

# 1 Markov Decision Process

## 1.1 Description

The Markov decision process is a process where the action of an agent only depends on the current state of the environment. There is a set of possible actions and a set of possible states. There is also an agent policy which chooses an action based on the given state. The agent receives information about the state, performs the action provided by its policy and may receive a reward for the action. The state is updated and executed another time. The MDP has fixed transition probabilities which depend as said only on the current state. The reward probability is also fixed.

The agent's policy will be adapted to gain a maximum Return meaning a maximum future reward. The reward is estimated by estimating the Return for the next step and adding the reward for the transition.

The best estimates for this transition are stored as  $V$  values depending on the state or  $Q$  depending on both the state and the action.

## 1.2 Example for a MDP

A robot navigates through a maze, the states are the positions and orientations of the robot. The actions are the movement commands. The reward is given based on the proximity to the goal position.

## 1.3 Example for a POMDP

A real life application where the robot from above is only able to measure a few of the values which are defining its state. Therefore it is only able to approximate a distribution of possible states and apply a more general form of the MDP.