Trabalho para casa 2 - Markov Decision Problems

a) A MDP é caracterizada pelo par ($\mathscr{X}$**;** **P**) em que $\mathscr{X}$ é o conjunto de estados possíveis tal que $\mathscr{X}$ = {1, 2, 3, 4, 5, 6} e **A =** { up; down**;** left; right}.

b) As probabilidades de transição para cada uma das acções são:

$$P_{up} = \begin{bmatrix} 0.8 & 0.1 & 0 & 0.1 & 0 & 0 \\ 0.1 & 0.7 & 0.1 & 0 & 0.1 & 0 \\ 0 & 0.1 & 0.8 & 0 & 0 & 0.1 \\ 0.6 & 0 & 0 & 0.3 & 0.1 & 0 \\ 0 & 0.6 & 0 & 0.1 & 0.2 & 0.1 \\ 0 & 0 & 0.6 & 0 & 0.1 & 0.3 \end{bmatrix} \quad P_{down} = \begin{bmatrix} 0.3 & 0.1 & 0 & 0.6 & 0 & 0 \\ 0.1 & 0.2 & 0.1 & 0 & 0.6 & 0 \\ 0 & 0.1 & 0.3 & 0 & 0 & 0.6 \\ 0.1 & 0 & 0 & 0.8 & 0.1 & 0 \\ 0 & 0.1 & 0 & 0.1 & 0.7 & 0.1 \\ 0 & 0 & 0.1 & 0 & 0.1 & 0.8 \end{bmatrix}$$

$$P_{left} = \begin{bmatrix} 0.8 & 0.1 & 0 & 0.1 & 0 & 0 \\ 0.6 & 0.2 & 0.1 & 0 & 0.1 & 0 \\ 0 & 0.6 & 0.3 & 0 & 0 & 0.1 \\ 0.1 & 0 & 0 & 0.8 & 0.1 & 0 \\ 0 & 0.1 & 0 & 0.6 & 0.2 & 0.1 \\ 0 & 0 & 0.1 & 0 & 0.6 & 0.3 \end{bmatrix} \quad P_{right} = \begin{bmatrix} 0.3 & 0.6 & 0 & 0.1 & 0 & 0 \\ 0.1 & 0.2 & 0.6 & 0 & 0.1 & 0 \\ 0 & 0.1 & 0.8 & 0 & 0 & 0.1 \\ 0.1 & 0 & 0 & 0.3 & 0.6 & 0 \\ 0 & 0.1 & 0 & 0.1 & 0.2 & 0.6 \\ 0 & 0 & 0.1 & 0 & 0.1 & 0.8 \end{bmatrix}$$

Em relação à função de custo, uma possibilidade é penalizar consoante o estado em que o jogador acaba (1 no dragão, 0 na princesa, e 0.05 no resto):

$$C = \begin{bmatrix} 0.05 & 0.05 & 0.05 & 0.05 \\ 0.05 & 0.05 & 0.05 & 0.05 \\ 0.05 & 0.05 & 0.05 & 0.05 \\ 0.05 & 0.05 & 0.05 & 0.05 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Grupo 3
Daniel Correia, 80967
Pedro Orvalho, 81151

c) Com a policy "The knight always goes up", a cost-to-go function (J^π) calcula-se por:

$$J^\pi = C_\pi + \gamma P_\pi J^\pi$$

$$J^\pi = (I - \gamma P_\pi)^{-1} C_\pi$$

Em que I é a matriz identidade, gamma é igual a 0.9, e sendo a policy ir sempre para cima usamos a matriz de transição de probabilidades da ação "up" (P_π) e o custo associado à acção up (C_π= C[:,0]):

$$\gamma = 0.9$$

$$C_\pi = \begin{bmatrix} 0.05 \\ 0.05 \\ 0.05 \\ 0.05 \\ 1 \\ 0 \end{bmatrix} \qquad P_\pi = \begin{bmatrix} 0.8 & 0.1 & 0 & 0.1 & 0 & 0 \\ 0.1 & 0.7 & 0.1 & 0 & 0.1 & 0 \\ 0 & 0.1 & 0.8 & 0 & 0 & 0.1 \\ 0.6 & 0 & 0 & 0.3 & 0.1 & 0 \\ 0 & 0.6 & 0 & 0.1 & 0.2 & 0.1 \\ 0 & 0 & 0.6 & 0 & 0.1 & 0.3 \end{bmatrix} \qquad I = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Logo,

$$J^\pi = \left(\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} - 0.9 \begin{bmatrix} 0.8 & 0.1 & 0 & 0.1 & 0 & 0 \\ 0.1 & 0.7 & 0.1 & 0 & 0.1 & 0 \\ 0 & 0.1 & 0.8 & 0 & 0 & 0.1 \\ 0.6 & 0 & 0 & 0.3 & 0.1 & 0 \\ 0 & 0.6 & 0 & 0.1 & 0.2 & 0.1 \\ 0 & 0 & 0.6 & 0 & 0.1 & 0.3 \end{bmatrix} \right)^{-1} \begin{bmatrix} 0.05 \\ 0.05 \\ 0.05 \\ 0.05 \\ 1 \\ 0 \end{bmatrix}$$

$$J^\pi = \begin{bmatrix} 0.80525055 \\ 1.02797112 \\ 0.77636737 \\ 0.92169726 \\ 2.08893003 \\ 0.83183847 \end{bmatrix}$$

Grupo 3
Daniel Correia, 80967
Pedro Orvalho, 81151