

Team 8

value iteration:

Matrices till convergence:

```
0.000000 0.000000 8.000000 0.000000
0.000000 -8.000000 0.000000 0.000000
0.000000 0.000000 0.000000 0.000000
-----
-0.400000 5.200000 8.000000 6.000000
-0.400000 -8.000000 0.000000 -0.400000
-0.400000 -0.400000 -0.400000 -0.400000
-----
3.680000 5.720000 8.000000 6.560000
-0.800000 -8.000000 0.000000 4.320000
-0.800000 -0.800000 -0.800000 -0.800000
-----
4.464000 5.772000 8.000000 7.088000
1.664000 -8.000000 0.000000 5.712000
-1.200000 -1.200000 -1.200000 2.896000
-----
4.830400 5.777200 8.000000 7.280000
2.537600 -8.000000 0.000000 6.412800
0.691200 -1.600000 1.676800 4.339200
-----
4.958560 5.777720 8.000000 7.369280
2.918080 -8.000000 0.000000 6.706560
1.539200 -0.018560 3.406720 5.331840
-----
5.009840 5.777772 8.000000 7.407584
3.058656 -8.000000 0.000000 6.836736
2.086528 1.523520 4.546816 5.839104
-----
5.029067 5.777777 8.000000 7.424432
3.113738 -8.000000 0.000000 6.893414
2.407930 2.589805 5.180646 6.107981
-----
5.036502 5.777778 8.000000 7.431785
3.134627 -8.000000 0.000000 6.918228
2.590764 3.203497 5.522514 6.243594
-----
5.039335 5.777778 8.000000 7.435001
```

3.142665 -8.000000 0.000000 6.929073
2.735337 3.538361 5.699378 6.311193

Expected Reward:

The Final Expected Reward is 2.735337

Optimal Path from start to end:

Current State: 2 0
Action to take: Right

Current State: 2 1
Action to take: Right

Current State: 2 2
Action to take: Right

Current State: 2 3
Action to take: Above

Current State: 1 3
Action to take: Above

Current State: 0 3
Action to take: Left

Current State: 0 2

Linear Programming:

values of x:

State, Action pair	Value of X
1,1	0
1,2	0
1,3	0

1,4	0.1217656012
2,1	0
2,2	0
2,3	0
2,4	0.10823609
4,1	0
4,2	0
4,3	0.987654321
4,4	0
5,1	0.1369863014
5,2	0
5,3	0
5,4	0
8,1	1.111111111
8,2	0
8,3	0
8,4	0
9,1	0
9,2	0
9,3	0
9,4	1.111111111
10,1	0
10,2	0
10,3	0
10,4	0.987654321
11,1	0
11,2	0

11,3	0
11,4	1.1111111111
12,1	0.987654321
12,2	0
12,3	0
12,4	0

Expected Reward:

6.5401657365

Description of why the rewards didn't match:

The rewards didn't match. The reward we got from value iteration is 2.735337, and the reward we got from LP is 6.5401657365. The reason why these don't match is because the reward in each case is calculated differently. In value iteration the reward corresponds to the utility (maximum) possible from the start state. In LP, the reward is the collection (sum) of all the possible rewards (R) taken as many times as possible (X).