

Code Explanation:

P_learner.py – Class for creating the Q function with zero values for any (s,a) pair.

q_learning.py – Class where the q learning algorithm is written, it implements the epsilon greedy decision, eligibility tracing, evaluating a given policy (by sampling)

plotter.py – Class for running the q learning algorithm with different hyper-parameters, saves the results and plot the graphs.

Play_simulation_agent.py – Given a policy (saved or trained) it will run the game and plot the location, action and reward.

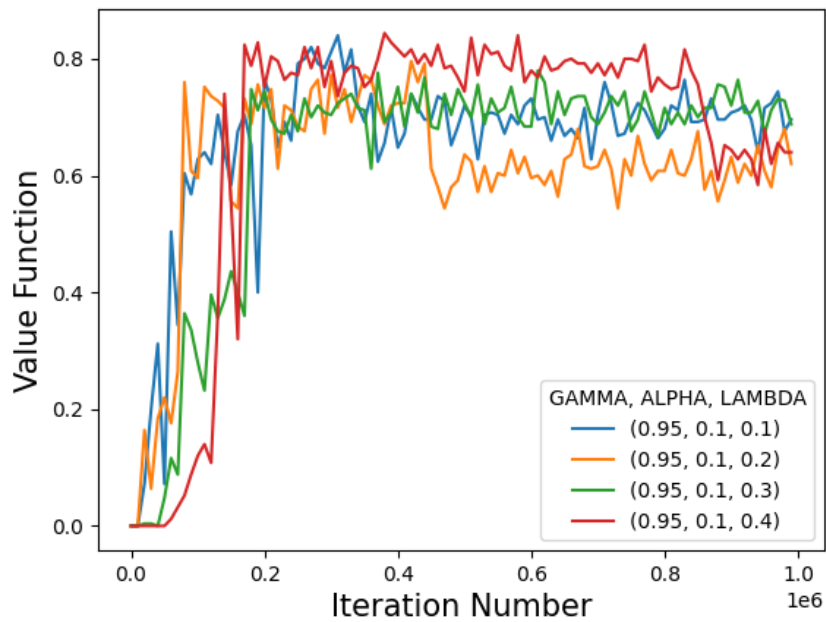
Graphs:

ALPHA = 0.1

LAMBDA = 0.1, 0.2, 0.3, 0.4

GAMMA = 0.95

Policy Evaluation Mean Over Iterations

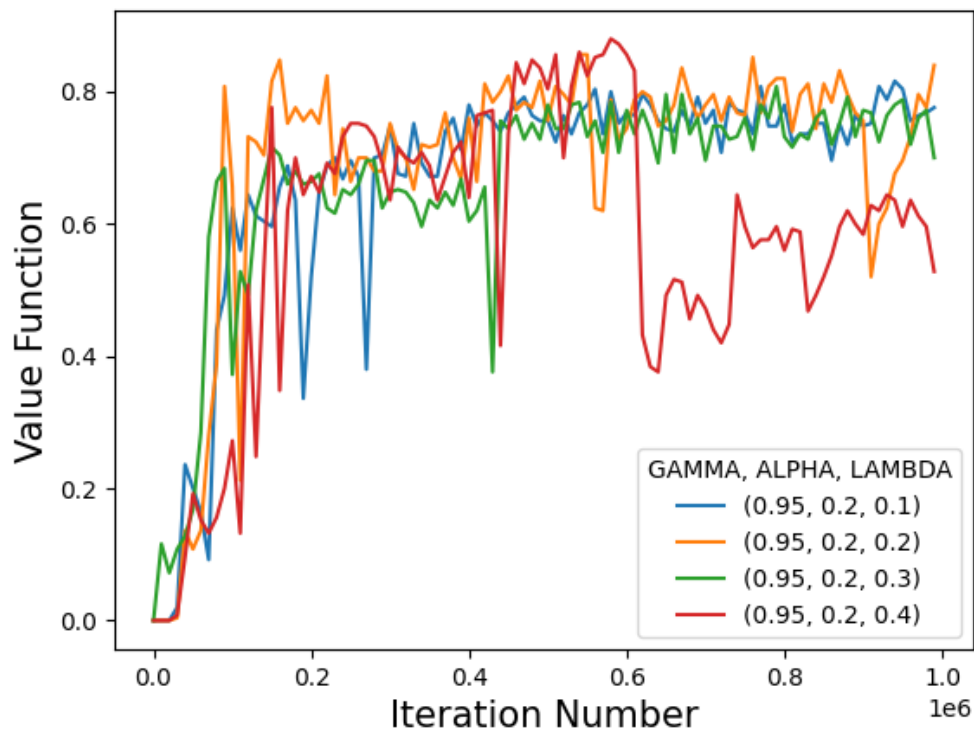


ALPHA = 0.2

LAMBDA = 0.1, 0.2, 0.3, 0.4

GAMMA = 0.95

Policy Evaluation Mean Over Iterations

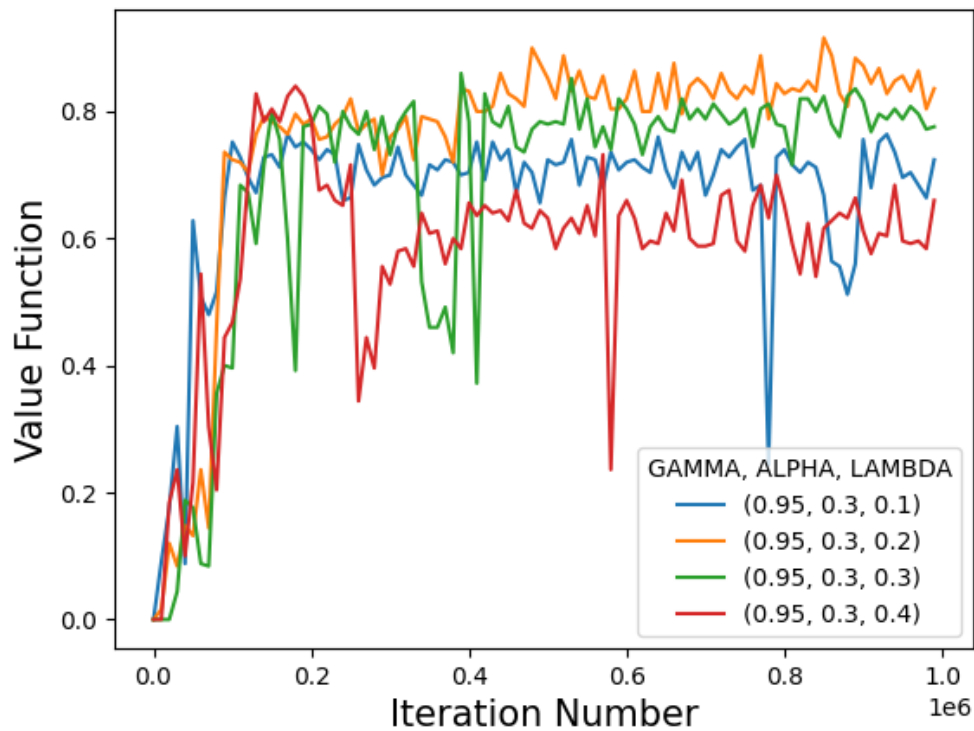


ALPHA = 0.3

LAMBDA = 0.1, 0.2, 0.3, 0.4

GAMMA = 0.95

Policy Evaluation Mean Over Iterations



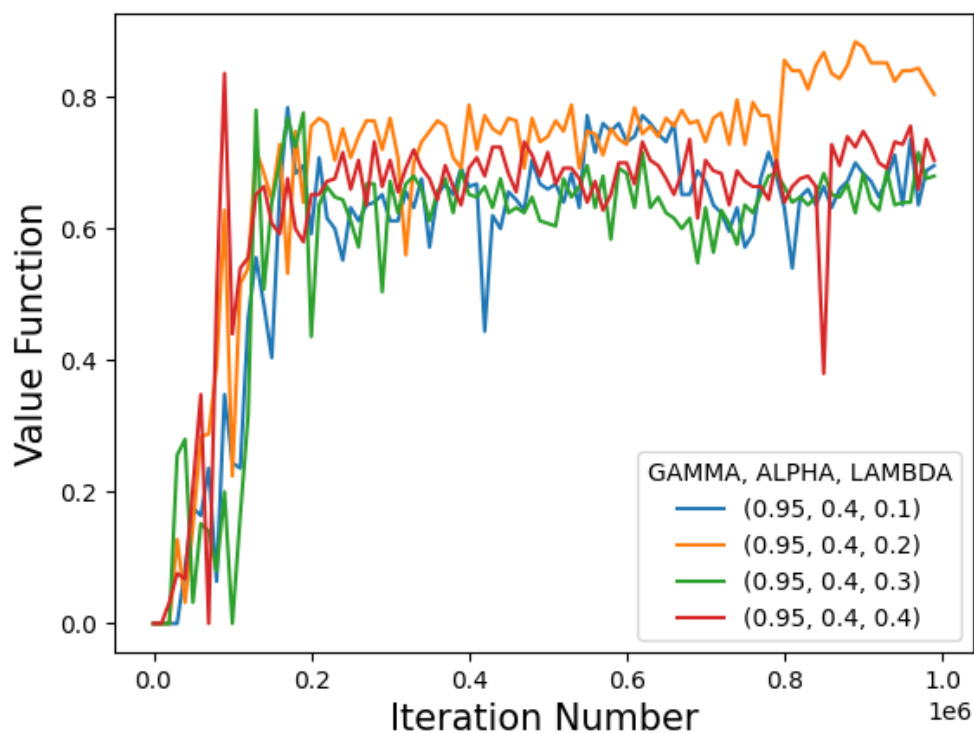
- **Best policy was achieved by using**
ALPHA = 0.3 and LAMBDA = 0.2

ALPHA = 0.4

LAMBDA = 0.1, 0.2, 0.3, 0.4

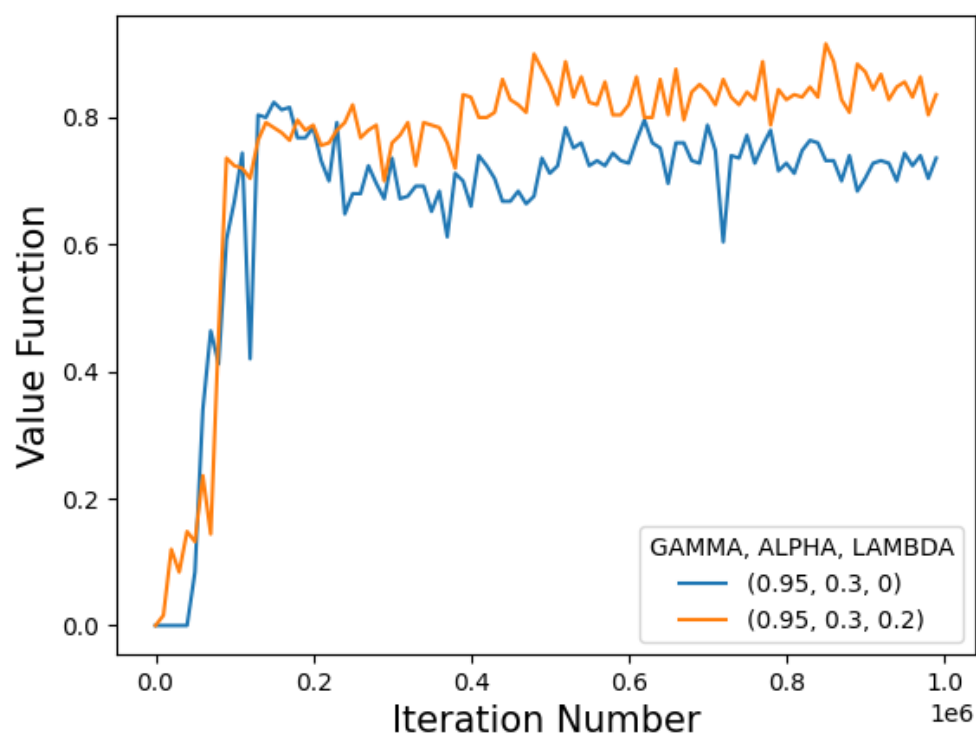
GAMMA = 0.95

Policy Evaluation Mean Over Iterations



Comparing best parameters with and without eligibility traces:

Policy Evaluation Mean Over Iterations

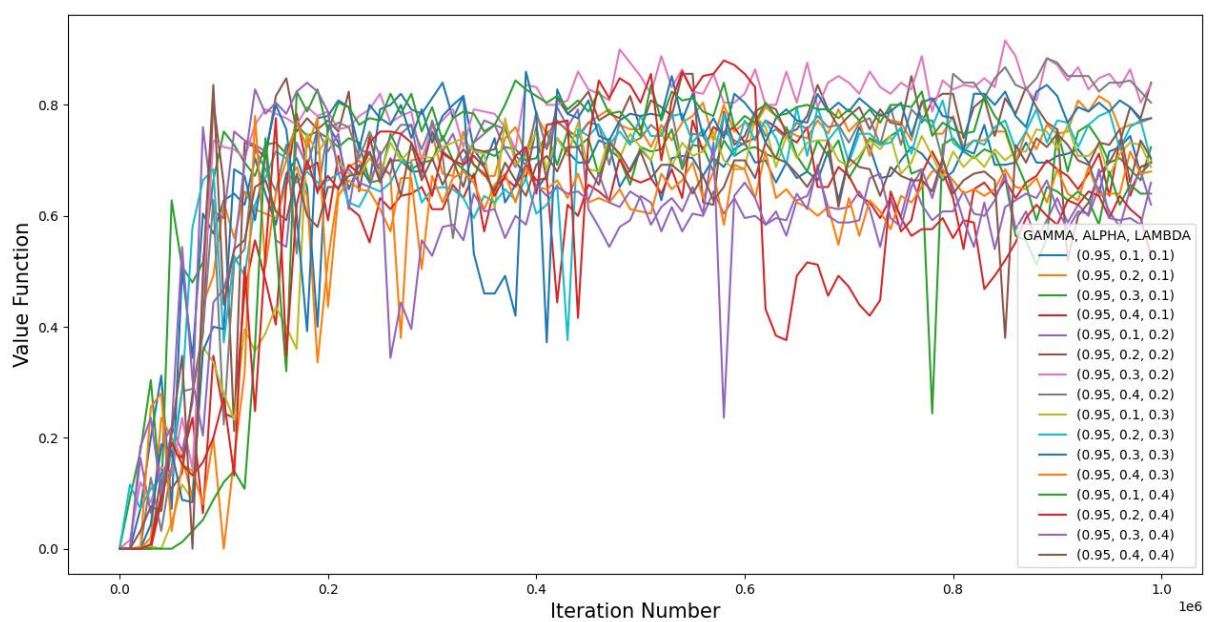


Yellow line is Q-learning with eligibility trace and LAMBDA = 0.2 and ALPHA = 0.3

Blue line is Q-learning with no eligibility trace (so LAMBDA = 0) and ALPHA = 0.3

All graphs together:

Policy Evaluation Mean Over Iterations



Simulations:

First run:

Number of steps: 42

Total reward: 1.0

0 (0, 0) b'S' (7, 7) __

1 (0, 0) b'S' (7, 7) UP 0.0

2 (0, 0) b'S' (7, 7) UP 0.0

3 (0, 1) b'F' (7, 7) UP 0.0

4 (0, 2) b'F' (7, 7) RIGHT 0.0

5 (0, 2) b'F' (7, 7) RIGHT 0.0

6 (0, 3) b'F' (7, 7) RIGHT 0.0

7 (1, 3) b'F' (7, 7) RIGHT 0.0

8 (0, 3) b'F' (7, 7) UP 0.0

9 (0, 4) b'F' (7, 7) RIGHT 0.0

10 (0, 4) b'F' (7, 7) RIGHT 0.0

11 (0, 4) b'F' (7, 7) RIGHT 0.0

12 (0, 4) b'F' (7, 7) RIGHT 0.0
13 (0, 4) b'F' (7, 7) RIGHT 0.0
14 (0, 5) b'F' (7, 7) RIGHT 0.0
15 (0, 6) b'F' (7, 7) RIGHT 0.0
16 (0, 7) b'F' (7, 7) RIGHT 0.0
17 (1, 7) b'F' (7, 7) DOWN 0.0
18 (1, 7) b'F' (7, 7) RIGHT 0.0
19 (0, 7) b'F' (7, 7) RIGHT 0.0
20 (1, 7) b'F' (7, 7) DOWN 0.0
21 (1, 7) b'F' (7, 7) RIGHT 0.0
22 (2, 7) b'F' (7, 7) RIGHT 0.0
23 (2, 7) b'F' (7, 7) DOWN 0.0
24 (2, 7) b'F' (7, 7) DOWN 0.0
25 (3, 7) b'F' (7, 7) DOWN 0.0
26 (3, 7) b'F' (7, 7) RIGHT 0.0
27 (3, 7) b'F' (7, 7) RIGHT 0.0
28 (4, 7) b'F' (7, 7) RIGHT 0.0
29 (5, 7) b'F' (7, 7) RIGHT 0.0
30 (4, 7) b'F' (7, 7) RIGHT 0.0
31 (5, 7) b'F' (7, 7) RIGHT 0.0
32 (4, 7) b'F' (7, 7) RIGHT 0.0
33 (4, 7) b'F' (7, 7) RIGHT 0.0
34 (5, 7) b'F' (7, 7) RIGHT 0.0
35 (5, 7) b'F' (7, 7) RIGHT 0.0
36 (5, 7) b'F' (7, 7) RIGHT 0.0
37 (6, 7) b'F' (7, 7) RIGHT 0.0
38 (6, 7) b'F' (7, 7) RIGHT 0.0
39 (5, 7) b'F' (7, 7) RIGHT 0.0
40 (6, 7) b'F' (7, 7) RIGHT 0.0
41 (6, 7) b'F' (7, 7) RIGHT 0.0
42 (7, 7) b'G' (7, 7) RIGHT 1.0

Second run:

Number of steps: 57

Total reward: 1.0

0 (0, 0) b'S' (8, 8) __

1 (0, 0) b'S' (7, 7) UP 0.0

2 (0, 1) b'F' (7, 7) UP 0.0

3 (0, 1) b'F' (7, 7) RIGHT 0.0

4 (0, 2) b'F' (7, 7) RIGHT 0.0

5 (0, 3) b'F' (7, 7) RIGHT 0.0

6 (0, 3) b'F' (7, 7) RIGHT 0.0

7 (0, 3) b'F' (7, 7) RIGHT 0.0

8 (0, 4) b'F' (7, 7) RIGHT 0.0

9 (0, 4) b'F' (7, 7) RIGHT 0.0

10 (0, 5) b'F' (7, 7) RIGHT 0.0

11 (1, 5) b'F' (7, 7) RIGHT 0.0

12 (2, 5) b'F' (7, 7) RIGHT 0.0

13 (2, 4) b'F' (7, 7) UP 0.0

14 (2, 5) b'F' (7, 7) RIGHT 0.0

15 (2, 6) b'F' (7, 7) UP 0.0

16 (2, 7) b'F' (7, 7) RIGHT 0.0

17 (2, 7) b'F' (7, 7) DOWN 0.0

18 (2, 6) b'F' (7, 7) DOWN 0.0

19 (1, 6) b'F' (7, 7) RIGHT 0.0

20 (2, 6) b'F' (7, 7) DOWN 0.0

21 (1, 6) b'F' (7, 7) RIGHT 0.0

22 (1, 7) b'F' (7, 7) DOWN 0.0

23 (1, 7) b'F' (7, 7) RIGHT 0.0

24 (0, 7) b'F' (7, 7) RIGHT 0.0
25 (1, 7) b'F' (7, 7) DOWN 0.0
26 (2, 7) b'F' (7, 7) RIGHT 0.0
27 (2, 7) b'F' (7, 7) DOWN 0.0
28 (2, 6) b'F' (7, 7) DOWN 0.0
29 (1, 6) b'F' (7, 7) RIGHT 0.0
30 (1, 5) b'F' (7, 7) DOWN 0.0
31 (2, 5) b'F' (7, 7) RIGHT 0.0
32 (1, 5) b'F' (7, 7) UP 0.0
33 (1, 6) b'F' (7, 7) RIGHT 0.0
34 (2, 6) b'F' (7, 7) DOWN 0.0
35 (3, 6) b'F' (7, 7) RIGHT 0.0
36 (2, 6) b'F' (7, 7) RIGHT 0.0
37 (2, 7) b'F' (7, 7) RIGHT 0.0
38 (2, 7) b'F' (7, 7) DOWN 0.0
39 (2, 7) b'F' (7, 7) DOWN 0.0
40 (3, 7) b'F' (7, 7) DOWN 0.0
41 (3, 7) b'F' (7, 7) RIGHT 0.0
42 (3, 7) b'F' (7, 7) RIGHT 0.0
43 (4, 7) b'F' (7, 7) RIGHT 0.0
44 (4, 7) b'F' (7, 7) RIGHT 0.0
45 (4, 7) b'F' (7, 7) RIGHT 0.0
46 (5, 7) b'F' (7, 7) RIGHT 0.0
47 (4, 7) b'F' (7, 7) RIGHT 0.0
48 (4, 7) b'F' (7, 7) RIGHT 0.0
49 (3, 7) b'F' (7, 7) RIGHT 0.0
50 (3, 7) b'F' (7, 7) RIGHT 0.0
51 (4, 7) b'F' (7, 7) RIGHT 0.0
52 (4, 7) b'F' (7, 7) RIGHT 0.0
53 (5, 7) b'F' (7, 7) RIGHT 0.0
54 (6, 7) b'F' (7, 7) RIGHT 0.0

55 (6, 7) b'F' (7, 7) RIGHT 0.0
56 (6, 7) b'F' (7, 7) RIGHT 0.0
57 (7, 7) b'G' (7, 7) RIGHT 1.0

Third run:

Number of steps: 44

Total reward: 1.0

0 (0, 0) b'S' (8, 8) __
1 (0, 1) b'F' (7, 7) UP 0.0
2 (1, 1) b'F' (7, 7) RIGHT 0.0
3 (0, 1) b'F' (7, 7) RIGHT 0.0
4 (1, 1) b'F' (7, 7) RIGHT 0.0
5 (0, 1) b'F' (7, 7) RIGHT 0.0
6 (0, 1) b'F' (7, 7) RIGHT 0.0
7 (0, 1) b'F' (7, 7) RIGHT 0.0
8 (0, 1) b'F' (7, 7) RIGHT 0.0
9 (1, 1) b'F' (7, 7) RIGHT 0.0
10 (2, 1) b'F' (7, 7) RIGHT 0.0
11 (2, 0) b'F' (7, 7) UP 0.0
12 (1, 0) b'F' (7, 7) UP 0.0
13 (0, 0) b'S' (7, 7) RIGHT 0.0
14 (0, 0) b'S' (7, 7) UP 0.0
15 (0, 0) b'S' (7, 7) UP 0.0
16 (0, 0) b'S' (7, 7) UP 0.0
17 (0, 0) b'S' (7, 7) UP 0.0
18 (0, 0) b'S' (7, 7) UP 0.0
19 (0, 0) b'S' (7, 7) UP 0.0
20 (0, 0) b'S' (7, 7) UP 0.0
21 (0, 0) b'S' (7, 7) UP 0.0

22 (0, 1) b'F' (7, 7) UP 0.0
23 (0, 2) b'F' (7, 7) RIGHT 0.0
24 (0, 3) b'F' (7, 7) RIGHT 0.0
25 (0, 4) b'F' (7, 7) RIGHT 0.0
26 (0, 4) b'F' (7, 7) RIGHT 0.0
27 (1, 4) b'F' (7, 7) RIGHT 0.0
28 (1, 5) b'F' (7, 7) UP 0.0
29 (1, 6) b'F' (7, 7) RIGHT 0.0
30 (1, 5) b'F' (7, 7) DOWN 0.0
31 (1, 6) b'F' (7, 7) RIGHT 0.0
32 (2, 6) b'F' (7, 7) DOWN 0.0
33 (3, 6) b'F' (7, 7) RIGHT 0.0
34 (3, 7) b'F' (7, 7) RIGHT 0.0
35 (4, 7) b'F' (7, 7) RIGHT 0.0
36 (4, 7) b'F' (7, 7) RIGHT 0.0
37 (5, 7) b'F' (7, 7) RIGHT 0.0
38 (4, 7) b'F' (7, 7) RIGHT 0.0
39 (5, 7) b'F' (7, 7) RIGHT 0.0
40 (4, 7) b'F' (7, 7) RIGHT 0.0
41 (5, 7) b'F' (7, 7) RIGHT 0.0
42 (5, 7) b'F' (7, 7) RIGHT 0.0
43 (6, 7) b'F' (7, 7) RIGHT 0.0
44 (7, 7) b'G' (7, 7) RIGHT 1.0