Strong Convexity of Affine Phase Retrieval

Meng Huang and Zhiqiang Xu

Abstract—Affine phase retrieval refers to the process of recovering a signal from intensity measurements, with some entries known in advance. In this paper, we demonstrate that a natural least squares formulation for affine phase retrieval is strongly convex on the complete space under certain mild conditions, given that the measurement vectors are complex Gaussian random vectors and that the number of measurements is \( m \geq O(d \log d) \), where \( d \) is the dimension of the signals. Based on the result, we prove that the simple gradient descent method for the affine phase retrieval converges linearly to the target solution with high probability from an arbitrary initial point. These results highlight a fundamental difference between affine phase retrieval and classical phase retrieval, where the least squares formulations for classical phase retrieval are non-convex.

Index Terms—Phase retrieval, strong convexity, linear convergence.

I. INTRODUCTION

A. Problem Setup

The problem of recovering \( x \) from the intensity-only measurements

\[
y_j = |\langle a_j, x \rangle + b_j|^2, \quad j = 1, \ldots, m
\]

is termed as affine phase retrieval. Here, \( x \in \mathbb{C}^d \) is an arbitrary unknown vector, \( a_j \in \mathbb{C}^d, j = 1, \ldots, m \) are known sampling vectors, \( b := (b_1, \ldots, b_m) \in \mathbb{C}^m \) is the bias vector and \( y_j \in \mathbb{R}, j = 1, \ldots, m \) are observed measurements. The affine phase retrieval is of significant importance to a number of fields, such as holography [3], [18], [25], [27], [37] and Fourier phase retrieval problem [4], [5], [30], where a “reference” is situated or a part of signal is a priori known before capturing the intensity-only measurements. It has been shown theoretically that \( m \geq 4d - 1 \) generic measurements are sufficient to recover all the signals \( x \) exactly [16], [21].

A natural approach to recover the signal \( x \) is to solve the following program:

\[
\min_{z \in \mathbb{C}^d} f(z) := \frac{1}{2m} \sum_{j=1}^{m} \left( |\langle a_j, z \rangle + b_j|^2 - y_j \right)^2. \tag{I.1}
\]

If all \( b_j \) are zeros, then the optimization problem in (I.1) simplifies to:

\[
\min_{z \in \mathbb{C}^d} g(z) := \frac{1}{2m} \sum_{j=1}^{m} \left( |\langle a_j, z \rangle|^2 - y_j \right)^2, \tag{I.2}
\]

which is the intensity-based model [8], [31], [38] for solving classical phase retrieval. Due to the non-convex nature of \( g \), gradient based algorithms for solving (I.2) heavily rely on carefully-designed initialization [8], [35], [38] or require \( g \) to have a benign geometric landscape [6], [7], [31]. In this paper, our primary focus lies on the program (I.1), and we are particularly interested in exploring the following questions:

- How does the bias vector \( b \) influence the overall geometric structure of the function \( f \)?
- Is it feasible to efficiently solve the program (I.1) using the straightforward gradient descent method? Is it possible to determine the rate of convergence?

B. Related Work

1) Phase Retrieval: The classical phase retrieval problem aims to recover a signal \( x \in \mathbb{C}^d \) from the intensity-only measurements

\[
y_j = |\langle a_j, x \rangle|^2, \quad j = 1, \ldots, m. \tag{I.3}
\]

It arises in various disciplines and has been investigated recently due to its wide range of practical applications in fields of physical sciences and engineering, such as X-ray crystallography [19], [29], microscopy [28], astronomy [14], optics and acoustics [2], [34] etc, where the detector can record only the diffracted intensity while losing the phase information.

Note that \( |\langle a_j, x \rangle|^2 = |\langle a_j, e^{i\theta}x \rangle|^2 \) for any \( \theta \in \mathbb{R} \). Therefore the recovery of \( x \) is up to a global phase for classical phase retrieval. It was shown that \( m \geq 4d - 4 \) generic measurements suffice to recover \( x \) for the complex case [13], [17], [36] and \( m \geq 2d - 1 \) are sufficient for the real case [2]. From an algorithmic perspective, some efficient gradient descent methods have been proposed to solve the classical phase retrieval problem based on some natural least squares formulations. Due to the
Mathematically, when a known reference signal with a known structure is included in the diffraction pattern, Nobel Prize in Physics in 1971. In holographic optics, a reference signal with a known structure is included in the diffraction pattern alongside the signal of interest [3], [18], [25]. Mathematically, when a known reference \( \mathbf{x}' \in \mathbb{C}^k \) is situated to the object \( \mathbf{x} \in \mathbb{C}^d \), it gives \( \hat{\mathbf{x}} := \begin{pmatrix} \hat{\mathbf{x}}_1 \\ \hat{\mathbf{x}}_2 \end{pmatrix} \in \mathbb{C}^{d+k} \). The intensity measurements that we obtain is

\[
y_j = |\langle \mathbf{a}_j, \hat{\mathbf{x}} \rangle|^2 = |\langle \mathbf{a}_j, \mathbf{x} \rangle + \langle \mathbf{a}_j', \mathbf{x}' \rangle|^2 = |\langle \mathbf{a}_j, \mathbf{x} \rangle + b_j|^2, \quad j = 1, \ldots, m.
\]

Here, \( \hat{\mathbf{a}}_j := \begin{pmatrix} \mathbf{a}_j \\ \mathbf{a}_j' \end{pmatrix} \in \mathbb{C}^{d+k} \) are the sensing vectors and \( b_j := \langle \mathbf{a}_j', \mathbf{x}' \rangle \) are known. The recovery of \( \mathbf{x} \in \mathbb{C}^d \) from the measurements \( y_j \) is dubbed as affine phase retrieval. Unlike the classical phase retrieval where one can only recover \( \mathbf{x} \) up to a unimodular constant, it is possible to recover the true signals \( \mathbf{x} \) exactly for the affine phase retrieval problem. The theoretical foundations of affine phase retrieval have been extensively explored in the seminal work [16], wherein they rigorously establish the injective properties and stability results. Notably, the authors of [16] demonstrate that the minimum number of measurements required for affine phase retrieval is \( m = 3d \), while a sufficient condition for exact recovery of all signals is achieved with \( m \geq 4d - 1 \) generic measurements. In [1], an algorithm is presented for solving the affine phase retrieval problem based on the Amplitude model. The algorithm utilizes an alternating minimization method with spectral initialization. In the case where the measurements \( \mathbf{a}_j \) correspond to the Fourier transform, various recovery algorithms, such as referenced deconvolution method [3] and unrolled network method [23], have been proposed to solve the affine phase retrieval problem for a specific subclass of the bias vector \( \mathbf{b} \), wherein \( b_j := \langle \mathbf{a}_j', \mathbf{x}' \rangle \) is defined with respect to a known reference signal \( \mathbf{x}' \in \mathbb{C}^k \) satisfying \( k \geq d \).

### C. Motivation

In the context of the classical phase retrieval where we aim to recover a signal \( \mathbf{x} \in \mathbb{C}^d \) from the intensity-only measurements

\[
y_j := |\langle \mathbf{a}_j, \mathbf{x} \rangle|^2, \quad j = 1, \ldots, m. \tag{I.4}
\]

Solving this nonlinear system of equations presents a formidable challenge due to its inherently combinatorial nature. A commonly used optimization for recovering \( \mathbf{x} \) from (I.4) is the following intensity-based model [8], [11], [38]:

\[
\min_{\mathbf{z} \in \mathbb{C}^d} g(\mathbf{z}) := \frac{1}{2m} \sum_{j=1}^{m} \left( |\langle \mathbf{a}_j, \mathbf{z} \rangle|^2 - y_j \right)^2. \tag{I.5}
\]

Sun, Qu and Wright study the global geometry structure of the loss function and show \( g \) does not have any spurious local minima under \( m = O(d \log^3 d) \) complex Gaussian random measurements [31]. With this benign geometric landscape in place, the authors of [31] develop a trust-region method to find a global solution of \( g \) with random initialization.

This paper focuses on the problem of affine phase retrieval, which aims to recover the signal \( \mathbf{x} \in \mathbb{C}^d \) from \( y_j = |\langle \mathbf{a}_j, \mathbf{x} \rangle + b_j|^2, \quad j = 1, \ldots, m \). Intuitively, the incorporation of prior information of \( \mathbf{b} \) instills the expectation that the recovery of \( \mathbf{x} \in \mathbb{C}^d \) by solving

\[
\min_{\mathbf{z} \in \mathbb{C}^d} f(\mathbf{z}) := \frac{1}{2m} \sum_{j=1}^{m} \left( |\langle \mathbf{a}_j, \mathbf{z} \rangle + b_j|^2 - y_j \right)^2, \tag{I.6}
\]

becomes more manageable or less challenging. Recall that when all \( b_j \) are zeros, the loss function \( f(\mathbf{z}) \) possesses an infinite number of minimizers of the form \( \mathbf{x}e^{i\theta} \) for any \( \theta \in \mathbb{R} \). However, for \( m \geq O(d) \) generic measurements \( \{\langle \mathbf{a}_j, b_j \rangle\}_{j=1}^{m} \), the target signal \( \mathbf{x} \) is the only minimizer of \( f(\mathbf{z}) \). As a consequence of the continuity of \( f \) with respect to \( \mathbf{b} \), the presence of spurious local minima is expected when the magnitude of \( \mathbf{b} \) is small. Consequently, two intriguing questions arise: Do spurious local minima persist when the magnitude of \( \mathbf{b} \) is large? Can the program (I.6) be efficiently solved by using the simple gradient descent method?

### D. Our Contributions

We analyze the global geometry structure of the loss function \( f \) presented in equation (I.6), and show that under mild conditions on the bias vector \( \mathbf{b} \), and given that \( m \geq O(d \log d) \), the function \( f \) exhibits strong convexity on the entire space. Furthermore, we demonstrate that the simple gradient descent method converges linearly to the global solution, regardless of the initial point chosen. This is summarized below.

**Theorem 1.1 (Informal):** Assume that \( \mathbf{x} \in \mathbb{C}^d \) is a fixed vector. Assume that the vector \( \mathbf{b} \in \mathbb{C}^m \) satisfies \( \|\mathbf{b}\|_2 \geq c_0 \sqrt{m} \|\mathbf{x}\|_2 \), \( \sum_{j=1}^{m} |b_j|^4 \leq c_1 m \|\mathbf{x}\|_4^4 \) and \( \|\mathbf{b}\|_\infty \leq c_2 \sqrt{\log m} \|\mathbf{x}\|_2 \) for some constants \( c_1 \geq c_0 \geq (3/4)^2 \) and \( c_2 > 0 \). Suppose that \( \mathbf{a}_j \in \mathbb{C}^d, j = 1, \ldots, m \), are i.i.d. complex Gaussian random vectors with \( m \geq C d \log d \). Then with high probability the function \( f \) given in (I.6) is strongly convex on the entire space \( \mathbb{C}^d \). Moreover, the Wirtinger flow method with a fixed step size converges linearly to the global solution, from an arbitrary initialization which lies in the complex ball with radius \( R_0 := 2 \left( \frac{1}{m} \sum_{j=1}^{m} y_j - \frac{\|\mathbf{b}\|_2^2}{m} \right)^{1/2} \).

Here, \( C > 0 \) is a constant depending only on \( c_0, c_1 \) and \( c_2 \).

**Remark 1.2:** Our result requires that the bias vector \( \mathbf{b} \in \mathbb{C}^m \) satisfies the conditions \( \|\mathbf{b}\|_2 \geq c_0 \sqrt{m} \|\mathbf{x}\|_2 \), \( \sum_{j=1}^{m} |b_j|^4 \leq c_1 m \|\mathbf{x}\|_4^4 \) and \( \|\mathbf{b}\|_\infty \leq c_2 \sqrt{\log m} \|\mathbf{x}\|_2 \) for some constants.
In fields of the physical sciences such as holography, these conditions are widely accepted and considered to be valid. To illustrate this, in holographic optics, the bias vector $b$ has the form $b_j := (a'_j, x')$, where $x' \in \mathbb{C}^d$ is known a reference and $a'_j \in \mathbb{C}^d$, $j = 1, \ldots, m$, are measurement vectors. Assume that $a'_j \in \mathbb{C}^d$ are generated independently according to the standard Gaussian distribution, if the “energy” of the reference $x'$ and the true signal $x$ are of comparable magnitudes, i.e., $\|x'|_2 \approx \|x\|_2$, then the vector $b$ satisfies those conditions with high probability.

Remark 1.3: Our result asserts that the loss function of affine phase retrieval has excellent geometric landscape, namely, it possesses the strong convexity property over the entire space $\mathbb{C}^d$. As a result, the gradient descent method is likely to succeed from any starting point. However, it is worth noting that the Lipschitz constant of the gradient tends to infinity as $\|x\|_2 \to \infty$. Therefore, solving the program (I.6) by the commonly used gradient descent method, it is advisable to choose a sufficiently small step size. In order to provide a succinct estimate of the necessary step size for convergence, and to establish an explicit rate of convergence, it is necessary for the initial point to be situated within the complex ball of radius $R_0$. It is worth emphasizing that a similar technique has also been employed in the work presented in [31].

E. Notations and Definitions

We say a vector $a \in \mathbb{C}^d$ is a complex Gaussian random vector if $a \sim 1/\sqrt{2} \cdot \mathcal{N}(0, I_d) + i \sqrt{2} \cdot \mathcal{N}(0, I_d)$. We set $S_{\mathbb{C}^d}^{-1} := \{ z \in \mathbb{C}^d : \| z \|_2 = 1 \}$. For a complex number $b$ we use $b_R$ and $b_i$ to denote the real and imaginary part of $b$, respectively. For any $A, B \in \mathbb{R}$, we use $A \leq B$ to denote $A \leq C_0 B$ where $C_0 \in \mathbb{R}_+^*$ is an absolute constant. The notion $\geq$ can be defined similarly. We use the notations $\| \cdot \|_2$ and $\| \cdot \|_*$ to denote the operator norm and nuclear norm of a matrix, respectively.

For any real-valued function $f(z) : \mathbb{C}^d \to \mathbb{R}$, if $f$ is differentiable as a function of $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$, where $x := z_R, y := z_I$, then the Wirtinger gradient and the Hessian matrix are well-defined. We refer the reader to papers [8], [31] for a more detailed account of Wirtinger calculus.

Definition 1.4: A real-valued function $f : \mathbb{C}^d \to \mathbb{R}$ is called strongly convex with a constant $c_0 > 0$ if

$$
\begin{align*}
\left( \begin{array}{c}
v \\
v \end{array} \right)^* \nabla^2 f(z) \left( \begin{array}{c}
v \\
v \end{array} \right) & \geq c_0 \| v \|_2^2, & \text{for all } z, v \in \mathbb{C}^d.
\end{align*}
$$

Here, $\nabla^2 f(z)$ is the Hessian matrix of $f$ in Wirtinger calculus.

Remark 1.5: For a differentiable function $h : \mathbb{R}^d \to \mathbb{R}$, the standard definition of strongly convex with constant $c_0 > 0$ is

$$
\begin{align*}
h(u) & \geq h(w) + \nabla h(w)^T (u - w) + \frac{c_0}{2} \| u - w \|_2^2 \quad (1.7)
\end{align*}
$$

for all $u, w \in \mathbb{R}^d$.

Here, the $\nabla h$ is the standard gradient of the function $h$. In fact, after performing some tedious calculations, we can demonstrate that Definition 1.4 is equivalent to equation (1.7).

For the loss function $f$ defined in (1.1), direct calculation gives the Wirtinger gradient

$$
\nabla f(z) = \frac{1}{m} \sum_{j=1}^m \left[ (a_j^* z + b_j)^2 - |y_j|^2 \right] \left( \begin{array}{c}
a_j^* \bar{z} + b_j \end{array} \right) \bar{a}_j \quad (1.8)
$$

and the Hessian matrix

$$
\nabla^2 f(z) = \frac{1}{m} \sum_{j=1}^m \left[ \begin{array}{cc}
|a_j|^2 z + b_j^2 & a_j^* a_j \\
a_j^* a_j & 2a_j^* z + b_j \end{array} \right] \quad (1.9)
$$

F. Organization

The paper is organized as follows. In Section II, we demonstrate that the loss function (1.1) is strongly convex. Building upon this characterization, in Section III, we establish that the Wirtinger flow algorithm is linearly convergent from any initial point. In Section IV, we present a series of numerical experiments to investigate the empirical performance of our algorithm. In Section V, we provide a brief discussion of future research directions. Appendix contains the technical lemmas required in our analysis.

II. THE STRONG CONVEXITY OF THE OBJECTIVE FUNCTION

In this section we demonstrate that the loss function $f$ given in (1.1) is strongly convex on the entire space. The intuition is as follows: Given that $a_j$ are complex Gaussian random vectors, it becomes straightforward to verify that the Hessian matrix is positive definite in expectation, under some mild conditions on $b$. To derive non-asymptotic results, we truncate the terms involving third and fourth powers of Gaussian random variables into two parts. The first part is well-behaved, while the second part is heavy-tailed. However, we are able to bound the heavy-tailed part by utilizing another nonnegative term, where its deviation below the expectation is bounded. By exploiting this technique, we can prove the subsequent result:

Theorem II.1: Assume that $x \in \mathbb{C}^d$ is an arbitrary fixed vector and the vector $b \in \mathbb{C}^m$ satisfies $\| b \|_2 \geq c_0 \sqrt{m} \| x \|_2$, $\sum_{j=1}^m |b_j|^4 \leq c_1 m \| x \|_2^4$ and $\| b \|_\infty \leq c_2 \sqrt{\log m} \| x \|_2$ for some constants $c_2 > 0$ and $c_1 \geq c_0 \geq (110/49)^2 \approx 5.04$. Suppose that $a_j \in \mathbb{C}^d, j = 1, \ldots, m$, are i.i.d. complex Gaussian random vectors and $y_j = |a_j, x| + b_j^2$, $j = 1, \ldots, m$. If $m \geq C d \log d$ then with probability at least $1 - c_m m^{-1} - 18 \exp(-c_d d) - c_e \exp(-c_r m/\log m)$ the Hessian matrix of $f$ given in (1.9) obeys

$$
\left( \begin{array}{c}
v \\
v \end{array} \right)^* \nabla^2 f(z) \left( \begin{array}{c}
v \\
v \end{array} \right) \geq \left( 1.9c_0^2 - 4.4 \right) \| x \|_2^2
$$

for all $v \in S_{\mathbb{C}^d}^{-1}, z \in \mathbb{C}^d$. Here, $C, c_1, c_2, c_3$ and $c_4$ are positive constants depending only on $c_0, c_1$ and $c_2$.

Proof: Without loss of generality, we assume that $\| x \|_2 = 1$. For any unit vector $v \in \mathbb{C}^d$, let $u := \left( \begin{array}{c}
v \\
v \end{array} \right) \in \mathbb{C}^{2d}$.
It then follows from (1.9) that

\[ u^* \nabla^2 f(z)u = \frac{2}{m} \sum_{j=1}^{m} (2|a_j^*z + bj|^2 - |a_j^*x + bj|^2) |a_j^*v|^2 \]

\[ + \frac{2}{m} \sum_{j=1}^{m} ((a_j^*z + bj)^2(v^*a_j)^2) \]

\[ = \frac{2}{m} \sum_{j=1}^{m} |a_j^*z + bj|^2 |a_j^*v|^2 - \frac{2}{m} \sum_{j=1}^{m} |a_j^*x + bj|^2 |a_j^*v|^2 \]

\[ + \frac{4}{m} \sum_{j=1}^{m} ((a_j^*z + bj)(v^*a_j))^2 \]

\[ = \frac{2}{m} \sum_{j=1}^{m} ( |a_j^*z|^2 |a_j^*v|^2 - |a_j^*x|^2 |a_j^*v|^2 + 2(v^*a_j^*a_j^*z)^2 \]

\[ + \frac{4}{m} \sum_{j=1}^{m} ((b_j^*a_j^*z)|a_j^*v|^2 - (b_j^*a_j^*x)|a_j^*v|^2 \]

\[ + \frac{4}{m} \sum_{j=1}^{m} ((b_j^*(a_j^*v)|a_j^*v|^2) \]

\[ \cdot (v^*a_j^*a_j^*z) \]  

(II.1)

Here, the second equality follows from the fact that \(|z|^2 + (z^2) = 2(z^2)|^2\) for any \(z \in \mathbb{C}\). For convenience, we set

\[ A(z, v) := \frac{1}{m} \sum_{j=1}^{m} |a_j^*z|^2 |a_j^*v|^2. \]

We claim that if \(m \geq c(e)d \log d\) then with probability at least \(1 - c_1 \exp(-c_2(m/\log m) - c_3(m^{-1} - 18 \exp(-c_4d))\), it holds that

\[ \frac{1}{2} u^* \nabla^2 f(z)u \geq A(z, v) - 6\epsilon \left( \frac{\|b\|_2^2}{\sqrt{m}} + 1 \right) \sqrt{A(z, v)} \]

\[ + (1 - \epsilon) \cdot \frac{\|b\|_2^2}{m} - |x^*v|^2 - 6\epsilon \left( \frac{\|b\|_2^2}{\sqrt{m}} + 1 \right) \|z\|^2 \]

\[ - 2\epsilon \left( \frac{\|b\|_2^2}{\sqrt{m}} + 1 \right) - 2\epsilon - (1 + \epsilon). \]  

(II.2)

Here, \(\epsilon\) is any constant in \((0, 1)\), \(c_1, c_2, c_4\) are positive universal constants, \(c_3(\epsilon)\) and \(c_3(\epsilon)\) are positive constants depending only on \(\epsilon\). Set

\[ \mathcal{R} := \{ (z, v) \in \mathbb{C}^d \times \mathbb{S}^{d-1} : A(z, v) \geq 1 \} . \]

To give a lower bound for \(u^* \nabla^2 f(z)u\), we divide the space \(\mathbb{C}^d \times \mathbb{S}^{d-1}\) into two regimes: \((z, v) \in \mathcal{R}\) and \((z, v) \notin \mathcal{R}\).

**Regime 1:** If \((z, v) \in \mathcal{R}\) then we have

\[ (A(z, v))^2 \leq A(z, v) . \]  

(III.3)

By Lemma A.2, we obtain that when \(m \geq c(e)d \log d\), with probability at least \(1 - c' \exp(-c''(\epsilon)m)\), it holds

\[ A(z, v) \geq (1 - \epsilon) \left( \|z\|^2 + |z^*v|^2 \right) \]

for all \(z \in \mathbb{C}^d\), \(v \in \mathbb{S}^{d-1} \).  

(II.4)

Note that \(\|b\|_2 \geq c_0 \sqrt{m}\) and \(\|b\|_2 \leq \sqrt{\frac{1}{m} \sum_{j=1}^{m} |b_j|^4} \leq C_1 \sqrt{m}\) for a universal constant \(C_1 \geq 1\). Putting (II.3) and (II.4) into (II.2) and taking \(\epsilon := \frac{1}{60(c_1 + 1)}\), we obtain that, when \(m \geq c(e)d \log d\), with probability at least \(1 - c_9 m^{-1} - c_6 \exp(-c_8 m/\log m) - 18 \exp(-c_6 d)\), it holds that

\[ \frac{1}{2} u^* \nabla^2 f(z)u \]

\[ \geq \left( 1 - 6\epsilon \left( \frac{\|b\|_2^2}{\sqrt{m}} + 1 \right) \right) \left( 1 - \epsilon \right) \left( \|z\|^2 + |z^*v|^2 \right) \]

\[ + 5 \left( 1 - \epsilon \right) \cdot \frac{\|b\|_2^2}{m} - \left( 1 + \epsilon \right) \cdot \|x^*v|^2 - 2\epsilon \]

\[ - 2\epsilon \left( \frac{\|b\|_2^2}{\sqrt{m}} + 1 \right) - 6\epsilon \left( \frac{\|b\|_2^2}{\sqrt{m}} + 1 \right) \|z\|^2 \]

\[ \geq 0.88 \|z\|^2 + 0.1\|z\|^2 + 0.98 \cdot \frac{\|b\|_2^2}{m} - 2.1 \]

\[ \geq 0.98 c_0^2 - 2.2, \]

where the last inequality follows from the fact that \(0.88 \epsilon^2 - 0.1\epsilon + 0.1 > 0\) for any \(\epsilon \geq 0\). Here, \(c_9, c_6, c_8\) and \(c_4\) are positive universal constants.

**Regime 2:** If \((z, v) \notin \mathcal{R}\) then we have

\[ A(z, v) < 1. \]

Similarly, taking \(\epsilon := \frac{1}{60(c_1 + 1)}\) in (II.2), we obtain that

\[ \frac{1}{2} u^* \nabla^2 f(z)u \]

\[ \geq A(z, v) - 6\epsilon \left( \frac{\|b\|_2^2}{\sqrt{m}} + 1 \right) \left( 1 - \epsilon \right) \cdot \frac{\|b\|_2^2}{m} - \left( 1 + \epsilon \right) \cdot \|x^*v|^2 - 2\epsilon \]

\[ - 2\epsilon \left( \frac{\|b\|_2^2}{\sqrt{m}} + 1 \right) - 6\epsilon \left( \frac{\|b\|_2^2}{\sqrt{m}} + 1 \right) \|z\|^2 \]

\[ \geq 0.98 c_0^2 - 2.2. \]

Here, the second inequality follows from (II.4) and the fact that \(|x^*v| \leq 1\). Combining the results above, we arrive at the conclusion.

It remain to prove (II.2). According to Lemma A.1, when \(m \geq c(e)d \log d\), with probability at least \(1 - c_5 \epsilon^{-2} m^{-1} - c_9 \exp(-c_7 \epsilon^2 m/\log m)\), it holds that

\[ \frac{1}{m} \sum_{j=1}^{m} (a_j^*x|^2|a_j^*v|^2 - v^* \left( \frac{1}{m} \sum_{j=1}^{m} a_j^*x|^2|a_j^*v|^2 \right) \]

\[ \leq 1 + \epsilon + |x^*v|^2 \]

for all \(v \in \mathbb{S}^{d-1}\), where we use the fact that \(\|z\|_2 = 1\) in the inequality. Here, \(\epsilon\) is any constant in \((0, 1)\), \(c_5, c_6\) and \(c_7\) are positive universal constants. From Lemma A.2, we obtain that the following holds with probability at least \(1 - c' \exp(-c''(\epsilon)m)\)

\[ \frac{1}{m} \sum_{j=1}^{m} (v^* a_j^* a_j^*z)^2 \geq 1 - \frac{\epsilon}{2} \left( \|z\|^2 + 3(z^*v)_{\mathbb{R}} - (z^*v)_3 \right) \]

\[ \geq 1 - \frac{\epsilon}{2} \left( \|z\|^2 - |z^*v|^2 \right) \]

for all \(z \in \mathbb{C}^d\), \(v \in \mathbb{S}^{d-1}\).
provided \( m \geq c(e) d \log d \). Here, \( c', c'' \) are positive universal constants and \( c''(e) \) is a constant depending only on \( e \). Recall that \( \| x \|_2 = 1 \), \( \sum_{j=1}^{m} |b_j|^4 \leq m \) and \( \| b \|_\infty \leq \sqrt{\log m} \). It follows from Lemma A.6 that, for \( m \geq c(e) d \log m \), with probability at least \( 1 - 6 \exp(-c_1 d) - 6 \exp(-c''(e) m / \log m) \), it holds

\[
\left| \frac{1}{m} \sum_{j=1}^{m} \langle \overline{b}_j(a_j^* z) \rangle_{\mathbb{R}} |a_j^* v|^2 \right| \leq \epsilon \left( \frac{\| b \|_2}{\sqrt{m}} + 1 \right) \left( \| z \|_2 + (A(z,v))^\frac{3}{2} \right)
\]

for all \( z \in \mathbb{C}^d \), \( v \in \mathbb{S}^{d-1}_C \). Here, \( c_4 > 0 \) is a universal constant. Applying Lemma A.8, the following holds with probability at least \( 1 - 6 \exp(-c_7 \epsilon^2 m / \log m) - 6 \exp(-c_3 d) \):

\[
\frac{1}{m} \sum_{j=1}^{m} \left( \overline{b}_j(a_j^* v)^2 \right)_{\mathbb{R}} \frac{1 - \epsilon}{2} \epsilon \left( \frac{\| b \|_2}{\sqrt{m}} - \epsilon \right)
\]

for all \( v \in \mathbb{S}^{d-1}_C \), provided \( m \geq C' \epsilon^{-2} \log(1/\epsilon) d \log m \), where \( C' > 0 \) is a universal constant. Recognize that \( \| b \|_2 \leq \sqrt{m} \). It can be deduced from Lemma A.7 that when \( m \geq C' \epsilon^{-2} d \log m \), the following holds with probability at least \( 1 - c_5 \epsilon^{-2} m^{-1} - 2 \exp(-c_7 \epsilon^{-2} m / \log m) \):

\[
\left| \frac{1}{m} \sum_{j=1}^{m} \langle \overline{b}_j(a_j^* x) \rangle_{\mathbb{R}} |a_j^* v|^2 \right| \leq \epsilon \left( \frac{\| b \|_2}{\sqrt{m}} + 1 \right)
\]

for all \( v \in \mathbb{S}^{d-1}_C \). Substituting the results above into (II.1), we obtain (II.2). 

### III. Optimization by Wirtinger Gradient Descent

Due to the strong convexity of \( f \), we can solve program (I.1) using the vanilla Wirtinger gradient descent algorithm, which is given by:

\[
z_{k+1} = z_k - \mu \nabla f(z_k).
\]

(III.1)

Here, with abuse of notation, we set

\[
\nabla f(z) := \frac{1}{m} \sum_{j=1}^{m} \left( (a_j^* z + b_j)^2 - y_j \right) (a_j^* z + b_j) a_j \in \mathbb{C}^d.
\]

(III.2)

This is because the second part of equation (I.8) is the conjugate of the first, and thus, only the first \( d \) entries are kept in equation (II.2).

Note that the Lipschitz constant of the gradient tends to infinity as \( \| x \|_2 \to \infty \). Therefore, when solving the program (I.1) by the update rule (III.1), it is essential to select a sufficiently small step size \( \mu \). For the sake of analysis, we will assume the initialization \( x_0 \) is an arbitrary point over \( \mathbb{B}_C^d(R_0) \) for some radius \( R_0 \). Therefore, an upper bound for \( \| x \|_2 \) becomes necessary, as presented below.

**Lemma III.1**: Assume that \( x \in \mathbb{C}^d \) is an arbitrary fixed vector and \( b = (b_1, \ldots, b_m)^T \in \mathbb{C}^m \) is a vector satisfying \( \| b \|_2 \leq \sqrt{m} \). Suppose \( y_j = a_j^* x + b_j^2, j = 1, \ldots, m \) where \( a_j \in \mathbb{C}^d \) are i.i.d complex Gaussian random vectors. Then, with probability at least \( 1 - 4 \exp(-cm) \), the following holds

\[
R_0 / 3 \leq \| x \|_2 \leq R_0.
\]

Here, \( R_0 := \sqrt{2\left( \frac{1}{m} \sum_{j=1}^{m} y_j - \| b \|_2^2 / m \right)^{1/2}} / \epsilon > 0 \) is a universal constant.

**Proof**: See Appendix.

Based on Lemma III.1, we have the following algorithm:

Next, we will prove that Algorithm 1 converges linearly to the target solution \( x \). To this end, we need to provide the Lipschitz constant of the gradient \( \nabla f(z) \), as shown below.

**Lemma III.2 (Local Smoothness Property)**: Suppose that \( a_j \in \mathbb{C}^d, j = 1, \ldots, m \), are i.i.d complex Gaussian random vectors. Let \( S_R := \{ z \in \mathbb{C}^d : \| z \|_2 \leq R \} \) be a bounded region where \( R \) is any positive constant. If \( m \geq C d \) then with probability at least \( 1 - 4 \exp(-cm) - c_8 m^{-d} \), the gradient \( \nabla f(z) \) given in (I.8) is Lipschitz continuous over \( S_R \), i.e.,

\[
\| \nabla f(z) - \nabla f(z') \|_2 \leq C_R \| z - z' \|_2 \quad \text{for all} \quad z, z' \in S_R,
\]

where

\[
C_R = 6 \sqrt{2} \left( 2 Rd \log m + \| b \|_\infty \sqrt{d \log m} \right) \left( R + \frac{\| b \|_2}{\sqrt{m}} \right) + 8 \sqrt{2} \left( 2 d \log m (R^2 + \| z \|_2^2 + \| b \|_\infty^2) \right).
\]

Here, \( C, c \) and \( c_8 \) are positive universal constants.

**Proof**: See Appendix.

Based on strongly convex and local smoothness properties as stated in Theorem II.1 and Lemma III.2 respectively, we are ready to present the convergence property of Algorithm 1.

**Theorem III.3**: Assume that \( x \in \mathbb{C}^d \) is an arbitrary fixed vector. Assume that the vector \( b \in \mathbb{C}^m \) satisfies \( \| b \|_2 \geq c_0 \sqrt{m} \| x \|_2 \), \( \sum_{j=1}^{m} \| b_j \|^4 \leq c_0 m \| x \|_2^4 \) and \( \| b \|_\infty \leq c_2 \sqrt{\log m} \| x \|_2 \) for some constants \( c_2 > 0 \) and \( c_1 \geq c_6 \geq (4.4/1.96)^2 \). Suppose that \( a_j \in \mathbb{C}^d, j = 1, \ldots, m \),
are i.i.d complex Gaussian random vectors and $y_j = [\langle a_j, x \rangle + b_j]$, $j = 1, \ldots, m$. If $m \geq Cd \log d$ then with probability at least $1 - c_d m^{-1} - c_b \exp(-c_d m/\log m) - 26 \exp(-c_d d)$, the iteration $z_k$ given by Algorithm 1 with a fixed step size $\mu \leq c_3/(d \log m \|x\|_2^2)$ obeys

$$\|z_k - x\|_2^2 \leq 16(1 - \rho)^k \|x\|_2^2,$$

where $\rho = \mu(1.96c_0^2 - 4.4) < 1$. Here, $C, c_0, c_b, c_c, c_d, c_3$ are positive constants depending only on $c_0, c_1, c_2, c_3$.

Proof: Set $\mathcal{R} := \{z \in \mathbb{C}^d : \|z\|_2 \leq 5\|x\|_2\}$. Lemma III.2 implies that, with probability at least $1 - 4 \exp(-cm) - c_4 m^{-d}$, the Wirtinger gradient $\nabla f(z)$ given in (I.8) is Lipschitz continuous over $\mathcal{R}$, namely,

$$\|\nabla f(z) - \nabla f(z')\|_2 \leq L_{\mathcal{R}} \|z - z'\|_2 \quad \forall z, z' \in \mathcal{R}, \quad (III.3)$$

where $L_{\mathcal{R}} := C_1 d \log m \|x\|_2^2$ for a constant $C_1 > 0$ depending only on $c_1$ and $c_2$. Here, we use $|b|_2 \leq \sqrt{\sqrt{m} \sum_{j=1}^m |b_j|^4} \leq \sqrt{c_1 \sqrt{m}} \|x\|_2$ which follows from $\sum_{j=1}^m |b_j|^4 \leq c_1 m \|x\|_2^2$ and the Cauchy-Schwarz inequality. We next claim that, with probability at least $1 - 4 \exp(-cm) - c_4 m^{-d}$, it holds that

$$lf(z') \leq f(z) + 2(\nabla f(z), z' - z)_{\mathbb{R}} + \frac{\sqrt{2} L_{\mathcal{R}}}{2} \|z' - z\|_2^2$$

for all $z', z \in \mathcal{R}$. \quad (III.4)

Here, $\nabla f(z)$ is given in (III.2). Theorem II.1 implies that the loss function $f(z)$ is strongly convex with probability at least $1 - c_0 m^{-1} - c_b \exp(-c_d m/\log m) - 18 \exp(-c_d d)$. Hence, we have

$$f(x) \geq f(z) + 2(\nabla f(z), x - z)_{\mathbb{R}} + \frac{\beta}{2} \|z - x\|_2^2$$

for all $z \in \mathbb{C}^d$, \quad (III.5)

where $\beta := 1.96c_0^2 - 4.4$ (see Remark I.5 for detail).

Based on (III.4) and (III.5), we can prove the conclusion recursively. Indeed, since the initial point $z_0 \in \mathbb{B}_{R_0}^d(R_0)$, we have $\|z_0\|_2 \leq R_0$. According to Lemma III.1, we obtain that, with probability at least $1 - 4 \exp(-cm)$, it holds that $\|x\|_2 \leq R_0 \leq 3\|x\|_2$, which implies $z_0, x \in \mathcal{R}$. Next, if $z_k \in \mathcal{R}$ then

$$\|z_{k+1} - x\|_2^2 \leq \|z_k - x - \mu \nabla f(z_k)\|_2^2$$

$$\leq (1 - \frac{\beta}{2}) \|z_k - x\|_2^2 - \mu(f(z_k) - f(x)) + \mu^2 \|\nabla f(z_k)\|_2^2$$

$$\leq (1 - \frac{\beta}{2}) \|z_k - x\|_2^2 - \mu(f(z_k) - f(x)) + \mu \|\nabla f(z_k)\|_2^2$$

$$\leq (1 - \frac{\beta}{2}) \|z_k - x\|_2^2 - \frac{2 \mu^2 L_{\mathcal{R}}}{4 - \sqrt{2}} (f(z_k) - f(x))$$

$$\leq (1 - \frac{\beta}{2}) \|z_k - x\|_2^2 - \frac{2 \mu^2 L_{\mathcal{R}}}{4 - \sqrt{2}} (f(z_k) - f(x))$$

$$\leq (1 - \frac{\beta}{2}) \|z_k - x\|_2^2,$$

where $\beta := 1.96c_0^2 - 4.4$. The proof is complete.

IV. NUMERICAL SIMULATIONS

In this section, we demonstrate experimentally that the loss function $f$ given in (I.1) is well structured even when the number of measurements $m = O(d)$. A. Recovery of 1D Signals

In our numerical experiments, the target vector $x \in \mathbb{C}^d$ and the measurement vectors $a_j, j = 1, \ldots, m$ are generated independently according to the standard complex Gaussian distribution or coded diffraction patterns (CDP) model. For the CDP model, we use masks of octonary patterns as in [8]. The bias vector $b \sim \mathcal{N}(0, \lambda \|x\|_2^2 a_j)$, where $\lambda = 0, 1, \ldots, 4$. At iteration $k$, the step size is $\mu(k) = \min(1 - \exp(-k/330), 0.2 \exp(-0.75k))$. We shall emphasize that, when $\lambda = 0$, our algorithm reduces to the Wirtinger Flow method [8] with random initialization for solving the classical phase retrieval problem.

Example IV.1: In this example, we investigate the empirical success rate of Algorithm 1 versus the number of measurements. We set $d = 1000$. For the complex Gaussian case, we set the maximum number of iterations $T = 2500$, and vary provided the step size $\mu \leq (4 - \sqrt{2})/(2L_{\mathcal{R}})$, where the first inequality follows from (III.5) and the second inequality follows from the fact of

$$f(z_k) - f(x) \geq \frac{4 - \sqrt{2}}{2L_{\mathcal{R}}} \|\nabla f(z_k)\|_2^2.$$
Fig. 1. The empirical success rate for different \( m/d \) based on 100 random trails. (a) Success rate for complex Gaussian case, (b) success rate for CDP case.

\( m \) within the range \([3d, 8d]\). For the CDP case, we set the maximum number of iterations \( T = 3000 \) and set the number of masks \( m/d = L \) from 2 to 8. The initial point \( z_0 \) is chosen uniformly from the complex ball \( \mathbb{B}_C^d(R_0) \), where \( R_0 \) is defined in Lemma III.1. For each \( m \), we run 100 times trials to calculate the success rate. Here, we define a trial as successful if the algorithm can produce a vector \( z_T \) satisfying \( \|z_T - x\|/\|x\| \leq 10^{-5} \) for \( \lambda \neq 0 \). The results are plotted in Fig. 1. We observe that the recovery performance improves with an increase in the value of \( \lambda \). Notably, Algorithm 1 achieves a 100% success rate when the number of measurements \( m \geq 4.5d \) and \( \lambda = 1, \ldots, 4 \). This suggests that having at least \( O(d) \) samples may be sufficient to ensure that the strong convexity property holds.

**Example IV.2:** In this example, we test the convergence rate of Algorithm 1 for the complex Gaussian case. We set the values of \( d \) and \( m \) to be 1000 and 5\( d \) respectively. To demonstrate its robustness, we consider both the noiseless and the noisy data models. In the noisy data model, we set \( y_j = |\langle a_j, x \rangle + b_j|^2 + \eta_j \), where the noise \( \eta_j \sim \mathcal{N}(0, 0.01^2) \). The results are presented in Fig. 2, which confirms the linear convergence of Algorithm 1.

**B. Recovery of Natural Image**

We next test the performance of Algorithm 1 on recovering a natural image from masked Fourier intensity measurements. The image is the Milky Way Galaxy with resolution 1080 × 1920. The colored image has RGB channels. We use \( L = 20 \) random octanary patterns to obtain the Fourier intensity measurements for each R/G/B channel as in [8]. At iteration \( k \), the step size is \( \mu(k) = \min(1 - \exp(-k/330), 0.4 \exp(-0.75\lambda)) \). Table I lists the averaged time elapsed and the number of iterations needed to achieve the relative error \( 10^{-5} \) and \( 10^{-10} \) over the three RGB channels. We can see that Algorithm 1 with \( \lambda = 1, \ldots, 3 \) outperform WF method in both iterations and the computational time cost.

**V. DISCUSSION**

In this paper, we provide the characterization of a natural least squares formulation (I.1) for affine phase retrieval. Particularly, we show the loss function \( f \) given in (I.1) is strongly convex on the entire space \( \mathbb{C}^d \). This benign geometric structure allows the simple gradient descent algorithm to reconstruct the target signals with a linear convergence rate.
TABLE I
TIME ELAPSED AND NUMBER OF ITERATIONS AMONG ALGORITHMS ON RECOVERY OF GALAXY IMAGE

| Algorithm | The Milky Way Galaxy | 10⁻² | 10⁻¹₀ |
|-----------|----------------------|------|-------|
|           | # Iters | Time(s) | # Iters | Time(s) |
| WF        | 212     | 183.5   | 286   | 231.6   |
| λ = 1     | 77      | 69.6    | 162   | 132.9   |
| λ = 2     | 81      | 72.2    | 171   | 159.1   |
| λ = 3     | 83      | 76.6    | 184   | 171.8   |

There are some interesting problems for future research. Firstly, Theorem II.1 establishes that \( m \geq d \log d \) samples are required for guaranteeing strong convexity. However, based on our numerical experiments, we conjecture that \( m \geq d \) samples might be sufficient to ensure this property. It would be fascinating to investigate whether this gap can be closed. Secondly, in this paper, we mainly focus on analyzing the global strong convexity property of intensity-based empirical loss. An alternative is to consider the amplitude-based loss. Due to the presence of the bias vector \( b \) and the non-differentiability of the amplitude-based loss, analyzing the global strong convexity property becomes a challenging task. Thirdly, our current analysis is limited to Gaussian random vectors as measurement vectors. Some numerical evidence have shown that affine phase retrieval from CDP settings or from masked Fourier intensity measurements works well if we start from a random initialization. Therefore, it would be of practical interest to extend the result in this paper to these types of measurement vectors. Exploring the behavior of the function \( f \) under different measurement vector distributions could provide valuable insights and broaden the applicability of the results.

APPENDIX A
PRELIMINARIES AND SUPPORTING LEMMAS

Lemma A.1 [31, Lemma 21]: Let \( a_j \in \mathbb{C}^d, j = 1, \ldots, m \), be i.i.d complex Gaussian random vectors. Suppose that \( x \in \mathbb{C}^d \) is a fixed vector. For any \( \epsilon \in (0, 1) \) the following holds with probability at least \( 1 - c_\epsilon \epsilon^{-2} m^{-1} - c_\epsilon \exp(-c_\epsilon \epsilon^2 m/ \log m) \):

\[
\left\| \frac{1}{m} \sum_{j=1}^{m} a_j^* x^2 a_j^* - (x^* x^2 I) \right\| \leq \epsilon \|x\|^2
\]

provided \( m \geq C(\epsilon) d \log d \). Here \( C(\epsilon) \) is a constant depending on \( \epsilon, c_\epsilon, c_\epsilon \) and \( c_\epsilon \) are positive absolute constants.

Lemma A.2 [31, Lemma 22]: Let \( a_j \in \mathbb{C}^d, j = 1, \ldots, m \), be i.i.d complex Gaussian random vectors. For any \( \epsilon \in (0, 1) \) it hold with probability at least \( 1 - c_\epsilon \epsilon^{-d} - c_\epsilon \exp(-c_\epsilon (\epsilon m)) \):

\[
\frac{1}{m} \sum_{j=1}^{m} a_j^* z^2 a_j^* v^2 \geq (1 - \epsilon) \left( \|v\|^2 \|z\|^2 + |v^* z|^2 \right)
\]

\[
\frac{1}{m} \sum_{j=1}^{m} \left( (a_j^* z)(v^* a_j) \right)^2 \geq (1 - \epsilon) \left( \frac{1}{2} \|z\|^2 \|v\|^2 + \frac{1}{4} (\Re(z^* v))^2 - \frac{1}{2} (\Im(z^* v))^2 \right)
\]

for all \( z, v \in \mathbb{C}^d \), provided \( m \geq C(\epsilon) d \log d \). Here \( C(\epsilon) \) and \( c(\epsilon) \) are constants depending on \( \epsilon \) and \( c, c' \) are positive absolute constants.

Lemma A.3 [22, Lemma 3.3]: Let \( \eta \in \mathbb{R}^m \) be a fixed vector. Suppose that \( a_j \in \mathbb{C}^d, j = 1, \ldots, m \), are i.i.d complex Gaussian random vectors. Then with probability at least \( 1 - 2 \exp(-c_0 d) \) it holds

\[
\left\| \sum_{j=1}^{m} n_j (a_j a_j^* - I) \right\|_2 \leq C \left( \sqrt{d} \|\eta\|_2 + d \|\eta\|_\infty \right).
\]

Here, \( C, c_0 \geq 0 \) are universal constants.

Lemma A.4: Assume that \( a_j \in \mathbb{C}^d, j = 1, \ldots, m \), are i.i.d complex Gaussian random vectors. If \( m \geq O(d) \) then with probability at least \( 1 - 3 \exp(-c m) \) it holds

\[
\frac{1}{m} \sum_{j=1}^{m} |a_j^* z + b_j|^2 - |a_j^* w + b_j|^2 
\leq \left( \frac{1}{2} \left( \|z\|_2 + \|w\|_2 + 2\|b\|_m \right) \right) \|z - w\|_2
\]

for all \( z, w \in \mathbb{C}^d \) and \( b = (b_1, \ldots, b_m)^T \in \mathbb{C}^m \). Here, \( c \) is a positive absolute constants.

Proof: A simple calculation shows that

\[
\frac{1}{m} \sum_{j=1}^{m} |a_j^* z + b_j|^2 - |a_j^* w + b_j|^2 
\leq \left( \frac{1}{2} \left( \|z\|_2 + \|w\|_2 + 2\|b\|_m \right) \right) \|z - w\|_2
\]

for all \( z, w \in \mathbb{C}^d \). Here, \( \| \| \) denotes the nuclear norm.

For the second term, we have

\[
\frac{1}{m} \sum_{j=1}^{m} |a_j^* (z - w)|^2 \leq \frac{3}{2} \|z^* - w^*\|_2
\]

\[
\leq \frac{3}{2} \|z^* - w^*\|_2
\]

for all \( z, w \in \mathbb{C}^d \). Here, \( \| \| \) denotes the nuclear norm.

In order to establish the global strong convexity of the loss function, not just the local strong convexity around the true signal \( x \) as demonstrated in references employing spectral initialization, we need to establish an upper bound for the quantity \( \frac{1}{m} \sum_{j=1}^{m} (b_j (a_j^* z))_{\mathbb{R}} |a_j^* v|^2 \) for all \( z \in \mathbb{C}^d, v \in \mathbb{S}^{d-1} \). It is
worth noting that $E \left( \tilde{b}_j(a_j^* z) \right)^2 | a_j^* v|^2 = 0$. For the purpose of the proof, we aim to show that the upper bound should be close to zero. However, due to the involvement of third powers of Gaussian random variables, the heavy-tailed behavior prevents it from uniformly concentrating around its expectation with only $d \log d$ samples. Establishing a suitable uniform upper bound for the quantity, which involves the third and fourth powers of Gaussian random variables, while simultaneously avoiding excessive increase in sampling complexity, poses a significant challenge in the field of statistics. To address this issue, we divide this quantity into the following three parts:

$$
\left| \frac{1}{m} \sum_{j=1}^{m} \left( \tilde{b}_j(a_j^* z) \right) R | a_j^* v|^2 \right|
\leq \left| \frac{1}{m} \sum_{j=1}^{m} b_j R(a_j^* z) |a_j^* v|^2 \phi\left(\frac{|a_j^* v|}{\beta}\right) \right|
+ \left| \frac{1}{m} \sum_{j=1}^{m} b_j R(a_j^* z) |a_j^* v|^2 \phi\left(\frac{|a_j^* v|}{\beta}\right) \right|
+ \frac{1}{m} \sum_{j=1}^{m} |b_j||a_j^* v||a_j^* v|^2 \{ |a_j^* v| \geq \beta \},
$$

where the first and second parts are well-behaved, while the third part still exhibits heavy-tailed behavior. Let $J_{\beta} = \{ j \in [m] : |a_j^* v| \geq \beta \}$. Intuitively, if we take $\beta$ to be a large constant, then with high probability $|J_{\beta}|/m \leq c_0 m$ holds uniformly over $v \in S_{d-1}^+$, where $c_0$ is a small positive constant. Roughly speaking, the third term can be upper bounded as follows:

$$
\frac{1}{m} \sum_{j=1}^{m} |b_j||a_j^* z||a_j^* v|^2 \{ |a_j^* v| \geq \beta \} \leq c_1 \left( \frac{1}{m} \sum_{j=1}^{m} |a_j^* z|^2 |a_j^* v|^2 \right)
$$

for a sufficient small constant $c_1 > 0$. Although we cannot provide a uniform upper bound for the term $\frac{1}{m} \sum_{j=1}^{m} |a_j^* z|^2 |a_j^* v|^2$, the influence can be well controlled under sufficiently small $c_1$, and it is enough for our proof. In what follows, the explicit expression of $c_1$ is given in Lemma A.5. Based on this lemma, the uniform upper bound of $\frac{1}{m} \sum_{j=1}^{m} \left( \tilde{b}_j(a_j^* z) \right) R |a_j^* v|^2$ is given in Lemma A.6.

Lemma A.5: Suppose that $\beta \geq 1$ is fixed constant. Assume $a_j \in \mathbb{C}^d, j = 1, \ldots, m$, are i.i.d. complex Gaussian random vectors and $b \in \mathbb{C}^m$ obeys $\sum_{j=1}^{m} |b_j|^4 \leq m$ and $\|b\|_{\infty} \leq \sqrt{\log m}$. For any $\epsilon \in (0, 1)$, if $m > C \epsilon^{-2} \log(1/\epsilon) \log m$ then with probability at least $1 - 2 \exp(-c_0^2 d/\log m) - 2 \exp(-c'' \epsilon d)$ it holds that

$$
\frac{1}{m} \sum_{j=1}^{m} |b_j||a_j^* v|^2 \{ |a_j^* v| \geq \beta \} \leq \left( 2 \beta e^{-0.49 \beta^2} + \epsilon \right) \frac{\|b\|_{\infty}^2}{m} + \epsilon
$$

for all $v \in S_{d-1}^+$. Here, $C, c_0, c'' > 0$ are universal constants.

Proof: We introduce an auxiliary Lipschitz function:

$$
\chi(t) = \begin{cases} t, & \text{if } t \geq \beta^2; \\
\frac{1}{\delta} t - \left( \frac{1}{\delta} - 1 \right) \beta^2, & \text{if } (1 - \delta) \beta^2 \leq t \leq \beta^2; \\
0, & \text{otherwise.}
\end{cases}
$$

Here $\delta \in (0, 1)$ is a constant which will be chosen later. Then

$$
\frac{1}{m} \sum_{j=1}^{m} |b_j||a_j^* v|^2 \{ |a_j^* v| \geq \beta \} \leq \frac{1}{m} \sum_{j=1}^{m} |b_j|^2 \chi(|a_j^* v|^2). \quad (A.1)
$$

For any fixed $v_0 \in S_{d-1}^+$, the terms $\chi(|a_j^* v_0|^2)$ are independent sub-exponential random variables. According to Bernstein’s inequality, for any fixed $t \geq 0$, it holds that

$$
P \left\{ \frac{1}{m} \sum_{j=1}^{m} |b_j|^2 \chi(|a_j^* v_0|^2) - \frac{|b|^2}{m} E \left( \chi(|a_j^* v_0|^2) \right) \right\} \leq 2 \exp \left( -c_1 \epsilon^2 m/\log m \right), \quad (A.2)
$$

where $c > 0$ is a universal constant. Recall that $\sum_{j=1}^{m} |b_j|^4 \leq m$ and $\|b\|_{\infty}^2 \leq \log m$. For any $0 < \epsilon < 1$, taking $t := \epsilon/2$ in (A.2), we obtain that with probability at least $1 - 2 \exp(-c_1 \epsilon^2 m/\log m)$ it holds

$$
\frac{1}{m} \sum_{j=1}^{m} |b_j|^2 \chi(|a_j^* v_0|^2) \leq \frac{|b|^2}{m} E \left( \chi(|a_j^* v_0|^2) \right) + \epsilon. \quad (A.3)
$$

where $c_1 > 0$ is a universal constant.

We next show that (A.3) holds for all $v \in S_{d-1}^+$. Suppose that $\mathcal{N}$ is a $\epsilon_0$-net over $S_{d-1}^+$ with $\# \mathcal{N} \leq (1 + \frac{1}{\epsilon_0})^2$. Then for any $v \in S_{d-1}^+$, there exists a $v_0 \in \mathcal{N}$ such that $\|v - v_0\|_2 \leq \epsilon_0$. Note that $\chi(t)$ is a Lipschitz function with Lipschitz constant $1/\delta$. We obtain that if $m \geq C_1 d \log m$ for a universal constant $C_1 > 0$ then

$$
\frac{1}{m} \sum_{j=1}^{m} |b_j|^2 \chi(|a_j^* v|^2) - \frac{1}{m} \sum_{j=1}^{m} |b_j|^2 \chi(|a_j^* v_0|^2)
\leq \frac{1}{m \delta} \sum_{j=1}^{m} |b_j|^2(|a_j^* v| + |a_j^* v_0|) \left| a_j^*(v - v_0) \right|
\leq \frac{1}{\delta} \left( \frac{1}{m} \sum_{j=1}^{m} |b_j|^2 |a_j^* v|^2 + \frac{1}{m} \sum_{j=1}^{m} |b_j|^2 |a_j^* v_0|^2 \right)
\cdot \left( \frac{1}{m} \sum_{j=1}^{m} |b_j|^2 \chi(|v - v_0|^2) \right)
\leq \frac{1}{\delta} \left( \frac{\|b\|_{\infty}^2}{m} + \frac{1}{m} \sum_{j=1}^{m} |b_j|^4 + \frac{d \|b\|_{\infty}^2}{m} \right) \epsilon_0
\leq \frac{1}{\delta} \left( \frac{\|b\|_{\infty}^2}{m} + 1 \right) \epsilon_0. \quad (A.4)
$$

Here, the inequality (a) follows from Lemma A.3 which says that, with probability at least $1 - 2 \exp(-c_3 d)$, it holds that

$$
\sum_{j=1}^{m} |b_j|^2 |a_j^* w|^2 \leq \epsilon \sum_{j=1}^{m} |b_j|^2 a_j a_j^* w
\leq \epsilon \left( \|b\|_{\infty}^2 + \sqrt{d} \sum_{j=1}^{m} |b_j|^4 + d \|b\|_{\infty}^2 \right) \|w\|_2^2
$$

for all $w \in \mathbb{C}^d$, where $c_3 > 0$ is a universal constant. Here, we use the fact that $\sum_{j=1}^{m} |b_j|^4 \leq m$ and $\|b\|_{\infty} \leq \log m$. Choosing
\( \varepsilon_0 := c_3\varepsilon \delta \) in (A.4) for some universal constant \( c_3 > 0 \) and taking the union bound over \( \mathcal{N} \), we obtain that
\[
\frac{1}{m} \sum_{j=1}^{m} |b_j| \chi(\langle a_j^* v \rangle^2) \leq \frac{\|b\|_2^2}{m} \cdot (\mathbb{E}(\chi(\langle a_j^* v \rangle^2)) + \varepsilon)
\]  
(A.5)
holds for all \( v \in S_{d-1}^{c_\ell} \) with probability at least \( 1 - 2 \exp(-c_\ell^2m/\log m) - 2 \exp(-c_2d) \) provided \( m \geq C' \varepsilon^{-2}\log(\varepsilon^{-1}d^{-1}) \) \( m \log m \), where \( C', c_\ell > 0 \) are universal constants.

We next turn to \( \mathbb{E}(\chi(\langle a_j^* v \rangle^2)) \) in (A.5). A simple calculation shows that for any \( \gamma > 0 \) it holds
\[
\mathbb{E}(\langle a_j^* v \rangle^2 | \langle a_j^* v \rangle \geq \gamma) \leq \frac{\gamma + 1}{\sqrt{\pi}} e^{-\gamma^2/2}. 
\]  
(A.6)
Note that \( \mathbb{E}(\chi(\langle a_j^* v \rangle^2)) \leq \mathbb{E}(\langle a_j^* v \rangle^2 | \langle a_j^* v \rangle \geq (1-\delta)^2 \gamma \). Set \( \delta := 0.01 \) and \( \gamma := (1-\delta)^2 \gamma \). It then follows from (A.1), (A.5) and (A.6) that with probability at least \( 1 - 2 \exp(-c_\ell^2m/\log m) - 2 \exp(-c_2d) \), it holds
\[
\frac{1}{m} \sum_{j=1}^{m} |b_j|^2 |\langle a_j^* v \rangle | |\langle a_j^* v \rangle | \geq \beta \leq (2\beta e^{-0.49\beta^2} + \varepsilon) \cdot \frac{\|b\|_2^2}{m} + \varepsilon
\]  
for all \( v \in S_{d-1}^{c_\ell} \), provided \( m \geq C e^{-2}\log(1/\varepsilon) / \log m \). Here, \( C > 0 \) is a universal constant. We complete the proof. \( \square \)

**Lemma A.6**: Suppose that \( a_j \in \mathbb{C}^d, j = 1, \ldots, m \), are i.i.d. complex Gaussian random vectors. Assume that \( a_0 \in \mathbb{C}^m \) is a vector obeying \( \sum_{j=1}^{m} |b_j|^4 \leq m \) and \( \|b\|_\infty \leq \sqrt{\log m} \).

For any \( \varepsilon \in (0, 1) \), if \( m \geq c(\varepsilon)d\log m \) then the following holds with probability at least \( 1 - 6 \exp(-c(\varepsilon)m/\log m) - 6 \exp(-c'\varepsilon)d) \):
\[
\left| \frac{1}{m} \sum_{j=1}^{m} \left( b_j(a_j^* z) \right) \cdot |a_j^* v| \right|^2 
\leq \varepsilon \left( \frac{\|b\|_2^2}{m} + 1 \right) \left( 1 + \left( \frac{1}{m} \sum_{j=1}^{m} |a_j^* z|^2 |a_j^* v|^2 \right)^{\frac{1}{2}} \right)
\]  

for all \( z, v \in S_{d-1}^{c_\ell} \), where \( c'' \) is a positive universal constant and \( c(\varepsilon), c'(\varepsilon) \) are positive constants depending only on \( \varepsilon \).

**Proof**: Suppose that \( \phi \in C_c^\infty(\mathbb{R}) \) is a Lipschitz continuous function satisfying \( 0 \leq \phi(x) \leq 1 \) for all \( x \in \mathbb{R} \). We furthermore require \( \phi(x) = 1 \) for \( |x| \leq 1 \) and \( \phi(x) = 0 \) for \( |x| \geq 2 \). For any \( \beta > 0 \), we have
\[
\left| \frac{1}{m} \sum_{j=1}^{m} \left( b_j(a_j^* z) \right) \cdot |a_j^* v| \right|^2 
\leq \frac{1}{m} \sum_{j=1}^{m} b_j \phi \left( \frac{a_j^* v}{\beta} \right) \|a_j^* v\|_2^2 
\leq \frac{1}{m} \sum_{j=1}^{m} b_j \phi \left( \frac{a_j^* v}{\beta} \right) \|\nabla\phi(\frac{a_j^* v}{\beta})\|_2 \|a_j^* v\|_2 
\leq \frac{1}{m} \sum_{j=1}^{m} b_j |\nabla\phi(\frac{a_j^* v}{\beta})| \|a_j^* v\|_2 
\leq 2 \left( \frac{\|b\|_2^2}{m} + 1 \right) \beta^2 \delta,
\]  
(A.7)
We claim that for any \( \varepsilon \in (0, 1) \) there exists a sufficiently large \( \beta > 1 \) such that if \( m \geq c(\varepsilon)d\log m \) then, with probability at least \( 1 - 6 \exp(-c(\varepsilon)m/\log m) - 6 \exp(c'\varepsilon d) \), it holds
\[
T_1 \leq \frac{\varepsilon}{2} \left( \frac{\|b\|_2^2}{\sqrt{m}} + 1 \right), \quad T_2 \leq \frac{\varepsilon}{2} \left( \frac{\|b\|_2^2}{\sqrt{m}} + 1 \right), \quad (A.8)
\]
\[
r \leq \varepsilon \left( \frac{\|b\|_2^2}{\sqrt{m}} + 1 \right) \left( \frac{1}{m} \sum_{j=1}^{m} |a_j^* z|^2 |a_j^* v|^2 \right)^{\frac{1}{2}} \quad (A.9)
\]
for all \( z, v \in S_{d-1}^{c_\ell} \). Here \( c(\varepsilon), c'(\varepsilon) \) are constants depending on \( \varepsilon \) and \( c''(\varepsilon) \) is a positive universal constant. Combining (A.7), (A.8) and (A.9), we arrive at the conclusion.

It remains to prove the claims (A.8) and (A.9). We first present an upper bound for \( T_1 \). For any fixed \( z_0, v_0 \in S_{d-1}^{c_\ell} \), due to the cut-off \( \phi \left( \frac{|a_j^* v|}{\beta} \right) \), the terms \( (a_j^* z_0)\langle a_j^* v \rangle^2 \phi \left( \frac{|a_j^* v_0|}{\beta} \right) \) are centered, independent sub-gaussian random variables with the sub-gaussian norm \( O(\beta^2) \). According to Hoeffding’s inequality, we obtain that the following holds with probability at least \( 1 - 2 \exp(-c\beta^2 |d| m) \):
\[
\left| \frac{1}{m} \sum_{j=1}^{m} b_j \langle a_j^* z_0 \rangle |a_j^* v_0|^2 \phi \left( \frac{|a_j^* v_0|}{\beta} \right) \right| \leq \frac{\|b\|_2^2}{4\sqrt{m}}, \quad (A.10)
\]
where \( c > 0 \) is a universal constant. We next show that (A.10) holds for all unit vectors \( z, v \in S_{d-1}^{c_\ell} \). Assume that \( \mathcal{N} \) is a \( \delta \)-net of the unit complex sphere in \( S_{d-1}^{c_\ell} \) and hence the covering number \#\( \mathcal{N} \) \( \leq (1 + \frac{2}{\delta})^{2d} \). For any \( z, v \in S_{d-1}^{c_\ell} \), there exists a \( (z_0, v_0) \in \mathcal{N} \times \mathcal{N} \) such that \( \|z - z_0\|_2 \leq \delta \) and \( \|v - v_0\|_2 \leq \delta \). Noting \( f(\tau) := \tau^2 \phi(\tau/\beta) \) is a bounded function with Lipschitz constant \( O(\beta) \), we obtain that
\[
\left| \frac{1}{m} \sum_{j=1}^{m} b_j \langle a_j^* z_0 \rangle |a_j^* v_0|^2 \phi \left( \frac{|a_j^* v_0|}{\beta} \right) \right| 
\leq \frac{1}{m} \sum_{j=1}^{m} b_j \langle a_j^* z_0 \rangle |a_j^* v_0|^2 \phi \left( \frac{|a_j^* v_0|}{\beta} \right) 
\leq \frac{\beta^2}{m} \left| \sum_{j=1}^{m} b_j |a_j^* v_0|^2 \right| 
\leq \frac{\beta^2}{m} \left| \sum_{j=1}^{m} b_j |a_j^* v_0|^2 \right| 
\leq 2 \left( \frac{\|b\|_2^2}{\sqrt{m}} + 1 \right) \beta^2 \delta,
\]  
where the fourth inequality follows from Lemma A.3 which says, with probability at least \( 1 - 2 \exp(-c_1d) \) for a universal
constant $c_1 > 0$, the following holds

$$\sum_{j=1}^{m} |b_j| |a_j^*| \leq \left( \sum_{j=1}^{m} |b_j| + \sqrt{d} \|b\|_2 + d \|b\|_\infty \right) \|w\|_2^2 \leq \left( \sqrt{m} \|b\|_2 + m \right) \|w\|_2^2 \quad \text{for all } w \in \mathbb{C}^d.$$ 

Choosing $\delta = c_2 \epsilon / \beta^2$ for some universal constant $c_2 > 0$ and taking the union bound, we obtain that

$$T_1 = \frac{1}{m} \sum_{j=1}^{m} b_j R(a_j^* z) \big| a_j^* v \big|^2 \phi \left( \frac{\big| a_j^* v \big|}{\beta} \right)$$

$$\leq \left( \frac{\|b\|_2}{\sqrt{m}} + 1 \right) \frac{\epsilon}{2} \quad \text{for all } z, v \in \mathbb{S}^{d-1}_{\mathbb{C}}$$

holds with probability at least $1 - 2 \exp(-c_3 \epsilon^2 \beta^{-2} m) - 2 \exp(-c_1 d)$ provided $m \geq C \cdot (\beta / \epsilon)^2 \log(\beta / \epsilon) d \log m$. Here, $C$ and $c_3$ are positive universal constants. Using the similar argument as above, we obtain that the following holds with probability at least $1 - 2 \exp(-c_1 \epsilon^2 \beta^{-2} m) - 2 \exp(-c_1 d)$:

$$T_2 = \frac{1}{m} \sum_{j=1}^{m} b_j R(a_j^* z) \big| a_j^* v \big|^2 \phi \left( \frac{\big| a_j^* v \big|}{\beta} \right)$$

$$\leq \left( \frac{\|b\|_2}{\sqrt{m}} + 1 \right) \frac{\epsilon}{2} \quad \text{for all } z, v \in \mathbb{S}^{d-1}_{\mathbb{C}}$$

provided $m \geq C \cdot (\beta / \epsilon)^2 \log(\beta / \epsilon) d \log m$. Finally, we turn to prove the claim $(A.9)$. Applying Cauchy-Schwarz inequality, we obtain

$$\frac{1}{m} \sum_{j=1}^{m} |b_j| |a_j^* z| |a_j^* v|^2 \chi \{ |a_j^* v| \geq \beta \}$$

$$\leq \left( \frac{1}{m} \sum_{j=1}^{m} |a_j^* z|^2 |a_j^* v|^2 \right) \left( \frac{1}{m} \sum_{j=1}^{m} |b_j|^2 |a_j^* v|^2 \chi \{ |a_j^* v| \geq \beta \} \right).$$

(A.11)

According to Lemma A.5, we obtain that if $m \geq C_1 \epsilon^{-4} \log(1 / \epsilon^2) d \log m$ then, with probability at least $1 - 2 \exp(-c_4 \epsilon^4 m / \log m) - 2 \exp(-c_5 d)$, the following holds

$$\frac{1}{m} \sum_{j=1}^{m} |b_j|^2 |a_j^* v|^2 \chi \{ |a_j^* v| \geq \beta \} \leq \frac{2 \beta \epsilon^{-0.49 \beta^2 + \epsilon^2 / m}}{2} \frac{\|b\|_2^2}{m} + \frac{\epsilon^2}{2}$$

$$\leq \left( \frac{\|b\|_2}{\sqrt{m}} + 1 \right) \frac{\epsilon}{2} \leq \left( \frac{\|b\|_2}{\sqrt{m}} + 1 \right)^2 \frac{\epsilon}{2} \quad \text{for all } v \in \mathbb{S}^{d-1}_{\mathbb{C}},$$

(A.12)

for all $v \in \mathbb{S}^{d-1}_{\mathbb{C}}$, where $C_1, c_4$ and $c_5$ are positive universal constants. Here, in the second inequality we take $\beta$ to be sufficiently large (depending only on $\epsilon$). Combining (A.11) and (A.12), we arrive at (A.9).

For the sake of completeness, the following lemma provides an upper bound on the spectral norm of

$$\frac{1}{m} \sum_{j=1}^{m} \big( \bar{b}_j(a_j^* x) \big) a_j a_j^*.$$ 

This estimate is commonly employed for a finite sequence of Gaussian random variables, with the only difference being the coefficients $b_j$ instead of the constant 1 in previous results.

**Lemma A.7:** Suppose $a_j \in \mathbb{C}^d$, $j = 1, \ldots, m$, are i.i.d. complex Gaussian random vectors. Assume that $z \in \mathbb{C}^d$ is a fixed vector and $b \in \mathbb{C}^m$ satisfies $|b_j|^2 \leq m$ and $\|b\|_\infty \leq \sqrt{\log m}$. For any $\epsilon \in (0, 1)$, if $m \geq C_1 \epsilon^{-2} \log m$ then, with probability at least $1 - 2 \exp(-c_2 \epsilon^2 m / \log m) - c_1 \epsilon^{-2} m^{-1}$, it holds that

$$\left\| \frac{1}{m} \sum_{j=1}^{m} \big( \bar{b}_j(a_j^* x) \big) a_j a_j^* \right\|_2 \leq \epsilon \left( \frac{\|b\|_2}{\sqrt{m}} + 1 \right) \|x\|_2.$$

Here, $C_1, c_2$ and $c_3$ are positive universal constants.

**Proof:** Assume that $N$ is a $1/4$-net of the complex unit sphere $\mathbb{S}^{d-1}_{\mathbb{C}}$ with the cardinality $\#N \leq \delta^{2d}$. According to [33, Lemma 4.4.3], we have

$$\left\| \frac{1}{m} \sum_{j=1}^{m} \big( \bar{b}_j(a_j^* x) \big) a_j a_j^* \right\|_2 \leq 2 \max_{w \in N} \frac{1}{m} \sum_{j=1}^{m} \big( \bar{b}_j(a_j^* x) \big) a_j a_j^* |w|_2.$$

Without loss of generality, we assume $x = e_1$. Then

$$\frac{1}{m} \sum_{j=1}^{m} \big( \bar{b}_j(a_j^* x) \big) a_j a_j^* |w|_2$$

$$= \frac{1}{m} |v_1|^2 \sum_{j=1}^{m} \big( \bar{b}_j(a_j) \big) a_j |w|_2$$

$$= \frac{1}{m} \frac{1}{m} \sum_{j=1}^{m} \big( \bar{b}_j(a_j) \big) a_j |w|_2$$

where $a_j, v \in \mathbb{C}^{d-1}$ are generated by deleting the first entry of the vector $a_j$ and $v$, respectively. For the first term, a simple calculation shows that

$$\text{Var} \left( \big( \bar{b}_j a_j \big) \right) \leq 3 |b_j|^2.$$

It then follows from Chebyshev’s inequality that

$$\mathbb{P} \left( \left| \frac{1}{m} \sum_{j=1}^{m} \big( \bar{b}_j(a_j) \big) a_j |w|_2 \right| \geq t \right) \leq \frac{3 |b_j|^2}{m^2 t^2}. \quad (A.13)$$

For any $0 < \epsilon < 1$, taking $t := \epsilon \|b\|_2 / \sqrt{m}$ in (A.13), we obtain that with probability at least $1 - 3 \epsilon^2 m^{-1}$ it holds

$$\frac{1}{m} \sum_{j=1}^{m} \big( \bar{b}_j(a_j) \big) a_j |w|_2 \leq \epsilon \|b\|_2 / \sqrt{m}.$$

For the second term $\frac{1}{m} \sum_{j=1}^{m} \big( \bar{b}_j(a_j) \big) a_j |w|_2$, note that

$$\frac{1}{m} \sum_{j=1}^{m} \big( \bar{b}_j(a_j) \big) a_j |w|_2$$

$$= \frac{1}{m} \sum_{j=1}^{m} b_j R(a_j, \Re \{ |a_j^* v| \}^2 - \frac{1}{m} \sum_{j=1}^{m} b_j \Re \{ a_j, \Re \{ |a_j^* v| \}^2 \}.$$

For any fixed $\tilde{v}_0 \in \mathbb{C}^{d-1}$, the terms $|a_j^* \tilde{v}_0|^2 - \|\tilde{v}_0\|_2^2$ are centered subexponential random variables with the maximal subexponential norm $K := O(\|\tilde{v}_0\|_2^2)$. Furthermore, $|a_j^* \tilde{v}_0|^2$ are
independent with $a_{j,1}$. Bernstein’s inequality gives

$$
\Pr\left( \frac{1}{m} \sum_{j=1}^{m} b_{j,\mathbb{R}} a_{j,1,\mathbb{R}} \left( \left| \tilde{a}_{j} \tilde{v} \right|^2 - \left\| \tilde{v} \right\|^2 \right) \geq t \left\| \tilde{v} \right\|^2 \right) \\
\leq 2 \exp\left( -c \min \left( \frac{m^2 t^2}{\sum_{j=1}^{m} b_{j,\mathbb{R}} a_{j,1,\mathbb{R}} \left\| \tilde{v} \right\|^2}, \max_{j \in [m]} \left\| b_{j,\mathbb{R}} a_{j,1,\mathbb{R}} \right\| \right) \right),
$$

where $c > 0$ is a universal constant. Taking $t := \epsilon / 2$, together with the union bound over $\mathcal{N}$, we obtain that

$$
\Pr\left( \frac{1}{m} \sum_{j=1}^{m} b_{j,\mathbb{R}} a_{j,1,\mathbb{R}} \left( \left| \tilde{a}_{j} \tilde{v} \right|^2 - \left\| v \right\|^2 \right) \geq \frac{\epsilon}{2} \left\| v \right\|^2 \right) \\
\leq 2 \exp\left( -4c \min \left( \frac{m^2 t^2}{\sum_{j=1}^{m} b_{j,\mathbb{R}} a_{j,1,\mathbb{R}} \left\| v \right\|^2}, \max_{j \in [m]} \left\| b_{j,\mathbb{R}} a_{j,1,\mathbb{R}} \right\| \right) + 5d \right),
$$

(A.14)

for all $v \in \mathcal{N}$. By Chebyshev’s inequality, we obtain that the following holds with probability at least $1 - \epsilon^{-2} m^{-1}$

$$
\left\| \frac{1}{m} \sum_{j=1}^{m} b_{j,\mathbb{R}} a_{j,1,\mathbb{R}} \right\|^2 \leq \frac{c \left\| b_{\mathbb{R}} \right\| \left\| v \right\|^2}{\sqrt{m}}.
$$

(A.15)

Similarly, the Chebyshev’s inequality implies

$$
\Pr\left( \sum_{j=1}^{m} b_{j,\mathbb{R}} a_{j,1,\mathbb{R}} - \frac{1}{2} \left\| b_{\mathbb{R}} \right\|^2 \geq m \right) \leq \sum_{j=1}^{m} |b_j|^4 \leq m
$$

(A.16)

provided $\sum_{j=1}^{m} |b_j|^4 \lesssim m$. Moreover, a union bound gives

$$
\max_{1 \leq j \leq m} |a_{j,1}| \leq \sqrt{10 \log m}
$$

(A.17)

with probability at least $1 - m^{-2}$. Putting (A.15), (A.16) and (A.17) into (A.14), we obtain that, with probability at least $1 - 2 \exp(-c_1 \epsilon^2 m / \log m) - c_2 \epsilon^{-2} m^{-1} - m^{-2}$, it holds

$$
\left\| \frac{1}{m} \sum_{j=1}^{m} b_{j,\mathbb{R}} a_{j,1,\mathbb{R}} \tilde{a}_{j} \tilde{v} \right\|^2 \leq \left( \frac{\left\| b_{\mathbb{R}} \right\|}{\sqrt{m}} + \frac{\epsilon}{2} \right) \left\| \tilde{v} \right\|^2
$$

provided $\left\| b \right\| \leq \sqrt{m} \sum_{j=1}^{m} |b_j|^4 \lesssim \sqrt{m}$, $\left\| b \right\|_{\infty} \lesssim \sqrt{\log m}$ and $m \geq C_1 \epsilon^{-2} d \log m$. Here, $C_1, c_1$ and $c_2$ are positive universal constants. Using the same argument as above, we obtain

$$
\left\| \frac{1}{m} \sum_{j=1}^{m} b_{j,\mathbb{R}} a_{j,1,\mathbb{R}} \tilde{a}_{j} \tilde{v} \right\|^2 \leq \left( \frac{\left\| b_{\mathbb{R}} \right\|}{\sqrt{m}} + \frac{\epsilon}{2} \right) \left\| \tilde{v} \right\|^2
$$

for all $v \in \mathcal{N}$.

Combining the above estimators, we have that

$$
\frac{1}{m} \sum_{j=1}^{m} \left( \tilde{b}_j \left( \tilde{a}_j^* \tilde{x} \right) \right)_{\mathbb{R}} \left| \tilde{a}_j^* v \right|^2 \leq \left( \frac{\left\| b_{\mathbb{R}} \right\|}{\sqrt{m}} + 1 \right) \left\| v \right\|^2
$$

for all $v \in \mathcal{N}$ holds with probability at least $1 - 2 \exp(-c_2' \epsilon^2 m / \log m) - c'' \epsilon^{-2} m^{-1}$, provided $m \geq C_2 \epsilon^{-2} d \log m$. Here, $c_2' > 0$ and $c'' > 0$ are universal constants. This completes the proof.

Lemma A.8: Suppose that $a_j \in \mathbb{C}^d$, $j = 1, \ldots, m$, are i.i.d. complex Gaussian random vectors. Assume that $b \in \mathbb{C}^m$ which satisfies $\sum_{j=1}^{m} |b_j|^4 \lesssim m$ and $\left\| b \right\|_{\infty} \leq \sqrt{\log m}$. For any $\epsilon \in (0, 1)$, if $m \geq C'' \epsilon^{-2} \log(1/\epsilon) / \log m$ then, with probability at least $1 - 6 \exp(-c'' \epsilon^2 m / \log m) - 6 \exp(-c'' d)$, it holds that

$$
\frac{1}{m} \sum_{j=1}^{m} \left( \tilde{b}_j \left( \tilde{a}_j^* v \right) \right)^2 \geq \frac{1 - \epsilon}{2} \frac{\left\| b \right\|^2}{m} - \epsilon
$$

for all $v \in \mathcal{S}_C^{-1}$.

Here, $C'', c', c'' > 0$ are universal constants.

Proof: A simple calculation shows that

$$
\left( \tilde{b}_j \left( \tilde{a}_j^* v \right) \right)^2 = b_j^2 \left( a_j^* v \right)^2 + b_j^2 \left( a_j^* v \right)_\mathbb{R} + 2 b_j v_j \left( a_j^* v \right)_\mathbb{R} \left( a_j^* v \right)_\mathbb{C}
$$

(A.18)

We first give a lower bound for the first term $\frac{1}{m} \sum_{j=1}^{m} b_j^2 \left( a_j^* v \right)^2 _\mathbb{R}$. Note that $E \left( \tilde{a}_j^* v \right)^2 = 1 / 2$. For any fixed $v_0 \in \mathcal{S}_C^{-1}$, by Bernstein’s inequality, we have

$$
\Pr\left( \frac{1}{m} \sum_{j=1}^{m} b_j^2 \left( a_j^* v_0 \right)^2 _\mathbb{R} \geq \frac{\left\| b_{\mathbb{R}} \right\|^2}{2m} \right) \geq \frac{\epsilon}{6}
$$

(A.19)

provided $m \geq C'' \epsilon^{-2} \log(1/\epsilon) / \log m$, where $C''$ and $c'$ are universal positive constants.

Similarly, for the second and third terms of (A.18), when $m \geq C'' \epsilon^{-2} \log(1/\epsilon) / \log m$, with probability at least $1 - 2 \exp(-c'' \epsilon^2 m / \log m) - 2 \exp(-c'' d)$, the followings hold:

$$
\rho \sum_{j=1}^{m} b_j \left( a_j^* v \right)_\mathbb{R} \left( a_j^* v \right) \mathbb{R} \geq \frac{3 - \epsilon}{6} \frac{\left\| b_{\mathbb{R}} \right\|^2}{m} - \epsilon
$$

(A.20)

provided $m \geq C'' \epsilon^{-2} \log(1/\epsilon) / \log m$, where $C''$ and $c'$ are universal positive constants.

Similarly, for the second and third terms of (A.18), when $m \geq C'' \epsilon^{-2} \log(1/\epsilon) / \log m$, with probability at least $1 - 2 \exp(-c'' \epsilon^2 m / \log m) - 2 \exp(-c'' d)$, the followings hold:

$$
\rho \sum_{j=1}^{m} b_j \left( a_j^* v \right)_\mathbb{R} \left( a_j^* v \right) \mathbb{R} \geq \frac{3 - \epsilon}{6} \frac{\left\| b_{\mathbb{R}} \right\|^2}{m} - \epsilon
$$

(A.21)

and

$$
\rho \sum_{j=1}^{m} b_j \left( a_j^* v \right)_\mathbb{R} \left( a_j^* v \right) \mathbb{R} \leq \frac{\epsilon}{3} \left( \frac{\left\| b_{\mathbb{R}} \right\|^2}{m} + 1 \right).
$$

(A.22)

Combining (A.21), (A.22) and (A.20), we complete the proof.

The purpose of the following lemma is to address the quantity that incorporates the third powers of Gaussian random vectors.
variables. The proof follows a similar approach as that of Lemma A.6, and thus, we omit the detailed proof here.

Lemma A.9: Suppose that $a_j \in \mathbb{C}^d, j = 1, \ldots, m$, are i.i.d. complex Gaussian random vectors. Assume that $b \in \mathbb{C}^m$ is a vector obeying $\sum_{j=1}^{m}|b_j|^4 \leq m$ and $\|b\|_\infty \leq \sqrt{\log m}$. For any $\epsilon \in (0, 1)$, if $m \geq c(\epsilon)d^2\log m$ then the following holds with probability at least $1 - 6\exp(-c'(\epsilon)m/\log m) - 6\exp(-c''d)$:

$$
\left| \frac{1}{m} \sum_{j=1}^{m} (b_j(a_j^*v)^\Re (v^*a_ja_j^*z)^\Re) \right| \\
\leq \epsilon \left( \left\| \frac{\|b\|_2}{\sqrt{m}} + 1 \right\| + \left( \frac{1}{m} \sum_{j=1}^{m} |a_j^*z|^2 |a_j^*v|^2 \right)^{\frac{1}{2}} \right)
$$

for all $z, v \in \mathbb{S}^{d-1}_-$, where $c''$ is a positive universal constant and $c(\epsilon), c'(\epsilon)$ are positive constants depending only on $\epsilon$.

APPENDIX B

PROOFS OF TECHNICAL RESULTS IN SECTION III

Proof of Lemma III.1: A simple calculation shows that

$$
\frac{1}{m} \sum_{j=1}^{m} |a_j^*x + b_j|^2 = \frac{1}{m} \sum_{j=1}^{m} |a_j^*x|^2 + \frac{2}{m} \sum_{j=1}^{m} (b_j(a_j^*x))_{\Re} + \frac{\|b\|^2}{m}.
$$

We first consider the term $\frac{1}{m} \sum_{j=1}^{m} |a_j^*x|^2$. We use Bernstein’s inequality to obtain that

$$(1 - \epsilon)\|x\|_2^2 \leq \frac{1}{m} \sum_{j=1}^{m} |a_j^*x|^2 \leq (1 + \epsilon)\|x\|_2^2$$

for any $0 < \epsilon \leq 1$, holds with probability at least $1 - 2\exp(-c_1\epsilon^2m)$. Here, $c_1 > 0$ is a universal constant. For the second term, noting that

$$
\frac{1}{m} \sum_{j=1}^{m} (b_j(a_j^*x))_{\Re} = \frac{1}{m} \sum_{j=1}^{m} b_j, a_j(a_j^*x)_{\Re} + \frac{1}{m} \sum_{j=1}^{m} b_j, a_j^*a_j x_{\Re}.
$$

we use Hoeffding’s inequality to obtain that

$$
\left| \frac{1}{m} \sum_{j=1}^{m} b_j, a_j(a_j^*x)_{\Re} \right| \leq \frac{\epsilon \|b\|_2}{\sqrt{m}} \|x\|_2,
$$

and

$$
\left| \frac{1}{m} \sum_{j=1}^{m} b_j, a_j^*a_j x_{\Re} \right| \leq \frac{\epsilon \|b\|_2}{\sqrt{m}} \|x\|_2
$$

hold with probability at least $1 - 2\exp(-c_2\epsilon^2m)$, where $c_2 > 0$ is a universal constant.

Recall that $\|b\|_2 \leq c_3\sqrt{m}\|x\|_2$ for a universal constant $c_3 > 0$. Collecting the above results, we obtain that, with probability at least $1 - 4\exp(-c_4\epsilon^2m)$, the following holds

$$(1 - c_0\epsilon)\|x\|_2^2 \leq \frac{1}{m} \sum_{j=1}^{m} |a_j^*x + b_j|^2 - \frac{\|b\|^2}{m} \leq (1 + c_0\epsilon)\|x\|_2^2$$

for a universal constant $c_0 := 2\sqrt{2}c_3 + 1$. Here, $c_4 > 0$ is a universal constant. Taking $\epsilon := \frac{\epsilon}{4c_0}$ we obtain that the following holds with probability at least $1 - 4\exp(-cm)$:

$$
\frac{1}{4} \|x\|_2^2 \leq \frac{1}{m} \sum_{j=1}^{m} y_j - \frac{\|b\|^2}{m} \leq \frac{7}{4} \|x\|_2^2,
$$

where $c > 0$ is a universal constant, which implies

$$
\frac{1}{3} R_0 \leq \|x\|_2 \leq R_0,
$$

where $R_0 := 2 \left( \frac{1}{m} \sum_{j=1}^{m} y_j - \frac{\|b\|^2}{m} \right)^{1/2}$. This completes the proof.

$\square$

Proof of Lemma III.2: For any $z, z' \in S_R$, we have

$$
\|\nabla f(z) - \nabla f(z')\|_2
$$

$$
= \sqrt{2} \frac{1}{m} \left| \sum_{j=1}^{m} (a_j^*z + b_j - y_j)(a_j^*z + b_j - y_j) a_j \right|
$$

$$
- \sum_{j=1}^{m} (a_j^*z' + b_j - y_j)(a_j^*z' + b_j - y_j) a_j \right|
$$

$$
\leq \sqrt{2} \left| \frac{1}{m} \sum_{j=1}^{m} (a_j^*z + b_j - a_j^*z' - b_j) a_j a_j^*z' \right|
$$

$$
+ \sqrt{2} \left| \frac{1}{m} \sum_{j=1}^{m} (a_j^*z + b_j - a_j^*z' - b_j) b_j a_j \right|
$$

$$
+ \sqrt{2} \left| \frac{1}{m} \sum_{j=1}^{m} (a_j^*z + b_j - a_j^*z' - b_j) b_j a_j \right|
$$

$$
+ \sqrt{2} \left| \frac{1}{m} \sum_{j=1}^{m} (a_j^*z + b_j - a_j^*z' - b_j) b_j a_j \right|
$$

. (B.1)

Since $a_j \in \mathbb{C}^d$ are complex Gaussian random vectors, with probability at least $1 - c_m^{-d}$, the following holds

$$
\max_{1 \leq j \leq m} \|a_j\|_2 \leq 2\sqrt{d \log m},
$$

where $c_m > 0$ is a universal constant. Lemma A.4 implies that when $m \geq Cd$ for a universal constant $C > 0$, with probability at least $1 - 3\exp(-cm)$, it holds that

$$
\frac{1}{m} \sum_{j=1}^{m} |a_j^*z + b_j|^2 - |a_j^*z' + b_j|^2 |
$$

$$
\leq 3 \left( R + \frac{\|b\|_2}{\sqrt{m}} \right) \|z - z'\|_2$$

for all $z, z' \in S_R$,

where $c > 0$ is a universal constant. Combining the above two estimators, we obtain that the following holds with probability
provided $m \geq Cd$. Here, we use the fact that $\|z\|_2 \leq R$ due to $z' \in S_R$. Using the same argument as above, we obtain that
\[
\left\| \frac{1}{m} \sum_{j=1}^m \left( |a_j^* z + b_j|^2 - |a_j^* z'| + j b_j \right) a_j \right\|_2 \\
\leq \left\| \frac{1}{m} \sum_{j=1}^m \left( |a_j^* z + b_j|^2 - |a_j^* z'| + j b_j \right) a_j \right\|_2 \\
\leq 12 \cdot d \cdot \log m \cdot R \cdot \left( R + \frac{\|b\|_2}{\sqrt{m}} \right) \|z - z'\|_2, \quad (B.2)
\]
Noting that $\|\frac{1}{m} \sum_{j=1}^m a_j a_j^*\|_2 \leq 2$ holds with probability at least $1 - \exp(-cm)$, we obtain that
\[
\left\| \frac{1}{m} \sum_{j=1}^m \left( |a_j^* z + b_j|^2 - |a_j^* z'| + j b_j \right) a_j \right\|_2 \\
\leq \left\| \frac{1}{m} \sum_{j=1}^m \left( |a_j^* z + b_j|^2 - |a_j^* z'| + j b_j \right) a_j \right\|_2 \\
\leq \left( \max_{1 \leq j \leq m} \left\| \frac{1}{\sqrt{m}} \right\|_2 \right) \left\| \frac{1}{m} \sum_{j=1}^m a_j a_j^* \right\|_2 \|z - z'\|_2 \leq 4 \left( 4R^2d \log m + \|b\|_\infty \right) \|z - z'\|_2. \quad (B.4)
\]
Here, we use the inequality $|a_j^* z + b_j|^2 \leq 2(|a_j^* z|^2 + |b_j|^2)$ for any $j$. Similarly,
\[
\left\| \frac{1}{m} \sum_{j=1}^m \left( |a_j^* z + b_j|^2 - |a_j^* z'| + j b_j \right) a_j \right\|_2 \\
\leq 4 \left( 4R^2d \log m + \|b\|_\infty \right) \|z - z'\|_2. \quad (B.5)
\]
Substituting (B.2), (B.3), (B.4) and (B.5) into (B.1), we obtain that when $m \geq Cd$, with probability at least $1 - 4 \exp(-cm)$ $- cm^{-d}$, it holds that
\[
\|\nabla f(z) - \nabla f(z')\|_2 \leq C_R \|z - z'\|_2 \quad \text{for all} \quad z, z' \in S_R,
\]
where
\[
C_R = 6\sqrt{2} \left( 2Rd \log m + \|b\|_\infty \sqrt{d \log m} \right) \left( R + \frac{\|b\|_2}{\sqrt{m}} \right) \] \[+ 8\sqrt{2} \left( 2d \log m (R^2 + \|x\|_2^2 + \|b\|_\infty^2) \right).
\]
This completes the proof. \hfill \Box

REFERENCES

[1] M. S. Asif, and C. Hegde, “The benefits of side information for structured phase retrieval,” in *Proc. 28th Eur. Signal Process. Conf. (EUSIPCO)*, 2021, pp. 775–778.

[2] R. Balan, P. Casazza, and D. Edidin, “On signal reconstruction without phase,” *Appl. Comput. Harmon. Anal.*, vol. 20, no. 3, pp. 345–356, 2006.

[3] D. A. Barmherzig, J. Sun, P. N. Li, T. J. Lane, and E. J. Candes, “Holographic phase retrieval and reference design,” *Inverse Prob.*, vol. 35, no. 9, 2019, Art. no. 094001.

[4] R. Beinert and G. Plonka, “Enforcing uniqueness in one-dimensional phase retrieval by additional signal information in time domain,” *Appl. Comput. Harmon. Anal.*, vol. 45, no. 3, pp. 505–525, 2018.

[5] T. Bendory, R. Beinert, and Y. C. Eldar, “Fourier phase retrieval: Uniqueness and algorithms,” in *Compressed Sensing and its Applications*, Cham, Switzerland: Springer International Publishing AG, 2017, pp. 55–91.

[6] J. F. Cai, M. Huang, D. Li, and Y. Wang, “Solving phase retrieval with random initial guess is nearly as good as by spectral initialization,” *Appl. Comput. Harmon. Anal.*, vol. 58, pp. 60–84, May 2022.

[7] J. F. Cai, M. Huang, D. Li, and Y. Wang, “Nearly optimal bounds for the global geometric landscape of phase retrieval,” *Inverse Prob.*, vol. 39, no. 7, 2023, Art. no. 075011.

[8] E. J. Candes, X. Li, and M. Soltanolkotabi, “Phase retrieval via Wirtinger flow: Theory and algorithms,” *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1985–2007, Apr. 2015.

[9] E. J. Candes, T. Strohmer, and V. Voroninski, “Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming,” *Commun. Pure Appl. Math.*, vol. 66, no. 8, pp. 1241–1274, 2013.

[10] Y. Chen and E. J. Candes, “Solving random quadratic systems of equations is nearly as easy as solving linear systems,” *Commun. Pure Appl. Math.*, vol. 70, no. 5, pp. 822–833, 2017.

[11] Y. Chen, Y. Chi, J. Fan, and C. Ma, “Gradient descent with random initialization: Fast global convergence for nonconvex phase retrieval,” *Math. Program.*, vol. 176, no. 1, pp. 5–37, 2019.

[12] Y. Chi, Y. M. Lu, and Y. Chen, “Nonconvex optimization meets low-rank matrix factorization: An overview,” *IEEE Trans. Signal Process.*, vol. 67, no. 20, pp. 5239–5269, Oct. 2019.

[13] A. Conca, D. Edidin, M. Hering, and C. Vinzant, “An algebraic characterization of injectivity in phase retrieval,” *Appl. Comput. Harmon. Anal.*, vol. 38, no. 2, pp. 346–356, 2015.

[14] J. C. Dainty and J. R. Fienup, “Phase retrieval and image reconstruction for astronomy,” in *Image Recovery: Theory and Application*, vol. 231, Orlando, FL, USA: Academic Press, Inc., 1987, p. 275.

[15] D. Gabor, “A new microscopic principle,” *Nature*, vol. 161, pp. 777–778, Dec. 1948.

[16] B. Gao, Q. Sun, Y. Wang, Z. Xu, “Phase retrieval from the magnitudes of affine linear measurements,” *Adv. Appl. Math.*, vol. 93, pp. 121–141, Feb. 2018.

[17] B. Gao and Z. Xu, “Phaseless recovery using the Gauss–Newton method,” *IEEE Trans. Signal Process.*, vol. 65, no. 22, pp. 5885–5896, Nov. 2017.

[18] M. Guizar-Sicairos and J. R. Fienup, “Holography with extended reference by autocorrelation linear differential operation,” *Opt. Exp.*, vol. 15, no. 26, pp. 17592–17612, 2007.

[19] R. W. Harrison, “Phase problem in crystallography,” *J. Opt. Soc. Am. A.*, vol. 10, no. 5, pp. 1046–1055, 1993.

[20] M. Huang and Y. Wang, “Linear convergence of randomized Kaczmarz method for solving complex-valued phaseless equations,” *SIAM J. Imag. Sci.*, vol. 15, no. 2, pp. 989–1016, 2022.

[21] M. Huang and Z. Xu, “Phase retrieval from the norms of affine transformations,” *Adv. Appl. Math.*, vol. 130, 2021, Art. no. 102243.

[22] M. Huang and Z. Xu, “Performance bound of the intensity-based model for noisy phase retrieval,” submitted for publication.

[23] R. Hyder, Z. Cai, and M. S. Asif, “Solving phase retrieval with a learned reference,” in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 425–441.

[24] K. Jaganathan, Y. C. Eldar, and B. Hassibi, “Phase retrieval: An overview of recent developments,” *Optical Compressive Imaging*, Boca Raton, FL, USA: CRC Press, 2016, pp. 279–312.

[25] T. Latchevskii, “Iterative phase retrieval for digital holography: Tutorial,” *J. Opt. Soc. Am. A.*, vol. 36, no. 12, pp. 31–40, 2019.

[26] Z. Li, J. F. Cai, and K. Wei, “Towards the optimal construction of a loss function without spurious local minima for solving quadratic equations,” *IEEE Trans. Inf. Theory*, vol. 66, no. 5, pp. 3242–3260, May 2020.

[27] M. Liebling, T. Blu, E. Cuche, P. Marquet, C. Depeursinge, and P. Marquet, “Local amplitude and phase retrieval method for digital holography applied to microscopy,” in *Proc. Eur. Conf. Biomed. Opt.*, 2003, p. 5143–210.

[28] J. Miao, T. Ishikawa, Q. Shen, and T. Earnest, “Extending X-ray crystallography to allow the imaging of noncrystalline materials, cells, and single protein complexes,” *Annu. Rev. Phys. Chem.*, vol. 59, pp. 387–410, May 2008.
[29] R. P. Millane, “Phase retrieval in crystallography and optics,” J. Opt. Soc. Am. A, vol. 7, no. 3, pp. 394–411, 1990.

[30] H. Sahinoglou and S. D. Cabrera, “On phase retrieval of finite-length sequences using the initial time sample,” IEEE Trans. Circuits Syst., vol. 38, no. 8, pp. 954–958, Aug. 1991.

[31] J. Sun, Q. Qu, and J. Wright, “A geometric analysis of phase retrieval,” Found. Comput. Math., vol. 18, no. 5, pp. 1131–1198, 2018.

[32] Y. S. Tan and R. Vershynin, “Phase retrieval via randomized Kaczmarz: Theoretical guarantees,” Inf. Inference A J. IMA, vol. 8, no. 1, pp. 97–123, 2019.

[33] R. Vershynin, High-Dimensional Probability: An Introduction With Applications in Data Science. Cambridge, U.K.: Cambridge Univ. Press, 2018.

[34] A. Walther, “The question of phase retrieval in optics,” J. Mod. Opt., vol. 10, no. 1, pp. 41–49, 1963.

[35] G. Wang, G. B. Giannakis, and Y. C. Eldar, “Solving systems of random quadratic equations via truncated amplitude flow,” IEEE Trans. Inf. Theory, vol. 64, no. 2, pp. 773–794, Feb. 2018.

[36] Y. Wang and Z. Xu, “Generalized phase retrieval: Measurement number, matrix recovery and beyond,” Appl. Comput. Harmon. Anal., vol. 47, no. 2, pp. 423–446, 2019.

[37] Z. Yuan and H. Wang, “Phase retrieval with background information,” Inverse Prob., vol. 35, no. 5, 2019, Art. no. 054003.

[38] H. Zhang, Y. Zhou, Y. Liang, and Y. Chi, “A nonconvex approach for phase retrieval: Reshaped Wirtinger flow and incremental algorithms,” J. Mach. Learn. Res., vol. 18, no. 1, pp. 5164–5198, 2017.

Meng Huang received the M.S. and Ph.D. degrees from the Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China, in 2017 and 2019, respectively. He is an Associate Professor with the School of Mathematical Sciences, Beihang University. His research interests include phase retrieval, matrix sensing, and mathematical data science.

Zhiqiang Xu received the Ph.D. degree from the Department of Applied Mathematics, Dalian University of Technology, Dalian, China, in 2003. He is currently a Professor with the Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China. His research interests include approximation theory, compressed sensing, and phase retrieval.