Minimax Optimal Fixed-Budget Best Arm Identification in Linear Bandits

Junwen Yang[1], and Vincent Y. F. Tan[123]

[1]Institute of Operations Research and Analytics, [2]Department of Mathematics, and [3]Department of ECE, National University of Singapore, Singapore

Linear Bandits

- An arm set $\mathcal{A}=[K]$, which corresponds to arm vectors
$$\{a(1),a(2),\ldots,a(K)\}\subset\mathbb{R}^d.$$
- At each time t, the agent chooses an arm A_t from the arm set $\mathcal{A}$ and then observes a noisy reward
$$X_t=\langle\theta^*,a(A_t)\rangle+\eta_t,$$
where $\theta^*\in\mathbb{R}^d$ is the unknown parameter vector and η_t is independent zero-mean 1-subgaussian random noise.

Best Arm Identification in the Fixed-Budget Setting

- To maximize the probability of identifying the best arm with no more than T arm pulls.
- The agent uses an online algorithm π to decide the arm A_t to pull at each time step t, and the arm $i_{\text{out}}\in\mathcal{A}$ to output as the identified best arm by time T.
- We seek to minimize the error probability
$$\Pr\left[i_{\text{out}}\neq\arg\max_{j\in\mathcal{A}}\langle\theta^*,a(j)\rangle\right].$$

Notations

- For any arm $i\in\mathcal{A}$, let $p(i)=\langle\theta^*,a(i)\rangle$ denote the expected reward.
- Assume that $p(1)>p(2)\geq\cdots\geq p(K)$.
- For any suboptimal arm i, we denote $\Delta_i=p(1)-p(i)$ as the optimality gap. Also set $\Delta_1=\Delta_2$.

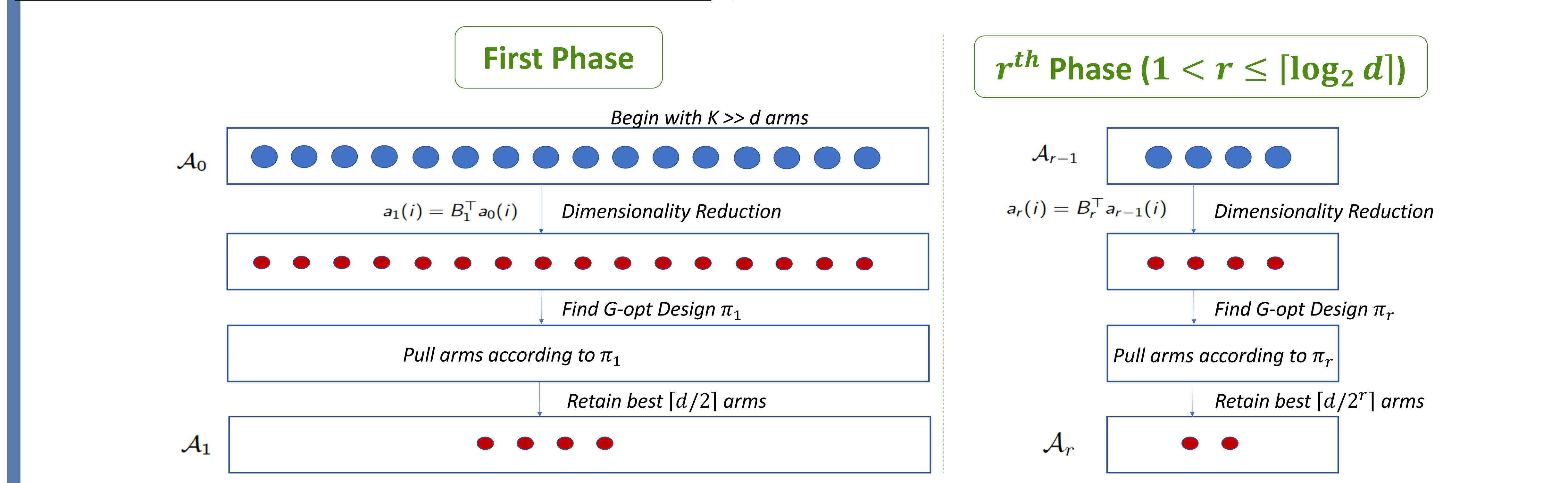
Hardness Quantities

$H_1=\sum_{1\leq i\leq K}\Delta_i^{-2}$	$H_2=\max_{2\leq i\leq K}\frac{i}{\Delta_i^2}$	$1\leq\frac{H_1}{H_2}\leq\log(2K)$
$H_{1,\text{lin}}=\sum_{1\leq i\leq d}\Delta_i^{-2}$	$H_{2,\text{lin}}=\max_{2\leq i\leq d}\frac{i}{\Delta_i^2}$	$1\leq\frac{H_{1,\text{lin}}}{H_{2,\text{lin}}}\leq\log(2d)$
$1\leq\frac{H_1}{H_{1,\text{lin}}}\leq\frac{K}{d}$	$1\leq\frac{H_2}{H_{2,\text{lin}}}\leq\frac{K}{d}$	

Algorithm: Optimal Design-based Linear Best Arm Identification (OD-LinBAI)

Input: time budget T, arm set $\mathcal{A}=[K]$ and arm vectors $\{a(1),a(2),\ldots,a(K)\}\subset\mathbb{R}^d$.

Initialize $t_0=1$, $\mathcal{A}_0\leftarrow\mathcal{A}$ and $d_0=d$.
For each arm $i\in\mathcal{A}_0$, set $a_0(i)=a(i)$.
Calculate $m=\Theta(T/\log_2 d)$.
for $r=1$ to $\lceil\log_2 d\rceil$ **do**
  Set $d_r=\dim(\text{span}(\{a_{r-1}(i):i\in\mathcal{A}_{r-1}\}))$.
  if $d_r=d_{r-1}$ **then**
    For each arm $i\in\mathcal{A}_{r-1}$, set $a_r(i)=a_{r-1}(i)$.
  else
    Find matrix $B_r\in\mathbb{R}^{d_{r-1}\times d_r}$ whose columns form a orthonormal basis of the subspace spanned by $\{a_{r-1}(i):i\in\mathcal{A}_{r-1}\}$.
    For each arm $i\in\mathcal{A}_{r-1}$, set $a_r(i)=B_r^\top a_{r-1}(i)$.
  end if
  if $r=1$ **then**
    Find a G-optimal design $\pi_r:\{a_r(i):i\in\mathcal{A}_{r-1}\}\rightarrow[0,1]$ with $|\text{Supp}(\pi_r)|\leq\frac{d(d+1)}{2}$.
  else
    Find a G-optimal design $\pi_r:\{a_r(i):i\in\mathcal{A}_{r-1}\}\rightarrow[0,1]$.
  end if
  Set
$$T_r(i)=\lceil\pi_r(a_r(i))\cdot m\rceil\quad\text{and}\quad T_r=\sum_{i\in\mathcal{A}_{r-1}}T_r(i).$$
  Choose each arm $i\in\mathcal{A}_{r-1}$ exactly $T_r(i)$ times.
  Calculate the OLS estimator:
$$\hat{\theta}_r=V_r^{-1}\sum_{t=t_r}^{t_r+T_r-1}a_r(A_t)X_t\text{ with }V_r=\sum_{i\in\mathcal{A}_{r-1}}T_r(i)a_r(i)a_r(i)^\top.$$
  For each arm $i\in\mathcal{A}_{r-1}$, estimate the expected reward:
$$\hat{p}_r(i)=\langle\hat{\theta}_r,a_r(i)\rangle.$$
  Let $\mathcal{A}_r$ be the set of $\lceil d/2^r\rceil$ arms in $\mathcal{A}_{r-1}$ with the largest estimates of the expected rewards.
  Set $t_{r+1}=t_r+T_r$.
end for
Output: the only arm i_{out} in $\mathcal{A}_{\lceil\log_2 d\rceil}$.



Minimax Lower Bound

Let $\mathcal{E}(a)$ denote the set of linear bandit instances whose $H_{1,\text{lin}}$ is bounded by $a>0$, i.e.,
$$\mathcal{E}(a)=\{\nu:H_{1,\text{lin}}(\nu)\leq a\}.$$

If $T\geq a^2\log(6Td)/900$, then
$$\min_{\pi}\max_{\nu\in\mathcal{E}(a)}\Pr[i_{\text{out}}^\pi\neq1]\geq\frac{1}{6}\exp\left(-\frac{240T}{a}\right).$$

Further if $a\geq15d^2$, then
$$\min_{\pi}\max_{\nu\in\mathcal{E}(a)}\left(\Pr[i_{\text{out}}^\pi\neq1]\cdot\exp\left(\frac{2700T}{H_{1,\text{lin}}(\nu)\log_2 d}\right)\right)\geq\frac{1}{6}.$$

Error Probability of OD-LinBAI

For any linear bandit instance ν, OD-LinBAI outputs an arm i_{out} satisfying
$$\Pr[i_{\text{out}}\neq1]\leq\left(\frac{4K}{d}+3\log_2 d\right)\exp\left(-\frac{m}{32H_{2,\text{lin}}}\right).$$

Minimax Optimality

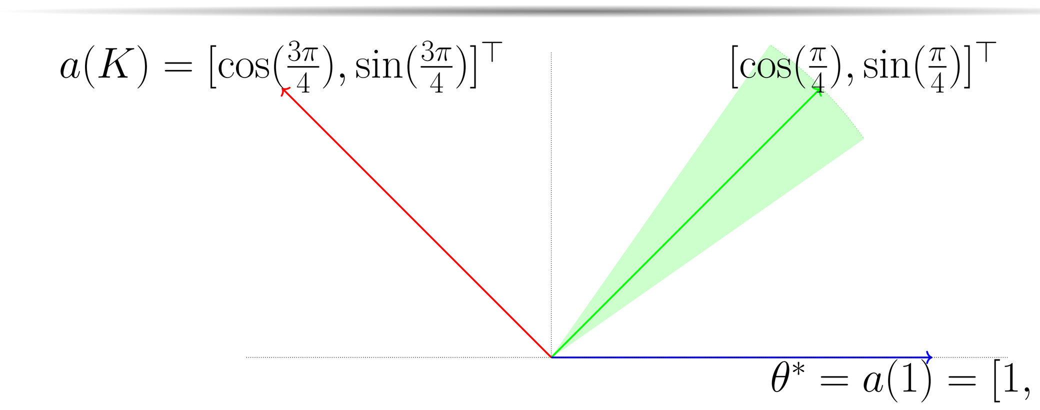
As the time budget T tends to infinity,

- Upper bound: $\exp\left(-\Omega\left(\frac{T}{H_{2,\text{lin}}\log_2 d}\right)\right)$
- Lower bound: $\exp\left(-O\left(\frac{T}{H_{1,\text{lin}}\log_2 d}\right)\right)$ $\quad H_{2,\text{lin}}\leq H_{1,\text{lin}}$

Comparison to Existing Art

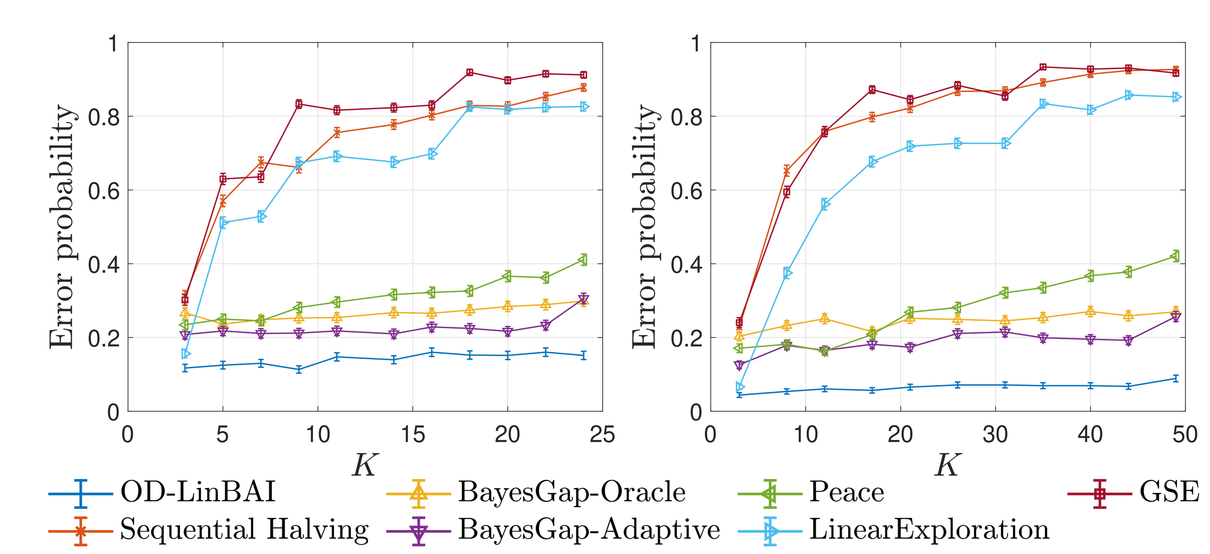
- BayesGap [1]:
  - Not parameter-free (require the knowledge of the instance)
  - Error probability: $\exp\left(-\Omega\left(\frac{T}{H_1}\right)\right)$
- Peace [2]:
  - Not parameter-free (require the knowledge of the instance)
  - Not minimax optimal
- LinearExploration [3]:
  - Error probability: $\exp\left(-\Omega\left(\frac{T}{H_2\log_2 K}\right)\right)$
- GSE [4]:
  - Error probability: $\exp\left(-\Omega\left(\frac{T\Delta_1^2}{d\log_2 K}\right)\right)$

Numerical Experiments



- One best arm, one worst arm and $K-2$ almost second best arms.
- $a(i)=[\cos(\pi/4+\phi_i),\sin(\pi/4+\phi_i)]^\top$ with $\phi_i\sim\mathcal{N}(0,0.09^2)$ for $i=2,3,\ldots,K-1$

Figure 1: Results for different numbers of arms K with $T=25,50$.



References

[1] M. Hoffman, B. Shahriari, and N. Freitas, "On correlation and budget constraints in model-based bandit optimization with application to automatic machine learning," in *AISTATS*, 2014.

[2] J. Katz-Samuels, L. P. Jain, Z. S. Karnin, and K. G. Jamieson, "An empirical process approach to the union bound: Practical algorithms for combinatorial and linear bandits," in *NeurIPS*, 2020.

[3] A. Alieva, A. Cutkosky, and A. Das, "Robust pure exploration in linear bandits with limited budget," in *ICML*, 2021.

[4] M. Azizi, B. Kveton, and M. Ghavamzadeh, "Fixed-budget best-arm identification in structured bandits," in *IJCAI*, 2022.

Full Paper is Available at:

https://arxiv.org/abs/2105.13017