Sequential Experimental Design for Transductive Linear Bandits

Tanner Fiez
Electrical and Computer Engineering
fiezt@uw.edu

Lalit Jain
Foster School of Business
lalitj@cs.washington.edu

Kevin Jamieson
Allen School of Computer Science
jamieson@cs.washington.edu

Lillian Ratliff
Electrical and Computer Engineering
ratliffl@uw.edu

Introduction

In many problems, there is a set of items $\mathcal{Z}$ with underlying structure, and the goal is to find which one maximizes a response transductively through noisy measurements of a set of probes $\mathcal{X}$.

Recommendations: $\mathcal{X} \subset \mathcal{Z} \subset \mathbb{R}^d$

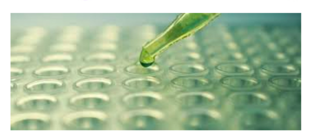
$\subset$

$\mathcal{X}$ $\mathcal{Z}$

$\mathcal{X} \rightarrow$ Popular songs that can be safely presented to users.

$\mathcal{Z} \rightarrow$ Music catalog to be evaluated including esoteric titles.

Drug Discovery: $\mathcal{Z} \subset \mathcal{X} \subset \mathbb{R}^d$



$\supset$ FDA ✓

$\mathcal{X}$ $\mathcal{Z}$

$\mathcal{X} \rightarrow$ Drugs including any experimental compounds verifiable in lab.

$\mathcal{Z} \rightarrow$ Drugs approved to be administered to patients.

How do we sequentially and adaptively decide which measurements to take?

Problem Statement

Given: items $\mathcal{Z} \subset \mathbb{R}^d$, probes $\mathcal{X} \subset \mathbb{R}^d$, unknown parameters $\theta^* \in \mathbb{R}^d$

Measure: At each time t, observe $r_t = x_t^\top \theta^* + \eta_t$, where η_t is 1-subGaussian

Find: $z^* = \arg\max_{z \in \mathcal{Z}} z^\top \theta^*$

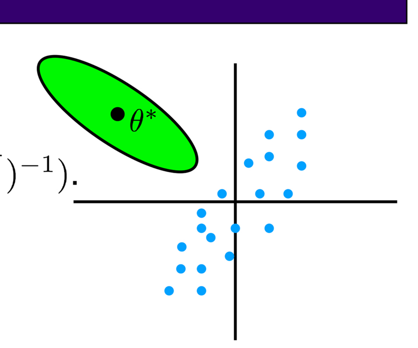
Transductive Linear Bandit Environment

Input: $\mathcal{X} \subset \mathbb{R}^d$, $\mathcal{Z} \subset \mathbb{R}^d$, $\delta \in (0,1)$.
Until learner invokes stopping time τ
  Learner selects $x_t \in \mathcal{X}$
  Nature reveals $r_t \leftarrow x_t^\top \theta^* + \eta_t$
Output: Learner invokes recommendation $\hat{z} \in \mathcal{Z}$

Generalization of Multi-Armed Bandits $\rightarrow \mathcal{X} = \mathcal{Z} = \{e_1, \dots, e_d\} \subset \mathbb{R}^d$
Generalization of Linear Bandits $\rightarrow \mathcal{X} = \mathcal{Z} \subset \mathbb{R}^d$.
Generalization of Combinatorial Bandits $\rightarrow \mathcal{X} = \{e_1, \dots, e_d\} \subset \mathbb{R}^d, \mathcal{Z} \subset \{0,1\}^d$.

Problem Intuition

Consider a learner selects a non-adaptive fixed design $\{x_t\}_{t=1}^T$ and observes rewards $\{r_t\}_{t=1}^T$ and constructs a least squares estimate $\widehat{\theta} = (\sum_{t=1}^T x_t x_t^\top)^{-1}(\sum_{t=1}^T r_t x_t)$. Then, $\widehat{\theta} - \theta^* \sim \mathcal{N}(0, (\sum_{t=1}^T x_t x_t^\top)^{-1})$.
Strategically sample to shape covariance!

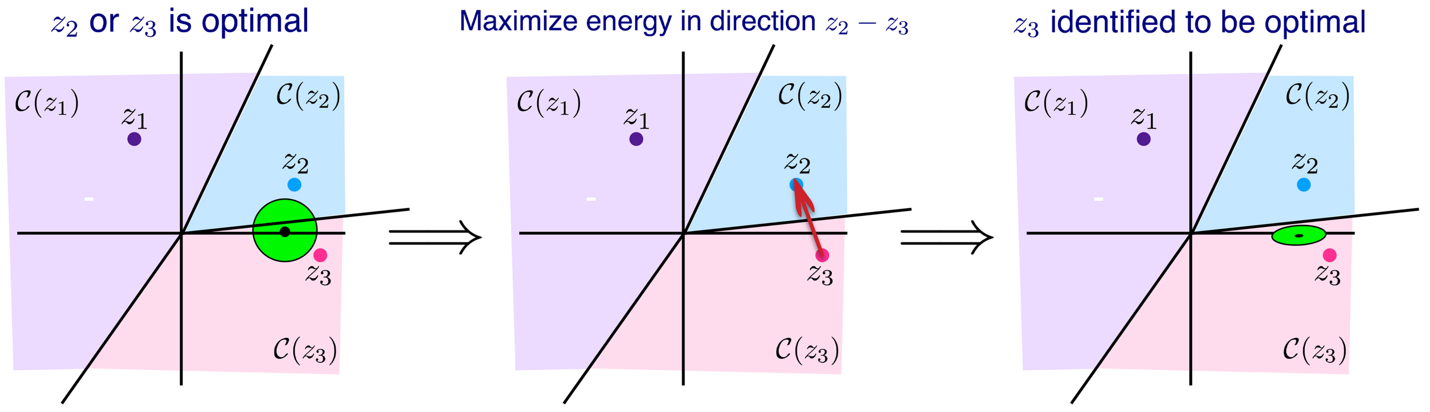


$z_* = \arg\max_{z \in \mathcal{Z}} \Longrightarrow (z_* - z)^\top \theta^* > 0 \quad \forall z \in \mathcal{Z} \setminus z_*$

$\mathcal{C}(z) = \{\theta \in \mathbb{R}^d : (z - z')^\top \theta > 0 \; \forall z' \in \mathcal{Z} \setminus z\} \Longrightarrow$ Cone of parameters for which $z = z_*$

Goal: Efficiently shrink confidence set into $\mathcal{C}(z_*)$ to identify z_* w.p. $\geq 1 - \delta$

Illustrative Toy Example



Algorithm

Algorithm 1: RAGE$(\mathcal{X}, \mathcal{Z}, \delta)$: **R**andomized **A**daptive **G**ap **E**limination

Input: $\mathcal{X} \subset \mathbb{R}^d$, $\mathcal{Z} \subset \mathbb{R}^d$, $\delta \in (0,1)$.
Initalize: $\widehat{\mathcal{Z}}_1 \leftarrow \mathcal{Z}, t \leftarrow 1$

while $|\widehat{\mathcal{Z}}_t| > 1$ **do**
  Experimental Design: $\lambda_t^* \leftarrow \arg\min_{\lambda \in \triangle_{\mathcal{X}}} \max_{z,z' \in \widehat{\mathcal{Z}}_t} \|z - z'\|^2_{(\sum_{x \in \mathcal{X}} \lambda_x x x^\top)^{-1}}$
  $\rho_t \leftarrow \min_{\lambda \in \triangle_{\mathcal{X}}} \max_{z,z' \in \widehat{\mathcal{Z}}_t} \|z - z'\|^2_{(\sum_{x \in \mathcal{X}} \lambda_x x x^\top)^{-1}}$

  Sample: $N_t \leftarrow \lceil 2(2^t)^2 \rho_t \log(t^2|\mathcal{Z}|^2/\delta) \rceil \}$
  Pull arms $x_1, \dots, x_{N_t}$ according to λ_t^* and obtain rewards $r_1, \dots, r_{N_t}$

  Eliminate: Let $\widehat{\theta}_t = A_t^{-1} b_t$
  $\widehat{\mathcal{Z}}_{t+1} \leftarrow \widehat{\mathcal{Z}}_t \setminus \{z \in \widehat{\mathcal{Z}} | \exists\, z' \in \widehat{\mathcal{Z}} : \|z' - z\|_{A_t^{-1}} \sqrt{2\log(t^2|\mathcal{Z}|^2/\delta)} < (z' - z)^\top \widehat{\theta}_t\}$
  $t \leftarrow t + 1$
Output: $\widehat{\mathcal{Z}}_t$

Methods

Optimal Sampling Distribution $\Longrightarrow \lambda^* := \arg\min_{\lambda \in \triangle_{\mathcal{X}}} \max_{z \in \mathcal{Z} \setminus \{z_*\}} \frac{\|z_* - z\|^2_{(\sum_{x \in \mathcal{X}} \lambda_x x x^\top)^{-1}}}{((z_* - z)^\top \theta^*)^2}$

Fact: Sampling according to rounded λ^* is sufficient to achieve $\rho^* \log(|\mathcal{X}|/\delta)$.

Challenge: Optimal sampling allocation cannot be computed without knowledge of θ^*!

Question: Can the optimal sampling allocation be closely approximated using an adaptive strategy?

Idea: Apply experimental design repeatedly in stages to converge toward optimal allocation.

Define

- $\mathcal{S}_t = \{z : (z_* - z)^\top \theta > 2^{-t}\}$
- For any $\mathcal{S} \subset \mathcal{Z}$, $\rho(\mathcal{S}) := \arg\min_{\lambda \in \triangle_{\mathcal{X}}} \max_{z,z' \in \mathcal{S}} \|z - z'\|^2_{(\sum_{x \in \mathcal{X}} \lambda_x x x^\top)^{-1}}$

Algorithm Guarantee: Arms with gaps bigger than 2^{-t} removed and $z_* \in \widehat{\mathcal{Z}}_t$ in each round.

$$\widehat{\mathcal{Z}}_t \subset S_t \text{ which implies } \rho_t \leq \rho(S_t)$$

Algorithm Sample Complexity: At most $\rho_t (2^t)^2 \log(t^2|\mathcal{Z}|^2/\delta) < \rho(S_t)(2^t)^2 \log(t^2|\mathcal{Z}|^2/\delta)$ per round

$$\sum_{t=1}^{\lfloor \log_2(1/\Delta_{\min}) \rfloor} \rho(S_t)(2^t)^2 \log(t^2|\mathcal{Z}|^2/\delta)$$

Need to compare $\sum_{t=1}^{\lfloor \log(1/\Delta_{\min}) \rfloor} (2^t)^2 \rho(\mathcal{S}_t)$ to the lower bound ρ^*!

$$\rho^* = \min_{\lambda \in \triangle_{\mathcal{X}}} \max_{t \leq \lfloor \log(1/\Delta_{\min}) \rfloor} \max_{z \in \mathcal{S}_t \setminus \{z_*\}} \frac{\|z_* - z\|^2_{(\sum_{x \in \mathcal{X}} \lambda_x x x^\top)^{-1}}}{((z_* - z)^\top \theta^*)^2}$$

$$\geq \min_{\lambda \in \triangle_{\mathcal{X}}} \max_{t \leq \lfloor \log(1/\Delta_{\min}) \rfloor} \max_{z \in \mathcal{S}_t \setminus \{z_*\}} \frac{\|z_* - z\|^2_{(\sum_{x \in \mathcal{X}} \lambda_x x x^\top)^{-1}}}{(2^{-t})^2}$$

$$\geq \frac{1}{\log_2(1/\Delta_{\min})} \min_{\lambda \in \triangle_{\mathcal{X}}} \sum_{t=1}^{\lfloor \log_2(1/\Delta_{\min}) \rfloor} (2^t)^2 \max_{z \in \mathcal{S}_t \setminus \{z_*\}} \|z_* - z\|^2_{(\sum_{x \in \mathcal{X}} \lambda_x x x^\top)^{-1}}$$

$$\geq \frac{1}{\log_2(1/\Delta_{\min})} \sum_{t=1}^{\lfloor \log_2(1/\Delta_{\min}) \rfloor} (2^t)^2 \min_{\lambda \in \triangle_{\mathcal{X}}} \max_{z \in \mathcal{S}_t \setminus \{z_*\}} \|z_* - z\|^2_{(\sum_{x \in \mathcal{X}} \lambda_x x x^\top)^{-1}}$$

$$\geq \frac{1}{4\log_2(1/\Delta_{\min})} \sum_{t=1}^{\lfloor \log_2(1/\Delta_{\min}) \rfloor} (2^t)^2 \underbrace{\min_{\lambda \in \triangle_{\mathcal{X}}} \max_{z,z' \in \mathcal{S}_t} \|z - z'\|^2_{(\sum_{x \in \mathcal{X}} \lambda_x x x^\top)^{-1}}}_{\rho(\mathcal{S}_t)}$$

Theoretical Guarantees

Key Problem-Dependent Quantity $\Longrightarrow \rho^* := \min_{\lambda \in \triangle_{\mathcal{X}}} \max_{z \in \mathcal{Z} \setminus \{z_*\}} \frac{\|z_* - z\|^2_{(\sum_{x \in \mathcal{X}} \lambda_x x x^\top)^{-1}}}{((z_* - z)^\top \theta^*)^2}$

Theorem (Lower Bound)
For $\mathcal{N}(0,1)$ noise, any δ-PAC algorithm must satisfy $\mathbb{E}_{\theta^*}[\tau] \geq \rho^* \log(1/(2.4\delta))$.

$\rightarrow$ Generalizes previous lower bounds from linear bandits and combinatorial bandits.

Theorem (RAGE Sample Complexity Bound)
Algorithm 1 identifies z_* w.p. $\geq 1 - \delta$ using a sample complexity no worse than
$\rho^* [\log(1/\delta) + \log(|\mathcal{Z}|) + \log(\log(1/\Delta_{\min}))] \log(1/\Delta_{\min}) + d\log(1/\Delta_{\min})$.
Matches lower bound up to log factors!

$\rightarrow$ Uniformly tighter bound than previous work and only existing non-asymptotic algorithm that nearly matches the problem-dependent lower bound.

$\rightarrow$ There exists $\mathcal{X}, \mathcal{Z}, \theta^*$ such that any static allocation (such as G-optimal design) requires d times the sample complexity of best known adaptive algorithm.

$\rightarrow$ $d\log(1/\Delta_{\min})$ term is an artifact of efficient rounding procedure in each round.

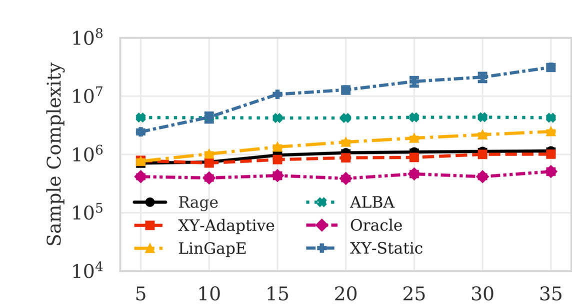
Numerical Experiments

Benchmark: $\mathcal{X} = \mathcal{Z} = \{e_1, \dots, e_d, x'\} \subset \mathbb{R}^d$, $x' = \cos(.01)e_1 + \sin(.01)e_2$ and $\theta^* = 2e_1$ so $x_* = x_1$.

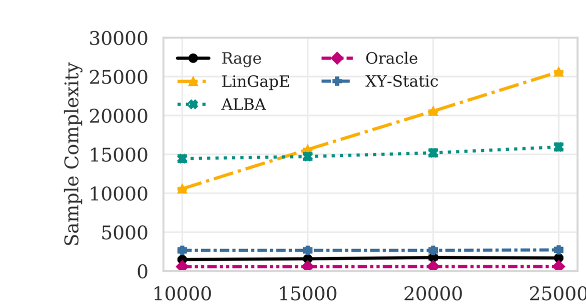
Duplicate Arms: $\mathcal{X} \subset \mathbb{R}^2$, where $\mathcal{X} = \mathcal{Z} = \{e_1, \cos(3\pi/4)e_1 + \sin(3\pi/4)e_2\} \cup \{\cos(\pi/4 + \phi_i)e_1 + \sin(\pi/4 + \phi_i)e_2\}_{i=3}^n$ with $\phi_i \sim \mathcal{N}(0, .09)$ for each $i \in \{3, \dots, n\}$ and $\theta^* = e_1$ so that $x_* = x_1$.

Uniform Sphere: $\mathcal{X} = \mathcal{Z} \sim$ unit sphere $\mathbb{S}^9$. The closest arms $x, x' \in \mathcal{X}$ are selected and $\theta^* = x$.

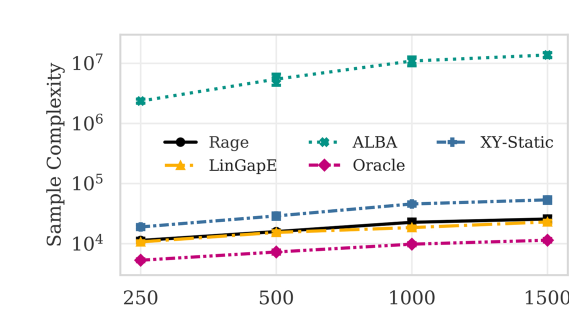
Yahoo! Click-Through: $\mathcal{X} = \mathcal{Z} \subset \mathbb{R}^{36}$ constructed from user and article features and θ^* learned from data. Rewards generated from Bernoulli$(x^\top \theta^*)$ for any arm selection $x \in \mathcal{X}$ and $|\mathcal{X}| = 40$.



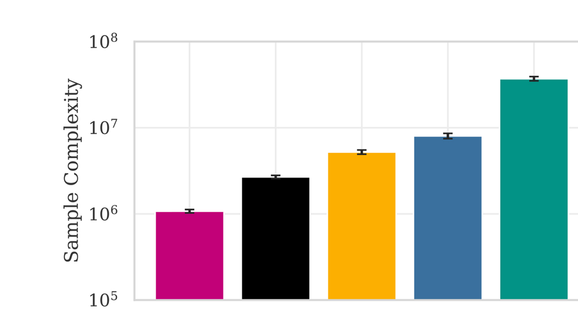
Benchmark Conclusion: Highlights the potential for gains of adaptive sampling over non-adaptive sampling.



Duplicate Arms Conclusion: Highlights the sample complexity can be independent of the number of arms.



Uniform Sphere Conclusion: Highlights the gains from computing experimental design on the differences between vectors.



Click-Through Conclusion: Highlights the empirical performance of RAGE on a real-world application.

About Me

Tanner Fiez is a 4th-year Electrical and Computer Engineering PhD student at the University of Washington advised by Lillian Ratliff. He previously obtained a B.S. in Electrical and Computer Engineering from Oregon State University. His research interests include multi-armed bandits and broadly sequential decision-making along with game theory. Contact: fiezt@uw.edu.

Reference: "Sequential Experimental Design for Transductive Linear Bandits," Fiez, Jain, Jamieson, Ratliff. NeurIPS, 2020.