

NANYANG TECHNOLOGICAL UNIVERSITY
SPECIAL SEMESTER ASSIGNMENT 2024-2025
EE6231 – REINFORCEMENT LEARNING

July 2025

Time Allowed: 1 Week

INSTRUCTIONS

1. Option 1: A computation-based approach that provides the necessary details, intermediate steps, and intermediate results for obtaining the answer.
 2. Option 2: A programming-based approach that provides executable original code and output results. The specific programming language is not restricted and can be Python, C++, or MATLAB.
-

1. Find the optimal policy for the following four-state grid world example, with five-action space, as shown in Fig.Q1a, where S_0 is the target state and the discount rate $\gamma = 0.9$, predefined threshold $\theta = 0.0001$ to check the convergence of the state values.

Hints: Bellman equation, BOE, state value, action value, policy evaluation, policy improvement.

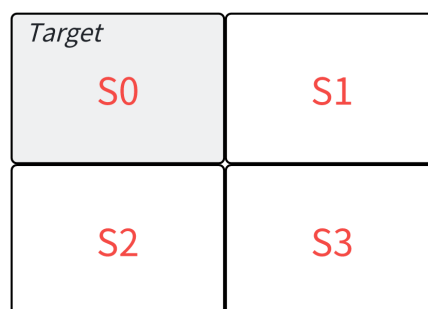


Figure Q1a: Four-state grid world example

Note: Question No. 1 continues on Page 2

(a) The action space is shown in Fig.Q1b.

- a_1 : move upward;
- a_2 : move rightward;
- a_3 : move downward;
- a_4 : move leftward;
- a_5 : stay still;

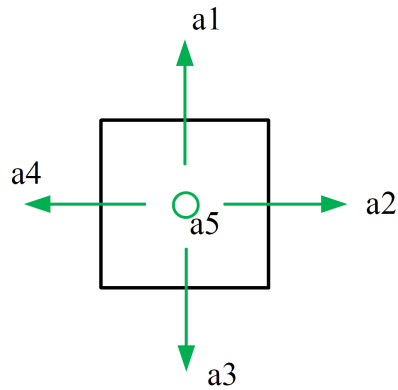


Figure Q1b: Action space of a state

(b) The rewards are designed as follows:

- If the agent attempts to stay still (action a_5), let $r_{\text{stay}} = -0.5$
- If the agent reaches the target cell (state S_0), let $r_{\text{target}} = +10$
- Otherwise, the agent gets a reward of $r = -1$.

(100 marks)

END OF PAPER