Linear MDP $(S, A, \{P_h\}, \{R_h\}, H, d_0)$.

$P_h(s'|s,a) = \phi_h(s,a)^T \psi(s')$ $\to \mathbb{R}^d$. $\quad \|\phi\|_2 \le 1$.

$R_h(s,a) = \phi_h(s,a)^T \theta_R$. $\quad R_h \in [0,1]$.

$V_{max} = H$.

$\Rightarrow \forall f_{h+1}$, $\mathcal{T} f_{h+1}$ is linear in ϕ_h.

Alg (UCB-LSVI): for $t = 1, 2, \cdots, T$

for $h = H, H-1, \cdots, 1$.

1. Define: $\Lambda_h^t = I + \sum_{i=1}^{t-1} \phi_h^i (\phi_h^i)^T$.

where $\phi_h^i := \phi_h(s_h^i, a_h^i)$ $\to$ episode index.

2. $\tilde{Q}_h^t(s,a) = \phi_h(s,a)^T (\Lambda_h^t)^{-1} \sum_{i=1}^{t-1} \phi_h^i \left(r_h^i + \hat{V}_{h+1}^t(s_{h+1}^i) \right)$

3. $\hat{Q}_h^t(s,a) = \tilde{Q}_h^t(s,a) + \beta \sqrt{\phi_h(s,a)^T (\Lambda_h^t)^{-1} \phi_h(s,a)}$.

(optism).

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

4. interact w/ MDP using π^t greedy w.r.t. $\hat{Q}_h^t$.

$$\text{Regret}_T := \sum_{t=1}^{T} \left(J(\pi^*) - J(\pi^t) \right).$$

Lemma. $\forall h, t$. Define $b_h^t(s,a) :=$

$$\left| \underbrace{\tilde{Q}_h^t(s,a)}_{(w.h.p.)\ \Delta} - \left(R_h(s,a) + P_h(s,a)^T \underbrace{\hat{V}_{h+1}^t}_{\beta} \right) \right|.$$

$$b_h^t(s,a) \le \|\phi(s,a)\|_{(\Lambda_h^t)^{-1}} \tilde{O}\left(H\left(d + \sqrt{\log \tfrac{1}{\alpha}}\right)\right).$$

$$\Rightarrow \hat{Q}_h^t(s,a) \ge R_h(s,a) + P_h(s,a)^T \hat{V}_{h+1}^t. \quad \Delta.$$

$X_1, \cdots, X_n$.

$$\hat{\mathbb{E}}[f(X_i)] \longrightarrow \mathbb{E}[f(X_i)]$$

$$f = f(X_1, \cdots, X_n) \in \mathcal{F}.$$

$x \in \mathbb{R}^d$. $\mathcal{F} = \{ f(x) = x^T A x : \forall A \in \mathbb{R}^{d \times d}, \}$.

$x^T \theta$.

$$x^T A x = \sum_{i,j} x_i x_j A_{ij}.$$

Lemma: Optism: $\forall h, t$, $J(\pi^*) = \mathbb{E}_{d_0}[V_1^*]$.

$\rightarrow$ $\hat{Q}_h^t \geq Q_h^*$, $\hat{V}_h^t \geq V_h^*$. ✓

Proof: $h = H+1$ ✓.

$$\hat{Q}_h^t(s,a) \geq R_h(s,a) + P_h(s,a)^T \hat{V}_{h+1}^t$$

$$\geq R_h + P_h^T V_{h+1}^*$$

$$= Q_h^*(s,a).$$

$$\hat{V}_h(s) = \min\left(H, \max_a \hat{Q}_h^t(s,a)\right).$$

$\Downarrow$ $V_h^*(s)$ $\qquad$ $\max_a Q_h^*(s,a) = V_h^*(s)$

$$\geq V_h^*(s).$$

$$\text{Regret}_T = \sum_{t=1}^{T} J(\pi^*) - J(\pi^t)$$

$$\leq \sum_{t=1}^{T} \mathbb{E}_{d_0}\left[\hat{Q}_1^t(s_1, \pi^t)\right] - J(\pi^t).$$

$$\leq \sum_{t=1}^{T} \sum_{h=1}^{H} \mathbb{E}_{d_h^{\pi^t}}\left[\hat{Q}_h^t(s,a) - R_h(s,a) - P_h(s,a)^T \left[\max_a \hat{Q}_{h+1}^t(\cdot, a)\right]\right]$$

$\min(\cdot, H)$

V_{h+1}

$$\leq \sum_{t=1}^{T} \sum_{h=1}^{H} \mathbb{E}_{d_h^{\pi^t}}\left[\tilde{Q}_h^t(s,a) + \beta \cdot \|\phi(s,a)\|_{(\Gamma_h^t)^{-1}} - R_h - P_h^T \hat{V}_{h+1}\right]$$

$$\leq \sum_{t=1}^{T} \sum_{h=1}^{H} \mathbb{E}_{d_h^{\pi^t}}\left[2\beta \|\phi(s,a)\|_{(\Gamma_h^t)^{-1}}\right].$$

$$= 2\beta \sum_{h=1}^{H} \sum_{t=1}^{T} \mathbb{E}_{d_h^{\pi^t}} \left[\|\phi(s,a)\|_{(\Lambda_h^t)^{-1}} \right].$$

$$\leq 2\beta \sum_{h=1}^{H} \left(\sum_{t=1}^{T} \|\phi_h(s_h^t, a_h^t)\|_{(\Lambda_h^t)^{-1}} \right.$$

$\mathbb{E}[\cdots \mid s_h^1, a_h^1, \ldots, s_h^{t-1}, a_h^t]$

$\left. + \text{Azuma's.} \right)$

$\Big\}$ · elliptical potential.