1 Environment Dynamics

The stochastic maze environment is shown in the figure. There are a total of 12 states in the environment represented by the indices $\{0, 1, ..., 11\}$. The agent starts in the initial state 0. Four actions are possible in each state: left, right, up, and down. The environment is stochastic and we take the intended action only with a probability $p = 0.8$. We take an orthogonal action with a probability $p = 0.2$. If we collide with the edge of the environment or the wall present in state 5, the agent comes back to the initial state. The transition into the goal state 3 has a +1 reward and transition into the hole state 7 has a -1 reward. All other transitions have a -0.01 reward associated with them.

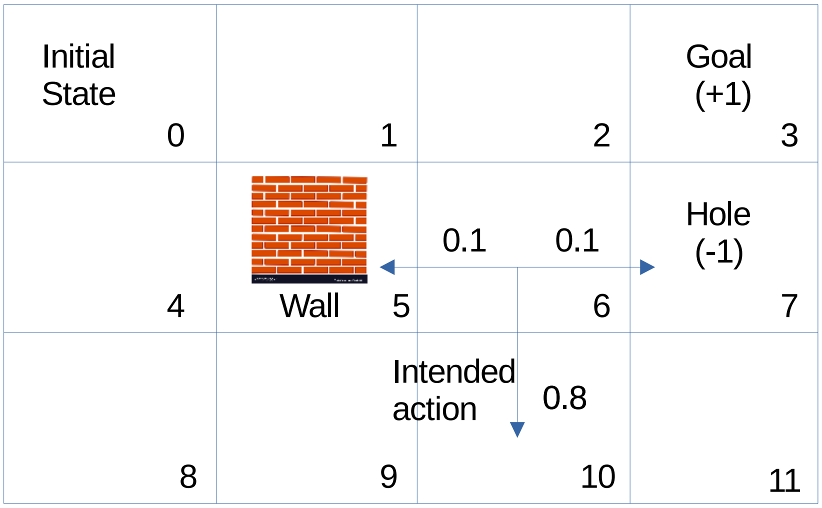

Figure 1: Stochastic Maze Environment

2 Problems

1. Implement the environment and check the implementation with some test cases.
2. Implements Policy Iteration (PI) and Value Iteration (VI).
3. Compare PI and VI in terms of convergence. Is the policy obtained by both same?