$Q^* = \mathcal{T} Q^*$. $\forall f \in \mathbb{R}^{S\times A}$, $(\mathcal{T} f)(s,a) = R(s,a) + \gamma \mathbb{E}_{s'\sim P(s,a)}\left[\max_{a'} f(s',a')\right]$.

$\boxed{f(s', \arg\max_{a'} f(s',a'))}$

Alg: repeatedly apply $\mathcal{T}$ to arbitrary $\boxed{f_0}$.

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

Type 1: Q^π, Q^*

Type 2: "$\forall f \in \mathbb{R}^{S\times A}$"

$\sum_{t\ge 1}^{\infty} \gamma^{t-1} r_t \in [0, \frac{R_{max}}{1-\gamma}]$.

Fact: $\exists f \in \mathbb{R}^{S\times A}$, where f is not Q^π for any π.

Example: MDP M, $R \equiv 0$. $\Rightarrow Q^\pi \equiv 0$, $Q^* \equiv 0$.

Value Iteration for V^*, Q^π, V^π.

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

$\forall f, f' \in \mathbb{R}^S$, $\|\mathcal{T} f - \mathcal{T} f'\|_\infty \le \gamma \|f - f'\|_\infty$

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

$\boxed{\pi(s')}$

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

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

Several ways to calculate $V^\pi(s)$.

- $V^\pi(s) = \underline{\underline{e_s^T}} (I - \gamma P^\pi)^{-1} R^\pi$.
- Init $f_0 \in \mathbb{R}^S$. $\underline{f_k \leftarrow \mathcal{T}^\pi f_{k-1}}$. $\xrightarrow{\text{Converge to}} V^\pi$.
- (Monte-Carlo): start from s, rollout actual trajectories, take average over random returns. $\sum_{t=1}^{\infty} \gamma^{t-1} r_t$.

Alt. proof for convergence of Value Iter.

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

$f_0 = \vec{0}$.

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

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

Generally: $f_k(s) = \max_{\text{non-stationary } \pi} \underbrace{\mathbb{E}\left[\sum_{t=1}^{k} \gamma^{t-1} r_t \mid \pi \right]}_{\boxed{V^{\pi,k}(s)}}$.

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

Proof sketch: $\underline{f_k \le V^*}$.

$$f_k(s) = \max_{\text{non-stat } \pi} V^{\pi,k}(s) \ge V^{\pi^*,k}(s).$$

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

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

$$- \underbrace{\mathbb{E}\left[\sum_{t=k+1}^{\infty} \gamma^{t-1} \underline{\underline{r_t}} \,\middle|\, \pi^*, s_1 = s\right]}_{\in [0, R_{max}].}$$

$$\geq V^*(s) - \underline{\sum_{t=k+1}^{\infty} \gamma^{t-1} R_{max}}$$

$$= V^*(s) - \gamma^k \underbrace{\sum_{t=k+1}^{\infty} \gamma^{t-k-1} R_{max}}_{= \frac{R_{max}}{1-\gamma}.}$$

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

Together: $\forall s$, $V^*(s) - \frac{\gamma^k R_{max}}{1-\gamma} \leq f_k(s) \leq V^*(s)$

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

---

if $f = Q^*$, $\pi_f(s) = \arg\max_{a \in A} f(s,a)$. $\rightarrow$ optimal.

In value Iter, we get $f_k \approx Q^*$.

output π_{f_k}. Q: any guarantee for the optimality of π_{f_k}?

A: $\forall \underline{f \in \mathbb{R}^{S \times A}}$. [Singh & Yee '94].

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

Q^{π_f} ? f