HW3

111652040. 吳泓瑞

Problem 1

(i)
$$L_{\pi_{\theta_1}}(\pi_{\theta_1}) = \eta(\pi_{\theta_1}) + \sum_{s\in S} d_\mu^{\pi_{\theta_1}}(s)\cdot \underline{\sum_{a\in A} \pi_{\theta_1}(a|s)\cdot A^{\pi_{\theta_1}}(s,a)}$$

(by lemma 1)
$$= \eta(\pi_{\theta_1}) + \sum_{s\in S} d_\mu^{\pi_{\theta_1}}(s)\cdot 0$$
$$= \eta(\pi_{\theta_1}) \quad \#$$

(ii)
$$\nabla_\theta L_{\pi_{\theta_1}}(\pi_\theta)\Big|_{\theta=\theta_1} = \nabla_\theta \eta(\pi_{\theta})\Big|_{\theta=\theta_1} + \nabla_\theta \Big(\underline{\sum_{s\in S} d_\mu^{\pi_{\theta_1}}(s)\cdot \sum_{a\in A} \pi_\theta(a|s)\cdot A^{\pi_{\theta_1}}(s,a)}\Big)\Big|_{\theta=\theta_1}$$

(by lemma 2)
$$= \nabla_\theta \eta(\pi_{\theta_1})\Big|_{\theta=\theta_1} + \nabla_\theta \Big(\sum_{s\in S} d_\mu^{\pi_{\theta_1}}(s)\cdot \sum_{a\in A} \pi_\theta(a|s)\cdot A^{\pi_{\theta_1}}(s,a)\Big)\Big|_{\theta=\theta_1}$$
$$= \nabla_\theta \Big(\underline{\eta(\pi_{\theta_1}) + \sum_{s\in S} d_\mu^{\pi_{\theta_1}}(s)\cdot \sum_{a\in A} \pi_\theta(a|s)\cdot A^{\pi_{\theta_1}}(s,a)}\Big)\Big|_{\theta=\theta_1}$$

(by performance difference lemma)
$$= \nabla_\theta (\eta(\pi_\theta))\Big|_{\theta=\theta_1}$$

Lemma 1. $\sum_{a\in A} \pi_{\theta_1}(a|s)\cdot A^{\pi_{\theta_1}}(s,a) = 0$

$$\sum_{a\in A} \pi_{\theta_1}(a|s)\cdot \underline{A^{\pi_{\theta_1}}(s,a)}$$
$$= \sum_{a\in A} \pi_{\theta_1}(a|s)\cdot \big(\underline{Q^{\pi_{\theta_1}}(s,a) - V^{\pi_{\theta_1}}(s)}\big)$$
$$= \underline{\sum_{a\in A} \pi_{\theta_1}(a|s)\cdot Q^{\pi_{\theta_1}}(s,a)} - \Big(\sum_{a\in A} \pi_{\theta_1}(a|s)\Big)\cdot V^{\pi_{\theta_1}}(s)$$

(by bellman equation)
$$= \underline{V^{\pi_{\theta_1}}(s)} - 1\cdot V^{\pi_{\theta_1}}(s) = 0 \quad \blacksquare$$

lemma 2. $\nabla_\theta \left(\sum_{s \in S} d_\mu^{\pi_\theta}(s) \cdot \sum_{a \in A} \pi_\theta(a|s) \cdot A^{\pi_{\theta'}}(s,a) \right) \Big|_{\theta=\theta_1}$

$= \nabla_\theta \left(\sum_{s \in S} d_\mu^{\pi_{\theta_1}}(s) \cdot \sum_{a \in A} \pi_\theta(a|s) \cdot A^{\pi_{\theta'}}(s,a) \right) \Big|_{\theta=\theta_1}$

$\nabla_\theta \left(\sum_{s \in S} d_\mu^{\pi_\theta}(s) \cdot \sum_{a \in A} \pi_\theta(a|s) \cdot A^{\pi_{\theta'}}(s,a) \right) \Big|_{\theta=\theta_1}$

$= \sum_{s \in S} \nabla_\theta \left(d_\mu^{\pi_\theta}(s) \cdot \sum_{a \in A} \pi_\theta(a|s) \cdot A^{\pi_{\theta'}}(s,a) \right) \Big|_{\theta=\theta_1}$

$= \sum_{s \in S} \Big(\nabla_\theta \left(d_\mu^{\pi_\theta}(s) \right) \Big|_{\theta=\theta_1} \cdot \sum_{a \in A} \pi_{\theta_1}(a|s) \cdot A^{\pi_{\theta'}}(s,a) \Big)$ $= 0$

$+ d_\mu^{\pi_{\theta_1}}(s) \cdot \nabla_\theta \left(\sum_{a \in A} \pi_\theta(a|s) \cdot A^{\pi_{\theta'}}(s,a) \right) \Big|_{\theta=\theta_1} \Big)$

(by lemma 1)

$= \sum_{s \in S} d_\mu^{\pi_{\theta_1}}(s) \cdot \nabla_\theta \left(\sum_{a \in A} \pi_\theta(a|s) \cdot A^{\pi_{\theta'}}(s,a) \right) \Big|_{\theta=\theta_1}$

$= \nabla_\theta \left(\sum_{s \in S} d_\mu^{\pi_{\theta_1}}(s) \cdot \sum_{a \in A} \pi_\theta(a|s) \cdot A^{\pi_{\theta'}}(s,a) \right) \Big|_{\theta=\theta_1}$ ∎

Problem 2

(a)

① Find $\min_{\theta \in \mathbb{R}^d} L(\theta, \lambda)$

(take derivative)

$$\because \nabla_\theta L(\theta, \lambda) = -\left(\nabla_\theta L_{\theta_K}(\theta)\Big|_{\theta=\theta_K}\right) + \lambda \cdot H_{\theta_K} \cdot (\theta - \theta_K) = 0$$

$$\therefore \quad \theta - \theta_K = \frac{1}{\lambda} \cdot (H_{\theta_K})^{-1} \cdot \nabla_\theta L_{\theta_K}(\theta)\Big|_{\theta=\theta_K}$$

$$\Rightarrow D(\lambda) = -\left(\nabla_\theta L_{\theta_K}(\theta)\Big|_{\theta=\theta_K}\right)^T \cdot \frac{1}{\lambda} \cdot (H_{\theta_K})^{-1} \cdot \left(\nabla_\theta L_{\theta_K}(\theta)\Big|_{\theta=\theta_K}\right) - \lambda \cdot \delta$$

$$+ \lambda \cdot \frac{1}{\lambda} \cdot \left(\nabla_\theta L_{\theta_K}(\theta)\Big|_{\theta=\theta_K}\right)^T \cdot (H_{\theta_K})^{-1} \cdot H_{\theta_K} \cdot (H_{\theta_K})^{-1} \cdot \left(\nabla_\theta L_{\theta_K}(\theta)\Big|_{\theta=\theta_K}\right)$$

$$= \frac{-1}{2\lambda} \left(\nabla_\theta L_{\theta_K}(\theta)\Big|_{\theta=\theta_K}\right)^T \cdot (H_{\theta_K})^{-1} \cdot \left(\nabla_\theta L_{\theta_K}(\theta)\Big|_{\theta=\theta_K}\right) - \lambda \cdot \delta$$

② Find $\max_{\lambda \geq 0} D(\lambda)$

$$\because \nabla_\lambda D(\lambda) = \frac{1}{2 \cdot \lambda^2} \cdot \left(\nabla_\theta L_{\theta_K}(\theta)\Big|_{\theta=\theta_K}\right)^T \cdot (H_{\theta_K})^{-1} \cdot \left(\nabla_\theta L_{\theta_K}(\theta)\Big|_{\theta=\theta_K}\right) - \delta = 0$$

$$\therefore \lambda = \pm \sqrt{\frac{1}{2 \cdot \delta} \cdot \left(\nabla_\theta L_{\theta_K}(\theta)\Big|_{\theta=\theta_K}\right)^T \cdot (H_{\theta_K})^{-1} \cdot \left(\nabla_\theta L_{\theta_K}(\theta)\Big|_{\theta=\theta_K}\right)}$$

(take $\lambda > 0$)

Since $\lambda = 0$ boundary value doesn't exist,

$$\lambda^* = \sqrt{\frac{1}{2 \cdot \delta} \cdot \left(\nabla_\theta L_{\theta_K}(\theta)\Big|_{\theta=\theta_K}\right)^T \cdot (H_{\theta_K})^{-1} \cdot \left(\nabla_\theta L_{\theta_K}(\theta)\Big|_{\theta=\theta_K}\right)} \quad \#$$

(b)

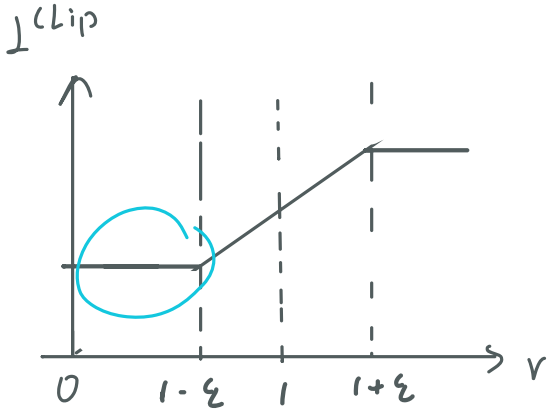
By (a)①, $\theta^* - \theta_K = \frac{1}{\lambda^*} \cdot (H_{\theta_K})^{-1} \cdot \nabla_\theta L_{\theta_K}(\theta)\Big|_{\theta=\theta_K}$

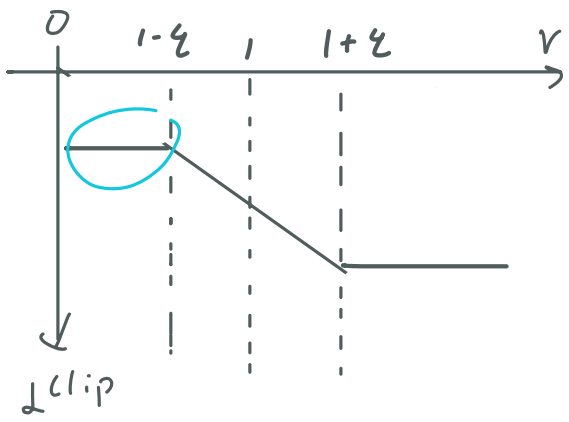
$\frac{1}{\lambda^*} = \alpha$

$$\Rightarrow \alpha = \sqrt{\frac{2 \cdot \delta}{\nabla_\theta L_{\theta_K}(\theta)\Big|_{\theta=\theta_K})^T \cdot (H_{\theta_K})^{-1} \cdot \left(\nabla_\theta L_{\theta_K}(\theta)\Big|_{\theta=\theta_K}\right)}} \quad \#$$

Problem 3

$P_t(\theta) > 0$	A_t	Return value of min	objective is clipped	sign of objective	Gradient
$P_t(\theta) \in [1-\epsilon, 1+\epsilon]$	+	$P_t(\theta) \cdot A_t$	No	+	✓
$P_t(\theta) \in [1-\epsilon, 1+\epsilon]$	−	$P_t(\theta) \cdot A_t$	No	−	✓
$P_t(\theta) < 1-\epsilon$	+	$(1-\epsilon) \cdot A_t$	yes	+	0
$P_t(\theta) < 1-\epsilon$	−	$(1-\epsilon) \cdot A_t$	yes	−	0
$P_t(\theta) > 1+\epsilon$	+	$(1+\epsilon) \cdot A_t$	yes	+	0
$P_t(\theta) > 1+\epsilon$	−	$(1+\epsilon) \cdot A_t$	yes	−	0





explain

Compared to Figure 1, we observe that $L^{clip}_{s,a}$ give a severe punishment on loss that is negative, while $\widetilde{L}^{clip}_{s,a}$ is clipped on both sides

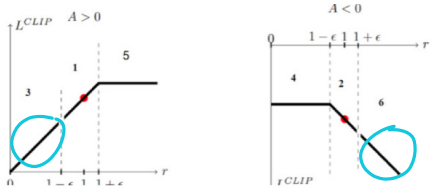


Figure 1: Behavior of the original PPO-clip objective.