

# TP2 - ROB311 - Learning for Robotics

Sarah Curtit Caroline Pascal

September 2019

## 1 Question 1

All the possible policies are :

$$\begin{aligned}\Pi(S_0) &= \begin{cases} a_1 : P(S_1) = 1 \\ a_2 : P(S_2) = 1 \end{cases} \\ \Pi(S_1) &= \{ a_0 : P(S_1) = 1 - x, P(S_3) = x \\ \Pi(S_2) &= \{ a_0 : P(S_0) = 1 - y, P(S_3) = y \\ \Pi(S_3) &= \{ a_0 : P(S_0) = 1 \end{aligned}$$

## 2 Question 2

The optimal value functions for each state are given by the following formulas :

$$\begin{aligned}V^*(S_0) &= R(S_0) + \max_a \gamma \sum_{S'} T(S, a, S') V^*(S') \\ &= \max \gamma (V^*(S_1), V^*(S_2))\end{aligned}$$

$$\begin{aligned}V^*(S_1) &= R(S_1) + \max_a \gamma \sum_{S'} T(S, a, S') V^*(S') \\ &= \gamma \left( (1 - x) V^*(S_1) + x V^*(S_3) \right)\end{aligned}$$

$$\begin{aligned}V^*(S_2) &= R(S_2) + \max_a \gamma \sum_{S'} T(S, a, S') V^*(S') \\ &= 1 + \gamma \left( (1 - y) V^*(S_0) + y V^*(S_3) \right)\end{aligned}$$

$$\begin{aligned}V^*(S_3) &= R(S_3) + \max_a \gamma \sum_{S'} T(S, a, S') V^*(S') \\ &= 10 + \gamma V^*(S_0)\end{aligned}$$

### 2.1 Question 3

Let  $x, y$  and  $\gamma$  be  $y, x, \gamma \in [0, 1]$ . We have the following iterative relation on the utility of state  $S_2$ ,  $\forall t \in \mathbb{N}$  :

$$\begin{aligned}V_{t+1}(S_2) &= \gamma(1 - y)V_t(S_0) + \gamma y V_t(S_3) + 1 \\ &= \gamma(1 - y)V_t(S_0) + 10\gamma y + \gamma^2 y V_{t-1}(S_0) + 1 \\ V_{t+1}(S_2) &> \gamma(1 - y)V_t(S_0) + \gamma^2 y V_{t-1}(S_0)\end{aligned}$$

We also know that,  $\forall t \in \mathbb{N}$  :

$$V_{t+1}(S_0) = \gamma \max_a (V_t(S_1), V_t(S_2)) \geq \gamma V_t(S_1)$$

So, the previous inequality becomes :

$$V_{t+1}(S_2) > \gamma^2(1-y)V_{t-1}(S_1) + \gamma^3yV_{t-2}(S_1) \quad (1)$$

On the other hand, we also have the following iterative relation on the utility of state  $S_1$ ,  $\forall t \in \mathbb{N}$  :

$$V_{t+1}(S_1) = \gamma(1-x)V_t(S_1) + \gamma x V_t(S_3)$$

In the specific case where  $x = 0$ , this relation becomes :

$$V_{t+1}(S_1) = \gamma V_t(S_1)$$

We can therefore re-write the inequality 1 :

$$\begin{aligned} V_{t+1}(S_2) &> (1-y)V_t(S_1) + yV_t(S_1) \\ V_{t+1}(S_2) &> V_{t+1}(S_1) \end{aligned}$$

Therefore, knowing that  $\lim_{x \rightarrow 0} V_t = V^*$ , by reaching the limit on the value of  $t$ , we finally have :

$$V^*(S_2) > V^*(S_1)$$

Meaning :

$${}_a V^*(S_0) = a_2$$

This final inequality shows that for any value of  $y$  and  $\gamma \in [0, 1]$ , the choice  $\boxed{x = 0}$  ensures that  $\Pi^*(S_0) = a_2$ .

## 2.2 Question 4

Let's first suppose that  $\Pi^*(S_0) = a_1$ , and so that  $V^*(S_0) = \gamma V^*(S_1)$ , as  $V^*(S_1) > V^*(S_2)$ . In this case, the formulas of the second questions are turning into a linear system of 4 equations, with 4 unknowns, that can be easily solved :

$$\begin{cases} V(S_1) &= \frac{10\gamma x}{1-\gamma(1-x)-\gamma^3x} \\ V(S_2) &= \gamma^2(1-y)V(S_1) + 10\gamma y + \gamma^3yV(S_1) + 1 \\ V(S_0) &= \gamma V(S_1) \\ V(S_3) &= 10 + \gamma^2V(S_1) \end{cases}$$

The assumption  $V^*(S_1) > V^*(S_2)$  is then translated as :

$$\begin{aligned} V(S_1) &> \gamma^2(1-y)V(S_1) + 10\gamma y + \gamma^3yV(S_1) + 1 \\ V(S_1)(1 - \gamma^2(1-y) - \gamma^3y) &> 1 + 10\gamma y \end{aligned}$$

We have  $1 - \gamma^2(1-y) - \gamma^3y = (1-\gamma)(1+\gamma(1+y)) > 0$ , so, the inequality becomes :

$$\begin{aligned} 10\gamma x(1 - \gamma^2(1-y) - \gamma^3y) &> (1 + 10\gamma y)(1 - \gamma(1-x) - \gamma^3x) \\ 10\gamma y(\gamma - 1)(1 + \gamma x) &> (1 - \gamma)(1 + 9\gamma^2x + 9\gamma x) \end{aligned}$$

As  $\gamma - 1 < 0$ , we finally get :

$$y < \frac{9\gamma^2x + 9\gamma x - 1}{10\gamma(\gamma x + 1)}$$

As a matter of fact, the same procedure carried with the opposite assumption,  $\Pi^*(S_0) = a_2$ , with  $V^*(S_0) = \gamma V^*(S_2)$ , and  $V^*(S_2) > V^*(S_1)$ , leads to the opposite result.

Since the equivalence hasn't been broken during our calculations, we finally have the following relations :

$$\begin{cases} \Pi^*(S_0) = a_1 & \Longleftrightarrow y < \frac{9\gamma^2x+9\gamma x-1}{10\gamma(\gamma x+1)} \\ \Pi^*(S_0) = a_2 & \Longleftrightarrow y > \frac{9\gamma^2x+9\gamma x-1}{10\gamma(\gamma x+1)} \end{cases}$$

Therefore,  $\forall x, \gamma \in [0, 1]$ , if  $y < \frac{9\gamma^2x+9\gamma x-1}{10\gamma(\gamma x+1)}$ , we will have  $\Pi^*(S_0) = a_1$ .

Unfortunately, we can show that the minimum value reached by  $y_{crit} = \frac{9\gamma^2x+9\gamma x-1}{10\gamma(\gamma x+1)}$  is  $\frac{-1}{10} < 0$ , meaning that there are values of  $x$  and  $\gamma$  for which  $y_{crit}$  is negative, and in such case, there is no possible value of  $y$  leading to  $\Pi^*(S_0) = a_1$ .

For example, with  $x = 0.1$  and  $\gamma = 0.1$ ,

$$y_{crit} = \frac{9\gamma^2x+9\gamma x-1}{10\gamma(\gamma x+1)} \simeq -0.89$$

which cannot be reached for  $y \in [0, 1]$ .

Therefore, there's no value of  $y$  for which  $\Pi(S_0) = a_1 \forall x, \gamma \in [0, 1]$

N.B. We could have used the very same method to answer the previous question. For the value of  $x$ , we have the following relations :

$$\begin{cases} \Pi^*(S_0) = a_1 & \Longleftrightarrow x > \frac{10\gamma y+1}{\gamma(-10\gamma y+9\gamma+9)} \\ \Pi^*(S_0) = a_2 & \Longleftrightarrow x < \frac{10\gamma y+1}{\gamma(-10\gamma y+9\gamma+9)} \end{cases}$$

In this case, the minimum value of  $x_{crit} = \frac{10\gamma y+1}{\gamma(-10\gamma y+9\gamma+9)}$  is  $\frac{1}{18} > 0$ , as a consequence, if  $x = 0$ , we will be always in the case where  $x < x_{crit}$ , regardless of the values of  $\gamma$  and  $y$ , implying that  $\Pi^*(S_0) = a_2$ , as shown previously.

## 2.3 Question 5

Running our value iteration implementation, we obtain, for  $x = y = 0.25$  and  $\gamma = 0.9$  :

$$\begin{cases} V^*(S_0) & \simeq & 14.19 \\ V^*(S_1) & \simeq & 15.76 \\ V^*(S_1) & \simeq & 15.70 \\ V^*(S_1) & \simeq & 22.77 \end{cases}$$

$$\begin{cases} \Pi^*(S_0) & = & a_1 \\ \Pi^*(S_1) & = & a_0 \\ \Pi^*(S_1) & = & a_0 \\ \Pi^*(S_1) & = & a_0 \end{cases}$$

We could have computed these results by ourselves, noticing that for these values,  $y_{crit} \simeq 0.258 > y$ , meaning that  $\Pi^*(S_0) = a_1$ .

Knowing this, we could have simply used the equations established at the beginning of the question 4, with the proper values of  $x$ ,  $y$  and  $\gamma$ . After verification, the values obtained by the two methods are the same!