ASSIGNMENT - 2

from the description in question we get that Vkn is the last iteration of the algorithm, and also

1 VKH - VK 100 = E

let B be the Bellman evaluation backup

11 Vr - VII = 11 Vr - VKH 1100 + 11 VKH - VIII00

(teangular mequality over norms)

= 1/1/x - Vxy 1/00 + 1/BVx - BV/1/00.

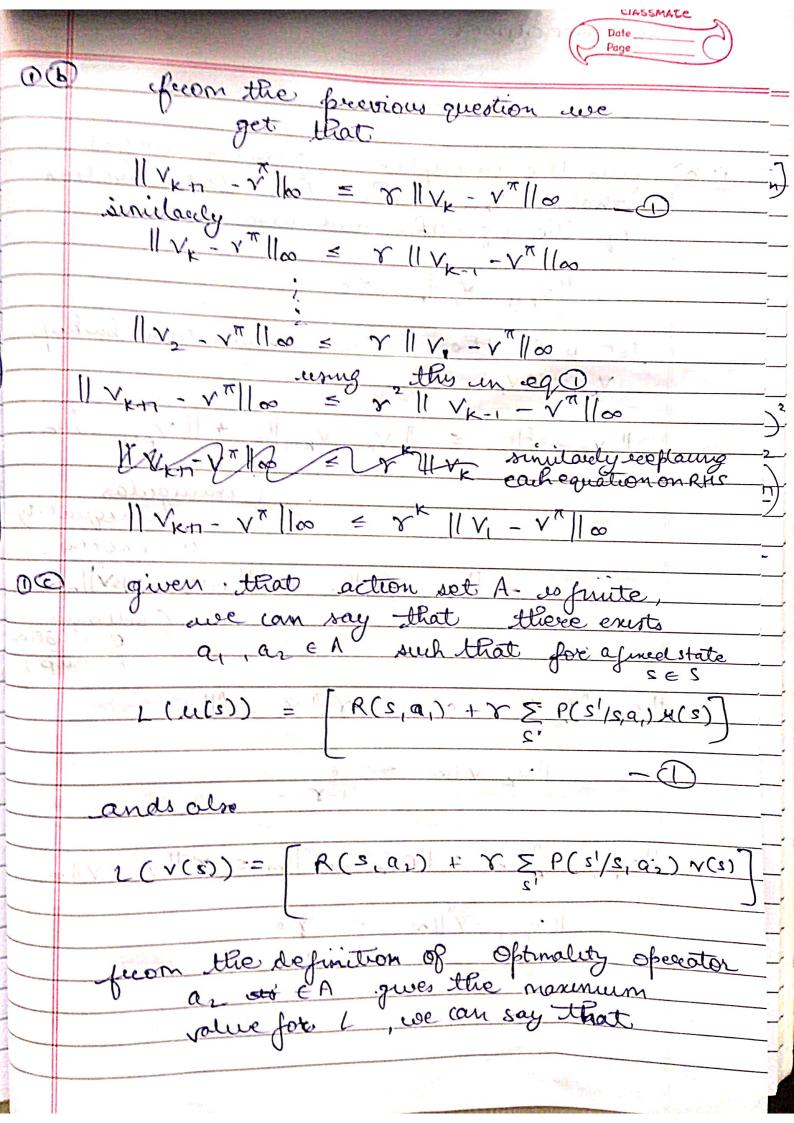
(Bellman Biraliation backub) backup)

|V_K - √"||_∞ = ε + 2 8 || V_K - √"||_∞

llv_k-vll∞ ≤ €

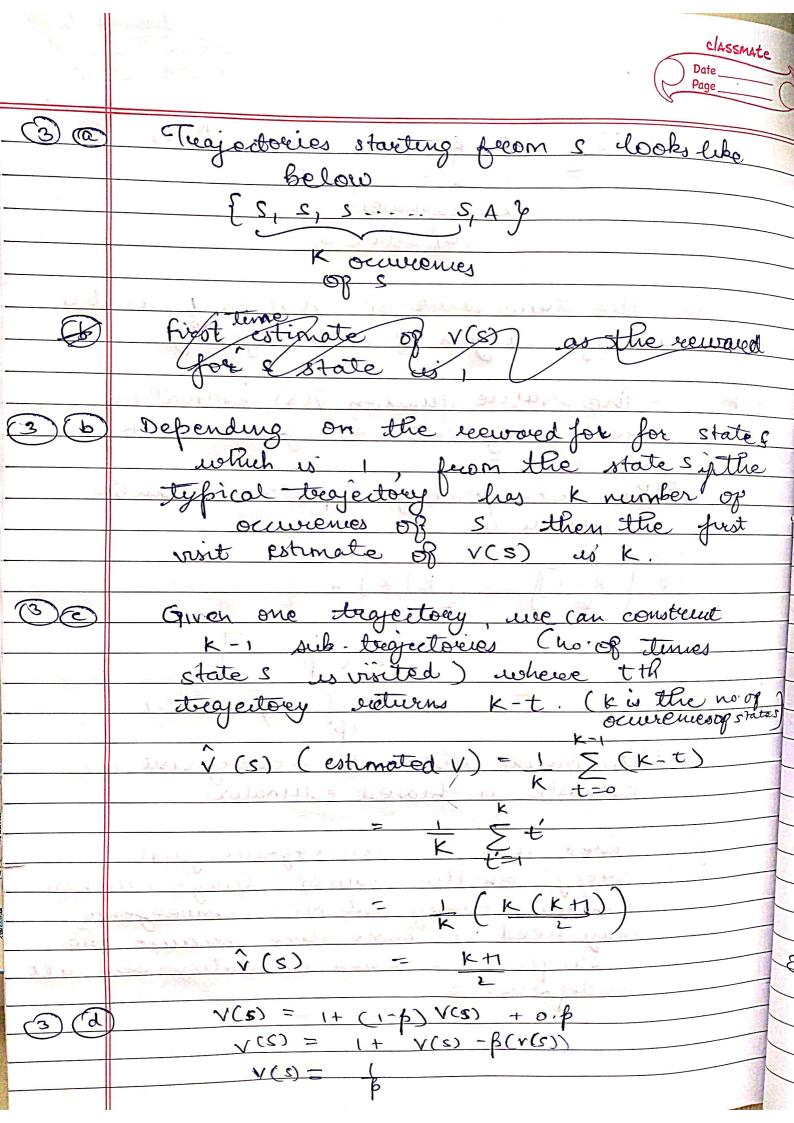
|| VKT -V | | | | = || BVK - BV | | ∞ ≤ X || VK - V || ∞

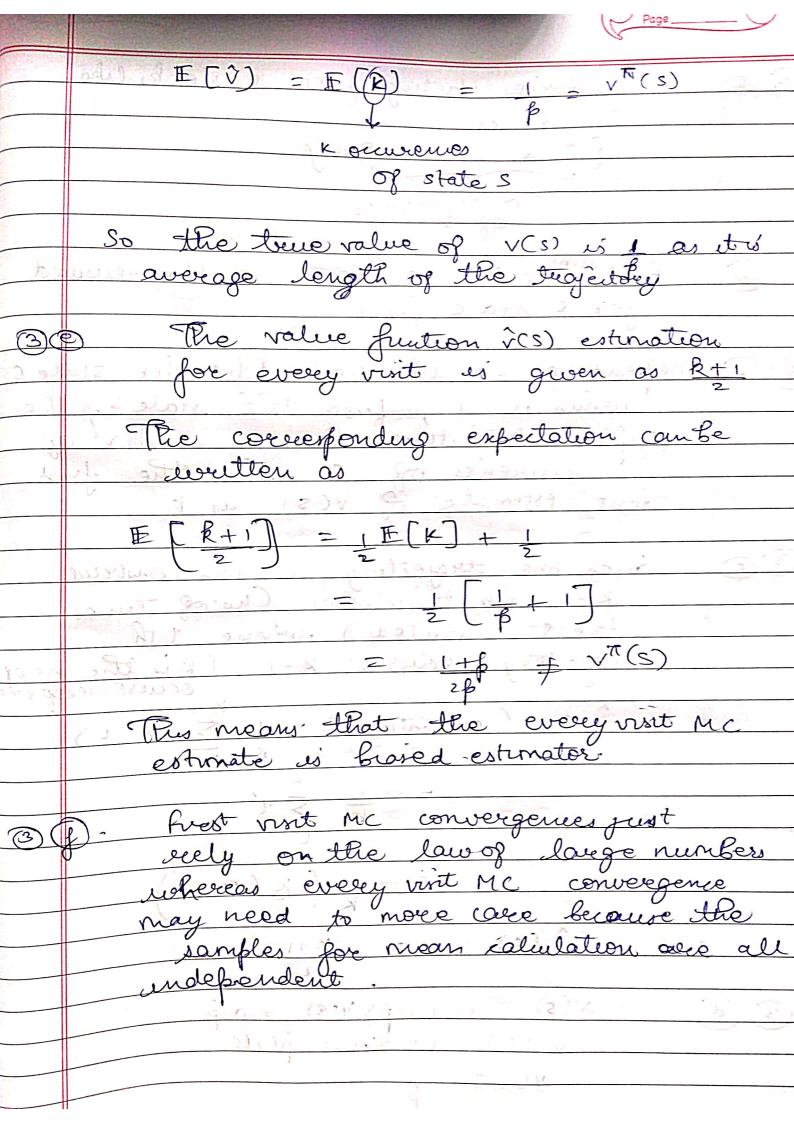
11 VKM - VT/100 = YE



R(s,a,) + 8 5 P(s'/s,a,) v(s) $L(V(s)) \ge$ subtract 2 joon (1 -γ 5 P(s'/s;a.) (x(s)-v(s) L(u(s)) - L(v(s)) = since u(s) < v(s). $L(u(s)) - L(v(s)) = \begin{cases} \gamma > \rho(s'/s, a_i) \\ s' \end{cases}$ L(u(s)) - L(V(s)) <0 L(u(s)) = L(v(s)) L(u) < L(v) hence browned

for all u, v e V, we can write 00 1 POQ(v) - POQ(u) [= || P(Q(u)) - P(Q(u)) || = 10 (v) -0 (u) 1 11 POO(v) - POO(us) = rpra 11 v - u11 To and ra Both belong to [0,1) hence there product also belongs to (0,1) which makes Pool composition to be contraction on the same normed rector space. 11 (208 (v) -008 (u) 11 = 110(P(v)) -0 (P(u)) 11 < rg / P(v) - P(M) / 1100P(N) - 00P(M)11= rprop 11v-M11 similarly, Oof comfortion is also contraction on the same normed veilor space. ferom the Da ese can see that The contration coefficient for loo and 26 Our ous spra. for a unique solution to exist, the 20 operator B as FOI must be contraction which means that both F and I must be contraction under the max-norm







$$S_{t} = 2 + v v (S_{t+1}) - v (S_{t+1})$$

$$\mathbb{E}\left[S_{t} \mid S_{t} = S\right] = \mathbb{E}\left[x_{tn} + \gamma \sqrt{(s_{tn})} - \sqrt{(s_{t})} \mid S_{t} = S\right]$$

$$= \sqrt{\pi(s)} - \sqrt{\pi(s)}$$

$$= \sqrt{\pi(s)} - \sqrt{\pi(s)}$$

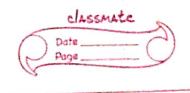
$$\mathbb{E}\left[\mathbf{S}_{t}\mid\mathbf{S}_{t}=\mathbf{S},\;\mathbf{A}_{t}=\mathbf{a}\right]$$

$$= \mathbb{E}\left[\varepsilon_{1} + \gamma \sqrt{(s_{t+1})} - \sqrt{(s_{t})}\right]$$

$$= \left[s_{t} = s, A_{t} = a\right]$$

$$E[S_{t} | S_{t} = S, A_{t} = a] = Q^{T}(S_{t}a) - V^{T}(S)$$

half of its unital value.
The half life owns when weighting deeps by holf. Therefore



$$\eta(\lambda) = \frac{\ln(\frac{1}{2})}{\ln(\lambda)}$$

for 3 step recturen the time taken 42

$$=) \quad \lambda^2 = \frac{1}{2}$$

The havemone serves governdization is
the f-revies defined of the f-

The f-series converges for all \$71 and diverges for all \$ < 1.

$$xt = \frac{1}{2} \times \frac{1}{2} \times$$

Algo doesnot converge with given of

(iii)
$$\alpha_t = 1$$
 \Rightarrow $\Sigma \alpha_t = \infty$
 $\Sigma \alpha_t = \infty$

cis

Algo converges with given value of of.

$$(iii) \quad \alpha_{+} = 1 \quad \Rightarrow \quad \sum_{k=1}^{\infty} \alpha_{+}^{2} = \infty$$

$$\sum_{k=1}^{\infty} \alpha_{+}^{2} = \infty$$

$$\sum_{k=1}^{\infty} \alpha_{+}^{2} = \infty$$

Algo converges with given value of of

$$\alpha_{t} = \frac{1}{t^{\gamma_{2}}} \Rightarrow \frac{\delta}{t} = \infty$$

Algo, converges with given value of of.

generalizing above result we can write

The Robbins - Moneroe condition y True when pe (\frac{1}{2},1)