AI Assignment 2 Report

Team 30

Part B.

Initial Rewards:

3.0	0.0	0.0	30.0
0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0
0.0	0.0	-6.0	0.0

1. A. Discount Factor = 0.1, Step Cost = -3.0

Final Utility Map:

3.0	0.0	-0.637	30.0
-2.821	-3.291	-3.088	-0.637
-3.291	-3.328	0.0	-3.113
-3.328	-3.33	-6.0	-3.328

Final Policy:

F	-	E	F
N	W	E	N
N	N	-	N
N	W	F	E

(F - End States, W - Wall, N - North, S - South, E - East, W - West)

Number of Iterations till convergence = 3

The algorithm converges rapidly, and the final utilities are all mostly negative. This is because our discount factor is much less, and as such, the farther rewards are given much less favour. So, the immediate negative step cost dominates the utility values.

B. Discount Factor = 0.99, Step Cost = -3.0

Final Utility Map:

3.0	0.0	25.439	30.0
11.841	17.264	21.827	25.439
8.823	12.811	0.0	21.382
5.181	7.057	-6.0	14.806

Final Policy:

F	-	E	F
E	E	E	N
E	N	-	N
N	N	F	N

(F - End States, W - Wall, N - North, S - South, E - East, W - West)

Number of Iterations till convergence = 16

Since the discount factor ≈ 1, there is not much decay in the rewards of the farther states. As such, the policy always tends to direct us to the goal state at (0,3), as it has a very high value.

2. A. Step Cost = 30, Discount Factor = 0.99

Final Utility Map:

3.0	0.0	1525.916	30.0
1525.916	1525.916	1525.916	1525.883
1525.916	1525.916	0.0	1525.881
1525.916	1525.916	-6.0	1525.845

Final Policy:

F	-	W	F
S	E	W	S
E	N	-	N
N	W	F	E

(F - End States, W - Wall, N - North, S - South, E - East, W - West)

Number of Iterations till convergence = 70

Since the step cost is very large, larger than the reward of the best goal state, the utilities are skewed and the policy sends us in circles. The cell at (1,3) has the highest utility, so it assumes that we cannot move from there.

B. Step Cost = -6, Discount Factor = 0.99

Final Utility Map:

3.0	0.0	21.293	30.0
-2.062	5.705	14.398	21.293
-8.782	-2.621	0.0	13.546
-15.551	-10.235	-6.0	4.588

Final Policy:

F	-	E	F
E	E	E	N
N	N	-	N
N	N	F	N

(F - End States, W - Wall, N - North, S - South, E - East, W - West)

Number of Iterations till convergence = 13

Here, the negative factor of the step cost initially outweighs all positive benefits, as they occur much farther away. However, as we approach the goal state at (0,3), the utilities become larger as it has a very high reward. So, the policy directs us towards that state.

C. Step Cost = -7.5, Discount Factor = 0.99

Final Utility Map:

3.0	0.0	19.22	30.0
-5.69	-0.025	10.683	19.22
-14.417	-9.93	0.0	9.629
-22.611	-14.689	-6.0	-0.519

Final Policy:

F	-	E	F
N	E	E	N
N	N	-	N
N	E	F	N

('F' - End States, '-' - Wall, 'N' - North, 'S' - South, 'E' - East, 'W' - West)

Number of Iterations till convergence = 17

This case is also almost exact same as 2.b. except for the fact that, since the step cost is higher, the negative values are present till farther along. As such, for example at (1, 0), the policy directs it to the north as opposed to the east as in the previous case.

D. Step Cost = -30, Discount Factor = 0.99

Final Utility Map:

3.0	0.0	-11.854	30.0
-39.012	-76.551	-44.977	-11.854
-76.551	-83.399	0.0	-49.051
-83.399	-47.67	-6.0	-43.942

Final Policy:

F	-	E	F
N	W	E	N
N	S	-	N
E	E	F	W

('F' - End States, '-' - Wall, 'N' - North, 'S' - South, 'E' - East, 'W' - West)

Number of Iterations till convergence = 7

Here, the unit step cost is extremely high, and as such, even the negative termination state is a valid place for the agent. So, the agent directs itself towards it's closest termination state, as the step cost is high for it to even consider the farther states.

Part C.

The LP solver in excel does not have discount factor, so there will be difference of magnitude <= step reward between the result of solver and VI algorithm. But if discount factor is kept as 1 in VI algorithm, both give same result.