Thresholded Lasso Bandit

Kaito Ariu
KTH
ariu@kth.se

Kenshi Abe
CyberAgent, Inc.
abe_kenshi@cyberagent.co.jp

Alexandre Proutière
KTH
alepro@kth.se

Abstract

In this paper, we revisit the regret minimization problem in sparse stochastic contextual linear bandits, where feature vectors may be of large dimension d, but where the reward function depends on a few, say $s_0 \ll d$, of these features only. We present Thresholded Lasso bandit, an algorithm that (i) estimates the vector defining the reward function as well as its sparse support, i.e., significant feature elements, using the Lasso framework with thresholding, and (ii) selects an arm greedily according to this estimate projected on its support. The algorithm does not require prior knowledge of the sparsity index s_0. For this simple algorithm, we establish non-asymptotic regret upper bounds scaling as $\mathcal{O}(\log d + \sqrt{T})$ in general, and as $\mathcal{O}(\log d + \log T)$ under the so-called margin condition (a setting where arms are well separated). The regret of previous algorithms scales as $\mathcal{O}(\log d + \sqrt{T \log(dT)})$ and $\mathcal{O}(\log T \log d)$ in the two settings, respectively. Through numerical experiments, we confirm that our algorithm outperforms existing methods.

1 Introduction

The linear contextual bandit [2, 13] is a sequential decision-making problem that generalizes the classical stochastic Multi-Armed Bandit (MAB) problem [12, 16], where (i) in each round, the decision maker is provided with a *context* in the form of a feature vector for each arm and where (ii) the expected reward is a linear function of these vectors. More precisely, at the beginning of round $t \geq 1$, the decision-maker receives for each arm k, a feature vector $A_{t,k} \in \mathbb{R}^d$. She then selects an arm, say k, and observes a sample of a random reward with mean $\langle A_{t,k}, \theta \rangle$. The parameter vector $\theta \in \mathbb{R}^d$ is fixed but initially unknown. Linear contextual bandits have been extensively applied in online services such as online advertisement and recommendation systems [13, 14, 22], and constitute arguably the most relevant structured bandit model in practice.

The major challenge in the design of efficient algorithms for contextual linear bandits stems from the high dimensionality of the feature space. For example, for display ad systems as studied in [8, 21], the joint information about a user, an ad and its publisher is encoded in a feature vector of dimension $d = 2^{24}$. Fortunately, typically only a few features significantly impact the expected reward. This observation has motivated the analysis of problems where the unknown parameter vector θ is sparse [3, 11, 15, 20]. In this paper, we also investigate sparse contextual linear bandits, and assume that θ only has at most $s_0 \ll d$ non-zero components. The set of these components and its cardinality s_0 are unknown to the decision-maker. Sparse contextual linear bandits have attracted a lot of attention recently. State-of-the-art algorithms developed to exploit the sparse structure achieve regrets scaling as $\mathcal{O}(\log d + \sqrt{T \log(dT)})$ in general, and $\mathcal{O}(\log d \log T)$ under the co-called margin condition (a setting where arms are well separated); refer to Section 2 for details.

We develop a novel algorithm, referred to as Thresholded Lasso bandit, with improved regret guarantees. Our algorithm first uses the Lasso framework with thresholding to maintain and update in each round estimates of the parameter vector θ and of its support. It then greedily picks an arm based on these estimates (the *thresholded* estimates of θ). The regret of the algorithm strongly depends on the accuracy of these estimates. We derive strong guarantees on this accuracy, which in turn leads to regret guarantees. Our contributions are as follows.

(i) Thresholded Lasso estimation performance. The performance of the Lasso-based estimation procedure is now fairly well understood, see e.g., [6, 17, 23]. For example, [23] provides an analysis of the estimation of the support of θ, and specifically, gives upper bounds on the number of false positives (components that are not in the support, but estimated as part of it) and false negatives (components that are in the support, but not estimated as part of it). These analyses, however, critically rely on the assumption that the observed data is i.i.d.. This assumption does not hold for the bandit problem, as the algorithm adapts its arm selection strategy depending on the past observations. Despite the non i.i.d.

Table 1: Comparison with existing studies. $\mathcal{O}$ notations are hiding logarithmic factors in s_0 and assuming that both T and d grow large. The 'Compatibility' and 'Margin' conditions refer to Assumptions 2 and 6.

Paper	Regret	Compatibility	Margin
Abbasi-Yadkori et al. (2012) [1]	$\mathcal{O}(ds_0(\log T)^2)$		
Bastani and Bayati (2020) [4]	$\mathcal{O}(s_0^2(\log d + \log T)^2)$	✓	✓
Wang et al. (2018) [20]	$\mathcal{O}(s_0^2(\log d + s_0)\log T)$	✓	✓
Kim and Paik (2019) [11]	$\mathcal{O}(s_0\sqrt{T}\log(dT))$	✓	
Oh et al. (2020) [15]	$\mathcal{O}(s_0^2\log d + s_0\sqrt{T\log(dT)})$	✓	
This work: Theorem 1	$\mathcal{O}(s_0^2\log d + s_0\log T)$	✓	✓
This work: Theorem 2	$\mathcal{O}(s_0^2\log d + \sqrt{s_0T})$	✓	

nature of the data, we manage to derive performance guarantees of the estimate of θ. In particular, we establish high probability guarantees that are independent of the dimension d (see Lemma 3).

(ii) Regret guarantees. Based on the analysis of the Thresholded Lasso estimation procedure, we provide a finite-time analysis of the regret of our algorithm. The regret scales at most as $\mathcal{O}(\log d + \sqrt{T})$ in general and $\mathcal{O}(\log d + \log T)$ under the margin condition. More precisely, the estimation error of θ induces a regret scaling as $\mathcal{O}(\sqrt{s_0T})$ (or $\mathcal{O}(s_0\log T)$ under the margin condition). The additional term $\mathcal{O}(\log d)$ in our regret upper bounds comes from the errors made when estimating the support of θ. It is worth noting that when using the plain Lasso estimator (without thresholding), one would obtain weaker performance guarantees for the estimation of θ, typically depending on d, see e.g., [3]. This dependence causes an additional multiplicative term $\log d$ in the regret.

(iii) Numerical experiments. We present extensive numerical experiments to illustrate the performance of the Thresholded Lasso bandit algorithm. We compare the estimation accuracy for θ and the regret of our algorithm to those of the Doubly-Robust Lasso bandit and Sparsity-Agnostic Lasso bandit algorithms [11, 15]. These experiments confirm the benefit of the use of the Lasso procedure with thresholding.

2 Related work

Stochastic linear bandit problems have attracted a lot of attention over the last decade. [7] addresses sparse linear bandits where $\|\theta\|_0 \le s_0$ and where the set of arms is restricted to the ℓ_2 unit ball. For regimes where the time horizon is much smaller than the dimension d, i.e., $T \ll d$, the authors propose an algorithm whose regret scales at most as $\mathcal{O}(s_0\sqrt{T}\log(dT))$. [10] studied high-dimensional linear bandit problems where the number of actions is larger than or equal to d. Under some signal strength conditions, they propose an algorithm achieves regret of $\mathcal{O}(s_0\log d + \sqrt{s_0T\log(dT)})$. Those studies, however, do not consider problems with contextual information.

Recently, high-dimensional contextual linear bandits have been investigated under the sparsity assumption $\|\theta\|_0 \le s_0$. In this line of research, the decision-maker is provided in each round with a set of arms defined by a finite set of feature vectors. This set is uniformly bounded across rounds. In this setting, the authors of [1] devise an algorithm with both a minimax (problem independent) regret upper bound $\widetilde{\mathcal{O}}(\sqrt{s_0dT})$ and problem dependent upper bound $\mathcal{O}(ds_0(\log T)^2)$ (the notation $\widetilde{\mathcal{O}}$ hides the polylogarithmic terms) without any assumption on the distribution (other than the assumptions similar to our Assumption 1).

In [4] (initially published in 2015 as [3]), the authors address a high-dimensional contextual linear bandit problem where the unknown parameter defining the reward function is arm-specific (θ is different for the various arms). In the proposed algorithm, arms are explored uniformly at random for $\mathcal{O}(s_0^2\log d\log T)$ prespecified rounds. Under the margin condition, similar to our Assumption 6, the algorithm achieves a regret of $\mathcal{O}(s_0^2(\log d + \log T)^2)$. For the same problem, [20] develops the so-called MCP-Bandit algorithm. The latter also uses the uniform exploration for $\mathcal{O}(s_0^2\log d\log T)$ prespecified rounds, and has improved regret guarantees: the regret scales as $\mathcal{O}(s_0^2(\log d + s_0)\log T)$.

High-dimensional contextual linear bandits have been also studied without the margin condition, but with a unique parameter θ defining the reward function. [11] designs an algorithm with uniform exploration phases of $\mathcal{O}(\sqrt{T\log(dT)\log T})$ rounds, and with regret $\mathcal{O}(s_0\sqrt{T}\log(dT))$. All the aforementioned algorithms require the knowledge of the sparsity index s_0. In [15], the authors propose an algorithm, referred to as SA Lasso bandit, that does not require this knowledge, and with regret $\mathcal{O}(s_0^2 \log d + s_0\sqrt{T\log(dT)})$. In addition, SA Lasso bandit does not include any uniform exploration phase. Its regret guarantees are derived under specific assumptions on the context distribution, the so-called relaxed symmetry assumption and the balanced covariance assumption. The authors also establish a $\mathcal{O}(s_0^2 \log d + \sqrt{s_0 T\log(dT)})$ regret upper bound under the so-called *restricted eigenvalue* condition.

In this paper, we develop an algorithm with improved regret guarantees with and without the margin condition. The algorithm does not rely on the knowledge of s_0. We have summarized the relevant studies and our work in Table 1. We will come back to this table when we discuss the assumptions.

3 Model and assumptions

3.1 Model and notation

We consider a contextual linear stochastic bandit problem in a high-dimensional space. In each round $t \in [T] := \{1, \dots, T\}$, the algorithm is given a set of context vectors $\mathcal{A}_t = \{A_{t,k} \in \mathbb{R}^d : k \in [K]\}$. The successive sets $(\mathcal{A}_t)_{t\geq 1}$ form an i.i.d. sequence with distribution p_A. In round t, the algorithm selects an arm $A_t \in \mathcal{A}_t$ based on past observations, and collects a random reward r_t. Formally, if $\mathcal{F}_t$ is the σ-algebra generated by random variables $(\mathcal{A}_1, A_1, r_1, \dots, \mathcal{A}_{t-1}, A_{t-1}, r_{t-1}, \mathcal{A}_t)$, A_t is $\mathcal{F}_t$-measurable. We assume that $r_t = \langle A_t, \theta\rangle + \varepsilon_t$, where ε_t is a zero mean sub-Gaussian random variable with variance proxy σ^2 given $\mathcal{F}_t$ and A_t. Our objective is to devise an algorithm with minimal regret, where regret is defined as:

$$R(T) := \mathbb{E}\left[\sum_{t=1}^{T} \max_{A\in\mathcal{A}_t} \langle A, \theta\rangle - r_t\right] = \mathbb{E}\left[\sum_{t=1}^{T} \max_{A\in\mathcal{A}_t} \langle A - A_t, \theta\rangle\right].$$

Notation. The ℓ_0-norm of a vector $\theta \in \mathbb{R}^d$ is $\|\theta\|_0 = \sum_{i=1}^d \mathbb{1}\{\theta_i \neq 0\}$. We denote $\hat{\Sigma}_t = \frac{1}{t}\sum_{s=1}^t A_s A_s^\top$ as the empirical Gram matrix generated by the arms selected under a specific algorithm. For any $B \subset [d]$, we define $\theta_B := (\theta_{1,B}, \dots, \theta_{d,B})^\top$ where for all $i \in [d]$, $\theta_{i,B} := \theta_i \mathbb{1}\{i \in B\}$. For each $B \subset [d]$, we define the submatrix $A(B) \in \mathbb{R}^{n\times|B|}$ of $A \in \mathbb{R}^{n\times d}$ where for $A(B)$, we extract the rows that are in B. We denote $\mathrm{supp}(x)$ as the set of the non-zero element indices of $x \in \mathbb{R}^d$. We also define $\theta_{\min}$ as the minimal value of $|\theta_i|$ on the support: $\theta_{\min} := \min_{i\in\mathrm{supp}(\theta)} |\theta_i|$.

3.2 Assumptions

We present a set of assumptions made throughout the paper. We also discuss and compare these assumptions to those made in the related literature.

Assumption 1 (Sparsity and parameter constraints) *The parameter θ defining the reward function is sparse, i.e., $\|\theta\|_0 \leq s_0$ for some fixed but unknown integer s_0 (s_0 does not depend on d). We denote by $S = \{i \in [d] : \theta_i \neq 0\}$ the support of θ. We further assume that $\|\theta\|_2 \leq s_2$ for some unknown constant s_2. Finally, we assume that for any subset S' of $[d]$ such that $|S'| \leq \mathcal{O}(s_0)$, the ℓ_2-norm of the context vector is bounded: for all t and for all $A \in \mathcal{A}_t$, for all $S' \subset [d]$ s.t. $|S'| \leq s_0 + \frac{4\nu C_\mathrm{b}\sqrt{s_0}}{\phi_0^2}$, $\|A_{S'}\|_2 \leq s_A$, where $s_A > 0$ is a constant and $\phi_0 > 0$, $\nu \geq 1$, $C_\mathrm{b} > 1$ are constants defined in Assumption 2, 3, and 4.*

Assumption 2 (Compatibility condition) *For a matrix $M \in \mathbb{R}^{d\times d}$ and a set $S_0 \subset [d]$, we define the compatibility constant $\phi(M, S_0)$ as:*

$$\phi^2(M, S_0) := \min_{x:\|x_{S_0}\|_1\neq 0} \left\{\frac{s_0 x^\top M x}{\|x_{S_0}\|_1^2} : \|x_{S_0^c}\|_1 \leq 3\|x_{S_0}\|_1\right\}.$$

We assume that for the Gram matrix of the action set $\Sigma := \frac{1}{K}\sum_{k=1}^K \mathbb{E}_{\mathcal{A}\sim p_A}\left[A_k A_k^\top\right]$ satisfies $\phi^2(\Sigma, S) \geq \phi_0^2$, where $\phi_0 > 0$ is some positive constant.

The compatibility condition was introduced in the high-dimensional statistics literature [6]. It ensures that the Lasso estimate [17] of the parameter θ approaches to its true value as the number of samples grows large. Note that it is easy to check that the compatibility condition is strictly weaker than assuming the positive definiteness of Σ. It allows us to consider feature vectors with strongly correlated components. Assumption 2 is considered to be essential for the Lasso estimate to be consistent and assumed in many of the relevant studies. See Table 1 for studies using the compatibility condition.

Assumption 3 (Relaxed symmetry [15]) *For the distribution p_A of $\mathcal{A}$, there exists a constant $\nu \geq 1$ such that for all $\boldsymbol{A} \in \mathbb{R}^{K\times d}$ such that $p_A(\boldsymbol{A}) > 0$, $\frac{p_A(\boldsymbol{A})}{p_A(-\boldsymbol{A})} \leq \nu$.*

Assumption 4 (Balanced covariance [15]) *For any permutation γ of $[K]$, for any integer $k \in \{2, ..., K-1\}$ and a fixed θ, there exists a constant $C_{\mathrm{b}} > 1$ such that*

$$C_{\mathrm{b}}\mathbb{E}_{\mathcal{A}\sim p_A}\left[(A_{\gamma(1)}A_{\gamma(1)}^\top + A_{\gamma(K)}A_{\gamma(K)}^\top)\mathbb{1}\{\langle A_{\gamma(1)}, \theta\rangle < \ldots < \langle A_{\gamma(K)}, \theta\rangle\}\right]$$
$$\succeq \mathbb{E}_{\mathcal{A}\sim p_A}\left[A_{\gamma(k)}A_{\gamma(k)}^\top \mathbb{1}\{\langle A_{\gamma(1)}, \theta\rangle < \ldots < \langle A_{\gamma(K)}, \theta\rangle\}\right].$$

Assumption 5 (Sparse positive definiteness) *Define for each $B \subset [d]$, $\Sigma_B := \frac{1}{K}\sum_{k=1}^K \mathbb{E}_{\mathcal{A}\sim p_A}\left[A_k(B)A_k(B)^\top\right]$, where $A_k(B)$ is a $|B|$-dimensional vector extracted the elements of A_k with indices in B. There exists a positive constant $\alpha > 0$ such that $\forall B \subset [d]$,*

$$\left(|B| \leq s_0 + (4\nu C_{\mathrm{b}}\sqrt{s_0})/\phi_0^2 \text{ and } S \subset B\right) \Longrightarrow \left(\min_{v\in\mathbb{R}^d : \|v\|_2 = 1} v^\top \Sigma_B v \geq \alpha\right).$$

The parameters $\phi_0, \nu, C_{\mathrm{b}}$ are those of Assumptions 2, 3, 4.

Assumption 3 comes from [15]. This assumption is satisfied for the wide range of distributions including multivariate Gaussian, uniform, and Bernoulli distributions. Assumption 4 is also adopted from [15]. This assumption holds for a wide range of distributions including multivariate Gaussian distribution, uniform distribution on sphere, an arbitrary independent distribution for each arm [15]. Assumption 5 implies that the context distribution is diverse enough in the neighborhood of the support of θ. Note that Assumption 5 is standard in low dimensional linear bandit literature. There, if $S = [d]$, the only choice for B is $[d]$, and the set of action has to span $\mathbb{R}^d$ (hence Assumption 5 is satisfied). We will see that after an accurate estimate of the support S (Lemma 1), Assumption 5 is used only to analyze the performance of the least square estimator of low-dimensional (order of $\mathcal{O}(s_0)$) vector. Assumption 5 is strictly weaker than the covariate diversity condition of [5], where the positive definiteness must be guaranteed for the Gram matrix generated by the greedy algorithms.

4 Algorithm

In this section, we present the Thresholded (TH) Lasso bandit algorithm. The algorithm adapts the idea of Lasso with thresholding proposed in [23] to estimate θ and its support. The main challenge in the analysis of the Lasso with thresholding stems from the fact that here, the data is non i.i.d. (the arm selection is adaptive).

Algorithm 1 TH Lasso Bandit

Require: λ_0

for $t = 1, \cdots, T$ **do**

  Receive a context set $\mathcal{A}_t := \{A_{t,k} : k \in [K]\}$

  Pull arm $A_t = \underset{A\in\mathcal{A}_t}{\mathrm{argmax}}\langle A, \hat{\theta}_t\rangle$ (ties are broken uniformly at random) and observe r_t

  $\lambda_t \leftarrow \lambda_0\sqrt{\frac{2\log t \log d}{t}}, \quad A \leftarrow (A_1, A_2, \ldots, A_t)^\top, \quad R \leftarrow (r_1, r_2, \ldots, r_t)^\top$

  $\hat{\theta}_0^{(t)} \leftarrow \underset{\theta}{\mathrm{argmin}}\ \frac{1}{t}\|R - A\theta\|_2^2 + \lambda_t\|\theta\|_1$

  $\hat{S}_0^{(t)} \leftarrow \{j \in [d] : |(\hat{\theta}_0^{(t)})_j| > 4\lambda_t\}, \quad \hat{S}_1^{(t)} \leftarrow \{j \in \hat{S}_0^{(t)} : |(\hat{\theta}_0^{(t)})_j| \geq 4\lambda_t\sqrt{|\hat{S}_0^{(t)}|}\}$

  $A_S \leftarrow (A_{1,\hat{S}_1^{(t)}}, A_{2,\hat{S}_1^{(t)}}, \ldots, A_{t,\hat{S}_1^{(t)}})^\top$

  $\hat{\theta}_{t+1} \leftarrow \underset{\theta}{\mathrm{argmin}}\ \|R - A_S\theta\|_2^2$

end for

The pseudo-code of our algorithm is presented in Algorithm 1. In round t, the algorithm pulls the arm in a greedy way using the estimated value $\hat{\theta}_t$ of θ. From the past selected arms and rewards, we get via the Lasso a first estimate $\hat{\theta}_0^{(t)}$ of θ. This estimate is then used to estimate the support of θ using appropriate thresholding. The regularizer $\lambda_t := \lambda_0\sqrt{(2\log t \log d)/t}$ is set at a much larger value than that in the previous work (they typically have the order of $\sqrt{(\log d + \log t)/t}$), as we are only focusing on the support recovery here. Note that we apply a thresholding procedure twice to $\hat{\theta}_0^{(t)}$ to provide the support estimate $\hat{S}_1^{(t)}$. The final estimate $\hat{\theta}_{t+1}$ is obtained as the Least-Squares estimator of θ, when restricted to $\hat{S}_1^{(t)}$. The initial support estimate done by Lasso contains too many false positives. By including thresholding steps in the algorithm, we remove the unnecessary false positives and improve the support estimate. We quantify this improvement in the next section.

5 Performance guarantees

We analyze the regret of the Thresholded Lasso bandit algorithm both when the margin condition holds and when it does not.

5.1 With the margin condition

Assumption 6 (Margin condition) *There exists a constant $C_{\mathrm{m}} > 0$ such that for all $\kappa > 0$,*

$$\forall k \neq k', \quad \mathbb{P}_{A\sim p_A}(0 < |\langle A_k - A_{k'}, \theta\rangle| \leq \kappa) \leq C_{\mathrm{m}}\kappa.$$

The margin condition controls the probability under p_A that two arms yield very similar rewards (and hence are hard to separate). It was first introduced in [9] in the (low-dimensional) linear bandit literature. The margin condition holds for the most usual context distributions (including the uniform distribution and Multivariate Gaussian distributions).

The following theorem provides a non-asymptotic regret upper bound of TH Lasso bandit under the margin condition. To simplify the presentation of our regret upper bound, we use the following quantities, functions of the parameters involved in Assumptions 1, 2, and 5. Define $h_0 = \lfloor \sqrt{\log(4(s_0 + \frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}))+1} \rfloor$ and $\tau = \max\left(\lfloor \frac{2\log(2d^2)}{C_0^2} \rfloor, \lfloor (\log\log d)\log d \rfloor\right)$, where $C_0 = \min\left\{\frac{1}{2}, \frac{\phi_0^2}{512 s_0 s_A^2 \nu C_{\mathrm{b}}}\right\}$.

Theorem 1 *Set $\lambda_0 = 4\sigma s_A\sqrt{c}$ with $c > 0$ such that $4\left(\frac{4\nu C_{\mathrm{b}} s_0}{\phi_0^2} + \sqrt{\left(1 + \frac{4\nu C_{\mathrm{b}}}{\phi_0^2}\right) s_0}\right)\lambda_\tau \leq \theta_{\min}$. Assume $\tau \geq \exp(4/c)$. Under Assumptions 1, 2, 3, 4, 5, and 6, we have:*

$$R(T) \leq 2s_A s_2 \tau + \frac{1408\sigma^2 s_A^4 C_{\mathrm{m}}(K-1)h_0^3\nu^2 C_{\mathrm{b}}^2\left(s_0 + \frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}\right)}{\alpha^2}(1+\log T)$$
$$+ 2(K-1)s_A s_2\left(\frac{\pi^2}{3} + \frac{2}{C_0^2} + \left(s_0 + \frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}\right)\frac{20 s_A^2\nu C_{\mathrm{b}}}{\alpha}\right).$$

In particular, we have, as d and T grows large, $R(T) = \mathcal{O}(s_0^2\log d + s_0\log T + s_0^2)$, where the $\mathcal{O}(\cdot)$ notation hides all the multiplicative and additive constants that do not depend on s_0, d, and T and hides $\log\log d$ and $(\log(s_0))^{\frac{3}{2}}$ multiplicative factors for the first and second term. Moreover, when $\lambda_0 = 1/(\log\log d)^{\frac{1}{4}}$ and d is large enough so that (i) $\log\log d \geq 4096\sigma^4 s_A^4$, (ii): $4\left(\frac{4\nu C_{\mathrm{b}} s_0}{\phi_0^2} + \sqrt{\left(1 + \frac{4\nu C_{\mathrm{b}}}{\phi_0^2}\right) s_0}\right)/(\theta_{\min}(\log\log d)^{\frac{1}{4}})\sqrt{2\log\tau\log d/\tau} \leq 1$, and (iii) $d \geq 100$, as d and T grow large,

$$R(T) = \mathcal{O}(s_0^2\log d + s_0\log T + s_0^2).$$

5.2 Without the margin condition

Theorem 2 *Set $\lambda_0 = 4\sigma s_A\sqrt{c}$ with $c > 0$ such that $4\left(\frac{4\nu C_{\mathrm{b}} s_0}{\phi_0^2} + \sqrt{\left(1 + \frac{4\nu C_{\mathrm{b}}}{\phi_0^2}\right) s_0}\right)\lambda_\tau \leq \theta_{\min}$. Assume $\tau \geq \exp(4/c)$. Under Assumptions 1, 2, 3, 4, and 5, we have:*

$$R(T) \leq 2s_A s_2 \tau + \frac{36\sigma s_A(K-1)h_0^2\nu C_{\mathrm{b}}\sqrt{8\left(s_0 + \frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}\right)}}{\alpha}\sqrt{T}$$
$$+ 2(K-1)s_A s_2\left(\frac{\pi^2}{3} + \frac{2}{C_0^2} + \left(s_0 + \frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}\right)\frac{20 s_A^2\nu C_{\mathrm{b}}}{\alpha}\right).$$

In particular, we have, as d and T grows large, $R(T) = \mathcal{O}(s_0^2\log d + \sqrt{s_0 T} + s_0^2)$, where the $\mathcal{O}(\cdot)$ notation hides all the multiplicative and additive constants that do not depend on s_0, d, and T and hides $\log\log d$

and $\log s_0$ *multiplicative factors for the first and second term respectively.*
Moreover, when $\lambda_0 = 1/(\log\log d)^{\frac{1}{4}}$ *and* d *is large enough so that the conditions (i), (ii), and (iii) of Theorem 1 are satisfied. As* d *and* T *grow large,*

$$R(T) = \mathcal{O}(s_0^2 \log d + \sqrt{s_0 T} + s_0^2).$$

Theorems 1 and 2 state that TH Lasso bandit achieves much lower regret than existing algorithms. Indeed, upper regret bounds for the latter had a term scaling as $\log d \log T$ (resp. $\log d + \sqrt{T\log(dT)}$) with (resp. without) the margin condition. TH Lasso bandit removes the $\log d$ and $\log T$ multiplicative factors. In most applications of the sparse linear contextual bandit, both T and d are typically very large, and the regret improvement obtained by TH Lasso bandit is significant. Note that the dependency in s_0 has been also improved from linear to square root compared with the analysis of Theorem 1 and 3 in SA Lasso bandit [15]. This stems from the fact that using the compatibility condition, one can only control the ℓ_1 norm of the estimation error of θ, while using the OLS leads to an ℓ_2 guarantee (see Lemma 3).

5.3 Sketch of the proof of theorems

We sketch below the proof of Theorem 1 and 2. A complete proof is presented in appendix.

(1) Estimation of the support of θ. First, we prove that the estimated support contains the true support S with high probability.

Lemma 1 *Let* $t \geq \frac{2\log(2d^2)}{C_0^2}$ *such that* $4\left(\frac{4\nu C_{\mathrm{b}} s_0}{\phi_0^2} + \sqrt{\left(1+\frac{4\nu C_{\mathrm{b}}}{\phi_0^2}\right)s_0}\right)\lambda_t \leq \theta_{\min}$. *Under Assumptions 1, 2, 3, and 4,*
$\mathbb{P}\left(S \subset \hat{S}_1^{(t)} \text{ and } |\hat{S}_1^{(t)} \setminus S| \leq \frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}\right) \geq 1 - 2\exp\left(-\frac{t\lambda_t^2}{32\sigma^2 s_A^2} + \log d\right) - \exp\left(-\frac{tC_0^2}{2}\right).$

Lemma 1 extends the support recovery result of the Thresholded Lasso [23] to the case of non-i.i.d data (generated by the bandit algorithm). The dependence on s_0 is analogous to the offline result Theorem 3.1 of [23]. Define the event $\mathcal{E}_t$ as:

$$\mathcal{E}_t := \left\{S \subset \hat{S}_1^{(t)} \text{ and } |\hat{S}_1^{(t)} \setminus S| \leq \frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}\right\}.$$

In the remaining of the proof, in view of Lemma 1, we can assume that the event $\mathcal{E}_t$ holds.
(2) Minimal eigenvalue of the empirical Gram matrix. We write $\hat{S}_1^{(t)} = \hat{S}$ for the simplicity. Let $\hat{\Sigma}_{\hat{S}} := \frac{1}{t}\sum_{s=1}^t A_s(\hat{S})A_s(\hat{S})^\top$ be the empirical Gram matrix on the estimated support. We prove that the positive definiteness of the empirical Gram matrix on the estimated support is guaranteed.

Lemma 2 *Let* $t \in [T]$. *Under Assumptions 1 and 5, we have:*

$$\mathbb{P}\left(\lambda_{\min}(\hat{\Sigma}_{\hat{S}}) \geq \frac{\alpha}{4\nu C_{\mathrm{b}}} \mid \mathcal{E}_t\right) \geq 1 - \exp\left(\log\left(s_0 + \frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}\right) - \frac{t\alpha}{20 s_A^2 \nu C_{\mathrm{b}}}\right).$$

(3) Estimation of θ after thresholding. Next, we study the accuracy of $\hat{\theta}_t$.

Lemma 3 *Let* $t \in [T]$ *and* $s' = s_0 + 4\nu C_{\mathrm{b}}\sqrt{s_0}/\phi_0^2$. *Under Assumption 1, we have for all* $x, \lambda > 0$*:*

$$\mathbb{P}\left(\|\hat{\theta}_{t+1} - \theta\|_2 \geq x \text{ and } \lambda_{\min}(\hat{\Sigma}_{\hat{S}}) \geq \lambda \mid \mathcal{E}_t\right) \leq 2s' \exp\left(-\frac{\lambda^2 t x^2}{2\sigma^2 s_A^2 s'}\right).$$

From the above lemma, we conclude that θ is well estimated with high probability.

(4) Instantaneous regret upper bound with Margin condition. For the previous lemmas, we can derive an upper bound on the instantaneous regret with Margin condition.

Lemma 4 *Define* $\mathcal{G}_t^{\frac{\alpha}{4\nu C_{\mathrm{b}}}} := \left\{\lambda_{\min}(\hat{\Sigma}_{\hat{S}}) \geq \frac{\alpha}{4\nu C_{\mathrm{b}}}\right\}$. *Let* $t \geq 2$. *Under Assumptions 1, 2, 3, 4, 5, and 6, the expected instantaneous regret satisfies:* $\mathbb{E}[\max_{A\in\mathcal{A}_t}\langle A - A_t, \theta\rangle] \leq \frac{1408\sigma^2 s_A^4 C_{\mathrm{m}}(K-1)h_0^3\nu^2 C_{\mathrm{b}}^2\left(s_0 + \frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}\right)}{\alpha^2}\frac{1}{t-1} +$ $2(K-1)s_A s_2\left(\mathbb{P}(\mathcal{E}_t^c) + \mathbb{P}\left(\left(\mathcal{G}_t^{\frac{\alpha}{4\nu C_{\mathrm{b}}}}\right)^c \middle| \mathcal{E}_t\right)\right).$

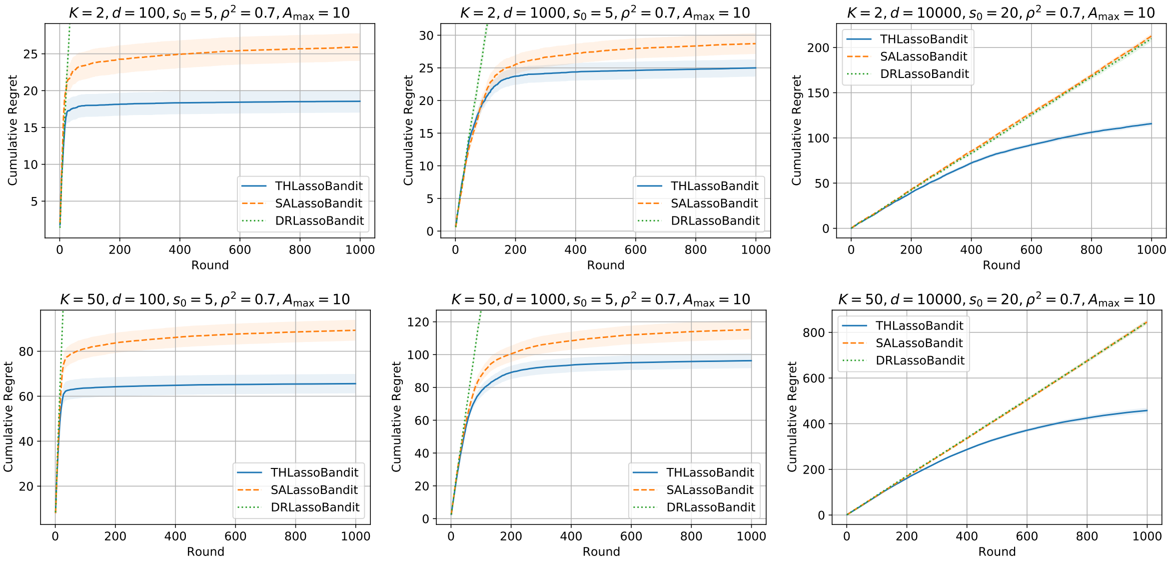

Figure 1: Cumulative regret of the three algorithms with $\rho^2 = 0.7$, $A_{\max} = 10$ in six scenarios selected using $K \in \{2, 50\}$, $d \in \{100, 1000, 10000\}$, and $s_0 \in \{5, 20\}$. The shaded area represents the standard errors.

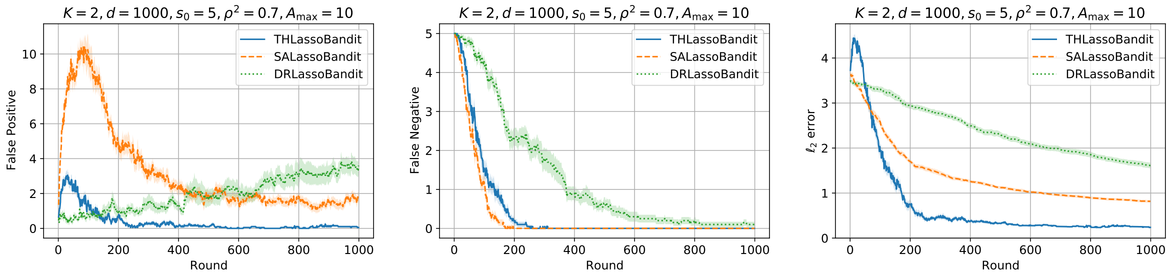

Figure 2: (Left) Number of false positives $|\hat{S}_1^{(t)} \setminus S|$, (center) false negatives $|S \setminus \hat{S}_1^{(t)}|$, (right) ℓ_2-norm error $\|\hat{\theta}_t - \theta\|_2$ of the three algorithms with $\rho^2 = 0.7$, $A_{\max} = 10$, $K = 2$, $s_0 = 5$, and $d = 1000$. The shaded area represents the standard errors.

On the other hand, without Margin condition, we also remark the key lemma, which is proven by discretization technique:

Lemma 5 (Without Margin condition) *Under Assumptions 1, 2, 3, 4, and 5, for any* $t \in [T]$, $\mathbb{E}[\max_{A \in \mathcal{A}_t} \langle A - A_t, \theta \rangle]$ *is upper bounded by*

$$\frac{36\sigma s_A (K-1) h_0^2 \nu C_{\mathrm{b}}}{\alpha} \sqrt{\frac{2\left(s_0 + \frac{4\nu C_{\mathrm{b}} \sqrt{s_0}}{\phi_0^2}\right)}{t-1}} + 2(K-1) s_A s_2 \left(\mathbb{P}(\mathcal{E}_t^c) + \mathbb{P}\left(\left(\mathcal{G}_t^{\frac{\alpha}{4\nu C_{\mathrm{b}}}} \right)^c \middle| \mathcal{E}_t \right) \right).$$

By summing up these instantaneous regret bounds, we get Theorem 1 and 2. Furthermore, in Appendix, we also provide regret analysis without Assumption 4 but when $K = 2$.

6 Experiments

In this section, we empirically evaluate the TH Lasso bandit algorithm. We compare its performance to those of the Doubly-Robust (DR) Lasso bandit [11] and SA Lasso bandit [15] algorithms. Note that we do not compare the proposed algorithm to the Lasso bandit algorithm [4], because this algorithm is designed for a problem different than ours (θ varies across arms in their setting).

Reward parameter and contexts. We consider problems where $\theta \in \mathbb{R}^d$ is sparse, i.e., $\|\theta\|_0 = s_0$. We generate each non-zero components of θ in an i.i.d. manner using the uniform distribution on $[1, 2]$. In each round t, for each component $i \in [d]$, we sample $((A_{t,1})_i, \cdots, (A_{t,K})_i)^\top \in \mathbb{R}^K$ from a Gaussian distribution $\mathcal{N}(\mathbf{0}_K, V)$ where $V_{i,i} = 1$ for all $i \in [K]$ and $V_{i,k} = \rho^2 = 0.7$ for all $i \neq k \in [K]$. We then

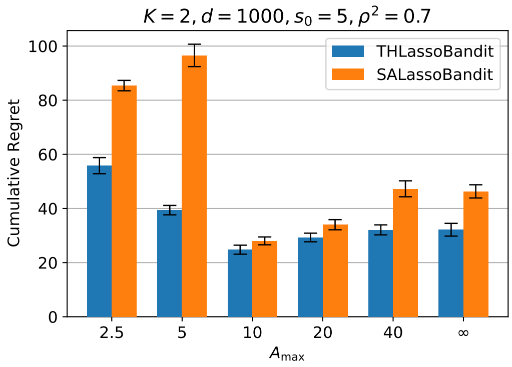

Figure 3: Cumulative regret at round $t = 1000$ of TH Lasso bandit and SA Lasso bandit with $\rho^2 = 0.7$, $K = 2$, $d = 1000$, $s_0 = 5$, and varying $A_{\max} \in \{2.5, 5, 10, 20, 40, \infty\}$. The error bars represent the standard errors.

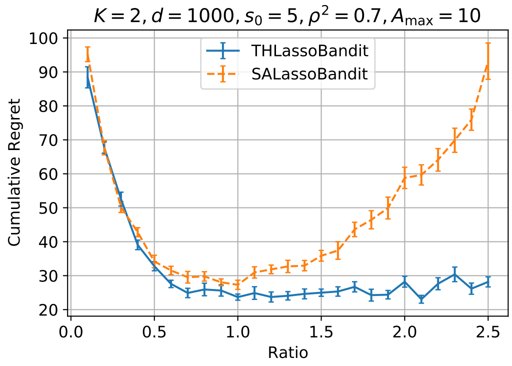

Figure 4: Cumulative regret at round $t = 1000$ of TH Lasso bandit and SA Lasso bandit with $\rho^2 = 0.7$, $A_{\max} = 10$, $K = 2$, $d = 1000$, $s_0 = 5$, and varying $\lambda_0/\lambda^* \in [0.1, 2.5]$. The error bars represent the standard errors.

normalize each $A_{t,k} = ((A_{t,k})_1, \ldots, (A_{t,k})_d)^\top \in \mathbb{R}^d$ so that its ℓ_2-norm is at most $A_{\max}$ for all $k \in [K]$. Note that the components of the feature vectors are correlated over $[d]$ and over $[K]$. The noise process is Gaussian, i.i.d. over rounds: $\varepsilon_t \sim \mathcal{N}(0, 1)$. We test the algorithms for different values of K, d, s_0, and $A_{\max}$. For each experimental setting, we averaged the results for 20 instances. We also provide additional experimental results with non-Gaussian distributions in the appendix.

Algorithms. For DR Lasso bandit, we use the tuned hyperparameter at `https://github.com/gisoo1989/Doubly-Robust-Lasso-Bandit`. For the SA Lasso bandit and TH Lasso bandit algorithms, we tune the hyperparameter λ_0 in $[0.01, 0.5]$ to roughly optimize the algorithm performance when $K = 2, d = 1000, A_{\max} = 10$, and $s_0 = 5$. As a result, we set $\lambda_0 = 0.16$ for SA Lasso bandit, and set $\lambda_0 = 0.03$ for TH Lasso bandit.

Results. We first compare the regret of each algorithm with $A_{\max} = 10$, $K \in \{2, 50\}$, $d = \{100, 1000, 10000\}$, and $s_0 \in \{5, 20\}$. We experimented with larger values of d, in addition to the one in existing studies. Figure 1 shows the average cumulative regret for each algorithm. We find that TH Lasso bandit outperforms the other algorithms in all scenarios. We provide additional experimental results, including experiments with different correlation levels $\rho^2 (\in \{0, 0.3\})$ and dimension d, in the appendix.

Next, we compare the estimation accuracy for θ under the three algorithms in the scenario: $K = 2, d = 1000, A_{\max} = 10, s_0 = 5$. Figure 2 shows the number of false positives $|\hat{S}_1^{(t)} \setminus S|$, the number of false negatives $|S \setminus \hat{S}_1^{(t)}|$, and ℓ_2-norm error $\|\hat{\theta}_t - \theta\|_2$. Note that, for DR Lasso bandit and SA Lasso bandit, we define the estimated support as $\hat{S}_1^{(t)} = \{i \in [d] : \hat{\theta}_{t,i} \neq 0\}$. We can observe that the number of false positives of our algorithm converge to zero faster than those of DR Lasso bandit and SA Lasso bandit. Furthermore, our algorithm yields a smaller estimation error $\|\hat{\theta}_t - \theta\|_2$ than the two other algorithms, as is shown in right column of Figure 2.

We also conduct experiments varying $A_{\max} \in \{2.5, 5, 10, 20, 40, \infty\}$. As in the previous experiments, for each $A_{\max}$, we normalize each feature vector $A_{t,k}$ so that its ℓ_2-norm is at most $A_{\max}$ for all $k \in [K]$. We set $K = 2, d = 1000$, and $s_0 = 5$. Figure 3 shows the average cumulative regret at $t = 1000$ of TH Lasso bandit and SA Lasso bandit for each $A_{\max}$. This experiment confirms that TH Lasso exhibits lower regret than SA Lasso bandit.

Finally, we examine the robustness of TH Lasso bandit and SA Lasso bandit with respect to the hyperparameter λ_0. We vary $\lambda_0 \in [0.1\lambda^*, 2.5\lambda^*]$ where $\lambda^* = 0.03$ for TH Lasso bandit and $\lambda^* = 0.16$ for SA Lasso bandit. We set $K = 2, d = 1000, s_0 = 5$, and $A_{\max} = 10$. Figure 4 shows the average cumulative regret at $t = 1000$ for TH Lasso bandit and SA Lasso bandit for different ratios λ_0/λ^*. Observe that the regret of TH Lasso bandit is lower than that of SA Lasso bandit for various ratios. Moreover, note the performance of TH Lasso bandit is not very sensitive to the choice of λ_0: it is robust. This contrasts with the SA Lasso bandit algorithm, for which a careful tuning of λ_0 is needed to get good performance.

7 Conclusion

In this paper, we studied the high-dimensional contextual linear bandit problem with sparsity. We devised TH Lasso bandit, a simple algorithm that applies a Lasso procedure with thresholding to estimate the support of the unknown parameter. We established finite-time regret upper bounds under various assumptions, and in particular with and without the margin condition. These bounds exhibit a better regret scaling than those derived for previous algorithms. We also numerically compared TH Lasso bandit to previous algorithms in a variety of settings, and showed that it outperformed other algorithms in these settings.

In future work, it would be interesting to consider scenarios where the assumptions made in this paper may not hold. In particular, it is worth investigating the case where the relaxed symmetry condition (Assumption 3) is not satisfied. In this case, being greedy in the successive arm selections may not work. It is intriguing to know whether devising an algorithm without forced uniform exploration and with reasonable regret guarantees is possible.

References

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

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

[3] H. Bastani and M. Bayati. Online decision-making with high-dimensional covariates. *Available at SSRN 2661896*, 2015.

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

[5] H. Bastani, M. Bayati, and K. Khosravi. Mostly exploration-free algorithms for contextual bandits. *Management Science, Articles in Advance*, 2020.

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

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

[8] O. Chapelle and L. Li. An empirical evaluation of thompson sampling. In *Advances in Neural Information Processing Systems*, pages 2249–2257, 2011.

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

[10] B. Hao, T. Lattimore, and M. Wang. High-dimensional sparse linear bandits. In *Advances in Neural Information Processing Systems*, 2020.

[11] G.-S. Kim and M. C. Paik. Doubly-robust lasso bandit. In *Advances in Neural Information Processing Systems*, pages 5877–5887, 2019.

[12] T. L. Lai and H. Robbins. Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics*, 6(1):4–22, 1985.

[13] L. Li, W. Chu, J. Langford, and R. E. Schapire. A contextual-bandit approach to personalized news article recommendation. In *International conference on World wide web*, pages 661–670, 2010.

[14] S. Li, A. Karatzoglou, and C. Gentile. Collaborative filtering bandits. In *International ACM SIGIR conference on Research and Development in Information Retrieval*, pages 539–548, 2016.

[15] M.-h. Oh, G. Iyengar, and A. Zeevi. Sparsity-agnostic lasso bandit. In *International Conference on Machine Learning*, 2021.

[16] H. Robbins. Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58(5):527–535, 1952.

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

[18] J. A. Tropp. User-friendly tail bounds for matrix martingales. Technical report, 2011.

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

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

[21] K. Weinberger, A. Dasgupta, J. Langford, A. Smola, and J. Attenberg. Feature hashing for large scale multitask learning. In *International Conference on Machine Learning*, page 1113–1120, 2009.

[22] C. Zeng, Q. Wang, S. Mokhtari, and T. Li. Online context-aware recommendation with time varying multi-armed bandit. In *International conference on Knowledge discovery and data mining*, pages 2025–2034, 2016.

[23] S. Zhou. Thresholded lasso for high dimensional variable selection and statistical estimation. *arXiv preprint arXiv:1002.1583*, 2010.

Appendix

A Table of Notations

Table 2 summarizes the notations used in the paper.

Table 2: Table of notations

Problem-specific notations	
$A_{t,k}$	Feature vector associated with the arm k
θ	Parameter vector
d	Dimension of feature vectors
s_0	Sparsity index
T	Total number of rounds
$\mathcal{A}_t$	Set of context vectors at round t
p_A	Distribution for $\mathcal{A}_t$
r_t	Reward at round t
$\mathcal{F}_t$	σ-algebra generated by random variables $(\mathcal{A}_1, A_1, r_1, \ldots, \mathcal{A}_{t-1}, A_{t-1}, r_{t-1}, \mathcal{A}_t)$
ε_t	Zero mean sub-Gaussian noise
σ^2	Variance proxy of ε_t
$R(T)$	Regret
$\hat{\Sigma}_t$	Empirical Gram matrix generated by the arms selected under a specific algorithm, i.e., $\frac{1}{t}\sum_{s=1}^{t} A_s A_s^\top$
S	Support $\{i \in [d] : \theta_i \neq 0\}$
$\theta_{\min}$	$\min_{i \in S} \lvert\theta_i\rvert$
s_A	ℓ_2 norm bound on $A_{S'}$ (see Assumptions 1 and 7)
$\phi^2(M, S_0)$	Compatibility constant (see Assumption 2)
Σ	Expected Gram matrix $\frac{1}{K}\sum_{k=1}^{K} \mathbb{E}_{\mathcal{A}\sim p_A}\left[A_k A_k^\top\right]$
ϕ_0^2	Lower bound on $\phi^2(\Sigma, S)$
ν	Constant for Relaxed symmetry (see Assumption 3)
C_{b}	Constant for Balanced covariance (see Assumption 4)
α	Constant for Sparse positive definiteness (see Assumption 5 and 8)
λ_t	Regularizer at round t
λ_0	Coefficient of the regularizer
$\hat{S}_0^{(t)}, \hat{S}_1^{(t)}$	Estimate of the support after the first and the second thresholding, respectively.
$\hat{\theta}_t$	Estimated vector of θ
C_{m}	Constant for the margin condition (see Assumption 6)
h_0	Constant whose order is $\mathcal{O}((\log s_0)^{\frac{1}{2}})$ (see definitions before the Theorems)
C_0	Constant whose order is $\mathcal{O}(1/s_0)$ (see definitions before the Theorems)
τ	$\max\left(\lfloor\frac{2\log(2d^2)}{C_0^2}\rfloor, \lfloor(\log\log d)\log d\rfloor\right)$
$\mathcal{E}_t$	Event $\left\{S \subset \hat{S}_1^{(t)} \text{ and } \lvert\hat{S}_1^{(t)} \setminus S\rvert \leq \frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}\right\}$
$\hat{S}$	Estimate of the support after the second thresholding (Equivalent to $\hat{S}_1^{(t)}$)
$\hat{\Sigma}_{\hat{S}}$	$\frac{1}{t}\sum_{s=1}^{t} A_s(\hat{S}) A_s(\hat{S})^\top$
$\mathcal{G}_t^\lambda$	Event $\left\{\lambda_{\min}(\hat{\Sigma}_{\hat{S}}) \geq \lambda\right\}$
$A_{\max}$	ℓ_2 norm bound on $A_{t,k}$ (used in the experiments)
Generic notations	
$\lVert x\rVert_0$	ℓ_0 norm of x, i.e., $\lVert x\rVert_0 = \sum_{i=1}^{d} \mathbb{1}\{\theta_i \neq 0\}$
$[x]$	Set of positive integers upto x, i.e., $[x] = \{1, \ldots, x\}$
$\langle x, y\rangle$	Inner product of x and y
$\mathbb{P}(A)$	Probability that event A occurs
$\mathbb{E}[a]$	Expected value of a
$\theta_{i,B}$	$\theta_i \mathbb{1}\{i \in B\}$
θ_B	$(\theta_{1,B}, \ldots, \theta_{d,B})^\top$
$A(B)$	$n \times \lvert B\rvert$ submatrix of $A \in \mathbb{R}^{n\times d}$ where $B \subset [d]$
$\mathrm{supp}(x)$	Set of the non-zero element indices of x

B Additional Theorems

Before presenting the additional theorems, we introduce following Assumptions (which are slightly modified versions of Assumption 1 and 5).

Assumption 7 (Sparsity and parameter constraints) *The parameter θ defining the reward function is sparse, i.e., $\|\theta\|_0 \leq s_0$ for some fixed but unknown integer s_0 (s_0 does not depend on d). We denote by $S = \{i \in [d] : \theta_i \neq 0\}$ the support of θ. We further assume that $\|\theta\|_2 \leq s_2$ for some unknown constant s_2. Finally, we assume that for* *any subset S' of $[d]$ such that $|S'| \le \mathcal{O}(s_0)$, the ℓ_2-norm of the context vector is bounded: for all t and for all $A \in \mathcal{A}_t$, for all $S' \subset [d]$ s.t. $|S'| \le s_0 + \frac{4\nu\sqrt{s_0}}{\phi_0^2}$, $\|A_{S'}\|_2 \le s_A$, where $s_A > 0$ is a constant and $\phi_0 > 0$, $\nu \ge 1$ are constants defined in Assumptions 2 and 3.*

Assumption 8 (Sparse positive definiteness) *Define $\Sigma_B := \frac{1}{K}\sum_{k=1}^{K} \mathbb{E}_{\mathcal{A}\sim p_A}\left[A_k(B)A_k(B)^\top\right]$, for any $B \subset [d]$, where $A_k(B)$ is a $|B|$-dimensional vector extracted the elements of A_k with indices in B. There exists a positive constant $\alpha > 0$ such that $\forall B \subset [d]$,*

$$\left(|B| \le s_0 + (4\nu\sqrt{s_0})/\phi_0^2 \text{ and } S \subset B\right) \Longrightarrow \left(\min_{v\in\mathbb{R}^d:\|v\|_2=1} v^\top \Sigma_B v \ge \alpha\right).$$

The parameters ϕ_0, ν are those of Assumptions 2, 3.

We redefine the parameters: $h_0 = \lfloor\sqrt{\log(4(s_0 + \frac{2\nu\sqrt{s_0}}{\phi_0^2}))} + 1\rfloor$ and $\tau = \max\left(\lfloor\frac{2\log(2d^2)}{C_0^2}\rfloor, \lfloor(\log\log d)\log d\rfloor\right)$, where $C_0 = \min\left\{\frac{1}{2}, \frac{\phi_0^2}{256 s_0 s_A^2 \nu}\right\}$. Next, we present regret upper bounds without Assumption 4 when $K = 2$.

Theorem 3 *Assume $K = 2$. Set $\lambda_0 = 4\sigma s_A\sqrt{c}$ with $c > 0$ such that $4\left(\frac{2\nu s_0}{\phi_0^2} + \sqrt{\left(1+\frac{2\nu}{\phi_0^2}\right)s_0}\right)\lambda_\tau \le \theta_{\min}$. Assume $\tau \ge \exp(4/c)$. Under Assumptions 7, 2, 3, 8, and 6, we have:*

$$\begin{aligned}R(T) &\le 2s_A s_2 \tau \\ &+ \frac{352\sigma^2 s_A^4 C_{\mathrm{m}} h_0^3 \nu^2\left(s_0 + \frac{2\nu\sqrt{s_0}}{\phi_0^2}\right)}{\alpha^2}(\log T + 1) + 2s_A s_2\left(\frac{\pi^2}{3} + \frac{2}{C_0^2} + \left(s_0 + \frac{2\nu\sqrt{s_0}}{\phi_0^2}\right)\frac{10 s_A^2\nu}{\alpha}\right).\end{aligned}$$

In particular, we have, as d and T grow large,

$$R(T) = \mathcal{O}(s_0^2\log d + s_0\log T + s_0^2),$$

where the $\mathcal{O}(\cdot)$ notation hides all the multiplicative and additive constants that do not depend on s_0, d, and T and hides $\log\log d$ and $(\log(s_0))^{\frac{3}{2}}$ multiplicative factors for the first and second term. Moreover, when $\lambda_0 = 1/(\log\log d)^{\frac{1}{4}}$ and d is large enough so that
(i): $\log\log d \ge 64\sigma^2 s_A^2(\log\log d)^{\frac{1}{2}}$, (ii): $4\left(\frac{2\nu s_0}{\phi_0^2} + \sqrt{\left(1+\frac{2\nu}{\phi_0^2}\right)s_0}\right)/(\theta_{\min}(\log\log d)^{\frac{1}{4}})\sqrt{2\log\tau\log d/\tau} \le 1$, and (iii): $d \ge 100$, as d and T grow large,

$$R(T) = \mathcal{O}(s_0^2\log d + s_0\log T + s_0^2).$$

Theorem 4 *Assume $K = 2$. Let $\lambda_0 = 4\sigma s_A\sqrt{c}$ with some constant $c > 0$, such that $4\left(\frac{2\nu s_0}{\phi_0^2} + \sqrt{\left(1+\frac{2\nu}{\phi_0^2}\right)s_0}\right)\lambda_\tau \le \theta_{\min}$. Assume $\tau \ge \exp(4/c)$. Under Assumptions 7, 2, 3, and 8, we have:*

$$\begin{aligned}R(T) &\le 2s_A s_2 \tau \\ &+ \frac{36\sigma s_A h_0^2 \nu\sqrt{2\left(s_0 + \frac{2\nu\sqrt{s_0}}{\phi_0^2}\right)}}{\alpha}\sqrt{T} + 2s_A s_2\left(\frac{\pi^2}{3} + \frac{2}{C_0^2} + \left(s_0 + \frac{2\nu\sqrt{s_0}}{\phi_0^2}\right)\frac{10 s_A^2\nu}{\alpha}\right).\end{aligned}$$

In particular, we have, as d and T grow large,

$$R(T) = \mathcal{O}(s_0^2\log d + \sqrt{s_0 T} + s_0^2),$$

where the $\mathcal{O}(\cdot)$ notation hides all the multiplicative and additive constants that do not depend on s_0, d, and T and hides $\log\log d$ and $\log s_0$ multiplicative factors for the first and second term respectively. Moreover, when $\lambda_0 = 1/(\log\log d)^{\frac{1}{4}}$ and d is large enough so that the conditions (i), (ii), and (iii) of Theorem 3 is satisfied. As d and T grows large,

$$R(T) = \mathcal{O}(s_0^2\log d + \sqrt{s_0 T} + s_0^2).$$

C Proof of Theorems

C.1 Proof of Theorem 1 (with margin)

We first upper bound the instantaneous regret in round $t \geq 1$. We have:

$$\begin{aligned}
\mathbb{E}[\max_{A\in\mathcal{A}_t}\langle A - A_t, \theta\rangle] &= \mathbb{E}[\max_{A\in\mathcal{A}_t}\langle A - A_t, \theta_S\rangle] \\
&\leq \mathbb{E}[|\max_{A\in\mathcal{A}_t}\langle A, \theta_S\rangle|] + \mathbb{E}[|\langle A_t, \theta_S\rangle|] \\
&\leq s_A s_2 + s_A s_2 \\
&= 2 s_A s_2, \qquad (1)
\end{aligned}$$

where the second inequality stems from Cauchy–Schwarz inequality. We deduce the following upper bound on the expected regret up to round T:

$$\begin{aligned}
R(T) &= \mathbb{E}\left[\sum_{t=1}^{T}\max_{A\in\mathcal{A}_t}\langle A - A_t, \theta\rangle\right] \\
&\overset{(a)}{\leq} 2s_As_2\tau + \sum_{t=\tau+1}^{T}\mathbb{E}\left[\max_{A\in\mathcal{A}_t}\langle A - A_t, \theta\rangle\right] \\
&\overset{(b)}{\leq} 2s_As_2\tau + \sum_{t=\tau+1}^{T}\Bigg(\frac{1408\sigma^2 s_A^4 C_{\mathrm{m}}(K-1)h_0^3\nu^2C_{\mathrm{b}}^2\left(s_0 + \frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}\right)}{\alpha^2}\frac{1}{t-1} \\
&\qquad + 2(K-1)s_As_2\left(\mathbb{P}(\mathcal{E}_t^c) + \mathbb{P}\left(\left(\mathcal{G}_t^{\frac{\alpha}{4\nu C_{\mathrm{b}}}}\right)^c\middle|\mathcal{E}_t\right)\right)\Bigg) \\
&\overset{(c)}{\leq} 2s_As_2\tau + \sum_{t=\tau+1}^{T}\Bigg(\frac{1408\sigma^2 s_A^4 C_{\mathrm{m}}(K-1)h_0^3\nu^2C_{\mathrm{b}}^2\left(s_0 + \frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}\right)}{\alpha^2}\frac{1}{t-1} \\
&\qquad + 2(K-1)s_As_2\Bigg(2\exp\left(-\frac{t\lambda_t^2}{32\sigma^2s_A^2} + \log d\right) + \exp\left(-\frac{tC_0^2}{2}\right) \\
&\qquad + \exp\left(\log\left(s_0 + \frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}\right) - \frac{t\alpha}{20s_A^2\nu C_{\mathrm{b}}}\right)\Bigg)\Bigg),
\end{aligned}$$

where for (a), we used equation (1) for $1 \leq t \leq \tau$; for (b), we used Lemma 4; for (c), we used Lemma 1 (for $\mathcal{E}_t$) and Lemma 2 (for $\mathcal{G}_t^{\frac{\alpha}{4\nu C_{\mathrm{b}}}}$).

Now we have:

$$\begin{aligned}
&\sum_{t=\tau+1}^{T}\frac{1408\sigma^2 s_A^4 C_{\mathrm{m}}(K-1)h_0^3\nu^2C_{\mathrm{b}}^2\left(s_0 + \frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}\right)}{\alpha^2}\frac{1}{t-1} \\
&= \sum_{t=\tau}^{T-1}\frac{1408\sigma^2 s_A^4 C_{\mathrm{m}}(K-1)h_0^3\nu^2C_{\mathrm{b}}^2\left(s_0 + \frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}\right)}{\alpha^2}\frac{1}{t} \\
&\leq \frac{1408\sigma^2 s_A^4 C_{\mathrm{m}}(K-1)h_0^3\nu^2C_{\mathrm{b}}^2\left(s_0 + \frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}\right)}{\alpha^2}\left(1 + \int_1^T \frac{1}{t}dt\right) \\
&= \frac{1408\sigma^2 s_A^4 C_{\mathrm{m}}(K-1)h_0^3\nu^2C_{\mathrm{b}}^2\left(s_0 + \frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}\right)}{\alpha^2}(1 + \log T),
\end{aligned}$$

and

$$
\begin{aligned}
\sum_{t=\tau+1}^{T} \exp\left(-\frac{t\lambda_t^2}{32\sigma^2 s_A^2}+\log d\right) &= \sum_{t=\tau+1}^{T} \exp\left(-c\log t\log d+\log d\right) \\
&\overset{(a)}{\leq} \sum_{t=\tau+1}^{T} \exp\left(-\frac{c\log d\log t}{2}\right) \\
&\overset{(b)}{\leq} \sum_{t=\tau+1}^{T} \exp\left(-2\log t\right) \\
&= \sum_{t=\tau+1}^{T} \frac{1}{t^2} \\
&\leq \sum_{t=1}^{\infty} \frac{1}{t^2} \\
&= \frac{\pi^2}{6},
\end{aligned}
$$

where for (a) and (b), we used the assumtion $\tau \geq \exp(4/c)$. In addition,

$$
\begin{aligned}
&\sum_{t=\tau+1}^{T} \exp\left(-\frac{tC_0^2}{2}\right)+\exp\left(\log\left(s_0+\frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}\right)-\frac{t\alpha}{20s_A^2\nu C_{\mathrm{b}}}\right) \\
&\leq \int_0^{\infty}\left(\exp\left(-\frac{tC_0^2}{2}\right)+\left(s_0+\frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}\right)\exp\left(-\frac{t\alpha}{20s_A^2\nu C_{\mathrm{b}}}\right)\right)dt \\
&= \frac{2}{C_0^2}+\left(s_0+\frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}\right)\frac{20s_A^2\nu C_{\mathrm{b}}}{\alpha}.
\end{aligned}
$$

In summary, we obtain:

$$
\begin{aligned}
R(T) \leq 2s_A s_2\tau &+ \frac{1408\sigma^2 s_A^4 C_{\mathrm{m}}(K-1)h_0^3\nu^2 C_{\mathrm{b}}^2\left(s_0+\frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}\right)}{\alpha^2}(1+\log T) \\
&+ 2(K-1)s_A s_2\left(\frac{\pi^2}{3}+\frac{2}{C_0^2}+\left(s_0+\frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}\right)\frac{20s_A^2\nu C_{\mathrm{b}}}{\alpha}\right).
\end{aligned}
$$

This concludes the proof of the regret upper bound. Regarding the scaling in d, T and s_0, note that $1/C_0^2 = \mathcal{O}(s_0^2)$ and $h_0^3 = \mathcal{O}((\log(s_0))^{\frac{3}{2}})$. Therefore, we have:

$$
R(T) = \mathcal{O}(s_0^2\log d + s_0\log T + s_0^2).
$$

Regarding the theorem when $\lambda_0 = 1/(\log\log d)^{\frac{1}{4}}$,

$$
\begin{aligned}
\sum_{t=\tau+1}^{T} \exp\left(-\frac{t\lambda_t^2}{32\sigma^2 s_A^2}+\log d\right) &= \sum_{t=\tau+1}^{T} \exp\left(-\left(\frac{\log t}{16\sigma^2 s_A^2(\log\log d)^{\frac{1}{2}}}-1\right)\log d\right) \\
&\overset{(a)}{\leq} \sum_{t=\tau+1}^{T} \exp\left(-\frac{\log t\log d}{32\sigma^2 s_A^2(\log\log d)^{\frac{1}{2}}}\right) \\
&\overset{(b)}{\leq} \sum_{t=\tau+1}^{T} \exp\left(-\frac{\log t\log \tau}{32\sigma^2 s_A^2(\log\log d)^{\frac{1}{2}}}\right) \\
&\overset{(c)}{\leq} \sum_{t=\tau+1}^{T} \exp\left(-2\log t\right) \\
&= \sum_{t=\tau+1}^{T} \frac{1}{t^2} \\
&\leq \sum_{t=1}^{\infty} \frac{1}{t^2} \\
&= \frac{\pi^2}{6},
\end{aligned}
$$

where for (a), we used the fact that $\log\tau \geq 64\sigma^2 s_A^2(\log\log d)^{\frac{1}{2}}$; for (b), we used the fact that $\log\tau \leq \log d$ from $d \geq 100$; for (c), we used again $\log\tau \geq 64\sigma^2 s_A^2(\log\log d)^{\frac{1}{2}}$. Therefore, similarly, we get the regret bound

$$\begin{aligned}R(T) \leq 2s_A s_2\tau &+ \frac{1408\sigma^2 s_A^4 C_{\mathrm{m}}(K-1)h_0^3\nu^2 C_{\mathrm{b}}^2\left(s_0+\frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}\right)}{\alpha^2}(1+\log T)\\ &+ 2(K-1)s_A s_2\left(\frac{\pi^2}{3}+\frac{2}{C_0^2}+\left(s_0+\frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}\right)\frac{20s_A^2\nu C_{\mathrm{b}}}{\alpha}\right).\end{aligned}$$

This concludes the proof.

□

C.2 Proof of Theorem 2 (without margin)

Using Lemma 5, we proceed as in the proof of Theorem 1, and deduce that:

$$\begin{aligned}R(T) &= \mathbb{E}\left[\sum_{t=1}^{T}\max_{A\in\mathcal{A}_t}\langle A-A_t,\theta\rangle\right]\\ &\leq 2s_A s_2\tau + \sum_{t=\tau+1}^{T}\mathbb{E}\left[\max_{A\in\mathcal{A}_t}\langle A-A_t,\theta\rangle\right]\\ &\leq 2s_A s_2\tau\\ &+ \sum_{t=\tau+1}^{T}\left(\frac{36\sigma s_A(K-1)h_0^2\nu C_{\mathrm{b}}}{\alpha}\sqrt{\frac{2\left(s_0+\frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}\right)}{t-1}} + 2(K-1)s_A s_2\left(\mathbb{P}(\mathcal{E}_t^c)+\mathbb{P}\left(\left(\mathcal{G}_t^{\frac{\alpha}{4\nu C_{\mathrm{b}}}}\right)^c\middle|\mathcal{E}_t\right)\right)\right)\end{aligned}$$

We also show that:

$$\begin{aligned}\sum_{t=\tau+1}^{T} 2(K-1)s_A s_2&\left(\mathbb{P}(\mathcal{E}_t^c)+\mathbb{P}\left(\left(\mathcal{G}_t^{\frac{\alpha}{4\nu C_{\mathrm{b}}}}\right)^c\middle|\mathcal{E}_t\right)\right)\\ &\leq 2(K-1)s_A s_2\left(\frac{\pi^2}{3}+\frac{2}{C_0^2}+\left(s_0+\frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}\right)\frac{20s_A^2\nu C_{\mathrm{b}}}{\alpha}\right)\end{aligned}$$

Now we have:

$$\begin{aligned}\sum_{t=\tau+1}^{T}\frac{36\sigma s_A(K-1)h_0^2\nu C_{\mathrm{b}}}{\alpha}\sqrt{\frac{2\left(s_0+\frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}\right)}{t-1}} &= \sum_{t=\tau}^{T-1}\frac{36\sigma s_A(K-1)h_0^2\nu C_{\mathrm{b}}}{\alpha}\sqrt{\frac{2\left(s_0+\frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}\right)}{t}}\\ &\leq \frac{36\sigma s_A(K-1)h_0^2\nu C_{\mathrm{b}}\sqrt{2\left(s_0+\frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}\right)}}{\alpha}(1+\int_1^T\sqrt{\frac{1}{t}}dt)\\ &\leq \frac{36\sigma s_A(K-1)h_0^2\nu C_{\mathrm{b}}\sqrt{8\left(s_0+\frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}\right)}}{\alpha}\sqrt{T},\end{aligned}$$

In summary, we get:

$$\begin{aligned}R(T) \leq 2s_A s_2\tau &+ \frac{36\sigma s_A(K-1)h_0^2\nu C_{\mathrm{b}}\sqrt{8\left(s_0+\frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}\right)}}{\alpha}\sqrt{T}\\ &+ 2(K-1)s_A s_2\left(\frac{\pi^2}{3}+\frac{2}{C_0^2}+\left(s_0+\frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}\right)\frac{20s_A^2\nu C_{\mathrm{b}}}{\alpha}\right).\end{aligned}$$

This concludes the proof of the regret upper bound. Regarding the scaling in d, T and s_0, note that $1/C_0^2 = \mathcal{O}(s_0^2)$ and $h_0^2 = \mathcal{O}(\log(s_0))$. Therefore, we have:

$$R(T) = \mathcal{O}\left(s_0^2\log d + \sqrt{s_0 T} + s_0^2\right).$$

Furthermore, when $\lambda_0 = 1/(\log\log d)^{\frac{1}{4}}$, we proceed as in the proof of Theorem 1, and get the regret bound

$$\begin{aligned}R(T) \leq 2s_A s_2\tau &+ \frac{36\sigma s_A(K-1)h_0^2\nu C_{\mathrm{b}}\sqrt{8\left(s_0+\frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}\right)}}{\alpha}\sqrt{T}\\ &+ 2(K-1)s_A s_2\left(\frac{\pi^2}{3}+\frac{2}{C_0^2}+\left(s_0+\frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}\right)\frac{20s_A^2\nu C_{\mathrm{b}}}{\alpha}\right).\end{aligned}$$

This concludes the proof. □

C.3 Proof of Theorem 3 (with margin, without balanced covariance)

This proof follows that of Theorem 1 to some extent. We first upper bound the instantaneous regret in round $t \geq 1$. We have:

$$\begin{aligned}
\mathbb{E}[\max_{A\in\mathcal{A}_t}\langle A - A_t, \theta\rangle] &= \mathbb{E}[\max_{A\in\mathcal{A}_t}\langle A - A_t, \theta_S\rangle] \qquad (2)\\
&\leq \mathbb{E}[|\max_{A\in\mathcal{A}_t}\langle A, \theta_S\rangle|] + \mathbb{E}[|\langle A_t, \theta_S\rangle|]\\
&\leq s_A s_2 + s_A s_2\\
&= 2 s_A s_2, \qquad (3)
\end{aligned}$$

where the second inequality stems from Cauchy–Schwarz inequality. We deduce the following upper bound on the expected regret up to round T:

$$\begin{aligned}
R(T) &= \mathbb{E}\left[\sum_{t=1}^{T}\max_{A\in\mathcal{A}_t}\langle A - A_t, \theta\rangle\right]\\
&\overset{(a)}{\leq} 2s_A s_2\tau + \sum_{t=\tau+1}^{T}\mathbb{E}\left[\max_{A\in\mathcal{A}_t}\langle A - A_t, \theta\rangle\right]\\
&\overset{(b)}{\leq} 2s_A s_2\tau + \sum_{t=\tau+1}^{T}\left(\frac{352\sigma^2 s_A^4 C_{\mathrm{m}} h_0^3\nu^2\left(s_0 + \frac{2\nu\sqrt{s_0}}{\phi_0^2}\right)}{\alpha^2}\frac{1}{t-1} + 2s_A s_2\left(\mathbb{P}(\mathcal{E}_t^c) + \mathbb{P}\left(\left(\mathcal{G}_t^{\frac{\alpha}{2\nu}}\right)^c\middle|\mathcal{E}_t\right)\right)\right)\\
&\overset{(c)}{\leq} 2s_A s_2\tau + \sum_{t=\tau+1}^{T}\Bigg(\frac{352\sigma^2 s_A^4 C_{\mathrm{m}} h_0^3\nu^2\left(s_0 + \frac{2\nu\sqrt{s_0}}{\phi_0^2}\right)}{\alpha^2}\frac{1}{t-1}\\
&\quad + 2s_A s_2\left(2\exp\left(-\frac{t\lambda_t^2}{32\sigma^2 s_A^2} + \log d\right) + \exp\left(-\frac{tC_0^2}{2}\right) + \exp\left(\log\left(s_0 + \frac{2\nu\sqrt{s_0}}{\phi_0^2}\right) - \frac{t\alpha}{10 s_A^2\nu}\right)\right)\Bigg),
\end{aligned}$$

where for (a), we used equation (3) for $1 \leq t \leq \tau$; for (b), we used Lemma 18; for (c), we used Lemma 15 (for $\mathcal{E}_t$) and Lemma 16 (for $\mathcal{G}_t^{\frac{\alpha}{2\nu}}$).

Now we have:

$$\begin{aligned}
\sum_{t=\tau+1}^{T}\frac{352\sigma^2 s_A^4 C_{\mathrm{m}} h_0^3\nu^2\left(s_0 + \frac{2\nu\sqrt{s_0}}{\phi_0^2}\right)}{\alpha^2}\frac{1}{t-1} &= \sum_{t=\tau}^{T-1}\frac{352\sigma^2 s_A^4 C_{\mathrm{m}} h_0^3\left(s_0 + \frac{2\nu\sqrt{s_0}}{\phi_0^2}\right)}{\alpha^2}\frac{1}{t}\\
&\leq \frac{352\sigma^2 s_A^4 C_{\mathrm{m}} h_0^3\nu^2\left(s_0 + \frac{2\nu\sqrt{s_0}}{\phi_0^2}\right)}{\alpha^2}(1 + \int_1^T \frac{1}{t}dt)\\
&= \frac{352\sigma^2 s_A^4 C_{\mathrm{m}} h_0^3\nu^2\left(s_0 + \frac{2\nu\sqrt{s_0}}{\phi_0^2}\right)}{\alpha^2}(1 + \log T),
\end{aligned}$$

and

$$\begin{aligned}
\sum_{t=\tau+1}^{T}\exp\left(-\frac{t\lambda_t^2}{32\sigma^2 s_A^2} + \log d\right) &= \sum_{t=\tau+1}^{T}\exp\left(-c\log t\log d + \log d\right)\\
&\overset{(a)}{\leq} \sum_{t=\tau+1}^{T}\exp\left(-\frac{c\log d\log t}{2}\right)\\
&\overset{(b)}{\leq} \sum_{t=\tau+1}^{T}\exp\left(-2\log t\right)\\
&= \sum_{t=\tau+1}^{T}\frac{1}{t^2}\\
&\leq \sum_{t=1}^{\infty}\frac{1}{t^2}\\
&= \frac{\pi^2}{6},
\end{aligned}$$

where for (a) and (b), we used the assumption $\tau \geq \exp(4/c)$. In addition,

$$\begin{aligned}
&\sum_{t=\tau+1}^{T}\left(\exp\left(-\frac{tC_0^2}{2}\right)+\exp\left(\log\left(s_0+\frac{2\nu\sqrt{s_0}}{\phi_0^2}\right)-\frac{t\alpha}{10s_A^2\nu}\right)\right)\\
&\leq \int_0^\infty \left(\exp\left(-\frac{tC_0^2}{2}\right)+\left(s_0+\frac{2\nu\sqrt{s_0}}{\phi_0^2}\right)\exp\left(-\frac{t\alpha}{10s_A^2\nu}\right)\right)dt\\
&=\frac{2}{C_0^2}+\left(s_0+\frac{2\nu\sqrt{s_0}}{\phi_0^2}\right)\frac{10s_A^2\nu}{\alpha}.
\end{aligned}$$

In summary, we obtain that:

$$\begin{aligned}
R(T) \leq\ & 2s_As_2\tau+\frac{352\sigma^2s_A^4C_{\mathrm{m}}h_0^3\nu^2\left(s_0+\frac{2\nu\sqrt{s_0}}{\phi_0^2}\right)}{\alpha^2}(\log T+1)\\
&+2s_As_2\left(\frac{\pi^2}{3}+\frac{2}{C_0^2}+\left(s_0+\frac{2\nu\sqrt{s_0}}{\phi_0^2}\right)\frac{10s_A^2\nu}{\alpha}\right).
\end{aligned}$$

Moreover, when $\lambda_0 = 1/(\log\log d)^{\frac{1}{4}}$ and conditions (i), (ii), and (iii) on d are satisfied,

$$\begin{aligned}
\sum_{t=\tau+1}^{T}\exp\left(-\frac{t\lambda_t^2}{32\sigma^2s_A^2}+\log d\right) &= \sum_{t=\tau+1}^{T}\exp\left(-\left(\frac{\log t}{16\sigma^2s_A^2(\log\log d)^{\frac{1}{2}}}-1\right)\log d\right)\\
&\overset{(a)}{\leq}\sum_{t=\tau+1}^{T}\exp\left(-\frac{\log t\log d}{32\sigma^2s_A^2(\log\log d)^{\frac{1}{2}}}\right)\\
&\overset{(b)}{\leq}\sum_{t=\tau+1}^{T}\exp\left(-\frac{\log t\log \tau}{32\sigma^2s_A^2(\log\log d)^{\frac{1}{2}}}\right)\\
&\overset{(c)}{\leq}\sum_{t=\tau+1}^{T}\exp\left(-2\log t\right)\\
&=\sum_{t=\tau+1}^{T}\frac{1}{t^2}\\
&\leq\sum_{t=1}^{\infty}\frac{1}{t^2}\\
&=\frac{\pi^2}{6},
\end{aligned}$$

where for (a), we used the fact that $\log\tau \geq 64\sigma^2s_A^2(\log\log d)^{\frac{1}{2}}$; for (b), we used the fact that $\log\tau \leq \log d$ from $d \geq 100$; for (c), we used again $\log\tau \geq 64\sigma^2s_A^2(\log\log d)^{\frac{1}{2}}$. Therefore, a similar regret upper bound can be obtained in this case:

$$\begin{aligned}
R(T) \leq\ & 2s_As_2\tau+\frac{352\sigma^2s_A^4C_{\mathrm{m}}h_0^3\nu^2\left(s_0+\frac{2\nu\sqrt{s_0}}{\phi_0^2}\right)}{\alpha^2}(\log T+1)\\
&+2s_As_2\left(\frac{\pi^2}{3}+\frac{2}{C_0^2}+\left(s_0+\frac{2\nu\sqrt{s_0}}{\phi_0^2}\right)\frac{10s_A^2\nu}{\alpha}\right).
\end{aligned}$$

This concludes the proof of the regret upper bound. Regarding the scaling in d, T and s_0, note that $1/C_0^2 = \mathcal{O}(s_0^2)$ and $h_0^3 = \mathcal{O}((\log(s_0))^{\frac{3}{2}})$. Therefore, we have:

$$R(T) = \mathcal{O}(s_0^2\log d + s_0\log T + s_0^2).$$

This concludes the proof. □

C.4 Proof of Theorem 4 (without margin, without balanced covariance)

This proof follows the proof of Theorem 2 mostly. Using Lemma 19, we proceed as in the proof of Theorem 2, and deduce that:

$$
\begin{aligned}
R(T) &= \mathbb{E}[\sum_{t=1}^{T} \max_{A\in\mathcal{A}_t} \langle A - A_t, \theta\rangle] \\
&\le 2s_A s_2 \tau + \sum_{t=\tau+1}^{T} \mathbb{E}[\max_{A\in\mathcal{A}_t} \langle A - A_t, \theta\rangle] \\
&\le 2s_A s_2 \tau + \sum_{t=\tau+1}^{T} \left(\frac{18\sigma s_A h_0^2 \nu}{\alpha} \sqrt{\frac{2\left(s_0 + \frac{2\nu\sqrt{s_0}}{\phi_0^2}\right)}{t-1}} + 2s_A s_2 \left(\mathbb{P}(\mathcal{E}_t^c) + \mathbb{P}\left(\left(\mathcal{G}_t^{\frac{\alpha}{2\nu}}\right)^c |\mathcal{E}_t\right)\right)\right).
\end{aligned}
$$

We have:

$$
\begin{aligned}
&\sum_{t=\tau+1}^{T} 2s_A s_2 \left(\mathbb{P}(\mathcal{E}_t^c) + \mathbb{P}\left(\left(\mathcal{G}_t^{\frac{\alpha}{2\nu}}\right)^c \middle| \mathcal{E}_t\right)\right) \\
\le &\sum_{t=\tau+1}^{T} 2s_A s_2 \left(2\exp\left(-\frac{t\lambda_t^2}{32\sigma^2 s_A^2} + \log d\right) + \exp\left(-\frac{tC_0^2}{2}\right) + \exp\left(\log\left(s_0 + \frac{2\nu\sqrt{s_0}}{\phi_0^2}\right) - \frac{t\alpha}{10 s_A^2 \nu}\right)\right),
\end{aligned}
$$

Regarding the series involving $\sqrt{1/t}$, we have:

$$
\begin{aligned}
\sum_{t=\tau+1}^{T} \sqrt{\frac{1}{t-1}} &= \sum_{t=\tau}^{T-1} \sqrt{\frac{1}{t}} \\
&\le 1 + \int_1^T \sqrt{\frac{1}{t}} dt \\
&\le 2\sqrt{T}.
\end{aligned}
$$

The bounds for

$$
\sum_{t=\tau+1}^{T} 2s_A s_2 \left(2\exp\left(-\frac{t\lambda_t^2}{32\sigma^2 s_A^2} + \log d\right) + \exp\left(-\frac{tC_0^2}{2}\right) + \exp\left(\log\left(s_0 + \frac{2\nu\sqrt{s_0}}{\phi_0^2}\right) - \frac{t\alpha}{10 s_A^2 \nu}\right)\right)
$$

hold similarly as is in Theorem 3.

In summary, we get:

$$
\begin{aligned}
R(T) &\le 2s_A s_2 \tau \\
&\quad + \frac{36\sigma s_A h_0^2 \nu \sqrt{2\left(s_0 + \frac{2\nu\sqrt{s_0}}{\phi_0^2}\right)}}{\alpha} \sqrt{T} + 2s_A s_2 \left(\frac{\pi^2}{3} + \frac{2}{C_0^2} + \left(s_0 + \frac{2\nu\sqrt{s_0}}{\phi_0^2}\right)\frac{10 s_A^2 \nu}{\alpha}\right).
\end{aligned}
$$

This concludes the proof of the regret upper bound. Regarding the scaling, note that $1/C_0^2 = \mathcal{O}(s_0^2)$. Therefore, we have:

$$
R(T) = \mathcal{O}\left(s_0^2 \log d + \sqrt{s_0 T \log T} + s_0^2\right).
$$

This concludes the proof. □

D Proof of Lemmas

D.1 Proof of Lemma 1

We define $v := \hat{\theta}_0^{(t)} - \theta$. We first analyze the performance of the initial Lasso estimate.

Lemma 6 *Let $\hat{\Sigma}_t := \frac{\sum_{s=1}^t A_s A_s^\top}{t}$ be the empirical covariance matrix of the selected context vectors. Suppose $\hat{\Sigma}_t$ satisfies the compatibility condition with the support S with the compatibility constant ϕ_t. Then, under Assumption 1, we have:*

$$
\mathbb{P}\left(\|v\|_1 \le \frac{4s_0\lambda_t}{\phi_t^2}\right) \ge 1 - 2\exp\left(-\frac{t\lambda_t^2}{32\sigma^2 s_A^2} + \log d\right).
$$

The next lemma then states that the compatibility constant of $\hat{\Sigma}_t$ does not deviate much from the compatibility constant of Σ.

Lemma 7 *Let $C_0 := \min\left\{\frac{1}{2}, \frac{\phi_0^2}{512 s_0 s_A^2 \nu C_\mathrm{b}}\right\}$. For all $t \geq \frac{2\log(2d^2)}{C_0^2}$, we have:*

$$\mathbb{P}\left(\phi^2(\hat{\Sigma}_t, S) \geq \frac{\phi_0^2}{4\nu C_\mathrm{b}}\right) \geq 1 - \exp\left(-\frac{tC_0^2}{2}\right).$$

Then, we follow the steps of the proof given by [23]. Let us define the event $\mathcal{G}_t$ as:

$$\mathcal{G}_t := \left\{\|v\|_1 \leq \frac{4s_0\lambda_t}{\phi_t^2}\right\}.$$

For the rest of this section, we assume that the event $\mathcal{G}_t$ holds. Note that:

$$\begin{aligned}
\|v\|_1 \geq \|v_{S^c}\|_1 &= \sum_{j\in S^c} |(\hat{\theta}_0^{(t)})_j| \\
&\geq \sum_{j\in S^c \cap \hat{S}_0^{(t)}} |(\hat{\theta}_0^{(t)})_j| \\
&= \sum_{j\in \hat{S}_0^{(t)} \setminus S} |(\hat{\theta}_0^{(t)})_j| \\
&\overset{(a)}{\geq} |\hat{S}_0^{(t)} \setminus S| 4\lambda_t,
\end{aligned}$$

where for (a), we used the construction of $\hat{S}_0^{(t)}$ in the algorithm. We get:

$$\begin{aligned}
|\hat{S}_0^{(t)} \setminus S| &\leq \frac{\|v\|_1}{4\lambda_t} \\
&\overset{(a)}{\leq} \frac{s_0}{\phi_t^2},
\end{aligned}$$

where for (a), we used the definition of $\mathcal{G}_t$.

We have: $\forall j \in S$,

$$\begin{aligned}
|(\hat{\theta}_0^{(t)})_j| &\geq \theta_{\min} - \|v_S\|_\infty \\
&\geq \theta_{\min} - \|v_S\|_1 \\
&\geq \theta_{\min} - \frac{4s_0\lambda_t}{\phi_t^2}.
\end{aligned}$$

Therefore, when t is large enough so that $4\lambda_t \leq \theta_{\min} - \frac{4s_0\lambda_t}{\phi_t^2}$, we have: $S \subset \hat{S}_0^{(t)}$. Using a similar argument, when t is large enough so that $4\lambda_t\sqrt{\left(1+\frac{1}{\phi_t^2}\right)s_0} \leq \theta_{\min} - \frac{4s_0\lambda_t}{\phi_t^2}$, it holds that $S \subset \hat{S}_1^{(t)}$. From the construction of $\hat{S}_1^{(t)}$ in the algorithm, it also holds that: $\hat{S}_1^{(t)} \subset \hat{S}_0^{(t)}$. Therefore,

$$\begin{aligned}
\|v\|_1 &\geq \sum_{i\in \hat{S}_0^{(t)} \setminus S} |(\hat{\theta}_0^{(t)})_i| \\
&\geq \sum_{i\in \hat{S}_1^{(t)} \setminus S} |(\hat{\theta}_0^{(t)})_i| \\
&\geq |\hat{S}_1^{(t)} \setminus S| 4\lambda_t \sqrt{|\hat{S}_0^{(t)}|},
\end{aligned}$$

and

$$\begin{aligned}
|\hat{S}_1^{(t)} \setminus S| &\leq \frac{\|v\|_1}{4\lambda_t\sqrt{|\hat{S}_0^{(t)}|}} \\
&\leq \frac{1}{4\lambda_t\sqrt{|\hat{S}_0^{(t)}|}} \cdot \frac{4s_0\lambda_t}{\phi_t^2} \\
&\leq \frac{\sqrt{s_0}}{\phi_t^2}.
\end{aligned}$$

Note that the condition $4\lambda_t\sqrt{\left(1+\frac{1}{\phi_t^2}\right)s_0} \leq \theta_{\min} - \frac{4s_0\lambda_t}{\phi_t^2}$ is equivalent to $4\lambda_t\left(\sqrt{\left(1+\frac{1}{\phi_t^2}\right)s_0} + \frac{s_0}{\phi_t^2}\right) \leq \theta_{\min}$. This concludes the proof of Lemma 1 by substituting $\phi_t^2 = \phi_0^2/(4\nu C_\mathrm{b})$. □

D.2 Proof of Lemmas used in the proof of Lemma 1

D.2.1 Proof of Lemma 6

The proof is similar to that given by [15]. Let us define the loss function:

$$\ell_t(\theta) := \frac{1}{t}\sum_{s=1}^{t}(r_t - \langle\theta, A_t\rangle)^2.$$

The initial Lasso estimate is given by:

$$\hat{\theta}_t := \arg\min_{\theta'}\left\{\ell_t(\theta') + \lambda_t\|\theta'\|_1\right\}.$$

From this definition, we get:

$$\ell_t(\hat{\theta}_t) + \lambda_t\|\hat{\theta}_t\|_1 \le \ell_t(\theta) + \lambda_t\|\theta\|_1.$$

Let us denote $\mathbb{E}[\cdot]$ as the expectation over r_t in this section. Note that in view of the previous inequality, we have:

$$\ell_t(\hat{\theta}_t) - \mathbb{E}[\ell_t(\hat{\theta}_t)] + \mathbb{E}[\ell_t(\hat{\theta}_t)] - \mathbb{E}[\ell_t(\theta)] + \lambda_t\|\hat{\theta}_t\|_1 \le \ell_t(\theta) - \mathbb{E}[\ell_t(\theta)] + \lambda_t\|\theta\|_1.$$

Denoting $v_t(\theta) := \ell_t(\theta) - \mathbb{E}[\ell_t(\theta)]$ and $\mathcal{E}(\theta') := \mathbb{E}[\ell_t(\theta')] - \mathbb{E}[\ell_t(\theta)]$,

$$v_t(\hat{\theta}_t) + \mathcal{E}(\hat{\theta}_t) + \lambda_t\|\hat{\theta}_t\|_1 \le v_t(\theta) + \lambda_t\|\theta\|_1.$$

Let us define the event $\mathcal{T}_t$:

$$\mathcal{T}_t := \{|v_t(\hat{\theta}_t) - v_t(\theta)| \le \frac{1}{2}\lambda_t\|\hat{\theta}_t - \theta\|_1\}.$$

We can condition on this event in the rest of the proof:

Lemma 8 *We have:*

$$\mathbb{P}\left(|v_t(\hat{\theta}_t) - v_t(\theta)| \le \frac{1}{2}\lambda_t\|\hat{\theta}_t - \theta\|_1\right) \ge 1 - 2\exp\left(-\frac{t\lambda_t^2}{32\sigma^2 s_A^2} + \log d\right).$$

Given the event $\mathcal{T}_t$, we have:

$$2\mathcal{E}(\hat{\theta}_t) \le 2\lambda_t(\|\theta\|_1 - \|\hat{\theta}_t\|_1) + \lambda_t\|\hat{\theta}_t - \theta\|_1.$$

By the triangle inequality,

$$\begin{aligned}\|\hat{\theta}_t\|_1 &= \|\hat{\theta}_{t,S}\|_1 + \|\hat{\theta}_{t,S^c}\|_1\\ &\ge \|\theta_S\|_1 - \|\hat{\theta}_{t,S} - \theta_S\| + \|\hat{\theta}_{t,S^c}\|_1.\end{aligned}$$

We also have:

$$\begin{aligned}\|\hat{\theta}_t - \theta\|_1 &= \|(\hat{\theta}_t - \theta)_S\|_1 + \|(\hat{\theta}_t - \theta)_{S^c}\|_1\\ &= \|\hat{\theta}_{t,S} - \theta_S\|_1 + \|\hat{\theta}_{t,S^c}\|_1.\end{aligned}$$

Therefore, we get:

$$\begin{aligned}2\mathcal{E}(\hat{\theta}_t) &\le 2\lambda_t\|\theta\|_1 - 2\lambda_t(\|\theta_S\|_1 - \|\hat{\theta}_{t,S} - \theta_S\|_1 + \|\hat{\theta}_{t,S^c}\|_1) + \lambda_t(\|\hat{\theta}_{t,S} - \theta_S\|_1 + \|\hat{\theta}_{t,S^c}\|_1)\\ &= 3\lambda_t\|\hat{\theta}_{t,S} - \theta_S\|_1 - \lambda_t\|\hat{\theta}_{t,S^c}\|_1.\end{aligned} \tag{4}$$

From the compatibility condition, we get:

$$\|\hat{\theta}_{t,S} - \theta_S\|_1^2 \le \frac{s_0(\hat{\theta}_t - \theta)^\top\hat{\Sigma}_t(\hat{\theta}_t - \theta)}{\phi_t^2} \tag{5}$$

Using inequality (4), we get:

$$\begin{aligned}2\mathcal{E}(\hat{\theta}_t) + \lambda_t\|\hat{\theta}_t - \theta\|_1 &= 2\mathcal{E}(\hat{\theta}_t) + \lambda_t\|\hat{\theta}_{t,S^c}\|_1 + \lambda_t\|\hat{\theta}_{t,S} - \theta_S\|_1\\ &\le 3\lambda_t\|\hat{\theta}_{t,S} - \theta_S\|_1 + \lambda_t\|\hat{\theta}_{t,S} - \theta_S\|_1\\ &= 4\lambda_t\|\hat{\theta}_{t,S} - \theta_S\|_1\\ &\le \frac{4\lambda_t}{\phi_t}\sqrt{s_0(\hat{\theta}_t - \theta)^\top\hat{\Sigma}_t(\hat{\theta}_t - \theta)}\\ &\le (\hat{\theta}_t - \theta)^\top\hat{\Sigma}_t(\hat{\theta}_t - \theta) + \frac{4\lambda_t^2 s_0}{\phi_t^2}\\ &\le 2\mathcal{E}(\hat{\theta}_t) + \frac{4\lambda_t^2 s_0}{\phi_t^2},\end{aligned}$$

where for the third inequality, we used $4uv \le u^2 + 4v^2$ with $u = \sqrt{(\hat{\theta}_t - \theta)^\top\hat{\Sigma}_t(\hat{\theta}_t - \theta)}$ and $v = \frac{\lambda_t\sqrt{s_0}}{\phi_t}$. The last inequality is due to Lemma 9:

Lemma 9 *We have:*

$$\mathcal{E}(\hat{\theta}_t) \geq \frac{1}{2}(\hat{\theta}_t - \theta)^\top \hat{\Sigma}_t(\hat{\theta}_t - \theta) \ .$$

Thus, we get:

$$\|\hat{\theta}_t - \theta\|_1 \leq \frac{4\lambda_t s_0}{\phi_t^2}.$$

This concludes the proof. □

D.2.2 Proof of Lemma 7

First we define the adapted Gram matrix $\Sigma_t := \frac{1}{t}\sum_{s=1}^t \mathbb{E}[A_s A_s^\top | \mathcal{F}_{s-1}]$. From the construction of the algorithm, $\mathbb{E}[A_s A_s^\top | \mathcal{F}_{s-1}] = \mathbb{E}[\sum_{k=1}^K A_{s,k} A_{s,k}^\top \mathbb{1}\{k = \operatorname{argmax}_{k'} \langle A_{s,k}, \hat{\theta}_s\rangle\} | \, \hat{\theta}_s]$. The following lemma characterizes the expected Gram matrix generated by the algorithm.

Lemma 10 (Lemma 10 of [15]) *Under Assumptions 3 and 4, for each fixed vector $\theta' \in \mathbb{R}^d$, we have:*

$$\mathbb{E}_{\mathcal{A} \sim p_A}\left[\sum_{k=1,2} A_k A_k^\top \mathbb{1}\{k = \operatorname*{argmax}_{k'} \langle A_k, \theta'\rangle\}\right] \succeq \frac{1}{2\nu C_{\mathrm{b}}}\Sigma,$$

where $A \succeq B$ means that $A - B$ is positive semidefinite.

Using Lemma 10, we have

$$\Sigma_t \succeq \frac{1}{2\nu C_{\mathrm{b}}}\Sigma. \tag{6}$$

By Lemma 6.18 of [6], Assumption 2, and the definition of the compatibility constant, we get:

$$\phi^2(\Sigma_t, S) \geq \phi^2(\frac{1}{2\nu C_{\mathrm{b}}}\Sigma, S) \geq \frac{\phi_0^2}{2\nu C_{\mathrm{b}}}. \tag{7}$$

Furthermore, we have a following adaptive matrix concentration results for $\hat{\Sigma}_t$:

Lemma 11 *Let $C_0 := \min\left\{\frac{1}{2}, \frac{\phi_0^2}{512 s_0 s_A^2 \nu C_{\mathrm{b}}}\right\}$. We have, for all $t \geq \frac{2\log(2d^2)}{C_0^2}$,*

$$\mathbb{P}\left(\frac{1}{2s_A^2}\|\hat{\Sigma}_t - \Sigma_t\|_\infty \geq \frac{\phi^2(\Sigma_t, S)}{64 s_0 s_A^2}\right) \leq \exp\left(-\frac{tC_0^2}{2}\right).$$

We use a following result from [6]:

Lemma 12 (Corollary 6.8 in [6]) *Suppose Σ_0 satisfies the compatibility condition for the set S with $|S| = s_0$, with the compatibility constant $\phi^2(\Sigma_0, S) > 0$, and that $\|\Sigma_0 - \Sigma_1\|_\infty \leq \lambda$, where $\frac{32\lambda s_0}{\phi^2(\Sigma_0,S)} \leq 1$. Then, the compatibility condition also holds for Σ_1 with the compatibility constant $\frac{\phi^2(\Sigma_0,S)}{2}$, i.e., $\phi^2(\Sigma_1, S) \geq \frac{\phi^2(\Sigma_0,S)}{2}$.*

Combining the above results, we get, for all $t \geq \frac{2\log(2d^2)}{C_0^2}$:

$$\begin{aligned}\phi^2(\hat{\Sigma}_t, S) &\geq \frac{\phi^2(\Sigma_t, S)}{2} \\ &\geq \frac{\phi_0^2}{4\nu C_{\mathrm{b}}},\end{aligned}$$

with probability at least $1 - \exp\left(-\frac{tC_0^2}{2}\right)$. This concludes the proof. □

D.2.3 Proof of Lemma 8

Let us denote $\hat{\theta} = \hat{\theta}_t$ for simplicity. We compute $v_t(\hat{\theta})$ as:

$$\begin{aligned}v_t(\hat{\theta}) &= \ell_t(\hat{\theta}) - \mathbb{E}[\ell_t(\hat{\theta})] \\ &= \frac{1}{t}\sum_{s=1}^t (r_s - \langle\hat{\theta}, A_s\rangle)^2 - \frac{1}{t}\sum_{s=1}^t \mathbb{E}\left[(r_s - \langle\hat{\theta}, A_s\rangle)^2\right] \\ &= \frac{1}{t}\sum_{s=1}^t (\langle\theta, A_s\rangle + \varepsilon_s - \langle\hat{\theta}, A_s\rangle)^2 - \frac{1}{t}\sum_{s=1}^t \mathbb{E}\left[(\langle\theta, A_s\rangle + \varepsilon_s - \langle\hat{\theta}, A_s\rangle)^2\right] \\ &= \frac{1}{t}\sum_{s=1}^t (2\varepsilon_s\langle\theta - \hat{\theta}, A_s\rangle + \varepsilon_s^2 - \mathbb{E}[\varepsilon_s^2]).\end{aligned}$$

We also have that:

$$v_t(\theta) = \frac{1}{t}\sum_{s=1}^{t}(\varepsilon_s^2 - \mathbb{E}[\varepsilon_s^2]).$$

Therefore, we can compute:

$$\begin{aligned}
v_t(\hat{\theta}) - v_t(\theta) &= \frac{1}{t}\sum_{s=1}^{t} 2\varepsilon_s\langle \theta - \hat{\theta}, A_s\rangle \\
&\le \frac{2}{t}\left\|\sum_{s=1}^{t}\varepsilon_s A_s\right\|_\infty \|\theta - \hat{\theta}\|_1,
\end{aligned}$$

where we used Hölder's inequality in the above inequality. We have that:

$$\mathbb{P}\left(\frac{2}{t}\left\|\sum_{s=1}^{t}\varepsilon_s A_s\right\|_\infty \le \lambda\right) \ge 1 - \sum_{i=1}^{d}\mathbb{P}\left(\frac{2}{t}\left|\sum_{s=1}^{t}\varepsilon_s (A_s)_i\right| > \lambda\right),$$

where $(A_s)_i$ is the i-th element of A_s. Define $\widetilde{\mathcal{F}}_t$ as the σ-algebra generated by the random variables $(A_1, \mathcal{A}_1, \varepsilon_1, \ldots, A_t, \mathcal{A}_t, \varepsilon_t, \mathcal{A}_{t+1})$. For each $i \in [d]$, we get $\mathbb{E}[\varepsilon_s(A_s)_i|\widetilde{\mathcal{F}}_{s-1}] = (A_s)_i\,\mathbb{E}[\varepsilon_s|\widetilde{\mathcal{F}}_{s-1}] = 0$. Thus, for each $i \in [d]$, $\{\varepsilon_s(A_s)_i\}_{s=1}^t$ is a martingale difference sequence adapted to the filtration $\widetilde{\mathcal{F}}_1 \subset \ldots \subset \widetilde{\mathcal{F}}_{t-1}$. By Assumption 1, we have $|(A_s)_i| \le s_A$. We compute, for each $\alpha \in \mathbb{R}$,

$$\begin{aligned}
\mathbb{E}[\exp(\alpha\varepsilon_s(A_s)_i)|\widetilde{\mathcal{F}}_{s-1}] &\le \mathbb{E}[\exp(\alpha\varepsilon_s s_A)|\widetilde{\mathcal{F}}_{s-1}] \\
&\le \exp\left(\frac{\alpha^2 s_A^2\sigma^2}{2}\right).
\end{aligned}$$

Therefore $\varepsilon_s(A_s)_i$ is also a sub-Gaussian random variable with the variance proxy $(s_A\sigma)^2$. Next, we use the concentration results by [19], Theorem 2.19:

Theorem 5 *Let $(Z_t, \widetilde{\mathcal{F}}_t)_{t=1}^\infty$ be a martingale difference sequence, and assume that for all $\alpha \in \mathbb{R}$, $\mathbb{E}[\exp(\alpha Z_s)|\widetilde{\mathcal{F}}_{s-1}] \le \exp(\frac{\alpha^2\sigma^2}{2})$ with probability one. Then, for all $x \ge 0$, we get:*

$$\mathbb{P}\left(\left|\sum_{s=1}^{t} Z_s\right| \ge x\right) \le 2\exp\left(-\frac{x^2}{2t\sigma^2}\right).$$

From these results, we get:

$$\mathbb{P}\left(\left|\sum_{s=1}^{t}\varepsilon_s(A_s)_i\right| > \frac{t\lambda}{2}\right) \le 2\exp\left(-\frac{t\lambda^2}{8\sigma^2 s_A^2}\right).$$

Taking $\lambda = \frac{1}{2}\lambda_t$,

$$\begin{aligned}
\mathbb{P}\left(\frac{2}{t}\left\|\sum_{s=1}^{t}\varepsilon_s A_s\right\|_\infty \le \frac{\lambda_t}{2}\right) &\ge 1 - 2d\exp\left(-\frac{t\lambda_t^2}{32\sigma^2 s_A^2}\right) \\
&= 1 - 2\exp\left(-\frac{t\lambda_t^2}{32\sigma^2 s_A^2} + \log d\right).
\end{aligned}$$

□

D.2.4 Proof of Lemma 9

We denote $\hat{\theta}_t = \hat{\theta}$ for brevity. From the definitions of $\mathcal{E}(\theta')$ and $\ell_t(\theta')$,

$$\begin{aligned}
\mathcal{E}(\hat{\theta}) &= \mathbb{E}[\ell_t(\hat{\theta})] - \mathbb{E}[\ell_t(\theta)] \\
&= \frac{1}{t}\,\mathbb{E}\left[\sum_{s=1}^{t}(\langle\hat{\theta} - \theta, A_s\rangle + \varepsilon_s)^2 - \sum_{s=1}^{t}\varepsilon_s^2\right] \\
&= \frac{1}{t}\sum_{s=1}^{t}\langle\hat{\theta} - \theta, A_s\rangle^2 \\
&= \frac{1}{t}\sum_{s=1}^{t}(\hat{\theta} - \theta)^\top A_s A_s^\top(\hat{\theta} - \theta) \\
&= (\hat{\theta} - \theta)^\top\hat{\Sigma}_t(\hat{\theta} - \theta) \\
&\ge \frac{1}{2}(\hat{\theta} - \theta)^\top\hat{\Sigma}_t(\hat{\theta} - \theta),
\end{aligned}$$

where for the inequality, we used the positive semi-definiteness of $\hat{\Sigma}_t$. □

D.2.5 Proof of Lemma 10

The proof is almost identical to the proof of Lemma 10 in [15]. $\square$

D.2.6 Proof of Lemma 11

Let us define $\gamma_t^{ij}(A_t)$ as:

$$\gamma_t^{ij}(A_t) \coloneqq \frac{1}{2s_A^2}\left((A_t)_i(A_t)_j - \mathbb{E}[(A_t)_i(A_t)_j \mid \mathcal{F}_{t-1}]\right),$$

where $(A_t)_i$ is the i-th element of A_t. We have, following a Bernstein-like inequality for the adapted data:

Lemma 13 (Bernstein-like inequality for the adapted data [15]) *Suppose for all $t \geq 1$, for all $1 \leq i \leq j \leq d$, $\mathbb{E}[\gamma_t^{ij}(A_t)|\mathcal{F}_{t-1}] = 0$ and $\mathbb{E}[|\gamma_t^{ij}(A_t)|^m \mid \mathcal{F}_{t-1}] \leq m!$ for all integer $m \geq 2$. Then, for all $x > 0$, and for all integer $t \geq 1$, we have:*

$$\mathbb{P}\left(\max_{1\leq i\leq j\leq d}\left|\frac{1}{t}\sum_{s=1}^{t}\gamma_s^{ij}(A_s)\right| \geq x + \sqrt{2x} + \sqrt{\frac{4\log(2d^2)}{t}} + \frac{2\log(2d^2)}{t}\right) \leq \exp\left(-\frac{tx}{2}\right).$$

Note that $\frac{1}{2s_A^2}\|\hat{\Sigma}_t - \Sigma_t\|_\infty = \max_{1\leq i\leq j\leq d}\left|\frac{1}{t}\sum_{s=1}^{t}\gamma_s^{ij}(A_s)\right|$, $\mathbb{E}[\gamma_t^{ij}(A_t)|\mathcal{F}_{t-1}] = 0$, and $\mathbb{E}[|\gamma_t^{ij}(A_t)|^m \mid \mathcal{F}_{t-1}] \leq 1$ for all integer $m \geq 2$. Therefore, we can apply Lemma 13:

$$\mathbb{P}\left(\frac{1}{2s_A^2}\|\hat{\Sigma}_t - \Sigma_t\|_\infty \geq x + \sqrt{2x} + \sqrt{\frac{4\log(2d^2)}{t}} + \frac{2\log(2d^2)}{t}\right) \leq \exp\left(-\frac{tx}{2}\right).$$

For all $t \geq \frac{2\log(2d^2)}{C_0^2}$ with $C_0 \coloneqq \min\left\{\frac{1}{2}, \frac{\phi_0^2}{256 s_0 s_A^2 \nu C_\mathrm{b}}\right\}$, taking $x = C_0^2$,

$$\begin{aligned}
x + \sqrt{2x} + \sqrt{\frac{4\log(2d^2)}{t}} + \frac{2\log(2d^2)}{t} &\leq 2C_0^2 + 2\sqrt{2}C_0 \\
&\leq 4C_0 \\
&\leq \frac{\phi_0^2}{128 s_0 s_A^2 \nu C_\mathrm{b}} \\
&\leq \frac{\phi^2(\Sigma_t, S)}{64 s_0 s_A^2}.
\end{aligned}$$

In summary, for all $t \geq \frac{2\log(2d^2)}{C_0^2}$, we get:

$$\begin{aligned}
\mathbb{P}\left(\frac{1}{2s_A^2}\|\hat{\Sigma}_t - \Sigma_t\|_\infty \geq \frac{\phi^2(\Sigma_t, S)}{64 s_0 s_A^2}\right) &\leq \mathbb{P}\left(\frac{1}{2s_A^2}\|\hat{\Sigma}_t - \Sigma_t\|_\infty \geq C_0^2 + \sqrt{2}C_0 + \sqrt{\frac{4\log(2d^2)}{t}} + \frac{2\log(2d^2)}{t}\right) \\
&\leq \exp\left(-\frac{tC_0^2}{2}\right).
\end{aligned}$$

This concludes the proof. $\square$

D.3 Proof of Lemma 2

For a fixed $\hat{S}$, first we define the adapted Gram matrix on the estimated support as $\Sigma_t \coloneqq \frac{1}{t}\sum_{s=1}^{t}\mathbb{E}[A_s(\hat{S})A_s(\hat{S})^\top|\mathcal{F}_{s-1}]$. From the construction of the algorithm, $\mathbb{E}[A_s(\hat{S})A_s(\hat{S})^\top|\mathcal{F}_{s-1}] = \mathbb{E}[\sum_{k=1}^{K} A_{s,k}(\hat{S})A_{s,k}(\hat{S})^\top \mathbb{1}\{k = \operatorname*{argmax}_{k'}\langle A_{s,k}, \hat{\theta}_s\rangle\}|\ \hat{\theta}_s]$. Recall that for each $B \subset [d]$, $\Sigma_B \coloneqq \frac{1}{K}\sum_{k=1}^{K}\mathbb{E}_{\mathcal{A}\sim p_A}\left[A_k(B)A_k(B)^\top\right]$, where $A_k(B)$ is a $|B|$-dimensional vector extracted the elements of A_k with indices in B. The following lemma characterizes the expected Gram matrix generated by the algorithm.

Lemma 14 *Fix $\hat{S}$ such that $S \subset \hat{S}$ and $|\hat{S}| \leq s_0 + (4\nu C_\mathrm{b}\sqrt{s_0})/\phi_0^2$. Fix $\theta' \in \mathbb{R}^d$. Under Assumption 2, 3, and 4, we have:*

$$\mathbb{E}_{\mathcal{A}\sim p_A}\left[\sum_{k\in[K]} A_k(\hat{S})A_k(\hat{S})^\top \mathbb{1}\{k = \operatorname*{argmax}_{k'}\langle A_k, \theta'_{\hat{S}}\rangle\}\right] \succeq \frac{1}{2\nu C_\mathrm{b}}\Sigma_{\hat{S}},$$

where $A \succeq B$ means that $A - B$ is positive semidefinite.

First, we prove the lower bound on the smallest eigenvalue of the expected covariance matrices. Let $\Sigma_{\hat{S}} := \frac{1}{t}\sum_{s=1}^{t}\mathbb{E}[A_s(\hat{S})A_s(\hat{S})^\top \mid \mathcal{F}_{t-1}]$. By Assumption 5 and the construction of the algorithm, under the event $\mathcal{E}_t$, we get:

$$\begin{aligned}
\lambda_{\min}(\Sigma_{\hat{S}}) &= \lambda_{\min}\left(\frac{1}{t}\sum_{s=1}^{t}\mathbb{E}\left[\sum_{k=1}^{K} A_{s,k,\hat{S}}A_{s,k,\hat{S}}\mathbb{1}\{k = \operatorname*{argmax}_{k'}\langle A_{k'}, \hat{\theta}_s\rangle\} \mid \hat{\theta}_s\right]\right)\\
&\geq \sum_{s=1}^{t}\lambda_{\min}\left(\frac{1}{t}\mathbb{E}\left[\sum_{k=1}^{K} A_{s,k,\hat{S}}A_{s,k,\hat{S}}\mathbb{1}\{k = \operatorname*{argmax}_{k'}\langle A_{k'}, \hat{\theta}_s\rangle\} \mid \hat{\theta}_s\right]\right)\\
&\geq \frac{\alpha}{2\nu C_{\mathrm{b}}},
\end{aligned}$$

where for the first inequality, we used the concavity of $\lambda_{\min}(\cdot)$ over the positive semi-definite matrices. Next, we prove the upper bound on the largest eigenvalue of $A_s(\hat{S})A_s(\hat{S})^\top$:

$$\begin{aligned}
\lambda_{\max}(A_s(\hat{S})A_s(\hat{S})^\top) &= \max_{\|v\|=1} v^\top A_s(\hat{S})A_s(\hat{S})^\top v\\
&\overset{(a)}{\leq} \max_{\|v\|=1} \|v\|_2^2\|A_s(\hat{S})\|_2^2\\
&\leq s_A^2,
\end{aligned}$$

where for (a), we used the Cauchy-Schwarz inequality. Now recall the matrix Chernoff inequality by [18]:

Theorem 6 (Matrix Chernoff, Theorem 3.1 of [18]) *Let $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \ldots \subset \mathcal{F}_t$ be a filtration and a consider a finite sequence $\{X_s\}$ of positive semi-definite matrices with dimension d, adapted to the filtration. Suppose $\lambda_{\max}(X_k) \leq R$ almost surely. Define the finite series: $Y := \sum_{s=1}^{t} X_s$ and $W := \sum_{s=1}^{t}\mathbb{E}[X_s \mid \mathcal{F}_{s-1}]$. Then, for all $\mu \geq 0$, for all $\delta \in [0,1)$, we have:*

$$\mathbb{P}\left(\lambda_{\min}(Y) \leq (1-\delta)\mu \text{ and } \lambda_{\min}(W) \geq \mu\right) \leq d\left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\frac{\mu}{R}}.$$

Taking $R = s_A^2$, $X_s = A_{s,\hat{S}}A_{s,\hat{S}}^\top$, $Y = t\hat{\Sigma}_{\hat{S}}$, $W = t\Sigma_{\hat{S}}$, $\delta = 1/2$, $\mu = t\frac{\alpha}{2\nu C_{\mathrm{b}}}$:

$$\begin{aligned}
\mathbb{P}\left(\lambda_{\min}(t\hat{\Sigma}_{\hat{S}}) \leq \frac{1}{2}t\frac{\alpha}{2\nu C_{\mathrm{b}}} \text{ and } \lambda_{\min}(t\Sigma_{\hat{S}}) \geq t\frac{\alpha}{2\nu C_{\mathrm{b}}}\right) &\leq \left(s_0 + \frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}\right)\left(\frac{e^{-0.5}}{0.5^{0.5}}\right)^{\frac{t}{s_A^2}\frac{\alpha}{2\nu C_{\mathrm{b}}}}\\
&\leq \exp\left(\log\left(s_0 + \frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}\right) - \frac{t\alpha}{20 s_A^2\nu C_{\mathrm{b}}}\right),
\end{aligned}$$

where for the last inequality, we used $-0.5 - 0.5\log(0.5) < -\frac{1}{10}$. This concludes the proof. □

D.4 Proof of Lemma 3

In this proof, we denote $\hat{S} = \hat{S}_1^{(t)}$ and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_t)^\top$. Assume $\lambda_{\min}(\hat{\Sigma}_{\hat{S}}) \geq \lambda$. We have:

$$\begin{aligned}
\|\hat{\theta}_{t+1} - \theta\|_2 &= \|(A(\hat{S})^\top A(\hat{S}))^{-1}A(\hat{S})^\top R - \theta\|_2\\
&= \|(A(\hat{S})^\top A(\hat{S}))^{-1}A(\hat{S})^\top(A\theta + \varepsilon) - \theta\|_2\\
&= \|(A(\hat{S})^\top A(\hat{S}))^{-1}A(\hat{S})^\top(A(\hat{S})\theta(\hat{S}) + \varepsilon) - \theta\|_2\\
&= \|(A(\hat{S})^\top A(\hat{S}))^{-1}A(\hat{S})^\top\varepsilon\|_2\\
&\leq \|(A(\hat{S})^\top A(\hat{S}))^{-1}\|_2\|A(\hat{S})^\top\varepsilon\|_2\\
&\leq \frac{1}{\lambda t}\|A(\hat{S})^\top\varepsilon\|_2.
\end{aligned}$$

We get (note that we are conditioning on a fixed $\hat{S}$ during the proof):

$$
\begin{aligned}
\mathbb{P}\left(\|\hat{\theta}_{t+1}-\theta\|_2 \geq x \text{ and } \lambda_{\min}(\hat{\Sigma}_{\hat{S}}) \geq \lambda\right) &= \mathbb{P}\left(\|\hat{\theta}_{t+1}-\theta\|_2 \geq x \;\middle|\; \lambda_{\min}(\hat{\Sigma}_{\hat{S}}) \geq \lambda\right)\mathbb{P}(\lambda_{\min}(\hat{\Sigma}_{\hat{S}}) \geq \lambda) \\
&\leq \mathbb{P}\left(\|A(\hat{S})^\top \varepsilon\|_2 \geq \lambda t x \;\middle|\; \lambda_{\min}(\hat{\Sigma}_{\hat{S}}) \geq \lambda\right)\mathbb{P}(\lambda_{\min}(\hat{\Sigma}_{\hat{S}}) \geq \lambda) \\
&\leq \mathbb{P}\left(\|A(\hat{S})^\top \varepsilon\|_2 \geq \lambda t x\right) \\
&\leq \sum_{i=1}^{d} \mathbb{P}\left(\left|\sum_{s=1}^{t} \varepsilon_s (A_s)_i \mathbb{1}\left\{i \in \hat{S}\right\}\right| \geq \frac{\lambda t x}{\sqrt{s_0 + \frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}}}\right) \\
&= \sum_{i \in \hat{S}} \mathbb{P}\left(\left|\sum_{s=1}^{t} \varepsilon_s (A_s)_i\right| \geq \frac{\lambda t x}{\sqrt{s_0 + \frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}}}\right) \\
&\overset{(a)}{\leq} 2\left(s_0 + \frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}\right)\exp\left(-\frac{\lambda^2 t x^2}{2\sigma^2 s_A^2\left(s_0 + \frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}\right)}\right),
\end{aligned}
$$

where for (a), we used Theorem 5. This concludes the proof. □

D.5 Proof of Lemma 4

We follow the proof strategy of Lemma 6 in [5]. Let r_t^π be the instantaneous expected regret of algorithm π at round t defined as:

$$r_t^\pi \coloneqq \mathbb{E}\left[\max_{A \in \mathcal{A}_t} \langle A - A_t, \theta\rangle\right].$$

Let us define the events $\mathcal{R}_k \coloneqq \{\mathcal{A}_t \in \mathbb{R}^{K\times d} : k \in \operatorname{argmax}_{k'} \langle A_{t,k'}, \theta\rangle\}$ and $\mathcal{G}_t^\lambda \coloneqq \left\{\lambda_{\min}(\hat{\Sigma}_{\hat{S}}) \geq \lambda\right\}$. We have:

$$r_t^\pi \leq \sum_{k=1}^{K} \mathbb{E}\left[r_t^\pi \mid \mathcal{A}_t \in \mathcal{R}_k\right]\mathbb{P}\left(\mathcal{A}_t \in \mathcal{R}_k\right).$$

The term $\mathbb{E}\left[r_t^\pi \mid \mathcal{A}_t \in \mathcal{R}_k\right]$ can be further computed as:

$$
\begin{aligned}
\mathbb{E}\left[r_t^\pi \mid \mathcal{A}_t \in \mathcal{R}_k\right] &= \mathbb{E}\left[\langle A_{t,k} - A_t, \theta\rangle \mid \mathcal{A}_t \in \mathcal{R}_k\right] \\
&\leq \mathbb{E}\left[\mathbb{1}\left\{\langle A_t, \hat{\theta}_t\rangle \geq \langle A_{t,k}, \hat{\theta}_t\rangle\right\}\langle A_{t,k} - A_t, \theta\rangle \mid \mathcal{A}_t \in \mathcal{R}_k\right] \\
&\leq \sum_{\ell \neq k} \mathbb{E}\left[\mathbb{1}\left\{\langle A_{t,\ell}, \hat{\theta}_t\rangle \geq \langle A_{t,k}, \hat{\theta}_t\rangle\right\}\langle A_{t,k} - A_{t,\ell}, \theta\rangle \mid \mathcal{A}_t \in \mathcal{R}_k\right] \\
&\leq \sum_{\ell \neq k} \mathbb{E}\left[\mathbb{1}\left\{\langle A_{t,\ell}, \hat{\theta}_t\rangle \geq \langle A_{t,k}, \hat{\theta}_t\rangle\right\}\langle A_{t,k} - A_{t,\ell}, \theta\rangle \mid \mathcal{A}_t \in \mathcal{R}_k,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{4\nu C_{\mathrm{b}}}}\right] \\
&\quad + 2(K-1)s_A s_2\left(\mathbb{P}(\mathcal{E}_t^c) + \mathbb{P}\left(\left(\mathcal{G}_t^{\frac{\alpha}{4\nu C_{\mathrm{b}}}}\right)^c \middle| \mathcal{E}_t\right)\right).
\end{aligned}
$$

Let us denote the event $I_h \coloneqq \{\mathcal{A}_t \in \mathbb{R}^{K\times d} : \langle A_{t,k} - A_{t,\ell}, \theta\rangle \in (2\delta s_A h, 2\delta s_A(h+1)]\}$ where

$$\delta = \frac{\sigma s_A \nu C_{\mathrm{b}}}{\alpha}\sqrt{\frac{32\left(s_0 + \frac{4\nu C_{\mathrm{b}} s_0}{\phi_0^2}\right)}{t-1}}.$$

By conditioning on I_h, we get:

$$
\begin{aligned}
&\mathbb{E}\left[\mathbb{1}\left\{\langle A_{t,\ell}, \hat{\theta}_t\rangle \geq \langle A_{t,k}, \hat{\theta}_t\rangle\right\}\langle A_{t,k} - A_{t,\ell}, \theta\rangle \mid \mathcal{A}_t \in \mathcal{R}_k \cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{4\nu C_{\mathrm{b}}}}\right] \\
&\leq \sum_{h=0}^{\lceil s_2/\delta\rceil} \mathbb{E}\left[\mathbb{1}\left\{\langle A_{t,\ell}, \hat{\theta}_t\rangle \geq \langle A_{t,k}, \hat{\theta}_t\rangle\right\}\langle A_{t,k} - A_{t,\ell}, \theta\rangle \mid \mathcal{A}_t \in \mathcal{R}_k \cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{4\nu C_{\mathrm{b}}}}\right]\mathbb{P}(\mathcal{A}_t \in I_h) \\
&\overset{(a)}{\leq} \sum_{h=0}^{\lceil s_2/\delta\rceil} 2\delta s_A(h+1)\,\mathbb{E}\left[\mathbb{1}\left\{\langle A_{t,\ell}, \hat{\theta}_t\rangle \geq \langle A_{t,k}, \hat{\theta}_t\rangle\right\} \mid \mathcal{A}_t \in \mathcal{R}_k \cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{4\nu C_{\mathrm{b}}}}\right]\mathbb{P}\left(\langle A_{t,k} - A_{t,\ell}, \theta\rangle \in (0, 2\delta s_A(h+1)]\right) \\
&\overset{(b)}{\leq} \sum_{h=0}^{\lceil s_2/\delta\rceil} 4\delta^2 s_A^2 (h+1)^2 C_{\mathrm{m}}\mathbb{P}\left(\langle A_{t,\ell}, \hat{\theta}_t\rangle \geq \langle A_{t,k}, \hat{\theta}_t\rangle \mid \mathcal{A}_t \in \mathcal{R}_k \cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{4\nu C_{\mathrm{b}}}}\right),
\end{aligned}
$$

where for (a), we used the definition of I_h and for (b), we used Assumption 6. Under the event $\mathcal{A}_t \in I_h$, the event $\langle A_{t,\ell}, \hat{\theta}_t\rangle \geq \langle A_{t,k}, \hat{\theta}_t\rangle$ happens only when at least one of the events $\langle A_{t,k}, \theta - \hat{\theta}\rangle \geq \delta s_A h$ or $\langle A_{t,\ell}, \hat{\theta} - \theta\rangle \geq \delta s_A h$ holds. Therefore,

$$
\begin{aligned}
&\mathbb{P}\left(\langle A_{t,\ell}, \hat{\theta}_t\rangle \geq \langle A_{t,k}, \hat{\theta}_t\rangle \mid \mathcal{A}_t \in \mathcal{R}_k \cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{4\nu C_{\mathrm{b}}}}\right) \\
&\leq \mathbb{P}\left(\langle A_{t,k}, \theta - \hat{\theta}\rangle \geq \delta s_A h \mid \mathcal{A}_t \in \mathcal{R}_k \cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{4\nu C_{\mathrm{b}}}}\right) + \mathbb{P}\left(\langle A_{t,\ell}, \hat{\theta} - \theta\rangle \geq \delta s_A h \mid \mathcal{A}_t \in \mathcal{R}_k \cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{4\nu C_{\mathrm{b}}}}\right) \\
&\overset{(a)}{\leq} \mathbb{P}\left(\|\theta - \hat{\theta}\|_2 \geq \delta h \mid \mathcal{A}_t \in \mathcal{R}_k \cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{4\nu C_{\mathrm{b}}}}\right) + \mathbb{P}\left(\|\theta - \hat{\theta}\|_2 \geq \delta h \mid \mathcal{A}_t \in \mathcal{R}_k \cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{4\nu C_{\mathrm{b}}}}\right) \\
&= 2\mathbb{P}\left(\|\theta - \hat{\theta}\|_2 \geq \delta h \mid \mathcal{A}_t \in \mathcal{R}_k \cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{4\nu C_{\mathrm{b}}}}\right),
\end{aligned}
$$

where for (a), we used the Cauchy–Schwarz inequality. Let us denote $s' = s_0 + \frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}$. Then, using Lemma 3, we get:

$$
\begin{aligned}
\mathbb{P}\left(\|\theta - \hat{\theta}\|_2 \geq \delta h \mid \mathcal{A}_t \in \mathcal{R}_k \cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{4\nu C_{\mathrm{b}}}}\right) &\leq 2s' \exp\left(-\frac{\alpha^2 t \delta^2 h^2}{32\sigma^2 s_A^2 \nu^2 C_{\mathrm{b}}^2 s'}\right) \\
&= 2s' \exp\left(-h^2\right).
\end{aligned}
$$

We also trivially have that:

$$
\mathbb{P}\left(\|\theta - \hat{\theta}\|_2 \geq \delta h \mid \mathcal{A}_t \in \mathcal{R}_k \cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{4\nu C_{\mathrm{b}}}}\right) \leq 1.
$$

Therefore, we can upper bound the expected instantaneous regret as:

$$
\begin{aligned}
r_t^\pi &\leq \sum_{k=1}^K \mathbb{E}\left[r_t^\pi \mid \mathcal{A}_t \in \mathcal{R}_k\right] \mathbb{P}\left(\mathcal{A}_t \in \mathcal{R}_k\right) \\
&\leq \sum_{k=1}^K \left(\sum_{\ell \neq k}\left(\sum_{h=0}^{\lceil s_2/\delta\rceil}\left(4\delta^2 s_A^2 (h+1)^2 C_{\mathrm{m}} \min\left\{1, 4s' \exp\left(-h^2\right)\right\}\right)\right)\right. \\
&\quad \left. +2(K-1)s_A s_2\left(\mathbb{P}(\mathcal{E}_t^c) + \mathbb{P}\left(\left(\mathcal{G}_t^{\frac{\alpha}{4\nu C_{\mathrm{b}}}}\right)^c\right)\right)\right) \cdot \mathbb{P}\left(\mathcal{A}_t \in \mathcal{R}_k\right) \\
&\overset{(a)}{\leq} G \sum_{h=0}^{\lceil s_2/\delta\rceil} (h+1)^2 \min\left\{1, 4s' \exp\left(-h^2\right)\right\} + 2(K-1)s_A s_2\left(\mathbb{P}(\mathcal{E}_t^c) + \mathbb{P}\left(\left(\mathcal{G}_t^{\frac{\alpha}{4\nu C_{\mathrm{b}}}}\right)^c \middle| \mathcal{E}_t\right)\right) \\
&\leq G\left(\sum_{h=0}^{h_0}(h+1)^2 + \sum_{h=h_0+1}^{h_{\max}} 4s'(h+1)^2 \exp\left(-h^2\right)\right) + 2(K-1)s_A s_2\left(\mathbb{P}(\mathcal{E}_t^c) + \mathbb{P}\left(\left(\mathcal{G}_t^{\frac{\alpha}{4\nu C_{\mathrm{b}}}}\right)^c \middle| \mathcal{E}_t\right)\right)
\end{aligned}
$$

where for brevity $G = 4\delta^2 s_A^2 C_{\mathrm{m}}(K-1)$, where for (a), we used $\sum_{k=1}^K \mathbb{P}\left(\mathcal{A}_t \in \mathcal{R}_k\right) = 1$ from Assumption 6 and we set $h_0 := \lfloor\sqrt{\log 4s'} + 1\rfloor$. We have:

$$
\begin{aligned}
\sum_{h=h_0+1}^{h_{\max}} (h+1)^2 \exp(-h^2) &= \sum_{h=h_0+1}^{h_{\max}} h^2 \exp(-h^2) + 2\sum_{h=h_0+1}^{h_{\max}} h \exp(-h^2) + \sum_{h=h_0+1}^{h_{\max}} \exp(-h^2) \\
&\leq \int_{h_0}^\infty h^2 \exp(-h^2) dh + 2\int_{h_0}^\infty h \exp(-h^2) dh + \int_{h_0}^\infty \exp(-h^2) dh.
\end{aligned}
$$

Using an integration by parts, the inequality $\int_{h_0}^\infty \exp(-h_0^2) dh \leq \exp(-h_0^2)/(h_0 + \sqrt{h_0^2 + 4/\pi}) \leq \exp(-h_0^2)$, and $h_0 \geq 1$ from $s_0 \geq 1$, we get:

$$
\begin{aligned}
\int_{h_0}^\infty h^2 \exp(-h^2) dh &\leq \frac{1}{2} h_0 \exp(-h_0^2) + \frac{1}{2}\exp(-h_0^2) \\
2\int_{h_0}^\infty h \exp(-h^2) dh &= \exp(-h_0^2) \\
\int_{h_0}^\infty \exp(-h^2) dh &\leq \exp(-h_0^2).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\sum_{h=h_0+1}^{h_{\max}} (h+1)^2 \exp(-h^2) &\leq \frac{1}{2} h_0 \exp(-h_0^2) + \frac{5}{2}\exp(-h_0^2) \\
&\leq h_0 \exp(-h_0^2) + 5\exp(-h_0^2).
\end{aligned}
$$

We get:

$$\sum_{h=0}^{h_0}(h+1)^2+\sum_{h=h_0+1}^{h_{\max}}4s'(h+1)^2\exp\left(-h^2\right)\leq\frac{(h_0+1)(h_0+2)(2h_0+3)}{6}+4s'(h_0+5)\exp(-h_0^2)$$
$$\leq\frac{2h_0^3+9h_0^2+13h_0+6}{6}+4s'(h_0+5)\frac{1}{4s'}$$
$$\stackrel{(a)}{\leq}11h_0^3,$$

where for (a), we used $h_0\geq 1$. Finally, we get:

$$r_t^\pi\leq 44\delta^2 s_A^2 C_{\mathrm{m}}(K-1)h_0^3+2(K-1)s_As_2\left(\mathbb{P}(\mathcal{E}_t^c)+\mathbb{P}\left(\left(\mathcal{G}_t^{\frac{\alpha}{4\nu C_{\mathrm{b}}}}\right)^c\middle|\mathcal{E}_t\right)\right)$$
$$\leq\frac{1408\sigma^2s_A^4C_{\mathrm{m}}(K-1)h_0^3\nu^2C_{\mathrm{b}}^2\left(s_0+\frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}\right)}{\alpha^2}\frac{1}{t-1}+2(K-1)s_As_2\left(\mathbb{P}(\mathcal{E}_t^c)+\mathbb{P}\left(\left(\mathcal{G}_t^{\frac{\alpha}{4\nu C_{\mathrm{b}}}}\right)^c\middle|\mathcal{E}_t\right)\right).$$

This concludes the proof. □

D.6 Proof of Lemma 5

Let r_t^π be the instantaneous expected regret of algorithm π in round t defined as:

$$r_t^\pi\coloneqq\mathbb{E}\left[\max_{A\in\mathcal{A}_t}\langle A-A_t,\theta\rangle\right].$$

Let us define the events $\mathcal{R}_k\coloneqq\{\mathcal{A}_t\in\mathbb{R}^{K\times d}:k\in\operatorname{argmax}_{k'}\langle A_{t,k'},\theta\rangle\}$ and $\mathcal{G}_t^\lambda\coloneqq\left\{\lambda_{\min}(\hat{\Sigma}_{\hat{S}})\geq\lambda\right\}$. As in the proof of Lemma 4, we get:

$$r_t^\pi\leq\sum_{k=1}^K\mathbb{E}\left[r_t^\pi\mid\mathcal{A}_t\in\mathcal{R}_k\right]\mathbb{P}\left(\mathcal{A}_t\in\mathcal{R}_k\right),$$

and

$$\mathbb{E}\left[r_t^\pi\mid\mathcal{A}_t\in\mathcal{R}_k\right]=\mathbb{E}\left[\langle A_{t,k}-A_t,\theta\rangle\mid\mathcal{A}_t\in\mathcal{R}_k\right]$$
$$\leq\sum_{\ell\neq k}\mathbb{E}\left[\mathbb{1}\left\{\langle A_{t,\ell},\hat{\theta}_t\rangle\geq\langle A_{t,k},\hat{\theta}_t\rangle\right\}\langle A_{t,k}-A_{t,\ell},\theta\rangle\mid\mathcal{A}_t\in\mathcal{R}_k,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{4\nu C_{\mathrm{b}}}}\right]$$
$$+2(K-1)s_As_2\left(\mathbb{P}(\mathcal{E}_t^c)+\mathbb{P}\left(\left(\mathcal{G}_t^{\frac{\alpha}{4\nu C_{\mathrm{b}}}}\right)^c\middle|\mathcal{E}_t\right)\right).$$

Let us denote the event $I_h\coloneqq\{\mathcal{A}_t\in\mathbb{R}^{K\times d}:\langle A_{t,k}-A_{t,\ell},\theta\rangle\in(2\delta s_Ah,2\delta s_A(h+1)]\}$ where

$$\delta=\frac{\sigma s_A\nu C_{\mathrm{b}}}{\alpha}\sqrt{\frac{32\left(s_0+\frac{4\nu C_{\mathrm{b}}s_0}{\phi_0^2}\right)}{t-1}}.$$

By conditioning on I_h, we get:

$$\mathbb{E}\left[\mathbb{1}\left\{\langle A_{t,\ell},\hat{\theta}_t\rangle\geq\langle A_{t,k},\hat{\theta}_t\rangle\right\}\langle A_{t,k}-A_{t,\ell},\theta\rangle\mid\mathcal{A}_t\in\mathcal{R}_k\cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{4\nu C_{\mathrm{b}}}}\right]$$
$$\leq\sum_{h=0}^{\lceil s_2/\delta\rceil}\mathbb{E}\left[\mathbb{1}\left\{\langle A_{t,\ell},\hat{\theta}_t\rangle\geq\langle A_{t,k},\hat{\theta}_t\rangle\right\}\langle A_{t,k}-A_{t,\ell},\theta\rangle\mid\mathcal{A}_t\in\mathcal{R}_k\cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{4\nu C_{\mathrm{b}}}}\right]\mathbb{P}(\mathcal{A}_t\in I_h)$$
$$\stackrel{(a)}{\leq}\sum_{h=0}^{\lceil s_2/\delta\rceil}2\delta s_A(h+1)\,\mathbb{E}\left[\mathbb{1}\left\{\langle A_{t,\ell},\hat{\theta}_t\rangle\geq\langle A_{t,k},\hat{\theta}_t\rangle\right\}\mid\mathcal{A}_t\in\mathcal{R}_k\cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{4\nu C_{\mathrm{b}}}}\right]$$
$$\times\,\mathbb{P}\left(\langle A_{t,k}-A_{t,\ell},\theta\rangle\in(0,2\delta s_A(h+1)]\right)$$
$$\leq\sum_{h=0}^{\lceil s_2/\delta\rceil}2\delta s_A(h+1)\mathbb{P}\left(\langle A_{t,\ell},\hat{\theta}_t\rangle\geq\langle A_{t,k},\hat{\theta}_t\rangle\mid\mathcal{A}_t\in\mathcal{R}_k\cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{4\nu C_{\mathrm{b}}}}\right),$$

where for (a), we used the definition of I_h. Under the event $\mathcal{A}_t \in I_h$, the event $\langle A_{t,\ell}, \hat{\theta}_t \rangle \geq \langle A_{t,k}, \hat{\theta}_t \rangle$ happens only when at least one of the events $\langle A_{t,k}, \theta - \hat{\theta} \rangle \geq \delta s_A h$ or $\langle A_{t,\ell}, \hat{\theta} - \theta \rangle \geq \delta s_A h$ holds. Therefore,

$$
\begin{aligned}
&\mathbb{P}\left(\langle A_{t,\ell}, \hat{\theta}_t \rangle \geq \langle A_{t,k}, \hat{\theta}_t \rangle \mid \mathcal{A}_t \in \mathcal{R}_k \cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{4\nu C_\mathrm{b}}}\right) \\
&\leq \mathbb{P}\left(\langle A_{t,k}, \theta - \hat{\theta} \rangle \geq \delta s_A h \mid \mathcal{A}_t \in \mathcal{R}_k \cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{2\nu}}\right) \\
&\qquad\qquad + \mathbb{P}\left(\langle A_{t,\ell}, \hat{\theta} - \theta \rangle \geq \delta s_A h \mid \mathcal{A}_t \in \mathcal{R}_k \cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{4\nu C_\mathrm{b}}}\right) \\
&\overset{(a)}{\leq} \mathbb{P}\left(\|\theta - \hat{\theta}\|_2 \geq \delta h \mid \mathcal{A}_t \in \mathcal{R}_k \cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{4\nu C_\mathrm{b}}}\right) + \mathbb{P}\left(\|\theta - \hat{\theta}\|_2 \geq \delta h \mid \mathcal{A}_t \in \mathcal{R}_k \cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{4\nu C_\mathrm{b}}}\right) \\
&= 2\mathbb{P}\left(\|\theta - \hat{\theta}\|_2 \geq \delta h \mid \mathcal{A}_t \in \mathcal{R}_k \cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{4\nu C_\mathrm{b}}}\right),
\end{aligned}
$$

where for (a), we used the Cauchy–Schwarz inequality. Let us denote $s' = s_0 + \frac{4\nu C_\mathrm{b}\sqrt{s_0}}{\phi_0^2}$. Then, using Lemma 3, we get:

$$
\begin{aligned}
\mathbb{P}\left(\|\theta - \hat{\theta}\|_2 \geq \delta h \mid \mathcal{A}_t \in \mathcal{R}_k \cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{4\nu C_\mathrm{b}}}\right) &\leq 2s' \exp\left(-\frac{\alpha^2 t \delta^2 h^2}{32\sigma^2 s_A^2 \nu^2 C_\mathrm{b}^2 s'}\right) \\
&= 2s' \exp\left(-h^2\right).
\end{aligned}
$$

We also trivially have that:

$$
\mathbb{P}\left(\|\theta - \hat{\theta}\|_2 \geq \delta h \mid \mathcal{A}_t \in \mathcal{R}_k \cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{4\nu C_\mathrm{b}}}\right) \leq 1.
$$

Therefore, we can upper bound the expected instantaneous regret as:

$$
\begin{aligned}
r_t^\pi &\leq \sum_{k=1}^{K} \mathbb{E}\left[r_t^\pi \mid \mathcal{A}_t \in \mathcal{R}_k\right] \mathbb{P}\left(\mathcal{A}_t \in \mathcal{R}_k\right) \\
&\leq \sum_{k=1}^{K} \left(\sum_{\ell \neq k} \left(\sum_{h=0}^{\lceil s_2/\delta \rceil} \left(2\delta s_A (h+1) \min\left\{1, 4s' \exp\left(-h^2\right)\right\}\right) \right) \right. \\
&\qquad \left. + 2(K-1) s_A s_2 \left(\mathbb{P}(\mathcal{E}_t^c) + \mathbb{P}\left(\left(\mathcal{G}_t^{\frac{\alpha}{4\nu C_\mathrm{b}}}\right)^c\right)\right) \cdot \mathbb{P}\left(\mathcal{A}_t \in \mathcal{R}_k\right) \right) \\
&\overset{(a)}{\leq} 2\delta s_A (K-1) \sum_{h=0}^{\lceil s_2/\delta \rceil} (h+1) \min\left\{1, 4s' \exp\left(-h^2\right)\right\} \\
&\qquad\qquad + 2(K-1) s_A s_2 \left(\mathbb{P}(\mathcal{E}_t^c) + \mathbb{P}\left(\left(\mathcal{G}_t^{\frac{\alpha}{4\nu C_\mathrm{b}}}\right)^c \middle| \mathcal{E}_t\right)\right) \\
&\leq 2\delta s_A (K-1) \left(\sum_{h=0}^{h_0} (h+1) + \sum_{h=h_0+1}^{h_{\max}} 4s'(h+1) \exp\left(-h^2\right) \right) \\
&\qquad\qquad + 2(K-1) s_A s_2 \left(\mathbb{P}(\mathcal{E}_t^c) + \mathbb{P}\left(\left(\mathcal{G}_t^{\frac{\alpha}{4\nu C_\mathrm{b}}}\right)^c \middle| \mathcal{E}_t\right)\right)
\end{aligned}
$$

where for (a), we used $\sum_{k=1}^{K} \mathbb{P}\left(\mathcal{A}_t \in \mathcal{R}_k\right) = 1$ and we set $h_0 \coloneqq \lfloor \sqrt{\log 4s'} + 1 \rfloor$. We have:

$$
\begin{aligned}
\sum_{h=h_0+1}^{h_{\max}} (h+1) \exp(-h^2) &= \sum_{h=h_0+1}^{h_{\max}} h \exp(-h^2) + \sum_{h=h_0+1}^{h_{\max}} \exp(-h^2) \\
&\leq \int_{h_0}^{\infty} h \exp(-h^2) dh + \int_{h_0}^{\infty} \exp(-h^2) dh.
\end{aligned}
$$

Since $h_0 \geq 1$ from $s_0 \geq 1$, we get:

$$
\begin{aligned}
\int_{h_0}^{\infty} h \exp(-h^2) dh &= \frac{1}{2} \exp(-h_0^2) \\
\int_{h_0}^{\infty} \exp(-h^2) dh &\leq \exp(-h_0^2).
\end{aligned}
$$

Therefore,

$$
\sum_{h=h_0+1}^{h_{\max}} (h+1) \exp(-h^2) \leq \frac{3}{2} \exp(-h_0^2).
$$

We get:

$$\begin{aligned}\sum_{h=0}^{h_0}(h+1) + \sum_{h=h_0+1}^{h_{\max}} 4s'(h+1)\exp\left(-h^2\right) &\le \frac{(h_0+1)(h_0+2)}{2} + 4s'\frac{3}{2}\exp(-h_0^2)\\ &\le \frac{h_0^2+3h_0+2}{2} + \frac{6s'}{4s'}\\ &\overset{(a)}{\le} \frac{9}{2}h_0^2,\end{aligned}$$

where for (a), we used $h_0 \ge 1$. Finally, we get:

$$\begin{aligned}r_t^\pi &\le 9\delta s_A(K-1)h_0^2 + 2(K-1)s_A s_2\left(\mathbb{P}(\mathcal{E}_t^c) + \mathbb{P}\left(\left(\mathcal{G}_t^{\frac{\alpha}{4\nu C_{\mathrm{b}}}}\right)^c \middle| \mathcal{E}_t\right)\right)\\ &\le \frac{36\sigma s_A(K-1)h_0^2\nu C_{\mathrm{b}}}{\alpha}\sqrt{\frac{2\left(s_0+\frac{4\nu C_{\mathrm{b}}\sqrt{s_0}}{\phi_0^2}\right)}{t-1}} + 2(K-1)s_A s_2\left(\mathbb{P}(\mathcal{E}_t^c) + \mathbb{P}\left(\left(\mathcal{G}_t^{\frac{\alpha}{4\nu C_{\mathrm{b}}}}\right)^c \middle| \mathcal{E}_t\right)\right).\end{aligned}$$

This concludes the proof. □

E Proof of Lemmas (without balanced covariance)

Lemma 15 *Let $t \ge \frac{2\log(2d^2)}{C_0^2}$ such that $4\left(\frac{2\nu s_0}{\phi_0^2} + \sqrt{\left(1+\frac{2\nu}{\phi_0^2}\right)s_0}\right)\lambda_t \le \theta_{\min}$. Under Assumptions 7, 2, 3, and 4, $\mathbb{P}\left(S \subset \hat{S}_1^{(t)} \text{ and } |\hat{S}_1^{(t)} \setminus S| \le \frac{2\nu\sqrt{s_0}}{\phi_0^2}\right) \ge 1 - 2\exp\left(-\frac{t\lambda_t^2}{32\sigma^2 s_A^2} + \log d\right) - \exp\left(-\frac{tC_0^2}{2}\right)$.*

Lemma 16 *Let $t \in [T]$. Under Assumptions 7 and 8, we have:*

$$\mathbb{P}\left(\lambda_{\min}(\hat{\Sigma}_{\hat{S}}) \ge \frac{\alpha}{2\nu} \mid \mathcal{E}_t\right) \ge 1 - \exp\left(\log\left(s_0 + \frac{2\nu\sqrt{s_0}}{\phi_0^2}\right) - \frac{t\alpha}{10s_A^2\nu}\right).$$

Lemma 17 *Let $t \in [T]$ and $s' = s_0 + 2\nu\sqrt{s_0}/\phi_0^2$. Under Assumption 7, we have for all $x, \lambda > 0$:*

$$\mathbb{P}\left(\|\hat{\theta}_{t+1} - \theta\|_2 \ge x \text{ and } \lambda_{\min}(\hat{\Sigma}_{\hat{S}}) \ge \lambda \mid \mathcal{E}_t\right) \le 2s'\exp\left(-\frac{\lambda^2 t x^2}{2\sigma^2 s_A^2 s'}\right).$$

Lemma 18 *Define $\mathcal{G}_t^{\frac{\alpha}{2\nu}} \coloneqq \left\{\lambda_{\min}(\hat{\Sigma}_{\hat{S}}) \ge \frac{\alpha}{2\nu}\right\}$. Let $t \ge 2$. Under Assumptions 7, 2, 3, 4, 8, and 6, the expected instantaneous regret $\mathbb{E}[\max_{A\in\mathcal{A}_t}\langle A - A_t, \theta\rangle]$ is upper bounded by: $\frac{352\sigma^2 s_A^4 C_{\mathrm{m}} h_0^3\nu^2\left(s_0+\frac{2\nu\sqrt{s_0}}{\phi_0^2}\right)}{\alpha^2}\frac{1}{t-1} + 2s_A s_2\left(\mathbb{P}(\mathcal{E}_t^c) + \mathbb{P}\left(\left(\mathcal{G}_t^{\frac{\alpha}{2\nu}}\right)^c \middle| \mathcal{E}_t\right)\right)$.*

Lemma 19 *Under Assumptions 7, 2, 3, 4, and 8, for any $t \in [T]$, $\mathbb{E}[\max_{A\in\mathcal{A}_t}\langle A - A_t, \theta\rangle]$ is upper bounded by: $\frac{18\sigma s_A h_0^2\nu}{\alpha}\sqrt{\frac{2\left(s_0+\frac{2\nu\sqrt{s_0}}{\phi_0^2}\right)}{t-1}} + 2s_A s_2\left(\mathbb{P}(\mathcal{E}_t^c) + \mathbb{P}\left(\left(\mathcal{G}_t^{\frac{\alpha}{2\nu}}\right)^c \middle| \mathcal{E}_t\right)\right)$.*

E.1 Proof of Lemma 15

We define $v \coloneqq \hat{\theta}_0^{(t)} - \theta$. We first analyze the performance of the initial Lasso estimate.

Lemma 20 *Let $\hat{\Sigma}_t \coloneqq \frac{\sum_{s=1}^t A_s A_s^\top}{t}$ be the empirical covariance matrix of the selected context vectors. Suppose $\hat{\Sigma}_t$ satisfies the compatibility condition with the support S with the compatibility constant ϕ_t. Then, under Assumption 7, we have:*

$$\mathbb{P}\left(\|v\|_1 \le \frac{4s_0\lambda_t}{\phi_t^2}\right) \ge 1 - 2\exp\left(-\frac{t\lambda_t^2}{32\sigma^2 s_A^2} + \log d\right).$$

The next lemma then states that the compatibility constant of $\hat{\Sigma}_t$ does not deviate much from the compatibility constant of Σ.

Lemma 21 *Assume $K = 2$. Let $C_0 \coloneqq \min\left\{\frac{1}{2}, \frac{\phi_0^2}{256 s_0 s_A^2\nu}\right\}$. For all $t \ge \frac{2\log(2d^2)}{C_0^2}$, we have:*

$$\mathbb{P}\left(\phi^2(\hat{\Sigma}_t, S) \ge \frac{\phi_0^2}{2\nu}\right) \ge 1 - \exp\left(-\frac{tC_0^2}{2}\right).$$

Then, we follow the steps of the proof given by [23]. Let us define the event $\mathcal{G}_t$ as:

$$\mathcal{G}_t := \left\{ \|v\|_1 \leq \frac{4s_0\lambda_t}{\phi_t^2} \right\}.$$

For the rest of this section, we assume that the event $\mathcal{G}_t$ holds. Note that:

$$\begin{aligned}
\|v\|_1 &\geq \|v_{S^c}\|_1 \\
&= \sum_{j\in S^c} |(\hat{\theta}_0^{(t)})_j| \\
&\geq \sum_{j\in S^c\cap \hat{S}_0^{(t)}} |(\hat{\theta}_0^{(t)})_j| \\
&= \sum_{j\in \hat{S}_0^{(t)}\setminus S} |(\hat{\theta}_0^{(t)})_j| \\
&\overset{(a)}{\geq} |\hat{S}_0^{(t)} \setminus S| 4\lambda_t,
\end{aligned}$$

where for (a), we used the construction of $\hat{S}_0^{(t)}$ in the algorithm. We get:

$$|\hat{S}_0^{(t)} \setminus S| \leq \frac{\|v\|_1}{4\lambda_t} \overset{(a)}{\leq} \frac{s_0}{\phi_t^2},$$

where for (a), we used the definition of $\mathcal{G}_t$.

We have: $\forall j \in S$,

$$\begin{aligned}
|(\hat{\theta}_0^{(t)})_j| &\geq \theta_{\min} - \|v_S\|_\infty \\
&\geq \theta_{\min} - \|v_S\|_1 \\
&\geq \theta_{\min} - \frac{4s_0\lambda_t}{\phi_t^2}.
\end{aligned}$$

Therefore, when t is large enough so that $4\lambda_t \leq \theta_{\min} - \frac{4s_0\lambda_t}{\phi_t^2}$, we have: $S \subset \hat{S}_0^{(t)}$. Using a similar argument, when t is large enough so that $4\lambda_t\sqrt{\left(1+\frac{1}{\phi_t^2}\right)s_0} \leq \theta_{\min} - \frac{4s_0\lambda_t}{\phi_t^2}$, it holds that $S \subset \hat{S}_1^{(t)}$. From the construction of $\hat{S}_1^{(t)}$ in the algorithm, it also holds that: $\hat{S}_1^{(t)} \subset \hat{S}_0^{(t)}$. Therefore,

$$\begin{aligned}
\|v\|_1 &\geq \sum_{i\in \hat{S}_0^{(t)}\setminus S} |(\hat{\theta}_0^{(t)})_i| \\
&\geq \sum_{i\in \hat{S}_1^{(t)}\setminus S} |(\hat{\theta}_0^{(t)})_i| \\
&\geq |\hat{S}_1^{(t)} \setminus S| 4\lambda_t \sqrt{|\hat{S}_0^{(t)}|},
\end{aligned}$$

and

$$\begin{aligned}
|\hat{S}_1^{(t)} \setminus S| &\leq \frac{\|v\|_1}{4\lambda_t\sqrt{|\hat{S}_0^{(t)}|}} \\
&\leq \frac{1}{4\lambda_t\sqrt{|\hat{S}_0^{(t)}|}} \cdot \frac{4s_0\lambda_t}{\phi_t^2} \\
&\leq \frac{\sqrt{s_0}}{\phi_t^2}.
\end{aligned}$$

Note that the condition $4\lambda_t\sqrt{\left(1+\frac{1}{\phi_t^2}\right)s_0} \leq \theta_{\min} - \frac{4s_0\lambda_t}{\phi_t^2}$ is equivalent to $4\lambda_t\left(\sqrt{\left(1+\frac{1}{\phi_t^2}\right)s_0} + \frac{s_0}{\phi_t^2}\right) \leq \theta_{\min}$. This concludes the proof of Lemma 15 by substituting $\phi_t^2 = \phi_0^2/(2\nu)$. □

E.2 Proof of Lemmas used in the proof of Lemma 15

E.2.1 Proof of Lemma 20

The proof is identical to that of Lemma 6.

E.2.2 Proof of Lemma 21

First we define the adapted Gram matrix $\Sigma_t := \frac{1}{t}\sum_{s=1}^t \mathbb{E}[A_s A_s^\top | \mathcal{F}_{s-1}]$. From the construction of the algorithm, $\mathbb{E}[A_s A_s^\top | \mathcal{F}_{s-1}] = \mathbb{E}[\sum_{k=1}^K A_{s,k} A_{s,k}^\top \mathbb{1}\{k = \operatorname{argmax}_{k'} \langle A_{s,k}, \hat{\theta}_s \rangle\} | \, \hat{\theta}_s]$. The following lemma characterizes the expected Gram matrix generated by the algorithm.

Lemma 22 *Assume $K = 2$. Under Assumption 2 and Assumption 3, for each fixed vector $\theta' \in \mathbb{R}^d$, we have:*

$$\mathbb{E}_{\mathcal{A}\sim p_A}\left[\sum_{k=1,2} A_k A_k^\top \mathbb{1}\{k = \operatorname*{argmax}_{k'} \langle A_k, \theta' \rangle\}\right] \succeq \frac{1}{\nu}\Sigma,$$

where $A \succeq B$ means that $A - B$ is positive semidefinite.

Using Lemma 22, we have

$$\Sigma_t \succeq \frac{1}{\nu}\Sigma. \tag{8}$$

By Lemma 6.18 of [6], Assumption 2, and the definition of the compatibility constant, we get:

$$\phi^2(\Sigma_t, S) \geq \phi^2(\frac{1}{\nu}\Sigma, S) \geq \frac{\phi_0^2}{\nu}. \tag{9}$$

Furthermore, we have a following adaptive matrix concentration results for $\hat{\Sigma}_t$:

Lemma 23 *Let $C_0 := \min\left\{\frac{1}{2}, \frac{\phi_0^2}{256 s_0 s_A^2 \nu}\right\}$. We have, for all $t \geq \frac{2\log(2d^2)}{C_0^2}$,*

$$\mathbb{P}\left(\frac{1}{2s_A^2}\|\hat{\Sigma}_t - \Sigma_t\|_\infty \geq \frac{\phi^2(\Sigma_t, S)}{64 s_0 s_A^2 \nu}\right) \leq \exp\left(-\frac{tC_0^2}{2}\right).$$

Combining Lemmas 23 and 12, we get, for all $t \geq \frac{2\log(2d^2)}{C_0^2}$:

$$\begin{aligned}\phi^2(\hat{\Sigma}_t, S) &\geq \frac{\phi^2(\Sigma_t, S)}{2} \\ &\geq \frac{\phi_0^2}{2\nu},\end{aligned}$$

with probability at least $1 - \exp\left(-\frac{tC_0^2}{2}\right)$. This concludes the proof. □

E.2.3 Proof of Lemma 22

The proof is almost identical to the proof of Lemma 2 in [15]. □

E.2.4 Proof of Lemma 23

Let us define $\gamma_t^{ij}(A_t)$ as:

$$\gamma_t^{ij}(A_t) := \frac{1}{2C_A^2}\left((A_t)_i(A_t)_j - \mathbb{E}[(A_t)_i(A_t)_j \mid \mathcal{F}_{t-1}]\right),$$

where $(A_t)_i$ is the i-th element of A_t.

Note that $\frac{1}{2s_A^2}\|\hat{\Sigma}_t - \Sigma_t\|_\infty = \max_{1\leq i\leq j\leq d}\left|\frac{1}{t}\sum_{s=1}^t \gamma_s^{ij}(A_s)\right|$, $\mathbb{E}[\gamma_t^{ij}(A_t)|\mathcal{F}_{t-1}] = 0$, and $\mathbb{E}[|\gamma_t^{ij}(A_t)|^m \mid \mathcal{F}_{t-1}] \leq 1$ for all integer $m \geq 2$. Therefore, we can apply Lemma 13:

$$\mathbb{P}\left(\frac{1}{2s_A^2}\|\hat{\Sigma}_t - \Sigma_t\|_\infty \geq x + \sqrt{2x} + \sqrt{\frac{4\log(2d^2)}{t}} + \frac{2\log(2d^2)}{t}\right) \leq \exp\left(-\frac{tx}{2}\right).$$

For all $t \geq \frac{2\log(2d^2)}{C_0^2}$ with $C_0 := \min\left\{\frac{1}{2}, \frac{\phi_0^2}{256 s_0 s_A^2 \nu}\right\}$, taking $x = C_0^2$,

$$\begin{aligned}x + \sqrt{2x} + \sqrt{\frac{4\log(2d^2)}{t}} + \frac{2\log(2d^2)}{t} &\leq 2C_0^2 + 2\sqrt{2}C_0 \\ &\leq 4C_0 \\ &\leq \frac{\phi_0^2}{64 s_0 s_A^2 \nu} \\ &\leq \frac{\phi^2(\Sigma_t, S)}{64 s_0 s_A^2 \nu}.\end{aligned}$$

In summary, for all $t \geq \frac{2\log(2d^2)}{C_0^2}$, we get:

$$\mathbb{P}\left(\frac{1}{2s_A^2}\|\hat{\Sigma}_t - \Sigma_t\|_\infty \geq \frac{\phi^2(\Sigma_t, S)}{64 s_0 s_A^2 \nu}\right) \leq \mathbb{P}\left(\frac{1}{2s_A^2}\|\hat{\Sigma}_t - \Sigma_t\|_\infty \geq C_0^2 + \sqrt{2}C_0 + \sqrt{\frac{4\log(2d^2)}{t}} + \frac{2\log(2d^2)}{t}\right)$$
$$\leq \exp\left(-\frac{tC_0^2}{2}\right).$$

This concludes the proof. □

E.3 Proof of Lemma 16

For a fixed $\hat{S}$, first we define the adapted Gram matrix on the estimated support as $\Sigma_t := \frac{1}{t}\sum_{s=1}^t \mathbb{E}[A_s(\hat{S})A_s(\hat{S})^\top|\mathcal{F}_{s-1}]$. From the construction of the algorithm, $\mathbb{E}[A_s(\hat{S})A_s(\hat{S})^\top|\mathcal{F}_{s-1}] = \mathbb{E}[\sum_{k=1}^K A_{s,k}(\hat{S})A_{s,k}(\hat{S})^\top \mathbb{1}\{k = \operatorname{argmax}_{k'}\langle A_{s,k}, \hat{\theta}_s\rangle\}|\ \hat{\theta}_s]$. Recall that for each $B \subset [d]$, $\Sigma_B := \frac{1}{K}\sum_{k=1}^K \mathbb{E}_{\mathcal{A}\sim p_A}\left[A_k(B)A_k(B)^\top\right]$, where $A_k(B)$ is a $|B|$-dimensional vector extracted the elements of A_k with indices in B. The following lemma characterizes the expected Gram matrix generated by the algorithm.

Lemma 24 *Assume $K = 2$. Fix $\hat{S}$ such that $S \subset \hat{S}$ and $|\hat{S}| \leq s_0 + (2\nu\sqrt{s_0})/\phi_0^2$. Fix $\theta' \in \mathbb{R}^d$. Under Assumption 2 and Assumption 3, we have:*

$$\mathbb{E}_{\mathcal{A}\sim p_A}\left[\sum_{k=1,2} A_k(\hat{S})A_k(\hat{S})^\top \mathbb{1}\{k = \operatorname*{argmax}_{k'}\langle A_k, \theta'_{\hat{S}}\rangle\}\right] \succeq \frac{1}{\nu}\Sigma_{\hat{S}},$$

where $A \succeq B$ means that $A - B$ is positive semidefinite.

First, we prove the lower bound on the smallest eigenvalue of the expected covariance matrices. Let $\Sigma_{\hat{S}} := \frac{1}{t}\sum_{s=1}^t \mathbb{E}[A_s(\hat{S})A_s(\hat{S})^\top \mid \mathcal{F}_{t-1}]$. By Assumption 8 and the construction of the algorithm, under the event $\mathcal{E}_t$, we get:

$$\begin{aligned}\lambda_{\min}(\Sigma_{\hat{S}}) &= \lambda_{\min}\left(\frac{1}{t}\sum_{s=1}^t \mathbb{E}\left[\sum_{k=1}^K A_{s,k,\hat{S}}A_{s,k,\hat{S}}\mathbb{1}\{k = \operatorname*{argmax}_{k'}\langle A_{k'}, \hat{\theta}_s\rangle\} \mid \hat{\theta}_s\right]\right)\\ &\geq \sum_{s=1}^t \lambda_{\min}\left(\frac{1}{t}\mathbb{E}\left[\sum_{k=1}^K A_{s,k,\hat{S}}A_{s,k,\hat{S}}\mathbb{1}\{k = \operatorname*{argmax}_{k'}\langle A_{k'}, \hat{\theta}_s\rangle\} \mid \hat{\theta}_s\right]\right)\\ &= \frac{\alpha}{\nu},\end{aligned}$$

where for the inequality, we used the concavity of $\lambda_{\min}(\cdot)$ over the positive semi-definite matrices. Next, we prove the upper bound on the largest eigenvalue of $A_s(\hat{S})A_s(\hat{S})^\top$:

$$\begin{aligned}\lambda_{\max}(A_s(\hat{S})A_s(\hat{S})^\top) &= \max_{\|v\|=1} v^\top A_s(\hat{S})A_s(\hat{S})\top v\\ &\leq \max_{\|v\|=1}\|v\|_2^2\|A_s(\hat{S})\|_2^2\\ &\leq s_A^2,\end{aligned}$$

where for the second inequality, we used the Cauchy-Schwarz inequality and Assumption 7. Taking $R = s_A^2$, $X_s = A_{s,\hat{S}}A_{s,\hat{S}}^\top$, $Y = t\hat{\Sigma}_{\hat{S}}$, $W = t\Sigma_{\hat{S}}$, $\delta = 1/2$, $\mu = t\frac{\alpha}{\nu}$ in Theorem 6, we have:

$$\begin{aligned}\mathbb{P}\left(\lambda_{\min}(t\hat{\Sigma}_{\hat{S}}) \leq \frac{1}{2}t\frac{\alpha}{\nu} \text{ and } \lambda_{\min}(t\Sigma_{\hat{S}}) \geq t\frac{\alpha}{\nu}\right) &\leq \left(s_0 + \frac{2\nu\sqrt{s_0}}{\phi_0^2}\right)\left(\frac{e^{-0.5}}{0.5^{0.5}}\right)^{\frac{t\alpha}{s_A^2\nu}}\\ &\leq \exp\left(\log\left(s_0 + \frac{2\nu\sqrt{s_0}}{\phi_0^2}\right) - \frac{t\alpha}{10 s_A^2\nu}\right),\end{aligned}$$

where for the last inequality, we used $-0.5 - 0.5\log(0.5) < -\frac{1}{10}$. This concludes the proof. □

E.3.1 Proof of Lemma 24

The proof is almost identical to the proof of Lemma 2 in [15].

E.4 Proof of Lemma 17

In this proof, we denote $\hat{S} = \hat{S}_1^{(t)}$ and $\varepsilon = (\varepsilon_1, \dots, \varepsilon_t)^\top$. Assume $\lambda_{\min}(\hat{\Sigma}_{\hat{S}}) \geq \lambda$. We have:

$$\begin{aligned}
\|\hat{\theta}_{t+1} - \theta\|_2 &= \|(A(\hat{S})^\top A(\hat{S}))^{-1} A(\hat{S})^\top R - \theta\|_2 \\
&= \|(A(\hat{S})^\top A(\hat{S}))^{-1} A(\hat{S})^\top (A\theta + \varepsilon) - \theta\|_2 \\
&= \|(A(\hat{S})^\top A(\hat{S}))^{-1} A(\hat{S})^\top (A(\hat{S})\theta(\hat{S}) + \varepsilon) - \theta\|_2 \\
&= \|(A(\hat{S})^\top A(\hat{S}))^{-1} A(\hat{S})^\top \varepsilon\|_2 \\
&\leq \|(A(\hat{S})^\top A(\hat{S}))^{-1}\|_2 \|A(\hat{S})^\top \varepsilon\|_2 \\
&\leq \frac{1}{\lambda t} \|A(\hat{S})^\top \varepsilon\|_2.
\end{aligned}$$

We get (note that we are conditioning on a fixed $\hat{S}$ during the proof):

$$\begin{aligned}
\mathbb{P}\left(\|\hat{\theta}_{t+1} - \theta\|_2 \geq x \text{ and } \lambda_{\min}(\hat{\Sigma}_{\hat{S}}) \geq \lambda\right) &= \mathbb{P}\left(\|\hat{\theta}_{t+1} - \theta\|_2 \geq x \;\middle|\; \lambda_{\min}(\hat{\Sigma}_{\hat{S}}) \geq \lambda\right) \mathbb{P}(\lambda_{\min}(\hat{\Sigma}_{\hat{S}}) \geq \lambda) \\
&\leq \mathbb{P}\left(\|A(\hat{S})^\top \varepsilon\|_2 \geq \lambda t x \;\middle|\; \lambda_{\min}(\hat{\Sigma}_{\hat{S}}) \geq \lambda\right) \mathbb{P}(\lambda_{\min}(\hat{\Sigma}_{\hat{S}}) \geq \lambda) \\
&\leq \mathbb{P}\left(\|A(\hat{S})^\top \varepsilon\|_2 \geq \lambda t x\right) \\
&\leq \sum_{i=1}^{d} \mathbb{P}\left(\left|\sum_{s=1}^{t} \varepsilon_s (A_s)_i \mathbb{1}\left\{i \in \hat{S}\right\}\right| \geq \frac{\lambda t x}{\sqrt{s_0 + \frac{2\nu\sqrt{s_0}}{\phi_0^2}}}\right) \\
&= \sum_{i \in \hat{S}} \mathbb{P}\left(\left|\sum_{s=1}^{t} \varepsilon_s (A_s)_i\right| \geq \frac{\lambda t x}{\sqrt{s_0 + \frac{2\nu\sqrt{s_0}}{\phi_0^2}}}\right) \\
&\overset{(a)}{\leq} 2\left(s_0 + \frac{2\nu\sqrt{s_0}}{\phi_0^2}\right) \exp\left(-\frac{\lambda^2 t x^2}{2\sigma^2 s_A^2 \left(s_0 + \frac{2\nu\sqrt{s_0}}{\phi_0^2}\right)}\right),
\end{aligned}$$

where for (a), we used Theorem 5. This concludes the proof. □

E.5 Proof of Lemma 18

We follow the proof strategy of Lemma 6 in [5]. Let r_t^π be the instantaneous expected regret of algorithm π at round t defined as:

$$r_t^\pi := \mathbb{E}\left[\max_{A \in \mathcal{A}_t} \langle A - A_t, \theta\rangle\right].$$

Let us define the events $\mathcal{R}_k := \{\mathcal{A}_t \in \mathbb{R}^{K \times d} : k \in \operatorname*{argmax}_{k'} \langle A_{t,k'}, \theta\rangle\}$ and $\mathcal{G}_t^\lambda := \left\{\lambda_{\min}(\hat{\Sigma}_{\hat{S}}) \geq \lambda\right\}$. We have:

$$r_t^\pi \leq \sum_{k=1}^{2} \mathbb{E}\left[r_t^\pi \mid \mathcal{A}_t \in \mathcal{R}_k\right] \mathbb{P}\left(\mathcal{A}_t \in \mathcal{R}_k\right).$$

The term $\mathbb{E}\left[r_t^\pi \mid \mathcal{A}_t \in \mathcal{R}_k\right]$ can be further computed as:

$$\begin{aligned}
\mathbb{E}\left[r_t^\pi \mid \mathcal{A}_t \in \mathcal{R}_k\right] &= \mathbb{E}\left[\langle A_{t,k} - A_t, \theta\rangle \mid \mathcal{A}_t \in \mathcal{R}_k\right] \\
&\leq \mathbb{E}\left[\mathbb{1}\left\{\langle A_t, \hat{\theta}_t\rangle \geq \langle A_{t,k}, \hat{\theta}_t\rangle\right\} \langle A_{t,k} - A_t, \theta\rangle \mid \mathcal{A}_t \in \mathcal{R}_k\right] \\
&\leq \sum_{\ell \neq k} \mathbb{E}\left[\mathbb{1}\left\{\langle A_{t,\ell}, \hat{\theta}_t\rangle \geq \langle A_{t,k}, \hat{\theta}_t\rangle\right\} \langle A_{t,k} - A_{t,\ell}, \theta\rangle \mid \mathcal{A}_t \in \mathcal{R}_k\right] \\
&\leq \sum_{\ell \neq k} \mathbb{E}\left[\mathbb{1}\left\{\langle A_{t,\ell}, \hat{\theta}_t\rangle \geq \langle A_{t,k}, \hat{\theta}_t\rangle\right\} \langle A_{t,k} - A_{t,\ell}, \theta\rangle \mid \mathcal{A}_t \in \mathcal{R}_k,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{2\nu}}\right] \\
&\quad + 2 s_A s_2 \left(\mathbb{P}(\mathcal{E}_t^c) + \mathbb{P}\left(\left(\mathcal{G}_t^{\frac{\alpha}{2\nu}}\right)^c \middle| \mathcal{E}_t\right)\right).
\end{aligned}$$

Let us denote the event $I_h := \{\mathcal{A}_t \in \mathbb{R}^{K \times d} : \langle A_{t,k} - A_{t,\ell}, \theta\rangle \in (2\delta s_A h, 2\delta s_A (h+1)]\}$ where

$$\delta = \frac{\sigma s_A \nu}{\alpha} \sqrt{\frac{8\left(s_0 + \frac{2\nu\sqrt{s_0}}{\phi_0^2}\right)}{t-1}}.$$

By conditioning on I_h, we get:

$$
\begin{aligned}
&\mathbb{E}\left[\mathbb{1}\left\{\langle A_{t,\ell},\hat{\theta}_t\rangle \geq \langle A_{t,k},\hat{\theta}_t\rangle\right\}\langle A_{t,k}-A_{t,\ell},\theta\rangle \mid \mathcal{A}_t\in\mathcal{R}_k\cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{2\nu}}\right]\\
&\leq \sum_{h=0}^{\lceil s_2/\delta\rceil}\mathbb{E}\left[\mathbb{1}\left\{\langle A_{t,\ell},\hat{\theta}_t\rangle \geq \langle A_{t,k},\hat{\theta}_t\rangle\right\}\langle A_{t,k}-A_{t,\ell},\theta\rangle \mid \mathcal{A}_t\in\mathcal{R}_k\cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{2\nu}}\right]\mathbb{P}(\mathcal{A}_t\in I_h)\\
&\overset{(a)}{\leq} \sum_{h=0}^{\lceil s_2/\delta\rceil} 2\delta s_A(h+1)\,\mathbb{E}\left[\mathbb{1}\left\{\langle A_{t,\ell},\hat{\theta}_t\rangle \geq \langle A_{t,k},\hat{\theta}_t\rangle\right\} \mid \mathcal{A}_t\in\mathcal{R}_k\cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{2\nu}}\right]\\
&\qquad\qquad\qquad \times \mathbb{P}\left(\langle A_{t,k}-A_{t,\ell},\theta\rangle \in (0,2\delta s_A(h+1)]\right)\\
&\overset{(b)}{\leq} \sum_{h=0}^{\lceil s_2/\delta\rceil} 4\delta^2 s_A^2(h+1)^2 C_{\mathrm{m}}\mathbb{P}\left(\langle A_{t,\ell},\hat{\theta}_t\rangle \geq \langle A_{t,k},\hat{\theta}_t\rangle \mid \mathcal{A}_t\in\mathcal{R}_k\cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{2\nu}}\right),
\end{aligned}
$$

where for (a), we used the definition of I_h and for (b), we used Assumption 6. Under the event $\mathcal{A}_t\in I_h$, the event $\langle A_{t,\ell},\hat{\theta}_t\rangle \geq \langle A_{t,k},\hat{\theta}_t\rangle$ happens only when at least one of the events $\langle A_{t,k},\theta-\hat{\theta}\rangle \geq \delta s_A h$ or $\langle A_{t,\ell},\hat{\theta}-\theta\rangle \geq \delta s_A h$ holds. Therefore,

$$
\begin{aligned}
&\mathbb{P}\left(\langle A_{t,\ell},\hat{\theta}_t\rangle \geq \langle A_{t,k},\hat{\theta}_t\rangle \mid \mathcal{A}_t\in\mathcal{R}_k\cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{2\nu}}\right)\\
&\leq \mathbb{P}\left(\langle A_{t,k},\theta-\hat{\theta}\rangle \geq \delta s_A h \mid \mathcal{A}_t\in\mathcal{R}_k\cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{2\nu}}\right)\\
&\qquad\qquad + \mathbb{P}\left(\langle A_{t,\ell},\hat{\theta}-\theta\rangle \geq \delta s_A h \mid \mathcal{A}_t\in\mathcal{R}_k\cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{2\nu}}\right)\\
&\overset{(a)}{\leq} \mathbb{P}\left(\|\theta-\hat{\theta}\|_2 \geq \delta h \mid \mathcal{A}_t\in\mathcal{R}_k\cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{2\nu}}\right) + \mathbb{P}\left(\|\theta-\hat{\theta}\|_2 \geq \delta h \mid \mathcal{A}_t\in\mathcal{R}_k\cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{2\nu}}\right)\\
&= 2\mathbb{P}\left(\|\theta-\hat{\theta}\|_2 \geq \delta h \mid \mathcal{A}_t\in\mathcal{R}_k\cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{2\nu}}\right),
\end{aligned}
$$

where for (a), we used the Cauchy–Schwarz inequality. Let us denote $s' = s_0 + \frac{2\nu\sqrt{s_0}}{\phi_0^2}$. Then, using Lemma 17, we get:

$$
\begin{aligned}
\mathbb{P}\left(\|\theta-\hat{\theta}\|_2 \geq \delta h \mid \mathcal{A}_t\in\mathcal{R}_k\cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{2\nu}}\right) &\leq 2s'\exp\left(-\frac{\alpha^2 t\delta^2 h^2}{8\sigma^2 s_A^2\nu^2 s'}\right)\\
&= 2s'\exp\left(-h^2\right).
\end{aligned}
$$

We also trivially have that:

$$
\mathbb{P}\left(\|\theta-\hat{\theta}\|_2 \geq \delta h \mid \mathcal{A}_t\in\mathcal{R}_k\cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{2\nu}}\right) \leq 1.
$$

Therefore, we can upper bound the expected instantaneous regret as:

$$
\begin{aligned}
r_t^\pi &\leq \sum_{k=1}^{2}\mathbb{E}\left[r_t^\pi \mid \mathcal{A}_t\in\mathcal{R}_k\right]\mathbb{P}\left(\mathcal{A}_t\in\mathcal{R}_k\right)\\
&\leq \sum_{k=1}^{2}\left(\sum_{\ell\neq k}\left(\sum_{h=0}^{\lceil s_2/\delta\rceil}\left(4\delta^2 s_A^2(h+1)^2 C_{\mathrm{m}}\min\left\{1,4s'\exp\left(-h^2\right)\right\}\right)\right)\right.\\
&\qquad \left.+2s_A s_2\left(\mathbb{P}(\mathcal{E}_t^c)+\mathbb{P}\left(\left(\mathcal{G}_t^{\frac{\alpha}{2\nu}}\right)^c\right)\right)\cdot\mathbb{P}\left(\mathcal{A}_t\in\mathcal{R}_k\right)\right)\\
&\overset{(a)}{\leq} 4\delta^2 s_A^2 C_{\mathrm{m}}\sum_{h=0}^{\lceil s_2/\delta\rceil}(h+1)^2\min\left\{1,4s'\exp\left(-h^2\right)\right\} + 2s_A s_2\left(\mathbb{P}(\mathcal{E}_t^c)+\mathbb{P}\left(\left(\mathcal{G}_t^{\frac{\alpha}{2\nu}}\right)^c\middle|\mathcal{E}_t\right)\right)\\
&\leq 4\delta^2 s_A^2 C_{\mathrm{m}}\left(\sum_{h=0}^{h_0}(h+1)^2 + \sum_{h=h_0+1}^{h_{\max}}4s'(h+1)^2\exp\left(-h^2\right)\right) + 2s_A s_2\left(\mathbb{P}(\mathcal{E}_t^c)+\mathbb{P}\left(\left(\mathcal{G}_t^{\frac{\alpha}{2\nu}}\right)^c\middle|\mathcal{E}_t\right)\right)
\end{aligned}
$$

where for (a), we used $\sum_{k=1}^{2}\mathbb{P}\left(\mathcal{A}_t\in\mathcal{R}_k\right)=1$ from Assumption 6 and we set $h_0 := \lfloor\sqrt{\log 4s'}+1\rfloor$. We have:

$$
\begin{aligned}
\sum_{h=h_0+1}^{h_{\max}}(h+1)^2\exp(-h^2) &= \sum_{h=h_0+1}^{h_{\max}}h^2\exp(-h^2) + 2\sum_{h=h_0+1}^{h_{\max}}h\exp(-h^2) + \sum_{h=h_0+1}^{h_{\max}}\exp(-h^2)\\
&\leq \int_{h_0}^{\infty}h^2\exp(-h^2)dh + 2\int_{h_0}^{\infty}h\exp(-h^2)dh + \int_{h_0}^{\infty}\exp(-h^2)dh.
\end{aligned}
$$

Using an integration by parts, the inequality $\int_{h_0}^{\infty} \exp(-h_0^2)dh \leq \exp(-h_0^2)/(h_0 + \sqrt{h_0^2 + 4/\pi}) \leq \exp(-h_0^2)$, and $h_0 \geq 1$ from $s_0 \geq 1$, we get:

$$\begin{aligned} \int_{h_0}^{\infty} h^2 \exp(-h^2)dh &\leq \frac{1}{2} h_0 \exp(-h_0^2) + \frac{1}{2} \exp(-h_0^2) \\ 2\int_{h_0}^{\infty} h \exp(-h^2)dh &= \exp(-h_0^2) \\ \int_{h_0}^{\infty} \exp(-h^2)dh &\leq \exp(-h_0^2). \end{aligned}$$

Therefore,

$$\begin{aligned} \sum_{h=h_0+1}^{h_{\max}} (h+1)^2 \exp(-h^2) &\leq \frac{1}{2} h_0 \exp(-h_0^2) + \frac{5}{2} \exp(-h_0^2) \\ &\leq h_0 \exp(-h_0^2) + 5 \exp(-h_0^2). \end{aligned}$$

We get:

$$\begin{aligned} \sum_{h=0}^{h_0} (h+1)^2 + \sum_{h=h_0+1}^{h_{\max}} 4s'(h+1)^2 \exp\left(-h^2\right) &\leq \frac{(h_0+1)(h_0+2)(2h_0+3)}{6} + 4s'(h_0+5)\exp(-h_0^2) \\ &\leq \frac{2h_0^3 + 9h_0^2 + 13h_0 + 6}{6} + 4s'(h_0+5)\frac{1}{4s'} \\ &\overset{(a)}{\leq} 11h_0^3, \end{aligned}$$

where for (a), we used $h_0 \geq 1$. Finally, we get:

$$\begin{aligned} r_t^\pi &\leq 44\delta^2 s_A^2 C_{\mathrm{m}} h_0^3 + 2s_A s_2 \left(\mathbb{P}(\mathcal{E}_t^c) + \mathbb{P}\left(\left(\mathcal{G}_t^{\frac{\alpha}{2\nu}}\right)^c \middle| \mathcal{E}_t\right)\right) \\ &\leq \frac{352\sigma^2 s_A^4 C_{\mathrm{m}} h_0^3 \nu^2 \left(s_0 + \frac{2\nu\sqrt{s_0}}{\phi_0^2}\right)}{\alpha^2} \frac{1}{t-1} + 2s_A s_2 \left(\mathbb{P}(\mathcal{E}_t^c) + \mathbb{P}\left(\left(\mathcal{G}_t^{\frac{\alpha}{2\nu}}\right)^c \middle| \mathcal{E}_t\right)\right). \end{aligned}$$

This concludes the proof. □

E.6 Proof of Lemma 19

Let r_t^π be the instantaneous expected regret of algorithm π in round t defined as:

$$r_t^\pi \coloneqq \mathbb{E}\left[\max_{A \in \mathcal{A}_t} \langle A - A_t, \theta \rangle\right].$$

Let us define the events $\mathcal{R}_k \coloneqq \{\mathcal{A}_t \in \mathbb{R}^{K \times d} : k \in \underset{k'}{\operatorname{argmax}} \langle A_{t,k'}, \theta \rangle\}$ and $\mathcal{G}_t^\lambda \coloneqq \left\{\lambda_{\min}(\hat{\Sigma}_{\hat{S}}) \geq \lambda\right\}$. As in the proof of Lemma 4, we get:

$$r_t^\pi \leq \sum_{k=1}^{2} \mathbb{E}\left[r_t^\pi \mid \mathcal{A}_t \in \mathcal{R}_k\right] \mathbb{P}\left(\mathcal{A}_t \in \mathcal{R}_k\right),$$

and

$$\begin{aligned} \mathbb{E}\left[r_t^\pi \mid \mathcal{A}_t \in \mathcal{R}_k\right] &= \mathbb{E}\left[\langle A_{t,k} - A_t, \theta \rangle \mid \mathcal{A}_t \in \mathcal{R}_k\right] \\ &\leq \sum_{\ell \neq k} \mathbb{E}\left[\mathbb{1}\left\{\langle A_{t,\ell}, \hat{\theta}_t \rangle \geq \langle A_{t,k}, \hat{\theta}_t \rangle\right\} \langle A_{t,k} - A_{t,\ell}, \theta \rangle \mid \mathcal{A}_t \in \mathcal{R}_k,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{2\nu}}\right] \\ &\quad + 2s_A s_2 \left(\mathbb{P}(\mathcal{E}_t^c) + \mathbb{P}\left(\left(\mathcal{G}_t^{\frac{\alpha}{2\nu}}\right)^c \middle| \mathcal{E}_t\right)\right). \end{aligned}$$

Let us denote the event $I_h \coloneqq \{\mathcal{A}_t \in \mathbb{R}^{K \times d} : \langle A_{t,k} - A_{t,\ell}, \theta \rangle \in (2\delta s_A h, 2\delta s_A(h+1)]\}$ where

$$\delta = \frac{\sigma s_A \nu}{\alpha} \sqrt{\frac{8\left(s_0 + \frac{2\nu s_0}{\phi_0^2}\right)}{t-1}}.$$

By conditioning on I_h, we get:

$$
\begin{aligned}
&\mathbb{E}\left[\mathbb{1}\left\{\langle A_{t,\ell},\hat{\theta}_t\rangle \geq \langle A_{t,k},\hat{\theta}_t\rangle\right\}\langle A_{t,k}-A_{t,\ell},\theta\rangle \mid \mathcal{A}_t\in\mathcal{R}_k\cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{2\nu}}\right] \\
&\leq \sum_{h=0}^{\lceil s_2/\delta\rceil}\mathbb{E}\left[\mathbb{1}\left\{\langle A_{t,\ell},\hat{\theta}_t\rangle \geq \langle A_{t,k},\hat{\theta}_t\rangle\right\}\langle A_{t,k}-A_{t,\ell},\theta\rangle \mid \mathcal{A}_t\in\mathcal{R}_k\cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{2\nu}}\right]\mathbb{P}(\mathcal{A}_t\in I_h) \\
&\overset{(a)}{\leq} \sum_{h=0}^{\lceil s_2/\delta\rceil} 2\delta s_A(h+1)\,\mathbb{E}\left[\mathbb{1}\left\{\langle A_{t,\ell},\hat{\theta}_t\rangle \geq \langle A_{t,k},\hat{\theta}_t\rangle\right\} \mid \mathcal{A}_t\in\mathcal{R}_k\cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{2\nu}}\right] \\
&\qquad\qquad\qquad \times \mathbb{P}\left(\langle A_{t,k}-A_{t,\ell},\theta\rangle\in(0,2\delta s_A(h+1)]\right) \\
&\leq \sum_{h=0}^{\lceil s_2/\delta\rceil} 2\delta s_A(h+1)\mathbb{P}\left(\langle A_{t,\ell},\hat{\theta}_t\rangle \geq \langle A_{t,k},\hat{\theta}_t\rangle \mid \mathcal{A}_t\in\mathcal{R}_k\cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{2\nu}}\right),
\end{aligned}
$$

where for (a), we used the definition of I_h. Under the event $\mathcal{A}_t\in I_h$, the event $\langle A_{t,\ell},\hat{\theta}_t\rangle \geq \langle A_{t,k},\hat{\theta}_t\rangle$ happens only when at least one of the events $\langle A_{t,k},\theta-\hat{\theta}\rangle\geq\delta s_A h$ or $\langle A_{t,\ell},\hat{\theta}-\theta\rangle\geq\delta s_A h$ holds. Therefore,

$$
\begin{aligned}
&\mathbb{P}\left(\langle A_{t,\ell},\hat{\theta}_t\rangle \geq \langle A_{t,k},\hat{\theta}_t\rangle \mid \mathcal{A}_t\in\mathcal{R}_k\cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{2\nu}}\right) \\
&\leq \mathbb{P}\left(\langle A_{t,k},\theta-\hat{\theta}\rangle\geq\delta s_A h \mid \mathcal{A}_t\in\mathcal{R}_k\cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{2\nu}}\right) \\
&\qquad\qquad + \mathbb{P}\left(\langle A_{t,\ell},\hat{\theta}-\theta\rangle\geq\delta s_A h \mid \mathcal{A}_t\in\mathcal{R}_k\cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{2\nu}}\right) \\
&\overset{(a)}{\leq} \mathbb{P}\left(\|\theta-\hat{\theta}\|_2\geq\delta h \mid \mathcal{A}_t\in\mathcal{R}_k\cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{2\nu}}\right) + \mathbb{P}\left(\|\theta-\hat{\theta}\|_2\geq\delta h \mid \mathcal{A}_t\in\mathcal{R}_k\cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{2\nu}}\right) \\
&= 2\mathbb{P}\left(\|\theta-\hat{\theta}\|_2\geq\delta h \mid \mathcal{A}_t\in\mathcal{R}_k\cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{2\nu}}\right),
\end{aligned}
$$

where for (a), we used the Cauchy–Schwarz inequality. Let us denote $s' = s_0 + \frac{2\nu\sqrt{s_0}}{\phi_0^2}$. Then, using Lemma 3, we get:

$$
\begin{aligned}
\mathbb{P}\left(\|\theta-\hat{\theta}\|_2\geq\delta h \mid \mathcal{A}_t\in\mathcal{R}_k\cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{2\nu}}\right) &\leq 2s'\exp\left(-\frac{\alpha^2 t\delta^2 h^2}{8\sigma^2 s_A^2\nu^2 C_{\mathrm{b}}^2 s'}\right) \\
&= 2s'\exp\left(-h^2\right).
\end{aligned}
$$

We also trivially have that:

$$
\mathbb{P}\left(\|\theta-\hat{\theta}\|_2\geq\delta h \mid \mathcal{A}_t\in\mathcal{R}_k\cap I_h,\ \mathcal{E}_t,\ \mathcal{G}_t^{\frac{\alpha}{2\nu}}\right) \leq 1.
$$

Therefore, we can upper bound the expected instantaneous regret as:

$$
\begin{aligned}
r_t^\pi &\leq \sum_{k=1}^{2}\mathbb{E}\left[r_t^\pi \mid \mathcal{A}_t\in\mathcal{R}_k\right]\mathbb{P}\left(\mathcal{A}_t\in\mathcal{R}_k\right) \\
&\leq \sum_{k=1}^{2}\left(\sum_{\ell\neq k}\left(\sum_{h=0}^{\lceil s_2/\delta\rceil}\left(2\delta s_A(h+1)\min\left\{1,4s'\exp\left(-h^2\right)\right\}\right)\right)\right. \\
&\quad \left. +2s_A s_2\left(\mathbb{P}(\mathcal{E}_t^c)+\mathbb{P}\left(\left(\mathcal{G}_t^{\frac{\alpha}{2\nu}}\right)^c\right)\right)\cdot\mathbb{P}\left(\mathcal{A}_t\in\mathcal{R}_k\right)\right) \\
&\overset{(a)}{\leq} 2\delta s_A\sum_{h=0}^{\lceil s_2/\delta\rceil}(h+1)\min\left\{1,4s'\exp\left(-h^2\right)\right\} + 2s_A s_2\left(\mathbb{P}(\mathcal{E}_t^c)+\mathbb{P}\left(\left(\mathcal{G}_t^{\frac{\alpha}{2\nu}}\right)^c\middle|\mathcal{E}_t\right)\right) \\
&\leq 2\delta s_A\left(\sum_{h=0}^{h_0}(h+1)+\sum_{h=h_0+1}^{h_{\max}}4s'(h+1)\exp\left(-h^2\right)\right) + 2s_A s_2\left(\mathbb{P}(\mathcal{E}_t^c)+\mathbb{P}\left(\left(\mathcal{G}_t^{\frac{\alpha}{2\nu}}\right)^c\middle|\mathcal{E}_t\right)\right)
\end{aligned}
$$

where for (a), we used $\sum_{k=1}^{K}\mathbb{P}\left(\mathcal{A}_t\in\mathcal{R}_k\right)=1$ and we set $h_0 \coloneqq \lfloor\sqrt{\log 4s'}+1\rfloor$. We have:

$$
\begin{aligned}
\sum_{h=h_0+1}^{h_{\max}}(h+1)\exp(-h^2) &= \sum_{h=h_0+1}^{h_{\max}}h\exp(-h^2)+\sum_{h=h_0+1}^{h_{\max}}\exp(-h^2) \\
&\leq \int_{h_0}^{\infty}h\exp(-h^2)dh+\int_{h_0}^{\infty}\exp(-h^2)dh.
\end{aligned}
$$

Since $h_0 \geq 1$ from $s_0 \geq 1$, we get:

$$\int_{h_0}^{\infty} h \exp(-h^2) dh = \frac{1}{2} \exp(-h_0^2)$$
$$\int_{h_0}^{\infty} \exp(-h^2) dh \leq \exp(-h_0^2).$$

Therefore,

$$\sum_{h=h_0+1}^{h_{\max}} (h+1) \exp(-h^2) \leq \frac{3}{2} \exp(-h_0^2).$$

We get:

$$\begin{aligned}
\sum_{h=0}^{h_0}(h+1) + \sum_{h=h_0+1}^{h_{\max}} 4s'(h+1) \exp\left(-h^2\right) &\leq \frac{(h_0+1)(h_0+2)}{2} + 4s' \frac{3}{2} \exp(-h_0^2) \\
&\leq \frac{h_0^2 + 3h_0 + 2}{2} + \frac{6s'}{4s'} \\
&\overset{(a)}{\leq} \frac{9}{2} h_0^2,
\end{aligned}$$

where for (a), we used $h_0 \geq 1$. Finally, we get:

$$\begin{aligned}
r_t^{\pi} &\leq 9\delta s_A h_0^2 + 2 s_A s_2 \left(\mathbb{P}(\mathcal{E}_t^c) + \mathbb{P}\left(\left(\mathcal{G}_t^{\frac{\alpha}{2\nu}} \right)^c \middle| \mathcal{E}_t \right) \right) \\
&\leq \frac{18 \sigma s_A h_0^2 \nu}{\alpha} \sqrt{\frac{2\left(s_0 + \frac{2\nu\sqrt{s_0}}{\phi_0^2}\right)}{t-1}} + 2 s_A s_2 \left(\mathbb{P}(\mathcal{E}_t^c) + \mathbb{P}\left(\left(\mathcal{G}_t^{\frac{\alpha}{2\nu}} \right)^c \middle| \mathcal{E}_t \right) \right).
\end{aligned}$$

This concludes the proof. □

F Additional Experimental Results

F.1 Additional Results with Various Correlation Levels

Figures 5-13 show the numerical results with correlation levels between two arms $\rho^2 \in \{0.0, 0.3, 0.7\}$ and dimension $d \in \{100, 200, 1000, 2000, 10000\}$, respectively. We find that TH Lasso bandit exhibits lower regret than SA Lasso bandit and DR Lasso bandit in all scenarios. In particular, the difference between TH Lasso and SA Lasso becomes more apparent as the dimension d increases, just as the theorem shows.

Case 1: $\rho^2 = 0.0$

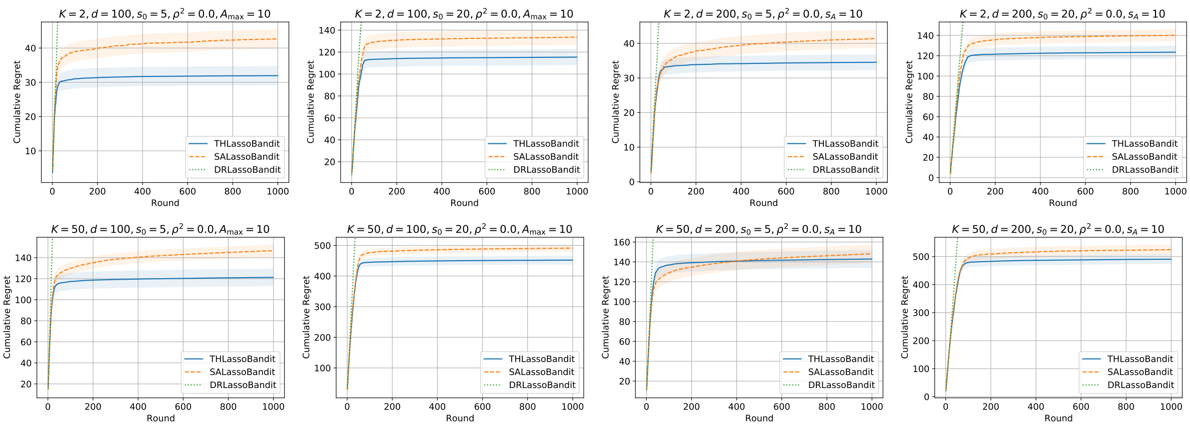

Figure 5: Cumulative regret of the three algorithms with $\rho^2 = 0.0$, $s_A = 10$, $K \in \{2, 50\}$, $d \in \{100, 200\}$, and $s_0 \in \{5, 20\}$. The shaded area represents the standard errors.

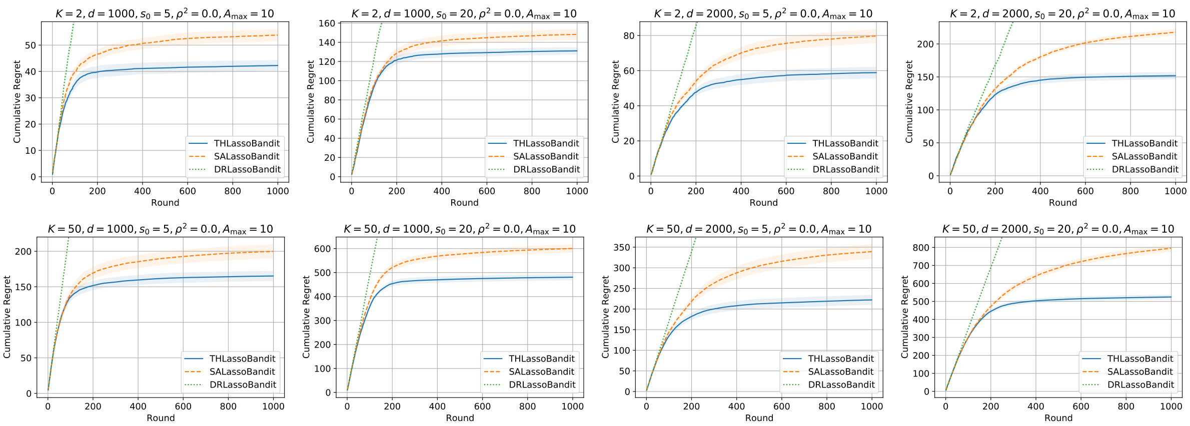

Figure 6: Cumulative regret of the three algorithms with $\rho^2 = 0.0$, $A_{\max} = 10$, $K \in \{2, 50\}$, $d \in \{1000, 2000\}$, and $s_0 \in \{5, 20\}$. The shaded area represents the standard errors.

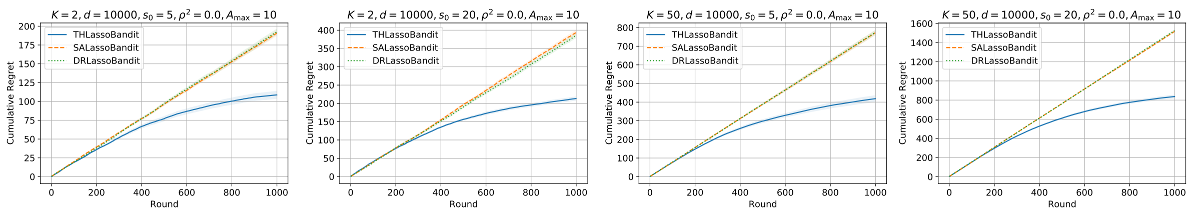

Figure 7: Cumulative regret of the three algorithms with $\rho^2 = 0.0$, $A_{\max} = 10$, $K \in \{2, 50\}$, $d = 10000$, and $s_0 \in \{5, 20\}$. The shaded area represents the standard errors.

Case 2: $\rho^2 = 0.3$

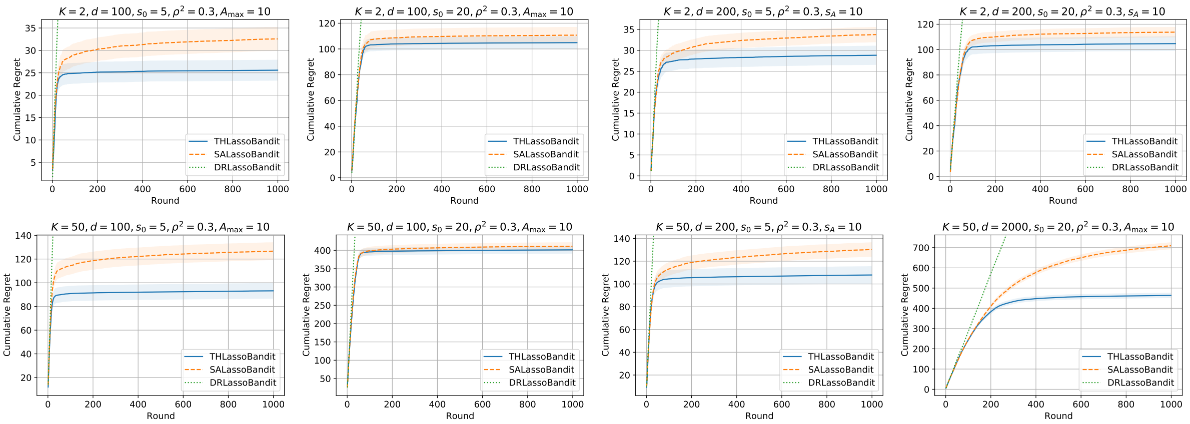

Figure 8: Cumulative regret of the three algorithms with $\rho^2 = 0.3$, $A_{\max} = 10$, $K \in \{2, 50\}$, $d \in \{100, 200\}$, and $s_0 \in \{5, 20\}$. The shaded area represents the standard errors.

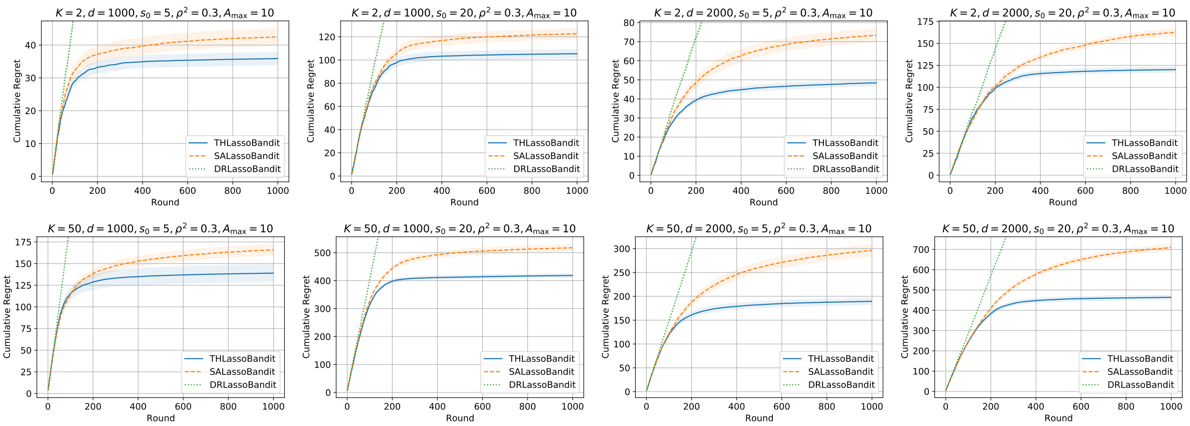

Figure 9: Cumulative regret of the three algorithms with $\rho^2 = 0.3$, $A_{\max} = 10$, $K \in \{2, 50\}$, $d \in \{1000, 2000\}$, and $s_0 \in \{5, 20\}$. The shaded area represents the standard errors.

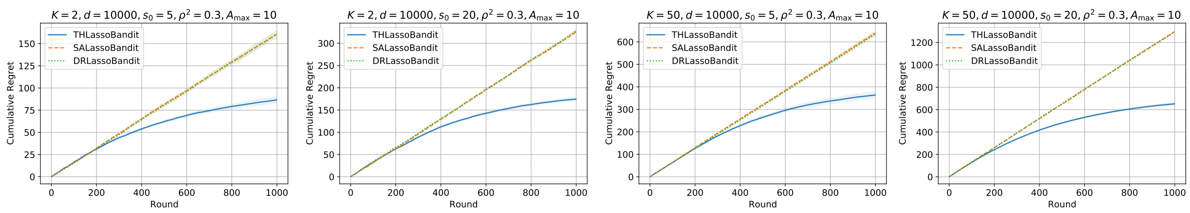

Figure 10: Cumulative regret of the three algorithms with $\rho^2 = 0.3$, $A_{\max} = 10$, $K \in \{2, 50\}$, $d = 10000$, and $s_0 \in \{5, 20\}$. The shaded area represents the standard errors.

Case 3: $\rho^2 = 0.7$

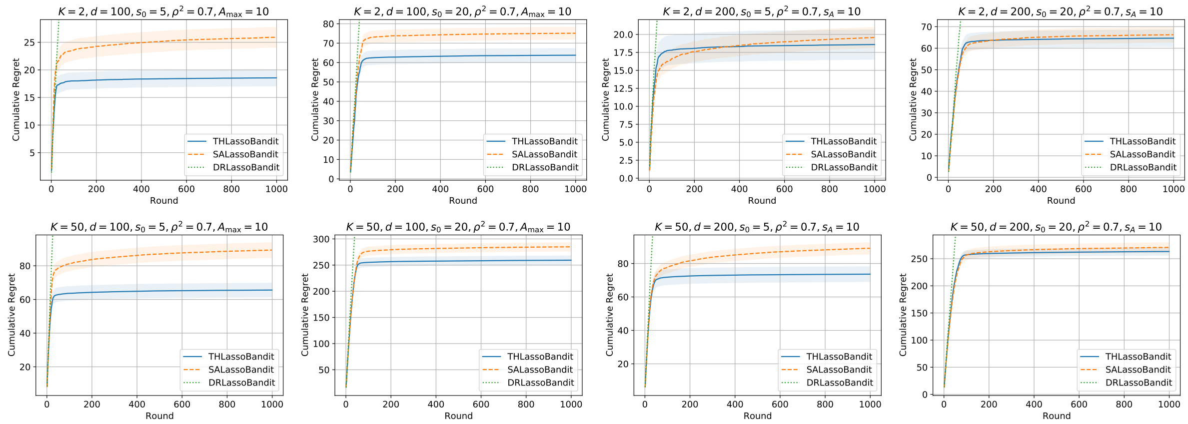

Figure 11: Cumulative regret of the three algorithms with $\rho^2 = 0.7$, $A_{\max} = 10$, $K \in \{2, 50\}$, $d \in \{100, 200\}$, and $s_0 \in \{5, 20\}$. The shaded area represents the standard errors.

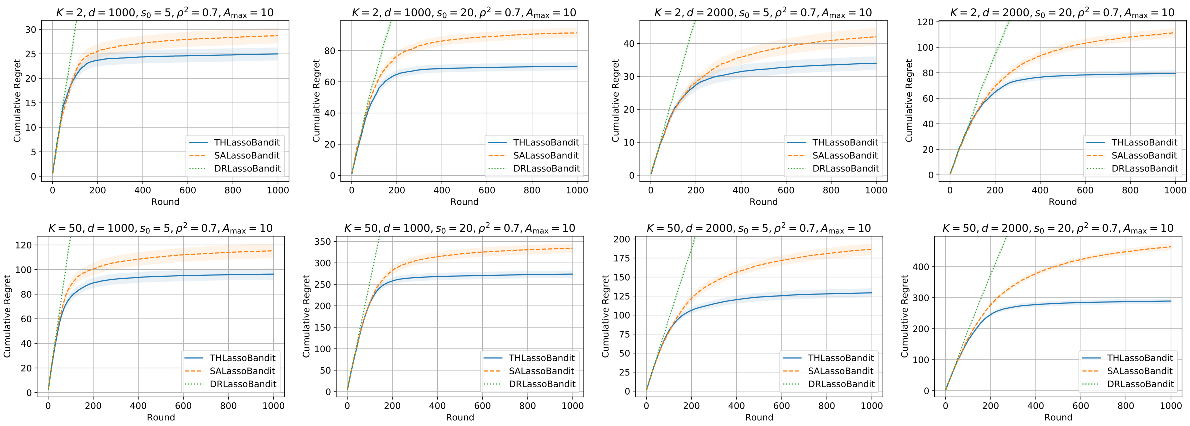

Figure 12: Cumulative regret of the three algorithms with $\rho^2 = 0.7$, $A_{\max} = 10$, $K \in \{2, 50\}$, $d \in \{1000, 2000\}$, and $s_0 \in \{5, 20\}$. The shaded area represents the standard errors.

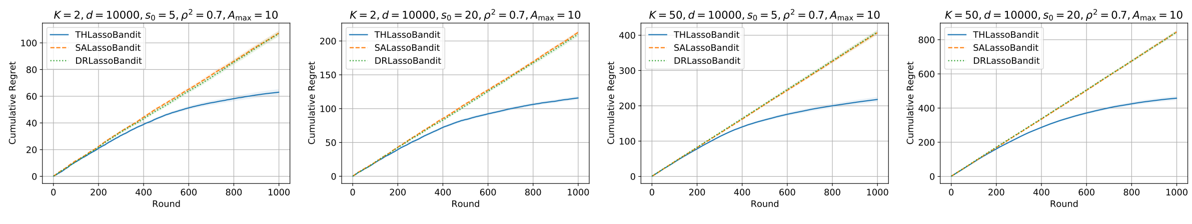

Figure 13: Cumulative regret of the three algorithms with $\rho^2 = 0.7$, $A_{\max} = 10$, $K \in \{2, 50\}$, $d = 10000$, and $s_0 \in \{5, 20\}$. The shaded area represents the standard errors.

F.2 Additional Results with Various $A_{\max}$ for K-Armed Bandits

We also present the experimental results with varying $A_{\max} \in \{2.5, 5, 10, 20, 40, \infty\}$ and a different parameter setting. We set $K = 50, d = 1000$, and $s_0 = 20$. Figure 14 shows the average cumulative regret at $t = 1000$ of TH Lasso bandit and SA Lasso bandit for each $A_{\max}$. We observe that TH Lasso bandit outperforms SA Lasso bandit for all $A_{\max}$.

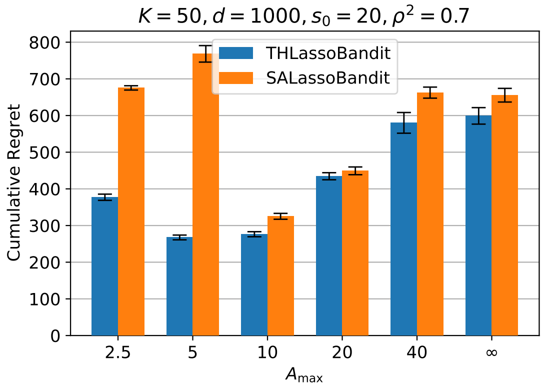

Figure 14: Cumulative regret at round $t = 1000$ of TH Lasso bandit and SA Lasso bandit with $\rho^2 = 0.7$, $K = 50$, $d = 1000$, $s_0 = 20$, and varying $A_{\max} \in \{2.5, 5, 10, 20, 40, \infty\}$. The error bars represent the standard errors.

F.3 Additional Results with Non-Gaussian Distributions

Figure 15 shows the numerical results with the uniform distribution and non-Gaussian elliptical distributions. For experiments with the uniform distribution, in each round t, we sample each feature vector $A_{t,k}$ independently from a d-dimensional hypercube $[-1, 1]^d$. For experiments with the elliptical distribution, we construct each feature vector $A_{t,k}$ by the following equation:

$$A_{t,k} = \mu + RAU^{(l)}$$

where $\mu \in \mathbb{R}^d$ is a mean vector, $U^{(l)} \in \mathbb{R}^l$ is uniformly distributed on the unit sphere in $\mathbb{R}^{(l)}$, $R \in \mathbb{R}$ is a random variable independent of $U^{(l)}$, and A is a $d \times l$-dimensional matrix with rank l. We sample R from Gaussian distribution $\mathcal{N}(0, 1)$, and sample each element of A in an i.i.d manner using the uniform distribution on $[-1, 1]$. We set $\mu = \mathbf{0}_d$ and set $l = 100$. As in the previous experiments, we generate each non-zero components of θ in an i.i.d manner using the uniform distribution on $[1, 2]$. The noise process is Gaussian, i.i.d. over rounds: $\epsilon_t \sim \mathcal{N}(0, 1)$. We find that TH Lasso bandit exhibits lower regret than SA Lasso bandit and DR Lasso bandit in experiments with both distributions.

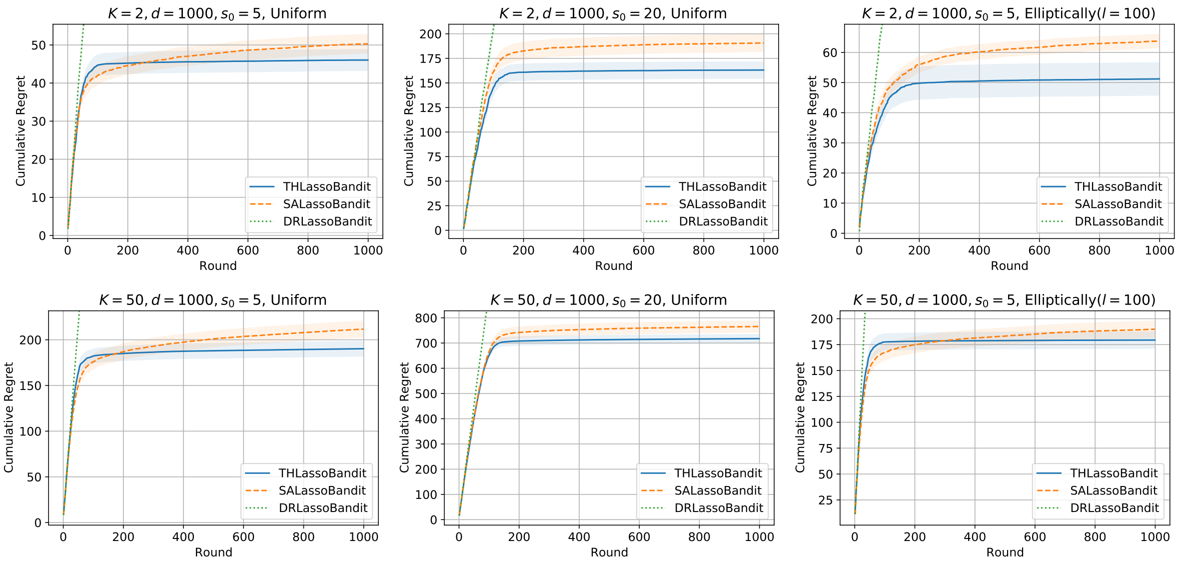

Figure 15: Cumulative regret of the three algorithms with non-Gaussian distributions. The figures in left and center columns show the experimental results with the uniform distribution. The figures in right column shows the experimental results with the elliptical distribution. The shaded area represents the standard errors.

F.4 Additional Results in Hard Instances

We further investigate the performance of TH Lasso bandit in hard instances where the feature vectors of the best arms do not cover the support of θ (situations where the covariate diversity condition [5] does not hold). In this experiment, we set $K = 3$ and $\theta = (1, 0.1, 1, 0, 0, \cdots, 0)^\top$ so that $S = \{1, 2, 3\}$. We generate the arm set by generating feature vectors separately on the support S and the non-support $S^c = [d] \setminus S$. First, in each round t, we generate the feature vectors on S as the following procedure: in each round t, the set of feature vectors $\mathcal{A}_t^S = \{A_{t,1}^S, A_{t,2}^S, A_{t,3}^S\}$ is set to $\mathcal{A}^1$ with probability 0.3 and is set to $\mathcal{A}^2$ with probability 0.7. We set $\mathcal{A}^1 = \{(1,0,0)^\top, (0,1,0)^\top, (0.9, 0.5, 0)^\top\}$ and $\mathcal{A}^1 = \{(0,1,0)^\top, (0,0,1)^\top, (0.0, 0.5, 0.9)^\top\}$. Second, for each component $i \in S^c$, we sample $((A_{t,1}^{S^c})_i, (A_{t,2}^{S^c})_i, (A_{t,3}^{S^c})_i)^\top \in \mathbb{R}^3$ from a Gaussian distribution $\mathcal{N}(\mathbf{0}_3, V)$ where $V_{j,j} = 1$ for all $j \in [3]$ and $V_{j,k} = \rho^2$ for all $j \neq k \in [3]$. We then define $A_{t,k}^{S^c} = ((A_{t,k}^{S^c})_1, \cdots, (A_{t,k}^{S^c})_{|S^c|})^\top \in \mathbb{R}^{|S^c|}$. Finally, by concatenating the feature vectors on S and S^c, we construct the feature vector $A_{t,k} = ((A_{t,k}^S)^\top, (A_{t,k}^{S^c})^\top)^\top \in \mathbb{R}^d$. Note that $(1,0,0)^\top \in \mathcal{A}^1$ and $(0,0,1)^\top \in \mathcal{A}^2$ are always included in the best arms, and they do not span $\mathbb{R}^3$. Figure 16 shows the numerical results with correlation levels between two arms $\rho^2 \in \{0.0, 0.3, 0.7\}$ and dimension $d \in \{1000, 2000\}$. We find that TH Lasso bandit exhibits lower regret than SA Lasso bandit and DR Lasso bandit in all scenarios.

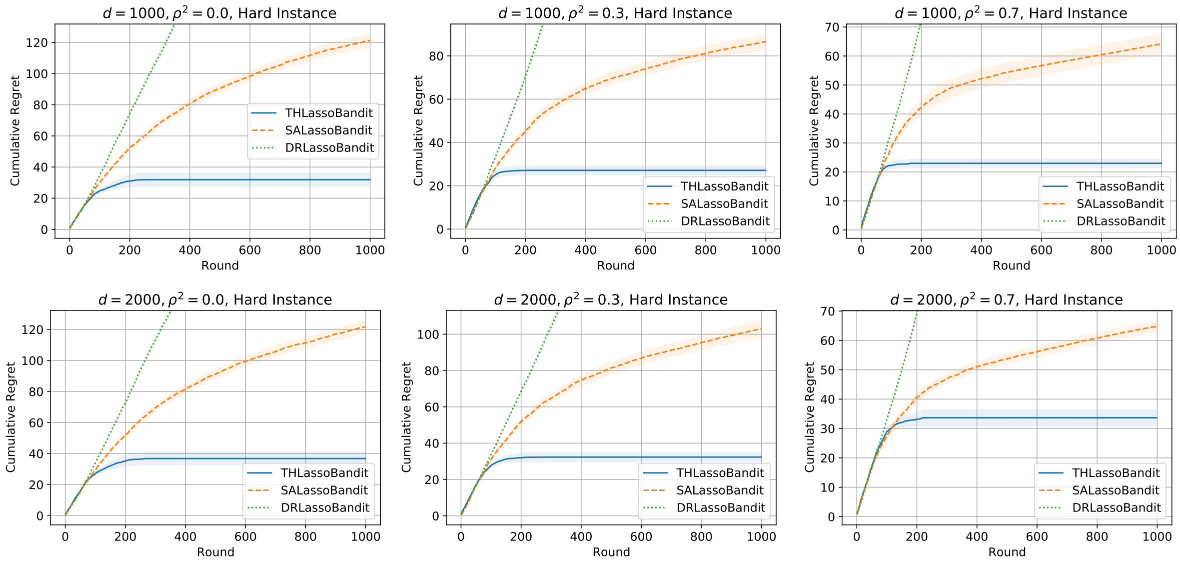

Figure 16: Cumulative regret of the three algorithms in hard instances. The shaded area represents the standard errors.