Local Planning

$$\boxed{A_{s_0} = \underset{a}{\arg\max}\, (\hat{T}^H_{s_0} v)(s_0, a)}$$

What does this buy us??

Localizing computation to s_0

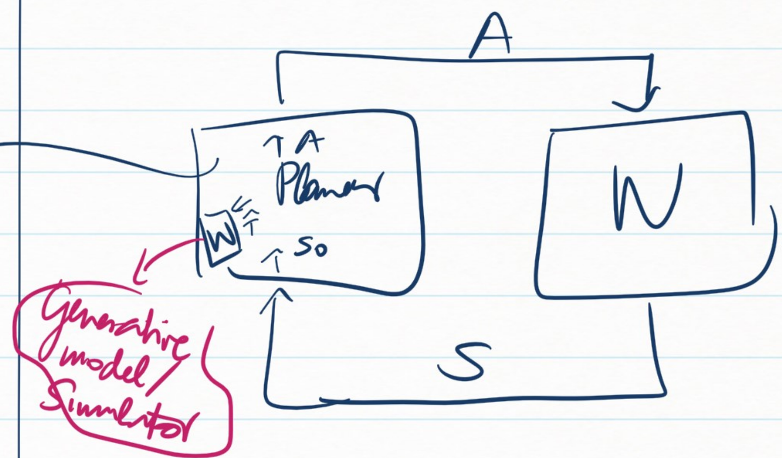
Goal? | $A = a_1$

① Little compute

② Input : $\delta > 0$

δ-optimality

π



Planner $\rightarrow$ $\pi(a \mid s_0) = \mathbb{P}_{s_0}(A_{s_0} = a)$

$$v^\pi \geq v^* - \delta \mathbf{1}$$

per-state compute cost

$$\boxed{O\left((A/\delta)^H\right) \ldots}$$

$\boxed{\langle P_a(s), v \rangle}$ $\boxed{\Omega(S)}$

$$C(s,a) \leftarrow$$

$$(s,a) \longmapsto \overbrace{S'_1(s,a), \dots, S'_m(s,a)}^{C(s,a)} \sim P_a(s)$$

independance

"call simulator"

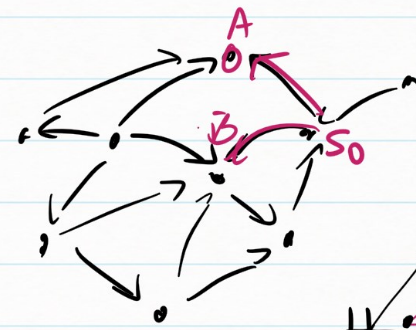
$$C(s) = (C(s,a))_{a\in A} \,/\, . = \bigcup_{a\in A} C(s,a)$$

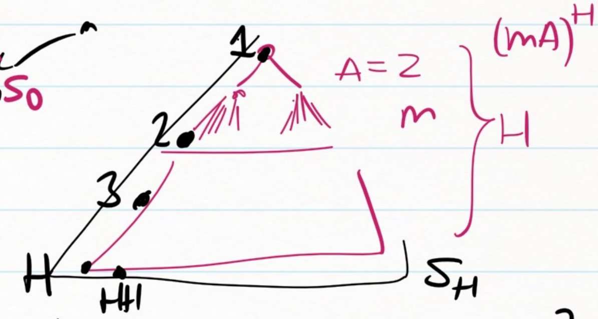
$$\hat{T} : \mathbb{R}^{S\times A} \to \mathbb{R}^{S\times A}$$

$$\boxed{\begin{aligned}(\hat{T}q)(s,a) &= r_a(s) + \gamma \frac{1}{m} \sum_{s'\in C(s,a)} \overbrace{\max_{a'} q(s',a')}^{(Mq)(s')} \\ (Tq)(s,a) &= r_a(s) + \gamma \langle P_a(s), Mq \rangle\end{aligned}}$$

$$G = (S, \mathcal{E})$$

$$\mathcal{E} = \{(s,s') \mid s' \in C(s)\}$$





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

$$h = 0 \qquad S_0 = \{s_0\}$$

$$S_1 = \{s_0, A, B\}$$

$$\vdots$$

$$-\hat{T}q^* + \hat{T}q^*$$

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

$$\delta_h \le \gamma \delta_h + \boxed{\underbrace{\| \hat{T}q^* - Tq^* \|_{S_H}}_{\le \frac{\varepsilon'}{1-\gamma}}} \quad \text{"whp"}$$

$$\delta_0 \le \frac{1}{1-\gamma}$$

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

$$\delta_0 \le \frac{1}{1-\gamma}$$

$$\delta_1 \le \frac{\gamma + \varepsilon'}{1-\gamma}$$

$$\delta_2 \le \frac{\gamma(\gamma+\varepsilon') + \varepsilon'}{1-\gamma} = \frac{\gamma^2 + \varepsilon'(1+\gamma)}{1-\gamma}$$

$$\vdots$$

$$\delta_H \le \frac{\gamma^H + \varepsilon' \overbrace{(1+\gamma+\dots+\gamma^{H-1})}^{\le \frac{1}{1-\gamma}}}{1-\gamma}$$

$$\le \left(\gamma^H + \frac{\varepsilon'}{1-\gamma}\right)\frac{1}{1-\gamma} \overset{?}{\le} \delta \quad \boxed{Mq^* = v^*}$$

$$\boxed{\varepsilon' = ?} \qquad \underbrace{\|\hat{T}q^* - Tq^*\|_{S_H}}$$

$$\max_{s \in S_H} \max_a \gamma \left| \frac{1}{m}\sum_{j=1}^{m} v^*(S_j'(s,a)) - \langle P_a(s), v^* \rangle \right| \Rightarrow$$

<u>Hoeffding's inequality</u>

Lemma: $0 \le X_i \le 1$ i.i.d. $i = 1, \dots, n$

$$(C)\begin{cases} \forall\, 0 \le \delta < 1 \quad \text{wp at least } 1-\delta, \\ \left| \frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E} X_1 \right| \le \sqrt{\frac{\log(2/\delta)}{2n}} \end{cases}$$

Lemma: $0 \le X_i \le 1$, $i = 1, \dots, n$

random element U

$(U, X_1, \dots, X_n)$ jointly distributed.

$X_1, \dots, X_n$ i.i.d. given U

$$\left[\mathbb{P}(X_i \in A_i,\ i = 1..n \mid U) = \prod^{n} \mathbb{P}(X_i \in A_i \mid U) \right]$$

$\forall\, 0 \le \delta < 1$ wp at least $1-\delta$

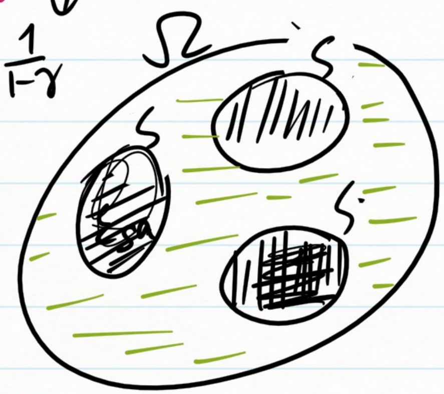
$$\left| \frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X_1 \mid U] \right| \le \sqrt{\frac{\log(2/\delta)}{2n}}$$

(s,a)

$$\|\hat{T}q^* - Tq^*\|_{S_H}$$

$$= \max_{s\in S_H}\ \max_{a\in A}\ \underbrace{\left|\frac{1}{m}\sum_{j=1}^{m} v^*(S_j'(s,a)) - \langle P_a(s), v^*\rangle\right|}_{\Delta(s,a)}$$

$\frac{1}{1-\gamma}$



Union bounds!

(outside of $\mathcal{E}_{sa}$)

$\forall (s,a)$ $|\Delta(s,a)| \le \frac{1}{1-\gamma}\sqrt{\frac{\log(2/\zeta)}{m}}$

$\mathbb{P}(\mathcal{E}_{sa}) \le \zeta$

Outside of $\bigcup_{sa}\mathcal{E}_{sa} \to \underbrace{\max_{(s,a)}|\Delta(s,a)| \le \frac{1}{1-\gamma}\sqrt{\frac{\log(2/\zeta)}{2m}}}_{\frac{\varepsilon'}{1-\gamma}}$

$$\mathbb{P}\Big(\big(\bigcup_{sa}\mathcal{E}_{sa}\big)^c\Big) = 1 - \mathbb{P}\Big(\bigcup_{sa}\mathcal{E}_{sa}\Big)$$

$$\ge 1 - \sum_{s,a}\mathbb{P}(\mathcal{E}_{sa}) \ge 1 - SA\zeta$$

wp $1-\delta_{failure}$

$\delta_{failure} = SA\zeta$

$\zeta = \frac{\delta_{failure}}{SA}$

random

$|S_H| \le (mA)^{H+1}$

take union bound over these?!

$$\frac{\varepsilon'}{1-\gamma} \le \frac{1}{1-\gamma}\sqrt{\frac{\log\left(\frac{2SA}{\delta_{failure}}\right)}{2m}}$$

$\delta/3$ $\delta/3$

$H = \frac{1}{(1-\gamma)^2}$

$$\frac{\gamma^H + \frac{1}{(1-\gamma)^2}\sqrt{\frac{\log\left(\frac{2SA}{\delta_{failure}}\right)}{2m}}}{1-\gamma}$$

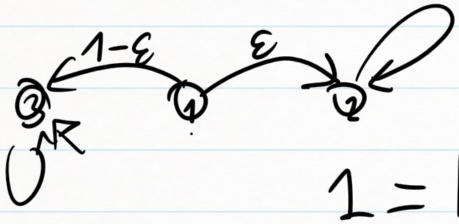
$(Am)^{H+1}$

$m \approx H\log(Am)$

$m \approx H\log(A)$

$$\frac{2\,\boxed{\delta_{failure}}}{(1-\gamma)^2} = \frac{\delta}{3}$$ ← target subopt.

$m = \Omega(\log(S))$



$S' \sim P_1$

S random

$1 = \mathbb{P}(S'=2 \mid S=2) = \underbrace{\mathbb{P}(S'=2)}_{=\varepsilon}$

$\boxed{S \doteq S'}$

$\{\hat{S}_1, \ldots, \hat{S}_N\}$ $\forall\, 1 \le i \le N$ n

$S_H = \{S_1, \ldots, S_n\}$

$n = (Am)^{H+1} - 1 / A-1$

$S_1, \ldots, S_n$ depth-first encounters of states in S_H

Lemma:

$\forall i \in [n]$ wp $1-\delta$
$\forall a \in [A]$

$$|\Delta(S_i, a)| \le \frac{1}{1-\gamma}\sqrt{\frac{\log(2/\delta)}{2m}}$$

Proof: Fix i, a

$J_i = \min\{1 \le k \le i \mid S_k = S_i\}$

$1 \le J_i \le i$

$$\Delta(S_i,a) = \underbrace{\frac{1}{m}\sum_{s' \in C(S_i,a)} v^*(s')}_{S_{J_i}} - \underbrace{\langle P_a(S_i), r^*\rangle}_{S_{J_i}}$$

$\Delta(S_i,a) = \Delta(S_{J_i},a)$

$C(S_{J_i},a) \perp S_{J_i}$ given S_{J_i}

$C(S_1), \ldots, C(S_{J_i-1})$

+ union bound over $i \in [n]$

+ everything together

Thm: Cost $= O\left((mA)^H \right)$

$H \approx H_{\gamma, (1-\gamma)^2 \delta}$

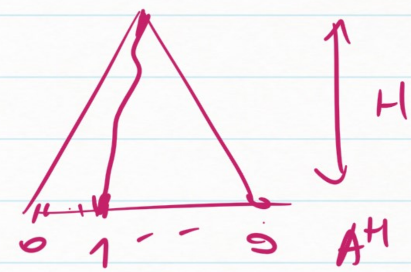
$m \approx H \log(A)$

local planning gives π:

$$v^\pi \geq v^* - \delta \mathbb{1}$$

Kearns, Mansour, Singh 20XXs

$\hat{T}$



[recursion
fresh samples

$\Omega(A^H)$