BY: S. VISHAL
CH18B020

CS6700 - RL - ASSIGNMENT - 1

(5) Before I answer the question, I would like to show $\|a+b\|_\xi < \|a\|_\xi + \|b\|_\xi$

$$\|a+b\|_\xi = \max_i \frac{|a(i)+b(i)|}{\xi(i)}$$

[triangle inequality] $< \max_i \left(\frac{|a(i)| + |b(i)|}{\xi(i)} \right)$

$\because \max((a+b)(i)) < \max(a(i)) + \max(b(i))$ $\quad < \max_i \frac{|a(i)|}{\xi(i)} + \max_j \frac{|b(j)|}{\xi(j)}$

$$= \|a\|_\xi + \|b\|_\xi$$

$$\Rightarrow \|a+b\|_\xi < \|a\|_\xi + \|b\|_\xi \quad \text{—— (1)}$$

a) Termination criterion: $\|J_{m+1} - J_m\|_\xi < \varepsilon\left(\frac{1-\gamma}{2\gamma}\right)$ —— (2)

T is a contraction operator

$$\Rightarrow \|TJ - TJ'\|_\xi \leq \gamma \|J - J'\|_\xi \quad \text{—— (3)}$$

Apply operator T in eqn (2)

$\Rightarrow \|TJ_{m+1} - TJ_m\| < \gamma \varepsilon\left(\frac{1-\gamma}{2\gamma}\right)$ ($\because$ T is a contraction! - eqn (3))

$$\Rightarrow \| J_{m+2} - J_{m+1} \|_{\xi} < \rho \left[\varepsilon \left(\frac{1-\rho}{2\rho} \right) \right]$$

Recursively applying T,

$$\| J_{m+3} - J_{m+2} \|_{\xi} < \rho^2 \left[\varepsilon \left(\frac{1-\rho}{2\rho} \right) \right]$$

$$\vdots$$

$$\| J_{m+n} - J_{m+n-1} \|_{\xi} < \rho^{n-1} \left[\varepsilon \left(\frac{1-\rho}{2\rho} \right) \right]$$

Add all these eqns.

$$\Rightarrow \sum_{k=2}^{n} \| J_{m+k} - J_{m+k-1} \|_{\xi} < \sum_{k=1}^{n-1} \rho^k \left[\varepsilon \left(\frac{1-\rho}{2\rho} \right) \right] \quad (4)$$

Now we use $\|a+b\|_{\xi} < \|a\|_{\xi} + \|b\|_{\xi}$ and push the sum inside the modulus

i.e. $\| J_{m+n} - J_{m+n-1} + J_{m+n-1} - J_{m+n-2} + \dots + J_{m+3} - J_{m+2} + J_{m+2} - J_{m+1} \|_{\xi}$

$$< \sum_{k=2}^{n} \| J_{m+k} - J_{m+k-1} \|_{\xi}$$

$$④ \Rightarrow \| J_{m+n} - J_{m+1} \|_{\xi} < \sum_{k=1}^{n-1} \rho^k \left(\frac{\varepsilon(1-\rho)}{2\rho} \right)$$

$\not{\ast}$- take $\lim_{n \to \infty}$

$$\lim_{n \to \infty} \| J_{m+n} - J_{m+1} \|_{\xi} < \lim_{n \to \infty} \sum_{k=1}^{n-1} \rho^k \left(\varepsilon \left(\frac{1-\rho}{2\rho} \right) \right)$$

$$= \frac{\rho}{1-\rho} \varepsilon \left(\frac{1-\rho}{2\rho} \right)$$

$$\Rightarrow \lim_{n \to \infty} \| J_{m+n} - J_{m+1} \|_{\xi} < \frac{\varepsilon}{2}$$

Also, we know that value iteration converges

$$\left(\lim_{k \to \infty} T^k J_0 = J^* \right)$$

$$\Rightarrow \| J_m^* - J_{m+1} \|_{\xi} < \varepsilon/2$$

$$\Rightarrow \| J_{m+1} - J_m^* \|_{\xi} < \frac{\varepsilon}{2} \quad ⑤$$

Hence proved.

b) From part (a) we have

$$\| J_{m+1} - J_m^* \|_{\xi} < \frac{\varepsilon}{2}$$

Applying T operator & using T is a contraction

$$< \| T J_{m+1} - J^* \| < \frac{\rho \varepsilon}{2} \quad (\because T J^* = J^*)$$

But $T_{\pi \varepsilon} J_{m+1} = T J_{m+1} \quad (\because T J^* = J^*)$

$$\Rightarrow \| T_{\pi \varepsilon} J_{m+1} - J^* \| < \frac{\rho \varepsilon}{2} \quad \text{—— (6)}$$

In showing each policy iteration indeed gives us a policy improvement, (which is what we have done here)

we showed (in class)

$$J_{\pi \varepsilon} \leq J_{m+1} \Rightarrow T_{\pi \varepsilon} J_{\pi \varepsilon} \leq T_{\pi \varepsilon} J_{m+1}$$

($\because T_{\pi \varepsilon}$ is a monotone operator.)

$$\Rightarrow T_{\pi \varepsilon} J_{\pi \varepsilon} - J^* \leq T_{\pi \varepsilon} J_{m+1} - J^*$$

But $T_{\pi k} J_{\pi k} = J_{\pi k} \Rightarrow J_{\pi \varepsilon} - J^* \leq T_{\pi \varepsilon} J_{m+1} - J^*$

$$\Rightarrow \| J_{\pi \varepsilon} - J^* \| \leq \| T_{\pi \varepsilon} J_{m+1} - J^* \| \quad \text{—— (7)}$$

plug (7) in (6)

$$\Rightarrow \| J_{\pi \varepsilon} - J^* \|_{\xi} \leq \| T_{\pi \varepsilon} J_{m+1} - J^* \|_{\xi} < \frac{\rho \varepsilon}{2}$$

$$\Rightarrow \| J_{\pi \varepsilon} - J^* \|_{\xi} < \frac{\rho \varepsilon}{2} ; \text{ But } 0 < \rho < 1 \Rightarrow \frac{\rho \varepsilon}{2} < \varepsilon$$

$$\Rightarrow \| J_{\pi \varepsilon} - J^* \|_{\xi} < \varepsilon$$

Hence proved