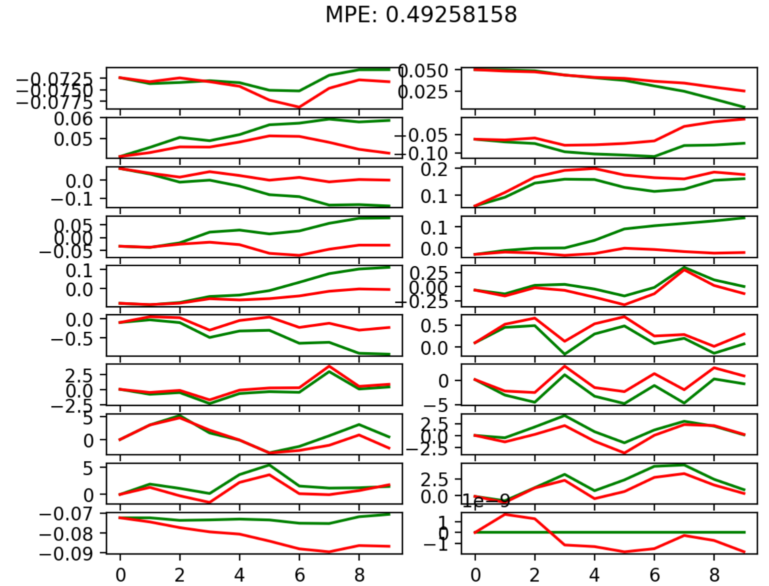
Assignment 4: Model-Based RL and Exploration

Andrew ID: `bharathh`
Collaborators: `ChatGPT`
NOTE: Please do **NOT** change the sizes of the answer blocks or plots.

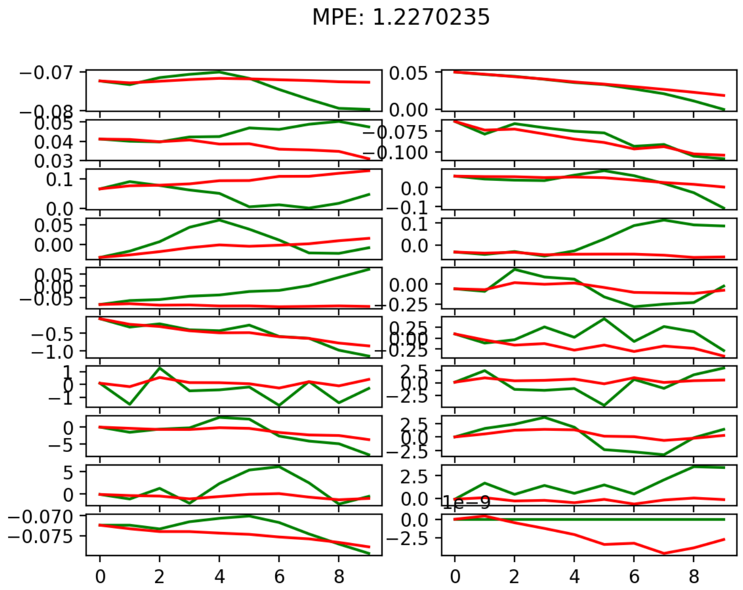
1 Problem 1: Dynamics Model Training [4pts]

The model `q1_cheetah_n500_arch2x200` performs best with least MPE as well as training loss. The other two, the smaller model that runs for same number of steps, and the same sized model that runs for only 10 steps, both perform worse. This indicates that a larger model can better approximate complex dynamics, at the same time it requires sufficient number of training steps to do so.

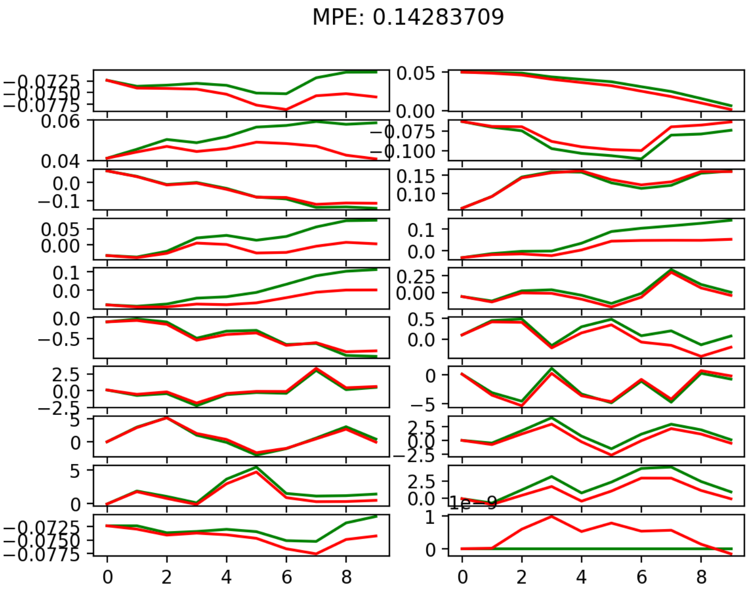
Plot 1



Plot 2

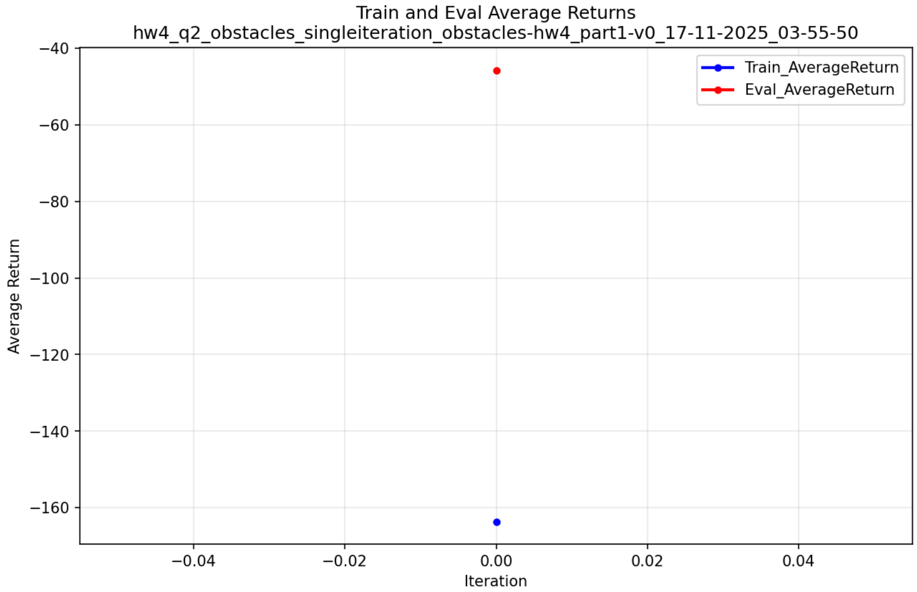

MPE: 1.2270235
−0.07
−0.08
0.05
0.04
0.03
0.1
0.0
0.05
0.00
0.05
0.00
−0.05
−0.5
−1.0
1
0
−1
0
−5
5
0
−0.070
−0.075
0.05
0.00
−0.075
−0.100
0.0
−0.1
0.1
0.0
0.00
−0.25
0.25
0.00
−0.25
2.5
0.0
−2.5
2.5
0.0
−2.5
2.5
0.0
1e−9
0.0
−2.5
0
2
4
6
8

Plot 3


MPE: 0.14283709
−0.0725
−0.0750
−0.0775
0.06
0.04
0.0
−0.1
0.05
0.00
0.1
0.0
0.0
−0.5
2.5
0.0
−2.5
5
0
5
0
−0.0725
−0.0750
−0.0775
0.05
0.00
−0.075
−0.100
0.15
0.10
0.1
0.0
0.25
0.00
0.5
0.0
0
−5
2.5
0.0
−2.5
2.5
0.0
1e−9
1
0
0
2
4
6
8

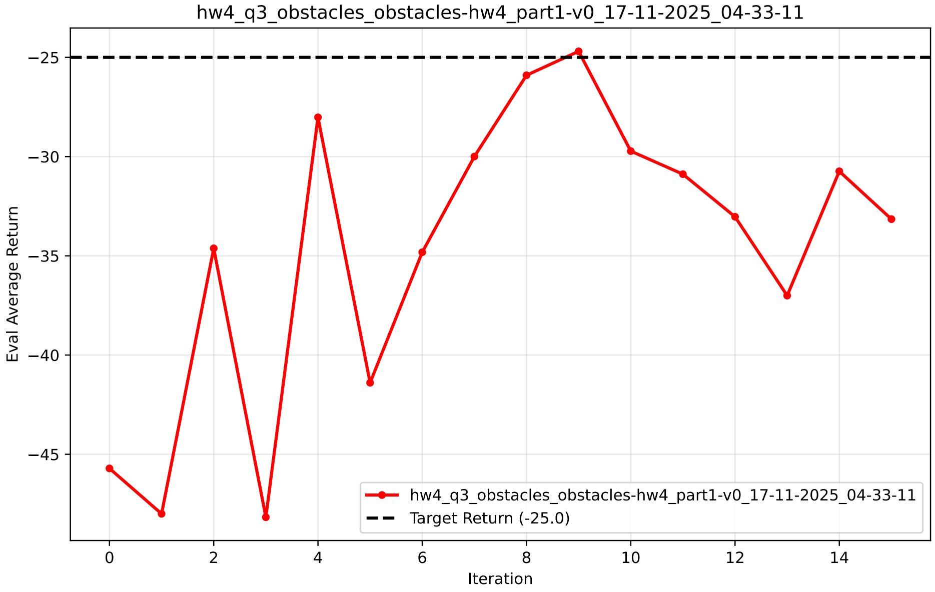
2 Problem 2: Action Selection [4pts]

Plot 1

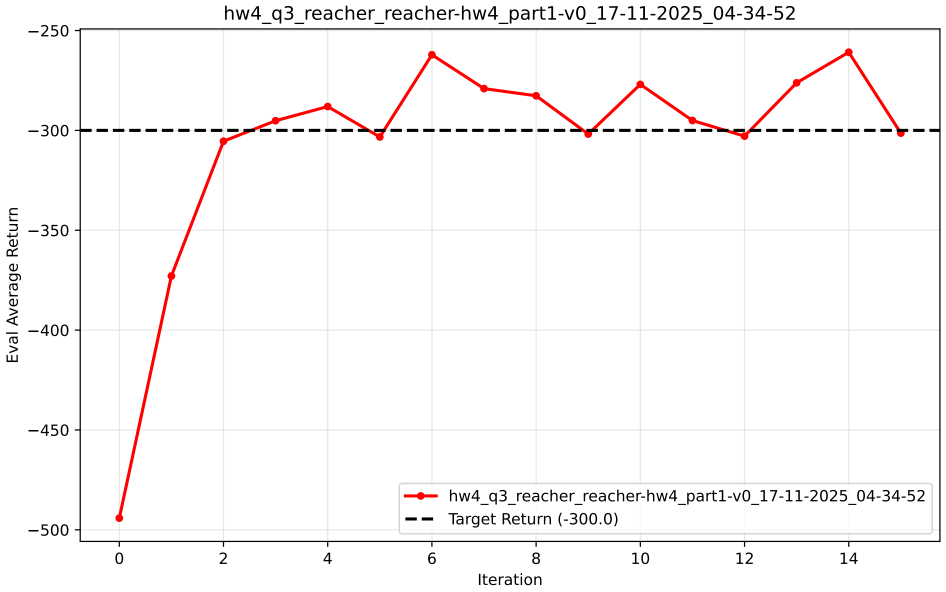


3 Problem 3: Iterative Model Training [3pts]

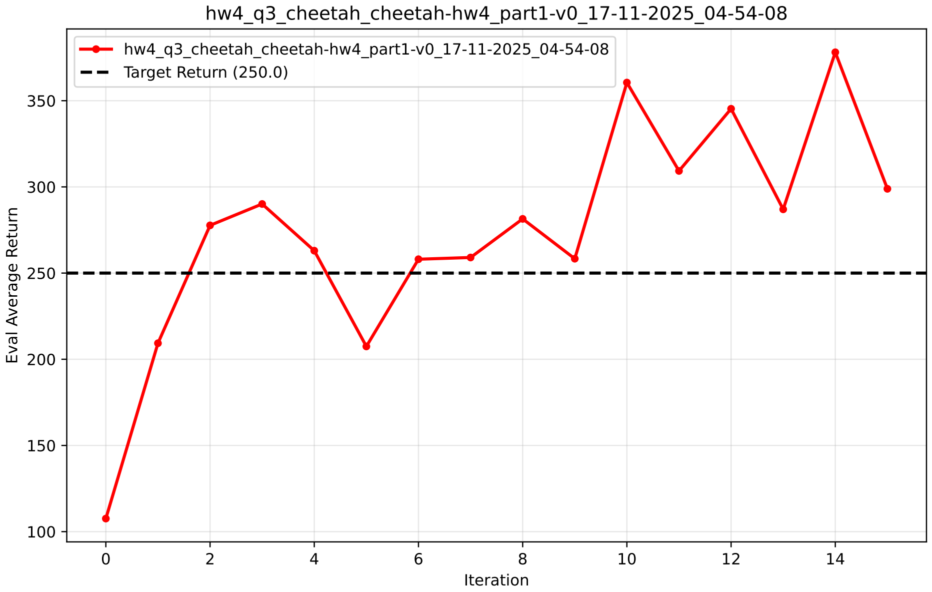
Plot 1



Plot 2

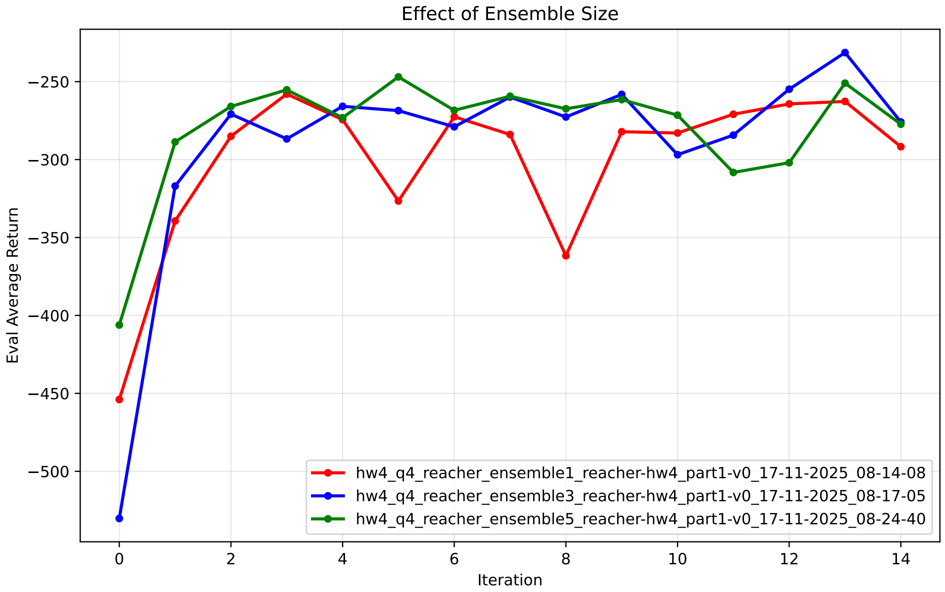

hw4_q3_reacher_reacher-hw4_part1-v0_17-11-2025_04-34-52
Eval Average Return
−250
−300
−350
−400
−450
−500
0
2
4
6
8
10
12
14
Iteration
hw4_q3_reacher_reacher-hw4_part1-v0_17-11-2025_04-34-52
Target Return (-300.0)

Plot 3


hw4_q3_cheetah_cheetah-hw4_part1-v0_17-11-2025_04-54-08
Eval Average Return
350
300
250
200
150
100
0
2
4
6
8
10
12
14
Iteration
hw4_q3_cheetah_cheetah-hw4_part1-v0_17-11-2025_04-54-08
Target Return (250.0)

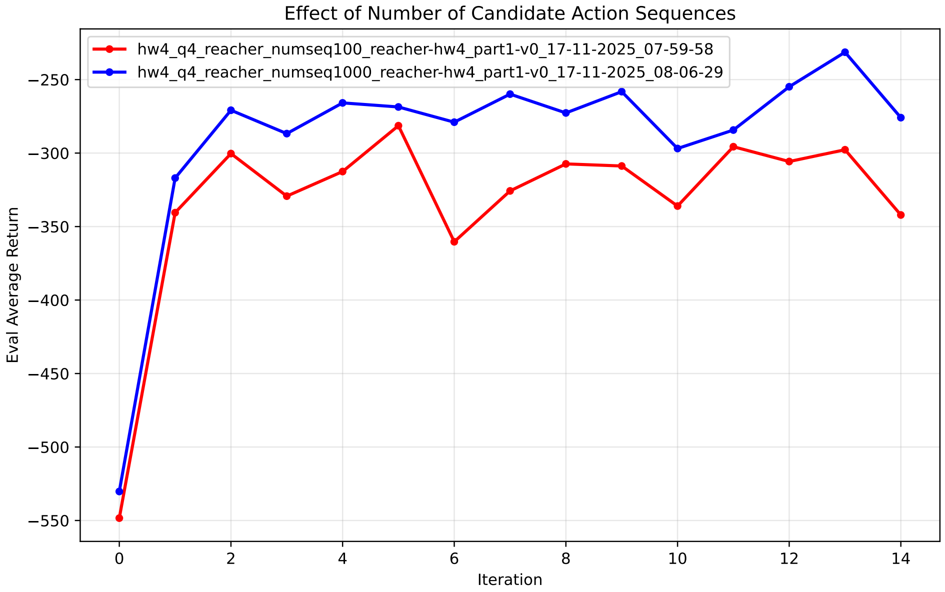
4 Problem 4: Hyper-parameter Comparison [4pts]

Plot 1



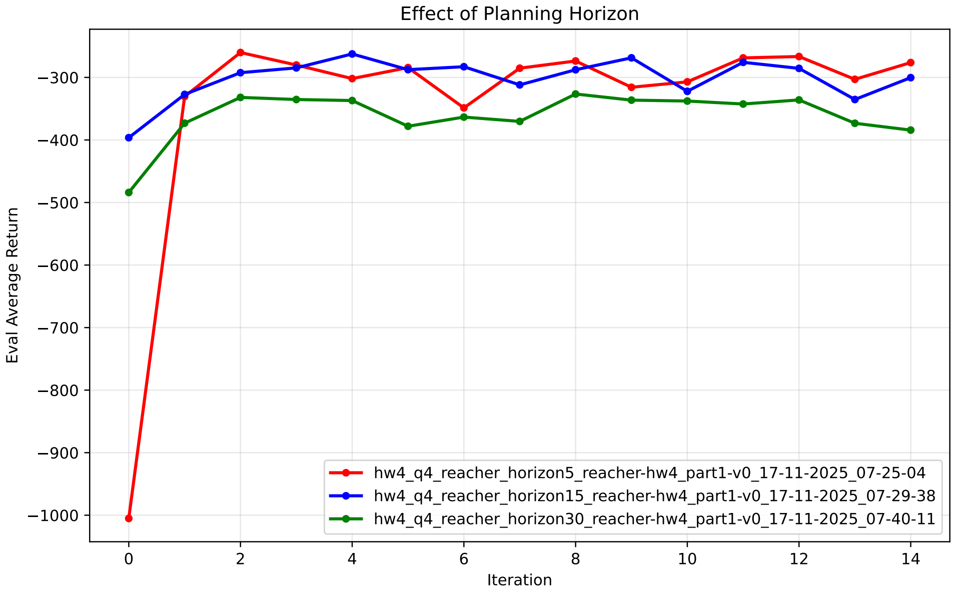
Ensemble size: While the initial performance of the larger ensemble is better, all converge to similar performance (∼-270). Ensemble 1 shows highest variance, while 3 and 5 are more stable.

Plot 2



Number of Candidate Action Sequences: More action sequences (N=1000) consistently outperforms fewer sequences (N=100) throughout training, indicating that higher N provides better coverage of action space, and thus action selection.

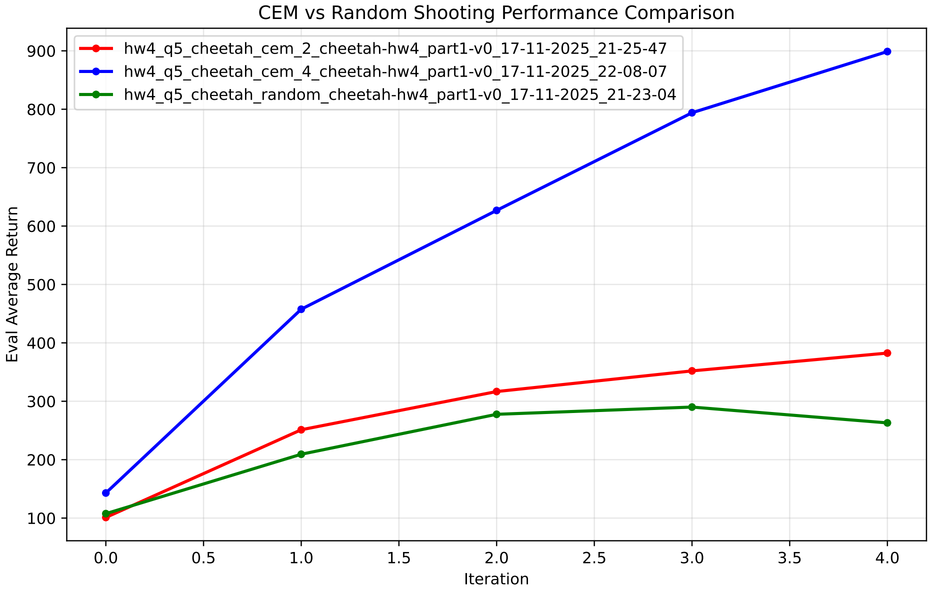
Plot 3



Planning Horizon: Shorter horizons (H=5, H=15) perform better than the longest horizon (H=30), indicating that very long horizons accumulate model prediction errors, degrading planning quality.

5 Problem 5: CEM (Bonus) [2.5pts]

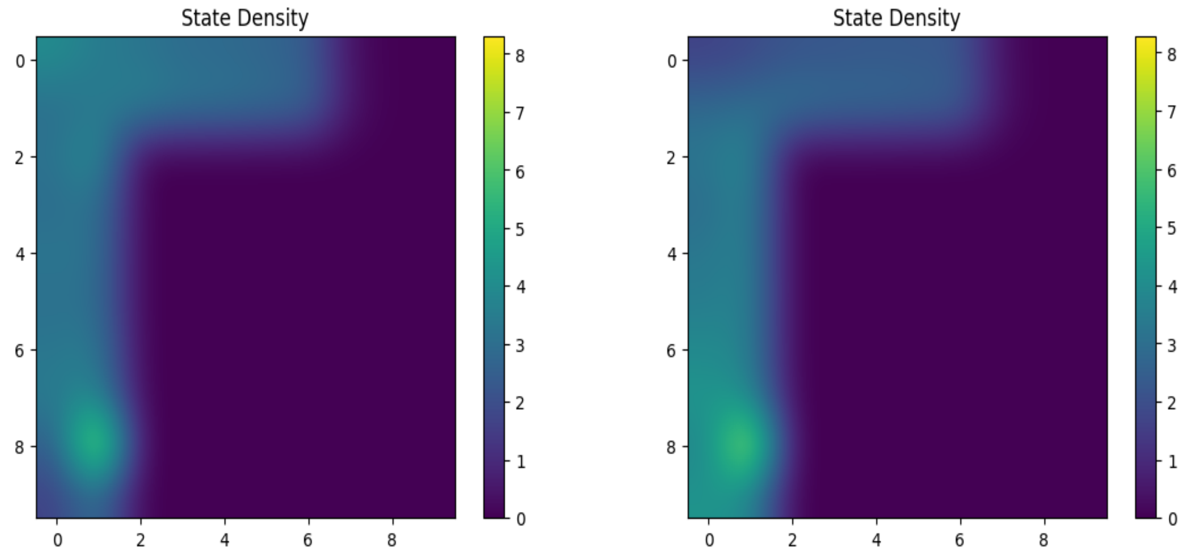
Plot 1



CEM: CEM outperforms random shooting, and increasing CEM iterations from 2 to 4 more than doubles performance, indicating that iterative refinement of the sampling distribution is useful for action optimsation.

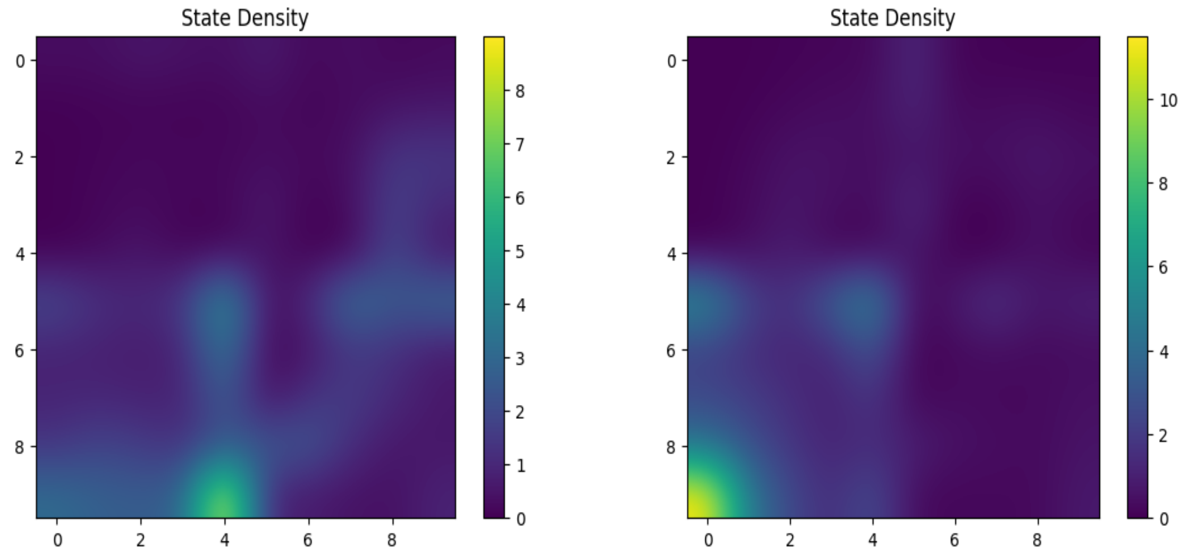
6 Problem 6: Exploration (Bonus) [2.5pts]

Plot 1



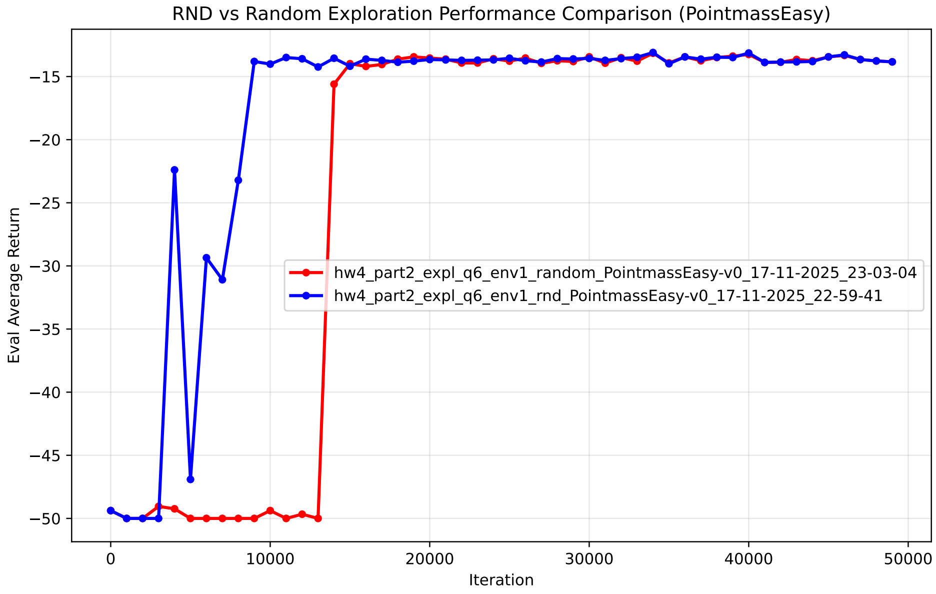
State Density Heatmaps in PointmassEasy env (Left: RND, Right: Random)

Plot 2



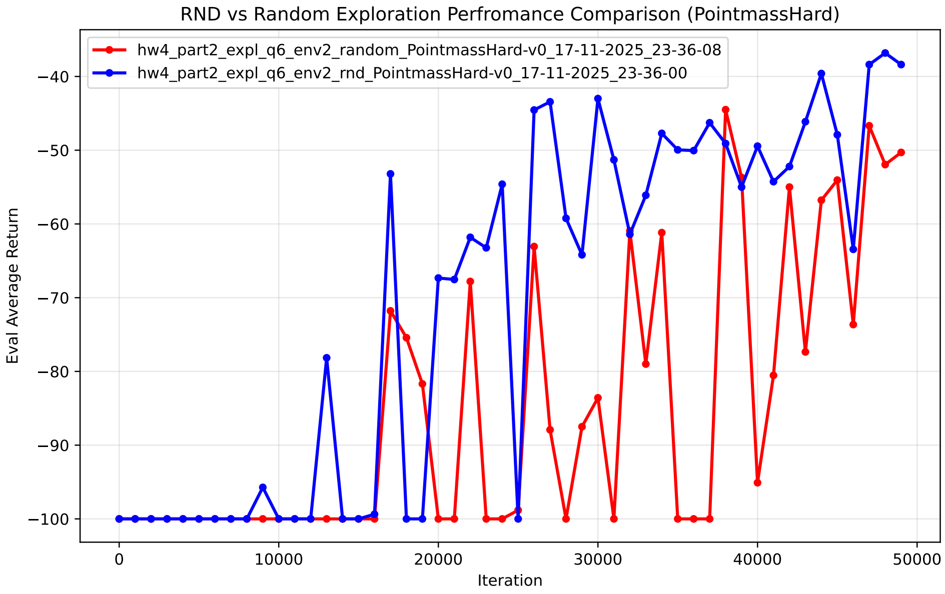
State Density Heatmaps in PointmassHard env (Left: RND, Right: Random)

Plot 3



PointmassEasy Performance Curves: RND converges faster than random exploration.

Plot 4



PointmassHard Performance Curves: The difference between performance is more pronounced in the hard env: not only does RND converge faster, it also performs better on average than random exploration.