Efficient Sparse Linear Bandits under High Dimensional Data

Xue Wang*
Damo Academy
Alibaba Group US
xue.w@alibaba-inc.com

Mike Mingcheng Wei*
School of Management
University at Buffalo
mcwei@buffalo.edu

Tao Yao*
School of Data Science
Shenzhen Research Inst. of Big Data
The Chinese University of HongKong, ShenZhen
yaotao@cuhk.edu.cn

ABSTRACT

We propose a computationally efficient Lasso Random Project Bandit (LRP-Bandit) algorithm for sparse linear bandit problems under high-dimensional settings with limited samples. LRP-Bandit bridges Lasso and Random Projection as feature selection and dimension reduction techniques to alleviate the computational complexity and improve the regret performance. We demonstrate that for the total feature dimension d, the significant feature dimension s, and the sample size T, the expected cumulative regret under LRP-Bandit is upper bounded by $\tilde{O}(T^{\frac{2}{3}} s^{\frac{3}{2}} \log^{\frac{7}{6}} d)$, where $\tilde{O}$ suppresses the logarithmic dependence on T. Further, we show that when available samples are larger than a problem-dependent threshold, the regret upper bound for LRP-Bandit can be further improved to $\tilde{O}(s\sqrt{T \log d})$. These regret upper bounds on T for both data-poor and data-rich regimes match the theoretical minimax lower bounds up to logarithmic factors. Through experiments, we show that LRP-Bandit is computationally efficient and outperforms other benchmarks on the expected cumulative regret.

CCS CONCEPTS

• **Theory of computation** → *Online learning algorithms.*

KEYWORDS

Online Learning, Multi-Armed Bandit, Dimensionality Reduction, Sparsity

ACM Reference Format:
Xue Wang, Mike Mingcheng Wei, and Tao Yao. 2023. Efficient Sparse Linear Bandits under High Dimensional Data. In ,. ACM, New York, NY, USA, 21 pages. https://doi.org/https://doi.org/10.1145/3580305.3599329

1 INTRODUCTION

The contextual bandit model has been extensively used to study the exploration-exploitation trade-off in a sequential learning and decision-making process [4, 7, 17, 20] and successfully applied to many practical problems [3, 35, 55]. In the big data era, contextual information for products and consumers is widely available and has been accumulated with extraordinary speed. Rich contextual information provides the decision-maker with unprecedented opportunities to learn and improve prediction accuracy.

Yet, in online settings, the decision maker's ability to effectively use all available high-dimensional contextual information to learn and select the reward-maximizing arm is often impaired by limited samples and high computational complexity. Recent works on the sparse contextual bandit framework have shown that the regret upper bound's dependence on the high-dimensional feature dimension d and the sample size dimension T can be reduced to sub-linear or poly-logarithmic orders by adopting sparse regularization (e.g., [5, 8, 16, 29, 31, 38, 39, 50, 53, 61]). However, these algorithms require strong assumptions on feature distributions or frequent parameter updates via non-smooth optimization [10], which are typically computationally expensive. [13, 15, 33, 63] adopt Random Projection (RP) or frequent directions methods to reduce the computational complexity, but special sampling procedures and the distortion and information loss intrinsic to these methods result in significant regret loss and lead to linear regret on T.

In this work, we propose a computationally efficient bandit algorithm based on the Upper Confidence Bound (UCB), i.e., the Lasso Random Projection Bandit (LRP-Bandit) algorithm. Specifically, with an epoch length that exponentially grows in time, we use Lasso [57] at the beginning of each epoch to construct (via thresholding Lasso estimators) a selected feature set, which includes features that potentially have strong influences on the decision-maker's reward; then, for each sample within an epoch, we adopt RP to compress high-dimensional features, excluding features in the selected feature set, to a low-dimensional space and then estimate coefficients for features in the selected feature set and randomly projected features. Through this process, parameter estimation can be performed in a low-dimensional space to significantly trim down the computational time while maintaining accuracy in predicting the decision-maker's reward.

Main Contributions: We demonstrate that feature selection via thresholding the Lasso estimator limits the negative influence of the information loss that is intrinsic and inevitable to RP and that RP can, in turn, alleviate the negative influence of model misspecification in Lasso due to limited samples. We propose a UCB-type algorithm based on Lasso and RP and theoretically prove that the expected cumulative regret of LRP-Bandit is upper bounded by $\tilde{O}(T^{\frac{2}{3}} s^{\frac{3}{2}} \log^{\frac{7}{6}} d)$, which matches the theoretical minimax lower-bound on T up to a logarithmic factor in the data-poor regime. Moreover, when available samples are larger than a problem-dependent threshold, the regret performance can be sharpened to $\tilde{O}(s\sqrt{T \log d})$,

*Authors contributed equally to this research. Correspondence to: Xue Wang <xue.w@alibaba-inc.com>

KDD 2023, 29th ACM SIGKDD CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING

ACM ISBN 979-8-4007-0103-0/23/08…$15.00
https://doi.org/https://doi.org/10.1145/3580305.3599329

Table 1: Regret comparisons for existing sparse linear bandit algorithms in data-rich regimes. Commonly required compatibility conditions and their variants are omitted to avoid duplication. τ is a problem-dependent parameter that has a complicated form and varies in different papers.

Bandit with feature selection	Regret (Data rich)	Additional requirement
Abbasi-Yadkori et al. [2]	$\tilde{O}(\sqrt{dT})$	
Sivakumar et al. [54]	$\tilde{O}(\sqrt{dT})$	Gaussian noise perturbed, adversary
Ren and Zhou [53]	$\tilde{O}(\log^{\frac{1}{2}} d \cdot T^{\frac{1}{2}+\tau})$	
Carpentier and Munos [13]	$\tilde{O}(\tau\sqrt{T})$	Uniformed exploring set
Lattimore et al. [34]	$\tilde{O}(\log d\sqrt{T})$	Solve combinatorial problems
Kim and Paik [31]	$\tilde{O}(\log d \cdot \sqrt{T})$	
Hao et al. [29]	Data poor: $\tilde{O}(\log^{\frac{1}{2}} d \cdot T^{\frac{2}{3}})$ Data rich: $\tilde{O}(\log d \cdot \sqrt{T})$	Pure exploration first, finite arms
Li et al. [38]	$\tilde{O}(\log^{\frac{1}{2}} d \cdot T^{\frac{2}{3}})$	Pure exploration first
Chen et al. [16]	$\tilde{O}(\text{poly-log}(d)\sqrt{T})$	Solve combinatorial problems
Bastani and Bayati [8]	$O(\tau \log^2 d \log^2 T)$	Margin condition
Wang et al. [61]	$O(\tau \log d \log T)$	Margin condition
Oh et al. [50]	$\tilde{O}(\sqrt{T \log d})$	Covariate diversity
Ariu et al. [5]	$\tilde{O}(\log d + \sqrt{T})$ $\tilde{O}(\log d + \log T)$	Covariate diversity Covariate diversity & Margin condition
Bandit with sketching		
Yu et al. [63]	$\tilde{O}(\sqrt{T} + \tau T)$	
Kuzborskij et al. [33]	$\tilde{O}(\tau d\sqrt{T})$	
Chen et al. [15]	$\tilde{O}((\sqrt{d} + \tau)\sqrt{T})$	
This paper		
Theorem 3.6	Data poor: $\tilde{O}(\log^{\frac{7}{6}} d \cdot T^{\frac{2}{3}})$	
Corollary 3.7	Data rich: $\tilde{O}(\sqrt{\log d \cdot T})$	

which matches the theoretical lower bound on T in the data-rich regime and further improves the poly-logarithmic dependence on d in the literature to sub-logarithmic $O(\sqrt{\log d})$ (e.g., [5, 8, 16, 29, 31, 50, 53, 61]). Through both synthetic experiments and Tencent's search advertising dataset, we further show that LRP-Bandit is computationally efficient (e.g., Figure 1c) and outperforms other benchmarking algorithms (Figure 1a, b, and d).

The LRP-Bandit builds on the idea of UCB and matches the theoretical minimax lower bounds on T in both data-poor and data-rich regimes under weaker assumptions than greedy-type algorithms [5, 50]. Yet, analyzing regret bounds for UCB-type algorithms with Lasso poses a unique theoretical challenge. Due to the bias in Lasso, the common unbiasedness requirement for UCB-type algorithms (e.g., [2, 36]) will no longer hold. [16] rely on the best subset selection to correct the bias, but it involves a time-consuming combinatorial optimization process. In this paper, we propose to use RP to bridge and control the selection bias issue in UCB and further restrain the information loss due to RP by only projecting high-dimensional features that are not in the feature set selected by thresholding Lasso estimators. By design, LRP-Bandit merely solves Lasso $O(\log T)$ times to trim down the computation complexity and is ready for real-world large-scale problems. To the best of our knowledge, LRP-Bandit is the first computationally efficient high-dimensional bandit algorithm that combines UCB with thresholding Lasso estimators and attends nearly optimal regret.

Related Literature: Our work is closely related to the contextual linear bandit literature. The classic UCB-type algorithms (e.g., [1, 17, 36]) are typically upper bounded by $\tilde{O}(d\sqrt{T})$. By extending to a high-dimensional sparse setting, [2] demonstrates a $\tilde{O}(\sqrt{dT})$ dependence, which is also achieved in [54] by designing a structured greedy algorithm with artificial Gaussian noise perturbation. Via RP and frequent directions, [13, 15, 33, 63] can break the polynomial dependence on d but may lead to a linear regret on T. By adopting sparse regularization, [5, 8, 16, 31, 50, 53, 61] establish ploy-logarithmic dependence on d, but the high computational complexity remains a challenge for these algorithms (e.g., see Figure 1c). With additional margin condition, [8, 61] show that the regret on T can be improved to poly-logarithmic dependence; yet, without this condition, sublinear dependence on T is the theoretical lower bound for the data-rich regime [17].

Recently, [29] propose the explore-the-sparsity-then-commit (ESTC) algorithm, which starts with a purely random exploration phase to establish a Lasso estimator and then commits to a greedy algorithm thereafter (a similar algorithm was proposed in [38] under a more general setting), and show that the ESTC algorithm achieves $\tilde{O}(T^{\frac{2}{3}} \log^{\frac{1}{3}} d)$ and its variant achieve $\tilde{O}(\sqrt{T} \log d)$ for the data-rich regime with finite arms, under which the margin condition is satisfied naturally. Instead of adopting a greedy algorithm, [16] propose a UCB-type algorithm and use the best subset selection, which requires a time-consuming combinatorial optimization procedure,

to de-bias and show that this algorithm reaches $\tilde{O}(\text{poly-log}(d)\sqrt{T})$. By introducing a covariate diversity assumption to explore the symmetric property of feature distribution, [5, 50] propose a greedy algorithm and a thresholding algorithm, both of which yield log-ploy dependence in d. [28] apply the information-directed sampling in sparse linear bandits and prove $O(\sqrt{dT})$ Bayesian regrets. In this paper, we prove that under relaxed/weaker assumptions, the UCB-based LRP-Bandit algorithm reaches $\tilde{O}(T^{\frac{2}{3}} s^{\frac{3}{2}} \log^{\frac{7}{6}} d)$ regret in the data-poor regime and $\tilde{O}(s\sqrt{T \log d})$ in the data-rich regime. To our best knowledge, it is the first UCB-type algorithm to achieve such nearly optimal bounds under high-dimensional data with a polynomial-time solution scheme.

As our algorithm uses both Lasso and random projection to perform feature selection and dimension reduction, this work is also related to these two streams of literature. In high-dimensional statistics, Lasso-type algorithms have been proposed and become standard approaches for high-dimensional feature selection [21, 41, 46, 47, 64]. In bandit setting, many algorithms are proposed to tackle the model selection problem [14, 19, 25, 27, 32, 44, 48, 49, 51, 65, 66]. Yet, these algorithms may miss some significant features in the true underlying model (i.e., model misspecification), especially under limited samples, and can be computationally challenging, therefore restraining these algorithms from being implemented in online settings. RP is one of the matrix sketching methods [18, 26, 43, 45] that approximate a high-dimensional matrix by a more compact low-dimensional one and has been proposed as a computationally efficient method to deal with high-dimensional data [24, 52]. Yet, distortions and information loss in projecting high-dimensional data into a low-dimensional space lead to significant regret loss. In this work, we couple Lasso and RP to limit both information loss and model misspecification while maintaining computational efficiency and achieving nearly optimal regret.

2 PROBLEM STATEMENT AND PRELIMINARIES

We consider a sequential decision-making process with stochastic arrivals: At each time $t \in \{1, 2, ...\}$, the decision-maker chooses an arm a_t, described by a feature vector $x_{t,a_t} \in \mathbb{R}^d$, from a decision set $\mathcal{K} = \{1, 2, ..., K\}$, and receives a reward y_t, which follows a linear form:

$$y_t = x_{t,a_t}^\top \beta^* + \epsilon_t, \tag{1}$$

where d is the total dimension, $\beta^* \in \mathbb{R}^d$ is the unknown true coefficient vector, ϵ_t is a σ^2-subgaussian random variable, and superscript $\top$ indicates the transposition operator.

The decision-maker's objective is to maximize the expected cumulative reward over T time periods. Denote the decision-maker's current policy as $\pi = \{\tilde{a}_t\}_{t \geq 1}$, where $\tilde{a}_t \in \mathcal{K}$ is the selected arm prescribed by policy π at time t. To benchmark this current policy, we define the decision-maker's expected cumulative regret up to time T under the policy π as

$$\text{Regret}(T) = \mathbb{E}\left[\sum_{t=1}^{T} [\max_{a_t \in \mathcal{K}} \{x_{t,a_t}^\top \beta^*\} - x_{t,\tilde{a}_t}^\top \beta^*]\right].$$

The decision-maker aims to select a policy π to minimize $\text{Regret}(T)$.

The contextual information is high-dimensional and exhibits a latent sparse structure. In particular, we use $\mathcal{S}^* = \{j : \beta_j^* \neq 0\}$ to denote the true index set for significant features that have non-zero coefficient values. The size of the index set (i.e., $s = |\mathcal{S}^*|$, where $|\cdot|$ denotes the cardinality of a set) is much smaller than the dimension of the feature vector (i.e., $s \ll d$), but the true index set $\mathcal{S}^*$ is unknown to the decision-maker.

Now, we first make the following technical assumption on the feature and coefficient vectors:

A. 1: There exist positive constants $x_{\max}$ and b such that $\|x_{t,a_t}\|_\infty \leq x_{\max}$ and $\|\beta\|_1 \leq b$ for all feasible t and a_t.

Assumption A.1 upper-bounds the feature and the coefficient vectors to avoid trivial decisions, which is a standard assumption in high-dimensional bandits (e.g., [5, 8, 29, 50]).

3 LRP-BANDIT ALGORITHM

In this section, we describe LRP-Bandit and establish its expected regret performance. §3.1 discusses the process of thresholding Lasso estimators to construct the selected feature set $\mathcal{S}$, §3.2 constructs the permutation matrix and the projection matrix to reduce the high-dimensional estimation problem into a low-dimensional space, and §3.3 formally present LRP-Bandit to establish its upper bound for the expected cumulative regret. For notation convenience, we will omit the subscript a_t for the chosen arm, as long as doing so will not cause any misinterpretation.

3.1 Lasso Estimator and Feature Selection

We denote $\mathcal{R}$ as the index set for *iid random samples*, i.e. , at any time $i \in \mathcal{R}$, the decision-maker randomly selects and plays an arm from his decision set with equal probability. §3.3 will detail how these random samples are generated via the random decay sampling scheme. Let n_t represent the size of the nonempty index set $\mathcal{R}$ up to time t, i.e., $n_t = |\mathcal{R}|$. The Lasso estimator for the unknown coefficient vector β^* can be defined as follows:

$$\hat{\beta} = \arg\min_{\beta} \frac{1}{n_t} \sum_{i \in \mathcal{R}} \|x_i^\top \beta - y_i\|_2^2 + \lambda \|\beta\|_1, \tag{2}$$

where $\lambda > 0$ is the regularization parameter and decreases in the random sample size n_t. Note that the Lasso estimator is identified in (2) by using only *iid random samples* (i.e., in $\mathcal{R}$), but not by using *all samples* observed up to time t. This is because that these random samples preserve the iid property necessary for the desired asymptotic performance of the Lasso estimator.

To ensure the identifiability of the Lasso estimator, we state the following compatibility condition, which is commonly adopted in the Lasso literature (e.g., [9, 11, 12]) to regulate the covariance matrix's behavior in a restricted region:

A. 2 There exists a positive constant κ such that for all vector u with $3\|u_{\mathcal{S}^*}\|_1 \geq \|u_{\mathcal{S}^{*c}}\|_1$, we have $\mathbb{E}_x\left[u^\top x x^\top u\right] \geq \frac{\kappa}{s}\|u_{\mathcal{S}^*}\|_1^2$, where $\mathcal{S}^{*c}$ denotes the complement set of $\mathcal{S}^*$.

When the random sample size n_t is large enough (i.e., on the order of $O(\log T)$), we can show that the Lasso estimator $\hat{\beta}$ will be close to the true feature coefficient β^* with high probability:

Theorem 3.1. *[Similar to Prop 1 in [8]] Let $n_t \geq O(\log T)$ and $\lambda = O\left(\sqrt{\frac{\log T + \log d}{n_t}}\right)$ for positive integers t and T. Per assumption*

A.1 and A.2, the event $\mathcal{E}_{lasso}(t) := \left\{\|\hat{\beta} - \beta^*\|_2 \le \mathcal{A}_0(t)\right\}$ *holds with probability at least* $1 - O(T^{-2})$*, where* $\mathcal{A}_0(t) = C_{lasso} \cdot s\sqrt{\frac{\log d + \log T}{n_t}}$ *and* C_{lasso} *is a positive constant.*

Proof. We first show the compatibility condition holds with high probability and then use the standard Lasso convergence result to complete the proof.

Step 1: Compatibility condition. We use the Lemma EC.6 in [8]. Note that Assumption 2 is equivalent to $\mathbb{E}[xx^\top] \in C(s, \kappa)$ in [8]. Then, for $n_t \ge 6C_2(\kappa)^{-2}\log d$ with $C_2(\kappa) = \min\left(\frac{1}{2}, \frac{\kappa^2}{256 s x_{\max}^2}\right)$, by Lemma EC.6 in [8], we have

$$\mathbb{P}\left[\frac{1}{n_t}\sum_{i\in\mathcal{R}} x_i x_i^\top \notin C(s, \frac{\kappa}{\sqrt{2}})\right] \le \exp(-C_2(\kappa)^2 n_t).$$

Step 2: Lasso convergence. We use Proposition 1 in [8]. If we set $\lambda = \frac{\chi\kappa^2}{8s}$, then the follow inequality hold:

$$\mathbb{P}\left[\left\|\hat{\beta} - \beta^*\right\|_1 > \chi\right] \le 2\exp\left[-C_1\left(\kappa/\sqrt{2}\right) n_t \chi^2 + \log d\right] + \exp(-C_2(\kappa)^2 n_t), \quad (3)$$

where $C_1\left(\kappa/\sqrt{2}\right) = \frac{\kappa^4}{2048 s^2 \sigma^2 x_{\max}^2}$. Next, when $\chi = \frac{32\sqrt{2}\sigma x_{\max}}{\kappa^2} \cdot s \cdot \sqrt{\frac{\log T + \log d}{n_t}}$ and $n_t \ge 2C_2(\kappa)^{-2}\log T$, inequality (3) implies

$$\mathbb{P}\left[\left\|\hat{\beta} - \beta^*\right\|_1 > \frac{32\sqrt{2}\sigma x_{\max}}{\kappa^2} \cdot s \cdot \sqrt{\frac{\log T + \log d}{n_t}}\right] \le \frac{3}{T^2}, \quad (4)$$

for $s \ge 2$. As we have $\|\cdot\|_1 \ge \|\cdot\|_2$, the inequality (4) holds when replacing $\left\|\hat{\beta} - \beta^*\right\|_1$ with $\left\|\hat{\beta} - \beta^*\right\|_2$. The desirable result directly follows by setting $C_{lasso} = \frac{32\sqrt{2}\sigma x_{\max}}{\kappa^2}$ and checking $n_t = \max\{6C_2(\kappa)^{-2}\log d, 2C_2(\kappa)^{-2}\log T\} \ge O(\log T)$[1] and $\lambda = \frac{\chi\kappa^2}{8s} = \frac{4\sqrt{2}\sigma x_{\max}}{\kappa^2}\sqrt{\frac{\log T + \log d}{n_t}} = O\left(\sqrt{\frac{\log T + \log d}{n_t}}\right)$. □

Note that in Theorem 3.1, as the random sample size n_t increases (e.g., by order of $O(t^c)$ for a positive constant c), $\mathcal{A}_0(t)$ decreases towards 0, which implies that the Lasso estimator asymptotically converges to its true value.

By thresholding the Lasso estimator $\hat{\beta}$, we construct a selected feature set $\mathcal{S}$, which has the following property:

Theorem 3.2. *Let the selected feature set be* $\mathcal{S} := \{j : |\hat{\beta}_j| \ge 2\mathcal{A}_0(t)\}$*. Under the event* $\mathcal{E}_{lasso}(t)$*, we have* $|\beta_j^*| \le 3\mathcal{A}_0(t)$ *for all* $j \notin \mathcal{S}$ *and* $|\mathcal{S}| \le s$*.*

Proof. By the definition of event $\mathcal{E}_{lasso}(t)$, the following inequality holds:

$$\|\hat{\beta} - \beta^*\| \le \mathcal{A}_0(t),$$

which implies $|\hat{\beta}_j| \ge |\beta_j^*| - \mathcal{A}_0(t)$ and $|\hat{\beta}_j| \le |\beta_j^*| + \mathcal{A}_0(t)$ for any j. By the definition of the index set $\mathcal{S} = \{j : |\hat{\beta}_j| \ge 2\mathcal{A}_0(t)\}$, we can show that if $j \notin \mathcal{S}$, then $|\hat{\beta}_j| < 2\mathcal{A}_0(t)$. Combining this inequality with the previous result that $|\hat{\beta}_j| \ge |\beta_j^*| - \mathcal{A}_0(t)$, we have

$$j \notin \mathcal{S} \Rightarrow |\beta_j^*| < 3\mathcal{A}_0(t).$$

Next, to prove $|\mathcal{S}| \le s$, we first note that $|\hat{\beta}_j| \ge 2\mathcal{A}_0(t)$ for $j \in \mathcal{S}$, combining which with the previous result that $|\hat{\beta}_j| \le |\beta_j^*| + \mathcal{A}_0(t)$ we have

$$j \in \mathcal{S} \Rightarrow |\beta_j^*| \ge \mathcal{A}_0(t) > 0.$$

Recall that by definition, we have $\mathcal{S}^* = \{j : \beta_j^* \ne 0\}$. Hence, $\mathcal{S} \subseteq \mathcal{S}^*$, which implies $|\mathcal{S}| \le |\mathcal{S}^*| = s$. □

The first part of Theorem 3.2 states that if a feature j is not in the selected feature set (i.e., $j \notin \mathcal{S}$), then its true coefficient value β_j^* will be small. In other words, features outside of the selected feature set $\mathcal{S}$ will have little influence on the reward. Further, note that $\mathcal{A}_0(t)$ decreases in the random sample size n_t. Therefore, if the random sample size is sufficiently large so that $3\mathcal{A}_0(t) < \min_{j\in\mathcal{S}^*}|\beta_j^*|$, then features outside of the selected feature set $\mathcal{S}$ are indeed insignificant (because $\mathcal{S} \subseteq \mathcal{S}^*$) and therefore have no influence on the reward. The second part of Theorem 3.2 suggests that the selected feature set has a lower dimension than the dimension of the true index set for significant features ($|\mathcal{S}| \le s = |\mathcal{S}^*|$). Therefore, parameter estimation for significant features can be efficiently performed in a lower dimension.

3.2 Random Projection and Coefficient Estimation

Theorem 3.2 ensures that when the random sample size is large enough, the selected feature set $\mathcal{S}$ could identify significant features that have strong influences on the decision-maker's reward. In practice, however, the decision-maker may not always have the luxury of obtaining sufficient random samples, due to high costs [8] or limited time [62]. Under these scenarios, the set $\mathcal{S}$ may include insignificant features and/or exclude significant features in the underlying true model, which causes the *model misspecification* problem. Hence, as many significant features/information will be hidden outside of the selected feature set $\mathcal{S}$, ignoring these details will lead to suboptimal arm selections. Yet, estimating coefficients for all features outside of the selected feature set $\mathcal{S}$ is still time-consuming, as these features remain high-dimensional.

To efficiently extract information contained in features outside of the selected feature set $\mathcal{S}$, we will project these high-dimensional features to a low-dimensional space via RP and then estimate coefficients for features both in the selected feature set $\mathcal{S}$ and in the projected low-dimensional space. In particular, we project high-dimensional $(d - |\mathcal{S}|)$ features outside of the selected feature set $\mathcal{S}$ into a low-dimensional m projected features by multiplying a Gaussian RP matrix $P \in \mathbb{R}^{m\times(d-|\mathcal{S}|)}$. The following Lemma shows that after RP, the distance between the original high-dimensional vector and the low-dimensional projected vector can be bounded.

Lemma 3.3. *[similar to Lemma 2 in [6]] Let* P *be a random* $n \times m$ *matrix with elements chosen independently from* $N(0, 1/m)$*. For any vector* $u \in \mathbb{R}^n$ *and* $\epsilon \in (0, \frac{1}{2}]$*, with probability at least* $1 - 2\exp(-\frac{1}{8}\epsilon^2 m)$*, the event* $\mathcal{E}_{rp}(m, n, \epsilon) := \{(1-\epsilon)\|u\|_2 \le \|Pu\|_2 \le (1+\epsilon)\|u\|_2\}$ *holds.*

[1] Here we ignore the dependence on s and $\log d$.

Proof. From Lemma 2 in [6], with probability $1-2\exp(-\frac{m}{4}(\epsilon^2-\epsilon^3))$, we have

$$\begin{aligned}(1-\epsilon)\|u\|_2^2 \le \|Pu\|_2^2 \le (1+\epsilon)\|u\|_2^2\\ \Rightarrow \sqrt{1-\epsilon}\|u\|_2 \le \|Pu\|_2 \le \sqrt{1+\epsilon}\|u\|_2\\ \Rightarrow (1-\epsilon)\|u\|_2 \le \|Pu\|_2 \le (1+\epsilon)\|u\|_2,\end{aligned}$$

where the last inequality uses the facts $\sqrt{1-\epsilon} > 1-\epsilon$ and $\sqrt{1+\epsilon} < 1+\epsilon$ when $\epsilon \in (0, \frac{1}{2}]$. The remaining task is to show the probability part, which follows the fact that $\epsilon^2 - \epsilon^3 \ge \frac{1}{2}\epsilon^2$ for $\epsilon \in (0, \frac{1}{2}]$. □

Lemma 3.3 demonstrates that RP can largely preserve the geometry structure of the original high-dimensional vector with acceptable distortions with high probability. Yet, the information loss in the process of projecting high-dimensional data to a low-dimensional space can lead to a straightly linear regret because such distortions will not vanish over time. Therefore, we will only project high-dimensional features *outside* of the selected feature set $\mathcal{S}$ to a low-dimensional space and then estimate coefficients for features in both the selected feature set $\mathcal{S}$ and the projected space. By doing so, the information loss (due to RP) will be limited by the set $\mathcal{S}$, constructed by thresholding the Lasso estimator $\hat{\beta}$, and the negative influence of model misspecification (due to Lasso) can be mitigated by including projected features outside of the selected feature set $\mathcal{S}$ via RP[2].

To this end, for a given selected feature set $\mathcal{S}$, we construct a permutation matrix $Q \in \mathbb{R}^{d\times d}$, which moves features in the selected feature set $\mathcal{S}$ of the original feature vector x to the top $|\mathcal{S}|$ places in the permuted feature vector Qx. To control the l_2 norm of feature covariate x after projection, we construct a binary sampling matrix $D_0 = \begin{pmatrix} I & 0 \\ 0 & D \end{pmatrix} \in \mathbb{R}^{d\times d}$, where $D \in \mathbb{R}^{(d-|\mathcal{S}|)\times(d-|\mathcal{S}|)}$ is a diagonal binary random matrix with q number of $(m/q)^{\frac{1}{2}}$ elements and 0 otherwise. At last, we generate a Gaussian RP matrix $P \in \mathbb{R}^{m\times(d-|\mathcal{S}|)}$ from Gaussian distribution $N(0, 1/m)$, and use this RP matrix P to construct the projection matrix $P_0 = \begin{pmatrix} I & 0 \\ 0 & P \end{pmatrix} \in \mathbb{R}^{(|\mathcal{S}|+m)\times d}$. By multiplying the projection matrix P_0 by the permuted feature vector Qx, we project the original d dimensional vector to a low-dimensional $(|\mathcal{S}| + m)$ vector, in which the first $|\mathcal{S}|$ elements are original features in the selected feature set $\mathcal{S}$ and the remaining m elements are the projected features from projecting the original high-dimensional features outside of the selected feature set $\mathcal{S}$ via the RP matrix P. Below, for simplicity, we use the following notations in the projected space: $z := P_0D_0Qx$, $\theta^* := P_0Q\beta^*$, and $\Sigma := Q^\top D_0^\top P_0^\top P_0 Q$.

Theorem 3.4. *Let t be a time index, matrices Q, D_0, and P_0 be constructed by the selected feature set $\mathcal{S} = \{j : |\hat{\beta}_j| \ge 2\mathcal{A}_0(t)\}$ and $q \ge m$. Under event $\mathcal{E}_{lasso}(t)$, when $m = O(\log T + s\log d)$, for a feasible x, the inequality $|x^\top(I-\Sigma)\beta^*| \le \mathcal{A}_1(m,t)$ with $\mathcal{A}_1(m,t) = O(s^{\frac{3}{2}}m^{-\frac{1}{2}}\log^{\frac{1}{2}} d\log T\cdot\mathcal{A}_0(t))$ holds with probability at least $1 - O(T^{-2})$.*

[2]It worth noting that besides RP, other matrix sketching methods, such as sparse random projection in [37] and frequency direction in [26], can be used as well. Due to the page limit, we will leave them for future research.

Proof. As Q is the permutation matrix and P_0 is the block diagonal matrix, we can show that

$$\begin{aligned}|x^\top(I-\Sigma)\beta^*| &= |(Qx)^\top(I - D_0P_0^\top P_0)Q\beta^*|\\ &= \underbrace{|x_{\mathcal{S}^c}^\top(I - DP^\top P)\beta^*_{\mathcal{S}^c}|}_{(a)}.\end{aligned} \tag{5}$$

As β^* is s-sparse, there will be at most s non-zero coefficients in $\beta^*_{\mathcal{S}^c}$. Without loss of generality, we assume that at most the first k elements of $\beta^*_{\mathcal{S}^c}$ are non-zero. By separating $x_{\mathcal{S}^c}$, P and D into $x_{\mathcal{S}^c} = \begin{pmatrix} x_{\mathcal{S}^c,k} \\ x_{\mathcal{S}^c,k^c}\end{pmatrix}$, $P = \begin{pmatrix} P_k & P_{k^c}\end{pmatrix}$ and $D = \begin{pmatrix} D_k & \\ & D_{k^c}\end{pmatrix}$ with $x_{\mathcal{S}^c,k} \in \mathcal{R}^{1\times k}$, $x_{\mathcal{S}^c,k^c} \in \mathcal{R}^{1\times(|\mathcal{S}^c|-k)}$, $P_k \in \mathbb{R}^{m\times k}$, $P_{k^c} \in \mathbb{R}^{m\times(|\mathcal{S}^c|-k)}$ and $D_k \in \mathbb{R}^{k\times k}$. We can directly show that (a) in (5) is upper bounded:

$$(a) \le |x_{\mathcal{S}^c,k}^\top(I - D_kP_k^\top P_k)\beta^*_{\mathcal{S}^c,k}| + |x_{\mathcal{S}^c,k^c}^\top D_{k^c}P_{k^c}^\top P_k\beta^*_{\mathcal{S}^c,k}|. \tag{6}$$

Finally, combining (5) and (6) with the Lemma 5.1 in the appendix, we can show that $|x^\top(I-\Sigma)\beta^*| \le O(s^{\frac{3}{2}}m^{-\frac{1}{2}}\log^{\frac{1}{2}} d\log T\mathcal{A}_0(t))$ holds with probability $1-O(T^{-2})$, when $m = O(\log T + s\log d)$. □

Theorem 3.4 shows that the expected reward is nearly invariant under Σ, which directly implies that our proposed projection scheme is nearly optimal in the sense that it will not introduce error when estimating the decision-maker's expected reward asymptotically. To demonstrate, note that the expected reward for a given arm k is $x_k^\top\beta^*$, whose counterpart after projection is $(P_0D_0Qx_k)^\top P_0Q\beta^* = x_k^\top\Sigma\beta^*$. Further, Theorem 3.4 demonstrates that the time dependence of the term $|x(I-\Sigma)\beta^*|$ is on the order of $\tilde{O}(\mathcal{A}_0(t)) \approx \tilde{O}(1/\sqrt{n_t})$. Therefore, if we can ensure that the order of the random sample size by time t is on the order of $O(t^c)$, where c is a positive constant, then $|x(I-\Sigma)\beta^*|$ will converge to 0 with high probability.

Using Lasso and random projection, we can project the original high-dimensional d features into a low-dimensional $(|\mathcal{S}| + m)$ for the parameter estimation. Now, we can present the estimator for the low-dimensional projected feature vector $z = P_0D_0Qx$ as follows:

$$\hat{\theta} = \arg\min_{\|\theta-\theta_0\|\le\tau}\sum_{i=1}^{t}\|z_i^\top\theta - y_i\|^2, \tag{7}$$

where τ is a positive constant selected by the decision-maker and $\theta_0 = \arg\min_\theta\|\theta - P_0Q\hat{\beta}\|$. The $\|\theta-\theta_0\| \le \tau$ is a local constraint added to (7) to prevent over-fitting. Note that we solve $\hat{\theta}$ only in the local space around θ_0, which is close to the true feature coefficient β^* with high probability.

Next, we formally state the performance of the projected estimator $\hat{\theta}$:

Theorem 3.5. *Let $\delta = \|\hat{\theta} - \theta^*\|$ and the time index $t \le T$. If conditions in Theorem 3.4 hold, then there exists a positive term C_3 such that the following event*

$$\begin{aligned}\sum_i^t|z_i^\top(\hat{\theta}-\theta^*)|^2 \le 18C_3(s+m)\log(T) + 4x_{\max}\sqrt{s+mt}\mathcal{A}_1(m,t)\delta\\ + 2(t\mathcal{A}_1(m,t)^2 + \sqrt{t}\log T\sigma\mathcal{A}_1(m,t))\end{aligned} \tag{8}$$

holds with probability at least $1 - O(T^{-2})$.

Proof. First, we define the following two functions to simplify notation:

$$f_t(\theta) := \mathbb{E}\left[\|(P_0Qx_t)^\top\theta - y_t\|^2 - \epsilon_t^2\right],$$
$$\hat{f}_t(\theta) := \|(P_0Qx_t)^\top\theta - y_t\|^2 - \epsilon_t^2.$$

By the standard covering number arguments (e.g., [59]), an ϵ covering set $\mathcal{H}(\epsilon)$ for $\|\theta - \theta^*\| \le \tau$ has a finite element upper bound of $\exp(\tilde{C}_3(s+m)\log(\tau/\epsilon))$, where parameters $\epsilon, \tau, \tilde{C}_3 > 0$. Using the union bound in Lemma 5.5 in Appendix, we can show that for any $\theta \in \mathcal{H}(\epsilon)$ and $\delta_4 > 0$,

$$\mathbb{P}\left(\left|\sum_{t=1}[\hat{f}_t(\theta) - f_t(\theta)]\right| \le \frac{2}{3}\log(1/\delta_4) + \sqrt{2}C_4\sqrt{\log(1/\delta_4)\sum_{t=1} f_t(\theta)}\right) \ge 1 - \exp(\tilde{C}_3(s+m)\log(\tau/\epsilon))\delta_4,$$

where $C_4 > 0$. By setting $\delta_4 = \exp(-2\tilde{C}_3(s+m)\log(T))$, $\epsilon = \frac{1}{2}\tau$ and $C_3 = \max\{C_4^2, 1\}\tilde{C}_3$, the above inequality implies that with probability at least $1 - O(T^{-2})$, we have

$$\left|\sum_{t=1}[\hat{f}_t(\theta) - f_t(\theta)]\right| \le \frac{4}{3}C_3(s+m)\log(T) + 2\sqrt{C_3(s+m)\log(T)\sum_{t=1} f_t(\theta)}. \quad (9)$$

Next, we can upper bound $|\sum_t f_t(\hat{\theta})|$ as follows:

$$|\sum_{t=1} f_t(\hat{\theta})| \le \left|\sum_{t=1}\hat{f}_t(\hat{\theta}) - \sum_{t=1} f_t(\hat{\theta})\right| + \max\left\{0, \sum_{t=1}\hat{f}_t(\hat{\theta})\right\}. \quad (10)$$

Let $\Gamma_T := \max\left\{0, \sum_{t=1}\hat{f}_t(\hat{\theta})\right\}$ and $x = \sqrt{\sum_t f_t(\hat{\theta})}$. We combine (9) and (10) to show that x^2, or equivalently $|\sum_{t=1} f_t(\hat{\theta})|$, can be bounded with probability $1 - O(T^{-2})$ as follows:

$$x^2 \le \frac{4}{3}C_3(s+m)\log(T) + 2\sqrt{C_3(s+m)\log(T)}\cdot x + \Gamma_T. \quad (11)$$

Note that the inequality (11) can be viewed as a quadratic inequality in x. Hence, we can solve for the upper bound of x

$$x \le \frac{2\sqrt{C_3(s+m)\log(T)} + \sqrt{(4+16/3)C_3(s+m)\log(T) + 4\Gamma_T}}{2}$$
$$\le \frac{2\sqrt{C_3(s+m)\log(T)} + 4\sqrt{C_3(s+m)\log(T)} + 2\sqrt{\Gamma_T}}{2} \quad (12)$$
$$\le 3\sqrt{C_3(s+m)\log(T)} + \sqrt{\Gamma_T}, \quad (13)$$

where in (12), we first enlarge $(4 + 16/3)$ to 16 and then uses the fact that $\sqrt{a^2+b^2} \le a + b$ for $a, b \ge 0$, and in (13), we uses the fact that $(a+b)^2 \le 2a^2 + 2b^2$ for all $a, b \in \mathbb{R}$.

Finally, combining (13) with Lemma 5.4, we can show that the following inequality holds with probability $1 - O(T^{-2})$:

$$\sum_{t=1}\|z_t^\top(\theta - \theta^*)\|^2 \le 18(C_3(s+m)\log(T) + 2\Gamma_T + 4x_{\max}\sqrt{s+m}T\mathcal{A}_1(m,t)\delta.$$

Finally, via Lemma 5.6 in the appendix, with probability at least $1 - O(T^{-2})$, $\Gamma_t \le t\mathcal{A}_1(m,t)^2 + \sqrt{t}\log T\sigma\mathcal{A}_1(m,t)$, and the statement in the theorem follows immediately. □

Theorem 3.5 describes the prediction accuracy. As the selected feature set $\mathcal{S}$ may not include all significant features, and since random projection may lead to information loss, the true model that characterizes the decision-maker's reward may not be within the projected space. In particular, the first term on the right-hand-side of Eq. (8) is of $\log(T)$ dependence, which is the typical result on *confidence ellipsoids* in bandit literature [2], which states the estimation accuracy through mixing random/non-random samples. The second and the third terms on the right-hand-side of Eq. (8) are errors generated by random projection. Note that when $n_t = \tilde{O}(t^{\frac{2}{3}})$, δ and $\mathcal{A}_1(m,t)$ can be bounded by $\tilde{O}(t^{-\frac{1}{3}})$ (See Lemma 5.9 in the appendix); then the second term will appear on the order of $\tilde{O}(t^{\frac{1}{3}})$. Moreover, the third term is also on the order of $\tilde{O}(t^{\frac{1}{3}})$. Hence, the right-hand-side of Eq. (8) will be on the order of $\tilde{O}(t^{\frac{1}{3}})$.

3.3 LRP-Bandit Algorithm

Before presenting LRP-Bandit, we will need to design a sampling scheme that generates sufficient random samples, which are essential to calibrate the parameter estimation. As bandit models involve exploitation and exploration, samples generated under exploitation typically are not purely random. Therefore, we propose a random decay sampling scheme[3] to generate sufficient, but not excessive to compromise the decision-maker's reward performance, random samples.

***Random Decay Sampling Scheme*:** At the beginning of time t, the decision-maker draws a random variable r_t that follows Bernoulli distribution with a success probability $P_{c_0,c_1}(t) = \min\{1, c_0 t^{-c_1}\}$, where $c_0 > 0$ and $c_1 \in (0,1)$ are positive constants selected by the decision-maker. If $r_t = 1$, then the decision-maker randomly selects and plays an arm from his decision set with equal probability. Otherwise, the decision-maker selects the arm with the highest upper confidence bound.

Now, we can present the proposed LRP-Bandit algorithm. This algorithm can be roughly described as follows: At time t, the decision maker will first check whether the current time t is the beginning of a new epoch: If yes, then run Lasso via Eq. (2) using only random samples in the set $\mathcal{R}$. After the initial check, the decision-maker will follow the random decay sampling scheme to draw a random Bernoulli variable: If this random variable equals 1, then the decision-maker will randomly select an arm from the decision set with equal probability; otherwise, the decision-maker will update the coefficient estimation for the projected feature vector via Eq. (7), using all samples in the set $\mathcal{W}$, and then adopt a UCB-type approach to select the arm with the highest upper confidence bound.

Note that LRP-Bandit divides the total time periods into consecutive epochs, and the size/length of each epoch increases exponentially. Only at the beginning of each epoch, the decision-maker will threshold Lasso (using only random samples in the set $\mathcal{R}$) to update the selected feature set $\mathcal{S}$, the permutation matrix Q, the projection matrix P_0, the local solution θ_0, and parameter τ. By construction, the frequency of using Lasso is decreasing at an exponential rate,

[3]Another way to generate sufficient random samples is to follow the explore sparsity then exploit structure (e.g., [29]). Yet, to select the optimal length for the pure exploration phase, these algorithms require the knowledge of the value of T, which is typically unavailable at the beginning.

LRP-Bandit Algorithm

Require: Input $c_0, c_1, m, q, \lambda_0$ and integer $u \geq 2$. Initialize $t = 1$, $epoch = 1$, $\mathcal{R} = \emptyset$, $\mathcal{W} = \emptyset$, $Q = I$, $\theta_0 = \mathbf{0}$, D_0 as diagonal q-sparse matrix with all nonzero elements equal $(m/q)^{1/2}$, $P_0 \in \mathbb{R}^{m\times d}$ with i.i.d. $N(0, 1/m)$ Gaussian elements, $\{\omega_t(m)\}$, and $\tau_1 = +\infty$.

for $t = 1, 2, \ldots$ **do**
  if $t = u^{epoch}$ **then**
    Solve Eq. (2) for $\hat{\beta}$ with samples in $\mathcal{R}$ and $\lambda = \lambda_0\sqrt{(\log d + \log t)/|\mathcal{R}|}$.
    Update set $\mathcal{S} = \{j : |\hat{\beta}_j| \geq 2\mathcal{A}_0(t)\}$ and reconstruct the matrices P_0 and Q.
    Update $epoch = epoch + 1$, $\theta_0 = \arg\min \|\theta - P_0Q\hat{\beta}\|$, and $\tau_{t-1} = \mathcal{A}_0(t)$.
  end if
  Draw a random variable r_t that follows Bernoulli distribution with success probability $P_{c_0}(t)$.
  if $r_t = 1$ **then**
    Randomly select an arm $a_t \in \mathcal{K}$; set $\mathcal{R} = \mathcal{R} \cup \{t\}$.
  else
    Solve Eq. (7) for $\hat{\theta}$ with samples in $\mathcal{W}$, θ_0, and $\tau = \tau_{t-1}$.
    Find $a_t = \arg\max_a z_{t,a}^\top\hat{\theta} + \omega_t(m)\|z_{t,a}\|_{X_{t-1}^{-1}}$, where $X_{t-1} = \sum_{i=1}^{t-1} z_i z_i$, $z_{t,a} = P_0D_0Qx_{t,a}$, and $z_t = P_0D_0Qx_t$ for all t, a.
  end if
  Offer arm a_t, observe y_t; update $x_t = x_{t,a_t}$, $\tau_t = \tau_{t-1}$, and $\mathcal{W} = \mathcal{W} \cup \{t\}$.
end for

which helps alleviate the computational burden associated with solving Lasso under high-dimensional data with large sample sizes while ensuring prediction accuracy. The following theorem establishes the expected cumulative regret upper bound for LRP-Bandit.

Theorem 3.6. *Let $c_0 = O(s^{\frac{3}{2}}\log^{\frac{7}{6}} d\log T)$, $c_1 = \frac{1}{3}$, $m = O(\log T + s\log d)$, $\omega_t(m) = O((s+m)\log^{\frac{1}{2}}(t) + (s+m)^{\frac{1}{4}}\sqrt{t\mathcal{A}_1(m,t)\tau_t\log t})$, $q \geq m$, $\tau_t = O(\mathcal{A}_0(t))$, and $\mathbb{E}[P_0D_0Qx_tx_t^\top Q^\top D_0P_0^\top]$ is positive definite for all P_0, D_0 and Q. Per assumption A.1-A.2, the expected cumulative regret for LRP-Bandit is upper bounded by $\tilde{O}(T^{\frac{2}{3}}\cdot s^{\frac{3}{2}}\log^{\frac{7}{6}} d)$, where $\tilde{O}(\cdot)$ suppresses the logarithmic dependence on T.*

Proof. We first separate the expected cumulative regret under random samples from that without random samples:

$$\begin{aligned}
&\text{Regret}(T)\\
=&\mathbb{E}\left[\sum_{t\in\mathcal{R}}^{T}[\max_{a\in\mathcal{K}}\{x_{t,a}^\top\beta^*\} - x_{t,a_t}^\top\beta^*] + \sum_{t\notin\mathcal{R}}^{T}[\max_{a\in\mathcal{K}}\{x_{t,a}^\top\beta^*\} - x_{t,a_t}^\top\beta^*]\right]\\
\leq&2x_{\max}b\mathbb{E}[n_T] + \underbrace{\mathbb{E}\left[\sum_{t\notin\mathcal{R}}^{T}[\max_{a\in\mathcal{K}}\{x_{t,a}^\top\beta^*\} - x_{t,a_t}^\top\beta^*]\right]}_{(b)}, \qquad (14)
\end{aligned}$$

where (14) uses $|x_t^\top\beta| \leq \|x_t\|_\infty\|\beta\|_1 \leq x_{\max}b$ for all x_t and feasible β in Assumption A.1. Let's assume events $\mathcal{E}_{lasso}(t)$ and $\mathcal{E}_{rp}(m, d, \frac{1}{2})$ hold. Here, we require $T_0 \leq t \leq T$ and $T_0 = O(c_0T^{\frac{2}{3}})$.

Using Lemma 5.8, we can show that $c_0 = O(\log T)$ implies that the random sample size n_T being on the order $O(c_0T^{\frac{2}{3}})$ with probability at least $1 - O(T^{-2})$. Hence, the first term in (14) can be bounded as follows:

$$2x_{\max}b\mathbb{E}[n_T] = \tilde{O}(s^{\frac{3}{2}}\log^{\frac{7}{6}} dT^{\frac{2}{3}}). \qquad (15)$$

Next, we need to bound the part (b) in (14). Without loss of generality, let's assume time T is the end of epoch j, and then we can rewrite the second part of the expected cumulative regret as the sum of regret of non-random decisions in all epochs. Using Lemma 5.3, we show that the regret of non-random decisions in epoch i, denoted as $\text{Regret}_{\text{epoch}}(i)$, can be bounded with probability $1 - O(T^{-2})$:

$$\text{Regret}_{\text{epoch}}(i) \leq \sum_{t=u^{i-1}}^{u^i-1} 2\mathcal{A}_1(m,t) + g(i)\cdot\omega_{\xi_{u^i}}(m)\sqrt{u^i},$$

where $g(i) = 2x_{\max}b\sqrt{1-u^{-1}}\sqrt{2(s+m)\log\left(\frac{8u^i dx_{\max}^2}{\mu m_{u^{i-1}-1}}\right)}$ and $\xi_{u^i} = \arg\max_{t\in[u^{i-1},u^i-1]}\omega_t(m)$. Then the regret of non-random decisions up to time T can be bounded as follows:

$$\begin{aligned}
&\sum_{i=1}^{j}\text{Regret}_{\text{epoch}}(i)\\
\leq&2(u-1)x_{\max}b + \sum_{i=2}^{j}\sum_{t=u^{i-1}}^{u^i-1} 2\mathcal{A}_1(m,t) + \sum_{i=2}^{j} g(i)\cdot\omega_{\xi_{u^i}}(m)\sqrt{u^i}\\
\leq&2(u-1)x_{\max}b + \sum_{t=u}^{u^j-1} 2\mathcal{A}_1(m,t) + \underbrace{\max_{k\in[2,j]}(g(k)\omega_{\xi_{u^k}}(m))}_{:=h_{\max}}\sum_{i=2}^{j}\sqrt{u^i}\\
\leq&2(u-1)x_{\max}b + \sum_{t=u}^{T} 2\mathcal{A}_1(m,t) + h_{\max}\frac{u}{u^{1/2}-1}\cdot\sqrt{T+1}, \qquad (16)
\end{aligned}$$

where the first inequality uses the fact that the first epoch length is $u-1$ and in the last inequality we use $T = u^j - 1$ per our assumption.

Next, we compute the upper bounds for $\sum_{t=1}^{T} 2\mathcal{A}_1(m,t)$. By using the definition of $\mathcal{A}_1(m,t)$ in Theorem 3.4 and $\mathcal{A}_0(t) = C_{lasso}s\sqrt{\frac{\log t+\log d}{n_t}} \leq O(s\log^{\frac{1}{2}} T\log^{\frac{1}{2}} d\cdot n_t^{-\frac{1}{2}})$, we can show that

$$\sum_{t=u}^{T} 2\mathcal{A}_1(m,t) \leq O\left(s^{\frac{5}{2}}m^{-\frac{1}{2}}\log d\log^{\frac{3}{2}} T\cdot\sum_{t=u}^{T} n_t^{-\frac{1}{2}}\right). \qquad (17)$$

Via Lemma 5.8, we have $n_t = O(c_0 t^{\frac{2}{3}}) = O(t^{\frac{2}{3}} \cdot s^{\frac{3}{2}} \log^{\frac{7}{6}} d \log T)$ with probabilities at least $1 - O(T^{-2})$, which implies

$$\sum_{t=u}^{T} n_t^{-\frac{1}{2}} = \tilde{O}\left(T^{\frac{2}{3}} \cdot s^{-\frac{3}{4}} \log^{-\frac{7}{12}} d\right). \tag{18}$$

Combining (17) with (18) and $m = O(\log T + s \log d)$, we have

$$\sum_{t=u}^{T} 2\mathcal{A}_1(m,t) \le \tilde{O}(s^{\frac{5}{4}} \log^{\frac{5}{12}} d \cdot T^{\frac{2}{3}}). \tag{19}$$

We then consider the upper bound for $h_{\max}$. It is direct to show that for any $i \in [2, j]$,

$$xwg(i) \le 2x_{\max} b\sqrt{1-u^{-1}} \sqrt{2(s+m)\log\left(\frac{8Tdx_{\max}^2}{\mu n_{u-1}}\right)}, \tag{20}$$

where we relax the terms u^i and $n_{u^{i-1}-1}$ in $g(i)$ by T and n_{u-1} respectively. On the other hand, from the monotonicity of $\omega_t(m)$, we have $\max_{k\in[2,j]} \omega_{\xi_{u^k}}(m) = \omega_T(m)$. Combining it with (20), $\omega_T(m) = O((s+m)\log^{\frac{1}{2}}(T) + (s+m)^{\frac{1}{4}}\sqrt{T\mathcal{A}_1(m,T)\tau_T \log T})$, $m = O(\log T + s\log d)$, $\tau_T = O(\mathcal{A}_0(T))$, $n_T = \tilde{O}(s^{3/2}\log^{7/6} d \cdot T^{2/3})$, and the definitions of $\mathcal{A}_1(m,T)$, we can show $h_{\max} \le \tilde{O}(s^{\frac{3}{2}} \log^{\frac{7}{6}} d \cdot T^{\frac{1}{6}})$. Finally, combining the upper bound of $h_{\max}$ with (14), (15), (16) and (19), we can conclude that with probability $1 - O(T^{-2})$, the cumulative regret up to epoch j can be bounded as follows:

$$\text{Regret}(T|\mathcal{E}_{lasso}(T) \cap \mathcal{E}_{rp}(m,d,\frac{1}{2})) \le \tilde{O}(s^{\frac{3}{2}} T^{\frac{2}{3}} \log^{\frac{7}{6}} d). \tag{21}$$

Note that in previous proofs, we assume events $\mathcal{E}_{lasso}(t)$ and $\mathcal{E}_{rp}(m,d,\frac{1}{2})$ hold. We then build the upper bound for the probability that those two events happen simultaneously. Using Theorem 3.1, Lemma 3.3, and union bound, we have

$$\begin{aligned}\mathbb{P}(\mathcal{E}_{lasso}(T) \cap \mathcal{E}_{rp}(m,d,\frac{1}{2})) &\ge 1 - 2\exp(-\frac{1}{8}\cdot\frac{1}{4}m) - T \cdot O(T^{-2}) \\ &\ge 1 - O(T^{-1}),\end{aligned}$$

where last inequality uses $m = O(\log T + s\log d)$. Finally, via the union bound, the unconditional expected cumulative regret can be upper bounded:

$$\begin{aligned}\text{Regret}\,(T) &\le \tilde{O}(s^{\frac{3}{2}} T^{\frac{2}{3}} \log^{\frac{7}{6}} d) \cdot (1 - O(T^{-1})) + x_{\max} bT \cdot O(T^{-1}) \\ &\le \tilde{O}(s^{\frac{3}{2}} T^{\frac{2}{3}} \log^{\frac{7}{6}} d).\end{aligned}$$

□

Theorem 3.6 demonstrates that under limited samples, LRP-Bandit achieves a tight logarithmic bound on the feature dimension $O(\log^{\frac{5}{3}} d)$ and attains $O(T^{\frac{2}{3}})$ upper bound on the sample size, which matches the regret minimax lower bound on T under limited samples [29]. It is worth noting that the $O(T^{\frac{2}{3}})$ dependence can be further improved, even under limited samples, by introducing the covariate diversity assumption (e.g., see [5, 50]): The covariate diversity assumption states the symmetric behavior of the feature distribution so that the Lasso-type estimator has better accuracy and leads to improved regret performance.

LRP-Bandit is also computationally efficient. Note that the frequency of solving Lasso problems decays exceptionally. Hence, if we apply the stochastic gradient descent with variance reduction techniques (e.g., SVRG [30]), the step-wise average computation cost will be $O(d)$. In addition, projecting the high-dimensional feature x into the low-dimension feature z, solving the local regression Eq. (7) with the gradient type method, and computing X_t^{-1} will cost $O(dm)$, $O((s+m)^2)$, and $O((s+m)^3)$, respectively. In sum, the average step-wise computation cost of LRP-Bandit is on the order of $O(dm + (s+m)^3)$, which can be further reduced in practice by using sparse random projection or other efficient optimization algorithms.

3.4 Improved upper bound for large samples

When the sample size is large so that $3\mathcal{A}_0(T) < \beta_{\min}$, where $\beta_{\min} := \min_{j\in\mathcal{S}^*} |\beta_j^*|$, the regret upper bound can be further improved to $\tilde{O}(s\sqrt{T\log d})$. This condition is commonly referred to as the information minimum signal condition in literature [22, 23, 29, 42, 61]. Here, we refer to the sample size under which inequality $3\mathcal{A}_0(T) < \beta_{\min}$ holds as the data-rich regime. Particularly, in the data-rich regime, we can show that $|\beta_j^*| = 0$ for $j \notin \mathcal{S}$ (see Theorem 3.2). Next, we can prove that the selected feature set $\mathcal{S}$ recovers the true significant feature set $\mathcal{S}^*$ (see Theorem 3.2), under which it is straightforward to show the $\tilde{O}(s\sqrt{T\log d})$ dependence. Moreover, when facing the perfect selection, the remaining dimension in β^* to be projected becomes strictly 0, which suggests that there will be no distortion and information loss. We formally summarized the result in the following corollary.

Corollary 3.7. *Let $c_0 = O(s^{\frac{1}{2}}\beta_{\min}^{-2}\log^{\frac{1}{2}} d \log T)$, $c_1 = \frac{1}{2}$, $m = O(\log T + s\log^{\frac{1}{2}} d)$, $\omega_t(m) = O((s+m)\log^{\frac{1}{2}}(t))$, $\tau_t = O(\mathcal{A}_0(t))$, $q \ge m$, and for all P_0, D_0 and Q, $\mathbb{E}[P_0 D_0 Q x_t x_t^\top Q^\top D_0^\top P_0^\top]$ is positive definite. If $3\mathcal{A}_0(T) < \beta_{\min}$, then per assumption A.1-A.2, the expected cumulative regret for LRP-Bandit is upper bounded by $\tilde{O}(s\sqrt{T\log d})$.*

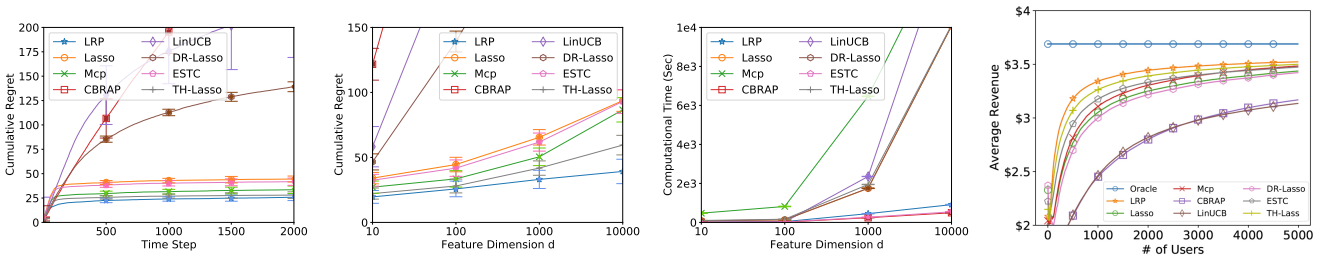
The proof steps for Corollary 3.7 are similar to that of Theorem 3.6 by changing the choices of c_0 and $\omega_t(m)$. We omit them for brevity.

Note that the information minimum signal condition will automatically hold with high probability for a large T. As the true underlying parameters β^* for any given problem is a constant for all T, we can directly show that the information minimum signal condition is satisfied for $n_t = \tilde{O}(s^2(\log T + \log d)\beta_{\min}^{-2})$, which can be attained for $t \ge O\left((s^2 \log d \beta_{\min}^{-2})^{\frac{1}{1-c_1}}\right)$. Moreover, if s and d are on the same order, the regrets in Corollary 3.7 reduce to the classic $O(d\sqrt{T})$ results (e.g.,see [1, 36]) up to some extra logarithmic terms.

Furthermore, the regret bound will remain the same, even if the decision-maker doesn't know the information minimal signal value $\beta_{\min}$ and uses a rough, or even wrong, estimation instead. Specifically, if $\beta_{\min}$ is underestimated/overestimated, then with a higher/lower sampling rate, the minimal signal condition will eventually hold so that the regret bound remains unchanged.

4 EMPIRICAL EXPERIMENTS

In this section , we benchmark LRP-Bandit to LinUCB [17], CBRAP [63], Lasso Bandit [8], MCP Bandit [61], Doubly Robust (DR), Lasso Bandit [31], ESTC [29], and Thresholded (TH) Lasso bandit [5]. LinUCB is a UCB-type contextual bandit algorithm without using dimension reduction techniques, which could potentially lead to high computational costs and poor regret performance. CBRAP uses



(a) Regret for Exp 1 (b) Regret for Exp 2 (c) Time for EXp 2 (d) Tencent Experiment

Figure 1: Hyperparameters for synthetic experiments: $c_0 = 1$, $m = \min\{30, d/2\}$, $\lambda_0 = 0.8$, $u = 2$.

RP to reduce computational time. Lasso Bandit, MCP Bandit, DR bandit, TH bandit, and ESTC algorithms replace the traditional OLS estimator with the sparse inducing estimator (e.g., Lasso and MCP) and are shown to perform well even with limited samples. Another possible benchmark is [50], which seems to have similar numerical performance in our experiments as TH Lasso Bandit and is omitted for better visual clarity. Next, we start with synthetic-data-based experiments to compare LRP-Bandit to these benchmarks in terms of regret performance and computational time. Then, we use the high-dimensional Tencent search advertising dataset to evaluate LRP-Bandit's performance in a real practice scenario where the technical assumptions specified early on may not hold. All experiments are run on a Macbook Pro Laptop with a 2.3 GHz Quad-Core Intel Core i5 CPU and 16G memory.

4.1 Synthetic Experiments

We consider a two-arm contextual linear bandit problem by varying $d = \{10, 10^2, 10^3, 10^4\}$ while keeping $s = 10$ to simulate different sparsity levels. The true coefficient vector is arbitrarily set to be $\beta^* = (1, 2, 3, 4, 5, 1.1, 2.1, 3.1, 4.1, 5.1, 0, 0...)$, and the error term ϵ_t is randomly draw from $N(0, 0.1)$. For each algorithm, we perform 100 trials and report the average cumulative regret.

The first experiment, Figure 1(a), illustrates the influence of the sample size T on the cumulative regret for the case where $d = 100$ (other cases exhibit a similar pattern and therefore are omitted). Overall, we observe that LRP-Bandit outperforms benchmarks in terms of cumulative regret. Particularly, LinUCB and CBRAP have significantly higher cumulative regret compared to other benchmarking algorithms. Further, we observe that CBRAP seems unable to converge in the experiment, which cautions the potential long-term negative influences of information loss in RP. Lasso-type bandits (i.e., Lasso, MCP, DR, TH, and ESTC) significantly reduce the decision-maker's cumulative regret from LinUCB and CBRAP. Yet, those may suffer from model misspecification due to limited samples. Therefore, by using RP to extract features outside of the significant feature set identified by thresholding the Lasso, LRP-Bandit can mitigate the negative influences of model misspecification and reduces the expected cumulative regret from Lasso-type bandits by 35% on average.

The second experiment presents the influence of the feature dimension d on the cumulative regret, Figure 1(b), and computational time, Figure 1(c), for the case $T = 1000$. As expected, the cumulative regret and computational time for all algorithms increases in d. Among all algorithms, LRP-Bandit has the lowest cumulative regret, which grows much milder than all other algorithms. Furthermore, from the computational time perspective, CBRAP, ESTC, and LRP-Bandit scale more efficient than the other algorithms and are more suitable for online decision-making.

4.2 Tencent Search Advertising Dataset

In the final experiment, we use the Tencent search advertising dataset [56] to scale up the experiment's dimensionality. For illustrative purposes, we focus on a three-ad experiment (ad IDs 21162526, 3065545, and 3827183) with 849338 entries/samples and 509256 features, and extending the experiment to include more ads or features will not qualitatively change our results and insights. For each algorithm, we perform 100 trials and report the average revenue for 5000 users. Figure 1(d) plots the average revenue performance. Due to the memory space and the computational time limits, LinUCB can not be directly implemented, so we have to resort to the matrix sketching-based LinUCB algorithm according to [33].

We observe that LRP-Bandit continues to outperform other benchmarks in terms of average revenue performance. In particular, by combining the Lasso and random projection, LRP-Bandit seems to be able to learn and accurately select the revenue-maximizing ad with a very small sample size and attain the highest revenue.

5 CONCLUSION AND FUTURE WORKS

Conclusion. In this work, we propose a computationally efficient LRP-Bandit algorithm for online contextual linear bandit problems under high-dimensional settings. We demonstrate that LRP-Bandit's expected cumulative regret is upper bounded by $\tilde{O}(T^{\frac{2}{3}} s^{\frac{3}{2}} \log^{\frac{7}{6}} d)$. With a large sample size, the expected cumulative regret can be further improved to $\tilde{O}(s\sqrt{T \log d})$. The T dependence of both results matches theoretical lower bounds. Through experiments based on synthetic and real-world datasets, we demonstrate that LRP-Bandit significantly improves the regret performance from existing benchmarking algorithms while remaining superior in computational efficiency.

Future works. In the current analysis, we require knowledge of the upper bound or the exact sparsity level s as known beforehand. It is worth exploring the situation that no such information

is available. The possible directions include utilizing the balanced covariance assumption (e.g., Assumption 6 in [50]) or switching to nonconvex sparse inducing penalty instead of Lasso [40]. Another limitation of the proposed algorithm is that we used the dense random projection matrix, which may still be computationally costly especially when the total dimension d is extremely large; hence, one possible future research direction is to further improve the computational efficiency by exploring other alternative matrix sketching methods (e.g., FD in [33]) under similar algorithmic frameworks.

REFERENCES

[1] Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. 2011. Improved algorithms for linear stochastic bandits. In *Advances in Neural Information Processing Systems*. 2312–2320.

[2] Yasin Abbasi-Yadkori, David Pal, and Csaba Szepesvari. 2012. Online-to-confidence-set conversions and application to sparse stochastic bandits. In *Artificial Intelligence and Statistics*. PMLR, 1–9.

[3] Deepak Agarwal, Bee-Chung Chen, Pradheep Elango, Nitin Motgi, Seung-Taek Park, Raghu Ramakrishnan, Scott Roy, and Joe Zachariah. 2009. Online models for content optimization. In *Advances in Neural Information Processing Systems*. 17–24.

[4] Shipra Agrawal and Navin Goyal. 2013. Thompson sampling for contextual bandits with linear payoffs. In *International Conference on Machine Learning*. 127–135.

[5] Kaito Ariu, Kenshi Abe, and Alexandre Proutière. 2022. Thresholded lasso bandit. In *International Conference on Machine Learning*. PMLR, 878–928.

[6] Rosa I. Arriaga and Santosh Vempala. 2006. An algorithmic theory of learning: Robust concepts and random projection. *Machine Learning* 63, 2 (2006), 161–182. https://doi.org/10.1007/s10994-006-6265-7

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

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

[9] Peter J Bickel, Ya'acov Ritov, Alexandre B Tsybakov, et al. 2009. Simultaneous analysis of Lasso and Dantzig selector. *The Annals of Statistics* 37, 4 (2009), 1705–1732.

[10] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al. 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning* 3, 1 (2011), 1–122.

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

[12] Emmanuel Candes, Terence Tao, et al. 2007. The Dantzig selector: Statistical estimation when p is much larger than n. *The annals of Statistics* 35, 6 (2007), 2313–2351.

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

[14] Niladri Chatterji, Vidya Muthukumar, and Peter Bartlett. 2020. Osom: A simultaneously optimal algorithm for multi-armed and linear contextual bandits. In *International Conference on Artificial Intelligence and Statistics*. PMLR, 1844–1854.

[15] Cheng Chen, Luo Luo, Weinan Zhang, Yong Yu, and Yijiang Lian. 2020. Efficient and Robust High-Dimensional Linear Contextual Bandits. IJCAI.

[16] Yi Chen, Yining Wang, Ethan X Fang, Zhaoran Wang, and Runze Li. 2022. Nearly dimension-independent sparse linear bandit over small action spaces via best subset selection. *J. Amer. Statist. Assoc.* (2022), 1–13.

[17] Wei Chu, Lihong Li, Lev Reyzin, and Robert Schapire. 2011. Contextual bandits with linear payoff functions. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. JMLR Workshop and Conference Proceedings, 208–214.

[18] Kenneth L Clarkson and David P Woodruff. 2017. Low-rank approximation and regression in input sparsity time. *Journal of the ACM (JACM)* 63, 6 (2017), 54.

[19] Ashok Cutkosky, Christoph Dann, Abhimanyu Das, Claudio Gentile, Aldo Pacchiano, and Manish Purohit. 2021. Dynamic balancing for model selection in bandits and rl. In *International Conference on Machine Learning*. PMLR, 2276–2285.

[20] Varsha Dani, Thomas P Hayes, and Sham M Kakade. 2008. Stochastic linear optimization under bandit feedback. (2008).

[21] Jianqing Fan and Runze Li. 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association* 96, 456 (2001), 1348–1360.

[22] Jianqing Fan, Han Liu, Qiang Sun, and Tong Zhang. 2018. I-LAMM for sparse learning: Simultaneous control of algorithmic complexity and statistical error. *Annals of statistics* 46, 2 (2018), 814.

[23] Jianqing Fan, Lingzhou Xue, and Hui Zou. 2014. Strong oracle optimality of folded concave penalized estimation. *Annals of statistics* 42, 3 (2014), 819.

[24] Xiaoli Z Fern and Carla E Brodley. 2003. Random projection for high dimensional data clustering: A cluster ensemble approach. In *Proceedings of the 20th international conference on machine learning (ICML-03)*. 186–193.

[25] Dylan J Foster, Akshay Krishnamurthy, and Haipeng Luo. 2019. Model Selection for Contextual Bandits. *Advances in Neural Information Processing Systems* 32 (2019), 14741–14752.

[26] Mina Ghashami, Edo Liberty, Jeff M Phillips, and David P Woodruff. 2016. Frequent directions: Simple and deterministic matrix sketching. *SIAM J. Comput.* 45, 5 (2016), 1762–1792.

[27] Avishek Ghosh, Abishek Sankararaman, and Ramchandran Kannan. 2021. Problem-complexity adaptive model selection for stochastic linear bandits. In *International Conference on Artificial Intelligence and Statistics*. PMLR, 1396–1404.

[28] Botao Hao, Tor Lattimore, and Wei Deng. 2021. Information Directed Sampling for Sparse Linear Bandits. *Advances in Neural Information Processing Systems* (2021).

[29] Botao Hao, Tor Lattimore, and Mengdi Wang. 2020. High-Dimensional Sparse Linear Bandits. *Advances in Neural Information Processing Systems* 33 (2020), 10753–10763.

[30] Rie Johnson and Tong Zhang. 2013. Accelerating stochastic gradient descent using predictive variance reduction. *Advances in neural information processing systems* 26 (2013), 315–323.

[31] Gi-Soo Kim and Myunghee Cho Paik. 2019. Doubly-Robust Lasso Bandit. In *Advances in Neural Information Processing Systems*. 5869–5879.

[32] Sanath Kumar Krishnamurthy and Susan Athey. 2021. Optimal Model Selection in Contextual Bandits with Many Classes via Offline Oracles. *arXiv preprint arXiv:2106.06483* (2021).

[33] Ilja Kuzborskij, Leonardo Cella, and Nicolò Cesa-Bianchi. 2019. Efficient linear bandits through matrix sketching. In *The 22nd International Conference on Artificial Intelligence and Statistics*. PMLR, 177–185.

[34] Tor Lattimore, Koby Crammer, and Csaba Szepesvári. 2015. Linear Multi-Resource Allocation with Semi-Bandit Feedback.. In *NIPS*. 964–972.

[35] Lihong Li, Wei Chu, John Langford, and Robert E Schapire. 2010. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*. ACM, 661–670.

[36] Lihong Li, Yu Lu, and Dengyong Zhou. 2017. Provably optimal algorithms for generalized linear contextual bandits. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org, 2071–2080.

[37] Ping Li, Trevor J Hastie, and Kenneth W Church. 2006. Very sparse random projections. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*. 287–296.

[38] Wenjie Li, Adarsh Barik, and Jean Honorio. 2022. A simple unified framework for high dimensional bandit problems. In *International Conference on Machine Learning*. PMLR, 12619–12655.

[39] Wenhao Li, Ningyuan Chen, and L Jeff Hong. 2020. Dimension reduction in contextual online learning via nonparametric variable selection. *arXiv preprint arXiv:2009.08265* (2020).

[40] Hongcheng Liu, Tao Yao, Runze Li, and Yinyu Ye. 2017. Folded concave penalized sparse linear regression: sparsity, statistical performance, and algorithmic theory for local solutions. *Mathematical programming* 166, 1-2 (2017), 207–240.

[41] Po-Ling Loh and Martin J Wainwright. 2013. Regularized M-estimators with nonconvexity: Statistical and algorithmic theory for local optima. In *Advances in Neural Information Processing Systems*. 476–484.

[42] Po-Ling Loh, Martin J Wainwright, et al. 2017. Support recovery without incoherence: A case for nonconvex regularization. *Annals of Statistics* 45, 6 (2017), 2455–2482.

[43] Haipeng Luo, Alekh Agarwal, Nicolo Cesa-Bianchi, and John Langford. 2016. Efficient second order online learning by sketching. In *Advances in Neural Information Processing Systems*. 902–910.

[44] Teodor Vanislavov Marinov and Julian Zimmert. 2021. The Pareto Frontier of model selection for general Contextual Bandits. *Advances in Neural Information Processing Systems* 34 (2021).

[45] Jiří Matoušek. 2008. On variants of the Johnson–Lindenstrauss lemma. *Random Structures & Algorithms* 33, 2 (2008), 142–156.

[46] Nicolai Meinshausen, Peter Bühlmann, et al. 2006. High-dimensional graphs and variable selection with the lasso. *The annals of statistics* 34, 3 (2006), 1436–1462.

[47] Nicolai Meinshausen, Bin Yu, et al. 2009. Lasso-type recovery of sparse representations for high-dimensional data. *The annals of statistics* 37, 1 (2009), 246–270.

[48] Ahmadreza Moradipari, Yasin Abbasi-Yadkori, Mahnoosh Alizadeh, and Mohammad Ghavamzadeh. 2021. Parameter and Feature Selection in Stochastic Linear Bandits. In *The 24nd International Conference on Artificial Intelligence and Statistics*.

[49] Vidya Muthukumar and Akshay Krishnamurthy. 2021. Universal and data-adaptive algorithms for model selection in linear contextual bandits. In *The 24nd International Conference on Artificial Intelligence and Statistics*.

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

[51] Aldo Pacchiano, Christoph Dann, Claudio Gentile, and Peter Bartlett. 2020. Regret bound balancing and elimination for model selection in bandits and rl. *arXiv preprint arXiv:2012.13045* (2020).

[52] Mert Pilanci and Martin J Wainwright. 2015. Randomized sketches of convex programs with sharp guarantees. *IEEE Transactions on Information Theory* 61, 9 (2015), 5096–5115.

[53] Zhimei Ren and Zhengyuan Zhou. 2020. Dynamic batch learning in high-dimensional sparse linear contextual bandits. *arXiv preprint arXiv:2008.11918* (2020).

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

[55] Liang Tang, Romer Rosales, Ajit Singh, and Deepak Agarwal. 2013. Automatic ad format selection via contextual bandits. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*. ACM, 1587–1594.

[56] Tencent. 2012. Predict the click-through rate of ads given the query and user information. https://www.kaggle.com/c/kddcup2012-track2. Accessed: Oct 22nd, 2018.

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

[58] Joel A Tropp et al. 2015. An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning* 8, 1-2 (2015), 1–230.

[59] SA van de Geer. 2000. Empirical process theory and applications.

[60] Roman Vershynin. 2010. Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027* (2010).

[61] Xue Wang, Mike Mingcheng Wei, and Tao Yao. 2018. Minimax Concave Penalized Multi-Armed Bandit Model with High-Dimensional Convariates. In *International Conference on Machine Learning*. 5187–5195.

[62] Xue Wang, Mike Mingcheng Wei, and Tao Yao. 2018. Online Learning and Decision-Making under Generalized Linear Model with High-Dimensional Data. *Available at SSRN 3294832* (2018).

[63] Xiaotian Yu, Michael R Lyu, and Irwin King. 2017. Cbrap: Contextual bandits with random projection. In *Thirty-First AAAI Conference on Artificial Intelligence*.

[64] Cun-Hui Zhang et al. 2010. Nearly unbiased variable selection under minimax concave penalty. *The Annals of statistics* 38, 2 (2010), 894–942.

[65] Yinglun Zhu, Julian Katz-Samuels, and Robert Nowak. 2022. Near instance optimal model selection for pure exploration linear bandits. In *International Conference on Artificial Intelligence and Statistics*. PMLR, 6735–6769.

[66] Yinglun Zhu and Robert Nowak. 2022. Pareto optimal model selection in linear bandits. In *International Conference on Artificial Intelligence and Statistics*. PMLR, 6793–6813.

Lemma 5.1. *Let $x_{\mathcal{S}^c,k}, x_{\mathcal{S}^c,k^c}, P_kP_{k^c}, D_k$ and D_{k^c} satisfy the definition in the proof of Theorem 3.4. Per conditions in Theorem 3.4 and set $m \geq 2\max\left\{C_1^2 s, \frac{2\log T+2s\log d}{C_1}\right\}$, with probability at least $1-\frac{9}{T^2}$ we have*

$$|x_{\mathcal{S}^c,k}^\top (I - D_k P_k^\top P_k)\beta^*_{\mathcal{S}^c,k}| \leq \frac{3C_1^{3/2}s^{3/2} + 3\sqrt{2}\sqrt{s^2\log T + s^3\log d} + \sqrt{C_1}(6s\sqrt{m} + 96sm^{-1/2}\log T)}{\sqrt{C_1 m}} x_{\max}\mathcal{A}_0(t) \tag{22}$$

and

$$|x_{\mathcal{S}^c,k^c}^\top P_{k^c}^\top P_k \beta^*_{\mathcal{S}^c,k}| \leq \frac{6\sqrt{s}m + 96\sqrt{s}\log T}{\sqrt{m}} x_{\max}\mathcal{A}_0(t). \tag{23}$$

Proof. We separately bound $|x_{\mathcal{S}^c,k}^\top (I - D_k P_k^\top P_k)\beta^*_{\mathcal{S}^c,k}|$ and $|x_{\mathcal{S}^c,k^c}^\top D_{k^c} P_{k^c}^\top P_k \beta^*_{\mathcal{S}^c,k}|$.

The bound for $|x_{\mathcal{S}^c,k}^\top (I - D_k P_k^\top P_k)\beta^*_{\mathcal{S}^c,k}|$. We first separate the term into two parts as follows

$$\begin{aligned}
|x_{\mathcal{S}^c,k}^\top (I - D_k P_k^\top P_k)\beta^*_{\mathcal{S}^c,k}| &= |x_{\mathcal{S}^c,k}^\top (I - P_k^\top P_k + P_k^\top P_k - D_k P_k^\top P_k)\beta^*_{\mathcal{S}^c,k}| \\
&\leq |x_{\mathcal{S}^c,k}^\top (I - P_k^\top P_k)\beta^*_{\mathcal{S}^c,k}| + |x_{\mathcal{S}^c,k}^\top (I - D_k) P_k^\top P_k \beta^*_{\mathcal{S}^c,k}| \\
&\leq \underbrace{\|x_{\mathcal{S}^c,k}\|\|(I - P_k^\top P_k)\|\|\beta^*_{\mathcal{S}^c,k}\|}_{(a_1)} + \underbrace{|x_{\mathcal{S}^c,k}^\top (I - D_k) P_k^\top P_k \beta^*_{\mathcal{S}^c,k}|}_{(a_2)}
\end{aligned} \tag{24}$$

(25)

The Remark 5.40 in [60] shows that there exists a constant $C_1 > 0$ such that for any $t > 0$, the following inequality holds with probability $1 - \exp\left(-C_1 t^2\right)$:

$$\|P_k^\top P_k - I\| \leq \max\{\delta, \delta^2\}, \tag{26}$$

where

$$\delta = C_1\sqrt{\frac{k}{m}} + \frac{t}{\sqrt{m}}.$$

Via (26), the first part of Theorem 3.2 (i.e., $|\beta_j^*| \leq 3\mathcal{A}_0(T)$ for all $j \notin \mathcal{S}$), and using the union bound, we can show that for any k dimensional subspace of the original d dimensional space, (i,e., $k < s$), the term (a_1) is upper bounded as follows with probability at least $1 - \binom{d}{s}\exp(-C_1t^2) \leq 1 - d^s\exp(-C_1t^2) = 1 - \exp(-C_1t^2 + s\log d)$:

$$(a_1) \leq \max\{\delta, \delta^2\}\|x_{\mathcal{S}^c,k}\|\|\beta^*_{\mathcal{S}^c,k}\| \leq \max\{\delta, \delta^2\} \cdot s \cdot x_{\max} \max_{j\in\mathcal{S}^c} |\beta_j^*|. \tag{27}$$

By Theorem 3.2, for all $j \in \mathcal{S}^c$, we have $|\beta_j^*| \leq 3\mathcal{A}_0(t)$, combining which with (27), we can show the following inequality:

$$(a_1) \leq 3sx_{\max}\max\{\delta, \delta^2\}\mathcal{A}_0(t). \tag{28}$$

Further, if we set $t = \sqrt{2(\log T + s\log d)/C_1}$, then we can immediately show that

$$1 - \exp(-C_1t^2 + s\log d) \geq 1 - \frac{1}{T^2}, \tag{29}$$

and

$$\delta = C_1\sqrt{\frac{k}{m}} + \sqrt{\frac{2\log T + 2s\log d}{C_1 m}}. \tag{30}$$

Finally, note that we have $k \leq s$ (see Theorem 3.2). Hence, when $m \geq 2\max\left\{C_1^2 s, \frac{2\log T+2s\log d}{C_1}\right\}$, we can show that $\delta \leq 1$, which implies $\delta^2 \leq \delta$. Then, via (28), (29), and (30), we can refine the upper bound of (a_1) in (27) as follows:

$$\begin{aligned}
(a_1) &\leq 3sx_{\max}\delta\mathcal{A}_0(t) \\
&= 3sx_{\max} \cdot \left(C_1\sqrt{\frac{k}{m}} + \sqrt{\frac{2\log T + 2s\log d}{C_1 m}}\right) \cdot \mathcal{A}_0(t) \\
&\leq \frac{3C_1^{3/2}s^{3/2} + 3\sqrt{2}\sqrt{s^2\log T + s^3\log d}}{\sqrt{C_1 m}} x_{\max}\mathcal{A}_0(t),
\end{aligned} \tag{31}$$

where we use $k \leq s$ in the last inequality.
We then build the upper bound for the term (a_2) in (24). As P_k is filled with i.i.d. $N(0, 1/m)$ random Gaussian elements, we apply Lemma 3.3 twice to have the following two inequalities:

$$\mathbb{P}\left(\|P_k\beta^*_{\mathcal{S}^c,k}\|_2 \geq (1+\epsilon)\|\beta^*_{\mathcal{S}^c,k}\|_2\right) \leq 2\exp\left(-\frac{1}{8}\epsilon^2 m\right) \tag{32}$$

and

$$\begin{aligned}&\mathbb{P}\left(\|P_k(I-D_k)x_{\mathcal{S}^c,k}\|_2 \geq (1+\epsilon)\|(I-D_k)x_{\mathcal{S}^c,k}\|_2\right) \leq 2\exp\left(-\frac{1}{8}\epsilon^2 m\right)\\ \Rightarrow&\mathbb{P}\left(\|P_k(I-D_k)x_{\mathcal{S}^c,k^c}\|_2 \geq (1+\epsilon)\sqrt{s}x_{\max}\right) \leq 2\exp\left(-\frac{1}{8}\epsilon^2 m\right),\end{aligned} \tag{33}$$

where the last inequality uses the fact that D_k is element-wise upper bounded by 1 and that $q \geq m$ and $k \leq s$.
Combining inequalities (32) and (33), we can show that

$$\begin{aligned}&\mathbb{P}\left((a_2) \geq (1+\epsilon)^2\sqrt{s}x_{\max}\|\beta^*_{\mathcal{S}^c,k}\|_2\right) \leq 4\exp\left(-\frac{1}{8}\epsilon^2 m\right)\\ \Rightarrow&\mathbb{P}\left((a_2) \geq 2(1+\epsilon^2)\sqrt{s}x_{\max}\|\beta^*_{\mathcal{S}^c,k}\|\right) \leq 4\exp\left(-\frac{1}{8}\epsilon^2 m\right),\end{aligned} \tag{34}$$

where we use the observation that $(a+b)^2 \leq 2(a^2+b^2)$ for all $a, b \in \mathbb{R}$. By setting $\epsilon = \sqrt{16\frac{\log T}{m}}$ for (34), we can show the following inequality:

$$\mathbb{P}\left((a_2) \geq 2\left(1+\frac{16\log T}{m}\right)\sqrt{s}x_{\max}\|\beta^*_{\mathcal{S}^c,k}\|\right) \leq 4\exp(-2\log T) = \frac{4}{T^2},$$

combining which with the factor that $\|\beta^*_{\mathcal{S}^c,k}\| \leq \sqrt{s}\cdot 3\mathcal{A}_0(t)$, we can establish the following probability bound for (a_2):

$$\mathbb{P}\left((a_2) \leq \frac{6sm+96s\log T}{m}x_{\max}\mathcal{A}_0(t)\right) \geq 1-\frac{4}{T^2}. \tag{35}$$

The desirable result follows by combining (24), (31), and (35).

The Bound for $|x^\top_{\mathcal{S}^c,k^c}D_{k^c}P^\top_{k^c}P_k\beta^*_{\mathcal{S}^c,k}|$. As matrix P_{k^c} is also filled with i.i.d $N(0,1/m)$ random Gaussian elements, we apply Lemma 3.3 to have

$$\begin{aligned}&\mathbb{P}\left(\|P_{k^c}D_{k^c}x_{\mathcal{S}^c,k^c}\|_2 \geq (1+\epsilon)\|D_{k^c}x_{\mathcal{S}^c,k^c}\|_2\right) \leq 2\exp\left(-\frac{1}{8}\epsilon^2 m\right)\\ \Rightarrow&\mathbb{P}\left(\|P_{k^c}D_{k^c}x_{\mathcal{S}^c,k^c}\|_2 \geq (1+\epsilon)\sqrt{m}x_{\max}\right) \leq 2\exp\left(-\frac{1}{8}\epsilon^2 m\right),\end{aligned} \tag{36}$$

where last inequality uses the fact that D_{k^c} contains at most q nonzero elements with strength $\sqrt{m/q}$. Combining inequalities (32) and (36) with the similar procedures for (34), we can show that

$$\mathbb{P}\left(|x^\top_{\mathcal{S}^c,k^c}D_{k^c}P^\top_{k^c}P_k\beta^*_{\mathcal{S}^c,k}| \geq 2(1+\epsilon^2)\sqrt{m}x_{\max}\|\beta^*_{\mathcal{S}^c,k}\|\right) \leq 4\exp\left(-\frac{1}{8}\epsilon^2 m\right), \tag{37}$$

where we use the observation that $(a+b)^2 \leq 2(a^2+b^2)$ for all $a, b \in \mathbb{R}$. By setting $\epsilon = \sqrt{16\frac{\log T}{m}}$ for (37), we can show the following inequality:

$$\mathbb{P}\left(|x^\top_{\mathcal{S}^c,k^c}D_{k^c}P^\top_{k^c}P_k\beta^*_{\mathcal{S}^c,k}| \geq \frac{2m+32\log T}{\sqrt{m}}x_{\max}\|\beta^*_{\mathcal{S}^c,k}\|\right) \leq 4\exp(-2\log T) = \frac{4}{T^2}.$$

Now, combining with $\|\beta^*_{\mathcal{S}^c,k}\| \leq \sqrt{s}\cdot 3\mathcal{A}_0(t)$, we can establish the following probability bound for $|x^\top_{\mathcal{S}^c,k^c}P^\top_{k^c}P_k\beta^*_{\mathcal{S}^c,k}|$:

$$\mathbb{P}\left(|x^\top_{\mathcal{S}^c,k^c}P^\top_{k^c}P_k\beta^*_{\mathcal{S}^c,k}| \leq \frac{6\sqrt{s}m+96\sqrt{s}\log T}{\sqrt{m}}x_{\max}\mathcal{A}_0(t)\right) \geq 1-\frac{4}{T^2}. \tag{38}$$

Finally, via union bound, we can show the desirable results hold with probability at least $1-\frac{9}{T^2}$. □

Lemma 5.2. *Let x_t^* and x_t be the optimal arm and the actually selected arm at time t, $\omega_t(m) \geq \sqrt{18(C_3(s+m)\log(t) + 2\Gamma_t + 4\sqrt{s+m}x_{\max}\delta_t t\mathcal{A}_1(m,t)}$, $\delta_t = \|\hat{\theta} - \theta^*\|$, $\Delta_t = \max_{x_t} x_t^\top \beta^* - z_t^\top \theta^*$, $n_t \geq O(\frac{(s+m)^2 \log T}{\mu})$, and $X_t = \sum_i^t z_i z_i^\top$. If there exists a μ such that $\mathbb{E}[z_t^\top z_t] \geq \mu > 0$ and conditions in Theorem 3.4 hold, then the following inequality holds with probability $1 - O(T^{-2})$:*

$$(x_T^*)^\top \beta^* - (x_T)^\top \beta^* \leq \min\left\{2x_{\max}b, 2\omega_T(m)\|z_t\|_{X_{t-1}^{-1}} + 2\Delta_T\right\}. \tag{39}$$

Proof. We denote the expected reward function in projected space as $\mathcal{R}_z(\theta) = z^\top\theta$. Then, the reward difference under the estimated coefficient vector $\hat{\theta}$ and under the true coefficient vector $\theta^* = P_0 Q\beta^*$ can be presented as

$$\begin{aligned}
|\mathcal{R}_z(\hat{\theta}) - \mathcal{R}_z(\theta^*)| &= \|z^\top(\hat{\theta} - \theta^*)\| \\
&= \sqrt{\|z^\top(\hat{\theta} - \theta^*)\|^2} \\
&= \sqrt{(\hat{\theta} - \theta^*)^\top zz^\top(\hat{\theta} - \theta^*)}.
\end{aligned} \tag{40}$$

By Lemma 5.7 in the Appendix, we know that $X_t = \sum_{i=1}^t z_i z_i^\top$ is invertible with high probability. Then, (40) implies

$$\begin{aligned}
|\mathcal{R}_z(\hat{\theta}) - \mathcal{R}_z(\theta^*)| &= \sqrt{(\hat{\theta} - \theta^*)^\top zz^\top(\hat{\theta} - \theta^*)} \\
&= \sqrt{(\hat{\theta} - \theta^*)^\top X_{t-1}^{1/2} X_{t-1}^{-1/2} zz^\top (X_{t-1}^{-1/2})^\top (X_{t-1}^{1/2})^\top (\hat{\theta} - \theta^*)} \\
&\leq \sqrt{\|(\hat{\theta} - \theta^*)^\top X_{t-1}^{1/2}\| \|X_{t-1}^{-1/2} zz^\top (X_{t-1}^{-1/2})^\top\| \|(X_{t-1}^{1/2})^\top (\hat{\theta} - \theta^*)\|} \\
&= \sqrt{\|(X_{t-1}^{1/2})^\top(\hat{\theta} - \theta^*)\|^2 \cdot z^\top X_{t-1}^{-1} z} \\
&= \sqrt{(\hat{\theta} - \theta^*)^\top X_{t-1}(\hat{\theta} - \theta^*) \cdot z^\top X_{t-1}^{-1} z} \\
&= \|z\|_{X_{t-1}^{-1}} \sqrt{\sum_{i=1}^{t-1} \|z_i^\top(\hat{\theta} - \theta^*)\|^2},
\end{aligned} \tag{41}$$

where $\|z\|_{X_{t-1}^{-1}}$ denotes the weighted 2-norm of z with matrix X_{t-1}^{-1}. Combining Theorem 3.5 and (41), with probability $1 - O(T^{-2})$ we can bound the reward difference by

$$\begin{aligned}
|\mathcal{R}_z(\hat{\theta}) - \mathcal{R}_z(\theta^*)| &\leq \|z\|_{X_{t-1}^{-1}} \sqrt{\sum_{i=1}^{t} \|z_i^\top(\hat{\theta} - \theta^*)\|^2} \\
&\leq \|z\|_{X_{t-1}^{-1}} \sqrt{18(C_3(s+m)\log(T) + 2\Gamma_t + 4\sqrt{s+m}x_{\max}\delta_t t\mathcal{A}_1(m,t)} \\
&\leq \|z\|_{X_{t-1}^{-1}} \omega_t(m),
\end{aligned} \tag{42}$$

where last inequality uses the definition of $\omega_t(m)$. Let $z^* = P_0 Q x^*$ and we then have the following bound:

$$\begin{aligned}
\mathcal{R}_{z^*}(\theta^*) - \mathcal{R}_{z_t}(\theta^*) &= \mathcal{R}_{z^*}(\theta^*) - \mathcal{R}_{z^*}(\hat{\theta}) + \mathcal{R}_{z^*}(\hat{\theta}) - \mathcal{R}_{z_t}(\hat{\theta}) + \mathcal{R}_{z_t}(\hat{\theta}) - \mathcal{R}_{z_t}(\theta^*) \\
&\leq |\mathcal{R}_{z^*}(\theta^*) - \mathcal{R}_{z^*}(\hat{\theta})| + \mathcal{R}_{z^*}(\hat{\theta}) - \mathcal{R}_{z_t}(\hat{\theta}) + |\mathcal{R}_{z_t}(\hat{\theta}) - \mathcal{R}_{z_t}(\theta^*)| \\
&\leq \|z^*\|_{X_{t-1}^{-1}} \omega_t(m) + \|z_t\|_{X_{t-1}^{-1}} \omega_t(m) + \mathcal{R}_{z^*}(\hat{\theta}) - \mathcal{R}_{z_t}(\hat{\theta}),
\end{aligned} \tag{43}$$

where (43) applies (42) on $|\mathcal{R}_{z^*}(\theta^*) - \mathcal{R}_{z^*}(\hat{\theta})|$ and $|\mathcal{R}_{z_t}(\hat{\theta}) - \mathcal{R}_{z_t}(\theta^*)|$. As z_t is selected by solving

$$\begin{aligned}
& z_t = \arg\max_z \mathcal{R}_z(\hat{\theta}) + \|z\|_{X_{t-1}^{-1}} \omega_t(m) \\
\Rightarrow & \mathcal{R}_{z^*}(\hat{\theta}) + \|z^*\|_{X_{t-1}^{-1}} \omega_t(m) \leq \mathcal{R}_{z_t}(\hat{\theta}) + \|z_t\|_{X_{t-1}^{-1}} \omega_t(m) \\
\Rightarrow & \|z^*\|_{X_{t-1}^{-1}} \omega_t(m) + \|z_t\|_{X_{t-1}^{-1}} \omega_t(m) + \mathcal{R}_{z^*}(\hat{\theta}) - \mathcal{R}_{z_t}(\hat{\theta}) \leq 2\|z_t\|_{X_{t-1}^{-1}} \omega_t(m).
\end{aligned} \tag{44}$$

Combining (43) and (44), we have

$$\mathcal{R}_{z^*}(\theta^*) - \mathcal{R}_{z_t}(\theta^*) \leq 2\|z_t\|_{X_{t-1}^{-1}} \omega_t(m).$$

As x_t^* and x_t are the optimal arm and the actually selected arm at time t, combining with the definition of Δ_t, we can bound the reward difference between the optimal arm and the actually selected arm at time t as follows

$$\begin{aligned}(x_t^*)^\top \beta^* - (x_t)^\top \beta^* &\le \mathcal{R}_{z^*}(\theta^*) - \mathcal{R}_{z_t}(\theta^*) + 2\Delta_t \\ &\le 2\|z_t\|_{X_{t-1}^{-1}} \omega_t(m) + 2\Delta_t. \end{aligned} \tag{45}$$

The remaining proof follows directly by using Assumption A.1. □

Lemma 5.3. *Let u and i be integers that are greater than 1, and $\omega_t(m)$ and n_t satisfy the same conditions as in Lemma 5.2. For the current epoch starting with $T_0 = u^{i-1}$ and ending with $T_1 = u^i - 1$, if $\omega_t(m) \ge 1$, there exists a $\mu > 0$ such that $\mathbb{E}[z_t^\top z_t] \ge \mu$ for all t, and conditions in Theorem 3.4 hold, then there exists a $\xi_T \in [T_0, T_1]$ such that the following inequality holds with probability $1 - O(T^{-2})$:*

$$\begin{aligned}&\textit{Regret of non-random decisions in epoch } i \\ &\le \sum_{t=T_0}^{T_1} 2\mathcal{A}_1(m,t) + 2x_{\max} b \omega_{\xi_T}(m)\sqrt{T_1+1}\sqrt{1-u^{-1}}\sqrt{2(s+m)\log\left(\frac{8Tdx_{\max}^2}{\mu n_{T_0-1}}\right)}. \end{aligned} \tag{46}$$

Moreover, if $\mathcal{A}_0(T) \ge \frac{1}{3}\beta_{\min}$ or $\mathcal{A}_0(T_0) < \frac{1}{3}\beta_{\min}$, then we have $\xi_T = T_1$.

Proof. Using Lemma 5.2, we have

$$\begin{aligned}&\text{Regret of non-random decisions in epoch } i \\ &\le \sum_{t=T_0}^{T_1} \min\left\{2x_{\max} b, 2\omega_t(m)\|z_t\|_{X_{t-1}^{-1}} + 2\Delta_t\right\} \\ &\le \sum_{t=T_0}^{T_1} \left(2\Delta_t + 2x_{\max} b \omega_t(m) \min\left\{1, \|z_t\|_{X_{t-1}^{-1}}\right\}\right) \\ &= \sum_{t=T_0}^{T_1} \left(2\Delta_t + 2x_{\max} b \omega_t(m) \min\left\{1, \sqrt{\lambda_{\max}(X_{t-1}^{-1/2} z_t z_t^\top X_{t-1}^{-1/2})}\right\}\right), \end{aligned} \tag{47}$$

where we denote $\lambda_{\max}(X_{t-1}^{-1/2} z_t z_t^\top X_{t-1}^{-1/2})$ as the largest eigenvalue of matrix $X_{t-1}^{-1/2} z_t z_t^\top X_{t-1}^{-1/2}$.

Let $\xi_T = \arg\max_{\xi \in [T_0, T_1]} \omega_\xi(m)$, and we can further simplify (47) as follows

$$\begin{aligned}&\text{Regret of non-random decisions in epoch } i \\ &\le 2\sum_{t=T_0}^{T_1} \Delta_t + 2x_{\max} b \omega_{\xi_T}(m) \sum_{t=T_0}^{T_1} \min\left\{1, \sqrt{\lambda_{\max}(X_{t-1}^{-1/2} z_t z_t^\top X_{t-1}^{-1/2})}\right\} \\ &\le 2\sum_{t=T_0}^{T_1} \Delta_t + 2x_{\max} b \omega_{\xi_T}(m)\sqrt{T_1 - T_0}\sqrt{\sum_{t=T_0}^{T_1} \min\left\{1, \lambda_{\max}(X_{t-1}^{-1/2} z_t z_t^\top X_{t-1}^{-1/2})\right\}} \qquad (48) \\ &\le 2\sum_{t=T_0}^{T_1} \Delta_t + 2x_{\max} b \omega_{\xi_T}(m)\sqrt{T_1+1}\sqrt{1-u^{-1}}\sqrt{\sum_{t=T_0}^{T_1} \min\left\{1, \lambda_{\max}(X_{t-1}^{-1/2} z_t z_t^\top X_{t-1}^{-1/2})\right\}}, \qquad (49) \end{aligned}$$

where (48) uses Cauchy-Schwarz inequality and (49) uses the fact that $T_1 - T_0 = u^i - u^{i-1} - 1 \le u^i(1-u^{-1}) = (T_1+1)(1-u^{-1})$.

Let $X_t = \sum_{i=1}^{T_1} z_i z_i^\top$. Per Lemma 5.7, we know that X_t is invertible with high probability, which implies

$$\begin{aligned} X_t = X_{t-1} + z_t z_t^\top &= X_{t-1}^{1/2}(I + X_{t-1}^{-1/2} z_t z_t^\top X_{t-1}^{-1/2}) X_{t-1}^{1/2} \\ \Rightarrow \log\det X_t &= \log\det X_{t-1} + \log\det(I + X_{t-1}^{-1/2} z_t z_t^\top X_{t-1}^{-1/2}) \\ \Rightarrow \log\det X_t - \log\det X_{t-1} &\ge \log(1 + \lambda_{\max}(X_{t-1}^{-1/2} z_t z_t^\top X_{t-1}^{-1/2})). \end{aligned}$$

Together with the observation that $2\log(1+x) \ge x$ for $x \in (0,1]$, we can show the following inequality holds:

$$\begin{aligned}\min\left\{1, \lambda_{\max}(X_{t-1}^{-1/2} z_t z_t^\top X_{t-1}^{-1/2})\right\} &\le 2\log\left(1 + \lambda_{\max}(X_{t-1}^{-1/2} z_t z_t^\top X_{t-1}^{-1/2})\right) \\ &\le 2\left(\log\det X_t - \log\det X_{t-1}\right) \end{aligned} \tag{50}$$

Note that we have the following two inequalities:

$$\lambda_{\max}\left(X_{T_1}\right) \le T_1 \max_{i\in[1,T]} \lambda_{\max}(z_i z_i^\top) \le T_1 \max_{i\in[1,T_1]} \|z_t\|^2 = T_1 \max_{i\in[1,T_1]} \|P_0 D_0 Q x_t\|^2 \le 4T_1 d x_{\max}^2, \tag{51}$$

$$\mathbb{P}\left(\lambda_{\min}\left(X_{T_0-1}\right) \ge \frac{1}{2} n_{T_0-1}\mu\right) \ge 1 - O((T_0-1)^{-2}) = 1 - O(T^{-2}), \tag{52}$$

where (51) uses event $\mathcal{E}_{rp}(m, d, \frac{1}{2})$ in last inequality, and (52) uses Lemma 5.7 and $T_0 = (T_1+1)/u$. Therefore, we can show that the following inequalities holds with probability $1 - O(T^{-2})$:

$$\log\det X_T \le (s+m)\log\lambda_{\max}(X_T) \le (s+m)\log(4Td x_{\max}^2) \tag{53}$$

$$\log\det X_{T_0-1} \ge (s+m)\log\lambda_{\min}(X_{T_0-1}) \ge (s+m)\log(\frac{1}{2} n_{T_0-1}\mu), \tag{54}$$

where X_t is at most $s+m$ dimension squared matrix. Combining (50), (53) and (54), we can show that

$$\begin{aligned}
\sum_{t=T_0}^{T_1} \min\left\{1, \lambda_{\max}(X_{t-1}^{-1/2} z_t z_t^\top X_{t-1}^{-1/2})\right\} &\le \sum_{t=T_0}^{T_1} 2\left(\log\det X_t - \log\det X_{t-1}\right) \\
&= 2\left(\log\det X_{T_1} - \log\det X_{T_0-1}\right) \\
&\le 2\left((s+m)\log(4T_1 d x_{\max}^2) - (s+m)\log(\frac{1}{2} n_{T_0-1}\mu)\right) \\
&\le 2(s+m)\log\left(\frac{8T_1 d x_{\max}^2}{n_{T_0-1}\mu}\right).
\end{aligned} \tag{55}$$

Finally, plugging (55) back to (49), we will have the following inequality:

$$\begin{aligned}
&\text{Regret of non-random decisions in epoch } i \\
&\le \sum_{t=T_0}^{T_1} 2\mathcal{A}_1(m,t) + 2x_{\max} b\omega_{\xi_T}(m)\sqrt{T_1+1}\sqrt{1-u^{-1}}\sqrt{2(s+m)\log\left(\frac{8T_1 d x_{\max}^2}{\mu n_{T_0-1}}\right)}.
\end{aligned} \tag{56}$$

The inequality in (56) also uses the fact that $\Delta_t = \max_x x^\top\beta^* - z^\top\theta^* \le |x^\top(I-\Sigma)\beta^*| \le \mathcal{A}_1(m,t)$, where the last inequality comes from Theorem 3.4.

The remaining part follows directly by using the monotonicity of $\omega_t(m)$, which monotonically increases in t when t satisfies $\mathcal{A}_0(t) \ge \frac{1}{3}\beta_{\min}$ or $\mathcal{A}_0(t) < \frac{1}{3}\beta_{\min}$. Thus, we have $\xi_T = \arg\max_{t\in[T_0,T]} \omega_t(m) = T$ for $\mathcal{A}_0(T_0) < \frac{1}{3}\beta_{\min}$ or $\mathcal{A}_0(T) \ge \frac{1}{3}\beta_{\min}$.

□

Lemma 5.4. *Let $\delta = \|\theta - \theta^*\|$ and T_1 be the current time step. Under conditions specified in Theorem 3.4, for any $T_0 \in (0, T_1)$, the following inequality holds with probability $1 - O(T^{-2})$:*

$$\sum_{t=T_0}^{T_1} \|z_t^\top(\theta-\theta^*)\|^2 \le \sum_{t=T_0}^{T_1} f_t(\theta) + 4\sqrt{s+m}\, x_{\max} T_1 \delta \mathcal{A}_1(m, T_1). \tag{57}$$

Proof.

$$\begin{aligned}
f_t(\theta) &= \mathbb{E}\left[|x_t^\top Q^\top D_0 P_0^\top\theta - y|^2 - \epsilon_t^2\right] \\
&= \mathbb{E}\left[|z_t^\top\theta - x_t^\top\beta^* - \epsilon_t|^2 - \epsilon_t^2\right] \\
&= \mathbb{E}\left[|z_t^\top\theta - x_t^\top\beta^*|^2 + \epsilon_t^2 - 2\epsilon_t(z_t^\top\theta - x_t^\top\beta^*) - \epsilon_t^2\right] \\
&= |z_t^\top\theta - x_t^\top\beta^*|^2 \\
&= |z_t^\top(\theta-\theta^*) + x^\top(\Sigma - I)\beta^*|^2 \qquad (58) \\
&= |z_t^\top(\theta-\theta^*)|^2 + |x_t^\top(\Sigma-I)\beta^*|^2 + 2z_t^\top(\theta-\theta^*)x_t^\top(\Sigma-I)\beta^* \\
&\ge |z_t^\top(\theta-\theta^*)|^2 + 2z_t^\top(\theta-\theta^*)x_t^\top(\Sigma-I)\beta^* \\
&\ge |z_t^\top(\theta-\theta^*)|^2 - 2\|z_t\|\|(\theta-\theta^*)\||x_t^2(\Sigma-I)\beta^*| \\
&\ge |z_t^\top(\theta-\theta^*)|^2 - 2\|z_t\|\delta|x_t^2(\Sigma-I)\beta^*|, \qquad (59)
\end{aligned}$$

where (58) uses $\theta^* = P_0 Q\beta^*$ and $z_t = P_0 D_0 Q x_t$, and the last inequality uses $\delta = \|\theta - \theta^*\|$.

Per $\mathcal{E}_{rp}(m,d,\frac{1}{2})$, there exists a feasible β such that

$$|z_t^\top \theta| = |(P_0D_0Qx_t)^\top P_0Q\beta| \le \|P_0D_0Qx_t\|_2\|P_0Q\beta\|_2 \le 1.5\|D_0Qx_t\|_2 \cdot 1.5\|\beta\|_2 \le 4\|D_0Qx_t\|_2\|\beta\|_2. \tag{60}$$

Moreover, since we assume that event $\mathcal{E}_{rp}(m,d,\frac{1}{2})$ in Theorem 3.4 holds , we have:

$$\|P_0D_0Qx\|_2 \le (1+\frac{1}{2})\|D_0Qx\|_2 \le 2\|D_0Qx\|_2.$$

Combing $D_0 = \begin{pmatrix} I & \\ & D \end{pmatrix}$, $D_{ii} \in \{0, \sqrt{\frac{m}{q}}\}$ and $|D| = q$ with Q being a permutation matrix and $\|x_t\|_\infty \le x_{\max}$ in Assumption A.1, we can upper $\|D_0Qx_t\|_2$ as follow

$$\|D_0Qx\| \le \sqrt{|\mathcal{S}|x_{\max}^2 + q(\sqrt{\frac{m}{q}}x_{\max})^2} \le \sqrt{s+m}x_{\max}, \tag{61}$$

where the last inequality uses $|\mathcal{S}| \le s$ in Theorem 3.2.

Above inequality (61) directly suggests that for all $i > 0$

$$\|z_i\| \le 2\sqrt{s+m}x_{\max}. \tag{62}$$

Combining (59), (62) and Theorem 3.4, we can show that the following inequality holds with probability $1 - O(T^{-2})$:

$$\sum_{t=T_0}^{T_1} f_t(\theta) \ge \sum_{t=T_0}^{T_1} \|z_t^\top(\theta-\theta^*)\|^2 - 4\sqrt{s+m}x_{\max}T_1\delta\mathcal{A}_1(m,T_1).$$

□

Lemma 5.5. *Under conditions specified in Theorem 3.4, for $\delta_4 > 0$, the following inequality holds with probability at least $1-\delta_4$:*

$$\left|\sum_t [\hat{f}_t(\theta) - f_t(\theta)]\right| \le \frac{2}{3}\log(1/\delta_4) + \sqrt{2}C_4\sqrt{\log(1/\delta_4)\sum_{i=1} f_t(\theta)}, \tag{63}$$

where $C_4 = \sqrt{288(s+m)x_{\max}^2b^2 + 18\sigma^2}$.

Proof. We first construct a Doob's martingale $\{M(i), i = 0, 1, 2, ..., T\}$ as follow:

$$M(i) = \mathbb{E}[\sum_t \hat{f}_t(\theta)|\mathcal{H}_i], \tag{64}$$

where $\mathcal{H}_i = \{\hat{f}_j(\theta), j \le i\}$. Using Bernstein's inequality, we have

$$\begin{aligned} &\mathbb{P}\left(|M(T) - M(0)| \ge t\right) \le \exp\left(-\frac{t^2}{2k+2t/3}\right) \\ \Rightarrow &\mathbb{P}\left(\left|\sum_t [\hat{f}_t(\theta) - f_t(\theta)]\right| \ge t\right) \le \exp\left(-\frac{t^2}{2k+2t/3}\right), \end{aligned} \tag{65}$$

where $k \ge \sum_{i=1}^{\top} Var[M(i) - M(i-1)|\mathcal{H}_{i-1}]$. We then prove the following claim:

Claim: $\sum_{i=1}^{\top}[(4x_{\max}b + x_{\max}b)^2 + 4\sigma^2]f_t(\theta) \ge \sum_{i=1}^{\top} Var[M(i) - M(i-1)|\mathcal{H}_{i-1}]$.

First, we show that the mean difference is zero:

$$\begin{aligned} &\mathbb{E}[M(i) - M(i-1)|\mathcal{H}_{i-1}] \\ &= \mathbb{E}\left[\mathbb{E}[\sum_t \hat{f}_t(\theta)|\mathcal{H}_i] - \mathbb{E}[\sum_t \hat{f}_t(\theta)|\mathcal{H}_{i-1}]\right] \\ &= \mathbb{E}[\sum_t \hat{f}_t(\theta)] - \mathbb{E}[\sum_t \hat{f}_t(\theta)] = 0 \end{aligned} \tag{66}$$

And, the variance is

$$\begin{aligned} &Var[M(i) - M(i-1)|\mathcal{H}_{i-1}] \\ &= \mathbb{E}[(M(i) - M(i-1)|\mathcal{H}_{i-1})^2] \\ &= \mathbb{E}[(\hat{f}_i(\theta) - f_i(\theta))^2] \\ &= \mathbb{E}[\hat{f}_i(\theta)^2] - f_i(\theta)^2 \\ &\leq \mathbb{E}[\hat{f}_i(\theta)^2], \end{aligned} \tag{67}$$

where the second to last equality follows from the fact that $\mathbb{E}[\hat{f}_i(\theta)] = f_t(\theta)$.

Next, we expand $\mathbb{E}[\hat{f}_i(\theta)^2]$ as follows:

$$\begin{aligned} \mathbb{E}[\hat{f}_t(\theta)^2] &= \mathbb{E}[(\|z_t^\top \theta - y_t\|^2 - \epsilon_t^2)^2] \\ &= \mathbb{E}[(\|z_t^\top \theta - x_t^\top \beta^* - \epsilon_t\|^2 - \epsilon_t^2)^2] \\ &= \mathbb{E}[(\|z_t^\top \theta - x_t^\top \beta^*\|^2 + \epsilon_t^2 - 2\epsilon_t\|z_t^\top \theta - x_t^\top \beta^*\| - \epsilon_t^2)^2] \\ &= \mathbb{E}[(\|z_t^\top \theta - x_t^\top \beta^*\|^2 - 2\epsilon_t\|z_t^\top \theta - x_t^\top \beta^*\|)^2] \\ &= \mathbb{E}[\|z_t^\top \theta - x_t^\top \beta^*\|^4 + 4\epsilon_t^2\|z_t^\top \theta - x_t^\top \beta^*\|^2 - 2\|z_t^\top \theta - x_t^\top \beta^*\|^2(2\epsilon_t\|z_t^\top \theta - x_t^\top \beta^*\|)] \\ &= \|z_t^\top \theta - x_t^\top \beta^*\|^4 + 4\sigma^2\|z_t^\top \theta - x_t^\top \beta^*\|^2 \qquad (68) \\ &= \|z_t^\top \theta - x_t^\top \beta^*\|^2(\|z_t^\top \theta - x_t^\top \beta^*\|^2 + 4\sigma^2), \qquad (69) \end{aligned}$$

where (68) uses ϵ_t being σ^2-subgaussian random variable.
Via the same procedures of (60)-(61) in Lemma 5.4, we have

$$\|z_t\|_2 \leq 2\sqrt{s+m}x_{\max} \tag{70}$$

and

$$\|\theta\|_2 \leq 2\|\beta\| \leq 2b. \tag{71}$$

Combining (70) and (71) with Assumption A.1, we have

$$\|z_t^\top \theta - x_t^\top \beta^*\|^2 \leq (4\sqrt{(s+m)}x_{\max}b + x_{\max}b)^2 \leq (8\sqrt{(s+m)}x_{\max}b)^2, \tag{72}$$

where the last inequality uses the fact that $s + m > 1$ by construction.
From (69) and (72), we can show the following inequality:

$$\begin{aligned} \mathbb{E}[\hat{f}_t(\theta)^2] &\leq \|z_t^\top - x_t^\top \beta^*\|^2[(8\sqrt{(s+m)}x_{\max}b)^2 + 4\sigma^2] \\ &= \|z_t^\top - x_t^\top \beta^*\|^2[64(s+m)x_{\max}^2 b^2 + 4\sigma^2]. \end{aligned} \tag{73}$$

On the other hand, we have

$$\begin{aligned} f_t(\theta) &= \mathbb{E}[\|z_t^\top \theta - y_t\|^2 - \epsilon_t^2] \\ &= \mathbb{E}[\|z_t^\top \theta - x_t^\top \beta^*\|^2 + \epsilon_t^2 - 2\epsilon_t\|z_t^\top \theta - x_t^\top \beta^*\| - \epsilon_t^2] \\ &= \|z_t^\top \theta - x_t^\top \beta^*\|^2. \end{aligned} \tag{74}$$

Combining (73) and (74), we have

$$\mathbb{E}[\hat{f}_t(\theta)^2] \leq f_t(\theta)[64(s+m)x_{\max}^2 b^2 + 4\sigma^2]\,.. \tag{75}$$

Finally, the **Claim** follows directly by combining (67) and (75).

Now, we set k in (65) as $k = C_4^2 \sum_{t=1} f_t(\theta)$, where $C_4 = \sqrt{288(s+m)x_{\max}^2 b^2 + 18\sigma^2}$. From the **Claim**, we can verify that

$$\begin{aligned} k &= \frac{9}{2}[64(s+m)x_{\max}^2 b^2 + 4\sigma^2] \sum_{i=1} f_t(\theta) \\ &\geq \sum_{i=1}[64(s+m)x_{\max}^2 b^2 + 4\sigma^2] f_t(\theta) \\ &\geq \sum_{i=1} Var[M(i) - M(i-1)|\mathcal{H}_{i-1}]. \end{aligned}$$

Then, we can plug $k = C_4^2 \sum_{t=1} f_t(\theta)$ back into (65) and show that

$$\mathbb{P}\left(\left|\sum_t [\hat{f}_t(\theta) - f_t(\theta)]\right| \geq \epsilon\right) \leq \exp\left(-\frac{\epsilon^2}{2C_4^2 \sum_{i=1} f_t(\theta) + 2\epsilon/3}\right)$$
$$\Rightarrow \mathbb{P}\left(\left|\sum_t [\hat{f}_t(\theta) - f_t(\theta)]\right| \leq \epsilon\right) \geq 1 - \delta_4, \tag{76}$$

where in (76) we set $\delta_4 = \exp\left(-\epsilon^2/(2C_4^2 \sum_{i=1} f_t(\theta) + 2\epsilon/3)\right)$.
Next, we solve for ϵ from the equation $\exp\left(-\epsilon^2/(2C_4^2 \sum_{i=1} f_t(\theta) + 2\epsilon/3)\right) = \delta_4$:

$$\exp\left(-\frac{\epsilon^2}{2C_4^2 \sum_{i=1} f_t(\theta) + 2\epsilon/3}\right) = \delta_4$$
$$\frac{\epsilon^2}{2\sum_{i=1} f_t(\theta) + 2\epsilon/3} = \log(1/\delta_4)$$
$$\log(1/\delta_4)(2C_4^2 \sum_{i=1} f_t(\theta) + 2\epsilon/3) = \epsilon^2$$
$$\epsilon^2 - \frac{2}{3}\log(1/\delta_4)\epsilon - 2C_4^2 \log(1/\delta_4) \sum_{i=1} f_t(\theta) = 0 \tag{77}$$
$$\Rightarrow \frac{\frac{2}{3}\log(1/\delta_4) + \sqrt{(\frac{2}{3}\log(1/\delta_4))^2 + 8C_4^2 \log(1/\delta_4) \sum_{i=1} f_t(\theta)}}{2} = \epsilon \tag{78}$$
$$\Rightarrow \frac{\frac{2}{3}\log(1/\delta_4) + \frac{2}{3}\log(1/\delta_4) + 2\sqrt{2}C_4\sqrt{\log(1/\delta_4) \sum_{i=1} f_t(\theta)}}{2} \geq \epsilon \tag{79}$$
$$\Rightarrow \frac{2}{3}\log(1/\delta_4) + \sqrt{2}C_4\sqrt{\log(1/\delta_4) \sum_{i=1} f_t(\theta)} \geq \epsilon, \tag{80}$$

where we solve the quadratic equation (77) in (78) and use the fact that $\sqrt{a^2 + b^2} \leq a + b$ for all $a, b \geq 0$ in (79).

Finally, combining (80) and (76), we can show that the following inequality holds with probability $1 - \delta_4$:

$$\left|\sum_t [\hat{f}_t(\theta) - f_t(\theta)]\right| \leq \frac{2}{3}\log(1/\delta_4) + \sqrt{2}C_4\sqrt{\log(1/\delta_4) \sum_{i=1} f_t(\theta)} \tag{81}$$

□

Lemma 5.6. *Under conditions specified in Theorem 3.4, if $\|\theta^* - \theta_0\| \leq \tau$ where θ_0 and τ are chosen based on* (7)*, then for time index t_1, the following inequality holds with probability at least $1 - O(T^{-2})$:*

$$\Gamma_{t_1} \leq t_1 \mathcal{A}_1(m, t_1)^2 + \sqrt{t_1}\log T \sigma \mathcal{A}_1(m, t_1). \tag{82}$$

Furthermore, when $\mathcal{A}_0(t_1) < \frac{1}{3}\beta_{\min}$, we have $\Gamma_{t_1} = 0$.

Proof. Since $\hat{\theta}$ is the optimal solution for (7) and θ^* is a feasible solution, we have

$$\begin{aligned}
\Gamma_{t_1} &\leq \max\left\{0, \sum_t^{t_1} |z_t^\top \theta^* - y_t|^2 - \epsilon_t^2\right\} \\
&= \max\left\{0, \sum_t^{t_1} |z_t^\top \theta^* - x_t^\top \beta^* - \epsilon_t|^2 - \epsilon_t^2\right\} \\
&= \max\left\{0, \sum_t^{t_1} |z_t^\top \theta^* - x_t^\top \beta^*|^2 - 2\epsilon_t |z_t^\top \theta - x_t^\top \beta^*|\right\} \\
&= \max\left\{0, \sum_t^{t_1} |x_t^\top (\Sigma - I)\beta^*|^2 - 2\epsilon_t |x_t^\top (\Sigma - I)\beta^*|\right\} \\
&\leq \max\left\{0, t_1 \mathcal{A}_1(m, t_1)^2 - 2\sum_t^{t_1} \epsilon_t |x_t^\top (\Sigma - I)\beta^*|)\right\},
\end{aligned} \tag{83}$$

where (83) uses Theorem 3.4.

From Hoeffding inequality and Theorem 3.4, we know that the following inequality holds for $\alpha > 0$

$$\mathbb{P}\left(\left|\sum_{t}^{t_1} \epsilon_t |x_t^\top (\Sigma - I)\beta^*|\right| \geq \alpha\right) \leq 2\exp\left(-\frac{2\alpha^2}{t_1\sigma^2\mathcal{A}_1(m,T_1)^2}\right)$$

Setting $\alpha = \sqrt{t_1}\log T\sigma\mathcal{A}_1(m,t_1)$, we have

$$\mathbb{P}\left(2\sum_{t}^{t_1} \epsilon_t \|x_t(\Sigma - I)\beta^*\| \geq \sqrt{t_1}\log T\sigma\mathcal{A}_1(m,t_1)\right) \leq O(T^{-2}) \tag{84}$$

Accordingly, the inequality stated in Lemma 5.6 directly follows from (84) and (83). Finally, as $\mathcal{A}_1(m,t_1) = 0$ for $\mathcal{A}_0(t_1) < \frac{1}{3}\beta_{\min}$, we can conclude that $\Gamma_{t_1} = 0$ in the data-rich regime. □

Lemma 5.7. *Let $z_{\max} = \max_t \|z_t\|$, $n_{t_1} \geq O\left((s+m)^2\mu^{-1}\log T\right)$ with $t_1 > 0$ and $\mu > 0$. If $\mathbb{E}[z_t^\top z_t] \geq \mu$ for all t, then with probability at least $1 - O(T^{-2})$, we have*

$$\sum_{t=1}^{t_1} z_t^\top z_t \geq \frac{1}{2}\mu n_{t_1} I. \tag{85}$$

Proof. Since $z_t^\top z_t$ is always positive semidefinite, we will have

$$\sum_{t=1}^{t_1} z_t^\top z_t \geq \sum_{t\in\text{random sample}}^{t_1} z_t^\top z_t. \tag{86}$$

We then use the Matrix Chernoff inequalities (e.g., Theorem 5.1.1 in [58]):

$$\mathbb{P}\left(\lambda_{\min}\left(\sum_{t\in\text{random sample}}^{t_1} z_t^\top z_t\right) \leq (1-\delta)\mu_{\min}\right) \leq (s+m)\left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu_{\min}/R}, \tag{87}$$

where $\mu_{\min} \leq \lambda_{\min}\left(\sum_{t\in\text{random sample}}^{t_1}\mathbb{E}[z_t^\top z_t]\right)$ and $R \geq \lambda_{\max}(z_t^\top z_t)$ for all t. Since vector z is at most $s+m$ dimension and $\|z\| \leq z_{\max}$, we have

$$\mathbb{E}[z_t^\top z_t] \leq (s+m)z_{\max}^2 I, \tag{88}$$

and

$$\lambda_{\min}\left(\sum_{t\in\text{random sample}}^{t_1}\mathbb{E}[z_t^\top z_t]\right) \geq \sum_{t\in\text{random sample}}^{t_1}\lambda_{\min}(\mathbb{E}[z_t^\top z_t]) = n_{t_1}\mu. \tag{89}$$

Thus, we can set $\mu_{\min} := n_{t_1}\mu$ and $R := (s+m)z_{\max}^2$. If we pick $\delta = 1/2$, we then have

$$\begin{aligned}
&\mathbb{P}\left(\lambda_{\min}\left(\sum_{i\in\text{random sample}}^{t_1} z_t^\top z_t\right) \leq \frac{1}{2}n_{t_1}\mu\right) \leq (s+m)\left(\frac{e^{-1/2}}{(1/2)^{1/2}}\right)^{n_{t_1}\mu/(s+m)z_{\max}^2} \\
\Rightarrow &\mathbb{P}\left(\sum_{i\in\text{random sample}}^{t_1} z_t^\top z_t \geq \frac{1}{2}n_{t_1}\mu\right) \leq (s+m)\left(\frac{e^{-1/2}}{(1/2)^{1/2}}\right)^{n_{t_1}\mu/(s+m)z_{\max}^2}.
\end{aligned}$$

The remaining part follows by using $n_{t_1} \geq \frac{2(s+m)z_{\max}^2(2\log T - \log(s+m))}{\mu\log(e/2)} = O((s+m)^2\log T/\mu)$. □

Lemma 5.8. *Under the Random Decay Sampling Scheme with $P_{c_0,c_1}(t) = \min\{1, c_0 t^{-c_1}\}$, where $c_0 > 0$, $c_1 \in (0,1)$, the following statement holds for $t > \max\left\{c_0^{1/c_1}, \left(\frac{20}{c_0}\log T\right)^{\frac{1}{1-c_1}}\right\}$:*

$$\mathbb{P}(n_t = O(c_0 t^{1-c_1})) \geq 1 - \frac{1}{T^2}. \tag{90}$$

Proof. Up to time t, the expected total number of random decisions is

$$\mathbb{E}[n_t] = \sum_{t=1}^{t}\min\left\{1, c_0 t^{-c_1}\right\}. \tag{91}$$

When $t \geq \lfloor c_0^{1/c_1} + 1 \rfloor := t_0$, we can show that

$$\mathbb{E}[n_t] = t_0 + c_0 \sum_{i=t_0}^{t} \left[i^{-c_1} \right]$$
$$\Rightarrow \sum_{i=t_0}^{t} i^{-c_1} \leq \frac{\mathbb{E}[n_t] - t_0}{c_0} \leq \sum_{i=t_0}^{t} 2i^{-c_1} \tag{92}$$

It is directly to show that for any $c_1 \in (0,1)$, we have

$$(t-1)^{1-c_1} - (t_0-1)^{1-c_1} \leq \sum_{i=t_0}^{t} i^{-c_1} \leq t^{1-c_1} - t_0^{1-c_1}. \tag{93}$$

Accordingly, we can show that

$$(t-1)^{1-c_1} - (t_0-1)^{1-c_1} \leq \frac{\mathbb{E}[n_t] - t_0}{c_0} \leq 2(t^{1-c_1} - t_0^{1-c_1})$$
$$\Rightarrow t_0 + c_0((t-1)^{1-c_1} - (t_0-1)^{1-c_1}) \leq \mathbb{E}[n_t] \leq t_0 + 2c_0(t^{1-c_1} - t_0^{1-c_1}). \tag{94}$$

Via Chernoff bound, we have

$$\mathbb{P}\left(\frac{1}{2}\mathbb{E}[n_t] \leq n_t \leq \frac{3}{2}\mathbb{E}[n_t]\right) \geq 1 - 2\exp\left(-\frac{1}{10}\mathbb{E}[n_t]\right). \tag{95}$$

Combining (95) with (94), we can conclude that

$$\mathbb{P}\left(\frac{1}{2}\left(t_0 + c_0((t-1)^{1-c_1} - (t_0-1)^{1-c_1})\right) \leq n_t \leq \frac{3}{2}\left(2c_0(t^{1-c_1} - t_0^{1-c_1}) + t_0\right)\right)$$
$$\geq 1 - 2\exp\left(-\frac{1}{10}\left(t_0 + c_0((t-1)^{1-c_1} - (t_0-1)^{1-c_1})\right)\right)$$
$$\Rightarrow \mathbb{P}(n_t = O(c_0 t^{1-c_1})) \geq 1 - 2\exp(-\frac{c_0}{10}(t-1)^{1-c_1}). \tag{96}$$

Then when $t > \left(\frac{20}{c_0}\log T\right)^{\frac{1}{1-c_1}}$, (96) implies

$$\mathbb{P}(n_t = O(c_0 t^{1-c_1})) \geq 1 - \frac{1}{T^2}. \tag{97}$$

□

Lemma 5.9. *Let $\hat{\beta}$ be the Lasso solution to* (2)*, $\hat{\theta}$ be the solution to* (7)*, and $\theta_0 = \arg\min \|\theta - P_0 Q\hat{\beta}\|$. Under the same conditions as in Theorem 3.4, if we set $n_T = \tilde{O}(T^{2/3})$ and $\tau = \tilde{O}(T^{-1/3})$, then we have $\|\hat{\theta} - \theta^*\| \leq \tilde{O}(T^{-1/3})$.*

Proof. We first use the triangle inequality:

$$\begin{aligned}\|\hat{\theta} - \theta^*\| &\leq \|\hat{\theta} - \theta_0\| + \|\theta_0 - P_0 Q\hat{\beta}\| + \|P_0 Q\hat{\beta} - \theta^*\| \\ &\leq \|\hat{\theta} - \theta_0\| + 2\|\theta^* - P_0 Q\hat{\beta}\| \\ &\leq \|\hat{\theta} - \theta_0\| + 2\|P_0 Q\beta^* - P_0 Q\hat{\beta}\| \\ &\leq \tau + 2\|P_0 Q(\beta^* - \hat{\beta})\|,\end{aligned}$$

where the second inequality uses the fact that θ_0 is the optimal solution to $\min \|\theta - P_0 Q\hat{\beta}\|$, the third inequality uses the definition of θ^*, and the last inequality uses the fact that $\|\hat{\theta} - \theta_0\| \leq \tau$ for all local regression solutions.

As we require $\tau = \tilde{O}(T^{-1/3})$, the remaining task is to show that $\|P_0 Q(\beta^* - \hat{\beta})\| = O(T^{-1/3})$. In the statement of Theorem 3.4, we assume that the condition $\mathcal{E}_{rp}(m, d, 1)$ holds. Combining this condition with the fact that Q is a permutation matrix that won't change the magnitude, we can show the following inequality:

$$\|P_0 Q(\beta^* - \hat{\beta})\| \leq 2\|Q(\beta^* - \hat{\beta})\| = 2\|\beta^* - \hat{\beta}\| \leq 2\mathcal{A}_0(T), \tag{98}$$

where the last inequality we use the $\mathcal{E}_{lasso}(T)$. Under the condition that $n_T = \tilde{O}(T^{2/3})$, it is direct to show that $\mathcal{A}_0(T) = \sqrt{(\log d + \log T)/n_T} = \tilde{O}(T^{-1/3})$, which completes the proof. □