Generalized Eluder Coefficient

Episodic Interactive Decision Making Problem: $(\mathcal{O}, \mathcal{A}, H, \mathbb{P}, R)$

- $\mathcal{O}, \mathcal{A}$: observation and action space
- H: length of horizon
- $\mathbb{P} = \{\mathbb{P}_h\}_{h\in[H]}$: $\mathbb{P}_h(o_{h+1} | \tau_h)$
- $R = \{R_h\}_{h\in[H]}$: $R_h(o_h, a_h)$

Example

- Markov Decision Process: $\mathcal{O} = S$
- Partially Observable MDP
- Predictive State Representation

Exploration-Exploitation Tradeoff

- optimism in the face of uncertainty
- posterior sampling

Unknown Model

Definition (Generalized Eluder Coefficient)

A model-based hypothesis class is a set of model $\mathcal{H}$ such that

$$f = (\mathbb{P}_f, R_f) \in \mathcal{H}$$

- $\pi_f = \{\pi_{h,f}\}_{h \in [H]}$
- $V_f = \{V_{h,f}\}_{h \in [H]}$

Generalized Eluder Coefficient

Definition (Generalized Eluder Coefficient)

Given a $\mathcal{H}$, a discrepancy function $\ell = \{\ell_f\}_{f \in \mathcal{H}}$, an exploration policy class $\Pi_{\exp}$, and $\epsilon > 0$, the generalized eluder coefficient $\mathrm{GEC}(\mathcal{H}, \ell, \Pi_{\exp}, \epsilon)$ is the smallest $d > 0$ such that

$$\sum_{k=1}^{K} V_{f^k} - V^{\pi_{f^k}} \leq \left[d \sum_{h=1}^{H} \sum_{k=1}^{K} \left(\sum_{s=1}^{k-1} \mathbb{E}_{\pi_{\exp}(f^s,h)} \ell_{f^s}(f^k, \xi_h) \right) \right]^{1/2} + 2\sqrt{dHT} + \epsilon HT$$

Optimism + Low GEC ≈ Sample Efficiency

$$
\begin{aligned}
\mathrm{Reg}(K) &= \sum_{k=1}^{K} V^* - V^{\pi_{f^k}} \\
&\overset{\text{optimism}}{\leq} \sum_{k=1}^{K} V_{f^k} - V^{\pi_{f^k}} \\
&\overset{\text{GEC}}{\leq} \left[d \sum_{h=1}^{H} \sum_{k=1}^{K} \left(\sum_{s=1}^{k-1} \mathbb{E}_{\pi_{\exp}(f^s,h)} \ell_{f^s}(f^k, \xi_h) \right) \right]^{1/2} + 2\sqrt{dHT} + \epsilon HT \\
&\overset{\text{training} \quad \text{error} \leq \beta}{\leq} \sqrt{dHT\beta}
\end{aligned}
$$

Maximize to Explore

For episode k, choose f^k such that

$$f^k = \underset{f \in \mathcal{H}}{\operatorname{argsup}} \{ V_{1,f}(x_1) - \eta L_f(\mathcal{D}^{k-1}) \}$$

Execute $\pi_{\mathrm{exp}}(f^k)$ to collect data.

Generic Posterior Sampling Algorithm

For episode k, sample $f^k \sim p^k$ such that

$$p^k(f) \propto p^0(f) \cdot \exp(V_{1,f}(x_1) - \eta L_f(\mathcal{D}^{k-1}))$$

Execute $\pi_{\exp}(f^k)$ to collect data.

Regret Analysis for MEX

$$\mathrm{Reg}(K) = \sum_{k=1}^{K} V^* - V^{\pi_{f^k}} = \sum_{k=1}^{K} V^* - V_{f^k} + \underbrace{\sum_{k=1}^{K} V_{f^k} - V^{\pi_{f^k}}}_{\mathrm{GEC}}$$

$$\begin{aligned}\sum_{k=1}^{K} V^* - V_{f^k} &= \sum_{k=1}^{K} V_{f^*} - V_{f^k} \\ &\le \eta \sum_{k=1}^{K} L_{f^*}(\mathcal{D}^{k-1}) - L_{f^k}(\mathcal{D}^{k-1}) \\ &\le -\eta d \sum_{h=1}^{H} \sum_{k=1}^{K} \left(\sum_{s=1}^{k-1} \mathbb{E}_{\pi_{\exp}(f^s,h)} \ell_{f^s}(f^k, \xi_h) \right) + \cdots\end{aligned}$$

Regret Analysis for GPS

$$\mathbb{E}_{f^k\sim p^k}(\mathrm{Reg}(K)) = \mathbb{E}\left(\sum_{k=1}^{K} V^* - V^{\pi_{f^k}}\right) = \mathbb{E}\left[\sum_{k=1}^{K} V^* - V_{f^k} + \underbrace{\sum_{k=1}^{K} V_{f^k} - V^{\pi_{f^k}}}_{\mathrm{GEC}}\right]$$

$$\mathbb{E}\left[\sum_{k=1}^{K} V^* - V_{f^k}\right] = \mathbb{E}\left[\sum_{k=1}^{K} V_{f^*} - V_{f^k}\right]$$

Lemma 1

$$\mathbb{E}[V_{f^*} - V_{f^k}] \le \mathbb{E}\left[-\eta d \sum_{h=1}^{H}\left(\sum_{s=1}^{k-1} \mathbb{E}_{\pi_{\exp}(f^s,h)} \ell_{f^s}(f^k, \xi_h)\right)\right] + \cdots$$