$(s_1, a_1, r_1, \cdots, s_H, a_H, r_H)$ $\quad s_1 \sim d_0,\ a_t \sim \pi_\theta.$

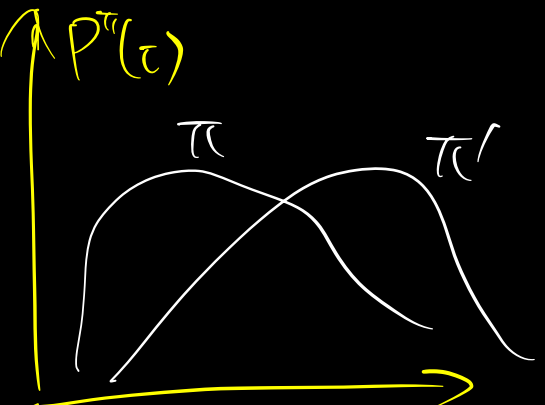
$$\nabla J(\pi) = \nabla \underbrace{\mathbb{E}_{\tau\sim\pi}[R(\tau)]}$$

$$R(\tau) = \sum_{t=1}^{H} \gamma^{t-1} r_t.$$

$$= \sum_\tau \left(\nabla P^\pi(\tau)\right) R(\tau).$$

$\rightarrow \tau \in (S\times A)^H.$

$$= \sum_\tau R(\tau)\, P^\pi(\tau) \nabla \log P^\pi(\tau).$$



$$= \sum_\tau R(\tau) P^\pi(\tau).$$

$$\nabla\left(\log d_0(s_1) + \log \pi(a_1|s_1) + \log P(s_2|s_1,a_1) + \log \pi(a_2|s_2) \cdots\right).$$

$$= \sum_\tau P^\pi(\tau) \cdot R(\tau) \cdot \sum_{t=1}^{H} \nabla \log \pi(a_t|s_t).$$

$$= \mathbb{E}_{\tau\sim\pi}\left[R(\tau) \sum_{t=1}^{H} \nabla \log \pi(a_t|s_t)\right] \leftarrow \text{"REINFORCE"}$$

$$P^\pi(\tau) = P^\pi(s_1, a_1, s_2, a_2, s_3, a_3, \cdots s_H, a_H)$$

$$= d_0(s_1) \cdot \pi(a_1|s_1) \cdot P(s_2|s_1,a_1)\, \pi(a_2|s_2) \cdots$$

REINFORCE

$$f'(x) = f(x) \cdot (\log f)'$$

IS $\left(\prod_t \frac{\pi_{\theta+\Delta\theta}(a_t|s_t)}{\pi_\theta(a_t|s_t)}\right) R(\tau)$

$$\nabla_\theta J(\pi_\theta) = \lim_{\Delta\theta \to 0} \frac{J(\pi_{\theta+\Delta\theta}) - J(\pi_\theta)}{\Delta\theta}$$

"AC" $$\nabla J(\pi) = \frac{1}{1-\gamma} \mathbb{E}_{(s,a)\sim d^\pi} [Q^\pi(s,a) \nabla \log \pi(a|s)] \quad (1)$$

"MC"

$$\rightarrow = \mathbb{E}_{\tau\sim\pi} \left[\sum_{t=1}^{H} \sum_{t'=t}^{H} \left(\gamma^{t'-1} r_{t'}\right) \cdot \nabla \log \pi(a_t|s_t) \right] \quad (2)$$

Equiv. b/t (1) & (2)

$$d^\pi = \sum_t \gamma^t d_t^\pi$$

$$(1) = \sum_t \gamma^{t-1} \mathbb{E}_{(s,a)\sim d_t^\pi} [Q^\pi(s,a) \nabla \log \pi(a|s)]$$

$$= \sum_t \gamma^{t-1} \mathbb{E}_{\tau\sim\pi} [Q^\pi(s_t,a_t) \nabla \log \pi(a_t|s_t)]$$

$$= \mathbb{E}_{\tau\sim\pi} \left[\sum_{t=1}^{H} \gamma^{t-1} \left(\sum_{t'=t}^{H} \gamma^{t'-t} r_t \right) \nabla \log \pi(a_t|s_t) \right]$$

"Proof 1": $\nabla J(\pi) = \nabla \sum_t \mathbb{E}_\pi[r_t]$.

$= \nabla \sum_t \sum_? P^\pi(?) \cdot r_t(?)$.

$? = (s_1, a_1, s_2, a_2, \dots, s_t, a_t)$

Proof 2. $\sum d_s^\pi$

$\nabla V^\pi(s) = \nabla \sum_a \pi(a|s) Q^\pi(s,a)$

$= \sum_a \left(\pi(a|s) \nabla Q^\pi(s,a) + (\nabla \pi(a|s)) Q^\pi(s,a) \right)$

$= Q^\pi(s,a) \cdot \sum_a \pi(a|s) \nabla \log \pi(a|s)$

$+ \left(\sum_a \pi(a|s) \nabla \left(R(s,a) + \gamma \mathbb{E}_{s' \sim P(\cdot|s,a)} [V^\pi(s')] \right) \right)$

$= \mathbb{E}_{a \sim \pi(\cdot|s)} [Q^\pi(s,a) \nabla \log \pi(a|s)]$

$+ \gamma \mathbb{E}_{s' \sim P(\cdot|s,\pi)} [\nabla V^\pi(s')]$.

$$\nabla J(\pi) = \nabla \mathbb{E}_{s_1 \sim d_0}[V^\pi(s_1)].$$

$$= \mathbb{E}_{s_1 \sim d_0, a \sim \pi}[Q^\pi \nabla \log \pi] + \gamma \mathbb{E}_{\substack{s_2 \sim d_2^\pi, a \sim \pi \\ s' \sim P(\cdot|s,a)}}[\nabla V^\pi(s')]$$

$$\| $$

$$\gamma \mathbb{E}_{s \sim d_2^\pi}[\nabla V^\pi(s)].$$

$$\vdots$$

$$\nabla J(\pi) = \sum_{t=1}^{\infty} \gamma^{t-1} \mathbb{E}_{s \sim d_t^\pi, a \sim \pi}[Q^\pi \nabla \log \pi]$$

$$\|$$

$$\frac{1}{1-\gamma} \mathbb{E}_{(s,a) \sim d^{\pi}} \left[\theta^{\pi} \nabla \log \pi \right]$$

$$\pi_\theta(a|s) \propto e^{\phi(s,a)^T \theta}.$$

$$= \frac{e^{\phi(s,a)^T \theta}}{\sum_{a'} e^{\phi(s,a')^T \theta}}$$

$$\nabla_\theta \log \pi_\theta(a|s) = \nabla_\theta \left(\log e^{\phi(s,a)^T \theta} - \log \sum_{a'} e^{\phi(s,a')^T \theta} \right).$$

$$= \nabla_\theta (\phi(s,a)^T \theta) - \nabla_\theta \left(\log \left(\sum_{a'} (\cdot) \right) \right)$$

$$= \phi(s,a) - \frac{\sum_{a'} e^{\phi(s,a')^T \theta} \cdot \phi(s,a')}{\sum_{a'} e^{\phi(s,a')^T \theta}}.$$

$$(a'X)' = a$$

$$= \phi(s,a) - \mathbb{E}_{a' \sim \pi_\theta} [\phi(s,a')].$$

$$\mathbb{E}_{a \sim \pi_\theta(\cdot|s)} [\phi(s,a)] -$$

$$\mathop{\mathbb{E}}_{a\sim\pi(\cdot|s)}\left[\nabla \log \pi(a|s)\right]$$

$$= \sum_a \pi(a|s) \cdot \frac{\nabla \pi(a|s)}{\pi(a|s)}$$

$$= \nabla \underbrace{\sum_a \pi(a|s)}_{} = \nabla \cdot 1 = 0.$$