Assignment 5 Part A Report

0a.

$$Q^*(s,a) = \sum_{s'} T(s,a,s')[R(s,a,s') + \gamma V^*(s')]$$

$$V^*(s) = \overset{max}{a} Q^*(s,a)$$

0b.

$$V_{k+1}(s) = \overset{max}{a} Q_{k+1}(s)$$

$$Q_{k+1}(s) = \sum_{s'} T(s,a,s')[R(s,a,s') + \gamma V_k(s')]$$

1a. 4 iterations of VI are required

1b. 8 iterations of VI are required

1c. No, the is moving the right all the time, so it will stay in a state forever.

2a. 7 iterations

2c. Yes, because this policy will allows you to get into a the final state in a relatively short path

2d. 56 iterations

2e. No, it has not changed because value iteration doesn't have to converge in order to find a best policy.

3a. The initial state has value 0.82. The policy indicates to go to the goal with reward 10 if you at the start state, in some other states, it determined that it should reach for the goal with

3b. The policy now indicates that it should always just reach for the goal with reward 100. Completely ignoring the goal with reward 10.

4a. 4 simulations

4b. 7 simulations

4c. There are three runs, which is 1, 1, 3 steps away

4d. The upper part of the state space is never reached

5a. No, it is not. The best policy is going to appear long before the convergence of the states.

5b. It’s not important because we don’t need to get the very accurate values for all states. Normally a rough bad value can already tell us that state doesn’t worth to explore again.