Practical and Optimal Algorithm for Linear Contextual Bandits with Rare Parameter Updates

Sanghoon Yu [1] **Min-hwan Oh** [1]

Abstract

We study linear contextual bandits under *rare parameter updates*: the learner may incorporate reward feedback into its parameter estimate only at a small number of update times, while still observing contexts online and selecting actions sequentially. This viewpoint clarifies a practical distinction that is often blurred in the literature: many "strictly batched" methods additionally restrict *within-interval context adaptivity*, meaning that the action rule inside an interval cannot depend on the *sequence* of realized contexts/actions in that interval (beyond the current round's context). For linear contextual bandits, we propose two practical algorithms with only $\mathcal{O}(\log \log T)$ parameter updates. Our first algorithm `BLCE-G` attains minimax-optimal regret (up to polylogarithmic factors in T) *simultaneously* in both the small-K and large-K regimes under a static schedule. Our second algorithm `BLCE` removes the near G-optimal design step—a dominant computational bottleneck in prior strictly batched static-grid methods—yet preserves minimax-optimal regret and achieves the lowest known runtime complexity among optimal algorithms. We further extend these rare-update and computational principles to generalized linear contextual bandits. Overall, our results yield statistically optimal algorithms under $\mathcal{O}(\log \log T)$ parameter updates that are also computationally efficient in practice.

1. Introduction

Stochastic *linear contextual bandits* are a cornerstone of sequential decision-making, where an agent repeatedly selects actions from a time-varying, feature-based set and observes rewards generated by an unknown linear model (Abe & Long, 1999; Auer, 2002; Abe et al., 2003; Dani et al., 2008; Li et al., 2010; Chu et al., 2011; Abbasi-Yadkori et al., 2011; Li et al., 2019; Lattimore & Szepesvári, 2020; Kirschner et al., 2021). The resulting low-dimensional structure enables efficient generalization and has seen wide application, from recommender systems (Li et al., 2010) and inventory control (Jin et al., 2021a) to clinical trials and precision medicine (Lu et al., 2021).

In deployments of bandit algorithms, the practical bottleneck is often not selecting an action given a context, but *updating* the underlying model once new reward feedback arrives: model retraining, confidence set recomputation, or posterior updates may require expensive pipelines, logging joins, privacy checks, or human-in-the-loop validation. This motivates learning under *rare parameter updates*, where the learner is allowed to incorporate rewards into its parameter estimate only at a small number of update times, while contexts continue to arrive online and decisions must still be made sequentially.

A common special case is the *batched-feedback* protocol, where rewards from an interval become available only at the end of that interval. However, in the contextual bandit literature, the term "batched" is also frequently used in a stronger sense that restricts *within-interval context adaptivity*. Concretely, many strictly batched methods commit at the start of an interval to an action rule that cannot depend on the *sequence* of realized contexts/actions within that interval. This stronger restriction is not inherent to batched feedback: contexts are observed online before choosing actions and are thus readily available. For example, Karbasi et al. (2021) formalize a batch policy at time t in the current interval as a mapping $\pi_t : H_{t_{\ell-1}} \times C_t \to \mathcal{A}$, where $H_{t_{\ell-1}}$ denotes the history up to the previous interval boundary and C_t denotes the set of contexts observed up to and including time t. This formalization permits dependence on within-interval contexts. Motivated by practice, we therefore separate these notions and focus on *rare parameter updates* as the key constraint, while allowing (and exploiting) reward-free within-interval updates based on the arriving contexts.

Recently, batched contextual bandit algorithms have been shown to achieve strong near-optimal regret guarantees with

[1]Seoul National University, Seoul, Republic of Korea. Correspondence to: Min-hwan Oh <minoh@snu.ac.kr>.

Proceedings of the 43^{rd} International Conference on Machine Learning, Seoul, South Korea. PMLR 306, 2026. Copyright 2026 by the author(s).

as few as $\mathcal{O}(\log\log T)$ parameter-update times under i.i.d. contexts (Ruan et al., 2021; Hanna et al., 2023a;b; Zhang et al., 2025). Yet, they rely on computationally expensive primitives such as G-optimal design, or suffer from exponential dependence on dimension due to discretization; see Section 3. As a result, despite attractive small number of update frequency and their regret guarantees, their computational complexity is impractically high (see Section 2 for runtime experiments), making these approaches hard to deploy in realistic pipelines where batched learning is meant to reduce cost. This mismatch motivates our method, which preserves rare parameter updates while avoiding these costly subroutines and exploiting reward-free within-interval context information to remain computationally practical.

Our main message in this work is simple: one can retain the *same* $\mathcal{O}(\log\log T)$ number of parameter updates while substantially improving practical efficiency: by allowing *reward-free* within-interval state updates based only on the arriving contexts and chosen actions, we can avoid heavy G-optimal design steps and obtain algorithms that are both minimax-optimal and fast. This is precisely the operational regime practitioners often want: infrequent reward-dependent retraining, but lightweight online processing of contexts that are already observed.

The discussion above leads to the following questions:

- For *linear contextual bandits*, can we design algorithms that achieve minimax-optimal regret in both small-K and large-K regimes using only $\mathcal{O}(\log\log T)$ parameter updates? Can we also remove the G-optimal design procedure while preserving minimax optimality and practical runtime?
- Can we extend the same rare-update and computational principles to *generalized linear* contextual bandits, obtaining near-optimal regret without any dependence on worst-case curvature parameters (such as κ), while keeping $\mathcal{O}(\log\log T)$ parameter updates?

Positive answers to these questions would unify theory and practice in linear and generalized linear contextual bandits, leading to algorithms that remain statistically optimal and computationally efficient under limited adaptivity. Our main contributions are summarized as follows:

- **Tightest regret bounds for linear contextual bandits under $\mathcal{O}(\log\log T)$ parameter updates.** We introduce `BLCE-G`, which combines near G-optimal design and arm elimination. It achieves the worst-case regret bound $\mathcal{O}\big(\sqrt{dT}(\sqrt{\log(KT)} \wedge \sqrt{d+\log T})\sqrt{\log d \log\log T}\big)$, where K is the number of arms, d is the feature dimension, and T is the horizon. `BLCE-G` simultaneously matches the minimax lower bounds $\Omega(d\sqrt{T})$ in the large-K regime ($K \geq \Omega(e^d)$) and $\Omega(\sqrt{dT\log K})$ in the small-K regime ($K \leq \mathcal{O}(e^d)$), up to logarithmic factors.
- **First minimax-optimal algorithm in rare updates without G-optimal design.** We propose `BLCE`, which replaces the G-optimal design step with uncertainty-driven exploration combined with arm elimination. It still achieves minimax-optimal regret and attains the lowest total time complexity $\mathcal{O}(Kd^2T\log\log T)$ among optimal algorithms. Its guarantees also extend beyond conventional i.i.d. contexts, as discussed in Remark 1.
- **Redefining the limited adaptability landscape: fully sequential vs. strictly batched vs. rare parameter updates.** We explicitly distinguish three operational settings that have often been conflated : (i) fully sequential (fully adaptive) contextual bandits; (ii) *strictly batched* contextual bandits, which disallow within-interval context adaptivity; and (iii) rare parameter updates, which restrict reward-dependent parameter updates but allow reward-free within-interval context adaptivity. We place prior work and our algorithms within this taxonomy to enable fair comparisons (see Section 2 and Table 1).
- **Extensions to generalized linear contextual bandits.** We extend the same rare-update and computational principles to generalized linear rewards and obtain an algorithm whose operation and regret bounds are independent of the worst-case curvature parameter κ, while retaining $\mathcal{O}(\log\log T)$ parameter updates and near-optimal regret guarantees.
- **Practical superiority.** Our experiments demonstrate that `BLCE-G`, `BLCE`, and `BGLE` consistently outperform prior linear and generalized linear contextual baselines across various instances, combining provable efficiency with strong empirical performance and substantially reduced runtime overhead.

2. Batched Feedback, Strict Batching, and Rare Parameter Updates

The term "batched bandits" is used in the literature in multiple, operationally distinct senses. Since our goal is to develop *practical* algorithms and enable fair comparisons, we make these distinctions explicit. In contextual problems, two axes are orthogonal: (i) when rewards are revealed (the feedback protocol), and (ii) whether the learner may adapt *within an interval* to the realized contexts and past actions. We organize the landscape into three settings—*fully sequential*, *strictly batched*, and *rare parameter updates*—and emphasize that *batched feedback* is a reward-delay constraint that can underlie either of the latter two.

Table 1. Worst-case regret, *parameter-update* complexity, and time complexity comparison for linear contextual bandits. We distinguish *rare parameter updates* (the number of times reward feedback is incorporated to update the parameter estimate) from the stronger *strict batching* restriction that the algorithm is *not* within-interval context-adaptive. **Both BLCE-G and BLCE achieve minimax-optimal regret**, matching the minimax lower bounds $\Omega(\sqrt{dT\log K} \wedge d\sqrt{T})$ (Dani et al., 2008; He et al., 2022) across all regimes while requiring only $\mathcal{O}(\log\log T)$ parameter updates. Among optimal algorithms, **BLCE-G attains the tightest regret, whereas BLCE achieves the lowest time complexity**. Here, $\mathcal{T}_{\mathrm{opt}}$ denotes the cost of a single call to the linear optimization oracle.

Paper	Worst-Case Regret	Parameter Updates	Context Adaptive	Time Complexity
Abbasi-Yadkori et al. (2011)	$\mathcal{O}(d\sqrt{T}\log T)$	$\mathcal{O}(d\log T)$	Yes	$\mathcal{O}((Kd+d^2)T + Kd^3\log T)$
Karbasi et al. (2021)	$\mathcal{O}(d^{3/2}\sqrt{T}\log T)$	$\mathcal{O}(K\log T)$	Yes	$\mathcal{O}((Kd+d^2)T)$
Ruan et al. (2021)	$\mathcal{O}(\sqrt{dT\log(dKT)\log d}\log\log T)$	$\mathcal{O}(\log\log T)$	No	$\mathcal{O}(Kd^4T(\log T+\log d))$
Hanna et al. (2023b)	$\mathcal{O}(d\sqrt{T\log T}\log\log T)$	$\mathcal{O}(\log\log T)$	No	$\Omega(T^d)$
Hanna et al. (2023a)	$\mathcal{O}(d^{3/2}\sqrt{T}\log T\log\log T)$	$\mathcal{O}(\log\log T)$	No	$\mathcal{O}(d^3\mathcal{T}_{\mathrm{opt}}T\log d\log^3 T\log\log T)$ $+\mathcal{O}(d^4\log d\log^3 T\log\log T)$
Zhang et al. (2025)	$\mathcal{O}(\sqrt{dT\log(dKT)\log T}\ \log(dT)\log\log T)$	$\mathcal{O}(\log\log T)$	No	$\mathcal{O}(Kd^2T\log\log T)$ $+\mathcal{O}(Kd^{7/2}\sqrt{T\log(dKT)\log T})$
Algorithm 1 (BLCE-G)	$\min\begin{cases}\mathcal{O}(\sqrt{dT\log(KT)\log d\log\log T})\\ \mathcal{O}(\sqrt{d(d+\log T)T\log d\log\log T})\end{cases}$	$\mathcal{O}(\log\log T)$	Yes	$\mathcal{O}(Kd^2T(d+\log\log T))$
Algorithm 2 (BLCE)	$\min\begin{cases}\mathcal{O}(\sqrt{dT\log(KT)\log T\log\log T})\\ \mathcal{O}(\sqrt{d(d+\log T)T\log T\log\log T})\end{cases}$	$\mathcal{O}(\log\log T)$	Yes	$\mathcal{O}(Kd^2T\log\log T)$

Context Adaptive: Yes means that, within each interval between parameter updates, the action choice may depend on reward-free state variables that evolve with the realized sequence of contexts/actions (e.g., cumulative Gram matrix updates affecting exploration). No indicates a *strictly batched* rule: within each interval, the action choice depends only on the interval-start state (learned from past rewards) and the current round's context set, with no dependence on past within-interval contexts/actions.

We say the problem has *batched feedback* if the horizon $[T]$ is partitioned into B disjoint intervals and the rewards collected within an interval are revealed only at its endpoint. This is purely a constraint on *reward availability*: regardless of how rewards are delayed, contexts are still observed online before actions are chosen.

In the *fully sequential* setting, rewards are revealed immediately each round and the learner may update its parameter estimate and decision rule every round. In batched-feedback terminology, this corresponds to the special case $B = T$.

In contrast, we use the term *strictly batched* to denote a stronger restriction than batched feedback: within each interval $(\mathcal{T}_{\ell-1}, \mathcal{T}_\ell]$, the learner must commit at time $\mathcal{T}_{\ell-1}$ to an action rule that is *not* allowed to depend on the *realized sequence* of contexts and past actions within the interval. Many static-grid contextual algorithms that are referred to as "batched" adopt (often implicitly) this strict restriction to obtain the minimal $\mathcal{O}(\log\log T)$ parameter-update complexity (Ruan et al., 2021; Hanna et al., 2023b; Zhang et al., 2025).

Finally, we say an algorithm has *rare parameter updates* if it updates its reward-based parameter estimate (e.g., via ridge regression, an MLE, or a posterior update) at most B times over T rounds. Crucially, this notion imposes *no restriction* on how the learner processes contexts within an interval: the learner may maintain and update reward-free state variables online using the realized contexts and chosen actions, and may let its action choices depend on these evolving state variables. This distinction is practically important in contextual bandits because contexts are *readily available* online: delaying the use of already observed contexts until the next parameter update is neither required by the feedback protocol nor typically natural. Even when reward-driven retraining is expensive, updating lightweight statistics (e.g., Gram matrices, uncertainty scores, or elimination candidate sets) using contexts and actions is usually cheap and leverages information the learner has *already* observed. Allowing such reward-free within-interval adaptivity therefore bridges a natural gap between the two extremes: fully sequential methods (maximal reward adaptivity) and strictly batched methods (no within-interval context adaptivity).

Our algorithms operate under a batched-feedback protocol with a static schedule and perform only $\mathcal{O}(\log\log T)$ parameter updates. They are *not strictly batched*: they perform lightweight, reward-free within-interval updates that influence action selection. This design choice preserves the desired operational constraint (rare reward-based parameter retraining) while avoiding the dominant computational bottlenecks that arise when one insists on strict batching (see Section 3).

3. Bottleneck of Existing Strictly Batched Algorithms

A recurring theme in the strictly batched contextual bandits literature is that the lack of within-interval context adaptivity forces the algorithm to "pre-plan" exploration within each interval. To guarantee sufficient information gain *without* reacting to the realized sequence of contexts/actions,

existing strictly batched methods often invoke computationally heavy primitives: repeated G-optimal design computations, discretization of continuous spaces via fine nets, or repeated oracle calls. These steps can dominate the overall runtime, even when the number of parameter updates is only $\mathcal{O}(\log\log T)$.

Concretely, Ruan et al. (2021) repeatedly compute a G-optimal design distribution on the observed arm set as contexts arrive; such per-round design computation is costly, leading to a large polynomial dependence on d in the overall runtime. In contrast, Hanna et al. (2023b) reduce contextual bandits to a non-contextual linear bandit instance by discretizing the space using a fine $(1/T)$-net, whose size scales as $\Omega(T^d)$; this yields an exponential dependence on d which is impractical. Hanna et al. (2023a) (which is theoretically suboptimal) address large (possibly infinite) action spaces via optimization-oracle access and discuss (approximate) optimal design primitives; their framework can require many oracle calls, and computing a 1-approximate G-optimal design is NP-hard, so even constant-factor approximations may still be computationally demanding in practice. Finally, Zhang et al. (2025) exhibit a particularly sharp bottleneck concentrated in the *first* batch: their Algorithm 1 executes a (near) G-optimal design policy at every round of batch 1, incurring $\mathcal{O}(Kd^3)$ time per round; since their first-batch length scales as $\widetilde{\Theta}(\sqrt{dT}\wedge T)$, this contributes an additional high-order cost of order $\widetilde{\mathcal{O}}(Kd^{7/2}\sqrt{T}\wedge Kd^3T)$, also leading to a large polynomial dependence on d. The numerical evaluations of these algorithms consistently show the inefficiency mentioned above (see Table 2).

As we focus on *rare parameter updates* setting instead of strictly batched setting, it has not been immediately clear as to whether optimal results can be derived while avoiding the computational inefficiency mentioned above.

4. Related Work

A substantial literature on *limited adaptivity* and *batched* bandits spans settings from multi-armed to linear (contextual) models. Early work established near-optimal learning with few policy updates in the multi-armed setting (Perchet et al., 2016; Gao et al., 2019; Jin et al., 2021b;c), later extended to linear bandits under Gaussian-type features (Han et al., 2020) and adversarial features (Esfandiari et al., 2021), culminating in algorithms that achieve near-optimal regret with the minimal batch complexity $\mathcal{O}(\log\log T)$ (Ren et al., 2024; Yu & Oh, 2025). Although Yu & Oh (2025) attain minimax-optimal regret in both regimes, their analysis is restricted to non-contextual batched bandits and does not extend to the linear contextual setting.

In linear *contextual* bandits, recent static-grid algorithms (Ruan et al., 2021; Hanna et al., 2023b; Zhang et al., 2025) achieve optimal regret with only $\mathcal{O}(\log\log T)$ update times under i.i.d. contexts. However, these approaches rely on G-optimal design, do not leverage within-interval context information, and achieve minimax-optimality only in a single regime. In contrast, Abbasi-Yadkori et al. (2011) require $\mathcal{O}(d\log T)$ updates, which exceeds the $\mathcal{O}(\log\log T)$ barrier. Similarly, Karbasi et al. (2021) requires $\mathcal{O}(K\log T)$ updates and provides a suboptimal regret guarantee of $\widetilde{\mathcal{O}}(d^{3/2}\sqrt{T})$. Finally, Hanna et al. (2023a) improve computational efficiency but also obtain a suboptimal regret dependence $\widetilde{\mathcal{O}}(d^{3/2}\sqrt{T})$.

5. Preliminaries

5.1. Notations

For a set, $|\cdot|$ denotes its cardinality. For $x\in\mathbb{R}^d$, $\|x\|_2$ denotes the Euclidean norm, and for a positive definite matrix H, we write $\|x\|_H := \sqrt{x^\top Hx}$. For a matrix, $\mathrm{tr}(\cdot)$ and $\det(\cdot)$ denote its trace and determinant. For $n\in\mathbb{N}$, we write $[n]:=\{1,\dots,n\}$, and use $\wedge$ for the minimum operator. For symmetric matrices A, B of the same dimension, $A\preceq B$ (resp. $A\succeq B$) means that $B-A$ (resp. $A-B$) is positive semidefinite. The indicator $\mathbb{1}_{\{E\}}$ equals 1 if the event E occurs and 0 otherwise. Finally, the natural filtration is $\mathcal{F}_t := \sigma(\mathcal{A}_1, x_{1,a_1}, r_1, \dots, \mathcal{A}_t, x_{t,a_t}, r_t)$, with $\mathcal{F}_0$ being the trivial σ-algebra.

5.2. Problem Setting: Linear Contextual Bandits with Batched Feedback

We study the stochastic linear contextual bandit problem. At each round $t\in[T]$, the agent observes contextual features of K arms $\mathcal{A}_t := \{x_{t,1},\dots,x_{t,K}\}\subseteq\mathbb{R}^d$ and selects one arm $x_{t,a_t}\in\mathcal{A}_t$, receiving the reward $r_t = \langle x_{t,a_t},\theta^*\rangle + \eta_t$, where $\theta^*\in\mathbb{R}^d$ is unknown and η_t is independent σ-subgaussian noise. Following the convention in the linear contextual bandit literature (Ruan et al., 2021; Hanna et al., 2023a;b; Zhang et al., 2025), we assume that each feature set $\mathcal{A}_t$ is drawn i.i.d. from an unknown distribution $\mathcal{D}$ across rounds, while allowing arbitrary correlation among features presented within the same round. (In Remark 1, we discuss how Algorithm 2 and its analysis further relax this i.i.d. assumption.) The performance of the agent is measured by the *cumulative expected regret*

$$\mathcal{R}(T) := \mathbb{E}\left[\sum_{t=1}^{T}\left(\langle x_t^*,\theta^*\rangle - \langle x_{t,a_t},\theta^*\rangle\right)\right],$$

where $x_t^*\in\mathrm{argmax}_{x\in\mathcal{A}_t}\langle x,\theta^*\rangle$ denotes an optimal arm. We make the following standard assumptions.

Assumption 1. *$\|x\|_2\le 1$ for all $x\in\mathcal{A}_t$ and $\|\theta^*\|_2\le 1$.*

Assumption 2. *The noise η_t is a 1-subgaussian random variable for all $t\in[T]$.*

In the batched-feedback setting, the horizon $[T]$ is partitioned into B disjoint *parameter-update intervals*,

$$\{1,\dots,T\} = \underbrace{[\mathcal{T}_0{+}1,\dots,\mathcal{T}_1]}_{\text{interval } 1} \cup \cdots \cup \underbrace{[\mathcal{T}_{B-1}{+}1,\dots,\mathcal{T}_B]}_{\text{interval } B}.$$

At time $t \in (\mathcal{T}_{\ell-1}, \mathcal{T}_\ell]$, the agent may use reward information from all previous intervals (up to $\mathcal{T}_{\ell-1}$), while contexts up to and including time t are always observable. The rewards for all rounds within the current interval become available only at the end of that interval (time $\mathcal{T}_\ell$). Importantly, we do *not* require the action rule within an interval to be strictly batched: the agent may update *reward-free* state variables using observed contexts and chosen actions, while the parameter update that incorporates rewards occurs only at interval boundaries.

6. Algorithm & Regret Analysis

6.1. Note on the Update Schedule

Regardless of whether an algorithm is strictly batched or not, there is also a notion of *update schedules* (which can be confusing to readers not well-versed in batched bandit literature). Three types of update schedules are commonly considered: the *static grid*, where update times are fixed in advance; the *adaptive grid*, where interval lengths are chosen adaptively; and the *rare switching* schedule studied by Abbasi-Yadkori et al. (2011), which allows arbitrary interaction while limiting the number of parameter recomputations. Following Ruan et al. (2021); Hanna et al. (2023a;b); Zhang et al. (2025), we adopt a *static grid* (most stringent one) that yields $B = \mathcal{O}(\log\log T)$ parameter updates.

6.2. Proposed Algorithms: BLCE-G and BLCE

We propose two algorithms BLCE-G and BLCE for linear contextual bandits under rare parameter updates. Let us start with explanation on our first algorithm, BLCE-G, which stands for the *Batched-Feedback Linear Contextual Bandit with Elimination and G-optimal design*, whose pseudocode is given in Algorithm 1. In the first interval, rounds are divided into two phases with ratio $c : (1-c)$. During the first c-fraction, arms are sampled according to a near G-optimal design over $\mathcal{A}_t$ (Line 5), while in the remaining $(1-c)$-fraction the algorithm selects the most informative direction with respect to the current Gram matrix (Line 7). To reduce computational and runtime cost, we adopt a relaxation of the G-optimal design, namely the near G-optimal design, which loosens the bound by at most a factor of two. Formally, for any arm set $X \subset \mathbb{R}^d$, there exists a design distribution $\mathcal{K}_X$ supported on X such that

$$\max_{x\in X} x^\top (\mathbb{E}_{z\sim\mathcal{K}_X}[zz^\top])^{-1} x \le 2d.$$

As shown in Corollary 4, such a design can be computed in time $\mathcal{O}(Kd^3)$. For fair comparison, we also account for this cost when evaluating other algorithms that rely on G-optimal design (Ruan et al., 2021; Hanna et al., 2023b; Zhang et al., 2025). After each arm pull, the inverse Gram matrix is updated using the Sherman-Morrison formula (reducing the cost from $\mathcal{O}(d^3)$ to $\mathcal{O}(d^2)$) (Line 8). At the end of the first interval, we set $V_1 \coloneqq H_{\mathcal{T}_1}$, compute the regression estimate $\hat\theta_1$, and reinitialize $H_{\mathcal{T}_1}$ for the next interval (Line 9).

For any interval $\ell \ge 2$, the algorithm eliminates suboptimal arms $\ell-1$ times using the past estimates $\hat\theta_1,\dots,\hat\theta_{\ell-1}$, yielding a nested sequence of feasible sets $\mathcal{A}_t^{(1)},\dots,\mathcal{A}_t^{(\ell-1)}$ (Line 13). The elimination threshold $\varepsilon_{t,k}$ for $k \in [\ell-1]$ is defined as

$$\max_{y\in\mathcal{A}_t^{(k-1)}} \|y\|_{V_k^{-1}} \Bigg(\sqrt{2\log\left(|\mathcal{A}_t^{(k-1)}|(B-1)T^2\right)} + \sqrt{\lambda} \wedge 2\sqrt{\log\left(\frac{2^{6d-5}\pi d(B-1)^2}{15^{d-1}/T^2}\right)} + 2\sqrt{\lambda} \Bigg).$$

Within interval $\ell \ge 2$, the rounds are partitioned in the ratio $c^2 : c(1-c) : (1-c)$. In the first c^2-fraction, arms are sampled according to a near G-optimal design over $\mathcal{A}_t^{(\ell-1)}$ (Line 15). In the next $c(1-c)$-fraction, the algorithm selects the most informative direction relative to the Gram matrix (Line 17). In the final $(1-c)$-fraction, arms are chosen greedily with respect to the latest estimate (Line 19). As before, after each pull the Gram matrix is updated (Line 20), and at the end of the interval we set $V_\ell \coloneqq H_{\mathcal{T}_\ell}$, compute $\hat\theta_\ell$, and reinitialize $H_{\mathcal{T}_\ell}$ for the next interval (Line 21).

Now, let us introduce our second algorithm, BLCE, which stands for *Batched-Feedback Linear Contextual Bandit with Elimination*. The pseudocode of BLCE is given in Algorithm 2. Relative to Algorithm 1, BLCE eliminates the near G-optimal design phase and instead lengthens the uncertainty-driven exploration phase to occupy those rounds. To our knowledge, BLCE is the first linear contextual bandit algorithm with *rare parameter updates* that achieves minimax-optimal regret without relying on G-optimal design, which has traditionally been regarded as essential in strictly batched static-grid settings (Ruan et al., 2021; Hanna et al., 2023b; Zhang et al., 2025). Notably, both BLCE-G and BLCE avoid enforcing any fixed choice of $c \in (0,1)$, thereby providing theoretical guarantees together with practical flexibility in balancing exploration and exploitation.

6.3. Regret Analysis for Linear Contextual Bandits

Theorem 1 (Regret of BLCE-G). *Consider running the BLCE-G algorithm for T rounds with K arms in d dimensions. The worst-case cumulative regret satisfies*

$$\mathcal{R}(T) = \mathcal{O}\Big(\sqrt{dT}\,(\sqrt{\log(KT)} \wedge \sqrt{d+\log T})\sqrt{\log d \log\log T}\Big) = \widetilde{\mathcal{O}}\Big(\sqrt{dT\log K} \wedge d\sqrt{T}\Big).$$

Algorithm 1 BLCE-G

Input: Horizon T; interval end times $\mathcal{T}_1 = \left\lceil \frac{\sqrt{T}}{\log_2 \log_2 T} \right\rceil + 1$, $\mathcal{T}_\ell = \left(\mathcal{T}_{\ell-1} + \left\lceil \frac{T^{1-2^{-\ell}}}{\log_2 \log_2 T} \right\rceil + 2 \right) \wedge T$ for $\ell \geq 2$; number of intervals B, with $\mathcal{T}_B = T$; within-interval allocation rate c

Initialize: $\lambda \leftarrow \log(dT)$, $H_0 \leftarrow \lambda I$;
for $t \leftarrow 1, 2, \ldots, \mathcal{T}_1$ **do**
  if $t \leq \lceil c\sqrt{T} / \log_2 \log_2 T \rceil$ **then**
    Pull arm $x_{t,a_t} \sim \pi_{G'}(\mathcal{A}_t)$;
  else
    Pull arm $x_{t,a_t} \in \arg\max_{x \in \mathcal{A}_t} \|x\|_{H_{t-1}^{-1}}$;
  $H_t^{-1} \leftarrow H_{t-1}^{-1} - H_{t-1}^{-1} x_{t,a_t} x_{t,a_t}^\top H_{t-1}^{-1} / (1 + x_{t,a_t}^\top H_{t-1}^{-1} x_{t,a_t})$;
$V_1^{-1} \leftarrow H_{\mathcal{T}_1}^{-1}$, $\hat{\theta}_1 \leftarrow V_1^{-1} \sum_{t=1}^{\mathcal{T}_1} r_t x_{t,a_t}$, $H_{\mathcal{T}_1} \leftarrow \lambda I$;
for $\ell \leftarrow 2, \ldots, B$ **do**
  for $t \leftarrow \mathcal{T}_{\ell-1} + 1, \ldots, \mathcal{T}_\ell$ **do**
    for $k \leftarrow 1, \ldots, \ell - 1$ **do**
      $x_t^{(k)} \leftarrow \arg\max_{x \in \mathcal{A}_t^{(k-1)}} \langle x, \hat{\theta}_k \rangle$, $\mathcal{A}_t^{(k)} \leftarrow \left\{ x \in \mathcal{A}_t^{(k-1)} \,\middle|\, \langle \hat{\theta}_k, x_t^{(k)} - x \rangle \leq 2\varepsilon_{t,k} \right\}$;
    if $t \leq \mathcal{T}_{\ell-1} + \lceil c^2 T^{1-2^{-\ell}} / \log_2 \log_2 T \rceil$ **then**
      Pull $x_{t,a_t} \sim \pi_{G'}(\mathcal{A}_t^{(\ell-1)})$;
    else if $t \leq \mathcal{T}_{\ell-1} + \lceil c^2 T^{1-2^{-\ell}} / \log_2 \log_2 T \rceil + \lceil c(1-c) T^{1-2^{-\ell}} / \log_2 \log_2 T \rceil$ **then**
      Pull $x_{t,a_t} \in \arg\max_{x \in \mathcal{A}_t^{(\ell-1)}} \|x\|_{H_{t-1}^{-1}}$;
    else
      Pull $x_{t,a_t} \in \arg\max_{x \in \mathcal{A}_t^{(\ell-1)}} \langle x, \hat{\theta}_{\ell-1} \rangle$;
    $H_t^{-1} \leftarrow H_{t-1}^{-1} - H_{t-1}^{-1} x_{t,a_t} x_{t,a_t}^\top H_{t-1}^{-1} / (1 + x_{t,a_t}^\top H_{t-1}^{-1} x_{t,a_t})$;
  $V_\ell^{-1} \leftarrow H_{\mathcal{T}_\ell}^{-1}$, $\hat{\theta}_\ell \leftarrow V_\ell^{-1} \sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell} r_t x_{t,a_t}$, $H_{\mathcal{T}_\ell} \leftarrow \lambda I$;

Discussion of Theorem 1. Theorem 1 shows that BLCE-G achieves minimax-optimal regret for fully sequential linear contextual bandits while performing only $\mathcal{O}(\log \log T)$ *parameter updates* under a static schedule. A notable feature of this bound is that it covers both regimes. In the small-K regime ($K \leq \mathcal{O}(e^d)$), the regret scales as $\widetilde{\mathcal{O}}(\sqrt{dT \log K})$, and in the large-$K$ regime ($K \geq \Omega(e^d)$), as $\widetilde{\mathcal{O}}(d\sqrt{T})$. Thus, BLCE-G provides the tightest known guarantees among linear contextual bandit algorithms with *rare parameter updates* and, to our knowledge, is the first algorithm to match the minimax lower bounds in both regimes.

Theorem 2 (Regret of BLCE)**.** *Consider running the BLCE algorithm for T rounds with K arms in d dimensions. The worst-case cumulative regret satisfies*

$$\mathcal{R}(T) = \mathcal{O}\left(\sqrt{dT} (\sqrt{\log(KT)} \wedge \sqrt{d + \log T}) \sqrt{\log T \log \log T} \right)$$

Algorithm 2 BLCE

Input: Horizon T; interval end times $\mathcal{T}_1 = \left\lceil \frac{\sqrt{T}}{\log_2 \log_2 T} \right\rceil$, $\mathcal{T}_\ell = \left(\mathcal{T}_{\ell-1} + \left\lceil \frac{T^{1-2^{-\ell}}}{\log_2 \log_2 T} \right\rceil + 1 \right) \wedge T$ for $\ell \geq 2$; number of intervals B, with $\mathcal{T}_B = T$; within-interval allocation rate c

Initialize: $\lambda \leftarrow 1$, $H_0 \leftarrow \lambda I$, $b_0 \leftarrow \mathbf{0}$;
for $t \leftarrow 1, 2, \ldots, \mathcal{T}_1$ **do**
  Pull arm $x_{t,a_t} \in \arg\max_{x \in \mathcal{A}_t} \|x\|_{H_{t-1}^{-1}}$;
  $H_t^{-1} \leftarrow H_{t-1}^{-1} - H_{t-1}^{-1} x_{t,a_t} x_{t,a_t}^\top H_{t-1}^{-1} / (1 + x_{t,a_t}^\top H_{t-1}^{-1} x_{t,a_t})$;
$V_1^{-1} \leftarrow H_{\mathcal{T}_1}^{-1}$, $\hat{\theta}_1 \leftarrow V_1^{-1} \sum_{t=1}^{\mathcal{T}_1} r_t x_{t,a_t}$, $H_{\mathcal{T}_1} \leftarrow \lambda I$;
for $\ell \leftarrow 2, \ldots, B$ **do**
  for $t \leftarrow \mathcal{T}_{\ell-1} + 1, \ldots, \mathcal{T}_\ell$ **do**
    for $k \leftarrow 1, \ldots, \ell - 1$ **do**
      $x_t^{(k)} \leftarrow \arg\max_{x \in \mathcal{A}_t^{(k-1)}} \langle x, \hat{\theta}_k \rangle$, $\mathcal{A}_t^{(k)} \leftarrow \left\{ x \in \mathcal{A}_t^{(k-1)} \,\middle|\, \langle \hat{\theta}_k, x_t^{(k)} - x \rangle \leq 2\varepsilon_{t,k} \right\}$;
    if $t \leq \mathcal{T}_{\ell-1} + \lceil cT^{1-2^{-\ell}} / \log_2 \log_2 T \rceil$ **then**
      Pull $x_{t,a_t} \in \arg\max_{x \in \mathcal{A}_t^{(\ell-1)}} \|x\|_{H_{t-1}^{-1}}$;
    else
      Pull $x_{t,a_t} \in \arg\max_{x \in \mathcal{A}_t^{(\ell-1)}} \langle x, \hat{\theta}_{\ell-1} \rangle$;
    $H_t^{-1} \leftarrow H_{t-1}^{-1} - H_{t-1}^{-1} x_{t,a_t} x_{t,a_t}^\top H_{t-1}^{-1} / (1 + x_{t,a_t}^\top H_{t-1}^{-1} x_{t,a_t})$;
  $V_\ell^{-1} \leftarrow H_{\mathcal{T}_\ell}^{-1}$, $\hat{\theta}_\ell \leftarrow V_\ell^{-1} \sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell} r_t x_{t,a_t}$, $H_{\mathcal{T}_\ell} \leftarrow \lambda I$;

$$= \widetilde{\mathcal{O}}\left(\sqrt{dT \log K} \wedge d\sqrt{T} \right).$$

Discussion of Theorem 2. Theorem 2 establishes that BLCE also achieves minimax-optimal regret with only $\mathcal{O}(\log \log T)$ parameter updates. Its key distinction lies in computation: by removing the G-optimal design step, BLCE significantly reduces both complexity and runtime while retaining minimax optimality. As shown in Table 1, this makes BLCE the first linear contextual bandit algorithm with *rare parameter updates* to combine minimax optimality across both regimes with *no reliance on G-optimal design*.

Remark 1. *While prior work on static-grid linear contextual bandits typically assumes i.i.d. contexts (Ruan et al., 2021; Hanna et al., 2023a;b; Zhang et al., 2025), we show that this assumption can be relaxed to the following batch-wise conditions (for any $\ell \geq 1$)*

(1) $\mathrm{Law}(\mathcal{A}_t | \mathcal{F}_{\mathcal{T}_{\ell-1}}) \sim \mathcal{D}_{\ell-1}, \forall t \in [\mathcal{T}_{\ell-1} + 1, \mathcal{T}_{\ell+1}]$
(2) $\mathcal{A}_t \perp\!\!\!\perp \{\mathcal{A}_s, x_{s,a_s}, r_s\}_{s=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell} | \mathcal{F}_{\mathcal{T}_{\ell-1}}, \forall t \in [\mathcal{T}_\ell + 1, \mathcal{T}_{\ell+1}]$
(3) $\mathcal{A}_s \perp\!\!\!\perp \{\mathcal{A}_u, x_{u,a_u}, r_u\}_{u=\mathcal{T}_{\ell-1}+1}^{s-1} | \mathcal{F}_{\mathcal{T}_{\ell-1}}, \forall s \in (\mathcal{T}_{\ell-1} + 1, \mathcal{T}_\ell]$

Given the history $\mathcal{F}_{\mathcal{T}_{\ell-1}}$, condition (1) requires inter-

vals ℓ and $\ell+1$ to share the same conditional context law; condition (2) enforces that contexts are conditionally independent of the contexts/actions/rewards realized in previous interval; and condition (3) imposes within-interval conditional independence of each context from earlier within-interval observations. These assumptions are strictly weaker than full i.i.d., requiring only (i) equality of the conditional context law across consecutive intervals and (ii) conditional independence across and within intervals. This relaxation offers greater modeling flexibility while preserving the guarantees established in Appendix B.

6.4. Time-Complexity of Algorithms

The main costs of BLCE-G are near G-optimal design and arm elimination. By Corollary 4, each near G-optimal design call costs $\mathcal{O}(Kd^3)$. In arm elimination, computing $\varepsilon_{t,k}$ costs $\mathcal{O}(Kd^2)$ per round; with at most $\mathcal{O}(\log\log T)$ rounds, this totals $\mathcal{O}(Kd^2T\log\log T)$. Hence, BLCE-G has overall complexity $\mathcal{O}(Kd^2T(d+\log\log T))$. For BLCE, only arm elimination is a bottleneck, giving $\mathcal{O}(Kd^2T\log\log T)$.

7. Extensions to Generalized Linear Contextual Bandits

In this section, we present our results extended *generalized* linear contextual bandits. The formal definition of the generalized linear contextual bandit with batched feedback is in Appendix C.

7.1. Proposed Algorithm: BGLE

We propose BGLE (*Batched-Feedback Generalized Linear Contextual Bandit with Elimination*), whose pseudocode is given in Algorithm 3. To extend our approach to the generalized linear setting, we build on the structure of Algorithm 2. In the first interval, the algorithm repeatedly pulls the most informative direction with respect to the current Gram matrix (Line 4) and updates its inverse via the Sherman–Morrison (Line 5). At the interval boundary, we set $V_1 \coloneqq H_{\mathcal{T}_1}$, compute the MLE $\hat{\theta}_1$ for the per–round log-loss $\ell_t(\theta) = m(\langle x_{t,a_t},\theta\rangle) - r_t\langle x_{t,a_t},\theta\rangle$, and reinitialize $H_{\mathcal{T}_1}$ for the next interval (Line 6). For each interval $\ell \geq 2$, the Gram matrix is *weighted* by $\alpha_{t,\ell-1}(\lambda)\,\dot{\mu}(\langle x_{t,a_t},\hat{\theta}_{\ell-1}\rangle)$, where

$$\alpha_{t,k}(\lambda) = \exp(-2RS)\,\mathbb{1}_{\{k=1\}} + \exp(-R(2S \wedge \|x_{t,a_t}\|_{V_k^{-1}}\beta(\lambda)))\,\mathbb{1}_{\{k\geq 2\}}$$

(Line 17). Beginning at interval $\ell \geq 3$, the algorithm performs $\ell-2$ elimination rounds using the estimates $\hat{\theta}_2,\ldots,\hat{\theta}_{\ell-1}$, yielding nested feasible sets $\mathcal{A}_t^{(2)},\ldots,\mathcal{A}_t^{(\ell-1)}$ (Line 11-12). Since no elimination is conducted with $\hat{\theta}_1$, we set $\mathcal{A}_t = \mathcal{A}_t^{(0)} = \mathcal{A}_t^{(1)}$. The elimination threshold $\varepsilon'_{t,k}(\lambda)$ for $k \in [\ell-1]\setminus\{1\}$ is defined as

$$\max_{y\in\mathcal{A}_t^{(k-1)}} \|y\|_{V_k^{-1}}\Big(24RS(\sqrt{d+\log T} + R(d+\log T)/\sqrt{\lambda}) + 2S\sqrt{\lambda}\Big)\,,$$

which, under the choice $\lambda = R^2(d+\log T)$, simplifies to $\max_{y\in\mathcal{A}_t^{(k-1)}}\|y\|_{V_k^{-1}}(50RS\sqrt{d+\log T})$. Within each interval $\ell \geq 2$, the action selection strategy follows that of BLCE, splitting the interval in the ratio $c:(1-c)$ between exploration and exploitation. The key difference is that arm selection is based on the *weighted* Gram matrix (Lines 14 and 16). At the end of interval ℓ, we set $V_\ell \coloneqq H_{\mathcal{T}_\ell}$, compute the MLE $\hat{\theta}_\ell$ for $\ell_t(\theta)$, and reinitialize $H_{\mathcal{T}_\ell}$ for the next interval (Line 18).

7.2. Worst-Case Regret of BGLE

To analyze BGLE, we introduce parameters that capture problem non-linearity. For any arm set $\mathcal{A}$, let $x^* \in \arg\max_{x\in\mathcal{A}}\mu(\langle x,\theta^*\rangle)$ denote an optimal arm, and define

$$\kappa \coloneqq \max_{\mathcal{A}\in\mathrm{supp}(\mathcal{D})}\max_{x\in\mathcal{A}}\frac{1}{\dot{\mu}(\langle x,\theta^*\rangle)},\ \hat{\kappa} \coloneqq \frac{1}{\mathbb{E}_{\mathcal{A}\sim\mathcal{D}}[\dot{\mu}(\langle x^*,\theta^*\rangle)]},$$
$$R_{\dot{\mu}} \coloneqq \max_{\mathcal{A}\in\mathrm{supp}(\mathcal{D})}\dot{\mu}(\langle x^*,\theta^*\rangle)$$

Here, κ captures the worst-case curvature (which can diverge in saturation), $\hat{\kappa}$ captures the average inverse curvature at the optimal arm, and $R_{\dot{\mu}}$ denotes the maximum derivative of the link function at optimal arms.

Theorem 3 (Regret of BGLE)**.** *Consider running the BGLE algorithm for T rounds with K arms in d dimensions. The worst-case cumulative regret satisfies*

$$\mathcal{R}(T) = \mathcal{O}\Big(RS\sqrt{d(d+\log T)T\log T\log\log T/\hat{\kappa}}\Big) \quad \text{(leading term)}$$
$$+\,\mathcal{O}\Big(\Big(R^2Se^{8RS}d(d+\log T)\log T\log\log T + \frac{R}{\log\log T}\Big)T^{\frac{1}{3}}\Big) \quad \text{(transient term)}$$
$$= \widetilde{\mathcal{O}}(RSd\sqrt{T}/\sqrt{\hat{\kappa}}) + \widetilde{\mathcal{O}}((R^2Se^{8RS}d^2+R)T^{1/3})\,.$$

Discussion of Theorem 3. Theorem 3 shows that both the leading and transient terms of BGLE are κ-*free*, in sharp contrast to (Sawarni et al., 2024), whose transient term depends on κ (see Table 3). Since $\dot{\mu}(z)\to 0$ in saturation, κ can grow arbitrarily large, so removing this dependence is a substantial improvement. Moreover, BGLE uses only $\mathcal{O}(\log\log T)$ parameter updates and is, to our knowledge, the first generalized linear contextual bandit algorithm with rare parameter updates to combine near-optimal guarantees with no reliance on G-optimal design. Finally, BGLE inherits substantially lower computational complexity by building on the BLCE framework.

Table 2. Average runtime (seconds) over 10 runs.

	Suboptimal algorithms		Optimal algorithms				
(K, d)	RS-OFUL	SoftBatch	BatchLinUCB-DG	Hanna et al. (2023b)	BatchLearning	BLCE-G	BLCE
$(1000, 5)$	0.85	4.18	290.87	Exponential	166.17	23.40	5.91
$(5000, 10)$	4.15	15.17	1300.01	Exponential	621.09	40.27	12.83
$(50, 20)$	0.42	3.74	1031.66	Exponential	45.85	2.26	1.06
$(100, 30)$	0.61	5.50	2987.07	Exponential	77.01	3.70	1.62

Remark 2. *The total time complexity of BGLE is $\mathcal{O}(Kd^2T \log\log T + \mathcal{C}_{\text{opt}}^{\text{tot}})$, where the first term comes from the BLCE, and the second from computing the MLE at interval boundaries (see Table 3).*

8. Numerical Experiments

We evaluate the performance of BLCE-G and BLCE over a horizon of $T = 10{,}000$ across 10 independent runs. At each round, K arms are sampled i.i.d. from a d-dimensional uniform distribution, and the parameter θ^* is drawn from a d-dimensional normal distribution. We consider four (K, d) pairs: $(1000, 5)$ and $(5000, 10)$, representing the large-K regime, and $(50, 20)$ and $(100, 30)$, representing the small-K regime. For comparison, we benchmark against state-of-the-art algorithms: RS-OFUL (Abbasi-Yadkori et al., 2011), BatchLinUCB-DG (Ruan et al., 2021), SoftBatch (Hanna et al., 2023b), and BatchLearning (Zhang et al., 2025). Hyperparameters are set consistently with theory, ensuring all choices satisfy the required conditions: BLCE-G and BLCE use within-interval allocation rate $c = 0.5$; RS-OFUL uses switching parameter $C = 1$; and SoftBatch employs discretization parameter $q = 1/(8\sqrt{d})$. Algorithms requiring G-optimal design are implemented using the same near G-optimal routine. Due to the substantial computational overhead reported in Table 1, we omit regret plots for the method of Hanna et al. (2023b).

We present three types of figures: (i) the average cumulative regret (solid line) with its standard deviation (shaded region) over 10 runs, (ii) zoomed-in views of regret curves to highlight differences between BLCE-G and BLCE, and (iii) the average parameter-update complexity across 10 runs, illustrating the frequency of parameter updates (for consistency with prior work, we label the horizontal axis using the term "batch"). As shown in Figure 1, both BLCE-G and BLCE consistently outperform all baselines in both large-K and small-K regimes, achieving the lowest regret with greater stability. Runtime comparisons in Table 2 further show that our methods incur substantially lower computational cost; in particular, BLCE, which eliminates G-optimal design entirely, achieves the fastest runtime among optimal algorithms, comparable even to suboptimal baselines. Overall, these results demonstrate that BLCE-G and BLCE combine minimax-optimal regret with practical efficiency.

For generalized linear contextual bandits, BGLE similarly outperforms the baseline, achieving lowest regret, stable performance, and reduced computation. See Appendix D.2 for details.

9. Conclusion

We studied contextual bandits under rare parameter updates, separating the cost of reward-dependent retraining from lightweight reward-free processing of contexts within an interval. For linear contextual bandits, BLCE-G achieves the tightest known rare-update regret guarantees across both small-K and large-K regimes, while BLCE removes the G-optimal design step and preserves minimax-optimality with substantially lower computational cost. For generalized linear contextual bandits, BGLE extends the same design principle and avoids explicit dependence on the worst-case curvature parameter in its regret bound. Together, these results suggest that limited reward adaptivity need not force computationally expensive pre-planned exploration, provided that within-interval context information can be used through reward-free updates.

Acknowledgments

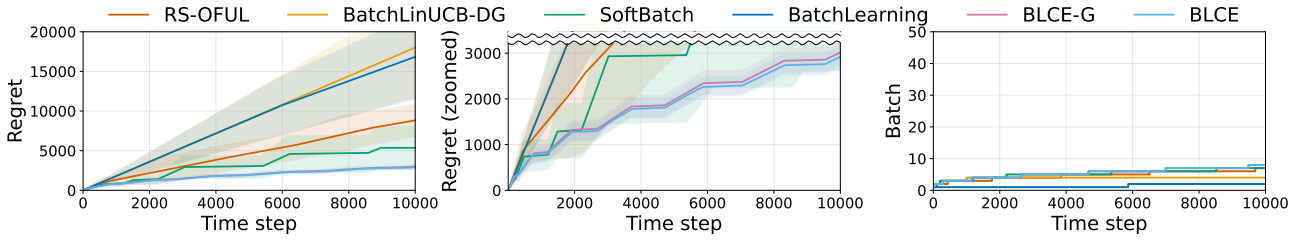
This work was supported by the National Research Foundation of Korea (NRF) grant and the Institute of Information & communications Technology Planning & Evaluation (IITP) grant both funded by the Korea government (MSIT) (No. RS-2022-NR071853, RS-2023-00222663, RS-2025-25418056, RS-2025-25463302, RS-2026-25507282).

Impact Statement

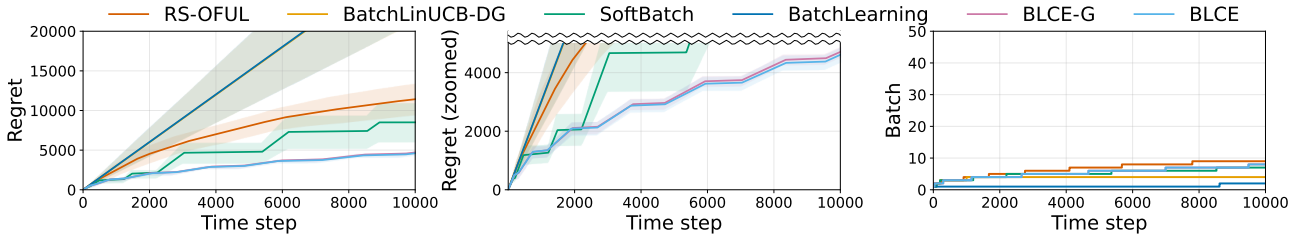
This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

References

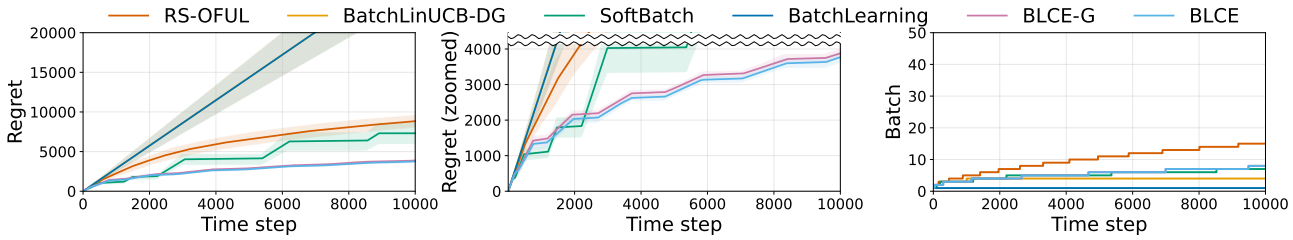
Abbasi-Yadkori, Y., Pál, D., and Szepesvári, C. Improved algorithms for linear stochastic bandits. *Advances in Neural Information Processing Systems*, 24:2312–2320, 2011.



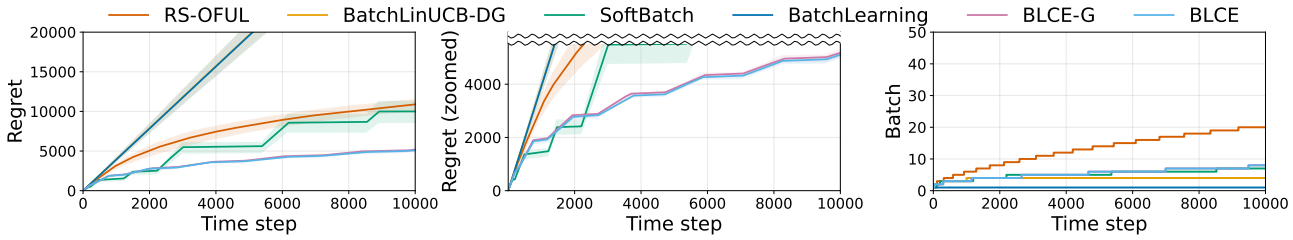
(a) $K = 1000, d = 5$



(b) $K = 5000, d = 10$



(c) $K = 50, d = 20$



(d) $K = 100, d = 30$

Figure 1. Regret, zoomed-in regret, and update (batch) complexity over time for different values of K and d.

Abe, N. and Long, P. M. Associative reinforcement learning using linear probabilistic concepts. In *International Conference on Machine Learning*, pp. 3–11, 1999.

Abe, N., Biermann, A. W., and Long, P. M. Reinforcement learning with immediate rewards and linear hypotheses. *Algorithmica*, 37:263–293, 2003.

Auer, P. Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research*, 3(Nov):397–422, 2002.

Chu, W., Li, L., Reyzin, L., and Schapire, R. Contextual bandits with linear payoff functions. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pp. 208–214. JMLR Workshop and Conference Proceedings, 2011.

Dani, V., Hayes, T. P., and Kakade, S. M. Stochastic linear optimization under bandit feedback. In *21st Annual Conference on Learning Theory*, number 101, pp. 355–366, 2008.

Esfandiari, H., Karbasi, A., Mehrabian, A., and Mirrokni, V. Regret bounds for batched bandits. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 7340–7348, 2021.

Gao, Z., Han, Y., Ren, Z., and Zhou, Z. Batched multi-armed bandits problem. *Advances in Neural Information Processing Systems*, 32, 2019.

Han, Y., Zhou, Z., Zhou, Z., Blanchet, J., Glynn, P. W., and Ye, Y. Sequential batch learning in finite-action linear

contextual bandits. *arXiv preprint arXiv:2004.06321*, 2020.

Hanna, O., Yang, L., and Fragouli, C. Efficient batched algorithm for contextual linear bandits with large action space via soft elimination. *Advances in Neural Information Processing Systems*, 36:56772–56783, 2023a.

Hanna, O. A., Yang, L., and Fragouli, C. Contexts can be cheap: Solving stochastic contextual bandits with linear bandit algorithms. In *The Thirty Sixth Annual Conference on Learning Theory*, pp. 1791–1821. PMLR, 2023b.

He, J., Zhang, J., and Zhang, R. Q. A reduction from linear contextual bandits lower bounds to estimations lower bounds. In *International Conference on Machine Learning*, pp. 8660–8677. PMLR, 2022.

Jin, R., Simchi-Levi, D., Wang, L., Wang, X., and Yang, S. Shrinking the upper confidence bound: A dynamic product selection problem for urban warehouses. *Management Science*, 67(8):4756–4771, 2021a.

Jin, T., Tang, J., Xu, P., Huang, K., Xiao, X., and Gu, Q. Almost optimal anytime algorithm for batched multi-armed bandits. In *International Conference on Machine Learning*, pp. 5065–5073. PMLR, 2021b.

Jin, T., Xu, P., Xiao, X., and Gu, Q. Double explore-then-commit: Asymptotic optimality and beyond. In *Conference on Learning Theory*, pp. 2584–2633. PMLR, 2021c.

Karbasi, A., Mirrokni, V., and Shadravan, M. Parallelizing thompson sampling. *Advances in Neural Information Processing Systems*, 34:10535–10548, 2021.

Kirschner, J., Lattimore, T., Vernade, C., and Szepesvári, C. Asymptotically optimal information-directed sampling. In *Conference on Learning Theory*, pp. 2777–2821. PMLR, 2021.

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

Li, L., Chu, W., Langford, J., and Schapire, R. E. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*, pp. 661–670, 2010.

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

Lu, Y., Xu, Z., and Tewari, A. Bandit algorithms for precision medicine. In *Handbook of Statistical Methods for Precision Medicine*, pp. 260–294. Chapman and Hall/CRC, 2021.

Perchet, V., Rigollet, P., Chassang, S., and Snowberg, E. Batched bandit problems. *The Annals of Statistics*, 44(2): 660–681, 2016.

Ren, X., Jin, T., and Xu, P. Optimal batched linear bandits. In *International Conference on Machine Learning*, pp. 42391–42416. PMLR, 2024.

Ruan, Y., Yang, J., and Zhou, Y. Linear bandits with limited adaptivity and learning distributional optimal design. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pp. 74–87, 2021.

Sawarni, A., Das, N., Barman, S., and Sinha, G. Generalized linear bandits with limited adaptivity. *Advances in Neural Information Processing Systems*, 37:8329–8369, 2024.

Todd, M. J. and Yıldırım, E. A. On khachiyan's algorithm for the computation of minimum-volume enclosing ellipsoids. *Discrete Applied Mathematics*, 155(13):1731–1744, 2007.

Yu, S. and Oh, M.-h. Optimal and practical batched linear bandit algorithm. In *Forty-Second International Conference on Machine Learning*, 2025.

Zhang, Z., Ji, X., and Zhou, Y. Almost optimal batch-regret tradeoff for batch linear contextual bandits. In *The Thirteenth International Conference on Learning Representations*, 2025.

Appendix

A. Proof of Theorem 1

Lemma 1. *(Yu & Oh, 2025) Let V_ℓ be the Gram matrix and $\hat{\theta}_\ell$ be the least squares estimator obtained from the contexts in the ℓ-th interval ($\ell \geq 1$). Then, for any $x \in \mathbb{R}^d$ and $0 < \delta < 1$, the following inequality holds with probability at least $1-\delta$*

$$|\langle x, \hat{\theta}_\ell - \theta^* \rangle| \leq \left(\sqrt{2 \log\left(\frac{1}{\delta}\right)} + \sqrt{\lambda} \right) \|x\|_{V_\ell^{-1}} .$$

Lemma 2. *(Yu & Oh, 2025) Let V_ℓ be the Gram matrix and $\hat{\theta}_\ell$ be the least squares estimator obtained from the contexts in the ℓ-th interval ($\ell \geq 1$). Then, for any $0 < \zeta, \delta < 1$ and $d \geq 2$, the following inequality holds with probability at least $1-\delta$*

$$\|\hat{\theta}_\ell - \theta^*\|_{V_\ell} \leq \frac{\sqrt{(d-1)\log(\frac{4}{4\zeta^2-\zeta^4}) + 2\log(\frac{\sqrt{2\pi d}}{\delta})} + \sqrt{\lambda}}{1-\zeta} .$$

Lemma 1 and Lemma 2 were originally proved in the linear bandit setting. However, the proofs do not rely on the non-contextual assumption that the feature vectors of arms remain fixed. Therefore, the results can be directly applied to the linear contextual bandit setting as well.

Corollary 1. *Let V_ℓ be the Gram matrix and $\hat{\theta}_\ell$ be the least squares estimator obtained from the contexts in the ℓ-th interval ($\ell \geq 1$). Then, for any $0 < \delta < 1$ and $d \geq 2$, the following inequality holds with probability at least $1-\delta$*

$$\|\hat{\theta}_\ell - \theta^*\|_{V_\ell} \leq 2\sqrt{\log\left(\frac{2^{6d-5}\pi d}{15^{d-1}\delta^2}\right)} + 2\sqrt{\lambda} .$$

Proof. Substituting $\zeta = 0.5$ into Lemma 2 yields the desired result. □

Lemma 3 (Good event)**.** *Define the following quantities:*

$$\beta_{t,\ell}^{(1)}(\delta) := \sqrt{2\log\left(\frac{2|\mathcal{A}_t^{(\ell-1)}|(B-1)T}{\delta}\right)} + \sqrt{\lambda},$$
$$\beta_{t,\ell}^{(2)}(\delta) := 2\sqrt{\log\left(\frac{2^{6d-3}\pi d(B-1)^2}{15^{d-1}\delta^2}\right)} + 2\sqrt{\lambda},$$
$$\varepsilon_{t,\ell}(\delta) := \max_{y \in \mathcal{A}_t^{(\ell-1)}} \|y\|_{V_\ell^{-1}} \cdot \left(\beta_{t,\ell}^{(1)}(\delta) \wedge \beta_{t,\ell}^{(2)}(\delta)\right) .$$

Here, δ is a constant in the interval $(0,1)$. Then, the following event E holds with probability at least $1-\delta$

$$E := \bigcap_{\ell=1}^{B-1} \bigcap_{t=\mathcal{T}_\ell+1}^{T} \left\{ |\langle x, \hat{\theta}_\ell - \theta^* \rangle| \leq \varepsilon_{t,\ell}(\delta),\ \forall x \in \mathcal{A}_t^{(\ell-1)} \right\} .$$

Proof. Fix arbitrary ℓ and t. By Lemma 1, the following inequality holds for all $x \in \mathcal{A}_t^{(\ell-1)}$ with probability at least $1 - \frac{\delta}{2(B-1)T}$

$$|\langle x, \hat{\theta}_\ell - \theta^* \rangle| \leq \max_{y \in \mathcal{A}_t^{(\ell-1)}} \|y\|_{V_\ell^{-1}} \cdot \left(\sqrt{2\log\left(\frac{2|\mathcal{A}_t^{(\ell-1)}|(B-1)T}{\delta}\right)} + \sqrt{\lambda} \right) .$$

Applying a union bound over all ℓ and t, the following event holds with probability at least $1 - \frac{\delta}{2}$

$$\bigcap_{\ell=1}^{B-1} \bigcap_{t=\mathcal{T}_\ell+1}^{T} \left\{ |\langle x, \hat{\theta}_\ell - \theta^* \rangle| \leq \max_{y \in \mathcal{A}_t^{(\ell-1)}} \|y\|_{V_\ell^{-1}} \cdot \beta_{t,\ell}^{(1)}(\delta),\ \forall x \in \mathcal{A}_t^{(\ell-1)} \right\} . \tag{1}$$

Next, for fixed ℓ, by the Cauchy-Schwarz inequality and Corollary 1, the following inequality holds for all t and $x \in \mathcal{A}_t^{(\ell-1)}$ with probability at least $1 - \frac{\delta}{2(B-1)}$

$$|\langle x, \hat{\theta}_\ell - \theta^* \rangle| \le \|x\|_{V_\ell^{-1}} \|\hat{\theta}_\ell - \theta^*\|_{V_\ell} \le \max_{y \in \mathcal{A}_t^{(\ell-1)}} \|y\|_{V_\ell^{-1}} \cdot \left(2\sqrt{\log\left(\frac{2^{6d-3}\pi d(B-1)^2}{15^{d-1}\delta^2}\right)} + 2\sqrt{\lambda} \right) .$$

A union bound over ℓ yields the following event with probability at least $1 - \frac{\delta}{2}$

$$\bigcap_{\ell=1}^{B-1} \bigcap_{t=\mathcal{T}_\ell+1}^{T} \left\{ |\langle x, \hat{\theta}_\ell - \theta^* \rangle| \le \max_{y \in \mathcal{A}_t^{(\ell-1)}} \|y\|_{V_\ell^{-1}} \cdot \beta_{t,\ell}^{(2)}(\delta),\ \forall x \in \mathcal{A}_t^{(\ell-1)} \right\} . \tag{2}$$

Combining event (1) and event (2), we conclude that the following event holds with probability at least $1 - \delta$

$$\bigcap_{\ell=1}^{B-1} \bigcap_{t=\mathcal{T}_\ell+1}^{T} \left\{ |\langle x, \hat{\theta}_\ell - \theta^* \rangle| \le \max_{y \in \mathcal{A}_t^{(\ell-1)}} \|y\|_{V_\ell^{-1}} \cdot \left(\beta_{t,\ell}^{(1)}(\delta) \wedge \beta_{t,\ell}^{(2)}(\delta) \right),\ \forall x \in \mathcal{A}_t^{(\ell-1)} \right\} .$$

□

Lemma 4. *Let E be the good event defined in Lemma 3 with $\delta = \frac{2}{T}$. Conditioned on E, the optimal arm $x_t^* \in \arg\max_{x \in \mathcal{A}_t} \langle x, \theta^* \rangle$ is never eliminated at any round t. In particular,*

$$x_t^* \in \mathcal{A}_t^{(\ell)}, \quad \textit{for all } 1 \le \ell \le B-1 \textit{ and } \mathcal{T}_\ell + 1 \le t \le T .$$

Proof. Fix $t \in [\mathcal{T}_s + 1, \mathcal{T}_{s+1}]$ for some $s \in [B-1]$. We show by induction on ℓ that $x_t^* \in \mathcal{A}_t^{(\ell)}$ for all $\ell \in [s]$.

Base case ($\ell = 1$). Since both x_t^* and $x_t^{(1)}$ belong to $\mathcal{A}_t^{(0)} (= \mathcal{A}_t)$, we have

$$\begin{aligned}
\langle \hat{\theta}_1, x_t^{(1)} - x_t^* \rangle &= \langle \hat{\theta}_1 - \theta^*, x_t^{(1)} - x_t^* \rangle + \langle \theta^*, x_t^{(1)} - x_t^* \rangle \\
&\le \langle \hat{\theta}_1 - \theta^*, x_t^{(1)} - x_t^* \rangle \\
&\le |\langle \hat{\theta}_1 - \theta^*, x_t^{(1)} \rangle| + |\langle \hat{\theta}_1 - \theta^*, x_t^* \rangle| \\
&\le 2\varepsilon_{t,1} ,
\end{aligned}$$

where the first inequality follows from the optimality of x_t^* and the last from the definition of the good event E. Hence $x_t^* \in \mathcal{A}_t^{(1)}$.

Inductive step. Assume $x_t^* \in \mathcal{A}_t^{(\ell-1)}$ for some $\ell \in \{2, \dots, s\}$. Since $x_t^{(\ell)} \in \mathcal{A}_t^{(\ell-1)}$, we similarly obtain

$$\begin{aligned}
\langle \hat{\theta}_\ell, x_t^{(\ell)} - x_t^* \rangle &= \langle \hat{\theta}_\ell - \theta^*, x_t^{(\ell)} - x_t^* \rangle + \langle \theta^*, x_t^{(\ell)} - x_t^* \rangle \\
&\le \langle \hat{\theta}_\ell - \theta^*, x_t^{(\ell)} - x_t^* \rangle \\
&\le |\langle \hat{\theta}_\ell - \theta^*, x_t^{(\ell)} \rangle| + |\langle \hat{\theta}_\ell - \theta^*, x_t^* \rangle| \\
&\le 2\varepsilon_{t,\ell} ,
\end{aligned}$$

which shows that $x_t^* \in \mathcal{A}_t^{(\ell)}$. By induction, the claim holds for all $\ell \in [s]$, completing the proof. □

Lemma 5. *(Abbasi-Yadkori et al., 2011) Let $\{x_1, \dots, x_n\} \subset \mathbb{R}^d$ be a sequence of vectors such that $\|x_i\|_2 \le 1$ for all $i \in [n]$. Let $H_0 \in \mathbb{R}^{d \times d}$ be a positive definite matrix, and define $H_t := H_0 + \sum_{i=1}^t x_i x_i^\top$. Then, the following inequality holds*

$$\sum_{t=1}^{n} \min\left\{1, \|x_t\|_{H_{t-1}^{-1}}^2\right\} \le 2 \log\left(\frac{\det(H_n)}{\det(H_0)}\right) .$$

Corollary 2. *Let $\{x_1, \dots, x_n\} \subset \mathbb{R}^d$ be a sequence of vectors such that $\|x_i\|_2 \le 1$ for all $i \in [n]$. Suppose $\lambda \ge 1$, and define $H_0 := \lambda I$ and $H_t := \lambda I + \sum_{i=1}^{t} x_i x_i^\top$ for each $t \in \{1, \dots, n\}$. Then, for any $1 \le m \le n$,*

$$\sum_{t=m}^{n} \|x_t\|^2_{H_{t-1}^{-1}} \le 2 \log \left(\frac{\det(H_n)}{\det(H_{m-1})} \right) .$$

Proof. Since $H_{t-1}^{-1} \preceq \lambda^{-1} I$ for all t, it follows that $H_{t-1}^{-1} \preceq \lambda^{-1} I \preceq I$. Consequently, $\|x_t\|^2_{H_{t-1}^{-1}} \le \|x_t\|_2^2 \le 1$ for each $t \in [m, n]$. Applying Lemma 5 over the interval $[m, n]$ yields the stated bound. □

Lemma 6. *Let H be a positive definite matrix. Suppose $x \in \mathbb{R}^d$ satisfies $\|x\|^2_{H^{-1}} \le c$. Then $cH \succeq xx^\top$.*

Proof. For any $z \in \mathbb{R}^d$, we have

$$z^\top (cH) z \ge \|x\|^2_{H^{-1}} \cdot \|z\|^2_H = \|H^{-\frac{1}{2}} x\|^2 \cdot \|H^{\frac{1}{2}} z\|^2 \ge \langle x, z \rangle^2 = z^\top (xx^\top) z .$$

The first inequality follows from the assumption $\|x\|^2_{H^{-1}} \le c$. The second inequality follows from the Cauchy–Schwarz inequality. Since this bound holds for all $z \in \mathbb{R}^d$, the matrix inequality $cH \succeq xx^\top$ follows. □

Lemma 7 (Pseudoinverse variance-to-matrix)**.** *Let $M \succeq 0$ and suppose $x \in \mathrm{range}(M)$. If $x^\top M^+ x \le c$, then*

$$xx^\top \preceq cM.$$

Proof. For any $z \in \mathbb{R}^d$,

$$(z^\top x)^2 = \left((M^{1/2} z)^\top M^{+/2} x \right)^2 \le (z^\top M z)(x^\top M^+ x) \le c\, z^\top M z.$$

Since this holds for every $z \in \mathbb{R}^d$, we obtain $xx^\top \preceq cM$. □

Lemma 8. *(Ruan et al., 2021) Let $x_1, \dots, x_n$ be independent and identically distributed (i.i.d.) random vectors drawn from a distribution $\mathcal{D}$ such that $\|x_i\|_2 \le 1$ almost surely. For any cutoff level $\lambda > 0$, the following inequality holds with probability at least $1 - 2d \exp(-\frac{\lambda n}{8})$*

$$3\lambda I + \frac{1}{n} \sum_{i=1}^{n} x_i x_i^\top \succeq \frac{1}{8} \mathbb{E}_{x \sim \mathcal{D}} \left[xx^\top \right]$$

Corollary 3. *Let $x_1, \dots, x_n$ be independent and identically distributed (i.i.d.) random vectors drawn from a distribution $\mathcal{D}$ such that $\|x_i\|_2 \le 1$ almost surely. Then, the following inequality holds with probability at least $1 - \frac{2}{T}$*

$$\frac{24 \log(dT)}{n} I + \frac{1}{n} \sum_{i=1}^{n} x_i x_i^\top \succeq \frac{1}{8} \mathbb{E}_{x \sim \mathcal{D}} \left[xx^\top \right]$$

Proof. The proof is a direct application of Lemma 8. We achieve the desired inequality by setting the cutoff level λ to $\lambda = \frac{8 \log(dT)}{n}$. □

Lemma 9. *(Todd & Yıldırım, 2007) Let $X = \{x_1, \dots, x_K\} \subset \mathbb{R}^d$ be a set of K points that spans $\mathbb{R}^d$, and fix $\varepsilon \in (0, 1]$. Then Khachiyan's barycentric coordinate descent algorithm computes a $(1 + \varepsilon)$-approximation to the minimum-volume enclosing ellipsoid of X in*

$$O\Big(Kd^2 \left([(1 + \varepsilon)^{2/(d+1)} - 1]^{-1} + \log d \right) \Big)$$

arithmetic operations. In particular, since $[(1 + \varepsilon)^{2/(d+1)} - 1]^{-1} = \Theta(d/\varepsilon)$ for $\varepsilon \in (0, 1]$, the total time complexity simplifies to

$$O\Big(\tfrac{Kd^3}{\varepsilon} \Big) .$$

Corollary 4 (Near G-optimal design)**.** *Let $X = \{x_1, \dots, x_K\} \subset \mathbb{R}^d$ and let $r = \operatorname{rank}(X) \le d$. Let $U \in \mathbb{R}^{d\times r}$ have orthonormal columns spanning $\operatorname{span}\{x_i\}$, and define $x'_i := U^\top x_i \in \mathbb{R}^r$ together with $\Sigma'(w) := \sum_{i=1}^K w_i\, x'_i x_i'^\top$ for $w \in \Delta_K$ (K-dimensional probability simplex). Applying the algorithm of Lemma 9 to $\{x'_i\}_{i=1}^K$ with accuracy parameter fixed to $\varepsilon = 1$ returns weights $w_\circ \in \Delta_K$ such that*

$$\max_{j\in[K]} x_j'^\top \Big(\Sigma'(w_\circ)\Big)^{-1} x'_j \;\le\; 2\,r\;,$$

in

$$O\Big(Kr^2\big([(1+1)^{2/(r+1)} - 1]^{-1} + \log r\big)\Big) \;=\; O(Kr^3)$$

arithmetic operations. Equivalently, in the original space we have

$$\max_{j\in[K]} x_j^\top \big(\Pi\,\Sigma(w_\circ)^+\Pi\big)\, x_j \;\le\; 2\,r\;, \qquad \textit{where}\;\; \Sigma(w) := \sum_{i=1}^K w_i\, x_i x_i^\top,\; \Pi := UU^\top\;,$$

with $^+$ denoting the Moore–Penrose pseudoinverse. Hence $w_\circ$ is a constant-factor near-optimal design in the effective dimension r, computed in polynomial time $O(Kr^3)$ by the method in Lemma 9.

Proof. Consider the subspace spanned by $\{x_i\}$ of dimension r, let U be an orthonormal basis, and project $x'_i := U^\top x_i \in \mathbb{R}^r$. The MVEE and D/G-optimal design problems are equivalently posed in $\mathbb{R}^r$, and guarantees for x'_i and $\Sigma'(w) := \sum_i w_i x'_i x_i'^\top$ transfer back to the original space through the projector $\Pi = UU^\top$ and the Moore–Penrose pseudoinverse. By Lemma 9, Khachiyan's barycentric coordinate descent computes an ε-approximate MVEE of $\{x'_i\}$ in

$$O\Big(Kr^2\big([(1+\varepsilon)^{2/(r+1)} - 1]^{-1} + \log r\big)\Big)$$

arithmetic operations; fixing $\varepsilon = 1$ gives $O(Kr^3)$.

Moreover, the standard MVEE $\leftrightarrow$ D/G-optimal duality and the algorithm's stopping rule imply that the returned weights $w_\circ \in \Delta_K$ satisfy the *constraint-violation guarantee*

$$\max_{j\in[K]} x_j'^\top \Big(\Sigma'(w_\circ)\Big)^{-1} x'_j \;\le\; (1+\varepsilon)\,r\;.$$

With $\varepsilon = 1$ this yields

$$\max_{j\in[K]} x_j'^\top \Big(\Sigma'(w_\circ)\Big)^{-1} x'_j \;\le\; 2\,r\;.$$

Finally, lifting back to the original space uses the identity

$$x_j'^\top \Big(\Sigma'(w_\circ)\Big)^{-1} x'_j \;=\; x_j^\top \big(\Pi\,\Sigma(w_\circ)^+\Pi\big)\, x_j\;, \qquad \Sigma(w) := \sum_{i=1}^K w_i\, x_i x_i^\top,\; \Pi := UU^\top,$$

so that

$$\max_{j\in[K]} x_j^\top \big(\Pi\,\Sigma(w_\circ)^+\Pi\big)\, x_j \;\le\; 2\,r\;.$$

Hence $w_\circ$ is a constant-factor near-G-optimal design in effective dimension r, computed in polynomial time $O(Kr^3)$ by the method in Lemma 9. □

Lemma 10 (Design transfer for the G-design prefix of `BLCE-G`)**.** *Fix an interval $j \in [B]$. Let*

$$n_{j,1} \coloneqq \mathcal{T}'_{j-1} - \mathcal{T}_{j-1}, \qquad n_{j,2} \coloneqq \mathcal{T}''_{j-1} - \mathcal{T}'_{j-1}.$$

For a fresh context draw in interval j, let $\mathcal{A}^{(j-1)}$ denote the post-elimination feasible set obtained by applying the first $j-1$ elimination steps using the estimates available at the beginning of interval j. For $j = 1$, set $\mathcal{A}^{(0)} = \mathcal{A}$.

Define the conditional mean G-design covariance

$$\bar{\Sigma}_j \coloneqq \mathbb{E}\big[\mathbb{E}_{z\sim\pi_{G'}(\mathcal{A}^{(j-1)})}[zz^\top] \,\big|\, \mathcal{F}_{\mathcal{T}_{j-1}}\big]\,.$$

Let

$$G_j \coloneqq \left\{ \bar{\Sigma}_j \preceq \frac{192}{n_{j,1}} H_{\mathcal{T}'_{j-1}} \right\}.$$

Then

$$\mathbb{P}(G_j \mid \mathcal{F}_{\mathcal{T}_{j-1}}) \geq 1 - \frac{2}{T}.$$

Moreover, on G_j,

$$\mathbb{E}\left[2 \log \frac{\det(H_{\mathcal{T}''_{j-1}})}{\det(H_{\mathcal{T}'_{j-1}})} \,\middle|\, \mathcal{F}_{\mathcal{T}'_{j-1}} \right] \leq 2d \log\left(1 + \frac{384 d\, n_{j,2}}{n_{j,1}} \right).$$

Proof. Condition on $\mathcal{F}_{\mathcal{T}_{j-1}}$. During the G-design prefix of interval j, the selected arms

$$g_s \coloneqq x_{s,a_s}, \qquad s = \mathcal{T}_{j-1} + 1, \ldots, \mathcal{T}'_{j-1},$$

are i.i.d. draws from the mixture distribution induced by first drawing a fresh context set and then sampling from $\pi_{G'}(\mathcal{A}_s^{(j-1)})$. The second moment of this mixture distribution is exactly $\bar{\Sigma}_j$. Therefore, by Corollary 3, with conditional probability at least $1 - 2/T$,

$$\frac{24 \log(dT)}{n_{j,1}} I + \frac{1}{n_{j,1}} \sum_{s=\mathcal{T}_{j-1}+1}^{\mathcal{T}'_{j-1}} g_s g_s^\top \succeq \frac{1}{8} \bar{\Sigma}_j.$$

Equivalently,

$$\bar{\Sigma}_j \preceq \frac{8}{n_{j,1}} \left(24 \log(dT) I + \sum_{s=\mathcal{T}_{j-1}+1}^{\mathcal{T}'_{j-1}} g_s g_s^\top \right).$$

Since $\lambda = \log(dT)$ and

$$H_{\mathcal{T}'_{j-1}} = \lambda I + \sum_{s=\mathcal{T}_{j-1}+1}^{\mathcal{T}'_{j-1}} g_s g_s^\top,$$

we have

$$24 \log(dT) I + \sum_{s=\mathcal{T}_{j-1}+1}^{\mathcal{T}'_{j-1}} g_s g_s^\top \preceq 24 H_{\mathcal{T}'_{j-1}}.$$

Thus

$$\bar{\Sigma}_j \preceq \frac{192}{n_{j,1}} H_{\mathcal{T}'_{j-1}},$$

which proves the first claim.

Now consider Phase 2 of interval j. For

$$s = \mathcal{T}'_{j-1} + 1, \ldots, \mathcal{T}''_{j-1},$$

write

$$\Sigma_s \coloneqq \mathbb{E}_{z \sim \pi_{G'}(\mathcal{A}_s^{(j-1)})}[z z^\top].$$

Since $x_{s,a_s} \in \mathcal{A}_s^{(j-1)}$, the near G-optimal design guarantee and Lemma 7 imply

$$x_{s,a_s} x_{s,a_s}^\top \preceq 2d\, \Sigma_s.$$

Let $H \coloneqq H_{\mathcal{T}'_{j-1}}$. Then

$$2 \log \frac{\det(H_{\mathcal{T}''_{j-1}})}{\det(H)} = 2 \log \det \left(I + H^{-1/2} \left(\sum_{s=\mathcal{T}'_{j-1}+1}^{\mathcal{T}''_{j-1}} x_{s,a_s} x_{s,a_s}^\top \right) H^{-1/2} \right)$$

$$\leq 2\log\det\left(I + 2d\sum_{s=\mathcal{T}'_{j-1}+1}^{\mathcal{T}''_{j-1}} H^{-1/2}\Sigma_s H^{-1/2}\right).$$

Conditional on $\mathcal{F}_{\mathcal{T}'_{j-1}}$, the Phase 2 context sets are fresh i.i.d. draws, and hence

$$\mathbb{E}[\Sigma_s \mid \mathcal{F}_{\mathcal{T}'_{j-1}}] = \bar{\Sigma}_j.$$

Using the concavity of $\log\det(\cdot)$ on the positive definite cone, we obtain

$$\mathbb{E}\left[2\log\frac{\det(H_{\mathcal{T}''_{j-1}})}{\det(H)}\,\middle|\,\mathcal{F}_{\mathcal{T}'_{j-1}}\right] \leq 2\log\det\Big(I + 2d\,n_{j,2}\,H^{-1/2}\bar{\Sigma}_j H^{-1/2}\Big).$$

On the event G_j,

$$H^{-1/2}\bar{\Sigma}_j H^{-1/2} \preceq \frac{192}{n_{j,1}} I.$$

Therefore,

$$\mathbb{E}\left[2\log\frac{\det(H_{\mathcal{T}''_{j-1}})}{\det(H)}\,\middle|\,\mathcal{F}_{\mathcal{T}'_{j-1}}\right] \leq 2d\log\left(1 + \frac{384d\,n_{j,2}}{n_{j,1}}\right).$$

This completes the proof. □

Proof of Theorem 1. To establish the claimed interval complexity bound, we analyze the lower bound on the length of the ℓ-th interval, where $\ell = \lceil\log_2\log_2 T\rceil$. By definition of the schedule, the length of this interval is at least

$$\left\lceil\frac{T^{1-2^{-\ell}}}{\log_2\log_2 T}\right\rceil + 2 \geq \frac{T^{1-2^{-\lceil\log_2\log_2 T\rceil}}}{\log_2\log_2 T} \geq \frac{T^{1-2^{-\log_2\log_2 T}}}{\log_2\log_2 T} = \frac{T^{1-\frac{1}{\log_2 T}}}{\log_2\log_2 T} = \frac{T}{2\log_2\log_2 T}\,.$$

Since the interval length is non-decreasing in ℓ, every interval with index $\ell \geq \lceil\log_2\log_2 T\rceil$ has length at least $\frac{T}{2\log_2\log_2 T}$. Given the total time horizon T, the number of such intervals is therefore at most $\lceil 2\log_2\log_2 T\rceil$. Including the initial $\lceil\log_2\log_2 T\rceil - 1$ intervals, the total number of intervals B is bounded by

$$B \leq \lceil 2\log_2\log_2 T\rceil + \lceil\log_2\log_2 T\rceil - 1 \leq 3\log_2\log_2 T + 1\,,$$

which implies that the interval complexity is $\mathcal{O}(\log\log T)$.

Next, we begin by decomposing the cumulative expected regret based on the good event E. The regret can be written as

$$\begin{aligned}\mathcal{R}(T) &= \sum_{k=1}^{B}\sum_{t=\mathcal{T}_{k-1}+1}^{\mathcal{T}_k}\mathbb{E}\left[\max_{x\in\mathcal{A}_t}\langle x,\theta^*\rangle - \langle x_{t,a_t},\theta^*\rangle\right]\\ &= \sum_{k=1}^{B}\sum_{t=\mathcal{T}_{k-1}+1}^{\mathcal{T}_k}\mathbb{E}\left[\langle x_t^* - x_{t,a_t},\theta^*\rangle\,\middle|\,E^c\right]\cdot\mathbb{P}(E^c) + \mathbb{E}\left[\langle x_t^* - x_{t,a_t},\theta^*\rangle\,\middle|\,E\right]\cdot\mathbb{P}(E)\,.\end{aligned}$$

Using the triangle inequality followed by the Cauchy-Schwarz inequality, we have

$$\begin{aligned}\langle x_t^* - x_{t,a_t},\theta^*\rangle &\leq |\langle x_t^* - x_{t,a_t},\theta^*\rangle|\\ &\leq |\langle x_t^*,\theta^*\rangle| + |\langle x_{t,a_t},\theta^*\rangle|\\ &\leq \|x_t^*\|\cdot\|\theta^*\| + \|x_{t,a_t}\|\cdot\|\theta^*\|\\ &\leq 2\,,\end{aligned}$$

where the last inequality follows from Assumption 1. Hence, the cumulative expected regret can be bounded as

$$\mathcal{R}(T) \le 2T \cdot \mathbb{P}(E^c) + \sum_{k=1}^{B} \sum_{t=\mathcal{T}_{k-1}+1}^{\mathcal{T}_k} \mathbb{E}\left[\langle x_t^* - x_{t,a_t}, \theta^* \rangle \middle| E\right] \cdot \mathbb{P}(E)$$
$$\le \mathcal{O}(1) + \sum_{k=1}^{B} \sum_{t=\mathcal{T}_{k-1}+1}^{\mathcal{T}_k} \mathbb{E}\left[\langle x_t^* - x_{t,a_t}, \theta^* \rangle \middle| E\right],$$

where the $\mathcal{O}(1)$ term follows from the high-probability guarantee $\mathbb{P}(E^c) = \mathcal{O}(1/T)$. Throughout the remainder of the analysis, we therefore condition on the good event E. Let Regret_ℓ denote the cumulative expected regret incurred during interval ℓ. We analyze the regret separately for the case $\ell = 1$ and for all subsequent intervals $\ell \ge 2$.

Case 1: $\ell = 1$**.** In the first interval, the number of rounds is $\mathcal{T}_1 = \left\lceil \frac{\sqrt{T}}{\log_2 \log_2 T} \right\rceil + 1$, and the instantaneous regret is bounded by 2. Therefore,

$$\text{Regret}_1 = \sum_{t=1}^{\mathcal{T}_1} \mathbb{E}[\langle x_t^* - x_{t,a_t}, \theta^* \rangle] \le 2\mathcal{T}_1 \le 2\left(\frac{\sqrt{T}}{\log_2 \log_2 T} + 2\right) = \mathcal{O}\big(\sqrt{T}\big) .$$

Case 2: $\ell \ge 2$**.** For each interval $\ell \ge 2$, the rounds following the arm elimination steps are divided into three phases:

Phases 1 & 2. In the first two phases of length $\left\lceil \frac{c^2 T^{1-2^{-\ell}}}{\log_2 \log_2 T} \right\rceil + \left\lceil \frac{c(1-c)T^{1-2^{-\ell}}}{\log_2 \log_2 T} \right\rceil$, the algorithm selects arm according to a near G-optimal design and the most informative direction with respect to the current Gram matrix. For any round t in these phases, the instantaneous regret satisfies

$$\begin{aligned}
\langle x_t^* - x_{t,a_t}, \theta^* \rangle &= \langle x_t^* - x_{t,a_t}, \theta^* - \hat{\theta}_{\ell-1} \rangle + \langle x_t^* - x_{t,a_t}, \hat{\theta}_{\ell-1} \rangle \\
&\le \langle x_t^* - x_{t,a_t}, \theta^* - \hat{\theta}_{\ell-1} \rangle + \langle x_t^{(\ell-1)} - x_{t,a_t}, \hat{\theta}_{\ell-1} \rangle \\
&\le |\langle x_t^*, \hat{\theta}_{\ell-1} - \theta^* \rangle| + |\langle x_{t,a_t}, \hat{\theta}_{\ell-1} - \theta^* \rangle| + 2\varepsilon_{t,\ell-1} \\
&\le 4\varepsilon_{t,\ell-1} . \qquad (3)
\end{aligned}$$

The first inequality uses the fact that $x_t^{(\ell-1)}$ is optimal with respect to $\hat{\theta}_{\ell-1}$ and that x_t^* belongs to $\mathcal{A}_t^{(\ell-2)}$ by Lemma 4. The second inequality follows since $x_{t,a_t} \in \mathcal{A}_t^{(\ell-1)}$ by construction of the arm elimination step. The final inequality follows from the definition of the good event E and again from Lemma 4, which guarantees $x_t^* \in \mathcal{A}_t^{(\ell-2)}$.

Phase 3. In the last phase, the algorithm selects arms greedily with respect to the estimated parameter, i.e., $x_{t,a_t} \in \arg\max_{x \in \mathcal{A}_t^{(\ell-1)}} \langle x, \hat{\theta}_{\ell-1} \rangle$. For any round t in this phase, the instantaneous regret satisfies

$$\begin{aligned}
\langle x_t^* - x_{t,a_t}, \theta^* \rangle &= \langle x_t^* - x_{t,a_t}, \theta^* - \hat{\theta}_{\ell-1} \rangle + \langle x_t^* - x_{t,a_t}, \hat{\theta}_{\ell-1} \rangle \\
&\le \langle x_t^* - x_{t,a_t}, \theta^* - \hat{\theta}_{\ell-1} \rangle \\
&\le |\langle x_t^*, \hat{\theta}_{\ell-1} - \theta^* \rangle| + |\langle x_{t,a_t}, \hat{\theta}_{\ell-1} - \theta^* \rangle| \\
&\le 2\varepsilon_{t,\ell-1} . \qquad (4)
\end{aligned}$$

The first inequality follows from the fact that x_{t,a_t} is optimal with respect to $\hat{\theta}_{\ell-1}$ and that x_t^* belongs to $\mathcal{A}_t^{(\ell-1)}$ by Lemma 4. The final inequality follows directly from the definition of the good event E and from Lemma 4, as in Phases 1 & 2.

Therefore, it suffices to upper bound the quantity $\sum_{t=\mathcal{T}_\ell+1}^{\mathcal{T}_{\ell+1}} \varepsilon_{t,\ell}$ for each $\ell \in [B-1]$, which reduces to bounding the sum $\sum_{t=\mathcal{T}_\ell+1}^{\mathcal{T}_{\ell+1}} \max_{y \in \mathcal{A}_t^{(\ell-1)}} \|y\|_{V_\ell^{-1}}$. Since both V_ℓ and the arm elimination rule—determined by $\hat{\theta}_1, \ldots, \hat{\theta}_{\ell-1}$—are measurable with respect to $\mathcal{F}_{\mathcal{T}_\ell}$, they can be treated as fixed quantities conditional on this filtration. Given that the contexts are drawn independently and identically, it follows by the tower property that for any $t, v \in [\mathcal{T}_\ell + 1, \mathcal{T}_{\ell+1}]$, we have

$$\mathbb{E}\left[\max_{y \in \mathcal{A}_t^{(\ell-1)}} y^\top V_\ell^{-1} y\right] = \mathbb{E}\left[\mathbb{E}\left[\max_{y \in \mathcal{A}_t^{(\ell-1)}} y^\top V_\ell^{-1} y \middle| \mathcal{F}_{\mathcal{T}_\ell}\right]\right]$$

$$= \mathbb{E}\left[\mathbb{E}\left[\max_{y\in\mathcal{A}_v^{(\ell-1)}} y^\top V_\ell^{-1} y \,\middle|\, \mathcal{F}_{\mathcal{T}_\ell}\right]\right]$$
$$= \mathbb{E}\left[\max_{y\in\mathcal{A}_v^{(\ell-1)}} y^\top V_\ell^{-1} y\right]. \tag{5}$$

Define $\mathcal{T}'_{\ell-1} \coloneqq \mathcal{T}_{\ell-1} + \left\lceil \frac{c^2 T^{1-2^{-\ell}}}{\log_2 \log_2 T} \right\rceil$ and $\mathcal{T}''_{\ell-1} \coloneqq \mathcal{T}'_{\ell-1} + \left\lceil \frac{c(1-c)T^{1-2^{-\ell}}}{\log_2 \log_2 T} \right\rceil$ for all $\ell \geq 2$, whereas for the first interval ($\ell = 1$), we set $\mathcal{T}'_0 \coloneqq \mathcal{T}_0 + \left\lceil \frac{c\sqrt{T}}{\log_2 \log_2 T} \right\rceil$ and $\mathcal{T}''_0 \coloneqq \mathcal{T}_1$. Now, fix any $t \in [\mathcal{T}_\ell + 1, \mathcal{T}_{\ell+1}]$ and consider the interval $s \in [\mathcal{T}'_{\ell-1}+1, \mathcal{T}''_{\ell-1}]$. Then, we have the following result

$$\sum_{s=\mathcal{T}'_{\ell-1}+1}^{\mathcal{T}''_{\ell-1}} \mathbb{E}\left[\max_{y\in\mathcal{A}_t^{(\ell-1)}} y^\top V_\ell^{-1} y\right] \leq \sum_{s=\mathcal{T}'_{\ell-1}+1}^{\mathcal{T}''_{\ell-1}} \mathbb{E}\left[\max_{y\in\mathcal{A}_t^{(\ell-1)}} y^\top H_{s-1}^{-1} y\right]$$
$$= \sum_{s=\mathcal{T}'_{\ell-1}+1}^{\mathcal{T}''_{\ell-1}} \mathbb{E}\left[\mathbb{E}\left[\max_{y\in\mathcal{A}_t^{(\ell-1)}} y^\top H_{s-1}^{-1} y \,\middle|\, \mathcal{F}_{s-1}\right]\right]$$
$$= \sum_{s=\mathcal{T}'_{\ell-1}+1}^{\mathcal{T}''_{\ell-1}} \mathbb{E}\left[\mathbb{E}\left[\max_{y\in\mathcal{A}_s^{(\ell-1)}} y^\top H_{s-1}^{-1} y \,\middle|\, \mathcal{F}_{s-1}\right]\right]$$
$$= \sum_{s=\mathcal{T}'_{\ell-1}+1}^{\mathcal{T}''_{\ell-1}} \mathbb{E}\left[\max_{y\in\mathcal{A}_s^{(\ell-1)}} y^\top H_{s-1}^{-1} y\right]. \tag{6}$$

The first inequality follows from the monotonicity of the matrices, as $H_{s-1} \preceq V_\ell$ for all s in the interval. The first and last equalities follow from the tower property. The second equality uses the fact that, conditional on $\mathcal{F}_{s-1}$, both H_{s-1} and the arm elimination rule (determined by $\hat{\theta}_1, \ldots, \hat{\theta}_{\ell-1}$) are fixed, and the distribution of the contexts remains unchanged due to their i.i.d. nature.

In Phase 2, the algorithm proceeds by selecting the most informative direction at each step with respect to the current Gram matrix. For all $\ell \geq 1$, we obtain

$$\sum_{s=\mathcal{T}'_{\ell-1}+1}^{\mathcal{T}''_{\ell-1}} \mathbb{E}\left[\max_{y\in\mathcal{A}_s^{(\ell-1)}} y^\top H_{s-1}^{-1} y\right] = \mathbb{E}\left[\sum_{s=\mathcal{T}'_{\ell-1}+1}^{\mathcal{T}''_{\ell-1}} \max_{y\in\mathcal{A}_s^{(\ell-1)}} y^\top H_{s-1}^{-1} y\right]$$
$$= \mathbb{E}\left[\sum_{s=\mathcal{T}'_{\ell-1}+1}^{\mathcal{T}''_{\ell-1}} x_{s,a_s}^\top H_{s-1}^{-1} x_{s,a_s}\right]$$
$$\leq \mathbb{E}\left[2\log\left(\frac{\det(H_{\mathcal{T}''_{\ell-1}})}{\det(H_{\mathcal{T}'_{\ell-1}})}\right)\right], \tag{7}$$

where the second equality follows from the arm-selection strategy of Phase 2, and the first inequality follows from Corollary 2.

Next, we bound the expected determinant growth during Phase 2. The previous argument reduces the problem to controlling

$$\mathbb{E}\left[2\log\left(\frac{\det(H_{\mathcal{T}''_{\ell-1}})}{\det(H_{\mathcal{T}'_{\ell-1}})}\right)\right].$$

We apply Lemma 10 with $j = \ell$. Let G_ℓ be the event defined in that lemma. Since

$$\mathbb{P}(G_\ell^c \mid \mathcal{F}_{\mathcal{T}_{\ell-1}}) \leq \frac{2}{T},$$

we have $\mathbb{P}(G_\ell^c) \leq 2/T$.

Define

$$Z_\ell \coloneqq 2\log\left(\frac{\det(H_{\mathcal{T}''_{\ell-1}})}{\det(H_{\mathcal{T}'_{\ell-1}})}\right).$$

By the tower property and Lemma 10,

$$\mathbb{E}[Z_\ell \mathbb{1}_{G_\ell}] = \mathbb{E}\Big[\mathbb{1}_{G_\ell}\mathbb{E}[Z_\ell \mid \mathcal{F}_{\mathcal{T}'_{\ell-1}}]\Big] \le 2d\log\left(1+\frac{384d\, n_{\ell,2}}{n_{\ell,1}}\right),$$

where

$$n_{\ell,1} \coloneqq \mathcal{T}'_{\ell-1} - \mathcal{T}_{\ell-1}, \qquad n_{\ell,2} \coloneqq \mathcal{T}''_{\ell-1} - \mathcal{T}'_{\ell-1}.$$

On the complement G_ℓ^c, we use the crude determinant bound. Since $H_{\mathcal{T}'_{\ell-1}} \succeq \lambda I$ and $\mathrm{tr}(H_{\mathcal{T}''_{\ell-1}}) \le d\lambda + T$, the AM–GM inequality gives

$$Z_\ell \le 2\log\left(\frac{(\mathrm{tr}(H_{\mathcal{T}''_{\ell-1}})/d)^d}{\lambda^d}\right) \le 2d\log\left(1+\frac{T}{d\lambda}\right) \le 2d\log(2T).$$

Therefore,

$$\begin{aligned}\mathbb{E}[Z_\ell] &= \mathbb{E}[Z_\ell \mathbb{1}_{G_\ell}] + \mathbb{E}[Z_\ell \mathbb{1}_{G_\ell^c}] \\ &\le 2d\log\left(1+\frac{384d\, n_{\ell,2}}{n_{\ell,1}}\right) + \frac{4d\log(2T)}{T}. \end{aligned} \tag{3}$$

For BLCE-G, $c \in (0,1)$ is fixed. By the definitions of $\mathcal{T}'_{\ell-1}$ and $\mathcal{T}''_{\ell-1}$, the ratio $n_{\ell,2}/n_{\ell,1}$ is bounded by a constant depending only on c. For instance, for all sufficiently large T,

$$\frac{n_{\ell,2}}{n_{\ell,1}} \le \frac{4}{c^2}.$$

Absorbing constants into the c-dependent constant, we get

$$\mathbb{E}[Z_\ell] \le 2d\log\left(\frac{1537d}{c^2}\right) + \frac{4d\log(2T)}{T}. \tag{8}$$

Now, we establish an upper bound on the cumulative expected regret. By combining the bounds derived in (3) and (4), we can bound the cumulative expected regret for interval ℓ, denoted by Regret_ℓ, for any $\ell \ge 2$ as follows

$$\begin{aligned}\mathrm{Regret}_\ell &= \sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell} \mathbb{E}[\langle x_t^* - x_{t,a_t}, \theta^*\rangle] \\ &\le 4\sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell} \mathbb{E}[\varepsilon_{t,\ell-1}] \\ &= 4\sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell} \mathbb{E}\left[\max_{y\in\mathcal{A}_t^{(\ell-2)}} \|y\|_{V_{\ell-1}^{-1}} \cdot \left(\beta_{t,\ell-1}^{(1)}\left(\frac{2}{T}\right) \wedge \beta_{t,\ell-1}^{(2)}\left(\frac{2}{T}\right)\right)\right] \\ &\le 4\sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell} \mathbb{E}\left[\max_{y\in\mathcal{A}_t^{(\ell-2)}} \|y\|_{V_{\ell-1}^{-1}}\right] \cdot \left(\sqrt{2\log\left(K(B-1)T^2\right)} + \sqrt{\lambda} \bigwedge 2\sqrt{\log\left(\frac{2^{6d-5}\pi d(B-1)^2T^2}{15^{d-1}}\right)} + 2\sqrt{\lambda}\right).\end{aligned}$$

The last inequality follows from the fact that the size of the candidate arm set satisfies $|\mathcal{A}_t^{(\ell-2)}| \le |\mathcal{A}_t| = K$ for all $t \in [\mathcal{T}_{\ell-1}+1, \mathcal{T}_\ell]$, allowing us to upper bound the confidence parameter $\beta_{t,\ell-1}^{(1)}$ uniformly over the action set.

Using (5), (6), and (8), we can bound the summation $\sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell} \mathbb{E}\left[\max_{y\in\mathcal{A}_t^{(\ell-2)}} \|y\|_{V_{\ell-1}^{-1}}\right]$ for any $\ell \ge 2$ as

$$\sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell} \mathbb{E}\left[\max_{y\in\mathcal{A}_t^{(\ell-2)}} \|y\|_{V_{\ell-1}^{-1}}\right] \le \sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell} \sqrt{\mathbb{E}\left[\max_{y\in\mathcal{A}_t^{(\ell-2)}} y^\top V_{\ell-1}^{-1} y\right]}$$

$$
\begin{aligned}
&= (\mathcal{T}_\ell - \mathcal{T}_{\ell-1}) \cdot \sqrt{\mathbb{E}\left[\max_{y \in \mathcal{A}_t^{(\ell-2)}} y^\top V_{\ell-1}^{-1} y\right]} \\
&\leq \frac{\mathcal{T}_\ell - \mathcal{T}_{\ell-1}}{\sqrt{\mathcal{T}''_{\ell-2} - \mathcal{T}'_{\ell-2}}} \cdot \sqrt{\sum_{s=\mathcal{T}'_{\ell-2}+1}^{\mathcal{T}''_{\ell-2}} \mathbb{E}\left[\max_{y \in \mathcal{A}_s^{(\ell-2)}} y^\top H_{s-1}^{-1} y\right]} \\
&\leq \frac{\mathcal{T}_\ell - \mathcal{T}_{\ell-1}}{\sqrt{\mathcal{T}''_{\ell-2} - \mathcal{T}'_{\ell-2}}} \cdot \sqrt{2d \log\left(\frac{1537d}{c^2}\right) + \frac{4d\log(2T)}{T}} \\
&\leq \frac{\frac{T^{1-2^{-\ell}}}{\log_2 \log_2 T} + 3}{\sqrt{\frac{c(1-c)T^{1-2^{1-\ell}}}{\log_2 \log_2 T}}} \cdot \sqrt{2d \log\left(\frac{1537d}{c^2}\right) + \frac{4d\log(2T)}{T}} \\
&= \left(\sqrt{\frac{T}{c(1-c)\log_2 \log_2 T}} + 3\sqrt{\frac{\log_2 \log_2 T}{c(1-c)T^{1-2^{1-\ell}}}}\right)\sqrt{2d \log\left(\frac{1537d}{c^2}\right) + \frac{4d\log(2T)}{T}} \\
&\leq \left(\sqrt{\frac{T}{c(1-c)\log_2 \log_2 T}} + \frac{3}{\sqrt{c(1-c)}}\right) \cdot \sqrt{2d \log d + 2\left(\log\left(\frac{1537}{c^2}\right) + 2\right) d} \\
&\leq c' \cdot \sqrt{\frac{dT \log d}{\log_2 \log_2 T}}\,,
\end{aligned}
$$

where t is arbitrary in the interval $[\mathcal{T}_{\ell-1}+1, \mathcal{T}_\ell]$, and we define $c' \coloneqq 4\sqrt{\frac{2(\log(1537/c^2)+3)}{c(1-c)}}$. The first inequality applies Jensen's inequality to move the square root outside the expectation. The first equality follows from (5), while the second inequality uses the bound in (6). The third inequality follows from (8). The fifth inequality follows from the facts that $\log_2 \log_2 T \leq T^{1-2^{1-\ell}}$ and $\log(2T) \leq T$ for all $T \geq 1$. The final inequality uses the bounds $1 \leq \frac{T}{\log_2 \log_2 T}$ and $d \leq d \log d$ for all $T \geq 1$ and $d \geq 3$.

We now derive the cumulative expected regret after the first interval. Based on the previously derived results, the total regret from intervals $\ell = 2$ to B can be bounded as follows

$$
\begin{aligned}
\sum_{\ell=2}^{B} \text{Regret}_\ell &\leq \sum_{\ell=2}^{B} 4c' \sqrt{\frac{dT \log d}{\log_2 \log_2 T}} \cdot \left(\sqrt{2\log\left(K(B-1)T^2\right)} + \sqrt{\lambda} \bigwedge 2\sqrt{\log\left(\frac{2^{6d-5}\pi d (B-1)^2 T^2}{15^{d-1}}\right)} + 2\sqrt{\lambda}\right) \\
&= \mathcal{O}\left(\sqrt{dT \log d \log\log T} \cdot \left(\sqrt{\log(KT)} + \sqrt{\log(dT)} \wedge \sqrt{d\log\left(\frac{64}{15}\right) + \log(dT)} + \sqrt{\log(dT)}\right)\right) \\
&= \mathcal{O}\left(\sqrt{dT \log d \log\log T} \cdot \left(\left(\sqrt{\log(KT)} + \sqrt{\log T}\right) \wedge \left(\sqrt{d + \log T} + \sqrt{\log T}\right)\right)\right) \\
&= \mathcal{O}\left(\sqrt{dT \log d \log\log T} \cdot \left(\sqrt{\log(KT)} \wedge \sqrt{d + \log T}\right)\right)\,.
\end{aligned}
$$

The first equality follows from substituting $B = \mathcal{O}(\log\log T)$ and $\lambda = \log(dT)$. The second equality holds in the regime where $T \geq d$. This is a safe assumption, as for $T \leq d$, the total regret is trivially bounded by $\mathcal{R}(T) = \sum_{t=1}^{T} \mathbb{E}[\langle x_t^* - x_{t,a_t}, \theta^* \rangle] \leq 2T \leq 2\sqrt{dT} = \mathcal{O}(\sqrt{dT})$, which is a much smaller bound.

Thus, the total worst-case regret is bounded as

$$
\begin{aligned}
\mathcal{R}(T) &\leq \sum_{k=1}^{B} \sum_{t=\mathcal{T}_{k-1}+1}^{\mathcal{T}_k} \mathbb{E}\left[\langle x_t^* - x_{t,a_t}, \theta^* \rangle \middle| E\right] \\
&= \text{Regret}_1 + \sum_{k=2}^{B} \text{Regret}_k
\end{aligned}
$$

$$= \mathcal{O}(\sqrt{T}) + \mathcal{O}\left(\sqrt{dT \log d \log\log T} \cdot \left(\sqrt{\log(KT)} \wedge \sqrt{d + \log T}\right)\right)$$
$$= \mathcal{O}\left(\sqrt{dT}\left(\sqrt{\log(KT)} \wedge \sqrt{d + \log T}\right)\sqrt{\log d \log\log T}\right) .$$

Therefore, the worst-case regret for the algorithm is given by

$$\mathcal{R}(T) = \mathcal{O}\left(\sqrt{dT}\left(\sqrt{\log(KT)} \wedge \sqrt{d + \log T}\right)\sqrt{\log d \log\log T}\right) = \widetilde{\mathcal{O}}\left(\sqrt{dT \log K} \wedge d\sqrt{T}\right) .$$

□

B. Proof of Theorem 2

Proof of Theorem 2. As Remark 1 encompasses the standard i.i.d. context assumption, we establish Theorem 2 under the more general conditions specified therein.

To establish the claimed interval complexity bound, we analyze the lower bound on the length of the ℓ-th interval, where $\ell = \lceil \log_2 \log_2 T \rceil$. By definition of the schedule, the length of this interval is at least

$$\left\lceil \frac{T^{1-2^{-\ell}}}{\log_2 \log_2 T} \right\rceil + 1 \geq \frac{T^{1-2^{-\lceil \log_2 \log_2 T \rceil}}}{\log_2 \log_2 T} \geq \frac{T^{1-2^{-\log_2 \log_2 T}}}{\log_2 \log_2 T} = \frac{T^{1-\frac{1}{\log_2 T}}}{\log_2 \log_2 T} = \frac{T}{2 \log_2 \log_2 T} \, .$$

Since the interval length is non-decreasing in ℓ, every interval with index $\ell \geq \lceil \log_2 \log_2 T \rceil$ has length at least $\frac{T}{2 \log_2 \log_2 T}$. Given the total time horizon T, the number of such intervals is therefore at most $\lceil 2 \log_2 \log_2 T \rceil$. Including the initial $\lceil \log_2 \log_2 T \rceil - 1$ intervals, the total number of intervals B is bounded by

$$B \leq \lceil 2 \log_2 \log_2 T \rceil + \lceil \log_2 \log_2 T \rceil - 1 \leq 3 \log_2 \log_2 T + 1 \, ,$$

which implies that the interval complexity is $\mathcal{O}(\log \log T)$.

Next, we begin by decomposing the cumulative expected regret based on the good event E. The regret can be written as

$$\begin{aligned} \mathcal{R}(T) &= \sum_{k=1}^{B} \sum_{t=\mathcal{T}_{k-1}+1}^{\mathcal{T}_k} \mathbb{E}\left[\max_{x \in \mathcal{A}_t} \langle x, \theta^* \rangle - \langle x_{t,a_t}, \theta^* \rangle \right] \\ &= \sum_{k=1}^{B} \sum_{t=\mathcal{T}_{k-1}+1}^{\mathcal{T}_k} \mathbb{E}\left[\langle x_t^* - x_{t,a_t}, \theta^* \rangle \middle| E^c \right] \cdot \mathbb{P}(E^c) + \mathbb{E}\left[\langle x_t^* - x_{t,a_t}, \theta^* \rangle \middle| E \right] \cdot \mathbb{P}(E) \, . \end{aligned}$$

Using the triangle inequality followed by the Cauchy-Schwarz inequality, we have

$$\begin{aligned} \langle x_t^* - x_{t,a_t}, \theta^* \rangle &\leq |\langle x_t^* - x_{t,a_t}, \theta^* \rangle| \\ &\leq |\langle x_t^*, \theta^* \rangle| + |\langle x_{t,a_t}, \theta^* \rangle| \\ &\leq \|x_t^*\| \cdot \|\theta^*\| + \|x_{t,a_t}\| \cdot \|\theta^*\| \\ &\leq 2 \, , \end{aligned}$$

where the last inequality follows from Assumption 1. Hence, the cumulative expected regret can be bounded as

$$\begin{aligned} \mathcal{R}(T) &\leq 2T \cdot \mathbb{P}(E^c) + \sum_{k=1}^{B} \sum_{t=\mathcal{T}_{k-1}+1}^{\mathcal{T}_k} \mathbb{E}\left[\langle x_t^* - x_{t,a_t}, \theta^* \rangle \middle| E \right] \cdot \mathbb{P}(E) \\ &\leq \mathcal{O}(1) + \sum_{k=1}^{B} \sum_{t=\mathcal{T}_{k-1}+1}^{\mathcal{T}_k} \mathbb{E}\left[\langle x_t^* - x_{t,a_t}, \theta^* \rangle \middle| E \right] \, , \end{aligned}$$

where the $\mathcal{O}(1)$ term follows from the high-probability guarantee $\mathbb{P}(E^c) = \mathcal{O}(1/T)$. Throughout the remainder of the analysis, we therefore condition on the good event E. Let Regret_ℓ denote the cumulative expected regret incurred during interval ℓ. We analyze the regret separately for the case $\ell = 1$ and for all subsequent intervals $\ell \geq 2$.

Case 1: $\ell = 1$. In the first interval, the number of rounds is $\mathcal{T}_1 = \left\lceil \frac{\sqrt{T}}{\log_2 \log_2 T} \right\rceil$, and the instantaneous regret is bounded by 2. Therefore,

$$\text{Regret}_1 = \sum_{t=1}^{\mathcal{T}_1} \mathbb{E}[\langle x_t^* - x_{t,a_t}, \theta^* \rangle] \leq 2\mathcal{T}_1 \leq 2\left(\frac{\sqrt{T}}{\log_2 \log_2 T} + 1 \right) = \mathcal{O}\big(\sqrt{T}\big) \, .$$

Case 2: $\ell \geq 2$. For each interval $\ell \geq 2$, the rounds following the arm elimination steps are divided into two phases:

Phase 1. In the first phase of length $\left\lceil \frac{cT^{1-2^{-\ell}}}{\log_2 \log_2 T} \right\rceil$, the algorithm selects the most informative direction with respect to the current Gram matrix. For any round t in these phases, the instantaneous regret satisfies

$$\begin{aligned}
\langle x_t^* - x_{t,a_t}, \theta^* \rangle &= \langle x_t^* - x_{t,a_t}, \theta^* - \hat{\theta}_{\ell-1} \rangle + \langle x_t^* - x_{t,a_t}, \hat{\theta}_{\ell-1} \rangle \\
&\leq \langle x_t^* - x_{t,a_t}, \theta^* - \hat{\theta}_{\ell-1} \rangle + \langle x_t^{(\ell-1)} - x_{t,a_t}, \hat{\theta}_{\ell-1} \rangle \\
&\leq |\langle x_t^*, \hat{\theta}_{\ell-1} - \theta^* \rangle| + |\langle x_{t,a_t}, \hat{\theta}_{\ell-1} - \theta^* \rangle| + 2\varepsilon_{t,\ell-1} \\
&\leq 4\varepsilon_{t,\ell-1} \, . \qquad (9)
\end{aligned}$$

The first inequality uses the fact that $x_t^{(\ell-1)}$ is optimal with respect to $\hat{\theta}_{\ell-1}$ and that x_t^* belongs to $\mathcal{A}_t^{(\ell-2)}$ by Lemma 4. The second inequality follows since $x_{t,a_t} \in \mathcal{A}_t^{(\ell-1)}$ by construction of the arm elimination step. The final inequality follows from the definition of the good event E and again from Lemma 4, which guarantees $x_t^* \in \mathcal{A}_t^{(\ell-2)}$.

Phase 2. In the last phase, the algorithm selects arms greedily with respect to the estimated parameter, i.e., $x_{t,a_t} \in \arg\max_{x \in \mathcal{A}_t^{(\ell-1)}} \langle x, \hat{\theta}_{\ell-1} \rangle$. For any round t in this phase, the instantaneous regret satisfies

$$\begin{aligned}
\langle x_t^* - x_{t,a_t}, \theta^* \rangle &= \langle x_t^* - x_{t,a_t}, \theta^* - \hat{\theta}_{\ell-1} \rangle + \langle x_t^* - x_{t,a_t}, \hat{\theta}_{\ell-1} \rangle \\
&\leq \langle x_t^* - x_{t,a_t}, \theta^* - \hat{\theta}_{\ell-1} \rangle \\
&\leq |\langle x_t^*, \hat{\theta}_{\ell-1} - \theta^* \rangle| + |\langle x_{t,a_t}, \hat{\theta}_{\ell-1} - \theta^* \rangle| \\
&\leq 2\varepsilon_{t,\ell-1} \, . \qquad (10)
\end{aligned}$$

The first inequality follows from the fact that x_{t,a_t} is optimal with respect to $\hat{\theta}_{\ell-1}$ and that x_t^* belongs to $\mathcal{A}_t^{(\ell-1)}$ by Lemma 4. The final inequality follows directly from the definition of the good event E and from Lemma 4, as in Phase 1.

Therefore, it suffices to upper bound the quantity $\sum_{t=\mathcal{T}_\ell+1}^{\mathcal{T}_{\ell+1}} \varepsilon_{t,\ell}$ for each $\ell \in [B-1]$, which reduces to bounding the sum $\sum_{t=\mathcal{T}_\ell+1}^{\mathcal{T}_{\ell+1}} \max_{y \in \mathcal{A}_t^{(\ell-1)}} \|y\|_{V_\ell^{-1}}$. Since both V_ℓ and the arm elimination rule—determined by $\hat{\theta}_1, \ldots, \hat{\theta}_{\ell-1}$—are measurable with respect to $\mathcal{F}_{\mathcal{T}_\ell}$, they can be treated as fixed quantities conditional on this filtration. Moreover, conditional on $\mathcal{F}_{\mathcal{T}_\ell}$ the action sets $\{\mathcal{A}_s\}_{s=\mathcal{T}_\ell+1}^{\mathcal{T}_{\ell+2}}$ are *identically distributed* with common law $\mathcal{D}_\ell$. Hence, by the tower property, for any $t, v \in [\mathcal{T}_\ell + 1, \mathcal{T}_{\ell+1}]$ we have

$$\begin{aligned}
\mathbb{E}\left[\max_{y \in \mathcal{A}_t^{(\ell-1)}} y^\top V_\ell^{-1} y \right] &= \mathbb{E}\left[\mathbb{E}\left[\max_{y \in \mathcal{A}_t^{(\ell-1)}} y^\top V_\ell^{-1} y \,\middle|\, \mathcal{F}_{\mathcal{T}_\ell} \right] \right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\max_{y \in \mathcal{A}_v^{(\ell-1)}} y^\top V_\ell^{-1} y \,\middle|\, \mathcal{F}_{\mathcal{T}_\ell} \right] \right] \\
&= \mathbb{E}\left[\max_{y \in \mathcal{A}_v^{(\ell-1)}} y^\top V_\ell^{-1} y \right] . \qquad (11)
\end{aligned}$$

Define $\mathcal{T}'_{\ell-1} \coloneqq \mathcal{T}_{\ell-1} + \left\lceil \frac{cT^{1-2^{-\ell}}}{\log_2 \log_2 T} \right\rceil$ for all $\ell \geq 2$, while for the first interval we set $\mathcal{T}'_0 \coloneqq \mathcal{T}_1$. Now fix any $t \in [\mathcal{T}_\ell + 1, \mathcal{T}_{\ell+1}]$ and consider the interval $s \in [\mathcal{T}_{\ell-1} + 1, \mathcal{T}'_{\ell-1}]$. Then we obtain

$$\begin{aligned}
\sum_{s=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}'_{\ell-1}} \mathbb{E}\left[\max_{y \in \mathcal{A}_t^{(\ell-1)}} y^\top V_\ell^{-1} y \right] &\leq \sum_{s=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}'_{\ell-1}} \mathbb{E}\left[\max_{y \in \mathcal{A}_t^{(\ell-1)}} y^\top H_{s-1}^{-1} y \right] \\
&= \sum_{s=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}'_{\ell-1}} \mathbb{E}\left[\mathbb{E}\left[\max_{y \in \mathcal{A}_t^{(\ell-1)}} y^\top H_{s-1}^{-1} y \,\middle|\, \mathcal{F}_{\mathcal{T}_{\ell-1}} \vee \sigma(H_{s-1}) \right] \right] \\
&= \sum_{s=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}'_{\ell-1}} \mathbb{E}\left[\mathbb{E}\left[\max_{y \in \mathcal{A}_s^{(\ell-1)}} y^\top H_{s-1}^{-1} y \,\middle|\, \mathcal{F}_{\mathcal{T}_{\ell-1}} \vee \sigma(H_{s-1}) \right] \right]
\end{aligned}$$

$$= \sum_{s=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}'_{\ell-1}} \mathbb{E}\left[\max_{y\in\mathcal{A}_s^{(\ell-1)}} y^\top H_{s-1}^{-1} y\right]. \tag{12}$$

The first inequality follows from monotonicity of the matrices, since $H_{s-1} \preceq V_\ell$ for all s in the interval. The first and last equalities follow from the tower property. The second equality relies on the fact that, conditional on $\mathcal{F}_{\mathcal{T}_{\ell-1}} \vee \sigma(H_{s-1})$, both H_{s-1} and the arm elimination rule (determined by $\hat{\theta}_1, \ldots, \hat{\theta}_{\ell-1}$) are fixed. Moreover, conditional on $\mathcal{F}_{\mathcal{T}_{\ell-1}}$, the action set $\mathcal{A}_t$ is independent of $\{\mathcal{A}_v, x_{v,a_v}, r_v\}_{v=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell}$ for $t \in [\mathcal{T}_\ell + 1, \mathcal{T}_{\ell+1}]$, and $\mathcal{A}_s$ is independent of $\{\mathcal{A}_u, x_{u,a_u}, r_u\}_{u=\mathcal{T}_{\ell-1}+1}^{s-1}$ for $s \in [\mathcal{T}_{\ell-1}+1, \mathcal{T}_\ell]$. Hence, conditional on $\mathcal{F}_{\mathcal{T}_{\ell-1}} \vee \sigma(H_{s-1})$, both $\mathcal{A}_t$ and $\mathcal{A}_s$ share the same conditional law $\mathcal{D}_{\ell-1}$, which justifies replacing $\mathcal{A}_t$ with $\mathcal{A}_s$ in the inner expectation.

In Phase 1, the algorithm proceeds by selecting the most informative direction at each step with respect to the current Gram matrix. For all $\ell \geq 1$, we obtain

$$\begin{aligned}
\sum_{s=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}'_{\ell-1}} \mathbb{E}\left[\max_{y\in\mathcal{A}_s^{(\ell-1)}} y^\top H_{s-1}^{-1} y\right] &= \mathbb{E}\left[\sum_{s=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}'_{\ell-1}} \max_{y\in\mathcal{A}_s^{(\ell-1)}} y^\top H_{s-1}^{-1} y\right] \\
&= \mathbb{E}\left[\sum_{s=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}'_{\ell-1}} x_{s,a_s}^\top H_{s-1}^{-1} x_{s,a_s}\right] \\
&\leq \mathbb{E}\left[2\log\left(\frac{\det(H_{\mathcal{T}'_{\ell-1}})}{\det(\lambda I)}\right)\right] \\
&\leq \mathbb{E}\left[2\log\left(\frac{(\mathrm{tr}(H_{\mathcal{T}'_{\ell-1}})/d)^d}{\lambda^d}\right)\right] \\
&\leq 2d\log\left(1+\frac{T}{d\lambda}\right) \\
&\leq 2d\log(2T)\,,
\end{aligned} \tag{13}$$

where the second equality follows from the arm-selection strategy of Phase 1, and the first inequality follows from Corollary 2. The second inequality replaces the determinant by the bound $\det(H) \leq (\mathrm{tr}(H)/d)^d$, which is a consequence of the AM–GM inequality. The third inequality applies the fact that $\|x\|_2 \leq 1$ for all contexts, which implies $\mathrm{tr}(H_{\mathcal{T}'_{\ell-1}}) \leq d\lambda + T$.

Now, we establish an upper bound on the cumulative expected regret. By combining the bounds derived in (9) and (10), we can bound the cumulative expected regret for interval ℓ, denoted by Regret_ℓ, for any $\ell \geq 2$ as follows

$$\begin{aligned}
\mathrm{Regret}_\ell &= \sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell} \mathbb{E}[\langle x_t^* - x_{t,a_t}, \theta^*\rangle] \\
&\leq 4 \sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell} \mathbb{E}[\varepsilon_{t,\ell-1}] \\
&= 4 \sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell} \mathbb{E}\left[\max_{y\in\mathcal{A}_t^{(\ell-2)}} \|y\|_{V_{\ell-1}^{-1}} \cdot \left(\beta_{t,\ell-1}^{(1)}\left(\frac{2}{T}\right) \wedge \beta_{t,\ell-1}^{(2)}\left(\frac{2}{T}\right)\right)\right] \\
&\leq 4 \sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell} \mathbb{E}\left[\max_{y\in\mathcal{A}_t^{(\ell-2)}} \|y\|_{V_{\ell-1}^{-1}}\right] \cdot \left(\sqrt{2\log\left(K(B-1)T^2\right)} + \sqrt{\lambda} \bigwedge 2\sqrt{\log\left(\frac{2^{6d-5}\pi d(B-1)^2T^2}{15^{d-1}}\right)} + 2\sqrt{\lambda}\right).
\end{aligned}$$

The last inequality follows from the fact that the size of the candidate arm set satisfies $|\mathcal{A}_t^{(\ell-2)}| \leq |\mathcal{A}_t| = K$ for all $t \in [\mathcal{T}_{\ell-1}+1, \mathcal{T}_\ell]$, allowing us to upper bound the confidence parameter $\beta_{t,\ell-1}^{(1)}$ uniformly over the action set.

Using (11), (12), and (13), we can bound the summation $\sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell} \mathbb{E}\left[\max_{y\in\mathcal{A}_t^{(\ell-2)}} \|y\|_{V_{\ell-1}^{-1}}\right]$ for any $\ell \geq 2$ as

$$\begin{aligned}
\sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell} \mathbb{E}\left[\max_{y\in\mathcal{A}_t^{(\ell-2)}} \|y\|_{V_{\ell-1}^{-1}}\right] &\leq \sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell} \sqrt{\mathbb{E}\left[\max_{y\in\mathcal{A}_t^{(\ell-2)}} y^\top V_{\ell-1}^{-1} y\right]} \\
&= (\mathcal{T}_\ell - \mathcal{T}_{\ell-1}) \cdot \sqrt{\mathbb{E}\left[\max_{y\in\mathcal{A}_t^{(\ell-2)}} y^\top V_{\ell-1}^{-1} y\right]} \\
&\leq \frac{\mathcal{T}_\ell - \mathcal{T}_{\ell-1}}{\sqrt{\mathcal{T}'_{\ell-2} - \mathcal{T}_{\ell-2}}} \cdot \sqrt{\sum_{s=\mathcal{T}_{\ell-2}+1}^{\mathcal{T}'_{\ell-2}} \mathbb{E}\left[\max_{y\in\mathcal{A}_s^{(\ell-2)}} y^\top H_{s-1}^{-1} y\right]} \\
&\leq \frac{\mathcal{T}_\ell - \mathcal{T}_{\ell-1}}{\sqrt{\mathcal{T}'_{\ell-2} - \mathcal{T}_{\ell-2}}} \cdot \sqrt{2d\log(2T)} \\
&\leq \frac{\frac{T^{1-2^{-\ell}}}{\log_2\log_2 T} + 2}{\sqrt{\frac{cT^{1-2^{1-\ell}}}{\log_2\log_2 T}}} \cdot \sqrt{2d\log(2T)} \\
&= \left(\sqrt{\frac{T}{c\log_2\log_2 T}} + 2\sqrt{\frac{\log_2\log_2 T}{cT^{1-2^{1-\ell}}}}\right) \cdot \sqrt{2d\log(2T)} \\
&\leq \left(\sqrt{\frac{T}{c\log_2\log_2 T}} + \frac{2}{\sqrt{c}}\right) \cdot 2\sqrt{d\log T} \\
&\leq c' \cdot \sqrt{\frac{dT\log T}{\log_2\log_2 T}},
\end{aligned}$$

where t is arbitrary in the interval $[\mathcal{T}_{\ell-1}+1, \mathcal{T}_\ell]$, and we define $c' \coloneqq \frac{6}{\sqrt{c}}$. The first inequality applies Jensen's inequality to move the square root outside the expectation. The first equality follows from (11), while the second inequality uses the bound in (12). The third inequality follows from (13). The fifth inequality follows from the facts that $\log_2\log_2 T \leq T^{1-2^{1-\ell}}$ and $\log(2T) \leq 2\log T$ for all $T \geq 2$. The final inequality uses the bounds $1 \leq \frac{T}{\log_2\log_2 T}$ for all $T \geq 1$.

We now derive the cumulative expected regret after the first interval. Building on the previously established results, the total regret incurred from intervals $\ell = 2$ to B can be bounded as

$$\begin{aligned}
\sum_{\ell=2}^{B} \text{Regret}_\ell &\leq \sum_{\ell=2}^{B} 4c' \sqrt{\frac{dT\log T}{\log_2\log_2 T}} \cdot \left(\sqrt{2\log\left(K(B-1)T^2\right)} + \sqrt{\lambda} \bigwedge 2\sqrt{\log\left(\frac{2^{6d-5}\pi d(B-1)^2T^2}{15^{d-1}}\right)} + 2\sqrt{\lambda}\right) \\
&= \mathcal{O}\left(\sqrt{dT\log T\log\log T} \cdot \left(\sqrt{\log(KT)} \wedge \sqrt{d\log\left(\frac{64}{15}\right) + \log(dT)}\right)\right) \\
&= \mathcal{O}\left(\sqrt{dT\log T\log\log T} \cdot \left(\sqrt{\log(KT)} \wedge \sqrt{d + \log T}\right)\right).
\end{aligned}$$

Here, the first equality follows from substituting $B = \mathcal{O}(\log\log T)$ and $\lambda = \mathcal{O}(1)$, while the second equality holds because the $\log(d)$ term is dominated by d.

Thus, the total worst-case regret is bounded as

$$\begin{aligned}
\mathcal{R}(T) &\leq \sum_{k=1}^{B} \sum_{t=\mathcal{T}_{k-1}+1}^{\mathcal{T}_k} \mathbb{E}\left[\langle x_t^* - x_{t,a_t}, \theta^*\rangle \middle| E\right] \\
&= \text{Regret}_1 + \sum_{k=2}^{B} \text{Regret}_k
\end{aligned}$$

$$= \mathcal{O}(\sqrt{T}) + \mathcal{O}\left(\sqrt{dT \log T \log\log T} \cdot \left(\sqrt{\log(KT)} \wedge \sqrt{d + \log T}\right)\right)$$
$$= \mathcal{O}\left(\sqrt{dT}\left(\sqrt{\log(KT)} \wedge \sqrt{d + \log T}\right)\sqrt{\log T \log\log T}\right) .$$

Therefore, the worst-case regret for the algorithm is given by

$$\mathcal{R}(T) = \mathcal{O}\left(\sqrt{dT}\left(\sqrt{\log(KT)} \wedge \sqrt{d + \log T}\right)\sqrt{\log T \log\log T}\right) = \widetilde{\mathcal{O}}\left(\sqrt{dT \log K} \wedge d\sqrt{T}\right) .$$

□

Algorithm 3 BGLE

Input: Horizon T; interval end times $\mathcal{T}_\ell = \ell\left\lceil \frac{\sqrt[3]{T}}{\log_2 \log_2 T} \right\rceil$ for $\ell \le 3$, $\mathcal{T}_\ell = \left(\mathcal{T}_{\ell-1} + \left\lceil \frac{T^{1-\frac{1}{3\cdot 2^{\ell-4}}}}{\log_2 \log_2 T} \right\rceil\right) \wedge T$ for $\ell \ge 4$; number of intervals B, with $\mathcal{T}_B = T$; within-interval allocation rate c;

Initialize: $\lambda \leftarrow R^2(d + \log T)$, $H_0 \leftarrow \lambda I$;
for $t \leftarrow 1, 2, \ldots, \mathcal{T}_1$ **do**
  Pull arm $x_{t,a_t} \in \arg\max_{x \in \mathcal{A}_t} \|x\|_{H_{t-1}^{-1}}$;
  $H_t \leftarrow H_{t-1} + x_{t,a_t} x_{t,a_t}^\top$;
$V_1 \leftarrow H_{\mathcal{T}_1}$, $\hat{\theta}_1 \leftarrow \arg\min_\theta \sum_{t=1}^{\mathcal{T}_1} \ell_t(\theta)$, $H_{\mathcal{T}_1} \leftarrow \lambda I$;
for $\ell \leftarrow 2, \ldots, B$ **do**
  for $t \leftarrow \mathcal{T}_{\ell-1} + 1, \ldots, \mathcal{T}_\ell$ **do**
    if $\ell \ge 3$ **then**
      for $k \leftarrow 2, \ldots, \ell - 1$ **do**
        $x_t^{(k)} \leftarrow \arg\max_{x \in \mathcal{A}_t^{(k-1)}} \langle x, \hat{\theta}_k \rangle$;
        $\mathcal{A}_t^{(k)} \leftarrow \left\{ x \in \mathcal{A}_t^{(k-1)} \,\middle|\, \langle \hat{\theta}_k, x_t^{(k)} - x \rangle \le 2\varepsilon'_{t,k}(\lambda) \right\}$;
    if $t \le \mathcal{T}_{\ell-1} + \left\lceil cT^{1-2^{((4-\ell)\wedge 1)}/3} / \log_2 \log_2 T \right\rceil$ **then**
      Pull arm $x_{t,a_t} \in \arg\max_{x \in \mathcal{A}_t^{(\ell-1)}} \|x\|_{H_{t-1}^{-1}}$;
    else
      Pull arm $x_{t,a_t} \in \arg\max_{x \in \mathcal{A}_t^{(\ell-1)}} \langle x, \hat{\theta}_{\ell-1} \rangle$;
    $H_t \leftarrow H_{t-1} + \alpha_{t,\ell-1}(\lambda) \dot{\mu}(\langle x_{t,a_t}, \hat{\theta}_{\ell-1} \rangle) x_{t,a_t} x_{t,a_t}^\top$;
  $V_\ell \leftarrow H_{\mathcal{T}_\ell}$, $\hat{\theta}_\ell \leftarrow \arg\min_\theta \sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell} \ell_t(\theta)$, $H_{\mathcal{T}_\ell} \leftarrow \lambda I$;

C. Proof of Theorem 3

Problem Setting: Generalized Linear Contextual Bandits with Batched Feedback We next consider generalized linear contextual bandits, where rewards follow a one-parameter exponential family distribution. Given an arm $x \in \mathbb{R}^d$ and an unknown parameter $\theta^* \in \mathbb{R}^d$, the reward r has density $p(r \mid x; \theta^*) = \exp\big(r\langle x, \theta^* \rangle - m(\langle x, \theta^* \rangle) + h(r)\big)\, \nu(\mathrm{d}r)$, with log-partition function m, base measure ν, and link function $\mu(z) := m'(z)$. We adopt the following standard assumptions.

Assumption 3. *$\|x\|_2 \le 1$ for all $x \in \mathcal{A}_t$, and $\|\theta^*\|_2 \le S$ for a known constant $S > 0$.*

Assumption 4. *The log-partition function m is convex and three times differentiable. Equivalently, $\dot{\mu} := m'' \ge 0$ and m''' exists.*

At each round $t \in [T]$, the learner observes an arm set $\mathcal{A}_t = \{x_{t,1}, \ldots, x_{t,K}\} \subseteq \mathbb{R}^d$ and selects $x_{t,a_t} \in \mathcal{A}_t$. The reward r_t is drawn from $p(r \mid x_{t,a_t}; \theta^*)$ with natural parameter $\langle x_{t,a_t}, \theta^* \rangle$, and satisfies $\mathbb{E}[r \mid x; \theta^*] = \mu(\langle x, \theta^* \rangle)$. Performance is measured by the *cumulative expected regret*

$$\mathcal{R}(T) = \mathbb{E}\left[\sum_{t=1}^{T} \big(\mu(\langle x_t^*, \theta^* \rangle) - \mu(\langle x_{t,a_t}, \theta^* \rangle) \big) \right],$$

where $x_t^* \in \operatorname{argmax}_{x \in \mathcal{A}_t} \mu(\langle x, \theta^* \rangle)$ is an optimal arm. Following Sawarni et al. (2024), we assume i.i.d. contexts drawn from an unknown distribution $\mathcal{D}$ and rewards almost surely supported on $[0, R]$, which implies the self-concordance condition $|\ddot{\mu}(z)| \le R\dot{\mu}(z)$ for all $z \in \mathbb{R}$. This property is crucial for our analysis.

We begin by assuming that the MLE estimator $\hat{\theta}$, obtained by minimizing the log-loss objective, always satisfies the boundedness condition $\|\hat{\theta}\|_2 \le S$. If this condition does not hold, one may instead apply the non-convex projection technique of Sawarni et al. (2024). The projected estimator preserves the same guarantees established in Sawarni et al. (2024), up to a multiplicative factor of 2. Therefore, the assumption $\|\hat{\theta}\|_2 \le S$ can be made without loss of generality.

Lemma 11. *For any $x \in [0, C]$, the following inequality holds*

$$e^x \le \frac{x(e^C - 1)}{C} + 1 \,.$$

Table 3. Worst-case regret, *parameter-update* complexity, and time complexity comparison for generalized linear contextual bandits. We distinguish *rare parameter updates* (the number of batch boundaries at which reward feedback is incorporated by re-solving the log-loss MLE) from the stronger *strict batching* restriction that the algorithm is *not* within-interval context-adaptive. **BGLE attains the tightest regret bound among the listed methods** using only $\mathcal{O}(\log\log T)$ batches, avoids any explicit dependence on the curvature parameter κ in the regret (beyond the data-driven $\hat{\kappa}$), and achieves the lowest time complexity. [1]

Paper	Worst-Case Regret	Parameter Updates	Context Adaptive	Time Complexity
Sawarni et al. (2024)	$\mathcal{O}\Big((RSd(\sqrt{d/\hat{\kappa}} \wedge \sqrt{R_{\dot{\mu}}\log d})\sqrt{\log T}\log\log T + R)\sqrt{T}\Big)$ $+\mathcal{O}\Big((\kappa R_{\dot{\mu}} R^5 S^2)^{1/3} e^{2RS} d^2 (\log T)^{2/3}\log\log T \cdot T^{\frac{1}{3}}\Big)$	$\mathcal{O}(\log\log T)$	No	$\mathcal{O}(Kd^4 T\log T + \mathcal{C}_{\text{opt}}^{\text{tot}})$ $+\mathcal{O}\big(Kd^5(\kappa R_{\dot{\mu}})^{1/3} e^{2RS}$ $(R^2S^2T\log^2 T)^{1/3}\big)$
Algorithm 3	$\mathcal{O}\Big(RS\sqrt{d(d+\log T)\log T\log\log T/\hat{\kappa}}\cdot\sqrt{T}\Big)$ $+\mathcal{O}\Big((R^2 S e^{8RS} d(d+\log T)\log T\log\log T + \frac{R}{\log\log T})T^{\frac{1}{3}}\Big)$	$\mathcal{O}(\log\log T)$	Yes	$\mathcal{O}(Kd^2 T\log\log T)$ $+\mathcal{O}(\mathcal{C}_{\text{opt}}^{\text{tot}})$

Proof. Apply the definition of convexity, $f\big((1-\alpha)a+\alpha b\big) \le (1-\alpha)f(a)+\alpha f(b)$, to $f(t)=e^t$ with $a=0$, $b=C$, and $\alpha = x/C \in [0,1]$. This gives

$$e^{(1-\alpha)0+\alpha C} \le (1-\alpha)e^0 + \alpha e^C \quad \Rightarrow \quad e^x \le 1 + \tfrac{x}{C}(e^C-1)\,,$$

which is the claim. □

Lemma 12. *For an exponential family distribution with log-partition function $m(\cdot)$, let $\mu(z) := m'(z)$. Then, for all $x_1, x_2 \in \mathbb{R}$, we have*

$$e^{-R|x_2-x_1|}\dot{\mu}(x_2) \le \dot{\mu}(x_1) \le e^{R|x_2-x_1|}\dot{\mu}(x_2)\,.$$

Proof. Without loss of generality, assume that $x_2 \ge x_1$. Define $h_1(x) := \dot{\mu}(x)e^{Rx}$ and $h_2(x) := \dot{\mu}(x)e^{-Rx}$. Differentiating these functions yields $h_1'(x) = (\ddot{\mu}(x) + R\dot{\mu}(x))e^{Rx}$ and $h_2'(x) = (\ddot{\mu}(x) - R\dot{\mu}(x))e^{-Rx}$. By the self-concordance property, we have $h_1'(x) \ge 0$ and $h_2'(x) \le 0$, which implies that $h_1(x)$ is non-decreasing and $h_2(x)$ is non-increasing. Consequently, $h_1(x_2) \ge h_1(x_1)$ and $h_2(x_2) \le h_2(x_1)$, which together establish the desired inequality. □

Lemma 13. *(Sawarni et al., 2024) For each interval $\ell \ge 1$, let $r_{\mathcal{T}_{\ell-1}+1}, \ldots, r_{\mathcal{T}_\ell}$ denote independent random variables drawn from the canonical exponential family such that $\mathbb{E}[r_s] = \mu(\langle x_{s,a_s}, \theta^*\rangle)$ for some $\theta^* \in \mathbb{R}^d$. Define the maximum likelihood estimator by $\hat{\theta}_\ell = \arg\min_\theta \sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell} \ell_t(\theta)$, and let $V_\ell^* := \lambda I + \sum_{s=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell} \dot{\mu}(\langle x_{s,a_s}, \theta^*\rangle) x_{s,a_s} x_{s,a_s}^\top$. Then, with probability at least $1-\frac{1}{T^2}$, the following inequality holds*

$$\|\hat{\theta}_\ell - \theta^*\|_{V_\ell^*} \le 24RS\left(\sqrt{d+\log T} + \frac{R(d+\log T)}{\sqrt{\lambda}}\right) + 2S\sqrt{\lambda} \triangleq \beta(\lambda)\,.$$

Lemma 14. *For any interval $\ell \ge 2$, define $V_\ell^* := \lambda I + \sum_{s=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell} \dot{\mu}(\langle x_{s,a_s}, \theta^*\rangle) x_{s,a_s} x_{s,a_s}^\top$ and $V_\ell := \lambda I + \sum_{s=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell} \alpha_{s,\ell-1}(\lambda)\dot{\mu}(\langle x_{s,a_s}, \hat{\theta}_{\ell-1}\rangle) x_{s,a_s} x_{s,a_s}^\top$. Then, for every $\ell \ge 2$, with probability at least $1-\frac{1}{T^2}$, the following matrix inequality holds*

$$V_\ell \preceq V_\ell^*\,.$$

Proof. We first consider the case $\ell = 2$, applying Lemma 12 yields

$$e^{-R|\langle x_{s,a_s}, \hat{\theta}_1 - \theta^*\rangle|}\dot{\mu}(\langle x_{s,a_s}, \hat{\theta}_1\rangle) \le \dot{\mu}(\langle x_{s,a_s}, \theta^*\rangle)\,.$$

By the assumptions $\|x_{s,a_s}\|_2 \le 1$, $\|\theta^*\|_2 \le S$, and $\|\hat{\theta}_1\|_2 \le S$, we further obtain

$$|\langle x_{s,a_s}, \hat{\theta}_1 - \theta^*\rangle| \le \|x_{s,a_s}\|_2 \cdot \|\hat{\theta}_1 - \theta^*\|_2 \le \|\hat{\theta}_1\|_2 + \|\theta^*\|_2 \le 2S\,.$$

[1] $\mathcal{C}_{\text{opt}}^{\text{tot}}$ denotes the total oracle cost of solving the log-loss minimization at interval boundaries. With B intervals, $\mathcal{C}_{\text{opt}}^{\text{tot}} = \sum_{\ell=1}^{B} \mathcal{C}_{\text{opt}}(\mathcal{T}_\ell - \mathcal{T}_{\ell-1}, d)$, where $\mathcal{C}_{\text{opt}}(n,d)$ is the cost of computing the unconstrained MLE from n samples in d dimensions.

Consequently,

$$\begin{aligned}\alpha_{s,1}(\lambda)\dot{\mu}(\langle x_{s,a_s},\hat{\theta}_1\rangle) &= e^{-2RS}\dot{\mu}(\langle x_{s,a_s},\hat{\theta}_1\rangle)\\ &\le e^{-R|\langle x_{s,a_s},\hat{\theta}_1-\theta^*\rangle|}\dot{\mu}(\langle x_{s,a_s},\hat{\theta}_1\rangle)\\ &\le \dot{\mu}(\langle x_{s,a_s},\theta^*\rangle)\ ,\end{aligned}$$

which establishes $V_2 \preceq V_2^*$.

For the general case $\ell \ge 3$, Lemma 12 gives

$$e^{-R|\langle x_{s,a_s},\hat{\theta}_{\ell-1}-\theta^*\rangle|}\dot{\mu}(\langle x_{s,a_s},\hat{\theta}_{\ell-1}\rangle) \le \dot{\mu}(\langle x_{s,a_s},\theta^*\rangle)\ .$$

Using Lemma 13 together with the assumptions $\|x_{s,a_s}\|_2 \le 1$, $\|\theta^*\|_2 \le S$ and $\|\hat{\theta}_{\ell-1}\|_2 \le S$, we obtain

$$|\langle x_{s,a_s},\hat{\theta}_{\ell-1}-\theta^*\rangle| \le (2S \wedge \|x_{s,a_s}\|_{V_{\ell-1}^{*-1}}\beta(\lambda)) \le (2S \wedge \|x_{s,a_s}\|_{V_{\ell-1}^{-1}}\beta(\lambda))\ ,$$

where the second inequality follows inductively from $V_{\ell-1} \preceq V_{\ell-1}^*$. Therefore,

$$\begin{aligned}\alpha_{s,\ell-1}(\lambda)\dot{\mu}(\langle x_{s,a_s},\hat{\theta}_{\ell-1}\rangle) &= e^{-R(2S\wedge\|x_{s,a_s}\|_{V_{\ell-1}^{-1}}\beta(\lambda))}\dot{\mu}(\langle x_{s,a_s},\hat{\theta}_{\ell-1}\rangle)\\ &\le e^{-R|\langle x_{s,a_s},\hat{\theta}_{\ell-1}-\theta^*\rangle|}\dot{\mu}(\langle x_{s,a_s},\hat{\theta}_{\ell-1}\rangle)\\ &\le \dot{\mu}(\langle x_{s,a_s},\theta^*\rangle)\ ,\end{aligned}$$

which completes the proof that $V_\ell \preceq V_\ell^*$ for all $\ell \ge 2$. □

Lemma 15 (Good event)**.** *Define the following quantities:*

$$\begin{aligned}\beta(\lambda) &\coloneqq 24RS\left(\sqrt{d+\log T}+\frac{R(d+\log T)}{\sqrt{\lambda}}\right)+2S\sqrt{\lambda}\,,\\ \varepsilon'_{t,\ell}(\lambda) &\coloneqq \max_{y\in\mathcal{A}_t^{(\ell-1)}}\|y\|_{V_\ell^{-1}}\cdot\beta(\lambda) \ \textit{ for } \ \ell\ge 2\ .\end{aligned}$$

Then, with probability at least $1-\frac{2(B-2)}{T^2}$*, the following event holds*

$$E' \coloneqq \bigcap_{\ell=2}^{B-1}\bigcap_{t=\mathcal{T}_\ell+1}^{T}\left\{|\langle x,\hat{\theta}_\ell-\theta^*\rangle| \le \varepsilon'_{t,\ell}(\lambda),\ \forall x\in\mathcal{A}_t^{(\ell-1)}\right\}\ .$$

Proof. For any interval $\ell \ge 2$, by the Cauchy-Schwarz inequality together with Lemma 13 and Lemma 14, it follows that for every round t and for all $x \in \mathcal{A}_t^{(\ell-1)}$, with probability at least $1-\frac{2}{T^2}$ we have

$$|\langle x,\hat{\theta}_\ell-\theta^*\rangle| \le \|x\|_{V_\ell^{*-1}}\|\hat{\theta}_\ell-\theta^*\|_{V_\ell^*} \le \|x\|_{V_\ell^{-1}}\cdot\beta(\lambda) \le \max_{y\in\mathcal{A}_t^{(\ell-1)}}\|y\|_{V_\ell^{-1}}\cdot\beta(\lambda)\ .$$

Applying a union bound over all intervals ℓ then guarantees that the event E' holds with probability at least $1-\frac{2(B-2)}{T^2}$, which completes the proof. □

Lemma 16. *Let E' be the good event defined in Lemma 15. Conditioned on E', the optimal arm $x_t^* \in \arg\max_{x\in\mathcal{A}_t}\langle x,\theta^*\rangle$ is never eliminated at any round t. In particular,*

$$x_t^* \in \mathcal{A}_t^{(\ell)}, \quad \textit{for all } 1\le\ell\le B-1 \textit{ and } \mathcal{T}_\ell+1\le t\le T\ .$$

Proof. Fix $t \in [\mathcal{T}_s+1, \mathcal{T}_{s+1}]$ for some $s \in [B-1]$. We show by induction on ℓ that $x_t^* \in \mathcal{A}_t^{(\ell)}$ for all $\ell \in [s]$.

Base case ($\ell = 1$). By definition we have $\mathcal{A}_t = \mathcal{A}_t^{(0)} = \mathcal{A}_t^{(1)}$, which immediately implies that $x_t^* \in \mathcal{A}_t^{(1)}$ holds trivially.

Inductive step. Assume $x_t^* \in \mathcal{A}_t^{(\ell-1)}$ for some $\ell \in \{2, \dots, s\}$. Since $x_t^{(\ell)} \in \mathcal{A}_t^{(\ell-1)}$, we similarly obtain

$$\begin{aligned}\langle \hat{\theta}_\ell, x_t^{(\ell)} - x_t^* \rangle &= \langle \hat{\theta}_\ell - \theta^*, x_t^{(\ell)} - x_t^* \rangle + \langle \theta^*, x_t^{(\ell)} - x_t^* \rangle \\ &\le \langle \hat{\theta}_\ell - \theta^*, x_t^{(\ell)} - x_t^* \rangle \\ &\le |\langle \hat{\theta}_\ell - \theta^*, x_t^{(\ell)} \rangle| + |\langle \hat{\theta}_\ell - \theta^*, x_t^* \rangle| \\ &\le 2\varepsilon'_{t,\ell}(\lambda) \ ,\end{aligned}$$

which shows that $x_t^* \in \mathcal{A}_t^{(\ell)}$. By induction, the claim holds for all $\ell \in [s]$, completing the proof. □

Proof of Theorem 3. To establish the claimed interval complexity bound, we analyze the lower bound on the length of the ℓ-th interval, where $\ell = \lceil \log_2 \log_2 T \rceil + 4$. By definition of the schedule, the length of this interval is at least

$$\left\lceil \frac{T^{1-\frac{1}{3\cdot 2^{\ell-4}}}}{\log_2 \log_2 T} \right\rceil \ge \frac{T^{1-2^{-\lceil \log_2 \log_2 T \rceil}}}{\log_2 \log_2 T} \ge \frac{T^{1-2^{-\log_2 \log_2 T}}}{\log_2 \log_2 T} = \frac{T^{1-\frac{1}{\log_2 T}}}{\log_2 \log_2 T} = \frac{T}{2 \log_2 \log_2 T} \ .$$

Since the interval length is non-decreasing in ℓ, every interval with index $\ell \ge \lceil \log_2 \log_2 T \rceil + 4$ has length at least $\frac{T}{2 \log_2 \log_2 T}$. Given the total time horizon T, the number of such intervals is therefore at most $\lceil 2 \log_2 \log_2 T \rceil$. Including the initial $\lceil \log_2 \log_2 T \rceil + 3$ intervals, the total number of intervals B is bounded by

$$B \le \lceil 2 \log_2 \log_2 T \rceil + \lceil \log_2 \log_2 T \rceil + 3 \le 3 \log_2 \log_2 T + 5 \ ,$$

which implies that the interval complexity is $\mathcal{O}(\log \log T)$.

Next, we decompose the cumulative expected regret with respect to the good event E'. The regret can be expressed as

$$\begin{aligned}\mathcal{R}(T) &= \sum_{k=1}^{B} \sum_{t=\mathcal{T}_{k-1}+1}^{\mathcal{T}_k} \mathbb{E}\left[\max_{x \in \mathcal{A}_t} \mu(\langle x, \theta^* \rangle) - \mu(\langle x_{t,a_t}, \theta^* \rangle) \right] \\ &= \sum_{k=1}^{B} \sum_{t=\mathcal{T}_{k-1}+1}^{\mathcal{T}_k} \mathbb{E}\left[\mu(\langle x_t^*, \theta^* \rangle) - \mu(\langle x_{t,a_t}, \theta^* \rangle) \middle| E'^c \right] \cdot \mathbb{P}(E'^c) + \mathbb{E}\left[\mu(\langle x_t^*, \theta^* \rangle) - \mu(\langle x_{t,a_t}, \theta^* \rangle) \middle| E' \right] \cdot \mathbb{P}(E') \ .\end{aligned}$$

Since the rewards are supported on $[0, R]$ and $\mathbb{E}[r \mid x; \theta^*] = \mu(\langle x, \theta^* \rangle)$ holds, the instantaneous regret is bounded by the support width, i.e.,

$$\mu(\langle x_t^*, \theta^* \rangle) - \mu(\langle x_{t,a_t}, \theta^* \rangle) \le R \ . \tag{14}$$

Hence, the cumulative expected regret can be bounded as

$$\begin{aligned}\mathcal{R}(T) &\le RT \cdot \frac{2(B-2)}{T^2} + \sum_{k=1}^{B} \sum_{t=\mathcal{T}_{k-1}+1}^{\mathcal{T}_k} \mathbb{E}\left[\mu(\langle x_t^*, \theta^* \rangle) - \mu(\langle x_{t,a_t}, \theta^* \rangle) \middle| E' \right] \cdot \mathbb{P}(E') \\ &\le \frac{2R(B-2)}{T} + \sum_{k=1}^{B} \sum_{t=\mathcal{T}_{k-1}+1}^{\mathcal{T}_k} \mathbb{E}\left[\mu(\langle x_t^*, \theta^* \rangle) - \mu(\langle x_{t,a_t}, \theta^* \rangle) \middle| E' \right] \\ &= \mathcal{O}\left(\frac{R \log \log T}{T} \right) + \sum_{k=1}^{B} \sum_{t=\mathcal{T}_{k-1}+1}^{\mathcal{T}_k} \mathbb{E}\left[\mu(\langle x_t^*, \theta^* \rangle) - \mu(\langle x_{t,a_t}, \theta^* \rangle) \middle| E' \right] \ .\end{aligned}$$

Throughout the remainder of the analysis, we condition on the good event E'. Let Regret_ℓ denote the cumulative expected regret incurred during interval ℓ. We analyze the regret by separating the discussion into four cases, namely $\ell = 1$, $\ell = 2$, $\ell = 3$, and the subsequent intervals with $\ell \ge 4$.

Case 1: $\ell \in \{1,2,3\}$**.** In the first three intervals, the number of rounds is given by $\left\lceil \frac{\sqrt[3]{T}}{\log_2 \log_2 T} \right\rceil$. By (14), each round incurs an instantaneous regret of at most R. Consequently, for the first interval we obtain

$$\text{Regret}_1 = \sum_{t=1}^{\mathcal{T}_1} \mathbb{E}[\mu(\langle x_t^*, \theta^* \rangle) - \mu(\langle x_{t,a_t}, \theta^* \rangle)] \le R\mathcal{T}_1 \le R\left(\frac{\sqrt[3]{T}}{\log_2 \log_2 T} + 1\right) = \mathcal{O}\left(\frac{R\sqrt[3]{T}}{\log \log T}\right) .$$

The same reasoning applies to the second and third intervals, which yields the same order of regret,

$$\text{Regret}_2 = \text{Regret}_3 = \mathcal{O}\left(\frac{R\sqrt[3]{T}}{\log \log T}\right) .$$

Case 2: $\ell \ge 4$**.** For each interval $\ell \ge 4$, the instantaneous regret can be controlled using the Mean Value Theorem. For some z lying between $\langle x_{t,a_t}, \theta^* \rangle$ and $\langle x_t^*, \theta^* \rangle$, we have

$$\begin{aligned}
\mu(\langle x_t^*, \theta^* \rangle) - \mu(\langle x_{t,a_t}, \theta^* \rangle) &= \dot{\mu}(z)\langle x_t^* - x_{t,a_t}, \theta^* \rangle \\
&= \dot{\mu}(z)(\langle x_t^* - x_{t,a_t}, \theta^* - \hat{\theta}_{\ell-1} \rangle + \langle x_t^* - x_{t,a_t}, \hat{\theta}_{\ell-1} \rangle) \\
&\le \dot{\mu}(z)(|\langle x_t^*, \hat{\theta}_{\ell-1} - \theta^* \rangle| + |\langle x_{t,a_t}, \hat{\theta}_{\ell-1} - \theta^* \rangle| + \langle x_t^{(\ell-1)} - x_{t,a_t}, \hat{\theta}_{\ell-1} \rangle) \\
&\le \underbrace{4\dot{\mu}(z)\varepsilon'_{t,\ell-1}(\lambda)}_{\triangleq A_t} . \qquad (15)
\end{aligned}$$

The first inequality uses the fact that $x_t^{(\ell-1)}$ is optimal with respect to $\hat{\theta}_{\ell-1}$, together with the guarantee from Lemma 16 that x_t^* belongs to $\mathcal{A}_t^{(\ell-2)}$. The final bound follows because $x_{t,a_t} \in \mathcal{A}_t^{(\ell-1)}$ by the arm elimination rule, combined with the definition of the good event E' and again Lemma 16, which ensures that $x_t^* \in \mathcal{A}_t^{(\ell-2)}$. As in the analysis of Algorithm 2, during the greedy selection step the regret can be bounded more tightly as $\mu(\langle x_t^*, \theta^* \rangle) - \mu(\langle x_{t,a_t}, \theta^* \rangle)$ by $2\dot{\mu}(z)\varepsilon'_{t,\ell-1}(\lambda)$, analogous to Equation (10). For simplicity in the subsequent analysis, we substitute the greedy selection step with uncertainty-driven exploration, so that in all intervals the algorithm may be analyzed under uncertainty-driven exploration alone. For rigor, interval end time $\mathcal{T}_\ell$ must be replaced by $\mathcal{T}'_{\ell-1} \coloneqq \mathcal{T}_{\ell-1} + \left\lceil \frac{cT^{1-2^{(4-\ell)}/3}}{\log_2 \log_2 T} \right\rceil$, as in the proof of Theorem 2, but this modification affects the regret bound only by a constant factor.

Bounding $\sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell} \mathbb{E}[A_t]$**.**

$$\begin{aligned}
\sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell} \mathbb{E}[A_t] &= 4 \sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell} \mathbb{E}[\dot{\mu}(z)\varepsilon'_{t,\ell-1}(\lambda)] \\
&\le 4 \sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell} \mathbb{E}\left[e^{R(\langle x_t^*, \theta^* \rangle - z)} \dot{\mu}(\langle x_t^*, \theta^* \rangle)\, \varepsilon'_{t,\ell-1}(\lambda)\right] \\
&\le 4 \sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell} \mathbb{E}\left[e^{R\langle x_t^* - x_{t,a_t}, \theta^* \rangle} \dot{\mu}(\langle x_t^*, \theta^* \rangle)\, \varepsilon'_{t,\ell-1}(\lambda)\right] \\
&\le 4 \sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell} \mathbb{E}\left[e^{4R(S \wedge \varepsilon'_{t,\ell-1}(\lambda))} \dot{\mu}(\langle x_t^*, \theta^* \rangle)\, \varepsilon'_{t,\ell-1}(\lambda)\right] \\
&\le 4 \sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell} \mathbb{E}\left[\underbrace{\dot{\mu}(\langle x_t^*, \theta^* \rangle)\, \varepsilon'_{t,\ell-1}(\lambda)}_{\triangleq B_t}\right] + \mathbb{E}\left[\underbrace{\frac{e^{4RS}(S \wedge \varepsilon'_{t,\ell-1}(\lambda))}{S} \dot{\mu}(\langle x_t^*, \theta^* \rangle)\, \varepsilon'_{t,\ell-1}(\lambda)}_{\triangleq C_t}\right] .
\end{aligned}$$

The first inequality is a direct application of Lemma 12. The second inequality holds because z lies between $\langle x_{t,a_t}, \theta^* \rangle$ and $\langle x_t^*, \theta^* \rangle$. The third inequality follows from Assumption 1 together with the bound on $\langle x_t^* - x_{t,a_t}, \theta^* \rangle$ derived in (15). Finally, the last inequality is obtained by invoking Lemma 11.

Bounding $\sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell} \mathbb{E}[B_t]$.

$$
\begin{aligned}
\sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell} \mathbb{E}[B_t] &= \beta(\lambda) \sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell} \mathbb{E}\left[\max_{y\in\mathcal{A}_t^{(\ell-2)}} \|\dot{\mu}(\langle x_t^*, \theta^*\rangle)y\|_{V_{\ell-1}^{-1}}\right] \\
&= \beta(\lambda)(\mathcal{T}_\ell - \mathcal{T}_{\ell-1})\mathbb{E}\left[\max_{y\in\mathcal{A}_t^{(\ell-2)}} \|\dot{\mu}(\langle x_t^*, \theta^*\rangle)y\|_{V_{\ell-1}^{-1}}\right] \qquad (16)\\
&= \frac{\beta(\lambda)(\mathcal{T}_\ell - \mathcal{T}_{\ell-1})}{\mathcal{T}_{\ell-1} - \mathcal{T}_{\ell-2}} \sum_{s=\mathcal{T}_{\ell-2}+1}^{\mathcal{T}_{\ell-1}} \mathbb{E}\left[\max_{y\in\mathcal{A}_t^{(\ell-2)}} \|\dot{\mu}(\langle x_t^*, \theta^*\rangle)y\|_{V_{\ell-1}^{-1}}\right] \\
&\le \frac{\beta(\lambda)(\mathcal{T}_\ell - \mathcal{T}_{\ell-1})}{\mathcal{T}_{\ell-1} - \mathcal{T}_{\ell-2}} \sum_{s=\mathcal{T}_{\ell-2}+1}^{\mathcal{T}_{\ell-1}} \mathbb{E}\left[\max_{y\in\mathcal{A}_t^{(\ell-2)}} \|\dot{\mu}(\langle x_t^*, \theta^*\rangle)y\|_{H_{s-1}^{-1}}\right] \\
&= \frac{\beta(\lambda)(\mathcal{T}_\ell - \mathcal{T}_{\ell-1})}{\mathcal{T}_{\ell-1} - \mathcal{T}_{\ell-2}} \sum_{s=\mathcal{T}_{\ell-2}+1}^{\mathcal{T}_{\ell-1}} \mathbb{E}\left[\mathbb{E}\left[\max_{y\in\mathcal{A}_t^{(\ell-2)}} \|\dot{\mu}(\langle x_t^*, \theta^*\rangle)y\|_{H_{s-1}^{-1}}\middle|\mathcal{F}_{s-1}\right]\right] \\
&= \frac{\beta(\lambda)(\mathcal{T}_\ell - \mathcal{T}_{\ell-1})}{\mathcal{T}_{\ell-1} - \mathcal{T}_{\ell-2}} \sum_{s=\mathcal{T}_{\ell-2}+1}^{\mathcal{T}_{\ell-1}} \mathbb{E}\left[\mathbb{E}\left[\max_{y\in\mathcal{A}_s^{(\ell-2)}} \|\dot{\mu}(\langle x_s^*, \theta^*\rangle)y\|_{H_{s-1}^{-1}}\middle|\mathcal{F}_{s-1}\right]\right] \\
&= \frac{\beta(\lambda)(\mathcal{T}_\ell - \mathcal{T}_{\ell-1})}{\mathcal{T}_{\ell-1} - \mathcal{T}_{\ell-2}} \sum_{s=\mathcal{T}_{\ell-2}+1}^{\mathcal{T}_{\ell-1}} \mathbb{E}\left[\max_{y\in\mathcal{A}_s^{(\ell-2)}} \|\dot{\mu}(\langle x_s^*, \theta^*\rangle)y\|_{H_{s-1}^{-1}}\right] \\
&= \frac{\beta(\lambda)(\mathcal{T}_\ell - \mathcal{T}_{\ell-1})}{\mathcal{T}_{\ell-1} - \mathcal{T}_{\ell-2}} \sum_{s=\mathcal{T}_{\ell-2}+1}^{\mathcal{T}_{\ell-1}} \mathbb{E}\Big[\underbrace{\|\dot{\mu}(\langle x_s^*, \theta^*\rangle)x_{s,a_s}\|_{H_{s-1}^{-1}}}_{\triangleq D_s}\Big] \ .
\end{aligned}
$$

The second equality mirrors the reasoning in (5), since both $V_{\ell-1}$ and the arm elimination rule—determined by $\hat{\theta}_1, \ldots, \hat{\theta}_{\ell-2}$—are measurable with respect to $\mathcal{F}_{\mathcal{T}_{\ell-1}}$, and can therefore be regarded as fixed conditional on this filtration. Given that the contexts are drawn independently and identically, their values are equal. The first inequality follows from the monotonicity, as $H_{s-1} \preceq V_{\ell-1}$ for all s in the interval. The fourth and sixth equalities use the tower property. The fifth equality relies on the fact that, conditional on $\mathcal{F}_{s-1}$, both H_{s-1} and the arm elimination rule (determined by $\hat{\theta}_1, \ldots, \hat{\theta}_{\ell-2}$) are fixed, while the distribution of the contexts remains unchanged. Finally, the last equality follows from the arm-selection strategy, since the factor $\dot{\mu}(\langle x_s^*, \theta^*\rangle)$ does not affect the maximization and can thus be pulled outside without altering the $\arg\max$.

Bounding $\sum_{s=\mathcal{T}_{\ell-2}+1}^{\mathcal{T}_{\ell-1}} \mathbb{E}[D_s]$.

$$
\begin{aligned}
\sum_{s=\mathcal{T}_{\ell-2}+1}^{\mathcal{T}_{\ell-1}} \mathbb{E}[D_s] &\le \sum_{s=\mathcal{T}_{\ell-2}+1}^{\mathcal{T}_{\ell-1}} \mathbb{E}\left[\left\| e^{\frac{R}{2}|\langle x_s^*, \theta^*\rangle - \langle x_{s,a_s}, \hat{\theta}_{\ell-2}\rangle|} \sqrt{\dot{\mu}(\langle x_s^*, \theta^*\rangle)\dot{\mu}(\langle x_{s,a_s}, \hat{\theta}_{\ell-2}\rangle)}\, x_{s,a_s}\right\|_{H_{s-1}^{-1}}\right] \\
&\le \sum_{s=\mathcal{T}_{\ell-2}+1}^{\mathcal{T}_{\ell-1}} \mathbb{E}\left[\left\| e^{\frac{R}{2}(2S\wedge 5\varepsilon'_{s,\ell-2}(\lambda))} \sqrt{\dot{\mu}(\langle x_s^*, \theta^*\rangle)\dot{\mu}(\langle x_{s,a_s}, \hat{\theta}_{\ell-2}\rangle)}\, x_{s,a_s}\right\|_{H_{s-1}^{-1}}\right] \\
&\le \sum_{s=\mathcal{T}_{\ell-2}+1}^{\mathcal{T}_{\ell-1}} \mathbb{E}\left[\left\| e^{\frac{R}{2}(2S\wedge 5\varepsilon'_{s,\ell-2}(\lambda)) + \frac{R}{2}(2S\wedge \varepsilon'_{s,\ell-2}(\lambda))} \sqrt{\dot{\mu}(\langle x_s^*, \theta^*\rangle)\alpha_{s,\ell-2}(\lambda)\dot{\mu}(\langle x_{s,a_s}, \hat{\theta}_{\ell-2}\rangle)}\, x_{s,a_s}\right\|_{H_{s-1}^{-1}}\right] \\
&\le \sum_{s=\mathcal{T}_{\ell-2}+1}^{\mathcal{T}_{\ell-1}} \mathbb{E}\left[\left\| e^{3R(S\wedge \varepsilon'_{s,\ell-2}(\lambda))} \sqrt{\dot{\mu}(\langle x_s^*, \theta^*\rangle)\alpha_{s,\ell-2}(\lambda)\dot{\mu}(\langle x_{s,a_s}, \hat{\theta}_{\ell-2}\rangle)}\, x_{s,a_s}\right\|_{H_{s-1}^{-1}}\right]
\end{aligned}
$$

$$\leq \sum_{s=\mathcal{T}_{\ell-2}+1}^{\mathcal{T}_{\ell-1}} \mathbb{E}\Bigg[\underbrace{\left\|\sqrt{\dot{\mu}(\langle x_s^*,\theta^*\rangle)\alpha_{s,\ell-2}(\lambda)\dot{\mu}(\langle x_{s,a_s},\hat{\theta}_{\ell-2}\rangle)}x_{s,a_s}\right\|_{H_{s-1}^{-1}}}_{\triangleq E_s}\Bigg] + \mathbb{E}\Bigg[\underbrace{\left\|\frac{e^{3RS}(S\wedge\varepsilon'_{s,\ell-2}(\lambda))}{S}\sqrt{\dot{\mu}(\langle x_s^*,\theta^*\rangle)\alpha_{s,\ell-2}(\lambda)\dot{\mu}(\langle x_{s,a_s},\hat{\theta}_{\ell-2}\rangle)}x_{s,a_s}\right\|_{H_{s-1}^{-1}}}_{\triangleq F_s}\Bigg].$$

The first inequality is obtained by applying Lemma 12 to compare $\dot{\mu}(\langle x_s^*,\theta^*\rangle)$ and $\dot{\mu}(\langle x_{s,a_s},\hat{\theta}_{\ell-2}\rangle)$. The second inequality follows from Assumption 3 together with the decomposition $|\langle x_s^*-x_{s,a_s},\theta^*\rangle|+|\langle x_{s,a_s},\theta^*-\hat{\theta}_{\ell-2}\rangle| \leq 4\varepsilon'_{s,\ell-2}(\lambda)+\varepsilon'_{s,\ell-2}(\lambda)$, as implied by (15) and the definition of the good event E'. The third inequality directly follows from the definition of $\alpha_{s,\ell-2}(\lambda)$ when $\ell \geq 4$. The final bound is obtained by invoking Lemma 11.

Bounding $\sum_{s=\mathcal{T}_{\ell-2}+1}^{\mathcal{T}_{\ell-1}}\mathbb{E}[E_s]$.

$$\begin{aligned}
\sum_{s=\mathcal{T}_{\ell-2}+1}^{\mathcal{T}_{\ell-1}}\mathbb{E}[E_s] &\leq \sum_{s=\mathcal{T}_{\ell-2}+1}^{\mathcal{T}_{\ell-1}}\sqrt{\mathbb{E}\left[\dot{\mu}(\langle x_s^*,\theta^*\rangle)\right]}\sqrt{\mathbb{E}\left[\left\|\sqrt{\alpha_{s,\ell-2}(\lambda)\dot{\mu}(\langle x_{s,a_s},\hat{\theta}_{\ell-2}\rangle)}x_{s,a_s}\right\|_{H_{s-1}^{-1}}^2\right]}\\
&\leq \frac{\sqrt{\mathcal{T}_{\ell-1}-\mathcal{T}_{\ell-2}}}{\sqrt{\hat{\kappa}}}\sqrt{\sum_{s=\mathcal{T}_{\ell-2}+1}^{\mathcal{T}_{\ell-1}}\mathbb{E}\left[\left\|\sqrt{\alpha_{s,\ell-2}(\lambda)\dot{\mu}(\langle x_{s,a_s},\hat{\theta}_{\ell-2}\rangle)}x_{s,a_s}\right\|_{H_{s-1}^{-1}}^2\right]}\\
&\leq \frac{\sqrt{\mathcal{T}_{\ell-1}-\mathcal{T}_{\ell-2}}}{\sqrt{\hat{\kappa}}}\sqrt{2\log\left(\frac{\det(H_{\mathcal{T}_{\ell-1}})}{\det(\lambda I)}\right)}\\
&\leq \frac{\sqrt{2d(\mathcal{T}_{\ell-1}-\mathcal{T}_{\ell-2})\log(2T)}}{\sqrt{\hat{\kappa}}}.
\end{aligned}$$

The first two inequalities are obtained by successive applications of the Cauchy-Schwarz inequality. The third inequality follows directly from Corollary 2. Finally, the last inequality is derived using the same reasoning as in (13), where the determinant is upper bounded by a trace argument, yielding a logarithmic dependence on T.

Bounding $\sum_{s=\mathcal{T}_{\ell-2}+1}^{\mathcal{T}_{\ell-1}}\mathbb{E}[F_s]$.

$$\begin{aligned}
\sum_{s=\mathcal{T}_{\ell-2}+1}^{\mathcal{T}_{\ell-1}}\mathbb{E}[F_s] &\leq \frac{e^{3RS}}{S}\sum_{s=\mathcal{T}_{\ell-2}+1}^{\mathcal{T}_{\ell-1}}\mathbb{E}\left[\varepsilon_{s,\ell-2}(\lambda)\left\|\sqrt{\dot{\mu}(\langle x_s^*,\theta^*\rangle)\alpha_{s,\ell-2}(\lambda)\dot{\mu}(\langle x_{s,a_s},\hat{\theta}_{\ell-2}\rangle)}x_{s,a_s}\right\|_{H_{s-1}^{-1}}\right]\\
&= \frac{e^{3RS}\beta(\lambda)}{S}\sum_{s=\mathcal{T}_{\ell-2}+1}^{\mathcal{T}_{\ell-1}}\mathbb{E}\left[\max_{y\in\mathcal{A}_s^{(\ell-3)}}\|y\|_{V_{\ell-2}^{-1}}\left\|\sqrt{\dot{\mu}(\langle x_s^*,\theta^*\rangle)\alpha_{s,\ell-2}(\lambda)\dot{\mu}(\langle x_{s,a_s},\hat{\theta}_{\ell-2}\rangle)}x_{s,a_s}\right\|_{H_{s-1}^{-1}}\right]\\
&\leq \frac{e^{3RS}\beta(\lambda)}{S}\sum_{s=\mathcal{T}_{\ell-2}+1}^{\mathcal{T}_{\ell-1}}\sqrt{\mathbb{E}\left[\max_{y\in\mathcal{A}_s^{(\ell-3)}}\|\sqrt{\dot{\mu}(\langle x_s^*,\theta^*\rangle)}y\|_{V_{\ell-2}^{-1}}^2\right]\mathbb{E}\left[\left\|\sqrt{\alpha_{s,\ell-2}(\lambda)\dot{\mu}(\langle x_{s,a_s},\hat{\theta}_{\ell-2}\rangle)}x_{s,a_s}\right\|_{H_{s-1}^{-1}}^2\right]}\\
&\leq \frac{e^{3RS}\beta(\lambda)\sqrt{2d(\mathcal{T}_{\ell-1}-\mathcal{T}_{\ell-2})\log(2T)}}{S}\sqrt{\underbrace{\mathbb{E}\left[\max_{y\in\mathcal{A}_s^{(\ell-3)}}\|\sqrt{\dot{\mu}(\langle x_s^*,\theta^*\rangle)}y\|_{V_{\ell-2}^{-1}}^2\right]}_{\triangleq G_s}}.
\end{aligned}$$

The second inequality follows from an application of the Cauchy-Schwarz inequality. The final inequality uses the fact that the expectation $\mathbb{E}\Big[\max_{y\in\mathcal{A}_s^{(\ell-3)}}\|\sqrt{\dot{\mu}(\langle x_s^*,\theta^*\rangle)}y\|_{V_{\ell-2}^{-1}}^2\Big]$ takes the same value for all $s\in[\mathcal{T}_{\ell-2}+1,\mathcal{T}_{\ell-1}]$, as established in (16), together with Corollary 2, which bounds the quadratic form by a log-determinant expression.

Bounding $\mathbb{E}[G_s]$ for $s \in [\mathcal{T}_{\ell-2}+1, \mathcal{T}_{\ell-1}]$.

$$\begin{aligned}
\mathbb{E}[G_s] &= \frac{1}{\mathcal{T}_{\ell-2}-\mathcal{T}_{\ell-3}} \sum_{u=\mathcal{T}_{\ell-3}+1}^{\mathcal{T}_{\ell-2}} \mathbb{E}\left[\max_{y\in\mathcal{A}_s^{(\ell-3)}} \|\sqrt{\dot\mu(\langle x_s^*,\theta^*\rangle)}y\|^2_{V_{\ell-2}^{-1}}\right] \\
&\le \frac{1}{\mathcal{T}_{\ell-2}-\mathcal{T}_{\ell-3}} \sum_{u=\mathcal{T}_{\ell-3}+1}^{\mathcal{T}_{\ell-2}} \mathbb{E}\left[\max_{y\in\mathcal{A}_s^{(\ell-3)}} \|\sqrt{\dot\mu(\langle x_s^*,\theta^*\rangle)}y\|^2_{H_{u-1}^{-1}}\right] \\
&= \frac{1}{\mathcal{T}_{\ell-2}-\mathcal{T}_{\ell-3}} \sum_{u=\mathcal{T}_{\ell-3}+1}^{\mathcal{T}_{\ell-2}} \mathbb{E}\left[\mathbb{E}\left[\max_{y\in\mathcal{A}_s^{(\ell-3)}} \|\sqrt{\dot\mu(\langle x_s^*,\theta^*\rangle)}y\|^2_{H_{u-1}^{-1}}\middle|\mathcal{F}_{u-1}\right]\right] \\
&= \frac{1}{\mathcal{T}_{\ell-2}-\mathcal{T}_{\ell-3}} \sum_{u=\mathcal{T}_{\ell-3}+1}^{\mathcal{T}_{\ell-2}} \mathbb{E}\left[\mathbb{E}\left[\max_{y\in\mathcal{A}_u^{(\ell-3)}} \|\sqrt{\dot\mu(\langle x_u^*,\theta^*\rangle)}y\|^2_{H_{u-1}^{-1}}\middle|\mathcal{F}_{u-1}\right]\right] \\
&= \frac{1}{\mathcal{T}_{\ell-2}-\mathcal{T}_{\ell-3}} \sum_{u=\mathcal{T}_{\ell-3}+1}^{\mathcal{T}_{\ell-2}} \mathbb{E}\left[\max_{y\in\mathcal{A}_u^{(\ell-3)}} \|\sqrt{\dot\mu(\langle x_u^*,\theta^*\rangle)}y\|^2_{H_{u-1}^{-1}}\right] \\
&= \frac{1}{\mathcal{T}_{\ell-2}-\mathcal{T}_{\ell-3}} \sum_{u=\mathcal{T}_{\ell-3}+1}^{\mathcal{T}_{\ell-2}} \mathbb{E}\Big[\underbrace{\|\sqrt{\dot\mu(\langle x_u^*,\theta^*\rangle)}x_{u,a_u}\|^2_{H_{u-1}^{-1}}}_{\triangleq I_u}\Big] \ .
\end{aligned}$$

The first inequality follows from the monotonicity of the matrices, since $H_{u-1} \preceq V_{\ell-2}$ for all u in the summation range. The second and fourth equalities are applications of the tower property. The third equality holds because, conditional on $\mathcal{F}_{u-1}$, both H_{u-1} and the arm elimination rule (determined by $\hat\theta_1,\dots,\hat\theta_{\ell-3}$) are fixed, while the distribution of the contexts remains unchanged. Finally, the last equality follows from the arm-selection strategy: the multiplicative factor $\dot\mu(\langle x_u^*,\theta^*\rangle)$ does not affect the maximization and can therefore be factored out without altering the $\arg\max$.

Bounding $\sum_{u=\mathcal{T}_{\ell-3}+1}^{\mathcal{T}_{\ell-2}} \mathbb{E}[I_u]$.

$$\begin{aligned}
\sum_{u=\mathcal{T}_{\ell-3}+1}^{\mathcal{T}_{\ell-2}} \mathbb{E}[I_u] &\le \sum_{u=\mathcal{T}_{\ell-3}+1}^{\mathcal{T}_{\ell-2}} \mathbb{E}\left[e^{R|\langle x_u^*,\theta^*\rangle - \langle x_{u,a_u},\hat\theta_{\ell-3}\rangle|}\left\|\sqrt{\dot\mu(\langle x_{u,a_u},\hat\theta_{\ell-3}\rangle)}x_{u,a_u}\right\|^2_{H_{u-1}^{-1}}\right] \\
&\le \sum_{u=\mathcal{T}_{\ell-3}+1}^{\mathcal{T}_{\ell-2}} \mathbb{E}\left[e^{4RS}\left\|\sqrt{\alpha_{u,\ell-3}(\lambda)\dot\mu(\langle x_{u,a_u},\hat\theta_{\ell-3}\rangle)}x_{u,a_u}\right\|^2_{H_{u-1}^{-1}}\right] \\
&\le 2e^{4RS} d\log(2T) \ .
\end{aligned}$$

The first inequality is obtained by applying Lemma 12, which allows us to compare $\dot\mu(\langle x_u^*,\theta^*\rangle)$ and $\dot\mu(\langle x_{u,a_u},\hat\theta_{\ell-3}\rangle)$. The second inequality uses Assumption 1, together with the fact that $\alpha_{u,\ell-3}(\lambda) \ge e^{-2RS}$ for all $\ell \ge 4$. Finally, the last inequality follows from Corollary 2, which bounds the quadratic form in terms of a log-determinant expression and yields the stated order.

We now combine the previous bounds step by step. First, from the result on $\mathbb{E}[G_s]$, we obtain

$$\mathbb{E}[G_s] \le \frac{1}{\mathcal{T}_{\ell-2}-\mathcal{T}_{\ell-3}} \sum_{u=\mathcal{T}_{\ell-3}+1}^{\mathcal{T}_{\ell-2}} \mathbb{E}[I_u] \le \frac{2e^{4RS} d\log(2T)}{\mathcal{T}_{\ell-2}-\mathcal{T}_{\ell-3}} \ .$$

Next, substituting this into the bound for $\sum_{s=\mathcal{T}_{\ell-2}+1}^{\mathcal{T}_{\ell-1}} \mathbb{E}[F_s]$, we obtain

$$\sum_{s=\mathcal{T}_{\ell-2}+1}^{\mathcal{T}_{\ell-1}} \mathbb{E}[F_s] \le \frac{e^{3RS}\beta(\lambda)\sqrt{2d(\mathcal{T}_{\ell-1}-\mathcal{T}_{\ell-2})\log(2T)}}{S} \sqrt{\mathbb{E}\left[G_s\right]}$$

$$\leq \frac{2e^{5RS}\beta(\lambda)d\log(2T)\sqrt{\mathcal{T}_{\ell-1}-\mathcal{T}_{\ell-2}}}{S\sqrt{\mathcal{T}_{\ell-2}-\mathcal{T}_{\ell-3}}}\,.$$

Furthermore, combining this with the result on $\sum_{s=\mathcal{T}_{\ell-2}+1}^{\mathcal{T}_{\ell-1}}\mathbb{E}[E_s]$, we can bound $\sum_{s=\mathcal{T}_{\ell-2}+1}^{\mathcal{T}_{\ell-1}}\mathbb{E}[D_s]$ as

$$\begin{aligned}\sum_{s=\mathcal{T}_{\ell-2}+1}^{\mathcal{T}_{\ell-1}}\mathbb{E}[D_s] &\leq \sum_{s=\mathcal{T}_{\ell-2}+1}^{\mathcal{T}_{\ell-1}}\mathbb{E}[E_s]+\sum_{s=\mathcal{T}_{\ell-2}+1}^{\mathcal{T}_{\ell-1}}\mathbb{E}[F_s]\\ &\leq \frac{\sqrt{2d(\mathcal{T}_{\ell-1}-\mathcal{T}_{\ell-2})\log(2T)}}{\sqrt{\hat{\kappa}}}+\frac{2e^{5RS}\beta(\lambda)d\log(2T)\sqrt{\mathcal{T}_{\ell-1}-\mathcal{T}_{\ell-2}}}{S\sqrt{\mathcal{T}_{\ell-2}-\mathcal{T}_{\ell-3}}}\,.\end{aligned}$$

Finally, substituting this bound into the expression for $\sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_{\ell}}\mathbb{E}[B_t]$ yields

$$\begin{aligned}\sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_{\ell}}\mathbb{E}[B_t] &\leq \frac{\beta(\lambda)(\mathcal{T}_{\ell}-\mathcal{T}_{\ell-1})}{\mathcal{T}_{\ell-1}-\mathcal{T}_{\ell-2}}\sum_{s=\mathcal{T}_{\ell-2}+1}^{\mathcal{T}_{\ell-1}}\mathbb{E}[D_s]\\ &\leq \frac{\beta(\lambda)(\mathcal{T}_{\ell}-\mathcal{T}_{\ell-1})}{\sqrt{\mathcal{T}_{\ell-1}-\mathcal{T}_{\ell-2}}}\left(\frac{\sqrt{2d\log(2T)}}{\sqrt{\hat{\kappa}}}+\frac{2e^{5RS}\beta(\lambda)d\log(2T)}{S\sqrt{\mathcal{T}_{\ell-2}-\mathcal{T}_{\ell-3}}}\right)\,.\end{aligned}$$

Bounding $\sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_{\ell}}\mathbb{E}[C_t]$.

$$\begin{aligned}\sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_{\ell}}\mathbb{E}[C_t] &\leq \frac{e^{4RS}}{S}\sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_{\ell}}\mathbb{E}\left[\dot{\mu}(\langle x_t^*,\theta^*\rangle)\,\varepsilon'^{2}_{t,\ell-1}(\lambda)\right]\\ &= \frac{e^{4RS}\beta(\lambda)^2}{S}\sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_{\ell}}\mathbb{E}\left[\max_{z\in\mathcal{A}_t^{(\ell-2)}}\|\sqrt{\dot{\mu}(\langle x_t^*,\theta^*\rangle)}z\|^2_{V_{\ell-1}^{-1}}\right]\\ &= \frac{e^{4RS}\beta(\lambda)^2(\mathcal{T}_{\ell}-\mathcal{T}_{\ell-1})}{S}\mathbb{E}\left[\max_{z\in\mathcal{A}_t^{(\ell-2)}}\|\sqrt{\dot{\mu}(\langle x_t^*,\theta^*\rangle)}z\|^2_{V_{\ell-1}^{-1}}\right]\\ &= \frac{e^{4RS}\beta(\lambda)^2(\mathcal{T}_{\ell}-\mathcal{T}_{\ell-1})}{S(\mathcal{T}_{\ell-1}-\mathcal{T}_{\ell-2})}\sum_{s=\mathcal{T}_{\ell-2}+1}^{\mathcal{T}_{\ell-1}}\mathbb{E}\left[\max_{z\in\mathcal{A}_t^{(\ell-2)}}\|\sqrt{\dot{\mu}(\langle x_t^*,\theta^*\rangle)}z\|^2_{V_{\ell-1}^{-1}}\right]\\ &\leq \frac{e^{4RS}\beta(\lambda)^2(\mathcal{T}_{\ell}-\mathcal{T}_{\ell-1})}{S(\mathcal{T}_{\ell-1}-\mathcal{T}_{\ell-2})}\sum_{s=\mathcal{T}_{\ell-2}+1}^{\mathcal{T}_{\ell-1}}\mathbb{E}\left[\max_{z\in\mathcal{A}_t^{(\ell-2)}}\|\sqrt{\dot{\mu}(\langle x_t^*,\theta^*\rangle)}z\|^2_{H_{s-1}^{-1}}\right]\\ &= \frac{e^{4RS}\beta(\lambda)^2(\mathcal{T}_{\ell}-\mathcal{T}_{\ell-1})}{S(\mathcal{T}_{\ell-1}-\mathcal{T}_{\ell-2})}\sum_{s=\mathcal{T}_{\ell-2}+1}^{\mathcal{T}_{\ell-1}}\mathbb{E}\left[\mathbb{E}\left[\max_{z\in\mathcal{A}_t^{(\ell-2)}}\|\sqrt{\dot{\mu}(\langle x_t^*,\theta^*\rangle)}z\|^2_{H_{s-1}^{-1}}\middle|\mathcal{F}_{s-1}\right]\right]\\ &= \frac{e^{4RS}\beta(\lambda)^2(\mathcal{T}_{\ell}-\mathcal{T}_{\ell-1})}{S(\mathcal{T}_{\ell-1}-\mathcal{T}_{\ell-2})}\sum_{s=\mathcal{T}_{\ell-2}+1}^{\mathcal{T}_{\ell-1}}\mathbb{E}\left[\mathbb{E}\left[\max_{z\in\mathcal{A}_s^{(\ell-2)}}\|\sqrt{\dot{\mu}(\langle x_s^*,\theta^*\rangle)}z\|^2_{H_{s-1}^{-1}}\middle|\mathcal{F}_{s-1}\right]\right]\\ &= \frac{e^{4RS}\beta(\lambda)^2(\mathcal{T}_{\ell}-\mathcal{T}_{\ell-1})}{S(\mathcal{T}_{\ell-1}-\mathcal{T}_{\ell-2})}\sum_{s=\mathcal{T}_{\ell-2}+1}^{\mathcal{T}_{\ell-1}}\mathbb{E}\left[\max_{z\in\mathcal{A}_s^{(\ell-2)}}\|\sqrt{\dot{\mu}(\langle x_s^*,\theta^*\rangle)}z\|^2_{H_{s-1}^{-1}}\right]\\ &= \frac{e^{4RS}\beta(\lambda)^2(\mathcal{T}_{\ell}-\mathcal{T}_{\ell-1})}{S(\mathcal{T}_{\ell-1}-\mathcal{T}_{\ell-2})}\sum_{s=\mathcal{T}_{\ell-2}+1}^{\mathcal{T}_{\ell-1}}\mathbb{E}\left[\|\sqrt{\dot{\mu}(\langle x_s^*,\theta^*\rangle)}x_{s,a_s}\|^2_{H_{s-1}^{-1}}\right]\\ &= \frac{e^{4RS}\beta(\lambda)^2(\mathcal{T}_{\ell}-\mathcal{T}_{\ell-1})}{S(\mathcal{T}_{\ell-1}-\mathcal{T}_{\ell-2})}\sum_{s=\mathcal{T}_{\ell-2}+1}^{\mathcal{T}_{\ell-1}}\mathbb{E}\left[I_s\right]\end{aligned}$$

$$\leq \frac{2e^{8RS}\beta(\lambda)^2(\mathcal{T}_\ell - \mathcal{T}_{\ell-1})d\log(2T)}{S(\mathcal{T}_{\ell-1} - \mathcal{T}_{\ell-2})}\,.$$

The second equality follows from the same reasoning as in (5), since both $V_{\ell-1}$ and the arm elimination rule—determined by $\hat{\theta}_1, \ldots, \hat{\theta}_{\ell-2}$—are measurable with respect to $\mathcal{F}_{\mathcal{T}_{\ell-1}}$ and can therefore be treated as fixed conditional on this filtration; given that the contexts are drawn independently and identically, their values coincide. The second inequality uses the monotonicity of the matrices, as $H_{s-1} \preceq V_{\ell-1}$ for all s in the interval. The fourth and sixth equalities apply the tower property. The fifth equality holds because, conditional on $\mathcal{F}_{s-1}$, both H_{s-1} and the arm elimination rule (determined by $\hat{\theta}_1, \ldots, \hat{\theta}_{\ell-2}$) are fixed, while the distribution of the contexts remains unchanged. The seventh equality is justified by the arm-selection strategy, as the multiplicative factor $\dot{\mu}(\langle x_s^*, \theta^*\rangle)$ does not affect the maximization and hence does not alter the $\arg\max$. The final inequality follows from the previously established bound on $\sum_{s=\mathcal{T}_{\ell-2}+1}^{\mathcal{T}_{\ell-1}} \mathbb{E}[I_s]$.

Combining the previously derived bounds for $\sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell} \mathbb{E}[B_t]$ and $\sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell} \mathbb{E}[C_t]$, we obtain

$$\begin{aligned}\sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell} \mathbb{E}[A_t] &\leq 4 \sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell} \mathbb{E}[B_t] + 4 \sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell} \mathbb{E}[C_t] \\ &\leq \frac{4\beta(\lambda)(\mathcal{T}_\ell - \mathcal{T}_{\ell-1})}{\sqrt{\mathcal{T}_{\ell-1} - \mathcal{T}_{\ell-2}}} \left(\frac{\sqrt{2d\log(2T)}}{\sqrt{\hat{\kappa}}} + \frac{2e^{5RS}\beta(\lambda)d\log(2T)}{S\sqrt{\mathcal{T}_{\ell-2} - \mathcal{T}_{\ell-3}}} \right) + \frac{8e^{8RS}\beta(\lambda)^2(\mathcal{T}_\ell - \mathcal{T}_{\ell-1})d\log(2T)}{S(\mathcal{T}_{\ell-1} - \mathcal{T}_{\ell-2})}\,.\end{aligned}$$

For the case $\ell = 4$, this simplifies to

$$\begin{aligned}\sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell} \mathbb{E}[A_t] &= \mathcal{O}\left(\frac{RS\sqrt{d+\log T} \cdot \frac{T^{\frac{2}{3}}}{\log\log T}}{\sqrt{\frac{\sqrt[3]{T}}{\log\log T}}} \left(\frac{\sqrt{d\log T}}{\sqrt{\hat{\kappa}}} + \frac{Re^{5RS}d\log T\sqrt{d+\log T}}{\sqrt{\frac{\sqrt[3]{T}}{\log\log T}}} \right) \right) \\ &\quad + \mathcal{O}\left(\frac{R^2Se^{8RS} \cdot \frac{dT^{\frac{2}{3}}(d+\log T)\log T}{\log\log T}}{\frac{\sqrt[3]{T}}{\log\log T}} \right) \\ &= \mathcal{O}\left(\frac{RS\sqrt{d(d+\log T)T\log T}}{\sqrt{\hat{\kappa}}\log\log T} + R^2Se^{8RS}d(d+\log T)T^{\frac{1}{3}}\log T \right)\,.\end{aligned}$$

For all subsequent intervals with $\ell \geq 5$, we similarly obtain

$$\begin{aligned}\sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell} \mathbb{E}[A_t] &= \mathcal{O}\left(\frac{RS\sqrt{d+\log T} \cdot \frac{T^{1-\frac{1}{3\cdot 2^{\ell-4}}}}{\log\log T}}{\sqrt{\frac{T^{1-\frac{1}{3\cdot 2^{\ell-5}}}}{\log\log T}}} \left(\frac{\sqrt{d\log T}}{\sqrt{\hat{\kappa}}} + \frac{Re^{5RS}d\log T\sqrt{d+\log T}}{\sqrt{\frac{T^{1-\frac{1}{3\cdot 2^{\ell-6}}}}{\log\log T}}} \right) \right) \\ &\quad + \mathcal{O}\left(\frac{R^2Se^{8RS} \cdot \frac{dT^{1-\frac{1}{3\cdot 2^{\ell-4}}}(d+\log T)\log T}{\log\log T}}{\frac{T^{1-\frac{1}{3\cdot 2^{\ell-5}}}}{\log\log T}} \right) \\ &= \mathcal{O}\left(\frac{RS\sqrt{d(d+\log T)T\log T}}{\sqrt{\hat{\kappa}}\log\log T} + R^2Se^{8RS}d(d+\log T)T^{\frac{1}{3\cdot 2^{\ell-5}}}\log T \right) \\ &= \mathcal{O}\left(\frac{RS\sqrt{d(d+\log T)T\log T}}{\sqrt{\hat{\kappa}}\log\log T} + R^2Se^{8RS}d(d+\log T)T^{\frac{1}{3}}\log T \right)\,.\end{aligned}$$

Therefore, the total worst-case regret is bounded as

$$\begin{aligned}\mathcal{R}(T) &= \sum_{\ell=1}^{3} \text{Regret}_\ell + \sum_{\ell=4}^{B} \text{Regret}_\ell \\ &= \mathcal{O}\left(\frac{RT^{\frac{1}{3}}}{\log\log T} \right) + \sum_{\ell=4}^{B} \sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell} \mathbb{E}[\mu(\langle x_t^*, \theta^*\rangle) - \mu(\langle x_{t,a_t}, \theta^*\rangle)]\end{aligned}$$

$$
\begin{aligned}
&\leq \mathcal{O}\left(\frac{RT^{\frac{1}{3}}}{\log\log T}\right) + \sum_{\ell=4}^{B} \sum_{t=\mathcal{T}_{\ell-1}+1}^{\mathcal{T}_\ell} \mathbb{E}[A_t] \\
&= \mathcal{O}\left(RS\sqrt{\frac{d(d+\log T)T\log T\log\log T}{\hat{\kappa}}} + \left(R^2Se^{8RS}d(d+\log T)\log T\log\log T + \frac{R}{\log\log T}\right)T^{\frac{1}{3}}\right) \\
&= \widetilde{\mathcal{O}}\left(\frac{dRS\sqrt{T}}{\sqrt{\hat{\kappa}}} + (R^2Se^{8RS}d^2 + R)T^{\frac{1}{3}}\right) .
\end{aligned}
$$

□

D. Additional Experiments

D.1. Linear Contextual Bandits: Experimental Results with Normal Contexts

We evaluate the performance of BLCE-G and BLCE by measuring the cumulative regret over $T = 10{,}000$ rounds. At each iteration, K arms are independently sampled from a d-dimensional normal distribution, and the parameter vector θ^* is drawn from a d-dimensional normal distribution. Each experiment is repeated 10 times. We consider $(K, d) \in \{(1000, 5), (5000, 10), (50, 20), (100, 30)\}$, where the first two pairs represent the large-K regime and the latter two correspond to the small-K regime.

For comparison, we benchmark against state-of-the-art algorithms: *Rarely Switching OFUL* (RS-OFUL; Abbasi-Yadkori et al. 2011), BatchLinUCB-DG (Ruan et al. 2021), *Efficient Batched Algorithm for linear contextual Bandits* (SoftBatch; Hanna et al. 2023b), and BatchLearning (Zhang et al. 2025). The within-interval allocation rate for BLCE-G and BLCE is set to $c = 0.5$. For RS-OFUL, the switching parameter is $C = 3$, and for SoftBatch, the discretization parameter is $q = 1/(8\sqrt{d})$. Algorithms of Hanna et al. (2023b) incur substantial computational overhead, as reflected in their time complexity reported in Table 1; we therefore omit their regret plots. Importantly, BLCE-G and BLCE are implemented with the exact theoretical hyperparameters specified in our main results, without additional tuning.

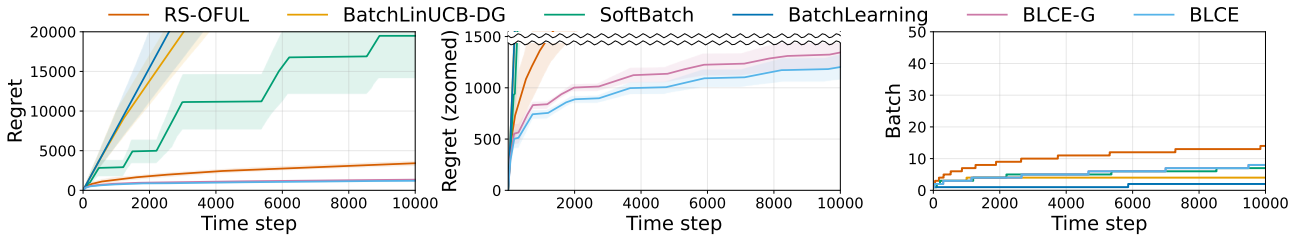
We report three types of figures. First, the average cumulative regret (solid line) together with its standard deviation (shaded region) over 10 runs. Second, zoomed-in views of the regret curves to highlight the differences between BLCE-G and BLCE. Third, the average update complexity across 10 runs, illustrating the frequency of parameter updates (for consistency with prior work, we label the horizontal axis using the term "batch").

As illustrated in Figure 2, both BLCE-G and BLCE consistently outperform all baselines in both the large-K and small-K regimes, achieving lowest regret and exhibiting greater stability. These results confirm that BLCE-G and BLCE not only attain the tightest theoretical guarantees but also deliver strong empirical performance, thereby fulfilling their design objectives of near-optimal regret and minimal interval complexity. Furthermore, runtime comparisons in Table 4 show that BLCE-G and BLCE incur substantially lower computation cost than other theoretically optimal algorithms. In particular, BLCE, which eliminates reliance on G-optimal design, attains the fastest runtime—comparable even to suboptimal algorithms.

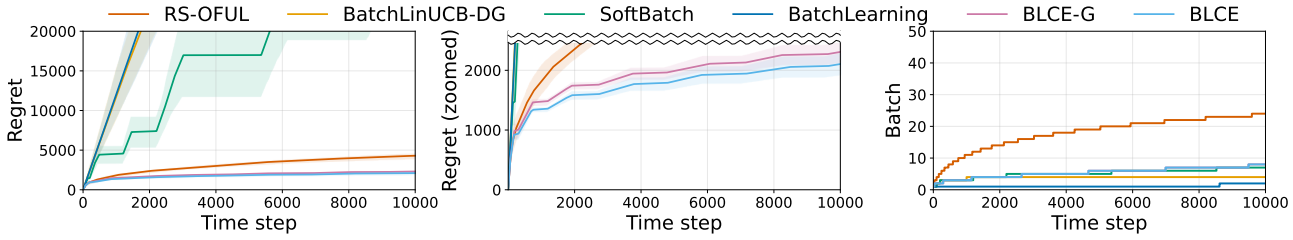
Overall, these experiments demonstrate a distinctive advantage of our approach: BLCE-G and BLCE combine minimax-optimal regret guarantees with practical efficiency. This dual benefit of theoretical optimality and empirical superiority sets them apart from all prior methods for linear contextual bandits.

Table 4. Average runtime (seconds) over 10 runs.

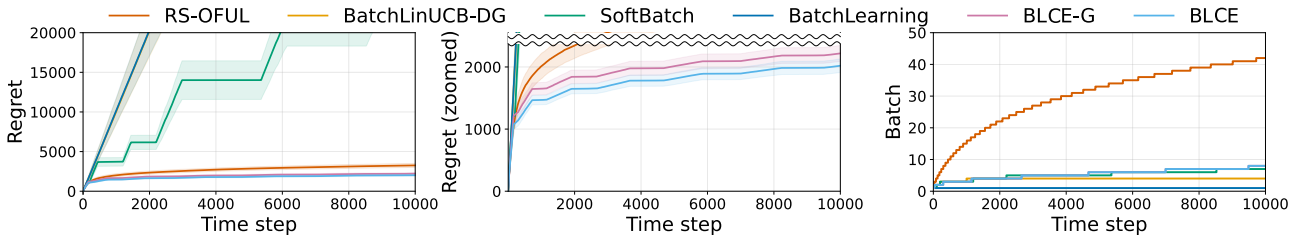
	Suboptimal algorithms		**Optimal algorithms**				
(K, d)	RS-OFUL	SoftBatch	BatchLinUCB-DG	Hanna et al. (2023b)	BatchLearning	BLCE-G	BLCE
$(1000, 5)$	1.37	1.15	148.19	Exponential	143.07	3.58	2.17
$(5000, 10)$	10.46	12.62	555.11	Exponential	590.69	9.16	6.19
$(50, 20)$	0.54	1.83	981.19	Exponential	46.24	1.39	1.05
$(100, 30)$	0.95	3.51	2773.66	Exponential	75.07	1.97	1.41



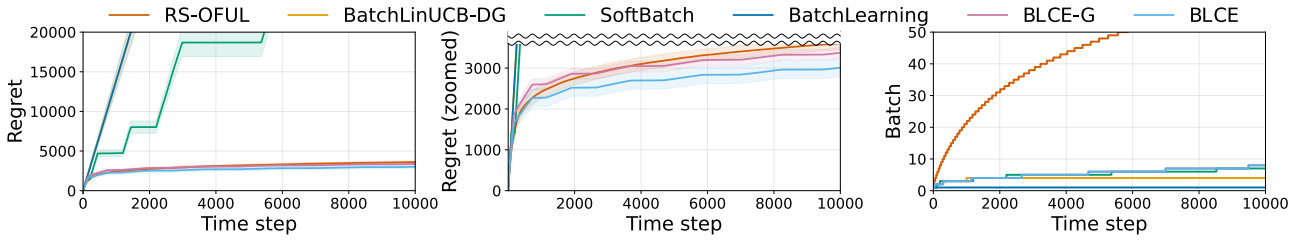
(a) $K = 1000, d = 5$



(b) $K = 5000, d = 10$



(c) $K = 50, d = 20$



(d) $K = 100, d = 30$

Figure 2. Regret, zoomed-in regret, and interval (batch) complexity over time for different values of K and d.

D.2. Experimental Results for Generalized Linear Contextual Bandits

We evaluate the performance of BGLE by measuring the cumulative regret over a horizon of $T = 10{,}000$ rounds. At each iteration, K arms are independently sampled from either a d-dimensional uniform or normal distribution, and the parameter vector θ^* is drawn from a d-dimensional normal distribution. Each experiment is repeated 20 times for the parameter pairs $(K, d) \in \{(20, 2), (50, 3)\}$, considering both uniform and normal contexts.

For comparison, we benchmark against a state-of-the-art algorithm: B-GLinCB (Sawarni et al. 2024). The within-interval allocation rate for BGLE is set to $c = 0.5$, and we conduct experiments on logistic bandits with $R = S = 1$. Importantly, BGLE is implemented with the exact theoretical hyperparameters from our main results, without any tuning.

We report two types of figures. First, the average cumulative regret (solid line) together with its standard deviation (shaded region) over 20 runs. Second, the average interval complexity across 20 runs, showing how frequently each algorithm updates its policy.

As shown in Figure 3, BGLE consistently outperforms the baseline, achieving lowest regret and demonstrating stable performance. Runtime comparisons in Table 5 further show that BGLE incurs lower computation cost than B-GLinCB.

An additional limitation of B-GLinCB is that it often uses only one interval (see Figure 3), even for small values of K and d. This behavior arises because its first interval length is determined by

$$\left(900R^2S\sqrt{\kappa}e^{3RS}d^3\log T\sqrt{T}\right)^{\frac{2}{3}},$$

which easily exceeds $T = 10{,}000$ when d is small, thereby preventing meaningful batching.

Overall, these experiments highlight a clear advantage of our approach: BGLE combines optimal regret guarantees with practical efficiency. This dual benefit of theoretical optimality and empirical superiority distinguishes our method from prior approaches to generalized linear contextual bandits.

Table 5. Average runtime (seconds) over 20 runs.

	Uniform distribution		**Normal distribution**	
(K, d)	B-GLinCB	BGLE	B-GLinCB	BGLE
$(20, 2)$	24.94	3.62	25.82	3.88
$(50, 3)$	27.45	3.91	28.78	4.06

D.3. Performance Comparison between BLCE-G and BLCE

We evaluate the performance of BLCE-G and BLCE over $T = 10{,}000$ rounds with $K = 1000$ arms and feature dimension $d = 5$. At the beginning of each run, the parameter vector θ^* is sampled from a d-dimensional normal distribution. We then conduct *four independent experiments*, each corresponding to a different fixed context distribution. In each experiment, and for every round, the d-dimensional feature vectors are drawn i.i.d. from one of the following distributions: (i) a Student-t distribution with 1.5 degrees of freedom, (ii) a Beta$(0.5, 0.5)$ distribution supported on $(0, 1)$, (iii) a Laplace distribution with location 0 and scale $1/\sqrt{2}$, or (iv) an Exponential distribution with unit scale supported on $[0, \infty)$. Thus, the four experiments differ only in the underlying feature distribution, while T, K, and d remain fixed across all settings.

The within-interval allocation rate for both BLCE-G and BLCE is fixed at $c = 0.5$. As in Appendix D.1, we report cumulative regret, zoomed-in regret, and interval (batch) complexity, averaged over 10 independent runs for each distributional setting.

As shown in Figure 4, BLCE-G and BLCE exhibit remarkably similar regret performance, with neither algorithm consistently dominating the other. Across most distributional settings, BLCE achieves slightly lower regret, whereas in the heavy-tailed Student-t distribution case this difference nearly vanishes and BLCE-G matches or occasionally exceeds the performance of BLCE. While theory guarantees that BLCE-G has a strictly smaller minimax regret bound, these guarantees characterize worst-case regimes and do not necessarily arise in benign stochastic environments. Hence, the empirical results do not contradict the theoretical results.

Finally, the runtime comparison in Table 6 shows that BLCE incurs lower computational cost than BLCE-G. By removing the G-optimal design component, BLCE achieves a simpler computational structure and correspondingly faster execution in practice.

Table 6. Average runtime (seconds) over 10 runs.

Method	**Student-t**	**Beta**	**Laplace**	**Exponential**
BLCE-G	9.33	47.57	9.91	48.96
BLCE	3.56	8.16	4.11	7.96

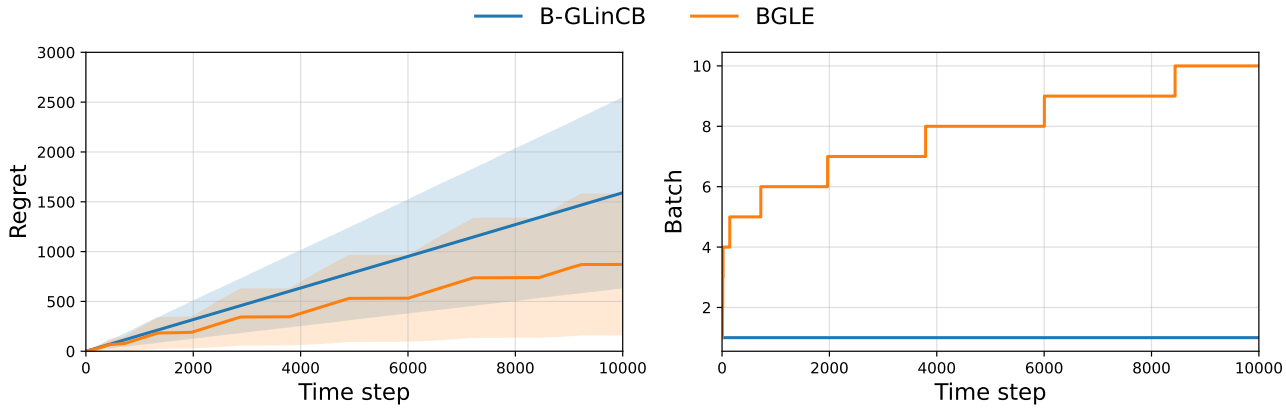
D.4. How the Performance of BLCE-G, BLCE, and BGLE Changes with the Allocation Rate c

We evaluate the performance of BLCE-G, BLCE, and BGLE over $T = 10{,}000$ rounds. For BLCE-G and BLCE, we set $K = 1000$ arms with feature dimension $d = 5$, where the arm contexts are independently sampled from either a d-dimensional uniform or normal distribution. For BGLE, we set $K = 50$ arms with feature dimension $d = 3$, again sampled independently from either a d-dimensional uniform or normal distribution. At the beginning of each run, the parameter vector θ^*

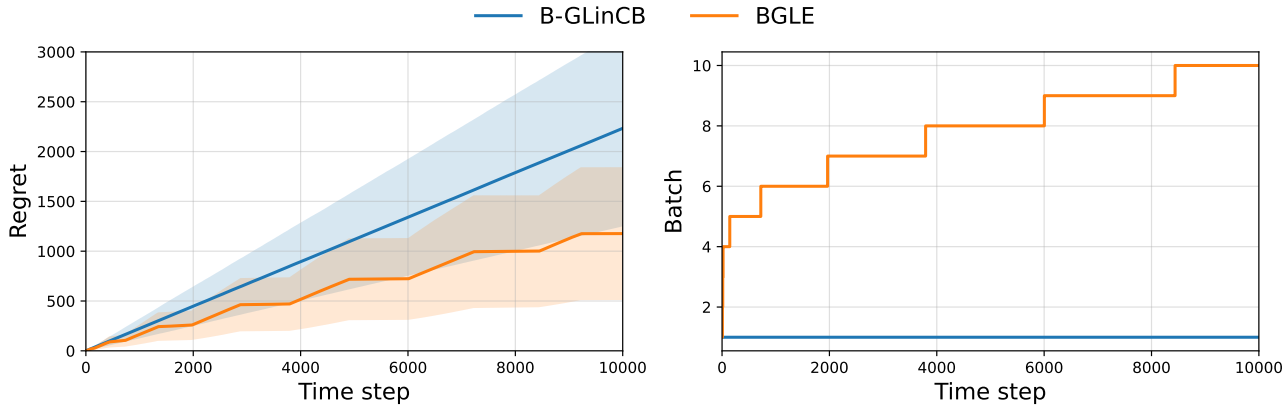
is sampled from a d-dimensional normal distribution. For each algorithm, we vary the within-interval allocation rate $c \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ and compare the resulting performance.

As in Appendix D.1, we report cumulative regret, zoomed-in regret, and interval complexity averaged over 10 independent runs for BLCE-G and BLCE. For BGLE, following Appendix D.2, we report cumulative regret and interval complexity averaged over 20 independent runs.

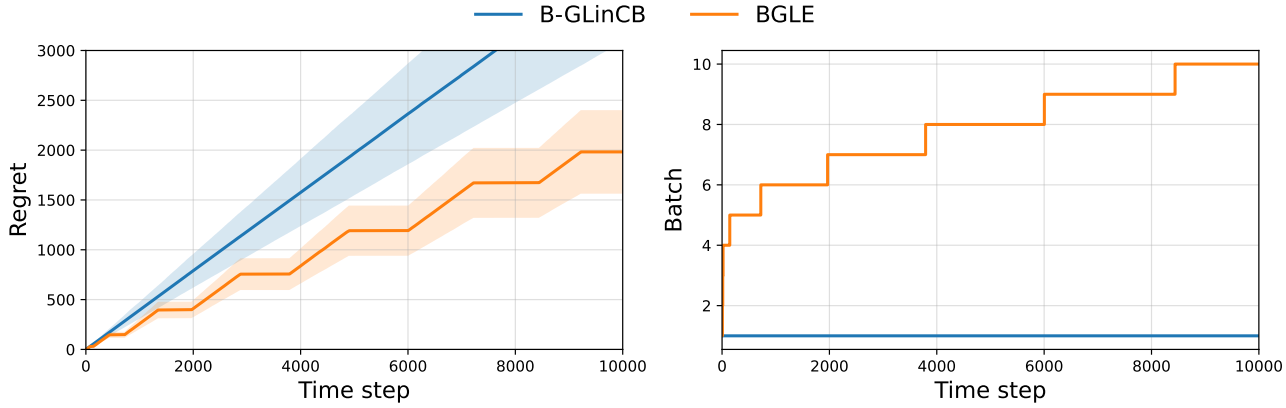
As shown in Figure 5 and Figure 6, decreasing the value of c consistently improves the empirical performance of BLCE-G, BLCE, and BGLE. This indicates that more greedy within-interval arm selection tends to yield better regret. Our theoretical analysis only requires that c be chosen in the interval $(0, 1)$ and does not predict how performance varies with c. Understanding why smaller c—that is, more greedy selection—leads to improved performance in practice remains an interesting direction for future work. The empirical findings suggest that, under stochastic contextual variation, prioritizing greedy exploitation may already provide strong performance even in the rare parameter update setting. Developing a rigorous mathematical explanation for this behavior would be an important step toward a deeper theoretical understanding of linear contextual bandits.



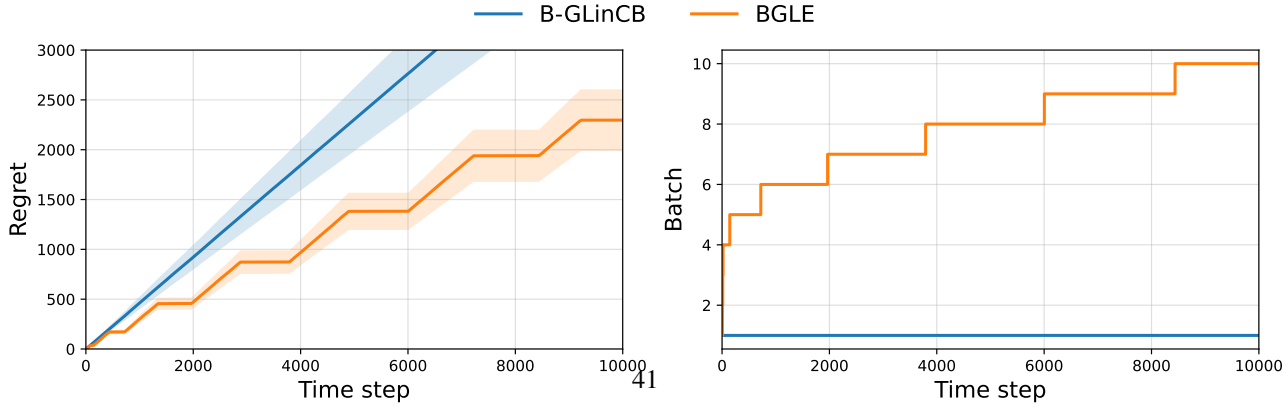
(a) $K = 20, d = 2$, uniform



(b) $K = 50, d = 3$, uniform



(c) $K = 20, d = 2$, normal



(d) $K = 50, d = 3$, normal

Figure 3. Regret and interval (batch) complexity over time for different values of K and d.

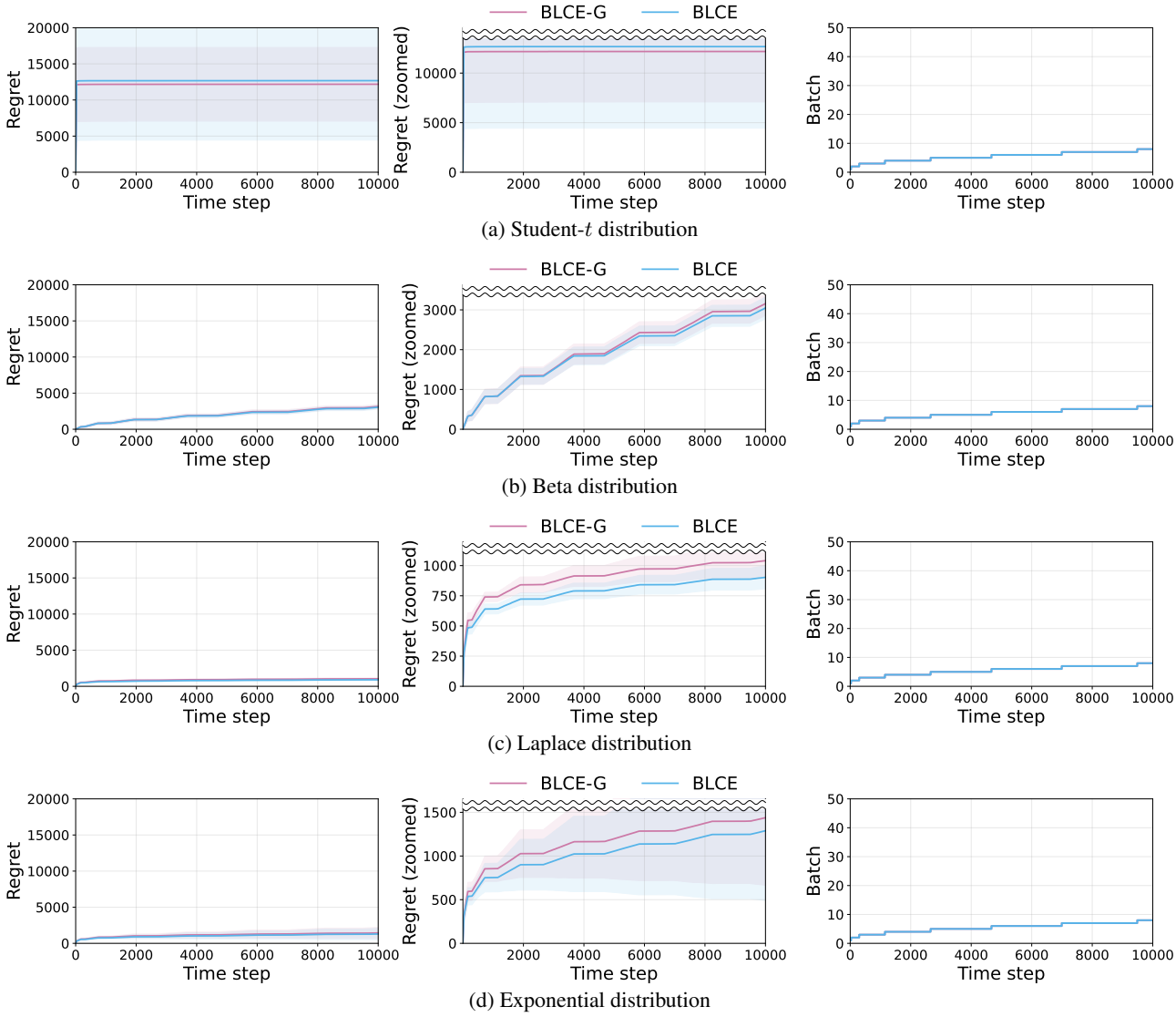
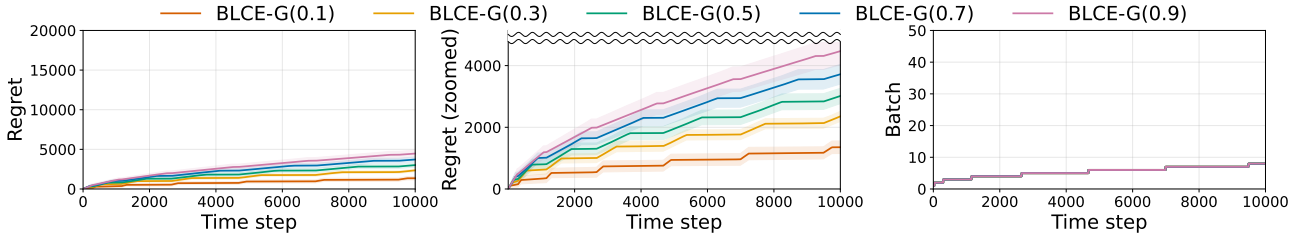
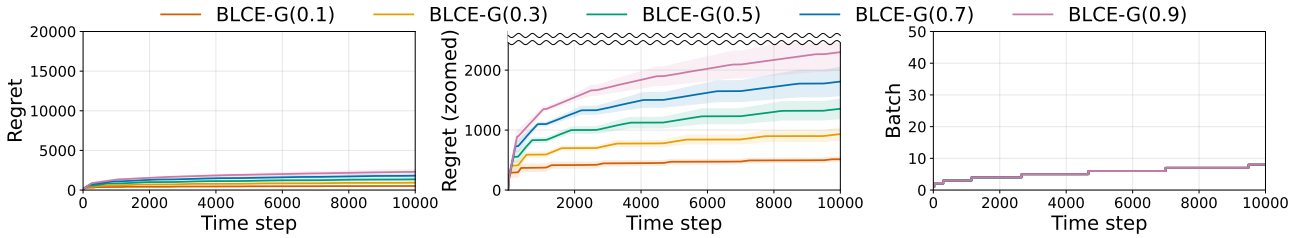


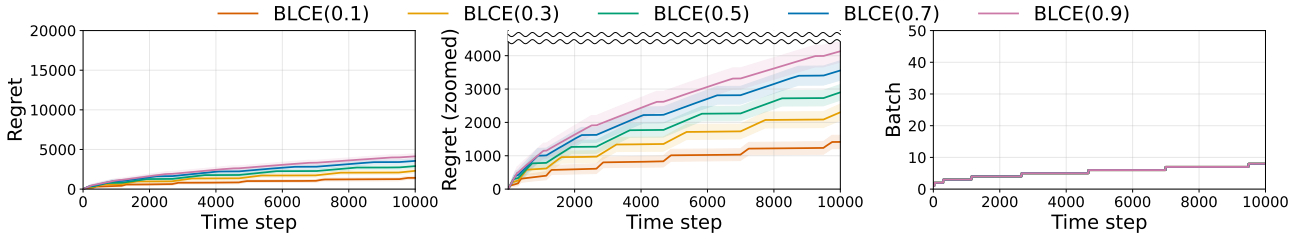
Figure 4. Regret, zoomed-in regret, and interval (batch) complexity over time for various context distributions.



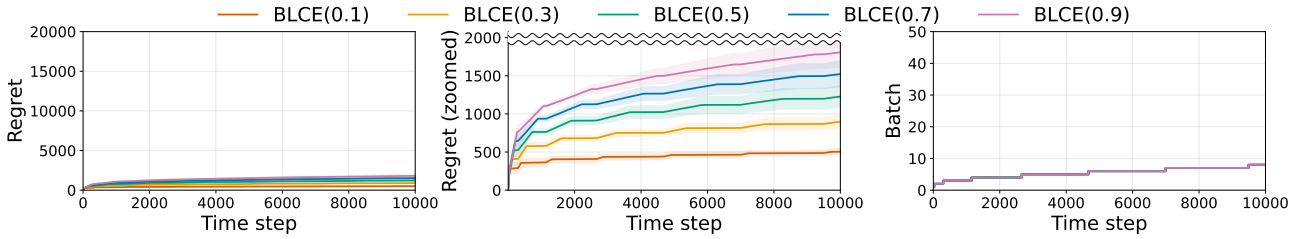
(a) BLCE–G, uniform



(b) BLCE–G, normal

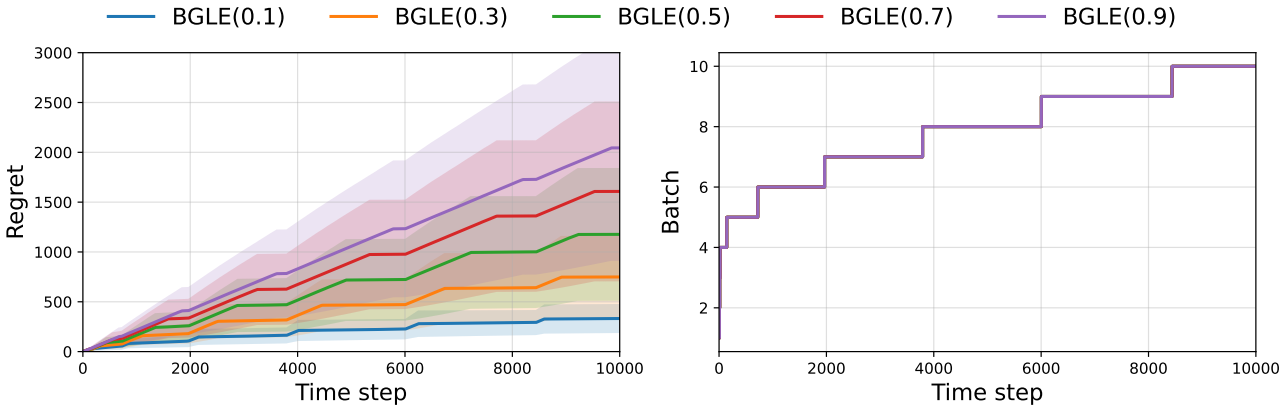


(c) BLCE, uniform

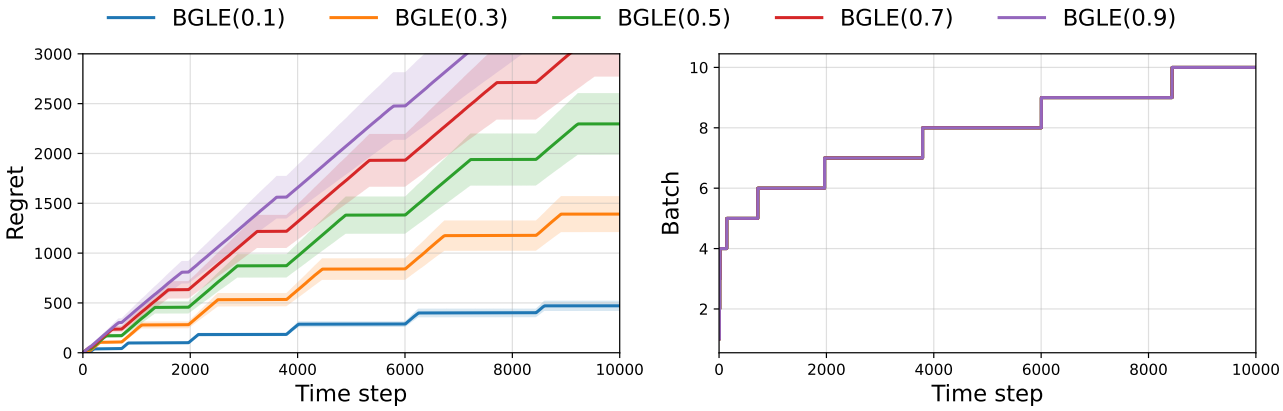


(d) BLCE, normal

Figure 5. Regret and interval (batch) complexity over time for different values of c.



(a) BGLE, uniform



(b) BGLE, normal

Figure 6. Regret and interval (batch) complexity over time for different values of c.