Sequential Experimental Design for Transductive Linear Bandits

Tanner Fiez, Lalit Jain, Kevin Jamieson, Lillian Ratliff
University of Washington

Transductive Linear Bandits

Input: $\mathcal{X}, \mathcal{Z} \in \mathbb{R}^d$, confidence $\delta \in (0, 1)$
for $t = 1, 2, \cdots$

1. Learner chooses $x_t \in \mathcal{X}$

$$\epsilon_t \sim \mathcal{N}(0, 1)$$

2. Nature reveals $y_t = \langle \theta_*, x_t \rangle + \epsilon_t$

Learner defined by **selection rule** $x_t \in \mathcal{X}$, **stopping rule** at which time the learner recommends $\hat{z} \in \mathcal{Z}$.

An algorithm is δ-correct if $\mathbb{P}(\hat{z} \neq \text{argmax}_{z \in \mathcal{Z}} \langle \theta_*, z \rangle) \geq 1 - \delta$

Contributions:
1. **Lower bounds for the Transductive Linear Bandit Problem.**
2. **RAGE Algorithm with matching sample complexity (up to logarithmic factors)**
3. **First matching upper and lower bounds for Pure Exploration for Linear Bandits.**

Examples

Example 1: Content Recommendation.

- $\mathcal{X} = \mathcal{Z} \subset \mathbb{R}^d$, corresponds to a set of songs
- Unknown $\theta_* \in \mathbb{R}^d$ encapsulates preferences of a user.
- How do we play songs to learn the users favorite songs?
- When $\mathcal{X} = \mathcal{Z}$, recover **pure exploration for linear bandits**.

Example 2: Drug Discovery

- $\mathcal{Z} \subset \mathcal{X} \subset \mathbb{R}^d$, corresponds to sets of compounds.
- θ_* feature vector of an antigen, $\langle \theta_*, x \rangle$, is effect of compound x on the antigen.
- Testing potentially unsafe compounds $\mathcal{X}$ that we would not use on patients may help us more quickly learn $\text{argmax}_{z \in \mathcal{Z}} \theta_*^\top z$

Theoretical Result Summary

$$\rho^* = \min_{\lambda \in \Delta_{\mathcal{X}}} \max_{z \in \mathcal{Z} \setminus z_*} \frac{(z^* - z)^\top (\sum_{x \in \mathcal{X}} \lambda_x x x^\top)^{-1} (z^* - z)}{|\langle \theta_*, z^* - z \rangle|^2}$$

Adaptive lower bound (Extends Soare 2015)

$$\rho^\star \log(1/\delta)$$

Adaptive upper bound (Fiez, Jain, Jamieson, Ratliff 2019) $\Delta = \min_{z \neq z^*} \langle \theta_*, z^* - z \rangle$

$$\rho^* [\log(1/\delta) + \log(|\mathcal{Z}|) + \log(\log(\Delta^{-1}))] \log(1/\Delta)$$

Non-adaptive (single round of experimental design):

$$\frac{d}{\Delta^2} [\log(1/\delta) + \log(|\mathcal{X}|)]$$

**When are these different?
When sampling its beneficial to sample along the differences.**

RAGE: Randomized Adaptive Gap Elimination

Input: $\mathcal{X}, \mathcal{Z} \subset \mathbb{R}^d$ **set** $\mathcal{Z}_1 = \mathcal{Z}$

for $\ell = 1, 2, \ldots$

1. Perform experimental design on $\mathcal{Z}_\ell$

$$\lambda_\ell = \text{argmin}_{\lambda \in \Delta_{\mathcal{X}}} \max_{z, z' \in \mathcal{Z}_\ell} (z' - z)^\top \left(\sum_{x \in \mathcal{X}} \lambda_x x x^\top \right)^{-1} (z' - z)$$

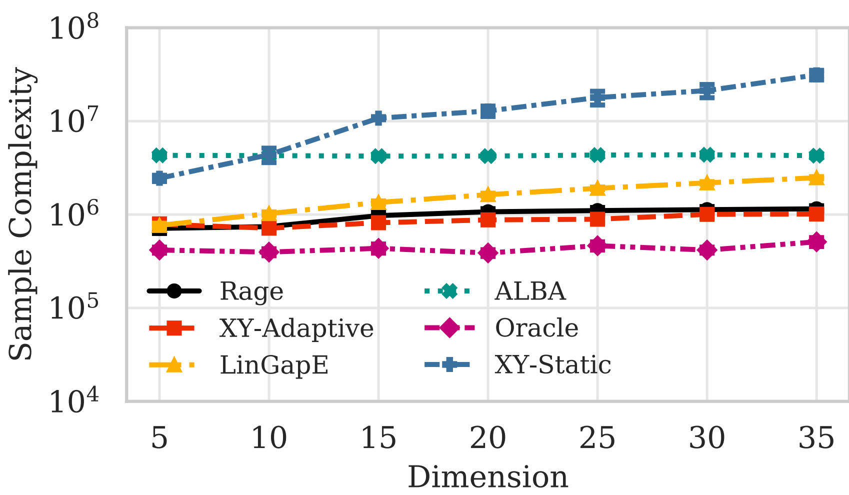
2. Compute $\widehat{\theta}_\ell$ **by sampling from** λ_ℓ **enough times to ensure that**

$$\langle \theta_*, z^* - z \rangle > 2^{-\ell} \implies \langle \widehat{\theta}_\ell, z^* - z \rangle > 2^{-(\ell+1)}$$

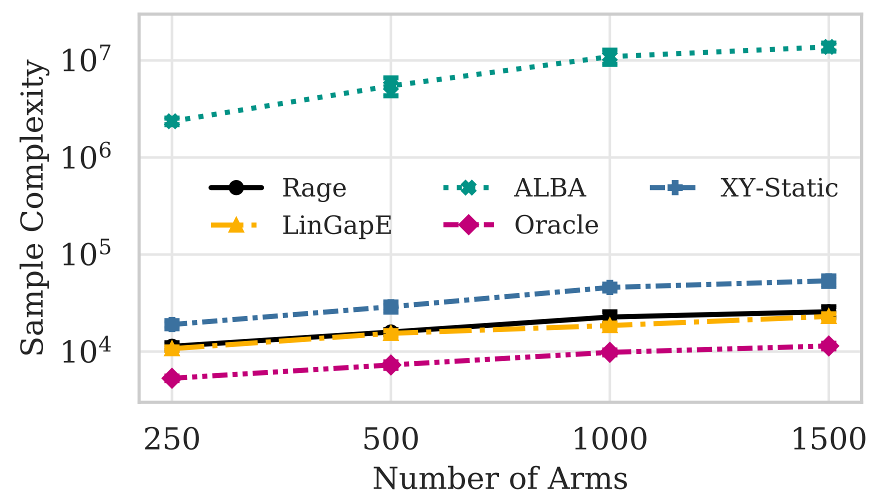
3. Update set

$$\mathcal{Z}_{\ell+1} = \mathcal{Z}_\ell \setminus \left\{ z \in \mathcal{Z}_\ell : \exists z' \in \mathcal{Z}_\ell, \langle \widehat{\theta}_\ell, z' - z \rangle > 2^{-(\ell+1)} \right\}$$

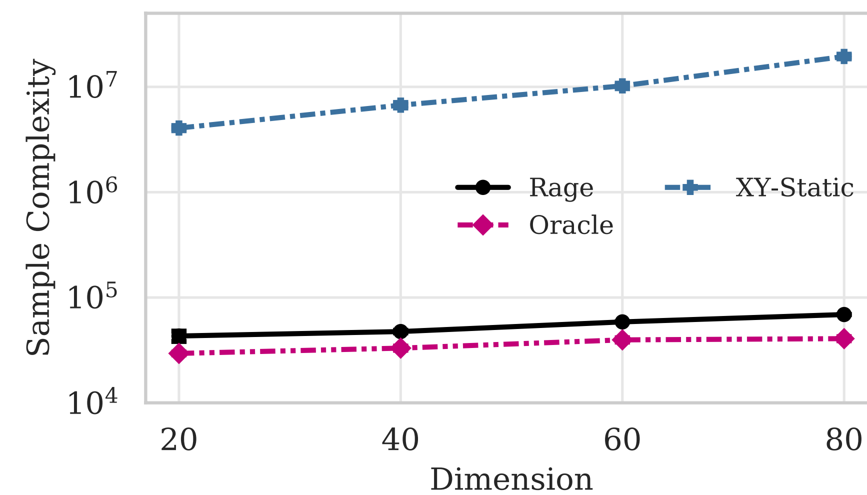
Experiments



Benchmark Example



Uniform Sphere Example



Transductive Example

Previous Work on Pure Exploration for Linear Bandits

Marta Soare, Alessandro Lazaric, and Rémi Munos. Best-arm identification in linear bandits. In *Advances in Neural Information Processing Systems*, pages 828–836, 2014.

Zohar S Karnin. Verification based solution for structured mab problems. In *Advances in Neural Information Processing Systems*, pages 145–153, 2016.

Chao Tao, Saúl Blanco, and Yuan Zhou. Best arm identification in linear bandits with linear dimension dependency. In *International Conference on Machine Learning*, pages 4884–4893, 2018.

Liyuan Xu, Junya Honda, and Masashi Sugiyama. A fully adaptive algorithm for pure exploration in linear bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 843–851, 2018.

Our result is uniformly tighter and first to match problem-dependent lower bound.