# Identifying All $\varepsilon$-Best Arms in (Misspecified) Linear Bandits

Zhekai Li
Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, zhekaili@mit.edu

Tianyi Ma
School of Operations Research and Information Engineering, Cornell University, tm693@cornell.edu

Cheng Hua
Antai College of Economics and Management, Shanghai Jiao Tong University, cheng.hua@sjtu.edu.cn

Ruihao Zhu
SC Johnson College of Business, Cornell University, ruihao.zhu@cornell.edu

Motivated by the need to efficiently identify multiple candidates in high trial-and-error cost tasks such as drug discovery, we propose a near-optimal algorithm to identify all $\varepsilon$-best arms (*i.e.*, those at most $\varepsilon$ worse than the optimum). Specifically, we introduce LinFACT, an algorithm designed to optimize the identification of all $\varepsilon$-best arms in linear bandits. We establish a novel information-theoretic lower bound on the sample complexity of this problem and demonstrate that LinFACT achieves instance optimality by matching this lower bound up to a logarithmic factor. A key ingredient of our proof is to integrate the lower bound directly into the scaling process for upper bound derivation, determining the termination round and thus the sample complexity. We also extend our analysis to settings with model misspecification and generalized linear models. Numerical experiments, including synthetic and real drug discovery data, demonstrate that LinFACT identifies more promising candidates with reduced sample complexity, offering significant computational efficiency and accelerating early-stage exploratory experiments.

*Key words*: ranking and selection; sequential decision making; simulation; adaptive experiment; model misspecification

## 1. Introduction

This paper addresses the problem of identifying the best set of options from a finite pool of candidates. A decision-maker sequentially selects candidates for evaluation, observing independent noisy rewards that reflect their quality. The goal is to strategically allocate measurement efforts to identify the desired candidates. This problem belongs to the class of pure exploration problems, which fall under the bandit framework but differ from traditional multi-armed bandits (MABs) that balance exploration and exploitation to minimize cumulative regret. Instead, pure exploration focuses on efficient information gathering to confidently apply the chosen best or best set of options. This approach is particularly relevant in applications such as drug discovery and product testing,

where identifying the most promising candidates is followed by utilizing them under high-cost conditions, such as clinical trials or large-scale manufacturing tests.

Conventional pure exploration focuses on identifying the optimal candidate, often referred to as the best arm in the bandit setting. However, in many real-world scenarios, candidates with rewards falling slightly below the optimum may later demonstrate advantageous traits, such as fewer side effects, simpler manufacturing processes, or lower resistance during implementation. Motivated by this insight, this paper aims to identify all $\varepsilon$-best candidates (*i.e.*, those whose performance is at most $\varepsilon$ worse than the optimum). This approach is especially valuable when exploring a range of nearly optimal options is necessary. Promoting multiple promising candidates not only mitigates risk but also increases the chances that at least one will prove successful.

This setting captures many applications in real-world scenarios. For example:

- *Drug Discovery*: In drug discovery, pharmaceutical companies aim to identify as many promising drug candidates as possible during preclinical stages. These candidates, known as preclinical candidate compounds (PCCs), are optimized compounds prepared for preclinical testing to assess efficacy, safety, and pharmacokinetics before advancing to clinical trials. Given the inherent risks and high failure rates in subsequent drug development (Das et al. 2021), starting with a larger pool of potential candidates increases the chances of identifying at least one successful, marketable drug.
- *Assortment Design*: In e-commerce (Boyd and Bilegan 2003, Elmaghraby and Keskinocak 2003, Feng et al. 2022), recommender systems (Huang et al. 2007, Peukert et al. 2023), and streaming services (Alaei et al. 2022, Godinho de Matos et al. 2018), expanding the consideration set (*e.g.*, products, movies, or songs) can improve user satisfaction and increase revenues. Offering a diverse range of recommendations helps cater to varying tastes and preferences.
- *Automatic Machine Learning*: In automatic machine learning (AutoML) (Thornton et al. 2013), the goal is to automate the process of selecting algorithms and tuning hyperparameters by providing multiple promising choices for predictive tasks. Due to the randomness of limited data, the best out-of-sample model may not always be optimal. Therefore, providing users with a diverse set of models that yield good results is critical to assisting them in selecting the best algorithm and hyperparameters.

### 1.1. Main Contributions

This paper focuses on identifying all $\varepsilon$-best arms in the linear bandit setting and presents contributions in both the algorithmic and theoretical dimensions.

- *$\delta$-Probably Approximately Correct (PAC) Algorithm:* On the algorithmic front, we introduce LinFACT (**Lin**ear **F**ast **A**rm **C**lassification with **T**hreshold estimation), a $\delta$-PAC (see Section

2.1 for a formal definition) phase-based algorithm for identifying all $\varepsilon$-best arms in linear bandit problems. LinFACT demonstrates superior effectiveness compared to existing pure exploration algorithms.

- *Matching Bound of Sample Complexity:* We make two key technical contributions. First, we derive an information-theoretic lower bound on the problem complexity for identifying all $\varepsilon$-best arms in linear bandits. To the best of our knowledge, this is the first such result in the literature. Second, we establish two distinct upper bounds on the expected sample complexity of LinFACT, illustrating the differences between various optimal design criteria used in its implementation. Notably, we demonstrate that LinFACT achieves instance optimality when using the $\mathcal{XY}$-optimal design criterion, matching the lower bound up to a logarithmic factor. The $\mathcal{XY}$-optimal design focuses on contrasting pairs of arms rather than evaluating each arm individually. Our analysis leverages the lower bound directly in defining the classification termination round and in scaling the upper bound.
- *Accounting for Misspecified Models and GLMs:* We extend our framework beyond linear models to handle misspecified linear bandits and generalized linear models (GLMs). For both cases, we provide theoretical upper bounds on the expected sample complexity. Furthermore, we analyze how prior knowledge of model misspecification impacts the algorithmic upper bounds and performance, and how the incorporation of GLMs influences the sample complexity.
- *Numerical Studies with Real-World Datasets:* We conduct extensive numerical experiments to demonstrate that our LinFACT algorithm outperforms existing methods in terms of sample complexity, computational efficiency, and reliable identification of all $\varepsilon$-best arms. In experiments with synthetic data, LinFACT outperforms other baselines in both adaptive and static settings. Using a real-world drug discovery dataset (Free and Wilson 1964), we further show that LinFACT achieves superior performance compared to previous algorithms. Notably, LinFACT is computationally efficient with time complexity of $O(Kd^2)$, which is lower than $O(nKd^2)$ of Lazy TTS ($n$ is a non-negligible number) (Rivera and Tewari 2024), $O(Kd^3)$ of top $m$ algorithms (Réda et al. 2021a), and $O(K^2 \log K)$ of KGCB (Negoescu et al. 2011).

### 1.2. Related Literature

**Pure Exploration.** The multi-armed bandits (MABs) model has been a critical paradigm for addressing the exploration-exploitation trade-off since its introduction by Thompson (1933) in the context of medical trials. While much of the research focuses on minimizing cumulative regret (Bubeck et al. 2012, Lattimore and Szepesvári 2020), our work focuses on the pure exploration setting (Koenig and Law 1985), where the goal is to select a subset of arms and evaluation is based on the final outcome. This distinction highlights the context-specific benefits of each approach:

MABs are suited for tasks where the goal is to optimize rewards in real-time, balancing exploration and exploitation, whereas pure exploration is focused on identifying a set of satisfactory arms, without the immediate concern for reward maximization. In pure exploration, the algorithm prioritizes information gathering over reward collection, transforming the objective from reward-centric to information-centric. The focus is on efficiently acquiring sufficient information about all arms for confident identification.

The origins of pure exploration problems date back to the 1950s in the context of stochastic simulation, specifically within ordinal optimization (Shin et al. 2018, Shen et al. 2021) or the Ranking and Selection (R&S) problem, first addressed by Bechhofer (1954). Various methodologies have since been developed to solve the canonical R&S problem, including elimination-type algorithms (Kim and Nelson 2001, Bubeck et al. 2013, Fan et al. 2016), Optimal Computing Budget Allocation (OCBA) (Chen et al. 2000), knowledge-gradient algorithms (Frazier et al. 2008, 2009, Ryzhov et al. 2012), UCB-type algorithms (Kaufmann and Kalyanakrishnan 2013), and the unified gap-based exploration (UGapE) algorithm (Gabillon et al. 2012). Comprehensive reviews of the R&S problem can be found in Kim and Nelson (2006), Hong et al. (2021), with the most recent overview provided by Li et al. (2024).

The general framework of pure exploration encompasses various exploration tasks (Qin and You 2025), including best arm identification (BAI) (Mannor and Tsitsiklis 2004, Even-Dar et al. 2006, Russo 2020, Komiyama et al. 2023, Simchi-Levi et al. 2024), top $m$ identification (Kalyanakrishnan and Stone 2010, Kalyanakrishnan et al. 2012), threshold bandits (Locatelli et al. 2016, Abernethy et al. 2016), and satisficing bandits (Feng et al. 2025). In applications such as drug discovery, pharmacologists aim to identify a set of highly potent drug candidates from potentially millions of compounds, with only the selected candidates advancing to more extensive testing. Given the uncertainty of final outcomes and the high cost of trial-and-error, identifying multiple promising candidates simultaneously is crucial. To minimize the cost of early-stage exploration, adaptive, sequential experimental designs are necessary, as they require fewer experiments compared to fixed designs.

**All $\varepsilon$-Best Arms Identification.** Conventional objectives, such as identifying the top $m$ best arms or all arms above a certain threshold, often face significant challenges. In the top $m$ task, selecting a small $m$ may exclude promising candidates, while choosing a large $m$ may include ineffective options, requiring an impractically large number of experiments. Similarly, setting a threshold too high can exclude viable candidates. Both approaches depend on prior knowledge of the problem to achieve good performance, which may not be available in real-world applications.

In contrast, identifying all $\varepsilon$-best arms (those within $\varepsilon$ of the best arm) overcomes these limitations. This approach promotes broader exploration while providing a robust guarantee: no significantly suboptimal arms will be selected, thereby improving the reliability of downstream tasks (Mason et al. 2020). The all $\varepsilon$-best arms identification problem generalizes both the top $m$ and threshold bandit problems. It reduces to the top $m$ problem if the number of $\varepsilon$-best arms is known in advance, and to a threshold bandit problem if the value of the best arm is known.

Mason et al. (2020) introduced the problem complexity for identifying all $\varepsilon$-best arms and derived a lower bound in the low-confidence regime. However, their lower bound involves a summation that may be unnecessary, indicating room for improvement. Building on Mason's work, Al Marjani et al. (2022) derived tighter lower bounds by fully characterizing the alternative bandit instances that an optimal sampling strategy must distinguish and eliminate. They also proposed the asymptotically optimal Track-and-Stop algorithm. However, both Mason et al. (2020) and Al Marjani et al. (2022) consider stochastic bandits without structures. In contrast, we study this problem in the linear bandit setting (Abe and Long 1999), which leverages structural relationships among arms. This presents new challenges, but also allows it to handle more complex scenarios. As a result, our work establishes the first information-theoretic lower bound for identifying all $\varepsilon$-best arms in linear bandits, applicable to any $\delta$-PAC algorithm.

An extended literature review of misspecified linear bandits and generalized linear bandit models can be found in Section EC.1 of the online appendix.

## 2. Problem Formulation

**Notations.** In this paper, we denote the set of positive integers up to $N$ by $[N] = \{1, \ldots, N\}$. Vectors and matrices are represented using boldface notation. The inner product of two vectors is denoted by $\langle \cdot, \cdot \rangle$. We define the weighted matrix norm $\|\boldsymbol{x}\|_{\boldsymbol{A}}$ as $\sqrt{\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}}$, where $\boldsymbol{A}$ is a positive semi-definite matrix that weights and scales the norm. For two probability measures $P$ and $Q$ over a common measurable space, if $P$ is absolutely continuous with respect to $Q$, the Kullback-Leibler divergence between $P$ and $Q$ is defined as

$$\mathrm{KL}(P, Q) = \begin{cases} \int \log\left(\frac{dP}{dQ}\right) dP, & \text{if } Q \ll P; \\ \infty, & \text{otherwise,} \end{cases} \tag{1}$$

where $\frac{dP}{dQ}$ is the Radon-Nikodym derivative of $P$ with respect to $Q$, and $Q \ll P$ indicates that $Q$ is absolutely continuous with respect to $P$.

**Setting.** We address the problem of identifying all $\varepsilon$-best arms from a finite set of $K$ arms where $K$ is a (possibly large) positive integer. Each arm $i \in [K]$ has an associated reward distribution with an unknown fixed mean $\mu_i$. Let the mean vector of all arms be denoted as $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_K)$,

which can only be estimated through bandit feedback from the selected arms. Without loss of generality, we assume $\mu_1 > \mu_2 \geq \ldots \geq \mu_K$. The gap $\Delta_i = \mu_1 - \mu_i$ (for $i \neq 1$) represents the difference in expected rewards between the optimal arm and arm $i$. To this end, we give a formal definition of the notion of $\varepsilon$-best arm.

DEFINITION 1. ($\varepsilon$-Best Arm). Given $\varepsilon > 0$, an arm $i$ is called *$\varepsilon$-best* if $\mu_i \geq \mu_1 - \varepsilon$.

Here, we adopt an additive framework to define $\varepsilon$-best arms. There also exists a multiplicative counterpart, where an arm $i$ is considered $\varepsilon$-best if $\mu_i \geq (1-\varepsilon)\mu_1$. While our study focuses on the additive model, the analysis for the multiplicative model follows similar reasoning. We denote the set of all $\varepsilon$-best arms[1] for a mean vector $\boldsymbol{\mu}$ as

$$G_\varepsilon(\boldsymbol{\mu}) \coloneqq \{i : \mu_i \geq \mu_1 - \varepsilon\}. \tag{2}$$

Define $\alpha_\varepsilon \coloneqq \min_{i \in G_\varepsilon(\boldsymbol{\mu})} (\mu_i - (\mu_1 - \varepsilon))$ as the distance from the smallest $\varepsilon$-best arm to the threshold $\mu_1 - \varepsilon$. Furthermore, if the complement of $G_\varepsilon(\boldsymbol{\mu})$, denoted as $G_\varepsilon^c(\boldsymbol{\mu})$, is non-empty, we define $\beta_\varepsilon \coloneqq \min_{i \in G_\varepsilon^c(\boldsymbol{\mu})} ((\mu_1 - \varepsilon) - \mu_i)$ as the closest distance from the threshold to the highest mean value of any arm that is not considered $\varepsilon$-best.

We study this problem under a linear structure, where the mean values depend on an unknown parameter vector $\boldsymbol{\theta} \in \mathbb{R}^d$. Each arm $i$ is associated with a feature vector $\boldsymbol{a}_i \in \mathbb{R}^d$. Let $\mathcal{A} \subset \mathbb{R}^d$ be the set of feature vectors, and let $\boldsymbol{\Psi} \coloneqq [\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_K] \in \mathbb{R}^{K \times d}$ be the feature matrix. With parameter $\boldsymbol{\theta}$, the mean value can be represented as $\boldsymbol{\mu}(\boldsymbol{\theta})$. For simplicity, we will consistently refer to the bandit instance as $\boldsymbol{\mu}$. When an arm $A_t$ with corresponding feature vector $\boldsymbol{a}_{A_t} \in \mathcal{A}$ is selected at time $t$, we observe the bandit feedback $X_t$, given by

$$X_t = \boldsymbol{a}_{A_t}^\top \boldsymbol{\theta} + \eta_t, \tag{3}$$

where $\mu_{A_t} = \boldsymbol{a}_{A_t}^\top \boldsymbol{\theta}$ is the true mean reward of the selected arm, and $\eta_t$ is a noise variable. We also make two additional standard assumptions on the norm of the parameters and the noise distributions (Abbasi-Yadkori et al. 2011).

ASSUMPTION 1. *Assume* $\max_{i \in [K]} \|\boldsymbol{a}_i\|_2 \leq L_1$, *where* $\|\cdot\|_2$ *denotes the* $\ell_2$*-norm and* $L_1$ *is a constant.*

ASSUMPTION 2. *The noise* $\eta_t$ *is conditionally 1-sub-Gaussian, i.e., for any* $\lambda \in \mathbb{R}$,

$$\mathbb{E}\left[e^{\lambda \eta_t} \,\middle|\, \boldsymbol{a}_{A_1}, \ldots, \boldsymbol{a}_{A_{t-1}}, \eta_1, \ldots, \eta_{t-1}\right] \leq \exp\left(\frac{\lambda^2}{2}\right). \tag{4}$$

[1] The set of all $\varepsilon$-best arms $G_\varepsilon(\boldsymbol{\mu})$ is also referred to as the good set in this paper.

### 2.1. Probably Approximately Correct Algorithm Framework

Our goal is to identify all $\varepsilon$-best arms with high confidence while minimizing the sampling budget. To achieve this, we employ three main components: stopping rule, sampling rule, and decision rule.

At each time step $t$, the stopping rule $\tau_\delta$ determines whether to continue or stop the process. If the process continues, an arm is selected according to the sampling rule, and the corresponding random reward is observed. When the process stops at $t = \tau_\delta$, a decision rule provides an estimate $\widehat{\mathcal{I}}_{\tau_\delta}$ of the true solution set $\mathcal{I}(\boldsymbol{\mu})$, which in our problem is the set of all $\varepsilon$-best arms, $G_\varepsilon(\boldsymbol{\mu})$.

We define the set of all viable mean vectors $\boldsymbol{\mu}$ as

$$M \coloneqq \left\{ \boldsymbol{\mu} \in \mathbb{R}^K \;\middle|\; \exists \boldsymbol{\theta} \in \mathbb{R}^d, \boldsymbol{\mu} = \boldsymbol{\Psi}\boldsymbol{\theta} \wedge \|\boldsymbol{a}_i\|_2 \leq L_1 \text{ for each } i \in [K] \right\}. \tag{5}$$

Here, the set $M$ consists of all possible mean vectors $\boldsymbol{\mu}$ that can be expressed as a linear combination of the parameter vector $\boldsymbol{\theta}$ through the matrix $\boldsymbol{\Psi}$.

We focus on algorithms that are probably approximately correct with high confidence, referred to as $\delta$-PAC algorithms.

Definition 2. ($\delta$-*PAC* Algorithm). An algorithm is $\delta$-*PAC* for all $\varepsilon$-best arms identification if it identifies the correct solution set with a probability of at least $1-\delta$ for any problem instance with mean $\boldsymbol{\mu} \in M$, *i.e.*,

$$\mathbb{P}_{\boldsymbol{\mu}} \left( \tau_\delta < \infty, \widehat{\mathcal{I}}_{\tau_\delta} = G_\varepsilon(\boldsymbol{\mu}) \right) \geq 1 - \delta, \qquad \forall \boldsymbol{\mu} \in M. \tag{6}$$

Upon stopping, $\delta$-PAC algorithms ensure the identification of all $\varepsilon$-best arms with high confidence. Therefore, our goal is to design a $\delta$-PAC algorithm that minimizes the stopping time, formulated as the following optimization problem.

$$\min \quad \mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta] \tag{7}$$

$$\text{s.t.} \quad \mathbb{P}_{\boldsymbol{\mu}} \left( \tau_\delta < \infty, \widehat{\mathcal{I}}_{\tau_\delta} = G_\varepsilon(\boldsymbol{\mu}) \right) \geq 1 - \delta, \qquad \forall \boldsymbol{\mu} \in M. \tag{8}$$

### 2.2. Optimal Design of Experiment

Linear bandit algorithms can be viewed as an online, adaptive counterpart to the classical optimal design problem. This section develops the key theoretical foundations, emphasizing the confidence bounds for parameter estimation that guide both the design of our algorithm and the subsequent analysis.

**Ordinary Least Squares.** Consider a sequence of pulled arms, denoted as $A_1, A_2, \ldots, A_t$, and the corresponding observed rewards $X_1, X_2, \ldots, X_t$. If the feature vectors of these arms,

$\boldsymbol{a}_{A_1}, \boldsymbol{a}_{A_2}, \ldots, \boldsymbol{a}_{A_t}$, span the space $\mathbb{R}^d$, the ordinary least squares (OLS) estimator for the parameter $\boldsymbol{\theta}$ is given by

$$\hat{\boldsymbol{\theta}}_t = \boldsymbol{V}_t^{-1} \sum_{n=1}^{t} \boldsymbol{a}_{A_n} X_n, \tag{9}$$

where $\boldsymbol{V}_t = \sum_{n=1}^{t} \boldsymbol{a}_{A_n} \boldsymbol{a}_{A_n}^\top \in \mathbb{R}^{d \times d}$ represents the information matrix. Using the properties of sub-Gaussian random variables, we can derive a confidence bound for the OLS estimator. This bound, denoted as $B_{t,\delta}$, is detailed in Proposition 1. The confidence region for the parameter $\boldsymbol{\theta}$ at time step $t$ is given by

$$\mathcal{C}_{t,\delta} = \left\{ \boldsymbol{\theta} : \|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}\|_{\boldsymbol{V}_t} \leq B_{t,\delta} \right\}. \tag{10}$$

PROPOSITION 1 **(Lattimore and Szepesvári (2020))**. *For any fixed sampling policy and any given vector $\boldsymbol{x} \in \mathbb{R}^d$, with probability at least $1-\delta$, the following holds.*

$$\left| \boldsymbol{x}^\top \left( \hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta} \right) \right| \leq \|\boldsymbol{x}\|_{\boldsymbol{V}_t^{-1}} B_{t,\delta}, \tag{11}$$

*where the anytime confidence bound $B_{t,\delta}$ is given by $B_{t,\delta} = 2\sqrt{2\left(d\log(6) + \log\left(\frac{1}{\delta}\right)\right)}$.*

In many practical scenarios, the observed data are not predetermined. To handle this, a martingale-based method can be employed, as described by Abbasi-Yadkori et al. (2011), to define an adaptive confidence bound for the OLS estimator. This accounts for the variability introduced by random rewards and adaptive sampling policies. The confidence interval in Proposition 1 highlights the connection between arm allocation policies in linear bandits and experimental design theory (Pukelsheim 2006). This connection serves as a fundamental component in constructing our algorithm.

**Feature Vector Projection.** At any time step where estimation needs to be made after sampling, if the feature vectors of the sampled arms do not span $\mathbb{R}^d$, we substitute them with dimensionality-reduced feature vectors (Yang and Tan 2022). Specifically, we project all feature vectors onto the subspace spanned by $\mathcal{A}$. Let $\boldsymbol{B} \in \mathbb{R}^{d \times d'}$ be an orthonormal basis for this subspace, where $d' < d$ is the dimension of the subspace. The new feature vector $\boldsymbol{a}'$ is then given by $\boldsymbol{a}' = \boldsymbol{B}^\top \boldsymbol{a}$. In this transformation, $\boldsymbol{B}\boldsymbol{B}^\top$ is a projection matrix, ensuring

$$\langle \boldsymbol{\theta}, \boldsymbol{a} \rangle = \left\langle \boldsymbol{\theta}, \boldsymbol{B}\boldsymbol{B}^\top \boldsymbol{a} \right\rangle = \left\langle \boldsymbol{B}^\top \boldsymbol{\theta}, \boldsymbol{B}^\top \boldsymbol{a} \right\rangle = \langle \boldsymbol{\theta}', \boldsymbol{a}' \rangle. \tag{12}$$

Equation (12) ensures that the mean values of all arms remain unchanged under the projection. The first equality holds because $\boldsymbol{B}\boldsymbol{B}^\top$ is a projection matrix and $\boldsymbol{a}$ lies in the subspace spanned by $\boldsymbol{B}$ (*i.e.*, $\boldsymbol{B}\boldsymbol{B}^\top \boldsymbol{a} = \boldsymbol{a}$). The second equality follows from the same matrix form $\boldsymbol{\theta}^\top \boldsymbol{B}\boldsymbol{B}^\top \boldsymbol{a}$. The third equality holds by definition.

**Optimal Design Criteria.** In contrast to stochastic bandits, where the mean values of the arms are estimated through repeated sampling of each arm, the linear bandit setting allows these values to be inferred from accurate estimation of the underlying parameter vector $\boldsymbol{\theta}$. As a result, pulling a single arm provides information about all arms.

A key sampling strategy in this context is the *G-optimal design*, which minimizes the maximum variance of the predicted responses across all arms by optimizing the fraction of times each arm is selected. Formally, the G-optimal design problem seeks a probability distribution $\pi$ on $\mathcal{A}$, where $\pi : \mathcal{A} \to [0,1]$ and $\sum_{\boldsymbol{a}\in\mathcal{A}} \pi(\boldsymbol{a}) = 1$, that minimizes

$$g(\pi) = \max_{\boldsymbol{a}\in\mathcal{A}} \|\boldsymbol{a}\|^2_{\boldsymbol{V}(\pi)^{-1}}, \tag{13}$$

where $\boldsymbol{V}(\pi) = \sum_{\boldsymbol{a}\in\mathcal{A}} \pi(\boldsymbol{a})\boldsymbol{a}\boldsymbol{a}^\top$ is the weighted information matrix, analogous to $\boldsymbol{V}_t$ in equation (9). The G-optimal design (13) ensures a tight confidence interval for mean value estimation. However, comparing the relative differences of mean values across different arms is more critical in identifying the best arms, rather than making the best estimation.

Therefore, we consider an alternative design criterion, the $\mathcal{XY}$*-optimal design*, that directly targets the estimation of these gaps. Consider $\mathcal{S} \subseteq \mathcal{A}$ as a subset of the arm space. We define

$$\mathcal{Y}(\mathcal{S}) \coloneqq \{\boldsymbol{a} - \boldsymbol{a}' : \forall \boldsymbol{a}, \boldsymbol{a}' \in \mathcal{S}, \boldsymbol{a} \neq \boldsymbol{a}'\} \tag{14}$$

as the set of vectors representing the differences between each pair of arms in $\mathcal{S}$. The $\mathcal{XY}$-optimal design minimizes

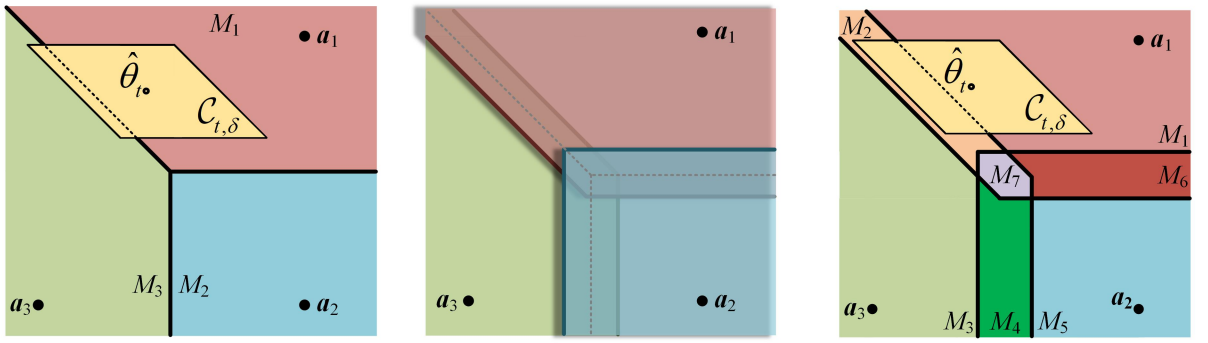
$$g_{\mathcal{XY}}(\pi) = \max_{\boldsymbol{y}\in\mathcal{Y}(\mathcal{A})} \|\boldsymbol{y}\|^2_{\boldsymbol{V}(\pi)^{-1}}. \tag{15}$$

As mentioned previously, the $\mathcal{XY}$-optimal design focuses on minimizing the maximum variance when estimating the differences (gaps) between pairs of arms. By doing so, it ensures differentiation between arms, rather than estimating each arm individually. This criterion is particularly useful when the goal is to identify relative performance rather than absolute quality.

## 3. Lower Bound and Problem Complexity

In this section, we present a novel information-theoretic lower bound for the problem of identifying all $\varepsilon$-best arms in linear bandits. Building on the approach of Soare et al. (2014), we extend the lower bound for best arm identification (BAI) to this more general setting. Figure 1 visualizes the structure of the stopping condition, with additional graphical insights provided in Section EC.4.1. These visualizations offer geometric intuition for the challenges involved in identifying all $\varepsilon$-best arms in linear bandits.

The sample complexity of an algorithm is quantified by the number of samples, denoted $\tau_\delta$, required to terminate the process. The goal of the algorithm design is to minimize the expected




(a) Best Arm Identification (BAI) in Linear Bandits

(b) Three Overlapping Regions Representing the Combination of All $\varepsilon$-Best Arms

(c) All $\varepsilon$-Best Arms Identification in Linear Bandits

**Figure 1** Illustration of the Stopping Condition: Best Arm Identification vs. All $\varepsilon$-Best Arms Identification

*Note. (a) Stopping occurs when the confidence region $\mathcal{C}_{t,\delta}$ for the estimated parameter $\hat{\boldsymbol{\theta}}_t$ contracts entirely within one of the three decision regions $M_i$ in a certain time step $t$. The boundaries between regions are defined by the hyperplanes $\boldsymbol{\vartheta}^\top(\boldsymbol{a}_i - \boldsymbol{a}_j) = 0$. Each dot represents an arm. (b) In the case of identifying all $\varepsilon$-best arms, the regions overlap. (c) These overlaps partition the space into seven distinct decision regions, increasing the difficulty of identification.*

sample complexity $\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta]$ across the entire set of algorithms $\mathcal{H}$. As introduced in Kaufmann et al. (2016), for $\delta \in (0,1)$, the non-asymptotic problem complexity of an instance $\boldsymbol{\mu}$ can be defined as

$$\kappa(\boldsymbol{\mu}) := \inf_{Algo \in \mathcal{H}} \frac{\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta]}{\log\left(\frac{1}{2.4\delta}\right)}, \tag{16}$$

which is the smallest possible constant such that the expected sample complexity $\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta]$ grows asymptotically in line with $\log\left(\frac{1}{2.4\delta}\right)$. The lower bound of the sample complexity $\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta]$ can be represented in a general form by the following proposition. Building on the analytical framework of Proposition 2, we formulate the $\varepsilon$-best arms identification problem in the linear bandit setting. This enables us to derive both lower and upper bounds on the sample complexity, thereby establishing the near-optimality of our algorithm.

PROPOSITION 2 **(Qin and You (2025))**. *For any $\boldsymbol{\mu} \in M$, there exists a set $\mathcal{X} = \mathcal{X}(\boldsymbol{\mu})$ and functions $\{C_x\}_{x \in \mathcal{X}}$ with $C_x : \mathcal{S}_K \times \mathcal{X} \to \mathbb{R}_+$ such that*

$$\kappa(\boldsymbol{\mu}) \geq \left(\Gamma^*_{\boldsymbol{\mu}}\right)^{-1}, \tag{17}$$

*where*

$$\Gamma^*_{\boldsymbol{\mu}} = \max_{\boldsymbol{p} \in \mathcal{S}_K} \min_{x \in \mathcal{X}} C_x(\boldsymbol{p}; \boldsymbol{\mu}). \tag{18}$$

In Proposition 2, $\mathcal{S}_K$ denotes the $K$-dimensional probability simplex, and $\mathcal{X} = \mathcal{X}(\boldsymbol{\mu})$ is referred to as the culprit set. This set comprises critical subqueries (or comparisons) that must be correctly resolved. An error in any of these comparisons may hinder the identification of the correct set.

For example, in the case of identifying the best arm, the culprit set is given by $\mathcal{X} = \{i : i \in [K] \setminus i^*\}$, where $i^*$ denotes the unique best arm, and each subquery involves distinguishing every arm from the best arm. In the threshold bandit problem, the culprit set consists of all arms, $\mathcal{X} = \{i : i \in [K]\}$, where each subquery requires accurately determining whether each arm exceeds the threshold. For the task of identifying the best $m$ arms, the culprit set is $\mathcal{X} = \{(i,j) : i \in \mathcal{I}, j \in \mathcal{I}^c\}$, where $\mathcal{I}$ represents the set of the best $m$ arms, and each subquery entails comparing the mean of each arm in $\mathcal{I}$ with those in the complement set $\mathcal{I}^c$.

The function $C_x(\boldsymbol{p}; \boldsymbol{\mu})$ represents the population version of the sequential generalized likelihood ratio statistic, which provides an information-theoretic measure of how easily each subquery, corresponding to the culprit $x \in \mathcal{X}$, can be answered.

In equation (18), the minimum aligns with the intuition that the instance posing the greatest challenge corresponds to the hardest subquery. The outer maximization seeks the optimal allocation of arms $\boldsymbol{p}$ to effectively address this subquery. For a more detailed introduction to this general pure exploration model, please refer to Section EC.2. For our setting, we establish the below lower bound.

THEOREM 1 **(Lower Bound)**. *Consider a set of arms where arm i follows a normal distribution* $\mathcal{N}(\mu_i, 1)$*, where* $\mu_i = \boldsymbol{a}_i^\top \boldsymbol{\theta}$*. Any* $\delta$*-PAC algorithm for identifying all* $\varepsilon$*-best arms in the linear bandit setting must satisfy*

$$\inf_{Algo \in \mathcal{H}} \frac{\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta]}{\log\left(\frac{1}{2.4\delta}\right)} \geq (\Gamma^*_{\boldsymbol{\mu}})^{-1} = \min_{\boldsymbol{p} \in \mathcal{S}_K} \max_{(i,j,m) \in \mathcal{X}} \max \left\{ \frac{2\|\boldsymbol{a}_i - \boldsymbol{a}_j\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}}}{\left(\boldsymbol{a}_i^\top \boldsymbol{\theta} - \boldsymbol{a}_j^\top \boldsymbol{\theta} + \varepsilon\right)^2}, \frac{2\|\boldsymbol{a}_1 - \boldsymbol{a}_m\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}}}{\left(\boldsymbol{a}_1^\top \boldsymbol{\theta} - \boldsymbol{a}_m^\top \boldsymbol{\theta} - \varepsilon\right)^2} \right\}, \tag{19}$$

*where* $\mathcal{X} = \{(i,j,m) : i \in G_\varepsilon(\boldsymbol{\mu}), j \neq i, m \notin G_\varepsilon(\boldsymbol{\mu})\}$*,* $\mathcal{S}_K$ *is the* $K$*-dimensional probability simplex.*

The detailed proof of the above theorem is presented in Section EC.4.3. For the lower bound derivation, we assume normally distributed rewards to obtain a closed-form expression. A similar bound can be derived under sub-Gaussian rewards, though the form is less explicit.

REMARK 1 (**Generality of the Lower Bound**). We note that the stochastic multi-armed bandit problem is a special case of the linear bandit problem. By setting $\mathcal{A} = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_d\}$, where $\boldsymbol{e}_i$ denotes the unit vector, the linear bandit model reduces to a stochastic setting. This relationship allows us to recover the lower bound result for identifying all $\varepsilon$-best arms in stochastic bandits (Al Marjani et al. 2022). Furthermore, the lower bound in Theorem 1 extends the lower bound for best arm identification in linear bandits (Fiez et al. 2019). This result is recovered by setting $\varepsilon = 0$ and redefining the culprit set as $\mathcal{X}(\boldsymbol{\mu}) = \{i : i \in [K] \setminus i^*\}$, where $i^*$ represents the best arm in the context of best arms identification.

## 4. Algorithm and Upper Bound

In this section, we propose the *LinFACT* algorithm (**Lin**ear **F**ast **A**rm **C**lassification with **T**hreshold estimation) to identify all $\varepsilon$-best arms in linear bandits efficiently. We then establish upper bounds on the expected sample complexity to demonstrate the optimality of the LinFACT algorithm. Specifically, the upper bound derived from the $\mathcal{XY}$-optimal sampling policy is shown to be instance optimal up to logarithmic factors.[2]

### 4.1. Algorithm

LinFACT is a phase-based, semi-adaptive algorithm in which the sampling rule remains fixed within each round and is updated only at the end based on the accumulated observations. As the algorithm proceeds through round $r$, LinFACT progressively refines two sets of arms:

- $G_r$: Arms empirically classified as $\varepsilon$-best (good).
- $B_r$: Arms empirically classified as not $\varepsilon$-best (bad).

This classification process continues until all arms have been assigned to either $G_r$ or $B_r$. Once complete, the decision rule returns $G_r$ as the final set of $\varepsilon$-best arms.

**Sampling Rule.** To minimize the sampling budget, we select arms that provide the maximum information about the mean values or the gaps between them. Unlike stochastic multi-armed bandits, where mean values are obtained exclusively by sampling specific arms, linear bandits allow these mean values to be inferred from the estimated parameters. In each round, arms are selected based on the G-optimal design (13) or the $\mathcal{XY}$-optimal design (15).

For G-optimal design, LinFACT-G refines an estimate of the true parameter $\boldsymbol{\theta}$ and uses this estimate to maintain an anytime confidence interval, such that for each arm's empirical mean value $\hat{\mu}_i$, we have

$$\mathbb{P}\left(\bigcap_{i\in\mathcal{A}_I(r-1)}\bigcap_{r\in\mathbb{N}}|\hat{\mu}_i(r)-\mu_i|\le C_{\delta/K}(r)\right)\ge 1-\delta. \tag{20}$$

The active set $\mathcal{A}(r)$ is defined as the set of uneliminated arms, as we continuously eliminate arms as round $r$ progresses. $\mathcal{A}_I(r)$ denotes the set of indices corresponding to $\mathcal{A}(r)$.

This confidence bound indicates that the algorithm maintains a probabilistic guarantee that the true mean value $\mu_i$ is within a certain range of the estimated mean value $\hat{\mu}_i$ for each arm $i$, uniformly over all rounds. The bound shrinks as more data is collected (since the confidence radius $C_{\delta/K}(r)$ decreases with more samples), thereby reducing uncertainty. The anytime confidence width $C_{\delta/K}(r)$ is maintained by the design of the sample budget in each round. We set $C_{\delta/K}(r)=2^{-r}=:\varepsilon_r$, which is halved with each iteration of the rounds.

[2] We refer to the algorithms based on G-optimal sampling and $\mathcal{XY}$-optimal sampling as LinFACT-G and LinFACT-$\mathcal{XY}$, respectively. A detailed comparison between the two approaches is provided in Section EC.3.

In LinFACT-G, the initial budget allocation policy is based on the G-optimal design and is defined as follows

$$\begin{cases} T_r(\boldsymbol{a}) = \left\lceil \frac{2d\pi_r(\boldsymbol{a})}{\varepsilon_r^2} \log\left(\frac{2Kr(r+1)}{\delta}\right) \right\rceil, \\ T_r = \sum_{\boldsymbol{a}\in\mathcal{A}(r-1)} T_r(\boldsymbol{a}) \end{cases} \tag{21}$$

where $T_r$ denotes the total sampling budget allocated in round $r$, and $\pi_r$ is the selection probability distribution over the remaining active arms $\mathcal{A}(r-1)$ from the previous round, obtained via the G-optimal design as defined in equation (13). The sampling procedure for each round $r$ is described in Algorithm 1.

**Algorithm 1** Subroutine: G-Optimal Sampling

1: **Input:** Projected active set $\mathcal{A}(r-1)$, round $r$, $\delta$.
2: Obtain $\pi_r \in \mathcal{P}(\mathcal{A}(r-1))$ with support size $\mathrm{Supp}(\pi_r) \le \frac{d(d+1)}{2}$ according to equation (13).
3: **for all** $\boldsymbol{a} \in \mathcal{A}(r-1)$ **do** ▷ Sampling
4: Sample arm $\boldsymbol{a}$ for $T_r(\boldsymbol{a})$ times in round $r$, as specified in equation (21).

We also adopt the $\mathcal{XY}$-optimal design because the G-optimal design is less effective for distinguishing between arms. While the G-optimal design minimizes the maximum variance in estimating individual arm means, it does not explicitly focus on the pairwise gaps that are critical for identifying good arms. In contrast, the $\mathcal{XY}$-optimal design is tailored to directly reduce the uncertainty in estimating these gaps.

We now introduce a sampling rule based on the $\mathcal{XY}$-optimal design, as defined in equation (15). Let $q(\epsilon)$ denote the error introduced by the rounding procedure. We have:

$$\begin{cases} T_r = \max\left\{ \left\lceil \frac{2g_{\mathcal{XY}}\left(\mathcal{Y}(\mathcal{A}(r-1))\right)(1+\epsilon)}{\varepsilon_r^2} \log\left(\frac{2K(K-1)r(r+1)}{\delta}\right) \right\rceil, q(\epsilon) \right\}, \\ T_r(\boldsymbol{a}) = \textsc{Round}(\pi_r, T_r) \end{cases} \tag{22}$$

In contrast to the G-optimal design, the $\mathcal{XY}$-optimal design focuses on bounding the confidence region of the pairwise differences between arms. The following inequality characterizes the corresponding high-probability event.

$$\mathbb{P}\left( \bigcap_{i\in\mathcal{A}_I(r-1)} \bigcap_{j\in\mathcal{A}_I(r-1), j\neq i} \bigcap_{r\in\mathbb{N}} |(\hat{\mu}_j(r) - \hat{\mu}_i(r)) - (\mu_j - \mu_i)| \le 2C_{\delta/K}(r) \right) \ge 1-\delta. \tag{23}$$

The rounding operation, denoted as Round, uses a $(1+\epsilon)$ approximation algorithm proposed by Allen-Zhu et al. (2017). The complete sampling procedure is outlined in Algorithm 2.

**Algorithm 2** Subroutine: $\mathcal{XY}$-Optimal Sampling

1: **Input:** Projected active set $\mathcal{A}(r-1)$, round $r$, $\delta$.

2: Obtain $\pi_r \in \mathcal{P}(\mathcal{A}(r-1))$ according to equation (15).
3: **for all** $\boldsymbol{a} \in \mathcal{A}(r-1)$ **do** ▷ Sampling
4: Sample arm $\boldsymbol{a}$ for $T_r(\boldsymbol{a})$ times in round $r$, as specified in equation (22).

**Estimation.** At the end of each round, after drawing $T_r(\boldsymbol{a})$ samples from the active set, we compute the empirical estimate of the parameter using standard ordinary least squares (OLS)

$$\hat{\boldsymbol{\theta}}_r = \boldsymbol{V}_r^{-1} \sum_{s=1}^{T_r} \boldsymbol{a}_{A_s} X_s, \tag{24}$$

where $\boldsymbol{V}_r = \sum_{\boldsymbol{a} \in \mathcal{A}(r-1)} T_r(\boldsymbol{a}) \boldsymbol{a}\boldsymbol{a}^\top$ is the information matrix. The estimator for the mean value of each arm $i \in \mathcal{A}_I(r-1)$ is then

$$\hat{\mu}_i = \hat{\mu}_i(r) = \boldsymbol{a}_i^\top \hat{\boldsymbol{\theta}}_r. \tag{25}$$

**Stopping Rule and Decision Rule.** As round $r$ progresses, LinFACT dynamically updates two sets of arms: $G_r$ and $B_r$, representing arms that are empirically considered $\varepsilon$-best (good) and those that are not (bad), respectively. The algorithm filters arms by maintaining an upper confidence bound $U_r$ and a lower confidence bound $L_r$ around the unknown threshold $\mu_1 - \varepsilon$, along with individual upper and lower confidence bounds for each arm.

The stopping rule and final decision procedure are described in Algorithm 3. For each arm $i$ in the active set, LinFACT eliminates the arm if its upper confidence bound falls below the threshold $L_r$ (line 6). Conversely, if the lower confidence bound of an arm exceeds $U_r$ (line 8), the arm is added to the set $G_r$. Additionally, any arm already in $G_r$ will be removed from the active set if its upper bound falls below the empirically largest lower bound among all active arms (line 10). This ensures that the best arm is always retained in the active set, which is necessary for estimating the threshold $\mu_1 - \varepsilon$. The classification process continues until all arms are categorized, that is, when $G_r \cup B_r = [K]$. At termination, the set $G_r$ is returned as the output of LinFACT, representing the arms identified as $\varepsilon$-best.

**Algorithm 3** Subroutine: Stopping Rule and Decision Rule

1: **Input:** Projected active set $\mathcal{A}_I(r-1)$, estimator $(\hat{\mu}_i(r))_{i \in \mathcal{A}_I(r-1)}$, round $r$, $\varepsilon$, confidence radius $C_{\delta/K}(r)$.
2: Let $U_r = \max_{i \in \mathcal{A}_I(r-1)} \hat{\mu}_i + C_{\delta/K}(r) - \varepsilon$.
3: Let $L_r = \max_{i \in \mathcal{A}_I(r-1)} \hat{\mu}_i - C_{\delta/K}(r) - \varepsilon$.
4: **for all** $i \in \mathcal{A}_I(r-1)$ **do** ▷ Arm Classification and Elimination
5: **if** $\hat{\mu}_i + C_{\delta/K}(r) < L_r$ **then**
6: Add $i$ to $B_r$ and eliminate $i$ from $\mathcal{A}_I(r-1)$.

**if** $\hat{\mu}_i - C_{\delta/K}(r) > U_r$ **then**
  Add $i$ to $G_r$.
**if** $i \in G_r$ **and** $\hat{\mu}_i + C_{\delta/K}(r) \leq \max_{j \in \mathcal{A}_I(r-1)} \hat{\mu}_j - C_{\delta/K}(r)$ **then**
  Eliminate $i$ from $\mathcal{A}_I(r-1)$.
**if** $G_r \cup B_r = [K]$ **then** ▷ Stopping Condition and Recommendation
  **Output:** the set $G_r$.

**The Complete LinFACT Algorithm.** The complete LinFACT algorithm is presented in Algorithm 4. The procedure proceeds as follows: based on the collected data, the decision-maker updates the parameter estimates and checks whether the stopping condition $\tau_\delta$ is satisfied. If so, the set $G_r$ is returned as the estimated set of all $\varepsilon$-best arms. If the stopping condition is not met, the process continues with further sampling and updates.

**Algorithm 4** LinFACT Algorithm

**Input:** $\varepsilon$, $\delta$, bandit instance.
Initialize $G_0 = \emptyset$, the set of good arms, and $B_0 = \emptyset$, the set of bad arms.
Initialize the active set $\mathcal{A}(0) = \mathcal{A}$, and $\mathcal{A}_I(0) = [K]$.
**for** $r = 1, 2, \ldots$ **do**
  Set $C_{\delta/K}(r) = \varepsilon_r = 2^{-r}$.
  Set $G_r = G_{r-1}$ and $B_r = B_{r-1}$.
  Project $\mathcal{A}(r-1)$ to a $d_r$-dimensional subspace that $\mathcal{A}(r-1)$ spans. ▷ Projection
  **if** Using G-optimal Sampling **then** ▷ Sampling
    Call Algorithm 1.
  **else if** Using $\mathcal{XY}$-optimal Sampling **then**
    Call Algorithm 2.
  Estimate $(\hat{\mu}_i(r))_{i \in \mathcal{A}_I(r-1)}$ using equations (24) and (25). ▷ Estimation
  Call Algorithm 3. ▷ Stopping Condition and Decision Rule

### 4.2. Upper Bounds of the LinFACT Algorithm

Theorems 2 and 3 establish upper bounds on the sample complexity of the proposed LinFACT algorithm.

Let $T_G$ and $T_{\mathcal{XY}}$ denote the number of samples required under the G-optimal and $\mathcal{XY}$-optimal designs, respectively. The formal statements of these theorems are given below.

THEOREM 2 **(Upper Bounds, G-Optimal Design)**. *For* $\xi = \min(\alpha_\varepsilon, \beta_\varepsilon)/16$*, there exists an event* $\mathcal{E}$ *such that* $\mathbb{P}(\mathcal{E}) \geq 1-\delta$*. On this event, the* LinFACT *algorithm with the G-optimal sampling policy achieves an expected sample complexity upper bound given by*

$$\mathbb{E}[T_G \mid \mathcal{E}] = \mathcal{O}\left(d\xi^{-2}\log\left(\frac{K}{\delta}\log_2(\xi^{-2})\right) + d^2\log(\xi^{-1})\right). \tag{26}$$

The detailed proof of Theorem 2 is presented in Section EC.5. However, the LinFACT algorithm based on the G-optimal design does not yield an upper bound that aligns with the lower bound. This limitation is discussed further in Section EC.6. In contrast, we will show that the algorithm using the $\mathcal{XY}$-optimal design achieves an upper bound that matches the lower bound up to a logarithmic factor.

THEOREM 3 **(Upper Bound, $\mathcal{XY}$-Optimal Design)**. *Assume that an instance of arms satisfies* $\min_{i\in G_\varepsilon\setminus\{1\}} \|\boldsymbol{a}_1 - \boldsymbol{a}_i\|^2 \geq L_2$ *and* $\max_{i\in[K]} |\mu_1 - \varepsilon - \mu_i| \leq 2$*. There exists an event* $\mathcal{E}$ *such that* $\mathbb{P}(\mathcal{E}) \geq 1-\delta$*. On this event, the* LinFACT *algorithm with the* $\mathcal{XY}$*-optimal sampling policy achieves an expected sample complexity upper bound given by*

$$\mathbb{E}[T_{\mathcal{XY}} \mid \mathcal{E}] = \mathcal{O}\left((\Gamma^*)^{-1}d\xi^{-1}\log(\xi^{-1})\log\left(\frac{K}{\delta}\log(\xi^{-2})\right) + \frac{d}{\epsilon^2}\log(\xi^{-1})\right), \tag{27}$$

*where* $\xi = \min(\alpha_\varepsilon, \beta_\varepsilon)/16$ *is the minimum gap of the problem instance,* $R_{\text{upper}} = \max\left\{\left\lceil\log_2\frac{4}{\alpha_\varepsilon}\right\rceil, \left\lceil\log_2\frac{4}{\beta_\varepsilon}\right\rceil\right\}$*, and* $(\Gamma^*)^{-1}$ *is the lower bound term defined in Theorem 1.*

The proof of the near-optimal upper bound in Theorem 3 is presented in Section EC.7 of the online appendix, where we make it clear how the lower bound helps to establish our upper bound.

## 5. Model Misspecification

In this section, we address the challenge of model misspecification, recognizing that real-world problems may deviate from perfect linearity. To account for such deviations, we propose an orthogonal parameterization-based algorithm, *i.e.*, LinFACT-MIS[3], a refined version of LinFACT to address model misspecification. We establish new upper bounds in the misspecified setting and provide insights into how such deviations impact algorithm performance.

Under model misspecification, we refine the linear model in equation (3) as

$$X_t = \boldsymbol{a}_{A_t}^\top\boldsymbol{\theta} + \eta_t + \Delta_m(\boldsymbol{a}_{A_t}), \tag{28}$$

where $\Delta_m : \mathbb{R}^d \to \mathbb{R}$ is a misspecification function quantifying the deviation from the true model.

[3] LinFACT-MIS builds upon LinFACT-G. We choose to extend LinFACT-G, rather than LinFACT-$\mathcal{XY}$, for reasons of analytical tractability and computational efficiency. Both variants exhibit comparable performance in terms of upper bounds in this setting, while G-optimal design offers a significantly lower computational burden.

ASSUMPTION 3. *Assume* $\|\boldsymbol{\mu}\|_\infty \le L_\infty$ *and* $\|\boldsymbol{\Delta}_m\|_\infty \le L_m$*, where* $\|\cdot\|_\infty$ *denotes the infinity norm and the bold* $K$*-dimensional vector* $\boldsymbol{\Delta}_m$ *represents the bias term of the misspecified model.*

Therefore, with this assumption, the set of realizable models is defined as

$$M := \left\{ \boldsymbol{\mu} \in \mathbb{R}^K \,\middle|\, \exists \boldsymbol{\theta} \in \mathbb{R}^d, \exists \boldsymbol{\Delta}_m \in \mathbb{R}^K, \boldsymbol{\mu} = \boldsymbol{\Psi}\boldsymbol{\theta} + \boldsymbol{\Delta}_m \wedge \|\boldsymbol{\mu}\|_\infty \le L_\infty \wedge \|\boldsymbol{\Delta}_m\|_\infty \le L_m \right\}. \tag{29}$$

The key distinction in the analysis under model misspecification lies in how the estimator $\hat{\boldsymbol{\mu}}_t$ is maintained. Specifically, we construct this estimator by projecting the empirical mean vector $\tilde{\boldsymbol{\mu}}_t$ at time $t$ onto the set of realizable models $M$ via the following optimization

$$\hat{\boldsymbol{\mu}}_t := \arg\min_{\boldsymbol{\vartheta} \in M} \|\boldsymbol{\vartheta} - \tilde{\boldsymbol{\mu}}_t\|^2_{\boldsymbol{D}_{\boldsymbol{N}_t}}, \tag{30}$$

where $\boldsymbol{N}_t = [N_{t1}, N_{t2}, \ldots, N_{tK}]^\top \in \mathbb{R}^K$ is the vector of sample counts for each arm at time $t$, and $\boldsymbol{D}_{\boldsymbol{N}_t} \in \mathbb{R}^{K \times K}$ is the diagonal matrix with $N_{t1}, N_{t2}, \ldots, N_{tK}$ as its diagonal entries.

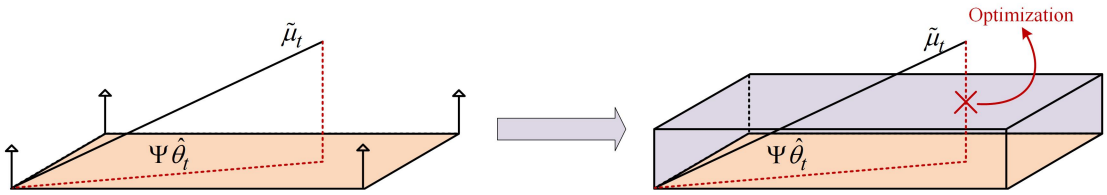



**Figure 2** Difference Between Standard OLS and Misspecification-Adjusted Projection Estimates

*Note. The left diagram shows the projection onto the span of pulled arms under a perfect linear model. The right diagram depicts the adjustment required under misspecification, where the projection must account for the deviation.*

In the absence of misspecification, this projection simplifies to the ordinary least squares (OLS) estimator. However, as shown in Figure 2, under model misspecification, the estimator can no longer be computed as a simple projection onto a hyperplane and falls outside the scope of standard OLS. Instead, it must be formulated as the optimization problem in equation (30), which minimizes a weighted quadratic objective over $K + d$ variables, subject to the constraints $\|\boldsymbol{\mu}\|_\infty \le L_\infty$ and $\|\boldsymbol{\Delta}_m\|_\infty \le L_m$.

### 5.1. Upper Bound with Misspecification

In this subsection, we present the upper bound on the expected sample complexity for LinFACT-MIS. This analysis highlights the influence of misspecification on the theoretical performance of the algorithm. Let $T_{G_{\text{mis}}}$ denote the number of samples taken under model misspecification.

THEOREM 4 **(Upper Bound, Misspecification)**. *Fix* $\varepsilon > 0$ *and suppose that the magnitude of misspecification satisfies* $L_m < \min\left\{\frac{\alpha_\varepsilon}{2\sqrt{d}}, \frac{\beta_\varepsilon}{2\sqrt{d}}\right\}$*. For* $\xi = \min\left(\alpha_\varepsilon - 2L_m\sqrt{d}, \beta_\varepsilon - 2L_m\sqrt{d}\right)/16$*, there*

*exists an event $\mathcal{E}$ such that $\mathbb{P}(\mathcal{E}) \geq 1-\delta$. On this event, LinFACT-MIS terminates and returns the correct solution with an expected sample complexity upper bound given by*

$$\mathbb{E}_{\boldsymbol{\mu}}\left[T_{G_{\text{mis}}} \mid \mathcal{E}\right] = \mathcal{O}\left(d\xi^{-2}\log\left(\frac{K}{\delta}\log(\xi^{-2})\right) + d^2\log(\xi^{-1})\right). \tag{31}$$

The proof of this theorem is provided in Section EC.8. The upper bound in Theorem 2, which assumes no model misspecification, can be viewed as a special case of Theorem 4 by setting $L_m = 0$. However, the bound in Theorem 4 becomes invalid when the misspecification magnitude $L_m$ is too large, as the logarithmic terms may involve negative arguments, violating the assumptions required for the bound to hold. If the variation in confidence radius across arms due to misspecification is not accounted for, Theorem 4 suggests that the sample complexity will increase. In particular, compared to the bound in Theorem 2, this result is looser because its denominator terms about $\varepsilon$ decrease from $\alpha_\varepsilon$ and $\beta_\varepsilon$ to $\alpha_\varepsilon - 2L_m\sqrt{d}$ and $\beta_\varepsilon - 2L_m\sqrt{d}$, respectively.

## 5.2. Orthogonal Parameterization

In this section, we present an alternative version of LinFACT-MIS based on orthogonal parameterization, designed to improve computational efficiency.

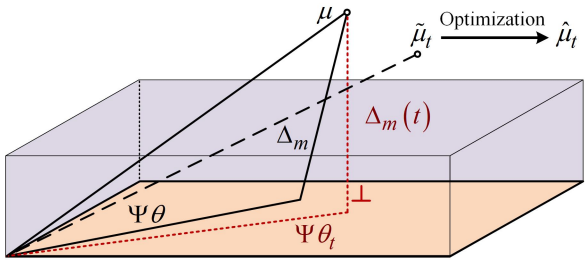



**Figure 3** Orthogonal Parameterization and Projection.

*Note. The estimator $\hat{\boldsymbol{\mu}}_t$ is obtained from the empirical mean $\tilde{\boldsymbol{\mu}}_t$ by solving an optimization problem. While the true mean vector $\boldsymbol{\mu}$ can be expressed as the sum of a linear component $\boldsymbol{\Psi\theta}$ and a non-linear model deviation $\boldsymbol{\Delta}_m$, it can also be decomposed at each time step $t$ via orthogonal projection into a linear part $\boldsymbol{\Psi\theta}_t$ on the hyperplane and a residual term $\boldsymbol{\Delta}_m(t)$ orthogonal to it.*

**Orthogonal Parameterization.** Under model misspecification, traditional confidence bounds for the mean estimator based on $\|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}\|^2_{\boldsymbol{V}_t}$, derived using either martingale-based methods (Abbasi-Yadkori et al. 2011) or covering arguments (Lattimore and Szepesvári 2020), are no longer directly applicable due to the presence of an additional misspecification term. To improve the concentration of the estimator in this setting, a key strategy is to adopt an orthogonal parameterization of the mean vectors within the realizable model $M$ (Réda et al. 2021b).

Rather than centering the confidence region around the true parameter $\boldsymbol{\theta}$, we focus on the quantity $\|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_t\|^2_{\boldsymbol{V}_t}$, where $\boldsymbol{\theta}_t$ is the orthogonal projection of the true mean vector onto the feature

space spanned by the pulled arms at time $t$. This $\boldsymbol{\theta}_t$-centered form corresponds to a self-normalized martingale and thus satisfies the same concentration bounds as in the classical linear bandit setting without misspecification. This approach offers an advantage over prior methods (Lattimore et al. 2020, Zanette et al. 2020), which require inflating the confidence radius between $\hat{\boldsymbol{\theta}}_t$ and $\boldsymbol{\theta}$ by a factor of $L_m^2 t$, leading to overly conservative bounds in misspecified settings where $L_m \gg 0$.

Specifically, we show that any mean vector $\boldsymbol{\mu} = \boldsymbol{\Psi}\boldsymbol{\theta} + \boldsymbol{\Delta}_m$ can be equivalently expressed at any time $t$ as $\boldsymbol{\mu} = \boldsymbol{\Psi}\boldsymbol{\theta}_t + \boldsymbol{\Delta}_{\boldsymbol{m}}(\boldsymbol{t})$, where

$$\boldsymbol{\theta}_t = \left(\boldsymbol{\Psi}_{N_t}^\top \boldsymbol{\Psi}_{N_t}\right)^{-1} \boldsymbol{\Psi}_{N_t}^\top \boldsymbol{D}_{N_t}^{1/2} \boldsymbol{\mu} = \boldsymbol{V}_t^{-1} \sum_{s=1}^{t} \mu_{A_s} \boldsymbol{a}_{A_s} \tag{32}$$

is the orthogonal projection of $\boldsymbol{\mu}$ onto the feature space spanned by the columns of $\boldsymbol{\Psi}_{N_t}$, and $\boldsymbol{\Delta}_{\boldsymbol{m}}(\boldsymbol{t}) = \boldsymbol{\mu} - \boldsymbol{\Psi}\boldsymbol{\theta}_t$ is the residual. Here, $\boldsymbol{\Psi}_{N_t} = \boldsymbol{D}_{N_t}^{1/2}\boldsymbol{\Psi}$ is the matrix of feature vectors weighted by the number of times each arm has been sampled up to time $t$, and $\boldsymbol{D}_{N_t}^{1/2}$ is a diagonal matrix with entries corresponding to the square roots of the number of samples for each arm.

**Upper Bound.** For LinFACT-MIS, the orthogonal parameterization involves updating the sampling rule in Algorithm 1 to the sampling rule in Algorithm 5. The estimation is no longer based on OLS but is achieved by calculating the estimator $(\hat{\mu}_i(r))_{i\in\mathcal{A}_I(r-1)}$ with observed data by solving the optimization problem described in equation (30) directly.

When model misspecification is accounted for and orthogonal parameterization is applied, the sampling policy is given by:

$$\begin{cases} T_r(\boldsymbol{a}) = \left\lceil \dfrac{8d\pi_r(\boldsymbol{a})}{\varepsilon_r^2} \left( d\log(6) + \log\left(\dfrac{Kr(r+1)}{\delta}\right)\right)\right\rceil \\ T_r = \sum_{\boldsymbol{a}\in\mathcal{A}(r-1)} T_r(\boldsymbol{a}) \end{cases}. \tag{33}$$

Let $T_{G_{\text{op}}}$ denote the total number of samples required when orthogonal parameterization is used. The corresponding upper bound is stated in the theorem below.

**Algorithm 5** Subroutine: Sampling With Orthogonal Parameterization

1: **Input:** Projected active set $\mathcal{A}_I(r-1)$, round $r$, $\delta$.
2: Find the G-optimal design $\pi_r \in \mathcal{P}(\mathcal{A}(r-1))$ with $\text{Supp}(\pi_r) \le \frac{d(d+1)}{2}$ according to equation (13).
3: **for all** $i \in \mathcal{A}_I(r-1)$ **do** ▷ Sampling
4: Sample arm $\boldsymbol{a}$ for $T_r(\boldsymbol{a})$ times in round $r$, as specified in equation (33).

THEOREM 5 **(Upper Bound, Orthogonal Parameterization)**. *Fix* $\varepsilon > 0$ *and suppose that the magnitude of misspecification satisfies* $L_m < \min\left\{\frac{\alpha_\varepsilon}{2(\sqrt{d}+2)}, \frac{\beta_\varepsilon}{2(\sqrt{d}+2)}\right\}$. *For* $\xi = \min\left(\alpha_\varepsilon - 2L_m(\sqrt{d}+2), \beta_\varepsilon - 2L_m(\sqrt{d}+2)\right)/16$, *there exists an event* $\mathcal{E}$ *such that* $\mathbb{P}(\mathcal{E}) \ge 1-\delta$.

*On this event, LinFACT-MIS terminates and returns the correct solution with an expected sample complexity upper bound given by*

$$\mathbb{E}_{\boldsymbol{\mu}}\left[T_{G_{\mathrm{op}}} \mid \mathcal{E}\right] = \mathcal{O}\left(d\xi^{-2}\log\left(\frac{K6^d}{\delta}\log(\xi^{-1})\right) + d^2\log(\xi^{-1})\right). \tag{34}$$

This theorem establishes that the upper bound remains of the same order as in Theorem 4, with the detailed proof provided in Section EC.9. As in Theorem 4, the use of orthogonal parameterization does not eliminate the expansion of the upper bound, which remains unavoidable. We provide further intuition in Section 5.3, arguing that without prior knowledge of the misspecification, it is not possible to recover full performance through algorithmic refinement alone.

### 5.3. Insights for Model Misspecification

*Lower Bounds in Linear and Stochastic Settings.* The lower bound becomes equivalent to the unstructured lower bound as soon as the misspecification upper bound $L_m > L_{\boldsymbol{\mu}}$, where $L_{\boldsymbol{\mu}}$ is an instance-dependent finite constant. This observation is formalized in Proposition 3, whose proof follows the same logic as Lemma 2 in Réda et al. (2021b).

Proposition 3. *There exists $L_{\boldsymbol{\mu}} \in \mathbb{R}$ with $L_{\boldsymbol{\mu}} \leq \max_k \mu_k - \min_k \mu_k$ such that if $L_m > L_{\boldsymbol{\mu}}$, then for any pure exploration task, the lower bound in the linear setting is equal to the unstructured lower bound.*

*Improvement with Unknown Misspecification is Not Possible.* Knowing that a problem is misspecified without access to an upper bound $L_m$ on $\|\boldsymbol{\Delta}_m\|_\infty$ is effectively equivalent to having no structural knowledge of the problem. As a result, improving algorithmic performance under such unknown model misspecification is infeasible. In particular, as shown in cumulative regret settings, sublinear regret guarantees are no longer achievable (Ghosh et al. 2017, Lattimore et al. 2020); similarly, in pure exploration, the theoretical lower bound cannot be attained.

*Prior Knowledge for the Misspecification.* When the upper bound $L_m$ on model misspecification is known in advance, LinFACT-MIS can be modified to account for this deviation. Specifically, we adjust the confidence radius $C_{\delta/K}(r)$ used in computing the lower and upper bounds (*i.e.*, $L_r$ and $U_r$) in Algorithm 4. With this modification, the number of rounds required to complete classification under misspecification, $R'_{\mathrm{upper}}$ and $R''_{\mathrm{upper}}$, coincides with $R_{\mathrm{upper}}$, the corresponding number of rounds under a perfectly linear model. This is achieved by replacing the confidence radius $\varepsilon_r$ with an inflated version $\varepsilon_r + L_m\sqrt{d}$ to compensate for the worst-case deviation due to misspecification. This adjustment preserves the validity of the original analysis and ensures that the same theoretical guarantees are retained. The following proposition formalizes this observation and is proved in Section EC.10.

PROPOSITION 4. *Suppose the misspecification magnitude $L_m > 0$ is known in advance. Adjusting the confidence radius $C_{\delta/K}(r)$ in Algorithm 4 at each round $r$ from its original value $\varepsilon_r$ to*

$$\varepsilon_r' = \varepsilon_r + L_m\sqrt{d}, \tag{35}$$

*ensures that the total number of rounds required under misspecification matches that under perfect linearity.*

## 6. Generalized Linear Model

In this section, we extend the linear bandits to a generalized linear model (GLM). In this setting, the reward function no longer follows the standard linear form in equation (3), but instead satisfies

$$\mathbb{E}\left[X_t \mid A_t\right] = \mu_{\text{link}}(\boldsymbol{a}_{A_t}^\top \boldsymbol{\theta}), \tag{36}$$

where $\mu_{\text{link}} : \mathbb{R} \to \mathbb{R}$ is the inverse link function. GLMs encompass a class of models that include, but are not limited to, linear models, allowing for various reward distributions beyond the Gaussian. For example, for binary-valued rewards, a suitable choice of $\mu_{\text{link}}$ is $\mu_{\text{link}}(x) = \exp(x)/(1+\exp(x))$, *i.e.*, sigmoid function, leading to the logistic regression model. For integer-valued rewards, $\mu_{\text{link}}(x) = \exp(x)$ leads to the Poisson regression model.

To keep this paper self-contained, we briefly review the main properties of GLMs (McCullagh 2019). A univariate probability distribution belongs to a canonical exponential family if its density with respect to a reference measure is given by

$$p_\omega(x) = \exp\left(x\omega - b(\omega) + c(x)\right), \tag{37}$$

where $\omega$ is a real parameter, $c(\cdot)$ is a real normalization function, and $b(\cdot)$ is assumed to be twice continuously differentiable. This family includes the Gaussian and Gamma distributions when the reference measure is the Lebesgue measure, and the Poisson and Bernoulli distributions when the reference measure is the counting measure on the integers. For a random variable $X$ with density defined in (37), $\mathbb{E}(X) = \dot{b}(\omega)$ and $\text{Var}(X) = \ddot{b}(\omega)$, where $\dot{b}$ and $\ddot{b}$ denote the first and second derivatives of $b$, respectively. Since the variance is always positive and $\mu_{\text{link}} = \dot{b}$ represents the inverse link function, $b$ is strictly convex and $\mu_{\text{link}}$ is increasing.

The canonical GLM assumes that $p_{\boldsymbol{\theta}}(X \mid \boldsymbol{a}_i) = p_{\boldsymbol{a}_i^\top \boldsymbol{\theta}}(X)$ for all arms $i$. The maximum likelihood estimator $\hat{\boldsymbol{\theta}}_t$, based on $\sigma$-algebra $\mathcal{F}_t = \sigma\left(A_1, X_1, A_2, X_2, \ldots, A_t, X_t\right)$, is defined as the maximizer of the function

$$\sum_{s=1}^{t} \log p_{\boldsymbol{\theta}}(X_s|\boldsymbol{a}_{A_s}) = \sum_{s=1}^{t} X_s \boldsymbol{a}_{A_s}^\top \boldsymbol{\theta} - b(\boldsymbol{a}_{A_s}^\top \boldsymbol{\theta}) + c(X_s). \tag{38}$$

This function is strictly concave in $\boldsymbol{\theta}$. By differentiating, we obtain that $\hat{\boldsymbol{\theta}}_t$ is the unique solution of the following estimating equation at time $t$,

$$\sum_{s=1}^{t} \left(X_s - \mu_{\text{link}}(\boldsymbol{a}_{A_s}^\top \boldsymbol{\theta})\right) \boldsymbol{a}_{A_s} = \mathbf{0}. \tag{39}$$

In practice, while the solution to equation (39) does not have a closed-form solution, it can be efficiently found using methods such as iteratively reweighted least squares (IRLS) (Wolke and Schwetlick 1988), which employs Newton's method. Here, $\check{\boldsymbol{\theta}}_t$ is a convex combination of $\boldsymbol{\theta}$ and its maximum likelihood estimate $\hat{\boldsymbol{\theta}}_t$ at time $t$. The existence of $c_{\min}$ can be ensured by performing forced exploration at the beginning of the algorithm, incurring a sampling cost of $O(d)$ (Kveton et al. 2023).

ASSUMPTION 4. *The derivative of the inverse link function, $\mu_{link}$, is bounded,* i.e., $c_{\min} \leq \dot{\mu}_{link}(\boldsymbol{a}^\top \check{\boldsymbol{\theta}}_t)$, *for some $c_{\min} \in \mathbb{R}^+$ and all arms.*

Assumption 4 is standard in the GLM literature (Li et al. 2017, Azizi et al. 2021), ensuring that the reward function is sufficiently smooth, with $c_{\min} > 0$ typically determined by the choice of link function.

### 6.1. Algorithm with GLM

In this section, we present a refined algorithm for the generalized linear model, referred to as LinFACT-GLM. This refinement involves modifying the sampling rule in Algorithm 1 and adjusting the estimation method. The designed sampling policy is described by

$$\begin{cases} T_r(\boldsymbol{a}) = \left\lceil \dfrac{2d\pi_r(\boldsymbol{a})}{\varepsilon_r^2 c_{\min}^2} \log\left(\dfrac{2Kr(r+1)}{\delta}\right) \right\rceil, \\ T_r = \sum_{\boldsymbol{a} \in \mathcal{A}(r-1)} T_r(\boldsymbol{a}) \end{cases} \tag{40}$$

where $c_{\min}$ is the known constant controlling the first-order derivative of the inverse link function.

**Algorithm 6** Subroutine: G-Optimal Sampling with GLM

1: **Input:** Projected active set $\mathcal{A}_I(r-1)$, round $r$, $\delta$.
2: Find the G-optimal design $\pi_r \in \mathcal{P}(\mathcal{A}(r-1))$ with support size $\text{Supp}(\pi_r) \leq \frac{d(d+1)}{2}$ according to equation (13).
3: **for all** $\boldsymbol{a} \in \mathcal{A}_I(r-1)$ **do** ▷ Sampling
4: Sample arm $\boldsymbol{a}$ for $T_r(\boldsymbol{a})$ times in round $r$, as specified in equation (40).

In the GLM setting, Ordinary Least Squares (OLS) is also not applicable. Instead, the estimator for each $i \in \mathcal{A}_I(r-1)$ is obtained by solving the optimization problem described in equation (38), using the estimating equation in (39) derived from the observed data.

### 6.2. Upper Bound for the GLM-Based LinFACT

Let $T_{G_{\mathrm{GLM}}}$ denote the number of samples collected under the GLM setting. The following theorem provides an upper bound on the expected sample complexity of LinFACT-GLM.

THEOREM 6 **(Upper Bound, Generalized Linear Model)**. *For* $\xi = \min(\alpha_\varepsilon, \beta_\varepsilon)/16$, *there exists an event* $\mathcal{E}$ *such that* $\mathbb{P}(\mathcal{E}) \geq 1-\delta$. *On this event, the* LinFACT-GLM *algorithm achieves an expected sample complexity upper bound given by*

$$\mathbb{E}[T_{\mathrm{GLM}} \mid \mathcal{E}] = \mathcal{O}\left(\frac{d}{c_{\min}^2}\xi^{-2}\log\left(\frac{K}{\delta}\log_2(\xi^{-2})\right) + d^2\log(\xi^{-1})\right). \tag{41}$$

The upper bound presented in Theorem 6, which generalizes the model to the GLM setting, can be viewed as an extension of Theorem 2. The detailed proof is provided in Section EC.11 of the online appendix.

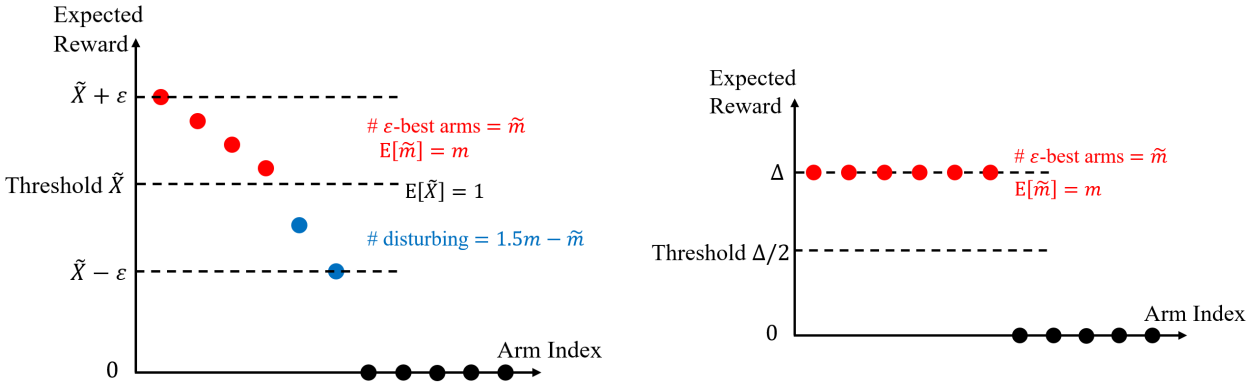
## 7. Numerical Experiments

In the numerical experiments, we compare our algorithm, LinFACT, with several baseline methods. These include the Bayesian optimization algorithm based on the knowledge-gradient acquisition function with correlated beliefs for best arm identification (KGCB) proposed by Negoescu et al. (2011); the gap-based algorithm for best arm identification (BayesGap) introduced by Hoffman et al. (2013); the track-and-stop algorithm for threshold bandits (Lazy TTS) developed by Rivera and Tewari (2024); and two gap-based algorithms for top $m$ arm identification, LinGIFA and $m$-LinGapE, presented by Réda et al. (2021a), which represent the state-of-the-art for returning multiple candidates.

Identifying all $\varepsilon$-best arms is often more challenging than identifying the top $m$ arms or arms above a given threshold. To address this, we adopt a random setting where both the number of $\varepsilon$-best arms and the $\varepsilon$-threshold are randomly sampled. In this setting, top $m$ algorithms and threshold bandit algorithms only have access to the expected reward values, ensuring a fair comparison. Since BayesGap and KGCB operate under a fixed-budget setting, we use the average sample complexity of LinFACT as the budget for comparison and evaluate their performance accordingly. For BAI algorithms, we select arms whose empirical means are within $\varepsilon$ of the empirical best arm once the budget is exhausted.

### 7.1. Synthetic Experiments

Following Soare et al. (2014), Xu et al. (2018), and Azizi et al. (2021), we categorize synthetic data into two types: adaptive and static settings. Figure 4 illustrates the construction of these synthetic datasets, with detailed configurations provided in Section EC.12. A summary of all settings is presented in Table 1.




(a) Synthetic I - Adaptive Setting (b) Synthetic II - Static Setting

**Figure 4** Illustration of the Synthetic Experiment Settings

*Note. In the adaptive setting (a), we randomly sample the threshold $\tilde{X}$ from a distribution with mean 1, and independently sample the number of $\varepsilon$-best arms $\tilde{m}$ from a distribution with mean $m$. The arms above and below the threshold are then uniformly drawn from the intervals $[\tilde{X}, \tilde{X}+\varepsilon]$ and $[\tilde{X}-\varepsilon, \tilde{X})$, respectively. In the static setting (b), we fix the threshold at $\Delta/2$ and randomly sample $\tilde{m}$ $\varepsilon$-best arms with reward $\Delta$.*

In the *adaptive setting*, arms are divided into three categories: (1) the arms to be selected (*i.e.*, the all $\varepsilon$-best arms), (2) disturbing arms that are slightly worse, and (3) base arms with zero rewards. The primary challenge for algorithms is to distinguish between the arms in categories (1) and (2), while the base arms in category (3) can be ignored. Adaptive algorithms that effectively leverage shared information to explore similar arms perform well in this setting.

In the *static setting*, arms are divided into two categories: the all $\varepsilon$-best arms and the base arms with zero rewards. In this case, algorithms must distinguish between all arms. Static algorithms that uniformly explore all arms are well-suited for this setting.

**Table 1** Synthetic Experiment Settings

| Setting Index | Setting Category | Setting Details |
|---|---|---|
| 1 | Adaptive | $(d, \mathbb{E}[m]) = (8, 4)$, $\varepsilon = 0.1$ |
| 2 | Adaptive | $(d, \mathbb{E}[m]) = (8, 4)$, $\varepsilon = 0.2$ |
| 3 | Adaptive | $(d, \mathbb{E}[m]) = (8, 4)$, $\varepsilon = 0.3$ |
| 4 | Adaptive | $(d, \mathbb{E}[m]) = (12, 4)$, $\varepsilon = 0.1$ |
| 5 | Adaptive | $(d, \mathbb{E}[m]) = (12, 4)$, $\varepsilon = 0.2$ |
| 6 | Adaptive | $(d, \mathbb{E}[m]) = (12, 4)$, $\varepsilon = 0.3$ |
| 7 | Static | $(d, \Delta) = (8, 1)$ |
| 8 | Static | $(d, \Delta) = (12, 1)$ |
| 9 | Static | $(d, \Delta) = (16, 1)$ |

Remark 2. A common misconception is that adaptive algorithms universally outperform static ones. While adaptive algorithms are typically efficient at focusing exploration on promising arms in many settings, they can be less effective in static environments. In such cases, adaptive methods

may inefficiently allocate samples between both candidate and baseline arms, leading to redundant exploration. In contrast, static algorithms can achieve the objective more efficiently by uniformly allocating samples across all arms, avoiding bias and over-exploration.

**Experiment Setup.** We benchmark our algorithms, LinFACT-G and LinFACT-$\mathcal{XY}$, against BayesGap, KGCB, LinGIFA, m-LinGapE, and Lazy TTS, focusing on both sample complexity and the $F1$ score.

We conduct each experiment using different data types (adaptive or static), arm dimensions ($d$), and numbers of arms ($K$). For each configuration, we generate 10 values of $m$ from a normal distribution centered at the expected value $\mathbb{E}[m]$ with a variance of 3.0, where $m$ is the input for a top-$m$ algorithm. For each sampled pair $(\tilde{m}, \tilde{X})$, where $\tilde{m}$ denotes the number of $\varepsilon$-best arms and $\tilde{X}$ represents the value of the best arm minus $\varepsilon$, we repeat the experiment 100 times. We then compute the average $F1$ score across the 10 $(\tilde{m}, \tilde{X})$ pairs, resulting in 1,000 total executions per algorithm.

In practice, we observe that KGCB, LinGIFA, m-LinGapE, and Lazy TTS are computationally intensive when the sampling budget is high. The time-consuming nature of KGCB has already been noted in the literature (Negoescu et al. 2011). For Lazy TTS, the algorithm requires repeatedly evaluating an objective function within an optimization problem, where each evaluation has a time complexity of $O(Kd^2 + d^3)$. As the optimization process involves a non-negligible number of iterations ($n$), the total time complexity becomes $O(nKd^2)$, making the algorithm inefficient.

For the two top-$m$ algorithms, the computational burden arises from performing matrix inversions for all arms, leading to a total time complexity of $O(Kd^3)$. In contrast, LinFACT achieves a significantly lower total time complexity of $O(Kd^2)$, as the optimization problem within our algorithm can be efficiently solved using a fixed-step gradient descent method. Table 2 presents the runtime for synthetic data, demonstrating that our algorithm is at least five times faster than all other methods. Notably, this performance gap is even more pronounced when using real data.

**Experiment Results.** Our experimental results are presented in Figures 5 and 6. In Figure 5, the vertical axis denotes the $F1$ score, with higher values indicating better algorithm performance. The first row of six plots shows the results under adaptive settings. As $\varepsilon$ increases, the non-optimal (disturbing) arms in the adaptive setting move progressively farther from the optimal arms, making them easier to distinguish. Consequently, the $F1$ score increases from left to right. An exception is BayesGap, which performs best when $\varepsilon = 0.2$. This occurs because best-arm identification algorithms, such as BayesGap, struggle to differentiate optimal arms from disturbing ones when they are close ($\varepsilon = 0.1$) and fail to fully explore the optimal arms when they are not closely clustered with the best arm ($\varepsilon = 0.3$).

**Table 2** Running Time (seconds) for Different Synthetic Experiments Among Algorithms

| Algorithm Settings | Adaptive Settings | | | | | | Static Settings | | |
|---|---|---|---|---|---|---|---|---|---|
| | $(d, \mathbb{E}[m]) = (8, 4)$ | | | $(d, \mathbb{E}[m]) = (12, 4)$ | | | $\Delta = 1$ | | |
| | $\varepsilon = 0.1$ | $\varepsilon = 0.2$ | $\varepsilon = 0.3$ | $\varepsilon = 0.1$ | $\varepsilon = 0.2$ | $\varepsilon = 0.3$ | $d = 8$ | $d = 12$ | $d = 16$ |
| LinFACT-G | **0.028** | **0.030** | **0.028** | **0.078** | **0.078** | **0.076** | **0.005** | **0.007** | **0.009** |
| LinFACT-$\mathcal{XY}$ | <u>0.037</u> | <u>0.038</u> | <u>0.038</u> | <u>0.163</u> | <u>0.155</u> | <u>0.145</u> | <u>0.015</u> | <u>0.028</u> | <u>0.049</u> |
| BayesGap | 0.112 | 0.110 | 0.110 | 0.306 | 0.304 | 0.307 | 0.036 | 0.071 | 0.112 |
| LinGIFA | 0.313 | 0.265 | 0.236 | 1.943 | 1.484 | 1.424 | 0.160 | 0.529 | 1.297 |
| m-LinGapE | 0.195 | 0.166 | 0.157 | 1.362 | 1.053 | 0.984 | 0.116 | 0.352 | 0.932 |
| Lazy TTS | 0.991 | 3.904 | 12.512 | 24.583 | 26.941 | 25.622 | 0.198 | 0.871 | 2.164 |
| KGCB | 1.990 | 1.936 | 1.670 | 6.245 | 6.611 | 6.082 | 0.303 | 0.745 | 1.386 |

*Note. The best result is in* ***bold*** *and the second best is* <u>*underlined*</u>*.*

Our LinFACT algorithms consistently outperform top-$m$ algorithms, BayesGap, and KGCB. While our algorithms perform slightly worse than Lazy TTS for some cases, they have much lower sample complexity, as shown in Figure 6, meaning that Lazy TTS requires substantially more samples to achieve these results. When comparing LinFACT-G and LinFACT-$\mathcal{XY}$, we observe that in adaptive settings, the $F1$ scores are similar, but LinFACT-$\mathcal{XY}$ achieves lower sample complexity. In static settings, however, LinFACT-G attains a higher $F1$ score with a reduced sample complexity. This difference stems from the distinct focus of the two designs: the $\mathcal{XY}$-optimal design prioritizes pulling arms to obtain better estimates along the directions representing differences between arms, while the G-optimal design aims to improve estimates along the directions representing all arms.

### 7.2. Experiments with Real Data - Drug Discovery

**Experiment Setup.** We adopt the Free-Wilson model (Katz et al. 1977) and use real data from a drug discovery task. The Free-Wilson model is a linear framework in which the overall efficacy of a compound is expressed as the sum of the contributions from each substituent on the base molecule, along with the effect of the base molecule itself (Negoescu et al. 2011).

Each compound is modeled as an arm represented by a binary indicator vector. Suppose there are $N$ modification sites, with site $n \in [N]$ offering $l_n$ alternative substituents. Then each arm $\boldsymbol{a}$ lies in $\mathbb{R}^{1+\sum_{n\in[N]} l_n}$, where the initial entry corresponds to the base molecule (*i.e.*, the intercept term). For each site, the corresponding segment in the vector has exactly one entry set to 1 (indicating the chosen substituent), with the remaining $l_n - 1$ entries set to 0. This results in a total of $\prod_{n\in[N]} l_n$ unique compound configurations.

We conducted our experiments using data as described in Katz et al. (1977), retaining only the non-zero entries. The base molecule, illustrated in Figure 7, contains five sites where substituents

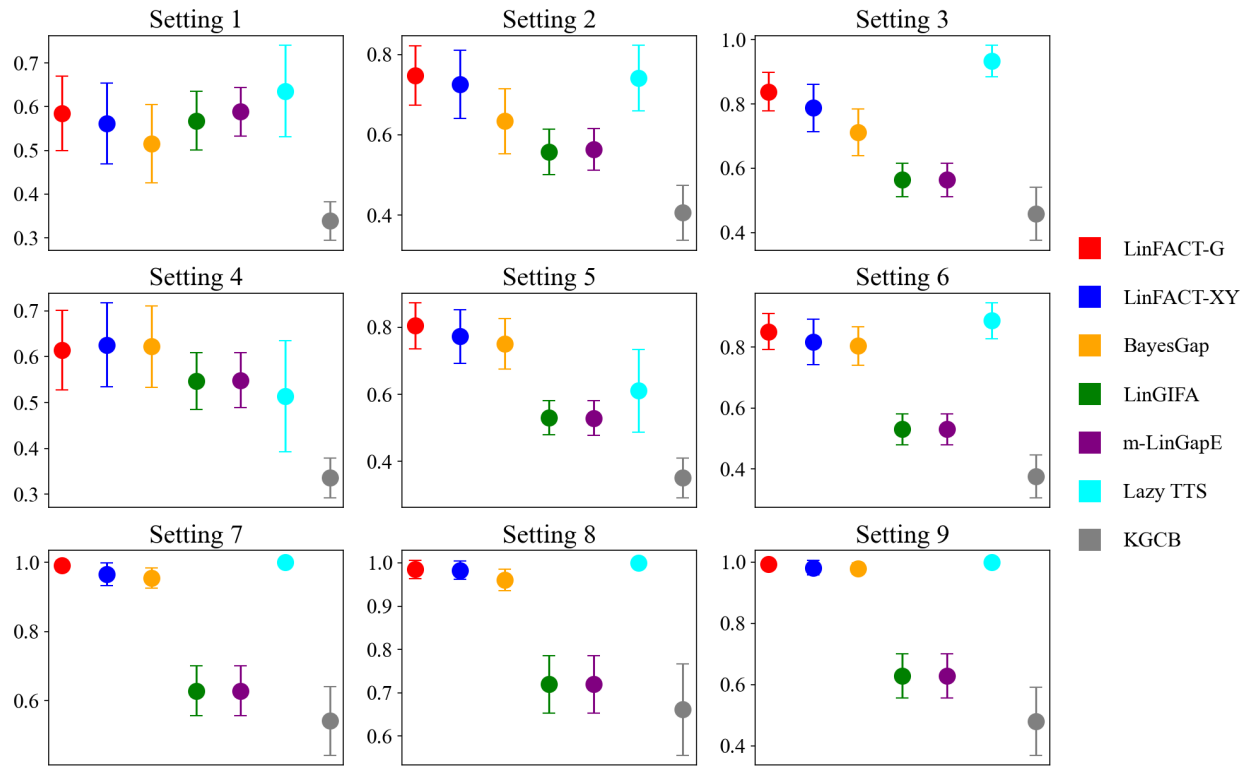



**Figure 5** $F1$ Scores for Different Synthetic Experiments Among Algorithms

*Note. The y-axis reports the F1 score, which reflects how accurately each algorithm identifies $\varepsilon$-best arms. LinFACT-G and LinFACT-$\mathcal{XY}$ consistently achieve high F1 scores. While Lazy TTS occasionally attains higher scores, we demonstrate in the next figure that it requires significantly more samples to do so. Detailed configurations of each experimental setting are provided in Table 1.*

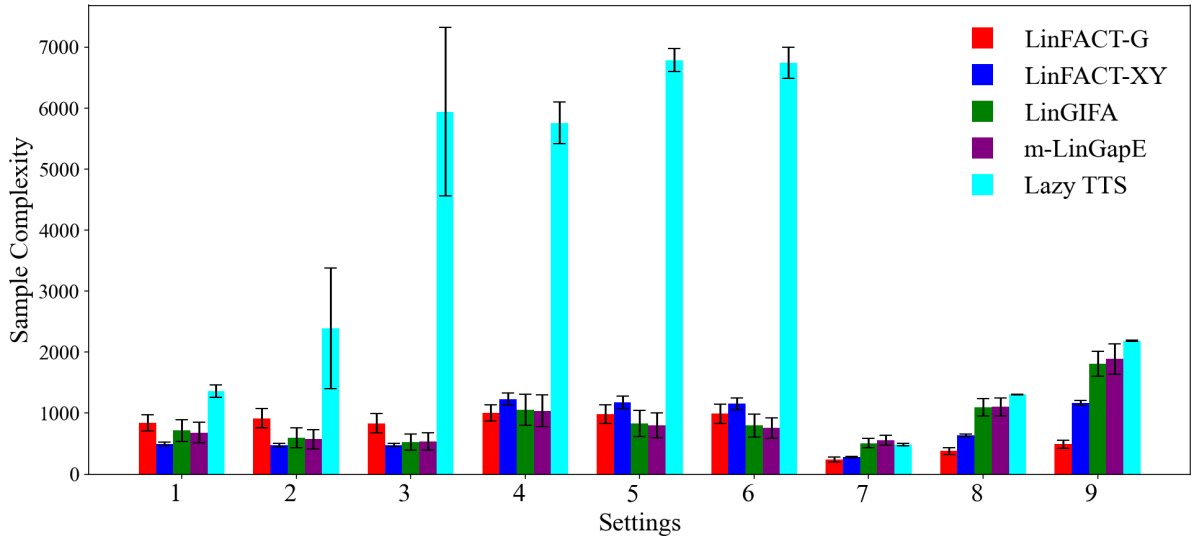



**Figure 6** Sample Complexity for Different Synthetic Experiments Among Algorithms

*Note. LinFACT-G and LinFACT-$\mathcal{XY}$ demonstrate sample complexities comparable to LinGIFA and m-LinGapE while achieving higher F1 scores in the adaptive settings (Settings 1 to 6), and consistently outperform all other algorithms in the static settings (Settings 7 to 9). BayesGap and KGCB are excluded from this comparison as they are designed for the fixed-budget setting, and thus their sample complexity is not well-defined.*

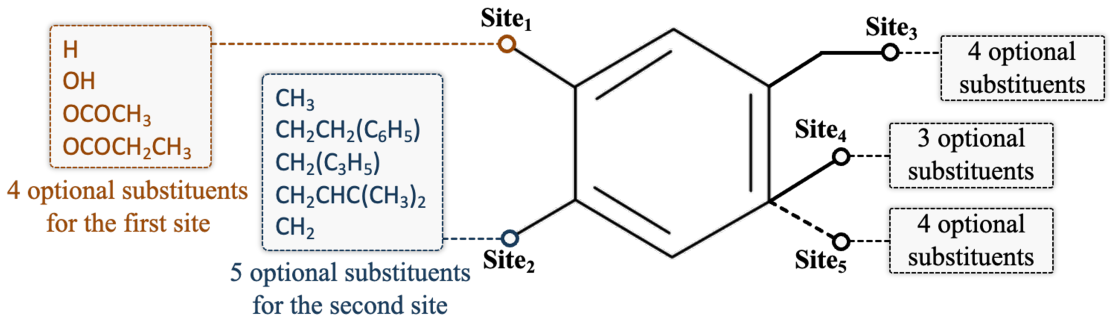



**Figure 7** An Example of Molecule and Substituent Locations with $\text{Site}_1, \ldots, \text{Site}_5$

*Note. We begin with a base molecule containing multiple attachment sites for chemical substituents. By varying these substituents, we generate a diverse set of compounds and aim to identify those with desirable properties.*

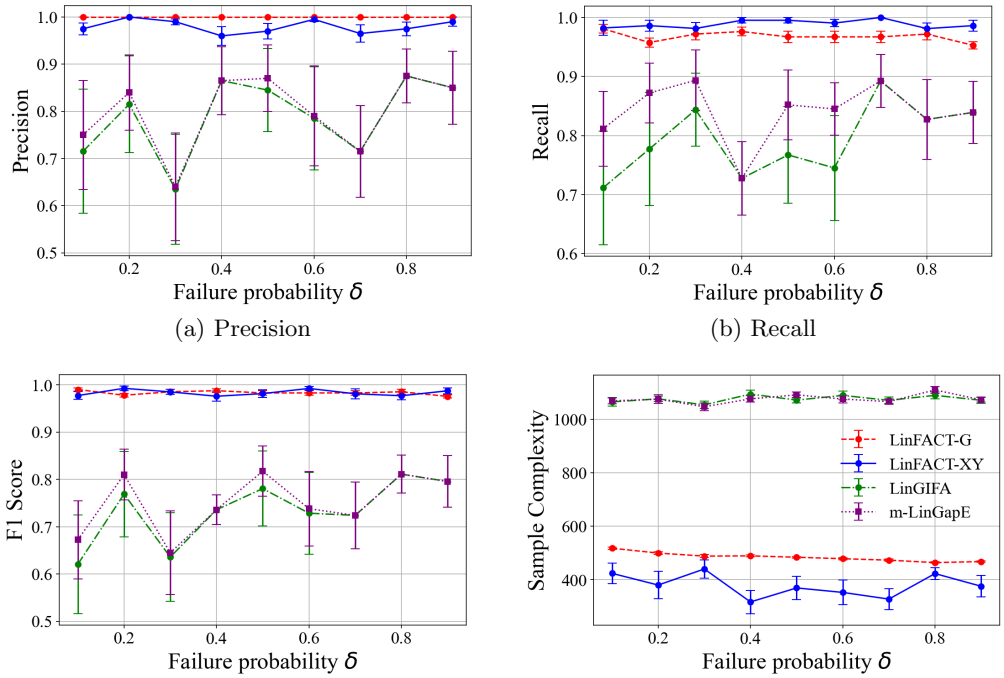
can be attached. Each site offers 4, 5, 4, 3, and 4 candidate substituents, respectively, resulting in 960 possible compounds. Each arm $\boldsymbol{a}$ is represented as a vector in $\mathbb{R}^{21}$.

We benchmarked our algorithms, LinFACT-G and LinFACT-$\mathcal{XY}$, against LinGIFA and m-LinGapE[4], evaluating how precision, recall, $F1$ score, and sample complexity vary with the failure probability $\delta$. In this study, the good set was defined to include 20 $\varepsilon$-best arms, which corresponds to $\varepsilon = 4.325$. All methods were tested over 10 trials, with $\delta$ ranging from 0.1 to 0.9 in increments of 0.1.

**Experiment results.** The experimental results are shown in Figure 8. LinFACT-G and LinFACT-$\mathcal{XY}$ consistently deliver high precision, recall, and $F1$ scores across various failure probabilities $\delta$, as shown in Figures 8a–8c. In particular, their precision remains close to 1.0, indicating that nearly all selected arms are truly $\varepsilon$-best. Their recall is also robust across $\delta$, suggesting strong coverage of the good set with minimal omission. This balance between high precision and recall leads to $F1$ scores that remain consistently near optimal, even as $\delta$ varies from 0.1 to 0.9. In contrast, LinGIFA and m-LinGapE exhibit larger fluctuations and overall lower values in all three metrics, with especially degraded recall and $F1$ performance at moderate values of $\delta$.

In addition, as illustrated in Figure 8d, LinFACT-$\mathcal{XY}$ achieves the lowest sample complexity, followed closely by LinFACT-G, with both outperforming all baseline methods. These performance disparities highlight the reliability and robustness of LinFACT methods. Finally, LinFACT-G and LinFACT-$\mathcal{XY}$ also demonstrate strong computational efficiency: both complete the task within one minute, whereas LinGIFA and m-LinGapE take about four times longer, and Lazy TTS requires over 30 minutes.

[4] KGCB and BayesGap are designed for fixed-budget settings, so their sample complexity is not well-defined. Lazy TTS, due to its prohibitive runtime, was excluded from these experiments.




(a) Precision (b) Recall

(c) $F1$ Score (d) Sample Complexity

**Figure 8** Precision, Recall, and $F1$ Score for Various Failure Probabilities $\delta$

*Note. LinFACT-G and LinFACT-$\mathcal{XY}$ consistently demonstrate the best performance across all metrics. LinFACT-$\mathcal{XY}$ achieves the lowest sample complexity while maintaining high accuracy. BayesGap and KGCB are excluded as they operate under a fixed-budget setting. Lazy TTS is also omitted due to excessive runtime.*

## 8. Conclusion

In this paper, we address the challenge of identifying all $\varepsilon$-best arms in linear bandits, motivated by applications such as drug discovery. We establish the first information-theoretic lower bound to characterize the problem's complexity and derive a matching upper bound. Our LinFACT algorithm achieves instance-optimal performance up to a logarithmic factor under the $\mathcal{XY}$-optimal design criterion.

We further extend our analysis to settings with model misspecification and generalized linear models (GLMs), deriving new upper bounds and providing insights into algorithmic behavior under these broader conditions. These results generalize and recover the guarantees from the perfectly linear case as special instances. Our numerical experiments confirm that LinFACT outperforms existing methods in both sample and computational efficiency, while maintaining high accuracy in identifying all $\varepsilon$-best arms.

**Future Research Directions.** First, while our current work focuses on the fixed-confidence setting, many real-world applications operate under a fixed sampling budget, where the objective is to achieve the best outcome within a limited number of trials. Extending the algorithm to this fixed-budget setting and establishing corresponding theoretical guarantees remains an important

avenue for exploration. Moreover, deriving fundamental lower bounds in the fixed-budget regime is still an open question. Second, although we have proposed extensions for both misspecified linear bandits and generalized linear models (GLMs), future work could benefit from developing separate algorithms tailored to each setting. Such targeted designs may offer improved performance.

## References


Abbasi-Yadkori Y, Pál D, Szepesvári C (2011) Improved algorithms for linear stochastic bandits. *Advances in neural information processing systems* 24.

Abe N, Long PM (1999) Associative reinforcement learning using linear probabilistic concepts. *ICML*, 3–11 (Citeseer).

Abernethy JD, Amin K, Zhu R (2016) Threshold bandits, with and without censored feedback. *Advances In Neural Information Processing Systems* 29.

Al Marjani A, Kocak T, Garivier A (2022) On the complexity of all $\varepsilon$-best arms identification. *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 317–332 (Springer).

Alaei S, Makhdoumi A, Malekian A, Pekeč S (2022) Revenue-sharing allocation strategies for two-sided media platforms: Pro-rata vs. user-centric. *Management Science* 68(12):8699–8721.

Allen-Zhu Z, Li Y, Singh A, Wang Y (2017) Near-optimal discrete optimization for experimental design: A regret minimization approach.

Allen-Zhu Z, Li Y, Singh A, Wang Y (2021) Near-optimal discrete optimization for experimental design: A regret minimization approach. *Mathematical Programming* 186:439–478.

Azizi MJ, Kveton B, Ghavamzadeh M (2021) Fixed-budget best-arm identification in structured bandits. *arXiv preprint arXiv:2106.04763* .

Bechhofer RE (1954) A single-sample multiple decision procedure for ranking means of normal populations with known variances. *The Annals of Mathematical Statistics* 16–39.

Boyd EA, Bilegan IC (2003) Revenue management and e-commerce. *Management science* 49(10):1363–1386.

Bubeck S, Cesa-Bianchi N, et al. (2012) Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends® in Machine Learning* 5(1):1–122.

Bubeck S, Wang T, Viswanathan N (2013) Multiple identifications in multi-armed bandits. *International Conference on Machine Learning*, 258–265 (PMLR).

Chen CH, Lin J, Yücesan E, Chick SE (2000) Simulation budget allocation for further enhancing the efficiency of ordinal optimization. *Discrete Event Dynamic Systems* 10:251–270.

Das P, Sercu T, Wadhawan K, Padhi I, Gehrmann S, Cipcigan F, Chenthamarakshan V, Strobelt H, Dos Santos C, Chen PY, et al. (2021) Accelerated antimicrobial discovery via deep generative models and molecular dynamics simulations. *Nature Biomedical Engineering* 5(6):613–623.

Elmaghraby W, Keskinocak P (2003) Dynamic pricing in the presence of inventory considerations: Research overview, current practices, and future directions. *Management science* 49(10):1287–1309.

Even-Dar E, Mannor S, Mansour Y, Mahadevan S (2006) Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *Journal of machine learning research* 7(6).

Fan W, Hong LJ, Nelson BL (2016) Indifference-zone-free selection of the best. *Operations Research* 64(6):1499–1514.

Feng Q, Ma T, Zhu R (2025) Satisficing regret minimization in bandits. *Proceedings of the 13th International Conference on Learning Representations*.

Feng Y, Caldentey R, Ryan CT (2022) Robust learning of consumer preferences. *Operations Research* 70(2):918–962.

Fiez T, Jain L, Jamieson KG, Ratliff L (2019) Sequential experimental design for transductive linear bandits. *Advances in neural information processing systems* 32.

Frazier P, Powell W, Dayanik S (2009) The knowledge-gradient policy for correlated normal beliefs. *INFORMS journal on Computing* 21(4):599–613.

Frazier PI, Powell WB, Dayanik S (2008) A knowledge-gradient policy for sequential information collection. *SIAM Journal on Control and Optimization* 47(5):2410–2439.

Free SM, Wilson JW (1964) A mathematical contribution to structure-activity studies. *Journal of medicinal chemistry* 7(4):395–399.

Gabillon V, Ghavamzadeh M, Lazaric A (2012) Best arm identification: A unified approach to fixed budget and fixed confidence. *Advances in Neural Information Processing Systems* 25.

Ghosh A, Chowdhury SR, Gopalan A (2017) Misspecified linear bandits. *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31.

Godinho de Matos M, Ferreira P, Smith MD (2018) The effect of subscription video-on-demand on piracy: Evidence from a household-level randomized experiment. *Management Science* 64(12):5610–5630.

Hoffman MW, Shahriari B, de Freitas N (2013) Exploiting correlation and budget constraints in bayesian multi-armed bandit optimization. *arXiv preprint arXiv:1303.6746* .

Hong LJ, Fan W, Luo J (2021) Review on ranking and selection: A new perspective. *Frontiers of Engineering Management* 8(3):321–343.

Huang Z, Zeng DD, Chen H (2007) Analyzing consumer-product graphs: Empirical findings and applications in recommender systems. *Management science* 53(7):1146–1164.

Kalyanakrishnan S, Stone P (2010) Efficient selection of multiple bandit arms: Theory and practice. *ICML*, volume 10, 511–518.

Kalyanakrishnan S, Tewari A, Auer P, Stone P (2012) Pac subset selection in stochastic multi-armed bandits. *ICML*, volume 12, 655–662.

Katz R, Osborne SF, Ionescu F (1977) Application of the free-wilson technique to structurally related series of homologs. quantitative structure-activity relationship studies of narcotic analgetics. *Journal of Medicinal Chemistry* 20(11):1413–1419.

Kaufmann E, Cappé O, Garivier A (2016) On the complexity of best arm identification in multi-armed bandit models. *Journal of Machine Learning Research* 17:1–42.

Kaufmann E, Kalyanakrishnan S (2013) Information complexity in bandit subset selection. *Conference on Learning Theory*, 228–251 (PMLR).

Kim SH, Nelson BL (2001) A fully sequential procedure for indifference-zone selection in simulation. *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 11(3):251–273.

Kim SH, Nelson BL (2006) Selecting the best system. *Handbooks in operations research and management science* 13:501–534.

Koenig LW, Law AM (1985) A procedure for selecting a subset of size m containing the l best of k independent normal populations, with applications to simulation. *Communications in Statistics-Simulation and Computation* 14(3):719–734.

Komiyama J, Ariu K, Kato M, Qin C (2023) Rate-optimal bayesian simple regret in best arm identification. *Mathematics of Operations Research* .

Kveton B, Zaheer M, Szepesvari C, Li L, Ghavamzadeh M, Boutilier C (2023) Randomized exploration in generalized linear bandits.

Lattimore T, Szepesvári C (2020) *Bandit algorithms* (Cambridge University Press).

Lattimore T, Szepesvari C, Weisz G (2020) Learning with good feature representations in bandits and in rl with a generative model.

Li L, Lu Y, Zhou D (2017) Provably optimal algorithms for generalized linear contextual bandits. *International Conference on Machine Learning*, 2071–2080 (PMLR).

Li Z, Fan W, Hong LJ (2024) The (surprising) sample optimality of greedy procedures for large-scale ranking and selection. *Management Science* .

Locatelli A, Gutzeit M, Carpentier A (2016) An optimal algorithm for the thresholding bandit problem. *International Conference on Machine Learning*, 1690–1698 (PMLR).

Mannor S, Tsitsiklis JN (2004) The sample complexity of exploration in the multi-armed bandit problem. *Journal of Machine Learning Research* 5(Jun):623–648.

Mason B, Jain L, Tripathy A, Nowak R (2020) Finding all $\epsilon$-good arms in stochastic bandits. *Advances in Neural Information Processing Systems* 33:20707–20718.

McCullagh P (2019) *Generalized linear models* (Routledge).

Negoescu DM, Frazier PI, Powell WB (2011) The knowledge-gradient algorithm for sequencing experiments in drug discovery. *INFORMS Journal on Computing* 23(3):346–363.

Peukert C, Sen A, Claussen J (2023) The editor and the algorithm: Recommendation technology in online news. *Management science* .

Pukelsheim F (2006) *Optimal design of experiments* (SIAM).

Qin C, You W (2025) Dual-directed algorithm design for efficient pure exploration. *Operations Research* .

Réda C, Kaufmann E, Delahaye-Duriez A (2021a) Top-m identification for linear bandits. *International Conference on Artificial Intelligence and Statistics*, 1108–1116 (PMLR).

Réda C, Tirinzoni A, Degenne R (2021b) Dealing with misspecification in fixed-confidence linear top-m identification. *Advances in Neural Information Processing Systems* 34:25489–25501.

Rivera EO, Tewari A (2024) Optimal thresholding linear bandit. *arXiv preprint arXiv:2402.09467* .

Russo D (2020) Simple bayesian algorithms for best-arm identification. *Operations Research* 68(6):1625–1647.

Ryzhov IO, Powell WB, Frazier PI (2012) The knowledge gradient algorithm for a general class of online learning problems. *Operations Research* 60(1):180–195.

Shen H, Hong LJ, Zhang X (2021) Ranking and selection with covariates for personalized decision making. *INFORMS Journal on Computing* 33(4):1500–1519.

Shin D, Broadie M, Zeevi A (2018) Tractable sampling strategies for ordinal optimization. *Operations Research* 66(6):1693–1712.

Simchi-Levi D, Wang C, Xu J (2024) On experimentation with heterogeneous subgroups: An asymptotic optimal $\delta$-weighted-pac design. *SSRN Electronic Journal* .

Soare M, Lazaric A, Munos R (2014) Best-arm identification in linear bandits. *Advances in Neural Information Processing Systems* 27.

Thompson WR (1933) On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika* 25(3-4):285–294.

Thornton C, Hutter F, Hoos HH, Leyton-Brown K (2013) Auto-weka: Combined selection and hyperparameter optimization of classification algorithms. *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, 847–855.

Wolke R, Schwetlick H (1988) Iteratively reweighted least squares: algorithms, convergence analysis, and numerical comparisons. *SIAM journal on scientific and statistical computing* 9(5):907–921.

Xu L, Honda J, Sugiyama M (2018) A fully adaptive algorithm for pure exploration in linear bandits. *International Conference on Artificial Intelligence and Statistics*, 843–851 (PMLR).

Yang J, Tan V (2022) Minimax optimal fixed-budget best arm identification in linear bandits. Koyejo S, Mohamed S, Agarwal A, Belgrave D, Cho K, Oh A, eds., *Advances in Neural Information Processing Systems*, volume 35, 12253–12266 (Curran Associates, Inc.).

Zanette A, Lazaric A, Kochenderfer M, Brunskill E (2020) Learning near optimal policies with low inherent bellman error. *International Conference on Machine Learning*, 10978–10989 (PMLR).

# E-Companion – Identifying All $\varepsilon$-Best Arms In Linear Bandits With Misspecification

## EC.1. Additional Literature

### EC.1.1. Misspecified Linear Bandits.

The linear bandit (LB) problem, introduced by Abe and Long (1999), extends the multi-armed bandits (MABs) framework by incorporating structural relationships among different arms. In the context of best arm identification, Garivier and Kaufmann (2016) established a classical lower bound, which was later extended to linear bandits by Fiez et al. (2019) using transportation inequalities.

The foundational study of linear bandits in the pure exploration framework was conducted by Hoffman et al. (2014), who addressed the best arm identification (BAI) problem in a fixed-budget setting while considering correlations among arm distributions. They proposed BayesGap, a Bayesian variant of the gap-based exploration algorithm (Gabillon et al. 2012). Although BayesGap outperformed methods that ignore correlations and structural relationships, its limitation of ceasing to pull arms deemed sub-optimal hindered its effectiveness in linear bandit pure exploration.

A key distinction between stochastic MABs and linear bandits is that, in MABs, once an arm's sub-optimality is confirmed with high probability, it is no longer pulled. In linear bandits, however, even sub-optimal arms can offer valuable information about the parameter vector, improving confidence in estimates and aiding the discrimination of near-optimal arms. This insight has led to the adoption of optimal linear experiment design as a crucial framework for linear bandit pure exploration (Abbasi-Yadkori et al. 2011, Soare et al. 2014, Fiez et al. 2019, Réda et al. 2021b, Yang and Tan 2021, Azizi et al. 2021a).

When applying linear models to real data, misspecification inevitably arises in situations where the data deviates from perfect linearity. The concept of misspecified bandit models was introduced in the context of cumulative regret by Ghosh et al. (2017), who demonstrated a significant limitation: any linear bandit algorithm (*e.g.*, OFUL (Abbasi-Yadkori et al. 2011) or LinUCB (Li et al. 2010)), which achieves optimal regret bounds on perfectly linear instances, can suffer linear regret on certain misspecified models. To address this, they proposed a hypothesis-test-based algorithm that avoids linear regret and achieves UCB-type sublinear regret for models with non-sparse deviations from linearity. Lattimore et al. (2020) further analyzed misspecification, showing that elimination-based algorithms with G-optimal design perform well under misspecification but incur an additional linear regret term proportional to the misspecification magnitude over the horizon.

In the pure exploration setting, misspecified linear models were first studied in the context of identifying the top $m$ best arms by Réda et al. (2021b), who introduced the MisLid algorithm, leveraging orthogonal parameterization to address misspecification. Subsequent research examined misspecification in ordinal optimization (Ahn et al. 2024), proposing prospective sampling methods that reduce the impact of misspecification as the sample size increases. Building on the definition of model misspecification and the optimization approach based on orthogonal parameterization from Réda et al. (2021b), we develop new algorithms for identifying all $\varepsilon$-best arms in misspecified linear bandits and establish new upper bounds.

### EC.1.2. Generalized Linear Bandits.

The generalized linear bandit (GLB) model (Filippi et al. 2010, Ahn and Shin 2020, Kveton et al. 2023) extends the multi-armed bandit framework by incorporating generalized linear models (GLMs) (McCullagh 2019) to model expected rewards. Specifically, the expected reward of each arm is given by a known link function applied to the inner product of a feature vector and an unknown parameter vector. Most existing algorithms for generalized linear bandits employ the upper confidence bound (UCB) approach, with randomized GLM algorithms (Chapelle and Li 2011, Russo et al. 2018, Kveton et al. 2023) demonstrating superior performance.

In the context of pure exploration, Azizi et al. (2021) introduced the first practical algorithm for best arm identification in generalized linear bandits, supported by theoretical analysis. Their work extends the best arm identification problem from linear models to more complex settings where the relationship between features and rewards follows generalized linear models (GLMs). Building on this foundation, we extend the pure exploration setting from best arm identification (BAI) to identifying all $\varepsilon$-best arms, providing analogous analyses and theoretical results for GLMs.

## EC.2. General Pure Exploration Model

In this section, we present a brief discussion about the general pure exploration problem. A more comprehensive explanation can be found in Qin and You (2025).

The decision-maker seeks to answer a query concerning the mean parameters $\boldsymbol{\mu}$ by adaptively allocating the sampling budget across the available arms. This query typically involves identifying a subset of arms that satisfy certain criteria, and the goal is to determine the correct answer with high probability. Let $\mathcal{I} = \mathcal{I}(\boldsymbol{\mu})$ denote the correct answer, which in our setting corresponds to the set of all $\varepsilon$-best arms. Define $\widetilde{M}$ as the set of parameters that yield a unique answer. Let $\Xi$ represent the collection of all possible answers, and for each $\mathcal{I}' \in \Xi$, define $M_{\mathcal{I}'} \coloneqq \left\{\boldsymbol{\vartheta} \in \widetilde{M} : \mathcal{I}(\boldsymbol{\vartheta}) = \mathcal{I}'\right\}$ as the set of parameters for which $\mathcal{I}'$ is the correct answer. The overall parameter space of interest is then given by $M \coloneqq \bigcup_{\mathcal{I}' \in \Xi} M_{\mathcal{I}'}$.

Recall that an algorithm is defined as a triplet $\mathcal{H} = (A_t,\, \tau_\delta,\, \hat{a}_\tau)$. The algorithm's sample complexity is quantified by the number of samples, denoted as $\tau_\delta$, at the point of termination. The objective is to formulate algorithms that minimize the expected sample complexity $\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta]$ across the set $\mathcal{H}$. As stated in Kaufmann et al. (2016), when $\delta \in (0,1)$, the non-asymptotic problem complexity of an instance $\boldsymbol{\mu}$ can be defined as

$$\kappa(\boldsymbol{\mu}) \coloneqq \inf_{Algo \in \mathcal{H}} \frac{\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta]}{\log(1/2.4\delta)}. \tag{EC.2.1}$$

This instance-dependent complexity indicates the smallest possible constant such that the expected sample complexity $\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta]$ scales in alignment with $\log(1/2.4\delta)$. The problem complexity $\kappa(\boldsymbol{\mu})$ is subject to an information-theoretic lower bound. This lower bound can be expressed as the optimal solution of an allocation problem, which we present in Proposition 2. To build this framework, we next introduce three important concepts: culprits, alternative sets, and $C_x$ function.

### EC.2.1. Culprits and Alternative Sets

Let $\mathcal{X}(\boldsymbol{\mu})$ denote the set of culprits under the true mean vector $\boldsymbol{\mu}$. These culprits are responsible for deviations from the correct answer $\mathcal{I}(\boldsymbol{\mu})$. The structure of $\mathcal{X}(\boldsymbol{\mu})$ varies depending on the specific exploration task, and identifying these culprits is essential for characterizing the problem's complexity and guiding the design of effective algorithms.

To identify the correct answer, an algorithm must distinguish among different instances within the parameter space $M$. Accordingly, for any instance $\boldsymbol{\mu} \in M$, the instance-dependent problem complexity $\kappa(\boldsymbol{\mu})$ is determined by the structure of the corresponding alternative set

$$\cup_{x \in \mathcal{X}(\boldsymbol{\mu})} \mathrm{Alt}_x(\boldsymbol{\mu}) = \mathrm{Alt}(\boldsymbol{\mu}) \coloneqq \{\boldsymbol{\vartheta} \in M : \mathcal{I}(\boldsymbol{\vartheta}) \neq \mathcal{I}(\boldsymbol{\mu})\}, \tag{EC.2.2}$$

which represents the set of parameters that return a solution that is different from the correct solution $\mathcal{I}(\boldsymbol{\mu})$.

As an example, consider the task of identifying the single best arm. In this case, the culprit set is $\mathcal{X}(\boldsymbol{\mu}) = [K] \backslash \{I^*(\boldsymbol{\mu})\}$, consisting of all arms except the current best arm. For each culprit $x \in \mathcal{X}(\boldsymbol{\mu})$, if there exists a parameter $\boldsymbol{\vartheta}$ under which arm $x$ has a higher mean than $I^*(\boldsymbol{\vartheta})$, then $\boldsymbol{\vartheta}$ leads to an incorrect identification caused by $x$. Each such culprit is associated with an alternative set—namely, the set of parameters that yield a wrong answer due to $x$—given by $\mathrm{Alt}_x(\boldsymbol{\mu}) = \left\{\boldsymbol{\vartheta} \in M : \vartheta_x \geq \vartheta_{I^*(\boldsymbol{\vartheta})}\right\}$ for $x \in \mathcal{X}(\boldsymbol{\mu})$.

### EC.2.2. $C_x$ unction

The task of identifying the correct answer can be formulated as a sequential hypothesis testing problem, which can be addressed using the Sequential Generalized Likelihood Ratio (SGLR) test (Kaufmann et al. 2016, Kaufmann and Koolen 2021). The SGLR statistic is defined to test a potentially composite null hypothesis $H_0 : (\boldsymbol{\mu} \in \Omega_0)$ against a potentially composite alternative hypothesis $H_1 : (\boldsymbol{\mu} \in \Omega_1)$, and is given by:

$$\mathrm{SGLR}_t = \frac{\sup_{\boldsymbol{\vartheta} \in \Omega_0 \cup \Omega_1} L(X_1, X_2, \ldots, X_t; \boldsymbol{\vartheta})}{\sup_{\boldsymbol{\vartheta} \in \Omega_0} L(X_1, X_2, \ldots, X_t; \boldsymbol{\vartheta})}, \tag{EC.2.3}$$

where $X_1, X_2, \ldots, X_t$ are the observed values from arm pulls, and $L(\cdot)$ denotes the likelihood function based on these observations and an unknown parameter $\boldsymbol{\vartheta}$. The set $\Omega_0$ corresponds to the restricted parameter space under the null hypothesis, while $\Omega_0 \cup \Omega_1$ defines the full parameter space under consideration, encompassing both the null and alternative hypotheses. These sets correspond to the alternative regions introduced in the previous section. A large value of $\mathrm{SGLR}_t$ indicates stronger evidence against the null hypothesis and supports rejecting it.

We consider distributions from a single-parameter exponential family parameterized by their means, following the formulation in Garivier and Kaufmann (2016). This family includes the Bernoulli, Poisson, and Gamma distributions with known shape parameters, as well as the Gaussian distribution with known variance. For each culprit $x \in \mathcal{X}(\hat{\boldsymbol{\mu}})$, where $\hat{\boldsymbol{\mu}}$ is the empirical mean based on observed data, we test the hypotheses: $H_{0,x} : \boldsymbol{\mu} \in \mathrm{Alt}_x(\hat{\boldsymbol{\mu}})$ versus $H_{1,x} : \boldsymbol{\mu} \notin \mathrm{Alt}_x(\hat{\boldsymbol{\mu}})$. When $\hat{\boldsymbol{\mu}}(t) \in \Omega_0 \cup \Omega_1$, the generalized likelihood ratio statistic in equation (EC.2.3) can be expressed in terms of a self-normalized sum. This leads to a formal expression of the SGLR statistic in Proposition EC.5, derived through maximum likelihood estimation and a reformulation of the KL divergence.

Proposition EC.5 **(Kaufmann and Koolen (2021))**. *The generalized likelihood ratio statistic for each culprit $x \in \mathcal{X}(\boldsymbol{\mu})$ at time step $t$ is defined as*

$$\begin{aligned} \hat{\Lambda}_{t,x} &= \ln(\mathrm{SGLR}_t) \\ &= \inf_{\boldsymbol{\vartheta} \in \mathrm{Alt}_x(\hat{\boldsymbol{\mu}}(t))} \sum_{i \in [K]} N_i(t) \mathrm{KL}(\hat{\mu}_i(t), \vartheta_i), \end{aligned} \tag{EC.2.4}$$

*where* $\mathrm{KL}(\cdot, \cdot)$ *represents the* KL *divergence of the two distributions parameterized by their means, and $N_i(t) = t \cdot p_i$ is the expected number of observations allocated to arm $i \in [K]$ up to time $t$.*

This proposition links the SGLR test to information-theoretic methods. To quantify the information and confidence required to assert that the true mean does not lie in $\mathrm{Alt}_x$ for all $x \in \mathcal{X}$, we

define the $C_x$ function as the population version of the SGLR statistic, sharing the same form as equation (EC.2.4).

$$C_x(\boldsymbol{p}) = C_x(\boldsymbol{p};\boldsymbol{\mu}) \coloneqq \inf_{\boldsymbol{\vartheta}\in \mathrm{Alt}_x} \sum_{i\in[K]} p_i \mathrm{KL}(\mu_i, \vartheta_i). \tag{EC.2.5}$$

With the introduction of culprits and the $C_x$ function, we arrive at the optimal allocation problem that defines the lower bound stated in Proposition 2. However, computing the lower bound can still be hard since it requires the solution of the minimax problem in equation (18). While the KL divergence in equation (EC.2.5) is convex for Gaussians, it can be non-convex to minimize the $C_x$ function over the culprit set $\mathcal{X}(\boldsymbol{\mu})$. To solve this problem, based on the following Proposition EC.6, we can write $\mathcal{X}(\boldsymbol{\mu})$ as a union of several convex sets. The following three equivalent expressions represent different ways of describing the lower bound, making the minimax problem in equation (18) tractable for every $x \in \mathcal{X}$.

$$\Gamma^*_{\boldsymbol{\mu}} = \max_{\boldsymbol{p}\in\mathcal{S}_K} \inf_{\boldsymbol{\vartheta}\in \mathrm{Alt}(\boldsymbol{\mu})} \sum_{i\in[K]} p_i \mathrm{KL}(\mu_i, \vartheta_i) = \max_{\boldsymbol{p}\in\mathcal{S}_K} \min_{x\in\mathcal{X}} \inf_{\boldsymbol{\vartheta}\in \mathrm{Alt}_x} \sum_{i\in[K]} p_i \mathrm{KL}(\mu_i, \vartheta_i) \tag{EC.2.6}$$

$$= \max_{\boldsymbol{p}\in\mathcal{S}_K} \min_{x\in\mathcal{X}} \sum_{i\in[K]} p_i \mathrm{KL}(\mu_i, \vartheta_i^x) \tag{EC.2.7}$$

$$= \max_{\boldsymbol{p}\in\mathcal{S}_K} \min_{x\in\mathcal{X}} C_x(\boldsymbol{p}), \tag{EC.2.8}$$

where we utilized the existence of a finite union set and a unique minimizer $\boldsymbol{\vartheta}^x$ in Proposition EC.6.

Proposition EC.6. *Assume that the distribution of each arm belongs to a canonical single-parameter exponential family, parameterized by its mean. Then, for each culprit $x \in \mathcal{X}(\boldsymbol{\mu})$,*

1. *(Wang et al. 2021) For each problem instance $\boldsymbol{\mu} \in M$, the alternative set $\mathrm{Alt}(\boldsymbol{\mu})$ is a finite union of convex sets. Namely, there exists a finite collection of convex sets $\{\mathrm{Alt}_x(\boldsymbol{\mu}) : x \in \mathcal{X}(\boldsymbol{\mu})\}$ such that $\mathrm{Alt}(\boldsymbol{\mu}) = \cup_{x\in\mathcal{X}(\boldsymbol{\mu})} \mathrm{Alt}_x(\boldsymbol{\mu})$.*
2. *Given a specific simplex distribution $\boldsymbol{p}$, there exists a unique $\boldsymbol{\vartheta}^x \in \mathrm{Alt}_x(\boldsymbol{\mu})$ that achieves the infimum in equation (EC.2.5).*

*Proof.* The proof proceeds in two parts. For the first part, the alternative set for any given culprit $x$ in our setting is given by

$$\mathrm{Alt}_x(\boldsymbol{\mu}) = \mathrm{Alt}_{i,j}(\boldsymbol{\mu}) \cup \mathrm{Alt}_m(\boldsymbol{\mu}), \tag{EC.2.9}$$

where $\mathrm{Alt}_{i,j}(\boldsymbol{\mu})$ and $\mathrm{Alt}_m(\boldsymbol{\mu})$ are defined in (EC.4.10) and (EC.4.19), respectively. Each $\mathrm{Alt}_x(\boldsymbol{\mu})$ is convex, as convexity is preserved under unions of convex sets. This can be verified by confirming that any convex combination of two points in $\mathrm{Alt}_x(\boldsymbol{\mu})$ remains in the set.

For the second part, when the reward distribution belongs to a single-parameter exponential family, the KL divergence $\mathrm{KL}(\chi, \chi')$ is continuous and strictly convex in $(\chi, \chi')$. This ensures that the infimum in equation (EC.2.5) is achieved uniquely by a single $\boldsymbol{\vartheta}$. □

### EC.2.3. Stopping Rule

In this section, we introduce the stopping rule, which suggests when to stop the algorithm and returns an answer that gives all $\varepsilon$-best arms with a probability of at least $1-\delta$. This stopping rule is based on the deviation inequalities that are linked to the generalized likelihood ratio test (Kaufmann and Koolen 2021).

For each $x_i \in \mathcal{X}(\boldsymbol{\mu})$ with $i \in [|\mathcal{X}|]$, let $M_i(\boldsymbol{\mu}) = \mathrm{Alt}_{x_i}(\boldsymbol{\mu})$ denote a partition of the realizable parameter space $M$ introduced in Section EC.2.1, where each partition $M_i(\boldsymbol{\mu})$ is associated with a distinct culprit in $\mathcal{X}(\boldsymbol{\mu})$. This implies that for any $\boldsymbol{\mu} \in M$, the parameter space $M$ can be uniquely partitioned to support a hypothesis testing framework. Let $M_0(\boldsymbol{\mu})$ denote the subset of parameters where $\boldsymbol{\mu}$ resides. If $\boldsymbol{\mu} \in M$, define $i^*(\boldsymbol{\mu})$ as the index of the unique element in the partition where the true mean value $\boldsymbol{\mu}$ belongs, and here $i^*(\boldsymbol{\mu}) = 0$. In other words, we have $\boldsymbol{\mu} \in M_0$ and $\mathrm{Alt}(\boldsymbol{\mu}) = M \backslash M_0$. Since the ordering among suboptimal arms is irrelevant, the sets $M_i(\boldsymbol{\mu})$ for $i \in \{0, 1, 2, \ldots, |\mathcal{X}|\}$ form a valid partition of $M$ for each $\boldsymbol{\mu}$. Accordingly, the alternative set can be further defined as

$$\mathrm{Alt}(\boldsymbol{\mu}) = \cup_{i:\boldsymbol{\mu} \notin M_i(\boldsymbol{\mu})} M_i(\boldsymbol{\mu}) = M \backslash M_{i^*(\boldsymbol{\mu})} = M \backslash M_0. \tag{EC.2.10}$$

Given a bandit instance $\boldsymbol{\mu}$, we consider a total of $|\mathcal{X}(\boldsymbol{\mu})| + 1$ hypotheses, defined as

$$H_0 = (\boldsymbol{\mu} \in M_0(\boldsymbol{\mu})), H_1 = (\boldsymbol{\mu} \in M_1(\boldsymbol{\mu})), \ldots, H_{|\mathcal{X}|} = (\boldsymbol{\mu} \in M_{|\mathcal{X}|}(\boldsymbol{\mu})). \tag{EC.2.11}$$

By substituting the true mean vector $\boldsymbol{\mu}$ with its empirical estimate $\hat{\boldsymbol{\mu}}(t)$, the SGLR test becomes data-dependent, relying on the empirical means at each time step $t$. Consequently, the hypotheses tested at time $t$ are also data-dependent. If $\hat{\boldsymbol{\mu}}(t) \in M$, we define $\hat{i}(t) = i^*(\hat{\boldsymbol{\mu}}(t))$ as the index of the partition to which $\hat{\boldsymbol{\mu}}(t)$ belongs—that is, $\hat{\boldsymbol{\mu}}(t) \in M_{\hat{i}(t)}$. If instead $\hat{\boldsymbol{\mu}}(t) \notin M$, we set $\hat{\Lambda}_{t,x} = 0$ for all culprits $x$, meaning no hypothesis test is conducted at that time step, and the process continues. In practice, when $\hat{\boldsymbol{\mu}}(t) \notin M$, the algorithm can revert to uniform exploration. Since the true mean vector $\boldsymbol{\mu} \in M$ and $M$ is assumed to be an open set, the law of large numbers guarantees that $\hat{\boldsymbol{\mu}}(t)$ will eventually re-enter the parameter space, i.e., $\hat{\boldsymbol{\mu}}(t) \in M$ after sufficient samples.

We run $|\mathcal{X}|$ time-varying SGLR tests in parallel, each testing $H_0$ against $H_i$ for $i \in [|\mathcal{X}(\hat{\boldsymbol{\mu}}(t))|]$. The procedure stops when any of these tests rejects $H_0$, indicating that the corresponding alternative set is empirically the easiest to reject. At this point, the accepted hypothesis for $\hat{\boldsymbol{\mu}}(t) \in M$ is identified as the most likely to be correct. Given a sequence of exploration rates $(\hat{\beta}_t(\delta))_{t \in \mathbb{N}}$, the SGLR stopping rule in the pure exploration setting is defined as follows:

$$\tau_\delta \coloneqq \inf \left\{ t \in \mathbb{N} : \min_{x \in \mathcal{X}(\hat{\boldsymbol{\mu}}(t))} \hat{\Lambda}_{t,x} > \hat{\beta}_t(\delta) \right\} = \inf \left\{ t \in \mathbb{N} : t \cdot \min_{x \in \mathcal{X}(\hat{\boldsymbol{\mu}}(t))} C_x(\boldsymbol{p}_t; \hat{\boldsymbol{\mu}}(t)) > \hat{\beta}_t(\delta) \right\}, \tag{EC.2.12}$$

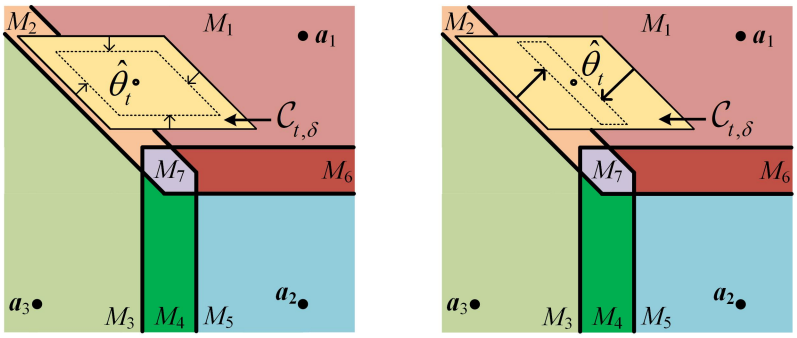
where the SGLR statistic $\hat{\Lambda}_{t,x}$ is defined in equation (EC.2.4). The testing process closely resembles the classical approach, except that the hypotheses are data-dependent and evolve over time.

We also provide insights from the perspective of the confidence region. Specifically, the event $\left\{\min_{x\in\mathcal{X}(\hat{\boldsymbol{\mu}}(t))}\hat{\Lambda}_{t,x} > \hat{\beta}_t(\delta)\right\} = \left\{\mathcal{C}_{t,\delta} \subseteq M_{\hat{i}(t)}\right\}$, where $\mathcal{C}_{t,\delta}$ denotes the confidence region of the mean vector, given by

$$\mathcal{C}_{t,\delta} := \left\{\boldsymbol{\vartheta} : \sum_{i=1}^{K} N_i(t)\mathrm{KL}(\hat{\mu}_i(t), \vartheta_i) \le \hat{\beta}_t(\delta)\right\}. \tag{EC.2.13}$$

Notably, although the confidence region $\mathcal{C}_{t,\delta}$ is defined for the mean vector $\boldsymbol{\mu}$, under the assumption of linear structure, it is equivalent to the confidence region for the parameter $\boldsymbol{\theta}$, as discussed in Sections 2.2 and EC.4.1. This equivalence allows the stopping rule to be interpreted as follows: the algorithm halts once the confidence region for the mean vector fully lies within a single partition region, aligning with the graphical interpretation of optimal allocation given in Section EC.4.2.

## EC.3. Difference between G-Optimal Design and $\mathcal{XY}$-Optimal Design




(a) Uniform Contraction Based on G-Optimal Design

(b) More Purposeful Contraction of $\mathcal{XY}$-Optimal Design

**Figure EC.3.1** Key Distinction Between G-Optimal and $\mathcal{XY}$-Optimal Designs in Terms of Stopping Criteria

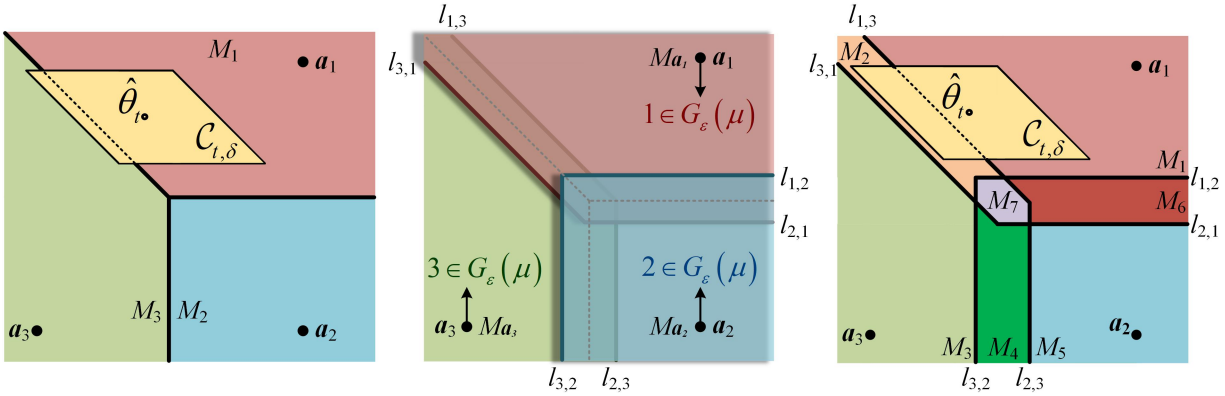
*Note. The contraction behavior and rate of the confidence region for the parameter $\hat{\boldsymbol{\theta}}_t$ differ: under G-optimal sampling (left), the region shrinks uniformly in all directions, whereas under $\mathcal{XY}$-optimal design (right), it contracts more strategically along directions critical for classification, allowing the confidence region to enter a decision region more rapidly and trigger earlier stopping.*

Returning to the intuition and visual explanation in Section EC.4.1 and Figure EC.4.1, we see that G-optimal sampling inevitably leads to inefficient sampling. Figure EC.3.1 illustrates the advantage of adopting the $\mathcal{XY}$-optimal design from the perspective of the stopping rule: the algorithm terminates once the yellow confidence region for $\hat{\boldsymbol{\theta}}$ enters one of the decision regions $M_i$ $(i = 1, \ldots, 7)$. Unlike the isotropic shrinkage of the confidence region under G-optimal design, the $\mathcal{XY}$-optimal design guides the region to contract more aggressively in directions critical for distinguishing arms. Rather than uniformly estimating $\boldsymbol{\theta}$, it prioritizes reducing uncertainty along directions that matter most for classification, leading to more efficient exploration.

## EC.4. Lower Bound for All $\varepsilon$-Best Arms Identification in Linear Bandits

This section provides both geometric insights regarding the stopping condition and formal proofs establishing the lower bound for identifying all $\varepsilon$-best arms in linear bandit settings.

### EC.4.1. Visual Illustration of the Stopping Condition




(a) Best Arm Identification (BAI) in Linear Bandits

(b) Three Overlapping Regions Representing the Combination of All $\varepsilon$-Best Arms

(c) All $\varepsilon$-Best Arms Identification in Linear Bandits

**Figure EC.4.1** Visual Illustration of Identifying the Best Arm vs. Identifying All $\varepsilon$-Best Arms.

*Note. (a) Stopping occurs when the confidence region $\mathcal{C}_{t,\delta}$ for the estimated parameter $\hat{\boldsymbol{\theta}}_t$ contracts entirely within one of the three decision regions $M_i$ in a certain time step $t$. The boundaries between regions are defined by the hyperplanes $\boldsymbol{\vartheta}^\top(\boldsymbol{a}_i - \boldsymbol{a}_j) = 0$. Each dot represents an arm. (b) In the case of identifying all $\varepsilon$-best arms, the regions overlap. (c) Due to these overlaps, the space is partitioned into seven distinct decision regions, increasing the difficulty of identification.*

The stopping condition is formulated as a hypothesis test conducted as data is collected, which can be interpreted as the process of the parameter confidence region contracting into one of the decision regions (*i.e.*, the set of parameters that yield the same decision). A more detailed version of Figure 1 is shown in Figure EC.4.1, which illustrates the core idea of the stopping condition for identifying all $\varepsilon$-best arms in linear bandits. The key distinction between identifying all $\varepsilon$-best arms and identifying the single best arm lies in how the decision regions are partitioned.

Figure EC.4.1(a) illustrates the best arm identification process in linear bandits, where $\boldsymbol{a}^* = \boldsymbol{a}^*(\boldsymbol{\mu})$ represents the arm with the largest mean value for each bandit instance $\boldsymbol{\mu}$. Let $M_i = \{\boldsymbol{\vartheta} \in \mathbb{R}^d \mid \boldsymbol{a}_i = \boldsymbol{a}^*\}$ be the set of parameters $\boldsymbol{\theta}$ for which $\boldsymbol{a}_i$ $(i = 1, 2, 3)$ is the optimal arm. Each $M_i$ forms a cone defined by the intersection of half-spaces.

Figure EC.4.1(b) represents an intermediate step, demonstrating the transition from best arm identification to identifying all $\varepsilon$-best arms. Let $M_{\boldsymbol{a}_i} = \{\boldsymbol{\vartheta} \in \mathbb{R}^d \mid i \in G_\varepsilon(\boldsymbol{\mu})\}$ be the set of parameters $\boldsymbol{\theta}$ include arm $i$ in the set $G_\varepsilon(\boldsymbol{\mu})$. $M_{\boldsymbol{a}_i}$ is similarly defined by the intersection of half-spaces. The

overlap of these three regions forms the decision regions $M_i$ $(i = 1, 2, \ldots, 7)$ in Figure EC.4.1(c), which correspond to the seven distinct types of $\varepsilon$-best arms sets. Besides, the BAI process in (a) is a special case of the $\varepsilon$-best arms identification in (c), occurring when the gap $\varepsilon$ approaches 0. The following statement provides a detailed explanation of how the decision regions in Figure EC.4.1(c) are constructed.

In a $d$-dimensional Euclidean space $\mathbb{R}^d$, hyperplanes $l_{i,j}$ can be defined for any pair of arms, partitioning the space into the following half-spaces:

$$H_{i,j}^{+} = \left\{ \boldsymbol{\vartheta} \in \mathbb{R}^d \mid (\boldsymbol{a}_i - \boldsymbol{a}_j)^\top \boldsymbol{\vartheta} > \varepsilon \right\}, \tag{EC.4.1}$$

$$H_{i,j}^{-} = \left\{ \boldsymbol{\vartheta} \in \mathbb{R}^d \mid (\boldsymbol{a}_i - \boldsymbol{a}_j)^\top \boldsymbol{\vartheta} \leq \varepsilon \right\}. \tag{EC.4.2}$$

The hyperplane $l_{i,j}$, which separates the half-spaces $H_{i,j}^{+}$ and $H_{i,j}^{-}$, is perpendicular to the direction vector $\boldsymbol{a}_i - \boldsymbol{a}_j$. The intersection of these half-spaces, represented by $\bigcap_{i,j \in [K],\, j \neq i} H_{i,j}$, partitions the space into distinct regions. Each region corresponds to a solution set, representing all $\varepsilon$-best arms if the true parameter $\boldsymbol{\theta}$ lies within that region. As the gap $\varepsilon$ approaches 0, the hyperplanes $l_{i,j}$ on both sides of the decision boundaries move closer together, causing some decision regions in Figure EC.4.1(c) to shrink until they vanish. The relationship between the true parameter $\boldsymbol{\theta}$ and the half-spaces determines the belonging of arms in the good set $G_\varepsilon(\boldsymbol{\mu})$.

For the case of three arms, the space is divided into three overlapping regions, as shown in Figure EC.4.1(b). These regions further generate seven $(2^3 - 1 = 7)$ decision regions, denoted $M_1$ through $M_7$, which are summarized in Table EC.4.1.

**Table EC.4.1** Three Overlapping Regions and Seven Decision Regions in the Case of Three Arms

| Decision Region | Set Expression | $\varepsilon$-Best Arms |
|---|---|---|
| $M_1$ | $M_{a_1} \cap M_{a_2}^c \cap M_{a_3}^c$ | $\{1\}$ |
| $M_2$ | $M_{a_1} \cap M_{a_2}^c \cap M_{a_3}$ | $\{1,3\}$ |
| $M_3$ | $M_{a_1}^c \cap M_{a_2}^c \cap M_{a_3}$ | $\{3\}$ |
| $M_4$ | $M_{a_1}^c \cap M_{a_2} \cap M_{a_3}$ | $\{2,3\}$ |
| $M_5$ | $M_{a_1}^c \cap M_{a_2} \cap M_{a_3}^c$ | $\{2\}$ |
| $M_6$ | $M_{a_1} \cap M_{a_2} \cap M_{a_3}^c$ | $\{1,2\}$ |
| $M_7$ | $M_{a_1} \cap M_{a_2} \cap M_{a_3}$ | $\{1,2,3\}$ |

The stopping condition verifies whether the confidence region $\mathcal{C}_{t,\delta}$ is entirely contained within a specific decision region $M_i$. It is important to note that, due to the definition of the confidence region and the property that $\hat{\boldsymbol{\theta}}_t \to \boldsymbol{\theta}$ as $t \to \infty$, any algorithm that continually samples all arms will eventually meet the stopping condition.

### EC.4.2. Optimal Allocation

The goal of the sampling policy is to construct an allocation sequence that drives the confidence set $C_{t,\delta}$ into the optimal region $M^*$ as efficiently as possible. Geometrically, this entails selecting arms that cause $C_{t,\delta}$ to contract into the optimal cone $M^*$ with minimal sampling effort. The condition $\mathcal{C}_{t,\delta} \subseteq M^*$ can be expressed as

$$\text{For all } i \in G_\varepsilon(\boldsymbol{\mu}),\, j \neq i,\, m \notin G_\varepsilon(\boldsymbol{\mu}) \text{ and } \forall \boldsymbol{\vartheta} \in \mathcal{C}_{t,\delta},\ \text{we have } \boldsymbol{\vartheta} \in H_{j,i}^- \text{ and } \boldsymbol{\vartheta} \in H_{1,m}^+. \tag{EC.4.3}$$

In words, every parameter vector $\boldsymbol{\vartheta}$ that remains plausible must preserve all required pairwise orderings: no $\varepsilon$-optimal arm $i$ can be overtaken by any rival $j$, and the best arm 1 must stay ahead of every suboptimal arm $m$. Equivalently, no $\boldsymbol{\vartheta} \in \mathcal{C}_{t,\delta}$ is allowed to flip these comparisons.

The relationships $\boldsymbol{\vartheta} \in H_{j,i}^-$ and $\boldsymbol{\vartheta} \in H_{1,m}^+$ are equivalent to the following inequalities by adding terms on both sides of the inequalities (EC.4.1) and (EC.4.2) and reorganizing:

$$\begin{cases} (\boldsymbol{a}_i - \boldsymbol{a}_j)^\top (\boldsymbol{\theta} - \boldsymbol{\vartheta}) \leq (\boldsymbol{a}_i - \boldsymbol{a}_j)^\top \boldsymbol{\theta} + \varepsilon \\ (\boldsymbol{a}_1 - \boldsymbol{a}_m)^\top (\boldsymbol{\theta} - \boldsymbol{\vartheta}) < (\boldsymbol{a}_1 - \boldsymbol{a}_m)^\top \boldsymbol{\theta} - \varepsilon \end{cases}. \tag{EC.4.4}$$

Now, we focus on the confidence region, which can be constructed following Cauchy's inequality and the definition of the confidence ellipse for parameter $\boldsymbol{\theta}$ in equation (10).

$$\mathcal{C}_{t,\delta} = \left\{ \boldsymbol{\vartheta} \in \mathbb{R}^d \,\middle|\, \forall i \in G_\varepsilon(\boldsymbol{\mu}),\, j \neq i,\, m \notin G_\varepsilon(\boldsymbol{\mu}),\, \begin{cases} (\boldsymbol{a}_i - \boldsymbol{a}_j)^\top (\boldsymbol{\theta} - \boldsymbol{\vartheta}) \leq \|\boldsymbol{a}_i - \boldsymbol{a}_j\|_{\boldsymbol{V}_t^{-1}} B_{t,\delta} \\ (\boldsymbol{a}_1 - \boldsymbol{a}_m)^\top (\boldsymbol{\theta} - \boldsymbol{\vartheta}) \leq \|\boldsymbol{a}_1 - \boldsymbol{a}_m\|_{\boldsymbol{V}_t^{-1}} B_{t,\delta} \end{cases} \right\}, \tag{EC.4.5}$$

where $\boldsymbol{V}_t$ is the information matrix as defined in equation (9) and the confidence bound for parameter $\boldsymbol{\theta}$, *i.e.*, $B_{t,\delta}$, can either be a fixed confidence bound as shown in Proposition 1 or a looser adaptive confidence bound introduced in Abbasi-Yadkori et al. (2011). The stopping condition $\mathcal{C}_{t,\delta} \subseteq M^*$ can thus be reformulated. For each $i \in G_\varepsilon(\boldsymbol{\mu})$, $j \neq i$, $m \notin G_\varepsilon(\boldsymbol{\mu})$, we have

$$\begin{cases} \|\boldsymbol{a}_i - \boldsymbol{a}_j\|_{\boldsymbol{V}_t^{-1}} B_{t,\delta} \leq (\boldsymbol{a}_i - \boldsymbol{a}_j)^\top \boldsymbol{\theta} + \varepsilon \\ \|\boldsymbol{a}_1 - \boldsymbol{a}_m\|_{\boldsymbol{V}_t^{-1}} B_{t,\delta} \leq (\boldsymbol{a}_1 - \boldsymbol{a}_m)^\top \boldsymbol{\theta} - \varepsilon \end{cases}. \tag{EC.4.6}$$

If equation (EC.4.6) holds, then for any $\boldsymbol{\theta} \in \mathcal{C}_{t,\delta}$, equation (EC.4.4) also holds, implying that $\mathcal{C}_{t,\delta} \subseteq M^*$. Given that $(\boldsymbol{a}_i - \boldsymbol{a}_j)^\top \boldsymbol{\theta} + \varepsilon > 0$ and $(\boldsymbol{a}_1 - \boldsymbol{a}_m)^\top \boldsymbol{\theta} - \varepsilon > 0$, and rearranging (EC.4.6), the oracle allocation strategy is determined as follows:

$$\{\boldsymbol{a}_{A_n}\}^* = \arg \min_{\{\boldsymbol{a}_{A_n}\}} \max_{i \in G_\varepsilon(\boldsymbol{\mu}),\, j \neq i,\, m \notin G_\varepsilon(\boldsymbol{\mu})} \max \left\{ \frac{2\|\boldsymbol{a}_i - \boldsymbol{a}_j\|^2_{\boldsymbol{V}_t^{-1}}}{\left(\boldsymbol{a}_i^\top \boldsymbol{\theta} - \boldsymbol{a}_j^\top \boldsymbol{\theta} + \varepsilon\right)^2}, \frac{2\|\boldsymbol{a}_1 - \boldsymbol{a}_m\|^2_{\boldsymbol{V}_t^{-1}}}{\left(\boldsymbol{a}_1^\top \boldsymbol{\theta} - \boldsymbol{a}_m^\top \boldsymbol{\theta} - \varepsilon\right)^2} \right\}, \tag{EC.4.7}$$

where $\{\boldsymbol{a}_{A_n}\} = (\boldsymbol{a}_{A_1}, \boldsymbol{a}_{A_1}, \ldots, \boldsymbol{a}_{A_n}) \in \mathcal{A}^n$ is a sequence of sampled arms and $\{\boldsymbol{a}_{A_n}\}^*$ is the oracle allocation strategy[5]. However, it is more convenient to demonstrate the sample complexity of

[5] When the arm elimination is considered, the active set of arms will gradually decrease as arms are removed, becoming a subset of $\mathcal{A}$.

the problem in terms of the continuous allocation proportion $\boldsymbol{p}$ instead of the discrete allocation sequence $\{\boldsymbol{a}_{A_n}\}$. Then, we can have the following optimal allocation proportion.

$$\boldsymbol{p}^* = \arg\min_{\boldsymbol{p}\in\mathcal{S}_K} \max_{i\in G_\varepsilon(\boldsymbol{\mu}),\, j\neq i,\, m\notin G_\varepsilon(\boldsymbol{\mu})} \max\left\{ \frac{2\|\boldsymbol{a}_i-\boldsymbol{a}_j\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}}}{\left(\boldsymbol{a}_i^\top\boldsymbol{\theta}-\boldsymbol{a}_j^\top\boldsymbol{\theta}+\varepsilon\right)^2}, \frac{2\|\boldsymbol{a}_1-\boldsymbol{a}_m\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}}}{\left(\boldsymbol{a}_1^\top\boldsymbol{\theta}-\boldsymbol{a}_m^\top\boldsymbol{\theta}-\varepsilon\right)^2} \right\}, \tag{EC.4.8}$$

where $\mathcal{S}_K$ denotes the $K$-dimensional probability simplex, and $\boldsymbol{V}_{\boldsymbol{p}} = \sum_{i=1}^K p_i \boldsymbol{a}_i \boldsymbol{a}_i^\top$ is the weighted information matrix, analogous to the Fisher information matrix (Chaloner and Verdinelli 1995). The intuition behind this optimal allocation strategy is as follows: to satisfy the inequality in (EC.4.6) as quickly as possible, the ratio of the left-hand side to the right-hand side should be minimized, leading to the outer minimization operation over the allocation probability $\boldsymbol{p}$. The middle maximization operation accounts for the fact that the inequalities in (EC.4.6) must hold for each possible triple $(i,j,m)$, which specifies the culprit set $\mathcal{X}=\mathcal{X}(\boldsymbol{\mu})$ (see Proposition 2 for a formal definition). Furthermore, since both inequalities in (EC.4.6) need to be satisfied, we take inner maximization operations to enforce this requirement.

### EC.4.3. Proof of the Lower Bound in Theorem 1

*Proof.* Deriving the lower bound requires constructing the largest possible alternative set and identifying the most challenging instance for distinguishing the arms. Specifically, to construct an alternative problem instance that possesses a distinct set of $\varepsilon$-best arms from the original problem instance $\boldsymbol{\mu}$, we systematically perturb the expected rewards of specific arms. This process can be achieved through two ways of construction, as further detailed in Figure EC.4.2:

- *Lowering an $\varepsilon$-Best Arm:* This approach decreases the expected reward of a currently $\varepsilon$-best arm $i$, while simultaneously increasing the expected reward of another arm $j$.
- *Elevating a Non-$\varepsilon$-Best Arm:* This approach increases the expected reward of a non-$\varepsilon$-best arm $m$, while concurrently reducing the expected rewards of the $\ell$ top-performing arms.

For all $\varepsilon$-best arms identification in linear bandits, we establish the correct answer, the culprit set, and the alternative set as follows. The general idea of the proof is to builds these alternatives explicitly and shows which one is the hardest to distinguish. Recall that $\mu_i = \langle\boldsymbol{\theta},\boldsymbol{a}_i\rangle$. For notational simplicity, we use $\boldsymbol{\mu}$ to represent the mean vector induced by true parameter $\boldsymbol{\theta}$.

- The correct answer $G_\varepsilon(\boldsymbol{\mu}) = \{i : \langle\boldsymbol{\theta},\boldsymbol{a}_i\rangle \geq \max_j \langle\boldsymbol{\theta},\boldsymbol{a}_j\rangle - \varepsilon\}$ for all $i\in[K]$. These are the $\varepsilon$-best arms under the true parameter.
- The culprit set $\mathcal{X}(\boldsymbol{\mu}) = \{(i,j,m,\ell) : i\in G_\varepsilon(\boldsymbol{\mu}), j\neq i, m\notin G_\varepsilon(\boldsymbol{\mu}), \ell\in\{1,2,\ldots,m-1\}\}$. The culprit set identifies potential sources of error. Intuitively, each tuple corresponds to two types of mistakes that can affect the output as mentioned above.
- The alternative set $\mathrm{Alt}(\boldsymbol{\mu}) = \cup_{x\in\mathcal{X}(\boldsymbol{\mu})}\mathrm{Alt}_x(\boldsymbol{\mu})$ and the form of $\mathrm{Alt}_x(\boldsymbol{\mu})$ is given by

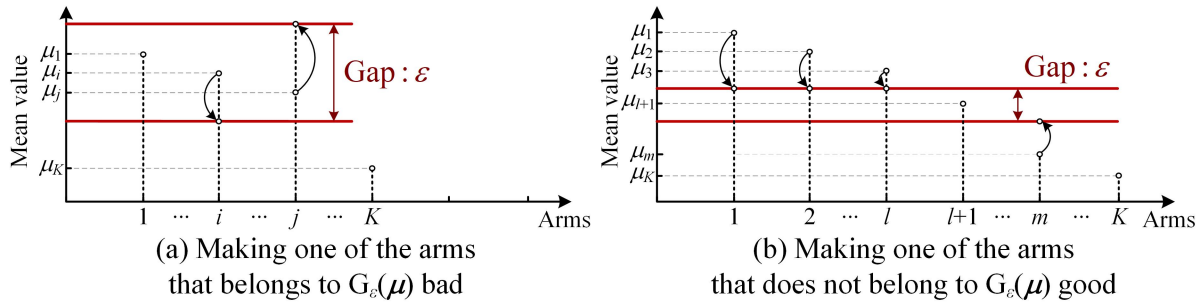



**Figure EC.4.2** Illustration of Constructing Alternative Set of All $\varepsilon$-Best Arms Identification Lower Bound

*Note. (a) Transforming an arm within $G_\varepsilon(\boldsymbol{\mu})$ into a non-$\varepsilon$-best arm by decreasing the mean of an $\varepsilon$-best arm $i$ while increasing the mean of another arm $j$. (b) Transforming an arm outside $G_\varepsilon(\boldsymbol{\mu})$ into an $\varepsilon$-best arm by increasing the mean of a non-$\varepsilon$-best arm $m$ and simultaneously decreasing the means of higher-performing arms.*

$$\mathrm{Alt}_x(\boldsymbol{\mu}) = \mathrm{Alt}_{i,j}(\boldsymbol{\mu}) \cup \mathrm{Alt}_{m,\ell}(\boldsymbol{\mu}) \text{ for all } x \in \mathcal{X}(\boldsymbol{\mu}). \tag{EC.4.9}$$

So for each culprit $x$, we build two kinds of alternative instances, each capable of flipping the answer. The two parts of the alternative sets can be expressed as

$$\mathrm{Alt}_{i,j}(\boldsymbol{\mu}) = \{\boldsymbol{\vartheta} : \langle \boldsymbol{\vartheta}, \boldsymbol{a}_i - \boldsymbol{a}_j \rangle < -\varepsilon\} \text{ for all } x \in \mathcal{X}(\boldsymbol{\mu}) \tag{EC.4.10}$$

and

$$\mathrm{Alt}_{m,\ell}(\boldsymbol{\mu}) = \left\{\boldsymbol{\vartheta} : \boldsymbol{\vartheta}^\top \boldsymbol{a}_1 = \cdots = \boldsymbol{\vartheta}^\top \boldsymbol{a}_\ell \geq \boldsymbol{\vartheta}^\top \boldsymbol{a}_m + \varepsilon > \boldsymbol{\vartheta}^\top \boldsymbol{a}_{\ell+1}\right\} \text{ for all } x \in \mathcal{X}(\boldsymbol{\mu}), \tag{EC.4.11}$$

which are the same as shown in the figure.

The first part $\mathrm{Alt}_{i,j}(\boldsymbol{\mu})$ forces a good arm $i$ to become at least $\varepsilon$ worse than arm $j$, causing $i$ to fall out of the $\varepsilon$-best set. In contrast, the second part ensures that a suboptimal arm is no more than $\varepsilon$ worse than the top $\ell$ arms with identical mean values, thereby including it in the $\varepsilon$-best set.

For a given culprit $x = (i, j, m, \ell)$, the first component of the alternative set can be written as

$$\mathrm{Alt}_{i,j}(\boldsymbol{\mu}) = \left\{ \boldsymbol{\vartheta}_{i,j}(\varepsilon, \boldsymbol{p}, \alpha) \mid \boldsymbol{\vartheta}_{i,j}(\varepsilon, \boldsymbol{p}, \alpha) = \boldsymbol{\theta} - \frac{\boldsymbol{y}_{i,j}^\top \boldsymbol{\theta} + \varepsilon + \alpha}{\|\boldsymbol{y}_{i,j}\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}}} \boldsymbol{V}_{\boldsymbol{p}}^{-1} \boldsymbol{y}_{i,j} \right\}, \tag{EC.4.12}$$

where $\boldsymbol{V}_{\boldsymbol{p}} = \sum_{i=1}^K p_i \boldsymbol{a}_i \boldsymbol{a}_i^\top$ is related to the Fisher information matrix (Chaloner and Verdinelli 1995), $\boldsymbol{y}_{i,j} = \boldsymbol{a}_i - \boldsymbol{a}_j$, and $\alpha > 0$ is a perturbation parameter. The form of this alternative set is derived from the solution to the following optimization problem:

$$\begin{aligned} \arg\min_{\boldsymbol{\vartheta} \in \mathbb{R}^d} \quad & \|\boldsymbol{\vartheta} - \boldsymbol{\theta}\|^2_{\boldsymbol{V}_{\boldsymbol{p}}} && \text{(EC.4.13)} \\ \text{s.t.} \quad & \boldsymbol{y}_{i,j}^\top \boldsymbol{\vartheta} = -\varepsilon - \alpha. && \text{(EC.4.14)} \end{aligned}$$

Here, $\alpha$ is introduced to construct the specific alternative set, providing an explicit expression for the alternative parameter $\boldsymbol{\vartheta}$. By letting $\alpha \to 0$, we realize the infimum in equation (EC.2.5) and equation (EC.2.6). Then, we have

$$\boldsymbol{y}_{i,j}^\top \boldsymbol{\vartheta}_{i,j}(\varepsilon, \boldsymbol{p}, \alpha) = -\varepsilon - \alpha < -\varepsilon, \tag{EC.4.15}$$

which satisfies the condition for being a parameter in an alternative set as defined in equation (EC.4.10). Under the Gaussian distribution assumption, *i.e.*, $\mu_i \sim \mathcal{N}(\boldsymbol{a}_i^\top \boldsymbol{\theta}, 1)$, the KL divergence between the mean value associated with the true parameter $\boldsymbol{\theta}$ and that associated with the alternative parameter $\boldsymbol{\vartheta}_{i,j}$ is

$$\begin{aligned}\mathrm{KL}(\boldsymbol{a}_i^\top \boldsymbol{\theta}, \boldsymbol{a}_i^\top \boldsymbol{\vartheta}_{i,j}) &= \frac{\left(\boldsymbol{a}_i^\top \left(\boldsymbol{\theta} - \boldsymbol{\vartheta}_{i,j}(\varepsilon, \boldsymbol{p}, \alpha)\right)\right)^2}{2(1)^2} \\ &= \boldsymbol{y}_{i,j}^\top \boldsymbol{V}_{\boldsymbol{p}}^{-1} \frac{\left(\boldsymbol{y}_{i,j}^\top \boldsymbol{\theta} + \varepsilon + \alpha\right)^2 \boldsymbol{a}_i \boldsymbol{a}_i^\top}{2\left(\|\boldsymbol{y}_{i,j}\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}}\right)^2} \boldsymbol{V}_{\boldsymbol{p}}^{-1} \boldsymbol{y}_{i,j}. \end{aligned} \tag{EC.4.16}$$

The last equation follows from substituting the expression for $\boldsymbol{\vartheta}_{i,j}$ given in (EC.4.12). Then, by Proposition 2 and the definition of the $C_x$ function in equation (EC.2.5), the lower bound can be expressed as

$$\begin{aligned}\frac{\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta]}{\log(1/2.4\delta)} &\geq \min_{\boldsymbol{p} \in S_K} \max_{\boldsymbol{\vartheta} \in \mathrm{Alt}(\boldsymbol{\mu})} \frac{1}{\sum_{n=1}^K p_n \mathrm{KL}(\boldsymbol{a}_n^\top \boldsymbol{\theta}, \boldsymbol{a}_n^\top \boldsymbol{\vartheta})} \\ &\geq \min_{\boldsymbol{p} \in S_K} \max_{x \in \mathcal{X}} \sup_{\alpha > 0} \frac{1}{\sum_{n=1}^K p_n \mathrm{KL}(\boldsymbol{a}_n^\top \boldsymbol{\theta}, \boldsymbol{a}_n^\top \boldsymbol{\vartheta}_{i,j}(\varepsilon, \boldsymbol{p}, \alpha))} \\ &= \min_{\boldsymbol{p} \in S_K} \max_{x \in \mathcal{X}} \frac{2\|\boldsymbol{y}_{i,j}\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}}}{\left(\boldsymbol{y}_{i,j}^\top \boldsymbol{\theta} + \varepsilon\right)^2}. \end{aligned} \tag{EC.4.17}$$

Note that we let $\alpha \to 0$ establish the result by realizing $\sup_{\alpha > 0}$. From this lower bound, we can also define the $C_x$ function for all $\varepsilon$-best arms identification in linear bandits, which is

$$C_{i,j}(\boldsymbol{p}) = \frac{\left(\boldsymbol{y}_{i,j}^\top \boldsymbol{\theta} + \varepsilon\right)^2}{2\|\boldsymbol{y}_{i,j}\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}}}. \tag{EC.4.18}$$

For the second part, *i.e.*, $\mathrm{Alt}_{m,\ell}(\boldsymbol{\mu})$, we assume that the mean value of arm 1 (*i.e.*, the best arm) remains fixed. This assumption partially sacrifices the completeness of the alternative set but allows the alternative set to be expressed in an explicit and symmetric form. This form remains tight.

Consequently, the culprit set is redefined as $\mathcal{X}(\boldsymbol{\mu}) = \{(i,j,m) : i \in G_\varepsilon(\boldsymbol{\mu}), j \neq i, m \notin G_\varepsilon(\boldsymbol{\mu})\}$ and the alternative set can be decomposed as $\mathrm{Alt}_x(\boldsymbol{\mu}) = \mathrm{Alt}_{i,j}(\boldsymbol{\mu}) \cup \mathrm{Alt}_m(\boldsymbol{\mu})$ for a given culprit $x = (i,j,m)$. Different from equation (EC.4.11), the second part of the alternative set becomes

$$\mathrm{Alt}_m(\boldsymbol{\mu}) = \{\boldsymbol{\vartheta} : \langle \boldsymbol{\vartheta}, \boldsymbol{a}_1 - \boldsymbol{a}_m \rangle < -\varepsilon\} \text{ for all } x \in \mathcal{X}(\boldsymbol{\mu}). \tag{EC.4.19}$$

Similarly, $\mathrm{Alt}_m(\boldsymbol{\mu})$ can be constructed as the set in equation (EC.4.12), given by

$$\mathrm{Alt}_m(\boldsymbol{\mu}) = \left\{ \boldsymbol{\vartheta}_m(\varepsilon, \boldsymbol{p}, \alpha) \mid \boldsymbol{\vartheta}_m(\varepsilon, \boldsymbol{p}, \alpha) = \boldsymbol{\theta} - \frac{\boldsymbol{y}_m^\top \boldsymbol{\theta} + \varepsilon + \alpha}{\|\boldsymbol{y}_m\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}}} \boldsymbol{V}_{\boldsymbol{p}}^{-1} \boldsymbol{y}_m \right\}, \tag{EC.4.20}$$

where $\boldsymbol{y}_m = \boldsymbol{a}_1 - \boldsymbol{a}_m$ and $\alpha > 0$. Then we have

$$\boldsymbol{y}_m^\top \boldsymbol{\vartheta}_m(\varepsilon, \boldsymbol{p}, \alpha) = \varepsilon - \alpha < \varepsilon. \tag{EC.4.21}$$

Hence, by following a derivation analogous to that of the first part, we obtain

$$C_m(\boldsymbol{p}) = \frac{\left(\boldsymbol{y}_m^\top \boldsymbol{\theta} - \varepsilon\right)^2}{2\|\boldsymbol{y}_m\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}}}. \tag{EC.4.22}$$

Deriving the tightest lower bound focuses on constructing alternative bandit instances that thoroughly challenge each $\varepsilon$-best arm configuration.

The minimum of the information function $C_x$ over all culprits $x \in \mathcal{X}(\boldsymbol{\mu})$ is then given by:

$$\min\left\{\min_{x\in\mathcal{X}} C_{i,j}(\boldsymbol{p}), \min_{x\in\mathcal{X}} C_m(\boldsymbol{p})\right\}. \tag{EC.4.23}$$

Then, by Proposition 2 and the definition of $C_x$ function, combining equation (EC.4.18) and equation (EC.4.22), the final lower bound can be expressed as

$$\frac{\mathbb{E}_{\boldsymbol{\mu}}[\tau_\delta]}{\log(1/2.4\delta)} \geq \min_{\boldsymbol{p}\in S_K} \max_{(i,j,m)\in\mathcal{X}} \max\left\{\frac{2\|\boldsymbol{y}_{i,j}\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}}}{\left(\boldsymbol{y}_{i,j}^\top \boldsymbol{\theta} + \varepsilon\right)^2}, \frac{2\|\boldsymbol{y}_m\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}}}{\left(\boldsymbol{y}_m^\top \boldsymbol{\theta} - \varepsilon\right)^2}\right\} \tag{EC.4.24}$$

$$= \min_{\boldsymbol{p}\in S_K} \max_{(i,j,m)\in\mathcal{X}} \max\left\{\frac{2\|\boldsymbol{a}_i - \boldsymbol{a}_j\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}}}{\left(\boldsymbol{a}_i^\top \boldsymbol{\theta} - \boldsymbol{a}_j^\top \boldsymbol{\theta} + \varepsilon\right)^2}, \frac{2\|\boldsymbol{a}_1 - \boldsymbol{a}_m\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}}}{\left(\boldsymbol{a}_1^\top \boldsymbol{\theta} - \boldsymbol{a}_m^\top \boldsymbol{\theta} - \varepsilon\right)^2}\right\}. \tag{EC.4.25}$$

□

## EC.5. Proof of Theorem 2

The proof of Theorem 2 consists of two parts. First, we establish the upper bound given in equation (EC.5.37). Then, we derive the final refined bound in equation (26), removing the summation in our first result.

To establish the result involving the summation in equation (EC.5.37), we define $R_{\max} = \min\{r : G_r = G_\varepsilon\}$ as the round in which the last $\varepsilon$-best arm is added to $G_r$. We divide the total number of samples into two phases: samples collected up to round $R_{\max}$ and those collected from round $R_{\max}+1$ until termination (if the algorithm does not stop at $R_{\max}$). The proof proceeds in eight steps, as outlined below.

### EC.5.1. Preliminary: Clean Events $\mathcal{E}_1$ and $\mathcal{E}_2$

We begin by defining two high-probability events, denoted as $\mathcal{E}_1$ and $\mathcal{E}_2$, which we refer to as clean events.

$$\mathcal{E}_1 = \left\{\bigcap_{i\in\mathcal{A}_I(r-1)} \bigcap_{r\in\mathbb{N}} |\hat{\mu}_i(r) - \mu_i| \leq C_{\delta/K}(r)\right\}. \tag{EC.5.1}$$

This event captures the condition that, for each round $r$, the estimated means $\hat{\mu}_i(r)$ of all arms $i$ in the active set $\mathcal{A}_I(r-1)$ lie within a confidence bound $C_{\delta/K}(r)$ of their true means $\mu_i$. It ensures uniform control over estimation errors across all arms and rounds with a prescribed confidence level.

Then, we introduce the following lemma to provide the confidence region for the estimated parameter $\hat{\boldsymbol{\theta}}$.

LEMMA EC.5.1 **(Lattimore and Szepesvári (2020))**. *Let confidence level $\delta \in (0,1)$, for each arm $\boldsymbol{a} \in \mathcal{A}$, we have*

$$\mathbb{P}\left\{\left|\left\langle \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}, \boldsymbol{a} \right\rangle\right| \geq \sqrt{2\left\|\boldsymbol{a}\right\|^2_{\boldsymbol{V}_t^{-1}} \log\left(\frac{2}{\delta}\right)}\right\} \leq \delta, \tag{EC.5.2}$$

*where $\boldsymbol{V}_t$ represents the information matrix defined in Section 2.2.*

Let $\pi_r$ denote the optimal allocation proportion computed for round $r$. The corresponding sampling budget is then determined according to equation (21). Thus, we obtain the information matrix in each round, denoted as $\boldsymbol{V}_r$[6].

$$\boldsymbol{V}_r = \sum_{\boldsymbol{a} \in \mathrm{Supp}(\pi_r)} T_r(\boldsymbol{a}) \boldsymbol{a}\boldsymbol{a}^\top \succeq^{7} \frac{2d}{\varepsilon_r^2} \log\left(\frac{2Kr(r+1)}{\delta}\right) \boldsymbol{V}(\pi). \tag{EC.5.3}$$

Then, by applying Lemma EC.5.1 with $\delta$ replaced by $\delta/(Kr(r+1))$, we obtain that for any arm $\boldsymbol{a} \in \mathcal{A}(r-1)$, with probability at least $1 - \delta/(Kr(r+1))$, we have

$$\begin{aligned}
\left|\left\langle \hat{\boldsymbol{\theta}}_r - \boldsymbol{\theta}, \boldsymbol{a} \right\rangle\right| &\leq \sqrt{2\left\|\boldsymbol{a}\right\|^2_{\boldsymbol{V}_r^{-1}} \log\left(\frac{2Kr(r+1)}{\delta}\right)} \\
&= \sqrt{2\boldsymbol{a}^\top \boldsymbol{V}_r^{-1} \boldsymbol{a} \log\left(\frac{2Kr(r+1)}{\delta}\right)} \\
&\leq \sqrt{2\boldsymbol{a}^\top \left(\frac{\varepsilon_r^2}{2d} \frac{1}{\log\left(\frac{2Kr(r+1)}{\delta}\right)} \boldsymbol{V}(\pi)^{-1}\right) \boldsymbol{a} \log\left(\frac{2Kr(r+1)}{\delta}\right)} \\
&\leq \varepsilon_r,
\end{aligned} \tag{EC.5.4}$$

where the third line follows from the matrix inequality in (EC.5.3), using the auxiliary result in Lemma EC.13.1, and the final line is derived from Lemma EC.13.3.

Thus, by applying the standard result of the G-optimal design in equation (EC.5.4), we define the confidence radius associated with the clean event $\mathcal{E}_1$ as

$$C_{\delta/K}(r) \coloneqq \varepsilon_r. \tag{EC.5.5}$$

[6] $\boldsymbol{V}_r$ and $\boldsymbol{V}_t$ can be used interchangeably when the context is unambiguous.

[7] The symbol $\succeq$ represents the relative magnitude relationship between matrices. Specifically, for any matrix $\boldsymbol{A}$, $\boldsymbol{A} \succeq 0$ means that the matrix $\boldsymbol{A}$ is positive semi-definite. $\boldsymbol{A} \succeq \boldsymbol{B}$ equals $\boldsymbol{A} - \boldsymbol{B} \succeq 0$.

Then, we have

$$\begin{aligned}\mathbb{P}(\mathcal{E}_1^c) &= \mathbb{P}\left\{\bigcup_{i\in\mathcal{A}_I(r-1)}\bigcup_{r\in\mathbb{N}}|\hat{\mu}_i(r)-\mu_i| > C_{\delta/K}(r)\right\}\\ &\le \sum_{r=1}^{\infty}\mathbb{P}\left\{\bigcup_{i\in\mathcal{A}_I(r-1)}|\hat{\mu}_i(r)-\mu_i| > \varepsilon_r\right\}\\ &\le \sum_{r=1}^{\infty}\sum_{i=1}^{K}\frac{\delta}{Kr(r+1)}\\ &= \delta, \end{aligned} \tag{EC.5.6}$$

where the second line follows from the union bound, and the third line combines the union bound with equation (EC.5.4). Therefore, we obtain

$$P(\mathcal{E}_1) \ge 1-\delta. \tag{EC.5.7}$$

Now consider another event, $\mathcal{E}_2$, which characterizes the gaps between different arms, defined as

$$\mathcal{E}_2 = \left\{\bigcap_{i\in G_\varepsilon}\bigcap_{j\in\mathcal{A}_I(r-1)}\bigcap_{r\in\mathbb{N}}|(\hat{\mu}_j(r)-\hat{\mu}_i(r))-(\mu_j-\mu_i)| \le 2\varepsilon_r\right\}. \tag{EC.5.8}$$

This event ensures the gap between the estimated mean rewards of arms $j$ and $i$ is uniformly close to their true gap for all rounds $r$, arms $i \in G_\varepsilon$, and arms $j$ in the active set $\mathcal{A}_I(r-1)$.

By (EC.5.4), for $i, j \in \mathcal{A}_I(r-1)$, we have

$$\begin{aligned}\mathbb{P}\left\{|(\hat{\mu}_j-\hat{\mu}_i)-(\mu_j-\mu_i)| > 2\varepsilon_r \mid \mathcal{E}_1\right\} &\le \mathbb{P}\left\{|\hat{\mu}_j-\mu_j|+|\hat{\mu}_i-\mu_i| > 2\varepsilon_r \mid \mathcal{E}_1\right\}\\ &\le \mathbb{P}\left\{|\hat{\mu}_j-\mu_j| > \varepsilon_r \mid \mathcal{E}_1\right\}+\mathbb{P}\left\{|\hat{\mu}_i-\mu_i| > \varepsilon_r \mid \mathcal{E}_1\right\}\\ &= 0, \end{aligned} \tag{EC.5.9}$$

which means

$$\mathbb{P}(\mathcal{E}_2 \mid \mathcal{E}_1) = 1. \tag{EC.5.10}$$

### EC.5.2. Step 1: Correctness

Recall that $G_\varepsilon$ denotes the true set of $\varepsilon$-best arms, and $G_r$ is the empirical good set identified by LinFACT-G in round $r$. Under event $\mathcal{E}_1$, we first show that if there exists a round $r$ such that $G_r \cup B_r = [K]$, then it must be that $G_r = G_\varepsilon$. This implies that under the clean event $\mathcal{E}_1$, the stopping condition of LinFACT-G ensures correct identification of $G_\varepsilon$.

Lemma EC.5.2. *Under event $\mathcal{E}_1$, we have $G_r \subseteq G_\varepsilon$ for all rounds $r \in \mathbb{N}$.*

*Proof.* We first show that $1 \in \mathcal{A}_I(r)$ for all $r \in \mathbb{N}$; that is, the best arm is never eliminated from the active set $\mathcal{A}(r-1)$ in any round $r$ on the event $\mathcal{E}_1$. For any arm $i$, we have

$$\hat{\mu}_1(r) + \varepsilon_r \geq \mu_1 \geq \mu_i \geq \hat{\mu}_i(r) - \varepsilon_r > \hat{\mu}_i(r) - \varepsilon_r - \varepsilon, \tag{EC.5.11}$$

which implies that $\hat{\mu}_1(r) + \varepsilon_r > \max_{i \in \mathcal{A}_I(r-1)} \hat{\mu}_i - \varepsilon_r - \varepsilon = L_r$ and $\hat{\mu}_1(r) + \varepsilon_r \geq \max_{i \in \mathcal{A}_I(r-1)} \hat{\mu}_i(r) - \varepsilon_r$. These inequalities confirm that arm 1 will not be removed from the active set $\mathcal{A}_I(r-1)$.

Secondly, we show that at all rounds $r$, $\mu_1 - \varepsilon \in [L_r, U_r]$. Since arm 1 never exists $\mathcal{A}_I(r-1)$,

$$U_r = \max_{i \in \mathcal{A}_I(r-1)} \hat{\mu}_i + \varepsilon_r - \varepsilon \geq \hat{\mu}_1(r) + \varepsilon_r - \varepsilon \geq \mu_1 - \varepsilon, \tag{EC.5.12}$$

and for any arm $i$,

$$\mu_1 - \varepsilon \geq \mu_i - \varepsilon \geq \hat{\mu}_i - \varepsilon_r - \varepsilon. \tag{EC.5.13}$$

Hence, taking the maximum over $i \in \mathcal{A}_I(r-1)$, we obtain

$$\mu_1 - \varepsilon \geq \max_{i \in \mathcal{A}_I(r-1)} \hat{\mu}_i - \varepsilon_r - \varepsilon = L_r. \tag{EC.5.14}$$

Next, we show that $G_r \subseteq G_\varepsilon$ for all $r \geq 1$. By contradiction, if $G_r \nsubseteq G_\varepsilon$, then it means that $\exists r \in \mathbb{N}$ and $\exists i \in G_\varepsilon^c \cap G_r$ such that,

$$\mu_i \geq \hat{\mu}_i - \varepsilon_r \geq U_r \geq \mu_1 - \varepsilon > \mu_i, \tag{EC.5.15}$$

which forms a contradiction. □

LEMMA EC.5.3. *Under event $\mathcal{E}_1$, we have $B_r \subseteq G_\varepsilon^c$ for all rounds $r \in \mathbb{N}$.*

*Proof.* Similarly, we proceed by contradiction. Consider the case that a good arm from $G_\varepsilon$ is added to $B_r$ for some round $r$. By definition, $B_0 = \emptyset$ and $B_{r-1} \subseteq B_r$ for all $r$. Then there must exist some $r \in \mathbb{N}$ and an $i \in G_\varepsilon$ such that $i \in B_r$ and $i \notin B_{r-1}$. Following line 6 of Algorithm 3, this occurs if and only if

$$\max_{j \in \mathcal{A}_I(r-1)} \hat{\mu}_j - \hat{\mu}_i > 2\varepsilon_r + \varepsilon. \tag{EC.5.16}$$

On the clean event $\mathcal{E}_1$, the above implies $\exists j \in \mathcal{A}_I(r-1)$ such that

$$\mu_j - \mu_i + 2\varepsilon_r \geq \hat{\mu}_j - \hat{\mu}_i \geq 2\varepsilon_r + \varepsilon, \tag{EC.5.17}$$

which yields $\mu_j - \mu_i \geq \varepsilon$, contradicting that $i \in G_\varepsilon$. □

The above Lemma EC.5.2 and Lemma EC.5.3 show that under $\mathcal{E}_1$, $G_r \cup B_r = [K]$ can lead to the result $G_r = G_\varepsilon$ and $B_r = G_\varepsilon^c$. Since $\mathbb{P}\{\mathcal{E}_1\} \geq 1 - \delta$, if LinFACT terminates, it can correctly provide the correct decision rule with a probability of least $1 - \delta$. Up to now, we have demonstrated the correctness of the algorithm's stopping rule. Then we will focus on bounding the sample complexity in the following parts.

### EC.5.3. Step 2: Total Sample Count

To bound the expected sampling budget, we decompose the total number of samples into two parts: the budget used before the round when the last arm is added to the good set $G_r$, and the budget used after this round until termination. The total number of samples drawn by the algorithm can thus be represented as

$$\begin{aligned} T &\le \sum_{r=1}^{\infty} \mathbb{1}\left[G_{r-1} \cup B_{r-1} \neq [K]\right] \sum_{\boldsymbol{a} \in \mathcal{A}(r-1)} T_r(\boldsymbol{a}) \\ &= \sum_{r=1}^{\infty} \mathbb{1}\left[G_{r-1} \neq G_\varepsilon\right] \mathbb{1}\left[G_{r-1} \cup B_{r-1} \neq [K]\right] \sum_{\boldsymbol{a} \in \mathcal{A}(r-1)} T_r(\boldsymbol{a}) \qquad \text{(EC.5.18)} \\ &\quad + \sum_{r=1}^{\infty} \mathbb{1}\left[G_{r-1} = G_\varepsilon\right] \mathbb{1}\left[G_{r-1} \cup B_{r-1} \neq [K]\right] \sum_{\boldsymbol{a} \in \mathcal{A}(r-1)} T_r(\boldsymbol{a}). \qquad \text{(EC.5.19)} \end{aligned}$$

In the following parts, we will bound these two terms separately. We begin by analyzing a single arm $i \in G_\varepsilon$, tracking which sets it belongs to as the rounds progress.

For $i \in G_\varepsilon$, let $R_i$ denote the number of rounds in which arm $i$ is sampled before being added to $G_r$ in line 8 of Algorithm 3. For $i \in G_\varepsilon^c$, let $R_i$ denote the number of rounds in which arm $i$ is sampled before being removed from $\mathcal{A}_I(r-1)$ and added to $B_r$ in line 6 of Algorithm 3. Then, by definition, $R_i$ can be expressed as

$$R_i = \min\left\{ r : \begin{cases} i \in G_r & \text{if } i \in G_\varepsilon \\ i \notin \mathcal{A}_I(r) & \text{if } i \in G_\varepsilon^c \end{cases} \right\} = \min\left\{ r : \begin{cases} \hat{\mu}_i - \varepsilon_r \ge U_r & \text{if } i \in G_\varepsilon \\ \hat{\mu}_i + \varepsilon_r \le L_r & \text{if } i \in G_\varepsilon^c \end{cases} \right\}. \qquad \text{(EC.5.20)}$$

**Bound $R_i$.** We define a helper function $h(x) = \log_2(1/|x|)$ to facilitate the proof. It can be observed that in round $r$, if $r \ge h(x)$, then $\varepsilon_r = 2^{-r} \le |x|$.

LEMMA EC.5.4. *For any $i \in G_\varepsilon$, we have $R_i \le \lceil h(0.25(\varepsilon - \Delta_i)) \rceil$, where $\Delta_i = \mu_1 - \mu_i$.*

*Proof.* Note that for $i \in G_\varepsilon$, the inequality $4\varepsilon_r < \mu_i - (\mu_1 - \varepsilon)$ holds when $r > h(0.25(\varepsilon - \Delta_i))$. This implies that for all $j \in \mathcal{A}_I(r-1)$,

$$\begin{aligned} \hat{\mu}_i - \varepsilon_r &\ge \mu_i - 2\varepsilon_r \\ &> \mu_1 + 2\varepsilon_r - \varepsilon \\ &\ge \mu_j + 2\varepsilon_r - \varepsilon \\ &\ge \hat{\mu}_j + \varepsilon_r - \varepsilon. \qquad \text{(EC.5.21)} \end{aligned}$$

Thus, in particular, $\hat{\mu}_i - \varepsilon_r > \max_{j \in \mathcal{A}_I(r-1)} \hat{\mu}_j(t) + \varepsilon_r - \varepsilon = U_r$. Therefore, we conclude that $R_i \le \lceil h(0.25(\varepsilon - \Delta_i)) \rceil$. □

After determining the latest round in which any arm $i \in G_\varepsilon$ is added to $G_r$, we define $R_{\max}$ as the round by which all good arms have been added to $G_r$, i.e., $R_{\max} \coloneqq \min\{r : G_r = G_\varepsilon\} = \max_{i \in G_\varepsilon} R_i$.

LEMMA EC.5.5. $R_{\max} \le \lceil h(0.25\alpha_\varepsilon) \rceil$.

*Proof.* Recall that $\alpha_\varepsilon = \min_{i \in G_\varepsilon} \mu_i - \mu_1 + \varepsilon = \min_{i \in G_\varepsilon} \varepsilon - \Delta_i$. By Lemma EC.5.4, $R_i \le \lceil h(0.25(\varepsilon - \Delta_i)) \rceil$ for $i \in G_\varepsilon$. Furthermore, $h(\cdot)$ is monotonically decreasing if $i \in G_\varepsilon$. Then for any $\delta > 0$, $R_{\max} = \max_{i \in G_\varepsilon} R_i \le \max_{i \in G_\varepsilon} \lceil h(0.25(\varepsilon - \Delta_i)) \rceil = \lceil h(\min_{i \in G_\varepsilon} \varepsilon - \Delta_i) \rceil = \lceil h(0.25\alpha_\varepsilon) \rceil$. □

**Bound the Total Samples up to $R_{max}$.** The total number of samples up to round $R_{\max}$ is $\sum_{r=1}^{R_{\max}} \sum_{\boldsymbol{a} \in \mathcal{A}(r-1)} T_r(\boldsymbol{a})$. By line 4 of the Algorithm 1, we have

$$T_r \le \frac{2d}{\varepsilon_r^2} \log\left(\frac{2Kr(r+1)}{\delta}\right) + \frac{d(d+1)}{2}. \tag{EC.5.22}$$

Hence,

$$\begin{aligned}
\sum_{r=1}^{R_{\max}} T_r &= \sum_{r=1}^{R_{\max}} \sum_{\boldsymbol{a} \in \mathcal{A}(r-1)} T_r(\boldsymbol{a}) \\
&\le \sum_{r=1}^{R_{\max}} \left( \frac{2d}{\varepsilon_r^2} \log\left(\frac{2Kr(r+1)}{\delta}\right) + \frac{d(d+1)}{2} \right) \\
&\le \sum_{r=1}^{R_{\max}} 2^{2r+1} d \log\left(\frac{2KR_{\max}(R_{\max}+1)}{\delta}\right) + \frac{d(d+1)}{2} R_{\max} \\
&\le c 2^{2R_{\max}+1} d \log\left(\frac{2KR_{\max}(R_{\max}+1)}{\delta}\right) + \frac{d(d+1)}{2} R_{\max},
\end{aligned} \tag{EC.5.23}$$

where $c$ is a universal constant, and recall that $R_{\max} \le \lceil h(0.25\alpha_\varepsilon) \rceil$. The second line follows from equation (EC.5.22). The third line applies a scaling argument based on the range of the summation. The last line replaces the term $2r+1$ with the largest round $R_{\max}$ and introduces a finite constant $c$.

Next, we bound two terms in equation (EC.5.18) and equation (EC.5.19) separately. The first term, which represents the samples taken before round $R_{\max}$, can be bounded in the following step.

**Bound (EC.5.18).** Recall that $R_{\max} \le \lceil h(0.25\alpha_\varepsilon) \rceil$ is the round where $G_{R_{\max}} = G_\varepsilon$. We have

$$\begin{aligned}
&\sum_{r=1}^{\infty} \mathbb{1}[G_{r-1} \ne G_\varepsilon] \mathbb{1}[G_{r-1} \cup B_{r-1} \ne [K]] \sum_{\boldsymbol{a} \in \mathcal{A}(r-1)} T_r(\boldsymbol{a}) \\
&= \sum_{r=1}^{R_{\max}} \mathbb{1}[G_{r-1} \cup B_{r-1} \ne [K]] \sum_{\boldsymbol{a} \in \mathcal{A}(r-1)} T_r(\boldsymbol{a}) \\
&\le c 2^{2R_{\max}+1} d \log\left(\frac{2KR_{\max}(R_{\max}+1)}{\delta}\right) + \frac{d(d+1)}{2} R_{\max},
\end{aligned} \tag{EC.5.24}$$

where the second line follows from the definition of $R_{\max}$, as no additional samples are collected after round $R_{\max}$. The third line follows directly from equation (EC.5.23).

**Bound (EC.5.19).** Next, we have

$$\begin{aligned}
&\sum_{r=1}^{\infty} \mathbb{1}\left[G_{r-1} = G_\varepsilon\right] \mathbb{1}\left[G_{r-1} \cup B_{r-1} \neq [K]\right] \sum_{\boldsymbol{a} \in \mathcal{A}(r-1)} T_r(\boldsymbol{a}) \\
&= \sum_{r=R_{\max}+1}^{\infty} \mathbb{1}\left[B_{r-1} \neq G_\varepsilon^c\right] \sum_{\boldsymbol{a} \in \mathcal{A}(r-1)} T_r(\boldsymbol{a}) \\
&\leq \sum_{r=R_{\max}+1}^{\infty} |G_\varepsilon^c \backslash B_{r-1}| \sum_{\boldsymbol{a} \in \mathcal{A}(r-1)} T_r(\boldsymbol{a}) \\
&= \sum_{r=R_{\max}+1}^{\infty} \sum_{i \in G_\varepsilon^c} \mathbb{1}\left[i \notin B_{r-1}\right] \sum_{\boldsymbol{a} \in \mathcal{A}(r-1)} T_r(\boldsymbol{a}) \\
&= \sum_{i \in G_\varepsilon^c} \sum_{r=R_{\max}+1}^{\infty} \mathbb{1}\left[i \notin B_{r-1}\right] \sum_{\boldsymbol{a} \in \mathcal{A}(r-1)} T_r(\boldsymbol{a}) \\
&\leq \sum_{i \in G_\varepsilon^c} \sum_{r=1}^{\infty} \mathbb{1}\left[i \notin B_{r-1}\right] \sum_{\boldsymbol{a} \in \mathcal{A}(r-1)} T_r(\boldsymbol{a}) \\
&\leq \sum_{i \in G_\varepsilon^c} \sum_{r=1}^{\infty} \mathbb{1}\left[i \notin B_{r-1}\right] \left( \frac{2d}{\varepsilon_r^2} \log\left( \frac{2Kr(r+1)}{\delta} \right) + \frac{d(d+1)}{2} \right), \qquad \text{(EC.5.25)}
\end{aligned}$$

where the second line follows from the definition of $R_{\max}$, as $G_r = G_\varepsilon$ for all rounds beyond $R_{\max}$. The third line uses the fact that as long as $G_\varepsilon^c \backslash B_{r-1}$ is not the empty set, the corresponding indicator function is 1. The fourth line considers the indicator function for each arm in $G_\varepsilon$. The fifth and sixth lines exchange the order of the double summation and enlarge the summation range over the rounds $r$.

### EC.5.4. Step 3: Bound the Expected Total Samples of LinFACT

We now take expectations over the total number of samples drawn for the given bandit instance $\boldsymbol{\mu}$. These expectations are conditioned on the high-probability event $\mathcal{E}_1$.

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\mu}}\left[T_G \,|\, \mathcal{E}_1\right] &\leq \sum_{r=1}^{\infty} \mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[G_r \cup B_r \neq [K]\right] | \, \mathcal{E}_1\right] \sum_{\boldsymbol{a} \in \mathcal{A}(r-1)} T_r(\boldsymbol{a}) \\
&= \sum_{r=1}^{\infty} \mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[G_{r-1} \neq G_\varepsilon\right] \mathbb{1}\left[G_{r-1} \cup B_{r-1} \neq [K]\right] | \, \mathcal{E}_1\right] \sum_{\boldsymbol{a} \in \mathcal{A}(r-1)} T_r(\boldsymbol{a}) \\
&+ \sum_{r=1}^{\infty} \mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[G_{r-1} = G_\varepsilon\right] \mathbb{1}\left[G_{r-1} \cup B_{r-1} \neq [K]\right] | \, \mathcal{E}_1\right] \sum_{\boldsymbol{a} \in \mathcal{A}(r-1)} T_r(\boldsymbol{a}) \\
&\overset{\text{(EC.5.24)}}{\leq} c 2^{2R_{\max}+1} d \log\left( \frac{2KR_{\max}(R_{\max}+1)}{\delta} \right) + \frac{d(d+1)}{2} R_{\max} \\
&+ \sum_{r=1}^{\infty} \mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[G_{r-1} = G_\varepsilon\right] \mathbb{1}\left[G_{r-1} \cup B_{r-1} \neq [K]\right] | \, \mathcal{E}_1\right] \sum_{\boldsymbol{a} \in \mathcal{A}(r-1)} T_r(\boldsymbol{a}) \\
&\overset{\text{(EC.5.25)}}{\leq} c 2^{2R_{\max}+1} d \log\left( \frac{2KR_{\max}(R_{\max}+1)}{\delta} \right) + \frac{d(d+1)}{2} R_{\max}
\end{aligned}$$

$$+\sum_{i\in G_\varepsilon^c}\sum_{r=1}^{\infty}\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[i\notin B_{r-1}\mid\mathcal{E}_1\right]\right]\left(\frac{2d}{\varepsilon_r^2}\log\left(\frac{2Kr(r+1)}{\delta}\right)+\frac{d(d+1)}{2}\right). \tag{EC.5.26}$$

The first line follows from the stopping condition of LinFACT-G and the additivity of the expectation. The second line applies the decomposition established in Step 2. The subsequent lines use the results from equation (EC.5.24) and equation (EC.5.25).

Next, we bound the last term. For a given $i\in G_\varepsilon^c$ and round $r$, we first bound the probability that $i\notin B_r$. By the Borel-Cantelli lemma, this implies that the probability of $i$ never being added to any $B_r$ is zero.

LEMMA EC.5.6. *For $i\in G_\varepsilon^c$ and $r\geq\left\lceil\log_2\left(\frac{4}{\Delta_i-\varepsilon}\right)\right\rceil$, we have $\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[i\notin B_r\right]\mid\mathcal{E}_1\right]=0$.*

*Proof.*

First, we have for any $i\in G_\varepsilon^c$,

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[i\notin B_r\right]\mid\mathcal{E}_1\right]&=\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[\max_{j\in\mathcal{A}_I(r-1)}\hat{\mu}_j-\hat{\mu}_i\leq 2\varepsilon_r+\varepsilon\right]\mid\mathcal{E}_1,i\notin B_m(m=\{1,2,\ldots,r-1\})\right]\\&\leq\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[\max_{j\in\mathcal{A}_I(r-1)}\hat{\mu}_j-\hat{\mu}_i\leq 2\varepsilon_r+\varepsilon\right]\mid\mathcal{E}_1\right],\end{aligned} \tag{EC.5.27}$$

where the first line follows from the fact that arm $i\in G_\varepsilon^c$ will never be added into $B_r$ from round 1 to round $r-1$ if $i\notin B_r$, and the second line accounts for the conditional expectation.

If $i\in B_{r-1}$, then $i\in B_r$ by definition. Otherwise, if $i\notin B_{r-1}$, then under event $\mathcal{E}_1$, for $i\in G_\varepsilon^c$ and $r\geq\left\lceil\log_2\left(\frac{4}{\Delta_i-\varepsilon}\right)\right\rceil$, we have

$$\max_{j\in\mathcal{A}_I(r-1)}\hat{\mu}_j-\hat{\mu}_i\geq\Delta_i-2^{-r+1}\geq\varepsilon+2\varepsilon_r, \tag{EC.5.28}$$

which implies that $i\in B_r$ by line 6 of the Algorithm 3. In other words, we have established the correctness of the algorithm when line 6 is triggered in Step 1. We now specify the exact condition under which line 6 will occur. In particular, under event $\mathcal{E}_1$, if $i\notin B_{r-1}$, then for all $i\in G_\varepsilon^c$ and $r\geq\left\lceil\log_2\left(\frac{4}{\Delta_i-\varepsilon}\right)\right\rceil$, we have

$$\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[\max_{j\in\mathcal{A}_I(r-1)}\hat{\mu}_j-\hat{\mu}_i>2\varepsilon_r+\varepsilon\right]\mid i\notin B_{r-1},\mathcal{E}_1\right]=1. \tag{EC.5.29}$$

Therefore, for all $i\in G_\varepsilon^c$ and $r\geq\left\lceil\log_2\left(\frac{4}{\Delta_i-\varepsilon}\right)\right\rceil$, we have

$$\begin{aligned}&\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[\max_{j\in\mathcal{A}_I(r-1)}\hat{\mu}_j-\hat{\mu}_i\leq 2\varepsilon_r+\varepsilon\right]\mid\mathcal{E}_1\right]\\&=\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[\max_{j\in\mathcal{A}_I(r-1)}\hat{\mu}_j-\hat{\mu}_i\leq 2\varepsilon_r+\varepsilon\right]\mathbb{1}\left[i\notin B_{r-1}\right]\mid\mathcal{E}_1\right]\\&+\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[\max_{j\in\mathcal{A}_I(r-1)}\hat{\mu}_j-\hat{\mu}_i\leq 2\varepsilon_r+\varepsilon\right]\mathbb{1}\left[i\in B_{r-1}\right]\mid\mathcal{E}_1\right]\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[\max_{j\in\mathcal{A}_I(r-1)}\hat{\mu}_j-\hat{\mu}_i\le 2\varepsilon_r+\varepsilon\right]\mathbb{1}\left[i\notin B_{r-1}\right]\mid\mathcal{E}_1\right]\\
&= \mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[\max_{j\in\mathcal{A}_I(r-1)}\hat{\mu}_j-\hat{\mu}_i\le 2\varepsilon_r+\varepsilon\right]\mathbb{1}\left[i\notin B_{r-1}\right]\mid i\notin B_{r-1},\mathcal{E}_1\right]\mathbb{P}\left(i\notin B_{r-1}\mid\mathcal{E}_1\right)\\
&+ \mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[\max_{j\in\mathcal{A}_I(r-1)}\hat{\mu}_j-\hat{\mu}_i\le 2\varepsilon_r+\varepsilon\right]\mathbb{1}\left[i\notin B_{r-1}\right]\mid i\in B_{r-1},\mathcal{E}_1\right]\mathbb{P}\left(i\in B_{r-1}\mid\mathcal{E}_1\right)\\
&= \mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[\max_{j\in\mathcal{A}_I(r-1)}\hat{\mu}_j-\hat{\mu}_i\le 2\varepsilon_r+\varepsilon\right]\mathbb{1}\left[i\notin B_{r-1}\right]\mid i\notin B_{r-1},\mathcal{E}_1\right]\mathbb{P}\left(i\notin B_{r-1}\mid\mathcal{E}_1\right)\\
&= \mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[\max_{j\in\mathcal{A}_I(r-1)}\hat{\mu}_j-\hat{\mu}_i\le 2\varepsilon_r+\varepsilon\right]\mid i\notin B_{r-1},\mathcal{E}_1\right]\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[i\notin B_{r-1}\right]\mid\mathcal{E}_1\right]\\
&= 0, \qquad \text{(EC.5.30)}
\end{aligned}
$$

where the second line follows from the additivity of expectation. The fourth line follows the deterministic result that $\mathbb{1}\left[i\notin B_r\right]\mathbb{1}\left[i\in B_{r-1}\right]=0$. The fifth line is the decomposition based on the conditional expectation. The eighth line uses the fact that the expectation of the indicator function is simply the probability. The last line follows the result in equation (EC.5.29). The lemma then follows by combining this result with equation (EC.5.27). □

Lemma EC.5.7. *For $i\in G_\varepsilon^c$, we have*

$$
\begin{aligned}
&\sum_{r=1}^{\infty}\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[i\notin B_{r-1}\right]\mid\mathcal{E}_1\right]\left(\frac{2d}{\varepsilon_r^2}\log\left(\frac{2Kr(r+1)}{\delta}\right)+\frac{d(d+1)}{2}\right)\\
&\le \frac{d(d+1)}{2}\log_2\left(\frac{8}{\Delta_i-\varepsilon}\right)+\xi\frac{256d}{(\Delta_i-\varepsilon)^2}\log\left(\frac{2K}{\delta}\log_2\frac{16}{\Delta_i-\varepsilon}\right). \qquad \text{(EC.5.31)}
\end{aligned}
$$

*Proof.*

$$
\begin{aligned}
&\sum_{r=1}^{\infty}\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[i\notin B_{r-1}\right]\mid\mathcal{E}_1\right]\left(\frac{2d}{\varepsilon_r^2}\log\left(\frac{2Kr(r+1)}{\delta}\right)+\frac{d(d+1)}{2}\right)\\
&= \sum_{r=1}^{\left\lceil\log_2\left(\frac{4}{\Delta_i-\varepsilon}\right)\right\rceil}\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[i\notin B_{r-1}\right]\mid\mathcal{E}_1\right]\left(\frac{2d}{\varepsilon_r^2}\log\left(\frac{2Kr(r+1)}{\delta}\right)+\frac{d(d+1)}{2}\right)\\
&+ \sum_{r=\left\lceil\log_2\left(\frac{4}{\Delta_i-\varepsilon}\right)\right\rceil+1}^{\infty}\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[i\notin B_{r-1}\right]\mid\mathcal{E}_1\right]\left(\frac{2d}{\varepsilon_r^2}\log\left(\frac{2Kr(r+1)}{\delta}\right)+\frac{d(d+1)}{2}\right)\\
&= \sum_{r=1}^{\left\lceil\log_2\left(\frac{4}{\Delta_i-\varepsilon}\right)\right\rceil}\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[i\notin B_{r-1}\right]\mid\mathcal{E}_1\right]\left(\frac{2d}{\varepsilon_r^2}\log\left(\frac{2Kr(r+1)}{\delta}\right)+\frac{d(d+1)}{2}\right)+0\\
&\le \sum_{r=1}^{\left\lceil\log_2\left(\frac{4}{\Delta_i-\varepsilon}\right)\right\rceil}\left(d2^{2r+1}\log\left(\frac{2Kr(r+1)}{\delta}\right)+\frac{d(d+1)}{2}\right)\\
&\le \frac{d(d+1)}{2}\log_2\left(\frac{8}{\Delta_i-\varepsilon}\right)+2d\log\left(\frac{2K}{\delta}\right)\sum_{r=1}^{\left\lceil\log_2\left(\frac{4}{\Delta_i-\varepsilon}\right)\right\rceil}2^{2r}+4d\sum_{r=1}^{\left\lceil\log_2\left(\frac{4}{\Delta_i-\varepsilon}\right)\right\rceil}2^{2r}\log(r+1)\\
&\le \frac{d(d+1)}{2}\log_2\left(\frac{8}{\Delta_i-\varepsilon}\right)+\xi\frac{256d}{(\Delta_i-\varepsilon)^2}\log\left(\frac{2K}{\delta}\log_2\frac{16}{\Delta_i-\varepsilon}\right), \qquad \text{(EC.5.32)}
\end{aligned}
$$

Here, $\xi$ is a sufficiently large universal constant. The second line follows from decomposing the summation across all rounds. The fourth line uses Lemma EC.5.6. The fifth line bounds the expectation of the indicator function to its maximum value of 1. The sixth and seventh lines replace the round $r$ with its maximum value $\left\lceil \log_2\left(\frac{4}{\Delta_i-\varepsilon}\right)\right\rceil$. □

Summarizing the aforementioned results, we have

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\mu}}\left[T_G \mid \mathcal{E}_1\right] \leq\, & c2^{2R_{\max}+1} d \log\left(\frac{2KR_{\max}(R_{\max}+1)}{\delta}\right) + \frac{d(d+1)}{2} R_{\max} \\ & + \sum_{i\in G_\varepsilon^c} \sum_{r=1}^{\infty} \mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[i \notin B_{r-1} \mid \mathcal{E}_1\right]\right] \left(\frac{2d}{\varepsilon_r^2} \log\left(\frac{2Kr(r+1)}{\delta}\right) + \frac{d(d+1)}{2}\right), \end{aligned} \tag{EC.5.33}$$

where

$$R_{\max} \leq \lceil h\left(0.25\alpha_\varepsilon\right)\rceil = \log_2 \frac{8}{\alpha_\varepsilon}. \tag{EC.5.34}$$

Also, we have

$$\begin{aligned}& \sum_{r=1}^{\infty} \mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[i \notin B_{r-1}\right] \mid \mathcal{E}_1\right] \left(\frac{2d}{\varepsilon_r^2} \log\left(\frac{2Kr(r+1)}{\delta}\right) + \frac{d(d+1)}{2}\right) \\ & \leq \frac{d(d+1)}{2} \log_2\left(\frac{8}{\Delta_i-\varepsilon}\right) + \xi \frac{256d}{(\Delta_i-\varepsilon)^2} \log\left(\frac{2K}{\delta} \log_2 \frac{16}{\Delta_i-\varepsilon}\right). \end{aligned} \tag{EC.5.35}$$

Then, we arrive at the final result as follows,

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\mu}}\left[T_G \mid \mathcal{E}_1\right] \leq\, & c2^{2R_{\max}+1} d \log\left(\frac{2KR_{\max}(R_{\max}+1)}{\delta}\right) + \frac{d(d+1)}{2} R_{\max} \\ & + \sum_{i\in G_\varepsilon^c} \sum_{r=1}^{\infty} \mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[i \notin B_{r-1} \mid \mathcal{E}_1\right]\right] \left(\frac{2d}{\varepsilon_r^2} \log\left(\frac{2Kr(r+1)}{\delta}\right) + \frac{d(d+1)}{2}\right) \\ & \leq c\frac{256d}{\alpha_\varepsilon^2} \log\left(\frac{2K}{\delta} \log_2\left(\frac{16}{\alpha_\varepsilon}\right)\right) + \frac{d(d+1)}{2} \log_2 \frac{8}{\alpha_\varepsilon} \\ & + \sum_{i\in G_\varepsilon^c} \left(\frac{d(d+1)}{2} \log_2\left(\frac{8}{\Delta_i-\varepsilon}\right) + \xi \frac{256d}{(\Delta_i-\varepsilon)^2} \log\left(\frac{2K}{\delta} \log_2 \frac{16}{\Delta_i-\varepsilon}\right)\right), \end{aligned} \tag{EC.5.36}$$

which can be further expressed as

$$\begin{aligned}\mathbb{E}[T_G \mid \mathcal{E}_1] = & \mathcal{O}\left(d\alpha_\varepsilon^{-2} \log\left(\frac{K}{\delta}\log(\alpha_\varepsilon^{-2})\right) + d^2 \log(\alpha_\varepsilon^{-1})\right) \\ & + \sum_{i\in G_\varepsilon^c} \left(\mathcal{O}\left(d^2 \log(\Delta_i-\varepsilon)^{-1} + d(\Delta_i-\varepsilon)^{-2} \log\left(\frac{K}{\delta}\log(\Delta_i-\varepsilon)^{-2}\right)\right)\right). \end{aligned} \tag{EC.5.37}$$

### EC.5.5. A Refined Bound

The result obtained in the previous steps involves a summation over the set $G_\varepsilon$, which can be further improved by eliminating this summation. Rather than focusing solely on the round $R_{\max}$, defined in Lemma EC.5.6 as the round in which all arms in $G_\varepsilon$ are classified into $G_r$, we now define the round at which all classifications are complete, *i.e.*, $G_r \cup B_r = [K]$.

LEMMA EC.5.8. *For $i \in G_\varepsilon$ and $r \geq \left\lceil \log_2\left(\frac{4}{\varepsilon-\Delta_i}\right)\right\rceil$, we have $\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[i \notin G_r\right] \mid \mathcal{E}_1\right] = 0$.*

*Proof.* First, for any $i \in G_\varepsilon$

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[i \notin G_r\right] \mid \mathcal{E}_1\right] &= \mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[\max_{j\in\mathcal{A}_I(r-1)} \hat{\mu}_j - \hat{\mu}_i \geq -2\varepsilon_r + \varepsilon\right] \mid \mathcal{E}_1, i \notin G_m (m = \{1,2,\ldots,r-1\})\right] \\ &\leq \mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[\max_{j\in\mathcal{A}_I(r-1)} \hat{\mu}_j - \hat{\mu}_i \geq -2\varepsilon_r + \varepsilon\right] \mid \mathcal{E}_1\right]. \end{aligned} \tag{EC.5.38}$$

If $i \in G_{r-1}$, then $i \in G_r$ by definition. Otherwise, if $i \notin G_{r-1}$, then under event $\mathcal{E}_1$, for $i \in G_\varepsilon$ and $r \geq \left\lceil \log_2\left(\frac{4}{\varepsilon-\Delta_i}\right)\right\rceil$, we have

$$\max_{j\in\mathcal{A}_I(r-1)} \hat{\mu}_j - \hat{\mu}_i \leq \mu_{\arg\max_{j\in\mathcal{A}_I(r-1)}\hat{\mu}_j} - \mu_i + 2^{-r+1} \leq \Delta_i + 2^{-r+1} \leq \varepsilon - 2\varepsilon_r, \tag{EC.5.39}$$

which implies that $i \in G_r$ by line 8 of the Algorithm 3. In other words, we now specify the exact condition under which line 8 will occur. In particular, under event $\mathcal{E}_1$, if $i \notin G_{r-1}$, for all $i \in G_\varepsilon$ and $r \geq \left\lceil \log_2\left(\frac{4}{\varepsilon-\Delta_i}\right)\right\rceil$, we have

$$\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[\max_{j\in\mathcal{A}_I(r-1)} \hat{\mu}_j - \hat{\mu}_i \leq -2\varepsilon_r + \varepsilon\right] \mid i \notin G_{r-1}, \mathcal{E}_1\right] = 1. \tag{EC.5.40}$$

Deterministically, $\mathbb{1}\left[i \notin G_r\right]\mathbb{1}\left[i \in G_{r-1}\right] = 0$. Therefore,

$$\begin{aligned}&\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[\max_{j\in\mathcal{A}_I(r-1)} \hat{\mu}_j - \hat{\mu}_i \geq -2\varepsilon_r + \varepsilon\right] \mid \mathcal{E}_1\right] \\ &= \mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[\max_{j\in\mathcal{A}_I(r-1)} \hat{\mu}_j - \hat{\mu}_i \geq -2\varepsilon_r + \varepsilon\right]\mathbb{1}\left[i \notin G_{r-1}\right] \mid \mathcal{E}_1\right] \\ &+ \mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[\max_{j\in\mathcal{A}_I(r-1)} \hat{\mu}_j - \hat{\mu}_i \geq -2\varepsilon_r + \varepsilon\right]\mathbb{1}\left[i \in G_{r-1}\right] \mid \mathcal{E}_1\right] \\ &= \mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[\max_{j\in\mathcal{A}_I(r-1)} \hat{\mu}_j - \hat{\mu}_i \geq -2\varepsilon_r + \varepsilon\right]\mathbb{1}\left[i \notin G_{r-1}\right] \mid \mathcal{E}_1\right] \\ &= \mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[\max_{j\in\mathcal{A}_I(r-1)} \hat{\mu}_j - \hat{\mu}_i \geq -2\varepsilon_r + \varepsilon\right]\mathbb{1}\left[i \notin G_{r-1}\right] \mid i \notin G_{r-1}, \mathcal{E}_1\right]\mathbb{P}\left(i \notin G_{r-1} \mid \mathcal{E}_1\right) \\ &+ \mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[\max_{j\in\mathcal{A}_I(r-1)} \hat{\mu}_j - \hat{\mu}_i \geq -2\varepsilon_r + \varepsilon\right]\mathbb{1}\left[i \notin G_{r-1}\right] \mid i \in G_{r-1}, \mathcal{E}_1\right]\mathbb{P}\left(i \in G_{r-1} \mid \mathcal{E}_1\right) \\ &= \mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[\max_{j\in\mathcal{A}_I(r-1)} \hat{\mu}_j - \hat{\mu}_i \geq -2\varepsilon_r + \varepsilon\right]\mathbb{1}\left[i \notin G_{r-1}\right] \mid i \notin G_{r-1}, \mathcal{E}_1\right]\mathbb{P}\left(i \notin G_{r-1} \mid \mathcal{E}_1\right) \\ &= \mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[\max_{j\in\mathcal{A}_I(r-1)} \hat{\mu}_j - \hat{\mu}_i \geq -2\varepsilon_r + \varepsilon\right] \mid i \notin G_{r-1}, \mathcal{E}_1\right]\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[i \notin G_{r-1}\right] \mid \mathcal{E}_1\right] \\ &= 0, \end{aligned} \tag{EC.5.41}$$

where the second line comes from the additivity of expectation. The fourth line follows the deterministic result that $\mathbb{1}\left[i \notin G_r\right]\mathbb{1}\left[i \in G_{r-1}\right] = 0$. The fifth line applies a decomposition based on the conditional expectation. The eighth line uses the fact that the expectation of an indicator function equals the corresponding probability. The final line follows from the result in equation (EC.5.40). The lemma then follows by combining this result with equation (EC.5.38). □

LEMMA EC.5.9. *The round at which all classifications are complete and the final answer is returned is* $R_{\text{upper}} = \max\left\{\left\lceil \log_2 \frac{4}{\alpha_\varepsilon}\right\rceil, \left\lceil \log_2 \frac{4}{\beta_\varepsilon}\right\rceil\right\}$.

*Proof.* Combining the result of Lemma EC.5.8 with that of Lemma EC.5.6, we have the following: for $i \in G_\varepsilon^c$ and $r \geq \left\lceil \log_2\left(\frac{4}{\Delta_i - \varepsilon}\right)\right\rceil$, we have $\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[i \notin B_r\right] \mid \mathcal{E}_1\right] = 0$; for $i \in G_\varepsilon$ and $r \geq \left\lceil \log_2\left(\frac{4}{\varepsilon - \Delta_i}\right)\right\rceil$, we have $\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[i \notin G_r\right] \mid \mathcal{E}_1\right] = 0$. Since $\alpha_\varepsilon = \min_{i \in G_\varepsilon}(\varepsilon - \Delta_i)$ and $\beta_\varepsilon = \min_{i \in G_\varepsilon^c}(\Delta_i - \varepsilon)$, for any round $r \geq R_{\text{upper}}$, all arms have been classified into either $G_r$ or $B_r$, marking the termination of the algorithm. □

LEMMA EC.5.10. *For the expected sample complexity conditioned on the high-probability event* $\mathcal{E}_1$*, we have*

$$\mathbb{E}_{\boldsymbol{\mu}}\left[T_{\mathcal{X}\mathcal{Y}} \mid \mathcal{E}_1\right] \leq \zeta \max\left\{\frac{256d}{\alpha_\varepsilon^2}\log\left(\frac{2K}{\delta}\log_2\frac{16}{\alpha_\varepsilon}\right), \frac{256d}{\beta_\varepsilon^2}\log\left(\frac{2K}{\delta}\log_2\frac{16}{\beta_\varepsilon}\right)\right\} + \frac{d(d+1)}{2}R_{\text{upper}}, \tag{EC.5.42}$$

*where* $\zeta$ *is a universal constant and* $R_{\text{upper}} = \max\left\{\left\lceil \log_2 \frac{4}{\alpha_\varepsilon}\right\rceil, \left\lceil \log_2 \frac{4}{\beta_\varepsilon}\right\rceil\right\}$.

*Proof.* We have

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\mu}}\left[T_{\mathcal{X}\mathcal{Y}} \mid \mathcal{E}_1\right] &\leq \sum_{r=1}^{\infty} \mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[G_r \cup B_r \neq [K]\right] \mid \mathcal{E}_1\right] \sum_{\boldsymbol{a} \in \mathcal{A}(r-1)} T_r(\boldsymbol{a}) \\
&\leq \sum_{r=1}^{R_{\text{upper}}} \left(d2^{2r+1}\log\left(\frac{2Kr(r+1)}{\delta}\right) + \frac{d(d+1)}{2}\right) \\
&\leq \frac{d(d+1)}{2}R_{\text{upper}} + 2d\log\left(\frac{2K}{\delta}\right)\sum_{r=1}^{R_{\text{upper}}} 2^{2r} + 4d\sum_{r=1}^{R_{\text{upper}}} 2^{2r}\log(r+1) \\
&\leq 4\log\left[\frac{2K}{\delta}\left(R_{\text{upper}}+1\right)\right]\sum_{r=1}^{R_{\text{upper}}} d2^{2r} + \frac{d(d+1)}{2}R_{\text{upper}} \qquad \text{(EC.5.43)} \\
&\leq \zeta \max\left\{\frac{256d}{\alpha_\varepsilon^2}\log\left(\frac{2K}{\delta}\log_2\frac{16}{\alpha_\varepsilon}\right), \frac{256d}{\beta_\varepsilon^2}\log\left(\frac{2K}{\delta}\log_2\frac{16}{\beta_\varepsilon}\right)\right\} + \frac{d(d+1)}{2}R_{\text{upper}}. \qquad \text{(EC.5.44)}
\end{aligned}$$

The second line follows from Lemma EC.5.9. Then, we have

$$\mathbb{E}[T_G \mid \mathcal{E}] = \mathcal{O}\left(d\xi^{-2}\log\left(\frac{K}{\delta}\log_2(\xi^{-2})\right) + d^2\log(\xi^{-1})\right), \tag{EC.5.45}$$

where $\xi = \min(\alpha_\varepsilon, \beta_\varepsilon)/16$ is the minimum gap between $\alpha_\varepsilon$ and $\beta_\varepsilon$, indicating the difficulty of the problem instance.

□

## EC.6. Additional Insights into the Algorithm Optimality

From another perspective, we consider the relationship between the lower bound and the upper bound in the following section and give some additional insights into the algorithm optimality. This relationship serves as the basis for the derivation of a near-optimal upper bound in Theorem 3.

For $\forall i \in (\mathcal{A}_I(r-1) \cap G_\varepsilon)$ and $\forall j \in \mathcal{A}_I(r-1)$, $j \neq i$, in round $r$, we have

$$\boldsymbol{y}_{j,i}^\top(\hat{\boldsymbol{\theta}}_r - \boldsymbol{\theta}) \leq 2\varepsilon_r \tag{EC.6.1}$$

and

$$\boldsymbol{y}_{j,i}^\top \hat{\boldsymbol{\theta}}_r - \varepsilon \leq \boldsymbol{y}_{j,i}^\top \boldsymbol{\theta} + 2\varepsilon_r - \varepsilon. \tag{EC.6.2}$$

LEMMA EC.6.1. *Define* $G_r' \coloneqq \left\{\exists j \in \mathcal{A}_I(r-1), j \neq i, i : \boldsymbol{y}_{j,i}^\top \boldsymbol{\theta} - \varepsilon > -4\varepsilon_r\right\}$. *We always have* $(\mathcal{A}_I(r-1) \cap G_\varepsilon \cap G_r^c) \subseteq G_r'$.

*Proof.* For $r = 1$, the lemma follows directly from the assumption in Theorem 3 that $\max_{i\in[K]} |\mu_1 - \varepsilon - \mu_i| \leq 2$. For $r \geq 2$, we proceed by contradiction. Suppose $i \in (\mathcal{A}_I(r-1) \cap G_\varepsilon \cap G_r^c) \cap (G_r')^c$, then for every $j \in \mathcal{A}_I(r-1)$ with $j \neq i$, we have

$$\boldsymbol{y}_{j,i}^\top \boldsymbol{\theta} \leq -4\varepsilon_r + \varepsilon. \tag{EC.6.3}$$

Hence, using equation (EC.6.2), for every $j \in \mathcal{A}_I(r-1)$ and $j \neq i$, we have

$$\boldsymbol{y}_{j,i}^\top \hat{\boldsymbol{\theta}}_r - \varepsilon \leq -2\varepsilon_r, \tag{EC.6.4}$$

which is exactly the condition for the algorithm to add arm $i$ into $G_r$ at line 8 of the algorithm, which yields a contradiction and completes the argument. Moreover, note that when $r \geq \left\lceil \log_2 \frac{4}{\alpha_\varepsilon} \right\rceil$, we have $G_r' \cap G_\varepsilon = \emptyset$. Furthermore, considering that $i \in G_\varepsilon$, we have $\boldsymbol{y}_{i,j}^\top \boldsymbol{\theta} + \varepsilon > 0$. □

There is, however, one exceptional case that invalidates the above derivation. Specifically, when $i$ is the index of the arm with the largest mean value, *i.e.*, $i = \arg\max_{i\in\mathcal{A}_I(r-1)}$, the proof no longer holds. For this situation to occur, it must be the case that $\arg\max_{i\in\mathcal{A}_I(r-1)} \in G_r^c$, which is equivalent to $\varepsilon \leq 2\varepsilon_r$. Since this condition can only hold for a limited number of rounds, its impact is negligible and can therefore be ignored.

On the other hand, for $i \in (\mathcal{A}_I(r-1) \cap G_\varepsilon^c)$, in any round $r$, we have

$$\boldsymbol{y}_{1,i}^\top(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_r) \leq 2\varepsilon_r \tag{EC.6.5}$$

and

$$\boldsymbol{y}_{1,i}^\top \hat{\boldsymbol{\theta}}_r - \varepsilon \geq \boldsymbol{y}_{1,i}^\top \boldsymbol{\theta} - 2\varepsilon_r - \varepsilon. \tag{EC.6.6}$$

LEMMA EC.6.2. *Define* $B'_r := \{i : \boldsymbol{y}_{1,i}^\top \boldsymbol{\theta} - \varepsilon < 4\varepsilon_r\}$. *We always have* $(\mathcal{A}_I(r-1) \cap G_\varepsilon^c) \subset B'_r$.

*Proof.* To establish this, first note that when $r = 1$, the lemma directly follows from the assumption in Theorem 3 that $\max_{i\in[K]} |\mu_1 - \varepsilon - \mu_i| \le 2$. For $r \ge 2$, using the same contradiction argument, assume that $i \in (\mathcal{A}_I(r-1) \cap G_\varepsilon^c) \cap (B'_r)^c$. Then we must have

$$\boldsymbol{y}_{1,i}^\top \boldsymbol{\theta} \ge 4\varepsilon_r + \varepsilon. \tag{EC.6.7}$$

Consequently, by equation (EC.6.6), it follows that

$$\boldsymbol{y}_{1,i}^\top \hat{\boldsymbol{\theta}}_r - \varepsilon \ge 2\varepsilon_r. \tag{EC.6.8}$$

This is precisely the condition for the algorithm to add arm $i$ into $B_r$ and eliminate it from $\mathcal{A}_I(r-1)$ as specified in line 6 of the algorithm. This contradiction leads to the desired result. Moreover, when $r \ge \left\lceil \log_2 \frac{4}{\beta_\varepsilon} \right\rceil$, we have $B'_r \cap G_\varepsilon^c = \emptyset$. Finally, for $i \in G_\varepsilon^c$, it follows that $\boldsymbol{y}_{1,i}^\top \boldsymbol{\theta} - \varepsilon > 0$. □

We now present a critical lemma that establishes the connection between the lower bound in Theorem 1 and the upper bound. Define $\mathcal{G}_\mathcal{Y}$ as the gauge of $\mathcal{Y}(\mathcal{A})$, where $\mathcal{A}$ denotes the initial set of all arm vectors. The details are provided in Lemma EC.13.2.

LEMMA EC.6.3. *Considering the lower bound* $(\Gamma^*)^{-1}$ *in Theorem 1, we have*

$$(\Gamma^*)^{-1} \ge \frac{{\mathcal{G}_\mathcal{Y}}^2 L_2}{4R_{\text{upper}} d L_1} \sum_{r=1}^{R_{\text{upper}}} 2^{2r-3} \min_{\boldsymbol{p}\in S_K} \max_{i,j\in\mathcal{A}_I(r-1)} \|\boldsymbol{y}_{i,j}\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}}, \tag{EC.6.9}$$

*where* $R_{\text{upper}} = \max\left\{\left\lceil \log_2 \frac{4}{\alpha_\varepsilon} \right\rceil, \left\lceil \log_2 \frac{4}{\beta_\varepsilon} \right\rceil\right\}$, *and* $L_1$ *and* $L_2$ *are constants.*

*Proof.* When round $r$ exceeds $R_{\text{upper}}$, we have $G'_r \cap G_\varepsilon = \emptyset$ and $B'_r \cap G_\varepsilon^c = \emptyset$, implying that $G_r \cup B_r = [K]$ and the algorithm terminates. From Theorem 1, we obtain

$$\begin{aligned}
(\Gamma^*)^{-1} &= \min_{\boldsymbol{p}\in S_K} \max_{(i,j,m)\in\mathcal{X}} \max\left\{ \frac{2\|\boldsymbol{y}_{i,j}\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}}}{\left(\boldsymbol{y}_{i,j}^\top \boldsymbol{\theta} + \varepsilon\right)^2}, \frac{2\|\boldsymbol{y}_{1,m}\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}}}{\left(\boldsymbol{y}_{1,m}^\top \boldsymbol{\theta} - \varepsilon\right)^2} \right\} \\
&= \min_{\boldsymbol{p}\in S_K} \max_{r\le R_{\text{upper}}} \max_{i\in G'_r\cap G_\varepsilon} \max_{\substack{j\in\mathcal{A}_I(r-1)\\ j\ne i}} \max_{m\in B'_r\cap G_\varepsilon^c} \max\left\{ \frac{2\|\boldsymbol{y}_{i,j}\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}}}{\left(\boldsymbol{y}_{i,j}^\top \boldsymbol{\theta} + \varepsilon\right)^2}, \frac{2\|\boldsymbol{y}_{1,m}\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}}}{\left(\boldsymbol{y}_{1,m}^\top \boldsymbol{\theta} - \varepsilon\right)^2} \right\} \\
&\ge \min_{\boldsymbol{p}\in S_K} \max_{r\le R_{\text{upper}}} \max_{i\in G'_r\cap G_\varepsilon} \max_{\substack{j\in\mathcal{A}_I(r-1)\\ j\ne i}} \max_{m\in B'_r\cap G_\varepsilon^c} \max\left\{ \frac{2\|\boldsymbol{y}_{i,j}\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}}}{\left(4\varepsilon_r\right)^2}, \frac{2\|\boldsymbol{y}_{1,m}\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}}}{\left(4\varepsilon_r\right)^2} \right\} \\
&\overset{\text{(i)}}{\ge} \frac{1}{R_{\text{upper}}} \min_{\boldsymbol{p}\in S_K} \sum_{r=1}^{R_{\text{upper}}} \max_{i\in G'_r\cap G_\varepsilon} \max_{\substack{j\in\mathcal{A}_I(r-1)\\ j\ne i}} \max_{m\in B'_r\cap G_\varepsilon^c} \max\left\{ \frac{2\|\boldsymbol{y}_{i,j}\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}}}{\left(4\varepsilon_r\right)^2}, \frac{2\|\boldsymbol{y}_{1,m}\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}}}{\left(4\varepsilon_r\right)^2} \right\} \\
&\overset{\text{(ii)}}{\ge} \frac{1}{R_{\text{upper}}} \sum_{r=1}^{R_{\text{upper}}} 2^{2r-3} \min_{\boldsymbol{p}\in S_K} \max_{i\in G'_r\cap G_\varepsilon} \max_{\substack{j\in\mathcal{A}_I(r-1)\\ j\ne i}} \max_{m\in B'_r\cap G_\varepsilon^c} \max\left\{ \|\boldsymbol{y}_{i,j}\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}}, \|\boldsymbol{y}_{1,m}\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}} \right\}
\end{aligned}$$

$$\overset{\text{(iii)}}{\geq} \frac{1}{R_{\text{upper}}} \sum_{r=1}^{R_{\text{upper}}} 2^{2r-3} \min_{\boldsymbol{p}\in S_K} \max_{i\in\mathcal{A}_I(r-1)\cap G_\varepsilon\cap G_r^c} \max_{\substack{j\in\mathcal{A}_I(r-1)\\ j\neq i}} \max_{m\in\mathcal{A}_I(r-1)\cap G_\varepsilon^c} \max\left\{\|\boldsymbol{y}_{i,j}\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}}, \|\boldsymbol{y}_{1,m}\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}}\right\}$$

$$\overset{\text{(iv)}}{\geq} \frac{{\mathcal{G}_{\mathcal{Y}}}^2 L_2}{dL_1} \frac{1}{R_{\text{upper}}} \sum_{r=1}^{R_{\text{upper}}} 2^{2r-3} \min_{\boldsymbol{p}\in S_K} \max_{i\in\mathcal{A}_I(r-1)\cap G_\varepsilon\setminus\{1\}} \max_{m\in\mathcal{A}_I(r-1)\cap G_\varepsilon^c} \max\left\{\|\boldsymbol{y}_{1,i}\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}}, \|\boldsymbol{y}_{1,m}\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}}\right\}$$

$$\overset{\text{(v)}}{\geq} \frac{{\mathcal{G}_{\mathcal{Y}}}^2 L_2}{4R_{\text{upper}}dL_1} \sum_{r=1}^{R_{\text{upper}}} 2^{2r-3} \min_{\boldsymbol{p}\in S_K} \max_{\substack{i,j\in\mathcal{A}_I(r-1)\\ i\neq j}} \|\boldsymbol{y}_{i,j}\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}}$$

$$= \frac{{\mathcal{G}_{\mathcal{Y}}}^2 L_2}{32R_{\text{upper}}dL_1} \sum_{r=1}^{R_{\text{upper}}} 2^{2r} g_{\mathcal{X}\mathcal{Y}}(\mathcal{Y}(\mathcal{A}(r-1))), \tag{EC.6.10}$$

where (i) follows from the fact that the maximum of positive numbers is always greater than or equal to their average, and (ii) uses the fact that the minimum of a sum is greater than or equal to the sum of the minimums. (iii) arises from the set inclusion relationships established in Lemma EC.6.1 and Lemma EC.6.2. (iv) is a direct consequence of Lemma EC.13.2 with $j=1$. Finally, for (v), note that for any $i,j\in\mathcal{A}_I(r-1)$, we have $\max_{i,j\in\mathcal{A}_I(r-1),i\neq j}\|\boldsymbol{y}_{i,j}\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}} \leq 4\max_{i\in\mathcal{A}_I(r-1)\setminus\{1\}}\|\boldsymbol{y}_{1,i}\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}}$. □

Moreover, to provide additional insights into the G-optimal design, from equation (EC.5.43), we obtain

$$\mathbb{E}_{\boldsymbol{\mu}}[T\mid\mathcal{E}_1] \leq 4\log\left[\frac{2K}{\delta}(R_{\text{upper}}+1)\right] \sum_{r=1}^{R_{\text{upper}}} d2^{2r} + \frac{d(d+1)}{2}R_{\text{upper}}. \tag{EC.6.11}$$

From inequalities (EC.6.10) and (EC.6.11), it follows that to align the upper and lower bounds, one must establish $d\leq c\min_{\boldsymbol{p}\in S_K}\max_{i,j\in\mathcal{A}_I(r-1),i\neq j}\|\boldsymbol{y}_{i,j}\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}}$ for some universal constant. However, applying the Kiefer–Wolfowitz Theorem from Lemma EC.13.3 yields only the reverse inequality

$$\min_{\boldsymbol{p}\in S_K}\max_{i,j\in\mathcal{A}_I(r-1),i\neq j}\|\boldsymbol{y}_{i,j}\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}} \leq 4\min_{\boldsymbol{p}\in S_K}\max_{i\in\mathcal{A}_I(r-1)}\|\boldsymbol{a}_i\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}} = 4d. \tag{EC.6.12}$$

This result indicates that LinFACT-G cannot achieve the lower bound established in Theorem 1.

## EC.7. Proof of Theorem 3

The central idea of this proof is to establish a direct relationship between the lower bound and the key minimax summation terms in the upper bound, thereby enabling the sample complexity to be bounded explicitly in terms of the lower bound term $(\Gamma^*)^{-1}$. We first show that the good event $\mathcal{E}_3$ defined below occurs with probability at least $1-\delta_r$ in each round $r$, where $\delta_r$ denotes the probability that the good event $\mathcal{E}_3$ does not hold in a certain round $r$. By taking the union bound across different rounds, we can complete the proof of Lemma EC.7.1 regarding the overall probability of event $\mathcal{E}_3$ occurring, denoted by $\delta$. We then demonstrate that the probability of this good event holding in every round is at least $1-\delta$. Consequently, we can sum the sample complexity

bounds from each round (conditioned on the good event) to obtain the overall bound on the sample complexity.

First, we define the good event $\mathcal{E}_3$:

$$\mathcal{E}_3 = \bigcap_{i\in\mathcal{A}_I(r-1)} \bigcap_{\substack{j\in\mathcal{A}_I(r-1)\\ j\neq i}} \bigcap_{r\in\mathbb{N}} |(\hat{\mu}_j(r)-\hat{\mu}_i(r))-(\mu_j-\mu_i)| \le 2\varepsilon_r. \tag{EC.7.1}$$

Since arms are sampled according to a preset allocation (*i.e.*, a fixed design), we introduce the following lemma to provide a confidence region for the estimated parameter $\boldsymbol{\theta}$.

LEMMA EC.7.1. *Let $\delta\in(0,1)$. Then, it holds that $P(\mathcal{E}_3)\ge 1-\delta$.*

*Proof.* Since $\hat{\boldsymbol{\theta}}_r$ is a ordinary least squares estimator of $\boldsymbol{\theta}$ and the noise is i.i.d., it follows that $\boldsymbol{y}^\top\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_r\right)$ is $\|\boldsymbol{y}\|^2_{\boldsymbol{V}_r^{-1}}$-sub-Gaussian for all $\boldsymbol{y}\in\mathcal{Y}(\mathcal{A}(r-1))$. Moreover, the guarantees of the rounding procedure ensure that

$$\|\boldsymbol{y}\|^2_{\boldsymbol{V}_r^{-1}} \le (1+\epsilon)\, g_{\mathcal{X}\mathcal{Y}}(\mathcal{Y}(\mathcal{A}(r-1)))/T_r \le \frac{2^{-2r-1}}{\log\frac{2K(K-1)r(r+1)}{\delta}} \tag{EC.7.2}$$

for all $\boldsymbol{y}\in\mathcal{Y}(\mathcal{A}(r-1))$, as ensured by our choice of $T_r$ in equation (22). Since the right-hand side is deterministic and does not depend on the randomness of the arm rewards, we have that for any $\rho>0$ and $\boldsymbol{y}\in\mathcal{Y}(\mathcal{A}(r-1))$,

$$\mathbb{P}\left\{|\boldsymbol{y}^\top\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_r\right)| > \sqrt{2^{-2r}\frac{\log(2/\rho)}{\log\frac{2K(K-1)r(r+1)}{\delta}}}\right\} \le \rho. \tag{EC.7.3}$$

Letting $\rho=\frac{\delta}{K(K-1)r(r+1)}$ and applying a union bound over all possible $\boldsymbol{y}\in\mathcal{Y}(\mathcal{A}(r-1))$, where $|\mathcal{Y}(\mathcal{A}(r-1))|\le|\mathcal{Y}(\mathcal{A}(0))|\le K(K-1)$, we obtain the desired probability guarantee:

$$\begin{aligned}
\mathbb{P}(\mathcal{E}_3^c) &= \mathbb{P}\left\{\bigcup_{i\in\mathcal{A}_I(r-1)}\bigcup_{\substack{j\in\mathcal{A}_I(r-1)\\ j\neq i}}\bigcup_{r\in\mathbb{N}} |(\hat{\mu}_j(r)-\hat{\mu}_i(r))-(\mu_j-\mu_i)| > \varepsilon_r\right\}\\
&\le \sum_{r=1}^{\infty}\mathbb{P}\left\{\bigcup_{i\in\mathcal{A}_I(r-1)}\bigcup_{\substack{j\in\mathcal{A}_I(r-1)\\ j\neq i}} |(\hat{\mu}_j(r)-\hat{\mu}_i(r))-(\mu_j-\mu_i)| > \varepsilon_r\right\}\\
&\le \sum_{r=1}^{\infty}\sum_{i=1}^{K}\sum_{\substack{j=1\\ j\neq i}}^{K}\frac{\delta}{K(K-1)r(r+1)}\\
&= \delta.
\end{aligned} \tag{EC.7.4}$$

Taking the union bound over all rounds $r\in\mathbb{N}$ completes the proof. □

Thus, by the standard result of the $\mathcal{XY}$-optimal design, we have

$$C_{\delta/K}(r) \coloneqq \varepsilon_r, \tag{EC.7.5}$$

which matches the expression in equation (EC.5.5) for the G-optimal design.

LEMMA EC.7.2. *On the event $\mathcal{E}_3$, the best arm $1 \in \mathcal{A}_I(r)$ for all $r \in \mathbb{N}$.*

*Proof.* If the event $\mathcal{E}_3$ holds, then for any arm $i \in \mathcal{A}_I(r-1)$, we have

$$\hat{\mu}_i(r) - \hat{\mu}_1(r) \leq \mu_i - \mu_1 + 2\varepsilon_r \leq 2\varepsilon_r < 2\varepsilon_r + \varepsilon, \tag{EC.7.6}$$

which implies that $\hat{\mu}_1(r) + \varepsilon_r > \max_{i \in \mathcal{A}_I(r-1)} \hat{\mu}_i - \varepsilon_r - \varepsilon = L_r$ and $\hat{\mu}_1(r) + \varepsilon_r \geq \max_{i \in \mathcal{A}_I(r-1)} \hat{\mu}_i(r) - C_{\delta/K}(r)$.

These inequalities ensure that arm 1 will not be eliminated from $\mathcal{A}_I(r)$ in LinFACT. □

LEMMA EC.7.3. *With probability at least $1-\delta$, and employing an $\varepsilon$-efficient rounding procedure, LinFACT-$\mathcal{XY}$ successfully identifies all $\varepsilon$-best arms and achieves instance-optimal sample complexity up to logarithmic factors, as given by*

$$T \leq c \left[ d R_{\text{upper}} \log\left( \frac{2K(R_{\text{upper}}+1)}{\delta} \right) \right] (\Gamma^*)^{-1} + q(\epsilon) R_{\text{upper}}, \tag{EC.7.7}$$

*where $c$ is a universal constant, $R_{\text{upper}} = \max\left\{ \left\lceil \log_2 \frac{4}{\alpha_\varepsilon} \right\rceil, \left\lceil \log_2 \frac{4}{\beta_\varepsilon} \right\rceil \right\}$.*

*Proof.* Combining the result of Lemma EC.7.1, we conclude that with probability at least $1-\delta$

$$\begin{aligned}
T &\leq \sum_{r=1}^{R_{\text{upper}}} \max\left\{ \left\lceil \frac{2 g_{\mathcal{XY}}(\mathcal{Y}(\mathcal{A}(r-1)))(1+\epsilon)}{\varepsilon_r^2} \log\left( \frac{2K(K-1)r(r+1)}{\delta} \right) \right\rceil, q(\epsilon) \right\} \\
&\leq \sum_{r=1}^{R_{\text{upper}}} 2 \cdot 2^{2r} g_{\mathcal{XY}}(\mathcal{Y}(\mathcal{A}(r-1)))(1+\epsilon) \log\left( \frac{2K(K-1)r(r+1)}{\delta} \right) + (1+q(\epsilon)) R_{\text{upper}} \\
&\leq \left[ 64(1+\epsilon) \log\left( \frac{2K(K-1)R_{\text{upper}}(R_{\text{upper}}+1)}{\delta} \right) \frac{R_{\text{upper}} d L_1}{{\mathcal{G}_{\mathcal{Y}}}^2 L_2} \right] (\Gamma^*)^{-1} + (1+q(\epsilon)) R_{\text{upper}} \\
&\leq \left[ 128(1+\epsilon) \log\left( \frac{2K(R_{\text{upper}}+1)}{\delta} \right) \frac{R_{\text{upper}} d L_1}{{\mathcal{G}_{\mathcal{Y}}}^2 L_2} \right] (\Gamma^*)^{-1} + (1+q(\epsilon)) R_{\text{upper}} \\
&\leq c \left[ d R_{\text{upper}} \log\left( \frac{2K(R_{\text{upper}}+1)}{\delta} \right) \right] (\Gamma^*)^{-1} + q(\epsilon) R_{\text{upper}},
\end{aligned} \tag{EC.7.8}$$

where $c$ is a universal constant, $R_{\text{upper}} = \max\left\{ \left\lceil \log_2 \frac{4}{\alpha_\varepsilon} \right\rceil, \left\lceil \log_2 \frac{4}{\beta_\varepsilon} \right\rceil \right\}$, and the third inequality follows from equation (EC.6.10).

Then, let $\xi = \min(\alpha_\varepsilon, \beta_\varepsilon)/16$ be the minimum gap of the problem instance. Considering that the approximation error term $q(\epsilon)$ is in the form of $\mathcal{O}(\frac{d}{\epsilon^2})$ (Allen-Zhu et al. 2021, Fiez et al. 2019), we have

$$\mathbb{E}[T_{\mathcal{XY}} \mid \mathcal{E}] = \mathcal{O}\left( \frac{d}{\Gamma^*} \xi^{-1} \log(\xi^{-1}) \log\left( \frac{K}{\delta} \log(\xi^{-2}) \right) + \frac{d}{\epsilon^2} \log(\xi^{-1}) \right). \tag{EC.7.9}$$

□

## EC.8. Proof of Theorem 4

### EC.8.1. Step 1: Define the Clean Events

The core of the proof lies in similarly defining the round at which all classifications are completed under the misspecified model. To this end, we reconstruct the anytime confidence radius for each arm in round $r$ and redefine the high-probability event over the entire execution of the algorithm. We denote this clean event as $\mathcal{E}_{1m}$.

$$\mathcal{E}_{1m} = \left\{ \bigcap_{i \in \mathcal{A}_I(r-1)} \bigcap_{r \in \mathbb{N}} |\hat{\mu}_i(r) - \mu_i| \le \varepsilon_r + L_m \sqrt{d} \right\}. \tag{EC.8.1}$$

Similarly, since arms are sampled according to a preset allocation (*i.e.*, the optimal design criterion), we invoke Lemma EC.5.1 to derive an adjusted confidence region for the estimated parameter $\hat{\boldsymbol{\theta}}_t$. When following the G-optimal sampling rule, as specified in lines 2 and 4 of the pseudocode, we obtain the following result for each round $r$:

$$\boldsymbol{V}_r = \sum_{\boldsymbol{a} \in \mathrm{Supp}(\pi_r)} T_r(\boldsymbol{a}) \boldsymbol{a}\boldsymbol{a}^\top \succeq \frac{2d}{\varepsilon_r^2} \log\left(\frac{2Kr(r+1)}{\delta}\right) \boldsymbol{V}(\pi). \tag{EC.8.2}$$

To give the confidence radius, we have the following decomposition.

$$\begin{aligned} \left| \left\langle \hat{\boldsymbol{\theta}}_r - \boldsymbol{\theta}, \boldsymbol{a} \right\rangle \right| &= \left| \boldsymbol{a}^\top \boldsymbol{V}_r^{-1} \sum_{s=1}^{T_r} m(\boldsymbol{a}_{A_s}) \boldsymbol{a}_{A_s} + \boldsymbol{a}^\top \boldsymbol{V}_r^{-1} \sum_{s=1}^{T_r} \eta_s \boldsymbol{a}_{A_s} \right| \\ &\le \left| \boldsymbol{a}^\top \boldsymbol{V}_r^{-1} \sum_{s=1}^{T_r} \Delta_m(\boldsymbol{a}_{A_s}) \boldsymbol{a}_{A_s} \right| + \left| \boldsymbol{a}^\top \boldsymbol{V}_r^{-1} \sum_{s=1}^{T_r} \eta_s \boldsymbol{a}_{A_s} \right|, \end{aligned} \tag{EC.8.3}$$

where the first term is bounded by

$$\begin{aligned} \left| \boldsymbol{a}^\top \boldsymbol{V}_r^{-1} \sum_{s=1}^{T_r} \Delta_m(\boldsymbol{a}_{A_s}) \boldsymbol{a}_{A_s} \right| &= \left| \boldsymbol{a}^\top \boldsymbol{V}_r^{-1} \sum_{\boldsymbol{b} \in \mathcal{A}(r-1)} T_r(\boldsymbol{b}) \Delta_m(\boldsymbol{b}) \boldsymbol{b} \right| \\ &\le L_m \sum_{\boldsymbol{b} \in \mathcal{A}(r-1)} T_r(\boldsymbol{b}) \left| \boldsymbol{a}^\top \boldsymbol{V}_r^{-1} \boldsymbol{b} \right| \\ &\le L_m \sqrt{\left( \sum_{\boldsymbol{b} \in \mathcal{A}(r-1)} T_r(\boldsymbol{b}) \right) \boldsymbol{a}^\top \left( \sum_{\boldsymbol{b} \in \mathcal{A}(r-1)} T_r(\boldsymbol{b}) \boldsymbol{V}_r^{-1} \boldsymbol{b}\boldsymbol{b}^\top \boldsymbol{V}_r^{-1} \boldsymbol{a} \right)} \\ &= L_m \sqrt{\sum_{\boldsymbol{b} \in \mathcal{A}(r-1)} T_r(\boldsymbol{b}) \|\boldsymbol{a}\|^2_{\boldsymbol{V}_r^{-1}}} \\ &\le L_m \sqrt{d}, \end{aligned} \tag{EC.8.4}$$

where the first inequality follows from Hölder's inequality, the second from Jensen's inequality, and the last from the guarantee of the G-optimal exploration policy, which ensures that $\|\boldsymbol{a}\|^2_{\boldsymbol{V}_r^{-1}} \le d/T_r$.

The second term is also bounded using Lemma EC.5.1 and the result in Lemma EC.13.1. For any arm $\boldsymbol{a} \in \mathcal{A}(r-1)$, with a probability of at least $1-\delta/Kr(r+1)$, we have

$$\begin{aligned}\left|\boldsymbol{a}^\top \boldsymbol{V}_r^{-1} \sum_{s=1}^{T_r} \eta_s \boldsymbol{a}_{A_s}\right| &\le \sqrt{2\left\|\boldsymbol{a}\right\|^2_{\boldsymbol{V}_r^{-1}} \log\left(\frac{2Kr(r+1)}{\delta}\right)} \\ &= \sqrt{2\boldsymbol{a}^\top \boldsymbol{V}_r^{-1}\boldsymbol{a} \log\left(\frac{2Kr(r+1)}{\delta}\right)} \\ &\le \sqrt{2\boldsymbol{a}^\top \left(\frac{\varepsilon_r^2}{2d}\frac{1}{\log\left(\frac{2Kr(r+1)}{\delta}\right)}\boldsymbol{V}(\pi)^{-1}\right)\boldsymbol{a} \log\left(\frac{2Kr(r+1)}{\delta}\right)} \\ &\le \varepsilon_r. \end{aligned} \tag{EC.8.5}$$

Thus, with the standard result of the G-optimal design, we also have

$$C_{\delta/K}(r) \coloneqq \varepsilon_r. \tag{EC.8.6}$$

To establish the probability guarantee for event $\mathcal{E}_{1m}$, we combine the results from equations (EC.8.4) and (EC.8.5), yielding

$$\begin{aligned}\mathbb{P}(\mathcal{E}_{1m}^c) &= \mathbb{P}\left\{\bigcup_{i\in\mathcal{A}_I(r-1)}\bigcup_{r\in\mathbb{N}} |\hat{\mu}_i(r)-\mu_i| > \varepsilon_r + L_m\sqrt{d}\right\} \\ &\le \sum_{r=1}^{\infty}\mathbb{P}\left\{\bigcup_{i\in\mathcal{A}_I(r-1)} |\hat{\mu}_i(r)-\mu_i| > \varepsilon_r + L_m\sqrt{d}\right\} \\ &\le \sum_{r=1}^{\infty}\sum_{i=1}^{K}\frac{\delta}{Kr(r+1)} \\ &= \delta. \end{aligned} \tag{EC.8.7}$$

Therefore, taking the union bounds over rounds $r \in \mathbb{N}$, we have

$$P(\mathcal{E}_{1m}) \ge 1-\delta. \tag{EC.8.8}$$

Considering an additional event that characterizes the gaps between different arms, defined as follows

$$\mathcal{E}_{2m} = \bigcap_{i\in G_\varepsilon}\bigcap_{j\in\mathcal{A}_I(r-1)}\bigcap_{r\in\mathbb{N}} |(\hat{\mu}_j(r)-\hat{\mu}_i(r)) - (\mu_j-\mu_i)| \le 2\varepsilon_r + 2L_m\sqrt{d}. \tag{EC.8.9}$$

By equation (EC.8.3), for $i, j \in \mathcal{A}_I(r-1)$, we have

$$\begin{aligned}\mathbb{P}\left\{|(\hat{\mu}_j-\hat{\mu}_i)-(\mu_j-\mu_i)| > 2\varepsilon_r + 2L_m\sqrt{d} \,|\, \mathcal{E}_{1m}\right\} &\le \mathbb{P}\left\{|\hat{\mu}_j-\mu_j| + |\hat{\mu}_i-\mu_i| > 2\varepsilon_r + 2L_m\sqrt{d} \,|\, \mathcal{E}_{1m}\right\} \\ &\le \mathbb{P}\left\{|\hat{\mu}_j-\mu_j| > \varepsilon_r + L_m\sqrt{d} \,|\, \mathcal{E}_{1m}\right\} \\ &+ \mathbb{P}\left\{|\hat{\mu}_i-\mu_i| > \varepsilon_r + L_m\sqrt{d} \,|\, \mathcal{E}_{1m}\right\} \\ &= 0, \end{aligned} \tag{EC.8.10}$$

which implies

$$\mathbb{P}(\mathcal{E}_{2m} \mid \mathcal{E}_{1m}) = 1. \tag{EC.8.11}$$

## EC.8.2. Step 2: Bound the Expected Sample Complexity

To bound the expected sample complexity under model misspecification, we aim to identify the round in which all arms $i \in G_\varepsilon$ have been included in $G_r$, and the round in which all arms $i \in G_\varepsilon^c$ have been added to $B_r$.

LEMMA EC.8.1. *For $i \in G_\varepsilon$ and $L_m < \frac{\alpha_\varepsilon}{2\sqrt{d}}$, if $r \geq \left\lceil \log_2 \left( \frac{4}{\varepsilon - \Delta_i - 2L_m\sqrt{d}} \right) \right\rceil$, then we have $\mathbb{E}_{\boldsymbol{\mu}} \left[ \mathbb{1}\left[ i \notin G_r \right] \mid \mathcal{E}_{1m} \right] = 0$.*

*Proof.* First, for any $i \in G_\varepsilon$

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\mu}} \left[ \mathbb{1}\left[ i \notin G_r \right] \mid \mathcal{E}_{1m} \right] &= \mathbb{E}_{\boldsymbol{\mu}} \left[ \mathbb{1}\left[ \max_{j \in \mathcal{A}_I(r-1)} \hat{\mu}_j - \hat{\mu}_i \geq -2\varepsilon_r + \varepsilon \right] \mid \mathcal{E}_{1m}, i \notin G_m (m = \{1, 2, \ldots, r-1\}) \right] \\
&\leq \mathbb{E}_{\boldsymbol{\mu}} \left[ \mathbb{1}\left[ \max_{j \in \mathcal{A}_I(r-1)} \hat{\mu}_j - \hat{\mu}_i \geq -2\varepsilon_r + \varepsilon \right] \mid \mathcal{E}_{1m} \right].
\end{aligned} \tag{EC.8.12}$$

If $i \in G_{r-1}$, then $i \in G_r$ by definition. Otherwise, if $i \notin G_{r-1}$, then under event $\mathcal{E}_{1m}$, for $i \in G_\varepsilon$ and $r \geq \left\lceil \log_2 \left( \frac{4}{\varepsilon - \Delta_i - 2L_m\sqrt{d}} \right) \right\rceil$, we have

$$\max_{j \in \mathcal{A}_I(r-1)} \hat{\mu}_j - \hat{\mu}_i \leq \mu_{\arg\max_{j \in \mathcal{A}_I(r-1)} \hat{\mu}_j} - \mu_i + 2^{-r+1} + 2L_m\sqrt{d} \leq \Delta_i + 2^{-r+1} + 2L_m\sqrt{d} \leq \varepsilon - 2\varepsilon_r, \tag{EC.8.13}$$

which implies that $i \in G_r$ by line 8 of the algorithm. In particular, under event $\mathcal{E}_{1m}$, if $i \notin G_{r-1}$, for all $i \in G_\varepsilon$ and $r \geq \left\lceil \log_2 \left( \frac{4}{\varepsilon - \Delta_i - 2L_m\sqrt{d}} \right) \right\rceil$, we have

$$\mathbb{E}_{\boldsymbol{\mu}} \left[ \mathbb{1}\left[ \max_{j \in \mathcal{A}_I(r-1)} \hat{\mu}_j - \hat{\mu}_i \leq -2\varepsilon_r + \varepsilon \right] \mid i \notin G_{r-1}, \mathcal{E}_{1m} \right] = 1. \tag{EC.8.14}$$

Consequently, $\mathbb{1}\left[ i \notin G_r \right] \mathbb{1}\left[ i \in G_{r-1} \right] = 0$. Therefore,

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\mu}} \left[ \mathbb{1}\left[ \max_{j \in \mathcal{A}_I(r-1)} \hat{\mu}_j - \hat{\mu}_i \geq -2\varepsilon_r + \varepsilon \right] \mid \mathcal{E}_{1m} \right] \\
&= \mathbb{E}_{\boldsymbol{\mu}} \left[ \mathbb{1}\left[ \max_{j \in \mathcal{A}_I(r-1)} \hat{\mu}_j - \hat{\mu}_i \geq -2\varepsilon_r + \varepsilon \right] \mathbb{1}\left[ i \notin G_{r-1} \right] \mid \mathcal{E}_{1m} \right] \\
&+ \mathbb{E}_{\boldsymbol{\mu}} \left[ \mathbb{1}\left[ \max_{j \in \mathcal{A}_I(r-1)} \hat{\mu}_j - \hat{\mu}_i \geq -2\varepsilon_r + \varepsilon \right] \mathbb{1}\left[ i \in G_{r-1} \right] \mid \mathcal{E}_{1m} \right] \\
&= \mathbb{E}_{\boldsymbol{\mu}} \left[ \mathbb{1}\left[ \max_{j \in \mathcal{A}_I(r-1)} \hat{\mu}_j - \hat{\mu}_i \geq -2\varepsilon_r + \varepsilon \right] \mathbb{1}\left[ i \notin G_{r-1} \right] \mid \mathcal{E}_{1m} \right] \\
&= \mathbb{E}_{\boldsymbol{\mu}} \left[ \mathbb{1}\left[ \max_{j \in \mathcal{A}_I(r-1)} \hat{\mu}_j - \hat{\mu}_i \geq -2\varepsilon_r + \varepsilon \right] \mathbb{1}\left[ i \notin G_{r-1} \right] \mid i \notin G_{r-1}, \mathcal{E}_{1m} \right] \mathbb{P}\left( i \notin G_{r-1} \mid \mathcal{E}_{1m} \right) \\
&+ \mathbb{E}_{\boldsymbol{\mu}} \left[ \mathbb{1}\left[ \max_{j \in \mathcal{A}_I(r-1)} \hat{\mu}_j - \hat{\mu}_i \geq -2\varepsilon_r + \varepsilon \right] \mathbb{1}\left[ i \notin G_{r-1} \right] \mid i \in G_{r-1}, \mathcal{E}_{1m} \right] \mathbb{P}\left( i \in G_{r-1} \mid \mathcal{E}_{1m} \right) \\
&= \mathbb{E}_{\boldsymbol{\mu}} \left[ \mathbb{1}\left[ \max_{j \in \mathcal{A}_I(r-1)} \hat{\mu}_j - \hat{\mu}_i \geq -2\varepsilon_r + \varepsilon \right] \mathbb{1}\left[ i \notin G_{r-1} \right] \mid i \notin G_{r-1}, \mathcal{E}_{1m} \right] \mathbb{P}\left( i \notin G_{r-1} \mid \mathcal{E}_{1m} \right)
\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[\max_{j\in\mathcal{A}_I(r-1)}\hat{\mu}_j-\hat{\mu}_i\geq -2\varepsilon_r+\varepsilon\right] \mid i\notin G_{r-1},\mathcal{E}_{1m}\right]\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[i\notin G_{r-1}\right]\mid\mathcal{E}_{1m}\right]\\
&= 0, \qquad \text{(EC.8.15)}
\end{aligned}
$$

where the second line follows from the additivity of expectation. The fourth line follows the deterministic result that $\mathbb{1}\left[i\notin G_r\right]\mathbb{1}\left[i\in G_{r-1}\right]=0$. The fifth line is the decomposition based on the conditional expectation. The eighth line comes from the fact that the expectation of the indicator function is simply the probability. The last line follows the result in equation (EC.8.14). The lemma can thus be concluded together with equation (EC.8.12).

In the perfectly linear model, $\varepsilon-\Delta_i>0$ always holds for any $i\in G_\varepsilon$. However, under model misspecification, the sign of this term within the logarithm must be verified. To ensure positivity for all $i\in G_\varepsilon$, it is necessary that $\alpha_\varepsilon>2L_m\sqrt{d}$. As the misspecification magnitude $L_m$ approaches $\alpha_\varepsilon/(2\sqrt{d})$, the upper bound on the expected sample complexity increases sharply, since the misspecification significantly impairs the identification of $\varepsilon$-best arms. Moreover, when $L_m\geq\alpha_\varepsilon/(2\sqrt{d})$, the sample complexity can no longer be bounded in this form, which is intuitive and consistent with the general insights discussed earlier in Section 5.3. □

Lemma EC.8.2. *For $i\in G_\varepsilon^c$ and $L_m<\frac{\beta_\varepsilon}{2\sqrt{d}}$, if $r\geq\left\lceil\log_2\left(\frac{4}{\Delta_i-\varepsilon-2L_m\sqrt{d}}\right)\right\rceil$, then we have $\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[i\notin B_r\right]\mid\mathcal{E}_{1m}\right]=0$.*

*Proof.* First, we for any $i\in G_\varepsilon^c$,

$$
\begin{aligned}
\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[i\notin B_r\right]\mid\mathcal{E}_{1m}\right] &= \mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[\max_{j\in\mathcal{A}_I(r-1)}\hat{\mu}_j-\hat{\mu}_i\leq 2\varepsilon_r+\varepsilon\right]\mid\mathcal{E}_{1m},i\notin B_m(m=\{1,2,\ldots,r-1\})\right]\\
&\leq \mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[\max_{j\in\mathcal{A}_I(r-1)}\hat{\mu}_j-\hat{\mu}_i\leq 2\varepsilon_r+\varepsilon\right]\mid\mathcal{E}_{1m}\right]. \qquad \text{(EC.8.16)}
\end{aligned}
$$

If $i\in B_{r-1}$, then $i\in B_r$ by definition. Otherwise, if $i\notin B_{r-1}$, then under event $\mathcal{E}_{1m}$, for $i\in G_\varepsilon^c$ and $r\geq\left\lceil\log_2\left(\frac{4}{\Delta_i-\varepsilon-2L_m\sqrt{d}}\right)\right\rceil$, we have

$$
\max_{j\in\mathcal{A}_I(r-1)}\hat{\mu}_j-\hat{\mu}_i\geq\Delta_i-2^{-r+1}-2L_m\sqrt{d}\geq\varepsilon+2\varepsilon_r, \qquad \text{(EC.8.17)}
$$

which implies that $i\in B_r$ by line 6 of the algorithm. In particular, under event $\mathcal{E}_{1m}$, if $i\notin B_{r-1}$, for all $i\in G_\varepsilon^c$ and $r\geq\left\lceil\log_2\left(\frac{4}{\Delta_i-\varepsilon-2L_m\sqrt{d}}\right)\right\rceil$, we have

$$
\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[\max_{j\in\mathcal{A}_I(r-1)}\hat{\mu}_j-\hat{\mu}_i>2\varepsilon_r+\varepsilon\right]\mid i\notin B_{r-1},\mathcal{E}_{1m}\right]=1. \qquad \text{(EC.8.18)}
$$

Deterministically, $\mathbb{1}\left[i\notin B_r\right]\mathbb{1}\left[i\in B_{r-1}\right]=0$. Therefore,

$$
\begin{aligned}
&\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[\max_{j\in\mathcal{A}_I(r-1)}\hat{\mu}_j-\hat{\mu}_i\leq 2\varepsilon_r+\varepsilon\right]\mid\mathcal{E}_{1m}\right]\\
&=\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[\max_{j\in\mathcal{A}_I(r-1)}\hat{\mu}_j-\hat{\mu}_i\leq 2\varepsilon_r+\varepsilon\right]\mathbb{1}\left[i\notin B_{r-1}\right]\mid\mathcal{E}_{1m}\right]
\end{aligned}
$$

$$
\begin{aligned}
&+\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[\max_{j\in\mathcal{A}_I(r-1)}\hat{\mu}_j-\hat{\mu}_i\le 2\varepsilon_r+\varepsilon\right]\mathbb{1}\left[i\in B_{r-1}\right]\mid\mathcal{E}_{1m}\right]\\
&=\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[\max_{j\in\mathcal{A}_I(r-1)}\hat{\mu}_j-\hat{\mu}_i\le 2\varepsilon_r+\varepsilon\right]\mathbb{1}\left[i\notin B_{r-1}\right]\mid\mathcal{E}_{1m}\right]\\
&=\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[\max_{j\in\mathcal{A}_I(r-1)}\hat{\mu}_j-\hat{\mu}_i\le 2\varepsilon_r+\varepsilon\right]\mathbb{1}\left[i\notin B_{r-1}\right]\mid i\notin B_{r-1},\mathcal{E}_{1m}\right]\mathbb{P}\left(i\notin B_{r-1}\mid\mathcal{E}_{1m}\right)\\
&+\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[\max_{j\in\mathcal{A}_I(r-1)}\hat{\mu}_j-\hat{\mu}_i\le 2\varepsilon_r+\varepsilon\right]\mathbb{1}\left[i\notin B_{r-1}\right]\mid i\in B_{r-1},\mathcal{E}_{1m}\right]\mathbb{P}\left(i\in B_{r-1}\mid\mathcal{E}_{1m}\right)\\
&=\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[\max_{j\in\mathcal{A}_I(r-1)}\hat{\mu}_j-\hat{\mu}_i\le 2\varepsilon_r+\varepsilon\right]\mathbb{1}\left[i\notin B_{r-1}\right]\mid i\notin B_{r-1},\mathcal{E}_{1m}\right]\mathbb{P}\left(i\notin B_{r-1}\mid\mathcal{E}_{1m}\right)\\
&=\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[\max_{j\in\mathcal{A}_I(r-1)}\hat{\mu}_j-\hat{\mu}_i\le 2\varepsilon_r+\varepsilon\right]\mid i\notin B_{r-1},\mathcal{E}_{1m}\right]\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[i\notin B_{r-1}\right]\mid\mathcal{E}_{1m}\right]\\
&=0. \qquad \text{(EC.8.19)}
\end{aligned}
$$

The second line follows from the linearity of expectation. The fourth line uses the deterministic fact that $\mathbb{1}\left[i\notin B_r\right]\mathbb{1}\left[i\in B_{r-1}\right]=0$. The fifth line applies the law of total expectation. The eighth line uses the identity that the expectation of an indicator function equals the corresponding probability. The final line follows from equation (EC.8.18). The lemma then follows by combining this result with equation (EC.8.16).

Similarly, we must ensure the positivity of the term inside the logarithm. To guarantee that $\Delta_i-\varepsilon-2L_m\sqrt{d}>0$ for every $i\in G_\varepsilon^c$, it is necessary that $\beta_\varepsilon>2L_m\sqrt{d}$. □

Lemma EC.8.3. *Suppose $L_m<\min\left\{\frac{\alpha_\varepsilon}{2\sqrt{d}},\frac{\beta_\varepsilon}{2\sqrt{d}}\right\}$, then the round by which all arms are classified is $R'_{upper}=\max\left\{\left\lceil\log_2\frac{4}{\alpha_\varepsilon-2L_m\sqrt{d}}\right\rceil,\left\lceil\log_2\frac{4}{\beta_\varepsilon-2L_m\sqrt{d}}\right\rceil\right\}$ under model misspecification.*

*Proof.* Combining the results of Lemmas EC.8.1 and EC.8.2, we define an auxiliary round $R_m$ to facilitate the derivation of the upper bound. Specifically, for $i\in G_\varepsilon^c$ and $r\ge\left\lceil\log_2\left(\frac{4}{\Delta_i-\varepsilon-2L_m\sqrt{d}}\right)\right\rceil$, we have $\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[i\notin B_r\right]\mid\mathcal{E}_{1m}\right]=0$. Similarly, for $i\in G_\varepsilon$ and $r\ge\left\lceil\log_2\left(\frac{4}{\varepsilon-\Delta_i-2L_m\sqrt{d}}\right)\right\rceil$, we have $\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[i\notin G_r\right]\mid\mathcal{E}_{1m}\right]=0$.

Noting that $\alpha_\varepsilon=\min_{i\in G_\varepsilon}(\varepsilon-\Delta_i)$ and $\beta_\varepsilon=\min_{i\in G_\varepsilon^c}(\Delta_i-\varepsilon)$, it follows that for any round $r\ge R'_{\text{upper}}$, all arms have been included in either $G_r$ or $B_r$, marking the termination of the algorithm. □

Lemma EC.8.4. *Under model specification, for the expected sample complexity conditioned on the high-probability event $\mathcal{E}_{1m}$, we have*

$$
\begin{aligned}
\mathbb{E}_{\boldsymbol{\mu}}\left[T_{G_{\text{mis}}}\mid\mathcal{E}_{1m}\right]\le c\max\Bigg\{&\frac{256d}{(\alpha_\varepsilon-2L_m\sqrt{d})^2}\log\left(\frac{2K}{\delta}\log_2\frac{16}{\alpha_\varepsilon-2L_m\sqrt{d}}\right),\\
&\frac{256d}{(\beta_\varepsilon-2L_m\sqrt{d})^2}\log\left(\frac{2K}{\delta}\log_2\frac{16}{\beta_\varepsilon-2L_m\sqrt{d}}\right)\Bigg\}+\frac{d(d+1)}{2}R'_{upper}, \qquad \text{(EC.8.20)}
\end{aligned}
$$

*where $c$ is a universal constant and $R'_{upper}=\max\left\{\left\lceil\log_2\frac{4}{\alpha_\varepsilon-2L_m\sqrt{d}}\right\rceil,\left\lceil\log_2\frac{4}{\beta_\varepsilon-2L_m\sqrt{d}}\right\rceil\right\}$.*

*Proof.* We can also decompose $T$ as in equation (EC.5.26), where all expectations are conditioned on the high-probability event $\mathcal{E}_{1m}$, as given by

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\mu}}\left[T_{G_{\text{mis}}} \mid \mathcal{E}_{1m}\right] &\leq \sum_{r=1}^{\infty} \mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[G_r \cup B_r \neq [K]\right] \mid \mathcal{E}_{1m}\right] \sum_{\boldsymbol{a}\in\mathcal{A}(r-1)} T_r(\boldsymbol{a}) \\
&\leq \sum_{r=1}^{R'_{\text{upper}}} \left(d2^{2r+1}\log\left(\frac{2Kr(r+1)}{\delta}\right) + \frac{d(d+1)}{2}\right) \\
&\leq \frac{d(d+1)}{2} R'_{\text{upper}} + 2d\log\left(\frac{2K}{\delta}\right) \sum_{r=1}^{R'_{\text{upper}}} 2^{2r} + 4d \sum_{r=1}^{R'_{\text{upper}}} 2^{2r}\log(r+1) \\
&\leq 4\log\left[\frac{2K}{\delta}\left(R'_{\text{upper}}+1\right)\right] \sum_{r=1}^{R'_{\text{upper}}} d2^{2r} + \frac{d(d+1)}{2} R'_{\text{upper}} \\
&\leq c\max\left\{\frac{256d}{(\alpha_\varepsilon - 2L_m\sqrt{d})^2}\log\left(\frac{2K}{\delta}\log_2\frac{16}{\alpha_\varepsilon - 2L_m\sqrt{d}}\right),\right. \\
&\quad \left.\frac{256d}{(\beta_\varepsilon - 2L_m\sqrt{d})^2}\log\left(\frac{2K}{\delta}\log_2\frac{16}{\beta_\varepsilon - 2L_m\sqrt{d}}\right)\right\} + \frac{d(d+1)}{2} R'_{\text{upper}}. \qquad \text{(EC.8.21)}
\end{aligned}$$

Then, let $\xi = \min\left(\alpha_\varepsilon - 2L_m\sqrt{d}, \beta_\varepsilon - 2L_m\sqrt{d}\right)/16$, we have

$$\mathbb{E}_{\boldsymbol{\mu}}\left[T_{G_{\text{mis}}} \mid \mathcal{E}_{1m}\right] = \mathcal{O}\left(d\xi^{-2}\log\left(\frac{K}{\delta}\log(\xi^{-2})\right) + d^2\log(\xi^{-1})\right). \qquad \text{(EC.8.22)}$$

□

## EC.9. Proof of Theorem 5

### EC.9.1. Step 1: Confidence Radius

In equation (EC.8.3), the first term—arising from the unknown model misspecification—is unavoidable without prior knowledge. However, rather than focusing on the distance between the true parameter $\boldsymbol{\theta}$ and its estimator $\hat{\boldsymbol{\theta}}_t$, we instead compare the orthogonal projection $\boldsymbol{\theta}_t$ with $\hat{\boldsymbol{\theta}}_t$ in the direction of $\boldsymbol{x} \in \mathbb{R}^d$.

Building on the definition of the empirically optimal vector $\hat{\boldsymbol{\mu}}_o(r)$, with $(\hat{\boldsymbol{\theta}}_o(r), \hat{\boldsymbol{\Delta}}_{mo}(r))$ as its associated solution, and the orthogonal parameterization $(\boldsymbol{\theta}_r, \boldsymbol{\Delta}_m(r))$, we derive the confidence radius for each arm via the following decomposition. For any $i \in \mathcal{A}_I(r-1)$, let $\hat{\mu}_i(r) = \hat{\mu}_{o,i}(r)$ denote the value of the optimal estimator on arm $i$ in round $r$, then we have

$$\begin{aligned}
|\hat{\mu}_i(r) - \mu_i| &= \left|\left\langle \hat{\boldsymbol{\theta}}_o(r) - \boldsymbol{\theta}, \boldsymbol{a}_i \right\rangle + \hat{\boldsymbol{\Delta}}_{mo}(r) - \Delta_{mi}\right| \\
&\leq \left|\left\langle \hat{\boldsymbol{\theta}}_o(r) - \boldsymbol{\theta}_r, \boldsymbol{a}_i \right\rangle\right| + \left|\langle \boldsymbol{\theta}_r - \boldsymbol{\theta}, \boldsymbol{a}_i\rangle\right| + \left|\hat{\boldsymbol{\Delta}}_{mo}(r) - \Delta_{mi}\right|, \qquad \text{(EC.9.1)}
\end{aligned}$$

where the third term is bounded by definition as $\left|\hat{\boldsymbol{\Delta}}_{mo}(r) - \Delta_{mi}\right| \leq 2L_m$, while the first two terms can be bounded using the auxiliary lemmas provided below.

LEMMA EC.9.1. *Let $r$ be any round such that $\boldsymbol{V}_r$ is invertible. Consider the orthogonal parameterization $(\boldsymbol{\theta}_r, \boldsymbol{\Delta}_m(r)$ for $\boldsymbol{\mu} = \boldsymbol{\Psi}\boldsymbol{\theta} + \boldsymbol{\Delta}_m$ with $\|\boldsymbol{\Delta}_{\boldsymbol{m}}\|_\infty \le L_m$. Then*

$$\|\boldsymbol{\theta}_r - \boldsymbol{\theta}\|_{\boldsymbol{V}_r} \le L_m \sqrt{T_r}, \tag{EC.9.2}$$

*where $T_r$ is defined in equation* (33).

*Proof.* We use the expression $\boldsymbol{\theta}_r = \boldsymbol{\theta} + \boldsymbol{V}_r^{-1}\boldsymbol{\Psi}^\top \boldsymbol{D}_{\boldsymbol{N}_r}\boldsymbol{\Delta}_m$ derived above. Let $\boldsymbol{P}_{\boldsymbol{N}_r} = \boldsymbol{\Psi}_{\boldsymbol{N}_r}\left(\boldsymbol{\Psi}_{\boldsymbol{N}_r}^\top \boldsymbol{\Psi}_{\boldsymbol{N}_r}\right)^\dagger \boldsymbol{\Psi}_{\boldsymbol{N}_r}^\top$ be a projection, we have

$$\begin{aligned}
\|\boldsymbol{\theta}_r - \boldsymbol{\theta}\|_{\boldsymbol{V}_r} &= \|\boldsymbol{V}_r^{-1}\boldsymbol{\Psi}^\top \boldsymbol{D}_{\boldsymbol{N}_r}\boldsymbol{\Delta}_{\boldsymbol{m}}\|_{\boldsymbol{V}_r} \\
&= \sqrt{\boldsymbol{\Delta}_{\boldsymbol{m}}^\top \boldsymbol{D}_{\boldsymbol{N}_r}\boldsymbol{\Psi}\boldsymbol{V}_r^{-1}\boldsymbol{\Psi}^\top \boldsymbol{D}_{\boldsymbol{N}_r}\boldsymbol{\Delta}_{\boldsymbol{m}}} \\
&= \left\|\boldsymbol{D}_{\boldsymbol{N}_r}^{1/2}\boldsymbol{\Delta}_{\boldsymbol{m}}\right\|_{\boldsymbol{P}_{\boldsymbol{N}_r}} \\
&\le \left\|\boldsymbol{D}_{\boldsymbol{N}_r}^{1/2}\boldsymbol{\Delta}_{\boldsymbol{m}}\right\| \\
&= \|\boldsymbol{\Delta}_{\boldsymbol{m}}\|_{D_{\boldsymbol{N}_r}} \\
&\le L_m\sqrt{T_r}.
\end{aligned} \tag{EC.9.3}$$

□

LEMMA EC.9.2. *(Réda et al. (2021b), Lemma 10). Let $\hat{\boldsymbol{\mu}}_o(r) = \boldsymbol{\Psi}\hat{\boldsymbol{\theta}}_o(r) + \hat{\boldsymbol{\Delta}}_{mo}(r)$ in round $r$, where $(\hat{\boldsymbol{\theta}}_o(r), \hat{\boldsymbol{\Delta}}_{mo}(r))$ are the solution of* (30). *Then the following relationship holds.*

$$\left\|\hat{\boldsymbol{\theta}}_o(r) - \boldsymbol{\theta}_r\right\|_{\boldsymbol{V}_r}^2 \le \left\|\hat{\boldsymbol{\theta}}_r - \boldsymbol{\theta}_r\right\|_{\boldsymbol{V}_r}^2. \tag{EC.9.4}$$

LEMMA EC.9.3. *(Lattimore and Szepesvári (2020), Section 20). Let $\delta \in (0,1)$. Then, with a probability of at least $1-\delta$, it holds that for all $t \in \mathbb{N}$,*

$$\left\|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}\right\|_{V_t} < 2\sqrt{2\left(d\log(6) + \log\left(\frac{1}{\delta}\right)\right)}. \tag{EC.9.5}$$

Equation (EC.9.5) is not directly applicable in our setting due to the deviation term introduced by model misspecification, as shown in equation (EC.8.3). However, by leveraging the orthogonal parameterization, we obtain

$$\|\hat{\boldsymbol{\theta}}_r - \boldsymbol{\theta}_r\|_{\boldsymbol{V}_r}^2 = \|\boldsymbol{V}_r^{-1}\Psi^\top S_r\|_{\boldsymbol{V}_r}^2 = \|\Psi^\top S_r\|_{\boldsymbol{V}_r^{-1}}^2, \tag{EC.9.6}$$

where $S_r = \sum_{s=1}^{T_r} X_{A_s} - \mu_{A_s}$ is the standard self-normalized quantity in the linear bandit literature, allowing existing techniques—such as various concentration inequalities—to be applied directly in the presence of model misspecification. This observation leads to the conclusion that, under misspecification, for any round $r \in \mathbb{N}$, it holds with probability at least $1-\delta$ that

$$\left\|\hat{\boldsymbol{\theta}}_r - \boldsymbol{\theta}_r\right\|_{V_t} < 2\sqrt{2\left(d\log(6) + \log\left(\frac{1}{\delta}\right)\right)}. \tag{EC.9.7}$$

This result serves as the basis for designing the sampling budget and conducting theoretical analyses.

Together with Lemmas EC.9.1, EC.9.2, and EC.9.3, the distance $|\tilde{\mu}_i(r)-\mu_i|$ can be bounded as follows: with probability at least $1-\delta/(Kr(r+1))$, we have

$$\begin{aligned}
|\tilde{\mu}_i(r)-\mu_i| &\le \left|\left\langle \hat{\boldsymbol{\theta}}_o(r)-\boldsymbol{\theta}_r, \boldsymbol{a}_i\right\rangle\right| + |\langle \boldsymbol{\theta}_r-\boldsymbol{\theta}, \boldsymbol{a}_i\rangle| + \left|\hat{\boldsymbol{\Delta}}_{mo}(r)-\Delta_{mi}\right| \\
&\le \left\|\hat{\boldsymbol{\theta}}_o(r)-\boldsymbol{\theta}_r\right\|_{\boldsymbol{V}_r} \|\boldsymbol{a}_i\|_{\boldsymbol{V}_r^{-1}} + \|\boldsymbol{\theta}_r-\boldsymbol{\theta}\|_{\boldsymbol{V}_r} \|\boldsymbol{a}_i\|_{\boldsymbol{V}_r^{-1}} + 2L_m \\
&\le \left\|\hat{\boldsymbol{\theta}}_r-\boldsymbol{\theta}_r\right\|_{\boldsymbol{V}_r} \|\boldsymbol{a}_i\|_{\boldsymbol{V}_r^{-1}} + \|\boldsymbol{\theta}_r-\boldsymbol{\theta}\|_{\boldsymbol{V}_r} \|\boldsymbol{a}_i\|_{\boldsymbol{V}_r^{-1}} + 2L_m \\
&\le \left(2\sqrt{2\left(d\log(6)+\log\left(\frac{Kr(r+1)}{\delta}\right)\right)} + L_m\sqrt{T_r}\right)\sqrt{\frac{d}{T_r}} + 2L_m \\
&\le \varepsilon_r + (\sqrt{d}+2)L_m. \qquad \text{(EC.9.8)}
\end{aligned}$$

Then, we define a new clean event

$$\mathcal{E}'_{1m} = \left\{ \bigcap_{i\in\mathcal{A}_I(r-1)} \bigcap_{r\in\mathbb{N}} |\hat{\mu}_i(r)-\mu_i| \le \varepsilon_r + (\sqrt{d}+2)L_m \right\}. \qquad \text{(EC.9.9)}$$

This mirrors the event defined in equation (EC.8.1), allowing us to derive all corresponding results in Section EC.8.

### EC.9.2. Step 2: Bound the Expected Sample Complexity

LEMMA EC.9.4. *For $i \in G_\varepsilon$ and $L_m < \frac{\alpha_\varepsilon}{2(\sqrt{d}+2)}$, if $r \ge \left\lceil \log_2\left(\frac{4}{\varepsilon-\Delta_i-2L_m(\sqrt{d}+2)}\right)\right\rceil$, then we have $\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[i\notin G_r\right] \mid \mathcal{E}'_{1m}\right] = 0$.*

LEMMA EC.9.5. *For $i \in G_\varepsilon^c$ and $L_m < \frac{\beta_\varepsilon}{2(\sqrt{d}+2)}$, if $r \ge \left\lceil \log_2\left(\frac{4}{\Delta_i-\varepsilon-2L_m(\sqrt{d}+2)}\right)\right\rceil$, then we have $\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[i\notin B_r\right] \mid \mathcal{E}'_{1m}\right] = 0$.*

LEMMA EC.9.6. *For the misspecification magnitude $L_m < \min\left\{\frac{\alpha_\varepsilon}{2(\sqrt{d}+2)}, \frac{\beta_\varepsilon}{2(\sqrt{d}+2)}\right\}$, the round $R''_{upper} = \max\left\{\left\lceil \log_2 \frac{4}{\alpha_\varepsilon-2L_m(\sqrt{d}+2)}\right\rceil, \left\lceil \log_2 \frac{4}{\beta_\varepsilon-2L_m(\sqrt{d}+2)}\right\rceil\right\}$ marks the point at which all classifications are completed and the algorithm terminates under model misspecification when using an estimation procedure based on orthogonal parameterization.*

The proofs of Lemmas EC.9.4, EC.9.5, and EC.9.6 follow similar arguments to those presented in Section EC.8 and are therefore omitted for brevity.

LEMMA EC.9.7. *For the expected sample complexity given the high probability event $\mathcal{E}'_{1m}$, we have*

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\mu}}\left[T_{G_{\text{op}}} \mid \mathcal{E}'_{1m}\right] \le c\max\Bigg\{ &\frac{256d}{(\alpha_\varepsilon-2L_m(\sqrt{d}+2))^2}\log\left(\frac{K6^d}{\delta}\log_2\frac{8}{\alpha_\varepsilon-2L_m(\sqrt{d}+2)}\right), \\
&\frac{256d}{(\beta_\varepsilon-2L_m(\sqrt{d}+2))^2}\log\left(\frac{K6^d}{\delta}\log_2\frac{8}{\beta_\varepsilon-2L_m(\sqrt{d}+2)}\right)\Bigg\} + \frac{d(d+1)}{2}R''_{upper}, \qquad \text{(EC.9.10)}
\end{aligned}$$

*where $c$ is a universal constant and $R''_{upper} = \max\left\{\left\lceil \log_2 \frac{4}{\alpha_\varepsilon - 2L_m\left(\sqrt{d}+2\right)}\right\rceil, \left\lceil \log_2 \frac{4}{\beta_\varepsilon - 2L_m\left(\sqrt{d}+2\right)}\right\rceil\right\}$ denotes the round in which all classifications are completed and the algorithm terminates under model misspecification with orthogonal parameterization.*

*Proof.* The decomposition of $T$ in equation (EC.5.26) can be reformulated, where the expectations are conditioned on the high-probability event $\mathcal{E}'_{1m}$, given by

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\mu}}\left[T_{G_{\text{op}}} \mid \mathcal{E}'_{1m}\right] &\leq \sum_{r=1}^{\infty} \mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[G_r \cup B_r \neq [K]\right] \mid \mathcal{E}'_{1m}\right] \sum_{\boldsymbol{a}\in\mathcal{A}(r-1)} T_r(\boldsymbol{a}) \\
&\leq \sum_{r=1}^{R''_{\text{upper}}} \left(d2^{2r+3}\left(d\log(6) + \log\left(\frac{Kr(r+1)}{\delta}\right)\right) + \frac{d(d+1)}{2}\right) \\
&\leq \frac{d(d+1)}{2} R''_{\text{upper}} + 8d\log\left(\frac{K6^d}{\delta}\right) \sum_{r=1}^{R''_{\text{upper}}} 2^{2r} + 16d \sum_{r=1}^{R''_{\text{upper}}} 2^{2r}\log(r+1) \\
&\leq 16\log\left[\frac{K6^d}{\delta}\left(R''_{\text{upper}}+1\right)\right] \sum_{r=1}^{R''_{\text{upper}}} d2^{2r} + \frac{d(d+1)}{2} R''_{\text{upper}} \qquad \text{(EC.9.11)} \\
&\leq c\max\left\{\frac{256d}{(\alpha_\varepsilon - 2L_m(\sqrt{d}+2))^2}\log\left(\frac{K6^d}{\delta}\log_2\frac{16}{\alpha_\varepsilon - 2L_m(\sqrt{d}+2)}\right),\right. \\
&\left.\frac{256d}{(\beta_\varepsilon - 2L_m(\sqrt{d}+2))^2}\log\left(\frac{K6^d}{\delta}\log_2\frac{16}{\beta_\varepsilon - 2L_m(\sqrt{d}+2)}\right)\right\} + \frac{d(d+1)}{2} R''_{\text{upper}}, \qquad \text{(EC.9.12)}
\end{aligned}$$

where $c$ is a universal constant. Then, let $\xi = \min\left(\alpha_\varepsilon - 2L_m(\sqrt{d}+2), \beta_\varepsilon - 2L_m(\sqrt{d}+2)\right)/16$, we have

$$\mathbb{E}_{\boldsymbol{\mu}}\left[T_{G_{\text{op}}} \mid \mathcal{E}\right] = \mathcal{O}\left(d\xi^{-2}\log\left(\frac{K6^d}{\delta}\log(\xi^{-1})\right) + d^2\log(\xi^{-1})\right). \qquad \text{(EC.9.13)}$$

□

## EC.10. Proof of Proposition 4

The proof of this proposition closely follows that of Theorem 4 in Section EC.8, with the only modification being the definition of $C_{\delta/K}(r)$, which is now set to $\varepsilon_r + L_m\sqrt{d}$ given that $L_m$ is known in advance. The conclusions in Section EC.8.1 remain valid. Additionally, the marginal round in Lemma EC.8.1 is updated from $\left\lceil \log_2\left(\frac{4}{\varepsilon - \Delta_i - 2L_m\sqrt{d}}\right)\right\rceil$ to $\left\lceil \log_2\left(\frac{4}{\varepsilon - \Delta_i}\right)\right\rceil$. The remainder of the proof follows directly by applying the same reasoning steps.

## EC.11. Proof of Theorem 6

### EC.11.1. Step 1: Rearrange the Clean Event

Following the derivation approach in Section EC.5.5, the core of the proof is to similarly identify the round in which all classifications are completed under the GLM setting. To this end, we reconstruct

the anytime confidence radius for the arms in each round $r$ and define a high-probability event that holds throughout the execution of the algorithm.

Let $\check{\boldsymbol{V}}_r = \sum_{s=1}^{T_r} \dot{\mu}_{\text{link}}(\boldsymbol{a}^\top \check{\boldsymbol{\theta}}_r)\boldsymbol{a}_s\boldsymbol{a}_s^\top$, where $\check{\boldsymbol{\theta}}_r$ is some convex combination of true parameter $\boldsymbol{\theta}$ and parameter $\hat{\boldsymbol{\theta}}_r$ based on maximum likelihood estimation (MLE). It can be checked that the unweighted matrix $\boldsymbol{V}_r = \sum_{s=1}^{T_r}(\boldsymbol{a}^\top \check{\boldsymbol{\theta}}_r)\boldsymbol{a}_s\boldsymbol{a}_s^\top$ in the standard linear model is the special case of this newly defined matrix $\check{\boldsymbol{V}}_r$ when the inverse link function $\mu_{\text{link}}(x) = x$.

For each arm $i$, we define the following auxiliary vector

$$W_i = (W_{i,1}, W_{i,2}, \ldots, W_{i,T_r}) = \boldsymbol{a}_i^\top \check{\boldsymbol{V}}_r^{-1} \left(\boldsymbol{a}_{A_1}, \boldsymbol{a}_{A_2}, \ldots, \boldsymbol{a}_{A_{T_r}}\right) \in \mathbb{R}^{T_r}, \tag{EC.11.1}$$

and thus we have

$$\|W_i\|_2^2 = W_i W_i^\top = \boldsymbol{a}_i^\top \check{\boldsymbol{V}}_r^{-1} \boldsymbol{V}_r \check{\boldsymbol{V}}_r^{-1} \boldsymbol{a}_i. \tag{EC.11.2}$$

To give the confidence radius under GLM, we have the following statement for any arm $i \in \mathcal{A}_I(r-1)$ in round $r$.

$$\begin{aligned} |\hat{\mu}_i(r) - \mu_i| &= \left| \boldsymbol{a}_i^\top (\hat{\boldsymbol{\theta}}_r - \boldsymbol{\theta}) \right| \\ &= \left| \boldsymbol{a}_i^\top \check{\boldsymbol{V}}_r^{-1} \sum_{s=1}^{T_r} \boldsymbol{a}_{A_s} \eta_s \right| \\ &= \left| \sum_{s=1}^{T_r} W_{i,s} \eta_{A_s} \right|, \end{aligned} \tag{EC.11.3}$$

where the second equality is established with Lemma 1 of Kveton et al. (2023) and $T_r$, which is defined in equation (40), is the adjusted sampling budget in each round $r$ for the GLM. Since $(\eta_{A_s})_{s \in T_r}$ are independent, mean zero, 1-sub-Gaussian random variables, then $\sum_{s=1}^{T_r} W_{i,s}\eta_{A_s}$ is a $\|W_i\|_2$-sub-Gaussian variable for each arm $i$, then we have

$$\mathbb{P}\left(|\hat{\mu}_i(r) - \mu_i| > \varepsilon_r\right) \leq 2\exp\left(-\frac{\varepsilon_r^2}{2\|W_i\|_2^2}\right). \tag{EC.11.4}$$

Since $\check{\boldsymbol{\theta}}_r$ is not known in the process, we need to find another way to represent this term. By assumption, we know $\dot{\mu}_{\text{link}} \geq c_{\min}$ for some $c_{\min} \in \mathbb{R}^+$ and for all $i \in \mathcal{A}_I(r-1)$. Therefore $c_{\min}^{-1}\boldsymbol{V}_r^{-1} \succeq \check{\boldsymbol{V}}_r^{-1}$ by definition of $\check{\boldsymbol{V}}_r$, and then we have

$$\begin{aligned} \|W_i\|_2^2 &= \boldsymbol{a}_i^\top \check{\boldsymbol{V}}_r^{-1} \boldsymbol{V}_r \check{\boldsymbol{V}}_r^{-1} \boldsymbol{a}_i \\ &\leq \boldsymbol{a}_i^\top c_{\min}^{-1} \boldsymbol{V}_r^{-1} \boldsymbol{V}_r c_{\min}^{-1} \boldsymbol{V}_r^{-1} \boldsymbol{a}_i \\ &= c_{\min}^{-2} \|\boldsymbol{a}_i\|_{\boldsymbol{V}_r^{-1}}^2. \end{aligned} \tag{EC.11.5}$$

Furthermore, if G-optimal design is considered, we have $\|\boldsymbol{a}_i\|^2_{\boldsymbol{V}_r^{-1}} \le \frac{d}{T_r}$. Together with equation (EC.11.4), we have

$$\begin{aligned}\mathbb{P}\left(|\hat{\mu}_i(r)-\mu_i|>\varepsilon_r\right) &\le 2\exp\left(-\frac{\varepsilon_r^2}{2c_{\min}^{-2}\|\boldsymbol{a}_i\|^2_{\boldsymbol{V}_r^{-1}}}\right)\\ &\le 2\exp\left(-\frac{\varepsilon_r^2 c_{\min}^2}{2d}T_r\right). \end{aligned} \tag{EC.11.6}$$

Finally, considering the definition of $T_r$ in equation (40), with a probability of at least $1-\delta/Kr(r+1)$, we have

$$|\hat{\mu}_i(r)-\mu_i| \le \varepsilon_r. \tag{EC.11.7}$$

Thus, with the standard result of the G-optimal design, we still have

$$C_{\delta/K}(r) \coloneqq \varepsilon_r, \tag{EC.11.8}$$

with which the events $\mathcal{E}_1$ and $\mathcal{E}_2$ in Section EC.5.1 hold with a probability of at least $1-\delta$.

### EC.11.2. Step 2: Bound the Expected Sample Complexity

LEMMA EC.11.1. *For $i \in G_\varepsilon$, if $r \ge \left\lceil \log_2\left(\frac{4}{\varepsilon-\Delta_i}\right)\right\rceil$, then we have $\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[i \notin G_r\right] \mid \mathcal{E}_1\right] = 0$.*

LEMMA EC.11.2. *For $i \in G_\varepsilon^c$, if $r \ge \left\lceil \log_2\left(\frac{4}{\Delta_i-\varepsilon}\right)\right\rceil$, then we have $\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[i \notin B_r\right] \mid \mathcal{E}_1\right] = 0$.*

LEMMA EC.11.3. *$R_{\mathrm{GLM}} = R_{\mathrm{upper}} = \max\left\{\left\lceil \log_2 \frac{4}{\alpha_\varepsilon}\right\rceil, \left\lceil \log_2 \frac{4}{\beta_\varepsilon}\right\rceil\right\}$ is the round where all the classifications have been finished and the answer is returned under GLM.*

The proofs of Lemmas EC.11.1, EC.11.2, and EC.11.3 closely follow those in Section EC.5.5.

LEMMA EC.11.4. *For the expected sample complexity with high probability event $\mathcal{E}_1$, we have*

$$\mathbb{E}_{\boldsymbol{\mu}}\left[T_{\mathrm{GLM}} \mid \mathcal{E}_1\right] \le c\max\left\{\frac{256d}{\alpha_\varepsilon^2 c_{\min}^2}\log\left(\frac{2K}{\delta}\log_2\frac{16}{\alpha_\varepsilon}\right), \frac{256d}{\beta_\varepsilon^2 c_{\min}^2}\log\left(\frac{2K}{\delta}\log_2\frac{16}{\beta_\varepsilon}\right)\right\} + \frac{d(d+1)}{2}R_{\mathrm{GLM}}, \tag{EC.11.9}$$

*where $c$ is a universal constant, $c_{\min}$ is the known constant controlling the first-order derivative of the inverse link function, and $R_{\mathrm{GLM}} = R_{\mathrm{upper}} = \max\left\{\left\lceil \log_2 \frac{4}{\alpha_\varepsilon}\right\rceil, \left\lceil \log_2 \frac{4}{\beta_\varepsilon}\right\rceil\right\}$ is the round where all the classifications have been finished and the answer is returned under GLM.*

*Proof.* We also consider the decomposition of $T$ in equation (EC.5.26), where all expectations are conditioned on the high-probability event $\mathcal{E}_1$, given by

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\mu}}\left[T_{\mathrm{GLM}} \mid \mathcal{E}_1\right] &\le \sum_{r=1}^{\infty}\mathbb{E}_{\boldsymbol{\mu}}\left[\mathbb{1}\left[G_r \cup B_r \ne [K]\right] \mid \mathcal{E}_1\right] \sum_{\boldsymbol{a}\in\mathcal{A}(r-1)} T_r(\boldsymbol{a})\\ &\le \sum_{r=1}^{R_{\mathrm{GLM}}}\left(\frac{d2^{2r+1}}{c_{\min}^2}\log\left(\frac{2Kr(r+1)}{\delta}\right) + \frac{d(d+1)}{2}\right)\end{aligned}$$

$$\leq \frac{d(d+1)}{2} R_{\mathrm{GLM}} + 2c_{\min}^{-2} d \log\left(\frac{2K}{\delta}\right) \sum_{r=1}^{R_{\mathrm{GLM}}} 2^{2r} + 4c_{\min}^{-2} d \sum_{r=1}^{R_{\mathrm{GLM}}} 2^{2r} \log(r+1)$$

$$\leq 4c_{\min}^{-2} \log\left[\frac{2K}{\delta}(R_{\mathrm{GLM}}+1)\right] \sum_{r=1}^{R_{\mathrm{GLM}}} d2^{2r} + \frac{d(d+1)}{2} R_{\mathrm{GLM}} \tag{EC.11.10}$$

$$\leq c \max\left\{\frac{256d}{\alpha_\varepsilon^2 c_{\min}^2} \log\left(\frac{2K}{\delta}\log_2\frac{16}{\alpha_\varepsilon}\right), \frac{256d}{\beta_\varepsilon^2 c_{\min}^2} \log\left(\frac{2K}{\delta}\log_2\frac{16}{\beta_\varepsilon}\right)\right\} + \frac{d(d+1)}{2} R_{\mathrm{GLM}}, \tag{EC.11.11}$$

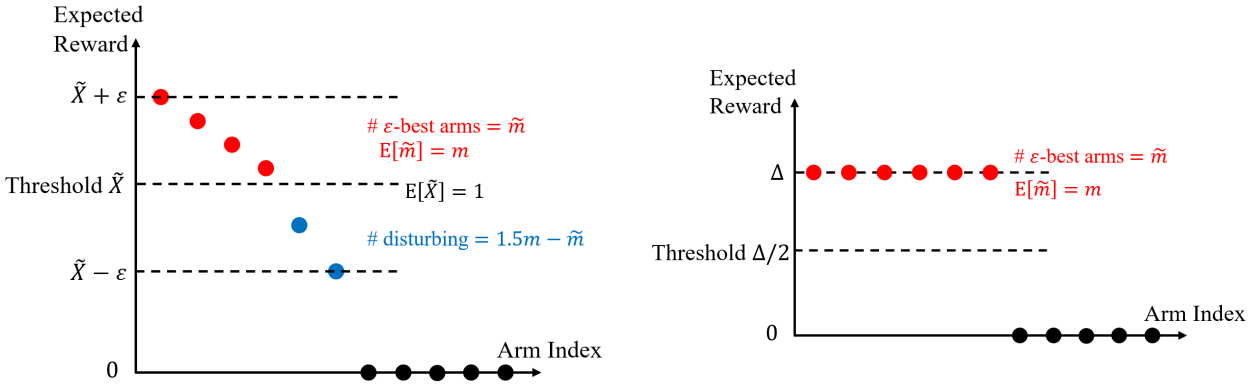
where $c$ is a universal constant. Then, let $\xi = \min(\alpha_\varepsilon, \beta_\varepsilon)/16$ denoting the minimum gap of the problem instance, we have

$$\mathbb{E}[T_{\mathrm{GLM}} \mid \mathcal{E}] = \mathcal{O}\left(\frac{d}{c_{\min}^2}\xi^{-2}\log\left(\frac{K}{\delta}\log_2(\xi^{-2})\right) + d^2\log(\xi^{-1})\right). \tag{EC.11.12}$$

□

## EC.12. Detailed Settings for Synthetic Experiments

We recap the figure here for better clarity.




(a) Synthetic I - Adaptive Setting

(b) Synthetic II - Static Setting

**Figure EC.12.1** Illustration on Synthetic Settings

### EC.12.1. Synthetic I - Adaptive Setting.

First, we randomly sample $\tilde{m}$, representing the number of $\varepsilon$-best arms, from a distribution with an expected value of $m$ (used as input for a top $m$ algorithm). Next, we randomly sample $\tilde{X}$, representing the best arm reward minus $\varepsilon$, from a distribution with an expected value of $X$ (used as input for a threshold bandit algorithm). We then assign $\tilde{m}$ $\varepsilon$-best arms with expected rewards uniformly distributed between $\tilde{X}+\varepsilon$ and $\tilde{X}$. Additionally, we assign $(1.5m - \tilde{m})$ arms that are not $\varepsilon$-best with expected rewards uniformly distributed between $\tilde{X}$ and $\tilde{X}-\varepsilon$, as illustrated in Figure EC.12.1.

Based on these designed arm rewards, we define the linear model parameter as:

$$\boldsymbol{\theta} = (\tilde{X} + \varepsilon, \tilde{X} + (\tilde{m} - 1)\varepsilon/\tilde{m}, \ldots, \tilde{X} + \varepsilon/\tilde{m}, 0, \ldots, 0)^\top.$$

Arms are $d$-dimensional canonical basis $e_1, e_2, \ldots, e_d$ and $(1.5m - \tilde{m})$ additional disturbing arms

$$\boldsymbol{x_i} = \left( \frac{\tilde{X} - (1.5m - \tilde{m} - i)\varepsilon/(1.5m - \tilde{m})}{\tilde{X} + \varepsilon}, 0, \cdots, 0, \sqrt{1 - \left( \frac{\tilde{X} - (1.5m - \tilde{m} - i)\varepsilon/(1.5m - \tilde{m})}{\tilde{X} + \varepsilon} \right)^2} \right)^\top$$

with $i \in [1.5m - \tilde{m}]$.

In the adaptive setting, pulling one arm can provide information about the distributions of other arms. The optimal policy in this setting should adaptively refine its sampling and stopping strategy based on historical data. This allows the algorithm to focus more on the disturbing arms, making adaptive strategies particularly effective as the algorithm progresses. In our experiments, we set $m = 4$, $X = 1$, with $d = 10$ and $\varepsilon \in \{0.1, 0.2, 0.3\}$. A total of six different problem instances are evaluated to compare the performance of the algorithms.

### EC.12.2. Synthetic II - Static Setting.

We consider a static synthetic setting, similar to the one proposed by Xu et al. (2018), where arms are represented as $d$-dimensional canonical basis vectors $e_1, e_2, \ldots, e_d$. We set the parameter vector $\boldsymbol{\theta} = (\Delta, \ldots, \Delta, 0, \ldots, 0)^\top$, where $\tilde{m}$ elements are $\Delta$ and $d - \tilde{m}$ elements are 0. In this setting, $\mathbb{E}[\tilde{m}] = m$, and only the value $m$ is provided as input to top $m$ algorithms. Consequently, the true mean values consist of some $\Delta$'s and some 0's.

If we set $\varepsilon = \Delta/2$, as $\Delta$ approaches 0, it becomes difficult to distinguish between the $\varepsilon$-best arms and the arms that are not $\varepsilon$-best. In the static setting, knowledge of the rewards does not alter the sampling strategy, as all arms must be estimated with equal accuracy to effectively differentiate between them. Therefore, a static policy is optimal in this case, and the goal of this setting is to assess the ability of our algorithm to adapt to such static conditions. In our experiment, we set $m = 4$ with $d \in \{8, 12, 16\}$ and $\Delta = 1$. A total of three different problem instances are evaluated to compare the algorithms.

## EC.13. Auxiliary Results

The following lemma shows that matrix inversion reverses the order relation.

LEMMA EC.13.1. *(Inversion Reverses Loewner orders) Let $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{d \times d}$. Suppose that $\boldsymbol{A} \succeq \boldsymbol{B}$ and $\boldsymbol{B}$ is invertible, we have*

$$\boldsymbol{A}^{-1} \preceq \boldsymbol{B}^{-1}. \tag{EC.13.1}$$

*Proof.* By definition, to show $\boldsymbol{B}^{-1}-\boldsymbol{A}^{-1}$ is a positive semi-definite matrix, it suffices to show that $\|\boldsymbol{x}\|^2_{\boldsymbol{B}^{-1}}-\|\boldsymbol{x}\|^2_{\boldsymbol{A}^{-1}}=\|\boldsymbol{x}\|^2_{\boldsymbol{B}^{-1}-\boldsymbol{A}^{-1}}\geq 0$ for any $\boldsymbol{x}\in\mathbb{R}^d$. Then, by the Cauchy-Schwarz inequality,

$$\|\boldsymbol{x}\|^2_{\boldsymbol{A}^{-1}}=\left\langle \boldsymbol{x},\boldsymbol{A}^{-1}\boldsymbol{x}\right\rangle \leq \|\boldsymbol{x}\|_{\boldsymbol{B}^{-1}}\left\|\boldsymbol{A}^{-1}\boldsymbol{x}\right\|_{\boldsymbol{B}}\leq \|\boldsymbol{x}\|_{\boldsymbol{B}^{-1}}\left\|\boldsymbol{A}^{-1}\boldsymbol{x}\right\|_{\boldsymbol{A}}=\|\boldsymbol{x}\|_{\boldsymbol{B}^{-1}}\|\boldsymbol{x}\|_{\boldsymbol{A}^{-1}}. \tag{EC.13.2}$$

Hence $\|\boldsymbol{x}\|_{\boldsymbol{A}^{-1}}\leq\|\boldsymbol{x}\|_{\boldsymbol{B}^{-1}}$ for all $\boldsymbol{x}$, which completes the lemma. □

The following lemma establishes an upper bound on the ratio between two optimization problems that incorporate instance-specific information from the bandit setting.

LEMMA EC.13.2. *We always have $\mathcal{A}_I=[K]$,* i.e., *the entire set of arms is under consideration. For any arm $i\in\mathcal{A}_I\cap G_\varepsilon\setminus\{1\}$, we have*

$$\frac{\min_{\boldsymbol{p}\in S_K}\max_{i\in\mathcal{A}_I\cap G_\varepsilon\setminus\{1\}}\|\boldsymbol{y}_{1,i}\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}}}{\min_{\boldsymbol{p}\in S_K}\min_{i\in\mathcal{A}_I\cap G_\varepsilon\setminus\{1\}}\|\boldsymbol{y}_{1,i}\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}}}\leq\frac{dL_1}{\mathcal{G}_{\mathcal{Y}}{}^2L_2}. \tag{EC.13.3}$$

*Proof.* For any arm $i\in[K]$, from a perspective of geometry quantity, let $\mathrm{conv}(\mathcal{A}\cup-\mathcal{A})$ denote the convex hull of symmetric $\mathcal{A}\cup-\mathcal{A}$. Then for any set $\mathcal{Y}\subset\mathbb{R}^d$ define the gauge of $\mathcal{Y}$ as

$$\mathcal{G}_{\mathcal{Y}}=\max\left\{c>0:c\mathcal{Y}\subseteq\mathrm{conv}(\mathcal{A}\cup-\mathcal{A})\right\}. \tag{EC.13.4}$$

We then provide a natural upper bound for $\min_{\boldsymbol{p}\in S_K}\max_{i\in\mathcal{A}_I\cap G_\varepsilon\setminus\{1\}}\|\boldsymbol{y}_{1,i}\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}}$, given by

$$\begin{aligned}\min_{\boldsymbol{p}\in S_K}\max_{i\in\mathcal{A}_I\cap G_\varepsilon\setminus\{1\}}\|\boldsymbol{y}_{1,i}\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}}&\leq\min_{\boldsymbol{p}\in S_K}\max_{y\in\mathcal{Y}(\mathcal{A}_I)}\|\boldsymbol{y}\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}}\\&=\frac{1}{\mathcal{G}_{\mathcal{Y}}{}^2}\min_{\boldsymbol{p}\in S_K}\max_{y\in\mathcal{Y}(\mathcal{A}_I)}\|\boldsymbol{y}\mathcal{G}_{\mathcal{Y}}\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}}\\&\leq\frac{1}{\mathcal{G}_{\mathcal{Y}}{}^2}\min_{\boldsymbol{p}\in S_K}\max_{\boldsymbol{a}\in\mathrm{conv}(\mathcal{A}\cup-\mathcal{A})}\|\boldsymbol{a}\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}}\\&=\frac{1}{\mathcal{G}_{\mathcal{Y}}{}^2}\min_{\boldsymbol{p}\in S_K}\max_{i\in\mathcal{A}_I}\|\boldsymbol{a}_i\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}}\\&\leq\frac{d}{\mathcal{G}_{\mathcal{Y}}{}^2},\end{aligned} \tag{EC.13.5}$$

where the third line follows from the fact that the maximum value of a convex function on a convex set must occur at a vertex. With the Kiefer-Wolfowitz Theorem for the G-optimal design, the last inequality is achieved.

Furthermore, for any arm $i\in\mathcal{A}_I\cap G_\varepsilon\setminus\{1\}$, we have

$$\begin{aligned}\min_{\boldsymbol{p}\in S_K}\min_{i\in\mathcal{A}_I\cap G_\varepsilon\setminus\{1\}}\|\boldsymbol{y}_{1,i}\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}}&\geq\min_{\boldsymbol{p}\in S_K}\min_{i\in\mathcal{A}_I\cap G_\varepsilon\setminus\{1\}}\mathrm{eig}_{\min}(\boldsymbol{V}_{\boldsymbol{p}}^{-1})\|\boldsymbol{y}_{1,i}\|_2^2\\&=\min_{\boldsymbol{p}\in S_K}\min_{i\in\mathcal{A}_I\cap G_\varepsilon\setminus\{1\}}\frac{1}{\mathrm{eig}_{\max}(\boldsymbol{V}_{\boldsymbol{p}})}\|\boldsymbol{y}_{1,i}\|_2^2\\&\geq\frac{1}{\max_{i\in\mathcal{A}_I}\|\boldsymbol{a}_i\|_2}\min_{\boldsymbol{p}\in S_K}\min_{i\in\mathcal{A}_I\cap G_\varepsilon\setminus\{1\}}\|\boldsymbol{y}_{1,i}\|_2^2,\end{aligned} \tag{EC.13.6}$$

where $\text{eig}_{\max}(\cdot)$ and $\text{eig}_{\min}(\cdot)$ are respectively the largest and smallest eigenvalues of a matrix. The first line follows from the Rayleigh Quotient and Rayleigh Theorem. The last line is derived by the relationship $\text{eig}_{\max}(\boldsymbol{V}_{\boldsymbol{p}}) \leq \max_{i\in\mathcal{A}_I} \|\boldsymbol{a}_i\|_2$. Recall the assumption in Theorem 3 that $\min_{i\in G_\varepsilon\setminus\{1\}} \|\boldsymbol{a}_1 - \boldsymbol{a}_i\|^2 \geq L_2$ and the assumption in Section 2 that $\|\boldsymbol{a}_i\|_2 \leq L_1$ for $\forall i \in [K]$, we have

$$\min_{\boldsymbol{p}\in S_K} \min_{i\in\mathcal{A}_I\cap G_\varepsilon\setminus\{1\}} \|\boldsymbol{y}_{1,i}\|^2_{\boldsymbol{V}_{\boldsymbol{p}}^{-1}} \geq \frac{L_2}{L_1}. \tag{EC.13.7}$$

Finally, combining inequalities (EC.13.5) and (EC.13.7) completes the lemma. □

LEMMA EC.13.3 **(Kiefer and Wolfowitz (1960))**. *If the arm vectors $\boldsymbol{a} \in \mathcal{A}$ span $\mathbb{R}^d$, then for any probability distribution $\pi \in \mathcal{P}(\mathcal{A})$, the following statements are equivalent:*

1. *$\pi^*$ minimizes the function $g(\pi) = \max_{\boldsymbol{a}\in\mathcal{A}} \|\boldsymbol{a}\|^2_{\boldsymbol{V}(\pi)^{-1}}$.*
2. *$\pi^*$ maximizes the function $f(\pi) = \log\det \boldsymbol{V}(\pi)$.*
3. *$g(\pi^*) = d$.*

*Additionally, there exists a $\pi^*$ of $g(\pi)$ such that the size of its support, $|Supp(\pi^*)|$, is at most $d(d+1)/2$.*

## Appendix References


Abbasi-Yadkori Y, Pál D, Szepesvári C (2011) Improved algorithms for linear stochastic bandits. *Advances in neural information processing systems* 24.

Abe N, Long PM (1999) Associative reinforcement learning using linear probabilistic concepts. *ICML*, 3–11 (Citeseer).

Ahn D, Shin D (2020) Ordinal optimization with generalized linear model. *2020 Winter Simulation Conference (WSC)*, 3008–3019 (IEEE).

Ahn D, Shin D, Zeevi A (2024) Feature misspecification in sequential learning problems. *Management Science* 0(0):null.

Azizi MJ, Kveton B, Ghavamzadeh M (2021a) Fixed-budget best-arm identification in contextual bandits: A static-adaptive algorithm. *CoRR* abs/2106.04763.

Azizi MJ, Kveton B, Ghavamzadeh M (2021b) Fixed-budget best-arm identification in structured bandits. *arXiv preprint arXiv:2106.04763* .

Chaloner K, Verdinelli I (1995) Bayesian experimental design: A review. *Statistical science* 273–304.

Chapelle O, Li L (2011) An empirical evaluation of thompson sampling. *Advances in neural information processing systems* 24.

Fiez T, Jain L, Jamieson KG, Ratliff L (2019) Sequential experimental design for transductive linear bandits. *Advances in neural information processing systems* 32.

Filippi S, Cappe O, Garivier A, Szepesvári C (2010) Parametric bandits: The generalized linear case. *Advances in neural information processing systems* 23.

Gabillon V, Ghavamzadeh M, Lazaric A (2012) Best arm identification: A unified approach to fixed budget and fixed confidence. *Advances in Neural Information Processing Systems* 25.

Garivier A, Kaufmann E (2016) Optimal best arm identification with fixed confidence. *Conference on Learning Theory*, 998–1027 (PMLR).

Ghosh A, Chowdhury SR, Gopalan A (2017) Misspecified linear bandits. *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31.

Hoffman M, Shahriari B, Freitas N (2014) On correlation and budget constraints in model-based bandit optimization with application to automatic machine learning. *Artificial Intelligence and Statistics*, 365–374 (PMLR).

Kaufmann E, Cappé O, Garivier A (2016) On the complexity of best arm identification in multi-armed bandit models. *Journal of Machine Learning Research* 17:1–42.

Kaufmann E, Koolen WM (2021) Mixture martingales revisited with applications to sequential tests and confidence intervals. *The Journal of Machine Learning Research* 22(1):11140–11183.

Kiefer J, Wolfowitz J (1960) The equivalence of two extremum problems. *Canadian Journal of Mathematics* 12:363–366.

Kveton B, Zaheer M, Szepesvari C, Li L, Ghavamzadeh M, Boutilier C (2023) Randomized exploration in generalized linear bandits.

Lattimore T, Szepesvári C (2020) *Bandit algorithms* (Cambridge University Press).

Lattimore T, Szepesvari C, Weisz G (2020) Learning with good feature representations in bandits and in rl with a generative model.

Li L, Chu W, Langford J, Schapire RE (2010) A contextual-bandit approach to personalized news article recommendation. *Proceedings of the 19th International Conference on World Wide Web*, 661–670, WWW '10 (New York, NY, USA: Association for Computing Machinery), ISBN 9781605587998.

McCullagh P (2019) *Generalized linear models* (Routledge).

Qin C, You W (2025) Dual-directed algorithm design for efficient pure exploration. *Operations Research* .

Réda C, Tirinzoni A, Degenne R (2021) Dealing with misspecification in fixed-confidence linear top-m identification. *Advances in Neural Information Processing Systems* 34:25489–25501.

Russo DJ, Van Roy B, Kazerouni A, Osband I, Wen Z, et al. (2018) A tutorial on thompson sampling. *Foundations and Trends® in Machine Learning* 11(1):1–96.

Soare M, Lazaric A, Munos R (2014) Best-arm identification in linear bandits. *Advances in Neural Information Processing Systems* 27.

Wang PA, Tzeng RC, Proutiere A (2021) Fast pure exploration via frank-wolfe. *Advances in Neural Information Processing Systems* 34:5810–5821.

Xu L, Honda J, Sugiyama M (2018) A fully adaptive algorithm for pure exploration in linear bandits. *International Conference on Artificial Intelligence and Statistics*, 843–851 (PMLR).

Yang J, Tan VY (2021) Minimax optimal fixed-budget best arm identification in linear bandits. *arXiv preprint arXiv:2105.13017* .