PopArt: Efficient Sparse Regression and Experimental Design for Optimal Sparse Linear Bandits

Kyoungseok Jang
University of Arizona
ksajks@arizona.edu

Chicheng Zhang
University of Arizona
chichengz@cs.arizona.edu

Kwang-Sung Jun
University of Arizona
kjun@cs.arizona.edu

Abstract

In sparse linear bandits, a learning agent sequentially selects an action and receive reward feedback, and the reward function depends linearly on a few coordinates of the covariates of the actions. This has applications in many real-world sequential decision making problems. In this paper, we propose a simple and computationally efficient sparse linear estimation method called PopArt that enjoys a tighter ℓ_1 recovery guarantee compared to Lasso (Tibshirani, 1996) in many problems. Our bound naturally motivates an experimental design criterion that is convex and thus computationally efficient to solve. Based on our novel estimator and design criterion, we derive sparse linear bandit algorithms that enjoy improved regret upper bounds upon the state of the art (Hao et al., 2020), especially w.r.t. the geometry of the given action set. Finally, we prove a matching lower bound for sparse linear bandits in the data-poor regime, which closes the gap between upper and lower bounds in prior work.

1 Introduction

In many modern science and engineering applications, high-dimensional data naturally emerges, where the number of features significantly outnumber the number of samples. In gene microarray analysis for cancer prediction [20], for example, tens of thousands of genes expression data are measured per patient, far exceeding the number of patients. Such practical settings motivate the study of high-dimensional statistics, where certain structures of the data are exploited to make statistical inference possible. One representative example is sparse linear models [13], where we assume that a linear regression task's underlying predictor depends only on a small subset of the input features.

On the other hand, online learning with bandit feedback, due to its practicality in many applications such as online news recommendations [17] or clinical trials [18, 22], has attracted a surge of research interests. Of particular interest is linear bandits, where in n rounds, the learner repeatedly takes an action A_t (e.g., some feature representation of a product or a medicine) from a set of available actions $\mathcal{A} \subset \mathbb{R}^d$ and receives a reward $r_t = \langle \theta^*, A_t \rangle + \eta_t$ as feedback where $\eta_t \in \mathbb{R}$ is an independent zero-mean, σ-sub-Gaussian noise. Sparsity structure is abundant in linear bandit applications: for example, customers' interests on a product depend only on a number of its key specs; the effectiveness of a medicine only depends on a number of key medicinal properties, which means that the unknown parameter θ^* sparse; i.e., it has a small number of nonzero entries.

Early studies [2, 6, 16] on sparse linear bandits have revealed that leveraging sparsity assumptions yields bandit algorithms with lower regret than those provided by full-dimensional linear bandit algorithms [3, 4, 8, 1]. However, most existing studies either rely on a particular arm set (e.g., a norm ball), which is unrealistic in many applications, or use computationally intractable algorithms. If we consider an arbitrary arm set, however, the optimal worst-case regret is $\Theta(\sqrt{sdn})$ where s is the sparsity level of θ^*, which means that as long as $n = O(sd)$, there exists an instance for which the algorithm suffers a linear regret [15]. This is in stark contrast to supervised learning where it is possible to enjoy nontrivial prediction error bounds for $n = o(d)$ [11]. This motivates a natural research question: Can we develop computationally efficient sparse linear bandit algorithms

Preprint. Under review.

	REGRET BOUND	DATA-POOR	ASSUMPTIONS
HAO ET AL. [12]	$\tilde{O}(s^{2/3}\mathcal{C}_{\min}^{-2/3}n^{2/3})$	✓	$\mathcal{A}$ SPANS $\mathbb{R}^d$
HAO ET AL. [12]	$\Omega(s^{1/3}\kappa^{-2/3}n^{2/3})$	✓	$\mathcal{A}$ SPANS $\mathbb{R}^d$
ALGORITHM 3 (OURS)	$\tilde{O}(s^{2/3}H_*^{2/3}n^{2/3})$	✓	$\mathcal{A}$ SPANS $\mathbb{R}^d$
THEOREM 5 (OURS)	$\Omega(s^{2/3}\kappa^{-2/3}n^{2/3})$	✓	$\mathcal{A}$ SPANS $\mathbb{R}^d$
HAO ET AL. [12]	$\tilde{O}(\sqrt{\mathcal{C}_{\min}^{-1}sn})$	✗	$\mathcal{A}$ SPANS $\mathbb{R}^d$, MIN. SIGNAL
ALGORITHM 4 (OURS)	$\tilde{O}(\sqrt{sn})$	✗	$\mathcal{A}$ SPANS $\mathbb{R}^d$, MIN. SIGNAL

Table 1: Regret bounds of our work and the prior art where s, d, n are the sparsity level, the feature dimension, and the number of rounds, respectively. The quantities $\mathcal{C}_{\min}$ and H_*^2 are the constants that captures the geometry of the action set (see Eq. (6) and (5)), and κ is a parameter for a specific family of arm sets that satisfies $\kappa^{-2} = \Theta(\mathcal{C}_{\min}^{-1}) = \Theta(H_*^2)$. In general, $H_*^2 \le \mathcal{C}_{\min}^{-1} \le \mathcal{C}_{\min}^{-2}$ (Propositon 2).

that allow a generic arm set yet enjoy nonvacuous bounds in the data-poor regime by exploiting problem-dependent characteristics?

The seminal work of Hao et al. [12] provides a positive answer to this question. They propose algorithms that enjoy nonvacuous regret bounds with an arbitrary arm set in the data poor regime using Lasso. Specifically, they have obtained a regret bound of $\tilde{O}(\mathcal{C}_{\min}^{-2/3}s^{2/3}n^{2/3})$ where $\mathcal{C}_{\min}$ is an arm-set-dependent quantity. However, their work still have left a few open problems. First, their regret upper bound does not match with their lower bound $\Omega(\mathcal{C}_{\min}^{-1/3}s^{1/3}n^{2/3})$. Second, it is not clear if $\mathcal{C}_{\min}$ is the right problem-dependent constant that captures the geometry of the arm set.

In this paper, we make a significant progress in high-dimensional linear regression and sparse linear bandits, which resolves or partly answers the aforementioned open problems.

First (Section 3), we propose a novel and computationally efficient estimator called PopArt (POPulation covariance regression with hARd Thresholding) that enjoys a tighter ℓ_1 norm recovery bound than the de facto standard sparse linear regression method Lasso in many problems. Motivated by the ℓ_1 norm recovery bound of PopArt, we develop a computationally-tractable design of experiment objective for finding the sampling distribution that minimize the error bound of PopArt, which is useful in settings where we have control on the sampling distribution (such as compressed sensing). Our design of experiments results in an ℓ_1 norm error bound that depends on the measurement set dependent quantity denoted by H_*^2 (see Eq. (5) for precise definition) that is provably better than $\mathcal{C}_{\min}^{-1}$ that appears in the ℓ_1 norm error bound used in Hao et al. [12], thus leading to an improved planning method for sparse linear bandits. **Second** (Section 4), Using PopArt, we design new algorithms for the sparse linear bandit problem, and improve the regret upper bound of prior work [12]; see Table 1 for the summary. **Third** (Section 5), We prove a matching lower bound in data-poor regime, showing that the regret rate obtained by our algorithm is optimal. The key insight in our lower bound is a novel application of the algorithmic symmetrization technique [21]. Unlike the speculation of Hao et al. [12, Remark 4.5], the improvable part was not the algorithm but the lower bound for sparsity s.

We empirically verify our theoretical findings in Section 6 where PopArt shows a favorable performance over Lasso. Finally, we conclude our paper with future research enabled by PopArt in Section 7. For space constraint, we discuss related work in Appendix A but closely related studies are discussed in depth throughout the paper.

2 Problem Definition and Preliminaries

Sparse linear bandits. We study the sparse linear bandit learning setting, where the learner is given access to an action space $\mathcal{A} \subset \{a \in \mathbb{R}^d : \|a\|_\infty \le 1\}$, and repeatedly interacts with the environment as follows: at each round $t = 1, \dots, n$, the learner chooses some action $A_t \in \mathcal{A}$, and receives reward feedback $r_t = \langle \theta^*, A_t \rangle + \eta_t$, where $\theta^* \in \mathbb{R}^d$ is the underlying reward predictor, and η_t is an independent zero-mean σ-subgaussian noise. We assume that θ^* is s-sparse; that is, it has at most s nonzero entries. The goal of the learner is to minimize its pseudo-regret defined as

$$\mathrm{Reg}(n) = n \max_{a \in \mathcal{A}} \langle \theta^*, a \rangle - \sum_{t=1}^{n} \langle \theta^*, A_t \rangle.$$

Experimental design for linear regression. In the experimental design for linear regression problem, one has a pool of unlabeled examples $\mathcal{X}$, and some underlying predictor θ^* to be learned. Querying the label of x, i.e. conducting experiment x, reveals a random label $y = \langle\theta^*, x\rangle + \eta$ associated with it, where η is a zero mean noise random variable. The goal is to accurately estimate θ^*, while using as few queries x as possible.

Definition 1. (Population covariance matrix Q) Let $\mathcal{P}(\mathcal{A})$ be the space of probability measures over $\mathcal{A}$ with the Borel σ-algebra, and define the population covariance matrix for the distribution $\mu \in \mathcal{P}(\mathcal{A})$ as follows:

$$Q(\mu) := \int_{a\in\mathcal{A}} aa^\top d\mu(a) \tag{1}$$

Classical approaches for experimental design focus on finding a distribution μ such that its induced population covariance matrix $Q(\mu)$ has properties amenable for building a low-error estimator, such as D-, A-, G-optimality [10].

Compatibility condition for Lasso. For a positive definite matrix $\Sigma \in \mathbb{R}^{d\times d}$ and a sparsity level $s \in [d] := \{1, \ldots, d\}$, we define its compatibility constant ϕ_0^2 as follows:

$$\phi_0^2 := \min_{S\subseteq[d]:|S|=s} \min_{v:\|v_S\|_1\le 3\|v_{-S}\|_1} \frac{sv^\top \Sigma v}{\|v_S\|_1^2}, \tag{2}$$

where $v_S \in \mathbb{R}^d$ denotes the vector that agrees with v in coordinates in S and 0 everywhere else and $v_{-S} \in \mathbb{R}^d$ denotes $v - v_S$.

Notation. Let e_i be the i-th indicator vector. We define $[x] = \{1, 2, \ldots, x\}$. Let $\mathrm{supp}(\theta)$ be the set of coordinate indices i where $\theta_i \neq 0$. We use $a \lesssim b$ to denote that there exists an absolute constant c such that $a \le cb$.

3 Improved Linear Regression and Experimental Design for Sparse Models

In this section, we discuss our novel sparse linear estimator PopArt for the setting where the population covariance matrix is known and show its strong theoretical properties. We then present a variation of PopArt called Warm-PopArt that amends a potential weakness of PopArt, followed by our novel experimental design for PopArt and discuss its merit over prior art.

PopArt (POPulation covariance regression with hARd Thresholding). Unlike typical estimators for the statistical learning setup, our main estimator PopArt described in Algorithm 1 takes the population covariance matrix denoted by Q as input. We summarize our assumption for PopArt.

Assumption 1. (Assumptions on the input of PopArt) There exists μ such that the input data points $\{(X_t, Y_t)\}_{t=1}^n$ satisfy that $X_t \overset{\text{i.i.d.}}{\sim} \mu$ and $Q = Q(\mu) := \mathbb{E}_{X\sim\mu}[XX^\top]$. Furthermore, $Y_t = \langle\theta^*, X_t\rangle + \eta_t$ with η_t being zero-mean σ-subgaussian noise. Also, $R_0 \ge \max_{a\in\mathcal{A}} |\langle a, \theta^* - \theta_0\rangle|$.

PopArt consists of several stages. In the first stage, for each (X_t, Y_t) pair, we create a one-sample estimator $\tilde\theta_t$ (step 4). The estimator, $\tilde\theta_t$, can be viewed as a generalization of doubly-robust estimator [7, 9] for linear models. Specifically, it is the sum of two parts: one is the pilot estimator θ_0 that is a hyperparameter of PopArt; the other is $Q(\mu)^{-1}X_t(Y_t - \langle X_t, \theta_0\rangle)$, an unbiased estimator of the difference $\theta^* - \theta_0$. Thus, it is not hard to see that $\tilde\theta_t$ is an unbiased estimator of θ^*. As we will see in Theorem 1, the variance of $\tilde\theta_t$ is smaller when θ_0 is closer to θ^*, showing the advantage of allowing a pilot estimator θ_0 as input. If no good pilot estimator is available a priori, one can set $\theta_0 = 0$.

From the discussion above, it is natural to take an average of $\tilde\theta_t$. Indeed, when n is large, the population covariance matrix Q is close to empirical covariance matrix $\hat Q := \frac{1}{n}\sum_{t=1}^n X_t X_t^\top$, which makes $\hat\theta_{\text{avg}} := \frac{1}{n}\sum_{t=1}^n \tilde\theta_t$ close to the ordinary least squares estimator $\hat\theta_{\text{OLS}} = \hat Q^{-1}(\frac{1}{n}\sum_{t=1}^n X_t Y_t)$. However, for technical reasons, the concentration property of $\tilde\theta_{\text{avg}}$ is hard to establish. This motivates PopArt's second stage (step 6), where, for each coordinate $i \in [d]$, we employ Catoni's estimator [19] (see Appendix B for a recap) to obtain an intermediate estimate for each θ_i^*, namely θ_i'.

To use Catoni's estimator, we need to have an upper bound of the variance of θ_i' for its α_i parameter. A direct calculation yields that, for all $i \in [d]$ and $t \in [n]$, $\mathrm{Var}(\tilde\theta_{ti}) \le \left(\max_{a\in\mathcal{A}} \langle\theta^* - \theta_0, a\rangle^2 + \sigma^2\right) \max_i (Q(\mu)^{-1})_{ii}$ where $\tilde\theta_{ti} := \langle\tilde\theta_t, e_i\rangle$. This implies that

Algorithm 1 PopArt (POPulation covariance regression with hARd Thresholding)

1: **Input:** Samples $\{(X_t, Y_t)\}_{t=1}^n$, the population covariance matrix $Q \in \mathbb{R}^{d\times d}$, pilot estimator $\theta_0 \in \mathbb{R}^d$, an upper bound R_0 of $\max_{a\in\mathcal{A}} |\langle a, \theta^* - \theta_0\rangle|$, failure rate δ.
2: **Output:** estimator $\hat{\theta}$
3: **for** $t = 1, \dots, n$ **do**
4: $\quad\tilde{\theta}_t = Q^{-1}X_t(Y_t - \langle X_t, \theta_0\rangle) + \theta_0$
5: **end for**
6: $\forall i \in [d], \theta'_i = \mathsf{Catoni}(\{\tilde{\theta}_{ti} := \langle\tilde{\theta}_t, e_i\rangle\}_{t=1}^n, \alpha_i, \frac{\delta}{2d})$ where $\alpha_i := \sqrt{\frac{2\log\frac{1}{\delta}}{n(R_0^2+\sigma^2)(Q^{-1})_{ii}(1+\frac{2\log\frac{1}{\delta}}{n-2\log\frac{1}{\delta}})}}$
7: $\hat{\theta} \leftarrow \mathsf{clip}_\lambda(\theta') := [\theta'_i \mathbb{1}(|\theta'_i| > \lambda_i)]_{i=1}^d$ where λ_i is defined in Proposition 1.
8: **return** $\hat{\theta}$

$\left(R_0^2 + \sigma^2\right)\max_i (Q(\mu)^{-1})_{ii}$ is an upper bound of $\mathrm{Var}(\tilde{\theta}_{ti})$. By the standard concentration inequality of Catoni's estimator (see Lemma 1), we obtain the following estimation error guarantee for θ'_i; the proof can be found in Appendix C.1. Hereafter, all proofs are deferred to appendix unless noted otherwise.

Proposition 1. Suppose Assumption 1 holds. In PopArt, for $i \in [d]$, if $n \geq 2\ln\frac{2d}{\delta}$, the following inequality holds with probability $1 - \frac{\delta}{d}$:

$$|\theta'_i - \theta^*_i| < \sqrt{\frac{4(R_0^2+\sigma^2)(Q(\mu)^{-1})_{ii}^2}{n}\log\frac{2d}{\delta}} =: \lambda_i$$

Proposition 1 shows that, for each coordinate i, $(\theta'_i - \lambda_i, \theta'_i + \lambda_i)$ forms a confidence interval for θ^*_i. Therefore, if $0 \notin (\theta'_i - \lambda_i, \theta'_i + \lambda_i)$, we can infer that $\theta^*_i \neq 0$, i.e., $i \in \mathrm{supp}(\theta^*)$. Based on the observation above, PopArt's last stage (step 7) performs a hard-thresholding for each of the coordinates of θ', using the threshold λ_i for coordinate i. Thanks to the thresholding step, with high probability, $\hat{\theta}$'s support is contained in that of θ^*, which means that all coordinates i outside the support of θ^* (typically the vast majority of the coordinates when $s \ll d$) satisfy $\hat{\theta}_i = \theta^*_i = 0$. Meanwhile, for coordinate i's in $\mathrm{supp}(\theta^*)$, the estimated value $\hat{\theta}_i$ is not too far from θ^*_i.

The following theorem states PopArt's estimation error bound in terms of its output $\hat{\theta}$'s ℓ_∞, ℓ_0, and ℓ_1 errors, respectively. We remark that replacing hard thresholding in the last stage with soft thresholding enjoys similar guarantees.

Theorem 1. Take Assumption 1. Let $H^2(Q) := \max_{i\in[d]}(Q^{-1})_{ii}$. Then, PopArt has the following guarantees with probability at least $1 - \delta$:

(i) $\forall i \in [d], |\hat{\theta}_i - \theta^*_i| < 2\sqrt{\frac{4(R_0^2+\sigma^2)(Q(\mu)^{-1})_{ii}}{n}\log\frac{2d}{\delta}}$ so $\|\hat{\theta} - \theta^*\|_\infty < 2\sqrt{\frac{4(R_0^2+\sigma^2)H^2(Q(\mu))}{n}\log\frac{2d}{\delta}}$,

(ii) $\mathrm{supp}(\hat{\theta}) \subset \mathrm{supp}(\theta^*)$ so $\|\hat{\theta} - \theta^*\|_0 \leq s$,

(iii) $\|\hat{\theta} - \theta^*\|_1 \leq 2s\sqrt{\frac{4(R_0^2+\sigma^2)H^2(Q(\mu))}{n}\log\frac{2d}{\delta}}$

Interestingly, PopArt has no false positive for identifying the sparsity pattern and enjoys an ℓ_∞ error bound, which is not available from Lasso, to our knowledge. Unfortunately, a direct comparison with Lasso is nontrivial since the largest compatibility constant ϕ_0 is defined as the solution of the optimization problem (2), let alone the fact that ϕ_0 is a function of the empirical covariance matrix. While we leave further investigation as future work, our experiment results in Section 6 suggest that there might be a case where PopArt makes a meaningful improvement over Lasso.

Proof of Theorem 1. From Proposition 1 and the union bound, one can check that

$$\|\theta' - \theta^*\|_\infty < \lambda \tag{3}$$

with probability $1 - \delta$. Therefore, the coordinates in $\mathrm{supp}(\theta^*)^c$ will be thresholded out because of $\|\theta' - \theta^*\|_\infty \leq \lambda$. Therefore, (ii) holds and for all $i \in \mathrm{supp}(\theta^*)^c$, $|\hat{\theta}_i - \theta^*_i| = 0$.

By definition, $\hat{\theta} = \mathsf{clip}_\lambda(\theta')$, we can say that $\|\hat{\theta} - \theta'\|_\infty \leq \lambda$. Plus, by Eq. (3), $\|\theta' - \theta^*\|_\infty \leq \lambda$. By the triangle inequality, $\|\theta^* - \hat{\theta}\|_\infty \leq 2\lambda$. Therefore, (i) holds.

Algorithm 2 WARM-POPART

1: **Input:** Samples $\{(X_t, Y_t)\}_{t=1}^n$, the population covariance matrix $Q \in \mathbb{R}^{d\times d}$, an upper bound $R_{\max}$ of $\max_{a\in\mathcal{A}} |\langle \theta^*, a\rangle|$, number of samples n_0, failure rate δ.
2: **Output:** $\hat{\theta}$, an estimate of θ^*
3: Run POPART$(\{(X_i, Y_i)\}_{i=1}^{\lfloor n_0/2\rfloor}, Q, \vec{0}, \delta, R_{\max})$ to obtain $\hat{\theta}_0$, a coarse estimate of θ^* for the next step.
4: Run POPART$(\{(X_i, Y_i)\}_{i=\lfloor n_0/2\rfloor+1}^{n_0}, Q, \hat{\theta}_0, \delta, \sigma)$ to obtain $\hat{\theta}$, an estimate of θ^*.

Lastly, (iii) can be argued as follows:

$$\|\hat{\theta} - \theta^*\|_1 = \sum_{i\in[d]} |\hat{\theta}_i - \theta_i^*| \le \sum_{i\in\text{supp}(\theta^*)^c} 0 + \sum_{i\in\text{supp}(\theta^*)} 2\lambda \le 2s\lambda. \qquad \square$$

WARM-POPART: Improved guarantee by warmup. One drawback of the POPART estimator is that its estimation error scales with $\sqrt{R_0^2 + \sigma^2}$, which can be very large when R_0 is large. One may attempt to use the fact that POPART allows a pilot estimator θ_0 to address this issue since R_0 gets smaller as θ_0 is closer to θ^*. However, it is a priori unclear how to obtain a θ_0 close to θ^* as θ^* is the unknown parameter that we wanted to estimate in the first place.

To get around this "chicken and egg" problem, we propose to introduce a warmup stage, which we call WARM-POPART (Algorithm 2). WARM-POPART consists of two stages. For the first warmup stage, the algorithm runs POPART with the zero vector as the pilot estimator and with the first half of the samples to obtain a coarse estimator denoted by $\hat{\theta}_0$ which guarantees that for large enough n_0, $\|\hat{\theta}_0 - \theta^*\|_1 \le \sigma$. In the second stage, using $\hat{\theta}_0$ as the pilot estimator, it runs POPART on the remaining half of the samples.

The following corollary states the estimation error bound of the output estimator $\hat{\theta}$. Compared with POPART's ℓ_1 recovery guarantee, WARM-POPART's ℓ_1 recovery guarantee (Equation (4)) has no dependence on $R_{\max}$; its dependence on $R_{\max}$ only appears in the lower bound requirement for n_0.

Corollary 1. Take Assumption 1 without the condition on R_0. Assume that $R_{\max} \ge \max_{a\in\mathcal{A}} |\langle a, \theta^*\rangle|$, and $n_0 > \frac{32s^2(R_{\max}^2+\sigma^2)H^2(Q(\mu))}{\sigma^2} \log\frac{2d}{\delta}$. Then, WARM-POPART has, with probability at least $1 - 2\delta$,

$$\|\hat{\theta} - \theta^*\|_1 \le 8s\sigma\sqrt{\frac{H^2(Q(\mu))\ln\frac{2d}{\delta}}{n_0}}. \tag{4}$$

Remark 1. In Algorithm 2, we choose POPART as our coarse estimator, but we can freely change the coarse estimation step (step 3) to other principled estimation methods (such as Lasso) without affecting the main estimation error bound (4); the only change will be the lower bound requirement of n_0 to another problem-dependent constant.

Remark 2. WARM-POPART requires the knowledge of $R_{\max}$, an upper bound of $\max_{a\in\mathcal{A}} |\langle\theta^*, a\rangle|$; this requirement can be relaxed by changing the last argument of the coarse estimation step (step 3) from $R_{\max}$, to say, $\sigma n_0^{\frac{1}{4}}$; with this change, a result analogous to Corollary 1 can be proved with a different lower bound requirement of n_0.

A novel and efficient experimental design for sparse linear estimation. In the experimental design setting where the learner has freedom to design the underlying sampling distribution μ, the ℓ_1 error bound of POPART and WARM-POPART naturally motivates a design criterion. Specifically, we can choose μ that minimizes $H^2(Q(\mu))$, which gives the lowest estimation error guarantee. We denote the optimal value of $H^2(Q(\mu))$ by

$$H_*^2 := \min_{\mu\in\mathcal{P}(\mathcal{A})} \max_{i\in[d]} (Q(\mu)^{-1})_{ii}\,. \tag{5}$$

The minimization of $H^2(Q(\mu))$ is a convex optimization problem, which admits efficient methods for finding the solution. Intuitively, H_*^2 captures the geometry of the action set $\mathcal{A}$.

To compare with previous studies that design a sampling distribution for Lasso, we first review the standard ℓ_1 error bound of Lasso.

Algorithm 3 Explore then commit with WARM-POPART

1: Input: time horizon n, action set $\mathcal{A}$, warm-up exploration length n_0, failure rate δ, reward threshold parameter $R_{\max}$, an upper bound of $\max_{a\in\mathcal{A}} |\langle\theta^*, a\rangle|$.
2: Solve the optimization problem in Eq. (5) and denote the solution as μ_*
3: **for** $t = 1, \ldots, n_0$ **do**
4: Independently pull the arm A_t according to μ_* and receives the reward r_t
5: **end for**
6: Run WARM-POPART$(\{A_t\}_{t=1}^{n_0}, \{r_t\}_{t=1}^{n_0}, Q(\mu_*), \delta, R_{\max})$ to obtain $\hat{\theta}$, an estimate of θ^*.
7: **for** $t = n_0 + 1, \ldots, n$ **do**
8: Take action $A_t = \arg\max_{a\in\mathcal{A}} \langle\hat{\theta}, a\rangle$, receive reward $r_t = \langle\theta^*, A_t\rangle + \eta_t$
9: **end for**

Theorem 2. (Buhlmann and van de Geer [5, Theorem 6.1]) With probability at least $1 - 2\delta$, the ℓ_1-estimation error of the optimal Lasso solution $\hat{\theta}_{\text{Lasso}}$ [5, Eq. (2.2)] with $\lambda = \sqrt{2\log(2d/\delta)/n}$ satisfies

$$\|\hat{\theta}_{\text{Lasso}} - \theta^*\|_1 \le \frac{s\sigma}{\phi_0^2}\sqrt{\frac{2\log(2d/\delta)}{n}},$$

where ϕ_0^2 is the compatibility constant with respect to the empirical covariance matrix $\Sigma = \frac{1}{n}\sum_{t=1}^n X_t X_t^\top$ and the sparsity s in Eq. (2).

Ideally, for Lasso, experiment design which minimizes the compatibility constant will guarantee the best estimation error bound within a fixed number of samples n. However, naively, the computation of the compatibility constant is intractable since Eq. (2) is a combinatorial optimization problem which is usually difficult to compute. One simple approach taken by Hao et al. [12] is to use the following computationally tractable surrogate of ϕ_0^2:

$$\mathcal{C}_{\min} := \max_{\mu\in\mathcal{P}(\mathcal{A})} \lambda_{\min}(Q(\mu)) \tag{6}$$

With high probability $\phi_0^2 \ge \frac{\mathcal{C}_{\min}}{2}$ [14], and one can replace ϕ_0 to $\mathcal{C}_{\min}/2$ in Theorem 2 as

Corollary 2. With probability at least $1 - \exp(-cn) - 2\delta$ for some universal constant c, the ℓ_1-estimation error of the optimal Lasso solution $\hat{\theta}_{\text{Lasso}}$ satisfies

$$\|\hat{\theta}_{\text{Lasso}} - \theta^*\|_1 \le \frac{2s\sigma}{\mathcal{C}_{\min}}\sqrt{\frac{2\log(2d/\delta)}{n}}, \tag{7}$$

The following proposition shows that our estimator has a better error bound compared to the surrogate experimental design for Lasso of Hao et al. [12].

Proposition 2. We have $H_*^2 \le \mathcal{C}_{\min}^{-1} \le dH_*^2$. Furthermore, there exist arm sets for which either of the inequalities is tight up to a constant factor.

Therefore, our new estimator has ℓ_1 error guarantees at least a factor $\mathcal{C}_{\min}^{-1/2}$ better than that provided by [12], as follows: when we choose the μ as the solution of the Eq. (5), then

$$(\text{RHS of (4)}) \lesssim s\sigma H_*\sqrt{\frac{\ln(2d/\delta)}{n}} \lesssim s\sigma\mathcal{C}_{\min}^{-1/2}\sqrt{\frac{\ln(2d/\delta)}{n}} \lesssim s\sigma\mathcal{C}_{\min}^{-1}\sqrt{\frac{\ln(2d/\delta)}{n}} \lesssim (\text{RHS of (7)})$$

4 Improved Sparse Linear Bandits using WARM-POPART

We now apply our new WARM-POPART sparse estimation algorithm to design new sparse linear bandit algorithms. Following prior work [12], we adopt the classical Explore-then-Commit (ETC) framework for algorithm design, and use POPART with experimental design to perform exploration. As we will see, the tighter ℓ_1 estimation error bound of our POPART-based estimators helps us obtain an improved regret bound.

Sparse linear bandit with WARM-POPART. Our first new algorithm, Explore then Commit with WARM-POPART (Algorithm 3), proceeds as follows. For the exploration stage, which consists of the first n_0 rounds, it solves the optimization problem (5) to find μ_*, the optimal sampling distribution

Algorithm 4 Restricted phase elimination with WARM-POPART

Input: time horizon n, finite action set $\mathcal{A}$, minimum signal m, failure rate δ, reward threshold parameter $R_{\max}$, an upper bound of $\max_{a \in \mathcal{A}} |\langle \theta^*, a \rangle|$
Solve the optimization problem in Eq. 5 and denote the solutions as Q and μ_*, respectively.
Let $n_2 = \max(\frac{256\sigma^2 H_*^2}{m^2} \log \frac{2d}{\delta}, \frac{32 s^2 (R_{\max}^2 + \sigma^2) H_*^2}{\sigma^2} \log \frac{2d}{\delta})$
for $t = 1, \dots, n_2$ **do**
  Independently pull the arm A_t according to μ_* and receives the reward r_t
end for
$\hat{\theta}_2 = \text{WARM-POPART}(\{A_t\}_{t=1}^n, \{R_t\}_{t=1}^n, Q, \delta, R_{\max})$
Identify the support $\hat{S} = \mathrm{supp}(\hat{\theta}_2)$
for $t = n_2 + 1, \dots, n$ **do**
  Invoke phased elimination algorithm for linear bandits on $\hat{S}$
end for

for POPART and samples from it to collect a dataset for the estimation of θ^*. Then, we use this dataset to compute the WARM-POPART estimator $\hat{\theta}$. Finally, in the commit stage, which consists of the remaining $n - n_0$ rounds, we take the greedy action with respect to $\hat{\theta}$. We prove the following regret guarantee of Algorithm 3:

Theorem 3. If Algorithm 3 has input time horizon $n > 16\sqrt{2}\frac{R_{\max}(R_{\max}^2 + \sigma^2)^{3/2} H_*^2 s^2}{s\sigma^4} \log \frac{2d}{\delta}$, action set $\mathcal{A} \subset [-1, +1]^d$, and exploration length $n_0 = 4(s^2\sigma^2 H_*^2 n^2 \log \frac{2d}{\delta} R_{\max}^{-2})^{\frac{1}{3}}$, $\lambda_1 = 4\sigma\sqrt{\frac{H_*^2}{n_0} \log \frac{2d}{\delta}}$, then with probability at least $1 - 2\delta$, $\mathrm{Reg}(n) \le 8 R_{\max}^{1/3} (s^2\sigma^2 H_*^2 n^2 \log \frac{2d}{\delta})^{\frac{1}{3}}$.

Proof. From the Corollary 1, $\|\hat{\theta} - \theta^*\|_1 \le 2s\lambda_1$ with probability at least $1 - 2\delta$. Therefore, with probability $1 - 2\delta$,

$$\mathrm{Reg}(n) \le R_{\max} n_0 + (n - n_0)\|\hat{\theta} - \theta^*\|_1 \le R_{\max} n_0 + 2sn\lambda_1 = R_{\max} n_0 + 8sn\sigma\sqrt{\frac{H_*^2}{n_0} \log \frac{2d}{\delta}}$$

and optimizing the right hand side with respect to n_0 leads to the desired upper bound. □

Compared with Hao et al. [12]'s regret bound $\tilde{O}((R_{\max} s^2 \sigma^2 \mathcal{C}_{\min}^{-2} n^2)^{1/3})$[1] , Algorithm 3's regret bound $\tilde{O}((R_{\max} s^2 \sigma^2 H_*^2 n^2)^{1/3})$ is at most $\tilde{O}((R_{\max} s^2 \sigma^2 \mathcal{C}_{\min}^{-1} n^2)^{1/3})$, which is at least a factor $\mathcal{C}_{\min}^{\frac{1}{3}}$ smaller. As we will see in Section 5, we show that the regret upper bound provided by Theorem 3 is unimprovable in general, answering an open question of [12].

Improved upper bound with minimum signal condition. Our second new algorithm, Algorithm 4, similarly uses WARM-POPART under an additional minimum signal condition.

Assumption 2 (Minimum signal)**.** There exists some known lower bound $m > 0$ such that $\min_{j \in \mathrm{supp}(\theta^*)} |\theta_j^*| > m$.

At a high level, Algorithm 4 uses the first n_2 rounds for identifying the support of θ^*; the ℓ_∞ recovery guarantee of WARM-POPART makes it suitable for this task. Under the minimal signal condition and a large enough n_2, it is guaranteed that $\hat{\theta}_2$'s support equals exactly the support of θ^*. After identifying the support of θ^*, Algorithm 4 treats this as a s-dimensional linear bandit problem by discarding the remaining $d - s$ coordinates of the arm covariates, and and perform phase elimination algorithm therein. The following theorem provides a regret upper bound of Algorithm 4.

Theorem 4. If Algorithm 4 has input time horizon $n > \max(\frac{2^8\sigma^2 H_*^2}{m^2}, \frac{2^5 s^2 (R_{\max}^2 + \sigma^2) H_*^2}{\sigma^2}) \log \frac{2d}{\delta}$, action set $\mathcal{A} \subset [-1, 1]^d$, upper bound of the reward $R_{\max}$, then with probability at least $1 - 2\delta$, the following regret upper bound of the Algorithm 4 holds: for universal constant $C > 0$,

$$\mathrm{Reg}(n) \le \max(\frac{2^8\sigma^2 H_*^2}{m^2} \log \frac{2d}{\delta}, \frac{2^5 s^2 (R_{\max}^2 + \sigma^2) H_*^2}{\sigma^2} \log \frac{2d}{\delta}) + C\sqrt{sn \log(|\mathcal{A}|n)}$$

[1]This is implicit in [12] – they assume that $\sigma = 1$ and do not keep track of the dependence on σ.

For sufficiently large n, the second term dominates, and we obtain an $O(\sqrt{sn})$ regret upper bound. Theorem 4 provides two major improvements compared to Hao et al. [12, Algorithm 2]. First, when m is moderately small (so that the first subterm in the first term dominates), it shortens the length of the exploration phase n_2 by a factor of $s \cdot \frac{\mathcal{C}_{\min}}{H_*^2}$. Second, compared with the regret bound $\tilde{O}(\sqrt{\frac{9\lambda_{\max}(\sum_{i=1}^{n_2} A_i A_i^\top/n_2)}{\mathcal{C}_{\min}}}\sqrt{sn})$ provided by [12], our main regret term $\tilde{O}(\sqrt{sn})$ is more interpretable and can be much lower.

5 Matching lower bound

We show the following theorem that establishes the optimality of Algorithm 3. This solves the open problem of Hao et al. [12, Remark 4.5] about the optimal order of regret in terms of sparsity and action set geometry in sparse linear bandits.

Theorem 5. For any algorithm, any s, d, κ that satisfies $d \geq n^{1/3}s^{4/3}\kappa^{-2/3}$, there exists a linear bandit environment an action set $\mathcal{A}$ and a s-sparse $\theta \in \mathbb{R}^d$, such that $\mathcal{C}_{\min}(\mathcal{A})^{-1} \leq \kappa^{-2}$, $R_{\max} \leq 2$, $\sigma = 1$, and

$$\mathrm{Reg}_n \geq \Omega(\kappa^{2/3}s^{2/3}n^{2/3}) .$$

We give an overview of our lower bound proof techniques, and defer the details to Appendix F.

Change of measure technique. Generally, researchers proved the lower bound by comparing two instances based on the information theory inequalities, such as Pinsker's inequality, or Bregtanolle-Huber inequality. In this proof, we also use two instances θ and θ', but we used the change of measure technique, to help lower bound the probability of events more freely. Specifically, for any event A,

$$\mathbb{P}_\theta(A) = \mathbb{E}_\theta[\mathbb{1}_A] = \mathbb{E}_{\theta'}\left[\mathbb{1}_A \prod_{t=1}^n \frac{p_\theta(r_t|a_t)}{p_{\theta'}(r_t|a_t)}\right] \gtrsim \mathbb{E}_{\theta'}\left[\mathbb{1}_A \exp\left(-\sum_{t=1}^n \langle A_t, \theta-\theta'\rangle^2\right)\right] . \tag{8}$$

Symmetrization. We utilize the algorithmic symmetrization technique of [21], which enables us to focusing on proving lower bounds against symmetric algorithms.

Definition 2 (Symmetric Algorithm)**.** An algorithm Alg is *symmetric* if for any permutation $\pi \in$ *Sym*(d), $\theta \in \mathbb{R}^d$, $\{a_t\}_{t=1}^n \in \mathcal{A}^n$,

$$\mathbb{P}_{\theta,\mathsf{Alg}}(A_1 = a_1, \cdots, A_n = a_n) = \mathbb{P}_{\pi(\theta),\mathsf{Alg}}(A_1 = \pi(a_1), \cdots, A_n = \pi(a_n))$$

where for vector v, $\pi(v) \in \mathbb{R}^d$ denotes its permuted version that moves v_i to the $\pi(i)$-th position.

This approach can help us to exploit the symmetry of θ' to lower bound the right hand side of (8); below, $\Pi := \{\pi' : \pi(\theta') = \theta'\}$ is the set of permutations that keep θ' invariant, and A is an event invariant under Π:

$$(8) \geq \frac{1}{|\Pi|}\sum_{\pi\in\Pi}\mathbb{E}_{\theta'}\left[\mathbb{1}_A \exp(-\sum_{t=1}^n\langle\pi^{-1}(A_t), \theta-\theta'\rangle^2)\right] \geq \mathbb{E}_{\theta'}\left[\mathbb{1}_A \exp\left(-\sum_{t=1}^n \frac{1}{|\Pi|}\sum_{\pi\in\Pi}\langle\pi^{-1}(A_t), \theta-\theta'\rangle^2\right)\right]$$

which helps us use combinatorial tools over the actions for the lower bound proof.

6 Experimental results

We evaluate the empirical performance of PopArt and our proposed experimental design, along with its impact on sparse linear bandits.

For sparse linear regression and experimental design, we compare our algorithm PopArt with μ being the solution of (5) with two baselines. The first baseline denoted by $C_{\min}$-Lasso is the method proposed by Hao et al. [12] that uses Lasso with sampling distribution μ defined by (6). The second baseline is H^2-Lasso, uses Lasso with sampling distribution μ defined by (5), which is meant to observe if Lasso can perform better with our experimental design and to see how PopArt is compared with Lasso as an estimator since they are given the same data. Of course, it is favored towards PopArt as we have optimized the design for it, so our intention is to observe if there ever exists a case where PopArt works better than Lasso.

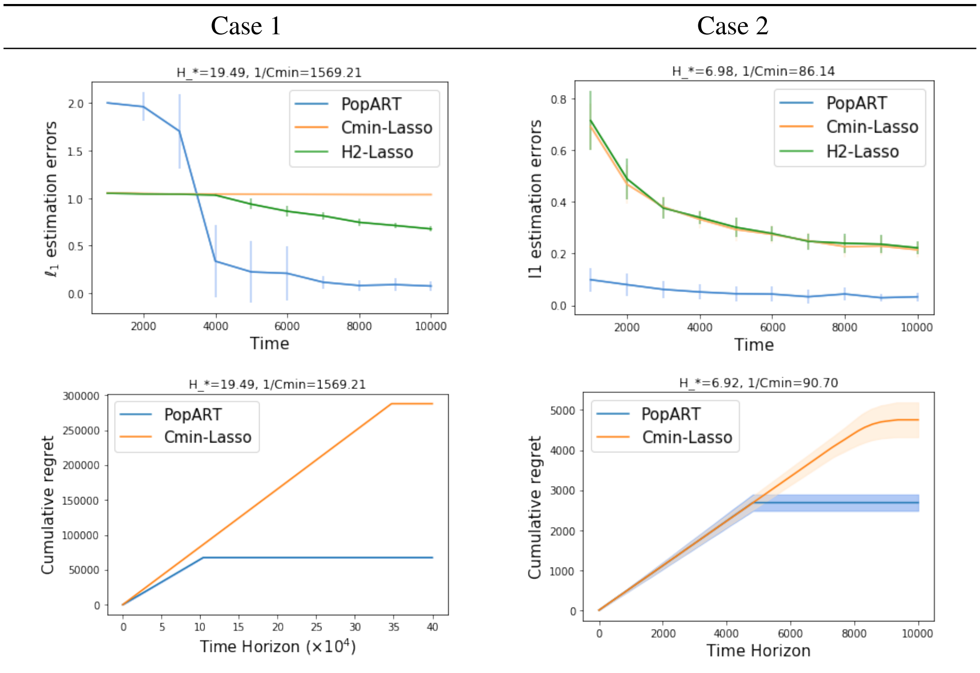

Figure 1: Experiment results on ℓ_1 estimation error cumulative regret.

For sparse linear bandits, we run a variant of our Algorithm 3 that uses WARM-POPART in place of POPART for simplicity. As a baseline, we use ESTC [12]. For both methods, we use the exploration length prescribed by theory. We consider two cases:

- **Case 1: Hard instance where $H_*^2 \ll \mathcal{C}_{\min}^{-1}$.** We use the action set constructed in Appendix D.1 where H_*^2 and $\mathcal{C}_{\min}$ shows a gap of $\Theta(d)$. We choose $d = 10$, $s = 2$, $\sigma = 0.1$.
- **Case 2. General unit vectors.** In this case, we choose $d = 30$, $s = 2$, $\sigma = 0.1$ and the action set $\mathcal{A}$ consists of $|\mathcal{A}| = 3d = 90$ uniformly random vectors on the unit sphere.

We run each method 30 times and report the average and standard deviation of the ℓ_1 estimation error and the cumulative regret in Figure 1.

Observation As we expected from the theoretical analysis, our estimator and bandit algorithm outperform the baselines. In terms of the ℓ_1 error, for both cases, we see that POPART converges much faster than $\mathcal{C}_{\min}$-Lasso for large enough n. Interestingly, H^2-Lasso also improves by just using the design computed for POPART in case 1. At the same time, H^2-Lasso is inferior than POPART even if they are given the same data points. While the design was optimized for POPART and POPART has the benefit of using the population covariance, which is unfair, it is still interesting to observe a significant gap between POPART and Lasso. For sparse linear bandit experiments, while ESTC requires exploration time almost the total length of the time horizon, ours requires a significantly shorter exploration phase in both cases and thus suffers much lower regret.

7 Conclusion

We have proposed a novel estimator POPART and experimental design for high-dimensional linear regression. POPART has not only enabled accurate estimation with computational efficiency but also led to improved sparse linear bandit algorithms. Furthermore, we have closed the gap between the lower and upper regret bound on an important family of instances in the data-poor regime.

Our work opens up numerous future directions. For POPART, we speculate that $(Q(\mu)^{-1})_{ii}$ is the statistical limit for testing whether $\theta_i^* = 0$ or not – it would be a valuable investigation to prove or disprove this. We believe this will also help investigate whether the dependence on H_*^2 in our regret upper bound is unimprovable (note our matching lower bound is only for a particular family of instances). Furthermore, it would be interesting to investigate whether we can use POPART without relying on the population covariance; e.g., use estimated covariance from an extra set of unlabeled data or find ways to use the empirical covariance directly. For sparse linear bandits, it would be interesting to develop an algorithm that achieves the data-poor regime optimal regret and data-rich regime optimal regret $\sqrt{sdn}$ simultaneously. Furthermore, it would be interesting to extend our result to changing arm set, which poses a great challenge in planning.

References

[1] Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. *Advances in neural information processing systems*, 24, 2011.

[2] Yasin Abbasi-Yadkori, David Pal, and Csaba Szepesvari. Online-to-Confidence-Set Conversions and Application to Sparse Stochastic Bandits. In *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2012.

[3] Naoki Abe and Philip M. Long. Associative reinforcement learning using linear probabilistic concepts. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 3–11, 1999.

[4] Peter Auer and M Long. Using Confidence Bounds for Exploitation-Exploration Trade-offs. *Journal of Machine Learning Research*, 3:2002, 2002.

[5] Peter Buhlmann and Sara van de Geer. *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer Publishing Company, Incorporated, 1st edition, 2011. ISBN 3642201911.

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

[7] Victor Chernozhukov, Mert Demirer, Greg Lewis, and Vasilis Syrgkanis. Semi-parametric efficient policy learning with continuous actions. *Advances in Neural Information Processing Systems*, 32, 2019.

[8] Varsha Dani, Thomas P Hayes, and Sham M Kakade. Stochastic Linear Optimization under Bandit Feedback. In *Proceedings of the Conference on Learning Theory (COLT)*, pages 355–366, 2008.

[9] Miroslav Dudík, John Langford, and Lihong Li. Doubly robust policy evaluation and learning. In *ICML*, 2011.

[10] Valerii Vadimovich Fedorov. *Theory of optimal experiments*. Elsevier, 2013.

[11] Dean P Foster and Edward I George. The risk inflation criterion for multiple regression. *The Annals of Statistics*, 22(4):1947–1975, 1994.

[12] Botao Hao, Tor Lattimore, and Mengdi Wang. High-dimensional sparse linear bandits. *Advances in Neural Information Processing Systems*, 33:10753–10763, 2020.

[13] Trevor Hastie, Robert Tibshirani, and Martin Wainwright. Statistical learning with sparsity. *Monographs on statistics and applied probability*, 143:143, 2015.

[14] Adel Javanmard and Andrea Montanari. Confidence intervals and hypothesis testing for high-dimensional regression. *The Journal of Machine Learning Research*, 15(1):2869–2909, 2014.

[15] Tor Lattimore and Csaba Szepesvári. Bandit Algorithms. 2018. URL `http://downloads.tor-lattimore.com/book.pdf`.

[16] Tor Lattimore, Koby Crammer, and Csaba Szepesvári. Linear multi-resource allocation with semi-bandit feedback. *Advances in Neural Information Processing Systems*, 28, 2015.

[17] Lihong Li, Wei Chu, John Langford, and Robert E. Schapire. A Contextual-Bandit Approach to Personalized News Article Recommendation. *Proceedings of the International Conference on World Wide Web (WWW)*, pages 661–670, 2010.

[18] Peng Liao, Predrag Klasnja, Ambuj Tewari, and Susan A Murphy. Sample size calculations for micro-randomized trials in mhealth. *Statistics in medicine*, 35(12):1944–1971, 2016.

[19] Gábor Lugosi and Shahar Mendelson. Mean estimation and regression under heavy-tailed distributions: A survey. *Foundations of Computational Mathematics*, 19(5):1145–1190, 2019.

[20] Sridhar Ramaswamy, Pablo Tamayo, Ryan Rifkin, Sayan Mukherjee, Chen-Hsiang Yeang, Michael Angelo, Christine Ladd, Michael Reich, Eva Latulippe, Jill P Mesirov, et al. Multiclass cancer diagnosis using tumor gene expression signatures. *Proceedings of the National Academy of Sciences*, 98(26):15149–15154, 2001.

[21] Max Simchowitz, Kevin Jamieson, and Benjamin Recht. The simulator: Understanding adaptive sampling in the moderate-confidence regime. In *Conference on Learning Theory*, pages 1794–1834. PMLR, 2017.

[22] Michael Woodroofe. A one-armed bandit problem with a concomitant variable. *Journal of the American Statistical Association*, 74(368):799–806, 1979.

Checklist

The checklist follows the references. Please read the checklist guidelines carefully for information on how to answer these questions. For each question, change the default **[TODO]** to [Yes] , [No] , or [N/A] . You are strongly encouraged to include a **justification to your answer**, either by referencing the appropriate section of your paper or providing a brief inline description. For example:

- Did you include the license to the code and datasets? [Yes] See Section **??**.
- Did you include the license to the code and datasets? [No] The code and the data are proprietary.
- Did you include the license to the code and datasets? [N/A]

Please do not modify the questions and only use the provided macros for your answers. Note that the Checklist section does not count towards the page limit. In your paper, please delete this instructions block and only keep the Checklist section heading above along with the questions/answers below.

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes]
   (c) Did you discuss any potential negative societal impacts of your work? [No] We are not aware of any negative impacts.
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes]
   (b) Did you include complete proofs of all theoretical results? [Yes]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [No]
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [No] But it was run on a 5-year-old laptop and no significant computation was required.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [No]
   (b) Did you mention the license of the assets? [No]

(c) Did you include any new assets either in the supplemental material or as a URL? [No]
(d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [No]
(e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [No]

5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [No]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [No]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [No]

Appendix

Table of Contents

A Related work

Sparse linear bandits. The sparse linear bandit problem is a natural extension of sparse linear regression to the bandit setup where the goal is to enjoy low regret in the high-dimensional setting by levering the sparsity of the unknown parameter θ^*. The first study we are aware of is Abbasi-Yadkori et al. [1] that achieves a $\tilde{O}(\sqrt{sdn})$ regret bound with a computationally intractable method, which is later shown to be optimal by Lattimore and Szepesvári [14, Section 24] yet is not computationally efficient. Since then, several approaches have been proposed. A large body of literature either assumes that the arm set is restricted to a continuous set (e.g., a norm ball) [4, 15] or that the set of available arms at every round is drawn in a time-varying manner, and playing arms greedily still induces a 'nice' arm distribution such as satisfying compatibility or restricted eigenvalue conditions [2, 13, 21, 17]. These assumptions allow them to leverage existing theoretical guarantees of Lasso. In contrast, we follow Hao et al. [10] and consider arm sets that are fixed throughout the bandit game without making further assumptions about the arm set. While this setup is interesting in its own for not having restrictive assumptions, it is also an important stepping stone towards efficient bandit algorithms for the more generic yet challenging setup of changing arm sets without any distributional assumptions. Our work is a direct improvement over Hao et al. [10], in that we close the gap between upper and lower bounds on the optimal worst-case regret; we refer to Table 1 for a detailed comparison.

Sparse linear regression. Natural attempts for solving sparse linear bandits are to turn to existing results from sparse linear regression. While best subset selection (BSS) is a straightforward approach of trying all the possible sparsity patterns that achieves good guarantees, its computational complexity is prohibitive [8]. As a computationally efficient alternative, Lasso is arguably the most popular approach for sparse linear regression for its simplicity and effectiveness [24]. However, Lasso has an inferior ℓ_1 norm error bound than BSS, perhaps due to its bias [25]. Rather than turning to existing results from sparse linear regression, we propose a novel estimator, PopArt, by leveraging the fact that the setup allows us to design the sampling distribution, which allows a better ℓ_1 norm error bound than Lasso except for the dependence on the range of the mean response variable.

Experimental design. In the linear bandit field, researchers often use experimental design to get the best estimator within the limited budget [22, 23, 3, 7, 16]. Especially, there were a few attempts using the population covariance based estimator instead of the traditional empirical covariance matrix [16, 23]. However, our study is the first approach that designs the experiment for minimizing the variance of each coordinate of the estimator uniformly, to the best of our knowledge.

For experimental design for sparse linear regression, Ravi et al. [18] propose heuristic approaches that ensures the design distribution satisfy incoherence conditions and restricted isometry property (RIP). Eftekhari et al. [6] study the design of c-optimal experiments in sparse regression models, where the goal is to estimate $\langle c, \theta^* \rangle$ with low error for some $c \in \mathbb{R}^d$; our experimental design task can be seen as simultaneously estimating $\langle c, \theta^* \rangle$ for all $c = e_1, \ldots, e_d$. Huang et al. [11] propose algorithms for optimal experimental design, tailored to minimizing the asymptotic variance of the debiased Lasso estimator [12]. In contrast, our results are based on finite-sample analyses.

In the theoretical computer science literature, a line of work on sketching also provides provably compressed sensing and sparse recovery algorithms [See 9, for an overview]; however, they mostly focus on using measurements (covariates) that are in $\{0,1\}^d$ and $\{-1,1\}^d$, as opposed to general measurement sets in $\mathbb{R}^d$.

B Catoni's Estimator

Definition 3 (Catoni's estimator [5])**.** For the i.i.d random variables $Z_1, \cdots, Z_n$, Catoni's mean estimator Catoni$(\{Z_i\}_{i=1}^n, \delta, \alpha)$ with error rate δ and the weight parameter α is defined as the unique value y which satisfies

$$\sum_{i=1}^n \psi(\alpha(Z_n - y)) = 0$$

where $\psi(x) \coloneqq \text{sign}(x)\log(1 + |x| + x^2/2)$.

Lemma 1 (Catoni's estimator concentration inequality [5])**.** For the i.i.d random variable $X_1, \cdots, X_n$ with mean μ, let $\hat{\mu}$ be their Catoni's estimator with error rate δ with the weight parameter $\alpha \coloneqq \sqrt{\frac{2\log\frac{1}{\delta}}{n\text{Var}(X_1)(1+\frac{2\log\frac{1}{\delta}}{n-2\log\frac{1}{\delta}})}}$. Then with probability at least $1 - 2\delta$, the following inequality holds:

$$|\hat{\mu} - \mu| < \sqrt{\frac{2\text{Var}(X_1)\log\frac{1}{\delta}}{n - \log\frac{1}{\delta}}}$$

C Proofs for PopArt and Warm-PopArt

C.1 Proof for Proposition 1

Proof. To lighten the notation, in this proof, we write $Q = Q(\mu)$, and let $(\tilde{\theta}, A, \eta)$ denote random vectors distributed identically to $(\tilde{\theta}_1, A_1, \eta_1)$. First, observe that $\mathbb{E}\left[\tilde{\theta}\right] = \theta^*$. We now use the law of total variance to decompose the covariance matrix of $\tilde{\theta}$, by first conditioning on X:

$$\mathbb{E}\left[(\tilde{\theta} - \theta^*)(\tilde{\theta} - \theta^*)^\top\right] = \mathbb{E}\left[(\mathbb{E}[\tilde{\theta} \mid X] - \theta^*)(\mathbb{E}[\tilde{\theta} \mid X] - \theta^*)^\top\right] + \mathbb{E}\left[(\tilde{\theta} - \mathbb{E}[\tilde{\theta} \mid X])(\tilde{\theta} - \mathbb{E}[\tilde{\theta} \mid X])^\top\right]$$

For the first term,

$$\begin{aligned}
\mathbb{E}\left[(\mathbb{E}[\tilde{\theta} \mid X] - \theta^*)(\mathbb{E}[\tilde{\theta} \mid X] - \theta^*)^\top\right] &= \mathbb{E}\left[(\mathbb{E}[\tilde{\theta} \mid X])(\mathbb{E}[\tilde{\theta} \mid X])^\top\right] - (\theta^* - \theta_0)(\theta^* - \theta_0)^\top \\
&\leq \mathbb{E}\left[(\mathbb{E}[\tilde{\theta} \mid X])(\mathbb{E}[\tilde{\theta} \mid X])^\top\right] = \mathbb{E}\left[Q^{-1}X(X^\top(\theta^* - \theta_0))^2 X^\top Q^{-1}\right] \\
&\leq R_0^2\,\mathbb{E}\left[Q^{-1}XX^\top Q^{-1}\right] = R_0^2 Q^{-1}
\end{aligned}$$

For the second term,

$$\mathbb{E}\left[(\tilde{\theta} - \mathbb{E}[\tilde{\theta} \mid X])(\tilde{\theta} - \mathbb{E}[\tilde{\theta} \mid X])^\top\right] = \mathbb{E}[Q^{-1}XX^\top Q^{-1}\eta^2] = \sigma^2 Q^{-1}.$$

Combining the above two bounds, we have $\mathrm{Var}(\tilde{\theta}) = \mathbb{E}\left[(\tilde{\theta}-\theta^*)(\tilde{\theta}-\theta^*)^\top\right] \le (R_0^2+\sigma^2)Q^{-1}$. Therefore, we can bound $\mathrm{Var}(\tilde{\theta}_i)$ as follows:

$$\mathrm{Var}(\tilde{\theta}_i) = \mathbb{E}[(e_i^\top(\tilde{\theta}-\theta^*))^2] \le (R_0^2+\sigma^2)(Q^{-1})_{ii}$$

By the theoretical guarantee of the Catoni's estimator (Lemma 1 in the Appendix), the desired inequality holds. □

C.2 Full version of Corollary 1 and its proof

Corollary 3. If WARM-POPART receives inputs $\{X_t, Y_t\}_{t=1}^{n_0}$ drawn from μ, $Q(\mu)$, failure probability δ, and $R_{\max}$ such that $R_{\max} \ge \max_{a\in\mathcal{A}} |\langle a, \theta^*\rangle|$, and $n_0 > \frac{32s^2(R_{\max}^2+\sigma^2)H^2(Q(\mu))}{\sigma^2}\log\frac{2d}{\delta}$, then all the following items hold with probability at least $1-2\delta$:

(i) $\|\hat{\theta}-\theta^*\|_\infty \le 8\sigma H(Q)\sqrt{\frac{\ln\frac{2d}{\delta}}{n_0}}$

(ii) $\mathrm{supp}(\hat{\theta}) \subset \mathrm{supp}(\theta^*)$ so $\|\hat{\theta}-\theta^*\|_0 \le s$

(iii) $\|\hat{\theta}-\theta^*\|_1 \le 8s\sigma H(Q)\sqrt{\frac{\ln\frac{2d}{\delta}}{n_0}}$

Proof. Since n_0 is sufficiently large, from the Theorem 1 with $R_0 = R_{\max}$ we can say that $\|\theta_0-\theta^*\|_1 \le \sigma$. Applying Theorem 1 again with $R_0 = \sigma$ we can get all (i), (ii), (iii) directly. □

D Proof of Proposition 2

First, we will prove $H_*^2 \le \mathcal{C}_{\min}^{-1} \le dH_*^2$. We will deal with equality in the next subsection.

Proof. For any positive definite matrix $Q \in \mathbb{R}^{d\times d}$,

$$H^2(Q) = \max_{i\in[d]}(Q^{-1})_{ii} = \max_{i\in[d]} e_i^\top Q^{-1} e_i \le \max_{v\in\mathbb{S}^{d-1}} v^\top Q^{-1} v = \lambda_{\max}(Q(\pi)^{-1}) \le \mathrm{tr}(Q(\pi)^{-1}) \le dH^2(Q) \tag{9}$$

Now, let the solution of the Eq. (5) and Eq. (6) as μ_H and μ_C, respectively. Then, by the rightmost inequality of (9) we have

$$\frac{1}{\mathcal{C}_{\min}} \le \lambda_{max}(Q(\mu_H)^{-1}) \le dH_*^2$$

and by the leftmost inequality of the (9) we have

$$H_*^2 \le H^2(Q(\mu_C)) \le \frac{1}{\mathcal{C}_{\min}}$$

Therefore, the inequality part of the Proposition 2 holds. □

D.1 Check the case of when the equality holds

For an example of $H_*^2 = \mathcal{C}_{\min}^{-1}$, consider $\mathcal{A} = \{e_i | i = 1, \cdots, d\}$; it can be seen that $H_*^2 = \mathcal{C}_{\min}^{-1} = d$.

We now present an example where $\mathcal{C}_{\min} = \Theta(dH_*^2)$.

Consider $\mathcal{A} = \{a_1, \ldots, a_d\}$, where

$$a_1 = \frac{1}{\sqrt{d}} e_1$$
$$a_i = e_1 + \frac{1}{\sqrt{d}} e_i.$$

and we will calculate $H^2(Q(\pi))$ and $\lambda_{\min}(Q(\pi))$ for the optimal sampling distributions π to achieve H_*^2 and $\mathcal{C}_{\min}$, respectively.

D.1.1 Prove that the optimal π satisfies $\pi(a_2) = \pi(a_3) = \cdots = \pi(a_d)$

We will first show that for both objectives $H^2(Q(\pi))$ and $\lambda_{\min}(Q(\pi))$, there exists an optimal sampling distribution π such that $\pi(a_2) = \pi(a_3) = \cdots = \pi(a_d)$.

Denote by $a := \pi(a_1)$. Fix a. For notational convenience, let $\pi(a_i) := b_i$ and $\mathbf{b} = (b_2, b_3, \cdots, b_d) \in \mathbb{R}^{d-1}$. Then the covariance matrix $Q(\pi)$ (abbreviated as Q) has the following form:

$$Q = \begin{bmatrix} \frac{a}{d} + \sum b_i & \frac{b_2}{\sqrt{d}} & \cdots & \frac{b_d}{\sqrt{d}} \\ \frac{b_2}{\sqrt{d}} & & & \\ \vdots & & \frac{1}{d} Diag(\mathbf{b}) & \\ \frac{b_d}{\sqrt{d}} & & & \end{bmatrix} \tag{10}$$

After some calculation, one can get the determinant

$$det(Q) = \frac{a(\Pi_{i=2}^d b_i)}{d^d}$$

and the cofactor

$$C_{ii} = \begin{cases} (\frac{\Pi_{i=1}^d b_i}{d^{d-1}}) & \text{if } i = 1 \\ (\frac{a}{d} + b_i)(\frac{\Pi_{s=2}^d b_s}{b_i d^{d-2}}) & \text{if } i = 2, \cdots, d \end{cases}$$

and therefore

$$(Q^{-1})_{ii} = \begin{cases} (\frac{d}{a}) & \text{if } i = 1 \\ (\frac{a}{d} + b_i) d^2/(a b_i) & \text{if } i = 2, \cdots, d \end{cases}$$

When a is a fixed parameter, $(Q^{-1})_{ii} = \frac{d^2}{a} + \frac{d}{b_i}$ and therefore the $\arg\max_i (Q^{-1})_{ii} = \arg\min b_i$. Under the constraint $\sum_{i=2}^d b_i = 1 - a$, the optimal solution is reached when $b_2 = b_3 = \cdots = b_d$.

For the C_{min} case, we will utilize symmetry of $\lambda_{\min}(Q(\pi))$. Note that $\lambda_{min}(Q)$ is a concave function w.r.t Q. Suppose that the $(b_2', b_3', \cdots, b_d') = \arg_b \max \lambda_{min}(Q(b))$. Then from the symmetry, for any cyclic permutation P, all $(b'_{P^i(2)}, b'_{P^i(3)}, \cdots, b'_{P^i(d)})$ $i = 1, \cdots, d-1$ are also the maximum. Therefore, by the Jensen's inequality,

$$C_{min} = \frac{1}{d} \sum_{i=0}^{d-1} \lambda_{min}(a, b'_{P^i(2)}, b'_{P^i(3)}, \cdots, b'_{P^i(d)}) \le \lambda_{min}(a, \frac{1-a}{d-1}, \frac{1-a}{d-1}, \cdots, \frac{1-a}{d-1}) \le C_{min}$$

Therefore, we can conclude $b_2 = b_3 = \cdots = b_d = \frac{1-a}{d-1}$ is indeed an optimal choice when a is fixed.

From now on, consider only the strategy π that satisfies $\pi(a_2) = \pi(a_3) = \cdots = \pi(a_d)$ for this section, and let $a = \pi(a_1)$ and $b = \pi(a_2)$. Then $a + (d-1)b = 1$. Now the covariance matrix induced by π is of the following form:

$$Q = \begin{bmatrix} \frac{a}{d} + (d-1)b & \frac{b}{\sqrt{d}} & \cdots & \frac{b}{\sqrt{d}} \\ \frac{b}{\sqrt{d}} & & & \\ \vdots & & \frac{b}{d} I_{d-1} & \\ \frac{b}{\sqrt{d}} & & & \end{bmatrix} \tag{11}$$

D.1.2 Calculating M^2

One can calculate $\det(Q) = \frac{a}{d}(\frac{b}{d})^{d-1}$ (using again the cofactor method) and the cofactor

$$C_{ii} = \begin{cases} (\frac{b}{d})^{d-1} & \text{if } i = 1 \\ (\frac{a}{d} + b)(\frac{b}{d})^{d-2} & \text{otherwise} \end{cases}$$

and therefore

$$(Q^{-1})_{ii} = \begin{cases} (\frac{d}{a}) & \text{if } i = 1 \\ (\frac{a}{d} + b)d^2/(ab) & \text{otherwise} \end{cases}$$

$(Q^{-1})_{ii}$ is always larger than $(Q^{-1})_{11}$, and by taking derivatives, the b that minimizes the $(Q^{-1})_{22}$ is $\sqrt{\frac{d}{d-1}} - 1 = \frac{1}{\sqrt{d-1}(\sqrt{d}+\sqrt{d-1})}$, and the corresponding $(Q^{-1})_{22} = d(\sqrt{d} + \sqrt{d-1})^2 = \Theta(d^2)$. In this case, $a = d - \sqrt{d(d-1)}$ (close to 1/2).

D.1.3 Calculating C_{min}

The eigenspectrum of Q is

$$\lambda_1, \lambda_2 = \frac{[\frac{1}{d} + \frac{(d^2-2d+2)b}{d}] \pm \sqrt{[\frac{1}{d} + \frac{(d^2-2d+2)b}{d}]^2 - \frac{4b-4(d-1)b^2}{d^2}}}{2}$$

$$\lambda_3 = \lambda_4 = \cdots = \lambda_d = \frac{b}{d}$$

Therefore, $\frac{1}{\lambda_{min}} = \max(\frac{d}{b}, \frac{d}{2b}(\frac{1+(d^2-2d+2)b+\sqrt{(1+(d^2-2d+2)b)^2-4b+4(d-1)b^2}}{1-(d-1)b}))$.

Note that for the optimal π, $a > 0$; otherwise $Q(\pi)$ is not invertible. Therefore, $b = \frac{1-a}{d-1} < \frac{1}{d-1}$. Hence, one can check that $\sqrt{(1 + (d^2 - 2d + 2)b)^2 - 4b + 4(d-1)b^2} \geq 1$, so $\frac{1}{\lambda_{min}} = \frac{d}{2b}(\frac{1+(d^2-2d+2)b+\sqrt{(1+(d^2-2d+2)b)^2-4b+4(d-1)b^2}}{1-(d-1)b})$. Now we investigate $\arg\min_b \frac{1}{\lambda_{min}(b)}$.

Recall that $a + (d-1)b = 1$, $b < 1/(d-1)$. Thus, $-4b + 4(d-1)b^2 < 0$. Therefore, the big square root term is smaller than $(1 + (d^2 - 2d + 2)b)$, which means if we let $f(b) = \frac{d(1+(d^2-2d+2)b)}{2b(1-(d-1)b)}$, then

$$f(b) \leq \frac{1}{\lambda_{min}(b)} \leq 2f(b)$$

It remains to calculate $\min_b f(b)$. A few derivative calculations show that $b^* = \arg\min_b f(b) = \frac{-(d-1)+\sqrt{d^3-1}}{d^3-1-(d-1)^2} = \Theta(d^{-3/2})$ and $\min_b f(b) = \Theta(d^3)$.

E Proofs for Sparse Linear Bandits

E.1 Proof of Theorem 4

Proof. From the (i) in Corollary 3, when $n_2 = \frac{256\sigma^2 H_*^2}{m^2} \log \frac{d}{\delta}$, with probability at least $1 - 2\delta$,

$$\|\hat{\theta} - \theta^*\|_\infty < 8\sigma\sqrt{\frac{H_*^2}{n_2} \log \frac{2d}{\delta}} = \frac{m}{2}$$

Therefore, with probability at least $1 - 2\delta$, for any index $i \in \text{supp}(\theta^*)^C$, $\hat{\theta}_i = 0$, and for any index $j \in \text{supp}(\theta^*)$, $|\hat{\theta}_j| > |\theta^*_j| - \frac{m}{2} > 0$. Thus, $\text{supp}(\theta^*) = \text{supp}(\hat{\theta})$ with probability at least $1 - 2\delta$. After that, we use the following result about the restricted phase retrieval [14]:

Theorem 6. (Lattimore and Szepesvári [14], Theorem 22.1) The n-steps regret of phase elimination algorithm satisfies

$$\text{Reg}_n \leq C\sqrt{nd\log(|\mathcal{A}|n)}$$

for an appropriately chosen universal constant $C > 0$.

□

F Proof of Lower Bound

In this section, we prove Theorem 5. We start with a restatement of it.

Theorem 7. (Restatement of the Theorem 5) For any algorithm, any s, d, κ that satisfies $s > 80, \kappa \in (0,1), d \geq \max(n^{1/3}s^{4/3}\kappa^{-4/3}, s^2)$, there exists a linear bandit environment an action set $\mathcal{A}$ and a s-sparse $\theta \in \mathbb{R}^d$, such that $\mathcal{C}_{\min}(\mathcal{A})^{-1} \leq \kappa^{-2}$, $R_{\max} \leq 2$, $\sigma = 1$, and

$$\mathrm{Reg}_n \geq \Omega(\kappa^{-2/3}s^{2/3}n^{2/3}) .$$

Remark 3. In Theorem 5 in the main text, there was a small error about the requirement of d as $d > \kappa^{-2/3}s^{4/3}n^{1/3}$, but the correct one is $d > \max(\kappa^{-4/3}s^{4/3}n^{1/3}, s^2)$. There was also a typo that the regret lower bound is $\Omega(\kappa^{2/3}s^{2/3}n^{2/3})$, but the correct bound is $\Omega(\kappa^{-2/3}s^{2/3}n^{2/3})$. We apologize for these errors.

In the lower bound instance that establishes Theorem 7, we will prove that $2\kappa^{-2} \geq \mathcal{C}_{\min}^{-1} \geq H_*^2$ (see Section F.2.7), and conclude that our $\tilde{O}(H_*^{2/3}s^{2/3}n^{2/3})$ regret upper bound of Algorithm 3 has a matching lower bound and conclude that the algorithm and the lower bound are both optimal in this setting.

For convenience, throughout the rest of this section, we prove the following slight variant of Theorem 7, where the dimensonality is $d+1$ as opposed to d, and the sparsity is $2s+1$ as opposed to s; note that the changes of these parameters do not affect the orders of the regret bounds in terms of them.

Theorem 8. For any algorithm, any s, d, κ that satisfies $s > 40, \kappa \in (0,1), d \geq \max(n^{1/3}s^{4/3}\kappa^{-4/3}, s^2)$, there exists a linear bandit environment an action set $\mathcal{A}$ and a $(2s+1)$-sparse $\theta \in \mathbb{R}^{d+1}$, such that $\mathcal{C}_{\min}(\mathcal{A})^{-1} \leq 2\kappa^{-2}$, $R_{\max} \leq 2$, $\sigma = 1$, and

$$\mathrm{Reg}_n \geq \Omega(\kappa^{-2/3}s^{2/3}n^{2/3}) .$$

Construction Following minimax lower bound and hypothesis testing terminology, we will often refer to an underlying reward predictor $\theta \in \mathbb{R}^{d+1}$ as a *hypothesis*. Let

$$\Theta_s = \left\{\theta \in \mathbb{R}^{d+1} | \theta_i \in \{-\epsilon, 0, \epsilon\} \text{ for } i \in [d], \theta_{d+1} = -1, \|\theta\|_0 = s+1\right\},$$

where $\epsilon = \kappa^{-2/3}s^{-1/3}n^{-1/3}$. We will use Θ_s and Θ_{2s} as our hypothesis space throughout the proof.

We construct a low-regret action set $\mathcal{A}$ and an informative action set $\mathcal{H}$ as follows:

$$\mathcal{I} = \left\{x \in \mathbb{R}^{d+1} | x_j \in \{-1, 0, 1\} \text{ for } j \in [d], \|x\|_1 = 2s, x_{d+1} = 0\right\}$$

$$\mathcal{H} = \left\{x \in \mathbb{R}^{d+1} | x_j \in \{-\kappa, \kappa\} \text{ for } j \in [d], |\sum_{j=1}^{d} x_j| \leq \kappa\sqrt{2d\ln 2d}, x_{d+1} = 1\right\}$$

where $\kappa \in (0,1)$ is a constant. The action set is the union $\mathcal{A} = \mathcal{I} \cup \mathcal{H}$.

The linear bandit environment parameterized by $\theta \in \mathbb{R}^{d+1}$ is defined as: given action taken A_t, its reward $r_t = \langle \theta, A_t \rangle + \eta_t$, where $\eta_t \sim N(0,1)$ is an independently drawn standard Gaussian noise. Note that by construction, η_t is σ^2-subgaussian with $\sigma = 1$.

Notations In this section, we will use $\boldsymbol{A} = (A_1, \cdots, A_n) \in \mathcal{A}^n$ as the random variable about the history of actions. For $\boldsymbol{a} \in \mathcal{A}^n$, let $T(\mathcal{H}; \boldsymbol{a}) = \sum_{t=1}^{n} \mathbb{1}(a_t \in \mathcal{H})$. For the brevity of the notation, we will write the random variable $T(\mathcal{H}; \boldsymbol{A})$ as $T(\mathcal{H})$, . Let $\mathsf{Sub}_a = \{S \subset [d] | |S| = a\}$, the set of subsets of $[d]$ which has a elements. In subsequent proofs, given a bandit algorithm Alg and an bandit environment θ, we use $\mathbb{P}_{\theta,\mathsf{Alg}}$ and $\mathbb{E}_{\theta,\mathsf{Alg}}$ to denote probability and expectation under the probability space induced by the interaction history between them. For any set of indices $S \in 2^{[d+1]}$, let $\mathrm{Sym}(S)$ be the symmetric group of the set S, and let $\Pi_S = \left\{\sigma \in \mathrm{Sym}([d+1]) : \sigma(j) = d+1\right\}$ be the set of permutations which permutes only the indices in S, and let $\Pi_{a:b} = \Pi_{\{a, a+1, \cdots, b\}}$.

Structure of the section Here is the high-level idea of the proof structure.

- Reduction using symmetrization (Section F.1) : First, we will prove that the regret lower bound of the symmetric sparse linear bandit algorithms (see Definition 6) is also the lower bound of the general sparse linear bandit algorithms (Lemma 5). Keen readers would note that our action set construction is symmetric, and this is for exploiting the symmetry. By focus on proving lower bounds for symmetric algorithms, we can exploit the favorable combinatorial properties of our action spaces to establish tighter lower bounds.
- Count the number of mistakes (Section F.2.1): Next, we will prove the core proposition of the lower bound proof, Proposition 3. This proposition can be summarized as, 'the learning agent has to pull sufficiently large number of arms in $\mathcal{H}$ (informative actions with high regret) to make less mistakes', where 'mistakes' refers to coordinates in the support of θ that has not been "touched" by the agent via pulling the low-regret arms $\mathcal{I}$ (See Equation (12) for a formal definition). This implies an inherent tension between pulling informative, high regret arms $\mathcal{H}$ and pulling low regret arms $\mathcal{I}$, which eventually leads to the desired lower bound in Theorem 5.
- Lower bound on symmetric algorithms (Section F.2.2) : Now it remains to show the proof of Proposition 3. Here, to improve the $\Omega(s^{1/3}n^{2/3})$ regret lower bound proved by Hao et al. [10] to $\Omega(s^{2/3}n^{2/3})$, we deviate from their usage of Bretagnolle-Huber inequality, and take a novel combination of various techniques such as change of measure technique, combinatorial calculation by utilizing symmetry (Claim 1).

F.1 Algorithmic symmetrization: reducing lower bounds for general algorithms to symmetric algorithms

In this section, we show how a proof of a lower bound for generic algorithms can be reduced to that of permutation-symmetric (abbrev. symmetric) (augmented) algorithms (Definition 6), specifically Lemma 5. To introduce symmetric algorithms, let us first define some useful terminology.

Definition 4 (Augmented bandit algorithm; a bandit algorithm's augmentation)**.**

1. Define an augmented bandit algorithm as: at time t, choose action A_t based on its historical observations $(A_s, r_s)_{s=1}^{t-1}$; finally it outputs $\hat{S} \subset [d]$, such that $|\hat{S}| = \frac{d}{2}$, and for all $i \in \hat{S}$,

$$\left|\left\{ j \in [d] : \sum_{t=1}^{n} \left|A_{t,j}\right| I(A_t \in \mathcal{I}) \le \sum_{t=1}^{n} \left|A_{t,i}\right| I(A_t \in \mathcal{I}) \right\}\right| \ge \frac{d}{2};$$

   in other words, all elements in $\hat{S}$ are among the top $\frac{d}{2}$ most frequently chosen coordinates (including ties) when restricted to arm pull history on $\mathcal{I}$.
2. Given a bandit algorithm Alg, define its augmentation $\widetilde{\mathsf{Alg}}$ as: at time step t, use Alg to output A_t based on all historical observations $(A_s, r_s)_{s=1}^{t-1}$; finally, output $\hat{S} \subset [d]$ as its top $\frac{d}{2}$ coordinates $i \in [d]$, sorted according to $\sum_{t=1}^{n} \left|A_{t,j}\right| I(A_t \in \mathcal{I})$ in descending order, breaking ties in dictionary order. In other words, the elements in $\hat{S}$ are the top $\frac{d}{2}$ most frequent chosen coordinates when restricted to arm pull history on $\mathcal{I}$.

Remark 4. From the above definitions, it can be readily seen that Alg's augmentation, $\widetilde{\mathsf{Alg}}$, is a valid augmented bandit algorithm.

Remark 5. $\widetilde{\mathsf{Alg}}$ outputs $\hat{S}$ by breaking ties in dictionary order. While this breaks symmetry by favoring coordinates with lower indices, as we will see in our reduction proof (proof of Lemma 5), we do not require $\widetilde{\mathsf{Alg}}$ to be symmetric (we will define symmetry momentarily in Definition 6); instead, we will work on a symmetrized version of $\widetilde{\mathsf{Alg}}$ (Definition 9).

Definition 5 (Permutaion over sets of coordinates, and vectors in $\mathbb{R}^{d+1}$)**.** Given a permutation $\sigma \in \Pi_{1:d}$:

- For a subset of coordinates $S \subset [d]$, define $\sigma(S) \subset [d]$ as $\sigma(S) := \{\sigma(i) : i \in S\}$.
- For vector $v \in \mathbb{R}^{d+1}$, define $\sigma(v) = (v_{\sigma^{-1}(1)}, \ldots, v_{\sigma^{-1}(d)}) = P_\sigma v \in \mathbb{R}^{d+1}$ as the permuted version of v using σ, where $P_\sigma = (e_{\sigma(1)}, \ldots, e_{\sigma(d)}, e_{\sigma(d+1)}) \in \mathbb{R}^{(d+1)\times(d+1)}$ is the permutation matrix[2] induced by σ and e_j denotes j-th standard basis. Note that for every $i \in [d+1]$, $\sigma(e_i) = e_{\sigma(i)}$.

[2]Here we use σ's row representation.

- For sequence of actions $\boldsymbol{a} = (a_1, \dots, a_n) \in \mathcal{A}^n$, define $\sigma(\boldsymbol{a}) = (\sigma(a_1), \dots, \sigma(a_n))$ as its permuted version using σ.

We will frequently apply the above vector permutation operation in our subsequent proofs, where the vector $v \in \mathbb{R}^{d+1}$ are often taken as actions A_t or hypotheses (underlying reward predictors) θ.

Now we are ready to define symmetric augmented bandit algorithms, a special class of bandit algorithms we will focus on.

Definition 6 (Symmetric augmented bandit algorithm)**.** An augmented bandit algorithm $\widetilde{\mathsf{AlgS}}$ is said to be *symmetric*, if for any $\sigma \in \Pi_{1:d}$ and any $\boldsymbol{a} = (a_1, \dots, a_n) \in \mathcal{A}^n$ and $U \subset [d]$,

$$\mathbb{P}_{\theta,\widetilde{\mathsf{AlgS}}}(\boldsymbol{A} = \boldsymbol{a}, \hat{S} = U) = \mathbb{P}_{\sigma(\theta),\widetilde{\mathsf{AlgS}}}(\boldsymbol{A} = \sigma(\boldsymbol{a}), \hat{S} = \sigma(U)).$$

Note that the above permutation symmetry notion is slightly different from Simchowitz et al. [20] – here we only consider permutations in $\Pi_{1:d}$, i.e., over the first d coordinates (out of all $d+1$ coordinates), whereas Simchowitz et al. [20] consider permutations over all coordinates (arms).

For symmetric augmented bandit algorithms, we have the following elementary property.

Lemma 2. For every symmetric augmented bandit algorithm $\widetilde{\mathsf{AlgS}}$, any $\sigma \in \Pi_{1:d}$ and any function $f : \mathcal{A}^n \times 2^{[d]} \to \mathbb{R}$,

$$\mathbb{E}_{\theta,\widetilde{\mathsf{AlgS}}}\left[f(\boldsymbol{A}, \hat{S})\right] = \mathbb{E}_{\sigma(\theta),\widetilde{\mathsf{AlgS}}}\left[f(\sigma^{-1}(\boldsymbol{A}), \sigma^{-1}(\hat{S}))\right]$$

Definition 7 (Permutation-invariant action space)**.** An action space $\mathcal{A}$ is said to be permutation-invariant, if for any $\pi \in \Pi_{1:d}$,

$$\pi(\mathcal{A}) \coloneqq \left\{\pi(a) : a \in \mathcal{A}\right\} = \mathcal{A}.$$

By our construction in the beginning of Section F, our action space $\mathcal{A} = \mathcal{I} \cup \mathcal{H}$ is permutation invariant.

Definition 8 (Permuted algorithm)**.** For an augmented bandit algorithm $\widetilde{\mathsf{Alg}}$ on a permutation-invariant action space $\mathcal{A}$, and a permutation $\pi \in \Pi_{1:d}$, define its π-permuted version $\widetilde{\mathsf{Alg}\pi}$ as: first permute the $[d]$ coordinates using π, and run $\widetilde{\mathsf{Alg}}$ with the permuted coordinates. Formally, at every time step t:

- $\widetilde{\mathsf{Alg}}$ outputs some action $A'_t \in \mathcal{A}$, and $\widetilde{\mathsf{Alg}\pi}$ accordingly outputs action $A_t = \pi^{-1}(A'_t) \in \mathcal{A}$
- Receives reward $r_t = \langle \theta, A_t \rangle + \eta_t = \langle \pi(\theta), \pi(A_t) \rangle + \eta_t = \langle \pi(\theta), A'_t \rangle + \eta_t$
- $\widetilde{\mathsf{Alg}}$ outputs $\hat{S}'$, and $\widetilde{\mathsf{Alg}\pi}$ outputs $\hat{S} = \pi^{-1}(\hat{S}')$.

Let $\mathcal{U}(\boldsymbol{a}) = \{U \in \mathsf{Sub}_{d/2} | \forall i \in U, \sum_{t=1}^n |a_{ti}| \mathbb{1}(a_t \in \mathcal{I}) \geq (d/2)\text{-}\max\{\sum_{t=1}^n |a_{tj}| \mathbb{1}(a_t \in \mathcal{I})\}_{j=1}^d\}$ where $k\text{-}\max S$ for a set $S \subseteq \mathbb{R}$ is the k-th largest member of S. We will mainly focus on $(\boldsymbol{a}, U) \in \mathcal{A}^n \times \mathsf{Sub}_{d/2}$ such that $U \in \mathcal{U}(\boldsymbol{a})$, but for the lemmas we keep the generality and consider any $U \subset [d]$.

The following lemma follows straightforwardly from the definition of $\widetilde{\mathsf{Alg}\pi}$:

Lemma 3.

- For any $\boldsymbol{a} = (a_1, \dots, a_n) \in \mathcal{A}^n$ and $U \subset [d]$,

$$\mathbb{P}_{\theta,\widetilde{\mathsf{Alg}\pi}}(\boldsymbol{A} = \boldsymbol{a}, \hat{S} = U) = \mathbb{P}_{\pi(\theta),\widetilde{\mathsf{Alg}}}(\boldsymbol{A} = \pi(\boldsymbol{a}), \hat{S} = \pi(U)).$$

- For any function $f : \mathcal{A}^n \times 2^{[d]} \to \mathbb{R}$,

$$\mathbb{E}_{\theta,\widetilde{\mathsf{Alg}\pi}}\left[f(\boldsymbol{A}, \hat{S})\right] = \mathbb{E}_{\pi(\theta),\widetilde{\mathsf{Alg}}}\left[f(\pi^{-1}(\boldsymbol{A}), \pi^{-1}(\hat{S}))\right].$$

Definition 9. For an augmented bandit algorithm $\widetilde{\mathsf{Alg}}$ on a permutation-invariant action space $\mathcal{A}$, define its symmetrized version $\widetilde{\mathsf{AlgP}}$ as: first, choosing π uniformly at random from $\Pi_{1:d}$, then, run $\widetilde{\mathsf{Alg}\pi}$ on the bandit environment for n rounds.

Lemma 4. We have the following:

1. $\mathbb{P}_{\theta,\widetilde{\mathsf{AlgP}}}(\cdot) = \frac{1}{|\Pi_{1:d}|} \sum_{\pi \in \Pi_{1:d}} \mathbb{P}_{\theta,\widetilde{\mathsf{Alg}\pi}}(\cdot)$, and $\mathbb{E}_{\theta,\widetilde{\mathsf{AlgP}}}[\cdot] = \frac{1}{|\Pi_{1:d}|} \sum_{\pi \in \Pi_{1:d}} \mathbb{E}_{\theta,\mathsf{Alg}\pi}[\cdot]$.

2. $\widetilde{\mathsf{AlgP}}$ is a symmetric augmented bandit algorithm.

The definition below formalizes the (pseudo-)regret notion under a specific hypothesis, which provides useful clarifications when using the averaging hammer to argue regret lower bounds.

Definition 10. Define

$$\mathrm{Reg}(n,\theta) = n \cdot \max_{a\in\mathcal{A}} \langle\theta, a\rangle - \sum_{t=1}^{n} \langle\theta, A_t\rangle$$

as the pseudo-regret of a sequence of actions $(A_t)_{t=1}^n$ under hypothesis θ.

The main result of this section is the following lemma that reduces proving lower bounds for general algorithms to proving lower bounds for symmetric augmented algorithms.

Lemma 5. If for all symmetric augmented bandit algorithms $\widetilde{\mathsf{AlgS}}$, there exists some $\theta \in \Theta_s \cup \Theta_{2s}$ such that $\mathbb{E}_{\theta,\widetilde{\mathsf{AlgS}}}\left[\mathrm{Reg}(n,\theta)\right] \geq R$, then, for all bandit algorithms Alg, there exists some $\theta' \in \Theta_s \cup \Theta_{2s}$ such that $\mathbb{E}_{\theta',\mathsf{Alg}}\left[\mathrm{Reg}(n,\theta')\right] \geq R$.

In light of this lemma, in Section F.2, we focus on showing regret lower bounds on symmetric augmented bandit algorithms under hypotheses in $\Theta_s \cup \Theta_{2s}$.

F.1.1 Deferred Proofs

Proof of Lemma 2. For any $f : \mathcal{A}^n \times 2^{[d]} \to \mathbb{R}$,

$$\begin{aligned}
\mathbb{E}_{\theta,\widetilde{\mathsf{AlgS}}}\left[f(\boldsymbol{A},\hat{S})\right] &= \sum_{(\boldsymbol{a},U)\in\mathcal{A}^n\times 2^{[d]}} \mathbb{P}_{\theta,\widetilde{\mathsf{AlgS}}}(\boldsymbol{A}=\boldsymbol{a},\hat{S}=U)f(\boldsymbol{a},U) && \text{(definition of expectation)}\\
&= \sum_{(\boldsymbol{a},U)\in\mathcal{A}^n\times 2^{[d]}} \mathbb{P}_{\sigma(\theta),\widetilde{\mathsf{AlgS}}}(\boldsymbol{A}=\sigma(\boldsymbol{a}),\hat{S}=\sigma(U))f(\boldsymbol{a},U) && \text{(symmetry)}\\
&= \sum_{(\boldsymbol{a},U)\in\mathcal{A}^n\times 2^{[d]}} \mathbb{P}_{\sigma(\theta),\widetilde{\mathsf{AlgS}}}(\sigma^{-1}(\boldsymbol{A})=\boldsymbol{a},\sigma^{-1}(\hat{S})=U)f(\boldsymbol{a},U) && \text{(algebra)}\\
&= \mathbb{E}_{\sigma(\theta),\widetilde{\mathsf{AlgS}}}\left[f(\sigma^{-1}(\boldsymbol{A}),\sigma^{-1}(\hat{S}))\right] && \text{(definition of expectation)}
\end{aligned}$$

□

Proof of Lemma 3. For the first item, denote by $\boldsymbol{A}' = (A'_1,\ldots,A'_n)$; for any $\boldsymbol{a} = (a_1,\ldots,a_n) \in \mathcal{A}^n$ and $U \subset [d]$,

$$\begin{aligned}
\mathbb{P}_{\theta,\widetilde{\mathsf{Alg}\pi}}(\boldsymbol{A}=\boldsymbol{a},\hat{S}=U) &= \mathbb{P}_{\theta,\widetilde{\mathsf{Alg}\pi}}(\pi^{-1}(\boldsymbol{A}')=\boldsymbol{a},\pi^{-1}(\hat{S}')=U) && \text{(definition of } \boldsymbol{A}'\text{)}\\
&= \mathbb{P}_{\theta,\widetilde{\mathsf{Alg}\pi}}(\boldsymbol{A}'=\pi(\boldsymbol{a}),\hat{S}'=\pi(U)) && \text{(algebra)}\\
&= \mathbb{P}_{\pi(\theta),\widetilde{\mathsf{Alg}}}(\boldsymbol{A}=\pi(\boldsymbol{a}),\hat{S}=\pi(U)) && \text{(switching to } \widetilde{\mathsf{Alg}}\text{'s perspective)}
\end{aligned}$$

The second item is the direct consequence of the first item by the following calculation.

$$\begin{aligned}
\mathbb{E}_{\theta,\widetilde{\mathsf{Alg}\pi}}\left[f(\boldsymbol{A},\hat{S})\right] &= \sum_{(\boldsymbol{a},U)\in\mathcal{A}^n\times 2^{[d]}} \mathbb{P}_{\theta,\widetilde{\mathsf{Alg}\pi}}(\boldsymbol{A}=\boldsymbol{a},\hat{S}=U)f(\boldsymbol{a},U) && \text{(definition of expectation)}\\
&= \sum_{(\boldsymbol{a},U)\in\mathcal{A}^n\times 2^{[d]}} \mathbb{P}_{\pi(\theta),\mathsf{Alg}}(\boldsymbol{A}=\pi(\boldsymbol{a}),\hat{S}=\pi(U))f(\boldsymbol{a},U) && \text{(the first item)}\\
&= \mathbb{E}_{\pi(\theta),\mathsf{Alg}}\left[f(\pi^{-1}(\boldsymbol{A}),\pi^{-1}(\hat{S}))\right]. && \text{(definition of expectation)}
\end{aligned}$$

□

Proof of Lemma 4. The first item follows from the definition of $\widetilde{\mathsf{AlgP}}$.

For the second item, for any permutation $\sigma \in \Pi_{1:d}$ and action history $\boldsymbol{a} \in \mathcal{A}^n$,

$$\mathbb{P}_{\theta,\widetilde{\mathsf{AlgP}}}(\boldsymbol{A}=\boldsymbol{a},\hat{S}=U) = \frac{1}{|\Pi_{1:d}|}\sum_{\pi\in\Pi_{1:d}} \mathbb{P}_{\theta,\widetilde{\mathsf{Alg}\pi}}(\boldsymbol{A}=\boldsymbol{a},\hat{S}=U) \qquad \text{(the first item)}$$

$$= \frac{1}{|\Pi_{1:d}|}\sum_{\pi\in\Pi_{1:d}} \mathbb{P}_{\pi(\theta),\widetilde{\mathsf{Alg}}}(\boldsymbol{A}=\pi(\boldsymbol{a}),\hat{S}=\pi(U)) \qquad \text{(Lemma 3)}$$

$$= \frac{1}{|\Pi_{1:d}|}\sum_{\pi\in\Pi_{1:d}} \mathbb{P}_{\pi\circ\sigma(\theta),\widetilde{\mathsf{Alg}}}(\boldsymbol{A}=\pi\circ\sigma(\boldsymbol{a}),\hat{S}=\pi\circ\sigma(U)) \qquad (*)$$

$$= \frac{1}{|\Pi_{1:d}|}\sum_{\pi\in\Pi_{1:d}} \mathbb{P}_{\sigma(\theta),\widetilde{\mathsf{Alg}\pi}}(\boldsymbol{A}=\sigma(\boldsymbol{a}),\hat{S}=\sigma(U)) \qquad \text{(Lemma 3)}$$

$$= \mathbb{P}_{\sigma(\theta),\widetilde{\mathsf{AlgP}}}(\boldsymbol{A}=\sigma(\boldsymbol{a}),\hat{S}=\sigma(U)), \qquad \text{(the first item)}$$

where in step (*), we use the observation that for any $\sigma\in\Pi_{1:d}$, $\{\pi\circ\sigma:\pi\in\Pi_{1:d}\}=\Pi_{1:d}$. □

Proof of Lemma 5. by assumption, we have, there exists some $\theta\in\Theta_s\cup\Theta_{2s}$,

$$\begin{aligned}
R &\le \mathbb{E}_{\theta,\widetilde{\mathsf{AlgP}}}[\mathrm{Reg}(n,\theta)] \\
&= \frac{1}{|\Pi_{1:d}|}\sum_{\pi\in\Pi_{1:d}} \mathbb{E}_{\theta,\widetilde{\mathsf{Alg}\pi}}[\mathrm{Reg}(n,\theta)] && \text{(Lemma 4)}\\
&= \frac{1}{|\Pi_{1:d}|}\sum_{\pi\in\Pi_{1:d}} \mathbb{E}_{\theta,\widetilde{\mathsf{Alg}\pi}}[n\cdot\max_{a\in\mathcal{A}}\langle\theta,a\rangle-\sum_{t=1}^{n}\langle\theta,A_t\rangle] && \text{(Definition 10)}\\
&= \frac{1}{|\Pi_{1:d}|}\sum_{\pi\in\Pi_{1:d}} \mathbb{E}_{\pi(\theta),\widetilde{\mathsf{Alg}}}[n\cdot\max_{a\in\mathcal{A}}\langle\theta,a\rangle-\sum_{t=1}^{n}\langle\theta,\pi^{-1}(A_t)\rangle] && \text{(Lemma 3)}\\
&= \frac{1}{|\Pi_{1:d}|}\sum_{\pi\in\Pi_{1:d}} \mathbb{E}_{\pi(\theta),\widetilde{\mathsf{Alg}}}[n\cdot\max_{a\in\mathcal{A}}\langle\pi(\theta),a\rangle-\sum_{t=1}^{n}\langle\pi(\theta),A_t\rangle] \\
& && \text{($\langle a,\pi^{-1}(b)\rangle=\langle\pi(a),b\rangle$, and $\mathcal{A}$'s permutation invariance)}\\
&= \frac{1}{|\Pi_{1:d}|}\sum_{\pi\in\Pi_{1:d}} \mathbb{E}_{\pi(\theta),\widetilde{\mathsf{Alg}}}[\mathrm{Reg}(n,\pi(\theta))] && \text{(Definition 10)}\\
&= \frac{1}{|\Pi_{1:d}|}\sum_{\pi\in\Pi_{1:d}} \mathbb{E}_{\pi(\theta),\mathsf{Alg}}[\mathrm{Reg}(n,\pi(\theta))] && \text{(Alg and $\widetilde{\mathsf{Alg}}$ take the same action sequence)}
\end{aligned}$$

By the pigeonhole principle, there exists $\pi\in\Pi_{1:d}$ which satisfies $\mathbb{E}_{\pi(\theta),\mathsf{Alg}}[\mathrm{Reg}(n,\pi(\theta))]\ge R$, and this $\pi(\theta)\in\Theta_s\cup\Theta_{2s}$ is the desired θ' in Lemma 5. □

F.2 Lower bound against symmetric algorithms

F.2.1 Count the number of mistake

From now on, by Lemma 5, we will focus on proving regret lower bound for any symmetric augmented algorithm $\widetilde{\mathsf{AlgS}}$ (Definition 6). For the brevity of notation, we omit the $\widetilde{\mathsf{AlgS}}$ notations from $\mathbb{P}$ and $\mathbb{E}$. For each history of action $\boldsymbol{A}\in\mathcal{A}^n$, we define the $M_\theta(\hat{S})$, the mistake respect to θ as follows:

$$M_\theta(\hat{S}) := |\mathrm{supp}(\theta)\setminus\hat{S}| \tag{12}$$

Let s be multiple of 4. For $\xi=\frac{1}{4}$, we like to show the following claim.

Proposition 3. If $P_\theta(M_\theta(\hat{S})\ge s/4)\le\xi$ for all $\theta\in\Theta_s\cup\Theta_{2s}$, then $\exists\theta'\in\Theta_s$ such that $\mathbb{E}_\theta[T(\mathcal{H})]\ge\Omega(\frac{1}{\kappa^2\epsilon^2})$

Given the above proposition, we are now ready to prove Theorem 8.

Proof of Theorem 8. If there exists a $\theta\in\Theta_s\cup\Theta_{2s}$ that satisfies $\mathbb{P}_\theta(M_\theta(\hat{S})\ge s/4)\ge\xi$, then if we note that $\sum_{t=1}^{n}|A_{tj}|\mathbb{1}(A_t\in\mathcal{I})\le\frac{4sn}{d}$ for $j\notin\hat{S}$ (if not, it contradicts $\sum_{j=1}^{d}\sum_{t=1}^{n}|A_{tj}|\mathbb{1}(A_t\in\mathcal{I})\le 2sn$), we can lower bound the regret as follows:

$$\mathbb{E}_\theta[\mathrm{Reg}_n]\ge\mathbb{E}_\theta[\mathbb{1}(M_\theta(\hat{S})\ge\frac{s}{4})\,\mathrm{Reg}_n]$$

$$\geq \xi \frac{s}{4}(n - \frac{4sn}{d})\epsilon = \Omega(\epsilon s n \xi)$$

On the other hand, if $P_\theta(M_\theta(\hat{S}) \geq s/4) \leq \xi$ for all $\theta \in \Theta_s \cup \Theta_{2s}$, then by Proposition 3 $\exists \theta'$ such that $\mathbb{E}_{\theta'}[T(\mathcal{H})] \geq \Omega(\frac{1}{\kappa^2\epsilon^2})$ so $\mathbb{E}_{\theta'}[\mathrm{Reg}_n] \geq \frac{1}{2}\mathbb{E}_{\theta'}[T(\mathcal{H})] \geq \Omega(\frac{1}{\kappa^2\epsilon^2})$. Because by construction, $\epsilon = \kappa^{-2/3}s^{-1/3}n^{-1/3}$, we get the desired regret lower bound $\Omega(\kappa^{-2/3}s^{2/3}n^{2/3})$ result for both cases. $\square$

F.2.2 Proof of Proposition 3

Let's assume that $\forall \theta \in \Theta_s \cup \Theta_{2s}, \mathbb{P}_\theta(M_\theta(\hat{S}) \geq s/4) \leq \xi$. Define

$$\theta' = (\underbrace{\epsilon, \cdots, \epsilon}_{s}, 0, \cdots, 0, -1) \in \Theta_s$$
$$\theta = (\underbrace{\epsilon, \cdots, \epsilon}_{2s}, 0, \cdots, 0, -1) \in \Theta_{2s}$$
$$\tilde{\theta} = \theta - \theta'$$

. The following two lemmas show the main advantage why we set θ' and θ in this way.

Lemma 6. Let $\phi \in \mathbb{R}^d$, $\pi \in \Pi$, $E_1, E_2 \in \mathsf{Sub}_{d/2}$ be the elements which satisfies $\pi(\phi) = \phi$ and $\pi(E_1) = E_2$. For any symmetric algorithm AlgS, we have that

$$\mathbb{P}_{\mathsf{AlgS},\phi}(\hat{S} = E_1, T(\mathcal{H}) \leq \tau) = \mathbb{P}_{\mathsf{AlgS},\phi}(\hat{S} = E_2, T(\mathcal{H}) \leq \tau)$$

.

Proof. To see this, note that:

$$\begin{aligned}
\mathbb{P}_{\mathsf{AlgS},\phi}(\hat{S} = E_1, T(\mathcal{H}) \leq \tau) &= \mathbb{E}_{\mathsf{AlgS},\phi}\left[\mathbb{1}\left(\hat{S} = E_1, \sum_{t=1}^{n} \mathbb{1}(A_t \in \mathcal{H}) \leq \tau\right)\right] \\
&= \mathbb{E}_{\mathsf{AlgS},\pi(\phi)}\left[\mathbb{1}\left(\pi^{-1}(\hat{S}) = E_1, \sum_{t=1}^{n} \mathbb{1}(\pi^{-1}(A_t) \in \mathcal{H}) \leq \tau\right)\right] \\
&= \mathbb{E}_{\mathsf{AlgS},\pi(\phi)}\left[\mathbb{1}\left(\hat{S} = \pi(E_1), T(\mathcal{H}) \leq \tau\right)\right] \\
&= \mathbb{E}_{\mathsf{AlgS},\phi}\left[\mathbb{1}\left(\hat{S} = E_2, T(\mathcal{H}) \leq \tau\right)\right]
\end{aligned}$$

$\square$

Let $Q_m := \{S \in \mathsf{Sub}_{d/2} | M_\theta(S) = m\}$ and $Q'_m = \{S \in \mathsf{Sub}_{d/2} | M_{\theta'}(S) = m\}$ be the set of size $d/2$ sets which has exactly m mistakes with θ and θ', respectively.

Lemma 7. For our θ, for each $E_1, E_2 \in Q_{\frac{1}{4}s - l}$, there exists $\pi \in \{\pi_1 \circ \pi_2 | \pi_1 \in \Pi_{1:2s}, \pi_2 \in \Pi_{2s+1:d}\}$ which satisfies $\pi(E_1) = E_2$. Similarly, for each $E'_1, E'_2 \in Q'_{\frac{3}{4}s-l}$, there exists $\pi \in \{\pi_1 \circ \pi_2 | \pi_1 \in \Pi_{1:s}, \pi_2 \in \Pi_{s+1:d}\}$ which satisfies $\pi(E_1) = E_2$.

Proof. Since $|E_1 \cap [1:2s]| = |E_2 \cap [1:2s]|$ and $|E_1 \cap [2s+1:d]| = |E_2 \cap [2s+1:d]|$, there exists $\pi \in \{\pi_1 \circ \pi_2 | \pi_1 \in \Pi_{1:2s}, \pi_2 \in \Pi_{2s+1:d}\}$ which satisfies $\pi(E_1) = E_2$. Similar proof holds also for E'_1, E'_2. $\square$

Then,

$$\mathbb{P}_\theta(M_\theta(\hat{S}) \geq s/4, T(\mathcal{H}) \leq \tau) \geq \sum_{l=0}^{\frac{1}{4}s} \mathbb{P}_\theta(M_\theta(\hat{S}) = \frac{3}{4}s - l, T(\mathcal{H}) \leq \tau)$$

Let $l \in \{1, \ldots, s/4\}$, and let $R_a = Q_{\frac{3}{4}s-a} \cap Q'_{\frac{1}{4}s-a}$. For $a \in \mathbb{R}^d$ and $r \in \mathbb{R}$, and $\phi \in \{\theta, \theta'\}$, let $p_\phi(r|a) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(r-\langle a,\phi\rangle)^2}{2}\right)$ be the probability density function of the reward r when the action a is given under hypothesis ϕ. Now pick one element $E \in R_l$. Then,

$$
\begin{aligned}
&\mathbb{P}_\theta(M = \frac{3}{4}s - l, T(\mathcal{H}) \le \tau) && (13)\\
&= \mathbb{P}_\theta(\hat{S} \in Q_{\frac{3}{4}s-l}, T(\mathcal{H}) \le \tau) && (14)\\
&= |Q_{\frac{3}{4}s-l}|\, \mathbb{P}_\theta(\hat{S} = E, T(\mathcal{H}) \le \tau) && \text{(Lemma 6 and 7)}\\
&= \frac{|Q_{\frac{3}{4}s-l}|}{|R_l|}\, \mathbb{P}_\theta(\hat{S} \in R_l, T(\mathcal{H}) \le \tau) && \text{(Lemma 6 and 7)}\\
&= \frac{|Q_{\frac{3}{4}s-l}|}{|R_l|}\, \mathbb{E}_{\theta'}[\mathbb{1}(\hat{S} \in R_l, T(\mathcal{H}) \le \tau) \prod_{t=1}^{n} \frac{p_\theta(r_t|A_t)}{p_{\theta'}(r_t|A_t)}] && \text{(change of measure)}\\
&= \frac{|Q_{\frac{3}{4}s-l}|}{|R_l|}\, \mathbb{E}_{\theta'}[\mathbb{1}(\hat{S} \in R_l, T(\mathcal{H}) \le \tau) \exp(-\sum_{t=1}^{n} \ln \frac{p_{\theta'}(r_t|A_t)}{p_\theta(r_t|A_t)})] && (15)
\end{aligned}
$$

Here, we will use the following Claim 1 to bound this empirical KL divergence to the fixed constant.

Claim 1.

$$
\begin{aligned}
&\mathbb{E}_{\theta'}[\mathbb{1}(\hat{S} \in R_l, T(\mathcal{H}) \le \tau) \exp(-\sum_{t=1}^{n} \ln \frac{p_{\theta'}(r_t|A_t)}{p_\theta(r_t|A_t)})]\\
&\qquad \ge \mathbb{P}_{\theta'}(\hat{S} \in R_l, T(\mathcal{H}) \le \tau) \exp(-KL(\rho, \delta, \tau)) - \delta^{1+\frac{1}{\rho}}
\end{aligned}
$$

where $KL(\rho, \delta, \tau) = \frac{1}{2}\epsilon^2(1+\rho)(\frac{4s^3 n}{d} + 77 s\kappa^2 \tau) + \frac{1}{\rho} \ln \frac{1}{\delta}$

Now, decide ρ and δ later and continuing from the previous inequality with $KL(\rho, \delta, \tau) = \frac{1}{2}\epsilon^2(1+\rho)(\frac{4s^3 n}{d} + 77 s\kappa^2 \tau) + \frac{1}{\rho} \ln \frac{1}{\delta}$,

$$
\begin{aligned}
(15) &\ge \frac{|Q_{\frac{3}{4}s-l}|}{|R_l|} \left(\mathbb{P}_{\theta'}(\hat{S} \in R_l, T(\mathcal{H}) \le \tau) \exp(-KL(\rho, \delta, \tau)) - \delta^{1+\frac{1}{\rho}}\right) && \text{(Claim 1)}\\
&= \frac{|Q_{\frac{3}{4}s-l}|}{|Q'_{\frac{s}{4}-l}|} \left(\mathbb{P}_{\theta'}(M_{\theta'}(\hat{S}) = \frac{s}{4} - l, T(\mathcal{H}) \le \tau) \exp(-KL(\rho, \delta, \tau))\right) - \frac{|Q_{\frac{3}{4}s-l}|}{|R_l|} \delta^{1+\frac{1}{\rho}} && \text{(Lemma 6 and 7)}\\
&= \frac{|Q_{\frac{3}{4}s-l}|}{|Q'_{\frac{s}{4}-l}|} \left(\mathbb{P}_{\theta'}(M_{\theta'}(\hat{S}) = \frac{s}{4} - l, T(\mathcal{H}) \le \tau) \exp(-KL(\rho, \delta, \tau)) - \frac{|Q'_{\frac{s}{4}-l}|}{|R_l|} \delta^{1+\frac{1}{\rho}}\right)\\
&\ge \frac{|Q_{\frac{3}{4}s-l}|}{|Q'_{\frac{s}{4}-l}|} \left(\mathbb{P}_{\theta'}(M_{\theta'}(\hat{S}) = \frac{s}{4} - l, T(\mathcal{H}) \le \tau) \exp(-KL(\rho, \delta, \tau)) - s\delta^{1+\frac{1}{\rho}}\right) && \text{(Lemma 8)}
\end{aligned}
$$

For the last inequality, we used the following lemma.

Lemma 8. For all $l \in [\frac{s}{4}]$, $\frac{|Q'_{\frac{s}{4}-l}|}{|R_l|} < s$.

In short,

$$
\mathbb{P}_\theta(M_\theta(\hat{S}) = \frac{3}{4}s - l, T(\mathcal{H}) \le \tau) \ge \frac{|Q_{\frac{3s}{4}-l}|}{|Q'_{\frac{s}{4}-l}|} \left(\mathbb{P}_{\theta'}(M_{\theta'}(\hat{S}) = \frac{s}{4} - l, T(\mathcal{H}) \le \tau) \exp(-KL(\rho, \delta, \tau)) - s\delta^{1+\frac{1}{\rho}}\right).
$$

Let $Y = \min_{l \in \frac{s}{4}} \frac{|Q_{\frac{3s}{4}-l}|}{|Q'_{\frac{s}{4}-l}|}$. Then,

$$
\mathbb{P}_\theta(M_\theta(\hat{S}) = \frac{3}{4}s - l, T(\mathcal{H}) \le \tau) \ge Y \left(\mathbb{P}_{\theta'}(M_{\theta'}(\hat{S}) = \frac{s}{4} - l, T(\mathcal{H}) \le \tau) \exp(-KL(\rho, \delta, \tau)) - s\delta^{1+\frac{1}{\rho}}\right).
$$

Summing up both sides for $l \in [\frac{s}{4}]$,

$$\xi \geq \sum_{l=1}^{\frac{s}{4}} \mathbb{P}_\theta(M = \frac{3}{4}s - l, T(\mathcal{H}) \leq \tau)$$
$$\geq Y\left(\mathbb{P}_{\theta'}(M \leq \frac{s}{4} - 1, T(\mathcal{H}) \leq \tau)\exp(-KL(\rho,\delta,\tau)) - \frac{s^2}{4}\delta^{1+\frac{1}{\rho}}\right)$$
$$\geq Y\left(\mathbb{P}_{\theta'}(M \leq \frac{s}{4} - 1) - \frac{\mathbb{E}_{\theta'}[T(\mathcal{H})]}{\tau}\right)\exp(-KL(\rho,\delta,\tau)) - \frac{s^2 Y}{4}\delta^{1+\frac{1}{\rho}}$$
$$(\mathbb{P}(A,B) \geq \mathbb{P}(A) - \mathbb{P}(\bar{B}); \text{Markov's ineq.})$$
$$\geq Y\left(1 - \xi - \frac{\mathbb{E}_{\theta'}[T(\mathcal{H})]}{\tau}\right)\exp(-KL(\rho,\delta,\tau)) - \frac{s^2 Y}{4}\delta^{1+\frac{1}{\rho}}$$
$$= Y(1-2\xi)\exp(-KL(\rho,\delta,\mathbb{E}_{\theta'}[T(\mathcal{H})]/\xi)) - \frac{s^2 Y}{4}\delta^{1+\frac{1}{\rho}} \qquad (\text{set } \tau = \mathbb{E}_{\theta'}[T(\mathcal{H})]/\xi)$$
$$\geq \frac{Y}{2}\exp(-KL(\rho,\delta,\mathbb{E}_{\theta'}[T(\mathcal{H})]/\xi)) - \frac{s^2 Y}{4}\delta^{1+\frac{1}{\rho}} \qquad (\text{setting } \xi \leq \tfrac{1}{4})$$

Setting $\delta = (\frac{4\xi}{s^2 Y})^{\frac{\rho}{\rho+1}}$, $\rho = 3$ and rearranging the last equation with sufficiently large s we get:

$$\mathbb{E}_{\theta'}[T(\mathcal{H})] \geq \frac{\xi}{77 s \kappa^2}\left(\frac{1}{4\epsilon^2}\ln\frac{Y}{4\xi} - \frac{4ns^3}{d}\right) \qquad (\text{Claim 3})$$

We are talking about the data-poor regime where $d > \kappa^{-4/3}s^{4/3}n^{1/3}$. So setting $\epsilon = \kappa^{-2/3}s^{-1/3}n^{1/3}$ and using the following Lemma 9 leads the conclusion that the order of $\mathbb{E}_{\theta'}[T(\mathcal{H})] \geq \Omega(\frac{1}{\kappa^2\epsilon^2})$.

Lemma 9. $\ln Y = \Omega(s)$, and for sufficiently large s, $\ln\frac{s^2 Y}{4\xi} \leq 2\ln\frac{Y}{4\xi}$

F.2.3 Proof of the Claim 1

This claim consists of two parts - prove the relationship between empirical KL and the true KL divergence, and the KL divergence calculation part. We will prove the first inequality as follows.

Lemma 10. For $\rho > 0$, let

$$B_{\theta',\theta}(\rho) := \left(\exists T \geq 1 \text{ s.t. } \sum_{t=1}^T \ln(\frac{p_{\theta'}(r_t|a_t)}{p_\theta(r_t|a_t)}) \geq (1+\rho)\sum_{t=1}^T \mathsf{KL}(p_{\theta'}(\cdot|a_t), p_\theta(\cdot|a_t)) + \frac{1}{\rho}\ln(\delta^{-1})\right)$$

Then, $\mathbb{P}_{\theta'}(B_{\theta,\theta'}(\rho)) \leq \delta$.

Proof. Let $J_t := \ln(\frac{p_{\theta'}(r_t|a_t)}{p_\theta(r_t|a_t)})$. Now let $H_t = \exp(\rho(\sum_{s=1}^t J_s - (1+\rho)\mathbb{E}_{\theta',s-1}[J_s|a_s])$ with $H_0 = 1$. Now we will prove that for all $\rho > 0$ the $\{H_t\}_{t=0}^T$ is a non-negative super-martingale.

$$\mathbb{E}_{\theta',t-1}[H_t] = \mathbb{E}_{\theta',t-1}[\exp(\rho(\sum_{s=1}^t J_s - (1+\rho)\mathbb{E}_{\theta',s-1}[J_s|a_s]))]$$
$$= H_{t-1}\,\mathbb{E}_{\theta',t-1}[\exp(\rho J_t - (\rho+\rho^2)\mathbb{E}_{\theta',t-1}[J_t|a_t])]$$

Now the proof boils down to check the case when $\mathbb{E}_{\theta',t-1}[\exp(\rho(J_t - (1+\rho)\mathbb{E}_{t-1}[J_t|a_t])] \leq 1$. Fortunately, we can explicitly calculate J_t. Let $\mu_t = \langle\theta, a_t\rangle$ and $\mu'_t = \langle\theta', a_t\rangle$. Then

$$J_t = \frac{-2r(\mu_t - \mu'_t) - (\mu'_t)^2 + \mu_t^2}{2}$$
$$\mathbb{E}_{\theta',t-1}[J_t|a_t] = \mathsf{KL}(p_{\theta'}(\cdot|a_t), p_\theta(\cdot|a_t))$$
$$= \mathsf{KL}(\mathcal{N}(\mu'_t, 1), \mathcal{N}(\mu_t, 1)) = \frac{(\mu'_t - \mu_t)^2}{2}$$

where $\mathsf{KL}(p_{\theta'}(\cdot|a_t), p_\theta(\cdot|a_t)) = \frac{(\mu_t - \mu'_t)^2}{2}$. Now by the 1-subgaussianity of $r_t - \mu'_t$,

$$\mathbb{E}_{\theta',t-1}\Big[\exp(\rho(J_t - (1+\rho)\mathbb{E}_{\theta',t-1}[J_t|a_t])\Big] = \mathbb{E}_{\theta',t-1}\Big[\mathbb{E}_{\theta',t-1}[\exp(\rho(J_t - (1+\rho)\mathbb{E}_{\theta',t-1}[J_t|a_t])|a_t]\Big]$$

$$= \mathbb{E}_{\theta',t-1}\left[\, \mathbb{E}_{\theta',t-1}[\exp(\rho \frac{-2r_t(\mu_t - \mu_t') - (\mu_t')^2 + \mu_t^2}{2})|a_t] \exp(-\rho(1+\rho)\frac{(\mu_t - \mu_t')^2}{2})\right]$$
$$\leq \mathbb{E}_{\theta',t-1}[\exp(\frac{\rho^2(\mu_t - \mu_t')^2}{2})\exp(\rho\frac{-2\mu_t'(\mu_t - \mu_t') - (\mu_t')^2 + \mu_t^2}{2})\exp(-\rho(1+\rho)\frac{(\mu_t - \mu_t')^2}{2})]$$
$$= 1$$

The last inequality holds because of the 1-subgaussian property of the η_t. Therefore, $\mathbb{E}_{\theta',t-1}[H_t] \leq H_{t-1}$, and therefore $\{H_t\}_{t=0}^T$ is a supermartingale. Finally, using Ville's maximal inequality on H_t we can achieve the desired probability inequality. □

By the above Lemma 10, one can induce the following relationship.

$$(15) = \frac{|Q_{\frac{3}{4}s-l}|}{|R_l|}\mathbb{E}_{\theta'}[\mathbb{1}(\hat{S} \in R_l, T(\mathcal{H}) \leq \tau)\exp(-\sum_{t=1}^n \ln\frac{p_{\theta'}(r_t|a_t)}{p_\theta(r_t|a_t)})]$$
$$= \frac{|Q_{\frac{3}{4}s-l}|}{|R_l|}\mathbb{E}_{\theta'}\left[\mathbb{1}\left(\hat{S} \in R_l, T(\mathcal{H}) \leq \tau, \neg B_{\theta,\theta'}(\rho)\right)\exp(-\frac{1}{2}(1+\rho)\sum_{t=1}^n \langle A_t, \tilde{\theta}\rangle^2 - \frac{1}{\rho}\ln\frac{1}{\delta})\right] \quad \text{(Lemma 10)}$$
$$= \frac{|Q_{\frac{3}{4}s-l}|}{|R_l|}\underbrace{\mathbb{E}_{\theta'}\left[\mathbb{1}\left(\hat{S} \in R_l, T(\mathcal{H}) \leq \tau, \neg B_{\theta,\theta'}(\rho)\right)\exp\left(-\frac{1}{2}(1+\rho)\sum_{t=1}^n \langle A_t, \tilde{\theta}\rangle^2\right)\right]}_{\text{(X)}}\exp(-\frac{1}{\rho}\ln\frac{1}{\delta})$$

Now the main part of the proof is bounding (X).

$$\begin{aligned}(X) =& \mathbb{E}_{\theta'}\left[\mathbb{1}\left(\hat{S} \in R_l, T(\mathcal{H}) \leq \tau, \neg B_{\theta,\theta'}(\rho)\right)\exp\left(-\frac{1}{2}(1+\rho)\sum_{t=1}^n \langle A_t, \tilde{\theta}\rangle^2\right)\right] \\ \geq& \mathbb{E}_{\theta'}\left[\mathbb{1}\left(\hat{S} \in R_l, T(\mathcal{H}) \leq \tau\right)\exp\left(-\frac{1}{2}(1+\rho)\sum_{t=1}^n \langle A_t, \tilde{\theta}\rangle^2\right)\right] - \delta\end{aligned}$$

Define $\Pi = \{\pi_1 \circ \pi_2 : \pi_1 \in \mathrm{Sym}([1:s]), \pi_2 \in \mathrm{Sym}([s+1:d])\}$. Importantly, for any $\pi \in \Pi$, $\pi(\theta') = \theta'$.

We focus on the first term in the above expression:

$$\mathbb{E}_{\theta'}\left[\mathbb{1}\left(\hat{S} \in R_l, T(\mathcal{H}) \leq \tau\right)\exp\left(-\frac{1}{2}(1+\rho)\sum_{t=1}^n \langle A_t, \tilde{\theta}\rangle^2\right)\right]$$
$$= \frac{1}{|\Pi|}\sum_{\sigma\in\Pi}\mathbb{E}_{\sigma(\theta')}\left[\mathbb{1}\left(\sigma^{-1}(\hat{S}) \in R_l\right)\cdot \mathbb{1}\left(\sum_{t=1}^n I(\sigma^{-1}(A_t) \in \mathcal{H}) \leq \tau\right)\exp\left(-\frac{1}{2}(1+\rho)\sum_{t=1}^n \langle \sigma^{-1}(A_t), \tilde{\theta}\rangle^2\right)\right] \quad \text{(Lemma 2)}$$
$$= \mathbb{E}_{\theta'}\left[\left(\frac{1}{|\Pi|}\sum_{\sigma\in\Pi}\mathbb{1}\left(\sigma^{-1}(\hat{S}) \in R_l\right)\exp\left(-\frac{1}{2}(1+\rho)\sum_{t=1}^n \langle \sigma^{-1}(A_t), \tilde{\theta}\rangle^2\right)\right)\mathbb{1}\left(\sum_{t=1}^n I(A_t \in \mathcal{H}) \leq \tau\right)\right] \quad (16)$$

Now for any realization of $\boldsymbol{A}, \hat{S}$, namely, $\boldsymbol{a} = (a_1, \ldots, a_n) \in \mathcal{A}^n, u \subset [d]$, we examine the quantity

$$\frac{1}{|\Pi|}\sum_{\sigma\in\Pi}\mathbb{1}\left(\sigma^{-1}(u) \in R_l\right)\exp\left(-\frac{1}{2}(1+\rho)\sum_{t=1}^n \langle \sigma^{-1}(a_t), \tilde{\theta}\rangle^2\right)$$

Claim 2. For any set u such that $|u| = \frac{d}{2}$, and any $a_1, \ldots, a_n$,

$$\frac{1}{|\Pi|}\sum_{\sigma\in\Pi}\mathbb{1}\left(\sigma^{-1}(u) \in R_l\right)\exp\left(-\frac{1}{2}(1+\rho)\sum_{t=1}^n \langle \sigma^{-1}(a_t), \tilde{\theta}\rangle^2\right)$$

$$\geq \mathbb{1}\left(u \in Q'_{\frac{1}{4}s-l}\right) \cdot \frac{|R_l|}{|Q'_{\frac{1}{4}s-l}|} \cdot \exp\left(-\frac{1}{2}(1+\rho)(\frac{4s^3 n}{d} + 77s\kappa^2 \mathcal{T}(\mathcal{H};a))\epsilon^2\right)$$

Proof. If $u \notin Q'_{\frac{1}{4}s-l}$, then for any permutation $\sigma \in \Pi$, it must be the case that $\sigma^{-1}(u) \notin Q'_{\frac{1}{4}s-l}$, and therefore, $\sigma^{-1}(u) \notin R_l$. In this case, both sides are equal to zero and the claim is trivially true.

Otherwise, $u \in Q'_{\frac{1}{4}s-l}$. Define $\Pi_{legal,u} = \left\{\sigma \in \Pi : \sigma^{-1}(u) \in R_l\right\}$. Using this notation, the left hand side can be equivalently written as:

$$\begin{aligned}
&\frac{1}{|\Pi|} \sum_{\sigma \in \Pi_{legal,u}} \exp\left(-\frac{1}{2}(1+\rho)\sum_{t=1}^{n} \langle \sigma^{-1}(a_t), \tilde{\theta}\rangle^2\right) \\
\geq &\frac{|\Pi_{legal,u}|}{|\Pi|} \cdot \frac{1}{|\Pi_{legal,u}|} \sum_{\sigma \in \Pi_{legal,u}} \exp\left(-\frac{1}{2}(1+\rho)\sum_{t=1}^{n} \langle \sigma^{-1}(a_t), \tilde{\theta}\rangle^2\right) \\
\geq &\frac{|\Pi_{legal,u}|}{|\Pi|} \cdot \exp\left(-\frac{1}{2}(1+\rho)\left(\frac{1}{|\Pi_{legal,u}|} \sum_{\sigma \in \Pi_{legal,u}} \sum_{t=1}^{n} \langle \sigma^{-1}(a_t), \tilde{\theta}\rangle^2\right)\right) && \text{(Jensen)} \\
\geq &\frac{|\Pi_{legal,u}|}{|\Pi|} \cdot \exp\left(-\frac{1}{2}(1+\rho)(\frac{4s^3 n}{d} + 77s\kappa^2 \mathcal{T}(\mathcal{H};a))\epsilon^2\right) && \text{(Lemma 12)} \\
= &\frac{|R_l|}{|Q'_{\frac{1}{4}s-l}|} \cdot \exp\left(-\frac{1}{2}(1+\rho)(\frac{4s^3 n}{d} + 77s\kappa^2 \mathcal{T}(\mathcal{H};a))\epsilon^2\right) && \text{(Lemma 11)}
\end{aligned}$$

□

We defer Lemma 12 and Lemma 11 to Section F.2.4. Proof of Claim 1 basically ends after combining Lemma 10 and the above claim. However, for the clear description how this was used in the main proof, we now continue Equation (16), which, using the above claim, is at least

$$\begin{aligned}
&\mathbb{E}_{\theta'}\left[\left(\frac{1}{|\Pi|}\sum_{\sigma \in \Pi} \mathbb{1}\left(\sigma^{-1}(\hat{S}) \in R_l\right) \exp\left(-\frac{1}{2}(1+\rho)\sum_{t=1}^{n} \langle \sigma^{-1}(A_t), \tilde{\theta}\rangle^2\right)\right) \mathbb{1}\left(\sum_{t=1}^{n} I(A_t \in \mathcal{H}) \leq \tau\right)\right] \\
\geq &\mathbb{E}_{\theta'}\left[\mathbb{1}\left(\hat{S} \in Q'_{\frac{1}{4}s-l}\right) \mathbb{1}\left(\sum_{t=1}^{n} I(A_t \in \mathcal{H}) \leq \tau\right)\right] \cdot \frac{|R_l|}{|Q'_{\frac{1}{4}s-l}|} \cdot \exp\left(-\frac{1}{2}(1+\rho)(\frac{4s^3 n}{d} + 77s\kappa^2 \tau)\epsilon^2\right) \\
= &\mathbb{E}_{\theta'}\left[\mathbb{1}\left(\hat{S} \in Q'_{\frac{1}{4}s-l}, T(\mathcal{H}) \leq \tau\right)\right] \cdot \frac{|R_l|}{|Q'_{\frac{1}{4}s-l}|} \cdot \exp\left(-\frac{1}{2}(1+\rho)(\frac{4s^3 n}{d} + 77s\kappa^2 \tau)\epsilon^2\right) \\
= &\mathbb{E}_{\theta'}\left[\mathbb{1}\left(\hat{S} \in R_l, T(\mathcal{H}) \leq \tau\right)\right] \cdot \exp\left(-\frac{1}{2}(1+\rho)(\frac{4s^3 n}{d} + 77s\kappa^2 \tau)\epsilon^2\right)
\end{aligned}$$

Therefore, in conclusion,

$$(15) \geq \frac{|Q_{\frac{3}{4}s-l}|}{|R_l|}\left(\mathbb{E}_{\theta'}\left[\mathbb{1}\left(\hat{S} \in R_l, T(\mathcal{H}) \leq \tau\right)\right] \cdot \exp\left(-\frac{1}{2}(1+\rho)(\frac{4s^3 n}{d} + 77s\kappa^2 \tau)\epsilon^2 - \frac{1}{\rho}\ln\frac{1}{\delta}\right) - \delta^{1+\frac{1}{\rho}}\right)$$

and now one can keep proceed from (15).

F.2.4 Deferred proof for Claim 1

Lemma 11. For any $u \in Q'_{\frac{1}{4}s-l}$,

$$\frac{|\Pi_{legal,u}|}{|\Pi|} = \frac{|R_l|}{|Q'_{\frac{1}{4}s-l}|}$$

Proof. It suffices to prove $|\Pi_{legal,u}||Q'_{\frac{1}{4}s-l}| = |R_l||\Pi|$. To see this, first note that for any $\hat{u} \in Q_{\frac{1}{4}s-l}$, $|\Pi_{legal,\hat{u}}| = |\Pi_{legal,u}|$. Next, note,

$$\begin{aligned}
(\text{LHS}) &= \sum_{\hat{u} \in Q'_{\frac{1}{4}s-l}} |\Pi_{legal,\hat{u}}| \\
&= \sum_{\hat{u} \in Q'_{\frac{1}{4}s-l}} \sum_{\sigma \in \Pi} I(\sigma(\hat{u}) \in R_l) \\
&= \sum_{\sigma \in \Pi} \sum_{\hat{u} \in Q'_{\frac{1}{4}s-l}} I(\sigma(\hat{u}) \in R_l) \\
&= \sum_{\sigma \in \Pi} |R_l| \\
&= (\text{RHS}).
\end{aligned}$$

□

Lemma 12. Assume $s \le \sqrt{d}$ and $d \ge 16$. For any $\boldsymbol{a} \in \mathcal{A}^n$ and $u \in \mathcal{U}(\boldsymbol{a})$ with $u \in Q'_{\frac{1}{4}s-l}$,

$$\frac{1}{|\Pi_{legal}(l,u)|} \sum_t \sum_{\sigma \in \Pi_{legal}(l,u)} \langle \sigma^{-1}(a_t), \tilde{\theta} \rangle^2 \le \left(\frac{4s^3 n}{d} + 77 s \kappa^2 \mathcal{T}(\mathcal{H}; a) \right) \epsilon^2$$

Proof. Let us first focus on $t : a_t \in \mathcal{I}$ where $\mathcal{I}$ is the reward arm set.

$$\begin{aligned}
&\frac{1}{|\Pi_{legal}(l,u)|} \sum_{t: a_t \in \mathcal{I}} \sum_{\sigma \in \Pi_{legal}(l,u)} \langle \sigma^{-1}(a_t), \tilde{\theta} \rangle^2 \\
&\le \frac{1}{|\Pi_{legal}(l,u)|} \sum_{t: a_t \in \mathcal{I}} s\epsilon^2 \sum_{j=s+1}^{2s} \sum_{\sigma \in \Pi_{legal}(l,u)} |a_{t\sigma^{-1}(j)}| \\
&\le \frac{1}{|\Pi_{legal}(l,u)|} \sum_{t: a_t \in \mathcal{I}} s\epsilon^2 \sum_{j=s+1}^{2s} \sum_{\sigma \in \Pi_{legal}(l,u)} \sum_{h=s+1}^{d} |a_{th}| \mathbb{1}\left(\sigma^{-1}(j) = h \right) \\
&\le s\epsilon^2 \sum_{j=s+1}^{2s} \sum_{t: a_t \in \mathcal{I}} \sum_{h=s+1}^{d} |a_{th}| \cdot \frac{1}{|\Pi_{legal}(l,u)|} \sum_{\sigma \in \Pi_{legal}(l,u)} \mathbb{1}\left(\sigma^{-1}(j) = h \right)
\end{aligned}$$

Let us compute $\sum_{\sigma \in \Pi_{legal}(l,u)} \mathbb{1}\left(\sigma^{-1}(j) = h \right)$.

WLOG, let us consider the following specific u: $u = \{1, \ldots, \frac{3}{4}s + l, \frac{3}{2}s + 1, \ldots, \frac{3}{2}s + k\}$ where $k = \frac{d}{2} - (\frac{3}{4}s + l)$. Define $\text{Seg}_1 = [1 : s]$, $\text{Seg}_2 = [s+1 : 2s]$, and $\text{Seg}_3 = [2s+1 : d]$. Note that u has exactly $s - (s/4 - l) = \frac{3}{4}s + l$ members in Seg_1, so a legal permutation σ^{-1} should choose $s/2$ coordinates from the remaining members of u (there are k of them) and choose $s/2$ coordinates from $\{s+1, \ldots, \frac{3}{2}s\} \cup \{\frac{3}{2}s + k + 1, \ldots, d\}$ then put those two into Seg_2.

Therefore, we need to consider the following two cases:

- $h \in \text{Seg}_2$: Permutations that take one correct coordinate and puts it into a correct coordinate (correct = Seg_2). Thus, we need to multiply the following two:
  - the segment 1: $s!$
  - the segment 2: $\binom{k-1}{\frac{s}{2}-1}\binom{d-s-k}{\frac{s}{2}}(s-1)!(d-2s)!$
- $h \in \text{Seg}_3$: take one correct coordinate and puts it into an incorrect coordinate.
  - the segment 1: $s!$
  - the segment 2: $\binom{k-1}{\frac{s}{2}}\binom{d-s-k}{\frac{s}{2}-1}(s-1)!(d-2s)!$

Note that $|\Pi_{legal}(l,u)| = s! \cdot \binom{k}{\frac{s}{2}}\binom{d-s-k}{\frac{s}{2}} s!(d-2s)!$.

For the first case,

$$\frac{s!\binom{k-1}{\frac{s}{2}-1}\binom{d-s-k}{\frac{s}{2}}(s-1)!(d-2s)!}{|\Pi_{legal}(l,u)|} = \frac{1}{2k}$$

For the second case, we get $\frac{1}{2(d-s-k-\frac{s}{2}+1)}$. Plugging in the definition of k, we get that

$$
\begin{aligned}
& s\epsilon^2 \sum_{j=s+1}^{2s} \sum_{t:a_t \in \mathcal{I}} \sum_{h=s+1}^{d} |a_{th}| \cdot \frac{1}{|\Pi_{legal}(l,u)|} \sum_{\sigma \in \Pi_{legal}(l,u)} \mathbb{1}\left(\sigma^{-1}(j) = h\right) \\
& \le s\epsilon^2 \sum_{j=s+1}^{2s} \sum_{t:a_t \in \mathcal{I}} \sum_{h=s+1}^{d} |a_{th}| \cdot \frac{1}{d-2s} \\
& \le \frac{2s^3\epsilon^2 n}{d-2s} \le \frac{4s^3\epsilon^2 n}{d} && (s \le \tfrac{d}{4})
\end{aligned}
$$

For the second part,

$$
\begin{aligned}
& \frac{1}{|\Pi_{legal}(l,u)|} \sum_{t:a_t \in \mathcal{H}} \sum_{\sigma \in \Pi_{legal}(l,u)} \langle \sigma^{-1}(a_t), \tilde{\theta} \rangle^2 \\
& = \frac{\kappa^2 \epsilon^2}{|\Pi_{legal}(l,u)|} \sum_{t:a_t \in \mathcal{H}} \sum_{\sigma \in \Pi_{legal}(l,u)} \left(\sum_{j=s+1}^{2s} |a_{t\sigma^{-1}(j)}|^2 + \sum_{a,b \in \mathsf{Seg}_2 : a \neq b} a_{t\sigma^{-1}(a)} a_{t\sigma^{-1}(b)} \right) \\
& = \epsilon^2 \cdot \left(s\kappa^2 \mathcal{T}(\mathcal{H}; a) + \sum_{t:a_t \in \mathcal{H}} \sum_{a,b \in \mathsf{Seg}_2 : a \neq b} \underbrace{\frac{1}{|\Pi_{legal}(l,u)|} \sum_{\sigma \in \Pi_{legal}(l,u)} a_{t\sigma^{-1}(a)} a_{t\sigma^{-1}(b)}}_{=: (Z_2')} \right)
\end{aligned}
$$

To compute (Z_2'), let us define the following where the first two are false positives w.r.t. $\mathrm{supp}(\theta')$ and the last two are true negatives w.r.t. $\mathrm{supp}(\theta')$:

$$
\begin{aligned}
C_+ &= \{i \in S' \backslash [s] | A'_{ti} = \kappa\} \\
C_- &= \{i \in S' \backslash [s] | A'_{ti} = -\kappa\} \\
M_+ &= \{i \in [s+1:d] \backslash S' | A'_{ti} = \kappa\} \\
M_- &= \{i \in [s+1:d] \backslash S' | A'_{ti} = -\kappa\} .
\end{aligned}
$$

Let c_+, c_-, m_+, m_- be $|C_+|, |C_-|, |M_+|, |M_-|$, respectively. Note that $c_+ + c_- = \frac{d}{2} - \frac{3}{4}s - l = k$ and $m_+ + m_- = \frac{d}{2} - \frac{1}{4}s + l = d - s - k$. Then,

- When $A_{t\pi^{-1}(a)} A_{t\pi^{-1}(b)} = \kappa^2$:
  - Common part $\mathsf{COM} = (s-2)!(d-2s)!s!$
  - $A_{t\pi^{-1}(a)} \in C_+, A_{t\pi^{-1}(b)} \in C_+$: $c_+ \times (c_+ - 1) \times \binom{k-2}{\frac{s}{2}-2}\binom{d-s-k}{\frac{s}{2}}\mathsf{COM}$
  - $A_{t\pi^{-1}(a)} \in M_+, A_{t\pi^{-1}(b)} \in M_+$: $m_+ \times (m_+ - 1) \times \binom{k}{\frac{s}{2}}\binom{d-s-k-2}{\frac{s}{2}-2}\mathsf{COM}$
  - $A_{t\pi^{-1}(a)} \in C_-, A_{t\pi^{-1}(b)} \in C_-$: $c_- \times (c_- - 1) \times \binom{k-2}{\frac{s}{2}-2}\binom{d-s-k}{\frac{s}{2}}\mathsf{COM}$
  - $A_{t\pi^{-1}(a)} \in M_-, A_{t\pi^{-1}(b)} \in M_-$:$m_- \times (m_- - 1) \times \binom{k}{\frac{s}{2}}\binom{d-s-k-2}{\frac{s}{2}-2}\mathsf{COM}$
  - $A_{t\pi^{-1}(a)} \in C_+, A_{t\pi^{-1}(b)} \in M_+$ and opposite: $2 \times c_+ \times m_+ \times \binom{k-1}{\frac{s}{2}-1}\binom{d-s-k-1}{\frac{s}{2}-1}\mathsf{COM}$
  - $A_{t\pi^{-1}(a)} \in C_-, A_{t\pi^{-1}(b)} \in M_-$ and opposite: $2 \times c_- \times m_- \times \binom{k-1}{\frac{s}{2}-1}\binom{d-s-k-1}{\frac{s}{2}-1}\mathsf{COM}$
- When $A_{t\pi^{-1}(a)} A_{t\pi^{-1}(b)} = -\kappa^2$:
  - $A_{t\pi^{-1}(a)} \in C_+, A_{t\pi^{-1}(b)} \in C_-$ and opposite: $2 \times c_+ \times c_- \times \binom{k-2}{\frac{s}{2}-2}\binom{d-s-k}{\frac{s}{2}}\mathsf{COM}$
  - $A_{t\pi^{-1}(a)} \in M_+, A_{t\pi^{-1}(b)} \in M_-$ and opposite: $2 \times m_+ \times m_- \times \binom{k}{\frac{s}{2}}\binom{d-s-k-2}{\frac{s}{2}-2}\mathsf{COM}$
  - $A_{t\pi^{-1}(a)} \in C_+, A_{t\pi^{-1}(b)} \in M_-$ and opposite: $2 \times c_+ \times m_- \times \binom{k-1}{\frac{s}{2}-1}\binom{d-s-k-1}{\frac{s}{2}-1}\mathsf{COM}$
  - $A_{t\pi^{-1}(a)} \in M_+, A_{t\pi^{-1}(b)} \in C_-$ and opposite: $2 \times m_+ \times c_- \times \binom{k-1}{\frac{s}{2}-1}\binom{d-s-k-1}{\frac{s}{2}-1}\mathsf{COM}$

To evaluate (Z_2'), note that

$$\frac{\mathsf{COM}}{|\Pi_{legal}(l,u)|} = \frac{1}{\binom{k}{s/2}\cdot\binom{d-s-k}{s/2}s(s-1)}$$

Assuming $d \geq s^2$ and $s \geq 4$, we have that $k \wedge (d-s-k) \geq \frac{d}{2\sqrt{2}}$. This also implies that $s \geq 2$, $k \geq 2$, and $d-s-k \geq 2$. Note that

$$\binom{k-2}{\frac{s}{2}-2}\binom{d-s-k}{\frac{s}{2}}\cdot\frac{\mathsf{COM}}{|\Pi_{legal}(l,u)|} = \frac{\frac{s}{2}(\frac{s}{2}-1)}{s(s-1)k(k-1)}$$

$$\binom{k}{\frac{s}{2}}\binom{d-s-k-2}{\frac{s}{2}-2}\cdot\frac{\mathsf{COM}}{|\Pi_{legal}(l,u)|} = \frac{\frac{s}{2}(\frac{s}{2}-1)}{s(s-1)(d-s-k)(d-s-k-1)}$$

$$\binom{k-1}{\frac{s}{2}-1}\binom{d-s-k-1}{\frac{s}{2}-1}\cdot\frac{\mathsf{COM}}{|\Pi_{legal}(l,u)|} = \frac{\frac{s}{2}\cdot\frac{s}{2}}{s(s-1)\cdot k(d-s-k)}$$

Then,

$$\begin{aligned}
(Z_2') &\leq \kappa^2\frac{\frac{s}{2}\left(\frac{s}{2}-1\right)}{s(s-1)}\left(\frac{c_+(c_+-1)+c_-(c_--1)-2c_+c_-}{k(k-1)} + \frac{m_+(m_+-1)+m_-(m_--1)-2m_+m_-}{(d-s-k)(d-s-k-1)}\right) \\
&\quad + \kappa^2\frac{\frac{s}{2}\cdot\frac{s}{2}}{s(s-1)}\left(2\cdot\frac{(c_+-c_-)(m_+-m_-)}{k(d-s-k)}\right) \\
&\leq \kappa^2\frac{\frac{s}{2}\left(\frac{s}{2}-1\right)}{s(s-1)}\left(\frac{(c_+-c_-)^2}{k(k-1)} + \frac{(m_+-m_-)^2}{(d-s-k)(d-s-k-1)}\right) \\
&\quad + \kappa^2\frac{\frac{s}{2}\cdot\frac{s}{2}}{s(s-1)}\left(2\cdot\frac{(c_+-c_-)(m_+-m_-)}{k(d-s-k)}\right) && (c_+ + c_- \geq 0, m_+ - m_- \geq 0) \\
&\leq \kappa^2\frac{\frac{s}{2}\cdot\frac{s}{2}}{s(s-1)}\left(\frac{(c_+-c_-)^2}{k(k-1)} + \frac{(m_+-m_-)^2}{(d-s-k)(d-s-k-1)} + 2\cdot\frac{(c_+-c_-)(m_+-m_-)}{k(d-s-k)}\right) \\
&= \underbrace{\kappa^2\frac{\frac{s}{2}\cdot\frac{s}{2}}{s(s-1)}\cdot\left(\frac{c_+-c_-}{k} + \frac{m_+-m_-}{d-s-k}\right)^2}_{=:\,(Z_{2,1}')} + \underbrace{\kappa^2\frac{\frac{s}{2}\cdot\frac{s}{2}}{s(s-1)}\left(\frac{(c_+-c_-)^2}{k^2(k-1)} + \frac{(m_+-m_-)^2}{(d-s-k)^2(d-s-k-1)}\right)}_{=:\,(Z_{2,2}')} \\
&&& \left(\tfrac{1}{x(x-1)} = \tfrac{1}{x^2} + \tfrac{1}{x^2(x-1)}\right)
\end{aligned}$$

For $(Z_{2,1}')$, let $w = \frac{(c_++m_+)-(c_-+m_-)}{2}$. Note $|2w| \leq \sqrt{2d\ln 2d} + s$. Then,

- $c_+ + c_- = k$, $m_+ + m_- = d-s-k$
- $c_+ + m_+ = \frac{d-s}{2} + w$, $c_- + m_- = \frac{d-s}{2} - w$.
- $c_+ = \frac{d-s}{2} + w - m_+$ and $c_- = \frac{d-s}{2} - w - m_-$

Thus, $c_+ - c_- = 2w - (m_+ - m_-)$. So,

$$\begin{aligned}
\frac{c_+-c_-}{k} + \frac{m_+-m_-}{d-s-k} &= \frac{2w-(m_+-m_-)}{k} + \frac{m_+-m_-}{d-s-k} \\
&= \frac{1}{k}\left(2w - (\frac{d-s-2k}{d-s-k})(m_+-m_-)\right) \\
&= \frac{1}{k}\left(2w - (\frac{\frac{7}{2}l-s}{d-s-k})(m_+-m_-)\right) \\
\implies \left|\frac{c_+-c_-}{k} + \frac{m_+-m_-}{d-s-k}\right| &\leq \frac{1}{k}\left(2w + \left|\frac{\frac{7}{2}l-s}{d-s-k}\right|(d-s-k)\right) \\
&\leq \frac{1}{k}\left(\sqrt{2d\ln(2d)}+s+s\right) \leq \frac{12\sqrt{\ln(2d)}}{\sqrt{d}} \\
&&& (k \geq \tfrac{d}{2\sqrt{2}}, \sqrt{2d\ln(2d)} > s)
\end{aligned}$$

Thus, using $s \geq 2$ and $d > \ln d$,

$$(Z'_{2,1}) = \kappa^2 \frac{\frac{s}{2} \cdot \frac{s}{2}}{s(s-1)} \cdot \frac{144 \ln(2d)}{d^2} \leq \kappa^2 \cdot \frac{72}{d}$$

For $(Z'_{2,2})$, using $k \wedge (d-s-k) \geq \frac{d}{4}$,

$$\begin{aligned}(Z'_{2,2}) &= \kappa^2 \cdot \frac{\frac{s}{2} \cdot \frac{s}{2}}{s(s-1)} \cdot \left(\frac{1}{k-1} + \frac{1}{d-s-k-1} \right) \\ &\leq \kappa^2 \cdot \frac{1}{2} \cdot \frac{8}{d} = \frac{4\kappa^2}{d}\end{aligned}$$

Thus, $(Z'_2) \leq \frac{76\kappa^2}{d}$.

Altogether,

$$\begin{aligned}&\frac{1}{|\Pi_{legal}(l,u)|} \sum_{t:a_t \in \mathcal{H}} \sum_{\sigma \in \Pi_{legal}(l,u)} \langle \sigma^{-1}(a_t), \tilde{\theta} \rangle^2 \\ &\leq \epsilon^2 \cdot \left(s\kappa^2 \mathcal{T}(\mathcal{H}; a) + \mathcal{T}(\mathcal{H}; a) \cdot s(s-1) \cdot 76 \frac{\kappa^2}{d} \right) \\ &< 77\epsilon^2 s\kappa^2 \mathcal{T}(\mathcal{H}; a)\end{aligned}$$

where the last inequality is by $(s-1) < d$. □

F.2.5 Proof of Lemma 9

Proof. We will use the following Stirling's approximation for calculating the scale of Y.

Lemma 13. (Stirling's approximation [19]) For $x \geq 1$,

$$\sqrt{2\pi x} \left(\frac{x}{e} \right)^x e^{\frac{1}{12n+1}} \leq x! \leq \sqrt{2\pi x} \left(\frac{x}{e} \right)^x e^{\frac{1}{12n}}$$

. To be more specific, there exists an absolute constant $s_0 = 10$ such that for all $x > s_0$,

$$\frac{x!}{\sqrt{2\pi x} \left(\frac{x}{e} \right)^x} \in (1, 1.01)$$

From this lemma, we can derive the following estimation of the combination. For $x, y > s_0$,

$$\frac{\binom{x+y}{x}}{\sqrt{\frac{1}{2\pi}(\frac{1}{x} + \frac{1}{y})} (\frac{x+y}{x})^x (\frac{x+y}{y})^y} \in (\frac{1}{1.01^2}, \frac{1.01}{1})$$

Now by definition, $|Q_a| = \binom{d-2s}{\frac{d}{2}-2s+a}\binom{2s}{a}$, and $|Q'_b| = \binom{d-s}{\frac{d}{2}-s+b}\binom{s}{b}$. For sufficiently large $s > 4s_0$ and $d > s^2$ we have the following approximation

$$\begin{aligned}\binom{d-2s}{\frac{d}{2} - \frac{3}{4}s + l} &\geq \frac{1}{1.01^2} \times \sqrt{\frac{1}{2\pi}(\frac{1}{\frac{d}{2} - \frac{3}{4}s + l} + \frac{1}{\frac{d}{2} - \frac{5}{4}s - l})} \left(\frac{d-2s}{\frac{d}{2} - \frac{3}{4}s + l} \right)^{\frac{d}{2} - \frac{3}{4}s + l} \left(\frac{d-2s}{\frac{d}{2} - \frac{5}{4}s - l} \right)^{\frac{d}{2} - \frac{5}{4}s - l} \\ &= \frac{1}{1.01^2} \times \sqrt{\frac{1}{2\pi}(\frac{1}{\frac{d}{2} - \frac{3}{4}s + l} + \frac{1}{\frac{d}{2} - \frac{5}{4}s - l})} 2^{d-2s} \left(1 - \frac{\frac{1}{4}s + l}{\frac{d}{2} - \frac{3}{4}s + l} \right)^{\frac{d}{2} - \frac{3}{4}s + l} \left(1 + \frac{\frac{1}{4}s + l}{\frac{d}{2} - \frac{5}{4}s - l} \right)^{\frac{d}{2} - \frac{5}{4}s - l}\end{aligned}$$

$$\begin{aligned}\binom{d-s}{\frac{d}{2} - \frac{1}{4}s + l} &\leq \frac{1.01}{1} \times \sqrt{\frac{1}{2\pi}(\frac{1}{\frac{d}{2} - \frac{1}{4}s + l} + \frac{1}{\frac{d}{2} - \frac{3}{4}s - l})} \left(\frac{d-s}{\frac{d}{2} - \frac{1}{4}s + l} \right)^{\frac{d}{2} - \frac{1}{4}s + l} \left(\frac{d-s}{\frac{d}{2} - \frac{3}{4}s - l} \right)^{\frac{d}{2} - \frac{3}{4}s - l} \\ &= \frac{1.01}{1} \times \sqrt{\frac{1}{2\pi}(\frac{1}{\frac{d}{2} - \frac{1}{4}s + l} + \frac{1}{\frac{d}{2} - \frac{3}{4}s - l})} 2^{d-s} \left(1 - \frac{\frac{1}{4}s + l}{\frac{d}{2} - \frac{1}{4}s + l} \right)^{\frac{d}{2} - \frac{1}{4}s + l} \left(1 + \frac{\frac{1}{4}s + l}{\frac{d}{2} - \frac{3}{4}s - l} \right)^{\frac{d}{2} - \frac{3}{4}s - l}\end{aligned}$$

From the Taylor's expansion, one can achieve the following inequality.

Lemma 14. For all $y > 2$ and $x > y^2$,

$$\frac{(1-\frac{y}{x})^x}{e^{-y}} \in (e^{-2}, 1)$$

Proof.

$$\begin{aligned}(1-\frac{y}{x})^x = \exp\left(x\ln(1-\frac{y}{x})\right) &\geq \exp\left(x(-\frac{y}{x})\right)\exp\left(-\frac{1}{(1-\frac{y}{x})^2}\times\frac{y^2}{2x}\right) && \text{(Taylor theorem with remainder)}\\ &\geq \exp(-y)e^{-2} && (e^{-a} \geq 1-a\text{, and } x^{1/2} > y > 2)\end{aligned}$$

Similarly,

$$(1-\frac{y}{x})^x = \exp\left(x\ln(1-\frac{y}{x})\right) \leq \exp\left(x\times(-\frac{y}{x})\right) \leq \exp(-y) \qquad (\ln a \leq a-1)$$

□

Therefore,

$$\binom{d-2s}{\frac{d}{2}-\frac{3}{4}s+l} \geq \frac{1}{1.01^2}e^{-2}\times\sqrt{\frac{1}{2\pi}(\frac{1}{\frac{d}{2}-\frac{3}{4}s+l}+\frac{1}{\frac{d}{2}-\frac{5}{4}s-l})}2^{d-2s}e^{-\frac{1}{4}s-l}e^{\frac{1}{4}s+l}$$

$$\binom{d-s}{\frac{d}{2}-\frac{1}{4}s+l} \leq \frac{1.01}{1}\times\sqrt{\frac{1}{2\pi}(\frac{1}{\frac{d}{2}-\frac{1}{4}s+l}+\frac{1}{\frac{d}{2}-\frac{3}{4}s-l})}2^{d-s}e^{-\frac{1}{4}s-l}e^{\frac{1}{4}s+l}$$

and thus $Y = \min_{l\in[s/4]}\frac{|Q_{\frac{3}{4}s-l}|}{|Q'_{\frac{s}{4}-l}|} \geq C2^{-s}\min_{l\in[s/4]}\frac{\binom{2s}{\frac{3}{4}s-l}}{\binom{s}{\frac{1}{4}s-l}}$ where $C = e^{-2}(\frac{1}{1.01})^3\frac{1}{4}$ is a universal constant.

From the formula $\binom{n}{k+1} = \binom{n}{k}\times\frac{n-k}{k+1}$, one can check that $\min_{l\in[\frac{s}{4}]}\frac{\binom{2s}{\frac{3}{4}s-l}}{\binom{s}{\frac{1}{4}s-l}} \geq \frac{\binom{2s}{\frac{3}{4}s}}{\binom{s}{\frac{1}{4}s}} \approx C'(2\times(\frac{8}{5})^5)^{\frac{1}{4}s}$ where the last one is from Lemma 13 again and $C' = \frac{1}{1.01^3}\times\sqrt{2}5$ is another universal constant. In short,

$$Y \geq C\times C'\times 2^{-s}(2\times(\frac{8}{5})^5)^{\frac{1}{4}s} \geq C\times C'\times(1.3)^{\frac{1}{4}s}$$

where the last inequality is from $2\times(\frac{8}{5})^5 = 20.97\cdots > 1.3\times 2^4$ Therefore, we can conclude that $\ln Y = \Omega(s)$.

□

F.2.6 Proof of Lemma 8

$$|R_l| = \binom{d-2s}{\frac{d}{2}-\frac{3}{4}s+l}\binom{s}{\frac{s}{2}}\binom{s}{\frac{s}{4}-l}$$

,

$$Q'_{\frac{1}{4}s-l} = \binom{d-s}{\frac{d}{2}-\frac{1}{4}s+l}\binom{s}{\frac{1}{4}s-l}$$

From the combinatorial formula, $\binom{d-s}{\frac{d}{2}-\frac{1}{4}s+l} = \sum_{a=0}^{s}\binom{d-2s}{\frac{d}{2}-\frac{1}{4}s+l-a}\binom{s}{a}$, and for sufficiently large d, $\max_{a\in[s]}\binom{d-2s}{\frac{d}{2}-\frac{1}{4}s+l-a}\binom{s}{a} = \binom{d-2s}{\frac{d}{2}-\frac{3}{4}s+l}\binom{s}{\frac{s}{2}}$. Plus, by the similar technique from the previous section, we can show that $\binom{d-2s}{\frac{d}{2}-\frac{3}{4}s+l}\binom{s}{\frac{s}{2}} > \binom{d-2s}{\frac{d}{2}-\frac{1}{4}s+l}\binom{s}{0} + \binom{d-2s}{\frac{d}{2}-\frac{5}{4}s+l}\binom{s}{s}$. Therefore, $|Q'_{\frac{1}{4}s-l}| \leq s|R_l|$

F.2.7 Scale of κ^2 with respect to $C_{\min}$

Last thing we have to deal with is connecting κ^2 to $\mathcal{C}_{\min}$ and H_*^2.

Lemma 15. $\mathcal{C}_{\min}(\mathcal{H}) \ge \frac{\kappa^2}{2}$

Proof. By the maximality of $\mathcal{C}_{\min}(\mathcal{H})$, $\mathcal{C}_{\min}(\mathcal{H}) \ge \lambda_{min}(Q(\mathsf{Unif}(\mathcal{H})))$. If we prove that $\lambda_{min}(Q(\mathsf{Unif}(\mathcal{H}))) \ge \frac{\kappa^2}{2}$, then the proof is done.

Let's define $\mathcal{H}_d$ as

$$\mathcal{H}_d = \left\{ a \in \{-\kappa, \kappa\}^d \;\middle|\; \left|\sum_{i=1}^d a_i\right| \le \kappa\sqrt{2d\ln(2d)} \right\}$$

It is the set of first d coordinate vectors of $\mathcal{H}$.

We can express $Q(\mathsf{Unif}(\mathcal{H}))$ in terms of $Q(\mathsf{Unif}(\mathcal{H}_d))$ as follows

$$Q(\mathsf{Unif}(\mathcal{H})) = \begin{bmatrix} Q(\mathsf{Unif}(\mathcal{H}_d)) & \vec{0} \\ \vec{0}^\top & 1 \end{bmatrix}$$

Therefore, the proof boils down to calculate $Q(\mathsf{Unif}(\mathcal{H}_d))$. We can connect this matrix to $Q(\mathsf{Unif}(\{-\kappa,\kappa\}^d))$ by the following method

$$Q(\mathsf{Unif}(\{-\kappa,\kappa\}^d)) = Q(\mathsf{Unif}(\mathcal{H}_d)) \times \mathbb{P}_{A\sim\mathsf{Unif}(\mathcal{H}_d)}(A \in \mathcal{H}_d) + Q(\mathsf{Unif}(\{-\kappa,\kappa\}^d \backslash \mathcal{H}_d)) \times \mathbb{P}_{A\sim\mathsf{Unif}(\{-\kappa,\kappa\}^d)}(A \notin \mathcal{H}_d)$$

Now, since Rademacher is 1 sub-Gaussian random variable, $\sum_{i=1}^n a_i$ is $\kappa\sqrt{d}$ sub-Gaussian random variable when $a \sim \mathsf{Unif}(\{-\kappa,\kappa\})$. Therefore,

$$\mathbb{P}_{A\sim\mathsf{Unif}(\{-\kappa,\kappa\}^d)}(A \notin \mathcal{H}_d) = \mathbb{P}_{A\sim\mathsf{Unif}(\{-\kappa,\kappa\}^d)}(|\sum_{i=1}^d A_i| \ge \kappa\sqrt{2d\ln(2d)}) \le \frac{1}{2d}$$

The last inequality is by the traditional Chernoff method [14]. Therefore, we can rewrite $Q(\mathsf{Unif}(\mathcal{H}_d))$ as

$$\begin{aligned} Q(\mathsf{Unif}(\mathcal{H}_d)) &\ge \mathbb{P}_{A\sim\mathsf{Unif}(\{-\kappa,\kappa\}^d)}(A \in \mathcal{H}_d) Q(\mathsf{Unif}(\mathcal{H}_d)) \\ &\ge Q(\mathsf{Unif}(\{-\kappa,\kappa\}^d)) - Q(\mathsf{Unif}(\{-\kappa,\kappa\}^d \backslash \mathcal{H}_d))\, \mathbb{P}_{A\sim\mathsf{Unif}(\{-\kappa,\kappa\}^d)}(A \notin \mathcal{H}_d) \\ &\ge Q(\mathsf{Unif}(\{-\kappa,\kappa\}^d)) - Q(\mathsf{Unif}(\{-\kappa,\kappa\}^d \backslash \mathcal{H}_d))\frac{1}{2d} \end{aligned}$$

Therefore, (Existing results about this positive definite matrix analysis)

$$\begin{aligned} \lambda_{\min}(Q(\mathsf{Unif}(\mathcal{H}_d))) &\ge \lambda_{\min}(Q(\mathsf{Unif}(\{-\kappa,\kappa\}^d)) - Q(\mathsf{Unif}(\{-\kappa,\kappa\}^d \backslash \mathcal{H}_d))\frac{1}{2d}) \\ &\ge \lambda_{\min}(Q(\mathsf{Unif}(\{-\kappa,\kappa\}^d))) - \lambda_{\max}(Q(\mathsf{Unif}(\{-\kappa,\kappa\}^d \backslash \mathcal{H}_d))\frac{1}{2d}) \end{aligned}$$

Since every element in $\{-\kappa,\kappa\}$ has ℓ_2-norm $\sqrt{d}\kappa$,

$$\begin{aligned} \lambda_{\max}(Q(\mathsf{Unif}(\{-\kappa,\kappa\}^d \backslash \mathcal{H}_d)) &= \max_{v\in\mathbb{S}^{d-1}} v^\top \sum_{a\in\{-\kappa,\kappa\}^d\backslash\mathcal{H}_d} \frac{1}{|\{-\kappa,\kappa\}^d\backslash\mathcal{H}_d|} aa^\top v \\ &= \max_{v\in\mathbb{S}^{d-1}} \sum_{a\in\{-\kappa,\kappa\}^d\backslash\mathcal{H}_d} \frac{1}{|\{-\kappa,\kappa\}^d\backslash\mathcal{H}_d|}(a^\top v)^2 \\ &\le \max_{v\in\mathbb{S}^{d-1}} \sum_{a\in\{-\kappa,\kappa\}^d\backslash\mathcal{H}_d} \frac{d\kappa^2}{|\{-\kappa,\kappa\}^d\backslash\mathcal{H}_d|} = d\kappa^2 \end{aligned}$$

and by simple symmetry one can calculate $\lambda_{\min}(Q(\mathsf{Unif}(\{-\kappa,\kappa\}^d))) = \kappa^2$. Therefore,

$$\lambda_{\min}(Q(\mathsf{Unif}(\mathcal{H}_d))) \ge \kappa^2 - d\kappa^2\frac{1}{2d} = \frac{\kappa^2}{2}$$

□

G Experiment details

- Case 1 - ℓ_1 estimation error experiment
  - $\theta = -e_1 + e_i$, $i \in \{2, \cdots, d\}$ chosen uniformly random before the start of the experiment.
  - Dimension $d = 10$, sparsity $s = 2$
  - Action set $\mathcal{A} = \{e_1 + \frac{1}{\sqrt{d}} e_i | i = 2, \cdots, d\} \cup \{\frac{1}{\sqrt{d}} e_1\}$
  - $T = 1000, 2000, \cdots, 10000$
  - $\sigma = 0.1$
  - Repetition: 30 times for each exploration time.
- Case 1 - bandit experiment
  - $\theta = e_1 + e_i$, $i \in \{2, \cdots, d\}$ chosen uniformly random before the start of the experiment.
  - Dimension $d = 10$, sparsity $s = 2$
  - Action set $\mathcal{A} = \{e_1 + \frac{1}{\sqrt{d}} e_i | i = 2, \cdots, d\} \cup \{\frac{1}{\sqrt{d}} e_1\}$
  - $T = 400000$
  - $\sigma = 0.1$
  - Repetition: 30 times
- Case 2 - ℓ_1 estimation error experiment
  - $\theta = e_i + e_j$, $i, j \in [d]$ chosen uniformly random before the start of the experiment.
  - Dimension $d = 30$, sparsity $s = 2$
  - Action set $\mathcal{A}$: 90 Uniform random vectors over $\mathbb{S}^{d-1}$ before the start of the round, where $\mathbb{S}^{d-1} = \{v \in \mathbb{R}^d | \|v\|_2 = 1\}$
  - $T = 1000, 2000, \cdots, 10000$
  - $\sigma = 0.1$
  - Repetition: 30 times for each exploration time.
- Case 2 - bandit experiment
  - $\theta = e_i + e_j$, $i, j \in [d]$ chosen uniformly random before the start of the experiment.
  - Dimension $d = 30$, sparsity $s = 2$
  - Action set $\mathcal{A}$: 90 Uniform random vectors over $\mathbb{S}^{d-1}$ before the start of the round.
  - $T = 10000$
  - $\sigma = 0.1$
  - Repetition: 30 times for each exploration time.

References

[1] Yasin Abbasi-Yadkori, David Pal, and Csaba Szepesvari. Online-to-Confidence-Set Conversions and Application to Sparse Stochastic Bandits. In *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2012.

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

[3] Romain Camilleri, Kevin Jamieson, and Julian Katz-Samuels. High-dimensional experimental design and kernel bandits. In *International Conference on Machine Learning*, pages 1227–1237. PMLR, 2021.

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

[5] Olivier Catoni. Challenging the empirical mean and empirical variance: a deviation study. In *Annales de l'IHP Probabilités et statistiques*, volume 48, pages 1148–1185, 2012.

[6] Hamid Eftekhari, Moulinath Banerjee, and Ya'acov Ritov. Design of c-optimal experiments for high dimensional linear models. *arXiv preprint arXiv:2010.12580*, 2020.

[7] Tanner Fiez, Lalit Jain, Kevin G Jamieson, and Lillian Ratliff. Sequential experimental design for transductive linear bandits. *Advances in neural information processing systems*, 32, 2019.

[8] Dean P Foster and Edward I George. The risk inflation criterion for multiple regression. *The Annals of Statistics*, 22(4):1947–1975, 1994.

[9] Anna Gilbert and Piotr Indyk. Sparse recovery using sparse matrices. *Proceedings of the IEEE*, 98(6):937–947, 2010.

[10] Botao Hao, Tor Lattimore, and Mengdi Wang. High-dimensional sparse linear bandits. *Advances in Neural Information Processing Systems*, 33:10753–10763, 2020.

[11] Yimin Huang, Xiangshun Kong, and Mingyao Ai. Optimal designs in sparse linear models. *Metrika*, 83(2):255–273, 2020.

[12] Adel Javanmard and Andrea Montanari. Confidence intervals and hypothesis testing for high-dimensional regression. *The Journal of Machine Learning Research*, 15(1):2869–2909, 2014.

[13] Gi-Soo Kim and Myunghee Cho Paik. Doubly-Robust Lasso Bandit. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 32, 2019.

[14] Tor Lattimore and Csaba Szepesvári. Bandit Algorithms. 2018. URL `http://downloads.tor-lattimore.com/book.pdf`.

[15] Tor Lattimore, Koby Crammer, and Csaba Szepesvári. Linear multi-resource allocation with semi-bandit feedback. *Advances in Neural Information Processing Systems*, 28, 2015.

[16] Blake Mason, Romain Camilleri, Subhojyoti Mukherjee, Kevin Jamieson, Robert Nowak, and Lalit Jain. Nearly optimal algorithms for level set estimation. *arXiv preprint arXiv:2111.01768*, 2021.

[17] Min-hwan Oh, Garud Iyengar, and Assaf Zeevi. Sparsity-agnostic lasso bandit. In *International Conference on Machine Learning*, pages 8271–8280. PMLR, 2021.

[18] Sathya Narayanan Ravi, Vamsi Ithapu, Sterling Johnson, and Vikas Singh. Experimental design on a budget for sparse linear models and applications. In *International Conference on Machine Learning*, pages 583–592. PMLR, 2016.

[19] Herbert Robbins. A remark on stirling's formula. *The American mathematical monthly*, 62(1): 26–29, 1955.

[20] Max Simchowitz, Kevin Jamieson, and Benjamin Recht. The simulator: Understanding adaptive sampling in the moderate-confidence regime. In *Conference on Learning Theory*, pages 1794–1834. PMLR, 2017.

[21] Vidyashankar Sivakumar, Steven Wu, and Arindam Banerjee. Structured linear contextual bandits: A sharp and geometric smoothed analysis. In *International Conference on Machine Learning*, pages 9026–9035. PMLR, 2020.

[22] Marta Soare, Alessandro Lazaric, and Rémi Munos. Best-arm identification in linear bandits. *Advances in Neural Information Processing Systems*, 27, 2014.

[23] Chao Tao, Saúl Blanco, and Yuan Zhou. Best arm identification in linear bandits with linear dimension dependency. In *International Conference on Machine Learning*, pages 4877–4886. PMLR, 2018.

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

[25] Sara A. van de Geer. On tight bounds for the lasso. *J. Mach. Learn. Res.*, 19:46:1–46:48, 2018.