ISLAMIC UNIVERSITY OF TECHNOLOGY (IUT)

ORGANISATION OF ISLAMIC COOPERATION (OIC)

Department of Computer Science and Engineering (CSE)

QUIZ 3 **WINTER SEMESTER, 2020-2021**

DURATION: 30 Minutes **FULL MARKS: 30**

CSE 4711: Artificial Intelligence

There are **2 (two)** questions. Answer **both** of them.
Figures in the right margin indicate marks for each question.
The square brackets at the start of each question denote the corresponding CO and PO.

---

1. Consider that an agent is in cell D of the Gridworld shown in Figure 1. The only action available in cells A, F, G, and H is `Exit` with rewards 6, 3, 0, and 0, respectively. When the agent exits from a cell, it gets the reward for that cell as specified. From the other cells, the agent can only take the action `Left` or `Right`, which results in the agent moving to the immediate left or right cell, respectively.

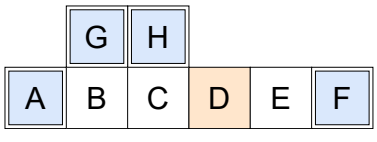

Figure 1: Gridworld for Question 1

If the agent is in cell B or C, and it takes a `Left` or `Right` action, it might fail sometimes. The agent will move in the desired direction with probability p, and it will fail and move up with probability $1 - p$. If the agent is in any other cell ($A, D, E, F, G,$ or H), the action will always be successful.

Assume that there is no living reward/penalty and the discount factor is γ.

a) [CO2, PO4] Model the problem as an MDP. Show the set of states, set of actions, start state, and terminal state(s). 8

Solution:

- States, $S = \{A, B, C, D, E, F, G, H\}$
- Actions, $A = \{\text{Left}, \text{Right}, \text{Exit}\}$
- Start State: D
- Terminal States: $\{A, F, G, H\}$

Rubric:

- 0.5 points for each correct value

b) [CO1, PO1] For each of the following policies, determine the value of each non-terminal state. Assume $\gamma \in [0, 1]$. 6×2

i. $\pi_{\text{Right}}(S) = \begin{cases} \texttt{Exit}, & \text{if } S \text{ is a Terminal State} \\ \texttt{Right}, & \text{otherwise} \end{cases}$

Solution:

- $B \to 3\gamma^4 p^2$
- $C \to 3\gamma^3 p$
- $D \to 3\gamma^2$
- $E \to 3\gamma$

Rubric:

- 0.5 points for each correct integer
- 0.5 points for each correct discount
- 0.5 points for each correct probability

ii. $\pi_{\text{Left}}(S) = \begin{cases} \texttt{Exit}, & \text{if } S \text{ is a Terminal State} \\ \texttt{Left}, & \text{otherwise} \end{cases}$

Solution:

- $B \to 6\gamma p$
- $C \to 6\gamma^2 p^2$
- $D \to 6\gamma^3 p^2$
- $E \to 6\gamma^4 p^2$

Rubric:

- 0.5 points for each correct integer
- 0.5 points for each correct discount
- 0.5 points for each correct probability

c) [CO3, PO2, PO3] What values of p will ensure that going Right from the start state is the optimal action? Provide a range in terms of γ. 5

Solution: For going right to be optimal, we need $V^{\pi_{\text{Right}}}$ of the start state to be greater or equal to $V^{\pi_{\text{Left}}}$. We get:

$$\begin{aligned} V^{\pi_{\text{Right}}}(D) &\geq V^{\pi_{\text{Left}}}(D) \\ 3\gamma^2 &\geq 6\gamma^3 p^2 \\ 6\gamma^3 p^2 &\leq 3\gamma^2 \\ p^2 &\leq \frac{3}{6\gamma} \\ p &\leq \frac{1}{\sqrt{2\gamma}} \end{aligned}$$

We get: $p \in \left[0, \frac{1}{\sqrt{2\gamma}}\right]$

Rubric:

- For defining the correct relation between the value of the start state: 3 points
- For the correct lower range: 1 point
- For the correct upper range: 1 point

2. [CO1, CO2, CO3, PO1, PO2, PO3, PO4] Write everything that you know about MDPs that was not asked in this quiz! 5

Solution: **Rubric:**

- Any correct information provided here that is not related to the questions given are accepted.