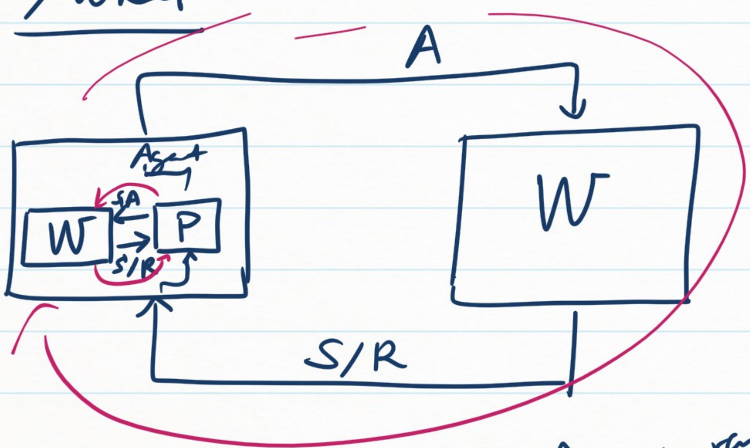
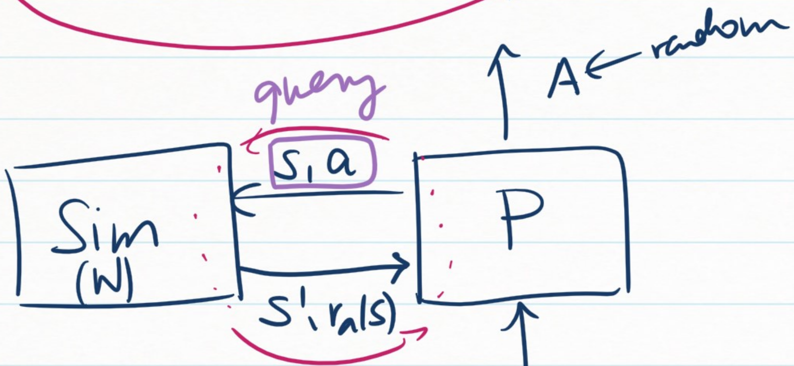
Local planning / Online planning

Why / What





$s' \sim P_a(s)$
↑ random

} Stochasticity leveraged to save on compute!

$\pi(a|s_0) = \mathbb{P}_{s_0}(A = a)$ Certainty Equivalence

No caching : memoryless planner
⇒ memoryless policy

$\boxed{Sim \overset{?}{=} W}$
sensitivity

Goal: ① $v^\pi \geq v^* - \delta 1$
$v^\pi(s)$

$\delta > 0$: input to the planner

② efficiency

* computation-time

* #query / query-cost

$v_0 = 0$

$v_{k+1} = T v_k \quad , k = 0, 1, \ldots$

δ- support policy?

$k = ?$

$v_k = T^k 0 \rightarrow v^*$

$k = H_{\gamma,(1-\gamma)^2\delta} \approx \frac{\log(1/\delta)}{1-\gamma}$

$$\boxed{\arg\max_a \; r_a(s_0) + \gamma \langle P_a(s_0), v^* \rangle}$$

optimal actions!

$\boxed{S_1, \ldots, S_m \sim P_a(s_0)}$ Sampling

$$\langle \hat{P}_a(s_0), v^* \rangle = \frac{1}{m} \sum_{i=1}^{m} v^*(S_i)$$

$$\arg\max_a \; r_a(s_0) + \gamma \langle \hat{P}_a(s_0), v_k \rangle$$

Independent of the size of the state space!

Deterministic MDP $\quad s' = g(s,a)$

$$v_k(s) = (T^k 0)(s)$$

$$= \max_a r_a(s) + \gamma \langle \underbrace{P_a(s)}_{\delta_{g(s,a)}}, T^{k-1} 0 \rangle$$

$$= \max_a \left(r_a(s) + \gamma \underbrace{(T^{k-1} 0)}_{v_{k-1}}(g(s,a)) \right)$$

```
def v(k, s) # v_k(s)
  if k=0 return 0;
  q = [r_a(s) + γ v(k-1, g(s,a)) for a∈A]
  return max(q)
```

queries

$O(A^k) \rightarrow O((mA)^k)$

k

$A = 3$

How big should be m?

$$\operatorname*{argmax}_a \left[\underbrace{r_a(s) + \gamma \langle P_a(s), v^* \rangle}_{q^*(s,a)} \right]$$

$\underbrace{q^*}$ "optimal value of a"

$$v^*(s) = \max_a q^*(s,a) \quad \forall s$$

[B.O.E.]

$$q^*(s,a) = r_a(s) + \gamma \langle P_a(s), \max_{a'} q^*(\cdot, a') \rangle$$

$$M : \mathbb{R}^{SA} \to \mathbb{R}^S$$
$$q \mapsto (Mq)(s) = \max_a q(s,a)$$

$$q^*(s,a) = r_a(s) + \gamma \langle P_a(s), Mq^* \rangle \quad \forall s,a$$

$\lfloor_{max}$:

B.O.E q^*

$$q^* = r + \gamma P M q^*$$

$$P : \mathbb{R}^S \to \mathbb{R}^{SA}$$
$$v \mapsto (Pv)(s,a) = \langle P_a(s), v \rangle$$

$$r : \mathbb{R}^{SA} \to \mathbb{R}$$
$$r(s,a) = r_a(s).$$

M_π

$$\tilde{T}q = r + \gamma P M q$$

$$q^* = \tilde{T} q^*$$

$$T \doteq \tilde{T} \quad | \quad q^* = T q^*$$

$$\underset{a}{\operatorname{argmax}} \; (T^k 0)(s_0, a)$$

$$(Tq)(s,a) = r_a(s) + \gamma \underbrace{\langle P_a(s), Mq \rangle}_{\text{costly!}}$$

$P_a(s) \to \hat{P}_a(s)$

$$C(s,a) = [S_{sa}^{(1)}, \ldots, S_{sa}^{(m)}]$$

$$S_{sa}^{(i)} \sim P_a(s), \quad i = 1 \ldots m$$

i.i.d.

$$(\hat{T}q)(s,a) = r_a(s) + \gamma \frac{1}{m} \sum_{s' \in C(s,a)} (Mq)(s')$$

$$= r_a(s) + \frac{\gamma}{m} \sum_{s' \in C(s,a)} \max_{a'} q(s', a)$$

$T \simeq \hat{T}$ "random approx"

$$\boxed{\underset{a}{\operatorname{argmax}} \; (\hat{T}^k 0)(s_0, a)}$$

Complexity? Branching: mA

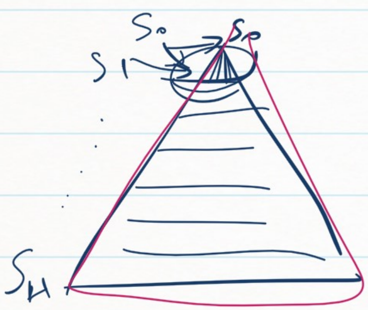
Depth k: $O((mA)^k)$

cost / query - cost

memoization to get $C(s,a)$

$\hat{T}^H 0 \approx q^*$ only true at s_0

high precision



low precision

$S_0 = \{s_0\}$

$S_1 = \{s_0\}$ + neighbors

⋮

$S_h = \{ s \in S \mid \text{dist}(s_0, s) \le h \}$

S_H = all the states encountered.

$|(\hat{T}^H 0)(s_0,a) - q^*(s_0,a)| \le ?$

$$\|q\|_S = \max_{\substack{s\in S\\ a\in A}} |q(s,a)|$$

small

$\delta_H = \|\hat{T}^H 0 - q^*\|_{S_0}$

$\delta_{H-1} = \|\hat{T}^{H-1} 0 - q^*\|_{S_1}$

⋮

$\delta_h = \|\hat{T}^h 0 - q^*\|_{S_{H-h}}$

⋮

$\delta_0 = \|\underbrace{\hat{T}^0 0}_{0} - q^*\|_{S_H} \le \frac{1}{1-\gamma}$

$h>0$:

$$\delta_h = \|\hat{T}^h 0 - q^*\|_{S_{H-h}}$$

$$\le \underbrace{\|\hat{T}^h 0 - \hat{T} q^*\|_{S_{H-h}}}_{\le \gamma \delta_{h-1}} + \boxed{\|\hat{T} q^* - T q^*\|_{S_H}}$$

$+ \frac{\varepsilon'}{1-\gamma}$

$u' = Mu$
$v' = Mv$

$\|\hat{T}^h 0 - \hat{T} q^*\|_{S_{H-h}}$

$= \| \hat{T} \underbrace{\hat{T}^{h-1} 0}_{u} - \hat{T} \underbrace{q^*}_{v} \|_{\underbrace{S_{H-h}}}$

$(\hat{T} u)(s,a) = r_a(s) + \frac{\gamma}{m} \sum_{s' \in C(s,a)} u'(s')$

$dist(s_0, s) \le H-h$

$dist(s_0, s') \le \underline{H-h+1}$

$\forall s' \in C(s,a)$

$\|\hat{T} u - \hat{T} v\|_{S_{H-h}} \le$

$\le \max_{s \in S_{H-h}, a \in A} \underbrace{\left| \frac{\gamma}{m} \sum_{s' \in C(s,a)} u'(s') - v'(s') \right|}$

$\le \gamma \max_{s' \in S_{H-h+1}} |u'(s') - v'(s')|$

$\le \gamma \underbrace{\|u - v\|_{S_{H-h+1}}}_{\delta_{h-1}}$

$\delta_h \le \gamma \delta_{h-1} + \frac{\varepsilon'}{1-\gamma}$

$\vdots$

$\delta_H \le \underbrace{\text{small} \ldots}_{\text{finish}}$

$\boxed{\varepsilon' ?}$

$(m, H) = f(\varepsilon)$

?