Value Iteration

$(\mathcal{T}f)(s,a) = R(s,a) + \gamma \mathbb{E}_{s' \sim P(\cdot|s,a)}[\max_{a'} f(s',a')]$

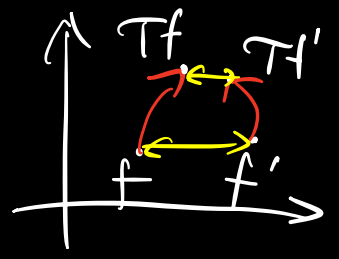
$Q^* = \mathcal{T}Q^*$.

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

$\|f\|_\infty = \max_{s,a} |f(s,a)|$

Contraction: $\forall f, f' \in \mathbb{R}^{S \times A}$,

$\|\mathcal{T}f - \mathcal{T}f'\|_\infty \le \gamma \|f - f'\|_\infty$.



Convergence: $\|f_k - Q^*\|_\infty = \|\mathcal{T}f_{k-1} - \mathcal{T}Q^*\|_\infty$

$\le \gamma \|f_{k-1} - Q^*\| \le \cdots \le \gamma^k \|f_0 - Q^*\|_\infty$

$\forall s,a, \quad |(\mathcal{T}f)(s,a) - (\mathcal{T}f')(s,a)|$

$= |R(s,a) + \gamma \mathbb{E}_{s' \sim P(\cdot|s,a)}[\max_{a'} f(s',a')]$

$- R(s,a) - \gamma \mathbb{E}_{s' \sim P(\cdot|s,a)}[\max_{a'} f'(s',a')]|$.

$= \gamma |\mathbb{E}_{s' \sim P(\cdot|s,a)}[\max_{a'} f(s',a') - \max_{a'} f'(s',a')]|$

$\le \gamma \max_{s'} |\max_{a'} f(s',a') - \max_{a'} f'(s',a')|$.

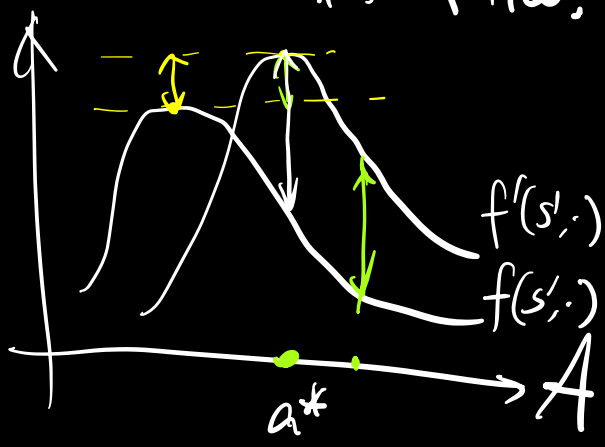
↓?

$\le$

$\leq \gamma \max_{s'} \max_{a'} |f(s',a') - f'(s',a')|$.

$\hookrightarrow \|f - f'\|_\infty$.

Proof of "$\leq$":



wlog: $\max_{a'} f'(s',a') \geq \max_{a'} f(s',a')$.

$a^* = \operatorname{argmax}_{a \in A} f'(s',a)$.

$\max_{a'} f'(s',\cdot) - \max_{a'} f(s',\cdot)$

$\leq f'(s',a^*) - f(s',a^*) \leq \max_{a'} |f'(s',a') - f(s',a')|$

$Q^* = \mathcal{T} Q^*$. $\quad (\mathcal{T}f)(s,a) = R(s,a) + \cdots$

$V^*(s) = \max_{a \in A} \left(R(s,a) + \gamma \mathbb{E}_{s' \sim P(\cdot|s,a)}[V^*(s')]\right)$.

define: $\forall f \in \mathbb{R}^S$.

$(\mathcal{T}f)(s) = \max_a (\cdots \cdots f(s'))$.

$V^* = \mathcal{T} V^*$.

$\mathbb{E}_{a \sim \pi(\cdot|s)}[R(s,a)]$.

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

$\forall f \in \mathbb{R}^S,\ (\mathcal{T}^\pi f)(s) = R(s,\pi) + \gamma \mathbb{E}_{s' \sim P(\cdot|s,\pi)}[f(s')]$.

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

$(\mathcal{T}^\pi f)(s,a) = R(s,a) + \gamma \mathbb{E}_{s' \sim P(\cdot|s,a)}[f(s',\pi)]$.

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

$\mathbb{E}_{a' \sim \pi(\cdot|s')}[f(s',a')]$.

$V^\pi = R^\pi + \gamma P^\pi V^\pi$.

$\forall f \in \mathbb{R}^S,\ \mathcal{T}^\pi f = R^\pi + \gamma P^\pi f$.

$\mathcal{T}^\pi f - \mathcal{T}^\pi f' \neq \mathcal{T}^\pi(f - f')$

Alt. proof to Convergence of VI:

$\forall f \in \mathbb{R}^S,\ (\mathcal{T}f)(s) = \max_a \left(R(s,a) + \gamma \mathbb{E}_{s' \sim P(\cdot|s,a)}[f(s')]\right)$

$V^* = \mathcal{T} V^*$.

$\vec{0} = f_0 \in \mathbb{R}^S$. $f_k \in \mathcal{T} f_{k-1}$.

---

① $V_H^\pi(s) = \mathbb{E}_\pi\left[\sum_{t=1}^{H} \gamma^{t-1} r_t \mid s_1 = s\right]$.

② $V_H^*(s) = \max_{\pi: S\times[H]\to A} V_H^\pi(s)$. ($[H] = \{1,2,\cdots,H\}$)

② $V_H^{\pi^*} \le V_H^*$.

$f_0 = 0 \iff V_0^*$.

$\boxed{f_k = V_k^*}$.

$f_1 = \mathcal{T} f_0 = \max_a (R(s,a)) \iff V_1^*$.

$f_2 = \mathcal{T} f_1 = \max_a \left(R(s,a) + \gamma \mathbb{E}_{s'\sim P(\cdot|s,a)}[f_1(s')] \right)$

$= \max_a \left(R(s,a) + \gamma \mathbb{E}_{s'\sim P(\cdot|s,a)}[\max_{a'} R(s',a')] \right)$.

$= V_2^*$.

$\vdots$

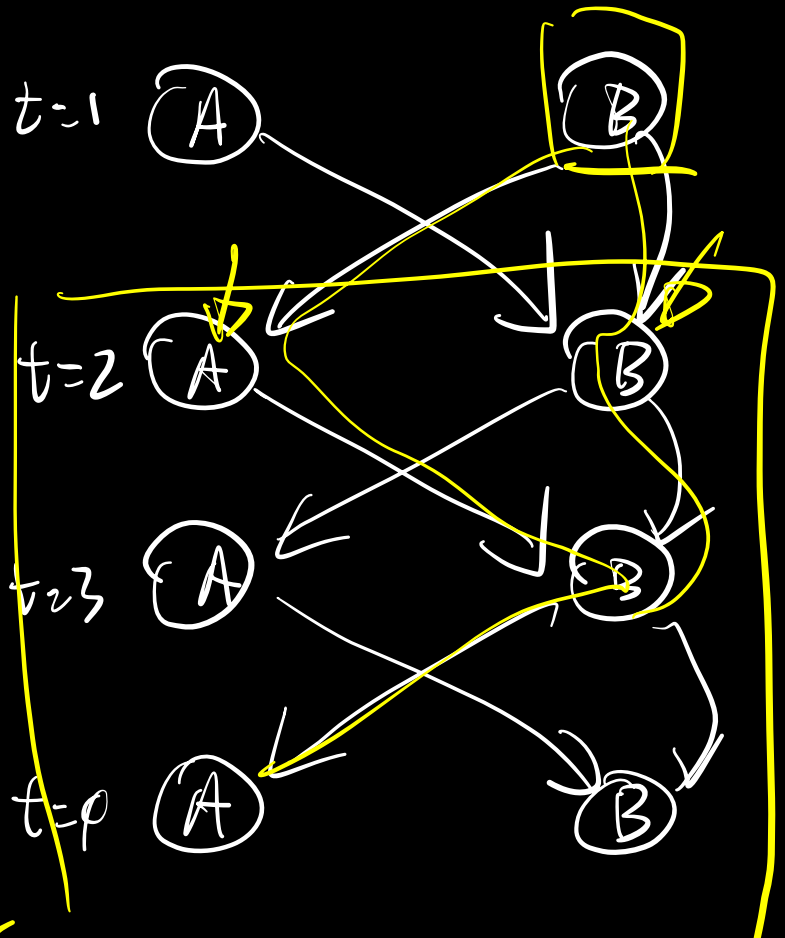
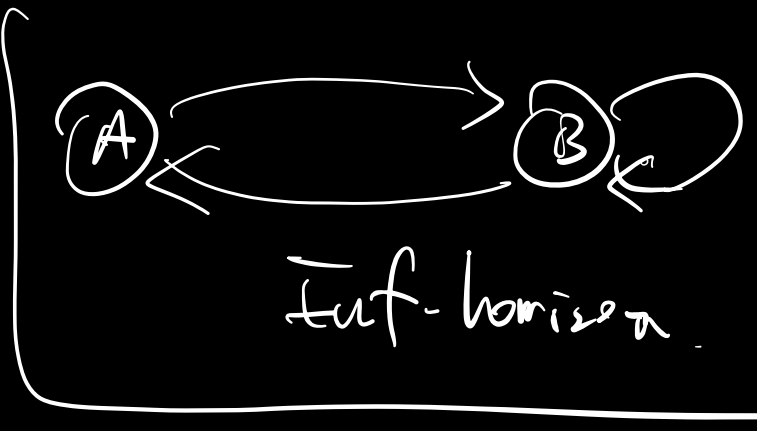
$\|f_k - V^*\|_\infty \le O(\gamma^k)$.

$V^* - \square \le V_k^* \le V^*$

① $V_k^* \le V^*$. $\left(V_k^*(s) \le V^*(s), \forall s \right)$.

② $V^*(s) = \mathbb{E}_{\pi^*}\left[\sum_{t=1}^{\infty} \gamma^{t-1} r_t \,\middle|\, s_1 = s \right]$.

$$= \mathbb{E}_{\pi^*}\left[\sum_{t=1}^{k} \gamma^{t-1} r_t \,\middle|\, s_1 = s \right] + \mathbb{E}_{\pi^*}\left[\sum_{t=k+1}^{\infty} \gamma^{t-1} r_t \right]$$

$$\le V_k^{\pi^*}(s) + \frac{\gamma^k R_{max}}{1-\gamma}$$

$$\le V_k^*(s) + \frac{\gamma^k R_{max}}{1-\gamma}$$