Function approx.

$$\Phi = \begin{bmatrix} \varphi^T(1) \\ \vdots \\ \varphi^T(S,A) \end{bmatrix} \in \mathbb{R}^{SA \times d}$$

(B2) $\quad \forall \pi \quad q^\pi \in \mathcal{F} = \text{span}(\Phi)$ $\qquad \varphi : S \times A \to \mathbb{R}^d$

$$\exists \theta \in \mathbb{R}^d : q^\pi = \Phi\theta \qquad [\forall (s,a) : \varphi(s,a)^T\theta = q^\pi(s,a)]$$

query + runtime indep. $\#S$

generative model

P.I.: $\quad \pi_0 \qquad k = 0, 1, 2, \ldots \quad : \pi_{k+1}$ greedy wrt q^{π_k}

$$q^{\pi_k} = \Phi\theta_k \quad , \quad \theta_k \in \mathbb{R}^d$$

$$d \ll \#S$$

extrapolation

Linear regression : $\quad \|\Phi\hat{\theta}_k - q^{\pi_k}\|_\infty \quad$ small

$\#S \times A \gg d$

KW-Theorem $\qquad \exists\, C \subseteq Z = S \times A \qquad \exists g : C \to [0,1]$, $\displaystyle\sum_{z' \in C} g(z') = 1$

1) $\quad |C| \le \dfrac{d(d+1)}{2}$

2.) $\quad \displaystyle\sup_{z \in Z} \|\varphi(z)\|_{G_g^{-1}} \le \sqrt{d}$

error blowup factor of least-squares.

$$G_g = \sum_{z \in C} g(z)\varphi(z)\varphi(z)^T$$

$S \times A$ finite

$\boxed{\varphi(S \times A) \subseteq \mathbb{R}^d \text{ compact}}$

$$\|u\|_P = (u^T P u)^{1/2}$$

$\boxed{\sqrt{d} \text{ unimprovable!}}$

Cor: $\forall \theta, \varepsilon: Z \to \mathbb{R}$

$$\hat{\theta} = G_g^{-1} \sum_{z' \in C} g(z') \underbrace{(\varphi(z')^T \theta + \varepsilon(z'))}_{\text{measured}} \varphi(z')$$

Then

$$\max_{z \in Z} |\varphi(z)^T \theta - \varphi(z)^T \hat{\theta}| \le \left(\max_{z' \in C} |\varepsilon(z')|\right) \sqrt{d}$$

Proof:

$$\hat{\theta} = \theta + \underbrace{G_g^{-1} \sum_{z' \in C} g(z') \varepsilon(z') \varphi(z')}$$

$z \in Z$ fixed

$$|\varphi(z)^T \hat{\theta} - \varphi(z)^T \theta| \le \sum_{z' \in C} |\varepsilon(z')| g(z') |\varphi(z)^T G_g^{-1} \varphi(z')|$$

$$\le \overbrace{\left(\max_{z' \in C} |\varepsilon(z')|\right)}^{\varepsilon_{apx}} \left(\sum_{z' \in C} g(z') |\varphi(z)^T G_g^{-1} \varphi(z')|\right)$$

$$\le \varepsilon_{apx} \left(\sum_{z' \in C} g(z') \underbrace{|\varphi(z)^T G_g^{-1} \varphi(z')|^2}\right)^{1/2}$$

$$= \varepsilon_{apx} \left(\sum_{z' \in C} g(z') \underbrace{\varphi(z)^T G_g^{-1} \varphi(z')} \underbrace{\varphi(z')^T G_g^{-1} \varphi(z)}\right)^{1/2}$$

$$= \varepsilon_{apx} \left(\varphi(z)^T G_g^{-1} \underbrace{\sum_{z' \in C} g(z') \varphi(z') \varphi(z')^T}_{G_g} G_g^{-1} \varphi(z)\right)^{1/2}$$

$$\boxed{\begin{aligned} &\left(\int f \, d\mu\right)^2 \le \int f^2 \, d\mu \\ &\text{Jensen's} \le \\ &\varphi\left(\int f \, d\mu\right) \le \int \varphi(f) \, d\mu \end{aligned}}$$

$$= \varepsilon_{apx} \Big(\underbrace{\varphi(z)^\top G_g^{-1} \varphi(z)}_{\left(\|\varphi(z)\|^2_{G_g^{-1}}\right)^{1/2}} \Big)^{1/2}$$

$$= \varepsilon_{apx} \cdot \underbrace{\|\varphi(z)\|_{G_g^{-1}}}_{\leq \sqrt{d}}$$

// Q.e.d.

(B2)

P.E. (Policy Evaluation)

relaxing

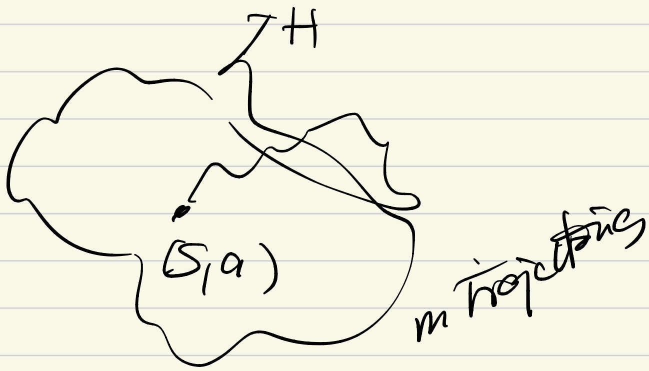
π $\qquad q^\pi = \Phi \theta + \varepsilon_\pi$

LS

$$\boxed{\hat\theta = G_g^{-1} \sum_{z' \in C} g(z') \hat R_m(z') \varphi(z')}$$

$$\hat R_m(s,a) = \frac{1}{m} \sum_{j=1}^{m} \underbrace{\sum_{t=0}^{H} \gamma^t r_{A_t^{(j)}}(S_t^{(j)})}_{\text{truncated return}}$$

$z = (s,a)$



$j = 1, \ldots, m$

$S_0^{(j)} = s$

$A_0^{(j)} = a$ $\qquad t \geq 1$

$S_t^{(j)} \sim P_{A_{t-1}^{(j)}}(S_{t-1}^{(j)})$

$A_t^{(j)} \sim \pi(\cdot \mid S_t^{(j)})$

$$|q^\pi(z) - \varphi(z)^T\hat{\theta}| = |\varphi(z)^T\theta + \varepsilon_\pi(z) - \varphi(z)^T\hat{\theta}|$$

$$\leq \underbrace{|\varphi(z)^T\theta - \varphi(z)^T\hat{\theta}|} + \underbrace{\max_{z'\in Z} |\varepsilon_\pi(z')|}_{\varepsilon_\pi^* \in \mathbb{R}}$$

$$\hat{R}_m(z) = \varphi(z)^T\theta + \underbrace{\hat{R}_m(z) - \varphi(z)^T\theta}_{\varepsilon(z)}$$

Using Corollary:

$$|\varphi(z)^T\theta - \varphi(z)^T\hat{\theta}| \leq \max_{z\in C} |\varepsilon(z)| \sqrt{d}$$

sampling error

$$\varepsilon(z) = \underbrace{\hat{R}_m(z) - (T_\pi^H 0)(z)}_{\mathrm{I}} + \underbrace{(T_\pi^H 0)(z) - q^\pi(z)}_{\mathrm{II}} - \underbrace{\varepsilon_\pi(z)}_{\mathrm{III}}$$

$\mathbb{E}[\ ] = (T_\pi^H 0)(z)$

$q^\pi(z) = \varphi(z)^T\theta + \varepsilon_\pi(z)$

$\leq \gamma^H / 1-\gamma$

Hoeffding's $\leq$: $|\hat{R}_m(z) - (T_\pi^H 0)(z)|$

$\forall z \in C$: wp $1-\delta$ $\leq \frac{1}{1-\gamma}\sqrt{\frac{\log(2/\delta)}{2m}}$

+ Union bound $z \in C$

$\delta \to \delta/|C|$

wp $1-\delta$ $\forall z \in C$:

$$\left|\hat{R}_m(z) - (T_\pi^H 0)(z)\right| \le \frac{1}{1-\gamma}\sqrt{\frac{\log\left(\frac{2|C|}{\delta}\right)}{2m}}$$

wp $1-\delta$: $\forall z$:

$$|\varepsilon(z)| \le \frac{1}{1-\gamma}\sqrt{\cdots} + \frac{\gamma^H}{1-\gamma} + \varepsilon_\pi^*$$

Lemma: wp $1-\delta$

$H \to \infty$ $\quad m \to \infty$ $\quad 0$

$$\max_{z \in Z} \left|q^\pi(z) - \varphi(z)^\top \hat{\theta}\right| \le \varepsilon_\pi^* + \sqrt{d}\left(\varepsilon_\pi^* + \frac{\gamma^H}{1-\gamma} + \frac{1}{1-\gamma}\sqrt{\frac{\cdots}{2m}}\right)$$

Notice $|C| = O(d^2)$

$\log |C| = O(\log d)$

$\boxed{\text{not depends on } \#S}$

$\varepsilon_\pi^* (1 + 3\sqrt{d})$

$\frac{\gamma^H}{1-\gamma} \le \varepsilon_{approx}$

$H = H_{\gamma, \varepsilon_{approx}}$

$$\frac{1}{1-\gamma}\sqrt{\frac{\log\left(\frac{2|C|}{\delta}\right)}{2m}} \le \varepsilon_{approx}$$

d^2

$$m \ge \left(\frac{1}{(1-\gamma)\varepsilon_{approx}}\right)^2 \log\left(\frac{2|C|}{\delta}\right)$$

π_0 $\quad \cancel{q^{\pi_0}}$ $\quad \theta_0$ $\quad q_0 = \Phi\theta_0 = q^{\pi_0} + \varepsilon_0$

π_1 greedy w.r.t. q_0 (not q^{π_0}!)

$\vdots$

π_k — " — $\quad q_k = \Phi\theta_k = q^{\pi_k} + \varepsilon_k$

$v^{\pi_k} \geq v^* - ??$ 1

k large enoug

$\boxed{\max_{1 \leq i \leq k} \|\varepsilon_i\|_\infty \leq \varepsilon}$

<u>Progress Lemma</u> A.P.I. $\quad \pi$ ML policy

$\geq T_{\pi^*} q \quad q = q^\pi + \varepsilon \quad \varepsilon : S \times A \to \mathbb{R}$

π' : $\boxed{T_{\pi'} q = T q}$ (π' greedy w.r.t. q)

$\left[M_{\pi'} q = M q : \quad \sum_{a \in A} \pi'(a|s) q(s,a) = \max q(s,a) \right] \quad \forall s$

$$\boxed{\|q^* - q^{\pi'}\|_\infty \leq \gamma \|q^* - q^\pi\|_\infty + \frac{2\|\varepsilon\|_\infty}{1-\gamma}}$$

<u>Proof</u>: $\quad \pi_*$ opt. ML policy $\quad T_{\pi^*} q^* = q^*$

$$q^* - q^{\pi'} = T_{\pi^*} q^* - T_{\pi^*} q^\pi + T_{\pi^*} q^\pi - T_{\pi^*} q + \underbrace{T_{\pi^*} q - T_{\pi'} q}_{\leq 0}$$

$$+ T_{\pi'} q - T_{\pi'} q^\pi + T_{\pi'} q^\pi - T_{\pi'} q^{\pi'}$$

$T_\pi q = r + \gamma \overset{P_\pi}{P M_\pi} q$

$$\le \gamma P_{\pi^*}(q^* - q^\pi) + \gamma P_{\pi^*}(q^\pi - q)$$

$$+ \gamma P_{\pi'}(q - q^\pi) + \gamma P_{\pi'}(q^\pi - q^{\pi'})$$

$$= \gamma P_{\pi^*}(q^* - q^\pi) + \gamma \left[(P_{\pi'} - P_{\pi^*})\varepsilon + P_{\pi'} \underbrace{(q^\pi - q^{\pi'})}_{\le A\varepsilon} \right]$$

$$q^\pi - \underbrace{q^{\pi'}}_{\substack{\| \\ (I-\gamma P_{\pi'})^{-1} r}} = (I - \gamma P_{\pi'})^{-1} \Big[\overbrace{\underbrace{(I - \gamma P_{\pi'}) q^\pi - r}_{q^\pi = q - \varepsilon}}^{I.} \Big]$$

$$\underbrace{\phantom{(I - \gamma P_{\pi'}) q^\pi - r}}_{\gamma(P_{\pi'} - P_\pi)\varepsilon}$$

$$I. = (I - \gamma P_{\pi'})(q - \varepsilon) - r$$

$$= q - (\underbrace{\gamma P_{\pi'} q + r}) - (I - \gamma P_{\pi'})\varepsilon$$

$$= q - T_{\pi'} q - (I - \gamma P_{\pi'})\varepsilon$$

$$q = q^\pi + \varepsilon \ / \ T_{\pi'}$$

$$T_{\pi'} q \ge T_\pi q = \underbrace{T_\pi q^\pi}_{q^\pi} + \gamma P_\pi \varepsilon$$

$$q - T_{\pi'} q \le (I - \gamma P_\pi)\varepsilon$$

$$\le (I - \gamma P_\pi)\varepsilon - (I - \gamma P_{\pi'})\varepsilon$$

$$= \gamma (P_{\pi'} - P_\pi)\varepsilon.$$