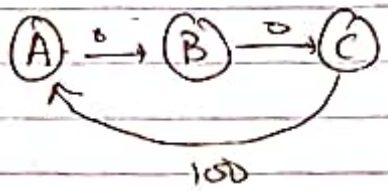
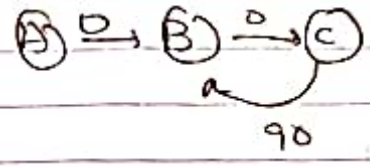
Homework - 6

Name :- Aishwarya Vinod Menon
Net ID :- AXV220062

Questions :-

1) 

2) 

2) For policy 1

$$V^*(C) = 100 + 100(\gamma^3) + 100\gamma^6 + \dots$$
$$= 100(1 + \gamma^3 + \gamma^6 + \dots)$$
$$= 100 \frac{1}{1-\gamma^3}$$
$$\approx 100.$$

$$V^*(B) = V^*(C) \cdot \gamma$$
$$= 100 \times 10^{-7}$$

$$V^*(A) = 100 \times 10^{-14} = V^*(B) \cdot \gamma$$

$$Q^*(A,B) = 0 + 10^{-14} \times 100 = 10^{-14} \times 100$$
$$Q^*(B,C) = 0 + 100 \times 10^{-7} = 100 \times 10^{-7}$$
$$Q^*(C,B) = 90 + 10^{-7} \times 10^{-7} \times 100$$
$$= 90 + 100 \times 10^{-14}$$
$$Q^*(C,A) = 100 + 10^{-7} \times 100 \times 10^{-14} = 100 + 100 \times 10^{-14}$$
$$Q^*(B,A) = 100 + 100 \times 10^{-7} \times 10^{-14} = 100 + 100 \times 10^{-21}$$

For policy 2

$$V^*(C) = 90 + \gamma^2 \cdot 90 + \gamma^4 \cdot 90 + \dots$$
$$= 90(1 + \gamma^2 + \gamma^4)$$
$$\approx 90$$

$$V^*(B) = \gamma \cdot V^*(C) = 90 \times 10^{-7}$$
$$V^*(A) = \gamma \cdot V^*(B) = 90 \times 10^{-14}$$
$$Q^*(B,A) = 0 + 10^{-7} \times 90 \times 10^{-14} = 90 \times 10^{-21}$$

$Q^*(A,B) = 90 \times 10^{-14}$

$Q^*(B,C) = 90 + 10^{-14} \times 90 = 90 + 90 \times 10^{-14}$

$Q^*(C,A) = 100 + 90 \times 10^{-21}$

~~Edge (C → A) has~~

$Q(C,A) > Q(C,B)$

∴ Initial guess wrong

Optimal policy is 1

3 $\gamma = 0.9999999$

Policy 1

$V^*(C) = 100 + 100\gamma^3 + 100\gamma^6$

$= 100 \frac{1}{1-\gamma^3}$

$= \frac{100}{3} \cdot 10^7$

$V^*(B) = \frac{100}{3}(10^7 - 1)$

$V^*(A) = \frac{100}{3}(10^7 - 2)$

$Q^*(A,B) = \frac{100}{3}(10^7 - 2)$

$Q^*(B,A) = \frac{100}{3}(10^7 - 3)$

$Q^*(B,C) = \frac{100}{3}(10^7 - 1)$

$Q^*(C,B) = 90 + \frac{100}{3}(10^7 - 2)$

$Q^*(C,A) = 100 + \frac{100}{3}(10^7 - 3)$

5 value from C to B > 8 value from C to A. Policy ~~1~~ 2 is optimal

Policy 2

$$V^*(C) = 90(1+\gamma^2+\gamma^4+\cdots)$$
$$= 90 \cdot \frac{1}{1-\gamma^2} = \frac{90}{2} \times 10^7$$
$$= 45 \cdot 10^7$$

$$V^*(B) = (1-10^{-7}) \cdot V^*(C)$$
$$= 45(10^7-1)$$

$$V^*(A) = (1-10^{-7}) * V^*(B)$$
$$= 45(10^7-2)$$

$$Q^*(A,B) = 0 + (1-10^{-7}) \cdot 45(10^7-1)$$
$$= 45(10^7-2)$$

$$Q^*(B,A) = (1-10^{-7})(10^7-1) \cdot 45$$

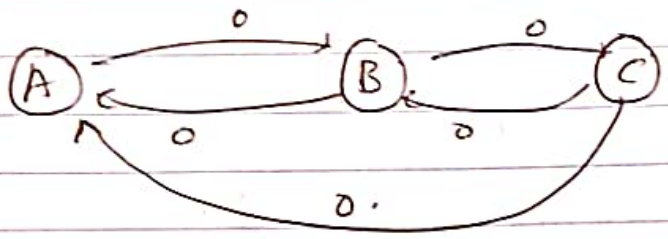
$$Q^*(B,C) = 45 \times 10^7 (1-10^{-7})$$
$$= 45(10^7-1)$$

$$Q^*(C,B) = 90 + (1-10^{-7}) \cdot 45(10^7-1)$$
$$= 45(2+10^7-3)$$
$$= 45(10^7-1)$$

$$Q^*(C,A) = 100 + (1-10^{-7}) \cdot 45(10^7-2)$$
$$= 100 + 45(10^7-3)$$

Policy 2 is optimal policy

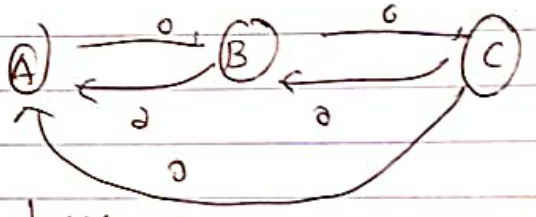
4. given



$\gamma = 1/2$

1. (B, left)

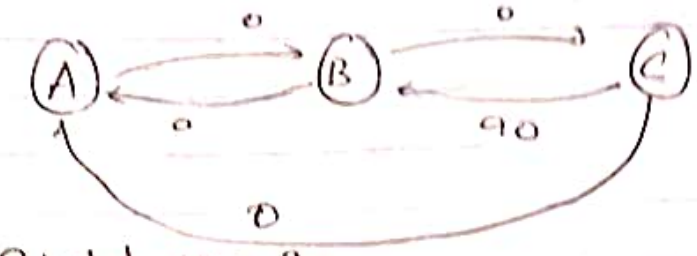
$\hat{Q}(s, a) =$



$0 + \frac{1}{2} \times 0 = 0$

2.

$(C, \text{left}) \rightarrow$

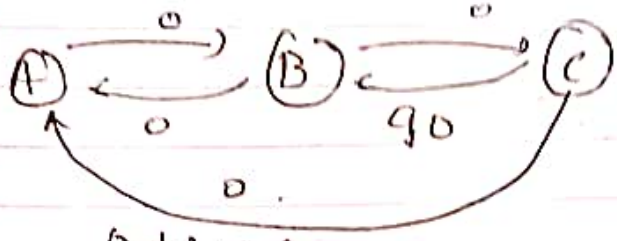


$90 + \frac{1}{2} \times 0 = 90$

3

(A, right)

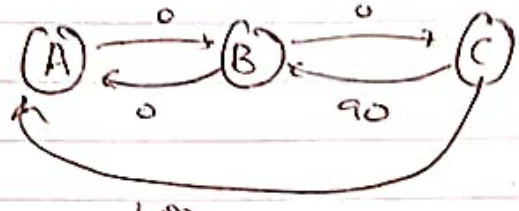
$\hat{\theta}(s, a) =$



$0 + \frac{1}{2} \times 0 = 0$

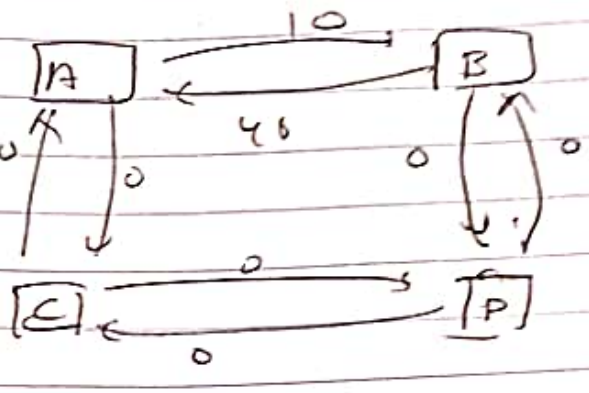
4. (C, down)

$\hat{\theta}(s, a) =$



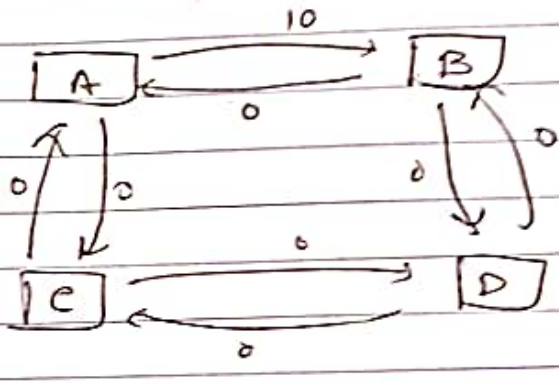
$100 + \frac{1}{2} \times 0 = 150.$

Question 2



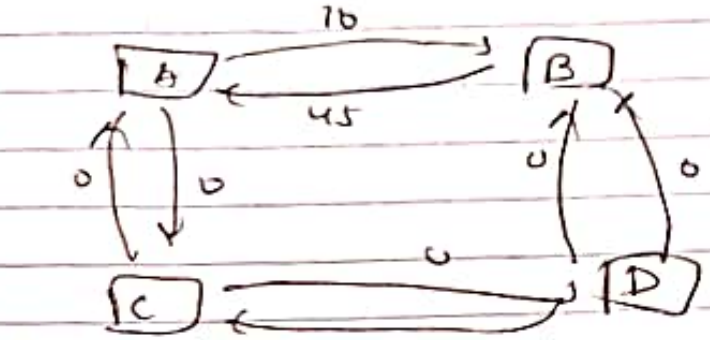
Part-A $\gamma = 0.5$

$V^{\pi}(A, \text{right}) =$



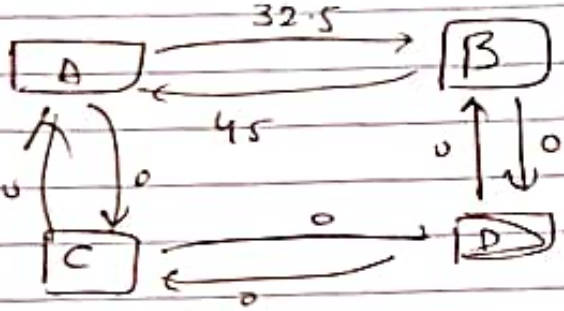
$$10 + \frac{1}{2} \times 0 = 10$$

2. $\hat{Q}(B, \text{left})$



$40 + \frac{1}{2} \times 10 = 45$ 0

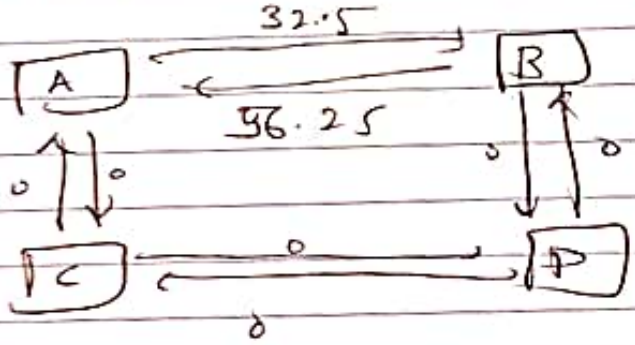
3. $\hat{Q}(A, \text{right})$



$10 + \frac{1}{2} \times 45$

$= 32.5$

4 $\hat{Q}(B, \text{left})$



$40 + \frac{1}{2} \times 32.5 = 56.5$

Part-B

It converges

1. $V^*_A = 10 + \frac{1}{2} \cdot 40 + \frac{1}{4} \cdot 10 + \frac{1}{8} \cdot 40 + \dots$

$10\left[1 + \frac{1}{4} + \dots\right] + \frac{40}{2}\left[1 + \frac{1}{4} + \dots\right]$

$= \frac{40}{1} \times \frac{80}{3} = 40$

$V^* B : \quad 40 + \frac{1}{2} \cdot 40 = 10$

The optimal policy Π^*

A ⇌ B $\quad \Theta^* :$ A ⇌ B (40 / 60)

Converges to