Sparse Linear Bandits with Blocking Constraints

Adit Jain
CollinearAI

Soumyabrata Pal
Adobe Research

Sunav Choudhary
Adobe Research

Ramasuri Narayanam
Adobe Research

Harshita Chopra
University of Washington

Vikram Krishnamurthy
Cornell University

Abstract

We investigate the high-dimensional sparse linear bandits problem in a data-poor regime where the time horizon is much smaller than the ambient dimension and number of arms. We study the setting under the additional *blocking constraint* where each unique arm can be pulled only once. The blocking constraint is motivated by practical applications in personalized content recommendation and identification of datapoints to improve annotation efficiency for complex learning tasks. With mild assumptions on the arms, our proposed online algorithm (`BSLB`) achieves a regret guarantee of $\widetilde{\mathsf{O}}((1+\beta_k)^2 k^{\frac{2}{3}}\mathsf{T}^{\frac{2}{3}})$ where the parameter vector has an (unknown) relative tail β_k - the ratio of ℓ_1 norm of the top-k and remaining entries of the parameter vector. To this end, we show novel offline statistical guarantees of the lasso estimator for the linear model that is robust to the sparsity modeling assumption. Finally, we propose a meta-algorithm (`C-BSLB`) based on corralling that does not need knowledge of optimal sparsity parameter k at minimal cost to regret. Our experiments on multiple real-world datasets demonstrate the validity of our algorithms and theoretical framework.

1 INTRODUCTION

Sparse linear bandits are a rich class of models for sequential decision-making in settings where only a few features contribute to the outcome. They have been applied to various domains such as personalized medicine and online advertising (Bastani and Bayati, 2020; Abbasi-Yadkori et al., 2012). Recently, the theoretical properties of sparse linear bandits have been explored in data-poor regimes (Wang et al., 2023). Formally, consider a time horizon of T rounds, M arms, each with a d-dimensional feature vector, and an unknown parameter vector with sparsity k, where $\mathsf{T} \ll d \ll \mathsf{M}$. In each round, the learner selects an arm and observes a noisy reward whose expectation is the inner product between the parameter vector and the arm's feature vector. The objective is to design an algorithm that sequentially selects arms to maximize cumulative reward. Notably, (Hao et al., 2020) established tight, dimension-free regret guarantees of $\mathsf{O}(\mathsf{T}^{2/3})$ in this setting. In this work, we study a more practical variant of sparse linear bandits that better models real-world, data-poor scenarios.

Concretely, there are three key practicalities: The first is the consideration of a novel constraint to the sparse bandit framework: each arm may be pulled only once, a restriction we refer to as the *blocking constraint*. Second, we address the challenge of *model misspecification* by providing regret guarantees that depend on how close the true parameter vector is to being k-sparse, thus broadening the applicability of our results beyond exactly sparse settings. Finally, unlike prior approaches, we develop an efficient algorithm that does not require prior knowledge of the sparsity parameter k.

Motivation. The blocking constraint is *inherent* to the data acquisition process in data-poor regimes. For instance, consider personalized applications such as movie/book recommendations on edge devices. Users typically consume an item and rate it only once. There is hardly any point in recommending an item that has been previously consumed, at least not for some time - this is captured via the *blocking constraint*. Existing theoretical approaches in recommendation litera-

Proceedings of the 29th International Conference on Artificial Intelligence and Statistics (AISTATS) 2026, Tangier, Morocco. PMLR: Volume 300.

ture (Bresler et al., 2014, 2016; Heckel and Ramchandran, 2017; Huleihel et al., 2021; Pal et al., 2024) use collaborative filtering (CF) via multiple users to address the blocking constraint. However, unlike CF which affects privacy, is limited on edge devices and relies on multiple users, our approach applies to a single-user setting while still addressing the blocking constraint.

A second application of our framework, in settings where labeling is expensive, is to create example banks comprising of *high quality hard datapoints* for in-context learning via large language models (LLMs). LLMs offer strong zero-shot capabilities, making it easier to prototype solutions for downstream tasks. However, they struggle with complex domain-specific queries, particularly when relevant training data is poor or evolving (Farr et al., 2024). In-context learning with a few-shot examples has emerged as a powerful approach, where a small set of high-quality examples improves model performance (Dong et al., 2022). Crucially, hard examples provide better domain-specific information (Baek et al., 2024; Liu et al., 2024; Mavromatis et al., 2023), but identifying them is challenging. Heuristic-based selection often leads to noisy, mislabeled, or outlier examples (Mindermann et al., 2022). Alternatively, we can leverage domain experts to assign a hardness score while annotating. This data-poor problem can be framed in a bandit framework, where unlabeled datapoints act as arms and are sequentially annotated while hardness scores are modeled as a sparse linear function of embeddings. Note that *blocking constraint* is *inevitable* here due to limited (even single) annotators. In Appendix A.1 to A.4, we present numerical results for these applications and motivate an application for material design experimentation which is also a data-scarce task (Fung et al., 2021) .

Overview of our Techniques and Contributions. We propose `BSLB` (Blocked Sparse Linear Bandits), an efficient algorithm in our framework which is a type of Explore-Then-Commit algorithm. In the exploration period, we sample a set of unique arms from a carefully chosen distribution and observe their rewards - the goal is to ensure that the expected covariance matrix of the sampled arms has a large minimum eigenvalue. Such a well-conditioned covariance matrix ensures that the confidence ball of the estimate shrinks in all directions. In the exploitation period, we use the Lasso estimator to estimate the unknown model parameters. The optimal exploration period depends on the correct sparsity level of the unknown parameter vector, which is difficult to set in practice. Therefore, we also present a meta-bandit algorithm based on corralling a set of base bandit algorithms (Agarwal et al., 2017) - obtaining the same order-wise regret guarantees but without needing the knowledge of sparsity hyperparameter. Below, we summarize the *main contributions* of this paper:

1. We propose the high-dimensional sparse linear bandits framework with the novel blocking constraint to model sequential decision-making tasks in data-poor regimes. Importantly, the time horizon is much smaller than the ambient dimension d. We account for model misspecification where the unknown parameter vector is *close to being sparse* with relative tail magnitude β_k at level k (ℓ_1 norm of bottom $d-k$ entries divided by ℓ_1 norm of top k ones).
2. We propose a computationally efficient "explore then commit" (ETC) algorithm `BSLB` for regret minimization in our framework (Theorem 3) that achieves a regret guarantee of $\widetilde{\mathsf{O}}((1+\beta_k)^2 k^{\frac{2}{3}}\mathsf{T}^{\frac{2}{3}})$ for fixed known k but unknown β_k - that is, the relative tail magnitude remains unknown to the algorithm. For the special case of exact sparsity that is $\beta_k = 0$, `BSLB` achieves a regret guarantee of $O(k^{2/3}\mathsf{T}^{2/3})$ (Corollary 2) which is tight in our setting (Theorem 4).
3. `BSLB` requires knowledge of the sparsity level k (which also controls the tail magnitude β_k) to set the length of the exploration period. This is challenging to establish in practice, especially for vectors that are not exactly sparse. Hence, we propose a meta-algorithm `C-BSLB` that combines base algorithms with different exploration periods. `C-BSLB` achieves same regret guarantees order-wise (Theorem 5) as `BSTB` but without knowing the optimal sparsity level.

To validate our theoretical results, we conduct experiments on both synthetic and real-world datasets. For personalized recommendations, we use MovieLens, Netflix, Jester, and Goodbooks. For intelligent annotation in data-poor settings, we evaluate our method on VOC 2012 image classification dataset and SST-2 text classification dataset. We also demonstrate our framework on a materials discovery task. Our algorithm, `BSLB`, outperforms baselines across all tasks. We summarize key results in Section 4 with full details in Appendix A.

Algorithmic Novelty: We consider a *naive modification of the algorithm proposed in* (Hao et al., 2020) and illustrate why blocking constraint makes the bandit instance non-trivially hard. Their `ESTC` algorithm in the sparse linear bandits framework provides tight regret guarantees of $O(k^{2/3}\mathsf{T}^{2/3})$ when the parameter vector is exactly k-sparse. Intuitively, one might consider modifying `ESTC` to incorporate the blocking constraint. In `ESTC`, a crucial initial step is to compute a distribution over the arms with the objective of maximizing the minimum eigenvalue of the expected covariance matrix. Subsequently, during the exploration period, arms are sampled from the computed distribution. We show that the sample covariance matrix is well-conditioned (the minimum eigenvalue of the sample covariance matrix is close to that of the expected one).

A naive modification to incorporate the blocking constraint involves using rejection sampling. If a previously pulled arm is sampled again, it is ignored and re-sampled from the same distribution. However, this significantly modifies the distribution over arms. For

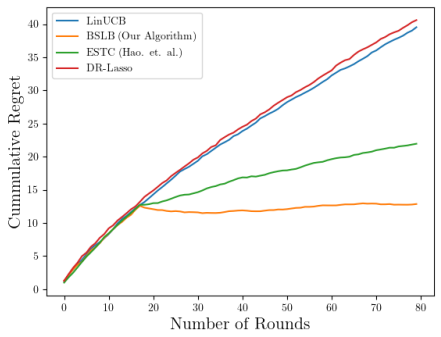

Figure 1: Simulation illustrating performance gap between our proposed algorithm `BSLB` and naive extensions of `ESTC`, LinUCB and DR-Lasso to incorporate blocking constraint. We consider instance with $\mathsf{M} = 500$ arms ($l = 5$ arms of unit ℓ_∞ norm and rest with 0.5), $d = 100$, and $k = 5$.

instance, if the computed distribution in `ESTC` concentrates probability mass on a few l arms (e.g., if l arms have significantly higher norm than others), the blocking constraint forces us to sample arms with low probability mass once those l arms are exhausted. This breaks the statistical guarantees proved for `ESTC` regarding the lasso estimator. We empirically illustrate and validate this special setting in Figure 1 where our algorithm is also benchmarked with similar naive extensions of DR-Lasso and LinUCB (Kim and Paik, 2019; Abbasi-yadkori et al., 2011) - the latter ones do not handle sparsity in any case. Hence, a naive modification of `ESTC` does not work, and the objective itself needs to incorporate the blocking constraint. However, this leads to a non-concave maximization problem, resulting in technical challenges. Finally, Hao et al. (2020) does not consider robustness to sparsity or unknown sparsity level, both of which are important in practice.

Technical Challenges and Novelty. (A) *Statistical guarantees:* Both in our setting and in (Hao et al., 2020), no assumptions are made on the arm vectors, unlike much of the existing sparsity literature, which requires the sample covariance matrix to satisfy the Restricted Isometry Property (RIP) (Boche et al., 2015). A weaker parameterized condition, the Restricted Eigenvalue (RE) property, has been shown to provide strong statistical guarantees for the Lasso estimator in sparse linear regression (Bickel et al., 2009; Rudelson and Zhou, 2013). Specifically, if the expected covariance matrix of the arm vectors satisfies RE with parameter K, then with high probability the sample covariance matrix of sufficiently large i.i.d. samples will also satisfy RE with parameter $K/2$. However, to the best of our knowledge, existing guarantees for the Lasso estimator hold only for exactly sparse parameter vectors when the sample covariance matrix of sampled arms just satisfies RE and therefore are of independent interest. In contrast, offline guarantees for soft sparse parameter vectors (where there is a non-zero tail) require the stronger RIP condition (Wainwright, 2019; Boche et al., 2015). Theorem 1 of this paper bridges the gap and provides general statistical guarantees which (a) holds under the blocking constraint (sampling without replacement from a distribution on arms) (b) holds for an arm set that satisfies only the RE condition, and (c) for a parameter vector that is close to being sparse with a non-zero tail.

(B) *Sampling in Exploration phase:* In linear bandits with arbitrary arms, uniform random sampling during exploration fails to obtain a well-conditioned covariance matrix when arms are distributed non-uniformly. Moreover, as discussed earlier, a naive adaptation of `ESTC`'s sampling technique does not work. To address this, we optimize the probability distribution over arms to maximize the minimum eigenvalue of the expected covariance matrix while incorporating the blocking constraint. However, this leads to a subset selection problem, where the distribution must assign equal probability mass to a subset of arms while excluding others. The resulting optimization problem is discrete and non-concave. To tackle this, we introduce a concave relaxation of the objective and propose a randomized rounding procedure that yields a feasible solution with strong approximation guarantees on the minimum eigenvalue of the covariance matrix. This contribution is of independent interest especially for experiment design literature (Madan et al., 2019), since this is a form of E-optimal design, which does not arise in (Hao et al., 2020). Even when working with a smaller search space using cardinality constraints E-optimal design is NP-hard as studied in (Allen-Zhu et al., 2021a), therefore the problem considered in the paper is computationally intractable and the approximation is novel.

(C) *Knowledge of hyper-parameters:* Our first proposed algorithm `BSLB` (similar to `ESTC`) requires as input a fixed sparsity k. However, the algorithm does not know the tail magnitude β_k at that sparsity level k. Importantly, if k is set too low or too high, then the regret guarantee will not be reasonable. However, it is challenging to set the sparsity parameter without knowing the parameter vector itself. To resolve this challenge, we use a corralling algorithm based on the techniques of (Agarwal et al., 2017) that combines several base algorithms and provides guarantees with respect to the optimal one. However, a naive application leads to an additional corralling cost with a linear dependence

on dimension d, making the regret vacuous ($\mathsf{T} \ll d$). We use a covering argument - we only choose a small representative subset (covering set) of base algorithms for input to corralling and show that for all remaining base algorithms, their regret is close enough to a base algorithm in the covering set.

Summary of differences with (Hao et al., 2020): In (Hao et al., 2020), (i) the exploration distribution is a solution to a concave program over the probability simplex; in our setting, the blocking constraint turns this into a discrete, non-concave subset selection problem requiring a novel concave relaxation and randomized rounding (Theorem 2). (ii) Their Lasso guarantees assume exact sparsity and i.i.d. sampling; ours extend to soft sparsity under the weaker RE condition with sampling without replacement, using McDiarmid's inequality (Theorem 1, Corollary 1). (iii) They require knowledge of k, whereas we introduce a meta-algorithm with a covering argument that avoids this (Theorem 5).

Note that *temporary* blocking has been studied before, where the arm is blocked for a few rounds and then becomes active again (Basu et al., 2019). However, our setup significantly differs in two respects - once an arm has been pulled, it remains permanently blocked, and second, we look at exact cumulative regret and not approximate regret as in (Bishop et al., 2020; Basu et al., 2021). Our permanent single-pull constraint is also related to *mortal bandits* (Chakrabarti et al., 2008), where arms expire over time. Key differences are: (a) in mortal bandits, the cardinality of the available arm set is kept fixed by introducing new arms, forcing continuous exploration akin to restless bandits, whereas our arm set only shrinks; (b) mortal bandit analyses are typically asymptotic, while ours provides finite-sample guarantees. We use the term "blocking constraint" following (Pal et al., 2024).

We denote vectors by bold small letters (say $\mathbf{x}$), scalars by plain letters (say x or X), sets by curly capital letters (say $\mathcal{X}$) and matrices by bold capital letters (say $\mathbf{X}$). We use $[m]$ to denote the set $\{1, 2, \ldots, m\}$, $\|\mathbf{x}\|_p$ to denote the p-norm of vector $\mathbf{x}$. For a set $\mathcal{T}$ of indices $\mathbf{v}_{\mathcal{T}}$ is used to denote the sub-vector of $\mathbf{v}$ restricted to the indices in $\mathcal{T}$ and 0 elsewhere. $\lambda_{\min}(\mathbf{A})$ denotes the minimum eigenvalue of the matrix $\mathbf{A}$ and $\mathsf{diag}(\mathbf{x})$ denotes a diagonal matrix with entries such as $\mathbf{x}$. We use $\mathcal{B}_\infty^d$ and $\mathcal{B}^d$ to denote the unit ball in d-dimensions with respect to the ℓ_∞-norm and ℓ_2 norm respectively. We write $\mathbb{E}X$ to denote the expectation of a random variable X. $\widetilde{O}(\cdot)$ notation hides logarithmic factors.

2 PROBLEM FORMULATION

Consider a set (more generally a multi-set) of M arms $\mathcal{A} \equiv \{\mathbf{a}^{(1)}, \mathbf{a}^{(2)}, \ldots, \mathbf{a}^{(\mathsf{M})}\} \subseteq \mathbb{R}^d$. Let $\mathbf{a}^{(j)} \in \mathcal{B}_\infty^d$ denote the d-dimensional vector embedding associated with the j^{th} arm. We make the same mild boundedness assumption as (Hao et al., 2020), that is, for all $\mathbf{a} \in \mathcal{A}$, $\|\mathbf{a}\|_\infty \leq 1$. We have a time horizon of T rounds. In the high-dimension regime, we have $\mathsf{T} \ll d \ll \mathsf{M}$; i.e., the horizon is much smaller than the ambient dimension, which, in turn, is much smaller than the number of arms. This is a suitable model for the data acquisition process in data-poor regimes. At each round $t \in [\mathsf{T}]$, an arm $\mathbf{a}_t$ which has not been pulled in the first $t-1$ rounds is selected by the online algorithm (decision-maker). Note that such a selection mechanism respects the blocking constraint. Subsequently, the arm a_t is pulled and the algorithm observes the stochastic reward r_t. We model the expected reward $\mathbb{E}r_t$ as a linear function of the arm embedding where $\boldsymbol{\theta} \in \mathbb{R}^d$ is an unknown parameter vector. In particular, the random variable r_t is generated according to $r_t = \langle \boldsymbol{\theta}, \mathbf{a}_t \rangle + \eta_t$ where $\{\eta_t\}_{t \in [T]}$ are zero-mean sub-gaussian random variables independent across rounds with variance proxy bounded from above by σ^2. For any sparsity level $k \leq d$ we define the tail of the (unknown) parameter vector $\boldsymbol{\theta}$ as, $\beta_k := \frac{\|\boldsymbol{\theta}_{\mathcal{T}_k^c}\|_1}{\|\boldsymbol{\theta}_{\mathcal{T}_k}\|_1}$ where $\mathcal{T}_k$ denotes the set of k largest coordinates of $\boldsymbol{\theta}$ by absolute value and $\mathcal{T}_k^c = [d] \setminus \mathcal{T}_k$. Note that β_k is unknown to the algorithm.

Now, we formally define the regret objective in our online learning set-up, which also respects the *blocking constraint*. Our regret definition captures the difference in the cumulative expected reward of arms selected by the online algorithm versus the cumulative expected reward of the T unique arms with the highest mean rewards. Consider a permutation $\pi : [|\mathcal{A}|] \to [|\mathcal{A}|]$ of arms such that for any $i < j$, we have $\langle \boldsymbol{\theta}, \mathbf{a}^{(\pi(i))} \rangle \geq \langle \boldsymbol{\theta}, \mathbf{a}^{\pi(j)} \rangle$. We can define the regret $\mathsf{Reg}(\mathsf{T})$ as,

$$\mathsf{Reg}(\mathsf{T}) := \sum_{t=1}^{\mathsf{T}} \langle \boldsymbol{\theta}, \mathbf{a}^{(\pi(t))} \rangle - \sum_{t=1}^{\mathsf{T}} \langle \boldsymbol{\theta}, \mathbf{a}_t \rangle. \tag{1}$$

We aim to design an algorithm that minimizes expected regret $\mathbb{E}[\mathsf{Reg}(\mathsf{T})]$ in our setting where the expectation is over the randomness in the algorithm.

3 OUR ALGORITHM AND MAIN RESULTS

Description of BSLB Algorithm. Our main contribution is to propose an Explore-Then-Commit (ETC) algorithm named `BSLB` which is summarized in Algorithm 1. `BSLB` takes as input a set of arms $\mathcal{A}$, the time horizon T, the exploration budget $\mathsf{T}_{\text{explore}}$, subset selection parameter $\hat{u}$ and the regularization parameter λ. Steps 1-6 of `BSLB` correspond to the exploration component in the algorithm. In Step 1, we first compute a good subset of arms $\mathcal{G} \subset \mathcal{A}$ (using the function

Algorithm 1 Explore then Commit for Blocked Sparse Linear Bandits (BSLB)

Require: Arms $\mathcal{A}$, time horizon T, Exploration Budget $\mathsf{T}_{\text{explore}}$, Regularization Parameter λ, Subset selection parameter $\hat{u}$

1: $\mathcal{G} = \text{GetGoodSubset}(\mathcal{A}, \hat{u})$ // Compute good subset of arms
2: **for** $t \in [\mathsf{T}_{\text{explore}}]$ **do**
3: Sample uniformly from $\mathbf{a}_t \sim \mathcal{G}$ and get reward r_t
4: $\mathcal{G} \leftarrow \mathcal{G} \setminus \{\mathbf{a}_t\}$ // Remove selected arm from remaining arms
5: **end for**
6: $\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{t \in [\mathsf{T}_{\text{explore}}]} (r_t - \langle \boldsymbol{\theta}, \mathbf{a}_t \rangle)^2 + \lambda ||\boldsymbol{\theta}||_1$ // Compute estimate using LASSO
7: $\mathcal{D} = \mathcal{A} \setminus \{\mathbf{a}_t |\ t \in [\mathsf{T}_{\text{explore}}]\}$ // Arms available for exploit phase
8: **for** $t \in [\mathsf{T}_{\text{explore}} + 1, \mathsf{T}]$ **do**
9: $\mathbf{a}_t = \arg\max_{\mathbf{a} \in \mathcal{D}} \langle \widehat{\boldsymbol{\theta}}, \mathbf{a} \rangle$
10: $\mathcal{D} = \mathcal{D} \setminus \{\mathbf{a}_t\}$
11: **end for**

Algorithm 2 GetGoodSubset: Subset selection for maximizing the minimum eigenvalue

Require: Set of Samples $\mathcal{A}$, Subset selection parameter $\hat{u}$

1: **Output:** Subset $\mathcal{G}$
2: Maximize the objective function defined in (5) with input $\hat{u}$ to obtain distribution $\hat{\boldsymbol{\mu}}$ over $\mathcal{A}$.
3: **for** $j \in [\mathsf{M}]$ **do**
4: $\mathcal{G} = \mathcal{G} \cup \{\mathbf{a}^{(j)}\}$ with probability $\hat{u}\hat{\boldsymbol{\mu}}_j$ // Add sample j to $\mathcal{G}$ with prob. $\hat{u} \cdot \hat{\boldsymbol{\mu}}_j$
5: **end for**
6: $\bar{\mathcal{G}} = \texttt{SubsetSearch}(\mathcal{A}, d, \hat{u})$ // Algorithm 3 in Appendix C.7
7: **return** $\arg\max_{\mathcal{H} \in \mathcal{G}, \bar{\mathcal{G}}} \lambda_{\min}(\mathcal{H})$

GetGoodSubset$(\mathcal{A}, \hat{u})$) which comprises representative arms that cover the d-dimensional space reasonably well. Subsequently, in Steps 2-5, we sample arms without replacement from the set of arms $\mathcal{G}$ for $\mathsf{T}_{\text{explore}}$ rounds. The goal in the exploration component is to select a subset of arms such that the image of sparse vectors under the linear transformation by the sample covariance matrix of the selected set has a sufficiently large magnitude (see Definition 1). As we prove, such a result ensures nice statistical guarantees of the parameters estimated using observations from the subset of arms pulled until the end of the exploration phase. Since the set of arms $\mathcal{A}$ can be arbitrary, note that sampling arms uniformly at random from the entire set might not have good coverage - especially when most arms are concentrated in a lower-dimensional subspace. Therefore, finding a good representative subset of arms leads to the following discrete optimization problem

$$\lambda^*_{\min} := \max_{\mathcal{G}' \subseteq \mathcal{A}} \lambda_{\min} \left(|\mathcal{G}'|^{-1} \sum_{\mathbf{a} \in \mathcal{G}'} \mathbf{a}\mathbf{a}^\mathsf{T} \right). \tag{2}$$

Algorithm 2: The function GetGoodSubset$(\mathcal{A}, \hat{u})$ approximates the solution to this computationally intractable discrete optimization. We maximize a relaxed concave program in equation 5 efficiently for a chosen input parameter $\hat{u}$ to obtain a distribution $\hat{\boldsymbol{\mu}}$ on the set of arms $\mathcal{A}$ - subsequently, we construct the subset $\mathcal{G}$ using randomized rounding (Step 4 of Algorithm 2) with $\hat{\boldsymbol{\mu}}$ to obtain a feasible solution to equation 2. In addition, in Step 6, we include a search over all subsets of size $\mathsf{O}(d)$ and identify the one having the highest minimum eigenvalue. Step 6 allows us to achieve strong theoretical results but is computationally expensive - in practice, it can be skipped. As we explain in Remark 3, skipping Step 6 leads to a slightly weaker regret guarantee, but significantly reduces computational complexity making Algorithm 2 highly efficient.

In Step 7 of Algorithm 1, we use the Lasso estimator to get an estimate $\widehat{\boldsymbol{\theta}}$ of the unknown parameter vector $\boldsymbol{\theta} \in \mathbb{R}^d$. Note that the number of samples used in obtaining the estimate $\widehat{\boldsymbol{\theta}}$ is much smaller than the dimension d. The second part of BSLB (Steps 8-11) corresponds to the exploitation component of the algorithm, we pull arms that are predicted to be the most rewarding according to our recovered estimate $\widehat{\boldsymbol{\theta}}$. At every round in BSLB, no arm is pulled more than once, thus respecting the *blocking constraint*. It is important to note that BSLB is two-shot. We change our data acquisition strategy only once after the exploration component - thus making our algorithm easy to implement in practice.

Next, we move on to our main theoretical results. In Section 3.1, we provide offline guarantees of the Lasso estimator for sparse linear regression that are robust to sparsity and hold only with weak conditions. In Section 3.2, we analyze an efficient algorithm for computing the sampling distribution in the exploration phase of BSLB. In Section 3.3, using the offline and approximation guarantees, we provide regret guarantees for our proposed algorithm BSLB. In Section 3.4, we discuss guarantees on C-BSLB (Algorithm 4) that does not need the knowledge of the sparsity level.

3.1 Offline Lasso Estimator Guarantees With Soft Sparsity and RE condition

To the best of our knowledge, there does not exist in the literature offline guarantees for sparse linear regression that is (A) robust to sparsity modeling assumption and (B) holds only under the mild RE condition on the sample covariance matrix. Our first theoretical result fills this gap to a certain extent with an upper

bound on error rate. We will start by introducing the definition of Restricted Eigenvalue (RE)

Definition 1. ***Restricted Eigenvalue (RE):*** *$\mathbf{X} \in \mathbb{R}^{n\times d}$ satisfies Restricted Eigenvalue property $\mathsf{RE}(k_0, \gamma, \mathbf{X})$, if there exists a constant $K(k_0, \gamma, \mathbf{X})$ such that for all $z \in \mathbb{R}^d$ and $z \neq \mathbf{0}$,*

$$0 < K(k_0, \gamma, \mathbf{X}) = \min_{J\subseteq\{1,\ldots,d\}|J|\leq k_0} \min_{\|z_{J^c}\|_1 \leq \gamma \|z_J\|_1} \frac{\|\mathbf{X}z\|_2}{\|z_J\|_2}.$$

Here $\|z_J\|_1 = \sum_{i\in J} |z_i|$ denotes the ℓ_1 norm of z restricted to the index set J, and $J^c = \{1, \ldots, d\} \setminus J$. Intuitively, RE requires that the design matrix $\mathbf{X}$ preserves the energy of all vectors whose mass is concentrated on at most k_0 coordinates (i.e., vectors in a restricted cone). (Bickel et al., 2009) showed that RE is among the weakest conditions imposed in the literature on the sample covariance matrix to ensure nice statistical guarantees on the Lasso estimator for sparse linear regression. We now state our first main result:

Theorem 1. *Let $\mathbf{X} \in \mathbb{R}^{n\times d}$ be the data matrix satisfying $|\mathbf{X}_{ij}| \leq 1 \forall i, j$. Let $\mathbf{r} \in \mathbb{R}^n$ be the corresponding observations such that $\mathbf{r} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\eta}$, where $\boldsymbol{\eta} \in \mathbb{R}^n$ is a zero-mean sub-gaussian random vector with i.i.d. components having bounded variance $\sigma^2 = O(1)$. Let $\boldsymbol{\theta}$ have an (unknown) relative tail β_k at fixed sparsity level k. Suppose $\mathbf{X}$ satisfies restricted eigenvalue property (Def. 1) $\mathsf{RE}(k, 4(1+\beta_k), \frac{\mathbf{X}}{\sqrt{n}})$ with constant $K > 0$. An estimate $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ recovered using Lasso (Line 9 in `BSLB`) with a regularization parameter $\lambda = \sqrt{\frac{\log d}{n}}$, satisfies following with probability $1 - o(d^{-2})$,*

$$\|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\|_1 = \widetilde{\mathsf{O}}\left(k(1+\beta_k)^2 K^{-2} n^{-1/2}\right). \tag{3}$$

Insight 1. Note that in equation 3, for a fixed sparsity k, the estimator error guarantee decays with the number of samples n (as $n^{-1/2}$) and RE constant K, and grows linearly with k. Existing error guarantees in literature focus only on exact sparse vectors $\boldsymbol{\theta}$ - the data matrix $\mathbf{X}$ satisfies $\mathsf{RE}(k, 3, \frac{\mathbf{X}}{\sqrt{n}})$ with constant K' and $\beta_k = 0$ (see Theorem 7.13 (Wainwright, 2019)). However, with moderately stronger assumption of $\mathsf{RE}(k, 6, \frac{\mathbf{X}}{\sqrt{n}})$ on the data matrix, guarantees of Theorem 1 hold for all $\beta_k \leq 1/2$. As stated in Theorem 1, for a larger tail with $\beta_k > 1/2$, $\mathbf{X}$, RE needs to hold on a larger cone.

Remark 1. Note that the statistical guarantee presented in Theorem 1 is an offline error rate that is *robust to sparsity modeling assumption* - similar to Theorem 7.19 in (Wainwright, 2019) and Theorem 1.6 in (Boche et al., 2015). However, the former holds only for the special case when $\mathbf{X}$ has i.i.d. Gaussian rows, and the latter requires the stronger RIP condition on the data matrix. Our error guarantee is much more general and holds for deterministic data matrices $\mathbf{X}$ satisfying RE.

Proof Outline: The lasso inequality states, $\frac{1}{n}\|\mathbf{X}\mathbf{h}\|_2^2 \leq \frac{\eta^\intercal \mathbf{X}\mathbf{h}}{n} + \lambda\|\mathbf{h}\|_1$. Using the RE condition, we show that the norm squared of the approximation error ($\mathbf{h} = \boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}$) is, $K^2\|\mathbf{h}_{\mathcal{T}_k}\|_2^2 = \mathsf{O}(\|\mathbf{h}\|_1 \|\frac{\mathbf{X}^\intercal\eta}{n}\|_\infty)$. Further by decomposing the approximation we show that $K^2\|\mathbf{h}\|_1 = \mathsf{O}(\|\frac{\mathbf{X}^\intercal\eta}{n}\|_\infty k(1+\beta_k)^2)$. Finally, we apply a concentration inequality for the noise vector η, $\|\frac{\mathbf{X}^\intercal\eta}{n}\|_\infty$ scales as $\mathsf{O}(\frac{1}{\sqrt{n}})$. Below we derive a corollary for the case when the rows of the design matrix are sampled without replacement from a set whose empirical covariance matrix has a minimum eigenvalue. The key step in extending the RE condition to the blocking constraint (sampling without replacement) is the use of concentration inequalities for sampling without replacement (Bardenet and Maillard, 2015): specifically, we apply symmetrization and McDiarmid's inequality to show that if the population covariance of the arm set $\mathcal{G}$ has a sufficiently large minimum eigenvalue, then the sample covariance of arms drawn without replacement from $\mathcal{G}$ satisfies RE with high probability. Due to space constraints, the proof is in Appendix C.2.

Corollary 1. *Let $\mathbf{X} \in \mathbb{R}^{n\times d}$ be the data matrix with n samples and dimension d, whose rows are sampled uniformly without replacement from a set $\mathcal{G} \subset \mathbb{R}^d$. Let $\Lambda = \lambda_{\min}(|\mathcal{G}|^{-1}\sum_{\mathbf{a}\in\mathcal{G}} \mathbf{a}\mathbf{a}^\intercal)$. Consider the same setup for observations $\mathbf{r}$ as in Theorem 1. Provided $n = \Omega(k\Lambda^{-4})$, an estimate $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ recovered using Lasso (Line 9 in `BSLB`), will satisfy with probability $1 - \exp(-\Omega(n)) - o(d^{-2})$,*

$$\|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\|_1 = \widetilde{\mathsf{O}}\left(k(1+\beta_k)^2\Lambda^{-1}n^{-1/2}\right). \tag{4}$$

Remark 2. In comparison to the existing statistical guarantees (Theorem A.3 of (Hao et al., 2020) which is a restatement of Theorem 7.13 of (Wainwright, 2019)), the multiplicative factor $(1+\beta_k)^2$ arise due to non-zero tail of the parameter vector and reduce to 1 for the exact sparsity case ($\beta_k = 0$). Note in particular that we do not have the RE assumption in Corollary 1. Instead, it is replaced by a lower bound on n - arms sampled without replacement from the set $\mathcal{G}$ whose sample covariance matrix has a sufficiently large minimum eigenvalue. This is possible because a lower bound on minimum eigenvalue for a positive semi-definite matrix implies a lower bound on RE with arbitrary parameters - concentration guarantees imply that the RE condition remains satisfied when sufficiently large (yet smaller than $|\mathcal{G}|$) number of samples are sampled from $\mathcal{G}$.

3.2 Subset Selection for Maximizing the Minimum Eigenvalue

Recall that in Step 4 of `BSLB`; we sample from a carefully chosen subset of arms that has good coverage - more precisely, our goal is to solve the optimization problem

in equation 2 to obtain a representative set of arms. Although (Hao et al., 2020) had a similar objective, the absence of blocking constraint in their framework implied that they could solve for a probability distribution on the set of arms such that the minimum eigenvalue of the expected covariance matrix is maximized. Since their solution space was the probability simplex, the objective was continuous and concave - implying that a solution can be found efficiently.

However, in our setting, due to the blocking constraint, we need to identify a subset of representative arms from which to sample uniformly at random without replacement in the exploration component - this leads to the objective in equation 2 being discrete and therefore non-concave. Note that a brute force solution to our objective implies a search over all subsets of $[\mathsf{M}]$ and will take time $\Omega(\exp(\mathsf{M}))$. To design an efficient algorithm for obtaining a good feasible solution to the non-concave objective in 2, our first step is to obtain a concave relaxation as described in equation 5 - in particular, instead of optimizing over a subset, we optimize over probability distributions over the set of arms such that the probability of any arm is upper-bounded,

$$\begin{aligned}\hat{\boldsymbol{\mu}}(\hat{u}) = &\underset{\boldsymbol{\mu}\in\mathcal{P}(\mathcal{A})}{\arg\max}\ \lambda_{\min}\left(\mathbf{A}\mathsf{diag}(\boldsymbol{\mu})\mathbf{A}^{\mathsf{T}}\right)\\ &\text{such that}\ \ \|\boldsymbol{\mu}\|_\infty \le \frac{1}{\hat{u}},\end{aligned} \tag{5}$$

Note that $\mathbf{A} = [\mathbf{a}^1, \ldots, \mathbf{a}^{\mathsf{M}}]^{\mathsf{T}} \in \mathbb{R}^{\mathsf{M}\times d}$ denotes the matrix with all arms and $\hat{u}$ is an additional parameter to the relaxed objective. The solution to equation 5(which is a convex relaxation) need not be a feasible solution for the original discrete optimization problem we intended to solve in equation 2. Therefore, we take the solution of the convex relaxation and use randomized rounding (Step 3-5 in Algorithm 2) to obtain an approximately optimal solution to the discrete optimization problem. Let $\mathcal{X} \subseteq \mathcal{A}$ be the optimal subset for which the RHS in Equation 2 is maximized and let $\lambda^*_{\min}$ be the corresponding objective value (minimum eigenvalue of corresponding sample covariance matrix). We present the following theorem on the approximation guarantees of the solution achieved by our procedure `GetGoodSubset` of Algorithm 1 - the theorem says that the minimum eigenvalue of the sample covariance matrix associated with arms in $\mathcal{G}$ (obtained post randomized rounding procedure) is close to $\lambda^*_{\min}$.

Theorem 2. *Let $\mathbf{A} = [\mathbf{a}^1, \ldots, \mathbf{a}^{\mathsf{M}}]^{\mathsf{T}} \in \mathbb{R}^{\mathsf{M}\times d}$ denote the matrix of all arms. Consider the concave optimization of equation 5 solved at $\hat{u} = \mathsf{O}(\frac{d}{(\lambda^l_{\min})^{2/3}})$. Let $\mathcal{G}$ be the output of the randomized rounding procedure (Algorithm 2) and $\widehat{\lambda}_{\min}$ be the minimum eigenvalue of the corresponding covariance matrix that is, $\widehat{\lambda}_{\min} = \lambda_{\min}(\left|\mathcal{G}^{-1}\right|\sum_{\mathbf{a}\in\mathcal{G}}\mathbf{a}\mathbf{a}^{\mathsf{T}})$. Then under the assumption $\lambda^*_{\min} \ge \lambda^l_{\min}$, we must have $\widehat{\lambda}_{\min} \ge \frac{1}{4}\lambda^*_{\min}$ with probability $1 - o(1)$.*

Remark 3. Note that the time complexity is polynomial in the number of arms $\mathsf{M}^{\mathsf{O}\left(d{\lambda^l_{\min}}^{-2/3}\right)}$ (refer to Appendix C.7 for details) which is significantly improved than the trivial brute force algorithm which has a running time of $\mathsf{O}(\exp\mathsf{M})$. One can further reduce the time complexity to $\mathsf{O}(\mathsf{M})$ and remove the exponential dependence on d, by choosing $\hat{u} = \mathsf{O}((\lambda^l_{\min})^{-2}d)$ at the cost of a slightly worse approximation guarantee of $\widehat{\lambda}_{\min} = \Omega\left((\lambda^l_{\min})^2\lambda^*_{\min}\right)$ [1] (See Appendix C.6).

Several existing techniques in experimental design deal with maximizing objectives such as minimum eigenvalue; however, they assume submodularity or matroid constraints (Allen-Zhu et al., 2021b), which the average minimum eigenvalue of equation 2 does not satisfy (see Appendix C.8 for a detailed discussion).

Proof Outline: We first show in Lemma 4 (using concentration guarantees) that the following objective values are close: (A) value of the maximized concave objective with distribution $\hat{\boldsymbol{\mu}} \in \mathcal{P}(\mathcal{A})$ and parameter $\hat{u}$ in Equation 5 (B) objective value of the set $\mathcal{G}$ (equation 2) obtained via randomized rounding procedure from $\hat{\boldsymbol{\mu}}$ at $\hat{u}$ (line 4 in Algorithm 2). Note that the value of the maximized concave objective in Equation 5 with parameter g_1 is larger than the value with parameter g_2 provided $g_1 \le g_2$. Therefore, we show our approximation guarantees with respect to objective in Equation 5 with parameter $\hat{u} = \mathsf{O}(d)$. We show that the approximation guarantee holds with high probability if $|\mathcal{X}| > \hat{u}$ and do a brute-force search otherwise. Finally, given that the concave objective with parameter d in Equation 5 is a relaxation of the discrete objective in Equation 2, the objective value of the former is going to be larger than the objective value of the latter.

3.3 Online Guarantees for `BSLB`

Our next result is the expected regret incurred by `BSLB` in the online setting. The key ingredient in the regret analysis lies in appropriately accounting for the blocking constraint in the exploitation component of `BSLB`. Below, we present the result detailing the regret guarantees of `BSLB` when the exploration period $\mathsf{T}_{\text{explore}}$ is set optimally using a known sparsity level k.

Theorem 3. ***(Regret Analysis of `BSLB`)*** *Consider the d-dimensional sparse linear bandits framework with blocking constraint having a set $\mathcal{A} \subset \mathcal{B}^d_\infty$ of M arms spanning $\mathbb{R}^d$ and T rounds ($\mathsf{T} \ll d \ll \mathsf{M}$). Let $\|\mathbf{a}\|_\infty \le 1 \forall \mathbf{a} \in \mathcal{A}$. In each round $t \in [\mathsf{T}]$, we choose arm $\mathbf{a}_t \in \mathcal{A}$ and observe reward $r_t = \langle\boldsymbol{\theta}, \mathbf{a}_t\rangle + \eta_t$ where $\boldsymbol{\theta} \in \mathbb{R}^d$ is unknown and η_t is zero-mean independent*

[1] Note that $\lambda^l_{\min} \le \lambda^*_{\min} \le 1$ since $\|\mathbf{a}\|_\infty \le 1$.

noise random variable with variance $\sigma^2 = O(1)$. *Let* $R_{\max} = \max_{\mathbf{a}} \langle \boldsymbol{\theta}, \mathbf{a} \rangle$. *Suppose* $\boldsymbol{\theta}$ *has (unknown) tail magnitude* β_k *at fixed sparsity level* k. *Let* $\lambda^*_{\min}$ *for the set* $\mathcal{A}$ *be as defined in equation 2 and let* $\lambda^l_{\min}$ *be its known lower bound. Let* $\widehat{\lambda}_{\min}$ *be the minimum eigenvalue of the normalized of the sampled subset from Step 1 of* BSLB. *In this framework,* BSLB *with exploration period* $\mathsf{T}_{explore} = \widetilde{\mathsf{O}}(R_{\max}^{-2/3} \widehat{\lambda}_{\min}^{-2/3} k^{\frac{2}{3}} \mathsf{T}^{\frac{2}{3}})$, *regularization parameter* $\lambda = \sqrt{\frac{\log d}{n}}$ *and subset selection parameter* $\hat{u} = \mathsf{O}(\frac{d}{(\lambda^l_{\min})^{2/3}})$, *achieves a regret guarantee*

$$\mathbb{E}[\mathsf{Reg}(\mathsf{T})] = \widetilde{\mathsf{O}}\left(R_{\max}^{1/3} (\lambda^*_{\min})^{-2/3} (1+\beta_k)^2 k^{\frac{2}{3}} \mathsf{T}^{\frac{2}{3}} \right).$$

Proof Outline: In exploration component of the regret decomposition, we use Corollary 1 to obtain error guarantees of Lasso and Theorem 2 to bound the minimum eigenvalue of the sample covariance matrix of the sampled arms. We optimize the exploration period to obtain our stated result (Proof in App. C.3).

Note that when $\beta_k = 0$ (when the parameter vector is exactly k-sparse), our regret guarantee has the same order in T, k, $\lambda^*_{\min}$, and $R_{\max}$ as (Hao et al., 2020, Theorem 4.2), which operates *without* the blocking constraint. This demonstrates that the blocking constraint does not impose a statistical degradation in the regret rate. We note that the regret definitions in the two settings differ (ours compares against the top-T unique arms, while theirs compares against the single best arm played T times), so the comparison is in terms of the order of the bounds rather than exact constants. It is reasonable to compare the two since any non-blocking instance can be mapped to a blocking instance on a multi-set of arms (see Appendix B.6).

Insight 2. BSLB enables diversity in selected arms by performing Step 2 in Algorithm 1 which ensures that $\lambda_{\min}$ of the covariance matrix of the subset used in exploration is approximately optimal. The exploration period in Theorem 3 is set to maximize reward but can be accordingly tuned in practice to increase diversity.

Remark 4. As our lower bounds of Theorem 4 show, the $\lambda^\star_{\min}$ dependence is information-theoretically unavoidable (similar to (Hao et al., 2020)). To mitigate this, our exploration does randomized rounding for E-optimal design to obtain $\mathcal{G}$ with $\lambda_{\min}(\mathcal{G}) \geq \frac{1}{4}\lambda^*_{\min}$. In practice, a valid lower bound $\lambda^l_{\min}$ can be obtained by computing the minimum eigenvalue of the covariance matrix of the *entire* arm set $\mathcal{A}$ (ignoring the max in equation 2), since $\lambda_{\min}(\mathcal{A}) \leq \lambda^*_{\min}$. Alternatively, one can estimate $\lambda^l_{\min}$ via matrix resampling (e.g., subsampling random arm subsets and taking the maximum observed minimum eigenvalue) as explained in Appendix B.5.

Remark 5 (**Computational Complexity**). We decompose the runtime of BSLB into three parts. *(i) Subset selection:* GetGoodSubset($\mathcal{A}$, $\hat{u}$) runs in $\mathsf{Poly}(\mathsf{M})$ time but exponential in d due to the brute-force search (Appendix C.7); this is significantly better than the trivial $\mathsf{O}(\exp(\mathsf{M}))$. One can remove the exponential dependence on d by choosing $\hat{u} = \mathsf{O}((\lambda^l_{\min})^{-2} d)$ (Appendix C.6), at the cost of a multiplicative factor of $(\lambda^l_{\min})^{-2}$ in regret. *(ii) Exploration:* LASSO (Step 7) runs in $\mathsf{Poly}(\mathsf{M}, d, \mathsf{T})$. *(iii) Exploitation:* In Step 9, one computes $\langle \widehat{\boldsymbol{\theta}}, \mathbf{a} \rangle$ for all remaining arms *once* and stores the top-$(\mathsf{T} - \mathsf{T}_{\text{explore}})$ arms using a min-heap in $\mathsf{O}(\mathsf{M}d + \mathsf{M}\log\mathsf{T})$ time. Each subsequent round retrieves the next best arm in $\mathsf{O}(\log \mathsf{T})$, giving an overall exploitation cost of $\mathsf{O}(\mathsf{M}d + \mathsf{T}\log\mathsf{T})$. A linear dependence on M and d is unavoidable for exact search and is also present in (Hao et al., 2020). For very large M, one can use approximate Maximum Inner Product Search (e.g., HNSW-based methods) to achieve sub-linear query times with minimal accuracy loss.

Following is the lower bound on regret for high-dimensional linear bandits with blocking constraint.

Theorem 4. *Consider the* d*-dimensional sparse linear bandits framework with blocking constraint having a set* $\mathcal{A} \subseteq \mathcal{B}^d$ *of* M *arms spanning* $\mathbb{R}^d$ *and* T *rounds* ($\mathsf{T} \ll d \ll \mathsf{M}$). *In each round* $t \in [\mathsf{T}]$, *we choose arm* $a_t \in \mathcal{A}$ *and observe reward* $r_t = \langle \theta, a_t \rangle + \eta_t$ *where* $\theta \in \mathbb{R}^d$, *is unknown and* η_t *is zero-mean independent noise random variable with variance* $\sigma^2 = 1$. *Assume that the parameter vector is* k*-sparse,* $\|\theta\|_0 = k$. *Then for any bandit algorithm the worst case regret is,*

$$\mathbb{E}[\mathsf{R}] = \Omega(\min((\lambda_{\min}(\mathcal{A}))^{-1/3} k^{1/3} \mathsf{T}^{2/3}), \sqrt{d\mathsf{T}})).$$

Proof in Appendix C.1 For the case when the true parameter satisfies the exact sparsity condition (the tail β_k is 0), our regret guarantee (see Corollary 2 in Appendix) achieves the same $\mathsf{T}^{2/3}$ regret dependence as in (Hao et al., 2020) without the *blocking constraint*.

Remark 6 (**On matching rates under stricter constraints**). That our regret matches the non-blocking rate of (Hao et al., 2020) does not diminish the contribution; it demonstrates that new techniques are needed to *maintain* optimal performance under a stricter constraint. An analogous phenomenon occurs in mean estimation under heavy-tailed noise: the minimax rate remains $O(1/\sqrt{n})$, yet the sample mean fails and robust estimators (e.g., median-of-means) are required. Similarly, the algorithms of (Hao et al., 2020) do not extend to the blocking constraint (Figure 1), and our techniques (McDiarmid-based RE guarantees, E-optimal design with randomized rounding, corralling with covering argument) are essential. We note that (Hao et al., 2020) does not consider soft sparsity or unknown k.

3.4 Corralling for unknown optimal sparsity

Note that for any unknown parameter vector $\boldsymbol{\theta}$, we can fix the sparsity level k and therefore the corresponding tail magnitude β_k - subsequently, we can obtain the guarantees of Theorem 3 by setting the exploration period optimally for the fixed k. However, if k is set too low, then β_k will be too high, and therefore, the multiplicative term containing β_k in the regret bound dominates. There is a trade-off, and therefore, there is an optimal choice of sparsity k. Therefore, we propose a meta-algorithm `C-BSLB` that exploits coralling (Agarwal et al., 2017) multiple versions of the `BSLB` algorithm 1 with different values of k used to set the exploration period $\mathsf{T}_{\text{explore}}$ - the meta-algorithm gradually learns to choose the best base algorithm.

However, naively applying CORRAL with all distinct base algorithms leads to a linear dependence on dimension d in the regret making it vacuous. Therefore we carefully choose $\log d$ base algorithms for search within CORRAL with corresponding sparsity parameters set on exponentially spaced points - such a restriction ensures that the overhead in regret is minimal (logarithmic dependence on dimension d). However, we still prove our regret guarantee with respect to the base algorithm with optimal sparsity - although the optimal base algorithm may not be in the set of carefully chosen base algorithms in the meta-algorithm.

Theorem 5. *Consider the d-dimensional sparse linear bandits framework with blocking constraint as described in Theorem 3. Let the `C-BSLB` algorithm (Algorithm 4 in Appendix C.10) run with an appropriate learning rate on multiple versions of `BSLB`, using distinct sparsity parameter k taking values in the set $\{2^i\}_{i=0}^{\lfloor \log_2(d) \rfloor + 1}$. Let the optimal sparsity parameter in Theorem 3 that achieves minimum regret be $k^\star \in \{1, 2, \ldots, d-1, d\}$, and let $\mathbb{E}[\mathsf{Reg}(\mathsf{T})]_*$ be the corresponding regret. Then the meta-algorithm `C-BSLB` achieves the following,*

$$\mathbb{E}[\mathsf{Reg}(\mathsf{T})] = \mathsf{O}(\sqrt{\mathsf{T}\log_2(d)} + \sqrt{k^\star}\log_2(d)\mathbb{E}[\mathsf{Reg}(\mathsf{T})]_*).$$

Note that the first term above and the multiplicative factor of $\sqrt{k^\star}\log_2(d)$ corresponds to the additional cost in combining the input base algorithms by Algorithm 4. We stress that the dependence on dimension d from the additional cost is only logarithmic.

Proof Outline: We use (Agarwal et al., 2017, Theorem 5) and our bound from Theorem 3 to obtain Theorem 5. The key novelty is to establish the following - when searching with the small curated set of base algorithms in CORRAL, we do not suffer a significant loss in the regret even if the base algorithm with the optimal sparsity parameter does not lie in the curated set. The crux of our proof lies in a covering argument. By using a recursive telescoping argument, we can bound the regret incurred between any base algorithm not used for the search while corraling versus the base algorithm with the closest sparsity parameter used in the search.

4 EXPERIMENTS

We demonstrate `BSLB` for selecting hard datapoints for annotation on the PASCAL VOC 2012 image classification dataset (Everingham et al., 2015) with human difficulty scores from (Ionescu et al., 2016). The model $\mathcal{M}$ is an SVM head on a frozen ViT backbone. We compare against random sampling, training on all data, and state-of-the-art active learning methods AnchorAL (Lesci and Vlachos, 2024) and SEALS (Coleman et al., 2022). Results in Table 1 (averaged over 5 folds) show that `BSLB` with only T annotations achieves 5–14% improvement over random sampling on hard validation data, 3.5–7% over active learning baselines, and matches or exceeds a model trained on upto 12× more data.

Val. Type	Object	AnchorAL	SEALS	Random	All	**BSLB**
easy	chair	94.0±1.7	90.6±1.8	96.4±1.0	**96.0±1.1**	94.6±1.6
	car	94.5±1.6	94.7±4.0	97.7±2.1	**98.7±0.1**	96.5±1.8
	bottle	93.0±2.5	92.8±2.3	96.8±1.1	**96.8±1.1**	94.8±2.0
	bott./chair	91.5±1.1	92.3±1.1	94.8±1.0	**94.6±2.1**	91.7±2.2
hard	chair	69.3±3.1	69.6±6.1	66.0±3.8	71.3±3.2	**73.3±3.3**
	car	70.3±4.0	70.0±5.7	60.0±5.4	65.4±4.0	**74.0±3.4**
	bottle	63.1±2.9	63.4±3.4	59.7±4.4	64.8±1.9	**66.8±2.6**
	bott./chair	67.1±3.5	66.3±1.4	68.0±4.0	72.3±2.0	**73.0±1.7**

Table 1: Test accuracy on PASCAL-VOC for 4 object detection tasks. `BSLB` significantly outperforms baselines on **hard** validation data while remaining competitive on easy data. "All" uses 6×–12× more labeled samples.

Recommendation and Materials Discovery. On personalized recommendation (MovieLens, Netflix, Goodbooks, Jester), `BSLB` and `C-BSLB` consistently outperform LinUCB, DR-Lasso, and random baselines (Appendix A.2). On a materials discovery task (MatBench dielectric (Dunn et al., 2020), 4764 candidates, 204 features), our method identifies 62% more high-value materials than random selection while using only 49 of 204 features (Appendix A.4).

5 FUTURE WORK

We studied regret minimization in high-dimensional sparse linear bandits with blocking constraints, establishing sub-linear regret for `BSLB` under soft sparsity and minimal assumptions, and proposing `C-BSLB` to eliminate knowledge of the sparsity hyperparameter. The linear reward assumption does not capture binary or non-linear rewards; extending to generalized linear bandits is a natural direction. Incorporating robust estimation for heavy-tailed noise is another avenue for future work.

References

Abbasi-yadkori, Y., Pál, D., and Szepesvári, C. (2011). Improved algorithms for linear stochastic bandits. In Shawe-Taylor, J., Zemel, R., Bartlett, P., Pereira, F., and Weinberger, K., editors, *Advances in Neural Information Processing Systems*, volume 24. Curran Associates, Inc.

Abbasi-Yadkori, Y., Pal, D., and Szepesvari, C. (2012). Online-to-confidence-set conversions and application to sparse stochastic bandits. In Lawrence, N. D. and Girolami, M., editors, *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*, volume 22 of *Proceedings of Machine Learning Research*, pages 1–9, La Palma, Canary Islands. PMLR.

Agarwal, A., Luo, H., Neyshabur, B., and Schapire, R. E. (2017). Corralling a band of bandit algorithms. In *Conference on Learning Theory*, pages 12–38. PMLR.

Albalak, A., Elazar, Y., Xie, S. M., Longpre, S., Lambert, N., Wang, X., Muennighoff, N., Hou, B., Pan, L., Jeong, H., et al. (2024). A survey on data selection for language models. *arXiv preprint arXiv:2402.16827*.

Allen-Zhu, Z., Li, Y., Singh, A., and Wang, Y. (2021a). Near-optimal discrete optimization for experimental design: a regret minimization approach. *Math. Program.*, 186(1–2):439–478.

Allen-Zhu, Z., Li, Y., Singh, A., and Wang, Y. (2021b). Near-optimal discrete optimization for experimental design: a regret minimization approach. *Math. Program.*, 186(1-2):439–478.

Baek, J., Lee, S. J., Gupta, P., Dalmia, S., Kolhar, P., et al. (2024). Revisiting in-context learning with long context language models. *arXiv preprint arXiv:2412.16926*.

Bardenet, R. and Maillard, O.-A. (2015). Concentration inequalities for sampling without replacement. *Bernoulli*, 21(3):1361–1385.

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

Basu, S., Papadigenopoulos, O., Caramanis, C., and Shakkottai, S. (2021). Contextual blocking bandits. In Banerjee, A. and Fukumizu, K., editors, *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pages 271–279. PMLR.

Basu, S., Sen, R., Sanghavi, S., and Shakkottai, S. (2019). Blocking bandits. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.

Bengio, Y., Louradour, J., Collobert, R., and Weston, J. (2009). Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*, pages 41–48.

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

Bishop, N., Chan, H., Mandal, D., and Tran-Thanh, L. (2020). Adversarial blocking bandits. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H., editors, *Advances in Neural Information Processing Systems*, volume 33, pages 8139–8149. Curran Associates, Inc.

Boche, H., Calderbank, R., Kutyniok, G., and Vybíral, J. (2015). A survey of compressed sensing. In *Compressed Sensing and its Applications: MATHEON Workshop 2013*, pages 1–39. Springer.

Bresler, G., Chen, G. H., and Shah, D. (2014). A latent source model for online collaborative filtering. *Advances in neural information processing systems*, 27.

Bresler, G., Shah, D., and Voloch, L. F. (2016). Collaborative filtering with low regret. In *Proceedings of the 2016 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Science*, pages 207–220.

Brown, A., Laddha, A., and Singh, M. (2024). Maximizing the minimum eigenvalue in constant dimension.

Chakrabarti, D., Kumar, R., Radlinski, F., and Upfal, E. (2008). Mortal multi-armed bandits. In *Advances in Neural Information Processing Systems*, volume 21.

Chakraborty, S., Roy, S., and Tewari, A. (2023). Thompson sampling for high-dimensional sparse linear contextual bandits. In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23. JMLR.org.

Coleman, C., Chou, E., Katz-Samuels, J., Culatana, S., Bailis, P., Berg, A. C., Nowak, R., Sumbaly, R., Zaharia, M., and Yalniz, I. Z. (2022). Similarity search for efficient active learning and search of rare concepts. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(6):6402–6410.

Dong, Q., Li, L., Dai, D., Zheng, C., Ma, J., Li, R., Xia, H., Xu, J., Wu, Z., Liu, T., et al. (2022). A survey on in-context learning. *arXiv preprint arXiv:2301.00234*.

Dunn, A., Wang, Q., Ganose, A., Ganose, A., Jain, A., and Persson, K. A. (2020). Benchmarking ma-

terials property prediction methods: the matbench test set and automatminer reference algorithm. *npj Computational Materials*, 6(1):138.

Ethayarajh, K., Choi, Y., and Swayamdipta, S. (2022). Understanding dataset difficulty with $\mathcal{V}$-usable information. In Chaudhuri, K., Jegelka, S., Song, L., Szepesvari, C., Niu, G., and Sabato, S., editors, *PMLR*, volume 162 of *Proceedings of Machine Learning Research*, pages 5988–6008.

Everingham, M., Eslami, S. M. A., Van Gool, L., Williams, C. K. I., Winn, J., and Zisserman, A. (2015). The pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision*, 111(1):98–136.

Farr, D., Manzonelli, N., Cruickshank, I., Starbird, K., and West, J. (2024). Llm chain ensembles for scalable and accurate data annotation. In *2024 IEEE International Conference on Big Data (BigData)*, pages 2110–2118. IEEE.

Fung, V., Zhang, J., Juarez, E., Seshadri, R., and Zhao, J. (2021). Benchmarking graph neural networks for materials chemistry. *npj Computational Materials*, 7(1):84.

Goldberg, K., Roeder, T., Gupta, D., and Perkins, C. (2001). Eigentaste: A constant time collaborative filtering algorithm. *Inf. Retr. Boston.*, 4(2):133–151.

Guo, C., Zhao, B., and Bai, Y. (2022). Deepcore: A comprehensive library for coreset selection in deep learning. In *International Conference on Database and Expert Systems Applications*, pages 181–195. Springer.

Hao, B., Lattimore, T., and Wang, M. (2020). High-dimensional sparse linear bandits. *Advances in Neural Information Processing Systems*, 33:10753–10763.

Heckel, R. and Ramchandran, K. (2017). The sample complexity of online one-class collaborative filtering. In *International Conference on Machine Learning*, pages 1452–1460. PMLR.

Huleihel, W., Pal, S., and Shayevitz, O. (2021). Learning user preferences in non-stationary environments. In *International Conference on Artificial Intelligence and Statistics*, pages 1432–1440. PMLR.

Ionescu, R. T., Alexe, B., Leordeanu, M., Popescu, M., Papadopoulos, D., and Ferrari, V. (2016). How hard can it be? Estimating the difficulty of visual search in an image. In *Proceedings of CVPR*, pages 2157–2166.

Kim, G.-S. and Paik, M. C. (2019). Doubly-robust lasso bandit. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.

Komiyama, J. and Imaizumi, M. (2024). High-dimensional contextual bandit problem without sparsity. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, NIPS '23, Red Hook, NY, USA. Curran Associates Inc.

Lesci, P. and Vlachos, A. (2024). AnchorAL: Computationally efficient active learning for large and imbalanced datasets. In Duh, K., Gomez, H., and Bethard, S., editors, *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 8445–8464, Mexico City, Mexico. Association for Computational Linguistics.

Li, D., Wang, Z., Chen, Y., Jiang, R., Ding, W., and Okumura, M. (2024). A survey on deep active learning: Recent advances and new frontiers. *IEEE Transactions on Neural Networks and Learning Systems*.

Li, W., Barik, A., and Honorio, J. (2022). A simple unified framework for high dimensional bandit problems. In Chaudhuri, K., Jegelka, S., Song, L., Szepesvari, C., Niu, G., and Sabato, S., editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 12619–12655. PMLR.

Liu, Y., Liu, J., Shi, X., Cheng, Q., Huang, Y., and Lu, W. (2024). Let's learn step by step: Enhancing in-context learning ability with curriculum learning. *arXiv preprint arXiv:2402.10738*.

Madan, V., Singh, M., Tantipongpipat, U., and Xie, W. (2019). Combinatorial algorithms for optimal design. In Beygelzimer, A. and Hsu, D., editors, *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pages 2210–2258. PMLR.

Maharana, A., Yadav, P., and Bansal, M. (2023). D2 pruning: Message passing for balancing diversity and difficulty in data pruning. *arXiv preprint arXiv:2310.07931*.

Mavromatis, C., Srinivasan, B., Shen, Z., Zhang, J., Rangwala, H., Faloutsos, C., and Karypis, G. (2023). Which examples to annotate for in-context learning? towards effective and efficient selection. *arXiv preprint arXiv:2310.20046*.

Mindermann, S., Brauner, J. M., Razzak, M. T., Sharma, M., Kirsch, A., Xu, W., Höltgen, B., Gomez, A. N., Morisot, A., Farquhar, S., et al. (2022). Prioritized training on points that are learnable, worth learning, and not yet learnt. In *International Conference on Machine Learning*, pages 15630–15649. PMLR.

Nikolaenko, V., Ioannidis, S., Weinsberg, U., Joye, M., Taft, N., and Boneh, D. (2013). Privacy-preserving matrix factorization. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security*, CCS '13, page 801–812, New York, NY, USA. Association for Computing Machinery.

Pal, S., Suggala, A., Shanmugam, K., and Jain, P. (2024). Blocked collaborative bandits: online collaborative filtering with per-item budget constraints. *Advances in Neural Information Processing Systems*, 36.

Pukowski, P. and Lu, H. (2024). Investigating the impact of hard samples on accuracy reveals in-class data imbalance. *arXiv preprint arXiv:2409.14401*.

Ramakrishnan, N., Keller, B., Mirza, B., Grama, A., and Karypis, G. (2001). Privacy risks in recommender systems. *IEEE Internet Computing*, 5(6):54–63.

Rudelson, M. and Zhou, S. (2013). Reconstruction from anisotropic random measurements. *IEEE Transactions on Information Theory*, 59(6):3434–3447.

Sener, O. and Savarese, S. (2018). Active learning for convolutional neural networks: A core-set approach. In *International Conference on Learning Representations*.

Settles, B. (2009). Active learning literature survey. *Computer Sciences Technical Report 1648*.

Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A., and Potts, C. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1631–1642, Seattle, Washington, USA. Association for Computational Linguistics.

Sorscher, B., Geirhos, R., Shekhar, S., Ganguli, S., and Morcos, A. (2022). Beyond neural scaling laws: beating power law scaling via data pruning. *Advances in Neural Information Processing Systems*, 35:19523–19536.

Soviany, P., Ionescu, R. T., Rota, P., and Sebe, N. (2022). Curriculum Learning: A Survey. *International Journal of Computer Vision*, 130(6):1526–1565.

Tolstikhin, I. O. (2017). Concentration inequalities for samples without replacement. *Theory of Probability & Its Applications*, 61(3):462–481.

Vershynin, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Wainwright, M. J. (2019). *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Wang, X., Wei, M. M., and Yao, T. (2023). Efficient sparse linear bandits under high dimensional data. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '23, page 2431–2443, New York, NY, USA. Association for Computing Machinery.

Wu, B., Xu, C., Dai, X., Wan, A., Zhang, P., Yan, Z., Tomizuka, M., Gonzalez, J., Keutzer, K., and Vajda, P. (2020). Visual transformers: Token-based image representation and processing for computer vision.

Zajac, Z. (2017). Goodbooks-10k: a new dataset for book recommendations. *FastML*.

Zhang, X., Kumar, G., Khayrallah, H., Murray, K., Gwinnup, J., Martindale, M. J., McNamee, P., Duh, K., and Carpuat, M. (2018). An empirical exploration of curriculum learning for neural machine translation. *CoRR*, abs/1811.00739.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. Yes
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. Yes
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. No

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. Yes
   (b) Complete proofs of all theoretical results. Yes
   (c) Clear explanations of any assumptions. Yes

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). No
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). Not Applicable
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). Yes

(d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). Not Applicable

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. Not Applicable
   (b) The license information of the assets, if applicable. Not Applicable
   (c) New assets either in the supplemental material or as a URL, if applicable. Not Applicable
   (d) Information about consent from data providers/curators. Not Applicable
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. Not Applicable

5. If you used crowdsourcing or conducted research with human subjects, check if you include:
   (a) The full text of instructions given to participants and screenshots. Not Applicable
   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. Not Applicable
   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. Not Applicable

Appendix & Supplementary Materials

Contents

A NUMERICAL EXPERIMENTS

A.1 Simulation Study

We first validate `BSLB` and `C-BSLB` on a synthetic blocked sparse linear bandit instance with $\mathsf{M} = 10000$, $d = 1000$, $\mathsf{T} = 300$, and sparsity $k = 10$ (tail parameter $\beta_k = 3$), averaged over 100 random initializations. Figure 2 shows that `C-BSLB` achieves sub-linear regret *without* knowledge of the optimal hyperparameters, outperforming corralling with ℓ_2-regularization (which does not exploit sparsity) and a random policy. The non-monotonic cumulative regret arises because our comparator (Eq. 1) assigns the i^{th} best arm to round i; during exploitation, the algorithm may select the j^{th} best arm at round i with $j > i$, yielding negative instantaneous regret. *Performance of C-BSLB (without knowledge of true parameters)*

Finally, we also run a simulation study to study the efficacy of our `BSLB` and `C-BSLB` algorithm and demonstrate how CORRAL can be used to achieve a sub-linear regret without the knowledge of the optimal parameters. We compute the cumulative regret at time t compared to the top$-t$ arms, and unlike the standard bandit setting, in a blocked setting, the cumulative regret need not be monotonic. To highlight how our method exploits the sparsity of the parameter, we also run CORRAL with multiple versions of our algorithm but with a simple linear regression estimator. We simulate the experimental set-up with the following parameters $\mathsf{M} = 10000, d = 1000$ and $\mathsf{T} = 300$. At sparsity level $k = 10$, the tail parameter is $\beta_k = 3$. The experiment is repeated with 100 different random parameter initialization. We plot the cumulative regret in Fig. 3, for algorithms run with different exploration period and two versions of the CORRAL algorithm. Our C-BSLB performs better than corralling with ℓ_2-regularization, showing that our method exploits the sparsity and does not require true knowledge of the hyperparameters. We also benchmark against a random policy and show that our method performs significantly better showing that the upper bound on regret is not *vacuous*.

A.2 Application 1: Personalized recommendation with Single Rating per Item

Next we demonstrate our bandit algorithm on real-world recommendation datasets. However, we construct the recommendation task such that a) each item receives only a single rating from a user and b) we can only use previous recommendations of a user for recommending content. This is in contrast to the standard collaborative filtering setting where an item can where the ratings of the other users is used to recommend content to you. Our setting makes this possible by exploiting the additional information from the embeddings obtained from a pre-trained network for the text (or image) features of the different items. We argue that our setting is be more relevant in recommendation scenarios where privacy is a concern.

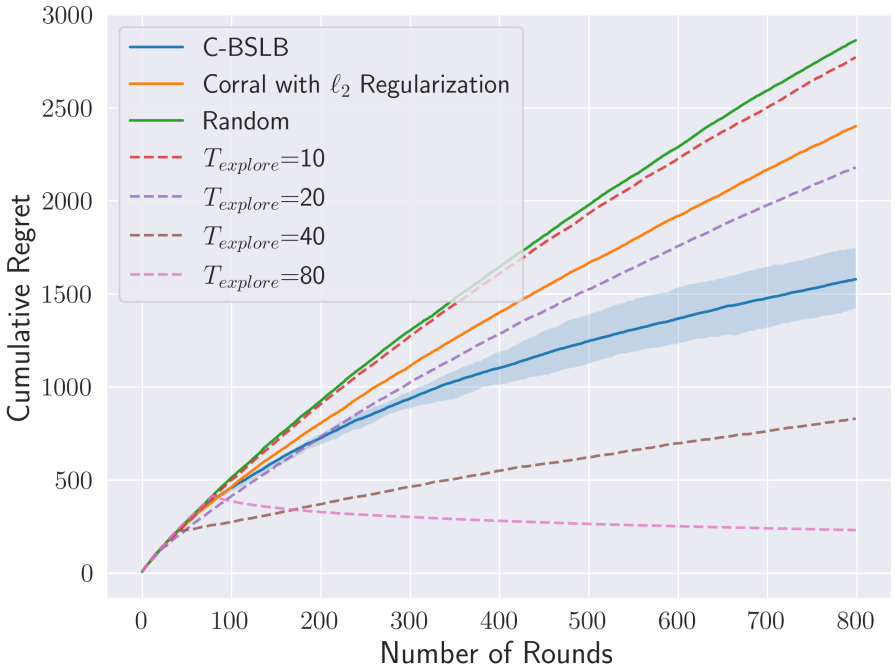

Figure 2: Cumulative regret on simulated instance (M=10000, d=1000, T=300). `C-BSLB` achieves sub-linear regret without knowledge of optimal hyperparameters.

We perform two sets experiments.

Experiment 1 on MovieLens and Netflix data (Movie Ratings) :

The MovieLens 100K dataset contains 100,000 5-star ratings from 1,000 users on 1,700 movies . For our analysis, we selected the 100 most active users (those with the highest rating counts) and the 100 most-rated movies. Using rating data from 50 users, we first applied matrix factorization to complete the user-item rating matrix, then derived 40-dimensional movie embeddings. To create synthetic high-dimensional embeddings, we extended these to 120 dimensions by randomly sampling additional coordinates uniformly from [0,1]. Results averaged across 5 test users (shown in Figure 4) demonstrate that our algorithm and ESTC outperform LinUCB and DR-Lasso, with the performance gap attributable to our method's effective exploitation of high-dimensional sparsity patterns . This aligns with theoretical expectations for sparse learning scenarios in recommendation systems. The Netflix Prize dataset contains 100 million 5-star ratings from 480,189 users across 17,770 movies, collected between 1998-2005. For analysis, we selected the 200 most active users (those with the highest rating counts) and 400 most-rated movies. Using ratings from 100 users, we applied matrix factorization to complete the user-item interaction matrix and derived 40-dimensional latent movie embeddings. To simulate high-dimensional sparse representations, we extended these embeddings to 120 dimensions by randomly sampling additional coordinates from [0,1]. Results averaged across 10 test users (shown in Figure 5) demonstrate that our algorithm and ESTC outperform baseline methods like LinUCB and DR-Lasso. This performance gap highlights the advantage of exploiting sparsity patterns in high-dimensional latent factor models.

Experiment 2 using Embeddings from Content Information: We run the corralling algorithm using copies = 4 copies of Algorithm 1 each with different exploration periods, $\mathsf{T}_{\text{explore}}$. Each of the instance, we first give $\mathsf{T}_{\text{explore}}$ random recommendations by sampling uniformly without replacement from a suitably constructed subset $\mathcal{G}$ to each user. Given the ratings obtained, we estimate the parameter $\boldsymbol{\theta}_{\text{user}}$ specific to the user using only their recommendations. For the remaining $\mathsf{T}_{\text{exploit}} = \mathsf{T} - \mathsf{T}_{\text{explore}}$ rounds, we give the top $\mathsf{T}_{\text{exploit}}$ recommendations based on the estimated parameter. To benchmark we run the algorithms independently and also against a random policy which randomly recommends. We next describe the two tasks that we report our results on for experiment 2,

Goodbooks-10k (Book Reviews): We use the Goodbooks-10k for a personalized book recommendation tasks (Zajac, 2017). For each book we use the title and the author to obtain embeddings using the MiniLM-L6-v2 sentence transformer which we use as the feature vectors for the arms. There are $\mathsf{M} = 1500$ books and we consider

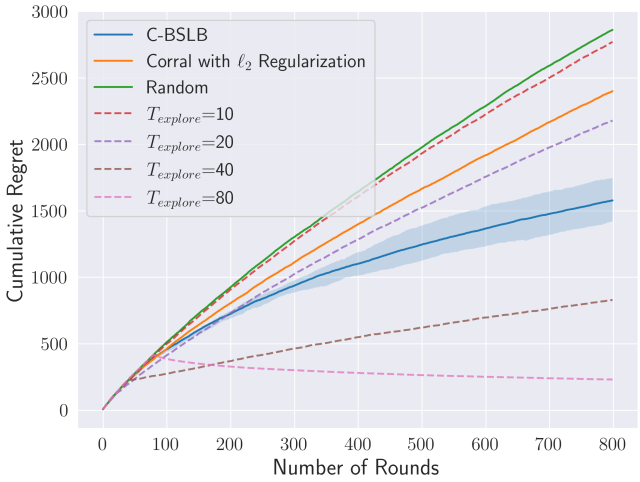

Figure 3: Regret of different algorithms in a Simulated Blocked Sparse Linear Bandit Setup.

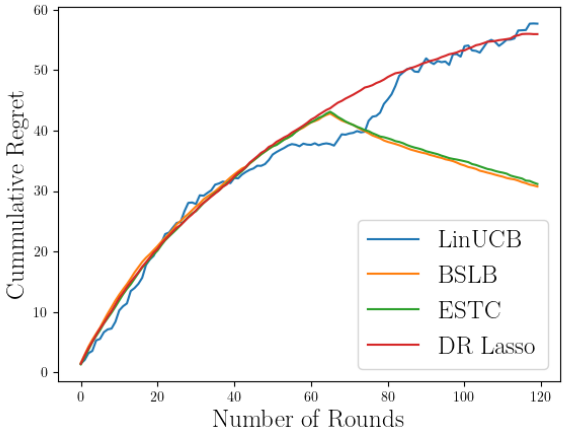

Figure 4: Numerical experiment on MovieLens illustrating performance gap between our proposed algorithm `BSLB` and naive extensions of LinUCB and DR-Lasso to incorporate blocking constraint. The performance of extended `ESTC` remains competitive.

10 users which have more than 600 ratings. The ratings are between 1 to 5. We consider the exploration periods as $[100, 150, 200, 300]$.

Jester (Joke Ratings): We use the Jester joke dataset 1 which has ratings on 100 jokes by 24,983 users (Goldberg et al., 2001). We obtain embedding for the jokes using the same transformer as above. For experimental purposes we filter out users which do not have ratings on all the jokes and are left with 7200 users. We run our algorithm with 10 different random seeds for each of the 7200 users and report the results averaged across all users. The joke ratings range from -10 to 10. For different algorithm instances $\mathsf{T}_{\text{explore}}$ is taken to be $[20, 40, 60, 80]$. **Results:** We summarize the cummulative regret from equation 1 of the algorithms in Figure 7. We add the random policy as a reference. We see that for the different dataset our the algorithm achieves a sub-linear regret. The reason the cumulative regret is not monotonic is due to the fact that the regret is with respect to the top-T arms. It can be seen that our algorithm with Corral achieves a performance close to the performance of the algorithm with the exploration period, $\mathsf{T}_{\text{explore}}$ out of the 5.

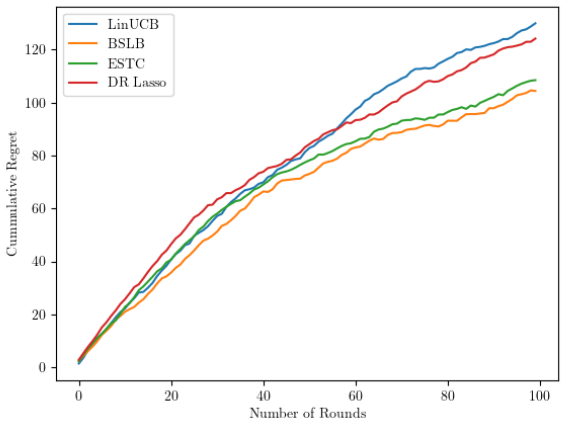

Figure 5: Numerical experiment on Netflix dataset illustrating performance gap between our proposed algorithm `BSLB` and naive extensions of LinUCB and DR-Lasso to incorporate blocking constraint. The performance of extended `ESTC` remains competitive.

A.3 Application 2: Adaptive Annotation using Difficulty Feedback from Annotators

Below, we demonstrate our methods for annotation in a label-scarce setting for image classification on the PASCAL VOC 2012 dataset. Additional experimental results on SST-2 (text dataset) can be found in Appendix A.3.4. Finally, experiments on the MovieLens, Netflix, and GoodBooks datasets in the context of personalized recommendation with few labeled data using our theoretical framework are in Appendix A.2. Finally, we provide detailed simulations in Appendix A.1.

We consider the setting where we have a total of M unlabelled samples (with $\mathsf{T} \ll \mathsf{M}$) and only T datapoints can be annotated (sequentially). For each unlabeled datapoint sent for annotation to the expert(s), we receive the ground truth label and the *difficulty score* r_t corresponding to the difficulty in annotating the datapoint. We showcase the effectiveness of `BSLB` (Algorithm 1) in our experimental set-up with real-world datasets. Given a model $\mathcal{M}$ to be trained on a downstream task, to benchmark `BSLB`, we consider the following set of baselines (to compare against) to choose subset of datapoints for annotation and subsequent training of the aforementioned model $\mathcal{M}$:

1. **Random**: Subset of T unlabeled datapoints chosen uniformly at random
2. **All**: All the samples in training data (except the validation fold)
3. **AnchorAL** (Lesci and Vlachos, 2024): an anchoring based active learning baseline (T samples).
4. **SEALS** (Coleman et al., 2022): a KNN based sub-sampling active learning baseline (T samples).

$\tau_{\text{easy}}, \tau_{\text{hard}}$ (thresholds on difficulty score to determine easy/hard samples) and $\mathsf{T}_{\text{explore}}$ (exploration rounds) are relevant hyper-parameters specified for the corresponding experiments [2]. We benchmark learning performance on 2 datasets: a) N_{valid} hard samples (samples with difficulty $> \tau_{\text{hard}}$) (**hard-valid**) b) N_{valid} easy samples (samples with difficulty ratings $< \tau_{\text{easy}}$) (**easy-valid**).

AnchorAL and **SEALS** are state-of-the-art active learning (AL) algorithms. In general, for a label-scarce complex task, AL might not be immediately applicable (see cold-start problem in (Li et al., 2024)) - especially for datapoints close to the decision boundary with noisy/less informative confidence intervals. This is because AL requires a reliably trained model on an even smaller subset of labeled datapoints - however, on datapoints far from the decision boundary (easy datapoints), noisy confidence signals are still useful. As we show in our experiment, this intuition holds, and the AL models, along with the **random** baseline, perform well on the **easy-valid** dataset. It is worth noting that complex (hard) datapoints often tend to be the main challenge in industrial applications. This is because it is *easy* to improve performance on easy data (cheaper to obtain) by simply increasing samples during training, but hard datapoints are difficult to generalize on (Pukowski and Lu, 2024).

[2] We consider the AL setup initialized with $\mathsf{T}_{\text{explore}}$ samples and $\mathsf{T} - \mathsf{T}_{\text{explore}}$ samples queried in a batch.

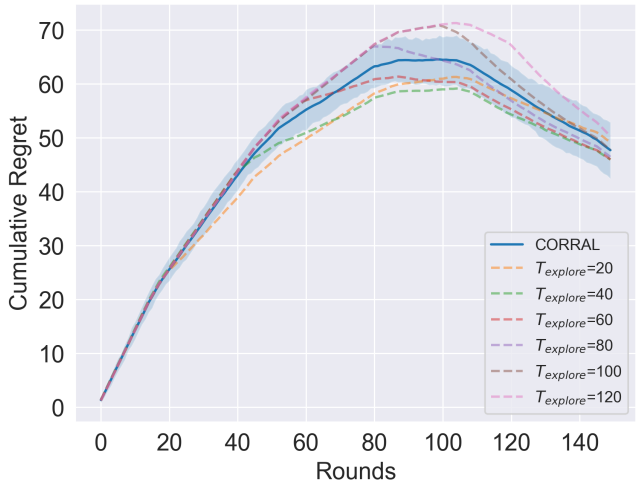

Figure 6: Cumulative Regret for recommendation using only single ratings using `BSLB` with different exploration periods and when run with CORRAL (Agarwal et al., 2017) in Books Dataset.

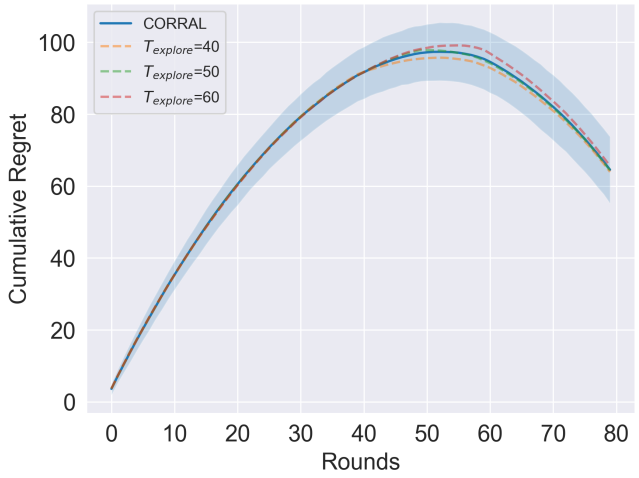

Figure 7: Cumulative Regret for recommendation using only single ratings using `BSLB` with different exploration periods and when run with CORRAL (Agarwal et al., 2017) in Jester.

Validation Type	Object Annotated	AnchorAL	SEALS	Random	All	**Our** (BSLB)
easy-valid	chair	94.0 ± 1.67	90.6 ±1.8	96.4 ±1.0	**96.0± 1.1**	94.6 ±1.6
	car	94.5± 1.6	94.7± 4.0	97.7±2.1	**98.7 ±0.1**	96.5± 1.8
	bottle	93.0±2.5	92.8±2.3	96.8±1.1	**96.8±1.1**	94.8±2.0
	bottle or chair	91.5±1.1	92.3±1.1	94.8±0.97	**94.6±2.1**	91.7±2.2
hard-valid	chair	69.3 ± 3.1	69.6± 6.1	66.0± 3.8	71.3± 3.2	**73.3± 3.3**
	car	70.3 ±4.0	70.0± 5.7	60.0± 5.4	65.4± 4.0	**74.0± 3.4**
	bottle	63.1±2.9	63.4±3.4	59.7±4.4	64.8±1.9	**66.8±2.6**
	bottle or chair	67.1±3.5	66.3±1.4	68.0±4.0	72.3±2.0	**73.0±1.7**

Table 2: Test accuracy of model $\mathcal{M}$ trained on different subsets of data annotated for 4 distinct object detection tasks in an image (PASCAL-VOC): The test performance of BSLB approach on the easy and hard validation dataset is at par with the $\mathcal{M}$ trained on all samples. We perform significantly better on the hard validation dataset compared to random sampling and active learning baselines.

A.3.1 Task 1: Image Classification on PASCAL VOC 2012

Our main result is for the image classification task on the public image dataset, PASCAL VOC 2012 (Everingham et al., 2015). The dataset has $11,540$ unique images and comprises segmentations for 20 objects. In addition to the image dataset, we use difficulty scores of annotations from (Ionescu et al., 2016) - the authors have provided the visual search difficulty by measuring the time taken to annotate in a controlled environment. The annotation task here was to identify if an image contains a particular object, e.g. "Does this image contain a car". The authors derive a difficulty score between 0 and 8 by normalizing the time to annotate.

In our experiment, the goal is to train a learning model for image classification - $\mathcal{M}$ is a support vector machine (SVM) head attached to a frozen pre-trained vision transformer (ViT) model pre-trained on ImageNet-21k dataset (Wu et al., 2020). We present results on the classification task - given an input image, predict if the image has an *object* or not. We consider 4 different objects, namely chair, car, bottle, and (bottle or chair). The last object is an OR conjunction of two labels. We consider the thresholds as $\tau_{\text{easy}} = 3.1$ and $\tau_{\text{hard}} = 3.9$ since the distribution of the difficulty scores in the dataset is heavy-tailed as shown in Figure 9. The (image, question) tuple with difficulty scores in the range $[3.1, 3.8]$ are highly noisy and therefore have been excluded. Table 3 contains the hyperparameters T, $\mathsf{T}_{\text{explore}}(\approx 0.6\mathsf{T})$ used and the number of samples in the **all** dataset for the different object classification tasks, along with the size of the validation datasets **hard-valid** and **easy-valid** and aggregaged accuracies. Table 4 contains results on the effect of varying $\mathsf{T}_{\text{explore}}$.

We present our results in Table 2 averaged over 5 validation folds. For this classification task, our method (BSLB) efficiently selects datapoints (to be annotated) compared to baselines with an equal number of samples. Regarding the quality of the final trained model $\mathcal{M}$, the learning performance of BSLB on **easy-valid** is within 2% of that obtained by the baseline **random**. However, there is an improvement of $5-14\%$ on the hard validation data **hard-valid**. When compared to the active learning baselines (**AnchorAL** and **SEALS**), BSLB performs better by $1-4\%$ on **easy-valid** and by $3.5-7\%$ on **hard-valid**. Finally, when compared to $\mathcal{M}$ trained on all datapoints (**all** baseline), which has $6\times$ to $12\times$ more samples, our method does better (0.7% to 8.6%) on the **hard-valid** and does decently on **easy-valid** ($< 3\%$ difference). These results validate our theory - in particular, we find that performance on **easy-valid** improves if the model $\mathcal{M}$ is trained on more samples (randomly chosen to improve coverage). However, improving the performance on **hard-valid** dataset is the main challenge where our simple approach BSLB with theoretical guarantees does reasonably well.

A.3.2 Validation on Different Hyperparameters

The hyperparameters are presented in Table 3. Note that the reason for selecting different T across different objects was because the validation datasets had to be big enough (so that the variance of accuracy is informative) and different objects had different total number of samples. The number of exploration rounds are set with respect to $\mathsf{T}(\sim 0.5\mathsf{T} - 0.7\mathsf{T})$ so that the approximation error after exploration is small enough. The active learning methods are also run with random initialization of $\mathsf{T}_{\text{explore}}$ explorations and one round of querying with

$\mathsf{T} - \mathsf{T}_{\text{explore}}$ queries. We present the results of a study where we vary only the $\mathsf{T}_{\text{explore}}$ for the same value of T in Table 4 and observe that with decreasing $\mathsf{T}_{\text{explore}}$ the estimation of difficulty scores deteriorates, and the performance on hard-valid deteriorates. The performance on easy-valid improves since samples are randomly chosen if the estimation is unsuccessful. Note that our method still performs better than AL baselines and random sampling.

A.3.3 Correlation between model difficulty and human annotation difficulty

In Figure 8, we show that for the chair images, if a model $\mathcal{M}$ is trained on all images, then the fraction of difficult samples at a certain distance from the classifier boundary goes down as the distance from the classifier body increases; especially after a certain distance from the decision boundary.

A.3.4 Task 2: Text Classification on SST-2

Next we present a result on text classification task on SST-2 (Socher et al., 2013). However since there were no human difficulty ratings available for this task, we use rarity of text (Zhang et al., 2018) as a heuristic for the difficulty ratings. The learning model is a SVM which classifies sentence embeddings obtained from the MiniLM-L6-v2 transformer. We consider $\mathsf{T}_{\text{explore}} = 100$ samples and $\mathsf{T}_{\text{exploit}} = 200$. The normalized rarity ranges from 0 to 1 and we set $\tau_{\text{hard}} = 0.5$ and $\tau_{\text{easy}} = 0.2$.

We observe a similar trend as the previous task where **BSLB** method performs better than a **random** subset by 3% and as good as the **random-large** subset on the **hard-valid** dataset. There is no regression on the **easy-valid**. The results on both the validation sets are comparable with **mixed** dataset which require all the difficulty ratings (which can be computed for the heuristic but not otherwise). **BSLB** performs better than both active learning methods on both the validation sets by 2%. However, since this is a standard sentiment analysis task, the embeddings are more informative, thereby improving the baseline performance for a **random** subset.

A.4 Application 3: Material Discovery

In materials discovery, each candidate composition or structure corresponds to an arm and a "pull" means running a synthesis/measurement or a high-fidelity simulation (Fung et al., 2021). These trials are costly and non-repeatable in the short term, so a per-candidate single-pull (blocking) constraint is inherent rather than optional.

Concretely, we demonstrated the framework on a MatBench dielectric task (Dunn et al., 2020): given thousands of crystalline candidates, the goal is to discover high–refractive-index materials using as few evaluations as possible. Blocking is unavoidable here—re-measuring the same sample adds cost without information—so we adopt one-shot pulls and featurize each structure with composition and site/structure descriptors (via matminer). In this real-world-scale study, BSLB (and its corralling variant C-BSLB for unknown sparsity k) substantially increased the hit rate of high-value discoveries and lowered cumulative regret relative to random and uncertainty-sampling baselines, while automatically focusing on a small subset of informative descriptors—illustrating both sample-efficiency and interpretability benefits for laboratory triage and high-throughput screening.

We evaluated the sparse linear bandit approach for discovering high refractive index materials using the MatBench dielectric dataset, which contains 4,764 crystalline materials with experimentally measured refractive indices ranging from 1.0 to 62.1 (mean = 2.43, std = 2.10).

The objective was to identify materials with high optical performance ($n > 2.5$), representing 28% of the dataset, using the minimum number of experimental evaluations. The **blocking constraint** is unavoidable in such a setting since one does not want to measure the same material twice, each experiment is costly due to acquisition and experimentation cost.

We extracted 204 high-dimensional features from crystal structures using matminer, including composition-based descriptors (Magpie elemental properties, stoichiometry, valence orbital characteristics) and structure-based descriptors (site statistics fingerprints, structural heterogeneity, chemical ordering). After preprocessing to remove highly correlated and constant features, all 4,764 samples were retained with a final dimensionality of 204 features that were standardized for analysis. For computational efficiency, experiments were conducted on a randomly selected subset of 1,000 materials with 253 (25.3%) exhibiting the target high refractive index property.

Across five independent experimental runs of 200 material evaluations each, the sparse linear bandit algorithm demonstrated substantial improvements over baseline approaches in discovering high-value materials. The sparse bandit identified 89.6 ± 6.7 materials with $n > 2.5$ compared to 55.2 ± 3.3 for random selection, representing a 62.3% improvement in discovery efficiency. The algorithm achieved an average refractive index of 2.828 ± 0.103 versus 2.41 ± 0.041 for random selection, with significantly lower cumulative regret (228.70 ± 13.26 vs 316.05 ± 22.77). A greedy uncertainty sampling baseline that selected materials farthest from known datapoints achieved intermediate performance (76.2 ± 3.4 high-value discoveries). Notably, the sparse bandit achieved effective dimensionality reduction by identifying and utilizing only 49 of the 204 available features (24% sparsity), demonstrating the algorithm's ability to automatically discover the most relevant materials descriptors for predicting optical properties while maintaining superior discovery performance.

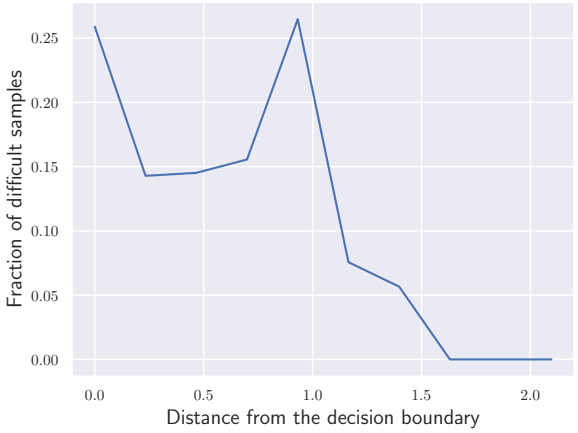

Figure 8: Fraction of difficult samples (labelled by humans) against the distance from decision boundary for SVM trained on all *chair* images. As the distance from the decision boundary increases the fraction of difficult samples (difficulty rating from humans > 3.5) decays to 0.

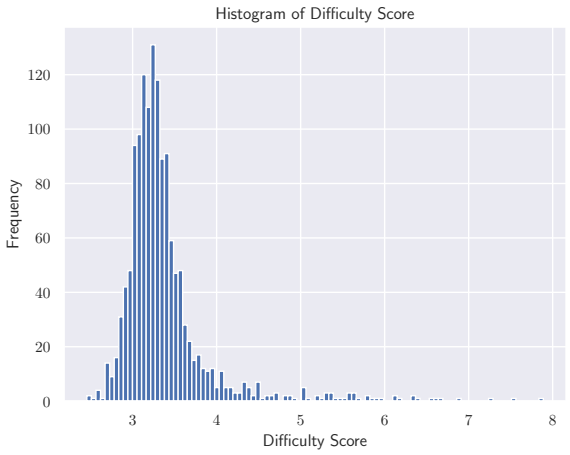

Figure 9: Histogram shows the heavy-tailed distribution of the difficulty score from (Ionescu et al., 2016) of the *chair* object of the PASCAL-VOC dataset. We clip the entries from the middle since they make the difficult estimation noisier, in practical implementation, one would need to develop a mechanism to flag samples with *ambiguous* difficulty and this is left for future work.

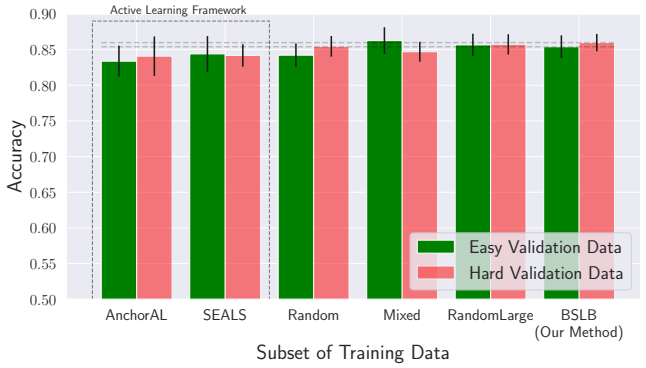

Figure 10: Text Classification on SST-2: The gains are not substantial on the text-classification, but show that our methods are task agnostic. Although conceptually active learning also does adaptive annotation, our method performs better (especially on **hard-valid**) in the label-scarce setting when $\mathsf{T} \ll d$ and the hardness of the samples considered.

Object Being Annotated	N_{valid} **(hard)**	N_{valid} **(easy)**	T	$\mathsf{T}_{\text{explore}}$	**All**		**BSLB**
					Num Samples	**Averaged Accuracies**	**Averaged Accuracies**
chair	60	80	100	80	960 (10x)	83.65	**83.95**
car	70	100	90	60	1227 (13x)	82.05	**85.25**
bottle	70	100	120	60	822 (6x)	80.8	**80.8**
bottle or chair	120	120	140	100	1807 (13x)	**83.45**	82.35

Table 3: Different hyperparameters used for the experiment of Sec∼A.3. The num samples show how our method achieves a similar accuracy (-1% to 4% improvement over all) by considering substantially less samples.

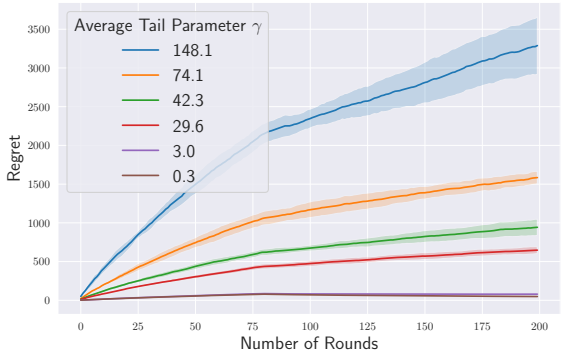

Figure 11: Effect of the tail parameter β_k on the performance of the `BSLB` algorithm with $\mathsf{T}_{\text{explore}} = 80$. As the tail increases in magnitude the cummulative regret worsens (increases). However observe that our algorithm is still robust to reasonably large tail $\beta_k = 75$.

A.5 How does the tail of the parameter matter?

In Figure 11, we investigate the effect of the tail parameter in the performance of `BSLB` with a fixed exploration period $\mathsf{T}_{\text{explore}} = 80$ and different sizes of the tail in the same setup as the simulation study of Appendix A.1. We observe that as the tail parameter β_k grows, the regret worsens, however we remark that even for a decent $\beta_k = 75$, the performance is reasonable.

	Validation Type	Object Annotated	AnchorAL	SEALS	Random	All	**Our** (BSLB)
$\mathsf{T}_{\text{explore}} = 80$	easy-valid	chair	94.5±1.0	93.2±1.5	95.7±1.0	**96.0±1.4**	92.3±1.4
		car	95.8±2.5	95.3±2.0	97.3±1.2	**98.2±1.1**	95.7±1.7
		bottle	95.0±0.9	95.0±1.1	96.2±1.9	96.7±1.8	**96.8±1.2**
	hard-valid	chair	72.0±3.0	71.0±1.4	67.0±5.0	69.4±4.1	**73.4±2.4**
		car	66.4±7.9	69.2±5.6	52.6±8.2	58.8±6.8	**74.0±2.7**
		bottle	61.2±1.5	61.8±2.0	51.2±2.9	52.6±3.1	**63.2±2.9**
$\mathsf{T}_{\text{explore}} = 60$	easy-valid	chair	94.8±2.1	93.7±1.7	95.5±2.0	**95.5±1.4**	94.8±1.8
		car	96.2±1.9	95.5±1.9	97.5±1.2	**98.3±1.1**	97.0±2.3
		bottle	95.0±1.2	94.8±1.0	97.2±1.7	**97.5±1.7**	96.0±1.9
	hard-valid	chair	70.0±5.5	70.0±3.8	68.0±4.2	69.8±4.0	**72.4±1.7**
		car	65.8±8.7	70.0±7.5	53.2±5.4	60.6±8.1	**72.6±3.9**
		bottle	61.6±1.0	61.6±2.3	53.4±2.6	54.0±2.6	**62.6±2.8**
$\mathsf{T}_{\text{explore}} = 30$	easy-valid	chair	94.8±1.5	94.5±1.2	96.3±1.4	**96.7±1.9**	94.5±3.2
		car	95.3±2.3	95.8±1.9	97.3±1.2	**98.2±1.1**	92.0±11.6
		bottle	95.3±0.8	95.0±0.5	96.2±1.9	96.7±1.8	**96.7±1.2**
	hard-valid	chair	69.4±2.1	**71.2±2.4**	67.6±4.9	69.8±4.0	70.8±4.5
		car	64.2±9.5	67.6±7.7	52.6±8.2	58.8±6.8	**70.8±6.5**
		bottle	60.2±1.9	62.0±2.1	51.2±2.9	52.6±3.1	**63.0±4.9**

Table 4: Learning Accuracies on Different Methods for Image Classification in PASCAL-VOC 2012: Effect of $\mathsf{T}_{\text{explore}}$ with the number of rounds fixed at $\mathsf{T} = 120$ and with 120 **easy-valid** and 100 **hard-valid** samples.

A.6 Sensitivity to Problem Parameters

We report the final-iterate regret of BSLB when varying individual problem parameters (β_k, k, d), with the remaining parameters held fixed, in Tables 5–7. These experiments use $\mathsf{M} = 1000$ arms, $\mathsf{T} = 50$ rounds, and $\mathsf{T}_{\text{explore}} = 80$ exploration rounds.

Table 5: Varying tail parameter β_k (d=100, k=5).

Tail β_k	**Regret**
0.01	43.00
0.1	48.67
1.0	164.82
2.5	402.35
5.0	823.01

As expected, regret increases with larger tail β_k (weaker sparsity), larger sparsity k, and larger dimension d, consistent with the dependencies in Theorem 3.

A.7 Convergence of CORRAL parameters in C-BSLB

We plot the convergence of the CORRAL parameters of C-BSLB in Figure 12 for the simulated experiment of Appendix A.1. We observe that the probability of sampling the best algorithm ($\mathsf{T}_{\text{explore}} = T_3 = 80$) increases with rounds. Note that since the experiments were run on a limited resource machine, we could only do $d = 1000$, and for our setup $\mathsf{T} \ll d$ has to be sufficiently low (500 in this case). This is not enough for the CORRAL algorithm to truely exploit the best possible algorithm in C-BSLB but as we see in Figure 3, however it still achieves the a decent performance.

Table 6: Varying sparsity k (d=100, β_k=0.4).

k	**Regret**
5	359.10
10	378.79
20	393.51

Table 7: Varying dimension d (k=10, β_k=0.4).

d	**Regret**
50	295.19
100	396.88
200	578.69

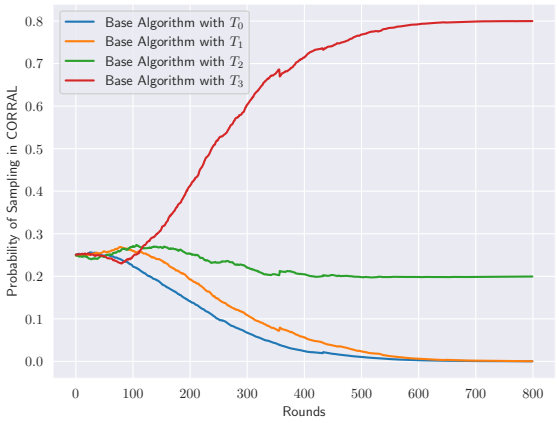

Figure 12: Convergence of the different sampling probabilities for the base algorithms of the `C-BSLB` (Algorithm 4). This plot is with respect to the simulation study parameters. We can observe that the probability for the best algorithm (T_3) improves with each iteration and for the worst performing algorithm (T_0) decays to 0.

B RELATED WORK

The Appendix comprises three sections: Section B discusses the related work, Section C details the detailed proofs of the technical results and Section A presents additional numerical results.

B.1 High Dimensional Sparse Bandits

Recent research on high-dimensional sparse linear bandits has focused on developing algorithms that can effectively handle the challenges posed by high-dimensional feature spaces while exploiting the underlying sparsity. (Hao et al., 2020) made significant contributions by establishing a novel $\Omega(n^{2/3})$ dimension-free minimax regret lower bound for sparse linear bandits in the data-poor regime, where the horizon is smaller than the ambient dimension. They complemented this with a nearly matching upper bound for an explore-then-commit algorithm, demonstrating that $\mathsf{O}(n^{2/3})$ is the optimal rate in the data-poor regime. Building upon these foundations, (Li et al., 2022) proposed a simple unified framework for high-dimensional bandit problems, including sparse linear bandits. Their Explore-the-Structure-Then-Commit (ESTC) algorithm achieved comparable regret bounds in the LASSO bandit problem and provided a general framework for the contextual bandit setting using low-rank matrices and group sparse matrices. (Wang et al., 2023) further improved the algorithm of (Hao et al., 2020) for computational complexity using ideas from random projection. On the other hand, (Chakraborty et al., 2023) has recently looked at contextual high dimensional bandits with exact sparsity constraints using Thompson sampling with a

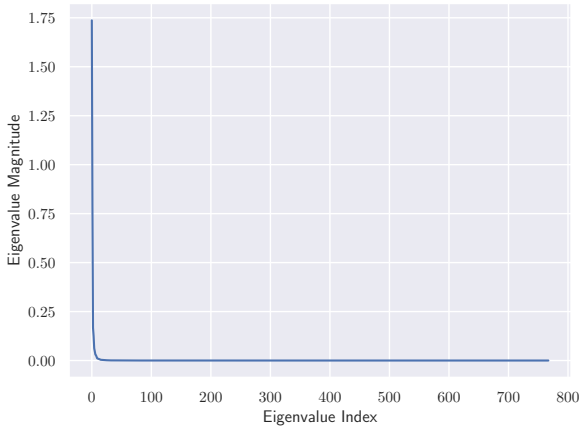

Figure 13: Eigenvalue spectrum of the embeddings of the two dataset show exponential decay in the eigenvalues, which implies that a uniformly random sample covers the set optimally with high probability because the data is primarily shaped by a few directions. For the PASCAL-2012 on object **chair** with ViT Base Patch16-224 embeddings.

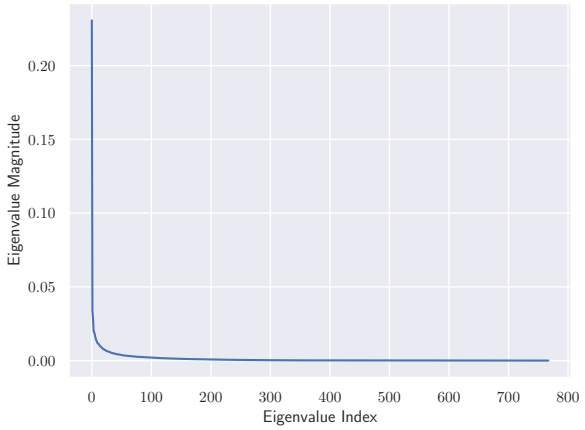

Figure 14: Eigenvalue spectrum of the embeddings of the two dataset show exponential decay in the eigenvalues, which implies that a uniformly random sample covers the set optimally with high probability because the data is primarily shaped by a few directions. Balanced Sample (2500 datapoints) of SST-2 with All MPNet Base V2 embeddings.

prior on the sparsity parameter and (Komiyama and Imaizumi, 2024) have studied high dimensional bandits without a strict sparsity constraint. However both these work give asymptotic bounds on the regret whereas our work provides finite-sample bounds on the regret. Further the existing work does not explicitly quantify the effect of the parameter tail on the regret or consider a *blocking* constraint, which are the two primary novel considerations in our paper.

B.2 Personalized Recommendation

Related work on privacy-preserving recommender systems highlights the tension between personalization and user privacy. Traditional recommender systems, such as collaborative filtering and matrix factorization, often rely on centralized data collection and processing, which creates privacy risks for users. These risks include potential re-identification of anonymized data and unauthorized access to sensitive information (Ramakrishnan et al., 2001). To address these concerns, researchers have explored various privacy-preserving techniques, including

differential privacy, secure multi-party computation, homomorphic encryption, and federated learning (Nikolaenko et al., 2013). These approaches aim to protect user data while maintaining recommendation accuracy. However, challenges remain in balancing privacy protection with system efficiency and recommendation quality. Our work is most closely related to recent studies on context-aware and hybrid recommendation systems that incorporate privacy-preserving mechanisms explicitly using *blocking* constraints (Pal et al., 2024).

B.3 Annotation in Niche Applications with limited expert annotators

A second application of our framework is identifying *hard datapoints* for annotation in expensive labeling tasks. Large Language Models (LLMs) offer strong zero-shot capabilities, making it easier to prototype solutions for downstream tasks. However, they struggle with complex domain-specific queries when relevant training data is scarce or evolving (Farr et al., 2024). In-context learning with few shot examples has emerged as a powerful approach, where a small set of high-quality examples improves model performance (Dong et al., 2022). Crucially, hard examples provide better domain-specific information (Baek et al., 2024; Liu et al., 2024; Mavromatis et al., 2023), but identifying them is challenging. Heuristic-based selection often leads to noisy, mislabeled, or outlier examples (Mindermann et al., 2022). Alternatively, we can leverage domain experts to assign a hardness score while annotating. This data-poor problem can be framed in a bandit framework, where unlabeled datapoints act as arms and are sequentially annotated while hardness scores are modeled as a sparse linear function of embeddings. In domains with very few annotators—sometimes only one—it is impractical to re-query the same datapoint, naturally leading to a *blocking constraint*. Beyond annotation, high-quality hard examples are also valuable in model training, where they improve generalization and efficiency (Sorscher et al., 2022; Maharana et al., 2023).

(Sorscher et al., 2022) demonstrated that selecting high-quality data can reduce the power-law association of test error with dataset size to an exponential law. In annotation-expensive tasks with large volumes of unlabeled data, the challenge is to select a representative subset of datapoints for labeling. Conceptually our work is similar to *active learning* (AL) (Settles, 2009; Lesci and Vlachos, 2024) where unlabeled samples are annotated adaptively, based on the confidence of a trained model (Coleman et al., 2022). Active learning works well with good initialization and informative confidence intervals. However, in our label-scarce setting, AL is particularly challenging with complex data due to the absence of a reliably trained model in the first place - this is more pronounced for difficult datapoints for which prediction is hard. AL needs an initial set of high-quality labeled samples to reasonably train a model - also known as the *cold-start* problem - when labels are scarce, uncertainty based sampling techniques are unsuitable (Li et al., 2024). Our goal is to identify informative samples with the help of the expert annotator(s), whom the final model aims to emulate. *Coreset selection* (Guo et al., 2022; Albalak et al., 2024; Sener and Savarese, 2018) aims to select a subset of datapoints for training. However, coreset selection assumes that a large amount of *labeled data* already exists, and the focus is on reducing computational costs. In contrast, our setting deals with the lack of labeled data, making existing coreset selection approaches, which rely on the entire labeled dataset, inapplicable. Our work also aligns with *curriculum learning* (Bengio et al., 2009), where a model is trained on samples of increasing difficulty/complexity. Due to the ambiguity in hardness definition, often heuristics are used to infer the difficulty of samples (Soviany et al., 2022) and can turn out unreliable and not generalizable. For niche tasks where an expert annotator is available, the difficulty ratings from the annotator are more informative than heuristics since the goal is to train a model to mimic the expert (Ionescu et al., 2016). In computer vision, there has been recent work regarding estimating the difficulty of a dataset for the model using implicit difficulty ratings of annotation. For NLP tasks, (Ethayarajh et al., 2022) constructs information theoretic metrics to estimate the difficulty of datapoints.

B.4 Applications where the blocking constraint is inevitable

In many personalized-recommendation and annotation settings, such as movie or book recommendations on edge devices users literally cannot consume the same item twice in immediate succession. On-device privacy-aware recommenders have no channel to query collaborative signals, so an immediate repeat shows the same item to the same user which a useless round both for learning and engagement. Likewise, sending an already-labelled image back to an expert consumes annotation budget without new information. We therefore treat blocking as an environmental constraint rather than a modelling choice: the learner never observes any reward for a second pull, so replaying cannot improve regret under the metric we adopt. Blocking constraints has been studied in different recommendation settings as discussed in the paper [1,2] and we treat blocking as a fundamental modeling

assumption rather than an optional restriction.

Under this modelling assumption, naively extending standard sparse bandit algorithms to a blocked setting fails as illustrated in Figure 1. As shown, rejection-sampling adaptations of ESTC, LinUCB, and DR-Lasso break their statistical guarantees. Once the high-mass arms are exhausted, the algorithm is forced to sample low-value arms, and regret explodes. In standard linear-bandit regret, the benchmark repeatedly selects the single best arm. Under blocking this benchmark is infeasible, so we compare against the top-T distinct arms (Eq. 1).

Moreover, under the blocking constraint, the traditional linear bandit regret objective allows pathological behavior, such as recommending a single high-reward movie repeatedly. Our new regret definition measures performance against the top T unique arms under blocking, incorporating realistic one shot constraints. Standard algorithms are not designed to optimize this metric, whereas our BSLB and C-BSLB algorithms are explicitly crafted to balance exploration, sparsity, and blocking, yielding provably optimal performance in data-poor, high-dimensional regimes.

Another application is materials discovery exemplifies the blocked regime: synthesizing or testing a candidate compound is costly, often destructive, and rarely repeated. The goal is therefore to select a small, diverse panel of candidates whose features well condition the linear model so that one-shot measurements yield an accurate parameter estimate and strong downstream recommendations. We include a synthetic vignette mirroring this setting—single-pull, high-dimensional features, sparse ground truth—to illustrate how our design improves conditioning and accelerates identification.

B.5 Discussion on $\lambda^*_{\min}$

On the role of $\lambda^\star_{\min}$. We denote by $\lambda^\star_{\min}$ the minimum eigenvalue of the empirical covariance of the subset of arms used for exploration; it coincides with the Restricted Eigenvalue (RE) constant that governs the statistical error of the Lasso in our analysis. In the blocked, high-dimensional, data-poor regime, both our upper and lower bounds necessarily depend on $\lambda^\star_{\min}$: the regret of BSLB scales as $\tilde{O}\big(k^{2/3}T^{2/3}(\lambda^\star_{\min})^{-2/3}\big)$ while the information-theoretic lower bound scales as $\Omega\big(k^{1/3}T^{2/3}(\lambda^\star_{\min})^{-1/3}\big)$. This shows the $\lambda^\star_{\min}$ dependence is fundamental rather than an artifact of algorithm design and mirrors prior high-dimensional sparse bandit results without blocking.

Algorithmic implications. Because regret degrades when $\lambda^\star_{\min}$ is small, our exploration subset is chosen by maximizing a convex surrogate for the E-optimal design objective, followed by randomized rounding; this yields a set G with $\lambda_{\min}(G) \geq \frac{1}{4}\,\lambda^\star_{\min}$ w.h.p., ensuring a well-conditioned design and enabling the stated regret rate under blocking. Under standard ℓ_∞-bounded arm assumptions, well-spread arm families (e.g., Hadamard-type constructions) give $\lambda^\star_{\min}$ bounded away from zero, illustrating regimes where nonvacuous guarantees are attainable.

Estimation and practice. In practice, $\lambda^\star_{\min}$ is unknown; we adopt a simple offline proxy: draw small random arm subsets and compute their empirical covariance eigenvalues to obtain a crude lower bound used for tuning. Empirically, even coarse estimates suffice to realize sublinear regret, while recognizing that extremely ill-conditioned arm sets can render bounds vacuous—an unavoidable limitation in this regime. We emphasize that the dependence on $\lambda^\star_{\min}$ arises from the underlying RE-controlled Lasso error and therefore persists across algorithmic choices.

B.6 Fair Comparator

Because the learner is prohibited from re-pulling an arm, a principled benchmark must obey the same constraint. We therefore define regret against the best set of T distinct arms, i.e., the set whose cumulative expected reward under the linear model,is maximal among all size-T subsets. Comparing to a policy that can re-pull arms would grant the comparator strictly more information than the learner can ever obtain under blocking, making the metric both unfair and misleading. This set based benchmark aligns the information structure of the oracle with that of the learner.

However if one wants, one can construct pseudo-instances to compare the two regimes. However, this is only useful for analytical purposes since this does not actually resemble the blocking constraint in practice. Our setting allows for multi-sets (we make use of this in our proof of the lower bound in Section C.1) and one can construct an arm set with independent copies of the best arm. Let $\mathcal{A}'$ be the extended set of arms with T independent copies of the best arm. Then over this instance the two comparators have the same expression. Therefore although

the setting no longer practically respects the blocking constraint, the regret expressions and the resulting regret bounds are the same.

B.7 Standalone contribution: offline Lasso under RE with blocked sampling

Beyond our bandit application, we establish high-probability prediction and support-recovery guarantees for the Lasso when data are gathered without replacement under a restricted-eigenvalue condition and soft sparsity. These results bridge classical RE-based analyses—usually stated for i.i.d. samples and practical design-of-experiments regimes where points are chosen adaptively and never revisited. We believe this transfer is of independent interest for practitioners designing one-shot data-collection campaigns.

B.8 Standalone Contribution

We propose a simple design procedure that targets the minimum eigenvalue (an E-optimal surrogate) by solving a tractable concave relaxation and then applying independent randomized rounding. Unlike more common A/D-optimal objectives, which control average or determinant-based criteria, our approach directly improves the worst-direction conditioning relevant for sparse recovery and regret. The method is easy to implement, parallelizable, and comes with concentration-based performance guarantees.

C TECHNICAL PROOFS

C.1 Proof of Lower Bound for Regret (Theorem 4)

Proof. We consider the hard-sparsity instance of the high-dimensional linear bandit setting with the *blocking* constraint. We prove this in three steps. First we show how one can construct an equivalent bandit problem with a *blocking* constraint for any problem without the *blocking* constraint. Next we use results from Hao et al. (2020) to show that the result holds true for this transformation. Finally, we show that this is the best

1. For any instance of the bandit problem without the *blocking* constraint, we can construct a bandit problem with the *blocking* constraint. This can be done by considering T copies of each of the arms, i.e. for the original arm set $\mathcal{A} = \{a^{(1)}, \ldots, a^{(\mathsf{M})}\}$ we construct the arm multiset $\mathcal{A}'$ as follows,

$$\mathcal{A}' = \cup_{i=1}^{\mathsf{M}} \cup_{j}^{\mathsf{T}} \{a_j^{(i)}\}$$

   where $a_j^{(i)}$ denotes the j^{th} copy of the i^{th} arm (such that it is a different arm with the same arm vector). Since it is a multiset, the union operation double-counts duplicate values. Now the bandit setting with blocking constraint with arm set $\mathcal{A}'$ is identical to the bandit setting without blocking constraint with arm set $\mathcal{A}$. Further the regret decomposition becomes identical in the first term, i.e. the top T arms have the same arm vector.

2. Now for any arm set $\mathcal{A}$ without the blocking constraint, the lower bound from Theorem 3.3 of Hao et al. (2020) holds for any algorithm, and the following bound holds

$$\mathbb{E}[\mathsf{R}] = \Omega(\min(\lambda_{\min}^{-1/3}(\mathcal{A}) k^{1/3} \mathsf{T}^{2/3}), \sqrt{d\mathsf{T}})).$$

3. Now if any algorithm operating with the blocking constraint could achieve the regret of order lesser than $\min(k^{1/3}\mathsf{T}^{2/3}), \sqrt{d\mathsf{T}})$, then the algorithm would solve the bandit problem with arm multiset $\mathcal{A}'$ (with the blocking constraint) with regret lower than $\min(k^{1/3}\mathsf{T}^{2/3}), \sqrt{d\mathsf{T}})$. But then we can solve the original problem with arm set $\mathcal{A}$ with the same regret, hence arriving at a contradiction.

In the data-poor regime $d \geq k^{1/3}\mathsf{T}^{2/3}$, which is the regime considered in the paper, this bound reduces to $\Omega(\lambda_{\min}^{-1/3}(\mathcal{A}) k^{1/3}\mathsf{T}^{2/3})$, which is the order that our upper bound achieves.

□

C.2 Proof of Offline Guarantees of Theorem 1

Preliminaries and Basis Pursuit Program

We need the following definition of a restricted eigenvalue of a matrix $\mathbf{X}$,

Definition 2. ***(Restricted Eigenvalue)*** *If $\mathbf{X}$ satisfies Restricted Eigenvalue property $\mathsf{RE}(k_0, \gamma, \mathbf{X})$, then there exists a constant $K(k_0, \gamma, \mathbf{X})$ such that for all $z \in \mathbb{R}^d$ and $z \neq \mathbf{0}$,*

$$K(k_0, \gamma, \mathbf{X}) = \min_{J \subseteq \{1,\dots,d\} | J| \leq k_0} \min_{\|z_{J^c}\|_1 \leq \gamma \|z_J\|_1} \frac{\|\mathbf{X}z\|_2}{\|z_J\|_2}.$$

This definition implies,

$$\begin{aligned} &K(k_0, \gamma, \mathbf{X})\|z_J\|_2 \leq \|\mathbf{X}z\|_2 \ \forall z, \\ &\text{such that} \quad \|z_{J^c}\|_1 \leq \gamma \|z_J\|_1 \ \forall J \subseteq \{1, \dots, d\}, |J| \leq k_0. \end{aligned}$$

We now prove the theorem.

Proof. Let $\mathbf{h} = \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}$ denote the error in estimation of the parameter, $\mathbf{w} = \mathbf{r} - \mathbf{X}\boldsymbol{\theta}$ denote the noise vector, λ is the Lagrange parameter in Line 9 of Algorithm 1 and n the number of samples.

For the vector $\mathbf{h}$, let $\mathcal{Y}_0$ denotes the top-k coordinates by absolute value , $\mathcal{Y}_1$ the next top k coordinates and so on . Let $\mathbf{v}_{\mathcal{Y}_i}$ denote the vector where the $\mathcal{Y}_i$ coordinates are equal to the coordinates of $\mathbf{v}$ and all the other coordinates are 0. Let $\alpha = \|\mathbf{h}_{(\mathcal{Y}_0 \cup \mathcal{Y}_1)}\|_2$.

The following inequality holds for a Lagrangian Lasso program (w/o any assumptions on sparsity) (See Eq.7.29 Proof of Theorem 7.13 Part (a) of Wainwright (2019); note this holds true without any assumptions on the sparsity or the design matrix but just by optimality argument of the Lasso program),

$$\frac{1}{n}\|\mathbf{X}\mathbf{h}\|_2^2 \leq \frac{\mathbf{w}^\mathsf{T}\mathbf{X}\mathbf{h}}{n} + \lambda(\|\boldsymbol{\theta}\|_1 - \|\widehat{\boldsymbol{\theta}}\|_1) \tag{6}$$

Now by triangle inequality $\|\boldsymbol{\theta}\|_1 - \|\widehat{\boldsymbol{\theta}}\|_1 \leq \|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\|_1 = \|\mathbf{h}\|_1$ and therefore,

$$\frac{1}{n}\|\mathbf{X}\mathbf{h}\|_2^2 \leq \frac{\mathbf{w}^\mathsf{T}\mathbf{X}\mathbf{h}}{n} + \lambda\|\mathbf{h}\|_1 \leq \left(\|\frac{\mathbf{X}^\mathsf{T}\mathbf{w}}{n}\|_\infty + \lambda\right)\|\mathbf{h}\|_1$$

(The last step follows as a result of using Hölder's inequality). Let $B = \left(\|\frac{\mathbf{X}^\mathsf{T}\mathbf{w}}{n}\|_\infty + \lambda\right)$, which gives us,

$$\frac{1}{n}\|\mathbf{X}\mathbf{h}\|_2^2 \leq B\|\mathbf{h}\|_1 \tag{7}$$

We lower bound $\frac{1}{n}\|\mathbf{X}\mathbf{h}\|_2^2$ using the RE condition and in order to do that we need to show that $\mathbf{h}$ lies in the cone $\{z | z \in \mathbb{R}^d, \|z_{\mathcal{Y}}^0\| \leq 4(1+\beta_k)\|z_{\mathcal{Y}_0^C}\|\}$, which is the subset of the cone the RE condition is satisfied on. Note that for the vector $\mathbf{h}$ the following condition holds,

$$\begin{aligned} \|\mathbf{h}_{\mathcal{Y}_0{}^C}\|_1 &= \|\mathbf{h}_{\mathcal{Y}_1}\|_1 + \|\mathbf{h}_{(\mathcal{Y}_0 \cup \mathcal{Y}_1)^C}\|_1 \\ &\overset{(a)}{\leq} \|\mathbf{h}_{\mathcal{Y}_1}\|_1 + \|\mathbf{h}_{(\mathcal{Y}_0 \cup \mathcal{Y}_1)}\|_1 + 2\beta_k \\ &\overset{(b)}{\leq} \|\mathbf{h}_{\mathcal{Y}_1}\|_1 + \|\mathbf{h}_{(\mathcal{Y}_0 \cup \mathcal{Y}_1)}\|_1 + 2\beta_k\|\boldsymbol{\theta}^*_{\mathcal{Y}_0}\|_1 \\ &\overset{(c)}{\leq} 2\|\mathbf{h}_{\mathcal{Y}_0}\|_1 + 2\|\mathbf{h}_{\mathcal{Y}_0}\|_1 + 4\beta_k\|\mathbf{h}_{\mathcal{Y}_0}\|_1 \\ &\leq 4(1+\gamma)\|\mathbf{h}_{\mathcal{Y}_0}\|_1, \end{aligned} \tag{8}$$

where (a) follows from the decomposition available in Theorem 1.6 in Boche et al. (2015),

$$\|\mathbf{h}_{(\mathcal{Y}_0 \cup \mathcal{Y}_1)^c}\|_1 \leq \|\mathbf{h}_{(\mathcal{Y}_0)^c}\|_1 \leq \|\mathbf{h}_{(\mathcal{Y}_0)}\|_1 + 2\beta_k \leq \|\mathbf{h}_{(\mathcal{Y}_0 \cup \mathcal{Y}_1)}\|_1 + 2\beta_k$$

(b) follows from the definition of the tail $\beta_k = \beta_k \|\boldsymbol{\theta}^*_{\mathcal{Y}_0}\|_1$

and (c) follows from $\|\boldsymbol{\theta}^*_{\mathcal{Y}_0}\|_1 \leq \|\boldsymbol{\theta}^*_{\mathcal{Y}_0} - \widehat{\boldsymbol{\theta}}_{\mathcal{Y}_0} + \widehat{\boldsymbol{\theta}}_{\mathcal{Y}_0}\|_1 \leq \|\boldsymbol{\theta}^*_{\mathcal{Y}_0} - \widehat{\boldsymbol{\theta}}_{\mathcal{Y}_0}\| + \|\widehat{\boldsymbol{\theta}}_{\mathcal{Y}_0}\|_1 \leq \|\mathbf{h}_{\mathcal{Y}_0}\| + \|\widehat{\boldsymbol{\theta}}_{\mathcal{Y}_0}\|_1 \leq 2\|\mathbf{h}_{\mathcal{Y}_0}\|_1$. The last inequality holds since $\widehat{\boldsymbol{\theta}}$ is a solution of equation 6 and $\|\widehat{\boldsymbol{\theta}}_{\mathcal{Y}_0}\|_1 \leq \|\widehat{\boldsymbol{\theta}}\|_1 \leq \|\mathbf{h}_{\mathcal{Y}_0} + \mathbf{0}_{\mathcal{Y}_0^c}\|_1 = \|\mathbf{h}_{\mathcal{Y}_0}\|_1$ otherwise $\mathbf{h}_{\mathcal{Y}_0} + \mathbf{0}_{\mathcal{Y}_0^c}$ would be the solution instead.

Now we can use the RE condition on the design matrix $\mathbf{X}$ using the vector $\mathbf{h}$, which is such that $\|\mathbf{h}_{\mathcal{Y}_0^c}\| \leq (4 + 4\beta_k)\|\mathbf{h}_{(\mathcal{Y}_0)}\|$ the following holds,

$$K^2(k_0, 4 + 4\beta_k, \frac{\mathbf{X}}{\sqrt{n}})\|\mathbf{h}_{\mathcal{Y}_0}\|_2^2 \leq \frac{\|\mathbf{X}\mathbf{h}\|_2^2}{n}. \tag{9}$$

Putting this into equation 7, and letting $K = K(k_0, 4 + 4\beta_k, \frac{\mathbf{X}}{\sqrt{n}})$,

$$K^2\|\mathbf{h}_{\mathcal{Y}_0}\|_2^2 \leq B\|\mathbf{h}\|_1 = B(\|\mathbf{h}_{(\mathcal{Y}_0)}\|_1 + \|\mathbf{h}_{(\mathcal{Y}_0)^C}\|_1) \overset{(a)}{\leq} B(5 + 4\beta_k)\|\mathbf{h}_{(\mathcal{Y}_0)}\|_1$$

where (a) follows from equation 8. Therefore,

$$\frac{K^2}{k}\|\mathbf{h}_{\mathcal{Y}_0}\|_1^2 \overset{(a)}{\leq} K^2\|\mathbf{h}_{\mathcal{Y}_0}\|_2^2 \leq B(5 + 4\beta_k)\|\mathbf{h}_{(\mathcal{Y}_0)}\|_1,$$

where (a) is due to Cauchy Schwarz. So we have,

$$\|\mathbf{h}_{\mathcal{Y}_0}\|_1 \leq \frac{Bk(5 + 4\beta_k)}{K^2}$$

which we combine with $\|\mathbf{h}\|_1 \leq (5 + 4\beta_k)\|\mathbf{h}_{(\mathcal{Y}_0)}\|_1$ to obtain,

$$\|\mathbf{h}\|_1 \leq \frac{Bk(5 + 4\beta_k)^2}{K^2} \tag{10}$$

Now we bound $\|\frac{\mathbf{X}^\mathsf{T}\mathbf{w}}{n}\|_\infty$ using the properties of a bounded zero-mean noise from Wainwright (2019). Consider the random variables, $\frac{(\mathbf{X}^\mathsf{T}\mathbf{w})_j}{n}$ each of which is a weighted sum of independent random variables. Each of the term of the sum is subgaussian with parameter bounded by $\frac{\sigma M}{n}$, where M is the ℓ_∞ bound. We have $M = 1$. Using a standard sub-Gaussian concentration bound $\mathbb{P}(\frac{(\mathbf{X}^\mathsf{T}\mathbf{w})_j}{n} \geq t) \leq 2\exp(-\frac{nt^2}{2\sigma^2})$. In Set $t = 2\zeta_2\sigma(\sqrt{\frac{\log d}{n}})$, then $\mathbb{P}(\frac{(\mathbf{X}^\mathsf{T}\mathbf{w})_j}{n} \geq 2\zeta_2\sigma(\sqrt{\frac{\log d}{n}}) \leq 2\exp(-2\zeta_2^2 \log d) = 2d^{-2\zeta_2^2}$. Using a union bound on d on the probability $\mathbb{P}(\|\frac{\mathbf{X}^\mathsf{T}\mathbf{w}}{n}\|_\infty \geq \zeta_2(\sqrt{\frac{\log d}{n}})) = \mathbb{P}(\max_{j\in[d]} \frac{(\mathbf{X}^\mathsf{T}\mathbf{w})_j}{n} \geq 2\zeta_2\sqrt{\frac{\log d}{n}}) \leq \sum_{j\in[d]} \mathbb{P}(\frac{(\mathbf{X}^\mathsf{T}\mathbf{w})_j}{n} \geq 2\zeta_2\sqrt{\frac{\log d}{n}}) \leq d^{1-2\zeta_2^2}$. Therefore for a large enough constant ζ_2, $B \leq 2\zeta_2\sigma(\sqrt{\frac{2\log d}{n}})$ with high probability $1 - d^{1-2\zeta_2^2}$.

Combining this with equation 10 and putting $\lambda = \sqrt{\frac{\log d}{n}}$, we have,

$$\|\mathbf{h}\|_1 \leq (2\sqrt{2}\zeta_2\sigma + 1)\frac{(5 + 4\beta_k)^2 k}{K^2}\sqrt{\frac{\log d}{n}}$$

with probability $1 - d^{1-2\zeta_2^2}$.

□

C.3 Proof of Corollary 1

The statement of the Corollary basically extends the estimation guarantees of Theorem 1 hold with high probability.

First let us denote the expected covariance matrix as (assume $\hat{\mathcal{G}}$ is the sampled subset without replacement from $\mathcal{A}$),

$$\Sigma = \mathbb{E}\left\{\frac{1}{|\hat{\mathcal{G}}|}\sum_{\mathbf{a}\in\hat{\mathcal{G}}} \mathbf{a}^\mathsf{T}\mathbf{a}\right\} = \frac{1}{|\hat{\mathcal{G}}|}\sum_{\mathbf{a}\in\hat{\mathcal{G}}} \mathbb{E}\{\mathbf{a}^\mathsf{T}\mathbf{a}\} \overset{(a)}{=} \frac{1}{|\mathcal{G}|}\sum_{\mathbf{a}\in\mathcal{G}} \mathbf{a}^\mathsf{T}\mathbf{a},$$

where (a) follows from sampling without replacement. The proof then hinges on the following key fact and using Theorem 6 stated later.

Fact 1. The minimum eigenvalue defined by $\Lambda = \lambda_{\min}\left(\frac{1}{|\mathcal{G}|}\sum_{\mathbf{a}\in\mathcal{G}}\mathbf{a}^\mathsf{T}\mathbf{a}\right)$ is such that,

$$\Lambda = \lambda_{\min}\left(\frac{1}{|\mathcal{G}|}\sum_{\mathbf{a}\in\mathcal{G}}\mathbf{a}^\mathsf{T}\mathbf{a}\right) = \min_{z\in\mathbb{R}^d}\frac{1}{|\mathcal{G}|}\frac{\|\mathbf{X}z\|_2^2}{\|z\|_2^2} \leq K^2(k, 4(1+\beta_k), \frac{\mathbf{X}}{\sqrt{|\mathcal{G}|}}) \tag{11}$$

We further show using results from Rudelson and Zhou (2013) (Proved in Theorem 6 below.) that the following holds with probability $1-\exp(-\zeta_5 n)$,

$$\frac{1}{2}K(k, 4(1+\beta_k), \Sigma^{1/2}) \leq K(k, 4(1+\beta_k), \hat{\Sigma}^{1/2}) \tag{12}$$

Therefore the following holds,

$$\frac{1}{2}K(k, 4(1+\beta_k), \frac{\mathbf{X}}{\sqrt{n}}) \geq \Lambda^{1/2}$$

with probability $1-\exp(-\zeta_5 n)$.

Therefore,

$$\|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\|_1 \leq 2\zeta_2\sigma\frac{(5+4\beta_k)^2 k}{K^2}\sqrt{\frac{2\log d}{n}} = \mathsf{O}\left(\frac{(5+4\beta_k)^2\sigma k}{\Lambda}\sqrt{\frac{2\log d}{n}}\right),$$

with probability $1 - d^{1-2\zeta_2^2} - \exp(-\zeta_5 n)$.

The only additional ingredient needed to prove the corollary is the following concentration inequality of equation 12, which shows that if the RE of the covariance matrix of the set which is used for sampling is bounded from below, then so is the RE of the sampled covariance matrix with high probability. However although the following theorem is adapted directly from Rudelson and Zhou (2013); to prove it we need to make changes accordingly for the sampling without replacement case. This introduces an additional multiplicative $\lambda_{\min}$ term in the exponential in the probability but does not change the order with respect to any other variable.

Concentration Inequality for RE condition to hold on sample covariance matrix given that it holds on the expected covariance matrix.

We use the following Theorem which is an extension of Theorem 8 from Rudelson and Zhou (2013).

Theorem 6. *Let $0 < \delta < 1$ and $0 < k_0 < d$. Let $Y \in \mathbb{R}^d$ be a random vector such that $\|Y\|_\infty \leq M$ a.s and denote $\Sigma = \mathbb{E}YY^T$. Let $\mathbf{X}$ be an $n \times d$ matrix, whose rows $Y_1, \ldots, Y_n$ are sampled without replacement from a set $\mathcal{G}$. Let Σ satisfy the $\mathsf{RE}(k_0, 3\gamma, \Sigma^{1/2})$ condition as in Definition 1. Set $k_1^\star = k_0 + k_0 \max_j \|\Sigma^{1/2}e_j\|_2^2 \times \left(\frac{16\mathsf{RE}(k_0,3\gamma,\Sigma^{1/2})^2(3\gamma)^2(3\gamma+1)}{\delta^2}\right)$. Assume that $k_1^\star \leq d$ and $\rho = \rho_{\min}(k_1^\star, \Sigma^{1/2}) > 0$, the ($k_1^\star-$sparse) minimum eigenvalue of $\Sigma^{1/2}$. Suppose the sample size satisfies for some absolute constant C,*

$$n \geq \frac{CM^2 k_1^\star \cdot \log d}{\rho\delta^2} \cdot \log^3\left(\frac{CM^2 k_1^\star \cdot \log d}{\rho\delta^2}\right).$$

Then, with probability at least $1-\exp(-\delta\rho^2 n/(6M^2 k_1^\star))$, $\mathsf{RE}(k_0^\star, \gamma, \mathbf{X})$ condition holds for matrix $\frac{1}{\sqrt{n}}\mathbf{X}$ with,

$$0 \leq (1-\delta)K(k_0, \gamma, \Sigma^{1/2}) \leq K\left(k_0, \gamma, \frac{1}{\sqrt{n}}\mathbf{X}\right).$$

(The inequality is reverse because our definition of Restricted Eigenvalue has the $1/K$ compared to the definition of Rudelson and Zhou (2013)). We use the Theorem with $\delta = \frac{1}{2}$

The proof of Theorem 6 is dependent on Theorem 23, which is reproduced below, and Theorem 10 (Reduction Principle), which is in the paper. Out of these two, only Theorem 23 has elements related to the randomness of

the design. In the proof of the above theorem, authors use concentration inequality to extend a RIP-like condition on a general cone (rather than sparse vectors). This concentration inequality results from the following theorem: an augmented version of Theorem 22 of Rudelson and Zhou (2013) to the sampling without replacement case.

The original proof of Theorem 22 is extremely involved (and mathematically rich). Reproducing the entire proof would have surmounted to reproducing the entire paper. We only highlight the key difference, it is recommended that the reader goes through the proof beforehand/side-by-side.

Theorem 7. *Set $1 > \delta > 0$, $0 < k_0^\star \leq d$ and $\Lambda_0 > 0$. Let $\mathcal{G}$ be a subset of vectors such that $||Y||_\infty \leq M$, with $\Sigma = \sum_{Y \in \mathcal{G}} \frac{1}{|\mathcal{G}|} YY^\mathsf{T}$. Σ satisfies $\mathsf{RE}(k_0^\star, 3\Lambda_0, \Sigma^{1/2})$. Rows of $\mathbf{X}$ are drawn uniformly without replacement from $\mathcal{G}$. Set $k_1^\star = k_0^\star + k_0^\star \max_j \|\Sigma^{1/2} e_j\|_2^2 \times \left(\frac{16\mathsf{RE}(k_0^\star, 3\Lambda_0, \Sigma^{1/2})^2 (3\Lambda_0)^2 (3\Lambda_0+1)}{\delta^2} \right)$. Assume $k_1^\star \leq d$ and $\lambda_{\min}(k_1^\star, \Sigma^{1/2}) > 0$. If for some absolute constant ζ_{12},*

$$n \geq \frac{\zeta_{12} M^2 k_1^\star \log d}{\lambda_{\min}(k_1^\star, \Sigma^{1/2}) \delta^2} \log^3 \left(\frac{\zeta_{12} M^2 k_1^\star \log d}{\lambda_{\min}(k_1^\star, \Sigma^{1/2}) \delta^2} \right)$$

then with probability $1 - \exp(\frac{-\delta \lambda_{\min}^2(k_1^\star, \Sigma^{1/2}) n}{6 M^2 k_1^\star})$, for all $\mathbf{v} \in \mathsf{C}(k_0^\star, \Lambda_0),\ \mathbf{v} \neq 0$

$$1 - \delta \leq \frac{1}{\sqrt{n}} \frac{\|\mathbf{X}\mathbf{v}\|_2}{\|\mathbf{v}\|_2} \leq 1 + \delta.$$

Proof. In the proof of Theorem 22 of Rudelson and Zhou (2013), two arguments require the sampling with replacement (i.i.d. samples), namely symmetrization and Talagrand's concentration inequality. We use a sampling with replacement version of the symmetrization argument and a sampling with replacement version of McDiarmid's concentration inequality to obtain comparable bounds. Therefore, to prove this argument, the following two lemmas.

Lemma 1. *(Symmetrization without Replacement)*

$$\mathbb{E} \sup_{x \in F} \left| \mathbb{E} f_j(x, Z_j) - \frac{1}{n} \sum f_j(x, Z_j) \right| \leq \frac{2}{n} \mathbb{E} \sup_{x \in F} \left| \sum \xi_j f_j(x, Z_j) \right|$$

where are i.i.d. Rademacher random variables and Z_j are random variables sampled uniformly without replacement from some set.

Proof. Let $Z_1, \ldots, Z_n$ be the random variables sampled uniformly without replacement from set $\mathcal{G}$. Z_1- Consider $Z_1^{'}, \ldots, Z_n^{'}$ be an independent sequence of random variables sampled uniformly without replacement from set $\mathcal{G}$. Then $\frac{1}{n}\sum f_j(x, Z_j) - \mathbb{E} f_j(x, Z_j)$ and $\frac{1}{n}\sum f_j(x, Z_j^{'}) - \mathbb{E} f_j(x, Z_j)$ are zero mean random variable. Then,

$$\mathbb{E}||\frac{1}{n}\sum f_j(x, Z_j) - \mathbb{E} f_j(x, Z_j)|| \leq \mathbb{E}||\frac{1}{n}\sum f_j(x, Z_j) - \mathbb{E} f_j(x, Z_j) - \frac{1}{n}\sum f_j(x, Z_j^{'}) + \mathbb{E} f_j(x, Z_j)||$$
$$\implies \mathbb{E}||\frac{1}{n}\sum f_j(x, Z_j) - \mathbb{E} f_j(x, Z_j)|| \leq \mathbb{E}||\frac{1}{n}\sum \left(f_j(x, Z_j) - f_j(x, Z_j^{'}) \right)||$$

(Since $\frac{1}{n}\sum f_j(x, Z_j) - \mathbb{E} f_j(x, Z_j)$ and $\frac{1}{n}\sum f_j(x, Z_j^{'}) - \mathbb{E} f_j(x, Z_j)$ are independent)

$$\mathbb{E}||\frac{1}{n}\sum f_j(x, Z_j) - \mathbb{E} f_j(x, Z_j)|| \leq \mathbb{E}||\frac{1}{n}\sum \left(f_j(x, Z_j) - f_j(x, Z_j^{'}) \right)||$$
$$\implies \mathbb{E}||\frac{1}{n}\sum \xi_j \left(f_j(x, Z_j) - f_j(x, Z_j^{'}) \right)|| \leq 2\mathbb{E}||\frac{1}{n}\sum \xi_j \left(f_j(x, Z_j) \right)||$$

(Symmetric random variables Vershynin (2018) since Z_j and Z_j' have same distribution; followed by Triangular Inequality). □

Lemma 2. *(Concentration using McDiarmid's inequality) If* $|f_j(x)| \leq \zeta_{13}$ *a.s.. And suppose* $W = \sup_{x\in F}\sum_{j=1}^{n} f_j(x, Z_j)$*, where* $Z_1, \ldots, Z_n$ *are sampled uniformly without replacement from some set. If* $\mathbb{E}W \leq 2\delta n$*, then the following holds,*

$$\mathbb{P}(W \geq 4\delta n) \leq \exp\left(\frac{-8\delta^2 n}{\zeta_{13}^2}\right)$$

Proof. We prove the result by using McDiarmid's inequality. First we bound the quantity,

$$\sup_{Z_i^{'}} |\sup_{x\in F}\sum_{j=1}^{n} f_j(x, Z_j) - \sup_{x\in F}\left(\sum_{j=1, j\neq i}^{n} f_j(x, Z_j) + f_i(x, Z_i^{'})\right)|$$

$$\leq \sup_{Z_i^{'}} |\sup_{x\in F}\sum_{j=1, j\neq i}^{n} f_j(x, Z_j) + \sup_{x\in F} f_j(x, Z_i) - \sup_{x\in F}\sum_{j=1, j\neq i}^{n} f_j(x, Z_j) - \inf_{x\in F} f_i(x, Z_i^{'})| \leq \sup_{Z_i^{'}} |\sup_{x\in F} f_j(x, Z_i)| \leq \zeta_{13}$$

We use the following version of McDiarmid's concentration inequality for random variable without replacement with $t = 2\delta n$ to obtain the result. The condition that needs to be verified is that W is symmetric under permutations of the individual $f_j(x, Z_j)$. This is obviously true since this is a unweighted sum of the individual $f_j(x, Z_j)$. We next state McDiarmid's concentration inequality without replacement from Tolstikhin (2017),

Lemma 3. *Suppose* $W = \sup_{x\in F}\sum_{j=1}^{n} f_j(x, Z_j)$*, where* $Z_1, \ldots, Z_n$ *are sampled uniformly without replacement from some set. Then,*

$$\mathbb{P}(W - \mathbb{E}W \geq t) \leq \exp\left(\frac{-2t^2}{n\zeta_{13}^2}\right).$$

The probability is $\exp\left(\frac{-8\delta^2 n}{\zeta_{13}^2}\right) \leq \exp\left(\frac{-\delta^2 n}{6\zeta_{13}^2}\right) \leq \exp\left(\frac{-\delta^2 n\lambda_{\min}^2(k_1^\star, \Sigma^{1/2})}{6M^4 m^2}\right) \leq \exp\left(\frac{-\delta^2 n\lambda_{\min}^2(k_1^\star, \Sigma^{1/2})}{6M^2 m}\right)$, which is same as that in the original theorem from Rudelson and Zhou (2013) except an additional $\lambda_{\min}(k_1^\star, \Sigma^{1/2})$ which is reflected in our Theorem statement. □

Comment on Dudley's inequality: Theorem 23 also uses Dudley's inequality, but there the $\Psi_1, \ldots, \Psi_n$ are treated as deterministic and so the proof goes through in our sampled without replacement case as well. □

We can now complete the proof by computing the probability of the following event,

$$\mathcal{E} = \left\{K(k, 4+4\beta_k, \frac{\mathbf{X}}{\sqrt{n}}) \geq K(k, 4+4\beta_k, \Sigma^{1/2}) \geq (\lambda_{\min}^*)^{-1/2}, \|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\|_1 = \widetilde{\mathsf{O}}\left(\sigma\frac{(5+4\beta_k)^2 k}{K(k, 4+4\beta_k, \frac{\mathbf{X}}{\sqrt{n}})}\sqrt{\frac{2\log d}{n}}\right)\right\}$$

has the probability, $\mathbb{P}(\mathcal{E}) \geq 1 - d^{1-2\zeta_2^2} - \exp(-\zeta_5 n)$. which completes the proof.

C.4 Proof of Approximation Guarantees of Theorem 2

For vectors $\mathbf{v}$ and $\mathbf{z}$, define $Z_{\mathbf{v}}(\mathbf{z}) = \mathbf{z}^\mathsf{T}\mathbf{v}\mathbf{v}^\mathsf{T}\mathbf{z}$.

Then the minimum eigenvalue for (covariance matrix of) a set of vectors $\mathcal{G}$ is given by, $\lambda_{\min}(\mathcal{G}) = \min_{\mathbf{z}\in\mathcal{B}^d}\frac{1}{|\mathcal{G}|}\sum_{\mathbf{v}\in\mathcal{G}} Z_{\mathbf{v}}(\mathbf{z})$.

Let randomized rounding be run with $\hat{u}$, and $\hat{\mathcal{G}}$ be the sampled set of arms. Of-course $|\hat{\mathcal{G}}|$ need not be equal to $\hat{u}$. However, we first assume that the denominator of $\lambda_{\min}(\hat{\mathcal{G}})$ is equal to $\hat{u}$. We later show that this assumption worsens the approximation guarantees by 2 with high probability. Under this assumption by construction of the randomized rounding procedure, the expected minimum eigenvalue of the sampled set is equal to the minimum eigenvalue corresponding to the solution of the convex optimization problem, since, $\mathbb{E}_{\boldsymbol{\mu}} Z_{\mathbf{v}}(\mathbf{z}) = \boldsymbol{\mu}_{\mathbf{v}} Z_{\mathbf{v}}(\mathbf{z})$.

Step 1: We therefore prove the following result to bound the approximation error between the $\lambda_{\min}(\hat{\mathcal{G}})$ obtained from the randomized rounding solution and the optimal solution (off by a factor of $\frac{\hat{u}}{|\hat{\mathcal{G}}|}$)of the convex relaxation from equation 5 run using $\hat{u}$.

Lemma 4. *Let $\mathcal{A}$ be a set of M arms where each arm is $\mathbf{a} \in \mathcal{B}_\infty^d$ and let $Z_{\mathbf{v}}(\mathbf{z}) = \mathbf{z}^\mathsf{T}\mathbf{v}\mathbf{v}^\mathsf{T}\mathbf{z}$. Let $\boldsymbol{\mu}$ be the solution of the convex relaxation of equation 5 at $\hat{u}$ and $\hat{\mathcal{G}}$ be the set sampled using randomized rounding (Step 18-20 in Alg. 1). Then for some constant ζ_{10} the following holds ,*

$$\mathbb{P}\left(\left|\inf_{z\in\mathcal{B}^d}\frac{1}{\hat{u}}\sum_{\mathbf{v}\in\hat{\mathcal{G}}}Z_{\mathbf{v}}(\mathbf{z}) - \inf_{z\in\mathcal{B}^d}\sum_{\mathbf{v}\in\mathcal{A}}\boldsymbol{\mu}_{\mathbf{v}}Z_{\mathbf{v}}(\mathbf{z})\right| \geq \frac{\zeta_{10}\sqrt{d}\log\mathsf{M}}{\sqrt{|\hat{u}|}}\right) \leq \frac{1}{\log\mathsf{M}} \tag{13}$$

Proof. By symmetrization over the sum of independent random variables $\{W_i, i \in [\mathsf{M}]\}$ each of which is $Z_{\mathbf{v}_i}(\mathbf{z})$ with probability $\hat{u}\boldsymbol{\mu}_{\mathbf{v}_i}$ and 0 otherwise. It can be seen that $\mathbb{E}\{\frac{1}{\hat{u}}\sum_{i\in[\mathsf{M}]}W_i\} = \sum_{\mathbf{v}\in\mathcal{A}}\boldsymbol{\mu}_{\mathbf{v}}Z_{\mathbf{v}}(\mathbf{z})$

$$\mathbb{E}\left[\sup_{z\in\mathcal{B}^d}\left|\frac{1}{\hat{u}}\sum_{i\in[\mathsf{M}]}W_i - \sum_{\mathbf{v}\in\mathcal{A}}\boldsymbol{\mu}_{\mathbf{v}}Z_{\mathbf{v}}(\mathbf{z})\right|\right] \leq 2\mathbb{E}\left[\sup_{z\in\mathcal{B}^d}\left|\frac{1}{\hat{u}}\sum_{i\in[\mathsf{M}]}\xi_i W_i\right|\right] = \frac{2}{\hat{u}}\mathbb{E}\left[\sup_{z\in\mathcal{B}^d}\left|\sum_{i\in[\mathsf{M}]}\xi_i W_i\right|\right] \tag{14}$$

where ξ_i are i.i.d. Rademacher random variables.

Now using Dudley's integral inequality on sum of independent RVs,

$$\mathbb{E}\left[\sup_{z\in\mathcal{B}^d}\left|\sum_i \xi_i W_i\right|\right] \leq \zeta_{11}\Psi\log^{1/2}(\frac{3}{\epsilon})\sqrt{d}$$

where Ψ is the constant which satisfies for all $\mathbf{z}_1$ and $\mathbf{z}_2$,

$$||\sum_{i\in[\mathsf{M}]}\xi_i W_i(\mathbf{z}_1) - \sum_{i\in[\mathsf{M}]}\xi_i W_i(\mathbf{z}_2)||_\psi \leq \Psi||z_1 - z_2||_2 \leq \sqrt{2}\Psi,$$

and where $\|\cdot\|$ is the sub Gaussian norm.

Now w.l.o.g.,

$$||\sum_{i\in[\mathsf{M}]}\xi_i W_i(\mathbf{z}_1) - \sum_{i\in[\mathsf{M}]}\xi_i W_i(\mathbf{z}_2)||_\psi \leq 2||\sum_{i\in[\mathsf{M}]}\xi_i W_i(\mathbf{z}_1)||_\psi \leq 2||\sum_i W_i(\mathbf{z}_1)||_\psi.$$

The last inequality follows since the $\xi_i W_i$ are bounded by W_i. Now from the definition of sub Gaussian norm,

$$||\sum_{i\in[\mathsf{M}]}W_i||_\psi = \inf\{t : \exp\frac{(\sum W_i)^2}{t^2} \leq 2\}$$

Now,

$$\exp\frac{\sum(W_i)^2}{t^2} \leq \exp\frac{4M^4 d}{\hat{u}t^2}\ (a)$$

(a) follows with high probability from the following argument (using Hoeffding's inequality on Bernoulli random variable $\mathbb{1}(\mathbf{v})$ (if $\mathbf{v}$ is present or not with probability $\hat{u}\boldsymbol{\mu}_{\mathbf{v}}$) with deviation equal to the mean),

$$\sum_{\mathbf{v}\in\hat{\mathcal{G}}}(W_i)^2 \leq \sum_{\mathbf{v}\in\mathcal{A}}2\mathbb{E}[\mathbb{1}(\mathbf{v})](Z_{\mathbf{v}}(\mathbf{z}_1))^2 \overset{(a)}{\leq} \sum_{\mathbf{v}\in\mathcal{A}}2\hat{u}\boldsymbol{\mu}_{\mathbf{v}}\frac{(\mathbf{z}_1^\mathsf{T}\mathbf{v}\mathbf{v}^\mathsf{T}\mathbf{z}_1)^2}{\hat{u}^2} \leq \frac{2M^4d^2}{\hat{u}}\sum\boldsymbol{\mu}_{\mathbf{v}} \leq \frac{2M^4d^2}{\hat{u}}, \tag{15}$$

with probability $1 - \exp(-2\mathsf{M})$. (a) follows from the fact that $\mathbf{a} \in \mathcal{B}_\infty^d$, $\|\mathbf{a}\|_2 \leq d$.

To find $\inf\{t : \exp(\frac{\cdot}{t^2}) \leq 2\}$,

$$\exp\frac{4M^4d^2}{\hat{u}t^2} \leq 2 \implies \sqrt{\frac{4M^4d^2}{\hat{u}\ln 2}} \leq t$$

Therefore $\Psi \leq \sqrt{\frac{4M^4d^2}{\hat{u}\ln 2}}$ and letting $\zeta_{10} = 4\zeta_{11}\sqrt{\frac{\log(\frac{3}{\epsilon})}{\ln(2)}}$, plugging this into equation 14 we obtain,

$$\mathbb{E}\left[\sup_{z\in\mathcal{B}^d}\left|\frac{1}{\hat{u}}\sum_{i\in[\mathsf{M}]}W_i - \sum_{\mathbf{v}\in\mathcal{A}}\boldsymbol{\mu}_{\mathbf{v}}Z_{\mathbf{v}}(\mathbf{z})\right|\right] \leq 2\zeta_{10}\sqrt{\frac{d^3M^4}{\hat{u}^3}}$$

Using Markov's inequality along with the union bound on the the probability of equation 15,

$$\mathbb{P}(\sup_{z\in\mathcal{B}^d}\left|\frac{1}{\hat{u}}\sum_{i\in[\mathsf{M}]}W_i-\sum_{\mathbf{v}\in\mathcal{A}}\boldsymbol{\mu}_{\mathbf{v}}Z_{\mathbf{v}}(\mathbf{z})\right|\geq 2\zeta_{10}\frac{\sqrt{d}t}{\sqrt{\hat{u}^3}})\leq\frac{1}{t}+\exp(-2\mathsf{M})$$

One can can set t appropriately.

Also finally note that showing the bound on the supremum implies the bound we set to show in equation (11). Note that $\frac{1}{\hat{u}}\sum_{i\in[\mathsf{M}]}W_i=\frac{1}{\hat{u}}\sum_{\mathbf{v}\in\hat{\mathcal{G}}}Z_{\mathbf{v}}$ Now if,

$$\sup_{z\in\mathcal{B}^d}\left|\frac{1}{\hat{u}}\sum_{\mathbf{v}\in\hat{\mathcal{G}}}Z_{\mathbf{v}}-\sum_{\mathbf{v}\in\mathcal{A}}\boldsymbol{\mu}_{\mathbf{v}}Z_{\mathbf{v}}(\mathbf{z})\right|\leq\epsilon$$

is true with probability $1-\delta$. Then $\forall z\in\mathcal{B}^d$,

$$\sum_{\mathbf{v}\in\mathcal{A}}\boldsymbol{\mu}_{\mathbf{v}}Z_{\mathbf{v}}(\mathbf{z})-\epsilon\leq\sum_{\mathbf{v}\in\hat{\mathcal{G}}}\frac{1}{\hat{u}}Z_{\mathbf{v}}(\mathbf{z})$$
$$\inf_{z\in\mathcal{B}^d}\mathbb{E}\sum_{\mathbf{v}\in\mathcal{A}}\boldsymbol{\mu}_{\mathbf{v}}Z_{\mathbf{v}}(\mathbf{z})-\epsilon\leq\inf_{z\in\mathcal{B}^d}\frac{1}{\hat{u}}\sum_{\mathbf{v}\in\hat{\mathcal{G}}}Z_{\mathbf{v}}(\mathbf{z})$$

with probability $1-\delta$. Similar to the other direction, we obtain the desired result. □

Step 2: Then, we bound the size of the actual sampled set with respect to $\hat{u}$, at which the convex relaxation is computed. We derive the following result which gives a probability bound on the number of arms sampled.

Lemma 5. *Let $\hat{u}$ be the subset size that the randomized rounding is run with (Line 20 in Alg. 1) and let $\hat{\mathcal{G}}$ be the true number of sampled arms. Then the following probability holds,*

$$\mathbb{P}(\frac{\hat{u}}{2}\leq|\hat{\mathcal{G}}|\leq 2\hat{u})\geq 1-2(\frac{2}{e})^{\frac{\hat{u}}{2}}\tag{16}$$

Proof. We prove the following two tail bounds and then take the union bound over them both,

$$\mathbb{P}(|\hat{\mathcal{G}}|\geq 2\hat{u})\leq(\frac{e}{3})^{\hat{u}},\mathbb{P}(|\hat{\mathcal{G}}|\leq\hat{u}/2)\leq(\frac{2}{e})^{\frac{\hat{u}}{2}}$$

First the size of the sampled subset is the sum of independent Bernoulli random variables, $|\hat{\mathcal{G}}|=\sum p_j$ where each $p_j=Ber(\boldsymbol{\mu}_j\hat{u})$. Using tail bound from Chernoff bound,

$$\mathbb{P}(|\hat{\mathcal{G}}|\geq 2\hat{u})\leq\inf_{t>0}\exp(-t2\hat{u})\mathbb{E}[\exp(t|\hat{\mathcal{G}}|)]=\inf_{t}\exp(-t2\hat{u})\prod_j\mathbb{E}[\exp(tp_j)]\text{ (independent rv)}$$
$$\inf_t\exp(-t2\hat{u})\prod_j\mathbb{E}[\exp(tp_j)]\leq\inf_t\exp(-t2\hat{u})\prod_j\exp(\hat{u}\boldsymbol{\mu}_j(\exp(t)-1))$$
$$=\inf_t\exp(-t2\hat{u})\exp(\hat{u}(\exp(t)-1))$$

Achieves infinum for $t=\ln 2$,

$$\mathbb{P}(|\hat{\mathcal{G}}|\geq 2\hat{u})\leq\exp(-2\hat{u}\ln 2+\hat{u})=(\frac{3}{e})^{\hat{u}}$$

Using a similar left tail bound,

$$\begin{aligned}\mathbb{P}(|\hat{\mathcal{G}}| \leq \hat{u}/2) &\leq \inf_{t<0} \exp(-t\hat{u}/2)\mathbb{E}[\exp(t|\hat{\mathcal{G}}|)] \\ &= \inf_{t} \exp(-t\hat{u}/2) \prod_j \mathbb{E}[\exp(tp_j)] \text{ (independent rv)} \\ &= \inf_{t} \exp(-t\hat{u}/2) \prod_j \mathbb{E}[\exp(tp_j)] \\ &\leq \inf_{t} \exp(-t\hat{u}/2) \prod_j \exp(\hat{u}\boldsymbol{\mu}_j(\exp(t)-1)) = \inf_{t} \exp(-t\hat{u}/2)\exp(\hat{u}(\exp(t)-1))\end{aligned}$$

Achieves infinum for $t = -\ln 2$,

$$\mathbb{P}(|\hat{\mathcal{G}}| \leq \hat{u}/2) \leq \exp(\hat{u}(-\frac{1}{2}) + \ln 2\hat{u}/2) = (\frac{2}{e})^{\frac{\hat{u}}{2}}$$

Now for $\hat{u} \geq 1$, $(\frac{2}{e})^{\frac{\hat{u}}{2}} \geq (\frac{3}{e})^{\hat{u}}$, and therefore applying the union bound we obtain the required result. □

Therefore the above lemma helps us prove the following statement,

$$\mathbb{P}(\frac{1}{2} \leq \frac{\min_{z\in\mathcal{B}^d} \sum_{\mathbf{v}} \frac{1}{|\hat{\mathcal{G}}|} p_{\mathbf{v}} z^{\mathsf{T}}\mathbf{v}\mathbf{v}^{\mathsf{T}} z}{\min_{z\in\mathcal{B}^d} \sum_{\mathbf{v}\in\mathcal{A}} \frac{1}{\hat{u}} p_{\mathbf{v}} z^{\mathsf{T}}\mathbf{v}\mathbf{v}^{\mathsf{T}} z} \leq 2) \geq 1 - 2(\frac{2}{e})^{\frac{\hat{u}}{2}}$$

Step 3: The above two lemmas help us prove the approximation error of the randomized rounding with respect to a fixed parameter $\hat{u}$. However, the approximation errors need to be with respect to the optimal choice of $\hat{u}$, u^*, which is the size of the optimal subset from equation 2.

We claim the following about the solution of the convex relaxation at $\hat{u}$ and $\zeta_{12}d$,

$$\begin{aligned}&\lambda^*_{\min}(\zeta_{12}d) \leq \operatorname*{arg\,max}_{\boldsymbol{\mu}\in\mathcal{P}(\mathcal{A});\|\boldsymbol{\mu}\|_\infty \leq \frac{1}{\zeta_{12}d}} \inf_{z\in\mathcal{B}^d} \sum_{\mathbf{v}\in\mathcal{A}} \boldsymbol{\mu}_{\mathbf{v}} Z_{\mathbf{v}}(\mathbf{z}) \leq \operatorname*{arg\,max}_{\boldsymbol{\mu}\in\mathcal{P}(\mathcal{A});\|\boldsymbol{\mu}\|_\infty \leq \frac{1}{\zeta_{12}\hat{u}}} \inf_{z\in\mathcal{B}^d} \sum_{\mathbf{v}\in\mathcal{A}} \frac{1}{\zeta_{12}d} Z_{\mathbf{v}}(\mathbf{z}) \\ &\leq \frac{\hat{u}}{\zeta_{12}d} \operatorname*{arg\,max}_{\boldsymbol{\mu}\in\mathcal{P}(\mathcal{A});\|\boldsymbol{\mu}\|_\infty \leq \frac{1}{\hat{u}}} \inf_{z\in\mathcal{B}^d} \sum_{\mathbf{v}\in\mathcal{A}} \frac{1}{\hat{u}} Z_{\mathbf{v}}(\mathbf{z}) \leq \frac{\hat{u}}{\zeta_{12}d} \operatorname*{arg\,max}_{\boldsymbol{\mu}\in\mathcal{P}(\mathcal{A});\|\boldsymbol{\mu}\|_\infty \leq \frac{1}{\hat{u}}} \inf_{z\in\mathcal{B}^d} \sum_{\mathbf{v}\in\mathcal{A}} \frac{1}{\hat{u}} Z_{\mathbf{v}}(\mathbf{z}) \\ &\leq \frac{\hat{u}}{\zeta_{12}d} \operatorname*{arg\,max}_{\boldsymbol{\mu}\in\mathcal{P}(\mathcal{A});\|\boldsymbol{\mu}\|_\infty \leq \frac{1}{\hat{u}}} \inf_{z\in\mathcal{B}^d} \sum_{\mathbf{v}\in\mathcal{A}} \frac{1}{\hat{u}} Z_{\mathbf{v}}(\mathbf{z}) \leq \frac{\hat{u}}{\zeta_{12}d} \operatorname*{arg\,max}_{\boldsymbol{\mu}\in\mathcal{P}(\mathcal{A});\|\boldsymbol{\mu}\|_\infty \leq \frac{1}{\hat{u}}} \inf_{z\in\mathcal{B}^d} \sum_{\mathbf{v}\in\mathcal{A}} \boldsymbol{\mu}_{\mathbf{v}} Z_{\mathbf{v}}(\mathbf{z}) \\ &\leq \frac{\hat{u}}{\zeta_{12}d} \lambda^*_{\min}(\hat{u})\end{aligned} \tag{17}$$

The last inequality follows from the fact that $\frac{1}{\hat{u}}$ lies is in the feasibility set.

Let u^* be the size of the optimal subset. If $u^* > \zeta_{12}d$, (otherwise we do a subset search), the convex relaxation at u^* is more constrained than the one at $\zeta_{12}d$,

$$\lambda^*_{\min}(u^*) \leq \lambda^*_{\min}(\zeta_{12}d) \leq \frac{\hat{u}}{\zeta_{12}d}\lambda^*_{\min}(\hat{u}) \implies \lambda^*_{\min}(\hat{u}) \geq \frac{\zeta_{12}d}{\hat{u}}\lambda^*_{\min}(u^*)$$

Putting Everything Together: We can now combine the two lemmas and the equation above to say that for an error $\epsilon = \zeta_{10}\frac{t\sqrt{d}}{\sqrt{\hat{u}^3}}$, to obtain the following,

$$\hat{\lambda}_{\min} \geq \frac{\zeta_{12}d}{2\hat{u}}\lambda^*_{\min}(u^*) - 2\zeta_{10}\sqrt{\frac{d^3M^4}{\hat{u}^3}}$$

Set ζ_{12} to be a large enough constant, $\hat{u} = \zeta_{12}d$ and $t = 2$,

$$\hat{\lambda}_{\min} \geq \frac{\lambda^*_{\min}(u^*)}{2} - 2\zeta_{10}\sqrt{\frac{1}{\zeta_{12}^3}}$$

with probability $\frac{1}{2} - \exp(-2\mathsf{M}) - 2(\frac{2}{e})^{\frac{\hat{u}}{2}}$.

If a lower bound $\lambda^l_{\min}$ of $\lambda^*_{\min}(u^*)$ is known, we set ζ_{12} to be large enough such that $2\zeta_{10}\sqrt{\frac{1}{\zeta^3_{12}}} \leq \frac{\lambda^l_{\min}}{4}$, making the inequality,

$$\hat{\lambda}_{\min} \geq \frac{1}{2}\lambda^*_{\min}(u^*) - \frac{\lambda^l_{\min}}{4} \implies \hat{\lambda}_{\min} \geq \frac{1}{4}\lambda^*_{\min}(u^*)$$

Therefore we need $\zeta_{12} \geq \frac{64^{2/3}(\zeta_{10})^{2/3}}{(\lambda^l_{\min})^{2/3}}$

To increase the probability of success , we can then run this multiple times and then take the maximum of the minimum eigenvalues of the different sampled subsets, as is standard when using randomized algorithms.

C.5 Proof of Regret Guarantees of Theorem 3

Proof. The regret bound can be proved in 3 steps. First, we decompose the regret, apply Corollary 1, and then optimize the exploration period.

Step 1. Regret Decomposition: $M = \sup_{\mathbf{a}\in\mathcal{A}} \|\mathbf{a}\|_\infty$. Define the maximum reward as, $R_{\max} = \max_{\mathbf{a}\in\mathcal{A}} |\boldsymbol{\theta}^\mathsf{T}\mathbf{a}|$ and $\mathbf{a}^*$ as the corresponding arm. The regret can be decomposed as,

$$\begin{aligned}
\mathbb{E}[\mathsf{Reg}(\mathsf{T})] =& \mathbb{E}_{\boldsymbol{\theta}}\left[\sum_{t=1}^{\mathsf{T}} \left\langle \boldsymbol{\theta}, \mathbf{a}^{\pi(t)} - \mathbf{a}_t \right\rangle\right] = \mathbb{E}_{\boldsymbol{\theta}}\left[\sum_{t=1}^{\mathsf{T}} \left\langle \boldsymbol{\theta}, \mathbf{a}^{\pi(t)} - \mathbf{a}_t \right\rangle\right] \\
&= \mathbb{E}_{\boldsymbol{\theta}}\left[\sum_{t=1}^{\mathsf{T}_{\text{explore}}} \left\langle \boldsymbol{\theta}, \mathbf{a}^{\pi(t)} - \mathbf{a}_t \right\rangle\right] + \mathbb{E}_{\boldsymbol{\theta}}\left[\sum_{t=\mathsf{T}_{\text{explore}}+1}^{\mathsf{T}} \left\langle \boldsymbol{\theta}, \mathbf{a}^{\pi(t)} - \mathbf{a}_t \right\rangle\right] \\
&\leq 2\mathsf{T}_{\text{explore}} R_{\max} + \mathbb{E}_{\boldsymbol{\theta}}\left[\sum_{t=\mathsf{T}_{\text{explore}}+1}^{\mathsf{T}} \left\langle \boldsymbol{\theta}, \mathbf{a}^{\pi(t)} - \mathbf{a}_t \right\rangle\right]
\end{aligned}$$

The decomposition step requires extra care since the regret is with respect to the top-$\mathsf{T}_{\text{explore}}$ arms. In the exploitation stage, the arms are selected such that the top $(\mathsf{T} - \mathsf{T}_{\text{explore}})-$arms are played according to $\widehat{\boldsymbol{\theta}}$ and are indexed by the permutation $\widehat{\pi}$ (that is $\mathbf{a}_t = \mathbf{a}^{\widehat{\pi}(j)}$). We next bound the regret for the j^{th} selected arm.

1. If $\boldsymbol{\theta}^\mathsf{T}\mathbf{a}^{(\widehat{\pi}(j))} \geq \boldsymbol{\theta}^\mathsf{T}\mathbf{a}^{(\pi(j))}$, then the regret for the j^{th} selected arm is negative.
2. If not i.e., $\boldsymbol{\theta}^\mathsf{T}\mathbf{a}^{(\widehat{\pi}(j))} \leq \boldsymbol{\theta}^\mathsf{T}\mathbf{a}^{(\pi(j))}$, then there exists an arm index, j_1 in the permutation π such that j_1 is shifted to the left in $\widehat{\pi}$. This implies that $\boldsymbol{\theta}^\mathsf{T}\mathbf{a}^{(\widehat{\pi}(j_1))} \geq \boldsymbol{\theta}^\mathsf{T}\mathbf{a}^{(\pi(j_1))}$. Note that by our estimation guarantees of estimating $\boldsymbol{\theta}$ upto ϵ accuracy with high probability, $(\widehat{\boldsymbol{\theta}}^\mathsf{T}\mathbf{a}^{(\widehat{\pi}(j))} - \boldsymbol{\theta}^\mathsf{T}\mathbf{a}^{(\widehat{\pi}(j))}) \leq \|\widehat{\boldsymbol{\theta}}^\mathsf{T} - \boldsymbol{\theta}^\mathsf{T}\|_1 \|\mathbf{a}^{(\widehat{\pi}(j))}\|_\infty \leq M\epsilon$ with probability $1 - \frac{1}{d^{\zeta_2 - 1}}$ (ζ_2 is a large enough constant). We decompose the regret for this case with respect to this index and bound the error:

$$\begin{aligned}
\boldsymbol{\theta}^\mathsf{T}\mathbf{a}^{(\pi(j))} - \boldsymbol{\theta}^\mathsf{T}\mathbf{a}^{(\widehat{\pi}(j))} =& \underbrace{(\boldsymbol{\theta}^\mathsf{T}\mathbf{a}^{(\pi(j))} - \boldsymbol{\theta}^\mathsf{T}\mathbf{a}^{(\widehat{\pi}(j_1))})}_{\leq 0} + \underbrace{(\boldsymbol{\theta}^\mathsf{T}\mathbf{a}^{(\widehat{\pi}(j_1))} - \widehat{\boldsymbol{\theta}}^\mathsf{T}\mathbf{a}^{(\widehat{\pi}(j_1))})}_{\leq M\epsilon} \\
&+ \underbrace{(\widehat{\boldsymbol{\theta}}^\mathsf{T}\mathbf{a}^{(\widehat{\pi}(j_1))} - (\widehat{\boldsymbol{\theta}}^\mathsf{T}\mathbf{a}^{(\widehat{\pi}(j))})}_{\leq 0} + \underbrace{(\widehat{\boldsymbol{\theta}}^\mathsf{T}\mathbf{a}^{(\widehat{\pi}(j))} - \boldsymbol{\theta}^\mathsf{T}\mathbf{a}^{(\widehat{\pi}(j))})}_{\leq M\epsilon} \\
&\leq \left\langle \boldsymbol{\theta}^\mathsf{T} - \widehat{\boldsymbol{\theta}}^\mathsf{T}, \mathbf{a}^{(\pi(j))} \right\rangle + \left\langle \boldsymbol{\theta}^\mathsf{T} - \widehat{\boldsymbol{\theta}}^\mathsf{T}, \mathbf{a}^{(\pi(j_1))} \right\rangle \leq 2M\epsilon
\end{aligned}$$

with probability $\nu = 1 - \frac{1}{d^{1-\zeta_2}} - \exp(-\zeta_5 n)$.

We therefore obtain the following,

$$\mathbb{E}[\mathsf{Reg}(\mathsf{T})] \leq 2M\mathsf{T}_{\text{explore}} R_{\max} + 2(\mathsf{T} - \mathsf{T}_{\text{explore}})\nu\epsilon + M(\mathsf{T} - \mathsf{T}_{\text{explore}})(1-\nu)R_{\max},$$

For a high enough ζ_3, $1-\nu = o(1)$ and the last term is very small. Further $\mathsf{T} \gg \mathsf{T}_{\text{explore}}$ (the number of exploration rounds is sublinear in T) we obtain,

$$\mathbb{E}[\mathsf{Reg}(\mathsf{T})] = \mathbb{E}_{\boldsymbol{\theta}}\left[\sum_{t=1}^{\mathsf{T}} \left\langle \boldsymbol{\theta}, \mathbf{a}^{\pi(t)} - \mathbf{a}_t \right\rangle\right] = \tilde{\mathsf{O}}\left(R_{\max}\mathsf{T}_{\text{explore}} + M\|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\|_1 \mathsf{T}\right)$$

Step 2. Fast Sparse Learning: We use Theorem 1, which is proved in the appendix, to obtain an estimation guarantee in terms of the number of exploration rounds. And we now apply the bound from Corollary 1 and obtain the following (Making an assumption similar to Hao et al. (2020) on the exploration rounds $\mathsf{T}_{\text{explore}} > \mathsf{O}(k\lambda_{\min}^{-4})$.

Step 3. Exploration Period Optimization: (The probability of error terms $(1-\nu)$ are left out in the expression.) $\epsilon = \widetilde{\mathsf{O}}\left(\sigma \frac{(5+4\beta_k)^2 k}{\lambda^*_{\min}} \sqrt{\frac{1}{\mathsf{T}_{\text{explore}}}}\right)$ equation **??** can then be bounded as,

$$\mathbb{E}[\mathsf{Reg}(\mathsf{T})] = \widetilde{\mathsf{O}}\left(R_{\max}\mathsf{T}_{\text{explore}} + \mathsf{T}(\sigma M \frac{(5+4\beta_k)^2 k}{\lambda^*_{\min}} \sqrt{\frac{1}{\mathsf{T}_{\text{explore}}}})\right)$$

Setting $\mathsf{T}_{\text{explore}} = R_{\max}^{-2/3} \hat{\lambda_{\min}}^{-2/3} k^{2/3} \mathsf{T}^{2/3}$ we obtain the following,

$$\mathbb{E}[\mathsf{Reg}(\mathsf{T})] = \widetilde{\mathsf{O}}\left(R_{\max}^{1/3}\left(\frac{1}{\hat{\lambda_{\min}}^{-2/3}} + \sigma M \frac{(5+4\beta_k)^2 \hat{\lambda_{\min}}^{1/3}}{\lambda^*_{\min}}\right) k^{2/3}\mathsf{T}^{2/3}\right)$$

Further using guarantees from Theorem 2, we use $\hat{\lambda_{\min}} = \Omega(\lambda^*_{\min})$ with high probability and $M = 1$,

$$\mathbb{E}[\mathsf{Reg}(\mathsf{T})] = \widetilde{\mathsf{O}}\left(R_{\max}^{1/3}\sigma(5+4\beta_k)^2 {\lambda^*_{\min}}^{-2/3} k^{2/3}\mathsf{T}^{2/3}\right).$$

□

We can also obtain a regret bound for the case of hard sparsity which is of the same order as Hao et al. (2020).

Corollary 2. *Let $\boldsymbol{\theta}$ be k-sparse, $\|\boldsymbol{\theta}\|_0 \le k$ in the sparse linear bandits framework of Theorem 3. Let $\lambda^*_{\min}$ be the minimum eigenvalue from equation 2 with the same assumptions as Theorem 3. Then Algorithm* **BSLB** *with exploration period $\mathsf{T}_{explore} = \mathsf{O}(k^{\frac{2}{3}}\mathsf{T}^{\frac{2}{3}})$, achieves a regret guarantee of $\mathbb{E}[\mathsf{Reg}(\mathsf{T})] = \mathsf{O}((\lambda^*_{\min})^{-1}k^{\frac{2}{3}}\mathsf{T}^{\frac{2}{3}})$.*

C.6 Alternative to Subset Algorithm

We can reduce the time complexity of Algorithm 2 by considering a bound on the eigenvalue of the sampled vectors in terms of a lower bound. The tradeoff here is in the regret gaurantee - the denominator gets an additional factor of $(\lambda^l_{\min})^2$. First note the following decomposition,

$$\frac{\lambda_{\min}(\mathcal{G}^*)}{\zeta_{12}} \le \frac{d\lambda_{\min}(\mathcal{G}^*)}{\zeta_{12}d} \le \frac{u^*\lambda_{\min}(\mathcal{G}^*) + (\hat{u}-u^*)\lambda_{\min}(\mathcal{J})}{\hat{u}}$$

$$\le \lambda_{\min}(\mathcal{G}^* \cup \mathcal{J}) \le \lambda^*_{\min}(\hat{u}) \overset{(a)}{\le} 2(\hat{\lambda_{\min}} + 2\zeta_{10}\sqrt{\frac{d^3}{\hat{u}^3}})$$

$$\implies \frac{\lambda_{\min}(\mathcal{G}^*)}{2\zeta_{12}} - \zeta_{10}\sqrt{\frac{1}{\zeta_{12}^3}} \le \hat{\lambda_{\min}}$$

where (a) follows from the approximation guarantees of the previous section.

Now plugging in $\zeta_{12} = \frac{16\zeta_{10}^2}{(\lambda^l_{\min})^2}$ we obtain,

$$\frac{\lambda^*_{\min}(u^*)(\lambda^l_{\min})^2}{32\zeta_{10}^2} - \frac{(\lambda^l_{\min})^3}{64\zeta_{10}^2} \le \hat{\lambda_{\min}}$$

$$\implies \hat{\lambda_{\min}} \ge \frac{(\lambda^l_{\min})^2(2\lambda^*_{\min}(u^*) - \lambda^l_{\min})}{64\zeta_{10}^2} \ge \frac{(\lambda^l_{\min})^2(\lambda^*_{\min}(u^*))}{64\zeta_{10}^2}$$

Therefore this choice of ζ_{12} (and hence $\hat{u}$, leads to a linear time algorithm however the regret bound is now $(\lambda^*_{\min})^{-1}(\lambda^l_{\min})^{-2}$ instead of $(\lambda^*_{\min})^{-1}$.

C.7 Subset Search Algorithm for Searching the Optimal Subset

From the previous previous proof we can set $\hat{u} = \zeta_{12}d$, and then search for subsets in the range $[d, \hat{u}]$ to obtain a minimum eigenvalue $\hat{\lambda_{\min}}$. We obtain the approximation guarantee, $\hat{\lambda_{\min}} \geq \frac{1}{2}\lambda^*_{\min}$ w.h.p., since we are only using the approximation guarantee from Lemma 4 and Lemma 5, and not from equation 17 because we are already searching the space $< \hat{u}$. Since the search space is dependent on $\mathsf{Poly}(d)$, the time complexity of the subset search algorithm, Algorithm 3 follows. This time complexity is substantially smaller than the complexity over the search over all arms which is of the order $\mathsf{O}(\exp(\mathsf{M}))$.

Algorithm 3 SubsetSearch : Subset Search for Optimal Subset

Input Approximation Factor ϵ, Search Bound $\hat{u}$
Output Subset $\bar{\mathcal{G}}$
Set $\bar{\lambda_{\min}} = 0$, $\bar{\mathcal{G}} = \phi$
for $\bar{d}$ in $\{d, \ldots, \hat{u}\}$ **do**
  for $\mathcal{G}'$ in $\{\mathcal{G} \subseteq \mathcal{A}; |\mathcal{A}| = \bar{d}\}$ **do**
    if $\lambda_{\min}(\sum_{\mathbf{a}\in\mathcal{G}'} |\mathcal{G}'^{-1}|\mathbf{a}\mathbf{a}^\mathsf{T}) > \bar{\lambda_{\min}}$ **then**
      Set $\bar{\lambda_{\min}} = \lambda_{\min}(\sum_{\mathbf{a}\in\mathcal{G}'} |\mathcal{G}'^{-1}|\mathbf{a}\mathbf{a}^\mathsf{T})$, $\bar{\mathcal{G}} = \mathcal{G}'$
    end if
  end for
end for

What if the maximum minimum eigenvalue $\lambda^*_{\min}$ is not known ? We can use a lower bound on the $\lambda^*_{\min}$. This is easy to obtain: Randomly sample subsets of the arm set $\mathcal{A}$ and compute the objective value in Equation 2 for each subset - the lower bound can be the maximum objective value across the sampled subsets.

What can the practitioner do to select a good subset size empirically? Additionally, if a practitioner wants to test out a particular choice of $\hat{u}$, the worst-case error can be empirically calculated (the difference between the convex relaxation at d and averaged across multiple randomized rounding runs for different values of $\hat{u}$). This is possible because the randomized rounding in GetGoodSubset($\mathcal{A}$) can be run offline.

C.8 Why Does the Average Minimum Eigenvalue Constraint Not Satisfy Matroid Constraints

Existing work in experiment design work with objective functions which often satisfy the matroid, submodularity or cardinality constraints to perform experiment design Allen-Zhu et al. (2021b); Brown et al. (2024). However we need to optimize the minimum eigenvalue averaged across the subset (because we want to avoid dependence of M in the regret term and also use the RE condition). Our objective clearly does not satisfy the cardinality constraint, and the feasible sets don't satisfy a matroid constraint (since removing a vector might improve the minimum eigenvalue averaged across the set, so there is no clear partitioning/structure to the feasible set). Finally, we tried to prove submodularity but were unable to do so for our objective function especially because the denominator is dependent on the subset size.

C.9 Details on corralling

Brief Overview on Corralling: The algorithm CORRAL Agarwal et al. (2017) is a meta-bandit algorithm that uses online mirror descent and importance sampling to sample different bandit algorithms that receive the reward. Using the rewards updates the probabilities used for sampling. The main objective is to achieve a regret which is as good as if the best base algorithm was run on its own.

To setup the context, we exactly reproduce the following excerpt, definition and theorems have been taken from Agarwal et al. (2017):

For an environment $\mathcal{E}$, we define the environment $\mathcal{E}'$ induced by importance weighting, which is the environment that results when importance weighting is applied to the losses provided by environment $\mathcal{E}$. More precisely, $\mathcal{E}'$ is defined as follows. On each round $t = 1, \ldots, T$,

1. $\mathcal{E}'$ picks an arbitrary sampling probability $p_t \in (0,1]$ and obtains $(x_t, f_t) = \mathcal{E}(\theta_1, \ldots, \theta_{t-1})$.

2. $\mathcal{E}'$ reveals x_t to the learner and the learner makes a decision θ_t.

3. With probability p_t, define $f'_t(\theta, x) = f_t(\theta, x)/p_t$ and $\theta'_t = \theta_t$; with probability $1 - p_t$, define $f'_t(\theta, x) = 0$ and $\theta'_t \in \Theta$ to be arbitrary.

4. $\mathcal{E}'$ reveals the loss $f'_t(\theta_t, x_t)$ to the learner, and passes θ'_t to $\mathcal{E}$.

Definition 3. *Agarwal et al. (2017) For some $\alpha \in (0,1]$ and non-decreasing function $R : \mathbb{N}_+ \to \mathbb{R}_+$, an algorithm with decision space $\mathcal{O}$ is called (α, R)-stable with respect to an environment $\mathcal{E}$ if its regret under $\mathcal{E}$ is $R(T)$, and its regret under any environment $\mathcal{E}'$ induced by importance weighting is*

$$\sup_{\theta \in \Theta} \mathbb{E}\left[\sum_{t=1}^{T} f_t(\theta, x_t) - f_t(\theta, x_t)\right] \leq \mathbb{E}[\rho^\alpha] R(T) \qquad (2)$$

where $\rho = \max_{t \in [T]} 1/p_t$ (with p_t as in the definition of $\mathcal{E}'$ above), and all expectations are taken over the randomness of both $\mathcal{E}'$ and the algorithm.

Similar too most reasonable Base Algorithms it can be seen that the BSLB algorithm satisfies is $(1, \mathbb{E}[\mathsf{Reg}(\mathsf{T})])$-stable by rescaling the losses.

We

Theorem 8. *(Theorem 4 in Agarwal et al. (2017)) For any $i \in [M]$, if base algorithm $\mathcal{B}_i$ (with decision space $\mathcal{O}_i$) is (α_i, R_i)-stable (recall Defn. 3) with respect to an environment $\mathcal{E}$, then under the same environment CORRAL satisfies*

$$\sup_{\theta \in \Theta,\ \pi \in \Pi} \mathbb{E}\left[\sum_{t=1}^{T} f_t(\theta_t, x_t) - f_t(\theta, x_t)\right] = \widetilde{O}\left(\frac{M}{\eta} + T\eta\frac{|\mathcal{O}_{\pi_t}|}{\eta} + \frac{\alpha_i}{\eta\beta} R_i(T)\right), \qquad (3)$$

where all expectations are taken over the randomness of CORRAL Algorithm, the base algorithms, and the environment.

Theorem 9. *(Theorem 5 in Agarwal et al. (2017)) Under the conditions of Theorem 7, if $\alpha_i = 1$, then with $\eta = \min\left\{\frac{1}{4\theta R_i(T)\ln T}, \sqrt{\frac{1}{T}}\right\}$ CORRAL satisfies: $\sup_{\theta \in \Theta, \pi \in \Pi}\left[\sum_{t=1}^{T} f_t(\theta_t, \pi_t) - f_t(\theta, \pi_t)\right] = \widetilde{O}\left(\sqrt{MT} + MR_i(T)\right)$.*

C.10 Proof of Corralling Guarantees for Theorem 5

Algorithm 4 Corralling with Blocked Linear Sparse Bandits (`C-BSLB`)

1: Input: Dimension d, Total Number of Rounds T, Regret Bound of Best Algorithm R_{best}
2: Set Learning rate $\eta = \min\left(\frac{1}{40\mathsf{T}R_{\text{best}}}, \sqrt{\frac{\lfloor \log_2(d) \rfloor}{\mathsf{T}}}\right)$
3: Set Exponential Grid $k \in [1, 2, \ldots 2^{\lfloor \log_2(d) \rfloor}]$
4: Initialize $\lfloor \log_2(d) \rfloor + 1$ Base Algorithms one for each sparsity parameter on an exponential grid, $\texttt{BSLB}(\mathsf{T}_{\text{explore}} = \zeta k^{1/3}\mathsf{T}^{2/3})$
5: Sample $\mathsf{M}_{\text{sampled}} = \zeta d^{1/3}\mathsf{T}^{2/3}$ arms without replacement to be used as proxy samples.
6: Run Corral($\lfloor \log_2(d) \rfloor + 1$ `BSLB` algorithms, η) from Agarwal et al. (2017) with Base Algorithms and time horizon T. If an arm is suggested which is already pulled, pull an arm from the remaining set of arms uniformly at random.

Before presenting the proof, we clarify what we mean by the exponential scale with an example. For dimension $d = 1024$, the exponential scale will be $k \in \{1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024\}$, and we initialize a base algorithm each with the exploration period set according to the k.

Remark: Also Step 5 and Step 6 in Algorithm 4 needs explanation. Note that the CORRAL algorithm as a whole has to respect the *blocking* constraint. Even though CORRAL does not require the bandit algorithms to be run independently, we want to avoid changing CORRAL or the base algorithm. Instead we just change the arms that are available to sample by exploiting the fact that our base algorithm is a two step algorithm and each of the steps can be performed offline. For each base algorithm:

1. For the explore phase of the base algorithm we take arms from the intersection of the subset sampled in step 5 and the subset of arms which have not been sampled.

2. For the exploit phase if the chosen base algorithm provides an arm which has already been sampled by CORRAL. Then we provide the feedback corresponding to that arm. And then we pull an arm from the remaining set of arms without replacement.

Note that with this modification, the exploration phase of each of the algorithm runs as if the algorithm was being run independently. Hence the regret bounds for each individual base algorithm still holds. We can now prove Theorem 9.

We now use Theorem 9 with $M = \log_2\lceil d\rceil + 1$ algorithms. However if we simply apply the theorem we can bound with respect to the sparsity parameter which lies on the grid, $s \in \{2^i\}_{i=0}^{\log_2\lceil d\rceil}$,

$$\mathbb{E}[\mathsf{Reg}(\mathsf{T})] \leq \tilde{\mathsf{O}}(\sqrt{\log_2\lceil d\rceil \mathsf{T}} + \log_2\lceil d\rceil \mathbb{E}[\mathsf{Reg}(\mathsf{T})]_s)$$

But the optimal sparsity parameter $k^\star$ may not lie on the grid and we need to bound $\mathbb{E}[\mathsf{Reg}(\mathsf{T})]_{k^\star}$ in terms of $\mathbb{E}[\mathsf{Reg}(\mathsf{T})]_s$. To that end we prove the following lemma,

Lemma 6. *Let $k^\star$ be the sparsity parameter at which the regret bound of* **??** *is minimized. And let $s \in \{2^i\}_{i=0}^{\log_2\lceil d\rceil}$ be the parameter on grid which is closest to $k^\star$ in absolute distance. Then the following holds (where ν is the probability of exploration round succeeding at sparsity level $k^\star$),*

$$\mathbb{E}[\mathsf{Reg}(\mathsf{T})]_s \leq \sqrt{2k^*}\mathbb{E}[\mathsf{Reg}(\mathsf{T})]_{k^\star} + \log_2(2k^*)\mathsf{O}(1-\nu)$$

Proof. Let the bound on the expected regret for sparsity level k be given by $\mathbb{E}[\mathsf{Reg}(\mathsf{T})]_k$.

From the statement of the theorem, we assume that for the optimal sparsity parameter $k^\star$, the nearest parameter (on the exponential scale) is s. $k^\star$ lies in the interval $[\lceil s/2\rceil, 2s]]$ (otherwise s would not be the closest parameter on the exponential scale.). Therefore if we were perform a binary search for $k^\star$, we would need at most $Y = \lfloor \log_2(4k^\star)\rfloor$ queries to search for $k^\star$. Let $k_1^*, k_2^*, \ldots, k_Y^*$ be the mid-points of these queries, where $k_Y^* = k^\star$. Now each of them is such that $k_j^* = \alpha k_{j-1}^*$, where $\alpha \in [0.75, 1.25]$.

First consider the case when $\alpha \in [0.75, 1]$, then by substituting $k = \lfloor \alpha k\rfloor$, in the regret bound of Theorem 3, the following inequalities can be obtained ,

$$\mathbb{E}[\mathsf{Reg}(\mathsf{T})]_{\lfloor \alpha k\rfloor} \leq (\frac{1}{\alpha})^{1/2}\mathbb{E}[\mathsf{Reg}(\mathsf{T})]_k + (1-\nu)\mathsf{O}(\mathsf{T}) \leq \sqrt{2}\mathbb{E}[\mathsf{Reg}(\mathsf{T})]_k + (1-\nu)\mathsf{O}(\mathsf{T}).$$

Now for the case, $\alpha \in [1, 1.25]$, we substitute for $k = \lceil \alpha k\rceil$

$$\mathbb{E}[\mathsf{Reg}(\mathsf{T})]_{\lceil \alpha k\rceil} \leq (\alpha)^{1/3}\mathbb{E}[\mathsf{Reg}(\mathsf{T})]_k + (1-\nu)\mathsf{O}(\mathsf{T}) \leq \sqrt{2}\mathbb{E}[\mathsf{Reg}(\mathsf{T})]_k + (1-\nu)\mathsf{O}(\mathsf{T})$$

The probability of success of each of them is $1 - o(1)$ and $\log(4k^*)$ times the probability of error is still $o(1)$.

Now we can take a cascade of products by decomposing $\mathbb{E}[\mathsf{Reg}(\mathsf{T})]_{k^*}$ using the above inequality in the direction of $k_1^*, k_2^*, \ldots, k_Y^*$. (i.e. we can decompose $k^\star = \alpha_1\alpha_2\ldots\alpha_Y k$,

$$\mathbb{E}[\mathsf{Reg}(\mathsf{T})]_s \leq \alpha\mathbb{E}[\mathsf{Reg}(\mathsf{T})]_{k_1^*} + O(1-\nu) \leq \cdots \leq (\sqrt{2})^Y\mathbb{E}[\mathsf{Reg}(\mathsf{T})]_{k^\star} + Y\mathsf{O}(1-\nu)$$
$$\mathbb{E}[\mathsf{Reg}(\mathsf{T})]_s \leq 2\sqrt{k^\star}\mathbb{E}[\mathsf{Reg}(\mathsf{T})]_{k^\star} + \log_2(4k^\star)\mathsf{O}(1-\nu)$$

□