Linear MDP A linear MDP $M = (S, A, P, R, H, d_0)$

- satisfies: $P(s'|s,a) = \phi(s,a)^T \psi(s')$ ($\phi \in \mathbb{R}^d$, $\psi \in \mathbb{R}^d$) ($d \ll |S \times A|$).
  $R(s,a) = \phi(s,a)^T \theta_R$. "low-rank MDP".
- learner knows $\phi: S \times A \to \mathbb{R}^d$.

$R_{max} = 1$.

Key property: $\forall f: S \times A \to \mathbb{R}$, $Tf, T^\pi f \in \mathcal{F}$. ($\forall \pi$)

$$\mathcal{F} := \{ (s,a) \mapsto \phi(s,a)^T \theta : \theta \in \mathbb{R}^d \}.$$

$$(Tf)(s,a) = R(s,a) + \langle P(s,a), V_f \rangle$$
$$= \phi^T(s,a)\,\theta_R + \phi(s,a)^T (\Psi\, V_f)$$
$$= \phi^T(s,a)\,(\underbrace{\theta_R + \Psi V_f}_{\theta_f})$$

$\Rightarrow Q^*, Q^\pi \in \mathcal{F}$.

Known vs. unknown. $\in \mathbb{R}^{d \times d}$.

$\forall h \in \{1, 2, \cdots, H\}$, $\Lambda_h = \mathbb{E}_{D_h}[\phi(s_h, a_h)\phi(s_h, a_h)^T]$. ($\in \mathbb{R}^d$)

$\{ s_1, a_1, r_1, \boxed{s_2, a_2} r_2, \cdots, s_H, a_H, r_H$

"known" (s_h, a_h): when $\phi(s_h,a_h)^\top \Lambda_h^{-1} \phi(s_h,a_h)$ is small

"unknown" (s_h, a_h): if big.

Special case: tabular: $\phi(s,a)$ is 1-hot.

Λ_h is diagonal with $\Lambda_{h_{(s,a),(s,a)}} = \frac{n(s,a)}{n}$.

Alg: at each episode.

$f_{H+1} \equiv 0$.

for $h = H, H-1, H-2, \dots, 1$. "Optimistic FQI"

$$\hat{f}_h = \operatorname*{argmin}_{f \in \mathcal{F}} \mathbb{E}_{\mathcal{D}_h}\left[\left(f(s,a) - r - V_{f_{h+1}}(s')\right)^2\right]$$

$$f_h(s,a) = \begin{cases} \hat{f}_h(s,a), & \text{if } (s,a) \in K. \\ V_{\max}, & \text{if } (s,a) \notin K \end{cases}$$

↳ non-smooth.

$f := f_1 \circ f_2 \circ f_3 \circ \dots \circ f_H$.

explore w/ $\pi_f := \pi_{f_1} \circ \pi_{f_2} \circ \dots \circ \pi_{f_H}$.

(collect a batch sample).

Simplification: batch size $\to \infty$. (only polynomial).

Want to bound iteration complexity.

$$\Rightarrow \begin{cases} (s,a) \in K: & \phi(s,a)^T \Lambda_h^{-1}(s,a)\, \phi(s,a) < +\infty. \\ (s,a) \notin K: & \phi(s,a)^T \Lambda_h^{-1}(s,a)\, \phi(s,a) = +\infty. \end{cases}$$

will show: in each round, $\exists h$, s.t. $\operatorname{rank}(\Lambda_h)$ inc by at least 1.

$V_{max} = (H-h+1)\cdot R_{max}$

Lemma. (Optimism): $f \geq Q^*$.

At H: $f_H(s,a) = \begin{cases} R(s,a) = Q_H^*(s,a). & \forall (s,a) \in K \\ V_{max} \geq Q_H^*(s,a). \end{cases}$

For $h < H$, by induction $f_{h+1} \geq Q_{h+1}^*$.

$$\Rightarrow V_{f_{h+1}} \geq V_{h+1}^*.$$

$$f_h(s,a) = \begin{cases} \mathbb{E}[r + V_{f_{h+1}}(s') \mid s,a] \geq \mathbb{E}[r + V_{h+1}^*(s') \mid s,a] \\ \qquad = Q_h^*(s,a). \\ V_{max} \geq Q_h^*(s,a) \end{cases}$$

↓ rely on

* $\mathbb{E}[\cdots \mid s,a] = (\mathcal{T} f_{h+1})(s,a)$
* $\forall f,\ \mathcal{T} f \in \mathcal{F}$.

"Optimal - or - explore".

$$\varepsilon < J(\pi^*) - J(\pi_f) = \mathbb{E}_{s\sim d_0}\left[\max_a Q^*(s,a)\right] - J(\pi_f).$$

$$\le \mathbb{E}_{s\sim d_0}\left[\max_a f_1(s,a)\right] - J(\pi_f).$$

$$= \mathbb{E}_{s\sim d_0}\left[f(s,\pi_f)\right] - J(\pi_f).$$

$f(s,\pi_f)$: viewed as prod of $J(\pi_f)$

$$= \sum_{h=1}^{H} \mathbb{E}_{\pi_f}\left[f(s_h,a_h) - r_h - V_f(s_{h+1})\right].$$

$$\Rightarrow \exists h,\ \frac{\varepsilon}{H} \le \mathbb{E}_{\pi_f}\left[f(s_h,a_h) - r_h - V_f(s_{h+1})\right]$$

$f \gtrsim \tilde{f}$

$$\le \mathbb{E}_{\pi_f}\left[f(s_h,a_h) - r_h - V_{\tilde{f}}(s_{h+1})\right].$$

$$\le \left|\mathbb{E}_{\pi_f}\left[f(s_h,a_h) - \mathbb{E}\left[r_h + V_f(s_{h+1})\middle|s_h,a_h\right]\right]\right|$$

$$\le \sqrt{\mathbb{E}_{\pi_f}\left[\left(f(s_h,a_h) - \mathbb{E}\left[r_h + V_f(s_{h+1})\middle|s_h,a_h\right]\right)^2\right]}$$

$$(*)$$

$$\mathbb{E}_{\pi_f}[\cdots] = \sum_{(s,a)} d_h^{\pi_f}(s,a) \cdot (\cdots).$$

$$= \sum_{(s,a)\in K} d_h^{\pi_f}(s,a)(\cdots) \quad \leftarrow \text{w.t.s. zero.}$$

$$+ \sum_{(s,a)\notin K} d_h^{\pi_f}(s,a)(\cdots).$$

$$\sum_{(s,a)\in K} d_h^{\pi_f}(s,a)\left(f(s,a) - \mathbb{E}[r + V_f(s') \mid s,a]\right)^2$$

$\forall (s,a) \in K.$

$$\Downarrow$$

$$f_h(s,a) = \tilde{f}_h(s,a) = \mathbb{E}[r + V_f(s') \mid s,a]$$

needs to be shown.

Linear regression. design $X \in \mathbb{R}^{n \times d}$ output $Y^{n \times 1}$.

$x^T \left((X^T X)^{-1} X^T Y \right)$

with $y = x^T \theta + \varepsilon$

ε: zero mean.

$x^T \left((X^T X)^{\dagger} X^T Y \right)$

$x_0 \in \mathbb{R}^{d \times 1}$.

$y_0 = x_0^T \theta$.

$a x_0^T \left(x_0 x_0^T \right)^{\dagger} x_0^T y_0$.

$= a y_0$.

$$\sum_{(s,a) \notin K} d^{\pi_f}(s,a) \underbrace{(\cdots)}_{\leq O(V_{max})} \geq \frac{\varepsilon}{H}.$$

$$\Rightarrow \sum_{s,a \notin K} d^{\pi_f}(s,a) \geq \frac{\varepsilon}{H V_{max}} > 0.$$

If we explore w/ π_f.

will collect data from $(s,a) \notin K$.

$$\Lambda_h' = \alpha \Lambda_h + (1-\alpha) \underbrace{\mathbb{E}_{\pi_f}[\phi(s,a)\phi(s,a)^T]}$$

Suppose $x^T \Lambda^{-1} x = +\infty$. ($\Lambda^{-1}$: $d \times d$.)

w.t.s. that $\operatorname{rank}(\Lambda + xx^T) > \operatorname{rank}(\Lambda)$

$\phi(s,a) \notin K$.

$\Downarrow$

$\phi(s,a)^T \Lambda_h^{-1} \phi(s,a) = +\infty$

Proof sketch: eigen-decompose

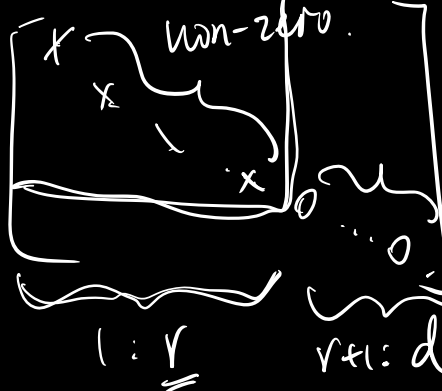
$\Lambda = U^T \Sigma U \in \mathbb{R}^{d \times d}$.

$\tilde{x} = U x$.

$x^T \Lambda^{-1} x = x^T U^T \Sigma^{-1} U x$.

$= \tilde{x}^T \Sigma^{-1} \tilde{x} = +\infty$.

$\Updownarrow$

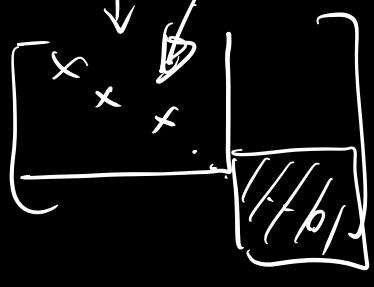


$\Rightarrow [\tilde{x}]_{r+1:d} \neq \vec{0}$.

$\operatorname{rank}(\Lambda + xx^T)$

$= \operatorname{rank}(U^T \Sigma U + U^T \tilde{x}\tilde{x}^T U)$

$= \operatorname{rank}(\Sigma + \tilde{x}\tilde{x}^T)$.



Remarks:

1. If we have explored everywhere.

i.e. $\mathbb{E}[\phi\phi^\top]$ is full-rank.

(Optimistic) FQI must output a near-optimal.

but in FQI lecture. we define exploratory data μ to be such that. $\max_\pi \frac{d_f^\pi(s,a)}{\mu(s,a)} \le C$.

Can unify: in FQI, we need C thru:

$$\sup_{f,f' \in \mathcal{F}} \frac{\|f - f'\|_{2,d_f^\pi}}{\|f - f'\|_{2,\mu}} \le \frac{d_f^\pi(s,a)}{\mu(s,a)} \le C.$$

$f - \mathcal{T}f, \quad \mathcal{T}f \in \mathcal{F}.$

∨

can show: when $\mathcal{F}$ is linear.

& $\mathbb{E}_\mu[\phi\phi^\top]$ has full rank

this quantity $< +\infty$.

2. UCB-LSVI [Jin et al '20].

→ $\exists f \in \mathcal{F} \quad \forall f$

→ use non-smooth.

Questions: can we do exploration with (say).

$Q^* \in \mathcal{F}$?

In Low-rank MDP. (i.e. linear MDP w/o ϕ given to learner)

$Q^* \in \mathcal{F} \quad \Rightarrow ?$

Bellman rank [JKALS '17].

1. $X \in \mathbb{R}^{n \times d}$, $Y \in \mathbb{R}^{n \times 1}$. $\hat{Y} = X\theta$, $\Lambda = X^T X$.

$\lim_{\lambda \to 0} (\Lambda + \lambda I)^{-1}$

Give $x_0 \in \mathbb{R}^d$ s.t. $x_0^T \Lambda^{-1} x_0 < +\infty$.

show that we can recover $x_0^T \theta$ even when Λ is rank deficient

Proof: eigen: $\Lambda = U^T \Sigma U$. $(Ux_0)^T \Sigma^{-1} (Ux_0) < +\infty$.

w.l.o.g. assume $\Sigma = \begin{bmatrix} x & & & \\ & \ddots & & \\ & & x & \\ & & & 0 \\ & & & & \ddots \\ & & & & & 0 \end{bmatrix}$ ($r = \text{rank}(\Lambda)$)

non-zero: $1 : r$; $r+1 : d$.

$\Rightarrow [Ux_0]_{r+1:d} = \vec{0}$.

---

w.t.s. $x_0^T \theta = x_0^T (\Lambda^\dagger X^T Y)$

RHS $= x_0^T \Lambda^\dagger X^T Y$ $\qquad x_0^T \Lambda^\dagger X^T$

$= x_0^T \Lambda^\dagger X^T X \theta$

$= x_0^T \Lambda^\dagger \Lambda \theta$

$= x_0^T U^T \Sigma^\dagger U U^T \Sigma U \theta$

$= x_0^T U^T \Sigma^\dagger \Sigma U \theta$.

$= (Ux_0)^T \Sigma^\dagger \Sigma U \theta$. $= (Ux_0)^T I_d U \theta$

$\begin{bmatrix} x \\ x \\ x \\ 0 \\ \vdots \\ 0 \end{bmatrix} \begin{bmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & 0 \\ & & & & \ddots \\ & & & & & 0 \end{bmatrix}$ $= x_0^T U^T U \theta$

$1 : r$. $= x_0^T \theta$ $\checkmark$.

2. Given $x_1^T \Lambda^{-1} x_1 = +\infty$.

show $\text{rank}(\Lambda + x_1 x_1^T) > \text{rank}(\Lambda)$.

(and similarly, if $x_1^T \Lambda^{-1} x_1 < +\infty$, rank doesn't change)

Proof: similar to above, we do everything in the eigen-decomposition of Λ, so.

W.l.o.g. we can assume Λ is diagonal.

$$\Lambda = \begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{bmatrix}$$

(also normalize the diagonals to 1).

$1:r$ $\quad r+1:d$.

$$x_1^T \Lambda^{-1} x_1 = \infty \iff [x_1]_{r+1:d} \neq \vec{0}.$$

Let $x_1 = [\underbrace{\bar{x}_1}_{1:r}, \underbrace{\tilde{x}_1}_{r+1:d}]$.

$$\Rightarrow \Lambda + x_1 x_1^T = \begin{bmatrix} I_r + \bar{x}_1 \bar{x}_1^T & \bar{x}_1 \tilde{x}_1^T \\ \tilde{x}_1 \bar{x}_1^T & \tilde{x}_1 \tilde{x}_1^T \end{bmatrix}.$$

now try to block-diagonalize this matrix.

$$\begin{bmatrix} I_r + \bar{x}_1 \bar{x}_1^T & \bar{x}_1 \tilde{x}_1^T \\ \tilde{x}_1 \bar{x}_1^T & \tilde{x}_1 \tilde{x}_1^T \end{bmatrix} = \begin{bmatrix} I_r & 0 \\ V & I_{d-r} \end{bmatrix} \begin{bmatrix} I_r + \bar{x}_1 \bar{x}_1^T & 0 \\ 0 & Z \end{bmatrix} \begin{bmatrix} I & V^T \\ 0 & I \end{bmatrix}.$$

W.t.s.: $Z \neq 0$

$$\begin{bmatrix} (I_r + \bar{X}_1\bar{X}_1^T) & 0 \\ V(I_r+\bar{X}_1\bar{X}_1^T) & Z \end{bmatrix} \begin{bmatrix} I & V^T \\ 0 & I \end{bmatrix}$$

$$= \begin{bmatrix} I_r + \bar{X}_1\bar{X}_1^T & (I_r+\bar{X}_1\bar{X}_1^T)V^T \\ V(I_r+\bar{X}_1\bar{X}_1^T) & V(I_r+\bar{X}_1\bar{X}_1^T)V^T + Z \end{bmatrix} \overset{\text{want}}{=} \begin{bmatrix} I_r+\bar{X}_1\bar{X}_1^T & \bar{X}_1\tilde{X}^T \\ \tilde{X}_1\bar{X}_1^T & \tilde{X}_1\tilde{X}_1^T \end{bmatrix}$$

$$\Rightarrow (I_r + \bar{X}_1\bar{X}_1^T)V^T = \bar{X}_1\tilde{X}_1^T$$

$$V^T = (I_r + \bar{X}_1\bar{X}_1^T)^{-1}(\bar{X}_1\tilde{X}_1^T).$$

$$\Rightarrow Z = \tilde{X}_1\tilde{X}_1^T - (\tilde{X}_1\bar{X}_1^T)(I_r + \bar{X}_1\bar{X}_1^T)^{-1}(\bar{X}_1\tilde{X}_1^T).$$

$(d-r)\times r$ $\quad r\times r$ $\quad r\times d$

to show $Z \neq 0$,

suffices to show $\exists y,\ y^T Z y > 0$.

let $y = \begin{pmatrix} \hat{} \\ \tilde{X}_1 \end{pmatrix}$. $\Rightarrow$ $y^T \tilde{X}_1\tilde{X}_1^T y = \|\tilde{X}_1\|^2$

$y^T\tilde{X}_1 (\bar{X}_1^T(I_r + \bar{X}_1\bar{X}_1^T)^{-1}\bar{X}_1)\tilde{X}_1^T y$

$\|\tilde{X}_1\|$ $\quad$ w.t.s. < 1. $\quad$ $\|\tilde{X}_1\|$.

$\bar{x}_1^T (I_r + \bar{x}_1 \bar{x}_1^T)^{-1} \bar{x}_1$.

rotate into basis w/ $\bar{x}_1$ as a basis vector

let Q be orthogonormal s.t.

$$\bar{x}_1 = Q \begin{bmatrix} \|x_1\| \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

$$\begin{bmatrix} \|x_1\| & 0 & \cdots & 0 \end{bmatrix} Q^T \left(I_r + Q \begin{bmatrix} \|x_1\|^2 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{bmatrix} Q^T\right)^{-1} Q \begin{bmatrix} \|x_1\| \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$= \frac{\|x_1\|^2}{\|x_1\|^2 + 1} < 1.$$