Assignment 6 Report for Part A

Hung Lo
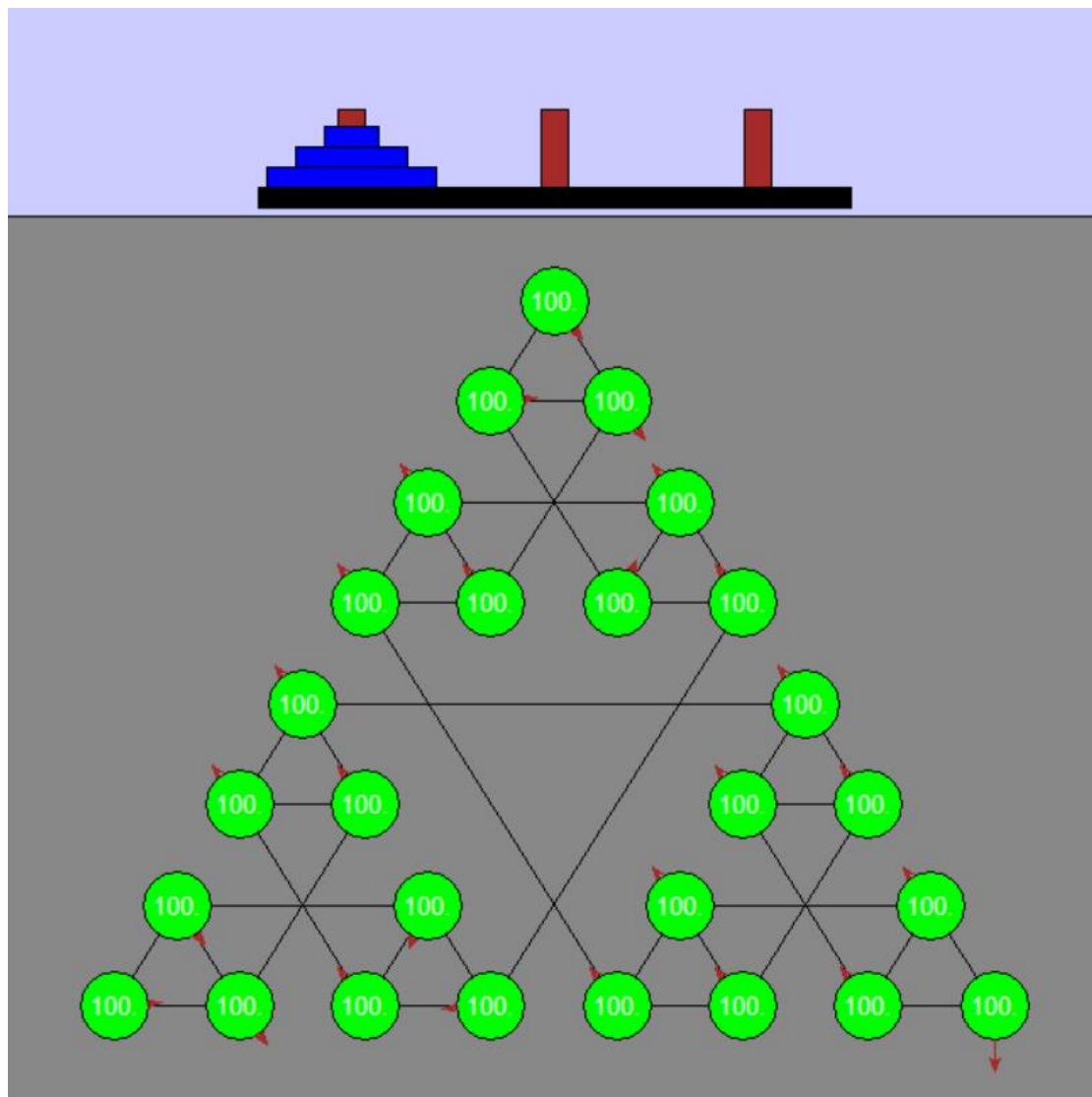
Problem1:

1a 4 iterations

1b 8 iterations

1c It is not a good policy, because no discount and no noise will make the V value of all state become 100 after converging. In this situation, the state will always try to stay in the same state (by moving illegal move) or move to nearby state (because reward is same as get to the terminal state). As result, the agent will not move to goal state.
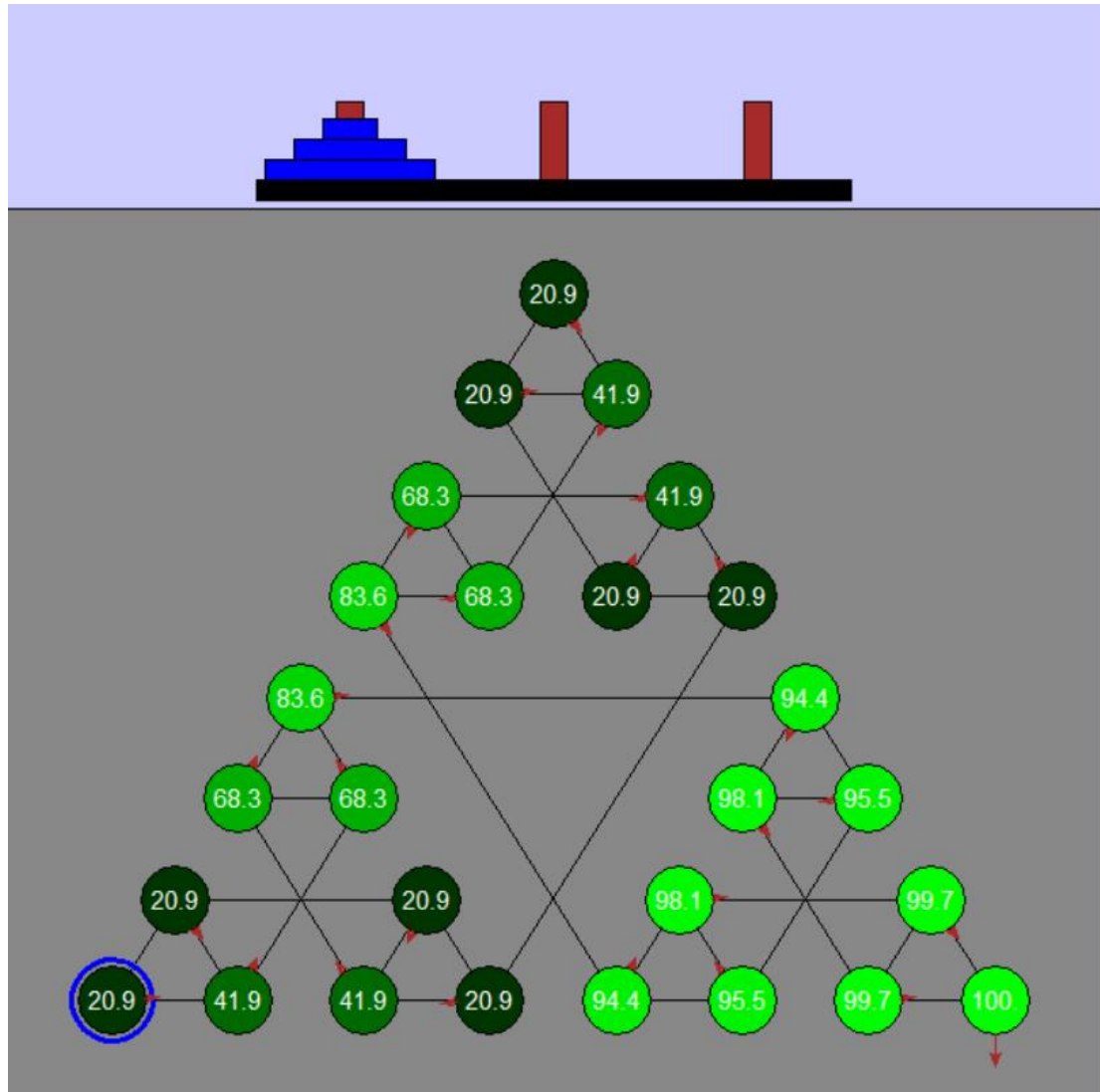
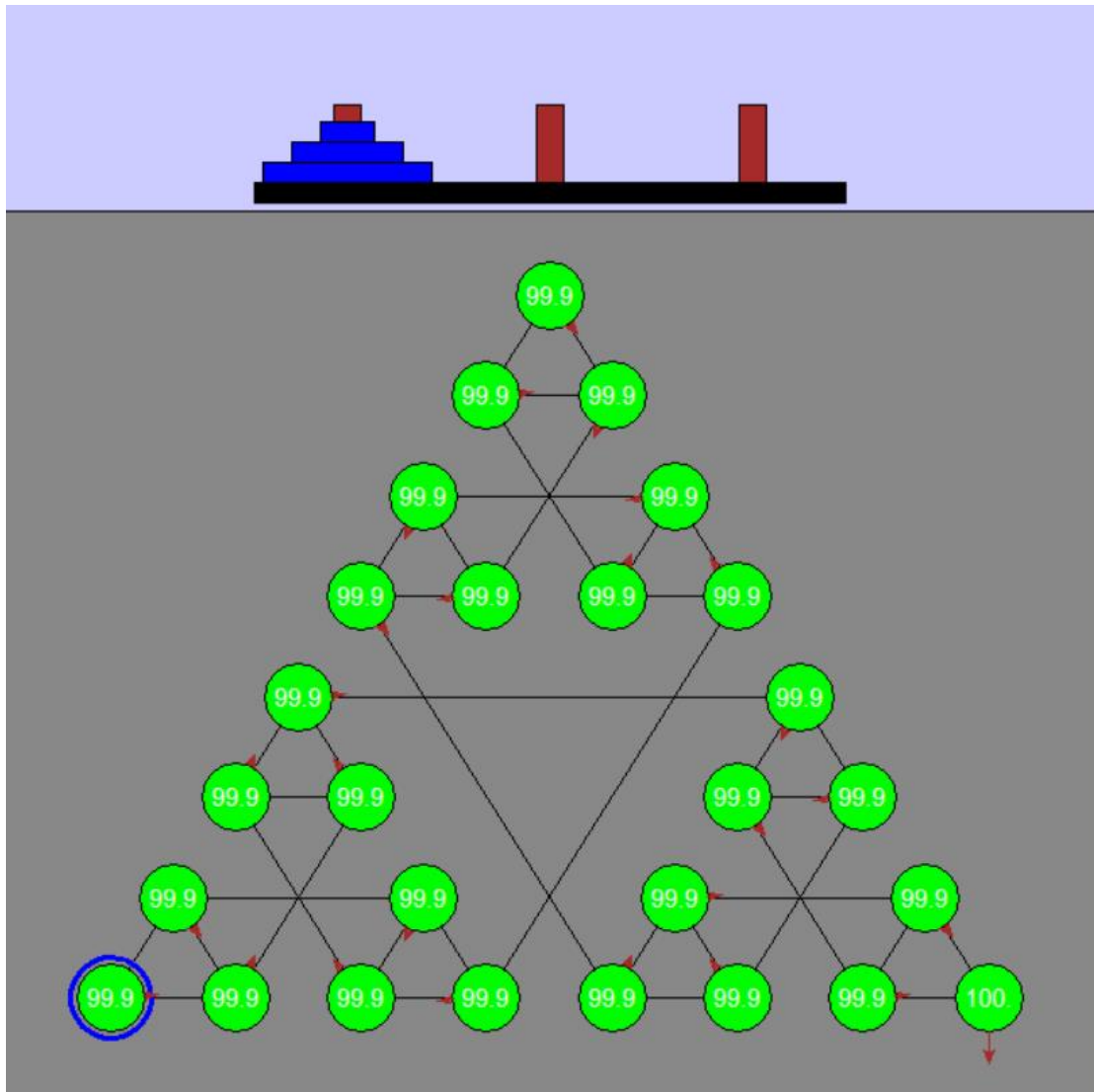Show policy below:

Problem2

2a 8 iterations

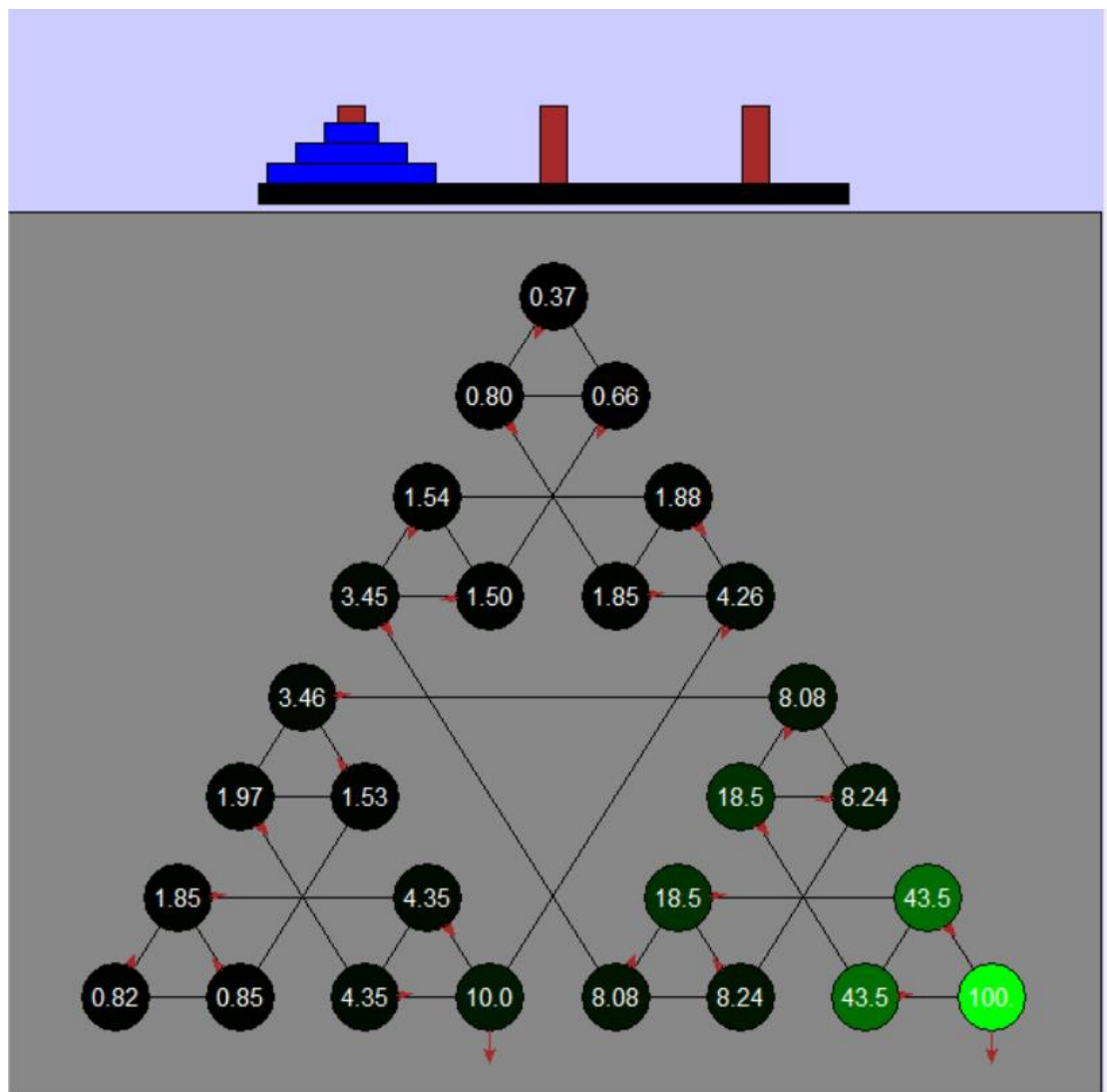2c Policy now is good, because the policy points the shortest route to goal state

2d After 56 iterations, VI will be converge.

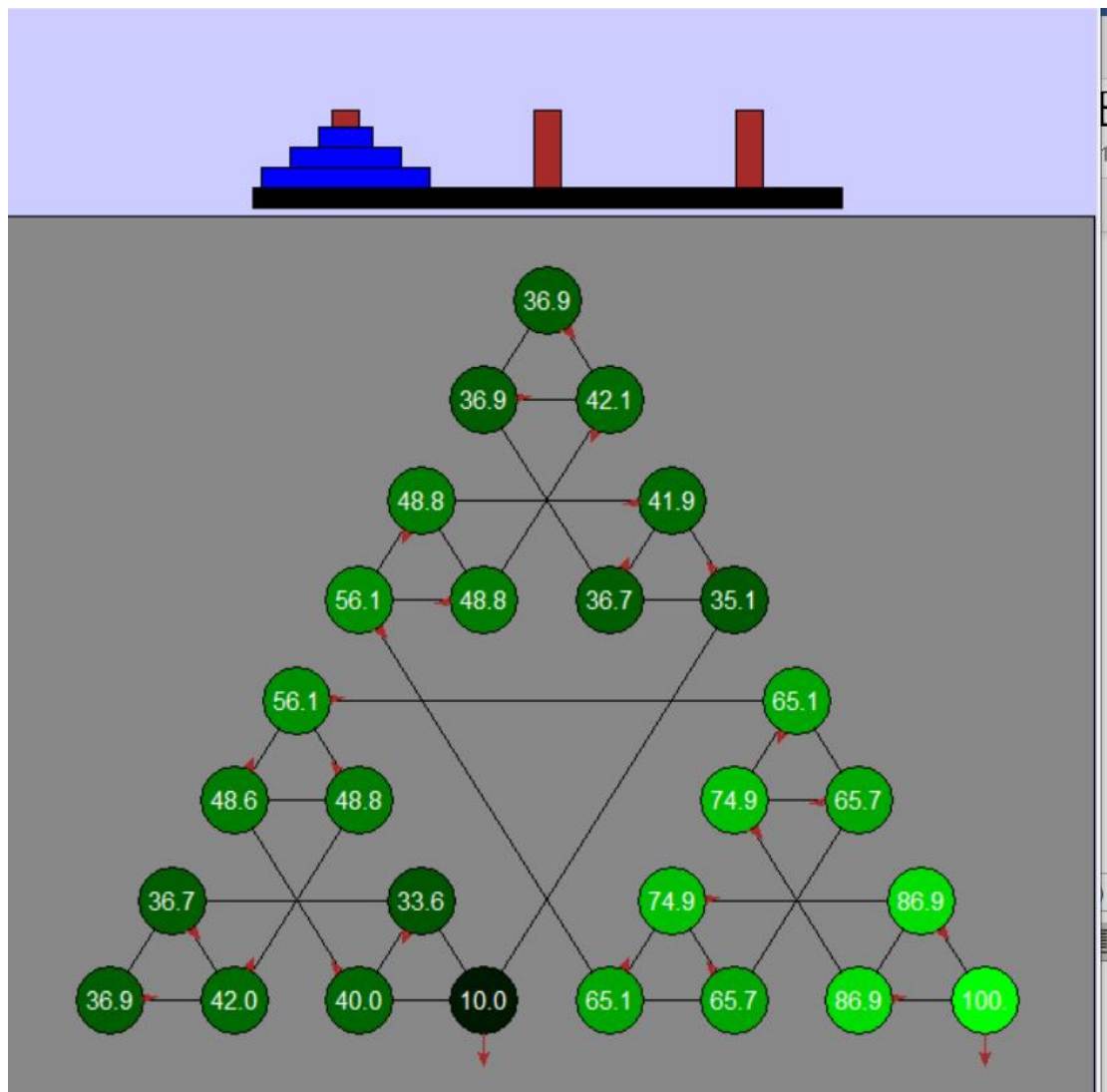2e Policy is same as 2c. Because Policy is return V_value of

Not change. Because gamma =1, so the value is always greater, while closer the goal state.
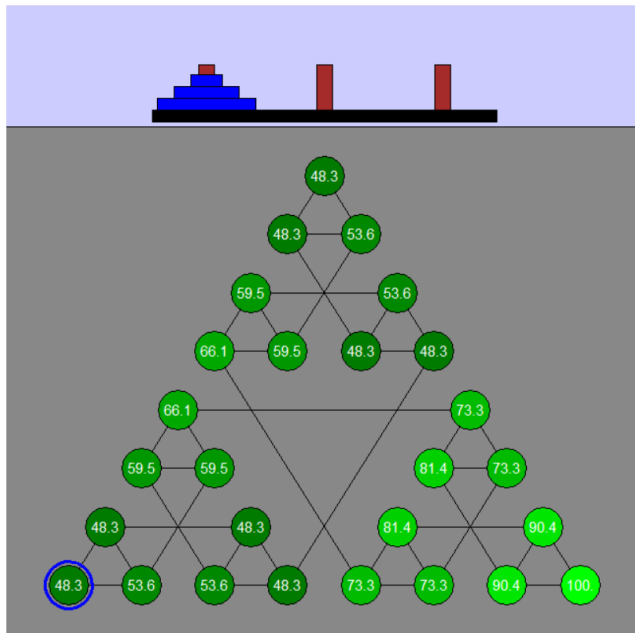
3a

3b

4a three of them

4b six of them

4c about 5 steps

4d States on the top seems to never be visited during simulation.

5a Converged value is not essential for state to get good policy. Once reward of goal state propagates to all state, the value will not be changed a lot between current state and converged state. For example, below is iteration=8:

And the other is iteration=100:
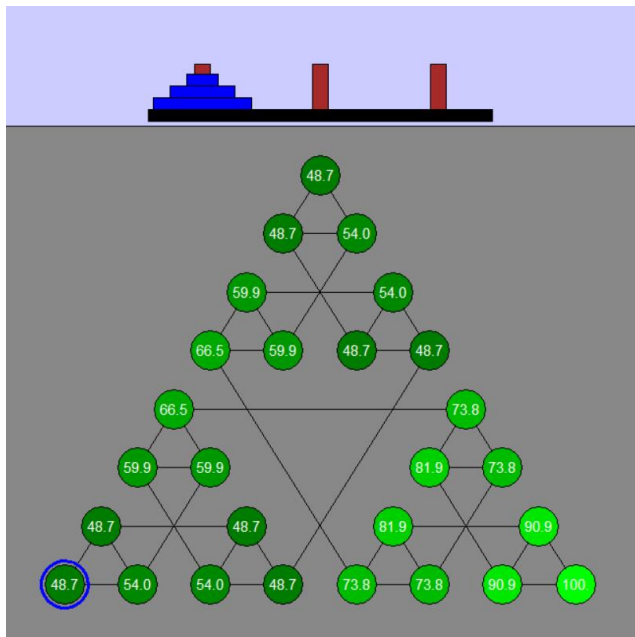
5b It is not important all states to be visited a lot after goal reward propagating all states, because value will not be changed a lot between converged situation and goal reward propagating state. (just like 5a pictures)