M1 IRSI

4AI04 Intelligence Artificielle

TP - Processus Décisionnels Markovien

Auteur :
Stéphane Sobucki
3300032

2018-2019

Table des matières

1 Introduction

Nous allons dans ce TP nous familiariser avec les Processus Décisionnels Markoviens et étudier deux méthodes pour la recherche de politique optimale.

2 Itération de valeur

Question 1) En utilisant l'algorithme d'itération de valeur avec horizon infini, on converge vers la solution en 40 itérations. Néanmoins, en traçant l'estimation des utilités en fonction du nombre d'itérations, on se rend compte que les valeurs n'évoluent presque pas après une quinzaine d'itérations.

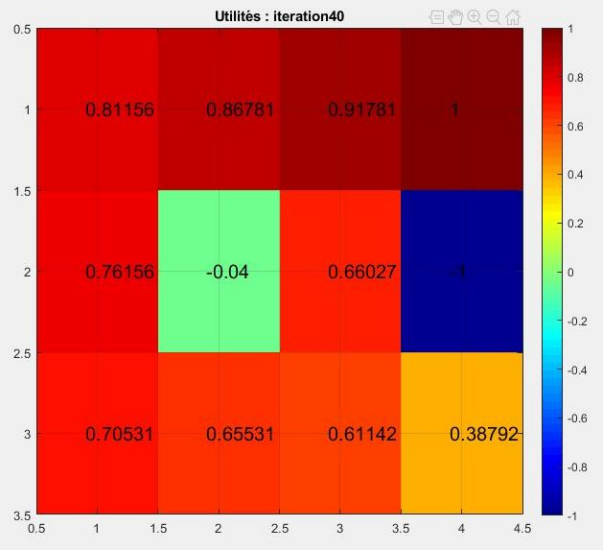

FIGURE 1 – Facteur d'escompte à 1

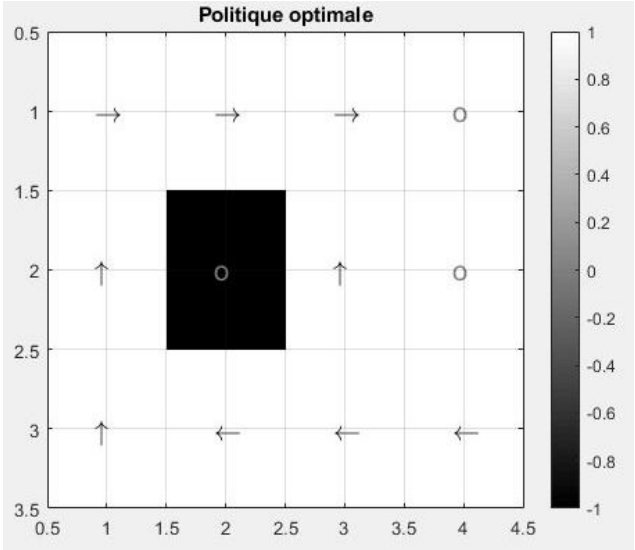

FIGURE 2 – Facteur d'escompte à 1

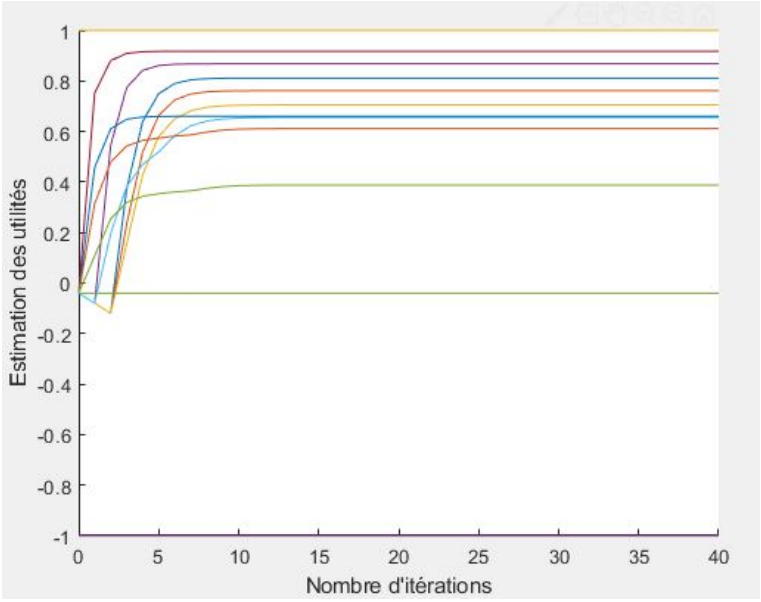

FIGURE 3 – Nombre d'itérations

Question 2) Nous allons étudier les différents paramètres qui influencent notre problème : le nombre d'itérations, le facteur d'escompte, l'initialisation des récompenses et l'influence de la case dangereuse. Nous traiterons ces problèmes séparément.

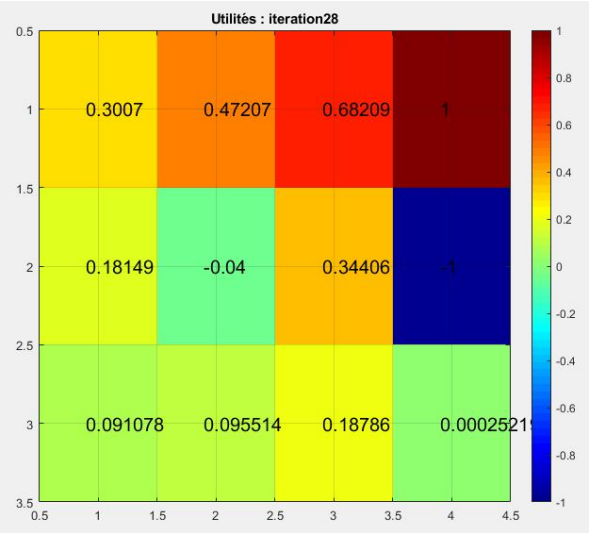

FIGURE 4 – Facteur d'escompte à 0.8

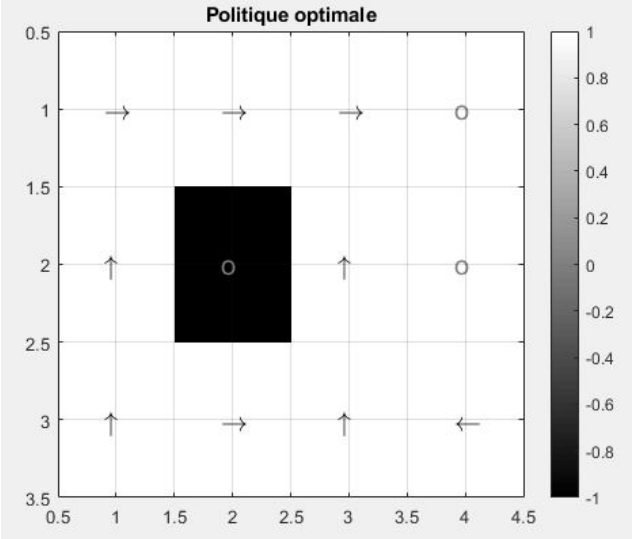

FIGURE 5 – Facteur d'escompte à 0.8

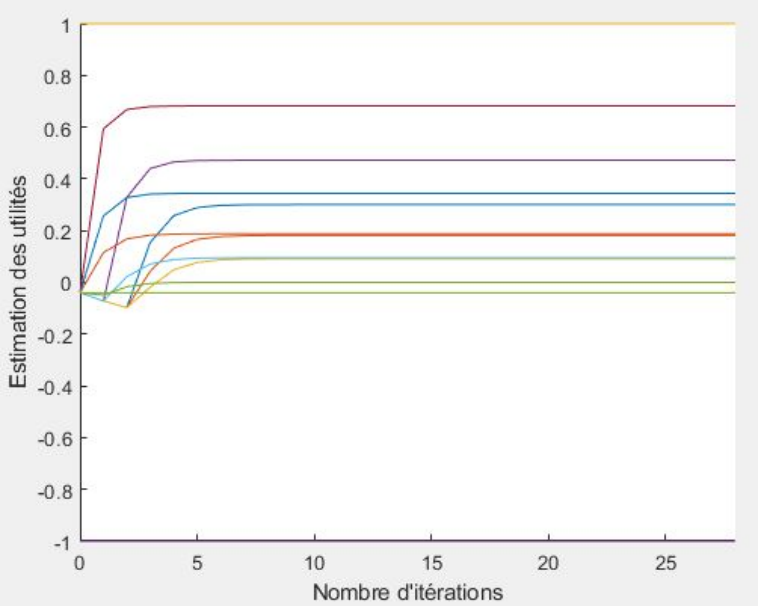

FIGURE 6 – Nombre d'itérations

En prenant le facteur d'escompte égal à 0.8 (les récompenses lointaines ont moins d'importance), on converge avec moins d'itérations (28) vers la solution. La solution est néanmoins différente de celle obtenue précédemment

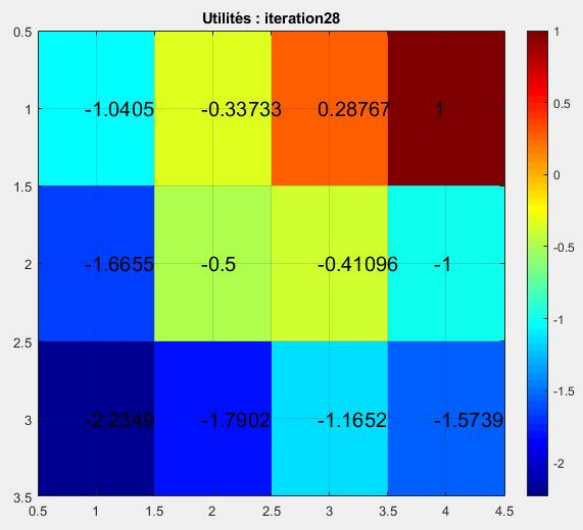

FIGURE 7 – Récompenses initialisées à -0.5

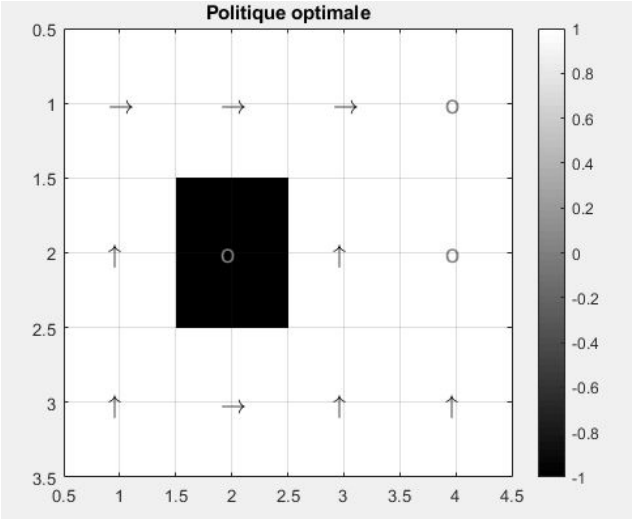

FIGURE 8 – Récompenses initialisées à -0.5

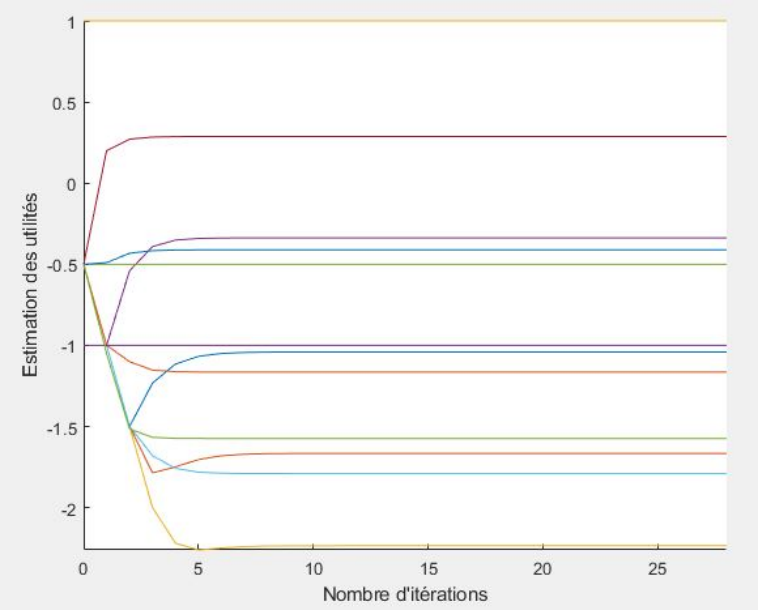

FIGURE 9 – Nombre d'itérations

Avec les récompenses initialisées à -0.5, on converge également vers la solution en moins d'itérations qu'avec l'initialisation à -0.4. Encore une fois, la solution obtenue est différente.

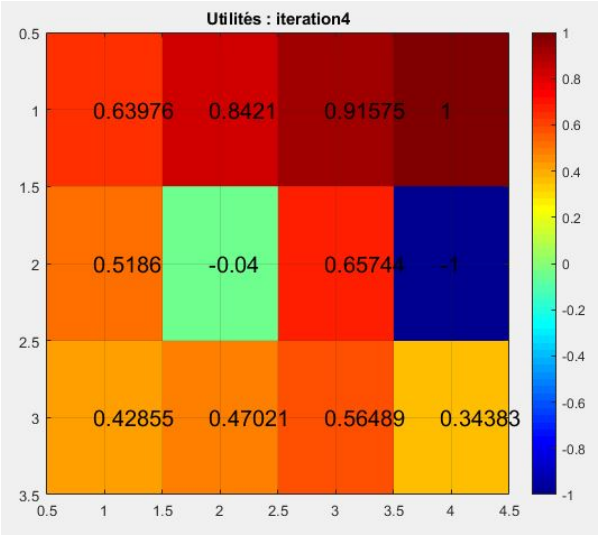

FIGURE 10 – 4 itérations

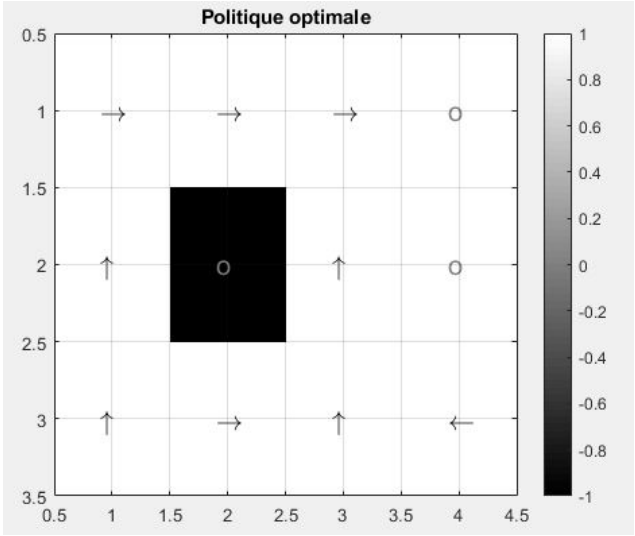

FIGURE 11 – 4 itérations

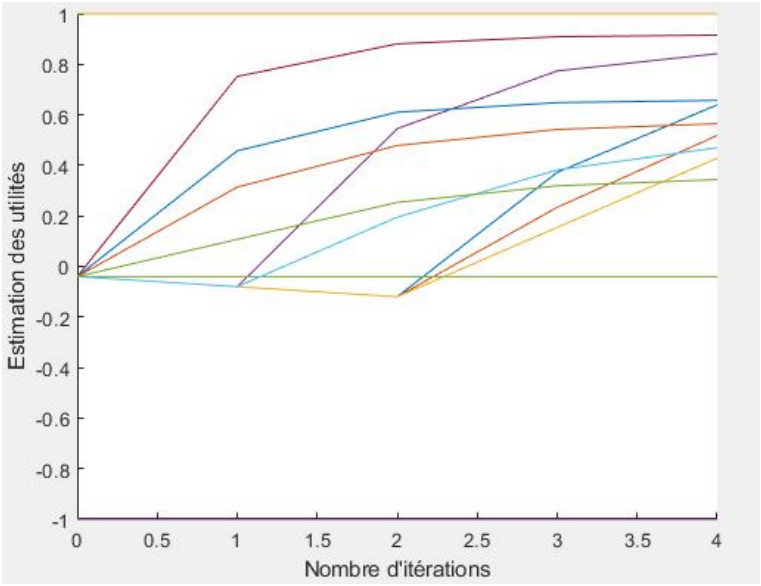

FIGURE 12 – Nombre d'itérations

En imposant la contrainte de 4 itérations maximales, on ne converge pas vers la solution optimale. En effet, on peut remarquer en traçant l'évolution des utilités en fonction du nombre d'itérations (figure 12) que leurs valeurs ne se sont pas encore stabilisées. La politique obtenue permet néanmoins d'atteindre la case associée à la récompense la plus élevée.

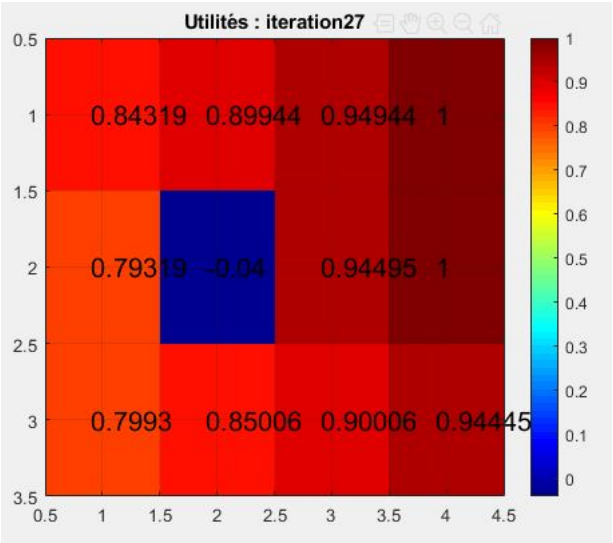

FIGURE 13 – Récompense R(11) = 1

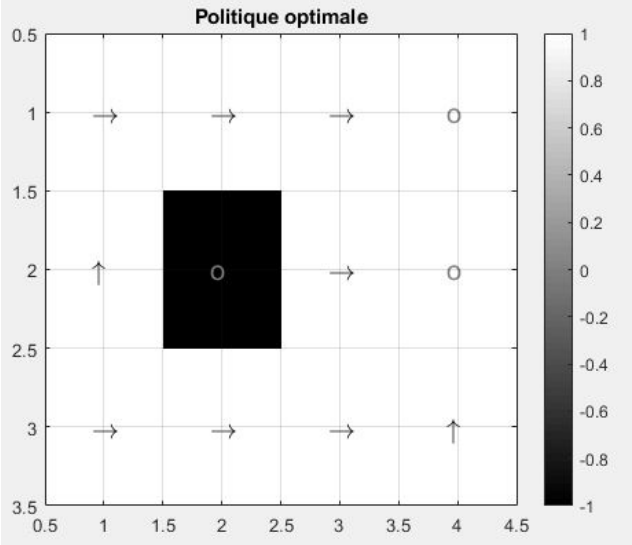

FIGURE 14 – Récompense R(11) = 1

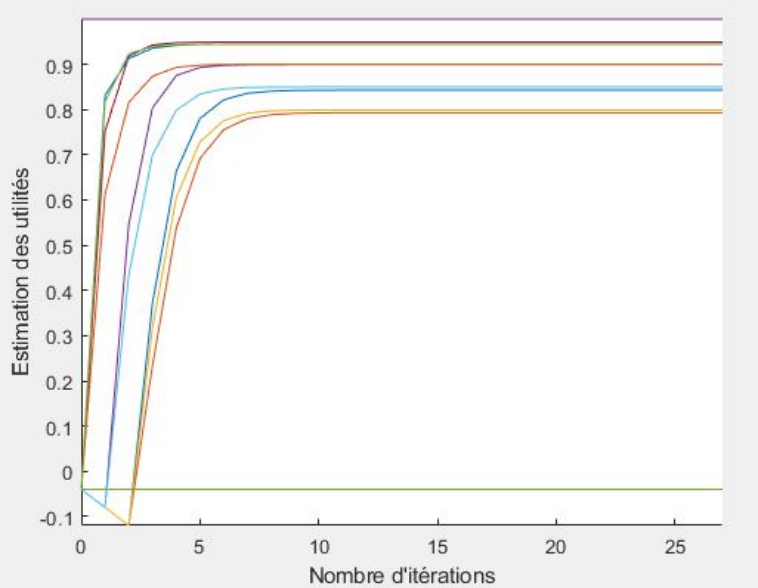

FIGURE 15 – Récompense R(11) = 1

En changeant la valeur de la case dangereuse, il y a maintenant cases "but". On remarque que les chemins mènent maintenant aux case 10 et 11. Si l'on veut atteindre une case précise, on ne peut donc pas avoir deux cases avec une grande récompense. On a également une convergence obtenue en moins d'itérations qu'avec la case dangereuse à -1.

3 Itération de politique

Question 3)

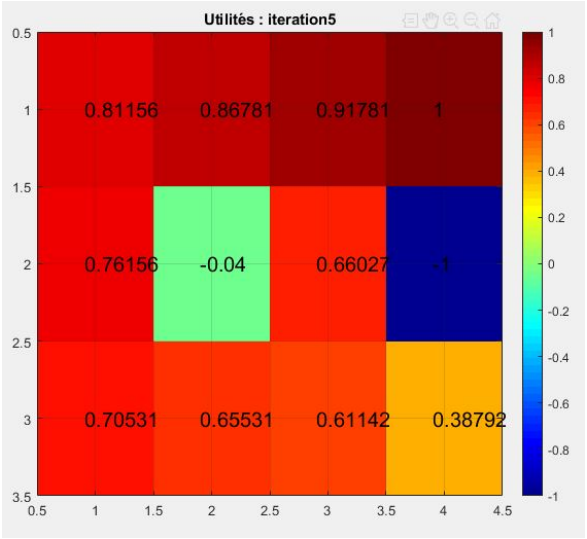

FIGURE 16 – Itération de politique

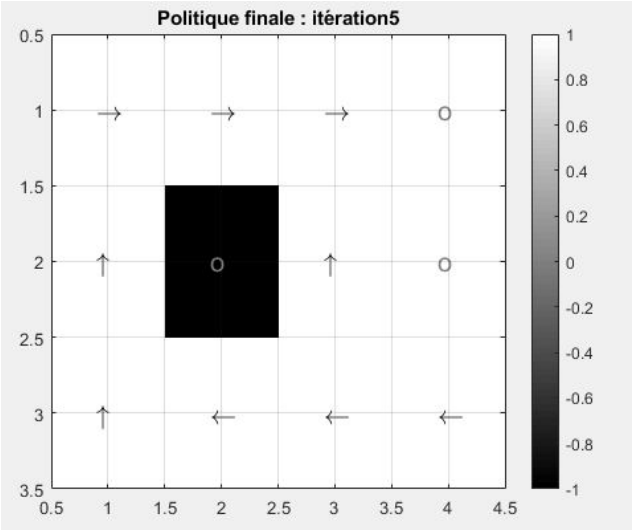

FIGURE 17 – Itération de politique

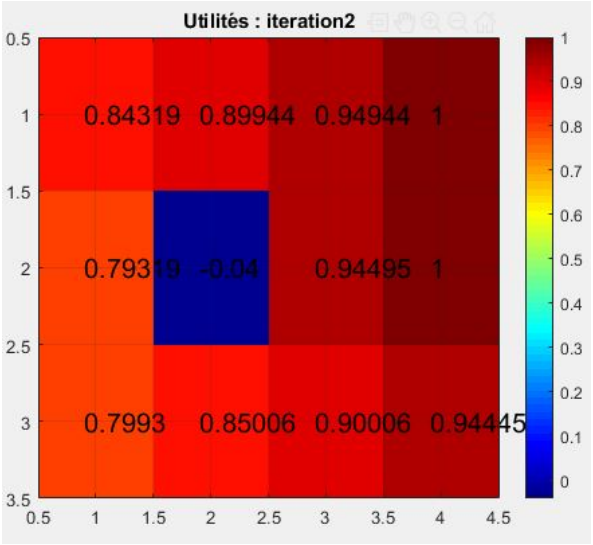

FIGURE 18 – Récompense R(11) = 1

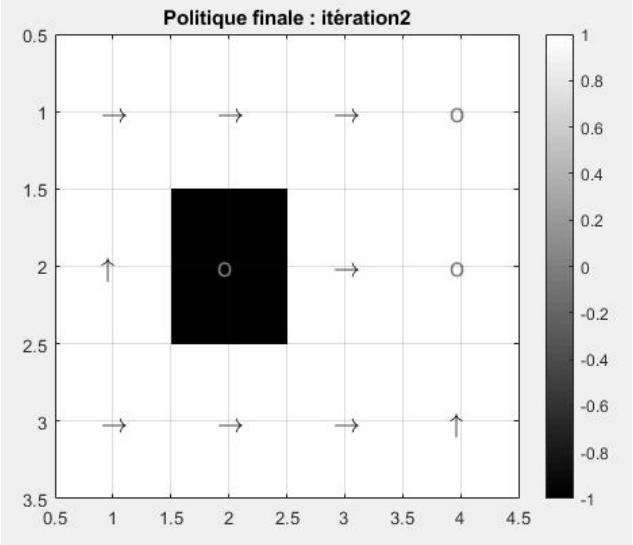

FIGURE 19 – Récompense R(11) = 1

Avec l'itération de politiques, on converge en beaucoup moins d'itérations à la solution, qui est la solution optimale, qu'avec l'itération de valeurs. On peut le vérifier avec, la récompense R(11) = 1. On converge vers la solution optimale en seulement 2 itérations contre 27 avec l'itération de valeurs.

4 Comparaison

Question 4) En comparant les temps de résolution des deux algorithmes (tic toc), on remarque qu'ils mettent plus ou moins autant de temps à converger. Bien que l'itération de politique converge ne moins d'itérations, elle nécessite d'inverser une matrice, ce qui peut être assez coûteux.