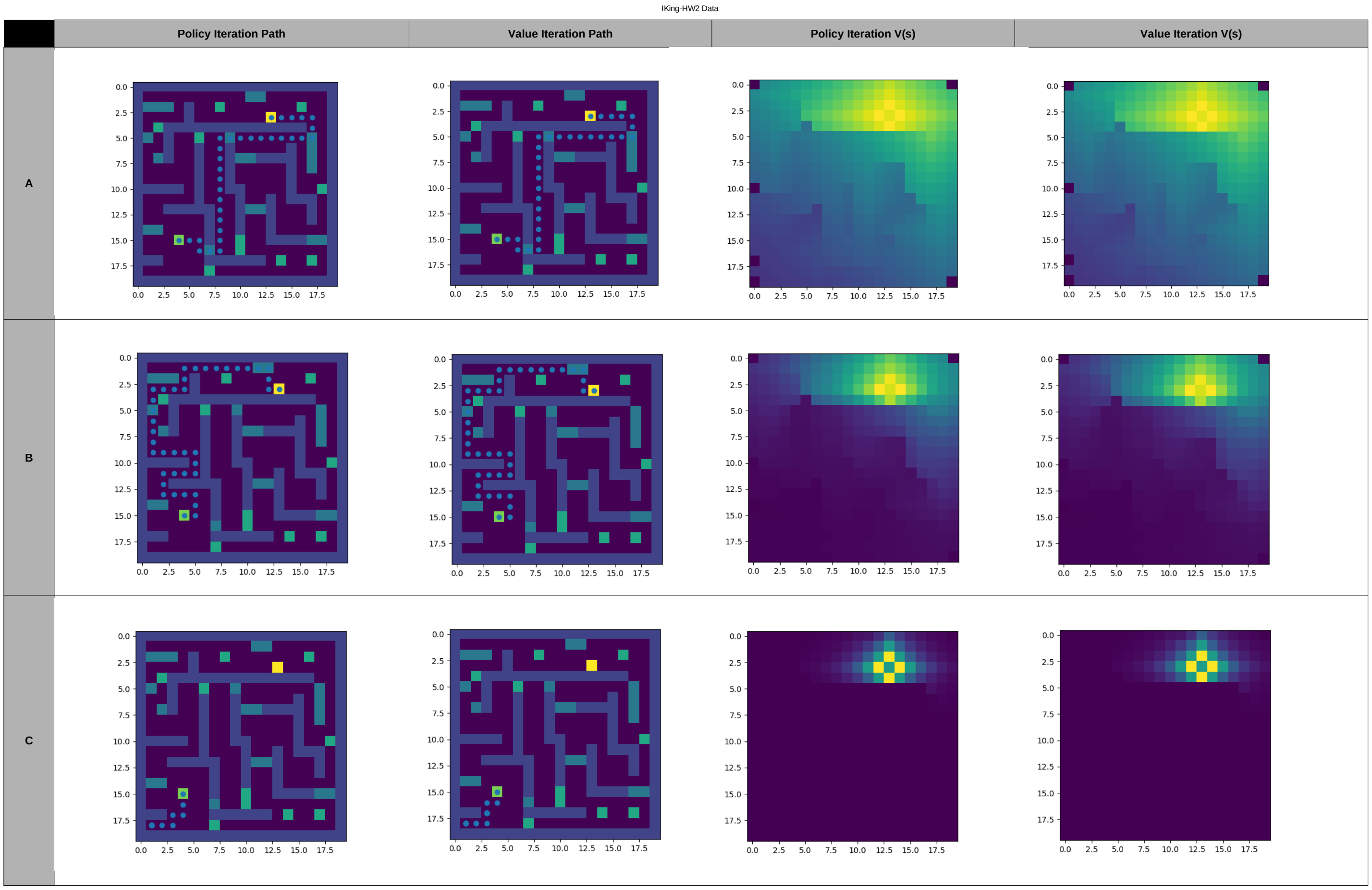
Homework #2

Isaiah King
03/10/20

Iterations:

Scenario	Number of Policy Evaluations	Number of Policy Improvements
A – Policy Iteration	748	13
A – Value Iteration	105	N/a
B – Policy Iteration	432	30
B – Value Iteration	106	N/a
C – Policy Iteration	84	38
C – Value Iteration	16	N/a

Data Visualization*:



*(Please see *data_visualization.pdf* for complete, uncompressed version)

As is evident from the best paths found by the agents, and the state-value tables, there is no difference in the results that come from policy or value iteration. The compute-time, however, does vary considerably. We note of particular interest that Scenario-A—whose parameters find the most efficient path to the goal—takes almost twice as many steps to find the solution than Scenario-B does using policy iteration. However, using value iteration, A and B take nearly identical compute time, and A finds a more efficient path to the goal.

From these two results alone, we observe that hyper-parameters, that is, the values for *p, γ,* and *θ*, determine the efficiency and computational complexity of the solution more than which particular dynamic programming approach is used. This is even more evident by observing the results from Scenario-C. Here, the discount factor is so small that even with an enormous reward for the goal, and a tiny value of θ, the value simply cannot propagate across all states. This is evident by the minuscule number of policy evaluations (the number of times the internal loop is computed in value iteration), but it is even more clear by observing the visualization of V(s) for each state on the 2D plane. Because γ is just over 0.5, any value in adjacent cells is halved with each subsequent displacement from its origin, which allows for negative rewards such as the holes and oil patches next to the start point, to overpower the positive rewards from far-away goals. As such, the path for Scenario-C shows the agent fleeing the nearby negative rewards into a corner, and remaining there until the simulation ends.