# **Robot Learning Exercise 3**

# Member 1

Name: Hojun Lim Mat. No: 3279159

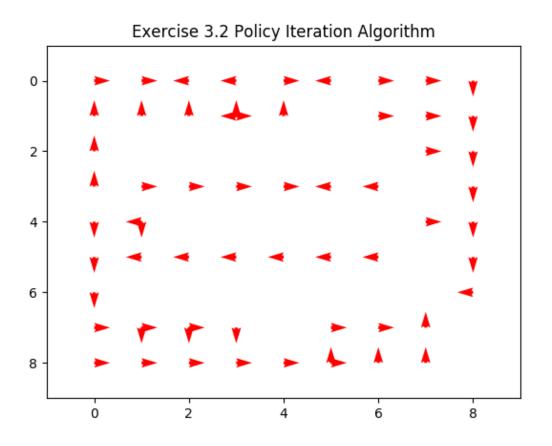
## Member 2

Name: Kajaree Das Mat. No: 3210311

# Solution for 3.1:

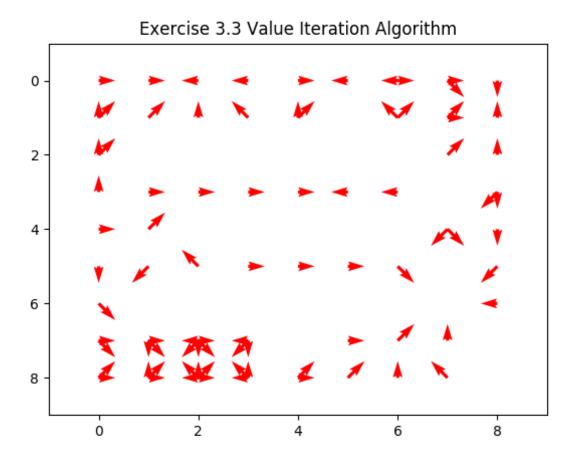
0 -	Exercise 3.1 expected value of all cells								
٥Ţ	-40.63	-65.95	-81.72	-90.46	-98.14	-101.2	2 -82.59	-57.64	-30.43
1 -	-40.77	-79.81	-96.57	-107.0	-115.4	v	-98.63	-67.07	-34.87
2 -	-40.77	-75.01	-30.57	-107.0	-113.4	^	-30.03	-07.07	-54.07
	-39.09	Х	Х	Х	Х	Х	Х	-81.94	-34.2
3 -	20.06	70.06	114.2	4 121 7	7 120 2	2 146 4	2 156 5	7 ∨	20.20
	-38.96	-70.06	-114.2	4-131./	/ -130.3	3-146.4	2-156.5	<b>'</b> ^	-29.28
4 -	-37.11	-74.07	Х	Х	Х	Х	Х	-97.02	-26.62
5 -									
	-33.39	-89.59	-127.0	3 -142.4	8 -149.8	1 -155.5	8 -164.9	3 X	-14.07
6 -	-24.77	х	х	Х	Х	Х	Х	G	-7.18
7 -		^							
´	-20.63	-14.48	-12.1	-18.26	Х	-19.86	-15.09	-6.54	Х
8 -	22.42	10.27	15.00	20.57	20.2	22.10	20.57	25.11	
	-23.43	-18.37	-15.93	-20.57	-28.2	-22.18	-20.57	-25.11	X
9 +	) 1	. 2	! 3	3 4	1 5	5 6	5 7	' 8	9

## Solution for 3.2:



0 ¬	Exercise 3.2 optimal value of all cells								
	49.18	49.18	49.26	49.26	49.18	49.26	53.61	54.01	54.46
1-	49.18	49.26	49.26	43.26	49.26	Х	48.01	54.46	54.95
2 -	43.26	Х	Х	Х	Х	Х	Х	54.95	62.17
3 -	37.86	43.09	49.18	49.18	49.18	49.26	49.26	Х	70.19
4 - 5 -	43.11	37.8	Х	Х	Х	Х	Х	70.19	79.1
	49.01	43.11	37.8	33.02	28.72	24.84	21.36	Х	89.0
6 -	55.57	Х	Х	Х	Х	Х	Х	G	100.0
7 -	62.85	64.28	65.87	67.63	Х	85.1	89.0	100.0	Х
8 -	64.28	65.87	67.63	69.59	78.43	81.59	85.1	89.0	Х
9 -	) 1	. 2	2 3	3 4	1 5	5 6	, <del>,</del>	7 8	3 9

Solution for 3.3:



Exercise 3.3 optimal value of all cells 0 -0.7 0.7 0.7 0.7 0.7 0.7 0.7 0.7 0.7 1 0.7 0.7 0.7 0.7 Х 0.7 0.7 0.7 0.7 2 · Х Х Х Х Х -0.05 Х 0.7 0.7 3 · -0.13 -0.05 0.7 0.7 0.7 0.7 0.7 Х 0.02 4 · -0.05 Х Х Х Χ -0.13 Χ 1.28 1.28 5 -0.05 -0.13 -0.14 Χ -0.05 0.02 1.28 12.5 12.5 6 0.7 Х Х Х Х Х Х G 12.5 7 0.7 0.7 0.7 0.7 Х 12.5 12.5 Χ 2.03 8 0.7 0.7 0.7

0.85

5

4

2.03

2.03

6

2.03

7

Х

8

#### Exercise 3.2 a n d 3.3 Comparison

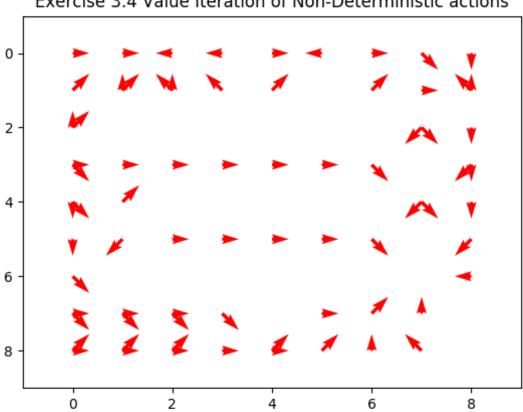
3

0.7

1

2

Compare to policy iteration algorithm of 3.2, we can clearly see that value iteration algorithm of 3.3 converge much earlyier than the algorithmn of 3.2 in terms of calculated optimal values. Moreover, as it is allowed to move in diagonal direction in 3.3, the value iteration algorithm choose a diagonal action when it can reach to the higher value states or the goal quicker, instead of traveling around(indirect way) like the optimal policy of 3.2 does.



Exercise 3.4 Value Iteration of Non-Deterministic actions

Exercise	3 /	ontimal	بتبلجيجا	a of al	احمااد
EXELLISE	0.4	obumai	vanu	- വ പ	i cens

	9.4	9.42	9.42	9.42	9.4	9.42	9.46	9.46	9.46
1-	9.42	9.42	9.42	9.42	9.42	Х	9.46	9.46	9.46
2 -	5.24	Х	Х	Х	Х	Х	Х	16.08	16.08
3 -	3.27	6.3	11.12	12.09	13.63	16.08	26.64	Х	26.64
4 -	3.7	6.3	Х	Х	Х	Х	Х	43.4	43.4
5 -	6.98	6.98	9.43	16.08	26.64	43.4	70.0	Х	70.0
6 -	12.19	Х	Х	Х	Х	Х	Х	G	70.0
7 -	12.19	13.8	16.35	20.4	Х	47.6	70.0	70.0	Х
8 -	12.19	13.8	16.35	20.4	33.49	47.6	47.6	47.6	Х
9 -		. 2	2 3	3 4	1 5	5 6	5 7	7 8	3 9

#### Exercise 3.3 and 3.4 Comparison

Compare to value iteration algorithm of 3.3, we can clearly see that value iteration algorithm of 3.4 converge much later than the algorithm of 3.3 in terms of calculated optimal values.

Moreover, as it is allowed to move in any desired direction in 3.4 and allowed to drift 45 degrees to the right or left with a probability, the value iteration algorithm chooses suitable action more freely to reach a higher value state or the goal quicker, instead of being bound by only desired or allowed actions like the optimal policy of 3.3 judging by the path our agent takes around the grid.