

A flexible convex relaxation for phase retrieval*

Sohail Bahmani[†] and Justin Romberg[†]

*School of Electrical and Computer Engineering,
Georgia Institute of Technology,
Atlanta, GA 30332-0250*

e-mail: sohail.bahmani@ece.gatech.edu; jrom@ece.gatech.edu

Abstract: We propose a flexible convex relaxation for the phase retrieval problem that operates in the natural domain of the signal. Therefore, we avoid the prohibitive computational cost associated with “lifting” and semidefinite programming (SDP) in methods such as *PhaseLift* and compete with recently developed non-convex techniques for phase retrieval. We relax the quadratic equations for phaseless measurements to inequality constraints each of which representing a symmetric “slab”. Through a simple convex program, our proposed estimator finds an extreme point of the intersection of these slabs that is best aligned with a given *anchor vector*. We characterize geometric conditions that certify success of the proposed estimator. Furthermore, using classic results in statistical learning theory, we show that for random measurements the geometric certificates hold with high probability at an optimal sample complexity. We demonstrate the effectiveness of the proposed method through numerical simulations using both independent random measurements and coded diffraction patterns.

We also extend this formulation to include *sparsity constraints* on the target vector. With this additional constraint, we show that, considering “nested” measurements, the number of phaseless measurements needed to recover the sparse vector is essentially the same (to within a logarithmic factor) as the number of linear measurements needed by standard compressive sensing techniques.

MSC 2010 subject classifications: Primary 62F10, 90C25; secondary 63P30.

Keywords and phrases: Phase retrieval, Vapnik-Chervonenkis theory, convex relaxation.

Received June 2017.

1. Introduction

Let $\mathbf{x}_\star \in \mathbb{C}^N$ be a signal that we would like to recover from noisy phaseless measurements

$$b_i = |\mathbf{a}_i^* \mathbf{x}_\star|^2 + \xi_i \quad i = 1, 2, \dots, M, \quad (1)$$

*Part of this manuscript appeared in Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS) 2017.

[†]The authors were supported in part by ONR grant N00014-11-1-0459, NSF grants CCF-1415498 and CCF-1422540, and the Packard Foundation.

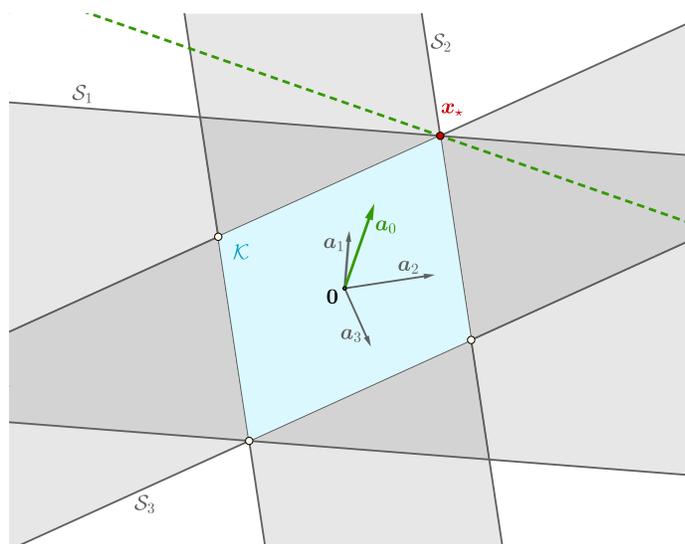


FIG 1. A two-dimensional illustration of slabs intersecting at \mathbf{x}_*

where the measurement vectors $\mathbf{a}_i \in \mathbb{C}^N$ are given. To solve this *phase retrieval* problem with provable accuracy, different methods that rely on semidefinite relaxation have been proposed previously [e.g., 8, 7, 36]. While these methods are guaranteed to produce an accurate solution in polynomial time, they are not scalable due to the use of semidefinite programming (SDP). This drawback of SDP-based methods has motivated development of alternative non-convex methods that operate in the natural domain of the signal and exhibit better scalability [e.g., 25, 10]. In this paper we follow a completely different approach and propose a convex relaxation of the phase retrieval problem that not only produces accurate solutions but also is scalable. Compared to the non-convex phase retrieval methods our approach inherits the flexibility of convex optimization both in analysis and application. In Section 3.3, we demonstrate this flexibility by showing how sparsity constraints can (almost seamlessly) be incorporated into the recovery algorithm.

The geometric idea at the core of our proposed method is the following. Relaxing each measurement equation in (1) to an inequality $|\mathbf{a}_i^* \mathbf{x}_*|^2 \leq b_i$ creates a symmetric slab \mathcal{S}_i of feasible solutions as illustrated in Figure 1. Collectively, these slabs describe a “complex polytope” \mathcal{K} of feasible solutions. In the noiseless regime (i.e., $\xi_i = 0$ for all i), the target signal \mathbf{x}_* would be one of the extreme points of \mathcal{K} . To distinguish \mathbf{x}_* among all of the extreme points, our idea is to find a hyperplane tangent to \mathcal{K} at \mathbf{x}_* . The crucial ingredient in this approach is an “anchor” vector $\mathbf{a}_0 \in \mathbb{C}^N \setminus \{\mathbf{0}\}$ that acts as the normal for the desired tangent hyperplane and it is required to have a non-vanishing correlation with \mathbf{x}_* in the sense that

$$|\mathbf{a}_0^* \mathbf{x}_*| \geq \delta \|\mathbf{a}_0\|_2 \|\mathbf{x}_*\|_2, \tag{2}$$

for some absolute constant $\delta \in (0, 1)$. The above geometric intuition is explained in more detail in Section 3.1. While our main result simply assumes that the anchor vector is given by an oracle, which may use the existing measurements, we discuss in Section 1.1 some realistic scenarios where a valid anchor vector exists or can be computed.

We assume that the noise is non-negative (i.e., $\xi_i \geq 0$) and we have $\|\boldsymbol{\xi}\|_\infty \leq \eta^{-1} \|\mathbf{x}_\star\|_2^2$ for some constant $\eta > 0$. Note that the non-negativity of the noise can be dropped at the cost of a slight reduction in the effective signal-to-noise ratio. In particular, one can add the noise upperbound (i.e., $\eta^{-1} \|\mathbf{x}_\star\|_2^2$) to each measurement to ensure the non-negativity. Throughout we treat \mathbb{C}^N as an inner-product space over \mathbb{R} equipped with the symmetric inner-product

$$\langle \cdot, \cdot \rangle : (\mathbf{x}_1, \mathbf{x}_2) \mapsto \operatorname{Re}(\mathbf{x}_1^* \mathbf{x}_2).$$

Clearly, in this setting \mathbb{C}^N will be a $2N$ -dimensional vector space.

With these assumptions in place, we propose the solution to the convex program¹

$$\begin{aligned} \max_{\mathbf{x}} \quad & \langle \mathbf{a}_0, \mathbf{x} \rangle \\ \text{subject to} \quad & |\mathbf{a}_i^* \mathbf{x}|^2 \leq b_i \quad 1 \leq i \leq M, \end{aligned} \tag{3}$$

as a computationally efficient estimator for \mathbf{x}_\star . Of course, the points equal to \mathbf{x} up to a global phase, namely,

$$\mathbb{T}\mathbf{x} \stackrel{\text{def}}{=} \{\omega \mathbf{x} : |\omega| = 1\},$$

yield the same phaseless measurements. Therefore, the goal is merely to estimate a point in $\mathbb{T}\mathbf{x}_\star$ accurately from the phaseless measurements (1).

In Lemma 2, below in Section 3, we establish a geometric condition that is sufficient to guarantee accurate estimation of \mathbf{x}_\star via the convex program (3).

The sufficient condition given by Lemma 2 can be interpreted in terms of (non-)existence of a particularly constrained halfspace that includes all of the points $\mathbf{a}_i \mathbf{a}_i^* \mathbf{x}_\star$. For random measurement vectors \mathbf{a}_i , this interpretation resembles the model and theory of *linear classifiers* studied in *statistical learning theory*, albeit in an unusual regime. Borrowing classic results from this area (summarized in Appendix A), we show that with high probability (3) produces an accurate estimate of \mathbf{x}_\star .

Specifically, in our main result, Theorem 1 in Section 3, we show that drawing

$$M \gtrsim N + \log \frac{1}{\varepsilon},$$

i.i.d. random measurements, with the hidden constant factor on the right-hand side depending on δ , would suffice for the conditions of Lemma 2 to hold with probability $\geq 1 - \varepsilon$. Consequently, solution $\hat{\mathbf{x}}$ of (3) would obey

$$\|\hat{\mathbf{x}} - \mathbf{x}_\star\|_2 \lesssim \eta^{-1} \|\mathbf{x}_\star\|_2.$$

¹In the real case, (3) reduces to a linear program.

1.1. Choosing the anchor vector

Our approach critically depends on the choice of the anchor vector \mathbf{a}_0 that obeys an inequality of the form (2). Below we discuss two interesting scenarios where such a vector would be accessible.

Non-negative signals: Perhaps the simplest scenario is when the target signal \mathbf{x}_* is known to be real and non-negative. In usual imaging modalities these model assumptions are realistic as natural images are typically represented by pixel intensities. For these types of signals we can choose $\mathbf{a}_0 = \frac{1}{\sqrt{N}}\mathbf{1}$ for which we obtain $|\mathbf{a}_0^* \mathbf{x}_*| = \|\mathbf{x}_*\|_1 / \sqrt{N}$. Then, for (2) to hold it suffices that $\|\mathbf{x}_*\|_1 \geq \delta \sqrt{N} \|\mathbf{x}_*\|_2$ for some absolute constant $\delta \in (0, 1)$. In particular, we need \mathbf{x}_* to have at least $\delta^2 N$ non-zero entries.

Random measurements: A more interesting scenario is when we can construct the vector \mathbf{a}_0 from the (random) measurements. An effective strategy is to set \mathbf{a}_0 to be the principal eigenvector of the matrix $\Sigma = \frac{1}{M} \sum_{i=1}^M b_i \mathbf{a}_i \mathbf{a}_i^*$. The principal eigenvector of Σ and its “truncated” variants have been used previously for initialization of the *Wirtinger Flow* algorithm [10] and its refined versions [11, 42]. For example, the following result is shown in [10, Section VII.H].

Lemma 1 (Candès et al. [10]). *For $1 \leq i \leq M$ let b_i be the phaseless measurements obtained from i.i.d. vectors $\mathbf{a}_i \sim \text{Normal}(\mathbf{0}, \frac{1}{2}\mathbf{I}) + i\text{Normal}(\mathbf{0}, \frac{1}{2}\mathbf{I})$ and no noise. If $M \stackrel{\delta}{\gtrsim} N \log N$ and \mathbf{a}_0 is the principal eigenvector of*

$$\Sigma = \frac{1}{M} \sum_{i=1}^M b_i \mathbf{a}_i \mathbf{a}_i^*,$$

then (2) holds with probability $\geq 1 - O(N^{-2})$.

While Lemma 1 can be refined or extended in various ways, we do not pursue these paths in this paper. We refer the interested reader to [23] which studies certain generalizations for Lemma 1. The only important extension to Lemma 1 for us is to consider the effect of noise. Here we use a simple argument based on eigenvalue perturbation. If b_i is given by (1) then we compute $\tilde{\mathbf{v}}$ a unit principal eigenvector of

$$\tilde{\Sigma} = \frac{1}{M} \sum_{i=1}^M |\mathbf{a}_i^* \mathbf{x}_*|^2 \mathbf{a}_i \mathbf{a}_i^* + \frac{1}{M} \sum_{i=1}^M \xi_i \mathbf{a}_i \mathbf{a}_i^*,$$

as the anchor. Let \mathbf{v} denote a unit principal eigenvector of

$$\Sigma = \frac{1}{M} \sum_{i=1}^M |\mathbf{a}_i^* \mathbf{x}_*|^2 \mathbf{a}_i \mathbf{a}_i^*,$$

that is aligned with $\tilde{\mathbf{v}}$ (i.e. $\mathbf{v}^* \tilde{\mathbf{v}} > 0$). Then, it follows from a variant of the Davis-Kahan theorem [41, Corollary 1] that

$$\|\tilde{\mathbf{v}} - \mathbf{v}\|_2 \lesssim \frac{\left\| \frac{1}{M} \sum_{i=1}^M \xi_i \mathbf{a}_i \mathbf{a}_i^* \right\|}{\|\mathbf{x}_*\|_2^2}.$$

Depending on the noise model we may show that $\left\| \frac{1}{M} \sum_{i=1}^M \xi_i \mathbf{a}_i \mathbf{a}_i^* \right\|$ is sufficiently small (e.g., using matrix concentration inequalities). In that case, if for instance $\|\tilde{\mathbf{v}} - \mathbf{v}\|_2 \leq \delta$, and $|\mathbf{v}^* \mathbf{x}_*| \geq 2\delta \|\mathbf{x}_*\|_2$ by Lemma 1, then we have

$$\begin{aligned} |\tilde{\mathbf{v}}^* \mathbf{x}_*| &= |(\tilde{\mathbf{v}} - \mathbf{v} + \mathbf{v})^* \mathbf{x}_*| \geq |\mathbf{v}^* \mathbf{x}_*| - |(\tilde{\mathbf{v}} - \mathbf{v})^* \mathbf{x}_*| \\ &\geq |\mathbf{v}^* \mathbf{x}_*| - \|\tilde{\mathbf{v}} - \mathbf{v}\|_2 \|\mathbf{x}_*\|_2 \\ &\geq (2\delta - \delta) \|\mathbf{x}_*\|_2 = \delta \|\mathbf{x}_*\|_2, \end{aligned}$$

which guarantees that $\tilde{\mathbf{v}}$ is also a valid anchor vector.

1.2. Sparse phase retrieval using an ℓ_1 regularizer

One of the benefits of posing the phase retrieval problem as a convex program in the natural domain of the signal is that the formulation is extensible. If we have additional information about the signal to be recovered that can be expressed using convex constraints and/or penalties, then the recovery procedure can take advantage of this information while remaining computationally tractable.

A natural model for many imaging applications is that the target vector \mathbf{x}_* is *sparse*. Here, we encourage the solution to be sparse by enforcing a constraint on the ℓ_1 norm:

$$\begin{aligned} &\max_{\mathbf{x}} \langle \mathbf{a}_0, \mathbf{x} \rangle \\ &\text{subject to } |\mathbf{a}_i^* \mathbf{x}|^2 \leq b_i \quad 1 \leq i \leq M \\ &\quad \|\mathbf{x}\|_1 \leq \|\mathbf{x}_*\|_1, \end{aligned} \tag{4}$$

For our analysis below, we make the simplifying assumption that $\|\mathbf{x}_*\|_1$ is known. In practice, the problem can always be put into Lagrange form, with an appropriate multiple of $\|\mathbf{x}\|_1$ subtracted from the functional $\langle \mathbf{a}_0, \mathbf{x} \rangle$.

As in the case of ordinary phase retrieval, a complete analysis of (4) requires (i) a rigorous method to construct the anchor, and (ii) a sample complexity analysis given the anchor vector. For generic Gaussian measurements, there are obstacles that make a tight mathematical analysis of these procedures complicated. To keep things simple, we will introduce some simple structure on the measurement vectors \mathbf{a}_i ; this structure will make the sample complexity analysis a relatively straightforward extension of the ordinary case, and will give us a concrete procedure for constructing an anchor vector from a nearly optimal number of samples. We note, though, that a near-optimal sample complexity analysis is possible using an analysis based on *Rademacher complexity* (see, e.g., the general framework in [3]) in place of the VC-based analysis in Section 3. With generic measurements, the main obstacle currently is construction of an anchor with optimal sample complexity. We are not aware of any computationally efficient construction that improves on the suboptimal sample complexity that is

quadratic in $\|\mathbf{x}_*\|_0$, but such constructions may exist (see, e.g., the discussion in [32]).

In this paper, we consider measurement vectors that have a “nested structure”, a model proposed in [20, 1, 39]. Specifically, we have

$$\mathbf{a}_i = \Psi^* \mathbf{w}_i \quad 1 \leq i \leq M, \tag{5}$$

where $\mathbf{w}_i \in \mathbb{C}^{M_0}$ are i.i.d. standard complex Gaussian vectors, and $\Psi \in \mathbb{C}^{N \times M_0}$ is a (random) matrix satisfying the restricted isometry property (RIP), i.e., for some constant $\gamma_K \in (0, 1)$ we have

$$(1 - \gamma_K) \|\mathbf{x}\|_2^2 \leq \|\Psi^* \mathbf{x}\|_2^2 \leq (1 + \gamma_K) \|\mathbf{x}\|_2^2, \tag{6}$$

for all K -sparse vectors \mathbf{x} . These conditions are usually fulfilled for random Ψ [4] when

$$M_0 \gtrsim K \log \frac{N}{K}.$$

Our second main result, stated precisely as Theorem 2 in Section 3.3 below, shows that if we take noisy phaseless measurements of a K_* -sparse vector \mathbf{x}_* of the form (5) and we have an anchor vector in place that satisfies (2), then solving (4) when

$$M \gtrsim K_* \log \frac{N}{K_*}, \tag{7}$$

produces (with high probability) an estimate $\hat{\mathbf{x}}$ that again obeys $\|\hat{\mathbf{x}} - \mathbf{x}_*\|_2 \lesssim \eta^{-1} \|\mathbf{x}_*\|_2$. Thus the number of phaseless nested measurements needed to recover \mathbf{x}_* is of the same order as if we had taken linear measurements $\langle \mathbf{a}_i, \mathbf{x}_* \rangle$ and used standard compressive sensing techniques to form an estimate.

The nested measurements also give us a technique for forming a data-driven anchor vector. To do this, we start by forming a matrix similar to the one used for the ordinary case above, but formed from the \mathbf{w}_i instead of the \mathbf{a}_i . We take

$$\Sigma = \frac{1}{M} \sum_{i=1}^M b_i \mathbf{w}_i \mathbf{w}_i^*,$$

compute the principal eigenvector \mathbf{w}_0 of Σ , and then form the anchor vector \mathbf{a}_0 by hard thresholding $\Psi \mathbf{w}_0$, retaining only its K_* largest terms. In Section 3.3, we show that when

$$M \gtrsim M_0 \log M_0,$$

this procedure results in an anchor vector that is sufficiently correlated with \mathbf{x}_* . Thus it appears we merely pay an extra log factor over the number of measurements in the sample complexity bound (7).

1.3. Other variations and extensions

There are several other ways the method proposed above can be extended. We will briefly discuss some of these here, and leave further exploration for

future research. While the core geometric ideas still apply to these extensions, significant modifications of our theoretical arguments would be necessary to analyze these extensions in full.

The gross noise model considered in this paper can be pessimistic in scenarios where we have random noise or deterministic noise with a different type of bound. In these scenarios, augmenting the estimator by a noise regularization term could result in accuracy bounds that gracefully vary with the considered noise.

Another interesting extension to the proposed method, is to adapt the current theory to the case of blockwise independent measurements as in coded diffraction imaging. Our numerical experiments in Section (2) suggest that the proposed method still performs well with these structured measurements. Nevertheless, to extend the analysis we may need to revise the current simple arguments based on Vapnik-Chervonenkis theory using more sophisticated tools from the theory of empirical processes.

Finally, it is possible to extend our convex formulation to the more general problem of recovering a low-rank positive semidefinite matrix from rank-one projections. Suppose that an $N \times N$ rank- R positive semidefinite matrix $\mathbf{S}_\star = \mathbf{X}_\star \mathbf{X}_\star^\top$ is observed through measurements of the form

$$\begin{aligned} b_i &= \mathbf{a}_i^\top \mathbf{S}_\star \mathbf{a}_i + \xi_i \\ &= \|\mathbf{X}_\star^\top \mathbf{a}_i\|_2^2 + \xi_i, \end{aligned} \quad i = 1, 2, \dots, M, \quad (8)$$

where the \mathbf{a}_i are measurement vectors and the ξ_i represent measurement noise — instead of measuring the magnitude of an inner product, we are measuring the norm of a matrix-vector multiplication. One application of measurements of this type is the problem of covariance sketching [6, 12, 2]. Similar to (3), we can estimate \mathbf{X}_\star via the convex program

$$\begin{aligned} \max_{\mathbf{X}} \quad & \text{tr}(\mathbf{A}_0^\top \mathbf{X}) \\ \text{subject to} \quad & \|\mathbf{X}^\top \mathbf{a}_i\|_2^2 \leq b_i \quad i = 1, 2, \dots, M, \end{aligned} \quad (9)$$

where $\mathbf{A}_0 \in \mathbb{R}^{N \times R}$ is an anchor matrix that is appropriately correlated with \mathbf{X}_\star ,

$$|\text{tr}(\mathbf{A}_0^\top \mathbf{U}_\star)| \geq \delta \|\mathbf{A}_0\|_F, \quad (10)$$

where \mathbf{U}_\star is a matrix that has orthogonal columns, each with ℓ_2 -norm $\frac{1}{\sqrt{R}}$, that span the column space of \mathbf{X}_\star .

1.4. Related work

There is a large body of research on phase retrieval addressing various aspect of the problem (see [21] and references therein). However, we focus only on the relevant results mostly developed in recent years. Perhaps, among the most

important developments are *PhaseLift* and similar methods that cast the phase retrieval problem as a particular semidefinite program [8, 7, 36]. The main idea used in [8] and [7] is that by *lifting* the unknown signal using the transformation $\mathbf{x}\mathbf{x}^* \mapsto \mathbf{X}$, the (noisy) phaseless measurements (1) that are quadratic in \mathbf{x} , can be converted to linear measurements of the rank-one positive semidefinite matrix $\mathbf{X}_* = \mathbf{x}_*\mathbf{x}_*^*$. With this observation, these SDP-based methods aim to solve the corresponding linear equations using the trace-norm to induce the rank-one structure in the solution. Inspired by the well-known convex relaxation of Max-Cut problem, *PhaseCut* method [36] considers the measurement phases as the unknown variables and applies a similar lifting transform to formulate a different semidefinite relaxation for phase retrieval. While these SDP-based methods are shown to produce accurate estimates of \mathbf{X}_* at optimal sample complexity for certain random measurement models, they become computationally prohibitive in medium- to large-scale problems where SDP is practically inefficient.

More recently, there has been a growing interest in non-convex iterative methods for phase retrieval [see e.g., 25, 10, 30, 11, 42, 37, 33]. These methods typically operate in the natural space of the signal and thus do not suffer the drawbacks of the SDP-based methods. In [25], using a specifically chosen initialization, some accuracy guarantees for a variant of the classic methods of Gerchberg and Saxton [15] and Fienup [14] are provided. This *alternating minimization* method of [25] basically updates an estimate of the signal assuming an estimate of the measurements' phase and vice versa in an alternating fashion through several iterations. The established sample complexity for this approach is (nearly) optimal in the dimension of the target signal, but it does not vary gracefully with the prescribed precision. Phase retrieval via the *Wirtinger Flow* (WF), a non-convex gradient descent method at core, is proposed in [10]. It is shown that for random measurements that have Normal distribution or certain *coded diffraction patterns*, with an appropriate initialization the WF iterates exhibit the linear rate of convergence to the target signal. More recent work on the WF method introduce better initialization by excluding the outlier measurements and achieve the optimal sample complexity [11, 42]. The WF class of algorithms and our proposed method both achieve optimal sample complexity (up to the constant factor) and have low computational cost. However, the WF methods need careful tuning of a step size parameter and their convergence analysis often relies on Gaussian measurements. This is partly because establishing robustness of non-convex methods generally requires stronger conditions. Our method provably works for a broader set of measurement distributions, has no tuning parameters, and can be implemented in various convex optimization software.

Shortly after a draft of this manuscript was first posted online, a few independent papers proposed and analyzed the same method and its variants. In [16], which refers to (3) as *PhaseMax*, a sharper constants in the sample complexity is obtained by assuming a stronger condition that the anchor is independent of the measurements in the analysis. Alternative proofs and variations that rely on matrix concentration inequalities appeared later in [19, 18, 17]. Another distinctive feature of our analysis compared to the mentioned results is that it is less sensitive to measurement distribution as it relies on VC-type bounds.

Until very recently, convex programming approaches to sparse phase retrieval considered some form of semidefinite relaxation combined with ℓ_1 -regularization to estimate the lifted sparse and rank-one target matrix [26, 22]. With standard Gaussian measurements, these methods are shown to require $M = O(K_\star^2)$ measurements to produce accurate solutions [22]. This suboptimal sample complexity is observed more generally in estimation of simultaneously sparse and low-rank matrices using naïve convex relaxations [27]. The prohibitive computational cost of (sparse) phase retrieval methods based on semidefinite programming, motivated the study of iterative non-convex approaches. Among the non-convex methods with provable guarantees, methods based on alternating minimization [25] and gradient descent [38] are shown to produce accurate solutions for sparse phase retrieval with suboptimal sample complexity $M = \tilde{O}(K_\star^2)$.

While achieving optimal sample complexity for sparse phase retrieval with generic Gaussian measurements is still open, a few techniques that exploit specially structured measurements are shown to reach optimal sample complexity with a computationally efficient algorithm. Examples of these methods include [28, 40] that construct measurements based on ideas from channel coding, as well as [20, 1, 39] that rely on “nested” measurements that are restricted to a low-dimensional subspace.

More recently, it is shown in [32] that, given a good initialization, the non-convex projected gradient descent with sparsity constraint can produce accurate solution to sparse phase retrieval with $M = O(K_\star \log(N/K_\star))$ Gaussian measurements. Furthermore, [17] considers the ℓ_1 -regularized form of (4) in the real-valued and noiseless scenario and shows that it achieves optimal sample complexity assuming the anchor vector is given. We emphasize that our general framework developed in [3] can establish optimal sample complexity for the ℓ_1 -regularized variant of (4), but it relies on tools other than VC-dimension to control the sample complexity.

2. Numerical experiments

We evaluated the performance of our proposed method on synthetic data with the target signal $\mathbf{x}_\star \sim \text{Normal}(\mathbf{0}, \frac{1}{2}\mathbf{I}) + \nu \text{Normal}(\mathbf{0}, \frac{1}{2}\mathbf{I})$ and measurements $\mathbf{a}_i \stackrel{\text{i.i.d.}}{\sim} \text{Normal}(\mathbf{0}, \frac{1}{2}\mathbf{I}) + \nu \text{Normal}(\mathbf{0}, \frac{1}{2}\mathbf{I})$ all having $N = 500$ coordinates. The noisy measurements follow (1) with the uniform noise $\xi_i \stackrel{\text{i.i.d.}}{\sim} \text{Uniform}([0, \eta^{-1}])$ in one experiment and the Gaussian noise $\xi_i \stackrel{\text{i.i.d.}}{\sim} \text{Normal}(0, \sigma^2)$ in the other. For the latter noise model we replaced any negative b_i by $b_i = 0$ to avoid negative measurements and defined the input signal-to-noise ratio as $\text{SNR} \stackrel{\text{def}}{=} 10 \log_{10} \frac{\|\mathbf{x}_\star\|_2^4}{\sigma^2}$. The vector \mathbf{a}_0 is constructed as in initialization of the Wirtinger Flow mentioned in Lemma 1 through 50 iterations of the power method. We implemented the convex program (3) by TFOCS [5] with smoothing parameter $\mu = 2 \times 10^{-3}$ and at most 500 iterations. Figure 2 illustrates the 0.9-quantile and median of the relative error $\min_{\phi \in [0, 2\pi)} \|\hat{\mathbf{x}} - e^{i\phi} \mathbf{x}_\star\|_2 / \|\mathbf{x}_\star\|_2$ observed over 100 trials of our algorithm for different sampling ratios M/N between 2 and 17. The plots also

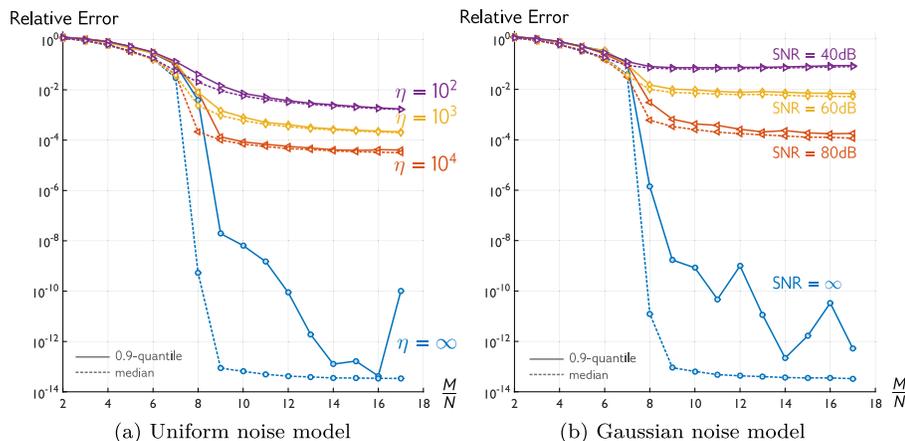


FIG 2. Empirical relative error of the proposed method at different sampling ratios (i.e., M/N) and noise levels with $N = 500$

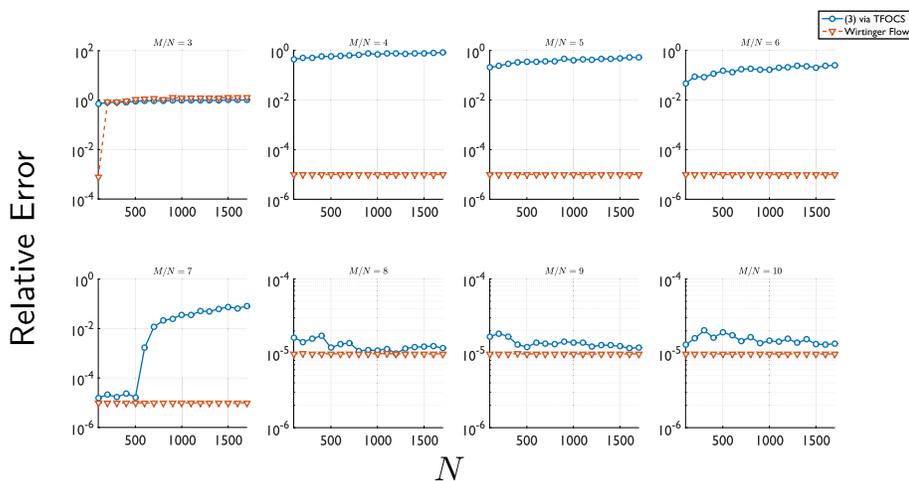
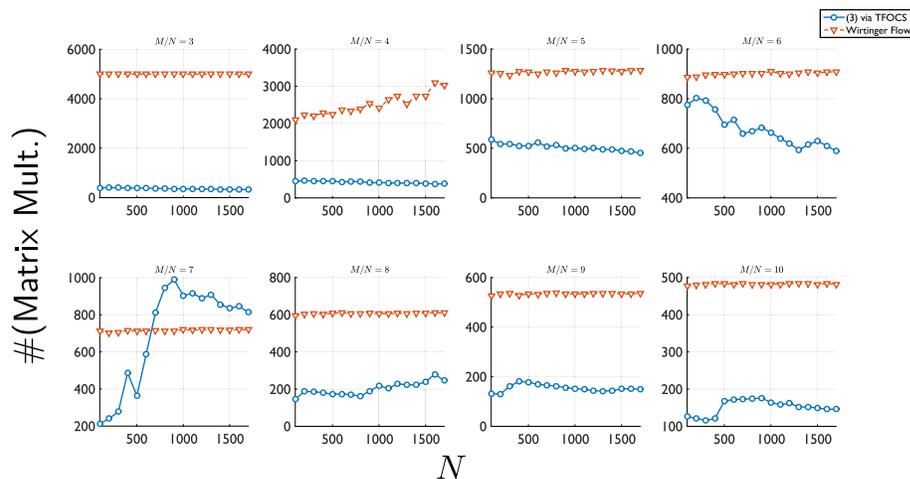


FIG 3. Relative error

show the effect of different levels of noise on the relative error for both of the considered noise models.

We also have another set of experiments comparing our method with the Wirtinger Flow method [10]. In these experiments we generate the measurement matrix $\mathbf{A} = [\mathbf{a}_1 \ \mathbf{a}_2 \ \cdots \ \mathbf{a}_M]^*$ with i.i.d. standard complex Gaussian entries for different signal dimensions N and sampling ratios M/N . For each choice of M and N we compared the performance of our approach and the Wirtinger Flow in the noiseless setting over 75 trials. Figure 3 compares the two methods in terms of the achieved accuracy. The graph shows the median of the rela-

FIG 4. Number of forward or adjoint applications of \mathbf{A}

tive error achieved by each of the two methods over the 75 trials. As can be seen, the Wirtinger Flow succeeds at a smaller sampling ratio (i.e., $M/N = 4$) whereas our approach requires essentially twice more measurements to succeed (i.e., $M/N = 8$). However, when both methods recover the signal accurately, our approach appears to be less demanding in terms of the computation and thus the running time. This claim is supported by Figure 4 that compares the median of the number of multiplications by \mathbf{A} or \mathbf{A}^* in each of the two approaches. We excluded the operation counts of initializing for the Wirtinger Flow and constructing the anchor for (3) from the reported numbers as they are both based on the method in Lemma 1. As Figure 4 suggests, the computational cost of our method peaks at $M/N = 7$ which can be explained as follows. At sampling ratios that our method fails (i.e., $M/N < 8$) the convex solver might not achieve the tolerance level set to halt and run for the entire prescribed maximum number of iterations. Therefore, for these sampling ratios the convex solver runs more iterations (without success) compared to the cases with $M/N \geq 8$ where the method succeeds.

We examined the numerical performance of (4) in the sparse phase retrieval problem as well. Let $M_0 = 2K_\star \log(N/K_\star)$ and $M = \lceil cM_0 \rceil$. For $N = 1000$ and different values of K_\star and c , we generated M measurement vectors according to (5) with $\Psi \in \mathbb{C}^{N \times \lceil M_0 \rceil}$ and $w_i \in \mathbb{C}^{\lceil M_0 \rceil}$ being populated with i.i.d. standard complex Gaussian entries. Figure 5 illustrated the median of the relative error achieved by (4) over 100. As the graphs suggest for the considered range of K_\star the proposed estimator recovers the signal (up to the numerical tolerance) for $M/M_0 \approx c \geq 14$.

Furthermore, we evaluated our method using noiseless measurements with coded diffraction patterns as described in [9]. Specifically, with indices $i = (k, \ell)$ for $1 \leq k \leq N$ and $1 \leq \ell \leq L$, we used measurements of the form $\mathbf{a}_i = \mathbf{f}_k \circ \phi_\ell$

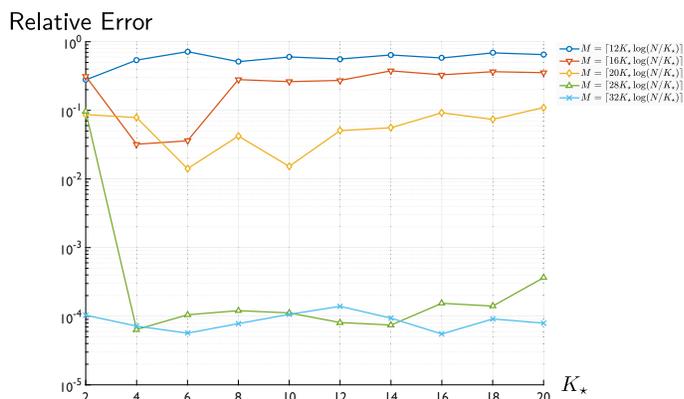


FIG 5. Relative error achieved (4) for sparse phase retrieval

which is the pointwise product of the k -th discrete Fourier basis vector (i.e., \mathbf{f}_k) and a random modulation pattern with i.i.d. symmetric Bernoulli entries (i.e., ϕ_ℓ). The target signal is an $N = 960 \times 1280 \approx 1.2 \times 10^6$ pixel image of a Persian Leopard.² We used $L = 20$ independent coded diffraction patterns $\{\phi_\ell\}_{1 \leq \ell \leq L}$. Therefore, the total number of (scalar) measurements is $M = LN \approx 2.5 \times 10^7$. Similar to the first simulation, the vector \mathbf{a}_0 is constructed as the (approximate) principal eigenvector of $\frac{1}{M} \sum_i b_i \mathbf{a}_i \mathbf{a}_i^*$ through 50 iterations of the power method. The convex program is also solved using TFOCS, but this time with smoothing parameter $\mu = 10^{-6}$ and restricting the total number of forward and adjoint coded diffraction operator to 500. The recovered image is depicted in Figure 6 which has a relative error of about 8.2×10^{-8} .

3. Theoretical analysis

In this section we provide the precise statement of our results. Main proofs are provided in Appendix B, and proofs of technical lemmas can be found in Appendix C. For the sake of simplicity in notation and derivation, but without loss of generality, we make the following assumptions. We assume that $\mathbf{a}_0^* \mathbf{x}_*$ is a positive real number since any point in $\mathbb{T} \mathbf{x}_*$ is a valid target. Furthermore, we assume that \mathbf{x}_* is unit-norm (i.e., $\|\mathbf{x}_*\|_2 = 1$) and thus the bound on the noise reduces to $\|\xi\|_\infty \leq \eta^{-1}$. We first establish, in Lemma 2, a geometric condition for success of phase retrieval through (3). Then we use this lemma to prove our main result for random measurements in Theorem 1. We also rely on tools from statistical learning theory that are outlined in Appendix A.

²Available online at: https://upload.wikimedia.org/wikipedia/commons/thumb/7/7d/Persian_Leopard_sitting.jpg/1280px-Persian_Leopard_sitting.jpg



FIG 6. *Persian Leopard at 960×1280 resolution. Relative error is 8.2×10^{-8} .*

3.1. Geometry of intersecting slabs

To understand the geometry of (3) it is worthwhile to first consider the noiseless scenario. The feasible set is the intersection of the sets

$$\mathcal{S}_i = \left\{ \mathbf{x} \in \mathbb{C}^N : |\mathbf{a}_i^* \mathbf{x}|^2 \leq b_i \right\}$$

corresponding to the pairs (\mathbf{a}_i, b_i) for $i = 1, 2, \dots, M$. The sets \mathcal{S}_i are effectively symmetric “complex slabs”. Denote their intersection by

$$\mathcal{K} = \bigcap_{i=1}^M \mathcal{S}_i.$$

In (3) the objective function is linear, thus its maximizer is an extreme point of the convex constraint set \mathcal{K} . Clearly, \mathbf{x}_\star as well as any other point in $\mathbb{T}\mathbf{x}_\star$ are extreme points of \mathcal{K} . However, \mathcal{K} typically has other extreme points that are not equivalent to \mathbf{x}_\star . Intuitively, using the non-vanishing correlation of \mathbf{a}_0 with \mathbf{x}_\star , the convex program (3) is effectively eliminating the superfluous extreme points of \mathcal{K} . The geometric interpretation is that the hyperplane normal to \mathbf{a}_0 that passes through \mathbf{x}_\star is also tangent to \mathcal{K} , as Figure 1 suggests. It is not difficult to show that an analogous interpretation from the dual point of view is that \mathbf{a}_0 is in the interior of cone $\{\mathbf{a}_i \mathbf{a}_i^* \mathbf{x}_\star\}_{1 \leq i \leq N}$ (i.e., the conical hull of the points $\mathbf{a}_i \mathbf{a}_i^* \mathbf{x}_\star$).

More generally, with noisy measurements, \mathcal{K} is still a symmetric complex polytope that is convex and includes $\mathbb{T}\mathbf{x}_\star$ due to non-negativity of the noise. We would like to find conditions that guarantee that the solution to (3) is close

to \mathbf{x}_* . More specifically, we would like to show that if $\widehat{\mathbf{x}} = \mathbf{x}_* + \mathbf{h}$ is any solution to (3) and $t > 0$ is some constant, then with $\|\mathbf{h}\|_2 > (t\eta)^{-1}$ the inequalities

$$\begin{aligned} \langle \mathbf{a}_0, \mathbf{h} \rangle &\geq 0 \\ |\mathbf{a}_i^* (\mathbf{x}_* + \mathbf{h})|^2 &\leq |\mathbf{a}_i^* \mathbf{x}_*|^2 + \xi_i \quad 1 \leq i \leq M, \end{aligned}$$

cannot hold simultaneously. The following lemma provides the desired sufficient condition.

Lemma 2. *Let*

$$\mathcal{R}_\delta = \{ \mathbf{h} \in \mathbb{C}^N : \|\mathbf{h} - (\mathbf{x}_*^* \mathbf{h}) \mathbf{x}_*\|_2 \geq \delta |\text{Im}(\mathbf{x}_*^* \mathbf{h})| \}, \quad (11)$$

and $\epsilon \geq 0$ be some constant. If every vector $\mathbf{h} \in \mathcal{R}_\delta$ with $\|\mathbf{h}\|_2 > \epsilon$ violates at least one of the inequalities

$$\begin{aligned} \langle \mathbf{a}_0, \mathbf{h} \rangle &\geq 0 \\ \langle \mathbf{a}_i \mathbf{a}_i^* \mathbf{x}_*, \mathbf{h} \rangle &\leq \frac{1}{2} \eta^{-1} \quad 1 \leq i \leq M, \end{aligned} \quad (12)$$

then any solution $\widehat{\mathbf{x}}$ to (3) obeys

$$\|\widehat{\mathbf{x}} - \mathbf{x}_*\|_2 \leq \epsilon.$$

Proof. It suffices to show that $\mathbf{h} = \widehat{\mathbf{x}} - \mathbf{x}_*$ obeys (12) and it belongs to \mathcal{R}_δ . Given that

$$\xi_i \geq |\mathbf{a}_i^* (\mathbf{x}_* + \mathbf{h})|^2 - |\mathbf{a}_i^* \mathbf{x}_*|^2 = |\mathbf{a}_i^* \mathbf{h}|^2 + 2\langle \mathbf{a}_i \mathbf{a}_i^* \mathbf{x}_*, \mathbf{h} \rangle \geq 2\langle \mathbf{a}_i \mathbf{a}_i^* \mathbf{x}_*, \mathbf{h} \rangle$$

and $\xi_i \leq \eta^{-1}$, we have $\langle \mathbf{a}_i \mathbf{a}_i^* \mathbf{x}_*, \mathbf{h} \rangle \leq \frac{1}{2} \eta^{-1}$. Feasibility of \mathbf{x}_* and optimality of $\widehat{\mathbf{x}}$ guarantee that $\langle \mathbf{a}_0, \mathbf{h} \rangle = \langle \mathbf{a}_0, \widehat{\mathbf{x}} \rangle - \langle \mathbf{a}_0, \mathbf{x}_* \rangle \geq 0$. Therefore, we have shown that \mathbf{h} satisfies (12).

The constraints of (3) are invariant under a global change of phase (i.e., the action of \mathbb{T}). It easily follows that the solution $\widehat{\mathbf{x}}$ to (3) should obey $\text{Im}(\mathbf{a}_0^* \widehat{\mathbf{x}}) = 0$. Therefore, we have $\text{Im}(\mathbf{a}_0^* \mathbf{h}) = 0$ as we assumed $\alpha = \mathbf{a}_0^* \mathbf{x}_* \in \mathbb{R}$. The same assumption also implies that $\mathbf{a}_0 = \alpha \mathbf{x}_* + \mathbf{a}_{0\perp}$ for $\mathbf{a}_{0\perp} = (\mathbf{I} - \mathbf{x}_* \mathbf{x}_*^*) \mathbf{a}_0$ which clearly obeys $\mathbf{x}_*^* \mathbf{a}_{0\perp} = 0$. Thus, using triangle inequality and the bound (2) we obtain

$$\begin{aligned} 0 = |\text{Im}(\mathbf{a}_0^* \mathbf{h})| &= |\alpha \text{Im}(\mathbf{x}_*^* \mathbf{h}) + \text{Im}(\mathbf{a}_{0\perp}^* \mathbf{h})| \\ &\geq \alpha |\text{Im}(\mathbf{x}_*^* \mathbf{h})| - |\text{Im}(\mathbf{a}_{0\perp}^* \mathbf{h})| \\ &\geq \delta \|\mathbf{a}_0\|_2 |\text{Im}(\mathbf{x}_*^* \mathbf{h})| - \|\mathbf{h}_\perp\|_2 \|\mathbf{a}_0\|_2, \end{aligned}$$

where $\mathbf{h}_\perp = (\mathbf{I} - \mathbf{x}_* \mathbf{x}_*^*) \mathbf{h} = \mathbf{h} - (\mathbf{x}_*^* \mathbf{h}) \mathbf{x}_*$. The above inequality completes the proof as it is equivalent to $\mathbf{h} \in \mathcal{R}_\delta$. \square

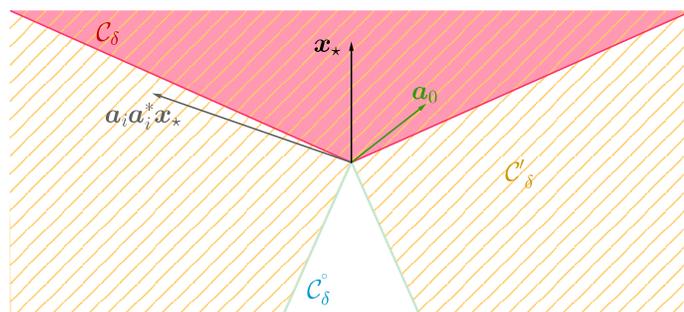


FIG 7. Relative positioning of \mathbf{a}_0 and $\mathbf{a}_i \mathbf{a}_i^* \mathbf{x}_*$ with respect to the cones \mathcal{C}_δ and \mathcal{C}'_δ

3.2. Guarantees for random measurements

In this section we will show that if the vectors \mathbf{a}_i for $1 \leq i \leq M$ are drawn from a random distribution and (2) holds for a sufficiently large constant δ , then with high probability (3) produces an accurate estimate of \mathbf{x}_* . Our strategy is to show that for a sufficiently large M the sufficient condition provided in Lemma 2 holds with high probability.

For $\delta \in (0, 1)$ let \mathcal{C}_δ be the convex cone given by

$$\mathcal{C}_\delta = \{ \mathbf{y} \in \mathbb{C}^N : \mathbf{x}_*^* \mathbf{y} \geq \delta \|\mathbf{y}\|_2 \},$$

where $\mathbf{x}_*^* \mathbf{y}$ is implicitly assumed to be a real number. The *polar cone* of a set \mathcal{C} is defined as

$$\mathcal{C}^\circ \stackrel{\text{def}}{=} \{ \mathbf{z} : \langle \mathbf{z}, \mathbf{y} \rangle \leq 0 \text{ for all } \mathbf{y} \in \mathcal{C} \}.$$

It is easy to verify that the polar cone of \mathcal{C}_δ is

$$\mathcal{C}_\delta^\circ = \left\{ \mathbf{z} \in \mathbb{C}^N : \delta \langle \mathbf{x}_*, \mathbf{z} \rangle \leq -\sqrt{1-\delta^2} \sqrt{\|\mathbf{z}\|_2^2 - |\mathbf{x}_*^* \mathbf{z}|^2} \right\}.$$

Since $\mathbf{a}_0 \in \mathcal{C}_\delta$ by assumption, it follows that for every $\mathbf{h} \in \mathcal{C}_\delta^\circ$ we have $\langle \mathbf{a}_0, \mathbf{h} \rangle \leq 0$. Therefore, the inequality $\langle \mathbf{a}_0, \mathbf{h} \rangle \geq 0$ can hold only for vectors \mathbf{z} in the closure of the complement of \mathcal{C}_δ° which we denote by

$$\mathcal{C}'_\delta = \left\{ \mathbf{z} \in \mathbb{C}^N : \delta \langle \mathbf{x}_*, \mathbf{z} \rangle \geq -\sqrt{1-\delta^2} \sqrt{\|\mathbf{z}\|_2^2 - |\mathbf{x}_*^* \mathbf{z}|^2} \right\}. \quad (13)$$

A typical positioning of \mathbf{a}_0 and $\mathbf{a}_i \mathbf{a}_i^* \mathbf{x}_*$ needed to guarantee unique recovery is illustrated in Figure 7.

Theorem 1. *Suppose that noisy phaseless measurements of a unit vector \mathbf{x}_* are given as in (1) under bounded non-negative noise $\xi_1, \xi_2, \dots, \xi_M \in [0, \eta^{-1}]$. Let \mathbf{a}_0 be an anchor vector obeying (2) for some constant $\delta \in (0, 1)$ and define \mathcal{R}_δ and \mathcal{C}'_δ respectively as in (11) and (13). Furthermore, given a constant $t > 0$,*

suppose that for $1 \leq i \leq M$ the measurement vectors \mathbf{a}_i are i.i.d. copies of a random variable $\mathbf{a} \in \mathbb{C}^N$ that obeys

$$\inf_{\substack{\mathbf{h} \in \mathcal{C}'_\delta \cap \mathcal{R}_\delta \\ \|\mathbf{h}\|_2 > (t\eta)^{-1}}} \mathbb{P} \left(\langle \mathbf{a}\mathbf{a}^* \mathbf{x}_*, \mathbf{h} \rangle > \frac{1}{2} \eta^{-1} \right) \geq p_{\min}(\delta, t),$$

for a constant $p_{\min}(\delta, t) \in (0, 1)$ depending only on parameters δ and t .³ For any $\varepsilon > 0$, if we have

$$M \gtrsim \frac{1}{p_{\min}^2(\delta, t)} \left(N + \log \frac{1}{\varepsilon} \right),$$

then with probability $\geq 1 - \varepsilon$ the estimate $\hat{\mathbf{x}}$ obtained through (3) obeys

$$\|\hat{\mathbf{x}} - \mathbf{x}_*\|_2 \leq (t\eta)^{-1}.$$

To have a concrete example, we can apply Theorem 1 to the case of measurements with normal distribution. Clearly, it suffices to quantify the constant $p_{\min}(\delta, t)$. For instance, Lemma 5 in Appendix C guarantees that we can choose

$$p_{\min}(\delta, t) = \left(\frac{1}{2} - \frac{\sqrt{1 - \delta^2}}{2} \right) e^{-2\sqrt{2}\delta^{-2}t}.$$

The dependence of the sample complexity on $p_{\min}(\delta, t)$ in Theorem 1 is not optimal and can be improved. In our analysis the most critical factor that determines the sample complexity is a uniform lower bound for a certain empirical process. While we used Theorem 3 to establish the desired lower bound, a sharper result can be obtained using “variance-normalized” variants of the theorem. For example, the term $1/p_{\min}^2(\delta, t)$ in the sample complexity guaranteed by Theorem 1 can be improved to $1/p_{\min}(\delta, t)$ by using [34, Theorem 4.4(2)] instead of Theorem 3 in the analysis.

3.3. Sparse phase retrieval

Theorem 2 below gives a precise recovery guarantee (in terms of number of samples needed) for recovering a K_* -sparse vector. Its proof, which can be found in Section B.2, follows that of Theorem 1, only with a modification of the set \mathcal{C}'_δ in (13) that reflects the additional constraint that \mathbf{h} be a descent vector for the ℓ_1 norm from \mathbf{x}_* .

To ease some of the notation, in what follows we denote the phase of any $z \in \mathbb{C}$ by

$$\langle z \rangle_{\mathbb{T}} \stackrel{\text{def}}{=} \begin{cases} \frac{z}{|z|} & , z \neq 0 \\ 0 & , z = 0. \end{cases}$$

For any complex vector \mathbf{z} , the operation $\langle \mathbf{z} \rangle_{\mathbb{T}}$ applies entry-wise.

³Clearly, the best $p_{\min}(\delta, t)$ decreases as t increases.

Theorem 2. *Suppose that noisy phaseless measurements of a K_* -sparse unit vector \mathbf{x}_* supported on \mathcal{X}_* are given as in (1) under bounded non-negative noise $\xi_1, \xi_2, \dots, \xi_M \in [0, \eta^{-1}]$. Let \mathbf{a}_0 be an anchor vector obeying (2) for some constant $\delta \in (0, 1)$ and define \mathcal{R}_δ and \mathcal{C}'_δ respectively as in (11) and (13). Furthermore, given a constant $t > 0$, suppose that for $1 \leq i \leq M$ we have $\mathbf{a}_i = \Psi \mathbf{w}_i$, where $\Psi \in \mathbb{C}^{N \times M_0}$ obeys (6) with $K = K_*$ and \mathbf{w}_i 's are i.i.d. copies of a random variable $\mathbf{w} \in \mathbb{C}^{M_0}$ that obeys*

$$\inf_{\substack{\mathbf{h} \in \mathcal{C}'_\delta \cap \mathcal{R}_\delta \\ \|\mathbf{h}\|_2 > (t\eta)^{-1} \\ \|\mathbf{h}_{\mathcal{X}_\varepsilon}\|_1 + \langle \mathbf{x}_* \mathbf{x}_* \rangle_{\mathcal{T}}, \mathbf{h}_{\mathcal{X}_*} \rangle \leq 0}} \mathbb{P} \left(\langle \Psi \mathbf{w} \mathbf{w}^* \Psi^* \mathbf{x}_*, \mathbf{h} \rangle > \frac{1}{2} \eta^{-1} \right) \geq p_{\min}(\delta, t), \quad (14)$$

for a constant $p_{\min}(\delta, t) \in (0, 1)$ that depend on parameters δ, t , and implicitly on Ψ . For any $\varepsilon > 0$, if we have

$$M \gtrsim \frac{1}{p_{\min}^2(\delta, t)} \left(M_0 + \log \frac{1}{\varepsilon} \right),$$

then with probability $\geq 1 - \varepsilon$ the estimate $\hat{\mathbf{x}}$ obtained through (4) obeys

$$\|\hat{\mathbf{x}} - \mathbf{x}_*\|_2 \leq (t\eta)^{-1}.$$

As mentioned above, with $M_0 \gtrsim K_* \log(N/K_*)$ the matrix Ψ can satisfy (6) for $K = K_*$ and a constant $\gamma_K < 1$. Therefore, Theorem 2 implies that it suffices to have

$$M \gtrsim \frac{1}{p_{\min}^2(\delta, t)} \left(K_* \log \frac{N}{K_*} + \log \frac{1}{\varepsilon} \right),$$

to guarantee accuracy of (4) with probability $\geq 1 - \varepsilon$.

There is one important thing that we are leaving unaddressed. Unlike ordinary unconstrained phase retrieval, we do not show the existence of a constant $p_{\min}(\delta, t)$ in (14) for a particular probability model for the \mathbf{w}_i . This appears to be a very technical calculation even in the case where the \mathbf{w}_i are iid Gaussian, and so we leave it for future work.

Constructing the anchor: As described above, the first step in forming the anchor vector is computing the principal eigenvector \mathbf{w}_0 of

$$\Sigma = \frac{1}{M} \sum_{i=1}^M b_i \mathbf{w} \mathbf{w}^*.$$

When $M \gtrsim M_0 \log M_0$, Lemma 1 above guarantees that

$$|\mathbf{w}_0^* \Psi^* \mathbf{x}_*| \geq \delta' \|\Psi^* \mathbf{x}_*\|_2. \quad (15)$$

Lemma 3 below shows that the best K_* -sparse approximation of $\Psi \mathbf{w}_0$ is sufficiently correlated with \mathbf{x}_* . The proof is provided in Appendix B.

Lemma 3. Let \mathbf{a}_0 denote the best K_* -sparse approximation of $\Psi \mathbf{w}_0$ where \mathbf{w}_0 is a unit vector that satisfies (15), and suppose that Ψ satisfies the (6) with $K = 2K_*$ and parameter γ_{2K_*} . Then \mathbf{a}_0 is a valid anchor for which (2) holds, if we have

$$\sqrt{\frac{1 + \gamma_{2K_*}}{1 - \gamma_{2K_*}}} \left(\gamma_{2K_*} + \sqrt{1 + \gamma_{2K_*}} \sqrt{2 - 2\delta'} \right)^2 \leq \frac{1 - \delta}{2}.$$

For concreteness, if $\delta' = \frac{7}{8}$ and $\gamma_{2K_*} = \frac{1}{20}$, then Lemma 3 guarantees that we can have $\delta = \frac{1}{3}$.

Appendix A: Tools from statistical learning theory

For reference, here we provide some of the classic results in statistical learning theory that we employed in our analysis. We mostly follow the exposition of the subject presented by [13, chapters 13 and 14].

Definition 1 (Shatter coefficient). The n -th shatter coefficient (or growth function) of a class \mathcal{F} of binary functions $f : \mathcal{X} \rightarrow \{0, 1\}$ is defined as

$$s(\mathcal{F}, n) \stackrel{\text{def}}{=} \max_{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n \in \mathcal{X}} |\{(f(\mathbf{x}_1), f(\mathbf{x}_2), \dots, f(\mathbf{x}_n)) : f \in \mathcal{F}\}|.$$

Intuitively, the shatter coefficient $s(\mathcal{F}, n)$ is the largest number of binary patterns that the functions in \mathcal{F} can induce on n points.

Definition 2 (VC-dimension). The Vapnik–Chervonenkis (VC) dimension of a class \mathcal{F} of binary functions is the largest number n such that $s(\mathcal{F}, n) = 2^n$, namely,

$$\dim_{\text{VC}}(\mathcal{F}) \stackrel{\text{def}}{=} \max \{n : s(\mathcal{F}, n) = 2^n\}.$$

Naturally, $\dim_{\text{VC}}(\mathcal{F}) = \infty$ if $s(\mathcal{F}, n) = 2^n$ for all n .

If \mathcal{F} can induce all binary patterns on n points, \mathcal{F} is said to “shatter” n points. Therefore, the VC-dimension of \mathcal{F} is the largest number of points that \mathcal{F} can shatter.

Lemma 4 (Vapnik and Chervonenkis [35], Sauer [29], Shelah [31]). For a class \mathcal{F} of binary functions with VC-dimension $d = \dim_{\text{VC}}(\mathcal{F})$ we have

$$s(\mathcal{F}, n) \leq \sum_{i=0}^d \binom{n}{i}.$$

In particular,

$$s(\mathcal{F}, n) \leq \left(\frac{en}{d}\right)^d. \tag{16}$$

The following theorem is originally due to Vapnik and Chervonenkis [35]. We restate the theorem as presented in [13].

Theorem 3 (Vapnik and Chervonenkis [35]). *Let \mathcal{F} be a class of binary functions and $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ be i.i.d. copies of an arbitrary random variable \mathbf{x} . Then for every $t > 0$ we have*

$$\mathbb{P} \left(\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(\mathbf{x}_i) - \mathbb{E}f(\mathbf{x}) \right| > t \right) \leq 8s(\mathcal{F}, n)e^{-nt^2/8}.$$

Appendix B: Main proofs

B.1. Theorem 1

The main ingredient in the proof below is the VC-theorem (i.e., Theorem 3) that helps bounding the deviation of an “empirical probability” from its expectation uniformly.

Proof of Theorem 1. Let $\mathbf{h} = \widehat{\mathbf{x}} - \mathbf{x}_*$. It suffices to show that for any $\mathbf{h} \in \mathcal{C}'_\delta \cap \mathcal{R}_\delta$ with $\|\mathbf{h}\|_2 > (t\eta)^{-1}$ there exists at least one $1 \leq i \leq M$ such that $\langle \mathbf{a}_i \mathbf{a}_i^* \mathbf{x}_*, \mathbf{h} \rangle > \frac{1}{2}\eta^{-1}$. Specifically, we would like to show that with high probability

$$\sum_{i=1}^M \mathbb{1} \left(\langle \mathbf{a}_i \mathbf{a}_i^* \mathbf{x}_*, \mathbf{h} \rangle > \frac{1}{2}\eta^{-1} \right) > 0.$$

Denote the empirical probability of $\langle \mathbf{a} \mathbf{a}^* \mathbf{x}_*, \mathbf{h} \rangle > \frac{1}{2}\eta^{-1}$ by

$$\widehat{p}_M(\mathbf{h}) = \frac{1}{M} \sum_{i=1}^M \mathbb{1} \left(\langle \mathbf{a}_i \mathbf{a}_i^* \mathbf{x}_*, \mathbf{h} \rangle > \frac{1}{2}\eta^{-1} \right),$$

which is an approximation of the true probability of the event denoted by

$$p(\mathbf{h}) = \mathbb{E} \mathbb{1} \left(\langle \mathbf{a} \mathbf{a}^* \mathbf{x}_*, \mathbf{h} \rangle > \frac{1}{2}\eta^{-1} \right) = \mathbb{P} \left(\langle \mathbf{a} \mathbf{a}^* \mathbf{x}_*, \mathbf{h} \rangle > \frac{1}{2}\eta^{-1} \right).$$

Considering the set of binary functions

$$\mathcal{F} = \left\{ z \mapsto \mathbb{1} \left(\langle z, \mathbf{h} \rangle > \frac{1}{2}\eta^{-1} \right) : \mathbf{h} \in \mathcal{C}'_\delta \cap \mathcal{R}_\delta \text{ and } \|\mathbf{h}\|_2 > (t\eta)^{-1} \right\},$$

whose shatter coefficient is denoted by $s(\mathcal{F}, M)$, a direct application of Theorem 3 in Appendix A shows that

$$\sup_{\substack{\mathbf{h} \in \mathcal{C}'_\delta \cap \mathcal{R}_\delta \\ \|\mathbf{h}\|_2 > (t\eta)^{-1}}} |\widehat{p}_M(\mathbf{h}) - p(\mathbf{h})| \leq \sqrt{\frac{8 \log \frac{s(\mathcal{F}, M)}{\varepsilon}}{M}}$$

with probability $\geq 1 - \varepsilon$. Since \mathcal{F} is a subset of \mathcal{H} the set of indicators of all half-spaces (with a common offset), it has a smaller VC-dimension than \mathcal{H} .

Moreover, it is well-known—as a direct implication of Radon’s theorem [see e.g., 24]—that the VC–dimension of half-spaces indicators is no more than the ambient dimension. In particular, we have $\dim_{\text{VC}}(\mathcal{F}) \leq \dim_{\text{VC}}(\mathcal{H}) \leq 2N$ as our domain is effectively a $2N$ -dimensional real vector space. Therefore, invoking Lemma 4 below we obtain

$$\sup_{\substack{\mathbf{h} \in \mathcal{C}'_\delta \cap \mathcal{R}_\delta \\ \|\mathbf{h}\|_2 > (t\eta)^{-1}}} |\widehat{p}_M(\mathbf{h}) - p(\mathbf{h})| \leq \sqrt{\frac{16N \log \frac{eM}{2N} + 8 \log \frac{8}{\epsilon}}{M}}.$$

Now, because $p(\mathbf{h}) \geq p_{\min}(\delta, t)$ for all $\mathbf{h} \in \mathcal{C}'_\delta \cap \mathcal{R}_\delta$ with $\|\mathbf{h}\|_2 > (t\eta)^{-1}$, the above inequality implies that

$$\inf_{\substack{\mathbf{h} \in \mathcal{C}'_\delta \cap \mathcal{R}_\delta \\ \|\mathbf{h}\|_2 > (t\eta)^{-1}}} \widehat{p}_M(\mathbf{h}) \geq p_{\min}(\delta, t) - \sqrt{\frac{16N \log \frac{eM}{2N} + 8 \log \frac{8}{\epsilon}}{M}}.$$

If $M = \frac{8}{p_{\min}^2(\delta, t)} (c \cdot 2N + 2 \log \frac{8}{\epsilon})$, then we have

$$\begin{aligned} \log \frac{eM}{2N} + \frac{\log \frac{8}{\epsilon}}{2N} &= \log \frac{8e}{p_{\min}^2(\delta, t)} + \log \left(c + \frac{\log \frac{8}{\epsilon}}{N} \right) + \frac{\log \frac{8}{\epsilon}}{2N} \\ &\leq \log \frac{8e}{p_{\min}^2(\delta, t)} + \frac{c}{2} + \frac{\log \frac{8}{\epsilon}}{2N} - 1 + \log 2 + \frac{\log \frac{8}{\epsilon}}{2N} \\ &< \log \frac{8e}{p_{\min}^2(\delta, t)} + \frac{c}{2} + \frac{\log \frac{8}{\epsilon}}{N}, \end{aligned}$$

where we used the inequality $\log u - \log 2 = \log \frac{u}{2} \leq \frac{u}{2} - 1$ in the second line. Setting $c = 2 \log \frac{8e}{p_{\min}^2(\delta, t)}$, it follows that

$$\begin{aligned} \frac{16N \log \frac{eM}{2N} + 8 \log \frac{8}{\epsilon}}{M} &= \frac{16N}{M} \left(\log \frac{eM}{2N} + \frac{\log \frac{8}{\epsilon}}{2N} \right) \\ &< \frac{16N}{M} \left(2 \log \frac{8e}{p_{\min}^2(\delta, t)} + \frac{\log \frac{8}{\epsilon}}{N} \right) = p_{\min}^2(\delta, t) \end{aligned}$$

and thus we can guarantee that

$$\inf_{\substack{\mathbf{h} \in \mathcal{C}'_\delta \cap \mathcal{R}_\delta \\ \|\mathbf{h}\|_2 > (t\eta)^{-1}}} \widehat{p}_M(\mathbf{h}) > p_{\min}(\delta, t) - p_{\min}(\delta, t) = 0.$$

This immediately implies that for $M \gtrsim N + \log \frac{1}{\epsilon}$ we have

$$\inf_{\substack{\mathbf{h} \in \mathcal{C}'_\delta \cap \mathcal{R}_\delta \\ \|\mathbf{h}\|_2 > (t\eta)^{-1}}} \sum_{i=1}^M \mathcal{K}(\langle \mathbf{a}_i \mathbf{a}_i^* \mathbf{x}_*, \mathbf{h} \rangle > \frac{1}{2} \eta^{-1}) = M \widehat{p}_M(\mathbf{h}) > 0,$$

as desired. □

B.2. Theorem 2

Proof of Theorem 2. Let $\mathbf{h} = \hat{\mathbf{x}} - \mathbf{x}_*$. By feasibility of $\hat{\mathbf{x}}$ we have $\|\mathbf{x}_* + \mathbf{h}\|_1 \leq \|\mathbf{x}_*\|_1$ which implies

$$\|\mathbf{h}_{\mathcal{X}_*^c}\|_1 + \langle \mathbf{x}_* \mathbf{x}_*^\top, \mathbf{h}_{\mathcal{X}_*} \rangle \leq 0. \quad (17)$$

Therefore, similar to the case of ordinary phase retrieval it suffices to show that for any $\mathbf{h} \in \mathcal{C}'_\delta \cap \mathcal{R}_\delta$ that obeys (17) and $\|\mathbf{h}\|_2 > (t\eta)^{-1}$ we should have $\langle \mathbf{a}_i \mathbf{a}_i^\top \mathbf{x}_*, \mathbf{h} \rangle > \frac{1}{2}\eta^{-1}$ for at least one $1 \leq i \leq M$. The goal is again to show that

$$\begin{aligned} \hat{p}_M(\mathbf{h}) &= \frac{1}{M} \sum_{i=1}^M \mathbb{1} \left(\langle \mathbf{a}_i \mathbf{a}_i^\top \mathbf{x}_*, \mathbf{h} \rangle > \frac{1}{2}\eta^{-1} \right) \\ &= \frac{1}{M} \sum_{i=1}^M \mathbb{1} \left(\langle \Psi \mathbf{w}_i \mathbf{w}_i^* \Psi^* \mathbf{x}_*, \mathbf{h} \rangle > \frac{1}{2}\eta^{-1} \right), \end{aligned}$$

is strictly positive for every $\mathbf{h} \in \mathcal{C}'_\delta \cap \mathcal{R}_\delta$ that obeys (17) and $\|\mathbf{h}\|_2 > (t\eta)^{-1}$. Let

$$p(\mathbf{h}) = \mathbb{E} \hat{p}_M(\mathbf{h}) = \mathbb{P} \left(\langle \Psi \mathbf{w} \mathbf{w}^* \Psi^* \mathbf{x}_*, \mathbf{h} \rangle > \frac{1}{2}\eta^{-1} \right).$$

The binary summands in $\hat{p}_M(\mathbf{h})$ can be equivalently expressed as a function of $\mathbf{h}' = \Psi^* \mathbf{h} \in \mathbb{C}^{M_0}$. Therefore, we can treat these binary functions as linear classifiers in \mathbb{C}^{M_0} and apply Theorem 3, in the same fashion that was applied for ordinary phase retrieval, to show that

$$\sup_{\substack{\mathbf{h} \in \mathcal{C}'_\delta \cap \mathcal{R}_\delta \\ \|\mathbf{h}\|_2 > (t\eta)^{-1} \\ \|\mathbf{h}_{\mathcal{X}_*^c}\|_1 + \langle \mathbf{x}_* \mathbf{x}_*^\top, \mathbf{h}_{\mathcal{X}_*} \rangle \leq 0}} |\hat{p}_M(\mathbf{h}) - p(\mathbf{h})| \leq \sqrt{\frac{16M_0 \log \frac{eM}{2M_0} + 8 \log \frac{8}{\varepsilon}}{M}},$$

holds with probability $\geq 1 - \varepsilon$. As in the proof of Theorem 1, straightforward calculations show that it suffices to have

$$M \gtrsim \frac{1}{p_{\min}^2(\delta, t)} (M_0 + \log \frac{1}{\varepsilon})$$

to guarantee with probability $\geq 1 - \varepsilon$ that $\hat{p}_M(\mathbf{h}) > 0$. \square

B.3. Lemma 3

The proof below primarily uses the equivalent formulations of RIP given by (6) and (18).

Proof of Lemma 3. By definition of \mathbf{a}_0 , for any $\beta \in \mathbb{C}$ we have

$$\|\Psi \mathbf{w}_0 - \mathbf{a}_0\|_2^2 \leq \|\Psi \mathbf{w}_0 - \beta \mathbf{x}_*\|_2^2,$$

which is equivalent to

$$\langle \mathbf{a}_0 - \beta \mathbf{x}_*, \mathbf{a}_0 + \beta \mathbf{x}_* - 2\Psi \mathbf{w}_0 \rangle \leq 0.$$

Adding and subtracting $\beta \mathbf{x}_*$ to the second argument of the inner product and rearranging the terms yields

$$\|\mathbf{a}_0 - \beta \mathbf{x}_*\|_2^2 \leq 2\langle \mathbf{a}_0 - \beta \mathbf{x}_*, \Psi \mathbf{w}_0 - \beta \mathbf{x}_* \rangle.$$

Let $\mathbf{e} = \mathbf{w}_0 - \beta \Psi^* \mathbf{x}_*$. Then, using the fact that (6) implies

$$\|P_K \Psi \Psi^* P_K^* - \mathbf{I}\| \leq \gamma_K, \quad (18)$$

for every P_K that is a projection onto K canonical coordinates, and the fact that $\mathbf{a}_0 - \beta \mathbf{x}_*$ has at most $2K_*$ nonzero entries we can write

$$\begin{aligned} \|\mathbf{a}_0 - \beta \mathbf{x}_*\|_2^2 &\leq 2\langle \mathbf{a}_0 - \beta \mathbf{x}_*, \beta \Psi \Psi^* \mathbf{x}_* + \Psi \mathbf{e} - \beta \mathbf{x}_* \rangle \\ &= 2\langle \mathbf{a}_0 - \beta \mathbf{x}_*, \beta (\Psi \Psi^* - \mathbf{I}) \mathbf{x}_* \rangle + 2\langle \mathbf{a}_0 - \beta \mathbf{x}_*, \Psi \mathbf{e} \rangle \\ &\leq 2\gamma_{2K_*} \|\mathbf{a}_0 - \beta \mathbf{x}_*\|_2 + 2\|\Psi^*(\mathbf{a}_0 - \beta \mathbf{x}_*)\|_2 \|\mathbf{e}\|_2 \\ &\leq 2\gamma_{2K_*} \|\mathbf{a}_0 - \beta \mathbf{x}_*\|_2 + 2\sqrt{1 + \gamma_{2K_*}} \|\mathbf{a}_0 - \beta \mathbf{x}_*\|_2 \|\mathbf{e}\|_2. \end{aligned}$$

Assuming that \mathbf{a}_0 is not perfectly aligned with \mathbf{x}_* , and by choosing $\beta = \langle \mathbf{x}_*^* \Psi \mathbf{w}_0 \rangle_{\mathbb{T}} / \|\Psi^* \mathbf{x}_*\|_2$ which minimizes $\|\mathbf{e}\|_2$, we deduce that

$$\begin{aligned} \|\mathbf{a}_0 - \beta \mathbf{x}_*\|_2 &\leq 2\gamma_{2K_*} + 2\sqrt{1 + \gamma_{2K_*}} \sqrt{2 - 2\frac{|\mathbf{w}_0^* \Psi^* \mathbf{x}_*|}{\|\Psi^* \mathbf{x}_*\|_2}} \\ &\leq 2\gamma_{2K_*} + 2\sqrt{1 + \gamma_{2K_*}} \sqrt{2 - 2\delta'}. \end{aligned}$$

Squaring both sides of the above inequality and expanding the squared norm then yields

$$\frac{|\mathbf{a}_0^* \mathbf{x}_*|}{\|\Psi^* \mathbf{x}_*\|_2} \geq \langle \mathbf{a}_0, \beta \mathbf{x}_* \rangle \geq \frac{1}{2} \|\mathbf{a}_0\|_2^2 + \frac{1}{2\|\Psi^* \mathbf{x}_*\|_2^2} - 2\left(\gamma_{2K_*} + \sqrt{1 + \gamma_{2K_*}} \sqrt{2 - 2\delta'}\right)^2.$$

By AM-GM inequality and (6) we obtain

$$\begin{aligned} |\mathbf{a}_0^* \mathbf{x}_*| &\geq \|\mathbf{a}_0\|_2 - 2\sqrt{1 + \gamma_{2K_*}} \left(\gamma_{2K_*} + \sqrt{1 + \gamma_{2K_*}} \sqrt{2 - 2\delta'}\right)^2 \\ &\geq \|\mathbf{a}_0\|_2 - \|\mathbf{a}_0\|_2 \cdot 2\sqrt{\frac{1 + \gamma_{2K_*}}{1 - \gamma_{2K_*}}} \left(\gamma_{2K_*} + \sqrt{1 + \gamma_{2K_*}} \sqrt{2 - 2\delta'}\right)^2, \end{aligned}$$

which completes the proof. \square

Appendix C: Technical lemmas

Lemma 5. *If $\mathbf{a} \sim \text{Normal}(\mathbf{0}, \frac{1}{2}\mathbf{I}) + i\text{Normal}(\mathbf{0}, \frac{1}{2}\mathbf{I})$ and \mathbf{x}_* is a unit vector, then for every $\mathbf{h} \in \mathcal{C}'_\delta \cap \mathcal{R}_\delta$ with $\|\mathbf{h}\|_2 > (t\eta)^{-1}$ we have*

$$\mathbb{P}\left(\langle \mathbf{a} \mathbf{a}^* \mathbf{x}_*, \mathbf{h} \rangle > \frac{1}{2}\eta^{-1}\right) \geq \left(\frac{1}{2} - \frac{\sqrt{1 - \delta^2}}{2}\right) e^{-2\sqrt{2}\delta^{-2}t}.$$

Proof. We can decompose \mathbf{h} as $\mathbf{h} = (\mathbf{x}_*^* \mathbf{h}) \mathbf{x}_* + \mathbf{h}_\perp$, where $\mathbf{x}_*^* \mathbf{h}_\perp = \mathbf{0}$. Therefore, we have $\langle \mathbf{a} \mathbf{a}^* \mathbf{x}_*, \mathbf{h} \rangle = \langle \mathbf{x}_*, \mathbf{h} \rangle |\mathbf{a}^* \mathbf{x}_*|^2 + \text{Re}(\overline{\mathbf{a}^* \mathbf{x}_*} \mathbf{a}^* \mathbf{h}_\perp)$. Using the facts

that $\mathbf{x}_*^* \mathbf{h}_\perp = \mathbf{0}$ and $\mathbf{a} \sim \text{Normal}(\mathbf{0}, \frac{1}{2} \mathbf{I}) + i \text{Normal}(\mathbf{0}, \frac{1}{2} \mathbf{I})$ it is straightforward to show that $\langle \mathbf{x}_*, \mathbf{h} \rangle |\mathbf{a}^* \mathbf{x}_*|^2 + \text{Re}(\overline{\mathbf{a}^* \mathbf{x}_*} \mathbf{a}^* \mathbf{h}_\perp)$ has the same distribution as $\frac{1}{2} \langle \mathbf{x}_*, \mathbf{h} \rangle \|\mathbf{g}_2\|^2 + \frac{1}{2} \|\mathbf{h}_\perp\|_2 \mathbf{g}_1^\top \mathbf{g}_2$ where $\mathbf{g}_1, \mathbf{g}_2 \in \mathbb{R}^2$ are independent standard Normal random variables:

$$\mathbb{P}\left(\langle \mathbf{a} \mathbf{a}^* \mathbf{x}_*, \mathbf{h} \rangle > \frac{1}{2} \eta^{-1}\right) = \mathbb{P}\left(\langle \mathbf{x}_*, \mathbf{h} \rangle \|\mathbf{g}_2\|_2^2 + \|\mathbf{h}_\perp\|_2 \mathbf{g}_1^\top \mathbf{g}_2 > \eta^{-1}\right).$$

Since \mathbf{g}_2 has a standard normal distribution, its norm and (normalized) direction are independent. Thus, we can treat $-\mathbf{g}_1^\top \frac{\mathbf{g}_2}{\|\mathbf{g}_2\|_2} = g$ as a standard Normal scalar which is independent of $\|\mathbf{g}_2\|_2 = v \sim \text{Rayleigh}(1)$. Therefore, we have

$$\begin{aligned} \mathbb{P}\left(\langle \mathbf{a} \mathbf{a}^* \mathbf{x}_*, \mathbf{h} \rangle > \frac{1}{2} \eta^{-1}\right) &= \mathbb{P}\left(\langle \mathbf{x}_*, \mathbf{h} \rangle \|\mathbf{g}_2\|_2^2 + \|\mathbf{h}_\perp\|_2 \|\mathbf{g}_2\|_2 \mathbf{g}_1^\top \frac{\mathbf{g}_2}{\|\mathbf{g}_2\|_2} > \eta^{-1}\right) \\ &= \mathbb{P}\left(\langle \mathbf{x}_*, \mathbf{h} \rangle v - \eta^{-1} v^{-1} > \|\mathbf{h}_\perp\|_2 g\right). \end{aligned}$$

Since $\mathbf{h} \in \mathcal{R}_\delta$ and $\|\mathbf{h}\|_2 > (t\eta)^{-1}$ we have

$$\begin{aligned} (t\eta)^{-2} < \|\mathbf{h}\|_2^2 &= \|\mathbf{h}_\perp\|_2^2 + (\text{Im}(\mathbf{x}_*^* \mathbf{h}))^2 + \langle \mathbf{x}_*, \mathbf{h} \rangle^2 \\ &\leq (1 + \delta^{-2}) \|\mathbf{h}_\perp\|_2^2 + \langle \mathbf{x}_*, \mathbf{h} \rangle^2. \end{aligned} \tag{19}$$

We consider two cases depending on $\|\mathbf{h}_\perp\|_2 = 0$ or not. If $\|\mathbf{h}_\perp\|_2 = 0$, then $|\langle \mathbf{x}_*, \mathbf{h} \rangle| > (t\eta)^{-1}$. The fact that $\mathbf{h} \in \mathcal{C}'_\delta$ as well, implies that $\langle \mathbf{x}_*, \mathbf{h} \rangle$ is non-negative and thereby $\langle \mathbf{x}_*, \mathbf{h} \rangle > (t\eta)^{-1}$. Consequently, we have

$$\begin{aligned} \mathbb{P}\left(\langle \mathbf{a} \mathbf{a}^* \mathbf{x}_*, \mathbf{h} \rangle > \frac{1}{2} \eta^{-1}\right) &= \mathbb{P}\left(\langle \mathbf{x}_*, \mathbf{h} \rangle v - \eta^{-1} v^{-1} > \|\mathbf{h}_\perp\|_2 g\right) \\ &\geq \mathbb{P}(t^{-1} v - v^{-1} > 0) = \mathbb{P}(v > \sqrt{t}) = e^{-\frac{t}{2}}. \end{aligned}$$

If $\|\mathbf{h}_\perp\|_2 > 0$, then we can invoke Lemma 6 below with $\alpha = \frac{\langle \mathbf{x}_*, \mathbf{h} \rangle}{\|\mathbf{h}_\perp\|_2}$ to show that

$$\begin{aligned} \mathbb{P}\left(\langle \mathbf{a} \mathbf{a}^* \mathbf{x}_*, \mathbf{h} \rangle > \frac{1}{2} \eta^{-1}\right) &\geq \mathbb{P}\left(\frac{\langle \mathbf{x}_*, \mathbf{h} \rangle}{\|\mathbf{h}_\perp\|_2} v - \frac{\eta^{-1}}{\|\mathbf{h}_\perp\|_2} v^{-1} > g\right) \\ &= \left(\frac{1}{2} + \frac{\alpha}{2\sqrt{\alpha^2 + 1}}\right) \exp\left(-\frac{\eta^{-1}}{\|\mathbf{h}_\perp\|_2 \sqrt{\alpha^2 + 1} + \alpha}\right). \end{aligned}$$

Then by rewriting (19) as $\frac{(t\eta)^{-2}}{\|\mathbf{h}_\perp\|_2^2} \leq 1 + \delta^{-2} + \alpha^2$ we have

$$\mathbb{P}\left(\langle \mathbf{a} \mathbf{a}^* \mathbf{x}_*, \mathbf{h} \rangle > \frac{1}{2} \eta^{-1}\right) \geq \left(\frac{1}{2} + \frac{\alpha}{2\sqrt{\alpha^2 + 1}}\right) \exp\left(-\frac{\sqrt{1 + \delta^{-2} + \alpha^2}}{\sqrt{\alpha^2 + 1} + \alpha} t\right).$$

The fact that $\mathbf{h} \in \mathcal{C}'_\delta$, guarantees that $\alpha \geq -\sqrt{\delta^{-2} - 1}$. Since $\frac{\alpha}{\sqrt{\alpha^2 + 1}}$ and $-\frac{\sqrt{1 + \delta^{-2} + \alpha^2}}{\sqrt{\alpha^2 + 1} + \alpha}$ are both increasing in α , we obtain

$$\mathbb{P}\left(\langle \mathbf{a} \mathbf{a}^* \mathbf{x}_*, \mathbf{h} \rangle > \frac{1}{2} \eta^{-1}\right) \geq \left(\frac{1}{2} - \frac{\sqrt{1 - \delta^2}}{2}\right) \exp\left(-\frac{\sqrt{2\delta^{-2}}}{\sqrt{\delta^{-2}} - \sqrt{\delta^{-2} - 1}} t\right),$$

$$\begin{aligned}
 &= \left(\frac{1}{2} - \frac{\sqrt{1-\delta^2}}{2} \right) \exp \left(-\frac{\sqrt{2}}{1-\sqrt{1-\delta^2}} t \right) \\
 &\geq \left(\frac{1}{2} - \frac{\sqrt{1-\delta^2}}{2} \right) \exp \left(-2\sqrt{2}\delta^{-2}t \right).
 \end{aligned}$$

The above lower bound is the smaller one of the two considered cases and thus the proof is complete. \square

Lemma 6. *Let $v \sim \text{Rayleigh}(1)$ and $g \sim \text{Normal}(0, 1)$ be independent random variables. Then we have*

$$\mathbb{P}(\alpha v + \beta v^{-1} > g) = \begin{cases} 1 - \frac{\sqrt{\alpha^2+1}-\alpha}{2\sqrt{\alpha^2+1}} e^{-\beta(\alpha+\sqrt{\alpha^2+1})} & \text{for } \beta \geq 0 \\ \frac{\sqrt{\alpha^2+1}+\alpha}{2\sqrt{\alpha^2+1}} e^{\beta/(\alpha+\sqrt{\alpha^2+1})} & \text{for } \beta < 0. \end{cases}$$

for all $\alpha, \beta \in \mathbb{R}$.

Proof. We denote the standard normal cumulative distribution function and its derivative by $\Phi(\cdot)$ and $\phi(\cdot)$, respectively. Let $F(\beta) = \mathbb{P}(\alpha v + \beta v^{-1} > \gamma) = \mathbb{E}\Phi(\alpha v + \beta v^{-1})$. By Leibniz’s rule we have

$$\begin{aligned}
 F'(\beta) &= \mathbb{E}(\phi(\alpha v + \beta v^{-1}) v^{-1}) \\
 &= \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{1}{2}(\alpha v + \beta v^{-1})^2} v^{-1} \cdot v e^{-\frac{1}{2}v^2} dv \\
 &= \frac{e^{-\alpha\beta}}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{1}{2}((\alpha^2+1)v^2 + \beta^2 v^{-2})} dv.
 \end{aligned}$$

Now let $G(\beta) = \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{1}{2}((\alpha^2+1)v^2 + \beta^2 v^{-2})} dv$, so that $F'(\beta) = e^{-\alpha\beta} G(\beta)$. Using Leibniz’s rule again, we can write

$$\begin{aligned}
 G'(\beta) &= -\frac{\beta}{\sqrt{2\pi}} \int_0^\infty v^{-2} e^{-\frac{1}{2}((\alpha^2+1)v^2 + \beta^2 v^{-2})} dv \\
 &= -\frac{\sqrt{\alpha^2+1}}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{1}{2}((\alpha^2+1)u^2 + \beta^2 u^{-2})} du \\
 &= -\sqrt{\alpha^2+1} G(\beta),
 \end{aligned}$$

where the second line follows from the change of variable $v = \frac{\beta}{\sqrt{\alpha^2+1}} u^{-1}$. It is straightforward to show that $G(0) = \frac{1}{2\sqrt{\alpha^2+1}}$. A simple integration then yields

$$G(\beta) = G(0)e^{-\beta\sqrt{\alpha^2+1}} = \frac{1}{2\sqrt{\alpha^2+1}} e^{-\beta\sqrt{\alpha^2+1}}$$

for $\beta \geq 0$, and since $G(\beta)$ is even for all β we have

$$G(\beta) = \frac{1}{2\sqrt{\alpha^2+1}} e^{-|\beta|\sqrt{\alpha^2+1}}.$$

It then follows that

$$F'(\beta) = \frac{1}{2\sqrt{\alpha^2 + 1}} e^{-(\alpha\beta + |\beta|\sqrt{\alpha^2 + 1})}$$

Integrating again we obtain

$$F(\beta) = \begin{cases} F(0) + \frac{\sqrt{\alpha^2 + 1} - \alpha}{2\sqrt{\alpha^2 + 1}} \left(1 - e^{-\beta(\alpha + \sqrt{\alpha^2 + 1})}\right) & , \beta \geq 0 \\ F(0) - \frac{\sqrt{\alpha^2 + 1} + \alpha}{2\sqrt{\alpha^2 + 1}} \left(1 - e^{-\beta(\alpha - \sqrt{\alpha^2 + 1})}\right) & , \beta < 0. \end{cases}$$

We can calculate $F(0)$ as

$$\begin{aligned} F(0) &= \mathbb{P}(\alpha v > g) \\ &= \begin{cases} \frac{1}{2} + \frac{1}{2} \mathbb{P}(\alpha^2 v^2 > g^2) & \text{for } \alpha \geq 0 \\ \frac{1}{2} \mathbb{P}(\alpha^2 v^2 < g^2) & \text{for } \alpha < 0 \end{cases} \\ &= \begin{cases} \frac{1}{2} + \frac{1}{2} \mathbb{P}\left(\frac{\alpha^2}{\alpha^2 + 1} > \frac{g^2}{v^2 + g^2}\right) & \text{for } \alpha \geq 0 \\ \frac{1}{2} - \frac{1}{2} \mathbb{P}\left(\frac{\alpha^2}{\alpha^2 + 1} > \frac{g^2}{v^2 + g^2}\right) & \text{for } \alpha < 0 \end{cases} \\ &= \frac{\sqrt{\alpha^2 + 1} + \alpha}{2\sqrt{\alpha^2 + 1}}, \end{aligned}$$

where the last line follows from the fact that $\frac{g}{\sqrt{v^2 + g^2}}$ has a uniform distribution over $[-1, 1]$. Replacing $F(0)$ in the expression of $F(\beta)$ and straightforward simplifications yield the desired result. \square

References

- [1] S. Bahmani and J. Romberg. Efficient compressive phase retrieval with constrained sensing vectors. In *Advances in Neural Information Processing Systems 28 (NIPS'15)*, pages 523–531. Curran Associates, Inc., Dec. 2015.
- [2] S. Bahmani and J. Romberg. Sketching for simultaneously sparse and low-rank covariance matrices. In *Proc. IEEE CAMSAP*, Cancun, Mexico, 2015.
- [3] S. Bahmani and J. Romberg. Solving equations of random convex functions via anchored regression. preprint [arXiv:1702.05327](https://arxiv.org/abs/1702.05327) [cs.LG], 2017.
- [4] R. G. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28(3):253–263, 2008. [MR2453366](https://doi.org/10.1007/s00365-008-0366-6)
- [5] S. R. Becker, E. J. Candès, and M. C. Grant. Templates for convex cone problems with applications to sparse signal recovery. *Mathematical Programming Computation*, 3(3):165–218, 2011. [MR2833262](https://doi.org/10.1007/s12532-011-0020-2)
- [6] T. T. Cai and A. Zhang. Rop: Matrix recovery via rank-one projections. *Ann. Statist.*, 43(1):102–138, 02 2015. [MR3285602](https://doi.org/10.1214/14-AOS1253)
- [7] E. J. Candès and X. Li. Solving quadratic equations via PhaseLift when there are about as many equations as unknowns. *Foundations of Computational Mathematics*, 14(5):1017–1026, 2014. [MR3260258](https://doi.org/10.1007/s00034-014-0528-2)

- [8] E. J. Candès, T. Strohmer, and V. Voroninski. PhaseLift: Exact and stable signal recovery from magnitude measurements via convex programming. *Communications on Pure and Applied Mathematics*, 66(8):1241–1274, 2013. [MR3069958](#)
- [9] E. J. Candès, X. Li, and M. Soltanolkotabi. Phase retrieval from coded diffraction patterns. *Applied and Computational Harmonic Analysis*, 39(2): 277–299, 2015. [MR3352016](#)
- [10] E. J. Candès, X. Li, and M. Soltanolkotabi. Phase retrieval via Wirtinger flow: Theory and algorithms. *Information Theory, IEEE Transactions on*, 61(4):1985–2007, Apr. 2015. [MR3332993](#)
- [11] Y. Chen and E. Candès. Solving random quadratic systems of equations is nearly as easy as solving linear systems. In *Advances in Neural Information Processing Systems 28 (NIPS’15)*, pages 739–747. Curran Associates, Inc., Dec. 2015. [MR3628877](#)
- [12] Y. Chen, Y. Chi, and A. Goldsmith. Exact and stable covariance estimation from quadratic sampling via convex programming. *IEEE Trans. Inform. Theory*, 61(7):4034–4059, 2015. [MR3367819](#)
- [13] L. Devroye, L. Györfi, and G. Lugosi. *A probabilistic theory of pattern recognition*, volume 31. Springer Science & Business Media, 2013. [MR1383093](#)
- [14] J. R. Fienup. Phase retrieval algorithms: a comparison. *Applied Optics*, 21(15):2758–2769, Aug 1982.
- [15] R. W. Gerchberg and W. O. Saxton. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik*, 35(2): 237–346, 1972.
- [16] T. Goldstein and C. Studer. Convex phase retrieval without lifting via PhaseMax. In D. Precup and Y. W. Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 1273–1281, International Convention Centre, Sydney, Australia, 06–11 Aug 2017. PMLR. earlier version: [arXiv:1610.07531 \[cs.IT\]](#).
- [17] P. Hand and V. Voroninski. Compressed sensing from phaseless gaussian measurements via linear programming in the natural parameter space. preprint [arXiv:1611.05985](#), 2016.
- [18] P. Hand and V. Voroninski. Corruption robust phase retrieval via linear programming. preprint [arXiv:1612.03547](#), 2016.
- [19] P. Hand and V. Voroninski. An elementary proof of convex phase retrieval in the natural parameter space via the linear program PhaseMax. preprint [arXiv:1611.03935](#), 2016.
- [20] M. Iwen, A. Viswanathan, and Y. Wang. Robust sparse phase retrieval made easy. *Applied and Computational Harmonic Analysis*, 42(1):135–142, 2017. earlier version: [arXiv:1410.5295 \[math.NA\]](#). [MR3574564](#)
- [21] K. Jaganathan, Y. C. Eldar, and B. Hassibi. Phase retrieval: An overview of recent developments. In A. Stern, editor, *Optical Compressive Imaging*, chapter 13, pages 263–296. CRC Press, 2016. preprint [arXiv:1510.07713 \[cs.IT\]](#). [MR3674475](#)

- [22] X. Li and V. Voroninski. Sparse signal recovery from quadratic measurements via convex programming. *SIAM Journal on Mathematical Analysis*, 45(5):3019–3033, 2013. [MR3106479](#)
- [23] Y. M. Lu and G. Li. Phase transitions of spectral initialization for high-dimensional nonconvex estimation. preprint [arXiv:1702.06435 \[cs.IT\]](#), 2017.
- [24] J. Matoušek. *Lectures on discrete geometry*, volume 212 of *Graduate Texts in Mathematics*. Springer-Verlag New York, 2002. [MR1899299](#)
- [25] P. Netrapalli, P. Jain, and S. Sanghavi. Phase retrieval using alternating minimization. In *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, pages 2796–2804, 2013. [MR3385838](#)
- [26] H. Ohlsson, A. Yang, R. Dong, and S. Sastry. CPRL—an extension of compressive sensing to the phase retrieval problem. In *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, pages 1367–1375, 2012.
- [27] S. Oymak, A. Jalali, M. Fazel, Y. Eldar, and B. Hassibi. Simultaneously structured models with application to sparse and low-rank matrices. *Information Theory, IEEE Transactions on*, 61(5):2886–2908, 2015. [MR3342310](#)
- [28] R. Pedarsani, K. Lee, and K. Ramchandran. Phasecode: Fast and efficient compressive phase retrieval based on sparse-graph codes. In *Communication, Control, and Computing (Allerton), 52nd Annual Allerton Conference on*, pages 842–849, Sep. 2014. Extended preprint [arXiv:1408.0034 \[cs.IT\]](#).
- [29] N. Sauer. On the density of families of sets. *Journal of Combinatorial Theory, Series A*, 13(1):145–147, 1972. [MR0307902](#)
- [30] P. Schniter and S. Rangan. Compressive phase retrieval via generalized approximate message passing. *IEEE Transactions on Signal Processing*, 63(4):1043–1055, Feb. 2015. [MR3311635](#)
- [31] S. Shelah. A combinatorial problem; stability and order for models and theories in infinitary languages. *Pacific J. Math.*, 41(1):247–261, 1972. [MR0307903](#)
- [32] M. Soltanolkotabi. Structured signal recovery from quadratic measurements: Breaking sample complexity barriers via nonconvex optimization. preprint [arXiv:1702.06175 \[cs.LG\]](#), 2017.
- [33] J. Sun, Q. Qu, and J. Wright. A geometric analysis of phase retrieval. In *2016 IEEE International Symposium on Information Theory (ISIT)*, pages 2379–2383, July 2016. full-length version: [arXiv:1602.06664 \[cs.IT\]](#).
- [34] V. N. Vapnik. *Statistical learning theory*. Wiley, 1998. [MR1641250](#)
- [35] V. N. Vapnik and A. Y. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability & Its Applications*, 16(2):264–280, 1971. [MR0288823](#)
- [36] I. Waldspurger, A. d’Aspremont, and S. Mallat. Phase recovery, MaxCut and complex semidefinite programming. *Mathematical Programming*, 149(1–2):47–81, 2015. [MR3300456](#)
- [37] G. Wang and G. Giannakis. Solving random systems of quadratic equations via truncated generalized gradient flow. In *Advances in Neural Information*

- Processing Systems 29 (NIPS'16)*, pages 568–576. Curran Associates, Inc., 2016.
- [38] G. Wang, L. Zhang, G. B. Giannakis, M. Akçakaya, and J. Chen. Sparse phase retrieval via truncated amplitude flow. *preprint* [arXiv:1611.07641](https://arxiv.org/abs/1611.07641) [[cs.IT](#)], 2016.
- [39] Ç. Yapar, V. Pohl, and H. Boche. Fast compressive phase retrieval from fourier measurements. In *Acoustics, Speech and Signal Processing (ICASSP'15), IEEE International Conference on*, pages 3267–3271, April 2015.
- [40] D. Yin, K. Lee, R. Pedarsani, and K. Ramchandran. Fast and robust compressive phase retrieval with sparse-graph codes. In *2015 IEEE International Symposium on Information Theory (ISIT)*, pages 2583–2587, June 2015. [MR3658550](#)
- [41] Y. Yu, T. Wang, and R. J. Samworth. A useful variant of the Davis-Kahan theorem for statisticians. *Biometrika*, 102(2):315–323, 2015.
- [42] H. Zhang, Y. Chi, and Y. Liang. Provable non-convex phase retrieval with outliers: Median truncated wirtinger flow. In *Proceedings of The 33rd International Conference on Machine Learning (ICML'16)*, volume 48, pages 1022–1031, New York, NY, USA., 2016. JMLR W&CP.