CSG700: Homework 3

Avinash. G. Kori
EDI513006.

(1) $f : \mathbb{R}^n \to \mathbb{R}^n$. $f \in C_L^{0,0}(\Omega)$

$\Rightarrow \|f(x) - f(y)\| \leq L\|x - y\|$. $\rightarrow$ (i)

A/Q $y = x^*$ $L = \alpha$.

$\Rightarrow \|f(x) - f(x^*)\| \leq \alpha \|x - x^*\|$.

$\Rightarrow \|f(x) - x^* + x - x\| \leq \alpha \|x - x^*\|$ $\langle \because f(x^*) = x^* \rangle$

$\Rightarrow \|f(x) - x\| + \|x - x^*\| \leq \alpha \|x - x^*\|$. $\langle \because \|a+b\| \leq \|a\| + \|b\| \rangle$

$\Rightarrow \|f(x) - x\| \leq (\alpha - 1)\|x - x^*\|$.

$\Rightarrow \|x - x^*\| \geq \frac{1}{(\alpha-1)} \|f(x) - x\|$.

$\Rightarrow$ $\boxed{\|x^* - x\| \leq \frac{1}{(1-\alpha)} \|f(x) - x\|.}$ $\rightarrow$ Proved

b) f is monotone $\Rightarrow f(x) > f(y)$ $\forall$ $x > y$.

given $x \leq y + \delta e$

$\Rightarrow f(x) \leq f(y + \delta e)$

As $f \in C_\alpha^{0,0}(\Omega)$

$\Rightarrow \|f(y+\delta e) - f(y)\| \leq \alpha \|\delta e\|$.

$\Rightarrow -\alpha \delta e \leq f(y+\delta e) - f(y) \leq \alpha \delta e$

$\Rightarrow f(y) - \alpha \delta e \leq f(y+\delta e) \leq f(y) + \delta \alpha e$

$\Rightarrow f(y+\delta e) \leq \alpha |\delta| e + f(y)$

$\Rightarrow$ $\boxed{f(x) \leq \alpha |\delta| e + f(y)}$ $\rightarrow$ Proved.

(2) $X = \{A, B\}$ $a = \{\text{cont}^n, \text{toggle}\}$

$R = \{r_A, r_B, -C\}$; $P_{ij}(\text{cont}^n) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$; $P_{ij}(\text{toggle}) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

Let π_1 be stationary policy $\pi_1 = \{\pi_{1A}, \pi_{1B}\}$

- for $\pi_1' = \{\text{cont}^n, \text{cont}^n\}$

  i.e. to continue to work in state they are in...

$$\Rightarrow J_{\pi_1'}(A) = r_A + \alpha J_{\pi_1'}(A)$$

$$\boxed{J_{\pi_1'} = T_{\pi_1'} J_{\pi_1'}}$$

$$\Rightarrow J_{\pi_1'}(A) = \frac{r_A}{1-\alpha}.$$

$$\text{Similarly } J_{\pi_1'}(B) = r_B + \alpha J_{\pi_1'}(B)$$

$$\Rightarrow J_{\pi_1'}(B) = \frac{r_B}{1-\alpha}.$$

$$J_{\pi_1'} = \left[\frac{r_A}{1-\alpha}, \frac{r_B}{1-\alpha}\right] \rightarrow \text{Policy evaluation for } \pi_1'$$

- for $\pi_1^2 = \{\text{cont}^n, \text{toggle}\}$

$$J_{\pi_1^2}(A) = r_A + \alpha J_{\pi_1^2}(A) \Rightarrow J_{\pi_1^2}(A) = \frac{r_A}{1-\alpha}.$$

$$J_{\pi_1^2}(B) = r_A - C + \alpha J_{\pi_1^2}(A) \Rightarrow J_{\pi_1^2}(B) = \frac{r_A}{1-\alpha} - C$$

$$J_{\pi_1^2} = \left[\frac{r_A}{1-\alpha}, \frac{r_A}{1-\alpha} - C\right] \rightarrow \text{Policy evaluation for } \pi_1^2$$

- for $\pi_1^3 = \{\text{toggle}, \text{cont}\}$.
  i.e jump to state B if in state A, and continue to stay in state B if in state B.

$$J_{\pi,3}(A) = r_B - C + \alpha J_{\pi,3}(B)$$

$$J_{\pi,3}(B) = r_B + \alpha J_{\pi,3}(B) \Rightarrow J_{\pi,3}(B) = \frac{r_B}{1-\alpha}$$

$$\Rightarrow J_{\pi,3}(A) = r_B - C + \frac{\alpha r_B}{1-\alpha} = \frac{r_B}{1-\alpha} - C.$$

$$J_{\pi,3} = \left[\frac{r_B}{1-\alpha} - C, \frac{r_B}{1-\alpha}\right] \rightarrow \text{Policy evaluation for } \pi,3$$

- for $\pi_1^4 = \{\text{toggle}, \text{toggle}\}$.
  i.e change state irrespective of current state.

$$J_{\pi,4}(A) = r_B - C + \alpha J_{\pi,4}(B)$$

$$J_{\pi,4}(B) = r_A - C + \alpha J_{\pi,4}(A)$$

$$\Rightarrow J_{\pi,4}(A) = \frac{r_A + \alpha r_B - (1+\alpha)C}{1-\alpha^2}$$

$$J_{\pi,4}(B) = \frac{r_B + \alpha r_A - (1+\alpha)C}{1-\alpha^2}$$

$$\Rightarrow J_{\pi,4} = \left[\frac{r_A + \alpha r_B - (1+\alpha)C}{1-\alpha^2}, \frac{r_B + \alpha r_A - (1+\alpha)C}{1-r^2}\right] \rightarrow \text{Policy evaluation for } \pi_1^4$$

if $\alpha \to 0$

$J_{T,1} = [r_A, r_B]$ will be optimal

Since in other policy cost at few states will be -ve.

$\Rightarrow$ $\alpha \to 0$ $T_1' =$ 1 cont, toggle is the best policy.

As $\alpha \to 1$.

$J_{T_1,1}\cdot(1-\alpha) = [r_A, r_B]$; $J_{T_1,2}\cdot(1-\alpha) = [r_A, r_A]$

$J_{T_1,3}\cdot(1-\alpha) = [r_B, r_B]$; $J_{T,4}\cdot(1-\alpha) = \left[\frac{r_A+r_B-C}{2}, \frac{r_A+r_B-C}{2}\right]$

As compared to all the J's

$J_{T_1^2}$ is optimal cost. Since $r_A > r_B$

$\Rightarrow T_1^2 =$ 1 cont, toggle move to state A is best policy for $\alpha \to 1$.

---

b) $C = 3$ $r_A = 2$ $r_B = 1$. $\alpha = 0.9$.

$J_{T_1,1} = [20, 10]$; $J_{T_1,2} = [20, 17]$

$J_{T_1,3} = [7, 10]$; $J_{T_1,4} = [-15.26, -14.74]$

from the above values it is clear that moving to state A is optimal.

(3)

$$m_j = \min_i \min_a P_{ij}(a)$$

$$\tilde{P}_{ij} = \frac{P_{ij} - m_j}{1 - \sum_k m_k}$$

a) for $\tilde{P}_{ij}$ to be transition probabilities

- $0 < \tilde{P}_{ij} < 1$. & $\sum_j \tilde{P}_{ij} = 1 \quad \forall i$

$$\sum_j \tilde{P}_{ij} = \sum_j \frac{P_{ij} - m_j}{1 - \sum_k m_k} = \frac{1}{(1 - \sum_k m_k)} \sum_j (P_{ij} - m_j)$$

$$\Rightarrow \frac{\sum_j P_{ij} - \sum_j m_j}{1 - \sum_k m_k} = \frac{1 - \sum_k m_k}{1 - \sum_k m_k} = 1$$

. P_{ij} is transition probability

$\longrightarrow$

$m_j \leq P_{ij}$ <∵ m_j involves min. over P_{ij}>

$\Rightarrow \frac{P_{ij} - m_j}{1 - \sum_k m_k} \geq 0$ <∵ $1 - \sum_k m_k > 0$ given>

$\longrightarrow$ (i).

We know that.

$$P_{ij} - m_j \leq \sum_j (P_{ij} - m_j)$$

$$\Rightarrow \frac{P_{ij} - m_j}{1 - \sum_k m_k} \leq 1.$$

$\longrightarrow$ (ii)

By (i) & (ii)

$0 \leq \tilde{P}_{ij} \leq 1$

Proved

b) $J^*(x) = \lim_{N\to\infty} T^N J(x) \quad \forall\, x \in S$

W.K.T be set of all states

$|g(i,a)| \le M. \Leftrightarrow g(i,a)$ is bounded.

~~$J^*(x_0) = \min_a \lim_{N\to\infty} \sum_{k=0}^{N+1} (\alpha_0 \sum P_{ij}(a))^k g(k,a)$~~

$$J^*(x_i) = \min_a \left[g(i,u) + \alpha \sum_j P_{ij}(a) J^*(x_j) \right] \quad \forall\, x_i, x_j \in S \longrightarrow (1)$$

Let $\tilde{J}(x_i) = \min_a \left[g(i,u) + \tilde{\alpha} \sum_j \tilde{P}_{ij}(a) \tilde{J}(x_j) \right]$

$$= \min_a \left[g(i,u) + \alpha \left(1 - \sum_k m_{ik}\right) \sum_j \frac{(P_{ij} - m_j)\tilde{J}(x_j)}{1 - \sum_k m_{ik}} \right]$$

$$= \min_a \left[g(i,u) + \alpha \sum_j (P_{ij} - m_j)\tilde{J}(x_j) \right]$$

$$= \min_a \left[g(i,u) + \alpha \sum_j P_{ij}(a)\tilde{J}(x_j) - \alpha \sum_j m_j \tilde{J}(x_j) \right]$$

adding $\alpha \dfrac{\sum_k m_{ik} \tilde{J}(x_k)}{1-\alpha}$ on both sides

$$\Rightarrow \tilde{J}(x_i) + \alpha \frac{\sum_k m_{ik}\tilde{J}(x_k)}{1-\alpha} = \min_a \left[g(i,u) + \alpha \sum_j P_{ij}(a)\tilde{J}(x_j) - \left[\alpha - \frac{\alpha}{1-\alpha}\right] \sum_k \tilde{J}(x_k) m_{ik} \right]$$

$$= \min \left[g(i,u) + \alpha \sum_j P_{ij}(a)\tilde{J}(x_j) + \frac{\alpha^2}{} \sum_i P_{ij}(a) \sum_k \tilde{J}(x_k) m_k \right]$$

$\left[\because \sum_j P_{ij} = 1 \right]$ ←

$$\Rightarrow \tilde{J}(x_i) + \frac{\alpha \sum_k m_k \tilde{J}(x_k)}{1-\alpha} = \min_a \left[g(i,u) + \alpha \sum_j P_{ij}(u) \left[\tilde{J}(x_i) + \frac{\alpha \sum_k m_k \tilde{J}(x_k)}{1-\alpha} \right] \right]$$

By comparing ② with ① ⟶ ②

$$J^*(x_i) = \tilde{J}(x_i) + \frac{\alpha \sum_k m_k \tilde{J}(x_k)}{1-\alpha}$$

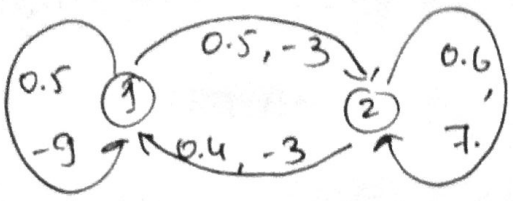
$$\Rightarrow \boxed{J^* = \tilde{J} + \frac{\alpha \sum_k m_k \tilde{J}(k)}{1-\alpha}} \longrightarrow \text{proved.}$$

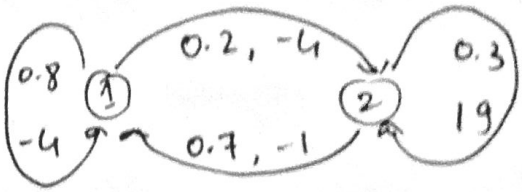
As $J^* = \tilde{J} + C$ where $C = \frac{\alpha \sum_k m_k \tilde{J}(x_k)}{1-\alpha}$.

Policy minimizing J^* will directly minimize $\tilde{J}$

$\Rightarrow$ J^* & $\tilde{J}$ has some optimal policy.

(4) $X = \{1, 2\}$; $A = \{a, b\}$; $\alpha = 0.9$. (5)





action a.

a & b} $\pi_{1,0} = \{a, a\}$

Policy evaluation @ $\pi_{1,0}$

$$J_{\pi_{1,0}}(1) = 0.5(-9 + \alpha J_{\pi_{1,0}}(1)) + 0.5(-3 + \alpha J_{\pi_{1,0}}(2))$$

$$= -6 + 0.5\alpha [J_{\pi_{1,0}}(1) + J_{\pi_{1,0}}(2)]$$

$$J_{\pi_{1,0}}(2) = 3 + 0.4\alpha J_{\pi_{1,0}}(1) + 0.6\alpha J_{\pi_{1,0}}(2)$$

$$J_{\pi_{1,0}} = [-15.5, -5.6]$$

Policy update ($T_{\pi_{1,1}} J_{\pi_{1,0}} = T J_{\pi_{1,0}}$)

$$TJ_{\pi_{1,0}}(1) = \min [-6 + 0.5\alpha[-21.1], -4 + \alpha(0.8(-15.5) + 0.2(-5.6)]$$

$$= \min [-15.56, -16.2]$$

$$= -16.2 \Rightarrow \pi_{1,1}(1) = b.$$

$$TJ_{\pi_{1,0}}(2) = \min [3 + \alpha(0.4(-15.5) + 0.6(-5.6)), 5 + \alpha(0.3(-5.6) + 0.7(-15.5))]$$

$$= \min [-5.56, -6.313]$$

$$= -6.313 \Rightarrow \pi_{1,1}(1) = b$$

$\pi_{1,1} = \{b, b\}$

Policy evaluation @ {b, b}

$$J_{\pi_1}(1) = 0.8(-4) + 0.2(-4) + \alpha[0.8 J_{\pi_1}(1) + 0.2 J_{\pi_1}(2)]$$

$$J_{\pi_1}(2) = 0.7(-1) + 0.3(19) + \alpha(0.7 J_{\pi_1}(1) + 0.3 J_{\pi_1}(2))$$

$$\Rightarrow J_{\pi_1} = [-22.2, -12.3]$$

Second update $(T_{\pi_2} J_{\pi_1} = T J_{\pi_1})$

$$T J_{\pi_1}(1) = \min\{-6 + 0.5\alpha[-34.5], -4 + \alpha[0.8(-22.2) + 0.2(-12.3)]$$

$$= \min\{-21.52, -22.2\}$$

$$= -22.2 \quad ; \pi_2(1) = b$$

$$T J_{\pi_1}(2) = \min\{3 + \alpha(0.4(-22.2) + 0.6(-12.3)),$$
$$5 + \alpha(0.3(-12.3) + 0.7(-22.2))\}$$

$$= \min\{-11.6, -12.7\}$$

$$= -12.7 \quad ; \pi_2(2) = b$$

$$\Rightarrow \pi_2 = \{b, b\}$$

$\therefore \pi_1 = \pi_2 = \{b, b\}$ π_2 is an optimal policy

c] $J = [0, 0]$

$$TJ = \min_a \sum P_{ij}(a)\left[g(i,a,j) + \alpha J(j)\right]$$

$TJ(1) = \min[-6, -4] = -6.$

$TJ(2) = \min[3, 5] = 3.$

$TJ = [-6, 3] \Rightarrow \pi_1(a, a)$

$$T^2J = \min_a \sum P_{ij}(a)\left[g(i,a,j) + \alpha TJ(j)\right]$$

$T^2J(1) = \min[-6 + 0.4(-3), -4 + 0.8(0.8(-6) + 0.2(3))]$

$= \min[-7.35, -7.78] = -7.78$ (b)

$T^2J(2) = \min[3 + \alpha(0.4(-6) + 0.6(3)), 5 + \alpha(0.3(3) + 0.7(-4))]$

$= \min[2.46, 2.03] \Rightarrow 2.03$ (b).

$T^2J = [-7.78, 2.03]$ $\pi_1 = (b, b)$

$T^3J(1) = \min[-6 + 0.4(-5.75), -4 + 0.8[0.8(-7.78) + 0.2(2.03)]]$

$= \min[-8.58, -9.23] = -9.23$

$T^3J(2) = \min[3 + \alpha(0.4(-7.78) + 0.6(2.03)),$
$5 + \alpha(0.3(-9.23) + 0.7(-8.58))]$

$= \min[1.3, 0.06] = 0.06$

$T^3J = [-9.23, 0.06]$ $\pi_1 = (b, b)$

$T^4J(1) = \min[-6 + 0.4(-9.17),\ -4 + \alpha(0.8(-9.23) + 0.2(0.06))]$

$= \min[-10.127,\ -10.638] = \underline{-10.638}$

$T^4J(2) = \min[3 + \alpha[0.4(-9.23) + 0.6(0.06)],$

$5 + \alpha[0.7(-9.23) + 0.3(0.06)]$

$= \min[-0.29,\ -0.80] = \underline{-0.8}$

$T^4J = [-10.63, -0.80] \quad \pi_1 = \{b, b\}$

a] finding optimal policy with $\alpha = 0.1$

let $\pi_0 = \{a, a\}$

$J_{\pi_0}(1) = -6 + 0.05(J_{\pi_0}(1) + J_{\pi_0}(2))$

$J_{\pi_0}(2) = 3 + 0.04 J_{\pi_0}(1) + 0.06 J_{\pi_0}(2)$

$J_{\pi_0} = [-6.16,\ 2.92]$

Policy update

$TJ_{\pi_0}(1) = \min\{-6 + 0.05(-3.24),\ -4 + 0.1(0.8(-6.16) + 0.2(2.92))\}$

$= \min\{-6.16,\ 2.92\} = \underline{-6.16}$

$TJ_{\pi_0}(2) = \min\{3 + 0.1(0.4(-6.16) + 0.6(2.92)),$

$5 + 0.1(0.3(2.92) + 0.7(-6.16))\}$

$= \min\{2.9,\ 4.6\} = 2.9$

$\Rightarrow \pi_1 = \{a, a\} = \pi_0 \Rightarrow \{a, a\}$ is optimal.

As $\alpha \to 0 \Rightarrow$ policy biased towards immediate reward / cost.

- @ state ① immediate minimum cost is to select action @ & stay in state ①.
- @ state ② immediate minimum cost is achieved by selecting action @ & jump to state ①