**Assignment 2 Question 3**

Alexander Khrulev 500882732

Mahan Pandey 500881861

**Using the following parameters**

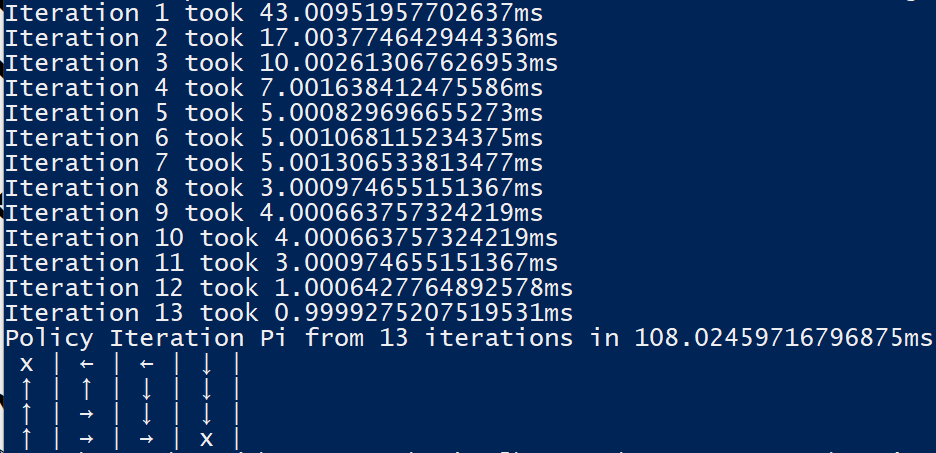
p1 = 0.8

p2 = 0.1

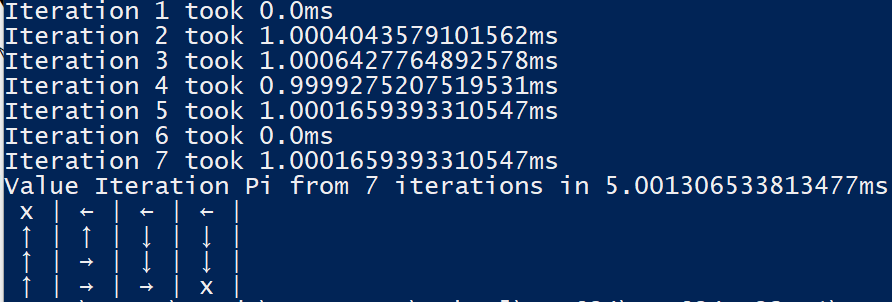
gamma = 0.95

theta = 0.001

**Policy iteration results**

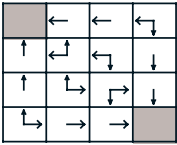
****

**Value iteration results**

****

**Discussion**

Both policy iteration and value iteration produce the optimal policy although they assign different actions for state 3. This is consistent with the girdworld optimal policy solutions given in the textbook:



Although both policy iteration and value iteration algorithms produce optimal policies the value iteration algorithm is significantly more performant. Not only does value iteration algorithm produce the optimal policy in 6 less iterations than the policy iteration algorithm the iterations of the value iteration algorithm is significantly faster than those of the policy iteration algorithm. This is because for each iteration of the policy iteration algorithm it will repeatedly execute “sweeps” or policy evaluation for each state until the V(s) converges and the delta between iterations of policy evaluation is less than theta. In contrast, the value iteration algorithm applies a truncated policy evaluation where it only executes only one “sweep” or one update of each state. Thus, the value iteration algorithm is significantly more performant than the policy iteration algorithm.