

Assignment - 03

K. Surya Prakash
EE18BTECH11026

Q2)

$$\nabla_{\theta} J(\theta) = E_{\pi_{\theta}} \left[\sum_{t=0}^{\infty} \nabla_{\theta} \log \pi_{\theta}(a_t | s_t) \cdot \Phi_t \right]$$

$$\Phi_t = \sum_{t'=t}^{\infty} \gamma^{t'-t} r_{t'} - b(s_t)$$

consider,

$$b = \frac{E_{\pi_{\theta}} \left[\nabla_{\theta} \log \pi(a_t | s_t)^2 G_{t:\infty}^{(\pi)} \right]}{E_{\pi_{\theta}} \left[\nabla_{\theta} \log (\pi(a_t | s_t)^2) \right]}$$

$$\text{Var}(\nabla_{\theta} J(\theta)) = E_{\pi_{\theta}} \left[\sum_{t=0}^{\infty} (\nabla_{\theta} \log \pi_{\theta}(a_t | s_t) \Phi_t)^2 \right] - \left[E_{\pi_{\theta}} \left(\sum_{t=0}^{\infty} \nabla_{\theta} \log \pi_{\theta}(a_t | s_t) \Phi_t \right) \right]^2$$

$$\text{Var}(\nabla_{\theta} J(\theta)) = E_{\pi \sim \pi_{\theta}(\tau)} \left[(\nabla_{\theta} \log \pi_{\theta}(a_{\tau} | s_{\tau})) \left[G_{\tau:\infty}^{(\pi)} - b \right] \right]^2$$

$$= \left(E_{\pi \sim \pi_{\theta}(\tau)} \left[\nabla_{\theta} \log \pi_{\theta}(a_{\tau} | s_{\tau}) (G_{\tau:\infty}^{(\pi)} - b) \right] \right)^2$$

\Rightarrow consider derivation w.r.t. b

$$\frac{\partial \text{Var}(\nabla_{\theta} J(\theta))}{\partial b} = -2 E \left[\nabla_{\theta} \log \pi_{\theta}(a_{\tau} | s_{\tau}) (G_{\tau:\infty}^{(\pi)} - b) \right]$$

Optimum occurs at

$$\frac{\partial (\text{Var}(\nabla_{\theta} \log \pi_{\theta}))}{\partial b} = 0$$

$$\Rightarrow -2E \left[\nabla_{\theta} \log \pi_{\theta} (a_{\tau|s_{\tau}})^{\top} (G(\tau)b) \right] = 0$$

$$\Rightarrow \cancel{E \nabla_{\theta} \log \pi_{\theta}}$$

$$\Rightarrow E \left[\nabla_{\theta} \log \pi_{\theta} (a_{\tau|s_{\tau}})^{\top} G(\tau) \right] - b E \left[\nabla_{\theta} \log \pi_{\theta} (a_{\tau|s_{\tau}})^{\top} \right]^2 = 0$$

$$\Rightarrow \left[\frac{E \left[\nabla_{\theta} \log \pi_{\theta} (a_{\tau|s_{\tau}})^{\top} G(\tau) \right]}{E \left(\nabla_{\theta} \log \pi_{\theta} (a_{\tau|s_{\tau}})^{\top} \right)^2} \right]$$

↪