Online Learning to Approach a Person with No-Regret: Supplementary Material

Hyemin Ahn, Yoonseon Oh, Sungjoon Choi, Claire J. Tomlin, and Songhwai Oh

I. Proof of Theorem 1

Lemma 1 (Lemma 5.1 in [1]): For $\delta \in (0,1)$, if $\beta_k = 2\log(|\mathcal{Q}|\pi_k/\delta)$, where $\sum \pi_k^{-1} = 1$ and $\pi_k = \pi^2 k^2/6$,

$$|\mathcal{P}(\mathbf{q}) - \mu_{k-1}(\mathbf{q})| \leq \beta_k^{1/2}\sigma_{k-1}(\mathbf{q})$$

$\forall \mathbf{q} \in \mathcal{Q}$, with probability $1-\delta$.

Lemma 2: If $|\mathcal{P}(\mathbf{q}) - \mu_{k-1}(\mathbf{q})| \leq \beta_k^{1/2}\sigma_{k-1}(\mathbf{q})$ $\forall \mathbf{q} \in \mathcal{Q}$,

$$r_k \leq \sum_{t=1}^{T_k} 2\beta_k^{1/2}\sigma_{k-1}(\xi_k(t)),$$

where $T_k = |\xi_k|$.

Proof: For ξ_k chosen at the kth round, the GP-UCB algorithm is applied such that:

$$\xi_k = \arg\max_{\xi\in\Xi} \sum_{t=1}^{|\xi|} \left(\mu_{k-1}(\xi(t)) + \beta_k^{\frac{1}{2}}\sigma_{k-1}(\xi(t))\right).$$

Therefore, it is clear that

$$\begin{aligned}
&\sum_{t=1}^{T_k}\big(\mu_{k-1}(\xi_k(t)) + \beta_k^{1/2}\sigma_{k-1}(\xi_k(t))\big) \\
&\qquad \geq \sum_{t=1}^{T^*}\big(\mu_{k-1}(\xi^*(t)) + \beta_k^{1/2}\sigma_{k-1}(\xi^*(t))\big) \geq f(\xi^*)
\end{aligned}$$

Hence, we have

$$\begin{aligned}
r_k &= f(\xi^*) - f(\xi_k) \\
&\leq \sum_{t=1}^{T_k}\big(\mu_{k-1}(\xi_k(t)) + \beta_t^{1/2}\sigma_{k-1}(\xi_k(t))\big) - f(\xi_k) \\
&\leq \sum_{t=1}^{T_k}\big(\mu_{k-1}(\xi_k(t)) - \mathcal{P}(\xi_k(t))\big) + \beta_t^{1/2}\sigma_{k-1}(\xi_k(t)) \\
&\leq \sum_{t=1}^{T_k} 2\beta_k^{1/2}\sigma_{k-1}(\xi_k(t))
\end{aligned}$$

■

Let $\Xi_k \in \Xi$ be a set of all k-combinations in Ξ. For $A \in \Xi_k$, we define $\mathbf{q}(A) = \cup_{\xi\in A}\cup_{t=1}^{|\xi|}\xi(t)$, the set of all states of all paths in A.

This work was supported in part by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2017R1A2B2006136) and by NSF CPS:Large:ActionWebs, award number 0931843.

H. Ahn, Y. Oh, S. Choi, and S. Oh are with the Department of Electrical and Computer Engineering and ASRI, Seoul National University, Seoul, Korea (e-mail: {hyemin.ahn, yoonseon.oh, sungjoon.choi}@cpslab.snu.ac.kr, songhwai@snu.ac.kr). C.J. Tomlin is with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, USA (e-mail: tomlin@berkeley.edu).

Lemma 3: For $\delta \in (0,1)$ and β_k defined as in Lemma 1, with probability at least $1-\delta$,

$$\sum_{k=1}^{K} r_k^2 \leq C_1\beta_K\gamma_K \tag{1}$$

where $C_1 = 8T_{\max}/\log(1+\sigma_\epsilon^{-2})$ and $\gamma_K = \max_{A\in\Xi_K}\mathbb{I}(p_{\mathbf{q}(A)};\mathcal{P}_{\mathbf{q}(A)})$ is the maximum information gain after K rounds. Here, $\mathcal{P}_{\mathbf{q}(A)}$ and $p_{\mathbf{q}(A)}$ are sets of comfort scores and corresponding observations at states in A, respectively.

Proof: From Lemma 2, we have

$$\begin{aligned}
r_k^2 &\leq \left(\sum_{t=1}^{T_k} 2\beta_k^{1/2}\sigma_{k-1}(\xi_k(t))\right)^2 \\
&\leq 4\beta_K\left(\sum_{t=1}^{T_k}\sigma_{k-1}(\xi_k(t))\right)^2 \leq 4\beta_K T_k\sum_{t=1}^{T_k}\sigma_{k-1}^2(\xi_k(t))
\end{aligned}$$

since β_k is nondecreasing. The last inequality is due to the Cauchy-Schwarz inequality. By defining $C_2 = \sigma_\epsilon^{-2}/\log(1+\sigma_\epsilon^{-2}) \geq 1$ as done in [1], we have

$$\begin{aligned}
r_k^2 &\leq 4\beta_K T_k\sigma_\epsilon^2\sum_{t=1}^{T_k}\sigma_\epsilon^{-2}\sigma_{k-1}^2(\xi_k(t)) \\
&\leq 4\beta_K T_k\sigma_\epsilon^2\left(\sum_{t=1}^{T_k} C_2\log(1+\sigma_\epsilon^{-2}\sigma_{k-1}^2(\xi_k(t)))\right) \\
&= 8\sigma_\epsilon^2 C_2 T_k\beta_K\left(\frac{1}{2}\sum_{t=1}^{T_k}\log(1+\sigma_\epsilon^{-2}\sigma_{k-1}^2(\xi_k(t)))\right)
\end{aligned}$$

Using Lemma 5.3 in [1], for $A_k \in \Xi_k$, we have

$$\mathbb{I}(p_{\mathbf{q}(A_k)};\mathcal{P}_{\mathbf{q}(A_k)}) = \sum_{\xi\in A_k}\left(\frac{1}{2}\sum_{t=1}^{|\xi|}\log(1+\sigma_\epsilon^{-2}\sigma_{k-1}^2(\xi(t)))\right)$$

Noting that $|A_k| = k$, we arrive at

$$\sum_{k=1}^{K} r_k^2 \leq 8\sigma_\epsilon^2 C_2 T_{\max}\beta_K\mathbb{I}\big(p_{\mathbf{q}(A_K)};\mathcal{P}_{\mathbf{q}(A_K)}\big) \leq C_1\beta_K\gamma_K$$

Lastly, C_1 can be simplified to $C_1 = 8T_{\max}/\log(1+\sigma_\epsilon^{-2})$

■

Since $R_K^2 \leq K\sum_{k=1}^{K} r_k^2$ using the Cauchy-Schwarz inequality, Theorem 1 has been proven.

References

[1] N. Srinivas, A. Krause, S. M. Kakade, and M. Seeger, "Gaussian process optimization in the bandit setting: No regret and experimental design," in *Proc. of the International Conference on Machine Learning*, 2010.