Phase Transition for Support Recovery from Gaussian Linear Measurements

Lekshmi Ramesh   Chandra R. Murthy   Himanshu Tyagi

Abstract—**We study the problem of recovering the common k-sized support of a set of n samples of dimension d, using m noisy linear measurements per sample. Most prior work has focused on the case when m exceeds k, in which case n of the order $(k/m)\log(d/k)$ is both necessary and sufficient. Thus, in this regime, only the total number of measurements across the samples matter, and there is not much benefit in getting more than k measurements per sample. In the measurement-constrained regime where we have access to fewer than k measurements per sample, we show an upper bound of $O((k^2/m^2)\log d)$ on the sample complexity for successful support recovery. Along with the lower bound from our previous work, this shows a sharp phase transition for the sample complexity of this problem around $k/m = 1$. In fact, our proposed algorithm is sample-optimal in both the regimes. It follows that, in the $m \ll k$ regime, multiple measurements from the same sample are more valuable than measurements from different samples.**

I. Introduction

The problem of support recovery in the single sample setting considers the following question: given noisy linear measurements $Y = \Phi x + W \in \mathbb{R}^m$ of a k-sparse vector $x \in \mathbb{R}^d$, can we recover the locations of its nonzero entries when $m < d$? The set of indices corresponding to the nonzero entries of x is called the support of x, and is denoted by $\mathrm{supp}(x)$. The measurement matrix $\Phi \in \mathbb{R}^{m \times d}$ is a design parameter that is chosen to enable exact or approximate recovery of $\mathrm{supp}(x)$, and $W \sim \mathcal{N}(0, \sigma^2 I)$ is noise. This problem (also sometimes referred to as model selection or variable selection) has received a lot of attention in the past decade [17], [4], [1], [14], [7], with a focus on designing recovery algorithms and on determining the number of measurements m required for successful recovery. In particular, it is known that $m = \Theta(k \log(d - k))$ measurements are necessary and sufficient for support recovery with high probability using a Gaussian measurement matrix [17]. It is important to note that this tight scaling holds in the low signal to noise ratio (SNR) regime of $x_{\min}/\sigma^2 = \Theta(1/k)$, where $x_{\min} \overset{\mathrm{def}}{=} \min_{i \in [d]} x_i$. In other

regimes of SNR, either the $\log$ dependence changes or the upper and lower bounds are known to differ by a factor of $(\log(1 + kx_{\min}^2/\sigma^2))^{-1}$; see [7] for a detailed discussion.

Parallel to the results in the single sample setting, there has been work on the natural extension of this problem to the multiple sample setting, which is the focus of this work. In this setting, there are multiple samples $x_1, \ldots, x_n$, all sharing a common unknown support $\mathcal{S}$ of cardinality k. For each sample x_i, we observe measurements $Y_i = \Phi_i x_i + W_i$, and the goal is to recover $\mathcal{S}$. We can ask the question of how the number of measurements per sample (m) and the number of samples (n) can be traded off for each other, and whether it is useful to take more samples or more measurements per sample.

While there have been several works in the multiple sample setting [18], [16], [5], [6], [15], [10], they focus on the regime where one has access to roughly $m \geq k$ measurements per sample. In particular, omitting the dependence on SNR, [10] shows that $mn = \Theta(k \log(d/k))$ is necessary and sufficient assuming $m = \Omega(k)$ and $k = o(d)$. While the sufficient condition in [10] is obtained via analysis of an exhaustive search decoder, algorithms such as the group LASSO also show a similar scaling of $mn = \Theta(k \log(d - k))$ provided $m > k$ [8]. From the discussion in the previous paragraph, it is clear that if we have $m = \Omega(k \log(d - k))$, then a single sample is sufficient for support recovery. Therefore, given that we have access to multiple samples now, a more interesting question to consider is whether we can perform recovery with $m < k$ measurements per sample. This measurement-constrained regime has received some attention in the past [2], [9], [12] and it was recently shown for the case of *random* inputs drawn from a subgaussian distribution that the tradeoff (ignoring noise variance and parameters dependent on the generative model for the samples) is $n = \Theta((k^2/m^2)\log d)$ [13].

In this work, we focus on the case of fixed inputs and show that the same tradeoff holds in this case as well. In particular, the lower bound from [13] for Gaussian inputs applies to this worst-case, fixed input setting, since an instance in the Gaussian case can be thought of as the input in the fixed case. Showing that the upper
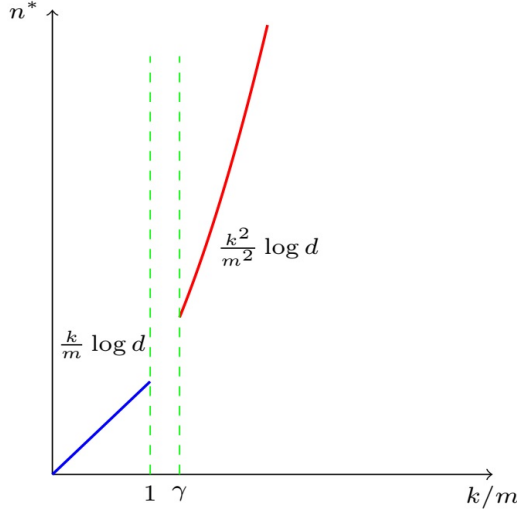
Fig. 1: Sample complexity of support recovery as a function of k/m.

bound also remains the same requires more work, and is the main focus of this paper. Specifically, we analyze the performance of the estimator from [13] in the fixed input setting, which requires improved bounds on the tail probability of heavy-tailed random variables[1] than the one used in [13].

In summary, we settle the question of tradeoff between m and n in the $m < k$ regime, and show that there exists a phase transition for the sample complexity of this problem at $k/m = 1$ as depicted in Figure I. Roughly, around this point, the sample complexity for support recovery undergoes a change from being linear in the ratio k/m to being quadratic in k/m (up to a factor $\log d$).

We note that the current lower bound proof from [13] requires some separation between k and m; namely, it requires $k/m > \gamma$ for some $\gamma > 1$. While the lower bound of $n = \Omega((k/m) \log(d/k))$ [10] continues to hold for $m < k$, it is not clear if a tighter lower bound on sample complexity in the regime $1 < k/m \leq \gamma$ can be obtained. Such a separation between k and m is, however, not required when deriving the upper bound.

Notation. We use upper case letters to denote either random variables or deterministic matrices, and lower-case letters to denote deterministic scalars or vectors. For a vector x, we use x_i to denote its ith entry, and for matrices $\{A_i\}_{i=1}^n$ we use A_{ij} to denote the jth column of A_i. We use $\|x\|_p^p \overset{\text{def}}{=} \sum_{i=1}^d x_i^p$ to denote the ℓ_p norm of a vector $x \in \mathbb{R}^d$, and $\|Z\|_{\mathcal{L}_p}^p \overset{\text{def}}{=} \mathbb{E}\left[|Z|^p\right]$ to denote the $\mathcal{L}_p$ norm of a random variable Z.

[1] We refer to a random variable as heavy-tailed if its moment generating function is unbounded.

Organization. In the next section, we formally state the problem and present our main result. We present the proof of our main result in Section III, and end with a discussion on directions for further work in Section IV.

II. PROBLEM FORMULATION AND MAIN RESULT

Let vectors $x_1, \ldots, x_n$ in $\mathbb{R}^d$ have a common support $\mathcal{S} \subset [d]$ of cardinality k. For each of these vectors, we have access to noisy linear measurements of the form $Y_i = \Phi_i x_i + W_i$, $i \in [n]$. Here, $\Phi_i \in \mathbb{R}^{m \times d}$ with $m < d$ are called the measurement matrices and $W_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2 I)$ is noise. The goal is to recover the support $\mathcal{S}$ using $\{Y_i, \Phi_i\}_{i=1}^n$. An estimator for $\mathcal{S}$ is a mapping $\hat{\mathcal{S}} : \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times d \times n} \to \binom{[d]}{k}$, where $\binom{[d]}{k}$ denotes the set of all subsets of $[d]$ of cardinality k. We assume that the estimator has knowledge of k and consider the probability of exact recovery, $\Pr\left(\hat{\mathcal{S}} \neq \mathcal{S}\right)$, as our recovery criterion. We note that one could also consider the setting where $|\mathcal{S}| \leq k$. The estimator that we consider here would output an $\hat{\mathcal{S}}$ that contains the true support with high probability. In this work, however, we assume that the true support has cardinality exactly k.

We make the following two assumptions on the measurement matrices and the input samples:

Assumption 1. *The $m \times d$ measurement matrices $\Phi_1, \ldots, \Phi_n$ are independent, with entries that are independent and distributed as $\mathcal{N}(0, 1/m)$.*

Assumption 2. *The d-dimensional inputs $x_1, \ldots, x_n$ are such that $\text{supp}(x_i) = \mathcal{S}$, for all $i \in [n]$, where $\mathcal{S} \subset [d]$ is a fixed set of cardinality k. Further, $|x_{iu}| \in [x_{\min}, x_{\max}]$, for all $i \in [n]$, $u \in \mathcal{S}$, where $x_{\min}, x_{\max} \in \mathbb{R}$.*

We focus on a measurement-constrained setting where we obtain only $m < k$ measurements per sample. The fundamental quantity of interest for us in this paper is the *sample complexity of support recovery,* defined below.

Definition 1. *For $m, k, d \in \mathbb{N}$, the sample complexity of support recovery $n^*(m, k, d, x_{\min}, x_{\max}, \sigma^2, \delta)$ is the minimum number of samples n for which we can find an estimator that can recover $\mathcal{S}$ with probability of error at most δ. Mathematically,*

$$\Pr\left(\hat{\mathcal{S}} \neq \mathcal{S}\right) \leq \delta, \quad \forall \mathcal{S} \in \binom{[d]}{k}. \tag{1}$$

For notational convenience, we use $n^*(m, k, d, x_{\min}, x_{\max}, \sigma^2, \delta) = n^*$ in the rest of the paper. Our main result shows a phase transition that occurs at $k/m = 1$ for the problem of support recovery. In particular, the dependence of sample complexity of support recovery on k/m undergoes a sharp change from linear to quadratic as we move from the $k/m \leq 1$ regime to the $k/m > 1$ regime. As mentioned before, a tight characterization of n^* in the interval $1 < k/m \leq \gamma$

is not known, although our upper bound stated in Theorem 1 below continues to hold for this case. Our main result is the following.

Theorem 1. *The sample complexity of support recovery under Assumptions 1 and 2, for $m \geq 1$, $k \leq d/2$, and $d \geq 4$, satisfies*

$$n^* = O\left(\frac{x_{\max}^4}{x_{\min}^4} \max \left\{ \left(\frac{k}{m} + \frac{\sigma^2}{x_{\max}^2} \right) \log \frac{d}{\delta}, \right. \right.$$
$$\left. \left. \left(\frac{k}{m} + \frac{\sigma^2}{x_{\max}^2} \right)^2 \log \frac{d}{\delta} \right\} \right).$$

As a special case, in the noiseless setting with $m < k$, we have the following corollary, which follows from Theorem 1 above and the lower bound in [13][2]. Note that the lower bound is stated for a constant probability of error.

Corollary 1. *In the noiseless setting, with $m < k \leq d/2$, $d \geq 4$ and $\delta = 1/3$, we have,*

$$n^* = \Theta\left(\frac{x_{\max}^4}{x_{\min}^4} \frac{k^2}{m^2} \log d \right).$$

We provide the proof of Theorem 1 in the next section.

III. ANALYSIS OF THE ESTIMATOR

We will analyze the closed form estimator from [13], but instead of random inputs, here we will consider fixed inputs $x_1, \ldots, x_n$. To see why the analysis in [13] does not extend in a straightforward way to this case, we first recall the form of the estimator. Let $\Phi_{iu} \in \mathbb{R}^m$ denote the uth column of Φ_i. We first compute proxy samples $\hat{X}_1, \ldots, \hat{X}_n$ with entries

$$\hat{X}_{iu} \overset{\text{def}}{=} \Phi_{iu}^\top Y_i = \Phi_{iu}^\top \Phi_i x_i + \Phi_{iu}^\top W_i, \quad u \in [d], \quad (2)$$

and then the sample variance along each coordinate as

$$\tilde{\lambda}_u \overset{\text{def}}{=} \frac{1}{n} \sum_{i=1}^n \hat{X}_{iu}^2, \quad u \in [d]. \quad (3)$$

The support estimate $\tilde{\mathcal{S}}$ consists of the k indices of $\tilde{\lambda}$ with the largest value. Analyzing the estimator would basically involve obtaining tail bounds for the random variable above. Considering the noiseless case first, note that each summand in (3) is of the form $(\Phi_{iu}^\top \Phi_i x_i)^2$, and can be viewed as a quadratic in either x_i or $\Phi_{iu}^\top \Phi_i$.

When x_is are random and subgaussian with independent coordinates, we can exploit the quadratic form in x_i to obtain a tail bound using the Hanson-Wright inequality (after conditioning on Φ_i). On the other hand, when x_i are deterministic, the summands in (3) are quadratic in $\Phi_{iu}^\top \Phi_i$, resulting in a heavy-tailed random

[2]When $k \leq d/2$ and $d \geq 4$, the $\log(k(d-k))$ factor in the lower bound is equal, upto constants, to $\log d$.

variable, and standard methods based on bounding the moment generating function (MGF) do not work.

We explain in the next section how a careful analysis involving conditioning on a certain *column* of Φ_i followed by a moment based bound can be used to get exponential tail bounds for heavy-tailed random variables. The analysis in [13] also deals with heavy-tailed random variables, but using a more elementary approach (see Lemmas B.2 and B.3 for instance) which would not work here.

A. A separation condition for support recovery

We will analyze the error probability of a threshold-based version of the estimator described in the previous section. In particular, we will use the estimate $\hat{\lambda} \overset{\text{def}}{=} \mathbb{1}_{\{\tilde{\lambda} \geq \tau\}}$, for an appropriate threshold τ, since

$$\Pr\left(\tilde{\mathcal{S}} \neq \mathcal{S} \right) \leq \Pr\left(\hat{\mathcal{S}} \neq \mathcal{S} \right),$$

where $\tilde{\mathcal{S}}$ and $\hat{\mathcal{S}}$ denote, respectively, the supports of $\tilde{\lambda}$ and $\hat{\lambda}$. The error probability $\Pr\left(\hat{\mathcal{S}} \neq \mathcal{S} \right)$ will essentially be determined by the tail behaviour of the variance estimate $\tilde{\lambda}$. Recall from the last section that variance estimate is an average of random variables of the form $(\Phi_{iu}^\top \Phi_i x_i + \Phi_{iu}^\top W_i)^2$. The $\Phi_{iu}^\top \Phi_i x_i$ term will be indicative of whether the coordinate u lies in the support or not, since it will have a $\|\Phi_{iu}\|_2^2$ term only when $u \in \mathcal{S}$.

The analysis is greatly simplified once we condition on Φ_{iu}, because then the summands in (3) are noncentral chi-square distributed, for which tail bounds can be obtained using standard methods. The error probability can be made small provided these tail probabilities (parameterized by Φ_{iu}) can be made small, which eventually leads to a condition on the measurement ensemble. We will show, using tail bounds for heavy-tailed random variables, that this condition is satisfied with high probability for the Gaussian ensemble when the parameters (n, m, k, d) scale as indicated in Theorem 1, thus finishing the proof.

The probability of error can be bounded as

$$\Pr\left(\hat{\mathcal{S}} \neq \mathcal{S} \right) \leq \sum_{u \in \mathcal{S}} \Pr\left(\tilde{\lambda}_u < \tau | E \right)$$
$$+ \sum_{u' \in \mathcal{S}^c} \Pr\left(\tilde{\lambda}_{u'} \geq \tau | E \right) + \Pr\left(E^c \right),$$
$$(4)$$

where E denotes the event that the measurement ensemble satisfies a certain condition, which we will describe shortly. For the right hand side to remain below δ, we require the summands in the first two terms to be at most $\delta/(3 \max\{k, d-k\})$. For simplicity, we will work with a requirement of $\delta/3d$.

Now, using (2) and (3), we can see that $\hat{X}_{iu}|\Phi_{iu} \sim \mathcal{N}(\mu_i, \nu_i^2)$ for $u \in \mathcal{S}$ with

$$\mu_i = \|\Phi_{iu}\|_2^2 x_{iu},$$

and

$$\nu_i^2 = \frac{\|\Phi_{iu}\|_2^2}{m} \sum_{v \in \mathcal{S}\backslash\{u\}} x_{iv}^2 + \sigma^2 \|\Phi_{iu}\|_2^2.$$

Similarly, we have $\hat{X}_{iu'}|\Phi_{iu'} \sim \mathcal{N}(0, \nu_i'^2)$, for $u' \in \mathcal{S}^c$, where

$$\nu_i'^2 = \frac{\|\Phi_{iu'}\|_2^2}{m} \sum_{v \in \mathcal{S}} x_{iv}^2 + \sigma^2 \|\Phi_{iu'}\|_2^2.$$

A direct application of Lemma A.1 then yields, for every $u \in \mathcal{S}$,

$$\Pr\left(\tilde{\lambda}_u < \tau | \{\Phi_{iu}\}_{i=1}^n\right) \le \exp\left(\frac{-n^2(\mu-\tau)^2}{4(\sum_{i=1}^n \nu_i^4 + \nu_i^2\mu_i^2)}\right),$$

where $\mu \overset{\text{def}}{=} \mathbb{E}\left[\tilde{\lambda}_u | \{\Phi_{iu}\}_{i=1}^n\right]$. For $u' \in \mathcal{S}^c$, we can obtain in a similar manner from Lemma A.1,

$$\Pr\left(\tilde{\lambda}_{u'} \ge \tau | \{\Phi_{iu'}\}_{i=1}^n\right)$$
$$\le \exp\left(-\min\left\{\frac{n^2(\tau-\mu')^2}{16\sum_{i=1}^n \nu_i'^4}, \frac{n(\tau-\mu')}{8\max_{i\in[n]}\nu_i'^2}\right\}\right),$$

where $\mu' \overset{\text{def}}{=} \mathbb{E}\left[\tilde{\lambda}_{u'} | \{\Phi_{iu}\}_{i=1}^n\right]$. For the missed detection and false alarm probabilities above to remain bounded above by $\delta/3d$, we require

$$\tau \le \mu - \sqrt{\frac{4}{n^2}\sum_{i=1}^n (\nu_i^4 + \mu_i^2\nu_i^2)\log\frac{3d}{\delta}},$$

and

$$\tau \ge \mu' + \max\left\{\sqrt{\frac{16}{n^2}\sum_{i=1}^n \nu_i'^4 \log\frac{3d}{\delta}}, \frac{8}{n}\max_{i\in[n]}\nu_i'^2 \log\frac{3d}{\delta}\right\}.$$

Therefore, for the existence of a threshold τ, we can see upon simplification that it suffices to have

$$\mu - \mu' > \sqrt{\frac{4}{n^2}\sum_{i=1}^n (\nu_i^4 + \nu_i^2\mu_i^2)\log\frac{3d}{\delta}}$$
$$+ \max\left\{\sqrt{\frac{16}{n^2}\sum_{i=1}^n \nu_i'^4 \log\frac{3d}{\delta}}, \frac{8}{n}\max_{i\in[n]}\nu_i'^2 \log\frac{3d}{\delta}\right\}. \tag{5}$$

A simple calculation shows that the conditional mean of the estimator under the $u \in \mathcal{S}$ and $u' \in \mathcal{S}^c$ cases are separated roughly by a constant term (after averaging over the measurement matrices), which makes the distinction

between the two cases possible. In particular,

$$\mu = \frac{1}{n}\sum_{i=1}^n \left(x_{iu}^2\|\Phi_{iu}\|_2^4 \right.$$
$$\left. + \|\Phi_{iu}\|_2^2\left(\frac{1}{m}\sum_{v\in\mathcal{S}\backslash\{u\}} x_{iv}^2 + \sigma^2\right)\right),$$

and

$$\mu' = \frac{1}{n}\sum_{i=1}^n \|\Phi_{iu}\|_2^2\left(\frac{1}{m}\sum_{v\in\mathcal{S}} x_{iv}^2 + \sigma^2\right).$$

Substituting this into (5) and simplifying, we can rewrite the condition as

$$\frac{x_{\min}^2}{x_{\max}^2}\frac{1}{n}\sum_{i=1}^n\left(\|\Phi_{iu}\|_2^4 - \frac{1}{m}\|\Phi_{iu}\|_2^2\right)$$
$$> \sqrt{\frac{4}{n^2}\left(\frac{k-1}{m} + \frac{\sigma^2}{x_{\max}^2}\right)^2\sum_{i=1}^n \|\Phi_{iu}\|_2^4 \log\frac{3d}{\delta}}$$
$$+ \sqrt{\frac{4}{n^2}\left(\frac{k-1}{m} + \frac{\sigma^2}{x_{\max}^2}\right)\sum_{i=1}^n \|\Phi_{iu}\|_2^6 \log\frac{3d}{\delta}}$$
$$+ \sqrt{\frac{16}{n^2}\left(\frac{k}{m} + \frac{\sigma^2}{x_{\max}^2}\right)^2\sum_{i=1}^n \|\Phi_{iu'}\|_2^4 \log\frac{3d}{\delta}}$$
$$+ \frac{8}{n}\left(\frac{k}{m} + \frac{\sigma^2}{x_{\max}^2}\right)\max_{i\in[n]}\|\Phi_{iu'}\|_2^2 \log\frac{3d}{\delta}, \tag{6}$$

for every $(u, u') \in \mathcal{S} \times \mathcal{S}^c$.

B. Separation condition for the Gaussian ensemble

We will show that when the measurement ensemble is Gaussian as described in Assumption 1, the separation condition in (6) is satisfied with high probability for a certain regime of the parameters (n, m, k, d). We will derive upper and lower bounds on the right hand side and left hand side respectively in (6), that hold with high probability, which after simplification will finally result in a condition on the parameters as stated in Theorem 1. Note that this translates to obtaining tail bounds for the random variable $(1/n)\sum_{i=1}^n \|\Phi_{iu}\|_2^{2q}$ with $q = 2, 3$. It is easy to see that $\|\Phi_{iu}\|_2^2$ is chi-square distributed (after scaling by m), and $\|\Phi\|_2^{2q}$ is therefore a heavy-tailed random variable, and so MGF based methods cannot be used here. We will see that a bound on the moments can be used to get exponential tail bounds (even when the MGF is unbounded). We will fix $q = 3$ and derive our results; the same arguments can be used for the $q = 2$ case as well. Define $Z \overset{\text{def}}{=} |(1/n)\sum_{i=1}^n(\|\Phi_{iu}\|_2^6 - \mathbb{E}\left[\|\Phi_{iu}\|_2^6\right]|$ and note that for all $p \ge 1$,

$$\Pr\left(Z \ge e(\mathbb{E}\left[Z^p\right])^{\frac{1}{p}}\right) = \Pr\left(Z^p \ge e^p\mathbb{E}\left[Z^p\right]\right) \le e^{-p}. \tag{7}$$

Further, for all $p \geq 2$, if we can show that $(\mathbb{E}[Z^p])^{\frac{1}{p}} \leq cp^\beta$ for some $\beta > 0$, then together with the previous inequality it implies that $\Pr\left(Z \geq ecp^\beta\right) \leq e^{-p}$, or, equivalently, for $t > 0$, that

$$\Pr\left(Z \geq t\right) \leq \exp(-(t/ec)^{\frac{1}{\beta}}). \tag{8}$$

We now need to determine an upper bound on $\|Z\|_{\mathcal{L}_p} \overset{\text{def}}{=} (\mathbb{E}[Z^p])^{\frac{1}{p}}$. We show such a moment bound, resulting in the following lemma.

Lemma 1. *For every $t > 0$,*

$$\Pr\left(\left|\frac{1}{n}\sum_{i=1}^n (\|\Phi_{iu}\|_2^6 - \mathbb{E}\left[\|\Phi_{iu}\|_2^6\right])\right| \geq t\right)$$
$$\leq \exp\left(-C\min\left\{nt, (m^3 nt)^{\frac{1}{4}}, nt^2\right\}\right).$$

A similar result can be obtained for the $(1/n)\sum_{i=1}^n \|\Phi_{iu}\|_2^4$ term in (6). This, together with the fact that $\mathbb{E}\left[\|\Phi_{iu}\|_2^4\right] = 1 + 2/m$ and $\mathbb{E}\left[\|\Phi_{iu}\|_2^6\right] = 1 + 6/m + 8/m^2$, gives upper and lower bounds that hold with high probability on all but the $\max_{i \in [n]} \|\Phi_{iu'}\|_2^2$ term in (6). The latter can be bounded with high probability using concentration for chi-squared random variables and a union bounding step, as given by the following lemma.

Lemma 2. *Let $\mu_{\max} \overset{\text{def}}{=} \mathbb{E}\left[\max_{i \in [n]} \|\Phi_{iu}\|_2^2\right]$. Then, for every $t > 0$,*

$$\Pr\left(\max_{i \in [n]} \|\Phi_{iu}\|_2^2 \geq \mu_{\max} + t\right)$$
$$\leq n \exp\left(\frac{-m}{8}\min\left\{(\mu_{\max} + t - 1)^2, \mu_{\max} + t - 1\right\}\right).$$

To ensure that the random variable on the left hand side of (6) exceeds the one on the right hand side with large probability, we can substitute the bounds from Lemmas 1 and 2, and check when the inequality holds. This results (up to some constant loss in the δ factor) in a condition on the problem parameters under which (6) holds for a fixed $(u, u') \in \mathcal{S} \times \mathcal{S}^c$. Applying a union bound over all $k(d-k)$ pairs gives the final requirement on n. Note that the leading terms on the right hand side of (6) would be roughly $\sqrt{(k^2/m^2 n)\log d/\delta}$, while the left hand side would be roughly a constant, leading to the following result.

Lemma 3. *The separation condition (6) holds for every $(u, u') \in \mathcal{S} \times \mathcal{S}^c$, with probability at least $1-\delta$, provided*

$$n \geq c\frac{x_{max}^4}{x_{\min}^4}\max\left\{\left(\frac{k}{m} + \frac{\sigma^2}{x_{\max}^2}\right)\log\frac{d}{\delta},\right.$$
$$\left.\left(\frac{k}{m} + \frac{\sigma^2}{x_{\max}^2}\right)^2\log\frac{d}{\delta}\right\},$$

for an absolute constant c.

By defining E as the event that the measurement matrices satisfy condition (6) for every $(u, u') \in \mathcal{S} \times \mathcal{S}^c$, we can see that the probability of error in (4) is at most δ, provided n satisfies the condition in Lemma 3. This completes the proof of Theorem 1.

IV. DISCUSSION

We showed a phase transition for the problem of support recovery from multiple samples. While the closed form estimator that we analyzed here is sample-optimal, it would be interesting to design other estimators that can work in the measurement-constrained regime without knowledge of the support size, and for which guarantees can be obtained with worst-case inputs. Finally, extending the lower bound on n^* to include the $1 < k/m \leq \gamma$ regime for $\gamma > 1$ would provide a better understanding of the problem.

APPENDIX A

Lemma A.1. *Let $X_1, \ldots, X_n$ be drawn i.i.d. from $\mathcal{N}(\mu_i, \sigma_i^2)$. Then, for every $t > 0$,*

$$\Pr\left(\frac{1}{n}\sum_{i=1}^n X_i^2 \leq \frac{1}{n}\sum_{i=1}^n (\sigma_i^2 + \mu_i^2) - t\right)$$
$$\leq \exp\left(\frac{-n^2 t^2}{4\sum_{i=1}^n (\sigma_i^4 + \sigma_i^2\mu_i^2)}\right),$$

and

$$\Pr\left(\frac{1}{n}\sum_{i=1}^n X_i^2 \geq \frac{1}{n}\sum_{i=1}^n (\sigma_i^2 + \mu_i^2) + t\right)$$
$$\leq \exp\left(-\min\left\{\frac{n^2 t^2}{16\sum_{i=1}^n (\sigma_i^4 + \sigma_i^2\mu_i^2)}, \frac{nt}{8\max_{i \in [n]}\sigma_i^2}\right\}\right).$$

Proof. The proof is similar to that of [3] for $\sigma^2 = 1$, and follows by upper bounding the MGF of a noncentral chi-squared random variable and then using the Chernoff method. We include the proof here for completeness. We will first show the left tail bound. To that end, we note that for $t > 0$ and $\lambda < 0$, the following holds for $Y \overset{\text{def}}{=} (1/n)\sum_{i=1}^n X_i^2$:

$$\Pr\left(Y \leq \mathbb{E}[Y] - t\right) \leq e^{\lambda t}\mathbb{E}\left[e^{\lambda(Y - \mathbb{E}[Y])}\right]. \tag{9}$$

To upper bound the MGF, first note that for $X \sim \mathcal{N}(\mu, \sigma^2)$,

$$\mathbb{E}\left[e^{\lambda(X^2 - \mathbb{E}[X^2])}\right] = e^{-\lambda(\sigma^2 + \mu^2)}\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{\lambda x^2} e^{\frac{-(x-\mu)^2}{2\sigma^2}}dx$$
$$= \frac{e^{-\lambda(\sigma^2 + \mu^2)}}{\sqrt{1 - 2\lambda\sigma^2}}e^{\frac{\lambda\mu^2}{1-2\lambda\sigma^2}},$$

for all $\lambda < 1/2\sigma^2$. Taking logarithms we have

$$\log\mathbb{E}\left[e^{\lambda(X^2 - \mathbb{E}[X^2])}\right]$$

$$= \frac{1}{2}\left(-\log(1 - 2\lambda\sigma^2 - 2\lambda\sigma^2)\right) + \frac{2\lambda^2\mu^2\sigma^2}{1 - 2\lambda\sigma^2} \quad (10)$$

$$\leq \lambda^2\sigma^4 + \frac{2\lambda^2\mu^2\sigma^2}{1 - 2\lambda\sigma^2}$$

$$\leq \lambda^2(\sigma^4 + 2\mu^2\sigma^2),$$

where we used $-\log(1 - x) - x \leq x^2/2$ for $x < 0$ in the second step. This gives

$$\log \mathbb{E}\left[e^{\lambda(Y - \mathbb{E}[Y])}\right] \leq \frac{\lambda^2}{n^2}\sum_{i=1}^{n}(\sigma_i^4 + \sigma_i^2\mu_i^2)$$

which upon substituting into (9) and optimizing over $\lambda < 0$ gives $\lambda = -n^2 t/(2\sum_{i=1}^{n}(\sigma_i^4 + \sigma_i^2\mu_i^2)))$ resulting in the left tail bound claimed in the lemma.

For the right tail bound, we continue from (10) and note that for $0 \leq \lambda \leq 1/4\sigma^2$,

$$\log \mathbb{E}\left[e^{\lambda(X^2 - \mathbb{E}[X^2])}\right] \leq \frac{2\lambda^2\sigma^4}{1 - 2\lambda\sigma^2} + \frac{2\lambda^2\mu^2\sigma^2}{1 - 2\lambda\sigma^2}$$
$$\leq 4\lambda^2(\sigma^4 + \mu^2\sigma^2),$$

where in the first step we used $-\log(1 - x) - x \leq x^2/2(1 - x)$ for all $x \in [0, 1)$. Extending as before to the normalized sum $(1/n)\sum_{i=1}^{n}X_i^2$, substituting into (9) and optimizing over $\lambda \in [0, 1/4\sigma^2)$, it can be seen that the minimum is attained at $\lambda = nt/(8\sum_{i=1}^{n}(\sigma_i^4 + \mu_i^2\sigma^2))$ if $t < 2\sum_{i=1}^{n}(\sigma^2 + \mu^2)$, and at $\lambda = 1/(4\sum_{i=1}^{n}\sigma_i^2)$ otherwise. This gives the right tail bound claimed in the lemma. $\square$

Proof of Lemma 1. First, note that

$$\left\|\frac{1}{n}\sum_{i=1}^{n}(\|\Phi_{iu}\|_2^6 - \mathbb{E}\left[\|\Phi_{iu}\|_2^6\right])\right\|_{\mathcal{L}_p}$$

$$= \frac{1}{nm^3}\left\|\sum_{i=1}^{n}(V_i^3 - \mathbb{E}\left[V_i^3\right])\right\|_{\mathcal{L}_p}, \quad (11)$$

where $V_i \overset{\text{def}}{=} m\|\Phi_{iu}\|_2^2 \sim \chi_m^2$, and χ_m^2 denotes the chi-square distribution with m degrees of freedom. To bound the moment of the sum, we use the following form of Rosenthal's inequality stated in [11].

Lemma A.2 ([11]). *Let $Z_1, \ldots, Z_n$ be independent and identically distributed random variables with mean zero. Then, for every $p \geq 2$,*

$$\left\|\sum_{i=1}^{n}Z_i\right\|_{\mathcal{L}_p} \leq c\left(pn^{\frac{1}{p}}\|Z_1\|_{\mathcal{L}_p} + \sqrt{pn}\|Z_1\|_{\mathcal{L}_2}\right),$$

for an absolute constant c.

In view of Lemma A.2, we now upper bound the $\mathcal{L}_p$ norm of each summand on the right side of (11) as follows:

$$\|V_i^3 - \mathbb{E}\left[V_i^3\right]\|_{\mathcal{L}_p}$$
$$\leq \|V_i^3\|_{\mathcal{L}_p} + \mathbb{E}\left[V_i^3\right]$$

$$= (\mathbb{E}\left[V_i^{3p}\right])^{\frac{1}{p}} + \mathbb{E}\left[V_i^3\right]$$

$$= \left(2^{3p}\frac{\Gamma(3p + m/2)}{\Gamma(m/2)}\right)^{\frac{1}{p}} + 2^3\frac{\Gamma(3 + m/2)}{\Gamma(m/2)}$$

$$\leq 2^3\left(e^{\frac{1}{p}}(3p + m/2)^3 + e(3 + m/2)^3\right)$$

$$\leq 2^6(3p + m/2)^3,$$

where we used the fact that $V_i \sim \chi_m^2$ in the third step and $\Gamma(x + a)/\Gamma(x) \leq e(x + a)^a$ for all $x \geq 1$, $a > 0$ in the fourth step. Together with Lemma A.2, this yields for $p \geq 2$,

$$\left\|\frac{1}{n}\sum_{i=1}^{n}(\|\Phi_{iu}\|_2^6 - \mathbb{E}\left[\|\Phi_{iu}\|_2^6\right])\right\|_{\mathcal{L}_p}$$

$$\leq \frac{c2^6}{nm^3}\left(pn^{\frac{1}{p}}(3p + m/2)^3 + \sqrt{pn}(6 + m/2)^3\right)$$

$$\leq c2^6\left(\frac{p}{n^{1-\frac{1}{p}}}\max\left\{1, \frac{(6p)^3}{m^3}\right\} + 7^3\sqrt{\frac{p}{n}}\right)$$

$$\leq c'\max\left\{\frac{p}{n^{1-\frac{1}{p}}}, \frac{p^4}{m^3 n^{1-\frac{1}{p}}}, \sqrt{\frac{p}{n}}\right\}. \quad (12)$$

Note that from (7), we expect p to be of the form $n^{c''}$ for some constant c'', in which case $p/n^{1-\frac{1}{p}} = (p/n)e^{\frac{1}{ec''}}$. We focus on this regime, and obtain using (8) and (12),

$$\Pr\left(\left|\frac{1}{n}\sum_{i=1}^{n}(\|\Phi_{iu}\|_2^6 - \mathbb{E}\left[\|\Phi_{iu}\|_2^6\right]) \geq t\right|\right)$$

$$\leq \exp\left(-C\min\left\{nt, (m^3 nt)^{\frac{1}{4}}, nt^2\right\}\right),$$

for every $t > 0$. $\square$

References

[1] S. Aeron, V. Saligrama, and M. Zhao, "Information theoretic bounds for compressed sensing," *IEEE Trans. on Inf. Theory*, vol. 56, no. 10, pp. 5111–5130, 2010.

[2] O. Balkan, K. Kreutz-Delgado, and S. Makeig, "Localization of more sources than sensors via jointly-sparse bayesian learning," *IEEE Signal Process. Lett.*, vol. 21, no. 2, pp. 131–134, 2014.

[3] L. Birgé, *An alternative point of view on Lepski's method*, ser. Lecture Notes–Monograph Series. Beachwood, OH: Institute of Mathematical Statistics, 2001, vol. Volume 36, pp. 113–133.

[4] A. K. Fletcher, S. Rangan, and V. K. Goyal, "Necessary and sufficient conditions for sparsity pattern recovery," *IEEE Trans. on Inf. Theory*, vol. 55, no. 12, pp. 5758–5772, 2009.

[5] S. Foucart, "Recovering jointly sparse vectors via hard thresholding pursuit," in *Sampling Theory and Applications SAMPTA 2011, Singapore*, 2011.

[6] Y. Jin and B. D. Rao, "Support recovery of sparse signals in the presence of multiple measurement vectors," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 3139–3157, May 2013.

[7] M. Ndaoud and A. B. Tsybakov, "Optimal variable selection and adaptive noisy compressed sensing," *IEEE Transactions on Information Theory*, vol. 66, no. 4, pp. 2517–2532, 2020.

[8] G. Obozinski, M. J. Wainwright, and M. I. Jordan, "Support union recovery in high-dimensional multivariate regression," *Ann. Statist.*, vol. 39, no. 1, pp. 1–47, 2011.

[9] P. Pal and P. P. Vaidyanathan, "Pushing the limits of sparse support recovery using correlation information," *IEEE Trans. on Sig. Proc.*, vol. 63, no. 3, pp. 711–726, 2015.

[10] S. Park, N. Y. Yu, and H. Lee, "An information-theoretic study for joint sparsity pattern recovery with different sensing matrices," *IEEE Trans. Inf. Theory*, vol. 63, no. 9, pp. 5559–5571, Sep. 2017.

[11] I. F. Pinelis and S. A. Utev, "Estimates of the moments of sums of independent random variables," *Theory of Probability & Its Applications*, vol. 29, no. 3, pp. 574–577, 1985. [Online]. Available: https://doi.org/10.1137/1129075

[12] L. Ramesh, C. R. Murthy, and H. Tyagi, "Sample-measurement tradeoff in support recovery under a subgaussian prior," in *2019 IEEE International Symposium on Information Theory (ISIT)*, July 2019, pp. 2709–2713.

[13] L. Ramesh, C. R. Murthy, and H. Tyagi, "Sample-measurement tradeoff in support recovery under a subgaussian prior," December 2019. [Online]. Available: http://arxiv.org/abs/1912.11247

[14] G. Reeves and M. Gastpar, "Sampling bounds for sparse support recovery in the presence of noise," in *IEEE International Symposium on Information Theory*, 2008, pp. 2187–2191.

[15] J. Scarlett and V. Cevher, "Limits on support recovery with probabilistic models: An information-theoretic framework," *IEEE Transactions on Information Theory*, vol. 63, no. 1, pp. 593–620, 2017.

[16] G. Tang and A. Nehorai, "Performance analysis for sparse support recovery," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1383–1399, 2010.

[17] M. J. Wainwright, "Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5728–5741, 2009.

[18] D. P. Wipf and B. D. Rao, "An empirical bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Trans. Signal Process.*, vol. 55, no. 7-2, pp. 3704–3716, 2007.