An Empirical Process Approach to the Union Bound: Practical Algorithms for Combinatorial and Linear Bandits

Julian Katz-Samuels  
Allan School of Computer Science & Engineering  
University of Washington  
jkatzsam@cs.washington.edu

Lalit Jain  
Foster School of Business  
University of Washington  
lalitj@uw.edu

Zohar Karnin  
Amazon Web Services  
zkarnin@gmail.com

Kevin Jamieson  
Allan School of Computer Science & Engineering  
University of Washington  
jamieson@cs.washington.edu

Abstract

This paper proposes near-optimal algorithms for the pure-exploration linear bandit problem in the fixed confidence and fixed budget settings. Leveraging ideas from the theory of suprema of empirical processes, we provide an algorithm whose sample complexity scales with the geometry of the instance and avoids an explicit union bound over the number of arms. Unlike previous approaches which sample based on minimizing a worst-case variance (e.g., G-optimal design), we define an experimental design objective based on the Gaussian-width of the underlying arm set. We provide a novel lower bound in terms of this objective that highlights its fundamental role in the sample complexity. The sample complexity of our fixed confidence algorithm matches this lower bound, and in addition is computationally efficient for combinatorial classes, e.g., shortest-path, matchings and matroids, where the arm sets can be exponentially large in the dimension. Finally, we propose the first algorithm for linear bandits in the fixed budget setting. Its guarantee matches our lower bound up to logarithmic factors.

1 Introduction

The pure exploration stochastic multi-armed bandit (MAB) problem has received attention in recent years because it offers a useful framework for designing algorithms for sequential experiments. In this paper, we consider a very general formulation of the pure exploration MAB problem, namely, pure exploration (transductive) linear bandits \cite{12}: given a set of measurement vectors $\mathcal{X} \subseteq \mathbb{R}^d$, a set of candidate items $\mathcal{Z} \subseteq \mathbb{R}^d$, and an unknown parameter vector $\theta \in \mathbb{R}^d$, an agent plays a sequential game where at each round she chooses a measurement vector $x \in \mathcal{X}$ and observes a stochastic random variable whose expected value is $\langle x, \theta \rangle$. The goal is to identify $z_* \in \arg\max_{z \in \mathcal{Z}} \langle z, \theta \rangle$. This problem generalizes many well-studied problems in the literature including best arm identification \cite{11, 21, 23, 25, 6}, Top-K arm identification \cite{22, 28, 9}, the thresholding bandit problem \cite{27}, combinatorial bandits \cite{10, 13, 8, 5, 20}, and linear bandits where $\mathcal{X} = \mathcal{Z}$ \cite{30, 33, 31}.

The recent work of \cite{12} proposed an algorithm that is within a $\log(|\mathcal{Z}|)$ multiplicative factor of previously known lower bounds \cite{39} on the sample complexity. This term reflects a naive union bound over all informative directions $\{z_* - z: z \in \mathcal{Z} \setminus \{z_*\}\}$. Although one might be inclined to dismiss $\log(|\mathcal{Z}|)$ as a small factor, in many practical problems it can be extremely large. For example,
A/B testing. For instance, in content recommendation which devises a fixed sampling scheme using knowledge of which describes the deviation of the suprema of a Gaussian process from its expectation, leading us to while we show that the true sample complexity does not depend on $\log(|Z|)$. Finally, in many applications of linear bandits such as content recommendation $|Z|$ can be enormous and thus the factor $\log(|Z|)$ can have a dramatic effect on the sample complexity.

The high-level goal of this paper is to study how the geometry of the measurement vectors $X$ and the candidate items $Z$ influences the sample complexity of the pure exploration transductive linear bandit problem in the moderate confidence regime. We appeal to the fundamental TIS-inequality \cite{Abbe17} which describes the deviation of the suprema of a Gaussian process from its expectation, leading us to propose an experimental design based on minimizing the expected suprema. We make the following contributions. First, we show a novel lower bound for the non-interactive oracle MLE algorithm, which devises a fixed sampling scheme using knowledge of $\theta$. While this non-interacting lower bound is not a lower bound for adaptive algorithms, it is suggestive of what union bounds are necessary and can be a multiplicative dimension factor larger than known adaptive lower bounds. Second, we develop a new algorithm for the fixed confidence setting (defined below) that nearly matches the performance of this oracle algorithm. Moreover, this algorithm recovers many of the state-of-the-art sample complexity results for combinatorial bandits as special cases. Third, applied specifically to the combinatorial bandit setting, we develop a practical and computationally efficient algorithm. We include experiments that show that our algorithm outperforms existing algorithms, often by an order of magnitude. Finally, we show that our techniques extend to the fixed budget setting where we provide the first fixed budget algorithm for transductive linear bandits. This algorithm matches the lower bound up to a factor that in most standard settings is bounded by $\log(d)$.

## 2 Preliminaries

In the (transductive) linear bandit problem, the agent is given a set $X \subset \mathbb{R}^d$ and a set of items $Z \subset \mathbb{R}^d$. At each round $t$, an algorithm $A$ selects a measurement $X_t \in X$ which is measurable with respect to the history $F_{t-1} = (X_1, Y_{s<t})$ and observes a noisy observation $Y_t = X_t^\top \theta + \eta$ where $\theta \in \mathbb{R}^d$ is the unknown model parameter and $\eta$ is independent mean-0 Gaussian noise. We assume that $\arg\max_{z \in Z} \langle z, \theta \rangle = \{z_1\}$, and the goal is to identify $z_1$. We consider two distinct settings.

**Definition 1.** Fixed-Confidence: Fix $X, Z, \Theta \subset \mathbb{R}^d$. An algorithm $A$ is $\delta$-PAC for $(X, Z, \Theta)$ if 1) the algorithm has a stopping time $\tau$ wrt $(F_t)_{t \in \mathbb{N}}$ and 2) at time $\tau$ it makes a recommendation $\hat{z} \in Z$ and for all $\theta \in \Theta$ it satisfies $\mathbb{P}_\theta(\hat{z} = z_1) \geq 1 - \delta$.

**Definition 2.** Fixed-Budget: Fix $X, Z, \Theta \subset \mathbb{R}^d$ and a budget $T$. An algorithm $A$ for fixed-budget returns a recommendation $\hat{z} \in Z$ after $T$ rounds.

Linear bandits is popular for applications such as content recommendation, digital advertisements, and A/B testing. For instance, in content recommendation $X = Z \subset \mathbb{R}^d$ may be sets of feature vectors describing songs (e.g., beats per minute, genre, etc.) and $\theta \in \mathbb{R}^d$ may represent an individual user’s preferences over the song library. An important sub-class of linear bandits is known as combinatorial bandits which is a focus of this work.

**Combinatorial Bandits:** In the combinatorial bandit setting, $X = \{e_1, \ldots, e_d\}$ (where $e_i$ is the $i$-th canonical basis vector) and $Z \subset \{0, 1\}^d$. We will sometimes overload notation by treating $Z$ as a collection of sets, e.g., for $z \in Z$ writing $i \in z$ iff $\langle e_i, z \rangle = 1$. We next give some examples of the combinatorial bandit setting.

**Example 1 (Matroid).** $M = (S, I)$ is a matroid where $S$ is a set of ground elements and $I \subset 2^S$ is a collection of independent sets. This setting includes best arm identification, Top-$K$ arm identification, identifying the minimum spanning tree with largest expected reward in a graph, and other important applications (see \cite{Abbe17} for a list of applications).

**Example 2 (Matching).** For a balanced bipartite graph with $d$ edges and $2\sqrt{d}$ vertices let $Z$ denote the set of $\sqrt{d}$ perfect bipartite matchings. The goal is to identify the matching $z \in Z$ that maximizes $\langle \theta, z \rangle$.

---

1 Our results still apply in the case where the noise is sub-Gaussian, but for simplicity here we assume that the noise is Gaussian (see the Supplementary Material).
In some of these settings, \( |Z| \) is exponential in the dimension \( d \). For example, in the problem of finding a best matching in a bipartite graph, \( |Z| = (\sqrt{d})! \). In this setting a naive evaluation of \( \arg\max_{z \in Z}(z, \theta) \) by enumerating \( Z \) becomes impossible even if \( \theta \) were known. For such problems, we assume access to a linear maximization oracle

\[
\text{ORACLE}(w) = \arg \max_{z \in Z} \langle z, w \rangle,
\]

which is available in many cases, including matroids, MATCHING, and identifying a shortest path in a directed acyclic graph (DAG). We will characterize the computational complexity of an algorithm in terms of the number of calls to the maximization oracle.

3 Review of Gaussian Processes

We now discuss how our work departs from previous approaches to the pure exploration linear bandit problem. Consider for a moment a fixed design where \( n \geq d \) measurements \( x_1, \ldots, x_n \) were decided before observing any data, and subsequently for each \( 1 \leq i \leq n \) we observe \( y_i = \langle x_i, \theta \rangle + \eta_i \) with \( \eta_i \sim N(0, 1) \). In this setting the maximum likelihood estimator (MLE) is given by ordinary least squares as

\[
\hat{\theta} = (\sum_{i=1}^n x_i x_i^\top)^{-1} \sum_{i=1}^n y_i x_i.
\]

Substituting the value of \( y_i \) into this expression, we obtain \( \hat{\theta} = \theta + (\sum_{i=1}^n x_i x_i^\top)^{-1/2} \eta \) in distribution where \( \eta \sim N(0, I_d) \). After collecting \( \{(x_i, y_i)\}_{i=1}^n \) and computing \( \hat{\theta} \), the most reasonable estimate for \( z_\star = \arg\max_{z \in Z} \langle z, \hat{\theta} \rangle \) is just \( \hat{z} = \arg\max_{z \in Z} \langle z, \hat{\theta} \rangle \).

The good event that \( \hat{z} = z_\star \) occurs if and only if \( \langle z_\star - z, \hat{\theta} \rangle \geq 0 \) for all \( z \in Z \setminus \{z_\star\} \). Since \( \hat{\theta} \) is a Gaussian random vector, for each \( z \in Z \), \( \langle z_\star - z, \hat{\theta} \rangle \sim N(0, (z_\star - z)^\top (\sum_{i=1}^n x_i x_i^\top)^{-1} (z_\star - z)) \).

If we apply a standard sub-Gaussian tail-bound with a union bound over all \( z \in Z \setminus \{z_\star\} \), then we have with probability greater than \( 1 - \delta \) that

\[
\langle z_\star - z, \hat{\theta} \rangle \geq \langle z_\star - z, \theta \rangle - \sqrt{2\|z_\star - z\|^2_{A^{-1}} \log(|Z|/\delta)}
\]

for all \( z \in Z \setminus \{z_\star\} \) simultaneously, where we have taken \( A = \sum_{i=1}^n x_i x_i^\top \) and used the notation \( \|x\|_W^2 = x^\top W x \) for any square \( W \). Thus, we conclude that if \( n \) and \( \{x_1, \ldots, x_n\} \) are chosen such that

\[
\max_{z \in Z} \frac{\|z_\star - z\|^2_{(\sum_{i=1}^n x_i x_i^\top)^{-1} A^{-1}}} {\langle z_\star - z, \hat{\theta} \rangle} > 2 \log(|Z|/\delta)
\]

then with probability at least \( 1 - \delta \) we will have

\[
\langle z_\star - z, \hat{\theta} \rangle > 0 \quad \text{for all} \quad z \in Z \quad \text{and consequently,} \quad \hat{z} = z_\star.
\]

This simple argument is the core of all approaches to pure exploration linear bandits until this paper \cite{tsirelson1971inequality, ibramov1972inequality, sudakov1981inequality, dudley2010uniform}. However, applying a naive union bound over all \( z \in Z \) can be extremely weak and does not exploit the geometry of \( Z \) that induces many correlations among the random variables \( \langle z_\star - z, \hat{\theta} \rangle \). At the heart of our approach is the following concentration inequality for the suprema of a Gaussian process (Theorem 5.8 in \cite{dudley2010uniform}).

**Theorem 1** (Tsirelson-Ibramov-Sudakov Inequality \cite{tsirelson1971inequality}). Let \( S \subset \mathbb{R}^d \) be bounded. Let \( (V_s)_{s \in S} \) be a Gaussian process such that \( \mathbb{E}[V_s] = 0 \) for all \( s \in S \). Define \( \sigma^2 = \sup_{s \in S} \mathbb{E}[V_s^2] \). Then, for all \( u > 0 \),

\[
\mathbb{P}(\sup_{s \in S} |V_s - \mathbb{E}[V_s]| \geq u) \leq 2 \exp \left( -\frac{u^2}{2\sigma^2} \right).
\]

Setting \( S = Z \), we can apply this to the Gaussian process \( V_z := (z_\star - z)^\top (\hat{\theta} - \theta) = (z_\star - z)^\top (\sum_{i=1}^n x_i x_i^\top)^{-1/2} \eta \) where, again, \( \eta \sim N(0, I_d) \). We then have with probability at least \( 1 - \delta \)

\[
(z_\star - z)^\top \hat{\theta} \geq (z_\star - z)^\top \theta - \mathbb{E}\eta \left[ \sup_{z \in Z \setminus \{z_\star\}} (z_\star - z)^\top A^{-1/2} \eta \right] - \sqrt{2\sup_{z \in Z} \|z_\star - z\|^2_{A^{-1}} \log(1/\delta)}
\]

for all \( z \in Z \setminus \{z_\star\} \) simultaneously. This bound naturally breaks into two components. The second term is the high-probability term, and as the discussion above implies, naturally motivates the experimental design objective \( \min_{x_1, \ldots, x_n} \max_{z \in Z \setminus \{z_\star\}} \|z_\star - z\|^2_{(\sum_{i=1}^n x_i x_i^\top)^{-1}} \) from past works on linear-bandit pure exploration. The first term, \( \mathbb{E}_{\eta \sim N(0, I_d)} \left[ \sup_{z \in Z \setminus \{z_\star\}} (z_\star - z)^\top (\sum_{i=1}^n x_i x_i^\top)^{-1/2} \eta \right] \) is the Gaussian-width of the set \( \{(z_\star - z)^\top (\sum_{i=1}^n x_i x_i^\top)^{-1/2} \mid z \in Z \setminus \{z_\star\}\} \) \cite{dudley2010uniform}. This term represents the penalty we pay for the union bound over the possible values of \( Z \) and reflects the underlying geometry.


of our arm set. For moderately sized values of $\delta \in (0, 1)$ such as the science-stalwart $\delta = 0.05$, the Gaussian width term can be substantially larger than the high probability term. Analogous to above, this motivates choosing $x_1, \ldots, x_n$ to minimize the Gaussian width term.

Relaxation to Continuous Experimental Designs. In practice, optimizing over all finite sets of $\mathcal{X}$ of size $n$ to minimize an experimental design objective is NP-hard. Define $\Delta := \{ \lambda \in \mathbb{R}^{[|X|]} : \sum \lambda_i = 1, \lambda_i \geq 0 \}$ to be the simplex over elements $X$ and define $A(\lambda) = \sum_{x \in X} \lambda_x x^\top$ where $\lambda \in \Delta$ denotes a convex combination of the measurement vectors. Defining the design that minimizes the high probability term motivates the definition

$$\rho^* := \inf_{\lambda \in \Delta} \rho^*(\lambda) \quad \text{where} \quad \rho^*(\lambda) := \sup_{z \in \mathbb{Z} \setminus \{z_\ast\}} \frac{\|z^\ast - z\|^2}{\langle \theta, z^\ast - z \rangle^2}.$$  

On the other hand, minimizing the Gaussian width term motivates the definition

$$\gamma^* := \inf_{\lambda \in \Delta} \gamma^*(\lambda) \quad \text{where} \quad \gamma^*(\lambda) := \mathbb{E}_{\eta \sim N(0, I)} \left[ \sup_{z \in \mathbb{Z} \setminus \{z_\ast\}} \langle z^\ast - z, \eta \rangle^2 \right].$$

While the above suggests the importance of the quantities $\rho^*$ and $\gamma^*$, we will show later how they are intrinsic to the problem hardness. For now, we point out that these quantities are easily relatable.

**Proposition 1.** There exists universal constants $c, c^\prime > 0$ such that for any $\mathcal{X}$ and $\mathcal{Z}$ we have $c \rho_\ast - \inf_{z \neq z_\ast} \inf_{\lambda \in \Delta} \frac{\|z^\ast - z\|^2}{\langle \theta, z^\ast - z \rangle^2} \leq \gamma^* \leq \min\{c^\prime \log(|\mathcal{Z}|) \rho^*, d \rho^*\}$.

Typically, $\inf_{z \neq z_\ast} \inf_{\lambda \in \Delta} \frac{\|z^\ast - z\|^2}{\langle \theta, z^\ast - z \rangle^2} \ll \rho^*$, in which case $\rho_\ast \lesssim \gamma_\ast$. While there are instances where $\gamma^* = \Theta(d \rho^*)$, the upper bound is not necessarily tight.

**Proposition 2.** There exists an instance of transductive linear bandits where $\gamma^* \geq c d \rho^*$, and a separate instance for which $\gamma^* \leq c^\prime \log(d) \rho^*$ where $c, c^\prime > 0$ are universal constants.

4 Towards the true sample complexity

This section formally justifies the quantities $\rho_\ast$ and $\gamma_\ast$ defined above. The following result holds for any $\mathcal{X}$ and $\mathcal{Z}$ and was first proven in this generality in [12], extending [30][29][8].

**Theorem 2 (Lower bound for any adaptive algorithm [12]).** For any $\delta \in (0, 1)$, any $\delta$-PAC algorithm wrt $(\mathcal{X}, \mathcal{Z}, \mathbb{R}^d)$ with stopping time $\tau$ satisfies $\mathbb{E}_\theta[\tau] \geq \log(\frac{1}{\delta}) \rho_\ast$.

Mirroring the approaches developed in [24][8][14], it is possible to develop an algorithm that satisfies $\lim_{\delta \rightarrow 0} \mathbb{E}_\theta[\tau] \log(\frac{1}{\delta}) = \rho_\ast$, demonstrating the tightness of Theorem 2 in the regime of $\delta$ tending towards 0. However, for fixed $\delta \in (0, 1)$, algorithms for linear bandits to date have only been able to match this lower bound up to additive factors of $d \rho_\ast$ or $\log(|\mathcal{Z}|) \rho_\ast$ [24][14] (note, this does not rule out optimality as $\delta \rightarrow 0$). In particular, the lower and the upper bounds of linear bandits do not reflect the underlying geometry of general sets $\mathcal{X}$ and $\mathcal{Z}$ in union bounds and are loose in general. For example, in the well-studied case of Top-K, these bounds do not capture some additive factors that are necessary and achievable in addition to $\rho_\ast$ alone [28][9].

As a step towards characterizing the true sample complexity, we next demonstrate a lower bound that incorporates the geometry of $\mathcal{X}$ and $\mathcal{Z}$ for, presumably, the best possible non-interactive algorithm. Precisely, the procedure is given access to $\theta$, chooses an allocation $\{x_1, x_2, \ldots, \} \in \mathcal{X}$, then observes $(y_1, y_2, \ldots) \in \mathcal{R}$ where $y_i \sim \mathcal{N}(\langle x_i, \theta \rangle, 1)$, and finally forms the MLE $\hat{\theta} = \arg \min_{\theta} \sum_i (y_i - \langle x_i, \theta \rangle)^2$ and outputs $\hat{Z} = \arg \max_{z \in \mathcal{Z}} \langle z, \theta \rangle$. We emphasize that this procedure can pick any allocation it desires using full knowledge of $\theta$; in particular, it can use the allocation that achieves $\rho_\ast$.

**Theorem 3 (Lower bound for non-interactive MLE).** Let $\delta \in (0, 1/8)$. Fix $\mathcal{X}, \mathcal{Z} \subset \mathbb{R}^d$ and $\Theta = \mathbb{R}^d$. Fix a problem $\theta \in \Theta$. Then, if the non-interactive MLE is $\delta$-PAC wrt $(\mathcal{X}, \mathcal{Z}, \Theta)$ where a different allocation can be used for each $\theta \in \Theta$, then it uses at least $c(\gamma_\ast + \rho_\ast \log(1/\delta))$ samples for the instance $(\mathcal{X}, \mathcal{Z}, \Theta)$ where $c > 0$ is a universal constant.
We note that while our upper bound has an extra additive factor of \(d\) we show that after a suitable transformation, it is convex in the combinatorial bandit setting. We

Theorem 4. With probability at least 1 − \(\delta\), Algorithm 7 terminates and returns \(z^*\) after a number of samples no more than

\[
\left[ \gamma^* + \rho^* \log(\log(\frac{B}{\Delta_{\min}}) / \delta) \right] \min(\log(\frac{B}{\Delta_{\min}}), \log(\frac{B}{\min_{z \in Z} \min_{z^* \in Z^*} \tau(z; z^*)})) + cd \log(\frac{B}{\Delta_{\min}}).
\]

We note that while our upper bound has an extra additive factor of \(d\) compared to the lower bound of Theorem 3, this factor is necessary in many instances of interest (see the Supplementary Material). \(\log(B) = O(\log(d))\) when \(X = Z\) and in combinatorial bandits, and \(B\) can be replaced by an upper bound on \(\max_{z \in Z} \Delta_z\) when one is known. \(\tau(\lambda; Z_k)\) can be optimized using stochastic mirror descent; we show that after a suitable transformation, it is convex in the combinatorial bandit setting. We conjecture that it is convex in the general case, as well.

5 Fixed Confidence Setting Algorithms

In this section, we present Algorithm 1 Peace, that achieves the state-of-the-art sample complexity for (transductive) linear bandits in the fixed confidence setting. In each round \(k\) we eliminate from the set of candidates \(Z\) all the elements that are roughly \(2^{-k}\) suboptimal. In each round the query allocation is fixed according to the best non-adaptive strategy.

Our algorithm must round a design to an integral solution. It uses an efficient rounding procedure \(\text{ROUND}(\lambda, N, \epsilon)\) that for \(\lambda \in \mathcal{D} \) and \(N \geq q(\epsilon)\) returns \(\kappa \in \mathbb{N}^{|\mathcal{A}|}\) such that \(\sum_{x \in \mathcal{X}} \kappa_x = N\) and \(\tau(\kappa; Z') \leq (1+\epsilon)\tau(N\lambda; Z')\). It suffices to take \(q(\epsilon) = O(d/\epsilon^2)\) (see the Supplementary Material). Define \(S_k := \{ z \in Z : \theta^\top(z^* - z) \leq B 2^{-k}\}\), \(\Delta_z := \theta^\top(z^* - z)\), and \(\Delta_{\min} := \min_{z \in \mathcal{Z} \setminus \{z^*\}} \Delta_z\).

**Theorem 4.** With probability at least 1 − \(\delta\), Algorithm 7 terminates and returns \(z^*\) after a number of samples no more than

\[
[\gamma^* + \rho^* \log(\log(\frac{B}{\Delta_{\min}}) / \delta)] \epsilon \min(\log(\frac{B}{\Delta_{\min}}), \log(\frac{B}{\min_{z \in Z} \min_{z^* \in Z^*} \tau(z; z^*))} + cd \log(\frac{B}{\Delta_{\min}}).
\]

We note that while our upper bound has an extra additive factor of \(d\) compared to the lower bound of Theorem 3, this factor is necessary in many instances of interest (see the Supplementary Material). \(\log(B) = O(\log(d))\) when \(X = Z\) and in combinatorial bandits, and \(B\) can be replaced by an upper bound on \(\max_{z \in Z} \Delta_z\) when one is known. \(\tau(\lambda; Z_k)\) can be optimized using stochastic mirror descent; we show that after a suitable transformation, it is convex in the combinatorial bandit setting. We conjecture that it is convex in the general case, as well.

5.1 Computationally Efficient Algorithm for Combinatorial Bandits

A drawback of Algorithm 1 is that it is computationally inefficient when \(|Z|\) is exponentially large in the dimension. In this section, we develop an algorithm for combinatorial bandits that is computationally efficient when the linear maximization oracle defined in (1) is available. We introduce the following notation for a set \(Z' \subset Z\):

\[
\gamma(Z') := \min_{\mathcal{A} \in \mathcal{D}} \mathbb{E}[\sup_{z, z' \in Z'} (z - z')^\top A^{-1/2} (\lambda) \eta]^2.
\]

Algorithm 2 proceeds by estimating the gap of each \(z \in Z\) at a progressively finer level of granularity. The objective in line 4 is carefully designed to find a sample-efficient allocation that ensures that at
whereas the former scales like $\gamma$ suboptimal item

Next, we turn to the fixed budget setting, where the goal is to minimize the probability of returning a

call to the linear maximization oracle can check whether the constraint in (5) is satisfied so the

$\gamma$ suboptimal item

samples and if

above linear program can be solved using binary search and multiple calls to the maximization oracle.

problem as the following equivalent linear program:

$h \in \{\eta \sim N(0, I) \mid \max_{z \in \mathcal{Z}} (\hat{z}_k - z)^\top A(\lambda)^{-1/2} \eta - s [2^{-k} \Gamma + \hat{\theta}_k^\top (\hat{z}_k - z)] \leq 0 \}$ (4)

Set $N_k \leftarrow \alpha [\tau_k \log((1/\delta_k)(1 + \epsilon)) \vee q(\epsilon)]$ and find $\{x_1, \ldots, x_{N_k}\} \leftarrow \text{ROUND}(\lambda_k, N_k)$;

Let $\tilde{\theta}_{k+1} \leftarrow (\sum_{s=1}^{N_k} x_s x_s^\top)^{-1} \sum_{s=1}^{N_k} x_s y_s$;

if $\text{UNIQUE}(\mathcal{Z}, \tilde{\theta}_k, 2^{-k} \Gamma)$ then return $\tilde{z}_k$

Algorithm 2: Fixed Confidence Peace with a linear maximization oracle.

round $k$, with high probability $\tilde{\theta}_k^\top (\hat{z}_k - z) \approx \Delta_z$ for all $z \in \mathcal{Z}$ such that $\Delta_z \geq 2^{-k} \Gamma$. The subroutine

$\text{UNIQUE}(\mathcal{Z}, \tilde{\theta}_k, 2^{-k} \Gamma)$ uses calls to the linear maximization oracle to determine whether the gaps are

sufficiently well-estimated to conclude that $\hat{z}_k$ is $z_\ast$ (see the Supplementary Material).

In the Supplementary Material, we provide procedures for computing $\gamma(\mathcal{Z})$ and (4) only using calls to the linear maximization oracle. The main challenge is to compute an unbiased estimate of the gradient

of the objective in equation (4) (for an appropriate first-order optimization procedure such as stochastic

mirror descent), which we now sketch. Since the expectation in (4) is non-negative, it suffices to
draw $\eta \sim N(0, I)$, and compute $\nabla_\lambda \max_{z \in \mathcal{Z}} g(\lambda; z)$. Since for a collection of differentiable functions $\{h_1, \ldots, h_l\}$, a sub-gradient $\nabla_\lambda \max_{z \in \mathcal{Z}} h_i(y)$ is simply $\nabla_y h_i(y)$ where $h_i(y) = \arg \max_{z \in \mathcal{Z}} h_i(y)$, it suffices to find $\arg \max_{z \in \mathcal{Z}} g(\lambda; z)$. We re-formulate this optimization problem as the following equivalent linear program:

$$\min_s \quad \text{subject to} \max_{z \in \mathcal{Z}} (\hat{z}_k - z)^\top A(\lambda)^{-1/2} \eta - s [2^{-k} \Gamma + \hat{\theta}_k^\top (\hat{z}_k - z)] \leq 0$$ (5)

A call to the linear maximization oracle can check whether the constraint in (5) is satisfied so the

above linear program can be solved using binary search and multiple calls to the maximization oracle.

Theorem 5. Consider the combinatorial bandit setting. With probability at least $1 - 4\delta$ Algorithm 2 terminates and returns $z_\ast$, after at most

$$[(\gamma_\ast + \rho_\ast) \log(\log(\gamma(\mathcal{Z})/\Delta_{m_{\min}})/\delta)] + d \epsilon \log(\gamma(\mathcal{Z})/\Delta_{m_{\min}})$$ samples and if $\delta \in (\frac{1}{2^5}, 1)$, then with probability at least $1 - 4\delta$, the number of oracle calls is upper bounded by

$$O([d + \log(\max_{z \in \mathcal{Z}} \Delta_z) + \Gamma] + \log(\frac{d}{\Delta_{m_{\min}}^2} \frac{d^3 \log(\gamma(\mathcal{Z})/\Delta_{m_{\min}})^5}{\delta^2}))$$

Theorem 5 neatly matches the sample complexity of Theorem 4. The latter scales like $\gamma_\ast + \rho_\ast \log(1/\delta)$ whereas the former scales like $(\gamma_\ast + \rho_\ast)\log(1/\delta)$, reflecting a tradeoff of statistical efficiency for computational efficiency. It is unknown if this tradeoff is necessary.

6 Fixed Budget Setting

Next, we turn to the fixed budget setting, where the goal is to minimize the probability of returning a

suboptimal item $z \in \mathcal{Z} \setminus \{z_\ast\}$ given a budget of $T$ total measurements. Algorithm 3 is a generalization of the successive halving algorithm [23] and the first algorithm for fixed-budget linear bandits. It divides the budget into equally sized epochs and progressively shrinks the set of candidates $\mathcal{Z}_k$. In each epoch, it computes a design that minimizes $\gamma(\mathcal{Z}_k)$ and samples according to a rounded solution. At the end of an epoch, it sorts the remaining items in $\mathcal{Z}_k$ by their estimated rewards and eliminates enough of the items with the smallest estimated rewards to ensure that $\gamma(\mathcal{Z}_{k+1}) \leq \frac{\gamma(\mathcal{Z}_k)}{2}$.
We note that the combinatorial bandit setting satisfies the assumption that
\log(k \log(\arg \max_{\gamma} \phi)) \geq 0 \text{ for all } \gamma \text{ in the transductive linear bandit problem, as well as an upper bound of } |Z| \text{ for all } \gamma.
Proposition 3. then a lower bound is known to scale as
\sum_{i=1}^{z^*} \hat{\Delta}_i \text{ with default value of } \lambda.

On the other hand, it is possible to come up with examples where
\text{ of Theorem 4, the authors developed the first algorithm to provably achieve}
\cite{30, 33, 31} culminating in the formulation of the transductive linear bandit problem in \cite{12} where
Fixed Budget Peace

Algorithm 3: Fixed Budget Peace

**Theorem 6.** Suppose that \gamma((z, z_e)) \geq 1 for all \z \in Z \setminus \{z_e\}. Then, if \gamma \geq c \max(\gamma, \gamma^*), d) \log(\gamma(Z)), Algorithm 3 returns \z \in Z such that
\mathbb{P}(|Z| \neq z_e) \leq 2 \log(\gamma(Z)) \exp(-c \gamma^* \log(\gamma(Z)) = \frac{T}{c \gamma^* \log(\gamma(Z))}).

We note that the combinatorial bandit setting satisfies the assumption that \gamma((z, z_e)) \geq 1 for all \z \in Z \setminus \{z_e\}, but this lower bound is unessential and the algorithm can be modified to accommodate another lower bound. Theorem 3 implies that if \gamma \geq c \log(1/\delta) \log(\gamma(Z)) \log(\log(\gamma(Z))), then Algorithm 3 returns \z with probability at least 1 - \delta. Finally, \log(\gamma(Z)) is \mathcal{O}(\log(d)) in many cases, e.g., combinatorial bandits and in linear bandits when \mathcal{X} = Z.

7 Discussion and Prior Art

**Transductive Linear Bandits:** There is a long line of work in pure-exploration linear bandits \cite{30, 33, 31} culminating in the formulation of the transductive linear bandit problem in \cite{12} where
the authors developed the first algorithm to provably achieve \gamma \log(|Z|/\delta). The sample complexity of Theorem 4, \gamma_e \log(1/\delta), is never worse than \cite{12} since \gamma_e \leq \gamma \log(|Z|) by Proposition 1. On the other hand, it is possible to come up with examples where \gamma_e does not scale with |Z|, but just \gamma_e (see experiments). While our algorithms work for arbitrary \mathcal{X}, Z \subset \mathbb{R}^d, problem instances of combinatorial bandits most clearly illustrate the advances of our new results over prior art.

**Combinatorial Bandits:** The pure exploration combinatorial bandit was introduced in \cite{10}, and followed by \cite{13}. These papers are within a \log(d) factor of the lower bound for the setting where \mathcal{Z} is a matroid. If \Delta_i = \Theta^T z - \max_{z \in \mathcal{Z}} \Theta^T z \text{ when } i \notin z_e \text{ and } \Theta^T z - \max_{z \in \mathcal{Z}} \Theta^T z \text{ otherwise}, then a lower bound is known to scale as \sum_{i=1}^{d} \Delta_i^{-2} \log(1/\delta). The following result shows that \gamma^* is within \log(d) of the lower bound, implying that our sample complexity scales as \sum_{i=1}^{d} \Delta_i^{-2} \log(d/\delta).

**Proposition 3.** Consider the combinatorial bandit setting and suppose that \mathcal{Z} is a matroid. Then, \gamma^* \leq c \log(d) \sum_{i=1}^{d} \Delta_i^{-2} \text{ for some absolute constant } c.

However, in the general setting where \mathcal{Z} is not necessarily a matroid, \cite{8} points out a class with |\mathcal{Z}| = 2 where the sample complexity of \cite{10, 13} is loose by a multiplicative factor of d. Chen et al. \cite{8} was the first to provide a lower bound equivalent to \rho \gamma \log(1/\delta) for the general combinatorial bandit problem, as well as an upper bound of \gamma \log(|\mathcal{Z}|/\delta). However, as stressed in the current work, the log(|\mathcal{Z}|) term is not necessary in many scenarios; for example, in Top-K, \gamma \log(|\mathcal{Z}|) is larger than the best achievable sample complexity by a multiplicative factor of k \cite{9, 8}. This is not in contradiction with the lower bound provided in Theorem 1.9 of \cite{8} which provides a specific worst-case class of instances where the log(|\mathcal{Z}|) is needed.

The next technological leap in combinatorial bandits is the algorithm of \cite{5} (and the follow-up \cite{20}). They provided an algorithm with a novel sample complexity that replaces \log(|\mathcal{Z}|) with a more geometrically inspired term. Define the sphere \mathcal{B}(z, r) = \{z' \in \mathcal{Z} : \|z' - z\|_2 = r\}, and the complexity parameter \varphi_i := \max_{z \in \mathcal{Z} \setminus \{z_e\}} \|z - z_e\|_2 \log(|\mathcal{Z}| \|z - z_e\|_2). Then \cite{5} provide a
We consider two paths
Matching: we use a balanced complete bipartite graph whereas our algorithm pays a cost of roughly $1/|Z| \geq \text{width}-2$ layers that then lead into a sink (see Figure 1 for an illustration). Note that (iii) depicts the relative performance of uniform sampling to FBPeace. Panel (iv) compares the performance of RAGE to Peace on the linear bandits experiment.

Figure 2: Panels (i) and (ii) depict the relative performance of CLUCB and UA to PEACE. Panel (iii) depicts the relative performance of uniform sampling to FBPeace. Panel (iv) compares the performance of RAGE to Peace on the linear bandits experiment.

The first two panels of Figure 2 depict ratio of the average performance of the competing algorithms to the average performance of all our experiments. We consider two combinatorial structures:
(i) Matching: we use a balanced complete bipartite graph $G = (U \cup V, E)$ where $|U| = |V| = 14$. Note that $|Z| = 14! \geq 8 \cdot 10^{10}$. We took two disjoint matchings $M_1$ and $M_2$ and set $\theta_e = 1$ if $e \in M_1$ and $\theta_e = 1 - h$ if $e \in M_2$ for $h \in \{0.15, \ldots, 0.05\}$. Otherwise, $\theta_e = 0$. (ii) Shortest Path: we consider a DAG where a source leads into two disjoint feed-forward networks with 26 width-2 layers that then lead into a sink (see Figure 1 for an illustration). Note that $|Z| \geq 10^8$. We consider two paths $P_1$ and $P_2$ such that they are in the disjoint feed-forward networks. We set $\theta_e = 1$ if $e \in P_1$ and $\theta_e = 1 - h$ if $e \in P_2$ for $h \in \{0.2, 0.15, 0.1, 0.05\}$. Otherwise, $\theta_e = -1$.

The first two panels of Figure 2 depict ratio of the average performance of the competing algorithms to the average performance of our algorithm. In the matching experiment, as the gap between the best matching $M_1$ and the second best matching $M_2$ get smaller, CLUCB pays a cost of roughly $|U|/h^2$ to distinguish $M_1$ from $M_2$ whereas our algorithm pays a cost of roughly $1/h^2$. A similar phenomenon occurs in the shortest path problem.

Multivariate Testing: We consider multivariate testing [16, 15] in which there are $d$ options, each having $k$ possible levels. For example, consider determining the optimal content for a display-ad with slots such as headline, body, etc. and each slot has several variations. A layout is specified by a $d$-tuple $f = (f_1, \ldots, f_d) \in \{1, \ldots, k\}^d$ indicating the level chosen for each option. For each option $I$, $1 \leq I \leq d$ and level $f$, $1 \leq f \leq k$, there is a weight $W^I_f \in \mathbb{R}$, and for each pair of options $I, J$ and factors

We also outperform [5] on a bi-clique problem (see the Supplementary Material).
There is a weight $W_{I,J} \in \mathbb{R}$ capturing linear and quadratic interaction terms respectively. The total reward of a layout $f = (f_1, \cdots, f_d)$ is given by $W_0 + \sum_{I=1}^{d} W_{I,1} f_I + \sum_{I=1}^{d} \sum_{J=1}^{d} W_{I,J} f_I f_J$.

The fixed budget experiment in Figure 2 considered a scenario when $k = 6$ and $d = 3$ and compares Algorithm 3 (FBPeace) to uniform sampling. We set $W_{1,1} = .8$ and $W_{2,3} = .1$ and all other weights to zero, capturing a setting where the three options must be synchronized. At 10000 samples, FBPeace is 30% more likely to return the true optimal layout.

**Linear Bandits.** We considered a setting in $\mathbb{R}^2$, where $\mathcal{X} = \{e_1, \cos(3\pi/4)e_1 + \sin(3\pi/4)e_2\}$ and $\mathcal{Z} = \{\cos(\pi/4 + \phi_i)e_1 + \sin(\pi/4 + \phi_i)e_2\}_{i=1}^{n}$ where $\phi_i \sim \text{Uniform}[0, .05]$. The parameter vector is fixed at $\theta = e_1$. In Figure 2, we see that as the number of arms increases (from $10^3$ to $10^6$), the number of samples by our algorithms is constant, yet grows linearly in $\log(|\mathcal{Z}|)$ for RAGE [12]. This reflects the main goal of the paper - optimal union bounding for large classes.
References

[1] Zeyuan Allen-Zhu, Yuanzhi Li, Aarti Singh, and Yining Wang. Near-optimal discrete optimization for experimental design: A regret minimization approach. Mathematical Programming, pages 1–40, 2020.

[2] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. Concentration inequalities: A nonasymptotic theory of independence. Oxford university press, 2013.

[3] Richard A Brualdi. Comments on bases in dependence structures. Bulletin of the Australian Mathematical Society, 1(2):161–167, 1969.

[4] Sébastien Bubeck et al. Convex optimization: Algorithms and complexity. Foundations and Trends® in Machine Learning, 8(3-4):231–357, 2015.

[5] Tongyi Cao and Akshay Krishnamurthy. Disagreement-based combinatorial pure exploration: Sample complexity bounds and an efficient algorithm. In Conference on Learning Theory, pages 558–588, 2019.

[6] Lijie Chen and Jian Li. On the optimal sample complexity for best arm identification. arXiv preprint arXiv:1511.03774, 2015.

[7] Lijie Chen, Anupam Gupta, and Jian Li. Pure exploration of multi-armed bandit under matroid constraints. In Conference on Learning Theory, pages 647–669, 2016.

[8] Lijie Chen, Anupam Gupta, Jian Li, Mingda Qiao, and Ruosong Wang. Nearly optimal sampling algorithms for combinatorial pure exploration. In Conference on Learning Theory, pages 482–534, 2017.

[9] Lijie Chen, Jian Li, and Mingda Qiao. Nearly instance optimal sample complexity bounds for top-k arm selection. In Artificial Intelligence and Statistics, pages 101–110, 2017.

[10] Shouyuan Chen, Tian Lin, Irwin King, Michael R Lyu, and Wei Chen. Combinatorial pure exploration of multi-armed bandits. In Advances in Neural Information Processing Systems, pages 379–387, 2014.

[11] Eyal Even-Dar, Shie Mannor, and Yishay Mansour. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. Journal of machine learning research, 7(Jun):1079–1105, 2006.

[12] Tanner Fiez, Lalit Jain, Kevin G Jamieson, and Lillian Ratliff. Sequential experimental design for transductive linear bandits. In Advances in Neural Information Processing Systems, pages 10666–10676, 2019.

[13] Victor Gabillon, Alessandro Lazaric, Mohammad Ghavamzadeh, Ronald Ortner, and Peter Bartlett. Improved learning complexity in combinatorial pure exploration bandits. In Artificial Intelligence and Statistics, pages 1004–1012, 2016.

[14] Aurélien Garivier and Emilie Kaufmann. Optimal best arm identification with fixed confidence. In Conference on Learning Theory, pages 998–1027, 2016.

[15] Barbara Heller. Statistics for experimenters, an introduction to design, data analysis, and model building: GEP Box, WG Hunter and JS Hunter, John Wiley and Sons, New York, NY. 1978. Pergamon, 1986.

[16] Daniel N Hill, Houssam Nassif, Yi Liu, Anand Iyer, and SVN Vishwanathan. An efficient bandit algorithm for real-time multivariate optimization. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 1813–1821, 2017.

[17] Steven R Howard, Aaditya Ramdas, Jon McAuliffe, and Jasjeet Sekhon. Uniform, nonparametric, non-asymptotic confidence sequences. arXiv preprint arXiv:1810.08240, 2018.

[18] Daniel Hsu and Sivan Sabato. Heavy-tailed regression with a generalized median-of-means. In International Conference on Machine Learning, pages 37–45, 2014.

[19] IA Ibragimov, VN Sudakov, and BS Tsirelson. Norms of gaussian sample functions. In Proceedings of the third Japan USSR symposium on probability theory, lecture notes in math, volume 550, pages 20–41, 1976.

[20] Lalit Jain and Kevin G Jamieson. A new perspective on pool-based active classification and false-discovery control. In Advances in Neural Information Processing Systems, pages 13992–14003, 2019.
[21] Kevin Jamieson, Matthew Malloy, Robert Nowak, and Sébastien Bubeck. lil’ucb: An optimal exploration algorithm for multi-armed bandits. In Conference on Learning Theory, pages 423–439, 2014.

[22] Shivaram Kalyanakrishnan, Ambuj Tewari, Peter Auer, and Peter Stone. PAC subset selection in stochastic multi-armed bandits. In Proceedings of the 29th International Conference on Machine Learning, ICML 2012, Edinburgh, Scotland, UK, June 26 - July 1, 2012, 2012.

[23] Zohar Karnin, Tomer Koren, and Oren Somekh. Almost optimal exploration in multi-armed bandits. In Sanjoy Dasgupta and David Mcallester, editors, Proceedings of the 30th International Conference on Machine Learning (ICML-13), volume 28, pages 1238–1246. JMLR Workshop and Conference Proceedings, May 2013.

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

[25] Emilie Kaufmann, Olivier Cappé, and Aurélien Garivier. On the complexity of best-arm identification in multi-armed bandit models. The Journal of Machine Learning Research, 17(1):1–42, 2016.

[26] Tor Lattimore and Csaba Szepesvári. Bandit Algorithms. Cambridge University Press, 2020.

[27] Andrea Locatelli, Maurilio Gutzeit, and Alexandra Carpentier. An optimal algorithm for the thresholding bandit problem. In International Conference on Machine Learning, pages 1690–1698, 2016.

[28] Max Simchowitz, Kevin Jamieson, and Benjamin Recht. The simulator: Understanding adaptive sampling in the moderate-confidence regime. In Conference on Learning Theory, pages 1794–1834, 2017.

[29] Marta Soare. Sequential resource allocation in linear stochastic bandits. PhD thesis, 2015.

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

[31] Chao Tao, Saúl Blanco, and Yuan Zhou. Best arm identification in linear bandits with linear dimension dependency. In International Conference on Machine Learning, pages 4877–4886, 2018.

[32] Roman Vershynin. High-Dimensional Probability. 2019.

[33] 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.
A Outline and Notation

Section [B] gives the proof of Theorem [3]. Section [C] presents proofs of the two results for the fixed confidence setting. Section [D] proves the main results on the computational efficiency of Algorithm [2]. Section [E] provides the proof of our upper bound for the fixed budget setting. Section [F] proves various results related to \( \gamma^* \). Section [G] gives additional lower bounds for the transductive linear bandit problem. Section [H] provides a discussion of rounding. Section [I] presents technical lemmas. Section [J] discusses the convexity of \( \gamma^* \). Section [K] discusses the sample complexity results of other papers. Section [L] gives further details on the other experiments.

For the combinatorial bandit setting, we assume wlog that for all \( i \in [d] \) there exist \( z, z' \in \mathcal{Z} \) such that \( i \in z \) and \( i \notin z' \). We will sometimes write \( z \cap z' \) to denote \( (z_1, z'_1, \ldots, z_d, z'_d) \). In a similar way, we will use \( z \Delta z' \) to denote the symmetric difference of \( z \) and \( z' \), viewed as sets. We use \( c, c', \ldots \) to denote positive universal constants whose values may change from line to line.

B Proof of Theorem [3]

Proof of Theorem [3] Define \( \Delta_z = \theta^T(z^* - z) \). Let \( \mathcal{X} = \{x_1, \ldots, x_m\} \). Fix \( \{x_{I_1}, \ldots, x_{I_T}\} \subset \mathcal{X} \) to be the measurement vectors pulled by the algorithm. Define the matrix

\[
X = \begin{pmatrix}
   x_{I_1}^T \\
   \vdots \\
   x_{I_T}^T
\end{pmatrix}
\]

Define \( \hat{\theta} = (X^TX)^{-1}X^TY \).

Let \( \lambda \in \Delta \) be the associated allocation: \( \lambda_i = \frac{1}{T} \sum_{s=1}^T \mathbb{1}\{I_s = i\} \). Note that

\[
E_{\eta \sim N(0, I)} \left[ \sup_{z \in \mathcal{Z} \setminus \{z^*\}} \frac{(z - z^T)(X^TX)^{-1/2}\eta}{\Delta_z} \right] \leq \frac{1}{\sqrt{T}} E_{\eta \sim N(0, I)} \left[ \sup_{z \in \mathcal{Z} \setminus \{z^*\}} \frac{(z - z^T)A(\lambda)^{-1/2}\eta}{\Delta_z} \right]
\]

and

\[
\sup_{z \in \mathcal{Z} \setminus \{z^*\}} \frac{\|z^* - z\|_{(X^TX)^{-1}}}{{\Delta}_z} = \frac{1}{\sqrt{T}} \sup_{z \in \mathcal{Z} \setminus \{z^*\}} \frac{\|z^* - z\|_{A(\lambda)^{-1}}}{{\Delta}_z}.
\]

Since \( x_{I_1}, \ldots, x_{I_T} \) are fixed, the same argument from the proof of Theorem 1 in [12] implies that since \( \mathcal{A} \) is \( \delta \)-PAC,

\[
T \geq \log \left( \frac{1}{4.5\delta} \right) \sup_{z \in \mathcal{Z} \setminus \{z^*\}} \frac{\|z^* - z\|^2_{A(\lambda)^{-1}}}{\theta^T(z^* - z)^2} \geq \sup_{z \in \mathcal{Z}} \frac{\|z^* - z\|^2_{A(\lambda)^{-1}}}{\theta^T(z^* - z)^2}.
\]

(6)

where the second inequality follows since \( \delta \in (0, 1/8) \). Next, we will show that that

\[
T \geq \frac{1}{4} E_{\eta \sim N(0, I)} \left[ \sup_{z \in \mathcal{Z} \setminus \{z^*\}} \frac{(z - z^T)A(\lambda)^{-1/2}\eta}{\theta^T(z^* - z)} \right]^2.
\]

(7)

Note that \( \hat{\theta} \sim N(\theta, (X^TX)^{-1}) \) so that

\[
E \sup_{z \in \mathcal{Z} \setminus \{z^*\}} \frac{(z - z^* - \theta)}{{\Delta}_z} = E_{\eta \sim N(0, I)} \left[ \sup_{z \in \mathcal{Z} \setminus \{z^*\}} \frac{(z - z^* - \theta)(X^TX)^{-1/2}\eta}{\Delta_z} \right] = E_{\eta \sim N(0, I)} \left[ \sup_{z \in \mathcal{Z} \setminus \{z^*\}} \frac{(z - z^* - \theta)\theta^T(z^* - z)^2}{\Delta_z} \right]
\]

where we used the fact that \( (z - z^*)^T(X^TX)^{-1/2}\eta \) and \( (z - z^*)^T(X^TX)^{-1/2}\eta \) are equal in distribution.
By Theorem 5.8 in [2], with probability at least $1 - e^{-1/2}$,

$$
\mathbb{E}_{\eta \sim N(0, I)} \left[ \sup_{z \in \mathcal{Z} \setminus \{z_*\}} \frac{(z^* - z)^T (X^T X)^{-1/2} \eta}{\Delta_z} \right] - \mathbb{E}_{\eta \sim N(0, I)} \left[ \sup_{z \in \mathcal{Z} \setminus \{z_*\}} \frac{(z - z_*)^T (\hat{\theta} - \theta)}{\Delta_z} \right]
$$

$$
\leq \sup_{z \in \mathcal{Z} \setminus \{z_*\}} \frac{\|z^* - z\|(X^T X)^{-1}}{\Delta_z}
$$

Towards a contradiction, suppose that inequality (7) does not hold. Then, with probability at least $1 - e^{-1/2}$ we have

$$
\sup_{z \in \mathcal{Z}} \frac{(z^* - z)^T (\hat{\theta} - \theta)}{\Delta_z} \geq \mathbb{E}_{\eta \sim N(0, I)} \left[ \sup_{z \in \mathcal{Z} \setminus \{z_*\}} \frac{(z^* - z)^T (X^T X)^{-1/2} \eta}{\Delta_z} \right] - \mathbb{E}_{\eta \sim N(0, I)} \left[ \sup_{z \in \mathcal{Z} \setminus \{z_*\}} \frac{(z - z_*)^T (\hat{\theta} - \theta)}{\Delta_z} \right] + \mathbb{E}_{\eta \sim N(0, I)} \left[ \sup_{z \in \mathcal{Z} \setminus \{z_*\}} \frac{(z - z_*)^T A(\lambda)^{-1/2} \eta}{\Delta_z} \right] - 1 \geq 1
$$

where inequality (5) follows from inequality (6) and inequality (9) follows from the negation of inequality (7). Rearranging the above inequality, if (7) fails to hold, then there exists a $z \in \mathcal{Z} \setminus \{z_*\}$ such that

$$
(z - z_*)^T \hat{\theta} > 0.
$$

But, this contradicts the assumption that $\mathcal{A}$ is $\delta$-PAC with $\delta \in (0, 1/8)$. Thus, (7) holds.

We conclude that

$$
T \geq \frac{1}{2} \left( \frac{1}{4} \mathbb{E}_{\eta \sim N(0, I)} \left[ \sup_{z \in \mathcal{Z} \setminus \{z_*\}} \frac{(z - z')^T A(\lambda)^{-1/2} \eta}{\Delta_z} \right] + \log \left( \frac{1}{2.4 \delta} \right) \sup_{z \in \mathcal{Z} \setminus \{z_*\}} \frac{\|z - z\|^2}{\theta^T (z - z)^2} \right)
$$

$$
\geq \frac{1}{2} \left[ \frac{1}{4} \gamma^* + \log \left( \frac{1}{2.4 \delta} \right) \rho^* \right]
$$

\[\square\]

C \quad \text{Fixed Confidence Upper Bound Proofs}

C.1 Peace Algorithm Proofs

Proof of Theorem 14

Step 1: Define a good event. Define $\delta_k = \frac{\delta}{N_k}$. Let $x_1, \ldots, x_{N_k}$ denote the pulled measurement vectors in round $k$. By Theorem 5.8 in [2], with probability at least $1 - e^{-1/2}$

$$
\sup_{z, z' \in \mathcal{Z}_k} \|z - z'\|^2 (\hat{\theta}_k - \theta)
$$

\[\leq \mathbb{E} \left[ \sup_{z, z' \in \mathcal{Z}_k} \|z - z'\|^2 (\hat{\theta}_k - \theta) \right] + \sqrt{2 \log \left( \frac{2k^2}{\delta} \right) \max_{z, z' \in \mathcal{Z}_k} \|z - z'\|^2 (\sum_{i=1}^{N_k} x_i x_i^T)^{-1}}
$$

\[= \mathbb{E}_{\eta \sim N(0, I)} \left[ \sup_{z, z' \in \mathcal{Z}_k} (z - z')^T (\sum_{i=1}^{N_k} x_i x_i^T)^{-1/2} \eta \right] + \sqrt{2 \log \left( \frac{2k^2}{\delta} \right) \max_{z, z' \in \mathcal{Z}_k} \|z - z'\|^2 (\sum_{i=1}^{N_k} x_i x_i^T)^{-1}}
$$

$$
\leq \sqrt{\frac{(1 + \epsilon)}{N_k}} \left[ \mathbb{E}_{\eta \sim N(0, I)} \left[ \sup_{z, z' \in \mathcal{Z}_k} (z - z')^T A(\lambda)^{-1/2} \eta \right] + \sqrt{2 \log \left( \frac{2k^2}{\delta} \right) \max_{z, z' \in \mathcal{Z}_k} \|z - z'\|^2 A(\lambda)^{-1}} \right]
$$

$$
\leq \frac{2(1 + \epsilon) \tau_k}{N_k}
$$

(10)
where inequality (10) follows by the guarantee on the the rounding subroutine ROUND and Lemma 10 and the line (11) uses $\sqrt{a} + \sqrt{b} \leq \sqrt{2a + 2b}$ and the definition of $\tau_k$. Define the events

$$E_k = \{ \sup_{z, z' \in Z_k} |(z - z')^\top (\hat{\theta}_k - \theta)| \leq \sqrt{\frac{2(1+\epsilon)\tau_k}{N_k}} \}$$

$$E = \cap_{k=1}^\infty E_k.$$ 

Note that line (11) implies that $\mathbb{P}(E_k) \geq 1 - \frac{\delta}{k^2}$. Thus, we have

$$\mathbb{P}(E) = \prod_{k=1}^\infty \mathbb{P}(E_k | \cap_{i=1}^{k-1} E_i) \geq \prod_{k=1}^\infty (1 - \frac{\delta}{k^2}) = \frac{\sin(\pi\delta)}{\pi\delta} \geq 1 - \delta$$

where the last line used $\delta \in (0, 1)$. We suppose $E$ holds for the remainder of the proof.

**Step 2: Correctness.** Define $S_k := \{ z \in Z : \theta^\top (z^* - z) \leq B2^{-k} \}$. We show that $z_* \in Z_k$ and $Z_k \subset S_k$ for $k = 2, 3, \ldots$. Using the event $E$, we have that

$$\sup_{z, z' \in Z_k} |(z - z')^\top (\hat{\theta}_1 - \theta)| \leq \sqrt{\frac{2(1+\epsilon)\tau_1}{N_1}} \leq \frac{B}{4}$$

where we used $N_k \geq 2\tau_k (\frac{2(1+\epsilon)}{B})^2 (1 + \epsilon)$. First, fix any $z \notin S_1$. We will then show that $z \notin Z_1$. By definition, $\theta^\top (z_* - z) \geq \frac{B}{2}$. Note that

$$(z_* - z)^\top \hat{\theta}_1 - B2^{-2} = (z_* - z)^\top (\hat{\theta}_1 - \theta) + \theta^\top (z_* - z) - B2^{-2}$$

$$\geq (z_* - z)^\top (\hat{\theta}_1 - \theta) + \frac{B}{4}$$

$$\geq \frac{B}{4} + \frac{B}{4}$$

$$\geq 0$$

where we applied the assumption that $z \in S_1$ and the event. Thus, by the elimination rule, $z \notin Z_1$.

Now, we show that $z_* \in Z_1$. Let $z \in Z_0$. Then, using the event we have that

$$(z - z_*)^\top \hat{\theta}_1 - B2^{-2} = (z - z_*)^\top (\hat{\theta}_1 - \theta) + \theta^\top (z_* - z) - B2^{-2}$$

$$< (z - z_*)^\top (\hat{\theta}_1 - \theta) - B2^{-2}$$

$$\leq \frac{B}{4} - \frac{B}{4}$$

$$= 0.$$ 

This proves the base case.

Next, we prove the inductive step. Suppose that $Z_{k-1} \subset S_{k-1}$; we show that $Z_k \subset S_k$. For any $z, z' \in Z_{k-1}$,

$$|(z - z')^\top (\hat{\theta}_{k-1} - \theta)| \leq \sqrt{\frac{2(1+\epsilon)\tau_{k-1}}{N_{k-1}}} \leq B2^{-(k+1)}.$$ 

Let $z \in S_k^c$ so that $\theta^\top (z^* - z) > B2^{-k}$. Then,

$$(z^* - z)^\top \hat{\theta}_{k-1} - B2^{-(k+1)} = (z^* - z)^\top (\hat{\theta}_{k-1} - \theta) + (z^* - z)^\top \theta - B2^{-(k+1)}$$

$$\geq (z^* - z)^\top (\hat{\theta}_{k-1} - \theta) + B2^{-(k+1)}$$

$$\geq -B2^{-(k+1)} + B2^{-(k+1)}$$

$$= 0.$$ 

Thus, $z \notin Z_k$, proving one part of the inductive step.
Next, we show \( z_* \in Z_k \). By the inductive hypothesis, \( z_* \in Z_{k-1} \). Let \( z \in Z_{k-1} \). Then,
\[
(z - z_*)^\top \tilde{\theta}_{k-1} - B 2^{-(k+1)} = (z - z_*)^\top (\tilde{\theta}_{k-1} - \theta) + (z - z_*)^\top \theta - B 2^{-(k+1)} \\
< (z - z_*)^\top (\tilde{\theta}_{k-1} - \theta) - B 2^{-(k+1)} \\
\leq B 2^{-(k+1)} - B 2^{-(k+1)} = 0
\]

**Step 3: Upper bounding the sample complexity.** Now, we bound the number of samples taken until the algorithm terminates. Since \( Z_k \subset S_k \) for \( k = 2, 3, \ldots \) as we showed in the previous step, once \( k \geq c \log(B/\Delta_{\min}) \), we have that \( Z_k = \{z_1\} \) and thus there are at most \( c \log(B/\Delta_{\min}) \) rounds. In round \( k \), the algorithm takes \( N_k = \lceil 2\tau_k(\frac{z + 1}{B})^2(1 + \epsilon) \rceil \lor q(\epsilon) \) samples and, thus, the sample complexity is bounded by the following sum
\[
\sum_{k=1}^{c \log(B/\Delta_{\min})} N_k \leq c'[\log(B/\Delta_{\min})d + \sum_{k=1}^{c \log(B/\Delta_{\min})} \tau_k \left(\frac{2^k}{B}\right)^2]
\]
where we used \( q(\epsilon) = O(d) \) by the guarantees on the rounding procedure and \( \epsilon = 1/10 \). Now, we focus on upper bounding the second term in the above expression. For \( k = 1 \), then
\[
\tau_1 \left(\frac{2^1}{B}\right)^2 \leq \frac{c}{B} \leq c'
\]
where we used the relation \( B = \tau_1 \lor 1 \).

Next, we bound the terms \( k > 1 \). Note that
\[
\tau_k \left(\frac{2^k}{B}\right)^2 = \mathbb{E}_{\eta \sim N(0, I)} \left[ \max_{z, z' \in Z_k} (z - z')^\top A(\lambda)^{-1/2} \eta^2 \left(\frac{2^k}{B}\right)^2 + 2 \log(\frac{1}{\delta_k}) \right] \max_{z, z' \in Z_k} \|z - z'\|^2_{A(\lambda)^{-1}} \left(\frac{2^k}{B}\right)^2
\]

We begin by bounding the second term. Fix \( \lambda \). Then,
\[
\max_{z, z' \in Z_k} \|z - z'\|^2_{A(\lambda)^{-1}} \left(\frac{2^k}{B}\right)^2 \leq \max_{z, z' \in S_k} \|z - z'\|^2_{A(\lambda)^{-1}} \left(\frac{2^k}{B}\right)^2 \leq c \max_{z \in S_k \setminus \{z_*\}} \|z_* - z\|^2_{A(\lambda)^{-1}} \left(\frac{2^k}{B}\right)^2 \\
\leq c \max_{z \in S_k \setminus \{z_*\}} \|z_* - z\|^2_{A(\lambda)^{-1}} \left(\frac{2^k}{B}\right)^2
\]
where line (14) follows since \( Z_k \subset S_k \) for \( k = 2, 3, \ldots \), line (15) follows since the triangle inequality implies \( \max_{z, z' \in S_k} \|z - z'\|^2_{A(\lambda)^{-1}} \leq \max_{z \in S_k \setminus \{z_*\}} \|z_* - z\|^2_{A(\lambda)^{-1}} \), and line (16) follows since for all \( z \in S_k \setminus \{z_*\} \), \( \Delta_z \leq 2^{-k} B \) by definition. Next, we bound the first term:
\[
\mathbb{E}_{\eta \sim N(0, I)} \left[ \max_{z, z' \in Z_k} \frac{(z - z')^\top A(\lambda)^{-1/2} \eta^2}{2^k B} \right] = 4 \mathbb{E}_{\eta \sim N(0, I)} \left[ \max_{z \in Z_k} \frac{(z_* - z)^\top A(\lambda)^{-1/2} \eta^2}{2^k B} \right] \leq 4 \mathbb{E}_{\eta \sim N(0, I)} \left[ \max_{z \in Z_k} \frac{(z_* - z)^\top A(\lambda)^{-1/2} \eta^2}{2^k B} \right] \leq 4 \mathbb{E}_{\eta \sim N(0, I)} \left[ \max_{z \in Z_k \setminus \{z_*\}} \frac{(z_* - z)^\top A(\lambda)^{-1/2} \eta^2}{2^k B} \right] \leq 8 \mathbb{E}_{\eta \sim N(0, I)} \left[ \max_{z \in Z_k \setminus \{z_*\}} \frac{(z_* - z)^\top A(\lambda)^{-1/2} \eta^2}{2^k B} \right] + \max_{z \in Z_k \setminus \{z_*\}} \frac{\|z_* - z\|^2_{A(\lambda)^{-1}}}{2^k B} \left(\frac{2^k}{B}\right)^2
\]
where line (17) follows by $Z_k \subset S_k$, line (18) follows by Lemma 13, for all $z \in S_k \setminus \{z_\ast\}$, $\Delta z \leq 2^{-k} B$, and $z_\ast \in S_k$, and line (19) follows by Lemma 15. Thus, combining (16) and (19), and taking the infimum over $\lambda$, we obtain

$$
t_k \left( \frac{2^k}{B} \right)^2 \leq c \left[ \inf \theta \sum_{z \in Z \setminus \{z_\ast\}} \left( \frac{(z_\ast - z)\top A(\lambda)^{-1/2} \eta}{\theta} \right)^2 \right]
$$

$$
+ \max_{z \in Z \setminus \{z_\ast\}} \left\| z_\ast - z \right\|^2_{A(\lambda)^{-1}} \log(k^2/\delta)
$$

$$
\leq c' [\gamma_\ast + \rho_\ast \log(k^2/\delta)]
$$

(20)

where line (20) follows by Lemma 12. Thus, combining (12), (13), and (20), we obtain

$$
(21) \text{ and } (22) \text{ together would imply the result. By a similar argument used to establish (21), it suffices to prove}
$$

$$
e \log(B/\Delta_{\min}) \sum_{k=1} \log(B/\Delta_{\min}) - \log\left( \frac{B}{\min_{k:|S_k|>1} F_k} \right) \left[ \gamma_\ast + \rho_\ast \log(\log(B/\Delta_{\min})/\delta) \right].
$$

(22)

Next, we will prove

$$
e \log(B/\Delta_{\min}) \sum_{k=1} \log(B/\Delta_{\min}) - \log\left( \frac{B}{\min_{k:|S_k|>1} F_k} \right) \left[ \gamma_\ast + \rho_\ast \log(\log(B/\Delta_{\min})/\delta) \right].
$$

(22)

where

$$
F_k := \left\{ \begin{array}{ll}
\inf_{\lambda} & \max_{z, z' \in S_k} \| z - z' \|^2_{A(\lambda)^{-1}} \log(2^k/\delta) + \mathbb{E}_{\eta} \left[ \max_{z \in S_k} (z - z')\top A(\lambda)^{-1/2} \eta \right]^2 & k \geq 1 \\
B & k = 0
\end{array} \right.
$$

(21) and (22) together would imply the result. By a similar argument used to establish (21), it suffices to prove

$$
e \log(B/\Delta_{\min}) \sum_{k=2} \log(B/\Delta_{\min}) - \log\left( \frac{B}{\min_{k:|S_k|>1} F_k} \right) \left[ \gamma_\ast + \rho_\ast \log(\log(B/\Delta_{\min})/\delta) \right]
$$

Let $L$ be the largest integer such that $|S_L| > 1$. Define

$$
H_i = \{ k \in [L] : F_k \in \left( \frac{F_0}{2^{-(i+1)}}, \frac{F_0}{2^{-i}} \right) \}.
$$

and define

$$
k_i := \max(k : k \in H_i)
$$

for $i \in \{ \log_2(F_0/F_L) \}$. Then, the sample complexity is upper bounded by

$$
e \log_2(F_0/F_L) \sum_{k=2} \log_2(F_0/F_L) - \log_2(F_0/F_L) \sum_{k=2} \log_2(F_0/F_L) \left[ \gamma_\ast + \rho_\ast \log(\log(B/\Delta_{\min})/\delta) \right]
$$

(23)

$$
= c \sum_{i=1}^{\log_2(F_0/F_L)} \sum_{k \in H_i} F_k \left( \frac{2^k}{B} \right)^2
$$

$$
\leq c' \sum_{i=1}^{\log_2(F_0/F_L)} \max_{k \in H_i} F_k \sum_{k \in H_i} \left( \frac{2^k}{B} \right)^2
$$

$$
\leq c'' \sum_{i=1}^{\log_2(F_0/F_L)} \max_{k \in H_i} F_k \left( \frac{2^{k_i}}{B} \right)^2
$$

(24)

$$
\leq c''' \sum_{i=1}^{\log_2(F_0/F_L)} F_{k_i} \left( \frac{2^{k_i}}{B} \right)^2
$$

$$
\leq c'''' \left[ \log_2(F_0/F_L) \right] \left[ \inf \theta \sum_{z \in Z \setminus \{z_\ast\}} \left( \frac{(z_\ast - z)\top A(\lambda)^{-1/2} \eta}{\theta} \right)^2 \right]
$$

$$
+ \max_{z \in Z \setminus \{z_\ast\}} \left\| z_\ast - z \right\|^2_{A(\lambda)^{-1}} \log(\log(B/\Delta_{\min})/\delta)
$$

(25)
We also define the U

\[ \gamma(z) := \inf_{\lambda \in \Delta} \mathbb{E}_{\eta \sim N(0, I)} [\max_{z \in Z} z^\top A(\lambda)^{-1/2} \eta]^2 \]

\[ \Gamma \leftarrow \Gamma \lor 1, \tilde{\theta}_0 \leftarrow 0 \in \mathbb{R}^d, \hat{\gamma}_0 \leftarrow \frac{\delta}{2\pi^2} ; \]

for \( k = 0, 1, 2, \ldots \) do

\[ \tilde{z}_k \leftarrow \arg \max_{z \in \mathbb{R}^d} \tilde{\theta}_k \cdot z ; \]

\[ \lambda_k, \tau_k \leftarrow \text{ComputeAlloc}(\tilde{z}_k, \tilde{\theta}_k, 2^{-k} \Gamma, \frac{6\delta}{4\pi^2(k+1)^2}) ; \]

\[ \inf_{\lambda \in \Delta} \mathbb{E}_{\eta \sim N(0, I)} [\max_{z \in Z} (\tilde{z}_k - z)^\top A(\lambda)^{-1/2} \eta]^2 ; \]

Set \( N_k \leftarrow 4 \tau_k \log(1/\delta_k) (1 + \epsilon) \lor q(\epsilon) \) and find \( \{x_1, \ldots, x_{N_k}\} \leftarrow \text{ROUND}(\lambda_k, N_k) ; \)

Pull arms \( x_1, \ldots, x_{N_k} \) and receive rewards \( y_1, \ldots, y_{N_k} \) ;

Let \( \tilde{\theta}_{k+1} \leftarrow (\sum_{s=1}^{N_k} x_s x_s^\top)^{-1} \sum_{s=1}^{N_k} x_s y_s ; \)

if \( \text{UNIQUE}(Z, \tilde{\theta}_k, 2^{-k} \Gamma) \) then return \( \tilde{z}_k \).

\[ \text{Algorithm 4: Fixed Confidence Peace with a linear maximization oracle.} \]

We briefly note that the optimization problem in (26) includes \( \gamma(Z) \) as a special case by the following identity:

\[ \mathbb{E}_{\eta \sim N(0, I)} [\max_{z,z' \in Z} (z - z')^\top A(\lambda)^{-1/2} \eta]^2 = 4\mathbb{E}_{\eta \sim N(0, I)} [\max_{z \in Z} z^\top A(\lambda)^{-1/2} \eta]^2. \]

We also define the UNIQUE subroutine (Algorithm 5), originally provided in [8]. It finds the empirical best \( \tilde{z} \) and the empirical second best \( \tilde{z}' \) and determines whether enough samples have been collected to conclude that \( \tilde{z} \) is the best. It uses at most \( d \) calls to the linear maximization oracle.

\[ \text{Input: } Z, \text{ estimate } \hat{\theta}, \text{ shift } b > 0 ; \]

\[ \tilde{z} \leftarrow \arg \max_{z \in Z} \hat{\theta}^\top z ; \]

for \( i = 1, 2, \ldots, d \) s.t. \( i \in \tilde{z} \) do

\[ \tilde{g}^{(i)} = \begin{cases} \hat{\theta}_j & j \neq i \\ -\infty & j = i \end{cases} \]

\[ \tilde{z}^{(i)} \leftarrow \arg \max_{z \in Z} (\tilde{g}^{(i)})^\top z ; \]

if \( \tilde{g}^\top (\tilde{z} - \tilde{z}^{(i)}) - b \leq 0 \) then return False

return True

\[ \text{Algorithm 5: UNIQUE.} \]

**Proof of Theorem 5**: We will first show that if we can solve the optimization problem

\[ \mathbb{E}_{\max_{z \in Z} (z_0 - z)^\top A(\lambda)^{-1/2} \eta} \]

for arbitrary \( \theta_0 \in \mathbb{R}^d, z_0 \in Z, \) and \( b > 0 \), then the sample complexity claim follows. In particular, this implies solving the optimization problems \( \gamma(Z) \) and (26). Then, we will show that solving it
approximately using the subroutine ComputeAlloc only affects up to a constant factor and bound the number of oracle calls.

Step 1: Good event holds with high probability. Define the sets
\[ S_k = \begin{cases} \{ z \in \mathcal{Z} : \Delta_z \leq \Gamma 2^{-k} \} & k \geq 1 \\ \emptyset & k = 0 \end{cases} \]
and define \( \delta_k = \frac{\delta}{2^k} \). Define the events for all \( j \in [k] \)
\[ \Sigma_{k,j} = \{ \sup_{z,z' \in S_j} \| (z - z')^\top (\hat{\theta}_k - \theta) \| \leq \sqrt{2(1 + \epsilon)(1 + \pi \log(1/\delta_k))} \mathbb{E}[\sup_{z,z' \in S_j} (z - z')^\top A(\lambda_k)^{-1/2}]^2 / N_k \} \]
\[ \Sigma_k = \cap_{j=0}^k \Sigma_{k,j} \]
\[ \Sigma = \cap_{k=1}^{\log(T/\Delta_{min})} \cap_{j=0}^k \Sigma_{k,j} \]
Let \( x_1, \ldots, x_{N_k} \) denote the measurement vectors selected in round \( k \). Theorem 5.8 from [2] implies that with probability at least \( 1 - \delta \),
\[ \sup_{z,z' \in S_j} \| (z - z')^\top (\hat{\theta}_{k+1} - \theta) \| \leq \mathbb{E} \sup_{z,z' \in S_j} (z - z')^\top (\hat{\theta}_k - \theta) + \sqrt{2(1 + \frac{1}{\delta_k}) \max_{z,z' \in S_j} \| z - z' \|^2 / (\Sigma_{j=1}^{N_k} z_i x_i^\top)^{-1}} \]
\[ = \mathbb{E} \sup_{z,z' \in S_j} (z - z')^\top (\sum_{i=1}^{N_k} x_i x_i^\top)^{-1/2} \eta + \sqrt{2(1 + \frac{1}{\delta_k}) \max_{z,z' \in S_j} \| z - z' \|^2 / (\Sigma_{j=1}^{N_k} z_i x_i^\top)^{-1}} \]
\[ \leq \mathbb{E} \sup_{z,z' \in S_j} (z - z')^\top (\sum_{i=1}^{N_k} x_i x_i^\top)^{-1/2} \eta + \frac{\pi \log(1/\delta_k) \mathbb{E} \sup_{z,z' \in S_j} (\sum_{i=1}^{N_k} x_i x_i^\top)^{-1/2} \eta^2}{2(1 + \pi \log(1/\delta_k)) \mathbb{E} \sup_{z,z' \in S_j} (\sum_{i=1}^{N_k} x_i x_i^\top)^{-1/2} \eta^2} \]
\[ \leq \mathbb{E} \sup_{z,z' \in S_j} (z - z')^\top (\sum_{i=1}^{N_k} x_i x_i^\top)^{-1/2} \eta \]
\[ \leq \frac{\pi \log(1/\delta_k) \mathbb{E} \sup_{z,z' \in S_j} (\sum_{i=1}^{N_k} x_i x_i^\top)^{-1/2} \eta^2}{2(1 + \pi \log(1/\delta_k)) \mathbb{E} \sup_{z,z' \in S_j} (z - z')^\top A(\lambda_k)^{-1/2} \eta^2 / N_k} \]
where line (27) follows by Lemma [11], line (28) follows by \( \sqrt{a} + \sqrt{b} < \sqrt{2(a + b)} \), and line (29) follows by Lemma [10]. Therefore, \( \mathbb{P}(\Sigma_{k,j}) \leq \frac{\delta}{2^k} \). By law of total probability,
\[ \mathbb{P}(\Sigma) \leq \sum_{k=1}^{\infty} \sum_{j=0}^{k-1} \mathbb{P}(\Sigma_{k,j}) \leq \sum_{k=1}^{\infty} (k + 1) \frac{\delta}{2^k} \leq 3\delta. \]
We suppose the event \( \Sigma \) holds for the rest of the proof.

Step 2: gaps are well estimated every round \( k \) Now, we show that the following hold: at every round \( k \geq 1 \),
1. if \( z \in S_k^c \),
\[ \| (z - z)^\top (\hat{\theta}_k - \theta) \| \leq \frac{\Delta_z}{8} \]
2. if \( z \in S_k \),
\[ \| (z - z)^\top (\hat{\theta}_k - \theta) \| \leq \frac{\Delta}{8} \]
We proceed inductively. First, we prove the base case $k = 1$. On the event $\Sigma_{1,1}$, we have using the definition of $N_1$, for all $z \in \mathcal{Z}$,

$$
| (z_* - z) \top (\hat{\theta}_1 - \theta) | \leq \sup_{z,z' \in \mathcal{Z}} | (z - z') \top (\hat{\theta}_1 - \theta) | \leq \frac{2(1 + \epsilon)(1 + \pi \log(1/\delta_k)) \mathbb{E}[\sup_{z,z' \in \mathcal{Z}} (z - z') \top A(\lambda)^{-1/2} \eta]^2}{N_0} \leq \frac{8(1 + \epsilon)(1 + \pi \log(1/\delta_k)) \mathbb{E}[\sup_{z \in \mathcal{Z}} (\hat{z}_0 - z) \top A(\lambda)^{-1/2} \eta]^2}{N_0} \leq \frac{2^{1-\Gamma}}{8} \quad (30)
$$

where in the last line we used $N_k = \alpha \left[ \tau_k \log(1/\delta_k)(1 + \epsilon) \right] \lor q(\epsilon)$. Observe that whether $z \in S_1$ or $z \in S_1^c$ the base case follows. Next, we show the inductive step. Suppose that at round $k \geq 1$, if $z \in S_k$,

$$
| (z_* - z) \top (\hat{\theta}_k - \theta) | \leq \frac{\Delta_k}{8}
$$

and if $z \in S_k$,

$$
| (z_* - z) \top (\hat{\theta}_k - \theta) | \leq \frac{2^{-k-\Gamma}}{8}.
$$

Now, consider round $k + 1$. Fix $z_0 \in S_{k+1}$. If $\Delta_k \geq \frac{\Gamma}{2}$, there is nothing to show by (30). Thus, suppose $\Delta_k \leq \frac{\Gamma}{2}$. Then, there exists $j \leq k$ such that $\Gamma 2^{-(j+1)} \leq \Delta_{z_0} \leq \Gamma 2^{-j}$. Then,

$$
\frac{| (z_* - z_0) \top (\hat{\theta}_{k+1} - \theta) |}{\Delta_{z_0}} \leq \sup_{z,z' \in \mathcal{Z}_j} \left| (z - z') \top (\hat{\theta}_{k+1} - \theta) \right| \leq \frac{2(1 + \epsilon)(1 + \pi \log(1/\delta_k)) \mathbb{E}[\sup_{z,z' \in \mathcal{Z}_j} (z - z') \top A(\lambda)^{-1/2} \eta]^2}{N_k} \leq \frac{8(1 + \epsilon)(1 + \pi \log(1/\delta_k)) \mathbb{E}[\sup_{z \in \mathcal{Z}_j} (\hat{z}_k - z) \top A(\lambda)^{-1/2} \eta]^2}{N_k} \leq \frac{36(1 + \epsilon)(1 + \pi \log(1/\delta_k)) \mathbb{E}[\sup_{z \in \mathcal{Z}_j} (\hat{z}_k - z) \top A(\lambda)^{-1/2} \eta]^2}{N_k} \leq \frac{36(1 + \epsilon)(1 + \pi \log(1/\delta_k)) \mathbb{E}[\sup_{z \in \mathcal{Z}} (\hat{z}_k - z) \top A(\lambda)^{-1/2} \eta]^2}{N_k} \leq \frac{162(1 + \epsilon)(1 + \pi \log(1/\delta_k)) \mathbb{E}[\sup_{z \in \mathcal{Z}} (\hat{z}_k - z) \top \theta_k + 2^{-k}\Gamma]^2}{N_k} \leq \frac{1}{8} \quad (31)
$$

$$
\frac{1}{8} \quad (32)
$$

$$
\frac{1}{8} \quad (33)
$$

$$
\frac{1}{8} \quad (34)
$$

$$
\frac{1}{8} \quad (35)
$$

where line (31) follows by the event $\Sigma$, line (32) follows from Lemma 13 since $\hat{z}_k \in S_j$ and for all $z \in S_j$, $3\Delta_{z_0} \geq \Delta_z + 2^{-k} \Gamma$, (34) follows by the inductive hypothesis and Lemma 1 and (35) follows by the definition of $N_k$. Next, fix $z_0 \in S_{k+1}$; a similar series of inequalities shows that

$$
| (z_* - z_0) \top (\hat{\theta}_{k+1} - \theta) | \leq \frac{2^{-(k+1)} \Gamma}{8}.
$$

yielding the claim.

**Step 3: Correctness.** To show correctness, it suffices to show that at round $k$, if $z_* \neq z_*$, then the $\text{UNIQUE}(\mathcal{Z}, \hat{\theta}_k, 2^{-k} \Gamma)$ returns false. Inspection of the subroutine reveals that it suffices to show that
We bound the first term as follows. Fix where we used the relation \( \Gamma = (\tilde{z}_k - z_*)^\top (\tilde{\theta}_k - \theta) - \Delta_{\tilde{z}_k} - 2^{-k}\Gamma \)

\[
\leq \max\left(\frac{\Delta_{\tilde{z}_k}}{8}, \frac{2^{-k}\Gamma}{8}\right) - \Delta_{\tilde{z}_k} - 2^{-k}\Gamma \\
\leq 0
\]

proving correctness.

**Step 4: Upper bound the sample complexity.** Note that at round \( k \), UNIQUE(\( Z, \tilde{\theta}_k, 2^{-k}\Gamma \)) checks whether the gap between \( \tilde{z}_k \) and \( \arg\max_{z \neq \tilde{z}_k} \tilde{\theta}_k^\top z \) is at least \( 2^{-k}\Gamma \), and terminates if it is. Thus, by the claim in Step 2, the algorithm terminates and outputs \( z_k \) once \( k \geq c \log(\Gamma/\Delta_{\min}) \). Thus, the sample complexity is upper bounded by

\[
\sum_{k=1}^{\log(\Gamma/\Delta_{\min})} N_k \leq c'[\log(\Gamma/\Delta_{\min})d \sum_{k=1}^{c\log(\Gamma/\Delta_{\min})} \tau_k(\frac{2^k}{\Gamma})^2]
\]

(36)

where we used \( q(\epsilon) = O(d) \) by the guarantees on the rounding procedure and \( \epsilon = 1/10 \). Now, we focus on upper bounding the second term in the above expression. For \( k = 1 \), then

\[
\tau_1(\frac{2^1}{\Gamma})^2 \leq c \leq c'
\]

(37)

where we used the relation \( \Gamma = \tau_1 \lor 1 \). Thus, to obtain the upper bound on the sample complexity, it suffices to upper bound

\[
\tau_k = \inf_{\lambda \in \Delta} \mathbb{E}_{\eta \sim N(0, I)}[\max_{z \in Z} \frac{(\tilde{z}_k - z)^\top A(\lambda)^{-1/2}\eta}{2^{-k}\Gamma + \tilde{\theta}_k^\top (\tilde{z}_k - z)}]^2
\]

for \( k > 1 \). Fix \( \lambda \in \Delta \). We have that

\[
\mathbb{E}_{\eta \sim N(0, I)}[\max_{z \in Z} \frac{(\tilde{z}_k - z)^\top A(\lambda)^{-1/2}\eta}{2^{-k}\Gamma + \tilde{\theta}_k^\top (\tilde{z}_k - z)}]^2 \leq c\mathbb{E}_{\eta \sim N(0, I)}[\max_{z \in Z} \frac{(\tilde{z}_k - z)^\top A(\lambda)^{-1/2}\eta}{2^{-k}\Gamma + \Delta_z}]^2
\]

\[
\leq c'[\mathbb{E}_{\eta \sim N(0, I)}[\max_{z \in Z} \frac{(z_* - z)^\top A(\lambda)^{-1/2}\eta}{2^{-k}\Gamma + \Delta_z}]^2
\]

\[
+ \mathbb{E}_{\eta \sim N(0, I)}[\max_{z \in Z} \frac{(z_* - z)^\top A(\lambda)^{-1/2}\eta}{2^{-k}\Gamma + \Delta_z}]^2
\]

(38)

We bound the first term as follows. Fix \( z_0 \in Z \setminus \{z_*\} \).

\[
\mathbb{E}_{\eta \sim N(0, I)}[\max_{z \in Z} \frac{(z_* - z)^\top A(\lambda)^{-1/2}\eta}{2^{-k}\Gamma + \Delta_z}]^2
\]

\[
= \mathbb{E}_{\eta \sim N(0, I)}[\max_{z \in Z \setminus \{z_*\}} \frac{(z_* - z)^\top A(\lambda)^{-1/2}\eta}{2^{-k}\Gamma + \Delta_z}]^2
\]

\[
\leq \mathbb{E}_{\eta \sim N(0, I)}[\max_{z \in Z \setminus \{z_*\}} \frac{(z_* - z)^\top A(\lambda)^{-1/2}\eta}{2^{-k}\Gamma + \Delta_z}]^2
\]

\[
\leq 8\mathbb{E}_{\eta \sim N(0, I)}[\max_{z \in Z \setminus \{z_*\}} \frac{(z_* - z)^\top A(\lambda)^{-1/2}\eta}{2^{-k}\Gamma + \Delta_z}]^2 + 8\frac{\|z_* - z_0\|_{A(\lambda)^{-1}}^2}{(2^{-k}\Gamma + \Delta_z)^2}
\]

\[
\leq 8\mathbb{E}_{\eta \sim N(0, I)}[\max_{z \in Z \setminus \{z_*\}} \frac{(z_* - z)^\top A(\lambda)^{-1/2}\eta}{\Delta_z}]^2
\]

(39)

where line (38) follows by exercise 7.6.9 in [32].
It remains to bound the second term. Note that
\[
E_{\eta \sim N(0, I)} \left[ \max_{z \in Z} \frac{(z_k - \hat{z}_k)^T A(\lambda)^{-1/2} \eta}{2^{-k} \Gamma} + \Delta_z \right]^2 
\leq E_{\eta \sim N(0, I)} \left[ \max_{z \in Z} \frac{(z_k - \hat{z}_k)^T A(\lambda)^{-1/2} \eta}{2^{-k} \Gamma} \right]^2
\leq c \frac{\|z_k - \hat{z}_k\|^2}{(2^{-k} \Gamma)^2} A(\lambda)^{-1}
\leq c \frac{\|z_k - \hat{z}_k\|^2}{\Delta_{\hat{z}_k}} A(\lambda)^{-1}
\leq c \max_{z \in Z \setminus \{z_k\}} \frac{\|z_k - z\|^2}{\Delta_z^2}
\] (40)

where line (40) follows since \( \hat{z}_k \in S_{k+2} \) by Lemma 1.

Thus, combining (36), (37), (39), and (41) yield the upper bound
\[
\log(\Gamma / \Delta_{\min}) \sum_{k=1}^N N_k \leq c \log(\Gamma / \Delta_{\min}) [d + \gamma^* + \rho^*].
\]

**Step 5: Computation.** Next, we show that we can solve the optimization problems \( \gamma(Z) \) and (26) approximately and bound the number of oracle calls. In the interest of brevity, define
\[
g_k(\lambda) := E_{\eta \sim N(0, I)} \left[ \max_{z \in Z} \frac{(z_k - z)^T A(\lambda)^{-1/2} \eta}{2^{-k} \Gamma + \theta_k} (z_k - z) \right]^2
\]

Let \( D_{k,1} \) denote the event that GetAlloc\((\hat{z}_k, \hat{\theta}_k, 2^{-k} \Gamma, \frac{6\delta}{\pi^2 (k+1)^2}) \) returns \( \lambda_k \in \Delta \) such that
\[
g_k(\lambda_k) \leq c \left[ \inf_{\lambda \in \Delta} g_k(\lambda) + 1 \right]
\] (42)

Let \( D_{k,1} \) denote the event that GetAlloc\((\hat{z}_k, \hat{\theta}_k, 2^{-k} \Gamma, \frac{6\delta}{\pi^2 (k+1)^2}) \) returns at most the following number of oracle calls
\[
c [d + \log(\phi \cdot k^2) + \log(\log(d) \cdot 2^{3} \frac{1}{(2-k)^2} d^2 \delta^2)] \log(d) \cdot 2^3 \frac{k^4}{(2-k)^2} \delta^2
\] (43)

where \( \phi \leq \max, \Delta, +, \Gamma \). Furthermore, define \( D_1 = \cap_k D_{k,1} \) and \( D_2 = \cap_k D_{k,1} \).

GetAlloc is applied with confidence level \( \frac{6\delta}{\pi^2 (k+1)^2} \), and thus by Theorem 3 and a standard union bound argument, with probability at least \( \mathbb{P}(D_1) \geq 1 - \frac{\phi}{d \delta} \) and \( \mathbb{P}(D_2) \geq 1 - \frac{\phi}{d \delta} \).

Next, let \( C_k \) denote that event that EvalAlloc\((\hat{z}_k, \hat{\theta}_k, 2^{-k} \Gamma, \frac{6\delta}{\pi^2 (k+1)^2}) \) that the algorithm outputs a \( \tau_k \) such that
\[
g_k(\lambda_k) \leq \tau_k \leq c [g_k(\lambda_k) + 1]
\] (44)

and the number of oracle calls is upper bounded by
\[
O\left( \frac{d^2}{(2-k)^2} \log(k/d) \log\left( \frac{dk}{2^{-k} \delta} \right) \right)
\]

Define \( C = \cap_k C_k \). Since EvalAlloc is applied with confidence level \( \delta = \frac{6\delta}{4\pi^2 k^2} \) and by Lemma 3 and a standard union bound argument, \( \mathbb{P}(C) \geq 1 - \frac{\phi}{d \delta} \).

Suppose that \( D_1 \cap C \cap E \) occurs. Inspection of the proof reveals that nothing is lost by the approximation in (62) and (44). Thus, by a union bound, it follows that with probability at least \( 1 - 4\delta \), the algorithm terminates and returns \( z_k \) after the stated number of samples in the theorem.

Now, suppose \( D_1 \cap D_2 \cap C \cap E \) holds. Since there are \( c \log(\Gamma / \Delta_{\min}) \) rounds, the bound on the number of oracle calls follows by the dominant term appearing in line (43). Thus, by the union bound and assuming \( \delta \geq \frac{1}{2^2} \), the event \( D_1 \cap D_2 \cap C \cap E \) occurs with probability at least \( 1 - 4\delta \). This completes the proof.
The following Lemma is an essential ingredient in the proof of the upper bound for the computationally efficient algorithm for combinatorial bandits.

**Lemma 1.** Let \( k \geq 1 \). Consider the \( k \)th round of Algorithm 2. Suppose that

- if \( z \in S_k^c \),
  \[
  |(z_\ast - z)^\top (\hat{\theta}_k - \theta)| \leq \frac{\Delta_z}{8} \tag{45}
  \]
- if \( z \in S_k \),
  \[
  |(z_\ast - z)^\top (\hat{\theta}_k - \theta)| \leq \frac{2^{-k}\Gamma}{8}. \tag{46}
  \]

Then, the following hold:

1. \( \bar{z}_k \in S_{k+2} \), \tag{47}

2. if \( z \in S_k^c \)
  \[
  |(\bar{z}_k - z)^\top \hat{\theta}_k - (z_\ast - z)^\top \theta| \leq \frac{1}{2}\Delta_z. \tag{48}
  \]

3. if \( z \in S_k \)
  \[
  |(\bar{z}_k - z)^\top \hat{\theta}_k - (z_\ast - z)^\top \theta| \leq \frac{1}{2}2^{-k}\Gamma. \tag{49}
  \]

4. There exist universal constants \( c, c' > 0 \) such that
  \[
  c\mathbb{E}[\sup_{z \in \tilde{z}} \frac{(\bar{z}_k - z)^\top A(\lambda)^{-1/2}\eta}{\Delta_z + 2^{-k}\Gamma}]^2 \leq \mathbb{E}[\sup_{z \in \tilde{z}} \frac{(\bar{z}_k - z)^\top \hat{\theta}_k}{\Delta_z + 2^{-k}\Gamma}]^2 
  \leq c'E[\sup_{z \in \tilde{z}} \frac{(\bar{z}_k - z)^\top A(\lambda)^{-1/2}\eta}{\Delta_z + 2^{-k}\Gamma}]^2
  \]

**Proof.**

**Step 1: 1 holds at round \( k \).** Note that if \( z \in S_{k+2}^c \cap S_k \), then

\[
\hat{\theta}_k^\top (z_\ast - z) \geq \Delta_z - \frac{2^{-k}\Gamma}{8} > 0
\]

by (46) and since \( z \in S_{k+2}^c \cap S_k \) implies that \( \Delta_z \geq \frac{2^{2k+4}}{4} \). Thus, \( z \neq \bar{z}_k \). On the other hand, if \( z \in S_{k}^c \),

\[
\hat{\theta}_k^\top (z_\ast - z) \geq \Delta_z - \frac{\Delta_z}{8} > 0
\]

by (45), so that \( z \neq \bar{z}_k \). Together, these cases together imply that \( \bar{z}_k \in S_{k+2} \).

**Step 2: 2 and 3 hold at round \( k \).** First, suppose \( z \in S_k^c \). We have that

\[
|(\bar{z}_k - z)^\top \hat{\theta}_k - (z_\ast - z)^\top \theta| \leq |(\bar{z}_k - z)^\top (\hat{\theta} - \theta)| + |\theta^\top (\bar{z}_k - z) - \theta^\top (z_\ast - z)|
\leq |(\bar{z}_k - z_\ast)^\top (\hat{\theta} - \theta)| + |(z_\ast - z)^\top (\hat{\theta} - \theta)| + |\theta^\top (\bar{z}_k - z_\ast)|
\leq \frac{1}{8}(2^{-k}\Gamma + \Delta_z) + \frac{1}{4}2^{-k}\Gamma \tag{50}
\]

where line (50) follows by (45) and by (47) which we have shown holds at round \( k \). By a similar argument, if \( z \in S_k \),

\[
|(\bar{z}_k - z)^\top \hat{\theta}_k - (z_\ast - z)^\top \theta| \leq \frac{1}{2}2^{-k}\Gamma.
\]
Therefore, since in addition yielding one of the inequalities. By a similar argument, we obtain the other inequality, proving the mixture with the uniform distribution.

In this section, we present the computational subroutines for the computationally efficient algorithm for Combinatorial Bandits.

**D Computational Results for Computationally Efficient Algorithm for Combinatorial Bandits**

In this section, we present the computational subroutines for the computationally efficient algorithm for combinatorial bandits. The main optimization problem in Algorithm 2 is given in line (26). Fix \( z_0 \in \mathcal{Z}, b > 0, \) and \( \theta_0 \in \mathbb{R}^d \) for the remainder of the section; we will omit dependence on these quantities because they are fixed. Since the Gaussian width is nonnegative, it suffices to solve:

\[
\inf_{\lambda \in \Delta} g(\lambda) := E_{\max z \in \mathcal{Z}} \frac{(z_0 - z)^\top A(\lambda)^{-1/2} \eta}{b + \theta_0^\top (z_0 - z)}.
\]

Define the following functions

\[
g(\lambda; \eta) := \max_{z \in \mathcal{Z}} \frac{(z_0 - z)^\top A(\lambda)^{-1/2} \eta}{b + \theta_0^\top (z_0 - z)}
\]

\[
g(\lambda; \eta; z) := \frac{(z_0 - z)^\top A(\lambda)^{-1/2} \eta}{b + \theta_0^\top (z_0 - z)}
\]

\[
g(\lambda; \eta; r) := \max_{z \in \mathcal{Z}} z^\top (A(\lambda)^{-1/2} \eta + r \theta_0) - r(b + \theta_0^\top z_0) - z_0^\top A(\lambda)^{-1/2} \eta
\]

\[
g(\lambda; \eta; r; z) := z^\top (A(\lambda)^{-1/2} \eta + r \theta_0) - r(b + \theta_0^\top z_0) - z_0^\top A(\lambda)^{-1/2} \eta
\]

**D.1 Main Subroutine**

**Algorithm 6: ComputeAlloc\((z_0, \theta_0, b, \delta)\)**

\[\text{ComputeAlloc}(z_0, \theta_0, b, \delta)\] is the main subroutine; it solves and evaluates \( \inf_{\lambda} g(\lambda). \) GetAlloc\((z_0, \theta_0, b, \delta)\) and EvalAlloc\((z_0, \theta_0, b, \lambda, \delta)\) only use calls to the linear maximization oracle. GetAlloc\((z_0, \theta_0, b, \delta)\) finds a solution within a constant additive factor of the optimal solution to the optimization problem \( \inf_{\lambda \in \Delta} g(\lambda) \) with probability at least \( 1 - \delta. \) EvalAlloc\((z_0, \theta_0, b, \lambda, \delta)\) determines the value of \( g(\lambda) \) within a constant additive factor with probability at least \( 1 - \delta. \)

GetAlloc (Algorithm 7) performs stochastic mirror descent over the subset of the simplex that is a mixture with the uniform distribution

\[
\tilde{\Delta} := \{ \lambda \in \mathbb{R}^d : \lambda = \frac{1}{2}(\kappa + \kappa') \text{ where } \kappa \in \Delta \text{ and } \kappa' = (1/d, \ldots, 1/d)^\top \}.
\]
Define the Bregman divergence associated with a function $f$:

$$D_f(x, y) = f(x) - f(y) - \nabla f(y)^T (x - y).$$

GetAlloc calls estimateGradient (Algorithm 8) to obtain an unbiased estimate of the gradient. estimateGradient needs to solve a maximization problem, for which it calls computeMax (Algorithm 9), a subroutine that essentially performs binary search.

**Algorithm 7: GetAlloc($z_0, \theta_0, b, \delta$): Stochastic Mirror Descent for Transductive Bandits with linear maximization oracle**

**Algorithm 8: estimateGradient($z_0, \theta_0, b, \lambda$): Compute unbiased stochastic subgradient**

**Algorithm 9: computeMax($z_0, \theta_0, b, \lambda, \eta, \text{TOL}$): Compute $g(\lambda; \eta)$**

EvalAlloc (Algorithm 10) estimates the number of samples to take in a round, only using calls to the linear maximization oracle. Because it estimates the mean of estimator that is not necessarily sub-Gaussian, but has controlled variance, this subroutine uses the median-of-means estimator.
We claim that we can interchange the expectation and differentiation. Note that
\[
\text{arg max}
\]
Algorithm 8 identifies

As a consequence, to show that estimateGradient returns an unbiased gradient, it suffices to show that

The estimateGradient algorithm terminates once it finds

\[
\phi
\]
\[
\begin{align*}
\text{Input: } & \lambda \in \Delta, z_0 \in \mathbb{Z}, \theta_0 \in \mathbb{R}^d, \text{ Offset } b \in \mathbb{R}; \\
& T \leftarrow 864 \frac{d^2}{\xi^2} \log(1/\delta); \\
& \text{Draw } \eta_1, \ldots, \eta_T \sim N(0, I); \\
& y_s \leftarrow \text{computeMax}(z_0, \theta_0, b, \lambda, \eta_s, \text{TOL} = 1/2) \text{ for } s = 1, \ldots, T; \\
& \text{Let } \tau \text{ be the output of the median of means estimator applied to } y_1, \ldots, y_T; \\
& \text{Return } |\tau + 1|^2
\end{align*}
\]

Algorithm 10: EvalAlloc(z_0, \theta_0, b, \lambda): Estimate g(\lambda)

D.2 Proofs

Recall the definitions:
\[
g(\lambda; \eta) := \max_{z \in \mathbb{Z}} \frac{(z_0 - z)^\top A(\lambda)^{-1/2} \eta}{b + \theta_0^\top (z_0 - z)}
\]
\[
g(\lambda; \eta; z) := \frac{(z_0 - z)^\top A(\lambda)^{-1/2} \eta}{b + \theta_0^\top (z_0 - z)}
\]
\[
g(\lambda; \eta; r) := \max_{z \in \mathbb{Z}} z^\top (A(\lambda)^{-1/2} \eta + r \theta_0) - r(b + \theta_0^\top z_0) - z_0^\top A(\lambda)^{-1/2} \eta
\]
\[
g(\lambda; \eta; r; z) := z^\top (A(\lambda)^{-1/2} \eta + r \theta_0) - r(b + \theta_0^\top z_0) - z_0^\top A(\lambda)^{-1/2} \eta
\]

The following Lemma provides the guarantee for estimateGradient. Define \( \phi = \sqrt{\max_{z \in \mathbb{Z}} \theta_0^\top (z_0 - z) + b}. \)

**Lemma 2.** Consider the combinatorial bandit setting. Fix \( z_0 \in \mathbb{Z}, b > 0, \theta_0 \in \mathbb{R}^d, \text{ and } \lambda \in \Delta. \)

estimateGradient(z_0, \theta_0, b, \lambda) returns an unbiased stochastic gradient of the function g(\lambda) with probability 1. Let \( \xi > 0. \) With probability at least \( 1 - \frac{25}{2^9}, \) it terminates after \( O(d + \log(\frac{d}{\xi}) + \log(\frac{\xi}{\delta})) \) oracle calls.

**Proof.**

**Step 1: Correctness.** Let \( \eta \sim N(0, I). \) Note that \( \mathbb{E} g(\lambda; \eta) = g(\lambda). \) Since \( \eta \sim N(0, I), \) with probability 1 arg max_{z \in \mathbb{Z}} g(\lambda; \eta; z) is unique and, therefore,
\[
\nabla_{\lambda} \max_{z \in \mathbb{Z}} g(\lambda; \eta; z) = \nabla_{\lambda} g(\lambda; \eta; \text{arg max}_{z \in \mathbb{Z}} g(\lambda; \eta; z)).
\]

We claim that we can interchange the expectation and differentiation. Note that
\[
\mathbb{E} \nabla_{\lambda} g(\lambda; \eta; \text{arg max}_{z \in \mathbb{Z}} g(\lambda; \eta; z)) = (\nabla_{\lambda} \mathbb{E} g(\lambda; \eta))_{i} \quad \forall i.
\]

Since
\[
|\nabla_{\lambda} g(\lambda; \eta)_i| \leq \frac{\lambda_i^{-3/2} \eta_i}{b + \theta_0^\top (z_0 - z)}
\]
and \( \mathbb{E}_{\eta} \left| \frac{\lambda_i^{-3/2} \eta_i}{b + \theta_0^\top (z_0 - z)} \right| < \infty, \) we have by standard results on exchanging differentiation and expectation that the claim follows. Thus, we have
\[
\mathbb{E} \nabla_{\lambda} g(\lambda; \eta; \text{arg max}_{z \in \mathbb{Z}} g(\lambda; \eta; z)) = \mathbb{E} \nabla_{\lambda} \max_{z \in \mathbb{Z}} g(\lambda; \eta; z) = \nabla_{\lambda} \mathbb{E} \max_{z \in \mathbb{Z}} g(\lambda; \eta; z).
\]

As a consequence, to show that estimateGradient returns an unbiased gradient, it suffices to show that Algorithm 8 identifies \( \text{arg max}_{z \in \mathbb{Z}} g(\lambda; \eta; \bar{r}; z) = 0. \)

Let \( \bar{z} \in \text{arg max}_{z \in \mathbb{Z}} g(\lambda; \eta; \bar{r}; z). \) Then,
\[
0 = \text{max}_{z \in \mathbb{Z}} \frac{g(\lambda; \eta; \bar{r}; z)}{g(\lambda; \eta; \bar{r}; \bar{z})} = \frac{\bar{z}^\top (A^{-1/2}(\lambda) \eta + \bar{r} \theta_0) - \bar{r}(b + \theta_0^\top z_0) - z_0^\top A(\lambda)^{-1/2} \eta}{\bar{z}^\top (A^{-1/2}(\lambda) \eta + \bar{r} \theta_0) - \bar{r}(b + \theta_0^\top z_0) - z_0^\top A(\lambda)^{-1/2} \eta}
\]
\[
> \bar{z}^\top (A^{-1/2}(\lambda) \eta + \bar{r} \theta_0) - \bar{r}(b + \theta_0^\top z_0) - z_0^\top A(\lambda)^{-1/2} \eta.
\]
where the strict inequality holds with probability 1 since \( \eta \sim N(0, I) \). Rearranging the above inequality, this implies that for every all \( z \in \mathcal{Z} \) \( \notin \{ \hat{z} \} \)

\[
(z_0 - z) \top A(\lambda)^{-1/2} \eta \leq \bar{r} = (z_0 - \hat{z}) \top A(\lambda)^{-1/2} \eta
\]

implying that \( \hat{z} = \arg \max_z g(\lambda; \bar{r}; z) \), showing \texttt{estimateGradient} returns an unbiased gradient.

**Step 2: Running time.** Next, we bound the number of oracle calls. Define \( \hat{y} = \sup_{z \in \mathcal{Z}} \frac{(z_0 - z)^\top A(\lambda)^{-1/2} \eta}{b + \hat{\theta}_0^\top (z_0 - z)} \). By Theorem 5.8 of \cite{2}, we have that

\[
\mathbb{V}(\hat{y}) \leq 4 \sup_{z \in \mathcal{Z}} \mathbb{V}(\frac{(z_0 - z)^\top A(\lambda)^{-1/2} \eta}{b + \hat{\theta}_0^\top (z_0 - z)}) \leq 8 \frac{d^2}{b^2}.
\]

where we used \( \lambda \in \bar{\Delta} \). Define the event

\[
\mathcal{E} = \{ \sup_{z \in \mathcal{Z}} (z_0 - z)^\top A(\lambda)^{-1/2} \eta \leq 8 \frac{d^2}{b^2} \}
\]

Thus, by Chebyshev’s inequality, we have that

\[
\mathbb{P}(\mathcal{E}^c) \leq \delta. \tag{51}
\]

Thus, choosing \( \delta = \frac{\delta}{2} \), we have with probability at least \( \mathbb{P}(\mathcal{E}^c) \leq \frac{\delta}{2} \). Then, by Lemma 5 the first while loop requires

\[
O(\log(\frac{d}{b^2})) = O(\log(\frac{d}{b}) + d + \log(\frac{1}{\xi}))
\]

oracle calls.

Next, we consider the second while loop. Define the event

\[
\mathcal{D} = \{ |g(\lambda; \eta; z) - g(\lambda; \eta; z')| > \frac{\xi}{\phi^{2d+1}} \text{, } \forall z \neq z' \in \mathcal{Z} \}.
\]

By Lemma 4 we have that with probability at least \( \mathbb{P}(\mathcal{D}) \geq 1 - \frac{\xi}{\phi^{2d+1}} \). Then, by Lemma 5 the second while loop requires at most \( O(d + \log(\frac{d}{\xi}) + \log(\frac{d}{\xi})) \) oracle calls. A standard union bound argument for event \( \mathcal{E} \cap \mathcal{D} \) yields the result.

\( \Box \)

The following Theorem provides the guarantee for GetAlloc.

**Theorem 7.** Consider the combinatorial bandit setting. Fix \( z_0 \in \mathcal{Z} \), \( b > 0 \), and \( \theta_0 \in \mathbb{R}^d \). With probability at least \( 1 - \delta \) GetAlloc\( (z_0, \theta_0, b, \delta) \) returns \( \hat{\lambda} \in \Delta \) such that

\[
g(\hat{\lambda})^2 \leq c \min_{\lambda \in \Delta} g(\lambda)^2 + 1.
\]

Let \( \xi > 0 \). Furthermore, with probability at least \( 1 - \frac{2\xi}{\eta^{3/2}} \), the number of oracle calls is bounded above by

\[
c[d + \log(\phi/\xi) + \log(\log(d))2^{d^3/\delta^2} \frac{1}{\delta^2}] \log(d)^2 \frac{d^3}{\delta^2} \frac{1}{\delta^2}.
\]

**Proof.**

**Step 1: Guarantee on final allocation.** Note that for any \( z \in \mathcal{Z} \),

\[
|\nabla_\lambda g(\lambda; \eta; z)| = \mathbb{1} \{ i \in |z_0 \Delta z| \} |\frac{\lambda_i^{-3/2} \eta_i}{b + \hat{\theta}_0^\top (z_0 - z)}|
\]

and thus

\[
\mathbb{E}_{\max_{z \in \mathcal{Z}}} \|\nabla_\lambda g(\lambda; \eta; z)\|^2 \leq c \frac{d^3}{b^2} \mathbb{E}_{\max_{z \in \mathcal{Z}}} \eta_i^2 \leq \log(d) c \frac{d^3}{b^2}
\]

26
Applying the median of means estimator (see [18]) to 

\[ \lambda \]

where we used the fact that 

Then, by Markov’s inequality, 

|\[ \tau \]|

Thus, it follows that 

The following Lemma provides the guarantee for Algorithm 10.

\[ \tilde{\lambda} \]
\[ \delta \]
\[ \Phi(\lambda) = \sum_{i=1}^{d} \lambda_i \log(\lambda_i) \]

It is not hard to see that 

\[ \sup_{\lambda} \Phi(\lambda) - \min_{\lambda'} \Phi(\lambda') \leq \log(d) \]

By Theorem 6.1 in [4],

\[ \mathbb{E}g(\tilde{\lambda}) - \min_{\lambda \in \Delta} g(\lambda) \leq c \log(d) \frac{d^{3/2}}{b} \sqrt{\frac{1}{T}}. \]

Then, by Markov’s inequality,

\[ \mathbb{P}(g(\tilde{\lambda}) - \min_{\lambda \in \Delta} g(\lambda) \geq 1) \leq \mathbb{E}g(\tilde{\lambda}) - \min_{\lambda \in \Delta} g(\lambda) \]

\[ \leq c \log(d) \frac{d^{3/2}}{b} \sqrt{\frac{1}{T}} \]

\[ \leq c \log(d) \frac{d^{3/2}}{b} \sqrt{\frac{1}{T}} \]

\[ = \delta \]

by our choice of \( T \). Noting that \( \min_{\lambda \in \Delta} g(\lambda) \leq \sqrt{2} \min_{\lambda \in \Delta} g(\lambda) \) yields the result.

**Step 2: Bound the number of oracle calls.** Using Lemma [2] with \( \xi' = \frac{\xi}{2} \) and union bounding over each of the \( T \) iterations, with probability at least \( 1 - \frac{\eta}{2d} \) the number of oracle calls is at most

\[ c[d + \log(\frac{d}{b\delta}) + \log(\phi/\xi \cdot T)]T = c[d + \log(\phi/\xi) + \log(\log(d)\frac{d^3}{b^2}) \log(d) \frac{d^3}{b^2} \frac{1}{\delta^2}]. \]

The following Lemma provides the guarantee for Algorithm [10]

**Lemma 3.** When with probability at least \( 1 - \delta \), Algorithm [10] returns \( \tau \) such that \( g(\lambda) \leq (\tau + 1)^2 \leq g(\lambda) + 4 \). Furthermore, with probability at least, \( 1 - \delta \), it uses \( O(\frac{d^2}{\delta^2} \log(1/\delta) \log(\frac{d}{b\delta})) \) oracle calls.

**Proof.** Let \( \tilde{y}_s = \sup_{z \in \mathcal{Z}} \frac{(z_0 - \bar{z})^T A(\lambda)^{-1/2} \eta_s}{b + \theta_0^1 (z_0 - \bar{z})} \). By Theorem 5.8 of [2], we have that

\[ \mathbb{V}(\tilde{y}_s) \leq 4 \sup_{z \in \mathcal{Z}} \mathbb{V} \left( \frac{(z_0 - \bar{z})^T A(\lambda)^{-1/2} \eta_s}{b + \theta_0^1 (z_0 - \bar{z})} \right) \leq 8 \frac{d^2}{b^2} \]

Applying the median of means estimator (see [18]) to \( \tilde{y}_1, \ldots, \tilde{y}_T \) yields that with probability at least \( 1 - \delta \) \( \tilde{\tau} \) satisfies

\[ |\tilde{\tau} - \mathbb{E} \sup_{z \in \mathcal{Z}} \frac{(z_0 - \bar{z})^T A(\lambda)^{-1/2} \eta_s}{b + \theta_0^1 (z_0 - \bar{z})} | \leq 1/2 \]

by our choice of \( T \) and standard results for median of means estimation. Since the procedure computeMax a tolerance of \( 1/2 \), by Lemma [5] we have that \( |y_s - \tilde{y}_s| \leq 1/2 \) for all \( s = 1, \ldots, T \). Thus, it follows that \( |\tilde{\tau} - \tau| \leq 1/2 \). Thus,

\[ |\tau - \mathbb{E} \sup_{z \in \mathcal{Z}} \frac{(z_0 - \bar{z})^T A(\lambda)^{-1/2} \eta_s}{b + \theta_0^1 (z_0 - \bar{z})} | \leq 1. \]

Manipulating the above inequality yields the result.
It remains to bound the number of oracle calls. Consider  \( \bar{y}_s = \sup_{z \in Z} \left( \frac{(z_0 - z) \top A(\lambda) - 1/2}{b + \theta_0 (z_0 - z)} \right) \). By the same argument made in inequality (51), we have that with probability at least \( 1 - \frac{\delta}{T} \), \( \bar{y}_s \leq O(\frac{d^2 T}{\delta \pi}) \). Union bounding over all \( y_s, s \in [T] \), we have that with probability at least \( 1 - \delta \), \( \sup_s y_s \leq O(\frac{d^2 T \log(1/\delta)}{\delta \pi}) \). Since the procedure computeMax uses a tolerance of 1/2, by Lemma 5 we have that each call of computeMax uses at most \( O(\log(\frac{d^2}{\delta \pi})) \) calls to the linear maximization oracle, yielding the result.

D.3 Technical Lemmas

**Lemma 4.** Consider the combinatorial bandit setting. Fix \( \theta_0 \in \mathbb{R}^d \) and \( b \geq 0 \). Let \( \xi > 0 \). Then,

\[
\Pr(\exists z \neq z' \in Z : |g(\lambda \eta; z) - g(\lambda \eta; z')| \leq \frac{\xi}{\sqrt{2d}}) \leq \frac{\xi}{2d}.
\]

**Proof.** Let \( m = |Z| \). Fix \( z \neq z' \in Z \). Fix \( z_0 \in Z \), \( k \in \mathbb{N} \), and \( \theta_0 \in \mathbb{R}^d \). For the sake of brevity, define \( h(z) := g(\lambda \eta; z) \). Note that \( |h(z) - h(z')| \) is a truncated normal distribution. Now, lower bound its variance.

\[
\text{Var}(h(z) - h(z')) = \left\| \frac{(z_0 - z)^\top A(\lambda)^{-1/2}}{2 - k B + \theta_0 (z_0 - z)} - \frac{(z_0 - z')^\top A(\lambda)^{-1/2}}{b + \theta_0 (z_0 - z)} \right\|_2^2 \\
\geq \left\| \frac{(z_0 - z)^\top A(\lambda)^{-1/2}}{2 - k B + \theta_0 (z_0 - z)} - \frac{(z_0 - z')^\top A(\lambda)^{-1/2}}{b + \theta_0 (z_0 - z)} \right\|_\infty
\geq \min_{v \in Z} \left\| \frac{1}{b + \theta_0 (z_0 - v)} \right\|_\infty
\geq \frac{1}{\varphi^2}
\]

where we used the fact that for every \( \|z - z'\|_\infty \geq 1 \) for combinatorial bandits and and the definition of \( \phi \).

Then, using the cdf of the half normal, we have that

\[
\Pr(|h(z) - h(z')| \leq \frac{\xi}{\sqrt{2d}}) = \int_0^{\frac{\xi}{\sqrt{2d}}} \frac{1}{\sqrt{\pi}} \exp\left( - \frac{y^2}{2 \sqrt{2d}} \right) dy \\
\leq \int_0^{\frac{\xi}{\sqrt{2d}}} \frac{1}{\sqrt{\pi}} \exp\left( - \frac{y^2}{2 \sqrt{2d}} \right) dy \\
\leq \sqrt{2/\pi} \frac{\xi}{2d}.
\]

Thus, using a union bound, we have that

\[
\Pr(\exists z \neq z' \in Z : |h(z) - h(z')| \leq \frac{\xi}{\sqrt{2d}}) \leq \frac{|Z|}{2d} \leq \frac{\xi}{2d}.
\]

**Lemma 5.** The following two claims holds regarding Algorithm 9

1. At the end of the first while loop of Algorithm 9 \( g(\lambda \eta) \in [\text{LOW}, \text{HIGH}] \) and it takes at most \( O(\log(\frac{g(\lambda \eta)}{\varepsilon})) \) oracle calls.

2. In the second while loop of Algorithm 9 it always holds that \( g(\lambda \eta) \in [\text{LOW}, \text{HIGH}] \). Furthermore, define \( \tilde{z} = \arg \max_{z \in Z} g(\lambda \eta; z) \). Then, if \( g(\lambda \eta) - \max_{z \neq \tilde{z}} g(\lambda \eta; z) > \varepsilon \), then it terminates after \( O(\log(\frac{g(\lambda \eta)}{\varepsilon})) \) oracle calls.

28
Proof. We begin by proving the first claim. By Lemma 6, if HIGH < g(λ; η), then g(λ; η; HIGH) > 0 and HIGH keeps increasing. At some point, we have HIGH > g(λ; η), which by Lemma 6 implies that g(λ; η; HIGH) < 0 and the while loop terminates. Notice that since z₀ ∈ Z, g(λ; η) ≥ 0 = LOW. Furthermore, since HIGH doubles at each round the first while loop takes at most O(log(g(λ; η))) oracle calls. This completes the proof of the first claim.

Next, we prove the second claim regarding the second while loop. At the beginning of the second while loop, g(λ; η) ∈ [LOW, HIGH]. It is a straightforward consequence of Lemma 6 that at the end of the if else statement in the second while loop it holds that g(λ; η) ∈ [LOW, HIGH]. In the last line of the while loop where

\[ \text{LOW} \leftarrow g(\lambda; \eta; z') \text{ for some } z' \in \arg \max \ g(\lambda; \eta; \text{LOW}) \]

it follows from Lemma 7 that g(λ; η; LOW) ≥ 0. Then, by Lemma 6, it follows that g(λ; η) ≥ LOW. Thus, the claim that g(λ; η) ∈ [LOW, HIGH] during the second while loop holds.

Finally, we bound the number of oracle calls. Assume g(λ; η) − maxₖ∈Z g(λ; η; z) > ε where z = arg maxₖ∈Z g(λ; η; z). Since at the end of the first while loop HIGH ≤ 2g(λ; η) and the second while loop performs binary search, we have that after O(log(\frac{g(λ, η)}{ε})) oracle calls,

\[ g(λ; η) ≥ \text{LOW} > g(λ; η) − \epsilon. \]

Let y ∈ arg maxₖ∈Z g(λ; η; LOW; z); we claim that y = arg maxₖ∈Z g(λ; η; z). By Lemma 7 we have that g(λ; η; LOW; y) ≥ 0. Rearranging, we obtain

\[ \frac{(z₀ − y)ᵀ A(λ)^{-1/2} η}{b + θ₀ᵀ (z₀ − y)} ≥ \text{LOW} > g(λ; η) − \epsilon > \maxₖ∈Z g(λ; η; z), \]

which implies that

\[ y = \arg \max_z g(λ; η; z) = \arg \max_z \frac{(z₀ − z)ᵀ A(λ)^{-1/2} η}{b + θ₀ᵀ (z₀ − z)}, \]

proving the claim.

Thus, inspection of the algorithm shows that it suffices to show that g(λ; η; g(λ; η))) = 0, but this follows directly from Lemma 6.

\[ \square \]

Lemma 6. If g(λ; η; r) < 0, then r > g(λ; η) and if g(λ; η; r) > 0, then r < g(λ; η).

Proof. Suppose g(λ; η; r) < 0. Then, by definition,

\[ \max_{z ∈ Z} zᵀ (A^{−1/2} (λ) η + rθ₀) − r(b + θ₀ᵀ z₀) − z₀ᵀ A(λ)^{-1/2} η < 0. \]

Rearranging, we have that for all z ∈ Z,

\[ \frac{(z₀ − z)ᵀ A(λ)^{-1/2} η}{b + θ₀ᵀ (z₀ − z)} < r, \]

thus proving the first claim. Next, suppose g(λ; η; r) > 0. Then, rearranging as above, there exists a z ∈ Z such that

\[ \frac{(z₀ − z)ᵀ A(λ)^{-1/2} η}{b + θ₀ᵀ (z₀ − z)} > r, \]

proving the second claim.

\[ \square \]

Lemma 7. If \( \max_{z ∈ Z} g(λ; η; z; L₁) ≥ 0 \), then letting \( L₂ = g(λ; η; z') \) for some \( z' ∈ \arg \max_{z ∈ Z} g(λ; η; z; L₁) \), we have that \( L₂ ≥ L₁ \) and \( g(λ; η; L₂) ≥ 0 \). Furthermore, \( g(λ; η; \text{LOW}) ≥ 0 \) throughout the execution of Algorithm 6.

\[ \square \]
Towards a contradiction, suppose 
\[ 0 = g(\lambda; \eta; z'; L_1) = (z')^\top (A^{-1/2}(\lambda)\eta + L_1\theta_0) - L_1(b + \theta_0^\top z_0) - z_0^\top A(\lambda)^{-1/2}\eta \geq 0. \]

Rearranging, we have that
\[ L_2 := \frac{(z_0 - z')^\top A(\lambda)^{-1/2}\eta}{b + \theta_0^\top (z_0 - z')} \geq L_1, \]
proving the first claim. Furthermore, rearranging the equality
\[ \frac{(z_0 - z')^\top A(\lambda)^{-1/2}\eta}{b + \theta_0^\top (z_0 - z')} = L_2 \]
yields \[ 0 = g(\lambda; \eta; z'; L_2) \leq \max_{z \in \mathcal{Z}} g(\lambda; \eta; z; L_2), \] yielding the second inequality.

Finally, \[ g(\lambda; \eta; 1_{\mathrm{low}}) \geq 0 \] follows inductively. In the base case, \( 1_{\mathrm{low}} = 0 \) and we observe that for \[ 0 = g(\lambda; \eta; z_0; 0) \leq \max_{z \in \mathcal{Z}} g(\lambda; \eta; z; 0). \] The inductive step follows by the update and the above claims.

\[ \square \]

## E Fixed Budget Upper Bound Proofs

Lemma [8] is the main step in the proof of the upper bound for the fixed budget algorithm.

**Lemma 8.** Suppose \( T \geq c R \max([\rho^* + \gamma^*], d) \). If \( z_* \in \mathcal{Z}_k \), then \( z_* \) is eliminated in round \( k \) with probability at most
\[ 2 \exp\left(\frac{-T}{c' [\rho^* + \gamma^*]}\right). \]

**Proof.** Let \( N = \lfloor T/R \rfloor \). Let \( \mathcal{X} = \{x_1, \ldots, x_m\} \). Let \( \lambda_k \) denote the design chosen by the algorithm in round \( k \). Let \( x_{I_1}, \ldots, x_{I_m} \) denote the measurement vectors selected in round \( k \) and define \( \lambda \in \Delta \) by \( \bar{\lambda}_i = \frac{1}{N} \sum_{s=1}^N 1\{I_s = i\} \). Let \( \xi > 0 \) (a constant to be chosen later). Define
\[ \Delta = \arg\min \Delta' \]
\[ \text{s.t.} \quad \frac{\|z - z'\|^2}{(\Delta')^2} \leq \xi [\rho^* + \gamma^*]. \]

Define the event
\[ \mathcal{E} = \{ \sup_{z, z' \in \mathcal{Z}_k} \frac{|(z - z')^\top (\hat{\theta}_k - \theta)|}{\Delta} \leq \sqrt{\frac{\mathbb{E}[\sup_{z, z' \in \mathcal{Z}_k} \frac{(z - z')^\top A(\lambda)^{-1/2}\eta}{\Delta}\frac{\Delta}{T/R}]}{\mathbb{E}[\Delta^2]}} + \frac{1}{2} \}. \]

By Theorem 5.8 in [2] with probability at least
\[ \mathbb{P}(\mathcal{E}^c) \leq 2 \exp\left(\frac{-[T/R]}{8 \sup_{z, z' \in \mathcal{Z}_k} \|z - z'\|^2 A(\lambda)^{-1/2}\eta} \right) \leq 2 \exp\left(\frac{-[T/R]}{8 \xi [\rho^* + \gamma^*]} \right) \]
where we used the definition of \( \Delta \). Suppose \( \mathcal{E} \) occurs for the remainder of the proof.

Define
\[ \mathcal{Z}_{k, \text{wrong}} = \{z \in \mathcal{Z}_k : \hat{\theta}_k^\top (z_* - z) < 0\}. \]

Towards a contradiction, suppose \( z_* \) is eliminated at round \( k \). Then, by definition of the algorithm,
\[ \gamma(\mathcal{Z}_{k, \text{wrong}} \cup \{z_*\}) \geq \frac{\gamma(\mathcal{Z}_k)}{2} = \frac{1}{2} \mathbb{E} \sup_{z, z' \in \mathcal{Z}_k} (z - z')^\top A(\lambda_k)^{-1/2}\eta. \]
Define $z_0 = \arg \max_{z \in Z_{k,wrong}} \Delta z$. Then,

$$
\frac{1}{2(1 + \epsilon)} \mathbb{E} \sup_{z, z' \in Z_k} (z - z')^\top A(\hat{\lambda})^{-1/2} \eta \leq \frac{1}{2} \mathbb{E} \sup_{z, z' \in Z_k} (z - z')^\top A(\lambda_k)^{-1/2} \eta
$$

(52)

$$
\leq \min_{\lambda} \mathbb{E} \sup_{z \in Z_{k,wrong} \cup \{z_*\}} (z - z')^\top A(\lambda)^{-1/2} \eta
$$

$$
\leq c \min_{\lambda} \mathbb{E} \sup_{z \in Z_{k,wrong} \cup \{z_*\}} (z_* - z)^\top A(\lambda)^{-1/2} \eta
$$

$$
\leq c' \min_{\lambda} \mathbb{E} \sup_{z \in Z_{k,wrong}} (z_* - z)^\top A(\lambda)^{-1/2} \eta
$$

$$
+ \|z_* - z_0\|_{A(\lambda)^{-1}}^2
$$

(53)

where line (52) follows by the guarantees of the rounding procedure and Lemma 10, and line (53) follows by Lemma 15. Thus,

$$
\mathbb{E} \left[ \sup_{z, z' \in Z_k} (z - z')^\top A(\hat{\lambda})^{-1/2} \eta \right]^2 \leq c \min_{\lambda} \mathbb{E} \left[ \sup_{z \in Z_{k,wrong}} (z_* - z)^\top A(\lambda)^{-1/2} \eta \right]^2
$$

$$
+ \frac{\|z_* - z_0\|_{A(\lambda)^{-1}}^2}{\Delta_{z_0}^2}
$$

$$
\leq c' [\gamma^* + \rho^*]
$$

(54)

where line (54) follows by Lemma 12. Furthermore, we have that

$$
\mathbb{E} \left[ \sup_{z, z' \in Z_k} (z - z')^\top A(\hat{\lambda})^{-1/2} \eta \right]^2 \geq c \sup_{z, z' \in Z_k} \|z - z'\|_{A(\lambda)^{-1}}^2 \frac{\Delta_{z_0}^2}{\Delta_{z_0}^2}
$$

(55)

by Lemma 11. Combining inequalities (54) and (55), we have that there exists a universal constant $\xi > 0$ such that $\Delta_{z_0} \geq \Delta$ (choose this $\xi$).

Then,

$$
|z_* - z_0| (\hat{\theta}_k - \theta) \leq \sup_{z, z' \in Z_k} |(z - z')^\top (\hat{\theta}_k - \theta)|
$$

$$
\leq \Delta_{z_0} \sqrt{\mathbb{E} \left[ \sup_{z, z' \in Z_k} (z - z')^\top A(\hat{\lambda})^{-1/2} \eta \right]^2} \frac{T/R}{T/R} + \Delta
$$

(56)

$$
\leq \Delta_{z_0} c' \sqrt{\gamma^* + \rho^*} \frac{T/R}{T/R} + \Delta
$$

(57)

$$
\leq \frac{\Delta_{z_0}}{2} + \frac{\Delta}{2}
$$

(58)

where line (56) follows by the event $\mathcal{E}$, line (57) follows by (54), and line (58) follows since $T \geq c R [\rho^* + \gamma^*]$ for an appropriately large universal constant $c > 0$. Rearranging the above inequality implies that

$$
(z_* - z_0)^\top \hat{\theta}_k > 0
$$

and thus $z_0 \notin Z_{k,wrong}$, a contradiction. Therefore, on $\mathcal{E}$, $z_*$ is not eliminated.

\[ \square \]

**Proof of Theorem 6** Define the event

$$
E_k = \{ z_* \text{ is not eliminated in round } k \},
$$

$$
E = \cap_{k=0}^R E_k.
$$

31
Then, by the law of total probability, Lemma 8 and the definition of $R = \lceil \log(\gamma(Z)) \rceil$,

$$\mathbb{P}(E^c) \leq \mathbb{P}(E_1^c) + \sum_{k=2}^{R} \mathbb{P}(E_k^c \cap \bigcap_{i=1}^{k-1} E_i) \leq \lceil \log(\gamma(Z)) \rceil \exp\left(\frac{-T}{32R[\rho^* + \gamma^*]}\right).$$

Assume the event $E$ holds. Recall the assumption that $\gamma(\{z, z^*\}) \geq 1$ for all $z \in Z \setminus \{z^*\}$. Since by the definition of the algorithm and $R$,

$$\gamma(Z_R) \leq \frac{\gamma(Z)}{2^R} \leq 1$$

the algorithm must terminate in one of the $\lceil \log(\gamma(Z)) \rceil$ rounds and return $z^*$, completing the proof.

\[\square\]

### F $\gamma^*$ Results

In this Section, we prove various results related to $\gamma^*$.

**Proof of Proposition 2:** Define $\theta = e_1$ and $z^* = e_1$. Let

$$Z = \{v \in \mathbb{R}^d : \|v\|_2 = 1, v_1 = 0\} \cup \{z^*\}.$$

Let $\mathcal{X} = \{e_1, \ldots, e_d\}$. Then, for any $\lambda \in \Delta$,

$$E_{\eta \sim N(0, I)}[\max_{z \in Z \setminus \{z^*\}}(z^* - z)^\top A(\lambda)^{-1/2} A(\lambda)^{-1/2} A(\lambda)^{-1/2} \eta] = E_{\eta \sim N(0, I)}[\max_{z \in Z \setminus \{z^*\}}(z^* - z)^\top A(\lambda)^{-1/2} \eta] \geq (d - 1)E_{\eta \sim N(0, I)}[\max_{z \in Z \setminus \{z^*\}} z^\top \eta] \geq (d - 1)(d + c)$$

where $c$ is a universal constant where the second to last inequality follows by symmetry and the last inequality follows by example 7.5.7 in [32]. On the other hand,

$$\rho^* = \inf_{\lambda} \max_{z \in Z \setminus \{z^*\}} \frac{\|z^* - z\|^2}{\theta^\top (z^* - z)^2} \leq 4d$$

where we took $\lambda = (1/d, \ldots, 1/d)^\top$. Thus, there exists an instance where $\rho^* \leq dc$ and $\gamma^* \geq c'd^2$, proving the result.

Top-K is an example of a problem instance where $\gamma^* \leq c \log(d)\rho^*$ (see Proposition 6).

\[\square\]

**Proposition 6.** Consider an instance of Top-K. Assume wlog $\theta_1 \geq \theta_2 \geq \ldots \geq \theta_d$.

$$\gamma^* \leq c \log(d) \left[ \sum_{i \leq k} (\theta_i - \theta_{k+1})^{-2} + \sum_{i > k} (\theta_k - \theta_i)^{-2} \right].$$

**Proof of Proposition 6:** Define

$$\Delta_i = \begin{cases} 
\theta_i - \theta_{k+1} & \text{if } i \leq k \\
\theta_k - \theta_i & \text{if } i > k 
\end{cases}$$
Set $\lambda_i = \frac{\Delta_{i}^2}{\sum_{j \in [d]} \Delta_{j}^2}$. Note that $Z = \{z \in [d] : |z| = k\}$. Then,

$$
\gamma^* \leq \mathbb{E}_{\eta \sim N(0, I)} \left( \max_{z \subset Z} \sum_{i \in \{k\} \setminus z} \frac{\lambda_i}{\sum_{j \in \{k\} \setminus z} \theta_1 - \sum_{j \in \{k\} \setminus z} \theta_j} \right) \left( \sum_{i \in \{k\} \setminus z} \theta_1 - \sum_{j \in \{k\} \setminus z} \theta_j \right)^2
$$

$$
= \sum_{j \in [n]} \Delta_j \mathbb{E}_{\eta \sim N(0, I)} \left( \max_{z \subset Z} \sum_{i \in \{k\} \setminus z} (\theta_1 - \theta_k) \eta_i + \sum_{j \in \{k\} \setminus z} (\theta_{k+1} - \theta_j) \eta_j \right)^2
$$

$$
= \sum_{j \in [n]} \Delta_j \mathbb{E}_{\eta \sim N(0, I)} \left( \max_{z \subset Z} v_z^T \eta + w_z^T \eta \right)^2
$$

where we defined the vectors

$$(v_z)_i = \frac{\theta_i - \theta_k}{\sum_{j \in \{k\} \setminus z} \theta_j} \mathbb{1}_{\{i \in \{k\} \setminus z\}}$$

$$(w_z)_i = \frac{\theta_{k+1} - \theta_i}{\sum_{j \in \{k\} \setminus z} \theta_j} \mathbb{1}_{\{i \in [z \setminus [k]\}}$$

Note that

$$
\|v_z\|_1 = \frac{\sum_{i \in \{k\} \setminus z} (\theta_i - \theta_k)}{\sum_{i \in \{k\} \setminus z} \theta_j} \sum_{i \in \{k\} \setminus z} \theta_j = \frac{\sum_{i \in \{k\} \setminus z} \theta_1 - \sum_{j \in \{k\} \setminus z} \theta_j}{\sum_{i \in \{k\} \setminus z} \theta_j} = 1
$$

where we used the fact that $|[k] \setminus z| = |z \setminus [k]|$ and the assumption $\theta_1 \geq \theta_2 \geq \ldots \geq \theta_n$. Similarly,

$$
\|w_z\|_1 \leq 1.
$$

Thus,

$$
\gamma^* \leq \sum_{j \in [d]} \Delta_j \mathbb{E}_{\eta \sim N(0, I)} \left( \max_{z \subset Z} v_z^T \eta + w_z^T \eta \right)^2
$$

$$
\leq \sum_{j \in [d]} \Delta_j \mathbb{E}_{\eta \sim N(0, I)} \left( \max_{v: \|v\|_1 \leq 1} v^T \eta + \max_{w: \|w\|_1 \leq 1} w^T \eta \right)^2
$$

$$
\leq c \log(d) \sum_{j \in [d]} \Delta_j^2
$$

where in the final inequality we used Example 7.5.9 of [32].

Proof of Proposition 7

$$
\gamma^* = \inf_{\lambda} \mathbb{E}_{\eta \sim N(0, I)} \left( \max_{z \subset Z} \left| \frac{A(\lambda)^{-1/2}(z^* - z)}{\theta^{\top}(z^* - z)} \right| \right)^2
$$

$$
\leq \inf_{\lambda} c \log(|Z|) \text{diam} \left\{ \frac{A(\lambda)^{-1/2}(z^* - z)}{\theta^{\top}(z^* - z)} : z \in Z \setminus \{z^*\} \right\}^2
$$

$$
\leq c' \log(|Z|) \inf_{\lambda} \max_{z \subset Z \setminus z^*} \left| \frac{A(\lambda)^{-1/2}(z^* - z)}{\theta^{\top}(z^* - z)} \right|^2
$$

$$
= c' \log(|Z|) \inf_{\lambda} \max_{z \subset Z \setminus z^*} \frac{\|z^* - z\|^2}{\theta^{\top}(z^* - z)^2} \|
$$

$$
= c' \log(|Z|) \rho^*
$$

where we used exercise 7.5.10 of [32] in line (59). On the other hand, Proposition 7.5.2 of [32] implies that

$$
\gamma^* \leq d \rho^*
$$
Now, we prove the lower bound. There exists $\xi > 0$, $z_1 \in \mathcal{Z}$, and $\lambda_1 \in \Delta$ such that
\[
\xi + \inf_{z \neq z_1} \inf_{\lambda \in \Delta} \frac{\|z - z_1\|^2}{\Delta^2} \geq \frac{\|z - z_1\|_{A(\lambda)^{-1}}}{\Delta_{z_1}^2}.
\]

Let $\lambda_2 \in \Delta$ attain $\gamma^*$. Let $\lambda = \frac{1}{2}(\lambda_1 + \lambda_2)$ Then,
\[
\min_{\lambda \in \Delta} \max_{z \neq z_1} \frac{\|z - z_1\|^2}{\Delta_{z_1}^2} \geq 4\left(\frac{\pi}{2}\right)^{\eta N(0,1)} \max_{z \in \mathcal{Z} \setminus \{z_1\}} \left(\frac{z - z_1}{\Delta_{z_1}}\right)^\top A(\lambda_2)^{-1/2} \eta \|z - z_1\|^2_{A(\lambda_2)^{-1}} + \xi.
\]

where line (60) follows by Lemma 11 and line (61) follows by the Sudakov-Fernique inequality (Theorem 7.2.11 of [32]) since $A(\lambda)^{-1} \leq 2A(\lambda_2)^{-1}$. Since $\xi > 0$ is arbitrary, sending $\xi \to 0$ yields the lower bound.

\[ \square \]

**Proof of Proposition 4** Recall the definition $B(z, r) = \{z' \in \mathcal{Z} : \|z - z'\|_2 = r\}$. Let $\lambda_i = \frac{\bar{z}_i}{\bar{z}_i^r}$. Further, define
\[
A_i = \{j \in [d] : \log(\|d|B(z_*, j)|) \in [2^{i-1}, 2^i]\}.
\]
Let $v > 0$ a constant to be chosen later. Then,

$$\sqrt{v} \leq E_{\eta \sim N(0,I)}[\max_{z \in \mathbb{Z} \setminus z_0} \frac{\sum_{i \in z, \Delta z} \frac{1}{\sqrt{\Delta z}} \eta_i}{\Delta z}]$$

$$= \frac{1}{v} E_{\eta \sim N(0,I)}[\max_{z \in \mathbb{Z} \setminus z_0} \frac{\sum_{i \in z, \Delta z} \frac{1}{\sqrt{\Delta z}} \eta_i}{\Delta z}]$$

$$= \frac{\sqrt{v^2}}{v} E_{\eta \sim N(0,I)}[\log(\max_{z \in \mathbb{Z} \setminus z_0} \exp(v \frac{\sum_{i \in z, \Delta z} \frac{1}{\sqrt{\Delta z}} \eta_i}{\Delta z}))]$$

$$\leq \frac{\sqrt{v^2}}{v} \log(E_{\eta \sim N(0,I)}[\max_{z \in \mathbb{Z} \setminus z_0} \exp(v \frac{\sum_{i \in z, \Delta z} \frac{1}{\sqrt{\Delta z}} \eta_i}{\Delta z}))])$$

$$\leq \frac{\sqrt{v^2}}{v} \log(4 \log(d) \max_{i \in [4 \log(d)]} E_{\eta} [\max_{z \neq z_0, |z \Delta z| \in A_i} \exp(v \frac{\sum_{i \in z, \Delta z} \frac{1}{\sqrt{\Delta z}} \eta_i}{\Delta z}))])$$

(62)

where line [62] follows by Jensen’s inequality, where line [63] follows by the definition of $A_i$, and line [64] follows since the max is upper bounded by the sum.

Notice that line [65] contains the moment generating function of a Gaussian random variable. We upper bound its variance as follows. Suppose $|z_0 \Delta z| \in A_i$. Then,

$$\mathbb{V}(\frac{\sum_{i \in z, \Delta z} \frac{1}{\sqrt{\Delta z}} \eta_i}{\Delta z})$$

$$= \mathbb{V}(\frac{\sum_{i \in z, \Delta z} \min_{z' \in z \Delta z'} \frac{\Delta z'}{\Delta z} \eta_i}{\Delta z})$$

$$= \sum_{i \in z, \Delta z} \frac{\Delta z'}{\Delta z} \mathbb{V}(\frac{\sum_{i \in z, \Delta z'} \min_{z' \in z \Delta z'} \frac{\Delta z'}{\Delta z} \eta_i}{\Delta z})$$

$$= \frac{1}{|z_0 \Delta z| \log(d|B(z_0, |z_0 \Delta z|)|)} \sum_{i \in z, \Delta z} \frac{\Delta z'}{\Delta z} \mathbb{V}(\frac{\sum_{i \in z, \Delta z'} \min_{z' \in z \Delta z'} \frac{\Delta z'}{\Delta z} \eta_i}{\Delta z})$$

$$\leq \frac{1}{|z_0 \Delta z| \log(d|B(z_0, |z_0 \Delta z|)|)}$$

$$\leq \frac{1}{2v^2}$$

(68)
where line (66) follows by the definition of $\varphi$, line (67) follows since $\eta \sim N(0, I)$, and line (68) follows since $|z_s \Delta z| \in A_i$. Now, continuing and using this upper bound on the variance, we have

$$\sqrt{\varphi^*} \log(4 \log(d)) \max_{i \in [4 \log(d)]} \mathbb{E}_{\eta} \left[ \max_{z \neq z_s, |z \Delta z| \in A_i} \exp \left( c \sum_{i \in z_s \Delta z} \frac{1}{\sqrt{\varphi^*}} \right) \right]$$

$$\leq \sqrt{\varphi^*} \log(4 \log(d)) \max_{i \in [4 \log(d)]} \left| \bigcup_{j \in A_i} B(z_s, j) \right| \exp \left( v^2 \frac{c}{2^{i+1}} \right)$$

$$= \max_{i \in [4 \log(d)]} \sqrt{\varphi^*} \log(4 \log(d)) / \sqrt{2^{i+1}} \left( \bigcup_{j \in A_i} B(z_s, j) \right)$$

$$= \max_{i \in [4 \log(d)]} \sqrt{\varphi^*} \log(4 \log(d)) \log \left( \max_{j \in A_i} \log(\left| A_i \right| B(z_s, j)) \right) \frac{c}{2^{i+1}}$$

$$\leq \max_{i \in [4 \log(d)]} \sqrt{\varphi^*} \log(4 \log(d)) \max_{j \in A_i} \log \left( \max_{j \in A_i} \log(\left| A_i \right| B(z_s, j)) \right) \frac{c}{2^{i+1}}$$

$$\leq c \sqrt{\varphi^*} \log(4 \log(d))$$

where (69) follows by Lemma [16] and $\{ z \in \mathcal{Z} : z \neq z_s, |z \Delta z| \in A_i \} \subseteq \bigcup_{j \in A_i} B(z_s, j)$, line (70) follows by maximizing the constant $v$, (71) follows since $|A_i| \leq d$, and line (72) follows by definition of $A_i$.

**Proof of Proposition**

Define the allocation

$$\lambda_i \propto \tilde{\Delta}_i^{-2}$$

where

$$\tilde{\Delta}_i = \begin{cases} \theta^T z_s - \max_{z \neq z_s, |z \Delta z| \in A_i} \theta^T z & i \not\in z_s \\ \theta^T z_s - \max_{z \neq z_s, |z \Delta z| \in A_i} \theta^T z & i \in z_s \end{cases}$$

Then,

$$\gamma^* \leq \mathbb{E}_{\eta \sim N(0, I)} \left[ \max_{z \in \mathcal{Z}} \sum_{i \in z_s \Delta z} \frac{1}{\sqrt{\varphi^*}} \right]$$

$$= \sum_{i=1}^d \tilde{\Delta}_i^{-2} \mathbb{E}_{\eta \sim N(0, I)} \left[ \max_{z \in \mathcal{Z}} \sum_{i \in z_s \Delta z} \frac{1}{\sqrt{\varphi^*}} \right]$$

$$= \sum_{i=1}^d \tilde{\Delta}_i^{-2} \mathbb{E}_{\eta \sim N(0, I)} \left[ \max_{z \in \mathcal{Z}} \sum_{i \in z_s \Delta z} \frac{1}{\sqrt{\varphi^*}} \right]$$

$$= \sum_{i=1}^d \tilde{\Delta}_i^{-2} \mathbb{E}_{\eta \sim N(0, I)} \left[ \max_{z \in \mathcal{Z}} \sum_{i \in z_s \Delta z} \frac{1}{\sqrt{\varphi^*}} \right]$$

where we defined the vectors

$$(v_z)_i = \frac{\theta^T z_s - \max_{z \neq z_s, |z \Delta z| \in A_i} \theta^T z}{\sum_{j \in z_s \Delta z} \theta_j - \sum_{j \in z_s \Delta z} \theta_j} 1 \{i \in z_s \}$$

$$(w_z)_i = \frac{\theta^T z_s - \max_{z \neq z_s, |z \Delta z| \in A_i} \theta^T z}{\sum_{j \in z_s \Delta z} \theta_j - \sum_{j \in z_s \Delta z} \theta_j} 1 \{i \in z \}$$

It remains to bound the expected suprema. Suppose wlog $z_s = \{1, \ldots, r \}$. By Lemma [9] there exists a bijection $\sigma : z_s \rightarrow z$ such that for every $i \in z_s$, $z^{(i)} := (z_s \setminus \{i\}) \cup \{\sigma(i)\} \in \mathcal{Z}$. Note that

$$\theta^T z_s - \max_{z \neq z_s, |z \Delta z| \in A_i} \theta^T z \leq \theta^T (z_s - z^{(i)}) = \theta_i - \theta_{\sigma(i)}.$$
Therefore,

\[ \|v_z\|_1 = \sum_{i \in z \setminus z} \left( \theta^T z_{-i} - \max_{z, i \in z} \theta^T z \right) \leq \sum_{i \in z} \left( \theta_i - \theta_{\sigma(i)} \right) \leq 1. \]

A similar argument shows that \( \|w_z\|_1 \leq 1 \). Thus,

\[ \mathbb{E}_{\eta \sim \mathcal{N}(0,I)} \left[ \max_{z \subseteq Z, \|v\|_1 \leq 1} v^T \eta + w^T \eta \right] \leq \mathbb{E}_{\eta \sim \mathcal{N}(0,I)} \left[ \max_{w : \|w\|_1 \leq 1} w^T \eta \right] \leq c \log(d) \]

where in the final inequality we used Example 7.5.9 of [32].

The following Lemma appears as Corollary 3 in [3].

**Lemma 9.** Given two bases \( B_1 \) and \( B_2 \) of a matroid \( \mathcal{M} = (E, I) \), there exists a bijection \( \sigma : B_1 \rightarrow B_2 \) such that \( (B_2 \setminus \sigma(e)) \cup e \in I \) for all \( e \in I \).

### G Additional Lower Bounds

In this section, we show that in several common situations \( \Omega(d) \) samples are required. The following Theorem applies to combinatorial bandits.

**Theorem 8.** Let \( \delta \in (0, 1/4) \). Consider the combinatorial bandit setting. Fix \( \Theta \subseteq [d] \) such that there is a unique best arm. Suppose \( \Theta \) satisfies the following property: for all \( i \in [d] \)

\( \left( \theta + e_i \cdot \min_{i \in z \setminus z} \Delta_z \right) \in \Theta \) or \( \left( \theta - e_i \cdot \min_{i \in z \setminus z} \Delta_z \right) \in \Theta \) is true.

If an algorithm \( \mathcal{A} \) is \( \delta \)-pac wrt \( (\mathcal{X}, Z, \Theta) \), then

\[ \mathbb{E}_{\theta} \left[ \sum_{i=1}^d |T_i| \right] \geq \frac{d}{2}, \]

where \( T_i \) denote the number of times that \( \mathcal{A} \) pulls \( e_i \).

The intuition behind the argument in Theorem 8 is that if \( \Omega(d) \) directions are not explored with constant probability, then there is some \( \theta \) that the algorithm has no information about with constant probability and this could alter \( \arg \max_{z \subseteq Z} \theta^T z \).

**Remark 1.** Note that if \( \Theta = \mathbb{R}^d \), then \( \Theta \) satisfies the condition in the above Theorem.

**Proof of Theorem 8** Without loss of generality, suppose \( 1 = \arg \max_i \theta^T z_i \). Towards a contradiction, suppose there is some arm \( i \) such that \( \mathbb{E}_\theta[T_i] \leq \frac{1}{2} \). Let \( z_j \) such that \( i \in z_j \setminus z_1 \) and suppose that \( i \in z_j \setminus z_1 \) (the other case is similar). Define

\[ \tilde{\theta}_k = \begin{cases} \theta_k & \text{if } k \neq i \\ \theta_i + 2\theta^T (z_1 - z_j) & \text{if } k = i. \end{cases} \]

Note that \( (z_1 - z_j)^\top \tilde{\theta} < 0 \). Observe that

\[ \frac{1}{2} \geq \mathbb{E}_\theta[T_i] \geq \mathbb{P}_\theta(T_i > 0). \]

Define the event \( A = \{T_i = 0\} \cap \{I = 1\} \), where \( I \) denotes the index of the set output by \( \mathcal{A} \) as its answer for the best set. Note that

\[ \mathbb{P}_\theta(A^c) \leq \mathbb{P}_\theta(T_i > 0) + \mathbb{P}_\theta(I \neq 1) \leq \frac{1}{2} + \delta \leq \frac{3}{4}. \]
so that $\mathbb{P}_\theta(A) \geq \frac{1}{2}$.

Define

$$\tilde{K}_{i, T_i} = \sum_{s=1}^{T_i} \log (f_\theta(Z_s) / f_\tilde{\theta}(Z_s))$$

where $Z_s$ is the observation on the $s$th pull of $e_i$, $f_\theta$ denotes the density of the distribution associated with $e_i \in \mathcal{X}$ under $\theta$, and $f_\tilde{\theta}$ denotes the density of the distribution associated with $e_i \in \mathcal{X}$ under $\tilde{\theta}$. Then, by the change of measure identity (Lemma 18) from [25],

$$\mathbb{P}_\tilde{\theta}(I = 1) \geq \mathbb{P}_\theta(A)$$

$$= \mathbb{E}_\theta[\mathbb{1}\{A\} \exp(-T_i \tilde{K}_{i, T_i})]$$

$$\geq \frac{1}{4},$$

where we used the fact that the only difference between problem $\theta$ and problem $\tilde{\theta}$ is the $i$th arm and on the event $A$, $T_i = 0$. Thus, on problem instance $(Z, \tilde{\theta})$, $A$ gives the incorrect answer with probability $1/4 > \delta$, which is a contradiction.

The following Theorem gives a lower bound for best arm identification in linear bandits.

**Theorem 9.** Let $\delta \in (0, 1)$. Let $\mathcal{X} \subset \mathbb{R}^d$, such that $\|x_i\|_2 \leq 1$ for all $i \in [\mathcal{X}]$, $Z = \mathcal{X}$, and $\Theta = \mathbb{R}^d$. Fix $\theta \in \Theta$ such that there is a unique best arm and let $x_1 = \arg\max_{x} \theta^\top x$. If an algorithm $A$ is $\delta$-pac wrt $(\mathcal{X}, \Theta)$ and $d \geq 3$, then

$$\mathbb{E}_\theta[\sum_{x \in \mathcal{X}} T_x] \geq c \log\left(\frac{1}{\delta}\right) \min_{i \neq 1} \frac{d}{\theta^\top (x_1 - x_i)^2}$$

where $T_x$ denotes the number of times that $A$ pulls $x \in \mathcal{X}$.

**Proof of Theorem 9.** By Theorem 1 of [12], we have that

$$\mathbb{E}_\theta[\sum_{x \in \mathcal{X}} T_x] \geq \log\left(\frac{1}{\delta}\right) c \rho^*$$

so it suffices to lower bound $\rho^*$.

Since

$$\rho^* = \min_{\lambda} \max_{i \neq 1} \frac{\|x_1 - x_i\|_{A(\lambda)^{-1}}^2}{\theta^\top (x_1 - x_i)^2} \geq \left[\min_{j \neq 1} \frac{cd}{\theta^\top (x_1 - x_j)^2}\right] \min_{i \neq 1} \|x_1 - x_i\|_{A(\lambda)^{-1}}^2,$$

it suffices to show that

$$\min_{\lambda} \max_{i \neq 1} \|x_1 - x_i\|_{A(\lambda)^{-1}}^2 \geq cd.$$

Let

$$\lambda^* = \arg\min_{\lambda} \max_{i \neq 1} \|x_1 - x_i\|_{A(\lambda)^{-1}}.$$ 

Then,

$$\sqrt{d} = \min_{\lambda} \max_{i \in [\mathcal{X}]} \|x_i\|_{A(\lambda)^{-1}}$$

$$\leq \min_{\lambda} \max_{i \neq 1} \|x_1 - x_i\|_{A(\lambda)^{-1}} + \|x_1\|_{A(\lambda)^{-1}}$$

$$\leq \max_{i \neq 1} \|x_1 - x_i\|_{A(\lambda^*)^{-1}} + \|x_1\|_{A(\lambda^*)^{-1}}$$

$$\leq \sqrt{2} \max_{i \neq 1} \|x_1 - x_i\|_{A(\lambda^*)^{-1}} + \sqrt{2} \|x_1\|_{A(\lambda^*)^{-1}}$$

$$= \sqrt{2} \max_{i \neq 1} \|x_1 - x_i\|_{A(\lambda^*)^{-1}} + \sqrt{2}. \quad (73)$$
The first line follows by Keifer-Wolfowitz (Theorem 21.1 in [26]). The second to last inequality follows because
\[ \frac{1}{2} A(\lambda^*) + \frac{1}{2} x_1 x_1^\top \preceq \frac{1}{2} A(\lambda^*) \]
which implies
\[ (\frac{1}{2} A(\lambda^*))^{-1} \succeq (\frac{1}{2} A(\lambda^*) + \frac{1}{2} x_1 x_1^\top)^{-1}. \]
Also, since \( x_1 \in \text{span}(x_1) \), the same fact implies that
\[ \| x_1 \| (\frac{1}{2} A(\lambda^*) + \frac{1}{2} x_1 x_1^\top)^{-1} \succeq \| x_1 \| A(\lambda)^{-1} \]
Rearranging the inequality (73), we obtain
\[ 2(\sqrt{d} - \sqrt{2}) \leq \min_{\lambda} \max_{i \neq 1} \| x_1 - x_i \| \]
and thus the result follows.

\[ \square \]

H Rounding

In this Section, we justify the application of the rounding procedure from [1]. Define
\[ S_N = \{ v \in \mathbb{N}^{[X]} : \sum_{i=1}^{[X]} v_i \leq N \} \]
\[ C_N = \{ v \in [0, N]^{[X]} : \sum_{i=1}^{[X]} v_i \leq N \} \]
The following Theorem appears in [1].

**Theorem 10.** Let \( F : S^+_d \rightarrow \mathbb{R} \) such that
- For any \( A, B \in S^+_d \), if \( A \preceq B \), then \( F(A) \geq F(B) \),
- for any \( A \in S^+_d \) and \( t \in (0, 1) \), \( F(tA) = t^{-1} F(A) \).

Let \( \epsilon \in (0, 1/6] \). Then, if \( |X| \geq N \geq 5 \frac{d}{\epsilon} \), for any \( \pi \in C_N \), there exists an algorithm that in \( \tilde{O}(|X|d^2) \) time rounds \( \pi \) to \( \kappa \in S_N \) such that
\[ F(A(\kappa)) \leq (1 + 6\epsilon) F(A(\pi)). \]
The following result shows that the optimization problem

**Lemma 10.** Fix \( V \subset \mathbb{R}^d \). Define the functions \( F, G : S^+_d \rightarrow \mathbb{R} \)
\[ F(A) = \mathbb{E}_{v \sim N(0, I)} [\max_{v \in V} v^\top A^{-1/2} \eta] \]
\[ G(A) = \max_{v \in V} v^\top A v. \]
\( F \) and \( G \) satisfy the conditions of Theorem 10.

**Proof.** It is trivial to see that \( G \) satisfies the conditions of Theorem 10. Thus, we focus on the function \( F \). Let \( A, B \in S^+_d \) such that \( A \preceq B \). Then, \( A^{-1} \succeq B^{-1} \). Fix \( v, w \in V \). Then,
\[ \mathbb{E}[(w - w)^\top A^{-1/2} \eta] = \mathbb{E}[(v - w)^\top B^{-1/2} \eta] \]
Then, by Sudakov-Fernique inequality (Theorem 7.2.11 in [32]), it follows that \( F(A) \geq F(B) \).
The second condition is trivial.
I Technical Lemmas related to $\gamma^*$

In this Section, we state and prove several useful technical lemmas.

**Lemma 11.** Let $S \subset \mathbb{Z}$. Then,

$$\mathbb{E}_{\eta \sim \text{N}(0, I)}[\max_{z, z' \in S}[A(\lambda)^{-1/2}(z - z')]^\top \eta]^2 \geq \frac{2}{\pi} \max_{z, z' \in S} \|z - z'\|^2_{A(\lambda)^{-1}}.$$

**Proof.** Fix $z_1, z_2 \in S$. Then,

$$\mathbb{E}_{\eta \sim \text{N}(0, I)}[\max_{z, z' \in S}[A(\lambda)^{-1/2}(z - z')]^\top \eta] \geq \mathbb{E}_{\eta \sim \text{N}(0, I)}[A(\lambda)^{-1/2}(z_1 - z_2)]^\top \eta] = \|z_1 - z_2\| \sqrt{\frac{2}{\pi}} \tag{74}$$

**Lemma 12.** Let $\alpha > 0$ be a constant. Then,

$$\inf_{\lambda} \mathbb{E}_{\eta \sim \text{N}(0, I)}[\max_{z \in \mathbb{Z}\setminus\{z^*\}} \frac{(z^* - z)^\top A(\lambda)^{-1/2} \eta}{\theta^\top (z^* - z)}]^2 + \max_{z \in \mathbb{Z}\setminus\{z^*\}} \frac{\|z^* - z\|^2_{A(\lambda)^{-1}} - \alpha}{\theta^\top (z^* - z)^2} \leq c[\gamma^* + \rho^* \alpha]$$

**Proof.** Let $\lambda_1$ denote the solution to $\gamma^*$ and $\lambda_2$ the solution to $\rho^*$. Define $\lambda = \frac{1}{2}(\lambda_1 + \lambda_2)$. It suffices to show that

$$\mathbb{E}_{\eta \sim \text{N}(0, I)}[\max_{z \in \mathbb{Z}\setminus\{z^*\}} \frac{(z^* - z)^\top A(\lambda)^{-1/2} \eta}{\theta^\top (z^* - z)}]^2 \leq c\mathbb{E}_{\eta \sim \text{N}(0, I)}[\max_{z \in \mathbb{Z}\setminus\{z^*\}} \frac{(z^* - z)^\top A(\lambda_1)^{-1/2} \eta}{\theta^\top (z^* - z)}]^2$$

and

$$\max_{z \in \mathbb{Z}\setminus\{z^*\}} \frac{\|z^* - z\|^2_{A(\lambda_1)^{-1}} - \alpha}{\theta^\top (z^* - z)^2} \leq c\max_{z \in \mathbb{Z}\setminus\{z^*\}} \frac{\|z^* - z\|^2_{A(\lambda_2)^{-1}}}{\theta^\top (z^* - z)^2}.$$

Note that

$$1/2 \sum_{x \in \mathcal{X}} (\lambda_{1,x} + \lambda_{2,x}) xx^\top \succeq 1/2 \sum_{x \in \mathcal{X}} \lambda_{i,x} xx^\top$$

for $i = 1, 2$. Therefore,

$$2A(\lambda_i)^{-1} \succeq A(\lambda)^{-1} \tag{74}$$

for $i = 1, 2$.

(74) immediately implies

$$\max_{z \in \mathbb{Z}\setminus\{z^*\}} \frac{\|z^* - z\|^2_{A(\lambda_1)^{-1}} - \alpha}{\theta^\top (z^* - z)^2} \leq c\max_{z \in \mathbb{Z}\setminus\{z^*\}} \frac{\|z^* - z\|^2_{A(\lambda_2)^{-1}}}{\theta^\top (z^* - z)^2}.$$

(74) implies via Sudakov-Fernique inequality (Theorem 7.2.11 in [32]) that

$$\mathbb{E}_{\eta \sim \text{N}(0, I)}[\max_{z \in \mathbb{Z}\setminus\{z^*\}} \frac{(z^* - z)^\top A(\lambda)^{-1/2} \eta}{\theta^\top (z^* - z)}]^2 \leq c\mathbb{E}_{\eta \sim \text{N}(0, I)}[\max_{z \in \mathbb{Z}\setminus\{z^*\}} \frac{(z^* - z)^\top A(\lambda_1)^{-1/2} \eta}{\theta^\top (z^* - z)}]^2.$$
Lemma 13. Let $V = \{v_1, \ldots, v_l\} \subset \mathbb{R}^d$ and suppose $0 \in V$. Let $a_i \geq 1$ for all $i$. Then,

$$
\mathbb{E}_{\eta \sim N(0, I)} \sup_{v_i \in V} v_i^\top \eta \leq \mathbb{E}_{\eta \sim N(0, I)} \sup_{v_i \in V} a_i v_i^\top \eta
$$

Proof. Fix $\eta \in \mathbb{R}^d$. Then, clearly,

$$
\sup_{v_i \in V} v_i^\top \eta \leq \sup_{v_i \in V} a_i v_i^\top \eta.
$$

Taking the expectation wrt $\eta \sim N(0, I)$ yields the result.

Lemma 14. Fix $V \subset \mathbb{R}^d$. Then,

$$
\mathbb{E}_{\eta \sim N(0, I)} \sup_{v \in V} v^\top \eta \geq 0.
$$

Proof. Fix $v_0 \in V$. Then,

$$
\mathbb{E}_{\eta \sim N(0, I)} \sup_{v \in V} v^\top \eta \geq \mathbb{E}_{\eta \sim N(0, I)} v_0^\top \eta = 0.
$$

Lemma 15. Let $V \subset \mathbb{R}^d$ and suppose $0 \in V$. Fix $v_0 \in V$. Then,

$$
\mathbb{E} \sup_{v \in V} v^\top \eta \leq 2(\|v_0\|_2 + \mathbb{E} \sup_{v \in V \setminus \{0\}} v^\top g)
$$

Proof. Fix $v_0 \in V$. Then,

$$
\mathbb{E} \sup_{v \in V} v^\top \eta \leq \mathbb{E} \sup_{v \in V \setminus \{0\}} |v^\top \eta| \leq 2(\|v_0\|_2 + \mathbb{E} \sup_{v \in V \setminus \{0\}} v^\top g)
$$

where the last inequality follows by exercise 7.6.9 of [32].

Lemma 16. Consider a sub-Gaussian random process $X_t$ indexed by $t \in T$ such that for any $\nu$ we have $\mathbb{E}[\exp(\nu X_t)] \leq \exp(\nu^2 \sigma^2_t / 2)$. Then $\mathbb{E} \sup_{t \in T} X_t \leq \sqrt{2 \sup_{t \in T} \sigma^2_t \log (|T|)}$.

Proof.

$$
\mathbb{E} \left[ \sup_{t \in T} X_t \right] = \frac{1}{\nu} \mathbb{E} \left[ \sup_{t \in T} \nu X_t \right]
$$

$$
= \frac{1}{\nu} \mathbb{E} \left[ \log \left( \sup_{t \in T} \exp(\nu X_t) \right) \right]
$$

$$
\leq \frac{1}{\nu} \log \left( \mathbb{E} \left[ \sup_{t \in T} \exp(\nu X_t) \right] \right)
$$

$$
\leq \frac{1}{\nu} \log \left( |T| \sup_{t \in T} \mathbb{E} \left[ \exp(\nu X_t) \right] \right)
$$

$$
\leq \frac{1}{\nu} \log \left( |T| \sup_{t \in T} \exp(\nu^2 \sigma^2_t / 2) \right)
$$

$$
= \frac{1}{\nu} \log (|T|) + \nu \sup_{t \in T} \sigma^2_t / 2
$$

$$
\leq \sqrt{2 \sup_{t \in T} \sigma^2_t \log (|T|)}
$$


The following result shows that after a suitable monotonic transformation, the objective function
in the optimization problems for finding a good allocation in Algorithms 1 and 2 is convex when
\(X = \{e_1, \ldots, e_d\}\), which holds in the combinatorial bandit problem. We note that Lemma 14 shows
that the gaussian width is nonnegative and thus it suffices consider the squareroot of the objective
function.

**Proposition 7.** Fix \(V \subset \mathbb{R}^d\).

\[
f(\lambda) = \mathbb{E}_{\eta \sim N(0, I)} \left[ \max_{v \in V} v^T \text{diag} \left( \frac{1}{\lambda_i^{1/2}} \right) \eta \right]
\]
is convex.

**Proof.** Fix \(\lambda, \kappa \in \mathbb{S}^{n-1}\) and \(\alpha \in [0, 1]\). By matrix convexity,

\[
\text{diag} \left( \frac{1}{\alpha \lambda_i + (1 - \alpha) \kappa_i} \right)^{1/2} \preceq \text{diag} \left( \frac{1}{\lambda_i} \right)^{1/2} + (1 - \alpha) \text{diag} \left( \frac{1}{\kappa_i} \right)^{1/2}.
\]
Furthermore, since the above matrices are diagonal,

\[
\text{diag} \left( \frac{1}{\alpha \lambda_i + (1 - \alpha) \kappa_i} \right) \preceq (\text{diag} \left( \frac{1}{\lambda_i} \right)^{1/2} + (1 - \alpha) \text{diag} \left( \frac{1}{\kappa_i} \right)^{1/2})^2.
\]
Then, by Sudakov-Fernique inequality (Theorem 7.2.11 [32]),

\[
f(\alpha \lambda + (1 - \alpha) \kappa) = \mathbb{E}_{\eta \sim N(0, I)} \left[ \max_{v \in V} v^T \eta \right] \sup_{v \in V} (\text{diag} \left( \frac{1}{\lambda_i} \right)^{1/2} + (1 - \alpha) \text{diag} \left( \frac{1}{\kappa_i} \right)^{1/2}) \eta
\]
\[
\leq \mathbb{E}_{\eta \sim N(0, I)} \left[ \max_{v \in V} v^T \eta \right] \sup_{v \in V} (\text{diag} \left( \frac{1}{\lambda_i} \right)^{1/2} + (1 - \alpha) \text{diag} \left( \frac{1}{\kappa_i} \right)^{1/2}) \eta
\]
\[
= \mathbb{E}_{\eta \sim N(0, I)} \left[ \max_{v \in V} v^T \eta \right] \text{diag} \left( \frac{1}{\lambda_i} \right)^{1/2} \eta + (1 - \alpha) \mathbb{E}_{\eta \sim N(0, I)} \left[ \max_{v \in V} v^T \eta \right] \text{diag} \left( \frac{1}{\kappa_i} \right)^{1/2} \eta
\]
\[
= \alpha f(\lambda) + (1 - \alpha) f(\kappa)
\]
\[
\square
\]

**K Comparison Results**

In this Section, we prove various results related to the sample complexities proposed in other works.
Recall the notation for the sphere \(B(z, r) = \{z' \in Z : \|z - z'\|_2 = r\}\).

**Proof of Proposition 5** Define \(\theta_1 = \ldots = \theta_k = 1/2, \theta_{k+1} = \ldots = \theta_{2k-1} = \frac{1}{2} - \frac{1}{\sqrt{k^2 + 2}}\) and \(\theta_{2k} = \ldots = \theta_d = 0\) and \(d = k^2\). Define

\[
\Delta_i = \begin{cases} 
\theta_i - \theta_{i+1} & : i \leq k \\
\theta_k - \theta_i & : i > k
\end{cases}
\]

\[
\bar{\lambda}_i = \frac{\Delta_i^{-2}}{\sum_{i=1}^{d} \Delta_i^{-2}}
\]

Note that

\[
\rho^* \leq \sum_i \Delta_i^{-2} \text{max}_{z \neq z_i} \frac{\sum_{z \neq z_i} \Delta_i^2}{\Delta_i^2}
\]
\[
\leq c \sum_i \Delta_i^{-2}
\]
\[
\leq c(k^2 + d)
\]
\[
\leq c'd.
\]
Consider arm $d$. We will show that $\varphi_d \geq ck \log(d)$. Fix $\tilde{z} = \{k + 1, k + 2, \ldots, 2k - 1, d\}$ and $z_* = [k]$. It suffices to show that
\[
\|z_* - \tilde{z}\|_1 \log(|B(z_*, |z_* \Delta \tilde{z}|)|) \geq c \log(d)k,
\]
from which the claim will follow. Note that
\[
\frac{\|z_* - \tilde{z}\|_1}{\theta^T (z_* - \tilde{z})^2} = \frac{2k}{(k - 1 \sqrt{k} + \frac{1}{2})^2} \geq c.
\]
Furthermore,
\[
\log(|B(z_*, |z_* \Delta \tilde{z}|)|) \geq \log\left(\frac{d - 2k}{k}\right)
\geq \log\left(\frac{(d - 2k)^{k}}{k!}\right)
\geq k \log\left(\frac{d - 2k}{k}\right)
\geq k \log(k - 2)
\geq \frac{1}{4} k \log(d)
\]
where in the last inequality we used $d = k^2$. Thus, the claim follows and $\varphi_d \geq ck$. A similar argument applies to arms $\{2k, \ldots, d - 1\}$ yielding the result. \qed

The following proposition shows that $\rho^*$ is lower bounded by the typical measure of hardness for top-k \cite{RY15}. It implies that the sample complexity of \cite{BBDKE08,CKT15} is off by a factor of $k$.

**Proposition 8.** Consider the top-k problem where $\theta_1 \geq \ldots \theta_k > \theta_{k+1} \geq \ldots \geq \theta_n$.

\[
\rho^* \geq \sum_{i \leq k} \frac{1}{(\theta_i - \theta_{k+1})^2} + \sum_{i > k} \frac{1}{(\theta_k - \theta_i)^2}
\]

**Proof.**

\[
\rho^* = \inf_{\lambda} \max_{z \in Z} \frac{\|z^* - z\|_{A(\lambda)^{-1}}}{\theta^T (z^* - z)^2}
\geq \inf_{\lambda} \max_{A \neq [k]} \frac{\sum_{i \in A \Delta [k]} \frac{1}{\lambda_i}}{\sum_{i \in [k]} \theta_i - \sum_{i \in A} \theta_i)^2}
\geq \min_{\lambda} \max_{i \in [k]} \frac{1}{\lambda_i} \left(\frac{1}{\lambda_{k+1}}\right), \max_{i \in [d \setminus [k]} \frac{1}{\lambda_{k+1}} \left(\frac{1}{\lambda_i}\right)^2
\geq \min_{\lambda} \max_{i \in [k]} \frac{1}{\lambda_i} \left(\frac{1}{\lambda_{k+1}}\right), \max_{i \in [d \setminus [k]} \frac{1}{\lambda_{k+1}} \left(\frac{1}{\lambda_i}\right)^2
\]

To minimize the RHS, we set it to a constant $c$. Then,

\[
\lambda_i = \begin{cases} \frac{1}{c(\theta_i - \theta_{k+1})^2} & i \leq k \\ \frac{1}{c(\theta_i - \theta_{k+1})^2} & \text{otherwise} \end{cases}
\]

Then, the result follows from the below and solving for $c$.

\[
1 = \sum_{i=1}^{k} \lambda_i = \frac{1}{c} \sum_{i \leq k} \frac{1}{(\theta_i - \theta_{k+1})^2} + \sum_{i > k} \frac{1}{(\theta_k - \theta_i)^2}.
\]

\qed

The following gives an instance where $|Z|$ is linear in the dimension $d$, but $\varphi^*$ is loose by a $\sqrt{d}$ factor.
Proposition 9. Consider the combinatorial bandit setting. There exists a problem where $|Z|$ is linear in the dimension $d$ and $\varphi^* \geq c \rho^* \sqrt{d}$.

Proof. Fix $k < d$. Define $z_1 = [k], z_2 = [k+1], z_3 = [k+3], \ldots, z_{d-k} = [d]$ and let $Z = \{z_1, \ldots, z_{d-k}\}$. Note $|Z| \leq d$ and thus satisfies the hypothesis. Fix $\epsilon > 0$ and let

$$\theta_i = \begin{cases} \epsilon & i \leq k \\ 0 & \text{otherwise} \end{cases}.$$

Then, $z_* = z_1$. The upper bound guarantee of [5][20] is at least

$$\sum_{i=1}^{k} \max_{z \neq z_*} \frac{|z_* \Delta z|}{\theta_i (z_* - z)^2} + \sum_{i=k+1}^{d} \frac{|z_* \Delta z_i|}{\theta_i (z_* - z_i)^2} = d \left( \frac{k+1}{(ke)^2} \right) \geq \frac{d}{ke^2}$$

(75)

On the other hand, we have that

$$\rho^* = \max_{z \neq z_*} \frac{\|z_* - z\|^2_{A(\lambda)^{-1}}}{\theta_i (z_* - z)^2} \leq 2 \left( \frac{k^2 + d}{(ke)^2} \right)$$

(76)

where we took

$$\lambda_i = \begin{cases} \frac{1}{2k^2} + \frac{1}{2d^2} & i \leq k \\ \frac{1}{2d^2} & \text{otherwise} \end{cases}.$$

Putting $k = \sqrt{d}$ into (75) and (76) yields the result. \qed

In the matching problem, if $\theta = \mathbb{1}\{i \in z\} \Delta$ for some $z \in Z$ and $\Delta > 0$, we say that it is an instance of HOMOGENOUS MATCHING. The following result appears in [10]. It shows that the sample complexity of [5][20] is correct for the homogeneous matching problem.

Proposition 10. Consider the homogenous MATCHING problem. Then, $\rho^* = \Theta(d/\Delta^2)$. Further, letting

$$\varphi_i = \max_{z \in Z \setminus \{z_*\}} z_* \Delta z \frac{|z_* \Delta z| \log(|B(z_* | z_* \Delta z|)|)}{\Delta z^2}$$

we have that $\sum_{i=1}^{n} \varphi_i = O(d/\Delta^2)$.

Remark 2. It follows from Proposition 4 that for the homogenous MATCHING problem, $\gamma^* \leq O(\log(\log(d))d/\Delta^2)$

The following result appears in [8]. It shows that there is a gap of order $d$ between the sample complexities in [10] and [13] and the lower bound.

Proposition 11. Let $d$ be even. Consider the combinatorial bandit setting where $X = \{e_1, \ldots, e_d\}$ and $Z = \{d/2, \{d/2 + 1, \ldots, d\}$ and $\theta_i = \mathbb{1}\{i \leq d/2\}$. Then, the guarantee of the CLUCB in [10] and the algorithm in [13] is $\Omega(d \epsilon^{-1} \ln(1/d))$. On the other hand, $\rho^* = \epsilon^2$.

The following result shows that the sample complexity cannot depend on $\log(Z)$ because $|Z|$ can be arbitrarily large while $\gamma^* \leq 1$.

Proposition 12. For any $N \in \mathbb{N}$, there exists an instance of the transductive linear bandit problem where $|Z| \geq N$ and $\gamma^* \leq 1$.

Proof of Proposition 12 Let $X = \{e_1, \ldots, e_d\}$. Let $\theta = a e_1$ for a constant $a > 0$ to be chosen later. Fix $\epsilon > 0$. Let $z_1 = e_1$. There exists $z_2, \ldots, z_N$ such that for every $i$, $\|z_i - z_i\|_2 = \epsilon$, $e_1^T z_i = 0$, and
We briefly sketch the extension to SubGaussian noise. First, we define some notation: If $c > a$, where $a$ is a universal constant, we assume $y$ is independent mean-0 subGaussian noise. Define the matrix $X = \begin{pmatrix} x_{1,1}^\top \\ \vdots \\ x_{d,1}^\top \end{pmatrix}$. Let $n \geq d$ and fix a set of measurements $x_{I,1}, \ldots, x_{I,n}$ and let $y_1, \ldots, y_n$ be the associated observations where we assume $y_i = x_i^\top \theta + \eta_i$ for $\eta_i$ is independent mean-0 subGaussian noise. Define the matrix $X = \begin{pmatrix} x_{1,1}^\top \\ \vdots \\ x_{d,1}^\top \end{pmatrix}$.

Define $\tilde{\theta} = (X^\top X)^{-1} X^\top Y$. Note that $\tilde{\theta} - \theta = (X^\top X)^{-1} X^\top \eta$. Note that $\|\eta\|_{\psi_2} \leq 1$. For any $v \in \mathbb{R}^d$,

$$\|v^\top (X^\top X)^{-1} X^\top \eta\|_{\psi_2} = \|X (X^\top X)^{-1} v\|_2 \frac{1}{\|X (X^\top X)^{-1} v\|_2} v^\top (X^\top X)^{-1} X^\top \eta \|_{\psi_2} \leq \|X (X^\top X)^{-1} v\|_2$$

This shows that $\|v^\top (X^\top X)^{-1} X^\top \eta\|_{\psi_2} \leq \|v^\top (X^\top X)^{-1} X^\top \tilde{\eta}\|_{\psi_2}$ where $\tilde{\eta} \sim N(0, I)$. Thus, applying Theorem 8.5.5 and Talagrand's majorizing measure theorem (Theorem 8.6.1) from \cite{32} yields for all $z \in Z \setminus \{z_*\}$

$$(z_* - z)^\top \tilde{\theta} \geq (z_* - z)^\top \theta - c \left( \sqrt{\sup_{z \notin Z \setminus \{z_*\}} (z_* - z)^\top A^{-1/2} \eta} \right)$$

where $c > 0$ is a universal constant, which is the essential concentration inequality used for the arguments in this paper.
Figure 3: The relative performance of CLUCB, UA, and the algorithm from [5] (denoted DisRegion) to PEACE on an instance of biclique.

**M Experiment Details**

**Biclique Experiment:** In the biclique problem, we are given a complete balanced bipartite graph with $\sqrt{d}$ nodes in each group and a total of $d$ edges. $Z$ is the set of bicliques with $\sqrt{s}$ nodes from each group in the bipartite graph. This problem is NP-hard, so there is no linear maximization oracle. Therefore, we consider a small instance where $\sqrt{d} = 8$ and $\sqrt{s} = 2$. We use a similar setup to the other combinatorial bandit experiments. We pick two random non-overlapping bicliques and let $B_1$ and $B_2$ denote the set of their respective edges. If $e \in B_1$, we set $\theta_e = 1$, and if $e \in B_2$, we set $\theta_e = 1 - h$ for $h \in \{.1, .8, .6, .4, .2\}$. Otherwise, we set $\theta_e = 0$.

We also compare to the Algorithm 4 from [5] (denoted DisRegion), which attains the best sample complexity result from that paper. Figure 3 shows that as the gap between the best biclique and the second best biclique decreases, the performance of the competing algorithms degrades relative to Peace. For example, for large $h$, Peace and DisRegion have similar performance but for $h = .2$, DisRegion requires more than 3 times as many samples as Peace.

**Combinatorial Bandit Experiments:** We used Python 3 and parallelized the simulations on an Intel(R) Xeon(R) CPU E5-2690. For each experiment, we generate noise from a standard normal distribution. We used the stochastic mirror descent algorithm described in Section K, but let $\lambda \in \Delta$ (instead of $\tilde{\Delta}$). We ran the algorithm for 1000 iterations with a batch size of 10 on all experiments. Once we obtained a $\lambda \in \Delta$, we used 2,000 samples to form an empirical mean to estimate the Gaussian width. We considered the setting where it is known that $\max_z \Delta_z \leq 2d$, which holds for example when $\theta \in [-1, 1]$, and thus solved

$$\inf_{\lambda \in \Delta} \mathbb{E}_{\eta \sim N(0, I)} \left[ \max_{z \in Z} \frac{(\tilde{z}_k - z)^T A(\lambda)^{-1/2} \eta - \theta^T_k (\tilde{z}_k - z)}{2^{-k} \cdot 2d + \theta^T_k (\tilde{z}_k - z)} \right]^2$$

instead of (26). We rounded our designs $\tau_k \lambda$ simply by taking the ceiling (which only incurs a loss of an additive factor of $d$ because $|\lambda| \leq d$).

To implement CLUCB, we use a state-of-the-art anytime confidence bound (inequality (2) from [17]), which is much better than the one used in [10]. For the uniform allocation algorithm, we use the termination condition that one obtains from applying the TIS inequality (Theorem 5.8 in [2]) to the process $\theta^T (z - z')$.

We used 20 trials for the matching experiment, 30 trials for the shortest path experiment, and 60 trials for the biclique experiment. We generated 95% confidence intervals using the bootstrap.
Transductive Linear Bandits: We made two main changes to the algorithm as written, both focused on computing the objective $\inf_{\lambda \in \Delta} \tau(\lambda; Z_k)$ more effectively. Firstly, we considered two different subproblems: $\min_{\lambda \in \Delta} \mathbb{E}_{\eta \sim N(0, I)} [\max_{z', z \in Z_k} (z - z')^T A(\lambda)^{-1/2} \eta]^2$ and $\min_{\lambda} \max_{z, z'} ||z' - z||^2_{A(\lambda)^{-1}}$. In the setting where there are extremely large number of arms, it is not practical to take a max over all pairs of them - so in both subproblems we only took the max over $\hat{z}_k - Z_k$ where $\hat{z}_k = \text{argmax}_{z \in Z_k} \hat{\theta}_k^T z$. To justify this, we point out that by Theorem 7.5.2 of [32] $\mathbb{E}_{\eta \sim N(0, I)} [\max_{z', z \in Z_k} (z - z')^T A(\lambda)^{-1/2} \eta] = 2\mathbb{E}_{\eta \sim N(0, I)} [\max_{z \in Z_k} (\hat{z}_k - z)^T A(\lambda)^{-1/2} \eta]$, and $\min_{\lambda} \max_{z, z'} ||z' - z||^2_{A(\lambda)^{-1}} \leq 4 \min_{\lambda} \max_{z \in Z_k} ||\hat{z}_k - z||^2_{A(\lambda)^{-1}}$. Motivated by this, we computed the distribution $\lambda' = \arg\min_{\lambda} \mathbb{E}_{\eta \sim N(0, I)} [\max_{z \in Z_k} (\hat{z}_k - z)^T A(\lambda)^{-1/2} \eta]$ and $\lambda'' = \min_{\lambda} \max_{z} ||\hat{z}_k - z||^2_{A(\lambda)^{-1}}$ and set $\lambda_k = (\lambda' + \lambda'')/2$. Note that using this distribution only makes the algorithm perform worst than if the optimal - it does not affect correctness in anyway.

Fixed Budget: As in the previous, we computed an allocation not using $\gamma(Z_k)$ but rather a minimum over the differences $\hat{z}_k - Z_k$. 

47
References

[1] Zeyuan Allen-Zhu, Yuanzhi Li, Aarti Singh, and Yining Wang. Near-optimal discrete optimization for experimental design: A regret minimization approach. *Mathematical Programming*, pages 1–40, 2020.

[2] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

[3] Richard A Brualdi. Comments on bases in dependence structures. *Bulletin of the Australian Mathematical Society*, 1(2):161–167, 1969.

[4] Sébastien Bubeck et al. Convex optimization: Algorithms and complexity. *Foundations and Trends® in Machine Learning*, 8(3-4):231–357, 2015.

[5] Tongyi Cao and Akshay Krishnamurthy. Disagreement-based combinatorial pure exploration: Sample complexity bounds and an efficient algorithm. In *Conference on Learning Theory*, pages 558–588, 2019.

[6] Lijie Chen and Jian Li. On the optimal sample complexity for best arm identification. *arXiv preprint arXiv:1511.03774*, 2015.

[7] Lijie Chen, Anupam Gupta, and Jian Li. Pure exploration of multi-armed bandit under matroid constraints. In *Conference on Learning Theory*, pages 647–669, 2016.

[8] Lijie Chen, Anupam Gupta, Jian Li, Mingda Qiao, and Ruosong Wang. Nearly optimal sampling algorithms for combinatorial pure exploration. In *Conference on Learning Theory*, pages 482–534, 2017.

[9] Lijie Chen, Jian Li, and Mingda Qiao. Nearly instance optimal sample complexity bounds for top-k arm selection. In *Artificial Intelligence and Statistics*, pages 101–110, 2017.

[10] Shouyuan Chen, Tian Lin, Irwin King, Michael R Lyu, and Wei Chen. Combinatorial pure exploration of multi-armed bandits. In *Advances in Neural Information Processing Systems*, pages 379–387, 2014.

[11] Eyal Even-Dar, Shie Mannor, and Yishay Mansour. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *Journal of machine learning research*, 7(Jun):1079–1105, 2006.

[12] Tanner Fiez, Lalit Jain, Kevin G Jamieson, and Lillian Ratliff. Sequential experimental design for transductive linear bandits. In *Advances in Neural Information Processing Systems*, pages 10666–10676, 2019.

[13] Victor Gabillon, Alessandro Lazaric, Mohammad Ghavamzadeh, Ronald Ortner, and Peter Bartlett. Improved learning complexity in combinatorial pure exploration bandits. In *Artificial Intelligence and Statistics*, pages 1004–1012, 2016.

[14] Aurélien Garivier and Emilie Kaufmann. Optimal best arm identification with fixed confidence. In *Conference on Learning Theory*, pages 998–1027, 2016.

[15] Barbara Heller. *Statistics for experimenters, an introduction to design, data analysis, and model building: GEP Box, WG Hunter and JS Hunter, John Wiley and Sons, New York, NY. 1978. Pergamon, 1986.*

[16] Daniel N Hill, Houssam Nassif, Yi Liu, Anand Iyer, and SVN Vishwanathan. An efficient bandit algorithm for realtime multivariate optimization. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1813–1821, 2017.

[17] Steven R Howard, Aaditya Ramdas, Jon McAuliffe, and Jasjeet Sekhon. Uniform, nonparametric, non-asymptotic confidence sequences. *arXiv preprint arXiv:1810.08240*, 2018.

[18] Daniel Hsu and Sivan Sabato. Heavy-tailed regression with a generalized median-of-means. In *International Conference on Machine Learning*, pages 37–45, 2014.

[19] IA Ibragimov, VN Sudakov, and BS Tsirelson. Norms of gaussian sample functions. In *Proceedings of the third Japan USSR symposium on probability theory, lecture notes in math*, volume 550, pages 20–41, 1976.

[20] Lalit Jain and Kevin G Jamieson. A new perspective on pool-based active classification and false-discovery control. In *Advances in Neural Information Processing Systems*, pages 13992–14003, 2019.
[21] Kevin Jamieson, Matthew Malloy, Robert Nowak, and Sébastien Bubeck. lil’ucb: An optimal exploration algorithm for multi-armed bandits. In Conference on Learning Theory, pages 423–439, 2014.

[22] Shivaram Kalyanakrishnan, Ambuj Tewari, Peter Auer, and Peter Stone. PAC subset selection in stochastic multi-armed bandits. In Proceedings of the 29th International Conference on Machine Learning, ICML 2012, Edinburgh, Scotland, UK, June 26 - July 1, 2012, 2012.

[23] Zohar Karnin, Tomer Koren, and Oren Somekh. Almost optimal exploration in multi-armed bandits. In Sanjoy Dasgupta and David Mcallester, editors, Proceedings of the 30th International Conference on Machine Learning (ICML-13), volume 28, pages 1238–1246. JMLR Workshop and Conference Proceedings, May 2013.

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

[25] Emilie Kaufmann, Olivier Cappé, and Aurélien Garivier. On the complexity of best-arm identification in multi-armed bandit models. The Journal of Machine Learning Research, 17(1):1–42, 2016.

[26] Tor Lattimore and Csaba Szepesvári. Bandit Algorithms. Cambridge University Press, 2020.

[27] Andrea Locatelli, Maurilio Gutzeit, and Alexandra Carpentier. An optimal algorithm for the thresholding bandit problem. In International Conference on Machine Learning, pages 1690–1698, 2016.

[28] Max Simchowitz, Kevin Jamieson, and Benjamin Recht. The simulator: Understanding adaptive sampling in the moderate-confidence regime. In Conference on Learning Theory, pages 1794–1834, 2017.

[29] Marta Soare. Sequential resource allocation in linear stochastic bandits. PhD thesis, 2015.

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

[31] Chao Tao, Saúl Blanco, and Yuan Zhou. Best arm identification in linear bandits with linear dimension dependency. In International Conference on Machine Learning, pages 4877–4886, 2018.

[32] Roman Vershynin. High-Dimensional Probability. 2019.

[33] 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.