On the Finite-Time Performance of the Knowledge Gradient Algorithm

Yanwen Li[1] & Siyang Gao[1,2]

[1] Department of Advanced Design and Systems Engineering, City University of Hong Kong
[2] School of Data Science, City University of Hong Kong

The 39th International Conference on Machine Learning (ICML2022)

Fixed-Budget Best Arm Identification (BAI) Problem

- **Parameters available to the agent:** the number of arms k and the number of rounds n.
- **Parameters unknown to the agent:** the mean rewards of the underlying distributions $\mathcal{D}_1,\ldots,\mathcal{D}_k$ corresponding to the arms $\{1,\ldots,k\}$. In this research, we assume that the arm b with the largest mean reward is **unique**.
- In each round $t = 1,\ldots,n$,
  (1) The agent pulls an arm $I_t \in \{1,\ldots,k\}$.
  (2) A noisy reward $X_{I_t,t}$ is drawn from $\mathcal{D}_{I_t}$ (independently from the past given I_t).

  After the n rounds, the agent recommends an arm $J_n \in \{1,\ldots,k\}$.
- **Goal:** Identify the best arm b, i.e., the arm with the largest mean reward.

Methods for Solving Fixed-Budget BAI:

- Knowledge Gradient (**KG**): Gupta & Miescke (1996); Frazier et al. (2008)
- Expected Improvement (**EI**): Jones et al. (1998); Ryzhov (2016)
- Optimal Computing Budget Allocation (**OCBA**): Chen et al. (2000); Gao & Shi (2015); Li & Gao (2022)
- Upper Confidence Bound Exploration (**UCB-E**): Audibert et al. (2010)
- Successive Rejects (**SR**): Audibert et al. (2010); Bubeck et al. (2013)
- Gap-Based Exploration (**GapE**): Gabillon et al. (2011)
- Top-Two Thompson Sampling (**TTTS**): Russo (2016); Shang et al. (2019)
- ......

Research on Knowledge Gradient:

- **Excellent empirical performance:** Frazier et al. (2008, 2009); Frazier (2009); Powell (2011); Wang & Powell (2018); ......
- **Wide application range:**
  - drug discovery (Negoescu et al., 2011)
  - urban delivery (Huang et al., 2019)
  - risk quantification (Cakmak et al., 2020)
  - experimental design in material science (Chen et al., 2015) and biotechnology (Li et al., 2018)
  - ......
- **Extension to other types of BAI problems:**
  - parallel BAI (Wu & Frazier, 2016)
  - contextual bandits (Ding et al., 2021)
  - ......
- **Few theoretical results:** Frazier et al. (2008); Ryzhov (2016); Wang & Powell (2018)

Summary of the Talk

- We study the finite-time performance of the KG algorithm under independent and normally distributed rewards with known variances.
- We derive the **upper and lower bounds** of the **proportions of rounds** allocated to pull the arms. With these bounds, existing asymptotic results become simple corollaries.
- We derive the **upper and lower bounds** of the KG algorithm on the **commonly used performance measures** for best arm identification and multi-armed bandits.

Algorithm Description:

Algorithm 1 Knowledge Gradient

Input: number of arms k, number of rounds n.

Initialize the knowledge set $\mathcal{S}_1 = \left\{ \left(\theta_{i,1}, \lambda_{i,1}^2\right) \middle| \, i = 1, \ldots, k \right\}$ and set $N_{1,1} = \cdots = N_{k,1} = 0$.

for $t = 1$ **to** $n - 1$ **do**

  Compute $I_t = \arg\max_i v_{i,t}^{\mathrm{KG}}$.

  Observe a reward $X_{I_t,t} \sim \mathcal{N}\left(\mu_{I_t}, \sigma_{I_t}^2\right)$.

  Update $\theta_{i,t+1}$ and $\lambda_{i,t+1}^2$, and update the knowledge set $\mathcal{S}_{t+1}$.

  $N_{I_t,t+1} = N_{I_t,t} + 1$, $N_{i,t+1} = N_{i,t}$ for $\forall i \neq I_t$.

  $t \leftarrow t + 1$.

end for

Output: $J_n = \arg\max_i \theta_{i,n}$.

Proposition

Under the KG algorithm, $\exists T > T_0$, it holds with a probability of at least $\left[1 - q\left(\frac{3}{4}t\right)\right]^k$ that for $\forall i \neq b$, $\forall t > T$,

$$\underline{\rho}_{i,b,t} \leq \frac{N_{i,t}}{N_{b,t}} \leq \overline{\rho}_{i,b,t}$$

where $q(s) = 4\sigma_{\max} k^{\frac{3}{8}} s^{-\frac{1}{8}} \exp\left\{-\frac{s^{\frac{1}{4}}}{8\sigma_{\max}^2 k^{\frac{3}{4}}}\right\}$.

For $\forall i \neq b$, $\lim_{t\to\infty} \underline{\rho}_{i,b,t} = \lim_{t\to\infty} \overline{\rho}_{i,b,t} = \frac{\sigma_i\left(\mu_b - \max_{j\neq b}\mu_j\right)}{\sigma_b(\mu_b - \mu_i)}$.

For $\forall i$, $\alpha_{i,t} = \frac{N_{i,t}}{t} = \frac{N_{i,t}/N_{b,t}}{\sum_{j=1}^{k} N_{j,t}/N_{b,t}}$.

Theorem

Under the KG algorithm, $\exists T > T_0$, it holds with a probability of at least $\left[1 - q\left(\frac{3}{4}t\right)\right]^k$ that for $\forall t > T$,

$$\frac{1}{1+\sum_{i\neq b}\overline{\rho}_{i,b,t}} \leq \alpha_{b,t} \leq \frac{1}{1+\sum_{i\neq b}\underline{\rho}_{i,b,t}},$$

$$\frac{\underline{\rho}_{i,b,t}}{1+\sum_{j\neq b}\overline{\rho}_{j,b,t}} \leq \alpha_{i,t} \leq \frac{\overline{\rho}_{i,b,t}}{1+\sum_{j\neq b}\underline{\rho}_{j,b,t}}, \ \forall i \neq b,$$

where $q(\cdot)$, $\underline{\rho}_{i,b,t}$, and $\overline{\rho}_{i,b,t}$ are from the Proposition.

Theorem

Under the KG algorithm, $\exists T > T_0$, $\forall t > T$, the probability of error $e_t = \mathbb{P}(J_t \neq b)$ is **upper-bounded** *by*

$$\frac{\sigma_b \sqrt{2\left(1+\sum_{i\neq b}\overline{\rho}_{i,b,t}\right)}}{\delta_{\min}\sqrt{\pi t}} \exp\left\{-\frac{\delta_{\min}^2}{8\sigma_b^2\left(1+\sum_{i\neq b}\overline{\rho}_{i,b,t}\right)} t\right\}$$
$$+\frac{\sqrt{2}k^{\frac{3}{8}}\sigma_b}{\sqrt{\pi}\delta_{\min} t^{\frac{3}{8}}}\left[1-\left[1-q\left(\frac{3}{4}t\right)\right]^k\right]\exp\left\{-\frac{\delta_{\min}^2}{8\sigma_b^2 k^{\frac{3}{4}}} t^{\frac{3}{4}}\right\}$$
$$+\sum_{i\neq b}\left\{\frac{\sigma_i\sqrt{\left(1+\sum_{i\neq b}\overline{\rho}_{i,b,t}\right)}}{\left(\mu_b-\mu_i-\frac{\delta_{\min}}{2}\right)\sqrt{2\pi\underline{\rho}_{i,b,t} t}}\exp\left\{-\frac{\left(\mu_b-\mu_i-\frac{\delta_{\min}}{2}\right)^2\underline{\rho}_{i,b,t}}{2\sigma_i^2\left(1+\sum_{i\neq b}\overline{\rho}_{i,b,t}\right)} t\right\}\right.$$
$$\left.+\frac{k^{\frac{3}{8}}\sigma_i\left[1-\left[1-q\left(\frac{3}{4}t\right)\right]^k\right]}{\sqrt{2\pi}\left(\mu_b-\mu_i-\frac{\delta_{\min}}{2}\right)t^{\frac{3}{8}}}\exp\left\{-\frac{\left(\mu_b-\mu_i-\frac{\delta_{\min}}{2}\right)^2}{2\sigma_i^2 k^{\frac{3}{4}}} t^{\frac{3}{4}}\right\}\right\},$$

where $q(\cdot)$, $\underline{\rho}_{i,b,t}$, and $\overline{\rho}_{i,b,t}$ are from the Proposition.

Theorem

Under the KG algorithm, $\exists T > T_0$, $\forall t > T$, the probability of error $e_t = \mathbb{P}(J_t \neq b)$ is **lower-bounded** *by*

$$\frac{\left[1-q\left(\frac{3}{4}t\right)\right]^{2k}}{2\pi}\min_{j\neq b}\left\{\frac{\frac{\delta_{\min}}{2\sigma_j}\sqrt{\frac{\underline{\rho}_{j,b,t}}{1+\sum_{i\neq b}\overline{\rho}_{i,b,t}}}}{1+\frac{\delta_{\min}^2\overline{\rho}_{j,b,t}t}{4\sigma_j^2\left(1+\sum_{i\neq b}\underline{\rho}_{i,b,t}\right)}}\frac{\frac{\left(\mu_b-\mu_j-\frac{\delta_{\min}}{2}\right)t}{\sigma_b\sqrt{1+\sum_{i\neq b}\overline{\rho}_{i,b,t}}}}{1+\frac{\left(\mu_b-\mu_j-\frac{\delta_{\min}}{2}\right)^2t}{\sigma_b^2\left(1+\sum_{i\neq b}\underline{\rho}_{i,b,t}\right)}}\exp\left\{-\frac{\delta_{\min}^2\overline{\rho}_{j,b,t}}{8\sigma_j^2\left(1+\sum_{i\neq b}\underline{\rho}_{i,b,t}\right)}t\right\}\cdot\exp\left\{-\frac{\left(\mu_b-\mu_j-\frac{\delta_{\min}}{2}\right)^2}{2\sigma_b^2\left(1+\sum_{i\neq b}\underline{\rho}_{i,b,t}\right)}t\right\}\right\}$$

where $q(\cdot)$, $\underline{\rho}_{i,b,t}$, and $\overline{\rho}_{i,b,t}$ are from the Proposition.

For $\forall t$, the simple regret $r_t = \mathbb{E}\left[\mu_b - \mu_{J_t}\right]$ satisfies that

$$\delta_{\min} e_t \leq r_t \leq \delta_{\max} e_t.$$

Theorem

Under the KG algorithm, the following statements hold:

- $\exists T > T_0$, *after round* $t > T$, *the cumulative regret* $R_t = t\mu_b - \sum_{s=1}^{t} \mathbb{E}\left[\mu_{I_t}\right]$ *is* **upper-bounded** *by*

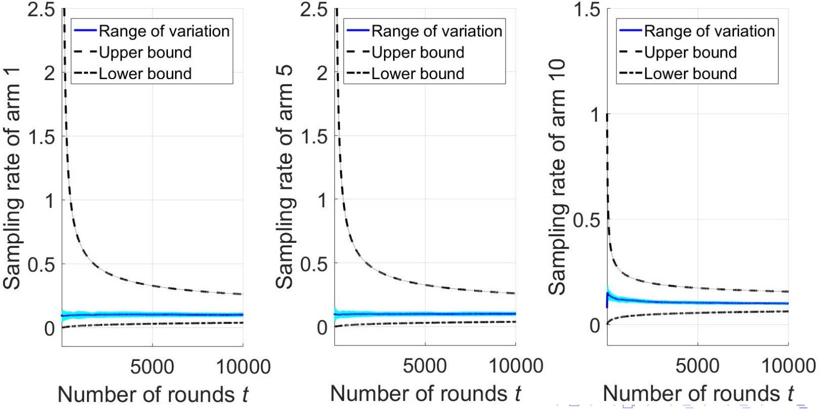
$$\frac{\sum_{i\neq b}\left(\mu_b - \mu_i\right)\overline{\rho}_{i,b,t}}{1 + \sum_{i\neq b}\underline{\rho}_{i,b,t}}t + k\sum_{i\neq b}\left(\mu_b - \mu_i\right)q\left(\frac{3}{4}t\right)t.$$

- $\lim_{t\to\infty}\frac{R_t}{t} = \frac{\sum_{i\neq b}\sigma_i}{\frac{\sigma_b}{\mu_b - \max_{j\neq b}\mu_j} + \sum_{i\neq b}\frac{\sigma_i}{\mu_b - \mu_i}}.$

where $q(\cdot)$, $\underline{\rho}_{i,b,t}$, *and* $\overline{\rho}_{i,b,t}$ *are from the Proposition.*

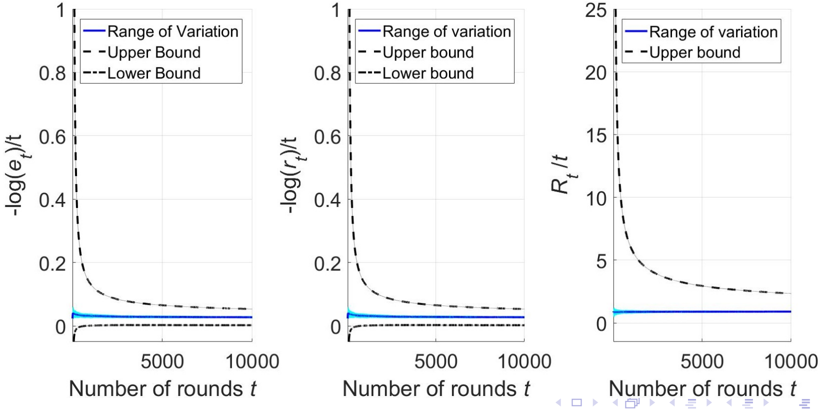
Numerical Results

Instance: We consider a set of ten arms $\{1, 2, \ldots, 10\}$. Set $\mu_i = 1$ for $i = 1, \ldots, 9$, $\mu_{10} = 2$, and $\sigma_i = 1$ for $i = 1, \ldots, 10$. The best arm $b = 10$.



Numerical Results

Instance: We consider a set of ten arms $\{1, 2, \ldots, 10\}$. Set $\mu_i = 1$ for $i = 1, \ldots, 9$, $\mu_{10} = 2$, and $\sigma_i = 1$ for $i = 1, \ldots, 10$. The best arm $b = 10$.



Conclusions and Discussions

- We consider the measures of the **probability of error** and **simple regret** in the BAI problem, and the measure of **cumulative regret** in the MAB problem, and derive **upper and lower bounds** of them.
- Our analysis can serve as the ground for **future research on the KG algorithm and BAI**.
- The analysis in this research might be extended to **study the performance of other KG-type and improvement-based algorithms**.
- The validity of the bounds highly depends on quantity T which is **difficult to compute**. Therefore, it is an important future research direction to study **how to quantify T**.

Thanks for listening!