Assignment 1

Author: **Paolo Renzi**

SAPIENZA
UNIVERSITÀ DI ROMA

MARR, RL
November 8, 2023

Theory

Problem 1

We can derive how many steps are needed from this equation:

$$2\gamma^i/(1-\gamma)||Q0 - Q*|| \leq \epsilon \tag{1}$$

Because $2\gamma^i/(1-\gamma)||Q0 - Q*||$ is the distance between $v^{\pi i}(s)$ and $v^*(s)$.

First of all we know both $Q0$ and $Q* \in [0, 1/(1-\gamma)]$ By the infinity norm, the maximum value is $1/(1-\gamma)$.

$$2\gamma^i/(1-\gamma)^2 \leq \epsilon \tag{2}$$

Then we add and subtract 1

$$2(1-(1-\gamma))^i/(1-\gamma)^2 \leq \epsilon \tag{3}$$

Then we substitute $-(1-\gamma)$ with $e^-(1-\gamma)^2$ because it makes solving it easier even if it makes the approsimation less accurate

$$2e^{-(1-\gamma)i}/(1-\gamma)^2 \leq \epsilon \tag{4}$$

Then we divide by 2 and multiply by $(1-\gamma)^2$ and get

$$e^{-(1-\gamma)i} \leq (1-\gamma)^2/2 \tag{5}$$

Then we take the log of both sides

$$-i(1-\gamma) \leq -\log(2/(\epsilon(1-\gamma)^2)) \tag{6}$$

In the end we divide by $-(1-\gamma)$ (so changing also the $\leq$ in $\geq$)

$$i \geq \log(2/(\epsilon(1-\gamma)^2))/(1-\gamma) \tag{7}$$

Problem 2

MDP definition
S1, ..., S7
r(S,a)

$$\begin{cases} 1/2 & S = S_1 \\ 5 & S = S_7 \\ 0 & otherwhise \end{cases}$$

$P(S_6|S_6, a_1) = 0.3$
$P(S_7|S_6, a_1) = 0.7$
$\pi(s) = a_1 \forall s \; v^1 = [0.5, 0, 0, 0, 0, 0, 5] \; \gamma = 0.9$
We calculate the discounted reward from S_6 and add it to the current reward of S_6 to get the value of S_6
$V(S_6) = r(S_6) + \gamma \mathbb{E}(V(S')) =$
$= 0 + 0.9(\mathbb{P}(S_6|S6, a_1)V(S_6) + \mathbb{P}(S_7|S6, a_1)V(S_7)) = 0.9(0.3 * 0 + 0.7 * 5) = 0.9 * 3.5 = 3.15$

Code

Policy iteration

To implement the reward function I used an if to check if i was in the goal state and return 1 and an else to return 0 in all other cases
To check if the transition was feasibile I used a couple of ifs, one to check if it was trying to go out of the grid and the second to check if it's going against an obstacle
To return the transition probability i first checked if the transition was feasibile and then I assigned 1/3 to each transition (if the transition isn't feasibile with the action it would would just remain the state it was)

iLQR

For iLQR I just implemented the formulas being careful about which operator i was using between * and @.