HW 8

(one-pt grad estimator)

$g_t = \frac{1}{\sigma^2}(a_t - \mu_{\theta_t}(x_t))(r_t(x_t, a_t) - b_t(x_t))$ $\qquad b_t : \mathcal{X} \to \mathbb{R}$; time varying baseline

$\nabla_a r_t(x_0, a_0) = \nabla_a r_t(x_0, a)\big|_{a=a_0}$

a) $r_t(x_t, \cdot)$; $\exists v_t(x_t) \in \mathbb{R}^d$ and $c_t(x_t) \in \mathbb{R}$ s.t.

$\forall a,\ r_t(x_t, a) = c_t(x_t) + v_t(x_t)^T a$; <u>Prove $\mathbb{E}_{a_t}(g_t) = v_t(x_t)$</u>

$$\mathbb{E}_{a_t}[g_t] = \mathbb{E}_{a_t}\left[\frac{1}{\sigma^2}(a_t - \mu_{\theta_t}(x_t))(r_t(x_t, a_t) - b_t(x_t))\right]$$

$$= \mathbb{E}_{a_t}\left[\frac{1}{\sigma^2}(a_t - \mu_{\theta_t}(x_t))(c_t(x_t) + v_t(x_t)^T a - b_t(x_t))\right] = \mathbb{E}_{a_t}\left[\frac{1}{\sigma^2}(a_t - \mu_{\theta_t}(x_t))(v_t(x_t)^T a + \cancel{c_t(x_t) - b_t(x_t)})\right]$$

$\to$ independent of action and $\mu_{z_t} = 0$

$$= \mathbb{E}_{a_t}\left[\frac{1}{\sigma^2}(a_t - \mu_{\theta_t}(x_t))v_t(x_t)^T a\right] = \frac{1}{\sigma^2}\mathbb{E}_{a_t}\left[\underbrace{(a_t - \mu_{\theta_t}(x_t))}_{z_t = a_t - \mu_{\theta_t}(x_t)} v_t(x_t)^T a_t\right] = \frac{1}{\sigma^2}\mathbb{E}_{a_t}\left[z_t v_t(x_t)^T(z_t + \mu_{\theta_t}(x_t))\right] = \frac{1}{\sigma^2}\mathbb{E}_{a_t}\left[z_t v_t(x_t)^T z_t + z_t v_t(x_t)^T \mu_{\theta_t}(x_t)\right]$$

$$= \frac{1}{\sigma^2}\left(\mathbb{E}_{a_t}[z_t v_t(x_t)^T z_t] + \underbrace{\mathbb{E}[z_t v_t(x_t)^T \cancel{\mu_{\theta}(x_t)}]}_{\text{independence}}\right) = \frac{1}{\sigma^2}\left(v_t(x_t)\mathbb{E}_{a_t}[z_t z_t^T] + \underbrace{\cancel{v_t(x_t)\mathbb{E}_{a_t}[z_t]}}_{0;\ \hookrightarrow \mu_{z_t} = 0}\right) = \frac{1}{\sigma^2}v_t(x_t)\underbrace{\mathbb{E}_{a_t}[(a - \mu_\theta(x_t))(a - \mu_\theta(x_t))]}_{\hookrightarrow \sigma^2 \text{ by definition of covariance}}$$

$$= \frac{\cancel{\sigma^2}}{\cancel{\sigma^2}} v_t(x_t)^T = \boxed{v_t(x_t)^T}$$

$$\boxed{\therefore \mathbb{E}_{a_t}[g_t] = v_t(x_t)^T}$$

<u>Prove:</u>

b) $\|\mathbb{E}_{a_t}[g_t] - \nabla_a r_t(x_t, \mu_{\theta_t}(x_t))\| \leq \sqrt{\frac{d(d+2)(d+4)}{4}} L\sigma$ $\qquad z_t = a_t - \mu_{\theta_t}(x_t)$ $\qquad z_t = a_t - \mu_{\theta_t}(x_t)$

<u>Note:</u> using Lemma 1: $\underbrace{\left|r_t(x_t, a) - [r_t(x_t, \mu_{\theta_t}(x_t)) + \nabla_a r_t(x_t, \mu_{\theta_t}(x_t))^T(a - \mu_{\theta_t}(x_t))]\right|}_{\hookrightarrow R(z)} \leq \frac{L}{2}\|a - \mu_{\theta_t}(x_t)\|^2$

for $r_t(x_t, \cdot)$ as provided in the question

<u>$\mathbb{E}[g_t]$</u>

$g_t = \frac{1}{\sigma^2}(a_t - \mu_{\theta_t}(x_t))(r_t(x_t, a_t) - b_t(x_t))$

$\mu_{z_t} = 0$; b_t is a constant

$$\mathbb{E}[g_t] = \mathbb{E}\left[\frac{1}{\sigma^2}(a_t - \mu_{\theta_t}(x_t))(r_t(x_t, a_t) - b_t(x_t))\right] = \frac{1}{\sigma^2}\mathbb{E}\left[z_t(r_t(x_t, z_t + \mu_{\theta_t}) - b_t(x_t))\right] = \frac{1}{\sigma^2}\mathbb{E}_{z_t}[z_t r_t(x_t, z_t + \mu_\theta) - z_t b_t(x_t)] = \frac{1}{\sigma^2}\left(\mathbb{E}_{z_t}[z_t r_t(x_t, z_t + \mu_\theta)] - \cancel{\mathbb{E}_{z_t}[z_t b_t(x_t)]}\right)$$

$$= \frac{1}{\sigma^2}\mathbb{E}\left[z_t r_t(x_t, z_t + \mu_{\theta_t})\right]$$

$$= \frac{1}{\sigma^2}\mathbb{E}_{z_t}\Big[z_t\underbrace{(r_t(x_t, \mu_{\theta_t}(x_t)) + \nabla_a r_t(x_t, \mu_{\theta_t}(x_t))^T(a_t - \mu_{\theta_t}(x_t)) + R(a_t))}_{\text{first order Taylor expansion}}\Big]$$

$$= \frac{1}{\sigma^2}\mathbb{E}_{z_t}\left[z_t(r_t(x_t, \mu_{\theta_t}(x_t)) + z_t\nabla_a r_t(x_t, \mu_{\theta_t}(x_t))^T z_t + z_t R(z_t)\right]$$

$$= \frac{1}{\sigma^2}\left(\underbrace{\cancel{\mathbb{E}_{z_t}[z_t(r_t(x_t, \mu_{\theta_t}(x_t))]}}_{0} + \mathbb{E}_{z_t}[\overset{z_t}{\cancel{z_t}}\nabla_a r_t(x_t, \mu_{\theta_t}(x_t))^T z_t] + \mathbb{E}[z_t R(z_t)]\right)$$

$$\frac{1}{\sigma^2}\left(\mathbb{E}_{z_t}\left[z_t \nabla_a r_t(x_t, \mu_\theta)^T z_t\right] + \mathbb{E}_{z_t}\left[z_t R(z_t)\right]\right) = \frac{1}{\sigma^2}\left(\nabla_a r_t(x_t, \mu_\theta)\, \mathbb{E}_{z_t}\left[z_t z_t^T\right] + \mathbb{E}_{z_t}\left[z_t R(z_t)\right]\right) = \frac{1}{\sigma^2}\left(\nabla_a r_t(x_t, \mu_\theta)\, \sigma^2 I + \mathbb{E}_{z_t}\left[z_t R(z_t)\right]\right]$$

$$= \nabla_a r_t(x_t, \mu_\theta) + \frac{1}{\sigma^2}\mathbb{E}_{z_t}\left[z_t R(z_t)\right]$$

$$\mathbb{E}_{a_t}[g_t] = \nabla_a r_t(x_t, \mu_\theta) + \frac{1}{\sigma^2}\mathbb{E}_{z_t}\left[z_t R(z_t)\right]$$

$$\left\|\mathbb{E}_{a_t}[g_t] - \nabla_a r_t(x_t, \mu_\theta(x_t))\right\| = \left\|\frac{1}{\sigma^2}\mathbb{E}_{z_t}\left[z_t R(z_t)\right]\right\|$$

$$\mathbb{E}_{z_t}\left[\|z_t\| \, \|R(z_t)\|\right] \le \frac{1}{\sigma^2}\frac{L}{2}\mathbb{E}_{z_t}\left[\|z_t\| \, \|z_t\|^2\right] = \frac{L}{\sigma^2 2}\mathbb{E}_{z_t}\left[\|z_t\|^3\right]$$ using Lemma 1

let $w_t \sim N(0, I)$; $z_t = \sigma w_t$

$$\mathbb{E}_{z_t}\left[\|z_t\| \, \|R(z_t)\|\right] \le \frac{L}{2\sigma^2}\mathbb{E}_{w_t}\left[\|\sigma w_t\|^3\right] = \frac{L}{2\cancel{\sigma^2}}\sigma^{\cancel{3}}\mathbb{E}_{w_t}\left[\|w_t\|^3\right] = \frac{L\sigma}{2}\sqrt{\left(\mathbb{E}_{w_t}\left[\|w_t\|^3\right]\right)^2} = \frac{L\sigma}{2}\sqrt{d(d+2)(d+4)}$$ using Lemma 2

$$\boxed{\mathbb{E}_{z_t}\left[\|z_t\| \, \|R(z_t)\|\right] \le L\sigma\sqrt{\frac{d(d+2)(d+4)}{4}}}$$

$$\boxed{\therefore \left\|\mathbb{E}_{a_t}[g_t] - \nabla_a r_t(x_t, \mu_{\theta_t}(x_t))\right\| \le \sqrt{\frac{d(d+2)(d+4)}{4}}\, L\sigma}$$

$$\pi_\theta(a|x) = \frac{1}{(2\pi\sigma^2)^{\frac{d}{2}}} \exp\left(-\frac{\|a-\mu_\theta(x)\|^2}{2\sigma^2}\right)$$

c) $$\underbrace{\frac{\pi_\theta(a_t|x_t)}{\pi_{\theta_t}(a_t|x_t)}(r_t(x_t,a) - b_t(x_t)) - \frac{1}{\eta} KL(\pi_\theta(\cdot|x_t), \pi_{\theta_t}(\cdot|x_t))} \cong \langle \mu_\theta(x_t) - \mu_{\theta_t}(x_t), g_t \rangle - \frac{1}{2\eta\sigma^2}\|\mu_\theta(x_t) - \mu_\theta(x_t)\|^2$$

$$\frac{\pi_\theta(a_t|x_t)}{\pi_{\theta_t}(a_t|x_t)}(r_t(x_t,a_t) - b_t(x_t)) - \frac{1}{\eta}\left(\log\frac{\sigma_{\pi_{\theta_t}}}{\sigma_{\pi_\theta}} + \frac{(\mu_{\pi_\theta} - \mu_{\theta_t})^2 + \sigma_\theta^2}{2\sigma_{\theta_t}} - \frac{1}{2}\right)$$

$$\frac{\frac{\exp\left(-\frac{\|a_t-\mu_\theta(x_t)\|^2}{2\sigma^2}\right)}{(2\pi\sigma_\theta^2)^{\frac{d}{2}}}}{\frac{\exp\left(-\frac{\|a_t-\mu_{\theta_t}(x_t)\|^2}{2\sigma^2}\right)}{(2\pi\sigma_{\theta_t}^2)^{\frac{d}{2}}}}(r_t(x_t,a) - b_t(x_t)) - \frac{1}{\eta}\left(\log\frac{\sigma_{\theta_t}}{\sigma_\theta} + \frac{(\mu_\theta - \mu_{\theta_t})^2 + \sigma_\theta^2}{2\sigma_{\theta_t}} - \frac{1}{2}\right)$$

$$\frac{\exp\left(-\frac{\|a_t-\mu_\theta(x_t)\|^2}{2\sigma_\theta^2}\right)(\cancel{2\pi}\sigma_{\theta_t}^{\cancel{2}})^{\frac{d}{\cancel{2}}}}{\exp\left(-\frac{\|a_t-\mu_{\theta_t}(x_t)\|^2}{2\sigma_{\theta_t}^2}\right)(\cancel{2\pi}\sigma_\theta^{\cancel{2}})^{\frac{d}{\cancel{2}}}}(r_t(x_t,a_t) - b_t(x_t)) - \frac{1}{\eta}\left(\log\frac{\sigma_{\theta_t}}{\sigma_\theta} + \frac{(\mu_\theta - \mu_{\theta_t})^2 + \sigma_\theta^2}{2\sigma_{\theta_t}} - \frac{1}{2}\right)$$

$$\exp\left(-\frac{\|a_t-\mu_\theta(x_t)\|^2}{2\sigma_\theta^2} + \frac{\|a_t-\mu_{\theta_t}(x_t)\|^2}{2\sigma_{\theta_t}^2}\right)\frac{\sigma_{\theta_t}^d}{\sigma_\theta^d}(r_t(x_t,a_t) - b_t(x_t)) - \frac{1}{\eta}\left(\log\frac{\sigma_{\theta_t}}{\sigma_\theta} + \frac{(\mu_\theta - \mu_{\theta_t})^2 + \sigma_\theta^2}{2\sigma_{\theta_t}} - \frac{1}{2}\right)$$

$$\exp\left(\frac{\|a_t-\mu_{\theta_t}(x_t)\|^2 - \|a_t-\mu_\theta(x_t)\|^2}{2\sigma_\theta^2}\right)\overset{1}{\cancel{\frac{\sigma_{\theta_t}^d}{\sigma_\theta^d}}}(r_t(x_t,a_t) - b_t(x_t)) - \frac{1}{\eta}\left(\overset{0}{\cancel{\log\left(\frac{\sigma_{\theta_t}}{\sigma_\theta}\right)}} + \frac{(\mu_\theta - \mu_{\theta_t})^2 + \sigma_\theta^2}{2\sigma_{\theta_t}} - \frac{1}{2}\right)$$

$$\|a_t - \mu_{\theta_t}(x_t)\|^2 = (a_t - \mu_{\theta_t}(x_t))^T(a_t - \mu_{\theta_t}(x_t)) = \|a_t\|^2 - 2a_t^T\mu_{\theta_t}(x_t) + \|\mu_{\theta_t}(x_t)\|^2$$

focused on first term

$$\exp\left(\frac{\cancel{\|a_t\|^2} - 2a_t^T\mu_{\theta_t}(x_t) + \|\mu_{\theta_t}(x_t)\|^2 - \cancel{\|a_t\|^2} + 2a_t^T\mu_\theta(x_t) - \|\mu_\theta(x_t)\|^2}{2\sigma_\theta^2}\right)(r_t(x_t,a_t) - b_t(x_t)) - \ldots$$

$$\exp\left(\frac{1}{2\sigma_\theta^2}\left(2a_t^T(\mu_\theta(x_t) - \mu_{\theta_t}(x_t)) + \|\mu_{\theta_t}(x_t)\|^2 - \|\mu_\theta(x_t)\|^2\right)\right)(r_t(x_t,a_t) - b_t(x_t)) - \ldots$$

let $\Delta\mu = \mu_\theta(x_t) - \mu_{\theta_t}(x_t)$

$$\|\mu_\theta(x_t)\|^2 = \|\Delta\mu + \mu_{\theta_t}(x_t)\| = (\Delta\mu + \mu_{\theta_t}(x_t))^T(\Delta\mu + \mu_{\theta_t}(x_t)) =$$

$$\|\Delta\mu\|^2 + 2\Delta\mu^T\mu_{\theta_t}(x_t) + \|\mu_{\theta_t}(x_t)\|^2$$

$$\exp\left(\frac{1}{2\sigma_\theta^2}\left(2a_t^T\Delta\mu + \cancel{\|\mu_{\theta_t}(x_t)\|^2} - \|\Delta\mu\|^2 - 2\Delta\mu^T\mu_{\theta_t}(x_t) - \cancel{\|\mu_{\theta_t}(x_t)\|^2}\right)\right)(r_t(x_t,a_t) - b_t(x_t)) - \ldots$$

$$\exp\left(\frac{1}{2\sigma_\theta^2}\left(2(a_t^T - \mu_{\theta_t}(x_t))\Delta\mu - \|\Delta\mu\|^2\right)\right)(r_t(x_t,a_t) - b_t(x_t)) - \ldots$$

$$(r_t(x_t,a_t) - b_t(x_t)) + \frac{1}{2\sigma_\theta^2}\left(2(a_t - \mu_{\theta_t}(x_t))^T \Delta\mu - \|\Delta\mu\|^2\right)(r_t(x_t,a_t) - b_t(x_t)) - \dots$$

$$(r_t(x_t,a_t) - b_t(x_t)) + \frac{\cancel{2}(a_t - \mu_{\theta_t}(x_t))^T \Delta\mu\,(r_t(x_t,a_t) - b_t(x_t))}{\cancel{2}\sigma_\theta^2} - \frac{(r_t(x_t,a_t) - b_t(x_t))\|\Delta\mu\|^2}{2\sigma_\theta^2} - \dots$$

constant w.r.t. θ so it doesn't affect argmax

$$\cancel{(r_t(x_t,a_t) - b_t(x_t))} + \Delta\mu\underbrace{\left(\frac{1}{\sigma_\theta^2}(a_t - \mu_{\theta_t}(x_t))^T(r_t(x_t,a_t) - b_t(x_t))\right)}_{g_t} - \cancel{\frac{1}{2\sigma_\theta^2}(r_t(x_t,a_t) - b_t(x_t))\|\Delta\mu\|^2} - \dots$$

Since $\Delta\mu$ is small, it grows smaller than the linear term

$$\langle \Delta\mu, g_t\rangle - \dots = \underbrace{\langle \mu_\theta(x_t) - \mu_{\theta_t}(x_t), g_t\rangle}_{\text{first term}} - \underbrace{\dots}_{\text{KL}}$$

KL Portion

$$\langle \mu_\theta(x_t) - \mu_{\theta_t}(x_t), g_t\rangle - \frac{1}{\eta}\left(\frac{(\mu_\theta - \mu_{\theta_t})^2 + \sigma_\theta^2}{2\sigma_{\theta_t}^2} - \frac{1}{2}\right)$$

since $\theta_t \approx \theta_{t+1}$

$$\langle \mu_\theta(x_t) - \mu_{\theta_t}(x_t), g_t\rangle - \frac{1}{\eta}\left(\frac{(\mu_\theta - \mu_{\theta_t})^2}{2\sigma_{\theta_t}^2} + \frac{\cancel{\sigma_\theta^2}}{2\cancel{\sigma_{\theta_t}^2}} - \frac{1}{2}\right)$$

$$\langle \mu_\theta(x_t) - \mu_{\theta_t}(x_t), g_t\rangle - \frac{1}{\eta}\left(\frac{(\mu_\theta - \mu_{\theta_t})^2}{2\sigma_{\theta_t}^2} + \underbrace{\cancel{\frac{1}{2} - \frac{1}{2}}}^{0}\right)$$

same as $(\mu_\theta - \mu_{\theta_t})^2$; just a vectorized version

$$\langle \mu_\theta(x_t) - \mu_{\theta_t}(x_t), g_t\rangle - \frac{1}{\eta}\left(\frac{(\mu_\theta - \mu_{\theta_t})^2}{2\sigma_{\theta_t}^2}\right) = \langle \mu_\theta(x_t) - \mu_{\theta_t}(x_t), g_t\rangle - \frac{1}{2\eta\sigma_{\theta_t}^2}\|\mu_\theta - \mu_{\theta_t}\|^2$$

some constant unrelated to θ

$$\boxed{\therefore \operatorname*{argmax}_\theta\left(\langle \mu_\theta(x_t) - \mu_{\theta_t}(x_t)\rangle - \frac{1}{2\eta\sigma_{\theta_t}^2}\|\mu_\theta - \mu_{\theta_t}\|^2 + C\right) \cong \operatorname*{argmax}_\theta\left(\langle \mu_\theta(x_t) - \mu_{\theta_t}(x_t), g_t\rangle - \frac{1}{2\eta\sigma_{\theta_t}^2}\|\mu_\theta - \mu_{\theta_t}\|^2\right)}$$

d) PG

$$\theta_{t+1} \leftarrow \theta_t + \eta \nabla_\theta \log \pi_\theta(a_t|x_t)\Big|_{\theta=\theta_t}(r_t(x_t,a_t) - b_t(x_t))$$

$$\theta_{t+1} \leftarrow \underset{\theta}{\operatorname{argmax}} \left\{ \langle \mu_\theta(x_t) - \mu_{\theta_t}(x_t), g_t \rangle - \frac{1}{2\eta}\|\theta - \theta_t\|^2 \right\}$$

$$\theta_t + \eta \nabla_\theta \log \pi_\theta(a_t|x_t)\Big|_{\theta=\theta_t}(r_t(x_t,a_t) - b_t(x_t)) \cong \underset{\theta}{\operatorname{argmax}} \left\{ \langle \mu_\theta(x_t) - \mu_{\theta_t}(x_t), g_t \rangle - \frac{1}{2\eta}\|\theta - \theta_t\|^2 \right\}$$

$$\theta_t + \eta \nabla_\theta \log\left(\frac{\exp\left(-\frac{\|a_t - \mu_\theta(x_t)\|^2}{2\sigma^2}\right)}{(2\pi\sigma^2)^{\frac{d}{2}}} \right)\Bigg|_{\theta=\theta_t}(r_t(x_t,a_t) - b_t(x_t))$$

$$\nabla_\theta \log\left(\frac{1}{(2\pi\sigma^2)^{\frac{d}{2}}} \exp\left(-\frac{\|a_t - \mu_\theta(x_t)\|^2}{2\sigma^2}\right) \right) = \nabla_\theta\left(\ln\left(\frac{1}{(2\pi\sigma^2)^{\frac{d}{2}}}\right) + \cancel{\ln}\left(\cancel{\exp}\left(-\frac{\|a_t - \mu_\theta(x_t)\|^2}{2\sigma^2}\right)\right)\right) = \nabla_\theta\left(\overbrace{\cancel{-\frac{d}{2}\ln(2\pi\sigma^2)}}^{\text{constant}} + -\frac{\|a_t - \mu_\theta(x_t)\|^2}{2\sigma^2} \right)$$

$$= \nabla_\theta\left(-\frac{\|a_t - \mu_\theta(x_t)\|^2}{2\sigma^2}\right) = \nabla_\theta\left(-\frac{1}{2\sigma^2}(a_t - \mu_\theta(x_t))^T(a_t - \mu_\theta(x_t))\right) = -\frac{1}{2\sigma^2}\nabla_\theta\left(\overset{0}{\cancel{\|a_t\|^2}} - 2a_t^T\mu_\theta(x_t) + \|\mu_\theta(x_t)\|^2\right)$$

$$= -\frac{1}{\cancel{2}\sigma^2}\left(-\cancel{2}a_t^T\frac{\partial \mu_\theta(x_t)}{\partial\theta} + \cancel{2}\mu_\theta(x_t)\frac{\partial \mu_\theta}{\partial\theta}\right) = \frac{1}{\sigma^2}\left(a_t^T\frac{\partial \mu_\theta(x_t)}{\partial\theta} - \mu_\theta(x_t)\frac{\partial \mu_\theta(x_t)}{\partial\theta}\right)\Bigg|_{\theta=\theta_t} = \frac{1}{\sigma^2}(a_T - \mu_\theta(x_t))\underbrace{\frac{\partial\mu_\theta(x_t)}{\partial\theta}}_{J_{\theta_t} \text{ for Jacobian}}\Bigg|_{\theta=\theta_t}$$

$$\theta_t + \frac{\eta}{\sigma^2}(a_t - \mu_\theta(x_t))J_{\theta_t}(r_t(x_t,a_t) - b_t) = \theta_t + \eta\underbrace{\frac{1}{\sigma^2}(a_t - \mu_\theta(x_t))(r_t(x_t,a_t) - b_t)}_{g_t}J_{\theta_t}$$

$\theta_{t+1} = \theta_t + \eta g_t J_{\theta_t}$ PG update

. . - - - - - . . Right Side of Term

$$\nabla_\theta\left(\langle \mu_\theta(x_t) - \mu_{\theta_t}(x_t), g_t\rangle - \frac{1}{2\eta}\|\theta - \theta_t\|^2\right) = 0$$

note: $\theta = \theta_{t+1}$

$$\nabla_\theta\left((\mu_\theta(x_t) - \mu_{\theta_t}(x_t))^T g_t - \frac{1}{2\eta}(\theta - \theta_t)^T(\theta - \theta_t)\right) = 0$$

$$\nabla_\theta\left((\theta - \theta_t)^T \nabla_\theta \mu_{\theta_t}(x_t) g_t - \frac{1}{2\eta}(\theta - \theta_t)^T(\theta - \theta_t)\right) = 0 \Rightarrow \nabla_\theta\left((\theta - \theta_t)^T \frac{\partial\mu_\theta(x_t)}{\partial\theta} g_t - \frac{1}{2\eta}(\theta - \theta_t)^T(\theta - \theta_t)\right) = 0 \Rightarrow \nabla_\theta\left((\theta - \theta_t)^T J_{\theta_t} g_t - \frac{1}{2\eta}(\theta - \theta_t)^T(\theta - \theta_t)\right)$$

$$= (1 - 0)^T J_{\theta_t} g_t - \frac{1}{2\eta}\left(\|\theta\|^2 - 2\theta^T\theta_t + \overset{0}{\cancel{\|\theta_t\|^2}}\right)$$

$$= J_{\theta_t} g_t - \frac{1}{\cancel{2}\eta}(\cancel{2}\theta - \cancel{2}\theta_t) = J_{\theta_t} g_t - \frac{1}{\eta}(\theta - \theta_t) = 0$$

$$J_{\theta_t} g_t - \frac{1}{\eta}(\theta_{t+1} - \theta_t) = 0 \rightarrow J_{\theta_t} g_t = \frac{1}{\eta}(\theta_{t+1} - \theta_t) \rightarrow \eta J_{\theta_t} g_t = \theta_{t+1} - \theta_t \rightarrow \boxed{\theta_{t+1} = \theta_t + \eta J_{\theta_t} g_t}$$

The PG update is equivalent to its argmax version since $\underbrace{\theta_t + \eta J_{\theta_t} g_t}_{\text{PG Update}} = \underbrace{\theta_t + \eta J_{\theta_t} g_t}_{\text{equivalence}}$