Supplementary Material of Global Multi-armed Bandits with Hölder Continuity

Onur Atan
UCLA

Cem Tekin
UCLA

Mihaela van der Schaar
UCLA

Lemma 1. *Given any $\theta_* \in \Theta$, there exists a constant $\epsilon_{\theta_*} = \delta_{\min}(\theta_*)^{1/\gamma_2}/(2D_2)^{1/\gamma_2}$, where D_2 and γ_2 are the constants given in Assumption 1 such that $\Delta_{\min}(\theta_*) \geq \epsilon_{\theta_*}$. In other words, the minimum suboptimality distance is always positive.*

Proof. For any suboptimal arm $k \in \mathcal{K} - k^*(\theta)$, we have $\mu_{k^*(\theta)}(\theta) - \mu_k(\theta) \geq \delta_{\min}(\theta) > 0$. We also know that $\mu_k(\theta') \geq \mu_{k^*(\theta)}(\theta')$ for all $\theta' \in \Theta_k$. Hence for any $\theta' \in \Theta_k$ at least one of the following should hold: (i) $\mu_k(\theta') \geq \mu_k(\theta) - \delta_{\min}(\theta)/2$, (ii) $\mu_{k^*(\theta)}(\theta') \leq \mu_{k^*(\theta)}(\theta) + \delta_{\min}(\theta)/2$. If both of the below does not hold, then we must have $\mu_k(\theta') < \mu_{k^*(\theta)}(\theta')$, which is false. This implies that we either have $\mu_k(\theta) - \mu_k(\theta') \leq \delta_{\min}(\theta)/2$ or $\mu_{k^*(\theta)}(\theta) - \mu_{k^*(\theta)}(\theta') \geq -\delta_{\min}(\theta)/2$, or both. Recall that from Assumption 1 we have $|\theta - \theta'| \geq |\mu_k(\theta) - \mu_k(\theta')|^{1/\gamma_2}/D_2^{1/\gamma_2}$. This implies that $|\theta - \theta'| \geq \epsilon_\theta$ for all $\theta' \in \Theta_k$. □

Lemma 2. *Consider a run of the greedy policy until time t. Then, the following relation between $\hat{\theta}_t$ and θ_* holds with probability one: $|\hat{\theta}_t - \theta_*| \leq \sum_{k=1}^K w_k(t) D_1 |\hat{X}_{k,t} - \mu_k(\theta_*)|^{\gamma_1}$*

Proof. Assumption 2 ensures that the reward functions are either monotonically increasing or decreasing. We generate imaginary functions that are $\mu_k(\theta) = \tilde{\mu}_k(\theta)$ for $\theta \in \Theta$ and for $y, y' \in [0,1]$,

$$|\tilde{\mu}_k^{-1}(y) - \tilde{\mu}_k^{-1}(y')| \leq D_1 |y - y'|^{\gamma_1} \tag{1}$$

We have also $\tilde{\mu}_k^{-1}(y) > 1$ when $y > \max_{\theta \in \Theta} \mu_k(\theta)$ and $\tilde{\mu}_k^{-1}(y) < 0$ when $y < \min_{\theta \in \Theta} \mu_k(\theta)$.Then,

$$
\begin{aligned}
|\theta_* - \hat{\theta}_t| &= |\sum_{k=1}^K w_k(t) \hat{\theta}_{k,t} - \theta_*| \\
&= \sum_{k=1}^K w_k(t) |\theta_* - \hat{\theta}_{k,t}| \\
&\leq \sum_{k=1}^K w_k(t) |\tilde{\mu}_k^{-1}(\hat{X}_{k,t}) - \tilde{\mu}_k^{-1}(\tilde{\mu}_k(\theta_*))| \\
&\leq \sum_{k=1}^K w_k(t) D_1 |\hat{X}_{k,t} - \mu_k(\theta_*)|^{\gamma_1},
\end{aligned} \tag{2}
$$

where we need to look at following two cases for the first inequality. The first case is $\hat{X}_{k,t} \in \mathcal{Y}_k$ where the statement immediately follows. The second case is $\hat{X}_{k,t} \notin \mathcal{Y}_k$, the global parameter estimator $\hat{\theta}_{k,t}$ is either 0 or 1. □

Lemma 3. *For given global parameter θ_*, the one step regret of the greedy policy is bounded by $r_t(\theta_*) = \mu^*(\theta_*) - \mu_{I_t}(\theta_*) \leq 2D_2 |\theta_* - \hat{\theta}_t|^{\gamma_2}$ with probability 1, where I_t is the arm selected by the greedy policy at time $t \geq 2$.*

Proof. Note that $I_t \in \arg\max_{k \in \mathcal{K}} \mu_k(\hat{\theta}_t)$. Therefore, we have

$$\mu_{I_t}(\hat{\theta}_t) - \mu_{k^*(\theta_*)}(\hat{\theta}_t) \geq 0. \tag{3}$$

We have $\mu^*(\theta_*) = \mu_{k^*(\theta_*)}(\theta_*)$. Then, we can bound

$$
\begin{aligned}
&\mu^*(\theta_*) - \mu_{I_t}(\theta_*) \\
&= \mu_{k^*(\theta_*)}(\theta_*) - \mu_{I_t}(\theta_*) \\
&\leq \mu_{k^*(\theta_*)}(\theta_*) - \mu_{I_t}(\theta_*) + \mu_{I_t}(\hat{\theta}_t) - \mu_{k^*(\theta_*)}(\hat{\theta}_t) \\
&= \mu_{k^*(\theta_*)}(\theta_*) - \mu_{k^*(\theta_*)}(\hat{\theta}_t) + \mu_{I_t}(\hat{\theta}_t) - \mu_{I_t}(\theta_*) \\
&\leq 2D_2 |\theta_* - \hat{\theta}_t|^{\gamma_2}
\end{aligned} \tag{4}
$$

, where the first inequality followed by inequality 3 and second inequality by Assumption 1. □

Lemma 4. *For any $t \geq 2$ and given global parameter θ_*, we have $\mathcal{G}^x_{\theta_*, \hat{\theta}_t} \subseteq \cup_{k=1}^K \mathcal{F}^k_{\theta_*, \hat{\theta}_t}((\frac{x}{D_1})^{\frac{1}{\gamma_1}})$ with probability 1.*

Proof.

$$
\begin{aligned}
&\{|\theta_* - \hat{\theta}_t| \geq x\} \\
&\subseteq \{\sum_{k=1}^K w_k(t) D_1 |\hat{X}_{k,t} - \mu_k(\theta_*)|^{\gamma_1} \geq x\} \\
&\subseteq \cup_{k=1}^K \{w_k(t) D_1 |\hat{X}_{k,t} - \mu_k(\theta_*)|^{\gamma_1} \geq w_k(t) x\}
\end{aligned}
$$

$$= \cup_{k=1}^{K} \{ |\hat{X}_{k,t} - \mu_k(\theta_*)| \geq (\frac{x}{D_1})^{\frac{1}{\gamma_1}} \} \quad (5)$$

, where the first inequality followed by Lemma 2 and second inequality by the fact that $\sum_{k=1}^{K} w_k(t) = 1$. □