AI Report Question 2
Team Number 40 (X = 40)

$$U(s_0 s_1 \ldots s_n) = \sum_{i=0}^{n} \gamma^i R(s_i)$$

Question 2.1.a:

Discount Factor = 0.1
Step Cost = -X/10 (-4)

Policy obtained from Value Iteration Algorithm:
['-', '-', 'E', '-']
['N', 'W', 'E', 'N']
['N', 'N', '-', 'N']
['N', 'W', '-', 'N']
(where [(0,0),(0,3),(3,2)] are terminal states and [(0,1),(2,2) are walls)

Utilities:
4.0 0.0 2.827 40.0
-0.085 -0.415 -0.147 2.827
-0.415 -0.442 0.0 -0.177
-0.442 -0.444 -8.0 -0.499

Observation:
As value of discount factor is small, the agent gives less preference to future values along the path.
So it gives more preference to current states which is easily observed as in states near (0,3), the agent tend to go towards (0,3).

The values of the utilities are negative showing that the agent will tend to go to terminal state as soon as possible, as shown in block (1,0) which tries to go N instead of E.

Question 2.1.b:

Discount Factor = 0.99

Step Cost = -X/10 (-4)

Policy obtained from Value Iteration Algorithm:
['-', '-', 'E', '-']
['E', 'E', 'E', 'N']
['E', 'N', '-', 'N']
['N', 'N', '-', 'N']
(where [(0,0),(0,3),(3,2)] are terminal states and [(0,1),(2,2) are walls)

Utilities:
4.0 0.0 33.974 40.0
15.974 23.173 29.202 33.974
12.015 17.289 0.0 28.613
7.199 9.642 -8.0 19.877

Observation:
As discount factor is close to 1 the agent gives more preference to future values along the path.
As a result it gives more preference to (0,3) as it has highest reward value of 40.

The values of the utilities are positive showing that the agent tries to get to the max utility state.

Question 2.2.a:

Discount Factor = 0.99

Step Cost = X (40)

Policy obtained from Value Iteration Algorithm:

['-', '-', 'S', '-']
['S', 'W', 'W', 'S']
['S', 'W', '-', 'N']
['N', 'W', '-', 'N']
(where [(0,0),(0,3),(3,2)] are terminal states and [(0,1),(2,2) are walls)

Utilities:
4.0 0.0 1772.043 40.0
2008.179 2008.178 1982.529 1991.092
2008.178 2008.179 0.0 1991.624
2008.179 2008.178 -8.0 1774.751

Observation:
As now the value of the step cost is +ve, we expect that the agent does not want to leave the environment.
We observe that it is in fact true, as shown by the policy.

The high values of the utilities indicate that the agent does not want to go to relatively low valued terminal states.

Question 2.2.b:

Discount Factor = 0.99

Step Cost = -X/5 (-8)

Policy obtained from Value Iteration Algorithm:
['-', '-', 'E', '-']
['E', 'E', 'E', 'N']
['N', 'N', '-', 'N']
['N', 'N', '-', 'N']
(where [(0,0),(0,3),(3,2)] are terminal states and [(0,1),(2,2) are walls)

Utilities:

4.0 0.0 28.502 40.0
-2.381 7.915 19.395 28.502
-11.252 -3.086 0.0 18.271
-20.196 -13.19 -8.0 6.39

Observation:
As the step cost is moderately negative the agent wants to go to the best terminal state ie (0,3).
We also observe that the states around (0,3) have +ve values and other states have -ve values.

Question 2.2.c:

Discount Factor = 0.99

Step Cost = -X/4 (-10)

Policy obtained from Value Iteration Algorithm:
['-', '-', 'E', '-']
['N', 'E', 'E', 'N']
['N', 'N', '-', 'N']
['N', 'E', '-', 'N']
(where [(0,0),(0,3),(3,2)] are terminal states and [(0,1),(2,2) are walls)

Utilities:
4.0 0.0 25.764 40.0
-7.445 0.306 14.487 25.764
-18.982 -12.873 0.0 13.094
-29.889 -19.453 -8.0 -0.367

Observations:
As now the value of step cost is more negative, the agent exits the environment through the less profitable terminal state as shown by the policy.

It tries to reduce the number of steps and wants to exit the environment.

We also observe that at (3,3) it chooses N instead of W, so it wants to leave the environment but still wants to exit through a terminal state with +ve reward.

Question 2.2.d:

Discount Factor = 0.99

Step Cost = -X (-40)

Policy obtained from Value Iteration Algorithm:
['-', '-', 'E', '-']
['N', 'W', 'E', 'N']
['N', 'S', '-', 'N']
['E', 'E', '-', 'W']
(where [(0,0),(0,3),(3,2)] are terminal states and [(0,1),(2,2) are walls)

Utilities:
4.0 0.0 -15.253 40.0
-51.458 -101.021 -58.981 -15.253
-101.021 -110.151 0.0 -64.359
-110.151 -63.002 -8.0 -58.031

Observation:
Due to the large negative step cost, the agent tries to exit the environment as soon as possible.
It does not differentiate if the if the terminal state is +ve or not as shown by (3,3).

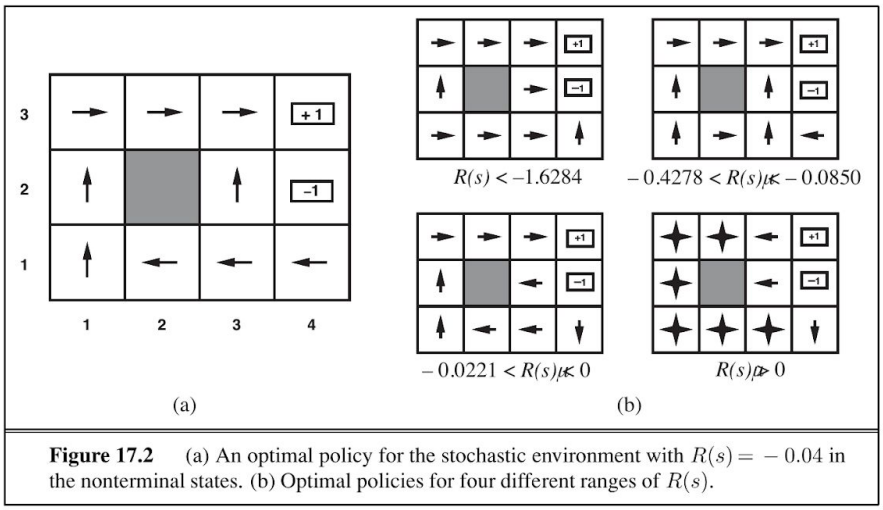

Figure 17.2 (a) An optimal policy for the stochastic environment with $R(s) = -0.04$ in the nonterminal states. (b) Optimal policies for four different ranges of $R(s)$.