# ANNOTATION EFFICIENCY: IDENTIFYING HARD SAM PLES VIA BLOCKED SPARSE LINEAR BANDITS

Anonymous authors

Paper under double-blind review

### ABSTRACT

This paper considers the problem of annotating datapoints using an expert with only a few annotation rounds in a *label-scarce* setting. We propose soliciting reliable feedback on difficulty in annotating a datapoint from the expert in addition to ground truth label. Existing literature in active learning or coreset selection turns out to be less relevant to our setting since they presume the existence of a reliable trained model, which is absent in the label-scarce regime. However, the literature on coreset selection emphasizes the presence of difficult data points in the training set to perform supervised learning in downstream tasks (Mindermann et al., 2022). Therefore, for a given fixed annotation budget of T rounds, we model the sequential decision-making problem of which (difficult) datapoints to choose for annotation in a sparse linear bandits framework with the constraint that no arm can be pulled more than once (*blocking constraint*). With mild assumptions on the datapoints, our (computationally efficient) Explore-Then-Commit algorithm BSLB achieves a regret guarantee of  $\widetilde{O}(k^{\frac{1}{3}}T^{\frac{2}{3}} + k^{-\frac{1}{2}}\beta_k + k^{-\frac{1}{12}}\beta_k^{\frac{1}{2}}T^{\frac{5}{6}})$  where the unknown parameter vector has tail magnitude  $\beta_k$  at sparsity level k. To this end, we show offline statistical guarantees of Lasso estimator with mild Restricted Eigenvalue (RE) condition that is also robust to sparsity. Finally, we propose a meta-algorithm C-BSLB that does not need knowledge of the optimal sparsity parameters at a no-regret cost. We demonstrate the efficacy of our BSLB algorithm for annotation in the label-scarce setting for an image classification task on the PASCAL-VOC dataset, where we use real-world annotation difficulty scores.

031 032

033 034

004

010 011

012

013

014

015

016

017

018

019

021

023

024

025

026

027

028

029

### 1 INTRODUCTION

In niche industrial applications such as Named Entity Recognition (Nguyen et al., 2023) and learning tasks on low-resource languages (Hedderich et al., 2021), obtaining high-quality labels is challenging due to the lack of expert annotators. However, high-quality labels are critical for effective model 037 training (Li et al., 2023), and thus, expert annotators must provide ground truth labels. (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 040 of unlabeled data, the challenge is to select a representative subset of datapoints for labeling. In 041 label-scarce tasks, where the number of expert annotators is extremely low, often only one, it is im-042 practical to query the same datapoint multiple times. While crowd-sourcing literature reduces noise 043 by aggregating labels from multiple annotators (Verroios & Garcia-Molina, 2015), the annotation 044 from a single or aggregated expert is considered the final ground truth label in our setting. We term this restriction, where a datapoint cannot be re-queried after annotation, the *blocking constraint*. Additionally, the annotation budget is typically much smaller than the datapoint embedding dimension. 046 An efficient annotation strategy should be sequential (instead of one-shot) in such cases, as each 047 annotation informs future decisions and helps identify more informative datapoints. 048

In addition, data-pruning techniques like coreset selection emphasize selecting hard examples (Maharana et al., 2023), while (Sorscher et al., 2022) justifies this for perceptron learning. Curriculum
learning (Bengio et al., 2009) also uses increasingly difficult examples, but defining the 'hardness' of unlabeled data is ambiguous. Hard examples identified by heuristics are often noisy, mislabeled, or outliers (Mindermann et al., 2022). *To address this, we propose soliciting annotation difficulty feedback directly from expert annotators*. This adds minimal annotation overhead and helps identify

hard examples more reliably than heuristics. Additionally, leveraging expert feedback to learn and predict hardness (via an auxillary model) is simpler than addressing complex downstream task itself.

Thus, the primary problem we aim to solve is: "Given a small annotation budget T, how should we 057 sequentially identify unique and difficult unlabeled datapoints for annotation as we receive expert 058 feedback on hardness of annotation?" Theoretically, we model this sequential decision-making process using sparse linear bandits with a blocking constraint. Specifically, suppose we have a 060 large set A of M unlabeled datapoints, each associated with a d-dimensional embedding. With T 061 rounds available, at each round  $t \in [T]$ , a datapoint  $\mathbf{a}_t$  is selected for annotation, and the expert 062 provides noisy feedback  $r_t$  on the difficulty of labeling that datapoint, along with the ground truth 063 label. To address the limited annotation budget and the relevance of a few important features in high-064 dimensional spaces (Hao et al., 2020), we model the expected hardness as a sparse linear function of the embedding. We ensure robustness to sparsity - our theoretical guarantees degrade gracefully as 065 the parameter vector becomes less sparse, i.e., as it deviates from sparsity due to an increasing tail. 066

067 Although our primary motivation comes from data selection for annotation in a label-scarce regime, 068 our theoretical framework is broadly applicable. For example, resource-constrained edge devices, 069 such as smartphones, consider the problem of recommendation of personalized products. In applications such as movie/book recommendations, typical users will hardly consume an item more 071 than once - so any feedback provided is not subject to change. Furthermore, the user will provide feedback for every item (books/movies) in a large repository of items. Here, our theoretical frame-072 work (sparse linear bandits with *blocking constraint*) is directly applicable as it involves learning 073 unknown user taste from sequential item recommendations based on user feedback. While existing 074 literature addresses this by combining feedback from multiple users (Bresler et al., 2014; Pal et al., 075 2024), such algorithms compromise user privacy. Our framework will not have this limitation. 076

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

084 085

### 1.1 PROBLEM FORMULATION AND PRELIMINARIES

Consider a dataset with a set (or more generally, a multi-set) of M unlabeled data-points  $\mathcal{A} \equiv \{\mathbf{a}^{(1)}, \mathbf{a}^{(2)}, \dots, \mathbf{a}^{(M)}\} \subseteq \mathbb{R}^d$ , each of which will be referred to as an *arm* in the bandit setup. Let  $\mathbf{a}^{(j)}$  denote the *d*-dimensional vector embedding associated with the  $j^{\text{th}}$  data-point (arm). The arm embedding vectors are contained in a ball of radius R that is  $\|\mathbf{a}\|_2 \leq R \forall \mathbf{a} \in \mathcal{A}$ .

091 We have an annotation budget of T rounds. In our annotation-poor regime, we have T  $\ll d \ll$ 092 M; that is, the annotation budget is much smaller than the ambient dimension, which, in turn, is significantly smaller than the number of arms. At each round  $t \in [T]$ , an unlabelled data-point  $\mathbf{a}_t$ 094 (corresponding to the arm which has not been pulled in the first t-1 rounds) is selected by the 095 online algorithm (decision-maker) and sent to the expert annotator for labeling. Note that such a 096 selection mechanism respects the blocking constraint, which stipulates that no unlabeled datapoint will be sent for annotation more than once. The expert <sup>1</sup>, as feedback, labels the datapoint  $a_t$  and 097 provides  $r_t$  that corresponds to the difficulty experienced in providing the ground truth label for  $\mathbf{a}_t$ . 098 We model the expected hardness  $\mathbb{E}r_t \langle \theta, \mathbf{a}_t \rangle$  as a linear function of the arm embedding where  $\theta \in \mathbb{R}^d$ 099 is an unknown parameter vector. In particular, the random variable  $r_t$  is obtained as  $r_t = \theta^T \mathbf{a}_t + \eta_t$ 100 where  $\eta_t$  is zero-mean i.i.d noise random variable with bounded variance  $\sigma^2$ . More precisely, 101

- 1.  $\mathbb{E}[\eta_t | \mathcal{F}_t] = 0$  and  $\mathbb{E}[\eta_t^2 | \mathcal{F}_t] \leq \sigma^2$ , where  $\mathcal{F}_t = \{(\mathbf{a}_1, r_1), \dots, (\mathbf{a}_{(t-1)}, r_{(t-1)})\}$  denotes the filtration till round  $t \in [\mathsf{T}]$ .
- 2. For any sparsity level  $k \leq d$  we define the tail of the parameter vector  $\boldsymbol{\theta}$  as,  $\beta_k := \|\boldsymbol{\theta}_{\mathcal{T}_k^c}\|_1$  where  $\mathcal{T}_k$  denotes the set of k largest coordinates of  $\boldsymbol{\theta}$  by absolute value and  $\mathcal{T}_k^c = [d] \setminus \mathcal{T}_k$ .
- 105 106 107

102 103

<sup>&</sup>lt;sup>1</sup>If there are multiple experts, we can consider the final feedback to be aggregated, which does not change.

Note that in the special case of k = 0 for some sparsity level  $k \ll d$ ,  $\theta$  will be referred to as a hardsparse vector. In our set-up, we account for soft sparsity when the tail is non-zero and unknown - our statistical guarantees degrade gracefully as the tail magnitude increases. Now, we formally define the objective (commonly known as regret) in our online learning set-up, which also respects the *blocking constraint*. Our regret definition captures the difference in cumulative expected hardness of datapoints selected by the online algorithm versus the cumulative expected hardness of top T unique datapoints. Consider a permutation  $\pi : [|\mathcal{A}|] \to [|\mathcal{A}|]$  of arms such that for any i < j, we have  $\langle \theta, \mathbf{a}^{(\pi(i))} \rangle \geq \langle \theta, \mathbf{a}^{\pi(j)} \rangle$ . We define the regret Reg(T) for our setting as,

116

119

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

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

Our Contributions and Techniques: The sparse linear bandits framework in the data-sparse regime 123 was first studied by Hao et al. (2020); Jang et al. (2022) - however, these works do not consider either 124 the blocking constraint or robustness to sparsity - both of which pose novel technical challenges. We 125 propose BSLB (Blocked Sparse Linear Bandits), an efficient algorithm in this setting which is pri-126 marily an Explore-Then-Commit algorithm. In the exploration period, we carefully choose a subset 127 of arms from which we sample without replacement - the goal is to ensure that the expected covari-128 ance matrix has a large minimum eigenvalue. In the exploitation period, we use a lasso estimator 129 to estimate the unknown model parameters while being robust to sparsity. The optimal exploration period depends on the correct sparsity level of the unknown parameter vector, which is difficult 130 131 to set in practice. Therefore, we also present a meta-bandit algorithm based on corralling a set of base bandits algorithms (Agarwal et al., 2017) - obtaining the same order-wise regret guarantees but 132 without needing the knowledge of hyperparameters. Below, we summarize our main contributions: 133

- We theoretically model the problem of identifying hard (informative) unlabeled datapoints by high dimensional sparse linear bandits framework with blocking constraint - the number of rounds or annotation budget is very small, even smaller than the ambient dimension. The hardness is modeled as a sparse linear function of the datapoint embedding - we allow for robustness to sparsity in the parameters where the tail magnitude is non-zero and unknown.
- 138 2. We propose an efficient "explore then commit" (ETC) algorithm BSLB for regret minimization in 139 our framework (Theorem 2) that achieves a regret guarantee of  $O(k^{1/3}T^{2/3} + k^{-1/12}\beta_k^{1/2}T^{5/6})$ 140 for a fixed sparsity k and the corresponding tail magnitude  $\beta_k$  - for the special case of hard 141 sparsity (tail magnitude  $\beta_k = 0$ ), BSLB achieves a regret guarantee of  $O(k^{1/3}T^{2/3})$  (Corollary 142 2). The run-time of BSLB is polynomial in the number of datapoints and is therefore efficient.
- A novel ingredient (function GETGOODSUBSET()) in the proposed algorithm BSLB is to select *a good* set of representative datapoints from the large set of unlabelled datapoints. This objective
  entails solving a discrete non-convex optimization problem in Equation 2. We propose a convex
  relaxation of the objective and obtain a feasible solution via randomized rounding Theorem 3
  states that our computationally efficient algorithm has good approximation guarantees.
- 4. BSLB requires knowledge of sparsity k (which also controls the tail magnitude β<sub>k</sub>) to set exploration period length. Since this is challenging to set in practice, we propose a meta-algorithm C-BSLB that combines base algorithms with different exploration periods. C-BSLB achieves same regret guarantees order-wise (Theorem 4) as BSTB but without knowledge of sparsity.
  5. To unlide our humothesis and theorem we present experiments on the public impact closed for a statement.
- 5. To validate our hypothesis and theory, we present experiments on the public image classification dataset PASCAL VOC 2012. Additional experiments on text classification on SST-2 and personalized recommendation with MovieLens, Jester, and Goodbooks datasets are in Appendix A.2.

Technical Challenges: At a high level, our proof follows a similar structure as in the analysis of Algorithm ESTC proposed in (Hao et al., 2020). First of all, both in (Hao et al., 2020) and our setting, there are no assumptions on the datapoint embeddings - unlike most existing literature on sparsity where the Gram matrix is assumed to satisfy Restricted Isometry Property (RIP) (Boche et al., 2015). A much weaker condition (namely Restricted Eigenvalue (RE)) in characterizing nice statistical guarantees of the Lasso estimator for sparse linear regression was shown in (Bickel et al., 2009; Rudelson & Zhou, 2013) - it was shown that if the covariance matrix of a distribution satisfies RE, then a sufficient number of independent samples will also satisfy RE. However, in our setting, the additional blocking constraint and our desired robustness to sparsity present three significant

162 technical challenges. (A) Sampling: While the above techniques can be used directly in the analysis 163 of Algorithm ESTC, the blocking constraint in our setting entails that sampling of datapoints cannot 164 be independent in the exploration phase. (B) Statistical guarantees: To the best of our knowledge, 165 when the Gram matrix only satisfies RE, the statistical guarantees of the Lasso estimator hold only 166 for hard sparse vectors - for these vectors, all but  $k \ll d$  entries are zero. Existing guarantees for soft sparsity (non-zero tail) hold only for Gram matrices satisfying the stronger RIP condition 167 (Wainwright, 2019). (C) Knowledge of hyper-parameters: Our first proposed algorithm BSLB 168 (similar to ESTC) requires as input a fixed sparsity k (corresponding tail magnitude  $\beta_k$ ). If the input sparsity is too low or too high, then the regret guarantee will not be reasonable. However, it is 170 challenging to set the sparsity parameter without having the knowledge of parameter vector itself. 171 We resolve all technical challenges with novel techniques - in particular, the latter two might be of 172 independent interest in other applications. 173

To resolve (A), we need to optimize for a probability distribution on datapoints (for sampling) with 174 a well-conditioned expected covariance matrix under the constraint that it is uniform on a subset 175 of datapoints and has zero mass on others. The related optimization problem is discrete and non-176 convex - we describe a convex relaxation of the objective and show a randomized rounding proce-177 dure to obtain a feasible solution with good approximation guarantees on the minimum eigenvalue 178 of the expected covariance matrix. Subsequently, we show that subsampling from the recovered 179 distribution in the exploration component ensures that the data covariance matrix satisfies RE with high probability. To resolve (B), we follow two steps 1) We extend guarantees in (Rudelson & Zhou, 181 2013) (Theorem 23) that demonstrates a variant of restricted isometry - the idea is that the data ma-182 trix acts as a near isometry on the image of all sparse vectors under a linear transformation. Such a 183 technique, in turn, allows us to extend RE guarantees for design matrices whose rows are sampled without replacement 2) using the RE condition we derive guarantees for soft sparsity from the re-184 sults on hard sparse vectors using an iterative approach to ensure robustness ((Boche et al., 2015) 185 Theorem 1.6). For (C), 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 187 one. However, a naive application leads to an additional corralling cost with a linear dependence on 188 dimension d, making the regret vacuous (T  $\ll$  d). We use a covering argument - we only choose a 189 small representative subset (covering set) of base algorithms for input to corralling and show that for 190 all remaining base algorithms, their regret is close enough to a base algorithm in the covering set. 191

Related Work: Conceptually our work is similar to active learning (Settles, 2009; Lesci & Vla-192 chos, 2024) where unlabeled samples are annotated adaptively, based on the confidence of a trained 193 model (Coleman et al., 2022). Active learning works well with good initialization and informative 194 confidence intervals. However, in our label-scarce setting, active learning is particularly challenging 195 with complex data due to the absence of a reliably trained model in the first place - this is more pro-196 nounced for difficult datapoints for which prediction is hard. Active Learning (AL) needs an initial 197 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 (AL) are unsuitable (Li et al., 199 2024). Our goal is to identify informative samples with the help of the expert annotator(s), whom 200 the final model aims to emulate. Coreset selection (Guo et al., 2022; Albalak et al., 2024; Sener & Savarese, 2018) aims to select a subset of datapoints for training. However, coreset selection 201 assumes that a large amount of labeled data already exists, and the focus is on reducing compu-202 tational costs. In contrast, our setting deals with the lack of labeled data, making existing coreset 203 selection approaches, which rely on the entire labeled dataset, inapplicable. Our work also aligns 204 with *curriculum learning* (Bengio et al., 2009), where a model is trained on samples of increasing 205 difficulty/complexity. Due to the ambiguity in hardness definition, often heuristics are used to infer 206 the difficulty of samples (Soviany et al., 2022) and can turn out unreliable and not generalizable. 207 For niche tasks where an expert annotator is available, the difficulty ratings from the annotator are 208 more informative since the goal is to train a model to mimic the expert. In computer vision, there 209 has been recent work regarding estimating the difficulty of a dataset for the model using implicit 210 difficulty ratings of annotation (Ionescu et al., 2016; Mayo et al., 2023). For NLP tasks, Ethayarajh 211 et al. (2022) constructs information theoretic metrics to estimate the difficulty of data points.

- 212
- 213
- 214
- 215

216 Algorithm 1 Blocked Sparse Linear Bandits (BSLB) for Efficient Annotation 217 1: Input: Unlabeled datapoints A, Annotation Budget T, Exploration Budget T<sub>explore</sub>, Regularization 218 **Parameter**  $\lambda$ , **Subset selection parameter**  $\hat{g}$ 219 2:  $\mathcal{G} = \text{GetGoodSubset}(\mathcal{A}, \hat{g})$ Compute good subset of datapoints (arms) 3:  $\mathcal{C}=\{\}, \mathcal{R}=\{\}$ Initialize Arm and Reward Set 220 4: for  $t \in [\mathsf{T}_{explore}]$  do Sample randomly  $\mathbf{a}_t \sim \mathcal{G}$  and get difficulty score  $r_t$ 5:  $\triangleright$  Pull arm and get feedback 222  $\mathcal{C} \leftarrow \mathcal{C} \cup \{\mathbf{a}_t\}, \ \mathcal{R} \leftarrow \mathcal{R} \cup \{r_t\}$ Store datapoint/arm and Difficulty Score 6:  $\mathcal{G} = \mathcal{G} \setminus \{\mathbf{a}_t\}$ 7: ▷ Update unlabeled good subset 224 8: end for 9:  $\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} ||\boldsymbol{\theta}||_1 \text{ s.t. } \sum_{t \in [\mathsf{T}_{explore}]} (\mathcal{R}[t] - \langle \boldsymbol{\theta}, \mathcal{C}[t] \rangle)^2 \leq \lambda$ ▷ Compute estimate using LASSO 225 10:  $\mathcal{D} = \mathcal{A} \setminus \mathcal{C}$ > Datapoints (arms) available for exploit phase 226 11: for  $t \in [\mathsf{T}_{explore} + 1, \mathsf{T}]$  do  $\triangleright$  Take Top-(T - T<sub>explore</sub>) difficult samples 227  $\mathbf{a}_t = \arg \max_{\mathbf{a} \in \mathcal{D}} \boldsymbol{\theta}^\mathsf{T} \mathbf{a}$ 12: 228  $\mathcal{C} \leftarrow \mathcal{C} \cup \{\mathbf{a}_t\}, \ \mathcal{D} = \mathcal{D} \setminus \{\mathbf{a}_t\}$ 13: 229 14: end for 230 15: procedure GETGOODSUBSET(Set of Samples  $\mathcal{A}$ , Subset selection parameter  $\hat{q}$ ) 231 **Output:** Sampled Subset  $\mathcal{G}$ 16: 232 17: Maximize the objective function defined in (6) with input  $\hat{g}$  to obtain distribution  $\hat{\mu}$  over  $\mathcal{A}$ . 18: for  $j \in [M]$  do 233  $\mathcal{G} = \mathcal{G} \cup \{\mathbf{a}^{(j)}\}$  with probability  $\hat{g}\hat{\mu}_j$ 19:  $\triangleright$  Add sample j to  $\mathcal{G}$  with prob.  $\hat{g}\hat{\mu}_j$ 234 20: end for 235 21: end procedure 236

### 2 OUR ALGORITHM AND MAIN RESULTS

240 Our main contribution is to propose an Explore-Then-Commit (ETC) algorithm named BSLB which 241 is summarized in Algorithm 1. BSLB takes as input a set of unlabeled datapoints (arms) A, the 242 annotation budget (time horizon) T, the exploration budget  $T_{explore}$  and Subset selection parameter 243  $\hat{g}$ . Steps 2-8 of BSLB correspond to the exploration component in the algorithm. In Step 2, we first 244 compute a good subset of arms  $\mathcal{G} \subset \mathcal{A}$  (using function GETGOODSUBSET $(\mathcal{A}, \hat{g})$ ) which comprises 245 of representative arms that cover the *d*-dimensional space reasonably well. Subsequently, in Steps 246 4-8, we sample arms without replacement from the set of arms  $\mathcal{G}$  for  $T_{explore}$  rounds. The goal in 247 the exploration component is to select a subset of arms such that the image of sparse vectors under 248 the linear transformation by the gram matrix of the selected set has a sufficiently large magnitude 249 (see Definition 1). As we prove, such a result ensures nice statistical guarantees of the parameter 250 (difficulty) estimation with the subset labeled (with difficulty scores) at the end of the exploration component. Since the set of arms, A can be arbitrary, note that sampling arms uniformly at random 251 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 253 following discrete optimization problem 254

256

237 238

239

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

The function GETGOODSUBSET( $\mathcal{A}, \hat{g}$ ) approximates the solution to this computationally infeasible discrete optimization. We maximize a relaxed concave program in equation 6 efficiently for a chosen input parameter  $\hat{g}$  to obtain a distribution  $\hat{\mu}$  on the set of arms  $\mathcal{A}$  - subsequently, we construct the subset  $\mathcal{G}$  using randomized rounding (Step 19) with  $\hat{\mu}$  to obtain a feasible solution to 2.

The second part of BSLB (Steps 9-14) corresponds to the exploitation component of the algorithm. In Step 9, we use the Lasso estimator to get an estimate  $\hat{\theta}$  of the unknown parameter vector  $\theta \in \mathbb{R}^d$ . Note that the number of samples used in obtaining the estimate  $\hat{\theta}$  is much smaller than the ambient dimension *d*. Finally, in Steps 11-14, we choose datapoints that are predicted to be hard according to our recovered estimate  $\hat{\theta}$  and submit them for annotation. 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 twoshot. That is, we change our data acquisition strategy only once after the exploration component thus making our algorithm easy to use in practice. Next, we move on to our main theoretical results.

### 270 2.1 OFFLINE LASSO ESTIMATOR GUARANTEES WITH SOFT SPARSITY AND RE CONDITION

To the best of our knowledge, there do 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 Gram 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 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  $z \neq \mathbf{0}$ ,

- 278 279
- 280

291 292  $0 < K(k_0, \gamma, \mathbf{X}) = \min_{J \subseteq \{1, \dots, d\} | J | \le k_0} \min_{\|z_{J^c}\|_1 \le \gamma \| z_J \|_1} \frac{\|\mathbf{X} z\|_2}{\|z_J\|_2}.$ 

281 Bickel et al. (2009) showed that RE is among the weakest conditions imposed in the literature on the 282 Gram matrix to ensure nice statistical guarantees on the Lasso estimator for sparse linear regression. **Theorem 1.** Let  $\mathbf{X} \in \mathbb{R}^{n \times d}$  be the data matrix with n samples, dimension d. Let  $\mathbf{r} \in \mathbb{R}^n$  be 283 284 the corresponding observations such that  $\mathbf{r} = \mathbf{X} \boldsymbol{\theta} + \boldsymbol{\eta}$ , where  $\boldsymbol{\eta} \in \mathbb{R}^n$  is a zero-mean random vector with i.i.d. components having bounded variance  $\sigma^2 = O(1)$ . Suppose X satisfies restricted 285 eigenvalue property (Def. 1) with  $\mathsf{RE}(k, 4(1 + \gamma_1), \frac{\mathbf{X}}{\sqrt{n}})$  with constant K. Let  $\boldsymbol{\theta}$  have a tail  $\beta_k$  at sparsity level k that is,  $\beta_k := \|\boldsymbol{\theta}_{\mathcal{T}_k^c}\|_1 \leq \gamma_2 \|\boldsymbol{\theta}_{\mathcal{T}_k}\|_1$  for some  $\gamma_2 \in \mathbb{R}$  satisfying  $\gamma_2 \leq \gamma_1$ , where  $\mathcal{T}_k$ 286 287 288 is the set of k largest coordinates by absolute value. An estimate  $\hat{\theta}$  of  $\theta$  recovered using Lasso (Line 289 9 in BSLB), satisfies following with probability  $1 - \exp(-\Omega(n))$ 290

$$\|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\|_{2} = \widetilde{O}\left(n^{-1/2}k^{1/2}K^{-2} + k^{-1/2}\beta_{k} + n^{-1/4}\beta_{k}^{1/2}K^{-1}\right)$$
(3)

293 Due to space constraints, the proof of Theorem 1 is deferred to Appendix A.1.2.

*Insight* 1. Note that in Equation 3, for a fixed sparsity k, the estimator error guarantee decays with datapoints n and RE constant K while growing linearly with the tail  $\beta_k$ . Existing error guarantees in literature focus only on hard sparse  $\theta$  - the data matrix X satisfies RE $(k, 3, \frac{X}{\sqrt{n}})$  with constant K' and  $\gamma_2 = 0$  (see Theorem 7.13 Wainwright (2019)). However, with moderately stronger assumption of RE $(k, 6, \frac{X}{\sqrt{n}})$  on the data matrix, guarantees of Theorem 1 hold for all  $\gamma_2 \le 1/2$ . As stated in Theorem 1, for a larger tail with  $\gamma_2 > 1/2$ , X needs to satisfy RE on a larger cone of vectors.

*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 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 X satisfying RE.

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.

**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  $\Lambda = \lambda_{\min}(|\mathcal{G}|^{-1}\sum_{\mathbf{a}\in\mathcal{G}}\mathbf{a}\mathbf{a}^{\mathsf{T}})$ . Consider the same setup for observations  $\mathbf{r}$  as in Theorem 1. Provided  $n = \Omega(k\Lambda^{-4})$ , an estimate  $\hat{\theta}$  of  $\theta$  recovered using Lasso (Line 9 in BSLB), will satisfy with probability  $1 - \exp(-\Omega(n))$ ,

$$\|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\|_{2} \le \widetilde{\mathsf{O}}\left(n^{-1/2}k^{1/2}\Lambda^{-1} + k^{-1/2}\beta_{k} + n^{-1/4}\beta_{k}^{1/2}\Lambda^{-1/2}\right).$$
(4)

Note in particular that we do not have the  $(\gamma_2 \leq \gamma_1)$  assumption on the parameter vector  $\boldsymbol{\theta}$  in Corollary 1. Instead, it is replaced by a lower bound on n - datapoints sampled without replacement from the set  $\mathcal{G}$  whose gram 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 sufficient (yet smaller than  $|\mathcal{G}|$ ) number of datapoints are sampled from  $\mathcal{G}$ .

320 321

322

312 313

2.2 ONLINE GUARANTEES - REGRET BOUND 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

353

354

368 369 370

component of BSLB. Below, we present the result detailing the regret guarantees of BSLB when the exploration period  $T_{explore}$  is set optimally using a known sparsity level k. The proof is deferred to Appendix A.1.3 We invoke Corollary 1 at the end of the exploration component to obtain error guarantees of Lasso, which in turn bounds the maximum regret incurred in each step of the exploitation component. We optimize the exploration/exploitation trade-off to obtain our stated result.

**Theorem 2.** (*Regret Analysis of BSLB*) Consider the d-dimensional sparse linear bandits framework with blocking constraint having a set  $\mathcal{A} \subset \mathcal{B}^d$  of M arms spanning  $\mathbb{R}^d$  and T rounds ( $T \ll d \ll M$ ). In each round  $t \in [T]$ , 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  $\eta_t$  is zero-mean independent noise random variable with variance  $\sigma^2 = O(1)$ . Suppose  $\boldsymbol{\theta}$  has tail magnitude  $\beta_k := \|\boldsymbol{\theta}_{\mathcal{T}_k^c}\|_1$  at sparsity level k where  $\mathcal{T}_k \subseteq \{1, \ldots, d\}$ is the set of k largest coordinates by absolute value. Let  $\lambda_{\min}^*$  for the set  $\mathcal{A}$  be as defined in equation 2 and assume that  $\lambda_{\min}^* = \Omega(\log^2 M)$ . In this framework, BSLB with exploration period  $T_{explore} = \widetilde{O}(k^{\frac{1}{3}}T^{\frac{2}{3}})$ , achieves a regret guarantee

$$\mathbb{E}[\mathsf{Reg}(\mathsf{T})] = \widetilde{\mathsf{O}}\left(k^{\frac{1}{3}}(\lambda_{\min}^*)^{-1}\mathsf{T}^{\frac{2}{3}} + k^{-\frac{1}{2}}\beta_k + k^{-\frac{1}{12}}(\lambda_{\min}^*)^{-1/2}\beta_k^{\frac{1}{2}}\mathsf{T}^{\frac{5}{6}}\right).$$
(5)

Insight 2. BSLB enables diversity in selected arms by performing Step 2 in Alg. 1 - this step ensures 339 that  $\lambda_{\min}$  of the covariance matrix of the subset used in exploration is approximately optimal. The 340 exploration period in Theorem 2 is optimized to maximize annotation of hard samples. However, in 341 practice, the exploration period of BSLB can be increased further if diversity has more importance. 342 Insight 3. The runtime of the GETGOODSUBSET( $\mathcal{A}, \hat{g}$ ) and the optimization in Step 9 of BSLB 343 (LASSO) is Poly(M, d, T). However, if the mild assumption  $\lambda_{\min}^* = \Omega(\log^2 M)$  not satisfied, then 344 GETGOODSUBSET( $\mathcal{A}, \hat{g}$ ) can be replaced with a (modified) Brute Force Algorithm that runs in time 345  $O(M^d)$  (see Appendix A.1.6 for details), and the theorem statement still holds. Note that the stated 346 runtime in the latter part is still significantly lower than the trivial brute force search for the optimal 347 subset having a runtime of  $O(\exp(M))$ . This is possible because, as a result of the approximation guarantees of Theorem 3, we can restrict the size of the subset while performing a brute-force search. 349

For the case when the true parameter satisfies the hard sparsity condition (the tail  $\beta_k$  is 0), our regret guarantee (see Corollary 2 in Appendix) achieves the same T<sup>2/3</sup> regret dependence as in as Hao et al. (2020) without the *blocking constraint*.

### 2.3 SUBSET SELECTION FOR MAXIMIZING THE MINIMUM EIGENVALUE

355 Recall that in Step 5 of BSLB; we sample from a carefully chosen subset of arms that has good 356 coverage - more precisely, our goal is to solve the optimization problem in Eq. 2 to obtain a repre-357 sentative set of arms. Although Hao et al. (2020) had a similar objective, the absence of blocking 358 constraint in their framework implied that they could solve for a probability distribution on the set of 359 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 360 that a solution can be found efficiently. However, in our setting, due to the blocking constraint, we 361 need to identify a subset of representative arms from which to sample uniformly at random without 362 replacement in the exploration component - this leads to the objective in Eq. 2 being discrete and therefore non-convex. Note that a brute force solution to our objective implies a search over all 364 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-convex objective in 2, our first step is to obtain a convex relaxation as 366 described in Eq. 6 - in particular, instead of optimizing over a subset, we optimize over probability 367 distributions over the set of arms such that the probability mass over any arm is bounded from above.

$$\hat{\boldsymbol{\mu}}(\hat{g}) = \underset{\boldsymbol{\mu}\in\mathcal{P}(\mathcal{A})}{\operatorname{arg\,max}} \lambda_{\min} \left( \mathbf{A}\mathsf{diag}(\boldsymbol{\mu})\mathbf{A}^{\mathsf{T}} \right) \text{ such that } \|\boldsymbol{\mu}\|_{\infty} \leq \frac{1}{\hat{g}}, \tag{6}$$

Note that  $\mathbf{A} = [\mathbf{a}^1, \dots, \mathbf{a}^{\mathsf{M}}]^{\mathsf{T}} \in \mathbb{R}^{\mathsf{M} \times d}$  denotes the matrix with all arms and  $\hat{g}$  is an additional parameter to the relaxed objective. Since the solution to Eq. 6 might not be a feasible one for Eq. we use a randomized rounding (Step 19 in BSLB) procedure to obtain a feasible solution. In the randomized rounding procedure, each datapoint  $j \in [\mathsf{M}]$  is sampled into our feasible output set  $\mathcal{G}$ (used in the exploration component) independently with probability  $\hat{g}\mu_j$ .

**Let**  $\mathcal{X} \subseteq \mathcal{A}$  with  $g^* = |\mathcal{X}|$  be the optimal subset for which the RHS in Equation 2 is maximized and let  $\lambda_{\min}^*$  be the corresponding objective value (minimum eigenvalue). 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 gram matrix associated with datapoints in  $\mathcal{G}$  (obtained post randomized rounding procedure) is close to  $\lambda_{\min}^*$ .

Theorem 3. Let  $\mathbf{A} = [\mathbf{a}^1, \dots, \mathbf{a}^{\mathsf{M}}]^\mathsf{T} \in \mathbb{R}^{\mathsf{M} \times d}$  denote the matrix of all arms. Consider the convex optimization of equation 6 solved at  $\hat{g} = \mathsf{O}(d)$ . Let  $\mathcal{G}$  be the output of the randomized rounding procedure (Step 18-20 of Alg. 1) and  $\hat{\lambda}_{\min}$  be the minimum eigenvalue of the corresponding covariance matrix that is,  $\hat{\lambda}_{\min} = \lambda_{\min}(|\mathcal{G}^{-1}| \sum_{\mathbf{a} \in \mathcal{G}} \mathbf{a} \mathbf{a}^\mathsf{T})$ . Then under the assumption  $\lambda^*_{\min} = \Omega((\log \mathsf{M})^2)$ , we must have  $\lambda^*_{\min} \leq 2\hat{\lambda}_{\min}(\log \mathsf{M})^2$  with probability 1 - o(1).

If assumption of  $\lambda_{\min}^* = \Omega((\log M)^2)$  is not satisfied in its place., then we can implement a bruteforce search over all subsets whose size is in the range  $[d, \alpha d]$  (for some constant  $\alpha \ge 1$ ) to maximize the objective in Eq. 2 - note that the time complexity is still polynomial in the number of arms  $M^{O(d)}$ (refer to Appendix A.1.6 for details) which is significantly improved than the trivial brute force algorithm which has a running time of  $O(\exp M)$ . Note that the modified brute-force algorithm enjoys stronger approximation guarantees, with  $\frac{\lambda_{\min}^*}{2} \le \widehat{\lambda}_{\min}$  and has a running time that is still polynomial in the number of arms M but exponential in the dimension d.

*Remark* 2. We want to highlight that several existing techniques in experimental design deal with
 maximizing objectives such as minimum eigenvalue; however, existing work assumes submodularity
 or matroid constraints (Allen-Zhu et al., 2021), which the average minimum eigenvalue (normalized
 with size of the set) in our setting does not satisfy as discussed in Appendix A.1.7.

397 *Proof Outline:* We first show in Lemma 4 (using concentration guarantees) that the following two 398 objective values are close namely (A) value of the maximized concave objective with distribution 399  $\hat{\mu} \in \mathcal{P}(\mathcal{A})$  and parameter  $\hat{g}$  in Equation 6 (B) objective value of the set  $\mathcal{G}$  (eq. 2) obtained via 400 randomized rounding procedure from  $\hat{\mu}$  at  $\hat{g}$  (line 19 in BSLB). Note that the value of the maxi-401 mized concave objective in Equation 6 with parameter  $g_1$  is larger than the value with parameter  $g_2$ 402 provided  $g_1 \leq g_2$ . Therefore, we show our approximation guarantees with respect to objective in 403 Equation 6 with parameter d - which in turn also translates into guarantees for the optimal param-404 eter  $g^*$  even though  $g^*$  is unknown (since  $g^* \ge d$ ). Finally, given that the concave objective with parameter d in Equation 6 is a relaxation of the discrete objective in Equation 2, the objective value 405 of the former is going to be larger than the objective value of the latter. Combining all these key 406 407 ingredients, we have proved our theorem statement.

408 409

#### 2.4 CORRALLING WHEN OPTIMAL SPARSITY LEVEL IS NOT KNOWN

410 Note that for any unknown parameter vector  $\theta$ , we can fix the sparsity level k and therefore the 411 corresponding tail magnitude  $\beta_k$  - subsequently, we can obtain the guarantees of Theorem 2 by 412 setting the exploration period optimally for the fixed k. However, if k is set too low, then  $\beta_k$  will be 413 too high, and therefore, the second term in regret (Equation 5) dominates. On the other hand, if k is 414 set too high, then  $\beta_k$  is low but the first term in the regret bound dominates. There is a trade-off, and 415 therefore, there is an optimal choice of sparsity  $k^*$  and, tail magnitude  $\beta_{k^*}$ . Therefore, we propose 416 a meta-algorithm C-BSLB that exploits coralling (Agarwal et al., 2017) multiple versions of the 417 BSLB algorithm 1 with different values of k used to set the exploration period  $T_{explore}$  - the metaalgorithm gradually learns to choose the best base algorithm. However, naively applying CORRAL 418 with all distinct base algorithms leads to a linear dependence on dimension d in the regret making 419 it vacuous. Therefore we carefully choose  $\log d$  base algorithms for search within CORRAL with 420 corresponding sparsity parameters set on exponentially spaced points - such a restriction ensures 421 that the overhead in regret is minimal (logarithmic dependence on dimension d). However, we still 422 prove our regret guarantee with respect to the base algorithm with optimal sparsity - although it is 423 not guaranteed that the optimal base algorithm will be in the set of carefully chosen base algorithms 424 provided as input to the meta-algorithm.

**Theorem 4.** Consider the d-dimensional sparse linear bandits framework with blocking constraint as described in Theorem 2. Let the C-BSLB algorithm (Algorithm 3 in Appendix A.1.9) 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 2 that achieves minimum regret be  $k^* \in \{1, 2, ..., d-1, d\}$ , and let  $\mathbb{E}[\operatorname{Reg}(\mathsf{T})]_*$  be the corresponding regret. Then the meta-algorithm C-BSLB achieves the following regret guarantee,

$$\mathbb{E}[\operatorname{\mathsf{Reg}}(\mathsf{T})] = O(\sqrt{\mathsf{T}\log_2(d)} + \sqrt{k^*}\log_2(d)\mathbb{E}[\operatorname{\mathsf{Reg}}(\mathsf{T})]_*). \tag{7}$$

<sup>432</sup> Note that the first term in Equation 7 and the multiplicative factor of  $\sqrt{k^*} \log_2(d)$  corresponds to the additional cost in combining the input base algorithms by Algorithm 3. We stress that the dependence on dimension *d* from the additional cost is only logarithmic.

Proof Outline: We use Theorem 5 of (Agarwal et al., 2017) and our regret bound from Theorem 2 to obtain Theorem 4. The key novelty in our proof 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.

442 443

444

### 3 EXPERIMENTS

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.2.3. 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.2.1. Finally, we provide detailed simulations in Appendix A.2.4.

We consider the setting where we have a total of M unlabelled samples (with  $T \ll M$ ) and only T datapoints can be annotated (sequentially). For each unlabeled data point 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 data point. We showcase the effectiveness of BSLB (Alg. 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

458 2. All: All the samples in training data (except the validation fold)

3. AnchorAL (Lesci & Vlachos, 2024): an anchoring based active learning baseline (T samples).

460 4. SEALS (Coleman et al., 2022): a KNN based sub-sampling active learning baseline (T samples).

AnchorAL and SEALS are state-of-the-art active learning (AL) algorithms. In general, for a label-466 scarce complex task, AL might not be immediately applicable (see cold-start problem in Li et al. 467 (2024)) - especially for datapoints close to the decision boundary with noisy/less informative confi-468 dence intervals. This is because AL requires a reliably trained model on an even smaller subset of 469 labeled datapoints - however, on datapoints far from the decision boundary (easy datapoints), noisy 470 confidence signals are still useful. As we show in our experiment, this intuition holds, and the AL 471 models, along with the random baseline, perform well on the easy-valid dataset. It is worth noting 472 that complex (hard) datapoints often tend to be the main challenge in industrial applications. This 473 is because it is *easy* to improve performance on easy data (cheaper to obtain) by simply increasing 474 samples during training, but hard datapoints are difficult to generalize on (Pukowski & Lu, 2024).

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 pretrained on ImageNet-21k dataset (Wu et al., 2020). We present results on the classification task

<sup>&</sup>lt;sup>2</sup>We consider the AL setup initialized with  $T_{explore}$  samples and  $T - T_{explore}$  samples queried in a batch.

6	Validation	Object	AmahanAI	CEALC	Dandam	A 11	Our (BSLB)	
	Туре	Annotated	Anchoral	SEALS	Kandolli	All		
		chair	$94.0\pm1.67$	$90.6 \pm 1.8$	$96.4 \pm 1.0$	96.0±1.1	94.6 ±1.6	
	easy-valid	car	$94.5 \pm 1.6$	$94.7 \pm 4.0$	$97.7 \pm 2.1$	$98.7 \pm 0.1$	$96.5 \pm 1.8$	
		bottle	$93.0{\pm}2.5$	$92.8 {\pm} 2.3$	$96.8 {\pm} 1.1$	96.8±1.1	$94.8{\pm}2.0$	
		bottle or chair	$91.5 \pm 1.1$	$92.3 {\pm} 1.1$	$94.8 {\pm} 0.97$	94.6±2.1	$91.7 \pm 2.2$	
		chair	$69.3 \pm 3.1$	$69.6 \pm 6.1$	$66.0\pm3.8$	$71.3 \pm 3.2$	73.3±3.3	
	hard-valid	car	$70.3 \pm 4.0$	$70.0\pm5.7$	$60.0\pm5.4$	$65.4 \pm 4.0$	$\textbf{74.0}{\pm}\textbf{ 3.4}$	
		bottle	63.1±2.9	$63.4 \pm 3.4$	$59.7 \pm 4.4$	$64.8 {\pm} 1.9$	66.8±2.6	
		bottle or chair	67.1±3.5	$66.3 \pm 1.4$	$68.0{\pm}4.0$	$72.3{\pm}2.0$	73.0±1.7	

496Table 1: Test accuracy of model  $\mathcal{M}$  trained on different subsets of data annotated for 4 distinct object497detection tasks in an image (PASCAL-VOC): The test performance of BSLB approach on the easy498and hard validation dataset is at par with the  $\mathcal{M}$  trained on all samples. We perform significantly499better on the hard validation dataset compared to random sampling and active learning baselines.

- 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  $\tau_{easy} = 3.1$  and  $\tau_{hard} = 3.9$  since the distribution of the difficulty scores in the dataset is heavy-tailed as shown in Figure 4b. The (image, question) tuple with difficulty scores in the range [3.1, 3.8] are highly noisy and therefore have been excluded. Table 2 contains the hyperparameters T,  $T_{explore} (\approx 0.6T)$  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 3 contains results on the effect of varying  $T_{explore}$ .

510 We present our results in Table 1 averaged over 5 validation folds. For this classification task, our 511 method (BSLB) efficiently selects datapoints (to be annotated) compared to baselines with an equal 512 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 513 an improvement of 5 - 14% on the hard validation data **hard-valid**. When compared to the active 514 learning baselines (AnchorAL and SEALS), BSLB performs better by 1-4% on easy-valid and by 515 3.5-7% on **hard-valid**. Finally, when compared to  $\mathcal{M}$  trained on all datapoints (all baseline), which 516 has  $6 \times$  to  $12 \times$  more samples, our method does better (0.7% to 8.6%) on the **hard-valid** and does 517 decently on **easy-valid** (< 3% difference). These results validate our theory - in particular, we find 518 that performance on **easy-valid** improves if the model  $\mathcal{M}$  is trained on more samples (randomly 519 chosen to improve coverage). However, improving the performance on hard-valid dataset is the 520 main challenge where our simple approach BSLB with theoretical guarantees does reasonably well. 521

522

### 4 CONCLUSION AND FUTURE WORK

523 524

This work formulates the problem of adaptive annotation with expert-provided difficulty feedback 526 as regret minimization in blocked sparse linear bandits. The goal is to annotate difficult samples to 527 improve generalization on hard datapoints, assuming difficulty ratings are linked to sample features 528 by an unknown parameter. We consider the practical consideration when each sample can be anno-529 tated once and only a few rounds of annotations are available. We propose an explore-then-commit 530 algorithm BSLB, a two-shot algorithm for unlabelled data-point selection (for annotation). We show 531 theoretical results establishing the sub-linear regret of the proposed BSLB algorithm - we also ensure that our guarantees work with minimum assumptions on datapoints and are robust to sparsity. 532 Finally, we present a meta-algorithm C-BSLB that corrals BSLB algorithms with different input 533 parameters - the result is a bandit algorithm that suffers the same regret orders as BSLB without the 534 knowledge of the optimal sparsity hyper-parameter. Numerical studies on real-life image and text 535 datasets show the efficacy of our methods in the label-scarce regime. We demonstrate experiments in 536 the appendix for recommendation and simulation setup using our theoretical framework and BSLB. 537

Future Work: Our experimental results are done on classifying state-of-the-art embeddings; how ever, more extensive experimentation can be considered in future work, especially for creating datasets used for generative tasks.

### 540 REFERENCES

551

552

553

563

564

565 566

567

568

579

583

584

- Alekh Agarwal, Haipeng Luo, Behnam Neyshabur, and Robert E Schapire. Corralling a band of
   bandit algorithms. In *Conference on Learning Theory*, pp. 12–38. PMLR, 2017.
- Alon Albalak, Yanai Elazar, Sang Michael Xie, Shayne Longpre, Nathan Lambert, Xinyi Wang, Niklas Muennighoff, Bairu Hou, Liangming Pan, Haewon Jeong, et al. A survey on data selection for language models. *arXiv preprint arXiv:2402.16827*, 2024.
- Zeyuan Allen-Zhu, Yuanzhi Li, Aarti Singh, and Yining Wang. Near-optimal discrete optimization
  for experimental design: a regret minimization approach. *Math. Program.*, 186(1-2):439–478,
  March 2021.
  - Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*, pp. 41–48, 2009.
- Peter J. Bickel, Ya'acov Ritov, and Alexandre B. Tsybakov. Simultaneous analysis of Lasso and Dantzig selector. *The Annals of Statistics*, 37(4):1705 1732, 2009. doi: 10.1214/08-AOS620.
  URL https://doi.org/10.1214/08-AOS620.
- Holger Boche, Robert Calderbank, Gitta Kutyniok, and Jan Vybíral. A survey of compressed sensing. In *Compressed Sensing and its Applications: MATHEON Workshop 2013*, pp. 1–39. Springer, 2015.
- Guy Bresler, George H Chen, and Devavrat Shah. A latent source model for online collaborative
   filtering. Advances in neural information processing systems, 27, 2014.
  - Adam Brown, Aditi Laddha, and Mohit Singh. Maximizing the minimum eigenvalue in constant dimension, 2024. URL https://arxiv.org/abs/2401.14317.
  - Scott Shaobing Chen, David L Donoho, and Michael A Saunders. Atomic decomposition by basis pursuit. *SIAM review*, 43(1):129–159, 2001.
- Cody Coleman, Edward Chou, Julian Katz-Samuels, Sean Culatana, Peter Bailis, Alexander C.
   Berg, Robert Nowak, Roshan Sumbaly, Matei Zaharia, and I. Zeki Yalniz. Similarity search for efficient active learning and search of rare concepts. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(6):6402–6410, Jun. 2022. doi: 10.1609/aaai.v36i6.20591. URL https://ojs.aaai.org/index.php/AAAI/article/view/20591.
- Kawin Ethayarajh, Yejin Choi, and Swabha Swayamdipta. Understanding dataset difficulty with *V*-usable information. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari,
  Gang Niu, and Sivan Sabato (eds.), *PMLR*, volume 162 of *Proceedings of Machine Learning Research*, pp. 5988–6008, 17–23 Jul 2022. URL https://proceedings.mlr.press/
  v162/ethayarajh22a.html.
- M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision*, 111(1):98–136, January 2015.
  - Ken Goldberg, Theresa Roeder, Dhruv Gupta, and Chris Perkins. Eigentaste: A constant time collaborative filtering algorithm. *Inf. Retr. Boston.*, 4(2):133–151, 2001.
- Chengcheng Guo, Bo Zhao, and Yanbing Bai. Deepcore: A comprehensive library for coreset selection in deep learning. In *International Conference on Database and Expert Systems Applications*, pp. 181–195. Springer, 2022.
- Botao Hao, Tor Lattimore, and Mengdi Wang. High-dimensional sparse linear bandits. Advances in Neural Information Processing Systems, 33:10753–10763, 2020.
- F. Maxwell Harper and Joseph A. Konstan. The movielens datasets: History and context. ACM
   Trans. Interact. Intell. Syst., 5(4), dec 2015. ISSN 2160-6455. doi: 10.1145/2827872. URL
   https://doi.org/10.1145/2827872.

594 595 596 597 598 599 600 601	Michael A. Hedderich, Lukas Lange, Heike Adel, Jannik Strötgen, and Dietrich Klakow. A survey on recent approaches for natural language processing in low-resource scenarios. In Kristina Toutanova, Anna Rumshisky, Luke Zettlemoyer, Dilek Hakkani-Tur, Iz Beltagy, Steven Bethard, Ryan Cotterell, Tanmoy Chakraborty, and Yichao Zhou (eds.), <i>Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies</i> , pp. 2545–2568, Online, June 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.naacl-main.201. URL https://aclanthology.org/2021.naacl-main.201.
602 603 604 605	Radu Tudor Ionescu, Bogdan Alexe, Marius Leordeanu, Marius Popescu, Dim Papadopoulos, and Vittorio Ferrari. How hard can it be? Estimating the difficulty of visual search in an image. In <i>Proceedings of CVPR</i> , pp. 2157–2166, June 2016.
606 607 608	Kyoungseok Jang, Chicheng Zhang, and Kwang-Sung Jun. Popart: Efficient sparse regression and experimental design for optimal sparse linear bandits. <i>Advances in Neural Information Processing Systems</i> , 35:2102–2114, 2022.
609 610 611 612 613 614 615	Pietro Lesci and Andreas Vlachos. AnchorAL: Computationally efficient active learning for large and imbalanced datasets. In Kevin Duh, Helena Gomez, and Steven Bethard (eds.), <i>Proceedings</i> of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), pp. 8445–8464, Mexico City, Mexico, June 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024. naacl-long.467. URL https://aclanthology.org/2024.naacl-long.467.
616 617 618	Dongyuan Li, Zhen Wang, Yankai Chen, Renhe Jiang, Weiping Ding, and Manabu Okumura. A survey on deep active learning: Recent advances and new frontiers. <i>IEEE Transactions on Neural Networks and Learning Systems</i> , 2024.
619 620 621	Ming Li, Yong Zhang, Zhitao Li, Jiuhai Chen, Lichang Chen, Ning Cheng, Jianzong Wang, Tianyi Zhou, and Jing Xiao. From quantity to quality: Boosting llm performance with self-guided data selection for instruction tuning. <i>arXiv preprint arXiv:2308.12032</i> , 2023.
622 623 624	Adyasha Maharana, Prateek Yadav, and Mohit Bansal. D2 pruning: Message passing for balancing diversity and difficulty in data pruning. <i>arXiv preprint arXiv:2310.07931</i> , 2023.
625 626 627 628 629 630 631 632	David Mayo, Jesse Cummings, Xinyu Lin, Dan Gutfreund, Boris Katz, and Andrei Barbu. How hard are computer vision datasets? calibrating dataset difficulty to viewing time. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), Advances in Neural Information Processing Systems, volume 36, pp. 11008–11036. Curran Associates, Inc., 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/ 2432e9646556f3a98dc78c1f4a10481b-Paper-Datasets_and_Benchmarks. pdf.
633 634 635 636	Sören Mindermann, Jan M Brauner, Muhammed T Razzak, Mrinank Sharma, Andreas Kirsch, Win- nie Xu, Benedikt Höltgen, Aidan N Gomez, Adrien Morisot, Sebastian Farquhar, et al. Prioritized training on points that are learnable, worth learning, and not yet learnt. In <i>International Confer- ence on Machine Learning</i> , pp. 15630–15649. PMLR, 2022.
637 638 639 640	Ngoc Dang Nguyen, Wei Tan, Lan Du, Wray Buntine, Richard Beare, and Changyou Chen. Auc maximization for low-resource named entity recognition. In <i>Proceedings of the AAAI Conference on Artificial Intelligence</i> , volume 37, pp. 13389–13399, 2023.
641 642 643	Soumyabrata Pal, Arun Suggala, Karthikeyan Shanmugam, and Prateek Jain. Blocked collabora- tive bandits: online collaborative filtering with per-item budget constraints. <i>Advances in Neural</i> <i>Information Processing Systems</i> , 36, 2024.
644 645 646	Pawel Pukowski and Haiping Lu. Investigating the impact of hard samples on accuracy reveals in-class data imbalance. <i>arXiv preprint arXiv:2409.14401</i> , 2024.
647	Mark Rudelson and Shuheng Zhou. Reconstruction from anisotropic random measurements. <i>IEEE Transactions on Information Theory</i> , 59(6):3434–3447, 2013. doi: 10.1109/TIT.2013.2243201.

648 649 650 651	Ozan Sener and Silvio Savarese. Active learning for convolutional neural networks: A core-set approach. In <i>International Conference on Learning Representations</i> , 2018. URL https://openreview.net/forum?id=HlaIuk-RW.
652	Burr Settles. Active learning literature survey. Computer Sciences Technical Report 1648, 2009.
653 654 655 656 657 658	Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D. Manning, Andrew Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In <i>Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing</i> , pp. 1631–1642, Seattle, Washington, USA, October 2013. Association for Computational Linguistics. URL https://www.aclweb.org/anthology/D13-1170.
659 660 661	Ben Sorscher, Robert Geirhos, Shashank Shekhar, Surya Ganguli, and Ari Morcos. Beyond neural scaling laws: beating power law scaling via data pruning. <i>Advances in Neural Information Processing Systems</i> , 35:19523–19536, 2022.
663 664 665 666	Petru Soviany, Radu Tudor Ionescu, Paolo Rota, and Nicu Sebe. Curriculum Learning: A Survey. International Journal of Computer Vision, 130(6):1526–1565, June 2022. ISSN 1573-1405. doi: 10.1007/s11263-022-01611-x. URL https://doi.org/10.1007/ s11263-022-01611-x.
667 668 669 670	I. O. Tolstikhin. Concentration inequalities for samples without replacement. <i>Theory of Probability</i> & <i>Its Applications</i> , 61(3):462–481, 2017. doi: 10.1137/S0040585X97T988277. URL https://doi.org/10.1137/S0040585X97T988277.
671 672	Vasilis Verroios and Hector Garcia-Molina. Entity resolution with crowd errors. In 2015 IEEE 31st international conference on data engineering, pp. 219–230. IEEE, 2015.
674 675 676	Roman Vershynin. <i>High-Dimensional Probability: An Introduction with Applications in Data Science</i> . Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2018.
677 678 679	Martin J. Wainwright. <i>High-Dimensional Statistics: A Non-Asymptotic Viewpoint</i> . Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2019.
680 681 682	Bichen Wu, Chenfeng Xu, Xiaoliang Dai, Alvin Wan, Peizhao Zhang, Zhicheng Yan, Masayoshi Tomizuka, Joseph Gonzalez, Kurt Keutzer, and Peter Vajda. Visual transformers: Token-based image representation and processing for computer vision, 2020.
683 684	Zygmunt Zajac. Goodbooks-10k: a new dataset for book recommendations. FastML, 2017.
685 686 687 688 689 690	Xuan Zhang, Gaurav Kumar, Huda Khayrallah, Kenton Murray, Jeremy Gwinnup, Marianna J. Martindale, Paul McNamee, Kevin Duh, and Marine Carpuat. An empirical exploration of curriculum learning for neural machine translation. <i>CoRR</i> , abs/1811.00739, 2018. URL http://arxiv.org/abs/1811.00739.
691 692	A APPENDIX
693 694 695	The Appendix comprises two sections: Section A.1 details the detailed proofs of the technical results and Section A.2 presents additional numerical results.
696 697	A.1 BRIEF OVERVIEW AND TECHNICAL PROOFS
698 699	A.1.1 LOWER BOUND FOR REGRET
700	We provide the following theorem on the lower bound of regret for the high-dimensional linear

We provide the following theorem on the lower bound of regret for the high-dimensional linear
 bandit setting with blocking constraint. The following lower bound is derived for a parameter vector with hard sparsity but trivially extends to a parameter vector with soft sparsity.

**Theorem 5.** 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  $\mathsf{T}$  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  $\eta_t$  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 lower bounded as follows,

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

*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  $\mathcal{A} = \{a^{(1)}, \ldots, a^{(M)}\}$  we construct the arm multiset  $\mathcal{A}'$  as follows,
  - $\mathcal{A}' = \cup_{i=1}^{\mathsf{M}} \cup_{i=1}^{\mathsf{T}} \{a_{i}^{(i)}\}$

where  $a_j^{(i)}$  denotes the  $j^{\text{th}}$  copy of the  $i^{\text{th}}$  arm (such that it is the different arm with the same arm vector). 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 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

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

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  $d \ge k^{1/3} \mathsf{T}^{2/3}$ , which is the regime considered in the paper, this bound reduces to  $\Omega(k^{1/3} \mathsf{T}^{2/3})$ , which is the order that our upper bound achieves.

### 741 A.1.2 PROOF OF THEOREM 1

<sup>742</sup> We defined the restricted eigenvalue of a matrix as,

**Definition 2.** *Restricted Eigenvalue:* If **X** satisfies Restricted Eigenvalue property  $\mathsf{RE}(k_0, \gamma, \mathbf{X})$ , 744 then there exists a constant  $K(k_0, \gamma, \mathbf{X})$  such that for all  $z \in \mathbb{R}^d$  and  $z \neq \mathbf{0}$ , 

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

This definition implies,

$$K(k_0, \gamma, \mathbf{X}) \| z_J \|_2 \le \| \mathbf{X} z \|_2 \, \forall \| z_{J^c} \|_1 \le \gamma \| z_J \|_1 \, \forall J \subseteq \{1, \dots, d\}, |J| \le k_0$$

We prove the theorem by using the Basis Pursuit Program which is one of three equivalent formula tion for LASSO Chen et al. (2001); Wainwright (2019),

753  
754  
755  

$$\theta = \arg \min_{\theta} ||\theta||_1$$
  
s.t.  $\sum_{t \in [\mathsf{T}_{explore}]} (\mathcal{R}[t] - \langle \theta, \mathcal{C}[t] \rangle)^2 \le \lambda$ 
(8)

<sup>756</sup> <sup>757</sup> <sup>758</sup> *Proof.* For the purpose of the proof denote the true parameter by  $\theta^*$ . The following inequality holds for a general basis pursuit program (w/o any assumptions on sparsity) Wainwright (2019),

$$\frac{1}{n} \|\mathbf{X}\mathbf{h}\|_{2}^{2} \le 2\frac{\mathbf{w}^{\mathsf{T}}\mathbf{X}\mathbf{h}}{n} + 2(b^{2} - \frac{\|\mathbf{w}\|_{2}^{2}}{2n})$$
(9)

where  $\mathbf{h} = \hat{\theta} - \theta^*$ ,  $\mathbf{w} = \mathcal{R} - \mathbf{X}\theta^*$ ,  $b^2$  is the tolerance and *n* is the number of samples. For the proof, the tolerance is equal to the regularization parameter  $b^2 = \lambda$ .

Let  $A = 2(b^2 - \frac{\|\mathbf{w}\|_2^2}{2n}) B = \|\frac{\mathbf{X}\mathbf{w}}{n}\|_{\infty}$  and  $\alpha = \|\mathbf{h}_{(\mathcal{T}_0 \cup \mathcal{T}_1)}\|_2$ , where  $\mathcal{T}_0$  denotes the top-k coordinates,  $\mathcal{T}_1$  the next top k coordinates and so on, for this proof.

First note that,

$$\|\mathbf{h}_{(\mathcal{T}_{0}\cup\mathcal{T}_{1})^{c}}\|_{1} = \sum_{m\geq 2} \|\mathbf{h}_{\mathcal{T}_{m}}\|_{1} \leq \sum_{m\geq 2} k \max_{j\in\mathcal{T}_{m}} \mathbf{h}_{j} \leq \sum_{m\geq 2} k \min_{j\in\mathcal{T}_{m-1}} \mathbf{h}_{j} \leq \sum_{m\geq 1} \|\mathbf{h}_{\mathcal{T}_{m}}\|_{1} \leq \|\mathbf{h}_{(\mathcal{T}_{0})^{c}}\|_{1}$$

And from the decomposition available in Theorem 1.6 in Boche et al. (2015) we know that

 $\|\mathbf{h}_{(\mathcal{T}_0)^c}\|_1 \le \|\mathbf{h}_{(\mathcal{T}_0)}\|_1 + 2\beta_k$ 

773774 We also the following inequality,

$$\|\boldsymbol{\theta}_{\mathcal{T}_0}^*\|_1 = \|\boldsymbol{\theta}_{\mathcal{T}_0}^*\|_1 - \|\widehat{\boldsymbol{\theta}}_{\mathcal{T}_0}\|_1 + \|\widehat{\boldsymbol{\theta}}_{\mathcal{T}_0}\|_1 \le \|\boldsymbol{\theta}_{\mathcal{T}_0}^* - \widehat{\boldsymbol{\theta}}_{\mathcal{T}_0}\|_1 + \|\widehat{\boldsymbol{\theta}}_{\mathcal{T}_0}\|_1 = \|\mathbf{h}_{\mathcal{T}_0}\|_1 + \|\widehat{\boldsymbol{\theta}}_{\mathcal{T}_0}\|_1 \le 2\|\mathbf{h}_{\mathcal{T}_0}\|_1$$

The last inequality holds since  $\hat{\theta}$  is a solution of equation 8 and  $\|\hat{\theta}_{\tau_0}\|_1 \leq \|\hat{\theta}_{\tau_0}\|_1 \leq \|\mathbf{h}_{\tau_0}\|_1$ otherwise  $\mathbf{h}_{\tau_0}$  would be the solution instead.

Therefore,

$$\begin{aligned} \|\mathbf{h}_{\mathcal{T}_{0}^{C}}\|_{1} &\leq \|\mathbf{h}_{\mathcal{T}_{1}}\|_{1} + \|\mathbf{h}_{(\mathcal{T}_{0}\cup\mathcal{T}_{1})^{C}}\|_{1} = \|\mathbf{h}_{\mathcal{T}_{1}}\|_{1} + \|\mathbf{h}_{(\mathcal{T}_{0}\cup\mathcal{T}_{1})}\|_{1} + 2\beta_{k} \\ &\leq \|\mathbf{h}_{\mathcal{T}_{1}}\|_{1} + \|\mathbf{h}_{(\mathcal{T}_{0}\cup\mathcal{T}_{1})}\|_{1} + 2\gamma \|\boldsymbol{\theta}_{\mathcal{T}_{0}}^{*}\|_{1} \leq 2\|\mathbf{h}_{\mathcal{T}_{0}}\|_{1} + 2\|\mathbf{h}_{\mathcal{T}_{0}}\|_{1} + 4\gamma \|\mathbf{h}_{\mathcal{T}_{0}}\|_{1} \leq 4(1+\gamma)\|\mathbf{h}_{\mathcal{T}_{0}}\|_{1} \end{aligned}$$

Now by the RE condition,

$$\frac{\|\mathbf{X}\mathbf{h}\|_2^2}{n} \ge K^2(k_0, 4+4\gamma, \frac{\mathbf{X}}{\sqrt{n}}) \|\mathbf{h}_{\mathcal{T}_0 \cup \mathcal{T}_1}\|_2^2$$

790 Putting this into equation 9,

$$K^{2} \|\mathbf{h}_{\mathcal{T}_{0}}\|_{2}^{2} \leq 2 \|\mathbf{h}\|_{1} \|\frac{\mathbf{X}^{\mathsf{T}} \mathbf{w}}{n}\|_{\infty} + 2(b^{2} - \frac{\|\mathbf{w}\|_{2}^{2}}{2n}) \leq B \|\mathbf{h}\|_{1} + A$$

We now bound  $\|\mathbf{h}\|_1$ 

$$\|\mathbf{h}\|_{1} = \|\mathbf{h}_{(\mathcal{T}_{0})}\|_{1} + \|\mathbf{h}_{(\mathcal{T}_{0})^{C}}\|_{1}$$
$$\|\mathbf{h}_{(\mathcal{T}_{0})}\|_{1} \le \sqrt{k} \|\mathbf{h}_{(\mathcal{T}_{0})}\|_{2}$$

And the bound on  $\|\mathbf{h}_{(\mathcal{T}_0)^C}\|_1$  is,

$$\|\mathbf{h}_{(\mathcal{T}_{0})^{C}}\|_{1} \leq \|\mathbf{h}_{(\mathcal{T}_{1})}\|_{1} + \|\mathbf{h}_{(\mathcal{T}_{0}\cup\mathcal{T}_{1})^{c}}\|_{1} \leq \|\mathbf{h}_{(\mathcal{T}_{1})}\|_{1} + \|\mathbf{h}_{(\mathcal{T}_{0}\cup\mathcal{T}_{1})}\|_{1} + 2\beta_{k}(\boldsymbol{\theta}^{*})_{1} \leq 4\sqrt{k}\|\mathbf{h}_{(\mathcal{T}_{0})}\|_{2} + 2\beta_{k}(\boldsymbol{\theta}^{*})_{1}$$

Combining both bounds we find the bound on  $\|\mathbf{h}\|_1$ ,

 $\|\mathbf{h}\|_1 \le 5\sqrt{k}\|\mathbf{h}_{(\mathcal{T}_0)}\|_2 + 2\beta_k(\boldsymbol{\theta}^*)_1 = C\alpha + 2\beta$ 

where we make the following substitution  $(C = 5\sqrt{k}, \beta = \beta_k(\theta^*)_1)$ 

$$K^2 \alpha^2 \leq B(C\alpha + 2\beta) + A \implies K^2 \alpha^2 \leq (BC)\alpha + A + 2B\beta$$

810 Let 
$$F = K^2(k_0, 4 + 4\gamma, \frac{1}{\sqrt{n}}\mathbf{X}),$$

813 814

819

820 821

This is true when,

$$\alpha \le 2\frac{BC}{F} + \sqrt{\frac{A + 2B\beta}{F}}$$

 $\alpha^2 \le \frac{BC}{F}\alpha + \frac{A + 2B\beta}{F}$ 

Now applying concentration using the properties of a bounded zero-mean noise,  $B \leq C_1 \sigma(\sqrt{\frac{2 \log d}{n}} + \epsilon)$  w.h.p. and  $A \leq \sigma^2 \epsilon$  w.h.p. (Denote the union bound on the probability of both these events by  $1 - 3 \exp(-\zeta_1 n)$  for some constant  $\zeta_1$  dependent on  $\epsilon$  and other constants),

$$||\mathbf{h}||_{2} \leq ||\mathbf{h}_{(\mathcal{T}_{0})}||_{2} + ||\mathbf{h}_{(\mathcal{T}_{0})^{c}}||_{2} \leq 4\frac{(BC)}{F} + \sqrt{\frac{A+2B\beta}{F}} + \frac{1}{\sqrt{k}}\beta$$

$$||\mathbf{h}||_2 \le 4\frac{BC}{F} + \sqrt{\frac{A+2B\beta}{F}} + \frac{1}{\sqrt{k}}\beta$$

827 828 829

830 831

832

833

840 841 842

843

849

826

$$||\mathbf{h}||_{2} \leq 4 \frac{(C_{1}\sigma(\sqrt{\frac{2\log d}{n}} + \epsilon)(5\sqrt{k}))}{F} + \sqrt{\frac{\sigma^{2}\epsilon + 2C_{1}\sigma(\sqrt{\frac{2\log d}{n}} + \epsilon)\beta}{F}} + \frac{1}{\sqrt{k}}\beta$$

with a high probability  $1 - 3 \exp(-\zeta_2 n)$ .

We simplify the above expression in terms of the number of samples n, sparsity parameter k, the tail parameter  $\beta$  and the RE constant.

$$||\mathbf{h}||_{2} \leq \zeta_{3} n^{-1/2} k^{1/2} K^{-2} + \zeta_{4} k^{-1/2} \beta + \zeta_{5} n^{-1/4} \beta^{1/2} K^{-1}$$
$$||\mathbf{h}||_{2} \leq \mathsf{O}\left(n^{-1/2} k^{1/2} K^{-2} + k^{-1/2} \beta + n^{-1/4} \beta^{1/2} K^{-1}\right)$$

### A.1.3 PROOF OF COROLLARY 1

The only additional ingredient needed to prove the corollary is the following concentration inequality, 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. Since in the assumption of the corollary, the minimum eigenvalue of the covariance matrix of the set used for sampling is > 0, and the RE is bounded from below by the minimum eigenvalue, the corollary follows.

However although the below theorem is adapted directly from Rudelson & 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. We now

Concentration Inequality for RE condition to hold on sample covariance matrix We use the follow ing Theorem from Rudelson & 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 **X** be an  $n \times d$  matrix, whose rows  $X_1, \ldots, X_n$ are sampled without replacement from a set  $\mathcal{G}$ . Let  $\Sigma$  satisfy the  $\mathsf{RE}(k_0^*, 3k_0, \Sigma^{1/2})$  condition as in Definition 1. Set  $k_1^* = k_0^* + k_0^* \max_j ||\Sigma^{1/2}e_j||_2^2 \times \left(\frac{16\mathsf{RE}(k_0^*, 3\Lambda_0, \Sigma^{1/2})^2(3\Lambda_0)^2(3\Lambda_0+1)}{\delta^2}\right)$ . Assume that  $k_1^* \leq d$  and  $\rho = \rho_{\min}(d, \Sigma^{1/2}) > 0$ . Suppose the sample size satisfies for some absolute constant C862

863 
$$n \ge \frac{CM^2 k_1^{\star} \cdot \log d}{\rho \delta^2} \cdot \log^3 \left( \frac{CM^2 k_1^{\star} \cdot \log d}{\rho \delta^2} \right).$$

864 Then, with probability at least  $1 - \exp(-\delta \rho^2 n/(6M^2 d))$ ,  $\mathsf{RE}(k^{\star}_{\Delta}, k_0, \mathbf{X})$  condition holds for matrix 865  $\frac{1}{\sqrt{n}}\mathbf{X}$  with 866

867

868 870

871

(The inequality is reverse because our definition of Restricted Eigenvalue has the 1/K compared to the definition of Rudelson & Zhou (2013))

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

872 The proof of Theorem 6 is dependent on Theorem 23, which is reproduced below, and Theorem 10 873 (Reduction Principle), which is in the paper. Out of these two, only Theorem 23 has elements 874 related to the randomness of the design. In the proof of the above theorem, authors use concentra-875 tion inequality to extend a RIP-like condition on a general cone (rather than sparse vectors). This 876 concentration inequality results from the following theorem: an augmented version of Theorem 22 877 of Rudelson & Zhou (2013) to the sampling without replacement case. The original proof of The-878 orem 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 879 that the reader goes through the proof beforehand/side-by-side. 880

**Theorem 7.** Set  $1 > \delta > 0$ ,  $0 < k_0^* \le d$  and  $\Lambda_0 > 0$ . Let  $\mathcal{G}$  be a subset of vectors such that  $||Y||_{\infty} \le M$ , with  $\Sigma = \sum_{Y \in \mathcal{G}} \frac{1}{|\mathcal{G}|} YY^{\mathsf{T}}$ .  $\Sigma$  satisfies  $\mathsf{RE}(k_0^*, 3\Lambda_0, \Sigma^{1/2})$ . Rows of 881 882 **X** are drawn uniformly without replacement from  $\mathcal{G}$ . Set  $k_1^{\star} = k_0^{\star} + k_0^{\star} \max_j \|\Sigma^{1/2} e_j\|_2^2 \times$ 883  $\left(\frac{16\mathsf{RE}(k_0^\star,3\Lambda_0,\Sigma^{1/2})^2(3\Lambda_0)^2(3\Lambda_0+1)}{\delta^2}\right)$ . Assume  $k_1^\star \leq d$  and  $\lambda_{\min}(k_1^\star,\Sigma^{1/2}) > 0$ . If for some absorbasic 884 885 *lute constant*  $\zeta_{12}$ *,* 886

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

then with probability  $1 - \exp(\frac{-\delta \lambda_{\min}^2(k_1^{\star}, \Sigma^{1/2})n}{6M^2k_1^{\star}})$ , 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 & Zhou (2013), two arguments require the sampling with replacement (i.i.d. samples), namely symmetrization and Talagrands concentration inequality. We use the sampling without a 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)

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

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 905  $\mathcal{G}$ .  $Z_1$  – 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  $\frac{1}{n} \sum f_j(x, Z'_j) - \mathbb{E}f_j(x, Z_j)$ 906 are zero mean random variable. Then, 908

913 (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) 914

915 
$$\mathbb{E}||\frac{1}{n}\sum f_j(x, Z_j) - \mathbb{E}f_j(x, Z_j)|| \le \mathbb{E}||\frac{1}{n}\sum \left(f_j(x, Z_j) - f_j(x, Z_j')\right)||$$
916

917 
$$\implies \mathbb{E}\left|\left|\frac{1}{n}\sum_{j}\xi_{j}\left(f_{j}(x,Z_{j})-f_{j}(x,Z_{j}')\right)\right|\right| \leq 2\mathbb{E}\left|\left|\frac{1}{n}\sum_{j}\xi_{j}\left(f_{j}(x,Z_{j})\right)\right|\right|$$

899 900 901

902

903 904

907

887

888 889

895

896

(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  $|f_i(x)| \leq \zeta_{13}$  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  $\mathbb{E}W \leq 2\delta n$ .

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

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

$$\begin{split} \sup_{Z'_{i}} |\sup_{x \in F} \sum_{j=1}^{n} f_{j}(x, Z_{j}) - \sup_{x \in F} \left( \sum_{j=1, j \neq i}^{n} f_{j}(x, Z_{j}) + f_{i}(x, Z'_{i}) \right) | \\ &\leq \sup_{Z'_{i}} |\sup_{x \in F} \sum_{j=1, j \neq i}^{n} f_{j}(x, Z_{j}) + \sup_{x \in F} f_{j}(x, Z_{i}) - \sup_{x \in F} \sum_{j=1, j \neq i}^{n} f_{j}(x, Z_{j}) - \inf_{x \in F} f_{i}(x, Z'_{i}) | \\ &\leq \sup_{Z'_{i}} |\sup_{x \in F} f_{j}(x, Z_{i})| \leq \zeta_{13} \end{split}$$

We use the loose 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  $f_i(x, Z_i)$ . This is obviously true. 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,

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

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

$$\mathcal{E} = \{ \mathsf{RE}(k, 4 + 4\gamma, \hat{\Sigma}) \text{ constant } K \ge \lambda_{\min}^{*}^{1/2}, \\ \|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|_{2} = \widetilde{\mathsf{O}} \left( \mathsf{T}_{\text{explore}}^{-1/2} k^{1/2} K^{-2} + k^{-1/2} \beta + \mathsf{T}_{\text{explore}}^{-1/4} \beta^{1/2} K^{-1} \right) \}$$

has the probability,  $\mathbb{P}(\mathcal{E}) \geq 1 - 3 \exp(-\zeta_5 \mathsf{T}_{explore})$  which completes the proof.

A.1.4 PROOF OF THEOREM 2

*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: 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. As shown in the Appendix, the regret can be decomposed as, this step requires care since the regret is with respect to the top-T<sub>explore</sub> arms. In the exploitation stage, the arms are selected such that the top  $(T - T_{explore})$ -arms are played according to  $\hat{\theta}$  and are indexed by the permutation  $\hat{\pi}$ . We next bound the regret for the  $j^{\text{th}}$  selected arm.

1. If  $\theta^{\mathsf{T}} \mathbf{a}^{(\widehat{\pi}(j))} \geq \theta^{\mathsf{T}} \mathbf{a}^{(\pi(j))}$ , then the regret for the  $j^{\text{th}}$  selected arm is negative.

2. If not i.e.,  $\theta^{\mathsf{T}} \mathbf{a}^{(\hat{\pi}(j))} \leq \theta^{\mathsf{T}} \mathbf{a}^{(\pi(j))}$ , then there exists an arm index,  $j_1$  in the permutation  $\pi$  such that  $j_1$  is shifted to the left in  $\hat{\pi}$ . This implies that  $\theta^{\mathsf{T}} \mathbf{a}^{(\hat{\pi}(j_1))} \geq \theta^{\mathsf{T}} \mathbf{a}^{(\pi(j_1))}$ . We decompose the regret for this case with respect to this index and bound the error:

$$\boldsymbol{\theta}^{\mathsf{T}} \mathbf{a}^{(\pi(j))} - \boldsymbol{\theta}^{\mathsf{T}} \mathbf{a}^{(\hat{\pi}(j))} = \underbrace{(\boldsymbol{\theta}^{\mathsf{T}} \mathbf{a}^{(\pi(j))} - \boldsymbol{\theta}^{\mathsf{T}} \mathbf{a}^{(\hat{\pi}(j_1))})}_{\leq 0} + \underbrace{(\boldsymbol{\theta}^{\mathsf{T}} \mathbf{a}^{(\hat{\pi}(j_1))} - \boldsymbol{\theta}^{\mathsf{T}} \mathbf{a}^{(\hat{\pi}(j_1))})}_{\leq 2\epsilon} + \underbrace{(\boldsymbol{\theta}^{\mathsf{T}} \mathbf{a}^{(\hat{\pi}(j_1))} - \boldsymbol{\theta}^{\mathsf{T}} \mathbf{a}^{(\hat{\pi}(j))})}_{\leq 0} + \underbrace{(\boldsymbol{\theta}^{\mathsf{T}} \mathbf{a}^{(\hat{\pi}(j))} - \boldsymbol{\theta}^{\mathsf{T}} \mathbf{a}^{(\hat{\pi}(j))})}_{\leq 2\epsilon}}_{\leq 2\epsilon} \\ \leq \left\langle \boldsymbol{\theta}^{\mathsf{T}} - \hat{\boldsymbol{\theta}}^{\mathsf{T}}, \mathbf{a}^{(\pi(j))} \right\rangle + \left\langle \boldsymbol{\theta}^{\mathsf{T}} - \hat{\boldsymbol{\theta}}^{\mathsf{T}}, \mathbf{a}^{(\pi(j_1))} \right\rangle \leq 4\epsilon$$

Here  $\epsilon$  is the error of estimation after exploration. We therefore obtain the following,

$$\mathbb{E}[\mathsf{Reg}(\mathsf{T})] \leq \mathsf{T}_{\mathsf{explore}} R_{\max} + 4(\mathsf{T} - \mathsf{T}_{\mathsf{explore}})\nu\mathsf{O}(\epsilon) + (\mathsf{T} - \mathsf{T}_{\mathsf{explore}})(1 - \nu)R_{\max},$$

Using  $T \gg T_{explore}$  (the number of exploration rounds is sublinear in T) we obtain,

$$\mathbb{E}[\operatorname{\mathsf{Reg}}(\mathsf{T})] = \mathbb{E}_{\boldsymbol{\theta}}\left[\sum_{t=1}^{\mathsf{T}} \left\langle \boldsymbol{\theta}, \mathbf{a}^{\pi(t)} - \mathbf{a}_{t} \right\rangle\right] \leq \mathbb{E}_{\boldsymbol{\theta}}\left[2R_{\max}\mathsf{T}_{\operatorname{explore}} + 2R_{\max}\|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}\|_{2}\mathsf{T}\right]$$
(10)

**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  $T_{explore} > O(k \log^4 M) > O(k \lambda_{min}^{-4})^3$ .

1000 Step 2. Exploration Period Optimization: (The probability of error terms  $(1 - \nu)$  are left out in the expression.) equation 5 can then be bounded as,

$$\mathbb{E}[\mathsf{Reg}(\mathsf{T})] = \widetilde{\mathsf{O}}\left(\mathsf{T}_{\mathsf{explore}} + \mathsf{T}\mathsf{T}_{\mathsf{explore}}^{-1/2}\lambda_{\min}^{-1}k^{1/2} + \mathsf{T}_{\mathsf{explore}}^{-1/4}\beta^{1/2}\lambda_{\min}^{-1/2} + k^{-1/2}\beta\right)$$

Setting  $T_{\text{explore}} = \widetilde{O}(k^{1/3}T^{2/3})$  we obtain the desired result.

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  $\theta$  be k-sparse,  $\|\theta\|_0 \leq k$  in the sparse linear bandits framework of Theorem 2. 1011 Let  $\lambda_{\min}^*$  be the minimum eigenvalue from equation 2 with the same assumptions as Theorem 2. 1012 Then Algorithm BSLB with exploration period  $T_{explore} = O(k^{\frac{1}{3}}T^{\frac{2}{3}})$ , achieves a regret guarantee of 1013  $\mathbb{E}[\operatorname{Reg}(T)] = O((\lambda_{\min}^*)^{-1}k^{\frac{1}{3}}T^{\frac{2}{3}})$ .

1015 A.1.5 PROOF OF THEOREM 3

1017 For vectors **v** and **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,  $\lambda_{\min}(\mathcal{G}) = \min_{\mathbf{z} \in \mathcal{B}^d} \frac{1}{|\mathcal{G}|} \sum_{\mathbf{v} \in \mathcal{G}} Z_{\mathbf{v}}(\mathbf{z}).$ 

1021 Let randomized rounding be run with  $\hat{g}$ , and  $\hat{\mathcal{G}}$  be the sampled set of arms. Of-course  $|\hat{\mathcal{G}}|$  need to be 1022 equal to  $\hat{g}$ . However, we first assume that the denominator of  $\lambda_{\min}(\hat{\mathcal{G}})$  is equal to  $\hat{g}$ . We later show 1023 that this assumption worsens the approximation guarantees by 2 with high probability. Under this

<sup>&</sup>lt;sup>3</sup>Another way to analyze this would be to assume  $T_{explore} = O(kT^{2/3})$  which would have similar regret guarantees with a slightly better dependence on k ( $k^{-1/12}$  term becomes  $k^{-1/4}$ ) but the results still holds.

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,  $\mathbb{E}_{\frac{1}{\hat{g}}}^{1}Z_{\mathbf{v}}(\mathbf{z}) = \mu_{\mathbf{v}}Z_{\mathbf{v}}(\mathbf{z})$ .

1030 We therefore prove the following result to bound the approximation error between the  $\lambda_{\min}(\hat{\mathcal{G}})$ 1031 obtained from the randomized rounding solution and the optimal solution of the convex relaxation 1032 from equation 6.

**Lemma 4.** Let  $\mathcal{A}$  be a set of  $\mathbb{M}$  arms where each arm is  $\mathbf{a} \in \mathcal{B}^d$  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 6 at  $\hat{g}$  and  $\hat{\mathcal{G}}$  be the set sampled using randomized rounding (Step 18-20 in Alg. 1). Then the following holds,

 $\mathbb{P}\left(\left|\inf_{z\in\mathcal{B}^d}\frac{1}{\hat{g}}\sum_{\mathbf{v}\in\hat{\mathcal{G}}}Z_{\mathbf{v}}(\mathbf{z}) - \inf_{z\in\mathcal{B}^d}\sum_{\mathbf{v}\in\mathcal{A}}\boldsymbol{\mu}_{\mathbf{v}}Z_{\mathbf{v}}(\mathbf{z})\right| \ge \mathsf{O}\left(\frac{\sqrt{d}\log\mathsf{M}}{\sqrt{|\hat{g}|}}\right)\right) \le \frac{1}{\log\mathsf{M}}$ 

(11)

1036

1039

1043 1044 1045

1048

1049

1052 1053

1056

1059

1062

1064

1040 Proof. Using  $\mathbb{E}^{\frac{1}{2}}_{\hat{g}} Z_{\mathbf{v}}(\mathbf{z}) = \boldsymbol{\mu}_{\mathbf{v}} Z_{\mathbf{v}}(\mathbf{z})$  from the preceding paragraph.

1042 By symmetrization,

$$\mathbb{E}\left[\sup_{z\in\mathcal{B}^d}\left|\sum Z_{\mathbf{v}}(\mathbf{z}) - \mathbb{E}\sum Z_{\mathbf{v}}(\mathbf{z})\right|\right] \leq 2\mathbb{E}\left[\sup_{z\in\mathcal{B}^d}\left|\sum \xi_{\mathbf{v}} Z_{\mathbf{v}}(\mathbf{z})\right|\right]$$

where  $\xi_{\mathbf{v}}$  are i.i.d. Rademacher random variables (we overload  $\mathbf{v}$  as the index). Now using Dudley's integral inequality,

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

1050 where  $\Psi$  is the constant which satisfies,

$$||\sum \xi_{\mathbf{v}} Z_{\mathbf{v}}(\mathbf{z}_1) - \sum \xi_{\mathbf{v}} Z_{\mathbf{v}}(\mathbf{z}_2)||_{\psi} \le \Psi ||z_1 - z_2||_2 \le \sqrt{2}\Psi$$

1054 Now w.l.o.g., 1055

$$||\sum \xi_{\mathbf{v}} Z_{\mathbf{v}}(\mathbf{z}_1) - \sum \xi_{\mathbf{v}} Z_{\mathbf{v}}(\mathbf{z}_2)||_{\psi} \le 2||\sum \xi_{\mathbf{v}} Z_{\mathbf{v}}(\mathbf{z}_1)||_{\psi}$$

1057 Now from the definition of subgaussian norm,1058

$$||\sum \xi_{\mathbf{v}} Z_{\mathbf{v}}(\mathbf{z}_1)||_{\psi} = \inf\{t : \exp \frac{(\sum \xi_{\mathbf{v}} Z_{\mathbf{v}}(\mathbf{z}_1))^2}{t^2} \le 2\}$$

1061 Now,

$$\exp\frac{(\sum \xi_{\mathbf{v}} Z_{\mathbf{v}}(\mathbf{z}_1))^2}{t^2} \le \exp\frac{(\sum Z_{\mathbf{v}}(\mathbf{z}_1))^2}{t^2} \le \exp 2\frac{\sum (Z_{\mathbf{v}}(\mathbf{z}_1))^2}{t^2} = \prod \exp 2\frac{(Z_{\mathbf{v}}(\mathbf{z}_1))^2}{t^2} \text{(independence)}$$
  
Now by Chernoff's bound,

1065  
1066  
1067
$$\prod \exp 2\frac{(Z_{\mathbf{v}}(\mathbf{z}_{1}))^{2}}{t^{2}} \leq \prod \exp \left\{ |\mathcal{G}| \boldsymbol{\mu}_{\mathbf{v}} \left( \exp \left( 2\frac{(\mathbf{z}^{\mathsf{T}} \mathbf{v} \mathbf{v}^{\mathsf{T}} \mathbf{z})^{2}}{|\mathcal{G}|^{2} t^{2}} \right) - 1 \right) \right\} = \exp \left\{ |\mathcal{G}| \left( \exp \left( 2\frac{(\mathbf{z}^{\mathsf{T}} \mathbf{v} \mathbf{v}^{\mathsf{T}} \mathbf{z})^{2}}{|\mathcal{G}|^{2} t^{2}} \right) - 1 \right) \right\}$$
1068
To find  $\inf \{t : \exp(-) \leq 2\}$ 

1068 To find 
$$\inf\{t : \exp(\frac{1}{t^2}) \le 2\}$$
  
1069

$$\exp\left\{ |\mathcal{G}| \left( \exp\left(2\frac{(\mathbf{z}^{\mathsf{T}}\mathbf{v}\mathbf{v}^{\mathsf{T}}\mathbf{z})^{2}}{|\mathcal{G}|^{2}t^{2}}\right) - 1 \right) \right\} \leq 2 \implies \exp\left(2\frac{(\mathbf{z}^{\mathsf{T}}\mathbf{v}\mathbf{v}^{\mathsf{T}}\mathbf{z})^{2}}{|\mathcal{G}|^{2}t^{2}}\right) \leq \frac{\ln 2}{|\mathcal{G}|} + 1$$
$$(\mathbf{z}^{\mathsf{T}}\mathbf{v}\mathbf{v}^{\mathsf{T}}\mathbf{z})^{2}\frac{2}{\ln\left(\frac{\ln 2}{|\mathcal{G}|} + 1\right)|\mathcal{G}|^{2}} \leq t^{2} \implies \frac{\sqrt{2}\mathbf{z}^{\mathsf{T}}\mathbf{v}\mathbf{v}^{\mathsf{T}}\mathbf{z}}{\sqrt{\ln\left(\frac{\ln 2}{|\mathcal{G}|} + 1\right)|\mathcal{G}|}} \leq t$$

1074 1075

1070 1071

1076 Upper bounding  $\mathbf{z}^{\mathsf{T}} \mathbf{v} \mathbf{v}^{\mathsf{T}} \mathbf{z} \leq 1$ , we obtain the inf *t*,

$$\frac{\sqrt{2}}{\sqrt{2}} \le t$$

1079 
$$\sqrt{\ln\left(\frac{\ln 2}{|\mathcal{G}|}+1\right)|\mathcal{G}|}$$

with 
$$|\mathcal{G}| >> \ln 2$$
,  $\ln\left(\frac{\ln 2}{|\mathcal{G}|} + 1\right) \approx \frac{\ln 2}{|\mathcal{G}|}$ ,

Therefore combining, 

$$\mathbb{E}\left[\sup_{z\in\mathcal{B}^d}\left|\sum Z_{\mathbf{v}}(\mathbf{z}) - \mathbb{E}\sum Z_{\mathbf{v}}(\mathbf{z})\right|\right] \le 2\sqrt{2}\zeta_{11}\log^{1/2}(\frac{3}{\epsilon})\frac{d^{1/2}}{\sqrt{|\mathcal{G}|}} = \zeta_{10}\frac{\sqrt{d}}{\sqrt{|\mathcal{G}|}}$$

 $\frac{\sqrt{2}}{\sqrt{|\mathcal{G}|}} \le t$ 

Using Markov's inequality and the above lemma,

$$\mathbb{P}(\sup_{z\in\mathcal{B}^d}\left|\sum Z_{\mathbf{v}}(\mathbf{z}) - \mathbb{E}\sum Z_{\mathbf{v}}(\mathbf{z})\right| \geq \zeta_{10}\frac{\sqrt{d}\log\mathsf{M}}{\sqrt{|\mathcal{G}|}}) \leq \frac{1}{\log\mathsf{M}}$$

If,

$$\sup_{z \in \mathcal{B}^d} \left| \sum Z_{\mathbf{v}}(\mathbf{z}) - \mathbb{E} \sum Z_{\mathbf{v}}(\mathbf{z}) \right| \le \epsilon$$

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

1100  
1101  
1102  
1103  

$$\mathbb{E} \sum Z_{\mathbf{v}}(\mathbf{z}) - \epsilon \leq \sum Z_{\mathbf{v}}(\mathbf{z})$$

$$\inf_{z \in \mathcal{B}^d} \mathbb{E} \sum Z_{\mathbf{v}}(\mathbf{z}) - \epsilon \leq \inf_{z \in \mathcal{B}^d} \sum Z_{\mathbf{v}}(\mathbf{z})$$

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

(12)

Then, we bound the size of the actual sampled set with respect to  $\hat{q}$ , 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  $\hat{g}$  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, 

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

 $\mathbb{P}(\frac{\hat{g}}{2} \le |\hat{\mathcal{G}}| \le 2\hat{g}) \ge 1 - 2(\frac{2}{e})^{\frac{\hat{g}}{2}}$ 

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

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

$$\begin{split} \mathbb{P}(|\hat{\mathcal{G}}| \geq 2\hat{g}) &\leq \inf_{t>0} \exp(-t2\hat{g}) \mathbb{E}[\exp(t|\hat{\mathcal{G}}|)] \\ &= \inf_t \exp(-t2\hat{g}) \prod_j \mathbb{E}[\exp(tp_j)] \text{ (independent rv)} \end{split}$$

$$\inf_{t} \exp(-t2\hat{g}) \prod_{j} \mathbb{E}[\exp(tp_{j})] \leq \inf_{t} \exp(-t2\hat{g}) \prod_{j} \exp(\hat{g}\boldsymbol{\mu}_{j}(\exp(t) - 1))$$
$$= \inf_{t} \exp(-t2\hat{g}) \exp(\hat{g}(\exp(t) - 1))$$

Achieves infinum for  $t = \ln 2$ ,

1132  
1133 
$$\mathbb{P}(|\hat{\mathcal{G}}| \ge 2\hat{g}) \le \exp(-2\hat{g}\ln 2 + \hat{g}) = (\frac{3}{e})^{\hat{g}}$$

<sup>1134</sup> Using a similar left tail bound,

1136  
1136  
1137  
1138  
1139  

$$\mathbb{P}(|\hat{\mathcal{G}}| \le \hat{g}/2) \le \inf_{t < 0} \exp(-t\hat{g}/2) \mathbb{E}[\exp(t|\hat{\mathcal{G}}|)]$$

$$= \inf_{t} \exp(-t\hat{g}/2) \prod_{j} \mathbb{E}[\exp(tp_{j})] \text{ (independent rv)}$$
1139

$$\inf_{t} \exp(-t\hat{g}/2) \prod_{j} \mathbb{E}[\exp(tp_{j})] \leq \inf_{t} \exp(-t\hat{g}/2) \prod_{j} \exp(\hat{g}\boldsymbol{\mu}_{j}(\exp(t) - 1))$$
$$= \inf_{t} \exp(-t\hat{g}/2) \exp(\hat{g}(\exp(t) - 1))$$

1141 1142 1143

1145 1146

1140

1144 Achieves infinum for  $t = -\ln 2$ ,

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

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

1149 1150

1155

1172 1173

1174

1178 1179

1151 Therefore the above lemma helps us prove the following statement,

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

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

1159 We claim the following about the solution of the convex relaxation at  $\hat{g}$  and d,

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

1171 Also since the convex relaxation at  $g^*$  is more constrained than the one at d,

$$\lambda_{\min}^*(g^*) \le \lambda_{\min}^*(d) \le \frac{\hat{g}}{d} \lambda_{\min}^*(\hat{g}) \implies \lambda_{\min}^*(\hat{g}) \ge \frac{d}{\hat{g}} \lambda_{\min}^*(g^*)$$

1175 We can now combine the two lemmas and the equation above to say that for an error  $\epsilon$  and  $\hat{g} = \zeta \frac{d \log^2 M}{\epsilon^2}$  the following holds with high probability,

$$\frac{\epsilon^2}{\zeta \log^2 \mathsf{M}} \lambda_{\min}^* - \frac{1}{2} \epsilon \leq \hat{\lambda_{\min}}$$

1180 where  $\zeta$  is a constant.

1182 Further bounding the lower bound,

$$\frac{\epsilon^2}{\zeta \log^2 \mathsf{M}} \lambda_{\min}^* - \frac{1}{2} \epsilon \ge \frac{\lambda_{\min}^*}{\log^2 \mathsf{M}}$$
1185

1186 
$$\zeta_1 \epsilon^2 - \frac{1}{2} \epsilon - \zeta \zeta_1 \ge 0$$

which has a positive determinant and can only satisfied when  $\epsilon \notin \left[\frac{1/2-\sqrt{1/4+\zeta\zeta_1^2}}{2\zeta_1}, \frac{1/2+\sqrt{1/4+\zeta\zeta_1^2}}{2\zeta_1}\right]$ . Since  $\epsilon > 0$ , therefore for this to be true,  $\epsilon > \frac{1/2 + \sqrt{1/4 + \zeta \zeta_1^2}}{2\zeta_1}$  $\epsilon > \frac{1/2 + \sqrt{1/4 + \zeta(\frac{\lambda_{\min}^*}{\zeta \log^2 \mathbf{M}})^2}}{2\frac{\lambda_{\min}^*}{2\lambda_{\min}^*}} = \frac{\zeta \log^2(\mathbf{M})/2 + \sqrt{\zeta^2 \log^4 \mathbf{M}/4 + \zeta(\lambda_{\min}^*)^2}}{2\lambda_{\min}^*}$ If  $\lambda_{\min}^* > \zeta \log^2(M)/2$ , then we can assume the stronger condition (note that constant  $\zeta$  from Lemma 4 is > 1),  $\epsilon > (1 + \sqrt{2\zeta})$ The search bound, therefore becomes,  $\hat{g} \leq \frac{d \log^2 \mathsf{M}\zeta}{(1+\zeta^2+2\sqrt{2\zeta})} \leq \frac{d \log^2 \lambda_{\min}}{(1+\zeta^2+2\sqrt{2\zeta})}, \implies \hat{g} = \mathsf{O}(d \log^2 \mathsf{M}).$ This completes the proof of the theorem. However if  $\lambda_{\min}^* < \zeta \log^2(\mathsf{M})/2$ , Then  $\epsilon > \frac{(1+\sqrt{2\zeta})\zeta \log^2(\mathsf{M})}{4\lambda_{\min}^*}$ and the search bound becomes,  $\hat{g} \leq \frac{16d(\lambda_{\min}^*)^2}{(1+\zeta^2+2\sqrt{2}\zeta)\zeta\log^2(\mathsf{M})} \implies \hat{g} = \mathsf{O}\big(\frac{d(\lambda_{\min}^*)^2}{\log^2\mathsf{M}}\big)$ Therefore for  $\zeta_2 = \min(\lambda_{\min}^*, \log^2 \mathsf{M}), \hat{g} = \mathsf{O}(\frac{d\zeta_2}{\log^2 \mathsf{M}}).$ Another way to derive a tighter bound on  $\lambda^*$  is, We consider the stronger condition,  $\epsilon > \frac{2\sqrt{\zeta^2 \log^4 \mathsf{M}/4 + \zeta(\lambda^*_{\min})^2}}{2\lambda^*_{\min}}$ Plugging this into  $\hat{g}$ ,  $\hat{g} \le \frac{d \log^2 \mathsf{M}(\lambda_{\min}^*)^2}{(\zeta \log^4 \mathsf{M}/4 + (\lambda^* \cdot \cdot)^2)}$ From  $\hat{g} \ge d$ ,  $\lambda_{\min}^* \ge \sqrt{\frac{\zeta \log^4 \mathsf{M}}{\log^2 \mathsf{M} - 1}}$ Under a stronger condition,  $\lambda_{\min}^* \ge \sqrt{\frac{2\zeta \log^4 \mathsf{M}}{\log^2 \mathsf{M}}} = \sqrt{2\zeta} \log \mathsf{M} \implies \lambda_{\min} = \mathsf{O}(\log \mathsf{M})$ A.1.6 BRUTE FORCE ALGORITHM FOR SEARCHING THE OPTIMAL SUBSET From the previous provides proof we can set  $\hat{g} = \frac{4\zeta d \log^2 M}{(\lambda_{\min}^*)^2}$ , and then search for subsets in the range  $[d, \hat{g}]$  to obtain a minimum eigenvalue  $\hat{\lambda_{\min}}$ . We obtain the approximation guarantee,  $\frac{1}{2}\lambda_{\min}^* \leq \hat{\lambda_{\min}}$ w.h.p., since we are only using the approximation guarantee from Lemma 4 and Lemma 5, and not from equation 13 because we are already searching the space  $\langle \hat{g} \rangle$ . Since the search space is dependent on Poly(d), the time complexity of the brute force algorithm, Algorithm 2 follows. This

- 1240 of the order O(exp(M))
- 1241 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

time complexity is substanially smaller than the complexity over the search over all arms which is

Algorit	hm 2 Brute Force Search for Optimal Subset
Input	t Approximation Factor $\epsilon$ , Search Bound $\hat{g}$
Outp	$\underline{ut}$ Subset $\underline{\mathcal{G}}$
Set $\lambda$	$\min_{d} = 0, \mathcal{G} = \phi$
ior a fe	or $G'$ in $\{G \subseteq A :  A  = \overline{d}\}$ do
	if $\lambda_{\min}(\sum_{\mathbf{r} \in \mathcal{C}'}  \mathcal{G}'^{-1}  \mathbf{aa}^{T}) > \lambda_{\min}$ then
	Set $\lambda_{\min} = \lambda_{\min} (\sum_{\mathbf{r} \in \mathcal{C}'}  \mathcal{G}'^{-1}  \mathbf{aa}^{T}), \bar{\mathcal{G}} = \mathcal{G}'$
	end if
e	nd for
end	tor
value in the sam	a Equation 2 for each subset - the lower bound can be the maximum objective value across pled subsets.
What co	an the practitioner do to select a good subset size empirically? Additionally, if a practitioner
wants t	o test out a particular choice of $\hat{g}$ , the worst-case error can be empirically calculated (the
differer	the between the convex relaxation at d and averaged across multiple randomized rounding different values of $\hat{a}$ . This is possible because CETCOOD SUBSET(4) can be run offling.
	f different values of $g$ ). This is possible because GETGOODSOBSET(A) can be full officient.
A 1 7	Why Dong the Auguston Musical Events of Courses and Merce Courses
A.I./	WHY DOES THE AVERAGE MINIMUM EIGENVALUE CONSTRAINT NOT SATISFY Matroid Constraints
	MARKOLD CONSTRAINTS
Existing	g work in experiment design work with objective functions which often satisfy the matroid,
submod	lularity or cardinality constraints to perform experiment design (Allen-Zhu et al., 2021;
Brown	et al., 2024). However we need to optimize the minimum eigenvalue averaged across the (because we want to avoid dependence of M in the regret term and also use the PE condi-
tion). (	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 aver-
aged ac	cross the set, so there is no clear partitioning/structure to the feasible set). Finally, we tried
to prov	e submodularity but were unable to do so for our objective function especially because the
denomi	nator is dependent on the subset size.
A.1.8	DETAILS ON CORRALLING
Brief (	Overview on Corralling: The algorithm CORRAL (Agarwal et al., 2017) is a meta-bandit
algorith	im that uses online mirror descent and importance sampling to sample different bandit algo-
rithms 1	that receive the reward. Using the rewards updates the probabilities used for sampling. The
main ol	bjective is to achieve a regret which is as good as if the best base algorithm was run on its
own.	
To setu	p the context, we exactly reproduce the following excerpt, definition and theorems have been
taken fi	rom Agarwal et al. (2017):
For an	environment $\mathcal{E}$ , we define the environment $\mathcal{E}'$ induced by importance weighting, which is the
environ	ment that results when importance weighting is applied to the losses provided by environment
E. Mor	e precisely, $\mathcal{E}'$ is defined as follows. On each round $t = 1, \ldots, T$ ,
1. E' pi	cks an arbitrary sampling probability $p_t \in (0,1]$ and obtains $(x_t, f_t) = \mathcal{E}(\theta_1, \dots, \theta_{t-1})$ .
2. E' re	veals $x_t$ to the learner and the learner makes a decision $ heta_t$ .
3. With $f'_t(\theta, x)$	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 $\theta = 0$ and $\theta'_t \in \Theta$ to be arbitrary.
4. E' re	veals the loss $f'_t(\theta_t, x_t)$ to the learner, and passes $\theta'_t$ to $\mathcal{E}$ .
Definit	<b>ion 3.</b> (Agarwal et al., 2017) For some $\alpha \in (0, 1]$ and non-decreasing function $R : \mathbb{N}_+ \to$
$\mathbb{R}_+$ , an its regre	algorithm with decision space $\mathcal{O}$ is called $(\alpha, R)$ -stable with respect to an environment $\mathcal{E}$ if et under $\mathcal{E}$ is $R(T)$ , and its regret under any environment $\mathcal{E}'$ induced by importance weighting

is

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

1300 1301 1302 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  $(1, \mathbb{E}[\text{Reg}(\mathsf{T})])$ -stable by rescaling the losses.

1305 1306 We

1310 1311

1312 1313

**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  $(\alpha_i, R_i)$ -stable (recall Defn. 3) with respect to an environment  $\mathcal{E}$ , then under the same environment CORRAL satisfies

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

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 $_{\theta \in \Theta, \pi \in \Pi} \left[\sum_{t=1}^T f_t(\theta_t, \pi_t) - f_t(\theta, \pi_t)\right] = \widetilde{O}\left(\sqrt{MT} + MR_i(T)\right).$ 

- A.1.9 PROOF OF THEOREM 4
- 1321 1322 1323

1324

1328

1330

1331

1332

1333

1334

Algorithm 3 Corralling with Blocked Linear Sparse Bandits (C-BSLB)

1: Input: Dimension d, Total Number of Rounds T, Regret Bound of Best Algorithm R<sub>best</sub>

1325 1. Input Dimension a, rotal realised of realised v,  
1326 2: Set Learning rate 
$$\eta = \min\left(\frac{1}{40TR_{best}}, \sqrt{\frac{\lfloor \log_2(d) \rfloor}{T}}\right)$$

3: Set Exponential Grid  $k \in [1, 2, \dots 2^{\lfloor \log_2(d) \rfloor}]$ 

- 4: Initialize  $\lfloor \log_2(d) \rfloor + 1$  Base Algorithms one for each sparsity parameter on an exponential grid, BSLB( $T_{explore} = \zeta k^{1/3} T^{2/3}$ )
  - 5: Sample  $M_{\text{sampled}} = \zeta d^{1/3} T^{2/3}$  arms without replacement to be used as proxy samples.
  - 6: Run Corral( $\lfloor \log_2(d) \rfloor + 1$  BSLB algorithms,  $\eta$ ) 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 3 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:

1345

1347

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.

1348
1349
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  $k^*$  may not lie on the grid and we need to bound  $\mathbb{E}[\operatorname{Reg}(\mathsf{T})]_{k^*}$  in terms of  $\mathbb{E}[\operatorname{Reg}(\mathsf{T})]_s$ . To that end we prove the following lemma,

**Lemma 6.** Let  $k^*$  be the sparsity parameter at which the regret bound of 5 is minimized. And let  $s \in \{2^i\}_{i=0}^{\log_2[d]}$  be the parameter on grid which is closest to  $k^*$  in absolute distance. Then the following holds (where  $\nu$  is the probability of exploration round succeeding at sparsity level  $k^*$ ), 1363

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

1364 1365

1377 1378

1381 1382

1356 1357

1367 *Proof.* Let the bound on the expected regret for sparsity level k be given by  $\mathbb{E}[\text{Reg}(\mathsf{T})]_k$ .

From the statement of the theorem, we assume that for the optimal sparsity parameter  $k^*$ , the nearest parameter (on the exponential scale) is s.  $k^*$  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  $k^*$ , we would need at most  $Y = \lfloor \log_2(4k^*) \rfloor$  queries to search for  $k^*$ . Let  $k_1^*, k_2^*, \ldots, k_Y^*$  be the mid-points of these queries, where  $k_Y^* = k^*$ . 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 2, the following inequalities can be obtained,

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

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

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

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

$$\mathbb{E}[\operatorname{Reg}(\mathsf{T})]_{s} \leq \alpha \mathbb{E}[\operatorname{Reg}(\mathsf{T})]_{k_{1}^{*}} + O(1-\nu) \leq \cdots \leq (\sqrt{2})^{Y} \mathbb{E}[\operatorname{Reg}(\mathsf{T})]_{k^{*}} + Y \mathsf{O}(1-\nu)$$
$$\mathbb{E}[\operatorname{Reg}(\mathsf{T})]_{s} \leq 2\sqrt{k^{*}} \mathbb{E}[\operatorname{Reg}(\mathsf{T})]_{k^{*}} + \log_{2}(4k^{*}) \mathsf{O}(1-\nu)$$

1390 1391 1392

1395

1388 1389

1393

### 1394 A.2 ADDITIONAL NUMERICAL EXPERIMENTS

### 1396 A.2.1 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.

1404 We run the corralling algorithm using copies = 4 copies of Algorithm 1 each with different ex-1405 ploration periods,  $T_{explore}$ . Each of the instance, we first give  $T_{explore}$  random recommendations by 1406 sampling uniformly without replacement from a suitably constructed subset  $\mathcal{G}$  to each user. Given 1407 the ratings obtained, we estimate the parameter  $\theta_{user}$  specific to the user using only their recommen-1408 dations. For the remaining  $T_{exploit} = T - T_{explore}$  rounds, we give the top  $T_{exploit}$  recommendations based on the estimated parameter. . To benchmark we run the algorithms independently and also 1409 against a random policy which randomly recommends. We next describe the three tasks that we 1410 report our results on, 1411

**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 M = 1500 books and we consider 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].

MovieLens (Movie Ratings): MovieLens 100K dataset has 100,000 ratings (5-star) from 1000 users on 1700 movies (Harper & Konstan, 2015). We obtain the embeddings of the TMDB descriptions of the movies using the same sentence transformer model. For experimentation, we consider the users which have more than 300 ratings and average the results across the users, there are 300 such users and we run the experiment with 100 different seeds for each user. We consider T<sub>explore</sub> for the different instances in the range [50, 100, 150, 200].

**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  $T_{explore}$  is taken to be [20, 40, 60, 80] **Results:** We summarize



Figure 1: Cumulative Regret for recommendation using only single ratings using BSLB with different exploration periods and when run with CORRAL (Agarwal et al., 2017). Since the regret is with respect to ranking the arms, it can decrease (since there can be negative terms).

the cummulative regret from Eq. 1 of the algorithms in Figure 1. 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,  $T_{explore}$  out of the 5.

## 1448 A.2.2 Hyperparameters for Experiment of Section 3

1450 The hyperparameters are presented in Table 2. Note that the reason for selecting different T across 1451 different objects was because the validation datasets had to be big enough (so that the variance of 1452 accuracy is informative) and different objects had different total number of samples. The number of 1453 exploration rounds are set with respect to  $T(\sim 0.5T - 0.7T)$  so that the approximation error after 1454 exploration is small enough. The active learning methods are also run with random initialization of  $T_{explore}$  explorations and one round of querying with  $T - T_{explore}$  queries. We present the results of 1455 a study where we vary only the  $T_{explore}$  for the same value of T in Table 3 and observe that with 1456 decreasing T<sub>explore</sub> the estimation of difficulty scores deteriorates, and the performance on hard-1457 valid deteriorates. The performance on easy-valid improves since samples are randomly chosen if 1458 the estimation is unsuccessful. Note that our method still performs better than AL baselines and 1459 random sampling. 1460

Object Being Annotated	$N_{\mathrm{valid}}$ (hard)	$N_{\mathrm{valid}}$ (easy)	т	<b>T</b> <sub>explore</sub>		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

1468 Table 2: Different hyperparameters used for the experiment of  $Sec \sim 3$ . The num samples show 1469 how our method achieves a similar accuracy (-1% to 4% improvement over all) by considering 1470 substantially less samples.

	Validation Type	Object Annotated	AnchorAL	SEALS	Random	All	Our (BSLB)
	easy-valid	chair	94.5±1.0	93.2±1.5	95.7±1.0	96.0±1.4	92.3±1.4
		car	$95.8{\pm}2.5$	$95.3{\pm}2.0$	$97.3 \pm 1.2$	98.2±1.1	95.7±1.7
$T_{avplore} = 80$		bottle	$95.0{\pm}0.9$	$95.0{\pm}1.1$	96.2±1.9	96.7±1.8	96.8±1.2
explore 00	hard-valid	chair	72.0±3.0	71.0±1.4	67.0±5.0	69.4±4.1	73.4±2.4
		car	$66.4 \pm 7.9$	$69.2 \pm 5.6$	$52.6 \pm 8.2$	$58.8{\pm}6.8$	$74.0{\pm}2.7$
		bottle	$61.2{\pm}1.5$	$61.8{\pm}2.0$	$51.2{\pm}2.9$	$52.6 {\pm} 3.1$	63.2±2.9
	easy-valid	chair	94.8±2.1	93.7±1.7	95.5±2.0	95.5±1.4	94.8±1.8
		car	$96.2{\pm}1.9$	$95.5 {\pm} 1.9$	$97.5 \pm 1.2$	98.3±1.1	$97.0{\pm}2.3$
$T_{max} = 60$		bottle	$95.0{\pm}1.2$	$94.8 {\pm} 1.0$	$97.2 \pm 1.7$	97.5±1.7	$96.0{\pm}1.9$
r explore — 00	hard-valid	chair	$70.0{\pm}5.5$	70.0±3.8	68.0±4.2	69.8±4.0	72.4±1.7
		car	$65.8 {\pm} 8.7$	$70.0{\pm}7.5$	$53.2 \pm 5.4$	$60.6 \pm 8.1$	72.6±3.9
		bottle	$61.6{\pm}1.0$	$61.6{\pm}2.3$	$53.4{\pm}2.6$	$54.0{\pm}2.6$	$62.6{\pm}2.8$
$T_{explore} = 30$	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	$95.8{\pm}1.9$	97.3±1.2	98.2±1.1	92.0±11.6
		bottle	$95.3{\pm}0.8$	$95.0{\pm}0.5$	$96.2{\pm}1.9$	$96.7{\pm}1.8$	96.7±1.2
	hard-valid	chair	69.4±2.1	71.2±2.4	67.6±4.9	69.8±4.0	$70.8 {\pm} 4.5$
		car	$64.2 \pm 9.5$	$67.6 {\pm} 7.7$	$52.6 {\pm} 8.2$	$58.8{\pm}6.8$	70.8±6.5
		bottle	$602 \pm 19$	620+21	512+29	$52.6 \pm 3.1$	63.0+4.9

Table 3: Learning Accuracies on Different Methods for Image Classification in PASCAL-VOC 1494 2012: Effect of  $T_{explore}$  with the number of rounds fixed at T = 120 and with 120 easy-valid and 1495 100 hard-valid samples. 1496

1497 1498

1471 1472

#### 1499 A.2.3 **TEXT CLASSIFICATION ON SST-2** 1500

1501 Next we present a result on text classification task on SST-2 (Socher et al., 2013). However since 1502 there were no human difficulty ratings available for this task, we use rarity of text (Zhang et al., 1503 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  $T_{explore} = 100$  samples and 1504  $T_{\text{exploit}} = 200$ . The normalized rarity ranges from 0 to 1 and we set  $\tau_{\text{hard}} = 0.5$  and  $\tau_{\text{easy}} = 0.2$ . 1505

1506 We observe a similar trend as the previous task where **BSLB** method performs better than a **random** 1507 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** 1509 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 1510 2%. However, since this is a standard sentiment analysis task, the embeddings are more informative, 1511 thereby improving the baseline performance for a random subset.



Figure 2: 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  $T \ll d$  and the hardness of the samples considered.

A.2.4 SIMULATION STUDY





1554

1536 1537

1538 1539

1540

1541

1542

1543

1544

1545 1546 1547

1548 1549

1550

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

Finally, we also run a simulation study to study the efficacy of our BSLB and C-BSLB algorithm and 1555 demonstrate how CORRAL can be used to achieve a sub-linear regret without the knowledge of the 1556 optimal parameters. We compute the cumulative regret at time t compared to the top-t arms, and 1557 unlike the standard bandit setting, in a blocked setting, the cumulative regret need not be monotonic. 1558 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 1560 experimental set-up with the following parameters M = 10000, d = 1000 and T = 300. At sparsity 1561 level k = 10, the tail parameter is  $\gamma = 3$ . 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 1563 corralling with  $\ell_2$ -regularization, showing that our method exploits the sparsity and does not require 1564 true knowledge of the hyperparameters. We also benchmark against a random policy and show that 1565 our method performs significantly better showing that the upper bound on regret is not vacuous.

### 1566 A.2.5 CORRELATION BETWEEN MODEL DIFFICULTY AND HUMAN ANNOTATION DIFFICULTY

In Figure 4a, we show that for the chair images, if a model  $\mathcal{M}$  (from the numerical study of Section 3) 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.





(b) Histogram shows the heavy-tailed distribution of the difficulty score from Ionescu et al. (2016) of

(a) Fraction of difficult samples (labelled by hu- the *chair* object of the PASCAL-VOC dataset. We mans) against the distance from decision bound- clip the entries from the middle since they make ary for SVM trained on all *chair* images. As the the difficult estimation noisier, in practical impledistance from the decision boundary increases the mentation, one would need to develop a mechfraction of difficult samples (difficulty rating from anism to flag samples with *ambiguous* difficulty humans > 3.5) decays to 0. and this is left for future work.

### 1593 A.2.6 How does the tail of the parameter matter?

In Figure 5, we investigate the effect of the tail parameter in the performance of BSLB with a fixed exploration period  $T_{explore} = 80$  and different sizes of the tail in the same setup as the simulation study of Appendix A.2.4. We observe that as the tail parameter  $\gamma$  grows, the regret worsens, however we remark that even for a decent  $\gamma = 75$ , the performance is reasonable.

### A.2.7 CONVERGENCE OF CORRAL PARAMETERS IN C-BSLB

We plot the convergence of the CORRAL parameters of C-BSLB in Figure 6 for the simulated experiment of Appendix A.1.8. We observe that the probability of sampling the best algorithm  $(T_{explore} = T_3 = 80)$  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  $T \ll d$  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.

60

1587

1590

- 160
- 1609 1610
- 1611
- 1612
- 1613
- 1614
- 1615
- 1616
- 1617
- 1618
- 1619



Figure 6: Convergence of the different sampling probabilities for the base algorithms of the C-BSLB (Algorithm 3). 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.



Figure 7: 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.