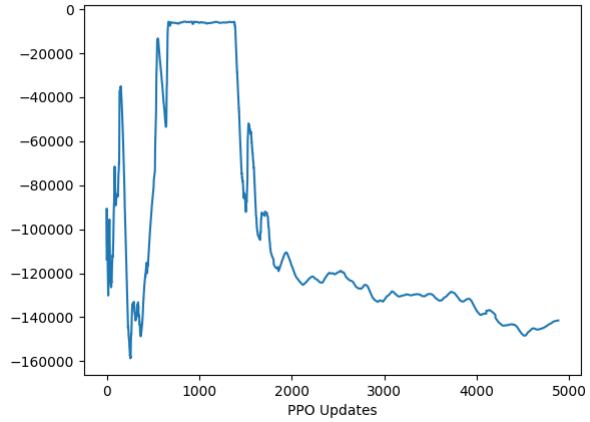
Seed 0 Seed 1 Seed 2

Adaptive Hand PPO: Goal Loc 8
learning rate: 3e-4; 10 million timesteps

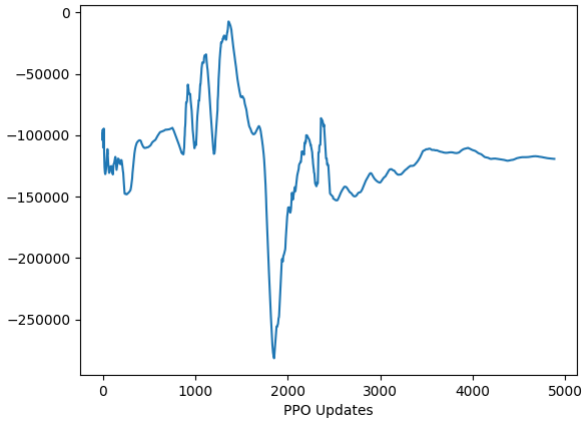
Without control reward

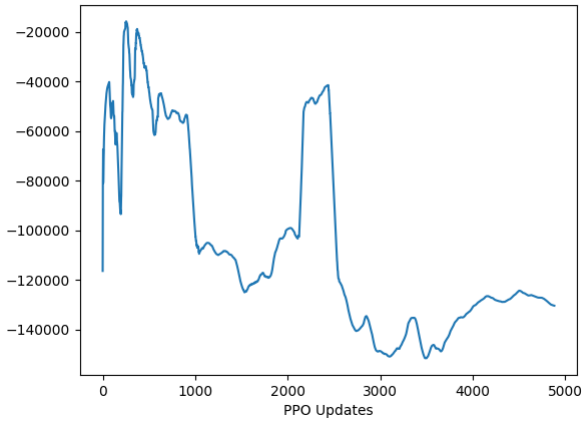


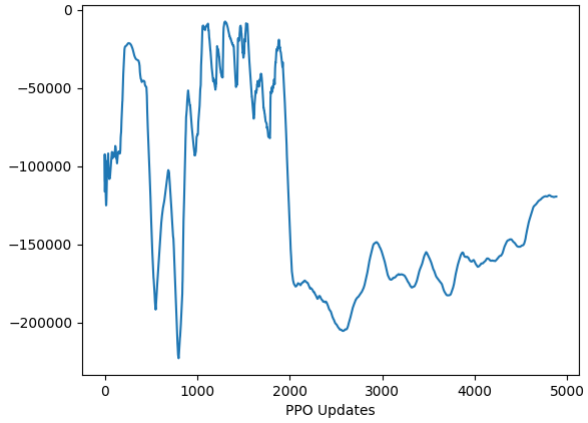




With control reward







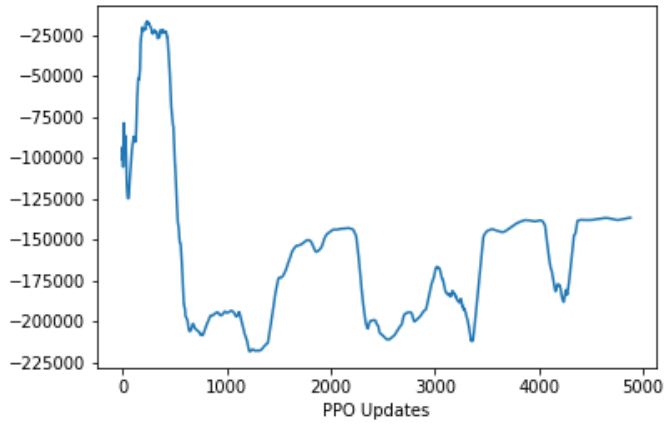
Seed 0

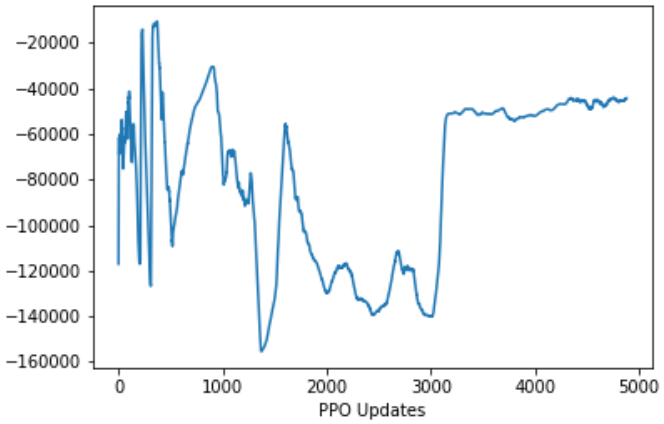
Seed 1

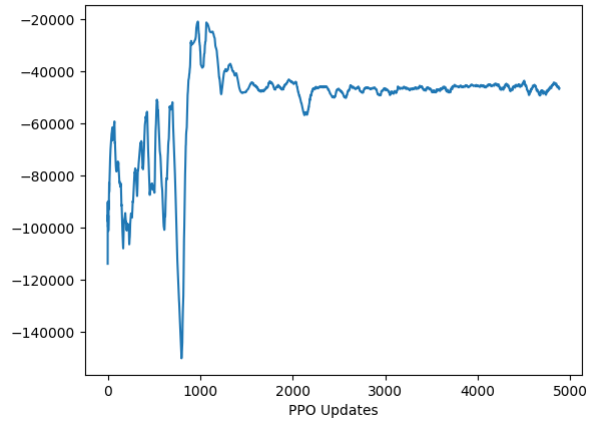
Seed 2

Adaptive Hand PPO: Goal Loc 8
learning rate: 1e-4; 10 million timesteps

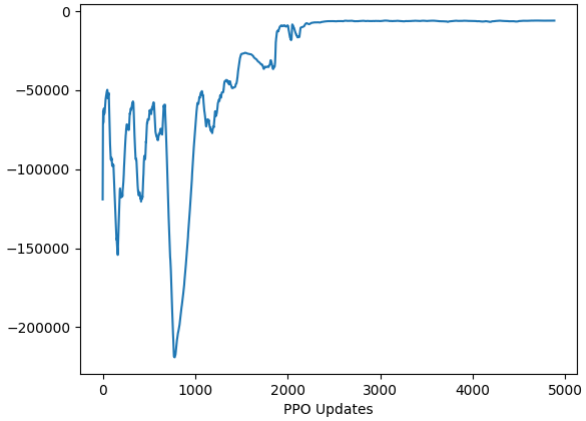
Without control reward

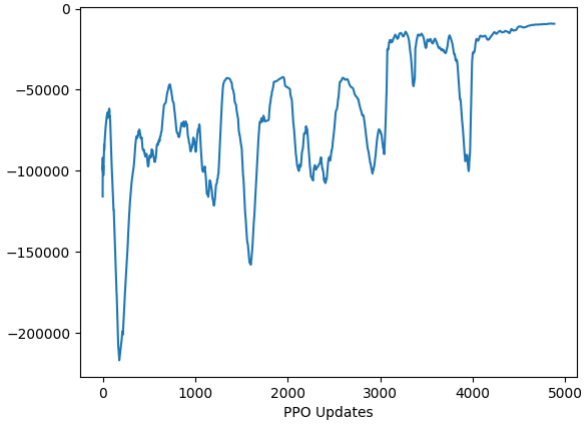


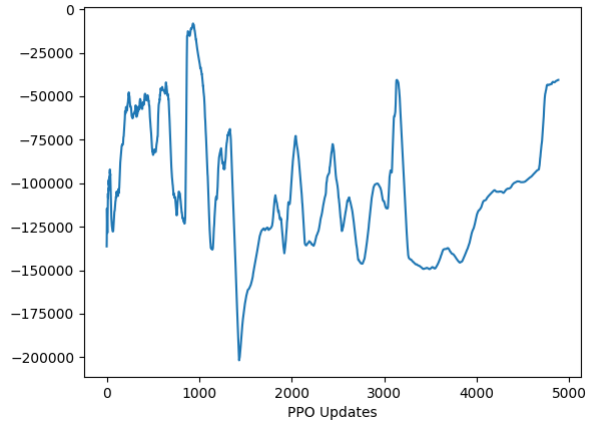




With control reward





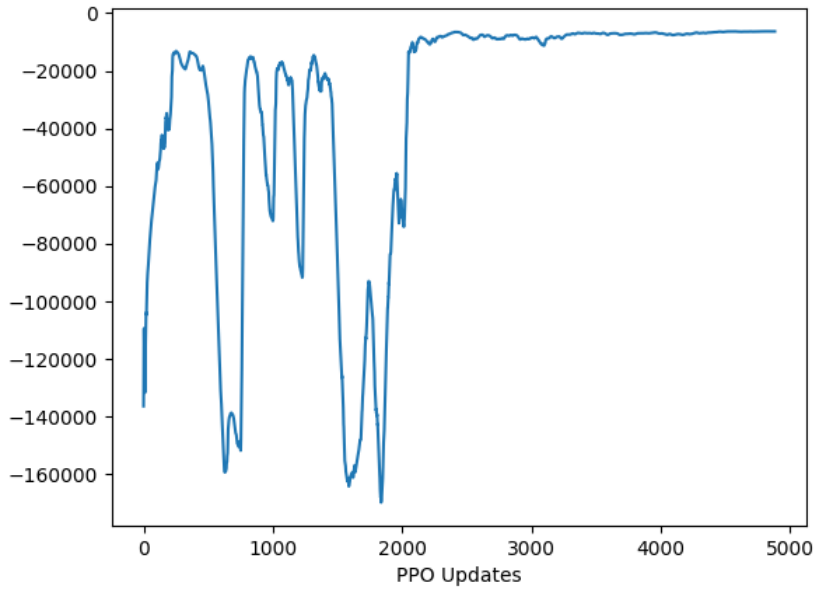


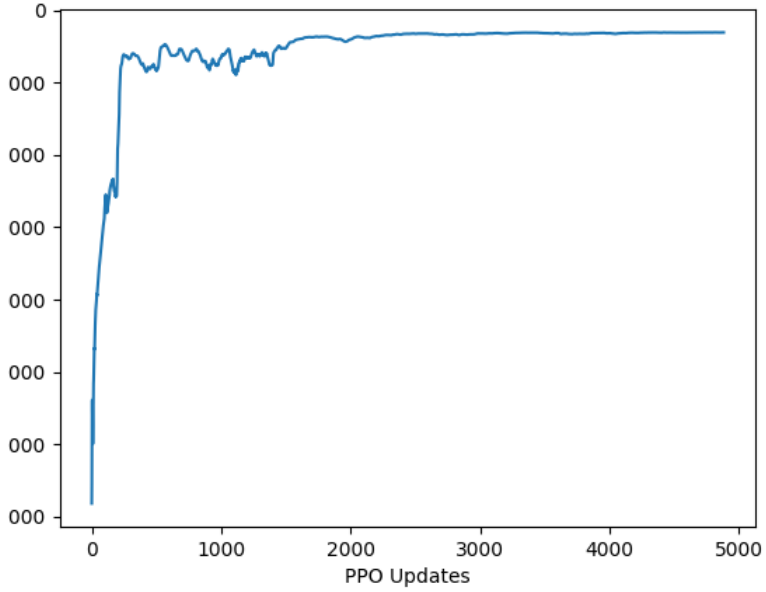
Seed 0

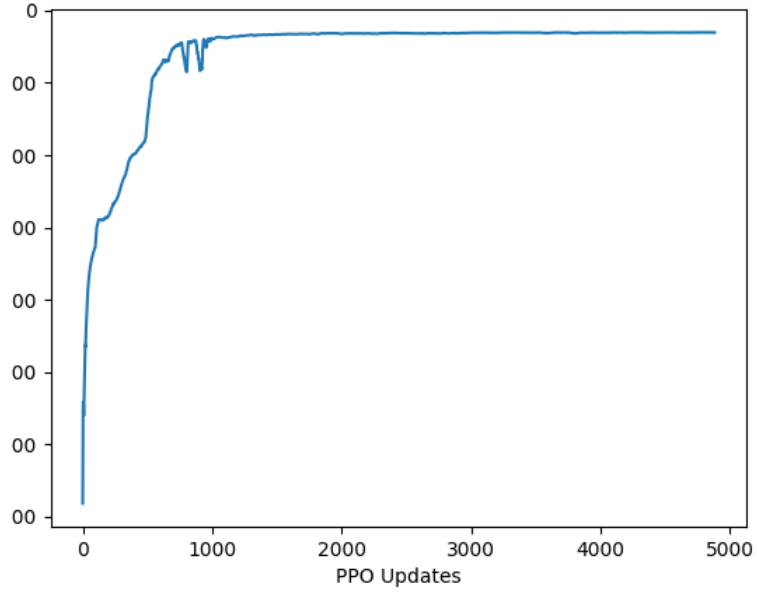
Seed 1

Seed 2

Adaptive Hand PPO: Goal Loc 8
10 million timesteps







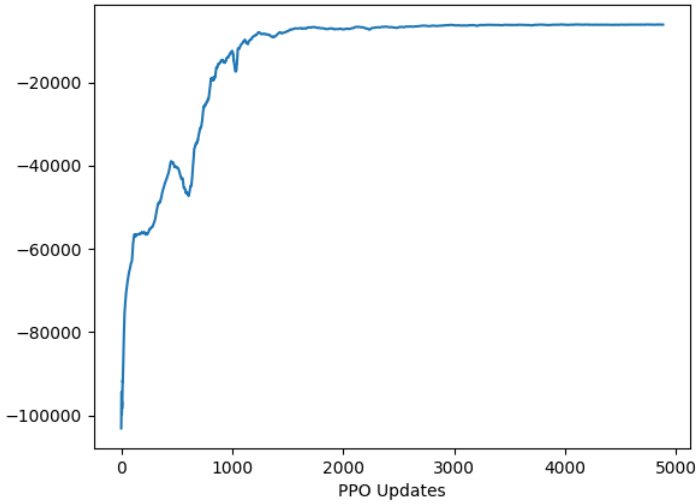
Learning rate
5e-5

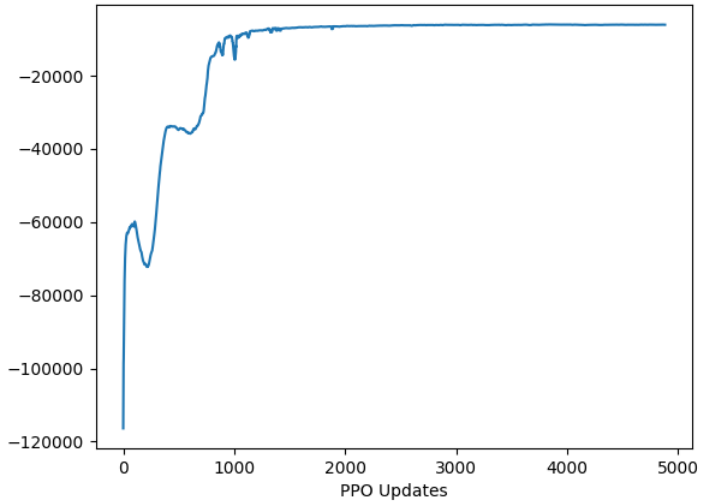
Learning rate
3e-5

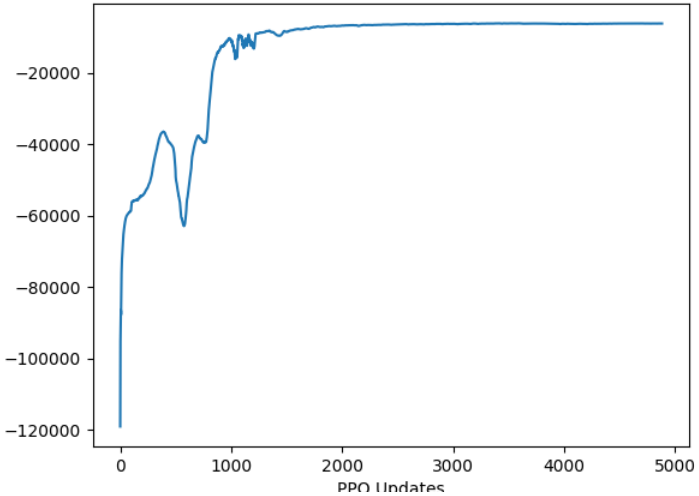
Learning rate
1e-5

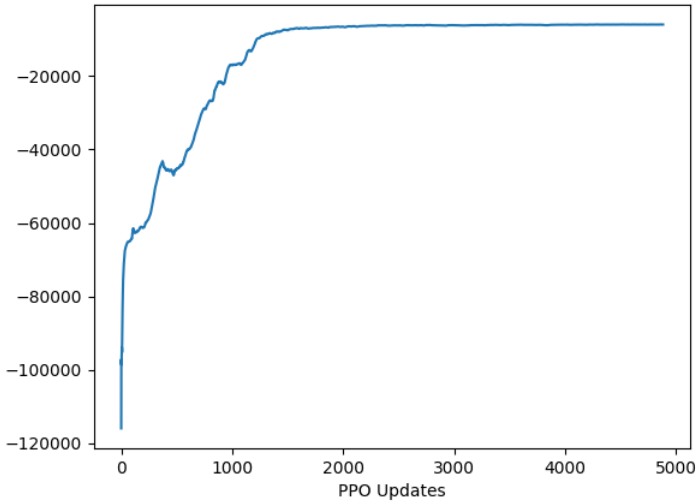
With control reward

Adaptive Hand PPO: Goal Loc 8
learning rate: 1e-5; 10 million timesteps

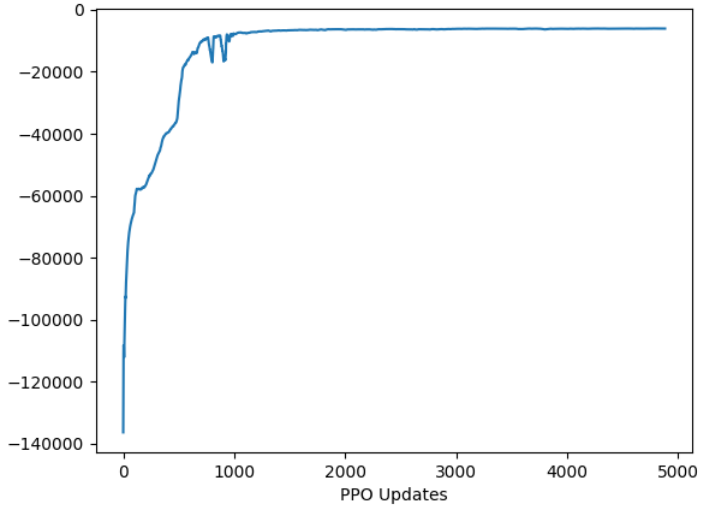




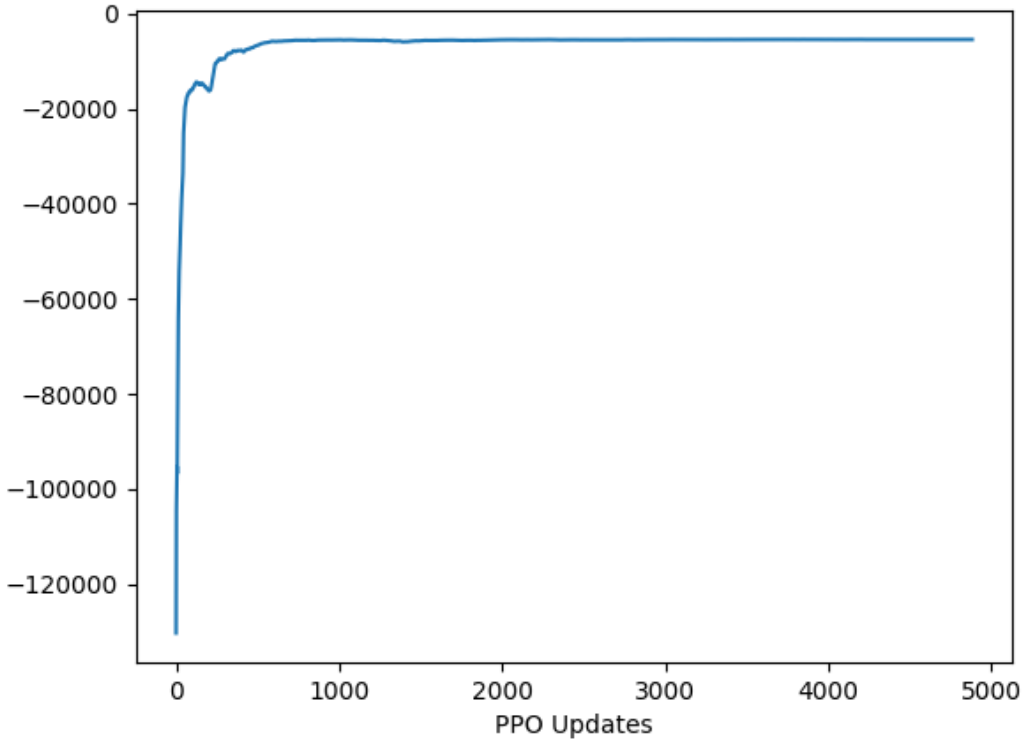




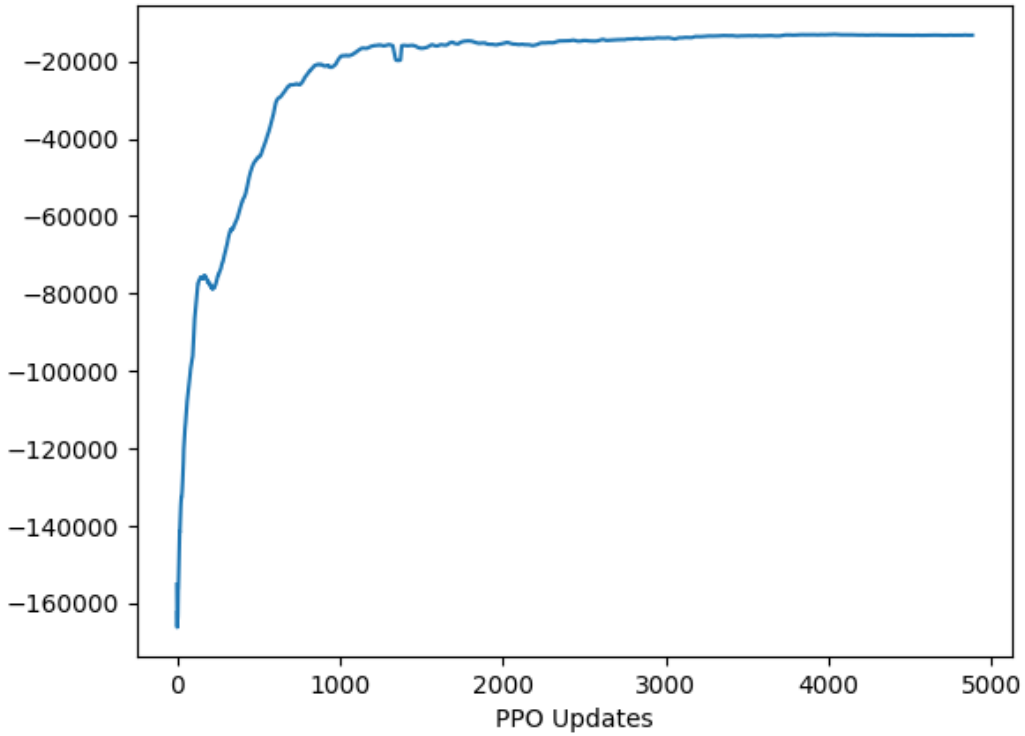
With control reward (from seed 0 to seed 4)



Adaptive Hand PPO: Different Goal Locations
learning rate: 1e-5; 10 million timesteps



With control reward: Goal Loc 7



With control reward: Goal Loc 0

To Do List

- Train PPO for gazebo hand in environment with obstacles
  - Normal obstacles scenario for 5 different goal locations
  - Horsehoe obstacles scenario

- Train PPO for gazebo hand in environment with obstacles (goal location as a part of state information)

- Rollout actions from trained policy on real gazebo:
  - 5 different goal locations
  - horsehoe scenario