# On the Information-Theoretic Limits of Noisy Sparse Phase Retrieval

Lan V. Truong
Department of Computer Science
National University of Singapore
Email: truongvl@comp.nus.edu.sg

Jonathan Scarlett
Depts. Computer Science & Mathematics
National University of Singapore
Email: scarlett@comp.nus.edu.sg

*Abstract*—The support recovery problem consists of determining a sparse subset of variables that is relevant in generating a set of observations. In this paper, we study the support recovery problem in the phase retrieval model consisting of noisy phaseless measurements, which arises in a diverse range of settings such as optical detection, $X$-ray crystallography, electron microscopy, and coherent diffractive imaging. Our focus is on information-theoretic fundamental limits under an approximate recovery criterion, with Gaussian measurements and a simple discrete model for the sparse non-zero entries. Our bounds provide sharp thresholds with near-matching constant factors in several scaling regimes on the sparsity and signal-to-noise ratio.

## I. INTRODUCTION

Recently, there has been a growing interest in recovering an unknown signal $\beta \in \mathbb{C}^p$ from phaseless quadratic observations of the form $Y = |\langle \beta, X \rangle|^2 + Z$, where $X \in \mathbb{C}^p$ is a measurement matrix, and $Z \in \mathbb{R}$ represents measurement noise. Since only the magnitude of $\langle \beta, X \rangle$ is measured, and not the phase (or the sign, in the real case), this problem is referred to as *phase retrieval*. The phase retrieval problem has many applications including optical detection, $X$-ray crystallography, electron microscopy, and coherent diffractive imaging [1]. Similarly to the basic linear model, various works have shown that the number of measurements can be reduced significantly if the signal $\beta \in \mathbb{C}^p$ is *sparse*, i.e., it has at most $k$ non-zero entries for some $k \ll p$ [1]–[5].

A distinct goal that has received less attention in phase retrieval, but considerable attention in other models, is the support recovery problem [6]–[8], where one wishes to exactly or approximately determine the support $S = \text{supp}(\beta)$ given a collection of observations $\mathbf{Y} \in \mathbb{R}^n$ and the corresponding measurement matrix $\mathbf{X} \in \mathbb{C}^{n \times p}$. This problem is of direct interest when the goal is to find *which* variables influence the output (rather than their weights), and may also be used as a first step towards estimating the values of $\beta$ (e.g., see [9]).

Under general linear and non-linear models, Scarlett and Cevher [10] provided achievability and converse bounds characterizing the trade-off between error probability and number of measurements. They applied their general bounds to the linear, 1-bit, and group testing models to obtain precise thresholds on the number of measurements required to achieve vanishing decoding error probability in the high-dimensional limit. Numerous other related works also exist, with the focus being mainly on linear models [11]–[15]; see [10] for a more

detailed overview. In particular, *approximate recovery* criteria were studied by Reeves and Gastpar [16], [17] in the regime $k = \Theta(p)$, and by Scarlett and Cevher [10] in the regime $k = o(p)$; we focus on the latter setting.

Although the initial bounds in [10] are very general, applying these bounds to new models can still be very challenging, due to the need to establish concentration bounds and mutual information bounds on a case-by-case basis. In this paper, we use this approach to establish fundamental limits for approximate support recovery in the phase retrieval model, under a log-concavity assumption on the noise distribution. To achieve this goal, we need to overcome at least two key challenges: establishing concentration bounds for information quantities in the phase retrieval model, and upper and lower bounding key conditional mutual information terms that have no closed form expressions.

## II. PROBLEM SETUP AND MAIN RESULT

### A. Model and Assumptions

Let $p$ denote the ambient dimension, $k$ the sparsity level, and $n$ the number of measurements. We let $\mathcal{S}$ be the set of subsets of $\{1, 2, \ldots, p\}$ having cardinality $k$. The key random variables in the support retrieval problem are the support set $S \in \mathcal{S}$, the unknown signal $\beta \in \mathbb{C}^p$, the measurement matrix $\mathbf{X} \in \mathbb{C}^{n \times p}$, and the observation vector $\mathbf{Y} \in \mathbb{R}^n$.

The support set $S$ is assumed to be equiprobable on the $\binom{p}{k}$ subsets within $\mathcal{S}$. Given $S$, the entries of $\beta_{S^c}$ (with $S^c = \{1, \ldots, p\} \setminus S$) are deterministically set to zero, and the remaining entries are generated according to some (discrete) distribution $\beta_S \sim P_{\beta_S}$. In this paper, we focus on the case that $\beta_S$ is a uniformly random permutation of a fixed complex vector $(b_1, \ldots, b_k)$. In the full version [18], we consider a more general setup also permitting *continuous* distributions on $\beta_S$.

We consider the setting of (complex) Gaussian measurements, in which the measurement matrix takes i.i.d. values on $\mathcal{CN}(0, 1)$, whose density is denoted by $f_X$. We write $f_X^{n \times p}$, to denote the corresponding i.i.d. distribution for matrices, and we write $f_X^k$ as a shorthand for $f_X^{k \times 1}$. Given $S = s$, each entry of the observation vector $\mathbf{Y}$ is generated in a conditionally independent manner according to the following model:

$$Y = |\langle X_s, \beta_s \rangle|^2 + Z, \tag{1}$$

where $X_s \sim f_X^k$, $\beta_s \in \mathbb{C}^k$, and $Z \sim f_Z$, with $f_Z$ being an arbitrary *log-concave* density function. This log-concavity assumption is made for mathematical convenience, but also captures a wide range of noise distributions, including Gaussian. We note that the permutation-invariance of $Y$, $X_S$ and $\beta_S$ with respect to $S$ allows us to condition on a fixed $S = s$ throughout the analysis (e.g. $s = \{1, \ldots, k\}$) without loss of generality; such conditioning should henceforth be assumed unless explicitly stated otherwise.

The relation (1) induces the following conditional joint distribution of $(X_s, Y)$ given $S = s$ and $\beta_S = b_s$:

$$f_{X_s Y | \beta_s}(x_s, y | b_s) = f_X^k(x_s) f_{Y|X_s\beta_s}(y | x_s, b_s) \qquad (2)$$
$$= f_X^k(x_s) f_Z(y - |\langle x_s, b_s \rangle|^2), \qquad (3)$$

and its multiple-observation counterpart

$$f_{\mathbf{X}_s \mathbf{Y} | \beta_s}(\mathbf{x}_s, \mathbf{y} | b_s) = f_X^{n \times k}(\mathbf{x}_s) f_{Y|X_s\beta_s}^n(\mathbf{y} | \mathbf{x}_s, b_s), \qquad (4)$$

where $f_{Y|X_s\beta_s}^n(\mathbf{y} | \mathbf{x}_s, b_s)$ is the $n$-fold product of $f_{Y|X_s\beta_s}(\cdot | \cdot, b_s)$. The remaining entries of the measurement matrix are distributed as $\mathbf{X}_{s^c} \sim f_X^{n \times (p-k)}$.

Given $\mathbf{X}$ and $\mathbf{Y}$, a decoder forms an estimate $\hat{S}$ of $S$. Like previous works studying the information-theoretic limits of support recovery (e.g., [10], [11]), we assume that the decoder knows the system model, including $f_{Y|X_s\beta_s}$ and $P_{\beta_s}$. We focus on the *approximate recovery criterion*, only requiring that at least $k - \lfloor \alpha^* k \rfloor + 1$ entries of $S$ are successfully identified (approximate recovery) for some $\alpha^* \in (0,1)$. Following [10], [16], the error probability is given by

$$\mathsf{P}_{\mathrm{e}}(\alpha^*) := \mathbb{P}\Big[\{|S \setminus \hat{S}| \geq \lfloor \alpha^* k \rfloor\} \cup \{|\hat{S} \setminus S| \geq \lfloor \alpha^* k \rfloor\}\Big]. \quad (5)$$

Note that if both $S$ and $\hat{S}$ have cardinality $k$ with probability one, then the two events in the union are identical, and hence either of the two can be removed.

Our main goal is to derive necessary and sufficient conditions on $n$ (as a function of $k$ and $p$) such that $\mathsf{P}_{\mathrm{e}}(\alpha^*)$ tends to zero as $p \to \infty$. Moreover, when considering converse results, we will not only be interested in conditions under which $\mathsf{P}_{\mathrm{e}}(\alpha^*) \nrightarrow 0$, but also conditions under which the stronger statement $\mathsf{P}_{\mathrm{e}}(\alpha^*) \to 1$ holds.

### B. Statement of Main Result

Here we state and discuss the main result of this paper. An analogous main result is stated in [18] for the case that the non-zero entries of $\beta_s$ are Gaussian, but that is omitted here due to space constraints.

Recall that $\beta_s$ is a uniformly random permutation of a fixed complex vector $(b_1, \ldots, b_k)$. We let $(b_1', \ldots, b_k')$ be the sorted version of $(b_1, \ldots, b_k)$ such that $|b_1'| \leq |b_2'| \leq \cdots |b_k'|$, and define the following mutual information quantities:

$$I_1(\alpha, k) := \frac{1}{2} \log\left[\left(\frac{4}{\exp(2h(Z))}\right) v'(\alpha) + 1\right], \qquad (6)$$

$$I_2(\alpha, k) := \frac{1}{2} \log\left[\left(\frac{2\pi e}{\exp(2h(Z))}\right) v'(\alpha) + 1\right]$$
$$+ \frac{1}{2} \log\left[1 + \frac{v'(\alpha)\left(\|b_s\|^2 - v'(\alpha)\right)}{v'(\alpha)^2 + \frac{\exp(2h(Z))}{2\pi e}}\right] + \frac{1}{2} \log\left(\frac{\pi e}{2}\right), \quad (7)$$

where $v'(\alpha) = \left(\sum_{i=1}^{\lfloor \alpha k \rfloor} |b_i'|^2\right)^2$ is the energy in the lowest (in magnitude) $\lfloor \alpha k \rfloor$ entries of $(b_1, \ldots, b_k)$. Here and subsequently, all logarithms have base $e$, and information measures are in units of nats.

**Theorem 1.** *Consider the preceding phase retrieval setup, with $\beta_s$ being a uniformly random permutation of a fixed complex vector $(b_1, b_2, \ldots, b_k)$. Let $|b_{\min}| = \min\{|b_i| : i \in \{1, \cdots, k\}\}$ and $|b_{\max}| = \max\{|b_i| : i \in \{1, \cdots, k\}\}$, and assume that $|b_{\min}| = \Theta(|b_{\max}|)$, and that $k \to \infty$ with $\|b_s\|_2 = \Theta(1)$ as $p \to \infty$. In addition, assume that there are $m_\beta \in \{1, \ldots, k\}$ distinct elements in $(b_1, \ldots, b_k)$.*

*We have $\mathsf{P}_{\mathrm{e}}(\alpha^*) \to 0$ as $p \to \infty$ provided that*

$$n \geq \max_{\alpha \in [\alpha^*, 1]} \frac{\alpha k \log(\frac{p}{k})}{I_1(\alpha, k)}(1 + \eta) \qquad (8)$$

*for arbitrarily small $\eta > 0$ if either of the following additional conditions hold: (i) $m_\beta = \Theta(1)$ and $k = o(p)$, or (ii) $\log k = o(\log p)$ (and $m_\beta$ is arbitrary).*

*Conversely, under the general scaling $k = o(p)$ and arbitrary $m_\beta$, we have $\mathsf{P}_{\mathrm{e}}(\alpha^*) \to 1$ as $p \to \infty$ whenever*

$$n \leq \max_{\alpha \in [\alpha^*, 1]} \frac{(\alpha - \alpha^*)k \log(\frac{p}{k})}{I_2(\alpha, k)}(1 - \eta), \qquad (9)$$

*for arbitrarily small $\eta > 0$.*

We observe that the upper and lower bounds are nearly in closed form, other than the optimization over a single scalar $\alpha$. Moreover, the two have a very similar form, with the main difference being the appearance of $\alpha$ vs. $(\alpha - \alpha^*)$ in the numerator, and $I_1$ vs. $I_2$ in the denominator. The bounds hold for an arbitrary log-concave noise distribution $f_Z$.

Since the noise variance $\sigma^2$ is fixed and the measurement matrix has normalized $\mathcal{CN}(0,1)$ entries, the assumption $\|b\|_2 = \Theta(1)$ corresponds to the case that the signal-to-noise ratio (SNR) is constant. We observe that under this assumption, the upper and lower bounds provide matching $\Theta(k \log \frac{p}{k})$ behavior. Perhaps more significantly, in the high-SNR limit (i.e., $\|b\|_2 \to \infty$), we obtain nearly identical *constant factors*. To see this, it suffices to crudely lower bound $I_1(\alpha, k)$ by $\frac{1}{2} \log\left[\left(\frac{4}{\exp(2h(Z))}\right)\left(\lfloor \alpha k \rfloor |b_{\min}|^2\right)^2 + 1\right]$, and upper bound $I_2(\alpha, k)$ by $\frac{1}{2} \log\left[\left(\frac{2\pi e}{\exp(2h(Z))}\right)\left(\lfloor \alpha k \rfloor |b_{\max}|^2\right)^2 + 1\right] + \frac{1}{2} \log\left[1 + \frac{\lfloor \alpha k \rfloor k |b_{\max}|^4}{\lfloor \alpha k \rfloor^2 |b_{\min}|^4}\right] + \frac{1}{2} \log\left(\frac{\pi e}{2}\right)$. For any $\alpha$ bounded away from zero, since $|b_{\min}| = \Theta(|b_{\max}|)$, these both behave as $\log(k |b_{\min}|^2)(1 + o(1))$ as $\|b\|_2 \to \infty$ (or equivalently $k |b_{\min}|^2 \to \infty$), which implies that the maxima in (8) and
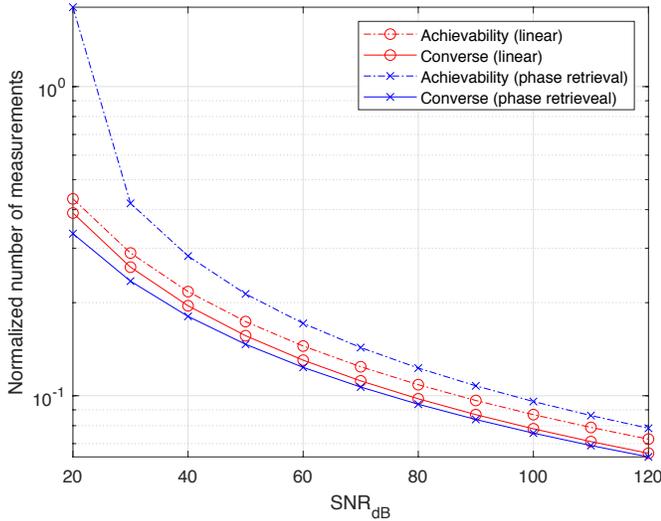
Fig. 1. Asymptotic thresholds on the number of measurements required for approximate support recovery for the linear model [10] and phase retrieval model with Gaussian noise, distortion level $\alpha^* = 0.1$, and non-zero entries $b_1 = \cdots = b_k = \sqrt{\frac{c_\beta}{k}}$. The asymptotic number of measurements is normalized by $k \log(\frac{p}{k})$, and $\mathrm{SNR}_{\mathrm{dB}} := 10 \log(2c_\beta^2)$.

(9) are attained by $\alpha = 1$ in this limit, and the upper and lower bounds coincide up to a factor of $\frac{1}{1-\alpha^*}$.

We believe that the extra assumptions on $m_\beta$ and $k$ in the achievability part are an artifact of our analysis, and note that similar assumptions were made for the linear model in [10].

### C. Comparison to the Linear Model

In Figure 1, we plot the upper and lower bounds of Theorem 1 for $\alpha^* = 0.1$ under various signal-to-noise ratios (SNRs), along with the counterparts for the linear model in [10]. We focus on the simple case that $Z \sim \mathcal{N}(0,1)$ and

$$b_1 = \cdots = b_k = \sqrt{\frac{c_\beta}{k}} \qquad (10)$$

for some $c_\beta > 0$, corresponding to $m_\beta = 1$ in Theorem 1. See [18, Appendix A] for a derivation of the SNR in the phase retrieval model, and a simplified statement of our sample complexity bounds when the non-zero entries of $\beta_s$ are given by (10).

As predicted by the discussion following Theorem 1, the upper and lower bounds are close (though still with a constant gap) when the SNR is sufficiently high. In addition, in this regime the information-theoretic limits of the phase retrieval model and the linear model are very similar.

However, at lower SNR, the gap for the phase retrieval model can widen more than that of the linear model. This appears to be because the key mutual information quantities arising in the analysis can only be expressed in closed form in the linear model, while requiring possibly-loose bounds in the phase retrieval model. However, all that is needed to close this gap (at least partially) is to deduce improved mutual information bounds for the phase retrieval setting (*cf.*, Section III-D below).

## III. OUTLINE OF THE PROOF

In this section, we outline the proof of Theorem 1; the details can be found in [18]. We first introduce some notation and recall the initial bounds for general observation models from [10], and then present the relevant log-concavity properties, mutual information bounds, and concentration bounds.

### A. Information-Theoretic Definitions

We first outline some information theoretic definitions from [10], recalling that we are conditioning on a fixed $S = s$ throughout. We consider partitions of the support set $s \in \mathcal{S}$ into two disjoint sets $s_{\mathrm{dif}} \neq \emptyset$ and $s_{\mathrm{eq}}$, where $s_{\mathrm{eq}}$ will typically correspond to an overlap between $s$ and some other set $\bar{s}$ (i.e., $s \cap \bar{s}$, the "equal" part), and $s_{\mathrm{dif}}$ will correspond to the indices in one set but not in the other (i.e., $s \setminus \bar{s}$, the "differing" part).

For fixed $s \in \mathcal{S}$ and a corresponding pair $(s_{\mathrm{dif}}, s_{\mathrm{eq}})$, we introduce the notation

$$f_{\mathbf{Y}|\mathbf{X}_{s_{\mathrm{dif}}}\mathbf{X}_{s_{\mathrm{eq}}}}(\mathbf{y}|\mathbf{x}_{s_{\mathrm{dif}}}, \mathbf{x}_{s_{\mathrm{eq}}}) := f_{\mathbf{Y}|\mathbf{X}_s}(\mathbf{y}|\mathbf{x}_s), \qquad (11)$$

$$f_{Y|X_{s_{\mathrm{dif}}}X_{s_{\mathrm{eq}}}}(y|x_{s_{\mathrm{dif}}}, x_{s_{\mathrm{eq}}}, b_s) := f_{Y|X_s\beta_s}(y|x_s, \beta_s), \quad (12)$$

where $f_{\mathbf{Y}|\mathbf{X}_s}$ is the marginal distribution of (4). While the left-hand sides of (11) and (12) represent the same quantities for any pair $(s_{\mathrm{dif}}, s_{\mathrm{eq}})$, it will still prove convenient to work with these in place of the right-hand sides. In particular, this allows us to introduce the marginal distributions

$$f_{\mathbf{Y}|\mathbf{X}_{s_{\mathrm{eq}}}}(\mathbf{y}|\mathbf{x}_{s_{\mathrm{eq}}}) := \sum_{\mathbf{x}_{s_{\mathrm{dif}}}} f_X^{n \times \ell}(\mathbf{x}_{s_{\mathrm{dif}}}) f_{\mathbf{Y}|\mathbf{X}_{s_{\mathrm{dif}}}\mathbf{X}_{s_{\mathrm{eq}}}}(\mathbf{y}|\mathbf{x}_{s_{\mathrm{dif}}}, \mathbf{x}_{s_{\mathrm{eq}}}),$$
$$(13)$$

$$f_{Y|X_{s_{\mathrm{eq}}}}(y|x_{s_{\mathrm{eq}}}) := \sum_{x_{s_{\mathrm{dif}}}} f_X^{\ell}(x_{s_{\mathrm{dif}}}) f_{Y|X_{s_{\mathrm{dif}}}X_{s_{\mathrm{eq}}}}(y|x_{s_{\mathrm{dif}}}, x_{s_{\mathrm{eq}}}),$$
$$(14)$$

where $\ell := |s_{\mathrm{dif}}|$. Using the preceding definitions, we introduce two information densities (in the terminology of the information theory literature, e.g., [19]). The first contains probabilities averaged over $\beta_s$,

$$i(\mathbf{x}_{s_{\mathrm{dif}}}; \mathbf{y}|\mathbf{x}_{s_{\mathrm{eq}}}) := \log \frac{f_{\mathbf{Y}|\mathbf{X}_{s_{\mathrm{dif}}}\mathbf{X}_{s_{\mathrm{eq}}}}(\mathbf{y}|\mathbf{x}_{s_{\mathrm{dif}}}, \mathbf{x}_{s_{\mathrm{eq}}})}{f_{\mathbf{Y}|\mathbf{X}_{s_{\mathrm{eq}}}}(\mathbf{y}|\mathbf{x}_{s_{\mathrm{eq}}})}, \quad (15)$$

whereas the second conditions on $\beta_s = b_s$:

$$i^n(\mathbf{x}_{s_{\mathrm{dif}}}; \mathbf{y}|\mathbf{x}_{s_{\mathrm{eq}}}, b_s) := \sum_{i=1}^n i(x_{s_{\mathrm{dif}}}^{(i)}; y^{(i)}|x_{s_{\mathrm{eq}}}^{(i)}, b_s), \qquad (16)$$

where $(x^{(i)}, y^{(i)})$ is the $i$-th measurement, and the single-letter information density is

$$i(x_{s_{\mathrm{dif}}}; y|x_{s_{\mathrm{eq}}}, b_s) := \log \frac{f_{Y|X_{s_{\mathrm{dif}}}X_{s_{\mathrm{eq}}}\beta_s}(y|x_{s_{\mathrm{dif}}}, x_{s_{\mathrm{eq}}}, b_s)}{f_{Y|X_{s_{\mathrm{eq}}}\beta_s}(y|x_{s_{\mathrm{eq}}}, b_s)}.$$
$$(17)$$

Averaging (17) with respect to the distribution in (12) conditioned on $\beta_s = b_s$ yields a conditional mutual information quantity, which is denoted by

$$I_{s_{\mathrm{dif}}, s_{\mathrm{eq}}}(b_s) := I(X_{s_{\mathrm{dif}}}; Y|X_{s_{\mathrm{eq}}}, \beta_s = b_s). \qquad (18)$$

## B. Initial Achievability and Converse Bounds

For the general support recovery problem with probabilistic models, the following achievability and converse bounds are given in [10]. Note that the versions of these bounds that we state here are simplified due to the fact that $\beta_s$ is a uniformly random permutation, so all (non-zero) values of $P_{\beta_s}(b_s)$ are identical. (The more general version requires defining a "typical set" of $\beta_s$ vectors.)

**Theorem 2.** *[10, Theorem 5] Fix any constants $\delta_1 > 0$ and $\delta_2 \in (0,1)$, and functions $\{\psi_\ell\}_{\ell=\lfloor \alpha^* k \rfloor}^{k} (\psi_\ell : \mathbb{Z} \times \mathbb{R} \to \mathbb{R})$ such that the following holds:*

$$\mathbb{P}\Big[ i^n(\mathbf{X}_{s_{\mathrm{dif}}}; \mathbf{Y} | \mathbf{X}_{s_{\mathrm{eq}}}, \beta_s) \le n(1-\delta_2) I_{s_{\mathrm{dif}}, s_{\mathrm{eq}}}(b_s) \,\big|\, \beta_s = b_s \Big]$$
$$\le \psi_{|s_{\mathrm{dif}}|}(n, \delta_2), \quad (19)$$

$$n \ge \frac{\log \binom{p-k}{|s_{\mathrm{dif}}|} + \log \left( \frac{k^2}{\delta_1^2} \binom{k}{|s_{\mathrm{dif}}|}^2 \right) + \log \frac{1}{\min_{b_s} P_{\beta_s}(b_s)}}{I_{s_{\mathrm{dif}}, s_{\mathrm{eq}}}(b_s)(1-\delta_2)}, \quad (20)$$

*for all $(s_{\mathrm{dif}}, s_{\mathrm{eq}})$ with $\lfloor \alpha^* k \rfloor \le |s_{\mathrm{dif}}| \le k$ and for all realizations $b_s$ of $\beta_s$. Then we have*

$$\mathsf{P}_{\mathrm{e}}(\alpha^*) \le \sum_{l=\lfloor \alpha^* k \rfloor}^{k} \binom{k}{\ell} \psi_\ell(n, \delta_2) + 2\delta_1. \quad (21)$$

**Theorem 3.** *[10, Theorem 6] Fix any constants $\delta_1 > 0, \delta_2 > 0$, and functions $\{\psi'_\ell\}_{\ell=\lfloor \alpha^* k \rfloor}^{k} (\psi'_\ell : \mathbb{Z} \times \mathbb{R} \to \mathbb{R})$ such that the following holds:*

$$\mathbb{P}\Big[ i^n(\mathbf{X}_{s_{\mathrm{dif}}}; \mathbf{Y} | \mathbf{X}_{s_{\mathrm{eq}}}, \beta_s) \le n(1+\delta_2) I_{s_{\mathrm{dif}}, s_{\mathrm{eq}}}(b_s) \,\big|\, \beta_s = b_s \Big]$$
$$\ge 1 - \psi'_{|s_{\mathrm{dif}}|}(n, \delta_2), \quad (22)$$

$$n \le \frac{\log \binom{p-k+|s_{\mathrm{dif}}|}{|s_{\mathrm{dif}}|} - \log \left( \sum_{d=0}^{\lfloor \alpha^* k \rfloor} \binom{p-k}{d} \binom{|s_{\mathrm{dif}}|}{d} \right) - \log \delta_1}{I_{s_{\mathrm{dif}}, s_{\mathrm{eq}}}(b_s)(1+\delta_2)},$$
$$(23)$$

*for all $(s_{\mathrm{dif}}, s_{\mathrm{eq}})$ with $|s_{\mathrm{dif}}| \in [\lfloor \alpha^* k \rfloor, k]$, and for all realizations $b_s$ of $\beta_s$. Then we have*

$$\mathsf{P}_{\mathrm{e}}(\alpha^*) \ge 1 - \max_{\ell=\lfloor \alpha^* k \rfloor, \ldots, k} \psi'_\ell(n, \delta_2) - \delta_1. \quad (24)$$

The steps for applying and simplifying these bounds are as follows:

1) Establish an explicit characterization of each mutual information term $I_{s_{\mathrm{dif}}, s_{\mathrm{eq}}}(b_s)$ (e.g., upper and lower bounds);
2) Use concentration of measure to find expressions for each function $\psi_\ell$ and $\psi'_\ell$ in Theorems 2 and 3, i.e., functions satisfying (19) and (22);
3) Combine and simplify the preceding steps to deduce the final sample complexity bound.

These steps turn out to be highly non-trivial in the phase retrieval setting; the details are outlined below.

## C. Log-Concavity Properties

Both our mutual information bounds and concentration bounds will crucially rely on the log-concavity properties stated in the following lemma.

**Lemma 4.** *Under the phase retrieval setup in Section II, we have the following:*

1) *Given $S = s$ and $\beta_s = b_s$, the conditional marginal density of $Y$ is log-concave;*
2) *Given $S = s$, $\beta_s = b_s$, and $X_{s_{\mathrm{eq}}} = x_{s_{\mathrm{eq}}}$ for some $s_{\mathrm{eq}} \subset s$, the conditional marginal density of $Y$ is log-concave.*

The proof (given in [18]) is very straightforward, and is based on the assumption that $f_Z$ is log-concave, along with well-known log-concavity properties.

## D. Mutual Information Bounds

While an exact expression for the mutual information $I_{s_{\mathrm{dif}}, s_{\mathrm{eq}}}(b)$ does not appear to be possible, the following theorem states closed-form upper and lower bounds. While there is a gap between the two in general, the asymptotic behavior is similar when $v_{\mathrm{dif}} = \sum_{i \in s_{\mathrm{dif}}} |b_i|^2$ grows large; this fact ultimately leads to tight sample complexity bounds in the high-SNR setting.

**Theorem 5.** *For the phase retrieval setup in Section II, the following holds for $I_{s_{\mathrm{dif}}, s_{\mathrm{eq}}}(b_s)$ defined in (18):*

$$\frac{1}{2} \log \left[ \left( \frac{4}{\exp(2h(Z))} \right) v_{\mathrm{dif}}^2 + 1 \right] \le I_{s_{\mathrm{dif}}, s_{\mathrm{eq}}}(b_s)$$
$$\le \frac{1}{2} \log \left( \frac{\pi e}{2} \right) + \frac{1}{2} \log \left[ \left( \frac{2\pi e}{\exp(2h(Z))} \right) v_{\mathrm{dif}}^2 + 1 \right]$$
$$+ \frac{1}{2} \log \left( 1 + \frac{v_{\mathrm{dif}} v_{\mathrm{eq}}}{v_{\mathrm{dif}}^2 + \frac{\exp(2h(Z))}{2\pi e}} \right), \quad (25)$$

*where $v_{\mathrm{eq}} = \sum_{i \in s_{\mathrm{eq}}} |b_i|^2$ and $v_{\mathrm{dif}} = \sum_{i \in s_{\mathrm{dif}}} |b_i|^2$.*

The upper bound is based on the entropy power inequality and the maximum entropy property of the Gaussian distribution, and the lower bound is based on (known) results that give nearly-matching lower bounds for log-concave random variables. The details are given in [18].

## E. Concentration Bounds

Perhaps the most technically challenging part of our analysis is to establish concentration bounds amounting to explicit expressions for $\psi_\ell$ and $\psi'_\ell$ in Theorems 2 and 3. These are stated in the following theorem.

**Theorem 6.** *Under the phase retrieval setup in Section II, the following bounds hold:*

$$\mathbb{P}\Big[ i^n(\mathbf{X}_{s_{\mathrm{dif}}}; \mathbf{Y} | \mathbf{X}_{s_{\mathrm{eq}}}, \beta_s = b_s) - n I_{s_{\mathrm{dif}}, s_{\mathrm{eq}}}(b_s) \le -2nC(b_s)\mu \Big]$$
$$\le \exp(-nC(b_s)r(\mu)) + \exp(-nC(b_s)r(-\mu)), \quad (26)$$
$$\mathbb{P}\Big[ i^n(\mathbf{X}_{s_{\mathrm{dif}}}; \mathbf{Y} | \mathbf{X}_{s_{\mathrm{eq}}}, \beta_s = b_s) - n I_{s_{\mathrm{dif}}, s_{\mathrm{eq}}}(b_s) \ge 2nC(b_s)\mu \Big]$$
$$\le \exp(-nC(b_s)r(\mu)) + \exp(-nC(b_s)r(-\mu)), \quad (27)$$

*for all $\mu > 0$, where $I_{s_{\mathrm{dif}},s_{\mathrm{eq}}}(b_s)$ is defined in (18), $C(b_s)$ is a constant depending on $b_s \in \mathbb{C}^k$, and*

$$r(u) = \begin{cases} u - \log(1+u) & \text{for} \quad -1 < u < \infty \\ +\infty & \text{otherwise.} \end{cases} \quad (28)$$

*In addition, we have $C = \Theta(1)$ whenever $\|b_s\|_2 = \Theta(1)$.*

We briefly discuss the ideas behind establishing this result. First note that

$$i^n(\mathbf{X}_{s_{\mathrm{dif}}}; \mathbf{Y}|\mathbf{X}_{s_{\mathrm{eq}}}, \beta_s = b_s) = \tilde{h}(\mathbf{Y}|\mathbf{X}_{s_{\mathrm{dif}}}, \mathbf{X}_{s_{\mathrm{eq}}}, \beta_s = b_s)$$
$$- \tilde{h}(\mathbf{Y}|\mathbf{X}_{s_{\mathrm{eq}}}, \beta_s = b_s), \quad (29)$$

where $\tilde{h}$ denotes the conditional negative log-density (e.g., $\tilde{h}(y|x) = -\log f_{Y|X}(y|x)$). Hence, to establish that $i^n$ concentrates, it suffices to show that both $\tilde{h}$ terms concentrate. The concentration of *unconditional* information random variables for log-concave distributions is well-known [20]. Unfortunately, such results cannot be used in our setting, because we are considering random variables that fail to be *jointly* log-concave. While we use similar high-level steps to [20] for establishing Theorem 6 based on the properties in Lemma 4 alone, this comes with considerable technical challenges. The interested reader is referred to [18] for the details.

*F. Wrapping Up*

Theorem 6 provides expressions for $\psi_\ell$ and $\psi'_\ell$ in Theorems 2 and 3, and allow us (with some tedious asymptotic analysis) to simplify the achievability condition on $n$ to

$$n \geq \max_{(s_{\mathrm{dif}},s_{\mathrm{eq}}) : \lfloor \alpha^* k \rfloor \leq |s_{\mathrm{dif}}| \leq k} \frac{\log\binom{p-k}{|s_{\mathrm{dif}}|}}{I_{s_{\mathrm{dif}},s_{\mathrm{eq}}}(b_s)}(1+\eta), \quad (30)$$

and the converse condition to

$$n \leq \max_{(s_{\mathrm{dif}},s_{\mathrm{eq}}) : \lfloor \alpha^* k \rfloor \leq |s_{\mathrm{dif}}| \leq k}$$
$$\frac{\log\binom{p-k+|s_{\mathrm{dif}}|}{|s_{\mathrm{dif}}|} - \log\left(\sum_{d=0}^{\lfloor \alpha^* k \rfloor} \binom{p-k}{d}\binom{|s_{\mathrm{dif}}|}{d}\right)}{I_{s_{\mathrm{dif}},s_{\mathrm{eq}}}(b_s)}(1-\eta) \quad (31)$$

for arbitrarily small $\eta > 0$. For the achievability part, these simplifications crucially use the assumptions on $k$ and $m_\beta$ stated in Theorem 2 to ensure that the term $\log \frac{1}{\min_{b_s} P_{\beta_s}(b_s)} \leq k \log m_\beta$ in (20) is asymptotically negligible.

To complete the achievability part, we use the mutual information lower bound in (25) of Theorem 5. Since this lower bound is increasing in $v_{\mathrm{dif}}$ and does not depend on $v_{\mathrm{eq}}$, we have the following whenever $|s_{\mathrm{dif}}| = \lfloor \alpha k \rfloor$:

$$I_{s_{\mathrm{dif}},s_{\mathrm{eq}}}(b_s) \geq I_1(\alpha, k), \quad (32)$$

recalling that $I_1(\alpha, k)$ defined in (6) replaces $v_{\mathrm{dif}}$ by the value corresponding to the lowest-magnitude entries of $b_s$. Hence, (8) of Theorem 1 follows from (30) by observing that the numerator of (30) behaves as $\left(\alpha k \log \frac{p}{k}\right)(1 + o(1))$ when $|s_{\mathrm{dif}}| = \alpha k$ (for $\alpha \in [\alpha^*, 1]$), and the denominator is lower bounded by $I_1(\alpha, k)$ via (32).

For the converse part, we use the upper bound in (25) of Theorem 5. While this bound depends on $v_{\mathrm{dif}}$ and $v_{\mathrm{eq}}$ in a more complicated fashion, the converse bound (31) remains valid when we replace the maximum over $(s_{\mathrm{dif}}, s_{\mathrm{eq}})$ by *any* fixed choice. Under the choice in which $s_{\mathrm{dif}}$ contains the indices corresponding to the $\lfloor \alpha k \rfloor$ entries of $b_s$ with the smallest magnitude, (25) yields

$$I_{s_{\mathrm{dif}},s_{\mathrm{eq}}}(b_s) \leq I_2(\alpha, k), \quad (33)$$

where $I_2(\alpha, k)$ is defined in (6). The numerator is already known to simplify to $(\alpha - \alpha^*)k \log(\frac{p}{k})(1 + o(1))$ whenever $|s_{\mathrm{dif}}| = \lfloor \alpha k \rfloor$ for $\alpha \in [\alpha^*, 1]$ [10], and combining these results yields (9) of Theorem 1.

## REFERENCES

[1] Y. C. Eldar and S. Mendelson, "Phase retrieval: Stability and recovery guarantees," *Applied and Computational Harmonic Analysis*, vol. 36, no. 3, pp. 473–494, 2014.

[2] M. Iwen, A. Viswanathan, and Y. Wang, "Robust sparse phase retrieval made easy," *Applied and Computational Harmonic Analysis*, vol. 42, no. 1, pp. 135 – 142, 2017.

[3] Y. M. Lu and M. Vetterli, "Sparse spectral factorization: Unicity and reconstruction algorithms," in *Proceedings of Int. Conf. Acoust. Speech, Signal Processing*, 2011, pp. 5976–5979.

[4] K. Jaganathan, S. Oymak, and B. Hassibi, "Sparse phase retrieval: Uniqueness guarantees and recovery algorithms," *IEEE Trans. Signal Process.*, vol. 65, no. 9, pp. 2402–2410, May 2017.

[5] Y. Li and V. Nakos, "Sublinear-time algorithms for compressive phase retrieval," in *Proc. of Intl. Symp. on Inform. Th.*, Vail, CO, 2018, pp. 2301–2305.

[6] M. B. Malyutov, "Search for sparse active inputs: A review," *Inf. Theory, Comb. and Search Theory*, pp. 609–647, 2013.

[7] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Springer New York, 2013.

[8] A. Miller, *Subset Selection in Regression*. Chapman Hall, 1990.

[9] G. Wang, L. Zhang, G. B. Giannakis, M. Akçakaya, and J. Chen, "Sparse phase retrieval via truncated amplitude flow," *arXiv preprint arXiv:1611.07641*, 2016.

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

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

[12] W. Wang, M. Wainwright, and K. Ramchandran, "Information theoretic bounds on model selection for gaussian markov random fields," in *Proc. of Intl. Symp. on Inform. Th.*, Austin, Texas, 2010.

[13] K. R. Rad, "Nearly sharp sufficient conditions on exact sparsity pattern recovery," in *Proc. of Intl. Symp. on Inform. Th.*, Austin, Texas, 2011, pp. 4672–4679.

[14] A. Tulino, G. Caire, S. Verdú, and S. Shamai, "Support recovery with sparsely sampled free random matrices," *IEEE Trans. on Inform. Th.*, vol. 59, no. 7, pp. 427–439, 2013.

[15] J. Scarlett, J. Evans, and S. Dey, "Compressed sensing with prior information: Information-theoretic limits and practical decoders," *IEEE Trans. Sig. Proc.*, vol. 61, no. 2, pp. 427–439, 2013.

[16] G. Reeves and M. Gastpar, "The sampling rate-distortion tradeoff for sparsity pattern recovery in compressed sensing," *IEEE Trans. on Inform. Th.*, vol. 58, no. 5, pp. 3065–3092, May 2012.

[17] ——, "Approximate sparsity pattern recovery: Information-theoretic lower bounds," *IEEE Trans. on Inform. Th.*, vol. 59, no. 6, pp. 3451–3465, June 2013.

[18] L. V. Truong and J. Scarlett, "Support recovery in the phase retrieval model: Information-theoretic fundamental limits," https://arxiv.org/abs/1901.10647.

[19] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. on Inform. Th.*, vol. 56, no. 5, pp. 2307–2359, 2010.

[20] M. Fradelizi, M. Madiman, and L. Wang, "Optimal concentration of information content for log-concave densities," *in High Dim. Prob. VII*, vol. 71, pp. 45–60, 2016.