On the Minimax Regret for Contextual Linear Bandits and Multi-Armed Bandits with Expert Advice

Shinji Ito
The University of Tokyo and RIKEN
shinji@mist.i.u-tokyo.ac.jp

Abstract

This paper examines two extensions of multi-armed bandit problems: multi-armed bandits with expert advice and contextual linear bandits. For the former problem, multi-armed bandits with expert advice, the previously known best upper and lower bounds have been $O(\sqrt{KT \log \frac{N}{K}})$ and $\Omega(\sqrt{KT \frac{\log N}{\log K}})$, respectively. Here, K, N, and T represent the numbers of arms, experts, and rounds, respectively. We provide a lower bound of $\Omega(\sqrt{KT \log \frac{N}{K}})$ for the setup in which the player chooses an expert before observing the advices in each round. For the latter problem, contextual linear bandits, we provide an algorithm that achieves $O(\sqrt{dT \log(K \min\{1, \frac{S}{d}\})})$ together with a matching lower bound, where d and S represent the dimensionality of feature vectors and the size of the context space, respectively.

1 Introduction

This paper considers problems of multi-armed bandits with expert advice (MwE) [Auer et al., 2002, Kale, 2014], linear bandits [Dani et al., 2008, Bubeck et al., 2012, Cesa-Bianchi and Lugosi, 2012], and contextual linear bandits (CLB) [Chu et al., 2011, Abbasi-Yadkori et al., 2011, Neu and Olkhovskaya, 2020, Hanna et al., 2023, 2024].

One contribution of this paper is to present the minimax regret for MwE, which addresses the open question posed by Seldin and Lugosi [2016]. For MwE, the well-known EXP4 algorithm [Auer et al., 2002] achieves $O\left(\sqrt{KT \log N}\right)$-regret, where K, T and N represent the numbers of arms, rounds and experts. This bound is, however, not always optimal for certain parameter settings of K and N. In fact, as discussed by Seldin and Lugosi [2016], when $N = K$, the problem is reduced to the standard K-armed bandit problem, for which the minimax regret is $\Theta(\sqrt{KT})$ [Audibert and Bubeck, 2009]. This means that the regret bound of EXP4 has a gap of an $O(\sqrt{\log K})$-factor when $N = K$. On the side of upper bounds, Kale [2014] addressed this issue by providing an algorithm achieving $O\left(\sqrt{KT \log_+ \frac{N}{K}}\right)$,[1] where we denote $\log_+ x := \max\{\log x, 1\}$. This is minimax optimal for the case of $N = O(K)$. However, the minimax optimal bound for arbitrary settings of K and N has been an open question. The best known lower bound $\Omega\left(\sqrt{KT \frac{\log N}{\log K}}\right)$ is shown by Seldin and Lugosi [2016], who conjectured that this lower bound is minimax optimal. This paper provides a solution to this open question by providing a lower bound of $\Omega\left(\sqrt{KT \log_+ \frac{N}{K}}\right)$, which, together with the upper

[1] Kale [2014] deals with more general settings of the multi-armed bandit with expert advice in which only a limited number of expert advice are accessible in each round.

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

Table 1: Upper bounds ($O(\cdot)$) and lower bounds ($\Omega(\cdot)$) on regret for three problems. BwE: multi-armed bandit with expert advice, LB: linear bandit, CLB: contextual linear bandit. K: number of arms, N: number of experts, d: dimensionality of feature vectors, S: size of the context space.

Setup	Reference	Bound	Parameter
BwE	[Auer et al., 2002]	$O\left(\sqrt{KT\log N}\right)$	
	[Kale, 2014]	$O\left(\sqrt{KT\log_+ \frac{N}{K}}\right)$	
	[Seldin and Lugosi, 2016]	$\Omega\left(\sqrt{KT\frac{\log N}{\log K}}\right)$	$N \geq K$
	[This work]	$\Omega\left(\sqrt{KT\log_+ \frac{N}{K}}\right)$	$N \geq K$
LB	[Bubeck et al., 2012]	$O\left(\sqrt{dT\log K}\right)$	
	[This work]	$O\left(\sqrt{dT\log_+ \frac{K}{d}}\right)$	
	[Auer et al., 2002]	$\Omega\left(\sqrt{dT}\right) = \Omega\left(\sqrt{dT\log_+ \frac{K}{d}}\right)$	$K = d$
	[Dani et al., 2008]	$\Omega\left(\sqrt{d^2T}\right) = \Omega\left(\sqrt{dT\log_+ \frac{K}{d}}\right)$	$K = 2^d$
CLB	[Liu et al., 2024]	$O\left(d\sqrt{T\log T}\right)$	
	[This work]	$O\left(\sqrt{dT\log_+\left(K\min\left\{1, \frac{S}{d}\right\}\right)}\right)$	
	[This work]	$\Omega\left(\sqrt{dT\log_+\left(K\min\left\{1, \frac{S}{d}\right\}\right)}\right)$	$K \leq 2^d \leq K^S$

bound by Kale [2014], implies that the minimax regret is $\Theta\left(\sqrt{KT\log_+ \frac{N}{K}}\right)$. This lower bound is shown for the problem setting in which the player need to choose an expert before observing the advices in each round. As noted in Remark 1 below, though this problem setup is more challenging than the "classical" setting where the player can observe all expert advice before selecting an arm, almost all known existing algorithms, including those in Table 1, work for this setting.

Contextual linear bandit problems are a generalization of multi-armed bandit problems in which each arm $i \in [K] = \{1, 2, \ldots, K\}$ is linked to a feature vector $\phi(X_t, i) \in \mathbb{R}^d$ that depends on the context X_t at the t-th round, and the expected loss suffered for choosing i is expressed as a linear function of the feature vector: $\langle \theta_t, \phi(X_t, i)\rangle$. In particular, this paper considers the problem setting of the stochastic-context and adversarial-loss model [Neu and Olkhovskaya, 2020, Liu et al., 2024], i.e., problems with stochastic X_t and adversarial θ_t. We also assume that we are given access to the distribution of contexts, similarly to previous studies, such as those by Neu and Olkhovskaya [2020]. Linear bandit problems can be considered as special cases of contextual linear bandits in which the context space is a singleton.

Upper and lower bounds on regret for (contextual) linear bandits are shown in Table 1. For linear bandit problems, we show a regret upper bound of $O\left(\sqrt{dT\log_+(K/d)}\right)$, which is slightly better than the known best bounds of $O(\sqrt{dT\log K})$ by previous studies [Bubeck et al., 2012, Cesa-Bianchi and Lugosi, 2012, Dani et al., 2007]. One notable aspect of this novel bound is that it matches lower bounds for two important special cases of $K = 2^d$ [Dani et al., 2008, Theorem 3], and $K = d$ (equivalent to the standard K-armed bandits). However, specifying the tight minimax bounds for arbitrary values of K remains an open problem.

For contextual linear bandits, this paper presents an upper bound of $O(\sqrt{dT\log_+(K\min\{1, S/d\})})$ together with a matching lower bound, where $S = |\mathcal{X}|$ is the cardinality of the context space. Note that the regret upper bound of $O(d\sqrt{T\log T})$ by Liu et al. [2024] applies to the more general setting of the adversarial-context and adversarial-loss model. We also note that the problem formulation of contextual linear bandit problems in some existing studies [Neu and Olkhovskaya, 2020, Kuroki et al., 2024, Olkhovskaya et al., 2024] is different from ours. As noted in Liu et al. [2024], however, this problem of different formulations can be reduced to our setting with dimension $d|\mathcal{A}|$, where $|\mathcal{A}|$ is the maximum number of actions in their settings. See [Liu et al., 2024] for a more detailed review of previous studies.

To show the regret lower bound for BwE, we follow the approach employed in the study [Chen et al., 2024] on the minimax regret for problems interpolating problems of (full-information) online learning with expert advice and multi-armed bandit. They have shown that, for the generalization of a K-armed bandit problem in which each arm is associated with ν different experts, the minimax regret is $\Theta\left(\sqrt{KT\log\nu}\right)$. Inspired by their construction of the problem instance, we consider a BwE instance in which $N' = \Theta(N/K)$ experts give advice for choosing one of two arms and there are independent $\Theta(K)$ copies of such structures. This yields a lower bound of $\Omega(\sqrt{KT\log N'}) = \Omega(\sqrt{KT\log(N/K)})$.

In the proof of regret upper bounds for (contextual) linear bandits, we develop an algorithm based on the follow-the-regularized-leader (FTRL) approach with a Tsallis entropy regularizer with a parameter $\alpha \in (0,1)$: $\psi_\alpha(w) = -\frac{1}{1-\alpha}\sum_{i\in[K]}(w(i)^\alpha - w(i))$. This approach can be interpreted as a generalization of the EXP2 algorithm [Bubeck et al., 2012]. In fact, EXP2 can be regarded as the FTRL approach with Shannon entropy regularization, which coincides with the limit of α-Tsallis entropy as α approaches 1. This paper shows that the regret bounds can be further improved by not fixing α to 1 but by adjusting it appropriately. Similar approaches have been used successfully in various online learning problems including multi-armed bandits [Audibert and Bubeck, 2009], bandits with expert advice [Kale, 2014], graph bandits [Eldowa et al., 2024], and sleeping bandits [Nguyen and Mehta, 2024]. In addition to FTRL with Tsallis entropy regularization, to achieve $O\left(\sqrt{KT\log_+\left(\frac{KS}{d}\right)}\right)$-regret in contextual linear bandits, we combine a novel technique of *context-dependent learning rate*. More precisely, we introduce learning rate parameters $\eta(X_t)$ that change depending on the observed context $X_t \in \mathcal{X}$. We show that tuning $\eta : \mathcal{X} \to \mathbb{R}_{>0}$ leads to improved regret bounds.

Notation For a natural number n, denote $[n] = \{1, 2, \ldots, n\}$. For two real vectors $x = (x(i))_{i\in[d]}, y = (y(i))_{i\in[d]} \in \mathbb{R}^d$, let $\langle x, y\rangle$ denote the inner product between x and y, i.e., $\langle x, y\rangle = \sum_{i\in[d]} x(i)y(i)$. Let $\mathcal{P}(K) = \left\{w \in [0,1]^K \mid \sum_{i\in[K]} w(i) = 1\right\}$ denote the set of probability distributions over $[K]$. For any symmetric matrices A and B, denote $A \succeq B$ if and only if $A - B$ is positive-semidefinite. Let $\mathrm{tr}(A)$ denote the trace of square matrices A.

2 Multi-armed bandit with expert advice

2.1 Problem setting

In the problem of *multi-armed bandits with expert advice* (BwE), the player is given the number of arms K and the number of experts N. In each round, each expert $j \in [N]$ select an advice $e_t(j) \in [K]$, and the player chooses an expert $J_t \in [N]$. Then, after the player observes the expert advice $(e_t(j))_{j\in[N]}$, the player pulls the arm $I_t = e_t(J_t) \in [K]$ and gets feedback of the suffered loss $\ell_t(I_t) \in [0,1]$. The performance of the player is measured by means of regret R_T defined as

$$R_T = \max_{j^*\in[N]} \mathbf{E}\left[\sum_{t=1}^T \ell_t(I_t) - \sum_{t=1}^T \ell_t(e_t(j^*))\right].$$

Remark 1. We consider the problem setting in which the player choose the expert J_t before observing expert advice $(e_t(j))_{j\in[N]}$. This setting is more challenging compared to the one where the player can observe all expert advice before selecting an arm, because the available information is more limited. On the other hand, existing algorithms by Auer et al. [2002], Kale [2014] can be applied to this more challenging setting and achieve the regret upper bounds shown in Table 1. Therefore, we have decided to adopt this setting in this study.

Best expert identification The *best expert identification* (BEI) problem is a variant of BwE problem in which the player aims to identify the expert attaining the minimum value of expected loss from as few feedbacks as possible. We here assume that (e_t, ℓ_t) follows an identical distribution $\mathcal{D}$ independently for all $t = 1, 2, \ldots$. Note that, for $(e, \ell) \sim \mathcal{D}$, all elements of e and ℓ may be dependent, and that we impose the independence assumption only between data at different t. For each expert $j \in [N]$ we define $\mu_j = \mathbf{E}_{(e,\ell)\sim\mathcal{D}}[\ell(e(j))]$ and let $j^* \in [K]$ denote the *best expert* in terms of the expectation, i.e., $j^* \in \arg\min_{j\in[N]} \mu_j$. In the BEI problem, the player repeats selecting arms and obtaining feedback for an arbitrary number of times T, and then outputs the estimated optimal expert $J_T \in [K]$.

For a positive number ε, an expert j is called an *ε-optimal expert* if $\mu_j < \mu_{j^*} + \varepsilon$. An algorithm $\mathcal{A}$ for BEI is called *(ε, δ)-probably approximately correct* (PAC) if it outputs an ε-optimal expert with probability at least $(1 - \delta)$. For any algorithm $\mathcal{A}$ for BEI, let $\mathcal{T}(\mathcal{A}, \mathcal{D})$ denote the expected value of the number of rounds T until $\mathcal{A}$ is terminated when it is applied to the problem instance associated with $\mathcal{D}$.

2.2 Reduction from BEI to BwE

This section shows that BEI problems can be reduced to BwE problems. More precisely, given an algorithm for BwE problems with a regret upper bound, we can construct an (ε, δ)-PAC algorithm with a bounded number of queries, as follows:

Lemma 1. *Suppose that there exists an algorithm $\mathcal{A}$ for BwE such that the regret is bounded as $R_T \leq r(T)$ for every T. For an arbitrary $\varepsilon \in (0, 1)$, let T^* be such that $T^* \geq \frac{2500 \cdot r(T^*)}{\varepsilon}$. Then, there exists an $(\varepsilon, 0.05)$-PAC algorithm $\mathcal{A}'$ for BEI such that $\mathcal{T}(\mathcal{A}, \mathcal{D}) \leq T^*$ for any $\mathcal{D}$.*

Proof. Consider the following algorithm $\mathcal{A}'$ for best-expert identification: (i) Run $\mathcal{A}$ for T^* rounds and let T_j be the number of rounds in which the expert j is chosen, i.e., $T_j = |\{t \in [T^*] \mid J_t = j\}|$. (ii) Output $\hat{J} = j$ with probability T_j/T^*, i.e., $\hat{J}$ is chosen so that $\Pr[J = j] = T_j/T^*$ for all $j \in [N]$. Let us show that this is an $(\varepsilon, 0.05)$-PAC algorithm.

Let $\mathcal{N}_\varepsilon \subseteq [N]$ denote the set of ε-optimal experts and let $\mathcal{N}_\varepsilon^c = [N] \setminus \mathcal{N}_\varepsilon$. Let $T_\varepsilon = \sum_{j \in \mathcal{N}_\varepsilon} T_j$ denote the number of choosing ε-optimal experts. We then have $R_{T^*} \geq \mathbf{E}\left[\varepsilon \cdot \sum_{j \in \mathcal{N}_\varepsilon^c} T_j\right] = \varepsilon\, \mathbf{E}\left[(T^* - T_\varepsilon)\right]$ as we suffer per-round regret of at least ε in expectation, in every round when an expert in $\mathcal{N}_\varepsilon^c$ is chosen. Hence, from Markov's inequality, we have

$$\begin{aligned}\Pr\left[T_\varepsilon \leq 0.99T^*\right] &= \Pr\left[T^* - T_\varepsilon \geq 0.01T^*\right] \leq \frac{100}{T^*} \mathbf{E}\left[T^* - T_\varepsilon\right] \\ &\leq \frac{100}{T^*} \cdot \frac{R_{T^*}}{\varepsilon} \leq \frac{100}{T^*} \cdot \frac{r(T^*)}{\varepsilon} \leq \frac{100}{2500} = 0.04,\end{aligned}$$

where the last inequality follows from the assumption of $T^* \geq \frac{2500 \cdot r(T^*)}{\varepsilon}$. Hence, from the construction of algorithm $\mathcal{A}'$,

$$\begin{aligned}\Pr\left[\hat{J} \in \mathcal{N}_\varepsilon^c\right] &= \sum_{j \in \mathcal{N}_\varepsilon^c} \mathbf{E}\left[\frac{T_j}{T^*}\right] = \mathbf{E}\left[\frac{T^* - T_\varepsilon}{T^*}\right] \\ &\leq \Pr\left[T_\varepsilon \leq 0.99T^*\right] \cdot 1 + \Pr\left[T_\varepsilon > 0.99T^*\right] \cdot 0.01 \leq 0.04 + 0.01 = 0.05,\end{aligned}$$

which means that $\mathcal{A}'$ is an $(\varepsilon, 0.05)$-PAC algorithm for BEI. □

2.3 Construction of problem instance

For any $p \in [0, 1]$, let $\mathrm{Ber}(p)$ represent a Bernoulli distribution of parameter p, i.e., if $X \sim \mathrm{Ber}(p)$, then $\Pr[X = 1] = p$ and $\Pr[X = 0] = 1 - p$.

Without loss of generality, we consider the case that N and K can be expressed as $N = N'm + 1$ and $K = 2m + 1$ for some positive integers N' and m. We denote the set of experts by $\mathcal{E} = \{0\} \cup \{(u, v)\}_{u \in [m], v \in [N']}$ and the set of arms by $\mathcal{A} = \{0\} \cup \{(u, b)\}_{u \in [m], b \in \{0,1\}}$. The expert advice e_t is given so that $e_t(0) = 0$ and $e_t((u, v)) \in \{(u, 0), (u, 1)\}$ for all $u \in [m]$ and $v \in [N']$.

Fix $u^* \in [m]$, $v^* \in [N']$, and $\varepsilon \in [0, 1)$. We define distributions of (e, ℓ) as follows:

- $\mathcal{D}(\varepsilon)$: When (e, ℓ) follows $\mathcal{D}(\varepsilon)$, $\ell(0)$ follows $\mathrm{Ber}((1 - \varepsilon)/2)$. For each $u \in [m]$, $\ell((u, 0))$ and $\ell((u, 1))$ are given by $\ell((u, 0)) = b_u$ and $\ell((u, 1)) = 1 - b_u$, where b_u follows $\mathrm{Ber}(1/2)$. The expert advice is given by $e((u, v)) = (u, b_{uv})$, where each b_{uv} follows $\mathrm{Ber}(1/2)$ for $u \in [m]$ and $v \in [N']$. All elements of $\ell(0)$, $(b_u)_{u \in [m]}$ and $(b_{uv})_{u \in [m], v \in [N']}$ are independent.

- $\mathcal{D}(\varepsilon, u^*, v^*)$: When (e, ℓ) follows $\mathcal{D}(\varepsilon, u^*, v^*)$, ℓ and e follows the same distribution as $\mathcal{D}(\varepsilon)$ except for $e((u^*, v^*))$. The value of $e((u^*, v^*))$ is given as $e((u^*, v^*)) = (u^*, b_{u^* v^*}) = (u^*, |\ell((u^*, 0)) - b'|)$, where b' is a random variable that follows $\mathrm{Ber}(1/2 - \varepsilon)$ independently of the other randomness. In other words, the second element of $e((u^*, v^*))$ is chosen so that $\ell(e(u^*, v^*))$ follows $\mathrm{Ber}(1/2 - \varepsilon)$.

The probability distribution of $\mathcal{D}(\varepsilon, u^*, v^*)$ is constructed so that the marginal distribution of e for $(e, \ell) \sim \mathcal{D}(\varepsilon, u^*, v^*)$ is same as that for $\mathcal{D}(\varepsilon)$. Indeed, the values of e is determined by $(b_{uv})_{u \in [m], v \in [N']}$, and the marginal distribution of b_{u^*, v^*} for $\mathcal{D}(\varepsilon, u^*, v^*)$ is a mixture distribution given as $\Pr\left[\ell((u^*, 0)) = 0\right] \cdot \mathrm{Ber}\left(\frac{1}{2} - \varepsilon\right) + \Pr\left[\ell((u^*, 0)) = 1\right] \cdot \mathrm{Ber}\left(\frac{1}{2} + \varepsilon\right) = \frac{1}{2}\mathrm{Ber}\left(\frac{1}{2} - \varepsilon\right) + \frac{1}{2}\mathrm{Ber}\left(\frac{1}{2} + \varepsilon\right) = \mathrm{Ber}\left(\frac{1}{2}\right)$, which is equivalent to distributions for $\mathcal{D}(\varepsilon)$. In addition, if $(e, \ell) \sim \mathcal{D}$ follows $\mathcal{D}(\varepsilon, u^*, v^*)$, the expected loss for choosing expert (u, v) is given as

$$\mathop{\mathbf{E}}_{(e,\ell) \sim \mathcal{D}(\varepsilon, u^*, v^*)} \left[\ell(e(u, v))\right] = \frac{1}{2} - \varepsilon \cdot \mathbf{1}\left[(u, v) = (u^*, v^*)\right],$$

where $\mathbf{1}[\cdot]$ represents the indicator function, i.e., $\mathbf{1}[E] = 1$ if E is true and $\mathbf{1}[E] = 0$ otherwise. This means that, for the problem instance of BEI associated with $\mathcal{D}(\varepsilon, u^*, v^*)$, the expert (u^*, v^*) is the only $(\varepsilon/2)$-optimal action.

We denote $\mathcal{D}_T(\varepsilon) = (\mathcal{D}(\varepsilon))^T$. Let $\mathcal{D}_T(\varepsilon, u^*)$ denote the uniform mixture of $\{(\mathcal{D}(\varepsilon, u^*, v^*))^T\}_{v^* \in [N']}$: $\mathcal{D}_T(\varepsilon, u^*) = \frac{1}{N'} \sum_{v^* \in [N']} (\mathcal{D}(\varepsilon, u^*, v^*))^T$.

2.4 Lower bound for best-expert identification problems

We first consider the special case of $m = 1$ and provide an instance-specific lower bound for BEI associated with $\mathcal{D}(\varepsilon)$ and $\{\mathcal{D}(\varepsilon, 1, v^*)\}_{v^* \in N'}$.

Lemma 2. *Suppose $m = 1$ and let $\mathcal{A}$ be an $(\varepsilon/2, 0.05)$-PAC algorithm for BEI. We then have $\mathcal{T}(\mathcal{A}, \mathcal{D}(\varepsilon)) \geq \frac{1}{2} \left\lfloor \frac{\ln(N'/4)}{4\varepsilon^2} \right\rfloor =: \frac{T^*}{2}$.*

Proof. Let T be the number of rounds the algorithm $\mathcal{A}$ proceeds. Let E denote the event that the algorithm $\mathcal{A}$ terminate at a round before $T^* + 1$ (i.e., $T \leq T^*$) and output $\hat{J} = 0$. Then, from Pinsker's inequality, we have $\left|\Pr\left[E | ((e_t, \ell_t))_{t=1}^{T^*} \sim \mathcal{D}_{T^*}(\varepsilon)\right] - \Pr\left[E | ((e_t, \ell_t))_{t=1}^{T^*} \sim \mathcal{D}_{T^*}(\varepsilon, 1)\right]\right| \leq D_{\mathrm{KL}}\left(\mathcal{D}_{T^*}(\varepsilon, 1) || \mathcal{D}_{T^*}(\varepsilon)\right) \leq 0.25$, where the last inequality follows from Lemma 9 in the supplementary. As $\mathcal{A}$ is an $(\varepsilon/2, 0.05)$-PAC algorithm, we have $\Pr\left[E | ((e_t, \ell_t))_{t=1}^{T^*} \sim \mathcal{D}_{T^*}(\varepsilon, 1)\right] \leq 0.05$. By applying the union bound, we obtain $1 - \Pr\left[E | ((e_t, \ell_t))_{t=1}^{T^*} \sim \mathcal{D}_{T^*}(\varepsilon)\right] = \Pr\left[T > T^* | ((e_t, \ell_t))_{t=1}^{T^*} \sim \mathcal{D}_{T^*}(\varepsilon)\right] + \Pr\left[T \leq T^*, \hat{J} \neq 0 | ((e_t, \ell_t))_{t=1}^{T^*} \sim \mathcal{D}_{T^*}(\varepsilon)\right] \leq \Pr\left[T > T^* | ((e_t, \ell_t))_{t=1}^{T^*} \sim \mathcal{D}_{T^*}(\varepsilon)\right] + 0.05$. Combining these inequalities, we obtain $\Pr\left[T > T^* | ((e_t, \ell_t))_{t=1}^{T^*} \sim \mathcal{D}_{T^*}(\varepsilon)\right] \geq 0.95 - \Pr\left[E | ((e_t, \ell_t))_{t=1}^{T^*} \sim \mathcal{D}_{T^*}(\varepsilon)\right] \geq 0.95 - 0.25 - \Pr\left[E | ((e_t, \ell_t))_{t=1}^{T^*} \sim \mathcal{D}_{T^*}(\varepsilon, 1)\right] \geq 0.95 - 0.25 - 0.05 = 0.65$, which implies that $\mathcal{T}(\mathcal{A}, \mathcal{D}(\varepsilon)) \geq \mathbf{E}\left[T \mid ((e_t, \ell_t))_{t=1}^{T^*} \sim \mathcal{D}_{T^*}(\varepsilon)\right] \geq 0.65 T^*$. This completes the proof. □

We can obtain a lower bound for BEI for general m by using Lemma 2.

Theorem 1. *Let $\mathcal{A}$ be an $(\varepsilon/2, 0.05)$-PAC algorithm for BEI. We then have $\mathcal{T}(\mathcal{A}, \mathcal{D}(\varepsilon)) \geq \frac{m}{2} \left\lfloor \frac{\ln(N'/4)}{4\varepsilon^2} \right\rfloor = \frac{mT^*}{2}$.*

2.5 Lower bound for multi-armed bandits with expert advice

We are now ready to provide a lower bound for BwE.

Theorem 2. *For any algorithm for BwE and for any sufficiently large $T > 0$, there exists a problem instance for which $R_T \geq C\sqrt{TK \log_+ \frac{N}{K}}$, where $C > 0$ is a universal constant.*

Proof. We fix T and set $m = \lfloor (K-1)/2 \rfloor$ and $N' = \lfloor (N-1)/m \rfloor$. Fix $\varepsilon = \Theta\left(\sqrt{\frac{m \log N'}{T}}\right)$ so that $T < \frac{m}{2}\left\lfloor \frac{\ln(N'/4)}{4\varepsilon^2} \right\rfloor$. Assume that $R_T < C\sqrt{TK \log_+ \frac{N}{K}} := r(T) = O\left(\sqrt{Tm \log N'}\right)$, where the universal constant $C > 0$ is sufficiently small so that $\frac{2500 \cdot r(T)}{\varepsilon} \leq T$ holds. Then, from Lemma 1, there exists an $(\varepsilon, 0.05)$-PAC BEI algorithm $\mathcal{A}$ achieving $\mathcal{T}(\mathcal{A}, \mathcal{D}) \leq T$ for any $\mathcal{D}$. This contradicts Theorem 1, which implies that $R_T \geq C\sqrt{TK \log_+ \frac{N}{K}}$. □

This lower bound, together with the upper bound presented by Kale [2014], means that the minimax regret for BwE is of $\Theta\left(\sqrt{KT \log_+ \frac{N}{K}}\right)$.

3 Contextual linear bandit

3.1 Problem setting

Before the game starts, the player is given a set $\mathcal{X}$ of contexts, a finite set $[K]$ of arms, a feature mapping $\phi : \mathcal{X} \times [K] \to \mathbb{R}^d$. In each round, the environment chooses loss vector $\theta_t \in \mathbb{R}^d$. Then a context $X_t \in \mathcal{X}$ is drawn from a fixed distribution $\mathcal{D}$, and is revealed to the player. The player chooses an action $I_t \in [K]$ and get feedback of $\ell_t \in [-1, 1]$, of which expectation is $\langle \theta_t, \phi(X_t, I_t) \rangle$. For any policy $\pi^* : \mathcal{X} \to [K]$, we define the regret by

$$R_T(\pi^*) = \mathbf{E}\left[\sum_{t=1}^{T} \langle \theta_t, \phi(X_t, I_t) \rangle - \sum_{t=1}^{T} \langle \theta_t, \phi(X_t, \pi^*(X_t)) \rangle\right], \quad R_T = \sup_{\pi^* : \mathcal{X} \to [K]} R_T(\pi^*).$$

We assume that $|\langle \theta_t, \phi(x, i) \rangle| \leq 1$ for all $x \in \mathcal{X}$ and $a \in [K]$. We also pose the following assumption regarding the distribution of the context.

Assumption 1. We assume that the set of contexts is a finite set $\mathcal{X} = [S]$ and that there exists $L \geq S$ such that the probability $g(x) := \Pr[X_t = x]$ is bounded from below as $g(x) \geq 1/L$ for all $x \in [S]$. Assume that the function g is given.

For any *randomized policy* $p : \mathcal{X} \to \mathcal{P}(K)$, we denote

$$V(p) = \mathop{\mathbf{E}}_{X \sim \mathcal{D}, I \sim p(X)}\left[\phi(X, I)\phi(X, I)^\top\right], \quad \lambda(p) = \sup_{i \in [K], x \in \mathcal{X}} \phi(x, i)^\top V(p)^{-1} \phi(x, i).$$

Assumption 2. We assume that there exists an *exploration policy* $p_0 : \mathcal{X} \to \mathcal{P}(K)$ such that $\lambda(p_0) < \infty$.

Let $\lambda_0 > 0$ denote (an upper bound of) the value of $\lambda(p_0)$. We may assume that $\lambda_0 \leq Ld$ without loss of generality. In fact, if we set $p_0(x)$ to be a g-optimal design (see, e.g., [Lattimore and Szepesvári, 2020, Section 21.1]) for $\mathcal{Z}_x = \{\phi(x, i) \mid i \in [K]\}$, we then have $\phi(x, i)^\top V(p)^{-1} \phi(x, i) \leq \frac{1}{g(x)} \phi(x, i)^\top V(p_0(x))^{-1} \phi(x, i) \leq dL$ holds for all $x \in \mathcal{X}$ and $i \in [K]$.

3.2 Algorithm

The proposed algorithm is based on the framework follow-the-regularized-leader (FTRL) with Tsallis entropy regularization. For $\alpha \in (0, 1)$, define a regularization function $\psi : \mathcal{P}(K) \to \mathbb{R}$ by

$$\psi(w) = \frac{1}{\alpha} \sum_{i \in [K]} (w(i) - w(i)^\alpha) = \frac{1}{\alpha}\left(1 - \sum_{i \in [K]} w(i)^\alpha\right). \quad (1)$$

Using this regularizer, we define a randomized policy $q_t : \mathcal{X} \to \mathcal{P}(K)$ on the basis of FTRL, and set an arm-selection policy p_t, as follows:

$$q_t(x) = \mathop{\arg\min}_{w \in \mathcal{P}(\mathcal{A})}\left\{\sum_{i \in [K]} w(i) \sum_{s=1}^{t-1} \left\langle \hat{\theta}_s, \phi(x, i) \right\rangle + \frac{1}{\eta(x)} \psi(w)\right\}, \quad p_t = (1 - \gamma) q_t + \gamma p_0 \quad (2)$$

Algorithm 1 Contextual linear bandit algorithm based on FTRL with Tsallis entropy

```
Input: Feature mapping $\phi$, learning rates $(\eta(x))_{x \in \mathcal{X}}$, exploration policy $p_0$, parameters $\gamma, \alpha$
for $t = 1, \ldots, T$ do
    Compute $p_t$ given by (1) and (2).
    Observe $X_t$, choose $I_t$ drawn from $p_t(X_t)$, and get feedback of $\ell_t$.
    Compute $\hat{\theta}_t$ defined by (3).
end for
```

where $\hat{\theta}_t$ is an unbiased estimator defined below, $\eta(x) > 0$ is a learning rate parameter that will be specified later, and $\gamma \in (0, 1/2)$ is a parameter that will be chosen depending on η. In each round t, we choose I_t following $p_t(X_t)$ and get feedback of ℓ_t. We then compute an unbiased estimator $\hat{\theta}_t$ of θ_t defined by

$$\hat{\theta}_t = \ell_t V(p_t)^{-1} \phi(X_t, I_t). \tag{3}$$

This is in fact an unbiased estimator as

$$\begin{aligned} \mathop{\mathbf{E}}_{X_t \sim \mathcal{D}, I_t \sim p_t(X_t)} \left[\hat{\theta}_t \right] &= \mathop{\mathbf{E}}_{X_t \sim \mathcal{D}, I_t \sim p_t(X_t)} \left[\ell_t V(p_t)^{-1} \phi(X_t, I_t) \right] \\ &= \mathop{\mathbf{E}}_{X_t \sim \mathcal{D}, I_t \sim p_t(X_t)} \left[V(p_t)^{-1} \phi(X_t, I_t) \phi(X_t, I_t)^\top \theta_t \right] = V(p_t)^{-1} V(p_t) \theta_t = \theta_t. \end{aligned} \tag{4}$$

The procedure of our proposed algorithm is summarized in Algorithm 1.

3.3 Regret upper bound

This section provides an upper bound on the regret for Algorithm 1. In the following, we use the symbol $\beta = 1 - \alpha \in (0, 1)$ for simplicity of notation. We will show a regret upper bound as follows:

Theorem 3. *Suppose that $1/2 \leq \alpha < 1$ and that γ satisfies*

$$\gamma = \lambda_0 \cdot \min \left\{ 8 \sup_{x \in \mathcal{X}} \eta(x), \ \sup_{x \in \mathcal{X}} \left(\frac{16 \eta(x)}{g(x)^\beta} \right)^{1/\alpha} \right\} \leq \frac{1}{2}. \tag{5}$$

Then, the regret for Algorithm 1 is bounded as follows:

$$R_T(\pi^*) = O \left(T \left(\frac{1}{\beta} \min \left\{ d \cdot \sup_{x \in \mathcal{X}} \eta(x), \ d^\alpha \cdot \sup_{x \in \mathcal{X}} \frac{\eta(x)}{g(x)^\beta} \right\} + \gamma \right) + \frac{K^\beta - 1}{\alpha} \sum_{x \in \mathcal{X}} \frac{g(x)}{\eta(x)} \right).$$

From this theorem, by tuning parameters α and $\eta(x)$ on the basis of T, d, K and g, we obtain the following:

Corollary 1. *For sufficiently large T, Algorithm 1 achieves the following:*

- *By setting $\eta(x) = \eta' g(x)^\beta =$ with $\eta' = \Theta \left(\sqrt{\frac{\beta (KS)^\beta}{\alpha d^\alpha T}} \right)$ and $\beta = \Theta \left(1 / \left(\log_+ \left(\frac{KS}{d} \right) \right) \right)$, we obtain $R_T = O \left(\sqrt{dT \log_+ \left(\frac{KS}{d} \right)} + \lambda_0 \sqrt{T^{1-\beta}} \right)$.*
- *By setting $\eta(x) = \eta = \Theta \left(\sqrt{\frac{\beta K^\beta}{\alpha d T}} \right)$ for all $x \in \mathcal{X}$ and $\beta = \Theta \left(\frac{1}{\log K} \right)$, we obtain $R_T = O \left(\sqrt{dT \log K} + \lambda_0 \min \left\{ \sqrt{\frac{T}{d \log K}}, L^{\frac{\beta}{\alpha}} \sqrt{T^{1-\beta}} \right\} \right)$.*

Remark 2. The proposed algorithm (Algorithm 1) can also work for the infinite context case, in which it enjoys the second regret upper bound in Corollary 1. In fact, this regret upper bound does not include $S = |\mathcal{X}|$ or L, and the value of $g(x)$ is not required to define $\eta(x)$ in showing this second upper bound. Hence, we can show this bound without the assumption that $\mathcal{X}$ is a finite set. In the infinite context case, however, further challenges regarding the computational complexity of the algorithm should be noted. For example, we need to compute $V(p_t)$ in the algorithm as it appears in the definition of $\hat{\theta}_t$ in (3), which tend to be computationally expensive, depending on the computational model and the setup of distributions.

To prove Theorem 3, we introduce some notations: Denote $\hat{\ell}_t(x) = \left(\left\langle \hat{\theta}_t, \phi(x,i) \right\rangle\right)_{i\in[K]} \in \mathbb{R}^K$, which is an unbiased estimator of $(\langle \theta_t, \phi(x,i)\rangle)_{i\in[K]}$. For any $w, w' \in \mathcal{P}(K)$, let $D(w,w')$ denote the Bregman divergence associated with ψ, i.e., $D(w,w') = \psi(w) - \psi(w') - \langle \nabla\psi(w'), w - w'\rangle$. From the standard analysis technique for FTRL (see, e.g., [Lattimore and Szepesvári, 2020, Chapter 28]) and the idea of *ghost sample* $X_0 \sim \mathcal{D}$ drawn independently from all other randomness (see, e.g., [Neu and Olkhovskaya, 2020]), we obtain the following upper bound:

Lemma 3. *For Algorithm 1, the regret is bounded as* $R_T \le 2\gamma T + \frac{K^\beta - 1}{\alpha}\sum_{x\in\mathcal{X}}\frac{g(x)}{\eta(x)} + \mathbf{E}\left[\sum_{t=1}^T \left(\left\langle \hat{\ell}_t(X_0), q_t(X_0) - q_{t+1}(X_0)\right\rangle - \frac{1}{\eta(X_0)} D(p_{t+1}(X_0), p_t(X_0))\right)\right]$.

We note that X_0 is a random variable that does not appear in the decision-making process or algorithms, but appear only in the analysis, and is defined to be independent of $X_1, X_2, \ldots, X_T$ (and therefore is independent of any other variables including q_t). The first component of the right-hand side of this formula can be bounded via the following lemma:

Lemma 4. *Suppose that* $\eta > 0$, $\ell \in \mathbb{R}^K$ *and* $w \in \mathcal{P}(K)$ *satisfy* $\eta|\ell(i)| \le \frac{1-\alpha}{4} w(i)^{\alpha-1}$ *for all* $i \in [K]$. *We then have* $\langle \ell, w - w'\rangle - \frac{1}{\eta} D(w', w) \le \frac{4\eta}{1-\alpha}\sum_{i\in[K]} w(i)^{2-\alpha}\ell(i)^2$.

This follows, e.g., directly from the first part of Lemma 9 given by Ito et al. [2024]. To check the sufficient conditions for applying Lemma 4, we use the following lemma:

Lemma 5. *It holds for any* $x \in \mathcal{X}$ *and* $i \in [K]$ *that*

$$\phi(x,i)^\top V(p_t)^{-1}\phi(x,i) \le \min\left\{\frac{\lambda_0}{\gamma}, \frac{1}{(1-\gamma)g(x)q_t(x,i)}\right\}. \tag{6}$$

Consequently, we have

$$\left|\hat{\ell}_t(x,i)\right| = \left|\left\langle \hat{\theta}_t, \phi(x,i)\right\rangle\right| \le \sqrt{\frac{\lambda_0}{\gamma}, \min\left\{\frac{\lambda_0}{\gamma}, \frac{1}{(1-\gamma)g(x)q_t(x,i)}\right\}}. \tag{7}$$

Combining Lemmas 4 and 5, we obtain the following:

Lemma 6. *Suppose that* $\alpha \ge 1/2$ *and that* γ *is given as* (5). *Then it holds for any* t *and* $x \in \mathcal{X}$ *that*

$$\left\langle \hat{\ell}_t(x), q_t(x) - q_{t+1}(x)\right\rangle - \frac{1}{\eta(x)} D(p_{t+1}(x), p_t(x)) \le \frac{4\eta(x)}{\beta}\sum_{i\in[K]} q_t(x,i)^{2-\alpha}\hat{\ell}_t(x,i)^2. \tag{8}$$

Further, the expectation of the right-hand side of (8) can be bounded as in the following, which is the key lemma for leading to an improved regret bound of $O(\sqrt{dT\log_+\frac{KS}{d}})$.

Lemma 7. *We have*

$$\mathbf{E}\left[\eta(X_0)\sum_{i\in[K]} q_t(X_0,i)^{2-\alpha}\hat{\ell}_t(X_0,i)^2\right] \le \frac{1}{1-\gamma}\min\left\{d\cdot\sup_{x\in\mathcal{X}}\eta(x),\ d^\alpha\cdot\sup_{x\in\mathcal{X}}\frac{\eta(x)}{g(x)^\beta}\right\}. \tag{9}$$

Proof. From the definition of $\hat{\ell}_t(x,i)$, for any fixed $x \in \mathcal{X}$ and $i \in [K]$, we have

$$\begin{aligned}
\mathbf{E}\left[\hat{\ell}_t(x,i)^2\right] &= \mathbf{E}\left[\left(\ell_t\phi(X_t,I_t)^\top V(p_t)^{-1}\phi(x,i)\right)^2\right] \le \mathbf{E}\left[\left(\phi(X_t,I_t)^\top V(p_t)^{-1}\phi(x,i)\right)^2\right] \\
&= \mathbf{E}\left[\phi(x,i)^\top V(p_t)^{-1}\phi(X_t,I_t)\phi(X_t,I_t)^\top V(p_t)^{-1}\phi(x,i)\right] \\
&= \phi(x,i)^\top V(p_t)^{-1}V(p_t)V(p_t)^{-1}\phi(x,i) \\
&= \phi(x,i)^\top V(p_t)^{-1}\phi(x,i) \le \frac{1}{1-\gamma}\phi(x,i)^\top V(q_t)^{-1}\phi(x,i).
\end{aligned} \tag{10}$$

Denote $\eta_1 = \sup_{x \in \mathcal{X}} \eta(x)/g(x)^\beta$. Then, from (10), we have

$$\begin{aligned}
[\text{LHS of (9)}] &\leq \frac{\eta_1}{1-\gamma} \mathbf{E} \left[g(X_0)^\beta \sum_{i \in [K]} q_t(X_0, i)^{1+\beta} \phi(X_0, i)^\top V(q_t)^{-1} \phi(X_0, i) \right] \\
&= \frac{\eta_1}{1-\gamma} \sum_{x \in \mathcal{X}} g(x)^{1+\beta} \sum_{i \in [K]} q_t(x, i)^{1+\beta} \phi(x, i)^\top V(q_t)^{-1} \phi(x, i) \\
&= \frac{\eta_1}{1-\gamma} \sum_{x \in \mathcal{X}} \sum_{i \in [K]} u(x, i)^\beta v(x, i), \qquad (11)
\end{aligned}$$

where we define $u(x, i) = g(x) q_t(x, i)$ and $v(x, i) = g(x) q_t(x, i) \phi(x, i)^\top V(q_t)^{-1} \phi(x, i)$. We then have (i) $u(x, i) \geq 0$ for all $x \in \mathcal{X}$ and $i \in [K]$; $\sum_{x' \in \mathcal{X}} \sum_{i' \in [K]} u(x', i') = 1$, (ii) $\sum_{x' \in \mathcal{X}} \sum_{i' \in [K]} v(x', i') = d$, and (iii) $0 \leq v(x, i) \leq 1$ for all $x \in \mathcal{X}$ and $i \in [K]$. Indeed, the condition (i) is clear from the fact that $u(x, i)$ is a probability mass function over $\mathcal{X} \times [K]$. The condition (ii) follows from

$$\begin{aligned}
\sum_{x' \in \mathcal{X}} \sum_{i' \in [K]} v(x', i') &= \operatorname{tr} \left(V(q_t)^{-1} \sum_{x \in \mathcal{X}} \sum_{i \in [K]} g(x) q_t(x, i) \phi(x, i) \phi(x, i)^\top \right) \\
&= \operatorname{tr} \left(V(q_t)^{-1} V(q_t) \right) = \operatorname{tr}(I_d) = d.
\end{aligned}$$

The condition (iii) follows from Lemma 8 and the fact that as $g(x) q_t(x, i) \phi(x, i) \phi(x, a)^\top \succeq V(q_t)$. Let $U \subseteq \mathcal{X} \times [K]$ be the top-d subset with respect to the values of $u(x, a)$, i.e., let U be such that $|U| = d$ and $u(x, i) \geq u(x', i')$ for any $(x, i) \in U$ and any $(x', i') \in (\mathcal{X} \times [K]) \setminus U$. Then, when we consider maximizing $\sum_{x \in \mathcal{X}} \sum_{i \in [K]} u(x, i)^\beta v(x, i)$ subject to the constraint of (ii) and (iii) on v, the maximum is attained by $v(x, i) = \mathbf{1}[(x, i) \in U]$. We hence have $\sum_{x \in \mathcal{X}} \sum_{i \in [K]} u(x, i)^\beta v(x, i) \leq \sum_{(x,i) \in U} u(x, i)^\beta \leq |U|^{1-\beta} \left(\sum_{(x,i) \in U} u(x, i) \right)^\beta \leq d^\alpha$, where the second inequality follows from Hölder's inequality and the last inequality follows from the condition (i). By combining this with (11), we obtain [LHS of (9)] $\leq \frac{\eta_1 d^\alpha}{1-\gamma}$. Similarly, denoting $\eta_0 = \sup_{x \in \mathcal{X}} \eta(x)$, we obtain [LHS of (9)] $\leq \frac{\eta_0}{1-\gamma} \mathbf{E} \left[\sum_{i \in [K]} q_t(X_0, i) \phi(X_0, i)^\top V(q_t)^{-1} \phi(X_0, i) \right] = \frac{\eta_0}{1-\gamma} \mathbf{E} \left[\operatorname{tr} \left(V(q_t)^{-1} \sum_{i \in [K]} q_t(X_0, i) \phi(X_0, i) \phi(X_0, i)^\top \right) \right] = \frac{\eta_0}{1-\gamma} \sum_{x \in \mathcal{X}} \sum_{i \in [K]} v(x, i) = \frac{\eta_0 d}{1-\gamma}$, which completes the proof. □

Now we are ready to provide an upper bound on regret.

We can easily see that Theorem 3 is a direct consequence of Lemmas 3, 6 and 7.

Proof of Corollary 1. Suppose $\eta(x) = \eta' g(x)^\beta$ with $\eta' = \Theta\left(\sqrt{\frac{\beta (KS)^\beta}{\alpha d^\alpha T}} \right)$ and $\beta = \Theta\left(1 / \left(\log_+ \left(\frac{KS}{d}\right)\right)\right)$. We then have $R_T = O\left(\frac{\eta' d^\alpha T}{\beta} + \gamma T + \frac{K^\beta}{\eta' \alpha} \sum_{x \in \mathcal{X}} g(x)^\alpha \right) = O\left(\frac{\eta' d^\alpha T}{\beta} + \gamma T + \frac{(KS)^\beta}{\eta' \alpha} \right) = O\left(\sqrt{\frac{dT}{\alpha\beta} \left(\frac{KS}{d}\right)^\beta} + \lambda_0 (\eta')^{1/\alpha} T \right) = O\left(\sqrt{dT \log_+ \left(\frac{KS}{d}\right)} + \lambda_0 \sqrt{T^{1-\beta}} \right)$.

Suppose $\eta(x) = \eta = \Theta\left(\sqrt{\frac{\beta K^\beta}{\alpha d T}} \right)$ for all $x \in \mathcal{X}$ and $\beta = \Theta\left(\frac{1}{\log K} \right)$. We then have $R_T = O\left(\frac{\eta d T}{\beta} + \gamma T + \frac{K^\beta}{\eta \alpha} \right) = O\left(\sqrt{\frac{d K^\beta T}{\alpha \beta}} + \lambda_0 \min\left\{ \sqrt{\frac{\beta K^\beta T}{d}}, L^{\frac{1-\alpha}{\alpha}} \left(\frac{\beta K^\beta}{d} \right)^{\frac{1}{2\alpha}} T^{1 - \frac{1}{2\alpha}} \right\} \right) = O\left(\sqrt{dT \log K} + \lambda_0 \min\left\{ \sqrt{\frac{T}{d \log K}}, L^{\frac{1-\alpha}{\alpha}} \left(\frac{1}{d \log K} \right)^{\frac{1}{2\alpha}} \sqrt{T^{1-\beta}} \right\} \right) = O\left(\sqrt{dT \log K} + \lambda_0 \min\left\{ \sqrt{\frac{T}{d \log K}}, L^{\frac{\beta}{\alpha}} \sqrt{T^{1-\beta}} \right\} \right)$. □

3.4 Regret lower bound

The following theorem implies that the regret upper bound given in Corollary 1 achieved by the Algorithm 1 is tight for $S \leq K \leq 2^d$.

Theorem 4. *Suppose any $d' \geq 1$, $S \geq 1$ and $T = \Omega(d'^2 S)$. Then, for any algorithm for contextual linear bandit problems with $K = 2^{d'}$, $d = d'S$, and $|\mathcal{X}| = S$ there exists a problem instance for which $R_T = \Omega\left(d'\sqrt{ST}\right)$.*

This theorem implies a regret lower bound of $\Omega\left(\sqrt{dT \log_+\left(K \min\left\{1, \frac{S}{d}\right\}\right)}\right)$. To see this, we first note Theorem 4 implies that, if some d' satisfies $K \geq 2^{d'}$ and $d \geq d'S$, we can obtain a regret lower bound of $R_T = \Omega(d'\sqrt{ST})$. Let (d, K, S) be an arbitrary given parameter set that satisfies $K \leq 2^d \leq K^S$. We then have $\log_2 K \leq d \leq S \log_2 K$. Define $d' := \lfloor \log_2 K \rfloor$ and $S' := \lfloor d / \log_2 K \rfloor \leq S$. Then, as we have $K \geq 2^{d'}$ and $d \geq S' \log_2 K \geq S'd' = \Omega(d)$, from Theorem 4, we obtain a regret lower bound of $R_T = \Omega(d'\sqrt{S'T}) = \Omega(\sqrt{S'd'Td'}) = \Omega(\sqrt{dTd'})$. By combining this with $d' = \Omega(\log K) = \Omega\left(\log_+\left(K \min\{1, \frac{S}{d}\}\right)\right)$, we obtain $R_T = \Omega\left(\sqrt{dT \log_+\left(K \min\{1, \frac{S}{d}\}\right)}\right)$. We hence have the following lower bound:

Corollary 2. *For any (d, K, S) such that $K \leq 2^d \leq K^S$ and for any algorithm for contextual linear bandit problems, there exists a problem instance for which $R_T = \Omega\left(\sqrt{dT \log_+\left(K \min\left\{1, \frac{S}{d}\right\}\right)}\right)$.*

Proof of Theorem 4. Theorem 4 can be shown by using the result of Dani et al. [2008, Theorem 3]. They provide a lower bound of $\Omega(d\sqrt{T})$ for (non-contextual) linear bandit problems with $K = 2^d$. We use S copies of their problem instance of the dimensionality d', to prove Theorem 4. The context X_t is drawn from uniform distribution over $\mathcal{X} = [S]$, each element of which corresponds to one of the copies of the linear bandit instance. Then, for any $x \in \mathcal{X}$, the number of rounds $t \leq T$ at which $X_t = x$ is of $\Omega(T/S)$ with a probability at least $1/2$. Hence, the expected cumulative regret suffered for rounds at which $X_t = x$ is of $\Omega(d'\sqrt{T/S})$ for each $x \in \mathcal{X}$. By summing this for all $x \in \mathcal{X}$, we obtain the lower bound of $\Omega(Sd'\sqrt{T/S}) = \Omega(d'\sqrt{ST})$. Note that features mapping ϕ need to be designed so that $\phi(x, i)$ and $\phi(x', i')$ are orthogonal for any $x \neq x'$ and $i, i' \in [K]$. We can satisfy this condition by setting the dimension of the entire feature space to $d = d'S$. □

4 Conclusion

In this study, we investigated the minimax regret in the contexts of the multi-armed bandit with expert advice and contextual linear bandit problems. For the former, we established a regret lower bound of $O(\sqrt{KT \log \frac{N}{K}})$ in the setting where the player selects an expert before observing expert advice. This bound matches, up to a constant factor, the upper bound provided by Kale [2014]. Additionally, for the contextual linear bandit problem, we proposed an algorithm that achieves a regret upper bound of $O(\sqrt{dT \log(K \min\{1, \frac{S}{d}\})})$. As illustrated in Table 1, this upper bound aligns with the lower bound under certain conditions on the parameters (K, S, D).

Remaining challenges in the problem with expert advice include establishing similar lower bounds when the player can observe expert advice before decision-making in each round. Furthermore, determining the minimax regret in a broader parameter setting for (contextual) linear bandits remains an open problem. Relaxing Assumption 1 on the contextual distribution is also an important direction for enhancing practical applicability.

References

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

J.-Y. Audibert and S. Bubeck. Minimax policies for adversarial and stochastic bandits. In *Conference on Learning Theory*, pages 217–226, 2009.

P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire. The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing*, 32(1):48–77, 2002.

S. Bubeck, N. Cesa-Bianchi, and S. M. Kakade. Towards minimax policies for online linear optimization with bandit feedback. In *Conference on Learning Theory*, pages 41.1–41.14, 2012.

N. Cesa-Bianchi and G. Lugosi. Combinatorial bandits. *Journal of Computer and System Sciences*, 78(5):1404–1422, 2012.

H. Chen, Y. He, and C. Zhang. On interpolating experts and multi-armed bandits. In *International Conference on Machine Learning*, pages 6776–6802, 2024.

W. Chu, L. Li, L. Reyzin, and R. Schapire. Contextual bandits with linear payoff functions. In *International Conference on Artificial Intelligence and Statistics*, pages 208–214, 2011.

V. Dani, S. M. Kakade, and T. Hayes. The price of bandit information for online optimization. *Advances in Neural Information Processing Systems*, 20, 2007.

V. Dani, T. P. Hayes, and S. M. Kakade. Stochastic linear optimization under bandit feedback. In *Conference on Learning Theory*, 2008.

K. Eldowa, E. Esposito, T. Cesari, and N. Cesa-Bianchi. On the minimax regret for online learning with feedback graphs. *Advances in Neural Information Processing Systems*, 36, 2024.

O. Hanna, L. Yang, and C. Fragouli. Efficient batched algorithm for contextual linear bandits with large action space via soft elimination. *Advances in Neural Information Processing Systems*, 36, 2024.

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

S. Ito, T. Tsuchiya, and J. Honda. Adaptive learning rate for follow-the-regularized-leader: Competitive analysis and best-of-both-worlds. In *Conference on Learning Theory*, pages 2522–2563, 2024.

S. Kale. Multiarmed bandits with limited expert advice. In *Conference on Learning Theory*, pages 107–122, 2014.

Y. Kuroki, A. Rumi, T. Tsuchiya, F. Vitale, and N. Cesa-Bianchi. Best-of-both-worlds algorithms for linear contextual bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 1216–1224, 2024.

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

H. Liu, C.-Y. Wei, and J. Zimmert. Bypassing the simulator: Near-optimal adversarial linear contextual bandits. *Advances in Neural Information Processing Systems*, 36, 2024.

G. Neu and J. Olkhovskaya. Efficient and robust algorithms for adversarial linear contextual bandits. In *Conference on Learning Theory*, pages 3049–3068, 2020.

Q. M. Nguyen and N. Mehta. Near-optimal per-action regret bounds for sleeping bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 2827–2835, 2024.

J. Olkhovskaya, J. Mayo, T. van Erven, G. Neu, and C.-Y. Wei. First-and second-order bounds for adversarial linear contextual bandits. *Advances in Neural Information Processing Systems*, 36, 2024.

Y. Seldin and G. Lugosi. A lower bound for multi-armed bandits with expert advice. In *13th European Workshop on Reinforcement Learning (EWRL)*, 2016.

A Auxiliary lemma

Lemma 8. *For any positive-definite matrix $A \in \mathbb{R}^{d\times d}$ and a real vector $x \in \mathbb{R}^d$, if $A \succeq xx^\top$, it holds that $xA^{-1}x \le 1$.*

Proof. From the assumption of $A \succeq xx^\top$, we have

$$x^\top A^{-1}x = (A^{-1}x)^\top A(A^{-1}x) \ge (A^{-1}x)^\top xx^\top (A^{-1}x) = (x^\top A^{-1}x)^2,$$

which implies $x^\top A^{-1}x \le 1$. □

B Omitted lemmas and proofs in Section 2

Lemma 9. *Suppose $m = 1$. It holds for any $N', T^* \in \mathbb{N}$ and $\varepsilon \in [0, 1/2]$ that*

$$D_{\mathrm{KL}}\left(\mathcal{D}_{T^*}(\varepsilon, 1) \mid\mid \mathcal{D}_{T^*}(\varepsilon)\right) \le \ln\left(\frac{N' - 1 + \left(1 + 4\varepsilon^2\right)^{T^*}}{N'}\right) \le \frac{\left(1 + 4\varepsilon^2\right)^{T^*} - 1}{N'}.$$

Consequently, if $T^* \le \frac{\ln(N'/4)}{4\varepsilon^2}$, then the above value of the KL divergence is at most $1/4$.

Proof. For (e, ℓ), let $b_1 \in \{0, 1\}$, $(c_v)_{v\in[N']} \in \{0, 1\}^{N'}$ be binary variables such that $\ell((1, 0)) = b_u$, $\ell((1, 1)) = 1 - b_u$, and $e((1, v)) = |b_u - c_v|$. Then, all the values of (e, ℓ) are determined by $\ell(0), b_1$ and $(c_v)_{v\in[N']}$. When (e, ℓ) follows $\mathcal{D}(\varepsilon)$, then $(\ell(0), b_1, (c_v)_{v\in[N']})$ follows $\mathcal{E}_0 := \mathrm{Ber}((1-\varepsilon)/2) \times \mathrm{Ber}(1/2) \times (\mathrm{Ber}(1/2))^{N'}$. On the other hand, when (e, ℓ) follows $\mathcal{D}(\varepsilon, 1, v^*)$ then $(\ell(0), b_1, (c_v)_{v\in[N']})$ follows $\mathcal{E}_{v^*} := \mathrm{Ber}((1 - \varepsilon)/2) \times \mathrm{Ber}(1/2) \times \mathcal{F}_{N'}(\varepsilon, v^*)$, where we define $\mathcal{F}_{v^*} = (\mathrm{Ber}(1/2))^{v^*-1} \times \mathrm{Ber}(1/2 + \varepsilon) \times (\mathrm{Ber}(1/2))^{N'-v^*}$. That is, if $(c_v)_{v\in[N']}$ follows $\mathcal{F}_{v^*}$, the v^*-th element c_{v^*} follows $\mathrm{Ber}(1/2)$ and the other elements follow $\mathrm{Ber}(1/2)$ independently. Hence, from the data processing inequality, recalling the definition of $\mathcal{D}_{T^*}(\varepsilon)$ and $\mathcal{D}_{T^*}(\varepsilon, v^*)$ given in Section 2.3, we obtain

$$\begin{aligned} D_{\mathrm{KL}}\left(\mathcal{D}_{T^*}(\varepsilon, 1) \mid\mid \mathcal{D}_{T^*}(\varepsilon)\right) &\le D_{\mathrm{KL}}\left(\frac{1}{N'}\sum_{v^*\in N'} \left(\mathcal{E}_{v^*}\right)^{T^*} \mid\mid \left(\mathcal{E}_0\right)^{T^*}\right) \\ &= D_{\mathrm{KL}}\left(\frac{1}{N'}\sum_{v^*\in N'} \left(\mathcal{F}_{v^*}\right)^{T^*} \mid\mid \left(\left(\mathrm{Ber}(1/2)\right)^{N'}\right)^{T^*}\right). \end{aligned} \qquad (12)$$

Let $p_0 : \{0, 1\}^{T^*\times N'} \to \mathbb{R}$ and $p_v : \{0, 1\}^{T^*\times N'}$ be the probability mass functions for $\left((\mathrm{Ber}(1/2))^{N'}\right)^{T^*}$ and $(\mathcal{F}_{v^*})^{T^*}$. Then, from the definition of the KL divergence, we have

$$\begin{aligned} \text{[RHS of (12)]} &= \frac{1}{N'}\sum_{v^*\in[N']} \mathop{\mathbf{E}}_{c\sim(\mathcal{F}_{v^*})^{T^*}}\left[\ln\left(\frac{1}{N'}\sum_{v\in[N']}\frac{p_v(c)}{p_0(c)}\right)\right] \\ &\le \frac{1}{N'}\sum_{v^*\in[N']} \ln\left(\frac{1}{N'}\sum_{v\in[N']} \mathop{\mathbf{E}}_{c\sim(\mathcal{F}_{v^*})^{T^*}}\left[\frac{p_v(c)}{p_0(c)}\right]\right), \end{aligned} \qquad (13)$$

where we used Jensen's inequality and the fact that $\ln(x)$ is a concave function. The ratio $\frac{p_{v^*}(c)}{p_0(c)}$ of probabilities can be expressed as

$$\frac{p_v(c)}{p_0(c)} = \prod_{t\in[T^*]} \left(\mathbf{1}[c_{tv} = 0] \cdot (1 - 2\varepsilon) + \mathbf{1}[c_{tv} = 1] \cdot (1 + 2\varepsilon)\right) = \prod_{t\in[T^*]} \left(1 + (4c_{tv} - 2)\varepsilon\right).$$

Hence, if $v \neq v^*$ then

$$\mathop{\mathbf{E}}_{c\sim(\mathcal{F}_{v^*})^{T^*}}\left[\frac{p_v(c)}{p_0(c)}\right] = \prod_{t\in[T^*]} \mathop{\mathbf{E}}_{(c_{tv})\sim\mathrm{Ber}(1/2)} \left(1 + (4c_{tv} - 2)\varepsilon\right) = 1,$$

where we used the condition that elements of (c_{tv}) are independent. Further, if $v = v^*$, we then have

$$\mathop{\mathbf{E}}_{c \sim (\mathcal{F}_{v^*})^{T^*}} \left[\frac{p_v(c)}{p_0(c)} \right] = \prod_{t \in [T^*]} \mathop{\mathbf{E}}_{(c_{tv}) \sim \mathrm{Ber}(1/2+\varepsilon)} (1 + (4c_{tv} - 2)\varepsilon) = \left(1 + 4\varepsilon^2\right)^T .$$

Therefore, we have

$$[\text{RHS of (13)}] = \frac{1}{N'} \sum_{v^* \in [N']} \ln \left(\frac{N' - 1 + \left(1 + 4\varepsilon^2\right)^{T^*}}{N'} \right) = \ln \left(\frac{N' - 1 + \left(1 + 4\varepsilon^2\right)^{T^*}}{N'} \right),$$

which completes the proof. □

B.1 Proof of Theorem 1

Proof. For each $u \in [m]$, let $\mathcal{T}_u(\mathcal{A}, \mathcal{D}(\varepsilon))$ denote the expected value of the number of rounds in which the algorithm $\mathcal{A}$ observes $\ell_t((u, 0))$ or $\ell_t((u, 1))$ before termination. Then, we will show that $\mathcal{T}_u(\mathcal{A}, \mathcal{D}(\varepsilon)) \geq T^*/2$ under the assumption that $\mathcal{A}$ is an $(\varepsilon/2, 0.05)$-PAC algorithm.

Fix any $u \in [m]$ and suppose that $\mathcal{A}$ is an $(\varepsilon/2, 0.05)$-PAC BEI algorithm for problems with general m. Then we can construct an $(\varepsilon/2, 0.05)$-PAC BEI algorithm $\mathcal{A}'$ for problems with $m = 1$ such that $\mathcal{T}(\mathcal{A}', \mathcal{D}'(\varepsilon)) \leq \mathcal{T}_u(\mathcal{A}, \mathcal{D}(\varepsilon))$, where $\mathcal{D}'(\varepsilon)$ and $\mathcal{D}(\varepsilon)$ represent distributions with $(m = 1, N' = N')$ and $(m = m, N' = N')$, respectively. To show this, consider the following procedure for solving BEI for $m = 1$ based on $\mathcal{A}$: Run algorithm $\mathcal{A}$. When algorithm $\mathcal{A}$ pulls an arm a_t other than $(u, 0)$ or $(u, 1)$, then generate e_t from $\mathcal{D}(\varepsilon)$ and $\ell(a_t)$ from $\mathrm{Ber}(1/2)$ independently of e_t, and feed them to $\mathcal{A}$. When algorithm $\mathcal{A}$ pulls $(u, 0)$ or $(u, 1)$, then query the instance of $m = 1$ and feed the observed losses to $\mathcal{A}$. Then, this construction provide an $(\varepsilon/2, 0.05)$-PAC BEI algorithm for the instance of $m = 1$, and if the instance is associated with $\mathcal{D}'(\varepsilon)$, the number of queries to this distributions has the expected value of $\mathcal{T}_u(\mathcal{A}, \mathcal{D}(\varepsilon))$. We hence have $\mathcal{T}(\mathcal{A}', \mathcal{D}'(\varepsilon)) \leq \mathcal{T}_u(\mathcal{A}, \mathcal{D}(\varepsilon))$.

From Lemma 2, we have $\mathcal{T}(\mathcal{A}', \mathcal{D}'(\varepsilon)) \geq T^*/2$. As we have $\mathcal{T}(\mathcal{A}, \mathcal{D}(\varepsilon)) \geq \sum_{u \in [m]} \mathcal{T}_u(\mathcal{A}, \mathcal{D}(\varepsilon)) \geq \sum_{u \in [m]} \mathcal{T}(\mathcal{A}', \mathcal{D}'(\varepsilon))$, we obtain $\mathcal{T}(\mathcal{A}, \mathcal{D}(\varepsilon)) \geq mT^*/2$. □

C Omitted proofs in Section 3

C.1 Proof of Lemma 3

Proof. Fix $\pi^* : \mathcal{X} \to [K]$. From the definitions of p_t in (2) and X_0, and the fact that $\hat{\theta}_t$ is an unbiased estimator of θ_t as shown in (4), we have

$$\begin{aligned}
R_T(\pi^*) &= \mathbf{E}\left[\sum_{t=1}^{T} \langle \theta_t, \phi(X_t, I_t) - \phi(X_t, \pi^*(X_t)) \rangle \right] \\
&= \mathbf{E}\left[\sum_{t=1}^{T} \left\langle \theta_t, \sum_{i \in [K]} p_t(X_t, i)\phi(X_t, i) - \phi(X_t, \pi^*(X_t)) \right\rangle \right] \\
&\leq \mathbf{E}\left[\sum_{t=1}^{T} \left\langle \theta_t, \sum_{i \in [K]} q_t(X_t, i)\phi(X_t, i) - \phi(X_t, \pi^*(X_t)) \right\rangle \right] + 2\gamma T \\
&= \mathbf{E}\left[\sum_{t=1}^{T} \left\langle \theta_t, \sum_{i \in [K]} q_t(X_0, i)\phi(X_0, i) - \phi(X_0, \pi^*(X_0)) \right\rangle \right] + 2\gamma T \\
&= \mathbf{E}\left[\sum_{t=1}^{T} \left\langle \hat{\theta}_t, \sum_{i \in [K]} q_t(X_0, i)\phi(X_0, i) - \phi(X_0, \pi^*(X_0)) \right\rangle \right] + 2\gamma T \\
&= \mathbf{E}\left[\sum_{t=1}^{T} \left(\left\langle \hat{\ell}_t, q_t(X_0) \right\rangle - \hat{\ell}_t(\pi^*(X_0)) \right) \right] + 2\gamma T. \qquad (14)
\end{aligned}$$

From [Lattimore and Szepesvári, 2020, Theorem 28.5], for any $x \in \mathcal{X}$, $(q_t(x))_{t \in [T]}$ defined by the FTRL procedure (2) satisfies

$$\begin{aligned} &\sum_{t=1}^{T} \left(\left\langle \hat{\ell}_t, q_t(X_0) \right\rangle - \hat{\ell}_t(\pi^*(X_0)) \right) \\ &\leq \sum_{t=1}^{T} \left(\left\langle \hat{\ell}_t(x), q_t(x) - q_{t+1}(x) \right\rangle - \frac{1}{\eta(x)} D(p_{t+1}(x), p_t(x)) \right) - \frac{\psi(q_1(x))}{\eta(x)}. \end{aligned} \tag{15}$$

Combining (14) and (8) with the fact that $\psi(q_1(x)) = \psi(\mathbf{1}/K) = -\frac{K^\beta - 1}{\alpha}$, we obtain the desired inequality. □

C.2 Proof of Lemma 5

Proof. From the definition of p_t in (2), as we have $V(p_t) \succeq \gamma V(p_0)$, we have

$$\phi(x,i)^\top V(p_t)^{-1} \phi(x,i) \leq \phi(x,i)^\top (\gamma V(p_0))^{-1} \phi(x,i) = \frac{1}{\gamma} \phi(x,i)^\top V(p_0)^{-1} \phi(x,i) \leq \frac{\lambda(p_0)}{\gamma} = \frac{\lambda_0}{\gamma}. \tag{16}$$

Further, as we have $V(p_t) \succeq (1-\gamma) V(q_t) \succeq (1-\gamma) g(x) q_t(x,i) \phi(x,i) \phi(x,i)^\top$, from Lemma 8, we have $\phi(x,i)^\top V(p_t)^{-1} \phi(x,i) \leq \frac{1}{(1-\gamma) g(x) q_t(x,i)}$. This, together with (16), yields (6). From the definition (3) of $\hat{\theta}_t$, we have

$$\begin{aligned} \left| \hat{\ell}_t(x,i) \right| &= \left| \left\langle \hat{\theta}_t, \phi(x,i) \right\rangle \right| = \left| \ell_t \phi(X_t, I_t)^\top V(p_t) \phi(x,i) \right| \\ &\leq \sqrt{\phi(X_t, I_t)^\top V(p_t) \phi(X_t, I_t) \phi(x,i)^\top V(p_t) \phi(x,i)}, \end{aligned}$$

where we used Cauchy-Schwarz inequality and the assumption that $|\ell_t| \leq 1$. From this and (6), we have (7). □

NeurIPS Paper Checklist

1. **Claims**

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

   Answer: [Yes]

   Justification: Table in the introduction describes the problem settings and results in this paper.

   Guidelines:

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

2. **Limitations**

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

   Answer: [Yes]

   Justification: Assumptions and limitations are explained in Sections 2.1 and 3.1.

   Guidelines:

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

3. **Theory Assumptions and Proofs**

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

   Answer: [Yes]

Justification: Assumptions are provided in Sections 2.1 and 3.1. Proofs are provided except in cases where it is obvious from other lemmas/theorems combinations or from existing literature.

Guidelines:

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

4. **Experimental Result Reproducibility**

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

   Answer: [NA]

   Justification: This study does not involve numerical experiments.

   Guidelines:

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

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

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

Answer: [NA]

Justification: This study does not involve numerical experiments.

Guidelines:

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

6. **Experimental Setting/Details**

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

   Answer: [NA]

   Justification: This study does not involve numerical experiments.

   Guidelines:

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

7. **Experiment Statistical Significance**

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

   Answer: [NA]

   Justification: This study does not involve numerical experiments.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)

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

8. **Experiments Compute Resources**

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

   Answer: [NA]

   Justification: This study does not involve numerical experiments.

   Guidelines:

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

9. **Code Of Ethics**

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

   Answer: [Yes]

   Justification: The authors have reviewed the NeurIPS Code of Ethics.

   Guidelines:

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

10. **Broader Impacts**

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

    Answer: [NA]

    Justification: This paper focuses on theoretical research and is not expected to have any societal impact.

    Guidelines:

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

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

11. **Safeguards**

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

    Answer: [NA]

    Justification: This paper focuses on theoretical research and is not expected to have any societal impact.

    Guidelines:

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

12. **Licenses for existing assets**

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

    Answer: [NA]

    Justification: This paper does not use existing assets.

    Guidelines:

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

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

13. **New Assets**

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

    Answer: [NA]

    Justification: This paper does not release new assets.

    Guidelines:

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

14. **Crowdsourcing and Research with Human Subjects**

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

    Answer: [NA]

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

    Guidelines:

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

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

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

    Answer: [NA]

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

    Guidelines:

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