April 8

$$H-A \approx O(H\sqrt{K})$$

$$R_K \lesssim \sum_{k=1}^{K} \xi_k + \sum_{k=1}^{K} \sum_{h=1}^{H-1} \eta_h^{(k)}$$

$$+ 2H C_\delta \boxed{\sum_{k=1}^{K} \sum_{h=0}^{H-1} \frac{1}{\sqrt{1 \vee N_k(S_h^{(k)}, A_h^{(k)})}}}$$

IV.

$\forall u \in \{0, 1, \ldots\}$

$$\beta(u) \leq \frac{C_\delta}{\sqrt{u}}$$

$$\boxed{\beta(0) = 1 \leq \frac{C_\delta}{\sqrt{1}}}$$

$||$ do

SA, K, H big
δ small

$$\underline{\text{IV.}} = \sum_{s,a} \sum_{k=1}^{K} \sum_{h=0}^{H-1} \frac{\mathbb{I}(S_h^{(k)} = s, A_h^{(k)} = a)}{\sqrt{1 \vee N_k(s,a)}}$$

$$M_k(s,a) = \sum_{h=0}^{H-1} \mathbb{I}(S_h^{(k)} = s, A_h^{(k)} = a)$$

$$N_k(s,a) = M_1(s,a) + \dots + M_k(s,a)$$

$$= \sum_{s,a} \sum_{k=1}^{K} \frac{\overbrace{M_k(s,a)}^{f'(k)}}{\sqrt{1 \vee \underbrace{(M_1(s,a) + \dots + M_k(s,a))}_{f(k)}}}$$

$$\int_1^K \frac{f'(x)}{\sqrt{f(x)}}\,dx = 2\Big[\sqrt{f(x)}\Big]_1^K$$

$$\le 2\sqrt{f(K)}$$

$$2\left(\sqrt{f}\right)' = \cancel{\frac{1}{2}} \cdot \frac{f'}{\sqrt{f}}$$

<u>Lemma</u>: $\forall\, m_1, \dots, m_k \geq 0$

$$\sum_{k=1}^{K} \frac{m_k}{\sqrt{1 \vee (m_1 + \dots + m_k)}} \leq 2\sqrt{m_1 + \dots + m_k}$$

Proof: telescoping.

$$\underline{\text{IV.}} \le 2\sum_{s,a} \sqrt{N_k(s,a)}$$

$$\le 2SA \boxed{\sum_{s,a} \frac{1}{SA} \sqrt{N_k(s,a)}}$$

Jensen's $\le$

$$\le 2SA \sqrt{\underbrace{\sum_{s,a} N_k(s,a)}_{HK} / SA}$$

$$= 2\sqrt{SAHK}$$

$\underline{\underline{\text{Theorem:}}}$ w.p. $1 - O(\delta)$

$$R_K \le 4 C_\delta H \sqrt{SAHK} +$$

$\sqrt{S \log(2) + \log HKSA/\delta}$

$\boxed{\text{UCRL(2)}}$ $c' H \sqrt{K \log \frac{1}{\delta}} +$

Jaksch
Ortner $c'' H \sqrt{HK \log \frac{1}{\delta}}$
Auer

Zihan Zhang / Xiangyang Ji / Simon S. Du

Theorem: Total reward per episode $\leq R$

$$R_K = O\left(R\left(\sqrt{SAK} + S^2 A\right) \operatorname{polylog} \frac{SAHK}{\delta}\right)$$

Theorem: $R_K = \Omega\left(R\sqrt{SAK}\right)$

Inhomogeneous : $P = (P_0), \ldots, (P_{H-1})$

$r = (r_0, \ldots r_{H-1})$

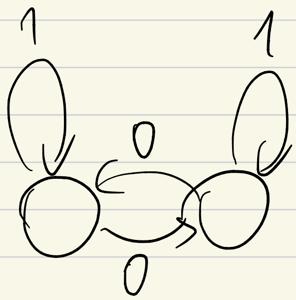
extra H

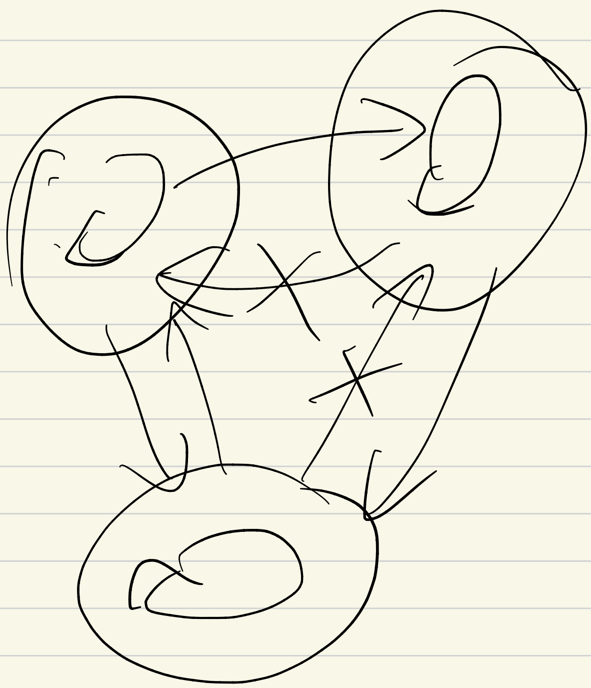
Optimistic Q-learning is opt.

Inf. horizon → episodes ← eat the cake!

- count doubles
- fixed schedule

necessary: linear regret otherwise!



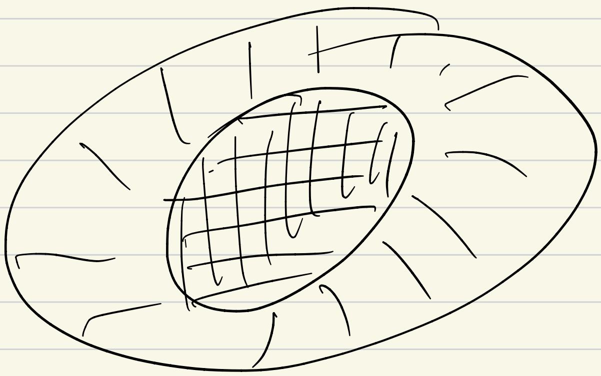


Randomize

$\downarrow$

PSRL / Thompson sampling

$\boxed{\text{whp you are optimistic}}$



Elimination

IDS

Finite-horizon setting: $H > 0$

$\varphi_h : S \times A \to \mathbb{R}^d$ $\qquad \|\varphi_h\|_2 \le 1$

Inhomogeneous case $\qquad h = 0, \ldots, H-1$

$q^* \in \mathcal{F}_\varphi$?

Idea: UCB on q^* + greedy.

$S_0^{(k)}, A_0^{(k)}, \ldots\ldots, S_{H-1}^{(k)}, A_{H-1}^{(k)}, S_H^{(k)}$

$$\underline{q_h^{(k)}} = \left[\underbrace{\Phi_h w_h^{(k)}}\right]^H + r_h + \beta_{kh} \qquad v_h^{(k)} = M q_h^{(k)}$$

known

$$q_h^* = r_h + P_h v_{h+1}^*, \quad v_h^* = M q_h^*$$

value iteration

Begining of episode k; data up to $k-1$

$$\underline{\underline{w_h^{(k)}(v)}} = \operatorname*{argmin}_{w \in \mathbb{R}^d} \sum_{i=1}^{k-1} \left(\varphi_h(Z_{ih})^T w - v(S_{h+1}^{(i)}) \right)^2 + \lambda \|w\|_2^2$$

$Z_{kh} = (S_h^{(k)}, A_h^{(k)})$

$w_h^{(k)} := w_h^{(k)}(v_{h+1}^{(k)})$

LSVI - UCB

Chi Jin
Zhuoran Yang Zhaoran Wang
M. Jordan

β_{kh} TBC

Ass: $\mathcal{V} \subseteq \mathbb{R}^{SA}$ $\quad \forall v \in \mathcal{V}: P_h v \in \mathcal{F}_{\varphi_h}$
$\forall h \in [H]$

Lemma:

$$X_1, \ldots, X_n, \ldots \in \mathbb{R}^d$$
$$Y_1, \ldots, Y_n, \ldots \in \mathbb{R}$$

$$Y_t = X_t^T \theta_* + \varepsilon_t ,$$

$$\mathbb{E}\left[e^{\lambda \varepsilon_t} \mid \mathcal{F}_{t-1}\right] \le e^{\frac{\lambda^2}{2}\sigma^2}$$

$$\hat{\theta}_t = \sum_{t=1}^{n} (X_t^T \theta - Y_t)^2 + \lambda \|\theta\|_2^2$$

wp $1-\delta$ $\quad \forall x \in \mathbb{R}^d$

$$|\langle x, \hat{\theta}_t \rangle - \langle x, \theta_* \rangle| \le \|x\|_{M_t^{-1}} \beta_t(\delta)$$

$$M_t = \sum_{s=1}^{t} X_s X_s^T + \lambda I$$

$$\beta_t(\delta) = \sqrt{\log \det \left(\left(\frac{M_t}{\lambda I}\right)^{1/2} / \delta \right) + \lambda \|\theta_*\|_2}$$

$$v \mapsto W_u^{(\ell)}(v) \qquad \boxed{v \mapsto \Phi\, W_u^{\ell}(v)} = P_u^{(\ell)} v$$

$$P_u v$$

linear in v