AI Assignment II Part B

Mohsin Mustafa (**20161131**) - Shyamgopal Karthik (**20161221**)

Team **I** - *OS is Karnel, Karnel is OS*

In the provided test-case substituting x for 1 (the team number), there were totally sixteen iterations (with the initial grid as iteration zero) when run with the Value Iteration algorithm (terminal case remaining the same as mentioned in the question).
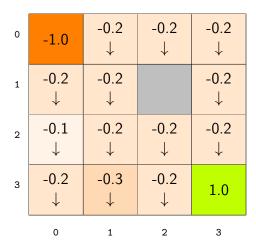
The iterations are shown in grid form, with arrows denoting the 'chosen' action for a particular state. There are explanations when prominent changes occur in a particular iteration. For each iteration, the loop runs through the states row-wise, left to right.
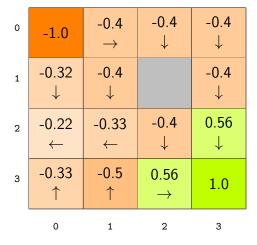
Iteration 0

The initial state of the board, where the two sinks in the two corners of the board have +1 and −1 as their values. The start state is $(3,0)$, the bottom-left corner. $(1,2)$ is gray-ed out as it is a wall, and $(2,0)$, $(3,1)$ have reward values that are not zero as well.

Iteration 1

	0	1	2	3
0	-1.0	-0.2 ↓	-0.2 ↓	-0.2 ↓
1	-0.2 ↓	-0.2 ↓		-0.2 ↓
2	-0.1 ↓	-0.2 ↓	-0.2 ↓	-0.2 ↓
3	-0.2 ↓	-0.3 ↓	-0.2 ↓	1.0

Since the values are being updated from the top-left corner, the states acquire negative values which cascade to the neighbouring states that are remaining to be calculated. The arrows point away from the top-left corner, which is the most negative part. The effects of the bottom-right positive sink is yet to be evident, since this was just the first iteration.

Iteration 2

	0	1	2	3
0	-1.0	-0.4 →	-0.4 ↓	-0.4 ↓
1	-0.32 ↓	-0.4 ↓		-0.4 ↓
2	-0.22 ←	-0.33 ←	-0.4 ↓	0.56 ↓
3	-0.33 ↑	-0.5 ↑	0.56 →	1.0

$(3, 2)$ and $(2, 3)$ have their values updated positively and both of their chosen actions directly point towards the positive sink. However since these values were calculated late, the other states aren't affected by their values in

the current iteration. Towards the left bottom corner, the states are being influence against the -0.1 reward value state and towards the 0.1 value state.

Iteration 3

	0	1	2	3
0	-1.0	-0.6 →	-0.6 ↓	-0.6 ↓
1	-0.45 ↓	-0.53 ←		0.168 ↓
2	-0.34 ←	-0.47 ←	0.271 ↓	0.616 ↓
3	-0.46 ↑	0.065 →	0.616 →	1.0

With this iteration, the third level (including $(3, 3)$ as level one) gets influenced by the positive sink's values. All the states in these levels clearly prefer the action that directly leads to the positive sink. Their values also become positive. On the other hand, the other states have their values become more negative as they are mainly influenced by the other negative reward states on the other side of the grid.

Iteration 4

	0	1	2	3
0	-1.0	-0.78 ↓	-0.8 ↓	-0.19 ↓
1	-0.57 ↓	-0.66 ←		0.326 ↓
2	-0.46 ←	-0.03 →	0.382 →	0.689 ↓
3	-0.23 →	0.153 →	0.689 →	1.0

Like the previous iteration, this one too increases the values of the states four 'levels' away from the bottom right sink, and the states further away become more negative as they are influenced by the negative sink in the top-left.

Iteration 5

	0	1	2	3
0	-1.0	-0.91 ↓	-0.51 →	-0.04 ↓
1	-0.69 ↓	-0.35 ↓		0.416 ↓
2	-0.23 →	0.054 →	0.458 →	0.707 ↓
3	-0.15 →	0.263 →	0.707 →	1.0

The breadth of influence of the positive sink increases, states become more positive (all from level 0 till the current level). However, the values of the states not influenced yet, $(1, 0)$ and $(0, 1)$, become more negative due to the direct influence from the negative sink and since they were calculated before their neighbours values increased in the current iteration.

Iteration 6

Now even the states directly opposite to the positive end state have started getting reduced in value, and all the states point towards the positive sink (away from the negative sink). The influence of the $+1$ and -1 reward states are lessening. Soon the values of the states will increase further as they prefer the route to the bottom-right. One interesting observation is that the $(0, 3)$ state has a higher value than the $(3, 0)$ state - this is due to the -0.1 reward value that the $(3, 1)$ state has.

	0	1	2	3
0	-1.0	-0.63 ↓	-0.33 →	0.078 ↓
1	-0.47 ↓	-0.26 ↓		0.449 ↓
2	-0.14 →	0.158 →	0.482 →	0.717 ↓
3	-0.02 →	0.297 →	0.717 →	1.0

Iteration 7 .. 16

	0	1	2	3
0	-1.0	-0.55 ↓	-0.2 →	0.134 ↓
1	-0.39 ↓	-0.15 ↓		0.463 ↓
2	-0.02 →	0.189 →	0.493 →	0.72 ↓
3	0.021 →	0.319 →	0.72 →	1.0

The following states till the very end follow the same process, the values get more and more precise till they converge to the final value. The Value Iteration algorithm was terminated when there was a $< 1\%$ change in all the values in a particular iteration. The arrows, or the preferred action, remains constant till the very end in this case and the values become more positive as the iterations increase.

	0	1	2	3
0	-1.0	-0.43 →	-0.13 →	0.163 ↓
1	-0.27 ↓	-0.1 ↓		0.468 ↓
2	0.015 →	0.212 →	0.497 →	0.721 ↓
3	0.055 →	0.327 →	0.721 →	1.0

	0	1	2	3
0	-1.0	-0.36 →	-0.1 →	0.178 ↓
1	-0.22 ↓	-0.07 ↓		0.471 ↓
2	0.048 →	0.22 →	0.499 →	0.722 ↓
3	0.068 →	0.331 →	0.722 →	1.0

	0	1	2	3
0	-1.0	-0.32 →	-0.08 →	0.185 ↓
1	-0.19 ↓	-0.05 ↓		0.472 ↓
2	0.06 →	0.225 →	0.5 →	0.722 ↓
3	0.076 →	0.333 →	0.722 →	1.0

	0	1	2	3
0	-1.0	-0.3 →	-0.07 →	0.188 ↓
1	-0.18 ↓	-0.04 ↓		0.472 ↓
2	0.069 →	0.228 →	0.5 →	0.722 ↓
3	0.08 →	0.333 →	0.722 →	1.0

	0	1	2	3
0	-1.0	-0.29 →	-0.06 →	0.19 ↓
1	-0.17 ↓	-0.04 ↓		0.472 ↓
2	0.072 →	0.229 →	0.5 →	0.722 ↓
3	0.082 →	0.334 →	0.722 →	1.0

	0	1	2	3
0	-1.0	-0.28 →	-0.06 →	0.19 ↓
1	-0.16 ↓	-0.04 ↓		0.472 ↓
2	0.074 →	0.229 →	0.5 →	0.722 ↓
3	0.082 →	0.334 →	0.722 →	1.0

	0	1	2	3
0	-1.0	-0.28 →	-0.06 →	0.191 ↓
1	-0.16 ↓	-0.04 ↓		0.472 ↓
2	0.075 →	0.23 →	0.5 →	0.722 ↓
3	0.083 →	0.334 →	0.722 →	1.0

	0	1	2	3
0	-1.0	-0.28 →	-0.06 →	0.191 ↓
1	-0.16 ↓	-0.04 ↓		0.472 ↓
2	0.076 →	0.23 →	0.5 →	0.722 ↓
3	0.083 →	0.334 →	0.722 →	1.0

	0	1	2	3
0	-1.0	-0.28 →	-0.06 →	0.191 ↓
1	-0.16 ↓	-0.04 ↓		0.472 ↓
2	0.076 →	0.23 →	0.5 →	0.722 ↓
3	0.083 →	0.334 →	0.722 →	1.0