Improved Regret of Linear Ensemble Sampling

Harin Lee
Seoul National University
Seoul, South Korea
harinboy@snu.ac.kr

Min-hwan Oh
Seoul National University
Seoul, South Korea
minoh@snu.ac.kr

Abstract

In this work, we close the fundamental gap of theory and practice by providing an improved regret bound for linear ensemble sampling. We prove that with an ensemble size logarithmic in T, linear ensemble sampling can achieve a frequentist regret bound of $\widetilde{\mathcal{O}}(d^{3/2}\sqrt{T})$, matching state-of-the-art results for randomized linear bandit algorithms, where d and T are the dimension of the parameter and the time horizon respectively. Our approach introduces a general regret analysis framework for linear bandit algorithms. Additionally, we reveal a significant relationship between linear ensemble sampling and Linear Perturbed-History Exploration (LinPHE), showing that LinPHE is a special case of linear ensemble sampling when the ensemble size equals T. This insight allows us to derive a new regret bound of $\widetilde{\mathcal{O}}(d^{3/2}\sqrt{T})$ for LinPHE, independent of the number of arms. Our contributions advance the theoretical foundation of ensemble sampling, bringing its regret bounds in line with the best known bounds for other randomized exploration algorithms.

1 Introduction

Ensemble sampling [16] has emerged as an empirically effective randomized exploration technique in various online decision-making problems, such as online recommendation [17, 27, 26] and deep reinforcement learning [18–20]. Despite its popularity, the theoretical understanding of ensemble sampling has lagged behind, even for the linear bandit problem, with previous results revealing suboptimal outcomes. For instance, a prior work [21] demonstrated that linear ensemble sampling could achieve $\mathcal{O}(\sqrt{T})$ *Bayesian* regret with an ensemble size growing at least linearly with T. However, the requirement for the ensemble size to be linear in T is highly unfavorable and prohibitive in many practical settings. A recent work [10] showed that a symmetrized version of linear ensemble sampling could provide an improvement in dependence on ensemble size of $\Theta(d \log T)$ and show a frequentist regret bound of $\widetilde{\mathcal{O}}(d^{5/2}\sqrt{T})$. However, this regret bound clearly falls short of the existing frequentist regret achieved by standard randomized algorithms such as Thompson Sampling (TS) [4, 2] and Perturbed-History Exploration (PHE) [11–13].[1]

In this work, we close this fundamental gap by providing an improved regret bound for linear ensemble sampling. We prove that linear ensemble sampling with an ensemble size logarithmic in T can still attain a frequentist regret bound of $\widetilde{\mathcal{O}}(d^{3/2}\sqrt{T})$, marking the first time that linear ensemble sampling achieves a state-of-the-art result for randomized linear bandit algorithms. Our

[1] LinTS [4, 2] has regret of $\widetilde{\mathcal{O}}\big(\min(d^{3/2}\sqrt{T}, d\sqrt{T\log K})\big)$, and LinPHE [12] has regret of $\widetilde{\mathcal{O}}(d\sqrt{T\log K})$. For exchanges between factors of $\mathcal{O}(\sqrt{d})$ and $\mathcal{O}(\sqrt{\log K})$, we refer to the analysis of Agrawal and Goyal [4]. Therefore, the existing result for linear ensemble sampling [10] has a gap of at least $\mathcal{O}(d)$ compared to LinTS and LinPHE. Besides this gap in regret bounds, there appears to be rather counter-intuitive dependence on ensemble size in Janz et al. [10] (see Remark 2 in Section 4).

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

https://doi.org/10.52202/079017-2947

approach not only improves upon the regret bound but also simplifies the algorithm by avoiding the use of symmetrized perturbations, making it more practical for implementation. For regret analysis, we present a general, concise framework for analyzing linear bandit algorithms, which may be of independent interest. Furthermore, we rigorously reveal the significant relationship between ensemble sampling and PHE for the first time, showing that in the regime where the ensemble size equals T, linear PHE (LinPHE) is a special case of linear ensemble sampling. With this new insight, we can use the regret analysis for ensemble sampling to derive a new regret bound of $\widetilde{\mathcal{O}}(d^{3/2}\sqrt{T})$ for LinPHE, which is independent of the number of arms K.

Our main contributions are summarized as follows:

- We prove a $\widetilde{\mathcal{O}}(d^{3/2}\sqrt{T})$ regret bound (Theorem 1) for linear ensemble sampling with an ensemble size of $m = \Omega(K \log T)$, where K denotes the number of arms. Importantly, our regret bound does not depend on K or m even logarithmically. Our result is the first to establish $\widetilde{\mathcal{O}}(d^{3/2}\sqrt{T})$ regret for linear ensemble sampling with an ensemble size sublinear in T, improving the previous bound by the factor d while maintaining the ensemble size to be logarithmic in T.

- As part of the regret analysis, we present a general regret analysis framework (Theorem 2) for linear bandit algorithms. This framework not only generalizes the regret analysis of randomized algorithms such as ensemble sampling and PHE but also applies to other optimism-based deterministic algorithms. This result can be of independent interest beyond ensemble sampling.

- We rigorously investigate the relationship between linear ensemble sampling and LinPHE. We show that in the regime of ensemble size $m = T$, LinPHE is a special case of the linear ensemble sampling algorithm. To our best knowledge, this is the first result to show the equivalence between linear ensemble sampling and LinPHE.

- As a byproduct, with this new insight into the relationship between linear ensemble sampling and LinPHE, we provide an alternate analysis for LinPHE as an extension of the analysis for linear ensemble sampling, achieving a $\widetilde{\mathcal{O}}(d^{3/2}\sqrt{T})$ regret bound with no dependence on K.

1.1 Related Work

The stochastic linear bandit problem [5, 1, 14] is a foundational sequential decision-making problem and a core model for multi-armed bandits with features. Numerous algorithms have been developed for this problem, including deterministic approaches such as UCB-based methods [5, 8, 1] and randomized algorithms such as Thompson sampling [24, 7, 4, 2] and PHE [11–13].

Thompson sampling [24], a classical randomized method, utilizes the posterior distribution of hidden parameters based on observed data. Initially proposed for Bayesian settings [22, 23], it has also demonstrated strong performance in frequentist settings [3, 4, 2]. For the stochastic linear bandit, Agrawal and Goyal [4] showed that Thompson sampling with a Gaussian prior achieves a regret bound of $\widetilde{\mathcal{O}}(d^{3/2}\sqrt{T})$, which can be reduced to $\widetilde{\mathcal{O}}(d\sqrt{T \log K})$ for small K. However, applying Thompson sampling to more complex problems remains challenging, especially when posterior computation becomes intractable, though approximate methods have been proposed [16, 25].

PHE [11–13] is another class of randomized algorithms that does not rely on posterior distributions, making it potentially applicable to more complex settings. In the finite-armed linear bandit model, PHE achieves a $\widetilde{\mathcal{O}}(d\sqrt{T \log K})$ regret bound, matching the performance of Thompson sampling for finite arms [4]. However, the relationship between PHE and ensemble sampling remains unexplored in previous studies.

Ensemble sampling [16] has gained popularity as a randomized exploration method across various decision-making tasks [17, 27, 26, 18–20]. Despite its empirical success, its theoretical foundation, particularly for linear bandits, is still relatively underdeveloped. Qin et al. [21] showed that linear ensemble sampling achieves $\mathcal{O}(\sqrt{T})$ Bayesian regret but requires an impractically large ensemble size that scales linearly with T. More recently, Janz et al. [10] reduced the dependence on ensemble size to $\Theta(d \log T)$ and achieved a frequentist regret bound of $\widetilde{\mathcal{O}}(d^{5/2}\sqrt{T})$. However, the frequentist regret bound of $\widetilde{\mathcal{O}}(d^{3/2}\sqrt{T})$ for ensemble sampling has yet to be achieved.

Algorithm 1 Linear Ensemble Sampling

Input : regularization parameter $\lambda > 0$, ensemble size $m \in \mathbb{N}$,
initial perturbation distribution $\mathcal{P}_I$ on $\mathbb{R}^d$, reward-perturbation distribution $\mathcal{P}_R$ on $\mathbb{R}$,
ensemble sampling distribution $\{\mathcal{J}_t\}_{t=1}^T$ on $[m]$
Sample $W^j \sim \mathcal{P}_I$ for each $j \in [m]$.
Initialize $V_0 = \lambda I_d$, $S_0^j = W^j$, $\theta_0^j = V_0^{-1} S_0^j$ for each $j \in [m]$
for $t = 1, 2, \ldots T$ **do**
  Sample $j_t \sim \mathcal{J}_t$
  Pull arm $X_t = \operatorname{argmax}_{x \in \mathcal{X}} x^\top \theta_{t-1}^{j_t}$ and observe Y_t
  Update $V_t = V_{t-1} + X_t X_t^\top$
  for $j = 1, 2, \ldots, m$ **do**
    Sample $Z_t^j \sim \mathcal{P}_R$
    Update $S_t^j = S_{t-1}^j + X_t(Y_t + Z_t^j)$ and $\theta_t^j = V_t^{-1} S_t^j$
  end for
end for

2 Preliminaries

2.1 Notations

$\mathbb{N}$ denotes the set of natural numbers starting from 1. For a positive integer M, $[M]$ denotes the set $\{1, 2, \ldots, M\}$. $\mathbf{0}_d$ denotes the zero vector in $\mathbb{R}^d$ and I_d denotes the identity matrix in $\mathbb{R}^{d \times d}$. We define and work within a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is the sample space, $\mathcal{F}$ is the event set, and $\mathbb{P}$ is the probability measure. $\mathcal{N}(\mu, \Sigma)$ denotes the uni- or multi-variate Gaussian distribution with mean μ and covariance Σ. $\wedge$ denotes logical conjunction ("and") and $\vee$ denotes logical disjunction ("or"). With slight abuse of notation, we write $\{\omega \in \Omega : A\}$ and A interchangeably when A is some condition, for simplicity. $\mathcal{O}(\cdot)$ denotes the asymptotic growth rate with respect to problem parameters d, T, and K. $\widetilde{\mathcal{O}}(\cdot)$ further hides logarithmic factors of T and d.

2.2 Problem Setting

We consider the stochastic linear bandit problem. The learning agent is presented with a non-empty arm set $\mathcal{X} \subset \mathbb{R}^d$. For T time steps, where $T \in \mathbb{N}$ is the time horizon, the agent selects an arm $X_t \in \mathcal{X}$ and receives a real-valued reward Y_t, where the reward is generated based on a hidden true parameter vector, $\theta^* \in \mathbb{R}^d$. Specifically, Y_t is defined as follows:

$$Y_t = X_t^\top \theta^* + \eta_t \,,$$

where η_t is a zero-mean random noise. The objective of the agent is to maximize the cumulative reward, or equivalently, to minimize the cumulative regret $R(T)$ defined as

$$R(T) := \sum_{t=1}^{T} \left(\sup_{x \in \mathcal{X}} x^\top \theta^* - X_t^\top \theta^* \right) .$$

3 Ensemble Sampling for Linear Bandits

Algorithm 1 describes linear ensemble sampling. The learner maintains an ensemble of m estimators, where each estimator fits perturbed rewards. For the j-th estimator, a random vector $W^j \in \mathbb{R}^d$ acts as an initial perturbation on the estimator, and a random variable Z_t^j perturbs the reward at time t. Specifically, θ_t^j is the solution of the following minimization problem:

$$\underset{\theta \in \mathbb{R}^d}{\text{minimize}} \, \lambda \left\| \theta - W^j/\lambda \right\|_2^2 + \sum_{i=1}^{t} \left(X_i^\top \theta - \left(Y_i + Z_i^j \right) \right)^2 \tag{1}$$

When selecting an arm, one of the m estimators is chosen according to an ensemble sampling distribution $\mathcal{J}_t$ and acts greedily with respect to the sampled estimator. Previous ensemble sampling

https://doi.org/10.52202/079017-2947

algorithms [16, 21, 10] sample the estimators uniformly from the ensemble, but we allow any policy for selecting the estimator. Further distinguishing from the algorithm presented in Janz et al. [10], we do not sample Rademacher random variables for symmetrization, making our algorithm simpler. Ensemble sampling is capable of being generalized to complex settings whenever solving minimization problem (1) is tractable. Especially when incremental updates of the minimization problem are cheap, for instance with neural networks or other gradient descent-based models, ensemble sampling can be an efficient exploration strategy. The algorithm may simply store an ensemble of models, sample one to select an action, and then update the models incrementally based on the observed reward and generated perturbation.

4 Regret Bound of Linear Ensemble Sampling

Before we present the regret bound of linear ensemble sampling (Algorithm 1), we present the following standard assumptions on the problem structure.

Assumption 1 (Arm set and parameter)**.** *$\mathcal{X}$ is closed and for all $x \in \mathcal{X}$, $\|x\|_2 \leq 1$. There exists $S > 0$ such that $\|\theta^*\|_2 \leq S$. Both bounds are known to the agent.*

Remark 1. Under Assumption 1, $\mathcal{X}$ is a compact set. Therefore, we can define $x^* := \mathrm{argmax}_{x \in \mathcal{X}} x^\top \theta^*$ and rewrite the definition of $R(T)$ as $\sum_{t=1}^{T} x^{*\top}\theta^* - X_t^\top \theta^*$.

For a rigorous statement of the second assumption, we define several filtrations. For $t \in [T] \cup \{0\}$, let $\mathcal{F}_t^X := \sigma(X_1, \dots X_t)$ and $\mathcal{F}_t^\eta := \sigma(\eta_1, \dots, \eta_t)$ be the σ-algebras generated by X_i and η_i up to time t respectively. We also define the σ-algebra generated by the algorithm's internal randomness up until the choice of X_t as $\mathcal{F}_t^A$. Let $\mathcal{F}_t := \sigma\left(\mathcal{F}_t^A \cup \mathcal{F}_t^X \cup \mathcal{F}_t^\eta\right)$ be the σ-algebra generated by the first t iterations of the interaction between the environment and the agent. In addition, let $\mathcal{F}_t^- := \sigma\left(\mathcal{F}_t^A \cup \mathcal{F}_t^X \cup \mathcal{F}_{t-1}^\eta\right)$ be the σ-algebra generated in the same way as $\mathcal{F}_t$, but excluding η_t.

Assumption 2 (Noise)**.** *There exists $\sigma \geq 0$ such that η_t is $\mathcal{F}_t^-$-conditionally σ-subGaussian for all $t \in [T]$, i.e., $\mathbb{E}\left[\exp(s\eta_t)|\mathcal{F}_t^-\right] \leq \exp\left(\sigma^2 s^2/2\right)$ holds almost surely for all $s \in \mathbb{R}$.*

Now, we define a value β_t to describe the variance of the generated perturbation values. Define a sequence $\{\beta_t(\delta)\}_{t=0}^\infty$ as $\beta_t(\delta) := \sigma\sqrt{d\log\left(1 + \frac{t}{d\lambda}\right) + 2\log\frac{1}{\delta}} + \sqrt{\lambda}S$. We may omit δ when its value is clear from the context. The definition of β_t comes from Abbasi-Yadkori et al. [1] as a confidence radius of the ridge estimator, which we later specify in Lemma 1. We now present the regret bound of linear ensemble sampling (Algorithm 1).

Theorem 1 (Regret bound of linear ensemble sampling)**.** *Fix $\delta \in (0, 1]$. Assume $|\mathcal{X}| = K < \infty$ and run Algorithm 1 with $\lambda \geq 1$, $m \geq C(K\log T + \log\frac{1}{\delta})$, $\mathcal{P}_I = \mathcal{N}(\mathbf{0}_d, \lambda\beta_T^2 I_d)$, $\mathcal{P}_R = \mathcal{N}(0, \beta_T^2)$, and $\mathcal{J}_t = Unif(m)$, where C is a universal constant and $Unif(m)$ denotes the uniform distribution over $[m]$. Then, with probability at least $1 - 4\delta$, the cumulative regret of Algorithm 1 is*

$$R(T) = \mathcal{O}\left((d\log T)^{\frac{3}{2}}\sqrt{T}\right).$$

Discussion of Theorem 1. Theorem 1 shows that Algorithm 1 achieves a $\widetilde{\mathcal{O}}(d^{3/2}\sqrt{T})$ frequentist regret bound with an ensemble size of $m = \Omega(K\log T)$. Importantly, our regret bound does not depend on K or m even logarithmically. Hence, this regret bound matches the state-of-the-art frequentist regret bound of linear Thompson sampling [4, 2]. Our result is the first to establish $\widetilde{\mathcal{O}}(d^{3/2}\sqrt{T})$ regret for linear ensemble sampling with an ensemble size sublinear in T, improving the previous bound by the factor d compared to the existing result in Janz et al. [10]. We conjecture that $\widetilde{\mathcal{O}}(d^{3/2}\sqrt{T})$ regret is highly likely to be the best bound for linear ensemble sampling based on the negative result in Hamidi and Bayati [9] for LinTS.[2] Comparing with the algorithm in Janz et al. [10], our version of linear ensemble sampling algorithm does not utilize Rademacher random variable for symmetrized perturbation. This allows our algorithm to be simpler than that of Janz et al. [10]. Partially due to this algorithmic difference, our regret analysis is quite distinct from the analysis of Janz et al. [10] (see the proof in Section 5.2).

[2]Hamidi and Bayati [9] have shown that LinTS without the posterior variance inflation by $\sqrt{d}$ factor (compared to optimism-based algorithms such as LinUCB or OFUL [1]) can lead to a linear regret in T. That is, the frequentist regret bound of $\widetilde{\mathcal{O}}(d^{3/2}\sqrt{T})$ for LinTS is the best one can derive for the algorithm.

Table 1: Comparison of regret bounds for linear ensemble sampling

Paper	Frequentist / Bayesian	Regret Bound	Ensemble Size
Lu and Van Roy [16]	Frequentist	Invalid	Invalid
Qin et al. [21]	Bayesian	$\widetilde{\mathcal{O}}(\sqrt{dT \log K})$	$\Omega(KT)$
Janz et al. [10]	Frequentist	$\widetilde{\mathcal{O}}(d^{5/2}\sqrt{T})$	$\Theta(d \log T)$
This work	Frequentist	$\widetilde{\mathcal{O}}(d^{3/2}\sqrt{T})$	$\Omega(K \log T)$

As in Lu and Van Roy [16] and Qin et al. [21], we study the finite-armed problem setting. Both studies analyze the excess regret of ensemble sampling compared to Thompson sampling through an information theoretical approach. However, direct comparisons of the regret bounds are non-trivial. The analysis by Lu and Van Roy [16] includes an error admitted by the authors and Qin et al. [21] analyze the Bayesian regret, which is a weaker notion of regret than the frequentist regret that we analyze in this work. Along with Janz et al. [10], our result also makes a progress in reducing the size of the ensemble compared to Lu and Van Roy [16] and Qin et al. [21]. The size of the ensemble required by Janz et al. [10] is $\Theta(d \log T)$.This requirement implies that their ensemble size may be smaller than ours when K is larger than d and also allows K to be infinite. However, it is important to note that their resulting regret bound of $\widetilde{\mathcal{O}}(d^{5/2}\sqrt{T})$ is clearly sub-optimal compared to regret bounds of other randomized exploration algorithms. Theorem 1 achieves the tighter regret bound while simultaneously reducing the size of the ensemble.

Remark 2 (**Counter-intuitive dependence on ensemble size** in Janz et al. [10])**.** The regret bound in Janz et al. [10] actually grows super-linearly with the ensemble size, which is counter-intuitive. Their regret bound implies that as the ensemble size increases, the performance of the algorithm deteriorates. This fails to explain the superior empirical performance observed for ensemble sampling even with a large ensemble. On the contrary, our result in Theorem 1 does not show any performance degradation as the ensemble size increases.

Remark 3 (**Generalizability of perturbation distributions**)**.** We show that Gaussian distribution for perturbation is not essential. The only properties of the Gaussian distribution we utilize are its tail probability and anti-concentration property, stated as Lemma 4 and Fact 1 in Section 5.2. Therefore, any other distributions exhibiting similar behaviors can instead be adopted. In Appendix H, we rigorously demonstrate that any symmetric subGaussian distribution with lower-bounded variance can be employed, possibly at a cost of a constant factor. A large class of distributions, including uniform distribution, spherical distribution, Rademacher distribution, and centered binomial distribution with $p = 1/2$ satisfy this condition. This result can be of independent interest.

5 Analysis

5.1 General Regret Analysis for Linear Bandits

We begin by presenting a general regret bound for any algorithm that selects the best arm based on an estimated parameter. This result can be of independent interest. This general bound and analysis serve as a general framework that includes the regret analysis of linear ensemble sampling (Theorem 1).

Theorem 2 (General regret bound for linear bandit algorithm)**.** *Fix $T \in \mathbb{N}$. Assume that at each time step $t \in [T]$, the agent chooses $X_t = \operatorname{argmax}_{x \in \mathcal{X}} x^\top \theta_t$, where $\theta_t \in \mathbb{R}^d$ is chosen by the agent under some (either deterministic or random) policy. Let $\lambda > 0$ and $V_t = \lambda I + \sum_{i=1}^t X_i X_i^\top$. Let $\{\mathcal{E}_{1,t}\}_{t=1}^T$ and $\{\mathcal{E}_{2,t}\}_{t=1}^T$ be sequences of events that satisfy two conditions:*

1. *(Concentration) There exists a constant $\gamma > 0$ such that*
$$\|\theta_t - \theta^*\|_{V_{t-1}} \mathbb{1}\{\mathcal{E}_{1,t}\} \leq \gamma \tag{2}$$
holds almost surely for all $t \in [T]$.

2. *(Optimism) $\mathcal{E}_{2,t} \in \mathcal{F}_{t-1}$ holds and there exists a constant $p \in (0, 1]$ such that*
$$\mathbb{P}\left(\left(x^{*\top}\theta^* \leq X_t^\top \theta_t \text{ and } \mathcal{E}_{1,t}\right) \text{ or } \mathcal{E}_{2,t}^{\mathsf{C}} \mid \mathcal{F}_{t-1}\right) \geq p \tag{3}$$
holds almost surely for all $t \in [T]$.

https://doi.org/10.52202/079017-2947

Take $\mathcal{E} = \cap_{t=1}^{T} (\mathcal{E}_{1,t} \cap \mathcal{E}_{2,t})$ and any $\delta \in (0, 1]$. Then, under the event $\mathcal{E}$ and an additional event whose probability is at least $1 - \delta$, the cumulative regret is bounded as follows:

$$R(T) \leq \gamma \left(1 + \frac{2}{p}\right) \sqrt{2dT \log\left(1 + \frac{T}{d\lambda}\right)} + \frac{\gamma}{p} \sqrt{\frac{2T}{\lambda} \log \frac{1}{\delta}}\,. \tag{4}$$

Discussion of Theorem 2. The audience well-versed in the regret analysis of randomized algorithms such as TS and PHE would recognize that bounding the regret using the probability of being optimistic is a standard procedure, also presented in Theorem 1 of Kveton et al. [12] and Theorem 2 of Janz et al. [10] as generalizations of the results in Agrawal and Goyal [4] and Abeille and Lazaric [2] respectively. However, our regret analysis offers much more concise approach than the existing techniques, which can be of independent interest beyond the analysis of ensemble sampling.

Theorem 2 states that if the two conditions, specifically *concentration* in (2) and *optimism* in (3), are met, then the algorithm achieves $\sqrt{T}$ regret. We provide the proof of Theorem 2 in Appendix B. Our proof technique generalizes the well-studied analysis of Abeille and Lazaric [2]. While their work poses conditions on a d-dimensional perturbation vector that is added to the ridge estimator, we do not assume the use of ridge regression nor we assume that the estimator is perturbed. Instead, we only pose conditions on the final estimator the algorithm exploits. Due to this generalization, Theorem 2 is even capable of inducing the regret bound of LinUCB [1], which always opts for an optimistic estimator, by setting $\gamma = 2\beta_T$ and $p = 1$ with appropriate concentration events assigned to $\mathcal{E}_{1,t}$ and $\mathcal{E}_{2,t}$. In addition, there are several improvements that simplify the proof which are worth noting. To exploit the optimism condition, we apply Markov's inequality on a well-defined random variable. The proof of Abeille and Lazaric [2] relies on defining a conditional distribution, conditioned on both history and the event of being optimistic. However, such distribution may not be well-defined if the probability of the event is 0 for given history. Janz et al. [10] try to solve this problem by separately handling such exceptional cases using conditional measures. However, their conditional measures depend on the random history, leading the probability p to be a random variable, which complicates the analysis. We also note that our proof does not require convex analysis studied in Abeille and Lazaric [2].

Remark 4 (Role of event $\mathcal{E}_{2,t}$)**.** Previous results that utilize the probability of being optimistic [4, 2, 12, 10] do not explicitly define events $\{\mathcal{E}_{2,t}\}_t$. However, their existence is crucial in our analysis of linear ensemble sampling. Since the perturbation sequences are also part of the history in ensemble sampling, the probability of θ_t being optimistic may be extremely small under some events in $\mathcal{F}_{t-1}$ that sample unfavorable sequences. The role of $\mathcal{E}_{2,t}$ is to confine our analysis to the case where such undesirable events do not occur.

5.2 Proof of Regret Bound in Theorem 1

We prove the regret bound of linear ensemble sampling stated in Theorem 1. To apply Theorem 2, high-probabilities of the sequences of events, namely $\{\mathcal{E}_{1,t}, \mathcal{E}_{2,t}\}_{t=1}^{T}$, should be guaranteed with an appropriate values of γ and p. We show that separate constraints can be imposed on the randomness of the rewards and the perturbations respectively to guarantee the probabilities of the events. We begin by decomposing the estimator into two parts: one that fits the observed rewards and the other that perturbs the estimator.

$$\begin{aligned}\theta_t^j = V_t^{-1} S_t^j &= V_t^{-1} \left(W^j + \sum_{i=1}^{t} X_i \left(Y_i + Z_i^j\right)\right) = V_t^{-1} \sum_{i=1}^{t} X_i Y_i + V_t^{-1} \left(W^j + \sum_{i=1}^{t} X_i Z_i^j\right) \\ &=: \hat{\theta}_t + \widetilde{\theta}_t^j\,,\end{aligned} \tag{5}$$

where we define $\hat{\theta}_t := V_t^{-1} \sum_{i=1}^{t} X_i Y_i$ and $\widetilde{\theta}_t^j := V_t^{-1}(W^j + \sum_{i=1}^{t} X_i Z_i^j)$. $\hat{\theta}_t$ is the ridge regression estimator of the observed data, and its randomness mainly comes from the noise of the rewards, $\{\eta_i\}_{i=1}^{t}$. $\widetilde{\theta}_t^j$ is the perturbation added to $\hat{\theta}_t$, and its randomness comes from the generated perturbation, W^j and $\{Z_i^j\}_{i=1}^{t}$. The following lemma states the well-known concentration result for the ridge estimator.

Lemma 1 (Theorem 2 of Abbasi-Yadkori et al. [1])**.** *Fix $\delta \in (0, 1]$. For $t \in \mathbb{N} \cup \{0\}$, define a sequence of events with $\beta_t(\delta) = \sigma \sqrt{d \log\left(1 + \frac{t}{d\lambda}\right) + 2 \log \frac{1}{\delta}} + \sqrt{\lambda} S$ as*

$$\hat{\mathcal{E}}_t := \left\{\omega \in \Omega : \left\|\hat{\theta}_t - \theta^*\right\|_{V_t} \leq \beta_t(\delta)\right\}$$

and their intersection $\hat{\mathcal{E}} := \bigcap_{t=0}^{\infty} \hat{\mathcal{E}}_t$. *Then,* $\mathbb{P}(\hat{\mathcal{E}}) \geq 1 - \delta$.

Now, we address $\widetilde{\theta}_t^j$. Define a *perturbation vector* that represents the perturbation sequence for each model as follows:

$$\mathbf{Z}_t^j := \begin{pmatrix} \frac{1}{\sqrt{\lambda}} W^{j\top} & Z_1^j & \cdots & Z_{t-1}^j \end{pmatrix}^\top \in \mathbb{R}^{d+t-1}, \forall j \in [m]\,. \tag{6}$$

Let $\mathbf{Z}_t := \mathbf{Z}_t^{j_t}$ so that $\mathbf{Z}_t$ is the perturbation vector of the model chosen at time t. The following lemma demonstrates that optimism condition (3) can be satisfied by an anti-concentration property of $\mathbf{Z}_t$ alone.

Lemma 2 (Sufficient condition for optimism)**.** *For* $t \in [T]$*, define a vector* $U_{t-1} \in \mathbb{R}^{d+t-1}$ *by*

$$U_{t-1}^\top := x^{*\top} V_{t-1}^{-1} \begin{pmatrix} \sqrt{\lambda} I_d & X_1 & \cdots & X_{t-1} \end{pmatrix}.$$

Then, $x^{*\top}\theta^* \leq X_t^\top \theta_t$ *holds whenever there exists a constant* $c > 0$ *such that* $U_{t-1}^\top \mathbf{Z}_t \geq c\|U_{t-1}\|_2$ *and* $\|\theta^* - \hat{\theta}_{t-1}\|_{V_{t-1}} \leq c$ *hold.*

We present a straightforward proof of Lemma 2 in Appendix C. We significantly deviate from the analyses of Abeille and Lazaric [2] and Janz et al. [10] in the method of guaranteeing the optimism condition. Their analyses require a d-dimensional perturbation vector to have a constant probability of having positive component, so-called *anti-concentrated*, in "every" possible direction in $\mathbb{R}^d$ since they only prove the existence of a direction that implies optimism. We observe and exploit the fact that it suffices to consider just "one" direction, specifically U_{t-1}, to produce an optimistic estimator. Since U_{t-1} depends only on the sequence of selected arms, the dependency between U_{t-1} and $\mathbf{Z}_t$ decouples when the dependency between $\{X_t\}_{t=1}^T$ and $\{\mathbf{Z}_t\}_{t=1}^T$ are decoupled, which we later achieve by taking the union bound in a unique way.

The following lemma shows that concentration and anti-concentration properties of the perturbation are sufficient conditions for Theorem 2. We provide a sketch of its proof and defer the remaining details to Appendix D.

Lemma 3. *Suppose the agent runs Algorithm 1 with some parameters. Fix* $\widetilde{\gamma} > 0$ *and* $p \in (0, 1]$*. For each* $t \in [T]$*, define two events*

$$\widetilde{\mathcal{E}}_{1,t} := \left\{\omega \in \Omega : \left\|\widetilde{\theta}_{t-1}^{j_t}\right\|_{V_{t-1}} \leq \widetilde{\gamma}\right\},$$
$$\widetilde{\mathcal{E}}_{2,t} := \left\{\omega \in \Omega : \mathbb{P}\left(U_{t-1}^\top \mathbf{Z}_t \geq \beta_{t-1} \|U_{t-1}\|_2 \text{ and } \widetilde{\mathcal{E}}_{1,t} \mid \mathcal{F}_{t-1}\right) \geq p\right\},$$

where U_{t-1} *is defined as in Lemma 2. Take* $\mathcal{E}_{1,t} = \widetilde{\mathcal{E}}_{1,t} \cap \hat{\mathcal{E}}_{t-1}$ *and* $\mathcal{E}_{2,t} = \widetilde{\mathcal{E}}_{2,t} \cap \hat{\mathcal{E}}_{t-1}$*. Then,* $\mathcal{E}_{1,t}$ *and* $\mathcal{E}_{2,t}$ *satisfy concentration condition* (2) *with* $\gamma = \widetilde{\gamma} + \beta_T$ *and optimism condition* (3) *with the same value of* p*. Consequently, with probability at least* $1 - 2\delta - \mathbb{P}(\widetilde{\mathcal{E}}^{\mathsf{C}})$*, where* $\widetilde{\mathcal{E}} := \cap_{t=1}^T (\widetilde{\mathcal{E}}_{1,t} \cap \widetilde{\mathcal{E}}_{2,t})$*, Algorithm 1 achieves regret bound* (4) *of Theorem 2.*

Remark 5. Lemma 3 applies to any perturbation-based algorithm that exploits $\theta_t = \hat{\theta}_{t-1} + \widetilde{\theta}_{t-1}$, where $\widetilde{\theta}_{t-1}$ is a linear transform of a random perturbation vector $\mathbf{Z}_t$. A version of linear Thompson sampling [2] and LinPHE also fall into this category.

Lemma 3 shifts the problem of constructing the regret bound of Algorithm 1 to lower-bounding the probabilities of the two sequences of events, $\{\widetilde{\mathcal{E}}_{1,t}\}_{t=1}^T$ and $\{\widetilde{\mathcal{E}}_{2,t}\}_{t=1}^T$. Note that $\widetilde{\mathcal{E}}_{1,t}$ and $\widetilde{\mathcal{E}}_{2,t}$ regard $\{X_i\}_{i=1}^{t-1}$ and $\mathbf{Z}_t$ only, and are independent of further randomness of $\{\eta_t\}_{t=1}^T$.

Sketch of Proof of Lemma 3. The concentration condition follows immediately by the triangle inequality. To show the optimism condition, we verify the following logical implication relationship:

$$\left(\left(U_{t-1}^\top \mathbf{Z}_t \geq \beta_{t-1} \|U_{t-1}\|_2\right) \wedge \widetilde{\mathcal{E}}_{1,t}\right) \vee \mathcal{E}_{2,t}^{\mathsf{C}} \Rightarrow \left(\left(x^{*\top}\theta^* \leq X_t^\top \theta_t\right) \wedge \mathcal{E}_{1,t}\right) \vee \mathcal{E}_{2,t}^{\mathsf{C}},$$

where Lemma 2 bridges the anti-concentration on the left hand side to the optimism on the right hand side. This implication relationship is converted to the following probability inequality:

$$\mathbb{P}\left(\left(\left(x^{*\top}\theta^* \leq X_t^\top \theta_t\right) \wedge \mathcal{E}_{1,t}\right) \vee \mathcal{E}_{2,t}^{\mathsf{C}} \mid \mathcal{F}_{t-1}\right) \geq \mathbb{P}\left(\left(\left(U_{t-1}^\top \mathbf{Z}_t \geq \beta_{t-1} \|U_{t-1}\|_2\right) \wedge \widetilde{\mathcal{E}}_{1,t}\right) \vee \mathcal{E}_{2,t}^{\mathsf{C}} \mid \mathcal{F}_{t-1}\right).$$

By the definition of $\widetilde{\mathcal{E}}_{2,t}$, the right hand side is bounded below by p, implying optimism condition (3). The probability of failure is bounded by the union bound and Lemma 1. □

https://doi.org/10.52202/079017-2947

Theorem 1 employs the Gaussian perturbation for a concrete instantiation of Algorithm 1. We define two values $\widetilde{\gamma}_T$ and γ_T, which serve as the confidence radii of $\widetilde{\theta}_t$ and θ_t for $t \in [T]$ under the Gaussian perturbation.

$$\widetilde{\gamma}_T := \beta_T \left(\sqrt{d \log \left(1 + \frac{T}{d\lambda}\right) + 2 \log \frac{2T}{\delta}} + \sqrt{d} + \sqrt{2 \log \frac{2T}{\delta}} \right), \qquad \gamma_T := \widetilde{\gamma}_T + \beta_T \,. \tag{7}$$

Note that in terms of d and T, both $\widetilde{\gamma}_T$ and γ_T are in $\mathcal{O}(d \log T)$. Lemma 4 illustrates the concentration result. Its proof is a simple application of Lemma 1, and is presented in Appendix E.

Lemma 4. *Suppose Algorithm 1 is run with parameters specified in Theorem 1. Fix $t \in [T]$ and $j \in [m]$. Suppose the sequence of arms $X_1, \ldots, X_t$ is chosen arbitrarily randomly, not necessarily by the the agent. Let $\widetilde{\theta}_t^j$ be defined as in Eq.* (5)*. Then, with probability at least $1 - \delta/T$, $\|\widetilde{\theta}_t^j\|_{V_t} \leq \widetilde{\gamma}_T$ holds.*

The following fact describes an anti-concentration property of Gaussian distribution, which follows from the fact that a linear combination of independent Gaussians is again Gaussian.

Fact 1. *If $Z \sim \mathcal{N}(0, \alpha^2 I_n)$ for some $\alpha \geq 0$ and $u \in \mathbb{R}^n$ is a fixed vector for some $n \in \mathbb{N}$, then $\mathbb{P}\left(u^\top Z \geq \alpha \|u\|_2\right) \geq \mathbb{P}(z \geq 1) =: p_N$, where $z \sim \mathcal{N}(0,1)$. We note that $p_N \geq 0.15$.*

All the building blocks we need to prove Theorem 1 is ready. The proof illustrates that the events specified in Lemma 3 occur with high probability.

Proof of Theorem 1. Define $\widetilde{\mathcal{E}}_{1,t}$ and $\widetilde{\mathcal{E}}_{2,t}$ as in Lemma 3 with $\widetilde{\gamma} = \widetilde{\gamma}_T$ and $p = p_N/4$, where $\widetilde{\gamma}_T$ is defined in Eq. (7) and p_N is defined in Fact 1. We show that these events occur with high probability, and the rest follows from Lemma 3. For the sake of the analysis, assume that $\delta/T \leq p_N/2 \approx 0.08$, which holds whenever $T \geq 14$ or $\delta < 0.07$.

Assume that the perturbation values W^j and Z_t^j are $\mathcal{F}_0^A$-measurable for all $j \in [m]$ and $t \in [T]$. An interpretation of this assumption is that the algorithm samples all the required values in advance. Note that we still obtain an equivalent algorithm and this modification need not actually take place in the execution. Under this assumption, the uniform sampling of $j_t \sim \mathcal{J}_t$ is the only source of randomness regarding the choice of θ_t and X_t when conditioned on the history $\mathcal{F}_{t-1}$. It may seem unintuitive, but this modification simplifies the proof because we only need to deal with j_t.

We first lower-bound the probability of $\widetilde{\mathcal{E}}_{1,t}$. When $j \in [m]$ is fixed, we can apply Lemma 4 and obtain $\mathbb{P}(\|\widetilde{\theta}_{t-1}^j\|_{V_{t-1}} \leq \widetilde{\gamma}_T) \geq 1 - \delta/T$. Since j_t is sampled independently of $\widetilde{\theta}_{t-1}^j$, it holds that

$$\mathbb{P}\left(\widetilde{\mathcal{E}}_{1,t}\right) = \sum_{j=1}^{m} \mathbb{P}\left(j_t = j, \left\|\widetilde{\theta}_{t-1}^j\right\|_{V_{t-1}} \leq \widetilde{\gamma}_T\right) = \sum_{j=1}^{m} \frac{1}{m} \mathbb{P}\left(\left\|\widetilde{\theta}_{t-1}^j\right\|_{V_{t-1}} \leq \widetilde{\gamma}_T\right) \geq 1 - \frac{\delta}{T} \,. \tag{8}$$

Now, we bound the probability of $\widetilde{\mathcal{E}}_{2,t}$. Fix $j \in [m]$. Recall that $\mathbf{Z}_t^j$ is the perturbation vector of the j-th model, defined in Eq. (6). The choice of Gaussian perturbation implies that $\mathbf{Z}_t^j \sim \mathcal{N}(\mathbf{0}_{d+t-1}, \beta_T^2 I_{d+t-1})$. Suppose that the sequence of arms $X_1, \ldots, X_T$ is fixed. Then, we can apply Fact 1, obtaining that $\mathbb{P}(U_{t-1}^\top \mathbf{Z}_t^j \geq \beta_{t-1} \|U_{t-1}\|_2) \geq p_N$, where the probability is measured over the randomness of the perturbation sequence $\mathbf{Z}_t^j$. Let $I_t^j := \mathbb{1}\{(U_{t-1}^\top \mathbf{Z}_t^j \geq \beta_{t-1}\|U_{t-1}\|_2) \wedge (\left\|\widetilde{\theta}_{t-1}^j\right\|_{V_{t-1}} \leq \widetilde{\gamma}_T)\}$. Then, we have that

$$\mathbb{P}\big(I_t^j = 1\big) \geq \mathbb{P}\big(U_{t-1}^\top \mathbf{Z}_t^j \geq \beta_{t-1}\|U_{t-1}\|_2\big) - \mathbb{P}\Big(\left\|\widetilde{\theta}_{t-1}^j\right\|_{V_{t-1}} > \widetilde{\gamma}_T\Big) \geq p_N - \delta/T \geq p_N/2 \,, \tag{9}$$

where the first inequality uses that $\mathbb{P}(A \cap B) \geq \mathbb{P}(A) - \mathbb{P}(B^{\mathsf{C}})$ holds for any events A and B, and the last inequality holds by the assumption $\delta/T \leq p_N/2$. However, as we assumed that the perturbation sequence is $\mathcal{F}_0^A$-measurable, it is $\mathcal{F}_{t-1}$-measurable, hence I_t^j is also $\mathcal{F}_{t-1}$-measurable. It means that the value of I_t^j is determined when the history up to time $t-1$ is fixed. The only remaining source of randomness in choosing θ_t conditioned on $\mathcal{F}_{t-1}$ is the sampling of $j_t \sim \mathcal{J}_t$. Therefore, it holds that

$$\mathbb{P}\left(\left(U_{t-1}^\top \mathbf{Z}_t^{j_t} \geq \beta_{t-1}\|U_{t-1}\|_2\right) \wedge \left(\left\|\widetilde{\theta}_{t-1}^{j_t}\right\|_{V_{t-1}} \leq \widetilde{\gamma}_T\right) \mid \mathcal{F}_{t-1}\right) = \frac{1}{m} \sum_{j=1}^{m} I_t^j \,. \tag{10}$$

Algorithm 2 Linear Perturbed-History Exploration (LinPHE)

Input : regularization parameter $\lambda > 0$, initial perturbation distribution $\mathcal{P}_I$ on $\mathbb{R}^d$, reward-perturbation distribution $\mathcal{P}_R$ on $\mathbb{R}$
Initialize $V_0 = \lambda I_d$
for $t = 1, 2, \dots T$ **do**
  Sample $W_t \sim \mathcal{P}_I$, $Z_{t,1}, \dots, Z_{t,t-1} \overset{i.i.d.}{\sim} \mathcal{P}_R$
  Update $\theta_t = V_{t-1}^{-1} \left(W_t + \sum_{i=1}^{t-1} X_i (Y_i + Z_{t,i}) \right)$
  Pull arm $X_t = \operatorname{argmax}_{x \in \mathcal{X}} x^\top \theta_t$, and observe Y_t
  Update $V_t = V_{t-1} + X_t X_t^\top$
end for

Since we have verified that the expectation of the right hand side is greater than $p_N/2$, Azuma-Hoeffding inequality (Lemma 12) implies that $\mathbb{P}(\frac{1}{m} \sum_{j=1}^{m} I_t^j < p_N/4) \leq \exp\left(-p_N^2 m/8\right)$. Recall that this result is obtained assuming that $X_1, \dots, X_T$ are fixed. We take the union bound over all possible sequences of arms. However, a naïve union bound multiplies K^T to the failure probability, which leads to an undesirable result of m scaling linearly with T. We present the following proposition inspired by an observation from Lu and Van Roy [16] that a permutation of selected arms can be regarded as equivalent. We note that the strong result of Lemma 6 in Lu and Van Roy [16] is not applicable to our setting since we do not assume that $\{\eta_t\}_{t=1}^T$ is distributed identically nor independently. Although we present all the main ideas to support Proposition 1 in this section, there may be a few points that readers find require further justification. Due to limited space, we provide a full, rigorous justification of Proposition 1 in Appendix F, where we present a different perspective on the sampling of perturbation.

Proposition 1. *There exists an event $\mathcal{E}_2^*$ such that under $\mathcal{E}_2^*$, $\frac{1}{m} \sum_{j=1}^{m} I_t^j \geq p_N/4$ holds for $t = 1, \dots, T$ and $\mathbb{P}(\mathcal{E}_2^{*\mathsf{C}}) \leq T^K \exp(-p_N^2 m/8)$.*

The key observation in Proposition 1 is that the perturbation vector consists of i.i.d. components, hence its distribution is invariant under independent permutations. Therefore, the distributions of $\widetilde{\theta}_{t-1}^j$ and $U_{t-1}^\top \mathbf{Z}_t^j$ remain invariant under the permutation of selected arms. Although the sequence of arms and the perturbation vector are not independent as a whole, the permutation that sorts the selected arms preserves the distribution of $\mathbf{Z}_t$ since X_t and Z_t are independent for all $t \in [T]$. The number of equivalence classes up to permutation over sequences of arms with lengths at most $T-1$ is less than T^K, since each arm can be selected 0 to $T-1$ times inclusively. Therefore, we take the union bound over the T^K sequences of arms and attain $\mathcal{E}_2^*$.
Taking $m = \frac{8}{p_N^2}(K \log T + \log \frac{1}{\delta})$, we obtain that $\mathbb{P}(\mathcal{E}_2^{*\mathsf{C}}) \leq \delta$. Eq. (10) implies that $\mathbb{P}(\cap_{t=1}^T \widetilde{\mathcal{E}}_{2,t}) \geq \mathbb{P}(\mathcal{E}_2^*) \geq 1 - \delta$. Therefore, by Lemma 3, with probability at least $1 - 4\delta$, the cumulative regret is bounded as follows:

$$R(T) \leq \gamma_T \left(1 + \frac{8}{p_N}\right) \sqrt{2dT \log\left(1 + \frac{T}{d\lambda}\right)} + \frac{4\gamma_T}{p_N} \sqrt{\frac{2T}{\lambda} \log \frac{1}{\delta}} = O\left((d \log T)^{\frac{3}{2}} \sqrt{T}\right).$$

□

6 Ensemble Sampling and Perturbed-History Exploration

In this section, we rigorously investigate the relationship between linear ensemble sampling and LinPHE. A generalized version of LinPHE is described in Algorithm 2. Note that the perturbed estimator, $\theta_t = V_{t-1}^{-1}\big(W_t + \sum_{i=1}^{t-1} X_i(Y_i + Z_{t,i})\big)$, resembles the estimator of linear ensemble sampling, which becomes evident when compared with Eq. (5). The main difference is that in LinPHE (Algorithm 2), the perturbation sequence is generated independently of the history at every time step, whereas in linear ensemble sampling (Algorithm 1), the sequence is not renewed but is incremented at each time step. However, we further observe that in linear ensemble sampling, as long as an estimator is not sampled for the arm selection, its perturbation sequence is independent of the selected arms and rewards. This implies that the estimator in the ensemble that is selected for the

https://doi.org/10.52202/079017-2947

first time is equivalent to the estimator computed by the policy of LinPHE. Specifically, if the j-th estimator is selected for the first time at time step t, then the perturbation values of the estimator, W^j and $\{Z_i^j\}_{i=1}^{t-1}$, have had no effect on previous interactions. Therefore, newly sampling them as $W_t^j \sim \mathcal{P}_I$ and $\{Z_{t,i}^j\}_{i=1}^{t-1} \overset{i.i.d.}{\sim} \mathcal{P}_R$, as in LinPHE, does not alter future interactions. We conclude that in the case where the ensemble size is greater than or equal to T, linear ensemble sampling becomes equivalent to LinPHE by selecting the estimators in a round robin.

Proposition 2. *Linear ensemble sampling (Algorithm 1) with $m = T$ and deterministic policy of choosing a model, e.g., $\mathcal{J}_t \equiv t$ for $t = 1, \ldots, T$, is equivalent to LinPHE (Algorithm 2).*

Proposition 2 shows that LinPHE is a special case of linear ensemble sampling and provides insightful consequences in both directions of the equivalence. To our best knowledge, Proposition 2 is the first result to formally demonstrate the relationship between linear ensemble sampling and LinPHE.

Linear ensemble sampling with T models is LinPHE: Since an ensemble of T models is equivalent to LinPHE which achieves a regret bound $\widetilde{\mathcal{O}}(d\sqrt{T \log K})$, the ensemble size larger than T is not necessary. This implication certainly emphasizes the sub-optimal requirements of the ensemble size in Lu and Van Roy [16], Qin et al. [21]. Even when $K > T$ in our problem setting, this equivalence provides the ground for upper bounding the ensemble size by T.

LinPHE is linear ensemble sampling with T models: Conversely, since LinPHE can be regarded as linear ensemble sampling with T models, it is possible to derive a regret bound of LinPHE by following the proof of Theorem 1. We present Corollary 1, which states a new regret bound of $\widetilde{\mathcal{O}}(d^{3/2}\sqrt{T})$ for LinPHE. Note that in this case, the regret bound is independent of K.

Corollary 1 (Regret bound of LinPHE)**.** *Fix $\delta \in (0, 1]$. Algorithm 2 with $\lambda \geq 1$, $\mathcal{P}_I = \mathcal{N}(\mathbf{0}_d, \lambda\beta_T^2 I_d)$ and $\mathcal{P}_R = \mathcal{N}(0, \beta_T^2)$ achieves $\mathcal{O}((d \log T)^{3/2}\sqrt{T})$ cumulative regret with probability at least $1 - 3\delta$.*

Discussion of Corollary 1. Kveton et al. [12] provide a $\widetilde{\mathcal{O}}(d\sqrt{T \log K})$ regret bound when the number of arms is finite. Our result is the first to prove that LinPHE achieves a $\widetilde{\mathcal{O}}(d^{3/2}\sqrt{T})$ regret bound that is independent of the number of arms. Hence, we view our overall results as a generalization of Kveton et al. [12]. It is widely observed that assuming the size of the arm set to be K may lead to an interchanging of a $\sqrt{d}$ factor with a $\sqrt{\log K}$ factor in the regret bound [4], although attaining such reduction may not always be done in a trivial manner [5, 8, 6]. Our focus is not merely on proving another regret bound for LinPHE, but rather on highlighting the close relationship between linear ensemble sampling and LinPHE, which, to our knowledge, has been overlooked in the literature.

The proof of Corollary 1 follows the proof of Theorem 1. Note that the latter part of the proof of Theorem 1 focuses on decoupling the dependency between $\{X_t\}_{t=1}^T$ and $\mathbf{Z}_t$. However, in the case of LinPHE, they are already independent since the perturbation sequence is freshly sampled at every time step, enabling a more elegant and concise proof. Especially, as it skips the parts that require the number of arms to be finite, for instance the use of Proposition 1, Corollary 1 holds even when the number of arms is infinite. The whole proof is presented in Appendix G.

7 Conclusion

We prove that linear ensemble sampling achieves a $\widetilde{\mathcal{O}}(d^{3/2}\sqrt{T})$ regret bound, marking the first such result in the frequentist setting and matching the best-known regret bound for randomized algorithms. The required ensemble size scales logarithmically with the time horizon as $\Omega(K \log T)$. Additionally, we expand our analysis to LinPHE, demonstrating that it is a special case of linear ensemble sampling with an ensemble of T models, achieving the same regret bound of $\widetilde{\mathcal{O}}(d^{3/2}\sqrt{T})$. While our work focuses on linear bandits, ensemble sampling applications have shown superior performance in more complex settings. This suggests that theoretical extensions beyond the linear setting are worth pursuing, with our results providing an important foundation for possibly understanding these extensions. Extending the results to general contextual settings, where the arm set may change over time with potentially non-linear reward functions, represents a promising direction for future work.

Acknowledgements

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2022R1C1C1006859, 2022R1A4A1030579, and RS-2023-00222663) and by AI-Bio Research Grant through Seoul National University.

References

[1] Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. *Advances in Neural Information Processing Systems*, 24:2312–2320, 2011.

[2] Marc Abeille and Alessandro Lazaric. Linear Thompson Sampling Revisited. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54, pages 176–184. PMLR, PMLR, 2017.

[3] Shipra Agrawal and Navin Goyal. Analysis of thompson sampling for the multi-armed bandit problem. In *Conference on learning theory*, pages 39–1. JMLR Workshop and Conference Proceedings, 2012.

[4] Shipra Agrawal and Navin Goyal. Thompson sampling for contextual bandits with linear payoffs. In *International conference on machine learning*, pages 127–135. PMLR, 2013.

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

[6] Sébastien Bubeck, Nicolo Cesa-Bianchi, and Sham M Kakade. Towards minimax policies for online linear optimization with bandit feedback. In *Conference on Learning Theory*, pages 41–1. JMLR Workshop and Conference Proceedings, 2012.

[7] Olivier Chapelle and Lihong Li. An empirical evaluation of thompson sampling. *Advances in neural information processing systems*, 24, 2011.

[8] Wei Chu, Lihong Li, Lev Reyzin, and Robert Schapire. Contextual bandits with linear payoff functions. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 208–214. JMLR Workshop and Conference Proceedings, 2011.

[9] Nima Hamidi and Mohsen Bayati. On frequentist regret of linear thompson sampling. *arXiv preprint arXiv:2006.06790*, 2020.

[10] David Janz, Alexander E Litvak, and Csaba Szepesvári. Ensemble sampling for linear bandits: small ensembles suffice. *arXiv preprint arXiv:2311.08376*, 2023.

[11] Branislav Kveton, Csaba Szepesvari, Mohammad Ghavamzadeh, and Craig Boutilier. Perturbed-history exploration in stochastic multi-armed bandits. In *International Joint Conference on Artificial Intelligence*, 2019. URL `https://api.semanticscholar.org/CorpusID:67856126`.

[12] Branislav Kveton, Csaba Szepesvári, Mohammad Ghavamzadeh, and Craig Boutilier. Perturbed-history exploration in stochastic linear bandits. In *Uncertainty in Artificial Intelligence*, pages 530–540. PMLR, 2020.

[13] Branislav Kveton, Manzil Zaheer, Csaba Szepesvari, Lihong Li, Mohammad Ghavamzadeh, and Craig Boutilier. Randomized exploration in generalized linear bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 2066–2076. PMLR, 2020.

[14] Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

[15] Beatrice Laurent and Pascal Massart. Adaptive estimation of a quadratic functional by model selection. *Annals of statistics*, pages 1302–1338, 2000.

[16] Xiuyuan Lu and Benjamin Van Roy. Ensemble sampling. *Advances in Neural Information Processing Systems*, 30, 2017.

 https://doi.org/10.52202/079017-2947

[17] Xiuyuan Lu, Zheng Wen, and Branislav Kveton. Efficient online recommendation via low-rank ensemble sampling. In *Proceedings of the 12th ACM Conference on Recommender Systems*, pages 460–464, 2018.

[18] Ian Osband, Charles Blundell, Alexander Pritzel, and Benjamin Van Roy. Deep exploration via bootstrapped dqn. *Advances in Neural Information Processing Systems*, 29, 2016.

[19] Ian Osband, John Aslanides, and Albin Cassirer. Randomized prior functions for deep reinforcement learning. *Advances in Neural Information Processing Systems*, 31, 2018.

[20] Ian Osband, Benjamin Van Roy, Daniel J Russo, Zheng Wen, et al. Deep exploration via randomized value functions. *Journal of Machine Learning Research*, 20(124):1–62, 2019.

[21] Chao Qin, Zheng Wen, Xiuyuan Lu, and Benjamin Van Roy. An analysis of ensemble sampling. *Advances in Neural Information Processing Systems*, 35:21602–21614, 2022.

[22] Daniel Russo and Benjamin Van Roy. Learning to optimize via posterior sampling. *Mathematics of Operations Research*, 39(4):1221–1243, 2014.

[23] Daniel J Russo, Benjamin Van Roy, Abbas Kazerouni, Ian Osband, Zheng Wen, et al. A tutorial on thompson sampling. *Foundations and Trends® in Machine Learning*, 11(1):1–96, 2018.

[24] William R Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294, 1933.

[25] Runzhe Wan, Haoyu Wei, Branislav Kveton, and Rui Song. Multiplier bootstrap-based exploration. In *International Conference on Machine Learning*, pages 35444–35490. PMLR, 2023.

[26] Jie Zhou, Botao Hao, Zheng Wen, Jingfei Zhang, and Will Wei Sun. Stochastic low-rank tensor bandits for multi-dimensional online decision making. *Journal of the American Statistical Association*, pages 1–14, 2024.

[27] Zheqing Zhu and Benjamin Van Roy. Deep exploration for recommendation systems. In *Proceedings of the 17th ACM Conference on Recommender Systems*, pages 963–970, 2023.

Appendix

A Notations

We summarizes the notations in this paper in Table 2 and Table 3

Table 2: Notations specific to this paper

Linear Bandit	
$\mathcal{X}$	Set of arms
θ^*	True parameter vector
x^*	Optimal arm
X_t	Chosen arm at time t
Y_t	Observed reward at time t
η_t	Zero-mean noise at time t
σ	SubGaussian variance proxy of η_t
d	Dimension of arms and true parameter vector
K	Number of arms (if $\lvert\mathcal{X}\rvert < \infty$)
T	Time horizon
$R(T)$	Cumulative regret
Algorithm	
λ	Regularization parameter
V_t	$\lambda I_d + \sum_{i=1}^{t} X_i X_i^\top$
θ_t	Perturbed estimator
W^j / W_t	Initial perturbation vector
Z_t^j / $Z_{t,i}$	Reward perturbation
$\mathcal{P}_I$	Distribution for initial perturbation W
$\mathcal{P}_R$	Distribution for reward perturbation Z_t
$\mathcal{J}_t$	Sampling policy at time t
Analysis	
δ	Probability of failure
$\mathcal{F}_t^X$	$\sigma(\{X_i\}_{i=1}^t)$
$\mathcal{F}_t^\eta$	$\sigma(\{\eta_i\}_{i=1}^t)$
$\mathcal{F}_t^A$	σ-algebra generated by algorithm's internal randomness until time t
$\mathcal{F}_t$	$\sigma(\mathcal{F}_t^A \cup \mathcal{F}_t^X \cup \mathcal{F}_t^\eta)$, σ-algebra generated by the first t-iterations of interaction
$\hat{\theta}_t$	Ridge regression estimator by t samples
$\widetilde{\theta}_t$	Perturbation in the estimator, $\theta_{t+1} - \hat{\theta}_t$
β_t	Confidence radius of $\hat{\theta}_t$
γ	Confidence radius of θ_t
$\widetilde{\gamma}$	Confidence radius of $\widetilde{\theta}_t$
p	Probability of optimistic arm selection
p_N	$\mathbb{P}(Z \geq 1) \approx 0.15$ where $Z \sim \mathcal{N}(0,1)$
$\mathcal{E}$	High probability event
$\mathcal{E}_{1,t}$	Concentration event
$\mathcal{E}_{2,t}$	Optimism event
$\hat{\mathcal{E}}_t$	Concentration event of $\hat{\theta}_t$
$\widetilde{\mathcal{E}}_{1,t}$	Concentration event of $\widetilde{\theta}_{t-1}$
$\widetilde{\mathcal{E}}_{2,t}$	Anti-concentration event of $\widetilde{\theta}_{t-1}$
$\mathbf{Z}_t$	Perturbation vector at time t
Φ_t	$\begin{pmatrix}\sqrt{\lambda} I_d & X_1 & \cdots & X_t\end{pmatrix} \in \mathbb{R}^{d+t}$
U_t	$x^{*\top} V_t^{-1} \Phi_t$

https://doi.org/10.52202/079017-2947

Table 3: Generic notations

Sets and functions	
$\mathbb{N}$	Set of natural numbers, starting from 1
$[M]$	Set of natural numbers up to M, i.e., $\{1, 2, \ldots, M\}$
$\mathbb{1}$	Indicator function
Vector and matrices	
$\|\cdot\|_2$	ℓ_2 norm of a vector
$(\cdot)_i$	i-th element of a vector
$\mathbf{0}_d$	Zero vector in $\mathbb{R}^d$
I_d	Identity matrix in $\mathbb{R}^{d\times d}$
Probability	
$(\Omega, \mathcal{F}, \mathbb{P})$	Probability space
$\mathbb{E}$	Expectation
$\mathcal{N}(\mu, \Sigma)$	Gaussian distribution with mean μ and covariance Σ
$\wedge$	Logical conjunction ("and")
$\vee$	Logical disjunction ("or")

B Proof of Theorem 2

Lemma 5 (Lemmas 10 and 11 in Abbasi-Yadkori et al. [1])**.** *Let $\lambda \geq 1$, $\{X_t\}_{t=1}^T$ be any sequence of d-dimensional vectors such that $\|X_t\|_2 \leq 1$ for all $t \in [T]$, and $V_t = \lambda I + \sum_{i=1}^t X_i X_i^\top$. Then, $\sum_{t=1}^T \|X_t\|^2_{V_{t-1}^{-1}} \leq 2d \log\left(1 + \frac{T}{d\lambda}\right)$.*

As alluded in the discussion of Theorem 2, we utilize Markov's inequality and the random variable of interest is $X_t^\top \theta_t \mathbb{1}\left\{\mathcal{E}_{1,t} \cap \mathcal{E}_{2,t}\right\}$. However, this random variable may not be non-negative, precluding the use of Markov's inequality. The following lemma shows that adding an appropriate term guarantees its non-negativity.

Lemma 6. *Assume the conditions of Theorem 2. Let $J(\theta) = \sup_{x\in\mathcal{X}} x^\top \theta$, where $\theta \in \mathbb{R}^d$. Let $\Theta_t = \{\theta \in \mathbb{R}^d \mid \|\theta - \theta^*\|_{V_{t-1}} \leq \gamma\}$. Define $\theta_t^- = \operatorname{argmin}_{\theta\in\Theta_t} J(\theta)$ and $X_t^- = \operatorname{argmax}_{x\in\mathcal{X}} x^\top \theta_t^-$. For any $\theta \in \mathbb{R}^d$ and an event $\mathcal{E}'$, we introduce the following notation:*

$$g_t\left(\theta, \mathcal{E}'\right) = \left(J(\theta) - J(\theta_t^-)\right) \mathbb{1}\{\mathcal{E}'\}.$$

Then, $g_t(\theta^*, \mathcal{E}') \geq 0$ holds for any event $\mathcal{E}' \in \mathcal{F}$, and $g_t(\theta_t, \mathcal{E}'') \geq 0$ holds almost surely for any event such that $\mathcal{E}'' \subset \mathcal{E}_{1,t}$.

Proof. We first prove $g_t(\theta^*, \mathcal{E}') \geq 0$. Since $\theta^* \in \Theta_t$,

$$J(\theta^*) \geq \inf_{\theta\in\Theta_t} J(\theta) = J(\theta_t^-)$$

always holds. Therefore, for any event $\mathcal{E}'$, $g_t(\theta^*, \mathcal{E}') \geq 0$ holds.
We now suppose $\mathcal{E}'' \subset \mathcal{E}_{1,t}$ and prove $g_t(\theta_t, \mathcal{E}'') \geq 0$. We consider two cases where $\mathcal{E}''$ does and does not hold. Under $\mathcal{E}''$, since $\mathcal{E}'' \subset \mathcal{E}_{1,t}$ and by concentration condition (2), $\theta_t \in \Theta_t$ holds almost surely. Then, $J(\theta_t) \geq \inf_{\theta\in\Theta_t} J(\theta) = J(\theta_t^-)$ holds. Under $\mathcal{E}''^{\mathsf{C}}$, $g_t(\theta_t, \mathcal{E}'') = 0 \geq 0$ trivially holds. Therefore, $g_t(\theta_t, \mathcal{E}'') \geq 0$ holds almost surely. □

Proof of Theorem 2. We show that with probability at least $1 - \delta$, it holds that

$$R(T)\mathbb{1}\left\{\mathcal{E}\right\} \leq \gamma\left(1 + \frac{2}{p}\right)\sqrt{2dT\log\left(1 + \frac{T}{d\lambda}\right)} + \frac{\gamma}{p}\sqrt{\frac{2T}{\lambda}\log\frac{1}{\delta}}.$$

We first bound the instantaneous regret under the event $\mathcal{E}$ at time t.

$$\begin{aligned}\left(x^{*\top}\theta^* - X_t^\top\theta^*\right)\mathbb{1}\left\{\mathcal{E}\right\} &= \left(x^{*\top}\theta^* - X_t^\top\theta_t + X_t^\top\theta_t - X_t^\top\theta^*\right)\mathbb{1}\left\{\mathcal{E}\right\} \\ &= \underbrace{\left(x^{*\top}\theta^* - X_t^\top\theta_t\right)\mathbb{1}\left\{\mathcal{E}\right\}}_{I_1} + \underbrace{\left(X_t^\top\theta_t - X_t^\top\theta^*\right)\mathbb{1}\left\{\mathcal{E}\right\}}_{I_2}\end{aligned}$$

I_2 is directly bounded under $\mathcal{E}$.

$$\begin{aligned} I_2 &= X_t^\top (\theta_t - \theta^*) \mathbb{1}\{\mathcal{E}\} \\ &\leq \|X_t\|_{V_{t-1}^{-1}} \|\theta_t - \theta^*\|_{V_{t-1}} \mathbb{1}\{\mathcal{E}\} \\ &\leq \gamma \|X_t\|_{V_{t-1}^{-1}} , \end{aligned}$$

where the first inequality is due to the Cauchy-Schwarz inequality and the second inequality comes from condition (2).
Now, we bound I_1. Let X_t^-, θ_t^-, and g_t be defined as in Lemma 6. By Lemma 6, $g_t(\theta_t, \mathcal{E}) \geq 0$ holds almost surely. Then,

$$\begin{aligned} I_1 &= x^{*\top}\theta^* \mathbb{1}\{\mathcal{E}\} - X_t^\top \theta_t \mathbb{1}\{\mathcal{E}\} \\ &= x^{*\top}\theta^* \mathbb{1}\{\mathcal{E}\} - X_t^{-\top}\theta_t^- \mathbb{1}\{\mathcal{E}\} + X_t^{-\top}\theta_t^- \mathbb{1}\{\mathcal{E}\} - X_t^\top \theta_t \mathbb{1}\{\mathcal{E}\} \\ &= g_t(\theta^*, \mathcal{E}) - g_t(\theta_t, \mathcal{E}) \\ &\leq g_t(\theta^*, \mathcal{E}) \\ &\leq g_t(\theta^*, \mathcal{E}_{2,t}) , \end{aligned}$$

where the last inequality holds since $\mathcal{E} \subset \mathcal{E}_{2,t}$. Again by Lemma 6, $g_t(\theta^*, \mathcal{E}_{2,t})$ and $g_t(\theta_t, \mathcal{E}_{1,t} \cap \mathcal{E}_{2,t})$ are non-negative almost surely. Note that by the first part of condition (3), $\mathcal{E}_{2,t}$ is $\mathcal{F}_{t-1}$-measurable and hence $g_t(\theta^*, \mathcal{E}_{2,t}) = (x^{*\top}\theta^* - X_t^{-\top}\theta_t^-)\mathbb{1}\{\mathcal{E}_{2,t}\}$ is also $\mathcal{F}_{t-1}$-measurable. Applying Markov's inequality conditioned on $\mathcal{F}_{t-1}$, we obtain that

$$g_t(\theta^*, \mathcal{E}_{2,t})\mathbb{P}\Big(g_t(\theta^*, \mathcal{E}_{2,t}) \leq g_t(\theta_t, \mathcal{E}_{1,t} \cap \mathcal{E}_{2,t}) \mid \mathcal{F}_{t-1}\Big) \leq \mathbb{E}\Big[g_t(\theta_t, \mathcal{E}_{1,t} \cap \mathcal{E}_{2,t}) \mid \mathcal{F}_{t-1}\Big] . \quad (11)$$

We lower-bound the probability on the left hand side utilizing condition (3). Suppose the event of interest in condition (3), namely $((x^{*\top}\theta^* \leq X_t^\top \theta_t) \wedge \mathcal{E}_{1,t}) \vee \mathcal{E}_{2,t}^\mathsf{C}$, holds. Under the event, either $\mathcal{E}_{2,t}^\mathsf{C}$ or $(x^{*\top}\theta^* \leq X_t^\top \theta_t) \wedge \mathcal{E}_{1,t} \wedge \mathcal{E}_{2,t}$ holds. Under $\mathcal{E}_{2,t}^\mathsf{C}$, $g_t(\theta^*, \mathcal{E}_{2,t}) \leq g_t(\theta_t, \mathcal{E}_{1,t} \cap \mathcal{E}_{2,t})$ becomes $0 \leq 0$, which trivially holds. Otherwise, we have $(x^{*\top}\theta^* \leq X_t^\top \theta_t) \wedge \mathcal{E}_{1,t} \wedge \mathcal{E}_{2,t}$. Under this event, we have that

$$\begin{aligned} x^{*\top}\theta^* \leq X_t^\top \theta_t &\Leftrightarrow x^{*\top}\theta^* - X_t^{-\top}\theta_t^- \leq X_t^\top \theta_t - X_t^{-\top}\theta_t^- \\ &\Leftrightarrow \left(x^{*\top}\theta^* - X_t^{-\top}\theta_t^-\right) \mathbb{1}\{\mathcal{E}_{2,t}\} \leq \left(X_t^\top \theta_t - X_t^{-\top}\theta_t^-\right) \mathbb{1}\{\mathcal{E}_{1,t} \cap \mathcal{E}_{2,t}\} \\ &\Leftrightarrow g_t(\theta^*, \mathcal{E}_{2,t}) \leq g_t(\theta_t, \mathcal{E}_{1,t} \cap \mathcal{E}_{2,t}) . \end{aligned}$$

Therefore, we have shown that

$$\left(\left(x^{*\top}\theta^* \leq X_t^\top \theta_t\right) \wedge \mathcal{E}_{1,t}\right) \vee \mathcal{E}_{2,t}^\mathsf{C} \Rightarrow g_t(\theta^*, \mathcal{E}_{2,t}) \leq g_t(\theta_t, \mathcal{E}_{1,t} \cap \mathcal{E}_{2,t}) ,$$

which implies

$$\begin{aligned} \mathbb{P}\left(g_t(\theta^*, \mathcal{E}_{2,t}) \leq g_t(\theta_t, \mathcal{E}_{1,t} \cap \mathcal{E}_{2,t}) \mid \mathcal{F}_{t-1}\right) &\geq \mathbb{P}\left(\left(x^{*\top}\theta^* \leq X_t^\top \theta_t \wedge \mathcal{E}_{1,t}\right) \vee \mathcal{E}_{2,t}^\mathsf{C} \mid \mathcal{F}_{t-1}\right) \\ &\geq p \end{aligned}$$

by condition (3). Therefore, we obtain that $g_t(\theta^*, \mathcal{E}_{2,t}) \leq \frac{1}{p}\mathbb{E}[g_t(\theta_t, \mathcal{E}_{1,t} \cap \mathcal{E}_{2,t}) \mid \mathcal{F}_{t-1}]$ from inequality (11). Lastly, we bound $g_t(\theta_t, \mathcal{E}_{1,t} \cap \mathcal{E}_{2,t})$.

$$\begin{aligned} g_t(\theta_t, \mathcal{E}_{1,t} \cap \mathcal{E}_{2,t}) &= \left(X_t^\top \theta_t - X_t^{-\top}\theta_t^-\right) \mathbb{1}\{\mathcal{E}_{1,t} \cap \mathcal{E}_{2,t}\} \\ &\leq \left(X_t^\top \theta_t - X_t^\top \theta_t^-\right) \mathbb{1}\{\mathcal{E}_{1,t} \cap \mathcal{E}_{2,t}\} \\ &\leq \|X_t\|_{V_{t-1}^{-1}} \left\|\theta_t - \theta_t^-\right\|_{V_{t-1}} \mathbb{1}\{\mathcal{E}_{1,t}\} \\ &\leq \|X_t\|_{V_{t-1}^{-1}} \left(\|\theta_t - \theta^*\|_{V_{t-1}} + \left\|\theta_t^- - \theta^*\right\|_{V_{t-1}}\right) \mathbb{1}\{\mathcal{E}_{1,t}\} \\ &\leq 2\gamma \|X_t\|_{V_{t-1}^{-1}} , \end{aligned}$$

where the first inequality uses that $X_t^- = \sup_{x \in \mathcal{X}} x^\top \theta_t^-$, the second inequality is due to the Cauchy-Schwarz inequality, the third inequality holds by the triangle inequality, and the last inequality comes

https://doi.org/10.52202/079017-2947

from condition (2) and that $\theta_t^- \in \Theta_t$ as defined in Lemma 6. Combining all, the instantaneous regret at time t under the event $\mathcal{E}$ is bounded as follows:

$$\begin{aligned}\left(x^{*\top}\theta^* - X_t^\top\theta^*\right)\mathbb{1}\left\{\mathcal{E}\right\} &\le \gamma\left\|X_t\right\|_{V_{t-1}^{-1}} + \frac{2\gamma}{p}\mathbb{E}\left[\left\|X_t\right\|_{V_{t-1}^{-1}} \mid \mathcal{F}_{t-1}\right] \\ &= \left(\gamma + \frac{2\gamma}{p}\right)\left\|X_t\right\|_{V_{t-1}^{-1}} + \frac{2\gamma}{p}\left(\mathbb{E}\left[\left\|X_t\right\|_{V_{t-1}^{-1}} \mid \mathcal{F}_{t-1}\right] - \left\|X_t\right\|_{V_{t-1}^{-1}}\right).\end{aligned}$$

Then, the cumulative regret under $\mathcal{E}$ is bounded as

$$R(T)\mathbb{1}\left\{\mathcal{E}\right\} \le \gamma\left(1+\frac{2}{p}\right)\sum_{t=1}^{T}\left\|X_t\right\|_{V_{t-1}^{-1}} + \frac{2\gamma}{p}\sum_{t=1}^{T}\left(\mathbb{E}\left[\left\|X_t\right\|_{V_{t-1}^{-1}} \mid \mathcal{F}_{t-1}\right] - \left\|X_t\right\|_{V_{t-1}^{-1}}\right).$$

To bound the first sum, we apply the Cauchy-Schwarz inequality and then Lemma 5.

$$\begin{aligned}\sum_{t=1}^{T}\left\|X_t\right\|_{V_{t-1}^{-1}} &\le \sqrt{T\sum_{t=1}^{T}\left\|X_t\right\|_{V_{t-1}^{-1}}^2} \\ &\le \sqrt{2dT\log\left(1+\frac{T}{d\lambda}\right)}.\end{aligned}$$

The second sum is bounded by Azuma-Hoeffding inequality. Note that $0 \le \left\|X_t\right\|_{V_{t-1}^{-1}} \le \sqrt{\lambda_{\max}(V_{t-1}^{-1})}\left\|X_t\right\|_2 \le \frac{1}{\sqrt{\lambda}}$, where $\lambda_{\max}(\cdot)$ denotes the maximum eigenvalue. By Lemma 12, with probability at least $1-\delta$, it holds that

$$\sum_{t=1}^{T}\left(\mathbb{E}\left[\left\|X_t\right\|_{V_{t-1}^{-1}} \mid \mathcal{F}_{t-1}\right] - \left\|X_t\right\|_{V_{t-1}^{-1}}\right) \le \sqrt{\frac{T}{2\lambda}\log\frac{1}{\delta}}.$$

Therefore, with probability at least $1-\delta$, the cumulative regret under $\mathcal{E}$ is bounded as follows:

$$R(T)\mathbb{1}\left\{\mathcal{E}\right\} \le \gamma\left(1+\frac{2}{p}\right)\sqrt{2dT\log\left(1+\frac{T}{d\lambda}\right)} + \frac{\gamma}{p}\sqrt{\frac{2T}{\lambda}\log\frac{1}{\delta}}.$$

□

C Proof of Lemma 2

Proof of Lemma 2. The proof is simple algebra utilizing a useful matrix Φ_t. Define Φ_t to be the matrix that stacks X_i in addition to an identity matrix as follows:

$$\Phi_t := \begin{pmatrix}\sqrt{\lambda}I_d & X_1 & \cdots & X_t\end{pmatrix} \in \mathbb{R}^{d\times(d+t)}.$$

We can express a relevant matrix and vectors using Φ_t, which are $V_t = \Phi_t\Phi_t^\top$, $\widetilde{\theta}_t^j = V_t^{-1}\Phi_t\mathbf{Z}_{t+1}^j$, and $U_t^\top = x^{*\top}V_t^{-1}\Phi_t$. Defining $\widetilde{\theta}_{t-1} = \widetilde{\theta}_{t-1}^{j_t}$ to be the perturbation in the selected estimator at time t, we obtain that

$$\begin{aligned}x^{*\top}\widetilde{\theta}_{t-1} &= x^{*\top}V_{t-1}^{-1}\Phi_{t-1}\mathbf{Z}_t \\ &= U_{t-1}^\top\mathbf{Z}_t.\end{aligned}$$

In addition, it holds that

$$\begin{aligned}\left\|U_{t-1}\right\|_2 &= \sqrt{x^{*\top}V_{t-1}^{-1}\Phi_{t-1}\Phi_{t-1}^\top V_{t-1}^{-1}x^*} \\ &= \sqrt{x^{*\top}V_{t-1}^{-1}x^*} \\ &= \left\|x^*\right\|_{V_{t-1}^{-1}}.\end{aligned}$$

By $X_t^\top \theta_t = \sup_{x \in \mathcal{X}} x^\top \theta_t$, it holds that $x^{*\top} \theta_t \leq X_t^\top \theta_t$. Then, we have that

$$\begin{aligned} X_t^\top \theta_t - x^{*\top} \theta^* &\geq x^{*\top} \theta_t - x^{*\top} \theta^* \\ &= x^{*\top} \left(\theta_t - \theta^* \right) \\ &= x^{*\top} \left(\widetilde{\theta}_{t-1} + \hat{\theta}_{t-1} - \theta^* \right) \\ &= U_{t-1} \mathbf{Z}_t + x^{*\top} \left(\hat{\theta}_{t-1} - \theta^* \right) . \end{aligned}$$

By the condition of the lemma, there exists a positive constant c such that $\|\hat{\theta}_{t-1} - \theta^*\|_{V_{t-1}} \leq c$ and $U_{t-1}^\top \mathbf{Z}_t - c \|U_{t-1}\|_2 \geq 0$. By the Cauchy-Schwarz inequality, it holds that

$$\begin{aligned} x^{*\top} \left(\hat{\theta}_{t-1} - \theta^* \right) &\geq - \|x^*\|_{V_{t-1}^{-1}} \left\| \hat{\theta}_{t-1} - \theta^* \right\|_{V_{t-1}} \\ &\geq -c \left\| U_{t-1} \right\|_2 . \end{aligned}$$

Therefore,

$$\begin{aligned} X_t^\top \theta_t - x^{*\top} \theta^* &\geq U_{t-1}^\top \mathbf{Z}_t - c \left\| U_{t-1} \right\|_2 \\ &\geq 0 , \end{aligned}$$

where we proved that $X_t^\top \theta_t \geq x^{*\top} \theta^*$. □

D Proof of Lemma 3

Proof of Lemma 3. Under $\mathcal{E}_{1,t} = \widetilde{\mathcal{E}}_{1,t} \cap \hat{\mathcal{E}}_{t-1}$, the concentration condition, specifically condition (2) in Theorem 2, holds by the triangle inequality.

$$\|\theta_t - \theta^*\|_{V_{t-1}} = \left\| \widetilde{\theta}_{t-1} + \hat{\theta}_{t-1} - \theta^* \right\|_{V_{t-1}} \leq \left\| \widetilde{\theta}_{t-1} \right\|_{V_{t-1}} + \left\| \hat{\theta}_{t-1} - \theta^* \right\|_{V_{t-1}} \leq \widetilde{\gamma} + \beta_{t-1} \leq \gamma .$$

To show the optimism condition, condition (3) in Theorem 2, we first show that $\mathcal{E}_{2,t} \in \mathcal{F}_{t-1}$. $\hat{\mathcal{E}}_{t-1} \in \mathcal{F}_{t-1}$ holds since it regards $\{X_i, \eta_i\}_{i=1}^{t-1}$ only. Since $\mathbb{P}((U_{t-1}^\top \mathbf{Z}_t \geq \beta_{t-1} \|U_{t-1}\|_2) \wedge \widetilde{\mathcal{E}}_{1,t} \mid \mathcal{F}_{t-1})$ is a $\mathcal{F}_{t-1}$-measurable random variable and $\widetilde{\mathcal{E}}_{2,t}$ is an event that the specified random variable is greater than or equal to p, $\widetilde{\mathcal{E}}_{2,t}$ is in $\mathcal{F}_{t-1}$. Therefore, we obtain that $\mathcal{E}_{2,t} = \hat{\mathcal{E}}_{t-1} \cap \widetilde{\mathcal{E}}_{2,t} \in \mathcal{F}_{t-1}$. To prove the remaining part of condition (3), we demonstrate the following logical implication relationships:

$$\begin{aligned} \left(\left(U_{t-1}^\top \mathbf{Z}_t \geq \beta_{t-1} \|U_{t-1}\|_2 \right) \wedge \widetilde{\mathcal{E}}_{1,t} \right) \vee \mathcal{E}_{2,t}^{\mathsf{C}} &\Rightarrow \left(\left(U_{t-1}^\top \mathbf{Z}_t \geq \beta_{t-1} \|U_{t-1}\|_2 \right) \wedge \widetilde{\mathcal{E}}_{1,t} \wedge \mathcal{E}_{2,t} \right) \vee \mathcal{E}_{2,t}^{\mathsf{C}} \\ &\Rightarrow \left(\left(U_{t-1}^\top \mathbf{Z}_t \geq \beta_{t-1} \|U_{t-1}\|_2 \right) \wedge \widetilde{\mathcal{E}}_{1,t} \wedge \hat{\mathcal{E}}_{t-1} \right) \vee \mathcal{E}_{2,t}^{\mathsf{C}} \\ &\Rightarrow \left(\left(x^{*\top} \theta^* \leq X_t^\top \theta_t \right) \wedge \widetilde{\mathcal{E}}_{1,t} \wedge \hat{\mathcal{E}}_{t-1} \right) \vee \mathcal{E}_{2,t}^{\mathsf{C}} \\ &\Rightarrow \left(\left(x^{*\top} \theta^* \leq X_t^\top \theta_t \right) \wedge \mathcal{E}_{1,t} \right) \vee \mathcal{E}_{2,t}^{\mathsf{C}} , \end{aligned}$$

where the first implication follows from $A \vee B^{\mathsf{C}} \Leftrightarrow (A \wedge B) \vee B^{\mathsf{C}}$, the second implication holds since $\mathcal{E}_{2,t} \subset \hat{\mathcal{E}}_{t-1}$, the third implication holds by Lemma 2, which states that $(U_{t-1}^\top \mathbf{Z}_t \geq \beta_{t-1} \|U_{t-1}\|_2 \wedge \hat{\mathcal{E}}_{t-1}) \Rightarrow (x^{*\top} \theta^* \leq X_t^\top \theta_t)$, and the last by $\mathcal{E}_{1,t} = \widetilde{\mathcal{E}}_{1,t} \cap \hat{\mathcal{E}}_{t-1}$. This implication relationship shows that

$$\mathbb{P}\left(\left(\left(x^{*\top} \theta^* \leq X_t^\top \theta_t \right) \wedge \mathcal{E}_{1,t} \right) \vee \mathcal{E}_{2,t}^{\mathsf{C}} \mid \mathcal{F}_{t-1} \right) \geq \mathbb{P}\left(\left(\left(U_{t-1}^\top \mathbf{Z}_t \geq \beta_{t-1} \|U_{t-1}\|_2 \right) \wedge \widetilde{\mathcal{E}}_{1,t} \right) \vee \mathcal{E}_{2,t}^{\mathsf{C}} \mid \mathcal{F}_{t-1} \right) .$$

https://doi.org/10.52202/079017-2947

We bound the right hand side using the definition of $\mathcal{E}_{2,t}$.

$$\begin{aligned}
&\mathbb{P}\left(\left(\left(U_{t-1}^{\top}\mathbf{Z}_t \geq \beta_{t-1}\left\|U_{t-1}\right\|_2\right) \wedge \widetilde{\mathcal{E}}_{1,t}\right) \vee \mathcal{E}_{2,t}^{\mathsf{C}} \mid \mathcal{F}_{t-1}\right) \\
&= \mathbb{E}\left[\mathbb{1}\left\{\left(\left(U_{t-1}^{\top}\mathbf{Z}_t \geq \beta_{t-1}\left\|U_{t-1}\right\|_2\right) \wedge \widetilde{\mathcal{E}}_{1,t}\right) \vee \mathcal{E}_{2,t}^{\mathsf{C}}\right\} \mid \mathcal{F}_{t-1}\right] \\
&= \mathbb{E}\left[\mathbb{1}\left\{\left(U_{t-1}^{\top}\mathbf{Z}_t \geq \beta_{t-1}\left\|U_{t-1}\right\|_2\right) \wedge \widetilde{\mathcal{E}}_{1,t}\right\}\mathbb{1}\left\{\mathcal{E}_{2,t}\right\} + \mathbb{1}\left\{\mathcal{E}_{2,t}^{\mathsf{C}}\right\} \mid \mathcal{F}_{t-1}\right] \\
&= \mathbb{E}\left[\mathbb{1}\left\{\left(U_{t-1}^{\top}\mathbf{Z}_t \geq \beta_{t-1}\left\|U_{t-1}\right\|_2\right) \wedge \widetilde{\mathcal{E}}_{1,t}\right\} \mid \mathcal{F}_{t-1}\right]\mathbb{1}\left\{\mathcal{E}_{2,t}\right\} + \mathbb{1}\left\{\mathcal{E}_{2,t}^{\mathsf{C}}\right\} \\
&\geq p\mathbb{1}\left\{\mathcal{E}_{2,t}\right\} + \mathbb{1}\left\{\mathcal{E}_{2,t}^{\mathsf{C}}\right\} \\
&\geq p\,,
\end{aligned}$$

where the third equality uses that $\mathcal{E}_{2,t} \in \mathcal{F}_{t-1}$ and the first inequality holds since under $\mathcal{E}_{2,t} \subset \widetilde{\mathcal{E}}_{2,t}$, $\mathbb{E}\left[\mathbb{1}\left\{\left(U_{t-1}^{\top}\mathbf{Z}_t \geq \beta_{t-1}\left\|U_{t-1}\right\|_2\right) \wedge \widetilde{\mathcal{E}}_{1,t}\right\} \mid \mathcal{F}_{t-1}\right] \geq p$ holds by the definition of $\widetilde{\mathcal{E}}_{2,t}$.
Therefore, $\mathcal{E}_{1,t}$ and $\mathcal{E}_{2,t}$ satisfy conditions (2) and (3). By Theorem 2, the regret bound stated in inequality (4) holds with probability at least $1-\delta-\mathbb{P}(\mathcal{E}^{\mathsf{C}})$, where the union bound is taken. Note that $\mathcal{E} = \cap_{t=1}^{T}\left(\mathcal{E}_{1,t} \cap \mathcal{E}_{2,t}\right) = \cap_{t=1}^{T}\left(\widetilde{\mathcal{E}}_{1,t} \cap \widetilde{\mathcal{E}}_{2,t} \cap \hat{\mathcal{E}}_{t-1}\right) \supset \hat{\mathcal{E}} \cap \widetilde{\mathcal{E}}$. The failure probability is bounded as

$$\begin{aligned}
\delta + \mathbb{P}\left(\mathcal{E}^{\mathsf{C}}\right) &\leq \delta + \mathbb{P}\left(\hat{\mathcal{E}}^{\mathsf{C}}\right) + \mathbb{P}\left(\widetilde{\mathcal{E}}^{\mathsf{C}}\right) \\
&\leq 2\delta + \mathbb{P}\left(\widetilde{\mathcal{E}}^{\mathsf{C}}\right)\,,
\end{aligned}$$

where the first inequality takes the union bound over $\mathcal{E}^{\mathsf{C}} \subset \hat{\mathcal{E}}^{\mathsf{C}} \cup \widetilde{\mathcal{E}}^{\mathsf{C}}$ and the second inequality is due to Lemma 1. □

E Proof of Lemma 4

Lemma 4 is a special case of Lemma 9, which generalizes Gaussian distribution to any subGaussian distribution. We first provide a general chi-squared concentration result, which is required to bound the perturbation induced by W. A generalized version of Lemma 7 is presented in Lemma 10.

Lemma 7. *If $Z \sim \mathcal{N}(0, I_d)$ is a d-dimensional multivariate Gaussian vector, then for any $\delta \in (0,1]$,*

$$\mathbb{P}\left(\left\|Z\right\|_2 \geq \sqrt{d} + \sqrt{2\log\frac{1}{\delta}}\right) \leq \delta\,.$$

Proof. By Lemma 13 with $x = \log\frac{1}{\delta}$, it holds that

$$\mathbb{P}\left(\left\|Z\right\|_2^2 - d \geq 2\sqrt{d\log\frac{1}{\delta}} + 2\log\frac{1}{\delta}\right) \leq \delta\,.$$

Since $d + 2\sqrt{d\log\frac{1}{\delta}} + 2\log\frac{1}{\delta} \leq \left(\sqrt{d} + \sqrt{2\log\frac{1}{\delta}}\right)^2$, it holds that

$$\begin{aligned}
\mathbb{P}\left(\left\|Z\right\|_2 \geq \sqrt{d} + \sqrt{2\log\frac{1}{\delta}}\right) &\leq \mathbb{P}\left(\left\|Z\right\|_2^2 \geq d + 2\sqrt{d\log\frac{1}{\delta}} + 2\log\frac{1}{\delta}\right) \\
&\leq \delta\,.
\end{aligned}$$

□

Proof of Lemma 4. By Lemma 7, with $\delta/2T$ instead of δ, yields

$$\mathbb{P}\left(\left\|W^j\right\|_2 \geq \sqrt{\lambda}\beta_T\left(\sqrt{d} + \sqrt{2\log\frac{2T}{\delta}}\right)\right) \leq \frac{\delta}{2T}\,,$$

since $W^j \sim \mathcal{N}(\mathbf{0}_d, \lambda\beta_T^2 I_d)$. Applying Lemma 9 with $\delta/2T$ instead of δ yields that

$$\left\|\widetilde{\theta}_{t-1}^j\right\|_{V_{t-1}} \leq \beta_T\sqrt{d\log\left(1+\frac{T}{d\lambda}\right)+2\log\frac{2T}{\delta}}+\beta_T\left(\sqrt{d}+\sqrt{2\log\frac{2T}{\delta}}\right)$$

holds with probability at least $1-\delta/T$. Note that the right hand side is equal to the definition of $\widetilde{\gamma}_T$, defined in Eq. (7). □

F Rigorous Justification of Proposition 1

In this section, we rigorously justify Proposition 1 that is stated in the proof of Theorem 1. To do so, we present a different viewpoint on the perturbation sequences.

Denote the arms as $\mathcal{X} = \{x^1, x^2, \ldots, x^K\}$. For the sake of the analysis, assume that $\delta/T \leq p_N/2 \approx 0.08$, which holds whenever $T \geq 14$ or $\delta < 0.07$.

We reconstruct the perturbation sampled by Algorithm 1. Assume that in addition to $\{W^j\}_{j=1}^m \overset{i.i.d.}{\sim} \mathcal{P}_I$, Algorithm 1 samples mKT samples of $\{\{Z_{k,t}^j\}_{(k,t)}\}_{j=1}^m \overset{i.i.d.}{\sim} \mathcal{P}_R$ at the beginning, where the subscript (k,t) enumerates from $(1,1)$ to (K,T). Define $N_{k,t} = \sum_{i=1}^t \mathbb{1}\left\{X_i = x^k\right\}$ to be the number of times arm k has been chosen up to time t. If the a_t-th arm, x^{a_t}, is selected at time t, then we assign $Z_t^j = Z_{a_t,N_{a_t,t}}^j$. Since $Z_{a_t,N_{a_t,t}}^j$ is still an i.i.d. sample of $\mathcal{P}_R$ conditioned on history, specifically on $\sigma(\mathcal{F}_t^X \cup \mathcal{F}_t^\eta \cup \sigma(\{\{Z_i^j\}_{i=1}^{t-1}\}_{j=1}^m))$, we attain an equivalent algorithm with Algorithm 1. We note that these modifications need not be taken in the execution of the algorithm, and their purpose is purely for the analysis. Define $\mathcal{F}_0^A = \sigma(\{W^j, \{Z_{k,t}^j\}_{(k,t)}\}_{j=1}^m)$, which reflects the fact that they are sampled in advance. For $t \in [T]$, define $\mathcal{F}_t^A = \sigma\left(\mathcal{F}_{t-1} \cup \sigma(j_t)\right)$, which indicates that the only additional randomness of the algorithm when choosing θ_t is the sampling of $j_t \sim \mathcal{J}_t$. Define the extended perturbation vector as follows:

$$\mathbf{Z}_{KT}^j = \begin{pmatrix} \frac{1}{\sqrt{\lambda}}W^{j\top} & Z_{1,1}^j & \ldots & Z_{1,T}^j & Z_{2,1}^j & \ldots & Z_{K,T}^j \end{pmatrix}^\top \in \mathbb{R}^{d+KT}$$

Removing some components and reordering $\mathbf{Z}_{KT}^j$ yields $\mathbf{Z}_t^j$, where the removal and reordering depend on the sequence of chosen arms, namely $a_1, \ldots, a_{t-1}$. Note that $\mathbf{Z}_{KT}^j \sim \mathcal{N}(\mathbf{0}_{d+KT}, \beta_T^2 I_{d+KT})$. We also define the corresponding extensions of Φ_t and U_t. Define a matrix that has n columns, first a of which are copies of $v \in \mathbb{R}^d$ and the rest are $\mathbf{0}_d$ as follows.

$$\mathrm{rep}(v, a, n) := \begin{pmatrix} v & \ldots & v & \mathbf{0}_d & \ldots & \mathbf{0}_d \end{pmatrix} \in \mathbb{R}^{d\times n}.$$

Define the extended version of Φ_t as follows:

$$\mathbf{\Phi}_t = \begin{pmatrix} \sqrt{\lambda}I_d & \mathrm{rep}(x^1, N_{1,t}, T) & \ldots & \mathrm{rep}(x^K, N_{K,t}, T) \end{pmatrix} \in \mathbb{R}^{d\times(d+KT)}.$$

$\mathbf{\Phi}_t$ extends Φ_t by permuting the columns so that the feature vectors from the same arm appear in consecutive columns, then inserting multiple $\mathbf{0}_d$ appropriately. Then, we have $V_t = \mathbf{\Phi}_t\mathbf{\Phi}_t^\top$ and $\widetilde{\theta}_t = V_t^{-1}\mathbf{\Phi}_t\mathbf{Z}_{KT}^j$, since $\mathbf{Z}_{KT}^j$ is permuted and extended from $\mathbf{Z}_t^j$ in a similar manner. We define the extended version of U_t as $\mathbf{U}_t = (x^{*\top}V_t^{-1}\mathbf{\Phi}_t)^\top \in \mathbb{R}^{d+KT}$. $\mathbf{U}_t$ is also a permutation of U_t with additional zeros inserted. It holds that $U_t^\top\mathbf{Z}_t^j = \mathbf{U}_t^\top\mathbf{Z}_{KT}^j$ and $\|U_t\|_2 = \|\mathbf{U}_t\|_2$.

Let $\mathbb{X}_t$ be the set of all possible $\mathbf{\Phi}_t$. Since $\mathbf{\Phi}_t$ is fully determined by $N_{1,t}, \ldots N_{K,t}$ and each $N_{k,t}$ takes value between 0 and $T-1$ inclusively when $0 \leq t \leq T-1$, we obtain that $\left|\cup_{t=0}^{T-1}\mathbb{X}_t\right| \leq T^K$. For any $t \in [T]$, take any $\mathbf{\Phi}_{t-1} \in \mathbb{X}_{t-1}$. Note that $\mathbf{U}_{t-1} = x^{*\top}(\mathbf{\Phi}_{t-1}\mathbf{\Phi}_{t-1}^\top)^{-1}\mathbf{\Phi}_{t-1}$ is fully determined by $\mathbf{\Phi}_{t-1}$, and $\widetilde{\theta}_{t-1}^j = (\mathbf{\Phi}_{t-1}\mathbf{\Phi}_{t-1}^\top)^{-1}\mathbf{\Phi}_{t-1}\mathbf{Z}_{KT}^j$ is determined by $\mathbf{\Phi}_{t-1}$ and $\mathbf{Z}_{KT}^j$. Assuming that $\mathbf{\Phi}_{t-1}$ is fixed, $\mathbf{U}_{t-1}$ is also fixed, therefore we can apply Fact 1 and obtain that $\mathbb{P}\big(\mathbf{U}_{t-1}^\top\mathbf{Z}_{KT}^j \geq \beta_T\|\mathbf{U}_{t-1}\|_2\big) \geq p_N$, where the only source of randomness comes from $\mathbf{Z}_{KT}^j$. Applying Lemma 4, we obtain that $\mathbb{P}(\|\widetilde{\theta}_{t-1}^j\|_{V_{t-1}} \leq \widetilde{\gamma}_T) \geq 1-\delta/T$. Let $I^j(\mathbf{\Phi}_{t-1}) = \mathbb{1}\{(\mathbf{U}_{t-1}^\top\mathbf{Z}_{KT}^j \geq \beta_T\|\mathbf{U}_{t-1}\|_2) \wedge (\|\widetilde{\theta}_{t-1}^j\|_{V_{t-1}} \leq \widetilde{\gamma}_T)\}$. Then, $\mathbb{P}(I^j(\mathbf{\Phi}_{t-1}) = 1) \geq p_N - \delta/T \geq p_N/2$. Since the only randomness on choosing the arm conditioned on $\mathcal{F}_{t-1}$ comes from sampling j_t, it holds that

$$\mathbb{P}\left(\left(\mathbf{U}_{t-1}^\top\mathbf{Z}_{KT}^{j_t} \geq \beta_T\left\|\mathbf{U}_{t-1}\right\|_2\right) \wedge \left(\left\|\widetilde{\theta}_{t-1}^{j_t}\right\|_{V_{t-1}} \leq \widetilde{\gamma}_T\right) \mid \mathcal{F}_{t-1}\right) = \frac{1}{m}\sum_{j=1}^m I^j\left(\mathbf{\Phi}_{t-1}\right).$$

https://doi.org/10.52202/079017-2947

We apply Azuma-Hoeffding inequality to show that $\frac{1}{m}\sum_{j=1}^{m} I^j\left(\mathbf{\Phi}_{t-1}\right)$ is bounded below with high probability. Since $\{I^j(\mathbf{\Phi}_{t-1})\}_{j=1}^m$ are i.i.d. Bernoulli random variables with the associated probability greater than $p_N/2$, it holds that

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{m}\sum_{j=1}^{m} I^j(\mathbf{\Phi}_{t-1}) \leq \frac{p_N}{4}\right) \\
&= \mathbb{P}\left(\frac{1}{m}\sum_{j=1}^{m} I^j(\mathbf{\Phi}_{t-1}) - \frac{1}{m}\sum_{j=1}^{m}\mathbb{E}[I^j(\mathbf{\Phi}_{t-1})] \leq \frac{p_N}{4} - \frac{1}{m}\sum_{j=1}^{m}\mathbb{E}[I^j(\mathbf{\Phi}_{t-1})]\right) \\
&\leq \mathbb{P}\left(\frac{1}{m}\sum_{j=1}^{m} I^j(\mathbf{\Phi}_{t-1}) - \frac{1}{m}\sum_{j=1}^{m}\mathbb{E}[I^j(\mathbf{\Phi}_{t-1})] \leq \frac{p_N}{4} - \frac{p_N}{2}\right) \\
&= \mathbb{P}\left(\frac{1}{m}\sum_{j=1}^{m} I^j(\mathbf{\Phi}_{t-1}) - \frac{1}{m}\sum_{j=1}^{m}\mathbb{E}[I^j(\mathbf{\Phi}_{t-1})] \leq -\frac{p_N}{4}\right) \\
&\leq \exp\left(-\frac{p_N^2 m}{8}\right),
\end{aligned}$$

where Lemma 12 is applied at the end. By taking the union bound over $\cup_{t=0}^{T-1}\mathbb{X}_t$, we obtain that

$$\mathbb{P}\left(\exists \mathbf{\Phi} \in \cup_{t=0}^{T-1}\mathbb{X}_t, \frac{1}{m}\sum_{j=1}^{m} I^j(\mathbf{\Phi}) \leq \frac{p_N}{4}\right) \leq T^K \exp\left(-\frac{p_N^2 m}{8}\right).$$

The event $\mathcal{E}_2^*$ is defined as the complement of the event above. The proof is complete.

$$\mathcal{E}_2^* := \left\{\omega \in \Omega : \forall \mathbf{\Phi} \in \cup_{t=0}^{T-1}\mathbb{X}_t, \frac{1}{m}\sum_{j=1}^{m} I^j(\mathbf{\Phi}) > \frac{p_N}{4}\right\}.$$

G Proof of Corollary 1

Proof of Corollary 1. Let $\widetilde{\mathcal{E}}_{1,t}$ and $\widetilde{\mathcal{E}}_{2,t}$ be defined as in Lemma 3 with $\widetilde{\gamma} = \widetilde{\gamma}_T$ and $p = p_N/2$. We redefine a couple of notations to adapt Algorithm 2. Let

$$\mathbf{Z}_t := \begin{pmatrix} \frac{1}{\sqrt{\lambda}} W_t^\top & Z_{t,1} & \dots & Z_{t,t-1} \end{pmatrix}^\top \in \mathbb{R}^{d+t-1}$$

to be the perturbation vector at time t, and

$$\widetilde{\theta}_{t-1} := V_{t-1}^{-1}\left(W_t + \sum_{i=1}^{t-1} X_i Z_{t,i}\right)$$

be the perturbation in the estimator θ_t. Regarding $\widetilde{\theta}_{t-1}$ as one of $\widetilde{\theta}_{t-1}^j$ in the proof of Theorem 1, we obtain that $\mathbb{P}(\widetilde{\mathcal{E}}_{1,t}) \geq 1-\delta/T$ and $\mathbb{P}((U_{t-1}^\top \mathbf{Z}_t \geq \beta_{t-1}\|U_{t-1}\|_2) \wedge \widetilde{\mathcal{E}}_{1,t}) \geq p_N/2$ hold, analogously to inequalities (8) and (9) respectively. Moreover, in contrast to the proof of Theorem 1, the perturbation vector $\mathbf{Z}_t$ is now independent of $\mathcal{F}_{t-1}$. Noting that U_{t-1} is $\mathcal{F}_{t-1}$-measurable, it always holds that

$$\mathbb{P}\left((U_{t-1}^\top \mathbf{Z}_t \geq \beta_{t-1}\|U_{t-1}\|_2) \wedge \widetilde{\mathcal{E}}_{1,t} \mid \mathcal{F}_{t-1}\right) \geq \frac{p_N}{2}.$$

This proves that $\widetilde{\mathcal{E}}_{2,t}$ is in fact the whole event. Taking the union bound, we obtain that $\mathbb{P}(\widetilde{\mathcal{E}}^{\mathsf{C}}) \leq \sum_{t=1}^{T}\mathbb{P}(\widetilde{\mathcal{E}}_{1,t}^{\mathsf{C}}) \leq \delta$. By Lemma 3, with probability at least $1-3\delta$, the cumulative regret is bounded by

$$R(T) \leq \gamma_T\left(1 + \frac{4}{p_N}\right)\sqrt{2dT\log\left(1 + \frac{T}{d\lambda}\right)} + \frac{2\gamma_T}{p_N}\sqrt{\frac{2T}{\lambda}\log\frac{1}{\delta}} = \mathcal{O}\left((d\log T)^{\frac{3}{2}}\sqrt{T}\right).$$

□

H Generalizability of Perturbation Distributions

In this section, we demonstrate that any distribution that is symmetric, subGaussian, and has lower-bounded variance satisfies the results of Lemma 4 and Fact 1, possibly up to a constant factor. As mentioned in Remark 3, it implies that our results are valid when the Gaussian distribution is replaced with any symmetric non-degenerate subGaussian distribution. The following lemma is a standard concentration result for vector martingales with subGaussian noises.

Lemma 8 (Theorem 1 in Abbasi-Yadkori et al. [1])**.** *Let $\{\mathcal{F}_t\}_{t=0}^{\infty}$ be a filtration. Let $\{\xi_t\}_{t=1}^{\infty}$ be a sequence of real-valued random variables such that ξ_t is $\mathcal{F}_t$-measurable and is $\mathcal{F}_{t-1}$-conditionally σ-subGaussian for some $\sigma \geq 0$. Let $\{X_t\}_{t=1}^{\infty}$ be a sequence of $\mathbb{R}^d$-valued random vectors such that X_t is $\mathcal{F}_{t-1}$-measurable and $\|X_t\|_2 \leq 1$ almost surely for all $t \geq 1$. Fix $\lambda \geq 1$. Let $V_t = \lambda I + \sum_{i=1}^{t} X_t X_t^\top$. Then, for any $\delta \in (0,1]$, with probability at least $1-\delta$, the following inequality holds for all $t \geq 0$:*

$$\left\| \sum_{i=1}^{t} \xi_i X_i \right\|_{V_t^{-1}} \leq \sigma \sqrt{d \log\left(1 + \frac{t}{d\lambda}\right) + 2\log\frac{1}{\delta}}\,.$$

Next lemma is a simple application of Lemma 8, which proves the concentration result of $\widetilde{\theta}_t$ under the subGaussianity of $\mathcal{P}_R$.

Lemma 9 (Sufficient condition for concentration)**.** *Fix any $\delta \in (0,1]$ and $t \in [T]$. Assume that $\mathbb{P}\left(\|W\|_2 > L_{t,0}^{\delta}\right) \leq \delta$, and $\{Z_i\}_{i=1}^{t-1}$ are mutually independent of each other and are $\mathcal{F}_i^X$-conditionally σ_R^2-subGaussian respectively for each $i \in [t-1]$. Then, with probability at least $1-2\delta$, it holds that*

$$\left\|\widetilde{\theta}_{t-1}\right\|_{V_{t-1}} \leq \sigma_R \sqrt{d \log\left(1 + \frac{T}{d\lambda}\right) + 2\log\frac{1}{\delta}} + \frac{L_{t,0}^{\delta}}{\sqrt{\lambda}}\,.$$

Proof. Recall that $\widetilde{\theta}_{t-1} = V_{t-1}^{-1}(W + \sum_{i=1}^{t-1} X_i Z_i)$. Therefore,

$$\begin{aligned}
\left\|\widetilde{\theta}_{t-1}\right\|_{V_{t-1}} &= \left\|V_{t-1}\widetilde{\theta}_{t-1}\right\|_{V_{t-1}^{-1}} \\
&= \left\|W + \sum_{i=1}^{t-1} X_i Z_{t,i}\right\|_{V_{t-1}^{-1}} \\
&\leq \|W\|_{V_{t-1}^{-1}} + \left\|\sum_{i=1}^{t-1} X_i Z_i\right\|_{V_{t-1}^{-1}}\,,
\end{aligned}$$

where the triangle inequality is used for the last inequality. To bound the first term, we use the fact that $\lambda_{\max}(V_{t-1}^{-1}) \leq \frac{1}{\lambda}$, where $\lambda_{\max}(\cdot)$ denotes the maximum eigenvalue. It implies that $\|W\|_{V_{t-1}^{-1}} \leq \frac{1}{\sqrt{\lambda}}\|W\|_2$. Since $\mathbb{P}(\|W\|_2 > L_{t,0}^{\delta}) \leq \delta$ by assumption, $\|W\|_{V_{t-1}^{-1}} \leq L_{t,0}^{\delta}/\sqrt{\lambda}$ holds with probability at least $1-\delta$. The second term is bounded by Lemma 8. With probability $1-\delta$, it holds that

$$\left\|\sum_{i=1}^{t-1} X_i Z_{t,i}\right\|_{V_{t-1}^{-1}} \leq \sigma_R \sqrt{d \log\left(1 + \frac{t}{d\lambda}\right) + 2\log\frac{1}{\delta}}\,.$$

By taking the union bound over the two events, we obtain that

$$\left\|\widetilde{\theta}_{t-1}\right\|_{V_{t-1}} \leq \frac{L_{t,0}^{\delta}}{\sqrt{\lambda}} + \sigma_R \sqrt{d \log\left(1 + \frac{t}{d\lambda}\right) + 2\log\frac{1}{\delta}}$$

holds with probability at least $1-2\delta$, which proves the the lemma. □

We also provide that the ℓ_2-norm of the vector W whose components are i.i.d. samples of a subGaussian distribution is upper-bounded with high-probability. This lemma, combined with Lemma 9, justifies $\mathcal{P}_I$ to be a distribution over $\mathbb{R}^d$ such that each component is an i.i.d. sample of a subGaussian distribution.

https://doi.org/10.52202/079017-2947

Lemma 10. *Suppose that $W \in \mathbb{R}^d$ and each component of W is sampled i.i.d. from a σ_I^2-subGaussian distribution. Take any $\delta \in (0, 1]$. Then, with probability $1 - \delta$, it holds that*

$$\|W\|_2 \leq \sigma_I \sqrt{2d + 4 \log \frac{1}{\delta}} \,.$$

Proof. Take $X_1 = \mathbf{e}_1, \ldots, X_d = \mathbf{e}_d$, where $\{\mathbf{e}_1, \ldots, \mathbf{e}_d\}$ is the standard basis of $\mathbb{R}^d$. By Lemma 8 with $\xi_i = (W)_i$ and $\lambda = 1$, it holds that

$$\left\| \sum_{i=1}^{d} (W)_i \mathbf{e}_i \right\|_{(2I_d)^{-1}} \leq \sigma_I \sqrt{d \log 2 + 2 \log \frac{1}{\delta}}$$

with probability at least $1 - \delta$. The proof is completed by noting that $\| \sum_{i=1}^{d} (W)_i \mathbf{e}_i \|_{(2I_d)^{-1}} = \frac{1}{\sqrt{2}} \|W\|_2$. □

Finally, we demonstrate that subGaussian distribution with lower-bounded variance satisfies the anti-concentration condition analogous to Fact 1. We normalize the distribution so that it is 1-subGaussian.

Lemma 11 (Sufficient condition for anti-concentration)**.** *Suppose that $\mathcal{P}$ is a real-valued distribution that is symmetric, 1-subGaussian, and has variance at least $1/2$. Suppose $Z \in \mathbb{R}^n$ for some $n \in \mathbb{N}$ and its components are i.i.d. samples of $\mathcal{P}$. Then, for any fixed $u \in \mathbb{R}^n$, it holds that*

$$\mathbb{P}\left(u^\top Z \geq \frac{1}{3} \|u\|_2 \right) \geq 0.01.$$

Proof. Without loss of generality, we may assume that $\|u\|_2 = 1$. Let $Y = u^\top Z$. Then, $\mathrm{Var}(Y) = \mathrm{Var}(\sum_{i=1}^{n} (u)_i (Z)_i) = \sum_{i=1}^{n} (u)_i^2 \mathrm{Var}((Z)_i) = \mathrm{Var}(\mathcal{P}) \geq \frac{1}{2}$. On the other hand, we attain an upper bound of $\mathrm{Var}(Y)$ as follows:

$$\begin{aligned}
\mathrm{Var}(Y) &= \mathbb{E}\left[Y^2\right] \\
&= \mathbb{E}\left[Y^2 \mathbb{1}\left\{|Y| \leq \frac{1}{3}\right\}\right] + \mathbb{E}\left[Y^2 \mathbb{1}\left\{\frac{1}{3} < |Y| \leq 4\right\}\right] + \mathbb{E}\left[Y^2 \mathbb{1}\left\{|Y| \geq 4\right\}\right] \\
&\leq \frac{1}{9} \mathbb{P}\left(|Y| \leq \frac{1}{3}\right) + 16 \mathbb{P}\left(\frac{1}{3} < |Y| \leq 4\right) + \mathbb{E}\left[Y^2 \mathbb{1}\left\{|Y| \geq 4\right\}\right] \\
&\leq \frac{1}{9} + 16 \mathbb{P}\left(\frac{1}{3} < |Y|\right) + \mathbb{E}\left[Y^2 \mathbb{1}\left\{|Y| \geq 4\right\}\right] \,. \qquad (12)
\end{aligned}$$

We upper-bound $\mathbb{E}[Y^2 \mathbb{1}\{|Y| \geq 4\}]$ using its subGaussian property. Note that $Y = u^\top Z$ is 1-subGaussian since $\|u\|_2 = 1$ and the components of Z are independent and 1-subGaussian. Applying the standard tail bound of subGaussian random variables, it holds that $\mathbb{P}\left(|Y| \geq x\right) \leq 2 \exp\left(-x^2/2\right)$, or equivalently, $\mathbb{P}\left(Y^2 \geq x\right) \leq 2 \exp\left(-x/2\right)$. Then, it holds that

$$\begin{aligned}
\mathbb{E}\left[Y^2 \mathbb{1}\left\{Y^2 \geq 16\right\}\right] &= \int_0^\infty \mathbb{P}\left(Y^2 \mathbb{1}\left\{Y^2 \geq 16\right\} \geq x\right) \, dx \\
&= \int_0^{16} \mathbb{P}\left(Y^2 \mathbb{1}\left\{Y^2 \geq 16\right\} \geq x\right) \, dx + \int_{16}^\infty \mathbb{P}\left(Y^2 \mathbb{1}\left\{Y^2 \geq 16\right\} \geq x\right) \, dx \\
&= 16 \mathbb{P}\left(Y^2 \geq 16\right) + \int_{16}^\infty \mathbb{P}\left(Y^2 \geq x\right) \, dx \\
&\leq 32 e^{-8} + \int_{16}^\infty 2 e^{-\frac{x}{2}} \, dx \\
&= 32 e^{-8} + 4 e^{-8} \\
&\leq 0.0121 \,. \qquad (13)
\end{aligned}$$

By plugging in the upper bound of (13) to inequality (12) and reordering the terms, we obtain that

$$\mathbb{P}\left(|Y| > \frac{1}{3}\right) \geq \frac{1}{16} \left(\frac{1}{2} - \frac{1}{9} - 0.0121\right) \geq 0.023 \,.$$

Finally, recall that Z is symmetric, therefore Y is symmetric. Therefore, we have $\mathbb{P}\left(Y \geq \frac{1}{3}\right) \geq \frac{0.023}{2} \geq 0.01$. □

I Auxiliary Lemmas

Lemma 12 (Azuma-Hoeffding Inequality)**.** *Fix $n \in \mathbb{N}$. Let $\{Z_i\}_{i=1}^n$ be a sequence of real-valued random variables adapted to a filtration $\{\mathcal{F}_i\}_{i=0}^n$. Suppose that there exists $a < b$ such that $Z_i \in [a, b]$ holds almost surely for all $i \in [n]$. Then, for any $\delta \in (0, 1]$, the following inequality holds with probability at least $1 - \delta$:*

$$\sum_{i=1}^{n} \left(Z_i - \mathbb{E}\left[Z_i \mid \mathcal{F}_{t-1}\right]\right) \leq (b-a)\sqrt{\frac{n}{2}\log\frac{1}{\delta}}\,.$$

Lemma 13 (Lemma 1 of Laurent and Massart [15])**.** *Let $Y_1, \ldots, Y_d$ be i.i.d. standard Gaussian variables. Set $Z = \sum_{i=1}^{d}(Y_i^2 - 1)$. Then, the following inequality holds for any $x > 0$,*

$$\mathbb{P}(Z \geq \sqrt{dx} + 2x) \leq e^{-x}\,.$$

https://doi.org/10.52202/079017-2947

NeurIPS Paper Checklist

1. **Claims**

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

   Answer: [Yes]

   Justification: We accurately describe the problem we set and the results we obtain in the abstract and introduction, specifically Section 1.

   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: We discuss the limitations of the work at the end of Section 7. Our work is limited to the linear bandit setting, while ensemble sampling also exhibit superior performance in complex settings. However, our results provide a necessary stepping stone to expand the theoretical analysis of ensemble sampling to such regimes.

   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: The full set of assumptions is presented in Section 4. The complete and correct proof is provided throughout Section 5 and the Appendix. All the important ideas are provided as a sketch in the main paper, if not provided as a whole.

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: The paper does not include 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.

https://doi.org/10.52202/079017-2947

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: The paper does not include 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: The paper does not include 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: The paper does not include 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: The paper does not include 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 research conducted in the paper conform, in every respect, with 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: There is no societal impact of the work performed.

    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.

https://doi.org/10.52202/079017-2947

- 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: The paper poses no such risks.

    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: The 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: The 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: 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.
    - 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.

https://doi.org/10.52202/079017-2947