Regret Analysis When Realizability Assumption Fails

BY JACK YANSONG LI
University of Illinois Chicago
Email: yli340@uic.edu

We consider the case where $\pi^* \notin \mathcal{H}$. We denote

$$\varepsilon_{\mathcal{H}} \triangleq \min_{\pi \in \mathcal{H}} |V^*(\pi) - V^*(\pi^*)| \tag{1}$$

and

$$\pi_{\mathcal{H}} \in \arg\min_{\pi \in \mathcal{H}} |V^*(\pi) - V^*(\pi^*)|.$$

The regret is defined as

$$\sum_{k=1}^{K} V(\psi(\pi^*), \pi^*) - V(\psi(\pi^k), \pi^*),$$

which can be decomposed into the sum of

$$\sum_{k=1}^{K} V(\psi(\pi^*), \pi^*) - V(\psi(\pi^k), \pi^k) \tag{2}$$

and

$$\sum_{k=1}^{K} V(\psi(\pi^k), \pi^k) - V(\psi(\pi^k), \pi^*). \tag{3}$$

In the following, we call (2) the type term and (3) the GEC term. For clarification, we also consider the following example:

Example 1. Consider a normal-form game defined as $(2, A_{\text{joint}} = [N]^2, r_{\text{joint}} = (r, r))$. The shared reward function r is defined as

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

The player 2 only takes pure strategy, i.e., $\mathcal{H}^* = [N]$. We consider the case where the hypothesis set is not complete, i.e., $\mathcal{H} = [N-1]$ and the true strategy of player 2 is out of the hypothesis set, i.e., $\pi^* = N$.

1 Analysis of the type term

By on the choice of π^k (based on the MEX algorithm), we have

$$V(\psi(\pi_{\mathcal{H}}), \pi_{\mathcal{H}}) - \eta \sum_{h=1}^{H} L_h^{k-1}(\pi_{\mathcal{H}}) \leq V(\psi(\pi^k), \pi^k) - \eta \sum_{h=1}^{H} L_h^{k-1}(\pi^k)$$

for all $k \in [K]$. By (1), $V(\psi(\pi_{\mathcal{H}}), \pi_{\mathcal{H}}) \geq V(\psi(\pi^*), \pi^*) - \varepsilon_{\mathcal{H}}$. Thus,

$$V(\psi(\pi^*), \pi^*) - V(\psi(\pi^k), \pi^k) \leq \eta \left(\sum_{h=1}^{H} L_h^{k-1}(\pi_{\mathcal{H}}) - \sum_{h=1}^{H} L_h^{k-1}(\pi^k) \right) + \varepsilon_{\mathcal{H}}.$$

The term $\sum_{h=1}^{H} L_h^{k-1}(\pi_{\mathcal{H}}) - \sum_{h=1}^{H} L_h^{k-1}(\pi^k)$ can be bounded using martingale exponential inequality[1].

Example 2. Adopt the same setting as in Example 1, it is clear that $\varepsilon_{\mathcal{H}} = 0$ and

$$V(\psi(\pi^*), \pi^*) - V(\psi(\pi^k), \pi^k) = 1 - 1 = 0.$$

2 Analysis of the GEC term

Example 3. Adopt the same setting as in Example 1, it is clear that

$$\sum_{k=1}^{K} V(\psi(\pi^k), \pi^k) - V(\psi(\pi^k), \pi^*) = K(1-0) = K.$$

The GEC term is bounded by the GEC assumption, i.e., there exist an $d(\varepsilon) > 0$, such that for all $\{\pi^k\} \subset \mathcal{H}$ and $\{\mu^k\} \subset \mathcal{U}$,

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

where $\varphi(\alpha, \varepsilon, H, K) = d(\varepsilon) / (2\alpha) + \sqrt{d(\varepsilon) HK} + \varepsilon HK$ and

$$\mathcal{L}_{\text{train}}(\pi^k) = \sum_{h,k} \sum_{s=1}^{k-1} \mathbb{E}_{\xi_h \sim \mu^s}[\ell(\pi^k; \xi_h)].$$

Given $\xi_h \triangleq (s_h, a_h, r_h, s_{h+1})$, ℓ is defined as

$$\ell(\pi^k; \xi_h) \triangleq D_H(\mathbb{P}(\cdot \mid s_h, a_h, \pi^k(s_h)), \mathbb{P}(\cdot \mid s_h, a_h, \pi^*(s_h))),$$

where $D_H(\cdot \mid \cdot)$ denotes the Hellinger distance. Intuitively, the low GEC assumption states that, in the long run, if the hypothesis $\{\pi^k\}$ has a small in-sample training error, i.e., the term

$$\mathbb{E}_{\xi_h \sim \mu^s}[\ell(\pi^k; \xi_h)]$$

is small. Then, the prediction error $V(\psi(\pi^k), \pi^k) - V(\psi(\pi^k), \pi^*)$ will also be small. However, the prediction error can never be small in the game defined in Example 1. Then, what is going wrong?

The original wrong proof in our AAMAS paper cancled out the term $\mathbb{E}_{\xi_h \sim \mu^s}[\ell(\pi^k; \xi_h)]$, the term is correct on the upper bound of the GEC term. However, the upper bound of the type term does not depends on $-\mathbb{E}_{\xi_h \sim \mu^s}[\ell(\pi^k; \xi_h)]$ when $\pi^* \notin \mathcal{H}$, thus, cannot cancle out $\mathbb{E}_{\xi_h \sim \mu^s}[\ell(\pi^k; \xi_h)]$ part in the upper bound of GEC term. To see this, recall that when $\pi^* \notin \mathcal{H}$, in the type term, we are trying to upper bound

$$\sum_{h=1}^{H} L_h^{k-1}(\pi_{\mathcal{H}}) - \sum_{h=1}^{H} L_h^{k-1}(\pi^k)$$

1. The martingale exponential inequality does not require π^* or $\pi_{\mathcal{H}}$ belongs to $\mathcal{H}$.

using martingale exponential inequality. However, when $\pi^* \in \mathcal{H}$, we are trying to upperbound

$$\sum_{h=1}^{H} L_h^{k-1}(\pi^*) - \sum_{h=1}^{H} L_h^{k-1}(\pi^k).$$

The upper bound of the later can cancle out the term $\mathbb{E}_{\xi_h \sim \mu^s}[\ell(\pi^k; \xi_h)]$ while the former cannot. One way to fix the proof is decompose $\sum_{h=1}^{H} L_h^{k-1}(\pi_{\mathcal{H}}) - \sum_{h=1}^{H} L_h^{k-1}(\pi^k)$ into the sum of $\sum_{h=1}^{H} L_h^{k-1}(\pi^*) - \sum_{h=1}^{H} L_h^{k-1}(\pi^k)$ and

$$\sum_{h=1}^{H} L_h^{k-1}(\pi_{\mathcal{H}}) - \sum_{h=1}^{H} L_h^{k-1}(\pi^*), \tag{4}$$

then, we upper bound the term (4). To do this, we need to introduce a new metric to measure the difference of the model induced by these policies.

3 Additional Issue

In the previous section, we show that a new metric should be introduced to upper bound the term (4). However, the term (4) is 0 in Example 1. This is because our current definition of L_h^k only consider difference in the transition kernel while ignoring the difference in the reward function. Formally, we need to redefine L to the form as

$$L_h^k(\pi) \triangleq -\log \mathbb{P}(s_{h+1}^k |\, s_h^k . a_h^k, \pi(s_h^k)) + \lambda |r(s_{h+1}^k |\, s_h^k . a_h^k, \pi(s_h^k)) - r(s_{h+1}^k |\, s_h^k . a_h^k, \pi^*(s_h^k))|.$$