Minimum Empirical Divergence for Sub-Gaussian Linear Bandits

Kapilan Balagopalan
University of Arizona
kapilanbgp@arizona.edu

Kwang-Sung Jun
University of Arizona
kjun@cs.arizona.edu

Abstract

We propose a novel linear bandit algorithm called LinMED (Linear Minimum Empirical Divergence), which is a linear extension of the MED algorithm that was originally designed for multi-armed bandits. LinMED is a randomized algorithm that admits a closed-form computation of the arm sampling probabilities, unlike the popular randomized algorithm called linear Thompson sampling. Such a feature proves useful for off-policy evaluation where the unbiased evaluation requires accurately computing the sampling probability. We prove that LinMED enjoys a near-optimal regret bound of $d\sqrt{n}$ up to logarithmic factors where d is the dimension and n is the time horizon. We further show that LinMED enjoys a $\frac{d^2}{\Delta}\left(\log^2(n)\right)\log\left(\log(n)\right)$ problem-dependent regret where Δ is the smallest suboptimality gap. Our empirical study shows that LinMED has a competitive performance with the state-of-the-art algorithms.

1 INTRODUCTION

The multi-armed bandit problem represents a stateless reinforcement learning framework with numerous real-world applications. One of its most prominent applications is in recommendation systems, which are extensively employed by e-commerce platforms (Gangan et al., 2021), digital streaming services (Gangan et al., 2021; Mary et al., 2015), news portals (Li et al., 2010), and a variety of other platforms experiencing significant economic growth. The multi-armed bandit problem has spawned several important variants, including stochastic linear bandits, adversarial bandits, and best-arm identification, all of which share a common underlying structure but are adapted to

different environments to achieve distinct goals. This has fostered a rich and robust area of research.

In bandit problems, the main objective is to minimize cumulative regret by learning to select optimal arms over time. A key challenge is maintaining a balance between exploration (gathering information about the mean rewards of various arms) and exploitation (leveraging gathered information to take arms with large estimated rewards). Focusing exclusively on either strategy would be sub-optimal for minimizing cumulative regret.

Stochastic linear bandits, in particular, generalize the classical multi-armed bandit framework. At each time step $t \in \{1, 2, \ldots\}$, the learner selects an arm A_t from a set of arms $\mathcal{A}_t \subset \mathbb{R}^d$, and observes a reward given by $Y_t = \langle A_t, \theta^* \rangle + \eta_t$, where $\theta^* \in \mathbb{R}^d$ is an unknown parameter and η_t is a zero-mean noise. The learner's objective is to minimize the cumulative (pseudo-)regret over the time horizon n, which is defined by:

$$\text{Reg}_n := \sum_{t=1}^{n} \max_{a \in \mathcal{A}_t} \langle a, \theta^* \rangle - \langle A_t, \theta^* \rangle . \quad (1)$$

Since linear bandits generalize multi-armed bandits, many linear bandit algorithms have been derived by adapting algorithmic principles from multi-armed bandits. For instance, LinUCB (Dani et al., 2008; Abbasi-Yadkori et al., 2011) extends the optimism principle from UCB (Auer et al., 2002), linear Thompson sampling (Agrawal and Goyal, 2014) extends Thompson sampling (Thompson, 1933; Agrawal and Goyal, 2017), and LinIMED (Bian and Tan, 2024) extends IMED (Honda and Takemura, 2015). Therefore, it is natural to explore the stochastic linear bandit version of the MED framework (Honda and Takemura, 2011).

Evaluating bandit algorithms for recommendation systems typically requires running it live with the customers. However, this severely costs the user experience if the algorithm has a poor performance. Off-policy evaluation (OPE) (Precup, 2000) aims to address this issue by evaluating an algorithm (i.e., target policy) using the data collected by another algorithm (i.e., log-

ging policy). The standard method for OPE is inverse propensity weighting (IPW) (Horvitz and Thompson, 1952), which provides an unbiased estimator for the target policy's performance. For this to work, the logging policy is required to satisfy two properties. First, it must assign a nonzero sampling probability to every arm because otherwise we obtain no information on the zero-probability arms, disallowing counterfactual inference on their rewards. This automatically excludes any algorithm that makes a deterministic arm selection conditioning on previous observations. Second, the logging policy must allow accurate computation of the sampling probability. This is because IPW uses inverse sampling probability as an importance weight to scale the observed rewards. Thus, when the sampling probability is small, even a small error can be detrimental. We remark that by accurate computation we mean the extra computation *in addition to* running the algorithm itself. For example, any algorithm that computes the assigned probability for each arm first and then samples an arm would require zero extra computation. On the other hand, (linear) Thompson sampling (Thompson, 1933; Agrawal and Goyal, 2014) itself does not compute the sampling probability, so extra computation is needed. We call algorithms that satisfy the two properties above to be *OPE-friendly*.

In this paper, we propose a novel linear bandit algorithm called LinMED (Linear Minimum Empirical Divergence), which is a linear version of Minimum Empirical Divergence (Honda and Takemura, 2011). LinMED has numerous merits.

First, LinMED is OPE-friendly. This is in stark contrast to the popular randomized algorithm linear Thompson sampling (LinTS) (Agrawal and Goyal, 2014) that is not OPE-friendly. LinTS does not have a known closed-form solution or efficient methods for computing arm sampling probabilities and may assign zero probabilities to many arms. Note that using Monte Carlo sampling for estimating the probability up to the target precision has the time complexity of $O(1/\sqrt{\text{precision}})$ where precision is the desired floating point precision, which is quite large for a numerical approximation method (vs, say, $\log(1/\text{precision})$ of the bisection method) In particular, as discussed above, the error in probability goes to the denominator and further amplifies the error for IPW. We numerically verify such a phenomenon in Figure 1. Specifically, we take the uniform policy, which assigns equal probabilities to each arm, as the target policy. We evaluate IPW based on two logging policies, LinMED and LinTS, respectively. Given the logged data $(A_t, p_t(A_t), Y_t)_{t=0}^n$ where $p_t(A_t)$ is probability of sampling arm A_t from the logging policy, the IPW score for the uniform policy
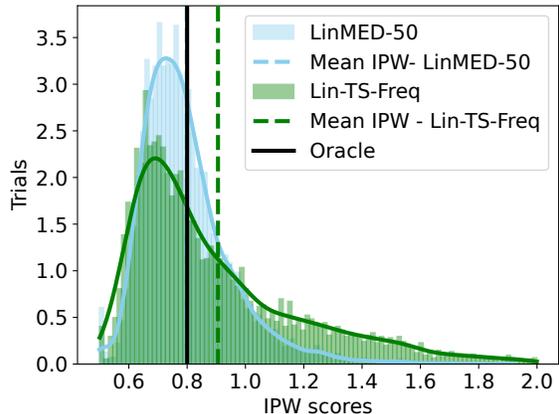
Figure 1: IPW scores of the uniform policy when the logging policy is LinMED and LinTS respectively. We used 1,000 Monte Carlo samples to estimate the sampling probabilities of LinTS. Oracle denotes the expected reward of the uniform policy. LinTS shows a nontrivial amount of bias, unlike LinMED (mean of LinMED is exactly aligned with the oracle, thus invisible in the plot). See Appendix G.3 for details.

is defined as

$$\text{IPW score} = \frac{1}{n} \sum_{t=1}^n \frac{\frac{1}{|\mathcal{A}|}}{p_t(A_t)} \cdot Y_t$$

As discussed for LinTS, the logged probability $p_t(A_t)$ must be estimated via Monte Carlo sampling. We selected the arm set $\mathcal{A} = \{a_1 = (1,0)^\top, a_2 = (0.6, 0.8)^\top\}$ and $\theta^* = (1,0)^\top$. It is expected that an OPE-friendly algorithm would yield an IPW score as an unbiased estimator of the expected reward of the uniform policy, which is 0.8. As shown in Figure 1, the mean IPW w.r.t. LinMED is almost identical to the true value 0.8 while that w.r.t. LinTS exhibits a significant bias.

Second, LinMED achieves not only a near-optimal minimax regret bound of $\tilde{O}(d\sqrt{n})$ (Dani et al., 2008), where $\tilde{O}$ omits logarithmic factors, but also an instance-dependent regret bound of $O(\frac{d^2}{\Delta} \log^2(n))$ where Δ is the smallest gap as defined in (9). To our knowledge, the only existing linear bandit algorithm with an nonasymptotic instance-dependent regret bound is OFUL (Abbasi-Yadkori et al., 2011). LinMED stands out even more when compared against randomized algorithms that allow closed-form computation of sampling probability, namely SquareCB (Foster and Rakhlin, 2020), EXP2 (Bubeck and Cesa-Bianchi, 2012), and SpannerIGW (Zhu et al., 2022), because they provably have sub-optimal instance-dependent regret of $\Omega(\Delta\sqrt{n})$, as we show later in Theorems 6 and 7 and numerically confirm in Section 7. We summarize the comparison of LinMED with other methods in Table 1.

Third, our analysis reveals that LinMED enjoys sub-linear regret bounds even if the sub-Gaussian noise parameter σ_*^2 is *under*-specified in the algorithm, albeit with an extra factor that grows with the degree of under-specification. This is in stark contrast to existing algorithms and their analyses that only provides a valid regret bound when the sub-Gaussian parameter is *over*-specified (Abbasi-Yadkori et al., 2011; Agrawal and Goyal, 2014). A more detailed discussion is provided in Section 5.

Finally, LinMED demonstrates outstanding empirical performance across various challenging scenarios, including delayed reward settings (see Appendix G.2) and "end of optimism" instance. A comprehensive discussion of these results is provided in Section 7.

Organization. In Section 2, we introduce the problem formulation and key notations. This is followed by the presentation of a warm up version of LinMED in Section 3 where we also provide a brief discussion on its connection to Maillard sampling (Bian and Jun, 2022) and SpannerIGW (Zhu et al., 2022) highlighting the importance of optimal experimental design for large arm sets. This is followed by the presentation of LinMED algorithm in Section 4. Next, we move to the main results in Section 5 where we establish the regret bounds of our algorithm. Additionally, in Section 6, we discuss the instance-dependent lower bounds for SpannerIGW and EXP2. Finally, Section 7 presents empirical studies to support our theoretical findings.

2 PRELIMINARIES

Notations. For any d dimensional vector $x \in \mathbb{R}^d$ and a $d \times d$ positive definite matrix A, we use $\|x\|_A$ to denote the Mahalanobis norm $\sqrt{x^\top A x}$ and we use $\|x\|$ to denote the Euclidean norm. We use $a \wedge b$ (resp. $a \vee b$) to denote the minimum (resp. maximum) of two real numbers a and b. For a set $\mathcal{B} \subset \mathbb{R}^d$, denote by $\Delta(\mathcal{B})$ the set of all probability measures on $\mathcal{B}$. The notation $\tilde{O}(\cdot)$ omits the logarithmic factors from the standard big-O notation $O(\cdot)$. For example, $A \log B = \tilde{O}(A)$. For any event $\mathcal{E}$, the complement of the event is denoted by $\overline{\mathcal{E}}$. We denote $a_i, a_{i+1}, \ldots, a_j$ by $a_{i:j}$.

The Stochastic Linear Bandit Model. In the stochastic linear bandit model, the learner chooses an arm A_t in each round t from the arm set $\mathcal{A}_t \subset \mathbb{R}^d$. After choosing arm A_t, the environment reveals a reward

$$Y_t = \langle \theta^*, A_t \rangle + \eta_t$$

to the learner where $\theta^* \in \mathbb{R}^d$ is an unknown coefficient of the linear model, η_t is a σ_*^2-sub-Gaussian noise conditioned on $A_{1:t}$ and $Y_{1:t-1}$. That is, for any $\lambda \in \mathbb{R}$,

almost surely,

$$\mathbb{E}\left[\exp\left(\lambda \eta_t\right) \mid A_{1:t}, Y_{1:t-1}\right] \le \exp\left(\frac{\lambda^2 \sigma_*^2}{2}\right).$$

Further, denote by $a_t^* := \arg\max_{a \in \mathcal{A}_t} \langle \theta^*, a \rangle$ the arm with the largest mean reward at time t. The goal of the learner is to minimize the cumulative (pseudo-)regret over the horizon n, which is precisely defined in (1). Throughout the paper, we focus on analyzing the expected (pseudo-)regret $\mathbb{E}\,\text{Reg}_n$ rather than a high probability bound. We also assume the following,

Assumption 1. For all $t \ge 1$, every arm $a \in \mathcal{A}_t$ satisfies $\|a\|_2 \le 1$. Furthermore, for some constant B, $\forall t \ge 1$ $\Delta_{a,t} := \langle \theta^*, a_t^* \rangle - \langle \theta^*, a \rangle \le B, \forall a \in \mathcal{A}_t$.

Note that prior linear bandit studies make the assumption of knowing the value of σ and S such that $\sigma_*^2 \le \sigma^2$ and $\|\theta^*\| \le S$, which accounts for the case of over-specification but not under-specification[1]. Instead, we analyze the regret of our proposed algorithm for arbitrarily given σ and S as guesses about σ_* and $\|\theta^*\|$, accounting for both over- and under-specification.

3 WARMUP: A LINEAR EXTENSION OF MINIMUM EMPIRICAL DIVERGENCE

In multi-armed bandits, we are given K arms and required to repeatedly choose an arm $A_t \in [K]$ to pull and observe its stochastic reward to maximize the cumulative reward. MED (Honda and Takemura, 2011) is a randomized multi-armed bandit algorithm that is optimized for bounded rewards (and achieved improved regret bounds compared to those that are optimized for sub-Gaussian rewards, which is a larger class of reward distributions) where the algorithm principle can be instantiated for (sub-)Gaussian rewards, which appeared first in Maillard (2013) and further analyzed in Bian and Jun (2022). Maillard sampling, or sub-Gaussian MED, samples arm from the following distribution:

$$p_{t,a} = \frac{\exp\left(-\frac{N_{t-1,a}}{2}\hat{\Delta}_{a,t-1}^2\right)}{\sum_{b \in \mathcal{A}} \exp\left(-\frac{N_{t-1,b}}{2}\hat{\Delta}_{b,t-1}^2\right)}$$

where $\hat{\Delta}_{a,t-1} = \max_{a' \in [K]} \hat{\mu}_{t-1,a'} - \hat{\mu}_{t-1,a}$ is the estimated reward gap at t, and $\hat{\mu}_{t,a}$ and $N_{t,a}$ are the empirical mean reward of arm a based on past rewards from arm a and the pull count of arm a, respectively. Throughout, we simply say MED for this instance and refer both of them interchangeably since they are the same in spirit.

Towards a linear extension of MED, one may consider

---

[1] Gales et al. (2022) adapt to the unknown norm $\|\theta^*\|$ but not the sub-Gaussian parameter σ_*^2.

Algorithms	Minimax regret	Instance-dependent regret	Efficiently computable probability	Probability assigned for all arms
OFUL (Abbasi-Yadkori et al., 2011)	$\tilde{O}(d\sqrt{n})$	$O(\frac{d^2}{\Delta}\log^3 n)$	N/A	No
LinIMED (Bian and Tan, 2024)	$\tilde{O}(d\sqrt{n})$	Unknown	N/A	No
LinTS (Agrawal and Goyal, 2014)	$\tilde{O}(d\sqrt{dn})$	Unknown	No	No
RandUCB (Vaswani et al., 2020)	$\tilde{O}(d\sqrt{n})$	Unknown	No	No
SquareCB (Foster and Rakhlin, 2020)	$\tilde{O}(\sqrt{Kdn})$	Unknown	Yes	No
E2D (Foster et al., 2023)	$\tilde{O}(d\sqrt{n})$	Unknown	Yes	No*
SpannerIGW (Zhu et al., 2022)	$\tilde{O}(d\sqrt{n})$	$\Omega(\Delta\sqrt{n})$	Yes	No*
EXP2 (Bubeck and Cesa-Bianchi, 2012)	$O(\sqrt{dn\log K})$	$\Omega(\Delta\sqrt{n})$	Yes	Yes
LinMED (ours)	$\tilde{O}(d\sqrt{n})$	$O(\frac{d^2}{\Delta}\log^2 n)$	Yes	Yes

Table 1: Comparison of linear bandit algorithms. 'No*' means that the algorithm can be modified to assign a nonzero probability to every arm. The term "efficiently computable probability" refers to the efficiency in extra computation in addition to running the algorithm.

the following counterparts for the linear model:

$$\hat{\Delta}_{a,t} \to \max_{a' \in \mathcal{A}_t} \langle \hat{\theta}_{t-1}, a' - a \rangle \text{ and } N_{t-1,a} \to \frac{1}{\|a\|^2_{V_{t-1}^{-1}}}. \quad (2)$$

The second term in (2) is justified since the leverage score ($\|a\|^2_{V_{t-1}^{-1}}$) decreases with amount of exploration performed in the direction of a. This leads to an algorithm that we call LinMEDNOPT (**Lin**ear **M**inimum **E**mpirical **D**ivergence with **No OPT**imal design of experiment) with the sampling distribution given by

$$p_t^{\text{LinMEDNOPT}}(a) = \frac{f_t(a)}{\sum_{b \in \mathcal{A}_t} f_t(b)}, \quad (3)$$

where $f(t)$ is defined in (5). Our attempts to analyze the regret of this algorithm resulted in a polynomial dependence on K, which is undesirable since the strength of linear bandits is the ability to handle a large or even an infinite number of arms.

Indeed, one can find a problem where the regret scales with K as follows. Specifically, consider a 2-dimensional problem where the best arm and θ^* are both $(1,0) \in \mathbb{R}^2$. The rest of the $K-1$ arms are all $(0,1) \in \mathbb{R}^2$; i.e., all the sub-optimal arms share the same feature representation. In the beginning, after a few arm pulls, LinMEDNOPT could misjudge one of the sub-optimal arms as the best arm with a constant probability (imagine $\hat{\theta}$ being around $(-1,0)$). Then, it assigns the same constant probability for choosing one of the sub-optimal arms in the next time step. Since there are $K-1$ such sub-optimal arms, the total probability assigned to them will be high. Consequently, this significantly reduces the probability assigned to the true optimal arm (at most $1/K$), resulting in not exploring in the direction of the optimal arm. Since pulling an arm in the direction of the suboptimal arm $(0,1)$ provides zero information on the best arm $(1,0)$, it will be not until the algorithm pulls the optimal arm a few times that it can recover from this undesirable state. The waiting
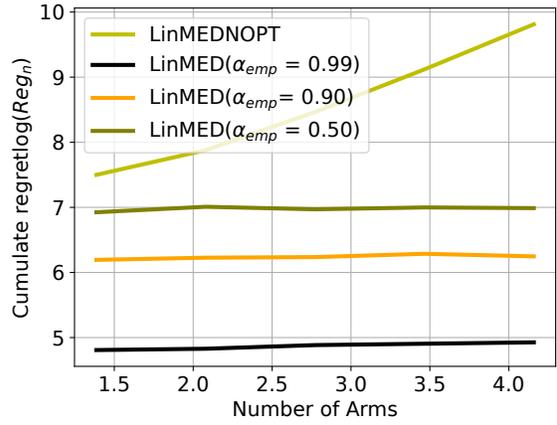
Figure 2: LinMED vs LinMEDNOPT, with $\sigma^2 = \sigma_*^2 = 3$ for $K \in \{4, 8, 16, 32, 64\}$, and $(\alpha_{\text{emp}}, \alpha_{\text{opt}}) \in \{(0.99, 0.005), (0.90, 0.05), (0.5, 0.25)\}$, $n = 20000$.

time for this is $\Omega(K)$ during which we suffer a linear regret. This indeed happens and leads to an order K regret numerically as can be seen in Figure 2.

Inspired by SpannerIGW (Zhu et al., 2022), we leverage the G-optimal design to avoid the dependence on K and propose an algorithm called LinMED in the next section. The key idea is that G-optimal design assigns probabilities to arms such that it will be informative for the linear model structure. Specifically, in the example above, G-optimal design will assign probabilities as if there are only two arms $(1,0)$ and $(0,1)$. This way, the probability will be assigned to these two arms almost equally at the beginning, ensuring that the waiting time to recover from the bad state discussed above is $\Theta(1)$ with respect to K rather than $\Theta(K)$. Our proposed algorithm LinMED will have a hyper-parameter $\alpha_{\text{opt}} \in (0,1)$ that controls how much we rely on the optimal design. Figure 2 shows that LinMED with various

choices of α_{opt} results in regret independent of K.

4 LINEAR MINIMUM EMPIRICAL DIVERGENCE (LINMED)

In this section, we describe our proposed algorithm Linear Minimum Empirical Divergence (LinMED; Algorithm 1). LinMED takes in guesses σ^2 and S on the unknown problem parameters σ_*^2 and $\|\theta^*\|$, but we do not require that these guesses are over-specified respectively, as we discussed in Section 2. At each time step t, the algorithm has maintained a ridge regression estimator $\hat{\theta}_{t-1}$ computed with a ridge parameter λ based on the samples collected up to time step $t-1$; see Algorithm 1 for their precise definitions. Let

$$\beta_t(\delta_t) := \left(\sigma \sqrt{\log\left(\frac{\det V_t}{\det V_0}\right) + 2\log\frac{1}{\delta_t}} + \sqrt{\lambda}S \right)^2 \quad (4)$$

where $V_t = \lambda I + \sum_{s=1}^{t} A_s A_s^T$.

LinMED first transforms the original arm set $\mathcal{A}_t$ into an augmented arm set $\overline{\mathcal{A}}_{(t)}$, see Algorithm 2. Although we present two different versions of LinMED, the version where the augmented arm set is generated by eliminating highly sub-optimal arms—while simpler to analyze—cannot be extended to cases where the true sub-Gaussian noise parameter is under-specified. Therefore, the main focus of this paper is the version 0, although detailed proofs for both version 0 and version 1 are provided in Appendix C. In version 0, the arms are rescaled as follows:

$$\overline{\mathcal{A}}_{(t)} = \{\sqrt{f_t(a)} \cdot a \mid a \in \mathcal{A}_t\}$$

where $f_t(a)$ is an exponential weight defined in (5).

In order to compute the arm sampling probability, we leverage the G-optimal design of experiments (Kiefer and Wolfowitz, 1960). Specifically, we assume that we have access to a computation oracle denoted by

$$\text{ApproxDesign}(\mathcal{B})$$

that takes in a set of vectors $\mathcal{B}$ and outputs a distribution over the set $\mathcal{B}$. We assume that ApproxDesign() satisfies the following two assumptions.

Assumption 2. (The design optimality) Given a set of vectors $\mathcal{B} \subset \mathbb{R}^d$, the oracle ApproxDesign($\mathcal{B}$) returns a C_{opt}-optimal design $q \in \Delta(\mathcal{B})$ for the set $\mathcal{B}$; i.e.

$$\|b\|_{V^{-1}(q)}^2 \leq C_{\text{opt}} d \log(d), \forall b \in \mathcal{B}$$

where $V(q) := \sum_{b \in \mathcal{B}} q_b b b^\top$ for $q \in \Delta(\mathcal{B})$. Furthermore, we assume that the support size of the design is small as follows:

Assumption 3. (Cardinality of design) Given a set of vectors $\mathcal{B} \subset \mathbb{R}^d$, the oracle ApproxDesign($\mathcal{B}$) returns a

---

Algorithm 1 LinMED

Input: regularization λ, failure rates $\{\delta_t\}_{t=0}^{\infty}$, optimal design fraction α_{opt}, empirical best fraction α_{emp}, ver $\in \{0, 1\}$, S (guess for $\|\theta^*\|_2$), and σ^2 (guess for σ_*^2)

1: Initialize $\hat{\theta}_0 = 0$, $V_0 = \lambda I$.
2: **for** $t = 1, 2, \ldots$ **do**
3:     Observe arm set $\mathcal{A}_t$.
4:     Estimate $\hat{a}_t = \max_{a' \in \mathcal{A}_t} \langle \hat{\theta}_{t-1}, a' \rangle$.
5:     Estimate $\hat{\Delta}_{a,t} := \langle \hat{\theta}_{t-1}, \hat{a}_t - a \rangle \quad \forall a \in \mathcal{A}_t$.
6:     Define $\forall a \in \mathcal{A}_t$

$$f_t(a) = \exp\left(-\frac{\hat{\Delta}_{a,t}^2}{\beta_{t-1}(\delta_{t-1})\|\hat{a}_t - a\|_{V_{t-1}^{-1}}^2} \right) \quad (5)$$

    where we take $\frac{0}{0} = 0$ and $\beta_t(\delta_t)$, defined in (4), is a function of S and σ^2.
7:     Compute a design:
    $q_t^{\text{opt}} = \text{ApproxDesignAugmented}(\mathcal{A}_t, f_t, \text{ver})$.

8:     Let $\forall a \in \mathcal{A}_t$

$$q_t(a) = \alpha_{\text{opt}} \cdot q_t^{\text{opt}}(a) + \alpha_{\text{emp}} \cdot \mathbb{1}\{a = \hat{a}_t\}$$
$$+ (1 - \alpha_{\text{opt}} - \alpha_{\text{emp}}) \cdot \frac{1}{|\mathcal{A}_t|}. \quad (6)$$

9:     Compute $p_t'(a)$:

$$p_t'(a) = \frac{q_t(a) f_t(a)}{\sum_{b \in \mathcal{A}_t} q_t(b) f_t(b)}. \quad (7)$$

10:     Define

$$\mathcal{B}_t = \{a \in \mathcal{A}_t : \|a\|_{V_{t-1}^{-1}}^2 > 1\}. \quad (8)$$

11:     **if** $|\mathcal{B}_t| > 0$ **then**
12:         $\forall a \in \mathcal{A}_t, \quad p_t(a) = \frac{1}{2}p_t'(a) + \frac{1}{2}\mathbb{1}\{a = B_t\}$
        where B_t is an arbitrarily chosen action $\in \mathcal{B}_t$.
13:     **else**
14:         $\forall a \in \mathcal{A}_t \quad p_t(a) = p_t'(a)$.
15:     **end if**
16:     Take action $A_t \sim p_t$.
17:     Observe the reward Y_t and update

$$V_t = V_{t-1} + A_t A_t^\top \text{ and } \quad \hat{\theta}_t = V_t^{-1} \sum_{s=1}^{t} A_s Y_s.$$

18: **end for**

---

design $q \in \Delta(\mathcal{B})$ for the set $\mathcal{B}$ such that

$$|\text{supp}(q)| = \tilde{\mathcal{O}}(d) .$$

Existence of such an oracle satisfying Assumptions 2 and 3 is guaranteed by Kiefer–Wolfowitz (Kiefer and Wolfowitz, 1960), and there are efficient algorithms for solving it (Todd, 2016). We present one such ApproxDesign() algorithm in Appendix F.

We compute $q_t^{\text{opt}} = \text{ApproxDesignAugmented}(\bar{\mathcal{A}}_{(t)})$. Subsequently, q_t is calculated as outlined in (6), wherein

Algorithm 2 ApproxDesignAugmented

Input: $\mathcal{A}_t$, f_t, ver $\in \{0, 1\}$

  if ver $= 0$ **then**

    Re-scale the arms:
$$\overline{\mathcal{A}}_{(t)} = \{\sqrt{f_t(a)} \cdot a \mid a \in \mathcal{A}_t\}$$

  else

    Eliminate highly sub-optimal arms:
$$\overline{\mathcal{A}}_{(t)} = \{a \in \mathcal{A}_t : f_t(a) \geq \frac{1}{e}\}$$

  end if

  Compute $q_t^{\text{opt}} = \text{ApproxDesign}(\overline{\mathcal{A}}_{(t)})$.

  return q_t^{opt}

a weight of α_{opt} is allocated to q_t^{opt}, α_{emp} is assigned to the empirical best arm, and the remaining weight is distributed among all the arms in $\mathcal{A}_t$. We then sample arm A_t according to the distribution $p_t^{'}$ defined in (7) whenever the set $\mathcal{B}_t$ defined in (8) is empty, otherwise we delegate one half of the probability to an arbitrarily chosen arm from $\mathcal{B}_t$. Finally, we observe the reward and update the estimator $\hat{\theta}_t$ for the next round.

5 MAIN RESULTS

We now provide regret guarantees of LinMED. For the instance-dependent regret bound, we will use the following assumption.

Assumption 4. (Lower bound for sub-optimality gap) There exists a constant $\Delta > 0$ such that
$$\Delta \leq \min_{t \in [n], a \in \mathcal{A}_t : \Delta_{a,t} > 0} \Delta_{a,t}, \quad \text{almost surely.} \quad (9)$$

Furthermore, we define the true confidence radius
$$\beta_t^*(\delta_t) := \left(\sigma_* \sqrt{\log\left(\frac{\det V_t}{\det V_0}\right) + 2 \log \frac{1}{\delta_t}} + \sqrt{\lambda} S_* \right)^2 \quad (10)$$

where $S_* := \|\theta^*\|_2$. We define
$$H_{\max} := \max_{t \in [n]} \exp\left(\frac{\beta_{t-1}^*(\delta_{t-1})}{\beta_{t-1}(\delta_{t-1})}\right).$$

We first state two generic theorems guaranteeing the regret bound of LinMED for any input λ, σ^2, and S, followed by a more concise results with a particular choices of λ under the over-specification and under-specification (of σ_*^2 and S_*) cases respectively.

Furthermore, for all upcoming instance-dependent results, we ignore all logarithmic factors except those related to n and omit terms that do not involve polylog(n) or $\frac{1}{\Delta}$. Similarly, for all upcoming minimax results, we ignore logarithmic factors except those related to n and omit terms that do not involve poly(n).

Theorem 1 (Instance-dependent bound). *Under Assumptions 1, 2, and 3, with $\delta_t = \frac{1}{t+1}$, LinMED satisfies, $\forall n \geq 1$,*
$$\mathbb{E} \operatorname{Reg}_n =$$
$$O\Bigg(\frac{1}{\Delta} d \log(n) \bigg(\Big(\sigma^2 d \log(n) + \lambda S^2 \Big) \log(\log n) +$$
$$\Big(\sigma_*^2 d \log(n) + \lambda S_*^2 \Big) H_{\max} \bigg) \Bigg)$$

Theorem 2 (Minimax bound). *Under Assumptions 1, 2, and 3, with $\delta_t = \frac{1}{t+1}$, LinMED satisfies, $\forall n \geq 1$,*
$$\mathbb{E} \operatorname{Reg}_n = O\Bigg(\sqrt{n} \bigg(\log^{\frac{1}{2}}(n) \Big(d\sigma \log(n) + \frac{\lambda S^2}{\sigma} \Big) + \frac{H_{\max}}{\sigma \log^{\frac{3}{2}}(n)} \Big(d\sigma_*^2 \log(n) + \lambda S_*^2 \Big) \bigg) \Bigg).$$

It is important to emphasize that in general the learner does not have access to the true sub-Gaussian parameter (σ_*^2) of the noise and S_*. The input sub-Gaussian parameter (σ^2) and S may either over-specified or under-specified with respect to their true values. Nevertheless, our algorithm provides a regret bound that remains valid across all such scenarios , a feature absent in the analysis of most of the state-of-the art algorithms such as OFUL, LinTS, and LinIMED. This constitutes one of the novel contributions of our analysis. It is noteworthy that, at first glance, one might be misled into believing that selecting smaller values for S and σ results in a smaller regret bound. However, this is not the case, as $H_{\max}$ increases exponentially as S and σ decrease.

Consider the case where the true sub-Gaussian parameter (σ_*^2) and S_* are over-specified, $H_{\max}$ tends to be less than $\exp(1)$, leading to the following corollaries:

Corollary 3 (Instance-dependent bound). Under Assumptions 1, 2, and 3, assuming $\sigma^2 \geq \sigma_*^2$, $S \geq S_*$ with $\lambda = \frac{\sigma^2}{S^2}$ and $\delta_t = \frac{1}{t+1}$, LinMED satisfies, $\forall n \geq 1$,
$$\mathbb{E} \operatorname{Reg}_n = O\left(\sigma^2 \frac{d^2}{\Delta} \log^2(n) \log(\log n) \right).$$

Corollary 4 (Minimax bound). Under Assumptions 1, 2, and 3, assuming $\sigma^2 \geq \sigma_*^2$, $S \geq S_*$ and with $\lambda = \frac{\sigma^2}{S^2}$ and $\delta_t = \frac{1}{t+1}$, LinMED satisfies, $\forall n \geq 1$,
$$\mathbb{E} \operatorname{Reg}_n = O\left(\sigma d\sqrt{n} \log^{\frac{3}{2}}(n) \right).$$

Instance-dependent bound of LinMED showcases a $\log(n)$ improvement over the instance-dependent bound of OFUL (Abbasi-Yadkori et al., 2011) and LinMED guarantees an optimal minimax bound up to logarithmic factors.

Next, we consider the case where the σ_*^2 is under-

specified and S_* is over-specified.

Corollary 5 (Minimax bound). Under Assumptions 1, 2, and 3, assuming $\sigma^2 < \sigma_*^2$, $S \geq S_*$ and with $\lambda = \frac{\sigma^2}{S^2}$ and $\delta_t = \frac{1}{t+1}$, $\forall n \geq 1$, LinMED satisfies

$$\mathbb{E}\operatorname{Reg}_n = O\left(\frac{\sigma d\sqrt{n}}{\log^{\frac{1}{2}}(n)}\left(\log^2(n) + \frac{\sigma_*^2}{\sigma^2}\exp\left(\frac{\sigma_*^2}{\sigma^2}\right)\right)\right)$$

One can derive the instance-dependent bound and bounds for under-specified S_* in a similar fashion. Proofs of the theorems and corollaries are deferred to the appendix.

The key steps of the proof of Theorem 1 and 2. Conceptually, our proof structure for Lemma 1 closely follows the framework of the Maillard sampling proof by Bian and Jun (2022). We define the following events: $\mathcal{U}_{t-1,\ell}(A_t) = \left\{\|A_t\|_{V_{t-1}^{-1}}^2 \geq \varepsilon_\ell\right\}$, $\mathcal{V}_{t-1}(A_t) = \left\{\hat{\Delta}_{A_t,t} \geq \frac{\Delta_{A_t,t}}{1+c}\right\}$, $\mathcal{W}_{t-1,\ell} = \left\{\max_{a' \in \mathcal{A}_t}\langle\hat{\theta}_{t-1}, a'\rangle \geq \langle\theta^*, a_t^*\rangle - \varepsilon_{2,\ell}\right\}$, where $\ell, \varepsilon_\ell, \varepsilon_{2,\ell}$ are parameters to be tuned. At an abstract level, the regret can be decomposed as follows:

$$\operatorname{Reg}_n = \mathbb{E}\left[\sum_{t=1}^n \Delta_{A_t,t}\right] = \mathbb{E}\left[\sum_{t=1}^n \Delta_{A_t,t}\mathbb{1}\left\{\mathcal{U}_{t-1,\ell}(A_t)\right\}\right]$$
$$\text{(Term 1)}$$

$$+ \mathbb{E}\left[\sum_{t=1}^n \Delta_{A_t,t}\mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(A_t)\right\}\mathbb{1}\left\{\mathcal{V}_{t-1}(A_t)\right\}\right]$$
$$\text{(Term 2)}$$

$$+ \mathbb{E}\left[\sum_{t=1}^n \Delta_{A_t,t}\mathbb{1}\left\{\overline{\mathcal{V}}_{t-1}(A_t)\right\}\mathbb{1}\left[\mathcal{W}_{t-1,\ell}\right]\right]$$
$$\text{(Term 3)}$$

$$+ \mathbb{E}\left[\sum_{t=1}^n \Delta_{A_t,t}\mathbb{1}\left[\overline{\mathcal{W}}_{t-1,\ell}\right]\right].$$
$$\text{(Term 4)}$$

We bound Term 1 and Term 3 using the elliptical potential count (EPC), as shown in Lemma 11. Term 2 is bounded by noting that the probability of selecting suboptimal arms is small when the events $\mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(A_t)\right\}$ and $\mathbb{1}\left\{\mathcal{V}_{t-1}(A_t)\right\}$ occur. Term 4 is the most challenging one where drawing an analogue from Maillard sampling's proof is nontrivial. Detailed proof is presented in Appendix C.

6 INSTANCE-DEPENDENT LOWER BOUNDS FOR SPANNERIGW AND EXP2

In this section, we analyze the instance-dependent regrets for EXP2 and SpannerIGW. We show that there are instances for which the above two algorithms have

(a)

(b)
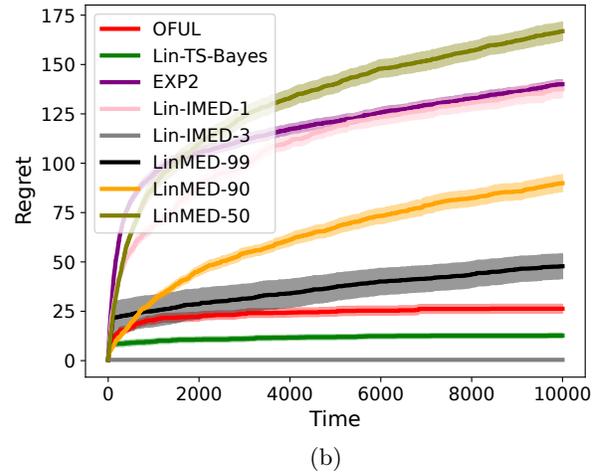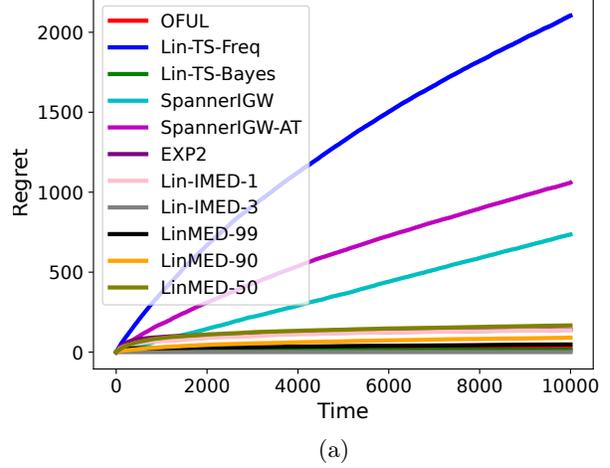
Figure 3: Large gap instance experiments

an instance-dependent bound of $\Omega(\Delta\sqrt{n})$. Hence, Lin-MED stands out as a leading randomized algorithm with closed-form arm sampling probabilities, achieving a logarithmic instance-dependent regret bound.

Theorem 6. *There exists a linear bandit problem for which the EXP2 algorithm satisfies*

$$\mathbb{E}\operatorname{Reg}_n \geq \Omega(\Delta\sqrt{n}).$$

Theorem 7. *There exists a linear bandit problem for which the SpannerIGW algorithm satisfies*

$$\mathbb{E}\operatorname{Reg}_n \geq \Omega(\Delta\sqrt{n}).$$

The proofs are deferred to the Appendix E.

7 EMPIRICAL STUDIES

This section is dedicated to demonstrating the effectiveness of our algorithm in comparison to several well-known algorithms across various scenarios, each of which evaluates different aspects of algorithmic perfor-

(a) $\varepsilon = 0.005,\ \sigma^2 = \sigma_*^2$

(b) $\varepsilon = 0.005,\ \sigma^2 = 2 \cdot \sigma_*^2$

(c) $\varepsilon = 0.01,\ \sigma^2 = \sigma_*^2$

(d) $\varepsilon = 0.01,\ \sigma^2 = 2 \cdot \sigma_*^2$
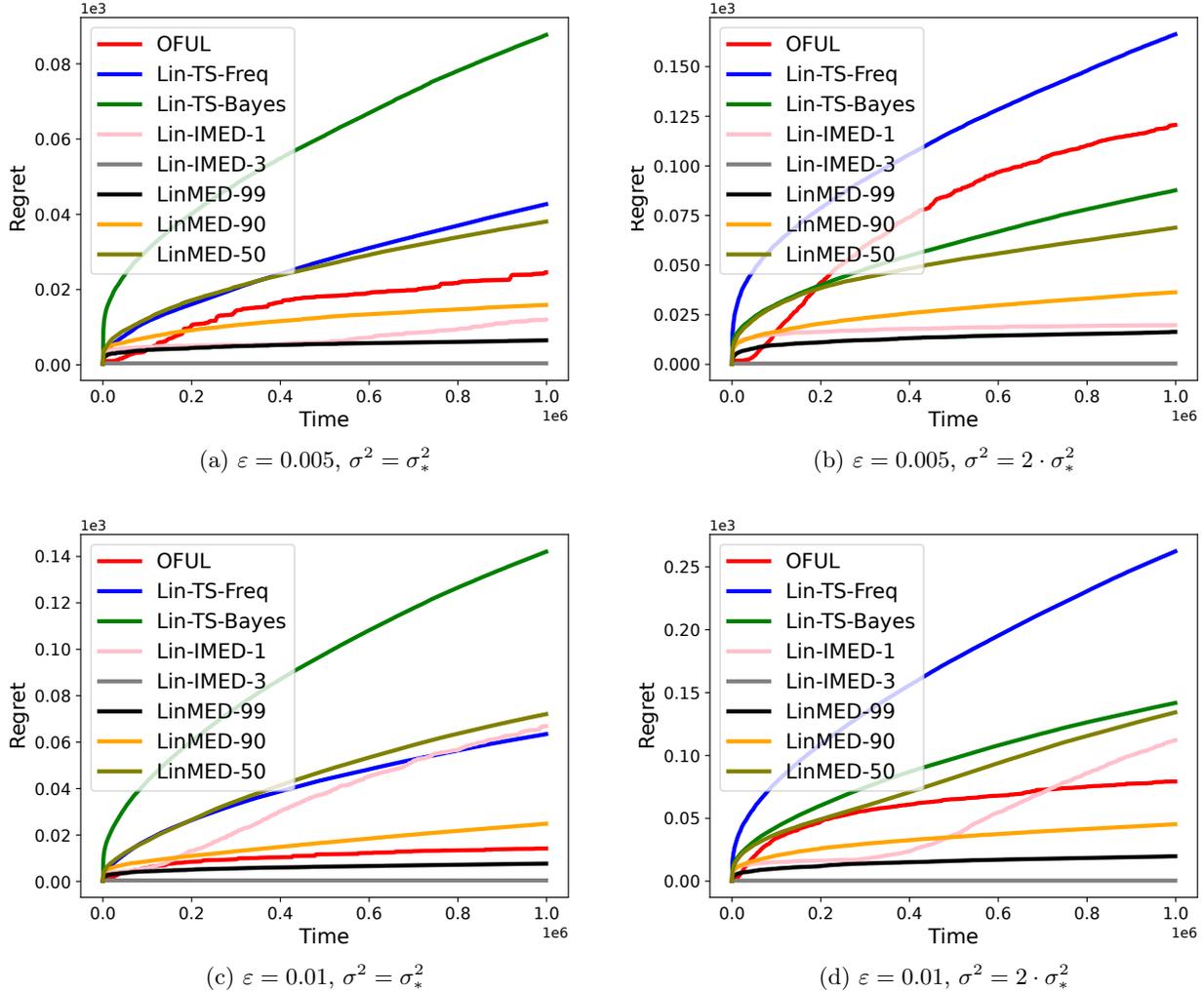
Figure 4: End of optimism experiments

mance. Throughout our empirical studies, we fine-tune $(\alpha_{\mathrm{emp}}, \alpha_{\mathrm{opt}})$ for LinMED algorithm using the following values: $(0.99, 0.005), (0.90, 0.05)$, and $(0.5, 0.25)$. We refer to the resulting variants as LinMED-99, LinMED-90, LinMED-50 respectively.

SpannerIGW utilizes exploration parameters γ and η, which are dependent on the time horizon n and remain fixed throughout all rounds. We modify these parameters to use t in place of n at each time step t, thereby deriving an anytime version of the algorithm, which we refer to as SpannerIGW-Anytime or SpannerIGW-AT. LinIMED has three variants, namely LinIMED-1, LinIMED-2, and LinIMED-3. In our study, we use LinIMED-1 and LinIMED-3 only, as prior experiments in Bian and Tan (2024) show that LinIMED-2 consistently performs between these two algorithms. Therefore, evaluating LinIMED-1 and LinIMED-3 sufficiently captures both ends of the performance spectrum. Additionally, LinIMED-3 has a parameter C, which we

set to 30, following Bian and Tan (2024). We also use a modified EXP2 algorithm (based on rewards instead of losses), presented in Algorithm 3, and refer to it simply as EXP2 for clarity.

Furthermore, it is noteworthy that, throughout our experiments, we select either the frequentist (Lin-TS-Freq) (Agrawal and Goyal, 2014) or Bayesian version (Lin-TS-Bayes) (Russo and Roy, 2014) of LinTS, or both. However, whenever we choose only one version, it implies that the selected version significantly outperforms the omitted one. We use LinTS to refer to both Lin-TS-Freq and Lin-TS-Bayes.

Large gap instance. Our algorithm achieves an instance-dependent regret bound of $O(\log^2 n)$ with respect to n omitting $\log\big(\log(n)\big)$ terms. This is much better than EXP2 and SpannerIGW, both of which have an instance-dependent lower bound in the order of $\Omega(\sqrt{n})$ with respect to n. Our instance-dependent

regret bound also shows a $\log(n)$ factor improvement over the original analysis of OFUL (Abbasi-Yadkori et al., 2011). The instance-dependent regret bounds for LinTS, LinIMED-1, and LinIMED-3 are not known to our knowledge.

The experimental setup of this scenario is as follows: $\mathcal{A} = \{(1,0), (0,1)\}$ and $\theta^* = (1,0)$. The noise follows a normal distribution $\mathcal{N}(0, \sigma^2_*)$ with $\sigma^2 = \sigma^2_* = 1$. The time horizon for each trial is $n = 10,000$ and conduct 10 such independent trials. We compare our algorithm against SpannerIGW (Zhu et al., 2022), SpannerIGW-Anytime, LinIMED-1, LinIMED-3 (Bian and Tan, 2024), OFUL (Abbasi-Yadkori et al., 2011), Lin-TS-Bayes (Bayesian version)(Russo and Roy, 2014), Lin-TS-Freq (Frequentest version) (Agrawal and Goyal, 2014), and EXP2 (Bubeck and Cesa-Bianchi, 2012).

Our simulations indicate that our algorithm outperforms SpannerIGW, SpannerIGW-Anytime, Lin-TS-Freq, EXP2, and LinIMED-1. Furthermore, our algorithm demonstrates performance that is sufficiently close to that of OFUL (Abbasi-Yadkori et al., 2011), LinIMED-3, and Lin-TS-Bayes. Figure 3a presents the primary plot of our results, while 3b displays the same data, with SpannerIGW, SpannerIGW-Anytime, and Lin-TS-Freq removed for a more precise comparison of the remaining algorithms. Furthermore, close visual inspection confirms the instance-dependent regret lower bound of $\Omega(\sqrt{n})$ that we proved in Section 6 for both EXP2 and SpannerIGW.

End of Optimism instance The "end of optimism instance" Lattimore and Szepesvári (2017) is often cited as a pitfall for optimism-based algorithms such as OFUL. Inspired by the end of optimism-based simulations conducted by Bian and Tan (2024), we perform similar experiments to evaluate the performance of our algorithm in comparison to OFUL, LinIMED-1, LinIMED-3, and LinTS. In this context, OFUL and LinTS are classified as optimistic algorithms, while LinIMED-1 and LinIMED-3 are minimum empirical divergence-based deterministic algorithms. We set the number of arms $K = 3$ and dimension $d = 2$ and $\mathcal{A} = \{a_1 = (1,0), a_2 = (0,1), a_3 = (1 - \varepsilon, 2\varepsilon)\}$ where $\varepsilon \in \{0.005, 0.01, 0.02\}$ and $\theta^* = (1,0)$. The noise follows $\mathcal{N}(0, \sigma^2)$ with $\sigma_* = 0.1$. The time horizon for each trial is $n = 1000,000$ and conduct 20 such independent trials. Furthermore, we conduct experiments for the cases where $\sigma^2 = \sigma^2_*$ and $\sigma^2 = 2 \cdot \sigma^2_*$.

Optimism-based algorithms typically identify the optimal arm (a_1) and the near-optimal arm (a_3) as the optimistic choices, frequently pulling these two arms. This behavior limits their ability to pull the highly suboptimal arm (a_2), which provides a crucial piece of information for distinguishing between the optimal and near-

optimal arms. As a result, optimistic algorithms struggle to differentiate effectively between these two arms, often incurring a small regret from repeatedly selecting the near-optimal arm (a_3) for an extended period.

In contrast, algorithms that do not follow optimistic principles explore adequately in the direction of the highly sub-optimal arm as well. Consequently, the trend of our algorithm reveals that it initially incurs significant regret by choosing the highly sub-optimal arm but ultimately converges on the optimal arm as a consistent choice.

From Figure 4, it is evident that LinIMED-3 and Lin-MED ($\alpha_{\text{emp}} = 0.99$) perform very well and significantly out performs LinTS whereas LinMED ($\alpha_{\text{emp}} = 0.90$) and LinMED ($\alpha_{\text{emp}} = 0.50$) exhibit comparable performance. However, LinIMED-1 and OFUL start with good performance, but their effectiveness deteriorates over time, especially when ε is too small. This effect is amplified when $\sigma^2 = 2 \cdot \sigma^2_*$. When the noise is overspecified, the performance of Lin-TS-Freq deteriorates significantly due to oversampling. We present detailed results in Appendix G.1.

8 CONCLUSION

Our proposed algorithm LinMED posesses many intriguing properties and shows excellent empirical performance, which opens up exciting avenues for future research. First, it would be interesting to explore ways to generalize the noise model to exponential family (generalized linear models) or generalize the linear class to generic hypothesis class. Identifying fundamental limits of adapting to the unknown sub-Gaussian noise level would be interesting and important. Second, we believe the challenge of coping with the unknown noise level is an important problem that has received less attention in the literature. Relatedly, Jun and Kim (2024) have shown that adapting to the unknown sub-Gaussian parameter σ^2_* is possible when it is overspecified. Finally, it would be intriguing to develop pure exploration or Bayesian optimization version of LinMED and explore the potential of the MED principle.

Acknowledgements

References

Abbasi-Yadkori, Y., Pal, D., and Szepesvari, C. Improved Algorithms for Linear Stochastic Bandits. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 1–19, 2011.

Abeille, M. and Lazaric, A. Linear Thompson Sampling Revisited. In *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 54, pages 176–184, 2017.

Agrawal, S. and Goyal, N. Thompson sampling for contextual bandits with linear payoffs, 2014.

Agrawal, S. and Goyal, N. Near-Optimal Regret Bounds for Thompson Sampling. *Journal of the ACM*, 64(5):1–24, 2017.

Auer, P., Cesa-Bianchi, N., and Fischer, P. Finite-time Analysis of the Multiarmed Bandit Problem. *Machine Learning*, 47(2–3):235–256, 2002.

Betke, U. and Henk, M. Approximating the volume of convex bodies. *Discrete & Computational Geometry*, 10(1):15–21, 1993.

Bian, J. and Jun, K.-S. Maillard Sampling: Boltzmann Exploration Done Optimally. *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2022.

Bian, J. and Tan, V. Y. Indexed minimum empirical divergence-based algorithms for linear bandits. *Transactions on Machine Learning Research (TMLR)*, 2024.

Bubeck, S. and Cesa-Bianchi, and Kakade, S. M. Towards minimax policies for online linear optimization with bandit feedback. pages 41–42. Microtome, 2012.

Dani, V., Hayes, T. P., and Kakade, S. M. Stochastic Linear Optimization under Bandit Feedback. In *Proceedings of the Conference on Learning Theory (COLT)*, pages 355–366, 2008.

Foster, D. J. and Rakhlin, A. Beyond UCB: Optimal and efficient contextual bandits with regression oracles. *Proceedings of the International Conference on Machine Learning (ICML)*, 2020.

Foster, D. J., Golowich, N., and Han, Y. Tight guarantees for interactive decision making with the decision-estimation coefficient, 2023.

Gales, S. B., Sethuraman, S., and Jun, K.-S. Norm-Agnostic Linear Bandits. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2022.

Gangan, E., Kudus, M., and Ilyushin, E. Survey of multi-armed bandit algorithms applied to recommendation systems. *International Journal of Open Information Technologies*, 9, 2021.

Honda, J. and Takemura, A. An asymptotically optimal policy for finite support models in the multiarmed bandit problem. *Machine Learning*, 85(3):361–391, 2011.

Honda, J. and Takemura, A. Non-asymptotic analysis of a new bandit algorithm for semi-bounded rewards. *J. Mach. Learn. Res.*, 16:3721–3756, 2015.

Horvitz, D. G. and Thompson, D. J. A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association*, 47 (260):663–685, 1952.

Jun, K.-S. and Kim, J. Noise-adaptive confidence sets for linear bandits and application to bayesian optimization. *Proceedings of the International Conference on Machine Learning (ICML)*, 2024.

Kiefer, J. and Wolfowitz, J. The equivalence of two extremum problems. *Canadian Journal of Mathematics*, 12:363–366, 1960.

Lattimore, T. and Szepesvári, C. The end of optimism? An asymptotic analysis of finite-armed linear bandits. In *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2017.

Li, L., Chu, W., Langford, J., and Schapire, R. E. A Contextual-Bandit Approach to Personalized News Article Recommendation. *Proceedings of the International Conference on World Wide Web*, pages 661–670, 2010.

Maillard, O.-A. *APPRENTISSAGE SÉQUENTIEL: Bandits, Statistique et Renforcement.* PhD thesis, Université des Sciences et Technologie de Lille-Lille I, 2013.

Mary, J., Gaudel, R., and Preux, P. Bandits and recommender systems. In Pardalos, P., Pavone, M., Farinella, G. M., and Cutello, V., editors, *Machine Learning, Optimization, and Big Data*, pages 325–336, Cham, 2015. Springer International Publishing.

Precup, D. Eligibility traces for off-policy policy evaluation. *Proceedings of The International Conference on Machine Learning (ICML)*, pages 759–766, 2000.

Russo, D. and Roy, B. V. Learning to Optimize via Posterior Sampling. *Mathematics of Operations Research*, 39(4):1221–1243, 2014.

Thompson, W. R. On the Likelihood that One Unknown Probability Exceeds Another in View of the Evidence of Two Samples. *Biometrika*, 25(3/4):285, 1933.

Todd, M. J. *Minimum-volume ellipsoids: Theory and algorithms.* SIAM, 2016.

Vaswani, S., Mehrabian, A., Durand, A., and Kveton, B. Old dog learns new tricks: Randomized ucb for bandit problems. *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2020.

Zhu, Y., Foster, D. J., Langford, J., and Mineiro, P. Contextual bandits with large action spaces: Made practical. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 27428–27453, 2022.

Checklist

1. For all models and algorithms presented, check if you include:

   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes]

   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes]

   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [No]

2. For any theoretical claim, check if you include:

   (a) Statements of the full set of assumptions of all theoretical results. [Yes]

   (b) Complete proofs of all theoretical results. [Yes]

   (c) Clear explanations of any assumptions. [Yes]

3. For all figures and tables that present empirical results, check if you include:

   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [No]

   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Not Applicable]

   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Not Applicable]

   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Not Applicable]

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:

   (a) Citations of the creator If your work uses existing assets. [Yes]

   (b) The license information of the assets, if applicable. [Not Applicable]

   (c) New assets either in the supplemental material or as a URL, if applicable. [Not Applicable]

   (d) Information about consent from data providers/curators. [Not Applicable]

   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:

   (a) The full text of instructions given to participants and screenshots. [Not Applicable]

   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]

   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

Supplementary Materials

Appendix

Table of Contents

A NOTATIONS

Let us define the relevant quantities: While some of these have already been introduced in the main body of the paper, we redefine them here for ease of reference.

- $a_t^* := \arg\max_{a \in \mathcal{A}_t} \langle \theta^*, a \rangle$

- $\varepsilon_\ell := 2^{-2\ell} \cdot \varepsilon$ where ε is a parameter to be determined later

- $\varepsilon_{2,\ell} := 2^{-\ell} \cdot \varepsilon_2$ where ε_2 is a parameter to be determined later

- $\Delta_{a,t} := \langle \theta^*, a_t^* \rangle - \langle \theta^*, a \rangle$

- $\Delta_t := \Delta_{A_t,t}$

- $\Delta := \min_{t \in [n], a \in \mathcal{A}_t : \Delta_{a,t} > 0} \Delta_{a,t}$

- $\overline{\Delta}_{a,t} := B \wedge \Delta_{a,t}$

- $\overline{\Delta}_t := B \wedge \Delta_{A_t,t}$

- $\sqrt{\beta_{t-1}(\delta_{t-1})} := \sigma \sqrt{2 \log\left(\frac{\det(V_{t-1})^{\frac{1}{2}} \det(\lambda I)^{-\frac{1}{2}}}{\delta_{t-1}} \right)} + \sqrt{\lambda} S$

- $\sqrt{\beta_{t-1}^*(\delta_{t-1})} := \sigma_* \sqrt{2 \log\left(\frac{\det(V_{t-1})^{\frac{1}{2}} \det(\lambda I)^{-\frac{1}{2}}}{\delta_{t-1}} \right)} + \sqrt{\lambda} S_*$

- $H_{\max} := \max_{t \in [n]} \exp\left(\frac{\beta_{t-1}^*(\delta_{t-1})}{\beta_{t-1}(\delta_{t-1})} \right)$

- $V(p_t) := \sum_{a \in \mathcal{A}_t} p_t(a) a a^\top$

- $\overline{V}(p_t) := \sum_{a \in \overline{\mathcal{A}}_{(t)}} p_t(a)(\bar{a}_{(t)})(\bar{a}_{(t)})^\top$ where $\bar{a}_{(t)} = \sqrt{f_t(a)} \cdot a$

Let $\mathcal{F}_t$ be the σ-algebra generated by $(A_1, Y_1, A_2, Y_2, \ldots\ldots A_t, Y_t)$. We define the following shortcuts:

- $\mathbb{P}_t(\mathcal{E}) := \mathbb{P}(\mathcal{E} \mid \mathcal{F}_t)$

- $\mathbb{E}_t[\mathcal{E}] := \mathbb{E}[\mathcal{E} \mid \mathcal{F}_t]$

- $\mathbb{1}_{\hat{a}_t}(a) := \mathbb{1}\{a = \hat{a}_t\}$

B ASSUMPTIONS

We would like to remind you of the Assumptions 1, 2, 3

C PROOFS

C.1 Good event

The following "good event" is derived from the work of Abbasi-Yadkori et al. (2011) and occurs with a probability of at least $1 - \delta_{t-1}$, as established in Lemma 13.

$$\mathcal{G}_1 = \left\{ \|\theta^* - \hat{\theta}_{t-1}\|_{V_{t-1}}^2 \leq \beta_{t-1}^*(\delta_{t-1}), \quad \forall t \geq 1 \right\}. \tag{11}$$

C.2 Conditioning events

In the main body of the paper, we presented a high-level proof sketch. Here, we provide a more detailed analysis by breaking down and evaluating the regret case by case, according to the following events:

$$\mathcal{U}_{t-1,\ell}(a) = \left\{ \|a\|^2_{V_{t-1}^{-1}} \geq \varepsilon_\ell \right\}$$

$$\mathcal{U}_{t-1}(a) = \left\{ \|a\|^2_{V_{t-1}^{-1}} \geq 1 \right\}$$

$$\mathcal{E}_t = \left\{ |\mathcal{B}_t| > 0 \right\}$$

$$\mathcal{V}_{t-1}(a) = \left\{ \hat{\Delta}_{a,t} \geq \frac{\Delta_{a,t}}{1+c} \right\}$$

$$\mathcal{W}_{t-1,\ell} = \left\{ \max_{a' \in \mathcal{A}_t} \langle \hat{\theta}_{t-1}, a' \rangle \geq \langle \theta^*, a_t^* \rangle - \varepsilon_{2,\ell} \right\}$$

$$\mathcal{D}_{t,\ell}(a) = \left\{ B \cdot 2^{-\ell} < \overline{\Delta}_{a,t} \leq B \cdot 2^{-\ell+1} \right\} \qquad (\overline{\Delta}_{a,t} := B \wedge \Delta_{a,t})$$

$$\overline{\mathcal{D}}_{t,L}(a) = \left\{ \overline{\Delta}_{a,t} \leq B \cdot 2^{-L} \right\}.$$

C.3 Proof of Lemma 1

In this section, we present the fundamental lemma underlying our regret analysis, which ultimately leads to the theorems and corollaries concerning both the instance-dependent and minimax bounds.

Lemma 1 (Regret Bound). *Under Assumptions 1, 2, and 3, with $\delta_t = \frac{1}{t+1}$, LinMED satisfies, $\forall n \geq 1$,*

$$\mathrm{Reg}_n \leq 6dB \left(3 + \frac{1}{\alpha_{\mathrm{emp}}^2} \right) \log\left(1 + \frac{2}{\lambda} \right) + 2B \log(n+1) + \frac{192\beta_n(\delta)\log(n)d}{2^{-L}B} \left(1 + \frac{1}{\alpha_{\mathrm{emp}}^2} \right) \log\left(1 + \frac{32\beta_n(\delta)\log(n)}{\lambda 2^{-2L}B^2} \right)$$

$$+ \frac{192\beta_n^*(\delta)d}{B \cdot 2^{-L}} \log\left(1 + \frac{32\beta_n^*(\delta)}{\lambda B^2 \cdot 2^{-2L}} \right) + \frac{512 H_{\max} \cdot C_{\mathrm{opt}} \cdot d\log(d)}{\alpha_{\mathrm{opt}}B \cdot 2^{-L}} \left(\frac{\lambda(S_*)^2}{2} + \sigma_*^2 d\log\left(1 + \frac{n}{d\lambda} \right) \right)$$

$$+ n \cdot B \cdot 2^{-L} \cdot \mathbb{1}\left\{ B \cdot 2^{-L} > \Delta \right\} + \frac{4B}{\alpha_{\mathrm{emp}}}.$$

Proof. First, we decompose the proof based on the occurrence of the event $\mathcal{E}_t$. This decomposition is crucial because, if the event occurs, we select an arm arbitrarily from the set $\mathcal{B}_t$ with a probability of at-least one-half, as described in Algorithm 1.

$$\mathrm{Reg}_n = \mathbb{E}\left[\sum_{t=1}^n \langle \theta^*, a_t^* \rangle - \langle \theta^*, A_t \rangle \right]$$

$$= \mathbb{E}\left[\sum_{t=1}^n \Delta_t \langle \theta^*, a_t^* \rangle - \langle \theta^*, A_t \rangle \right] \qquad (\Delta_t := \Delta_{A_t,t})$$

$$\leq \mathbb{E}\left[\sum_{t=1}^n (B \wedge \Delta_t) \langle \theta^*, a_t^* \rangle - \langle \theta^*, A_t \rangle \right]$$

$$= \mathbb{E}\left[\sum_{t=1}^n \overline{\Delta}_t \mathbb{1}\{A_t \neq a_t^*\} \right] \qquad (\overline{\Delta}_t := B \wedge \Delta_{A_t,t})$$

$$= \mathbb{E}\left[\sum_{t=1}^n \overline{\Delta}_t \mathbb{1}\{A_t \neq a_t^*\} \left(\mathbb{1}\left\{ \overline{\mathcal{E}}_t \right\} + \mathbb{1}\{\mathcal{E}_t\} \right) \right]$$

$$\leq \mathbb{E}\left[\sum_{t=1}^{n}\overline{\Delta}_t\,\mathbb{1}\left\{A_t \neq a_t^*\right\}\mathbb{1}\left\{\overline{\mathcal{E}}_t\right\}\right] + \mathbb{E}\left[\sum_{t=1}^{n}\overline{\Delta}_t\,\mathbb{1}\left\{A_t \neq a_t^*\right\}\mathbb{1}\left\{\mathcal{E}_t\right\}\right].$$

$$\underbrace{\phantom{\mathbb{E}\left[\sum_{t=1}^{n}\overline{\Delta}_t\,\mathbb{1}\left\{A_t \neq a_t^*\right\}\mathbb{1}\left\{\overline{\mathcal{E}}_t\right\}\right]}}_{A_1} \qquad \underbrace{\phantom{\mathbb{E}\left[\sum_{t=1}^{n}\overline{\Delta}_t\,\mathbb{1}\left\{A_t \neq a_t^*\right\}\mathbb{1}\left\{\mathcal{E}_t\right\}\right]}}_{A_2}$$

Consequently, if the selected arm A_t is in $\mathcal{B}_t$, it directly implies that $\|A_t\|_{V_{t-1}^{-1}}^2 > 1$, in accordance with the definition of the set $\mathcal{B}_t$. Moreover, the expected number of occurrences of the event $\|A_t\|_{V_{t-1}^{-1}}^2 > 1$ can be managed using the Elliptical Potential Count (EPC), as demonstrated in Lemma 11.

$$A_2 = \mathbb{E}\left[\sum_{t=1}^{n}\overline{\Delta}_t\,\mathbb{1}\left\{A_t \neq a_t^*\right\}\mathbb{1}\left\{\mathcal{E}_t\right\}\right]$$

$$= 2\,\mathbb{E}\left[\sum_{t=1}^{n}\overline{\Delta}_t\,\mathbb{1}\left\{A_t \neq a_t^*\right\}\mathbb{1}\left\{\mathcal{E}_t\right\}\frac{1}{2}\right]$$

$$\leq 2\,\mathbb{E}\left[\sum_{t=1}^{n}\overline{\Delta}_t\,\mathbb{1}\left\{A_t \neq a_t^*\right\}\mathbb{1}\left\{\mathcal{E}_t\right\}\mathbb{P}\left(A_t \in \mathcal{B}_t\right)\right]$$

$$= 2\,\mathbb{E}\left[\sum_{t=1}^{n}\overline{\Delta}_t\,\mathbb{1}\left\{A_t \neq a_t^*\right\}\mathbb{1}\left\{\mathcal{E}_t\right\}\mathbb{P}\left(\|A_t\|_{V_{t-1}^{-1}}^2 > 1\right)\right]$$

$$= 2\,\mathbb{E}\left[\sum_{t=1}^{n}\overline{\Delta}_t\,\mathbb{1}\left\{A_t \neq a_t^*\right\}\mathbb{1}\left\{\mathcal{E}_t\right\}\mathbb{E}_{t-1}\left[\mathbb{1}\left\{\|A_t\|_{V_{t-1}^{-1}}^2 > 1\right\}\right]\right]$$

$$\leq 2\,\mathbb{E}\left[\sum_{t=1}^{n}\overline{\Delta}_t\,\mathbb{1}\left\{A_t \neq a_t^*\right\}\mathbb{1}\left\{\|A_t\|_{V_{t-1}^{-1}}^2 > 1\right\}\right]$$

$$\leq 2B\cdot\mathbb{E}\left[\sum_{t=1}^{n}\mathbb{1}\left\{\|A_t\|_{V_{t-1}^{-1}}^2 > 1\right\}\right]$$

$$\leq 2B\cdot 3d\log\left(1+\frac{2}{\lambda}\right) \qquad\qquad \text{(by lemma 11)}$$

$$= 6Bd\log\left(1+\frac{2}{\lambda}\right).$$

This concludes the bounding of the term A_2.

Moving to A_1, we utilize the peeling technique on $\overline{\Delta}_t$ to enhance the precision of our analysis. This involves decomposing $\overline{\Delta}_t$ into piecewise ranges defined as $B\cdot 2^{-\ell} < \overline{\Delta}_t \leq B\cdot 2^{-\ell+1}$, facilitating a more granular analysis of regret. From a broader perspective, this approach can be likened to approximating the area under a graph using rectangles; the smaller the area of the rectangle, the more precise and accurate the resulting analysis becomes. In our context, this technique significantly reduces the looseness of the analysis by a factor of $\frac{1}{\Delta}$.

$$A_1 \leq \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\overline{\Delta}_t\,\mathbb{1}\left\{\mathcal{D}_{t,\ell}(A_t)\right\}\mathbb{1}\left\{A_t \neq a_t^*\right\}\mathbb{1}\left\{\overline{\mathcal{E}}_t\right\}\right] + \mathbb{E}\left[\sum_{t=1}^{n}\overline{\Delta}_t\,\mathbb{1}\left\{\overline{\mathcal{D}}_{t,L}(a)\right\}\mathbb{1}\left\{A_t \neq a_t^*\right\}\mathbb{1}\left\{\overline{\mathcal{E}}_t\right\}\right].$$

$$\underbrace{\phantom{\mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\overline{\Delta}_t\,\mathbb{1}\left\{\mathcal{D}_{t,\ell}(A_t)\right\}\mathbb{1}\left\{A_t \neq a_t^*\right\}\mathbb{1}\left\{\overline{\mathcal{E}}_t\right\}\right]}}_{B_1} \qquad \underbrace{\phantom{\mathbb{E}\left[\sum_{t=1}^{n}\overline{\Delta}_t\,\mathbb{1}\left\{\overline{\mathcal{D}}_{t,L}(a)\right\}\mathbb{1}\left\{A_t \neq a_t^*\right\}\mathbb{1}\left\{\overline{\mathcal{E}}_t\right\}\right]}}_{B_2}$$

The term B_2 can be bounded as follows: We maintain the analysis variable L unchanged throughout the lemma, as it serves as the critical parameter in deriving both instance-dependent and minimax regret bounds from this

lemma.

$$B_2 \leq \mathbb{E}\left[\sum_{t=1}^{n} \overline{\Delta}_t \, \mathbb{1}\left\{\overline{\mathcal{D}}_{t,L}(a)\right\} \mathbb{1}\left\{\overline{\mathcal{E}}_t\right\}\right]$$

$$\leq \mathbb{E}\left[\sum_{t=1}^{n} \overline{\Delta}_t \, \mathbb{1}\left\{\overline{\mathcal{D}}_{t,L}(a)\right\}\right]$$

$$= \mathbb{E}\left[\sum_{t=1}^{n} \overline{\Delta}_t \, \mathbb{1}\left\{\overline{\Delta}_t < B \cdot 2^{-L}\right\} \mathbb{1}\left\{B \cdot 2^{-L} \leq \Delta\right\}\right]$$

$$+ \mathbb{E}\left[\sum_{t=1}^{n} \overline{\Delta}_t \, \mathbb{1}\left\{\overline{\Delta}_t < B \cdot 2^{-L}\right\} \mathbb{1}\left\{B \cdot 2^{-L} > \Delta\right\}\right]$$

$$= 0 + nB2^{-L} \cdot \mathbb{1}\left\{B \cdot 2^{-L} > \Delta\right\}$$

$$= n \cdot B \cdot 2^{-L} \cdot \mathbb{1}\left\{B \cdot 2^{-L} > \Delta\right\}.$$

This concludes the bounding of the term B_2.

The first term B_1 can be further split into separate terms based on the condition $\mathcal{U}_{t-1,\ell}$ as follows:

$$B_1 = \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L} \overline{\Delta}_t \, \mathbb{1}\left\{\mathcal{D}_{t,\ell}(A_t)\right\} \mathbb{1}\left\{A_t \neq a_t^*\right\} \mathbb{1}\left\{\overline{\mathcal{E}}_t\right\}\right]$$

$$= \underbrace{\mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L} \overline{\Delta}_t \, \mathbb{1}\left\{\mathcal{D}_{t,\ell}(A_t)\right\} \mathbb{1}\left\{A_t \neq a_t^*\right\} \mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(A_t)\right\} \mathbb{1}\left\{\overline{\mathcal{E}}_t\right\}\right]}_{D_1}$$

$$+ \underbrace{\mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L} \overline{\Delta}_t \, \mathbb{1}\left\{\mathcal{D}_{t,\ell}(A_t)\right\} \mathbb{1}\left\{A_t \neq a_t^*\right\} \mathbb{1}\left\{\mathcal{U}_{t-1,\ell}(A_t)\right\} \mathbb{1}\left\{\overline{\mathcal{E}}_t\right\}\right]}_{D_2}.$$

The term D_2 can be bounded using Elliptical Potential Count (Lemma 11) as follows:

$$D_2 = \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L} \overline{\Delta}_t \, \mathbb{1}\left\{\mathcal{D}_{t,\ell}(A_t)\right\} \mathbb{1}\left\{A_t \neq a_t^*\right\} \mathbb{1}\left\{\mathcal{U}_{t-1,\ell}(A_t)\right\} \mathbb{1}\left\{\overline{\mathcal{E}}_t\right\}\right]$$

$$\leq \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L} \overline{\Delta}_t \, \mathbb{1}\left\{\mathcal{D}_{t,\ell}(A_t)\right\} \mathbb{1}\left\{A_t \neq a_t^*\right\} \mathbb{1}\left\{\mathcal{U}_{t-1,\ell}(A_t)\right\}\right]$$

$$= \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L} \overline{\Delta}_t \, \mathbb{1}\left\{\mathcal{D}_{t,\ell}(A_t)\right\} \mathbb{1}\left\{A_t \neq a_t^*\right\} \mathbb{1}\left\{\|A_t\|_{V_{t-1}^{-1}}^2 \geq \varepsilon_\ell\right\}\right]$$

$$\leq \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L} \overline{\Delta}_t \, \mathbb{1}\left\{\mathcal{D}_{t,\ell}(A_t)\right\} \mathbb{1}\left\{\|A_t\|_{V_{t-1}^{-1}}^2 \geq \varepsilon_\ell\right\}\right]$$

$$= \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L} \overline{\Delta}_t \, \mathbb{1}\left\{B \cdot 2^{-\ell} \leq \overline{\Delta}_t \leq B \cdot 2^{-\ell+1}\right\} \mathbb{1}\left\{\|A_t\|_{V_{t-1}^{-1}}^2 \geq \varepsilon_\ell\right\}\right]$$

$$\leq \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L} B \cdot 2^{-\ell+1} \cdot \mathbb{1}\left\{\|A_t\|_{V_{t-1}^{-1}}^2 \geq \varepsilon_\ell\right\}\right]$$

$$\leq \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}B\cdot 2^{-\ell+1}\cdot \mathbb{1}\left\{\|A_t\|_{V_{t-1}^{-1}}^2 \geq 2^{-2\ell}\varepsilon\right\}\right]$$

$$\leq \sum_{\ell=1}^{L}B\cdot 2^{-\ell+1}\cdot \mathbb{E}\left[\sum_{t=1}^{n}\mathbb{1}\left\{\|A_t\|_{V_{t-1}^{-1}}^2 \geq 2^{-2\ell}\varepsilon\right\}\right].$$

We can apply the EPC from Lemma 11 directly only when $2^{-2\ell}\varepsilon < 1$. Consequently, we must analyze it in two cases, as follows:

$$D_2 \leq \sum_{\ell=1}^{L}B\cdot 2^{-\ell+1}\cdot \mathbb{E}\left[\sum_{t=1}^{n}\mathbb{1}\left\{\|A_t\|_{V_{t-1}^{-1}}^2 \geq 2^{-2\ell}\varepsilon\right\}\right]$$

$$= \sum_{\ell=1}^{L}B\cdot 2^{-\ell+1}\cdot \mathbb{E}\left[\sum_{t=1}^{n}\mathbb{1}\left\{\|A_t\|_{V_{t-1}^{-1}}^2 \geq 2^{-2\ell}\varepsilon\right\}\left(\mathbb{1}\left\{2^{-2\ell}\varepsilon \leq 1\right\}+\mathbb{1}\left\{2^{-2\ell}\varepsilon > 1\right\}\right)\right]$$

$$\leq \sum_{\ell=1}^{L}B\cdot 2^{-\ell+1}\cdot \mathbb{E}\left[\sum_{t=1}^{n}\mathbb{1}\left\{\|A_t\|_{V_{t-1}^{-1}}^2 \geq 2^{-2\ell}\varepsilon\right\}\mathbb{1}\left\{2^{-2\ell}\varepsilon \leq 1\right\}\right]$$

$$+ \sum_{\ell=1}^{L}B\cdot 2^{-\ell+1}\cdot \mathbb{E}\left[\sum_{t=1}^{n}\mathbb{1}\left\{\|A_t\|_{V_{t-1}^{-1}}^2 \geq 2^{-2\ell}\varepsilon\right\}\mathbb{1}\left\{2^{-2\ell}\varepsilon > 1\right\}\right]$$

$$\leq \sum_{\ell=1}^{L}B\cdot 2^{-\ell+1}\cdot 3\frac{d}{2^{-2\ell}\varepsilon}\log\left(1+\frac{2}{\lambda 2^{-2\ell}\varepsilon}\right) \qquad \text{(by lemma 11)}$$

$$+ \sum_{\ell=1}^{L}B\cdot 2^{-\ell+1}\mathbb{E}\left[\sum_{t=1}^{n}\mathbb{1}\left\{\|A_t\|_{V_{t-1}^{-1}}^2 \geq 1\right\}\right]$$

$$\leq \sum_{\ell=1}^{L}B\cdot 2^{-\ell+1}\cdot 3\frac{d}{2^{-2\ell}\varepsilon}\log\left(1+\frac{2}{\lambda 2^{-2\ell}\varepsilon}\right)+\sum_{\ell=1}^{L}B\cdot 2^{-\ell+1}\cdot 3d\log\left(1+\frac{2}{\lambda}\right) \qquad \text{(by lemma 11)}$$

$$\leq \frac{12dB}{2^{-L}\varepsilon}\log\left(1+\frac{2}{\lambda 2^{-2L}\varepsilon}\right)+6dB\log\left(1+\frac{2}{\lambda}\right).$$

This concludes the bounding of the term D_2.

Moving on to D_1, we can write D_1 into 3 terms based on the conditions $\mathcal{V}_{t-1}$ and $\mathcal{W}_{t-1,\ell}$ as follows:

$$D_1 = \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\overline{\Delta}_t\,\mathbb{1}\left\{\mathcal{D}_{t,\ell}(A_t)\right\}\mathbb{1}\left\{A_t \neq a_t^*\right\}\mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(A_t)\right\}\mathbb{1}\left\{\overline{\mathcal{E}}_t\right\}\right]$$

$$= \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\overline{\Delta}_t\,\mathbb{1}\left\{\mathcal{D}_{t,\ell}(A_t)\right\}\mathbb{1}\left\{A_t \neq a_t^*\right\}\mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(A_t)\right\}\mathbb{1}\left\{\mathcal{V}_{t-1}(A_t)\right\}\mathbb{1}\left\{\overline{\mathcal{E}}_t\right\}\right]$$

$$+ \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\overline{\Delta}_t\,\mathbb{1}\left\{\mathcal{D}_{t,\ell}(A_t)\right\}\mathbb{1}\left\{A_t \neq a_t^*\right\}\mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(A_t)\right\}\mathbb{1}\left\{\overline{\mathcal{V}}_{t-1}(A_t)\right\}\mathbb{1}\left\{\overline{\mathcal{E}}_t\right\}\right]$$

$$= \underbrace{\mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\overline{\Delta}_t\,\mathbb{1}\left\{\mathcal{D}_{t,\ell}(A_t)\right\}\mathbb{1}\left\{A_t \neq a_t^*\right\}\mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(A_t)\right\}\mathbb{1}\left\{\mathcal{V}_{t-1}(A_t)\right\}\mathbb{1}\left\{\overline{\mathcal{E}}_t\right\}\right]}_{F_1}$$

$$+ \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\overline{\Delta}_t \mathbb{1}\left\{\mathcal{D}_{t,\ell}(A_t)\right\}\mathbb{1}\left\{A_t \neq a_t^*\right\}\mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(A_t)\right\}\mathbb{1}\left\{\overline{\mathcal{V}}_{t-1}(A_t)\right\}\mathbb{1}\left\{\mathcal{W}_{t-1,\ell}\right\}\mathbb{1}\left\{\overline{\mathcal{E}}_t\right\}\right]}_{F_2}$$

$$+ \underbrace{\mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\overline{\Delta}_t \mathbb{1}\left\{\mathcal{D}_{t,\ell}(A_t)\right\}\mathbb{1}\left\{A_t \neq a_t^*\right\}\mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(A_t)\right\}\mathbb{1}\left\{\overline{\mathcal{V}}_{t-1}(A_t)\right\}\mathbb{1}\left\{\overline{\mathcal{W}}_{t-1,\ell}\right\}\mathbb{1}\left\{\overline{\mathcal{E}}_t\right\}\right]}_{F_3}.$$

The term F_1 is conditioned on following events:

1. $\overline{\mathcal{U}}_{t-1,\ell}(a) = \left\{\|a\|_{V_{t-1}^{-1}}^2 < \varepsilon_\ell\right\}$, which will be partly useful for upper bounding the denominator of $f_t(a)$. Intuitively, this condition indicates that arm a has been sufficiently explored.

2. $\mathcal{V}_{t-1}(a) = \left\{\hat{\Delta}_{a,t} \geq \frac{\Delta_{a,t}}{1+c}\right\}$, which will provide a lower bound for the numerator of $f_t(a)$. This condition suggests that the empirical gap is larger than the true gap, thereby ensuring that arm a is appropriately distinguished as sub-optimal.

Thus, we must be able to control the probability of selecting a sub-optimal arm as follows:

$$F_1 = \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\overline{\Delta}_t \mathbb{1}\left\{\mathcal{D}_{t,\ell}(A_t)\right\}\mathbb{1}\left\{A_t \neq a_t^*\right\}\mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(A_t)\right\}\mathbb{1}\left\{\mathcal{V}_{t-1}(A_t)\right\}\mathbb{1}\left\{\overline{\mathcal{E}}_t\right\}\right]$$

$$\leq \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\overline{\Delta}_t \mathbb{1}\left\{\mathcal{D}_{t,\ell}(A_t)\right\}\mathbb{1}\left\{A_t \neq a_t^*\right\}\mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(A_t)\right\}\mathbb{1}\left\{\mathcal{V}_{t-1}(A_t)\right\}\right]$$

$$\leq \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\sum_{a\in\mathcal{A}_t}\overline{\Delta}_{a,t} \mathbb{1}\left\{\mathcal{D}_{t,\ell}(a)\right\}\mathbb{1}\left\{A_t = a\right\}\mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(a)\right\}\mathbb{1}\left\{\mathcal{V}_{t-1}(a)\right\}\right]$$

$$= \mathbb{E}\left[\mathbb{E}_{t-1}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\sum_{a\in\mathcal{A}_t}\overline{\Delta}_{a,t} \mathbb{1}\left\{\mathcal{D}_{t,\ell}(a)\right\}\mathbb{1}\left\{A_t = a\right\}\mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(a)\right\}\mathbb{1}\left\{\mathcal{V}_{t-1}(a)\right\}\right]\right] \qquad \text{(tower rule)}$$

$$= \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\sum_{a\in\mathcal{A}_t}\overline{\Delta}_{a,t} \mathbb{1}\left\{\mathcal{D}_{t,\ell}(a)\right\}\mathbb{E}_{t-1}\left[\mathbb{1}\left\{A_t = a\right\}\right]\mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(a)\right\}\mathbb{1}\left\{\mathcal{V}_{t-1}(a)\right\}\right]$$

$$= \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\sum_{a\in\mathcal{A}_t}\overline{\Delta}_{a,t} \mathbb{1}\left\{\mathcal{D}_{t,\ell}(a)\right\}p_t(a)\mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(a)\right\}\mathbb{1}\left\{\mathcal{V}_{t-1}(a)\right\}\right]$$

$$= \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\sum_{a\in\mathcal{A}_t}\overline{\Delta}_{a,t} \mathbb{1}\left\{\mathcal{D}_{t,\ell}(a)\right\}q_t(a)\frac{f_t(a)}{\sum_{b\in\mathcal{A}_t}q_t(b)f_t(b)}\mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(a)\right\}\mathbb{1}\left\{\mathcal{V}_{t-1}(a)\right\}\right]$$

$$\leq \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\sum_{a\in\mathcal{A}_t}\overline{\Delta}_{a,t} \mathbb{1}\left\{\mathcal{D}_{t,\ell}(a)\right\}q_t(a)f_t(a)\mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(a)\right\}\mathbb{1}\left\{\mathcal{V}_{t-1}(a)\right\}\right]\frac{1}{\alpha_{\text{emp}}} \qquad \text{(by lemma 3)}$$

$$= \underbrace{\mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\sum_{a\in\mathcal{A}_t}\overline{\Delta}_{a,t} \mathbb{1}\left\{\mathcal{D}_{t,\ell}(a)\right\}q_t(a)f_t(a)\mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(a)\right\}\mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(\hat{a}_t)\right\}\mathbb{1}\left\{\mathcal{V}_{t-1}(a)\right\}\right]}_{F_{11}}\frac{1}{\alpha_{\text{emp}}}$$

$$
+ \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\sum_{a\in\mathcal{A}_t}\overline{\Delta}_{a,t}\,\mathbb{1}\left\{\mathcal{D}_{t,\ell}(a)\right\}q_t(a)f_t(a)\,\mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(a)\right\}\mathbb{1}\left\{\mathcal{U}_{t-1,\ell}(\hat{a}_t)\right\}\mathbb{1}\left\{\mathcal{V}_{t-1}(a)\right\}\right]\underbrace{}_{F_{12}}\frac{1}{\alpha_{\mathrm{emp}}}.
$$

Since the denominator of $f_t(a)$ includes the Mahalanobis norm of the empirical best arm, it is essential to establish a bound on $\|\hat{a}_t\|^2_{V_{t-1}^{-1}}$. To achieve this, we further decompose the term F_1 into F_{11} and F_{12}. The term F_{11} comprises both the denominator and numerator as expected. Therefore, a modest application of algebra, alongside the triangle inequality, should yield the necessary bound. We will defer the tuning of ε until the conclusion to obtain a desirable bound.

$$
F_{11} = \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\sum_{a\in\mathcal{A}_t,a\neq\hat{a}_t}\overline{\Delta}_{a,t}\,\mathbb{1}\left\{\mathcal{D}_{t,\ell}(a)\right\}q_t(a)f_t(a)\,\mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(a)\right\}\mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(\hat{a}_t)\right\}\right.
$$

$$
\left.\mathbb{1}\left\{\mathcal{V}_{t-1}(a)\right\}\right]
$$

$$
+ \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\overline{\Delta}_{\hat{a}_t,t}\,\mathbb{1}\left\{\mathcal{D}_{t,\ell}(\hat{a}_t)\right\}q_t(\hat{a}_t)f_t(\hat{a}_t)\,\mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(\hat{a}_t)\right\}\mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(\hat{a}_t)\right\}\right.
$$

$$
\left.\mathbb{1}\left\{\mathcal{V}_{t-1}(\hat{a}_t)\right\}\right]
$$

$$
= \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\sum_{a\in\mathcal{A}_t,a\neq\hat{a}_t}\overline{\Delta}_{a,t}\,\mathbb{1}\left\{\mathcal{D}_{t,\ell}(a)\right\}q_t(a)f_t(a)\,\mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(a)\right\}\mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(\hat{a}_t)\right\}\right.
$$

$$
\left.\mathbb{1}\left\{\mathcal{V}_{t-1}(a)\right\}\right] \qquad\qquad (\mathbb{1}\left\{\mathcal{V}_{t-1}(\hat{a}_t)\right\}=0)
$$

$$
\leq \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\sum_{a\in\mathcal{A}_t}\overline{\Delta}_{a,t}\,\mathbb{1}\left\{\mathcal{D}_{t,\ell}(a)\right\}q_t(a)\exp\left(-\frac{\hat{\Delta}_{a,t}^2}{2\beta_{t-1}(\delta_{t-1})\left(\|\hat{a}_t\|^2_{V_{t-1}^{-1}}+\|a\|^2_{V_{t-1}^{-1}}\right)}\right)\mathbb{1}\left\{\|a\|^2_{V_{t-1}^{-1}}<\varepsilon_\ell\right\}\right.
$$

$$
(\text{ from lemma } 2)
$$

$$
\left.\mathbb{1}\left\{\|\hat{a}_t\|^2_{V_{t-1}^{-1}}<\varepsilon_\ell\right\}\mathbb{1}\left\{\mathcal{V}_{t-1}(a)\right\}\right]
$$

$$
\leq \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\sum_{a\in\mathcal{A}_t}\overline{\Delta}_{a,t}\,\mathbb{1}\left\{\mathcal{D}_{t,\ell}(a)\right\}q_t(a)\exp\left(-\frac{\hat{\Delta}_{a,t}^2}{2\beta_{t-1}(\delta_{t-1})\left(\varepsilon_\ell+\varepsilon_\ell\right)}\right)\mathbb{1}\left\{\mathcal{V}_{t-1}(a)\right\}\right]
$$

$$
= \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\sum_{a\in\mathcal{A}_t}\overline{\Delta}_{a,t}\,\mathbb{1}\left\{\mathcal{D}_{t,\ell}(a)\right\}q_t(a)\exp\left(-\frac{\hat{\Delta}_{a,t}^2}{4\beta_{t-1}(\delta_{t-1})\varepsilon_\ell}\right)\mathbb{1}\left\{\hat{\Delta}_{a,t}\geq\frac{\Delta_{a,t}}{1+c}\right\}\right]
$$

$$
\leq \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\sum_{a\in\mathcal{A}_t}\overline{\Delta}_{a,t}\,\mathbb{1}\left\{\mathcal{D}_{t,\ell}(a)\right\}q_t(a)\exp\left(-\frac{\left(\frac{\Delta_{a,t}}{1+c}\right)^2}{4\beta_{t-1}(\delta_{t-1})\varepsilon_\ell}\right)\right].
$$

This is where the peeling technique proves beneficial. By applying the peeling technique to $\overline{\Delta}_{a,t}$, we can effectively

cancel out the denominator and numerator as follows:

$$F_{11} \leq \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\sum_{a \in \mathcal{A}_t} \overline{\Delta}_{a,t} \mathbb{1}\left\{B \cdot 2^{-\ell} \leq \overline{\Delta}_{a,t} \leq B \cdot 2^{-\ell+1}\right\} q_t(a) \exp\left(-\frac{\left(\frac{\Delta_{a,t}}{1+c}\right)^2}{4\beta_{t-1}(\delta_{t-1})\varepsilon_\ell}\right)\right]$$

$$\leq \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\sum_{a \in \mathcal{A}_t} B \cdot 2^{-\ell+1} q_t(a) \exp\left(-\frac{\left(\frac{B \cdot 2^{-\ell}}{1+c}\right)^2}{4\beta_{t-1}(\delta_{t-1})\varepsilon_\ell}\right)\right]$$

$$= \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\left(\sum_{a \in \mathcal{A}_t} q_t(a)\right) B \cdot 2^{-\ell+1} \exp\left(-\frac{\left(\frac{B \cdot 2^{-\ell}}{1+c}\right)^2}{4\beta_{t-1}(\delta_{t-1})\varepsilon_\ell}\right)\right]$$

$$= \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L} B \cdot 2^{-\ell+1} \exp\left(-\frac{\left(\frac{B \cdot 2^{-\ell}}{1+c}\right)^2}{4\beta_{t-1}(\delta_{t-1})\varepsilon_\ell}\right)\right]$$

$$= \sum_{t=1}^{n}\sum_{\ell=1}^{L} B \cdot 2^{-\ell+1} \exp\left(-\frac{B^2}{4\varepsilon \cdot \beta_{t-1}(\delta_{t-1})(1+c)^2}\right)$$

$$= B \cdot \sum_{t=1}^{n} \exp\left(-\frac{B^2}{4\varepsilon \cdot \beta_{t-1}(\delta_{t-1})(1+c)^2}\right) \sum_{\ell=1}^{L} 2^{-\ell+1}$$

$$= 2B \cdot \sum_{t=1}^{n} \exp\left(-\frac{B^2}{4\varepsilon \cdot \beta_{t-1}(\delta_{t-1})(1+c)^2}\right) \sum_{\ell=1}^{L} 2^{-\ell}$$

$$\leq 2B \cdot \sum_{t=1}^{n} \exp\left(-\frac{B^2}{4\varepsilon \cdot \beta_{t-1}(\delta_{t-1})(1+c)^2}\right) \sum_{\ell=1}^{\infty} 2^{-\ell}$$

$$= 4B \sum_{t=1}^{n} \exp\left(-\frac{B^2}{4\varepsilon \cdot \beta_{t-1}(\delta_{t-1})(1+c)^2}\right). \qquad \text{(geometric sum)}$$

Since $\beta_{t-1}(\delta_{t-1})$ is an increasing function in t, we can upper bound $\beta_{t-1}(\delta_{t-1})$ with $\beta_n(\delta_n)$ which leads to,

$$F_{11} \leq 4B \sum_{t=1}^{n} \exp\left(-\frac{B^2}{4\varepsilon \cdot \beta_n(\delta_n)(1+c)^2}\right)$$

$$\leq 4Bn \exp\left(-\frac{B^2}{4\varepsilon \cdot \beta_n(\delta_n)(1+c)^2}\right)$$

$$\leq 4Bn \exp\left(-\frac{B^2}{16\varepsilon \cdot \beta_n(\delta_n)}\right). \qquad \text{(choose } c = 1\text{)}$$

This concludes the bounding of the term F_{11}.

The term F_{12} presents a challenge. Unlike F_{11}, we cannot bound the probability directly because $\|\hat{a}_t\|^2_{V_{t-1}^{-1}}$ is unbounded. However, we know that $\hat{a}_t$ is assigned a probability greater than the constant α_{emp} of being chosen at each round. Additionally, the elliptical potential count provides a bound on the number of times $\|A_t\|^2_{V_{t-1}^{-1}}$ can exceed ε_ℓ. Since the empirical best arm is chosen with significant probability, the elliptical potential count also indirectly limits the number of occurrences where $\|\hat{a}_t\|^2_{V_{t-1}^{-1}} \geq \varepsilon_\ell$ (See Claim 1). Therefore,

$$F_{12} = \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\sum_{a \in \mathcal{A}_t} \overline{\Delta}_{a,t} \mathbb{1}\left\{\mathcal{D}_{t,\ell}(a)\right\} q_t(a) f_t(a) \mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(a)\right\} \mathbb{1}\left\{\mathcal{U}_{t-1,\ell}(\hat{a}_t)\right\} \mathbb{1}\left\{\mathcal{V}_{t-1}(a)\right\}\right]$$

$$\leq \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\sum_{a\in\mathcal{A}_t}\overline{\Delta}_{a,t}\,\mathbb{1}\left\{\mathcal{D}_{t,\ell}(a)\right\}q_t(a)f_t(a)\,\mathbb{1}\left\{\mathcal{U}_{t-1,\ell}(\hat{a}_t)\right\}\right]$$

$$\leq \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\sum_{a\in\mathcal{A}_t}\overline{\Delta}_{a,t}\,\mathbb{1}\left\{\mathcal{D}_{t,\ell}(a)\right\}q_t(a)\,\mathbb{1}\left\{\mathcal{U}_{t-1,\ell}(\hat{a}_t)\right\}\right] \qquad (f_t(a)\leq 1)$$

$$= \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\sum_{a\in\mathcal{A}_t}\overline{\Delta}_{a,t}\,\mathbb{1}\left\{B\cdot 2^{-\ell}\leq\overline{\Delta}_{a,t}\leq B\cdot 2^{-\ell+1}\right\}q_t(a)\,\mathbb{1}\left\{\mathcal{U}_{t-1,\ell}(\hat{a}_t)\right\}\right]$$

$$\leq B\,\mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\sum_{a\in\mathcal{A}_t}2^{-\ell+1}q_t(a)\,\mathbb{1}\left\{\mathcal{U}_{t-1,\ell}(\hat{a}_t)\right\}\right]$$

$$= B\,\mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\left(\sum_{a\in\mathcal{A}_t}q_t(a)\right)2^{-\ell+1}\,\mathbb{1}\left\{\mathcal{U}_{t-1,\ell}(\hat{a}_t)\right\}\right]$$

$$= B\,\mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}2^{-\ell+1}\,\mathbb{1}\left\{\mathcal{U}_{t-1,\ell}(\hat{a}_t)\right\}\right]$$

$$= B\,\mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}2^{-\ell+1}\,\mathbb{1}\left\{\|\hat{a}_t\|_{V_{t-1}^{-1}}^2\geq\varepsilon_\ell\right\}\right]$$

$$= B\sum_{\ell=1}^{L}2^{-\ell+1}\,\mathbb{E}\left[\sum_{t=1}^{n}\mathbb{1}\left\{\|\hat{a}_t\|_{V_{t-1}^{-1}}^2\geq\varepsilon_\ell\right\}\right]$$

$$= \frac{B}{\alpha_{\mathrm{emp}}}\sum_{\ell=1}^{L}2^{-\ell+1}\,\mathbb{E}\left[\sum_{t=1}^{n}\mathbb{1}\left\{\|A_t\|_{V_{t-1}^{-1}}^2\geq\varepsilon_\ell\right\}\right]. \qquad \text{(by Claim 1)}$$

From this point onward, the remaining results follow directly from D_2 after applying Lemma 11,

$$F_{12}\leq\frac{1}{\alpha_{\mathrm{emp}}}\left(\frac{12dB}{2^{-L}\varepsilon}\log\left(1+\frac{2}{\lambda 2^{-2L}\varepsilon}\right)+6dB\log\left(1+\frac{2}{\lambda}\right)\right).$$

This concludes the bounding of the term F_{12}.

By combining the bounds obtained for F_{11} and F_{12}, we can now establish a bound for F_1.

$$F_1\leq\frac{1}{\alpha_{\mathrm{emp}}}F_{11}+\frac{1}{\alpha_{\mathrm{emp}}}F_{12}$$

$$\leq\frac{1}{\alpha_{\mathrm{emp}}}\cdot 4Bn\exp\left(-\frac{B^2}{16\varepsilon\cdot\beta_n(\delta_n)}\right)+\frac{1}{\alpha_{\mathrm{emp}}^2}\left(\frac{12dB}{2^{-L}\varepsilon}\log\left(1+\frac{2}{\lambda 2^{-2L}\varepsilon}\right)+6dB\log\left(1+\frac{2}{\lambda}\right)\right).$$

This concludes the bounding of the term F_1.

We now proceed to analyze the term F_2. Specifically, we will further decompose F_2 into two components, F_{21} and F_{22}, based on the occurrence of the favorable event $\mathcal{G}_1$ as follows:

$$F2=\mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\overline{\Delta}_t\,\mathbb{1}\left\{\mathcal{D}_{t,\ell}(A_t)\right\}\mathbb{1}\left\{A_t\neq a_t^*\right\}\mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(A_t)\right\}\mathbb{1}\left\{\overline{\mathcal{V}}_{t-1}(A_t)\right\}\mathbb{1}\left\{\mathcal{W}_{t-1,\ell}\right\}\mathbb{1}\left\{\overline{\mathcal{E}}_t\right\}\right]$$

$$\leq\mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\overline{\Delta}_t\,\mathbb{1}\left\{\mathcal{D}_{t,\ell}(A_t)\right\}\mathbb{1}\left\{A_t\neq a_t^*\right\}\mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(A_t)\right\}\mathbb{1}\left\{\overline{\mathcal{V}}_{t-1}(A_t)\right\}\mathbb{1}\left\{\mathcal{W}_{t-1,\ell}\right\}\right]$$

$$= \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\overline{\Delta}_t \, \mathbb{1}\left\{\mathcal{D}_{t,\ell}(A_t)\right\}\mathbb{1}\left\{A_t \neq a_t^*\right\}\mathbb{1}\left\{\mathcal{G}_1\right\}\mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(A_t)\right\}\mathbb{1}\left\{\overline{\mathcal{V}}_{t-1}(A_t)\right\}\mathbb{1}\left\{\mathcal{W}_{t-1,\ell}\right\}\right]}_{F_{21}}$$

$$+ \underbrace{\mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\overline{\Delta}_t \, \mathbb{1}\left\{\mathcal{D}_{t,\ell}(A_t)\right\}\mathbb{1}\left\{A_t \neq a_t^*\right\}\mathbb{1}\left\{\overline{\mathcal{G}}_1\right\}\mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(A_t)\right\}\mathbb{1}\left\{\overline{\mathcal{V}}_{t-1}(A_t)\right\}\mathbb{1}\left\{\mathcal{W}_{t-1,\ell}\right\}\right]}_{F_{22}}.$$

In the term F_{22}, the favorable event $\mathcal{G}_1$ does not occur. Given the low probability of this occurrence, we can bound F_{22} quite straightforwardly as follows:

$$F_{22} = \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\overline{\Delta}_t \, \mathbb{1}\left\{\mathcal{D}_{t,\ell}(A_t)\right\}\mathbb{1}\left\{A_t \neq a_t^*\right\}\mathbb{1}\left\{\overline{\mathcal{G}}_1\right\}\mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(A_t)\right\}\mathbb{1}\left\{\overline{\mathcal{V}}_{t-1}(A_t)\right\}\mathbb{1}\left\{\mathcal{W}_{t-1,\ell}\right\}\right]$$

$$\leq B\,\mathbb{E}\left[\sum_{t=1}^{n}\mathbb{1}\left\{\overline{\mathcal{G}}_1\right\}\right]$$

$$\leq B\sum_{t=1}^{n}\mathbb{E}\left[\mathbb{1}\left\{\overline{\mathcal{G}}_1\right\}\right]$$

$$\leq B\sum_{t=1}^{n}\mathbb{P}\left(\overline{\mathcal{G}}_1\right)$$

$$\leq B\sum_{t=1}^{n}\delta_t \qquad\qquad\qquad\qquad\qquad \text{(Lemma 13)}$$

$$= B\sum_{t=1}^{n}\frac{1}{t+1} \qquad\qquad\qquad\qquad\qquad (\delta_t = \tfrac{1}{t+1})$$

$$\leq B\log(n+1).$$

This concludes the bounding of the term F_{22}.

However, the term F_{21} is more complex. It encompasses the following primary events: It has the following main events:

1. $\mathcal{W}_{t-1,\ell} = \left\{\max_{a' \in \mathcal{A}_t}\langle\hat{\theta}_{t-1}, a'\rangle \geq \langle\theta^*, a_t^*\rangle - \varepsilon_{2,\ell}\right\}$. Intuitively, this event signifies that the estimated reward is sufficiently close to the maximum achievable reward.

2. $\mathcal{V}_{t-1}(a) = \left\{\hat{\Delta}_{a,t} \geq \frac{\Delta_{a,t}}{1+c}\right\}$.

3. $\mathcal{G}_1$.

Through a series of algebraic manipulations and parameter tuning, we demonstrate that the occurrence of all three events is contingent upon the condition $\|A_t\|_{V_{t-1}^{-1}}^2 \geq \frac{2^{-2\ell}}{16\beta_{t-1}(\delta_{t-1})}$. This condition can be effectively managed using the elliptical potential count.

$$F_{21} = \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\overline{\Delta}_t \, \mathbb{1}\left\{\mathcal{D}_{t,\ell}(A_t)\right\}\mathbb{1}\left\{A_t \neq a_t^*\right\}\mathbb{1}\left\{\mathcal{G}_1\right\}\mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(A_t)\right\}\mathbb{1}\left\{\overline{\mathcal{V}}_{t-1}(A_t)\right\}\mathbb{1}\left\{\mathcal{W}_{t-1,\ell}\right\}\right]$$

$$\leq \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\overline{\Delta}_t \, \mathbb{1}\left\{\mathcal{D}_{t,\ell}(A_t)\right\}\mathbb{1}\left\{\mathcal{G}_1\right\}\mathbb{1}\left\{\overline{\mathcal{V}}_{t-1}(A_t)\right\}\mathbb{1}\left\{\mathcal{W}_{t-1,\ell}\right\}\right]$$

$$= \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\overline{\Delta}_t \, \mathbb{1}\left\{\mathcal{D}_{t,\ell}(A_t)\right\}\mathbb{1}\left\{\mathcal{G}_1\right\}\mathbb{1}\left\{\hat{\Delta}_{A_t,t} < \frac{\Delta_{A_t,t}}{1+c}\right\}\mathbb{1}\left\{\mathcal{W}_{t-1,\ell}\right\}\right].$$

From this point onward, we adopt a proof style in which, if the occurrence of event A implies the occurrence of event B that is,

$$A \implies B,$$

then, we have

$$\mathbb{E}\left[\mathbb{1}\{A\}\right] \leq \mathbb{E}\left[\mathbb{1}\{B\}\right].$$

Similarly, if the occurrence of 2 events A, B implies the occurrence of a third event C that is,

$$A, B \implies C,$$

then, we have

$$\mathbb{E}\left[\mathbb{1}\{A\}\mathbb{1}\{B\}\right] \leq \mathbb{E}\left[\mathbb{1}\{C\}\right].$$

This approach to writing mathematical expressions facilitates a reduction in verbosity within the proof. Now, moving on to F_{21},

$$F_{21} = \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\overline{\Delta}_t\,\mathbb{1}\left\{\mathcal{D}_{t,\ell}(A_t)\right\}\mathbb{1}\{\mathcal{G}_1\}\,\mathbb{1}\left\{\max_{a'\in\mathcal{A}_t}\langle\hat{\theta}_{t-1},a'\rangle - \langle\hat{\theta}_{t-1},A_t\rangle < \frac{\Delta_{A_t,t}}{1+c}\right\}\right.$$
$$\left.\mathbb{1}\left\{\max_{a'\in\mathcal{A}_t}\langle\hat{\theta}_{t-1},a'\rangle \geq \langle\theta^*,a_t^*\rangle - \varepsilon_{2,\ell}\right\}\right]$$

$$\leq \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\overline{\Delta}_t\,\mathbb{1}\left\{\mathcal{D}_{t,\ell}(A_t)\right\}\mathbb{1}\{\mathcal{G}_1\}\,\mathbb{1}\left\{\langle\theta^*,a_t^*\rangle - \varepsilon_{2,\ell} - \langle\hat{\theta}_{t-1},A_t\rangle < \frac{\Delta_{A_t,t}}{1+c}\right\}\right]$$

$$= \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\overline{\Delta}_t\,\mathbb{1}\left\{\mathcal{D}_{t,\ell}(A_t)\right\}\mathbb{1}\{\mathcal{G}_1\}\,\mathbb{1}\left\{\langle\theta^*,a_t^*\rangle - \langle\theta^*,A_t\rangle - \varepsilon_{2,\ell} + \langle\theta^*,A_t\rangle - \langle\hat{\theta}_{t-1},A_t\rangle < \frac{\Delta_{A_t,t}}{1+c}\right\}\right]$$

$$= \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\overline{\Delta}_t\,\mathbb{1}\left\{\mathcal{D}_{t,\ell}(A_t)\right\}\mathbb{1}\{\mathcal{G}_1\}\,\mathbb{1}\left\{\Delta_{A_t,t} - \varepsilon_{2,\ell} + \langle\theta^*,A_t\rangle - \langle\hat{\theta}_{t-1},A_t\rangle < \frac{\Delta_{A_t,t}}{1+c}\right\}\right]$$

$$= \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\overline{\Delta}_t\,\mathbb{1}\left\{\mathcal{D}_{t,\ell}(A_t)\right\}\mathbb{1}\{\mathcal{G}_1\}\,\mathbb{1}\left\{\frac{c\Delta_{A_t,t}}{1+c} - \varepsilon_{2,\ell} < \langle\hat{\theta}_{t-1} - \theta^*,A_t\rangle\right\}\right]$$

$$= \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\overline{\Delta}_t\,\mathbb{1}\left\{\mathcal{D}_{t,\ell}(A_t)\right\}\mathbb{1}\{\mathcal{G}_1\}\,\mathbb{1}\left\{\frac{\Delta_{A_t,t}}{2} - \varepsilon_2\cdot 2^{-\ell} < \langle\hat{\theta}_{t-1} - \theta^*,A_t\rangle\right\}\right] \qquad \text{(choose } c=1\text{)}$$

$$= \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\overline{\Delta}_t\,\mathbb{1}\left\{B\cdot 2^{-\ell} \leq \overline{\Delta}_t \leq B\cdot 2^{-\ell+1}\right\}\mathbb{1}\{\mathcal{G}_1\}\,\mathbb{1}\left\{\frac{\Delta_{A_t,t}}{2} - \varepsilon_2\cdot 2^{-\ell} < \langle\hat{\theta}_{t-1} - \theta^*,A_t\rangle\right\}\right]$$

$$\leq \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}B\cdot 2^{-\ell+1}\,\mathbb{1}\left\{B\cdot 2^{-\ell} \leq \overline{\Delta}_t \leq B\cdot 2^{-\ell+1}\right\}\mathbb{1}\{\mathcal{G}_1\}\,\mathbb{1}\left\{\frac{\Delta_{A_t,t}}{2} - \varepsilon_2\cdot 2^{-\ell} < \langle\hat{\theta}_{t-1} - \theta^*,A_t\rangle\right\}\right]$$

$$\leq \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}B\cdot 2^{-\ell+1}\,\mathbb{1}\{\mathcal{G}_1\}\,\mathbb{1}\left\{\frac{B\cdot 2^{-\ell}}{2} - \varepsilon_2\cdot 2^{-\ell} < \langle\hat{\theta}_{t-1} - \theta^*,A_t\rangle\right\}\right]$$

$$= \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}B\cdot 2^{-\ell+1}\,\mathbb{1}\{\mathcal{G}_1\}\,\mathbb{1}\left\{\frac{B\cdot 2^{-\ell}}{4} < \langle\hat{\theta}_{t-1} - \theta^*,A_t\rangle\right\}\right] \qquad \text{(choose } \varepsilon_2 = \frac{B}{4}\text{)}$$

$$\leq \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}B\cdot 2^{-\ell+1}\,\mathbb{1}\{\mathcal{G}_1\}\,\mathbb{1}\left\{\frac{B^2\cdot 2^{-2\ell}}{16} < \|\hat{\theta}_{t-1} - \theta^*\|_{V_{t-1}}^2\|A_t\|_{V_{t-1}^{-1}}^2\right\}\right] \qquad \text{(Cauchy-Schwartz)}$$

$$= \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}B\cdot 2^{-\ell+1}\,\mathbb{1}\left\{\|\theta^* - \hat{\theta}_{t-1}\|_{V_{t-1}}^2 \leq \beta_{t-1}^*(\delta_{t-1})\right\}\mathbb{1}\left\{\frac{B^2\cdot 2^{-2\ell}}{16} < \|\hat{\theta}_{t-1} - \theta^*\|_{V_{t-1}}^2\|A_t\|_{V_{t-1}^{-1}}^2\right\}\right]$$

$$\leq \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L} B \cdot 2^{-\ell+1} \mathbb{1}\left\{\|A_t\|_{V_{t-1}^{-1}}^2 \geq \frac{B^2 \cdot 2^{-2\ell}}{16\beta_{t-1}^*(\delta_{t-1})}\right\}\right]$$

$$= \sum_{\ell=1}^{L} B \cdot 2^{-\ell+1} \mathbb{E}\left[\sum_{t=1}^{n} \mathbb{1}\left\{\|A_t\|_{V_{t-1}^{-1}}^2 \geq \frac{B^2 \cdot 2^{-2\ell}}{16\beta_{t-1}^*(\delta_{t-1})}\right\}\right].$$

We can apply the EPC from Lemma 11 directly only when $\frac{B^2 \cdot 2^{-2\ell}}{16\beta_{t-1}^*(\delta_{t-1})} \leq 1$. Consequently, we must analyze it in two cases as follows:

$$F_{21} \leq \sum_{\ell=1}^{L} B \cdot 2^{-\ell+1} \mathbb{E}\left[\sum_{t=1}^{n} \mathbb{1}\left\{\|A_t\|_{V_{t-1}^{-1}}^2 \geq \frac{B^2 \cdot 2^{-2\ell}}{16\beta_{t-1}^*(\delta_{t-1})}\right\}\right]$$

$$= \sum_{\ell=1}^{L} B \cdot 2^{-\ell+1} \mathbb{E}\left[\sum_{t=1}^{n} \mathbb{1}\left\{\|A_t\|_{V_{t-1}^{-1}}^2 \geq \frac{B^2 \cdot 2^{-2\ell}}{16\beta_{t-1}^*(\delta_{t-1})}\right\}\left(\mathbb{1}\left\{\frac{B^2 \cdot 2^{-2\ell}}{16\beta_{t-1}^*(\delta_{t-1})} \leq 1\right\} + \mathbb{1}\left\{\frac{B^2 \cdot 2^{-2\ell}}{16\beta_{t-1}^*(\delta_{t-1})} > 1\right\}\right)\right]$$

$$= \sum_{\ell=1}^{L} B \cdot 2^{-\ell+1} \mathbb{E}\left[\sum_{t=1}^{n} \mathbb{1}\left\{\|A_t\|_{V_{t-1}^{-1}}^2 \geq \frac{B^2 \cdot 2^{-2\ell}}{16\beta_{t-1}^*(\delta_{t-1})}\right\}\mathbb{1}\left\{\frac{B^2 \cdot 2^{-2\ell}}{16\beta_{t-1}^*(\delta_{t-1})} \leq 1\right\}\right]$$

$$+ \sum_{\ell=1}^{L} B \cdot 2^{-\ell+1} \mathbb{E}\left[\sum_{t=1}^{n} \mathbb{1}\left\{\|A_t\|_{V_{t-1}^{-1}}^2 \geq \frac{B^2 \cdot 2^{-2\ell}}{16\beta_{t-1}^*(\delta_{t-1})}\right\}\mathbb{1}\left\{\frac{B^2 \cdot 2^{-2\ell}}{16\beta_{t-1}^*(\delta_{t-1})} > 1\right\}\right]$$

$$\leq \sum_{\ell=1}^{L} B \cdot 2^{-\ell+1} \cdot 3\frac{d}{\frac{B^2 \cdot 2^{-2\ell}}{16\beta_{t-1}^*(\delta_{t-1})}} \log\left(1 + \frac{2}{\lambda \frac{B^2 \cdot 2^{-2\ell}}{16\beta_{t-1}^*(\delta_{t-1})}}\right) \qquad \text{(by lemma 11)}$$

$$+ \sum_{\ell=1}^{L} B \cdot 2^{-\ell+1} \mathbb{E}\left[\sum_{t=1}^{n} \mathbb{1}\left\{\|A_t\|_{V_{t-1}^{-1}}^2 \geq 1\right\}\right]$$

$$\leq \frac{192\beta_{t-1}^*(\delta_{t-1})d}{B \cdot 2^{-L}} \log\left(1 + \frac{32\beta_{t-1}^*(\delta_{t-1})}{\lambda B^2 \cdot 2^{-2L}}\right) + \sum_{\ell=1}^{L} B \cdot 2^{-\ell+1} \cdot 3d \log\left(1 + \frac{2}{\lambda}\right) \qquad \text{(by lemma 11)}$$

$$\leq \frac{192\beta_n^*(\delta_n)d}{B \cdot 2^{-L}} \log\left(1 + \frac{32\beta_n^*(\delta_n)}{\lambda B^2 \cdot 2^{-2L}}\right) + 6dB \log\left(1 + \frac{2}{\lambda}\right).$$

This concludes the bounding of the term F_{21}.

By combining the bounds obtained for F_{21} and F_{22}, we can now establish a bound for F_2.

$$F_2 = F_{21} + F_{22}$$

$$\leq B \log(n+1) + \frac{192\beta_n^*(\delta_n)d}{B \cdot 2^{-L}} \log\left(1 + \frac{32\beta_n^*(\delta_n)}{\lambda B^2 \cdot 2^{-2L}}\right) + 6dB \log\left(1 + \frac{2}{\lambda}\right).$$

This concludes the bounding of the term F_2.

We now proceed to analyze the term F_3. Similar to F_2, we will further decompose F_3 into two components, F_{31} and F_{32}, based on the occurrence of the favorable event $\mathcal{G}_1$ as follows:

$$F_3 = \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L} \overline{\Delta}_t \mathbb{1}\left\{\mathcal{D}_{t,\ell}(A_t)\right\} \mathbb{1}\left\{A_t \neq a_t^*\right\} \mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(A_t)\right\} \mathbb{1}\left\{\overline{\mathcal{V}}_{t-1}(A_t)\right\} \mathbb{1}\left\{\overline{\mathcal{W}}_{t-1,\ell}\right\} \mathbb{1}\left\{\overline{\mathcal{E}}_t\right\}\right]$$

$$= \mathbb{E}\left[\underbrace{\sum_{t=1}^{n}\sum_{\ell=1}^{L}\overline{\Delta}_t\,\mathbb{1}\left\{\mathcal{D}_{t,\ell}(A_t)\right\}\mathbb{1}\left\{A_t \neq a_t^*\right\}\mathbb{1}\left\{\mathcal{G}_1\right\}\mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(A_t)\right\}\mathbb{1}\left\{\overline{\mathcal{V}}_{t-1}(A_t)\right\}\mathbb{1}\left\{\overline{\mathcal{W}}_{t-1,\ell}\right\}\mathbb{1}\left\{\overline{\mathcal{E}}_t\right\}}_{F_{31}}\right]$$

$$+ \mathbb{E}\left[\underbrace{\sum_{t=1}^{n}\sum_{\ell=1}^{L}\overline{\Delta}_t\,\mathbb{1}\left\{\mathcal{D}_{t,\ell}(A_t)\right\}\mathbb{1}\left\{A_t \neq a_t^*\right\}\mathbb{1}\left\{\overline{\mathcal{G}}_1\right\}\mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(A_t)\right\}\mathbb{1}\left\{\overline{\mathcal{V}}_{t-1}(A_t)\right\}\mathbb{1}\left\{\overline{\mathcal{W}}_{t-1,\ell}\right\}\mathbb{1}\left\{\overline{\mathcal{E}}_t\right\}}_{F_{32}}\right].$$

Analogous to F_{22}, we can bound F_{32} with relative ease; therefore, we shall omit the details.

$$F_{32} \leq B\log(n+1). \hspace{4cm} (\delta_t = \tfrac{1}{t+1})$$

The analysis of the term F_{31} is the most complex and intricate among the terms. Initially, we exclude certain terms that are not pertinent to the specific analysis approach we will employ for F_{31}

$$F_{31} = \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\overline{\Delta}_t\,\mathbb{1}\left\{\mathcal{D}_{t,\ell}(A_t)\right\}\mathbb{1}\left\{A_t \neq a_t^*\right\}\mathbb{1}\left\{\mathcal{G}_1\right\}\mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(A_t)\right\}\mathbb{1}\left\{\overline{\mathcal{V}}_{t-1}(A_t)\right\}\mathbb{1}\left\{\overline{\mathcal{W}}_{t-1,\ell}\right\}\mathbb{1}\left\{\overline{\mathcal{E}}_t\right\}\right]$$

$$\leq \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\overline{\Delta}_t\,\mathbb{1}\left\{\mathcal{D}_{t,\ell}(A_t)\right\}\mathbb{1}\left\{\mathcal{G}_1\right\}\mathbb{1}\left\{\overline{\mathcal{W}}_{t-1,\ell}\right\}\mathbb{1}\left\{\overline{\mathcal{E}}_t\right\}\right].$$

In the derived simplified version above, it is evident that below 3 events are occurring, each serving a significant purpose in the analysis.

1. $\mathcal{G}_1$
2. $\overline{\mathcal{W}}_{t-1,\ell}$
3. $\overline{\mathcal{E}}_t$

Our primary objective is to bound the probability of the event $\overline{\mathcal{W}}_{t-1,\ell}$ occurring in conjunction with the other two events.

$$F_{31} \leq \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}\overline{\Delta}_t\,\mathbb{1}\left\{2^{-\ell} \leq \overline{\Delta}_t \leq 2^{-\ell+1}\right\}\mathbb{1}\left\{\mathcal{G}_1\right\}\mathbb{1}\left\{\overline{\mathcal{W}}_{t-1,\ell}\right\}\mathbb{1}\left\{\overline{\mathcal{E}}_t\right\}\right]$$

$$\leq \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}B\cdot 2^{-\ell+1}\,\mathbb{1}\left\{\mathcal{G}_1\right\}\mathbb{1}\left\{\overline{\mathcal{W}}_{t-1,\ell}\right\}\mathbb{1}\left\{\overline{\mathcal{E}}_t\right\}\right]$$

$$= \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}B\cdot 2^{-\ell+1}\,\mathbb{1}\left\{\mathcal{G}_1\right\}\mathbb{1}\left\{\max_{a'\in\mathcal{A}_t}\langle\hat{\theta}_{t-1},a'\rangle < \langle\theta^*,a_t^*\rangle - \varepsilon_{2,\ell}\right\}\mathbb{1}\left\{\overline{\mathcal{E}}_t\right\}\right]$$

$$\leq \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}B\cdot 2^{-\ell+1}\,\mathbb{1}\left\{\mathcal{G}_1\right\}\mathbb{1}\left\{\langle\hat{\theta}_{t-1},a_t^*\rangle < \langle\theta^*,a_t^*\rangle - \varepsilon_{2,\ell}\right\}\mathbb{1}\left\{\overline{\mathcal{E}}_t\right\}\right]$$

$$= \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}B\cdot 2^{-\ell+1}\,\mathbb{1}\left\{\mathcal{G}_1\right\}\mathbb{1}\left\{\varepsilon_{2,\ell} \leq \langle\theta^* - \hat{\theta}_{t-1},a_t^*\rangle\right\}\mathbb{1}\left\{\overline{\mathcal{E}}_t\right\}\right]$$

$$\leq \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L}B\cdot 2^{-\ell+1}\,\mathbb{1}\left\{\mathcal{G}_1\right\}\mathbb{1}\left\{\varepsilon_{2,\ell} \leq \|\theta^* - \hat{\theta}_{t-1}\|_{V(p_t)}\|a^*\|_{V(p_t)^{-1}}\right\}\mathbb{1}\left\{\overline{\mathcal{E}}_t\right\}\right]. \hspace{0.5cm} \text{(Cauchy-Schwartz)}$$

In the aforementioned derivation, it can be anticipated that the terms $\|\theta^* - \hat{\theta}_{t-1}\|_{V(p_t)}$ and $\|a^*\|_{V(p_t)^{-1}}$ cannot

assume significantly large values due to the following reasons:

1. $\|\theta^* - \hat{\theta}_{t-1}\|_{V(p_t)}$ - The online learning paradigm necessitates that $\hat{\theta}_{t-1}$ remains sufficiently close to θ^*.

2. $\|a^*\|_{V(p_t)^{-1}}$ - The ApproxDesign() ensures that exploration is adequately conducted in all relevant directions.

To facilitate easy understanding of the proof, we first bound $\|a^*\|_{V(p_t)^{-1}}$ by leveraging guarantees from the ApproxDesign(), which is established in detail in Lemma 5.

$$F_{31} \leq \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L} B \cdot 2^{-\ell+1} \mathbb{1}\{\mathcal{G}_1\} \mathbb{1}\left\{\varepsilon_{2,\ell} \leq \|\theta^* - \hat{\theta}_{t-1}\|_{V(p_t)} \cdot \sqrt{\frac{2}{\alpha_{\text{opt}}} \exp\left(\frac{\|\hat{\theta}_{t-1} - \theta^*\|_{V_{t-1}}^2}{\beta_{t-1}(\delta_{t-1})}\right) \cdot C_{\text{opt}} \cdot d\log(d)}\right\} \mathbb{1}\{\overline{\mathcal{E}}_t\}\right].$$

(by Lemma 5)

Next, we utilize the fact that $\mathcal{G}_1 = \left\{\|\theta^* - \hat{\theta}_{t-1}\|_{V_{t-1}}^2 \leq \beta_{t-1}^*(\delta_{t-1}), \quad \forall t \geq 1\right\}$ to simplify the analysis further as follows:

$$F_{31} \leq \mathbb{E}\left[\sum_{t=1}^{n}\sum_{\ell=1}^{L} B \cdot 2^{-\ell+1} \mathbb{1}\left\{\varepsilon_{2,\ell} \leq \|\theta^* - \hat{\theta}_{t-1}\|_{V(p_t)} \cdot \sqrt{\frac{2}{\alpha_{\text{opt}}} \exp\left(\frac{\beta_{t-1}^*(\delta_{t-1})}{\beta_{t-1}(\delta_{t-1})}\right) \cdot C_{\text{opt}} \cdot d\log(d)}\right\} \mathbb{1}\{\overline{\mathcal{E}}_t\}\right]$$

$$= \sum_{\ell=1}^{L} B \cdot 2^{-\ell+1} \mathbb{E}\left[\sum_{t=1}^{n} \mathbb{1}\left\{\varepsilon_{2,\ell} \leq \|\theta^* - \hat{\theta}_{t-1}\|_{V(p_t)} \cdot \sqrt{\frac{2}{\alpha_{\text{opt}}} \exp\left(\frac{\beta_{t-1}^*(\delta_{t-1})}{\beta_{t-1}(\delta_{t-1})}\right) \cdot C_{\text{opt}} \cdot d\log(d)}\right\} \mathbb{1}\{\overline{\mathcal{E}}_t\}\right]$$

$$= \sum_{\ell=1}^{L} B \cdot 2^{-\ell+1} \sum_{t=1}^{n} \mathbb{E}\left[\mathbb{1}\left\{\varepsilon_{2,\ell} \leq \|\theta^* - \hat{\theta}_{t-1}\|_{V(p_t)} \cdot \sqrt{\frac{2}{\alpha_{\text{opt}}} \exp\left(\frac{\beta_{t-1}^*(\delta_{t-1})}{\beta_{t-1}(\delta_{t-1})}\right) \cdot C_{\text{opt}} \cdot d\log(d)}\right\} \mathbb{1}\{\overline{\mathcal{E}}_t\}\right]$$

$$= \sum_{\ell=1}^{L} B \cdot 2^{-\ell+1} \sum_{t=1}^{n} \mathbb{E}\left[\mathbb{1}\left\{\|\theta^* - \hat{\theta}_{t-1}\|_{V(p_t)}^2 \geq \frac{\varepsilon_{2,\ell}^2}{\frac{2}{\alpha_{\text{opt}}} \exp\left(\frac{\beta_{t-1}^*(\delta_{t-1})}{\beta_{t-1}(\delta_{t-1})}\right) \cdot C_{\text{opt}} \cdot d\log(d)}\right\} \mathbb{1}\{\overline{\mathcal{E}}_t\}\right].$$

By the definition of $V(p_t)$, we have $\|\theta^* - \hat{\theta}_{t-1}\|_{V(p_t)}^2 = \mathbb{E}_{A_t \sim p_t}\left[\left(\left(\theta^* - \hat{\theta}_{t-1}\right) A_t^\top\right)^2\right]$.

$$F_{31} \leq \sum_{\ell=1}^{L} B \cdot 2^{-\ell+1} \sum_{t=1}^{n} \mathbb{E}\left[\mathbb{1}\left\{\mathbb{E}_{A_t \sim p_t}\left[\left(\left(\theta^* - \hat{\theta}_{t-1}\right) A_t^\top\right)^2\right] \geq \frac{\varepsilon_{2,\ell}^2}{\frac{2}{\alpha_{\text{opt}}} \exp\left(\frac{\beta_{t-1}^*(\delta_{t-1})}{\beta_{t-1}(\delta_{t-1})}\right) \cdot C_{\text{opt}} \cdot d\log(d)}\right\} \mathbb{1}\{\overline{\mathcal{E}}_t\}\right]$$

(Claim 2)

$$= \sum_{\ell=1}^{L} B \cdot 2^{-\ell+1} \sum_{t=1}^{n} \mathbb{E}\left[\mathbb{1}\left\{\sum_{a \in \mathcal{A}_t} p_t(a)\left[\left(\left(\theta^* - \hat{\theta}_{t-1}\right) a^\top\right)^2\right] \geq \frac{\varepsilon_{2,\ell}^2}{\frac{2}{\alpha_{\text{opt}}} \exp\left(\frac{\beta_{t-1}^*(\delta_{t-1})}{\beta_{t-1}(\delta_{t-1})}\right) \cdot C_{\text{opt}} \cdot d\log(d)}\right\}\right.$$

$$\left.\mathbb{1}\left\{\forall a \in \mathcal{A}_t, \|a\|_{V_{t-1}^{-1}}^2 \leq 1\right\}\right].$$

The event $\overline{\mathcal{E}}_t$ implies that, for all the arms in $\mathcal{A}_t$ their leverage score is bounded above by 1. Consequently, we can confidently incorporate the index function $\mathbb{1}\left\{\|a\|_{V_{t-1}^{-1}}^2 \leq 1\right\}$ within the summation $(\sum_{a \in \mathcal{A}_t})$ without affecting

the analysis.

$$F_{31} \leq \sum_{\ell=1}^{L} B \cdot 2^{-\ell+1} \sum_{t=1}^{n} \mathbb{E}\left[\mathbb{1}\left\{\sum_{a \in \mathcal{A}_t} p_t(a)\left[\left(\left(\theta^* - \hat{\theta}_{t-1}\right) a^\top\right)^2\right] \mathbb{1}\left\{\|a\|_{V_{t-1}^{-1}}^2 \leq 1\right\} \geq \frac{\varepsilon_{2,\ell}^2}{\frac{2}{\alpha_{\text{opt}}} \exp\left(\frac{\beta_{t-1}^*(\delta_{t-1})}{\beta_{t-1}(\delta_{t-1})}\right) \cdot C_{\text{opt}} \cdot d\log(d)}\right\}\right].$$

Furthermore, it is evident that

$$\sum_{a \in \mathcal{A}_t} p_t(a)\left[\left(\left(\theta^* - \hat{\theta}_{t-1}\right) a^\top\right)^2\right] \mathbb{1}\left\{\|a\|_{V_{t-1}^{-1}}^2 \leq 1\right\} = \mathbb{E}_{A_t \sim p_t}\left[\left(\left(\theta^* - \hat{\theta}_{t-1}\right) A_t^\top\right)^2 \mathbb{1}\left\{\|A_t\|_{V_{t-1}^{-1}}^2 \leq 1\right\}\right].$$

Hence,

$$F_{31} \leq \sum_{\ell=1}^{L} B \cdot 2^{-\ell+1} \sum_{t=1}^{n} \mathbb{E}\left[\mathbb{1}\left\{\mathbb{E}_{A_t \sim p_t}\left[\left(\left(\theta^* - \hat{\theta}_{t-1}\right) A_t^\top\right)^2 \mathbb{1}\left\{\|A_t\|_{V_{t-1}^{-1}}^2 \leq 1\right\}\right] \geq \frac{\varepsilon_{2,\ell}^2}{\frac{2}{\alpha_{\text{opt}}} \exp\left(\frac{\beta_{t-1}^*(\delta_{t-1})}{\beta_{t-1}(\delta_{t-1})}\right) \cdot C_{\text{opt}} \cdot d\log(d)}\right\}\right]$$

$$= \sum_{\ell=1}^{L} B \cdot 2^{-\ell+1} \sum_{t=1}^{n} \mathbb{E}\left[\mathbb{1}\left\{\mathbb{E}_{t-1}\left[\left(\left(\theta^* - \hat{\theta}_{t-1}\right) A_t^\top\right)^2 \mathbb{1}\left\{\|A_t\|_{V_{t-1}^{-1}}^2 \leq 1\right\}\right] \geq \frac{\varepsilon_{2,\ell}^2}{\frac{2}{\alpha_{\text{opt}}} \exp\left(\frac{\beta_{t-1}^*(\delta_{t-1})}{\beta_{t-1}(\delta_{t-1})}\right) \cdot C_{\text{opt}} \cdot d\log(d)}\right\}\right]$$

$$= \sum_{\ell=1}^{L} B \cdot 2^{-\ell+1} \sum_{t=1}^{n} \mathbb{P}\left(\mathbb{E}_{t-1}\left[\left(\left(\theta^* - \hat{\theta}_{t-1}\right) A_t^\top\right)^2 \mathbb{1}\left\{\|A_t\|_{V_{t-1}^{-1}}^2 \leq 1\right\}\right] \geq \frac{\varepsilon_{2,\ell}^2}{\frac{2}{\alpha_{\text{opt}}} \exp\left(\frac{\beta_{t-1}^*(\delta_{t-1})}{\beta_{t-1}(\delta_{t-1})}\right) \cdot C_{\text{opt}} \cdot d\log(d)}\right).$$

Applying Markov's inequality to the expression above, we can derive the following bound:

$$F_{31} \leq \sum_{\ell=1}^{L} B \cdot 2^{-\ell+1} \sum_{t=1}^{n} \frac{\mathbb{E}\left[\mathbb{E}_{t-1}\left[\left(\left(\theta^* - \hat{\theta}_{t-1}\right) A_t^\top\right)^2 \mathbb{1}\left\{\|A_t\|_{V_{t-1}^{-1}}^2 \leq 1\right\}\right]\right]}{\frac{\varepsilon_{2,\ell}^2}{\frac{2}{\alpha_{\text{opt}}} \exp\left(\frac{\beta_{t-1}^*(\delta_{t-1})}{\beta_{t-1}(\delta_{t-1})}\right) \cdot C_{\text{opt}} \cdot d\log(d)}} \qquad \text{(Markov's Inequality)}$$

$$\leq \sum_{\ell=1}^{L} B \cdot 2^{-\ell+1} \frac{\frac{2}{\alpha_{\text{opt}}} H_{\max} \cdot C_{\text{opt}} \cdot d\log(d)}{\varepsilon_{2,\ell}^2} \mathbb{E}\left[\sum_{t=1}^{n}\left(\left(\theta^* - \hat{\theta}_{t-1}\right) A_t^\top\right)^2 \mathbb{1}\left\{\|A_t\|_{V_{t-1}^{-1}} \leq 1\right\}\right].$$

This section marks a pivotal moment in our analysis, as we leverage the Online Learning Equality established in Lemma 7 to derive Lemma 10. This derived lemma subsequently enables us to bound the expression $\mathbb{E}\left[\sum_{t=1}^{n}\left(\left(\theta^* - \hat{\theta}_{t-1}\right) A_t^\top\right)^2 \mathbb{1}\left\{\|A_t\|_{V_{t-1}^{-1}} \leq 1\right\}\right]$ as follows:

$$F_{31} \leq \sum_{\ell=1}^{L} B \cdot 2^{-\ell+1} \frac{\frac{2}{\alpha_{\text{opt}}} H_{\max} \cdot C_{\text{opt}} \cdot d\log(d)}{\varepsilon_{2,\ell}^2}\left(2\lambda\|\theta^*\|_2^2 + 4\sigma_*^2 d\log\left(1 + \frac{n}{d\lambda}\right)\right) \qquad \text{(by Lemma 10)}$$

$$\leq \sum_{\ell=1}^{L} B \cdot 2^{-\ell+1} \frac{\frac{2}{\alpha_{\text{opt}}} H_{\max} \cdot C_{\text{opt}} \cdot d\log(d)}{\varepsilon_{2,\ell}^2}\left(2\lambda(S_*)^2 + 4\sigma_*^2 d\log\left(1 + \frac{n}{d\lambda}\right)\right)$$

$$= \frac{512 H_{\max} \cdot C_{\text{opt}} \cdot d\log(d)}{\alpha_{\text{opt}} B \cdot 2^{-L}}\left(\frac{\lambda(S^*)^2}{2} + \sigma_*^2 d\log\left(1 + \frac{n}{d\lambda}\right)\right).. \qquad (\varepsilon_2 = \frac{B}{4})$$

This concludes the bounding of the term F_{31}.

By combining the bounds obtained for F_{31} and F_{32}, we can now establish a bound for F_3.

$$F_3 = F_{31} + F_{32}$$

$$\leq B\log(n+1) + \frac{512 H_{\max} \cdot C_{\text{opt}} \cdot d\log(d)}{\alpha_{\text{opt}} B \cdot 2^{-L}}\left(\frac{\lambda(S^*)^2}{2} + \sigma_*^2 d\log\left(1 + \frac{n}{d\lambda}\right)\right).$$

In summary, by consolidating all the bounds derived throughout our analysis, we arrive at the final regret bound articulated as follows:

$$\text{Reg}_n \le 6Bd\log\left(1 + \frac{2}{\lambda}\right)$$

$$+ \frac{1}{\alpha_{\text{emp}}} \cdot 4Bn\exp\left(-\frac{B^2}{16\varepsilon \cdot \beta_n(\delta_n)}\right) + \frac{1}{\alpha_{\text{emp}}^2}\left(\frac{12dB}{2^{-L}\varepsilon}\log\left(1 + \frac{2}{\lambda 2^{-2L}\varepsilon}\right) + 6dB\log\left(1 + \frac{2}{\lambda}\right)\right)$$

$$+ B\log(n+1) + \frac{192\beta_n^*(\delta_n)d}{B \cdot 2^{-L}}\log\left(1 + \frac{32\beta_n^*(\delta_n)}{\lambda B^2 \cdot 2^{-2L}}\right) + 6dB\log\left(1 + \frac{2}{\lambda}\right)$$

$$+ B\log(n+1) + \frac{512H_{\max} \cdot C_{\text{opt}} \cdot d\log(d)}{\alpha_{\text{opt}}B \cdot 2^{-L}}\left(\frac{\lambda(S^*)^2}{2} + \sigma_*^2 d\log\left(1 + \frac{n}{d\lambda}\right)\right)$$

$$+ \frac{12dB}{2^{-L}\varepsilon}\log\left(1 + \frac{2}{\lambda 2^{-2L}\varepsilon}\right) + 6dB\log\left(1 + \frac{2}{\lambda}\right) + n \cdot B \cdot 2^{-L} \cdot \mathbb{1}\left\{B \cdot 2^{-L} > \Delta\right\}$$

$$\le 6dB\left(4 + \frac{1}{\alpha_{\text{emp}}^2}\right)\log\left(1 + \frac{2}{\lambda}\right) + 2B\log(n+1) + \frac{12dB}{2^{-L}\varepsilon}\left(1 + \frac{1}{\alpha_{\text{emp}}^2}\right)\log\left(1 + \frac{2}{\lambda 2^{-2L}\varepsilon}\right)$$

$$+ \frac{192\beta_n^*(\delta)d}{B \cdot 2^{-L}}\log\left(1 + \frac{32\beta_n^*(\delta)}{\lambda B^2 \cdot 2^{-2L}}\right) + \frac{512H_{\max} \cdot C_{\text{opt}} \cdot d\log(d)}{\alpha_{\text{opt}}B \cdot 2^{-L}}\left(\frac{\lambda(S_*)^2}{2} + \sigma_*^2 d\log\left(1 + \frac{n}{d\lambda}\right)\right)$$

$$+ n \cdot B \cdot 2^{-L} \cdot \mathbb{1}\left\{B \cdot 2^{-L} > \Delta\right\} + \frac{1}{\alpha_{\text{emp}}} \cdot 4Bn\exp\left(-\frac{B^2}{16\varepsilon \cdot \beta_n(\delta_n)}\right).$$

Furthermore, by tuning $\varepsilon = \frac{B^2}{16\beta_n(\delta)\log n}$ we can derive the final bound as follows:

$$\text{Reg}_n \le 6dB\left(3 + \frac{1}{\alpha_{\text{emp}}^2}\right)\log\left(1 + \frac{2}{\lambda}\right) + 2B\log(n+1) + \frac{192\beta_n(\delta)\log(n)d}{2^{-L}B}\left(1 + \frac{1}{\alpha_{\text{emp}}^2}\right)\log\left(1 + \frac{32\beta_n(\delta)\log(n)}{\lambda 2^{-2L}B^2}\right)$$

$$+ \frac{192\beta_n^*(\delta)d}{B \cdot 2^{-L}}\log\left(1 + \frac{32\beta_n^*(\delta)}{\lambda B^2 \cdot 2^{-2L}}\right) + \frac{512H_{\max} \cdot C_{\text{opt}} \cdot d\log(d)}{\alpha_{\text{opt}}B \cdot 2^{-L}}\left(\frac{\lambda(S_*)^2}{2} + \sigma_*^2 d\log\left(1 + \frac{n}{d\lambda}\right)\right)$$

$$+ n \cdot B \cdot 2^{-L} \cdot \mathbb{1}\left\{B \cdot 2^{-L} > \Delta\right\} + \frac{4B}{\alpha_{\text{emp}}}.$$

This bound encapsulates the cumulative effects of the various factors we have considered, providing a comprehensive measure of the regret associated with our algorithm. The implications of this bound are significant, as they delineate the performance guarantees of our approach under the specified conditions, ultimately contributing to a deeper understanding of the theoretical foundations underpinning our work.

Proof concludes. □

C.4 Proof for augmenting the arm set by eliminating highly sub-optimal arms (version 1)

We eliminate all the arms for which $f_t(a) < \frac{1}{e}$, where $f_t(a)$ is the quantity defined in Equation (5). The title of this section may be slightly misleading. In fact, Version 1 eliminates arms that exhibit one or both of the following characteristics:

1. A large estimated sub-optimality gap, $\hat{\Delta}_{a,t}^2$ and/or

2. Arms that have already been sufficiently explored, as indicated by the bound $\|\hat{a}_t - a\|_{V_{t-1}^{-1}}^2 \leq$
   $2\left(\|\hat{a}_t\|_{V_{t-1}^{-1}}^2 + \|a\|_{V_{t-1}^{-1}}^2\right)$

This approach is intuitively appealing, as it ensures that we avoid incurring regret by assigning probability to highly sub-optimal arms. Additionally, there is no need to allocate probability to directions that have already been sufficiently explored.

Before proceeding with the proof, it is important to note that this version of the augmenting arm set is ineffective when σ_*^2 and S_* are under-specified. We require $\sigma_*^2 \leq \sigma^2$ and $S_* \leq S$.

The majority of the proof for Version 0 applies to this version as well; however, we derive a different bound for $\|a_t^*\|_{V(p_t)^{-1}}^2$ compared to the one presented in Lemma 5

$$\overline{\mathcal{A}}_{(t)} = \{a \in \mathcal{A}_t : f_t(a) \geq \frac{1}{e}\}.$$

Step 1: We prove that with high probability $\forall t$, $a_t^* \in \overline{\mathcal{A}}_{(t)}$

$$f_t(a_t^*) = \exp\left(-\frac{\hat{\Delta}_{a_t^*,t}^2}{\beta_{t-1}(\delta_{t-1})\|\hat{a}_t - a_t^*\|_{V_{t-1}^{-1}}^2}\right).$$

Furthermore,

$$
\begin{aligned}
\hat{\Delta}_{a_t^*,t} &= \langle \hat{a}_t, \hat{\theta}_{t-1} \rangle - \langle a_t^*, \hat{\theta}_{t-1} \rangle \\
&= \langle \hat{a}_t, \hat{\theta}_{t-1} \rangle - \langle \hat{a}_t, \theta^* \rangle + \langle \hat{a}_t, \theta^* \rangle - \langle a_t^*, \hat{\theta}_{t-1} \rangle \\
&\leq \langle \hat{a}_t, \hat{\theta}_{t-1} \rangle - \langle \hat{a}_t, \theta^* \rangle + \langle a_t^*, \theta^* \rangle - \langle a_t^*, \hat{\theta}_{t-1} \rangle \\
&= \langle \hat{a}_t, \hat{\theta}_{t-1} - \theta^* \rangle - \langle a_t^*, \hat{\theta}_{t-1} - \theta^* \rangle \\
&\leq \|\hat{a}_t - a_t^*\|_{V_{t-1}^{-1}} \cdot \|\hat{\theta}_{t-1} - \theta^*\|_{V_{t-1}} \qquad \text{(Cauchy-Schwartz)} \\
\hat{\Delta}_{a_t^*,t}^2 &\leq \|\hat{a}_t - a_t^*\|_{V_{t-1}^{-1}}^2 \cdot \|\hat{\theta}_{t-1} - \theta^*\|_{V_{t-1}}^2.
\end{aligned}
$$

Combining the two displays above, we obtain the following result:

$$
\begin{aligned}
f_t(a_t^*) &\geq \exp\left(-\frac{\|\hat{a}_t - a_t^*\|_{V_{t-1}^{-1}}^2 \cdot \|\hat{\theta}_{t-1} - \theta^*\|_{V_{t-1}}^2}{\beta_{t-1}(\delta_{t-1})\|\hat{a}_t - a_t^*\|_{V_{t-1}^{-1}}^2}\right) \\
&= \exp\left(-\frac{\|\hat{\theta}_{t-1} - \theta^*\|_{V_{t-1}}^2}{\beta_{t-1}(\delta_{t-1})}\right).
\end{aligned}
$$

Moreover, with a probability of at least $1 - \delta_t$,
$$\forall t, \quad \|\hat{\theta}_{t-1} - \theta^*\|_{V_{t-1}}^2 \leq \beta_{t-1}(\delta_{t-1}).$$
Hence we can conclude that, with a probability of at least $1 - \delta_t$,
$$f_t(a_t^*) \geq \exp(-1) = \frac{1}{e} \implies a_t^* \in \overline{\mathcal{A}}_{(t)}.$$

This also implies that we can apply the guarantees ApproxDesign() on a_t^*

Step 2: Bounding $\|a_t^*\|_{V(p_t)^{-1}}^2$

This proof is analogous to the one presented in Lemma 5.

$$
\begin{aligned}
V(p_t) &= \sum_{a \in \overline{\mathcal{A}}_{(t)}} p_t(a) a a^\top \\
&\succeq \frac{1}{2} \sum_{a \in \overline{\mathcal{A}}_{(t)}} p_t'(a) a a^\top \\
&= \frac{1}{2} \sum_{a \in \overline{\mathcal{A}}_{(t)}} \frac{q_t(a) f_t(a)}{\sum_{b \in \mathcal{A}_t} q_t(b) f_t(b)} a a^\top \\
&\succeq \frac{1}{2} \sum_{a \in \overline{\mathcal{A}}_{(t)}} q_t(a) f_t(a) a a^\top & \text{(by Lemma 3)} \\
&\succeq \frac{1}{2} \sum_{a \in \overline{\mathcal{A}}_{(t)}} \alpha_{\text{opt}} \cdot q_t^{\text{opt}}(a) f_t(a) a a^\top \\
&\succeq \frac{1}{2e} \sum_{a \in \overline{\mathcal{A}}_{(t)}} \alpha_{\text{opt}} \cdot q_t^{\text{opt}}(a) a a^\top & (\forall a \in \overline{\mathcal{A}}_{(t)}, \ f_t(a) \geq \tfrac{1}{e}) \\
&= \frac{\alpha_{\text{opt}}}{2e} \sum_{a \in \overline{\mathcal{A}}_{(t)}} q_t^{\text{opt}}(a) a a^T \\
&= \frac{\alpha_{\text{opt}}}{2e} V(q_t^{\text{opt}}).
\end{aligned}
$$

In conclusion, we obtain the following bound:

$$
\|a_t^*\|_{V(p_t)^{-1}}^2 \leq \frac{2e}{\alpha_{\text{opt}}} \cdot C_{\text{opt}} \cdot d \log(d).
$$

With the exception of the aforementioned two steps, the remainder of the proof closely follows that of Version 0; therefore, we omit the details.

C.5 Proof of Theorem 1

Theorem 1 (Instance-dependent bound). Under Assumptions 1, 2, and 3, with $\delta_t = \frac{1}{t+1}$, LinMED satisfies, $\forall n \geq 1$,

$$\mathbb{E}\,\mathrm{Reg}_n = O\left(\frac{1}{\Delta} d \log(n)\left(\left(\sigma^2 d \log(n) + \lambda S^2\right) \log\left(\log n\right) + \left(\sigma_*^2 d \log(n) + \lambda S_*^2\right) H_{\max}\right)\right).$$

Proof. From Lemma 1 we have

$$\mathrm{Reg}_n \leq 6dB \left(3 + \frac{1}{\alpha_{\mathrm{emp}}^2}\right) \log\left(1 + \frac{2}{\lambda}\right) + 2B \log(n+1) + \frac{192\beta_n(\delta)\log(n)d}{2^{-L}B}\left(1 + \frac{1}{\alpha_{\mathrm{emp}}^2}\right) \log\left(1 + \frac{32\beta_n(\delta)\log(n)}{\lambda 2^{-2L}B^2}\right)$$

$$+ \frac{192\beta_n^*(\delta_n)d}{B \cdot 2^{-L}} \log\left(1 + \frac{32\beta_n^*(\delta_n)}{\lambda B^2 \cdot 2^{-2L}}\right) + \frac{512 H_{\max} \cdot C_{\mathrm{opt}} \cdot d \log(d)}{\alpha_{\mathrm{opt}} B \cdot 2^{-L}}\left(\frac{\lambda(S^*)^2}{2} + \sigma_*^2 d \log\left(1 + \frac{n}{d\lambda}\right)\right)$$

$$+ n \cdot B \cdot 2^{-L} \cdot \mathbb{1}\left\{B \cdot 2^{-L} > \Delta\right\} + \frac{4B}{\alpha_{\mathrm{emp}}}.$$

Where 2^{-L} is an analysis variable we introduced, such that

$$\mathcal{D}_{t,L}(a) = \left\{\Delta_{a,t} \leq B \cdot 2^{-L}\right\}.$$

By choosing L such that $\frac{\Delta}{2} \leq B \cdot 2^{-L} \leq \Delta$, we can show that,

$$\mathbb{1}\left\{B \cdot 2^{-L} > \Delta\right\} = \mathbb{1}\left\{\Delta > \Delta\right\} = 0.$$

Hence,

$$\mathrm{Reg}_n \leq 6dB \left(3 + \frac{1}{\alpha_{\mathrm{emp}}^2}\right) \log\left(1 + \frac{2}{\lambda}\right) + 2B \log(n+1) + \frac{384\beta_n(\delta)\log(n)d}{\Delta}\left(1 + \frac{1}{\alpha_{\mathrm{emp}}^2}\right) \log\left(1 + \frac{128\beta_n(\delta)\log(n)}{\lambda \Delta^2}\right)$$

$$+ \frac{384\beta_n^*(\delta)d}{\Delta} \log\left(1 + \frac{128\beta_n^*(\delta)}{\lambda \Delta^2}\right) + \frac{1024 H_{\max} \cdot C_{\mathrm{opt}} \cdot d \log(d)}{\alpha_{\mathrm{opt}} \Delta}\left(\frac{\lambda(S_*)^2}{2} + \sigma_*^2 d \log\left(1 + \frac{n}{d\lambda}\right)\right) + \frac{4B}{\alpha_{\mathrm{emp}}}$$

$$= O\left(\frac{1}{\Delta} d \log(n)\left(\left(\sigma^2 d \log(n) + \lambda S^2\right) \log\left(\log n\right) + \left(\sigma_*^2 d \log(n) + \lambda S_*^2\right) H_{\max}\right)\right).$$

Proof concludes. □

C.6 Proof of Theorem 2

Theorem 2 (Minimax bound). Under Assumptions 1, 2, and 3, with $\delta_t = \frac{1}{t+1}$, LinMED satisfies, $\forall n \geq 1$,

$$\mathbb{E}\operatorname{Reg}_n = O\left(\sqrt{n}\left(\log^{\frac{1}{2}}(n)\left(d\sigma\log(n) + \frac{\lambda S}{\sigma}\right) + \frac{H_{\max}}{\sigma\log^{\frac{3}{2}}(n)}\left(d\sigma_*^2\log(n) + \lambda S_*^2\right)\right)\right).$$

Proof. From Lemma 1 we have

$$\operatorname{Reg}_n \leq 6dB\left(3 + \frac{1}{\alpha_{\text{emp}}^2}\right)\log\left(1 + \frac{2}{\lambda}\right) + 2B\log(n+1) + \frac{192\beta_n(\delta)\log(n)d}{2^{-L}B}\left(1 + \frac{1}{\alpha_{\text{emp}}^2}\right)\log\left(1 + \frac{32\beta_n(\delta)\log(n)}{\lambda 2^{-2L}B^2}\right)$$

$$+ \frac{192\beta_n^*(\delta_n)d}{B \cdot 2^{-L}}\log\left(1 + \frac{32\beta_n^*(\delta_n)}{\lambda B^2 \cdot 2^{-2L}}\right) + \frac{512H_{\max} \cdot C_{\text{opt}} \cdot d\log(d)}{\alpha_{\text{opt}}B \cdot 2^{-L}}\left(\frac{\lambda(S^*)^2}{2} + \sigma_*^2 d\log\left(1 + \frac{n}{d\lambda}\right)\right)$$

$$+ n \cdot B \cdot 2^{-L} \cdot \mathbb{1}\left\{B \cdot 2^{-L} > \Delta\right\} + \frac{4B}{\alpha_{\text{emp}}}.$$

Where 2^{-L} is an analysis variable we introduced, such that

$$\mathcal{D}_{t,L}(a) = \left\{\Delta_{a,t} \leq B \cdot 2^{-L}\right\}.$$

Case 1 : $n \leq 4\sigma^2\left(\frac{d}{B}\right)^2\log^3(n)$

This is a trivial case. We can show that

$$\operatorname{Reg}_n \leq n$$
$$= \sqrt{n} \cdot \sqrt{n}$$
$$\leq 2\sigma\log^{\frac{3}{2}}(n)\frac{d}{B}\sqrt{n}.$$

Case 2 : $n > 4\sigma^2\left(\frac{d}{B}\right)^2\log^3(n)$

We can set $2^{-(L+1)} \leq \frac{\sigma d\log^{\frac{3}{2}}(n)}{B\sqrt{n}} \leq 2^{-L} < \frac{1}{2}$ ($L \geq 1$ will be assured).

Also,

$$\mathbb{1}\left\{B \cdot 2^{-L} > \Delta\right\} \leq 1.$$

Hence the regret bound is

$$\operatorname{Reg}_n \leq 6dB\left(3 + \frac{1}{\alpha_{\text{emp}}^2}\right)\log\left(1 + \frac{2}{\lambda}\right) + 2B\log(n+1) + \frac{192\beta_n(\delta)\sqrt{n}}{\sigma\log^{\frac{1}{2}}(n)}\left(1 + \frac{1}{\alpha_{\text{emp}}^2}\right)\log\left(1 + \frac{32\beta_n(\delta)n}{\sigma^2\lambda d^2\log^2(n)}\right)$$

$$+ \frac{192\beta_n^*(\delta)\sqrt{n}}{\sigma\log^{\frac{3}{2}}(n)}\log\left(1 + \frac{32\beta_n^*(\delta)n}{\lambda\sigma^2 d^2\log^3(n)}\right) + \frac{512\sqrt{n}H_{\max} \cdot C_{\text{opt}}\log(d)}{\sigma\log^{\frac{3}{2}}(n)\alpha_{\text{opt}}}\left(\frac{\lambda(S_*)^2}{2} + \sigma_*^2 d\log\left(1 + \frac{n}{d\lambda}\right)\right)$$

$$+ 2\sigma d\sqrt{n}\log^{\frac{3}{2}}(n) + \frac{4B}{\alpha_{\text{emp}}}$$

$$= O\left(\sqrt{n}\left(\log^{\frac{1}{2}}(n)\left(d\sigma\log(n) + \frac{\lambda S^2}{\sigma}\right) + \frac{H_{\max}}{\sigma\log^{\frac{3}{2}}(n)}\left(d\sigma_*^2\log(n) + \lambda S_*^2\right)\right)\right).$$

Proof concludes. □

C.7 Proof of Corollary 3

Corollary 3 (Instance-dependent bound). Under Assumptions 1, 2, and 3, assuming $\sigma^2 \geq \sigma_*^2$, $S \geq S_*$ with $\lambda = \frac{\sigma^2}{S^2}$ and $\delta_t = \frac{1}{t+1}$, LinMED satisfies, $\forall n \geq 1$,

$$\mathbb{E}\operatorname{Reg}_n = O\left(\sigma^2 \frac{d^2}{\Delta} \log^2(n) \log(\log n)\right).$$

Proof. The proof of this corollary follows directly from Theorem 1 and is straightforward. Given that $\sigma^2 \geq \sigma_*^2$ and $S \geq S_*$, it follows that $H_{\max} \leq \exp(1)$, resulting in the dominance of the first term over the second. Further substitution of $\lambda = \frac{\sigma^2}{S^2}$ yields the final result. $\qquad\square$

C.8 Proof of Corollary 4

Corollary 4 (Minimax bound). Under Assumptions 1, 2, and 3, assuming $\sigma^2 \geq \sigma_*^2$, $S \geq S_*$ and with $\lambda = \frac{\sigma^2}{S^2}$ and $\delta_t = \frac{1}{t+1}$, LinMED satisfies, $\forall n \geq 1$,

$$\mathbb{E}\operatorname{Reg}_n = O\left(\sigma d\sqrt{n} \log^{\frac{3}{2}}(n)\right).$$

Proof. The proof of this corollary follows directly from Theorem 2 and is straightforward. Given that $\sigma^2 \geq \sigma_*^2$ and $S \geq S_*$, it follows that $H_{\max} \leq \exp(1)$, resulting in the dominance of the first term over the second. Further substitution of $\lambda = \frac{\sigma^2}{S^2}$ yields the final result. $\qquad\square$

C.9 Proof of Corollary 5

Corollary 5 (Minimax bound). Under Assumptions 1, 2, and 3, assuming $\sigma^2 < \sigma_*^2$, $S \geq S_*$ and with $\lambda = \frac{\sigma^2}{S^2}$ and $\delta_t = \frac{1}{t+1}$, $\forall n \geq 1$, LinMED satisfies

$$\mathbb{E}\operatorname{Reg}_n = O\left(\frac{\sigma d\sqrt{n}}{\log^{\frac{1}{2}}(n)}\left(\log^2(n) + \frac{\sigma_*^2}{\sigma^2}\exp\left(\frac{\sigma_*^2}{\sigma^2}\right)\right)\right).$$

Proof. We can bound $\beta_t^*(\delta_t)$ as follows :

$$\beta_t^*(\delta_t) = \left(\sigma_*\sqrt{\log\left(\frac{\det V_t}{\det V_0}\right) + 2\log\frac{1}{\delta_t}} + \sqrt{\lambda}S_*\right)^2$$

$$\leq \left(\sigma_*\sqrt{\log\left(\frac{\det V_t}{\det V_0}\right) + 2\log\frac{1}{\delta_t}} + \sqrt{\lambda}S\right)^2$$

$$\leq \left(\sigma_*\sqrt{\log\left(\frac{\det V_t}{\det V_0}\right) + 2\log\frac{1}{\delta_t}} + \sigma\right)^2$$

$$= \sigma_*^2\left(\sqrt{\log\left(\frac{\det V_t}{\det V_0}\right) + 2\log\frac{1}{\delta_t}} + \frac{\sigma}{\sigma_*}\right)^2$$

$$\leq \sigma_*^2\left(\sqrt{\log\left(\frac{\det V_t}{\det V_0}\right) + 2\log\frac{1}{\delta_t}} + 1\right)^2.$$

Similarly,

$$\beta_t(\delta_t) = \sigma^2\left(\sqrt{\log\left(\frac{\det V_t}{\det V_0}\right) + 2\log\frac{1}{\delta_t}} + 1\right)^2.$$

Hence, we can conclude that,

$$H_{\max} \leq \exp\left(\frac{\sigma_*^2}{\sigma^2}\right).$$

By bounding $H_{\max}$ with the aforementioned quantity and S_* with S, and subsequently substituting $\lambda = \frac{\sigma^2}{S^2}$ into Theorem 2, the final results are obtained. $\qquad\square$

D LEMMATA

Lemma 2. Let $f_t(a)$ be the quantity defined in Equation (5) where $\hat{a}_t \neq a$. Then $\forall t > 1$,

$$f_t(a) \leq \exp\left(-\frac{\hat{\Delta}_{a,t}^2}{2\beta_{t-1}(\delta_{t-1})\left(\|\hat{a}_t\|_{V_{t-1}^{-1}}^2 + \|a\|_{V_{t-1}^{-1}}^2\right)}\right).$$

Proof.

$$f_t(a) = \exp\left(-\frac{\hat{\Delta}_{a,t}^2}{\beta_{t-1}(\delta_{t-1})\|\hat{a}_t - a\|_{V_{t-1}^{-1}}^2}\right)$$

$$\leq \exp\left(-\frac{\hat{\Delta}_{a,t}^2}{\beta_{t-1}(\delta_{t-1})\left(\|\hat{a}_t\|_{V_{t-1}^{-1}} + \|a\|_{V_{t-1}^{-1}}\right)^2}\right) \qquad \text{(triangle inequality)}$$

$$\leq \exp\left(-\frac{\hat{\Delta}_{a,t}^2}{2\beta_{t-1}(\delta_{t-1})\left(\|\hat{a}_t\|_{V_{t-1}^{-1}}^2 + \|a\|_{V_{t-1}^{-1}}^2\right)}\right). \qquad \text{(AM-GM)}$$

$\square$

Lemma 3. Let $f_t(a)$ be the quantity defined in Equation (5) where $\hat{a}_t \neq a$ and $q_t(a)$ be the quantity defined in Equation (6). Then $\forall t > 1$,

$$1 \geq \sum_{b \in \mathcal{A}_t} q_t(b)f_t(b) \geq \alpha_{\text{emp}}.$$

Proof.

$$\sum_{b \in \mathcal{A}_t} q_t(b)f_t(b) \geq q_t(\hat{a}_t)f_t(\hat{a}_t).$$

Furthermore, $f_t(\hat{a}_t) = \exp(0) = 1$ because $\hat{\Delta}_{\hat{a}_t,t} = 0$, which leads to

$$\sum_{b \in \mathcal{A}_t} q_t(b)f_t(b) \geq q_t(\hat{a}_t)$$

$$\geq \alpha_{\text{emp}}\, \mathbb{1}\{\hat{a}_t = \hat{a}_t\}$$

$$= \alpha_{\text{emp}}.$$

Hence,

$$\sum_{b \in \mathcal{A}_t} q_t(b)f_t(b) \geq \alpha_{\text{emp}}.$$

Also, note that

$$\sum_{b \in \mathcal{A}_t} q_t(b)f_t(b) \leq \sum_{b \in \mathcal{A}_t} q_t(b) \leq 1. \qquad \text{(Because, } f_t(b) \leq 1,\ \forall b)$$

$\square$

Lemma 4. In the context of the LinMED algorithm, $\forall t > 1$, the probability of choosing the empirical best arm by LinMED algorithm satisfies $\mathbb{P}(A_t = \hat{a}_t) \geq \alpha_{\text{emp}}$.

Proof.

$$\mathbb{P}(A_t = \hat{a}_t) = p_t(\hat{a}_t)$$

$$= q_t(\hat{a}_t) \cdot \frac{f_t(\hat{a}_t)}{\sum_{b \in \mathcal{A}} q_t(b)f_t(b)}$$

$$\geq q_t(\hat{a}_t) \cdot f_t(\hat{a}_t) \qquad \text{(by lemma 3)}$$

$$= q_t(\hat{a}_t)$$

$$\geq \alpha_{\text{emp}} \cdot \mathbb{1}\{\hat{a}_t = \hat{a}_t\}$$
$$= \alpha_{\text{emp}}.$$

$\square$

Lemma 5. In the context of the LinMED algorithm, we have

$$\|a_t^*\|_{V(p_t)^{-1}}^2 \leq \frac{2}{\alpha_{\text{opt}}} \exp\left(\frac{\|\hat{\theta}_{t-1} - \theta^*\|_{V_{t-1}}^2}{\beta_{t-1}(\delta_{t-1})}\right) \cdot C_{\text{opt}} \cdot d\log(d).$$

where a_t^* is true best arm at time t.

Proof.

$$V(p_t) = \sum_{a \in \mathcal{A}_t} p_t(a) a a^\top$$
$$\succeq \frac{1}{2} \sum_{a \in \mathcal{A}_t} p_t'(a) a a^\top$$
$$= \frac{1}{2} \sum_{a \in \mathcal{A}_t} \frac{q_t(a) f_t(a)}{\sum_{b \in \mathcal{A}_t} q_t(b) f_t(b)} a a^\top$$
$$\succeq \frac{1}{2} \sum_{a \in \mathcal{A}_t} q_t(a) f_t(a) a a^\top \qquad \text{(by Lemma 3)}$$
$$\succeq \frac{1}{2} \sum_{a \in \mathcal{A}_t} \alpha_{\text{opt}} \cdot q_t^{\text{opt}}(a) f_t(a) a a^\top$$
$$= \frac{\alpha_{\text{opt}}}{2} \sum_{a \in \mathcal{A}_t} q_t^{\text{opt}}(a) \left(\sqrt{f_t(a)}a\right)\left(\sqrt{f_t(a)}a\right)^\top$$
$$= \frac{\alpha_{\text{opt}}}{2} \sum_{a \in \mathcal{A}_t} q_t^{\text{opt}}(a) (\bar{a}_{(t)})(\bar{a}_{(t)})^\top$$
$$= \frac{\alpha_{\text{opt}}}{2} \overline{V}(q_t^{\text{opt}}).$$

Here, note that both $V(p_t)$ and $\overline{V}(q_t^{\text{opt}})$ are invertible.

$$\|a_t^*\|_{V(p_t)^{-1}}^2 \leq \frac{2}{\alpha_{\text{opt}}} \|a_t^*\|_{\overline{V}(q_t^{\text{opt}})^{-1}}^2$$
$$\leq \frac{2}{\alpha_{\text{opt}}} \frac{1}{f_t(a_t^*)} \|\overline{a^*}_t\|_{\overline{V}(q_t^{\text{opt}})^{-1}}^2$$
$$\leq \frac{2}{\alpha_{\text{opt}}} \frac{1}{f_t(a_t^*)} C_{\text{opt}} \cdot d\log(d). \qquad \text{(Assumption 2)}$$

$$\frac{1}{f_t(a_t^*)} = \exp\left(\frac{\hat{\Delta}_{a_t^*,t}^2}{\beta_{t-1}(\delta_{t-1})\|\hat{a}_t - a_t^*\|_{V_{t-1}^{-1}}^2}\right).$$

$$\hat{\Delta}_{a_t^*,t} = \langle \hat{a}_t, \hat{\theta}_{t-1} \rangle - \langle a_t^*, \hat{\theta}_{t-1} \rangle$$
$$= \langle \hat{a}_t, \hat{\theta}_{t-1} \rangle - \langle \hat{a}_t, \theta^* \rangle + \langle \hat{a}_t, \theta^* \rangle - \langle a_t^*, \hat{\theta}_{t-1} \rangle$$
$$\leq \langle \hat{a}_t, \hat{\theta}_{t-1} \rangle - \langle \hat{a}_t, \theta^* \rangle + \langle a_t^*, \theta^* \rangle - \langle a_t^*, \hat{\theta}_{t-1} \rangle$$
$$= \langle \hat{a}_t, \hat{\theta}_{t-1} - \theta^* \rangle - \langle a_t^*, \hat{\theta}_{t-1} - \theta^* \rangle$$
$$\leq \|\hat{a}_t - a_t^*\|_{V_{t-1}^{-1}} \cdot \|\hat{\theta}_{t-1} - \theta^*\|_{V_{t-1}} \qquad \text{(Cauchy-Schwartz)}$$
$$\hat{\Delta}_{a_t^*,t}^2 \leq \|\hat{a}_t - a_t^*\|_{V_{t-1}^{-1}}^2 \cdot \|\hat{\theta}_{t-1} - \theta^*\|_{V_{t-1}}^2.$$

$$\|a_t^*\|_{V(p_t)^{-1}}^2 \leq \frac{2}{\alpha_{\text{opt}}} \exp\left(\frac{\|\hat{a}_t - a_t^*\|_{V_{t-1}^{-1}}^2 \cdot \|\hat{\theta}_{t-1} - \theta^*\|_{V_{t-1}^{-1}}^2}{\beta_{t-1}(\delta_{t-1})\|\hat{a}_t - a_t^*\|_{V_{t-1}^{-1}}^2}\right) \cdot C_{\text{opt}} \cdot d \log(d).$$

$\square$

Claim 1. In the context of the LinMED algorithm, we have

$$\mathcal{U}_{t-1,\ell}(a) = \left\{\|a\|_{V_{t-1}^{-1}}^2 \geq \varepsilon_\ell\right\}.$$

Then,

$$\mathbb{E}\left[\sum_{t=1}^n \mathbb{1}\left\{\overline{\mathcal{U}}_{t,\ell}(\hat{a}_t)\right\}\right] \leq \frac{1}{\alpha_{\text{emp}}} \mathbb{E}\left[\sum_{t=1}^n \mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(A_t)\right\}\right].$$

Proof.

$$\mathbb{E}\left[\sum_{t=1}^n \mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(\hat{a}_t)\right\}\right] = \frac{1}{\alpha_{\text{emp}}} \mathbb{E}\left[\sum_{t=1}^n \alpha_{\text{emp}} \cdot \mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(\hat{a}_t)\right\}\right]$$

$$\leq \frac{1}{\alpha_{\text{emp}}} \mathbb{E}\left[\sum_{t=1}^n \mathbb{P}\left(A_t = \hat{a}_t\right) \cdot \mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(\hat{a}_t)\right\}\right] \qquad \text{(by lemma 4)}$$

$$= \frac{1}{\alpha_{\text{emp}}} \mathbb{E}\left[\sum_{t=1}^n \mathbb{E}_{t-1}\left[\mathbb{1}\left\{A_t = \hat{a}_t\right\}\right] \cdot \mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(\hat{a}_t)\right\}\right]$$

$$= \frac{1}{\alpha_{\text{emp}}} \mathbb{E}\left[\sum_{t=1}^n \mathbb{1}\left\{A_t = \hat{a}_t\right\} \cdot \mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(\hat{a}_t)\right\}\right] \qquad \text{(tower rule)}$$

$$= \frac{1}{\alpha_{\text{emp}}} \mathbb{E}\left[\sum_{t=1}^n \mathbb{1}\left\{\overline{\mathcal{U}}_{t-1,\ell}(A_t)\right\}\right].$$

$\square$

Lemma 6 (from Lemma 7.1 A Modern Introduction to Online Learning). Let $\Theta \subseteq \mathbb{R}^d$ be closed and non-empty. Denote by $F_t(\theta) = \psi_t(\theta) + \sum_{s=1}^{t-1} \ell_s(\theta)$. Assume that $\arg\min_{\theta \in \Theta} F_t(\theta)$ is not empty and set $\hat{\theta}_{t-1} \in \arg\min_{\theta \in \Theta} F_t(\theta)$. Then for any $\theta^* \in \mathbb{R}^d$, we have

$$\sum_{t=1}^n \left(\ell_t(\hat{\theta}_{t-1}) - \ell_t(\theta^*)\right) = \psi_{n+1}(\theta^*) - \min_{\theta \in \Theta} \psi_1(\theta) + \sum_{t=1}^n \left[F_t(\hat{\theta}_{t-1}) - F_{t+1}(\hat{\theta}_t) + \ell_t(\hat{\theta}_{t-1})\right]$$

$$+ F_{n+1}(\hat{\theta}_n) - F_{n+1}(\theta^*).$$

Lemma 7. Let $\Theta \subseteq \mathbb{R}^d$ be closed and non-empty, $\ell_t(\theta) = \frac{1}{2}\left(A_t^\top \theta - y_t\right)^2$ and $F_t(\theta) = \lambda\|\theta\|_2^2 + \sum_{s=1}^{t-1} \ell_s(\theta)$. Assume that $\arg\min_{\theta \in \Theta} F_t(\theta)$ is not empty and set $\hat{\theta}_{t-1} \in \arg\min_{\theta \in \Theta} F_t(\theta)$. Then for any $\theta^* \in \mathbb{R}^d$, we have

$$\sum_{t=1}^n \left(\ell_t(\hat{\theta}_{t-1}) - \ell_t(\theta^*)\right) = \frac{\lambda}{2}\|\theta^*\|_2^2 + \sum_{t=1}^n \ell_t(\hat{\theta}_{t-1})\|A_t\|_{V_t^{-1}}^2 - \frac{1}{2}\|\hat{\theta}_n - \theta^*\|_{V_n}^2.$$

Proof. By Lemma 6, we have

$$\sum_{t=1}^n \left(\ell_t(\hat{\theta}_{t-1}) - \ell_t(\theta^*)\right) = \psi_{n+1}(\theta^*) - \min_{\theta \in \Theta} \psi_1(\theta) + \sum_{t=1}^n \left[F_t(\hat{\theta}_{t-1}) - F_{t+1}(\hat{\theta}_t) + \ell_t(\hat{\theta}_{t-1})\right]$$

$$+ F_{n+1}(\hat{\theta}_n) - F_{n+1}(\theta^*).$$

$$\nabla \ell_t(\theta) = \left(A_t^\top \theta - y_t\right) \cdot A_t = \sqrt{2\ell_t(\theta)} \cdot A_t$$

$$\nabla^2 \ell_t(\theta) = A_t A_t^\top.$$

$$\nabla F_t(\theta) = \lambda\theta + \sum_{s=1}^{t-1}\left(A_s^\top\theta - y_s\right)\cdot A_s.$$

$$\nabla^2 F_t(\theta) = \lambda + \sum_{s=1}^{t-1}A_s A_s^\top = V_{t-1}.$$

Using the Taylor's theorem for a quadratic polynomial,

$$F_{n+1}(\theta^*) = F_{n+1}(\hat\theta_n) + \left(\nabla F_{n+1}(\hat\theta_n)\right)^\top\left(\theta^* - \hat\theta_n\right) + \frac{1}{2}\left(\theta^* - \hat\theta_n\right)^T\nabla^2 F_{n+1}(\hat\theta_n)\left(\theta^* - \hat\theta_n\right)$$

$$= F_{n+1}(\hat\theta_n) + 0 + \frac{1}{2}\left(\theta^* - \hat\theta_n\right)^T V_n\left(\theta^* - \hat\theta_n\right)$$

(second term is 0 by the optimality condition)

$$F_{n+1}(\hat\theta_n) - F_{n+1}(\theta^*) = -\frac{1}{2}\|\hat\theta_n - \theta^*\|_{V_n}^2.$$

Then,

$$\sum_{t=1}^{n}\left[F_t(\hat\theta_{t-1}) - F_{t+1}(\hat\theta_t) + \ell_t(\hat\theta_{t-1})\right] = \sum_{t=1}^{n}\left[F_{t+1}(\hat\theta_{t-1}) - F_{t+1}(\hat\theta_t)\right]$$

$$= \sum_{t=1}^{n}\frac{1}{2}\|\hat\theta_t - \hat\theta_{t-1}\|_{V_t}^2.$$

Similarly,

$$\sum_{t=1}^{n}\left[F_t(\hat\theta_{t-1}) - F_{t+1}(\hat\theta_t) + \ell_t(\hat\theta_{t-1})\right] = \sum_{t=1}^{n}\left[F_t(\hat\theta_{t-1}) - F_t(\hat\theta_t) + \ell_t(\hat\theta_{t-1}) - \ell_t(\hat\theta_t)\right]$$

$$= \sum_{t=1}^{n}\left[-\left(F_t(\hat\theta_t) - F_t(\hat\theta_{t-1})\right) + \ell_t(\hat\theta_{t-1}) - \ell_t(\hat\theta_t)\right]$$

$$= \sum_{t=1}^{n}-\frac{1}{2}\|\hat\theta_t - \hat\theta_{t-1}\|_{V_{t-1}}^2 - \left(\ell_t(\hat\theta_t) - \ell_t(\hat\theta_{t-1})\right)$$

$$= \sum_{t=1}^{n}-\frac{1}{2}\|\hat\theta_t - \hat\theta_{t-1}\|_{V_{t-1}}^2 - \left(\left(\hat\theta_t - \hat\theta_{t-1}\right)^\top\nabla\ell_t(\hat\theta_{t-1})\right.$$

$$\left.+ \frac{1}{2}\left(\hat\theta_t - \hat\theta_{t-1}\right)^\top\nabla^2\ell_t(\hat\theta_{t-1})\left(\hat\theta_t - \hat\theta_{t-1}\right)\right)$$

$$= \sum_{t=1}^{n}-\frac{1}{2}\|\hat\theta_t - \hat\theta_{t-1}\|_{V_{t-1}}^2 - \left(\left(\hat\theta_t - \hat\theta_{t-1}\right)^\top\nabla\ell_t(\hat\theta_{t-1})\right.$$

$$\left.+ \frac{1}{2}\left(\hat\theta_t - \hat\theta_{t-1}\right)^\top A_t A_t^\top\left(\hat\theta_t - \hat\theta_{t-1}\right)\right)$$

$$= \sum_{t=1}^{n}-\frac{1}{2}\|\hat\theta_t - \hat\theta_{t-1}\|_{V_t}^2 - \left(\hat\theta_t - \hat\theta_{t-1}\right)^\top\nabla\ell_t(\hat\theta_{t-1}).$$

Hence,

$$\|\hat\theta_t - \hat\theta_{t-1}\|_{V_t}^2 = \left(\hat\theta_{t-1} - \hat\theta_t\right)^\top\nabla\ell_t(\hat\theta_{t-1})$$

$$\iff \left(\hat\theta_{t-1} - \hat\theta_t\right)^\top V_t\left(\hat\theta_{t-1} - \hat\theta_t\right) = \left(\hat\theta_{t-1} - \hat\theta_t\right)^\top\nabla\ell_t(\hat\theta_{t-1})$$

$$\implies \hat\theta_{t-1} - \hat\theta_t = V_t^{-1}\nabla\ell_t(\hat\theta_{t-1})$$

$$= \sqrt{2\ell_t(\hat\theta_{t-1})}V_t^{-1}A_t$$

$$\frac{1}{2}\|\hat\theta_t - \hat\theta_{t-1}\|_{V_t}^2 = \ell_t(\hat\theta_{t-1})\|V_t^{-1}A_t\|_{V_t}^2$$

$$= \ell_t(\hat{\theta}_{t-1}) \|A_t\|^2_{V_t^{-1}}.$$

Finally, we have

$$\sum_{t=1}^{n} \left[F_t(\hat{\theta}_{t-1}) - F_{t+1}(\hat{\theta}_t) + \ell_t(\hat{\theta}_{t-1}) \right] = \ell_t(\hat{\theta}_{t-1}) \|A_t\|^2_{V_t^{-1}}.$$

Also, trivially we have

$$\min_{\theta \in \Theta} \psi_1(\theta) = 0.$$

Putting everything together, we have

$$\sum_{t=1}^{n} \left(\ell_t(\hat{\theta}_{t-1}) - \ell_t(\theta^*) \right) = \frac{\lambda}{2} \|\theta^*\|^2_2 + \sum_{t=1}^{n} \ell_t(\hat{\theta}_{t-1}) \|A_t\|^2_{V_t^{-1}} - \frac{1}{2} \|\hat{\theta}_n - \theta^*\|^2_{V_n}.$$

$$\square$$

Lemma 8. In the context of the LinMED algorithm, we have

$$\mathbb{E} \left[\sum_{t=1}^{n} \left(A_t^\top (\hat{\theta}_{t-1} - \theta^*) \right)^2 \left(1 - \|A_t\|^2_{V_t^{-1}} \right) \right] \leq \lambda \|\theta^*\|^2_2 + 2\sigma_*^2 d \log \left(1 + \frac{n}{d\lambda} \right).$$

where σ_*^2 is sub-gaussian parameter of the noise.

Proof. Let $r_t = A_t^\top (\hat{\theta}_{t-1} - \theta^*)$, $D_t = \|A_t\|^2_{V_t^{-1}}$.

$$\begin{aligned}
\ell_t(\hat{\theta}_{t-1}) - \ell_t(\theta^*) &= \frac{1}{2} \left(\left(A_t^T \hat{\theta}_{t-1} - y_t \right)^2 - \left(A_t^T \theta^* - y_t \right)^2 \right) \\
&= \frac{1}{2} \left(\left(A_t^T \left(\hat{\theta}_{t-1} - \theta^* \right) - \eta_t \right)^2 - \eta_t^2 \right) &\text{(because } y_t = A_t^\top \theta^* + \eta_t) \\
&= \frac{1}{2} \left(A_t^T \left(\hat{\theta}_{t-1} - \theta^* \right) \right)^2 - A_t^T \left(\hat{\theta}_{t-1} - \theta^* \right) \cdot \eta_t \\
&= \frac{1}{2} r_t^2 - \eta_t r_t.
\end{aligned}$$

Hence,

$$\sum_{t=1}^{n} \left(\ell_t(\hat{\theta}_{t-1}) - \ell_t(\theta^*) \right) = \frac{1}{2} \sum_{t=1}^{n} r_t^2 - \sum_{t=1}^{n} \eta_t r_t.$$

By Lemma 7, we have

$$\begin{aligned}
\sum_{t=1}^{n} \left(\ell_t(\hat{\theta}_{t-1}) - \ell_t(\theta^*) \right) &= \frac{\lambda}{2} \|\theta^*\|^2_2 + \sum_{t=1}^{n} \ell_t(\hat{\theta}_{t-1}) \|A_t\|_{V_t^{-1}} - \frac{1}{2} \|\hat{\theta}_n - \theta^*\|^2_{V_n} \\
&\leq \frac{\lambda}{2} \|\theta^*\|^2_2 + \sum_{t=1}^{n} \ell_t(\hat{\theta}_{t-1}) \|A_t\|^2_{V_t^{-1}}.
\end{aligned}$$

Hence,

$$\frac{1}{2} \sum_{t=1}^{n} r_t^2 - \sum_{t=1}^{n} \eta_t r_t \leq \frac{\lambda}{2} \|\theta^*\|^2_2 + \sum_{t=1}^{n} \ell_t(\hat{\theta}_{t-1}) \|A_t\|^2_{V_t^{-1}}$$

$$\frac{1}{2} \sum_{t=1}^{n} r_t^2 - \sum_{t=1}^{n} \eta_t r_t \leq \frac{\lambda}{2} \|\theta^*\|^2_2 + \sum_{t=1}^{n} \ell_t(\hat{\theta}_{t-1}) D_t.$$

Also note that $\ell_t(\hat{\theta}_{t-1})$ can be expanded as follows,

$$\begin{aligned}
\ell_t(\hat{\theta}_{t-1}) &= \frac{1}{2} \left(A_t^\top \hat{\theta}_{t-1} - y_t \right)^2 \\
&= \frac{1}{2} \left(A_t^\top \hat{\theta}_{t-1} - A_t^\top \theta^* - \eta_t \right)^2
\end{aligned}$$

$$= \frac{1}{2} \left(A_t^\top \left(\hat{\theta}_{t-1} - \theta^* \right) - \eta_t \right)^2$$

$$= \frac{1}{2} \left(r_t - \eta_t \right)^2$$

$$= \frac{1}{2} \left(r_t^2 - 2r_t\eta_t + \eta_t^2 \right).$$

Hence,

$$\frac{1}{2} \sum_{t=1}^n r_t^2 - \sum_{t=1}^n \eta_t r_t \le \frac{\lambda}{2} \|\theta^*\|_2^2 + \sum_{t=1}^n \frac{1}{2} \left(r_t^2 - 2r_t\eta_t + \eta_t^2 \right) D_t$$

$$\frac{1}{2} \sum_{t=1}^n r_t^2 \left(1 - D_t \right) \le \frac{\lambda}{2} \|\theta^*\|_2^2 + \frac{1}{2} \sum_{t=1}^n \eta_t^2 D_t + \frac{1}{2} \sum_{t=1}^n r_t\eta_t \left(1 - D_t \right).$$

We can take expectation both sides,

$$\mathbb{E}\left[\sum_{t=1}^n r_t^2 \left(1 - D_t \right) \right] \le \lambda \|\theta^*\|_2^2 + \mathbb{E}\left[\sum_{t=1}^n \eta_t^2 D_t \right] + \mathbb{E}\left[\sum_{t=1}^n r_t\eta_t \left(1 - D_t \right) \right]$$

$$= \lambda \|\theta^*\|_2^2 + \mathbb{E}\left[\sum_{t=1}^n \mathbb{E}_{t-1}\left[\eta_t^2 \right] D_t \right] + \mathbb{E}\left[\sum_{t=1}^n r_t \, \mathbb{E}_{t-1}\left[\eta_t \right] \left(1 - D_t \right) \right]$$

$$= \lambda \|\theta^*\|_2^2 + \sigma_*^2 \, \mathbb{E}\left[\sum_{t=1}^n D_t \right] + 0$$

$$= \lambda \|\theta^*\|_2^2 + \sigma_*^2 \, \mathbb{E}\left[\sum_{t=1}^n \|A_t\|_{V_t^{-1}}^2 \right]$$

$$\le \lambda \|\theta^*\|_2^2 + 2d\sigma_*^2 \log\left(1 + \frac{n}{d\lambda} \right). \qquad \text{(Lemma 12)}$$

Thus proved. $\qquad\qquad\square$

Lemma 9. In the context of the LinMED algorithm, we have

$$\|A_t\|_{V_{t-1}^{-1}}^2 \le 1 \iff \|A_t\|_{V_t^{-1}}^2 \le \frac{1}{2}.$$

Proof. Sherman-Morrison formula says,

$$\left(A + uv^\top \right)^{-1} = A^{-1} - \frac{A^{-1}uv^\top A^{-1}}{1 + v^\top A^{-1}u}.$$

replace $A = V_{t-1}$ and $u = v = A_t$,

$$\left(V_{t-1} + A_t A_t^\top \right)^{-1} = V_{t-1}^{-1} - \frac{V_{t-1}^{-1} A_t A_t^\top V_{t-1}^{-1}}{1 + A_t^\top V_{t-1}^{-1} A_t}$$

$$V_t^{-1} = V_{t-1}^{-1} - \frac{V_{t-1}^{-1} A_t A_t^\top V_{t-1}^{-1}}{1 + \|A_t\|_{V_{t-1}^{-1}}^2}$$

$$A_t^\top V_t^{-1} A_t = A_t^\top V_{t-1}^{-1} A_t - \frac{A_t^\top V_{t-1}^{-1} A_t A_t^\top V_{t-1}^{-1} A_t}{1 + \|A_t\|_{V_{t-1}^{-1}}^2}$$

$$\|A_t\|_{V_t^{-1}}^2 = \|A_t\|_{V_{t-1}^{-1}}^2 - \frac{\|A_t\|_{V_{t-1}^{-1}}^4}{1 + \|A_t\|_{V_{t-1}^{-1}}^2}$$

$$= \|A_t\|_{V_{t-1}^{-1}}^2 \left(1 - \frac{\|A_t\|_{V_{t-1}^{-1}}^2}{1 + \|A_t\|_{V_{t-1}^{-1}}^2} \right)$$

$$= \|A_t\|_{V_{t-1}^{-1}}^2 \left(\frac{1}{1 + \|A_t\|_{V_{t-1}^{-1}}^2} \right)$$

$$= 1 - \frac{1}{1 + \|A_t\|_{V_{t-1}^{-1}}^2}.$$

If $\|A_t\|_{V_{t-1}^{-1}}^2 \leq 1$,

$$\|A_t\|_{V_t^{-1}}^2 \leq 1 - \frac{1}{2} \qquad\qquad (\ \|A_t\|_{V_{t-1}^{-1}}^2 \leq 1)$$

$$= \frac{1}{2}.$$

Hence,

$$\|A_t\|_{V_{t-1}^{-1}}^2 \leq 1 \implies \|A_t\|_{V_t^{-1}}^2 \leq \frac{1}{2}.$$

If $\|A_t\|_{V_t^{-1}}^2 \leq \frac{1}{2}$

$$1 - \frac{1}{1 + \|A_t\|_{V_{t-1}^{-1}}^2} \leq \frac{1}{2}$$

$$\|A_t\|_{V_{t-1}^{-1}}^2 \leq 1.$$

Hence,

$$\|A_t\|_{V_t^{-1}}^2 \leq \frac{1}{2} \implies \|A_t\|_{V_{t-1}^{-1}}^2 \leq 1.$$

Thus proved. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Lemma 10. In the context of the LinMED algorithm, we have

$$\mathbb{E}\left[\sum_{t=1}^{n} \left(A_t^\top (\hat{\theta}_{t-1} - \theta^*) \right)^2 \mathbb{1}\left\{ \|A_t\|_{V_{t-1}^{-1}}^2 \leq 1 \right\} \right] \leq 2\lambda \|\theta^*\|_2^2 + 4\sigma_*^2 d \log \left(1 + \frac{n}{d\lambda} \right).$$

where σ_*^2 is sub-gaussian parameter of the noise.

Proof.

$$\|A_t\|_{V_{t-1}^{-1}}^2 \leq 1 \iff \|A_t\|_{V_t^{-1}}^2 \leq \frac{1}{2}. \qquad\qquad \text{(by Lemma 9)}$$

Hence,

$$\mathbb{1}\left\{ \|A_t\|_{V_{t-1}^{-1}}^2 \leq 1 \right\} = \mathbb{1}\left\{ \|A_t\|_{V_t^{-1}}^2 \leq \frac{1}{2} \right\}.$$

$$\mathbb{E}\left[\sum_{t=1}^{n} \left(A_t^\top (\hat{\theta}_{t-1} - \theta^*) \right)^2 \left(1 - \|A_t\|_{V_t^{-1}}^2 \right) \right]$$

$$\geq \mathbb{E}\left[\sum_{t=1}^{n} \left(A_t^\top (\hat{\theta}_{t-1} - \theta^*) \right)^2 \left(1 - \|A_t\|_{V_t^{-1}}^2 \right) \mathbb{1}\left\{ \|A_t\|_{V_{t-1}^{-1}}^2 \leq 1 \right\} \mathbb{1}\left\{ \|A_t\|_{V_{t-1}^{-1}}^2 \leq 1 \right\} \right]$$

$$= \mathbb{E}\left[\sum_{t=1}^{n} \left(A_t^\top (\hat{\theta}_{t-1} - \theta^*) \right)^2 \left(1 - \|A_t\|_{V_t^{-1}}^2 \right) \mathbb{1}\left\{ \|A_t\|_{V_t^{-1}}^2 \leq \frac{1}{2} \right\} \mathbb{1}\left\{ \|A_t\|_{V_{t-1}^{-1}}^2 \leq 1 \right\} \right]$$

$$\geq \frac{1}{2} \mathbb{E}\left[\sum_{t=1}^{n} \left(A_t^\top (\hat{\theta}_{t-1} - \theta^*) \right)^2 \mathbb{1}\left\{ \|A_t\|_{V_{t-1}^{-1}}^2 \leq 1 \right\} \right].$$

Hence,

$$\mathbb{E}\left[\sum_{t=1}^{n}\left(A_t^\top(\hat{\theta}_{t-1}-\theta^*)\right)^2 \mathbb{1}\left\{\|A_t\|_{V_{t-1}^{-1}}^2 \leq 1\right\}\right] \leq 2\,\mathbb{E}\left[\sum_{t=1}^{n}\left(A_t^\top(\hat{\theta}_{t-1}-\theta^*)\right)^2\left(1-\|A_t\|_{V_t^{-1}}^2\right)\right]$$

$$\leq 2\lambda\|\theta^*\|_2^2 + 4\sigma_*^2 d\log\left(1+\frac{n}{d\lambda}\right). \qquad \text{(by Lemma 8)}$$

$\square$

Claim 2. In the context of the LinMED algorithm, we have

$$\sum_{t=1}^{n}\|\theta^*-\hat{\theta}_{t-1}\|_{V(p_t)}^2 = \mathbb{E}_{A_t\sim p_t}\left[\left(\left(\theta^*-\hat{\theta}_{t-1}\right)A_t^\top\right)^2\right].$$

Proof.

$$\begin{aligned}
\|\theta^*-\hat{\theta}_{t-1}\|_{V(p_t)}^2 &= \|\theta^*-\hat{\theta}_{t-1}\|_{V(p_t)}^2 \\
&= \left(\theta^*-\hat{\theta}_{t-1}\right)^\top V(p_t)\left(\theta^*-\hat{\theta}_{t-1}\right) \\
&= \left(\theta^*-\hat{\theta}_{t-1}\right)^\top \sum_{a\in\mathcal{A}_t} p_t(a)aa^\top\left(\theta^*-\hat{\theta}_{t-1}\right) \\
&= \sum_{a\in\mathcal{A}_t} p_t(a)\left(\left(\theta^*-\hat{\theta}_{t-1}\right)a^\top\right)\left(\left(\theta^*-\hat{\theta}_{t-1}\right)a^\top\right)^\top \mathbb{1}\left\{\mathcal{U}_{t-1}(A_t)\right\} \\
&= \sum_{a\in\mathcal{A}_t} p_t(a)\left(\left(\theta^*-\hat{\theta}_{t-1}\right)a^\top\right)^2 \\
&= \mathbb{E}_{A_t\sim p_t}\left[\left(\left(\theta^*-\hat{\theta}_{t-1}\right)A_t^\top\right)^2\right].
\end{aligned}$$

$\square$

Lemma 11 (Elliptical potential count lemma adapted from Lemma C.2 of Jun and Kim (2024))**.** Let $x_1, x_2, ..., x_t \in \mathbb{R}^d$ be a sequence of vectors with $\|x_s\|_2 \leq 1, \forall s \in [t]$. Let $V_t = \lambda I + \sum_{s=1}^{t} x_s x_s^\top$ for some $\lambda > 0$. Let $J = \{s \in [t] : \|x_s\|_{V_{s-1}^{-1}}^2 \geq L^2\}$ for some $L^2 \leq 1$. Then,

$$|J| \leq 3\frac{d}{L^2}\ln\left(1+\frac{2}{L^2\lambda}\right).$$

Lemma 12 (Elliptical potential lemma adapted from Proposition 2 of Abeille and Lazaric (2017))**.** Let $x_1, x_2, ..., x_t \in \mathbb{R}^d$ be a sequence of vectors with $\|x_s\|_2 \leq 1, \forall s \in [t]$. Let $V_t = \lambda I + \sum_{s=1}^{t} x_s x_s^\top$ for some $\lambda > 0$. Then,

$$\sum_{s=1}^{t}\|x_s\|_{V_s^{-1}}^2 \leq 2d\log(1+\frac{t}{d\lambda}).$$

Lemma 13 (OFUL confidence bound lemma adapted from Theorem 2 of Abbasi-Yadkori et al. (2011))**.** Assume $\forall s \in [t], \quad \|a_s\| \leq 1$, and $\|\theta^*\|_2 \leq S$, for some fixed S > 0. We also assume $\Delta_a := \max_{a'\in\mathcal{A}_t}\langle a', \theta^*\rangle - \langle a, \theta^*\rangle \leq 1, \quad \forall a \in \mathcal{A}$

$$\forall t \geq 1, \quad \mathbb{P}\left(\|\hat{\theta}_{t-1}-\theta^*\|_{V_{t-1}} \leq \sqrt{\beta_{t-1}(\delta_{t-1})}\right) \geq 1-\delta.$$

E   LOWER BOUND ARGUMENTS

In this section, we establish an instance-dependent regret lower bound of order $\Omega(\Delta\sqrt{n})$ for the modified version of EXP2 (outlined below) as well as for SpannerIGW (Zhu et al., 2022).To demonstrate this, we consider the following instance:

Let the arm set $\mathcal{A} \subset \mathbb{R}^2$ be

$$\mathcal{A} := \{e_1, e_2\}. \qquad \text{(where as } e_1 = (1,0)^\top, e_2 = (0,1)^\top)$$

Let $\theta^* = (1,0)^\top$ and dimension of the arm set and θ^* be $d = 2$

E.1   Lower bound for EXP2 algorithm (modified version)

The original EXP2 algorithm (Bubeck and Cesa-Bianchi, 2012) was developed for the bounded loss model. However, we introduce a slightly modified version of this algorithm to accommodate the unbounded reward setting. This modified algorithm is essential for establishing the lower bound for regret in such environments.

Algorithm 3 EXP2 (Reward Version)

Input: Finite Arm set $\mathcal{A} \in \mathbb{R}^d$, learning rate η, exploration distribution π, exploration parameter γ Optimal design fraction: α_{opt},

1: **for** $t = 1, 2, \ldots n$ **do**
2:     Compute sampling distribution

$$P_t(a) = \gamma\pi(a) + (1-\gamma)\frac{\exp\left(\eta\langle a, \sum_{s=1}^{t-1}\hat{\theta}_s\rangle\right)}{\sum_{a'\in\mathcal{A}_t}\exp\left(\eta\langle a', \sum_{s=1}^{t-1}\hat{\theta}_s\rangle\right)}.$$

3:     Sample action
$$A_t \sim p_t.$$

4:     Observe the reward,
$$Y_t = \langle\theta^*, A_t\rangle + \eta_t.$$

5:     Update
$$\hat{\theta}_t = Q_t^{-1}A_tY_t.$$

    where $Q_t = \sum_{a\in\mathcal{A}_t} P_t(a)aa^\top$.
6: **end for**

Theorem 6. There exists a linear bandit problem for which the EXP2 algorithm satisfies
$$\mathbb{E}\operatorname{Reg}_n \geq \Omega(\Delta\sqrt{n}).$$

Proof. The EXP2 algorithm samples an arm according to the following probability expression:

$$P_t(a) = \gamma\pi(a) + (1-\gamma)\frac{\exp\left(\eta\sum_{s=1}^{t-1}\langle\hat{\theta}_s, a\rangle\right)}{\sum_{a'\in\mathcal{A}}\exp\left(\eta\sum_{s=1}^{t-1}\langle\hat{\theta}_s, a'\rangle\right)}.$$

Moreover, from Lemma 14, we know that, $\pi(e_1) = \pi(e_2) = \frac{1}{2}$ forms a valid G-optimal design (Kiefer and Wolfowitz, 1960).

Hence,

$$R_n = \mathbb{E}\left[\sum_{t=1}^n\langle\theta^*, a_t^*\rangle - \langle\theta^*, A_t\rangle\right]$$

$$= \mathbb{E}\left[\sum_{t=1}^{n} \overline{\Delta}_t \mathbb{1}\{A_t \neq a_t^*\}\right]$$

$$= \mathbb{E}\left[\sum_{t=1}^{n} \Delta_{e_2} \mathbb{1}\{A_t = e_2\}\right]$$

$$= \sum_{t=1}^{n} \Delta_{e_2} P_t(e_2)$$

$$\geq \sum_{t=1}^{n} \Delta_{e_2} \gamma \pi(e_2)$$

$$= \gamma \frac{n}{2} \Delta. \qquad\qquad (\Delta = \min_{a \in \mathcal{A}, a \neq a_t^*} \Delta_a = \Delta_{e_2})$$

Furthermore, to achieve an optimal minimax bound, we must appropriately tune the parameter γ as follows (Bubeck and Cesa-Bianchi, 2012):

$$\gamma = \sqrt{\frac{g(\pi)^2 \log K}{(2g(\pi) + d)n}}. \qquad\qquad (K = \text{ Number of arms } = 2)$$

Hence,

$$R_n \geq \Delta \cdot \sqrt{\frac{ng(\pi)^2 \log K}{(2g(\pi) + d)}} = \Omega(\Delta\sqrt{n}).$$

This concludes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

E.2 SpannerIGW

Theorem 7. There exists a linear bandit problem for which the SpannerIGW algorithm satisfies
$$\mathbb{E}\operatorname{Reg}_n \geq \Omega(\Delta\sqrt{n}).$$

Proof. Let $\overline{\mathcal{A}}_{(t)}$ be the transformed arm set at time t, SpannerIGW (Zhu et al., 2022) calculates an approximate design (q_t^{opt}), similar to G-optimal design. From Lemma 14,

$$q_t^{\mathrm{opt}}(e_2) = \frac{1}{2}.$$

$$q_t(a) = \frac{1}{2}q_t^{\mathrm{opt}}(a) + \frac{1}{2}\mathbb{I}_{\hat{a}_t}(a)$$

$$q_t(e_2) \geq \frac{1}{4}.$$

Let $n \geq 8$. Recall that $\Delta = 1$. Let us further assume that there is no noise in the reward: $Y_t = \langle A_t, \theta^* \rangle + \eta_t$ where $\eta_t = 0$ with probability 1. This noise η_t can be viewed as having 1-sub-Gaussian noise, and we assume that the algorithm only knows that the noise is 1-sub-Gaussian and does not know that the actual noise is deterministically 0.

For the regression oracle, let us use the online ridge regression $\hat{\theta}_t = (\tau I + \sum_{s=1}^{t} A_s A_s^\top)^{-1} \sum_{s=1}^{t} A_s Y_s \in \mathbb{R}^2$ with regularizer $\tau = 1/10$ and the initial parameter $\hat{\theta}_0 = 0 \in \mathbb{R}^2$. This will enjoy a regret bound of $\operatorname{Reg}_{\mathrm{Sq}}(t) \leq c \log(t)$ for some c. Let N_1 be the number of times arm e_1 has been pulled up to (and including) time step $t-1$. Then, if $N_1 \geq 1$, then the prediction at time t will be $\hat{f}_t(e_1) = \frac{N}{N+\tau}$ and

$$0.9 \leq \frac{1}{1 + \tau} \leq \hat{f}_t(e_1) \leq 1 .$$

Furthermore, the prediction for e_2 is always $\hat{f}_t(e_2) = 0$ at all time.

Note that once both arms have been pulled at least once up to (and including) time step $t-1$, then the probability

of pulling the arm e_2 is

$$\frac{q_t^{\text{opt}} + \frac{1}{2}\mathbb{I}_{\hat{a}_t}}{\lambda + \eta(\hat{f}_t(\hat{a}_t) - \hat{f}_t(e_2))} \geq \frac{\frac{1}{2} \cdot \frac{1}{2} + 0}{\lambda + \eta \cdot (1 - 0)} \geq \frac{\frac{1}{4}}{1 + \frac{\gamma}{C_{\text{opt}}d}} = \frac{\frac{1}{4}}{1 + \frac{\gamma}{d}} = \frac{\frac{1}{4}}{1 + \frac{1}{d} \cdot \sqrt{\frac{2n}{c\ln(n) + 32\log(2/\delta)}}}.$$

where we set $\delta = \frac{1}{4}$.

Let J be the first time step at the end of which we have pulled both e_1 and e_2 at least once; i.e.,

$$J := \min\left\{t \in \mathbb{N}_+ : \sum_{s=1}^{t} \mathbb{1}\{A_t = e_1\} \geq 1, \sum_{s=1}^{t} \mathbb{1}\{A_t = e_2\} \geq 1\right\}.$$

Then,

$$\mathbb{E}[\text{Reg}_n] \geq \mathbb{E}[\text{Reg}_n \mathbb{1}\{J \leq 4\}]$$

$$\geq \mathbb{E}[\mathbb{1}\{J \leq 4\} \sum_{t=1}^{n} \mathbb{1}\{A_t = e_2\}]$$

$$\geq \mathbb{E}[\mathbb{1}\{J \leq 4\} \sum_{t=5}^{n} \mathbb{1}\{A_t = e_2\}]$$

$$= \mathbb{E}[\mathbb{1}\{J \leq 4\} \sum_{t=5}^{n} \mathbb{E}[\mathbb{1}\{A_t = e_2\} \mid A_1, \ldots, A_{t-1}]]$$

$$\geq \mathbb{E}[\mathbb{1}\{J \leq 4\}] \cdot (n-4) \cdot \frac{\frac{1}{4}}{1 + \frac{1}{d} \cdot \sqrt{\frac{2n}{c\ln(n) + 32\log(2/\delta)}}}$$

$$\geq \mathbb{P}(J \leq 4) \cdot \Omega(\sqrt{n \log(n)}).$$

It remains to show that $\mathbb{P}(J \leq 4)$ is lowerbounded by an absolute constant. Note that
$$\mathbb{P}(J \leq 4) = 1 - \mathbb{P}(J \geq 5).$$

and

$$\mathbb{P}(J \geq 5) \leq \mathbb{P}(\forall t \in [4], A_t = e_1) + \mathbb{P}(\forall t \in [4], A_t = e_2).$$

For the first term,

$$\mathbb{P}(\forall t \in [4], A_t = e_1) = \mathbb{P}(A_1 = e_1) \prod_{t=2}^{4} \mathbb{P}(A_t = e_1 \mid A_{1:t-1} = e_1)$$

$$\leq \left(\frac{1}{4} + \frac{1}{2}\right) \cdot 1^3$$

$$\leq \frac{3}{4}.$$

For the second term,

$$\mathbb{P}(\forall t \in [4], A_t = e_2) \leq \left(\frac{\frac{1}{4}}{\lambda}\right)^4 \leq \left(\frac{\frac{1}{4}}{\frac{1}{2}}\right)^4 = \frac{1}{16}.$$

Thus,

$$\mathbb{P}(J \leq 4) \geq 1 - \frac{13}{16} = \frac{3}{16}.$$

This implies that

$$\mathbb{E}[\text{Reg}_n] \geq \Omega(\sqrt{n \log(n)}\Delta).$$

This concludes the proof. $\qquad\square$

E.3 Lemmata

The objective of the following lemma 14 is to demonstrate that, regardless of the scaling applied to the arm set in the previous section (orthogonal basis of $\mathbb{R}^2$), a probability distribution that assigns equal probability to each arm

will still satisfy Assumption 2 and Assumption 3. Moreover, such a distribution is a G-optimal design (Kiefer and Wolfowitz, 1960). This lemma plays a crucial role in establishing the lower bound proof.

Lemma 14. For an arm set $\mathcal{A} := \{p \cdot e_1, q \cdot e_2\}$ where $p, q > 0$ and $e_1 = (1,0)^\top, e_2 = (0,1)^\top$. Let π be a probability distribution that assigns probability to each arms as follows:

$$\pi(a_1) = \pi(a_2) = \frac{1}{2}. \qquad \text{(where } \forall i, \quad a_i \text{ denotes the } i\text{-th arm in } \mathcal{A})$$

Then, π is G-optimal design.

Proof.

$$\mathcal{A} := \{p \cdot e_1, q \cdot e_2\}. \qquad \text{(where } p, q > 0)$$

Here, dimension $d = 2$ and $|\mathcal{A}| = 2$. Let be π be probability distribution with,

$$\pi(a_1) = \pi(a_2) = \frac{1}{2}.$$

$$\sum_{i=1}^{2} \pi(a_i) a_i a_i^T = \frac{1}{2} \left(\begin{bmatrix} p^2 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & q^2 \end{bmatrix} \right)$$

$$V(\pi) := \begin{bmatrix} \frac{p^2}{2} & 0 \\ 0 & \frac{q^2}{2} \end{bmatrix}$$

$$V^{-1}(\pi) = \frac{4}{p^2 \cdot q^2} \begin{bmatrix} \frac{q^2}{2} & 0 \\ 0 & \frac{p^2}{2} \end{bmatrix}$$

$$= \frac{1}{p^2 \cdot q^2} \begin{bmatrix} 2p^2 & 0 \\ 0 & 2q^2 \end{bmatrix}.$$

$$\|a_1\|_{V^{-1}(\pi)}^2 = p^2 e_1^T V^{-1}(\pi) e_1$$

$$= 2.$$

$$\|a_2\|_{V^{-1}(\pi)}^2 = q^2 e_2^T V^{-1}(\pi) e_2$$

$$= 2.$$

Let

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

$$= 2 = d.$$

However, A G-optimal design (Kiefer and Wolfowitz, 1960) π^* should satisfy the following :

$$g(\pi^*) = d.$$

Hence we can conclude,

$$\pi = \pi^*.$$

Hence, π is G-optimal design. □

F ALGORITHM FOR APPROXIMATE OPTIMAL EXPERIMENTAL DESIGN

In this section, we present a procedure to obtain a ApproxDesign() which can satisfy the assumptions 2 and 3.

This procedure consist of two steps. The first step involves implementing the computationally efficient version of the BH sampling algorithm, as presented by Gales et al. (2022) (in appendix Section C.1, Algorithm 5), which refines the original algorithm by Betke and Henk (1993). We present this in Algorithm 4. This implementation outputs $\mathcal{A}_0 = \{a_1, ..., a_{|\mathcal{A}_0|}\} \subseteq \mathcal{A}$ such that, $|\mathcal{A}_0| \le 2d$ and the determinant of the matrix $\overline{V}_{|\mathcal{A}_0|}$ is sufficiently large, where $\overline{V}_k := \sum_{s=1}^{k} a_s a_s^\top$.

Algorithm 4 Computationally efficient BH algorithm

Input: Original arm set $\mathcal{A} \subset \mathbb{R}^d$ with $|\mathcal{A}| = K$

1: **if** $K \le 2d$ **then**
2:    $\mathcal{A}_0 \leftarrow \mathcal{A}$
3:    **return** $\mathcal{A}_0$
4: **end if**
5: $\Psi \leftarrow \{0\}$, $\mathcal{A}_0 \leftarrow \emptyset$, $i \leftarrow 0$, $v_0 \leftarrow (0, \cdots, 0)^\top \in \mathbb{R}^d$
6: **while** $\mathbb{R}^d \setminus \Psi \neq \emptyset$ **do**
7:    $i \leftarrow i + 1$
8:    **if** $i = 1$ **then**
9:       Set $b_i = e_i$ where e_i is the i-th index vector.
10:    **else**
11:       Set $v_{i-1}^\perp = v_{i-1} - \sum_{j=0}^{i-2} \frac{\langle v_j^\perp, v_{i-1} \rangle}{\langle v_j^\perp, v_j^\perp \rangle} v_j^\perp$
12:       Set $b_i = e_i - \sum_{j=0}^{i-1} \frac{\langle v_j^\perp, e_i \rangle}{\langle v_j^\perp, v_j^\perp \rangle} v_j^\perp$
13:    **end if**
14:    $p \leftarrow \arg\max_{a \in \mathcal{A}} \langle b_i, a \rangle$
15:    $q \leftarrow \arg\min_{a \in \mathcal{A}} \langle b_i, a \rangle$
16:    $\mathcal{A}_0 \leftarrow \mathcal{A}_0 \cup \{p\} \cup \{q\}$
17:    $v_i \leftarrow p - q$
18:    $\Psi \leftarrow \text{Span}(\Psi, v_i)$
19: **end while**
20: **return** $\mathcal{A}_0$

The second step of the procedure is detailed in Algorithm 5. It takes the output $\mathcal{A}_0$ from the computationally efficient BH sampling algorithm, refines the probability distribution for arms in $\mathcal{A}$.

Algorithm 5 ApproxDesign

Input: Original arm set $\mathcal{A} \subset \mathbb{R}^d$

1: $\mathcal{A}_0 \leftarrow$ Computationally efficient BH algorithm$(\mathcal{A})$
2: $k \leftarrow |\mathcal{A}_0| + 1$
3: **while** $\max_{a \in \mathcal{A}} \|a\|_{\overline{V}_{k-1}^{-1}}^2 > 1$ **do**
4:    $a_k = \arg\max_{a \in \mathcal{A}} \|a\|_{\overline{V}_{k-1}^{-1}}^2$
5:    $k \leftarrow k + 1$
6: **end while**
7: $\tau = k - 1$
8: **return** π such that $\forall a \in \mathcal{A}$, $\pi(a) = \frac{C_\tau^{\mathcal{A}}(a)}{\sum_{b \in \mathcal{A}} C_\tau^{\mathcal{A}}(b)}$ where $\mathcal{C}_k^{\mathcal{A}}(a) := \sum_{s=1}^{k} \mathbb{1}\{a_s = a\}$, $\forall a \in \mathcal{A}$.

Furthermore, Theorem 5 of Gales et al. (2022) shows that,

$$\tau = \mathcal{O}(d \log d). \tag{12}$$

Claim 3. Let π be the design from the ApproxDesign() Algorithm 5, then,

$$\|a\|_{V^{-1}(\pi)}^2 \le C_{\text{opt}} d \log(d), \forall a \in \mathcal{A}.$$

Proof. Note that, when the algorithm stops,

$$\max_{a \in \mathcal{A}} \|a\|^2_{\overline{V}_\tau^{-1}} \leq 1. \tag{13}$$

Furthermore, we have

$$
\begin{aligned}
\overline{V}_\tau &= \sum_{s=1}^{\tau} a_s a_s^\top \\
&= \sum_{a \in \mathcal{A}} C_\tau^{\mathcal{A}}(a) a a^\top \\
&= \left(\sum_{b \in \mathcal{A}} C_\tau^{\mathcal{A}}(b) \right) \sum_{a \in \mathcal{A}} \frac{C_\tau^{\mathcal{A}}(a)}{\sum_{b \in \mathcal{A}} C_\tau^{\mathcal{A}}(b)} a a^\top \\
&= \left(\sum_{b \in \mathcal{A}} C_\tau^{\mathcal{A}}(b) \right) \sum_{a \in \mathcal{A}} \pi(a) a a^\top \\
&= \left(\sum_{b \in \mathcal{A}} C_\tau^{\mathcal{A}}(b) \right) V(\pi) \\
&= \tau V(\pi) \\
&= \mathcal{O}(d \log d) V(\pi). \tag{from (12)}
\end{aligned}
$$

Then,

$$
\begin{aligned}
\|a\|^2_{\overline{V}_\tau^{-1}} &= \frac{1}{\mathcal{O}(d \log d)} \|a\|^2_{V^{-1}(\pi)} \\
\|a\|^2_{V^{-1}(\pi)} &\leq \mathcal{O}(d \log d) \cdot \max_{a \in \mathcal{A}} \|a\|^2_{\overline{V}_\tau^{-1}} \\
&\leq \mathcal{O}(d \log d) \cdot 1 \tag{from (13)} \\
&\leq C_{\text{opt}} d \log(d).
\end{aligned}
$$

$\square$

Claim 4.

$$|\text{supp}(\pi)| = \tilde{\mathcal{O}}(d) .$$

Proof.

$$
\begin{aligned}
|\text{supp}(\pi)| &\leq \tau \\
&= \mathcal{O}(d \log d) \tag{from (12)} \\
&= \tilde{O}(d).
\end{aligned}
$$

$\square$

Hence, it is proved that the Algorithm 5 along with computationally efficient BH sampling Algorithm 4 outputs a design that satisfies Assumptions 2 and 3.

G EMPIRICAL STUDIES

G.1 End of optimism experiments

Since this experiment has already been covered in the main body of the paper, we present it concisely here.

Experimental setup: We set the number of arms $K = 3$, dimension $d = 2$ and $\mathcal{A} = \{a_1 = (1,0)^\top, a_2 = (0,1)^\top, a_3 = (1 - \varepsilon, 2\varepsilon)^\top\}$ where $\varepsilon \in \{0.005, 0.01, 0.02\}$ and $\theta^* = (1,0)^\top$. The noise follows $\mathcal{N}(0, \sigma_*^2)$. The time horizon for each trial is $n = 1,000,000$ and conduct 20 such independent trials. Furthermore, we conduct experiments for the cases i) $\sigma^2 = \sigma_*^2$ where $\sigma_*^2 = 0.1$, ii) $\sigma^2 = 2 \cdot \sigma_*^2$ where $\sigma_*^2 = 0.1$, and iii) $\sigma^2 = 0.1 \cdot \sigma_*^2$ where $\sigma_*^2 = 10$.

Algorithms evaluated: We evaluate the following algorithms: OFUL (Abbasi-Yadkori et al., 2011), Lin-TS-Freq (Thompson sampling frequentest version) (Agrawal and Goyal, 2014), Lin-TS-Bayes (Thompson sampling Bayesian version) (Russo and Roy, 2014), Lin-IMED-1 (Bian and Tan, 2024), Lin-IMED-3 (Bian and Tan, 2024), LinMED-99 ($\alpha_{\mathrm{opt}} = 0.99$), LinMED-90 ($\alpha_{\mathrm{opt}} = 0.90$), and LinMED-50 ($\alpha_{\mathrm{opt}} = 0.50$). Note that, for Lin-TS-Bayes, we are still evaluating the frequentest regret as we do for every other algorithms. Lin-IMED-3 have a hyper-parameter C, which we set to $C = 30$ following Bian and Tan (2024).

(a) $\sigma^2 = \sigma_*^2$ (b) $\sigma^2 = 2 \cdot \sigma_*^2$ (c) $\sigma^2 = 0.1 \cdot \sigma_*^2$
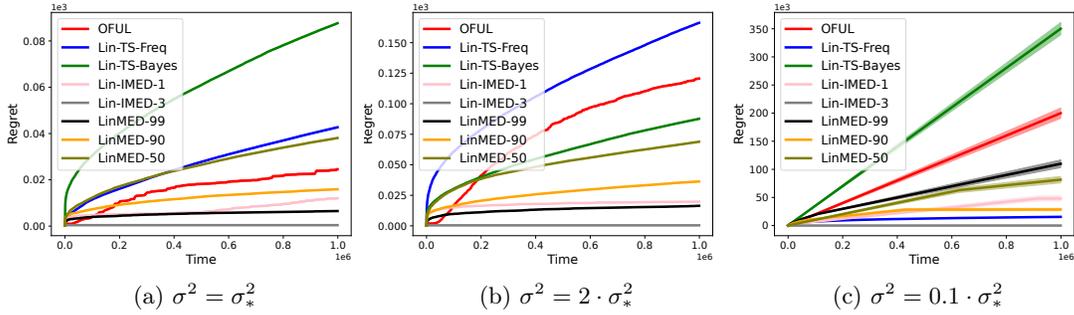
Figure 5: End of optimism experiments $\varepsilon = 0.005$

(a) $\sigma^2 = \sigma_*^2$ (b) $\sigma^2 = 2 \cdot \sigma_*^2$ (c) $\sigma^2 = 0.1 \cdot \sigma_*^2$
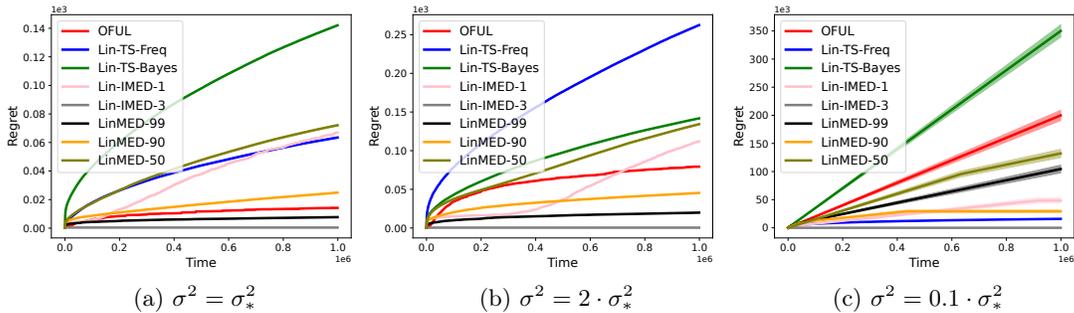
Figure 6: End of optimism experiments $\varepsilon = 0.01$

Remarks : This experiment has been discussed already in the main body of the paper for $\sigma^2 = \sigma_*^2$ and $\sigma^2 = 2 \cdot \sigma_*^2$. In addition to that, we observe the following: when the noise is under-specified ($\sigma^2 = 0.1 \cdot \sigma_*^2$), the performance of both OFUL and Lin-TS-Bayes degrades significantly, resulting in nearly linear regret. On the other hand, Lin-TS-Freq performs well due to the fact that the degree of oversampling is lesser due to small σ^2. Among the LinMED variants, LinMED-99 shows a more pronounced deterioration compared to LinMED-90 and LinMED-50, indicating that higher exploration is required in such scenarios. Notably, Lin-IMED-3 maintains strong performance across all conditions.

$$(a)\ \varepsilon = 0.02,\ \sigma^2 = \sigma_*^2 \qquad (b)\ \varepsilon = 0.02,\ \sigma^2 = 2 \cdot \sigma_*^2 \qquad (c)\ \sigma^2 = 0.1 \cdot \sigma_*^2$$
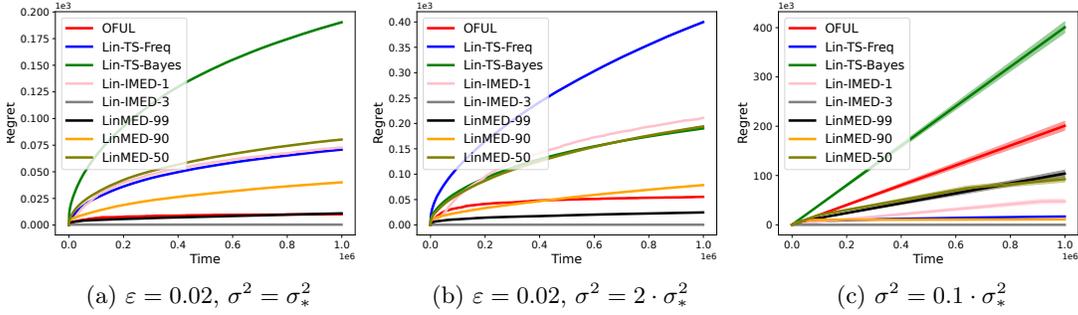
Figure 7: End of optimism experiments $\varepsilon = 0.02$

G.2 Delayed reward experiments on real-world data set

One of the advantages of randomized algorithms over deterministic algorithms is their superior performance in delayed reward settings, where immediate rewards are not accessible. To investigate this, we utilized the MovieLens real-world dataset. We extracted user and movie features and constructed our own true parameter θ^* for the reward calculation, as opposed to relying on the reported rewards in the dataset. This approach was necessary due to the sparsity of the reported rewards, where many users not providing ratings for numerous movies. Additionally, we performed Principal Component Analysis (PCA) to isolate the dominant features from both the movie and user datasets. Then we generated 2 dimensional movie feature vectors and 2 dimensional user feature vectors.

During the implementation we first fix randomly chosen $K = 10$ movies and for each time step, we randomly select a user as the context and generate arm set of size $K = 10$ by taking outer product of each movie vectors with the user vector. Since we calculated the outer product between 2 dimensional user vector and 2 dimensional movie vector, the resulting dimension of the arm set is $d = 4$. We set the noise parameter to $\sigma^2 = \sigma_*^2 = 1$ and varied the delay time across the set $\{0, 10, 20\}$. The time horizon for each trial is $n = 5,000$ and conduct 100 such independent trials.

Algorithms evaluated: We evaluate the following algorithms: OFUL (Abbasi-Yadkori et al., 2011), Lin-TS-Bayes (Thompson sampling Bayesian version) (Russo and Roy, 2014), LinMED-99 ($\alpha_{\mathrm{opt}} = 0.99$), LinMED-90 ($\alpha_{\mathrm{opt}} = 0.90$), and LinMED-50 ($\alpha_{\mathrm{opt}} = 0.50$). Note that, for Lin-TS-Bayes, we are still evaluating the frequentest regret as we do for every other algorithms.

$$(a)\ \text{Delay} = 0 \qquad (b)\ \text{Delay} = 10 \qquad (c)\ \text{Delay} = 20$$
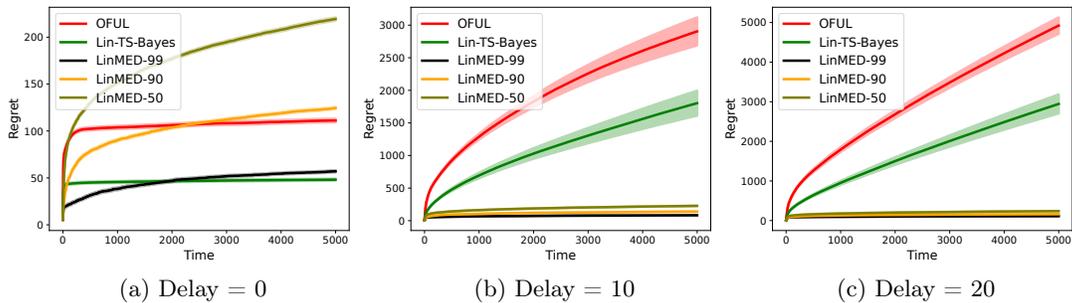
Figure 8: Delayed reward experiments

As expected, the performance of OFUL significantly deteriorates when rewards are delayed, in contrast to the randomized algorithms, LinMED and Lin-TS-Bayes. Moreover, all three variants of LinMED demonstrate strong performance under these conditions.

Lin-TS-Feq	$M = 10^3$	$M = 10^4$	$M = 10^5$	Oracle
Mean	0.906	0.819	0.799	0.800
Standard deviation	0.099	0.069	0.039	0

Table 2: Mean and standard deviation of rewards received by the Uniform policy using logged data from Lin-TS-Freq for offline evaluation. Here M stands for number of Monte-carlo samples

G.3 Offline evaluation experiments

This section presents our simulation results on offline evaluation using logged data. We utilize the logged data generated by our algorithms, LinMED and Lin-TS-Freq (frequentest version), to estimate the expected reward of a policy (call it Uniform target policy) that selects arms uniformly at random from $\mathcal{A}$ (we use fixed arm set for this experiment). The logged data takes the form $(A_t, p_t(A_t), Y_t)_{t=0}^n$, where A_t represents the chosen arm, $p_t(A_t)$ denotes the probability (either exact or approximate) of selecting that arm, and Y_t indicates the received reward at time step t. We consider the Inverse Propensity Weighting (IPW) estimator to estimate the cumulative reward of the Uniform target policy as follows:

$$\text{IPW score} = \frac{1}{n} \sum_{t=1}^{n} \frac{\frac{1}{|\mathcal{A}|}}{p_t(A_t)} \cdot Y_t \tag{14}$$

LinMED assigns a closed-form probability to the chosen arm, whereas Lin-TS-Freq estimates the probability of selecting an arm using Monte Carlo trials. We set the number of arms $K = 2$, dimension $d = 2$ and $\mathcal{A} = \{a_1 = (1,0)^\top, a_2 = (0.6, 0.8)^\top\}$, $\theta^* = (1,0)^\top$ while varying the number of Monte Carlo samples for estimating the probability of arm selection in Lin-TS-Freq across the set $\{10^3, 10^4, 10^5\}$. The noise is modeled as $\mathcal{N}(0, \sigma_*^2)$, with $\sigma^2 = \sigma_*^2 = 0.1$. Throughout this experiment, we evaluate LinMED-50 ($\alpha_{\text{opt}} = 0.5$). The time horizon for each trial is $n = 1,000$ and conduct $5,000$ such independent trials to calculate the histogram representation of IPW scores. See Figure 9.

(a) Monte-carlo samples $= 10^3$ (b) Monte-carlo samples $= 10^4$ (c) Monte-carlo samples $= 10^5$
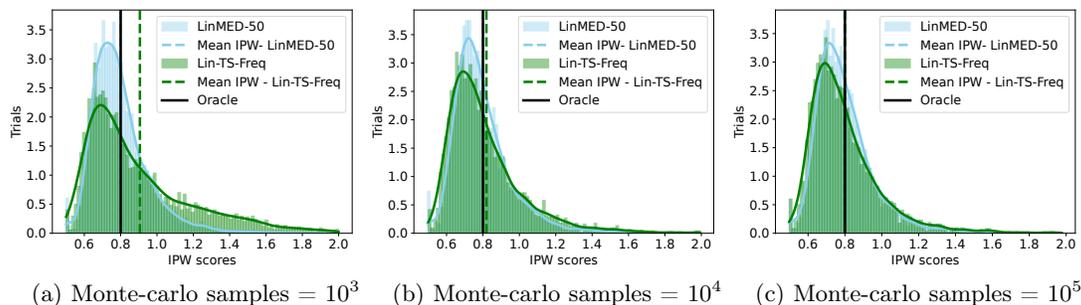
Figure 9: Offline evaluation experiments

In Figure 9, the oracle value of 0.8 represents the expected reward of the Uniform policy when real-time data is used. The mean reward received by the Uniform policy using logged data from LinMED-50 for offline evaluation matches the oracle value to a two-decimal place accuracy, with a standard deviation of approximately 0.03. The mean and standard deviation of Lin-TS-Freq are provided in Table 2. Although the performance of Lin-TS-Freq approaches that of LinMED-50 when the number of Monte Carlo samples reaches 10^5, there is a significant bias for sample sizes of 10^3 and 10^4. However, using 10^5 samples for estimating probabilities is computationally expensive. Additionally, there are cases where Lin-TS-Freq assigns zero probability to some arms. Due to LinMED's ability to assign a closed-form probability to each arm, LinMED is more suitable than Lin-TS-Freq for offline evaluation.

G.4 Synthetic unit ball arm set experiments

G.4.1 Fixed number of arms (K), different dimensions (d)

Experimental setup: We fix the number of arms $K = 10$. For different $d \in \{2, 20, 50\}$, we randomly sample $K = 10$ arms from d dimensional unit ball S^{d-1}. The noise follows $\mathcal{N}(0, \sigma_*^2)$ with $\sigma_*^2 = 1$. The time horizon for

each trial is $n = 5,000$ and conduct 50 such independent trials. Furthermore, we conduct experiments for the cases i) $\sigma^2 = \sigma_*^2$, ii) $\sigma^2 = 2 \cdot \sigma_*^2$, and iii) $\sigma^2 = 0.1 \cdot \sigma_*^2$.

Algorithms evaluated: We evaluate the following algorithms: OFUL (Abbasi-Yadkori et al., 2011), Lin-TS-Freq (Thompson sampling frequentest version) (Agrawal and Goyal, 2014), Lin-TS-Bayes (Thompson sampling Bayesian version) (Russo and Roy, 2014), Lin-IMED-1 (Bian and Tan, 2024), Lin-IMED-3 (Bian and Tan, 2024), LinMED-99 ($\alpha_{\mathrm{opt}} = 0.99$), LinMED-90 ($\alpha_{\mathrm{opt}} = 0.90$), LinMED-50 ($\alpha_{\mathrm{opt}} = 0.50$), SpannerIGW (Zhu et al., 2022), and SpannerIGW-AT (Zhu et al., 2022). Note that, for Lin-TS-Bayes, we are still evaluating the frequentest regret as we do for every other algorithms. Lin-IMED-3 have a hyper-parameter C, which we set to $C = 30$ following Bian and Tan (2024). Moreover SpannerIGW-AT is the anytime version of SpannerIGW.

(a) $d = 2$ (b) $d = 20$ (c) $d = 50$

Figure 10: $\sigma^2 = \sigma_*^2$

(a) $d = 2$ (b) $d = 20$ (c) $d = 50$

Figure 11: $\sigma^2 = 0.1 \cdot \sigma_*^2$

(a) $d = 2$ (b) $d = 20$ (c) $d = 50$

Figure 12: $\sigma^2 = 2 \cdot \sigma_*^2$

Remarks: Firstly, Lin-IMED-3, which demonstrated strong performance in most of the prior experiments, exhibits notably poor performance in this particular experiment. This decline could potentially be attributed to the choice of hyperparameter setting, specifically $C = 30$. However, it is reasonable to retain this setting, as the same value was consistently applied in the previous experiments.

Secondly, although the performance of all algorithms deteriorates with increasing dimensionality, the decline in Lin-TS-Freq is particularly pronounced due to the $\sqrt{d}$ oversampling, a factor also reflected in its theoretical regret guarantee. This downward trend in Lin-TS-Freq becomes more severe when the sub-Gaussian parameter of the noise is over-specified. However, an improvement in performance is observed when the sub-Gaussian parameter of the noise is under-specified. The latter trend is expected, as the oversampling rate of Lin-TS-Freq grows with σ^2.

OFUL performs well except when the sub-Gaussian parameter of the noise is over-specified.

All variants of LinMED perform competitively compared to other algorithms. Notably, the performance of LinMED is not significantly affected by noise misspecifications. Moreover, LinMED-90 outperforms LinMED-99 in high-dimensional contexts, suggesting that a higher degree of exploration is essential when dealing with such settings.

G.4.2 Fixed dimension (d), different numbers of arms (K)

Experimental setup: We fix the dimension $d = 2$. For different $K \in \{10, 100, 500\}$, we randomly samples K arms from d dimensional unit ball S^{d-1}. The noise follows $\mathcal{N}(0, \sigma_*^2)$ with $\sigma_*^2 = 1$. The time horizon for each trial is $n = 5000$ and conduct 50 such independent trials. Furthermore, we conduct experiments for the cases i) $\sigma^2 = \sigma_*^2$, ii) $\sigma^2 = 2 \cdot \sigma_*^2$, and iii) $\sigma^2 = 0.1 \cdot \sigma_*^2$.

Algorithms evaluated: We evaluate the following algorithms: OFUL (Abbasi-Yadkori et al., 2011), Lin-TS-Freq (Thompson sampling frequentest version) (Agrawal and Goyal, 2014), Lin-TS-Bayes (Thompson sampling Bayesian version) (Russo and Roy, 2014), Lin-IMED-1 (Bian and Tan, 2024), Lin-IMED-3 (Bian and Tan, 2024), LinMED-99 ($\alpha_{\mathrm{opt}} = 0.99$), LinMED-90 ($\alpha_{\mathrm{opt}} = 0.90$), LinMED-50 ($\alpha_{\mathrm{opt}} = 0.50$), SpannerIGW (Zhu et al., 2022), and SpannerIGW-AT (Zhu et al., 2022). Note that, for Lin-TS-Bayes, we are still evaluating the frequentest regret as we do for every other algorithms. Lin-IMED-3 have a hyper-parameter C, which we set to $C = 30$ following Bian and Tan (2024). Moreover SpannerIGW-AT is the anytime version of SpannerIGW.

(a) $K = 10$ (b) $K = 100$ (c) $K = 500$

Figure 13: $\sigma^2 = \sigma_*^2$

(a) $K = 10$ (b) $K = 100$ (c) $K = 500$

Figure 14: $\sigma^2 = 0.1 \cdot \sigma_*^2$

Remarks: Similar to the fixed K setting we analyzed previously, Lin-IMED-3 demonstrates noticeably poor performance. Additionally, the performance of Lin-TS-Freq deteriorates when the sub-Gaussian parameter of the

(a) $K = 10$

(b) $K = 100$

(c) $K = 500$
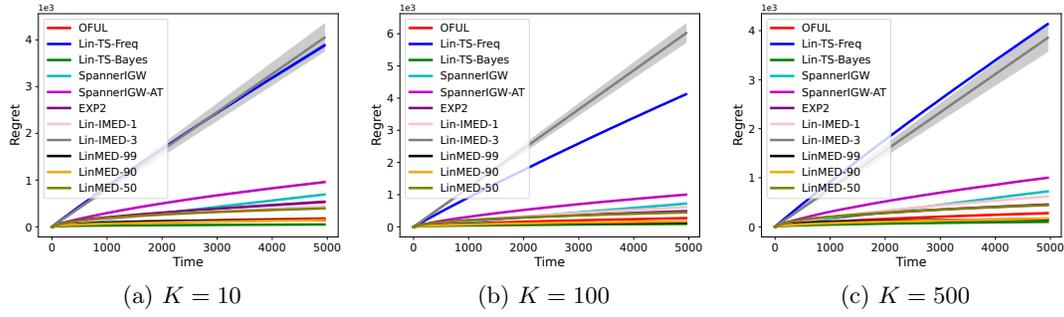
Figure 15: $\sigma^2 = 2 \cdot \sigma_*^2$

noise is over-specified. All variants of LinMED perform competitively compared to other algorithms. Moreover, across most algorithms, there are no substantial variations in performance with respect to K, suggesting that the regret does not exhibit dependency on K.