So far we have assumed $|S| < \infty$, $|A| < \infty$

And our regret scaled w/ $\text{poly}(|S|, |A|)\sqrt{T}$

Today: Two ways to handle infinite state + action spaces.

Let $\phi : S \times A \to \mathbb{R}^d$ be a feature map.

Def | Linear MDP. Assume $\exists$ $\{\theta_h\}_{h=1}^H \subset \mathbb{R}^d$, $\{\mu_h\}_h \subset \mathbb{R}^{|S| \times d}$

$$r_h(s,a) = \theta_h^T \phi(s,a) \qquad P_h(s' \mid s,a) = \mu_h(s')\,\phi(s,a).$$

Assume $\{\theta_h\}_h$ are known, and $\{\mu_h\}$ are unknown.

Things to note

- P_h has $|S|d$ parameters, and thus could be arbitrarily large.
- If we think of $\underbrace{P_h \in \mathbb{R}^{S \times SA}}$ as a matrix, it has rank d.

$$P_h = \underset{\mathbb{R}^{S \times d}}{\mu}\ \underset{\mathbb{R}^{d \times SA}}{\Phi} \qquad \Phi = \underbrace{[\phi(s_1,a_1) \cdots \phi(s_1,a_{|A|}), \phi(s_2,a_1) \cdots]}_{\mathbb{R}^{d \times SA}}$$

$\rightarrow$ P_h must be at most rank d.

• $Q_h(s,a)$ is linear in $\phi(s,a)$.

$$Q^*_h(s,a) = r_h(s,a) + \mathbb{E}_{s' \sim P(\cdot|s,a)}\left[V^*_{h+1}(s')\right]$$

$$= \theta_h^T \phi(s,a) + P(\cdot|s,a)^T V^*_{h+1}$$

$$= \theta_h^T \phi(s,a) + (V^*_{h+1})^T \mu_h \phi(s,a)$$

$$= \left(\theta_h^T + (V^*_{h+1})^T \mu_h\right) \phi(s,a)$$

$$= w_h^T \phi(s,a)$$

Suppose we have data of form

$$\left\{(s_h^i, a_h^i, s_{h+1}^i)\right\}_{i=1}^{k-1}$$

Want to estimate μ_h

$$\hat{\mu}_h^k = \underset{\mu \in \mathbb{R}^{S \times d}}{\operatorname{argmin}} \sum_{i=1}^{k-1} \left\| \delta(s_{h+1}^i) - \mu \phi(s_h^i, a_h^i) \right\|_2^2 + \lambda \|\mu\|_F^2$$

$$= \sum_{i=0}^{k-1} \delta(s_{h+1}^i) \phi(s_h^i, a_h^i)^T (\Lambda_h^k)^{-1}$$

where $\Lambda_h^h = \sum_{i=1}^{h-1} \phi(s_h^i, a_h^i)\, \phi(s_h^i, a_h^i)^T + \lambda I$

$$\mathbb{E}[\, \delta(s_{h+1}^i) \mid s_h^i, a_h^i \,] = \mu_h(s_{h+1}^i)\, \phi(s_h^i, a_h^i)$$

$$\mathbb{E}[\hat{\mu}_h^h] \approx \mu_h^h$$

$$\mathrm{Cov}\left(\hat{\mu}_h^h(s')\right) = \left(\Lambda_h^h\right)^{-1}$$

Key piece of analysis in finite MPS was bounding

$$\left| V^T\left(\hat{P}_{h+1}(\cdot \mid s,a) - P_{h+1}(\cdot \mid s,a)\right)\right| = \left| V^T\left(\hat{\mu}_{h+1}^k - \mu_{h+1}\right) \phi(s,a)\right|$$

$$= \left| V^T\left(\hat{\mu}_{h+1} - \mu_{h+1}\right)\left(\Lambda_h^h\right)^{1/2} \left(\Lambda_h^h\right)^{-1/2} \phi(s,a)\right|$$

$$\leq \left\| \left(\hat{\mu}_{h+1} - \mu_{h+1}\right)^T V \right\|_{\Lambda_h^h} \cdot \left\| \phi(s,a) \right\|_{(\Lambda_h^k)^{-1}}$$

$$\leq \beta \cdot \left\| \phi(s,a) \right\|_{(\Lambda_h^h)^{-1}}$$

hides factors of H, d.

But no dependence on $|S|$ or $|A|$

Algorithm

for $h = 1, 2, \ldots$

$$\hat{\mu}_h^k = \underset{\mu \in \mathbb{R}^{S \times d}}{\operatorname{argmin}} \sum_{i=1}^{k-1} \| \delta(s_{h+1}^i) - \mu \, \phi(s_h^i, a_h^i) \|_2^2 + \lambda \|\mu\|_F^2 \qquad \forall h$$

$V_{H+1} = 0$

for $h = H, H-1, \ldots, 1$

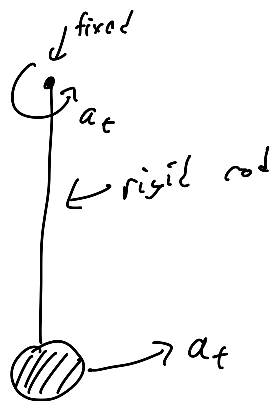
$$\hat{Q}_h^k(s,a) = \min \{ H, \; \beta \|\phi(s,a)\|_{(\Lambda_h^k)^{-1}} + \theta_h^T \phi(s,a) + \hat{V}_{h+1}^k \cdot \hat{\mu}_{h+1}^k \phi(s,a) \}$$

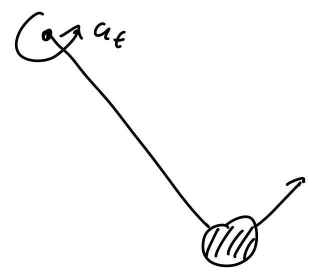
$$= \min \{ H, \; \beta \|\phi(s,a)\|_{(\Lambda_h^k)^{-1}} + r_h(s,a) + \hat{P}_{h+1}(s'|s,a)^T \hat{V}_{h+1}^k \}$$

$$\hat{V}_h^k(s,a) = \max_a \hat{Q}_h^k(s,a)$$

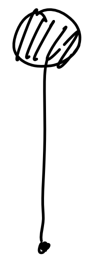
Roll-out collect $\{(s_h^k, a_h^k, s_{h+1}^k)\}_{h=1}^{H}$

Final Regret $\le \sqrt{d^3 K} \; \mathrm{poly}(H)$.





Goal state



Linear dynamical systems.

State $x_t \in \mathbb{R}^d$

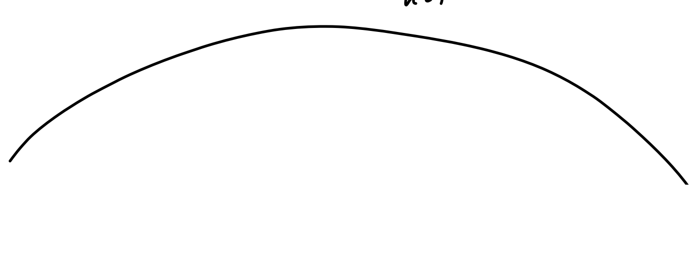
Action (input) $u_t \in \mathbb{R}^p$, $w_t \sim \mathcal{N}(0, I)$ noise

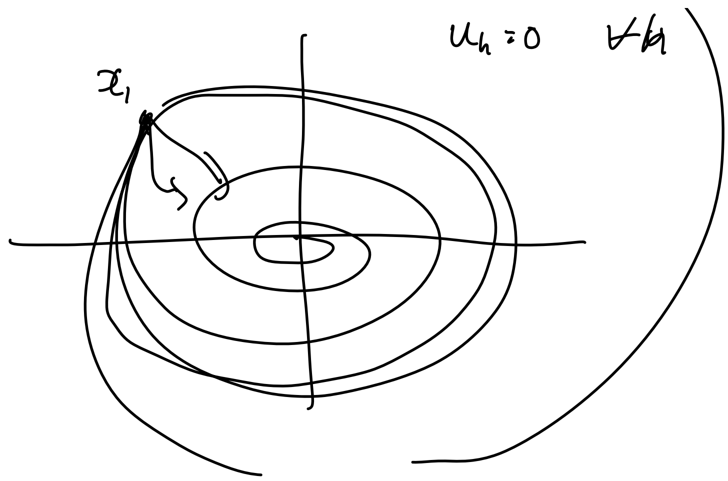
$$x_{t+1} = A x_t + B u_t + w_t$$

$$P_{h+1}(\cdot \mid x_h, u_h) \sim \mathcal{N}(A x_h + B u_h, I)$$

For known matrices $Q \in \mathbb{R}^{d \times d}$, $R \in \mathbb{R}^{p \times p}$

Loss function $\mathbb{E}\left[x_H^T Q x_H + \sum_{h=1}^{H-1} x_h^T Q x_h + u_h^T R u_h\right]$





$$Q_h^\pi(x,u) = \mathbb{E}\left[x_H^T Q x_H + \sum_{t=h}^{H-1} (x_t^T Q x_t + u_t^T R u_t) \,\middle|\, x_h = x, u_h = u \right]$$

$$Q_H^\pi(x,u) = \mathbb{E}[x_H^T Q x_H]$$

$$x_H = A x_{H-1} + B u_{H-1} + w_{H-1}$$

$$Q_{H-1}^*(x,u) = \mathbb{E}\big[(Ax + Bu + w_{H-1})^T Q (Ax + Bu + w_{H-1})$$
$$+ x^T Q x + u^T R u \big]$$

$$\arg\min_u Q_{H-1}^*(x,u) = -K_{H-1} x_{H-1}$$

Value iteration says in general $\exists K_t \; \forall t$

opt. $\quad u_t = -K_t x_t$