

Team 49

Value Iteration:

Matrices till convergence:

W	W	49.000	W
0.000	0.000	0.000	0.000
0.000	-49.000	W	0.000
0.000	0.000	0.000	0.000

<----->

W	W	49.000	W
-2.450	-2.450	36.750	-2.450
-2.450	-49.000	W	-2.450
-2.450	-2.450	-2.450	-2.450

<----->

W	W	49.000	W
-4.900	21.805	36.260	26.460
-4.900	-49.000	W	-4.900
-4.900	-4.900	-4.900	-4.900

<----->

W	W	49.000	W
14.014	23.839	41.577	28.714
-7.350	-49.000	W	17.738
-7.350	-7.350	-7.350	-7.350

<----->

W	W	49.000	W
17.287	28.295	42.005	35.456
3.126	-49.000	W	24.069
-9.800	-9.800	-9.800	10.270

<----->

W	W	49.000	W
22.227	29.084	43.125	37.107
6.792	-49.000	W	30.729
-1.909	-12.250	3.806	16.852

<----->

W W 49.000 W
23.719 30.058 43.369 38.834
11.111 -49.000 W 33.381
1.568 -5.530 11.793 24.199

<----->

W W 49.000 W
25.080 30.351 43.639 39.467
12.736 -49.000 W 35.293
6.043 1.531 19.268 27.854

<----->

W W 49.000 W
25.612 30.596 43.732 39.937
13.987 -49.000 W 36.182
8.496 8.217 23.687 30.497

<----->

W W 49.000 W
25.987 30.695 43.803 40.147
14.539 -49.000 W 36.736
10.411 12.421 26.685 31.914

<----->

W W 49.000 W
26.159 30.762 43.834 40.281
14.894 -49.000 W 37.015
11.464 15.240 28.418 32.799

<----->

W W 49.000 W
26.265 30.794 43.854 40.347
15.066 -49.000 W 37.178
12.378 16.908 29.473 33.284

<-----Iteration Stopped----->

Results for Delta = 0:

W	W	49.000	W
26.385	30.835	43.877	40.431
15.287	-49.000	W	37.369
16.101	19.265	30.861	33.923

policy:

```
-----
East East North West
North ---- --- North
East East East North
```

Expected Reward:

The Final Expected Reward is 12.378
(16.101 if delta=0)

Optimal Path from start to end:

```
-----
Current State: 3 0
Action to take: East
-----
Current State: 3 1
Action to take: East
-----
Current State: 3 2
Action to take: East
-----
Current State: 3 3
Action to take: North
-----
Current State: 2 3
Action to take: North
-----
Current State: 1 3
Action to take: West
-----
Current State: 1 2
Action to take: North
-----
Final State: 0 2
```

Linear Programming :

Values of x:

State,Action Pair	Value of x
3,5	0.864702570838806
5,1	0
5,2	0
5,3	0
5,4	0.121765601217656
6,1	0
6,2	0
6,3	0
6,4	0.228333669254417
7,1	1.08087821354851
7,2	0
7,3	0
7,4	0
8,1	0
8,2	0
8,3	1.12276409768122
8,4	0
9,1	0.136986301369863
9,2	0
9,3	0
9,4	0
10,5	0.135297429161193
12,1	1.12799983319781

12,2	0
12,3	0
12,4	0
13,1	0
13,2	0
13,3	0
13,4	1.11111111111111
14,1	0
14,2	0
14,3	0
14,4	0.987654320987654
15,1	0
15,2	0
15,3	0
15,4	1.11111111111111
16,1	0.987654320987654
16,2	0
16,3	0
16,4	0

Expected Reward:

16.10101842

Description of why the records match/don't match:

In both of the methods, value iteration and linear programming, the aim is to maximize utility/reward. Hence, both these methods will end up achieving the same results if the precision is high and accurate.

In value iteration method, the utility in the start state gives the reward of selecting the best path possible from the start state to the terminal states. In linear programming, the summation of reward*x gives the utility for each state,action pair. Thus, both these methods gives the same probabilities, policy and path.

The corresponding values obtained by value iteration and linear programming will match if the value of the delta is small enough. Larger deltas will provide less iterations and gives the approximate value of utilities at all the stages. So, if we use delta large enough (not

near 0) , then , results of both these methods do not match , however, if $\delta = 0$, then the results are bound to match.