AE504: Bonus Homework Assignment 1

Alex Faustino

March 15, 2019

Problem 1. For each state and each of the five possible actions, determine the reward that the agent achieves by performing the action at the corresponding state. (Note that the reward can be written as the negative of a cost.) Provide your answer in the form of five 7×7 tables, each for one of the agents actions.

You should assume that the actions which would make the agent leave the state space (e.g., go north from the top left corner) incur a reward of $-\infty$.

Solution: Since the reward for choosing the stay put action is always 0 no matter the state the table is simply 49 zeros. The rewards for choosing the remaining actions from every state can be seen in Tables 1-4.

$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$
-31.25	-11.24	-14.69	-55.76	-10.61	-23.09	-1.49
-26	-18.64	-3.25	-3.56	-8.84	-19.49	-2
-47.24	-11.89	-29.09	-1.36	-1	-3.56	-21.25
-25.01	-6.29	-43.25	-4.61	-34.64	-1.36	-24.04
-44.56	-2.69	-12.56	-57.25	-28.04	-5	-13.25
-74.96	-44.56	-1.16	-21.25	-8.29	-1.36	-2.21

Table 1: Reward for choosing the North action at every state.

Problem 2(a). Determine the optimal path for the agent, if it is required to arrive at its goal within **6** time steps. Determine the total amount of energy that the agent expends while following this path, expressed in units of energy.

Solution: Clearly, the optimal path is the only path. Choosing 'South' at every T is the only way to make it to the goal in 6 times steps. This yield a cost of $249.02e_t$.

-2.69	-2.21	-1	-4.61	-2.21	-1.16	$-\infty$
-2	-1.36	-14.69	-6.76	-1.25	-20.36	$-\infty$
-1.04	-27.01	-15.44	-47.24	-2	-2.21	$-\infty$
-14.69	-51.41	-1.81	-55.76	-1.36	-4.24	$-\infty$
-13.25	-3.89	-14.69	-1.09	-34.64	-6.76	$-\infty$
-4.24	-41.96	-1.16	-7.76	-7.76	-10.61	$-\infty$
-1.04	-1.04	-21.25	-22.16	-1.49	-2.96	$-\infty$

Table 2: Reward for choosing the East action at every state.

-31.25	-11.24	-14.69	-55.76	-10.61	-23.09	-1.49
-26	-18.64	-3.25	-3.56	-8.84	-19.49	-2
-47.24	-11.89	-29.09	-1.36	-1	-3.56	-21.25
-25.01	-6.29	-43.25	-4.61	-34.64	-1.36	-24.04
-44.56	-2.69	-12.56	-57.25	-28.04	-5	-13.25
-74.96	-44.56	-1.16	-21.25	-8.29	-1.36	-2.21
$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$	$-\infty$

Table 3: Reward for choosing the South action at every state.

$-\infty$	-2.69	-2.21	-1	-4.61	-2.21	-1.16
$-\infty$	-2	-1.36	-14.69	-6.76	-1.25	-20.36
$-\infty$	-1.04	-27.01	-15.44	-47.24	-2	-2.21
$-\infty$	-14.69	-51.41	-1.81	-55.76	-1.36	-4.24
$-\infty$	-13.25	-3.89	-14.69	-1.09	-34.64	-6.76
$-\infty$	-4.24	-41.96	-1.16	-7.76	-7.76	-10.61
$-\infty$	-1.04	-1.04	-21.25	-22.16	-1.49	-2.96

Table 4: Reward for choosing the West action at every state.

Problem 2(b). Determine the optimal path for the agent, if it is required to arrive at its goal within **10** time steps. Determine the total amount of energy that the agent expends while following this path, expressed in units of energy.

Solution: This requires actual calculation as the increased maximum time steps allows for the rover to circumvent the steep squares directly between itself and the goal. Using the Python script attached we find that the optimal cost when $T = 10$ is $70.4e_t$. This is achieved with the trajectory $a^{10} = \{E, S, S, S, S, E, S, S, W, W\}$.

Problem 2(c). Determine the optimal path for the agent, if it is required to arrive at its goal within **20** time steps. Determine the total amount of energy that the agent expends while following this path, expressed in units of energy.

Solution: Similar to part b the increase in maximum time steps allows for even more careful planning by the rover. Using the Python script attached we find that the optimal cost when $T = 20$ is $49.42e_t$. This is achieved with the trajectory $a^{20} = \{P, P, E, E, E, E, E, E, S, S, W, S, S, S, W, W, W, S, W, W\}$.

Problem 3. Using the terrain map above, informally explain why the paths in Problem 2 look the way that do.

Solution: As already mentioned in Problem 2, the paths look this way because as the rover is given more time it can choose longer paths that avoid steep terrain. If we continue to increase T the cost will eventually converge to its minimum value for this (s_0, s_T) pair.