Fast Compressive Phase Retrieval under Bounded Noise

Hongyang Zhang,[*1] **Shan You,**[*2,3] **Zhouchen Lin,**[†2,3] **Chao Xu**[2,3]
[1]Machine Learning Department, Carnegie Mellon University, U.S.A.
[2]Key Lab. of Machine Perception (MOE), School of EECS, Peking University, P. R. China
[3] Cooperative Medianet Innovation Center, Shanghai Jiao Tong University, P. R. China
hongyanz@cs.cmu.edu, youshan@pku.edu.cn, zlin@pku.edu.cn, xuchao@cis.pku.edu.cn

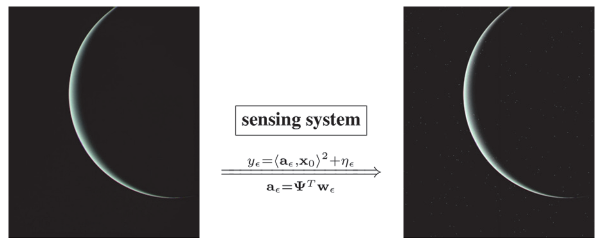
Abstract

We study the problem of recovering a t-sparse vector $\pm\mathbf{x}_0$ in $\mathbb{R}^n$ from m quadratic equations $\mathbf{y}_i = (\mathbf{a}_i^T\mathbf{x})^2$ with noisy measurements $\mathbf{y}_i$'s. This is known as the problem of compressive phase retrieval, and has been widely applied to X-ray diffraction imaging, microscopy, quantum mechanics, etc. The challenge is to design a a) *fast* and b) *noise-tolerant* algorithms with c) *near-optimal sample complexity*. Prior work in this direction typically achieved one or two of these goals, but none of them enjoyed the three performance guarantees simultaneously. In this work, with a particular set of sensing vectors $\mathbf{a}_i$'s, we give a provable algorithm that is robust to any bounded yet unstructured deterministic noise. Our algorithm requires roughly $\mathcal{O}(t)$ measurements and runs in $\mathcal{O}(tn\log(1/\epsilon))$ time, where ϵ is the error. This result advances the state-of-the-art work, and guarantees the applicability of our method to large datasets. Experiments on synthetic and real data verify our theory.

Introduction

Phase retrieval is a new research topic in machine learning, signal processing, and statistics. In machine learning and signal processing, the goal of phase retrieval is to recover a hidden signal $+\mathbf{x}$ or $-\mathbf{x}$ from a quadratic system $\{\mathbf{y}_i = (\mathbf{a}_i^T\mathbf{x})^2\}_{i=1}^m$ with known sensing vectors $\{\mathbf{a}_i\}_{i=1}^m$ and measurements $\{\mathbf{y}_i\}_{i=1}^m$. In statistics, the $\mathbf{x}$ indicates a set of parameters remaining estimated for some probability distribution. The problem becomes more challenging when the signal, or the unknown parameter vector $\mathbf{x}$, is t-sparse and only $m = \tilde{\mathcal{O}}(t)$[1] equations are available. This is known as the problem of *compressive phase retrieval*, and has been widely applied to X-ray diffraction imaging (Schniter and Rangan 2015), microscopy (Miao et al. 2008), quantum mechanics (Corbett 2006), and many other domains. See Figure 1 for an application of compressive phase retrieval in image sensing and restoration.

Despite a large amount of work on compressive phase retrieval, many fundamental problems remain unresolved.



(a) real scene (b) recovered image

Figure 1: Recovery results of our algorithm in the framework of compressive phase retrieval, where the high-resolution real image contains a crescent Uranus taken by Voyager.

One long-standing challenge is designing *fast* algorithms with *near-optimal sample complexity*. If there is no speed requirement, the recovery problem can be solved easily by dimension lifting approach, whose basic idea is to convert the problem of recovering a sparse signal $\pm\mathbf{x} \in \mathbb{R}^n$ to the problem of recovering a rank-one matrix $\mathbf{x}\mathbf{x}^T \in \mathbb{R}^{n\times n}$ by implementing nuclear norm minimization (Candès, Strohmer, and Voroninski 2013) or matrix completion (Candès, Li, and Soltanolkotabi 2015; Zhang, Lin, and Zhang 2016; Zhang et al. 2015b). While this method enjoys intriguing performance guarantee with near-optimal sample complexity, the computational cost is large (Bahmani and Romberg 2015). This definitely limits its applicability to large datasets.

When the measurements are noisy, the problem becomes even more challenging. Prior work in this direction mostly assumed certain statistical structures on the noise model, e.g., the noise is subject to Poisson distribution, and maximized the likelihood function corresponding to the structure. However, when no assumption is made on the nature of noise, the likelihood functions are unavailable and so these approaches do not work. In this paper, with a particular set of sensing vectors, we propose a provable algorithm that is *fast, noise-tolerant*, and performs well with *near-optimal sample complexity*.

Related Work

Lots of literatures have investigated the problem of general (a.k.a. non-sparse) phase retrieval (PR) in recent years.

---

*Equal contributions.

†Corresponding author.

[1]We denote by $\tilde{\mathcal{O}}(\cdot)$ the simplicity of $\mathcal{O}(\cdot)$ that omits the logarithm factor.

They proposed various algorithms with sample complexity of the same order as the ambient dimension, up to a logarithmic factor. Specifically, (Candès, Strohmer, and Voroninski 2013) proposed the dimension lifting. Although the method enjoys $\mathcal{O}(n \log n)$ sample complexity, the computational cost is too high to be applied to the real applications. To resolve the issue, (Candès, Li, and Soltanolkotabi 2015) designed Wirtinger Flow (WF) method (Candès, Li, and Soltanolkotabi 2015; Zhang and Liang 2016) with sample complexity as small as $\mathcal{O}(n \log n)$. In comparison with dimension lifting, the time complexity is only $\mathcal{O}(n^3 \log(1/\epsilon))$, which is significantly faster. Here ϵ indicates the algorithmic precision. Later, Truncated Wirtinger Flow (TWF) (Chen and Candès 2015) reduced the sample complexity to $\mathcal{O}(n)$ and time complexity to $\mathcal{O}(n^2 \log(1/\epsilon))$ by truncating large abnormal measurements.

The exploration of sparse structures (Zhang et al. 2015a; 2016) in the phase retrieval context (compressive phase retrieval, CPR) has drawn lots of attention as well. A natural approach is to inherit the techniques of general phase retrieval while imposing the sparsity constraint. Specifically, (Moravec, Romberg, and Baraniuk 2007) formulated the problem of sparse phase retrieval as an ℓ_1 minimization problem with a non-convex constraint. They empirically showed that the t-sparse signal can be successfully estimated by only $\mathcal{O}(t \log \frac{n}{t})$ measurements (Shechtman, Beck, and Eldar 2014). Unfortunately, these non-convex methods typically lack global convergence guarantees. To resolve the issue, some researchers proposed relaxing the non-convex problems to a convex one. Probably one of the most popular techniques is the sparse dimension lifting method. The method lifts the target vector $\mathbf{x}$ to a rank-1 sparse matrix $\mathbf{x}\mathbf{x}^T$ and do nuclear norm minimization to recover the hidden signal (Li and Voroninski 2013; Ohlsson et al. 2012). The sample complexity is as low as $\mathcal{O}(t^2 \log n)$ with Gaussian sensing vector (Li and Voroninski 2013). Recent work (Iwen, Viswanathan, and Wang 2015; Bahmani and Romberg 2015) further reduced the sample complexity to $\mathcal{O}(t \log \frac{n}{t})$ using constrained sensing vectors. They assumed that the measurement vectors lie in a random low-dimensional subspace and the recovery process can be decomposed into two steps, both of which are convex models. However, the computational cost of these methods is too large to be applied to high-dimensional signal processing.

Our Contributions

In this paper, we design a *fast, noise-tolerant* algorithm with *near-optimal sample complexity*. Our algorithm advances the state-of-the-art approaches in the following aspects:

- Regarding the sample complexity, our algorithm requires only $\tilde{\mathcal{O}}(t)$ measurements, which matches the information-theoretic limit up to a logarithm factor.
- The time complexity of our algorithm is $\mathcal{O}(tn \log(1/\epsilon))$ for error ϵ, which is significantly faster than the dimension lifting based methods. This result is comparable with the computational cost of compressive sensing problem, although the problem of compressive phase retrieval is more challenging than that of compressive sensing.
- Our algorithm is robust to any bounded yet unstructured noise, with provable intriguing performance guarantee. We show that the ℓ_2 error of our estimator decreases in the same order of $\mathcal{O}(1/m)$.

Table 1: Performance indexes in phase retrieval for t-sparse n-dimensional signals with error ϵ

	WF	TWF	TWF+Sparse
Noise	Possion	Possion	Sub-exponential
Sample	$\tilde{\mathcal{O}}(n)$	$\mathcal{O}(n)$	$\tilde{\mathcal{O}}(t^2)$
Time	$\tilde{\mathcal{O}}(n^3 \log(1/\epsilon))$	$\mathcal{O}(n^2 \log(1/\epsilon))$	ungiven*
	ECPR	SparsePR	Ours
Noise	Unstructured	Unstructured	**Unstructured**
Sample	$\tilde{\mathcal{O}}(t)$	$\tilde{\mathcal{O}}(t)$	$\boldsymbol{\tilde{\mathcal{O}}(t)}$
Time	ungiven*	ungiven*	$\boldsymbol{\tilde{\mathcal{O}}(tn \log(1/\epsilon))}$

The comparison of all these three aspects in various algorithms is presented in Table 1. The novelty of our algorithm is a refinement step inspired from disagreement based active learning. This step enables us to exactly recover the underlying signal $\mathbf{x}_0$ with finite samples in the noise-free setting, and significantly reduces the error of our estimator in the noisy scenario.

Preliminaries

Problem Description: The problem of phase retrieval derives from solving a linear system. In the noiseless case, given m linear equations $b_i = \mathbf{a}_i^T \mathbf{x}_0, \ i = 1, 2, ..., m$, where $\mathbf{a}_i$'s $\in \mathbb{R}^n$ are sensing vectors, y_i's > 0 are measurements and $\mathbf{x}_0 \in \mathbb{R}^n$ is the unknown target signal, the problem of phase retrieval assumes that the phases/signs of b_i's are unavailable and aims at recovering $\pm\mathbf{x}_0$ from the quadratic equations $b_i^2 = y_i = (\mathbf{a}_i^T \mathbf{x}_0)^2, \ i = 1, 2, ..., m$. This problem becomes more challenging when the underlying vector $\mathbf{x}_0$ is t-sparse and the number of equations m is far less than the ambient dimension n. This is known as the problem of *Compressive Phase Retrieval*. Our goal is to design an algorithm with sample complexity $\tilde{\mathcal{O}}(t)$ and time complexity $\tilde{\mathcal{O}}(tn)$, advancing the state-of-the-art results.

Noise Model: In the noisy case, an adversary can add any bounded yet unstructured noise $\{\eta_i\}_{i=1}^m$ to the clean measurements, namely,

$$y_i = (\mathbf{a}_i^T \mathbf{x}_0)^2 + \eta_i > 0, \ i = 1, 2, ..., m. \tag{1}$$

Here η_i's are deterministic constants, and the magnitude of noise is bounded such that $\|[\eta_1; ...; \eta_m]\|_2 \leq \eta$. Beyond that, no other assumptions are made on the nature of noise and so even adversarial noise is allowed. We investigate the possibility of recovering the hidden signal in this worst case.

Sampling Model: To efficiently sense the sparse signal, we study the problem of noisy compressive phase retrieval with a particular set of sensing vectors $\mathbf{a}_i$'s of the form

$$\mathbf{a}_i = \mathbf{\Psi}^T \mathbf{w}_i \in \mathbb{R}^n, \tag{2}$$

where $\mathbf{\Psi} \in \mathbb{R}^{d \times n}$ and $\mathbf{w}_i$'s $\in \mathbb{R}^d$ are known a priori. Specifically, we assume that $\mathbf{w}_i$'s are drawn i.i.d. from $\mathcal{N}(\mathbf{0}, \boldsymbol{I})$ of $\mathbb{R}^d$, and $\mathbf{\Psi}$ is a restricted isometry matrix such that

$$(1 - \delta_{2t})\|\mathbf{x}\|_2^2 \leq \|\mathbf{\Psi}\mathbf{x}\|_2^2 \leq (1 + \delta_{2t})\|\mathbf{x}\|_2^2, \tag{3}$$

for a constant $\delta_{2t} \in [0,1]$ and all $2t$-sparse vectors $\mathbf{x}$. Therefore, the sensing vectors $\{\mathbf{a}_i\}_{i=1}^m$ lie on a fixed d-dimensional subspace — the row space of matrix $\mathbf{\Psi}$. Such a sampling scheme is realistic in lots of interesting applications (Bahmani and Romberg 2015):

- In the application of imaging through scattering media, one usually models the optical transfer function by a random matrix, a.k.a. the transfer matrix. In this scenario, different LED excitations lead to illumination patterns that are in the row space of the transfer matrix (Bertolotti et al. 2012; Liutkus et al. 2014).
- In the application of estimating convariance matrix, convariance matrices that are simultaneously low-rank and sparse can be sketched by sensing vectors that lie in a low-dimensional subspace (Chen, Chi, and Goldsmith 2015).
- In the application of one-bit camera (Duarte et al. 2008), one modulates the sparse signal $\mathbf{x}_0$ in the frequency domain and obtain $\mathbf{F}^T\mathbf{D}\mathbf{F}\mathbf{x}_0$, where $\mathbf{F}$ is the unitary discrete Fourier transform matrix and $\mathbf{D}$ is a diagonal matrix with Rademacher diagonal entries. In this setting, we can characterize the sensing vectors by setting $\mathbf{\Psi} = \mathbf{F}$ in our sampling model.

Main Results

In this section, we propose a fast algorithm for robustly recovering a sparse signal from phase-quantized measurements, and develop our main theoretical contributions under bounded yet *unstructured* noise. Our algorithm is robust and requires few measurements: With $\mathcal{O}(t\,\mathrm{poly}(\log n))$ samples the algorithm suffices to converge to the underlying t-sparse signal in $\mathbb{R}^n$ with provable small error. The time complexity of our algorithm is $\mathcal{O}(tn\log(1/\epsilon))$, which is significantly faster than dimension lifting based methods (Candès, Li, and Soltanolkotabi 2015; Candès, Strohmer, and Voroninski 2013; Bahmani and Romberg 2015).

Recovery Algorithm

Given a sensing vector $\mathbf{a} \in \mathbb{R}^n$ and the measurement $y = \langle \mathbf{a}, \mathbf{x}_0\rangle^2$, to recover the underlying t-sparse signal $\mathbf{x}_0 \in \mathbb{R}^n$, we maximize the covariance between y and $\langle \mathbf{a}, \mathbf{x}\rangle^2$ over the candidate space $\mathcal{K} = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_0 \le t, \|\mathbf{x}\|_2 \le 1\}$. Namely, we are interested in the optimization problem

$$\max_{\mathbf{x}\in\mathcal{K}} \mathbb{E}(\langle \mathbf{a}, \mathbf{x}_0\rangle\langle \mathbf{a}, \mathbf{x}\rangle)^2. \tag{4}$$

Maximizing (4) over the candidate space is a typical non-convex program which cannot be solved in polynomial time. To mitigate the computational issue, we take into account the sampling strategy of (2). Roughly speaking, our main procedures are a) projecting the high-dimensional signal of $\mathbb{R}^n$ to a low-dimensional space $\mathbb{R}^d = \mathbb{R}^{\tilde{O}(t)}$ by our sampling scheme, and optimizing model (4) in the low-dimensional space without the constraint set $\mathcal{K}$, which can be done via a closed-form solution, and b) recovering the underlying sparse signal by compressive sensing techniques. To this end, our algorithm has an initialization step and a refinement step for procedure a), and a step of sparse recovery for procedure b).

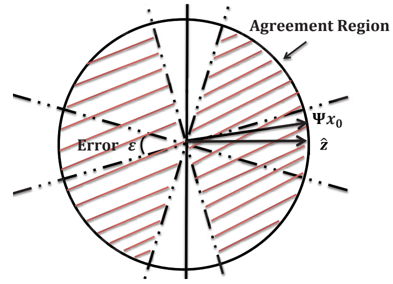


Figure 2: Illustration of agreement region of $\widehat{\mathbf{z}}$ ($\mathsf{AgrReg}(\widehat{\mathbf{z}})$) and $\mathbf{\Psi}\mathbf{x}_0$ modulo a global sign. In the red shadow area, $\mathrm{sign}(\mathbf{w}_i^T\mathbf{\Psi}\mathbf{x}_0) = \pm\mathrm{sign}(\mathbf{w}_i^T\widehat{\mathbf{z}})$ for all $\mathbf{a}_i$'s. It is not hard to see that the agreement region becomes larger as $\widehat{\mathbf{z}}$ gets closer to $\mathbf{\Psi}\mathbf{x}_0$.

Initialization Step: In the low-dimensional space $\mathbb{R}^d$, we approximate the expectation in the objective function of (4) with the empirical average. By our sampling oracle of (2), the objective function can be approximately rewritten as

$$\begin{aligned}\mathbb{E}y\langle \mathbf{a}, \mathbf{x}\rangle^2 &\approx \frac{1}{m}\sum_{i=1}^m y_i\langle \mathbf{\Psi}^T\mathbf{w}_i, \mathbf{x}\rangle^2 \\ &= \frac{1}{m}\sum_{i=1}^m y_i\langle \mathbf{w}_i, \mathbf{\Psi}\mathbf{x}\rangle^2 \triangleq \frac{1}{m}\sum_{i=1}^m y_i\langle \mathbf{w}_i, \mathbf{z}\rangle^2.\end{aligned} \tag{5}$$

Fortunately, the problem of maximizing (5) has a closed-form solution, which is the leading eigenvector of matrix $\frac{1}{m}\sum_{i=1}^m y_i\mathbf{w}_i\mathbf{w}_i^T$. Our theoretical analysis shows that the initialization step outputs a solution of error ϵ by a high probability, provided that the sample size is $\mathcal{O}(\epsilon^{-2}t\,\mathrm{poly}(\log n))$.
Recovery via Refinement: Although the initialization step enjoys solid theoretical guarantee, the sample complexity of $\mathcal{O}(\epsilon^{-2}t\,\mathrm{poly}(\log n))$ implies infinitely many measurements when the error ϵ goes to zero, even in the noise-free setting. To alleviate this issue, we are inspired from the fact that the problem of phase retrieval is as easy as solving a linear system, *provided that the phases of those measurements are known a priori*. To exploit this, we note that the output of initialization step already contains certain label information. Specifically, we first run the initialization step and obtain a solution $\widehat{\mathbf{z}}$ with small *constant* error, which implies a small constant angle between $\pm\mathbf{\Psi}\mathbf{x}_0$ and $\widehat{\mathbf{z}}$. For this step, we only need $\mathcal{O}(t\,\mathrm{poly}(\log n))$ samples. We then safely label all $\mathbf{w}_i$'s lying in the agreement region of $\mathbf{\Psi}\mathbf{x}_0$ and $\widehat{\mathbf{z}}$ with correct phases modulo a global sign (See Figure 2), and solve the resultant linear system by the least squares methods.
Sparse Recovery: As the operator $\mathbf{\Psi}$ is an almost isometric mapping from $\mathbb{R}^n$ to $\mathbb{R}^d$ for any t-sparse signal, we can hopefully recover the signal in the original $\mathbb{R}^n$ space by standard compressed sensing techniques. Our three-stage approach is summarized in Algorithm 1.
Time Complexity: The initialization step computes the leading eigenvector of a $d \times d$ matrix and runs in $\mathcal{O}(d^2)$ time. The refinement step solves a least squares problem, which requires at most $\mathcal{O}(d^3)$ time. For the step of sparse recovery, solving an ℓ_1 norm minimization problem needs

$\mathcal{O}(dn\log(1/\epsilon))$ time by alternating direction method of multipliers (Hong and Luo 2012), where ϵ indicates the error. As $d = \mathcal{O}(t\ \mathrm{poly}(\log n)) \ll n$, the running time of our algorithm is $\tilde{\mathcal{O}}(tn\mathrm{poly}(\log n)\log(1/\epsilon))$, which is significantly faster than dimension lifting methods for sparse recovery (Bahmani and Romberg 2015).

Algorithm 1 Robust Recovery of Sparse Signal by Phase Retrieval

---

Input: A set of sensing vectors $\{\mathbf{a}_i \in \mathbb{R}^n : i = 1, 2, ..., m\}$ drawn according to sampling oracle (2), where $\mathbf{w}_i$'s $\in \mathbb{R}^d$ are i.i.d. sampled from the Gaussian distribution $\mathcal{N}(\mathbf{0}, \boldsymbol{I})$; A set of measurements $\{y_i : i = 1, 2, ..., m\}$ generated by (1).

Initialize: Solve (6) and obtain the optimal $\widehat{\mathbf{z}}$:

$$\widehat{\mathbf{z}} := \operatorname*{argmax}_{\mathbf{z}} \frac{1}{m}\sum_{i=1}^{m} y_i(\mathbf{w}_i^T\mathbf{z})^2, \ \text{ s.t. } \|\mathbf{z}\|_2 \le 1. \quad (6)$$

Refine:

1. Determine the index set $\Omega = \{i : \mathbf{w}_i \in \mathsf{AgrReg}(\widehat{\mathbf{z}})\}$ according to $\mathsf{AgrReg}(\widehat{\mathbf{z}})$ in Figure 2 or (9).

2. Construct design matrix and b, where $\mathbf{W}_{\Omega:}$ is to extract the rows of $\mathbf{W}$ in the index set Ω:
$\mathbf{W} = \left[\mathbf{w}_1^T; \ldots; \mathbf{w}_m^T\right]_{\Omega:}; \mathbf{b} = \left[\sqrt{y_1}; \sqrt{y_2}; \ldots; \sqrt{y_m}\right]_{\Omega:}$.

3. Allocate phases to vector $\mathbf{b}$ according to $\widehat{\mathbf{z}}$, modulo a global sign.

4. Solve the least squares problem

$$\widetilde{\mathbf{z}} := \operatorname*{argmin}_{\mathbf{z}} \|\mathbf{W}\mathbf{z} - \mathbf{b}\|_2^2, \quad (7)$$

Recover: Sparse recovery by

$$\widetilde{\mathbf{x}} := \operatorname*{argmin}_{\mathbf{x}} \|\mathbf{x}\|_1, \ \text{ s.t. } \|\widetilde{\mathbf{z}} - \boldsymbol{\Psi}\mathbf{x}\|_2 \le \mathcal{O}(\sqrt{\eta/m}). \quad (8)$$

Output: Sparse estimator $\widetilde{\mathbf{x}}$.

Recovery Guarantee

The analysis on the optimization problems (6), (7), and (8) leads to the following guarantee on Algorithm 1.

Theorem 1. *Let $\{\mathbf{a}_i\}_{i=1}^m \in \mathbb{R}^n$ be random vectors sampled i.i.d. according to the sampling model in the preliminaries section, and $d \ge Ct\log(n/t)$. Assume that the measurements $\{y_i\}_{i=1}^m$ follow the model $y_i = \langle\boldsymbol{\Psi}^T\mathbf{w}_i, \mathbf{x}_0\rangle^2 + \eta_i$. Let $\mathbf{x}_0 \in \mathbb{R}^n$ be the underlying t-sparse signal such that $\|\boldsymbol{\Psi}\mathbf{x}_0\| = 1$, and $m \ge c_0 d\log^4\left(\frac{d}{\delta\epsilon}\right)$. Then with probability at least $1-\delta$, the output $\widetilde{\mathbf{x}}$ of Algorithm 1 satisfies*

$$\min_{\pm\mathbf{x}_0} \|\widetilde{\mathbf{x}} \pm \mathbf{x}_0\|_2 \le \sqrt{\eta}/m.$$

Theorem 4 implies strong guarantee on the recoverability of Algorithm 1: The algorithm can approximately recover the underlying signal with small error under the adversarial noise, if the sample size is of order $\mathcal{O}(t\ \mathrm{poly}(\log n))$. When there is no noise, our algorithm exactly recovers the target vector modulo a global sign.

We compare our result with several related line of research in the prior work. The first is the paper of Candès et al. (Candès, Li, and Soltanolkotabi 2015), that gives sample complexity bound in the order of $\mathcal{O}(n\log n)$ in the non-sparse case, via Wirtinger flow. Later, Chen et al. (Chen and Candès 2015) improved the result to $\mathcal{O}(n)$ by truncating those measurements of large magnitude. The noise model in their work is restricted to the Possion distribution. In comparison, our work improves over these results in two-fold: a) we take into account the sparsity structure of the underlying signal, reducing the sample complexity to only having a logarithm dependence on the ambient dimension n; b) our noise model may have arbitrary structure, so even adversarial errors are allowed.

There are several recent work that studies the problem of compressive phase retrieval. One of them is the paper of Bahmani and Romberg (Bahmani and Romberg 2015), which requires $\mathcal{O}(t\ \mathrm{poly}(\log n))$ samples and much time to approximately recover the target signal under the same sampling model as ours. Compare to this, our algorithm has comparable sample complexity, while our time complexity $\tilde{\mathcal{O}}(tn)$ is significantly lower than that of their approach. Dong et al. (Yin et al. 2015) used sparse-graph codes to formulate a PhaseCode-style algorithm with similar complexity; however, its measurement matrix is designed particularly and does not cohere with the real applications, such as imaging through scattering media. Cai et al. (Cai, Li, and Ma 2015) proposed applying Wirtinger flow based method to the problem of sparse phase retrieval, with sample complexity $\mathcal{O}(t^2\ \mathrm{poly}(\log n))$. In comparison, our result improves over theirs in the order of t.

Proof Outline

Consider the optimization problem in the initialization step:

$$\max_{\|\mathbf{z}\|_2\le 1} \frac{1}{m}\sum_{i=1}^{m}((\mathbf{w}_i^T\boldsymbol{\Psi}\mathbf{x}_0)^2 + \eta_i)(\mathbf{w}_i^T\mathbf{z})^2.$$

Denote by $f_{\mathbf{x}_0}(\mathbf{z})$ the objective function whose subscript $\mathbf{x}_0$ indicates that f is a random function with distribution depending on $\mathbf{x}_0$. We claim that for any $\mathbf{z}$ that is far away from $\pm\boldsymbol{\Psi}\mathbf{x}_0$, $f_{\mathbf{x}}(\mathbf{z})$ cannot be small. To see this, we begin with an analysis on the expectation of the objective function $f_{\mathbf{x}_0}(\mathbf{z})$.

Lemma 2 (Expectations)**.** *Let $\|\boldsymbol{\Psi}\mathbf{x}_0\|_2 = 1$. Suppose that $\{\mathbf{w}_i\}_{i=1}^m$ are drawn i.i.d. from the Gaussian distribution $\mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Then for any $\mathbf{z} \in \mathbb{R}^d$, we have*

$$\mathbb{E}f_{\mathbf{x}_0}(\mathbf{z}) = \|\mathbf{z}\|_2^2\left(1 + \frac{2\mathbf{z}^T\boldsymbol{\Psi}\mathbf{x}_0\mathbf{x}_0^T\boldsymbol{\Psi}\mathbf{z}}{\|\mathbf{z}\|_2^2} + \frac{1}{m}\sum_{i=1}^{m}\eta_i\right),$$

In particular, if we further have $\|\mathbf{z}\|_2 = 1$, then $\mathbb{E}[f_{\mathbf{x}_0}(\boldsymbol{\Psi}\mathbf{x}_0) - f_{\mathbf{x}_0}(\mathbf{z})] \ge \frac{1}{2}\min_{\pm\mathbf{x}_0}\|\mathbf{z} \pm \boldsymbol{\Psi}\mathbf{x}_0\|_2^2$.

Proof. Note that $\mathbb{E}(\mathbf{w}_i^T\boldsymbol{\Psi}\mathbf{x}_0)^2\mathbf{w}_i\mathbf{w}_i^T = \boldsymbol{I} + 2\boldsymbol{\Psi}\mathbf{x}_0\mathbf{x}_0^T\boldsymbol{\Psi}^T$. So we have

$$\begin{aligned}\mathbb{E}(\mathbf{w}_i^T\psi\mathbf{x}_0)^2(\mathbf{w}_i^T\mathbf{z})^2 &= \mathbf{z}^T[\mathbb{E}(\mathbf{w}_i^T\boldsymbol{\Psi}\mathbf{x}_0)^2\mathbf{w}_i\mathbf{w}_i^T]\mathbf{z} \\ &= \|\mathbf{z}\|_2^2 + 2\mathbf{z}^T\boldsymbol{\Psi}\mathbf{x}_0\mathbf{x}_0^T\boldsymbol{\Psi}^T\mathbf{z}.\end{aligned}$$

We also have $\mathbb{E}(\mathbf{w}_i^T\mathbf{z})^2 = \|\mathbf{z}\|_2^2$. So

$$\mathbb{E}f_{\mathbf{x}_0}(\mathbf{z}) = \|\mathbf{z}\|_2^2\left(1 + \frac{2\mathbf{z}^T\boldsymbol{\Psi}\mathbf{x}_0\mathbf{x}_0^T\boldsymbol{\Psi}^T\mathbf{z}}{\|\mathbf{z}\|_2^2} + \frac{1}{m}\sum_{i=1}^{m}\eta_i\right).$$

Furthermore, if $\|\mathbf{z}\|_2 = 1$, then

$$\mathbb{E}[f_{\mathbf{x}_0}(\mathbf{\Psi x}_0) - f_{\mathbf{x}_0}(\mathbf{z})] = 2\langle \mathbf{\Psi x}_0, \mathbf{\Psi x}_0\rangle^2 - 2\langle \mathbf{z}, \mathbf{\Psi x}_0\rangle^2$$
$$= 2\sin^2\left(\min_{\pm\mathbf{x}_0} \theta(\mathbf{z}, \pm\mathbf{\Psi x}_0)\right)$$
$$\geq \frac{1}{2}\min_{\pm\mathbf{x}_0} \theta^2(\mathbf{z}, \pm\mathbf{\Psi x}_0) \geq \frac{1}{2}\min_{\pm\mathbf{x}_0} \|\mathbf{z} \pm \mathbf{\Psi x}_0\|_2^2.$$

□

Lemma 2 asserts that the optimal solution $\widehat{\mathbf{z}}$ to the expected form of (6)

$$\max_{\mathbf{z}} \ \mathbb{E}\left[\frac{1}{m}\sum_{i=1}^{m} y_i(\mathbf{w}_i^T\mathbf{z})^2\right], \ \text{s.t.} \ \|\mathbf{z}\|_2 \leq 1,$$

is exactly the desired vector $\mathbf{\Psi x}_0$ modulo a global sign. By concentration of measure, the output of initialization step will be sufficiently close to $\mathbf{\Psi x}_0$ when the sample complexity m is large enough. To this end, the following result shows that $f_{\mathbf{x}_0}(\mathbf{z})$ does not deviate far away from $\mathbb{E}f_{\mathbf{x}_0}(\mathbf{z})$ uniformly for all $\mathbf{z} \in \{\mathbf{z} : \|\mathbf{z}\|_2 \leq 1\}$, when m is large. The proof can be found in the supplementary material.

Lemma 3 (Concentration of Measure)**.** *Let* $\mathbf{z} \in \{\mathbf{z} \in \mathbb{R}^d : \|\mathbf{z}\|_2 \leq 1\}$ *and* $\{\mathbf{w}_i\}_{i=1}^m$ *be random vectors i.i.d. sampled from the Gaussian distribution* $\mathcal{N}(\mathbf{0}, \boldsymbol{I})$*. Suppose that* $m \geq c_0 d \log^4\left(\frac{1}{\delta}\right)\epsilon^{-2}$ *with a constant* c_0*, then with probability at least* $1-\delta$*,* $\sup_{\mathbf{z}} |f_{\mathbf{x}_0}(\mathbf{z}) - \mathbb{E}f_{\mathbf{x}_0}(\mathbf{z})| \leq \epsilon$*, where the supremum is over all* $\mathbf{z} \in \{\mathbf{z} \in \mathbb{R}^d : \|\mathbf{z}\|_2 \leq 1\}$*.*

Now we prove the correctness of initialization step.

Theorem 4. *Let* $\{\mathbf{a}_i\}_{i=1}^m \in \mathbb{R}^n$ *be random vectors sampled i.i.d. from the sampling model in the preliminaries section. Assume that* $\{y_i\}_{i=1}^m$ *follow the model* $y_i = (\mathbf{a}_i^T\mathbf{x}_0)^2 + \eta_i$*. Let* $\delta > 0$ *and* $m \geq c_0 d \log^4\left(\frac{1}{\delta}\right)\epsilon^{-2}$*. Then with probability at least* $1-\delta$*, the solution* $\widehat{\mathbf{z}}$ *to the convex program* (6) *satisfies* $\min_{\pm\mathbf{x}_0} \|\widehat{\mathbf{z}} \pm \mathbf{\Psi x}_0\|_2^2 \leq \epsilon$*.*

Proof. The proof is an immediate result of Lemmas 2 and 3. Specifically, note that the optimal solution $\widehat{\mathbf{z}}$ to (6) satisfies $\|\widehat{\mathbf{z}}\|_2 = 1$. Thus

$$0 \leq f_{\mathbf{x}_0}(\widehat{\mathbf{z}}) - f_{\mathbf{x}_0}(\mathbf{\Psi x}_0) \quad (\widehat{\mathbf{z}} \text{ is the optimal solution})$$
$$\leq \mathbb{E}f_{\mathbf{x}_0}(\widehat{\mathbf{z}}) + \frac{\epsilon}{4} - \mathbb{E}f_{\mathbf{x}_0}(\mathbf{\Psi x}_0) + \frac{\epsilon}{4} \quad \text{(By Lemma 3)}$$
$$= \mathbb{E}[f_{\mathbf{x}_0}(\widehat{\mathbf{z}}) - f_{\mathbf{x}_0}(\mathbf{\Psi x}_0)] + \frac{\epsilon}{2}$$
$$\leq -\frac{1}{2}\min_{\pm\mathbf{x}_0} \|\widehat{\mathbf{z}} \pm \mathbf{\Psi x}_0\|_2^2 + \frac{\epsilon}{2} \quad \text{(By Lemma 2)}.$$

Thus we obtain that $\min_{\pm\mathbf{x}_0} \|\widehat{\mathbf{z}} \pm \mathbf{\Psi x}_0\|_2^2 \leq \epsilon$. □

We then prove the correctness of refinement step, i.e., model (7). We need the following lemma on the probability of Gaussian vector falling in the disagreement region.

Lemma 5 ((Balcan, Blum, and Vempala 2014))**.** *Denote by* D *an isotropic log-concave distribution in* $\mathbb{R}^n$*. Then there exist constant* c *and* c' *such that for any two vectors* $\mathbf{u}$ *and* $\mathbf{v}$ *in* $\mathbb{R}^n$ *we have that*

$$c\theta(\mathbf{u}, \mathbf{v}) \leq \Pr_{\mathbf{x}\sim D}[\text{sign}(\mathbf{u}^T\mathbf{x}) \neq \text{sign}(\mathbf{v}^T\mathbf{x})] \leq c'\theta(\mathbf{u}, \mathbf{v}).$$

Lemma 5 states that the disagreement probability only depends on the angle between two vectors, being independent of ambient dimension. Now we are ready to prove the the correctness of refinement step.

Proof. Let $\mathbf{z}_0 = \mathbf{\Psi x}_0$. By Theorem 4, with $m \geq c_0 d \log^4\left(\frac{1}{\delta}\right)$ the initialization step outputs a solution with small *constant* error ϵ. Since $\min_{\pm\mathbf{z}_0} \theta(\widehat{\mathbf{z}}, \pm\mathbf{z}_0) \leq 2\min_{\pm\mathbf{z}_0} \|\widehat{\mathbf{z}} \pm \mathbf{z}_0\|_2 \leq 2\sqrt{\epsilon}$, the agreement region of $\widehat{\mathbf{z}}$ and $\pm\mathbf{z}_0$ contains the set

$$\begin{aligned}\mathsf{AgrReg}(\widehat{\mathbf{z}}) = \{&\mathbf{w} \in \mathbb{R}^d : \text{sign}(\widehat{\mathbf{z}} \cdot \mathbf{w}) = \text{sign}(\boldsymbol{v} \cdot \mathbf{w}),\\ &\forall \boldsymbol{v} \text{ satisfies } \theta(\widehat{\mathbf{z}}, \boldsymbol{v}) \leq 2\sqrt{\epsilon}\} \cup\\ \{&\mathbf{w} \in \mathbb{R}^d : \text{sign}(-\widehat{\mathbf{z}} \cdot \mathbf{w}) = \text{sign}(\boldsymbol{v} \cdot \mathbf{w}),\\ &\forall \boldsymbol{v} \text{ satisfies } \theta(-\widehat{\mathbf{z}}, \boldsymbol{v}) \leq 2\sqrt{\epsilon}\} \quad (9)\end{aligned}$$

which, by Lemma 5, has probability at least $1 - c\sqrt{\epsilon}$ on the event that $\mathbf{w}_i$'s lie in this region for a constant c. So in average, there will be $m(1 - c\sqrt{\epsilon}) \geq c_0 d \log^4\left(\frac{1}{\delta}\right)(1 - c\sqrt{\epsilon})$ many points falling in $\mathsf{AgrReg}(\widehat{\mathbf{z}})$. Using a Chernoff bound, the number of examples falling in the agreement region grows in the same order of its expectation, i.e., $c'd\log^4\left(\frac{1}{\delta}\right)(1 - c\sqrt{\epsilon})$ for some constant c', with probability at least $1-\delta$. By standard analysis on sub-Gaussian matrix, the smallest singular value of $\mathbf{W}$ is in the order of $\Theta(\sqrt{m})$ (Vershynin 2010). On the other hand, the optimal solution to the least squares problem (7) is $\widetilde{\mathbf{z}} = \mathbf{W}^\dagger\mathbf{b}$. So $\|\widetilde{\mathbf{z}} - \mathbf{z}_0\|_2 \leq \sqrt{\eta}\|\mathbf{W}^\dagger\| = \sqrt{\eta}/\sigma_m(\mathbf{W}) = \mathcal{O}(\sqrt{\eta/m})$. □

Finally, a straightforward application of results on compressive sensing leads to the result of Theorem 1.

Numerical Experiments

In this section, we implement numerical simulations to evaluate the performance of the proposed Algorithm 1 and compare it with other competing methods.

Synthetic Simulations

We first investigate the recovery performance of the proposed algorithm 1. We fix the dimension of the ground-truth sparse real signal $\mathbf{x}_0$ to be 256 (i.e. $n = 256$). The non-zero indices of $\mathbf{x}_0$ are independently and randomly selected within uniform distribution; entry values are drawn i.i.d. from $\mathcal{N}(0, 1)$ and normalized, i.e. $\|\mathbf{x}_0\|_2 = 1$. Denote the estimated signal as $\widetilde{\mathbf{x}}$; since CPR focuses on the direction of the target signal, we evaluate the recovery performance by measuring the relative error defined by $\min_{\pm\widetilde{\mathbf{x}}} \frac{\|\widetilde{\mathbf{x}} - \mathbf{x}_0\|_2}{\|\mathbf{x}_0\|_2}$. The entries of noise vector $\boldsymbol{\eta}$ are drawn i.i.d from the Gaussian mixture distribution of four Gaussian components with identical variance σ^2. We denote σ^2 as the noise level and in our experiment we choose 10^{-2} or 10^{-4}. The sensing vectors $\{\mathbf{w}_i\}_{i=1}^m$ are generated i.i.d. from $\mathbf{w}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$ on $\mathbb{R}^d$ and the compressive matrix $\mathbf{\Psi}$ with restricted isometry property is also Gaussian with i.i.d. $\mathcal{N}(0, \frac{1}{d})$ entries.

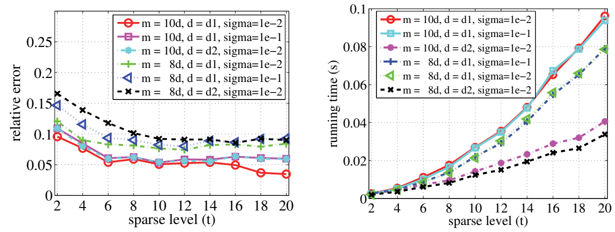
There are three parameters on the recovery accuracy, i.e. the support size t, the low dimension $d = O(t\log(n/t))$ and the sample size m. We select the support size t in set

Table 2: The performance of 100 trials of different algorithms with $n = 256$.

	relative error			
t	TWF	ECPR	SparsePR	Ours
6	0.6081	0.0568	**0.0566**	0.0605
8	0.5142	0.0629	0.0626	**0.0624**
10	0.4367	0.0559	0.0556	**0.0533**
12	0.3092	**0.0576**	0.0590	0.0585
14	0.1524	**0.0575**	0.0577	0.0580
	average running time (s)			
t	TWF	ECPR	SparsePR	Ours
6	0.0816	4.0276	4.0301	**0.0102**
8	0.0918	5.0880	5.0919	**0.0166**
10	0.1063	6.4807	6.4871	**0.0267**
12	0.1169	8.2197	8.2098	**0.0350**
14	0.1295	10.3199	10.3323	**0.0477**

$\{2, 4, 6, ..., 20\}$ and vary the values of d and m to investigate the influence of their choices on the recovery accuracy of our method. All experiments are implemented over 100 trials and for the "Recover" step in Algorithm 1 we utilize the SPAMS toolbox (Mairal et al. 2009). Figure 3 shows the 0.9 quantiles of the relative error versus k for different choices of d and m, and the corresponding average running time. As shown in Figure 3, the recovery error is satisfactory and relatively stable with noise for different sparsity levels. With the increase of sample size m, the recovery error significantly drops due to its contribution to a more accurate initial guess and better refinement results. Besides, larger embedding dimension d also contributes more to the improvement of recovery accuracy. It may be intuitive to think that the high noise level would have negative impact on the recovery performance, however, our algorithm still has significant robustness against high-level noise. As for the efficiency performance, it mainly depends on the values of m and d: m dominates the computational cost of refinement step while d is straightforwardly related to Steps 1 and 3. Note that there is a tradeoff between recovery accuracy and efficiency, reflected by m and d: the larger m and d are, the better the accuracy will be and the more time the algorithm will cost.

We also compare our algorithm to other state-of-the-art methods, including ECPR (Bahmani and Romberg 2015), SparsePR (Iwen, Viswanathan, and Wang 2015) and TWF (Chen and Candès 2015). The generation scheme of ground-truth signal $\mathbf{x}_0$, compressive matrix $\mathbf{\Psi}$ and sensing vectors $\mathbf{w}_i$s is the same as in the first experiment apart from the parameter choices. In specific, we choose $d = \lceil 2t(1+\log \frac{n}{t})\rceil$, $m = 10d$, $\sigma^2 = 10^{-2}$ in different $t \in \{6, 8, 10, 12, 14\}$. Table 2 shows the 0.9 quantiles of the relative error in 100 trails, and also reports the average running time for all competing methods. As shown in Table 2, the error results of TWF are much larger (> 0.6) than the other three methods (< 0.1), which results from that it does not explore the inherent sparse structure of real signals. Our algorithm is comparable with ECPR and SparsePR in the recovery performance, and all of them can achieve excellent recovery accuracies. However, our algorithm is exceedingly faster than the other algorithms, i.e. $200 \sim 300$ times than ECPR and SparsePR.



(a) Empirical 0.9 quantile of the relative error

(b) Average running time (s)

Figure 3: The recovery and efficiency performance of 100 trails w.r.t. different sparsity levels t for different choices of m and d with $n = 256$. Specifically, $d1 = \lceil 2t(1 + \log \frac{n}{t})\rceil$ and $d2 = \lceil 2t \log \frac{n}{t}\rceil$, where $\lceil\cdot\rceil$ denotes the smallest integer larger than a real number.

Real Experiment: Image Recovery

We implement our algorithm in the image recovery problem. We select an image of crescent Uranus taken by Voyager [2] shown in Figure 1 and vectorize it into a vector. Then we draw a Gaussian compressive matrix $\mathbf{\Psi}$ and sensing vectors $\mathbf{w}_i$ as the synthetic simulations, and we recover the whole image using our 3-stage algorithm with mixed Gaussian noise level $\sigma^2 = 10^{-2}$. As shown in Figure 1, the recovered image is of high quality and little distortion. Note that the additional light dots (emerging when zooming in the recovered image) in the recovered image can be effectively removed by various filtering methods, such as median filtering. Also note that the comparable compressive algorithms ECPR and SparsePR would be restricted on the laptop computer for the lifted signals would not fit into the limited memory budget.

Conclusions

In this paper, we investigate the problem of compressive phase retrieval. There are three challenges in this field: efficiency, robustness and sample complexity. Prior work mostly covers one or two of these aspects, which limits their applicability to real applications. To address the three challenges simultaneously, we propose a fast and robust algorithm for compressive phase retrieval, which is able to handle any bounded yet unstructured noise. For a t-sparse signal in $\mathbb{R}^n$, our algorithm can recover the ground truth with a small error using $\tilde{\mathcal{O}}(t)$ samples and finish in $\tilde{\mathcal{O}}(tn \log(1/\epsilon))$ time, based on a series of theoretical analysis. Synthetic and real experiments validate the superiority of our algorithm in both recovery accuracy and computational efficiency.

Acknowledgements Zhouchen Lin is supported by National Basic Research Program of China (973 Program) (grant no. 2015CB352502), National Natural Science Foundation (NSF) of China (grant nos. 61625301 and 61231002), and Qualcomm. Shan You and Chao Xu are supported by the National Natural Science Foundation of China under Grant NSFC 61375026 and 2015BAF15B00.

[2] https://www.nasa.gov/sites/default/files/thumbnails/image/pia00143.jpg

References

Bahmani, S., and Romberg, J. 2015. Efficient compressive phase retrieval with constrained sensing vectors. In *Advances in Neural Information Processing Systems*, 523–531.

Balcan, M.-F.; Blum, A.; and Vempala, S. 2014. Efficient representations for life-long learning and autoencoding. *arXiv preprint arXiv:1411.1490*.

Bertolotti, J.; van Putten, E. G.; Blum, C.; Lagendijk, A.; Vos, W. L.; and Mosk, A. P. 2012. Non-invasive imaging through opaque scattering layers. *Nature* 491(7423):232–234.

Cai, T. T.; Li, X.; and Ma, Z. 2015. Optimal rates of convergence for noisy sparse phase retrieval via thresholded wirtinger flow. *arXiv preprint arXiv:1506.03382*.

Candès, E. J.; Li, X.; and Soltanolkotabi, M. 2015. Phase retrieval via Wirtinger flow: Theory and algorithms. *IEEE Transactions on Information Theory* 61(4):1985–2007.

Candès, E. J.; Strohmer, T.; and Voroninski, V. 2013. Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. *Communications on Pure and Applied Mathematics* 66(8):1241–1274.

Chen, Y., and Candès, E. J. 2015. Solving random quadratic systems of equations is nearly as easy as solving linear systems. In *Advances in Neural Information Processing Systems*, 739–747.

Chen, Y.; Chi, Y.; and Goldsmith, A. J. 2015. Exact and stable covariance estimation from quadratic sampling via convex programming. *IEEE Transactions on Information Theory* 61(7):4034–4059.

Corbett, J. 2006. The Pauli problem, state reconstruction and quantum-real numbers. *Reports on Mathematical Physics* 57(1):53–68.

Duarte, M. F.; Davenport, M. A.; Takhar, D.; Laska, J. N.; Sun, T.; Kelly, K. E.; Baraniuk, R. G.; et al. 2008. Single-pixel imaging via compressive sampling. *IEEE Signal Processing Magazine* 25(2):83.

Hong, M., and Luo, Z.-Q. 2012. On the linear convergence of the alternating direction method of multipliers. *arXiv preprint arXiv:1208.3922*.

Iwen, M.; Viswanathan, A.; and Wang, Y. 2015. Robust sparse phase retrieval made easy. *Applied and Computational Harmonic Analysis*.

Li, X., and Voroninski, V. 2013. Sparse signal recovery from quadratic measurements via convex programming. *SIAM Journal on Mathematical Analysis* 45(5):3019–3033.

Liutkus, A.; Martina, D.; Popoff, S.; Chardon, G.; Katz, O.; Lerosey, G.; Gigan, S.; Daudet, L.; and Carron, I. 2014. Imaging with nature: Compressive imaging using a multiply scattering medium. *Scientific reports* 4.

Mairal, J.; Bach, F.; Ponce, J.; and Sapiro, G. 2009. Online dictionary learning for sparse coding. In *Proceedings of the 26th annual international conference on machine learning*, 689–696. ACM.

Miao, J.; Ishikawa, T.; Shen, Q.; and Earnest, T. 2008. Extending x-ray crystallography to allow the imaging of noncrystalline materials, cells, and single protein complexes. *Annu. Rev. Phys. Chem.* 59:387–410.

Moravec, M. L.; Romberg, J. K.; and Baraniuk, R. G. 2007. Compressive phase retrieval. In *Optical Engineering+ Applications*, 670120–670120. International Society for Optics and Photonics.

Ohlsson, H.; Yang, A.; Dong, R.; and Sastry, S. 2012. Cprl–an extension of compressive sensing to the phase retrieval problem. In *Advances in Neural Information Processing Systems*, 1367–1375.

Schniter, P., and Rangan, S. 2015. Compressive phase retrieval via generalized approximate message passing. *IEEE Transactions on Signal Processing* 63(4):1043–1055.

Shechtman, Y.; Beck, A.; and Eldar, Y. C. 2014. GESPAR: Efficient phase retrieval of sparse signals. *IEEE Transactions on Signal Processing* 62(4):928–938.

Vershynin, R. 2010. Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint: 1011.3027*.

Yin, D.; Lee, K.; Pedarsani, R.; and Ramchandran, K. 2015. Fast and robust compressive phase retrieval with sparse-graph codes. In *2015 IEEE International Symposium on Information Theory (ISIT)*, 2583–2587.

Zhang, H., and Liang, Y. 2016. Reshaped wirtinger flow for solving quadratic systems of equations. *arXiv preprint arXiv:1605.07719*.

Zhang, H.; Lin, Z.; Zhang, C.; and Chang, E. 2015a. Exact recoverability of robust PCA via outlier pursuit with tight recovery bounds. In *AAAI Conference on Artificial Intelligence*, 3143–3149.

Zhang, H.; Lin, Z.; Zhang, C.; and Gao, J. 2015b. Relations among some low rank subspace recovery models. *Neural Computation* 27:1915–1950.

Zhang, W.; Zhang, L.; Jin, R.; Cai, D.; and He, X. 2016. Accelerated sparse linear regression via random projection. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI)*, 2337–2343.

Zhang, H.; Lin, Z.; and Zhang, C. 2016. Completing low-rank matrices with corrupted samples from few coefficients in general basis. *IEEE Transactions on Information Theory* (8):4748–4768.