A Near-optimal High-probability Swap-Regret Upper Bound for Multi-agent Bandits in Unknown General-sum Games (Supplementary material)

Zhiming Huang[1] **Jianping Pan**[1]

[1]Department of Computer Science, University of Victoria, BC, Canada

We start by introducing the notations that will be used in the proofs of Lemma 5.1 and Theorem 5.3. As the proofs are for each individual agent n, without confusion, we drop the subscript n in some notations for brevity.

Recall that $\mathcal{G}_t$ the σ-algebra generated by the history information of all agents till round t, i.e., $\mathcal{G}_t := \sigma\left(\{a_n^1, r_n^1, \dots, a_n^t, r_n^t\}_{n\in\mathcal{N}}\right)$ and let $\mathbf{E}_t[\cdot] := \mathbf{E}[\cdot|\mathcal{G}_t]$ be the expectation conditioned on the history information by the end of round t. Recall that $y_a^t := 1 - u_n^t(a; \mathbb{A}_{-n}^t)$ is the instantaneous loss function if agent n plays arm $a \in A_n$ in round t, and thus $Y_{a,a'}^t := \frac{\mathbf{1}[a_n^t = a'] p_a^t q_{a,a'}^t y_{a'}^t}{p_{a'}^t}$ and $\hat{Y}_{a,a'}^t = \frac{Y_{a,a'}^t}{q_{a,a'}^t + \gamma_t}$. Denote by $\hat{L}_a^t := \sum_{t=1}^{T} \sum_{a'\in A_n} q_{a,a'}^t \hat{Y}_{a,a'}^t$ and $L_a^T := \sum_{t=1}^{T} \sum_{a'\in A_n} Y_{a,a'}^t$.

A PROOF OF LEMMA 5.1

Proof. Recall that $\tilde{Y}_{a,a'}^t := \mathbf{1}[a_n^t = a] y_{a'}^t$. We first prove that the process $\{Z_t\}_{t\geq 0}$, where $Z_t := \exp\left\{\sum_{s=1}^{t} \sum_{a\in A_n} \sum_{a'\in A_n} \beta_{a,a'}^s \left(\hat{Y}_{a,a'}^s - \tilde{Y}_{a,a'}^s\right)\right\}$ for $t > 0$ and $Z_0 = 1$, is a supermartingale with respect to filtration $\{\mathcal{G}_t\}_{t\geq 0}$ for all $a \in A_n$, i.e., $\mathbf{E}\left[Z_t|\mathcal{G}_{t-1}\right] \leq Z_{t-1}$. Denote by $\mathbb{A}_{-n}^t$ the actions of all agents except for agent n in round t. Then, we have that

$$
\begin{aligned}
&\mathbf{E}_{t-1}\left[\exp\left\{\sum_{a\in A_n}\sum_{a'\in A_n} \beta_{a,a'}^t \left(\hat{Y}_{a,a'}^t - \tilde{Y}_{a,a'}^t\right)\right\}\right] = \mathbf{E}_{t-1}\left[\frac{\exp\left\{\sum\limits_{a\in A_n}\sum\limits_{a'\in A_n} \beta_{a,a'}^t \hat{Y}_{a,a'}^t\right\}}{\exp\left\{\sum\limits_{a\in A_n}\sum\limits_{a'\in A_n} \beta_{a,a'}^t \tilde{Y}_{a,a'}^t\right\}}\right] = \mathbf{E}_{t-1}\left[\frac{\exp\left\{\sum\limits_{a\in A_n}\sum\limits_{a'\in A_n} \beta_{a,a'}^t \hat{Y}_{a,a'}^t\right\}}{\exp\left\{\sum\limits_{a\in A_n}\sum\limits_{a'\in A_n} \beta_{a,a'}^t \mathbf{1}[a_n^t = a] y_{a'}^t\right\}}\right] \\
&= \mathbf{E}_{t-1}\left[\mathbf{E}_{t-1}\left[\frac{\exp\left\{\sum\limits_{a\in A_n}\sum\limits_{a'\in A_n} \beta_{a,a'}^t \hat{Y}_{a,a'}^t\right\}}{\exp\left\{\sum\limits_{a\in A_n}\sum\limits_{a'\in A_n} \beta_{a,a'}^t \mathbf{1}[a_n^t = a] y_{a'}^t\right\}} \mid \mathbb{A}_{-n}^t\right]\right] \leq \mathbf{E}_{t-1}\left[\frac{\mathbf{E}_{t-1}\left[\exp\left\{\sum\limits_{a\in A_n}\sum\limits_{a'\in A_n} \beta_{a,a'}^t \hat{Y}_{a,a'}^t\right\} \mid \mathbb{A}_{-n}^t\right]}{\exp\left\{\sum\limits_{a\in A_n}\sum\limits_{a'\in A_n} \beta_{a,a'}^t p_a^t y_{a'}^t\right\}}\right],
\end{aligned}
\tag{11}
$$

where the third equality is due to the law of total expectation, and the fourth equality is due to that $y_{a'}^t$ is determined given $\mathbb{A}_{-n}^t$ and $\beta_{a,a'}^t$ is $\mathcal{G}_{t-1}$-measurable. Denote by $\mathbf{E}_{n,t-1}[\cdot] := \mathbf{E}_{t-1}\left[\cdot \mid \mathbb{A}_{-n}^t\right]$. Then, we show that $\mathbf{E}_{n,t-1}\left[\exp\left\{\sum_{a\in A_n}\sum_{a'\in A_n} \beta_{a,a'}^t \hat{Y}_{a,a'}^t\right\}\right] \leq \exp\left\{\sum_{a\in A_n}\sum_{a'\in A_n} \beta_{a,a'}^t p_a^t y_{a'}^t\right\}$ as follows:

Accepted for the 39th Conference on Uncertainty in Artificial Intelligence (UAI 2023).

$$
\begin{aligned}
&\mathbf{E}_{n,t-1}\left[\exp\left\{\sum_{a\in A_n}\sum_{a'\in A_n}\beta^t_{a,a'}\hat{Y}^t_{a,a'}\right\}\right] = \mathbf{E}_{n,t-1}\left[\exp\left\{\sum_{a\in A_n}\sum_{a'\in A_n}\beta^t_{a,a'}\frac{p^t_a\mathbf{1}[a^t_n=a']q^t_{a,a'}y^t_{a'}}{p^t_{a'}(q^t_{a,a'}+\gamma_t)}\right\}\right]\\
&\leq \mathbf{E}_{n,t-1}\left[\sum_{a\in A_n}p^t_a\exp\left\{\sum_{a'\in A_n}\beta^t_{a,a'}\frac{\mathbf{1}[a^t_n=a']q^t_{a,a'}y^t_{a'}}{p^t_{a'}(q^t_{a,a'}+\gamma_t)}\right\}\right] \leq \mathbf{E}_{n,t-1}\left[\sum_{a\in A_n}p^t_a\exp\left\{\sum_{a'\in A_n}\frac{\beta^t_{a,a'}}{2\gamma_t}\frac{2\gamma_t\mathbf{1}[a^t_n=a']q^t_{a,a'}y^t_{a'}}{p^t_{a'}(q^t_{a,a'}+\gamma_t\mathbf{1}[a^t_n=a']q^t_{a,a'}y^t_{a'})}\right\}\right]\\
&= \mathbf{E}_{n,t-1}\left[\sum_{a\in A_n}p^t_a\exp\left\{\sum_{a'\in A_n}\frac{\beta^t_{a,a'}}{2\gamma_t}\frac{2\gamma_t\mathbf{1}[a^t_n=a']y^t_{a'}/p^t_{a'}}{1+\gamma_t\mathbf{1}[a^t_n=a']y^t_{a'}/p^t_{a'}}\right\}\right] \leq \mathbf{E}_{n,t-1}\left[\sum_{a\in A_n}p^t_a\exp\left\{\sum_{a'\in A_n}\frac{\beta^t_{a,a'}}{2\gamma_t}\log(1+2\gamma_t\mathbf{1}[a^t_n=a']y^t_{a'}/p^t_{a'})\right\}\right]\\
&\leq \mathbf{E}_{n,t-1}\left[\sum_{a\in A_n}p^t_a\exp\left\{\sum_{a'\in A_n}\log(1+\beta^t_{a,a'}\mathbf{1}[a^t_n=a']y^t_{a'}/p^t_{a'})\right\}\right] = \mathbf{E}_{n,t-1}\left[\sum_{a\in A_n}p^t_a\prod_{a'\in A_n}(1+\beta^t_{a,a'}\mathbf{1}[a^t_n=a']y^t_{a'}/p^t_{a'})\right].
\end{aligned}
$$

where the first inequality is due to Jensen's inequality, the second inequality is due to that $0\leq\mathbf{1}[a^t_n=a']q^t_{a,a'}y^t_{a'}\leq 1$, the third inequality is due to the fact that $\frac{z}{1+z/2}\leq\log(1+z)$ for all $z>0$, and the last inequality is due to the inequality $x\log(1+y)\leq\log(1+xy)$ for all $y>-1$ and $x\in[0,1]$. As $\mathbf{1}[a^t_n=a']\mathbf{1}[a^t_n=a'']=0$ for any $a'\neq a''$, the last term in above equation can be further processed as follows:

$$
\begin{aligned}
&\mathbf{E}_{n,t-1}\left[\sum_{a\in A_n}p^t_a\prod_{a'\in A_n}(1+\beta^t_{a,a'}\mathbf{1}[a^t_n=a']y^t_{a'}/p^t_{a'})\right] = \mathbf{E}_{n,t-1}\left[\sum_{a\in A_n}p^t_a(1+\sum_{a'\in A_n}\beta^t_{a,a'}\mathbf{1}[a^t_n=a']y^t_{a'}/p^t_{a'})\right]\\
&= \mathbf{E}_{n,t-1}\left[1+\sum_{a\in A_n}\sum_{a'\in A_n}p^t_a\beta^t_{a,a'}\mathbf{1}[a^t_n=a']y^t_{a'}/p^t_{a'}\right] = 1+\sum_{a\in A_n}\sum_{a'\in A_n}p^t_a\beta^t_{a,a'}y^t_{a'} \leq \exp\left\{\sum_{a\in A_n}\sum_{a'\in A_n}\beta^t_{a,a'}p^t_a y^t_{a'}\right\},
\end{aligned}
$$

where the inequality is due to $1+x\leq\exp\{x\}$ for any $x\in\mathbb{R}$. Therefore, we have shown that $\mathbf{E}_{n,t-1}\left[\exp\left\{\sum_{a\in A_n}\sum_{a'\in A_n}\beta^t_{a,a'}\hat{Y}^t_{a,a'}\right\}\right]\leq\exp\left\{\sum_{a\in A_n}\sum_{a'\in A_n}\beta^t_{a,a'}p^t_a y^t_{a'}\right\}$, which indicates that (11) is bounded by 1. Thus,

$$
\mathbf{E}_{t-1}\left[Z_t\right] = \mathbf{E}_{t-1}\left[\exp\left\{\sum_{a\in A_n}\sum_{a'\in A_n}\beta^t_{a,a'}\left(\hat{Y}^t_{a,a'}-\tilde{Y}^t_{a,a'}\right)\right\}\right]\cdot Z_{t-1}\leq Z_{t-1},
$$

which shows that $\{Z_t\}_{t\geq 0}$ is a supermartingale with respect to filtration $\{\mathcal{G}_t\}_{t\geq 0}$. Thus, we have $\mathbf{E}\left[Z_T\right]\leq\mathbf{E}\left[Z_{T-1}\right]\ldots\leq\mathbf{E}\left[Z_0\right]=1$. By the Markov inequality, we have

$$
\begin{aligned}
\Pr\left(\sum_{t=1}^{T}\beta^t_{a,a'}\sum_{a\in A_n}\sum_{a'\in A_n}\left(\hat{Y}^t_{a,a'}-\tilde{Y}^t_{a,a'}\right)\geq\epsilon\right) &\leq \mathbf{E}\left[\exp\left\{\sum_{t=1}^{T}\beta^t_{a,a'}\sum_{a\in A_n}\sum_{a'\in A_n}\left(\hat{Y}^t_{a,a'}-\tilde{Y}^t_{a,a'}\right)\geq\epsilon\right\}\right]\cdot\exp\{-\epsilon\}\\
&\leq \exp\{-\epsilon\}.
\end{aligned}
$$

Then, the lemma follows by solving $\exp\{-\epsilon\}=\delta$ for ϵ.

□

B PROOF OF THEOREM 5.3

Proof. By the relationship between P^t_n and Q^t_a, we have the following equation held:

$$
\begin{aligned}
\sum_{a\in A_n}L^T_a &= \sum_{a\in A_n}\sum_{t=1}^{T}\sum_{a'\in A_n}Y^t_{a,a'} = \sum_{t=1}^{T}\sum_{a'\in A_n}\sum_{a\in A_n}\frac{\mathbf{1}[a^t_n=a']p^t_a q^t_{a,a'}}{p^t_{a'}}y^t_{a'}\\
&= \sum_{t=1}^{T}\sum_{a'\in A_n}\mathbf{1}[a^t_n=a']y^t_{a'} = \sum_{t=1}^{T}\sum_{a\in A_n}\mathbf{1}[a^t_n=a]y^t_a,
\end{aligned}
\tag{12}
$$

The regret defined in (3) can be rewritten in the loss form and can be decomposed as follows:

$$
\begin{aligned}
R_n^{\text{swa}}(T,\mathcal{F}) &= \max_{F\in\mathcal{F}} \sum_{t=1}^{T}\sum_{a\in A_n} \mathbf{1}[a_n^t = a] y_a^t - \sum_{t=1}^{T}\sum_{a\in A_n} \mathbf{1}[a_n^t = a] y_{F(a)}^t \\
&= \max_{F\in\mathcal{F}} \sum_{a\in A_n} L_a^T - \sum_{a\in A_n} \tilde{L}_{a,F(a)}^T = \underbrace{\sum_{a\in A_n} (L_a^T - \hat{L}_a^T)}_{=:(\text{a})} + \underbrace{\sum_{a\in A_n} (\hat{L}_a^T - \hat{L}_{a,F(a)}^T)}_{=:(\text{b})} + \underbrace{\sum_{a\in A_n} (\hat{L}_{a,F(a)}^T - \tilde{L}_{a,F(a)}^T)}_{=:(c)},
\end{aligned}
\tag{13}
$$

where the second equality is due to (12) and the definition of $\tilde{L}_{a,F(a)}^T := \sum_{t=1}^{T} \mathbf{1}[a_n^t = a] y_{F(a)}^t$.

We first show how to bound (a). By definition of L_a^T and $\hat{L}_a^T$, we have that

$$
L_a^T - \hat{L}_a^T = \sum_{t=1}^{T}\sum_{a'\in A_n} Y_{a,a'}^t - \sum_{t=1}^{T}\sum_{a'\in A_n} q_{a,a'}^t \hat{Y}_{a,a'}^t = \sum_{t=1}^{T}\sum_{a'\in A_n} Y_{a,a'}^t \left(1 - \frac{q_{a,a'}^t}{q_{a,a'}^t + \gamma_t}\right) = \sum_{t=1}^{T} \gamma_t \sum_{a'\in A_n} \hat{Y}_{a,a'}^t .
$$

Thus, (a) is bounded by $\sum_{t=1}^{T} \gamma_t \sum_{a\in A_n}\sum_{a'\in A_n} \hat{Y}_{a,a'}^t$.

Then, we show how to bound (b). Let $W_n^t := \prod_{a\in A_n}\sum_{a'\in A_n} \exp(-\eta_{t+1}\hat{L}_{a,a'}^t)$, and we have that $W_n^0 = \prod_{a\in A_n}\sum_{a'\in A_n} \exp(0) = (K_n)^{K_n}$. Note that $W_n^T = W_n^0 \frac{W_n^1}{W_n^0}\cdots\frac{W_n^T}{W_n^{T-1}} = (K_n)^{K_n}\prod_{t=1}^{T}\frac{W_n^t}{W_n^{t-1}}$. Then we have

$$
\exp(-\sum_{a\in A_n} \eta_{T+1}\hat{L}_{a,F(a)}^T) = \prod_{a\in A_n} \exp(-\eta_{T+1}\hat{L}_{a,F(a)}^T) \le \prod_{a\in A_n}\sum_{a'\in A_n} \exp(-\eta_{T+1}\hat{L}_{a,a'}^t) = (K_n)^{K_n}\prod_{t=1}^{T}\frac{W_n^t}{W_n^{t-1}},
\tag{14}
$$

where the inequality is due to that $\exp(-\eta_T \hat{L}_{w,w'}^T) \ge 0$. Then, by the definition of $q_{w,w'}^t$ in (5), we obtain that

$$
\begin{aligned}
\frac{W_n^t}{W_n^{t-1}} &= \frac{\prod_{a\in A_n}\sum_{a'\in A_n} \exp(-\eta_t\hat{L}_{a,a'}^{t-1})\exp(-\eta_t\hat{Y}_{a,a'}^t)}{\prod_{a\in A_n}\sum_{a'\in A_n} \exp(-\eta_t\hat{L}_{a,a'}^{t-1})} \\
&= \prod_{a\in A_n}\sum_{a'\in A_n} \frac{\exp(-\eta_t\hat{L}_{a,a'}^{t-1})}{\sum_{a'\in A_n}\exp(-\eta_t\hat{L}_{a,a'}^{t-1})}\exp(-\eta_t\hat{Y}_{a,a'}^t) \\
&= \prod_{a\in A_n}\sum_{a'\in A_n} q_{a,a'}^t \exp(-\eta_t\hat{Y}_{a,a'}^t) \le \prod_{a\in A_n}\sum_{a'\in A_n} q_{a,a'}^t \exp(-\eta_T\hat{Y}_{a,a'}^t) \\
&\le \prod_{a\in A_n}\left(\sum_{a'\in A_n} q_{a,a'}^t - \eta_T\sum_{a'\in A_n} q_{a,a'}^t\hat{Y}_{a,a'}^t + \frac{\eta_T^2}{2}\sum_{a'\in A_n} q_{a,a'}^t(\hat{Y}_{a,a'}^t)^2\right) \\
&\le \prod_{a\in A_n}\exp\left(-\eta_T\sum_{a'\in A_n} q_{a,a'}^t\hat{Y}_{a,a'}^t + \frac{\eta_T^2}{2}\sum_{a'\in A_n} q_{a,a'}^t(\hat{Y}_{a,a'}^t)^2\right) \\
&= \exp\left(-\eta_T\sum_{a\in A_n}\sum_{a'\in A_n} q_{a,a'}^t\hat{Y}_{a,a'}^t + \frac{\eta_T^2}{2}\sum_{a\in A_n}\sum_{a'\in A_n} q_{a,a'}^t(\hat{Y}_{a,a'}^t)^2\right),
\end{aligned}
\tag{15}
$$

where the first inequality is due to that η_t is a non-increasing parameter, the second inequality is due to that $\exp(x) \le 1 + x + \frac{x^2}{2}$ for any $x \le 0$, and the third inequality is due to that $1 + x \le \exp(x)$ for any $x \in \mathbb{R}$. Combining (15) and (14), and taking the logarithm for both sides of the above inequality, we have that

$$
-\sum_{a\in A_n}\eta_T\hat{L}_{a,F(a)}^T \le K_n\log(K_n) - \sum_{a\in A_n}\eta_T \underbrace{\sum_{t=1}^{T}\sum_{a'\in A_n} q_{a,a'}^t\hat{Y}_{a,a'}^t}_{=:\hat{L}_a^T\ (\text{by definition of } \hat{L}_a^T)} + \frac{\eta_T^2}{2}\sum_{t=1}^{T}\sum_{a\in A_n}\sum_{a'\in A_n} q_{a,a'}^t\left(\hat{Y}_{a,a'}^t\right)^2 .
$$

Dividing both sides by $\eta_T > 0$, with rearrangement, we have

$$\begin{aligned}\sum_{a\in A_n} \hat{L}_a^T - \sum_{a\in A_n} \hat{L}_{a,F(a)}^T &\leq \frac{K_n \log(K_n)}{\eta_T} + \frac{\eta_T}{2}\sum_{t=1}^{T}\sum_{a\in A_n}\sum_{a'\in A_n} q_{a,a'}^t \left(\hat{Y}_{a,a'}^t\right)^2 \\ &\leq \frac{K_n \log(K_n)}{\eta_T} + \sum_{t=1}^{T}\frac{\eta_t}{2}\sum_{a\in A_n}\sum_{a'\in A_n} \hat{Y}_{a,a'}^t,\end{aligned} \tag{16}$$

where the second inequality is due to that η_t is a non-increasing parameter and the fact that $q_{a,a'}^t \hat{Y}_{a,a'}^t \leq 1$. Combining with the bound of (a), we have

$$\sum_{a\in A_n}\left(L_a^T - \tilde{L}_{a,F(a)}^T\right) \leq \frac{K_n \log(K_n)}{\eta_T} + \sum_{t=1}^{T}\left(\frac{\eta_t}{2}+\gamma_t\right)\sum_{a\in A_n}\sum_{a'\in A_n}\hat{Y}_{a,a'}^t + \sum_{a\in A_n}\left(\hat{L}_{a,F(a)}^T - \tilde{L}_{a,F(a)}\right).$$

Let $\gamma_t = \eta_t/2$. By invoking Lemma 5.2, with probability at least $1-\delta$, we have the following inequality held:

$$\begin{aligned}\sum_{a\in A_n}\left(L_a^t - \tilde{L}_{a,a'}^T\right) &\leq \frac{K_n \log(K_n)}{\eta_T} + \sum_{t=1}^{T}\eta_t\left(\sum_{a\in A_n}\sum_{a'\in A_n}\tilde{Y}_{a,a'}^t\right) + \log(\frac{1}{\delta}) + \frac{1}{\eta_T}\log(\frac{K_n^{K_n}}{\delta}) \\ &\leq \frac{K_n\log(K_n) + K_n\log(K_n/\delta)}{\eta_T} + \sum_{t=1}^{T}\eta_t K_n + \log(\frac{1}{\delta}),\end{aligned}$$

where the last inequality is due to that $\sum\limits_{a\in A_n}\sum\limits_{a'\in A_n}\tilde{Y}_{a,a'}^t = \sum\limits_{a\in A_n}\sum\limits_{a'\in A_n}\mathbf{1}[a_n^t = a]y_{a'}^t \leq K_n$ and $\log(\frac{K_n^{K_n}}{\delta}) \leq K_n\log(K_n/\delta)$ for $\delta \in (0,1)$.

Letting $\eta_t = \sqrt{\frac{\log(K_n)}{t}}$, we have

$$R_n^T(T,\mathcal{F}) \leq 2K_n\sqrt{T\log(K_n)} + K_n\sqrt{\log(K_n)}\sum_{t=1}^{T}\sqrt{\frac{1}{t}} + \left(1 + K_n\sqrt{\frac{T}{\log K_n}}\right)\log(\frac{1}{\delta}).$$

When $\eta_t = \sqrt{\frac{\log(K_n)+\log(K_n/\delta)}{t}}$, the above inequality becomes

$$R_n^T(T,\mathcal{F}) \leq K_n\sqrt{T(\log(K_n)+\log(K_n/\delta))} + K_n\sqrt{(\log(K_n)+\log(K_n/\delta)}\sum_{t=1}^{T}\frac{1}{t} + \log(\frac{1}{\delta}).$$

Theorem 5.3 follows by $\sum\limits_{t=1}^{T}\sqrt{\frac{1}{t}} \leq 2\sqrt{T}$. $\square$