97040 Rostislav Andreev
97155 João Aparício

Homework 5. Reinforcement Learning

a) $\dot{J}^{\pi} = \mathbb{E}_{\pi}[c_t + (\gamma-1) J^{\pi}]$

$$= c_{\pi} + (\gamma-1) J^{\pi}$$

$$= c_{\pi} + (\gamma-1) \frac{c_{\pi}}{(1-\gamma)}$$

$$= c_{\pi} - c_{\pi}$$

$$= 0 \quad \text{cqd}$$

b) $\dot{\varepsilon}_t = \dfrac{\partial \varepsilon_t}{\partial t}$

$$= \frac{\partial}{\partial t} \frac{1}{2}(J^t - J^{\pi})^2$$

$$= \frac{1}{2} \cdot 2 \cdot (J^t - J^{\pi}) \cdot \frac{\partial}{\partial t}(J^t - J^{\pi})$$

$$= (J^t - J^{\pi}) \cdot \dot{J}$$

$$= (J^t - J^{\pi}) \cdot \mathbb{E}_{\pi}[c_t + (\gamma-1) J]$$

$$= (J^t - J^{\pi}) \cdot (c_{\pi} + (\gamma-1) J^t)$$

Given that $\dot{J}^{\pi} = 0$, we have

$$\dot{\varepsilon}_t = (J^t - J^{\pi}) \cdot (c_{\pi} + (\gamma-1) J^t - \dot{J}^{\pi})$$

$$= (J^t - J^{\pi}) \cdot (c_{\pi} + (\gamma-1) J^t - (c_{\pi} + (\gamma-1) J^{\pi}))$$

$$= (J^t - J^{\pi}) \cdot (c_{\pi} + (\gamma-1) J^t - c_{\pi} - (\gamma-1) J^{\pi})$$

$$= (J^t - J^{\pi}) \cdot ((\gamma-1) J^t - (\gamma-1) J^{\pi})$$

$$= (J^t - J^{\pi}) \cdot (\gamma-1)(J^t - J^{\pi})$$

$$= (\gamma-1) \cdot (J^t - J^{\pi})^2$$

Given that $(\gamma-1) \leq 0$ and $(J^t - J^{\pi})^2 \geq 0$, the product between the two will always be ≤ 0, giving the result $\dot{\varepsilon}_t \leq 0$.

c) With the result obtained in (b) we can conclude that the energy is always decreasing and eventually reaches 0. E_t only reaches 0 when $J^t = J^\pi$, which suggests that J^t will eventually converge to J^π and will permanently stay there. This will work for any $\gamma \in [0,1]$ as long as every state is visited infinitely often. As our case only has one state, we can apply this theorem. and conclude that the convergence occurs.