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

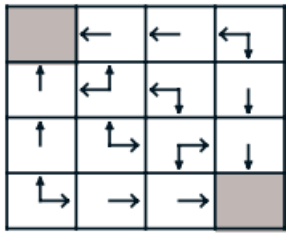
```
Iteration 1 took 43.00951957702637ms
Iteration 2 took 17.003774642944336ms
Iteration 3 took 10.002613067626953ms
Iteration 4 took 7.001638412475586ms
Iteration 5 took 5.000829696655273ms
Iteration 6 took 5.001068115234375ms
Iteration 7 took 5.001306533813477ms
Iteration 8 took 3.000974655151367ms
Iteration 9 took 4.000663757324219ms
Iteration 10 took 4.000663757324219ms
Iteration 11 took 3.000974655151367ms
Iteration 12 took 1.0006427764892578ms
Iteration 13 took 0.9999275207519531ms
Policy Iteration Pi from 13 iterations in 108.02459716796875ms
 x | ← | ← | ↓ |
 ↑ | ↑ | ↓ | ↓ |
 ↑ | → | ↓ | ↓ |
 ↑ | → | → | x |
```

Value iteration results

```
Iteration 1 took 0.0ms
Iteration 2 took 1.0004043579101562ms
Iteration 3 took 1.0006427764892578ms
Iteration 4 took 0.9999275207519531ms
Iteration 5 took 1.0001659393310547ms
Iteration 6 took 0.0ms
Iteration 7 took 1.0001659393310547ms
Value Iteration Pi from 7 iterations in 5.001306533813477ms
 x | ← | ← | ← |
 ↑ | ↑ | ↓ | ↓ |
 ↑ | → | ↓ | ↓ |
 ↑ | → | → | x |
```

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.