Faculty of Applied Sciences
University of Liege
February 12, 2019

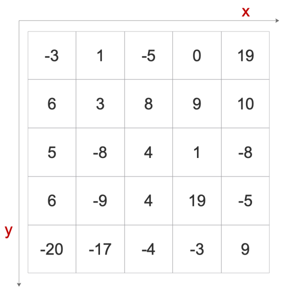
-

Optimal decision making for complex problems:

Assignment 1 - Section 1 to 3

s141770
Derroitte Natan

An important first task in order to start this project is to rigorously define the axes used. These will be widely used throughout the project as the agent's position and movements are defined along these axes. It has been chosen to position the point (0, 0) at the top left of the grid, the x-axis horizontally, directed to the right and the y-axis vertically, directed downwards. This is shown in figure 1.

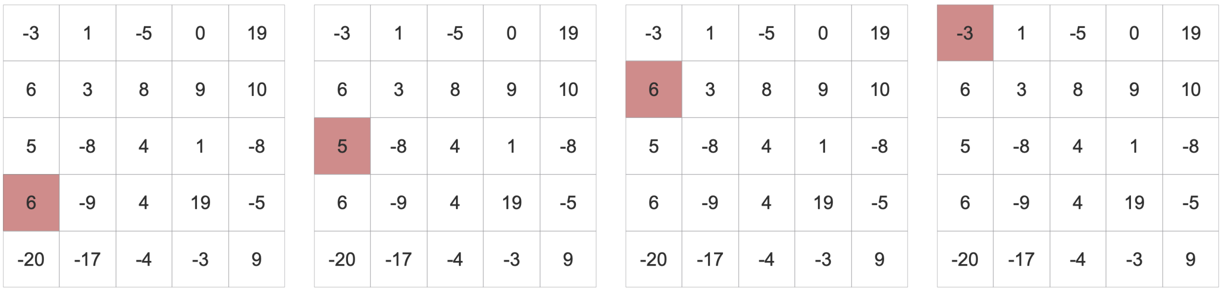


-3	1	-5	0	19
6	3	8	9	10
5	-8	4	1	-8
6	-9	4	19	-5
-20	-17	-4	-3	9

Figure 1: Representation of the axis system

1 Implementation of the domain

The first step of the project was to implement the domain presented in the statement. In order to do so, an agent following a simple policy of always moving upwards was first implemented.



-3	1	-5	0	19
6	3	8	9	10
5	-8	4	1	-8
6	-9	4	19	-5
-20	-17	-4	-3	9

-3	1	-5	0	19
6	3	8	9	10
5	-8	4	1	-8
6	-9	4	19	-5
-20	-17	-4	-3	9

-3	1	-5	0	19
6	3	8	9	10
5	-8	4	1	-8
6	-9	4	19	-5
-20	-17	-4	-3	9

-3	1	-5	0	19
6	3	8	9	10
5	-8	4	1	-8
6	-9	4	19	-5
-20	-17	-4	-3	9

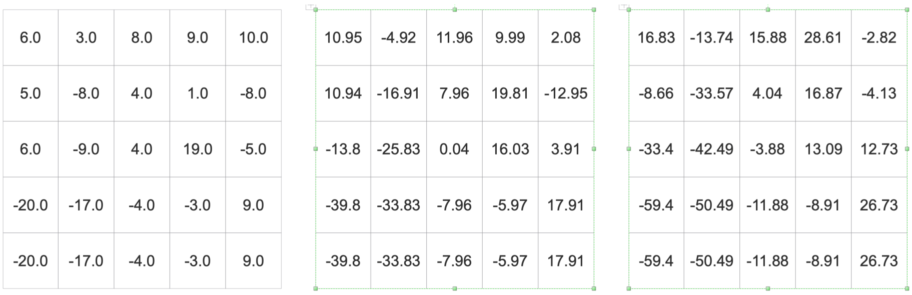
Figure 2: Movement of the agent when it follows a simple policy of always going up in deterministic mode

As expected, the figures 2 show the agent's movement in the grid.
Note that β is then 0 and that the program is therefore in deterministic mode. By increasing the value of β, the agent will have a $100\beta\%$ chance of directly reaching the cell (0, 0). It can therefore be understood that the difference between stochastic and deterministic mode only affects the first two iterations of the program for this question.

The sum of the rewards is also displayed when the code corresponding to this question is called. Knowing that the agent is blocked on a last negative reward cell, the sum of the rewards naturally tends towards -∞ when the number of iterations increases.

2 Expected return of a policy

The policy considered for this part of the project was to always carry out the action (0, 1), i. e. to go down according to the axis system defined in this report.
Therefore, following the definition of the expected return, the results calculated in this third question are shown in figures 3 and 4. The figures 3 first show the first iterations of the expected return.



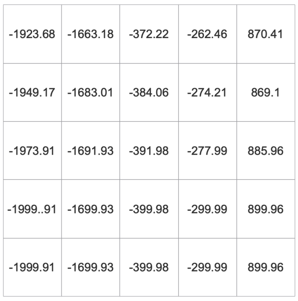
6.0	3.0	8.0	9.0	10.0
5.0	-8.0	4.0	1.0	-8.0
6.0	-9.0	4.0	19.0	-5.0
-20.0	-17.0	-4.0	-3.0	9.0
-20.0	-17.0	-4.0	-3.0	9.0

10.95	-4.92	11.96	9.99	2.08
10.94	-16.91	7.96	19.81	-12.95
-13.8	-25.83	0.04	16.03	3.91
-39.8	-33.83	-7.96	-5.97	17.91
-39.8	-33.83	-7.96	-5.97	17.91

16.83	-13.74	15.88	28.61	-2.82
-8.66	-33.57	4.04	16.87	-4.13
-33.4	-42.49	-3.88	13.09	12.73
-59.4	-50.49	-11.88	-8.91	26.73
-59.4	-50.49	-11.88	-8.91	26.73

Figure 3: Expected result for different number of iterations when $\beta = 0$. From left to right : $N = 1$, $N = 2$, $N = 3$.

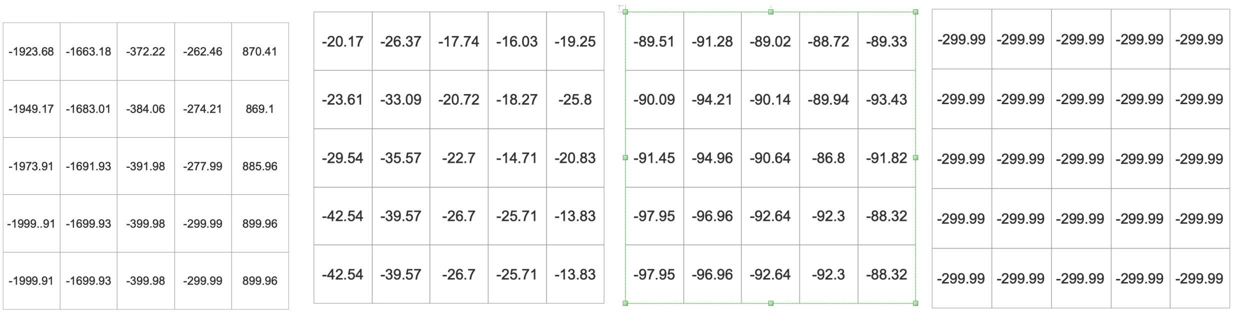
The figure 4 however illustrates the matrix of J_{μ}^{1000} when $\beta = 0$. The goal behind this calculation for a significant number of iterations is to tend towards J_μ. Indeed, $\lim_{N\to\infty} ||J_{\mu}^{N} - J_\mu||_\infty$. This can be verified while looking at the error bound and this will be done in the section 2.1.



-1923.68	-1663.18	-372.22	-262.46	870.41
-1949.17	-1683.01	-384.06	-274.21	869.1
-1973.91	-1691.93	-391.98	-277.99	885.96
-1999..91	-1699.93	-399.98	-299.99	899.96
-1999.91	-1699.93	-399.98	-299.99	899.96

Figure 4: The expected result matrix after 1000 iterations in deterministic setting.

The impact of the β parameter can now be studied: in the figure 5, the results of J_μ^{1000} are shown for different β values. It can be observed that when β is equals to 1, all the movements correspond to joining (0, 0). This results in homogeneous J_μ^N matrices.
To a lesser extent, it can be seen that the more β increases, the closer the values of $J_\mu^N(x)$ are to each other.



-1923.68	-1663.18	-372.22	-262.46	870.41
-1949.17	-1683.01	-384.06	-274.21	869.1
-1973.91	-1691.93	-391.98	-277.99	885.96
-1999..91	-1699.93	-399.98	-299.99	899.96
-1999.91	-1699.93	-399.98	-299.99	899.96

-20.17	-26.37	-17.74	-16.03	-19.25
-23.61	-33.09	-20.72	-18.27	-25.8
-29.54	-35.57	-22.7	-14.71	-20.83
-42.54	-39.57	-26.7	-25.71	-13.83
-42.54	-39.57	-26.7	-25.71	-13.83

-89.51	-91.28	-89.02	-88.72	-89.33
-90.09	-94.21	-90.14	-89.94	-93.43
-91.45	-94.96	-90.64	-86.8	-91.82
-97.95	-96.96	-92.64	-92.3	-88.32
-97.95	-96.96	-92.64	-92.3	-88.32

-299.99	-299.99	-299.99	-299.99	-299.99
-299.99	-299.99	-299.99	-299.99	-299.99
-299.99	-299.99	-299.99	-299.99	-299.99
-299.99	-299.99	-299.99	-299.99	-299.99
-299.99	-299.99	-299.99	-299.99	-299.99

Figure 5: Impact of β on the expected result for J_μ^{1000}. From left to right : $\beta = 0$, $\beta = 0.5$, $\beta = 0.75$, $\beta = 1$.

2.1 Error bound

The error bound can be calculated simply by using the following formula:

$$||J_\mu^N - J_\mu||_\infty \leq \frac{\gamma^N}{1-\gamma} Br \tag{1}$$

It appears that the approximation error depends on two parameters for a given grid: γ, the discount factor and N, the number of iterations.
The influence of the number of iterations on the error seems natural and can be studied in the following table:

	J_μ^5	J_μ^{100}	J_μ^{500}	J_μ^{752}	J_μ^{1000}
Error bound	1806.881	695.461	12.483	0.991	0.082

Table 1: Comparison of the error bound for different values of N, the number of iterations

Note that a large number of iterations is required to limit the error to a value less than 1, mainly due to the high value of γ, given in the statement.

Finally, it should be noted that it is possible to determine the number of iterations required to obtain a particular precision. By isolating N in the equation 1:

$$N = \log_\gamma \left(\frac{(1-\gamma)accuracy}{Br} \right)$$