# Structured Stochastic Linear Bandits (DRAFT)

Nicholas Johnson, Vidyashankar Sivakumar, Arindam Banerjee
{njohnson,sivakuma,banerjee@cs.umn.edu}
Department of Computer Science and Engineering
University of Minnesota

March 10, 2016

**Abstract**

In this paper, we consider the structured stochastic linear bandit problem which is a sequential decision making problem where at each round $t$ the algorithm has to select a $p$-dimensional vector $x_t$ from a convex set after which it observes a loss $\ell_t(x_t)$. We assume the loss is a linear function of the vector and an unknown parameter $\theta^*$. We consider the problem when $\theta^*$ is structured which we characterize as having a small value according to some norm, e.g., $s$-sparse, group-sparse, etc.

We precisely characterize how the regret grows for any norm structure in terms of the Gaussian width and show regret bounds which remove a $\sqrt{p}$ term. Additionally, we provide insight into the problem by introducing a new analysis technique which depends on recent developments in structured estimation.

## 1 Introduction

In this paper, we consider the stochastic linear bandit problem which proceeds in rounds $t = 1, \ldots, T$ where at each round $t$ the algorithm selects a vector $x_t$ from some compact, convex decision set $\mathcal{X} \subset \mathbb{R}^p$ and receives a stochastic loss of $\ell_t(x_t) = \langle x_t, \theta^* \rangle + \eta_t$ where $\theta^*$ is an unknown parameter and $\eta_t$ is a noise term defined as a martingale difference sequence. The algorithm observes only $\ell_t(x_t)$ at each round $t$ and can use all previous feedback to select $x_{t+1}$. The goal of the algorithm is to minimize the cumulative loss and we measure its performance in terms of the (pseudo) regret defined as

$$R_T = \sum_{t=1}^{T} \langle x_t, \theta^* \rangle - \operatorname*{argmin}_{x^* \in \mathcal{X}} \sum_{t=1}^{T} \langle x^*, \theta^* \rangle \tag{1}$$

which compares the algorithm's cumulative loss to that of the best fixed vector $x^*$ in hindsight.

We consider the setting where $\theta^*$ is structured which we characterize as having a small value according to some norm $R(\cdot)$. Typical examples of structure include sparsity, group-sparse, etc. For such a problem, we will utilize and extend recent developments in structured estimation.

We can view the stochastic linear bandit as a linear regression problem where the goal is to estimate the unknown parameter $\theta^*$. Once $\theta^*$ is known the algorithm can compute the optimal vector $x^*$ and incur no regret thereafter. However, unlike typical linear regression, the samples $x_t$ must be

selected actively in such a way to improved the estimation while at the same time incurring little regret. In the bandit literature [13], this is known as the exploration vs exploitation trade-off. Additionally, the samples have strong dependencies with one another and with the noise $\eta_t$ which makes such a problem non-trivial.

Typical approaches for such a problem are to design an algorithm which computes an estimate $\hat{\theta}_t$ of the unknown parameter $\theta^*$ in such a way where a confidence set, usually in the shape of an ellipsoid centered at $\hat{\theta}_t$, can be constructed which contains $\theta^*$ with high-probability. Afterwhich, the algorithm ignores the estimate $\hat{\theta}_t$ and simultaneously explores by selecting a $\widetilde{\theta}_t$ from the confidence ellipsoid and exploits by selecting a $x_t$ from the decision set in order to minimize their inner product. Selecting each $x_t$ in such a way forces the confidence ellipsoids to shrink fast enough such that the regret is sublinear while affording it the ability to explore within the confidence ellipsoid. If the algorithm were to select $x_t$ by minimizing the inner product with the estimate $\hat{\theta}_t$, the regret would not be sublinear since this amounts to exploiting in every round. Therefore, it is necessary to use the confidence ellipsoid for exploration.

In this paper, we follow such an approach however, unlike previous works, we explicitly take advantage of the structure of $\theta^*$ in order to improve the regret by $\sqrt{p}$. We are able to achieve such an improvement by utilizing recent techniques in structured estimation, e.g., generic chaining, and present a geometric argument based on the Gaussian width which is a geometric measure of the size of a set.

Such analysis becomes advantageous for problems where $\theta^*$ is structured. For example, the Gaussian width for an unstructured $\theta^*$ is equal to $\sqrt{p}$ so we do not get an improvement in such a problem however, for an $s$-sparse $\theta^*$ the Gaussian width is $\sqrt{s \log p}$ which is an improvement over $\sqrt{p}$.

### 1.1 Previous Works

The study of multiarmed bandit problems dates back to the 1930s [29] and 1950s [26]. The use of upper confidence bounds was presented in [6] with their well-known UCB1 algorithm. UCB1 was design for the standard $K$-arm stochastic bandit problem where the algorithm has to choose from $K$ decisions after which a stochastic loss is drawn independently from a distribution associated with the decision and the algorithm receives the loss. For such a problem, it was shown that a regret of the order $\frac{K}{\Delta} \log T$ is achievable where $\Delta$ is the gap in performance between the best and second best arms. Such a problem is a special case of the stochastic linear bandit problem by letting the decision set be the set of basis vectors in $\mathbb{R}^p$

Along the same lines, [5, 23, 18] studied the problem where now a $p$-dimensional feature vector is provided for each of the $K$ decisions and the expected loss is a linear function of the feature vector and an unknown parameter. A regret was shown for such a problem which is of the form $O\left(\log^{3/2}(K)\sqrt{p}\sqrt{T}\right)$. However, for modern applications of bandit algorithms in recommender systems, experiment design, advertisement scheduling, etc., dependence on the cardinality of the decision set $K$ becomes infeasible since many problems have large or possible infinite decision sets.

As such, [21] provided a solution in their paper which completely characterizes the regret in terms of lower and upper bounds of the stochastic linear bandit problem by providing a UCB-based algorithm called ConfidenceBall2 which generalizes previous algorithms to general, compact, convex decision sets in $\mathbb{R}^p$. It works by computing an estimate $\hat{\theta}_t$ of the unknown parameter $\theta^*$ using ridge regression which gives an estimation error of the form $\|\theta^* - \hat{\theta}_t\|_{2,D_t} \leq \sqrt{\beta_t}$. Such a bound is a confidence set in

the shape of an ellipsoid centered at $\hat{\theta}_t$. Once the confidence ellipsoid is computed, vectors $x_t$ and $\widetilde{\theta}_t$ are selected optimistically from the decision set and confidence ellipsoid respectively, such that their inner product is minimized. They show how to set $\beta_t$ such that $\theta^*$ stays within the confidence ellipsoid with high-probability for all $t$ while having enough control over the regret such that it is sublinear in $T$. Their analysis lends insight into the problem by showing the instantaneous regret is at most the width of the confidence ellipsoid and the cumulative regret is $\widetilde{O}(p\sqrt{T})$[1]. Note, the regret bound depends on the dimensionality of $p$ and not the cardinality of the decision set, which is infinite in this setting.

Building off the work of [21], a paper by [1] shows how to construct a tighter confidence ellipsoid using a novel self-normalized tail inequality for vector-valued martingales. As a result of a tighter confidence ellipsoid, they are able to shave off a log factor from the regret and show it performs well empirically.

Previous work up to around 2011 had only considered the problem with no structural assumptions about the optimal parameter $\theta^*$. In two papers published simultaneously, [16] and [2] both consider the sparse stochastic linear bandit problem where $\theta^*$ is assumed to be $s$-sparse. [16] use a different noise model where the noise is in the parameters, i.e., $\ell_t(x_t) = \langle x_t, \theta^* \rangle + \langle x_t, \eta_t \rangle$. They use techniques from compressed sensing where they randomly select vectors then perform hard-thresholding in order to identify the subspace where $\theta^*$ lives. They compute a new decision set by taking the intersection of the subspace and the general decision set $\mathcal{X}$ and feed the new decision set into ConfidenceBall2 of [21] which is run thereafter. They show they can estimate the subspace without incurring more than $O(\sqrt{T})$ regret which is the most they can afford for a $\sqrt{T}$ bound. Then the rest of the regret is inherited from the ConfidenceBall2 algorithm where the dimensionality is now $s$ instead of $p$. As such they show a $\widetilde{O}(s\sqrt{T})$ regret.

[2] use a different approach than [16] and do not assume the decisions come from the unit $L_2$ ball. They present a method which can use the predictions of any existing online algorithm with an upper bound on its regret and convert them into a confidence set. Their algorithm's regret depends on the online algorithm used for constructing the confidence set but they show when using the algorithm SeqSEW [22] of $\widetilde{O}(\sqrt{spT})$.

In a separate research thread, several papers [7, 24, 19, 20, 10, 4, 3, 14] have studied the adversarial linear bandit problem where the optimal parameter can change with time $\theta_t^*$. Similar to the stochastic linear bandit approach, typically a confidence ellipsoid is constructed around the optimal parameter. However, different tools are used, e.g., interior-point methods, Dikin ellipsoids, etc. and the analysis is significantly different so we do not consider such problems further.

### 1.2 Overview of Contributions

Our first contribution is in the regret bound. With the exception of a couple papers [2, 16], previous work was not able to take advantage of the structure of the unknown parameter $\theta^*$. In particular, even with knowledge that $\theta^*$ has a specific structure the regret was bounded as $\widetilde{O}(p\sqrt{T})$ . Our first contribution is a precise characterization of how the regret scales with the structure of $\theta^*$. Specifically, if $\theta^*$ is structured in terms of having a small value according to norm $R(\cdot)$, e.g., $L_1$, $L_{(1,2)}$, etc., then we show that the regret scales as $\widetilde{O}(\psi(E_r)w(\Omega_R)\sqrt{p}\sqrt{T})$ where $\psi(E_r)$ is the norm compatibility constant of the restricted error set $E_r$, $\Omega_R$ is the unit norm $R(\cdot)$ ball, and $w(\Omega_R)$

[1] The $\widetilde{O}(\cdot)$ hides log factors.

is the Gaussian width of $\Omega_R$, i.e., a geometric measure of the size of the set. We show for the unstructured $\theta^*$ the regret is $\widetilde{O}(p\sqrt{T})$ which matches the regret bounds of [21]. Moreover, we shave off a factor of $\sqrt{p}$ from the regret for *any* structured $\theta^*$. Note, [21] shows a lower bound of $\Omega(p\sqrt{T})$ (Omega-O notation) though, such a lower bound is for the worst-case decision set. We do not hit the lower bound since we assume the decision set is the unit $L_2$ ball.

Our second contribution which leads to the improved regret bounds is a new, geometry-based analysis technique. The analysis relies on generic chaining [27, 28] and associated techniques and provides a geometric perspective on the problem in terms of the Gaussian width. In particular, following previous approaches, such techniques allow us to show an upper bound on the estimation error and construct a confidence ellipsoid around the estimate $\hat{\theta}_t$ containing $\theta^*$ with high-probability. However, our bounds contain geometric quantities, e.g., $\psi(E_r)$ and $w(A)$, which hold for any structured $\theta^*$. As such, regret bounds are immediately seen in terms of such geometric quantities for any structure. Additionally, similar to [21], a key insight into the regret analysis is that the instantaneous regret is at most the Gaussian width of $\Omega_R$.

Our third contribution is an algorithm for any norm structured $\theta^*$. It only requires solving a norm regularized regression problem, e.g., Lasso, at each round $t$. As such, it can be implemented immediately using existing tools.

## 2 Problem Setting

In each round $t = 1, \dots, T$ the algorithm selects a vector $x_t$ from the unit $L_2$ ball $x_t \in \mathcal{X} := \{x \in \mathbb{R}^p : \|x\|_2 \leq 1\}$ and receives a loss of $\ell_t(x_t) = \langle x_t, \theta^* \rangle + \eta_t$. The unknown parameter $\theta^* \in \mathbb{R}^p$ is fixed for all $t$ and is assumed to have $\|\theta^*\|_2 = 1$ and be structured which we characterize as having a small value according to some norm $R(\cdot)$, e.g., $s$-sparse $\theta^*$ with $R(\theta^*) = \|\theta^*\|_1$. $\eta_t$ is a martingale difference sequence (MDS) noise term, i.e., $\mathbb{E}[\eta_t] < \infty$, $\mathbb{E}[\eta_t | \mathcal{F}_{t-1}] = 0$ where $\mathcal{F}_t = \{x_1, \dots, x_{t+1}, \eta_1, \dots \eta_t\}$ is a filtration which consists of a nested sequence of $\sigma$-algebras. Therefore, $x_t$ is $\mathcal{F}_{t-1}$ measurable and $\eta_t$ is $\mathcal{F}_t$ measurable. Additionally, we assume each $\eta_t$ is bounded as $|\eta_t| \leq B$ and is independent of $x_t$.

The goal of the algorithm is to minimize the cumulative loss $\sum_t \ell_t(x_t)$. We measure the performance of the algorithm in terms of the fixed cumulative (pseudo) regret

$$R_T = \sum_{t=1}^{T} \langle x_t, \theta^* \rangle - \min_{x^* \in \mathcal{X}} \sum_{t=1}^{T} \langle x^*, \theta^* \rangle \ . \tag{2}$$

We require the algorithm's regret grows sub-linearly in $T$, i.e., $R_T \leq o(T)$, and desire it grows with the structure of $\theta^*$ rather than the dimensionality $p$ with high-probability.

Our setting is different than previous settings [21, 1, 2, 16] because we assume $\theta^*$ is (generally) structured and explicitly use such an assumption. Additionally, we show the results hold for a simple decision set, i.e., the unit $L_2$ ball rather than arbitrary, convex decision sets used in some previous work. Moreover, we assume the noise is bounded following [21] and [16] rather than a more general assumption that it is a sub-Gaussian random variable as in [1, 2].

## 3 Structured Estimation

We rely heavily on recent developments in the analysis of non-asymptotic bounds for structured estimation in high-dimensional statistics. In this section, we will discuss the main developments and tools needed for the analysis of our algorithm which can be found in the following papers [15, 11, 17, 25, 30, 12, 8].

In high-dimensional statistical estimation, one is concerned with settings in which the dimension $p$ of the parameter $\theta^*$ to be estimated is significatly larger than the sample size $n$, i.e., $p \gg n$. It is well-known that for $n$ i.i.d. samples, one can compute an estimate $\hat{\theta}_n$ using least squares regression which converges to $\theta^*$ at a rate of $O\left(\sqrt{\frac{p}{n}}\right)$. Such a convergence rate can be improved when $\theta^*$ is structured which is usually characterized or approximated as a small value according to some norm $R(\cdot)$. For such problems, estimation is performed by solving a norm regularized regression problem of the form

$$\hat{\theta}_n := \underset{\theta \in \mathbb{R}^p}{\operatorname{argmin}} \, \mathcal{L}(\theta, Z_n) + \lambda_n R(\theta) \tag{3}$$

where $\mathcal{L}(\cdot, \cdot)$ is a convex loss function[2], $Z_n$ is the dataset consisting of $i = 1, \ldots, n$ pairs $(x_i, y_i)$ where $x_i \in \mathbb{R}^p$ is a sample, $y_i \in \mathbb{R}$ is the response, and $\lambda_n$ is the regularization parameter.

For such problems, let $\hat{\Delta}_n = \hat{\theta}_n - \theta^*$ be the estimation error vector. From [8], for a design matrix $X$ constructed from independent-isotropic sub-Gaussian[3] vectors and for a suitably large $\lambda_n$, the error vector belongs to the restricted error set

$$E_r = \left\{ \Delta \in \mathbb{R}^p : R(\theta^* + \Delta) \leq R(\theta^*) + \frac{1}{\rho} R(\Delta) \right\} \tag{4}$$

where $\rho > 1$ is a constant. For such a $\rho$, $E_r$ is a restricted set of directions, in particular, the error vector $\Delta$ cannot be in the direction of $\theta^*$. Additionally, for general norms $E_r$ may not be convex.

Using the restricted error set, bounds on the estimation error can be established which hold with high-probability under two assumptions. First, the regularization parameter $\lambda_n$ is suitable large. In particular, for any $\rho > 1$, the regularization parameter $\lambda_n$ needs to satisfy

$$\lambda_n \geq \rho R^*(\nabla \mathcal{L}(\theta^*; Z_n)) \ . \tag{5}$$

Second, the loss function must satisfy restricted strong convexity (RSC) in the restricted error set $E_r$ as illustrated in [25]. Specifically, there exists a suitable constant $\kappa > 0$ such that

$$\mathcal{L}(\theta^* + \Delta) - \mathcal{L}(\theta^*) - \langle \nabla \mathcal{L}(\theta^*), \Delta \rangle \geq \kappa \|\Delta\|_2^2 \quad \forall \Delta \in E_r \ . \tag{6}$$

For the setting with squared loss, the RSC condition simplifies to the restricted eigenvalue (RE) condition

$$\frac{1}{n} \|X\Delta\|_2^2 \geq \kappa \|\Delta\|_2^2 \quad \forall \Delta \in E_r \tag{7}$$

where $X \in \mathbb{R}^{n \times p}$ is the design matrix [17].

[2] We drop the second argument when it is clear from the context

[3] Definitions of sub-Gaussian vectors and related quantities is presented in the appendix.

Assuming $\lambda_n$ is large enough and the loss function has RSC/RE, we know from [25, 8] the following bound holds with high-probability

$$\|\hat{\Delta}_n\|_2 \leq c\psi(E_r)\frac{\lambda_n}{\kappa} \tag{8}$$

where $\psi(E_r) = \sup_{u \in E_r} \frac{R(u)}{\|u\|_2}$ is a norm compatibility constant and $c > 0$ is a constant.

For a design matrix $X$ with $w^2(A)$ independent-isotropic sub-Gaussian rows, we know from [8] when computing the estimate $\hat{\theta}_n$ by solving a norm regularized regression problem (3), the estimation error for any structured $\theta^*$ is upper bounded by the norm compatibility constant $\psi(E_R)$ and the Gaussian width of the unit norm ball $\Omega_R := \{u \in \mathbb{R}^p : R(u) \leq 1\}$ as

$$\|\hat{\Delta}_n\|_2 \leq c\psi(E_r)\frac{w(\Omega_R)}{\sqrt{n}} \ . \tag{9}$$

For the unstructured $\theta^*$: $\psi(E_r) = 1$, $w(\Omega_R) = \sqrt{p}$ so $\|\hat{\Delta}_n\|_2 \leq O\left(\sqrt{\frac{p}{n}}\right)$. For the $s$-sparse $\theta^*$: $\psi(E_r) = \sqrt{s}$, $w(\Omega_R) = \sqrt{\log p}$ so $\|\hat{\Delta}_n\|_2 \leq O\left(\sqrt{\frac{s \log p}{n}}\right)$. Similar results can be computed for any general structured $\theta^*$. The key insight is that the estimation error depends on the Gaussian width of the unit norm ball and converges at a rate of the order $\frac{1}{\sqrt{n}}$.

For the stochastic linear bandit problem, one typically considers confidence bounds on the estimation using an ellipsoid centered at the estimate $\hat{\theta}_n$ computed from the Mahalanobis distance defined as

$$\|\Delta\|_{2,D} = \sqrt{\Delta^\top D \Delta}$$

where $D = X^\top X$ is the sample covariance matrix. We can transform the bound in (9) to obtain the ellipsoidal bound

$$\|\Delta\|_{2,D} \leq c\psi(E_r)\frac{\lambda_n}{\kappa}\sqrt{n} \tag{10}$$

for constant $c > 0$.

**Proof:** Proof of (10).

By the definition of a convex function we have

$$\mathcal{L}(\theta^* + \Delta) - \mathcal{L}(\theta^*) \geq \langle \nabla\mathcal{L}(\theta^*), \Delta \rangle$$

and by the definition of a dual norm we have

$$|\langle \nabla\mathcal{L}(\theta^*), \Delta \rangle| \leq R^*(\nabla\mathcal{L}(\theta^*))R(\Delta) \ .$$

By construction following (5) we get

$$R^*(\nabla\mathcal{L}(\theta^*)) \leq \frac{\lambda_n}{\rho}$$

which implies

$$|\langle \nabla\mathcal{L}(\theta^*), \Delta \rangle| \leq \frac{\lambda_n}{\rho}R(\Delta)$$
$$\Rightarrow \langle \nabla\mathcal{L}(\theta^*), \Delta \rangle \geq -\frac{\lambda_n}{\rho}R(\Delta) \ .$$

Therefore,

$$\mathcal{L}(\theta^* + \Delta) - \mathcal{L}(\theta^*) \geq -\frac{\lambda_n}{\rho} R(\Delta)$$
$$\Rightarrow |\mathcal{L}(\theta^* + \Delta) - \mathcal{L}(\theta^*)| \leq \frac{\lambda_n}{\rho} R(\Delta)$$

By the definition of the norm compatibility constant $\psi(E_r) = \sup_{u \in E_r} \frac{R(u)}{\|u\|_2}$ we get $R(\Delta) \leq \|\Delta\|_2 \psi(E_r)$ which implies

$$|\mathcal{L}(\theta^* + \Delta) - \mathcal{L}(\theta^*)| \leq \frac{\lambda_n}{\rho} \|\Delta\|_2 \psi(E_r) \ .$$

Therefore, for the squared loss, since $\mathcal{L}(\theta^* + \Delta) - \mathcal{L}(\theta^*) = \frac{1}{n}\|X\Delta\|_2^2$ we get

$$|\mathcal{L}(\theta^* + \Delta) - \mathcal{L}(\theta^*)| = \left| \frac{1}{n} \|X\Delta\|_2^2 \right| = \frac{1}{n} \|X\Delta\|_2^2 \ .$$

Therefore,

$$\frac{1}{n} \|X\Delta\|_2^2 \leq \frac{\lambda_n}{\rho} \|\Delta\|_2 \psi(E_r) \ .$$

Using the bound in (8) we obtain

$$\frac{1}{n} \|X\Delta\|_2^2 \leq \frac{\lambda_n}{\rho} \psi(E_r) \frac{\lambda_n}{\kappa} \psi(E_r) \ .$$

Finally, noting $\frac{1}{n}\|X\Delta\|_2^2 = \frac{1}{n}\|\Delta\|_{2,D}^2$ and taking the square root of both sides we get the final bound

$$\|\Delta\|_{2,D} \leq c \psi(E_r) \frac{\lambda_n}{\kappa} \sqrt{n}$$

for constant $c > 0$. ∎

# 4 Algorithm

The typical approach one takes when designing an algorithm for the stochastic linear bandit problem [21, 1, 2, 16] is to construct a confidence ellipsoid $C_t$ from $x_1, \ldots, x_t$ and $\ell_1(x_1), \ldots, \ell_t(x_t)$ such that $C_t$ contains $\theta^*$ with high-probability. Given $C_{t-1}$, $x_t$ is selected optimistically by solving the following optimization problem

$$(x_t, \widetilde{\theta}_t) := \operatorname*{argmin}_{\substack{x \in \mathcal{X} \\ \theta \in C_{t-1}}} \langle x, \theta \rangle \ . \tag{11}$$

The main technical problem is in constructing a $C_t$ which ensures sublinear regret. Such a problem is difficult because of the complicated dependencies from constructing $C_t$ based on past information involving dependent randomness.

### 4.1 Algorithm

The algorithm is designed to take advantage of recent ideas from structured estimation (refer to Section 3). The main idea is to select as many random independent-isotropic vectors as we can afford in order to compute an initial estimator which has bounded estimation error with high-probability. After the initial estimation, we select vectors optimistically according to (11) to incur little regret whilst maintaining a bound on the estimation error which decreases with time. In the following sections, we describe each of the steps and in Section 5 we analyze the algorithm and prove bounds on the estimation error and regret.

**Algorithm 1** Structured Stochastic Linear Bandit

Input: $p$, $R(\cdot)$, $T$, $w^2(A)$
$C_n :=$ Random_Estimation$(p, R(\cdot), T, w^2(A))$
For $t = n+1, \ldots, T$
  Compute $x_t := \operatorname{argmin}_{\|x\|_2 \le 1,\, \theta \in C_{t-1}} \langle x, \theta \rangle$
  Play $x_t$ and receive loss $\ell_t(x_t)$
  Update $X_t$, set $D_t = X_t^\top X_t$, and update $y_t$.
  Set $\alpha = \sqrt{2 \log T} \frac{\phi}{2w(\Omega_R)}$
  Set $\lambda_t = 2LB(1+\alpha)\frac{w(\Omega_R)}{\sqrt{t}}$
  Set $\beta_t = \frac{2LB}{\kappa}(1+\alpha)\psi(E_r)w(\Omega_R)$
  Compute $\hat{\theta}_t = \operatorname*{argmin}_{\theta \in \mathbb{R}^p} \frac{1}{t}\|y_t - X_t\theta\|_2^2 + \lambda_t R(\theta)$
  Construct $C_t := \{\theta : \|\theta - \hat{\theta}_t\|_{2,D_t} \le \beta_t\}$
End For

**Algorithm 2** Random_Estimation

Input: $p$, $R(\cdot)$, $T$, $w^2(A)$
  Play $n = w^2(A)$ random vectors $x_{1:n}$
  Receive losses $\ell_{1:n}$
  Construct $X_n$, $y_n$, and set $D_n = X_n^\top X_n$
  Set $\alpha = \sqrt{2 \log T} \frac{\phi}{2w(\Omega_R)}$
  Set $\lambda_n = 2LB(1+\alpha)\frac{w(\Omega_R)}{\sqrt{n}}$
  Set $\beta_n = \frac{2LB}{\kappa}(1+\alpha)\psi(E_r)w(\Omega_R)$
  Compute $\hat{\theta}_n = \operatorname*{argmin}_{\theta \in \mathbb{R}^p} \frac{1}{n}\|y_n - X_n\theta\|_2^2 + \lambda_n R(\theta)$
  Construct $C_n := \{\theta : \|\theta - \hat{\theta}_n\|_{2,D_n} \le \beta_n\}$
End For

The algorithm consists of three main steps. First, a vector $x_t$ is selected and the loss $\ell_t(x_t)$ is observed. The design matrix $X_t$ and response vector $y_t$ are updated with $x_t$ and $\ell_t(x_t)$ respectively, the sample covariance is computed as $D_t = X_t^\top X_t$, and the quantities $\lambda_t$ and $\beta_t$ are updated. Second, a new estimate is computed as

$$\hat{\theta}_t := \operatorname*{argmin}_{\theta \in \mathbb{R}^p} \frac{1}{t}\|y_t - X_t\theta\|_2^2 + \lambda_t R(\theta) \tag{12}$$

where $R(\theta)$ is a suitable norm and $\lambda_t > 0$ is the regularization parameter. Third, a confidence ellipsoid is constructed as

$$C_t := \left\{\theta \in \mathbb{R}^p : \|\theta - \hat{\theta}_t\|_{2,D_t} \le \beta_t\right\} . \tag{13}$$

For the first step, the algorithm selects $x_t$ differently depending on if we are in the initial random estimation rounds or the simultaneous exploration and exploitation rounds.

#### 4.1.1 Random Estimation Rounds

For the initial rounds $t = 1, \ldots, n = w^2(A)$ the algorithm selects independent-isotropic sub-Gaussian vectors $x_{1:n} := \{x_1, \ldots, x_n\}$ and receives the corresponding losses $\ell_{1:n} := \{\ell_1(x_1), \ldots, \ell_n(x_n)\}$. After round $n$, the algorithm constructs a $(n \times p)$-dimensional design matrix $X_n$ of the vectors $x_{1:n}$ and a $n$-dimensional vector $y_n$ of the losses $\ell_{1:n}$. The first estimate $\hat{\theta}_1$ of $\theta^*$ is computed by solving (12). The confidence ellipsoid is then constructed via (13). The random estimation rounds can be considered the "burn-in" period similar to using a barycentric spanner as in Dani et al. [21].

However, we cannot continue to select random vectors because each incurs an instantaneous regret of $O(1)$ and will give a cumulative regret which is linear in $T$. Therefore, after the $n = w^2(A)$ estimation rounds, we need to actively select vectors for the remaining $T - n$ rounds in order to control the regret.

#### 4.1.2 Active Exploration and Exploitation Rounds

At round $n < t \leq T$, using the confidence ball $C_{t-1}$ the algorithm selects the vector $x_t$ via

$$(x_t, \widetilde{\theta}_t) := \underset{\substack{x \in \mathcal{X} \\ \theta \in C_{t-1}}}{\operatorname{argmin}} \langle x, \theta \rangle \tag{14}$$

and receives loss $\ell_t(x_t)$. The design matrix $X_t$, sample covariance $D_t$, and response vector $y_t$ are updated with $x_t$ and $\ell_t(x_t)$, and $\lambda_t$ and $\beta_t$ are also updated. A new estimate $\hat{\theta}_t$ is computed as (12) and using $\hat{\theta}_t$ a new confidence ellipsoid $C_t$ is computed via (13). The main technical contribution is showing that selecting vectors via (14) and for an appropriate $\beta_t$, the confidence ball $C_t$ shrinks fast enough to ensure the algorithm incurs a cumulative regret of the order $\sqrt{T}$ with high-probability. The algorithm for general structured stochastic linear bandits is presented in Algorithm 1.

### 4.2 Regret Upper Bounds

The fixed cumulative regret of Algorithm 1 is upper bounded by

$$R_T \leq \widetilde{O}\left(\psi(E_r) w(\Omega_R) \sqrt{p} \sqrt{T}\right) \tag{15}$$

with high-probability. We will prove such a result in Section 5.5. For such a proof we recall the error bound $\|\hat{\theta}_t - \theta^*\|_{2,D_t} \leq c\psi(E_r)\frac{\lambda_t}{\kappa}\sqrt{t} = \beta_t$. The bound depends deterministically on $\lambda_t$ and $\kappa$ and holds when the following two conditions are satisfied: (1) $\lambda_t$ being suitable large and (2) the loss function having RSC/RE with parameter $\kappa$. We present the analysis of such quantities in the next section.

## 5 Analysis

Recall the estimation error bound

$$\|\hat{\theta}_t - \theta^*\|_{2,D_t} \leq c\psi(E_r)\frac{\lambda_t}{\kappa}\sqrt{t} \tag{16}$$

for constant $c > 0$ where $\psi(E_R) = \sup_{u \in E_r} \frac{R(u)}{\|u\|_2}$ is the norm compatibility constant, $\lambda_t$ is the regularization parameter, and $\kappa$ is the RSC/RE constant. Such a bound is deterministic but depends on the values of $\kappa$ and $\lambda_t$. For the analysis, we will provide bounds on such terms in order for the estimation bound to hold.

The analysis will proceed as follows. First, in Section 5.2 we will show a bound on the RSC/RE parameter $\kappa$. Second, in Section 5.3 we will show a bound on regularization parameter $\lambda_t$. Third, in Section 5.4 we will show a high-probability bound that holds simulatenaously for all rounds. Finally, in Section 5.5 we will show a bound on the cumulative regret. All of the results will hold with high-probability. We begin by stating the assumptions under which our analysis holds.

### 5.1 Assumptions and Notation

We assume the number of rounds $T$ is known a priori, the norm regression loss function $\mathcal{L}(\theta^*, Z_t)$ is the squared loss, the decision set $\mathcal{X}$ is the unit $L_2$ ball, i.e., $\mathcal{X} := \{x \in \mathbb{R}^p : \|x\|_2 \leq 1\}$ thus, each $x_t$ has sub-Gaussian norm satisfying $\|x_t\|_{\psi_2} \leq 1$, the noise $\eta_t$ is a martingale difference sequence (MDS) with each $|\eta_t| \leq B$ and independent of $x_t$, and we assume the structure is known, e.g., we know the sparsity level for an $s$-sparse $\theta^*$. The following table includes notation used in the analysis and is provided for convenience.

| | |
|---|---|
| $\theta^* \in \mathbb{R}^p$ | Optimal parameter |
| $\hat{\theta}_t \in \mathbb{R}^p$ | Estimate |
| $x_t \in \mathcal{X}$ | Vector played |
| $\eta_t \in \mathbb{R}, \lvert\eta_t\rvert \leq B$ | Bounded MDS noise |
| $\ell_t(x_t) = \langle x_t, \theta^* \rangle + \eta_t$ | Loss |
| $X_t \in \mathbb{R}^{t \times p}$ | Design matrix |
| $y_t \in \mathbb{R}^t$ | Response vector |
| $\omega_t = [\eta_1 \dots \eta_t]^\top$ | Vector of noise terms |
| $\hat{\Delta}_t = \hat{\theta}_t - \theta^*$ | Error vector |
| $\beta_t \geq 0$ | Confidence ball radius |
| $C_t := \{\theta \in \mathbb{R}^p : \|\theta - \hat{\theta}_t\|_2 \leq \beta_t\}$ | Confidence ball |
| $R(\cdot), R^*(\cdot)$ | Norm, dual norm regularizer |
| $E_r := \{\Delta \in \mathbb{R}^p : R(\theta^* + \Delta) \leq R(\theta^*) + \frac{1}{\rho} R(\Delta)\}$ | Restricted error set |
| $w(A)$ | Gaussian width of set $A$ |
| $\Omega_R := \{u \in \mathbb{R}^p : R(u) \leq 1\}$ | Unit $R(\cdot)$ norm ball |
| $\mathcal{F}_t := \{x_1, \dots, x_{t+1}, \eta_1, \dots, \eta_t\}$ | Filtration |

Table 1: Notations.

### 5.2 Restricted Eigenvalue Condition

For the bound on $\|\Delta\|_{2,D_t}$, we need the restricted eigenvalue condition (RE) to hold for all vectors in the restricted error set $E_r$. For a design matrix $X$ with $n$ rows, a response vector $y$, and parameter $\kappa$ the RE condition for squared loss is

$$\begin{aligned} &\frac{1}{n}\|y - X(\theta^* + \Delta)\|_2^2 - \frac{1}{n}\|y - X\theta^*\|_2^2 - \frac{1}{n}\left\langle X^\top(y - X\theta^*), \Delta \right\rangle \geq \kappa\|\Delta\|_2^2 \\ &\Rightarrow \frac{1}{n}\|X\Delta\|_2^2 \geq \kappa\|\Delta\|_2^2 \,. \end{aligned} \tag{17}$$

We need the above equation to be satisfied $\forall \Delta \in E_r$ for the error bound in (16) to hold. To that end, we consider the following problem

$$\inf_{\Delta \in \text{cone}(E_r)} \frac{1}{n}\|X\Delta\|_2^2 \geq \kappa\|\Delta\|_2^2 \,. \tag{18}$$

Clearly if (18) is true then it is true for all $\Delta \in E_r$ since $E_r \subseteq \text{cone}(E_r)$. Additionally, since only the direction matters and not the magnitude we consider just the vectors on the spherical cap

$A = \text{cone}(E_r) \cap S^{p-1}$

$$\inf_{u \in A} \frac{1}{n} \|Xu\|_2^2 \geq \kappa \|u\|_2^2 \tag{19}$$

where $S^{p-1}$ is the unit sphere in $\mathbb{R}^p$. Since $\|u\|_2 = 1$ for all $u \in A$ we simply focus on

$$\inf_{u \in A} \frac{1}{n} \|Xu\|_2^2 \geq \kappa \tag{20}$$

which suffices in proving the RE condition for the restricted error set. We will show that when the design matrix has enough independent-isotropic sub-Gaussian rows that it can contain any number of dependent rows and still satisfy the RE condition.

Let $n$ be the number of independent-isotropic sub-Gaussian rows where each row satisfies $\|\|x_i\|\|_{\psi_2} \leq 1$ and $\mathbb{E}[x_i x_i^\top] = \mathbb{I}_{p \times p}$ and $m$ be the number of dependent rows selected via (11). Then denote $X \in \mathbb{R}^{(n+m) \times p}$ as the full design matrix, $X_n \in \mathbb{R}^{n \times p}$ as the design matrix with only independent-isotropic rows, and $X_m \in \mathbb{R}^{m \times p}$ as the design matrix with only dependent rows. Let $A := \text{cone}(E_r) \cap S^{p-1}$ be a spherical cap and assume $n \geq O(w^2(A))$.

In the sequel, we will prove the following main result.

**Theorem 1** *For $n \geq O(w^2(A))$ and $0 \leq m \leq T < \infty$, $X$ satisfies the following condition*

$$\inf_{u \in A} \frac{1}{n+m} \|Xu\|_2^2 \geq \kappa \tag{21}$$

*for $\kappa \leq \frac{n}{n+m}\left(1 - c\frac{w(A)}{\sqrt{n}}\right) + \frac{m}{n+m}\left(\lambda_{min}(\Sigma|A)\left(1 - c\frac{w(A)}{\sqrt{m}}\right) - \frac{4L\tau w(A)}{\sqrt{m}}\right)$ with probability at least $1 - 2\exp(-c_0 w^2(A)) - 2\exp(-c_0 w^2(A)) - L\exp\left(-\left(\frac{\epsilon}{L\tau\Delta(A)}\right)^2\right)$ for absolute constants $c_0, c, L, \epsilon > 0$ and where $\|\hat{\theta}_n - \theta^*\|_2 \leq \tau$ is the estimation error after the first $n$ rounds and $\Delta(A) = \sup_{u,v \in A} \|u - v\|_2$ is the diameter of the error set.*

To prove Theorem 1, we consider the term $\frac{1}{n+m}\|Xu\|_2^2$. We can decompose it as

$$\frac{1}{n+m}\|Xu\|_2^2 = \frac{1}{n+m}\|X_n u\|_2^2 + \frac{1}{n+m}\|X_m u\|_2^2 \ , \tag{22}$$

which implies

$$\inf_{u \in A} \frac{1}{n+m}\|Xu\|_2^2 = \inf_{u \in A} \frac{1}{n+m}\|X_n u\|_2^2 + \inf_{u \in A} \frac{1}{n+m}\|X_m u\|_2^2 \ . \tag{23}$$

Therefore to derive bounds on $\frac{1}{n+m}\|Xu\|_2^2$ we need to bound $\|X_n u\|_2^2$ and $\|X_m u\|_2^2$ which we consider in the next two sections.

### 5.2.1 Bound on $\|X_n u\|_2^2$

To show a bound on $\|X_n u\|_2^2$, we make use of the following theorem from [8] (Theorem 11)

**Theorem 2** *Let $X \in \mathbb{R}^{n\times p}$ be a design matrix with independent-isotropic sub-Gaussian rows, i.e., $\|\|x_i\|\|_{\psi_2} \leq 1$ and $\mathbb{E}[x_i x_i^\top] = \mathbb{I}_{p\times p}$. Then, for absolute constants $c_0, c > 0$ with probability at least $1 - 2\exp(-c_0 w^2(A))$, we have*

$$1 + c\frac{w(A)}{\sqrt{n}} \geq \sup_{u\in A} \frac{1}{n}\|Xu\|_2^2 \geq \inf_{u\in A} \frac{1}{n}\|Xu\|_2^2 \geq 1 - c\frac{w(A)}{\sqrt{n}} \ . \tag{24}$$

For $n = O(w^2(A))$ the following lemma follows from a straightforward application of Theorem 2.

**Lemma 1** *If $n = O(w^2(A))$, $c_0$ is an absolute constant then the following is true with probability atleast $1 - 2\exp(-c_0 w^2(A))$,*

$$\inf_{u\in A} \|X_n u\|_2^2 \geq n\kappa_n \tag{25}$$

*where $\kappa_n = 1 - c\frac{w(A)}{\sqrt{n}}$.*

### 5.2.2 Bounds on $\|X_m u\|_2$

Since $X_m$ is the design matrix with only dependent rows, Theorem 2 is inapplicable. The analysis requires showing bounds on a martingale difference sequence term and a term which involves the distribution of the rows of $X_m$ which change with time. Given such analysis, we obtain the following theorem.

**Theorem 3** *With probability at least $1-2\exp(-c_0 w^2(A)) - L\exp\left(-\left(\frac{\epsilon}{L\tau\Delta(A)}\right)^2\right)$, where $c_0, L, \epsilon > 0$ are constants and $\tau = \|\theta^* - \hat{\theta}\|_2$ is the error after the first $n$ rounds and $\Delta(A) = \sup_{u,v\in A}\|u-v\|_2$ is the diameter of the error set $A$*

$$\inf_{u\in A} \|X_m u\|_2^2 \geq m\kappa_m \tag{26}$$

*where $\kappa_m = \lambda_{min}(\Sigma|A)\left(1 - c\frac{w(A)}{\sqrt{m}}\right) - \frac{4L\tau w(A)}{\sqrt{m}}$.*

**Proof:** Proof of Theorem 3

To prove bounds on $\|X_m u\|_2^2$ we start with the following observation

$$\frac{1}{m}\|X_m u\|_2^2 = \frac{1}{m}\sum_{t=1}^{m}\langle x_t, u\rangle^2 \tag{27}$$

$$= \frac{1}{m}\sum_{t=1}^{m}\langle x_t - \mu_t, u\rangle^2 - \frac{1}{m}\sum_{t=1}^{m}\langle \mu_t, u\rangle^2 + \frac{2}{m}\sum_{t=1}^{m}\langle x_t, u\rangle\langle \mu_t, u\rangle \tag{28}$$

$$= \frac{1}{m}\sum_{t=1}^{m}\langle x_t - \mu_t, u\rangle^2 + \frac{2}{m}\sum_{t=1}^{m}\langle \mu_t, u\rangle\langle x_t - \mu_t, u\rangle + \frac{1}{m}\sum_{t=1}^{m}\langle \mu_t, u\rangle^2 \ , \tag{29}$$

where $\mu_t = \mathbb{E}[x_t|\mathcal{F}_{t-2}, \eta_{t-1}]$ is the expectation of $x_t$ given data from all previous rounds. By subtracting the mean of $x_t$, we are centering the variable and, in effect, constructing a martingale

difference sequence. Hence we get,

$$\inf_{u\in A}\frac{1}{m}\|X_m u\|_2^2 \geq \inf_{u\in A}\frac{1}{m}\sum_{t=1}^{m}\langle x_t-\mu_t,u\rangle^2 + \inf_{u\in A}\frac{2}{m}\sum_{t=1}^{m}\langle \mu_t,u\rangle\langle x_t-\mu_t,u\rangle + \frac{1}{m}\sum_{t=1}^{m}\langle \mu_t,u\rangle^2 \tag{30}$$

$$\geq \inf_{u\in A}\frac{1}{m}\sum_{t=1}^{m}\langle x_t-\mu_t,u\rangle^2 - \sup_{u\in A}\frac{2}{m}\sum_{t=1}^{m}\langle x_t-\mu_t,u\rangle \tag{31}$$

where the second inequality follows from $|\langle \mu_t,u\rangle| \leq 1$. To obtain the bounds we have to first bound the quantities $\frac{1}{m}\sum_{t=1}^{m}\langle x_t-\mu_t,u\rangle^2$ and $\frac{2}{m}\sum_{t=1}^{m}\langle x_t-\mu_t,u\rangle$. Note that by design after the first $n$ rounds, we can form the bound in (16). As such, when selecting $x_t$ via (14), each $x_t$ is being selected from $-\text{cone}(E_r)$ which is the restricted error set reflected about the origin. If $\|\hat{\theta}_n-\theta^*\|_2 = O\left(\psi(E_r)\frac{w(\Omega_R)}{\sqrt{n}}\right) \leq \tau$ is the error bound after the first $n$ rounds then after the first $n$ rounds, $\mathbb{E}[x_t-\mu_t|\mathcal{F}_{t-2},\eta_{t-1}] = 0$ since it is a MDS and $\|x_t-\mu_t\|_2 = O(\|\hat{\theta}_n-\theta^*\|_2) \leq \tau$.

**1. Bound for** $\sup_{u\in A}\frac{2}{m}\sum_{t=1}^{m}\langle x_t-\mu_t,u\rangle$

Since $x_t-\mu_t$ is a bounded vector-valued MDS which is bounded as $\|x_t-\mu_t\|_2 \leq \tau$ and $u\in A$ is a unit vector $\|u\|_2=1$ then the product $\langle x_t-\mu_t,u\rangle$ is a bounded, symmetric MDS with $|\langle x_t-\mu_t,u\rangle| \leq \|x_t-\mu_t\|_2\|u\|_2 \leq \tau\|u\|_2$. Therefore, by the Azuma-Hoeffding inequality we obtain

$$P\left(\left|\sum_{t=1}^{m}\langle x_t-\mu_t,u\rangle\right| \geq \gamma\sqrt{m}\right) \leq 2\exp\left(\frac{-\gamma^2}{2\tau^2\|u\|_2^2}\right) \tag{32}$$

$$\Rightarrow P\left(\frac{1}{\sqrt{m}}\left|\sum_{t=1}^{m}\langle x_t-\mu_t,u\rangle\right| \geq \gamma\right) \leq 2\exp\left(\frac{-\gamma^2}{2\tau^2\|u\|_2^2}\right)\ . \tag{33}$$

Therefore, for any $u,v\in A$

$$P\left(\frac{1}{\sqrt{m}}\left|\sum_{t=1}^{m}\langle x_t-\mu_t,u-v\rangle\right| \geq \gamma\right) \leq 2\exp\left(\frac{-\gamma^2}{2\tau^2\|u-v\|_2^2}\right)\ . \tag{34}$$

From (34) and using the generic chaining argument [28] following Theorem 5 it follows that for a constant $L$,

$$\mathbb{E}\left[\sup_{u\in A}\frac{1}{\sqrt{m}}\sum_{t=1}^{m}\langle x_t-\mu_t,u\rangle\right] \leq L\tau w(A)\ . \tag{35}$$

By a similar argument as stated in Theorem 8,

$$P\left(\sup_{u\in A}\left|\frac{1}{\sqrt{m}}\langle x_t-\mu_t,u\rangle\right| \geq 2L\tau w(A)+\epsilon\right) \leq L\exp\left(-\left(\frac{\epsilon}{L\tau\Delta(A)}\right)^2\right)\ , \tag{36}$$

where $\Delta(A) = \sup_{u,v\in A}\|u-v\|_2$ is the diameter of the error set $A$. Therefore, we get the following high-probability bound

$$P\left(\sup_{u\in A}\left|\frac{2}{m}\langle x_t-\mu_t,u\rangle\right| \leq \frac{4L\tau w(A)}{\sqrt{m}}+\epsilon\right) \geq 1-L\exp\left(-\left(\frac{\epsilon}{2L\tau\Delta(A)}\right)^2\right)\ . \tag{37}$$

**2. Bound on** $\inf_{u\in A}\frac{1}{m}\sum_{t=1}^{m}\langle x_t-\mu_t,u\rangle^2$

To prove a bound on $\inf_{u\in A}\frac{1}{m}\sum_{t=1}^{m}\langle x_t-\mu_t,u\rangle^2$ we use similar arguments as in [9] (Theorem 12) which we include below.

**Theorem 4** *Let $X$ be a design matrix with independent anisotropic sub-Gaussian rows, i.e., $\mathbb{E}[x_i^\top x_i]=\Sigma$ and $\|\|x_i\Sigma^{-1/2}\|\|_{\psi_2}\leq\kappa$. Then, for absolute constants $c_0,c>0$, with probability at least $(1-2\exp(-c_0w^2(A)))$, we have*

$$\sup_{u\in A}\left|\frac{1}{n}\frac{1}{u^\top\Sigma u}\|Xu\|_2^2-1\right|=\sup_{u\in A}\left|\frac{1}{n}\frac{1}{u^\top\Sigma u}\sum_{i=1}^{n}\langle x_i,u\rangle^2-1\right|\leq c\frac{w(A)}{\sqrt{n}}\ . \tag{38}$$

*Further,*

$$\lambda_{min}(\Sigma|A)\left(1-c\frac{w(A)}{\sqrt{n}}\right)\leq\inf_{u\in A}\frac{1}{n}\|Xu\|_2^2\leq\sup_{u\in A}\frac{1}{n}\|Xu\|_2^2\leq\lambda_{max}(\Sigma|A)\left(1+c\frac{w(A)}{\sqrt{n}}\right)\ , \tag{39}$$

*where $\lambda_{min}(\Sigma|A)=\inf_{u\in A}u^\top\Sigma u$ and $\lambda_{max}(\Sigma|A)=\sup_{u\in A}u^\top\Sigma u$ are the restricted minimum and maximum eigenvalues of $\Sigma$ restricted to $A\subseteq S^{p-1}$.*

The above theorem can be easily extended to consider $x_i$s which are MDS.

In our problem, each $x_t$ in the design matrix $X_m$ is chosen by solving (11) which is an optimization problem involving the confidence ellipsoid $C_{t-1}$. Each row $x_t$ in the design matrix $X_m$ will have a covariance matrix $\Sigma_t$ which depends on the ellipsoid $C_{t-1}$ such that $\|\|(x_t-\mu_t)\Sigma_t^{-1/2}\|\|_{\psi_2}\leq\|(x_t-\mu_t)\Sigma_t^{-1/2}\|_2\leq c_1$.

We will assume that there is a matrix $\Sigma$ for all $\Sigma_t$ such that the following bounds hold

$$\lambda_{min}(\Sigma|A)\leq\lambda_{min}(\Sigma_t|A_t)=\inf_{u\in A_t}u^T\Sigma_tu\quad\forall t, \tag{40}$$

$$\lambda_{max}(\Sigma|A)\geq\lambda_{max}(\Sigma_t|A_t)=\sup_{u\in A_t}u^T\Sigma_tu\quad\forall t. \tag{41}$$

Therefore, using Theorem 4, $w(A_{max})\geq w(A_t),\forall t$ and $w(A_{min})\leq w(A_t),\forall t$ with probability at least $1-2\exp(-c_0w^2(A))$

$$\inf_{u\in A}\frac{1}{m}\|X_mu\|_2^2\geq\lambda_{min}(\Sigma|A)\left(1-c\frac{w(A_{max})}{\sqrt{m}}\right)\ . \tag{42}$$

Combining 30, 37, and 42 with probability at least $1-2\exp(-c_0w^2(A))-L\exp\left(-\left(\frac{\epsilon}{L\tau\Delta(A)}\right)^2\right)$,

$$\inf_{u\in A}\frac{1}{m}\|X_mu\|_2^2\geq\lambda_{min}(\Sigma|A)\left(1-c\frac{w(A)}{\sqrt{m}}\right)-\frac{4L\tau w(A)}{\sqrt{m}}\ . \tag{43}$$

For $m=O(w^2(A)\lambda_{min}^2(\Sigma|A))$ this implies,

$$\inf_{u\in A}\|X_mu\|_2^2\geq m\kappa_m>0 \tag{44}$$

where $\kappa_m \lambda_{min}(\Sigma|A)\left(1 - c\frac{w(A)}{\sqrt{m}}\right) - \frac{4L\tau w(A)}{\sqrt{m}}$. Combining (23), (25), and (44) we get the following result with high probability,

$$\inf_{u \in A} \frac{1}{n+m}\|Xu\|_2^2 \geq \frac{n\kappa_n}{n+m} + \frac{m\kappa_m}{n+m} \geq \kappa > 0 \ , \tag{45}$$

for some constant $\kappa$ which completes the proof. ∎

### 5.3 Bound on Regularization Parameter $\lambda_t$

Recall the regularization parameter $\lambda_t$ needs to satisfy the inequality

$$\lambda_t \geq \rho R^*(\nabla \mathcal{L}(\theta^*; Z_t)) = \rho R^*\left(\frac{1}{t} X_t^\top (y_t - X_t\theta^*)\right) \tag{46}$$

for $\rho > 1$. Two issues of the right hand side are (1) the expression depends on the unknown parameter $\theta^*$ and (2) the expression is a random variable since it depends on $n$ independent-isotropic sub-Gaussian vectors and a sequence of random noise terms $\eta_1, \ldots, \eta_t$. We can remove the dependence on $\theta^*$ by observing that $y_t - X_t\theta^*$ is precisely the $t$-dimensional noise vector $\omega_t = [\eta_1 \ldots \eta_t]^\top$. Therefore,

$$R^*\left(\frac{1}{t} X_t^\top (y_t - X_t\theta^*)\right) = R^*\left(\frac{1}{t} X_t^\top \omega_t\right) \ . \tag{47}$$

We will show a high-probability upper bound on $R^*\left(\frac{1}{t} X_t^\top \omega_t\right)$ which holds simultaneously for all rounds $t = 1, \ldots, T$. Note, by the definition of the dual norm $R^*\left(\frac{1}{t} X_t^\top \omega_t\right) = \sup_{R(u) \leq 1} \frac{1}{t}\left\langle X_t^\top \omega_t, u\right\rangle$. The proof involves showing that $\frac{1}{t}\langle X_t^\top \omega_t, u\rangle$ is a martingale difference sequence (MDS) which concentrates as a sub-Gaussian random variable. Then, using a generic chaining argument, we show the supremum of such a quantity also concentrates as a sub-Gaussian random variable.

We begin by observing that

$$\frac{1}{t}\langle X_t^\top \omega_t, u\rangle = \frac{1}{\sqrt{t}}\frac{1}{\sqrt{t}}\langle X_t^\top \omega_t, u\rangle \ . \tag{48}$$

We will save one of the $\frac{1}{\sqrt{t}}$ terms for later and now proceed to show how $\frac{1}{\sqrt{t}}\left\langle X_t^\top \omega_t, u\right\rangle$ concentrates.

#### 5.3.1 $\frac{1}{\sqrt{t}}\left\langle X_t^\top \omega_t, u\right\rangle$ Concentrates as a Sub-Gaussian

First, let

$$\frac{1}{\sqrt{t}}\left\langle X_t^\top \omega_t, u\right\rangle = \|u\|_2 \frac{1}{\sqrt{t}}\left\langle X_t^\top \omega_t, \frac{u}{\|u\|_2}\right\rangle = \|u\|_2 \frac{1}{\sqrt{t}}\left\langle X_t^\top \omega_t, q\right\rangle \tag{49}$$

where $q = \frac{u}{\|u\|_2}$. We focus on the term $\frac{1}{\sqrt{t}}\left\langle X_t^\top \omega_t, q\right\rangle$. We can construct a martingale difference sequence (MDS) by observing that

$$\left\langle X_t^\top \omega_t, q\right\rangle = \langle \omega_t, X_t q\rangle = \sum_{\tau=1}^{t} \eta_\tau \langle x_\tau, q\rangle = \sum_{\tau=1}^{t} z_\tau \tag{50}$$

for $z_\tau = \eta_t \langle x_\tau, q \rangle$. Recall the filtration defined as

$$\mathcal{F}_t := \{x_1, \ldots, x_{t+1}, \eta_1, \ldots, \eta_t\} \ . \tag{51}$$

Each $z_\tau$ can be seen as a MDS since

$$\mathbb{E}[z_\tau | \mathcal{F}_{\tau-1}] = \mathbb{E}[\eta_\tau \langle x_\tau, q \rangle | \mathcal{F}_{\tau-1}] = \langle x_\tau, q \rangle \cdot \mathbb{E}[\eta_\tau | \mathcal{F}_{\tau-1}] = 0 \tag{52}$$

because $x_\tau$ is $\mathcal{F}_{\tau-1}$ measurable and $\eta_\tau$ is $\mathcal{F}_\tau$ measurable. Additionally, each $z_\tau$ follows a sub-Gaussian distribution with parameter $B$ because we assume $|\eta_\tau| \leq B$, $\|x_\tau\|_2 \leq 1$, and $\|q\|_2 = 1$ therefore, their product is bounded as $|\eta_\tau \langle x_\tau, q \rangle| \leq B$ and thus from Lemma 10 and Definition 2 it is sub-Gaussian with $\| \eta_\tau \langle x_\tau, q \rangle \|_{\psi_2} \leq B$. Since each $z_\tau$ is a bounded MDS, we can use the Azuma-Hoeffding inequality to show that the sum $\sum_{\tau=1}^t z_\tau$ concentrates as a sub-Gaussian with parameter $B$. For all $\gamma \geq 0$

$$\begin{aligned} P\left( \left| \sum_{\tau=1}^t z_\tau \right| \geq \gamma \right) &= P\left( \left| \langle X_t^\top \omega_t, q \rangle \right| \geq \gamma \right) \\ &= P\left( \left| \left\langle X_t^\top \omega_t, \frac{u}{\|u\|_2} \right\rangle \right| \geq \gamma \right) \leq 2 \exp\left( \frac{-\gamma^2}{2tB^2} \right) \\ &= P\left( \frac{1}{\sqrt{t}} \left| \left\langle X_t^\top \omega_t, \frac{u}{\|u\|_2} \right\rangle \right| \geq \zeta \right) \leq 2 \exp\left( \frac{-\zeta^2}{2B^2} \right) \end{aligned} \tag{53}$$

where $\zeta = \gamma / \sqrt{t}$ which implies $\gamma = \sqrt{t}\zeta$. From (53) and (93) in Definition 1 we can see that the term $\frac{1}{\sqrt{t}} \left\langle X_t^\top \omega_t, \frac{u}{\|u\|_2} \right\rangle$ concentrations as a sub-Gaussian with $\| \langle X_t^\top \omega_t, \frac{u}{\|u\|_2} \rangle \|_{\psi_2} \leq B$.

Next, we show that the term $\frac{1}{\sqrt{t}} \left\langle X_t^\top \omega_t, u \right\rangle$ also concentrations as a sub-Gaussian with $\| \langle X_t^\top \omega_t, u \rangle \|_{\psi_2} \leq \|u\|_2 B$ using (53) as

$$\begin{aligned} & P\left( \frac{1}{\sqrt{t}} \left| \left\langle X_t^\top \omega_t, \frac{u}{\|u\|_2} \right\rangle \right| \geq \zeta \right) \\ =& P\left( \|u\|_2 \frac{1}{\sqrt{t}} \left| \left\langle X_t^\top \omega_t, \frac{u}{\|u\|_2} \right\rangle \right| \geq \|u\|_2 \zeta \right) \\ =& P\left( \frac{1}{\sqrt{t}} \left| \left\langle X_t^\top \omega_t, u \right\rangle \right| \geq \epsilon \right) \leq 2 \exp\left( \frac{-\epsilon^2}{2\|u\|_2^2 B^2} \right) \end{aligned} \tag{54}$$

where $\epsilon = \|u\|_2 \zeta$ which implies $\zeta = \epsilon / \|u\|_2$. The reason we went through showing the above is because the generic chaining argument we will invoke to bound $\sup_{R(u) \leq 1} \frac{1}{\sqrt{t}} \left\langle X_t^\top \omega_t, u \right\rangle$ requires that $\frac{1}{\sqrt{t}} \left\langle X_t^\top \omega_t, u \right\rangle$ is a sub-Gaussian random variable.

### 5.3.2 Bound on $\sup_{R(u) \leq 1} \frac{1}{\sqrt{t}} \langle X_t^\top \omega_t, u \rangle$ via Generic Chaining

We obtain a high-probability bound on $\sup_{R(u) \leq 1} \frac{1}{\sqrt{t}} \langle X_t^\top \omega_t, u \rangle$ using a generic chaining argument from [27, 28]. This involves (1) showing that the absolute difference of two sub-Gaussian processes concentrates as a sub-Gaussian, (2) showing the expectation over the supremum of the absolute difference of two sub-Gaussian processes is upper bounded by the sub-Gaussian width of a set

from which the processes are indexed from, and (3) showing the supremum of a sub-Gaussian process is concentrated around its expectation and therefore, around the sub-Gaussian width with high-probability.

**(1) Sub-Gaussian Process Concentration**

First, we show that the absolute difference of two sub-Gaussian processes concentrates as a sub-Gaussian. Let $Y_u = \frac{1}{\sqrt{t}}\langle X_t^\top \omega_t, u\rangle$ indexed by $u \in \Omega_R$ and $Y_v = \frac{1}{\sqrt{t}}\langle X_t^\top \omega_t, v\rangle$ indexed by $v \in \Omega_R$ be two zero-mean (since they are both a MDS sum), random symmetric processes (since $(Y_u)_{u\in\Omega_r}$ has the same law as $(-Y_u)_{u\in\Omega_R}$ via (53) and $\omega_t$ is symmetric and similarly for $Y_v$). Then by construction

$$|Y_u - Y_v| = \frac{1}{\sqrt{t}}\left|\left\langle X_t^\top \omega_t, u - v\right\rangle\right| .$$

Using the bound we established in (54), we obtain the following bound on the absolute difference of two sub-Gaussian random processes $Y_u$ and $Y_v$ as

$$P\left(\frac{1}{\sqrt{t}}\left|\left\langle X_t^\top \omega_t, u - v\right\rangle\right| \geq \epsilon\right) \leq 2\exp\left(\frac{-\epsilon^2}{2\|u-v\|_2^2 B^2}\right) \tag{55}$$

which shows $|Y_u - Y_v|$ concentrates as a sub-Gaussian random variable with $\|\|Y_u - Y_v\|\|_{\psi_2} = \|u-v\|_2 B$.

**(2) Bound on $\mathbb{E}\left[\sup_{R(u)\leq 1}\frac{1}{t}\langle X_t^\top \omega_t, u\rangle\right]$**

In order to establish a high-probability bound on $\sup_{R(u)\leq 1}\frac{1}{t}\langle X_t^\top \omega_t, u\rangle$ we need to prove a bound on $\mathbb{E}\left[\sup_{R(u)\leq 1}\frac{1}{t}\langle X_t^\top \omega_t, u\rangle\right]$. To prove such a bound, we will apply a generic chaining argument for upper bounds on such sub-Gaussian processes. For the generic chaining argument, we will need the result in (55) and the following theorem.

**Theorem 5** ***(Talagrand [27], Theorem 2.1.5)*** *Consider two processes $(Y_u)_{u\in\Omega_R}$ and $(X_u)_{u\in\Omega_R}$ indexed by the same set. Assume that the process $(X_u)_{u\in\Omega_R}$ is Gaussian and that the process $(Y_u)_{u\in\Omega_R}$ satisfies the condition*

$$\forall \epsilon > 0, \forall u, v \in \Omega_R, P(|Y_u - Y_v| \geq \epsilon) \leq 2\exp\left(-\frac{\epsilon^2}{d(u,v)^2}\right) \tag{56}$$

*where $d(u,v)$ is a distance function which we assume is $d(u,v) = \|u-v\|_2$ for the set $\Omega_R$. Then we have*

$$\mathbb{E}\left[\sup_{u,v\in\Omega_R}|Y_u - Y_v|\right] \leq L\mathbb{E}\left[\sup_{v\in\Omega_R} X_v\right] \tag{57}$$

*where $L$ is an absolute constant.*

First, notice that $\mathbb{E}\left[\sup_{v\in\Omega_R} X_v\right]$ is exactly the Gaussian width $w(\Omega_R)$ of the set $\Omega_R$ as seen by Definition 3. For our purposes, we make one modification to the above theorem similar to [9] (Theorem 8). In (55), we see that $|Y_u - Y_v|$ concentrates as a sub-Gaussian with parameter $B$. To bound the expectation of two sub-Gaussian processes, we scale the Gaussian width by the sub-Gaussian parameter B to get

$$\mathbb{E}\left[\sup_{u,v\in\Omega_R}|Y_u - Y_v|\right] \leq LB\mathbb{E}\left[\sup_{v\in\Omega_R} X_v\right] = LBw(\Omega_R)\ . \tag{58}$$

This shows for two sub-Gaussian processes $Y_u$ and $Y_v$, the expectation of the supremum of their absolute difference is upper bounded by the Gaussian width scaled by the sub-Gaussian norm, i.e., the sub-Gaussian width.

The second result we need is the following lemma.

**Lemma 2** ***(Talagrand [27], Lemma 1.2.8)*** *If the process $(Y_u)_{u\in\Omega_R}$ is symmetric then*

$$\mathbb{E}\left[\sup_{u,v\in\Omega_R}|Y_u - Y_v|\right] = 2\mathbb{E}\left[\sup_{u\in\Omega_R} Y_u\right]\ . \tag{59}$$

We know from above that our processes $Y_u = \frac{1}{\sqrt{t}}\langle X_t^\top \omega_t, u\rangle$ and $Y_v = \frac{1}{\sqrt{t}}\langle X_t^\top \omega_t, v\rangle$ are symmetric. As such we get the following lemma.

**Lemma 3** *From (55) we can see that the condition of Theorem 5 is satisfied in the sub-Gaussian case so using Theorem 5 and Lemma 2 for some absolute constant $L$ we obtain*

$$\mathbb{E}\left[\sup_{u\in\Omega_R}\frac{1}{t}\left|\left\langle X_t^\top \omega_t, u\right\rangle\right|\right] \leq LB\frac{w(\Omega_R)}{\sqrt{t}}\ . \tag{60}$$

**Proof:** Proof of Lemma 3.

$$\begin{aligned}\mathbb{E}\left[\sup_{u,v\in\Omega_R}|Y_u - Y_v|\right] &= 2\mathbb{E}\left[\sup_{u\in\Omega_R}|Y_u|\right] \\ &= 2\mathbb{E}\left[\sup_{u\in\Omega_R}\frac{1}{\sqrt{t}}\left|\left\langle X_t^\top \omega_t, u\right\rangle\right|\right] \\ &\leq LBw(\Omega_R)\ . \end{aligned}\tag{61}$$

Therefore,

$$\mathbb{E}\left[\sup_{u\in\Omega_R}\frac{1}{t}\left|\langle X_t^\top \omega_t, u\rangle\right|\right] = \frac{1}{\sqrt{t}}\mathbb{E}\left[\sup_{u\in\Omega_R}\frac{1}{\sqrt{t}}\left|\langle X_t^\top \omega_t, u\rangle\right|\right] \tag{62}$$

$$\leq LB\frac{w(\Omega_R)}{\sqrt{t}}\ . \tag{63}$$

■

**(3) Concentration of** $\sup_{R(u)\leq 1}\frac{1}{\sqrt{t}}\left\langle X_t^\top \omega_t, u\right\rangle$

To complete the argument, we need the following theorem.

**Theorem 6** ***(Talagrand [28], Theorem 2.2.27)*** *If the process $(Y_u)$ satisfies (56) or similarly (55) for the sub-Gaussian case then for $\epsilon > 0$ one has*

$$P\left(\sup_{u,v\in\Omega_R}|Y_u - Y_v| \geq L\big(\gamma_2(\Omega_R, d(u,v)) + \epsilon\Delta(\Omega_R)\big)\right) \leq L\exp(-\epsilon^2)\ . \tag{64}$$

we get $\Delta(\Omega_R) = \sup_{u\in\Omega_R}\|2u\|_2 = \sup_{v\in\Omega_R}\|-2v\|_2$. Additionally, we can simplify Theorem 6 by using the following theorem.

**Theorem 7** ***(Talagrand [28], Theorem 2.4.1)*** *For some universal constant $L$ we have*

$$\frac{1}{L}\gamma_2(\Omega_R, d(u,v)) \leq \mathbb{E}\left[\sup_{u\in\Omega_R} Y_u\right] \leq L\gamma_2(\Omega_R, d(u,v))\ . \tag{65}$$

Combining Theorem 6 with Theorem 7, using Lemma 2, (58), and our definitions of $Y_u$ and $Y_v$ for any $\epsilon > 0$ we get

**Theorem 8**

$$P\left(\sup_{R(u)\leq 1}\frac{1}{\sqrt{t}}\left|\left\langle X_t^\top\omega_t, u\right\rangle\right| \geq 2cLB(1+\epsilon)w(\Omega_R)\right) \leq L\exp\left(-\left(\frac{\epsilon}{2cLBw(\Omega_R)}\right)^2\right)\ . \tag{66}$$

**Proof:** Proof of Theorem 8.

$$\begin{aligned}
&P\left(\sup_{u,v\in\Omega_R}|Y_u - Y_v| \geq L\big(\gamma_2(\Omega_R, d(u,v)) + \zeta\Delta(\Omega_R)\big)\right)\\
&= P\left(\sup_{u,v\in\Omega_R}|Y_u - Y_v| \geq L\gamma_2(\Omega_R, d(u,v)) + \epsilon\right)\\
&\leq P\left(\sup_{u,v\in\Omega_R}|Y_u - Y_v| \geq \mathbb{E}\left[\sup_{u,v\in\Omega_R}|Y_u - Y_v|\right] + \epsilon\right)\\
&= P\left(\sup_{u\in\Omega_R}|Y_u| \geq 2\mathbb{E}\left[\sup_{u\in\Omega_R}|Y_u|\right] + \epsilon\right)\\
&= P\left(\sup_{R(u)\leq 1}\frac{1}{\sqrt{t}}\left|\left\langle X_t^\top\omega_t, u\right\rangle\right| \geq 2LBw(\Omega_R) + \epsilon\right) \leq L\exp\left(-\left(\frac{\epsilon}{LB\Delta(\Omega_R)}\right)^2\right)\ .
\end{aligned}$$

■

Putting everything together, we get the following main theorem.

**Theorem 9** *Let $X_t = [x_1 \dots x_t]^\top$ be a design matrix where each row satisfies $\|\|x_i\|\|_{\psi_2} \leq 1$, $\omega_t = [\eta_1 \dots \eta_t]$ be a noise vector where each $|\eta_i| \leq B$, $\Omega_R = \{u : R(u) \leq 1\}$ be the unit norm ball of $R(\cdot)$, and define $\phi = \sup_{R(u)\leq 1}\|2u\|_2$ then for any $\alpha > 0$*

$$P\left(R^*\left(\frac{1}{t}X_t^\top\omega_t\right) \geq (1+\alpha)2LB\frac{w(\Omega_R)}{\sqrt{t}}\right) \leq L\exp\left(-\left(\frac{\alpha w(\Omega_R)}{\phi}\right)^2\right)\ . \tag{67}$$

*where $L$ is a universal constant.*

**Proof:** Proof of Theorem 9

$$P\left(R^*\left(\frac{1}{\sqrt{t}}X_t^\top\omega_t\right) \geq 2LBw(\Omega_R)+\epsilon\right) \leq L\exp\left(-\left(\frac{\epsilon}{LB\phi}\right)^2\right) \tag{68}$$

$$P\left(R^*\left(\frac{1}{t}X_t^\top\omega_t\right) \geq 2LB\frac{w(\Omega_R)}{\sqrt{t}}+\gamma\right) \leq L\exp\left(-\left(\frac{\sqrt{t}\gamma}{LB\phi}\right)^2\right) \tag{69}$$

$$P\left(R^*\left(\frac{1}{t}X_t^\top\omega_t\right) \geq 2LB\frac{w(\Omega_R)}{\sqrt{t}}+\alpha 2LB\frac{w(\Omega_R)}{\sqrt{t}}\right) \leq L\exp\left(-\left(\frac{\sqrt{t}\alpha 2LB\frac{w(\Omega_R)}{\sqrt{t}}}{LB\phi}\right)^2\right) \tag{70}$$

$$P\left(R^*\left(\frac{1}{t}X_t^\top\omega_t\right) \geq (1+\alpha)2LB\frac{w(\Omega_R)}{\sqrt{t}}\right) \leq L\exp\left(-\left(\frac{\alpha 2w(\Omega_R)}{\phi}\right)^2\right)\ . \tag{71}$$

where the first inequality is from Theorem 8, the second inequality is from multiplying both sides by $\frac{1}{\sqrt{t}}$ and setting $\gamma=\frac{\epsilon}{\sqrt{t}}$, and the third inequality is from setting $\gamma=\alpha 2LB\frac{w(\Omega_R)}{\sqrt{t}}$. ■

### 5.3.3 High-Probability Bound on $R^*\left(\frac{1}{T}X_T^\top\omega_T\right)$ for all $T$

Theorem 9 gives a high-probability bound on the value of $R^*\left(X_t^\top\omega_t\right)$ for round $t$ but we need a bound which holds simultaneously for all rounds $T$ with high-probability.

**Theorem 10** *Using Theorem 9, for all $T$ with $\alpha^2=2\log T\left(\frac{\phi}{2w(\Omega_R)}\right)^2$*

$$P\left(R^*\left(\frac{1}{t}X_t^\top\omega_t\right) \leq (1+\alpha)2LB\frac{w(\Omega_R)}{\sqrt{t}}\right) \geq 1-\frac{L}{T}\ . \tag{72}$$

**Proof:** Proof of Theorem 10

From Theorem 9, for any round $t$ we have the bound

$$P\left(R^*\left(\frac{1}{t}X_t^\top\omega_t\right) \geq (1+\alpha)2LB\frac{w(\Omega_R)}{\sqrt{t}}\right) \leq L\exp\left(-\left(\frac{\alpha 2w(\Omega_R)}{\phi}\right)^2\right)\ . \tag{73}$$

We desire a bound on the probability that holds simultaneously for all $t=1,\ldots,T$. We can obtain such a bound by setting $\alpha^2=2\log T\left(\frac{\phi}{2w(\Omega_R)}\right)^2$ and applying a union bound for all $t$

$$\bigcup_{t=1}^{T} P\left(R^*\left(\frac{1}{t}X_t^\top\omega_t\right) \geq (1+\alpha)2LB\frac{w(\Omega_R)}{\sqrt{t}}\right) \leq \sum_{t=1}^{T} L\exp\left(-\left(\frac{\alpha 2w(\Omega_R)}{\phi}\right)^2\right) \tag{74}$$

$$= L\sum_{t=1}^{T}\exp\left(-2\log T\right) \tag{75}$$

$$= L\sum_{t=1}^{T}\frac{1}{T^2} \tag{76}$$

$$= \frac{L}{T}\ . \tag{77}$$

■

### 5.3.4 Setting the Value of $\lambda_t$

Now, recall from (46) ultimately we need $\lambda_t \geq \rho R^* \left(\frac{1}{t} X_t^\top \omega_t\right)$ for $a > 1$. From Theorem 10, we can set $\lambda_t$ to be

$$\lambda_t \geq 2LB(1+\alpha)\frac{w(\Omega_R)}{\sqrt{t}} \tag{78}$$

with $\alpha = \sqrt{2\log T}\frac{\phi}{2w(\Omega_R)}$.

## 5.4 Estimation Error Bound with High-Probability

From Section 5.2 and Section 5.3 we showed high-probability bounds on the values of $\kappa$ and $\lambda_t$ which were required for the bound

$$\|\hat{\theta}_t - \theta^*\|_{2,D_t} \leq c\psi(E_r)\frac{\lambda_t}{\kappa}\sqrt{t}\ . \tag{79}$$

for $\kappa > 0$ and $\lambda_t \geq 2LB(1+\alpha)\frac{w(\Omega_R)}{\sqrt{t}}$. Let $c = \frac{2LB}{\kappa}$ and $\alpha = \sqrt{2\log T}\frac{\phi}{2w(\Omega_R)}$, then

$$\beta_t = c(1+\alpha)\psi(E_r)w(\Omega_R)\ . \tag{80}$$

If we construct the confidence ball as $C_t := \{\theta \in \mathbb{R}^p : \|\theta - \hat{\theta}_t\|_{2,D_t} \leq \beta_t\}$, then $\theta^* \in C_t$ with probability $1 - 2\exp(-c_0 w^2(A)) - 2\exp(-c_0 w^2(A)) - L\exp\left(-\left(\frac{\epsilon}{L\tau\Delta(A)}\right)^2\right) - \frac{L}{T}$.

## 5.5 Regret Analysis

The following analysis is conditioned on $\theta^* \in C_t$ for all $t$ which occurs with high-probability as shown in Section 5.4. Let the optimal arm to pull be defined as $x^* := \text{argmin}_{x\in\mathcal{X}} \langle x, \theta^*\rangle = -\frac{\theta^*}{\|\theta^*\|_2}$. The main result we will prove is the following theorem.

**Theorem 11** *In round $t$, let $\beta_t = \frac{2LB}{\kappa}(1+\alpha)\psi(E_r)w(\Omega_R)$, $|\eta_t| \leq B$, $\alpha = \sqrt{2\log T}\frac{\phi}{2w(\Omega_R)}$, and $\phi = \sup_{R(u)\leq 1}\|2u\|_2$. For all sufficiently large $T$, the fixed cumulative regret of Algorithm 1 is*

$$R_T \leq \widetilde{O}\left(\psi(E_r)w(\Omega_R)\sqrt{p}\sqrt{T}\right) \tag{81}$$

*with high-probability.*

To prove the theorem, we first show an upper bound on the instantaneous and cumulative regret for the initial $n$ estimation rounds and then separately for remaining $T-n$ exploration and exploitation rounds. Then, we show an upper bound on the fixed cumulative regret for Algorithm 1.

### 5.5.1 Estimation Rounds

The following lemma shows the instantaneous and cumulative regret for the estimation rounds.

**Lemma 4** *The algorithm initially selects $n = O(w^2(A))$ independent-isotropic sub-Gaussian vectors to satisfy RSC/RE and computes the initial estimate. Each of the $n = O(w^2(A))$ vectors incurs at most $2\|\theta^*\|_2$ instantaneous regret which gives $2\|\theta^*\|_2 O(w^2(A)))$ cumulative regret.*

**Proof:** Proof of Lemma 4. The instantaneous regret for each estimation round is

$$\begin{aligned}
&\langle x_t, \theta^* \rangle - \langle x^*, \theta^* \rangle \\
&\leq \left\langle \frac{\theta^*}{\|\theta^*\|_2}, \theta^* \right\rangle - \left\langle -\frac{\theta^*}{\|\theta^*\|_2}, \theta^* \right\rangle \\
&= 2\|\theta^*\|_2 \ .
\end{aligned}$$

The cumulative regret for all estimation rounds is $R_n = 2\|\theta^*\|_2 O(w^2(A))$. ∎

### 5.5.2 Exploration and Exploitation Rounds

For ease of exposition, we start the notation counter at 1 for the rounds after the initial $n$ estimation rounds and end at $T$ rather than $n+1$ and end at $T-n$. For the rest of the analysis, we will rely on some results established in [21] which we repeat here for completeness. Note, our $\beta_t$ has the square root included whereas [21] does not.

**Theorem 12** ***(Sum of Squares Regret Bound Dani et al. [21], Theorem 6)***
*Let $r_t = \langle x_t, \theta^* \rangle - \langle x^*, \theta^* \rangle$ denote the instantaneous regret acquired by the algorithm on round $t$. For Algorithm 1, if $\theta^* \in C_{t-1}$ for all $t \leq T$ then*

$$\sum_{t=1}^{T} r_t^2 \leq 8p\beta_T^2 \log T \ . \tag{82}$$

A key insight from [21] is that on any round $t$ where $\theta^* \in C_{t-1}$, the instantaneous regret is at most the width of the ellipsoid in the direction of $x_t$ as shown in the following lemmas. In addition, the algorithm's choice of decisions forces the ellipsoids to shrink at a rate such that the sum of the squares of the widths is small.

**Lemma 5** ***(Dani et al. [21], Lemma 7)***
*For Algorithm 1, if $\theta \in C_{t-1}$ and $x \in \mathcal{X}$, then*

$$\left| (\theta - \hat{\theta}_t)^\top x \right| \leq \beta_t \sqrt{x^\top D_t^{-1} x} \ . \tag{83}$$

Define

$$w_t := \sqrt{x_t^\top D_t^{-1} x_t} \tag{84}$$

which is interpreted as the normalized width at time $t$ in the direction of the selected decision $x_t$. The true width $2\beta_t w_t$ is an upper bound for the instantaneous regret.

**Proof:** Proof of Lemma 5.

$$\begin{aligned}
\left|(\theta-\hat{\theta}_t)^\top x\right| &= \left|(\theta-\hat{\theta}_t)^\top D_t^{1/2} D_t^{-1/2} x\right| \\
&= \left|\left(D_t^{1/2}(\theta-\hat{\theta}_t)\right)^\top D_t^{-1/2} x\right| \\
&\le \left\|D_t^{1/2}(\theta-\hat{\theta}_t)\right\|_2 \left\|D_t^{-1/2} x\right\|_2 \quad \text{(by Cauchy-Schwarz)} \\
&= \left\|D_t^{1/2}(\theta-\hat{\theta}_t)\right\|_2 \sqrt{x^\top D_t^{-1} x} \\
&\le \beta_t \sqrt{x^\top D_t^{-1} x} \qquad (85)
\end{aligned}$$

where the last inequality holds since $\theta \in C_{t-1}$. ■

**Lemma 6** ***(Dani et al. [21], Lemma 8)***
*For Algorithm 1, if $\theta^* \in C_{t-1}$, then*

$$r_t \le 2\min(\beta_t w_t, 1)\ . \qquad (86)$$

**Proof:** Proof of Lemma 6.
Let $\widetilde{\theta}_t \in C_{t-1}$ denote the vector which minimizes the dot product $\langle \widetilde{\theta}_t, x_t\rangle$. By choice of $x_t$, we have

$$\langle \widetilde{\theta}_t, x_t\rangle = \min_{\theta\in C_{t-1}, x\in\mathcal{X}} \langle x, \theta\rangle \le \langle x^*, \theta^*\rangle \qquad (87)$$

where the inequality used the hypothesis $\theta^* \in C_{t-1}$. Hence,

$$\begin{aligned}
r_t &= \langle x_t, \theta^*\rangle - \langle x^*, \theta^*\rangle \\
&\le \langle x_t, \theta^* - \widetilde{\theta}_t\rangle \\
&= \langle x_t, \theta^* - \hat{\theta}_t\rangle + \langle x_t, \hat{\theta}_t - \widetilde{\theta}_t\rangle \\
&\le 2\beta_t w_t
\end{aligned}$$

where the last step follows from Lemma 5 ■

Next we show that the sum of the squares of the widths does not grow too fast.

**Lemma 7** ***(Dani et al. [21], Lemma 9)***
*We have for all $t$*

$$\sum_{\tau=1}^{t} \min(w_\tau^2, 1) \le 2p\log t\ . \qquad (88)$$

*The following lemmas are used for the proof.*

**Lemma 8** ***(Dani et al. [21], Lemma 10)***
*For every $t \le T$*

$$\det D_{t+1} = \prod_{\tau=1}^{t} (1 + w_t^2)\ . \qquad (89)$$

**Proof:** Proof of Lemma 8.
By the definition of $D_{t+1}$, we have

$$\begin{aligned}\det D_{t+1} &= \det(D_t + x_t x_t^\top) \\ &= \det\left(D_t^{1/2}\left(I + D_t^{-1/2} x_t x_t^\top D_t^{-1/2}\right) D_t^{1/2}\right) \\ &= \det(D_t)\det\left(I + D_t^{-1/2} x_t \left(D_t^{-1/2} x_t\right)^\top\right) \\ &= \det(D_t)\det\left(I + v_t v_t^\top\right)\end{aligned}$$

where $v_t = D_t^{-1/2} x_t$. Now observe that $v_t^\top v_t = w_t^2$ and

$$(I + v_t v_t^\top) v_t = v_t + v_t(v_t^\top v_t) = (1 + w_t^2) v_t \ . \tag{90}$$

Hence $(1+w_t^2)$ is an eigenvalue of $I+v_t v_t^\top$. Since $v_t v_t^\top$ is a rank one matrix, all the other eigenvalues of $I + v_t v_t^\top$ equal 1. It follows that $\det(I + v_t v_t^\top)$ is $(1 + w_t^2)$, and so

$$\det D_{t+1} = (1 + w_t^2)\det D_t \ . \tag{91}$$

Since we are constructing $D_1$ after $w^2(A)$ random samples and we assume there is a matrix $\Sigma$ such that the following bounds hold

$$\begin{aligned}0 < \lambda_{min}(\Sigma|A) \leq \lambda_{min}(\Sigma_t|A_t) &= \inf_{u \in A_t} u^\top \Sigma_t u \quad \forall t, \\ \infty > \lambda_{max}(\Sigma|A) \geq \lambda_{max}(\Sigma_t|A_t) &= \sup_{u \in A_t} u^\top \Sigma_t u \quad \forall t,\end{aligned}$$

then $\det D_1$ is non-zero and the result follows by induction. ∎

**Lemma 9** ***(Dani et al. [21], Lemma 11)***
*For all $t$,* $\det D_t \leq t^p$.

**Proof:** Proof of Lemma 9
The rank one matrix $x_t x_t^\top$ has $x_t^\top x_t = \|x_t\|_2^2$ as its unique non-zero eigenvalue. Since we have sampled for $w^2(A)$ rounds it follows that

$$\begin{aligned}\operatorname{trace} D_t &\leq \operatorname{trace}\left(I + \sum_{\tau<t} x_t x_t^\top\right) \\ &= p + \sum_{\tau<t} \operatorname{trace}(x_t x_t^\top) \\ &= p + \sum_{\tau<t} \|x_\tau\|_2^2 \\ &\leq pt \ .\end{aligned}$$

Now, recall that trace $D_t$ equals the sum of the eigenvalues of $D_t$. On the other hand $\det(D_t)$ equals the product of the eigenvalues. Since $A_t$ is positive definite, its eigenvalues are all positive. Subject

to these constraints, $\det(D_t)$ is maximized when all the eigenvalues are equal and the desired bound follows. ∎

**Proof:** Proof of Lemma 7

Using the fact that for $0 \leq y \leq 1$, $\log(1+y) \geq y/2$ we have

$$\begin{aligned}\sum_{\tau=1}^{t} \min(w_t^2, 1) &\leq \sum_{\tau=1}^{t} 2\log(1+w_\tau^2) \\ &= 2\log(\det D_{t+1}) \\ &\leq 2p\log t\end{aligned}$$

by Lemmas 8 and 9. ∎

We can now prove Theorem 12.

**Proof:** Proof of Theorem 12.

Assume that $\theta^* \in C_{t-1}$ for all $t$. Then

$$\begin{aligned}\sum_{t=1}^{T} r_t^2 &\leq \sum_{t=1}^{T} 4\beta_t^2 \min(w_t^2, 1) \quad \text{(by Lemma 6)} \\ &\leq 4\beta_T^2 \sum_{t=1}^{T} \min(w_t^2, 1) \quad (\text{since } \beta_1^2 = \beta_2^2 = \cdots = \beta_T^2) \\ &\leq 8\beta_T^2 p \log T \quad \text{(by Lemma 7)} .\end{aligned}$$

∎

Given such results from [21], we now proof Theorem 11.

**Proof:** Proof of Theorem 11.

From the RSC/RE and $\lambda_t$ analysis and Theorem 12, we know that with probability $1 - 2\exp(-c_0 w^2(A)) - 2\exp(-c_0 w^2(A)) - L\exp\left(-\left(\frac{\epsilon}{L\tau\Delta(A)}\right)^2\right) - \frac{L}{T}$ where the cumulative regret is a sum of the initial estimation rounds $R_n$ and the exploration and exploitation rounds as

$$\begin{aligned}R_T = R_n + \sum_{t=1}^{T} r_t &\leq 2\|\theta^*\|_2 O(w^2(A)) + \left(T\sum_{t=1}^{T} r_t^2\right)^{1/2} \\ &\leq 2\|\theta^*\|_2 O(w^2(A)) + \beta_t\sqrt{8pT\log T} .\end{aligned}$$

Setting $\beta_t = \frac{2LB}{\kappa}(1+\alpha)\psi(E_r)w(\Omega_R)$ and $\alpha = \sqrt{2\log T}\frac{\phi}{2w(\Omega_R)}$ completes the proof.

We highlight the regret bounds of a few common types of structure which follows easily from Theorem 11 and the values of $\psi(E_r)$ and $w(A)$ from Banerjee et al. [8].

**Corollary 1 (Unstructured $\theta^*$)** *Plugging in $\psi(E_r) = O(1)$ and $w(\Omega_R) = O(\sqrt{p})$ gives the cumulative regret of an unstructured $\theta^*$ with high-probability as $\widetilde{O}\left(p\sqrt{T}\right)$ which matches the regret from Dani et al. [21] up to log and constant factors.*

**Corollary 2 ($s$-sparse $\theta^*$)** *Plugging in $\psi(E_r) = O(\sqrt{s})$ and $w(\Omega_R) = O(\sqrt{\log p})$ gives the cumulative regret of an $s$-sparse $\theta^*$ with high-probability as $\widetilde{O}\left(\sqrt{s \log p}\sqrt{p}\sqrt{T}\right)$ which matches the regret from Abbasi-Yadkori et al. [2] up to log and constant factors. The regret from Carpentier and Munos [16] of $\widetilde{O}(s \log p\sqrt{T})$ beats our rate however, they consider a different noise model, and make more simplifying assumptions including i.i.d. noise.*

**Corollary 3 (Group-sparse $\theta^*$)** *Let $K$ be the number of groups $\{\mathcal{G}_1, \ldots, \mathcal{G}_K\}$ where each $\mathcal{G}_i \subset \{1, \ldots, p\}$ are partitioned into disjoint groups with $m = \max_i |\mathcal{G}_i|$ and for a given subset $S_{\mathcal{G}} \subset \{1, \ldots, K\}$. Then plugging in $\psi(E_r) = O(\sqrt{S_{\mathcal{G}}})$ and $w(\Omega_R) = O(\sqrt{m + \log K})$ gives the cumulative regret of a group-sparse $\theta^*$ with high-probability as $\widetilde{O}\left(\sqrt{S_{\mathcal{G}}(m + \log K)}\sqrt{p}\sqrt{T}\right)$.*

# 6 Conclusions

We proved that the regret is upper bounded by $\widetilde{O}(\psi(E_r)w(\Omega_R)\sqrt{p}\sqrt{T})$ which for an unstructured $\theta^*$ matches results in [21]. Additionally, such regret bounds save a factor of $\sqrt{p}$ for structured $\theta^*$ such as $s$-sparse, group-sparse, etc. The regret analysis depends on recent techniques in structure estimation, in particular, generic chaining and provides a geometric perspective on the stochastic linear bandit problem. We presented an algorithm which solved a norm regularized regression problem in each step for which there are several existing tools which implies our algorithm can be implemented with ease.

The main open problem is whether one can shave off the other $\sqrt{p}$ term which comes from the sum of squared widths of the confidence ellipsoid, i.e., $\sum_t (1 + w_t^2)$. We conjecture that such a term can be replaced by the Gaussian width using the statistical dimension.

**Acknowledgements:** The research was supported by NSF grants IIS-1447566, IIS-1422557, CCF-1451986, CNS-1314560, IIS-0953274, IIS-1029711, and by NASA grant NNX12AQ39A.

# A Definitions and Properties

## A.1 Definitions

The following definitions and lemmas can be found in [8, 9, 30].

**Definition 1** A random variable $x$ is sub-Gaussian if the moments satisfies

$$[\mathbb{E}|x|^p]^{\frac{1}{p}} \leq K\sqrt{p} \tag{92}$$

for any $p \geq 1$ with constant $K$. The minimum value of $K$ is called the sub-Gaussian norm of $x$ and denoted by $\|\|x\|\|_{\psi_2}$.
Additionally, every sub-Gaussian random variable satisfies

$$P(|x| > t) \leq \exp\left(1 - c\frac{t^2}{\|x\|_{\psi_2}^2}\right) \tag{93}$$

for all $t \geq 0$.

**Definition 2** A random vector $X \in \mathbb{R}^p$ is sub-Gaussian if the one-dimensional marginals $\langle X, x\rangle$ are sub-Gaussian random variables for all $x \in \mathbb{R}^p$. The sub-Gaussian norm of $X$ is defined as

$$\|\|X\|\|_{\psi_2} = \sup_{x \in S^{p-1}} \|\langle X, x\rangle\|_{\psi_2} \tag{94}$$

**Definition 3** For any set $A \in \mathbb{R}^p$, the Gaussian width of the set $A$ is defined as

$$w(A) = \mathbb{E}\left[\sup_{u \in A}\langle g, u\rangle\right] \tag{95}$$

where the expectation is over $g \sim N(0, \mathbb{I}_{p\times p})$ which is a vector of independent zero-mean unit-variance Gaussian random variables.

**Lemma 10** *For any bounded random variable $|X| \leq B$, then $X$ is a sub-Gaussian random variable with $\|X\|_{\psi_2} \leq B$.*

**Lemma 11** *Consider a sub-Gaussian random vector $X$ with sub-Gaussian norm $K = \max_i \|\|X_i\|\|_{\psi_2}$, then, for vector $a$, $Z = \langle X, a\rangle$ is a sub-Gaussian random variable with sub-Gaussian norm $\|\|Z\|\|_{\psi_2} \leq CK\|a\|_2$ for absolute constant $C$.*

# References


[1] Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. In J. Shawe-Taylor, R.S. Zemel, P.L. Bartlett, F. Pereira, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 24*, pages 2312–2320. Curran Associates, Inc., 2011.

[2] Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Online-to-Confidence-Set Conversions and Application to Sparse Stochastic Bandits. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2012.

[3] J. Abernethy and A. Rakhlin. Beating the adaptive bandit with high probability. In *COLT*, 2009.

[4] J. Abernethy, E. Hazan, and A. Rakhlin. Competing in the dark: An efficient algorithm for bandit linear optimization. In *COLT*, 2008.

[5] P. Auer. Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research*, 3:2002, 2002.

[6] Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47(2-3):235–256, 2002.

[7] Baruch Awerbuch and Robert D. Kleinberg. Adaptive Routing with End-to-end Feedback: Distributed Learning and Geometric Approaches. In *ACM Symposium on Theory of computing (STOC)*, 2004.

[8] Arindam Banerjee, Sheng Chen, Farideh Fazayeli, and Vidyashankar Sivakumar. Estimation with Norm Regularization. In *Neural Information Processing Systems (NIPS)*, 2014.

[9] Arindam Banerjee, Sheng Chen, Farideh Fazayeli, and Vidyashankar Sivakumar. Estimation with Norm Regularization. *arXiv:1505.02294v3*, nov 2015.

[10] P. Bartlett, V. Dani, T. Hayes, S. Kakade, A. Rakhlin, and A. Tewari. High-probability regret bounds for bandit online linear optimization. In *COLT*, July 2008.

[11] Peter J. Bickel, Yaacov Ritov, and Alexandre B. Tsybakov. Simultaneous analysis of Lasso and Dantzig selector. *The Annals of Statistics*, 37(4):1705–1732, 2009. ISSN 0090-5364.

[12] Stephane Boucheron, Gabor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.

[13] Sébastien Bubeck and Nicolò Cesa-Bianchi. *Regret Analysis of Stochastic and Nonstochastic Multi-armed Bandit Problems*, volume 5. NOW, 2012.

[14] Sbastien Bubeck, Nicol Cesa-Bianchi, and Sham M. Kakade. Towards minimax policies for online linear optimization with bandit feedback. In Shie Mannor, Nathan Srebro, and Robert C. Williamson, editors, *COLT*, volume 23 of *JMLR Proceedings*. JMLR.org, 2012.

[15] Emmanuel J. Candes and Terence Tao. The Dantzig selector : statistical estimation when p is much larger than n. *The Annals of Statistics*, 35(6):2313–2351, 2007.

[16] Alexandra Carpentier and Remi Munos. Bandit Theory Meets Compressed Sensing for High-dimensional Stochastic Linear Bandit. *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2012.

[17] Venkat Chandrasekaran, Benjamin Recht, Pablo A. Parrilo, and Alan S. Willsky. The Convex Geometry of Linear Inverse Problems. *Foundations of Computational Mathematics*, 12(6): 805–849, oct 2012. ISSN 1615-3375.

[18] Wei Chu, Lihong Li, Lev Reyzin, and Robert E. Schapire. Contextual bandits with linear payoff functions. In *AISTATS*, volume 15 of *JMLR Proceedings*, pages 208–214. JMLR.org, 2011.

[19] V. Dani and T. Hayes. Robbing the bandit: Less regret in online geometric optimization against an adaptive adversary. In *SODA*, pages 937–943, 2006.

[20] V. Dani, S. Kakade, and T. Hayes. The price of bandit information for online optimization. In *NIPS*, pages 345–352. MIT Press, 2007.

[21] Varsha Dani, Thomas P. Hayes, and Sham M. Kakade. Stochastic Linear Optimization Under Bandit Feedback. In *Conference on Learning Theory (COLT)*, 2008.

[22] Sebastien Gerchinovitz. Sparsity regret bounds for individual sequences in online linear regression. In *COLT*, JMLR Proceedings, 2011.

[23] Lihong Li, Wei Chu, John Langford, and Robert E. Schapire. A Contextual-Bandit Approach to Personalized News Article Recommendation. *International World Wide Web Conference (WWW)*, 2010.

[24] B. McMahan and A. Blum. Online geometric optimization in the bandit setting against an adaptive adversary. In *COLT*, pages 109–123, 2004.

[25] Sahand N. Negahban, Pradeep Ravikumar, Martin J. Wainwright, and Bin Yu. A Unified Framework for High-Dimensional Analysis of $M$-Estimators with Decomposable Regularizers. *Statistical Science*, 27(4):538–557, 2012. ISSN 0883-4237.

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

[27] Michel Talagrand. *The Generic Chaining*. Springer Monographs in Mathematics. Springer Berlin, 2005.

[28] Michel Talagrand. *Upper and Lower Bounds for Stochastic Processes*. Springer-Verlag, 2014.

[29] William R. Thompson. On the Likelihood that One Unknown Probability Exceeds Another in View of the Evidence of Two Samples. *Biometrika*, 25(3/4):285–294, 1933.

[30] Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices. In Y Eldar and G. Kutyniok, editors, *Compressed Sensing*, pages 210–268. Cambridge University Press, Cambridge, nov 2012.