a)

Want to show: $\nabla_\theta L_{\pi_{\theta_1}}(\pi_\theta)|_{\theta=\theta_1} = \nabla_\theta \eta(\pi_\theta)|_{\theta=\theta_1}$

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

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

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

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

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

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

$$= \sum_s [\nabla_\theta(d_\mu^{\pi_\theta}(s))|_{\theta=\theta_1} \cdot [\sum_a \pi_{\theta_1}(a|s)(Q^{\pi_{\theta_1}}(s,a) - V^{\pi_{\theta_1}}(s))]] + \sum_s [d_\mu^{\pi_{\theta_1}}(s) (\sum_a \nabla_\theta (a|s) A^{\pi_{\theta_1}}(s,a)|_{\theta=\theta_1})]$$

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

$$= \sum_s [\nabla_\theta(d_\mu^{\pi_\theta}(s))|_{\theta=\theta_1} \cdot 0] + \sum_s [d_\mu^{\pi_{\theta_1}}(s) \sum_a \nabla_\theta (a|s) A^{\pi_{\theta_1}}(s,a)|_{\theta=\theta_1}]$$

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

$= \underline{LHS}$ #.

1.

(b-1)

Want to show: $\forall f: S \to R$, $\forall \pi$,

$$(1-\gamma) E_{s\sim\mu}[f(s)] + E_{s\sim d_\mu^\pi, a\sim\pi, s'\sim P(\cdot|s,a)}[\gamma f(s')] - E_{s\sim d_\mu^\pi}[f(s)] = 0.$$

proof:

Leverage formulation (18) shown in the original paper:

$$d_\mu^\pi = (1-\gamma) \sum_{t=0}^{\infty} (\gamma P_\pi)^t \mu = (1-\gamma)(I-\gamma P_\pi)^{-1} \mu$$

(where $P_\pi \in R^{|S|\times|S|}$ denotes the transition matrix w. components $P_\pi(s'|s) = \int P(s'|s,a)\pi(a|s)da$)

Multiply both sides of (18) by $(I-\gamma P_\pi)$:

$$(I-\gamma P_\pi) d_\mu^\pi = (1-\gamma)\cancel{(I-\gamma P_\pi)}\cancel{(I-\gamma P_\pi)^{-1}}\mu$$

$$\Rightarrow (1-\gamma)\mu + \gamma P_\pi d_\mu^\pi - d_\mu^\pi = 0.$$

Take inner product with the vector $f \in R^{|S|}$:

$$(1-\gamma)\mu \cdot f + P_\pi d_\mu^\pi \gamma f - d_\mu^\pi \cdot f = 0$$

Rewrite this vector form into function function:

$$(1-\gamma) E_{s\sim\mu}[f(s)] + E_{s\sim d_\mu^\pi, a\sim\pi(\cdot|s), s'\sim P(\cdot|s,a)}[\gamma f(s')] - E_{s\sim d_\mu^\pi}[f(s)] = 0 \quad \#$$

(b-2)

Want to show: $\eta(\pi) = \frac{1}{1-\gamma} E_{s\sim d_\mu^\pi, a\sim\pi(\cdot|s)}[R(s,a)]$

$$= E_{s\sim\mu}[f(s)] + \frac{1}{1-\gamma} E_{s\sim d_\mu^\pi, a\sim\pi, s'\sim p(\cdot|s,a)}[R(s,a) + \gamma f(s') - f(s)]$$

proof:

Divide the result in (b-1) by $(1-\gamma)$:

$$E_{s\sim\mu}[f(s)] + \frac{1}{1-\gamma} E_{s\sim d_\mu^\pi, a\sim\pi, s'\sim p(\cdot|s,a)}[\gamma f(s')] - \frac{1}{1-\gamma} E_{s\sim d_\mu^\pi}[f(s)] = 0$$

$$\therefore \eta(\pi) = \frac{1}{1-\gamma} E_{s\sim d_\mu^\pi, a\sim\pi(\cdot|s)}[R(s,a)]$$

$$= \frac{1}{1-\gamma} E_{s\sim d_\mu^\pi, a\sim\pi(\cdot|s)}[R(s,a)] + 0$$

$$= \frac{1}{1-\gamma} E_{s\sim d_\mu^\pi, a\sim\pi(\cdot|s)}[R(s,a)] + \Big(E_{s\sim\mu}[f(s)] + \frac{1}{1-\gamma} E_{s\sim d_\mu^\pi, a\sim\pi, s'\sim p(\cdot|s,a)}[\gamma f(s')] - \frac{1}{1-\gamma} E_{s\sim d_\mu^\pi}[f(s)]\Big)$$

$$= E_{s\sim\mu}[f(s)] + \frac{1}{1-\gamma}\Big(E_{s\sim d_\mu^\pi, a\sim\pi(\cdot|s)}[R(s,a)] + E_{s\sim d_\mu^\pi, a\sim\pi, s'\sim p(\cdot|s,a)}[\gamma f(s')] - E_{s\sim d_\mu^\pi}[f(s)]\Big)$$

$$= E_{s\sim\mu}[f(s) + \frac{1}{1-\gamma} E_{s\sim d_\mu^\pi, a\sim\pi, s'\sim p(\cdot|s,a)}[R(s,a) + \gamma f(s') - f(s)]$$

$\#$

(C-1)

Want to show: $\forall \pi, \pi'$, $d_\mu^{\pi'} - d_\mu^{\pi} = \gamma (I - \gamma P^{\pi'})^{-1} (P^{\pi'} - P^{\pi}) d_\mu^{\pi}$.

proof:

By definition, $d_\mu^{\pi} = (1-\gamma) \sum_{t=0}^{N} (\gamma P^{\pi})^t \mu$

$= (1-\gamma)(I - \gamma P^{\pi})^{-1} \mu$

(where $P^{\pi}(s'|s) = \int P(s'|s,a)\pi(a|s)\,da \Rightarrow P_\pi^t = P_\pi P_\pi^{t-1} = P_\pi^t \mu$)

$d_\mu^{\pi'} - d_\mu^{\pi} = (1-\gamma)(I - \gamma P^{\pi'})^{-1} \mu - (1-\gamma)(I - \gamma P^{\pi})^{-1} \mu$

$= (1-\gamma)[(I - \gamma P^{\pi'})^{-1} - (I - \gamma P^{\pi})^{-1}] \mu$.

Let $G' = (I - \gamma P^{\pi'})^{-1}$, $G = (I - \gamma P^{\pi})^{-1}$, $\Delta = P^{\pi'} - P^{\pi}$

Then $G^{-1} - G'^{-1} = (I - \gamma P^{\pi}) - (I - \gamma P^{\pi'})$

$= \gamma \Delta$

Left multiply by G and right multiply by G',

$G(G^{-1} - G'^{-1})G' = G' - G = \gamma G \Delta G' = \gamma G' \Delta G$

$\therefore d_\mu^{\pi'} - d_\mu^{\pi} = (1-\gamma)[G' - G]\mu$

$= (1-\gamma)\gamma G' \Delta G \mu$

$= (1-\gamma)\gamma (I - \gamma P^{\pi'})^{-1} (P^{\pi'} - P^{\pi})(I - \gamma P^{\pi})^{-1} \mu$

$= \gamma (I - \gamma P^{\pi'})^{-1} (P^{\pi'} - P^{\pi}) d_\mu^{\pi}$ #

(C-2)

Want to show: $\| d^{\pi'} - d^{\pi} \|_1 \le \frac{2\gamma}{1-\gamma} E_{s \sim d_\mu^{\pi}} [D_{TV}(\pi'(\cdot|s) \| \pi(\cdot|s))]$

proof:

Leverage the result in (C-1):

$\| d_\mu^{\pi'} - d_\mu^{\pi} \|_1 = \gamma \| G' \Delta d_\mu^{\pi} \|_1 \le \gamma \|G'\|_1 \|\Delta d_\mu^{\pi}\|_1$

For the $\|G'\|_1$ term:

$\|G'\|_1 = \|(I - \gamma P^{\pi'})^{-1}\|_1 = \| \sum_{t=0}^{\infty} (\gamma P^{\pi'})^t \|_1 \le \sum_{t=0}^{\infty} \gamma^t \|P^{\pi'}\|_1^t = \frac{1}{1-\gamma}$

For the $\|\Delta d_\mu^{\pi}\|_1$ term:

$\|\Delta d_\mu^{\pi}\|_1 = \|(P^{\pi'} - P^{\pi}) d_\mu^{\pi}\|_1 = \sum_{s'} \left| \sum_{s} (P^{\pi'}(s'|s) - P^{\pi}(s'|s)) d_\mu^{\pi}(s) \right|$

$\le \sum_{s,s'} |P^{\pi'}(s'|s) - P^{\pi}(s'|s)| \, d_\mu^{\pi}(s)$

$= \sum_{s,s'} \left| \sum_a P(s'|s,a)(\pi'(a|s) - \pi(a|s)) \right| d_\mu^{\pi}(s)$

$\le \sum_{s,a,s'} P(s'|s,a) |\pi'(a|s) - \pi(a|s)| \, d_\mu^{\pi}(s)$

$= \sum_{s,a} |\pi'(a|s) - \pi(a|s)| \, d_\mu^{\pi}(s) = 2 E_{s \sim d_\mu^{\pi}} [D_{TV}(\pi'(\cdot|s) \| \pi(\cdot|s))]$

1.

(d)

Want to show: $\eta(\pi') - \eta(\pi) \geq \frac{1}{1-\gamma}\left(Y_{\pi,f}(\pi') - 2\varepsilon_f^{\pi'} D_{TV}(d_\mu^{\pi'} \| d_\mu^{\pi})\right)$

proof:

$$Y_{\pi,f}(\pi') := E_{s\sim d_\mu^\pi,\, a\sim\pi(\cdot|s),\, s'\sim p(\cdot|s,a)}\left[\left(\frac{\pi'(a|s)}{\pi(a|s)} - 1\right)(R(s,a) + \gamma f(s') - f(s))\right]$$

$$\varepsilon_f^{\pi'} := \max_s \left| E_{a\sim\pi'(\cdot|s),\, s'\sim p(\cdot|s,a)}[R(s,a) + \gamma f(s') - f(s)]\right|$$

Let $\delta_f(s,a,s') = R(s,a) + \gamma f(s') - f(s)$,

By the result in (b-2),

$$\eta(\pi') - \eta(\pi) = \frac{1}{1-\gamma}\left[E_{s\sim d_\mu^{\pi'},\, a\sim\pi',\, s'\sim p(\cdot|s,a)}[\delta_f(s,a,s')] - E_{s\sim d_\mu^{\pi},\, a\sim\pi,\, s'\sim p(\cdot|s,a)}[\delta_f(s,a,s')]\right]$$

Let $\bar{\delta}_f^{\pi'} \in \mathbb{R}^{|S|}$ denotes the vector of components $\bar{\delta}_f^{\pi'}(s) = E_{a\sim\pi'(\cdot|s),\, s'\sim p(\cdot|s,a)}[\delta_f(s,a,s')]$

$$E_{s\sim d_\mu^{\pi'},\, a\sim\pi',\, s'\sim p}[\delta_f(s,a,s')] = \langle d^{\pi'}, \bar{\delta}_f^{\pi'}\rangle = \langle d^{\pi}, \bar{\delta}_f^{\pi'}\rangle + \langle d^{\pi'} - d^{\pi}, \bar{\delta}_f^{\pi'}\rangle$$

By Hölder's inequality,

$$\langle d_\mu^{\pi}, \bar{\delta}_f^{\pi'}\rangle + \| d_\mu^{\pi'} - d_\mu^{\pi}\|_p \|\bar{\delta}_f^{\pi'}\|_q \geq E_{s\sim d_\mu^{\pi'},\, a\sim\pi',\, s\sim p}[\delta_f(s,a,s')]$$

$$\geq \langle d_\mu^{\pi}, \bar{\delta}_f^{\pi'}\rangle - \| d_\mu^{\pi'} - d_\mu^{\pi}\|_p \|\bar{\delta}_f^{\pi'}\|_q,$$

(where $p, q \in [1,\infty]$, s.t. $\frac{1}{p} + \frac{1}{q} = 1$).

$\because \| d_\mu^{\pi'} - d_\mu^{\pi}\|_1 = 2D_{TV}(d_\mu^{\pi'} \| d_\mu^{\pi})$, and

$\|\bar{\delta}_f^{\pi'}\|_\infty = \varepsilon_f^{\pi'}$, and

$$\langle d_\mu^{\pi}, \bar{\delta}_f^{\pi'}\rangle = E_{s\sim d_\mu^\pi\, a\sim\pi',\, s'\sim p(\cdot|s,a)}[\delta_f(s,a,s')]$$

$$= E_{s\sim d_\mu^\pi,\, a\sim\pi,\, s'\sim p(\cdot|s,a)}\left[\frac{\pi'(a|s)}{\pi(a|s)}\delta_f(s,a,s')\right] \text{...... by importance sampling method.}$$

$\therefore \eta(\pi') - \eta(\pi) \geq \frac{1}{1-\gamma}\left(Y_{\pi,f}(\pi') - 2\varepsilon_f^{\pi'} D_{TV}(d_\mu^{\pi'} \| d_\mu^{\pi})\right)$ #

1.

(c)

Want to show: $\eta(\pi') - \eta(\pi) \geq \frac{1}{1-\gamma} E_{s\sim d_\mu^\pi, a\sim\pi'(\cdot|s)}\left[A^\pi(s,a) - \frac{2\varepsilon^{\pi'}\gamma}{(1-\gamma)}\left(D_{TV}(\pi'(\cdot|s)\|\pi(\cdot|s))\right)\right]$

proof:

observe that $A^\pi(s,a) = E_{s'\sim p}[\delta_{V^\pi}(s,a,s')|s,a]$,

By (d):

$$\eta(\pi') - \eta(\pi) \geq \frac{1}{1-\gamma}\left(E_{s\sim d_\mu^\pi, a\sim\pi'(\cdot|s), s'\sim p(\cdot|s,a)}[\delta_f(s,a,s')] - 2\varepsilon^{\pi'} D_{TV}(d_\mu^{\pi'}\|d_\mu^\pi)\right)$$

$$= \frac{1}{1-\gamma} E_{s\sim d_\mu^\pi, a\sim\pi'(\cdot|s)}\left[A^\pi(s,a) - 2\varepsilon^{\pi'} D_{TV}(d_\mu^{\pi'}\|d_\mu^\pi)\right]$$

$$= \frac{1}{1-\gamma} E_{s\sim d_\mu^\pi, a\sim\pi'(\cdot|s)}\left[A^\pi(s,a) - \varepsilon^{\pi'}\|d_\mu^{\pi'} - d_\mu^\pi\|_1\right] \ldots\ldots \boxed{1}$$

By (c-2):

$$\|d_\mu^{\pi'} - d_\mu^\pi\|_1 \leq \frac{2\gamma}{1-\gamma} E_{s\sim d_\mu^\pi}[D_{TV}(\pi'(\cdot|s)\|\pi(\cdot|s))]$$

$$\Rightarrow -\|d_\mu^{\pi'} - d_\mu^\pi\|_1 \geq \frac{2\gamma}{1-\gamma} E_{s\sim d_\mu^\pi}[D_{TV}(\pi'(\cdot|s)\|\pi(\cdot|s))] \ldots\ldots\ldots\ldots \boxed{2}$$

Combining $\boxed{1}$ and $\boxed{2}$:

$$\eta(\pi') - \eta(\pi) \geq \frac{1}{1-\gamma} E_{s\sim d_\mu^\pi, a\sim\pi'(\cdot|s)}\left[A^\pi(s,a) - \varepsilon^{\pi'}\|d_\mu^{\pi'} - d_\mu^\pi\|_1\right]$$

$$\geq \frac{1}{1-\gamma} E_{s\sim d_\mu^\pi, a\sim\pi'(\cdot|s)}\left[A^\pi(s,a) - \frac{2\varepsilon^{\pi'}\gamma}{(1-\gamma)}\left(D_{TV}(\pi'(\cdot|s)\|\pi(\cdot|s))\right)\right]$$

#

Want to show : $D(\lambda, \nu) := \min_{\theta \in \mathbb{R}^d} \left\{ -g^T(\theta - \theta_0) + \nu^T(c + B^T(\theta - \theta_0)) + \lambda\left(\frac{1}{2}(\theta - \theta_k)^T H(\theta - \theta_k) - \delta\right)\right\}$

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

proof :

The dual problem of (OPT) is : $\max_{\lambda \geq 0, \nu \geq 0} D(\lambda, \nu)$.

$$\max_{\substack{\lambda\geq 0\\ \nu\geq 0}} D(\lambda, \nu) = \max_{\substack{\lambda\geq 0\\ \nu\geq 0}} \min_{\theta \in \mathbb{R}^d} \left\{ -g^T(\theta-\theta_0) + \nu^T(c + B^T(\theta - \theta_0)) + \frac{\lambda}{2}\left((\theta-\theta_k)^T H(\theta-\theta_k) - \delta\right)\right\}$$

$$= \max_{\substack{\lambda\geq 0\\ \nu\geq 0}} \min_{\theta \in \mathbb{R}^d} \left\{ \frac{\lambda}{2}(\theta-\theta_k)^T H(\theta-\theta_k) + (-g^T + \nu^T B^T)(\theta - \theta_0) + (\nu^T c - \frac{\lambda}{2}\delta)\right\}$$

Let $x = \theta - \theta_k \Rightarrow x^* = \frac{1}{\lambda} H^{-1}(g - B\nu)$, by solving $\nabla_x L(x, \lambda, \nu) = 0$

plugin x^* :

$$\max_{\substack{\lambda\geq 0\\ \nu\geq 0}} D(\lambda, \nu) = \max_{\substack{\lambda\geq 0\\ \nu\geq 0}} \left\{ \frac{1}{2\lambda}(g - B\nu)^T H^{-1}(g - B\nu) - (g - \nu B)^T \frac{1}{\lambda} H^{-1}(g - B\nu) + (\nu^T c - \frac{\lambda}{2}\delta)\right\}$$

$$= \max_{\substack{\lambda\geq 0\\ \nu\geq 0}} \frac{-1}{2\lambda}(g - B\nu)^T H^{-1}(g - B\nu) + (\nu^T c - \frac{\lambda\delta}{2})$$

$$= \max_{\substack{\lambda\geq 0\\ \nu\geq 0}} \frac{-1}{2\lambda}\left(g^T H^{-1} g - 2g^T H^{-1} B\nu + \nu^T B^T H^{-1} B\nu\right) + \nu^T c - \frac{\lambda\delta}{2}$$

$$\therefore D(\lambda, \nu) = \frac{-1}{2\lambda}\left(g^T H^{-1} g - 2g^T H^{-1} B\nu + \nu^T B^T H^{-1} B\nu\right) + \nu^T c - \frac{\lambda\delta}{2}$$

#.