Case Study

Consider a normal-form game defined as $(2, A_{\text{joint}} = [N]^2, r_{\text{joint}} = (V, V))$. The shared reward function r is defined as

$$V(a,b) = \begin{cases} 1 & \text{if} \quad a = b \\ 0 & \text{if} \quad a \neq b \end{cases}.$$

Now, assume the player 2 only takes pure strategy, i.e., $\mathcal{H}^* = [N]$. From player 1's perspective, the possible reward functions assigned to her are

$$\mathcal{F} \triangleq \{V_b | \, V_b(\cdot) = V(\cdot, b), b \in [N]\}.$$

In the following, we will reduce the proof used for regret bound of MEX algorithm in this simple example. We set $\mathcal{H} = [N-1]$ and $\pi^* = N$. Also, since there is no state transition, we apply the Hellinger distance and the loss function for estimation on the reward.

1 GEC term

First to note that the Hellinger distance defined as

$$\ell(\pi^k; (a,r)) \triangleq D_H(V(\cdot|\, a, \pi^k), V(\cdot|\, a, \pi^*))$$

always equals $1/\sqrt{2}$ for all $k \in [K]$. The training error $\mathcal{L}_{\text{train}}(\pi^k)$ defined as

$$\mathcal{L}_{\text{train}}(\pi^k) = \sum_{k=1}^{K} \sum_{s=1}^{k-1} \mathbb{E}_{a \sim \psi(\pi^k)}[\ell(\pi^k; (a,r))]$$

equals

$$\mathcal{L}_{\text{train}}(\pi^k) = \frac{K^2 - K}{2\sqrt{2}}.$$

Also, define φ as

$$\varphi(\alpha, \varepsilon, K) \triangleq \frac{d(\varepsilon)}{2\alpha} + \sqrt{d(\varepsilon)\, K} + \varepsilon K.$$

Now, the GEC assumption states that: there exist an $d(\varepsilon) > 0$, such that for all $\{\pi^k\} \subset \mathcal{H}$,

$$\sum_{k=1}^{K} V(\psi(\pi^k), \pi^k) - V(\psi(\pi^k), \pi^*) \leq \inf_{\alpha > 0} \left\{ \frac{\alpha}{2} \mathcal{L}_{\text{train}}(\pi^k) + \varphi(\alpha, \varepsilon, K) \right\}.$$

Combining all above equations gives us

$$\begin{aligned} \sum_{k=1}^{K} V(\psi(\pi^k), \pi^k) - V(\psi(\pi^k), \pi^*) &\leq \inf_{\alpha > 0} \left\{ \frac{\alpha(K^2 - K)}{4\sqrt{2}} + \frac{d(\varepsilon)}{2\alpha} + \sqrt{d(\varepsilon)\, K} + \varepsilon K \right\} \\ &= \sqrt{\frac{(K^2 - K) d(\varepsilon)}{2\sqrt{2}}} + \sqrt{d(\varepsilon)\, K} + \varepsilon K. \end{aligned}$$

2 Type term

The estimation loss function L is defined as

$$L^k(\pi) = \begin{cases} 0 & \text{if } r(\psi(\pi), \pi^*) = 1 \\ 1 & \text{if } r(\psi(\pi), \pi^*) = 0 \end{cases}.$$

It is clear that $L^k(\pi) = 1$ for all $k \in [K]$ and $\pi \in \mathcal{H}$ in our setting. By on the choice of π^k (based on the MEX algorithm), the type term is bounded by

$$V(\psi(\pi^*), \pi^*) - V(\psi(\pi^k), \pi^k) \leq \eta \sum_k \; (L^{k-1}(\pi_{\mathcal{H}}) - L^{k-1}(\pi^k)) + \varepsilon_{\mathcal{H}},$$

where $\varepsilon_{\mathcal{H}} = 0$. The term

$$L^{k-1}(\pi_{\mathcal{H}}) - L^{k-1}(\pi^k) = 0.$$

Now, we decompose the term $L^{k-1}(\pi_{\mathcal{H}}) - L^{k-1}(\pi^k)$ into the sum of $L_h^{k-1}(\pi^*) - L_h^{k-1}(\pi^k)$ and $L_h^{k-1}(\pi_{\mathcal{H}}) - L_h^{k-1}(\pi^*)$. It is clear that

$$L^{k-1}(\pi^*) - L^{k-1}(\pi^k) = -1$$

and

$$L^{k-1}(\pi_{\mathcal{H}}) - L^{k-1}(\pi^*) = 1$$

since $L_h^{k-1}(\pi^*) = 0$. Thus, the upper bound of the type term can be expressed as

$$V(\psi(\pi^*), \pi^*) - V(\psi(\pi^k), \pi^k) \leq -\eta K + \eta K$$

where $-\eta K$ term upper bounds $L^{k-1}(\pi^*) - L^{k-1}(\pi^k)$ and the ηK term upper bounds $L^{k-1}(\pi_{\mathcal{H}}) - L^{k-1}(\pi^*)$. Combining all these equations gives us

$$\mathrm{Reg}(K) \leq \underbrace{\sqrt{\frac{(K^2 - K)d(\varepsilon)}{2\sqrt{2}}} + \varepsilon K}_{\mathrm{GEC}} \underbrace{-\eta K}_{L^{k-1}(\pi^*) - L^{k-1}(\pi^k)} \underbrace{+\eta K}_{L^{k-1}(\pi_{\mathcal{H}}) - L^{k-1}(\pi^*)} + \text{sublinear term.}$$

The upper bound of GEC term together with the upper bound of $L^{k-1}(\pi^*) - L^{k-1}(\pi^k)$ are cancled out if we choose η properly. However the linear term ηK remains due to the upper bound of $L^{k-1}(\pi_{\mathcal{H}}) - L^{k-1}(\pi^*)$.