1 MEX on infinite hypothesis set

Given an infinity hypothesis set $\mathcal{H}_{\text{inf}}$, how to run MEX algorithm on it? In this case, we can get a regret

$$\text{Reg}(K) \lesssim \sqrt{HK} + K\varepsilon_\psi$$

by running MEX on a random set $\mathcal{H}$ where

$$\varepsilon_\psi \triangleq \min_{\pi \in \mathcal{H}} |V(\psi(\pi), \pi) - V(\psi(\pi^*), \pi^*)|.$$

We also denote

$$\pi^*_{\text{det}} \triangleq \text{argmin}_{\pi \in \mathcal{H}} |V(\psi(\pi), \pi) - V(\psi(\pi^*), \pi^*)|$$

Lemma 1. *If* $\pi^* \in \mathcal{H}$, $\text{Reg}(K) \lesssim \sqrt{HK}$.

2 Proof of the regret bound

Proof. We decompose the regret into two terms,

$$\begin{aligned}
\text{Regret}(K) &\triangleq \sum_{k=1}^{K} V(\psi(\pi^*), \pi^*) - V(\psi(\pi^k), \pi^*) \\
&= \underbrace{\sum_{k=1}^{K} V(\psi(\pi^*), \pi^*) - V(\psi(\pi^k), \pi^k)}_{\text{Term (i)}} + \underbrace{\sum_{k=1}^{K} V(\psi(\pi^k), \pi^k) - V(\psi(\pi^k), \pi^*)}_{\text{Term (ii)}}.
\end{aligned}$$

Term (i). By the choice of π^k, we have

$$V(\psi(\pi^*_{\text{det}}), \pi^*_{\text{det}}) - \eta \sum_{h=1}^{H} L_h^{k-1}(\pi^*_{\text{det}}) \leq V(\mu^k, \pi^k) - \eta \sum_{h=1}^{H} L_h^{k-1}(\pi^k)$$

for all $k \in [K]$. By definition

$$V(\psi(\pi^*_{\text{det}}), \pi^*_{\text{det}}) \geq V(\psi(\pi^*), \pi^*) - \varepsilon_\psi.$$

Thus,

$$V(\psi(\pi^*), \pi^*) - V(\mu^k, \pi^k) \leq \eta \sum_{h=1}^{H} L_h^{k-1}(\pi^*_{\text{det}}) - \eta \sum_{h=1}^{H} L_h^{k-1}(\pi^k) + \varepsilon_\psi. \tag{1}$$

for any $\pi^k \overset{\psi}{\sim} \pi^{k'}$, we have

$$V(\psi(\pi^k), \pi^k) = V(\psi(\pi^{k'}), \pi^{k'}).$$

Thus, an upper bound for $V(\psi(\pi^*), \pi^*) - V(\mu^k, \pi^k)$ is also an upper bound for $V(\psi(\pi^*), \pi^*) - V(\mu^{k'}, \pi^{k'})$. Applying Lemma 2, we have that with probability at least $1 - \delta$, for any $(h, k) \in [H] \times [K]$, $\mu^s = \psi(\pi^s)$ and $\pi^k \in \mathcal{H}$,

$$L_h^{k-1}(\pi^*_{\text{det}}) - L_h^{k-1}(\pi^k) \leq -2 \sum_{s=1}^{k-1} \mathbb{E}_{\xi_h \sim \mu^s}[\ell_{\pi^s}(\pi; \xi_h)] + 2\log(H n_{\text{type}}^{\psi}(\mathcal{H}) / \delta).$$

Substituting the above equation into (1) gives us that with probability at least $1-\delta$, for any $k\in[K]$, $\mu^s=\psi(\pi^s)$ and $\pi^k\in\mathcal{H}$

$$V(\psi(\pi^*),\pi^*)-V(\mu^k,\pi^k)\leq -2\eta\sum_{h=1}^{H}\sum_{s=1}^{k-1}\mathbb{E}_{\xi_h\sim\mu^s}[\ell_{\pi^s}(\pi;\xi_h)]+2H\eta\log(Hn_{\text{type}}^{\psi}(\mathcal{H})/\delta)+\varepsilon_{\psi}.$$

Summing over $[K]$ gives us

$$\text{Term (i)}\leq -2\eta\sum_{k=1}^{K}\sum_{h=1}^{H}\sum_{s=1}^{k-1}\mathbb{E}_{\xi_h\sim\mu^s}[\ell_{\pi^s}(\pi;\xi_h)]+2\eta KH\log(Hn_{\text{type}}^{\psi}(\mathcal{H})/\delta)+K\varepsilon_{\psi}.$$

Term (ii). Follow the proof of Theorem 4.4 in [1], we have that for all $\mu^s=\psi(\pi^s)$

$$\text{Term (ii)}\leq 2\eta\sum_{k=1}^{K}\sum_{h=1}^{H}\sum_{s=1}^{k-1}\mathbb{E}_{\xi_h\sim\mu^s}[\ell_{\pi^s}(\pi;\xi_h)]+\frac{d_{\text{GEC}}(\varepsilon_{\text{conf}})}{8\eta}+\sqrt{d_{\text{GEC}}(\varepsilon_{\text{conf}})HK}+\varepsilon_{\text{conf}}HK.$$

Combining Term (i) and Term (ii).

$$\begin{aligned}\text{Regret}(K)&=\text{Term (i)}+\text{Term (ii)}\\&\leq 2\eta KH\log(Hn_{\text{type}}^{\psi}(\mathcal{H})/\delta)+\frac{d_{\text{GEC}}(\varepsilon_{\text{conf}})}{8\eta}+\sqrt{d_{\text{GEC}}(\varepsilon_{\text{conf}})HK}+\varepsilon_{\text{conf}}HK+K\varepsilon_{\psi}.\end{aligned}$$

Set $\varepsilon_{\text{conf}}=\frac{1}{\sqrt{HK}}-\frac{\varepsilon_{\psi}}{H}$. For $\varepsilon_{\text{conf}}>0$, we need

$$\varepsilon_{\psi}\leq\sqrt{\frac{H}{K}}$$

□

3 Appendix

Lemma 2. *With probability at least $1-\delta$, for any $(h,k)\in[H]\times[K]$, $\mu^s\in\text{BR}(\pi^s)$, and $\pi\in\Pi$*

$$L_h^{k-1}(\pi^*)-L_h^{k-1}(\pi)\leq -2\sum_{s=1}^{k-1}\mathbb{E}_{\xi_h\sim\mu^s}[\ell_{\pi^s}(\pi;\xi_h)]+2\log(H|\Pi|/\delta).$$

Proof. Given $\pi\in\mathcal{H}$, we denote the random variable $X_{h,\pi}^k$ as

$$X_{h,\pi}^k=\log\left(\frac{\mathbb{P}_{h,\pi^*}(s_{h+1}^k|\,s_h^k,a_h^k)}{\mathbb{P}_{h,\pi}(s_{h+1}^k|\,s_h^k,a_h^k)}\right).$$

Now we define a filtration $\{\mathcal{F}_{h,k}\}_{k=1}^K$ as (B.25) in [1]. Thus we have $X_{h,\pi}^k\in\mathcal{F}_{h,k}$. Therefore, by applying Lemma D.1 in [1], we have that with probability at least $1-\delta$, for any $(h,k)\in[H]\times[K]$, and $\pi\in\Pi$, we have

$$-\frac{1}{2}\sum_{s=1}^{k-1}X_{h,\pi}^s\leq\sum_{s=1}^{k-1}\log\mathbb{E}\left[\exp\left\{-\frac{1}{2}X_{h,\pi}^s\right\}|\,\mathcal{F}_{h,s-1}\right]+\log(H|\Pi|/\delta). \tag{2}$$

Meanwhile, by (B.27) in [1], for any $\mu^s\in\text{BR}(\pi^s)$, the conditional expectation equals to

$$\mathbb{E}\left[\exp\left\{-\frac{1}{2}X_{h,\pi}^s\right\}|\,\mathcal{F}_{h,s-1}\right]=1-\mathbb{E}_{(s_h^s,a_h^s)\sim\mu^s}[D_H(\mathbb{P}_{h,\pi^*}(\cdot|\,s_h^s,a_h^s)||\mathbb{P}_{h,\pi}(\cdot|\,s_h^s,a_h^s))]. \tag{3}$$

Denote $\mathbb{E}_{(s_h^s,a_h^s)\sim\mu^s}[D_H(\mathbb{P}_{h,\pi^*}(\cdot|\, s_h^s,a_h^s)||\mathbb{P}_{h,\pi^s}(\cdot|\, s_h^s,a_h^s))]$ as $\mathbb{E}_{\xi_h\sim\mu^s}[\ell_{\pi^s}(\pi;\xi_h)]$. Using the fact $\log(x)\leq x-1$ and substituting (3) into (2) finishes the proof. $\square$

Bibliography

[1] Zhihan Liu, Miao Lu, Wei Xiong, Han Zhong, Hao Hu, Shenao Zhang, Sirui Zheng, Zhuoran Yang, and Zhaoran Wang. One Objective to Rule Them All: A Maximization Objective Fusing Estimation and Planning for Exploration. may 2023.