Problem 1

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

$$\begin{aligned}
L_{\pi_{\theta_1}}(\pi_{\theta_1}) &= \eta(\pi_{\theta_1}) + \sum_s d_\mu^{\pi_{\theta_1}}(s) \sum_a \pi_{\theta_1}(a|s) A^{\pi_{\theta_1}}(s,a) \\
&= \eta(\pi_{\theta_1}) + \sum_s d_\mu^{\pi_{\theta_1}}(s) \sum_a \pi_{\theta_1}(a|s) \left(Q^{\pi_{\theta_1}}(s,a) - V^{\pi_{\theta_1}}(s)\right) \\
&= \eta(\pi_{\theta_1}) + \sum_s d_\mu^{\pi_{\theta_1}}(s) \left(\underbrace{V^{\pi_{\theta_1}}(s) - V^{\pi_{\theta_1}}(s)}_{=0}\right) \\
&= \underline{\eta(\pi_{\theta_1})}
\end{aligned}$$

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

$\eta(\pi_\theta) = \eta(\pi_{\theta_1}) + \sum_s d_\mu^{\pi_\theta}(s) \sum_a \pi_\theta(a|s) A^{\pi_{\theta_1}}(s,a)$ — Average performance difference lemma

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

$$\begin{aligned}
\nabla_\theta \eta(\pi_\theta)\Big|_{\theta=\theta_1} &= \left(\sum_s \nabla_\theta d_\mu^{\pi_\theta}(s) \sum_a \pi_\theta(a|s) + \sum_s d_\mu^{\pi_\theta}(s) \sum_a \nabla_\theta \pi_\theta(a|s)\right)\Big|_{\theta=\theta_1} A^{\pi_{\theta_1}}(s,a) \\
&= \sum_s \nabla_\theta d_\mu^{\pi_\theta}(s) \sum_a \pi_\theta(a|s)\Big|_{\theta=\theta_1} A^{\pi_{\theta_1}}(s,a) + \sum_s d_\mu^{\pi_\theta}(s) \sum_a \nabla_\theta \pi_\theta(a|s)\Big|_{\theta=\theta_1} A^{\pi_{\theta_1}}(s,a) \\
&= \sum_s \nabla_\theta d_\mu^{\pi_\theta}(s)\Big|_{\theta=\theta_1} \underbrace{\boxed{\sum_a \pi_{\theta_1}(a|s) A^{\pi_{\theta_1}}(s,a)}}_{=0} + \sum_s d_\mu^{\pi_\theta}(s) \sum_a \nabla_\theta \pi_\theta(a|s)\Big|_{\theta=\theta_1} A^{\pi_{\theta_1}}(s,a) \\
&= \sum_s d_\mu^{\pi_{\theta_1}}(s) \sum_a \nabla_\theta \pi_\theta(a|s)\Big|_{\theta=\theta_1} A^{\pi_{\theta_1}}(s,a)
\end{aligned}$$

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

NAN PAO

Problem 2

(a) Set $\frac{d}{d\theta}\mathcal{L}(\theta,\lambda)=0$ to find $D(\lambda)$

$$\mathcal{L}(\theta,\lambda) = -\left(\nabla_\theta L_{\theta_k}(\theta)\big|_{\theta=\theta_k}\right)^T(\theta-\theta_k) + \lambda\left(\tfrac{1}{2}(\theta-\theta_k)^T H(\theta-\theta_k)-\delta\right)$$

$$\frac{d}{d\theta}\mathcal{L}(\theta,\lambda)=0$$

From matrix calculus, we know that

$$\frac{\partial a^T x}{\partial x} = a$$

$$\frac{\partial x^T A x}{\partial x} = 2Ax \quad \text{when } A \text{ is symmetric.}$$

So $\frac{d}{d\theta}\mathcal{L}(\theta,\lambda) = -\left(\nabla_\theta L_{\theta_k}(\theta)\big|_{\theta=\theta_k}\right) + \lambda\left(\tfrac{1}{\cancel{2}}\cdot\cancel{2}\,H\cdot(\theta-\theta_k)\right)=0$

since H is symmetric

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

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

Replace $\theta-\theta_k$ in $\mathcal{L}(\theta,\lambda)$ with ①:

$$\mathcal{L}(\theta,\lambda) = -\left(\nabla_\theta L_{\theta_k}(\theta)\big|_{\theta=\theta_k}\right)^T\cdot\frac{1}{\lambda}H^{-1}\left(\nabla_\theta L_{\theta_k}(\theta)\big|_{\theta=\theta_k}\right)$$

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

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

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

Then set $\frac{d}{d\lambda} D(\lambda) = 0$ to find λ^*

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

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

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

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

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

(b) By ①, we know that

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

$$\Rightarrow \theta^* = \theta_k + \underbrace{\boxed{\frac{1}{\lambda^*}}}_{=\alpha} H^{-1} \nabla_\theta L_{\theta_k}(\theta)\big|_{\theta=\theta_k}$$

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