LINEAR CONVERGENCE OF RANDOMIZED KACZMARZ METHOD FOR SOLVING COMPLEX-VALUED PHASELESS EQUATIONS

MENG HUANG AND YANG WANG

Abstract. A randomized Kaczmarz method was recently proposed for phase retrieval, which has been shown numerically to exhibit empirical performance over other state-of-the-art phase retrieval algorithms both in terms of the sampling complexity and in terms of computation time. While the rate of convergence has been studied well in the real case where the signals and measurement vectors are all real-valued, there is no guarantee for the convergence in the complex case. In fact, the linear convergence of the randomized Kaczmarz method for phase retrieval in the complex setting is left as a conjecture by Tan and Vershynin. In this paper, we provide the first theoretical guarantees for it. We show that for random measurements $a_j \in \mathbb{C}^n$, $j = 1, \ldots, m$ which are drawn independently and uniformly from the complex unit sphere, or equivalent are independent complex Gaussian random vectors, when $m \geq Cn$ for some universal positive constant $C$, the randomized Kaczmarz scheme with a good initialization converges linearly to the target solution (up to a global phase) in expectation with high probability. This gives a positive answer to that conjecture.

1. Introduction

1.1. Problem setup. Let $x \in \mathbb{C}^n$ (or $\mathbb{R}^n$) be an arbitrary unknown vector. We consider the problem of recovering $x$ from the phaseless equations:

$$b_j = |\langle a_j, x \rangle|, \quad j = 1, \ldots, m,$$

where $a_j \in \mathbb{C}^n$ (or $\mathbb{R}^n$) are known sampling vectors and $b_j \in \mathbb{R}$ are observed measurements. This problem, termed as phase retrieval, has been a topic of study from 1980s due to its wide range of practical applications in fields of physical sciences and engineering, such as X-ray crystallography [18, 28], diffraction imaging [10, 31], microscopy [26], astronomy [12], optics and acoustics [1, 2, 38] etc, where the detector can record only the diffracted intensity while losing the phase information. Despite its simple mathematical form, it has been shown that to reconstruct a finite-dimensional discrete signal from its Fourier transform magnitudes is generally NP-complete [30]. Another special case of solving this phaseless equations is the well-known stone problem in combinatorial optimization, which is also NP-complete [3].

Y. Wang was supported in part by the Hong Kong Research Grant Council grants 16306415 and 16308518.
To solve (1), we employ the randomized Kaczmarz method where the update rule is given by

\[
z_{k+1} = z_k - \left(1 - \frac{b_{i_k}}{\|a_{i_k}^* z_k\|^2}\right) a_{i_k}^* z_k, \tag{2}
\]

where \(i_k\) is chosen randomly from the \(\{1, \ldots, m\}\) with probability proportional to \(\|a_{i_k}\|^2\) at the \(k\)-th iteration. Actually, the update rule above is a natural adaptation of the classical randomized Kaczmarz method [22] for solving linear equations. The idea behind the scheme is simple. When the iteration is close enough to the signal vector \(x\), the phase information can be approximated by that of the current estimate. Thus, in each iteration, we first select a measurement vector \(a_{i_k}\) randomly, and then project the current estimate \(z_k\) onto the hyperplane

\[
\left\{ z \in \mathbb{C}^n : \langle a_{i_k}, z \rangle = b_{i_k} \cdot \frac{a_{i_k}^* z_k}{\|a_{i_k}^* z_k\|} \right\}.
\]

That gives the scheme (2).

We are interested in the following questions:

Does the randomized Kaczmarz scheme (2) converge to the target solution \(x\) (up to a global phase) in the complex setting? Can we establish the rate of convergence?

1.2. Motivation. The randomized Kaczmarz method for solving phase retrieval problem was proposed by Wei [41] in 2015. It has been demonstrated in [41] using numerical experiments that the randomized Kaczmarz method exhibits empirical performance over other state-of-the-art phase retrieval algorithms both in terms of the sampling complexity and in terms of computation time, when the measurements are real or complex Gaussian random vectors, or when they follow the coded diffraction pattern (CDP) model. However, no adequate theoretical guarantee for the convergence was established in [41]. To bridge the gap, for the real Gaussian measurement vectors, Li et al. [24] establish an asymptotic convergence of the randomized Kaczmarz method for phase retrieval, but it requires infinite number of samples, which is unrealistic in practical. Lately, Tan and Vershynin [34] use the chain argument coupled with bounds on Vapnik-Chervonenkis (VC) dimension and metric entropy, and then prove theoretically that the randomized Kaczmarz method for phase retrieval is linearly convergent with \(O(n)\) Gaussian random measurements, where \(n\) is the dimension of the signal. A result almost same to that of [34] is also obtained independently by Jeong and Gunturk [21] using
the tools of hyperplane tessellation and “drift analysis”. Another similar conditional error contrac-
tivity result is also established by Zhang et al. [42], which is called incremental reshaped Wirtinger
flow.

We shall emphasize that all results concerning the convergence of the randomized Kaczmarz
method for phase retrieval are for the real case where the signals and measurement vectors are all
real-valued. Since the phase can only be $+1$ or $-1$ in the real case, then the measurement vectors
can be divided into “good measurements” with correct phase and “bad measurements” with incorrect
phase. When the initial point is close enough to the true solution, the total influence of “bad mea-
surements” can be well controlled. However, this is not true for the complex measurements because
$xe^{i\theta}$ is continuous with respect to $\theta \in [0, 2\pi)$. For this reasons, the proofs for the real case can not
be generalized to the complex setting easily. As stated in [34, Section 7.2], the linear convergence
of the randomized Kaczmarz method for phase retrieval in the complex setting is left as a CONJEC-
TURE. We shall point out that the convergence of randomized Kaczmarz method for phase retrieval
in complex setting is of more practical interest.

In this paper, we aim to prove this conjecture by introducing a deterministic condition on measure-
ment vectors called “Restricted Strong Convexity” and then showing that the random measurements
drawn independently and uniformly from the complex-valued sphere, or equivalently for the complex
Gaussian random vectors, satisfy this condition with high probability, as long as the measurement
number $m \geq O(n)$.

1.3. Related Work.

1.3.1. Phase retrieval. The phase retrieval problem, which aims to recover $x$ from phaseless equa-
tions (1), has received intensive investigations recently. Note that if $z$ is a solution to (1) then $ze^{i\theta}$
is also the solution of this problem for any $\theta \in \mathbb{R}$. Therefore the recovery of the solution $x$ is up to
a global phase. It has been shown theoretically that $m \geq 4n - 4$ generic measurements suffice to
recover $x$ for the complex case [11, 40] and $m \geq 2n - 1$ are sufficient for the real case [2].

Many algorithms with provable performance guarantees have been designed to solve the phase
retrieval problem. One line of research relies on a “matrix-lifting” technique, which lifts the phase
retrieval problem into a low rank matrix recovery problem, and then a nuclear norm minimization
is adopted as a convex surrogate of the rank constraint. Such methods include PhaseLift [6, 8], PhaseCut [37] etc. While this convex methods have a substantial advance in theory, they tend to be computationally inefficient for large scale problems. Another line of research seeks to optimize a non-convex loss function in the natural parameter space, which achieves significantly improved computational performance. The first non-convex algorithm with theoretical guarantees was given by Netrapalli et al who proved that the AltMinPhase [29] algorithm, based on a technique known as spectral initialization, converges linearly to the true solution up to a global phase with $O(n \log^3 n)$ resampling Gaussian random measurements. This work led to further several other non-convex algorithms based on spectral initialization [4, 7, 9, 19, 36]. Specifically, Candès et al developed the Wirtinger Flow (WF) [7] method and proved that the WF algorithm can achieve linear convergence with $O(n \log n)$ Gaussian random measurements. Lately, Chen and Candès improved the result to $O(n)$ Gaussian random measurements by incorporating a truncation, namely the Truncated Wirtinger Flow (TWF) [9] algorithm. Other non-convex methods with provable guarantees include the Gauss-Newton [17], the trust-region [33], Smoothed Amplitude Flow [5], Truncated Amplitude Flow (TAF) algorithm [39], Reshaped Wirtinger Flow (RWF) [42] algorithm and Perturbed Amplitude Flow (PAF) [16] algorithm, to name just a few. We refer the reader to survey papers [20, 31] for accounts of recent developments in the theory, algorithms and applications of phase retrieval.

1.3.2. Randomized Kaczmarz method for linear equations. Kaczmarz method is one of the most popular algorithms for solving overdetermined system of linear equations [22], which iteratively project the current estimate onto the hyperplane of chosen equation at a time. Suppose the system of linear equations we want to solve is given by $Ax = y$, where $A \in \mathbb{C}^{m \times n}$. In each iteration of the Kaczmarz method, one row $a_{ik}$ of $A$ is selected and then the new iterate $z_{k+1}$ is obtained by projecting the current estimate $z_k$ orthogonally onto the solution hyperplane of $\langle a_{ik}, z \rangle = y_{ik}$ as follows:

$$z_{k+1} = z_k + \frac{y_{ik} - \langle a_{ik}, z_k \rangle}{\|a_{ik}\|^2} a_{ik}.$$  

The classical version of Kaczmarz method sweeps through the rows of $A$ in a cyclic manner, however, it lacks useful theoretical guarantees. Existing results in this manner are based on quantities of matrix $A$ which are hard to compute [13–15]. In 2009, Strohmer and Vershynin [32] propose a randomized Kaczmarz method where the row of $A$ is selected in random order and they prove that this randomized
Kaczmarz method is convergent with expected exponential rate. More precisely, at each step \( k \), if the index \( i_k \) is chosen randomly from the \( \{1, \ldots, m\} \) with probability proportional to \( \|a_{i_k}\|_2^2 \) then for any initial \( z_0 \), the iteration \( z_k \) given by randomized Kaczmarz scheme (3) obeys
\[
\mathbb{E}\|z_k - x\|_2 \leq \left( 1 - \frac{1}{k(A) \cdot n} \right)^{k/2} \cdot \|z_0 - x\|_2,
\]
where \( k(A) \) is the condition number of \( A \).

1.3.3. Randomized Kaczmarz method for phase retrieval. As stated before, the randomized Kaczmarz method for phase retrieval is proposed by Wei in 2015. He was able to show numerically [41] that the randomized Kaczmarz method exhibits empirical performance over other state-of-the-art phase retrieval algorithms, but lack of adequate theoretical performance guarantee. Lately, in the real case, when the measurements \( a_j \) are drawn independently and uniformly from the unit sphere, several results have been established independently to guarantee the linear convergence of the randomized Kaczmarz method under appropriate initialization.

For instance, Tan and Vershynin [34] prove that for any \( 0 < \delta, \delta_0 \leq 1 \) if \( m \gtrsim n \log(m/n) + \log(1/\delta_0) \) and \( a_j \in \mathbb{R}^n \) are drawn independently and uniformly from the unit sphere, then with probability at least \( 1 - \delta_0 \) it holds: the \( k \)-th step randomized Kaczmarz estimate \( z_k \) given by (2) satisfies
\[
\mathbb{E}_{\mathcal{I}_k} [\text{dist}(z_k, x) 1_{\tau = \infty}] \leq (1 - \frac{1}{2n})^{k/2} \text{dist}(z_0, x),
\]
provided \( \text{dist}(z_0, x) \leq c\delta \|x\|_2 \) for some constant \( c > 0 \). Furthermore, the probability \( \mathbb{P}(\tau < \infty) \leq \delta^2 \). Here \( \tau \) is the stopping time and \( \mathbb{E}_{\mathcal{I}_k} \) denotes the expectation with respect to randomness \( \mathcal{I}_k := \{i_1, i_2, \ldots, i_k\} \) conditioned on the high probability event of random measurements \( \{a_j\}_{j=1}^m \).

1.4. Our Contributions. As stated before, randomized Kaczmarz method is a popular and convenient method for solving phase retrieval problem due to its fast convergence and low computational complexity. For the real setting, the theoretical guarantee of linear convergence has been established, however, there is no result concerning the rate of convergence in the complex setting. Since there is an essential difference between the real setting and complex setting, the convergence of randomized Kaczmarz method in the complex setting has been left as a conjecture [34, Section 7.2]. The goal of this paper is to give a positive answer to this conjecture, as shown below.
Theorem 1.1. Assume that the measurement vectors $a_1, \ldots, a_m \in \mathbb{C}^n$ are drawn independently and uniformly from the unit sphere $S^{n-1}_\mathbb{C}$. For any $0 < \delta < 1$, let $z_0$ be an initial estimate to $x$ such that $\text{dist}(z_0, x) \leq 0.01\|x\|_2$. There exist universal constants $C_0, c_0 > 0$ such that if $m \geq C_0n$ then with probability at least $1 - 14\exp(-c_0n)$ it holds: the iteration $z_k$ given by randomized Kaczmarz update rule (2) obeys
\[
\mathbb{E}_{T_k}[\text{dist}(z_k, x)1_{\tau=\infty}] \leq (1 - 0.03/n)^{k/2}\text{dist}(z_0, x),
\]
where $\tau$ is the stopping time defined by
\[
\tau := \min \{ k : z_k \notin B \} \quad \text{with} \quad B := \{ z : \text{dist}(z, x) \leq 0.01\|x\|_2 \}.
\]
Furthermore, the probability $\mathbb{P}(\tau < \infty) \leq \delta^2$. Here $\mathbb{E}_{T_k}$ denotes the expectation with respect to randomness $T_k := \{i_1, i_2, \ldots, i_k\}$ conditioned on the high probability event of random measurements $\{a_j\}_{j=1}^m$.

The theorem asserts that the randomized Kaczmarz method converges linearly to the global solution $x$ (up to a global phase) in expectation for random measurements $a_j \in \mathbb{C}^n$ which are drawn independently and uniformly from the complex unit sphere, or equivalent are independent complex Gaussian random vectors, with an optimal sample complexity.

Remark 1.2. Theorem 1.1 requires an initial estimate $z_0$ which is close to the target solution. In fact, a good initial estimate can be obtained easily by spectral initialization which is widely used in non-convex algorithms for phase retrieval. For instance, when $a_j \in \mathbb{C}^n$ are complex Gaussian random vectors, Gao and Xu [17] develop a spectral method based on exponential function, and prove that with probability at least $1 - \exp(-cn)$ the spectral initialization can give an initial guess $z_0$ satisfying $\text{dist}(z_0, x) \leq \epsilon\|x\|_2$ for any fixed $\epsilon$, provided $m \geq Cn$ for a positive constant $C$. We refer the reader to [9, 39, 42] for others spectral initialization and [25, 27] for the optimal design of a spectral initialization.

1.5. Notations. Throughout this paper, we assume the measurements $a_j \in \mathbb{C}^n$, $j = 1, \ldots, m$ are drawn independently and uniformly from the complex unit sphere. We say $\xi \in \mathbb{C}^n$ is a complex Gaussian random vector if $\xi \sim 1/\sqrt{2} \cdot \mathcal{N}(0, I_n) + i/\sqrt{2} \cdot \mathcal{N}(0, I_n)$. We write $z \in S^{n-1}_\mathbb{C}$ if $z \in \mathbb{C}^n$ and $\|z\|_2 = 1$. Let $\Re(z) \in \mathbb{R}$ and $\Im(z) \in \mathbb{R}$ denote the real and imaginary part of a complex number $z \in \mathbb{C}$. For any $A, B \in \mathbb{R}$, we use $A \lesssim B$ to denote $A \leq C_0B$ where $C_0 \in \mathbb{R}_+$ is an
absolute constant. The notion $\gtrsim$ can be defined similarly. In this paper, we use $C, c$ and the subscript (superscript) form of them to denote universal constants whose values vary with the context.

1.6. **Organization.** The paper is organized as follows. In Section 2, we introduce some notations and definitions that will be used in our paper. In particular, the restricted strong convexity condition plays a key role in the proof of main result. In Section 3, we first show that under the restricted strong convexity condition, a convergence result for a single step can be established, and then we show the main result can be proved by using the tools from stochastic process. In Section 4, we demonstrate that the random measurements drawn independently and uniformly from the complex unit sphere satisfies restricted strong convexity condition with high probability. A brief discussion is presented in Section 5. Appendix collects the technical lemmas needed in the proofs.

2. **Preliminaries**

The aim of this section is to introduce some definitions that will be used in our paper. Let $x \in \mathbb{C}^n$ be the target signal we want to recover. The measurements we obtain are

\begin{equation}
    b_j = |\langle a_j, x \rangle|, \quad j = 1, \ldots, m,
\end{equation}

where $a_j \in \mathbb{C}^n$ are measurement vectors. In this paper, we assume without loss of generality that $a_j \in S_{\mathbb{C}}^{n-1}$ for all $j = 1, \ldots, m$. For the recovery of $x$ we consider the randomized Kaczmarz method given by

\begin{equation}
    z_{k+1} = z_k - \left( 1 - \frac{b_{i_k}}{|a_{i_k}^* z_k|} \right) a_{i_k} a_{i_k}^* z_k,
\end{equation}

where $i_k$ is chosen uniformly from the $\{1, \ldots, m\}$ at random at the $k$-th iteration.

Obviously, for any $z$ if $z$ is a solution to (4) then $ze^{i\phi}$ is also a solution to it for any $\phi \in \mathbb{R}$. Thus the set of solutions to (4) is $\{xe^{i\phi} : \phi \in \mathbb{R}\}$, which is a one-dimensional circle in $\mathbb{C}^n$. For this reason, we define the distance between $z$ and $x$ as

\[ \text{dist}(z, x) = \min_{\phi \in \mathbb{R}} \|z - xe^{i\phi}\|_2. \]

For convenience, we also define the phase $\phi(z)$ as

\begin{equation}
    \phi(z) := \arg\min_{\phi \in \mathbb{R}} \|z - xe^{i\phi}\|_2
\end{equation}
for any $z \in \mathbb{C}^n$. Moreover, for any $\epsilon \geq 0$ we define the $\epsilon$-neighborhood of $x$ as

$$E(\epsilon) := \{ z \in \mathbb{C}^n : \text{dist}(z, x) \leq \epsilon \}.$$  

The following auxiliary loss function plays a key role in the proof of main result:

$$f(z) = \frac{1}{m} \sum_{j=1}^{m} (|a_j^*z| - |a_j^*x|)^2.$$  

Since it is not differentiable, we shall need the directional derivative. For any vector $v \neq 0$ in $\mathbb{C}^n$, the one-sided directional derivative of $f$ at $z$ along the direction $v$ is given by

$$D_v f(z) := \lim_{t \to 0^+} \frac{f(z + tv) - f(z)}{t}$$

if the limit exists. It is not difficult to compute that the one-sided directional derivative of $f$ in (8) along any direction $v$ is

$$D_v f(z) = \frac{2}{m} \sum_{j=1}^{m} \left( 1 - \frac{|a_j^*x|}{|a_j^*z|} \right) \Re(a_j^*vz^*a_i).$$

Finally, we need the assumption that $f$ satisfies a local restricted strong convexity on $E(\epsilon)$, which essentially states that the function is well behaved along the line connecting the current point to its nearest global solution. Here, $E(\epsilon)$ and $f$ are defined in (7) and (8), respectively.

**Definition 2.1 (Restricted Strong Convexity).** The function $f$ is said to obey the restricted strong convexity $\text{RSC}(\gamma, \epsilon)$ for some $\gamma, \epsilon > 0$ if

$$D_{z - xe^{i\phi(z)}} f(z) \geq \gamma \| z - xe^{i\phi(z)} \|^2_2 + f(z)$$

for all $z \in E(\epsilon)$.

3. **Proof of The Main Result**

In this section we present the detailed proof of the main result. We first prove that under the assumption of $f$ satisfying restricted strong convexity, a bound for the expected decrement in distance to the solution set can be established for the randomized Kaczmaz scheme in a single step. Next, we show that for random measurements $a_j \in \mathbb{C}^n, j = 1, \ldots, m$ which are drawn independently and uniformly from the complex unit sphere, the function $f$ defined in (8) satisfies the restricted strong convexity with high probability, provided $m \geq Cn$ for some constant $C > 0$. Finally, using the tools from stochastic process, we could prove the randomized Kaczmaz method is linearly convergent in expectation, which concludes the proof of the main result.
Theorem 3.1. Assume that $f$ defined in (8) satisfies the restricted strong convexity $\text{RSC} (\gamma, \epsilon)$. Then the iteration $z_{k+1}$ given by randomized Kaczmarz update rule (5) obeys

$$
\mathbb{E}_{i_k} \left[ \text{dist}^2(z_{k+1}, x) \right] \leq (1 - \gamma) \text{dist}^2(z_k, x)
$$

for all $z_k$ satisfying $\text{dist}(z_k, x) \leq \epsilon \|x\|_2$. Here, $\mathbb{E}_{i_k}$ denotes the expectation with respect to randomness of $i_k$ at iteration $k$.

Proof. Recognize that $\|a_{i_k}\|_2 = 1$. Using the restricted strong convexity condition $\text{RSC} (\gamma, \epsilon)$, we have

$$
\mathbb{E}_{i_k} \text{dist}^2(z_{k+1}, x) = \mathbb{E}_{i_k} \|z_{k+1} - x e^{i\phi(z_{k+1})}\|^2_2 \\
\leq \mathbb{E}_{i_k} \left\| z_k - x e^{i\phi(z_k)} - \left(1 - \frac{y_{i_k}}{\|a_{i_k}^* x_k\|} \right) a_{i_k}^* z_k a_{i_k}\right\|^2_2 \\
= \|z_k - x e^{i\phi(z_k)}\|^2_2 + \mathbb{E}_{i_k} \left(1 - \frac{y_{i_k}}{\|a_{i_k}^* z_k\|} \right)^2 \|a_{i_k}^* z_k\|^2 \\
- 2\mathbb{E}_{i_k} \mathbb{R} \left(1 - \frac{y_{i_k}}{\|a_{i_k}^* z_k\|} \right) a_{i_k}^* a_{i_k}^* (z_k - x e^{i\phi(z_k)}) \\
= \|z_k - x e^{i\phi(z_k)}\|^2_2 + f(z_k) - D_{z_k - x e^{i\phi(z_k)}} f(z_k) \\
\leq (1 - \gamma) \|z_k - x e^{i\phi(z_k)}\|^2_2,
$$

where the third equation follows from the expression of directional derivative as shown in (9). This completes the proof. \qed

Theorem 3.2. Assume the measurement vectors $a_1, \ldots, a_m \in \mathbb{C}^n$ are drawn uniformly from the unit sphere $S^{n-1}_\mathbb{C}$. Suppose that $m \geq C_0 n$ and $f$ is defined in (8). Then $f$ satisfies the restricted strong convexity $\text{RSC}(0.03/n, 0.01)$ with probability at least $1 - 14 \exp(-c_0 n)$, where $C_0, c_0$ are universal positive constants.

Proof. The proof of this theorem is deferred to Section 4. \qed

Based on Theorem 3.1 and Theorem 3.2, we obtain that if $m \geq C_0 n$ for some universal constant $C_0 > 0$, then with probability at least $1 - 14 \exp(-c_0 n)$, the $(k+1)$-th iteration obeys

$$
\mathbb{E}_{i_{k+1}} \left[ \text{dist}^2(z_{k+1}, x) \right] \leq (1 - 0.03/n) \text{dist}^2(z_k, x),
$$
provided \( \text{dist}(z_k, x) \leq 0.01\|x\|_2 \) at \( k \) step. To be able to iterate this result recursively we need the condition \( \text{dist}(z_k, x) \leq 0.01\|x\|_2 \) holds for all \( k \), however, it does not hold arbitrarily. Hence, we introduce a stopping time

\[
\tau := \min \{ k : z_k \notin B \},
\]

where \( B := \{ z : \text{dist}(z, x) \leq 0.01\|x\|_2 \} \). With this in place, we can give the proof of Theorem 1.1. We restate our main result here for convenience.

**Theorem 3.3.** Suppose \( m \geq C_0 n \) for some universal constant \( C_0 > 0 \). Assume the measurement vectors \( a_1, \ldots, a_m \in \mathbb{C}^n \) are drawn independently and uniformly from the unit sphere \( S_{\mathbb{C}}^{n-1} \). For any \( 0 < \delta < 1 \), let \( z_0 \) be an initial estimate to \( x \) such that \( \|z_0 - x\|_2 \leq 0.01\delta\|x\|_2 \). Let \( \tau \) be the stopping time defined in (10). Then with probability at least \( 1 - 14\exp(-C_0 n) \) it holds: the iteration \( z_k \) given by randomized Kaczmarz update rule (5) obeys

\[
\mathbb{E}_{\mathcal{I}^k} [\text{dist}(z_k, x) 1_{\tau = \infty}] \leq (1 - 0.03/n)^k/2 \text{dist}(z_0, x).
\]

Furthermore, the probability \( \mathbb{P}(\tau < \infty) \leq \delta^2 \). Here \( \mathbb{E}_{\mathcal{I}^k} \) denotes the expectation with respect to randomness \( \mathcal{I}^k := \{i_1, i_2, \ldots, i_k\} \) conditioned on the high probability event of random measurements \( \{a_j\}_{j=1}^m \) and \( C_0 > 0 \) is a universal constant.

**Proof.** From Theorem 3.1 and Theorem 3.2, we obtain that if \( m \geq C_0 n \) then with probability at least \( 1 - 14\exp(-C_0 n) \) it holds

\[
\mathbb{E}_{\mathcal{I}^k+1} \left[ \text{dist}^2(z_{k+1}, x) 1_{\tau > k+1} \mid z_k \in B \right] \leq \mathbb{E}_{\mathcal{I}^k+1} \left[ \text{dist}^2(z_{k+1}, x) 1_{\tau > k} \mid z_k \in B \right] = \mathbb{E}_{\mathcal{I}^k+1} \left[ \text{dist}^2(z_{k+1}, x) \mid z_k \in B \right] 1_{\tau > k} \leq (1 - 0.03/n) \text{dist}^2(z_k, x) 1_{\tau > k}.
\]

Note that \( z_k \in B \) is an event with respect to randomness \( \mathcal{I}^k \). Taking expectation gives

\[
\mathbb{E}_{\mathcal{I}^k+1} \left[ \text{dist}^2(z_{k+1}, x) 1_{\tau > k+1} \right] = \mathbb{E}_{\mathcal{I}^k} \left[ \mathbb{E}_{i_{k+1}} \left[ \text{dist}^2(z_{k+1}, x) 1_{\tau > k+1} \mid z_k \in B \right] \right] \leq (1 - 0.03/n) \mathbb{E}_{\mathcal{I}^k} \left[ \text{dist}^2(z_k, x) 1_{\tau > k} \right].
\]

By induction, we arrive at the first part of the conclusion.

For the second part, define \( Y_k := \|z_k \land x\|_2^2 \). Using the similar idea of Theorem 3.1 in [34], we can check that \( Y_k \) is a non-negative supermartingale. It then follows from the supermartingale
maximum inequality that
\[
\mathbb{P}\left( \sup_{1 \leq k < \infty} Y_k \geq 0.01^2 \|x\|_2^2 \right) \leq \frac{\mathbb{E}Y_0}{0.01^2 \|x\|_2^2} \leq \delta^2.
\]
This completes the proof. \(\square\)

4. Proof of Theorem 3.2

**Proof of Theorem 3.2.** For any \(z \in \mathbb{C}^n\), set \(h = e^{-i\phi(z)}z - x\) where \(\phi(z)\) is defined in (6). It is easy to check that the function \(f\) given in (8) can be rewritten as
\[
f(z) = \frac{1}{m} \sum_{j=1}^{m} \left( |a_j^* h|^2 + 2\Re(h^* a_j a_j^* x) + 2 |a_j^* x|^2 - 2 |a_j^* z| |a_j^* x| \right).
\]
To show that the function \(f\) satisfies the restricted strong convexity, from the definition, it suffices to give a lower bound for \(D_{z-xe^{i\phi(z)}} f(z) - f(z)\). Note that
\[
|a_j^* z|^2 = |a_j^* x|^2 + 2\Re(h^* a_j a_j^* x) + |a_j^* h|^2.
\]
By some algebraic computation, we immediately have
\[
D_{z-xe^{i\phi(z)}} f(z) - f(z)
\]
\[
= \frac{2}{m} \sum_{j=1}^{m} \left( 1 - \frac{|a_j^* x|}{|a_j^* z|} \right) \left( |a_j^* h|^2 + \Re(h^* a_j a_j^* x) \right) \frac{1}{m} \sum_{j=1}^{m} \left( |a_j^* z| - |a_j^* x| \right)^2
\]
\[
= \frac{1}{m} \sum_{j=1}^{m} |a_j^* h|^2 + \frac{2}{m} \sum_{j=1}^{m} \left( |a_j^* z| |a_j^* x| - |a_j^* x|^2 - \frac{|a_j^* x| |a_j^* h|^2}{|a_j^* z|} \right)
\]
\[
= \frac{1}{m} \sum_{j=1}^{m} |a_j^* h|^2 + \frac{2}{m} \sum_{j=1}^{m} \frac{|a_j^* z| ^3 - |a_j^* z| |a_j^* x|^2 + |a_j^* x| |\Re(h^* a_j a_j^* x)|}{|a_j^* z|}
\]
\[
= \frac{1}{m} \sum_{j=1}^{m} |a_j^* h|^2 + \frac{2}{m} \sum_{j=1}^{m} \frac{|a_j^* z| |a_j^* x| |\Re(h^* a_j a_j^* x)| - |a_j^* x|^2 |\Re(h^* a_j a_j^* x)| - |a_j^* x|^2 |a_j^* h|^2}{|a_j^* z| ^2 (|a_j^* z| + |a_j^* x|)}
\]
\[
= \frac{4}{m} \sum_{j=1}^{m} \frac{|a_j^* x|^2 |a_j^* h|^2}{|a_j^* z|^2 (|a_j^* z| + |a_j^* x|)^2}.
\]
(11)

We next divide the indexes into two groups: \(j \in I_\alpha\) and \(j \in I_\alpha^c\), where \(I_\alpha := \left\{ j : |a_j^* x| \geq \alpha |a_j^* h| \right\}\) for some fixed parameter \(\alpha > 0\). For convenience, we denote \(D_{z-xe^{i\phi(z)}} f(z) - f(z) := \frac{1}{m} \sum_{j=1}^{m} T_j\).
We claim that for any \( \alpha > 1 \) it holds
\[
T_j \geq \frac{4\alpha^3}{(\alpha + 1)(2\alpha + 1)^2} \cdot \frac{\Re^2(h^*a_j\bar{a}_j^*x)}{|a_j^*x|^2} - \frac{8\alpha^2 - 5\alpha + 1}{(\alpha - 1)(2\alpha - 1)^2} \cdot |a_j^*h|^2 \quad \text{for } j \in I_\alpha
\]
and
\[
T_j \geq -3|a_j^*h|^2 \quad \text{for } j \in I_\alpha^c.
\]
This taken collectively with the identity (11) leads to a lower estimate
\[
D_{z-xe^{i\phi(z)}}f(z) - f(z) \geq \frac{4\alpha^3}{(\alpha + 1)(2\alpha + 1)^2} \cdot \frac{1}{m} \sum_{j \in I_\alpha} \frac{\Re^2(h^*a_j\bar{a}_j^*x)}{|a_j^*x|^2}
-
\frac{8\alpha^2 - 5\alpha + 1}{(\alpha - 1)(2\alpha - 1)^2} \cdot \frac{1}{m} \sum_{j \in I_\alpha} |a_j^*h|^2 - \frac{3}{m} \sum_{j \in I_\alpha^c} |a_j^*h|^2,
\]
leaving us with three quantities in the right-hand side to deal with. Let \( \rho := \|h\|_2 \). From the definition of \( h \), it is easy to check \( \Im(h^*x) = 0 \). According to Lemma 6.2, we immediately obtain that for any \( 0 < \delta \leq 1 \) there exist universal constants \( C, c > 0 \) such that if \( \alpha \rho \leq 1/3 \) and \( m \geq C\delta^{-2}\log(1/\delta)n \) then with probability at least \( 1 - 6\exp(-c\delta^2 n) \) it holds
\[
\frac{1}{m} \sum_{j \in I_\alpha} \frac{\Re^2(h^*a_j\bar{a}_j^*x)}{|a_j^*x|^2} \geq \frac{1}{n} \cdot \left( \frac{3}{8} - \frac{\alpha^2 \rho^2}{(0.99 + \alpha \rho)^2} - \delta \right) \|h\|_2^2.
\]
For the second term, it follows from Lemma 6.1 that for \( m \geq C\delta^{-2}n \), with probability at least \( 1 - 2\exp(-c\delta^2 n) \),
\[
\frac{1}{m} \sum_{j=1}^m |a_j^*h|^2 \leq \frac{1 + \delta}{n} \|h\|_2^2.
\]
Finally, for the third term, applying Lemma 6.3, we have that when \( m \geq C\delta^{-2}\log(1/\delta)n \) and \( 0 < \alpha \rho \leq 0.4 \), with probability at least \( 1 - 6\exp(-c\delta^2 n) \),
\[
\frac{1}{m} \sum_{j \in I_\alpha^c} |a_j^*h|^2 \leq \frac{1}{n} \cdot \left( \frac{2\alpha^2 \rho^2}{0.99 + \alpha^2 \rho^2} + \delta \right) \|h\|_2^2.
\]
Setting \( \alpha := 12 \), \( \delta := 0.001 \) and putting (15), (16), (17) into (14), we obtain the conclusion that with probability at least \( 1 - 14\exp(-c_0n) \) it holds
\[
D_{z-xe^{i\phi(z)}}f(z) - f(z) \geq \frac{0.03}{n} \|h\|_2^2 \quad \text{for all } \quad \|h\|_2 \leq 0.01,
\]
provided \( m \geq C_0n \). Here, \( C_0, c_0 \) are universal positive constants.
It remains to prove the claims. We first consider the case where $j \in I_\alpha$. It follows from (11) that

$$T_j = |a_j^* h|^2 - \frac{2|a_j^* x|^2 |a_j^* h|^2}{|a_j^* z|(|a_j^* z| + |a_j^* x|)} + \frac{2|a_j^* x| |a_j^* h|^2 \Re(h^* a_j a_j^* x)}{|a_j^* z|(|a_j^* z| + |a_j^* x|)^2} + \frac{4|a_j^* x| |\Re^2(h^* a_j a_j^* x)}{|a_j^* z|(|a_j^* z| + |a_j^* x|)^2}. $$

From the definition of $I_\alpha$, it is easy to see that when $j \in I_\alpha$ we have

$$1 - 1/\alpha |a_j^* x| \leq |a_j^* x| - |a_j^* h| \leq |a_j^* x| \leq |a_j^* x| + |a_j^* h| \leq (1 + 1/\alpha) |a_j^* x|. $$

Thus the second term of $T_j$ obeys

$$\frac{|a_j^* x|^2 |a_j^* h|^2}{|a_j^* z|(|a_j^* z| + |a_j^* x|)} \leq \frac{\alpha^2}{(\alpha - 1)(2\alpha - 1)} |a_j^* h|^2. $$

Similarly, the third term of $T_j$ satisfies

$$\frac{|a_j^* x| |a_j^* h|^2 |\Re(h^* a_j a_j^* x)|}{|a_j^* z|(|a_j^* z| + |a_j^* x|)^2} \leq \frac{|a_j^* x|^2 |a_j^* h|^3}{|a_j^* z|(|a_j^* z| + |a_j^* x|)^2} \leq \frac{\alpha^2}{(\alpha - 1)(2\alpha - 1)^2} |a_j^* h|^2. $$

Finally, using the upper bound in (18), we have

$$\frac{|a_j^* x| |\Re^2(h^* a_j a_j^* x)}{|a_j^* z|(|a_j^* z| + |a_j^* x|)^2} \geq \frac{\alpha^3}{(\alpha + 1)(2\alpha + 1)^2} \frac{\Re^2(h^* a_j a_j^* x)}{|a_j^* x|^2}. $$

Collecting the above three estimators, we have

$$T_j \geq \frac{4\alpha^3}{(\alpha + 1)(2\alpha + 1)^2} \frac{\Re^2(h^* a_j a_j^* x)}{|a_j^* x|^2} - \frac{8\alpha^2 - 5\alpha + 1}{(\alpha - 1)(2\alpha - 1)^2} |a_j^* h|^2, $$

which proves the claim (12).

We next turn to consider the case where $j \notin I_\alpha$. From the definition, we know $T_j$ can be denoted as

$$T_j = 2 \left(1 - \frac{|a_j^* x|}{|a_j^* z|}\right) \Re(e^{-i\phi} h^* a_j a_j^* z) - (|a_j^* z| - |a_j^* x|)^2 \quad \text{for all } j. $$

It then immediately gives

$$|T_j| \leq \frac{2|a_j^* z| - |a_j^* x|}{|a_j^* z|} \cdot |h^* a_j a_j^* z| + (|a_j^* z| - |a_j^* x|)^2 \leq 3 |a_j^* h|^2, $$

where we use the Cauchy-Schwarz inequality and the fact that $|a_j^* z| - |a_j^* x|$ $\leq |a_j^* h|$ in the last inequality. This completes the claim (13). \[\square\]
5. Discussions

This paper considers the convergence of randomized Kaczmarz method for phase retrieval in the complex setting. A linear convergence rate has been established by combining restricted strong convexity condition and tools from stochastic process, which gives a positive answer for the conjecture given in [34, Section 7.2].

There are some interesting problems for future research. First, it has been shown numerically that randomized Kaczmarz method is also efficient for solving Fourier phase retrieval problem, at least when the measurements follow the coded diffraction pattern (CDP) model, it is of practical interest to provide some theoretical guarantees for it. Second, the convergence of randomized Kaczmarz method relies on a spectral initialization. Some numerical evidence have shown that randomized Kaczmarz method works well even if we start from an arbitrary initialization. It is interesting to provide some theoretical justifications for it.

6. Appendix

Lemma 6.1. Suppose that the vectors \(a_1, \ldots, a_m \in \mathbb{C}^n\) are drawn uniformly from the unit sphere \(S^{n-1}_\mathbb{C}\). For any \(0 < \delta \leq 1\), if \(m \geq C\delta^{-2}n\) then with probability at least \(1 - 2\exp(-c\delta^2m)\) it holds

\[
\left\| \frac{1}{m} \sum_{j=1}^{m} a_j a_j^* - \frac{1}{n} \cdot I \right\|_2 \leq \frac{\delta}{n}.
\]

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

Proof. Assume that \(\mathcal{N}\) is an \(1/4\)-net of the complex unit sphere \(S^{n-1}_\mathbb{C} \subset \mathbb{C}^n\). It then follows from [35, Lemma 4.4.3] that

\[
\left\| \frac{1}{m} \sum_{j=1}^{m} a_j a_j^* - \frac{1}{n} \cdot I \right\|_2 \leq 2\max_{h \in \mathcal{N}} \left| \frac{1}{m} \sum_{j=1}^{m} |a_j^* h|^2 - \frac{1}{n} \right|.
\]

Here, the cardinality \(|\mathcal{N}| \leq 9^{2n}\). Due to the unitary invariance of \(a_j\), for any fixed \(h \in S^{n-1}_\mathbb{C}\), we have

\[
\mathbb{E}|a_j^* h|^2 = \mathbb{E}|a_j^* e_1|^2 = \mathbb{E}|a_j^* e_2|^2 = \cdots = \mathbb{E}|a_j^* e_n|^2 = \frac{1}{n},
\]

\[
\mathbb{E}\|a_j\|^2 = \frac{1}{n}.
\]

It means that for any fixed \(h \in S^{n-1}_\mathbb{C} \subset \mathbb{C}^n\), the terms \(|a_j^* h|^2 - 1/n\) are independent, mean zero, sub-exponential random variables with sub-exponential norm bounded by \(K = c_1/n\) for some universal
constant $c_1 > 0$ [35, Theorem 3.4.6]. Using Bernstein’s inequality, we obtain that for any $0 < \delta \leq 1$ with probability at least $1 - 2\exp(-c_2\delta^2m)$,

$$\left| \frac{1}{m} \sum_{j=1}^{m} |a_j^* h|^2 - \frac{1}{n} \right| \leq \frac{\delta}{2n}$$

holds for some positive constant $c_2$. Taking the union bound over $\mathcal{N}$, we obtain that

$$\left\| \frac{1}{m} \sum_{j=1}^{m} a_j a_j^* - \frac{1}{n} \mathbf{I} \right\|_2 \leq \frac{\delta}{n}$$

holds with probability at least

$$1 - 2\exp(-c_2\delta^2m) \cdot 9^{2n} \geq 1 - 2\exp(-c\delta^2m),$$

provided $m \geq C\delta^{-2}n$ for some constants $C, c > 0$. This completes the proof.

\[ \square \]

**Lemma 6.2.** Let $x$ be a vector in $\mathbb{C}^n$ with $\|x\|_2 = 1$ and $\lambda \geq 3$. Assume that the vectors $a_1, \ldots, a_m \in \mathbb{C}^n$ are drawn uniformly from the unit sphere $S_{\mathbb{C}}^{n-1}$. For any fixed $0 < \delta \leq 1$, there exist universal constants $C, c > 0$ such that if $m \geq C\delta^{-2} \log(1/\delta)n$ then with probability at least $1 - 6\exp(-c\delta^2n)$ it holds

$$\frac{1}{m} \sum_{j=1}^{m} \Re^2(h^* a_j a_j^* x) \mathbf{1}_{\{\lambda |a_j^* x| \geq |a_j^* h|\}} \geq \frac{1}{n} \left( \frac{3}{8} - \frac{1}{(1 + 0.99\lambda)^2} - \frac{\delta}{2} \right)$$

for all $h \in \mathbb{C}^n$ with $\|h\|_2 = 1$ and $\Im(h^* x) = 0$.

**Proof.** We first prove the result for any fixed $h$ and then apply an $\varepsilon$-net argument to develop a uniform bound for it. To begin with, we introduce a series of auxiliary random Lipschitz functions to approximate the indicator functions. For any $j = 1, \ldots, m$, define

$$\chi_j(t) := \begin{cases} 1, & \text{if } t \leq 0.99\lambda |a_j^* x|; \\ \frac{100}{\lambda |a_j^* x|^2} t + 100, & \text{if } 0.99\lambda |a_j^* x| \leq t \leq \lambda |a_j^* x|; \\ 0, & \text{otherwise}. \end{cases}$$

It then gives

$$\frac{\Re^2(h^* a_j a_j^* x)}{|a_j^* x|^2} \mathbf{1}_{\{\lambda |a_j^* x| \geq |a_j^* h|\}} \geq \frac{\Re^2(h^* a_j a_j^* x)}{|a_j^* x|^2} \chi_j(|a_j^* h|) \geq \frac{\Re^2(h^* a_j a_j^* x)}{|a_j^* x|^2} \mathbf{1}_{\{0.99\lambda |a_j^* x| \geq |a_j^* h|\}}.$$

For any fixed $h$, since $a_1, \ldots, a_m$ are random vectors uniformly distributed on the unit sphere, it means that the terms $\frac{\Re^2(h^* a_j a_j^* x)}{|a_j^* x|^2} \chi_j(|a_j^* h|)$ are independent sub-exponential random variables with
the maximal sub-exponential norm $K = c_1/n$ for some universal constant $c_1 > 0$ [35, Theorem 3.4.6]. Apply Bernstein’s inequality gives that for any fixed $0 < \delta \leq 1$ the following holds

$$\frac{1}{m} \sum_{j=1}^{m} \frac{\Re^2(h^* a_j a_j^* x)}{|a_j^* x|^2} \chi_j(|a_j^* h|) \geq \mathbb{E} \left( \frac{\Re^2(h^* a a^* x)}{|a^* x|^2} \chi_j(|a^* h|) \right) - \frac{\delta}{4n}$$

with probability at least $1 - 2 \exp(-c_2\delta^2 m)$, where $c_2$ is a universal positive constant. Here, $a \in \mathbb{C}^n$ is a vector uniformly distributed on the unit sphere.

Next, we give a uniform bound for the estimate (20). Construct an $\epsilon$-net $\mathcal{N}$ over the unit sphere in $\mathbb{C}^n$ with cardinality $|\mathcal{N}| \leq (1 + \frac{2}{\epsilon})^{2n}$. Then we have

$$\frac{1}{m} \sum_{j=1}^{m} \frac{\Re^2(h^* a_j a_j^* x)}{|a_j^* x|^2} \chi_j(|a_j^* h|) \geq \mathbb{E} \left( \frac{\Re^2(h^* a a^* x)}{|a^* x|^2} \chi_j(|a^* h|) \right) - \frac{\delta}{4n} \quad \text{for all} \quad h \in \mathcal{N}$$

with probability at least

$$1 - 2 \exp(-c_2\delta^2 m) \cdot (1 + \frac{2}{\epsilon})^{2n}.$$

For any $h$ with $\|h\|_2 = 1$, there exists a $h_0 \in \mathcal{N}$ such that $\|h - h_0\|_2 \leq \epsilon$. We claim that there exist universal constants $C', c_3 > 0$ such that if $m \geq C'n$ then with probability at least $1 - 2 \exp(-c_3m)$ it holds

$$\left| \frac{1}{m} \sum_{j=1}^{m} \frac{\Re^2(h^* a_j a_j^* x)}{|a_j^* x|^2} \chi_j(|a_j^* h|) - \frac{1}{m} \sum_{j=1}^{m} \frac{\Re^2(h_0^* a_j a_j^* x)}{|a_j^* x|^2} \chi_j(|a_j^* h_0|) \right| \leq \frac{205\epsilon}{n}.$$  

Choosing $\epsilon := \delta/820$, we then obtain that

$$\frac{1}{m} \sum_{j=1}^{m} \frac{\Re^2(h^* a_j a_j^* x)}{|a_j^* x|^2} \chi_j(|a_j^* h|) \geq \mathbb{E} \left( \frac{\Re^2(h^* a a^* x)}{|a^* x|^2} \chi_j(|a^* h|) \right) - \frac{\delta}{2n} \quad \text{for all} \quad \|h\|_2 = 1$$

holds with probability at least

$$1 - 2 \exp(-c_3m) - 2 \exp(-c_2\delta^2 m)(1 + \frac{2}{\epsilon})^{2n} \geq 1 - 4 \exp(-c_4\delta^2 m),$$

provided $m \geq C\log(1/\delta)\delta^{-2}n$ for some positive constant $C$. Here $c_4$ is a universal positive constant. To give a lower bound for the expectation in (22), recognize that if $\xi \in \mathbb{C}^n$ is a complex Gaussian random vector then $\xi/\|\xi\|_2$ is a vector uniformly distributed on the unit sphere. Since $\|\xi\|_2 \leq (1 + \delta_1)\sqrt{n}$ holds for any fixed $0 < \delta_1 \leq 1$ [35, Theorem 3.1.1] with probability at least
1 - 2 \exp(-c_5\delta_1^2 n) for some universal constant \(c_5 > 0\), it then follows from Lemma 6.4 that

\[
\mathbb{E}\left(\frac{\|h^*a^*x\|_2^2}{\|a^*x\|^2} \cdot 1_{\{\|a^*x\| \geq \|a^*h\|\}}\right) = \mathbb{E}\left(\frac{\|h^*\xi^* x\|_2^2}{\|\xi\|^2} \cdot 1_{\{\|\xi^* x\| \geq \|\xi^* h\|\}}\right)
\geq \frac{1}{(1 + 3\delta_1)n} \cdot \mathbb{E}\left(\frac{\|h^*\xi^* x\|_2^2}{\|\xi^* x\|^2} \cdot 1_{\{\|\xi^* x\| \geq \|\xi^* h\|\}}\right)
\geq \frac{1}{(1 + 3\delta_1)n} \cdot \left(\frac{3}{8} - \frac{1}{(\lambda + 1)^2}\right).
\]

Taking \(\delta_1 := \delta/2\), we obtain that for any \(\lambda \geq 2.95\) with probability at least \(1 - 2 \exp(-c_6\delta^2 n)\) it holds

\[
\mathbb{E}\left(\frac{\|h^*a^*a^*x\|_2^2}{\|a^*x\|^2} \cdot 1_{\{\|a^*x\| \geq \|a^*h\|\}}\right) \geq \frac{1}{n} \cdot \left(\frac{3}{8} - \frac{1}{(1 + 0.99\lambda)^2}\right),
\]

where \(c_6 > 0\) is a universal constant. Collecting (19), (22) and (23) together, we obtain the conclusion that for any \(\lambda \geq 3\) with probability at least \(1 - 6 \exp(-c\delta^2 n)\) it holds

\[
\mathbb{E}\left(\frac{\|h^*a^*j^*x\|_2^2}{\|a^*j^*x\|^2} \cdot 1_{\{\|a^*j^*x\| \geq \|a^*j^*h\|\}}\right) \geq \frac{1}{n} \cdot \left(\frac{3}{8} - \frac{1}{(1 + 0.99\lambda)^2}\right),
\]

provided \(m \geq C \log(1/\delta)\delta^{-2}n\). Here, \(c\) is a universal positive constant.

Finally, it remains to prove the claim (21). To this end, we claim that for all \(j = 1, \ldots, m\) it holds

\[
\left|\frac{\mathbb{E}(h^*a^*_j a^*_j x)}{\|a^*_j x\|^2} - \frac{\mathbb{E}(h^*_0 a^*_j a^*_j x)}{\|a^*_j x\|^2} - \chi_j(|a^*_j h|)\right| \leq 101|a^*_j h||a^*_j (h - h_0)| + 101|a^*_j h_0||a^*_j (h - h_0)|.
\]

Indeed, from the definition of \(\chi_j(t)\), if both \(|a^*_j h| > \lambda|a^*_j x|\) and \(|a^*_j h_0| > \lambda|a^*_j x|\) then the above inequality holds directly. Thus, we only need to consider the case where \(|a^*_j h| \leq \lambda|a^*_j x|\) or \(|a^*_j h_0| \leq \lambda|a^*_j x|\). Without loss of generality, we assume \(|a^*_j h| \leq \lambda|a^*_j x|\). Then we have

\[
\left|\frac{\mathbb{E}(h^*a^*_j a^*_j x)}{\|a^*_j x\|^2} - \frac{\mathbb{E}(h^*_0 a^*_j a^*_j x)}{\|a^*_j x\|^2} - \chi_j(|a^*_j h|)\right| \leq \frac{100}{\lambda|a^*_j x|} (|a^*_j h|^2 \chi_j(|a^*_j h|) - \chi_j(|a^*_j h_0|) + (|a^*_j h| + |a^*_j h_0|) |a^*_j (h - h_0)|)
\leq 101|a^*_j h||a^*_j (h - h_0)| + |a^*_j h_0||a^*_j (h - h_0)|.
\]
which gives (24). According to Lemma 6.1, we obtain that for $m \geq C'n$ with probability at least $1 - 2 \exp(-c_3n)$ it holds
\[
\left| \frac{1}{m} \sum_{j=1}^{m} \Re^2(h^*a_ja_j^*x) \chi_j(|a_j^*h|) - \frac{1}{m} \sum_{j=1}^{m} \Re^2(h_0^*a_ja_j^*x) \chi_j(|a_j^*h_0|) \right| 
\leq \frac{101}{m} \sum_{j=1}^{m} |a_j^*h||a_j^*(h-h_0)| + \frac{101}{m} \sum_{j=1}^{m} |a_j^*h_0||a_j^*(h-h_0)| 
\leq 101 \left[ \frac{1}{m} \sum_{j=1}^{m} |a_j^*h|^2 \right] \left[ \frac{1}{m} \sum_{j=1}^{m} |a_j^*(h-h_0)|^2 \right] + 101 \left[ \frac{1}{m} \sum_{j=1}^{m} |a_j^*h_0|^2 \right] \left[ \frac{1}{m} \sum_{j=1}^{m} |a_j^*(h-h_0)|^2 \right] 
\leq \frac{205\epsilon}{n},
\]
which proves the claim (21).

Lemma 6.3. Let $x$ be a vector in $\mathbb{C}^n$ with $\|x\|_2 = 1$ and $0 < \lambda \leq 0.4$. Assume that the vectors $a_1, \ldots, a_m \in \mathbb{C}^n$ are drawn uniformly from the unit sphere $\mathbb{S}^{n-1}_{\mathbb{C}}$. For any fixed $0 < \delta \leq 1$, there exist universal constants $C, c > 0$ such that for $m \geq C\delta^{-2} \log(1/\delta)n$, with probability at least $1 - 6 \exp(-c_3\delta^2 n)$, it holds
\[
\frac{1}{m} \sum_{j=1}^{m} |a_j^*h|^2 \cdot 1\{|a_j^*x|\leq \lambda |a_j^*h|\} \leq \frac{2\lambda^2}{(\lambda^2 + 0.99)n} + \frac{\delta}{n}
\]
for all $h \in \mathbb{C}^n$ with $\|h\|_2 = 1$ and $\Re(h^*x) = 0$.

Proof. Due to the non-Lipschitz of indicator functions, we introduce a series of auxiliary random Lipschitz functions to approximate them. For any $j = 1, \ldots, m$, define
\[
\chi_j(t) := \begin{cases} 
t, & \text{if } t \geq |a_j^*x|^2/\lambda^2; \\
100t - \frac{99|a_j^*x|^2}{\lambda^2}, & \text{if } 0.99|a_j^*x|^2/\lambda^2 \leq t \leq |a_j^*x|^2/\lambda^2; \\
0, & \text{otherwise.}
\end{cases}
\]

Then it is easy to check that
\[
(25) \quad \frac{1}{m} \sum_{j=1}^{m} |a_j^*h|^2 \cdot 1\{|a_j^*x|\leq \lambda |a_j^*h|\} \leq \frac{1}{m} \sum_{j=1}^{m} \chi_j(|a_j^*h|^2) \leq \frac{1}{m} \sum_{j=1}^{m} |a_j^*h|^2 \cdot 1\{|0.99|a_j^*x|\leq \lambda |a_j^*h|\}.
\]
For any fixed $h$, since $a_1, \ldots, a_m$ are drawn uniformly from the unit sphere $\mathbb{S}^{n-1}_{\mathbb{C}}$, thus the terms $\chi_j(|a_j^*h|^2)$ are independent sub-exponential random variables with the maximal sub-exponential
norm $K = c_1/n$ for some universal constant $c_1 > 0$. According to Bernstein’s inequality, for any fixed $0 < \delta \leq 1$, with probability at least $1 - 2\exp(-c_2\delta^2m)$, it holds

$$
\left| \frac{1}{m} \sum_{j=1}^{m} \chi_j(|a_j^*h|^2) - \mathbb{E} \left[ \chi_j(|a_j^*h|^2) \right] \right| \leq \frac{\delta}{4n},
$$

where $c_2$ is a universal positive constant. Here, $a \in \mathbb{C}^n$ is a vector uniformly distributed on the unit sphere.

To give a uniform bound for the the estimate (26), we construct an $\varepsilon$-net $\mathcal{N}$ over the unit sphere in $\mathbb{C}^n$ with cardinality $|\mathcal{N}| \leq (1 + \frac{2}{\varepsilon})^{2n}$. Then for any $h$ with $\|h\|_2 = 1$, there exists a $h_0 \in \mathcal{N}$ such that $\|h - h_0\|_2 \leq \varepsilon$. Note that $\chi_j(t)$ is a Lipschitz function with Lipschitz constant 100. It then follows from Lemma 6.1 that for $m \geq C'n$ with probability at least $1 - 2\exp(-c_3m)$ it holds

$$
\left| \frac{1}{m} \sum_{j=1}^{m} \chi_j(|a_j^*h|^2) - \frac{1}{m} \sum_{j=1}^{m} \chi_j(|a_j^*h_0|^2) \right| \leq \frac{100}{m} \sum_{j=1}^{m} (|a_j^*h| + |a_j^*h_0|) |a_j^*(h - h_0)|
\leq 100 \sqrt{\frac{1}{m} \sum_{j=1}^{m} |a_j^*h|^2} \sqrt{\frac{1}{m} \sum_{j=1}^{m} |a_j^*(h - h_0)|^2} + 100 \sqrt{\frac{1}{m} \sum_{j=1}^{m} |a_j^*h_0|^2} \sqrt{\frac{1}{m} \sum_{j=1}^{m} |a_j^*(h - h_0)|^2}
\leq \frac{202\varepsilon}{n},
$$

where the third line follows from the Cauchy-Schwarz inequality. Choosing $\varepsilon := \delta/808$ and taking the union bound over $\mathcal{N}$, we obtain that

$$
\frac{1}{m} \sum_{j=1}^{m} \chi_j(|a_j^*h|^2) \geq \mathbb{E} \left( \frac{1}{m} \sum_{j=1}^{m} \chi_j(|a_j^*h|^2) \right) - \frac{\delta}{2n} \quad \text{for all } \|h\|_2 = 1
$$

holds with probability at least

$$
1 - 2\exp(-c_3m) - 2\exp(-c_2\delta^2m)(1 + \frac{2}{\varepsilon})^{2n} \geq 1 - 4\exp(-c_4\delta^2m),
$$

provided $m \geq C\delta^{-2}\log(1/\delta)n$ for some positive constants $C, c_4$.

Finally, we need to lower bound the expectation. To this end, let $\xi \in \mathbb{C}^n$ be a complex Gaussian random vector. We claim that for any $0 < \lambda \leq \sqrt{\frac{5 - \sqrt{21}}{2}}$ it holds

$$
\mathbb{E} \left( |\xi^*h|^2 \cdot 1_{\{|\xi^*e| \leq \lambda |\xi^*h|\}} \right) \leq \frac{2\lambda^2}{\chi^2 + 1}.
$$
Since $\xi/\|\xi\|_2$ is a vector uniformly distributed on the unit sphere and $\|\xi\|_2 \leq (1 - \delta_0) \sqrt{n}$ holds for any fixed $0 < \delta_0 \leq 1$ with probability at least $1 - 2 \exp(-c_5 \delta_0^2 n)$ [35, Theorem 3.1.1]. It then gives that for any $0 < \lambda \leq \sqrt{\frac{3 - 2\eta}{2}}$ with probability at least $1 - 2 \exp(-c_5 \delta_0^2 n)$ we have

$$
\mathbb{E} \left( \left| \mathbf{a}^* \mathbf{h} \right|^2 \cdot \mathbb{I}_{\{ |\mathbf{a}^* \mathbf{x}| \leq \lambda |\mathbf{a}^* \mathbf{h}| \}} \right) = \mathbb{E} \left( \frac{\left| \mathbf{e} \right|^2}{\|\xi\|^2} \cdot \mathbb{I}_{\{ |\mathbf{e} \mathbf{x}| \leq \lambda |\mathbf{e} \mathbf{h}| \}} \right)
\leq \frac{1}{(1 - \delta_0)n} \cdot \mathbb{E} \left( |\mathbf{e} \mathbf{h}|^2 \cdot \mathbb{I}_{\{ |\mathbf{e} \mathbf{x}| \leq \lambda |\mathbf{e} \mathbf{h}| \}} \right)
\leq \frac{1}{(1 - \delta_0)n} \cdot \frac{2\lambda^2}{\lambda^2 + 1}.
$$

(29)

Taking the constant $\delta_0 = \delta/3$, it then follows from (25), (27) and (29) that for any fixed $0 < \lambda \leq 0.4$, with probability at least $1 - 6 \exp(-c\delta^2 n)$, it holds

$$
\frac{1}{m} \sum_{j=1}^{m} \left| \mathbf{a}_j^* \mathbf{h} \right|^2 \cdot \mathbb{I}_{\{ |\mathbf{a}_j^* \mathbf{x}| \leq \lambda |\mathbf{a}_j^* \mathbf{h}| \}} \leq \frac{2\lambda^2}{(\lambda^2 + 0.99)n} + \frac{\delta}{n},
$$

provided $m \geq C\delta^{-2} \log(1/\delta)n$, where $C$ is a universal positive constant. This completes the proof.

It remains to prove the claim (28). Indeed, due to the unitary invariance of Gaussian random vector, without loss of generality, we assume $\mathbf{h} = \mathbf{e}_1$ and $\mathbf{x} = \sigma \mathbf{e}_1 + \tau e^{i\phi} \mathbf{e}_2$, where $\sigma = \mathbf{h}^* \mathbf{x} \in \mathbb{R}$, $|\sigma| \leq 1$ and $\tau = \sqrt{1 - \sigma^2}$. Let $\xi_1, \xi_2$ be the first and second entries of $\xi$. Denote $\xi_1 = \xi_{1,\Re} + i\xi_{1,\Im}$ and $\xi_2 = \xi_{2,\Re} + i\xi_{2,\Im}$ where $\xi_{1,\Re}, \xi_{1,\Im}, \xi_{2,\Re}, \xi_{2,\Im}$ are independent Gaussian random variables with distribution $\mathcal{N}(0, 1/2)$. Then the inequality $|\mathbf{e}^* \mathbf{x}| \leq \lambda |\mathbf{e}^* \mathbf{h}|$ is equivalent to

$$(\sigma \xi_{1,\Re} + \tau (\cos \phi \xi_{2,\Re} + \sin \phi \xi_{2,\Im}))^2 + (\sigma \xi_{1,\Im} - \tau (\sin \phi \xi_{2,\Re} - \cos \phi \xi_{2,\Im}))^2 \leq \lambda (\xi_{1,\Re}^2 + \xi_{1,\Im}^2).$$

To prove the inequality (28), we take the polar coordinates transformations and denote

$$
\begin{align*}
\xi_{1,\Re} &= r_1 \cos \theta_1 \\
\xi_{1,\Im} &= r_1 \sin \theta_1 \\
\xi_{1,\Re} &= r_2 \cos \theta_2 \\
\xi_{1,\Im} &= r_2 \sin \theta_2
\end{align*}
$$

with $r_1, r_2 \in (0, +\infty)$, $\theta_1, \theta_2 \in [0, 2\pi]$. Then the expectation can be written as

$$
G(\lambda, \sigma) := \mathbb{E} \left( |\mathbf{e}^* \mathbf{h}|^2 \cdot \mathbb{I}_{\{ |\mathbf{e}^* \mathbf{x}| \leq \lambda |\mathbf{e}^* \mathbf{h}| \}} \right)
= \frac{1}{\pi^2} \int_0^{2\pi} \int_0^{2\pi} \int_0^{+\infty} \int_0^{\lambda r_1 r_2} \int_0^{r_1^2 + r_2^2} e^{- (r_1^2 + r_2^2)/\tau^2} \cdot e^{2\sigma r_1 r_2 \cos(\theta_1 - \theta_2)/\tau^2} d\rho d\rho_1 d\theta_1 d\theta_2.
$$
We next prove that $f$. In fact, through some basic algebraic computation, we have
\[ \text{Lemma 6.4.} \quad \mu \begin{cases} 
\frac{\partial f}{\partial \lambda} = 1 \int_0^{2\pi} \int_0^{2\pi} \int_0^{+\infty} \lambda \cdot r_1^5 / r_2^2 \cdot e^{-(1+\lambda^2)r_1^2 / r_2^2} \cdot f_{\lambda 2} r_1^2 \cos(\theta_1 - \theta_2) r_1^2 \, dr_1 d\theta_1 d\theta_2 \\
= 1 \int_0^{2\pi} \int_0^{2\pi} \int_0^{+\infty} \lambda r_1^5 \cdot (1 + \lambda^2 - 2\lambda \cos(\theta_1 - \theta_2))^3 \, dr_1 d\theta_1 d\theta_2 \\
= 4\tau^4 \cdot \frac{\lambda(1 + \lambda^4 + 2\lambda^2 + 2\lambda^2 \sigma^2)}{\sqrt{(1 + \lambda^2 + 2\lambda \sigma)^6(1 + \lambda^2 - 2\lambda \sigma)^6}} \\
\leq 2\tau^4 \frac{(\mu_+^2 + \mu_-^2)}{(\mu_+ + \mu_-)^{5/2}}, 
\end{cases} \]
where $\mu_+ := 1 + \lambda^2 + 2\lambda \sigma \geq 0$ and $\mu_- := 1 + \lambda^2 - 2\lambda \sigma \geq 0$. Let
\[ f(\lambda, \sigma) := \frac{\tau^4 (\mu_+^2 + \mu_-^2)}{(\mu_+ + \mu_-)^{5/2}}. \]
We next prove that $f(\lambda, \sigma)$ is a decreasing function with respect to $\sigma$ for any fixed $\lambda \leq \sqrt{\frac{5 - \sqrt{21}}{2}}$. In fact, through some basic algebraic computation, we have
\[ \frac{\partial f(\lambda, \sigma)}{\partial \sigma} = (1 - \sigma^2) \cdot \frac{\lambda(1 - \sigma^2)(\mu_+ - \mu_-) (5\mu_+^2 + 4\mu_+ - 5\mu_-^2) - 4\sigma(\mu_+^3 + \mu_-^3)\mu_+ - \mu_-}{(\mu_+ + \mu_-)^{7/2}} \]
\[ \leq (1 - \sigma^2)\sigma(\mu_+^2 + \mu_-^2) \cdot \frac{28\lambda^2 (1 - \sigma^2) - 4\mu_+ - \mu_-}{(\mu_+ + \mu_-)^{7/2}} \]
\[ = (1 - \sigma^2)\sigma(\mu_+^2 + \mu_-^2) \cdot \frac{28\lambda^2 - 12\lambda^2 \sigma^2 - 4(1 + \lambda^2)^2}{(\mu_+ + \mu_-)^{7/2}} \]
\[ \leq 0, \]
provided $\lambda \leq \sqrt{\frac{5 - \sqrt{21}}{2}}$. Note that $G(0, \sigma) = 0$. It then immediately gives
\[ G(\lambda, \sigma) \leq 2 \int_0^{\lambda} f(t, \sigma) dt \leq 2 \int_0^{\lambda} f(t, 0) dt = 4 \int_0^{\lambda} \frac{t}{(1 + t^2)^3} dt = \frac{\lambda^2(\lambda^2 + 2)}{(\lambda^2 + 1)^2}, \]
which completes the claim (28). 

\[ \square \]

**Lemma 6.4.** Assume $\lambda \geq 2.95$. Let $x, h$ be two fixed vectors in $\mathbb{C}^n$ with $\|x\|_2 = \|h\|_2 = 1$ and $\Re(h^* x) = 0$. Suppose $\xi \in \mathbb{C}^n$ is a complex Gaussian random vector. Then we have
\[ \mathbb{E} \left( \frac{\Re^2 (h^*_\xi \xi^*_x)}{|\xi^*_x|^2} \cdot \mathbb{I}_{\{\lambda|\xi^*_x| \geq |\xi^*_h|\}} \right) \geq \frac{3}{8} - \frac{1}{(\lambda + 1)^2}. \]

**Proof.** Due to the unitary invariance of Gaussian random vector, without loss of generality, we assume $h = e_1$ and $x = \sigma e_1 + \tau e^{i\phi} e_2$, where $\sigma = h^* x \in \mathbb{R}, |\sigma| \leq 1$ and $\tau = \sqrt{1 - \sigma^2}$. Let
ξ₁, ξ₂ be the first and second entries of ξ. Denote ξ₁ = ξ₁,₁ + iξ₁,₂ and ξ₂ = ξ₂,₁ + iξ₂,₂ where ξ₁,₁, ξ₁,₂, ξ₂,₁, ξ₂,₂ are independent Gaussian random variables with distribution \(\mathcal{N}(0, 1/2)\). Then the inequality \(\lambda |\xi^* x| \geq |\xi^* h|\) is equivalent to

\[
\lambda \sqrt{(\sigma \xi_{1,1} + \tau (\cos \phi \xi_{2,1} + \sin \phi \xi_{2,2}))^2 + (\sigma \xi_{1,2} - \tau (\sin \phi \xi_{2,1} - \cos \phi \xi_{2,2}))^2} \geq \sqrt{(\xi_{1,1}^2 + \xi_{1,2}^2)}.
\]

To obtain the conclusion, we take the polar coordinates transformations and denote

\[
\begin{align*}
\xi_{1,1} &= r_1 \cos \theta_1 \\
\xi_{1,2} &= r_1 \sin \theta_1 \\
\sigma \xi_{1,1} + \tau (\cos \phi \xi_{2,1} + \sin \phi \xi_{2,2}) &= r_2 \cos \theta_2 \\
\sigma \xi_{1,2} - \tau (\sin \phi \xi_{2,1} - \cos \phi \xi_{2,2}) &= r_2 \sin \theta_2
\end{align*}
\]

with \(r_1, r_2 \in (0, +\infty), \theta_1, \theta_2 \in [0, 2\pi]\). It is easy to check that

\[
\Re(h^* \xi^* x) = \Re \left( (\sigma \xi_1 + \tau e^{-i\phi} \xi_2) \right) 
= \xi_{1,1} (\sigma \xi_{1,1} + \tau (\cos \phi \xi_{2,1} + \sin \phi \xi_{2,2})) + \xi_{1,2} (\sigma \xi_{1,2} - \tau (\sin \phi \xi_{2,1} - \cos \phi \xi_{2,2})) 
= r_1 r_2 \cos(\theta_1 - \theta_2).
\]

It means the expectation can be written as

\[
F(\lambda, \sigma) := \mathbb{E} \left( \frac{\Re^2(h^* \xi^* x)}{|\xi^* x|^2} \cdot 1_{\{|\xi^* x| \geq |\xi^* h|\}} \right)
= \frac{1}{\pi^2} \int_0^{2\pi} \int_0^{2\pi} \int_0^{+\infty} \int_0^{+\infty} \rho_1^3 r_2 \cdot \cos^2(\theta_1 - \theta_2) \cdot e^{-(r_1^2 + r_2^2)/\tau^2} \cdot e^{2\sigma \tau_1 r_2 \cos(\theta_1 - \theta_2)/\tau^2} dr_1 dr_2 d\theta_1 d\theta_2
= 2 \sum_{k=0}^{+\infty} \frac{2k + 1}{(k!)^2 \cdot (k + 1)} \cdot \frac{\sigma^{2(k + 1)}}{\tau^{4k + 2}} \int_0^{r_1^2 + r_2^2} \cdot e^{-(r_1^2 + r_2^2)/\tau^2} dr_1 dr_2;
\]

where the last equation follows from the fact

\[
\int_0^{2\pi} \int_0^{2\pi} \cos^2(\theta_1 - \theta_2) d\theta_1 d\theta_2 = \frac{(2k - 1)!!}{2^{k - 2} \cdot k!}
\]

for any integer \(k\). To evaluate \(F(\lambda, \sigma)\), we first take the derivative and then obtain

\[
\frac{\partial F(\lambda, \sigma)}{\partial \lambda} := 2 \sum_{k=0}^{+\infty} \frac{2k + 1}{(k!)^2 \cdot (k + 1)} \cdot \frac{\sigma^{2(k + 1)}}{\tau^{4k + 2}} \int_0^{+\infty} \lambda^{2k + 3} r_1^{4k + 5} \cdot e^{-(1 + \lambda^2) r_2^2/\tau^2} dr_2
= 2 \sum_{k=0}^{+\infty} \frac{(2k + 1)!!(2k + 1)}{(k!)^2} \cdot \frac{\sigma^{2k(1 - \sigma^2)^2}}{\tau^{4k + 2}} \cdot \left( \frac{\lambda}{1 + \lambda^2} \right)^{2k + 3}.
\]

Since \(F(0, \sigma) = 0\), it implies that

\[
F(\lambda, \sigma) = 2 \sum_{k=0}^{+\infty} \frac{(2k + 1)!!(2k + 1)}{(k!)^2} \cdot \frac{\sigma^{2k(1 - \sigma^2)^2}}{\tau^{4k + 2}} \cdot \int_0^{+\infty} \left( \frac{t}{1 + t^2} \right)^{2k + 3} dt.
\]
With this in place, all we need to do is to lower bound the integral \( \int_0^\lambda \left( \frac{t}{1+t^2} \right)^{2k+3} dt \). Note that
\[
\int_0^\lambda \left( \frac{t}{1+t^2} \right)^{2k+3} dt = \int_0^1 \left( \frac{t}{1+t^2} \right)^{2k+3} dt + \int_1^\lambda \left( \frac{t}{1+t^2} \right)^{2k+3} dt := I + II.
\]
For the first term, let \( t = \tan \theta \). It then gives
\[
I = \int_0^{\pi/4} \sin^{2k+3} \theta \cos^{2k+1} \theta \, d\theta
= \frac{1}{(2k+1)!!} \cdot \frac{1}{2(2k+3)} - \frac{1}{2(k+1) \cdot 2^{2k+3}}.
\]
For the second term, noting that \( \lambda \geq 1 \), we have
\[
II \geq \int_1^\lambda (1+t)^{2k-3} dt = \frac{1}{2(k+1)} \cdot \left( \frac{1}{2^{2k+2}} - \frac{1}{(\lambda+1)^{2k+2}} \right).
\]
Putting (31) and (32) into (30), we have
\[
F(\lambda, \sigma) \geq \sum_{k=0}^{\infty} \frac{(2k+1)!!(2k+7)}{(k+1)!} \cdot \sigma^{2k}(1-\sigma^2)^2 \cdot \left( \frac{(k+1)!}{4(2k+1)!!} + \frac{1}{2^{k+3}} - \frac{2^k}{(1+\lambda)^{2k+2}} \right).
\]
Let \( \beta := (1+\lambda)^2 \). Expand \( F(\lambda, \sigma) \) into a series with respect to \( \sigma \) and we have
\[
(33) \quad F(\lambda, \sigma) \geq \frac{3}{8} - \frac{1}{\beta} + \frac{9}{16}\sigma^2 - \sum_{k=1}^{\infty} \frac{(2k-1)!!(2k+7)}{2^{k+4}(k+2)!} \sigma^{2k+2} - \frac{4}{\beta^2} \sum_{k=1}^{\infty} \frac{2^k(2k-1)!!}{(k-1)!\beta^k} \sigma^{2k+2},
\]
where we use the fact that \( \lambda \geq 2.95 \) in the above inequality. Next, we need to upper bound the last two series. From Wallis’ inequality [23], we know
\[
\frac{(2k-1)!!}{2^k k!} \leq \frac{1}{\sqrt{2k}}
\]
Thus
\[
\sum_{k=1}^{\infty} \frac{(2k-1)!!(2k+7)}{2^{k+4}(k+2)!} \sigma^{2k+2} \leq \frac{\sigma^2}{16} \sum_{k=1}^{\infty} \frac{2k+7}{\sqrt{2k}(k+2)(k+1)} \leq \frac{\sigma^2}{8} \sum_{k=1}^{\infty} \frac{1}{k^{3/2}} \leq \frac{\sigma^2}{8} \left( 1 + \int_1^{\infty} \frac{1}{t^{3/2}} \, dt \right) = \frac{3}{8}\sigma^2,
\]
where the second inequality follows from the fact that \( \frac{2k+7}{(k+2)(k+1)} \leq \frac{2\sqrt{k}}{k} \) for all \( k \geq 1 \).
On the other hand, using Wallis’ inequality again, we have

\[
(1 - \frac{4}{\beta})^2 \cdot \sum_{k=1}^{\infty} \frac{2^k (2k - 1)!!}{(k - 1)! \beta^k} \sigma^{2k+2} \leq (1 - \frac{4}{\beta})^2 \cdot \sigma^2 \sum_{k=1}^{\infty} \frac{4^k \cdot k}{\sqrt{2} k \beta^k} \\
\leq (1 - \frac{4}{\beta})^2 \cdot \frac{\sigma^2}{\sqrt{2}} \sum_{k=1}^{\infty} k \left(\frac{4}{\beta}\right)^k \\
= \frac{2\sqrt{2}}{\beta} \sigma^2,
\]

(35)

where the last equation follows from the fact that

\[
\sum_{k=1}^{\infty} k x^{k-1} = \frac{1}{(1 - x)^2} \quad \text{for all} \quad 0 \leq x < 1.
\]

Putting (34) and (35) into (33), we know that for \( \lambda \geq 2.95 \), it holds

\[
F(\lambda, \sigma) \geq \frac{3}{8} - \frac{1}{\beta}.
\]

This completes the proof. □

**Bibliography**

[1] B. Balan, “On signal reconstruction from its spectrogram,” *44th Annual Conference on Information Sciences and Systems (CISS)*, pp. 1–4, IEEE, 2010.

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

[3] A. Ben-Tal and A. Nemirovski, “Lectures on modern convex optimization: analysis, algorithms, and engineering applications,” SIAM, 2001.

[4] T. Bendory, Y. C. Eldar, and N. Boumal, “Non-convex phase retrieval from STFT measurements,” *IEEE Trans. Inf. Theory*, vol. 64, no. 1, pp. 467–484, 2017.

[5] J. Cai, M. Huang, D. Li, and Y. Wang, Solving phase retrieval with random initial guess is nearly as good as by spectral initialization, 2021 [Online]. Available: http://arxiv.org/abs/2101.03540.

[6] E. J. Candès and X. Li, “Solving quadratic equations via PhaseLift when there are about as many equations as unknowns,” *Found. Comut. Math.*, vol. 14, no. 5, pp. 1017–1026, 2014.

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

[8] E. J. Candès, 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.

[9] Y. Chen and E. J. Candès, “Solving random quadratic systems of equations is nearly as easy as solving linear systems,” *Commun. Pure Appl. Math.*, vol. 70, no. 5, pp. 822–883, 2017.

[10] A. Chai, M. Moscoso, and G. Papanicolaou, “Array imaging using intensity-only measurements,” *Inverse Probl.*, vol. 27, no. 1, 015005, 2010.

[11] 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.

[12] J. C. Dainty and J.R. Fienup, “Phase retrieval and image reconstruction for astronomy,” *Image Recovery: Theory and Application*, vol. 231, pp. 275, 1987.
[13] F. Deutsch, “Rate of convergence of the method of alternating projections,” *Parametric optimization and approximation*, pp. 96–107, 1984.
[14] F. Deutsch and H. Hundal, “The rate of convergence for the method of alternating projections: II,” *J. Math. Anal. Appl.* vol. 205, no. 2, pp. 381–405, 1997.
[15] A. Galántai, “On the rate of convergence of the alternating projection method in finite dimensional spaces,” *J. Math. Anal. Appl.*, vol. 310, no. 1, pp. 30–44, 2005.
[16] B. Gao, X. Sun, Y. Wang, and Z. Xu, “Perturbed Amplitude Flow for Phase Retrieval,” *IEEE Trans. Signal Process.*, vol. 68, pp. 5427–5440, 2020.
[17] B. Gao and Z. Xu, “Phaseless recovery using the Gauss–Newton method,” *IEEE Trans. Signal Process.*, vol. 65, no. 22, pp. 5885–5896, 2017.
[18] R. W. Harrison, “Phase problem in crystallography,” *JOSA A*, vol. 10, no. 5, pp. 1046–1055, 1993.
[19] M. Huang and Z. Xu, “Solving systems of quadratic equations via exponential-type gradient descent algorithm,” *Journal of Computational Mathematics*, vol. 38, no. 4, 2020.
[20] K. Jaganathan, Y. C. Eldar, and B. Hassibi, “Phase retrieval: An overview of recent developments,” *Optical Compressive Imaging*, pp. 279–312, 2016.
[21] H. Jeong and C. S. Güntürk, Convergence of the randomized Kaczmarz method for phase retrieval, 2017 [Online]. Available: http://arxiv.org/abs/1706.10291.
[22] S. Kaczmarz, “Angenaherte auflosung von systemen linearer gleichungen,” *Bull. Int. Acad. Pol. Sic. Let., Cl. Sci. Math. Nat.*, pp. 355–357, 1937.
[23] N. D. Kazarinoff, *Analytic inequalities*. Courier Corporation, 2014.
[24] G. Li, Y. Gu, and Y. M. Lu, “Phase retrieval using iterative projections: Dynamics in the large systems limit,” *In The 53rd Annual Allerton Conference on Communication, Control, and Computing*, 2015.
[25] W. Luo, W. Alghamdi, and Y. M. Lu, “Optimal spectral initialization for signal recovery with applications to phase retrieval,” *IEEE Trans. Signal Process.*, vol. 67, no. 9, pp. 2347–2356, 2019.
[26] 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, 2008.
[27] M. Mondelli and A. Montanari, “Fundamental limits of weak recovery with applications to phase retrieval,” *Foundations Comput. Math.*, vol. 19, pp. 703–773, 2019.
[28] R. P. Millane, “Phase retrieval in crystallography and optics,” *J. Optical Soc. America A*, vol. 7, no. 3, pp. 394-411, 1990.
[29] P. Netrapalli, P. Jain, and S. Sanghavi, “Phase retrieval using alternating minimization,” *IEEE Trans. Signal Process.*, vol. 63, no. 18, pp. 4814–4826, 2015.
[30] H. Sahinoglou and S. D. Cabrera, “On phase retrieval of finite-length sequences using the initial time sample,” *IEEE Trans. Circuits and Syst.*, vol. 38, no. 8, pp. 954–958, 1991.
[31] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, “Phase retrieval with application to optical imaging: a contemporary overview,” *IEEE Signal Process. Mag.*, vol. 32, no. 3, pp. 87–109, 2015.
[32] T. Strohmer and R. Vershynin, “A randomized Kaczmarz algorithm with exponential convergence,” *J. Fourier Anal. Appl.*, vol. 15, no. 2, pp. 262–278, 2009.
[33] J. Sun, Q. Qu, and J. Wright, “A geometric analysis of phase retrieval,” *Found. Comput. Math.*, vol. 18, no. 5, pp. 1131–1198, 2018.
[34] Y. S. Tan and R. Vershynin, “Phase retrieval via randomized kaczmarz: Theoretical guarantees,” *Information and Inference: A Journal of the IMA*, vol. 8, no. 1, pp. 97–123, 2019.
[35] R. Vershynin, *High-dimensional probability: An introduction with applications in data science*. U.K.:Cambridge Univ. Press, 2018.
[36] I. Waldspurger, “Phase retrieval with random gaussian sensing vectors by alternating projections,” *IEEE Trans. Inf. Theory*, vol. 64, no. 5, pp. 3301–3312, 2018.
[37] I. Waldspurger, A. d’Aspremont, and S. Mallat, “Phase recovery, maxcut and complex semidefinite programming,” *Math. Prog.*, vol. 149, no. 1-2, pp. 47–81, 2015.
[38] A. Walther, “The question of phase retrieval in optics,” *J. Mod. Opt.*, vol. 10, no. 1, pp. 41–49, 1963.
[39] G. Wang, G. B. Gianakis, 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, 2018.
[40] 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.

[41] K. Wei, “Solving systems of phaseless equations via kaczmarz methods: a proof of concept study,” *Inverse Probl.*, vol. 31, no. 12, 125008, 2015.

[42] H. Zhang, Y. Zhou, Y. Liang, and Y. Chi, “A nonconvex approach for phase retrieval: Reshaped wirtinger flow and incremental algorithms,” *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 5164–5198, 2017.

**Department of Mathematics, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China**

*Email address: menghuang@ust.hk*

**Department of Mathematics, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China**

*Email address: yangwang@ust.hk*