## Étienne Debard

## Corentin Maréchal

# Partie 5 - TP1 MDP

### Question 1 :

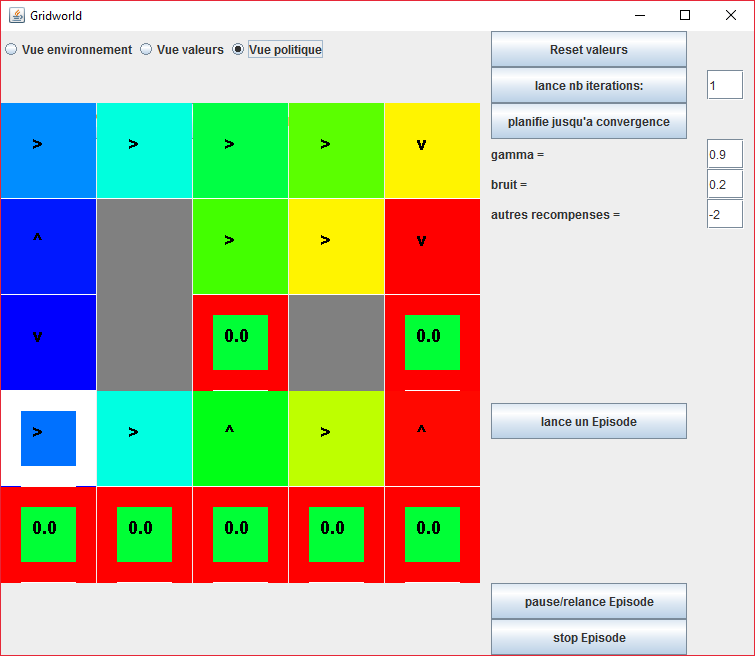
**Bruit = 0**

On se met dans un environnement détermiste où les dangers n’importent pas. Dans ce cas, on d’aller le plus rapidement vers la solution, donc vers l’autre côté du pont.

### Question 2 :

1. **Récompense = -2**

Avec une récompense négative entre -2 et -4, on choisit de partir sur le chemin dangereux. Du coup, étant proche de l’état déterministe avec un bruit à 0.2, ces dangers ne sont pas pris en compte et on se retrouve à prendre le chemin le plus court et plus dangereux. Néanmoins, si on dépasse un certain seuil (environ -4), on aura un meilleure résultat d’aller dans les éléments absorbant à -10.

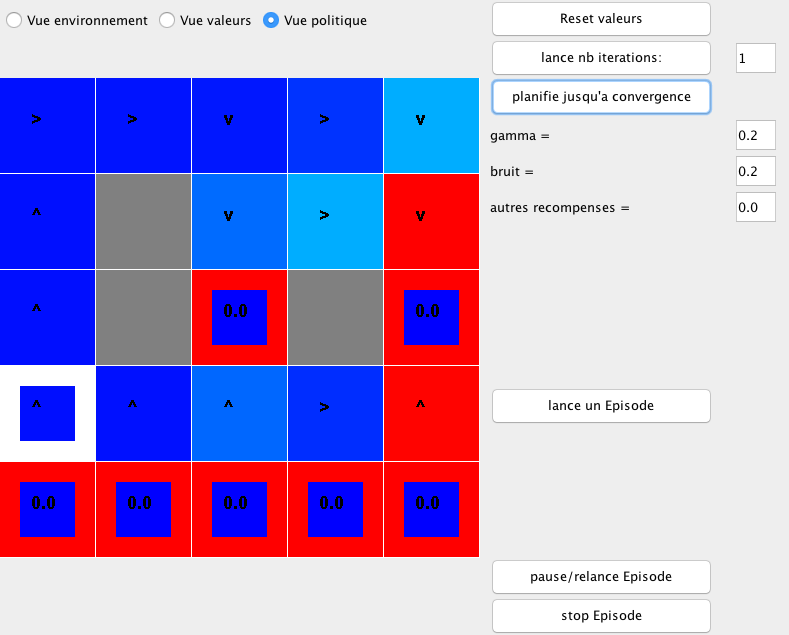


2. **Bruit = 0**  
Ici, on se met dans un environnement déterministe. Les dangers ne sont pas pris en compte et le choix de l’agent sera de prendre le plus court chemin vers la plus grande valeur (car gamme=0.9 donc champ de vision élevé).



3. **Gamma = 0.2**

Gamma permet de jouer sur le champ de vision. En le réduisant, l’agent va donc prendre le chemin le moins risqué (car bruit=0.2) et plus court (car gamma faible).



4. **Récompense = 2**  
Sachant que la valeur maximale des états d’absorbance est de 10, il faut donc que les valeurs de certains états soient supérieures à cette valeur maximale.

Avec les valeurs initiales, on sait que l’agent va prendre le chemin le moins risqué (bruit faible) et avec une vision grande (gamma élevé). Si on augmente la récompense, l’agent ira toujours dans cette zone où le risque est moindre et on pourra le faire « boucler » entre plusieurs cases, car les récompenses seront plus intéressantes que celle de l’état absorbant maximal.

