

$$\begin{aligned}
 \text{Improvement: } \nabla_{\theta} v^{\pi_{\theta}} &= \nabla \left(\sum_{h=1}^H \gamma^{h-1} \sum_{\tau_{1:h}} r_h \cdot P^{\pi}(\tau_{1:h}) \right) \\
 &= \sum_{h'=1}^H \gamma^{h'-1} \sum_{\tau_{1:h'}} r_{h'} P^{\pi}(\tau_{1:h'}) \left(\sum_{h=1}^{h'} \nabla \log(\pi(a_h | s_h)) \right) \\
 &= \sum_{h'=1}^H \gamma^{h'-1} \sum_{\tau_{1:h'}} r_{h'} P^{\pi}(\tau_{1:h'}) \left(\begin{array}{c} \downarrow \\ \dots \end{array} \right)
 \end{aligned}$$

$$\begin{aligned}
 &= \sum_{\tau_{1:H}} P^{\pi}(\tau_{1:H}) \sum_{h'=1}^H \gamma^{h'-1} r_{h'} \sum_{h=1}^{h'} \nabla \log \pi(a_h | s_h) \\
 &= \mathbb{E} \left[\sum_{h'=1}^H \gamma^{h'-1} r_{h'} \sum_{h=1}^{h'} \nabla \log \pi(a_h | s_h) \right].
 \end{aligned}$$

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

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

$$\begin{aligned}
&= \mathbb{E} \left[\sum_{h=1}^H \nabla \log \pi(a_h | s_h) \cdot \underbrace{\sum_{h'=h}^H \gamma^{h'-1} r_{h'}} \right] \\
&= \sum_{h=1}^H \mathbb{E} \left[\nabla \log \pi(a_h | s_h) \cdot \sum_{h'=h}^H \gamma^{h'-1} r_{h'} \right] \\
&= \sum_{h=1}^H \mathbb{E} \left[\nabla \log \pi(a_h | s_h) \cdot \left(\gamma^{h-1} Q^{\pi}(s_h, a_h) \right) \right] \\
&= \frac{1}{1-\gamma} \mathbb{E}_{s, a \sim \eta^{\pi}} \left[\nabla \log \pi(a | s) \cdot \left(Q^{\pi}(s, a) - f(s) \right) \right].
\end{aligned}$$

s is deterministic.

$$\begin{aligned}
&\mathbb{E}_{a \sim \pi} \left[\nabla \log \pi(a | s) \right] \\
&= \sum_{a \in A} \pi(a | s) \cdot \nabla \log \pi(a | s) \\
&= \sum_{a \in A} \pi(a | s) \cdot \frac{\nabla \pi(a | s)}{\pi(a | s)} \\
&= \nabla \sum_{a \in A} \pi(a | s) = 0.
\end{aligned}$$