Logistic Bandits with $\tilde{O}(\sqrt{dT})$ Regret without Context Diversity Assumptions

Seoungbin Bae
Department of Industrial & Systems Engineering
KAIST
sbbae31@kaist.ac.kr

Dabeen Lee
Department of Mathematical Sciences
Seoul National University
dabeenl@snu.ac.kr

Abstract

We study the K-armed logistic bandit problem, where at each round, the agent observes K feature vectors associated with K actions. Existing approaches that achieve a rate-optimal $\widetilde{\mathcal{O}}(\sqrt{dT})$ regret bound rely heavily on context diversity assumptions, such as strict positivity of the minimum eigenvalue of a context covariance matrix. These assumptions, however, impose strong restrictions on the context process, as they rule out the situation where the context vectors are concentrated in a low-dimensional subspace. In this paper, we propose SupSplitLog, which, to the best of our knowledge, is the first algorithm for logistic bandits that achieves $\widetilde{\mathcal{O}}(\sqrt{dT})$ regret without any context diversity assumption. The key idea is to split the collected samples into two disjoint subsets when constructing estimators; one is used to compute an *initial-point estimator*, while the other is used to apply a Newton-type *one-step correction* procedure. The splitting rule is carefully designed to balance the accuracy requirements of the initial-point estimator and the one-step correction procedure. Moreover, SupSplitLog strictly improves on the existing algorithms in terms of the dependence on dimension d in the regret upper bound. Furthermore, SupSplitLog can be adapted simply to deduce a regret bound that grows with a data-dependent complexity measure, avoiding a direct dependence on d, which is favorable when the context vectors are concentrated in a low-dimensional subspace. We also provide experimental results that demonstrate numerically the superiority of our algorithm, validating the theoretical results.

1 Introduction

Logistic bandits provide modeling frameworks for sequential decision-making problems with binary feedback. They naturally arise in many real-world applications and have therefore been extensively studied in recent years (Jun et al., 2017; Liu et al., 2024; Zhang et al., 2025; Boudart et al., 2026). Much of the theoretical literature has focused on improving the dependence of regret bounds on problem-dependent quantities, such as κ (Faury et al., 2020; Abeille et al., 2021), where $1/\kappa$ characterizes the minimum variance of the reward distribution, and the norm bound B on the unknown parameter (Sawarni et al., 2024; Lee et al., 2024b). However, the question of whether we can further reduce the dependence on the context dimension d remains relatively underexplored.

Existing approaches that achieve a rate-optimal $\widetilde{\mathcal{O}}(\sqrt{dT})$ regret bound for K-armed logistic bandits rely on strong assumptions on the context process (Li et al., 2017; Jun et al., 2021; Kim et al., 2023). They assume that either contexts are i.i.d. or there exists a strictly positive lower bound on the minimum eigenvalue of a context covariance matrix, given by $\lambda_{\min}(\mathbb{E}[\frac{1}{K}\sum_{a\in[K]}\boldsymbol{x}_{t,a}\boldsymbol{x}_{t,a}^\top]) > \sigma_0^2$. These assumptions not only impose strong restrictions on the context process, but also lead to overly conservative regret bounds in high-dimensional problems. In particular, such assumptions forbid the

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

Table 1: Comparison with prior work. "Data-dependent" indicates whether the regret bound admits a refinement in terms of a data-dependent complexity measure, such as $\log \frac{\det \mathbf{V}_T}{\det \kappa\lambda \mathbf{I}}$, where $\mathbf{V}_T$ denotes the Gram matrix formed by the realized feature vectors up to round T.

Algorithm	Diversity assumption ($\sigma_0^{-2} = \Omega(d)$)	Regret $\widetilde{\mathcal{O}}(\cdot)$	Data-dependent
SupCB-GLM (Li et al., 2017)	$\lambda_{\min}(\mathbb{E}[\frac{1}{K}\sum_{a\in[K]} \boldsymbol{x}_{t,a}\boldsymbol{x}_{t,a}^\top]) > \sigma_0^2$	$\kappa\sqrt{dT} + d\sigma_0^{-8} + \kappa^8 d^3 \sigma_0^{-4}$	✗
SupLogistic (Jun et al., 2021)	$\lambda_{\min}(\mathbb{E}[\frac{1}{K}\sum_{a\in[K]} \boldsymbol{x}_{t,a}\boldsymbol{x}_{t,a}^\top]) > \sigma_0^2$	$\sqrt{dT} + d\sigma_0^{-8} + \kappa^2 d\sigma_0^{-4}$	✗
DDRTS-GLM (Kim et al., 2023)	$\lambda_{\min}(\mathbb{E}[\frac{1}{K}\sum_{a\in[K]} \boldsymbol{x}_{t,a}\boldsymbol{x}_{t,a}^\top]) > \sigma_0^2$	$\sqrt{\kappa\sigma_0^{-2}T} + \kappa^3\sigma_0^{-6}$	✗
SupSplitLog (ours)	-	$\kappa\sqrt{dT} + \kappa^{2.5} d^{2.5}$	✓

case where context vectors are concentrated around a low-dimensional subspace, in which case the ambient dimension d can be much larger than the dimension of the subspace.

In this paper, we address this gap by developing a new algorithm that guarantees $\widetilde{\mathcal{O}}(\sqrt{dT})$ regret without any context diversity assumption. Our algorithm, SupSplitLog, not only lifts the assumption but also attains a regret bound whose dependence on d is strictly better than the existing algorithms. Moreover, SupSplitLog may deal with context vectors concentrated around a low-dimensional subspace, and it can be adapted to deduce a regret bound that scales with a data-dependent complexity measure, avoiding a direct dependence on d. SupSplitLog is based on an innovative algorithmic component that splits the incoming samples according to their roles in constructing estimators and applies a Newton-type one-step correction procedure. Our contributions are summarized as follows:

- We propose SupSplitLog, which is the first algorithm for logistic bandits that achieves a $\widetilde{\mathcal{O}}(\sqrt{dT})$ regret bound without any context diversity assumption. Moreover, the regret bound achieves the smallest dependence on d compared to the existing algorithms (see Table 1).
- Moreover, the algorithm can be adapted to deduce a regret bound that grows with a data-dependent complexity measure such as $\log \frac{\det \mathbf{V}_T}{\det \kappa\lambda \mathbf{I}}$, where $\mathbf{V}_T$ denotes the Gram matrix formed by the realized feature vectors, thereby avoiding a direct dependence on d.
- The main technical idea is to split the collected samples into two disjoint subsets when constructing estimators: one subset is used to compute an *initial-point estimator*, while the other is used to perform a Newton-type *one-step correction* procedure. The splitting rule is carefully designed to balance the accuracy requirements between the initial-point estimator and the one-step correction framework. This novel technique of one-step correction leads to a high-probability bound on the estimation error that does not depend on d.
- We complement our theoretical analysis with experiments on high-dimensional problems with low-dimensional structure, where the realized complexity $\log \frac{\det \mathbf{V}_T}{\det \kappa\lambda \mathbf{I}}$ is small. The results show that SupSplitLog performs most favorably among the baselines in such regimes, consistent with our theory.

2 Related work

Although there is a regret lower bound of $\Omega(d\sqrt{T})$ for the general (infinite-armed) setting of logistic bandits (Abeille et al., 2021), the K-armed setting admits sharper bounds of order $\widetilde{\mathcal{O}}(\sqrt{dT})$ (Li et al., 2017; Jun et al., 2021), and these match a lower bound of order $\Omega(\sqrt{dT})$ for K-armed generalized linear bandits (Li et al., 2019). This improvement is made possible by what we call the Auer framework (Auer, 2002), which constructs independent sample buckets and thereby allows one to apply scalar-valued concentration inequalities for independent noises, instead of the vector-valued concentration arguments typically used in the general setting. Building on the Auer framework, Li et al. (2017) first established a regret bound of order $\widetilde{\mathcal{O}}(\kappa\sqrt{dT} + \kappa^8 d^3\sigma_0^{-4} + d\sigma_0^{-8})$ for K-armed logistic bandits. Later, Jun et al. (2021) improved the dependence on the problem-dependent quantity κ, where $1/\kappa$ characterizes the minimum variance of the reward distribution (see Assumption 2 for a formal definition), and obtained a regret bound of order $\widetilde{\mathcal{O}}(\sqrt{dT} + \kappa^2 d\sigma_0^{-4} + d\sigma_0^{-8})$. However, both

results rely on the context diversity assumption $\lambda_{\min}(\mathbb{E}[\frac{1}{K}\sum_{a\in[K]}\boldsymbol{x}_{t,a}\boldsymbol{x}_{t,a}^\top]) > \sigma_0^2$, which becomes problematic in high-dimensional regimes since $\sigma_0^{-2} = \Omega(d)$.

More recently, Kim et al. (2023) proposed a different approach based on doubly robust estimation and Thompson sampling, taking a detour from the Auer framework, and their regret bound is $\widetilde{\mathcal{O}}(\sigma_0^{-1}\sqrt{\kappa T} + \kappa^3\sigma_0^{-6})$. Nevertheless, their analysis still requires the context diversity assumption, and $\sigma_0^{-1}\sqrt{T}$ can be worse than $\sqrt{dT}$ as $\sigma_0^{-2} = \Omega(d)$.

There exist algorithms for generalized linear bandits and infinite-armed logistic bandits that do not rely on the context diversity assumption, but their regret bounds are of order $\widetilde{\mathcal{O}}(d\sqrt{T})$ (Filippi et al., 2010). As in Table 1, our algorithm achieves a regret bound of order $\widetilde{\mathcal{O}}(\kappa\sqrt{dT} + \kappa^{2.5}d^{2.5})$ without the context diversity assumption. Moreover, as $\sigma_0^{-2} = \Omega(d)$, our regret bound improves on the previous bounds in terms of dependence on d. We list additional related work in Section A.

3 Problem setting

We consider the K-armed contextual bandit problem where for each round $t \in [T]$, the agent observes an arm set $\mathcal{X}_t \coloneqq \{\boldsymbol{x}_{t,a}\}_{a\in[K]} \subset \mathbb{R}^d$, chooses $a_t \in [K]$, and receives a stochastic reward $r_t \in \{0, 1\}$ given by $r_t \mid \mathcal{F}_t, \boldsymbol{x}_{t,a_t} \sim \mathrm{Bern}(\mu(\boldsymbol{x}_{t,a_t}^\top\boldsymbol{\theta}^*))$. Here, $\mathcal{F}_t = \sigma(\{\mathcal{X}_i, a_i\}_{i=1}^t, \{r_i\}_{i=1}^{t-1})$ defines a filtration where $\boldsymbol{x}_{t,a_t}$ is $\mathcal{F}_t$-measurable and r_t is $\mathcal{F}_{t+1}$-measurable. In addition, $\mu(z) = (1 + e^{-z})^{-1}$ is the logistic function, and $\boldsymbol{\theta}^* \in \mathbb{R}^d$ is an unknown parameter. Let a_t^* be an optimal arm in round t, i.e., $a_t^* = \arg\max_{a\in[K]} \boldsymbol{x}_{t,a}^\top\boldsymbol{\theta}^*$. The agent's goal is to minimize cumulative regret, defined as $R_T \coloneqq \sum_{t=1}^T [\mu(\boldsymbol{x}_{t,a_t^*}^\top\boldsymbol{\theta}^*) - \mu(\boldsymbol{x}_{t,a_t}^\top\boldsymbol{\theta}^*)]$.

The following are some standard assumptions used throughout the paper.

Assumption 1. *$\|\boldsymbol{x}\|_2 \le 1$ for all $\boldsymbol{x} \in \mathcal{X}_t, t \in [T]$, and $\boldsymbol{\theta}^* \in \Theta \coloneqq \{\boldsymbol{\theta} : \|\boldsymbol{\theta}\|_2 \le B\}$ for some $B > 0$. Here, $\boldsymbol{\theta}^*$ is unknown while B is known to the agent.*

Assumption 2. *For some $\kappa, L_\mu > 0$, we have $1/\kappa \le \dot{\mu}(\boldsymbol{x}^\top\boldsymbol{\theta}) \le L_\mu$ for all $\boldsymbol{x} \in \mathcal{X}_t, t \in [T]$, $\boldsymbol{\theta} \in \Theta$.*

Assumption 1 impose bounds on the size of feature vectors and the unknown parameter $\boldsymbol{\theta}^*$. Assumption 2 introduces instance-dependent constants that characterize the local behavior of $\dot{\mu}(\cdot)$. Here, we set $L_\mu = 1/4$ to simplify the proof for the base case of Lemma 14.

Assumption 3. *We assume that the context sequence is generated independently of the agent's past actions and realized rewards.*

Remark 4 (Remark on Assumption 3)**.** *Previous works (Li et al., 2017; Jun et al., 2021; Kim et al., 2023) assume that (i) $\mathcal{X}_1, \dots, \mathcal{X}_T$ are i.i.d. with a fixed distribution ν, and (ii) the minimum eigenvalue of the context covariance matrix is uniformly lower bounded by a strictly positive quantity, i.e., $\lambda_{\min}(\mathbb{E}[\frac{1}{K}\sum_{a\in[K]}\boldsymbol{x}_{t,a}\boldsymbol{x}_{t,a}^\top]) > \sigma_0^2$ with $\sigma_0^{-2} = \Omega(d)$. We impose none of these assumptions. Assumption 3 is strictly weaker than assuming i.i.d. contexts together with the covariance lower bound.*

4 Challenges in achieving $\widetilde{O}(\sqrt{dT})$ regret in logistic bandits

In this section, we identify the key requirement for achieving $\widetilde{O}(\sqrt{dT})$ regret in logistic bandits, explain the main technical obstacle in establishing this requirement, and review the approach of previous work together with its limitations. This will help us to better motivate our framework.

To achieve $\widetilde{O}(\sqrt{dT})$ regret, we need a high-probability upper bound on the error term $|\boldsymbol{x}^\top(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*)|$ that does not depend on d for a fixed $\boldsymbol{x} \in \mathbb{R}^d$ and a given estimator $\widehat{\boldsymbol{\theta}}_t$ after round t (Li et al., 2017; Jun et al., 2021). When $\widehat{\boldsymbol{\theta}}_t$ is the maximum likelihood estimator of the logistic loss $\mathcal{L}_t(\boldsymbol{\theta}) = -\sum_{i=1}^t [r_i \log(\mu(\boldsymbol{x}_{i,a_i}^\top\boldsymbol{\theta})) + (1 - r_i)\log(1 - \mu(\boldsymbol{x}_{i,a_i}^\top\boldsymbol{\theta}))]$, we may argue that $|\boldsymbol{x}^\top(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*)| = |\sum_{i=1}^t z_i\epsilon_i|$ where $z_i = (\boldsymbol{x}^\top\check{\mathbf{H}}_t^{-1}\boldsymbol{x}_{i,a_i})$, $\check{\mathbf{H}}_t = \sum_{i=1}^t \dot{\mu}(\boldsymbol{x}_{i,a_i}^\top\check{\boldsymbol{\theta}}_{t,i})\boldsymbol{x}_{i,a_i}\boldsymbol{x}_{i,a_i}^\top$ is an estimator-induced Hessian matrix, $\check{\boldsymbol{\theta}}_{t,i} = (1 - c_i)\widehat{\boldsymbol{\theta}}_t + c_i\boldsymbol{\theta}^*$ for some $c_i \in [0, 1]$ due to the mean-value theorem, and $\epsilon_i = r_i - \mu(\boldsymbol{x}_{i,a_i}^\top\boldsymbol{\theta}^*)$. Then one may attempt to apply a scalar-valued concentration inequality

to $|\sum_{i=1}^{t} z_i \epsilon_i|$ to yield a high-probability upper bound that does not depend on d. However, the difficulty is two-fold. First, under a general bandit policy, the realized rewards $r_1, \ldots, r_i$ affect future selected actions and observed rewards. Second, even if the policy is independent of the realized rewards, the definition of $\check{\mathbf{H}}_t$ shows that z_i still depends on future intermediate points $\check{\boldsymbol{\theta}}_{t,i+1}, \ldots, \check{\boldsymbol{\theta}}_{t,t}$. Therefore, scalar-valued concentration inequalities cannot be applied directly.

To circumvent this issue, the existing algorithms (Li et al., 2017; Jun et al., 2021) replaced $\check{\mathbf{H}}_t$ with the true Hessian $\mathbf{H}_t^* = \sum_{i=1}^{t} \dot{\mu}(\boldsymbol{x}_{i,a_i}^\top \boldsymbol{\theta}^*) \boldsymbol{x}_{i,a_i} \boldsymbol{x}_{i,a_i}^\top$, leading to

$$\left| \boldsymbol{x}^\top \left(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^* \right) \right| \leq \underbrace{\left| \sum_{i=1}^{t} \left(\boldsymbol{x}^\top \mathbf{H}_t^{*-1} \boldsymbol{x}_{i,a_i} \right) \cdot \epsilon_i \right|}_{\text{(noise)}} + \underbrace{\left| \boldsymbol{x}^\top \mathbf{H}_t^{*-1} \left(\check{\mathbf{H}}_t - \mathbf{H}_t^* \right) \left(\boldsymbol{\theta}^* - \widehat{\boldsymbol{\theta}}_t \right) \right|}_{\text{(remainder)}}. \tag{1}$$

For the (noise) term, they adapted the Auer framework (Auer, 2002) to construct level-wise buckets so that samples within each bucket are selected independently of the realized rewards. One can then apply a scalar-valued concentration inequality within each bucket. For the (remainder) term, the key is to keep $\check{\mathbf{H}}_t$ and $\mathbf{H}_t^*$ close to each other. Since each $\check{\boldsymbol{\theta}}_{t,i}$ lies between $\widehat{\boldsymbol{\theta}}_t$ and $\boldsymbol{\theta}^*$, it is enough to keep $\max_{i \in [t]} |\boldsymbol{x}_{i,a_i}^\top (\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*)|$ small (Jun et al., 2021). Applying the Cauchy-Schwarz inequality, we have

$$\max_{i \in [t]} \left| \boldsymbol{x}_{i,a_i}^\top (\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*) \right| \leq \max_{i \in [t]} \|\boldsymbol{x}_{i,a_i}\|_{\mathbf{V}_t^{-1}} \cdot \|\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\|_{\mathbf{V}_t} \leq \lambda_{\min}^{-1/2}(\mathbf{V}_t) \cdot \|\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\|_{\mathbf{V}_t} \tag{2}$$

where $\mathbf{V}_t = \sum_{i=1}^{t} \boldsymbol{x}_{i,a_i} \boldsymbol{x}_{i,a_i}^\top$. Here, to ensure that $\lambda_{\min}(\mathbf{V}_t)$ is sufficiently large, the existing methods apply an initial random-sampling phase whose length grows as a function of d.

However, the above approach relies on the assumption that $\lambda_{\min}(\mathbb{E}[\frac{1}{K} \sum_{a=1}^{K} \boldsymbol{x}_{i,a} \boldsymbol{x}_{i,a}^\top]) \geq \sigma_0^2$ and cannot avoid additional random-sampling rounds. These requirements are restrictive in two respects. First, they exclude settings where the features are concentrated around a lower-dimensional subspace. Second, the random-sampling phase is statistically inefficient and can be impractical. More fundamentally, both ingredients enforce uniform exploration across all d dimensions of the feature space. Consequently, the regret guarantees scale with the ambient dimension d and can be overly conservative in high-dimensional settings. To overcome these limitations, in the next section we propose a new algorithm that separates the roles of the collected samples, using one part to obtain an initial estimator and the other to carry out a one-step correction, thereby avoiding the context diversity assumption and a random-sampling phase.

5 Main idea and algorithm design

In this section, we introduce our algorithm, SupSplitLog, in Algorithm 1. As in the Auer framework, the algorithm maintains, for each level $s \in [S]$, a bucket $\Psi^{(s)}$ that stores the time indices of the arms selected at that level. However, the key distinction from the existing sup-style algorithms is that we partition the samples of each bucket into two disjoint subsets, namely the pilot set $P^{(s)}$ and the estimation set $E^{(s)}$, i.e., $\Psi^{(s)} = P^{(s)} \cup E^{(s)}$ and $P^{(s)} \cap E^{(s)} = \emptyset$. These two subsets are used to construct an initial-point estimator $\bar{\boldsymbol{\theta}}_{t-1}^{(s)}$ and a one-step corrected estimator $\hat{\boldsymbol{\theta}}_{t-1}^{(s)}$, respectively. A newly collected sample is assigned to $P^{(s)}$ or $E^{(s)}$ according to a level-dependent threshold $\tau^{(s)}$, which balances the need to improve the initial-point estimator against the need to perform an accurate one-step correction. This sample-splitting design is the key structural ingredient behind our analysis.

We set $\lambda = 1$. At level s in round t, we first construct the design matrices $\mathbf{V}_{P,t-1}^{(s)} = \kappa\lambda\mathbf{I} + \sum_{i \in P^{(s)}(t-1)} \boldsymbol{x}_{i,a_i} \boldsymbol{x}_{i,a_i}^\top$ and $\mathbf{V}_{E,t-1}^{(s)} = \kappa\lambda\mathbf{I} + \sum_{i \in E^{(s)}(t-1)} \boldsymbol{x}_{i,a_i} \boldsymbol{x}_{i,a_i}^\top$, and then we use the samples in $P^{(s)}(t-1)$ to compute the *initial-point estimator* $\bar{\boldsymbol{\theta}}_{t-1}^{(s)}$ as follows:

$$\bar{\boldsymbol{\theta}}_t^{(s)} = \arg\min_{\boldsymbol{\theta} \in \Theta} \left\{ \sum_{i \in P^{(s)}(t)} \left[-r_i \log\big(\mu(\boldsymbol{x}_{i,a_i}^\top \boldsymbol{\theta})\big) - (1 - r_i) \log\big(1 - \mu(\boldsymbol{x}_{i,a_i}^\top \boldsymbol{\theta})\big) \right] + \frac{\lambda}{2} \|\boldsymbol{\theta}\|_2^2 \right\}. \tag{3}$$

We then update $\bar{\boldsymbol{\theta}}_{t-1}^{(s)}$ using the samples in $E^{(s)}(t-1)$ through a *one-step correction*, which yields

$$\widehat{\boldsymbol{\theta}}_t^{(s)} = \bar{\boldsymbol{\theta}}_t^{(s)} + \left(\mathbf{H}_t^{(s)}(\bar{\boldsymbol{\theta}}_t^{(s)}) \right)^{-1} \boldsymbol{g}_t^{(s)}(\bar{\boldsymbol{\theta}}_t^{(s)}), \tag{4}$$

Algorithm 1 SupSplitLog

input $d, T, K, B, \kappa, \lambda, \delta$
initialize $S \leftarrow \lfloor \log_2 T \rfloor$, α as in (5), β as in (7)
initialize For $s \in [S]$, set $\tau^{(s)}$ as in (6), $\Psi^{(s)} \leftarrow \emptyset$, $P^{(s)} \leftarrow \emptyset$, $E^{(s)} \leftarrow \emptyset$, $\bar{\boldsymbol{\theta}}_0^{(s)} \leftarrow \mathbf{0}$

for $t = 1, \ldots, T$ **do**
  Initialize $A^{(1)} \leftarrow [K]$, $s \leftarrow 1$, $a_t \leftarrow$ `None`
  while $a_t =$ `None` **do**
    $\mathbf{V}_{P,t-1}^{(s)} \leftarrow \kappa\lambda\mathbf{I} + \sum_{i \in P^{(s)}(t-1)} \boldsymbol{x}_{i,a_i} \boldsymbol{x}_{i,a_i}^\top$
    $\mathbf{V}_{E,t-1}^{(s)} \leftarrow \kappa\lambda\mathbf{I} + \sum_{i \in E^{(s)}(t-1)} \boldsymbol{x}_{i,a_i} \boldsymbol{x}_{i,a_i}^\top$
    Update $\bar{\boldsymbol{\theta}}_{t-1}^{(s)}$ as in (3), $\widehat{\boldsymbol{\theta}}_{t-1}^{(s)}$ as in (4)
    For $a \in A^{(s)}$, compute $m_{t,a}^{(s)} \leftarrow \boldsymbol{x}_{t,a}^\top \widehat{\boldsymbol{\theta}}_{t-1}^{(s)}$, $w_{t,a}^{(s)} \leftarrow \alpha \|\boldsymbol{x}_{t,a}\|_{(\mathbf{V}_{E,t-1}^{(s)})^{-1}}$
    if $\exists a \in A^{(s)} \ : \ w_{t,a}^{(s)} > 2^{-s}$ **then** ▷ condition (a)
      Choose $a_t \leftarrow \arg\max_{a \in A^{(s)}} w_{t,a}^{(s)}$, receive r_t
      $\Psi^{(s)} \leftarrow \Psi^{(s)} \cup \{t\}$
      if $\|\boldsymbol{x}_{t,a_t}\|^2_{(\mathbf{V}_{P,t-1}^{(s)})^{-1}} > \tau^{(s)}$ **then** ▷ condition (a1)
        $P^{(s)} \leftarrow P^{(s)} \cup \{t\}$
      else ▷ condition (a2)
        $E^{(s)} \leftarrow E^{(s)} \cup \{t\}$
      end if
    else if $\forall a \in A^{(s)}$, $w_{t,a}^{(s)} \le 1/\sqrt{T}$ **then** ▷ condition (b)
      Choose, $a_t \leftarrow \arg\max_{a \in A^{(s)}} m_{t,a}^{(s)}$, receive r_t
    else ▷ condition (c)
      $A^{(s+1)} \leftarrow \{a \in A^{(s)} \ : \ m_{t,a}^{(s)} \ge \max_{a \in A^{(s)}} m_{t,a}^{(s)} - 2 \cdot 2^{-s}\}$
      $s \leftarrow s + 1$
    end if
  end while
end for

where the gradient $\boldsymbol{g}_t^{(s)}(\boldsymbol{\theta})$ and the Hessian $\mathbf{H}_t^{(s)}(\boldsymbol{\theta})$ of the loss function are given by

$$\boldsymbol{g}_t^{(s)}(\boldsymbol{\theta}) = \sum_{i \in E^{(s)}(t)} \big(r_i - \mu(\boldsymbol{x}_{i,a_i}^\top \boldsymbol{\theta})\big)\boldsymbol{x}_{i,a_i} - \lambda\boldsymbol{\theta}, \quad \mathbf{H}_t^{(s)}(\boldsymbol{\theta}) = \lambda\mathbf{I} + \sum_{i \in E^{(s)}(t)} \dot{\mu}(\boldsymbol{x}_{i,a_i}^\top \boldsymbol{\theta}) \boldsymbol{x}_{i,a_i} \boldsymbol{x}_{i,a_i}^\top.$$

Remark 5 (Discussion on the one-step estimator $\widehat{\boldsymbol{\theta}}$)**.** *A one-step estimator starts from an initial estimator and applies a single Newton-type correction step, using the gradient and curvature information, instead of fully re-optimizing the objective. This idea is classical in asymptotic and semiparametric statistics, where a single correction is often enough to recover much of the accuracy of the exact estimator (Bickel et al., 1993; Van der Vaart, 2000). We adapt this idea in a bandit-specific way through sample splitting. The initial-point estimator $\bar{\boldsymbol{\theta}}_t^{(s)}$ is computed from the pilot set, while the one-step correction $\widehat{\boldsymbol{\theta}}_t^{(s)}$ is carried out on the estimation set. As a result, the matrix quantity $\mathbf{H}_t^{(s)}(\bar{\boldsymbol{\theta}}_t^{(s)})$ is determined by the pilot estimator and the observed contexts, and it is separated from the realized rewards in the estimation set. This separation is the key to our concentration argument.*

Remark 6 (Comment on regret analysis)**.** *An additional residual term appears in regret analysis (see the (remainder) term in Lemma 9), due to the correction step applied to the exact estimator for the estimation set. Our splitting rule, through conditions (a1) and (a2), is designed to control this, by ensuring that the pilot samples make the initial-point estimator sufficiently accurate before the correction step. Thus, the same split simultaneously enables concentration of the main stochastic term and keeps the residual term small. We explain more details in Section 6.*

For each arm $a \in A^{(s)}$, we compute $m_{t,a}^{(s)} = \boldsymbol{x}_{t,a}^\top \widehat{\boldsymbol{\theta}}_{t-1}^{(s)}$ and $w_{t,a}^{(s)} = \alpha \|\boldsymbol{x}_{t,a}\|_{(\mathbf{V}_{E,t-1}^{(s)})^{-1}}$, where

$$\alpha = 2(\alpha_1 + \alpha_2), \quad \alpha_1 = \sqrt{2L_\mu \kappa^2 \log(4TSK/\delta)} + \sqrt{\kappa/(9\lambda)} \log(4TSK/\delta), \quad \alpha_2 = B\sqrt{\kappa\lambda}. \quad (5)$$

The confidence level associated with level s is set to 2^{-s}. The algorithm then proceeds as follows: [condition (a)] If the uncertainty with respect to $E^{(s)}$ is still large in the direction of $\boldsymbol{x}_{t,a}$, i.e., $w_{t,a}^{(s)} > 2^{-s}$, then the algorithm explores and adds the current time index to the bucket, $\Psi^{(s)} \leftarrow \Psi^{(s)} \cup \{t\}$. [condition (b)] If condition (a) does not hold and the uncertainty of every action is below $1/\sqrt{T}$, then we stop and proceed to exploitation. [condition (c)] Otherwise, if the uncertainty of every action is below the current confidence level, then the algorithm eliminates arms that cannot be optimal, namely those satisfying $m_{t,a}^{(s)} < \max_{a \in A^{(s)}} m_{t,a}^{(s)} - 2 \cdot 2^{-s}$, and moves to the next level, i.e., $s \leftarrow s+1$.

When condition (a) is triggered, we further decide whether the newly collected sample should be assigned to the pilot set or the estimation set. [condition (a1)] If the uncertainty with respect to $P^{(s)}$ in the direction $\boldsymbol{x}_{t,a_t}$ exceeds the level-dependent threshold $\tau^{(s)}$, i.e. $\|\boldsymbol{x}_{t,a_t}\|^2_{(\mathbf{V}_{P,t-1}^{(s)})^{-1}} > \tau^{(s)}$,

$$\tau^{(s)} = \frac{1}{\beta^2} \min\left\{1, \frac{2^{-s}}{\sqrt{32 L_\mu \kappa d \log(1+T/(\kappa\lambda d))}}\right\}, \tag{6}$$

$$\beta = \kappa\sqrt{0.5 d \log(1+T/(\kappa\lambda d)) + 0.5\log(2S/\delta)} + B\sqrt{\kappa\lambda}, \tag{7}$$

then we add the corresponding time index to $P^{(s)}$ to improve the accuracy of the initial-point estimator. [condition (a2)] Otherwise, we add it to $E^{(s)}$ so that it can be used for the one-step correction.

We next state the main regret bound for SupSplitLog. The full proof is deferred to Section 6.2.

Theorem 7. *With probability at least $1-\delta$, the regret of Algorithm 1 is upper bounded by*

$$R_T \le 16 L_\mu d \log(1+T/(\kappa\lambda d))\beta^2 \left(1 + S\sqrt{32 L_\mu \kappa d \log(1+T/(\kappa\lambda d))}\right) + 8 L_\mu \alpha \sqrt{2STd\log(1+T/(\kappa\lambda d))} + 2L_\mu\sqrt{T}.$$

Remark 8 (Data-dependent regret of Theorem 7)**.** *While Theorem 7 is stated in terms of the ambient dimension d, this dependence can be refined. By allowing β and $\tau^{(s)}$ to be chosen adaptively based on the data observed up to round t, e.g., through $\beta_t^{(s)}$ and $\tau_t^{(s)}$, one can refine the bound so that d is replaced by a data-dependent complexity measure of $\log \frac{\det \mathbf{V}_T}{\det \kappa\lambda \mathbf{I}}$ where $\mathbf{V}_T = \sum_{t=1}^T \boldsymbol{x}_{t,a_t}\boldsymbol{x}_{t,a_t}^\top + \kappa\lambda\mathbf{I}$. We defer the details to Section B.*

6 Analysis

In this section, we analyze Theorem 7. We first establish a uniform concentration result for the logit estimates and then use it to derive the regret bound. Full proofs are deferred to Section D.

6.1 Uniform concentration of the logit estimates

We first establish a dimension-free high-probability upper bound on $|\boldsymbol{x}_{t,a}^\top(\widehat{\boldsymbol{\theta}}_{t-1}^{(s)} - \boldsymbol{\theta}^*)|$ simultaneously for all $t \in [T]$, $s \in [S]$, $a \in [K]$.

One-step estimator decomposition. By the definitions of the initial-point estimator and the one-step corrected estimator in (3) and (4), we obtain the following decomposition.

Lemma 9. *For all $\boldsymbol{x} \in \mathbb{R}^d$, $t \in [0, T-1]$, $s \in [S]$, we have*

$$\left|\boldsymbol{x}^\top(\widehat{\boldsymbol{\theta}}_t^{(s)} - \boldsymbol{\theta}^*)\right| \le \underbrace{\left|\boldsymbol{x}^\top\left(\left(\mathbf{H}_t^{(s)}(\bar{\boldsymbol{\theta}}_t^{(s)})\right)^{-1} \sum_{i \in E^{(s)}(t)} \epsilon_i \boldsymbol{x}_{i,a_i}\right)\right|}_{\text{(noise)}} + \underbrace{\left|\boldsymbol{x}^\top\left(\mathbf{H}_t^{(s)}(\bar{\boldsymbol{\theta}}_t^{(s)})\right)^{-1}\lambda\boldsymbol{\theta}^*\right|}_{\text{(bias)}}$$

$$+ \underbrace{\left|\boldsymbol{x}^\top\left(\mathbf{H}_t^{(s)}(\bar{\boldsymbol{\theta}}_t^{(s)})\right)^{-1}\left(\sum_{i \in E^{(s)}(t)} \left(\dot{\mu}(\boldsymbol{x}_{i,a_i}^\top \check{\boldsymbol{\theta}}_{t,i}^{(s)}) - \dot{\mu}(\boldsymbol{x}_{i,a_i}^\top \bar{\boldsymbol{\theta}}_t^{(s)})\right)\boldsymbol{x}_{i,a_i}\boldsymbol{x}_{i,a_i}^\top\right)(\boldsymbol{\theta}^* - \bar{\boldsymbol{\theta}}_t^{(s)})\right|}_{\text{(remainder)}},$$

where $\check{\boldsymbol{\theta}}_{t,i}^{(s)} = c_i\boldsymbol{\theta}^* + (1-c_i)\bar{\boldsymbol{\theta}}_t^{(s)}$ is induced by the mean-value theorem for some $c_i \in [0,1]$.

Noise term. A key feature of our one-step estimator is that, in the (noise) term in Lemma 9, the Hessian $\mathbf{H}_t^{(s)}(\bar{\boldsymbol{\theta}}_t^{(s)})$ and the noise ϵ_i are built from disjoint sample subsets. This separation allows us to condition on a σ-field with respect to which the coefficients are measurable, while the noises in $E^{(s)}(t)$ remain conditionally zero-mean and conditionally independent. Define

$$\begin{aligned}\mathcal{H}_{s,t} \coloneqq &\sigma\big(\{\boldsymbol{x}_{i,a}\}_{i\in[t],a\in[K]},\ \{r_j\}_{j\in\Psi^{(1)}(t)\cup\cdots\cup\Psi^{(s-1)}(t)}\big)\\ &\vee\sigma\big(\{\boldsymbol{x}_{i,a_i}\}_{i\in E^{(s)}(t)}\big)\vee\sigma\big(\{\boldsymbol{x}_{i,a_i},r_i\}_{i\in P^{(s)}(t)}\big).\end{aligned}$$

Then, for all $s\in[S]$ and $t\in[T]$, conditional on $\mathcal{H}_{s,t}$, $\{\epsilon_i\}_{i\in E^{(s)}(t)}$ are independent and $\{c_i\}_{i\in E^{(s)}(t)}$ are deterministic (see Lemma 27). Therefore, unlike the previous approach, we can directly apply a scalar-valued concentration inequality (Bernstein inequality, Lemma 33).

Lemma 10. *Fix $\boldsymbol{x}\in\mathbb{R}^d$, $s\in[S]$, $t\in[0,T-1]$. Then with probability at least $1-\delta$, we have*

$$(\text{noise})\le\left(\sqrt{2L_\mu\kappa^2\log(2/\delta)}+\frac{\sqrt{\kappa}\log(2/\delta)}{3\sqrt{\lambda}}\right)\|\boldsymbol{x}\|_{(\mathbf{V}_{E,t}^{(s)})^{-1}}.$$

Remainder term. As in (1) and (2), controlling the (remainder) term reduces to control

$$r_t^{(s)}\coloneqq\max_{i\in E^{(s)}(t)}\left|\boldsymbol{x}_{i,a_i}^\top(\bar{\boldsymbol{\theta}}_t^{(s)}-\boldsymbol{\theta}^*)\right|.$$

Let $\mathcal{E}_1$ denote the event that $\|\bar{\boldsymbol{\theta}}_t^{(s)}-\boldsymbol{\theta}^*\|_{\mathbf{V}_{P,t}^{(s)}}\le\beta$ simultaneously for all $s\in[S]$ and $t\in[0,T-1]$. By the well-known self-normalized bound of Abbasi-Yadkori et al. (2011), this event holds with high probability (see Lemma 30). The next lemma shows that $r_t^{(s)}$ is uniformly bounded on this event.

Lemma 11. *On the event $\mathcal{E}_1$, for all $t\in[0,T-1]$, $s\in[S]$, we have $r_t^{(s)}\le\beta\sqrt{\tau^{(s)}}\le 1$.*

Proof of Lemma 11. The event $\{i\in E^{(s)}(t)\}$ implies that in round i, condition (a2) is satisfied,

$$\|\boldsymbol{x}_{i,a_i}\|^2_{(\mathbf{V}_{P,i-1}^{(s)})^{-1}}\le\tau^{(s)}\tag{8}$$

Therefore, on the event $\mathcal{E}_1$, we have

$$\begin{aligned}r_t^{(s)}&\le\max_{i\in E^{(s)}(t)}\|\boldsymbol{x}_{i,a_i}\|_{(\mathbf{V}_{P,i-1}^{(s)})^{-1}}\cdot\|\bar{\boldsymbol{\theta}}_t^{(s)}-\boldsymbol{\theta}^*\|_{\mathbf{V}_{P,i-1}^{(s)}} && \text{(Cauchy-Schwarz)}\\ &\le\|\bar{\boldsymbol{\theta}}_t^{(s)}-\boldsymbol{\theta}^*\|_{\mathbf{V}_{P,t}^{(s)}}\cdot\max_{i\in E^{(s)}(t)}\|\boldsymbol{x}_{i,a_i}\|_{(\mathbf{V}_{P,i-1}^{(s)})^{-1}} && (\mathbf{V}_{P,t}^{(s)}\succeq\mathbf{V}_{P,i-1}^{(s)},\forall i\in E^{(s)}(t)\subset[t])\\ &\le\|\bar{\boldsymbol{\theta}}_t^{(s)}-\boldsymbol{\theta}^*\|_{\mathbf{V}_{P,t}^{(s)}}\cdot\sqrt{\tau^{(s)}} && ((8))\\ &\le\beta\cdot\sqrt{\tau^{(s)}}\le 1, && (\mathcal{E}_1)\end{aligned}$$

where the last inequality follows from the definition of $\tau^{(s)}$. □

Lemma 12. *On the event $\mathcal{E}_1$, for all $\boldsymbol{x}\in\mathbb{R}^d$, $t\in[0,T-1]$, $s\in[S]$, $(\text{remainder})\le\frac{\alpha}{2}\|\boldsymbol{x}\|_{(\mathbf{V}_{E,t}^{(s)})^{-1}}$.*

Final bound. The (bias) term arises from the regularization parameter λ and can be bounded straightforwardly, so we omit its proof. Combining Lemmas 9, 10 and 12, together with the bound on the (bias) term, yields the following uniform concentration result.

Theorem 13. *With probability at least $1-\delta$, simultaneously for all $t\in[T]$, $s\in[S]$, $a\in[K]$, we have $|m_{t,a}^{(s)}-\boldsymbol{x}_{t,a}^\top\boldsymbol{\theta}^*|\le w_{t,a}^{(s)}$.*

6.2 Regret analysis

We now prove Theorem 7. Let $\mathcal{E}$ denote the good event when Theorem 13 holds. The following lemma is a standard consequence of the Auer framework and are included for completeness.

Lemma 14. *On the good event $\mathcal{E}$, the per-round regret is upper bounded by*

$$\mu(\boldsymbol{x}_{t,a_t^*}^\top\boldsymbol{\theta}^*)-\mu(\boldsymbol{x}_{t,a_t}^\top\boldsymbol{\theta}^*)\le\begin{cases}8L_\mu 2^{-s_t} & \text{if } a_t \text{ is selected by condition (a),}\\ 2L_\mu/\sqrt{T} & \text{if } a_t \text{ is selected by condition (b).}\end{cases}$$

The next lemma bounds the cardinality of the pilot set.

Lemma 15. *For all $s \in [S]$, we have $|P^{(s)}(T)| \leq 2d\log(1+T/(\kappa\lambda d))(\tau^{(s)})^{-1}$.*

We are now ready to prove Theorem 7.

Proof of Theorem 7. Let $\Delta_t := \mu(\boldsymbol{x}_{t,a_t^*}^\top \boldsymbol{\theta}^*) - \mu(\boldsymbol{x}_{t,a_t}^\top \boldsymbol{\theta}^*)$. Then, we have

$$R_T = \underbrace{\sum_{s=1}^{S} \sum_{t \in P^{(s)}(T)} \Delta_t}_{A_1} + \underbrace{\sum_{s=1}^{S} \sum_{t \in E^{(s)}(T)} \Delta_t}_{A_2} + \underbrace{\sum_{t \in ([T] \setminus (\Psi^{(1)}(T) \cup \cdots \cup \Psi^{(S)}(T)))} \Delta_t}_{A_3}.$$

For the term A_1

$$\begin{aligned}
A_1 &\leq \sum_{s=1}^{S} \sum_{t \in P^{(s)}(T)} 8L_\mu 2^{-s_t} \leq 8L_\mu \sum_{s=1}^{S} \left|P^{(s)}(T)\right| 2^{-s} && \text{(Lemma 14)} \\
&\leq 8L_\mu \sum_{s=1}^{S} \frac{2^{-s} 2d\log(1+T/(\kappa\lambda d))}{\tau^{(s)}} && \text{(Lemma 15)} \\
&\leq 8L_\mu \sum_{s=1}^{S} 2d\log(1+T/(\kappa\lambda d))\beta^2 \left(2^{-s} + \sqrt{32L_\mu \kappa d \log(1+T/(\kappa\lambda d))}\right) && \text{((6))} \\
&\leq 16L_\mu d\log(1+T/(\kappa\lambda d))\beta^2 \left(1 + S\sqrt{32L_\mu \kappa d \log(1+T/(\kappa\lambda d))}\right).
\end{aligned}$$

For the term A_2, we have $A_2 \leq \sum_{s=1}^{S} \sum_{t \in E^{(s)}(T)} 8L_\mu 2^{-s_t} \leq 8L_\mu \sum_{s=1}^{S} \sum_{t \in E^{(s)}(T)} w_{t,a_t}^{(s)} = 8L_\mu \alpha \sum_{s=1}^{S} \sum_{t \in E^{(s)}(T)} \|\boldsymbol{x}_{t,a_t}\|_{(\mathbf{V}_{E,t-1}^{(s)})^{-1}}$, where the first inequality follows from Lemma 14, and the second inequality follows from the condition (a). Denote $E^{(s)}(T)$ by $E^{(s)}$, and enumerate its elements in chronological order as $t^{(1)}, \ldots, t^{(|E^{(s)}|)}$. Let $\mathbf{V}_0^{(s)} = \kappa\lambda\mathbf{I}$, $\mathbf{V}_j^{(s)} = \mathbf{V}_{j-1}^{(s)} + \boldsymbol{x}_{t^{(j)},a_{t^{(j)}}} \boldsymbol{x}_{t^{(j)},a_{t^{(j)}}}^\top$. Then, we can proceed as

$$\begin{aligned}
A_2 &\leq 8L_\mu \alpha \sum_{s=1}^{S} \sum_{j=1}^{|E^{(s)}|} \min\left\{1, \left\|\boldsymbol{x}_{t^{(j)},a_t^{(j)}}\right\|_{(\mathbf{V}_{j-1}^{(s)})^{-1}}\right\} && (\lambda \geq 1/\kappa,\ \|\boldsymbol{x}_{t^{(j)},a_{t^{(j)}}}\|_{(\mathbf{V}_{j-1}^{(s)})^{-1}} \leq 1) \\
&\leq 8L_\mu \alpha \sqrt{S \sum_{s=1}^{S} |E^{(s)}| \sum_{j=1}^{|E^{(s)}|} \min\left\{1, \left\|\boldsymbol{x}_{t^{(j)},a_t^{(j)}}\right\|^2_{(\mathbf{V}_{j-1}^{(s)})^{-1}}\right\}} && \text{(Cauchy-Schwarz)} \\
&\leq 8L_\mu \alpha \sqrt{S \sum_{s=1}^{S} |E^{(s)}| 2d\log(1+T/(\kappa\lambda d))} && \text{(Lemma 34)} \\
&\leq 8L_\mu \alpha \sqrt{2STd\log(1+T/(\kappa\lambda d))}.
\end{aligned}$$

For the term A_3, following from Lemma 14, we have $A_3 \leq T \times 2L_\mu/\sqrt{T} = 2L_\mu\sqrt{T}$. Substituting term A_1, term A_2, and term A_3 back completes the proof. □

7 Experiments

In this section, we empirically evaluate the performance of SupSplitLog. In all experiments, we use the refined version in Algorithm 2, which replaces β and $\tau^{(s)}$ with the data-dependent quantities $\beta_t^{(s)}$ and $\tau_t^{(s)}$; see Section B for details. We compare SupSplitLog with the three baselines listed in Table 1: SupCB-GLM (Li et al., 2017), SupLogistic (Jun et al., 2021), and DDRTS-GLM (Kim et al., 2023). For each algorithm, we report the mean cumulative regret with ± 1 standard deviation.

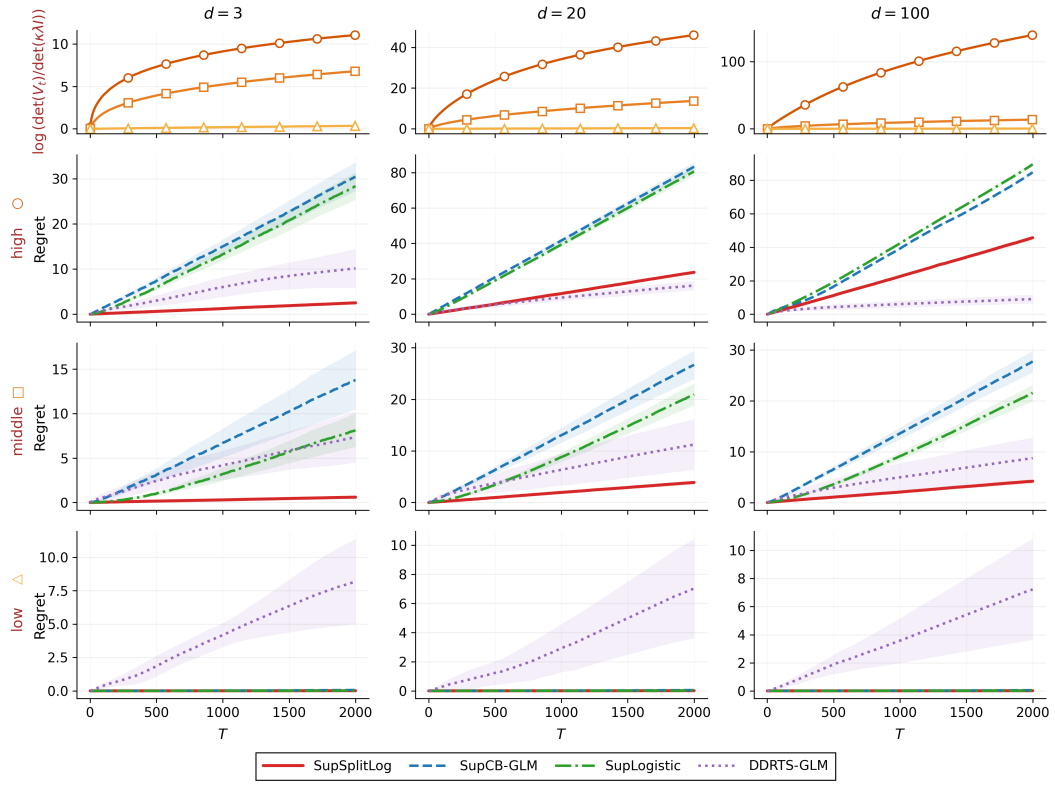


Figure 1: Comparison across ambient dimension d and realized geometry. Top row: trajectories of $\log(\det \mathbf{V}_t / \det \kappa\lambda\mathbf{I})$ under three regimes, `high`, `middle`, and `low`. Bottom rows: cumulative regret of SupSplitLog, SupCB-GLM, SupLogistic, and DDRTS-GLM on the corresponding instances.

To study the dependence on the ambient feature dimension d, we vary both the value of d and the geometry of the incoming features. For each value of d, we consider three regimes, labeled `high`, `middle`, and `low`. These regimes correspond to different effective subspace structures of the incoming features and hence different values of $\log \frac{\det \mathbf{V}_t}{\det \kappa\lambda\mathbf{I}}$, where $\mathbf{V}_t = \kappa\lambda\mathbf{I} + \sum_{i=1}^{t} \sum_{a\in[K]} \boldsymbol{x}_{i,a}\boldsymbol{x}_{i,a}^\top$.

Figure 1 reports the results. The top row shows the trajectory of $\log \frac{\det \mathbf{V}_t}{\det \kappa\lambda\mathbf{I}}$ over time under the three regimes, and the bottom rows show the corresponding cumulative regrets. In the `high` regime, where the incoming features effectively span the ambient space, DDRTS-GLM is competitive. In the `middle` and `low` regimes, where the incoming features concentrate on a lower-dimensional structure, SupSplitLog performs most favorably. These results highlight the improved dimension dependence of our method and show that its advantage is most visible when the effective feature dimension is much smaller than the ambient dimension. Full implementation details of the bandit environments and baseline algorithms are deferred to Section C.

8 Conclusion and future work

We propose SupSplitLog, which, to the best of our knowledge, is the first algorithm for logistic bandits that achieves $\widetilde{\mathcal{O}}(\sqrt{dT})$ regret without any context diversity assumption. The key idea is a sample-splitting design that separates the construction of an initial-point estimator from a one-step correction, yielding a dimension-free high-probability upper bound on the estimation error. This leads to a regret bound with improved dependence on d over previous work.

There are several interesting directions for future work. One is to reduce the $d^{2.5}$ factor in the non-leading term of our regret bound. A sufficiently improved non-leading dependence on d may make it possible to obtain sublinear regret even in regimes where d grows proportionally with T. Another direction is to improve the dependence on κ, which remains a major theme in logistic bandits.

Acknowledgements

This work was supported by the Global-LAMP Program of the National Research Foundation of Korea (NRF) grant funded by the Ministry of Education (No. RS-2023-00301976). The authors declare no competing interests.

References

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. *Advances in neural information processing systems*, 24, 2011.

Marc Abeille, Louis Faury, and Clément Calauzènes. Instance-wise minimax-optimal algorithms for logistic bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 3691–3699. PMLR, 2021.

Arya Akhavan, Amitis Shidani, Alex Ayoub, and David Janz. Bernstein-type dimension-free concentration for self-normalised martingales. *arXiv preprint arXiv:2507.20982*, 2025.

Peter Auer. Using confidence bounds for exploitation-exploration trade-offs. *Journal of machine learning research*, 3(Nov):397–422, 2002.

Seoungbin Bae and Dabeen Lee. Neural logistic bandits. *arXiv preprint arXiv:2505.02069*, 2025.

Seoungbin Bae, Garyeong Kang, and Dabeen Lee. Queue length regret bounds for contextual queueing bandits. *arXiv preprint arXiv:2601.19300*, 2026.

Viktor Bengs, Róbert Busa-Fekete, Adil El Mesaoudi-Paul, and Eyke Hüllermeier. Preference-based online learning with dueling bandits: A survey. *Journal of Machine Learning Research*, 22(7): 1–108, 2021.

Peter J Bickel, Chris AJ Klaassen, Peter J Bickel, Ya'acov Ritov, J Klaassen, Jon A Wellner, and YA'Acov Ritov. *Efficient and adaptive estimation for semiparametric models*, volume 4. Johns Hopkins University Press, Baltimore, 1993.

Pierre Boudart, Pierre Gaillard, and Alessandro Rudi. Enjoying non-linearity in multinomial logistic bandits: A minimax-optimal algorithm. In *37th International Conference on Algorithmic Learning Theory*, 2026.

Sayak Ray Chowdhury and Aditya Gopalan. On kernelized multi-armed bandits. In *International Conference on Machine Learning*, pages 844–853. PMLR, 2017.

Wei Chu, Lihong Li, Lev Reyzin, and Robert Schapire. Contextual bandits with linear payoff functions. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 208–214. JMLR Workshop and Conference Proceedings, 2011.

Qiwei Di, Tao Jin, Yue Wu, Heyang Zhao, Farzad Farnoud, and Quanquan Gu. Variance-aware regret bounds for stochastic contextual dueling bandits. *arXiv preprint arXiv:2310.00968*, 2023.

Shi Dong, Tengyu Ma, and Benjamin Van Roy. On the performance of thompson sampling on logistic bandits. In *Conference on Learning Theory*, pages 1158–1160. PMLR, 2019.

Louis Faury, Marc Abeille, Clément Calauzènes, and Olivier Fercoq. Improved optimistic algorithms for logistic bandits. In *International Conference on Machine Learning*, pages 3052–3060. PMLR, 2020.

Louis Faury, Marc Abeille, Kwang-Sung Jun, and Clément Calauzènes. Jointly efficient and optimal algorithms for logistic bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 546–580. PMLR, 2022.

Sarah Filippi, Olivier Cappe, Aurélien Garivier, and Csaba Szepesvári. Parametric bandits: The generalized linear case. *Advances in neural information processing systems*, 23, 2010.

Yichao Fu, Siqi Zhu, Runlong Su, Aurick Qiao, Ion Stoica, and Hao Zhang. Efficient llm scheduling by learning to rank. *Advances in Neural Information Processing Systems*, 37:59006–59029, 2024.

Amaury Gouverneur, Tobias Oechtering, and Mikael Skoglund. An information-theoretic analysis of thompson sampling for logistic bandit problems. *Transactions on Machine Learning Research*, 2026. ISSN 2835-8856. URL https://openreview.net/forum?id=94y5XfiJ7N.

Tanmay Goyal and Gaurav Sinha. Efficient algorithms for logistic contextual slate bandits with bandit feedback. *arXiv preprint arXiv:2506.13163*, 2025.

Kwang-Sung Jun, Aniruddha Bhargava, Robert Nowak, and Rebecca Willett. Scalable generalized linear bandits: Online computation and hashing. *Advances in Neural Information Processing Systems*, 30, 2017.

Kwang-Sung Jun, Lalit Jain, Blake Mason, and Houssam Nassif. Improved confidence bounds for the linear logistic model and applications to bandits. In *International Conference on Machine Learning*, pages 5148–5157. PMLR, 2021.

Yue Kang, Mingshuo Liu, Bongsoo Yi, Jing Lyu, Zhi Zhang, Doudou Zhou, and Yao Li. Single index bandits: Generalized linear contextual bandits with unknown reward functions. *arXiv preprint arXiv:2506.12751*, 2025.

Parnian Kassraie and Andreas Krause. Neural contextual bandits without regret. In *International Conference on Artificial Intelligence and Statistics*, pages 240–278. PMLR, 2022.

Jung-hun Kim and Min-hwan Oh. Queueing matching bandits with preference feedback. *Advances in Neural Information Processing Systems*, 37:64649–64702, 2024.

Wonyoung Kim, Kyungbok Lee, and Myunghee Cho Paik. Double doubly robust thompson sampling for generalized linear contextual bandits. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 8300–8307, 2023.

Joongkyu Lee and Min hwan Oh. Improved online confidence bounds for multinomial logistic bandits. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=FXQ09DpwXt.

Junghyun Lee, Se-Young Yun, and Kwang-Sung Jun. Improved regret bounds of (multinomial) logistic bandits via regret-to-confidence-set conversion. In *International Conference on Artificial Intelligence and Statistics*, pages 4474–4482. PMLR, 2024a.

Junghyun Lee, Se-Young Yun, and Kwang-Sung Jun. A unified confidence sequence for generalized linear models, with applications to bandits. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024b. URL https://openreview.net/forum?id=MDdOQayWTA.

Lihong Li, Yu Lu, and Dengyong Zhou. Provably optimal algorithms for generalized linear contextual bandits. In *International Conference on Machine Learning*, pages 2071–2080. PMLR, 2017.

Yingkai Li, Yining Wang, and Yuan Zhou. Nearly minimax-optimal regret for linearly parameterized bandits. In *Conference on Learning Theory*, pages 2173–2174. PMLR, 2019.

Yingkai Li, Yining Wang, Xi Chen, and Yuan Zhou. Tight regret bounds for infinite-armed linear contextual bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 370–378. PMLR, 2021.

Shuai Liu, Alex Ayoub, Flore Sentenac, Xiaoqi Tan, and Csaba Szepesvári. Almost free: Self-concordance in natural exponential families and an application to bandits. *Advances in Neural Information Processing Systems*, 37:70828–70874, 2024.

Blake Mason, Kwang-Sung Jun, and Lalit Jain. An experimental design approach for regret minimization in logistic bandits. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 7736–7743, 2022.

Alberto Maria Metelli, Simone Drago, and Marco Mussi. Generalized kernelized bandits: A novel self-normalized bernstein-like dimension-free inequality and regret bounds. *arXiv preprint arXiv:2508.01681*, 2025.

Min-hwan Oh and Garud Iyengar. Multinomial logit contextual bandits: Provable optimality and practicality. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 9205–9213, 2021.

Matteo Papini, Andrea Tirinzoni, Marcello Restelli, Alessandro Lazaric, and Matteo Pirotta. Leveraging good representations in linear contextual bandits. In *International Conference on Machine Learning*, pages 8371–8380. PMLR, 2021.

Barna Pásztor, Parnian Kassraie, and Andreas Krause. Bandits with preference feedback: A stackelberg game perspective. *Advances in Neural Information Processing Systems*, 37:11997–12034, 2024.

Aadirupa Saha. Optimal algorithms for stochastic contextual preference bandits. *Advances in Neural Information Processing Systems*, 34:30050–30062, 2021.

Ayush Sawarni, Nirjhar Das, Siddharth Barman, and Gaurav Sinha. Generalized linear bandits with limited adaptivity. *Advances in Neural Information Processing Systems*, 37:8329–8369, 2024.

Niranjan Srinivas, Andreas Krause, Sham M Kakade, and Matthias W Seeger. Information-theoretic regret bounds for gaussian process optimization in the bandit setting. *IEEE transactions on information theory*, 58(5):3250–3265, 2012.

Sattar Vakili, Kia Khezeli, and Victor Picheny. On information gain and regret bounds in gaussian process bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 82–90. PMLR, 2021.

Michal Valko, Nathaniel Korda, Rémi Munos, Ilias Flaounas, and Nelo Cristianini. Finite-time analysis of kernelised contextual bandits. *arXiv preprint arXiv:1309.6869*, 2013.

Aad W Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.

Arun Verma, Zhongxiang Dai, Xiaoqiang Lin, Patrick Jaillet, and Bryan Kian Hsiang Low. Neural dueling bandits: Preference-based optimization with human feedback. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=VELhv9BBfn.

Weiqiang Wu, Jing Yang, and Cong Shen. Stochastic linear contextual bandits with diverse contexts. In *International Conference on Artificial Intelligence and Statistics*, pages 2392–2401. PMLR, 2020.

Yu-Jie Zhang and Masashi Sugiyama. Online (multinomial) logistic bandit: Improved regret and constant computation cost. *Advances in Neural Information Processing Systems*, 36:29741–29782, 2023.

Yu-Jie Zhang, Sheng-An Xu, Peng Zhao, and Masashi Sugiyama. Generalized linear bandits: Almost optimal regret with one-pass update. *arXiv preprint arXiv:2507.11847*, 2025.

Dongruo Zhou, Lihong Li, and Quanquan Gu. Neural contextual bandits with ucb-based exploration. In *International conference on machine learning*, pages 11492–11502. PMLR, 2020.

A Additional related work

Logistic bandits naturally model sequential decision-making problems with binary feedback, and have therefore been studied in a range of application domains. Recent examples include queueing and service systems (Kim and Oh, 2024; Fu et al., 2024; Bae et al., 2026), as well as preference-based and pairwise-feedback settings (Verma et al., 2025; Pásztor et al., 2024; Bengs et al., 2021; Di et al., 2023).

One of the main directions in theoretical research on logistic bandits has focused on infinite-armed generalized linear and logistic bandit problems. Filippi et al. (2010) initiated this line by establishing a $\widetilde{\mathcal{O}}\left(d\sqrt{T}\right)$ regret bound for generalized linear bandits. Subsequent work improved the dependence on problem-dependent quantities such as κ, derived sharper instance-dependent guarantees, and extended the analysis to broader generalized linear settings (Faury et al., 2020; Abeille et al., 2021; Faury et al., 2022; Lee et al., 2024b; Sawarni et al., 2024; Liu et al., 2024). There are also several related directions, such as developing jointly efficient online algorithms (Jun et al., 2017; Zhang and Sugiyama, 2023; Zhang et al., 2025), improving the dependence on the norm bound B (Lee et al., 2024a; Lee and hwan Oh, 2025), designing Thompson sampling-based algorithms (Dong et al., 2019; Gouverneur et al., 2026), extending logistic bandits to multinomial logit models (Boudart et al., 2026), and taking care of unknown reward functions (Kang et al., 2025).

Another direction is to provide improved regret bounds for the finite-armed setting where at each round the agent takes a set of K actions. This line of research is also closely related to finite-action linear bandits. In particular, the Auer framework (Auer, 2002) and its refinements (Valko et al., 2013; Chu et al., 2011; Li et al., 2019, 2021) provide the basic independent-sample construction on which later finite-armed logistic bandit algorithms are built. For the K-armed logistic bandit setting, Li et al. (2017) came up with an algorithm with a regret bound of order $\widetilde{O}(\sqrt{dT})$. Later, Jun et al. (2021) refined the algorithm and improved the dependence on κ on the regret bound. More recently, Kim et al. (2023) developed a Thompson sampling-based algorithm, which improved practicality and numerical performance. However, as mentioned earlier, these works heavily rely on context-diversity assumptions. We note that context-diversity assumptions also appear in several related generalized linear settings (Wu et al., 2020; Papini et al., 2021; Goyal and Sinha, 2025).

In fact, context-diversity assumptions are the main hurdle in extending the K-armed logistic bandit frameworks to high-dimensional settings. For K-armed linear bandits, we may obtain a $\widetilde{O}(\sqrt{dT})$ regret bound without a context-diversity assumption, based on which we can extend such result to richer high-dimensional models such as kernelized and neural bandits (Zhou et al., 2020; Kassraie and Krause, 2022). In contrast, to our knowledge, existing kernelized and neural extensions of logistic bandit models do not admit an analogous $\widetilde{O}(\sqrt{dT})$ regret bound. The obstacle is that the context-diversity assumptions used in the K-armed setting do not extend directly to kernelized and neural settings without incurring direct dependence on the feature dimension d (Verma et al., 2025; Bae and Lee, 2025; Akhavan et al., 2025; Metelli et al., 2025).

Lastly, we note that there are other finite-action settings with similar feedback models, including dueling bandits (Saha, 2021), multinomial logit bandits (Oh and Iyengar, 2021), and the fixed-arm set setting (Mason et al., 2022). However, these works also assume some richness and diversity conditions or consider more restricted settings. Taken together, these directions show that removing context diversity assumptions while retaining $\widetilde{\mathcal{O}}(\sqrt{dT})$ regret in finite-armed logistic bandits has remained open prior to our work.

B Refined algorithm for data-dependent regret

As mentioned in Remark 8, in this section we present a refined version of Algorithm 1, illustrated in Algorithm 2, which yields a data-dependent regret bound.

Algorithm 2 SupSplitLog with data-dependent regret

input $d, T, K, B, \kappa, \lambda, \delta$
initialize $S \leftarrow \lfloor \log_2 T \rfloor$, α as in (5)
initialize For $s \in [S]$, $\Psi^{(s)} \leftarrow \emptyset$, $P^{(s)} \leftarrow \emptyset$, $E^{(s)} \leftarrow \emptyset$, $\bar{\boldsymbol{\theta}}_0^{(s)} \leftarrow \mathbf{0}$

for $t = 1, \ldots, T$ **do**
  Initialize $A^{(1)} \leftarrow [K]$, $s \leftarrow 1$, $a_t \leftarrow$ `None`
  while $a_t =$ `None` **do**
    Run Algorithm 3 with $s, t-1, P^{(s)}, E^{(s)}$
    Update $\bar{\boldsymbol{\theta}}_{t-1}^{(s)}$ as in (3), $\widehat{\boldsymbol{\theta}}_{t-1}^{(s)}$ as in (4)
    For $a \in A^{(s)}$, compute $m_{t,a}^{(s)} \leftarrow \boldsymbol{x}_{t,a}^\top \widehat{\boldsymbol{\theta}}_{t-1}^{(s)}$, $w_{t,a}^{(s)} \leftarrow \alpha \|\boldsymbol{x}_{t,a}\|_{(\mathbf{V}_{E,t-1}^{(s)})^{-1}}$
    if $\exists a \in A^{(s)} \ : \ w_{t,a}^{(s)} > 2^{-s}$ **then** ▷ condition (a)
      Choose $a_t \leftarrow \arg\max_{a \in A^{(s)}} w_{t,a}^{(s)}$, receive r_t
      $\Psi^{(s)} \leftarrow \Psi^{(s)} \cup \{t\}$
      if $\|\boldsymbol{x}_{t,a_t}\|^2_{(\mathbf{V}_{P,t-1}^{(s)})^{-1}} > \tau_{t-1}^{(s)}$ **then** ▷ condition (a1)
        $P^{(s)} \leftarrow P^{(s)} \cup \{t\}$
      else ▷ condition (a2)
        $E^{(s)} \leftarrow E^{(s)} \cup \{t\}$
      end if
    else if $\forall a \in A^{(s)}$, $w_{t,a}^{(s)} \le 1/\sqrt{T}$ **then** ▷ condition (b)
      Choose, $a_t \leftarrow \arg\max_{a \in A^{(s)}} m_{t,a}^{(s)}$, receive r_t
    else ▷ condition (c)
      $A^{(s+1)} \leftarrow \{a \in A^{(s)} \ : \ m_{t,a}^{(s)} \ge \max_{a \in A^{(s)}} m_{t,a}^{(s)} - 2 \cdot 2^{-s}\}$
      $s \leftarrow s + 1$
    end if
  end while
end for

Algorithm 3 Rebalancing

input $s, t, P^{(s)}, E^{(s)}$
while `true` **do**
  $\mathbf{V}_{P,t}^{(s)} \leftarrow \kappa\lambda\mathbf{I} + \sum_{i \in P^{(s)}} \boldsymbol{x}_{i,a_i} \boldsymbol{x}_{i,a_i}^\top$
  $\mathbf{V}_{E,t}^{(s)} \leftarrow \kappa\lambda\mathbf{I} + \sum_{i \in E^{(s)}} \boldsymbol{x}_{i,a_i} \boldsymbol{x}_{i,a_i}^\top$
  Update $\beta_t^{(s)}$ as in (10), $\tau_t^{(s)}$ as in (9)
  if $\exists i \in E^{(s)} \ : \ \|\boldsymbol{x}_{i,a_i}\|^2_{(\mathbf{V}_{P,t}^{(s)})^{-1}} > \tau_t^{(s)}$ **then**
    Choose the smallest such index i
    $E^{(s)} \leftarrow E^{(s)} \setminus \{i\}$, $P^{(s)} \leftarrow P^{(s)} \cup \{i\}$
  else
    break
  end if
end while

$$\tau_t^{(s)} = \frac{1}{(\beta_t^{(s)})^2} \min\left\{1,\ 2^{-s}\left(32L_\mu\kappa\log\left(\frac{\det \mathbf{V}_{E,t}^{(s)}}{\det \kappa\lambda\mathbf{I}}\right)\right)^{-1/2}\right\} \tag{9}$$

$$\beta_t^{(s)} = \frac{\kappa}{\sqrt{2}}\left(\log\left(\frac{\det \mathbf{V}_{P,t}^{(s)}}{\det \kappa\lambda\mathbf{I}}\right) + \log(2S/\delta)\right)^{1/2} + B\sqrt{\kappa\lambda}, \tag{10}$$

In Algorithm 2, the main changes from Algorithm 1 are highlighted in blue. Since $\tau_t^{(s)}$ changes with the collected samples, an index previously assigned to $E^{(s)}$ may no longer satisfy $\|\boldsymbol{x}_{i,a_i}\|^2_{(\mathbf{V}_{P,t}^{(s)})^{-1}} \le \tau_t^{(s)}$. Before computing $\widehat{\boldsymbol{\theta}}_t^{(s)}$, Algorithm 3 repeatedly moves such indices from $E^{(s)}$ to $P^{(s)}$, updating both design matrices, $\beta_t^{(s)}$, and $\tau_t^{(s)}$ after each move, until the subsets $P^{(s)}(t)$ and $E^{(s)}(t)$ satisfy

$$\|\boldsymbol{x}_{i,a_i}\|^2_{(\mathbf{V}_{P,t}^{(s)})^{-1}} \le \tau_t^{(s)}, \qquad \forall i \in E^{(s)}(t). \tag{11}$$

This ensures that the initial-point estimator is sufficiently accurate on the remaining estimation samples, thereby controlling the (remainder) term in Lemma 9. When $\log(\det \mathbf{V}_{E,t}^{(s)}/\det \kappa\lambda\mathbf{I}) = 0$, we set $\tau_t^{(s)} = 1/(\beta_t^{(s)})^2$. We denote the final subsets by $P^{(s)}(T)$ and $E^{(s)}(T)$ and define $r_t^{(s)} = 0$ when $E^{(s)}(t) = \emptyset$.

In addition, let $\mathbf{V}_t = \kappa\lambda\mathbf{I} + \sum_{i=1}^t \boldsymbol{x}_{i,a_i}\boldsymbol{x}_{i,a_i}^\top$ be the design matrix formed by all features selected up to round t, and define $\beta_t = \frac{\kappa}{\sqrt{2}}(\log(\frac{\det \mathbf{V}_t}{\det \kappa\lambda\mathbf{I}}) + \log(2S/\delta))^{1/2} + B\sqrt{\kappa\lambda}$. Then, $\mathbf{V}_t \succeq \mathbf{V}_{P,t}^{(s)}$, $\mathbf{V}_t \succeq \mathbf{V}_{E,t}^{(s)}$, and $\beta_t \ge \beta_t^{(s)}$ for all $s \in [S]$ and $t \in [T]$.

Theorem 16. *Define a data-dependent complexity measure $\widetilde{d}$ as*

$$\widetilde{d} := \log\left(\frac{\det \mathbf{V}_T}{\det \kappa\lambda\mathbf{I}}\right) = \log\det\left(\mathbf{I} + \frac{1}{\kappa\lambda}\sum_{t=1}^T \boldsymbol{x}_{t,a_t}\boldsymbol{x}_{t,a_t}^\top\right). \tag{12}$$

Then, with probability at least $1-\delta$, the regret of Algorithm 2 is upper bounded by

$$R_T \le 16L_\mu\widetilde{d}\beta_T^2\left(1 + S\sqrt{32L_\mu\kappa\widetilde{d}}\right) + 8L_\mu\alpha\sqrt{2ST\widetilde{d}} + 2L_\mu\sqrt{T}.$$

Remark 17 (Remark on $\widetilde{d}$)**.** *This log-determinant quantity $\widetilde{d}$ is a data-dependent measure of the realized feature geometry. In particular, under $\|\boldsymbol{x}_{t,a_t}\|_2 \le 1$, we have $\widetilde{d} \le d\log(1 + T/(\kappa\lambda d))$, so $\widetilde{d}$ is always upper bounded by the ambient-dimension scale up to logarithmic factors, while it can be substantially smaller when the selected features are concentrated on a low-dimensional subspace and have small norm.*

This log-determinant quantity is also standard in kernelized and Gaussian-process bandits (Srinivas et al., 2012; Chowdhury and Gopalan, 2017; Vakili et al., 2021): an analogous information-gain quantity is given by $\frac{1}{2}\log\det(I + \lambda^{-1}K_T)$, where $K_T \in \mathbb{R}^{T\times T}$ is the kernel Gram matrix of the selected points, defined by $(K_T)_{ij} = k(\boldsymbol{x}_{i,a_i}, \boldsymbol{x}_{j,a_j})$ for $i,j \in [T]$. Such a quantity is essential in kernelized settings because the feature dimension induced by the kernel can be very large or even infinite, so regret bounds stated directly in terms of d become unavailable or uninformative.

Remark 18 (Computational complexity)**.** *Each call to Algorithm 3 takes at most $|E^{(s)}|+1$ iterations, where $E^{(s)}$ is the set at the start of the call. Each index moves from $E^{(s)}$ to $P^{(s)}$ at most once, giving at most T moves in total. Both algorithms compute the initial-point estimator in (3) at most once per round and level, hence at most TS times. Their total running time is the time for these computations plus $O(T^2d^2 + TSd^3 + TSKd^2)$.*

Remark 19 (Conditional independence in Algorithm 3)**.** *Conditional on $\mathcal{G}_{s,T}$ defined in Section D.2, the assignments to $P^{(s)}$ and $E^{(s)}$ and the moves in Algorithm 3 are fixed because they use only the selected features. Further conditioning on $\{r_i : i \in P^{(s)}(t)\}$ fixes the coefficients in the (noise) term and leaves $\{\epsilon_i : i \in E^{(s)}(t)\}$ independent with mean zero.*

B.1 Proof of Theorem 16

Let $\mathcal{E}_1'$ denote the event on which Lemma 20 holds.

Lemma 20. *With probability at least $1-\delta/2$, simultaneously for all $s \in [S], t \in [0, T-1]$, we have $\|\bar{\boldsymbol{\theta}}_t^{(s)} - \boldsymbol{\theta}^*\|_{\mathbf{V}_{P,t}^{(s)}} \le \beta_t^{(s)}$.*

Proof. Use the proof of Lemma 30 in Section E.3, replacing $d\log(1 + T/(\kappa\lambda d))$ with $\log(\det \mathbf{V}_{P,t}^{(s)}/\det \kappa\lambda\mathbf{I})$ in Lemma 35. Order the samples by when they are added to $P^{(s)}$, including moves in Algorithm 3, and apply Lemma 25 and remark 19. □

Lemma 21. *On the event $\mathcal{E}_1'$, for all $t \in [0, T-1], s \in [S]$, after rebalancing in Algorithm 3, we have $r_t^{(s)} \le \beta_t^{(s)}\sqrt{\tau_t^{(s)}} \le 1$.*

Proof. On the event $\mathcal{E}_1'$, by the Cauchy-Schwarz inequality and (11), for all $i \in E^{(s)}(t)$,

$$\left|\boldsymbol{x}_{i,a_i}^\top(\bar{\boldsymbol{\theta}}_t^{(s)} - \boldsymbol{\theta}^*)\right| \le \left\|\bar{\boldsymbol{\theta}}_t^{(s)} - \boldsymbol{\theta}^*\right\|_{\mathbf{V}_{P,t}^{(s)}} \|\boldsymbol{x}_{i,a_i}\|_{(\mathbf{V}_{P,t}^{(s)})^{-1}} \le \beta_t^{(s)}\sqrt{\tau_t^{(s)}} \le 1.$$

The last inequality follows from (9), giving the desired bound on $r_t^{(s)}$. □

Lemma 22. *For all $t \in [0, T-1], s \in [S]$, we have $|E^{(s)}(t)| \le 2\alpha^2 4^s \log(\frac{\det \mathbf{V}_{E,t}^{(s)}}{\det \kappa\lambda\mathbf{I}})$.*

Proof. Use the proof of Lemma 29 in Section E.2, taking $E = E^{(s)}(t)$ and using $\log(\det \mathbf{V}_{E,t}^{(s)}/\det \kappa\lambda\mathbf{I})$ instead of $d\log(1+T/(\kappa\lambda d))$ in Lemma 34. For the indices still in $E^{(s)}(t)$, $\mathbf{V}_{j-1} \preceq \mathbf{V}_{E,t^{(j)}-1}^{(s)}$, so condition (a) gives $\alpha\|\boldsymbol{x}_{t^{(j)},a_{t^{(j)}}}\|_{\mathbf{V}_{j-1}^{-1}} > 2^{-s}$ as in that proof. □

Lemma 23. *On the event $\mathcal{E}_1'$, for all $\boldsymbol{x} \in \mathbb{R}^d, t \in [0, T-1], s \in [S]$, after rebalancing in Algorithm 3, we have* (remainder) $\le \frac{\alpha}{2}\|\boldsymbol{x}\|_{(\mathbf{V}_{E,t}^{(s)})^{-1}}$.

Proof of Lemma 23. We fix t and s and drop the relevant notations for simplicity. From the result of Lemma 28,

$$\begin{aligned}
&\left|\boldsymbol{x}^\top\mathbf{H}^{-1}(\bar{\boldsymbol{\theta}})\left(\sum_{i\in E}\left(\dot{\mu}(\boldsymbol{x}_{i,a_i}^\top\check{\boldsymbol{\theta}}_i) - \dot{\mu}(\boldsymbol{x}_{i,a_i}^\top\bar{\boldsymbol{\theta}})\right)\boldsymbol{x}_{i,a_i}\boldsymbol{x}_{i,a_i}^\top\right)(\boldsymbol{\theta}^* - \bar{\boldsymbol{\theta}})\right| \\
&\quad\le r(e^r-1)\sqrt{|E|}\sqrt{L_\mu\kappa}\|\boldsymbol{x}\|_{\mathbf{V}_E^{-1}} \\
&\quad\le r(e^r-1)\alpha 2^s\sqrt{2\log\left(\frac{\det \mathbf{V}_{E,t}^{(s)}}{\det\kappa\lambda\mathbf{I}}\right)}\sqrt{L_\mu\kappa}\|\boldsymbol{x}\|_{\mathbf{V}_E^{-1}} && \text{(Lemma 22)} \\
&\quad\le r^2\alpha 2^{s+1}\sqrt{2\log\left(\frac{\det \mathbf{V}_{E,t}^{(s)}}{\det\kappa\lambda\mathbf{I}}\right)}\sqrt{L_\mu\kappa}\|\boldsymbol{x}\|_{\mathbf{V}_E^{-1}} && \text{(Lemma 21, } e^r-1 \le 2r, \forall r \le 1) \\
&\quad\le \alpha\tau_t^{(s)}(\beta_t^{(s)})^2 2^{s+1}\sqrt{2\log\left(\frac{\det \mathbf{V}_{E,t}^{(s)}}{\det\kappa\lambda\mathbf{I}}\right)}\sqrt{L_\mu\kappa}\|\boldsymbol{x}\|_{\mathbf{V}_E^{-1}} && \text{(Lemma 21)} \\
&\quad\le \frac{\alpha}{2}\|\boldsymbol{x}\|_{\mathbf{V}_E^{-1}},
\end{aligned}$$

where the last inequality follows from the definition of $\tau_t^{(s)}$. □

By Remark 19, the (noise) bound in Lemma 10 also holds for Algorithm 2. Using Lemma 23 in place of Lemma 12, the proof of Theorem 13 gives, with probability at least $1-\delta$,

$$\left|m_{t,a}^{(s)} - \boldsymbol{x}_{t,a}^\top\boldsymbol{\theta}^*\right| \le w_{t,a}^{(s)},$$

simultaneously whenever Algorithm 2 computes $m_{t,a}^{(s)}$ and $w_{t,a}^{(s)}$. Thus, Lemma 14 also holds. Lastly, we state the analogue of Lemma 15 for the pilot set.

Lemma 24. *For all $s \in [S]$, we have $|P^{(s)}(T)| \le 2\log(\frac{\det \mathbf{V}_{P,T}^{(s)}}{\det \kappa\lambda\mathbf{I}})\beta_T^2 \max\{1,\ 2^s\sqrt{32L_\mu\kappa\widetilde{d}}\}$.*

Proof. Use the proof of Lemma 15, ordering the indices by when they are added to $P^{(s)}$, including moves in Algorithm 3. Since $\mathbf{V}_{P,t}^{(s)}, \mathbf{V}_{E,t}^{(s)} \preceq \mathbf{V}_T$, replace $\tau^{(s)}$ by $\beta_T^{-2}\min\{1, 2^{-s}/\sqrt{32L_\mu\kappa\widetilde{d}}\}$ using (9). In Lemma 34, use $\log(\det \mathbf{V}_{P,T}^{(s)}/\det \kappa\lambda\mathbf{I})$ instead of $d\log(1+T/(\kappa\lambda d))$. □

Proof of Theorem 16. We have,

$$R_T = \underbrace{\sum_{s=1}^{S}\sum_{t\in P^{(s)}(T)}\left[\mu(\boldsymbol{x}_{t,a_t^*}^\top\boldsymbol{\theta}^*) - \mu(\boldsymbol{x}_{t,a_t}^\top\boldsymbol{\theta}^*)\right]}_{A_1} + \underbrace{\sum_{s=1}^{S}\sum_{t\in E^{(s)}(T)}\left[\mu(\boldsymbol{x}_{t,a_t^*}^\top\boldsymbol{\theta}^*) - \mu(\boldsymbol{x}_{t,a_t}^\top\boldsymbol{\theta}^*)\right]}_{A_2}$$
$$+ \underbrace{\sum_{t\in([T]\setminus(\Psi^{(1)}(T)\cup\cdots\cup\Psi^{(S)}(T)))}\left[\mu(\boldsymbol{x}_{t,a_t^*}^\top\boldsymbol{\theta}^*) - \mu(\boldsymbol{x}_{t,a_t}^\top\boldsymbol{\theta}^*)\right]}_{A_3}.$$

For the term A_1

$$\begin{aligned}
A_1 &\le \sum_{s=1}^{S}\sum_{t\in P^{(s)}(T)} 8L_\mu 2^{-s_t} \le 8L_\mu\sum_{s=1}^{S}\left|P^{(s)}(T)\right|2^{-s} && \text{(Lemma 14)}\\
&\le 16L_\mu\beta_T^2\sum_{s=1}^{S}\log\left(\frac{\det\mathbf{V}_{P,T}^{(s)}}{\det\kappa\lambda\mathbf{I}}\right)\max\left\{2^{-s},\ \sqrt{32L_\mu\kappa\widetilde{d}}\right\} && \text{(Lemma 24)}\\
&\le 16L_\mu\widetilde{d}\beta_T^2\sum_{s=1}^{S}\left(2^{-s}+\sqrt{32L_\mu\kappa\widetilde{d}}\right)\\
&\le 16L_\mu\widetilde{d}\beta_T^2\left(1+S\sqrt{32L_\mu\kappa\widetilde{d}}\right).
\end{aligned}$$

For the term A_2,

$$\begin{aligned}
A_2 &\le \sum_{s=1}^{S}\sum_{t\in E^{(s)}(T)} 8L_\mu 2^{-s_t} && \text{(Lemma 14)}\\
&\le 8L_\mu\sum_{s=1}^{S}\sum_{t\in E^{(s)}(T)} w_{t,a_t}^{(s)} = 8L_\mu\alpha\sum_{s=1}^{S}\sum_{t\in E^{(s)}(T)}\|\boldsymbol{x}_{t,a_t}\|_{(\mathbf{V}_{E,t-1}^{(s)})^{-1}}. && \text{(condition (a))}
\end{aligned}$$

Denote $E^{(s)}(T)$ by $E^{(s)}$, and enumerate its elements in chronological order as $t^{(1)},\dots,t^{(|E^{(s)}|)}$. Let $\mathbf{V}_0^{(s)} = \kappa\lambda\mathbf{I}$, $\mathbf{V}_j^{(s)} = \mathbf{V}_{j-1}^{(s)} + \boldsymbol{x}_{t^{(j)},a_{t^{(j)}}}\boldsymbol{x}_{t^{(j)},a_{t^{(j)}}}^\top$. Since $\mathbf{V}_{j-1}^{(s)} \preceq \mathbf{V}_{E,t^{(j)}-1}^{(s)}$, we have

$\|\boldsymbol{x}_{t^{(j)},a_{t^{(j)}}}\|_{(\mathbf{V}^{(s)}_{E,t^{(j)}-1})^{-1}} \le \|\boldsymbol{x}_{t^{(j)},a_{t^{(j)}}}\|_{(\mathbf{V}^{(s)}_{j-1})^{-1}}$. Then, we can proceed as

$$
\begin{aligned}
A_2 &\le 8L_\mu\alpha \sum_{s=1}^{S}\sum_{j=1}^{|E^{(s)}|} \min\left\{1, \left\|\boldsymbol{x}_{t^{(j)},a_{t^{(j)}}}\right\|_{(\mathbf{V}^{(s)}_{j-1})^{-1}}\right\} && (\lambda \ge 1/\kappa,\ \|\boldsymbol{x}_{t^{(j)},a_{t^{(j)}}}\|_{(\mathbf{V}^{(s)}_{j-1})^{-1}} \le 1)\\
&\le 8L_\mu\alpha \sqrt{S\sum_{s=1}^{S}\left(|E^{(s)}| \sum_{j=1}^{|E^{(s)}|} \min\left\{1, \left\|\boldsymbol{x}_{t^{(j)},a_{t^{(j)}}}\right\|_{(\mathbf{V}^{(s)}_{j-1})^{-1}}\right\}^2\right)} && \text{(Cauchy-Schwarz)}\\
&\le 8L_\mu\alpha \sqrt{2S\sum_{s=1}^{S}\left(|E^{(s)}| \log\left(\frac{\det \mathbf{V}^{(s)}_{E,T}}{\det \kappa\lambda\mathbf{I}}\right)\right)} && \text{(Lemma 34)}\\
&\le 8L_\mu\alpha\sqrt{2ST\widetilde{d}}.
\end{aligned}
$$

For the term A_3,

$$A_3 \le T \times 2L_\mu/\sqrt{T} = 2L_\mu\sqrt{T}. \qquad \text{(Lemma 14)}$$

Substituting term A_1, term A_2, and term A_3 back completes the proof. □

C Experimental details

This section provides implementation details for the synthetic experiment in Figure 1. All experiments were conducted on a server equipped with an AMD EPYC 9354 32-Core Processor, 251 GiB of RAM, and one NVIDIA RTX A6000 GPU.

Environment. We use a synthetic logistic bandit environment in which contexts are generated at initialization. The generator explicitly enforces $\|\boldsymbol{x}_{t,a}\|_2 \le 1$, $\|\boldsymbol{\theta}^*\|_2 \le B$, and $\dot{\mu}(\boldsymbol{x}_{t,a}^\top\boldsymbol{\theta}^*) \ge 1/\kappa$. For Figure 1, we fix $T = 2000$, $K = 5$, $\kappa = 20$, $\lambda = 1$, $B = 1$, and vary the ambient dimension over $d \in \{3, 20, 100\}$. For each value of d, we consider three geometry regimes, labeled `high`, `middle`, and `low`. These regimes are implemented by changing the effective subspace dimension r and the norm range of the incoming features:

$$
\begin{aligned}
\texttt{low}: &\quad r = 2,\ \|\boldsymbol{x}\|_2 \in [0, 0.05],\\
\texttt{middle}: &\quad r = 10,\ \|\boldsymbol{x}\|_2 \in [0.3, 0.5],\\
\texttt{high}: &\quad r = d,\ \|\boldsymbol{x}\|_2 \in [0.8, 1.0].
\end{aligned}
$$

Thus, the three regimes correspond to different effective subspace structures and lead to different values of the log-determinant quantity shown in Figure 1. All experiments are repeated over 10 independent runs, and the plots report the mean performance with $\pm 1\sigma$ error bars.

Baseline algorithms. We compare the refined version of our method in Algorithm 2 with three baselines: SupCB-GLM (Li et al., 2017), SupLogistic (Jun et al., 2021), and DDRTS-GLM (Kim et al., 2023). During implementation, Algorithm 2 maintains pilot and estimation buckets at each level, computes an initial-point estimator on the pilot data, and then applies a one-step Newton correction on the estimation data. The thresholds $\tau_t^{(s)}$ are updated adaptively using the current log-determinant statistics.

SupCB-GLM and SupLogistic are both implemented with independent level-wise buckets. In both cases, each level has its own warm-up size $\tau = \sqrt{dT}$, and samples are not reused across levels. For SupLogistic, the analysis in Appendix G of Jun et al. (2021) requires a sufficiently large horizon depending on σ_0 for the choice $\tau = \sqrt{dT}$. If the context covariance is singular, this analysis gives no finite such horizon. Thus, our implementation gives the baselines a finite exploration period even on instances outside their regret guarantees. DDRTS-GLM is implemented as a practical Thompson-sampling-style baseline that repeatedly fits a regularized logistic estimator, forms a Hessian-based covariance matrix, and samples actions using Monte Carlo argmax probabilities.

All methods use the same projected gradient descent routine for fitting regularized logistic objectives, and we use 20 gradient steps with learning rate 0.3. For DDRTS-GLM, we use one Monte Carlo sample to approximate argmax probabilities.

D Deferred proofs for Section 6

D.1 Proof of Lemma 9

Proof. Consider the term $\widehat{\boldsymbol{\theta}}_t^{(s)} - \boldsymbol{\theta}^*$. In this section, we fix t and s and drop the notations for simplicity. Then, we have

$$\begin{aligned}\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* &= \bar{\boldsymbol{\theta}} + \mathbf{H}^{-1}(\bar{\boldsymbol{\theta}})\boldsymbol{g}(\bar{\boldsymbol{\theta}}) - \boldsymbol{\theta}^* \\ &= \bar{\boldsymbol{\theta}} + \mathbf{H}^{-1}(\bar{\boldsymbol{\theta}})\left(\sum_{i\in E}\big(r_i - \mu(\boldsymbol{x}_{i,a_i}^\top\bar{\boldsymbol{\theta}})\big)\boldsymbol{x}_{i,a_i} - \lambda\bar{\boldsymbol{\theta}}\right) - \boldsymbol{\theta}^*\end{aligned}$$

By the definition,

$$\begin{aligned}r_i - \mu(\boldsymbol{x}_{i,a_i}^\top\bar{\boldsymbol{\theta}}) &= \epsilon_i + \mu(\boldsymbol{x}_{i,a_i}^\top\boldsymbol{\theta}^*) - \mu(\boldsymbol{x}_{i,a_i}^\top\bar{\boldsymbol{\theta}}) \\ &= \epsilon_i + \dot{\mu}(\boldsymbol{x}_{i,a_i}^\top\check{\boldsymbol{\theta}}_i)\boldsymbol{x}_{i,a_i}^\top(\boldsymbol{\theta}^* - \bar{\boldsymbol{\theta}}),\end{aligned}$$

where the last equality holds by the mean-value theorem with some $c_i \in [0,1]$ and $\check{\boldsymbol{\theta}}_i = c_i\boldsymbol{\theta}^* + (1-c_i)\bar{\boldsymbol{\theta}}$. Substituting this result back gives

$$\begin{aligned}\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* &= \mathbf{H}^{-1}(\bar{\boldsymbol{\theta}})\sum_{i\in E}\epsilon_i\boldsymbol{x}_{i,a_i} + \mathbf{H}^{-1}(\bar{\boldsymbol{\theta}})\left(\sum_{i\in E}\dot{\mu}(\boldsymbol{x}_{i,a_i}^\top\check{\boldsymbol{\theta}}_i)\boldsymbol{x}_{i,a_i}\boldsymbol{x}_{i,a_i}^\top(\boldsymbol{\theta}^* - \bar{\boldsymbol{\theta}})\right) - \mathbf{H}^{-1}(\bar{\boldsymbol{\theta}})\lambda\bar{\boldsymbol{\theta}} - (\boldsymbol{\theta}^* - \bar{\boldsymbol{\theta}}) \\ &= \mathbf{H}^{-1}(\bar{\boldsymbol{\theta}})\sum_{i\in E}\epsilon_i\boldsymbol{x}_{i,a_i} + \mathbf{H}^{-1}(\bar{\boldsymbol{\theta}})\left(\sum_{i\in E}\Big(\dot{\mu}(\boldsymbol{x}_{i,a_i}^\top\check{\boldsymbol{\theta}}_i) - \dot{\mu}(\boldsymbol{x}_{i,a_i}^\top\bar{\boldsymbol{\theta}})\Big)\boldsymbol{x}_{i,a_i}\boldsymbol{x}_{i,a_i}^\top(\boldsymbol{\theta}^* - \bar{\boldsymbol{\theta}})\right) \\ &\quad \underbrace{+ \mathbf{H}^{-1}(\bar{\boldsymbol{\theta}})\left(\sum_{i\in E}\dot{\mu}(\boldsymbol{x}_{i,a_i}^\top\bar{\boldsymbol{\theta}})\boldsymbol{x}_{i,a_i}\boldsymbol{x}_{i,a_i}^\top(\boldsymbol{\theta}^* - \bar{\boldsymbol{\theta}})\right) - \mathbf{H}^{-1}(\bar{\boldsymbol{\theta}})\lambda\bar{\boldsymbol{\theta}} - (\boldsymbol{\theta}^* - \bar{\boldsymbol{\theta}})}_{A_1}\end{aligned}$$

For the term A_1,

$$\begin{aligned}A_1 &= \mathbf{H}^{-1}(\bar{\boldsymbol{\theta}})(\mathbf{H}(\bar{\boldsymbol{\theta}}) - \lambda\mathbf{I})(\boldsymbol{\theta}^* - \bar{\boldsymbol{\theta}}) - \mathbf{H}^{-1}(\bar{\boldsymbol{\theta}})\lambda\bar{\boldsymbol{\theta}} - (\boldsymbol{\theta}^* - \bar{\boldsymbol{\theta}}) \\ &= -\mathbf{H}^{-1}(\bar{\boldsymbol{\theta}})\lambda(\boldsymbol{\theta}^* - \bar{\boldsymbol{\theta}}) - \mathbf{H}^{-1}(\bar{\boldsymbol{\theta}})\lambda\bar{\boldsymbol{\theta}} \\ &= -\mathbf{H}^{-1}(\bar{\boldsymbol{\theta}})\lambda\boldsymbol{\theta}^*.\end{aligned}$$

Now, plugging this result back, and compute $|\boldsymbol{x}^\top(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)|$ and split the term with triangle inequality completes the proof. □

D.2 Proof of Lemma 10

In this section, we fix t and s and drop the notations for simplicity. Denote

$$p_i = \boldsymbol{x}^\top\mathbf{H}^{-1}(\bar{\boldsymbol{\theta}})\boldsymbol{x}_{i,a_i}, \quad \forall i \in E.$$

Then, we have

$$\left|\boldsymbol{x}^\top\left(\mathbf{H}^{-1}(\bar{\boldsymbol{\theta}})\sum_{i\in E}\epsilon_i\boldsymbol{x}_{i,a_i}\right)\right| = \left|\sum_{i\in E}p_i\epsilon_i\right|. \tag{13}$$

We want to apply the Bernstein inequality to (13). To apply this, we need to show that being conditioned on a proper filtration, (i) $\{\epsilon_i\}_{i\in E}$ are independent, and (ii) $\{p_i\}_{i\in E}$ are fixed.

First, define a filtration

$$\mathcal{G}_{s,t} \coloneqq \sigma\big(\{\boldsymbol{x}_{i,a}\}_{i\in[t],a\in[K]},\ \{r_j\}_{j\in\Psi^{(1)}(t)\cup\Psi^{(2)}(t)\cup\cdots\cup\Psi^{(s-1)}(t)}\big).$$

Lemma 25 (Level-s fixed design, (Auer, 2002))**.** *For all $s \in [S], t \in [T]$, $\Psi^{(s)}(t)$ and $\{x_{i,a_i}\}_{i\in\Psi^{(s)}(t)}$ are $\mathcal{G}_{s,t}$-measurable. Therefore, conditional on $\mathcal{G}_{s,t} \vee \sigma(\{\boldsymbol{x}_{i,a_i}\}_{i\in\Psi^{(s)}(t)})$, $\{r_i\}_{i\in\Psi^{(s)}(t)}$ are independent.*

Proof. Consider any $i \in [t]$. The event $\{i \in \Psi^{(s)}(t)\}$ implies that in round i, the algorithm (i) reaches level s and then (ii) falls in to condition (a). These two requirements, (i) and (ii), only depend on arm sets $\leq t$ i.e., $\{\boldsymbol{x}_{i,a}\}_{i\in[t],a\in[K]}$, and the rewards of levels $< s$ i.e., $r_j \in \Psi^{(1)}(t) \cup \Psi^{(2)}(t) \cup \cdots \cup \Psi^{(s-1)}(t)$. This means that $\Psi^{(s)}(t)$ and $\{x_{i,a_i}\}_{i\in\Psi^{(s)}(t)}$ are $\mathcal{G}_{s,t}$-measurable. Therefore, conditional on $\mathcal{G}_{s,t} \vee \sigma(\{\boldsymbol{x}_{i,a_i}\}_{i\in\Psi^{(s)}(t)})$, $\{r_i\}_{i\in\Psi^{(s)}(t)}$ are independent Bernoulli random variables. □

Proposition 26. *For all $s \in [S], t \in [T]$, $P^{(s)}(t)$ and $E^{(s)}(t)$ are also $\mathcal{G}_{s,t}$-measurable.*

Proof. Consider any $i \in [t]$ and the event $\{i \in P^{(s)}(t)\}$. The event implies that $\{i \in \Psi^{(s)}(t)\}$, which is $\mathcal{G}_{s,t}$-measurable by Lemma 25, and that the condition (a1) holds, which is $\|\boldsymbol{x}_{i,a_i}\|^2_{(\mathbf{V}^{(s)}_{P,i-1})^{-1}} > \tau^{(s)}$. Since whether such condition holds only depends on $\{\boldsymbol{x}_{i,a_i}\}_{i\in\Psi^{(s)}(t)}$, it is also $\mathcal{G}_{s,t}$-measurable. Same applies to the event $\{i \in E^{(s)}(t)\}$. □

Now, we establish (i) and (ii) conditioning on a new filtration

$$\mathcal{H}_{s,t} \coloneqq \mathcal{G}_{s,t} \vee \sigma(\{\boldsymbol{x}_{i,a_i}\}_{i\in E}) \vee \sigma(\{x_{i,a_i}, r_i\}_{i\in P}).$$

Lemma 27. *For all $s \in [S], t \in [T]$, conditional on $\mathcal{H}_{s,t}$, $\{\epsilon_i\}_{i\in E^{(s)}(t)}$ are independent, and $\{p_i\}_{i\in E^{(s)}(t)}$ are deterministic.*

Proof. For (i), by Lemma 25 and proposition 26, $\{r_i\}_{i\in P\cup E}$ are independent, conditioned on $\mathcal{G}_{s,t} \vee \sigma(\{\boldsymbol{x}_{i,a_i}\}_{i\in E}) \vee \sigma(\{\boldsymbol{x}_{i,a_i}\}_{i\in P})$. Therefore, further conditioning on $\sigma(\{r_i\}_{i\in P})$, which is same as $\mathcal{H}_{s,t}$, does not affect the independence on $\{r_i\}_{i\in E}$. Also the definition of $\epsilon_i = r_i - \mu(\boldsymbol{x}_{i,a_i}^\top \boldsymbol{\theta}^*)$ implies that $\{\epsilon_i\}_{i\in E}$ are independent. Next, for (ii), p_i consists of $\{\boldsymbol{x}_{i,a_i}\}_{i\in E}$, which is $\mathcal{H}_{s,t}$-measurable, and $\mathbf{H}(\bar{\boldsymbol{\theta}})$. From the definition of $\mathbf{H}(\bar{\boldsymbol{\theta}})$, it consists of $\{\boldsymbol{x}_{i,a_i}\}_{i\in E}$ and $\bar{\boldsymbol{\theta}}$ where $\bar{\boldsymbol{\theta}}$ is determined by $\{\boldsymbol{x}_{i,a_i}, r_i\}_{i\in P}$. Therefore, p_i is $\mathcal{H}_{s,t}$-measurable. □

Now we are ready to start the proof.

Proof of Lemma 10. We prepare to apply the Bernstein inequality to (13). We start by upper bounding $\max_{i\in E} |p_i\epsilon_i|$ and $\sum_{i\in E} \mathrm{Var}(p_i\epsilon_i \mid \mathcal{H}_{s,t})$: For $\max_{i\in E} |p_i\epsilon_i|$, we have

$$\begin{aligned}
\max_{i\in E} |p_i\epsilon_i| &\leq \max_{i\in E} |\boldsymbol{x}^\top \mathbf{H}^{-1}(\bar{\boldsymbol{\theta}})\boldsymbol{x}_{i,a_i}| \\
&\leq \max_{i\in E} \|\boldsymbol{x}^\top \mathbf{H}^{-1/2}(\bar{\boldsymbol{\theta}})\|_2 \cdot \|\mathbf{H}^{-1/2}(\bar{\boldsymbol{\theta}})\|_2 \cdot \|\boldsymbol{x}_{i,a_i}\|_2 && \text{(Cauchy-Schwarz)} \\
&\leq \frac{1}{\sqrt{\lambda}} \|\boldsymbol{x}\|_{\mathbf{H}^{-1}(\bar{\boldsymbol{\theta}})} && (\|\boldsymbol{x}_{i,a_i}\|_2 \leq 1, \lambda_{\min}(\mathbf{H}(\bar{\boldsymbol{\theta}})) \geq \lambda) \\
&\leq \sqrt{\frac{\kappa}{\lambda}} \|\boldsymbol{x}\|_{\mathbf{V}_E^{-1}}, && (14)
\end{aligned}$$

where the last inequality holds since $\bar{\boldsymbol{\theta}} \in \Theta$, and $\mathbf{V}_E \preceq \kappa\mathbf{H}(\bar{\boldsymbol{\theta}})$.

For $\sum_{i\in E} \mathrm{Var}(p_i\epsilon_i \mid \mathcal{H}_{s,t})$, we have

$$\begin{aligned}
\sum_{i\in E} \mathrm{Var}(p_i\epsilon_i \mid \mathcal{H}_{s,t}) &= \sum_{i\in E} \mathbb{E}[(p_i\epsilon_i)^2 \mid \mathcal{H}_{s,t}] && (\mathbb{E}[\epsilon_i \mid \mathcal{H}_{s,t}] = 0) \\
&= \sum_{i\in E} \boldsymbol{x}^\top \mathbf{H}^{-1}(\bar{\boldsymbol{\theta}})\boldsymbol{x}_{i,a_i}\boldsymbol{x}_{i,a_i}^\top \mathbf{H}^{-1}(\bar{\boldsymbol{\theta}})\boldsymbol{x} \cdot \mathbb{E}[\epsilon_i^2 \mid \mathcal{H}_{s,t}] \\
&\leq L_\mu \sum_{i\in E} \boldsymbol{x}^\top \mathbf{H}^{-1}(\bar{\boldsymbol{\theta}})\boldsymbol{x}_{i,a_i}\boldsymbol{x}_{i,a_i}^\top \mathbf{H}^{-1}(\bar{\boldsymbol{\theta}})\boldsymbol{x} && \text{(Assumption 2)} \\
&= L_\mu \boldsymbol{x}^\top \mathbf{H}^{-1}(\bar{\boldsymbol{\theta}}) \left(\sum_{i\in E} \boldsymbol{x}_{i,a_i}\boldsymbol{x}_{i,a_i}^\top\right) \mathbf{H}^{-1}(\bar{\boldsymbol{\theta}})\boldsymbol{x} \\
&\leq L_\mu \boldsymbol{x}^\top \mathbf{H}^{-1}(\bar{\boldsymbol{\theta}})\kappa\mathbf{H}(\bar{\boldsymbol{\theta}})\mathbf{H}^{-1}(\bar{\boldsymbol{\theta}})\boldsymbol{x} && (\mathbf{V}_E \preceq \kappa\mathbf{H}(\bar{\boldsymbol{\theta}})) \\
&\leq L_\mu \kappa^2 \|\boldsymbol{x}\|^2_{\mathbf{V}_E^{-1}}. && (15)
\end{aligned}$$

Applying Bernstein inequality (Lemma 33) with

$$b \leftarrow \sqrt{\frac{\kappa}{\lambda}}\|\boldsymbol{x}\|_{\mathbf{V}_E^{-1}}, \quad V \leftarrow L_\mu \kappa^2 \|\boldsymbol{x}\|^2_{\mathbf{V}_E^{-1}}, \quad L \leftarrow \log(2/\delta), \qquad \text{((14) and (15))}$$

with probability at least $1-\delta$,

$$\left|\sum_{i\in E} p_i \epsilon_i\right| \leq \sqrt{2L_\mu \kappa^2 \log(2/\delta)}\|\boldsymbol{x}\|_{\mathbf{V}_E^{-1}} + \frac{\sqrt{\kappa}\log(2/\delta)}{3\sqrt{\lambda}}\|\boldsymbol{x}\|_{\mathbf{V}_E^{-1}},$$

which completes the proof. □

D.3 Proof of Lemma 12

Before starting the proof, we introduce two technical lemmas:

Lemma 28. *For all $\boldsymbol{x} \in \mathbb{R}^d$, $t \in [0, T-1]$, $s \in [S]$,*

$$\left|\boldsymbol{x}^\top \left(\mathbf{H}_t^{(s)}(\bar{\boldsymbol{\theta}}_t^{(s)})\right)^{-1} \left(\sum_{i\in E^{(s)}(t)} \left(\dot{\mu}(\boldsymbol{x}_{i,a_i}^\top \check{\boldsymbol{\theta}}_{t,i}^{(s)}) - \dot{\mu}(\boldsymbol{x}_{i,a_i}^\top \bar{\boldsymbol{\theta}}_t^{(s)})\right)\boldsymbol{x}_{i,a_i}\boldsymbol{x}_{i,a_i}^\top\right)(\boldsymbol{\theta}^* - \bar{\boldsymbol{\theta}}_t^{(s)})\right|$$
$$\leq r_t^{(s)}(e^{r_t^{(s)}} - 1)\sqrt{|E^{(s)}(t)|}\sqrt{L_\mu\kappa}\|\boldsymbol{x}\|_{(\mathbf{V}_{E,t}^{(s)})^{-1}}.$$

Proof. The proof is deferred to Section E.1. □

Lemma 29. *For all $t \in [0, T-1]$, $s \in [S]$, we have $|E^{(s)}(t)| \leq 2\alpha^2 d 4^s \log(1 + T/(\kappa\lambda d))$.*

Proof. The proof is deferred to Section E.2. □

Now we are ready to start the proof. In this section, we fix t and s and drop the notations for simplicity.

Proof of Lemma 12. We can start from the result of Lemma 28, which is

$$\begin{aligned}
&\left|\boldsymbol{x}^\top \mathbf{H}^{-1}(\bar{\boldsymbol{\theta}})\left(\sum_{i\in E}\left(\dot{\mu}(\boldsymbol{x}_{i,a_i}^\top \check{\boldsymbol{\theta}}_i) - \dot{\mu}(\boldsymbol{x}_{i,a_i}^\top \bar{\boldsymbol{\theta}})\right)\boldsymbol{x}_{i,a_i}\boldsymbol{x}_{i,a_i}^\top\right)(\boldsymbol{\theta}^* - \bar{\boldsymbol{\theta}})\right| \\
&\leq r(e^r - 1)\sqrt{|E|}\sqrt{L_\mu\kappa}\|\boldsymbol{x}\|_{\mathbf{V}_E^{-1}} \\
&\leq r(e^r - 1)\alpha 2^s \sqrt{2d\log(1 + T/(\kappa\lambda d))}\sqrt{L_\mu\kappa}\|\boldsymbol{x}\|_{\mathbf{V}_E^{-1}} && \text{(Lemma 29)} \\
&\leq r^2\alpha 2^{s+1}\sqrt{2d\log(1 + T/(\kappa\lambda d))}\sqrt{L_\mu\kappa}\|\boldsymbol{x}\|_{\mathbf{V}_E^{-1}} && (e^r - 1 \leq 2r, \forall r \leq 1) \\
&\leq \alpha\tau^{(s)}\beta^2 2^{s+1}\sqrt{2d\log(1 + T/(\kappa\lambda d))}\sqrt{L_\mu\kappa}\|\boldsymbol{x}\|_{\mathbf{V}_E^{-1}} && \text{(Lemma 11)} \\
&\leq \frac{\alpha}{2}\|\boldsymbol{x}\|_{\mathbf{V}_E^{-1}},
\end{aligned}$$

where the last inequality follows from the definition of $\tau^{(s)}$. □

D.4 Proof of Theorem 13

Before we start the proof, we introduce one technical lemma:

Lemma 30. *With probability at least $1 - \delta/2$, simultaneously for all $s \in [S], t \in [0, T-1]$,*

$$\left\|\bar{\boldsymbol{\theta}}_t^{(s)} - \boldsymbol{\theta}^*\right\|_{\mathbf{V}_{P,t}^{(s)}} \leq \beta, \quad \beta := \frac{\kappa}{\sqrt{2}}\sqrt{d\log(1 + T/(\kappa\lambda d)) + \log(2S/\delta)} + B\sqrt{\kappa\lambda}.$$

Proof. The proof is deferred to Section E.3. □

Remark 31. *In Lemma 30, one could use a fixed-design tail inequality and thereby remove the* $\log(T)$ *factor in* β*. However, our main interest is in reducing the dimension factor* d*, and more importantly in extending this dependence to an effective dimension in future work. Since the results obtained under a fixed-design analysis are tied to the ambient dimension, such an extension would not be possible. For this reason, we instead use a self-normalized inequality for martingale difference sequences, as in Lemma 30.*

Recall the definition of $\mathcal{E}_1$ when Lemma 30 holds. Then, we can see that Lemma 12 holds with probability at least $1 - \delta/2$. Now we are ready to start the proof.

Proof of Theorem 13. We are going to show that

$$\left|\boldsymbol{x}_{t+1,a}^\top(\widehat{\boldsymbol{\theta}}_t^{(s)} - \boldsymbol{\theta}^*)\right| \le w_{t+1,a}^{(s)}$$

holds simultaneously for all $t \in [0, T-1], s \in [S], a \in [K]$ with probability at least $1 - \delta$.

Notice that by Lemma 9, we have

$$\left|\boldsymbol{x}_{t+1,a}^\top(\widehat{\boldsymbol{\theta}}_t^{(s)} - \boldsymbol{\theta}^*)\right| \le (\text{noise}) + (\text{bias}) + (\text{remainder}). \tag{16}$$

First, for the (noise) term, from Lemma 10, we replace $\delta \leftarrow \delta/(2TSK)$, and taking union bound gives that, with probability at least $1 - \delta/2$, simultaneously for all $t \in [0, T-1], s \in [S], a \in [K]$,

$$(\text{noise}) \le \alpha_1 \|\boldsymbol{x}_{t+1,a}\|_{(\mathbf{V}_{E,t}^{(s)})^{-1}}, \quad \alpha_1 := \left(\sqrt{2L_\mu\kappa^2\log(4TSK/\delta)} + \frac{\sqrt{\kappa}\log(4TSK/\delta)}{3\sqrt{\lambda}}\right),$$

We denote the event as $\mathcal{E}_2$ when the above inequality holds. For the (bias) term, for any $t \in [0, T-1], s \in [S], a \in [K]$,

$$\begin{aligned}
(\text{bias}) &\le \|\boldsymbol{x}_{t+1,a}\|_{(\mathbf{H}_t^{(s)}(\bar{\boldsymbol{\theta}}_t^{(s)}))^{-1}} \cdot \lambda\|\boldsymbol{\theta}^*\|_{(\mathbf{H}_t^{(s)}(\bar{\boldsymbol{\theta}}_t^{(s)}))^{-1}} && \text{(Cauchy-Schwarz)} \\
&\le \sqrt{\kappa}\|\boldsymbol{x}_{t+1,a}\|_{(\mathbf{V}_{E,t}^{(s)})^{-1}} \cdot \lambda\|\boldsymbol{\theta}^*\|_2 \cdot \lambda_{\min}^{-1/2}(\mathbf{H}_t^{(s)}(\bar{\boldsymbol{\theta}}_t^{(s)})) && (\mathbf{V}_{E,t}^{(s)} \preceq \kappa\mathbf{H}_t^{(s)}(\bar{\boldsymbol{\theta}}_t^{(s)})) \\
&\le B\sqrt{\kappa\lambda}\|\boldsymbol{x}_{t+1,a}\|_{(\mathbf{V}_{E,t}^{(s)})^{-1}} \\
&= \alpha_2\|\boldsymbol{x}_{t+1,a}\|_{(\mathbf{V}_{E,t}^{(s)})^{-1}},
\end{aligned}$$

where the last equality follows from the definition of α_2. For the (remainder) term, from the result of Lemma 12, with probability at least $1 - \delta/2$ (or on the event $\mathcal{E}_1$), simultaneously for all $t \in [0, T-1], s \in [S], a \in [K]$,

$$(\text{remainder}) \le \frac{\alpha}{2}\|\boldsymbol{x}_{t+1,a}\|_{(\mathbf{V}_{E,t}^{(s)})^{-1}}.$$

Finally, substituting the (noise), (bias), and (remainder) term back to (16), and taking union bound for $\mathcal{E}_1, \mathcal{E}_2$, and shifting the index of t by 1 ($t \leftarrow t-1$), we have that, with probability at least $1 - \delta$, simultaneously for all $t \in [T], s \in [S], a \in [K]$,

$$\begin{aligned}
\left|\boldsymbol{x}_{t,a}^\top(\widehat{\boldsymbol{\theta}}_{t-1}^{(s)} - \boldsymbol{\theta}^*)\right| &\le (\alpha_1 + \alpha_2 + \alpha/2)\|\boldsymbol{x}_{t,a}\|_{(\mathbf{V}_{E,t-1}^{(s)})^{-1}} \\
&= \alpha\|\boldsymbol{x}_{t,a}\|_{(\mathbf{V}_{E,t-1}^{(s)})^{-1}} \\
&= w_{t,a}^{(s)},
\end{aligned}$$

completes the proof. □

D.5 Proof of Lemma 14

We first introduce the following lemma, showing that the optimal arm is never eliminated until we choose the action at certain level.

Lemma 32. *On the good event* $\mathcal{E}$*, if the action* a_t *is chosen in level* s_t*, then the optimal arm is never eliminated up to* s_t*, i.e.* $a_t^* \in A^{(s)}$ *for all* $s \in [s_t]$*.*

Proof. The proof is deferred to Section E.4 □

Now we start the proof.

Proof. First, consider the case when a_t is selected by condition (a). For the base case $s_t = 1$, we can see that $\mu(\boldsymbol{x}_{t,a_t^*}^\top \boldsymbol{\theta}^*) - \mu(\boldsymbol{x}_{t,a_t}^\top \boldsymbol{\theta}^*) \le 8L_\mu 2^{-1} = 1$. For the case $s_t > 1$, by condition (a) in level $s_t - 1$, we have

$$\left|\boldsymbol{x}_{t,a}^\top \boldsymbol{\theta}^* - m_{t,a}^{(s_t-1)}\right| \le w_{t,a}^{(s_t-1)} \le 2^{-(s_t-1)}. \tag{17}$$

Also, by condition (c) in level $s_t - 1$,

$$m_{t,a}^{(s_t-1)} \ge \max_{a \in A^{(s_t-1)}} m_{t,a}^{(s_t-1)} - 2 \cdot 2^{-(s_t-1)}. \tag{18}$$

Now, for the per-round regret

$$\begin{aligned}
\mu(\boldsymbol{x}_{t,a_t^*}^\top \boldsymbol{\theta}^*) - \mu(\boldsymbol{x}_{t,a_t}^\top \boldsymbol{\theta}^*) &\le L_\mu \left(\boldsymbol{x}_{t,a_t^*}^\top \boldsymbol{\theta}^* - \boldsymbol{x}_{t,a_t}^\top \boldsymbol{\theta}^*\right) && \text{(Assumption 2, } \boldsymbol{x}_{t,a_t^*}^\top \boldsymbol{\theta}^* \ge \boldsymbol{x}_{t,a_t}^\top \boldsymbol{\theta}^*) \\
&= L_\mu \left[\underbrace{\left(\boldsymbol{x}_{t,a_t^*}^\top \boldsymbol{\theta}^* - m_{t,a_t^*}^{(s_t-1)}\right)}_{A_1} + \underbrace{\left(m_{t,a_t^*}^{(s_t-1)} - m_{t,a_t}^{(s_t-1)}\right)}_{A_2} + \underbrace{\left(m_{t,a_t}^{(s_t-1)} - \boldsymbol{x}_{t,a_t}^\top \boldsymbol{\theta}^*\right)}_{A_3} \right]
\end{aligned}$$

For term A_1 and term A_3, by (17)

$$A_1, A_3 \le 2^{-(s_t-1)} = 2 \cdot 2^{-s_t}.$$

For term A_2, by (18)

$$A_2 \le 2 \cdot 2^{-(s_t-1)} = 4 \cdot 2^{-s_t}$$

Substituting term A_1, term A_2, and term A_3 back gives the desired result.

Next, consider the case when a_t is selected by condition (b), which means in level s_t,

$$\left|\boldsymbol{x}_{t,a}^\top \boldsymbol{\theta}^* - m_{t,a}^{(s_t)}\right| \le w_{t,a}^{(s_t)} \le 1/\sqrt{T}. \tag{19}$$

Then, for the per-round regret,

$$\begin{aligned}
\mu(\boldsymbol{x}_{t,a_t^*}^\top \boldsymbol{\theta}^*) - \mu(\boldsymbol{x}_{t,a_t}^\top \boldsymbol{\theta}^*) &\le L_\mu \left(\boldsymbol{x}_{t,a_t^*}^\top \boldsymbol{\theta}^* - \boldsymbol{x}_{t,a_t}^\top \boldsymbol{\theta}^*\right) \\
&= L_\mu \left[\underbrace{\left(\boldsymbol{x}_{t,a_t^*}^\top \boldsymbol{\theta}^* - m_{t,a_t^*}^{(s_t)}\right)}_{A_4} + \underbrace{\left(m_{t,a_t^*}^{(s_t)} - m_{t,a_t}^{(s_t)}\right)}_{A_5} + \underbrace{\left(m_{t,a_t}^{(s_t)} - \boldsymbol{x}_{t,a_t}^\top \boldsymbol{\theta}^*\right)}_{A_6} \right].
\end{aligned}$$

For term A_4 and term A_6, by (19),

$$A_4, A_6 \le 1/\sqrt{T}.$$

For term A_5, $A_5 \le 0$ by our choice a_t. Substituting term A_4, term A_5, and term A_6 back gives the desired result, completing the proof. □

D.6 Proof of Lemma 15

Proof. Fix $s \in [S]$. Denote $P^{(s)}(T)$ as P for simplicity. Redefine $P = \{t^{(1)}, \ldots, t^{(|P|)}\}$ chronologically, $\mathbf{V}_0 = \kappa\lambda\mathbf{I}$, $\mathbf{V}_j = \mathbf{V}_{j-1} + x_{t^{(j)},a_{t^{(j)}}} x_{t^{(j)},a_{t^{(j)}}}^\top$. Then for $t^{(j)} \in P$, by condition (a1),

$$\left\|\boldsymbol{x}_{t^{(j)},a_{t^{(j)}}}\right\|_{\mathbf{V}_{j-1}^{-1}}^2 \ge \tau^{(s)}. \tag{20}$$

Then,

$$\begin{aligned}
|P| \cdot \tau^{(s)} &\le \sum_{j=1}^{|P|} \min\left\{1, \left\|\boldsymbol{x}_{t^{(j)},a_{t^{(j)}}}\right\|_{\mathbf{V}_{j-1}^{-1}}^2\right\} && (\lambda \ge 1/\kappa, \|\boldsymbol{x}_{t^{(j)},a_{t^{(j)}}}\|_{\mathbf{V}_{j-1}^{-1}} \le 1) \\
&\le 2d\log(1 + |P|/(\kappa\lambda d)) && \text{(Lemma 34)} \\
&\le 2d\log(1 + T/(\kappa\lambda d)),
\end{aligned}$$

rearranging $\tau^{(s)}$ to the right-hand side completes the proof. □

E Deferred proof for Section D

E.1 Proof of Lemma 28

In this section, we fix t and s and drop the notation for it for simplicity. Also, redefine $r \coloneqq \max_{i\in E} |\boldsymbol{x}_{i,a_i}^\top(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)|$.

Proof. In order to control the term $|\dot{\mu}(\boldsymbol{x}_{i,a_i}^\top \check{\boldsymbol{\theta}}_i) - \dot{\mu}(\boldsymbol{x}_{i,a_i}^\top \bar{\boldsymbol{\theta}})|$, we start by using the fact that $(\log \dot{\mu}(v))' = \frac{\ddot{\mu}(v)}{\dot{\mu}(v)} = 1 - 2\mu(v)$, we have $|(\log \dot{\mu}(v))'| \le 1$ and it means that $\log \dot{\mu}(v)$ is a 1-Lipschitz function. Therefore, we have

$$\left|\log \dot{\mu}(\boldsymbol{x}_{i,a_i}^\top \check{\boldsymbol{\theta}}_i) - \log \dot{\mu}(\boldsymbol{x}_{i,a_i}^\top \bar{\boldsymbol{\theta}})\right| \le \left|\boldsymbol{x}_{i,a_i}^\top(\check{\boldsymbol{\theta}}_i - \bar{\boldsymbol{\theta}})\right|,$$

where $\check{\boldsymbol{\theta}}_i = c_i\bar{\boldsymbol{\theta}} + (1-c_i)\boldsymbol{\theta}^*$ for some constant $c_i \in [0,1]$. For the right-hand side

$$\left|\boldsymbol{x}_{i,a_i}^\top(\check{\boldsymbol{\theta}}_i - \bar{\boldsymbol{\theta}})\right| = (1-c_i)\left|\boldsymbol{x}_{i,a_i}^\top(\boldsymbol{\theta}^* - \bar{\boldsymbol{\theta}})\right| \le (1-c_i)r \le r.$$

Substituting this back gives

$$\begin{aligned}
&\left|\log \dot{\mu}(\boldsymbol{x}_{i,a_i}^\top \check{\boldsymbol{\theta}}_i) - \log \dot{\mu}(\boldsymbol{x}_{i,a_i}^\top \bar{\boldsymbol{\theta}})\right| \le r \\
&\quad\implies (e^{-r}-1)\dot{\mu}(\boldsymbol{x}_{i,a_i}^\top \bar{\boldsymbol{\theta}}) \le \dot{\mu}(\boldsymbol{x}_{i,a_i}^\top \check{\boldsymbol{\theta}}_i) - \dot{\mu}(\boldsymbol{x}_{i,a_i}^\top \bar{\boldsymbol{\theta}}) \le (e^r-1)\dot{\mu}(\boldsymbol{x}_{i,a_i}^\top \bar{\boldsymbol{\theta}}).
\end{aligned}$$

Also, since $|e^{-r}-1| \le |e^r - 1|$ for all $r \ge 0$, we have

$$\left|\dot{\mu}(\boldsymbol{x}_{i,a_i}^\top \check{\boldsymbol{\theta}}_i) - \dot{\mu}(\boldsymbol{x}_{i,a_i}^\top \bar{\boldsymbol{\theta}})\right| \le (e^r-1)\dot{\mu}(\boldsymbol{x}_{i,a_i}^\top \bar{\boldsymbol{\theta}}). \tag{21}$$

Now, we begin the proof:

$$\begin{aligned}
&\left|\boldsymbol{x}^\top \mathbf{H}^{-1}(\bar{\boldsymbol{\theta}})\left(\sum_{i\in E}\Big(\dot{\mu}(\boldsymbol{x}_{i,a_i}^\top \check{\boldsymbol{\theta}}_i) - \dot{\mu}(\boldsymbol{x}_{i,a_i}^\top \bar{\boldsymbol{\theta}})\Big)\boldsymbol{x}_{i,a_i}\boldsymbol{x}_{i,a_i}^\top\right)(\boldsymbol{\theta}^* - \bar{\boldsymbol{\theta}})\right| \\
&= \left|\sum_{i\in E}\Big(\dot{\mu}(\boldsymbol{x}_{i,a_i}^\top \check{\boldsymbol{\theta}}_i) - \dot{\mu}(\boldsymbol{x}_{i,a_i}^\top \bar{\boldsymbol{\theta}})\Big)\boldsymbol{x}^\top \mathbf{H}^{-1}(\bar{\boldsymbol{\theta}})\boldsymbol{x}_{i,a_i}\boldsymbol{x}_{i,a_i}^\top(\boldsymbol{\theta}^* - \bar{\boldsymbol{\theta}})\right| \\
&\le \sum_{i\in E}\left|\Big(\dot{\mu}(\boldsymbol{x}_{i,a_i}^\top \check{\boldsymbol{\theta}}_i) - \dot{\mu}(\boldsymbol{x}_{i,a_i}^\top \bar{\boldsymbol{\theta}})\Big)\right| \cdot \left|\boldsymbol{x}^\top \mathbf{H}^{-1}(\bar{\boldsymbol{\theta}})\boldsymbol{x}_{i,a_i}\right| \cdot \left|\boldsymbol{x}_{i,a_i}^\top(\boldsymbol{\theta}^* - \bar{\boldsymbol{\theta}})\right| \\
&\le \sum_{i\in E}(e^r-1)\dot{\mu}(\boldsymbol{x}_{i,a_i}^\top \bar{\boldsymbol{\theta}}) \cdot \left|\boldsymbol{x}^\top \mathbf{H}^{-1}(\bar{\boldsymbol{\theta}})\boldsymbol{x}_{i,a_i}\right| \cdot \left|\boldsymbol{x}_{i,a_i}^\top(\boldsymbol{\theta}^* - \bar{\boldsymbol{\theta}})\right| && ((21)) \\
&\le (e^r-1)\underbrace{\sqrt{\sum_{i\in E}\dot{\mu}(\boldsymbol{x}_{i,a_i}^\top \bar{\boldsymbol{\theta}})\big(\boldsymbol{x}^\top \mathbf{H}^{-1}(\bar{\boldsymbol{\theta}})\boldsymbol{x}_{i,a_i}\big)^2}}_{A_1}\underbrace{\sqrt{\sum_{i\in E}\dot{\mu}(\boldsymbol{x}_{i,a_i}^\top \bar{\boldsymbol{\theta}})\big(\boldsymbol{x}_{i,a_i}^\top(\boldsymbol{\theta}^* - \bar{\boldsymbol{\theta}})\big)^2}}_{A_2}. \\
& && \text{(Cauchy-Schwarz)}
\end{aligned}$$

For the term A_1,

$$\begin{aligned}
A_1^2 &= \boldsymbol{x}^\top \mathbf{H}^{-1}(\bar{\boldsymbol{\theta}})\left(\sum_{i\in E}\dot{\mu}(\boldsymbol{x}_{i,a_i}^\top \bar{\boldsymbol{\theta}})\boldsymbol{x}_{i,a_i}\boldsymbol{x}_{i,a_i}^\top\right)\mathbf{H}^{-1}(\bar{\boldsymbol{\theta}})\boldsymbol{x} \\
&\le \boldsymbol{x}^\top \mathbf{H}^{-1}(\bar{\boldsymbol{\theta}})\mathbf{H}(\bar{\boldsymbol{\theta}})\mathbf{H}^{-1}(\bar{\boldsymbol{\theta}})\boldsymbol{x} \\
&\le \kappa\|\boldsymbol{x}\|^2_{\mathbf{V}_E^{-1}}. && (\mathbf{V}_E \preceq \kappa\mathbf{H}(\bar{\boldsymbol{\theta}}))
\end{aligned}$$

For the term A_2,

$$A_2^2 \le \sum_{i\in E} L_\mu\big(\boldsymbol{x}_{i,a_i}^\top(\boldsymbol{\theta}^* - \bar{\boldsymbol{\theta}})\big)^2 \le |E|r^2 L_\mu.$$

Substituting term A_1 and term A_2 back yields the desired result, completing the proof. □

E.2 Proof of Lemma 29

Proof. In this section, we fix t and s and drop the notation for it for simplicity. Redefine $E = \{t^{(1)}, t^{(2)}, \dots, t^{(|E|)}\}$ chronologically and set $\mathbf{V}_0 = \kappa\lambda\mathbf{I}$ and $\mathbf{V}_j = \mathbf{V}_{j-1} + \boldsymbol{x}_{t^{(j)},a_{t^{(j)}}} \boldsymbol{x}_{t^{(j)},a_{t^{(j)}}}^\top$. If $t^{(j)} \in E$, then condition (a) was triggered at level s, which means

$$\begin{aligned} w^{(s)}_{t^{(j)},a_{t^{(j)}}} = \alpha \|\boldsymbol{x}_{t^{(j)},a_{t^{(j)}}}\|_{\mathbf{V}_{j-1}^{-1}} &\geq 2^{-s} \\ \implies \alpha^2 \|\boldsymbol{x}_{t^{(j)},a_{t^{(j)}}}\|^2_{\mathbf{V}_{j-1}^{-1}} &\geq 4^{-s} \\ \implies \alpha^2 \min\left\{1,\ \|\boldsymbol{x}_{t^{(j)},a_{t^{(j)}}}\|^2_{\mathbf{V}_{j-1}^{-1}}\right\} &\geq 4^{-s}, \end{aligned}$$

where the last inequality holds since $\|\boldsymbol{x}_{t^{(j)},a_{t^{(j)}}}\|^2_{\mathbf{V}_{j-1}^{-1}} \leq \|\boldsymbol{x}_{t^{(j)},a_{t^{(j)}}}\|_2^2 / \lambda_{\min}(\mathbf{V}_{j-1}) \leq 1/(\kappa\lambda) \leq 1$ where the last inequality follows from our choice of $\lambda \geq 1/\kappa$. We can proceed as

$$\begin{aligned} \implies |E|4^{-s} &\leq \alpha^2 \sum_{j=1}^{|E|} \min\left\{1,\ \|\boldsymbol{x}_{t^{(j)},a_{t^{(j)}}}\|^2_{\mathbf{V}_{j-1}^{-1}}\right\} \\ &\leq 2\alpha^2 d \log(1 + |E|/(\kappa\lambda d)) && \text{(Lemma 34)} \\ &\leq 2\alpha^2 d \log(1 + T/(\kappa\lambda d)). \end{aligned}$$

Finally, moving 4^{-s} to the right-hand side completes the proof. □

E.3 Proof of Lemma 30

Recall the definition of the loss function

$$\mathcal{L}_t^{(s)}(\boldsymbol{\theta}) \coloneqq \sum_{i \in P^{(s)}(t)} \ell(\boldsymbol{x}_{i,a_i}, r_i; \boldsymbol{\theta}) + \frac{\lambda}{2}\|\boldsymbol{\theta}\|_2^2.$$

In this section, we fix t and s and drop the notation for it for simplicity. So we simply write as $\mathcal{L}(\boldsymbol{\theta}) = \sum_{i \in P} \ell(\boldsymbol{x}_{i,a_i}, r_i; \boldsymbol{\theta}) + \frac{\lambda}{2}\|\boldsymbol{\theta}\|_2^2$.

Proof. $\mathcal{L}(\boldsymbol{\theta})$ is a strongly convex, differentiable function and $\bar{\boldsymbol{\theta}}$ minimizes $\mathcal{L}(\boldsymbol{\theta})$ over a convex set Θ. Therefore, we have $\langle \nabla\mathcal{L}(\bar{\boldsymbol{\theta}}),\ \boldsymbol{\theta} - \bar{\boldsymbol{\theta}} \rangle \geq 0$ for all $\boldsymbol{\theta} \in \Theta$. Substituting

$$\nabla\mathcal{L}(\bar{\boldsymbol{\theta}}) = \sum_{i \in P} \big(\mu(\boldsymbol{x}_{i,a_i}^\top \bar{\boldsymbol{\theta}}) - r_i\big)\boldsymbol{x}_{i,a_i} + \lambda\bar{\boldsymbol{\theta}}, \quad \boldsymbol{\theta} \leftarrow \boldsymbol{\theta}^* \in \Theta,$$

we have

$$\begin{aligned} &- \left\langle \sum_{i \in P} \big(\mu(\boldsymbol{x}_{i,a_i}^\top \bar{\boldsymbol{\theta}}) - r_i\big)\boldsymbol{x}_{i,a_i},\ \bar{\boldsymbol{\theta}} - \boldsymbol{\theta}^* \right\rangle \geq \lambda\langle \bar{\boldsymbol{\theta}},\ \bar{\boldsymbol{\theta}} - \boldsymbol{\theta}^* \rangle \\ \implies &- \left\langle \sum_{i \in P} \big[\big(\mu(\boldsymbol{x}_{i,a_i}^\top \bar{\boldsymbol{\theta}}) - \mu(\boldsymbol{x}_{i,a_i}^\top \boldsymbol{\theta}^*)\big) - \big(r_i - \mu(\boldsymbol{x}_{i,a_i}^\top \boldsymbol{\theta}^*)\big)\big]\boldsymbol{x}_{i,a_i},\ \bar{\boldsymbol{\theta}} - \boldsymbol{\theta}^* \right\rangle \geq \lambda\langle \bar{\boldsymbol{\theta}},\ \bar{\boldsymbol{\theta}} - \boldsymbol{\theta}^* \rangle \\ \implies &\left\langle \sum_{i \in P} \epsilon_i \boldsymbol{x}_{i,a_i},\ \bar{\boldsymbol{\theta}} - \boldsymbol{\theta}^* \right\rangle \geq \left\langle \sum_{i \in P} \big(\mu(\boldsymbol{x}_{i,a_i}^\top \bar{\boldsymbol{\theta}}) - \mu(\boldsymbol{x}_{i,a_i}^\top \boldsymbol{\theta}^*)\big)\boldsymbol{x}_{i,a_i},\ \bar{\boldsymbol{\theta}} - \boldsymbol{\theta}^* \right\rangle + \lambda\langle \bar{\boldsymbol{\theta}},\ \bar{\boldsymbol{\theta}} - \boldsymbol{\theta}^* \rangle. \\ \implies &\underbrace{\left\langle \sum_{i \in P} \epsilon_i \boldsymbol{x}_{i,a_i},\ \bar{\boldsymbol{\theta}} - \boldsymbol{\theta}^* \right\rangle}_{A_1} \geq \underbrace{\left\langle \sum_{i \in P} \big(\mu(\boldsymbol{x}_{i,a_i}^\top \bar{\boldsymbol{\theta}}) - \mu(\boldsymbol{x}_{i,a_i}^\top \boldsymbol{\theta}^*)\big)\boldsymbol{x}_{i,a_i} + \lambda(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}^*),\ \bar{\boldsymbol{\theta}} - \boldsymbol{\theta}^* \right\rangle}_{A_2} + \underbrace{\lambda\langle \boldsymbol{\theta}^*,\ \bar{\boldsymbol{\theta}} - \boldsymbol{\theta}^* \rangle}_{A_3}. \end{aligned}$$

For the term A_1,

$$A_1 \leq \left\| \sum_{i \in P} \epsilon_i \boldsymbol{x}_{i,a_i} \right\|_{\mathbf{V}_P^{-1}} \cdot \|\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_{\mathbf{V}_P}. \qquad \text{(Cauchy-Schwarz)}$$

For the term A_2,

$$\begin{aligned}
A_2 &= \sum_{i\in P}(\bar{\boldsymbol{\theta}}-\boldsymbol{\theta}^*)^\top \dot{\mu}(\boldsymbol{x}_{i,a_i}^\top \check{\boldsymbol{\theta}}_i)\boldsymbol{x}_{i,a_i}\boldsymbol{x}_{i,a_i}^\top(\bar{\boldsymbol{\theta}}-\boldsymbol{\theta}^*) + (\bar{\boldsymbol{\theta}}-\boldsymbol{\theta}^*)^\top\lambda\mathbf{I}(\bar{\boldsymbol{\theta}}-\boldsymbol{\theta}^*) && \text{(mean-value)}\\
&= (\bar{\boldsymbol{\theta}}-\boldsymbol{\theta}^*)^\top\left(\sum_{i\in P}\dot{\mu}(\boldsymbol{x}_{i,a_i}^\top \check{\boldsymbol{\theta}}_i)\boldsymbol{x}_{i,a_i}\boldsymbol{x}_{i,a_i}^\top + \lambda\mathbf{I}\right)(\bar{\boldsymbol{\theta}}-\boldsymbol{\theta}^*)\\
&\geq \frac{1}{\kappa}(\bar{\boldsymbol{\theta}}-\boldsymbol{\theta}^*)^\top\left(\sum_{i\in P}\boldsymbol{x}_{i,a_i}\boldsymbol{x}_{i,a_i}^\top + \kappa\lambda\mathbf{I}\right)(\bar{\boldsymbol{\theta}}-\boldsymbol{\theta}^*) && (\check{\boldsymbol{\theta}}_i\in\Theta)\\
&\geq \frac{1}{\kappa}\|\bar{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\|_{\mathbf{V}_P}^2.
\end{aligned}$$

For the term A_3,

$$A_3 \geq -\lambda\|\boldsymbol{\theta}^*\|_{\mathbf{V}_P^{-1}}\cdot\|\bar{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\|_{\mathbf{V}_P} \geq -B\sqrt{\frac{\lambda}{\kappa}}\|\bar{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\|_{\mathbf{V}_P},$$

where the first inequality follows from the Cauchy-Schwarz inequality, and the second inequality holds since $\|\boldsymbol{\theta}^*\|_{\mathbf{V}_P^{-1}} \leq \|\boldsymbol{\theta}^*\|_2/\lambda_{\min}^{1/2}(\mathbf{V}_P) \leq B/\sqrt{\kappa\lambda}$.

Substituting term A_1, term A_2, and term A_3 back, and dividing both sides with $\frac{1}{\kappa}\|\bar{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\|_{\mathbf{V}_P}$ gives

$$\|\bar{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\|_{\mathbf{V}_P} \leq \kappa\underbrace{\left\|\sum_{i\in P}\epsilon_i\boldsymbol{x}_{i,a_i}\right\|_{\mathbf{V}_P^{-1}}}_{A_4} + B\sqrt{\kappa\lambda}.$$

For term A_4, we prepare to apply the self-normalized bound for vector-valued martingales, which is the well known result from Theorem 1 of Abbasi-Yadkori et al. (2011).

Redefine $P = \{t^{(1)}, t^{(2)}, \ldots, t^{(|P|)}\}$ chronologically, and let $\mathbf{V}_0 = \kappa\lambda\mathbf{I}$ and $\mathbf{V}_j = \mathbf{V}_{j-1} + \boldsymbol{x}_{t^{(j)},a_{t^{(j)}}}\boldsymbol{x}_{t^{(j)},a_{t^{(j)}}}^\top$. Then

$$A_4 = \left\|\sum_{j=1}^{|P|}\epsilon_{t^{(j)}}\boldsymbol{x}_{t^{(j)},a_{t^{(j)}}}\right\|_{\mathbf{V}_{|P|}^{-1}}.$$

We can see that $\boldsymbol{x}_{t^{(j)},a_{t^{(j)}}}$ is $\mathcal{F}_{t^{(j)}}$-measurable, and $\epsilon_{t^{(j)}}$ is $\mathcal{F}_{t^{(j+1)}}$-measurable. Also, $\mathbb{E}[\epsilon_{t^{(j)}} \mid \mathcal{F}_{t^{(j)}}] = 0$, $|\epsilon_{t^{(j)}}| \leq 1$, therefore $1/2$-subGaussian noise. Applying Lemma 35, simultaneously for all $j \geq 0$, with probability at least $1-\delta$,

$$A_4 \leq \frac{1}{\sqrt{2}}\sqrt{d\log(1+|P|/(\kappa\lambda d)) + \log(1/\delta)} \leq \frac{1}{\sqrt{2}}\sqrt{d\log(1+T/(\kappa\lambda d)) + \log(1/\delta)}.$$

Finally, substituting term A_4 back, replacing $\delta \leftarrow \delta/(2S)$, and taking the union bound for all $s \in [S]$ yields that, with probability at least $1-\delta/2$, simultaneously for all $t \in [T]$, $s \in [S]$, the desired result holds. □

E.4 Proof of Lemma 32

Denote the good event $\mathcal{E}$ when Theorem 13 holds.

Proof. We prove this by induction.

First, for $s = 1$, $a_t^* \in A^{(1)} = [K]$. Set $s' \in [s_t]$ and assume that for $s = s'$, $a_t^* \in A^{(s')}$.

Now, Consider $s = s' + 1$. On the good event $\mathcal{E}$,

$$\left|m_{t,a^*}^{(s')} - \boldsymbol{x}_{t,a^*}^\top\boldsymbol{\theta}^*\right| \leq w_{t,a^*}^{(s')} \leq 2^{-s'},$$

where the first inequality follows from Theorem 13, and the second inequality holds since the algorithm reach level $s'+1$ implies that the condition (a) in level s' is not satisfied. Denote $\hat{a} = \arg\max_{a \in A^{(s')}} m_{t,a}^{(s')}$. Then,

$$\begin{aligned} m_{t,a^*}^{(s')} \geq \boldsymbol{x}_{t,a^*}^\top \boldsymbol{\theta}^* - 2^{-s'} \geq \boldsymbol{x}_{t,\hat{a}}^\top \boldsymbol{\theta}^* - 2^{-s'} \geq m_{t,\hat{a}}^{(s')} - 2 \cdot 2^{-s'} \\ \implies m_{t,a^*}^{(s')} \geq \max_{a \in A^{(s')}} m_{t,a}^{(s')} - 2 \cdot 2^{-s'}, \end{aligned}$$

which implies that $a^* \in A^{(s'+1)}$ by condition (c) in level s', completing the proof. □

F Auxiliary Lemmas

Lemma 33 (Bernstein inequality)**.** *For independent random variables $X_1, \ldots, X_n$ with $\mathbb{E}[X_i] = 0$, $|X_i| \leq b$ almost surely, and $\sum_{i=1}^n \mathrm{Var}(X_i) \leq V$, then for all $L \geq 0$,*

$$\mathbb{P}\left(\left| \sum_{i=1}^n X_i \right| \geq \sqrt{2VL} + \frac{bL}{3} \right) \leq 2e^{-L}.$$

Lemma 34 (Elliptical potential lemma, Lemma 11 of Abbasi-Yadkori et al. (2011))**.** *For any $\lambda > 0$ and sequence $\{\boldsymbol{x}_t\}_{t=1}^T \in \mathbb{R}^d$, define $\mathbf{V}_0 = \lambda \mathbf{I}$ and $\mathbf{V}_t = \mathbf{V}_{t-1} + \boldsymbol{x}_i \boldsymbol{x}_i^\top$. Then, provided that $\|\boldsymbol{x}_t\|_2 \leq L$ holds for all $t \in [T]$, we have*

$$\sum_{t=1}^T \min\left\{1, \ \|\boldsymbol{x}_t\|_{\mathbf{V}_{t-1}^{-1}}^2\right\} \leq 2 \log \frac{\det \mathbf{V}_T}{\det \mathbf{V}_0} \leq 2d \log\left(1 + \frac{TL^2}{d\lambda}\right).$$

Lemma 35 (Theorem 1 of Abbasi-Yadkori et al. (2011))**.** *Let $\{\mathcal{F}_t\}_{t=1}^\infty$ be a filtration. Let $\{\boldsymbol{x}_t\}_{t=1}^\infty$, $\{\epsilon_t\}_{t=1}^\infty$ be stochastic processes such that $\boldsymbol{x}_t \in \mathbb{R}^d$ is $\mathcal{F}_t$-measurable with $\|\boldsymbol{x}_t\|_2 \leq L$, and $\epsilon_t \in \mathbb{R}$ is $\mathcal{F}_{t+1}$-measurable and R-subGaussian condition on $\mathcal{F}_t$. Define $\mathbf{V}_0 = \lambda \mathbf{I}$ and $\mathbf{V}_t = \lambda \mathbf{I} + \sum_{i=1}^t \boldsymbol{x}_i \boldsymbol{x}_i^\top$. Fix $\delta \in (0, 1]$. Then with probability at least $1 - \delta$, simultaneously for all $t \geq 0$,*

$$\left\| \sum_{i=1}^t \epsilon_i \boldsymbol{x}_i \right\|_{\mathbf{V}_t^{-1}}^2 \leq 2R^2 \left(\log \frac{\det \mathbf{V}_t}{\det \mathbf{V}_0} + \log(1/\delta) \right) \leq 2R^2 \left(d \log\left(1 + \frac{tL^2}{d\lambda}\right) + \log(1/\delta) \right).$$

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: We state the paper's contributions and scope in the abstract and introduction.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We discuss potential improvements to our results in the Conclusion and future work section.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: We state the assumptions and a complete proof.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: We present the experimental results and details in the Experiments and Experimental details sections.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: We release the code.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: Check the Experimental details section.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: The figure reports error bars.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.

- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Check the Experimental details section.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?

   Answer: [Yes]

   Justification: Our research conducted in the paper conforms to the NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: There is no societal impact of the work.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper poses no such risks.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [N/A]

    Justification: The paper does not use existing assets.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: The paper does not release new assets.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: The core method development in this research does not involve LLMs as any important components.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.