Fixed-Budget Best-Arm Identification in Sparse Linear Bandits

Recep Can Yavas

June 28, 2023

Create Meeting

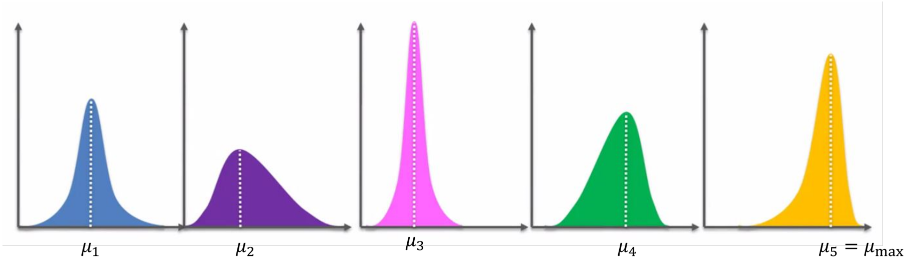
Joint work with Vincent Tan

Outline

- Introduction to multi-armed bandits framework
- Problem formulation
- Proposed algorithm
- Error probability bound and experiments

Multi-armed bandits

Motivation: A suitable mathematical model for applications such as drug design, online advertisement, and online recommendation systems



There are two main problems:

1. **Regret minimization:** Maximize the total reward after T pulls (exploration and exploitation)
2. **Best arm identification:** Find the arm with the largest mean (pure exploration)

Linear Best Arm Identification Problem

- K arms: each associated with a vector $a(k) \in \mathbb{R}^d$, $k \in [K]$, known to the agent
- Let at time t, we pull arm $A(t) \in \{1, \ldots, K\}$ and observe random reward

$$Y_t = \langle a(A(t)), \theta^* \rangle + Z_t$$

  where Z_t's are independent 1-subgaussian and θ^* is *unknown global* feature vector of dimension d
- The agent must make decisions based on the previous selections $A_1, A_2, \ldots, A_{t-1}$ and the observed rewards so far $Y_1, \ldots, Y_{t-1}$
- This problem is a sequential decision problem
- Best arm is the arm with the largest mean

$$\mu_1 \triangleq \max_{k \in [K]} \langle a(k), \theta^* \rangle$$

- WLOG, $\mu_1 > \mu_2 \geq \cdots \geq \mu_K$

Two Common Objectives in BAI

T: agent estimates the "best" arm after T arm pulls as $\hat{I} \in [K]$

Two different objectives:

1. fixed confidence: Target $\mathbb{P}\left[\hat{I} \neq 1\right]$ is fixed. T is a random variable. Minimize $\mathbb{E}[T]$.
2. fixed budget: T is fixed. Minimize $\mathbb{P}\left[\hat{I} \neq 1\right]$. Our focus is this setting

Sparse Linear Bandits

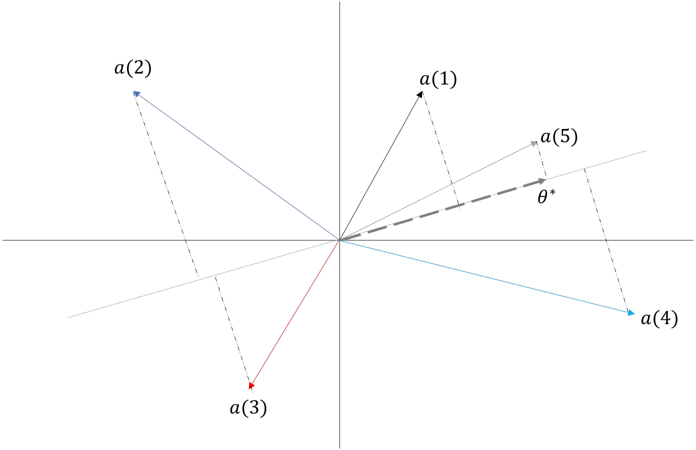
Global feature vector $\theta^* \in \mathbb{R}^d$

Motivation: Drug tests
$d \approx 100$: the number of variables measured for blood samples
$s = 5$: only 5 out of 100 variables significantly affect the performance of the drug
the rest is irrelevant to us

We model this phenomenon as

- Sparsity: $\|\theta^*\|_0 = \sum_{i=1}^{d} 1\{\theta_i^* \neq 0\} = s$ where $s \ll d$

Example



- $a(4)$ has the largest inner product with θ^*, therefore is the best arm.

Literature review

- Fixed confidence: For general bandits, [Garivier and Kaufmann (2016)] derive the asymptotically optimal algorithm
- Fixed confidence for linear bandits: [Soare et al. (2014)], [Xu et al. (2018)], [Xu et al. (2018)], ...
- Fixed budget for linear bandits: [Hoffman et al. (2014)], [Katz-Samuels et al. (2020)], [Yang and Tan (2022)]
- Sparsity in regret minimization: [Abbasi-Yadkori et al. (2012)], [Bastani and Bayati (2020)], [Hao et al. (2021)], [Ariu et al. (2022)]

Our algorithm: Lasso-OD

The high-level idea is

- First, explore the sparsity of θ^* by estimating its support

  $$S(\theta^*) = \{i \in [d] \colon \theta_i^* \neq 0\}$$

  With high probability, the estimated support captures the true support, i.e., $\hat{S} \supset S(\theta^*)$ and its size isn't too big, i.e., $|\hat{S}| - s \ll d$

- Then, we use the G-optimal design to decide on the arm pulls and to sequentially halve the remaining best arm candidates in each round (Non-sparse solution from [Yang and Tan, (2022)])

Phase 1: Estimate $S(\theta^*)$	Phase 2: G-Optimal design restricted to $S(\theta^*)$

T_1 T

Technique: Lasso (Tibshirani, 1996)

Let $Y = (Y_1, \ldots, Y_{T_1}) \in \mathbb{R}^{T_1}$ be rewards. Let $\mathrm{A} \in \mathbb{R}^{T_1 \times d}$ be the design matrix, each row being the arm vector we pull.

The s-sparse θ^* can be estimated by the convex program

$$\hat{\theta}_{\mathrm{init}} = \arg\max_{\theta} \frac{1}{T_1} \left\| Y - \mathrm{A}\theta \right\|_2^2 + \lambda_{\mathrm{init}} \left\| \theta \right\|_1$$

λ_{init}: a free parameter to be optimized according to performance criteria

Thresholding [Ariu et al. (2022)]: We estimate the support after thresholding $\hat{\theta}_{\mathrm{init}}$

$$\hat{S} = \{ j \in [d] \colon |(\hat{\theta}_{\mathrm{init}})_j| \geq \lambda_{\mathrm{thres}} \}$$

Then, we run the algorithm only on the variables in $\hat{S}$

Main Result

Theorem

$$\mathbb{P}\left[\hat{I} \neq 1\right] \leq (K + \log_2 d + 2d) \exp\left\{-\frac{T}{16\lfloor \log_2(s+s^2)\rfloor (1+\epsilon) H_{2,\text{lin}}(s+s^2)(1+c_0)}\right\}$$

$$H_{2,\text{lin}}(s) = \max_{i \in \{2,\dots,s\}} \frac{i}{(\mu_1 - \mu_i)^2}$$

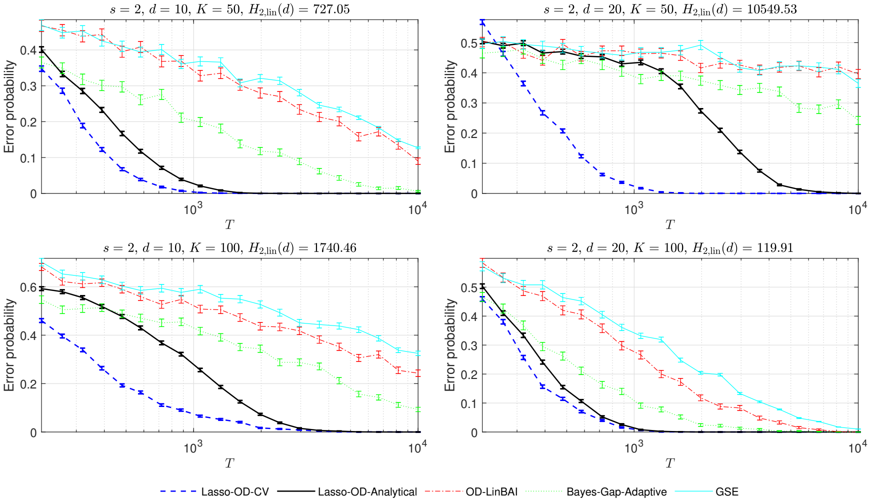
c_0 is a constant and $\epsilon \to 0$ as $T \to \infty$.

Main message: Error probability scales as

$$\exp\left\{-\Omega\left(\frac{T}{\log_2(s) H_{2,\text{lin}}(s+s^2)}\right)\right\}$$

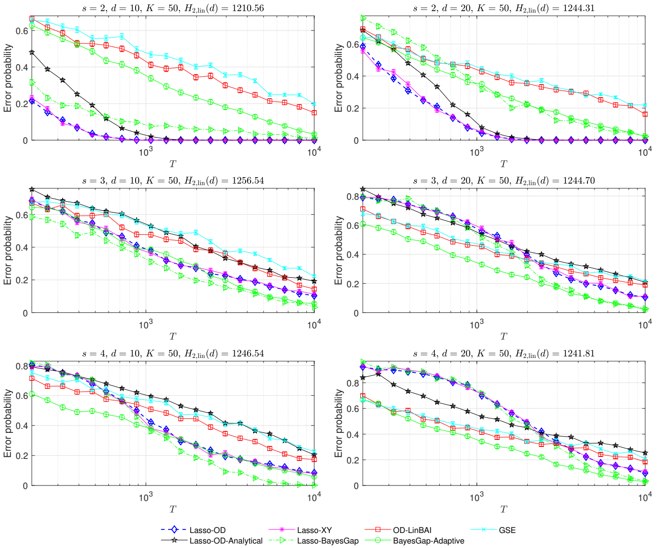
which is independent of d!

Experiments

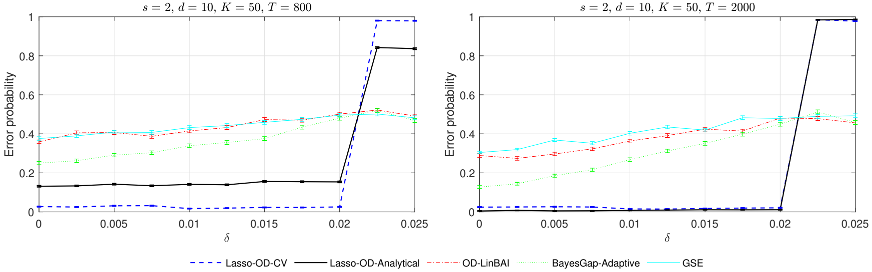


Arm vectors are drawn i.i.d. Gaussian.

Experiments



Robustness test



Non-zero coordinates in θ^* are replaced with $\pm\epsilon$.

Alternative to Thresholded Lasso (Jang et al. (2023))

Algorithm 1 PopArt (POPulation covariance regression with hARd Thresholding)

1: **Input:** Samples $\{(X_t, Y_t)\}_{t=1}^n$, the population covariance matrix $Q \in \mathbb{R}^{d\times d}$, pilot estimator $\theta_0 \in \mathbb{R}^d$, an upper bound R_0 of $\max_{a\in\mathcal{A}} |\langle a, \theta^* - \theta_0\rangle|$, failure rate δ.
2: **Output:** estimator $\hat{\theta}$
3: **for** $t = 1, \dots, n$ **do**
4: $\quad \tilde{\theta}_t = Q^{-1} X_t (Y_t - \langle X_t, \theta_0\rangle) + \theta_0$
5: **end for**
6: $\forall i \in [d], \theta'_i = \mathsf{Catoni}(\{\tilde{\theta}_{ti} := \langle \tilde{\theta}_t, e_i\rangle\}_{t=1}^n, \alpha_i, \frac{\delta}{2d})$ where $\alpha_i := \sqrt{\frac{2\log\frac{2d}{\delta}}{n(R_0^2+\sigma^2)(Q^{-1})_{ii}\left(1+\frac{2\log\frac{2d}{\delta}}{n-2\log\frac{2d}{\delta}}\right)}}$
7: $\hat{\theta} \leftarrow \mathsf{clip}_\lambda(\theta') := [\theta'_i \mathbb{1}(|\theta'_i| > \lambda_i)]_{i=1}^d$ where λ_i is defined in Proposition 1.
8: **return** $\hat{\theta}$

Catoni mean estimator of $(Z_1, \dots, Z_n)$ are defined as the unique solution to

$$\sum_{i=1}^{n} \phi(\alpha(Z_i - y)\hat{\mu}) = 0,$$

where $\phi(x) = \mathrm{sign}(x)\log(1 + |x| + x^2)$