Remarks about VI

1. ex.1. $Q^* = \mathcal{T} Q^*$. and simarly V^π, Q^π, V^*.

ex.2. $f_k \leftarrow \mathcal{T} f_{k-1}$. w/ f_0 init arbitrarily.

ex.3. $\forall f, f' \in \mathbb{R}^{S\times A}$, $\|\mathcal{T} f - \mathcal{T} f'\|_\infty \le \gamma \cdot \|f - f'\|_\infty$

2. Convergence: $\|f_k - Q^*\|_\infty \propto \gamma^k$

How good is policy $\pi_{f_k}(s) := \arg\max_a f_k(s,a)$

$$\|V^* - V^{\pi_f}\|_\infty \le \frac{2 \cdot \|f - Q^*\|_\infty}{1-\gamma}, \quad \forall f \in \mathbb{R}^{S\times A}.$$

$$f(s,a) = \mathbb{I}[\pi^*(s) = a].$$

3. Bellman eq. VI alg.

Bellman eq.	VI alg.
$Q^* = \mathcal{T} Q^*$.	$f_k \leftarrow \mathcal{T} f_{k-1}, \forall f_0 \in \mathbb{R}^{S\times A}$
$V^* = \mathcal{T} V^*$	$f_k \leftarrow \mathcal{T} f_{k-1}, \forall f_0 \in \mathbb{R}^S$.
$V^\pi = \mathcal{T}^\pi V^\pi$	$f_k \leftarrow \mathcal{T}^\pi f_{k-1}, \forall f_0 \in \mathbb{R}^S$.

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

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

$\mathcal{T}^{\pi}$ is also γ- contraction under l_∞.

$Q^{\pi} = \mathcal{T}^{\pi} Q^{\pi}$. $\qquad f_k \leftarrow \mathcal{T}^{\pi} f_{k-1}, \ \forall f_0 \in \mathbb{R}^{S \times A}$.

Alternative proof for VI.

$V^* = \mathcal{T} V^*$. $\qquad f_0 \in \mathbb{R}^S$. $\qquad f_k \leftarrow \mathcal{T} f_{k-1}$.

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

$f_0 = \vec{0}$.

$$f_1 = \mathcal{T} \vec{0} \quad \Rightarrow f_1(s) = \max_a R(s,a)$$

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

Claim: $\underbrace{f_k(s)}_{= \mathcal{T}^k f_0} = \max \underbrace{\mathbb{E}\left[\sum_{t=1}^{k} \gamma^{t-1} r_t \,\middle|\, \pi, s_1 = s \right]}_{\Downarrow \, V^{\pi,k}(s)}$

π: non-stationary policy

Result: $\| f_K - V^* \|_\infty \le \frac{\gamma^K R_{max}}{1-\gamma}$. $\measuredangle$

Proof: (1) $f_K \le V^*$.

(2) $f_K \ge V^* - (\quad)$

$$f_K(s) = \max_{\text{non-stat } \pi} V^{\pi,K}(s).$$

$$\ge V^{\pi^*,K}(s).$$

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

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

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

$$\ge V^*(s) - \mathbb{E}\left[\sum_{t=K+1}^{\infty} \gamma^{t-1} R_{max} \,\middle|\, \dots \right].$$

$$= V^*(s) - \gamma^k \cdot \underbrace{\left(\frac{R_{max}}{1-\gamma}\right)}.$$