March 25

Batch RL $\forall S, A$

Thm $\forall \pi_{beh.} \in ML(S, A)$ — π designs

$\forall \varepsilon, \delta$ $\forall A$ sound $H = H_{\delta,\varepsilon}$

$\exists M = (S, [A], P, r, \gamma)$ $n(A, M, \varepsilon, \delta) \geq c_0 \cdot A^H$

$S \geq H_{\delta,\varepsilon}$ $d = SA = HA$ $= c_0 \left(\frac{d}{H}\right)^H$

$Z = (Z_1, Z_2, \ldots, Z_n, \ldots)$ } model-based

$Z_n \in (S \times A)^n$

2-design / fixed-design

$\boxed{Z = S \times A}$ $\text{poly}(S, A, H, \frac{1}{\varepsilon^2})$

$\varphi: S \to \mathbb{R}^d$, $\varphi: S \times A \to \mathbb{R}^d$

(M, φ)

→ q^π-realizability: $\forall \pi: q^\pi \in \mathcal{F}_\varphi$ ✓

q^*-realizability: $q^* \in \mathcal{F}_\varphi$ ✗

V^π

V^*

Bi ⊂ Planning: Adaptively choosing (sa) ⊂ Batch learning: Passively choosing π/(sa)

① Way-Foster-Kakade → (1.5) Amortila-Jiang-Xie-... ② Andrea Zanette

Theorem: $\forall d>0$, $H>0$ (horizon) $\quad P=(P_h)_{h=1}^H$ — def.

OPE $\quad \exists (M,\varphi)$, $M=(S,A,P,r)$; $\varphi: S\times A \to \mathbb{R}^d$, $\|\varphi\|_2 \le 1$

$s_0 \in S$ (=) $\quad r\in[0,1]$ $\quad |S| = O(dH)$, $|A|=2$

radar design → $\exists\, \underline{\nu} = (\nu_1, \dots, \nu_H)$, $\nu_i \in \mathcal{M}_1(S\times A)$

s.t.

① $q_h^\pi \in \mathcal{F}_\varphi = \mathrm{span}(\Phi)$ $\quad \forall h, \forall \pi \in ML$

② $\mathbb{E}_{\nu_h}[\varphi\varphi^T] \succeq \frac{1}{d} I$ $\quad \forall h\in[H]$ (good design)

③ $\forall A$ alg. OPE sound

$\forall \pi \in ML$ $\quad n(A,M,\varphi,\pi) = \Omega\left(\left(\frac{d}{2}\right)^H\right)$ ∞

π-design ?

Theorem: $\forall d>0$, $H>0$ $\quad \exists (M,\varphi)$, $M=(S,A,P,r)$ (stoch. / det.)

OPE. $\quad \varphi: S\times A\to\mathbb{R}^d$ $\|\varphi\|_2\le 1$, $r\in[0,1]$

$|A|=2$, $|S| = O(dH)$

$\exists \pi_{beh}$, $\exists \pi_{trg}$ s.t. $\quad \pi_{beh} \to \nu_h^{beh}$, $h\in[H]$

① $q_h^{\pi_{trg}} \in \mathcal{F}_\varphi$ $\quad \forall h$

② $\mathbb{E}_{\nu_h^{beh}}[\varphi\varphi^T] \succeq \frac{1}{d} I$, $\forall h$

(3.) $\forall A$ sound $\quad n(A, M, \varphi, \pi_{trg}, \pi_{ben}) = \Omega\left(\left(\frac{d}{2}-1\right)^H\right)$

OPO ?≤? OPE

Prop: $n_{OPO} \geq n_{OPE}$ | Z-designs, π-designs

A is "good" at OPO

$\Rightarrow A'(A)$ is "good" at OPE

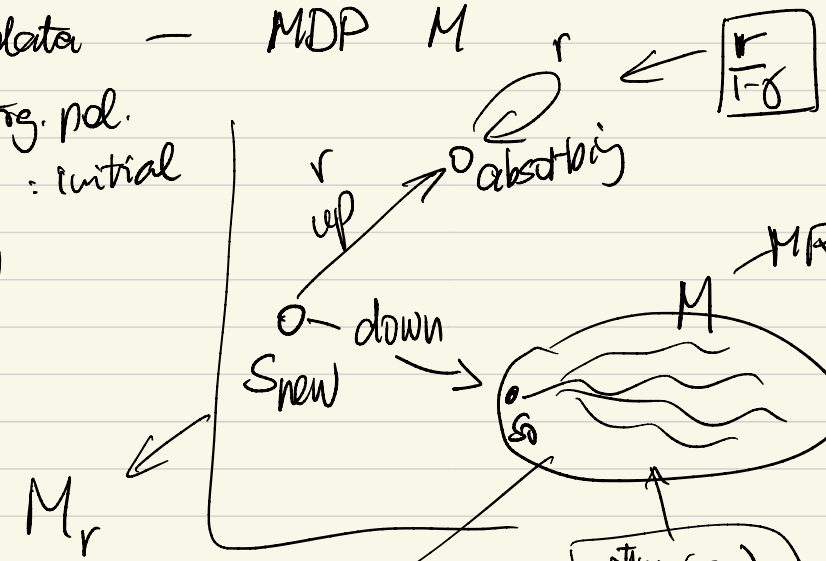
q^π realizability

$\mathcal{D}$: data — MDP M

π_{trg} : trg. pol.

$s_0 \in S$: initial

$V^{\pi_{trg}}(s_0)$



$r \in [0,1]$

$r := r_0 = \frac{1}{2}$

down

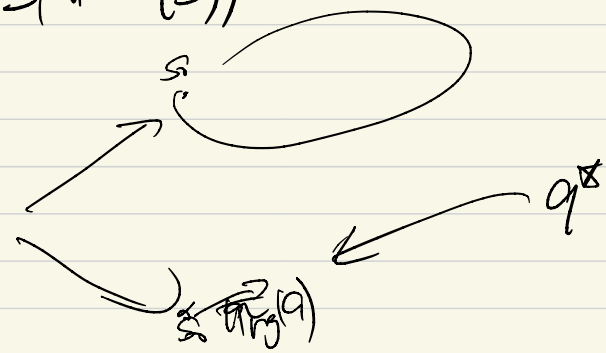
up

$V^{\pi_{trg}}(s_0) > \frac{1}{2}\frac{1}{1-\gamma}$

$r_0 = \frac{3}{4}$

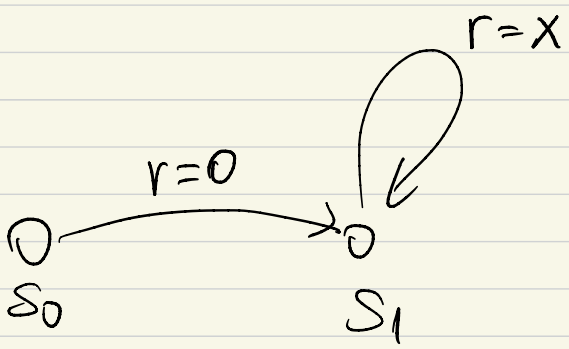
(s,a)

$(s, \pi^{trg}(s))$



Discounted inf. horizon

Z-design



$$q^* = V^* = \left[\frac{\gamma x}{1-\gamma} \;,\; \frac{x}{1-\gamma}\right] = \underbrace{\begin{bmatrix}\gamma & 1\end{bmatrix}}_{\Phi} \frac{x}{1-\gamma}$$

$\varphi(s_0) = \gamma$

$\varphi(s_1) = 1$

Obs: from s_0 only!

$\theta = \frac{x}{1-\gamma}$

$$\mathbb{E}\left[\varphi(S)\varphi(S)\right]^T = \sigma^2$$

unidentifiability

$P(S = s_0) = 1$

Π-design ??

Andrea Zanette

S, A fixed.

π- design $\quad g = (g_i)_{i=1}^{\infty}$

$$g_i = (\pi_1^{(i)}, \dots, \pi_{k_n}^{(i)}, c_1^{(i)}, \dots, c_{k_n}^{(i)})$$

$$\pi_j^{(i)} : S \to A \quad , \quad c_j^{(i)} \in \mathbb{N}$$

$$\to (s_1^{(i)}, \dots, s_{k_n}^{(i)})$$

$Z_n =$ $\{(s_1, a_1), \dots, (s_n, a_n)\}$

$k_n = n$

$s_1^{(i)} = s_1, \dots$

$s_n^{(i)} = s_n$

$\pi_1^{(i)}(s_1) = a_1$

$\vdots$

$\pi_n^{(i)}(s_n) = a_n$

$c_1^{(i)} = \dots = c_n^{(i)} = 1$

$Z \in S \times A$

fix n $\quad \tilde{g} := g_n$

Z_n = set of reachable states

$$\sum_{j=1}^{k_n} c_j^{(i)} \leq n$$

Observation $\quad r|_Z \quad P|_Z$

Theorem 1: OPE fixed target.

$\forall r, d \quad \exists S, A$ s.t. $\forall g$ π-design

$\forall A$ alg sound

$\exists (M, Q, \pi_{reg})$ $\quad M = (S, A, P, r)$ dist.

$\|Q\|_2 \le 1$, $Q: S \times A \to \mathbb{R}^d$ $\quad r \in [-1, 1]$

s.t. $q^{\pi_{reg}} \in \mathcal{F}_Q$

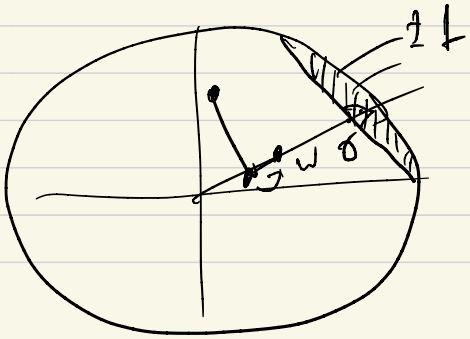
$$n(A, M, Q, \pi_{reg}, g) =$$

$$= \Omega\left(\sqrt{d}\left(\frac{1}{2(1-r)}\right)^d\right)$$

A^H
d^H

$$= \Omega\left(\sqrt{d}\left(\frac{H}{2}\right)^d\right)$$

$S = A = B_2^d = B = \{x \in \mathbb{R}^d \mid \|x\|_2 \le 1\}$



$f(s, a) = a$ $\quad Q(s, a) = a$

π_{reg}

$$\mu(v^* - v^{\hat{\pi}}) \leq f(n, \nu^*, M, \dots)$$

$$\mu v^{\pi} \geq \underbrace{\mu v^* - f(n, \nu^*, M, \dots)}_{F(n, \nu^*, M, \mu, \dots)}$$