Feb 9

① API: Finish! ② Lower bounds

$\pi_0 \qquad \theta_0 \qquad q_0 = \Phi\theta_0 = q^{\pi_0} + \varepsilon_0$

π_1 greedy w.r.t. q_0 (not q^{π_0}!)

$\vdots$

$\pi_k \qquad$ — " — $\qquad q_k = \Phi\theta_k = q^{\pi_{k-1}} + \varepsilon_{k-1}$

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

k large enough

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

1. Policy evaluation
2. Error propagation

Feb 9

$H \approx H_{\gamma, \varepsilon_{apx}}$, $m \approx H^2_{\gamma,\varepsilon_{apx}} \log\left(\frac{2d^2}{\delta}\right)$

Policy eval.

Lemma: wp $1-\delta$

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

$$\max_{z \in Z} \left| q^\pi(z) - \varphi(z)^T \theta \right| \le \varepsilon^*_\pi + \sqrt{d}\left(\varepsilon^*_\pi + \frac{\gamma^H}{1-\gamma} + \frac{1}{1-\gamma}\sqrt{\frac{\log(2|C|/\delta)}{2m}}\right)$$

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

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

note depends on $\#S$

Progress Lemma A.P.I. $\forall \pi$ ML policy

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

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

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

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

$\delta_k = \| q^* - q^{\pi_k} \|_\infty$

$\delta_{k+1} \le \gamma \delta_k + \varepsilon/_{1-\gamma}$ $\Rightarrow$ $\boxed{\delta_k \le \gamma^k \delta_0 + \frac{\varepsilon}{(1-\gamma)^2}}$

$\le \gamma\left(\gamma \delta_{k-1} + \frac{\varepsilon}{1-\gamma}\right) + \frac{\varepsilon}{1-\gamma} = \gamma^2 \delta_{k-1} + \frac{\varepsilon(1+\gamma)}{1-\gamma}$

Policy error bound: $\quad q : S \times A \to \mathbb{R}$

π greedy wrt q

$$v^{\pi} \geq v^* - \frac{2\|q-q^*\|_\infty}{1-\gamma}$$

Putting it all together!

$$\pi_0 \to \theta_0 \to \pi_1 \to \theta_1 \to \cdots \quad \theta_k \to \pi_k$$

$$v^{\pi_k} \geq v^* - \frac{2\|q_k-q^*\|_\infty}{1-\gamma}$$

$$\boxed{\text{wp } 1-\delta: \quad v^{\pi_k} \geq v^* - 6\,\frac{\varepsilon_{apx}(1+3\sqrt{d})}{(1-\gamma)^3}}$$

$$\|q_k - q^*\|_\infty \leq \underbrace{\|q^{\pi_k} - q^*\|_\infty}_{\leq \underbrace{\frac{\gamma^k}{1-\gamma}}_{} + \underbrace{\frac{\varepsilon_{apx}(1+3\sqrt{d})}{1-\gamma^2}}_{}} + \underbrace{\|q^{\pi_k} - q_k\|_\infty}_{\varepsilon_{apx}(1+3\sqrt{d})}$$

$r_a(s) \in [0,1]$

$$k = H_{\gamma,(1-\gamma)\varepsilon_{apx}(1+3\sqrt{d})}$$

$$\leq \frac{3\cancel{2}\,\varepsilon_{apx}(1+3\sqrt{d})}{1-\gamma^2}$$

$$m = H^2_{\gamma,(1-\gamma)\varepsilon_{apx}} \log\left(\frac{2d^2}{\delta/H_{\gamma,(1-\gamma)\dots}}\right)$$

$$= H^2_{\gamma,(1-\gamma)\varepsilon_{apx}} \log\left(\frac{2d^2 H_{\gamma,\dots}}{\delta}\right)$$

$0 \le \delta \le 1$

Thm:

$k = H_{\gamma, (1-\gamma)\varepsilon_{\cancel{apx}}} (1 + 3\sqrt{d})$

$m = H^2_{\gamma,(1-\gamma)\varepsilon_{\cancel{apx}}} \log\left(\frac{2d^2 H_{\gamma,} \ldots}{\delta}\right)$

$H = \underbrace{H_{\gamma, \varepsilon_{\cancel{apx}}}} \simeq \frac{\log\left(1/(1-\gamma)\varepsilon_{\cancel{apx}}\right)}{1-\gamma}$

$d^2 \times m \times k \times H$

$\varepsilon_{apx} = \sup_{\pi} \inf_{\theta} \| \Phi\theta - q^{\pi} \|_{\infty}$

$\Rightarrow$ w.p. $1 - \delta$,

$v^{\pi_\varepsilon} \ge v^* - \boxed{\frac{6\,\varepsilon_{apx}(1 + 3\sqrt{d})}{(1-\gamma)^3}}$

$6(\varepsilon_{apx}(1+\sqrt{d}) + 2\varepsilon\sqrt{d}) / (1-\gamma)^3$

#Query
#Compute $\Big\}$ $\mathrm{poly}(d, \log(1/\varepsilon), H_{\gamma,\varepsilon}, \log(1/\delta))$

Caveat: [C how do we get it ??]

Necessary!

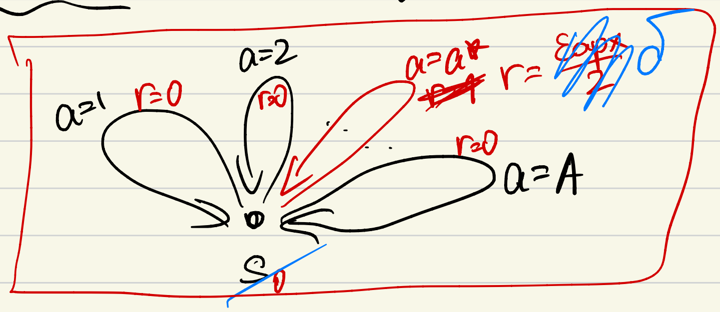
$\boxed{\frac{\varepsilon_{apx}(1+\sqrt{d})}{(1-\gamma)^2}}$: $\Leftarrow$ one that necessary

Du-Kakade-Wang-Yang ~~2020~~ 2019

$\varepsilon_{apx}\sqrt{d}$: $\sqrt{d}$ is necessary if #queries must be $\mathrm{poly}(d, \dots)$

Bandit $\quad \gamma = 0 \qquad A$ actions

$A \gg d$



$r_a^{a^*} = \frac{\varepsilon_{apx}}{2} \cdot \mathbb{I}(a = a^*)$

$\sup_\pi \inf_\theta \| \Phi\theta - q^\pi \|_\infty \le \varepsilon_{apx}$ $\qquad \exists ? \Phi$

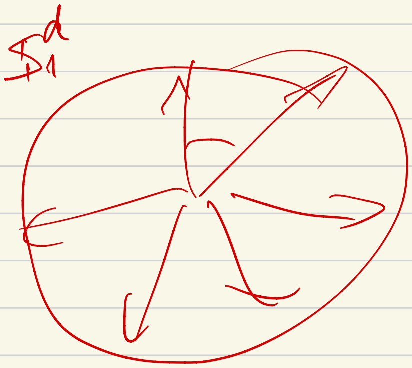
$q^\pi(s_0, a) = r_a(\cancel{s})$

(C) $\boxed{\max_{a^* \in [A]} \inf_\theta \max_{a \in [A]} | \varphi(a)^\top \theta - r_a^{a^*} | \le \varepsilon_{apx}}$

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

$\delta < \varepsilon_{apx}/2 \implies$ Alg needs to find a^*!

$\implies$ Alg needs $\Omega(A)$ queries.



J-L Lemma:

$\forall \delta > 0,\ d, k \quad \boxed{\left\lceil \frac{8 \ln k}{\delta^2} \right\rceil \le d \le k}$

$\implies \exists \varphi_1, \dots, \varphi_k \in S_1^d$ s.t.

$\forall i \ne j \qquad |\langle \varphi_i, \varphi_j \rangle| \le \delta$

$k \le \exp\left(\frac{\delta^2}{8} d\right)$

$$\boxed{A = \left\lfloor \exp\left(\frac{\delta^2}{8} d\right) \right\rfloor} \; ; \quad \delta = \varepsilon_{apx}$$

$$\overset{JL}{\Rightarrow} \quad \exists\, \varphi(1), \dots, \varphi(A) \in S_1^d \quad \text{s.t.}$$

$$\max_{i \neq j} |\langle \varphi(i), \varphi(j) \rangle| \leq \varepsilon_{apx}$$

Choose $a^* \in [A]$; $\quad \theta_{a^*} = \frac{\varepsilon_{apx}}{2} \varphi(a^*)$.

Does (C) hold?

$$\varepsilon_{apx} \geq |\varphi(a)^T \theta_{a^*} - r_a^{a^*}| = |\underbrace{\langle \varphi(a), \varphi(a^*) \rangle}_{} \frac{\varepsilon_{apx}}{2} - r_a^{a^*}|$$

$$= \begin{cases} |1 \frac{\varepsilon_{apx}}{2} - \frac{\varepsilon_{apx}}{2}| = 0, & a = a^* \\ |\frac{\varepsilon_{apx}^2}{2} - 0| & a \neq a^* \end{cases}$$

$0 \leq \varepsilon_{apx} \leq 1 \quad \Rightarrow \quad$ (C) does hold!

$\Rightarrow \quad \forall\, \mathcal{A}$ alg. $\quad \delta$-subopt. $\quad \delta \leq \frac{\varepsilon_{apx}}{2}$, d fixed

$\exists\, \Phi \in \mathbb{R}^{A \times d}$, $\quad A = \lfloor \exp(c \cdot \varepsilon_{apx}^2 d) \rfloor$

s.t. $\mathcal{A}$ makes $\Omega(A)$ queries.

$$A = \left\lfloor \exp\left(c \left(\frac{\varepsilon_{apx}}{\delta}\right)^2 d\right) \right\rfloor$$

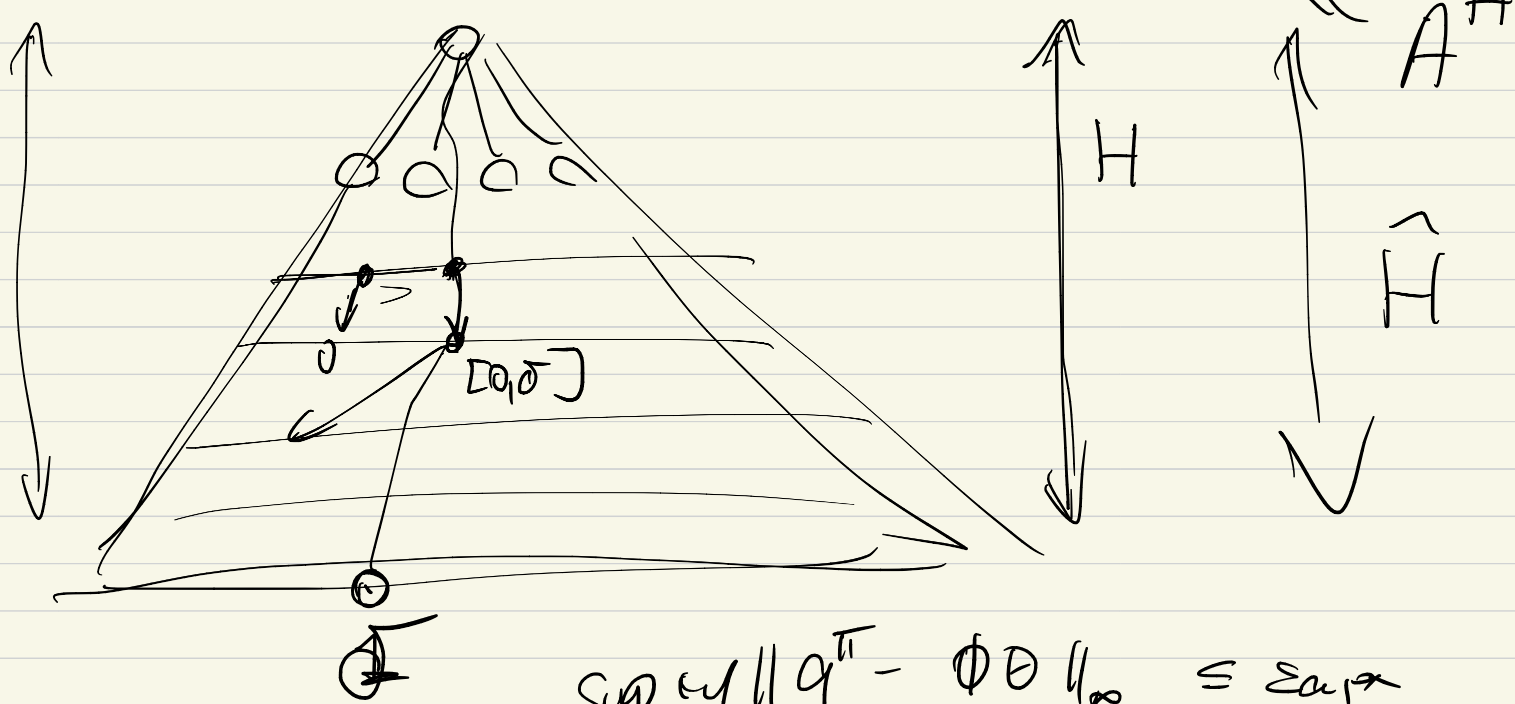
$\gamma = 0$ $\qquad$ $\gamma > 0$

fixed-horizon

$$\circ \to \circ \to \circ \to \circ \cdots \circ \to \circ$$
$$1 \quad 2 \quad \cdots \quad H$$

Thm: $\Omega\left(\min\left(A^H, \exp\left(c \cdot \left(\frac{\varepsilon_{apx}}{\delta}\right)^2 d\right)\right)\right)$

$= A^{\hat{H}}$



$$\sup_{\pi} \inf_{\theta} \| q^{\pi} - \Phi\theta \|_{\infty} \le \varepsilon_{apx}$$

$U \sim S_1^d$ uniform $\quad \mathbb{P}\left(U_1^2 \ge \beta/d\right) \le \exp\left(-\frac{\beta}{4}\right)$

$\beta \ge 6$

$\boxed{\text{J-L}}$ Probabilistic method $\quad \varphi_1, \ldots, \varphi_k$ Unif.

$$1 \overset{?}{>} \mathbb{P}\left(\max_{i \ne j} |\langle \varphi_i, \varphi_j \rangle| \ge \delta\right) \le \frac{d(d-1)}{2} \mathbb{P}\left(|\langle \varphi_1, \varphi_2 \rangle| \ge \delta\right)$$

< 1