$$V^* = \mathcal{T} V^*, \qquad Q^* = \mathcal{T} Q^*.$$

VI: $\forall k,\ f_k \leftarrow \mathcal{T} f_{k-1}.$

$$\|f_k - Q^*\|_\infty \le \gamma^k \frac{R_{max}}{1-\gamma}.$$

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

$$=: \left(\mathcal{T}^\pi V^\pi\right)(s).$$

$\forall k,\ f_k \leftarrow \mathcal{T}^\pi f_{k-1}.$ $\quad \|f_k - V^\pi\|_\infty \le \gamma^k \frac{R_{max}}{1-\gamma}$

$$Q^\pi(s,a) = \mathbb{E}\left[\sum_{t=1}^{\infty} \gamma^{t-1} r_t \,\middle|\, s_1 = s,\ a_1 = a,\ a_{2:\infty} \sim \pi\right].$$

$$Q^* = Q^{\pi^*}$$

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

$$= \mathcal{T}^\pi Q^\pi.$$

$$V^*,\ Q^*,\quad V^\pi,\ Q^\pi$$

$$V^{\pi}(s) = \underline{\underline{Q}}^{\pi}(s, \pi).$$

$$V^{*}(s) = \max_{a} Q^{*}(s,a)$$

$$\pi^{*}(s) = \arg\max_{a} Q^{*}(s,a).$$

$$M = (S, A, P, R, \gamma) \quad \mathbb{E}\left[\sum_{t=1}^{\infty} \gamma^{t-1} r_t \,\middle|\, \pi\right].$$

$$M = (S, A, P, R, H) \quad \mathbb{E}\left[\sum_{t=1}^{H} r_t \,\middle|\, \pi\right].$$

$$V_t^{\pi}(s) = R(s,\pi) + \mathbb{E}_{s' \sim P(\cdot|s,a)}\left[V_{t+1}^{\pi}(s')\right].$$

$$V_{H+1}^{\pi}(s) \equiv 0.$$

$$V_t^{*}(s) = \max_{a} \left[R(s,a) + \ \cdots \ V_{t+1}^{*}(s') \right]$$

① $Q^* = \mathcal{T} Q^*$. $\quad \forall k,\ f_k \leftarrow \mathcal{T} f_{k-1}$.

$$\boxed{\|f_k - Q^*\|_\infty \le \gamma^k \|f_0 - Q^*\|_\infty}$$

② $\pi^* = \boxed{\pi_{Q^*}}$. $\quad \pi_f(s) = \arg\max_a f(s,a)$.

Lemma. $\forall f \in \mathbb{R}^{S \times A}$, π_f.

$$\|V^* - V^{\pi_f}\|_\infty \le \frac{2 \cdot \|f - Q^*\|_\infty}{1-\gamma}.$$

if reward is const 1, $\quad V^* \equiv Q^* \equiv V^\pi \equiv Q^\pi \equiv \frac{1}{1-\gamma}$.

$\forall s$, $\quad V^*(s) - V^{\pi_f}(s) \qquad Q^{\pi_f}(s,\pi^*)$.

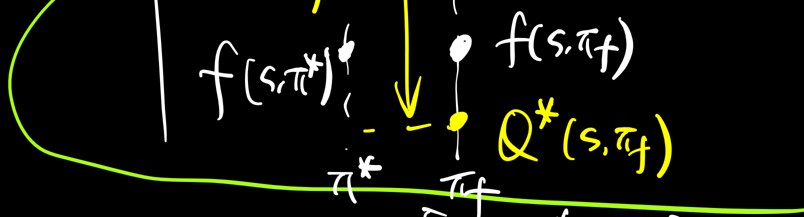
$$= \underbrace{Q^*(s,\pi^*) - Q^*(s,\pi_f)}_{(\mathrm{I})} + \underbrace{Q^*(s,\pi_f) - Q^{\pi_f}(s,\pi_f)}_{(\mathrm{II})}.$$

$$(\mathrm{I}) = Q^*(s,\pi^*) - f(s,\pi^*) + f(s,\pi^*) - Q^*(s,\pi_f)$$

$$\le Q^*(s,\pi^*) - f(s,\pi^*) + f(s,\pi_f) - Q^*(s,\pi_f)$$

$$\le 2 \cdot \|f - Q^*\|_\infty.$$

$Q^*(s,\pi^*)$



$$(\mathrm{II}) = Q^*(s,\pi_f) - Q^{\pi_f}(s,\pi_f).$$

$$= \cancel{R(s,\pi_f)} + \gamma\, \mathbb{E}_{s' \sim P(\cdot|s,\pi_f)}\left[V^*(s')\right]$$

$$- \cancel{R(s,\pi_f)} - \gamma\, \mathbb{E}_{s' \sim P(\cdot|s,\pi_f)}\left[V^{\pi_f}(s')\right].$$

$$= \gamma\, \mathbb{E}_{s' \sim P(\cdot|s,\pi_f)}\left[V^*(s') - V^{\pi_f}(s')\right]$$

$$\leq \gamma\, \|V^* - V^{\pi_f}\|_\infty$$

$$\|V^* - V^{\pi_f}\|_\infty \leq \underset{(\mathrm{I})}{2 \cdot \|f - Q^*\|_\infty} + \underset{(\mathrm{II}).}{\gamma \cdot \|V^* - V^{\pi_f}\|_\infty}$$

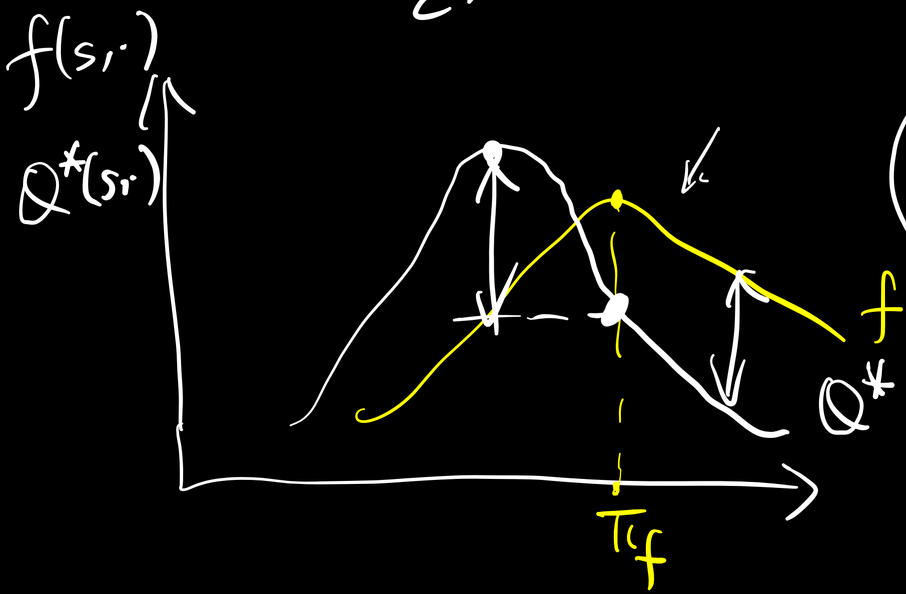
$$\|f_k - Q^*\|_\infty \leq \gamma^k \frac{R_{max}}{1-\gamma}.$$

Output : $\hat{\pi} = \pi_{f_k}.$

$$\|V^* - V^{\hat{\pi}}\|_\infty \leq \frac{2 \cdot \|f_k - Q^*\|_\infty}{1-\gamma}$$

$$\leq \frac{2\, \gamma^k R_{max}}{(1-\gamma)^2}.$$

$$\frac{R_{max}^2}{\varepsilon^3}$$



$$Q^*(s, \pi^*) - Q^*(s, \pi_f)$$