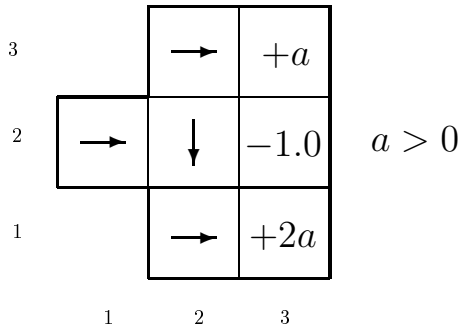
1. An agent is situated in a grid environment illustrated below. The agent is always in one of the states $(1\ldots3, 1\ldots3)$ excluding states $(1,1)$ and $(1,3)$ (the coordinates mention column first).



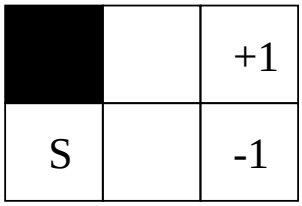
In each state, the agent may perform one of four actions $\downarrow$, $\uparrow$, $\leftarrow$ and $\rightarrow$ in order to move in the respective direction in the grid. These actions are uncertain so that the probability of the intended outcome (the agent moves in the correct direction) is 0.5. Otherwise, the agent moves with probabilities of 0.25 and 0.25 at right angles to the intended direction. Attempts to move outside the grid make the only exceptions to this rule, i.e., the agent stays in the same state. E.g. in state (1,2), the action $\rightarrow$ moves the agent to state (2,2) with a probability 0.5 and the agent remains in state (1,2) with a probability 0.25+0.25=0.5 because the agent would move outside the grid otherwise.

(a) In state (2,3), the agent performs a sequence of actions $[\downarrow, \downarrow]$. Which states are reachable by this sequence and with which probabilities?

(b) The utilities of states (3,1...3) appear in the figure and the constant a is positive. The cost associated with each action is 0.25. The arrows in the figure indicate a policy for the agent. Given this policy, determine the expected values for the utilities $U(1,2)$, $U(2,1)$ and $U(2,3)$ as well as the value of a assuming that the expected value for $U(2,2)$ is 0.

(c) Why a policy according to which the action $\leftarrow$ is performed in state (1,2) is problematic?

2. Consider a simplified operating environment for the agent:



The same actions are available for the agent but the probabilities are revised to 0.8 (the intended direction) and 0.1 (the other two directions) whereas the cost associated with is each action is 0.2.

Calculate the optimal policy π^* for the agent using (a) the value iteration algorithm and (b) the policy iteration algorithm.