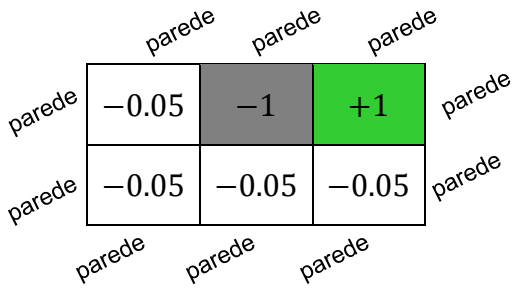
Exercício E1 – Processos de Decisão de Markov (MDPs)

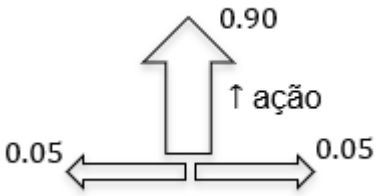
1) Dado um ambiente do tipo *Grid World* composto por 6 casas em duas fileiras, conforme a figura abaixo:

Considere um robô móvel na posição s_0 do ambiente com capacidade de se movimentar nas 4 direções: $\mathcal{A} = \{N = \uparrow, E = \rightarrow, S = \downarrow, W = \leftarrow\}$. A posição de destino do robô é representada pela casa verde, enquanto a casa cinza representa um obstáculo (ambos são estados terminais, qualquer ação tomada mantém o agente no próprio estado). Uma Função de Recompensa $\mathcal{R}$ que representa a tarefa de posicionamento do robô é ilustrada abaixo, indicando a recompensa obtida por um agente ao tomar qualquer ação na casa correspondente ($\mathcal{R}(s,.)$):



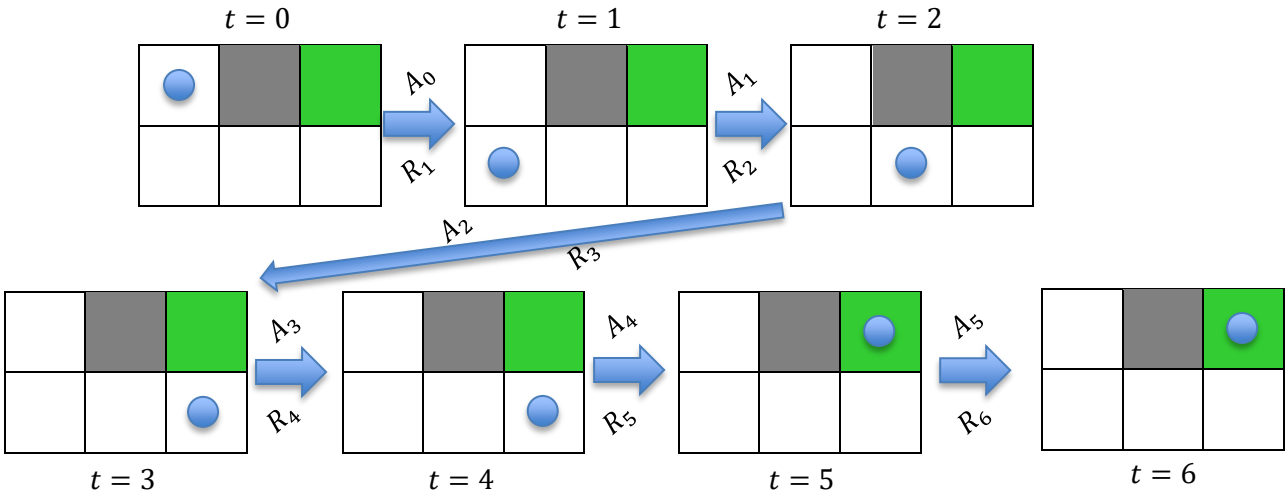
Quando o robô executa uma ação, ele se move para a casa vizinha na direção escolhida em 90% das vezes, com 5% de chance de ocorrer um escorregamento em cada direção perpendicular à ação tomada. Se o robô fosse colidir com uma parede externa do ambiente ele em vez disso permanece na mesma posição.

Todas as ações podem ser tomadas em qualquer estado, somente após o agente escolher a ação tomada que a função de probabilidades de transição do ambiente ($\mathcal{P}_{ss'}^{a}$) determina o estado seguinte de acordo com o escorregamento.



a) Faça um diagrama parcial do MDP associado a esse problema, representando todas as transições de estados, ações e recompensas **a partir do estado inicial apenas**. Nomeie os demais estados e indique as recompensas e probabilidades de transição associadas a cada uma das 4 ações tomadas no estado inicial.

b) Considerando um fator de desconto $\gamma = 0.9$ e o seguinte episódio, determine o retorno G_0 obtido a partir do estado inicial. Considere o MDP com horizonte T finito.



c) Considere um agente aleatório π_{rand} $(\pi(s,a) = 0.25, \forall a)$ e a função Valor dos estados V_π indicada abaixo. Utilize a equação de Bellman para calcular o valor do estado s_0.

V_{s_0}	-1	1
-0.598	-0.495	0.116

Função Valor dos Estados $V_\pi(s)$ para agente aleatório.