cme241 - Assignment 3

Frederic Ladouceur

January 2021

1 Question 1

$$V^{\pi_D}(s) = Q^{\pi_D}(s, \pi_D(s))) \tag{1}$$

$$Q^{\pi_D}(s,a) = R(s,a) + \gamma \sum_{s' \epsilon N} P(s, \pi_D(s), s') V^{\pi_D}(s') \tag{2}$$

$$V^{\pi_D}(s) = R(s, \pi_D(s)) + \gamma \sum_{s' \epsilon N} P(s, \pi_D(s), s') V^{\pi_D}(s') \tag{3}$$

$$Q^{\pi_D}(s,a) = R(s,a) + \gamma \sum_{s' \epsilon N} P(s, \pi_D(s), s') Q^{\pi_D}(s', \pi_D(s'))) \tag{4}$$

The four equations above represents the 4 MDP Bellman equations for a deterministic policy. Equation 1 represents V^{π_D} in terms of Q^{π_D}, Equation 2 represents Q^{π_D} in terms of V^{π_D}, Equation 3 represents V^{π_D} in terms of V^{π_D}, and Equation 4 represents Q^{π_D} in terms of Q^{π_D}.

2 Question 2

The MDP State-Value Function Bellman Optimality Equation is:

$$V^*(s) = max_{a \epsilon A}\{R(s,a) + \gamma \sum_{s' \epsilon N} P(s,a,s') V^*(s')\} \tag{5}$$

From the values given in the problem statement, the equation becomes:

$$V^*(s) = max_{a \epsilon A}\{(1-a)(1+2a) + 1/2(aV^*(s+1) + (1-a)V^*(s))\} \tag{6}$$

Since the optimal value function doesn't depend on the value of s, i.e. $V^*(s) = V^*(s')$ for all $s, s' \epsilon S$, we have:

$$\begin{aligned} V^*(s) &= max_{a \epsilon A}\{(1-a)(1+2a) + 1/2(aV^*(s) + (1-a)V^*(s))\} & (7) \\ V^*(s) &= 2max_{a \epsilon A}\{(1-a)(1+2a)\} & (8) \\ V^*(s) &= 9/4 & (9) \end{aligned}$$

We also have:

$$\pi^*(s) = \text{argmax}_{a \epsilon A}\{(1-a)(1+2a)\} = 1/4 \tag{10}$$