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

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

---

Define: Bellman optimality operator $\mathcal{T}$.

$\mathcal{T}: \mathbb{R}^{S\times A} \to \mathbb{R}^{S\times A}$. $\forall f \in \mathbb{R}^{S\times A}$

$(\mathcal{T}f) \in \mathbb{R}^{S\times A}$

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

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

Abuse notation & define

$\mathcal{T}: \mathbb{R}^S \to \mathbb{R}^S$. $\forall f \in \mathbb{R}^S$.

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

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

init. $f_0 \in \mathbb{R}^S$.

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

## Value Iteration

- Init $f_0 \in \mathbb{R}^{S \times A}$. (e.g.: $f_0 = \vec{0}$).
- For $k = 0, 1, 2, \ldots$, $f_{k+1} \leftarrow \mathcal{T} f_k$.

---

Convergence: "$\mathcal{T}$ is a $\gamma$-contraction under $\ell_\infty$"

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

---

$$
\begin{aligned}
\|f_{k+1} - Q^*\|_\infty &= \|\mathcal{T} f_k - \mathcal{T} Q^*\|_\infty. \\
&\le \gamma \cdot \|f_k - Q^*\|_\infty. \\
&\le \gamma^2 \cdot \|f_{k-1} - Q^*\|_\infty \\
&\le \cdots \le \gamma^{k+1} \underbrace{\|f_0 - Q^*\|_\infty}.
\end{aligned}
$$

$r_t \in [0, R_{max}]$.

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

$Q^*, V^*, Q^\pi, V^\pi \in \left[0, \frac{R_{max}}{1-\gamma}\right]$.

if $f_0 = \vec{0}$, $\|f_0 - Q^*\|_\infty = \|Q^*\|_\infty \le \frac{R_{max}}{1-\gamma}$.

$\therefore \|f_{k+1} - Q^*\|_\infty \le \gamma^{k+1} \cdot \frac{R_{max}}{1-\gamma}$.

$f_0 \in \left[0, \frac{R_{max}}{1-\gamma}\right]^{S \times A}$.

Establish $\gamma$-contraction of $\mathcal{T}$.

i.e., $\forall f, f'$, $\|\mathcal{T}f - \mathcal{T}f'\|_\infty \le \boxed{\gamma \cdot \|f - f'\|_\infty}$.

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

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

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

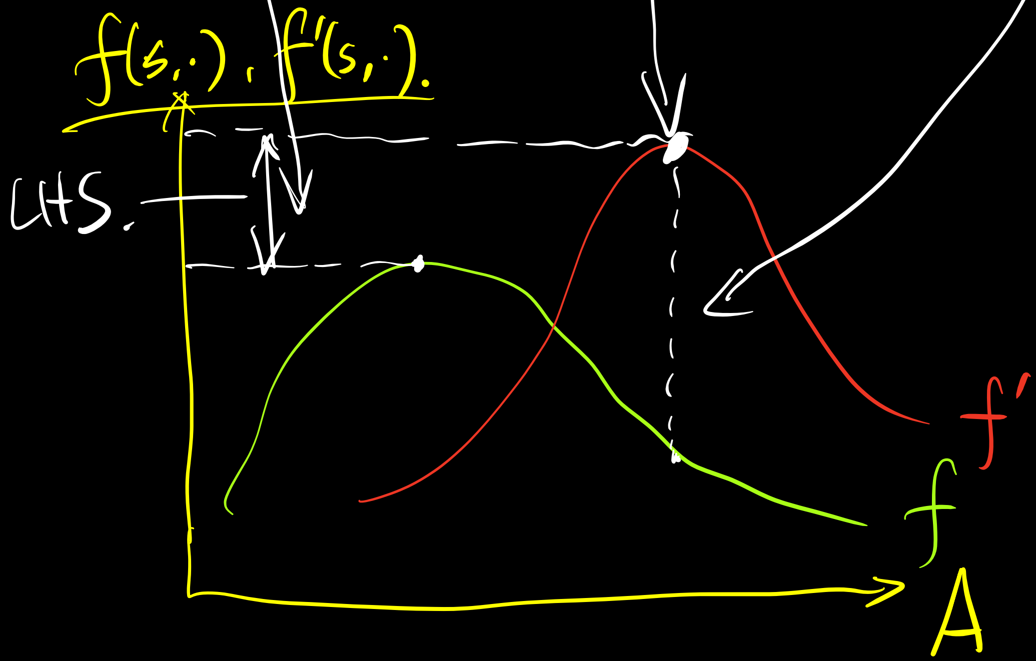
$$\le \gamma \max_{s'} \big| \max_{a'} f(s',a') - \max_{a'} f'(s',a') \big|.$$

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

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

it suffices to show $\forall s$.

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

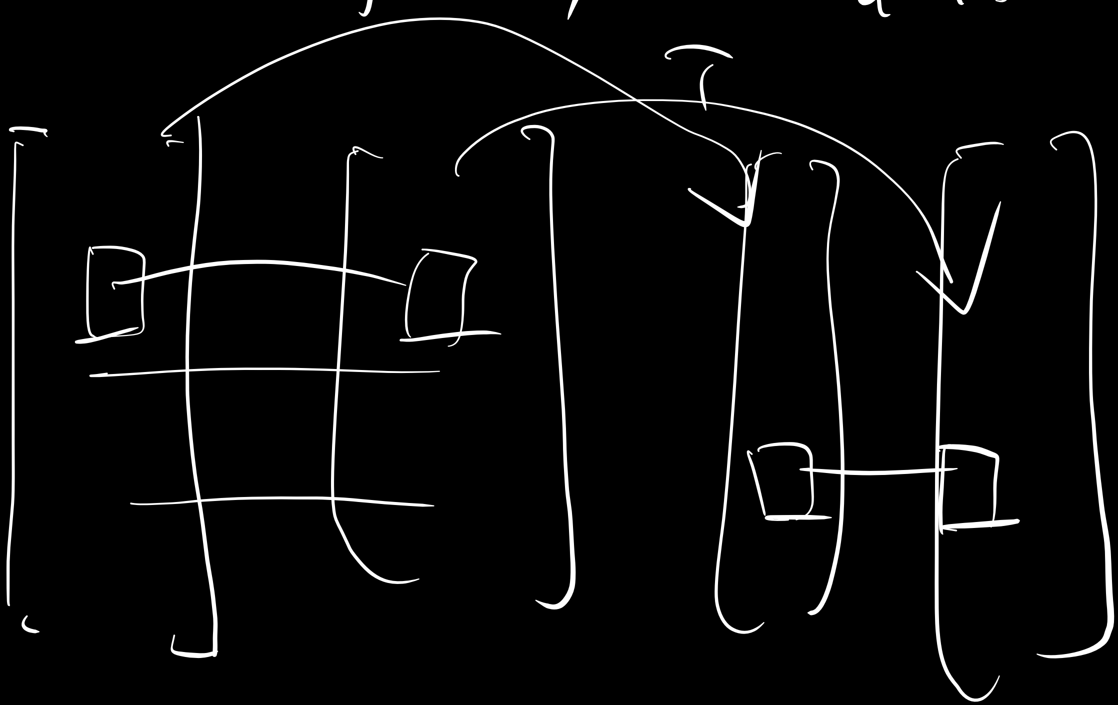



w.l.o.g. $\underline{\max_a f'(s,a)} \geq \max_a f(s,a)$.

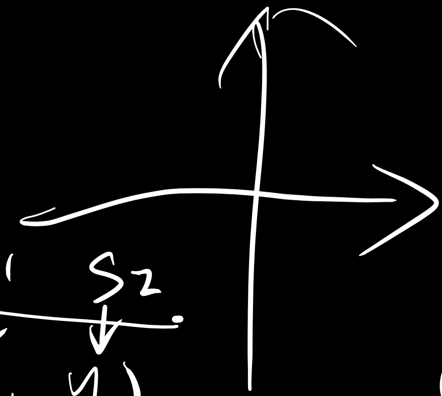
def $a^* := \arg\max_a f'(s,a)$

LHS=

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



$\mathbb{R}^2$



$s_1 \quad s_2$

$(x, y)$

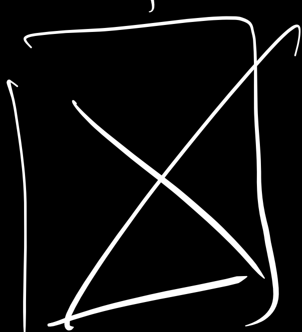
$s_1 \to s_2$

$\mathbb{R}^{\underline{|S\times A|}}$

$\begin{bmatrix} f(s_1) \\ f(s_2) \end{bmatrix}$ $\begin{bmatrix} V^\pi(s_1) \\ V^\pi(s_2) \end{bmatrix}$

A

S



$\begin{bmatrix} \theta \\ \vdots \end{bmatrix} \to \begin{matrix} (s_1, a_1) \\ (s_1, a_2) \\ (s_2, a_1) \\ (s_2, a_2) \\ \vdots \end{matrix}$

$v_1, v_2 \in \mathbb{R}^d$.

$i = 1, 2, \dots, d$.

$\max_i \left| v[i] \right| =: \|v\|_\infty$.