Improvement: $\nabla_\theta V^{\pi_\theta} = \nabla\left(\sum_{h=1}^{H} \gamma^{h-1} \sum_{\tau_{1:h}} r_h \cdot P^{\pi}(\tau_{1:h})\right)$

$= \sum_{h'=1}^{H} \gamma^{h'-1} \sum_{\tau_{1:h'}} r_{h'} \, P^{\pi}(\tau_{1:h'}) \left(\sum_{h=1}^{h'} \nabla \log(\pi(a_h \mid s_h))\right)$

$= \sum_{h'=1}^{H} \gamma^{h'-1} \sum_{\tau_{1:H}} r_{h'} \, P^{\pi}(\tau_{1:H}) \cdot (\quad \cdots \quad)$

$= \sum_{\tau_{1:H}} P^{\pi}(\tau_{1:H}) \sum_{h'=1}^{H} \gamma^{h'-1} r_{h'} \sum_{h=1}^{h'} \nabla \log \pi(a_h \mid s_h$

$= \mathbb{E}\left[\sum_{h'=1}^{H} \gamma^{h'-1} r_{h'} \sum_{h=1}^{h'} \nabla \log \pi(a_h \mid s_h)\right].$

$\nabla \log \pi(a_1 \mid s_1)$: $h' = 1, 2, \cdots, H$

$\nabla \log \pi(a_2 \mid s_2)$: $h' = 2, 3, \cdots, H$.

$$= \mathbb{E}\left[\sum_{h=1}^{H} \nabla \log \pi(a_h|s_h) \cdot \sum_{h'=h}^{H} \gamma^{h'-1} r_{h'} \right].$$

$$= \sum_{h=1}^{H} \mathbb{E}\left[\nabla \log \pi(a_h|s_h) \cdot \sum_{h'=h}^{H} \gamma^{h'-1} r_{h'} \right].$$

$$= \sum_{h=1}^{H} \mathbb{E}\left[\nabla \log \pi(a_h|s_h) \cdot \left(\gamma^{h-1} Q^{\pi}(s_h, a_h) \right) \right].$$

$$= \frac{1}{1-\gamma} \mathbb{E}_{s,a \sim \eta^{\pi}}\left[\nabla \log \pi(a|s) \cdot \left(Q^{\pi}(s,a) - f(s) \right) \right].$$

s is deterministic.

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

$$= \sum_{a \in A} \pi(a|s) \cdot \nabla \log \pi(a|s)$$

$$= \sum_{a \in A} \pi(a|s) \cdot \frac{\nabla \pi(a|s)}{\pi(a|s)}$$

$$= \nabla \sum_{a \in A} \pi(a|s) = \nabla 1 = 0.$$