*Appl. Math. J. Chinese Univ.*
2020, 35(4): 479-502

# Phase retrieval with PhaseLift algorithm

LI Hui-ping LI Song

**Abstract**. This paper provides a contemporary overview of phase retrieval problem with PhaseLift algorithm and summarizes theoretical results which have been derived during the past few years. Based on the lifting technique, the phase retrieval problem can be transformed into the low rank matrix recovery problem and then be solved by convex programming known as PhaseLift. Thus, stable guarantees for such problem have been gradually established for measurements sampled from sufficiently random distribution, for instance, the standard normal distribution. Further, exact recovery results have also been set up for masked Fourier measurements which are closely related to practical applications.

## §1 Introduction

### 1.1 Background

The problem of phase retrieval i.e., the recovery of a signal from its Fourier magnitude [61, 62], arises in many areas of engineering and applied physics, including optics [74], X-ray crystallography [55], astronomical imaging [20], speech processing [64], computational biology [70], blind deconvolution [6] and more details in [38]. Its origin comes from the fact that detectors oftentimes can record only the squared modulus of the Fresnel or Fraunhofer diffraction pattern of the radiation that is scattered from an object. In such settings, one cannot measure the phase of the optical wave reaching the detector, and, therefore, much information about the scattered object or the optical field is lost since, as is well known, the phase encodes a lot of the structural content of the image we wish to form.

There are many ways in which one can pose the phase retrieval problem, for instance, depending on whether one assumes a continuous or discrete space model for the signal. In this paper, we focus on the discretized one-dimensional (1D) setting for simplicity and because numerical algorithms ultimately operate with digital data. Let $\boldsymbol{x}_0 = [x_0[0], x_0[1], \cdots, x_0[N-1]]^\top$

Received: 2020-05-06. Revised: 2020-06-12.
MR Subject Classification: 94A08, 94A12, 94A20, 78A45, 90C59.
Keywords: phase retrieval, PhaseLift algorithm, random measurements, masked Fourier measurements, sparsity.
Digital Object Identifier(DOI): https://doi.org/10.1007/s11766-020-4136-3.
Supported by the National Natural Science Foundation of China (11531013, U1630116) and the fundamental research funds for the central universities.

be a signal of length $N$ such that it has non-zero values only within the interval $[0, N-1]$. Denote $\boldsymbol{y} = [y[0], y[1], \cdots, y[N-1]]^\top$ as its $N$ point discrete Fourier transform (DFT) and let $\boldsymbol{z} = [z[0], z[1], \cdots, z[N-1]]^\top$ be the Fourier magnitude-square measurements, namely, $|z[k]| = |y[k]|^2$. Then the phase retrieval problem can be mathematically stated as:

$$\begin{aligned} &\text{Find } \boldsymbol{x} \\ &\text{s. t. } |z[k]| = |y[k]|^2, \quad 0 \le k \le N-1, \end{aligned} \tag{1}$$

where $y[k] = \langle \boldsymbol{f}_k, \boldsymbol{x}_0 \rangle$ and $\boldsymbol{f}_k$ being the conjugate of the $k$-th column of $N$ point DFT matrix, with elements $e^{\frac{i \cdot 2\pi kn}{N}}$, $n = 0, 1, \cdots, N-1$.

In the 1D setting, this problem is ill-posed since there are many different signals whose Fourier transforms have the same magnitude. Clearly, if $\boldsymbol{x}$ is a solution to the phase retrieval problem, then (i) $c \cdot \boldsymbol{x}$ for any scalar $c \in \mathbb{C}$ obeying $|c| = 1$ is also a solution, (ii) the time-reversed signal or (iii) the shifted signal is also a solution. Usually, these "trivial associates" of $\boldsymbol{x}$ are acceptable ambiguities in view of physical aspect. But in general infinitely many solutions can be obtained from $\{|y[k]|, 0 \le k \le N-1\}$ beyond these trivial ambiguities [67]. Furthermore, it has been shown, using spectral factorization, that there is no uniqueness and the feasible set of (1) can include up to $2^N$ non-equivalent solutions and adding support constraints on $\boldsymbol{x}$ does not help to ensure uniqueness [33]. For higher dimensions (for instance, 2D signal), it has been shown by Hayes [32] using dimension counting that, with the exception of a set of signals of measure zero, phase retrieval with oversampling is well posed up to trivial ambiguities.

As we further discuss below, to guarantee unique identification in the 1D case it is necessary to assume additional constraints on the unknown signal such as sparsity or to introduce specific redundancy into the measurements. Meanwhile, even when uniqueness of the underlying signal is theoretically possible, it is important to recover this signal efficiently and robustly. By and large, earlier methods for phase retrieval from oversampled data are alternating projection algorithms pioneered by Gerchberg and Saxton [25] and Fienup [22, 23]. In their framework, the key idea is to reformulate (1) into the following least-squares problem and apply alternating projection to solve this non-convex problem:

$$\min_{\boldsymbol{x}} \sum_{k=0}^{N-1} (|y[k]|^2 - |\langle \boldsymbol{f}_k, \boldsymbol{x} \rangle|^2)^2. \tag{2}$$

Often, the objective is shown to be monotonically decreasing as the iterations progress. While this algorithm is simple to implement and amenable to additional constraints such as the positivity of $\boldsymbol{x}_0$, its convergence remains problematic, due to the projections onto non-convex sets. Thus such algorithms need careful exploitation of signal constraints and delicate parameter selection to increase the likelihood of convergence to a correct solution, see [19, 49, 50, 59].

Recently, one of the popular approaches to treat phase retrieval problems is to use semidefinite programming (SDP) methods. Such algorithms have been shown to yield robust solutions to various quadratic-constrained optimization problems, see [10–12, 17, 26, 73]. As we can see from (1), the quadratic constraints can be lifted up and interpreted as linear measurements about the rank-one matrix $\boldsymbol{X} = \boldsymbol{x}\boldsymbol{x}^*$. Namely,

$$|z[k]| = |y[k]|^2 = |\langle \boldsymbol{f}_k, \boldsymbol{x} \rangle|^2 = Tr(\boldsymbol{x}^* \boldsymbol{f}_k \boldsymbol{f}_k^* \boldsymbol{x}) = Tr(\boldsymbol{f}_k \boldsymbol{f}_k^* \boldsymbol{x}\boldsymbol{x}^*) = Tr(\boldsymbol{f}_k \boldsymbol{f}_k^* \boldsymbol{X}). \tag{3}$$

In what follows, we shall let $\mathcal{A}$ be the linear operator mapping positive semidefinite matrices into $\{Tr(\boldsymbol{f}_k\boldsymbol{f}_k^*\boldsymbol{X}), k=0,\cdots,N-1\}$. Hence, the phase retrieval problem (1) is equivalent to

$$\begin{aligned} &\text{Find } \boldsymbol{X} \\ &\text{s. t. } \mathcal{A}(\boldsymbol{X})=\boldsymbol{y}, \\ &\quad \operatorname{rank}(\boldsymbol{X})=1, \\ &\quad \boldsymbol{X}\succeq\boldsymbol{0}. \end{aligned} \quad \Leftrightarrow \quad \begin{aligned} &\min \operatorname{rank}(\boldsymbol{X}) \\ &\text{s. t. } \mathcal{A}(\boldsymbol{X})=\boldsymbol{y}, \\ &\quad \boldsymbol{X}\succeq\boldsymbol{0}. \end{aligned} \tag{4}$$

As such, the problem (4) is a rank-minimization problem over an affine slice of the positive semidefinite cone. The idea of linearizing the phase retrieval problem by reformulating it as a problem of recovering a matrix from linear measurements can be found in [3]. Then we can factorize the solution $\boldsymbol{X}$ of (4) as $\boldsymbol{x}\boldsymbol{x}^*$ in order to obtain the solutions to the phase retrieval problem (1). Since the rank-minimization problem (4) is NP-hard and non-convex, many researchers have recently attempted to replace the the rank functional $\operatorname{rank}(\boldsymbol{X})$ by the convex surrogate $Tr(\boldsymbol{X})$ [16,65], giving the familiar SDP

$$\begin{aligned} &\min_{\boldsymbol{X}} \ Tr(\boldsymbol{X}) \\ &\text{s. t. } \mathcal{A}(\boldsymbol{X})=\boldsymbol{y}, \\ &\quad \boldsymbol{X}\succeq\boldsymbol{0}. \end{aligned} \tag{5}$$

The convex program (5) is known as PhaseLift algorithm which was proposed in [10,17] and it is a semidefinite program (SDP) in standard form, so there is a rapidly growing list of algorithms for solving problems of this kind as efficiently as possible. Furthermore, SDP algorithms are known to be robust in general. Consequently, when we consider noisy measurements with bounded noise, namely, $\boldsymbol{y}=\mathcal{A}(\boldsymbol{X})+\boldsymbol{\varepsilon}$ with $\|\boldsymbol{\varepsilon}\|_{\ell_2}\leq\eta$, a robust variant PhaseLift algorithm was provided in [17],

$$\begin{aligned} &\min_{\boldsymbol{X}} \ Tr(\boldsymbol{X}) \\ &\text{s. t. } \|\boldsymbol{y}-\mathcal{A}(\boldsymbol{X})\|_{\ell_2}\leq\eta, \\ &\quad \boldsymbol{X}\succeq\boldsymbol{0}. \end{aligned} \tag{6}$$

In this case, the authors in [11,17] do not claim $\boldsymbol{X}$ that has low rank. Instead, they suggest estimating $\boldsymbol{x}_0$ by extracting the largest rank-one component. Namely, if we suppose the SVD of $\boldsymbol{X}$ is

$$\boldsymbol{X}=\sum_{i=1}^{n}\lambda_i\boldsymbol{u}_i\boldsymbol{u}_i^*, \quad \lambda_1\geq\lambda_2\geq\cdots\geq\lambda_n, \tag{7}$$

and $\boldsymbol{u}_1,\cdots,\boldsymbol{u}_n$ are mutually orthogonal, then let $\boldsymbol{x}=\sqrt{\lambda_1}\cdot\boldsymbol{u}_1$ be an estimator of $\boldsymbol{x}_0$. As is stated above, the PhaseLift algorithm can transform what appears to be a combinatorial problem into a convex program which offers tractable solutions. Moreover, the existence of a noise-aware recovery procedure (6)-also based on a tractable convex program-is robust vis-à-vis additive noise.

More recent methods described in [9,10,13,18,75–77,82,83] retrieve the phase by applying techniques such as matrix completion, non-convex formulations, and so on. Specifically, one of the non-convex formulations based on the intensity measurements $\{|y[k]|^2\}_{k=0}^{N-1}$ or amplitude mea-

surements $\{|y[k]|\}_{k=0}^{N-1}$, called the Wirtinger Flow (WF) [13] or Truncated Amplitude Flow (TAF) [75], and another type of non-convex approaches are based on combining lifting technique and matrix recovery [9], which are all gradient descent method based on the Wirtinger derivative. In addition, there are several alternates proposed and studied in [2, 27, 48, 57, 71, 80]. However, these methods mentioned above greatly depend on choosing good initializers.

## 1.2 Notations

We shall introduce some basic notations which will be used throughout this paper. Let $\boldsymbol{y}^*$ and $\boldsymbol{y}^\top$ be the conjugate transpose and transpose of a vector $\boldsymbol{y} \in \mathbb{C}^n$ or $\mathbb{R}^n$, respectively. Write $\mathbb{H}^{n\times n}$ and $\mathbb{S}^{n\times n}$ to mean the space of all Hermitian matrices and symmetric matrices. Let $\boldsymbol{x}_0 \in \mathbb{R}^n$ or $\mathbb{C}^n$ be a signal we aim to recover, we define the space

$$T := \{\boldsymbol{x}_0\boldsymbol{z}^\top + \boldsymbol{z}\boldsymbol{x}_0^\top : \forall \boldsymbol{z} \in \mathbb{R}^n\} \subset \mathbb{S}^{n\times n} \text{ or } \tilde{T} := \{\boldsymbol{x}_0\boldsymbol{z}^* + \boldsymbol{z}\boldsymbol{x}_0^* : \forall \boldsymbol{z} \in \mathbb{C}^n\} \subset \mathbb{H}^{n\times n}, \quad (8)$$

as the tangent space of the manifold of all rank one symmetric matrices or Hermitian matrices at the point $\boldsymbol{X}_0 = \boldsymbol{x}_0\boldsymbol{x}_0^\top$ or $\boldsymbol{x}_0\boldsymbol{x}_0^*$. Also, we denote the projection of a matrix $\boldsymbol{Y}$ onto the space $T(\tilde{T})$ by $\boldsymbol{Y}_T(\boldsymbol{Y}_{\tilde{T}})$ and its orthogonal complement is expressed as $\boldsymbol{Y}_{T^\perp}(\boldsymbol{Y}_{\tilde{T}^\perp})$. We use the projector $\Pi_{\boldsymbol{Y}} : \mathbb{H}^{n\times n} \to \mathbb{H}^{n\times n}$ onto some matrix $\boldsymbol{Y} \in \mathbb{H}^{n\times n}$ which is defined as

$$\Pi_{\boldsymbol{Y}}(\boldsymbol{Z}) = \boldsymbol{Y}(\boldsymbol{Y}, \boldsymbol{Z}) = tr(\boldsymbol{Y}\boldsymbol{Z})\boldsymbol{Y}, \quad \forall \boldsymbol{Z} \in \mathbb{H}^{n\times n}.$$

Let $\mathcal{I} : \mathbb{H}^{n\times n} \to \mathbb{H}^{n\times n}$ be the identity map. Here, write " $\geq$, $\leq$ " and " $\succeq$, $\preceq$ " to mean the positive or negative semi-definiteness of matrices and matrix-valued operators, respectively. $\{\boldsymbol{e}_j\}_{1\leq j\leq n}$ represent the standard orthogonal basis in $\mathbb{C}^n$. Denote $\|\cdot\|_1$, $\|\cdot\|_2$, $\|\cdot\|_{1,1}$ as the nuclear norm, Frobenius norm, and entry-wise $\ell_1$ norm of matirces, respectively. Denote $\|\cdot\|_{\ell_p}$ as the $\ell_p$-norm of vectors, where $p = 1, 2$. Particularly, $\|\cdot\|_{\ell_0}$ denotes the sparsity of a vector. $C > 0$, (or $c, c', C_0, C_1, C', C''$) denotes a universal constant that might be different in each occurrence.

# §2 PhaseLift under Random Measurements

## 2.1 Non-uniform Guarantees for PhaseLift algorithm

As is well known to us, the phase retrieval problem has attracted great attention in the past few years, due to its wide occurrence in many physical applications. In the classical case discussed above, the $\{\boldsymbol{f}_k\}_{k=0}^{N-1}$ are complex exponentials at frequency $e^{\frac{i\cdot 2\pi kn}{N}}$ so that one collects the squared modulus of the Fourier transform of $\boldsymbol{x}_0$. In such case, we often call the phase problem as the classical phase retrieval problem. Later, many other choices for the measurement vectors $\boldsymbol{a}_k$ are frequently discussed in the literature; see [3, 24] for instance. In this case, we call the phase problem as the general phase retrieval problem. A frame-theoretic approach to signal recovery from magnitude measurements has been proposed in [3–5], where the authors derived various necessary and sufficient conditions for the uniqueness of the solution, as well as various polynomial time numerical algorithms for very specific choices of $\boldsymbol{a}_k$ . While theoretically quite appealing, the drawbacks are that the methods are $(i)$ either algebraic in nature, thus severely

limiting their stability in the presence of noise or slightly inexact data, or (*ii*) the number of measurements is on the order of $n^2$, which is much too large compared to the number of unknowns.

Hence, in a pioneering work by Candès et al. [11,17], a convex relaxation via trace norm minimization, known as PhaseLift, was studied. As is introduced in the Section 1, This algorithm consists in lifting up the original problem of vector recovery from a quadratic system into that of recovering a rank-one matrix by a trace-norm relaxation. The crucial questions are whether and under which conditions the combinatorially hard problem (4) and the convex problem (5) are formally equivalent in the noiseless case and robustly recover the true solution by the convex program (6). Thus, many researchers have attemptted to take advantage of sufficiently random measurements to overcome these problems, so that to show the feasibility and robustness of the PhaseLift algorithms. To the best of our knowledge, these analyses are mainly based on two kinds of technical tools, namely, the standard duality arguments in semidefinite programming and the Mendelson's Small Ball Method [40,45,52–54,72]. In what follows, before introducing stable guarantees for PhaseLift algorithm, we redefine the linear operator $\mathcal{A}$ as follows:

$$\begin{aligned} &\mathcal{A}:\ \mathbb{H}^{n\times n}\to\mathbb{R}^m \\ &\boldsymbol{X}\to\mathcal{A}(\boldsymbol{X})=\sum_{k=1}^{n} tr(\boldsymbol{A}_k\boldsymbol{X})\boldsymbol{e}_k, \end{aligned} \tag{9}$$

where $\boldsymbol{A}_k=\boldsymbol{a_k}\boldsymbol{a_k^*}$ with $\boldsymbol{a}_k$ being different measurement vectors in each occurrence. Correspondingly, the adjoint operator $\mathcal{A}^*$ maps real-valued inputs into Hermitian matrices and is given by

$$\begin{aligned} &\mathcal{A}^*:\ \mathbb{R}^m\to\mathbb{H}^{n\times n} \\ &\boldsymbol{z}\to\mathcal{A}^*(\boldsymbol{z})=\sum_{k=1}^{n} z_k\cdot\boldsymbol{A}_k. \end{aligned} \tag{10}$$

Then the first recovery result under random Gaussian measurements or uniformly distributed measurements was established by Candès et al. in [17] as follows:

**Theorem 2.1.** *( [17]) Consider an arbitrary signal $\boldsymbol{x}_0\in\mathbb{R}^n$ or $\mathbb{C}^n$ and suppose that the number of measurements obeys $m=c_0\cdot n\log n$, where $c_0$ is a sufficiently large constant. Then in both the real and complex cases, the solution to the trace-minimization program is exact with high probability in the sense that* (5) *has a unique solution obeying*

$$\hat{\boldsymbol{X}}=\boldsymbol{x}_0\boldsymbol{x}_0^*. \tag{11}$$

*This holds with probability at least $1-3e^{-\gamma\frac{m}{n}}$, where $\gamma$ is a positive absolute constant.*

This result shows that the convex program recovers $\boldsymbol{x}_0$ exactly (up to global phase), provided the number $m$ of magnitude measurements is on the order of $n\log n$ in the noiseless case. Expressed differently, Theorem 2.1 establishes a rigorous equivalence between a class of phase retrieval problems and a class of semidefinite programs. Clearly, any phase retrieval algorithm, no matter how complicated or intractable, would need at least $2n$ quadratic measurements to recover a complex-valued object $\boldsymbol{x}_0\in\mathbb{C}^n$.

In the real world, measurements are often contaminated by noise. Using the frame-works

developed in [14,31], we can extend Theorem 2.1 to accommodate noisy measurements. Since in many applications of interest in optics and other areas of physics, we often observe

$$y_k = |\langle \boldsymbol{a}_k, \boldsymbol{x}_0\rangle|^2 + \varepsilon_k, \quad 1 \le k \le m, \tag{12}$$

which leads to the noisy measurements $\boldsymbol{y} = \mathcal{A}(\boldsymbol{X}_0) + \boldsymbol{\varepsilon}$ with $\|\boldsymbol{\varepsilon}\|_{\ell_2} \le \eta$ and $\boldsymbol{\varepsilon} = [\varepsilon_1, \cdots, \varepsilon_m]^\top$ where $\eta > 0$ is an absolute constant. So by the convex program (6), robust recovery was established:

**Theorem 2.2.** *([17]) Fix $\boldsymbol{x}_0 \in \mathbb{R}^n$ or $\mathbb{C}^n$ and assume the $\boldsymbol{a}_k'$s are uniformly sampled on the sphere of radius $\sqrt{n}$. Under the hypotheses of Theorem* 2.1*, the solution to* (6) *obeys*

$$\|\hat{\boldsymbol{X}} - \boldsymbol{x}_0\boldsymbol{x}_0^*\|_2 \le C_0\eta, \tag{13}$$

*for some positive numerical constant $C_0$. We also get*

$$\|\hat{\boldsymbol{x}} - e^{i\cdot\phi}\boldsymbol{x}_0\|_{\ell_2} \le C_0 \cdot \min\{\|\boldsymbol{x}_0\|_{\ell_2}, \frac{\eta}{\|\boldsymbol{x}_0\|_{\ell_2}}\}. \tag{14}$$

*for some $\phi \in [0, 2\pi]$. Both these estimates hold with nearly the same probability as that in the noiseless case. Here $\hat{\boldsymbol{x}}$ is derived from extracting the largest rank-one component of $\hat{\boldsymbol{X}}$ as introduced in Section* 1.

The strategy to prove Theorem 2.1 and Theorem 2.2 hinges on the fact that a strengthening of the injectivity property allows the authors in [17] to relax the properties of the dual certificate, as that in the approach pioneered in [28] for matrix completion. For example, when the original signal $\boldsymbol{x}_0 \in \mathbb{R}^n$, the crucial lemma is required as below. As for the complex case, one only requires a slight adjustment in the numerical constants, see Lemma 5.4 in [17]. We omit it for simplicity.

**Lemma 2.3.** *([17]) Suppose that the mapping $\mathcal{A}$ obeys the following two properties: for all positive semidefinite matrices $\boldsymbol{X}$*

$$m^{-1}\|\mathcal{A}(\boldsymbol{X})\|_{\ell_1} < (1 + \frac{1}{9})\|\boldsymbol{X}\|_1 \tag{15}$$

*and for all matrices $\boldsymbol{X} \in T$ where $T$ is defined as* (8)

$$m^{-1}\|\mathcal{A}(\boldsymbol{X})\|_{\ell_1} > 0.94(1 - \frac{1}{9})\|\boldsymbol{X}\|. \tag{16}$$

*Suppose further that there exists an approximate dual certificate $\boldsymbol{Y}$ in the range of $\mathcal{A}^*$ obeying*

$$\|\boldsymbol{Y}_T - \boldsymbol{e}_1\boldsymbol{e}_1^*\|_2 \le \frac{1}{3}, \quad \|\boldsymbol{Y}_{T^\perp}\| \le \frac{1}{2}. \tag{17}$$

*Then $\boldsymbol{e}_1\boldsymbol{e}_1^*$ is the unique minimizer to* (5).

The first property (15) is reminiscent of the (one-sided) restricted isometry property (RIP) in the area of compressed sensing [15]. The difference is that it is expressed in the 1-norm rather than the 2-norm. Having said this, it's noted that RIP-1 properties have also been used in the compressed sensing literature; see [7] for example. The authors in [17] use this property instead of a property about $\|\mathcal{A}(\boldsymbol{X})\|_{\ell_2}$, because a RIP property in the 2-norm does not hold here (essentially because $\|\mathcal{A}(\boldsymbol{X})\|_{\ell_2}$ involves fourth moments of Gaussian variables), as they demonstrated in the Appendix. The second property (16) is a form of local RIP-1 since it holds only for matrices in $T$ which is the tangent space of the manifold of all rank one symmetric

matrices at the point $\boldsymbol{X}$. Then they emphasized that the bound for the dual certificate in (17) was loose in the sense that $\boldsymbol{Y}_T$ and $\boldsymbol{e}_1\boldsymbol{e}_1^*$ may not be that close, a fact that would play a crucial role in their proof. Finally, they sufficiently utilized the some statistics arguments about random Gaussian distribution or uniform distribution to ensure $\mathcal{A}$ and the constructed dual certificate $\boldsymbol{Y}$ obeying the properties stated in Lemma 2.3.

Later, the authors in [11] showed that the convex relaxation (5) with a number of equations on the order of the number of unknowns succeeded to recovering the original signal $\boldsymbol{x}_0$ with much higher probability. Moreover, they provided a universal result stating that once the vectors $\boldsymbol{a}_k$ have been selected, all input signals $\boldsymbol{x}_0$ can be recovered simultaneously.

**Theorem 2.4.** *([11]) Suppose the $\boldsymbol{a}_k'$s are generated independently and identically according the normal distribution or the uniform distribution. Assume that $m \geq c_0 \cdot n$ where $c_0$ is a sufficiently large constant. Then the following holds with probability at least $1 - O(e^{-\gamma \cdot m})$: for all $\boldsymbol{x}_0 \in \mathbb{R}^n$ or $\mathbb{C}^n$, the PhaseLift feasibility problem*

$$\{\boldsymbol{X}: \quad \mathcal{A}(\boldsymbol{X}) = \boldsymbol{y}, \quad \boldsymbol{X} \succeq \boldsymbol{0}\} \tag{18}$$

*has a unique point, namely, $\boldsymbol{x}_0\boldsymbol{x}_0^*$. Thus, exact recovery holds simultaneously over all input signals.*

In words, the solution to most systems of quadratic equations can be obtained by solving a semidefinite programming feasibility problem; this is true as long as the number of equations is at least a constant times the number of unknowns. Further, the probability of failure is exponentially small in the number of measurements, a significant sharpening of Theorem 2.1. A second departure from Theorem 2.1 is that we only need to solve a feasibility problem. Indeed, trace minimization is unnecessary since the feasible set reduces to a single element, an observation independently presented in [17] when the number of measurements obeys $m = c_0 \log n$. The third departure is that exact recovery holds universally as explained above. To be sure, Theorem 2.1 states that with high probability, the null space of $\mathcal{A}$ is tangent to the positive semidefinite (PSD) cone $\{\boldsymbol{X}: \boldsymbol{X} \succeq \boldsymbol{0}\}$ at a fixed rank-one matrix $\boldsymbol{X}_0 = \boldsymbol{x}_0\boldsymbol{x}_0^*$ whereas Theorem 2.4 asserts that this null-space is tangent to the PSD cone at all rank-one elements. Mathematically, what makes this possible is the sharpening of the probability bounds; that is to say, the fact that for a fixed $\boldsymbol{x}_0$, recovery holds with probability at least $1 - O(e^{-\gamma \cdot m})$. Importantly, these improvements are derived from adaptable truncations in the statistics arguments for establishing similar properties as that stated in Lemma 2.3, see Lemma 2.1 and Lemma 2.2 in [11].

Further, in the noise case, the authors in [11] suggested recovering the signal by

$$\begin{aligned} &\min_{\boldsymbol{X}} \ \|\boldsymbol{y} - \mathcal{A}(\boldsymbol{X})\|_{\ell_1} \\ &\text{s. t.} \quad \boldsymbol{X} \succeq \boldsymbol{0}. \end{aligned} \tag{19}$$

The proposal cannot be simpler: find the positive semidefinite matrix $\boldsymbol{X}$ that best fits the observed data in an $\ell_1$ sense. One can then extract the best-rank one approximation to recover the signal $\boldsymbol{x}_0$.

**Theorem 2.5.** *([11]) Suppose the $\boldsymbol{a}_k'$s are generated independently and identically according*

*the normal distribution or the uniform distribution. Assume that* $m \geq c_0 \cdot n$ *where* $c_0$ *is a sufficiently large constant. Then for all* $\boldsymbol{x}_0 \in \mathbb{R}^n$ *or* $\mathbb{C}^n$*, the solution to* (19) *obeys*

$$\|\hat{\boldsymbol{X}} - \boldsymbol{x}_0\boldsymbol{x}_0^*\|_2 \leq C_0 \frac{\|\boldsymbol{\varepsilon}\|_{\ell_1}}{m} \tag{20}$$

*for some numerical constant* $C_0$*. For the Gaussian models, this holds with the same probability as in the noiseless case whereas the probability of failure is exponentially small in* $n$ *in the uniform model. By finding the largest eigenvector with largest eigenvalue of* $\hat{\boldsymbol{X}}$*, one can also construct an estimate obeying*

$$\|\hat{\boldsymbol{x}} - e^{i\phi}\boldsymbol{x}_0\|_{\ell_2} \leq C_0 \min\{\|\boldsymbol{x}_0\|_{\ell_2}, \frac{\|\boldsymbol{\varepsilon}\|_{\ell_1}}{m\|\boldsymbol{x}_0\|_{\ell_2}}\} \tag{21}$$

*for some* $\phi \in [0, 2\pi]$*.*

The authors in [11] stressed that the bounds (20) and (21) considerably strengthened the Theorem 2.2. To be sure, if we assume that the noise $\boldsymbol{\varepsilon}$ is known to be bounded, $\|\boldsymbol{\varepsilon}\|_{\ell_2} \leq \eta$, then (6) yields an estimate $\hat{\boldsymbol{X}}$ obeying

$$\|\hat{\boldsymbol{X}} - \boldsymbol{x}_0\boldsymbol{x}_0^*\|_2^2 \leq C_0^2\eta^2. \tag{22}$$

In contrast, since $\|\boldsymbol{\varepsilon}\|_{\ell_1} \leq \sqrt{m}\|\boldsymbol{\varepsilon}\|_{\ell_2} \leq \sqrt{m}\eta$, the Theorem 2.5 gives

$$\|\hat{\boldsymbol{X}} - \boldsymbol{x}_0\boldsymbol{x}_0^*\|_2^2 \leq C_0^2 \frac{\eta^2}{m}, \tag{23}$$

this represents a substantial improvement. Importantly, these results stated in Theorem 2.5 are optimal and cannot possibly be improved. To see why the stability result is optimal, suppose without loss of generality that $\|\boldsymbol{x}_0\|_{\ell_2} = 1$. Further, imagine that we are informed that $\|\boldsymbol{\varepsilon}\|_{\ell_1} \leq C_0\eta$ for some known $C_0$. Since $\|\mathcal{A}(\boldsymbol{x}_0\boldsymbol{x}_0)\|_{\ell_1} \approx m$, it would not be possible to distinguish between solutions of the form $(1+\lambda)\boldsymbol{x}_0\boldsymbol{x}_0$ for $\max\{0, 1 - C_0\} \lesssim 1 + \lambda \lesssim 1 + C_0$. Hence, the error in the Frobenius norm may be as large as $C_0\|\boldsymbol{x}_0\boldsymbol{x}_0^*\|_1 = C_0$, which is what the theorem gives.

PhaseLift algorithm is optimal in the sense that the number of amplitude measurements required for phase reconstruction scales linearly with the dimension of the signal. However, it specifically demands random Gaussian measurement vectors-a limitation that restricts practical utility and obscures the specific properties of measurement ensembles that enable phase retrieval. So the authors in [29] presented a partial de-randomization of PhaseLift that only required sampling from certain polynomial size vector configurations, called $t$-designs. Such configurations have been studied in algebraic combinatorics, coding theory, and quantum information. They proved reconstruction guarantees for a number of measurements that depended on the degree $t$ of the design. If the degree is allowed to grow logarithmically with the dimension, the bounds become tight up to polylog-factors. Beyond the specific case of PhaseLift, this work highlighted the utility of spherical designs for the de-randomization of data recovery schemes. Based on the PhaseLift algorithm (5) in the noiseless case and assume that the intensity $y' = \|\boldsymbol{x}_0\|_{\ell_2}^2$ is known, they suggested to recover $\boldsymbol{x}_0$ by solving the following feasibility

problem

$$\begin{aligned} &\text{Find} \quad \boldsymbol{X} \\ &\text{s. t.} \quad \boldsymbol{y} = \mathcal{A}(\boldsymbol{X}), \\ &\qquad tr(\boldsymbol{X}) = y^{'}, \\ &\qquad \boldsymbol{X} = \boldsymbol{X}^{*},\ \boldsymbol{X} \succeq \boldsymbol{0}. \end{aligned} \tag{24}$$

To begin with, let $\mathbb{V}_1, \cdots, \mathbb{V}_k$ be (finite dimensional, complex) vector spaces, and let $\mathbb{V}_1^*, \cdots, \mathbb{V}_k^*$ be their dual spaces. A function

$$f : \mathbb{V}_1 \times \cdots \times \mathbb{V}_k \longrightarrow \mathbb{C}$$

is multilinear, if it is linear in each $\mathbb{V}_i,\ i = 1, \cdots, k$. We denote the space of such functions by $\mathbb{V}_1^* \otimes \cdots \otimes \mathbb{V}_k^*$ and call it the tensor product of $\mathbb{V}_1, \cdots, \mathbb{V}_k$. Consequently, the tensor product $(\mathbb{V}^n)^{\otimes k}$ is the space of all multilinear functions

$$f : (\mathbb{V}^n)^* \times \cdots \times (\mathbb{V}^n)^* \longrightarrow \mathbb{C}$$

and we call the elementary elements $\boldsymbol{z}_1 \otimes \cdots \otimes \boldsymbol{z}_k$ the tensor product of the vectors $\boldsymbol{z}_1, \cdots, \boldsymbol{z}_k \in \mathbb{V}^n$. We define the (symmetrizer) map $P_{Sym^k} : (\mathbb{V}^n)^{\otimes k} \to (\mathbb{V}^n)^{\otimes k}$ via their action on elementary elements:

$$P_{Sym^k}(\boldsymbol{z}_1 \otimes \cdots \otimes \boldsymbol{z}_k) := \frac{1}{k!} \sum_{\pi \in S_k} \boldsymbol{z}_{\pi(1)} \otimes \cdots \otimes \boldsymbol{z}_{\pi(k)} \tag{25}$$

where $S_k$ denotes the group of permutations of $k$ elements and $\{\pi(1), \cdots, \pi(k)\}$ denotes these permutations. This map projects $(\mathbb{V}^n)^{\otimes k}$ onto the totally symmetric subspace $Sym^k$ of $(\mathbb{V}^n)^{\otimes k}$ whose dimension [44] is

$$dim(Sym^k) = \begin{pmatrix} n+k-1 \\ k \end{pmatrix}. \tag{26}$$

Roughly speaking, a complex projective $t$-design is a finite subset of the complex unit sphere in the complex vector space $\mathbb{C}^n$ with the property that the discrete average of any polynomial of degree t or less equals its uniform average. Many equivalent definitions-see e.g. [34, 41, 58]- capture this essence. Here, there is a more explicit definition of a $t$-design that is much more suitable for our purpose:

**Definition 2.6.** *( [29]) A finite set $\{\boldsymbol{w}_1, \cdots, \boldsymbol{w}_N\} \subset \mathbb{C}^n$ of normalized vectors is called a $t$-design of dimension $n$ if and only if*

$$\frac{1}{N} \sum_{i=1}^{N} (\boldsymbol{w}_i \boldsymbol{w}_i^*)^{\otimes t} = dim(Sym^t)^{-1} \cdot P_{Sym^t}, \tag{27}$$

*where $P_{Sym^t}$ denotes the projector onto the totally symmetric subspace* (25) *of $(\mathbb{C}^n)^{\otimes t}$ and consequently*

$$dimSym^t = \begin{pmatrix} n+t-1 \\ t \end{pmatrix}.$$

**Theorem 2.7.** *( [29]) Let $\boldsymbol{x}_0 \in \mathbb{C}^n$ be the unknown signal. Suppose that $\|\boldsymbol{x}_0\|_{\ell_2}$ is known and that $m$ measurement vectors $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_m$ have been sampled independently and uniformly at random from a $t$-design $D_t \subset \mathbb{C}^n (t \geq 3)$. Then, with probability at least $1 - e^{-\omega}$, PhaseLift*

(24) *recovers* $\boldsymbol{x}_0$ *up to a global phase, provided that the sampling rate exceeds*

$$m \geq c_0 \omega \cdot t n^{1+\frac{2}{t}} \log^2 n.$$

*Here* $\omega \geq 1$ *is an arbitrary parameter and* $c_0$ *is a universal constant.*

As the discussion of the previous subsection suggests, the bounds on the sampling rate decrease as the order of the design increases. For fixed $t$, and up to poly-log factors, it is proportional to $O(n^{1+\frac{2}{t}})$. This is sub-quadratic for the regime $t \geq 3$ where their arguments apply. If the degree is allowed to grow logarithmialy with the dimension (as $t = 2\log n$), they recovered an optimal, linear scaling up to a poly-log overhead, $m = O(n \log^3 n)$.

In light of the highly structured, analytical and exact designs known for degree 2 and 3, it is of great interest to ask whether a linear scaling can already be achieved for some small, fixed $t$. As shown by the following theorem, however, for $t = 2$ not even a sub-quadratic scaling is possible if no additional assumptions are made, irrespective of the reconstruction algorithm used.

**Theorem 2.8.** *(* [29]*) Let* $n$ *be a prime power larger than* 2*. Then there exists a* 2*-design* $D_2 \subset \mathbb{C}^n$ *and orthogonal, normalized vectors* $\boldsymbol{x}_0, \boldsymbol{z}$ *which have the following property: Suppose that* $m$ *measurement vectors* $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_m$ *are sampled independently and uniformly at random from* $D_2$*. Then, for any* $\omega \geq 0$*, the number of measurements must obey*

$$m \geq \frac{\omega \cdot n(n+1)}{4},$$

*or the event*

$$|\langle \boldsymbol{a}_k, \boldsymbol{x}_0 \rangle|^2 = |\langle \boldsymbol{a}_k, \boldsymbol{z} \rangle|^2, \quad \forall\ k \in \{1, \cdots, m\}$$

*will occur with probability at least* $e^{-\omega}$*.*

To prove the Theorem 2.7, the authors [29] first followed [17,28] to establish a certain injectivity property of the measurement operator $\mathcal{A}$. Compared to [17], their injectivity properties are somewhat weaker. The proof in [17] used the independence of the components of the Gaussian measurement operator, which is not available in this setting in [29], where individual vector components might be strongly correlated. Then they constructed an "approximate dual certificate" that proved that the sought-for signal indeed minimized the nuclear norm. Owing to the weaker bounds found here, the construction was more complicated than that in [17]. In the language of [28], they attempted to carry out the full "golfing scheme", as opposed to the "single leg" that proved sufficient in [17]. In addition, the defining properties of a maximal set of MUBs [39,41,66,81] allow them to derive the converse bound, namely, Theorem 2.8.

## 2.2 Uniform Guarantees for PhaseLift algorithm

Signal reconstruction from random measurements is a central preoccupation in contemporary signal processing. In this problem, we often acquire linear measurements of an unknown, structured signal through a random sampling process. Recently, Tropp [72] has described a simple approach that addresses a wide class of convex signal reconstruction problems involving

random sampling and allows us to analyze general measurements in a unified way. To understand these questions, the core challenge is to produce a lower bound on a nonnegative empirical process. This approach is referred as "the bowling scheme" relying on a powerful framework, called the Small Ball Method, that was developed by Mendelson and coauthors in a sequence of papers, including [40, 45, 52–54]. Although an important feature of the phase retrieval problem is that the signal $\boldsymbol{x}_0$ enters the measurement process (1) quadratically, which leads to a non-linear and non-convex inverse problem. Fortunately, this apparent obstacle of having nonlinear measurements can be overcome by noting that the measurement process -while quadratic in $\boldsymbol{x}_0$-is linear in the outer product $\boldsymbol{x}_0\boldsymbol{x}_0^*$ as described as (3). This connects the phase retrieval problem to the already extensive field of low-rank matrix recovery. Indeed, it is just a special case of low rank matrix recovery, where both the signal $\boldsymbol{X} = \boldsymbol{x}_0\boldsymbol{x}_0$ and the measurement matrices $\boldsymbol{A}_k = \boldsymbol{a}_k\boldsymbol{a}_k^*$ are constrained to be proportional to rank-one projectors. Further, we generally concentrate on Hermitian matrices $\boldsymbol{X} \in \mathbb{C}^{n\times n}$ and consider such rank-one measurements with noise vector $\boldsymbol{\varepsilon} \in \mathbb{R}^m$ and $\|\boldsymbol{\varepsilon}\|_{\ell_2} \leq \eta$. Then we can recovery $\boldsymbol{X}$ via nuclear norm minimization corresponds to

$$\min_{\boldsymbol{Z}\in\mathbb{H}^{n\times n}} \|\boldsymbol{Z}\|_1 \text{ s. t. } \|\boldsymbol{y} - \mathcal{A}(\boldsymbol{Z})\|_{\ell_2} \leq \eta. \tag{28}$$

Then such special low-rank matrix recovery problem can be uniformly solved by the Tropp's bowling scheme in [72] or a uniform variant of Tropp's bowling scheme which was further provided in [43]. The crucial ingredient of the bowling scheme is a new technique due to Mendelson [53] and Koltchiskii, Mendelson [40] to obtain lower bounds for quantities of the form $\inf_{\boldsymbol{u}\in E}\sum_{k=1}^{m}|\langle\boldsymbol{\phi}_k,\boldsymbol{x}\rangle|^2$ where the $\boldsymbol{\phi}_k$ are independent random vectors in $\mathbb{R}^d$ and $E$ is a subset of $\boldsymbol{u} \in E$ in $\mathbb{R}^d$. We start by recalling from [72] the notions and results underlying this technique.

Suppose we measure $\boldsymbol{x}_0 \in \mathbb{R}^d$ via measurements $\boldsymbol{y} = \Phi\boldsymbol{x}_0 + \boldsymbol{\varepsilon} \in \mathbb{R}^m$, where $\Phi$ is an $m \times d$ measurement matrix and $\boldsymbol{\varepsilon} \in \mathbb{R}^m$ a vector of unknown errors with $\|\boldsymbol{\varepsilon}\|_{\ell_2} \leq \eta$ where $\eta \geq 0$. Let $f : \mathbb{R}^d \to \mathbb{R}\cup\{\infty\}$ be a proper convex function, they aimed at recovering $\boldsymbol{x}_0$ by solving the convex program

$$\min f(\boldsymbol{x}) \text{ s. t. } \|\Phi\boldsymbol{x}_0 - \boldsymbol{y}\|_{\ell_2} \leq \eta. \tag{29}$$

Let $K \subset \mathbb{R}^d$ be a cone. Then we define the minimum singular value of $\Phi$ with respect to $K$ as

$$\lambda_{min}(\Phi; E) = \inf\{\|\Phi\boldsymbol{u}\|_{\ell_2}, \boldsymbol{u} \in K \cap \mathbb{S}^{d-1}\}. \tag{30}$$

where $\mathbb{S}^{d-1}$ is the unit sphere in $\mathbb{R}^d$. For $\boldsymbol{x} \in \mathbb{R}^d$, we consider the (convex) descent cone

$$\mathcal{D}(f, \boldsymbol{x}) = \bigcup_{\tau>0}\{\boldsymbol{y} \in \mathbb{R}^d : f(\boldsymbol{x} + \tau\boldsymbol{y}) \leq f(\boldsymbol{x})\}. \tag{31}$$

With these notions, the success of the convex program (29) can be estimated as follows.

**Lemma 2.9.** *( [72]) Let $\boldsymbol{x}_0 \in \mathbb{R}^d$, $\Phi \in \mathbb{R}^{m\times d}$ and $\boldsymbol{y} = \Phi\boldsymbol{x}_0+\boldsymbol{\varepsilon} \in \mathbb{R}^m$ with $\|\boldsymbol{\varepsilon}\|_{\ell_2} \leq \eta$ where $\eta \geq 0$. Let $f : \mathbb{R}^d \to \mathbb{R}\cup\{\infty\}$ be a proper convex function and $\hat{\boldsymbol{x}}$ be a solution of the corresponding convex program* (29)*. Then*

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}_0\|_{\ell_2} \leq \frac{2\eta}{\lambda_{min}(\Phi; \mathcal{D}(f, \boldsymbol{x}_0))}. \tag{32}$$

Then a crucial point is that in the situation that $\Phi$ is a random matrix with i.i.d. rows, the following result can be applied to estimate $\lambda_{min}(\Phi; \mathcal{D}(f, \boldsymbol{x}_0))$(see also [40, 72]).

**Lemma 2.10.** *(Koltchinskii, Mendelson; Tropp's version) Fix $E \subset \mathbb{R}^d$ and let $\boldsymbol{\phi}_1, \cdots, \boldsymbol{\phi}_m$ be independent copies of a random vector $\boldsymbol{\phi} \in \mathbb{R}^d$. For $\xi > 0$, let*

$$\begin{aligned} Q_\xi(E; \boldsymbol{\phi}) &= \inf_{\boldsymbol{u} \in E} \mathbb{P}\{|\langle \boldsymbol{\phi}, \boldsymbol{u} \rangle| \geq \xi\}, \\ W_m(E; \boldsymbol{\phi}) &= \mathbb{E} \sup_{\boldsymbol{u} \in E} \langle \boldsymbol{h}, \boldsymbol{u} \rangle, \ \textit{where } \boldsymbol{h} = \frac{1}{\sqrt{m}} \sum_{k=1}^m \epsilon_k \boldsymbol{\phi}_k. \end{aligned} \tag{33}$$

*with $(\epsilon_k)$ being a Rademacher sequence. Then for any $\xi > 0$ and any $t \geq 0$, with probability at least $1 - e^{-2t^2}$,*

$$\inf_{\boldsymbol{u} \in E} \Big(\sum_{k=1}^m |\langle \boldsymbol{\phi}_k, \boldsymbol{x} \rangle|^2\Big)^{\frac{1}{2}} \geq \xi \sqrt{m} Q_{2\xi}(E; \boldsymbol{\phi}) - 2W_m(E; \boldsymbol{\phi}) - \xi \cdot t. \tag{34}$$

Finally, they applied the notions in these results in the context of Theorems 2.11 and 2.13 as follows:

i) identify $\mathbb{H}^{n \times n}$ with $\mathbb{R}^d = \mathbb{R}^{n^2}$.

ii) $\Phi$ is the matrix of $\mathcal{A}$ in the standard basis, i.e., $\Phi(\boldsymbol{X})_k = tr(\boldsymbol{a}_k \boldsymbol{a}_k^* \boldsymbol{X})$.

iii) $f : \mathbb{H}^{n \times n} \to \mathbb{R} \cup \{\infty\}$ is the nuclear norm, i.e., $f(\boldsymbol{X}) = \|\boldsymbol{X}\|_1$.

The first main result gives a uniform and stable guarantee for recovering rank-$r$ matrices with $O(rn)$ rank one measurements that are proportional to projectors onto standard Gaussian random vectors.

**Theorem 2.11.** *( [43]) Consider the measurement process $\boldsymbol{y} = \mathcal{A}(\boldsymbol{X}) + \boldsymbol{\varepsilon}$ with $\|\varepsilon\|_{\ell_2} \leq \eta, \eta \geq 0$, and measurement matrices $A_k = \boldsymbol{a}_k \boldsymbol{a}_k^*$, where $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_m \in \mathbb{C}^n$ are independent standard Gaussian distributed random vectors. Furthermore assume that the number of measurements $m \geq C \cdot nr$ for $1 \leq r \leq n$ arbitrary. Then with probability at least $1 - e^{-C'm}$, it holds that for any positive semidefinite matrix $\boldsymbol{X} \in \mathbb{H}^{n \times n}$ with rank at most $r$, any solution $\hat{\boldsymbol{X}}$ to the convex optimization problem* (28) *obeys*

$$\|\hat{\boldsymbol{X}} - \boldsymbol{X}\|_2 \leq \frac{C_1 \eta}{\sqrt{m}}. \tag{35}$$

*Here, $C, C'$ and $C_1$ denote universal positive constants. (In particular, for $\eta = 0$, one has exact reconstruction.)*

For the rank one case $r = 1$, Theorem 2.11 essentially reproduced the main result in [11] which used completely different proof techniques. (More precisely, in [11] instead of the program (28), one minimized $\|\mathcal{A}(\boldsymbol{Z}) - \boldsymbol{y}\|_{\ell_1}$ where $\boldsymbol{Z}$ is positive semidefinite. Then with high probability for positive semidefinite $\boldsymbol{X}$ of rank-one and any minimizer $\hat{\boldsymbol{X}}$, the error estimate is $\|\hat{\boldsymbol{X}} - \boldsymbol{X}\|_1 \leq \frac{C_1 \|\varepsilon\|_{\ell_1}}{\sqrt{m}}$. Here, the error estimate $\frac{C_1 \|\varepsilon\|_{\ell_2}}{\sqrt{m}}$ is slightly weaker.) A variant of the above statement was shown in [72] to hold (in the real case) for a fixed matrix $\boldsymbol{X}$ of rank one. (More precisely, in [72] it was assumed that $\boldsymbol{X}$ was positive semidefinite and the optimization was performed wrt. the function $f$.) In fact, the proof reorganized and extended the arguments of [ [72], Section 8] in such a way, that Theorem 8.1 of [72] was shown to hold even uniformly (that is simultaneously

for all $\boldsymbol{X}$) and for arbitrary rank. In contrast to [11], they didn't need $\epsilon$-nets to show uniformity.

As we will see, similar result can also be obtained from more general measurements drawn independently from a weighted complex projective 4-design in the sense of Definition 2.12.

**Definition 2.12.** *([43]) For $t \in \mathbb{N}$, a finite set $\{\boldsymbol{w}_1, \cdots, \boldsymbol{w}_N\} \subset \mathbb{C}^n$ of normalized vectors with corresponding weights $\{p_1, \cdots, p_N\}$ such that $p_i \geq 0$ and $\sum_{i=1}^{N} p_i = 1$ is called a weighted complex projective $t$-design of dimension $n$ if and only if*

$$\frac{1}{N}\sum_{i=1}^{N} p_i(\boldsymbol{w}_i\boldsymbol{w}_i^*)^{\otimes t} = \int\limits_{\mathbb{CP}^{n-1}} (\boldsymbol{w}_i\boldsymbol{w}_i^*)^{\otimes t} dw, \tag{36}$$

*where the integral on the right hand side is taken with respect to the unique unitarily-invariant probability measure on the complex projective space $\mathbb{CP}^{n-1}$.*

In [29] exact complex projective $t$-designs have been applied to the problem of phase retrieval. The main result (Theorem 2.7) in [29] is a non-uniform exact recovery guarantee for phase retrieval via PhaseLift algorithm that requires $m = O(t \cdot n^{1+2/t} \log^2 n)$ measurement vectors that are drawn uniformly from a proper $t$-design ($t \geq 3$). The bowling scheme which was employed in [43], allowed for considerably generalizing and improving this statement.

**Theorem 2.13.** *([43]) Let $\{p_i, w_i\}_{i=1}^N$ be a weighted 4-design and consider the measurement matrices $A_k = \sqrt{n \cdot (n+1)}\boldsymbol{a}_k\boldsymbol{a}_k^*$, where $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_m \in \mathbb{C}^n$ are drawn independently from $\{p_i, w_i\}_{i=1}^N$. Furthermore assume that the number of measurements $m$ obeys $m \geq C \cdot nr \log n$ for $1 \leq r \leq n$ arbitrary. Then with probability at least $1 - e^{-C'm}$, it holds that for any positive semidefinite matrix $\boldsymbol{X} \in \mathbb{H}^{n\times n}$ with rank at most $r$, any solution $\hat{\boldsymbol{X}}$ to the convex optimization problem (28) with noisy measurements $\boldsymbol{y} = \mathcal{A}(\boldsymbol{X}) + \boldsymbol{\varepsilon}$ where $\|\varepsilon\|_{\ell_2} \leq \eta, \eta \geq 0$, obeys*

$$\|\hat{\boldsymbol{X}} - \boldsymbol{X}\|_2 \leq \frac{C_1\eta}{\sqrt{m}}. \tag{37}$$

*Here, $C, C'$ and $C_1$ denote universal positive constants. (In particular, for $\eta = 0$, one has exact reconstruction.)*

The normalization factor $\sqrt{n \cdot (n+1)}$ leads to approximately the same normalization of the $A_k$ (wrt. the Frobenius norm) as in expectation in the Gauss case. This theorem is a stable, uniform guarantee for recovering arbitrary Hermitian matrices of rank at most $r$ with high probability using the convex optimization problem (28) and $m = O(nr \log n)$ measurements drawn independently (according to the design's weights) from a weighted 4-design. It obviously covers sampling from 4-designs as a special case.

Also, Theorem 2.13 is close to optimal in terms of the design order $t$ required. In the context of the phase retrieval problem, it was shown from Theorem 2.8, that choosing measurements uniformly from a proper 2-design does not allow for a sub-quadratic sampling rate $m$ without additional structural assumptions on the measurement ensemble. It is presently open whether Theorem 2.13 also holds for 3-designs.

In a word, the so-called bowling scheme was developed in [43, 72] as a useful method that guaranteed successful uniform recovery of Hermitian rank $r$ matrices, either for the vectors

$\boldsymbol{a}_k, k = 1, \cdots, m$, being chosen independently at random according to a standard Gaussian distribution, or $\boldsymbol{a}_k$ being sampled independently from an complex projective $t$-design with $t = 4$. Particularly, when the matrix $\boldsymbol{X} = \boldsymbol{x}\boldsymbol{x}^*$ to be recovered is of rank one, then such low rank matrix recovery problem actually reduces to the problem of phaseless estimation via the PhaseLift algorithm. Accordingly, the Theorem 2.11 and 2.13 remain valid. In addition, many researchers have also been working on the general phase retrieval problem via PhaseLift algorithm or the extended low rank matrix recovery from more general rank-one measurements, for instance, sub-gaussian measurements, Bernoulli measurements with erasures, and so forth, see [42, 51].

## §3 PhaseLift under Additional Structured Measurements

In this section, we are going to study the phase retrieval problem from additional structured measurements. As we can see from Section 2, a line of works establish that if the sampling vectors $\boldsymbol{a}_k$ are sufficiently randomized, then the original signal $\boldsymbol{x}_0$ can be uniquely or stably recovered from on the order of $n$ equations only or up to log-factors via the PhaseLift algorithm. While this is all reassuring, the problem is that the Gaussian model, in which each measurement gives us the magnitude of the dot product $\sum_{i=1}^{n} x[i]a_k[i]$ between the signal and (complex-valued) Gaussian white noise, is very far from the kind of data one can collect in an X-ray imaging and many related experiments. The purpose of this section is to show that the PhaseLift relaxation is still exact in a physically inspired setup where one can modulate the signal of interest and then let diffraction occur.

In practical applications, common approaches include the use of modulated light beams and masks right after the sample, see [35, 84]. Hence, a structured measurements setup closely related to applications was originally proposed in [21]. Mathematically, such structured measurements we consider here can be expressed as follows:

$$y_{k,l} = \left| \sum_{j=1}^{n} x_j \bar{\epsilon}_{l,j} e^{-\frac{i2\pi kj}{n}} \right|^2, \quad 1 \le k \le n, 1 \le l \le L. \tag{38}$$

Here $\epsilon_{l,j}$ is the code for modulating the signal. Namely, we collect the magnitudes of the discrete Fourier transform of $L$ modulations of the signal $\boldsymbol{x}_0 = [x_0[1], \ldots, x_0[n]]^\top$. Denote the $k$-th discrete Fourier vector by $\boldsymbol{f}_k = \sum_{j=1}^{n} \omega^{jk} \boldsymbol{e}_j$ with $\omega := e^{\frac{2\pi i}{n}}$ a $n$-th root of unity and set $\boldsymbol{D}_l = \sum_{i=1}^{n} \epsilon_{l,i} \boldsymbol{e}_i \boldsymbol{e}_i^*$. Then the measurements (38) can be realized by the composition of diagonal matrices and Fourier transforms, namely,

$$y_{k,l} = |\boldsymbol{f}_k^* \boldsymbol{D}_l^* \boldsymbol{x}_0|^2, \quad 1 \le k \le n, 1 \le l \le L. \tag{39}$$

These diagonal matrices are often called as masks. Then combined with the "lifting" technique [3, 10], we can translate the measurements (38) into the matrix form

$$y_{k,l} = tr(\boldsymbol{F}_{k,l} \boldsymbol{X}_0), \quad 1 \le k \le n, 1 \le l \le L, \tag{40}$$

where $\boldsymbol{F}_{k,l} = \boldsymbol{D}_l \boldsymbol{f}_k \boldsymbol{f}_k^* \boldsymbol{D}_l^*$. Finally, $\boldsymbol{y} = \mathcal{A}(\boldsymbol{X})$ gives the noiseless structured measurements (38) with the linear operator $\mathcal{A}$ related to the measure vector $\boldsymbol{D}_l \boldsymbol{f}_k$.

Correspondingly, there has been a lot of interest in PhaseLift algorithm to solve the phase retrieval problem under Fourier measurements with masks. In what follows, two kinds of random masks have been studied. First, Candès et al. [12] worked with diagonal matrices $\boldsymbol{D}_l,\ 1 \le l \le L$, which are i.i.d. copies of a matrix $\boldsymbol{D}$, whose entries $\epsilon_{l,j}$ $(1 \le j \le n,\ 1 \le l \le L)$ are i.i.d. copies of a complex random variable $\epsilon$ which obeys

$$\begin{aligned} &\mathbb{E}[\epsilon] = \mathbb{E}[\epsilon^2] = 0, \\ &|\epsilon| \le M \quad \textit{almost surely for some } M > 0, \\ &\mathbb{E}[|\epsilon|^4] = 2\mathbb{E}[|\epsilon|^2]^2. \end{aligned} \tag{41}$$

Thus they demonstrated that with high probability, $O(\log^4 n)$ random complex masks are enough to exactly reconstruct complex signals.

**Theorem 3.1.** *( [12]) Let $\boldsymbol{x}_0 \in \mathbb{C}^n$ be an unknown signal and suppose that the number $L$ masks (41) obeys $L \ge c\gamma \log^4 n$. Then with probability at least $1 - \frac{1}{n^\gamma}$, the PhaseLift (5) reduces to a unique point $\boldsymbol{x}_0 \boldsymbol{x}_0^*$, and thus recovers $\boldsymbol{x}_0$ up to a global phase. Here $\gamma \ge 1$ and $c > 0$ are absolute constants.*

Remarkably, the phase recovery problem is different than that in which the sampling vectors are Gaussian as in [11, 17]. The reason is that the measurements in Theorem 3.1 are far more structured and far "less random". Loosely speaking, the random modulation model in [12] used on the order of $m := nL$ random bits whereas the Gaussian model with the same number of quadratic equations would use on the order of $m \cdot n$ random bits. A consequence of this difference is that the proof of the theorem requires new techniques and ideas. Having said this, an open and interesting research direction is to close the gap-remove the log factors-and show whether or not perfect recovery can be achieved from a number of masks independent of dimension.

In order to reduce the number of masks required to estimate the true signal via PhaseLift algorithm, Gross et al. [30] designed different masks $\boldsymbol{D}_l,\ 1 \le l \le L$, which are i.i.d. copies of a matrix $\boldsymbol{D}$, whose entries $\epsilon_{l,j}$ $(1 \le j \le n,\ 1 \le l \le L)$ are i.i.d. copies of a real random variable $\epsilon$ which obeys

$$\begin{aligned} &\mathbb{E}[\epsilon] = \mathbb{E}[\epsilon^3] = 0, \\ &|\epsilon| \le M \quad \textit{almost surely for some } M > 0, \\ &\mathbb{E}[\epsilon^4] = 2\mathbb{E}[\epsilon^2]^2, \ \textit{and} \ \ \nu := \mathbb{E}[\epsilon^2]. \end{aligned} \tag{42}$$

Combined with this type of random masks, they showed that recovery of $\boldsymbol{x}_0$ would be guaranteed for $L \ge c \log^2 n$, provided that the signal's intensity $y' = \|\boldsymbol{x}_0\|_{\ell_2}^2$ is known.

**Theorem 3.2.** *( [30]) Let $\boldsymbol{x}_0 \in \mathbb{R}^n$ be an unknown signal with $y' = 1$ and let $n > 3$ be an odd number. Suppose that $L$ real masks are used with entries being independent copies of a random variable obeying (42). Then with probability at least $1 - e^{-\omega}$, the feasibility problem (24) exactly recovers $\boldsymbol{x}_0$ up to a global phase, provided that $L \ge c\omega \log^2 n$. Here $\omega \ge 1$ is an*

*arbitrary parameter and c is a constant which is independent of the signal's dimension.*

It's noteworthy that the Theorem 3.2 implies that $O(\log^2 n)$ real masks with entires obeying (42) are sufficient to recover complex signals of odd dimensions. Meanwhile, they thought that this result remained valid for complex signals of even dimensions. Consequently, we have recently worked on the problem about reducing sampling complexity of masked Fourier measurements. We found that similar result as stated as Theorem 3.2 holds true for real signals of even dimensions, but doesn't hold for complex and even dimensional signals, see [47]. Meanwhile, since in the real world, measurements are often contaminated by noise. We studied the stability and robustness of the PhaseLift algorithm under the masked Fourier measurements with random masks. Then we demonstrated that the PhaseLift algorithm can stably estimate the true signal from a number of random masks, which is poly-logarithmic in the number of unknowns, see [47].

In addition, the authors in [37] considered specific masks instead of random masks. They showed that two specific simple masks (each mask provides $2n$ measurements) or five specific simple masks (each mask provides $n$ measurements) are sufficient to provably and stably recover almost all signals. For the first setting, namely, assume that the $2n$-DFT was applied in the measurement process with specific masks $\boldsymbol{D}_i,\ i = 1, 2$ with entries defined as follows,

$$\epsilon_{1,i} = 1, \quad 1 \le i \le n, \qquad \epsilon_{2,i} := \begin{cases} 0 & \text{if } i = 1, \\ 1 & \text{if } 2 \le i \le n, \end{cases} \tag{43}$$

then the related result is

**Theorem 3.3.** *( [37]) Consider any arbitrary signal $\boldsymbol{x}_0 \in \mathbb{C}^n$ such that $x_0[1] \neq 0$. Suppose measurements are taken with the masks defined by $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$, the convex program* (5) *has a unique feasible point, namely, $\boldsymbol{x}_0\boldsymbol{x}_0^*$, and hence $\boldsymbol{x}_0$ can be uniquely recovered (up to a global phase).*

For the second setting, namely, assume that the $n$-DFT was applied in the measurement process with specific masks $\boldsymbol{D}_i,\ i = 3, \cdots, 7$ with entries defined as follows:

$$\epsilon_{3,i} := \begin{cases} 1 & \text{if } 1 \le i \le \frac{n}{2}, \\ 0 & \text{if } \frac{n}{2} \le i \le n, \end{cases} \qquad \epsilon_{4,i} := \begin{cases} 0, & i = 1 \\ 1 & \text{if } 2 \le i \le \frac{n}{2}, \\ 0 & \text{if } \frac{n}{2} \le i \le n, \end{cases} \tag{44}$$

$$\epsilon_{5,i} := \begin{cases} 0 & \text{if } 1 \le i \le \frac{n}{2} + 1, \\ 1 & \text{if } \frac{n}{2} + 2 \le i \le n, \end{cases} \qquad \epsilon_{6,i} := \begin{cases} 0 & \text{if } 1 \le i \le \frac{n}{2}, \\ 1 & \text{if } \frac{n}{2} \le i \le n, \end{cases} \tag{45}$$

and

$$\epsilon_{7,i} := \begin{cases} 0, & 1 \le i \le \frac{n}{4} \\ 1 & \text{if } \frac{n}{4} + 1 \le i \le \frac{3n}{4}, \\ 0 & \text{if } \frac{3n}{4} + 1 \le i \le n. \end{cases} \tag{46}$$

**Theorem 3.4.** *( [37]) Consider any arbitrary signal $\boldsymbol{x}_0 \in \mathbb{C}^n$ such that $x_0[1], x[\frac{n}{2}], x[\frac{n}{2}+1] \neq 0$. Suppose measurements are taken with the masks defined by $\boldsymbol{D}_i,\ i = 3, \cdots, 7$, the convex program*

(5) *has a unique feasible point, namely,* $\boldsymbol{x}_0\boldsymbol{x}_0^*$*, and hence* $\boldsymbol{x}_0$ *can be uniquely recovered (up to a global phase).*

In the noisy case, suppose the noise corresponding to each measurement is bounded by $\eta$. Then they considered the solution to

$$\begin{aligned} &\min_{\boldsymbol{X}} \ Tr(\boldsymbol{X}) \\ &\text{s. t. } \|\boldsymbol{y}-\mathcal{A}(\boldsymbol{X})\|_\infty \le \eta, \\ &\qquad \boldsymbol{X} \succeq \boldsymbol{0}. \end{aligned} \tag{47}$$

to estimate $\boldsymbol{x}_0$.

**Theorem 3.5.** *(* [37]*) Consider any arbitrary signal* $\boldsymbol{x}_0 \in \mathbb{C}^n$ *such that* $\|\boldsymbol{x}_0\|_{\ell_1} \le \beta$ *and* $|x_0[1]| \ge \gamma > 0$ *for some* $\beta, \gamma$*. Suppose measurements are taken with the masks defined by* $\boldsymbol{D}_i,\ i = 1, 2$*, the solution to the convex program* (47) $\hat{\boldsymbol{X}}$ *obeys*

$$\|\boldsymbol{x}_0\boldsymbol{x}_0^* - \hat{\boldsymbol{X}}\|_2 \le C(\beta,\gamma)\cdot\eta,$$

*for some numerical constant* $C(\beta, \gamma)$*.*

## §4 PhaseLift for Sparse Signal

In this section, we want to investigate the sparse phase retrieval problem

$$y_k = |\langle \boldsymbol{a}_k, \boldsymbol{x}_0\rangle|^2,\ 1 \le k \le m, \tag{48}$$

for a $s$-sparse signal $\boldsymbol{x}_0 \in \mathbb{C}^n$. As stated in [46], if the unknown vector $\boldsymbol{x}_0$ is assumed to be $s$-sparse, then under some mild conditions on the number of measurements, system (48) becomes well-posed.

**Theorem 4.1.** *(* [46]*) Let* $\boldsymbol{x}_0 \in \mathbb{R}^n$ *be a* $s$*-sparse signal,* $\boldsymbol{a}_k \in \mathbb{R}^n, k = 1, \cdots, m_1$ *be generic measurement vectors and let* $\boldsymbol{z}_0 \in \mathbb{C}^n$ *be a* $s$*-sparse complex signal and* $\boldsymbol{b}_k \in \mathbb{C}^n, k = 1, \cdots, m_2$ *be generic measurement vectors. Then* $m_1 \ge 4s-1$*,* $m_2 \ge 8s-2$ *quadratic measurements* $\{|\langle \boldsymbol{a}_k, \boldsymbol{x}_0\rangle|^2\}_{k=1}^{m_1}$*,* $\{|\langle \boldsymbol{b}_k, \boldsymbol{z}_0\rangle|^2\}_{k=1}^{m_2}$ *are sufficient to recover* $\boldsymbol{x}_0$ *and* $\boldsymbol{z}_0$ *up to a global phase.*

Here, by generic they mean an open dense subset of the set of all $m$-element frames in $\mathbb{R}^n$ or $\mathbb{C}^n$.

Later, the authors in [78] studied the problem of minimal number of samples (measurements) required for $s$-sparse phase retrieval. To begin with, they say that a set of vectors $\mathcal{A}_m$ has the phase retrieval property, or is phase retrievable, if $M_{\mathcal{A}_m}$ is injective on $\tilde{H} := \mathbb{R}^n/\sim$ (or $\mathbb{C}^n/\sim$) where the operator $M_{\mathcal{A}_m}$ is defined as following:

$$\begin{aligned} M_{\mathcal{A}_m}:\ \tilde{H} &\longrightarrow \mathbb{R}^m \\ \boldsymbol{x}_0 &\longrightarrow M_{\mathcal{A}_m}(\boldsymbol{x}_0) = [|\langle \boldsymbol{a}_1, \boldsymbol{x}_0\rangle|^2, \cdots, |\langle \boldsymbol{a}_m, \boldsymbol{x}_0\rangle|^2]^\top, \end{aligned}$$

and the equivalence relation $\sim$ on $\mathbb{R}^n$(or $\mathbb{C}^n$): $\boldsymbol{x} \sim \boldsymbol{z}$ if and only if there is a constant $c \in \mathbb{R}$(or $\mathbb{C}$) with $|c| = 1$ such that $\boldsymbol{x} = c\cdot\boldsymbol{z}$. When $\tilde{H} := \mathbb{R}^n/\sim$, they have completely settled the minimality question for $s$-sparse phase retrieval in the real case.

**Theorem 4.2.** *( [78]) Let $\mathcal{A}_m = \{\boldsymbol{a}_1, \cdots, \boldsymbol{a}_m\}$ be a set of vectors in $\mathbb{R}^n$. Assume that $\mathcal{A}_m$ is $s$-sparse phase retrievable on $\mathbb{R}^n$. Then $m \geq \min\{2s, 2n-1\}$. Furthermore, a set $\mathcal{A}_m$ of $m \geq \min\{2s, 2n-1\}$ generically chosen vectors in $\mathbb{R}^n$ is $s$-sparse phase retrievable.*

When $\tilde{H} := \mathbb{C}^n/\sim$, they showed that

**Theorem 4.3.** *( [78]) A set $\mathcal{A}_m$ of $m \geq 4s-2$ generically chosen vectors in $\mathbb{C}^n$ is $s$-sparse phase retrievable.*

Although the above theorem showed that in the complex case any $m \geq 4s-2$ generically chosen vectors are $s$-sparse phase retrievable, it is unknown whether $4s-2$ is in fact the minimal number required. It is still an open question to be solved.

Meanwhile, it is well known that there have been many kinds of convex methods [1,36,46,60] and a group of non-convex methods to tackle with the sparse phase retrieval problem, see [8,56,57,63,68,69,79]. One of convex optimization methods in [46] has recently been proven to be very successful in solving the sparse phase retrieval problem from random Gaussian measurements. Inspired by the success of convex relaxations in compressed sensing and the PhaseLift algorithm for phase retrieval, it is natural to leverage the sparsity assumption to try to efficiently recover signals from fewer than $n$ intensity measurements. A convex formulation in this direction, which, to the best of our knowledge, was first proposed in [60] to solve (48), is the following program:

$$\begin{aligned} \min_{\boldsymbol{X}} \ & Tr(\boldsymbol{X}) + \lambda \cdot \|\boldsymbol{X}\|_{1,1} \\ \text{s. t. } & \mathcal{A}(\boldsymbol{X}) = \boldsymbol{y}, \\ & \boldsymbol{X} \succeq \boldsymbol{0}. \end{aligned} \tag{49}$$

Then the following theorem provided in [46] showed that if $\{\boldsymbol{a}_k\}_{k=1}^m$ are i.i.d standard normal random vectors, the solution to (49) for an appropriate choice of $\lambda$, is exactly $\boldsymbol{x}_0\boldsymbol{x}_0^\top$, provided that $s \leq O(\sqrt{\frac{m}{\log n}})$.

**Theorem 4.4.** *( [46]) Fix a signal $\boldsymbol{x}_0 \in \mathbb{R}^n$ with $\|\boldsymbol{x}_0\|_{\ell_2} = 1$ and $\|\boldsymbol{x}_0\|_0 = s$. Let $\boldsymbol{a}_k \in \mathbb{R}^n$ be i.i.d standard normal random vectors. Then the solution to the convex program (49) is exact with probability at least $1 - (2\log n + 3)(4e^{-\gamma \cdot \frac{m}{2\log n+3}} + \frac{1}{n^3}) - (5 + 2n^2)e^{-\gamma \cdot m}$, provided $\lambda > \sqrt{s}\|\boldsymbol{x}_0\|_{\ell_1} + 1$, $\lambda < \frac{n^2}{4}$ and $m \geq C\lambda^2 \log n$. Here $C$ and $\gamma$ are numerical constants.*

**Remark 4.5.** *By choosing $\lambda = \sqrt{\frac{m}{4C\log n}}$, we have exact recovery with probability at least $1 - (2\log n + 3)(4e^{-\gamma \cdot \frac{m}{2\log n+3}} + \frac{1}{n^3}) - (5 + 2n^2)e^{-\gamma \cdot m}$ if the number of measurements obeys $m \geq O(\|\boldsymbol{x}_0\|_{\ell_1}^2 \cdot s\log n)$. Moreover, by choosing $\boldsymbol{x}_0$ to be a $s$-sparse vector with components $x_k = \pm\frac{1}{\sqrt{s}}$, $k = 1, \cdots, n$, this reads $m \geq O(s^2 \log n)$.*

This following theorem obtained sharp theoretical results on the performance of (49) in the Gaussian quadratic measurement setting, which may be surprising since it implies that there is a substantial gap between the sufficient number of measurements for injectivity and the necessary number of measurements for recovery via a class of natural convex relaxations.

**Theorem 4.6.** *( [46]) Under the setting of Theorem* 4.2*, assuming* $4 \leq s \leq m \leq \frac{n}{40\log n}$*, then there is an event* $E$ *with probability at least* $1 - \frac{m}{n^5} - m \cdot e^{-0.09n+0.09s+0.79m}$*, such that the following property holds: If there exists a lambda* $\in \mathbb{R}$ *such that* $\boldsymbol{x}_0\boldsymbol{x}_0^\top$ *is a minimizer of* (49)*, then we have*

$$m \geq \min\{(\frac{s}{4}-1)^2, \frac{\max\{\|\boldsymbol{x}_0\|_{\ell_1}^2 - \frac{s}{2}, 0\}^2}{500\log^2 n}\}. \tag{50}$$

**Remark 4.7.** *Taking* $\boldsymbol{x}_0$ *to be a* $s$*-sparse vector with components* $x_k = \pm\frac{1}{\sqrt{s}}$*,* $k = 1, \cdots, n$*, this reads* $m \geq O(\frac{s^2}{\log^2 n})$*.*

However, such methods are either lack of analyses of optimal sample complexity or recovery guarantees for this problem under masked Fourier measurements. Recently, the authors in [1,36] independently proposed a two-stage algorithm which consisted of a low-rank recovery stage and a sparse recovery stage to recover a sparse and rank-one matrix. They show that the $s$-sparse $\boldsymbol{x}_0 \in \mathbb{R}^n$ can be estimated robust to noise from $O(s\log(\frac{n}{s}))$ composite random Gaussian measurements. So in [47], we were inspired to apply this two-stage algorithm and more structured measurements to recover the $s$-sparse signal $\boldsymbol{x}_0$. The structured measurements we used there, namely, $\boldsymbol{a}_i$ $(i = 1, \ldots, m)$, were composed of masked Fourier measurement and random matrices (for instance, Bernoulli random matrix or partial Fourier matrix). Based on the stable guarantees established for general signals, we proved that the $s$-sparse signal $\boldsymbol{x}_0$ can be stably estimated from $O(s\log(\frac{en}{s})\log^4(s\log(\frac{en}{s})))$ corrupted composite measurements.

## References

[1] S Bahmani, J Romberg. *Efficient compressive phase retrieval with constrained sensing vectors*, IEEE Neural Information Processing Systems, 2015: 523-531.

[2] S Bahmani, J Romberg. *A flexible convex relaxation for phase retrieval*, Electron J Stat, 2017, 11(2): 5254-5281.

[3] R Balan, B G Bodmann, P G Cassazza, D Edidin. *Painless reconstruction from magnitudes of frame coefficients*, J Fourier Anal Appl, 2015, 15(4): 488-501.

[4] R Balan, B G Bodmann, P G Cassazza, D Edidin. *Fast algorithms for signal reconstruction without phase*, Proc SPIE, 2007, 6701: 67011L-67011L-9.

[5] R Balan, P Casazza D Edidin. *On signal reconstruction without noisy phase*, Appl Comp Harm Anal, 2006, 20: 345-356.

[6] B Baykal. *Blind channel estimation via combining autocorrelation and blind phase estimation*, IEEE Transactions on Circuits and Systems, 2004, 51(6): 1125-1131.

[7] R Berinde, A C Gilbert, P Indyk, H Karloff, M J Strauss. *Combining geometry and combinatorics: A unified approach to sparse signal recovery*, 2008 Forty-sixth Annual Allerton Conference on Communication, Control, and Computing, 2008: 798-805.

[8] T Cai, X Li, Z Ma. *Optimal rates of convergence for noisy sparse phase retrieval via thresholded Wirtinger flow*, Ann Statist, 2016, 44(5): 2221-2251.

[9] J F Cai, K Wei. *Solving systems of phaseless equations via Riemannian optimization with optimal sampling complexity*, arXiv preprint arXiv:1809.02773, 2018.

[10] E J Candès, Y C Eldar, T Strohmer, V Voroninshi. *Phase retrieval via matrix completion*, SIAM J Imaging Sci, 2013, 6(1): 199-225.

[11] E J Candès, X Li. *Solving quadratic equations via phaselift when there are about as many equations as unknowns*, Found Comput Math, 2014, 14(5): 1017-1026.

[12] E J Candès, X Li, M Soltanolkotabi. *Phase retrieval from coded diffraction patterns*, Appl Comput Harmon Anal, 2015, 39(2): 277-299.

[13] E J Candès, X Li, M Soltanolkotabi. *Phase retrieval via Wirtinger Flow: theory and algorithms*, IEEE Trans Inform Theory, 2015, 61(4): 1985-2007.

[14] E J Candès, Y Plan. *Matrix completion with noise*, IEEE Proc, 2010, 98(6): 925-936.

[15] E J Candès, T Tao. *Decoding by linear programming*, IEEE Trans Inform Theory, 2005, 51(12): 4203-4215.

[16] E J Candès, B Recht. *Exact matrix completion via convex optimization*, Found Comput Math, 2009, 9(6): 717-772.

[17] E J Candès, T Stromher, V Voroninshi. *PhaseLift: exact and stable signal recovery from magnitude measurements via convex programming*, Comm Pure Appl Math, 2013, 66(8): 1241-1274.

[18] Y Chen, E J Candès. *Solving random quadratic systems of equations is nearly as easy as solving linear systems*, Comm Pure Appl Math, 2015, 70(5): 739-747.

[19] C C Chen, J Miao, C W Wang, T K Lee. *Application of the optimization technique to noncrystalline X-ray diffraction microscopy: Guided hybrid input-output method(GHIO)*, Phys Rev B, 2007, 76: 064113.

[20] J C Dainty, J R Fienup. *Array imaging using intensity-only measurements, phase retrieval and image reconstruction for astronomy*, H. Stark (Ed), Image Recovery: Theory and Application, Academic Press, San Diego, 1987: 231-275.

[21] A Fannjiang, W Liao. *Phase retrieval with random phase illumination*, J Opt Soc Amer A, 2012, 29(9): 1847-1859.

[22] J R Fienup. *Reconstruction of an object from the modulus of its Fourier transform*, Opt Lett, 1978, 3(1): 27-29.

[23] J R Fienup. *Phase retrieval algorithms: a comparison*, 1982, Appl Opt, 21(15): 2758-2769.

[24] J Finkelstein. *Pure-state informationally complete and "really" complete measurements*, Phys Rev A, 2004, 70(5): 052107.

[25] R W Gerchberg, W O Saxton. *A Practical algorithm for the determination of phase from image and diffraction plane pictures*, Optik, 1972, 35(2): 237-246.

[26] M X Goemans, D P Williamson. *Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming*, Journal of the ACM, 1995, 42(6): 1115-1145.

[27] T Goldstein, C Studer. *PhaseMax: convex phase retrieval via basis pursuit*, IEEE Trans Inform Theory, 2018, 64(4):2675-2689.

[28] D Gross. *Recovering low-rank matrices from few coefficients in any basis*, IEEE Trans Inform Theory, 2011, 57(3): 1548-1566.

[29] D Gross, F Krahmer, R Kueng. *A partial de-randomization of PhaseLift using spherical designs*, J Fourier Anal Appl, 2015, 21(2): 229-266.

[30] D Gross, F Krahmer, R Kueng. *Improved recovery guarantees for phase retrieval from coded diffraction patterns*, Appl Comput Harmon Anal, 2017, 42(1): 37-64.

[31] D Gross, Y K Liu, S T Flammia, S Becker, J Eisert. *Quantum-state tomography via compressed sensing*, Phys Rev Lett, 2010, 105(15): 150401.

[32] M H Hayes. *The reconstruction of a multidimensional sequence from the phase or magnitude of its Fourier transform*, IEEE Transactions on Acoustics, Speech and Signal Processing, 1982, 30(2): 140-154.

[33] E M Hofstetter. *Construction of time-limited functions with specified autocorrelation functions*, IEEE Trans Inform Theory, 1964, 10(2): 119-126.

[34] S Hoggar. *t-Designs in projective spaces*, Eur J Comb, 1982, 3(3): 233-254.

[35] M J Humphry, B Kraus, A C Hurst, A M Maiden, J M Rodenburg. *Ptychographic electron microscopy using high-angle dark-field scattering for sub-nanometre resolution imaging*, Nature Communications, 2012, 3(1): 730.

[36] M Iwen, A Viswanathan, Y Wang. *Robust sparse phase retrieval made easy*, Appl Comput Harmon Anal, 2017, 42(1): 135-142.

[37] K Jaganathan, Y Eldar, B Hassibi. *Phase retrieval with masks using convex optimization*, IEEE International Symposium on Information Theory, 2015: 1655-1659.

[38] K Jaganathan, Y Eldar, B Hassibi. *Phase Retrieval: An Overview of Recent Developments*, arXiv preprint arXiv: Information Theory, 2015.

[39] A Klappenecker, M Rotteler. *Mutually unbiased bases are complex projective 2-designs*, IEEE International Symposium on Information Theory, 2005, 1: 1740-1744.

[40] V Koltchinskii, S Mendelson. *Bounding the smallest singular value of a random matrix without concentration*, International Mathematics Research Notices, 2015, 23: 12991-13008.

[41] H König. *Cubature formulas on spheres*, Advances in Multivariate Approximation, Proceedings of the 3rd International Conference on Multivariate Approximation Theory, 1999: 201-211.

[42] F Kramher, Y Liu. *Phase retrieval without small-ball probability assumptions*, IEEE Trans Inform Theory, 2018, 64(1): 485-500.

[43] R Kueng, H Rauhut, U Terstiege. *Low rank matrix recovery from rank one measurements*, Appl Comput Harmon Anal, 2014, 42(1): 88-116.

[44] J M Landsberg. *Tensors: Geometry and Applications*, American Mathematical Society (AMS), 2012.

[45] G Lecué, S Mendelson. *Compressed sensing under weak moment assumptions*, arXiv preprint arXiv:1401.2188, 2014.

[46] X Li, V Voroninski. *Sparse signal recovery from quadratic measurements via convex programming*, SIAM J Math Anal, 2013, 45(5): 3019-3033.

[47] H Li, S Li. *Phase retrieval from coded diffraction patterns with noise*, Submitted, 2019.

[48] H Li, S Li, Y Xia. *PhaseMax: Stable guarantees from noisy sub-gaussian measurements*, Anal Appl, Online, 2019.

[49] S Marchesini. *Phase retrieval and saddle-point optimization*, J Opt Soc Amer A, 2007, 24(10): 3289-3296.

[50] S Marchesini. *A unified evaluation of iterative projection algorithms for phase retrieval*, Rev Sci Instrum, 2007, 78(1): 011301.

[51] K Maryia, K Richard, R Holger, U Terstiege. *Stable low rank matrix recovery from null space property*, Information and Inference: A Journal of the IMA, 2016, 5(4): 405-441.

[52] S Mendelson. *A remark on the diameter of random sections of convex bodies*, arXiv preprint arXiv: 1312.3608, 2014.

[53] S Mendelson. *Learning without concentration*, J Assoc Comput Mach, 2015, 62(3): 21.

[54] S Mendelson. *Learning without concentration for general loss functions*, Probability Theory and Related Fields, 2018, 171(1): 459-502.

[55] R Millane. *Phase retrieval in crystallography and optics*, J Opt Soc Amer A, 1990, 7(3): 394-411.

[56] M L Moravec, J K Romberg, R G Baraniuk. *Compressive phase retrieval*, Proceedings of SPIE, 2007, 6701: 6701201-67012011.

[57] P Netrapalli, P Jain, S Sanghavi. *Phase retrieval using alternating minimization*, IEEE Trans Signal Process, 2015, 63(18): 4814-4826.

[58] A Neumaier. *Combinatorial Configurations in Terms of Distances*, Department of Mathematics Memorandum, Eindhoven, 1981.

[59] K A Nugent, A G Peele, H N Chapman, A P Manusco. *Unique phase recovery for non-periodic objects*, Phys Rev Lett, 2003, 91(20): 203902.

[60] H Ohlsson, A Yang, R Dong, S Sastry. *CPRL-an extension of compressive sensing to the phase retrieval problem*, IEEE Neural Information Processing Systems, 2012: 1367-1375.

[61] A L Patterson. *A Fourier series method for the determination of the components of interatomic distances in crystals*, Phys Rev, 1934, 46(5): 372-376.

[62] A L Patterson. *Ambiguities in the X-ray analysis of crystal structures*, Phys Rev, 1944, 46: 195-201.

[63] R Pedarsani, D Yin, K Lee, K Ramchandran. *PhaseCode: fast and efficient compressive phase retrieval based on sparse-graph codes*, IEEE Trans Inform Theory, 2017, 63(6): 3663-3691.

[64] L Rabiner, B H Juang. *Fundamentals of speech recognition*, Prentice Hall, 1993.

[65] B Recht, M Fazel, P Parrilo. *Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization*, SIAM Rev, 2010, 52(3): 471-501.

[66] B Sanderson. *Immersions and embeddings of projective spaces*, Proc Lond Math Soc, 1964, 1: 137-153.

[67] J L C Sanz. *Mathematical considerations for the problem of Fourier transform phase retrieval from magnitude*, SIAM J Appl Math, 1985, 45(4): 651-664.

[68] Y Shechtman, A Beck, Y C Eldar. *GESPAR: efficient phase retrieval of sparse signals*, IEEE Trans Signal Process, 2014, 62(4): 928-938.

[69] Y Shechtman, Y C Eldar, A Szameit, M Segev. *Sparsity based sub-wavelength imaging with partially incoherent light via quadratic compressed sensing*, 2011, Optics Express, 19(16): 14807-14822.

[70] M Stefik. *Inferring DNA structures from segmentation data*, Artificial Intelligence, 1978, 11(1): 85-114.

[71] Y Tan, R Vershynin. *Phase retrieval via randomized Kaczmarz: theoretical guarantees*, arXiv preprint arXiv:1706.09993, 2018.

[72] J R Tropp. *Convex Recovery of a Structured Signal from Independent Random Linear Measurements*, arXiv preprint arXiv: Information Theory, 2014: 67-101.

[73] I Waldspurger, A d'Aspremont, S Mallat. *Phase recovery, maxcut and complex semidefinite programming*, Math Program, 2015, 149(1): 47-81.

[74] A Walther. *The question of phase retrieval in optics*, Journal of Modern Optics, 1963, 10(1): 41-49.

[75] G Wang, G B Giannakis, Y C Eldar. *Solving systems of random quadratic equations via truncated amplitude flow*, IEEE Trans Inform Theory, 2018, 64(2): 773-794.

[76] G Wang, G B Giannakis, Y Saad, J Chen. *Phase retrieval via reweighted amplitude flow*, IEEE Trans Signal Process, 2018, 66(11): 2818-2833.

[77] G Wang, G B Giannakis, J Chen. *Scalable solvers of random quadratic equations via stochastic truncated amplitude flow*, IEEE Trans Signal Process, 2017, 65(8): 1961-1974.

[78] Y Wang, Z Q Xu. *Phase retrieval for sparse signals*, Appl Comput Harmon Anal, 2014, 37(3): 531-544.

[79] G Wang, L Zhang, G B Giannakis, M Akcakaya, J Chen. *Sparse phase retrieval via truncated amplitude flow*, IEEE Trans Signal Process, 2018, 66(2): 479-491.

[80] K Wei. *Solving systems of phaseless equations via Kaczmarz methods: a proof of concept study*, Inverse Problem, 2015, 31(12): 125008-125030.

[81] G Zauner. *Quantendesigns: Grundzüge einer nichtkommutativen Designtheorie*, PhD dissertation, University of Vienna, 1999.

[82] H Zhang, Y Zhou, Y Liang, Y Chi. *Reshaped Wirtinger flow and incremental algorithm for solving quadratic system of equations*, arXiv preprint arXiv:1605.07719, 2016.

[83] H Zhang, Y Chi, Y Liang. *Provable non-convex phase retrieval with outliers: Median truncated Wirtinger flow*, arXiv preprint arXiv:1603.03805, 2016.

[84] G Zheng, R Horstmeyer, C Yang. *Wide-field, high-resolution Fourier pty-chographic microscopy*, Nature Photonics, 2013, 7: 739-745.

School of Mathematical Sciences, Zhejiang University, Hangzhou 310027, China.
Email: lihuiping@zju.edu.cn, songli@zju.edu.cn