

RL HW2 DDPG, TRPO, and PPO

110704054 謝弘廷

Problem 1.

(i)

$$\text{代入 } \pi_{\theta} = \pi_{\theta_1}$$

$$\mathcal{L}_{\pi_{\theta_1}}(\pi_{\theta_1}) = \eta(\pi_{\theta_1}) + \sum_s d_{\mu}^{\pi_{\theta_1}}(s) \sum_a \pi_{\theta_1}(a|s) A^{\pi_{\theta_1}}(s, a)$$

For Any State s ,

$$\sum_a \pi_{\theta_1}(a|s) A^{\pi_{\theta_1}}(s, a) = \sum_a \pi_{\theta_1}(a|s) [Q^{\pi_{\theta_1}}(s, a) - V^{\pi_{\theta_1}}(s)]$$

$$= V^{\pi_{\theta_1}}(s) - V^{\pi_{\theta_1}}(s) = 0$$

$$\Rightarrow \mathcal{L}_{\pi_{\theta_1}}(\pi_{\theta_1}) = \eta(\pi_{\theta_1}) + 0 = \eta(\pi_{\theta_1})_{\#}$$

(ii)

$$\nabla_{\theta} \mathcal{L}_{\pi_{\theta_1}}(\pi_{\theta})|_{\theta=\theta_1} = \nabla_{\theta} \eta(\pi_{\theta}) + \sum_s d_{\mu}^{\pi_{\theta_1}}(s) \sum_a \nabla_{\theta} [\pi_{\theta}(a|s)] A^{\pi_{\theta_1}}(s, a)|_{\theta=\theta_1}$$

$$A = Q - V \text{ 代入:}$$

$$\nabla_{\theta} \mathcal{L}_{\pi_{\theta_1}}(\pi_{\theta})|_{\theta=\theta_1} = \nabla_{\theta} \eta(\pi_{\theta}) + \sum_s d_{\mu}^{\pi_{\theta_1}}(s) \nabla_{\theta} (V^{\pi_{\theta_1}}(s) - V^{\pi_{\theta_1}}(s))$$

$$= \nabla_{\theta} \eta(\pi_{\theta_1})_{\#}$$

Problem 2.

Lagrangian: $\mathcal{L}(\theta, \lambda) = -g^T(\theta - \theta_k) + \lambda \left[\frac{1}{2}(\theta - \theta_k)^T H (\theta - \theta_k) - \delta \right]$

$$\nabla_{\theta} \mathcal{L}(\theta, \lambda) = -g + \lambda H(\theta - \theta_k)^T \stackrel{!}{=} 0$$

$$\Rightarrow \theta(\lambda) = \theta_k + \frac{1}{\lambda} H^{-1} g$$

$\theta(\lambda)$ 代 $\lambda \mathcal{L}$:

$$\nabla(\lambda) = \min_{\theta} \mathcal{L}(\theta, \lambda) = \mathcal{L}(\theta(\lambda), \lambda)$$

$$= -g^T \left(\frac{1}{\lambda} H^{-1} g \right) + \lambda \left[\frac{1}{2} \left(\frac{1}{\lambda} H^{-1} g \right)^T H \left(\frac{1}{\lambda} H^{-1} g \right) - \delta \right]$$

$$= -\frac{1}{2\lambda} g^T H^{-1} g - \lambda \delta$$

$$\frac{\partial \nabla}{\partial \lambda} = 0 \Rightarrow \lambda^* = \sqrt{\frac{g^T H^{-1} g}{2\delta}}$$

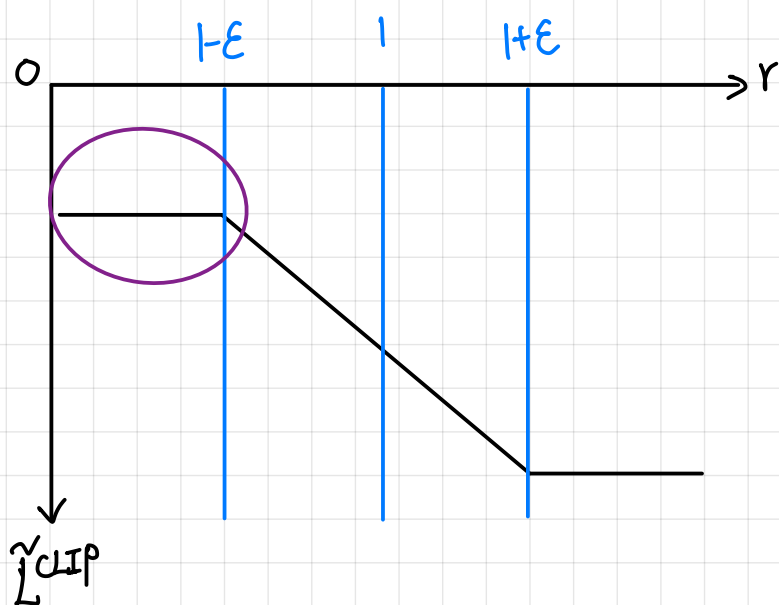
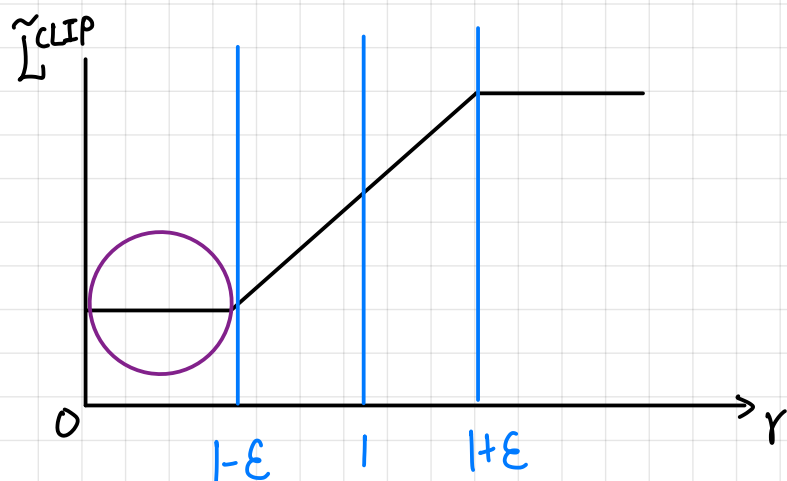
(b)

代回求 θ^*

\Rightarrow 步長 $\alpha = \frac{1}{\lambda^*}$, 則

$$\theta^* = \theta_k + \alpha H^{-1} g, \quad \alpha = \frac{1}{\lambda^*} = \sqrt{\frac{2\delta}{g^T H^{-1} g}}$$

Problem 3.



原始 L^{CLIP} : 根據優勢的正負動態調整裁剪策略, 優勢正 \rightarrow 保守更新
優勢負 \rightarrow 懲罰加重

變體 \tilde{L}^{CLIP} : 無論優勢正負, 都在 $r < 1 - \epsilon$ 或 $r > 1 + \epsilon$ 雙測鎖死