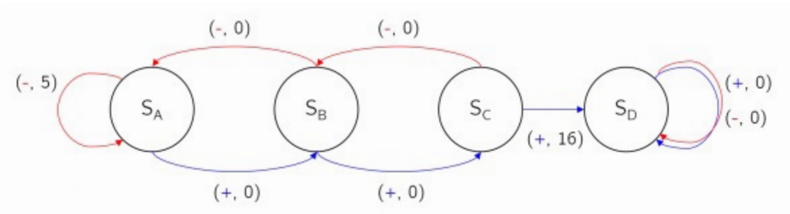

Using value iteration

$$V_{opt}^{(t)}(s) \leftarrow \max_{a \in Actions(s)} \sum_{s'} T(s,a,s')\left[Reward(s,a,s') + \gamma V_{opt}^{(t-1)}(s')\right]$$

prob 1a)

① $V_{opt}^{(1)}(S_A) = \max\begin{cases} a=+ \rightarrow 0+0.001\times 0 \\ a=- \rightarrow 5+0.001\times 0 \end{cases} = 5 \quad (-)$

$V_{opt}^{(1)}(S_B) = \max\begin{cases} a=+ \rightarrow 0+0.001\times 0 \\ a=- \rightarrow 0+0.001\times 0 \end{cases} = 0$

$V_{opt}^{(1)}(S_C) = \max\begin{cases} a=+ \rightarrow 16+0.001\times 0 \\ a=- \rightarrow 0+0.001\times 0 \end{cases} = 16 \quad (+)$

$V_{opt}^{(1)}(S_d) = \max\begin{cases} 0 \\ 0 \end{cases} = 0$

② $V_{opt}^{(2)}(S_A) = \max(0, 5+0.005) = 5.005 \quad (-)$

$V_{opt}^{(2)}(S_B) = \max(0+0.016, 0+0.005) = 0.016 \quad (+)$

$V_{opt}^{(2)}(S_C) = \max(16, 0) = 16 \quad (+)$

$V_{opt}^{(2)}(S_D) = \max(0, 0) = 0$

③ $V_{opt}^{(3)}(S_A) = \max(0+0.016\times 0.01, 5+0.01\times 5.005) \approx 5.005 \quad (-)$ almost same

$V_{opt}^{(3)}(S_B) = \max(0.001\times 16, 0.01\times 5.005) = 0.016 \quad (+)$ same

$V_{opt}^{(3)}(S_C) = \max(16, 0.01\times 5.005) = 16 \quad (+)$ same

$V_{opt}^{(3)}(S_D) = 0$ same

$\Rightarrow \therefore$ optimal policy for S_A: $(-)$

prob 1 b)

① $V_{opt}^{(1)}(S_A) = \max\begin{Bmatrix} a=+ \rightarrow 0+0.999\times 0 \\ a=- \rightarrow 5+0.999\times 0 \end{Bmatrix} = 5$ (−)

$V_{opt}^{(1)}(S_B) = \max\begin{Bmatrix} a=+ \rightarrow 0+0.999\times 0 \\ a=- \rightarrow 0+0.999\times 0 \end{Bmatrix} = 0$

$V_{opt}^{(1)}(S_C) = \max\begin{Bmatrix} a=+ \rightarrow 16+0.999\,0 \\ a=- \rightarrow 0+0.999\times 0 \end{Bmatrix} = 16$ (+)

$V_{opt}^{(1)}(S_d) = \max\begin{Bmatrix} 0 \\ 0 \end{Bmatrix} = 0$

② $V_{opt}^{(2)}(S_A) = \max(0,\ 5+0.999\times 5) = 9.995$ (−)

$V_{opt}^{(2)}(S_B) = \max(0+0.999\times 16,\ 0+0.999\times 5) = 15.984$ (+)

$V_{opt}^{(2)}(S_C) = \max(16,\ 0) = 16$ (+)

$V_{opt}^{(2)}(S_D) = \max(0, 0) = 0$

③ $V_{opt}^{(3)}(S_A) = \max(0+0.999\times 15.984,\ 0.999\times 9.995) = 15.968016$ (+)

$V_{opt}^{(3)}(S_B) = \max(0+0.999\times 16+0.0999\times 9.995) = 15.984$ (+) same

$V_{opt}^{(3)}(S_C) = \max(16,\ 0+0.999\times 15.984) = 16$ (+) same

$V_{opt}^{(3)}(S_D) = 0$ same

④ $V_{opt}^{(4)}(S_A) = \max(0+0.999\times 15.984,\ 0.999\times 15.968016) = 15.968016$ (+) same.

∴ optimal policy for S_A : (+)

prob 1 c)

$$V_{opt}^{(t)}(S_A) = \max\left(0 + \gamma V_{opt}^{(t-1)}(S_B),\ 5 + \gamma V_{opt}^{(t-1)}(S_A)\right)$$

$$V_{opt}^{(t)}(S_B) = \max\left(0 + \gamma V_{opt}^{(t-1)}(S_C),\ 0 + \gamma V_{opt}^{(t-1)}(S_A)\right)$$

$$V_{opt}^{(t)}(S_C) = \max\left(16 + \gamma V_{opt}^{(t-1)}(S_D) + 0 + \gamma V_{opt}^{(t-1)}(S_B)\right) = \max\left(16,\ \gamma V_{opt}^{(t-1)}(S_B)\right)$$

$$V_{opt}^{(t)}(S_D) = 0$$

$\Rightarrow$ ①

$V^{(1)}_{opt}(S_A) = 5$

$V^{(2)}_{opt}(S_B) = 0$

$V^{(3)}_{opt}(S_C) = 16$

$V^{(4)}_{opt}(S_D) = 0$

②

$V^{(2)}_{opt}(S_A) = (1+\gamma)5$

$V^{(2)}_{opt}(S_B) = 16\gamma$

$V^{(3)}_{opt}(S_C) = 16$

$V^{(4)}_{opt}(S_D) = 0$

③

$$V^{(3)}_{opt}(S_A) = \max\left(16\gamma^2,\ 5(1+\gamma(1+\gamma))\right)$$

$$V^{(3)}_{opt}(S_B) = \max\left(16\gamma,\ \gamma(1+\gamma)5\right)$$

$$V^{(3)}_{opt}(S_C) = \max\left(16,\ 16\gamma^2\right)$$

$$V^{(4)}_{opt}(S_D) = 0$$

...

because 16γ is always bigger than $\gamma V_{opt}^{(t-1)}(S_A)$,

optimal policy of S_B is $(+)$