1. Recap / Value Iteration

$\Rightarrow$ compute π, $\varepsilon > 0$:
$v^\pi \geq v^* - \varepsilon 1$.

2. Policy Iteration $\boxed{\tilde{O}\left(\frac{SA}{1-\gamma}\right)}$

New!

3. Comp. complexity of planning in finite MDPs?
Lower bound

1. $M = (S, A, P, r, \gamma) \quad 0 \leq \gamma < 1$

π — general / memoryless

Fundamental Theorem

1. π greedy w.r.t. v^*

$\boxed{T_\pi v^* = T v^*}$:

$T_\pi v = r_\pi + \gamma P_\pi v$

$T_a v = r_a + \gamma P_a v$

$(T v)(s) = \max_a (T_a v)(s)$

$\Rightarrow v^\pi = v^*$

2. $T v^* = v^*$

$$\tilde{v}^*(s) = \sup_{\pi \in ML} v^\pi(s)$$

Part 1: $v^* \leftarrow \tilde{v}^*$, prove thm

Part 2: $v^* = \tilde{v}^*$

Part 1: π is greedy w.r.t. $\tilde{v}^*$ $\Rightarrow$ $v^\pi = \tilde{v}^*$. ?

$\boxed{v^\pi \le \tilde{v}^*}$ ✓

$\boxed{\tilde{v}^* \le_? v^\pi}$

$\boxed{\tilde{v}^* \le T\tilde{v}^*}$?

✓

$\forall \pi$ ML: $v^\pi \overset{?}{\le} T\tilde{v}^*$ $/\sup_{\pi ML}$

$\Rightarrow \tilde{v}^* \le T\tilde{v}^*$

π ML

$$v^\pi = T_\pi v^\pi$$

$$v^\pi \le \tilde{v}^* \quad / T_\pi$$

$$T_\pi v^\pi \le T_\pi \tilde{v}^* \le T\tilde{v}^*$$

$$\Rightarrow v^\pi \le T\tilde{v}^* \quad ✓$$

Take π greedy wrt $\tilde{v}^*$:

$$T_\pi \tilde{v}^* = T\tilde{v}^* \ge \tilde{v}^* \quad / T_\pi$$

$$T_\pi^2 \tilde{v}^* \ge T_\pi \tilde{v}^* \ge \tilde{v}^*$$

$a_k \ge a$

$\Rightarrow \lim_{k \to \infty} a_k \ge a$

$\vdots$

$$T_\pi^k \tilde{v}^* \ge \tilde{v}^*$$

Banach's fixed point theorem $\downarrow k \to \infty$

$$v^\pi \ge \tilde{v}^*$$

// Qu. e. d.

part 1 FT

$v^* = \tilde{v}^*$, $\tilde{v}^* \le T\tilde{v}^*$

$v^* \le T v^*$

π greedy w.r.t. v^* :

$$v_\pi = T_\pi v_\pi = T_\pi v^* = T v^*$$

$$\|$$

$$v^*$$

// q.e.d.

Value-iteration

$v_0 \in \mathbb{R}^S$; $v_0 = 0$

$v_{k+1} = T v_k \Rightarrow O(S \times SA)$

Banach's FP Thm $\Rightarrow$

$$\|v_k - v^*\|_\infty \le \gamma^k \|v_0 - v^*\|$$

$v_0 = 0$

$$= \gamma^k \|v^*\|_\infty \le \frac{\gamma^k}{1-\gamma} \le \varepsilon$$

$r_a(s) \in [0,1]$

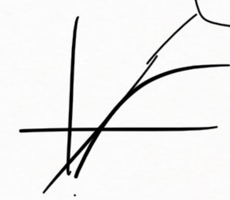
$$\Rightarrow 0 \le \sum \gamma^t \underbrace{r_{A_t}(S_t)}_{1} \le \frac{1}{1-\gamma}$$

$$k \ge \frac{\log\left(\frac{1}{\varepsilon(1-\gamma)}\right)}{\log(1/\gamma)}$$

$$\boxed{k \ge \boxed{\frac{\log\left(\frac{1}{\varepsilon(1-\gamma)}\right)}{1-\gamma}}} \ge \frac{\log\left(\frac{1}{\varepsilon(1-\gamma)}\right)}{\log(1/\gamma)}$$

$H_{\gamma,\varepsilon}$

$$\underbrace{\log x \le x - 1}_{x>0}$$



$k: \ v_k \geq v^* - \varepsilon 1$

$\left(\boxed{\|v_k - v^*\|_\infty \leq \varepsilon} \Rightarrow v_k \geq v^* - \varepsilon 1 \right)$

$\underline{v \geq v^* - 1\varepsilon}$ $\qquad 1(s) \equiv 1 \ \forall s$

Greedify!

$\pi : \ \boxed{T_\pi v = T v}$

Is π good policy?

$\underline{T_\pi v \geq T_\pi (v^* - 1\varepsilon)}$

Dead-end...

$T_\pi (v + a 1) \qquad a \in \mathbb{R}$

$= r_\pi + \gamma P_\pi (v + a 1)$

$= \underbrace{r_\pi + \gamma P_\pi v} + \gamma a \underbrace{P_\pi 1}_{1}$

$= T_\pi v + \gamma a 1$

$T(v + a1) = Tv + \underbrace{\gamma a 1}$

$v \geq v^* - 1\varepsilon \quad / T$

$Tv \geq T(v^* - \varepsilon 1) = Tv^* - \gamma\varepsilon 1$

$\| \qquad = \underline{v^* - \gamma \varepsilon 1}$

$T_\pi v$

$\underline{T_\pi v \geq v^* - \gamma \varepsilon 1}$

$\geq v - (\gamma + 1)\varepsilon 1 \ / T_\pi$

$\varepsilon(\gamma^2 + \gamma + 1) 1$

$v^* \geq v - 1\varepsilon$

$T_\pi^2 v \geq T_\pi v - \underbrace{\gamma(1+\gamma)}\varepsilon 1 = Tv - \dots$

$\geq v^* - \dots - \gamma \varepsilon 1$

$$T_\pi^2 v \geq v^* - \varepsilon\underbrace{(1+\gamma+\gamma^2)}\mathbb{1}$$

$$\vdots$$

$$T_\pi^k v \geq v^* - \varepsilon(1+\gamma+\dots+\gamma^k)\mathbb{1}$$

$/ k \to \infty$

$$\downarrow$$

$$v^\pi \geq v^* - \frac{\varepsilon}{1-\gamma}\mathbb{1}$$

Prop: π is greedy wrt. v :

$$\|v - v^*\|_\infty \leq \varepsilon$$

$$\Rightarrow v^\pi \geq v^* - \frac{\varepsilon}{1-\gamma}\mathbb{1}$$

$$\boxed{k \geq H_{\gamma,\, \varepsilon(1-\gamma)}}$$

$$\|v_k - v^*\| \leq \varepsilon(1-\gamma)$$

π_k greedy wrt v_k

$\Rightarrow$ Prop. $\quad v^{\pi_k} \geq v^* - \dfrac{\varepsilon\cancel{(1-\gamma)}}{\cancel{1-\gamma}}\cdot\mathbb{1}$

$$O\Big(S^2 A \underbrace{H_{\gamma,\, \varepsilon(1-\gamma)}}_{\frac{\log\left(1/\varepsilon(1-\gamma)^2\right)}{1-\gamma}}\Big)$$

$$\log(1/\varepsilon) \;; \quad \frac{1}{1-\gamma}$$

Yinyu Ye

P.I. $\boxed{O(\text{poly}(S,A)/_{1-\gamma})}$

outputs π : $v^{\pi} = v^*$!

No ε!

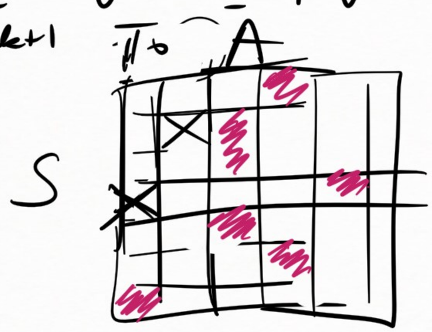
Policy Iteration

deterministic

π_0 ML arbitrarily

$k = 0, 1, \dots$

π_{k+1} : $T_{\pi_{k+1}} v^{\pi_k} = T v^{\pi_k}$



$\boxed{SA-A}$

Computation in round $k = 0, 1, \dots$:

$v^{\pi_k} = ?$ / $T^{i}_{\pi_k} v^{\pi_{k-1}} \dots$

$$\underbrace{T_{\pi_k}} v^{\pi_k} = v^{\pi_k}$$

$$r_{\pi_k} + \gamma P_{\pi_k} v^{\pi_k} = v^{\pi_k}$$

$$v^{\pi_k} = \underbrace{(I - \gamma P_{\pi_k})^{-1}}_{S \times S} r_{\pi_k}$$

---

1. $\boxed{\| v^{\pi_k} - v^* \|_\infty \le \gamma^k \| v^{\pi_0} - v^* \|_\infty}$

$\exists s_0 \in S$; $\boxed{\text{assuming } \pi_0 \text{ not optimal}}$

2. After $k^* = \tilde{O}\left(\frac{1}{1-\gamma}\right)$

$\forall k \ge k^*$ $\quad \pi_k(s_0) \ne \pi_0(s_0)$

$v^* \ge v^{\pi_{k+1}} \ge T v^{\pi_k} \ge T^k v^{\pi_0}$

$O(S^3) \vee O(S^2 A)$

per iteration

After $\tilde{O}\left(\frac{SA}{1-\gamma}\right)$

$\pi_k : v^{\pi_k} = v^*$

No $\log(1/\varepsilon)$

Lower bound: Yichen Chen
Mengdi Wang
2017

$\boxed{\Omega(S^2 A)}$

$\boxed{A^S}$

$\boxed{V.I}$

$\emptyset$

$\Omega(S)$