# Assignment 4

Team 7

June 15, 2021

## Contents

## List of Figures

## 1 Introduction

This is the submission file for Assignment 4.
The Authors:

- ME20B020 AKSHAT RAKESH GARHWAL
- ME20B032 ARCHISH S
- ME20B132 PRABHAT BEDIDA
- MM20B005 ALBIN GEORGE
- MM20B049 PRITHVIRAJ PRATAP BHOSLE

## 2 Archish S me20b032

A Markov decision process can be described as a tuple $\langle S, A, T, R \rangle$, where

- $S$ is a finite set of states of the world;
- $A$ is a finite set of actions;
- $T : S \times A \rightarrow \Pi(S)$ is the *state-transition function*, giving for each world state and agent action, a probability distribution over world states (we write $T(s, a, s')$ for the probability of ending in state $s'$, gievn that the agent starts in state $s$ and takes action $a$);
- $R : S \times A \rightarrow \mathbb{R}$ is the reward function, giving the expected immediate reward gained by the agent for taking each action in each state (we write $R(s, a)$ for the expected reward for taking action $a$ in state $s$);

- A stationary policy, $\pi : S \to A$, is a situation-action mapping that specifies, for each state, an action to be taken.
- $V_\pi(s)$ is the expected discounted sum of future reward for starting in state $s$ and executing policy $\pi$.

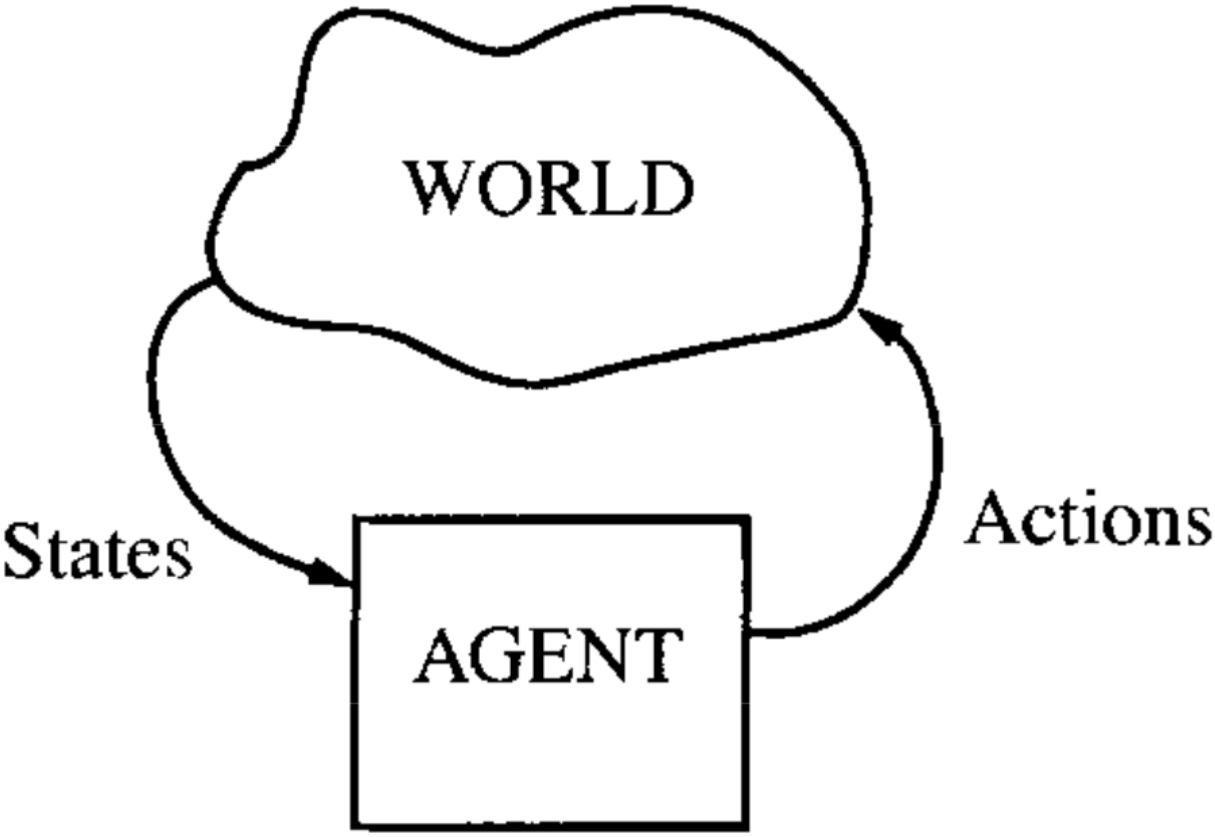

Figure 1: An MDP models the synchronous interaction between agent and world

In this model, as described by figure 1, the next state and the expected reward depend only on the previous state and the action taken; even if we were to condition on additional previous states, the transition probabilities and the expected rewards would remain the same. This is known as the Markov property.

$$V_\pi(s) = R(s, \pi(s)) + \gamma \sum_{s' \in S} T(s, \pi(s), s') V_\pi(s') \tag{1}$$

Given the Value Funcition 1 a greedy policy with respect to that value function, $\pi_V$, is defined as

$$\pi_V(s) = \operatorname*{argmax}_a \left[ R(s, a) + \gamma \sum_{s' \in S} T(s, \pi(s), s') V(s') \right] \tag{2}$$

# References

[1] L.P. Kaelbling, M.L. Littman, and A.R. Cassandra. Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101(1-2):99–134, 1998.