0786009 鄭景鴻

Problem 1

(i) from eq.(1):

$$L_{\pi_{\theta_1}}(\pi_{\theta_1}) = \eta(\pi_{\theta_1}) + \sum_s d_\mu^{\pi_{\theta_1}}(s) \underbrace{\left(\sum_a \pi_{\theta_1}(a|s) A^{\pi_{\theta_1}}(s,a)\right)}_{=0}$$

$$\therefore L_{\pi_{\theta_1}}(\pi_{\theta_1}) = \eta(\pi_{\theta_1})$$

(ii) ① $\nabla_\theta L_{\pi_{\theta_1}}(\theta) = \cancel{\nabla_\theta \eta(\pi_{\theta_1})}^{0} + \sum_s d_\mu^{\pi_{\theta_1}}(s) \sum_a \left(\nabla_\theta \pi_\theta(a|s)\right) A^{\pi_{\theta_1}}(s,a)$

$$\Rightarrow \nabla_\theta L_{\pi_{\theta_1}}(\theta)\Big|_{\theta=\theta_1} = \sum_s d_\mu^{\pi_{\theta_1}}(s) \sum_a \left(\nabla_\theta \pi_\theta(a|s)\Big|_{\theta=\theta_1}\right) A^{\pi_{\theta_1}}(s,a)$$

② originally, $\eta(\pi_\theta) = \eta(\pi_{\theta_1}) + \sum_s d_\mu^{\pi_\theta}(s) \sum_a \pi_\theta(a|s) A^{\pi_{\theta_1}}(s,a)$

$$\therefore \nabla_\theta \eta(\pi_\theta) = \cancel{\nabla_\theta \eta(\pi_{\theta_1})}^{0} + \left[\sum_s d_\mu^{\pi_\theta}(s) \sum_a \left(\nabla_\theta \pi_\theta(a|s)\right) A^{\pi_{\theta_1}}(s,a)\right]$$

$$\Rightarrow \nabla_\theta \eta(\pi_\theta)\Big|_{\theta=\theta_1} = \sum_s d_\mu^{\pi_{\theta_1}}(s) \sum_a \left(\nabla_\theta \pi_\theta(a|s)\Big|_{\theta=\theta_1}\right) A^{\pi_{\theta_1}}(s,a)$$

③ from ① and ②: $\nabla_\theta L_{\pi_{\theta_1}}(\pi_\theta)\Big|_{\theta=\theta_1} = \nabla_\theta \eta(\pi_\theta)\Big|_{\theta=\theta_1}$ #

---

check in another way:

$$\frac{\partial V^\pi(s)}{\partial\theta} = \frac{\partial}{\partial\theta}\sum_a \pi(s,a) Q^\pi(s,a) = \sum_a \left[\frac{\partial \pi(s,a)}{\partial\theta} Q^\pi(s,a) + \pi(s,a)\frac{\partial}{\partial\theta} Q^\pi(s,a)\right]$$

$$= \sum_a \left[\frac{\partial\pi(s,a)}{\partial\theta} Q^\pi(s,a) + \pi(s,a) \underbrace{\frac{\partial}{\partial\theta}\left[\cancel{R_s^a} + \sum_{s'} \gamma P_{ss'}^a V^\pi(s')\right]}_{\sum_{s'} \gamma P_{ss'}^a \frac{\partial}{\partial\theta} V^\pi(s')}\right]$$

$$= \sum_x \underbrace{\sum_k^\infty \gamma^k P_r(s\to x, k, \pi)}_{d_\mu^{\pi_\theta}} \sum_a \frac{\partial\pi(s,a)}{\partial\theta} Q^\pi(s,a)$$

$\Rightarrow \dfrac{\partial\eta(\pi_\theta)}{\partial\theta} = \sum_s d_\mu^{\pi_\theta}(s) \sum_a \dfrac{\partial\pi_\theta(s,a)}{\partial\theta} Q^\pi(s,a)$ and we choose $V^\pi(s)$ as the baseline

$$\Rightarrow \nabla_\theta \eta(\pi_\theta)\Big|_{\theta=\theta_1} = \sum_s d_\mu^{\pi_{\theta_1}}(s) \sum_a \left(\nabla_\theta \pi_\theta(a|s)\Big|_{\theta=\theta_1}\right) A^{\pi_{\theta_1}}(s,a)$$

0786009 廖家瑜

Problem 2

(a) (i) since $D(\lambda) := \min_{\theta \in \mathbb{R}^d} \mathcal{L}(\theta, \lambda)$, we find θ^* from $\nabla_\theta \mathcal{L}(\theta, \lambda) = 0$:

$$\Rightarrow 0 = -\left(\nabla_\theta L_{\theta_k}(\theta)\big|_{\theta=\theta_k}\right) + \lambda H(\theta - \theta_k)$$

$$\Rightarrow \theta^* - \theta_k = \frac{1}{\lambda} H^{-1}\left(\nabla_\theta L_{\theta_k}(\theta)\big|_{\theta=\theta_k}\right) \text{ --- ①}$$

substute ① into eq. (4): $\mathcal{L}(\theta, \lambda)$:

$$\Rightarrow D(\lambda) = -\left(\nabla_\theta L_{\theta_k}(\theta)\big|_{\theta=\theta_k}\right)^T \frac{H^{-1}}{\lambda}\left(\nabla_\theta L_{\theta_k}(\theta)\big|_{\theta=\theta_k}\right)$$

$$+ \underbrace{\frac{\cancel{\lambda}}{2} \cdot \left[\frac{H^{-1}}{\cancel{\lambda}}\left(\nabla_\theta L_{\theta_k}(\theta)\big|_{\theta=\theta_k}\right)\right]^T \cancel{H} \left[\frac{\cancel{H^{-1}}}{\lambda}\left(\nabla_\theta L_{\theta_k}(\theta)\big|_{\theta=\theta_k}\right)\right]}_{= \frac{1}{2\lambda}\left(\nabla_\theta L_{\theta_k}(\theta)\big|_{\theta=\theta_k}\right)^T \underbrace{(H^{-1})^T}_{= H^{-1}} \left(\nabla_\theta L_{\theta_k}(\theta)\big|_{\theta=\theta_k}\right)} - \lambda\delta$$

$$\Rightarrow D(\lambda) = \frac{-1}{2\lambda}\left[\left(\nabla_\theta L_{\theta_k}(\theta)\big|_{\theta=\theta_k}\right)^T H^{-1}\left(\nabla_\theta L_{\theta_k}(\theta)\big|_{\theta=\theta_k}\right)\right] - \lambda\delta \ \#$$

(ii) find λ^* by solving $\frac{\partial D(\lambda)}{\partial \lambda} = 0$

$$\Rightarrow \frac{\partial D(\lambda)}{\partial \lambda} = \frac{1}{2\lambda^2}\left[\left(\nabla_\theta L_{\theta_k}(\theta)\big|_{\theta=\theta_k}\right)^T H^{-1}\left(\nabla_\theta L_{\theta_k}(\theta)\big|_{\theta=\theta_k}\right)\right] - \delta = 0$$

$$\Rightarrow \lambda^* = \sqrt{\frac{1}{2\delta}\left[\left(\nabla_\theta L_{\theta_k}(\theta)\big|_{\theta=\theta_k}\right)^T H^{-1}\left(\nabla_\theta L_{\theta_k}(\theta)\big|_{\theta=\theta_k}\right)\right]} \ \#$$

(b) since (a)-①: $\theta^* = \theta_k + \frac{1}{\lambda^*} H^{-1}\left(\nabla_\theta L_{\theta_k}(\theta)\big|_{\theta=\theta_k}\right)$

$$= \theta_k + \alpha H^{-1}\left(\nabla_\theta L_{\theta_k}(\theta)\big|_{\theta=\theta_k}\right)$$

$$\therefore \alpha = \frac{1}{\lambda^*} = \sqrt{2\delta} \cdot \left[\left(\nabla_\theta L_{\theta_k}(\theta)\big|_{\theta=\theta_k}\right)^T H^{-1}\left(\nabla_\theta L_{\theta_k}(\theta)\big|_{\theta=\theta_k}\right)\right]^{\frac{-1}{2}} \ \#$$