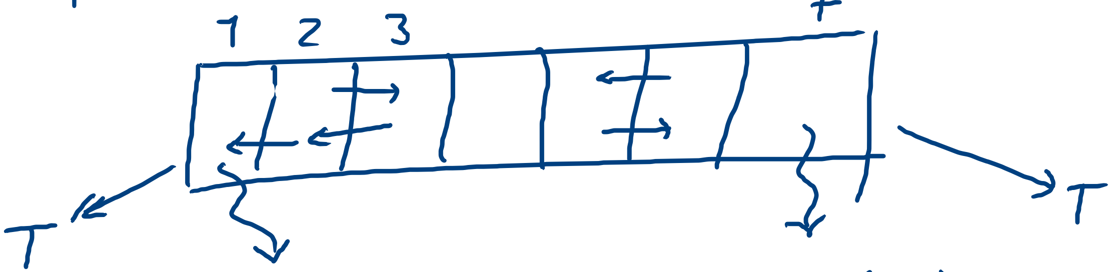
$$V_k^*(s) = \max_a \sum_{s'} p(s'|s,a)\left[R(s) + \gamma V_{k-1}^*(s')\right]$$

$$\overbrace{\vec{V}_k = \mathcal{T}\vec{V}_{k-1}}$$

$R(s)=0$

Example:

$\gamma = 0.9$



$R(1)=1$

$R(7)=10$

$$\vec{V}_0 \quad \begin{array}{|c|c|c|c|c|c|c|} \hline 1 & 0 & 0 & 0 & 0 & 0 & 10 \\ \hline \end{array} \longrightarrow \vec{V}_1$$

$\vec{V}_2$

1	0.9	0	0	0	9	10

2 6

$\max\{0.9 \times 10,\ 0.9 \times 8.1\}$

$\vec{V}_3$

1	0.9	0.9^2	0	8.1	9	10

$\vec{V}_4$

1	0.9	0.9^2	7...	8.1	9	10

2 3

$\vec{V}_5$

1	0.9	6..	7..	8.1	9	10

$\vec{V}_6$

1	5...	6..	7.	8.	9	10

Contraction map

$$\| T\vec{V}_1 - T\vec{V}_2 \|_\infty \leq \alpha \| \vec{V}_1 - \vec{V}_2 \|_\infty$$

$$\alpha < 1$$

$s \longrightarrow$ arbitrary

$$\underline{T\vec{V}_1(s)} = \max_a \sum_{s'} p(s' \mid s, a) \left[R(s) + \gamma \underline{V_1}(s') \right]$$

input

$$T\vec{V}_2(s) = \max_a \sum_{s'} p(s' \mid s, a) \left[R(s) + \gamma V_2(s') \right]$$

$$T\vec{V}_{1(s)} - T\vec{V}_{2(s)} \leq \sum_{s'} p(s' \mid s, a_1) \left[\cancel{R(s)} + \gamma V_1(s') \right] - \sum p(s' \mid s, a_1) \left[\cancel{R(s)} + \gamma V_2(s') \right]$$

$$= \gamma \sum_{s'} P(s'|s,a) \left[V_1(s') - V_2(s') \right]$$

$$\leq \underbrace{\gamma}_{<1} \underbrace{\max_{s'} \left\{ \left| V_1(s') - V_2(s') \right| \right\}}_{\| \vec{V}_1 - \vec{V}_2 \|_\infty} \underbrace{\sum_{s'} P(s'|s,a)}_{1}$$

$\Rightarrow$ T is a contraction mapping

$$\vec{V}_0 \xrightarrow{T} \vec{V}_1 \xrightarrow{T} \vec{V}_2 \quad \cdots\cdots$$

$$\searrow \| \vec{V}_k - \vec{V}^* \|_\infty = \| T\vec{V}_{k-1} - T\vec{V}^* \|_\infty$$

$$\leq$$

$$\leq \gamma \, \| \vec{V}_{k-1} - \vec{V}^{*} \|_\infty$$

$$\leq \gamma^2 \, \| \vec{V}_{k-2} - \vec{V}^{*} \|_\infty$$

$$\vdots$$

$$\leq \gamma^k \, \underbrace{\| \vec{V}_0 - \vec{V}^{*} \|_\infty}_{K}$$

$$\Rightarrow \quad \underbrace{0}_{\to 0} \leq \underbrace{\| \vec{V}_k - \vec{V}^{*} \|_\infty}_{\Rightarrow \; \to 0} \leq \underbrace{K\gamma^k}_{\to 0}$$

$$k \to \infty$$

Policy Improvement

$$\forall s:\ \underline{\underline{\pi_{k+1}(s)}} = \arg\max_a \sum_{s'} P(s'|s,a)\left[\underline{\underline{R(s)}} + \gamma \underline{\underline{V_{\pi_k}(s')}}\right]$$

$$\overbrace{\tau^{\pi_{k+1}} \cdots \tau^{\pi_{k+1}}\left(\tau^{\pi_{k+1}}\left(V_{\pi_k}\right)\right)}$$

Optimal

$$\underbrace{\quad}_{t=1}$$

Does it implement π_{k+1}

$$\mathcal{T}^{\pi_{k+1}} \vec{V}(s) = \sum_{s'} P(s' \mid s, \pi_{k+1}(s)) \left[R(s) + \gamma V(s') \right]$$

why it improves

$$\underbrace{\vec{V}^{\pi_{k+1}}} = \underline{\mathcal{T}^{\pi_{k+1}} \mathcal{T}^{\pi_{k+1}} \cdots \mathcal{T}^{\pi_{k+1}} \vec{V}^{\pi_k}}$$

$$\geq \underbrace{\vec{V}^{\pi_k}}$$

$$\pi_{k+1}(s) = \arg\max_a \sum_{s'} P(s' \mid s, a) \left[R(s) + \gamma V^{\pi_k}(s') \right]$$

1) $$\underbrace{T^{\pi_{k+1}} \vec{V}^{\pi_k}}_{v_1} \geq \underbrace{\vec{V}^{\pi_k}}_{v_2}$$

2) $$\vec{V}_1 \geq \vec{V}_2 \Rightarrow T^{\pi_{k+1}} \vec{V}_1 \geq T^{\pi_{k+1}} \vec{V}_2$$

$$T^{\pi_{k+1}} \; T^{\pi_{k+1}} \vec{V}^{\pi_k} \geq \vec{V}^{\pi_k}$$

$$\vdots$$

$$\underbrace{T^{\pi_{k+1}} \cdots \; T^{\pi_{k+1}} \vec{V}^{\pi_k}}_{\vec{V}^{\pi_{k+1}}} \geq \vec{V}^{\pi_k}$$