Question 2: MDP

1. Define the MDP:

a) State Space:

· Each cell represents a state, so because there are 25×25 cells, there are 625 potential states. However, some states may be blocked and there is only one goal state.

· So the state space =

$$S = \{(i,j)) \mid 1 \le i, j \le 25\}$$
blocked states $\cup$ {goal state}

b) Action Space

· The agent can only move up, down, left and right.

$$A = \{up, down, left, right\}$$

c) Transition Probabilities

· In a deterministic environment (not specified otherwise so I am just assuming), taking an action leads to a predictable next state, so:

$$P = \begin{cases} 1 & \text{if } S_{t+1} \text{ is reachable from } S_t \text{ using action } a_t \\ 0 & \text{otherwise} \end{cases}$$

d) Reward Function
   • For each step the agent takes to reach the goal, it receives a reward of -1.

$$R = \begin{cases} -1 & \text{for all transitions until the goal is reached} \\ 0 & \text{if goal state is reached} \end{cases}$$

e) Discount Factor
   • The specific value depends on how much future rewards are valued over immediate rewards. It is a value between 0 and 1 ($0 \leq \gamma \leq 1$). It is common practice to set $\gamma = 0.9$ to balance future and immediate rewards.