Assign 4.

P1

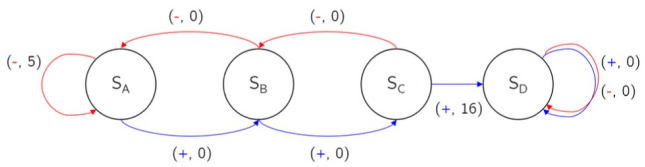

Figure 1: MDP problem with (action, reward) pairs.

episode length 3

1a i) $\pi(S_A) = -$

$$U = \begin{cases} 5 + \gamma 5 + \gamma^2 5 \\ 5 + \gamma 5 + 0 \\ 5 \\ 5 \end{cases}$$

$$V_\pi = \frac{1}{4}(20 + 10\gamma + 5\gamma^2)$$

ii) $\pi(S_A) = +$

$$U = \begin{cases} \gamma^2 16 \\ 0 \\ 0 \\ \gamma^2 5 \end{cases}$$

$$V_\pi = \frac{21}{4}\gamma^2$$

Optimal policy is $(-)$ when $\gamma = 0.001$

1b 위의 식에서 $\gamma = 0.999$ 를 대입하면

i) $8.74 > $ ii) 5.24

$\therefore$ $(-)$

1c i) $\pi(S_B) = -$

$$U = \begin{cases} 0 \\ \gamma 5 \\ 0 \\ \gamma 5 + \gamma^2 5 \end{cases} \qquad V_\pi = \frac{1}{4}(10\gamma + 5\gamma^2)$$

ii) $\pi(S_B) = +$

$$U = \begin{cases} 0 \\ 0 \\ \gamma 16 \\ \gamma 16 \end{cases} \qquad V_\pi = 8\gamma$$

$$V_{\pi(i)} - V_{\pi(ii)} = \frac{5}{2}\gamma + \frac{5}{4}\gamma^2 - 8\gamma$$

$$= \frac{5}{4}\gamma\left(\gamma - \frac{22}{5}\right) < 0$$

when $\gamma \in (0, 1)$

$\therefore$ Optimal policy is always $(+)$