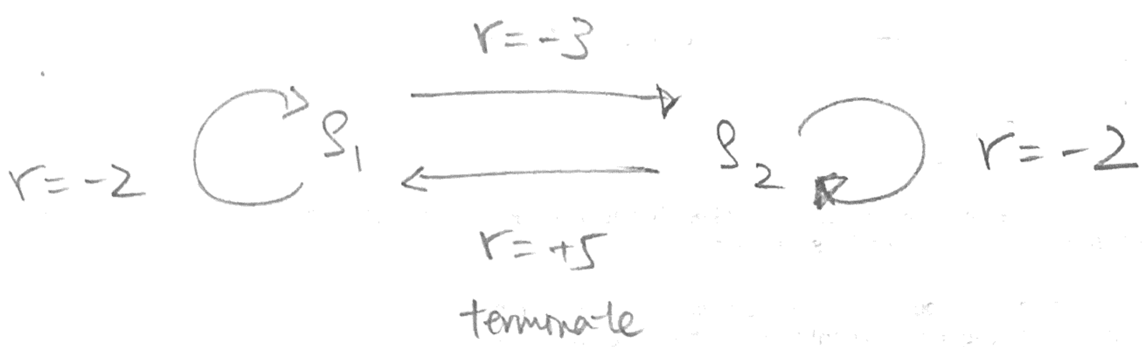
1] $S = \{s_1, s_2\}$ $A = \{\text{stay}, \text{go}\}$



a] $$\sum_{t=0}^{\infty} \gamma^t r_t(s_t, a_t) = \sum_{t=0}^{\infty} \gamma^t r_t(s_1, \text{stay})$$

$$= \sum_{t=0}^{\infty} \gamma^t (-2),$$ since $0 < \gamma \leq 1$, this is a geometric series.

$$= \boxed{\frac{-2}{1-\gamma}}$$

b] Consider π_e that goes to terminate ASAP.

i.e. $\pi_e(s_1) = \text{go}$, $\pi_e(s_2) = \text{go}$.

Now consider its sum of discounted rewards.

$$\sum_{t=0}^{\infty} \gamma^t r_t(s_t, a_t) = \gamma^0 r_0(s_1, go)$$

$$+ \gamma^1 r_1(s_2, go)$$

$$= \underline{-3 + \gamma(5)}$$

Let's compare this with result from 1a. (π_{1a})

We can use this π_e if the following holds.

$$-3 + \gamma(5) - \left(\frac{-2}{1-\gamma}\right) \geq 0$$

$$\frac{(-3+5\gamma)(1-\gamma)+2}{1-\gamma} \geq 0$$

$\xrightarrow{}$ denominator is always >0 since $\gamma < 1$.

So inequality depends on numerator.

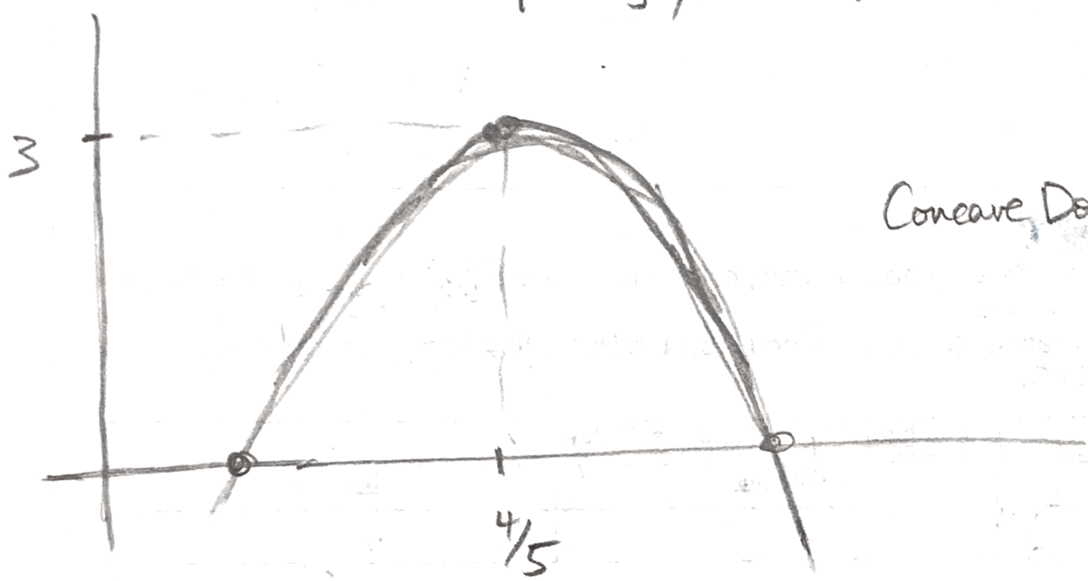
$$\text{numerator} = -3 + \cancel{5\gamma} + \cancel{3}\gamma - 5\gamma^2 + 2$$

$$f(\gamma) = -5\gamma^2 + 8\gamma - 1$$

$$f(\gamma) = -5\left(\gamma^2 - \frac{8}{5}\gamma + \frac{4}{5}\right) + 3$$

$$= -5\left(\gamma - \frac{4}{5}\right)^2 + 3.$$

$\therefore$ when $\gamma = \frac{4}{5}$, $f(\gamma) = 3$



Concave Down, since -5

need to find roots to determine when $f(\gamma) \geq 0$; this is when we should use π_e, o.w. when $f(\gamma) < 0$, use π_{1a}

$$\text{roots} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{-8 \pm \sqrt{64 - 4(-5)(-1)}}{-10}$$

$$= \frac{-8 \pm 2\sqrt{11}}{-10} = \frac{4 \pm \sqrt{11}}{5} \approx (.137, 1.46)$$

$$\text{Optimal Policy} = \begin{cases} (go, go) & \text{if } .137 \leq \gamma \leq 1.46 \\ (stay, stay, \ldots, stay) & \text{o.w.} \end{cases}$$

1c) $V^0 = [0, 0]$

$$V^1(s_1) = \max\{ r(s_1, stay) + V^0(s_1), \; r(s_1, go) + V^0(s_2) \}$$

$$= \max\{ -2 + 0, -3 + 0 \} = -2$$

$$V^1(s_2) = \max\{ r(s_2, stay) + V^0(s_2), r(s_2, go) \}$$

$$= \max\{ -2 + 0, 5 \} = 5$$

$\therefore V^1 = [-2, 5]$

$$V^2(s_1) = \max\{ r(s_1, stay) + V^1(s_1), r(s_1, go) + V^1(s_2) \}$$

$$= \max\{ -2 - 2, -3 + 5 \} = 2$$

$$V^2(s_2) = \max\{ r(s_2, stay) + V^1(s_2); r(s_2, go) \}$$

$$= \max\{ -2 + 5, 5 \} = 5$$

$\therefore V^2 = [2, 5]$

1c)

$V^3(s_1) = \max\{r(s_1, stay) + V^2(s_1),\ r(s_1, go) + V^2(s_2)\}$

$= \max\{-2+2,\ -3+5\} = 2$

$V^3(s_2) = \max\{r(s_2, stay) + V^2(s_2),\ r(s_2, go)\} = 5$

$\underbrace{V^3 = [2, 5]}$

$\boxed{\therefore V^* = [2, 5]}$

2) a) $\|v^i - V^*\|_\infty = \max_{s \in S} |v^i(s) - V^*(s)|$

$i=1;$ $\max\{\underbrace{|v^1(s_1) - V^*(s_1)|}_{4}, \underbrace{|v^1(s_2) - V^*(s_2)|}_{0}\} = 4$

$i=2;$ $\max\{|v^2(s_1) - V^*(s_1)|; |v^2(s_2) - V^*(s_2)|\} = 0$

$i=3;$ $\max\{|v^3(s_1) - V^*(s_1)|, |v^3(s_2) - V^*(s_2)|\} = 0$

b) $T(v) = \max_a \sum_{s'} p(s'|s,a)[r(s,a) + \gamma v(s')]$

$T(v') = \max_a \sum_{s'} p(s'|s,a)[r(s,a) + \gamma v'(s')]$

$\|Tv - Tv'\|_\infty$

$= \|\max_a \sum_{s'} p(s'|s,a)[r(s,a) + \gamma v(s')] - \max_a \sum_{s'} p(s'|s,a)$
$[r(s,a) + \gamma v'(s')]\|_\infty$

$\leq \max_a \|\sum_{s'} p(s'|s,a)[r(s,a) + \gamma v(s')] - \sum_{s'} p(s'|s,a)$
$[r(s,a) + \gamma v'(s')]\|_\infty$ $\quad \|\max_a f(a) - \max_a (g(a))\|_\infty$
$\leq \max_a \|f(a) - g(a)\|_\infty$

$$= \max_a \| \sum_{s'} p(s'|s a] [r v(s') - r v'(s')] \|_\infty$$

$$= \max_a \| \sum_{s'} p(s'|s,a) \, r(v(s') - v'(s')) \|_\infty$$

$$= r \max_a \| \sum_{s'} p(s'|s,a)(v(s') - v'(s')) \|_\infty$$

$$\leq \gamma \| v(s') - v'(s') \|_\infty$$

2c) $\| V^{n+1} - V^* \|_\infty = \| V^* - V^{n+1} \|_\infty$

$$= \| \sum_{t=1}^{\infty} V^{n+1+t} - V^{n+t} \|_\infty$$

$$\leq \sum_{t=1}^{\infty} \| V^{n+1+t} - V^{n+1} \|_\infty$$

$$\leq \sum_{t=1}^{\infty} \gamma^t \| V^{n+1} - V^n \|_\infty$$

$$= \| V^{n+1} - V^n \|_\infty \cdot \frac{r}{1-r}$$

$\therefore \| V^{n+1} - V^* \|_\infty \leq \frac{\gamma}{1-\gamma} \| V^{n+1} - V^n \|_\infty$

if $\| V^{n+1} - V^n \|_\infty < \epsilon$

then

$\| V^{n+1} - V^* \|_\infty \leq \frac{\gamma}{1-\gamma} \epsilon$

2d) Suppose $V^* = T^{(\infty)}(0)$

$$T(V^*) = T(T^{\infty}(0)) = T^{\infty}(0) = V^*$$

$$\therefore \quad T(V^*) = V^*$$

Suppose $V^* = T^{\infty}(0)$

$$\| T^n(v) - T^{(n)}(v') \|_\infty \le \gamma^n \|v - v'\|_\infty \quad \text{by } \underline{2b}$$

$$\| T^{\infty}(v) - T^{\infty}(0) \|_\infty \le \gamma \|v - 0\|_\infty = 0$$

$$\| T^{\infty}(v) - V^* \|_\infty = 0$$

$$T^{\infty}(v) = V^*$$

4a)

$$\nabla_\theta J(\theta) = \nabla_\theta \mathbb{E}_{\tau \sim \pi_\theta}[R(\tau)] \quad \text{—— (A)}$$

if $R(\tau)$ is changed to $R(\tau) - b$,

then $\nabla_\theta J(\theta) = \nabla_\theta \mathbb{E}_{\tau \sim \pi_\theta}[R(\tau) - b]$

s.t. b is not a function of θ

$$= \mathbb{E}_{\tau \sim \pi_\theta}[\nabla_\theta (R(\tau) - b)]$$

$$= \mathbb{E}_{\tau \sim \pi_\theta}[\nabla_\theta R(\tau)]$$

$$= \nabla_\theta \mathbb{E}_{\tau \sim \pi_\theta}[R(\tau)] \quad \text{—— (B)}$$

$\therefore$ (A) = (B)