Reinforcement Learning
HW2

309505018
資科碩二

Problem 1

Recall that $L_{\pi_{\theta'}}(\pi_\theta) = \eta(\pi_{\theta'}) + \sum_s d_\mu(s) \sum_a \pi_\theta(a|s) A^{\pi_{\theta'}}(s,a)$

Then we have

$$\nabla_\theta L_{\pi_{\theta'}}(\pi_\theta) = \sum_s d_\mu(s) \sum_a \nabla_\theta \pi_\theta(a|s) A^{\pi_{\theta'}}(s,a)$$

On the other hand, note that $\eta(\pi_\theta) = \eta(\pi_{\theta'}) + \sum_s d_\mu(s) \cdot \sum_a \pi_\theta(a|s) A^{\pi_\theta}(s,a)$

$$\nabla_\theta \eta(\pi_\theta) = \sum_s \nabla_\theta \left(d_\mu^{\pi}(s) \cdot \sum_a \pi_\theta(a|s) A^{\pi}(s,a) \right)$$

$$= \sum_s \left[(\nabla_\theta d_\mu(s)) \left(\sum_a \pi_\theta(a|s) A^{\pi}(s,a) \right) + d_\mu(s) \sum_a \nabla_\theta \pi_\theta(a|s) A^{\pi_{\theta'}}(s,a) \right]$$

Then,

$$\nabla_\theta \eta(\pi_\theta)\Big|_{\theta=\theta'} = \sum_s \left[(\nabla_\theta d_\mu^{\pi}(s)) \left(\sum_a \pi_{\theta'} A^{\pi_\theta}(s,a) \right) + d_\mu^{\pi}(s) \sum_a \nabla_\theta \pi_\theta(a|s) A^{\pi_\theta}(s,a) \right]$$

$$= \sum_s d_\mu^{\pi_\theta}(s) \sum_a \nabla_\theta \pi_\theta(a|s) A^{\pi_\theta}(s,a)$$

$$= \nabla_\theta L_{\pi_{\theta'}}(\pi_\theta)$$

309505018

Problem 2

$$D(\lambda) = \min \{ -(\nabla_\theta L_{\theta_k}(\theta)|_{\theta=\theta_k})^T(\theta-\theta_k) + \lambda(\tfrac{1}{2}(\theta-\theta_k)^T H(\theta-\theta_k) - \delta) \}$$

H is a positive matrix and hence $L(\theta,\lambda)$ is strictly convex

$$\nabla_\theta L(\theta,\lambda) = -(\nabla_\theta L_{\theta_k}(\theta)|_{\theta=\theta_k}) + \lambda H(\theta-\theta_k)$$

since $L(\theta,\lambda)$ is strictly convex, then a point θ^*, λ^* the global minimum if and only if $\nabla_\theta L(\theta,\lambda)|_{\theta=\theta^*,\lambda=\lambda^*} = 0$

$$\Rightarrow \nabla_\theta L(\theta^*,\lambda^*) = -(\nabla_\theta L_{\theta_k}(\theta)|_{\theta=\theta_k}) + \lambda^* H(\theta^*-\theta_k) = 0$$

$$\Leftrightarrow \theta^* - \theta_k = \frac{1}{\lambda^*} H^{-1}(\nabla_\theta L_{\theta_k}(\theta)|_{\theta=\theta_k})$$

Hence,

$$D(\lambda) = L(\theta^*,\lambda) = \frac{1}{\lambda^*}(\nabla_\theta L_{\theta_k}(\theta)|_{\theta=\theta_k})^T H^{-1}(\nabla_\theta L_{\theta_k}(\theta)|_{\theta=\theta_k})$$

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

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

(a)

$$\lambda^* = \sqrt{\frac{(\nabla_\theta L_{\theta_k}(\theta)|_{\theta=\theta_k})^T H^{-1}(\nabla L_{\theta_k}(\theta)|_{\theta=\theta_k})}{2\delta}}.$$

(b)

$$\alpha = \frac{1}{\lambda^*}$$