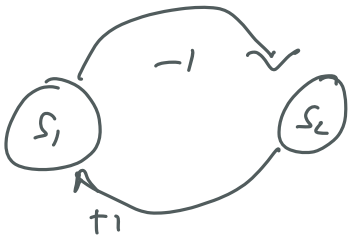
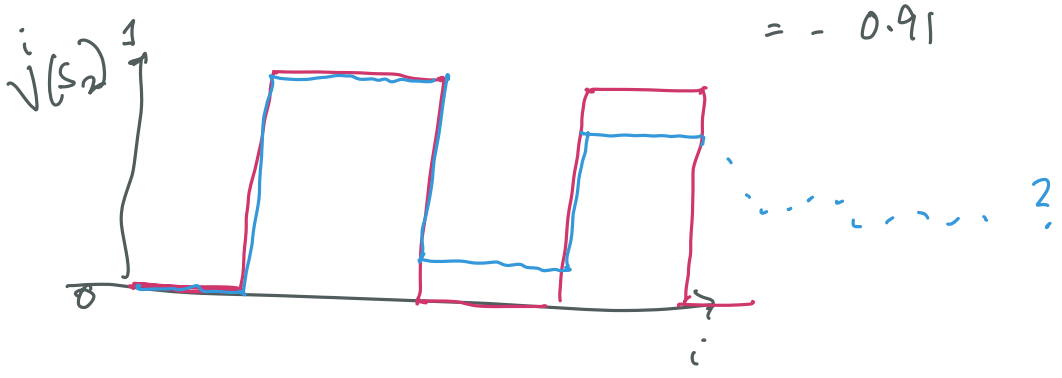


$$V^{i+1}(S_1) = -1 + \gamma V^{i}(S_2)$$
$$V^{i+1}(S_2) = 1 + \gamma V^{i}(S_1)$$

Iter	$V(S_1)$	$V(S_2)$
1	0	0
2	-1	1
3	0	0
4.	-1	1
5.	0	0

$\gamma = 0.9$

1.	0	0
2.	-1	1
3.	-0.1	0.1
4.	$-1 + 0.9 \times 0.1$ $-1 + 0.09$ $= -0.91$	0.91



$$V(S_1) = 1,\ 1-\gamma,\ 1-\gamma+\gamma^2,\ 1-\gamma+\gamma^2-\gamma^3, \ldots$$
$$0 \le \gamma < 1$$

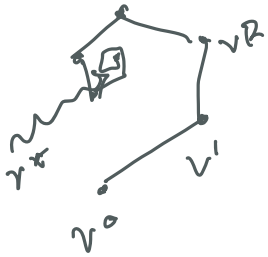
$$\frac{1}{1+\gamma}$$

$$= \frac{1}{1+0.9} \approx 0.5$$

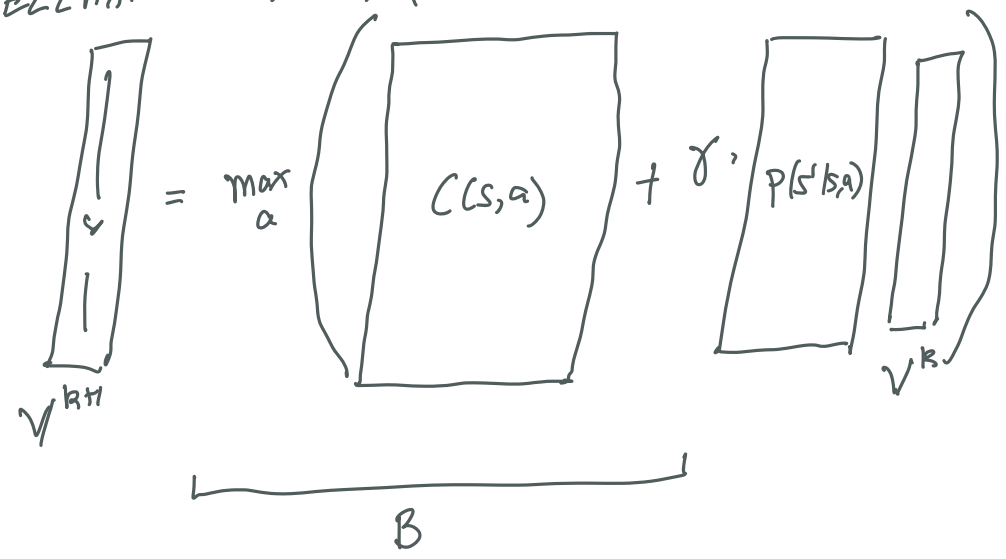
$$V(S_2) = -1 + \gamma - \gamma^2 + \gamma^3 - \cdots = \frac{-1}{1+\gamma}$$

Convergence of Value Iteration

$\gamma < 1$



Bellman Operator



Piece 1 Bellman operator is a contraction.

$$\| BU - BV \|_\infty \leq \gamma \| U - V \|_\infty$$

$$V^{k+1} = BV^k$$

Piece 2

$$\| V^* - V^{k+1} \|_\infty = \| BV^* - BV^k \|_\infty$$

$$\leq \gamma \| V^* - V^k \|_\infty \quad \text{(A)}$$

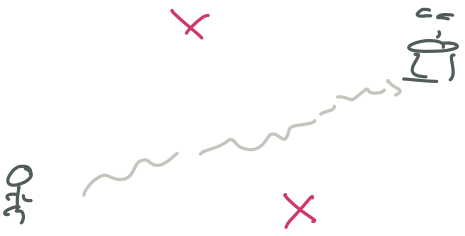
$$\| V^* - V^k \|_\infty \leq \gamma \| V^* - V^{k-1} \|_\infty$$

$$\leq \gamma^2 \| V^* - V^{k-1} \|_\infty$$

$$\vdots$$

$$\leq \gamma^{k+1} \| V^* - V^0 \|_\infty$$

Attributes of a Hard MDP



1. Plan deep into the future
2. Large State Space / Actia Space (continus acto)
3. Cycles
4. Stochastic MDP are harder (Noise)

(R-SAT v/s SAT)

Goal is to find a police $\hat{\pi}$ s.t.

$$\| J(\hat{\pi}) - J(\pi^*) \|_\infty \leq \epsilon$$

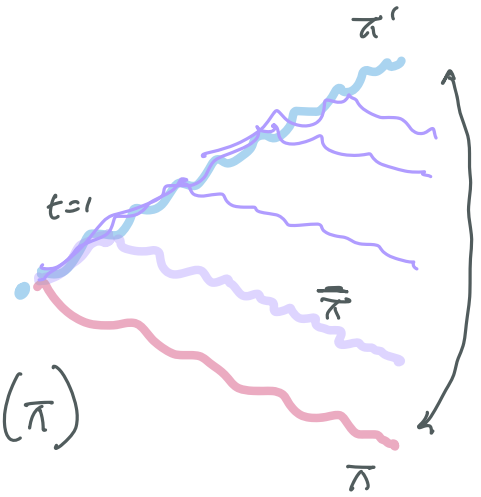
with minimum planning regret

Performance Difference Lemma

π π'

$$J(\pi') - J(\pi)$$

Performance of π' — Perf of π



Telescoping Trick

$$= J(\pi') - J(\bar{\pi}) + J(\bar{\pi}) - J(\pi)$$

$$= V^{\pi'}(s_0) - V^{\pi}(s_0)$$

$$= V^{\pi'}(s_0) - \left(c(s_0,a_0) + \gamma V^{\pi}(s_1)\right) + \left(c(s_0,a_0) + \gamma V^{\pi}(s_1)\right) - V^{\pi}(s_0)$$

$a_0 \sim \pi'$ $\qquad$ $a_0 \sim \pi'$

$$= \left(c(s_0,a_0) + \gamma V^{\pi'}(s_1)\right) - \left(c(s_0,a_0) + \gamma V^{\pi}(s_1)\right) + \cdots$$

$a_0 \sim \pi'$

$$= \gamma\left(V^{\pi'}(s_1) - V^{\pi}(s_1)\right) + Q^{\pi}(s_0, \pi'(s_0)) - V^{\pi}(s_0)$$

$$A^{\pi}(s_0, \pi'(s_0))$$

$$J(\pi') - J(\pi) = \sum_{t=0}^{T-1} \mathop{\mathbb{E}}_{s_t \sim \pi'} \gamma^t A^{\pi}(s_t, \pi'(s_t))$$

$$\pi' = \arg\max_a \left(c(s,a) + \gamma V^{\pi}(s_1) - V^{\pi}(s)\right)$$