HOMEWORK 3 – Partially Observable MDP

Exercise 1

a) **The problem described above can be modeled as an MDP. Identify the state space, X, and the action space, A, and write down the cost function for the MDP (you need not specify the transition probabilities). Consider throughout γ = 0.95.**

```
  0 1 2 3 4
0|_|_|_|_|_|
1|S|_|_|_|G|
2|_|_|_|_|_|
```

```
State Space:
 [(0, 0), (0, 1), (0, 2), (0, 3), (0, 4), (1, 0), (1, 1), (1, 2), (1, 3), (1, 4), (2, 0), (2, 1), (2, 2), (2, 3), (2, 4)]

Action Space:
 ('UP', 'RIGHT', 'DOWN', 'LEFT')

Cost Function:
 [[ 1.  1.  1.  1.]
 [ 1.  1.  1.  1.]
 [ 1.  1.  1.  1.]
 [ 1.  1.  1.  1.]
 [ 1.  1.  1.  1.]
 [ 1.  1.  1.  1.]
 [ 1.  1.  1.  1.]
 [ 1.  1.  1.  1.]
 [ 1.  1.  1.  1.]
 [ 1.  1.  1.  1.]
 [ 0.  0.  0.  0.]
 [ 1.  1.  1.  1.]
 [ 1.  1.  1.  1.]
 [ 1.  1.  1.  1.]
 [ 1.  1.  1.  1.]]
```

b) **Indicate the transition information resulting from the two actions of the agent (state, action, cost, next state).**

```
State:      (1, 0)
Action:     RIGHT
Cost:       1.0
New State:  (1, 1)
-----------------
State:      (1, 1)
Action:     RIGHT
Cost:       1.0
New State:  (0, 2)
```

Luís Morais 78416
João Rodrigues 78672

c) Suppose that the agent is following the Q-learning algorithm, with the Q-function initialized as an all-zeros function. Indicate the Q-values after the two Q-learning updates with step-size α = 0.1, resulting from the transitions in (b).

```
Updated Q-Matrix:
 [[ 0.   0.   0.   0. ]
 [ 0.   0.   0.   0. ]
 [ 0.   0.   0.   0. ]
 [ 0.   0.   0.   0. ]
 [ 0.   0.   0.   0. ]
 [ 0.   0.1  0.   0. ]
 [ 0.   0.1  0.   0. ]
 [ 0.   0.   0.   0. ]
 [ 0.   0.   0.   0. ]
 [ 0.   0.   0.   0. ]
 [ 0.   0.   0.   0. ]
 [ 0.   0.   0.   0. ]
 [ 0.   0.   0.   0. ]
 [ 0.   0.   0.   0. ]
 [ 0.   0.   0.   0. ]]
```

Luís Morais 78416
João Rodrigues 78672