Sparsity-Agnostic Lasso Bandit

Min-hwan Oh [1] **Garud Iyengar** [2] **Assaf Zeevi** [2]

Abstract

We consider a stochastic contextual bandit problem where the dimension d of the feature vectors is potentially large, however, only a sparse subset of features of cardinality $s_0 \ll d$ affect the reward function. Essentially all existing algorithms for sparse bandits require a priori knowledge of the value of the sparsity index s_0. This knowledge is almost never available in practice, and misspecification of this parameter can lead to severe deterioration in the performance of existing methods. The main contribution of this paper is to propose an algorithm that does *not* require prior knowledge of the sparsity index s_0 and establish tight regret bounds on its performance under mild conditions. We also comprehensively evaluate our proposed algorithm numerically and show that it consistently outperforms existing methods, even when the correct sparsity index is revealed to them but is kept hidden from our algorithm.

1. Introduction

In classical multi-armed bandits (MAB), one of the arms is pulled in each round and a reward corresponding to the chosen arm is revealed to the decision-making agent. The rewards are, typically, independent and identically distributed samples from an arm-specific distribution. The goal of the agent is to devise a strategy for pulling arms that maximizes cumulative rewards, suitably balancing between exploration and exploitation. Linear contextual bandits (Abe & Long, 1999; Auer, 2002; Chu et al., 2011) and generalized linear contextual bandits (Filippi et al., 2010; Li et al., 2017) are more recent important extensions of the basic MAB setting, where each arm a is associated with a known feature vector $x_a \in \mathbb{R}^d$, and the expected payoff of the arm is a (typically, monotone increasing) function of the inner product $x_a^\top \beta^*$ for a fixed and unknown parameter vector $\beta^* \in \mathbb{R}^d$. Unlike the traditional MAB problem, here pulling any one arm provides some information about the unknown parameter vector, and hence, insight into the average reward of all the other arms. These contextual bandit algorithms are applicable in a variety of problem settings, such as recommender systems, assortment selection in online retail, and healthcare analytics (Li et al., 2010; Oh & Iyengar, 2019; Tewari & Murphy, 2017), where the contextual information can be used for personalization and generalization.

In most application domains highlighted above, the feature space is high-dimensional ($d \gg 1$), yet typically only a small subset of the features influence the expected reward. That is, the unknown parameter vector is *sparse* with only elements corresponding to the relevant features being non-zero, i.e., the *sparsity index* $s_0 = \|\beta^*\|_0 \ll d$, where the zero norm $\|x\|_0$ counts non-zero entries in the vector x. There is an emerging body of literature on contextual bandit problems with sparse linear reward functions (Abbasi-Yadkori et al., 2012; Gilton & Willett, 2017; Bastani & Bayati, 2020; Wang et al., 2018; Kim & Paik, 2019) which propose methods to exploit the sparse structure under various conditions. However, there is a crucial shortcoming in almost all of these approaches: the algorithms require *prior* knowledge of the sparsity index s_0, information that is almost never available in practice. In the absence of such knowledge, the existing algorithms fail to fully leverage the sparse structure, and their performance does not guarantee the improvements in dimensionality-dependence which can be realized in the sparse problem setting (and can lead to extremely poor performance if s_0 is underspecified). The purpose of this paper is to demonstrate that a relatively simple contextual bandit algorithm that exploits ℓ_1-regularized regression using Lasso (Tibshirani, 1996) in a sparsity-agnostic manner, is provably near-optimal insofar as its regret performance (under suitable regularity). Our contributions are as follows:

(a) We propose the first general sparse bandit algorithm that does not require prior knowledge of the sparsity index s_0.

(b) We establish that the regret bound of our proposed algorithm is $\mathcal{O}(s_0\sqrt{T\log(dT)})$ for the two-armed case, which affords the most accessible exposition of the key analytical ideas. (Extensions to the general K-armed

*Equal contribution [1]Seoul National University, Seoul, South Korea [2]Columbia University, New York, NY, USA. Correspondence to: Min-hwan Oh <minoh@snu.ac.kr>.

Proceedings of the 38^{th} International Conference on Machine Learning, PMLR 139, 2021.

case are discussed later.) The regret bound scale in s_0 and d matches the equivalent terms in the *offline* Lasso results (see the discussions in Section 4.1).

(c) We comprehensively evaluate our algorithm on numerical experiments and show that it consistently outperforms existing methods, even when these methods are granted prior knowledge of the correct sparsity index (and can greatly outperform them if this information is misspecified).

The salient feature of our algorithm is that it does not rely on *forced sampling* which was used by almost all previous work, e.g., Bastani & Bayati (2020); Wang et al. (2018); Kim & Paik (2019), to satisfy certain regularity of the empirical Gram matrix. Forced sampling requires prior knowledge of s_0 because such schemes, the key ideas of which go back to Goldenshluger & Zeevi (2013), need to be fine-tuned using the *correct* sparsity index. (See further discussions in Section 2.4.)

2. Preliminaries

2.1. Notation

For a vector $x \in \mathbb{R}^d$, we use $\|x\|_1$ and $\|x\|_2$ to denote its ℓ_1-norm and ℓ_2 norm respectively, the notation $\|x\|_0$ is reserved for the cardinality of the set of non-zero entries of that vector. We define $[n]$ for a positive integer n to be a set containing positive integers up to n, i.e., $\{1, 2, ..., n\}$. For a real-valued function f , we use $\dot{f}$ and $\ddot{f}$ to denote its first and second derivatives.

2.2. Generalized Linear Contextual Bandits

We consider the stochastic generalized linear bandit problem with K arms. Let T be the problem horizon, namely the number of rounds to be played. In each round $t \in [T]$, the learning agent observes a context consisting of a set of K feature vectors $\mathcal{X}_t = \left\{X_{t,i} \in \mathbb{R}^d \mid i \in [K]\right\}$, where the tuple $\mathcal{X}_t$ is drawn i.i.d. over $t \in [T]$ from an unknown joint distribution with probability density $p_{\mathcal{X}}$ with respect to the Lebesgue measure. Note that the feature vectors for different arms are allowed to be correlated. Each feature vector $X_{t,i}$ is associated with an unknown stochastic reward $Y_{t,i} \in \mathbb{R}$. The agent then selects one arm, denoted by $a_t \in [K]$ and observes the reward $Y_t := Y_{t,a_t}$, corresponding to the chosen arm's feature $X_t := X_{t,a_t}$, as a bandit feedback. The policy consists of the sequence of actions $\pi = \{a_t : t = 1, 2, ...\}$ and is non-anticipating, namely each action only depends on past observations and actions.

In this work, we assume that the reward $Y_{t,i}$ of arm i is given by a generalized linear model (GLM), i.e.

$$Y_{t,i} = \mu(X_{t,i}^\top \beta^*) + \epsilon_{t,i}$$

where $\mu : \mathbb{R} \to \mathbb{R}$ (also known as *inverse link function*) is a *known* increasing function, $\beta^* \in \mathbb{R}^d$ is an *unknown* parameter, and each $\epsilon_{t,i}$ is an independent zero-mean noise. Therefore, $\mathbb{E}[Y_{t,i}|X_{t,i} = x] = \mu(x^\top \beta^*)$ for all $i \in [K]$ and $t \in [T]$. Widely used examples for μ are $\mu(z) = z$ which corresponds to the linear model, and $\mu(z) = 1/(1 + e^{-z})$ which corresponds to the logistic model. The parameter β^* and the feature vectors $\{X_{t,i}\}$ are potentially high-dimensional, i.e., $d \gg 1$, but β^* is *sparse*, that is, the number of non-zero elements in β^*, $s_0 = \|\beta^*\|_0 \ll d$. It is important to note that the agent *does not* know s_0 or the support of the unknown parameter β^*.

We assume that there is an increasing sequence of sigma fields $\{\mathcal{F}_t\}$ such that each $\epsilon_{t,i}$ is $\mathcal{F}_t$-measurable with $\mathbb{E}[\epsilon_{t,i}|\mathcal{F}_{t-1}] = 0$. In our problem, $\mathcal{F}_t$ is the sigma-field generated by random variables of chosen actions $\{a_1, ..., a_t\}$, their features $\{X_{1,a_1}, ..., X_{t,a_t}\}$, and the corresponding rewards $\{Y_{1,a_1}, ..., Y_{t,a_t}\}$. We assume the noise $\epsilon_{t,i}$ for all $i \in [K]$ is sub-Gaussian with parameter σ, where σ is a positive absolute constant, i.e., $\mathbb{E}[e^{\alpha \epsilon_{t,i}}] \le e^{\alpha^2 \sigma^2 / 2}$ for all $\alpha \in \mathbb{R}$. In practice, for bounded reward $Y_{t,i}$, the noise $\epsilon_{t,i}$ is also bounded and hence satisfies the sub-Gaussian assumption with an appropriate σ value.

The agent's goal is to maximize the cumulative expected reward $\mathbb{E}[\sum_{t=1}^T \mu(X_{t,a_t}^\top \beta^*)]$ over T rounds. Let $a_t^* = \text{argmax}_{i \in [K]} \left\{\mu(X_{t,i}^\top \beta^*)\right\}$ denote the optimal arm for each round t. Then, the expected cumulative *regret* of policy $\pi = \{a_1, ..., a_T\}$ is defined as

$$\mathcal{R}^\pi(T) := \sum_{t=1}^T \mathbb{E}\left[\mu(X_{t,a_t^*}^\top \beta^*) - \mu(X_{t,a_t}^\top \beta^*)\right] .$$

Hence, maximizing the expected cumulative rewards of policy π over T rounds is equivalent to minimizing the cumulative regret $\mathcal{R}^\pi(T)$. Note that all the expectations and probabilities throughout the paper are with respect to feature vectors and noise unless explicitly stated otherwise.

2.3. Lasso for Generalized Linear Models

For given samples $Y_1, ..., Y_n$ and corresponding features $X_1, ..., X_n$, the Lasso (Tibshirani, 1996) estimate for the generalized linear model can be defined as

$$\hat{\beta}_n \in \underset{\beta}{\text{argmin}} \left\{\ell_n(\beta) + \lambda \|\beta\|_1\right\} \tag{1}$$

where $\ell_n(\beta) := -\frac{1}{n} \sum_{j=1}^n \left[Y_j X_j^\top \beta - m(X_j^\top \beta)\right]$, $m(\cdot)$ is infinitely differentiable with $\dot{m}(X^\top \beta^*) = \mathbb{E}[Y|X] = \mu(X^\top \beta^*)$, and λ is a penalty parameter. Lasso is known to be an efficient (offline) tool for estimating the high-dimensional linear regression parameter. The "fast convergence" property of Lasso is guaranteed when data are i.i.d. and when the observed covariates are not highly correlated. The restricted eigenvalue condition (Bickel et al., 2009; Raskutti et al., 2010), the compatibility condition (Van De Geer & Bühlmann, 2009), and the restricted isometry property (Candes & Tao, 2007) have been used to ensure that such high correlations are avoided. In sequential learning settings, however, these conditions are often violated because the observations are adapted to the past and the feature variables of the chosen arms converge to a small region of the feature space as the learning agent updates its arm selection policy.

2.4. Why do existing sparse bandit algorithms require prior knowledge of the sparsity index?

The primary reason that a priori knowledge of sparsity is assumed throughout most of the literature is, roughly speaking, to ensure suitable "size" of the confidence bounds and concentration. For example, (Abbasi-Yadkori et al., 2012) require the parameter s_0 to explicitly construct a high probability confidence set with its radius proportional to s_0 rather than d. The recently proposed bandit algorithms of (Bastani & Bayati, 2020; Kim & Paik, 2019) and the variant with MCP estimator in (Wang et al., 2018) employ a logic that is similar in spirit (though different in execution). Specifically, the compatibility condition or restricted eigenvalue condition is assumed to hold only for the theoretical Gram matrix, and the empirical Gram matrix may not satisfy such condition (the difficulty in controlling that is due to the non-i.i.d. adapted samples of the feature variables). As a remedy to this issue, (Bastani & Bayati, 2020) and (Wang et al., 2018) utilize the forced-sampling technique of (Goldenshluger & Zeevi, 2013) to obtain a "sufficient" number of i.i.d. samples and use that to show that the empirical Gram matrices concentrate in the vicinity of the theoretical Gram matrix, and hence, satisfy the compatibility condition after a sufficient amount of forced-sampling. The forced-sampling duration needs to be predefined and scales at least polynomially in the sparsity s_0 to ensure concentration of the Gram matrices. That is, if the algorithm does not know s_0, the forced-sampling duration will have to scale polynomially in d. (Kim & Paik, 2019) propose an alternative to forced sampling that builds on doubly-robust techniques used in the missing data literature; however, their algorithm involves random arm selection with a probability that is calibrated using s_0, and initial uniform sampling whose duration requires knowledge of s_0 and scales polynomially with s_0 in order to establish their regret bounds. The sensitivity to the sparsity index specification is also evident in cases where its value is *misspecified* which may result in severe deterioration in the performance of the algorithm (see further discussion in Section 5.1).

The key observation in our analysis is that, under some mild conditions, i.i.d. samples, which are the key output of the forced sampling scheme, are in fact not essential. We show that the empirical Gram matrix satisfies the required regularity after a sufficient number of rounds, provided the theoretical Gram matrix is also regular; the details of this analysis are in Section 4. Numerical experiments support this findings, and moreover, demonstrate that the performance of the algorithm can be superior to forced-sampling-based schemes that are tuned with foreknowledge of the parameter s_0.

3. Proposed Algorithm

Our proposed SPARSITY-AGNOSTIC (SA) LASSO BANDIT algorithm for high-dimensional GLM bandits is summarized in Algorithm 1. As the name suggests, our algorithm does not require prior knowledge of the sparsity index s_0. It relies on Lasso for parameter estimation, and does not explicitly use exploration strategies or forced-sampling. Instead, in each round, we choose an arm which maximizes the inner product of a feature vector and the Lasso estimate. After observing the reward, we update the regularization parameter λ_t and update the Lasso estimate $\hat{\beta}_t$ which minimizes the penalized negative log-likelihood function defined in (1).

SA LASSO BANDIT requires only one input parameter λ_0. We show in Section 4 that $\lambda_0 = 2\sigma x_{\max}$ where $x_{\max}$ is a bound on the ℓ_2-norm of the feature vectors $X_{t,i}$. Thus, λ_0 does *not* depend on the sparsity index s_0 or the underlying parameter β^*. (Note that, in comparison, Kim & Paik (2019) require three tuning parameters, and Bastani & Bayati (2020) and Wang et al. (2018) require four tuning parameters, most of which are functions of the unknown sparsity index s_0.) It is worth noting that tuning parameters, while helping to achieve low regret, are challenging to specify in online learning settings. In contrast, our proposed algorithm is practical and easy to implement.

Algorithm 1 SA LASSO BANDIT

```
Input parameter: λ_0
for all t = 1 to T do
    Observe X_{t,i} for all i ∈ [K]
    Compute a_t = argmax_{i∈[K]} X_{t,i}^⊤ β̂_t
    Pull arm a_t and observe Y_t
    Update λ_t ← λ_0 sqrt((4 log t + 2 log d)/t)
    Update β̂_{t+1} ← argmin_β {ℓ_t(β) + λ_t ||β||_1}
end for
```

Input parameter: λ_0
for all $t = 1$ to T **do**
  Observe $X_{t,i}$ for all $i \in [K]$
  Compute $a_t = \operatorname{argmax}_{i\in[K]} X_{t,i}^\top \hat{\beta}_t$
  Pull arm a_t and observe Y_t
  Update $\lambda_t \leftarrow \lambda_0 \sqrt{\frac{4\log t + 2\log d}{t}}$
  Update $\hat{\beta}_{t+1} \leftarrow \operatorname{argmin}_\beta \{\ell_t(\beta) + \lambda_t \|\beta\|_1\}$
end for

Discussion of the algorithm. Algorithm 1 may appear to be an *exploration-free* greedy algorithm (see e.g., Bastani et al. 2020), but this is not the case. To better understand this we will compare the steps in Algorithm 1 to upper-confidence bound (UCB) algorithms. A UCB algorithm constructs a high-probability confidence ellipsoid around a *greedy* maximum likelihood estimate and chooses a parameter value within the ellipse that maximizes the reward.

Once the UCB estimate is chosen, the action selection is greedy with respect to the parameter estimate.[1] The UCB algorithms regularize parameter estimates by carefully controlling the size of the confidence ellipsoid to ensure convergence, thus, exploration is loosely equivalent to regularizing the parameter estimate. The algorithm we propose also computes the parameter estimate by *regularizing* the MLE with a sparsifying norm, and then, as in UCB, takes a greedy action with respect to this regularized parameter estimate. We adjust the penalty parameter associated with the sparsifying norm over time at carefully specified rate in order to ensure that our estimate is consistent as we collect more samples. (This adjustment and specification do not require knowledge of sparsity s_0.) Incorrect choice for the penalty parameter would lead to large regret, which is analogous to poor choice of confidence widths in UCB.

4. Regret Analysis

In this section, we establish an upper bound on the expected regret of SA LASSO BANDIT for the two-armed generalized linear bandits. We focus on the two-arm case primarily for clarity and accessibility of key analysis ideas. We later extend our analysis to the K-armed case with $K \geq 3$ in Section 5. It is important to note that our proposed algorithm does not change with the number of arms. We start with an assumption standard in the (generalized) linear bandit literature.

Assumption 1 (Feature set and parameter)**.** *There exists a positive constant $x_{\max}$ such that $\|x\|_2 \leq x_{\max}$ for all $x \in \mathcal{X}_t$ and all t, and a positive constant b such that $\|\beta^*\|_2 \leq b$.*

Assumption 2 (Link function)**.** *There exist $\kappa_0 > 0$ and $\kappa_1 < \infty$ such that the derivative $\dot{\mu}(\cdot)$ of the link function satisfies $\kappa_0 \leq \dot{\mu}(x^\top \beta) \leq \kappa_1$ for all x and β.*

Clearly for the linear link function, $\kappa_0 = \kappa_1 = 1$. For the logistic link function, we have $\kappa_1 = 1/4$.

Definition 1 (Active set and sparsity index)**.** *The active set $S_0 := \{j : \beta^*_j \neq 0\}$ is the set of indices j for which β^*_j is non-zero, and the sparsity index $s_0 = |S_0|$ denotes the cardinality of the active set S_0.*

For the active set S_0, and an arbitrary vector $\beta \in \mathbb{R}^d$, we can define

$$\beta_{j,S_0} := \beta_j \mathbb{1}\{j \in S_0\}\,, \quad \beta_{j,S_0^c} := \beta_j \mathbb{1}\{j \notin S_0\}\,.$$

Thus, $\beta_{S_0} = [\beta_{1,S_0}, ..., \beta_{d,S_0}]^\top$ has zero elements outside the set S_0 and the components of $\beta_{S_0^c}$ can only be non-zero in the complement of S_0. Let $\mathbb{C}(S_0)$ denote the set of vectors

$$\mathbb{C}(S_0) := \{\beta \in \mathbb{R}^d \mid \|\beta_{S_0^c}\|_1 \leq 3\|\beta_{S_0}\|_1\}\,. \tag{2}$$

[1]Likewise, in Thompson sampling (Thompson, 1933), the agent chooses the greedy action for the sampled parameter.

Let $\mathbf{X} \in \mathbb{R}^{K\times d}$ denote the design matrix where each row is a feature vector for an arm. (Although we focus on $K = 2$ case in this section, the definitions and the assumptions introduced here also apply to the case of $K \geq 3$.) Then, in keeping with the previous literature on sparse estimation and specifically on sparse bandits (Bastani & Bayati, 2020; Wang et al., 2018; Kim & Paik, 2019), we assume that the following compatibility condition is satisfied for the theoretical Gram matrix $\Sigma := \frac{1}{K}\mathbb{E}[\mathbf{X}^\top \mathbf{X}]$.

Assumption 3 (Compatibility condition)**.** *For active set S_0, there exists compatibility constant $\phi_0^2 > 0$ such that*

$$\phi_0^2 \|\beta_{S_0}\|_1^2 \leq s_0 \beta^\top \Sigma \beta \quad \textit{for all } \beta \in \mathbb{C}(S_0)\,.$$

We add to this the following mild assumption that is more specific to our analysis.

Assumption 4 (Relaxed symmetry)**.** *For a joint distribution $p_{\mathcal{X}}$, there exists $\nu < \infty$ such that $\frac{p_{\mathcal{X}}(-\mathbf{x})}{p_{\mathcal{X}}(\mathbf{x})} \leq \nu$ for all $\mathbf{x}$.*

Discussion of the assumptions. Assumptions 1 and 2 are the standard regularity assumptions used in the GLM bandit literature (Filippi et al., 2010; Li et al., 2017; Kveton et al., 2020). It is important to note that unlike the existing GLM bandit algorithms which explicitly use the value of κ_0, our proposed algorithm does not use κ_0 or κ_1 — this information is only needed to establish the regret bound. The compatibility condition in Assumption 3 is analogous to the standard positive-definite assumption on the Gram matrix for the ordinary least squares estimator for linear models but is less restrictive. The compatibility condition ensures that truly active components of the parameter vector are not "too correlated." As mentioned above, the compatibility condition is a standard assumption in the sparse bandit literature (Bastani & Bayati, 2020; Wang et al., 2018; Kim & Paik, 2019). Assumption 4 states that the joint distribution $p_{\mathcal{X}}$ can be skewed but this skewness is bounded. Obviously, if $p_{\mathcal{X}}$ is symmetrical, we have $\nu = 1$. Assumption 4 is satisfied for a large class of continuous and discrete distributions, e.g., elliptical distributions including Gaussian and truncated Gaussian distributions, multi-dimensional uniform distribution, and Rademacher distribution. Note that in the non-sparse low dimensional setting (i.e., $d = s$), the relaxed symmetry in Assumption 4 together with the positive definiteness of the theoretical Gram matrix is equivalent to the covariate diversity condition introduced in (Bastani et al., 2020). However, in the sparse high-dimensional setting considered here, the relaxed symmetry does not imply diversity in all covariates. Consequently, the greedy parameter estimation approach proposed by (Bastani et al., 2020) is not guaranteed to achieve a sublinear regret. As in the case of κ_0 and κ_1 in Assumption 2, the parameter ν is only needed to establish the regret bound, our proposed algorithm does not require knowledge of ν.

4.1. Regret Bound for SA Lasso Bandit

Theorem 1 (Regret bound for two arms)**.** *Suppose $K = 2$ and Assumptions 2-4 hold. Then the expected regret of the* SA Lasso Bandit *policy (π) over horizon T is upper-bounded by*

$$\mathcal{R}^\pi(T) \le 4\kappa_{\max} + \frac{2\log(2d^2)+2}{C_0(\phi_0, s_0)^2} + \frac{32\kappa_{\max}\rho_0\sigma s_0\sqrt{T\log(dT)}}{\kappa_{\min}\phi_0^2}$$

where $C_0(\phi_0, s_0) = \min\left(\frac{1}{2}, \frac{\phi_0^2}{128 s_0 \rho_0}\right)$.

Discussion of Theorem 1. In terms of key problem primitives, Theorem 1 establishes a regret bound of $\mathcal{O}(s_0\sqrt{T\log(dT)})$ without any prior knowledge on s_0. The bound shows that the regret of SA Lasso Bandit grows at most logarithmically in feature dimension d. The key takeaway from this theorem is that SA Lasso Bandit is sparsity-agnostic and is able to achieve *correct* dependence on parameters d and s_0. Based on the offline Lasso convergence results under the compatibility condition (e.g., Theorem 6.1 in (Bühlmann & Van De Geer, 2011)), we believe that the dependence on d and s_0 in Theorem 1 is best possible.[2]

The regret bound in Theorem 1 is tighter than the previously known bound in the same problem setting (Kim & Paik, 2019) although direct comparison is not immediate, given the difference in assumptions involved — compared to (Kim & Paik, 2019), we require Assumption 4 whereas they assume the sparsity index s_0 is known. Having said that, the numerical experiments in Section 6 support our theoretical claims and provide additional evidence that our proposed algorithm compares very favorably to other existing methods (which are tuned with the knowledge of the correct s_0), and moreover, the performance is not sensitive to several of our assumptions that were imposed primarily for technical tractability purposes. As mentioned earlier, the previous work on sparse bandits (Bastani & Bayati, 2020; Wang et al., 2018; Kim & Paik, 2019) require the knowledge of sparsity. In the absence of such knowledge, if sparsity is underspecified, then these algorithms would suffer a regret linear in T. On the other hand, if the sparsity is overspecified, the regret of these algorithms scales with d instead of s_0. Our proposed algorithm does not require such prior knowledge, hence there is no risk of under or over specification, and yet our analysis provides a sharper regret guarantee. Furthermore, our result also suggests that even when the sparsity is known, random sampling to satisfy the compatibility condition, invoked by all existing sparse bandit algorithms to date, can be wasteful since said conditions may be already satisfied even in the absence of such sampling. This finding is also supported by the numerical experiments in Section 6 and additional experiments in the appendix. We provide the outline of the proof and the key lemmas in the following section.

[2]Since the horizon T does not exist in offline Lasso results, it is not straightforward to see whether $\sqrt{T}$ dependence can be improved comparing only with the offline Lasso results. Clearly, without an additional assumption on the separability of the arms, we know that poly-logarithmic scalability in T is not feasible. We briefly discuss our conjecture in comparison with the lower bound result in the non-sparse linear bandits in Secton B.1 in the appendix where we discuss the regret bound under the RE condition.

4.2. Challenges and Proof Outlines

There are two essential challenges that prevent us from fully benefiting from the fast convergence property of Lasso:

(i) The samples induced by our bandit policy are not i.i.d., therefore the standard Lasso oracle inequality does not hold.

(ii) Empirical Gram matrices do not necessarily satisfy the compatibility condition even under Assumption 3. This is because the selected feature variables for which the rewards are observed do not provide an "even" representation for the entire distribution.

To resolve (i), we provide a Lasso oracle inequality for the GLM with non-i.i.d. adapted samples under the compatibility condition in Lemma 1. For (ii), we aim to provide a remedy without using the knowledge of sparsity or without using i.i.d. samples. Hence, this poses a greater challenge. In Section 4.2.2, we address this issue by showing that the empirical Gram matrix behaves "nicely" even when we choose arms adaptively without deliberate random sampling. In particular, we show that adapted Gram matrices can be controlled by the theoretical Gram matrix, and the empirical Gram matrix concentrates properly around the adapted Gram matrix as we collect more samples. Connecting this matrix concentration to the corresponding compatibility constants, we show that the empirical Gram matrix satisfies the compatibility condition with high probability.

4.2.1. Lasso Oracle Inequality for GLM with Non-i.i.d. Data.

We present an oracle inequality for the Lasso estimator for GLM under non-i.i.d. data. This is a generalization of the standard Lasso oracle inequality (Bühlmann & Van De Geer, 2011) that allows adapted sequences of observations. This result may be of independent interest.

Lemma 1 (Oracle inequality)**.** *Suppose the compatibility condition holds for the empirical covariance matrix* $\hat{\Sigma}_t = \frac{1}{t}\sum_{\tau=1}^{t} X_\tau X_\tau^\top$ *with active set S_0 and compatibility constant ϕ_t. For some $\delta \in (0,1)$, define the regularization parameter* $\lambda_t := 2\sigma\sqrt{\frac{2[\log(2/\delta)+\log d]}{t}}$. *Then with proba-*

bility at least $1-\delta$*, the Lasso estimate* $\hat{\beta}_t$ *defined in* (1) *satisfies*

$$\|\hat{\beta}_t - \beta^*\|_1 \le \frac{4s_0\lambda_t}{\kappa_{\min}\phi_t^2}\,.$$

Note that here we assume that the compatibility condition holds for the empirical Gram matrix $\hat{\Sigma}_t$. In the next section, we show that this holds with high probability. The Lasso oracle inequality holds without further assumptions on the underlying parameter β^* or its support. Therefore, if we show that $\hat{\Sigma}_t$ satisfies the compatibility condition absent knowledge of s_0, then the remainder of the result does not require this knowledge as well.

4.2.2. Compatibility Condition and Matrix Concentration.

For matrix M, we define $\phi^2(M, S_0) := \min_\beta \left\{ s_0\beta^\top M\beta/\|\beta_{S_0}\|_1^2 : \|\beta_{S_0^c}\|_1 \le 3\|\beta_{S_0}\|_1 \neq 0 \right\}$ as the (generic) compatibility constant. Hence, it suffices to show $\phi^2(M, S_0) > 0$ in order to show that matrix M satisfies the compatibility condition. Now, under Assumption 3, the theoretical Gram matrix $\Sigma = \frac{1}{K}\mathbb{E}[\mathbf{X}^\top\mathbf{X}]$ satisfies the compatibility condition i.e., $\phi_0^2 = \phi^2(\Sigma, S_0) > 0$.

Definition 2. *We define the adapted Gram matrix as* $\Sigma_t := \frac{1}{t}\sum_{\tau=1}^t \mathbb{E}[X_\tau X_\tau^\top | \mathcal{F}_\tau]$ *and the empirical Gram matrix as* $\hat{\Sigma}_t := \sum_{\tau=1}^t X_\tau X_\tau^\top$.

For each $\mathbb{E}[X_\tau X_\tau^\top|\mathcal{F}_\tau]$ in Σ_t, the history $\mathcal{F}_\tau$ affects how the feature vector X_τ is chosen. More specifically, our algorithm uses $\mathcal{F}_\tau$ to compute $\hat{\beta}_\tau$ and then chooses arm a_τ such that its (realized) feature x_{a_τ} maximizes $x_{a_\tau}^\top\hat{\beta}_\tau$. Therefore, we can rewrite Σ_t as

$$\Sigma_t = \frac{1}{t}\sum_{\tau=1}^t\sum_{i=1}^2 \mathbb{E}_{\mathcal{X}_\tau}\left[X_{\tau,i}X_{\tau,i}^\top \mathbb{1}\{X_{\tau,i} = \operatorname*{argmax}_{X\in\mathcal{X}_\tau} X^\top\hat{\beta}_\tau\}|\hat{\beta}_\tau\right].$$

Since the compatibility condition is satisfied only for the theoretical Gram matrix Σ and we need to show the empirical Gram matrix $\hat{\Sigma}_t$ satisfies the compatibility condition, the adapted Gram matrix Σ_t serves as a bridge between Σ and $\hat{\Sigma}_t$ in our analysis. We first lower-bound the compatibility constant $\phi^2(\Sigma_t, S_0)$ in terms of $\phi^2(\Sigma, S_0)$ so that we can show that Σ_t satisfies the compatibility condition as long as Σ satisfies the compatibility condition. Then, we show that $\hat{\Sigma}_t$ concentrates around Σ_t with high probability and that such matrix concentration guarantees the compatibility condition of $\hat{\Sigma}_t$.

In Lemma 2, we show that Σ_t can be controlled in terms of the theoretical Gram matrix Σ, which allows us to link the compatibility constant of Σ to compatibility constant of Σ_t. Note that Lemma 2 shows the result for any fixed vector β; hence can be applied to $\mathbb{E}[X_\tau X_\tau^\top|\mathcal{F}_\tau]$.

Lemma 2. *For fixed* $\beta \in \mathbb{R}^d$*, we have* $\sum_{i=1}^2 \mathbb{E}\left[X_{t,i}X_{t,i}^\top \mathbb{1}\{X_{t,i} = \operatorname{argmax}_{X\in\mathcal{X}_t} X^\top\beta\}\right] \succcurlyeq \frac{\Sigma}{\rho_0}$.

Therefore, we have $\Sigma_t \succcurlyeq \frac{\Sigma}{\rho_0}$ which implies that $\phi^2(\Sigma_t, S_0) \ge \frac{\phi^2(\Sigma, S_0)}{\rho_0} > 0$, i.e., Σ_t satisfies the compatibility condition. Note that both Σ and Σ_t can be singular. In Lemma 3, we show that $\hat{\Sigma}_t$ concentrates to Σ_t with high probability. This result is crucial in our analysis since it allows the matrix concentration without using i.i.d. samples. The proof of Lemma 3 utilizes a new Bernstein-type inequality for adapted samples (Lemma 8 in the appendix) which may be of independent interest.

Lemma 3. *For* $t \ge \frac{2\log(2d^2)}{C(\phi_0,s_0)^2}$ *where* $C(\phi_0, s_0) = \min\left(\frac{1}{2}, \frac{\phi_0^2}{128s_0\rho_0}\right)$*, we have*

$$\mathbb{P}\left(\|\Sigma_t - \hat{\Sigma}_t\|_\infty \ge \frac{\phi_0^2}{32s_0\rho_0}\right) \le \exp\left\{-\frac{tC(\phi_0,s_0)^2}{2}\right\}.$$

Then, we invoke the following corollary to use the matrix concentration results to ensure the compatibility condition for $\hat{\Sigma}_t$.

Corollary 1 (Corollary 6.8, (Bühlmann & Van De Geer, 2011))**.** *Suppose that* Σ_0*-compatibility condition holds for the index set* S *with cardinality* $s = |S|$*, with compatibility constant* $\phi^2(\Sigma_0, S)$*, and that* $\|\Sigma_1 - \Sigma_0\|_\infty \le \Delta$*, where* $32s\Delta \le \phi^2(\Sigma_0, S)$*. Then, for the set* S*, the* Σ_1*-compatibility condition holds as well, with* $\phi^2(\Sigma_1, S) \ge \phi^2(\Sigma_0, S)/2$*.*

In order to satisfy the hypotheses for Lemma 3 and Corollary 1, we define the *initial* period $t < T_0 := 2\log(2d^2)/C(\phi_0, s_0)^2$ during which the compatibility condition for the empirical Gram matrix is not guaranteed, and the event $\mathcal{E}_t := \left\{\|\Sigma_t - \hat{\Sigma}_t\|_\infty \le \phi_0^2/(32s_0\rho_0)\right\}$. Then for all $t \ge \lceil T_0 \rceil$ and Σ_t for which event $\mathcal{E}_t$ holds, we have

$$\phi_t^2 := \phi^2(\hat{\Sigma}_t, S_0) \ge \frac{\phi^2(\Sigma_t, S_0)}{2} \ge \frac{\phi_0^2}{2\rho_0} > 0\,.$$

Hence, the compatibility condition is satisfied for the empirical Gram matrix without using sparsity information.

4.2.3. Proof Sketch of Theorem 1

We combine the results above to analyze the regret bound of SA Lasso Bandit shown in Theorem 1. First, we divide the time horizon $[T]$ into three groups:

(a) $(t \le T_0)$. Here the compatibility condition is not guaranteed to hold.

(b) $(t > T_0)$ such that $\mathcal{E}_t$ holds.

(c) $(t > T_0)$ such that $\mathcal{E}_t$ does not hold.

These sets are disjoint, hence we bound the regret contribution from each separately and obtain an upper bound on the overall regret. It is important to note that SA LASSO BANDIT Algorithm does not rely in any way on this partitioning – it is introduced purely for the purpose of analysis. Set (a) is the initial period over which we do not have guarantees for the compatibility condition. Therefore, we cannot apply the Lasso convergence result; hence we can incur $\mathcal{O}(s_0^2 \log d)$ regret. Set (b) is where the compatibility condition is satisfied; hence the Lasso oracle inequality in Lemma 1 can apply. In fact, this group can be further divided to two cases: (b-1) when the high-probability Lasso result holds and (b-2) when it does not, where the regret of (b-2) can be bounded by $\mathcal{O}(1)$. For (b-1), using the Lasso convergence result and summing the regret over the time horizon gives $\mathcal{O}(s_0\sqrt{T\log(dT)})$ regret, which is the leading factor in the regret bound of Theorem 1. Lastly, (c) contains the failure events of Lemma 3 whose regret is $\mathcal{O}(s_0^2)$. The proofs of the lemmas are deferred to the appendix.

5. Extension to *K* Arms

Recall that SA LASSO BANDIT is valid for any number of arms; hence, no modifications are required to extend the algorithm to $K \geq 3$ arms. The analysis of SA LASSO BANDIT for the K-armed case tackles largely the same challenges described in Section 4.2: the need for a Lasso convergence result for adapted samples and ensuring the compatibility condition without knowing s_0 (and without relying on i.i.d. samples). The former challenge is again taken care of by the Lasso convergence result in Lemma 1. However, the latter issue is more subtle in the K-armed case than in the two-armed case. In particular, when controlling the adapted Gram matrix Σ_t with the theoretical Gram matrix Σ, the Gram matrix for the unobserved feature vectors could be incomparable with the Gram matrix for the observed feature vectors. For this issue, we introduce an additional regularity condition, which we denote as the "balanced covariance" condition.

Assumption 5 (Balanced covariance)**.** *Consider a permutation $(i_1, ..., i_K)$ of $(1, ..., K)$. For any integer $k \in \{2, ..., K-1\}$ and fixed vector β, there exists $C_{\mathcal{X}} < \infty$ such that*

$$\mathbb{E}\left[X_{i_k} X_{i_k}^\top \mathbb{1}\{X_{i_1}^\top \beta < ... < X_{i_K}^\top \beta\}\right] \preccurlyeq C_{\mathcal{X}} \mathbb{E}\left[(X_{i_1} X_{i_1}^\top + X_{i_K} X_{i_K}^\top) \mathbb{1}\{X_{i_1}^\top \beta < ... < X_{i_K}^\top \beta\}\right].$$

In Algorithm 1 we observe only the reward corresponding to arm i_1, and Assumption 4 implies that we have some control on the arm i_K. This balanced covariance condition implies that there is "sufficient randomness" in the observed features compared to non-observed features. The exact value of $C_{\mathcal{X}}$ depends on the joint distribution of $\mathcal{X}$ including the correlation between arms. In general, the more positive the correlation, the smaller $C_{\mathcal{X}}$ (obviously, with an extreme case of perfectly correlated arms having a constant $C_{\mathcal{X}}$ independent of any problem parameters). When the arms are independent and identically distributed, Assumption 5 holds with $C_{\mathcal{X}} = \mathcal{O}(1)$ for both the multivariate Gaussian distribution and a uniform distribution on a sphere, and for an arbitrary independent distribution for each arm, Assumption 5 holds for $C_{\mathcal{X}} = \binom{K-1}{K_0}$ where $K_0 = \lceil (K-1)/2 \rceil$. It is important to note that even in this pessimistic case, $C_{\mathcal{X}}$ does not exhibit dependence on dimensionality d or the sparsity index s_0. These are formalized in Proposition 1 in the appendix.[3] This balanced covariance condition is somewhat similar to "positive-definiteness" condition for observed contexts in the bandit literature (e.g., Goldenshluger & Zeevi (2013); Bastani et al. (2020)). However, notice that we allow the covariance matrices on both sides of the inequality to be singular. Hence, the positive-definiteness condition for observed context in our setting may not hold even when the balanced covariance condition holds. While this condition admittedly originates from our proof technique, it also provides potential insights on learnability of problem instances. That is, $C_{\mathcal{X}}$ close to infinity implies that the distribution of feature vectors is heavily skewed toward a particular direction. Hence, learning algorithms may require many more samples to learn the unknown parameter, leading to larger regret. It is important to note that our algorithm does not require any prior information on $C_{\mathcal{X}}$. The regret bound for the K-armed sparse bandits under Assumption 5 is as follows.

Theorem 2 (Regret bound for K arms)**.** *Suppose $K \geq 3$ and Assumptions 1-4, and 5 hold. Let $\lambda_0 = 2\sigma x_{\max}$. Then the expected cumulative regret of the* SA LASSO BANDIT *policy π over horizon $T \geq 1$ is upper-bounded by*

$$\mathcal{R}^\pi(T) \leq 4\kappa_1 + \frac{4\kappa_1 x_{\max} b(\log(2d^2)+1)}{C_1(s_0)^2} + \frac{64\kappa_1 \nu C_{\mathcal{X}} \sigma x_{\max} s_0 \sqrt{T\log(dT)}}{\kappa_0 \phi_0^2}$$

where $C_1(s_0) = \min\left(\frac{1}{2}, \frac{\phi_0^2}{256 s_0 \nu C_{\mathcal{X}} x_{\max}^2}\right)$.

Theorem 2 establishes $\mathcal{O}\left(s_0\sqrt{T\log(dT)}\right)$ regret without prior knowledge on s_0, achieving the same rate as Theorem 1 in terms of the key problem primitives. Since both multivariate Gaussian distributions and uniform distributions satisfy Assumption 4 with $\nu = 1$ and Assumption 5 with $C_{\mathcal{X}} = \mathcal{O}(1)$, the regret bound in Theorem 2 still holds

[3] While it is not our primary goal to derive general tight bounds on $C_{\mathcal{X}}$, we acknowledge that the bound on $C_{\mathcal{X}}$ for an arbitrary distribution for independent arms is very loose, and is the result of conservative analysis driven by lack of information on $p_{\mathcal{X}}$. Numerical evaluation on distributions other than Gaussian and uniform distributions, detailed in Section 5, buttress this point and indicate that the dependence on K is no greater than linear.

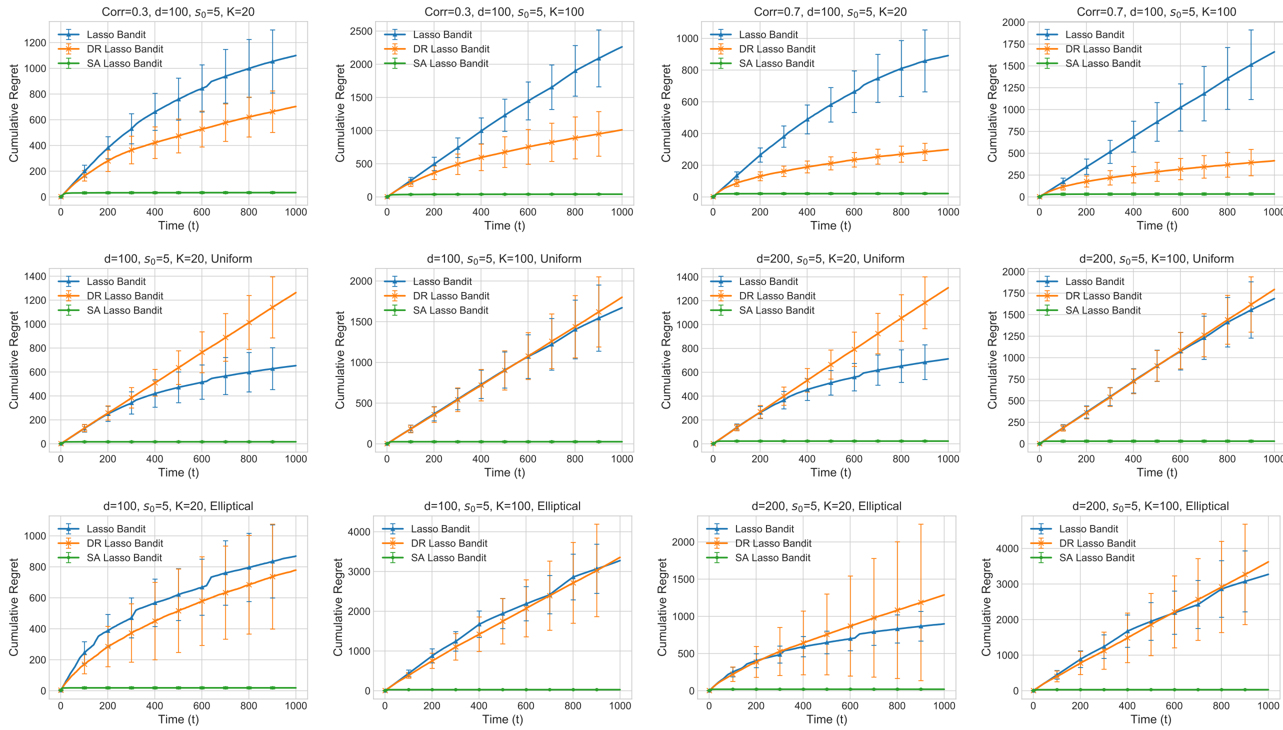

Figure 1. The evaluations of SA LASSO BANDIT (Algorithm 1), DR LASSO BANDIT (Kim & Paik, 2019), and LASSO BANDIT (Bastani & Bayati, 2020). The first row shows results for features drawn from a multivariate Gaussian distribution with varying correlation between arms. The second and third rows show results for uniform and non-Gaussian elliptical distributions respectively. The results provide clear evidence that SA LASSO BANDIT outperforms the benchmarks across various experiments.

under Assumptions 1-3 for these distributions. Therefore, to our knowledge, this is the first sparsity-agnostic regret bound for a general K-armed high-dimensional contextual bandit algorithm even for the Gaussian distribution or uniform distribution.

The proof of Theorem 2 largely follows that of Theorem 1. The main difference is how we control the adapted Gram matrix Σ_t with the theoretical Gram matrix Σ. Under the balanced covariance condition, we can ensure the lower bound of the adapted Gram matrix as a function of the theoretical Gram matrix, which is analogous to the result in Lemma 2. In particular, we show that for a fixed $\beta \in \mathbb{R}^d$,

$$\sum_{i=1}^{K} \mathbb{E}_{\mathcal{X}_t}\Big[X_{t,i}X_{t,i}^\top \mathbb{1}\{X_{t,i} = \operatorname*{argmax}_{X\in\mathcal{X}_t} X^\top\beta\}\Big] \succcurlyeq (2\nu C_{\mathcal{X}})^{-1}\Sigma\,.$$

The formal result is presented in Lemma 10 in the appendix along with its proof. Next, we again invoke the matrix concentration result in Lemma 3 to connect the compatibility constant of empirical Gram matrix $\hat{\Sigma}_t$ to that of Σ_t, and eventually to the theoretical Gram matrix Σ. Thus, we ensure the compatibility condition of $\hat{\Sigma}_t$. The additional regret in the K-armed case as compared to the two-armed case is essentially a scaling by $C_{\mathcal{X}}$ to ensure the balanced covariance condition.

6. Experiments

We conduct numerical experiments to evaluate SA LASSO BANDIT and compare with existing sparse bandit algorithms: DR LASSO BANDIT (Kim & Paik, 2019) and LASSO BANDIT (Bastani & Bayati, 2020). For each case with different experimental configurations, we conduct 20 independent runs. For performance evaluations, we report the average of the cumulative regret for each of the algorithms. The error bars represent the standard deviations. Each row of the plots show experiments using different distributions for feature vectors. Additional results are presented in the appendix. SA LASSO BANDIT exhibits superior performances across different distributions as well as other problem parameters.

The results provide convincing evidence that the performance of our proposed algorithm is superior to the existing sparse bandit methods that we compare with. SA LASSO BANDIT outperforms the existing sparse bandit algorithms by significant margins, even though the correct sparsity index s_0 is revealed to these algorithms and kept hidden from SA LASSO BANDIT. Furthermore, SA LASSO BANDIT is much more practical and simple to implement with a minimal number of a hyperparameter.

References

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

Abbasi-Yadkori, Y., Pal, D., and Szepesvari, C. Online-to-confidence-set conversions and application to sparse stochastic bandits. In *Artificial Intelligence and Statistics*, pp. 1–9, 2012.

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

Agrawal, S. and Goyal, N. Thompson sampling for contextual bandits with linear payoffs. In *International Conference on Machine Learning*, pp. 127–135, 2013.

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

Bang, H. and Robins, J. M. Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61 (4):962–973, 2005.

Bastani, H. and Bayati, M. Online decision making with high-dimensional covariates. *Operations Research*, 68 (1):276–294, 2020.

Bastani, H., Bayati, M., and Khosravi, K. Mostly exploration-free algorithms for contextual bandits. *Management Science*, 2020.

Bickel, P. J., Ritov, Y., and Tsybakov, A. B. Simultaneous analysis of lasso and dantzig selector. *The Annals of Statistics*, 37(4):1705–1732, 2009.

Bühlmann, P. and Van De Geer, S. *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media, 2011.

Cambanis, S., Huang, S., and Simons, G. On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis*, 11(3):368–385, 1981.

Candes, E. and Tao, T. The dantzig selector: Statistical estimation when p is much larger than n. *The annals of Statistics*, 35(6):2313–2351, 2007.

Carpentier, A. and Munos, R. Bandit theory meets compressed sensing for high dimensional stochastic linear bandit. In *Artificial Intelligence and Statistics*, pp. 190–198, 2012.

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

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

Filippi, S., Cappe, O., Garivier, A., and Szepesvári, C. Parametric bandits: The generalized linear case. In *Advances in Neural Information Processing Systems*, pp. 586–594, 2010.

Gilton, D. and Willett, R. Sparse linear contextual bandits via relevance vector machines. In *2017 International Conference on Sampling Theory and Applications (SampTA)*, pp. 518–522. IEEE, 2017.

Goldenshluger, A. and Zeevi, A. A linear response bandit problem. *Stochastic Systems*, 3(1):230–261, 2013.

Kim, G.-S. and Paik, M. C. Doubly-robust lasso bandit. In *Advances in Neural Information Processing Systems*, pp. 5869–5879, 2019.

Kveton, B., Zaheer, M., Szepesvari, C., Li, L., Ghavamzadeh, M., and Boutilier, C. Randomized exploration in generalized linear bandits. In *International Conference on Artificial Intelligence and Statistics*, pp. 2066–2076, 2020.

Lattimore, T. and Szepesvári, C. *Bandit Algorithms*. Cambridge University Press (preprint), 2019.

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

Li, L., Lu, Y., and Zhou, D. Provably optimal algorithms for generalized linear contextual bandits. In *International Conference on Machine Learning*, pp. 2071–2080, 2017.

Oh, M.-h. and Iyengar, G. Thompson sampling for multinomial logit contextual bandits. In *Advances in Neural Information Processing Systems*, pp. 3151–3161, 2019.

Raskutti, G., Wainwright, M. J., and Yu, B. Restricted eigenvalue properties for correlated gaussian designs. *Journal of Machine Learning Research*, 11(Aug):2241–2259, 2010.

Rusmevichientong, P. and Tsitsiklis, J. N. Linearly parameterized bandits. *Mathematics of Operations Research*, 35 (2):395–411, 2010.

Tewari, A. and Murphy, S. A. From ads to interventions: Contextual bandits in mobile health. In *Mobile Health*, pp. 495–517. Springer, 2017.

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

Tibshirani, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

Van De Geer, S. A. and Bühlmann, P. On the conditions used to prove oracle results for the lasso. *Electronic Journal of Statistics*, 3:1360–1392, 2009.

Wainwright, M. J. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press, 2019.

Wang, X., Wei, M., and Yao, T. Minimax concave penalized multi-armed bandit model with high-dimensional covariates. In *International Conference on Machine Learning*, pp. 5200–5208, 2018.

Zhang, C.-H. Nearly unbiased variable selection under minimax concave penalty. *The Annals of statistics*, 38(2): 894–942, 2010.

A. Proofs of Lemmas for Theorem 1

A.1. Proof of Lemma 1

The proof follows from modifying the proof of the standard Lasso oracle inequality (Bühlmann & Van De Geer, 2011) using martingale theory. Recall from (1) that the negative log-likelihood of the GLM is

$$\ell_t(\beta) = -\frac{1}{t}\sum_{\tau=1}^{t}\left[Y_\tau X_\tau^\top \beta - m(X_\tau^\top \beta)\right]$$

where m is a normalizing function with its gradient $\dot{m}(X^\top\beta) = \mu(X^\top\beta)$. Now, we denote the expectation of $\ell_t(\beta)$ over Y by $\bar{\ell}_t(\beta)$:

$$\bar{\ell}_t(\beta) := \mathbb{E}_Y[\ell_t(\beta)] = -\frac{1}{t}\sum_{\tau=1}^{t}\left[\mu(X_\tau^\top\beta^*)X_\tau^\top\beta - m(X_\tau^\top\beta)\right].$$

Note that $\nabla_\beta \bar{\ell}_t(\beta) = -\frac{1}{t}\sum_{\tau=1}^{t}\left[\mu(X_\tau^\top\beta^*) - \mu(X_\tau^\top\beta)\right]X_\tau$. Hence, we have $\nabla_\beta\bar{\ell}_t(\beta^*) = \vec{0}_d$ which implies that $\beta^* = \operatorname{argmin}_\beta \bar{\ell}_t(\beta)$ given the fact that m is convex in the GLM. Hence, for any parameter $\beta \in \mathbb{R}^d$, the excess risk is defined as

$$\mathcal{E}(\beta) := \bar{\ell}_t(\beta) - \bar{\ell}_t(\beta^*).$$

Note that by definition, $\mathcal{E}(\beta) \geq 0$, for all $\beta \in \mathbb{R}^d$ (with $\mathcal{E}(\beta^*) = 0$). The Lasso estimate $\hat{\beta}_t$ for the GLM is given by the minimization of the penalized negative log-likelihood

$$\hat{\beta}_t := \operatorname*{argmin}_\beta \left\{\ell_t(\beta) + \lambda_t\|\beta\|_1\right\}$$

where λ is the penalty parameter whose value needs to be chosen to control the noise of the model. Now, we define the empirical process of the problem as

$$v_t(\beta) := \ell_t(\beta) - \bar{\ell}_t(\beta).$$

Note that the randomness in $\{Y_\tau\}$ still plays a role on $\ell_t(\beta)$ and hence on $v_t(\beta)$. Then by the definition of $\hat{\beta}_t$, we have

$$\ell_t(\hat{\beta}_t) + \lambda_t\|\hat{\beta}_t\|_1 \leq \ell_t(\beta^*) + \lambda_t\|\beta^*\|_1.$$

Adding and subtracting terms, we have

$$\ell_t(\hat{\beta}_t) - \bar{\ell}_t(\hat{\beta}_t) + \bar{\ell}_t(\hat{\beta}_t) - \bar{\ell}_t(\beta^*) + \lambda_t\|\hat{\beta}_t\|_1 \leq \ell_t(\beta^*) - \bar{\ell}_t(\beta^*) + \lambda_t\|\beta^*\|_1.$$

Rearranging terms gives the following "basic inequality" for the GLM

$$\mathcal{E}(\hat{\beta}_t) + \lambda_t\|\hat{\beta}_t\|_1 \leq -[v_t(\hat{\beta}_t) - v_t(\beta^*)] + \lambda_t\|\beta^*\|_1.$$

The basic inequality implies that in order to provide an upper-bound for the penalized excess risk, we need to control the deviation of the empirical process $[v_t(\hat{\beta}_t) - v_t(\beta^*)]$ (Bühlmann & Van De Geer, 2011). And we bound this deviation of the empirical process in terms of the parameter estimation error $\|\hat{\beta}_t - \beta^*\|_1$. Essentially, $[v_t(\hat{\beta}_t) - v_t(\beta^*)]$ is where the random noise plays a role, and with large enough penalization (suitably large λ) we can control such randomness in the empirical process. We define the event of the empirical process being controlled by the penalization.

$$\mathcal{T} := \left\{|v_t(\hat{\beta}_t) - v_t(\beta^*)| \leq \lambda\|\hat{\beta}_t - \beta^*\|_1\right\}. \tag{3}$$

Lemma 4 ensures that we can control this empirical process deviation with high probability. Hence, in the rest of the proof, we restrict ourselves to the case where the empirical process behaves well, i.e., event $\mathcal{T}$ in (3) holds.

Lemma 4. *Assume X_t satisfies $\|X_t\|_2 \leq x_{\max}$ for all t. If $\lambda = \sigma x_{\max}\sqrt{\frac{2[\log(2/\delta)+\log d]}{t}}$, then with probability at least $1-\delta$ we have*

$$|v_t(\hat{\beta}_t) - v_t(\beta^*)| \leq \lambda\|\hat{\beta}_t - \beta^*\|_1.$$

On event $\mathcal{T}$, for $\lambda_t \geq 2\lambda$, we have

$$2\mathcal{E}(\hat{\beta}_t) + 2\lambda_t\|\hat{\beta}_t\|_1 \leq \lambda_t\|\hat{\beta}_t - \beta^*\|_1 + 2\lambda_t\|\beta^*\|_1\,. \tag{4}$$

Let $\hat{\beta} := \hat{\beta}_t$ for brevity. Using the active set S_0, we can define the following:

$$\beta_{j,S_0} := \beta_j \mathbb{1}\{j \in S_0\} \qquad \beta_{j,S_0^c} := \beta_j \mathbb{1}\{j \notin S_0\}$$

so that $\beta_{S_0} = [\beta_{1,S_0}, ..., \beta_{d,S_0}]^\top$ has zero elements outside the set S_0 and the elements of $\beta_{S_0^c}$ can only be non-zero in the complement of S_0. We can then lower-bound $\|\hat{\beta}\|_1$ using the triangle inequality,

$$\begin{aligned}\|\hat{\beta}\|_1 &= \|\hat{\beta}_{S_0}\|_1 + \|\hat{\beta}_{S_0^c}\|_1 \\ &\geq \|\beta^*_{S_0}\|_1 - \|\hat{\beta}_{S_0} - \beta^*_{S_0}\|_1 + \|\hat{\beta}_{S_0^c}\|_1\,.\end{aligned}$$

Also, we can rewrite

$$\begin{aligned}\|\hat{\beta} - \beta^*\|_1 &= \|\hat{\beta}_{S_0} - \beta^*_{S_0}\|_1 + \|\hat{\beta}_{S_0^c} - \beta^*_{S_0^c}\|_1 \\ &= \|\hat{\beta}_{S_0} - \beta^*_{S_0}\|_1 + \|\hat{\beta}_{S_0^c}\|_1\,.\end{aligned}$$

Then we continue from (4)

$$\begin{aligned}2\mathcal{E}(\hat{\beta}) + 2\lambda_t\|\beta^*_{S_0}\|_1 - 2\lambda_t\|\hat{\beta}_{S_0} - \beta^*_{S_0}\| + 2\lambda_t\|\hat{\beta}_{S_0^c}\|_1 &\leq \lambda_t\|\hat{\beta}_{S_0} - \beta^*_{S_0}\|_1 + \lambda_t\|\hat{\beta}_{S_0^c}\|_1 + 2\lambda_t\|\beta^*\|_1 \\ &= \lambda_t\|\hat{\beta}_{S_0} - \beta^*_{S_0}\|_1 + \lambda_t\|\hat{\beta}_{S_0^c}\|_1 + 2\lambda_t\|\beta^*_{S_0}\|_1\,.\end{aligned}$$

Therefore, we have

$$\begin{aligned}0 \leq 2\mathcal{E}(\hat{\beta}) &\leq 3\lambda_t\|\hat{\beta}_{S_0} - \beta^*_{S_0}\|_1 - \lambda_t\|\hat{\beta}_{S_0^c}\|_1 \\ &= \lambda_t\left(3\|\hat{\beta}_{S_0} - \beta^*_{S_0}\|_1 - \|\hat{\beta}_{S_0^c} - \beta^*_{S_0^c}\|_1\right)\end{aligned} \tag{5}$$

Then the compatibility condition can be applied to the vector $\hat{\beta} - \beta^*$ which gives

$$\|\hat{\beta}_{S_0} - \beta^*_{S_0}\|_1^2 \leq s_0(\hat{\beta} - \beta^*)^\top\hat{\Sigma}(\hat{\beta} - \beta^*)/\phi_t^2\,. \tag{6}$$

From (5), we have

$$2\mathcal{E}(\hat{\beta}) + \lambda_t\|\hat{\beta}_{S_0^c}\|_1 \leq 3\lambda_t\|\hat{\beta}_{S_0} - \beta^*_{S_0}\|_1\,.$$

Therefore, we have

$$\begin{aligned}2\mathcal{E}(\hat{\beta}) + \lambda_t\|\hat{\beta} - \beta^*\|_1 &= 2\mathcal{E}(\hat{\beta}) + \lambda_t\|\hat{\beta}_{S_0^c}\|_1 + \lambda_t\|\hat{\beta}_{S_0} - \beta^*_{S_0}\|_1 \\ &\leq 3\lambda_t\|\hat{\beta}_{S_0} - \beta^*_{S_0}\|_1 + \lambda_t\|\hat{\beta}_{S_0} - \beta^*_{S_0}\|_1 \\ &= 4\lambda_t\|\hat{\beta}_{S_0} - \beta^*_{S_0}\|_1 \\ &\leq 4\lambda_t\sqrt{s_0(\hat{\beta} - \beta^*)^\top\hat{\Sigma}(\hat{\beta} - \beta^*)}/\phi_t \\ &\leq \kappa_0(\hat{\beta} - \beta^*)^\top\hat{\Sigma}(\hat{\beta} - \beta^*) + \frac{4\lambda_t^2 s_0}{\kappa_0\phi_t^2} \\ &\leq 2\mathcal{E}(\hat{\beta}) + \frac{4\lambda^2 s_0}{\kappa_0\phi_t^2}\end{aligned}$$

where the second inequality is from applying the compatibility condition (6) and the third inequality is by using $4uv \leq u^2 + 4v^2$ with $u = \sqrt{\kappa_0(\hat{\beta} - \beta^*)^\top\hat{\Sigma}(\hat{\beta} - \beta^*)}$ and $v = \frac{\lambda_t\sqrt{s_0}}{\phi_t\sqrt{\kappa_0}}$. The last inequality is from Lemma 5. Hence, rearranging gives

$$\|\hat{\beta} - \beta^*\|_1 \leq \frac{4s_0\lambda_t}{\kappa_0\phi_t^2}\,.$$

This completes the proof.

A.2. Proof of Lemma 4

Proof. By the definitions of the negative log-likelihood $\ell_t(\beta)$ and its expectation $\bar{\ell}_t(\beta)$, we can rewrite the empirical process $v_t(\beta)$ as

$$\begin{aligned} v_t(\beta) &= \ell_t(\beta) - \bar{\ell}_t(\beta) \\ &= -\frac{1}{t}\sum_{\tau=1}^{t}\left[Y_\tau X_\tau^\top \beta - m(X_\tau^\top \beta)\right] + \frac{1}{t}\sum_{\tau=1}^{t}\left[\mu(X_\tau^\top \beta^*)X_\tau^\top \beta - m(X_\tau^\top \beta)\right] \\ &= -\frac{1}{t}\sum_{\tau=1}^{t}\left[Y_\tau X_\tau^\top \beta - \mu(X_\tau^\top \beta^*)X_\tau^\top \beta\right] \\ &= -\frac{1}{t}\sum_{\tau=1}^{t}\epsilon_\tau X_\tau^\top \beta \end{aligned}$$

where the last equality uses the definition of ϵ_τ. Then, the empirical process deviation is

$$v_t(\hat{\beta}_t) - v_n(\beta^*) = -\frac{1}{t}\sum_{\tau=1}^{t}\epsilon_\tau X_\tau^\top(\hat{\beta}_t - \beta^*).$$

Applying Hölder's inequality, we have

$$|v_t(\hat{\beta}_t) - v_t(\beta^*)| \le \frac{1}{t}\left\|\sum_{\tau=1}^{t}\epsilon_\tau X_\tau\right\|_\infty \|\hat{\beta}_t - \beta^*\|_1.$$

Then controlling the empirical process reduces to controlling $\frac{1}{t}\left\|\sum_{\tau=1}^{t}\epsilon_\tau X_\tau\right\|_\infty$. Then, using the union bound, it follows that

$$\begin{aligned} \mathbb{P}\left(\frac{1}{t}\left\|\sum_{\tau=1}^{t}\epsilon_\tau X_\tau\right\|_\infty \le \lambda\right) &= 1 - \mathbb{P}\left(\frac{1}{t}\left\|\sum_{\tau=1}^{t}\epsilon_\tau X_\tau\right\|_\infty > \lambda\right) \\ &\ge 1 - \sum_{j=1}^{d}\mathbb{P}\left(\frac{1}{t}\left|\sum_{\tau=1}^{t}\epsilon_\tau X_\tau^{(j)}\right| > \lambda\right) \end{aligned}$$

where $X_\tau^{(j)}$ is the j-th element of X_τ. For each $j \in [d]$, and $\tau \in [t]$, we let $Z_\tau^{(j)} := \epsilon_\tau X_\tau^{(j)}$. Let $\widetilde{\mathcal{F}}_{t-1}$ denote the sigma-field that contains all observed information prior to taking an action in round t, i.e., $\widetilde{\mathcal{F}}_{t-1}$ is generated by random variables of previously chosen actions $\{a_1, ..., a_{t-1}\}$, their features $\{X_1, ..., X_{t-1}\}$, the corresponding rewards $\{Y_1, ..., Y_{t-1}\}$ and the set of feature vectors $\mathcal{X}_t = \{X_{t,1}, ..., X_{t,K}\}$ in round t.

Then, each $\{Z_\tau^{(j)}\}_{\tau=1}^{t}$ for $j \in [d]$ is a martingale difference sequence adapted to the filtration $\widetilde{\mathcal{F}}_1 \subset ... \subset \widetilde{\mathcal{F}}_\tau$ since $\mathbb{E}[\epsilon_\tau X_\tau^{(j)}|\widetilde{\mathcal{F}}_{\tau-1}] = X_\tau^{(j)}\mathbb{E}[\epsilon_\tau|\widetilde{\mathcal{F}}_{\tau-1}] = 0$ for each j. Note that each $X_\tau^{(j)}$ is a bounded random variable with $|X_\tau^{(j)}| \le \|X_\tau\|_\infty \le \|X_\tau\|_2 \le x_{\max}$. Then from the fact that ϵ_τ is σ^2-sub-Gaussian, it follows that $Z_\tau^{(j)}$ is also σ^2-sub-Gaussian. That is,

$$\begin{aligned} \mathbb{E}\left[\exp(\alpha Z_\tau^{(j)}) \mid \widetilde{\mathcal{F}}_{\tau-1}\right] &= \mathbb{E}\left[\exp\left\{\left(\alpha X_\tau^{(j)}\right)\epsilon_\tau\right\} \mid \widetilde{\mathcal{F}}_{\tau-1}\right] \\ &\le \mathbb{E}\left[\exp(\alpha x_{\max}\epsilon_\tau) \mid \widetilde{\mathcal{F}}_{\tau-1}\right] \\ &\le \exp\left(\frac{\alpha^2 x_{\max}^2 \sigma^2}{2}\right) \end{aligned}$$

for any $\alpha \in \mathbb{R}$. Then, using the concentration result in Lemma 14, we have

$$\mathbb{P}\left(\left|\sum_{\tau=1}^{t}\epsilon_\tau X_\tau^{(j)}\right| > t\lambda\right) \le 2\exp\left(-\frac{t^2\lambda^2}{2t\sigma^2 x_{\max}^2}\right) \le 2\exp\left(-\frac{t\lambda^2}{2\sigma^2 x_{\max}^2}\right).$$

So, with $\lambda = \sigma x_{\max}\sqrt{\frac{2[\log(2/\delta)+\log d]}{t}}$, we have

$$\mathbb{P}\left(\frac{1}{t}\left\|\sum_{\tau=1}^{t}\epsilon_\tau X_\tau\right\|_\infty \leq \lambda\right) \geq 1 - 2d\exp\left(\log\frac{\delta}{2} - \log d\right) = 1 - \delta\,.$$

□

Lemma 5. *The excess risk is lower-bounded by*

$$\mathcal{E}(\hat{\beta}_t) \geq \frac{\kappa_0}{2}(\hat{\beta}_t - \beta^*)^\top \hat{\Sigma}(\hat{\beta}_t - \beta^*)\,.$$

Proof. By the definition of the excess risk $\mathcal{E}(\beta)$, we have

$$\begin{aligned}\mathcal{E}(\beta) &= \bar{\ell}_t(\beta) - \bar{\ell}_t(\beta^*)\\ &= -\frac{1}{t}\sum_{\tau=1}^{t}\left[\mu(X_\tau^\top \beta^*)X_\tau^\top \beta - m(X_\tau^\top \beta)\right] + \frac{1}{t}\sum_{\tau=1}^{t}\left[\mu(X_\tau^\top \beta^*)X_\tau^\top \beta^* - m(X_\tau^\top \beta^*)\right]\,.\end{aligned}$$

Since $\dot{m}(\cdot) = \mu(\cdot)$, we have $\nabla_\beta \bar{\ell}_t(\beta^*) = \vec{0}_d$. Hence, the gradient of the excess risk $\nabla_\beta \mathcal{E}(\beta)$ and the Hessian are given as

$$\begin{aligned}\nabla_\beta \mathcal{E}(\beta) &= -\frac{1}{t}\sum_{\tau=1}^{t}\left[\mu(X_\tau^\top \beta^*)X_\tau - \mu(X_\tau^\top \beta)X_\tau\right]\,,\\ H_{\mathcal{E}}(\beta) := \nabla_\beta^2 \mathcal{E}(\beta) &= \frac{1}{t}\sum_{\tau=1}^{t}\dot{\mu}(X_\tau^\top \beta)X_\tau X_\tau^\top\,.\end{aligned}$$

Using the Taylor expansion, with $\bar{\beta} = c\beta^* + (1-c)\hat{\beta}$ for some $c \in (0,1)$

$$\mathcal{E}(\hat{\beta}_t) = \mathcal{E}(\beta^*) + \nabla_\beta \mathcal{E}(\beta^*)^\top(\hat{\beta}_t - \beta^*) + \frac{1}{2}(\hat{\beta}_t - \beta^*)^\top H_{\mathcal{E}}(\bar{\beta})(\hat{\beta}_t - \beta^*)\,. \tag{7}$$

Note that by the definition of β^*, we have $\mathcal{E}(\beta^*) = 0$ and $\nabla_\beta \mathcal{E}(\beta^*) = \nabla_\beta \ell(\beta^*) = \vec{0}_d$. Hence, combining with the definition of the Hessian, we have

$$\begin{aligned}\mathcal{E}(\hat{\beta}_t) &= \frac{1}{2}(\hat{\beta}_t - \beta^*)^\top \left[\frac{1}{t}\sum_{\tau=1}^{t}\dot{\mu}(X_\tau^\top \bar{\beta})X_\tau X_\tau^\top\right](\hat{\beta}_t - \beta^*)\\ &\geq \frac{\kappa_0}{2}(\hat{\beta}_t - \beta^*)^\top \hat{\Sigma}(\hat{\beta}_t - \beta^*)\end{aligned}$$

where the last inequality is from Assumption 2 and $\hat{\Sigma} = \frac{1}{t}\sum_{\tau=1}^{t} X_\tau X_\tau^\top$. ∎

A.3. Proof of Lemma 2

Proof. Consider $\mathcal{X} = \{X_1, X_2\}$. Let the joint density function of x_1, x_2 as $p_{\mathcal{X}}(x_1, x_2)$. Then we have

$$\begin{aligned}\mathbb{E}[\mathbf{X}^\top \mathbf{X}] &= \int (x_1 x_1^\top + x_2 x_2^\top) p_{\mathcal{X}}(x_1, x_2) dx_1, x_2\\ &= \int x_1 x_1^\top \left[\mathbb{1}\left\{(x_1 - x_2)^\top \beta \geq 0\right\} + \mathbb{1}\left\{(x_1 - x_2)^\top \beta \leq 0\right\}\right] p_{\mathcal{X}}(x_1, x_2) dx_1, x_2\\ &\quad + \int x_2 x_2^\top \left[\mathbb{1}\left\{(x_1 - x_2)^\top \beta \geq 0\right\} + \mathbb{1}\left\{(x_1 - x_2)^\top \beta \leq 0\right\}\right] p_{\mathcal{X}}(x_1, x_2) dx_1, x_2\end{aligned}$$

Let's first look at the first integral.

$$\begin{aligned}
&\int x_1 x_1^\top \left[\mathbb{1}\left\{(x_1 - x_2)^\top \beta \geq 0\right\} + \mathbb{1}\left\{(x_1 - x_2)^\top \beta \leq 0\right\}\right] p_{\mathcal{X}}(x_1, x_2) dx_1, x_2 \\
&= \int x_1 x_1^\top \left[\mathbb{1}\left\{(x_1 - x_2)^\top \beta \geq 0\right\} p_{\mathcal{X}}(x_1, x_2) + \mathbb{1}\left\{-(x_1 - x_2)^\top \beta \geq 0\right\} p_{\mathcal{X}}(x_1, x_2)\right] dx_1, x_2 \\
&\preccurlyeq \int x_1 x_1^\top \mathbb{1}\left\{(x_1 - x_2)^\top \beta \geq 0\right\} p_{\mathcal{X}}(x_1, x_2) dx_1, x_2 \\
&\quad + \nu \int x_1 x_1^\top \mathbb{1}\left\{-(x_1 - x_2)^\top \beta \geq 0\right\} p_{\mathcal{X}}(-x_1, -x_2) dx_1, x_2 \\
&= \int x_1 x_1^\top \mathbb{1}\left\{(x_1 - x_2)^\top \beta \geq 0\right\} p_{\mathcal{X}}(x_1, x_2) dx_1, x_2 \\
&\quad + \nu \int x_1 x_1^\top \mathbb{1}\left\{(x_1 - x_2)^\top \beta \geq 0\right\} p_{\mathcal{X}}(x_1, x_2) dx_1, x_2 \\
&= (1+\nu) \int x_1 x_1^\top \mathbb{1}\left\{(x_1 - x_2)^\top \beta \geq 0\right\} p_{\mathcal{X}}(x_1, x_2) dx_1, x_2 \\
&= (1+\nu) \mathbb{E}\left[X_1 X_1^\top \mathbb{1}\{X_1 = \operatorname*{argmax}_{X \in \mathcal{X}} X^\top \beta\}\right]
\end{aligned}$$

where the inequality follows from Assumption 4. Likewise, we can show for the second integral that

$$\begin{aligned}
&\int x_2 x_2^\top \left[\mathbb{1}\left\{(x_1 - x_2)^\top \beta \geq 0\right\} + \mathbb{1}\left\{(x_1 - x_2)^\top \beta \leq 0\right\}\right] p_{\mathcal{X}}(x_1, x_2) dx_1, x_2 \\
&= (1+\nu) \mathbb{E}\left[X_2 X_2^\top \mathbb{1}\{X_1 = \operatorname*{argmax}_{X \in \mathcal{X}} X^\top \beta\}\right].
\end{aligned}$$

Hence,

$$\mathbb{E}[\mathbf{X}^\top \mathbf{X}] = (1+\nu)\left(\mathbb{E}\left[X_1 X_1^\top \mathbb{1}\{X_1 = \operatorname*{argmax}_{X \in \mathcal{X}} X^\top \beta\}\right] + \mathbb{E}\left[X_2 X_2^\top \mathbb{1}\{X_2 = \operatorname*{argmax}_{X \in \mathcal{X}} X^\top \beta\}\right]\right).$$

Therefore, with the fact that $\nu \geq 1$, we have

$$\sum_{i=1}^{2} \mathbb{E}\left[X_i X_i^\top \mathbb{1}\{X_i = \operatorname*{argmax}_{X \in \mathcal{X}} X^\top \beta\}\right] \succcurlyeq \frac{2}{1+\nu} \cdot \frac{1}{2} \mathbb{E}[\mathbf{X}^\top \mathbf{X}] \succcurlyeq \nu^{-1} \Sigma.$$

□

A.4. Bernstein-type Inequality for Adapted Samples

In this section, we derive a Bernstein-type inequality for adapted samples which is shown in Lemma 8. We first define the following function of a random variable X_t which is used throughout this section.

Definition 3. *For all i, j with $1 \leq i \leq j \leq d$, we define $\gamma_t^{ij}(X_t)$ to be a real-value function which take random variable $X_t \in \mathbb{R}^d$ as input:*

$$\gamma_t^{ij}(X_t) := \frac{1}{2x_{\max}^2}\left(X_t^{(i)} X_t^{(j)} - \mathbb{E}[X_t^{(i)} X_t^{(j)} \mid \mathcal{F}_{t-1}]\right) \tag{8}$$

where $X_t^{(i)}$ is the i-th element of X_t.

It is easy to see that $\mathbb{E}\left[\gamma_t^{ij}(X_t) \mid \mathcal{F}_{t-1}\right] = 0$ and $\mathbb{E}\left[|\gamma_t^{ij}(X_t)|^m \mid \mathcal{F}_{t-1}\right] \leq 1$ for all integer $m \geq 2$. While we introduce this specific function $\gamma_t^{ij}(X_t)$ in order to connect to the matrix concentration $\|\Sigma_\tau - \hat{\Sigma}_\tau\|_\infty$, Lemma 7 and Lemma 8 can be applied to any function $\gamma_t^{ij}(X_t)$ that satisfies the zero mean and the bounded m-th moment conditions.

Lemma 6 (Bühlmann & Van De Geer (2011), Lemma 14.1)**.** *Let $Z_t \in \mathbb{R}$ be a random variable with $\mathbb{E}[Z_t \mid \mathcal{F}_{t-1}] = 0$. Then it holds that*

$$\log \mathbb{E}\left[e^{Z_t} \mid \mathcal{F}_{t-1}\right] \leq \mathbb{E}\left[e^{|Z_t|} \mid \mathcal{F}_{t-1}\right] - 1 - \mathbb{E}\left[|Z| \mid \mathcal{F}_{t-1}\right] .$$

Proof. The proof follows directly from the proof of Lemma 14.1 in Bühlmann & Van De Geer (2011), applying their result to a conditional expectation. For any $c > 0$,

$$\begin{aligned}\exp(Z_t - c) - 1 &\leq \frac{\exp(Z_t)}{1+c} - 1 \\ &= \frac{e^{Z_t} - 1 - Z_t + Z_t - c}{1+c} \\ &\leq \frac{e^{|Z_t|} - 1 - |Z_t| + Z_t - c}{1+c} .\end{aligned}$$

Let $c = \mathbb{E}\left[e^{|Z_t|} \mid \mathcal{F}_{t-1}\right] - 1 - \mathbb{E}\left[|Z| \mid \mathcal{F}_{t-1}\right]$. Hence, since $\mathbb{E}[Z_t \mid \mathcal{F}_{t-1}] = 0$,

$$\mathbb{E}\left[\exp(Z_t - c) \mid \mathcal{F}_{t-1}\right] - 1 \leq \frac{\mathbb{E}\left[e^{|Z_t|} \mid \mathcal{F}_{t-1}\right] - 1 - \mathbb{E}\left[|Z_t| \mid \mathcal{F}_{t-1}\right] - c}{1+c} = 0 .$$

□

Lemma 7. *Suppose $\mathbb{E}\big[\gamma_t^{ij}(X_t) \mid \mathcal{F}_{t-1}\big] = 0$ and $\mathbb{E}\big[|\gamma_t^{ij}(X_t)|^m \mid \mathcal{F}_{t-1}\big] \leq m!$ for all integer $m \geq 2$, all $t \geq 1$ and all $1 \leq i \leq j \leq d$. Then, for $L > 1$ we have*

$$\mathbb{E}\left[\exp\left(\frac{1}{L}\sum_{t=1}^{\tau} \gamma_t^{ij}(X_t)\right)\right] \leq \exp\left(\frac{\tau}{L(L-1)}\right) .$$

Proof.

$$\begin{aligned}\mathbb{E}\left[\exp\left(\frac{1}{L}\sum_{t=1}^{\tau} \gamma_t^{ij}(X_t)\right)\right] &= \mathbb{E}\left[\mathbb{E}\left[\exp\left(\frac{1}{L}\sum_{t=1}^{\tau} \gamma_t^{ij}(X_t)\right) \mid \mathcal{F}_{\tau-1}\right]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\exp\left(\frac{\gamma_\tau^{ij}(X_\tau)}{L}\right) \mid \mathcal{F}_{\tau-1}\right]\exp\left(\frac{1}{L}\sum_{t=1}^{\tau-1} \gamma_t^{ij}(X_t)\right)\right] \\ &\leq e^{\frac{1}{L(L-1)}}\,\mathbb{E}\left[\exp\left(\frac{1}{L}\sum_{t=1}^{\tau-1} \gamma_t^{ij}(X_t)\right)\right]\end{aligned}$$

where the inequality is from Lemma 6 and noting that

$$\begin{aligned}\log \mathbb{E}\left[\exp\left(\frac{\gamma_\tau^{ij}(X_\tau)}{L}\right) \mid \mathcal{F}_{\tau-1}\right] &\leq \mathbb{E}\left[e^{\left|\gamma_\tau^{ij}(X_\tau)\right|/\tau} - 1 - \frac{\left|\gamma_\tau^{ij}(X_\tau)\right|}{L} \mid \mathcal{F}_{\tau-1}\right] \\ &= \mathbb{E}\left[\sum_{m=2}^{\infty} \frac{\left|\gamma_\tau^{ij}(X_\tau)\right|^m}{L^m m!} \mid \mathcal{F}_{\tau-1}\right] \\ &= \sum_{m=2}^{\infty} \frac{\mathbb{E}\left[\left|\gamma_\tau^{ij}(X_\tau)\right|^m \mid \mathcal{F}_{\tau-1}\right]}{L^m m!} \\ &\leq \frac{1}{L(L-1)} .\end{aligned}$$

Then, repeatedly applying this to the rest of the sum $\frac{1}{L}\sum_{t=1}^{\tau-1} \gamma_t^{ij}(X_t)$, we have

$$\mathbb{E}\left[\exp\left(\frac{1}{L}\sum_{t=1}^{\tau} \gamma_t^{ij}(X_t)\right)\right] \leq \exp\left(\frac{\tau}{L(L-1)}\right) .$$

□

Lemma 8 (Bernstein-type inequality for adapted samples)**.** *Suppose* $\mathbb{E}\big[\gamma_t^{ij}(X_t) \mid \mathcal{F}_{t-1}\big] = 0$ *and* $\mathbb{E}\big[|\gamma_t^{ij}(X_t)|^m \mid \mathcal{F}_{t-1}\big] \leq m!$ *for all integer* $m \geq 2$*, all* $t \geq 1$ *and all* $1 \leq i \leq j \leq d$*. Then for all* $w > 0$*, we have*

$$\mathbb{P}\left(\max_{1\leq i\leq j\leq d}\left|\frac{1}{\tau}\sum_{t=1}^{\tau}\gamma_t^{ij}(X_t)\right| \geq w + \sqrt{2w} + \sqrt{\frac{4\log(2d^2)}{\tau}} + \frac{2\log(2d^2)}{\tau}\right) \leq \exp\left(-\frac{\tau w}{2}\right) .$$

Proof. Using the Chernoff bound and Lemma 7, for any $L > 1$ we have

$$\begin{aligned}
\mathbb{P}\left(\sum_{t=1}^{\tau}\gamma_t^{ij}(X_t) \geq a\right) &= \mathbb{P}\left(\exp\left(\frac{1}{L}\sum_{t=1}^{\tau}\gamma_t^{ij}(X_t)\right) \geq \exp\left(\frac{a}{L}\right)\right) \\
&\leq \frac{\mathbb{E}\left[\exp\left(\frac{1}{L}\sum_{t=1}^{\tau}\gamma_t^{ij}(X_t)\right)\right]}{\exp\left(\frac{a}{L}\right)} \\
&\leq \exp\left(-\frac{a}{L}\right)\exp\left(\frac{\tau}{L(L-1)}\right) \\
&= \exp\left(-\frac{a}{L} + \frac{\tau}{L(L-1)}\right) .
\end{aligned}$$

Here, $L = \frac{\tau + a + \sqrt{\tau^2 + \tau a}}{a}$ minimizes the right hand side above for $L > 1$. Therefore,

$$\begin{aligned}
\mathbb{P}\left(\sum_{t=1}^{\tau}\gamma_t^{ij}(X_t) \geq a\right) &\leq \exp\left\{-\frac{a^2}{\tau + a + \sqrt{\tau^2 + \tau a}} + \frac{\tau a^2}{(\tau + a + \sqrt{\tau^2 + \tau a})(\tau + \sqrt{\tau^2 + \tau a})}\right\} \\
&= \exp\left\{-\left(\frac{\sqrt{1 + a/\tau}}{1 + \sqrt{1 + a/\tau}}\right)\frac{a^2}{\tau + a + \sqrt{\tau^2 + \tau a}}\right\} \\
&\leq \exp\left\{-\frac{a^2}{2\left(\tau + a + \sqrt{\tau^2 + \tau a}\right)}\right\} \\
&\leq \exp\left\{-\frac{a^2}{2\left(\tau + a + \sqrt{\tau^2 + 2\tau a}\right)}\right\} .
\end{aligned}$$

Choosing $a = \tau\left(w + \sqrt{2w}\right)$ gives

$$\mathbb{P}\left(\frac{1}{\tau}\sum_{t=1}^{\tau}\gamma_t^{ij}(X_t) \geq w + \sqrt{2w}\right) \leq \exp\left(-\frac{\tau w}{2}\right) . \tag{9}$$

Then for the maximal inequality, we first apply the union bound to (9).

$$\begin{aligned}
\mathbb{P}\left(\max_{1\leq i\leq j\leq d}\frac{1}{\tau}\left|\sum_{t=1}^{\tau}\gamma_t^{ij}(X_t)\right| \geq w + \sqrt{2w}\right) &\leq \sum_{1\leq i\leq j\leq d} 2\mathbb{P}\left(\frac{1}{\tau}\sum_{t=1}^{\tau}\gamma_t^{ij}(X_t) \geq w + \sqrt{2w}\right) \\
&\leq 2d^2\exp\left(-\frac{\tau w}{2}\right) \\
&= \exp\left(-\frac{\tau w}{2} + \log(2d^2)\right) .
\end{aligned}$$

Then,

$$
\begin{aligned}
&\mathbb{P}\left(\max_{1\le i\le j\le d}\frac{1}{\tau}\left|\sum_{t=1}^{\tau}\gamma_t^{ij}(X_t)\right| \ge w+\sqrt{2w}+\sqrt{\frac{4\log(2d^2)}{\tau}}+\frac{2\log(2d^2)}{\tau}\right)\\
&\le \mathbb{P}\left(\max_{1\le i\le j\le d}\frac{1}{\tau}\left|\sum_{t=1}^{\tau}\gamma_t^{ij}(X_t)\right| \ge \left(w+\frac{2\log(2d^2)}{\tau}\right)+\sqrt{2\left(w+\frac{2\log(2d^2)}{\tau}\right)}\right)\\
&= \mathbb{P}\left(\max_{1\le i\le j\le d}\frac{1}{\tau}\left|\sum_{t=1}^{\tau}\gamma_t^{ij}(X_t)\right| \ge w'+\sqrt{2w'}\right)\\
&\le \exp\left(-\frac{\tau w'}{2}+\log(2d^2)\right)\\
&= \exp\left(-\frac{\tau w}{2}\right)
\end{aligned}
$$

where $w' = w + \frac{2\log(2d^2)}{\tau}$. □

A.5. Proof of Lemma 3

Proof. Notice the difference between the unconditional theoretical Gram matrix Σ and its adapted version $\mathbb{E}[X_t X_t^\top|\mathcal{F}_{t-1}]$ which is a conditional covariance matrix conditioned on the history $\mathcal{F}_{t-1}$. Recall that from Algorithm 1, in each round t we choose X_t given the history $\mathcal{F}_{t-1}$. More precisely, we compute β_t based on $\mathcal{F}_{t-1}$ and choose X_t which maximizes the product $X_t^\top\hat{\beta}_t$, i.e., $\operatorname{argmax}_{X\in\mathcal{X}_t} X^\top\hat{\beta}_t$ where $\mathcal{X}_t = \{X_{t,1}, X_{t,2}\}$. Hence, we can write $\mathbb{E}[X_t X_t^\top|\mathcal{F}_{t-1}]$ as the following:

$$
\mathbb{E}[X_t X_t^\top|\mathcal{F}_{t-1}] = \sum_{i=1}^{2}\mathbb{E}_{\mathcal{X}_t}\left[X_{t,i}X_{t,i}^\top \mathbb{1}\{X_{ti} = \operatorname*{argmax}_{X\in\mathcal{X}_t} X^\top\hat{\beta}_t\} \mid \hat{\beta}_t\right].
$$

From Lemma 2, it follows that

$$
\mathbb{E}[X_t X_t^\top|\mathcal{F}_{t-1}] \succcurlyeq \nu^{-1}\Sigma\,.
$$

Now, taking an average over t gives,

$$
\Sigma_\tau = \frac{1}{\tau}\sum_{t=1}^{\tau}\mathbb{E}[X_t X_t^\top|\mathcal{F}_{t-1}] \succcurlyeq \nu^{-1}\Sigma\,.
$$

Then, we define $\widetilde{\beta}$ corresponding to compatibility constant $\phi^2(\Sigma_\tau, S_0)$, that is,

$$
\widetilde{\beta} := \operatorname*{argmin}_{\beta}\left\{\frac{\beta^\top\Sigma_\tau\beta}{\|\beta_{S_0}\|_1^2} : \|\beta_{S_0^c}\|_1 \le 3\|\beta_{S_0}\|_1 \ne 0\right\}.
$$

Therefore, it follows that

$$
\frac{\widetilde{\beta}^\top\Sigma_\tau\widetilde{\beta}}{\|\widetilde{\beta}_{S_0}\|_1^2} \ge \frac{\widetilde{\beta}^\top\Sigma\widetilde{\beta}}{\nu\|\widetilde{\beta}_{S_0}\|_1^2} \ge \frac{\phi_0^2}{\nu} \tag{10}
$$

where the second inequality is by the compatibility condition on Σ. Thus, Σ_τ satisfies the compatibility condition with compatibility constant $\phi^2(\Sigma_\tau, S_0) = \frac{\phi_0^2}{\nu}$.

Now, noting that $\frac{1}{2x_{\max}^2}\|\Sigma_\tau - \hat{\Sigma}_\tau\|_\infty = \max_{1\le i\le j\le d}\frac{1}{\tau}\left|\sum_{t=1}^{\tau}\gamma_t^{ij}(X_t)\right|$ for $\gamma_t^{ij}(\cdot)$ defined in (8), we can use a Bernstein-type inequality for adapted samples in Lemma 8 to get

$$
\mathbb{P}\left(\frac{\|\Sigma_\tau - \hat{\Sigma}_\tau\|_\infty}{2x_{\max}^2} \ge w+\sqrt{2w}+\sqrt{\frac{4\log(2d^2)}{\tau}}+\frac{2\log(2d^2)}{\tau}\right) \le \exp\left(-\frac{\tau w}{2}\right).
$$

For $\tau \geq \frac{2\log(2d^2)}{C_0(s_0)^2}$ where $C_0(s_0) = \min\left(\frac{1}{2}, \frac{\phi_0^2}{256 s_0 \nu x_{\max}^2}\right)$, letting $w = C_0(s_0)^2$ gives

$$\begin{aligned} w + \sqrt{2w} + \sqrt{\frac{4\log(2d^2)}{\tau}} + \frac{2\log(2d^2)}{\tau} &\leq 2\left(C_0(s_0)^2 + \sqrt{2}C_0(s_0)\right) \\ &\leq 4C_0(s_0) \\ &\leq \frac{\phi_0^2}{64 s_0 \nu x_{\max}^2} \\ &= \frac{\phi^2(\Sigma_\tau, S_0)}{64 s_0 x_{\max}^2} \,. \end{aligned}$$

Hence,

$$\begin{aligned} \mathbb{P}\left(\frac{\|\Sigma_\tau - \hat{\Sigma}_\tau\|_\infty}{2x_{\max}^2} \geq \frac{\phi^2(\Sigma_\tau, S_0)}{64 s_0 x_{\max}^2}\right) &\leq \mathbb{P}\left(\frac{\|\Sigma_\tau - \hat{\Sigma}_\tau\|_\infty}{2x_{\max}^2} \geq w + \sqrt{2w} + \sqrt{\frac{4\log(2d^2)}{\tau}} + \frac{2\log(2d^2)}{\tau}\right) \\ &\leq \exp\left(-\frac{\tau w}{2}\right) \\ &= \exp\left(-\frac{\tau C_0(s_0)^2}{2}\right) \,. \end{aligned}$$

□

Corollary 2. *For $t \geq \frac{2\log(2d^2)}{C_0(s_0)^2}$ where $C_0(s_0) = \min\left(\frac{1}{2}, \frac{\phi_0^2}{256 s_0 \nu x_{\max}^2}\right)$, the empirical Gram matrix $\hat{\Sigma}_t$ satisfies the compatibility condition with compatibility constant $\phi_t \geq \frac{\phi_0^2}{2\nu} > 0$ with probability at least $1 - \exp\left\{-tC_0(s_0)^2/2\right\}$.*

Proof. We can use Corollary 1 (Bühlmann & Van De Geer (2011), Corollary 6.8) to show that the empirical Gram matrix $\hat{\Sigma}_\tau$ satisfies the compatibility condition as long as Σ_τ satisfies the compatibility condition. From (10), we know Σ_τ satisfies the compatibility condition with compatibility constant $\frac{\phi_0^2}{\nu}$. Then, combining Lemma 3 and Corollary 1, it follows that given $\|\Sigma_t - \hat{\Sigma}_t\|_\infty \leq \frac{\phi_0^2}{32 s_0 \nu}$ for $t \geq \lceil T_0 \rceil$, we have

$$\phi^2(\hat{\Sigma}_t, S_0) \geq \frac{\phi^2(\Sigma_t, S_0)}{2} \geq \frac{\phi_0^2}{2\nu} > 0 \,.$$

That is, $\hat{\Sigma}_\tau$ satisfies the compatibility condition with compatibility constant which is at least $\frac{\phi_0^2}{2\nu} > 0$. □

B. Proof of Theorem 1

Proof. First, let $T_0 := \frac{2\log(2d^2)}{C_0(s_0)^2}$ where $C_0(s_0) = \min\left(\frac{1}{2}, \frac{\phi_0^2}{256 s_0 \nu x_{\max}^2}\right)$. Also, we define the high probability event $\mathcal{E}_t$:

$$\mathcal{E}_t := \left\{\|\Sigma_t - \hat{\Sigma}_t\|_\infty \geq \frac{\phi_0^2}{32 s_0 \nu}\right\} \,.$$

Hence, on this event $\mathcal{E}_t$, if $t \geq T_0$, then from Corollary 2 we have $\phi_t^2 \geq \frac{\phi_0^2}{2\nu}$, i.e., the compatibility condition holds in round t. Slightly overloading the subscript for brevity, let $X_t := X_{t,a_t}$ be a feature of the arm chosen in round t and $X_{a_t^*} := X_{t,a_t^*}$ be the feature of the optimal arm in round t. First, we look at the (non-expected) immediate regret $\text{Reg}(t)$ with $\mathcal{R}(t) = \mathbb{E}[\text{Reg}(t)]$ in round t. Notice that by Assumptions 1 and 2 and by the mean value theorem, $\text{Reg}(t)$ is bounded by

$$\text{Reg}(t) \leq \kappa_1\left(X_{a_t^*}^\top \beta^* - X_t^\top \beta^*\right) \leq \kappa_1 \|X_{a_t^*} - X_t\|_2 \|\beta^*\|_2 \leq 2\kappa_1 x_{\max} b$$

Then we can decompose the immediate regret as follows.

$$\begin{aligned} \text{Reg}(t) &= \text{Reg}(t)\mathbb{1}(t \leq T_0) + \text{Reg}(t)\mathbb{1}(t > T_0, \mathcal{E}_t) + \text{Reg}(t)\mathbb{1}(t > T_0, \mathcal{E}_t^c) \\ &\leq 2\kappa_1 x_{\max} b\mathbb{1}(t \leq T_0) + \text{Reg}(t)\mathbb{1}(t > T_0, \mathcal{E}_t) + 2\kappa_1 x_{\max} b\mathbb{1}(t > T_0, \mathcal{E}_t^c) \\ &= 2\kappa_1 x_{\max} b\mathbb{1}(t \leq T_0) + \text{Reg}(t)\mathbb{1}\left(\mu(X_t^\top \hat{\beta}_t) \geq \mu(X_{a_t^*}^\top \hat{\beta}_t), t > T_0, \mathcal{E}_t\right) \\ &\quad + 2\kappa_1 x_{\max} b\mathbb{1}(t > T_0, \mathcal{E}_t^c) \end{aligned}$$

where the last equality follows from the optimality of X_t with respect to parameter $\hat{\beta}_t$, i.e., $X_t = \text{argmax}_{X \in \mathcal{X}_t} \mu(X^\top \hat{\beta}_t)$. For the second term, we have

$$\begin{aligned}
\mathbb{P}\left(\mu(X_t^\top \hat{\beta}_t) \geq \mu(X_{a_t^*}^\top \hat{\beta}_t)\right) &= \mathbb{P}\left(\mu(X_t^\top \hat{\beta}_t) - \mu(X_{a_t^*}^\top \hat{\beta}_t) + \text{Reg}(t) \geq \text{Reg}(t)\right) \\
&= \mathbb{P}\left((\mu(X_t^\top \hat{\beta}_t) - \mu(X_t^\top \beta^*)) - (\mu(X_{a_t^*}^\top \hat{\beta}_t) - \mu(X_{a_t^*}^\top \beta^*)) \geq \text{Reg}(t)\right) \\
&\leq \mathbb{P}\left(|\mu(X_t^\top \hat{\beta}_t) - \mu(X_t^\top \beta^*)| + |\mu(X_{a_t^*}^\top \hat{\beta}_t) - \mu(X_{a_t^*}^\top \beta^*)| \geq \text{Reg}(t)\right) \\
&\leq \mathbb{P}\left(\kappa_1 \|\hat{\beta}_t - \beta^*\|_1 \|X_t\|_\infty + \kappa_1 \|\hat{\beta}_t - \beta^*\|_1 \|X_{a_t^*}\|_\infty \geq \text{Reg}(t)\right) \\
&\leq \mathbb{P}\left(2\kappa_1 \|\hat{\beta}_t - \beta^*\|_1 \geq \text{Reg}(t)\right)
\end{aligned}$$

where the last inequality is from the fact that each $X_{t,i}$ is bounded. For an arbitrary constant $g_t > 0$, we continue with expected regret $\mathcal{R}(t) = \mathbb{E}[\text{Reg}(t)]$ for $t > T_0$.

$$\begin{aligned}
\mathcal{R}(t) &\leq \mathbb{E}\left[\text{Reg}(t)\mathbb{1}\left(2\kappa_1 \|\hat{\beta}_t - \beta^*\|_1 \geq \text{Reg}(t), \mathcal{E}_t\right)\right] + 2\kappa_1 x_{\max} b \mathbb{P}(\mathcal{E}_t^c) \\
&= \mathbb{E}\left[\text{Reg}(t)\mathbb{1}\left(2\kappa_1 \|\hat{\beta}_t - \beta^*\|_1 \geq \text{Reg}(t), \text{Reg}(t) \leq \kappa_1 g_t, \mathcal{E}_t\right)\right] \\
&\quad + \mathbb{E}\left[\text{Reg}(t)\mathbb{1}\left(2\kappa_1 \|\hat{\beta}_t - \beta^*\|_1 \geq \text{Reg}(t), \text{Reg}(t) > \kappa_1 g_t, \mathcal{E}_t,\right)\right] + 2\kappa_1 x_{\max} b \mathbb{P}(\mathcal{E}_t^c) \\
&\leq \kappa_1 g_t + \kappa_1 \mathbb{P}\left(2\|\hat{\beta}_t - \beta^*\|_1 \geq g_t, \mathcal{E}_t\right) + 2\kappa_1 x_{\max} b \mathbb{P}(\mathcal{E}_t^c) \,.
\end{aligned}$$

Summing over all rounds after the initial T_0 rounds, we have

$$\sum_{t=\lceil T_0 \rceil}^{T} \mathcal{R}(t) \leq \kappa_1 \underbrace{\sum_{t=\lceil T_0 \rceil}^{T} g_t}_{(a)} + \underbrace{\kappa_1 \sum_{t=\lceil T_0 \rceil}^{T} \mathbb{P}\left(2\|\hat{\beta}_t - \beta^*\|_1 \geq g_t, \mathcal{E}_t\right)}_{(b)} + \underbrace{2\kappa_1 x_{\max} b \sum_{t=\lceil T_0 \rceil}^{T} \mathbb{P}(\mathcal{E}_t^c)}_{(c)} \,. \tag{11}$$

We first bound the term (b) in (11). We choose $g_t := \frac{2s_0\lambda_t}{\kappa_0\phi_t^2} = \frac{4\sigma x_{\max} s_0}{\kappa_0 \phi_t^2}\sqrt{\frac{4\log t + 2\log d}{t}}$. Then using Lemma 1, we have

$$\mathbb{P}\left(2\|\hat{\beta}_t - \beta^*\|_1 \geq g_t, \mathcal{E}_t\right) \leq \frac{2}{t^2} \,.$$

for all $t \geq T_0$. Therefore, it follows that

$$\sum_{t=\lceil T_0 \rceil}^{T} \mathbb{P}\left(2\|\hat{\beta}_t - \beta^*\|_1 \geq g_t, \mathcal{E}_t\right) \leq \sum_{t=\lceil T_0 \rceil}^{T} \frac{2}{t^2} \leq \sum_{t=1}^{\infty} \frac{2}{t^2} \leq \frac{\pi^2}{3} < 4 \,.$$

For the term (a) in (11), we have $\phi_t^2 \geq \frac{\phi_0^2}{2\nu}$ provided that event $\mathcal{E}_t$ holds. Hence, we have

$$\begin{aligned}
\sum_{t=\lceil T_0 \rceil}^{T} g_t &= \sum_{t=\lceil T_0 \rceil}^{T} \frac{4\sigma x_{\max} s_0}{\kappa_0 \phi_t^2}\sqrt{\frac{4\log t + 2\log d}{t}} \\
&\leq \sum_{t=\lceil T_0 \rceil}^{T} \frac{8\nu\sigma x_{\max} s_0}{\kappa_0 \phi_0^2}\sqrt{\frac{4\log t + 2\log d}{t}} \\
&\leq \frac{8\nu\sigma x_{\max} s_0 \sqrt{4\log T + 2\log d}}{\kappa_0 \phi_0^2} \sum_{t=\lceil T_0 \rceil}^{T} \frac{1}{\sqrt{t}} \\
&\leq \frac{8\nu\sigma x_{\max} s_0 \sqrt{4\log T + 2\log d}}{\kappa_0 \phi_0^2} \sum_{t=1}^{T} \frac{1}{\sqrt{t}} \\
&\leq \frac{16\nu\sigma x_{\max} s_0 \sqrt{4\log T + 2\log d}}{\kappa_0 \phi_0^2}\sqrt{T}
\end{aligned}$$

where the last inequality is from the fact that $\sum_{t=1}^{T} \frac{1}{\sqrt{t}} \leq \int_{t=0}^{T} \frac{1}{\sqrt{t}} = 2\sqrt{T}$.

Finally, for the term (c) in (11), we have from Lemma 3:

$$\begin{aligned}
\sum_{t=\lceil T_0 \rceil}^{T} \mathbb{P}(\mathcal{E}_t^c) &\leq \sum_{t=\lceil T_0 \rceil}^{T} \mathbb{P}\left(\|\Sigma_t - \hat{\Sigma}_t\|_\infty \geq \frac{\phi_0^2}{32 s_0 \nu} \right) \\
&\leq \sum_{t=\lceil T_0 \rceil}^{T} \exp\left(-\frac{t C_0(s_0)^2}{2} \right) \\
&\leq \sum_{t=1}^{\infty} \exp\left(-\frac{t C_0(s_0)^2}{2} \right) \\
&\leq \frac{2}{C_0(s_0)^2} .
\end{aligned}$$

□

B.1. Regret under the Restricted Eigenvalue Condition

In our analysis so far, we have presented the main results under the compatibility condition in order to be consistent with previous results in the sparse bandit literature. In this section, we present the regret bound for SA LASSO BANDIT under the restricted eigenvalue (RE) condition and briefly discuss its implication in terms of potentially matching lower bounds. Similar to the analysis under the compatibility condition, we assume that the RE condition is satisfied only for the theoretical Gram matrix $\Sigma = \frac{1}{K}\mathbb{E}[\mathbf{X}^\top \mathbf{X}]$.

Assumption 6 (RE condition)**.** *For active set S_0 and Σ, there exists restricted eigenvalue $\phi_1 > 0$ such that $\phi_1^2 \|\beta\|_2^2 \leq \beta^\top \Sigma \beta$ for all $\beta \in \mathbb{C}(S_0)$ defined in (2).*

The RE condition is very similar to the compatibility condition in Assumption 3 but uses the ℓ_2 norm instead of the ℓ_1 norm. Based on this condition, we can show the following regret bound.

Theorem 3 (Regret bound under RE condition)**.** *Suppose $K = 2$ and Assumptions 1, 2, 4, and 6 hold. Then the expected cumulative regret of the* SA LASSO BANDIT *policy is $\mathcal{O}\big(\sqrt{s_0 T \log(dT)}\big)$.*

Theorem 3 establishes $\mathcal{O}\big(\sqrt{s_0 T \log(dT)}\big)$ regret without any prior knowledge on s_0. The regret upper-bound based on the RE condition still enjoys logarithmic dependence on d and furthermore sub-linear dependence on s_0. Compared to Theorem 1, the regret bound in Theorem 3 is smaller by $\sqrt{s_0}$ factor, which is again consistent with the offline Lasso results under the RE condition (Theorem 7.19 in Wainwright 2019). The difference in the regret bounds in Theorem 1 and Theorem 3 is due to the RE condition being slightly stronger than the compatibility condition.

The RE condition is more directly analogous (as compared to the compatibility condition) to the standard positive-definiteness assumption for covariance matrices in GLM bandits (Li et al., 2017). That is, the RE condition is equivalent to positive-definite covariance when $s_0 = d$, i.e., non-sparse settings. Li et al. (2017) showed $\mathcal{O}\big((\log T)^{3/2}\sqrt{dT \log K}\big)$ regret bound of for GLM bandits, which matches the $\Omega(\sqrt{dT})$ minimax lower bound established (Chu et al., 2011) for linear bandits with finite arms, up to logarithmic factors. Therefore, in sparse settings, we conjecture that $\mathcal{O}\big(\sqrt{s_0 T \log(dT)}\big)$ regret is *best possible* up to logarithmic factors under the RE condition (and so is $\mathcal{O}\big(s_0\sqrt{T \log(dT)}\big)$ regret under the compatibility condition). While we present these conjectures, we do not claim our results are minimax.

C. Proof of Theorem 3

The proof follows similar arguments as the proof of Theorem 1. The key difference is that the RE condition involves ℓ_2 norm and therefore the analysis requires the Lasso oracle inequality of the GLM in ℓ_2 norm, which we provide as an extension of Lemma 1.

Corollary 3. *Assume that the RE condition holds for $\hat{\Sigma}_t$ with active set S_0 and restricted eigenvalue ϕ_t. For some $\delta \in (0,1)$, let the regularization parameter λ_t be*

$$\lambda_t := 2\sigma x_{\max}\sqrt{\frac{2[\log(2/\delta)+\log d]}{t}}\,.$$

Then with probability at least $1-\delta$, we have

$$\|\hat{\beta}_t - \beta^*\|_2 \leq \frac{3\sqrt{s_0}\lambda_t}{\kappa_0\phi_t^2}\,.$$

Proof. Continuing from (5) in Lemma 1, the RE condition can be applied to the vector $\hat{\beta}-\beta^*$ which gives

$$\|\hat{\beta}-\beta^*\|_2^2 \leq \frac{(\hat{\beta}-\beta^*)^\top\hat{\Sigma}_t(\hat{\beta}-\beta^*)}{\phi_t^2}\,. \tag{12}$$

Again from (5), we can use the margin condition in Lemma 5

$$\begin{aligned}
3\lambda_t\|\hat{\beta}_{S_0}-\beta^*_{S_0}\|_1 &\geq 2\mathcal{E}(\hat{\beta}_n)\\
&\geq \kappa_0(\hat{\beta}-\beta^*)^\top\hat{\Sigma}_t(\hat{\beta}-\beta^*)\\
&\geq \kappa_0\phi_t^2\|\hat{\beta}-\beta^*\|_2^2
\end{aligned}$$

where the last inequality is from (12) applying the RE condition. Then, it follows that

$$\begin{aligned}
\kappa_0\phi_t^2\|\hat{\beta}-\beta^*\|_2^2 &\leq 3\lambda_t\|\hat{\beta}_{S_0}-\beta^*_{S_0}\|_1\\
&\leq 3\lambda_t\sqrt{s_0}\|\hat{\beta}_{S_0}-\beta^*_{S_0}\|_2\\
&\leq 3\lambda_t\sqrt{s_0}\|\hat{\beta}-\beta^*\|_2\,.
\end{aligned}$$

Hence, dividing the both sides by $\|\hat{\beta}-\beta^*\|_2$ and rearranging gives

$$\|\hat{\beta}-\beta^*\|_2 \leq \frac{3\sqrt{s_0}\lambda_t}{\kappa_0\phi_t^2}\,.$$

This complete the proof. □

C.1. Ensuring the RE Condition for the Empirical Gram Matrix

To distinguish from the compatibility constant, we introduce the definition of a generic restricted eigenvalue of matrix M over active set S_0.

Definition 4. *The restricted eigenvalue of M over S_0 is*

$$\phi^2_{RE}(M,S_0) := \min_{\beta}\left\{\frac{\beta^\top M\beta}{\|\beta\|_2^2} : \|\beta_{S_0^c}\|_1 \leq 3\|\beta_{S_0}\|_1 \neq 0\right\}.$$

Note that Assumption 6 only provides the RE condition for the theoretical Gram matrix Σ. Then, we follow the same arguments as in the analysis under the compatibility condition to show that $\phi^2_{\mathrm{RE}}(\Sigma_t,S_0) \geq \frac{\phi^2_{\mathrm{RE}}(\Sigma,S_0)}{\nu} > 0$, i.e., Σ_t satisfies the RE condition. Then using Lemma 3, we can show that $\hat{\Sigma}_t$ concentrates to Σ_t with high probability. The following lemma (similar to Corollary 1) ensures the RE condition of $\hat{\Sigma}_t$ conditioned on the matrix concentration of the empirical Gram matrix $\hat{\Sigma}_t$.

Lemma 9. *Suppose that the RE condition holds for Σ_0 and the index set S with cardinality $s=|S|$, with restricted eigenvalue $\phi^2_{RE}(\Sigma_0,S)>0$, and that $\|\Sigma_1-\Sigma_0\|_\infty \leq \Delta$, where $32s\Delta \leq \phi^2_{RE}(\Sigma_0,S)$. Then, for the set S, the RE condition holds as well for Σ_1, with $\phi^2_{RE}(\Sigma_1,S) \geq \phi^2_{RE}(\Sigma_0,S)/2$.*

Proof. The proof is an adaptation of Lemma 6.17 in (Bühlmann & Van De Geer, 2011) to the RE condition.

$$\begin{aligned}\left|\beta^\top \Sigma_1 \beta - \beta^\top \Sigma_0 \beta\right| &= \left|\beta^\top (\Sigma_1 - \Sigma_0)\beta\right| \\ &\le \|\Sigma_1 - \Sigma_0\|_\infty \|\beta\|_1^2 \\ &\le \Delta \|\beta\|_1^2\end{aligned}$$

For β such that $\|\beta_{S^c}\| \le 3\|\beta_S\|$, we have the RE condition satisfied for Σ_0. Hence, we have

$$\|\beta\|_1 \le 4\|\beta_S\|_1 \le 4\sqrt{s}\|\beta_S\|_2 \le 4\sqrt{s}\|\beta\|_2 \le \frac{4\sqrt{s_0 \beta^\top \Sigma_0 \beta}}{\phi_{\text{RE}}(\Sigma_0, S)}\,.$$

Therefore, it follows that

$$\left|\beta^\top \Sigma_1 \beta - \beta^\top \Sigma_0 \beta\right| \le \frac{16 s \Delta \beta^\top \Sigma_0 \beta}{\phi_{\text{RE}}^2(\Sigma_0, S)}\,.$$

Since $\beta^\top \Sigma_0 \beta > 0$, dividing the both sides by $\beta^\top \Sigma_0 \beta$ gives

$$\left|\frac{\beta^\top \Sigma_1 \beta}{\beta^\top \Sigma_0 \beta} - 1\right| \le \frac{16 s \Delta}{\phi_{\text{RE}}^2(\Sigma_0, S)}$$

Now, since $32 s \Delta \le \phi_{\text{RE}}^2(\Sigma_0, S)$, it follows that

$$\frac{1}{2} \cdot \frac{\beta^\top \Sigma_0 \beta}{\|\beta\|_2^2} \le \frac{\beta^\top \Sigma_1 \beta}{\|\beta\|_2^2} \le \frac{3}{2} \cdot \frac{\beta^\top \Sigma_0 \beta}{\|\beta\|_2^2}\,.$$

Hence,

$$\phi_{\text{RE}}^2(\Sigma_1, S) \ge \frac{\phi_{\text{RE}}^2(\Sigma_0, S)}{2}\,.$$

□

C.2. Proof of Theorem 3

Proof. The proof of Theorem 3 follows the similar arguments as the proof of Theorem 1. The only difference is that we use ℓ_2 error bound $\|\hat\beta_t - \beta^*\|_2$ instead of $\|\hat\beta_t - \beta^*\|_1$. First, note that

$$\begin{aligned}\mathbb{P}\left(\mu(X_t^\top \hat\beta_t) \ge \mu(X_{a_t^*}^\top \hat\beta_t)\right) &\le \mathbb{P}\left(|\mu(X_t^\top \hat\beta_t) - \mu(X_t^\top \beta^*)| + |\mu(X_{a_t^*}^\top \hat\beta_t) - \mu(X_{a_t^*}^\top \beta^*)| \ge \text{Reg}(t)\right) \\ &\le \mathbb{P}\left(\kappa_1 \|\hat\beta_t - \beta^*\|_2 \|X_t\|_2 + \kappa_1 \|\hat\beta_t - \beta^*\|_2 \|X_t^*\|_2 \ge \text{Reg}(t)\right) \\ &\le \mathbb{P}\left(2\kappa_1 \|\hat\beta_t - \beta^*\|_2 \ge \text{Reg}(t)\right).\end{aligned}$$

For an arbitrary constant $g_t > 0$, we continue with expected regret $\mathbb{E}[\text{Reg}(t)]$ for $t > T_0$.

$$\mathcal{R}(t) \le \kappa_1 g_t + \kappa_1 \mathbb{P}\left(2\|\hat\beta_t - \beta^*\|_2 \ge g_t, \mathcal{E}_t\right) + 2\kappa_1 x_{\max} b \mathbb{P}(\mathcal{E}_t^c)\,.$$

Hence, the cumulative regret is bounded by

$$\sum_{t=1}^T \mathcal{R}(t) \le 2\kappa_1 x_{\max} b T_0 + \kappa_1 \sum_{t=\lceil T_0 \rceil}^T g_t + \kappa_1 \sum_{t=\lceil T_0 \rceil}^T \mathbb{P}\left(2\|\hat\beta_t - \beta^*\|_2 \ge g_t, \mathcal{E}_t\right) + 2\kappa_1 x_{\max} b \sum_{t=\lceil T_0 \rceil}^T \mathbb{P}(\mathcal{E}_t^c)\,.$$

Let $g_t := \frac{3\sqrt{s_0}\lambda_t}{2\kappa_0 \phi_t^2} = \frac{6\sigma x_{\max}}{\kappa_0 \phi_t^2}\sqrt{\frac{s_0(4\log t + 2\log d)}{t}}$. From Lemma 1, we have

$$\mathbb{P}\left(2\|\hat\beta_t - \beta^*\|_2 \ge g_t, \mathcal{E}_t\right) \le \frac{2}{t^2}$$

for all t. Therefore, it follows that

$$\sum_{t=\lceil T_0 \rceil}^{T} \mathbb{P}\left(2\|\hat{\beta}_t - \beta^*\|_2 \geq g_t, \mathcal{E}_t\right) \leq \sum_{t=1}^{T} \mathbb{P}\left(2\|\hat{\beta}_t - \beta^*\|_2 \geq g_t, \mathcal{E}_t\right) \leq \frac{\pi^2}{3} < 4\,.$$

For $t \geq T_0$, we have $\phi_t^2 \geq \frac{\phi_1^2}{2\nu}$ provided that event $\mathcal{E}_t$ holds. Hence, we have

$$\begin{aligned}
\sum_{t=\lceil T_0 \rceil}^{T} g_t &= \sum_{t=\lceil T_0 \rceil}^{T} \frac{6\sigma x_{\max}}{\kappa_0 \phi_t^2} \sqrt{\frac{s_0(4\log t + 2\log d)}{t}} \\
&\leq \sum_{t=\lceil T_0 \rceil}^{T} \frac{12\nu\sigma x_{\max}}{\kappa_0 \phi_1^2} \sqrt{\frac{s_0(4\log t + 2\log d)}{t}} \\
&\leq \frac{12\nu\sigma x_{\max}\sqrt{s_0(4\log T + 2\log d)}}{\kappa_0 \phi_1^2} \sum_{t=1}^{T} \frac{1}{\sqrt{t}} \\
&\leq \frac{24\nu\sigma x_{\max}\sqrt{s_0(4\log T + 2\log d)}}{\kappa_0 \phi_1^2} \sqrt{T}
\end{aligned}$$

where the last inequality is from the fact that $\sum_{t=1}^{T} \frac{1}{\sqrt{t}} \leq \int_{t=0}^{T} \frac{1}{\sqrt{t}} = 2\sqrt{T}$. Combining all the results with the bounds on T_0 and $\sum_{t=\lceil T_0 \rceil}^{T} \mathbb{P}(\mathcal{E}_t^c)$ from the proof of Theorem 1, the expected regret under the RE condition is bounded by

$$\mathcal{R}^{\pi}(T) \leq 4\kappa_1 + \frac{4\kappa_1 x_{\max} b(\log(2d^2) + 1)}{C_2(\phi_1, s_0)^2} + \frac{48\kappa_1 \nu\sigma x_{\max}\sqrt{s_0 T \log(dT)}}{\kappa_0 \phi_1^2}$$

where $C_2(\phi_1, s_0) = \min\left(\frac{1}{2}, \frac{\phi_1^2}{256 s_0 \nu x_{\max}^2}\right)$. □

D. Regret Analysis for *K*-Armed Case

D.1. Proof Outline of Theorem 2

As discussed in Section 5, the analysis for the *K*-armed bandit mostly follows the proof of the two-armed bandit analysis in Section 4. Assuming the compatibility condition of the empirical Gram matrix $\hat{\Sigma}_t$, the Lasso oracle inequality for adapted samples in Lemma 1 can be directly applied. Hence, what we have left is ensuring the compatibility condition of $\hat{\Sigma}_t$. As before, for each $\mathbb{E}[X_\tau X_\tau^\top | \mathcal{F}_\tau]$ in Σ_t, the history $\mathcal{F}_\tau$ affects how feature vector X_τ is chosen. Similar to the two-armed bandit case, we rewrite Σ_t as

$$\Sigma_t = \frac{1}{t}\sum_{\tau=1}^{t}\sum_{i=1}^{K} \mathbb{E}_{\mathcal{X}_t}\left[X_{\tau,i}X_{\tau,i}^\top \mathbb{1}\{X_{\tau,i} = \operatorname*{argmax}_{X \in \mathcal{X}_\tau} X^\top \hat{\beta}_\tau\} \mid \hat{\beta}_\tau\right].$$

Recall that the compatibility condition is only assumed for the theoretical Gram matrix Σ (Assumption 3). Again, the adapted Gram matrix Σ_t is used to bridge Σ and $\hat{\Sigma}_t$ to ensure the compatibility of $\hat{\Sigma}_t$. The key difference between the two-armed bandit analysis and the K-armed bandit analysis lies in how Σ_t is controlled by Σ. In particular, under the balanced covariance condition in Assumption 5, we show the following lemma which is a generalization of Lemma 2.

Lemma 10. *Suppose Assumption 5 holds. For a fixed vector $\beta \in \mathbb{R}^d$, we have*

$$\sum_{i=1}^{K} \mathbb{E}_{\mathcal{X}_t}\left[X_{t,i}X_{t,i}^\top \mathbb{1}\{X_i = \operatorname*{argmax}_{X \in \mathcal{X}_t} X^\top \beta\}\right] \succcurlyeq (2\nu C_{\mathcal{X}})^{-1}\Sigma\,.$$

With this result, we can lower-bound the compatibility constant $\phi^2(\Sigma_t, S_0)$ of the adapted Gram matrix in terms of the compatibility constant $\phi^2(\Sigma, S_0)$ for the theoretical Gram matrix. That is, we have $\Sigma_t \succcurlyeq (2\nu C_{\mathcal{X}})^{-1}\Sigma$ which implies that

$$\phi^2(\Sigma_t, S_0) \geq \frac{\phi^2(\Sigma, S_0)}{2\nu C_{\mathcal{X}}} > 0\,.$$

Hence, Σ_t satisfies the compatibility condition. Then, we can show that $\hat{\Sigma}_t$ concentrates to Σ_t with high probability which directly follows from applying Lemma 2, which is formally stated as follows.

Corollary 4. *For $t \geq \frac{2\log(2d^2)}{C_1(s_0)^2}$ where $C_1(s_0) = \min\left(\frac{1}{2}, \frac{\phi_0^2}{256 s_0 \nu C_{\mathcal{X}} x_{\max}^2}\right)$, we have*

$$\mathbb{P}\left(\|\Sigma_t - \hat{\Sigma}_t\|_\infty \geq \frac{\phi_0^2}{32 s_0 \nu C_{\mathcal{X}}}\right) \leq \exp\left\{-\frac{tC_1(s_0)^2}{2}\right\}.$$

Now, we can invoke Corollary 1 to connect this matrix concentration result to guaranteeing the compatibility condition of $\hat{\Sigma}_t$. Therefore, $\hat{\Sigma}_t$ satisfies the compatibility condition with compatibility constant $\phi_t^2 = \frac{\phi_0^2}{4\nu C_{\mathcal{X}}} > 0$. The rest of the proof of Theorem 2 directly follows the proof of Theorem 1 using this compatibility constant.

D.2. Proof of Lemma 10

Proof. Since the distribution of $\mathcal{X}_t = \{X_{t,1}, ..., X_{t,K}\}$ is time-invariant, we suppress the subscript on t and write $\mathcal{X} = \{X_1, ..., X_K\}$. Let joint distribution of $\mathcal{X}$ as $p_{\mathcal{X}}(x_1, ..., x_K) = p_{\mathcal{X}}(\mathbf{x})$ where we let $\mathbf{x} = (x_1, ..., x_K)$. All expectations in this proof is taken with respect to the tuple $\mathcal{X}$. Then the theoretical Gram matrix is defined as

$$\begin{aligned}\mathbb{E}[\mathbf{X}^\top \mathbf{X}] &= \mathbb{E}\left[\sum_{i=1}^K X_i X_i^\top\right] \\ &= \int (x_1 x_1^\top + ... + x_K x_K^\top) p_{\mathcal{X}}(\mathbf{x}) d\mathbf{x}\end{aligned}$$

Let's first focus on $\int x_1 x_1^\top p_{\mathcal{X}}(\mathbf{x}) d\mathbf{x}$.

$$\begin{aligned}\int x_1 x_1^\top p_{\mathcal{X}}(\mathbf{x}) d\mathbf{x} = &\int x_1 x_1^\top \mathbb{1}\left\{x_1 = \operatorname*{argmax}_{x_i \in \mathcal{X}} x_i^\top \beta\right\} p_{\mathcal{X}}(\mathbf{x}) d\mathbf{x} \\ &+ \int x_1 x_1^\top \mathbb{1}\left\{x_1 = \operatorname*{argmin}_{x_i \in \mathcal{X}} x_i^\top \beta\right\} p_{\mathcal{X}}(\mathbf{x}) d\mathbf{x} \\ &+ \int x_1 x_1^\top \mathbb{1}\left\{x_1 \neq \operatorname*{argmax}_{x_i \in \mathcal{X}} x_i^\top \beta, x_1 \neq \operatorname*{argmin}_{x_i \in \mathcal{X}} x_i^\top \beta\right\} p_{\mathcal{X}}(\mathbf{x}) d\mathbf{x}.\end{aligned}$$

We define three disjoint sets of possible orderings for $\{1, ..., K\}$ as follows.

Definition 5. *We define the following sets of permutations of* $(1, ..., K)$.

$$\begin{aligned}\mathcal{I}_1^{\max} &:= \{\textit{indices } (i_1, ..., i_K) \textit{ such that } i_K = 1\} \\ \mathcal{I}_1^{\min} &:= \{\textit{indices } (i_1, ..., i_K) \textit{ such that } i_1 = 1\} \\ \mathcal{I}_1^{mid} &:= \{\textit{indices } (i_1, ..., i_K) \textit{ such that } i_1 \neq 1 \textit{ and } i_K \neq 1\}.\end{aligned}$$

Then, for $\int x_1 x_1^\top \mathbb{1}\{x_1 = \operatorname{argmin}_{x_i \in \mathcal{X}} x_i^\top \beta\} p_{\mathcal{X}}(\mathbf{x}) d\mathbf{x}$, we can write

$$\int x_1 x_1^\top \mathbb{1}\left\{x_1 = \operatorname*{argmin}_{x_i \in \mathcal{X}} x_i^\top \beta\right\} p_{\mathcal{X}}(\mathbf{x}) d\mathbf{x} = \sum_{(i_1, ..., i_K) \in \mathcal{I}_1^{\min}} \int x_1 x_1^\top \mathbb{1}\left\{x_{i_1}^\top \beta \leq ... \leq x_{i_K}^\top \beta\right\} p_{\mathcal{X}}(\mathbf{x}) d\mathbf{x}$$

Then for any $(i_1, ..., i_K) \in \mathcal{I}_1^{\min}$,

$$\begin{aligned}\int x_1 x_1^\top \mathbb{1}\left\{x_{i_1}^\top \beta \leq ... \leq x_{i_K}^\top \beta\right\} p_{\mathcal{X}}(\mathbf{x}) d\mathbf{x} &= \int x_1 x_1^\top \mathbb{1}\left\{-x_{i_1}^\top \beta \geq ... \geq -x_{i_K}^\top \beta\right\} p_{\mathcal{X}}(\mathbf{x}) d\mathbf{x} \\ &\preccurlyeq \nu \int x_1 x_1^\top \mathbb{1}\left\{-x_{i_1}^\top \beta \geq ... \geq -x_{i_K}^\top \beta\right\} p_{\mathcal{X}}(-\mathbf{x}) d\mathbf{x} \\ &= \nu \int x_1 x_1^\top \mathbb{1}\left\{x_{i_1}^\top \beta \geq ... \geq x_{i_K}^\top \beta\right\} p_{\mathcal{X}}(\mathbf{x}) d\mathbf{x}\end{aligned}$$

where the inequality is again from Assumption 4. Since the elements in $\mathcal{I}_1^{\min}$ can be considered as reversed orderings of elements in $\mathcal{I}_1^{\max}$ (and obviously $|\mathcal{I}_1^{\min}| = |\mathcal{I}_1^{\max}|$),

$$\begin{aligned}
\mathbb{E}\left[X_1 X_1^\top \mathbb{1}\{X_1 = \operatorname*{argmin}_{X\in\mathcal{X}} X^\top\beta\}\right] &= \int x_1 x_1^\top \mathbb{1}\Big\{x_1 = \operatorname*{argmin}_{x_i\in\mathcal{X}} x_i^\top\beta\Big\} p_{\mathcal{X}}(\mathbf{x}) d\mathbf{x} \\
&= \sum_{(i_1,\dots,i_K)\in\mathcal{I}_1^{\min}} \int x_1 x_1^\top \mathbb{1}\Big\{x_{i_1}^\top\beta \le \dots \le x_{i_K}^\top\beta\Big\} p_{\mathcal{X}}(\mathbf{x}) d\mathbf{x} \\
&\preccurlyeq \sum_{(i_1,\dots,i_K)\in\mathcal{I}_1^{\min}} \nu \int x_1 x_1^\top \mathbb{1}\Big\{x_{i_1}^\top\beta \ge \dots \ge x_{i_K}^\top\beta\Big\} p_{\mathcal{X}}(\mathbf{x}) d\mathbf{x} \\
&= \nu \int x_1 x_1^\top \mathbb{1}\Big\{x_1 = \operatorname*{argmax}_{x_i\in\mathcal{X}} x_i^\top\beta\Big\} p_{\mathcal{X}}(\mathbf{x}) d\mathbf{x} \\
&= \nu\mathbb{E}\left[X_1 X_1^\top \mathbb{1}\{X_1 = \operatorname*{argmax}_{X\in\mathcal{X}} X^\top\beta\}\right] .
\end{aligned}$$

Also, using the definitions of $\mathcal{I}_1^{\min}, \mathcal{I}_1^{\text{mid}}$ and $\mathcal{I}_1^{\max}$, we can rewrite $\mathbb{E}\left[X_1 X_1^\top\right]$.

$$\begin{aligned}
\mathbb{E}\left[X_1 X_1^\top\right] &= \mathbb{E}\left[X_1 X_1^\top \mathbb{1}\{X_1 = \operatorname*{argmin}_{X\in\mathcal{X}} X^\top\beta\}\right] + \mathbb{E}\left[X_1 X_1^\top \mathbb{1}\{X_1 = \operatorname*{argmax}_{X\in\mathcal{X}} X^\top\beta\}\right] \\
&\quad + \mathbb{E}\left[X_1 X_1^\top \mathbb{1}\{X_1 \ne \operatorname*{argmin}_{X\in\mathcal{X}} X^\top\beta, X_1 \ne \operatorname*{argmax}_{X\in\mathcal{X}} X^\top\beta\}\right] \\
&= \sum_{(i_1,\dots,i_K)\in\mathcal{I}_1^{\text{min}}} \mathbb{E}\left[X_1 X_1^\top \mathbb{1}\{X_{i_1}^\top\beta < \cdots < X_{i_K}^\top\beta\}\right] \\
&\quad + \sum_{(i_1,\dots,i_K)\in\mathcal{I}_1^{\text{max}}} \mathbb{E}\left[X_1 X_1^\top \mathbb{1}\{X_{i_1}^\top\beta < \cdots < X_{i_K}^\top\beta\}\right] \\
&\quad + \sum_{(i_1,\dots,i_K)\in\mathcal{I}_1^{\text{mid}}} \mathbb{E}\left[X_1 X_1^\top \mathbb{1}\{X_{i_1}^\top\beta < \cdots < X_{i_K}^\top\beta\}\right] \\
&= \sum_{(i_1,\dots,i_K)\in\mathcal{I}_1^{\text{min}}} \mathbb{E}\left[X_{i_1} X_{i_1}^\top \mathbb{1}\{X_{i_1}^\top\beta < \cdots < X_{i_K}^\top\beta\}\right] \\
&\quad + \sum_{(i_1,\dots,i_K)\in\mathcal{I}_1^{\text{max}}} \mathbb{E}\left[X_{i_K} X_{i_K}^\top \mathbb{1}\{X_{i_1}^\top\beta < \cdots < X_{i_K}^\top\beta\}\right] \\
&\quad + \sum_{(i_1,\dots,i_K)\in\mathcal{I}_1^{\text{mid}}} \mathbb{E}\left[X_1 X_1^\top \mathbb{1}\{X_{i_1}^\top\beta < \cdots < X_{i_K}^\top\beta\}\right] .
\end{aligned}$$

sFrom Assumption 5, we have

$$\mathbb{E}\left[X_1 X_1^\top \mathbb{1}\{X_{i_1}^\top\beta < \cdots < X_{i_K}^\top\beta\}\right] \preccurlyeq C_{\mathcal{X}}\mathbb{E}\left[(X_{i_1} X_{i_1}^\top + X_{i_K} X_{i_K}^\top)\mathbb{1}\{X_{i_1}^\top\beta < \cdots < X_{i_K}^\top\beta\}\right] .$$

Then it follows that

$$\begin{aligned}
\mathbb{E}\left[X_1 X_1^\top\right] \preccurlyeq & \sum_{(i_1,\dots,i_K)\in\mathcal{I}_1^{\text{min}}} \mathbb{E}\left[X_{i_1}X_{i_1}^\top \mathbb{1}\{X_{i_1}^\top\beta < \cdots < X_{i_K}^\top\beta\}\right] \\
& + \sum_{(i_1,\dots,i_K)\in\mathcal{I}_1^{\text{max}}} \mathbb{E}\left[X_{i_K}X_{i_K}^\top \mathbb{1}\{X_{i_1}^\top\beta < \cdots < X_{i_K}^\top\beta\}\right] \\
& + \sum_{(i_1,\dots,i_K)\in\mathcal{I}_1^{\text{mid}}} C_{\mathcal{X}}\mathbb{E}\left[\left(X_{i_1}X_{i_1}^\top + X_{i_K}X_{i_K}^\top\right)\mathbb{1}\{X_{i_1}^\top\beta < \cdots < X_{i_K}^\top\beta\}\right] \\
\preccurlyeq & \sum_{(i_1,\dots,i_K)\in\mathcal{I}_1^{\text{min}}} C_{\mathcal{X}}\mathbb{E}\left[\left(X_{i_1}X_{i_1}^\top + X_{i_K}X_{i_K}^\top\right)\mathbb{1}\{X_{i_1}^\top\beta < \cdots < X_{i_K}^\top\beta\}\right] \\
& + \sum_{(i_1,\dots,i_K)\in\mathcal{I}_1^{\text{max}}} C_{\mathcal{X}}\mathbb{E}\left[\left(X_{i_1}X_{i_1}^\top + X_{i_K}X_{i_K}^\top\right)\mathbb{1}\{X_{i_1}^\top\beta < \cdots < X_{i_K}^\top\beta\}\right] \\
& + \sum_{(i_1,\dots,i_K)\in\mathcal{I}_1^{\text{mid}}} C_{\mathcal{X}}\mathbb{E}\left[\left(X_{i_1}X_{i_1}^\top + X_{i_K}X_{i_K}^\top\right)\mathbb{1}\{X_{i_1}^\top\beta < \cdots < X_{i_K}^\top\beta\}\right] .
\end{aligned}$$

Since $\mathcal{I}_1^{\min}, \mathcal{I}_1^{\text{mid}}$ and $\mathcal{I}_1^{\max}$ are disjoint sets, we can write

$$\begin{aligned}
\mathbb{E}\left[X_i X_i^\top \mathbb{1}\{X_i = \operatorname*{argmin}_{X\in\mathcal{X}} X^\top\beta\}\right] = & \sum_{(i_1,\dots,i_K)\in\mathcal{I}_1^{\text{min}}} \mathbb{E}\left[X_{i_1}X_{i_1}^\top \mathbb{1}\{X_{i_1}^\top\beta < \cdots < X_{i_K}^\top\beta\}\right] \\
& + \sum_{(i_1,\dots,i_K)\in\mathcal{I}_1^{\text{max}}} \mathbb{E}\left[X_{i_1}X_{i_1}^\top \mathbb{1}\{X_{i_1}^\top\beta < \cdots < X_{i_K}^\top\beta\}\right] \\
& + \sum_{(i_1,\dots,i_K)\in\mathcal{I}_1^{\text{mid}}} \mathbb{E}\left[X_{i_1}X_{i_1}^\top \mathbb{1}\{X_{i_1}^\top\beta < \cdots < X_{i_K}^\top\beta\}\right] .
\end{aligned}$$

We can also express $\mathbb{E}\left[X_i X_i^\top \mathbb{1}\{X_i = \operatorname{argmax}_{X\in\mathcal{X}} X^\top\beta\}\right]$ similarly. Therefore, we have

$$\begin{aligned}
\mathbb{E}\left[X_1X_1^\top\right] &\preccurlyeq C_{\mathcal{X}} \sum_{i=1}^K \left(\mathbb{E}\left[X_iX_i^\top \mathbb{1}\{X_i = \operatorname*{argmin}_{X\in\mathcal{X}} X^\top\beta\}\right] + \mathbb{E}\left[X_iX_i^\top \mathbb{1}\{X_i = \operatorname*{argmax}_{X\in\mathcal{X}} X^\top\beta\}\right]\right) \\
&\preccurlyeq C_{\mathcal{X}}(1+\nu)\sum_{i=1}^K \mathbb{E}\left[X_iX_i^\top \mathbb{1}\{X_i = \operatorname*{argmax}_{X\in\mathcal{X}} X^\top\beta\}\right] .
\end{aligned}$$

Then, summing $\mathbb{E}\left[X_jX_j^\top\right]$ over all $j = 1, ..., K$ gives

$$\mathbb{E}[\mathbf{X}^\top\mathbf{X}] = \sum_{j=1}^K \mathbb{E}\left[X_jX_j^\top\right] \preccurlyeq KC_{\mathcal{X}}(1+\nu)\sum_{i=1}^K \mathbb{E}\left[X_iX_i^\top \mathbb{1}\{X_i = \operatorname*{argmax}_{X\in\mathcal{X}} X^\top\beta\}\right] .$$

Hence,

$$\sum_{i=1}^K \mathbb{E}\left[X_iX_i^\top \mathbb{1}\{X_i = \operatorname*{argmax}_{X\in\mathcal{X}} X^\top\beta\}\right] \succcurlyeq \frac{1}{C_{\mathcal{X}}(1+\nu)}\cdot\frac{1}{K}\mathbb{E}[\mathbf{X}^\top\mathbf{X}] \succcurlyeq (2C_{\mathcal{X}}\nu)^{-1}\Sigma .$$

□

D.3. Proposition 1

Proposition 1. *In the case of independent arms, both a multivariate Gaussian distribution and a uniform distribution on a unit sphere satisfy Assumption 5 with $C_{\mathcal{X}} = \mathcal{O}(1)$. For an arbitrary distribution, it holds with $C_{\mathcal{X}} = \binom{K-1}{K_0}$ where $K_0 = \lceil (K-1)/2 \rceil$.*

The proof of Proposition 1 involves the following few technical lemmas.

Lemma 11. *Suppose each $X_i \in \mathbb{R}^d$ is i.i.d. Gaussian with mean μ and covariance matrix Γ. For any permutation $(i_1, ..., i_K)$ of $(1, ..., K)$, any integer $k \in \{2, ..., K-1\}$ and fixed β,*

$$\begin{aligned}\mathbb{E}\Big[X_{i_k} X_{i_k}^\top \mathbb{1}\{X_{i_1}^\top \beta < ... < X_{i_K}^\top \beta\}\Big] \preccurlyeq \mathbb{E}\Big[X_{i_1} X_{i_1}^\top \mathbb{1}\{X_{i_1}^\top \beta < ... < X_{i_K}^\top \beta\}\Big] \\ + \mathbb{E}\Big[X_{i_K} X_{i_K}^\top \mathbb{1}\{X_{i_1}^\top \beta < ... < X_{i_K}^\top \beta\}\Big] .\end{aligned}$$

Proof. It suffices to show that for any $y \in \mathbb{R}^d$

$$\begin{aligned}&\mathbb{E}\left[(X_{i_k}^\top y)^2 \mathbb{1}\{X_{i_1}^\top \beta < ... < X_{i_K}^\top \beta\}\right] \\ &\le \mathbb{E}\left[(X_{i_1}^\top y)^2 \mathbb{1}\{X_{i_1}^\top \beta < ... < X_{i_K}^\top \beta\}\right] + \mathbb{E}\left[(X_{i_K}^\top y)^2 \mathbb{1}\{X_{i_1}^\top \beta < ... < X_{i_K}^\top \beta\}\right] .\end{aligned}$$

Now, we can write

$$y = \widetilde{\beta}(\widetilde{\beta}^\top y) + \sum_{j=1}^{d-1} g_j g_j^\top y := \widetilde{\beta} w_0 + \sum_{j=1}^{d-1} g_j g_j^\top y \,.$$

where $w_0 = \widetilde{\beta}^\top y$ and $\widetilde{\beta} = \frac{\beta}{\|\beta\|}$ and $\left[\widetilde{\beta}, g_i, ..., g_{d-1}\right]$ form an orthonormal basis. For $i \in [N]$, we can write

$$\begin{aligned}X_i^\top y &= (X_i^\top \widetilde{\beta}) w_0 + X_i^\top \left(\sum_{j=1}^{d-1} g_j g_j^\top\right) y \\ &= (X_i^\top \widetilde{\beta}) w_0 + \left[\left(\sum_{j=1}^{d-1} g_j g_j^\top\right) X_i\right]^\top y \,.\end{aligned}$$

Then we define the following two random variables

$$U_i := X_i^\top \widetilde{\beta}, \qquad V_i := G X_i$$

where $G = \sum_{j=1}^{d-1} g_j g_j^\top$. Then we have

$$\begin{bmatrix} U_i \\ V_i \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu^\top \widetilde{\beta} \\ G\mu \end{bmatrix}, \quad \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}\right)$$

where

$$\begin{aligned}A_{11} &= \widetilde{\beta}^\top \Gamma \widetilde{\beta} \in \mathbb{R} \\ A_{12} &= A_{21}^\top = \widetilde{\beta}^\top \Gamma G^\top \in \mathbb{R}^{1\times d} \\ A_{22} &= G \Gamma G^\top \in \mathbb{R}^{d \times d} \,.\end{aligned}$$

Then, we know from Lemma 15 that the conditional distribution $V_i \mid U_i$ of a multivariate normal distribution is also a multivariate normal distribution. In particular,

$$V_i \mid U_i = u_i \sim \mathcal{N}\left(G\mu + A_{21} A_{11}^{-1}(u_i - \mu^\top \widetilde{\beta}), B\right)$$

where $B = A_{22} - A_{21} A_{11}^{-1} A_{12}$. Therefore, given $U_{i_k} = u_{i_k}$, we can write

$$\begin{aligned}X_{i_k}^\top y &= u_{i_k} w_0 + V_{i_k}^\top y \\ &= u_{i_k} w_0 + \left(G\mu + A_{21} A_{11}^{-1}(u_{i_k} - \mu^\top \widetilde{\beta}) + B^{1/2} Z\right)^\top y \,.\end{aligned}$$

where $Z \sim \mathcal{N}(0, I_d)$ and $Z \perp\!\!\!\perp U_{i_k}$. Rearranging gives

$$X_{i_k}^\top y = u_{i_k}\left(w_0 + A_{11}^{-1}A_{12}y\right) + \left(G\mu - A_{21}A_{11}^{-1}\mu^\top\widetilde{\beta}\right)^\top y + Z^\top B^{1/2}y\,.$$

Hence, $X_{i_k}^\top y$ is a linear function of u_{i_k}. Then it follows that

$$\begin{aligned}
\left(X_{i_k}^\top y\right)^2 &= \left[u_{i_k}\left(w_0 + A_{11}^{-1}A_{12}y\right) + \left(G\mu - A_{21}A_{11}^{-1}\mu^\top\widetilde{\beta}\right)^\top y + Z^\top B^{1/2}y\right]^2 \\
&\le \max\left\{\left[u_{i_1}\left(w_0 + A_{11}^{-1}A_{12}y\right) + \left(G\mu - A_{21}A_{11}^{-1}\mu^\top\widetilde{\beta}\right)^\top y + Z^\top B^{1/2}y\right]^2,\right. \\
&\qquad\left.\left[u_{i_K}\left(w_0 + A_{11}^{-1}A_{12}y\right) + \left(G\mu - A_{21}A_{11}^{-1}\mu^\top\widetilde{\beta}\right)^\top y + Z^\top B^{1/2}y\right]^2\right\} \\
&\le \left[u_{i_1}\left(w_0 + A_{11}^{-1}A_{12}y\right) + \left(G\mu - A_{21}A_{11}^{-1}\mu^\top\widetilde{\beta}\right)^\top y + Z^\top B^{1/2}y\right]^2 \\
&\quad + \left[u_{i_K}\left(w_0 + A_{11}^{-1}A_{12}y\right) + \left(G\mu - A_{21}A_{11}^{-1}\mu^\top\widetilde{\beta}\right)^\top y + Z^\top B^{1/2}y\right]^2.
\end{aligned}$$

Therefore, it follows that

$$\begin{aligned}
&\mathbb{E}\left[(X_{i_k}^\top y)^2\mathbb{1}\{X_{i_1}^\top\beta < \ldots < X_{i_K}^\top\beta\}\right] \\
&\le \mathbb{E}\left[(X_{i_1}^\top y)^2\mathbb{1}\{X_{i_1}^\top\beta < \ldots < X_{i_K}^\top\beta\}\right] + \mathbb{E}\left[(X_{i_K}^\top y)^2\mathbb{1}\{X_{i_1}^\top\beta < \ldots < X_{i_K}^\top\beta\}\right].
\end{aligned}$$

Hence,

$$\mathbb{E}\left[X_{i_k}X_{i_k}^\top\mathbb{1}\{X_{i_1}^\top\beta < \ldots < X_{i_K}^\top\beta\}\right] \preccurlyeq \mathbb{E}\left[(X_{i_1}X_{i_1}^\top + X_{i_K}X_{i_K}^\top)\mathbb{1}\{X_{i_1}^\top\beta < \ldots < X_{i_K}^\top\beta\}\right].$$

□

Lemma 12. *Suppose $X \in \mathbb{R}^d$ is uniformly distributed on the unit sphere $\mathcal{S}^{d-1}$ and $K = o(d)$. For fixed vector $\beta \in \mathbb{R}^d$ and a given integer $k \in \{2, \ldots, K-1\}$,*

$$\mathbb{E}\left[X_{i_k}X_{i_k}^\top\mathbb{1}\{X_{i_1}^\top\beta < \ldots < X_{i_K}^\top\beta\}\right] \preccurlyeq C_{\mathcal{X}}\mathbb{E}\left[(X_{i_1}X_{i_1}^\top + X_{i_K}X_{i_K}^\top)\mathbb{1}\{X_{i_1}^\top\beta < \ldots < X_{i_K}^\top\beta\}\right].$$

where $C_{\mathcal{X}} = \mathcal{O}(1)$.

Proof. Here, we instead show directly

$$\mathbb{E}[XX^\top] \preccurlyeq C\left(\mathbb{E}\Big[XX^\top\mathbb{1}\{X = \operatorname*{argmax}_{X_i\in\{X_1,\ldots,X_K\}} X_i^\top\widetilde{\beta}\}\Big] + \mathbb{E}\Big[XX^\top\mathbb{1}\{X = \operatorname*{argmin}_{X_i\in\{X_1,\ldots,X_K\}} X_i^\top\widetilde{\beta}\}\Big]\right)$$

for some constant C. It can be shown that if $C = \mathcal{O}(1)$, then the claim holds with $C_{\mathcal{X}} = \mathcal{O}(1)$. Suppose $X \in \mathbb{R}^d$ is uniformly distributed on the unit sphere $\mathcal{S}^{d-1} := \{s \in \mathbb{R}^d : \|s\|_2 = 1\}$. Then by Lemma 2 in (Cambanis et al., 1981), we can write for each X_i,

$$X_i \sim \left(B_iU_{i,1}, (1 - B_i^2)^{1/2}U_{i,2}\right)$$

where $B_i \sim \text{beta}\left(\frac{1}{2}, \frac{d-1}{2}\right)$, $U_{i,1} = \pm 1$ with probability $\frac{1}{2}$, $U_{i,2} \sim \text{unif}(\mathcal{S}^{d-2})$. $U_{i,1}, U_{i,2}$ and B_i are independent of each other. Similar to the analysis of the Gaussian case, we can normalize β so that $\widetilde{\beta} = \frac{\beta}{\|\beta\|}$. Without loss of generality, assume that $\widetilde{\beta} = [1, 0, \ldots, 0]^\top$. That is, only the first element is non-zero. We can do this since X is spherical and rotation invariant. Then we can write

$$\mathbb{E}\Big[XX^\top\mathbb{1}\{X = \operatorname*{argmax}_{X_i\in\{X_1,\ldots,X_K\}} X_i^\top\widetilde{\beta}\}\Big] = \mathbb{E}\Big[XX^\top\mathbb{1}\{X = \operatorname*{argmax}_{X_i\in\{X_1,\ldots,X_K\}} X_i^{(1)}\}\Big]$$

where $X_i^{(1)}$ is the first element of X_i. Similarly,

$$\mathbb{E}\Big[XX^\top \mathbb{1}\{X = \operatorname*{argmin}_{X_i\in\{X_1,\ldots,X_K\}} X_i^\top \widetilde{\beta}\}\Big] = \mathbb{E}\Big[XX^\top \mathbb{1}\{X = \operatorname*{argmin}_{X_i\in\{X_1,\ldots,X_K\}} X_i^{(1)}\}\Big].$$

Now, from the definition of X, for $B \sim \text{beta}\left(\frac{1}{2}, \frac{d-1}{2}\right)$ we have

$$X_iX_i^\top = \begin{bmatrix} B_i^2 & B_i\sqrt{1-B_i^2}U_{i,1}U_{i,2}^\top \\ B_i\sqrt{1-B_i^2}U_{i,1}U_{i,2} & (1-B_i^2)U_{i,2}U_{i,2}^\top \end{bmatrix}.$$

By the independence of U_1, U_2, and B, we have

$$\mathbb{E}\left[XX^\top\right] = \mathbb{E}\begin{bmatrix} B^2 & 0 \\ 0 & \frac{1}{d-1}(1-B^2)I_{d-1} \end{bmatrix}.$$

By the definitions of B_i and $U_{i,1}$, it follows that

$$\mathbb{E}\Big[XX^\top \mathbb{1}\{B = \max_{B_i\in\{B_1,\ldots,B_K\}} B_i\}\Big] \preccurlyeq \mathbb{E}\Big[XX^\top \mathbb{1}\{X = \operatorname*{argmax}_{X_i\in\{X_1,\ldots,X_K\}} X_i^{(1)}\}\Big] + \mathbb{E}\Big[XX^\top \mathbb{1}\{X = \operatorname*{argmin}_{X_i\in\{X_1,\ldots,X_K\}} X_i^{(1)}\}\Big].$$

Since $\mathbb{E}[B^2] = \frac{(\alpha+1)\alpha}{(\alpha+\beta+1)(\alpha+\beta)}$ for $B \sim \text{beta}(\alpha,\beta)$, we have $\mathbb{E}[B^2] = \frac{3}{d(d+2)}$ and $\frac{1-\mathbb{E}[B^2]}{d-1} = \frac{d+3}{d(d+2)}$ using $\alpha = \frac{1}{2}$ and $\beta = \frac{d-1}{2}$. Clearly, $\lambda_{\min}(\mathbb{E}\left[XX^\top\right]) = \frac{3}{d(d+2)}$. Similarly, for the matrix $\mathbb{E}\left[XX^\top\mathbb{1}\{B = \max_i B_i\}\right]$, we have

$$\mathbb{E}\left[XX^\top \mathbb{1}\{B = \max_i B_i\}\right] = \mathbb{E}\begin{bmatrix} B^2\mathbb{1}\{B = \max_i B_i\} & 0 \\ 0 & \frac{1}{d-1}(1-B^2)\mathbb{1}\{B = \max_i B_i\}I_{d-1} \end{bmatrix}.$$

Note that $\mathbb{E}[B^2\mathbb{1}\{B = \max_i B_i\}] = \sum_{j=1}^K \mathbb{E}[B_j^2\mathbb{1}\{B_j = \max_i B_i\}] \geq \mathbb{E}[B^2]$. Then, we need to show

$$C(1 - \mathbb{E}[B^2\mathbb{1}\{B = \max_i B_i\}]) \geq 1 - \mathbb{E}[B^2]$$

for some C. Note that $\mathbb{E}[B^2\mathbb{1}\{B = \max_i B_i\}] \leq N\mathbb{E}[B^2]$. Hence, we can show

$$C \geq \frac{1-\mathbb{E}[B^2]}{1-N\mathbb{E}[B^2]} = \frac{1-\frac{3}{d(d+2)}}{1-\frac{3K}{d(d+2)}} = \frac{d^2+d-3}{d^2+d-3K}.$$

Since $K = o(d)$, we have $C = \mathcal{O}(1)$. Hence,

$$\begin{aligned}
\mathbb{E}[XX^\top] &\preccurlyeq C\mathbb{E}\Big[XX^\top \mathbb{1}\{B = \max_{B_i\in\{B_1,\ldots,B_K\}} B_i\}\Big] \\
&\preccurlyeq C\left(\mathbb{E}\Big[XX^\top \mathbb{1}\{X = \operatorname*{argmax}_{X_i\in\{X_1,\ldots,X_K\}} X_i^{(1)}\}\Big] + \mathbb{E}\Big[XX^\top \mathbb{1}\{X = \operatorname*{argmin}_{X_i\in\{X_1,\ldots,X_K\}} X_i^{(1)}\}\Big]\right) \\
&= C\left(\mathbb{E}\Big[XX^\top \mathbb{1}\{X = \operatorname*{argmax}_{X_i\in\{X_1,\ldots,X_K\}} X_i^\top\widetilde{\beta}\}\Big] + \mathbb{E}\Big[XX^\top \mathbb{1}\{X = \operatorname*{argmin}_{X_i\in\{X_1,\ldots,X_K\}} X_i^\top\widetilde{\beta}\}\Big]\right)
\end{aligned}$$

which implies $C_{\mathcal{X}} = \mathcal{O}(1)$. □

Lemma 13. *Consider i.i.d. arbitrary distribution $p_{\mathcal{X}}$. Fix some vector $\beta \in \mathbb{R}^d$. For a given integer $k \in \{2, ..., K-1\}$,*

$$\mathbb{E}\left[X_kX_k^\top \mathbb{1}\{X_1^\top\beta < ... < X_k^\top\beta < ... < X_K^\top\beta\}\right] \preccurlyeq C_{K,k}\mathbb{E}\left[(X_1X_1^\top + X_KX_K^\top)\mathbb{1}\{X_1^\top\beta < ... < X_K^\top\beta\}\right]$$

where $C_{\mathcal{X}} = \binom{K-1}{(K-1)/2}$ assuming K is odd — if K is even, we can use $\lceil (K-1)/2 \rceil$.

Proof. First notice that

$$\begin{aligned}&\mathbb{E}\left[X_k X_k^\top \mathbb{1}\{X_1^\top \beta < \cdots < X_k^\top \beta < \cdots < X_K^\top \beta\}\right] \\ &= \mathbb{E}_V\left[VV^\top \mathbb{E}_{X_{1:K}/X_k}\left[\mathbb{1}\{X_1^\top \beta < \cdots < X_{k-1}^\top \beta < V^\top \beta < X_{k+1}^\top \beta < \cdots < X_K^\top \beta\} \mid V\right]\right]\end{aligned}$$

where $X_{1:K}/X_k$ denotes $X_1, ..., X_{k-1}, X_{k+1}, ..., X_K$. Also,

$$\begin{aligned}\mathbb{E}\left[X_1 X_1^\top \mathbb{1}\{X_1^\top \beta < \cdots < X_K^\top \beta\}\right] &= \mathbb{E}_V\left[VV^\top \mathbb{E}_{X_{2:K}}\left[\mathbb{1}\{V^\top \beta < X_2^\top \beta < \cdots < X_K^\top \beta\} \mid V\right]\right] \\ \mathbb{E}\left[X_K X_K^\top \mathbb{1}\{X_1^\top \beta < \cdots < X_K^\top \beta\}\right] &= \mathbb{E}_V\left[VV^\top \mathbb{E}_{X_{1:K-1}}\left[\mathbb{1}\{X_1^\top \beta < \cdots < X_{K-1}^\top \beta < V^\top \beta\} \mid V\right]\right]\end{aligned}$$

Let $\psi(y) := \mathbb{P}(X^\top \beta \le y)$ denote the CDF of $X^\top \beta$. Then

$$\begin{aligned}&\mathbb{P}\left(X_1^\top \beta < \cdots < X_{k-1}^\top \beta < V^\top \beta < X_{k+1}^\top \beta < \cdots < X_K^\top \beta\right) \\ &= \prod_{i=1}^{k-1} \mathbb{P}\left(X_i^\top \beta \le V^\top \beta\right) \frac{1}{(k-1)!} \prod_{i=k+1}^{N} \mathbb{P}\left(X_i^\top \beta \ge V^\top \beta\right) \frac{1}{(K-k)!} \\ &= \frac{1}{(k-1)!(K-k)!} \psi(V^\top \beta)^{k-1} \left(1 - \psi(V^\top \beta)\right)^{K-k}.\end{aligned}$$

Likewise

$$\begin{aligned}\mathbb{P}\left(V^\top \beta < X_2^\top \beta < \cdots < X_K^\top \beta\right) &= \frac{1}{(K-1)!} \left(1 - \psi(V^\top \beta)\right)^{K-1}, \\ \mathbb{P}\left(X_1^\top \beta < \cdots < X_{K-1}^\top \beta < V^\top \beta\right) &= \frac{1}{(K-1)!} \psi(V^\top \beta)^{K-1}.\end{aligned}$$

Then, we need to show there exists $C_{K,k}$ such that

$$\begin{aligned}&\mathbb{P}\left(X_1^\top \beta < \cdots < X_{k-1}^\top \beta < V^\top \beta < X_{k+1}^\top \beta < \cdots < X_K^\top \beta\right) \\ &\le C_{K,k}\left[\mathbb{P}\left(V^\top \beta < X_2^\top \beta < \cdots < X_K^\top \beta\right) + \mathbb{P}\left(X_1^\top \beta < \cdots < X_{K-1}^\top \beta < V^\top \beta\right)\right].\end{aligned}$$

That is,

$$\frac{\psi(V^\top \beta)^{k-1}\left(1 - \psi(V^\top \beta)\right)^{K-k}}{(k-1)!(K-k)!} \le \frac{C_{K,k}}{(K-1)!}\left[\left(1 - \psi(V^\top \beta)\right)^{K-1} + \psi(V^\top \beta)^{K-1}\right].$$

Hence,

$$C_{K,k} \ge \binom{K-1}{k-1} \frac{\psi(V^\top \beta)^{k-1}\left(1 - \psi(V^\top \beta)\right)^{K-k}}{\left(1 - \psi(V^\top \beta)\right)^{K-1} + \psi(V^\top \beta)^{K-1}}.$$

Since $\psi(V^\top \beta) \in [0, 1]$, we have

$$\frac{\psi(V^\top \beta)^{k-1}\left(1 - \psi(V^\top \beta)\right)^{K-k}}{\left(1 - \psi(V^\top \beta)\right)^{K-1} + \psi(V^\top \beta)^{K-1}} \le 1$$

for all K and k. Hence, for $C_{K,k} = \binom{K-1}{k-1}$,

$$\begin{aligned}&\mathbb{E}\left[X_k X_k^\top \mathbb{1}\{X_1^\top \beta < \cdots < X_k^\top \beta < \cdots < X_K^\top \beta\}\right] \\ &\preccurlyeq C_{K,k}\mathbb{E}\left[(X_1 X_1^\top + X_K X_K^\top)\mathbb{1}\{X_1^\top \beta < \cdots < X_K^\top \beta\}\right].\end{aligned}$$

□

E. Other lemmas

Lemma 14 (Wainwright (2019), Theorem 2.19)**.** *Let $\{Z_\tau, \mathcal{F}_\tau\}_\tau^\infty$ be a martingale difference sequence, and suppose that Z_τ is σ^2-sub-Gaussian in an adapted sense, i.e., for all $\alpha \in \mathbb{R}$, $\mathbb{E}[e^{\alpha Z_\tau} | \mathcal{F}_{\tau-1}] \leq e^{\alpha^2\sigma^2/2}$ almost surely. Then for all $\gamma \geq 0$, $\mathbb{P}\left[|\sum_{\tau=1}^n Z_\tau| \geq \gamma\right] \leq 2\exp[-\gamma^2/(2n\sigma^2)]$.*

Note that Lemma 15 is a well-known result, but for the sake of completeness, we present its formal statment and proof.

Lemma 15. *Let $X \in \mathbb{R}^d$ follow a multivariate Gaussian distribution with mean μ and covarance matrix Σ and consider the partition of X with*

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right).$$

Then the conditional distribution of X_1 given X_2 is also a multivariate Gaussian distribution. In particular

$$X_1 \mid X_2 = x_2 \sim \mathcal{N}\left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right).$$

Proof. Define $Z = X_1 + \mathbf{A}X_2$ where $\mathbf{A} = -\Sigma_{12}\Sigma_{22}^{-1}$. Now we can write

$$\begin{aligned} \operatorname{cov}(Z, X_2) &= \operatorname{cov}(X_1, X_2) + \operatorname{cov}(\mathbf{A}X_2, X_2) \\ &= \Sigma_{12} + \mathbf{A}\operatorname{var}(X_2) \\ &= \Sigma_{12} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{22} \\ &= 0 \end{aligned}$$

Therefore Z and X_2 are not correlated and, since they are jointly normal, they are independent[4]. Now, clearly we have $\mathbb{E}(Z) = \mu_1 + \mathbf{A}\mu_2$. Then

$$\begin{aligned} \mathbb{E}[X_1|X_2] &= \mathbb{E}[Z - \mathbf{A}X_2|X_2] \\ &= \mathbb{E}[Z|X_2] - \mathbb{E}[\mathbf{A}X_2|X_2] \\ &= \mathbb{E}[Z] - \mathbf{A}X_2 \\ &= \mu_1 + \mathbf{A}(\mu_2 - X_2) \\ &= \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2). \end{aligned}$$

For the covariance matrix, note that

$$\begin{aligned} \operatorname{var}(X_1|X_2) &= \operatorname{var}(Z - \mathbf{A}X_2|X_2) \\ &= \operatorname{var}(Z|X_2) + \operatorname{var}(\mathbf{A}X_2|X_2) - \mathbf{A}\operatorname{cov}(Z, -X_2) - \operatorname{cov}(Z, -X_2)\mathbf{A}^\top \\ &= \operatorname{var}(Z|X_2) \\ &= \operatorname{var}(Z) \end{aligned}$$

Hence, it follows that

$$\begin{aligned} \operatorname{var}(X_1|X_2) &= \operatorname{var}(Z) \\ &= \operatorname{var}(X_1 + \mathbf{A}X_2) \\ &= \operatorname{var}(X_1) + \mathbf{A}\operatorname{var}(X_2)\mathbf{A}^\top + \mathbf{A}\operatorname{cov}(X_1, X_2) + \operatorname{cov}(X_2, X_1)\mathbf{A}^\top \\ &= \Sigma_{11} + \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{22}\Sigma_{22}^{-1}\Sigma_{21} - 2\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \\ &= \Sigma_{11} + \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} - 2\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \\ &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \end{aligned}$$

□

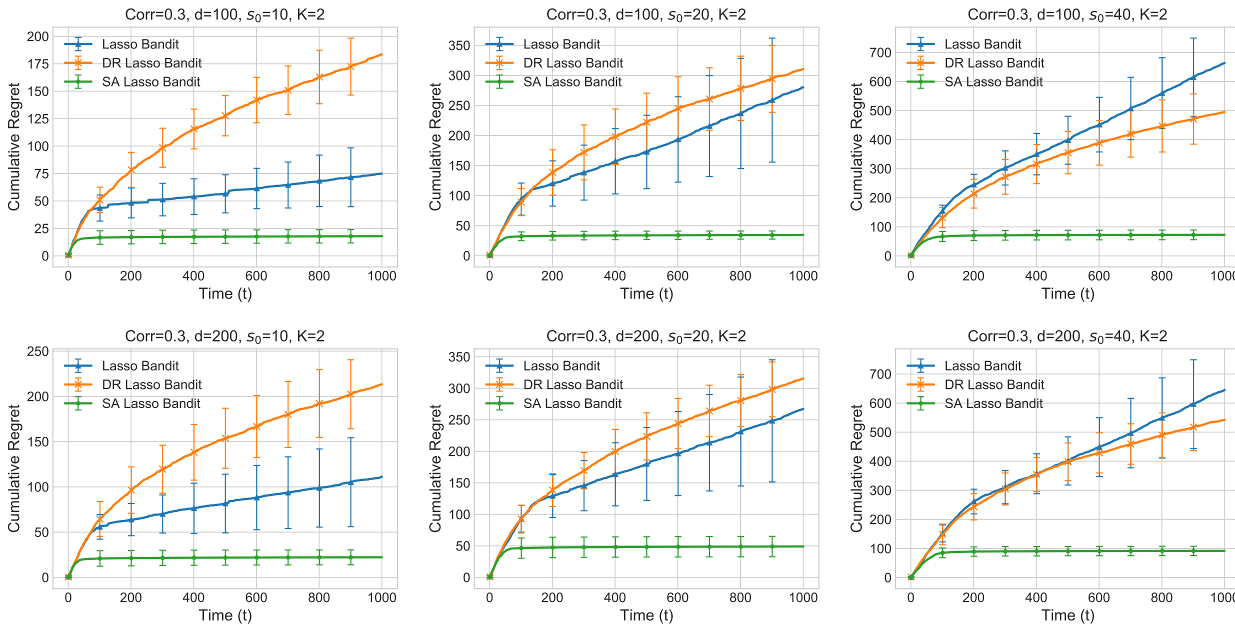

Figure 2. The plots show the t-round regret of SA LASSO BANDIT (Algorithm 1), DR LASSO BANDIT (Kim & Paik, 2019), and LASSO BANDIT (Bastani & Bayati, 2020) for $K = 2$, $d = 100$ (first row) and $d = 200$ (second row) with varying sparsity $s_0 = 10, 20, 40$ under a weak correlation.

F. Numerical Experiment Details

We conduct numerical experiments to evaluate SA LASSO BANDIT and compare with existing sparse bandit algorithms: DR LASSO BANDIT (Kim & Paik, 2019) and LASSO BANDIT (Bastani & Bayati, 2020). We follow the experimental setup of (Kim & Paik, 2019) to evaluate algorithms under different levels of correlation between arms. Although we consider $K = 2$ case in this section, the experimental setup introduced here also applies to numerical evaluations for $K \geq 3$ arm case in Section 5. For each dimension $i \in [d]$, we sample each element of the feature vectors $[x_{t,1}^{(i)}, ..., x_{t,K}^{(i)}]$ from multivariate Gaussian distribution $\mathcal{N}(\vec{0}_K, V)$ where covariance matrix V is defined as $V_{i,i} = 1$ for all diagonal elements i, i and $V_{i,j} = \rho^2$ for all off-diagonal elements $i \neq j \in [d]$. Hence, for $\rho^2 > 0$, feature vectors for each arm are allowed to be correlated. We consider two levels of correlation with $\rho^2 = 0.3$ (weak correlation) and $\rho^2 = 0.7$ (strong correlation). In these two sets of experiments, we consider feature dimensions $d = 100$ and $d = 200$. For comparison, we use a linear reward with the linear link function $\mu(z) = z$ since both LASSO BANDIT and DR LASSO BANDIT are proposed in linear reward settings. We generate β^* with varying sparsity $s_0 = \|\beta^*\|_0$. For a given s_0, we generate the non-zero elements of β^* from a uniform distribution in $[0, 1]$. For each case with different experimental configurations, we conduct 20 independent runs, and report the average of the cumulative regret for each of the algorithms. The error bars represent the standard deviations.

DR LASSO BANDIT is proposed for the same problem setting as ours. Therefore, it does not require any modifications. However, the problem setting of LASSO BANDIT is different from ours: it assumes that the context variable is the same for all arms but the underlying parameter differs for each arm. As was done in the experiments of (Kim & Paik, 2019), LASSO BANDIT can be applied in our setting by constructing a Kd-dimensional context vector $x_t = [x_{t,i}^\top, ..., x_{t,K}^\top]^\top \in \mathbb{R}^{Kd}$ and Kd-dimensional parameter β_i^* for each arm i where $\beta_i^* = [{\beta^*}^\top \mathbb{1}(i = 1), ..., {\beta^*}^\top \mathbb{1}(i = K)]^\top \in \mathbb{R}^{Kd}$. Note that despite the concatenation, the effective dimension of the unknown parameter β_i^* remains the same as far as estimation is concerned.

F.1. Evaluation for *2* Arms

We first discuss the numerical evaluation results for two-armed bandits, which are shown in Figure 2 and Figure 3 under various problem instances. It is important to note that we report the performances of the benchmarks (DR LASSO BANDIT

[4] If a random vector has a multivariate normal distribution then any two or more of its components that are uncorrelated are independent.

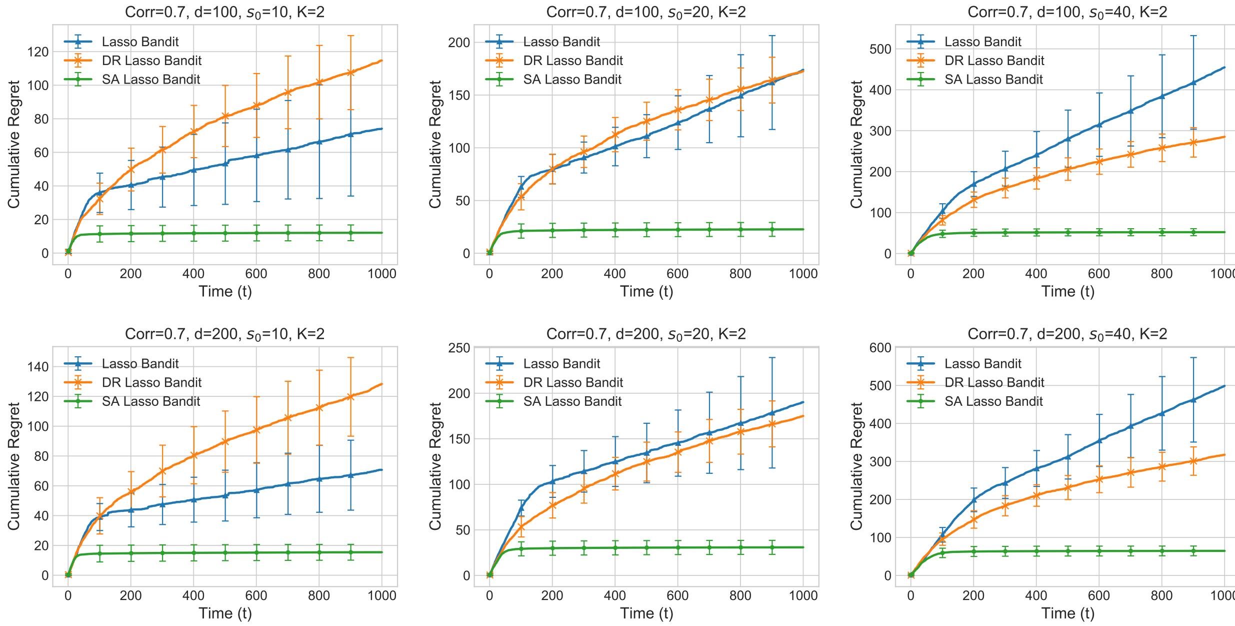

Figure 3. The plots show the t-round regret of SA LASSO BANDIT (Algorithm 1), DR LASSO BANDIT (Kim & Paik, 2019), and LASSO BANDIT (Bastani & Bayati, 2020) for $K = 2$, $d = 100$ (first row) and $d = 200$ (second row) with varying sparsity $s_0 = 10, 20, 40$ under a strong correlation.

and LASSO BANDIT) assuming that they have access to correct sparsity s_0 and this information is kept hidden from ours. Despite such advantage to the benchmarks, the experiment results shown in Figure 2 and Figure 3 demonstrate that SA LASSO BANDIT outperforms the other methods by significant margins consistently across various instances of experiments. We also verify that the performance of our proposed algorithm is the least sensitive and scales very well with changes in problem instances, which suggests that our algorithm is very effective for various high-dimensional bandit problem instances with a sparse structure. Regret scalability on sparsity s_0 appears to be at most linear while dependence on feature dimension d appears to be very minimal in most of the instances, which is consistent with our theoretical findings. We also observe that a higher correlation between arms (feature vectors) improves the overall performances of the algorithms. This finding is further evidenced by the experiments for the K-armed case.

F.2. Evaluation for K Arms

In this section, we validate the performance of SA LASSO BANDIT in K-armed sparse bandit settings via additional numerical experiments and provide comparison with the existing sparse bandit algorithms. The setup of the experiments is identical to the setup described in Section 6. We perform evaluations under various instances. In particular, we focus on the performances of algorithms as the number of arms increases (Figure 1). Additionally, to investigate the effect of the balanced covariance condition, we also evaluate algorithms on features drawn from a non-Gaussian elliptical distribution, for which we do not have a tight bound of $C_{\mathcal{X}}$. In Figure 4, we further investigate the effect of correlation between arms as well as the effect of ambient feature dimension in K-armed settings.

The evaluation results in Figure 1 and Figure 4 again provide the convincing evidence that the performance of our proposed algorithm is superior to the existing sparse bandit methods that we compare with in K-armed bandits. Again, SA LASSO BANDIT outperforms the existing sparse bandit algorithms by significant margins. Furthermore, SA LASSO BANDIT is much more practical and easy to implement with a minimal number of a hyperparameter (only noise variance parameter is needed for our algorithm). We again observe that under strong correlation, the algorithms generally perform better compared to weak correlation instances. This strong correlation would imply a smaller $C_{\mathcal{X}}$ as briefly discussed earlier when we introduce the balanced covariance condition. Hence, the results are consistent with our theoretical findings. Note that strong correlation does not immediately imply that the performances are generally better since it potentially decreases the value of compatibility constant. Thus, the regret would increase with an increase in correlation as far as the compatibility

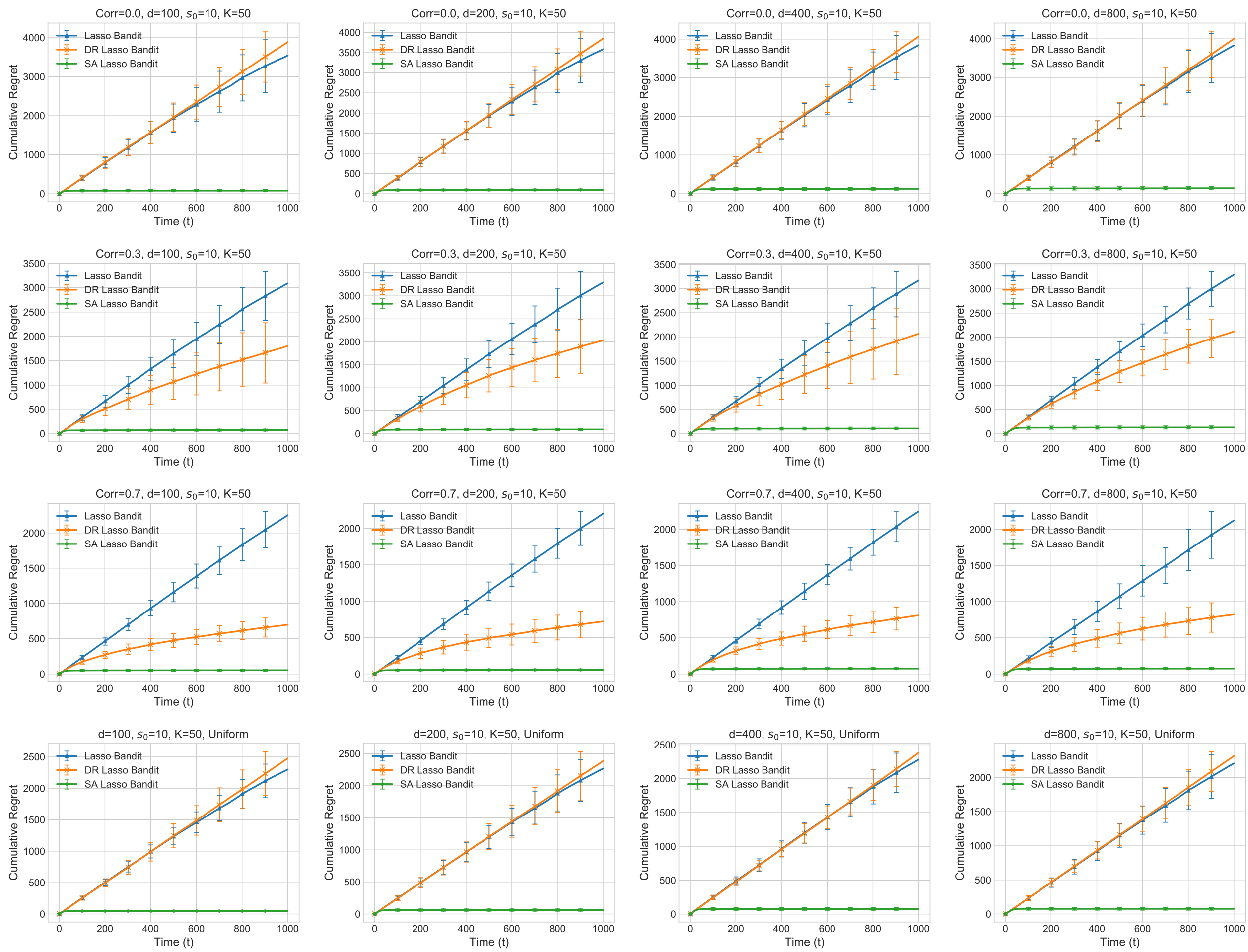

Figure 4. The plots show the t-round regret of SA LASSO BANDIT (Algorithm 1), DR LASSO BANDIT (Kim & Paik, 2019), and LASSO BANDIT (Bastani & Bayati, 2020) for $K = 50$ and $s_0 = 10$. The first three rows are the results with features drawn from a multivariate Gaussian distribution with varying levels of correlation between arms. In the fourth row, the features are drawn from a uniform distribution on a unit sphere. For each row, we present evaluations for varying feature dimensions, $d = 100, 200, 400, 800$.

condition is concerned. However, as evidenced by our experiments, there appears to be an offsetting effect, which we argue can potentially be explained by the balanced covariance condition. As for features i.i.d. from the uniform distribution (the fourth row in Figure 4) and i.i.d. from the Gaussian distribution (the first row in Figure 4), while the performance of existing algorithms (e.g., DR LASSO BANDIT from (Kim & Paik, 2019)) deteriorates significantly with the change of feature distributions (particularly without correlation between arms), SA LASSO BANDIT remains very robust, and still exhibits superior performances.

G. Related Work

Linear bandits and generalized linear bandits have been widely studied (Abe & Long, 1999; Auer, 2002; Dani et al., 2008; Rusmevichientong & Tsitsiklis, 2010; Abbasi-Yadkori et al., 2011; Filippi et al., 2010; Chu et al., 2011; Agrawal & Goyal, 2013; Li et al., 2017). However, these (generalized) linear contextual bandit strategies are not able to exploit sparse structure in the unknown parameter vector and hence may incur regret proportional to the full ambient dimension d rather than the sparse set of features of cardinality s_0. To exploit spare structure, Abbasi-Yadkori et al. (2012) propose a framework to construct high probability confidence sets for online linear prediction and establish the $\widetilde{\mathcal{O}}(\sqrt{s_0 dT})$ regret bound, where $\widetilde{\mathcal{O}}$ hides logarithmic terms. However, their algorithm needs to know the sparsity s_0. Furthermore, their algorithm is not computationally efficient; an implementable version of their framework is not yet known (Section 23.5 in (Lattimore & Szepesvári, 2019)). It is worth noting that the $\sqrt{d}$ dependence in the regret bound is unavoidable unless additional assumptions are imposed; see Theorem 24.3 in (Lattimore & Szepesvári, 2019). Gilton & Willett (2017) adapt Thompson sampling (Thompson, 1933) to sparse linear bandits; however, they also assume a priori knowledge of a small superset of the support for the parameter.

Bastani & Bayati (2020) address the contextual bandit problem with high-dimensional features. They propose a bandit algorithm which uses Lasso (Tibshirani, 1996) to estimate the parameter of each arm separately. To ensure compatibility of the empirical Gram matrices, they adapt the forced-sampling technique in (Goldenshluger & Zeevi, 2013) which is now tuned using the (a priori known) sparsity index, and is implemented for each arm at predefined time points. They establish a $\mathcal{O}(K^4 s_0^2 [\log d + \log T]^2)$ regret bound where K is the number of arms. Note that they invoke several additional assumptions: a margin condition that ensures that the density of the context distribution is bounded near the decision boundary, and arm-optimality which assumes a gap between the optimal and sub-optimal arms exists with some positive probability. In the same problem setting, Wang et al. (2018) propose an algorithm which uses forced-sampling along with the minimax concave penalty (MCP) estimator (Zhang, 2010) and improve the regret bound to $\mathcal{O}(K^3 s_0^2 [s_0 + \log d] \log T)$. Note that (Bastani & Bayati, 2020) and (Wang et al., 2018) achieve a poly-logarithmic dependence on T in regret, exploiting the arm optimality condition which assumes a gap between the optimal and sub-optimal arms exists with some probability. Since we do not assume such separability between arms, poly-logarithmic dependence on T is not attainable in our problem setting. Kim & Paik (2019) extend the Lasso bandit (Bastani & Bayati, 2020) to linear bandit settings and propose a different approach to address the non-compatibility of the empirical Gram matrices by using a doubly-robust technique (Bang & Robins, 2005) that originates with the missing data / imputation literature. They achieve $\mathcal{O}(s_0 \sqrt{T} \log(dT))$ regret.

All of the aforementioned algorithms require that the learning agent know the sparsity s_0 of the unknown parameter (or a non-trivial upper-bound on sparsity which is strictly less than d).[5] That is, only when the algorithm knows s_0, it can guarantee the regret bounds mentioned above. Otherwise, the regret bounds would scale polynomially with d instead of s_0 or potentially scale linearly with T. To our knowledge, the only work in sparse bandits which does not require this prior knowledge of sparsity is the work by Carpentier & Munos (2012) although the algorithm still requires to know the ℓ_2-norm of the unknown parameter. However, their analysis uses a non-standard definition of noise and is restricted to the case where the set of arms is the ℓ_2 unit ball and fixed over time, a structure they exploit in a significant manner, and which limits the scope of their algorithm.

[5] Besides sparsity, some algorithms require further knowledge, such as arm optimality lower bounding probability (Bastani & Bayati, 2020; Wang et al., 2018), which is also not readily available in practice.