

## Assignment 2: Policy Gradient

**Total points: 9 + 4.5 + 4**  
**Bonus: 2.5 (Q9)**

**Andrew ID:** Write your Andrew ID here.

**Collaborators:** Write the Andrew IDs of your collaborators here (if any).

**NOTE:** Please do NOT change the sizes of the answer blocks or plots.

## 5 Small-Scale Experiments

### 5.1 Experiment 1 (Cartpole) – [5 points total]

#### 5.1.1 Configurations

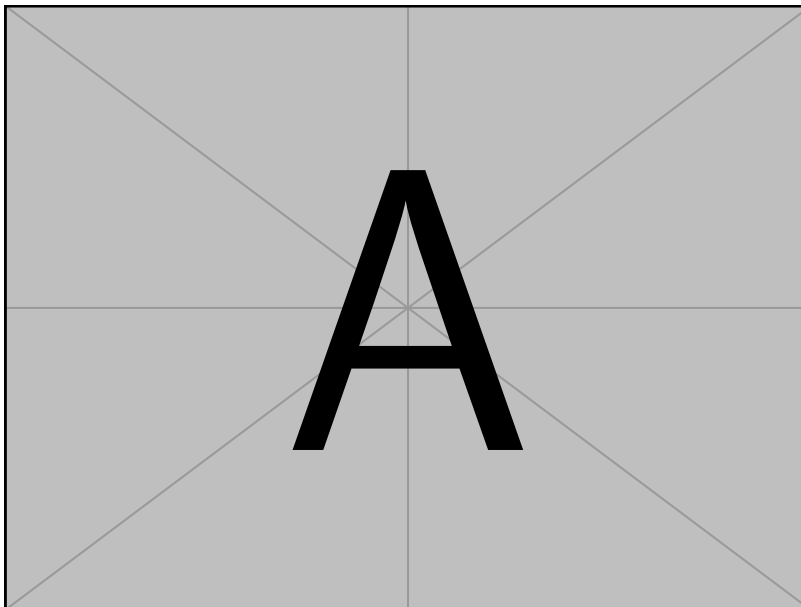
##### Q5.1.1

```
pythonrob831/scripts/run_hw2.py--env_name CartPole-v0-n150-b1500\  
-dsa --exp_name q1_sb_no_rtg_dsa  
  
pythonrob831/scripts/run_hw2.py--env_name CartPole-v0-n150-b1500\  
-rtg -dsa --exp_name q1_sb_rtg_dsa  
  
pythonrob831/scripts/run_hw2.py--env_name CartPole-v0-n150-b1500\  
-rtg --exp_name q1_sb_rtg_na  
  
pythonrob831/scripts/run_hw2.py--env_name CartPole-v0-n150-b6000\  
-dsa--exp_name q1_lb_no_rtg_dsa  
  
pythonrob831/scripts/run_hw2.py--env_name CartPole-v0-n150-b6000\  
-rtg -dsa --exp_name q1_lb_rtg_dsa  
  
pythonrob831/scripts/run_hw2.py--env_name CartPole-v0-n150-b6000\  
-rtg --exp_name q1_lb_rtg_na
```

#### 5.1.2 Plots

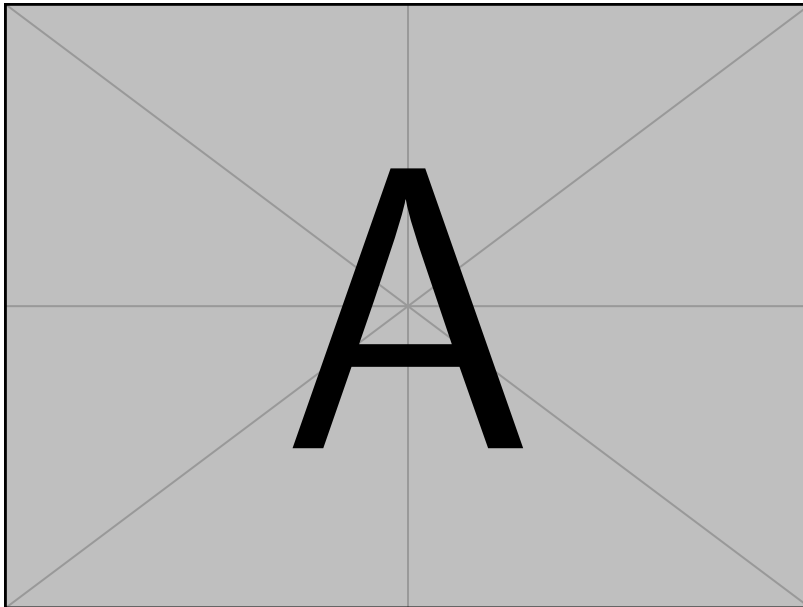
##### 5.1.2.1 Small batch – [1 points]

##### Q5.1.2.1



**5.1.2.2 Large batch – [1 points]**

Q5.1.2.2

**5.1.3 Analysis****5.1.3.1 Value estimator – [1 points]**

Q5.1.3.1

**5.1.3.2 Advantage standardization – [1 points]**

Q5.1.3.2

**5.1.3.3 Batch size – [1 points]**

Q5.1.3.3

**5.2 Experiment 2 (InvertedPendulum) – [4 points total]****5.2.1 Configurations – [1.5 points]**

Q5.2.1

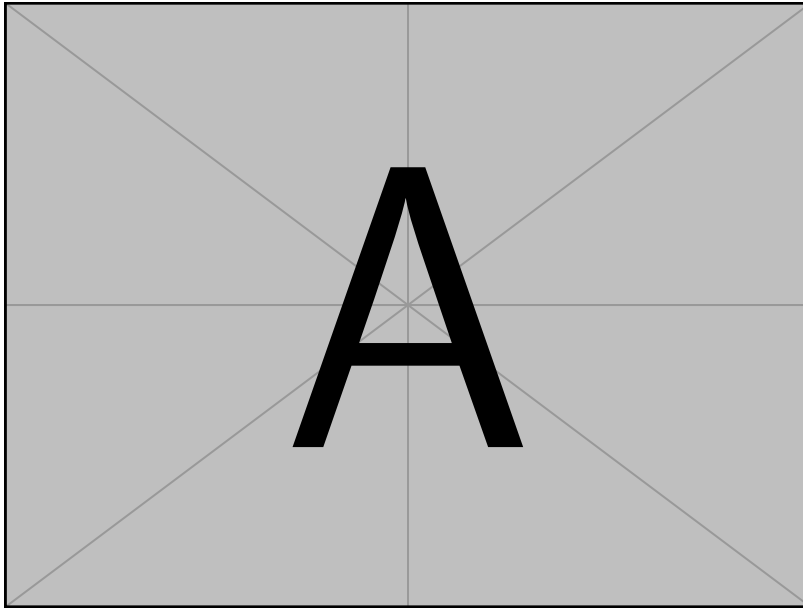
```
python rob831/scripts/run_hw2.py --env_name InvertedPendulum-v4 \
--ep_len1000--discount0.92-n100-l2-s64-b<b*>-lr<r*>-rtg\
--exp_name q2_b<b*>-r<r*>
```

**5.2.2 smallest  $b^*$  and largest  $r^*$  (same run) – [1.5 points]**

Q5.2.2

### 5.2.3 Plot – [1 points]

Q5.2.3



## 7 More Complex Experiments

### 7.1 Experiment 3 (LunarLander) – [1 points total]

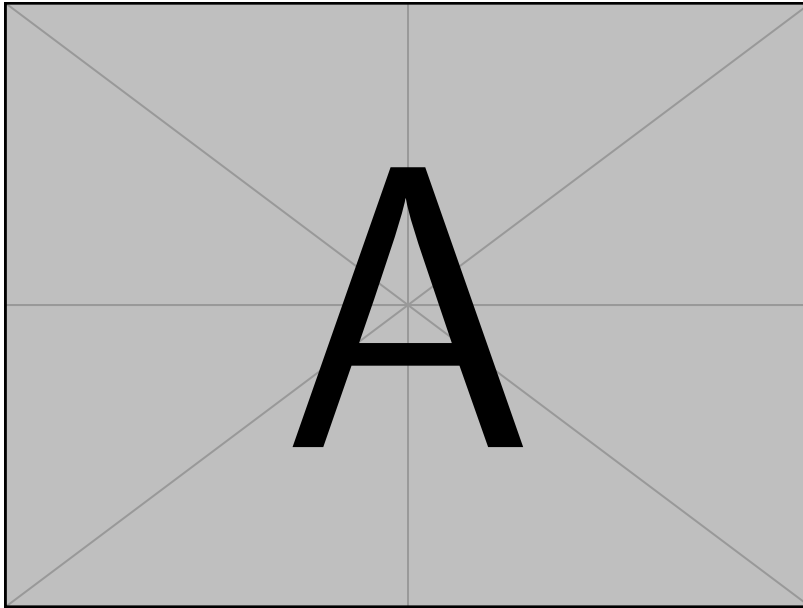
#### 7.1.1 Configurations

Q7.1.1

```
python rob831/scripts/run_hw2.py \  
  --env_name LunarLanderContinuous-v4 --ep_len 1000 \  
  --discount 0.99 -n 100 -l 2 -s 64 -b 10000 -lr 0.005 \  
  --reward_to_go --nn_baseline --exp_name q3_b10000_r0.005
```

### 7.1.2 Plot – [1 points]

Q7.1.2

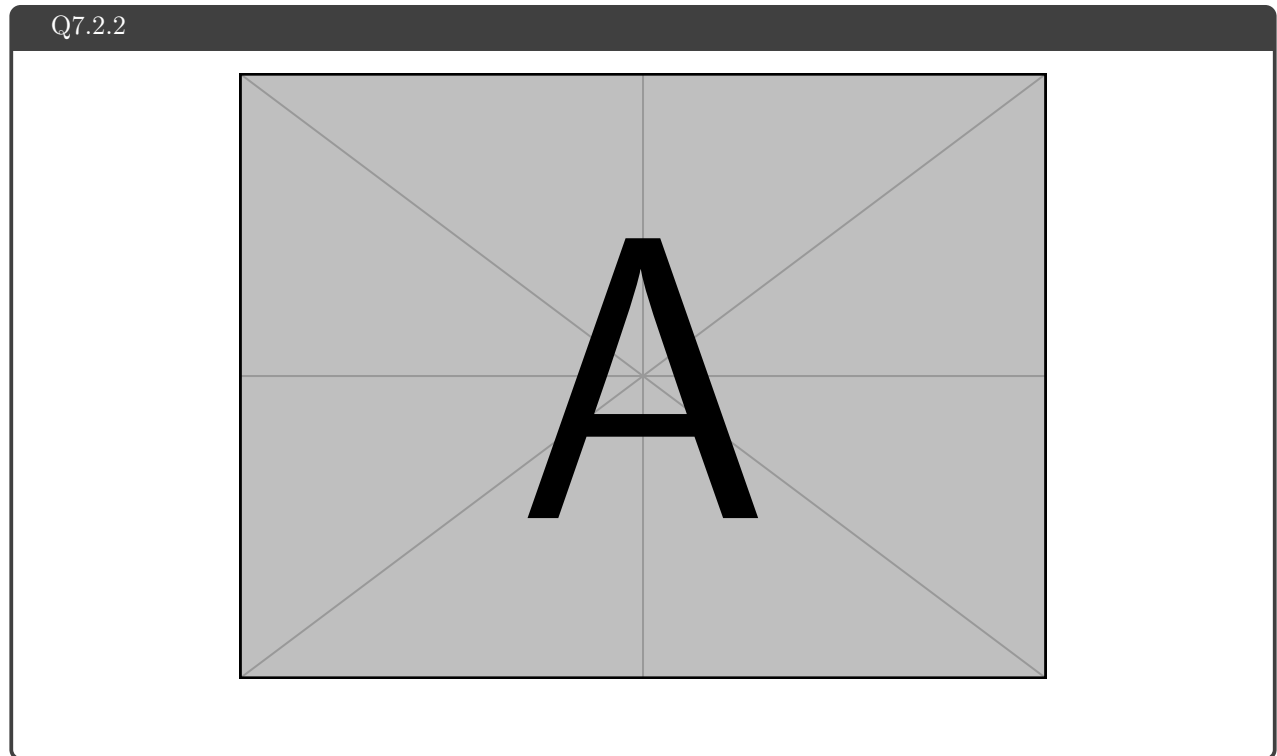


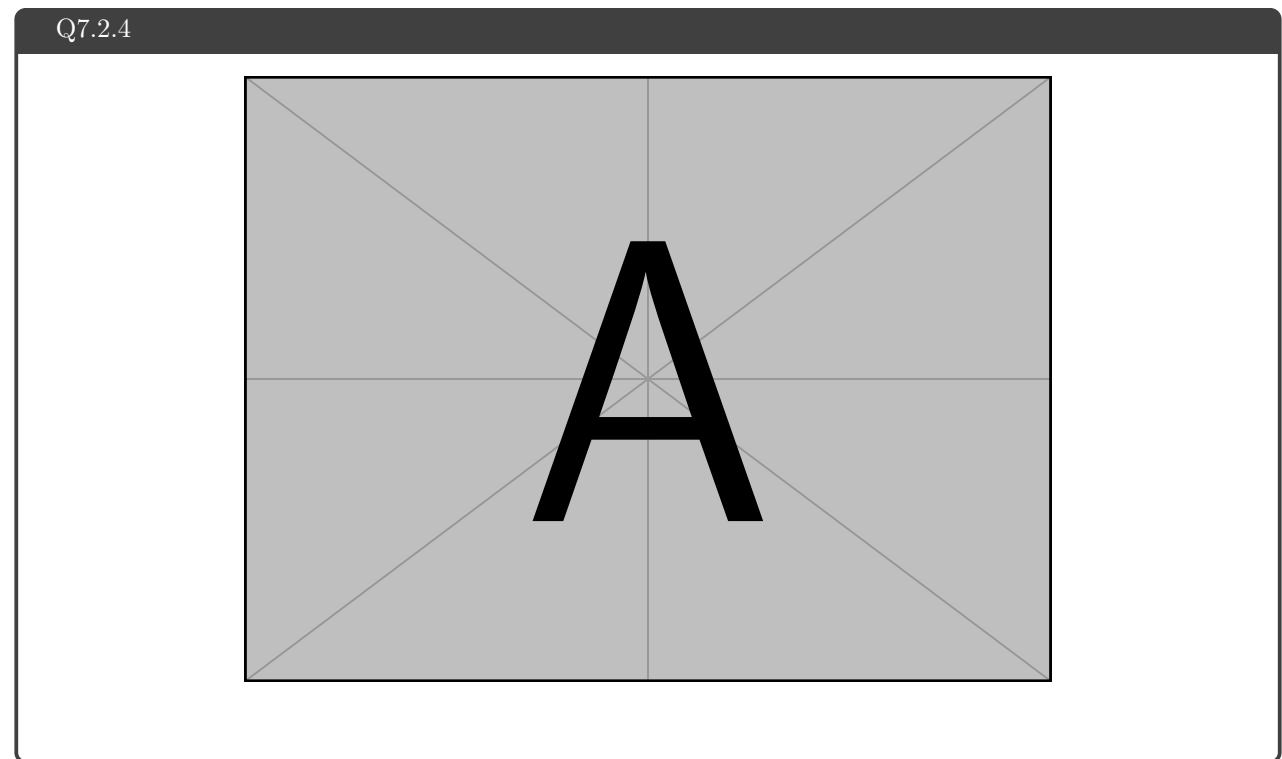
## 7.2 Experiment 4 (HalfCheetah) – [3.5 points]

### 7.2.1 Configurations

Q7.2.1

```
python rob831/scripts/run_hw2.py --env_name HalfCheetah-v4 --ep_len 150 \  
  --discount 0.95 -n 100 -l 2 -s 32 -b 10000 -lr 0.02 \  
  --exp_name q4_search_b10000_lr0.02  
python rob831/scripts/run_hw2.py --env_name HalfCheetah-v4 --ep_len 150 \  
  --discount 0.95 -n 100 -l 2 -s 32 -b 10000 -lr 0.02 -rtg \  
  --exp_name q4_search_b10000_lr0.02_rtg  
python rob831/scripts/run_hw2.py --env_name HalfCheetah-v4 --ep_len 150 \  
  --discount 0.95 -n 100 -l 2 -s 32 -b 10000 -lr 0.02 --nn_baseline \  
  --exp_name q4_search_b10000_lr0.02_nnbaseline  
python rob831/scripts/run_hw2.py --env_name HalfCheetah-v4 --ep_len 150 \  
  --discount 0.95 -n 100 -l 2 -s 32 -b 10000 -lr 0.02 -rtg --nn_baseline \  
  --exp_name q4_search_b10000_lr0.02_rtg_nnbaseline
```

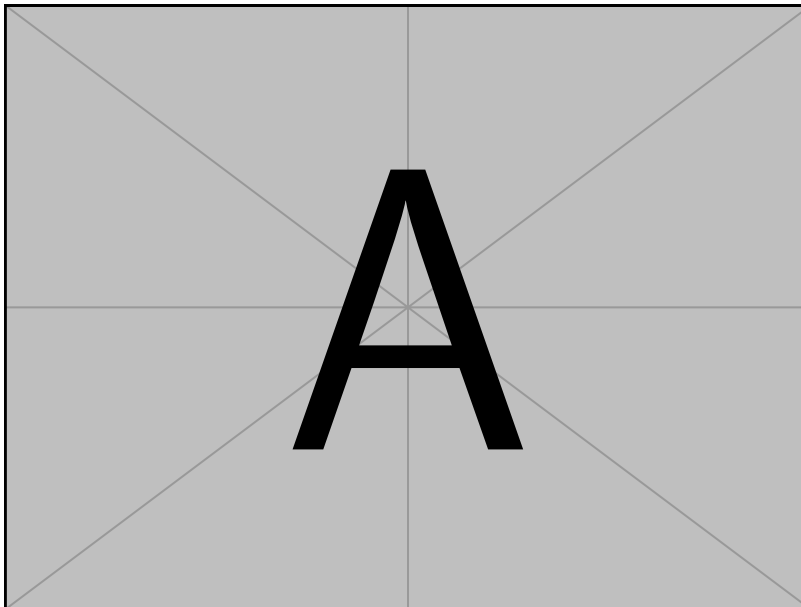
**7.2.2 Plot – [1 points]****7.2.3 Optimal  $b^*$  and  $r^*$  – [0.5 points]**

**7.2.4 Plot – [0.5 points]****7.2.5 Describe how  $b^*$  and  $r^*$  affect task performance – [0.5 points]**

Q7.2.5

**7.2.6 Configurations with optimal  $b^*$  and  $r^*$  – [0.5 points]****Q7.2.6**

```
python rob831/scripts/run_hw2.py --env_name HalfCheetah-v4 --ep_len 150 \  
--discount 0.95 -n 100 -l 2 -s 32 -b <b*> -lr <r*> \  
--exp_name q4_b<b*>_r<r*>  
  
python rob831/scripts/run_hw2.py --env_name HalfCheetah-v4 --ep_len 150 \  
--discount 0.95 -n 100 -l 2 -s 32 -b <b*> -lr <r*> -rtg \  
--exp_name q4_b<b*>_r<r*>_rtg  
  
python rob831/scripts/run_hw2.py --env_name HalfCheetah-v4 --ep_len 150 \  
--discount 0.95 -n 100 -l 2 -s 32 -b <b*> -lr <r*> --nn_baseline \  
--exp_name q4_b<b*>_r<r*>_nnbaseline  
  
python rob831/scripts/run_hw2.py --env_name HalfCheetah-v4 --ep_len 150 \  
--discount 0.95 -n 100 -l 2 -s 32 -b <b*> -lr <r*> -rtg --nn_baseline \  
--exp_name q4_b<b*>_r<r*>_rtg_nnbaseline
```

**7.2.7 Plot for four runs with optimal  $b^*$  and  $r^*$  – [0.5 points]****Q7.2.7****8 Implementing Generalized Advantage Estimation**



## 8.1 Experiment 5 (Hopper) – [4 points]

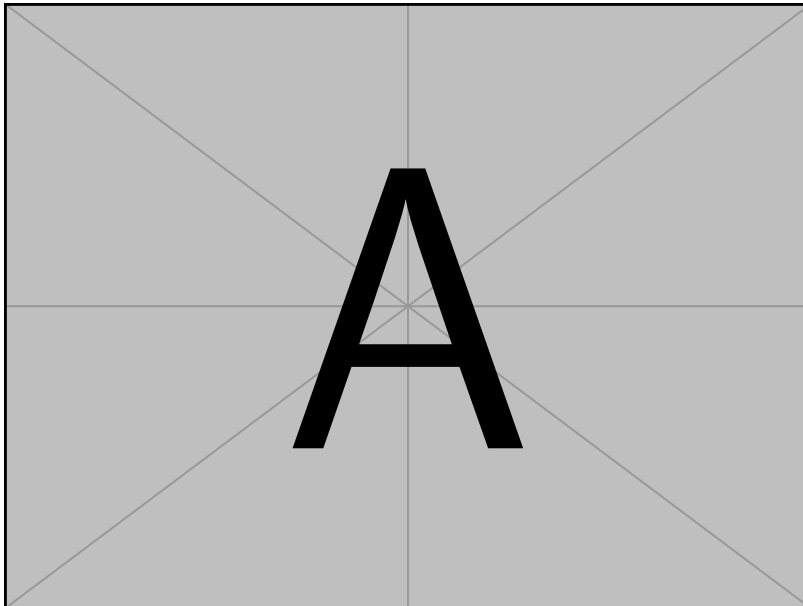
### 8.1.1 Configurations

#### Q8.1.1

```
#  $\lambda \in [0, 0.95, 0.99, 1]$ 
python rob831/scripts/run_hw2.py \
  --env_name Hopper-v4 --ep_len 1000
  --discount 0.99 -n 300 -l 2 -s 32 -b 2000 -lr 0.001 \
  --reward_to_go --nn_baseline --action_noise_std 0.5 --gae_lambda < $\lambda$ > \
  --exp_name q5_b2000_r0.001_lambda< $\lambda$ >
```

### 8.1.2 Plot – [2 points]

#### Q8.1.2



### 8.1.3 Describe how $\lambda$ affects task performance – [2 points]

#### Q8.1.3

## 9 More Bonus!

### 9.1 Parallelization – [1.5 points]

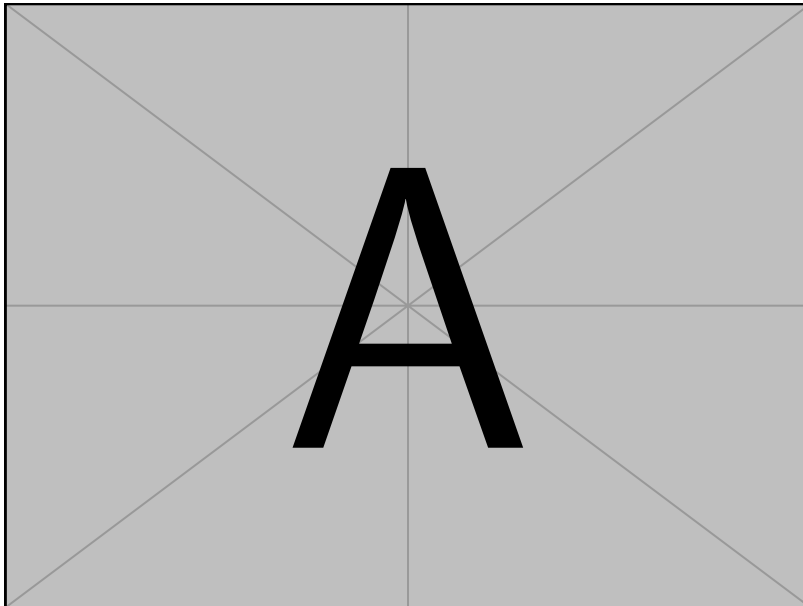
Q9.1

Difference in training time:

```
python rob831/scripts/run_hw2.py \
```

### 9.2 Multiple gradient steps – [1 points]

Q9.1



```
python rob831/scripts/run_hw2.py \
```