$\Delta : \mathcal{X} \times \Theta \to \mathbb{R}$ — discriminat fn.

Def 1: $x_{1:n} \in X^n$ is $\varepsilon_1/\varepsilon_2/\ell^\infty$ - eluder sequence if

$\forall i \in [n] \quad \exists \theta \in \Theta$ s.t. $|\Delta(x_i, \theta)| > \varepsilon_1$ & $\max_{j \le i-1} |\Delta(x_j, \theta)| \le \varepsilon_2$

Set of these: $S_\infty(\varepsilon_1, \varepsilon_2)$.

Def 2: $x_{1:n} \in \mathcal{X}^n$ is $\varepsilon_1/\varepsilon_2/\ell^2$ - eluder seq. if

$\forall i \in [n] \quad \exists \theta \in \Theta$ s.t. $|\Delta(x_i, \theta)| > \varepsilon_1$ & $\sum_{j=1}^{i-1} \Delta^2(x_j, \theta) \le \varepsilon_2^2$

Set of these: $S_2(\varepsilon_1, \varepsilon_2)$.

---

1. S_∞, S_2 are decreasing in ε_1, increasing in ε_2.
2. $S_\infty(\varepsilon_1, \varepsilon_2) \cap X^n \subseteq S_2(\varepsilon_1, \sqrt{n-1}\,\varepsilon_2)$, $\forall \varepsilon_1, \varepsilon_2, n$. [$\theta_1 \dots \theta_{n-1}$ witness for ℓ^∞ $\Rightarrow \sum_{j=1}^{i-1} \Delta^2(x_j, \theta_i) \le (i-1)\varepsilon_2^2 \le (n-1)\varepsilon_2^2$.

Claim: $\max\{n \mid S_\infty(\varepsilon, \frac{\varepsilon}{\sqrt{d}}) \cap X^n \neq \emptyset\} \le \max\{n \mid S_2(\varepsilon, \varepsilon) \cap X^n \neq \emptyset\} =: d$.

Proof: It suffices if $S_\infty(\varepsilon, \frac{\varepsilon}{\sqrt{d}}) \cap X^{d+1} = \emptyset$. And for this: $S_\infty(\varepsilon, \frac{\varepsilon}{\sqrt{d}}) \cap X^{d+1} \subseteq S_2(\varepsilon, \varepsilon) \cap X^{d+1} = \emptyset$ // Q.e.d.

We can directly bound $\max\{n \mid S_\infty(\varepsilon_1, \varepsilon_2) \cap \mathcal{X}^n \neq \emptyset\}$

when $\Delta(x, \theta) = x^\top \theta$, $\mathcal{X} = B_2^d(r)$, $\Theta = B_2^d(S)$.

How? Usual proof: Let $x_{1:n} \in S_\infty(\varepsilon_1, \varepsilon_2)$. Take $i \in [n]$.

$$\Rightarrow \varepsilon_1 < \max\left\{ |x_i^\top \theta| : \max_{1 \le j \le i-1} |x_j^\top \theta| \le \varepsilon_2, \ \|\theta\|_2 \le S \right\}$$

$$\le \max\left\{ |x_i^\top \theta| : \sum_{j=1}^{i-1} |x_j^\top \theta|^2 \le (i-1)\varepsilon_2^2, \ \|\theta\|_2^2 \le S^2 \right\}$$

$$= \max\left\{ |x_i^\top \theta| : \theta^\top (X_{i-1} + \lambda I)\theta \le i\,\varepsilon_2^2 \right\}$$

$X_i = \sum_{i \le i} x_j x_j^\top$

$\lambda = \varepsilon_2^2 / S^2$

$V_i = X_i + \lambda I$

$$= \sqrt{i}\,\varepsilon_2 \|x_i\|_{V_{i-1}^{-1}} \Rightarrow \|x_i\|_{V_{i-1}^{-1}} > \frac{\varepsilon_1}{\sqrt{i}\,\varepsilon_2}.$$

$$\left(\frac{\lambda d + i r^2}{d}\right)^d = \left(\frac{\operatorname{tr} V_i}{d}\right)^d \ge \det V_i = \lambda^d \prod_{j=1}^{i} \left(1 + \|x_j\|_{V_{j-1}^{-1}}\right) = \lambda^d \prod_{j=1}^{i}\left(1 + \frac{\varepsilon_1}{\sqrt{j}\,\varepsilon_2}\right)$$

/ log

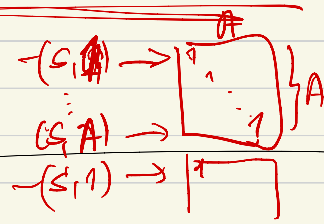
$$\Rightarrow d \log\left(1 + \frac{i r^2}{d\lambda}\right) \ge \sum_{j=1}^{i} \log\left(1 + \frac{\varepsilon_1}{\sqrt{j}\,\varepsilon_2}\right) \approx \frac{\varepsilon_1}{\varepsilon_2} \sum_{j=1}^{i} \frac{1}{\sqrt{j}} \approx \frac{\varepsilon_1}{\varepsilon_2}\sqrt{i}$$

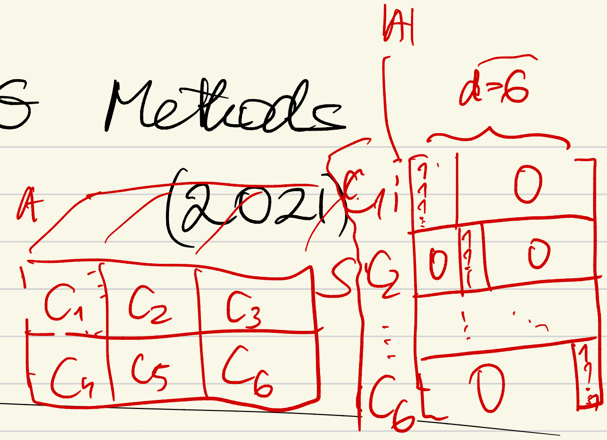
; solve for longest i

$$\Rightarrow \boxed{i_{max} = O(d^2 \log^2(\ldots))}$$

Approximation Benefits of PG Methods with Aggregated States (2021)

Dan Russo





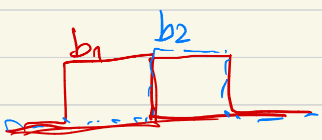
Lower bound for API

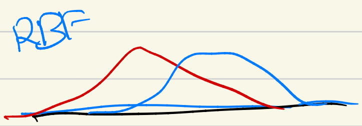
$$\tilde{\varepsilon}_{apx} = \sup_{\pi \in \Pi_\varphi} \inf_\theta \|\phi\theta - q^\pi\|_\infty$$

Theorem: $\forall \gamma \in [0,1)$, $\forall \varepsilon_{apx} > 0$ $\exists$ MDP $M = (S, A, P, r, \gamma)$, $\exists \Phi$

$\exists \pi_1$ policy of M; initial API policy.

$\exists \xi \in M_1(S)$



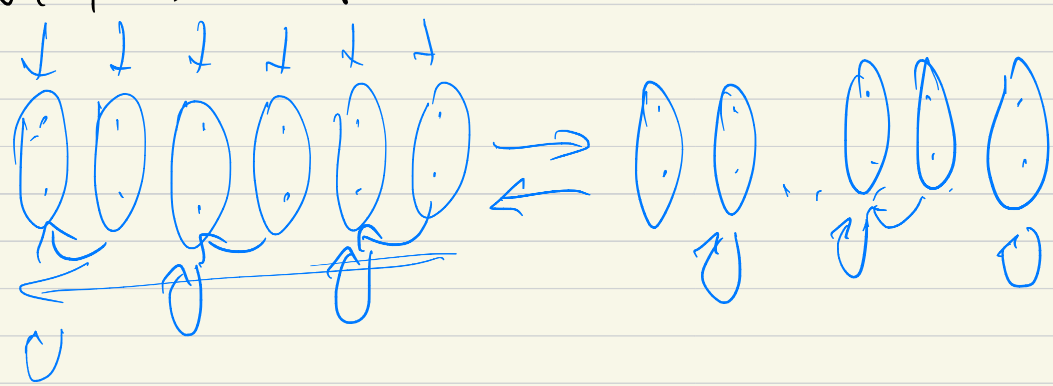


$\rightsquigarrow \pi_1, \pi_2, \pi_3, \dots$

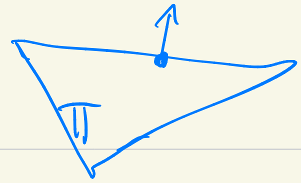
s.t.
$$\inf_{t \geq 1} v^*(\xi) - v^{\pi_t}(\xi) \geq \frac{c \cdot \tilde{\varepsilon}_{apx}}{\boxed{(1-\gamma)^2}}$$

Policy eval:
$$\hat{q}_t = \operatorname*{argmin}_{q \in \mathcal{F}_\varphi} \sum_{\substack{s \in S \\ a \in A}} w(s,a)\left(q(s,a) - q^{\pi_t}(s,a)\right)^2$$

$w(s,a) =$ uniform



$J(\pi) \doteq v^{\pi}(\rho) \longrightarrow \max$



Thm: π^{∞} stat. point of $J(\pi)$

$$\frac{1}{1-\gamma} \leftarrow v^*(\rho) - v^{\pi}(\rho) \le \frac{4\,\varepsilon_{apx}}{\boxed{1-\gamma}}$$

$$\Updownarrow$$

$$(1-\gamma)\left(v^*(\rho) - v^{\pi}(\rho)\right) \le 4\,\varepsilon_{apx}$$

$$\tilde{v}^*(\rho) - \tilde{v}^{\pi}(\rho) \le \underline{4\,\varepsilon_{apx}}$$

POLITEX Policy Iteration with Expert Advice

$\pi_1,\ \pi_2,\ \pi_3,\ \dots,\ \pi_n$

$\downarrow\ \ \downarrow\ \ \downarrow$

$\hat{q}_1\ \ \hat{q}_2\ \ \hat{q}_3$

$\hat{\theta}_1 \quad \hat{q}_1 \approx q^{\pi_1} \quad \hat{\theta}_2 \quad \hat{\theta}_3$

$$\bar{\theta}_{k-1}^{T}\varphi(s,a)$$

$$\pi_k(a|s) \propto \exp\left(\eta \sum_{j=1}^{k-1} \hat{q}_j(s,a)\right) = E_k(s,a)$$

$$\pi_k(a|s) = \frac{E_k(s,a)}{\sum_{a'} E_k(s,a')}$$

LSPE G-optimal design

m rollouts, H length,

$$\bar{\theta}_{k-1} = \sum_{j=1}^{k-1} \hat{\theta}_j$$

$$\hat{q}_k = \Phi \hat{\theta}_k$$

$$\sum \hat{q}_j = \Phi \underline{\sum_j \hat{\theta}_j}$$

Why does it work?

$\pi_1, \dots, \pi_n$ $\quad \frac{1}{n}(\pi_1 + \dots + \pi_n)$

$K \sim \mathrm{Unif}([n])$

$A = \pi_K(S_0)$

Politex in plain

π : policy induced by Politex. $\quad R = \sum_{t=0}^{\infty} \gamma^t r_{A_t}(S_t)$

$$v^\pi(s) = \mathbb{E}_s^\pi[R] = \frac{1}{n}\sum_{k=1}^n \mathbb{E}_s^{\pi_k}[R] = \frac{1}{n}\sum_{k=1}^n v^{\pi_k}(s)$$

$$\mathbb{P}_s^\pi = \frac{1}{n}\sum_{k=1}^n \mathbb{P}_s^{\pi_k}$$

$$\frac{1}{n}\sum_{k=1}^n v^{\pi^*} - v^{\pi_k}$$

$$\boxed{v^{\pi^*} - v^{\pi_k} = (I-\gamma P_{\pi^*})^{-1}\left[T_{\pi^*} v^{\pi_k} - v^{\pi_k}\right]}$$

$$= \frac{1}{n}(I-\gamma P_{\pi^*})^{-1}\sum_{k=1}^n \left[T_{\pi^*} v^{\pi_k} - v^{\pi_k}\right]$$

$$(M_\pi q)(s) = \sum_a \pi(a|s) q(s,a)$$

$$r_\pi = M_\pi r$$

$$P_\pi = M_\pi P$$

$$T_{\pi^*} v^{\pi_k} = r_{\pi^*} + \gamma P_{\pi^*} v^{\pi_k}$$

$$= M_{\pi^*}(r + \gamma P v^{\pi_k})$$

$$= M_{\pi^*} q^{\pi_k}$$

$$v^{\pi_k} = M_{\pi_k} q^{\pi_k}$$

$$\frac{\varepsilon_{approx}\sqrt{d}}{1-\gamma}$$

$$\varepsilon_{approx}\sqrt{d} + \dots$$

$$= \frac{1}{n}(I-\gamma P_*)^{-1}\sum_{k=1}^n M_{\pi^*} q^{\pi_k} - M_{\pi_k} q^{\pi_k}$$

$$= \frac{1}{n}(I-\gamma P_*)^{-1}\sum_{k=1}^n \underbrace{M_{\pi^*}\hat{q}_k - M_{\pi_k}\hat{q}_k}_{\mathrm{I.}\ \to\ ??} + \overbrace{\frac{1}{n}(I-\gamma P_{\pi^*})^{-1}\sum_{k=1}^n (M_{\pi^*} - M_{\pi_k}) \times (q^{\pi_k} - \hat{q}_k)}^{\mathrm{II.}}$$