March 30

Andrea Zanette $\qquad S, A$ fixed.

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

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

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

$$\left. s_1^{(i)}, \dots, s_{k_n}^{(i)}\right)$$

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

$k_n = n$

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

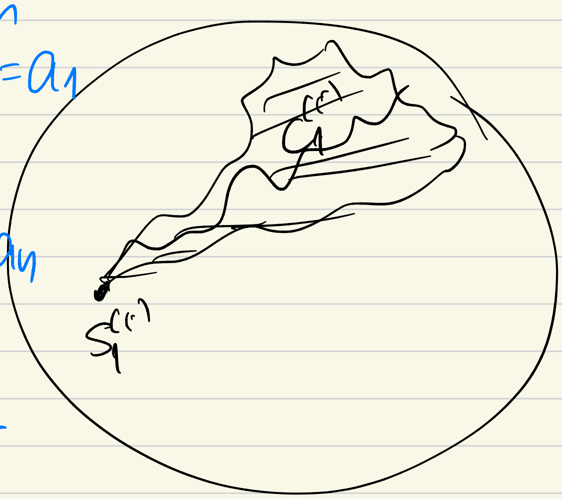
$s_n^{(i)} = s_n$

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

$\vdots$

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

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



Fix n $\qquad \tilde{g} := g_n$

Z_n = set of reachable states

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

$Z \subseteq S \times A$

Observation $\qquad r|_Z \qquad P|_Z$

Theorem 1: $\boxed{\text{OPE}}$ fixed target.

$\forall \gamma, d \quad \exists S, A \quad$ s.t. $\boxed{\forall \xi \quad \pi\text{-design}}$

$\forall \mathcal{A}$ alg $\boxed{\text{sound}}$

$$\exists (M, \varphi, \boxed{\pi_{trg}}) \qquad M = (S, A, \overset{\text{det.}}{P}, \underset{[-1, 1]}{r})$$

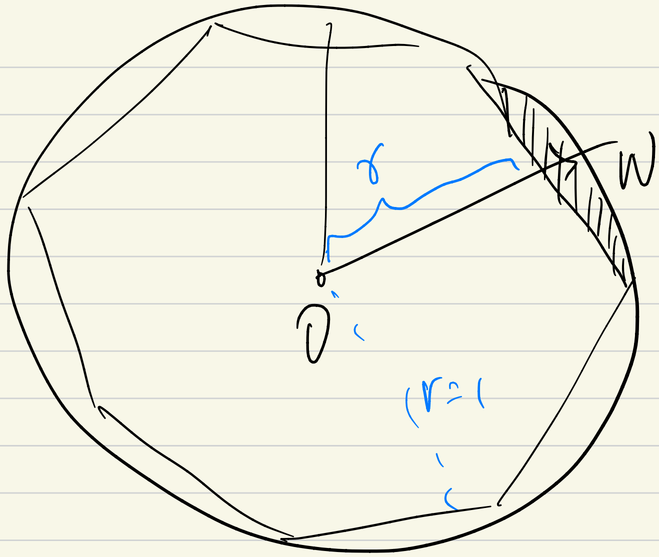
$\|\varphi\|_2 \le 1$, $\varphi : S \times A \to \mathbb{R}^d$

s.t. $\boxed{q^{\pi_{trg}} \in \mathcal{F}_\varphi}$

$$n(\mathcal{A}, M, \varphi, \pi_{trg}, \xi) =$$

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

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



$B = \{x \in \mathbb{R}^d \mid \|x\|_2 \leq 1\}$

$C_\gamma(w) = \{x \in B \mid x^T w \geq \gamma\}$

Lemma:

$\underline{N(\gamma, d)} = \Omega\left(\sqrt{d}\left(\frac{H_\gamma}{2}\right)^d\right)$

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

J-L lemma

Volume calculations

Proof: $\mathcal{A}$, ε π-design

$w \in \partial B$

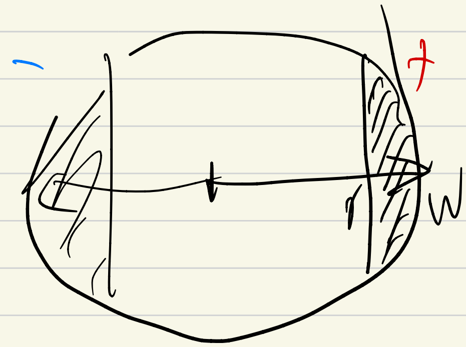
$\sigma \in \{\pm 1\}$

$M^{(w,\sigma)} = (S, \mathcal{A}, P, r^{(w,\sigma)})$

$S = \mathcal{A} = B$, $s_0 = 0$

P: $\underline{f(s,a) = a}$

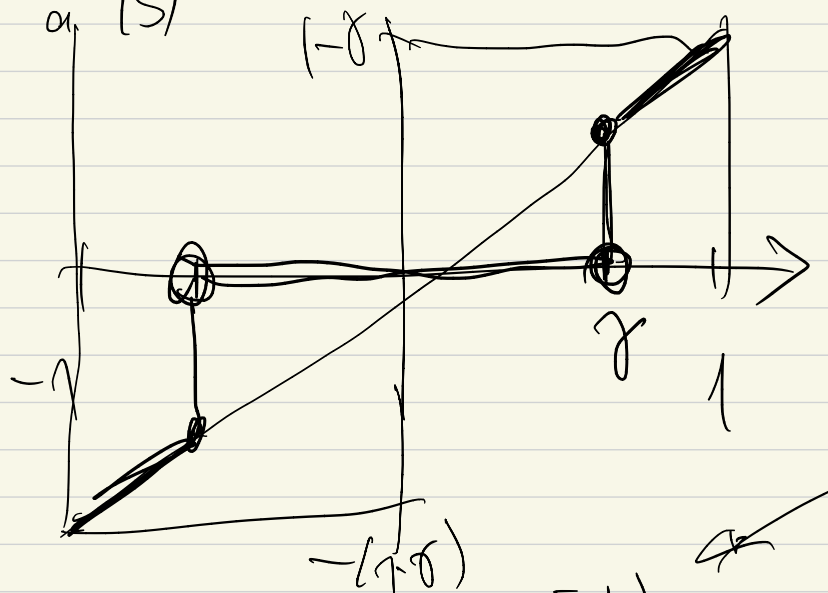
det. dyn.



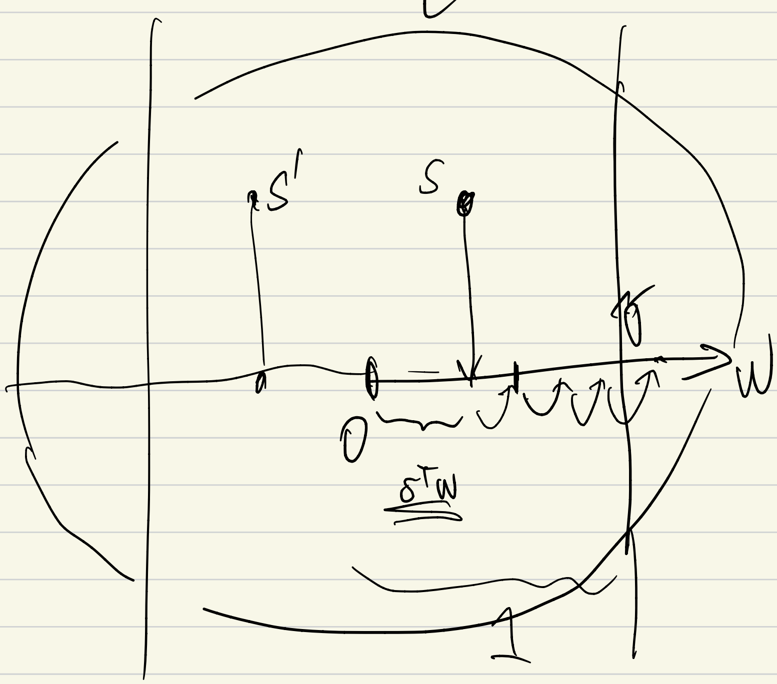
$$r_a^{(w,0)}(s) = \begin{cases} \dfrac{\sigma(1-\gamma)a^T w}{}, & |a^T w| \geq \gamma \\ 0, & \text{o.w.} \end{cases}$$

$r_a^{(w,+1)}(s)$

$\sigma = +1$



$$\Pi_{trg}^{w}(s) = \begin{cases} \left(\dfrac{s^T w}{\gamma}\right) w, & |s^T w| < \gamma \\ s, & \text{o.w.} \end{cases}$$



$$Q(s,a) = a$$

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

(*) $q^{\pi_{trg}^w}(s,a) = \varphi(s,a)^T \theta$

$\boxed{\begin{array}{c} q^{\pi}(s,a) = \sigma a^T w \\ a = w \\ \sigma ! \end{array}}$

Claim: $\theta = \sigma w$

Proof: 1.) Check directly

2.) Bellman equation

$\pi := \pi_{trg}^w$

$$q(s,a) = r_a(s) + \gamma\, q(\underbrace{f(s,a)}_{a}, \pi(\underbrace{f(s,a)}_{a}))$$

$$(s,a) \mapsto \sigma \varphi^T(s,a) w = \underline{\sigma a^T w}$$

$$\sigma a^T w \overset{?}{=} r_a(s) + \gamma \sigma \pi(a)^T w$$

Case 1: $|a^T w| \geq \gamma$

$$RHS = \sigma(1-\gamma) a^T w + \gamma \sigma a^T w = \sigma a^T w \checkmark$$

Case 2: $|a^T w| < \gamma$

$$RHS = 0 + \cancel{\gamma} \sigma \left(\frac{a^T w}{\cancel{\gamma}}\right) \cancel{w^T w} = \sigma a^T w \checkmark$$

sample size n

$S: (s_1, a_1), \dots, (s_n, a_n) \in B^2$

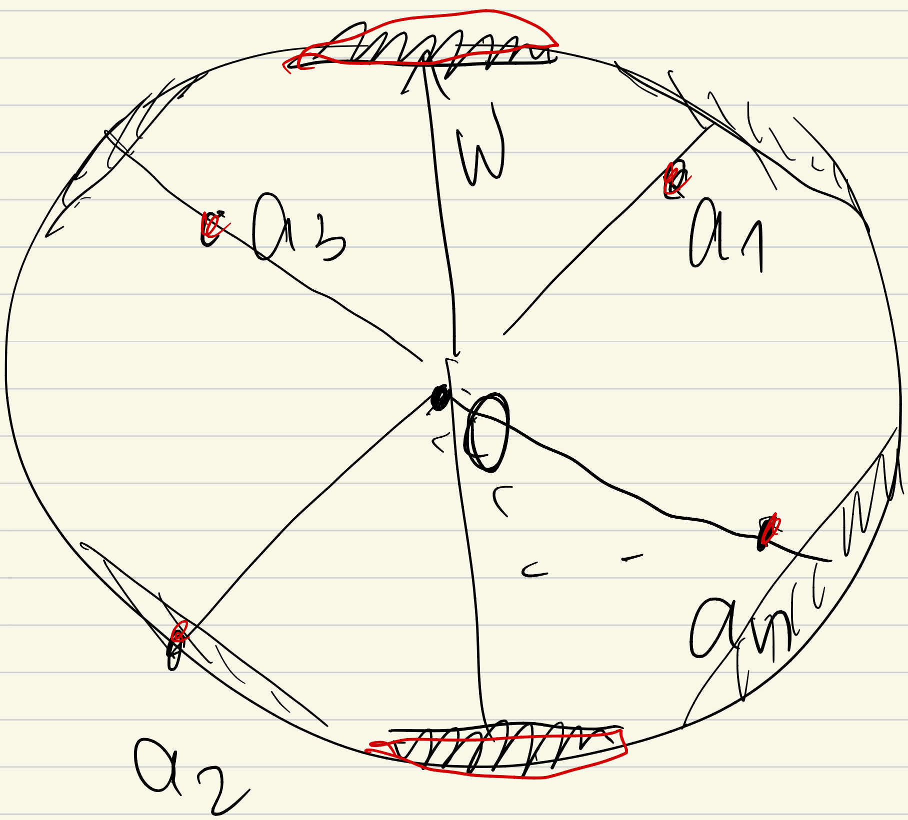
$\boxed{n < N(r, d)} \Longrightarrow \exists w \in \partial B$

s.t.

$C_r(w) \cap C_r(a_i) = \emptyset$

$C_r(-w) \cap C_r(a_i) = \emptyset$

$\forall i \in [n]$



$\sigma \in \{\pm 1\}$

// q.e.d.

First result: evaluating fixed target policy which has linear q^π.

Theorem 2: $\boxed{q^* \in \mathcal{F}_\varphi.}$

Plan: $2^{\Omega(d \wedge H)}$

$\forall \sigma, d$ $\forall \xi$ $\underline{\pi\text{-design}}$

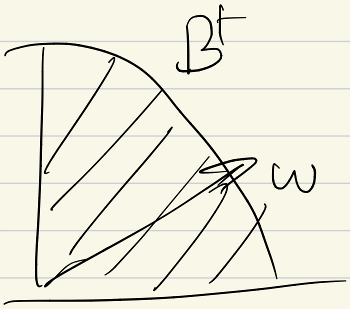
$\forall A$ alg. sound q^*-est.

$\exists (M, \varphi, s_0)$, $\varphi \in \mathbb{R}^{dH}$, $\|\varphi\|_2 \le 1$

$q^* \in \mathcal{F}_\varphi$ s.t.

$$n(A, M, \varphi, \xi, s_0) = \Omega\left(\frac{\sqrt{d}}{2d}\left(\frac{H}{2}\right)^{\frac{d-1}{2}}\right)$$

Proof: $B^+ = \{x \in B \mid x \geq 0\}$

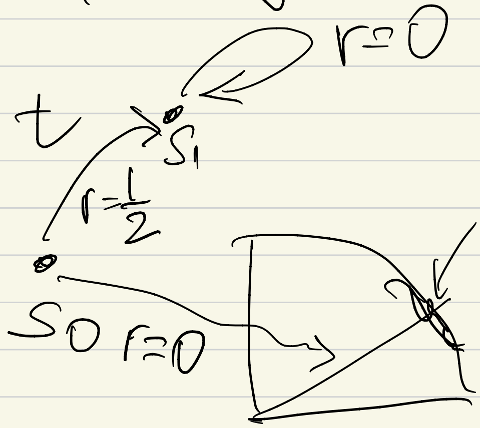


$$S = \{s_0, s_1\} \cup B^+$$

$$\mathcal{A} = \{t\} \cup B^+$$

$$w \in \partial B \cap [0,\infty)^d$$

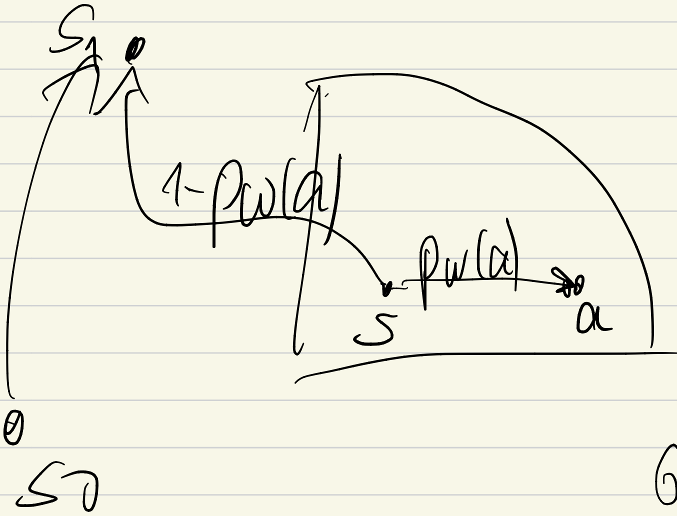
$$\sigma \in \{0,1\}$$



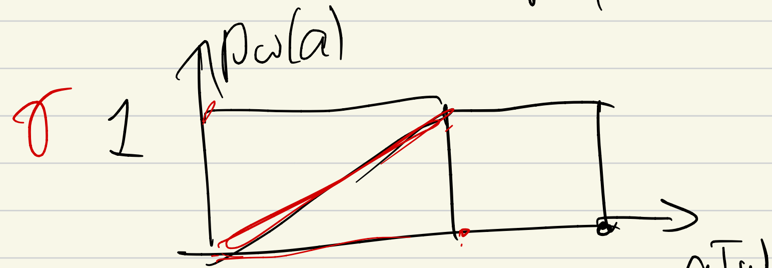
$s \in \{s_0\} \cup B^+$

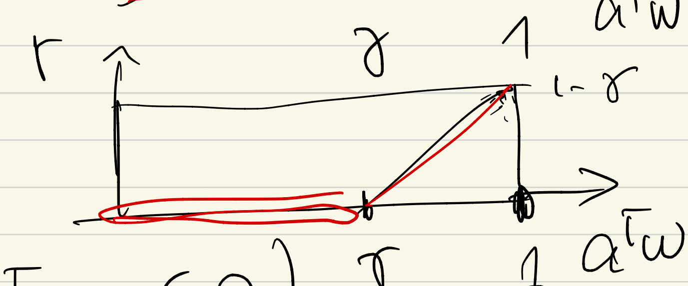
$$P_a^{(w)}(s, ds') = \delta_a(ds')\, p_w(a) + \delta_{s_1}(ds')(1 - p_w(a))$$

$$p_w(a) = \min\left(\frac{a^T w}{\gamma}, 1\right)$$



$\sigma = 1$





$$r_a^{(w,\sigma)}(s) = \sigma \max(a^T w - \gamma, 0)$$

$\sigma = 0$ $\quad q^*(s_0, t) = \frac{1}{2}$ $\cancel{\alpha}$

$q^*(\cancel{s_0}, a) = 0$,

$s \neq s_0$ $\quad q^*(s, a) = 0$

$\sigma = 1$ $\quad \pi^*(s) = w$ $\qquad s =$

$q^*(s_0, t) = \frac{1}{2}$

$q^*(s_1, a) = 0$

$s \neq s_0, s_1$ $\quad v^*(s) = 1$, $v^*(s_1) = 0$

$$q^*(s, a) = r_a(s) + \gamma p_w(a) v^*(a)$$

$$+ \gamma (1 - p_w(a)) v^*(s_1)$$

$s \neq s_1$

$$= \max(a^T w - \gamma, 0) + \gamma \min\left(\frac{a^T w}{\gamma}, 1\right)$$

$$= a^T w$$

$$\varphi(s, a) = \begin{bmatrix} a \\ \hline 0 \end{bmatrix} \mathbb{I}(a \in B^+) + \mathbb{I}(a = t) \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \hline 1 \end{bmatrix}$$

$s \neq s_1$

$$\varphi(s_1, a) = 0$$

$\sigma = 0$: $\quad \theta_{\sigma, w} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \hline 1/2 \end{bmatrix}$

$\sigma = 1$: $\quad \theta_{\sigma, w} = \begin{bmatrix} w \\ \\ \hline 1/2 \end{bmatrix}$

$$\Rightarrow \quad q^*_{(\sigma, w)}(s, a) = \varphi(s, a)^T \theta_{\sigma, w}$$