Function approximation + Planning

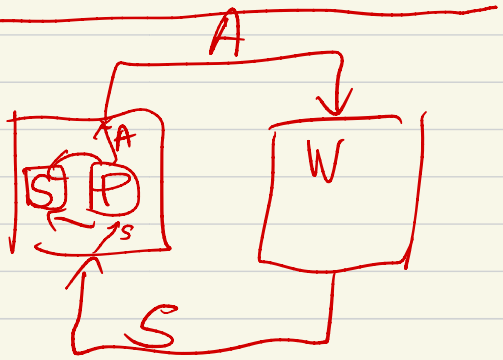
Review

Tabular MDPs

S states
A actions

$\Omega(S^2A)$

S big!

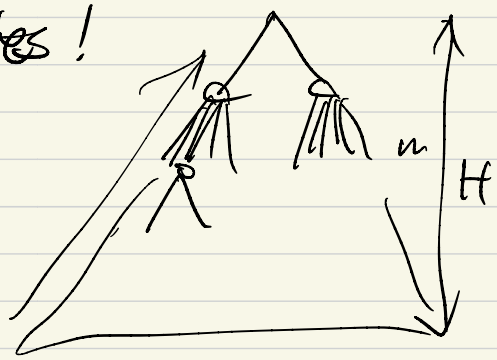


Local Planning

$O((H^5A)^H)$ $\quad H \approx \frac{1}{1-\gamma}$

runtime indep. of # states!

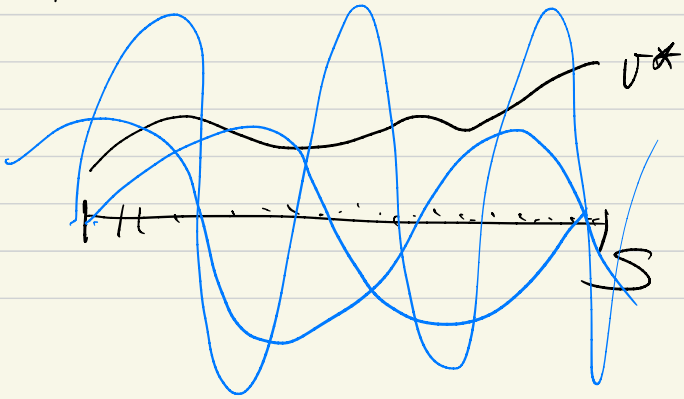
$\Omega(A^H)$



$\gamma = 0.99$ $\quad 1-\gamma = \frac{1}{\frac{1}{100}} = 100$

$A = 2$, $\boxed{2^{100}}$

Value function approximation



$\hat{V}(S)$

$$V^*(S) = \sum_{i=1}^{d} \theta_i \varphi_i(S)$$

ε

??

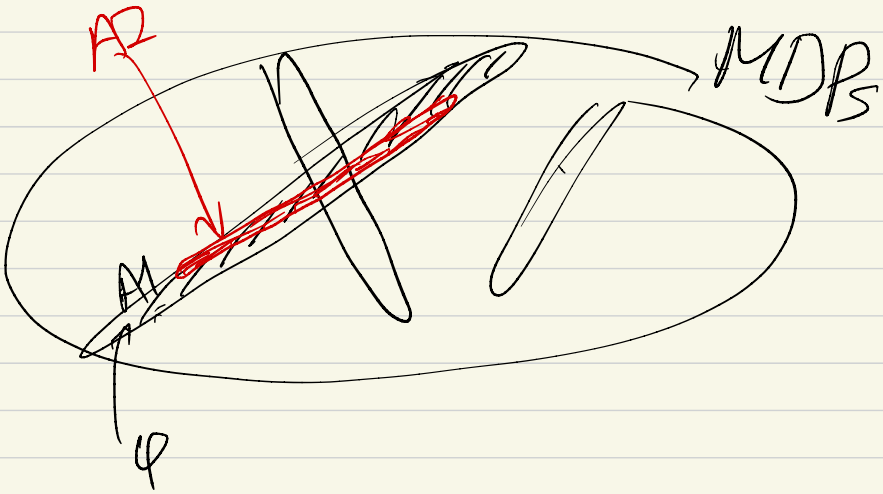
One-step lookahead

$$\hat{\pi}(s) = \arg\max_a r_a(s) + \gamma \underbrace{\langle P_a(s), \hat{v} \rangle}_{\text{sampling}}$$

$$\frac{1}{1-\gamma} \simeq \|\hat{v} - v^*\|_\infty \leq \varepsilon \equiv \frac{1}{1-\gamma}\{v^{\hat{\pi}} \geq v^* - \boxed{\frac{2\gamma\varepsilon}{1-\gamma}}1$$

$\gamma = 0,99$

$\theta \in \mathbb{R}^d \qquad d \ll S$



$\varphi(s) = [\varphi_1(s), \dots, \varphi_d(s)]^T$

$$n = |S| \left\{ \begin{array}{l} v^* = \begin{bmatrix} \varphi^T(s_1)\theta_* \\ \vdots \\ \varphi^T(s_n)\theta_* \end{bmatrix} = \begin{bmatrix} -\varphi^T(s_1)- \\ \vdots \\ -\varphi^T(s_n)- \end{bmatrix} \theta_* = \Phi\theta_* \\ \qquad\qquad \underbrace{\qquad\qquad}_{d} \end{array} \right.$$

$n \gg d$

$v^* \in \text{span}(\Phi) = \{\Phi\theta \mid \theta \in \mathbb{R}^d\} \leq \mathbb{R}^n$

$\mathcal{F}$

(A1) $v^* \in \text{span}(\Phi)$: v^* - realizability

(A2) $\forall \pi: v^\pi \in \text{span}(\Phi)$

Note: (A2) $\Rightarrow$ (A1); FT $\exists \pi: v^\pi = v^*$.

(B1) $q^* \in \text{span}(\Phi)$ $\quad \Phi := \tilde{\Phi}$

(B2) $\forall \pi : q^\pi \in \text{span}(\Phi)$ ⟵

(B2) $\Rightarrow$ (B1) $\quad \mathcal{F} = \text{span}(\Phi)$

Today: Assume (B2) holds.

Simulator access.

$\exists$ local planner s.t.

poly $(A, d, H, \frac{1}{\delta})$ queries/compute

$\rightsquigarrow \pi$: $V^\pi \geq V^* - \delta 1$.

Value Iteration: $Q_0 = 0$, $q_{k+1} = Tq_k$ ($q_k \in \mathcal{F} \Rightarrow q_{k+1} \in \mathcal{F}$)

Policy Iteration:

(B2)

π_0 arbitrary

π_{k+1} greedy wrt. V^{π_k}

π_{k+1} greedy wrt. q^{π_k}

$\forall q \in \mathcal{F} \Rightarrow Tq \in \mathcal{F}$

closedness/invariance $T\mathcal{F} \subseteq \mathcal{F}$

$q^{\pi_k} = r + \gamma P V^{\pi_k}$

$r(s,a)$ $\quad (PV)(s,a) = \langle P_{sa}(\cdot), V \rangle$

$\pi_{k+1}(s) = \operatorname{argmax}_a q^{\pi_k}(s,a) = \operatorname{argmax}_a \psi(s,a)^\top \theta_k$

$q^{\pi_k} = \Phi \theta_k$, $\theta_k = ?$

How to calculate θ_π ?

$\pi : S \to A$, (B2) , simulator

↳ point-wise evaluate π

Need: $\theta \in \mathbb{R}^d$ $\quad \Phi\theta = q^\pi$

Policy evaluation $\quad q^\pi = r + \gamma P M_\pi q^\pi$

$P_\pi : \mathbb{R}^{SA} \to \mathbb{R}^{SA}$

$q^\pi = \Phi\theta$ $\qquad$ $q^\pi = r + \gamma P_\pi q^\pi$

$q^\pi \in \mathbb{R}^{SA}$ SA unknowns

$\theta \in \mathbb{R}^d$ d unknowns

SA+d

$\Phi\theta = r + \gamma P_\pi \Phi\theta$

unknowns: d

$r \in \mathbb{R}^{SA}$ OOPS! we have SA equations!

$$\underbrace{\begin{bmatrix} \\ (I - \gamma P_\pi)\Phi \\ \\ \end{bmatrix}}_{d} \Big\}\, SA \qquad \begin{bmatrix} \theta \\ \vdots \\ \theta \end{bmatrix}_d = \underbrace{\begin{bmatrix} r \\ \vdots \\ r \end{bmatrix}}_{1}\, SA$$

- SVD
- LU decomposition
- * Subsample rows
- * Least-squares

What if (B2) does not even hold??

↳ Robust

$\text{poly}(d, A, H, \frac{1}{\delta}, \dots)$ indep. of S

Weighted least-squares

$\phi\theta = q^\pi$

$q^\pi(s,a) = \varphi(s,a)^\top \theta$

"Least-squares" / pervers. est. problem

$(s,a) \in C \subseteq S \times A$

$q^\pi \approx_\varepsilon T_\pi^k 0$

$k \geq H_{\gamma,\delta} \Rightarrow \varepsilon \leq \delta$

k-step truncated discounted return.

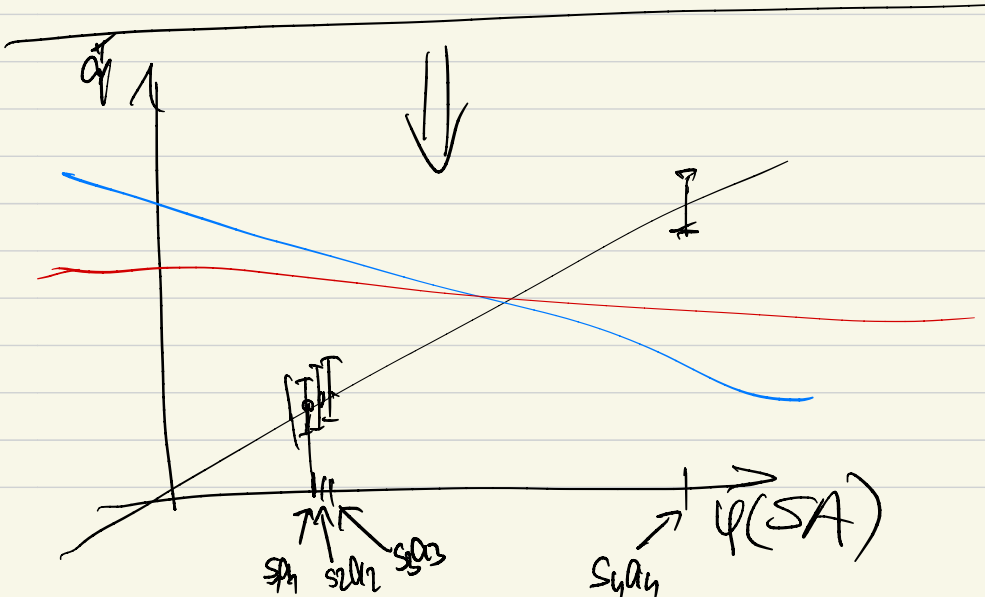
$$q^\pi(s,a) \approx \frac{1}{m} \sum_{i=1}^{m} \sum_{t=0}^{k} \gamma^t r_{A_t^{(i)}}(S_t^{(i)}) = \hat{R}(s,a)$$

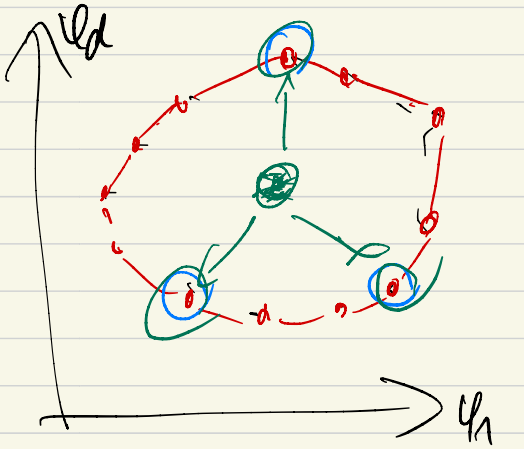
$\hat{R}(s,a)$

simulations:

$$A_t^{(i)} = \pi(S_t^{(i)}), \quad t \geq 1$$
$$S_{t+1}^{(i)} \sim P_{A_t^{(i)}}(S_t^{(i)})$$
$$S_0^{(i)} = s, \quad A_0^{(i)} = a$$

$$\mathbb{E}[\hat{R}(s,a)] = (T_\pi^k 0)(s,a)$$





Theorem (Kiefer - Wolfowitz):

$\varphi: S \times A \to \mathbb{R}^d$ $\qquad \sum_{(s,a)} g(s,a) = 1$

$\exists\, C \subseteq S \times A$, $g : C \to [0,1]$

$|C| \leq d(d+1)/2$

s.t. $\max_{(s,a)} \|\varphi(s,a)\|_{G_g^{-1}} \leq \sqrt{d}$

$G_g = \sum_{(s,a) \in C} g(s,a)\, \varphi(s,a)\, \varphi(s,a)^T$

$\|x\|_P^2 = x^T P x$

$|C| \approx 2d \log\log d + 16$ (Todd, 2016)

Optimal Experimental Design

How to find a small C?