$V^{\pi^*} = V^*$:

$$\left\{ \begin{array}{l} \underline{\underline{V^*}}(s) = \max_a \Big(\underbrace{R(s,a) + \gamma \mathbb{E}_{s' \sim P(\cdot|s,a)} [\underline{V^*}(s')]}_{Q^*(s,a)} \Big). \\ \boxed{\underline{\underline{Q^*}}(s,a) = R(s,a) + \gamma \mathbb{E}_{s' \sim P(\cdot|s,a)} \left[\max_{a'} Q^*(s',a') \right].} \end{array} \right.$$

Define: $\mathcal{T} : \mathbb{R}^{S \times A} \to \underline{\underline{\mathbb{R}^{S \times A}}}$

$\forall f \in \mathbb{R}^{S \times A}$, $\underline{(\mathcal{T} f)} \underline{(s,a)}$

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

Bellman opt. eq: $Q^* = \mathcal{T} Q^*$

Value Iteration: Init $f_0 \in \mathbb{R}^{S \times A}$. $(= \vec{0})$.

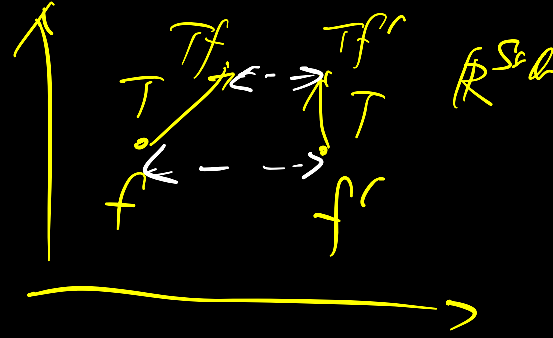
for $k = 1, 2, 3, \ldots$, $f_k \leftarrow \underline{\underline{\mathcal{T}}} f_{k-1}$

Lemma: $\mathcal{T}$ is γ-contraction under $\|\cdot\|_\infty$.

i.e. $\forall f, f' \in \mathbb{R}^{S \times A}$

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

Given Lemma, $Q_1^*, Q_2^*, \dots$,



$$\|f_k - Q_i^*\|_\infty$$

$$= \|Tf_{k-1} - TQ_i^*\|_\infty$$

$$\le \gamma \cdot \|f_{k-1} - Q^*\|_\infty \le \dots \le \gamma^k \underbrace{\|f_0 - Q^*\|_\infty} \le \gamma^k \frac{R_{max}}{1-\gamma}$$

$Q_1^* \ne Q_2^*$, $\|Q_1^* - Q_2^*\| = \varepsilon > 0$.

$$\|f_k - Q_1^*\|_\infty < \frac{\varepsilon}{2}$$

Proof of contraction: $\forall f, f'$. $\|Tf - Tf'\|_\infty \le \gamma \|f - f'\|_\infty$

Proof: $\forall (s,a)$

$$|Tf(s,a) - Tf'(s,a)|$$

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

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

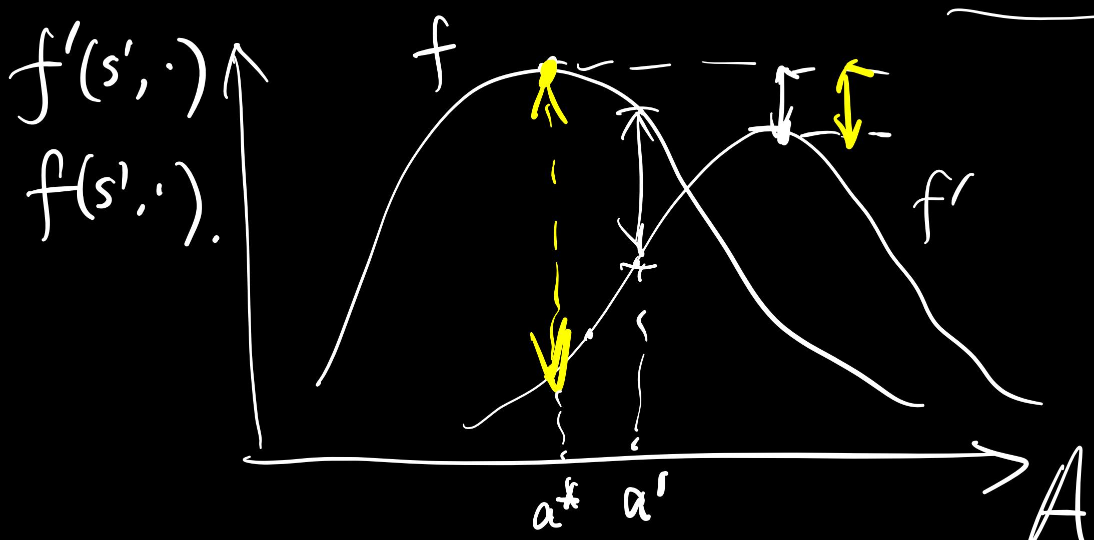
$$\leq \gamma \underline{\underline{\max_{s'}}} \left| \max_{a'} f(s',a') - \max_{a'} f'(s',a') \right| .$$

$$\boxed{\overset{?}{\leq}} \ \gamma \cdot \|f - f'\|_\infty = \gamma \max_{s', a'} |f(s',a') - f'(s',a')|$$

$$= \gamma \cdot \underline{\underline{\max_{s'}}} \left(\max_{a'} |f(s',a') - f'(s',a')| \right).$$

Suffices to show: $\forall s'$.

$$\left| \max_{a'} f(s',a') - \max_{a'} f'(s',a') \right| \overset{(*)}{\leq} \underline{\max_{a'} |f(s',a') - f'(s',a')|}$$



Proof of $(*)$: w.l.o.g. $\max_{a'} f(s',a') \geq \max_{a'} f'(s',a')$

also define $a^* = \arg\max_{a'} f(s',a')$

$$\left| \underline{\max_{a'} f(s',a')} - \max_{a'} f'(s',a') \right|$$
$$\downarrow$$

$$= f(s', a^*) - \underbrace{\max_{a'} f'(s', a') \triangleq}_{\geq}$$

$$\leq f(s', a^*) - f'(s', a^*).$$

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

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

Alt. proof (for V^*):

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

$$=: (\mathcal{T} \overset{f}{V^*})(s).$$

VI (for V^*): Init $f_0 \in \mathbb{R}^S$.

$\forall K = 1, 2, 3, \ldots \quad f_K \leftarrow \mathcal{T} f_{K-1}.$

Lemma: $\mathcal{T}$ is a γ-contraction under $\|\cdot\|_\infty$.

$$f_0 = \vec{0}.$$

$$f_1 = \mathcal{T} f_0 = \max_a R(s,a) + \gamma \mathbb{E}_{\ldots}[\underline{f_0(s')}])$$

$$= \underline{\max_a R(s,a)}.$$

$$f_2 = \max_a \left(R(s,a) + \gamma \mathbb{E}_{s' \sim P(\cdot|s,a)} \left[\max_{a'} R(s',a') \right] \right)$$

Claim: f_K is optimal for:

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

i.e.. $f_K(s) = \max_{\pi:\ \text{non-stationary}} V^{\pi,K}(s)$