SIAM J. IMAGING SCIENCES
Vol. 15, No. 2, pp. 989–1016

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 computation time. While the rate of convergence has been well studied 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 [*Inf. Inference*, 8 (2019), pp. 97–123]. In this paper, we provide the first theoretical guarantees for it. We show that for random measurements $\boldsymbol{a}_j \in \mathbb{C}^n$, $j = 1, \ldots, m$, which are drawn independently and uniformly from the complex unit sphere, or equivalently 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.

Key words. Kaczmarz method, phase retrieval, linear convergence, stochastic gradient descent

MSC codes. 68W15, 94A12, 65Y20

DOI. 10.1137/21M1450537

1. Introduction.

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

$$b_j = |\langle \boldsymbol{a}_j, \boldsymbol{x} \rangle|, \quad j = 1, \ldots, m, \tag{1.1}$$

where $\boldsymbol{a}_j \in \mathbb{C}^n$ (or $\mathbb{R}^n$) are known sampling vectors and $b_j \in \mathbb{R}$ are observed measurements. This problem, called *phase retrieval*, has been a topic of study since the 1980's due to its wide range of practical applications in fields of physical sciences and engineering, such as X-ray crystallography [18, 27], diffraction imaging [31, 9], microscopy [26], astronomy [12], optics and acoustics [38, 1, 2], 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

*Received by the editors October 4, 2021; accepted for publication (in revised form) March 9, 2022; published electronically June 30, 2022.
https://doi.org/10.1137/21M1450537
Funding: The work of the second author was supported in part by the Hong Kong Research Grants Council grants 16306415 and 16308518.
†School of Mathematical Sciences, Beihang University, Beijing, 100191, China (menghuang@buaa.edu.cn).
‡Department of Mathematics, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China (yangwang@ust.hk).

generally NP-complete [30]. Another special case of solving these phaseless equations is the well-known *stone problem* in combinatorial optimization, which is also NP-complete [3].

To solve (1.1), we employ the randomized Kaczmarz method where the update rule is given by

$$\boldsymbol{z}_{k+1} = \boldsymbol{z}_k - \left(1 - \frac{b_{i_k}}{|\boldsymbol{a}_{i_k}^* \boldsymbol{z}_k|}\right) \frac{\boldsymbol{a}_{i_k}^* \boldsymbol{z}_k}{\|\boldsymbol{a}_{i_k}\|_2^2} \boldsymbol{a}_{i_k}, \tag{1.2}$$

where i_k is chosen randomly from the $\{1, \ldots, m\}$ with probability proportional to $\|\boldsymbol{a}_{i_k}\|_2^2$ at the $(k+1)$th iteration. Actually, the update rule above is a natural adaption 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 $\boldsymbol{x}$, the phase information can be approximated by that of the current estimate. Thus, in each iteration, we first select a measurement vector $\boldsymbol{a}_{i_k}$ randomly and then project the current estimate $\boldsymbol{z}_k$ onto the hyperplane

$$\left\{ \boldsymbol{z} \in \mathbb{C}^n : \langle \boldsymbol{a}_{i_k}, \boldsymbol{z} \rangle = b_{i_k} \cdot \frac{\boldsymbol{a}_{i_k}^* \boldsymbol{z}_k}{|\boldsymbol{a}_{i_k}^* \boldsymbol{z}_k|} \right\}.$$

That gives the scheme (1.2).

We are interested in the following questions:

Does the randomized Kaczmarz scheme (1.2) converge to the target solution $\boldsymbol{x}$ (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 the 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 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, Gu, and Lu [24] established an asymptotic convergence of the randomized Kaczmarz method for phase retrieval, but it requires an infinite number of samples, which is unrealistic in practicality. Recently, Tan and Vershynin [34] used the chain argument coupled with bounds on Vapnik–Chervonenkis (VC) dimension and metric entropy and then proved 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 the same as that of [34] was also obtained independently by Jeong and Güntürk [21] using the tools of hyperplane tessellation and "drift analysis." Another similar conditional error contractivity result was 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, the measurement vectors can be divided into "good measurements" with a correct phase and "bad measurements" with an incorrect phase. When the initial point is close enough to the true solution, the total influence of "bad measurements" can be well controlled. However, this is not true for the complex measurements because $\boldsymbol{x}e^{i\theta}$ is continuous with respect to $\theta \in [0, 2\pi)$. For this reason, the proofs for the real case cannot 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 *conjecture*. We shall point out that the convergence of the randomized Kaczmarz method for phase retrieval in a complex setting is of more practical interest.

In this paper, we aim to prove this conjecture by introducing a deterministic condition on measurement 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 $\boldsymbol{x}$ from phaseless equations (1.1), has received intensive investigations recently. Note that if $\boldsymbol{z}$ is a solution to (1.1), then $\boldsymbol{z}e^{i\theta}$ is also the solution of this problem for any $\theta \in \mathbb{R}$. Therefore, the recovery of the solution $\boldsymbol{x}$ is up to a global phase. It has been shown theoretically that $m \geq 4n - 4$ generic measurements suffice to recover $\boldsymbol{x}$ 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 [8, 6], PhaseCut [37], etc. While these 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 nonconvex loss function in the natural parameter space, which achieves significantly improved computational performance. The first nonconvex algorithm with theoretical guarantees was given by Netrapalli, Jain, and Sanghavi, 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 nonconvex algorithms based on spectral initialization [4, 7, 10, 36, 19]. Specifically, Candès, Li, and Soltanolkotabi 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. Recently, Chen and Candès improved the result to $O(n)$ Gaussian random measurements by incorporating a truncation, namely the truncated Wirtinger flow (TWF) [10] algorithm. Other nonconvex methods with provable guarantees include the Gauss–Newton [17], the trust-region [33], smoothed amplitude flow [5], the truncated amplitude flow (TAF) [39] algorithm, the reshaped Wirtinger flow (RWF) [42] algorithm, and the 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. The Kaczmarz method is one of the most popular algorithms for solving overdetermined systems of linear equations [22], which iteratively project the current estimate onto the hyperplane of one chosen equation at a time. Suppose the system of linear equations we want to solve is given by $A\boldsymbol{x} = \boldsymbol{y}$, where $A \in \mathbb{C}^{m\times n}$. In each iteration of the Kaczmarz method, one row $\boldsymbol{a}_{i_k}$ of A is selected and then the new iterate $\boldsymbol{z}_{k+1}$ is obtained by projecting the current estimate $\boldsymbol{z}_k$ orthogonally onto the solution hyperplane of $\langle \boldsymbol{a}_{i_k}, \boldsymbol{z}\rangle = y_{i_k}$ as follows:

$$\boldsymbol{z}_{k+1} = \boldsymbol{z}_k + \frac{y_{i_k} - \langle \boldsymbol{a}_{i_k}, \boldsymbol{z}_k\rangle}{\|\boldsymbol{a}_{i_k}\|_2^2}\boldsymbol{a}_{i_k}. \tag{1.3}$$

The classical version of the 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, 14, 15]. In 2009, Strohmer and Vershynin [32] proposed a randomized Kaczmarz method where the row of A is selected in random order and they proved that this randomized Kaczmarz method is convergent with the 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 $\|\boldsymbol{a}_{i_k}\|_2^2$, then for any initial $\boldsymbol{z}_0$ the iteration $\boldsymbol{z}_k$ given by randomized Kaczmarz scheme (1.3) obeys

$$\mathbb{E}\|\boldsymbol{z}_k - \boldsymbol{x}\|_2 \le \left(1 - \frac{1}{k(A)\cdot n}\right)^{k/2} \cdot \|\boldsymbol{z}_0 - \boldsymbol{x}\|_2,$$

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 was 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 lacks adequate theoretical performance guarantees. Recently, in the real case, when the measurements $\boldsymbol{a}_j$ 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 $0 < \delta$, $\delta_0 \le 1$, if $m \gtrsim n\log(m/n) + \log(1/\delta_0)$ and $\boldsymbol{a}_j \in \mathbb{R}^n$ are drawn independently and uniformly from the unit sphere, then with probability at least $1 - \delta_0$ it holds that the kth step randomized Kaczmarz estimate $\boldsymbol{z}_k$ given by (1.2) satisfies

$$\mathbb{E}_{\mathcal{I}^k}\left[\mathrm{dist}(\boldsymbol{z}_k, \boldsymbol{x})\mathbb{1}_{\tau=\infty}\right] \le \left(1 - \frac{1}{2n}\right)^{k/2} \mathrm{dist}(\boldsymbol{z}_0, \boldsymbol{x}),$$

provided $\mathrm{dist}(\boldsymbol{z}_0, \boldsymbol{x}) \le c\delta\|\boldsymbol{x}\|_2$ for some constant $c > 0$. Furthermore, the probability $\mathbb{P}(\tau < \infty) \le \delta^2$. Here τ is the stopping time and $\mathbb{E}_{\mathcal{I}^k}$ denotes the expectation with respect to randomness $\mathcal{I}^k := \{i_0, i_1, \ldots, i_{k-1}\}$ conditioned on the high probability event of random measurements $\{\boldsymbol{a}_j\}_{j=1}^m$.

1.4. Our contributions.

As stated before, the randomized Kaczmarz method is a popular and convenient method for solving the 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 the 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* $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m \in \mathbb{C}^n$ *are drawn independently and uniformly from the unit sphere* $\mathbb{S}_{\mathbb{C}}^{n-1}$. *For any* $0 < \delta < 1$, *let* $\boldsymbol{z}_0$ *be an initial estimate to* $\boldsymbol{x}$ *such that* $\operatorname{dist}(\boldsymbol{z}_0, \boldsymbol{x}) \le 0.01\delta \|\boldsymbol{x}\|_2$. *There exist universal constants* $C_0, c_0 > 0$ *such that if* $m \ge C_0 n$, *then with probability at least* $1 - 14\exp(-c_0 n)$ *it holds that the iteration* $\boldsymbol{z}_k$ *given by randomized Kaczmarz update rule* (1.2) *obeys*

$$\mathbb{E}_{\mathcal{I}^k}\left[\operatorname{dist}(\boldsymbol{z}_k, \boldsymbol{x}) \mathbb{1}_{\tau=\infty}\right] \le (1 - 0.03/n)^{k/2} \operatorname{dist}(\boldsymbol{z}_0, \boldsymbol{x}),$$

where τ *is the stopping time defined by*

$$\tau := \min\left\{k : \boldsymbol{z}_k \notin B\right\} \quad \textit{with} \quad B := \left\{\boldsymbol{z} : \operatorname{dist}(\boldsymbol{z}, \boldsymbol{x}) \le 0.01\|\boldsymbol{x}\|_2\right\}.$$

Furthermore, the probability $\mathbb{P}(\tau < \infty) \le \delta^2$. *Here* $\mathbb{E}_{\mathcal{I}^k}$ *denotes the expectation with respect to randomness* $\mathcal{I}^k := \{i_0, i_1, \ldots, i_{k-1}\}$ *conditioned on the high probability event of random measurements* $\{\boldsymbol{a}_j\}_{j=1}^m$.

The theorem asserts that the randomized Kaczmarz method converges linearly to the global solution $\boldsymbol{x}$ (up to a global phase) in expectation for random measurements $\boldsymbol{a}_j \in \mathbb{C}^n$ which are drawn independently and uniformly from the complex unit sphere, or equivalently are independent complex Gaussian random vectors, with an optimal sample complexity. The proof of this theorem is much more direct than the Tan–Vershynin analysis of the randomized Kaczmarz algorithm for real Gaussian measurements [34]. Specifically, our analysis is based on the restricted strong convexity property of the loss function (2.5), while the Tan–Vershynin analysis is based on a chain argument coupled with bounds on Vapnik–Chervonenkis (VC) dimension and metric entropy.

Remark 1.2. In Theorem 1.1, there are two sources of randomness: one is from the measurements $\boldsymbol{a}_j$, and the other is from the selection of the equation at each iteration of the algorithm. For this reason, an important distinction between the randomized Kaczmarz method and other algorithms such as WF [7], TWF [10], and TAF [39] is that it is conditional on the event $\boldsymbol{z}_k \in B$ for all $k \ge 1$. From Theorem 1.1, this event holds with probability at least $1 - \delta^2$.

Remark 1.3. Theorem 1.1 requires an initial estimate $\boldsymbol{z}_0$ 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 nonconvex algorithms for phase retrieval. For instance, when $\boldsymbol{a}_j \in \mathbb{C}^n$ 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 $\boldsymbol{z}_0$ satisfying $\operatorname{dist}(\boldsymbol{z}_0, \boldsymbol{x}) \le \epsilon \|\boldsymbol{x}\|_2$ for any fixed ϵ, provided $m \ge Cn$ for a positive constant C. We refer the reader to [10, 39, 42] for the spectral initialization of others and [25, 28] for the optimal design of a spectral initialization.

1.5. Notations. Throughout this paper, we assume the measurements $\boldsymbol{a}_j \in \mathbb{C}^n$, $j = 1, \ldots, m$, 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 $\boldsymbol{z} \in \mathbb{S}_{\mathbb{C}}^{n-1}$ if $\boldsymbol{z} \in \mathbb{C}^n$ and $\|\boldsymbol{z}\|_2 = 1$. Let $\Re(z) \in \mathbb{R}$ and $\Im(z) \in \mathbb{R}$ denote the real and imaginary parts of a complex number $z \in \mathbb{C}$. For any $A, B \in \mathbb{R}$, we use $A \lesssim B$ to denote $A \le C_0 B$, 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 the 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 that the main result can be proved by using the tools from stochastic processes. In section 4, we demonstrate that the random measurements drawn independently and uniformly from the complex unit sphere satisfy the restricted strong convexity condition with high probability. In section 5, we carry out some numerical experiments to demonstrate the efficiency and robustness of the randomized Kaczmarz method. A brief discussion is presented in section 6. Section 7 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 $\boldsymbol{x} \in \mathbb{C}^n$ be the target signal we want to recover. The measurements we obtain are

$$b_j = |\langle \boldsymbol{a}_j, \boldsymbol{x} \rangle|, \quad j = 1, \ldots, m, \tag{2.1}$$

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

$$\boldsymbol{z}_{k+1} = \boldsymbol{z}_k - \left(1 - \frac{b_{i_k}}{|\boldsymbol{a}_{i_k}^* \boldsymbol{z}_k|}\right) \boldsymbol{a}_{i_k} \boldsymbol{a}_{i_k}^* \boldsymbol{z}_k, \tag{2.2}$$

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

Obviously, for any $\boldsymbol{z}$, if $\boldsymbol{z}$ is a solution to (2.1), then $\boldsymbol{z}e^{i\phi}$ is also a solution to it for any $\phi \in \mathbb{R}$. Thus, for $m \ge O(n)$ generic measurements $\boldsymbol{a}_j$, the set of solutions to (2.1) is $\left\{\boldsymbol{x}e^{i\phi} : \phi \in \mathbb{R}\right\}$, which is a one-dimensional circle in $\mathbb{C}^n$ [11, 40] . For this reason, we define the distance between $\boldsymbol{z}$ and $\boldsymbol{x}$ as

$$\operatorname{dist}(\boldsymbol{z}, \boldsymbol{x}) = \min_{\phi \in \mathbb{R}} \|\boldsymbol{z} - \boldsymbol{x}e^{i\phi}\|_2.$$

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

$$\phi(\boldsymbol{z}) := \operatorname*{argmin}_{\phi \in \mathbb{R}} \|\boldsymbol{z} - \boldsymbol{x}e^{i\phi}\|_2 \tag{2.3}$$

for any $\boldsymbol{z} \in \mathbb{C}^n$. Moreover, for any $\epsilon \geq 0$ we define the ϵ-neighborhood of $\boldsymbol{x}$ as

$$E(\epsilon) := \{\boldsymbol{z} \in \mathbb{C}^n : \mathrm{dist}(\boldsymbol{z}, \boldsymbol{x}) \leq \epsilon \|\boldsymbol{x}\|_2\}\,. \tag{2.4}$$

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

$$f(\boldsymbol{z}) = \frac{1}{m} \sum_{j=1}^{m} \left(|\boldsymbol{a}_j^* \boldsymbol{z}| - |\boldsymbol{a}_j^* \boldsymbol{x}|\right)^2. \tag{2.5}$$

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

$$D_{\boldsymbol{v}} f(\boldsymbol{z}) := \lim_{t \to 0^+} \frac{f(\boldsymbol{z} + t\boldsymbol{v}) - f(\boldsymbol{z})}{t}$$

if the limit exists. It is not difficult to compute that the one-sided directional derivative of f in (2.5) along any direction $\boldsymbol{v}$ is

$$D_{\boldsymbol{v}} f(\boldsymbol{z}) = \frac{2}{m} \sum_{j=1}^{m} \left(1 - \frac{|\boldsymbol{a}_j^* \boldsymbol{x}|}{|\boldsymbol{a}_j^* \boldsymbol{z}|}\right) \Re(\boldsymbol{a}_j^* \boldsymbol{v} \boldsymbol{z}^* \boldsymbol{a}_j). \tag{2.6}$$

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 (2.4) and (2.5), respectively.

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

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

for all $\boldsymbol{z} \in E(\epsilon)$.

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 Kaczmarz scheme in a single step. Next, we show that for random measurements $\boldsymbol{a}_j \in \mathbb{C}^n, j = 1, \ldots, m$, which are drawn independently and uniformly from the complex unit sphere the function f defined in (2.5) satisfies the restricted strong convexity with high probability, provided $m \geq Cn$ for some constant $C > 0$. Finally, using the tools from stochastic processes, we could prove that the randomized Kaczmarz method is linearly convergent in expectation, which concludes the proof of the main result.

Theorem 3.1. *Assume f defined in (2.5) satisfies the restricted strong convexity $RSC(\gamma, \epsilon)$. Then the iteration $\boldsymbol{z}_{k+1}$ given by randomized Kaczmarz update rule (2.2) obeys*

$$\mathbb{E}_{i_k}\left[\mathrm{dist}^2(\boldsymbol{z}_{k+1}, \boldsymbol{x})\right] \leq (1 - \gamma)\mathrm{dist}^2(\boldsymbol{z}_k, \boldsymbol{x})$$

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

Proof. Recognize that $\|\boldsymbol{a}_{i_k}\|_2 = 1$. Using restricted strong convexity condition RSC(γ, ϵ), we have

$$
\begin{aligned}
\mathbb{E}_{i_k} \mathrm{dist}^2(\boldsymbol{z}_{k+1}, \boldsymbol{x}) &= \mathbb{E}_{i_k} \|\boldsymbol{z}_{k+1} - \boldsymbol{x} e^{i\phi(\boldsymbol{z}_{k+1})}\|_2^2 \\
&\le \mathbb{E}_{i_k} \left\| \boldsymbol{z}_k - \boldsymbol{x} e^{i\phi(\boldsymbol{z}_k)} - \left(1 - \frac{b_{i_k}}{|\boldsymbol{a}_{i_k}^* \boldsymbol{z}_k|}\right) \boldsymbol{a}_{i_k}^* \boldsymbol{z}_k \boldsymbol{a}_{i_k} \right\|_2^2 \\
&= \|\boldsymbol{z}_k - \boldsymbol{x} e^{i\phi(\boldsymbol{z}_k)}\|_2^2 + \mathbb{E}_{i_k} \left(1 - \frac{b_{i_k}}{|\boldsymbol{a}_{i_k}^* \boldsymbol{z}_k|}\right)^2 |\boldsymbol{a}_{i_k}^* \boldsymbol{z}_k|^2 \\
&\quad -2\mathbb{E}_{i_k} \Re \left(\left(1 - \frac{b_{i_k}}{|\boldsymbol{a}_{i_k}^* \boldsymbol{z}_k|}\right) \boldsymbol{z}_k^* \boldsymbol{a}_{i_k} \boldsymbol{a}_{i_k}^* (\boldsymbol{z}_k - \boldsymbol{x} e^{i\phi(\boldsymbol{z}_k)}) \right) \\
&= \|\boldsymbol{z}_k - \boldsymbol{x} e^{i\phi(\boldsymbol{z}_k)}\|_2^2 + f(\boldsymbol{z}_k) - D_{\boldsymbol{z}_k - \boldsymbol{x} e^{i\phi(\boldsymbol{z}_k)}} f(\boldsymbol{z}_k) \\
&\le (1-\gamma) \|\boldsymbol{z}_k - \boldsymbol{x} e^{i\phi(\boldsymbol{z}_k)}\|_2^2,
\end{aligned}
$$

where the third equation follows from the expression of the directional derivative as shown in (2.6). This completes the proof. ∎

Theorem 3.2. *Assume the measurement vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m \in \mathbb{C}^n$ are drawn uniformly from the unit sphere $\mathbb{S}_{\mathbb{C}}^{n-1}$. Suppose that $m \ge C_0 n$ and f is defined in (2.5). Then f satisfies the restricted strong convexity RSC$(\frac{0.03}{n}, 0.01)$ 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. ∎

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

$$\mathbb{E}_{i_k}\left[\mathrm{dist}^2(\boldsymbol{z}_{k+1}, \boldsymbol{x})\right] \le (1 - 0.03/n)\, \mathrm{dist}^2(\boldsymbol{z}_k, \boldsymbol{x}),$$

provided $\mathrm{dist}(\boldsymbol{z}_k, \boldsymbol{x}) \le 0.01\|\boldsymbol{x}\|_2$ at k step. To be able to iterate this result recursively, we need that the condition $\mathrm{dist}(\boldsymbol{z}_k, \boldsymbol{x}) \le 0.01\|\boldsymbol{x}\|_2$ holds for all k; however, it does not hold arbitrarily. Hence, we introduce a stopping time

$$\tau := \min\{k : \boldsymbol{z}_k \notin B\}, \tag{3.1}$$

where $B := \{\boldsymbol{z} : \mathrm{dist}(\boldsymbol{z}, \boldsymbol{x}) \le 0.01\|\boldsymbol{x}\|_2\}$. 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 $m \ge C_0 n$ for some universal constant $C_0 > 0$. Assume the measurement vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m \in \mathbb{C}^n$ are drawn independently and uniformly from the unit sphere $\mathbb{S}_{\mathbb{C}}^{n-1}$. For any $0 < \delta < 1$, let $\boldsymbol{z}_0$ be an initial estimate to $\boldsymbol{x}$ such that $\|\boldsymbol{z}_0 - \boldsymbol{x}\|_2 \le 0.01\delta\|\boldsymbol{x}\|_2$. Let τ be the stopping time defined in (3.1). Then with probability at least $1 - 14\exp(-c_0 n)$ it holds that the iteration $\boldsymbol{z}_k$ given by randomized Kaczmarz update rule (2.2) obeys*

$$\mathbb{E}_{\mathcal{I}^k}\left[\mathrm{dist}(\boldsymbol{z}_k, \boldsymbol{x}) \mathbb{1}_{\tau=\infty}\right] \le (1 - 0.03/n)^{k/2} \mathrm{dist}(\boldsymbol{z}_0, \boldsymbol{x}).$$

Furthermore, the probability $\mathbb{P}(\tau < \infty) \le \delta^2$. Here $\mathbb{E}_{\mathcal{I}^k}$ denotes the expectation with respect to randomness $\mathcal{I}^k := \{i_0, i_1, \ldots, i_{k-1}\}$ conditioned on the high probability event of random measurements $\{\boldsymbol{a}_j\}_{j=1}^m$ and $c_0 > 0$ is a universal constant.

Proof. From Theorems 3.1 and 3.2, we obtain that if $m \geq C_0 n$, then with probability at least $1 - 14\exp(-c_0 n)$ it holds that

$$\begin{aligned}\mathbb{E}_{i_k}\left[\operatorname{dist}^2(\boldsymbol{z}_{k+1},\boldsymbol{x})\mathbb{1}_{\tau>k+1} \;\middle|\; \boldsymbol{z}_k \in B\right] &\leq \mathbb{E}_{i_k}\left[\operatorname{dist}^2(\boldsymbol{z}_{k+1},\boldsymbol{x})\mathbb{1}_{\tau>k} \;\middle|\; \boldsymbol{z}_k \in B\right] \\ &= \mathbb{E}_{i_k}\left[\operatorname{dist}^2(\boldsymbol{z}_{k+1},\boldsymbol{x}) \;\middle|\; \boldsymbol{z}_k \in B\right]\mathbb{1}_{\tau>k} \\ &\leq (1-0.03/n)\operatorname{dist}^2(\boldsymbol{z}_k,\boldsymbol{x})\mathbb{1}_{\tau>k}.\end{aligned}$$

Note that $\boldsymbol{z}_k \in B$ is an event with respect to randomness $\mathcal{I}^k$. Taking expectation gives

$$\begin{aligned}\mathbb{E}_{\mathcal{I}^{k+1}}\left[\operatorname{dist}^2(\boldsymbol{z}_{k+1},\boldsymbol{x})\mathbb{1}_{\tau>k+1}\right] &= \mathbb{E}_{\mathcal{I}^k}\left[\mathbb{E}_{i_k}\left[\operatorname{dist}^2(\boldsymbol{z}_{k+1},\boldsymbol{x})\mathbb{1}_{\tau>k+1} \;\middle|\; \boldsymbol{z}_k \in B\right]\right] \\ &\leq (1-0.03/n)\,\mathbb{E}_{\mathcal{I}^k}\left[\operatorname{dist}^2(\boldsymbol{z}_k,\boldsymbol{x})\mathbb{1}_{\tau>k}\right].\end{aligned}$$

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

For the second part, define $Y_k := \|\boldsymbol{z}_{k\wedge\tau} - \boldsymbol{x}\|_2^2$, where $k \wedge \tau := \min\{k, \tau\}$. Using the idea similar to that of Theorem 3.1 in [34], we can check that Y_k is a nonnegative supermartingale. It then follows from the supermartingale maximum inequality that

$$\mathbb{P}\left(\sup_{1\leq k<\infty} Y_k \geq 0.01^2\|\boldsymbol{x}\|_2^2\right) \leq \frac{Y_0}{0.01^2\|\boldsymbol{x}\|_2^2} \leq \delta^2.$$

This completes the proof. ∎

4. Proof of Theorem 3.2.

Proof of Theorem 3.2. For any $\boldsymbol{z} \in \mathbb{C}^n$, set $\boldsymbol{h} = e^{-i\phi(\boldsymbol{z})}\boldsymbol{z} - \boldsymbol{x}$, where $\phi(\boldsymbol{z})$ is defined in (2.3). Note that

$$|\boldsymbol{a}_j^*\boldsymbol{z}|^2 = |\boldsymbol{a}_j^*\boldsymbol{x}|^2 + 2\Re(\boldsymbol{h}^*\boldsymbol{a}_j\boldsymbol{a}_j^*\boldsymbol{x}) + |\boldsymbol{a}_j^*\boldsymbol{h}|^2.$$

Here $\Re(\cdot)$ denotes the real part for a complex number. It is easy to check that the function f given in (2.5) can be rewritten as

$$f(\boldsymbol{z}) = \frac{1}{m}\sum_{j=1}^m\left(\left|\boldsymbol{a}_j^*\boldsymbol{h}\right|^2 + 2\Re(\boldsymbol{h}^*\boldsymbol{a}_j\boldsymbol{a}_j^*\boldsymbol{x}) + 2\left|\boldsymbol{a}_j^*\boldsymbol{x}\right|^2 - 2\left|\boldsymbol{a}_j^*\boldsymbol{z}\right|\left|\boldsymbol{a}_j^*\boldsymbol{x}\right|\right).$$

To show that the function f satisfies the restricted strong convexity, from the definition it suffices to give a lower bound for $D_{\boldsymbol{z}-\boldsymbol{x}e^{i\phi(\boldsymbol{z})}}f(\boldsymbol{z}) - f(\boldsymbol{z})$. By some algebraic computation, we immediately have

$$
\begin{aligned}
&D_{\boldsymbol{z}-\boldsymbol{x}e^{i\phi(\boldsymbol{z})}} f(\boldsymbol{z}) - f(\boldsymbol{z}) \\
&= \frac{2}{m}\sum_{j=1}^m \left(1-\frac{|\boldsymbol{a}_j^*\boldsymbol{x}|}{|\boldsymbol{a}_j^*\boldsymbol{z}|}\right)\left(\left|\boldsymbol{a}_j^*\boldsymbol{h}\right|^2 + \Re(\boldsymbol{h}^*\boldsymbol{a}_j\boldsymbol{a}_j^*\boldsymbol{x})\right) - \frac{1}{m}\sum_{j=1}^m \left(|\boldsymbol{a}_j^*\boldsymbol{z}| - |\boldsymbol{a}_j^*\boldsymbol{x}|\right)^2 \\
&= \frac{1}{m}\sum_{j=1}^m \left|\boldsymbol{a}_j^*\boldsymbol{h}\right|^2 + \frac{2}{m}\sum_{j=1}^m \left(|\boldsymbol{a}_j^*\boldsymbol{z}|\,|\boldsymbol{a}_j^*\boldsymbol{x}| - |\boldsymbol{a}_j^*\boldsymbol{x}|^2 - \frac{|\boldsymbol{a}_j^*\boldsymbol{x}||\boldsymbol{a}_j^*\boldsymbol{h}|^2}{|\boldsymbol{a}_j^*\boldsymbol{z}|} - \frac{|\boldsymbol{a}_j^*\boldsymbol{x}|\Re(\boldsymbol{h}^*\boldsymbol{a}_j\boldsymbol{a}_j^*\boldsymbol{x})}{|\boldsymbol{a}_j^*\boldsymbol{z}|}\right) \\
&= \frac{1}{m}\sum_{j=1}^m \left|\boldsymbol{a}_j^*\boldsymbol{h}\right|^2 + \frac{2}{m}\sum_{j=1}^m \frac{|\boldsymbol{a}_j^*\boldsymbol{x}|^3 - |\boldsymbol{a}_j^*\boldsymbol{z}||\boldsymbol{a}_j^*\boldsymbol{x}|^2 + |\boldsymbol{a}_j^*\boldsymbol{x}|\Re(\boldsymbol{h}^*\boldsymbol{a}_j\boldsymbol{a}_j^*\boldsymbol{x})}{|\boldsymbol{a}_j^*\boldsymbol{z}|} \\
&= \frac{1}{m}\sum_{j=1}^m \left|\boldsymbol{a}_j^*\boldsymbol{h}\right|^2 + \frac{2}{m}\sum_{j=1}^m \frac{|\boldsymbol{a}_j^*\boldsymbol{z}||\boldsymbol{a}_j^*\boldsymbol{x}|\Re(\boldsymbol{h}^*\boldsymbol{a}_j\boldsymbol{a}_j^*\boldsymbol{x}) - |\boldsymbol{a}_j^*\boldsymbol{x}|^2\Re(\boldsymbol{h}^*\boldsymbol{a}_j\boldsymbol{a}_j^*\boldsymbol{x}) - |\boldsymbol{a}_j^*\boldsymbol{x}|^2|\boldsymbol{a}_j^*\boldsymbol{h}|^2}{|\boldsymbol{a}_j^*\boldsymbol{z}|(|\boldsymbol{a}_j^*\boldsymbol{z}| + |\boldsymbol{a}_j^*\boldsymbol{x}|)} \\
&= \frac{1}{m}\sum_{j=1}^m \left|\boldsymbol{a}_j^*\boldsymbol{h}\right|^2 - \frac{2}{m}\sum_{j=1}^m \frac{|\boldsymbol{a}_j^*\boldsymbol{x}|^2|\boldsymbol{a}_j^*\boldsymbol{h}|^2}{|\boldsymbol{a}_j^*\boldsymbol{z}|(|\boldsymbol{a}_j^*\boldsymbol{z}| + |\boldsymbol{a}_j^*\boldsymbol{x}|)} + \frac{2}{m}\sum_{j=1}^m \frac{|\boldsymbol{a}_j^*\boldsymbol{x}||\boldsymbol{a}_j^*\boldsymbol{h}|^2\Re(\boldsymbol{h}^*\boldsymbol{a}_j\boldsymbol{a}_j^*\boldsymbol{x})}{|\boldsymbol{a}_j^*\boldsymbol{z}|(|\boldsymbol{a}_j^*\boldsymbol{z}| + |\boldsymbol{a}_j^*\boldsymbol{x}|)^2} \\
&\quad + \frac{4}{m}\sum_{j=1}^m \frac{|\boldsymbol{a}_j^*\boldsymbol{x}|\Re^2(\boldsymbol{h}^*\boldsymbol{a}_j\boldsymbol{a}_j^*\boldsymbol{x})}{|\boldsymbol{a}_j^*\boldsymbol{z}|(|\boldsymbol{a}_j^*\boldsymbol{z}| + |\boldsymbol{a}_j^*\boldsymbol{x}|)^2}.
\end{aligned} \tag{4.1}
$$

Now we divide the indexes into two groups $j \in I_\alpha$ and $j \in I_\alpha^c$, where $I_\alpha := \left\{j : |\boldsymbol{a}_j^*\boldsymbol{x}| \ge \alpha|\boldsymbol{a}_j^*\boldsymbol{h}|\right\}$ for some fixed parameter $\alpha > 0$. For convenience, we denote $D_{\boldsymbol{z}-\boldsymbol{x}e^{i\phi(\boldsymbol{z})}} f(\boldsymbol{z}) - f(\boldsymbol{z}) := \frac{1}{m}\sum_{j=1}^m T_j$. We claim that for any $\alpha > 1$ it holds that

$$
T_j \ge \frac{4\alpha^3}{(\alpha+1)(2\alpha+1)^2}\cdot\frac{\Re^2(\boldsymbol{h}^*\boldsymbol{a}_j\boldsymbol{a}_j^*\boldsymbol{x})}{|\boldsymbol{a}_j^*\boldsymbol{x}|^2} - \frac{8\alpha^2-5\alpha+1}{(\alpha-1)(2\alpha-1)^2}\cdot|\boldsymbol{a}_j^*\boldsymbol{h}|^2 \quad \text{for } j \in I_\alpha \tag{4.2}
$$

and

$$
T_j \ge -3\left|\boldsymbol{a}_j^*\boldsymbol{h}\right|^2 \quad \text{for } j \in I_\alpha^c. \tag{4.3}
$$

This taken collectively with the identity (4.1) leads to a lower estimate

$$
\begin{aligned}
D_{\boldsymbol{z}-\boldsymbol{x}e^{i\phi(\boldsymbol{z})}} f(\boldsymbol{z}) - f(\boldsymbol{z}) \ge{}& \frac{4\alpha^3}{(\alpha+1)(2\alpha+1)^2}\cdot\frac{1}{m}\sum_{j\in I_\alpha}\frac{\Re^2(\boldsymbol{h}^*\boldsymbol{a}_j\boldsymbol{a}_j^*\boldsymbol{x})}{|\boldsymbol{a}_j^*\boldsymbol{x}|^2} \\
& - \frac{8\alpha^2-5\alpha+1}{(\alpha-1)(2\alpha-1)^2}\cdot\frac{1}{m}\sum_{j\in I_\alpha}|\boldsymbol{a}_j^*\boldsymbol{h}|^2 - \frac{3}{m}\sum_{j\in I_\alpha^c}\left|\boldsymbol{a}_j^*\boldsymbol{h}\right|^2,
\end{aligned} \tag{4.4}
$$

leaving us with three quantities in the right-hand side to deal with. Let $\rho := \|\boldsymbol{h}\|_2$. From the definition of $\boldsymbol{h}$, it is easy to check that $\Im(\boldsymbol{h}^*\boldsymbol{x}) = 0$. According to Lemma 7.2, we immediately obtain that for any $0 < \delta \le 1$ there exist universal constants $C, c > 0$ such that if $\alpha\rho \le 1/3$ and $m \ge C\delta^{-2}\log(1/\delta)n$, then with probability at least $1 - 6\exp(-c\delta^2 n)$ it holds that

$$
\frac{1}{m}\sum_{j\in I_\alpha}\frac{\Re^2(\boldsymbol{h}^*\boldsymbol{a}_j\boldsymbol{a}_j^*\boldsymbol{x})}{|\boldsymbol{a}_j^*\boldsymbol{x}|^2} \ge \frac{1}{n}\cdot\left(\frac{3}{8} - \frac{\alpha^2\rho^2}{(0.99+\alpha\rho)^2} - \delta\right)\|\boldsymbol{h}\|_2^2. \tag{4.5}
$$

For the second term, it follows from Lemma 7.1 that for $m \geq C\delta^{-2}n$, with probability at least $1-2\exp(-c\delta^2 n)$,

$$\frac{1}{m}\sum_{j=1}^{m} |\boldsymbol{a}_j^*\boldsymbol{h}|^2 \leq \frac{1+\delta}{n}\|\boldsymbol{h}\|_2^2. \tag{4.6}$$

Finally, for the third term, applying Lemma 7.3, we have that when $m \geq C\delta^{-2}\log(1/\delta)n$ and $0<\alpha\rho\leq 0.4$, with probability at least $1-6\exp(-c\delta^2 n)$,

$$\frac{1}{m}\sum_{j\in I_\alpha^c} \left|\boldsymbol{a}_j^*\boldsymbol{h}\right|^2 \leq \frac{1}{n}\cdot\left(\frac{2\alpha^2\rho^2}{0.99+\alpha^2\rho^2}+\delta\right)\|\boldsymbol{h}\|_2^2. \tag{4.7}$$

Setting $\alpha := 12$, $\delta := 0.001$ and putting (4.5), (4.6), (4.7) into (4.4), we obtain the conclusion that with probability at least $1-14\exp(-c_0 n)$ it holds that

$$D_{\boldsymbol{z}-\boldsymbol{x}e^{i\phi(\boldsymbol{z})}} f(\boldsymbol{z}) - f(\boldsymbol{z}) \geq \frac{0.03}{n}\|\boldsymbol{h}\|_2^2 \quad \text{for all} \quad \|\boldsymbol{h}\|_2 \leq 0.01,$$

provided $m \geq C_0 n$. 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 (4.1) that

$$T_j = \left|\boldsymbol{a}_j^*\boldsymbol{h}\right|^2 - \frac{2|\boldsymbol{a}_j^*\boldsymbol{x}|^2|\boldsymbol{a}_j^*\boldsymbol{h}|^2}{|\boldsymbol{a}_j^*\boldsymbol{z}|(|\boldsymbol{a}_j^*\boldsymbol{z}|+|\boldsymbol{a}_j^*\boldsymbol{x}|)} + \frac{2|\boldsymbol{a}_j^*\boldsymbol{x}||\boldsymbol{a}_j^*\boldsymbol{h}|^2\Re(\boldsymbol{h}^*\boldsymbol{a}_j\boldsymbol{a}_j^*\boldsymbol{x})}{|\boldsymbol{a}_j^*\boldsymbol{z}|(|\boldsymbol{a}_j^*\boldsymbol{z}|+|\boldsymbol{a}_j^*\boldsymbol{x}|)^2} + \frac{4|\boldsymbol{a}_j^*\boldsymbol{x}|\Re^2(\boldsymbol{h}^*\boldsymbol{a}_j\boldsymbol{a}_j^*\boldsymbol{x})}{|\boldsymbol{a}_j^*\boldsymbol{z}|(|\boldsymbol{a}_j^*\boldsymbol{z}|+|\boldsymbol{a}_j^*\boldsymbol{x}|)^2}.$$

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

$$(1-1/\alpha)|\boldsymbol{a}_j^*\boldsymbol{x}| \leq |\boldsymbol{a}_j^*\boldsymbol{x}| - |\boldsymbol{a}_j^*\boldsymbol{h}| \leq |\boldsymbol{a}_j^*\boldsymbol{z}| \leq |\boldsymbol{a}_j^*\boldsymbol{x}| + |\boldsymbol{a}_j^*\boldsymbol{h}| \leq (1+1/\alpha)|\boldsymbol{a}_j^*\boldsymbol{x}|. \tag{4.8}$$

Thus, the second term of T_j obeys

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

Similarly, the third term of T_j satisfies

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

Finally, using the upper bound in (4.8), we have

$$\begin{aligned}\frac{|\boldsymbol{a}_j^*\boldsymbol{x}|\Re^2(\boldsymbol{h}^*\boldsymbol{a}_j\boldsymbol{a}_j^*\boldsymbol{x})}{|\boldsymbol{a}_j^*\boldsymbol{z}|(|\boldsymbol{a}_j^*\boldsymbol{z}|+|\boldsymbol{a}_j^*\boldsymbol{x}|)^2} &= \frac{|\boldsymbol{a}_j^*\boldsymbol{x}|^3}{|\boldsymbol{a}_j^*\boldsymbol{z}|(|\boldsymbol{a}_j^*\boldsymbol{z}|+|\boldsymbol{a}_j^*\boldsymbol{x}|)^2}\cdot\frac{\Re^2(\boldsymbol{h}^*\boldsymbol{a}_j\boldsymbol{a}_j^*\boldsymbol{x})}{|\boldsymbol{a}_j^*\boldsymbol{x}|^2} \\ &\geq \frac{\alpha^3}{(\alpha+1)(2\alpha+1)^2}\cdot\frac{\Re^2(\boldsymbol{h}^*\boldsymbol{a}_j\boldsymbol{a}_j^*\boldsymbol{x})}{|\boldsymbol{a}_j^*\boldsymbol{x}|^2}.\end{aligned}$$

Collecting the above three estimators, we have

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

which proves the claim (4.2).

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

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

It then immediately gives

$$|T_j| \leq \frac{2\big||\boldsymbol{a}_j^* \boldsymbol{z}| - |\boldsymbol{a}_j^* \boldsymbol{x}|\big|}{|\boldsymbol{a}_j^* \boldsymbol{z}|} \cdot |\boldsymbol{h}^* \boldsymbol{a}_j \boldsymbol{a}_j^* \boldsymbol{z}| + \left(|\boldsymbol{a}_j^* \boldsymbol{z}| - |\boldsymbol{a}_j^* \boldsymbol{x}|\right)^2 \leq 3\left|\boldsymbol{a}_j^* \boldsymbol{h}\right|^2,$$

where we use the Cauchy–Schwarz inequality and the fact that $\big||\boldsymbol{a}_j^* \boldsymbol{z}| - |\boldsymbol{a}_j^* \boldsymbol{x}|\big| \leq |\boldsymbol{a}_j^* \boldsymbol{h}|$ in the last inequality. This completes the claim (4.3). ■

5. Numerical experiments. In this section, we demonstrate the efficiency and robustness of the randomized Kaczmarz method via a series of numerical experiments in comparison with WF [7], TWF [10], and TAF [39]. Here WF, TWF, and TAF are selected due to their wide applications and high efficiency for solving the phase retrieval problem. All experiments are carried out on a laptop computer with a 2.4GHz Intel Core i7 Processor, 8 GB 2133 MHz LPDDR3 memory, and MATLAB R2016a.

5.1. Recovery of signals with noiseless measurements. In our numerical experiments, the target vector $\boldsymbol{x} \in \mathbb{C}^n$ is chosen randomly from the standard complex Gaussian distribution, that is, $\boldsymbol{x} \sim \mathcal{N}(0, I_n) + i\mathcal{N}(0, I_n)$. The measurement vectors $\boldsymbol{a}_j$, $j = 1, \ldots, m$, are generated randomly from standard complex Gaussian distribution or the coded diffraction pattern (CDP) model. For the CDP model, we use masks of octanary patterns as in [7]. The code for WF, TWF, and TAF can be downloaded[1] with suggested parameters.

Example 5.1. In this example, we test the empirical success rate of the randomized Kaczmarz method versus the number of measurements. We set $n = 1000$. For the complex Gaussian case, we vary m within the range $[2n, 6n]$. For the CDP case, we set the number of masks $m/n = L$ from 2 to 6. For each m, we run 100 time trials to calculate the success rate. Here we say a trial has successfully reconstructed the target signal if the algorithm returns a vector $\boldsymbol{z}_T$ which has a small relative error, that is, when $\mathrm{dist}(\boldsymbol{z}_T - \boldsymbol{x})/\|\boldsymbol{x}\|_2 \leq 10^{-5}$. We use the initial point suggested in [7] for all the tested algorithms. The results are plotted in Figure 1. It can be seen that the empirical success rate of the randomized Kaczmarz method is comparable with that of TAF and even slightly better than TWF and WF.

[1] from https://viterbi-web.usc.edu/∼soltanol/WFcode.html, https://yuxinchen2020.github.io/Software.html, and https://gangwg.github.io/TAF/codes.html

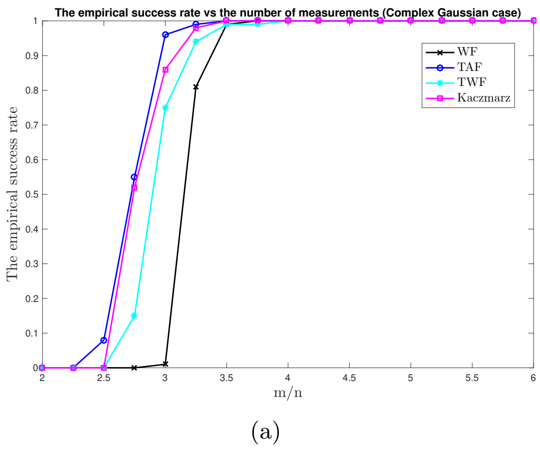


Figure 1. *The empirical success rate for different m/n based on* 100 *random trails.* (a) *Success rate for complex Gaussian case.* (b) *Success rate for CDP case.*

Example 5.2. In this example, we compare the computational complexity of the randomized Kaczmarz method with those of WF, TWF, and TAF for the complex Gaussian and CDP cases. We set $n = 1000$. For the complex Gaussian case, we choose $m = 5n$. For the CDP case, we choose the number of masks to be $L = 6$. Figure 2 shows the relative error versus the number of passes through the data (i.e., computational cost measured by the number of gradient computations divided by m). It can be seen that the randomized Kaczmarz method offers substantial improvements in computational complexity over state-of-the-art algorithms.

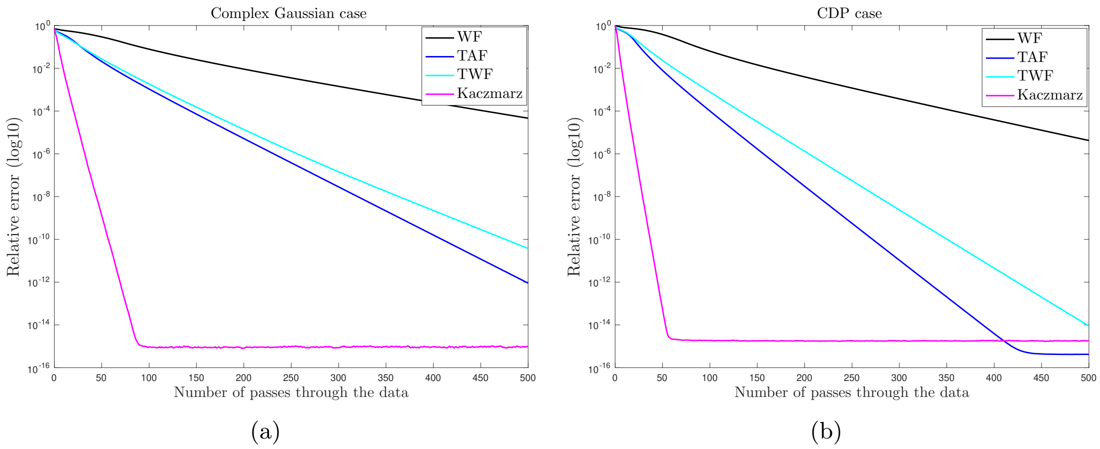
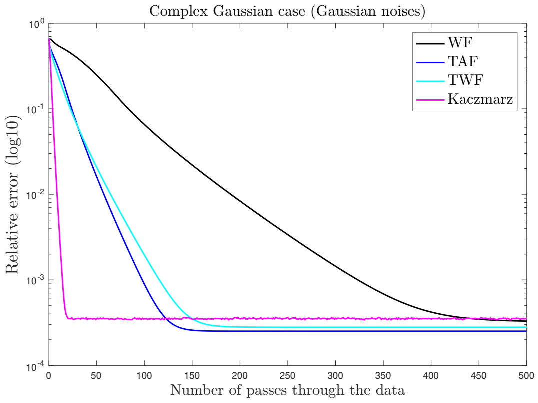


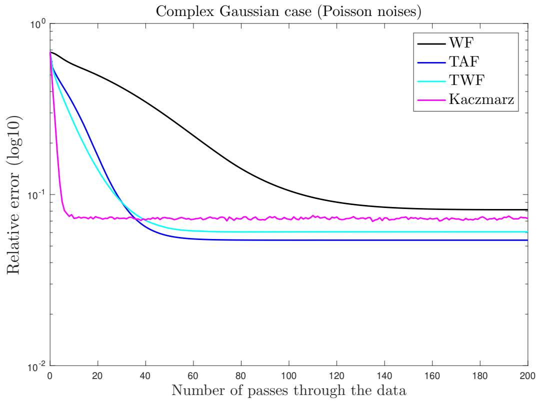
Figure 2. *Relative error versus number of passes for the randomized Kaczmarz, WF, TWF, and TAF methods:* (a) *The complex Gaussian case.* (b) *The CDP case.*

5.2. Robustness to Poisson and Gaussian noises. We now explore the performance of the randomized Kaczmarz method under noisy measurements. Two types of noise distributions are considered. One is Poisson noises, $b_j = \sqrt{\text{Poisson}(|\langle \boldsymbol{a}_j, \boldsymbol{x}_\zeta \rangle|^2)}$ for all $j = 1, \ldots, m$, where $\boldsymbol{x}_\zeta$ is the ground truth with $\boldsymbol{x}_\zeta = \zeta \boldsymbol{x}$. The other is additive white Gaussian noises, $b_j = |\langle \boldsymbol{a}_j, \boldsymbol{x} \rangle| + \eta_j$, where η_j are i.i.d. Gaussian random variables. The measurements $\boldsymbol{a}_j \in \mathbb{C}^n$ are complex Gaussian random vectors.

Example 5.3. In this example, we compare the computational complexity of the randomized Kaczmarz method with those of WF, TWF, and TAF under noisy measurements. We set $n = 1000$ and the number of measurements to be $m = 5n$. For Gaussian noises, $\eta_j \sim 0.01 \cdot N(0, 1)$. For Poisson noises, we choose $\zeta = 10$. The relative error versus the number of passes through the data is presented in Figure 3. We observe that all the algorithms reach almost the same accuracy. Moreover, the randomized Kaczmarz method requires the smallest number of passes to converge for both the Gaussian and the Poisson noises.



(a)



(b)

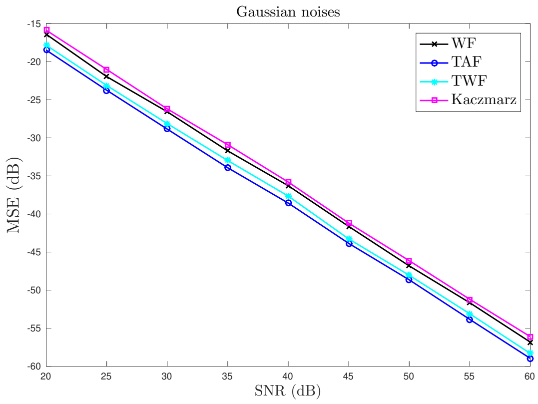
Figure 3. *Relative error versus number of passes for the randomized Kaczmarz, WF, TWF, and TAF methods under noisy measurements:* (a) *Gaussian noises.* (b) *Poisson noises.*

Example 5.4. In this example, we add different levels of Gaussian and Poisson noises to explore the relationship between the signal-to-noise rate (SNR) of the measurements and the mean square error (MSE) of the recovered signal. Specifically, SNR and MSE are evaluated by

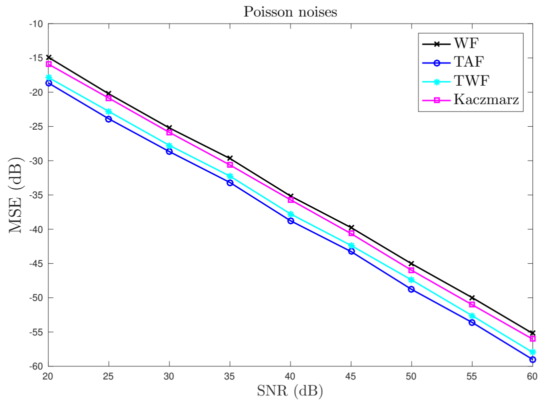
$$\text{MSE} := 10 \log_{10} \frac{\text{dist}^2(\boldsymbol{z}, \boldsymbol{x})}{\|\boldsymbol{x}\|^2} \quad \text{and} \quad \text{SNR} = 10 \log_{10} \frac{\sum_{i=1}^m |\boldsymbol{a}_j^* \boldsymbol{x}|^2}{\|\eta\|^2},$$

where $\boldsymbol{z}$ is the output of the algorithms. Here, for Poisson noises, we can think of $\eta_j := b_j - |\boldsymbol{a}_j^* \boldsymbol{x}|$ for all $j = 1, \ldots, m$, where $b_j = \sqrt{\text{Poisson}(|\boldsymbol{a}_j^* \boldsymbol{x}|^2)}$. We choose $n = 1000$ and $m = 5n$. The SNR varies from 20db to 60db. The results are shown in Figure 4. We can see that the randomized Kaczmarz method performs well for noisy phase retrieval.

5.3. Recovery of natural image. We next compare the performance of the randomized Kaczmarz method on recovering a natural image from masked Fourier intensity measurements. To take the advantage of the FFT, we consider a set of measurements corresponding to one mask instead of one measurement for each iteration of the randomized Kaczmarz method.



(a)



(b)

Figure 4. *SNR versus relative MSE on a dB-scale under the noisy measurements:* (a) *Gaussian noises.* (b) *Poisson noises.*

The image is the Milky Way Galaxy with resolution 1080×1920. The colored image has RGB channels. We use $L = 16$ random octanary patterns to obtain the Fourier intensity measurements for each R/G/B channel as in [7]. Table 1 lists the average time elapsed and the number of passes needed to achieve the relative errors 10^{-5} and 10^{-10} over the three RGB channels. We can see that the randomized Kaczmarz method runs faster than WF, TWF, and TAF. It outperforms the other three algorithms in both the number of passes and the computational time cost. Furthermore, the randomized Kaczmarz method performs well even with $L = 5$ under 100 passes. Figure 5 shows the image recovered by the randomized Kaczmarz method with $L = 5$.

Table 1
Time elapsed and number of passes among algorithms on recovery of galaxy image.

Algorithm	The Milky Way Galaxy			
	10^{-5}		10^{-10}	
	# Passes	Time(s)	# Passes	Time(s)
Kaczmarz	**7**	**87.8**	**10**	**93.4**
WF	161	356.7	241	467.6
TAF	71	228.8	132	368.2
TWF	82	280.1	145	378.6

5.4. Random initialization.

In this section, we investigate the performance of the randomized Kaczmarz method when the initial point is generated randomly according to the standard complex Gaussian distribution. We set $n = 1000$. The measurement vectors $\boldsymbol{a}_j$, $j = 1, \ldots, m$, are generated randomly from standard complex Gaussian distribution. Figure 6(a) compares the empirical success rate of the randomized Kaczmarz method under random and spectral initialization. It can be seen that the randomized Kaczmarz method with spectral initializa-

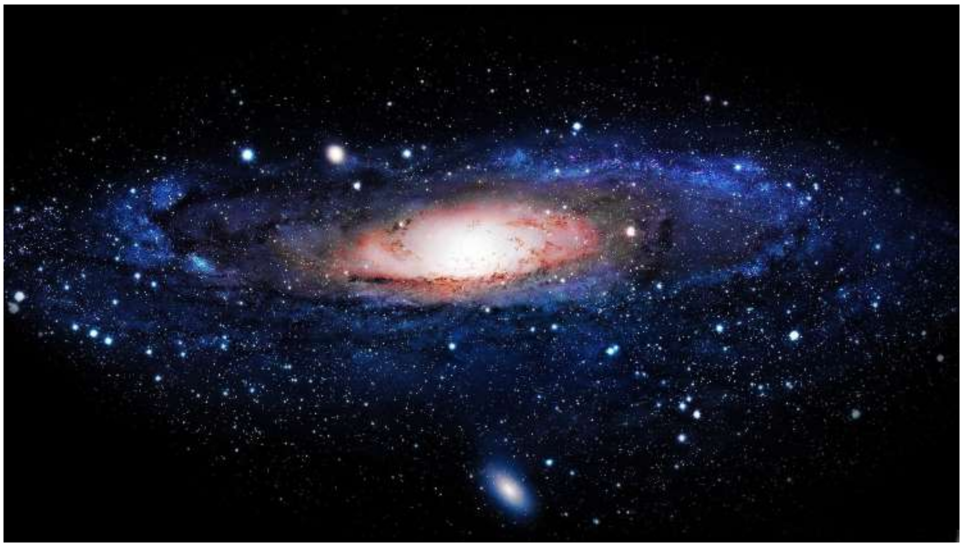
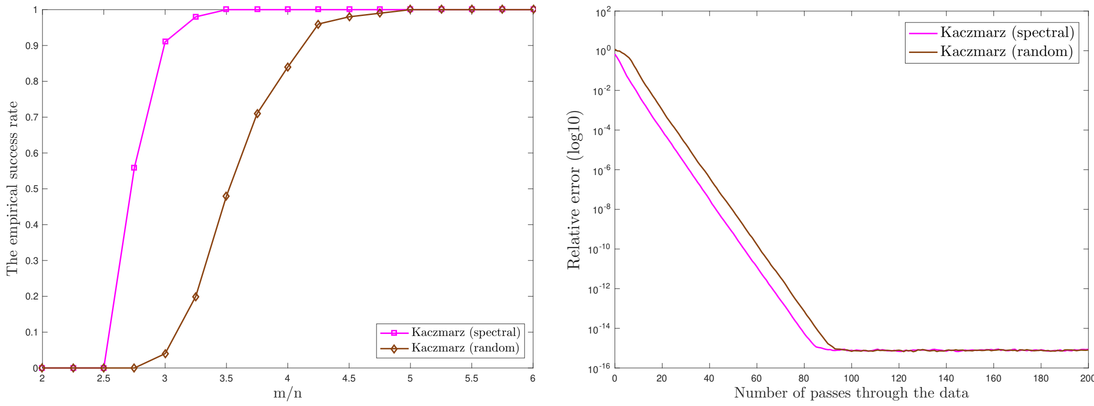


Figure 5. *The Milky Way Galaxy image: The randomized Kaczmarz method with $L = 5$ takes 100 passes, computation time is 64.7 s, and relative error is 1.36×10^{-15}.*



(a) (b)

Figure 6. *The performance of the randomized Kaczmarz method under random and spectral initialization:* (a) *The empirical success rate.* (b) *The relative error.*

tion has a higher success rate than the random initialization. However, it requires at most $1.5n$ more numbers of measurements for the random initialization to recover all the test signals. Figure 6(b) shows the relative error as a function of the number of passes through the data under $m = 5n$. The experiments show that the randomized Kaczmarz method is not sensitive to the initial point.

6. Discussions.

This paper considers convergence of the randomized Kaczmarz method for phase retrieval in the complex setting. A linear convergence rate has been established by combining a restricted strong convexity condition and tools from stochastic processes, 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 the randomized Kaczmarz method is also efficient for solving the Fourier phase retrieval problem, at least when the measurements follow the CDP model. It is of practical interest to provide some theoretical guarantees for it. Second, the convergence of the randomized Kaczmarz method relies on a spectral initialization. Some numerical evidence has shown that the randomized Kaczmarz method works well even if we start from a random initialization. It is interesting to provide some theoretical justifications for it.

7. Appendix.

The following lemma states that $\frac{1}{m}\sum_{j=1}^m \boldsymbol{a}_j\boldsymbol{a}_j^*$ is well behaved, provided $\boldsymbol{a}_j \in \mathbb{C}^n$ are drawn uniformly from the unit sphere. A similar result for the real case can be found in [35, Theorem 4.6.1].

Lemma 7.1. *Suppose that the vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m \in \mathbb{C}^n$ are drawn uniformly from the unit sphere $\mathbb{S}_{\mathbb{C}}^{n-1}$. For any $0 < \delta \le 1$, if $m \ge C\delta^{-2}n$, then with probability at least $1-2\exp(-c\delta^2 m)$ it holds*

$$\left\| \frac{1}{m}\sum_{j=1}^m \boldsymbol{a}_j\boldsymbol{a}_j^* - \frac{1}{n}\cdot I \right\|_2 \le \frac{\delta}{n}.$$

Here C and c are universal positive constants.

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

$$\left\| \frac{1}{m}\sum_{j=1}^m \boldsymbol{a}_j\boldsymbol{a}_j^* - \frac{1}{n}\cdot I \right\|_2 \le 2\max_{\boldsymbol{h}\in\mathcal{N}} \left| \frac{1}{m}\sum_{j=1}^m |\boldsymbol{a}_j^*\boldsymbol{h}|^2 - \frac{1}{n} \right|.$$

Here the cardinality $|\mathcal{N}| \le 9^{2n}$. Due to the unitary invariance of $\boldsymbol{a}_j$, for any fixed $\boldsymbol{h} \in \mathbb{S}_{\mathbb{C}}^{n-1}$ we have

$$\mathbb{E}|\boldsymbol{a}_j^*\boldsymbol{h}|^2 = \mathbb{E}|\boldsymbol{a}_j^*\boldsymbol{e}_1|^2 = \mathbb{E}|\boldsymbol{a}_j^*\boldsymbol{e}_2|^2 = \cdots = \mathbb{E}|\boldsymbol{a}_j^*\boldsymbol{e}_n|^2 = \frac{1}{n}\mathbb{E}\|\boldsymbol{a}_j\|_2^2 = \frac{1}{n},$$

where $\boldsymbol{e}_j \in \mathbb{C}^n$ are vectors whose jth entry is 1 and all the others entries are 0. It means that for any fixed $\boldsymbol{h} \in \mathbb{S}_{\mathbb{C}}^{n-1} \subset \mathbb{C}^n$ the terms $|\boldsymbol{a}_j^*\boldsymbol{h}|^2 - 1/n$ are independent, mean zero, subexponential random variables with their subexponential 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 $0 < \delta \le 1$ with probability at least $1 - 2\exp(-c_2\delta^2 m)$,

$$\left| \frac{1}{m}\sum_{j=1}^m |\boldsymbol{a}_j^*\boldsymbol{h}|^2 - \frac{1}{n} \right| \le \frac{\delta}{2n}$$

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

$$\left\| \frac{1}{m}\sum_{j=1}^m \boldsymbol{a}_j\boldsymbol{a}_j^* - \frac{1}{n}\cdot I \right\|_2 \le \frac{\delta}{n}$$

holds with probability at least

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

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

Lemma 7.2. *Let $\boldsymbol{x}$ be a vector in $\mathbb{C}^n$ with $\|\boldsymbol{x}\|_2=1$ and $\lambda\ge 3$. Assume that the vectors $\boldsymbol{a}_1,\dots,\boldsymbol{a}_m\in\mathbb{C}^n$ are drawn uniformly from the unit sphere $\mathbb{S}_{\mathbb{C}}^{n-1}$. For any fixed $0<\delta\le 1$, there exist universal constants $C,c>0$ such that if $m\ge C\delta^{-2}\log(1/\delta)n$, then with probability at least $1-6\exp(-c\delta^2 n)$ it holds that*

$$\frac{1}{m}\sum_{j=1}^m \frac{\Re^2(\boldsymbol{h}^*\boldsymbol{a}_j\boldsymbol{a}_j^*\boldsymbol{x})}{|\boldsymbol{a}_j^*\boldsymbol{x}|^2}\cdot\mathbb{1}_{\{\lambda|\boldsymbol{a}_j^*\boldsymbol{x}|\ge|\boldsymbol{a}_j^*\boldsymbol{h}|\}}\ge\frac{1}{n}\cdot\left(\frac{3}{8}-\frac{1}{(1+0.99\lambda)^2}-\delta\right)$$

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

Proof. We first prove the result for any fixed $\boldsymbol{h}$ and then apply an ε-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 $j=1,\dots,m$, define

$$\chi_j(t):=\begin{cases}1 & \text{if}\quad t\le 0.99\lambda|\boldsymbol{a}_j^*\boldsymbol{x}|;\\ -\frac{100}{\lambda|\boldsymbol{a}_j^*\boldsymbol{x}|}t+100 & \text{if}\quad 0.99\lambda|\boldsymbol{a}_j^*\boldsymbol{x}|\le t\le\lambda|\boldsymbol{a}_j^*\boldsymbol{x}|;\\ 0 & \text{otherwise.}\end{cases}$$

It then gives

$$\frac{\Re^2(\boldsymbol{h}^*\boldsymbol{a}_j\boldsymbol{a}_j^*\boldsymbol{x})}{|\boldsymbol{a}_j^*\boldsymbol{x}|^2}\cdot\mathbb{1}_{\{\lambda|\boldsymbol{a}_j^*\boldsymbol{x}|\ge|\boldsymbol{a}_j^*\boldsymbol{h}|\}}\ge\frac{\Re^2(\boldsymbol{h}^*\boldsymbol{a}_j\boldsymbol{a}_j^*\boldsymbol{x})}{|\boldsymbol{a}_j^*\boldsymbol{x}|^2}\chi_j(|\boldsymbol{a}_j^*\boldsymbol{h}|)\ge\frac{\Re^2(\boldsymbol{h}^*\boldsymbol{a}_j\boldsymbol{a}_j^*\boldsymbol{x})}{|\boldsymbol{a}_j^*\boldsymbol{x}|^2}\cdot\mathbb{1}_{\{0.99\lambda|\boldsymbol{a}_j^*\boldsymbol{x}|\ge|\boldsymbol{a}_j^*\boldsymbol{h}|\}}. \tag{7.1}$$

For any fixed $\boldsymbol{h}$, since $\boldsymbol{a}_1,\dots,\boldsymbol{a}_m$ are random vectors uniformly distributed on the unit sphere, it means that the terms $\frac{\Re^2(\boldsymbol{h}^*\boldsymbol{a}_j\boldsymbol{a}_j^*\boldsymbol{x})}{|\boldsymbol{a}_j^*\boldsymbol{x}|^2}\chi_j(|\boldsymbol{a}_j^*\boldsymbol{h}|)$ are independent subexponential random variables with the maximal subexponential norm $K=c_1/n$ for some universal constant $c_1>0$ [35, Theorem 3.4.6]. Applying Bernstein's inequality gives that for any fixed $0<\delta\le 1$ the following holds:

$$\frac{1}{m}\sum_{j=1}^m\frac{\Re^2(\boldsymbol{h}^*\boldsymbol{a}_j\boldsymbol{a}_j^*\boldsymbol{x})}{|\boldsymbol{a}_j^*\boldsymbol{x}|^2}\chi_j(|\boldsymbol{a}_j^*\boldsymbol{h}|)\ge\mathbb{E}\left(\frac{\Re^2(\boldsymbol{h}^*\boldsymbol{a}_1\boldsymbol{a}_1^*\boldsymbol{x})}{|\boldsymbol{a}_1^*\boldsymbol{x}|^2}\chi_1(|\boldsymbol{a}_1^*\boldsymbol{h}|)\right)-\frac{\delta}{4n} \tag{7.2}$$

with probability at least $1-2\exp(-c_2\delta^2 m)$, where c_2 is a universal positive constant.

Next, we give a uniform bound for the estimate (7.2). Construct an ε-net $\mathcal{N}$ over the unit sphere in $\mathbb{C}^n$ with cardinality $|\mathcal{N}|\le(1+\frac{2}{\varepsilon})^{2n}$. Then we have

$$\frac{1}{m}\sum_{j=1}^m\frac{\Re^2(\boldsymbol{h}^*\boldsymbol{a}_j\boldsymbol{a}_j^*\boldsymbol{x})}{|\boldsymbol{a}_j^*\boldsymbol{x}|^2}\chi_j(|\boldsymbol{a}_j^*\boldsymbol{h}|)\ge\mathbb{E}\left(\frac{\Re^2(\boldsymbol{h}^*\boldsymbol{a}_1\boldsymbol{a}_1^*\boldsymbol{x})}{|\boldsymbol{a}_1^*\boldsymbol{x}|^2}\chi_1(|\boldsymbol{a}_1^*\boldsymbol{h}|)\right)-\frac{\delta}{4n}\quad\text{for all}\quad\boldsymbol{h}\in\mathcal{N}$$

with probability at least

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

For any $\boldsymbol{h}$ with $\|\boldsymbol{h}\|_2 = 1$, there exists an $\boldsymbol{h}_0 \in \mathcal{N}$ such that $\|\boldsymbol{h} - \boldsymbol{h}_0\|_2 \leq \varepsilon$. We claim that there exist universal constants $C', c_3 > 0$ such that if $m \geq C'n$, then with probability at least $1 - 2\exp(-c_3 m)$ it holds that

$$\left| \frac{1}{m}\sum_{j=1}^{m} \frac{\Re^2(\boldsymbol{h}^*\boldsymbol{a}_j\boldsymbol{a}_j^*\boldsymbol{x})}{|\boldsymbol{a}_j^*\boldsymbol{x}|^2}\chi_j(|\boldsymbol{a}_j^*\boldsymbol{h}|) - \frac{1}{m}\sum_{j=1}^{m} \frac{\Re^2(\boldsymbol{h}_0^*\boldsymbol{a}_j\boldsymbol{a}_j^*\boldsymbol{x})}{|\boldsymbol{a}_j^*\boldsymbol{x}|^2}\chi_j(|\boldsymbol{a}_j^*\boldsymbol{h}_0|) \right| \leq \frac{205\varepsilon}{n}. \tag{7.3}$$

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

$$\frac{1}{m}\sum_{j=1}^{m} \frac{\Re^2(\boldsymbol{h}^*\boldsymbol{a}_j\boldsymbol{a}_j^*\boldsymbol{x})}{|\boldsymbol{a}_j^*\boldsymbol{x}|^2}\chi_j(|\boldsymbol{a}_j^*\boldsymbol{h}|) \geq \mathbb{E}\left(\frac{\Re^2(\boldsymbol{h}^*\boldsymbol{a}_1\boldsymbol{a}_1^*\boldsymbol{x})}{|\boldsymbol{a}_1^*\boldsymbol{x}|^2}\chi_1(|\boldsymbol{a}_1^*\boldsymbol{h}|)\right) - \frac{\delta}{2n} \quad \text{for all} \quad \|\boldsymbol{h}\|_2 = 1 \tag{7.4}$$

holds with probability at least

$$1 - 2\exp(-c_3 m) - 2\exp(-c_2\delta^2 m)\left(1 + \frac{2}{\varepsilon}\right)^{2n} \geq 1 - 4\exp(-c_4\delta^2 m),$$

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 (7.4), 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 7.4 that

$$\begin{aligned}
\mathbb{E}\left(\frac{\Re^2(\boldsymbol{h}^*\boldsymbol{a}_1\boldsymbol{a}_1^*\boldsymbol{x})}{|\boldsymbol{a}_1^*\boldsymbol{x}|^2} \cdot \mathbb{1}_{\{\lambda|\boldsymbol{a}_1^*\boldsymbol{x}| \geq |\boldsymbol{a}_1^*\boldsymbol{h}|\}}\right) &= \mathbb{E}\left(\frac{\Re^2(\boldsymbol{h}^*\xi\xi^*\boldsymbol{x})}{\|\xi\|_2^2|\xi^*\boldsymbol{x}|^2} \cdot \mathbb{1}_{\{\lambda|\xi^*\boldsymbol{x}| \geq |\xi^*\boldsymbol{h}|\}}\right) \\
&\geq \frac{1}{(1+3\delta_1)n} \cdot \mathbb{E}\left(\frac{\Re^2(\boldsymbol{h}^*\xi\xi^*\boldsymbol{x})}{|\xi^*\boldsymbol{x}|^2} \cdot \mathbb{1}_{\{\lambda|\xi^*\boldsymbol{x}| \geq |\xi^*\boldsymbol{h}|\}}\right) \\
&\geq \frac{1}{(1+3\delta_1)n} \cdot \left(\frac{3}{8} - \frac{1}{(\lambda+1)^2}\right).
\end{aligned}$$

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 that

$$\mathbb{E}\left(\frac{\Re^2(\boldsymbol{h}^*\boldsymbol{a}_1\boldsymbol{a}_1^*\boldsymbol{x})}{|\boldsymbol{a}_1^*\boldsymbol{x}|^2} \cdot \mathbb{1}_{\{\lambda|\boldsymbol{a}_1^*\boldsymbol{x}| \geq |\boldsymbol{a}_1^*\boldsymbol{h}|\}}\right) \geq \frac{1}{n} \cdot \left(\frac{3}{8} - \frac{1}{(\lambda+1)^2} - \frac{\delta}{2}\right), \tag{7.5}$$

where $c_6 > 0$ is a universal constant. Collecting (7.1), (7.4), and (7.5) together, we obtain the conclusion that for any $\lambda \geq 3$ with probability at least $1 - 6\exp(-c\delta^2 n)$ it holds that

$$\frac{\Re^2(\boldsymbol{h}^*\boldsymbol{a}_j\boldsymbol{a}_j^*\boldsymbol{x})}{|\boldsymbol{a}_j^*\boldsymbol{x}|^2} \cdot \mathbb{1}_{\{\lambda|\boldsymbol{a}_j^*\boldsymbol{x}| \geq |\boldsymbol{a}_j^*\boldsymbol{h}|\}} \geq \frac{1}{n} \cdot \left(\frac{3}{8} - \frac{1}{(1+0.99\lambda)^2} - \delta\right),$$

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

Finally, it remains to prove the claim (7.3). To this end, we claim that for all $j = 1, \ldots, m$ it holds that

$$\begin{aligned}&\left|\frac{\Re^2(\boldsymbol{h}^*\boldsymbol{a}_j\boldsymbol{a}_j^*\boldsymbol{x})}{|\boldsymbol{a}_j^*\boldsymbol{x}|^2}\chi_j(|\boldsymbol{a}_j^*\boldsymbol{h}|) - \frac{\Re^2(\boldsymbol{h}_0^*\boldsymbol{a}_j\boldsymbol{a}_j^*\boldsymbol{x})}{|\boldsymbol{a}_j^*\boldsymbol{x}|^2}\chi_j(|\boldsymbol{a}_j^*\boldsymbol{h}_0|)\right| \\ &\leq 101|\boldsymbol{a}_j^*\boldsymbol{h}||\boldsymbol{a}_j^*(\boldsymbol{h}-\boldsymbol{h}_0)| + 101|\boldsymbol{a}_j^*\boldsymbol{h}_0||\boldsymbol{a}_j^*(\boldsymbol{h}-\boldsymbol{h}_0)|.\end{aligned} \tag{7.6}$$

Indeed, from the definition of $\chi_j(t)$, if both $|\boldsymbol{a}_j^*\boldsymbol{h}| > \lambda|\boldsymbol{a}_j^*\boldsymbol{x}|$ and $|\boldsymbol{a}_j^*\boldsymbol{h}_0| > \lambda|\boldsymbol{a}_j^*\boldsymbol{x}|$, then the above inequality holds directly. Thus, we only need to consider the case where $|\boldsymbol{a}_j^*\boldsymbol{h}| \leq \lambda|\boldsymbol{a}_j^*\boldsymbol{x}|$ or $|\boldsymbol{a}_j^*\boldsymbol{h}_0| \leq \lambda|\boldsymbol{a}_j^*\boldsymbol{x}|$. Without loss of generality, we assume $|\boldsymbol{a}_j^*\boldsymbol{h}| \leq \lambda|\boldsymbol{a}_j^*\boldsymbol{x}|$. Then we have

$$\begin{aligned}&\left|\frac{\Re^2(\boldsymbol{h}^*\boldsymbol{a}_j\boldsymbol{a}_j^*\boldsymbol{x})}{|\boldsymbol{a}_j^*\boldsymbol{x}|^2}\chi_j(|\boldsymbol{a}_j^*\boldsymbol{h}|) - \frac{\Re^2(\boldsymbol{h}_0^*\boldsymbol{a}_j\boldsymbol{a}_j^*\boldsymbol{x})}{|\boldsymbol{a}_j^*\boldsymbol{x}|^2}\chi_j(|\boldsymbol{a}_j^*\boldsymbol{h}_0|)\right| \\ &\leq |\boldsymbol{a}_j^*\boldsymbol{h}|^2\left|\chi_j(|\boldsymbol{a}_j^*\boldsymbol{h}|) - \chi_j(|\boldsymbol{a}_j^*\boldsymbol{h}_0|)\right| + \left(|\boldsymbol{a}_j^*\boldsymbol{h}| + |\boldsymbol{a}_j^*\boldsymbol{h}_0|\right)|\boldsymbol{a}_j^*(\boldsymbol{h}-\boldsymbol{h}_0)| \\ &\leq \frac{100|\boldsymbol{a}_j^*\boldsymbol{h}|^2}{\lambda|\boldsymbol{a}_j^*\boldsymbol{x}|}|\boldsymbol{a}_j^*(\boldsymbol{h}-\boldsymbol{h}_0)| + \left(|\boldsymbol{a}_j^*\boldsymbol{h}| + |\boldsymbol{a}_j^*\boldsymbol{h}_0|\right)|\boldsymbol{a}_j^*(\boldsymbol{h}-\boldsymbol{h}_0)| \\ &\leq 101|\boldsymbol{a}_j^*\boldsymbol{h}||\boldsymbol{a}_j^*(\boldsymbol{h}-\boldsymbol{h}_0)| + |\boldsymbol{a}_j^*\boldsymbol{h}_0||\boldsymbol{a}_j^*(\boldsymbol{h}-\boldsymbol{h}_0)|,\end{aligned}$$

which gives (7.6). According to Lemma 7.1, we obtain that for $m \geq C'n$ with probability at least $1 - 2\exp(-c_3 n)$ it holds that

$$\begin{aligned}&\left|\frac{1}{m}\sum_{j=1}^m \frac{\Re^2(\boldsymbol{h}^*\boldsymbol{a}_j\boldsymbol{a}_j^*\boldsymbol{x})}{|\boldsymbol{a}_j^*\boldsymbol{x}|^2}\chi_j(|\boldsymbol{a}_j^*\boldsymbol{h}|) - \frac{1}{m}\sum_{j=1}^m \frac{\Re^2(\boldsymbol{h}_0^*\boldsymbol{a}_j\boldsymbol{a}_j^*\boldsymbol{x})}{|\boldsymbol{a}_j^*\boldsymbol{x}|^2}\chi_j(|\boldsymbol{a}_j^*\boldsymbol{h}_0|)\right| \\ &\leq \frac{101}{m}\sum_{j=1}^m |\boldsymbol{a}_j^*\boldsymbol{h}||\boldsymbol{a}_j^*(\boldsymbol{h}-\boldsymbol{h}_0)| + \frac{101}{m}\sum_{j=1}^m |\boldsymbol{a}_j^*\boldsymbol{h}_0||\boldsymbol{a}_j^*(\boldsymbol{h}-\boldsymbol{h}_0)| \\ &\leq 101\sqrt{\frac{1}{m}\sum_{j=1}^m |\boldsymbol{a}_j^*\boldsymbol{h}|^2}\sqrt{\frac{1}{m}\sum_{j=1}^m |\boldsymbol{a}_j^*(\boldsymbol{h}-\boldsymbol{h}_0)|^2} + 101\sqrt{\frac{1}{m}\sum_{j=1}^m |\boldsymbol{a}_j^*\boldsymbol{h}_0|^2}\sqrt{\frac{1}{m}\sum_{j=1}^m |\boldsymbol{a}_j^*(\boldsymbol{h}-\boldsymbol{h}_0)|^2} \\ &\leq \frac{205\varepsilon}{n},\end{aligned}$$

which proves the claim (7.3). ■

Lemma 7.3. *Let $\boldsymbol{x}$ be a vector in $\mathbb{C}^n$ with $\|\boldsymbol{x}\|_2 = 1$ and $0 < \lambda \leq 0.4$. Assume that the vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m \in \mathbb{C}^n$ are drawn uniformly from the unit sphere $\mathbb{S}_{\mathbb{C}}^{n-1}$. 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 $1 - 6\exp(-c\delta^2 n)$, it holds that*

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

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

Proof. Due to the non-Lipschitz property of indicator functions, we introduce a series of auxiliary random Lipschitz functions to approximate them. For any $j = 1, \ldots, m$, define

$$\chi_j(t) := \begin{cases} t & \text{if} \quad t \geq |\boldsymbol{a}_j^* \boldsymbol{x}|^2/\lambda^2; \\ 100t - \frac{99|\boldsymbol{a}_j^*\boldsymbol{x}|^2}{\lambda^2} & \text{if} \quad 0.99|\boldsymbol{a}_j^*\boldsymbol{x}|^2/\lambda^2 \leq t \leq |\boldsymbol{a}_j^*\boldsymbol{x}|^2/\lambda^2; \\ 0 & \text{otherwise.} \end{cases}$$

Then it is easy to check that

$$\frac{1}{m}\sum_{j=1}^m |\boldsymbol{a}_j^*\boldsymbol{h}|^2 \cdot \mathbb{1}_{\{|\boldsymbol{a}_j^*\boldsymbol{x}| \leq \lambda |\boldsymbol{a}_j^*\boldsymbol{h}|\}} \leq \frac{1}{m}\sum_{j=1}^m \chi_j(|\boldsymbol{a}_j^*\boldsymbol{h}|^2) \leq \frac{1}{m}\sum_{j=1}^m |\boldsymbol{a}_j^*\boldsymbol{h}|^2 \cdot \mathbb{1}_{\{0.99|\boldsymbol{a}_j^*\boldsymbol{x}| \leq \lambda |\boldsymbol{a}_j^*\boldsymbol{h}|\}}. \tag{7.7}$$

For any fixed $\boldsymbol{h}$, since $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m$ are drawn uniformly from the unit sphere $\mathbb{S}_{\mathbb{C}}^{n-1}$, the terms $\chi_j(|\boldsymbol{a}_j^*\boldsymbol{h}|^2)$ are independent subexponential random variables with the maximal subexponential 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 $1 - 2\exp(-c_2\delta^2 m)$, it holds that

$$\left| \frac{1}{m}\sum_{j=1}^m \chi_j(|\boldsymbol{a}_j^*\boldsymbol{h}|^2) - \mathbb{E}\left[\chi_1(|\boldsymbol{a}_1^*\boldsymbol{h}|^2)\right] \right| \leq \frac{\delta}{4n}, \tag{7.8}$$

where c_2 is a universal positive constant.

To give a uniform bound for the estimate (7.8), we construct an ε-net $\mathcal{N}$ over the unit sphere in $\mathbb{C}^n$ with cardinality $|\mathcal{N}| \leq (1 + \frac{2}{\varepsilon})^{2n}$. Then, for any $\boldsymbol{h}$ with $\|\boldsymbol{h}\|_2 = 1$, there exists an $\boldsymbol{h}_0 \in \mathcal{N}$ such that $\|\boldsymbol{h} - \boldsymbol{h}_0\|_2 \leq \varepsilon$. Note that $\chi_j(t)$ is a Lipschitz function with Lipschitz constant 100. It then follows from Lemma 7.1 that for $m \geq C'n$ with probability at least $1 - 2\exp(-c_3 m)$ it holds that

$$\begin{aligned} &\left| \frac{1}{m}\sum_{j=1}^m \chi_j(|\boldsymbol{a}_j^*\boldsymbol{h}|^2) - \frac{1}{m}\sum_{j=1}^m \chi_j(|\boldsymbol{a}_j^*\boldsymbol{h}_0|^2) \right| \\ &\leq \frac{100}{m}\sum_{j=1}^m \left(|\boldsymbol{a}_j^*\boldsymbol{h}| + |\boldsymbol{a}_j^*\boldsymbol{h}_0|\right) |\boldsymbol{a}_j^*(\boldsymbol{h} - \boldsymbol{h}_0)| \\ &\leq 100\sqrt{\frac{1}{m}\sum_{j=1}^m |\boldsymbol{a}_j^*\boldsymbol{h}|^2}\sqrt{\frac{1}{m}\sum_{j=1}^m |\boldsymbol{a}_j^*(\boldsymbol{h} - \boldsymbol{h}_0)|^2} + 100\sqrt{\frac{1}{m}\sum_{j=1}^m |\boldsymbol{a}_j^*\boldsymbol{h}_0|^2}\sqrt{\frac{1}{m}\sum_{j=1}^m |\boldsymbol{a}_j^*(\boldsymbol{h} - \boldsymbol{h}_0)|^2} \\ &\leq \frac{202\varepsilon}{n}, \end{aligned}$$

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

$$\frac{1}{m}\sum_{j=1}^m \chi_j(|\boldsymbol{a}_j^*\boldsymbol{h}|^2) \geq \mathbb{E}\left(\frac{1}{m}\sum_{j=1}^m \chi_1(|\boldsymbol{a}_1^*\boldsymbol{h}|^2)\right) - \frac{\delta}{2n} \quad \text{for all} \quad \|\boldsymbol{h}\|_2 = 1 \tag{7.9}$$

holds with probability at least

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

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 that

$$\mathbb{E}\left(|\xi^*\boldsymbol{h}|^2\cdot\mathbb{1}_{\{|\xi^*\boldsymbol{x}|\leq\lambda|\xi^*\boldsymbol{h}|\}}\right)\leq\frac{2\lambda^2}{\lambda^2+1}. \tag{7.10}$$

Note that $\xi/\|\xi\|_2$ is a vector uniformly distributed on the unit sphere, and the inequality $\|\xi\|_2\leq(1-\delta_0)\sqrt{n}$ holds for any fixed $0\leq\delta_0\leq1$ with probability at least $1-2\exp(-c_5\delta_0^2n)$ [35, Theorem 3.1.1]. It then gives that for any $0<\lambda\leq\sqrt{\frac{5-\sqrt{21}}{2}}$ with probability at least $1-2\exp(-c_5\delta_0^2n)$ we have

$$\begin{aligned}\mathbb{E}\left(|\boldsymbol{a}_1^*\boldsymbol{h}|^2\cdot\mathbb{1}_{\{|\boldsymbol{a}_1^*\boldsymbol{x}|\leq\lambda|\boldsymbol{a}_1^*\boldsymbol{h}|\}}\right)&=\mathbb{E}\left(\frac{|\xi^*\boldsymbol{h}|^2}{\|\xi\|_2^2}\cdot\mathbb{1}_{\{|\xi^*\boldsymbol{x}|\leq\lambda|\xi^*\boldsymbol{h}|\}}\right)\\&\leq\frac{1}{(1-\delta_0)n}\cdot\mathbb{E}\left(|\xi^*\boldsymbol{h}|^2\cdot\mathbb{1}_{\{|\xi^*\boldsymbol{x}|\leq\lambda|\xi^*\boldsymbol{h}|\}}\right)\\&\leq\frac{1}{(1-\delta_0)n}\cdot\frac{2\lambda^2}{\lambda^2+1}.\end{aligned}\tag{7.11}$$

Taking the constant $\delta_0=\delta/3$, it then follows from (7.7), (7.9), and (7.11) that for any fixed $0<\lambda\leq0.4$, with probability at least $1-6\exp(-c\delta^2n)$, it holds that

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

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 (7.10). Indeed, due to the unitary invariance of the Gaussian random vector, without loss of generality, we assume $\boldsymbol{h}=\boldsymbol{e}_1$ and $\boldsymbol{x}=\sigma\boldsymbol{e}_1+\tau e^{i\phi}\boldsymbol{e}_2$, where $\sigma=\boldsymbol{h}^*\boldsymbol{x}\in\mathbb{R}$, $|\sigma|\leq1$, and $\tau=\sqrt{1-\sigma^2}$. Let ξ_1,ξ_2 be the first and second entries of ξ. 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 $|\xi^*\boldsymbol{x}|\leq\lambda|\xi^*\boldsymbol{h}|$ 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 (7.10), we take the polar coordinates transformations and denote

$$\left.\begin{aligned}\xi_{1,\Re}&=r_1\cos\theta_1\\\xi_{1,\Im}&=r_1\sin\theta_1\\\sigma\xi_{1,\Re}+\tau(\cos\phi\xi_{2,\Re}+\sin\phi\xi_{2,\Im})&=r_2\cos\theta_2\\\sigma\xi_{1,\Im}-\tau(\sin\phi\xi_{2,\Re}-\cos\phi\xi_{2,\Im})&=r_2\sin\theta_2\end{aligned}\right\}$$

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

$$\begin{aligned} G(\lambda, \sigma) &:= \mathbb{E}\left(|\xi^* \boldsymbol{h}|^2 \cdot \mathbb{1}_{\{|\xi^* \boldsymbol{x}| \le \lambda |\xi^* \boldsymbol{h}|\}}\right) \\ &= \frac{1}{\pi^2} \int_0^{2\pi} \int_0^{2\pi} \int_0^{+\infty} \int_0^{\lambda \cdot r_1} \frac{r_1^3 r_2}{\tau^2} e^{-(r_1^2 + r_2^2)/\tau^2} \cdot e^{2\sigma r_1 r_2 \cos(\theta_1 - \theta_2)/\tau^2} dr_2 dr_1 d\theta_1 d\theta_2. \end{aligned}$$

It gives

$$\begin{aligned} \frac{\partial G(\lambda, \sigma)}{\partial \lambda} &= \frac{1}{\pi^2} \int_0^{2\pi} \int_0^{2\pi} \int_0^{+\infty} \lambda \cdot r_1^5/\tau^2 \cdot e^{-(1+\lambda^2) r_1^2/\tau^2} \cdot e^{2\sigma \lambda r_1^2 \cos(\theta_1 - \theta_2)/\tau^2} dr_1 d\theta_1 d\theta_2 \\ &= \frac{1}{\pi^2} \int_0^{2\pi} \int_0^{2\pi} \frac{\lambda \tau^4}{(1 + \lambda^2 - 2\lambda\sigma \cos(\theta_1 - \theta_2))^3} d\theta_1 d\theta_2 \\ &= 4\tau^4 \cdot \frac{\lambda(1 + \lambda^4 + 2\lambda^2 + 2\lambda^2\sigma^2)}{\sqrt{(1 + \lambda^2 + 2\lambda\sigma)^5 (1 + \lambda^2 - 2\lambda\sigma)^5}} \\ &\le \frac{2\tau^4(\mu_+^2 + \mu_-^2)}{(\mu_+ \mu_-)^{5/2}}, \end{aligned}$$

where $\mu_+ := 1 + \lambda^2 + 2\lambda\sigma \ge 0$ and $\mu_- := 1 + \lambda^2 - 2\lambda\sigma \ge 0$. 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 σ for any fixed $\lambda \le \sqrt{\frac{5 - \sqrt{21}}{2}}$. In fact, through some basic algebraic computation, we have

$$\begin{aligned} \frac{\partial f(\lambda, \sigma)}{\partial \sigma} &= (1 - \sigma^2) \cdot \frac{\lambda(1 - \sigma^2)(\mu_+ - \mu_-)(5\mu_+^2 + 4\mu_+\mu_- + 5\mu_-^2) - 4\sigma(\mu_+^2 + \mu_-^2)\mu_+\mu_-}{(\mu_+\mu_-)^{7/2}} \\ &\le (1 - \sigma^2)\sigma(\mu_+^2 + \mu_-^2) \cdot \frac{28\lambda^2(1 - \sigma^2) - 4\mu_+\mu_-}{(\mu_+\mu_-)^{7/2}} \\ &= (1 - \sigma^2)\sigma(\mu_+^2 + \mu_-^2) \cdot \frac{28\lambda^2 - 12\lambda^2\sigma^2 - 4(1 + \lambda^2)^2}{(\mu_+\mu_-)^{7/2}} \\ &\le 0, \end{aligned}$$

provided $\lambda \le \sqrt{\frac{5 - \sqrt{21}}{2}}$. Note that $G(0, \sigma) = 0$. It then immediately gives

$$G(\lambda, \sigma) \le 2 \int_0^\lambda f(t, \sigma) dt \le 2 \int_0^\lambda f(t, 0) dt = 4 \int_0^\lambda \frac{t}{(1 + t^2)^3} dt = \frac{\lambda^2(\lambda^2 + 2)}{(\lambda^2 + 1)^2},$$

which completes the claim (7.10). ■

Lemma 7.4. *Assume $\lambda \ge 2.95$. Let $\boldsymbol{x}, \boldsymbol{h}$ be two fixed vectors in $\mathbb{C}^n$ with $\|\boldsymbol{x}\|_2 = \|\boldsymbol{h}\|_2 = 1$ and $\Im(\boldsymbol{h}^* \boldsymbol{x}) = 0$. Suppose $\xi \in \mathbb{C}^n$ is a complex Gaussian random vector. Then we have*

$$\mathbb{E}\left(\frac{\Re^2(\boldsymbol{h}^* \xi \xi^* \boldsymbol{x})}{|\xi^* \boldsymbol{x}|^2} \cdot \mathbb{1}_{\{\lambda |\xi^* \boldsymbol{x}| \ge |\xi^* \boldsymbol{h}|\}}\right) \ge \frac{3}{8} - \frac{1}{(\lambda + 1)^2}.$$

Proof. Due to the unitary invariance of the Gaussian random vector, without loss of generality, we assume $\boldsymbol{h} = \boldsymbol{e}_1$ and $\boldsymbol{x} = \sigma \boldsymbol{e}_1 + \tau e^{i\phi} \boldsymbol{e}_2$, where $\sigma = \boldsymbol{h}^* \boldsymbol{x} \in \mathbb{R}$, $|\sigma| \le 1$, and $\tau = \sqrt{1-\sigma^2}$. Let ξ_1, ξ_2 be the first and second entries of ξ. 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 $\lambda |\xi^* \boldsymbol{x}| \ge |\xi^* \boldsymbol{h}|$ is equivalent to

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

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

$$\left.\begin{aligned} \xi_{1,\Re} &= r_1 \cos\theta_1 \\ \xi_{1,\Im} &= r_1 \sin\theta_1 \\ \sigma \xi_{1,\Re} + \tau(\cos\phi \xi_{2,\Re} + \sin\phi \xi_{2,\Im}) &= r_2 \cos\theta_2 \\ \sigma \xi_{1,\Im} - \tau(\sin\phi \xi_{2,\Re} - \cos\phi \xi_{2,\Im}) &= r_2 \sin\theta_2 \end{aligned}\right\}$$

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

$$\begin{aligned} \Re(\boldsymbol{h}^* \xi \xi^* \boldsymbol{x}) &= \Re\left((\sigma \xi_1 + \tau e^{-i\phi} \xi_2) \bar{\xi}\right) \\ &= \xi_{1,\Re} \left(\sigma \xi_{1,\Re} + \tau(\cos\phi \xi_{2,\Re} + \sin\phi \xi_{2,\Im})\right) + \xi_{1,\Im} \left(\sigma \xi_{1,\Im} - \tau(\sin\phi \xi_{2,\Re} - \cos\phi \xi_{2,\Im})\right) \\ &= r_1 r_2 \cos(\theta_1 - \theta_2). \end{aligned}$$

It means the expectation can be written as

$$\begin{aligned} F(\lambda, \sigma) \; &:= \; \mathbb{E}\left(\frac{\Re^2(\boldsymbol{h}^* \xi \xi^* \boldsymbol{x})}{|\xi^* \boldsymbol{x}|^2} \cdot \mathbb{1}_{\{\lambda |\xi^* \boldsymbol{x}| \ge |\xi^* \boldsymbol{h}|\}}\right) \\ &= \frac{1}{\pi^2} \int_0^{2\pi} \int_0^{2\pi} \int_0^{+\infty} \int_0^{\lambda \cdot r_2} \frac{r_1^3 r_2}{\tau^2} \\ &\quad \cdot \cos^2(\theta_1 - \theta_2) \cdot e^{-(r_1^2 + r_2^2)/\tau^2} \cdot e^{2\sigma r_1 r_2 \cos(\theta_1 - \theta_2)/\tau^2} dr_1 dr_2 d\theta_1 d\theta_2 \\ &= 2 \sum_{k=0}^{\infty} \frac{2k+1}{(k!)^2 \cdot (k+1)} \cdot \frac{\sigma^{2k}}{\tau^{4k+2}} \int_0^{+\infty} \int_0^{\lambda \cdot r_2} r_1^{2k+3} r_2^{2k+1} \cdot e^{-(r_1^2 + r_2^2)/\tau^2} dr_1 dr_2, \end{aligned}$$

where the last equation follows from the fact that

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

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

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

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

$$F(\lambda,\sigma)=2\sum_{k=0}^{\infty}\frac{(2k+1)!(2k+1)}{(k!)^2}\cdot\sigma^{2k}(1-\sigma^2)^2\cdot\int_0^{\lambda}\left(\frac{t}{1+t^2}\right)^{2k+3}dt. \tag{7.12}$$

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

$$\begin{aligned}I&=\int_0^{\frac{\pi}{4}}\sin^{2k+3}\theta\cos^{2k+1}\theta\ d\theta\\&=\frac{1}{2^{2k+2}}\int_0^{\frac{\pi}{4}}\sin^{2k+1}(2\theta)(1-\cos(2\theta))\ d\theta\\&=\frac{k!}{(2k+1)!!\cdot 2^{k+3}}-\frac{1}{2(k+1)\cdot 2^{2k+3}}.\end{aligned}\tag{7.13}$$

For the second term, noting that $\lambda\geq 1$, we have

$$II\geq\int_1^{\lambda}(1+t)^{-2k-3}dt=\frac{1}{2(k+1)}\cdot\left(\frac{1}{2^{2k+2}}-\frac{1}{(\lambda+1)^{2k+2}}\right). \tag{7.14}$$

Putting (7.13) and (7.14) into (7.12), we have

$$F(\lambda,\sigma)\geq\sum_{k=0}^{\infty}\frac{(2k+1)!!(2k+1)}{(k+1)!}\cdot\sigma^{2k}(1-\sigma^2)^2\cdot\left(\frac{(k+1)!}{4(2k+1)!!}+\frac{1}{2^{k+3}}-\frac{2^k}{(1+\lambda)^{2k+2}}\right).$$

Let $\beta:=(1+\lambda)^2$. Expand $F(\lambda,\sigma)$ into a series with respect to σ, and we have

$$F(\lambda,\sigma)\geq\frac{3}{8}-\frac{1}{\beta}+\frac{9}{16}\sigma^2-\sum_{k=1}^{\infty}\frac{(2k-1)!!(2k+7)}{2^{k+4}(k+2)!}\sigma^{2k+2}-\left(1-\frac{4}{\beta}\right)^2\cdot\sum_{k=1}^{\infty}\frac{2^k(2k-1)!!}{(k-1)!\beta^k}\sigma^{2k+2}, \tag{7.15}$$

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 that

$$\frac{(2k-1)!!}{2^k k!}\leq\frac{1}{\sqrt{2k}}.$$

Thus,

$$\begin{aligned}\sum_{k=1}^{\infty}\frac{(2k-1)!!(2k+7)}{2^{k+4}(k+2)!}\sigma^{2k+2} &\le \frac{\sigma^2}{16}\cdot\sum_{k=1}^{\infty}\frac{2k+7}{\sqrt{2k}(k+2)(k+1)}\\ &\le \frac{\sigma^2}{8}\cdot\sum_{k=1}^{\infty}\frac{1}{k^{3/2}}\\ &\le \frac{\sigma^2}{8}\left(1+\int_1^{\infty}\frac{1}{t^{3/2}}\,dt\right)\\ &= \frac{3}{8}\sigma^2, \qquad (7.16)\end{aligned}$$

where the second inequality follows from the fact that $\frac{2k+7}{(k+2)(k+1)} \le \frac{2\sqrt{2}}{k}$ for all $k \ge 1$.

On the other hand, using Wallis' inequality again, we have

$$\begin{aligned}\left(1-\frac{4}{\beta}\right)^2\cdot\sum_{k=1}^{\infty}\frac{2^k(2k-1)!!}{(k-1)!\beta^k}\sigma^{2k+2} &\le \left(1-\frac{4}{\beta}\right)^2\cdot\sigma^2\cdot\sum_{k=1}^{\infty}\frac{4^k\cdot k}{\sqrt{2k}\beta^k}\\ &\le \left(1-\frac{4}{\beta}\right)^2\cdot\frac{\sigma^2}{\sqrt{2}}\cdot\sum_{k=1}^{\infty}k\left(\frac{4}{\beta}\right)^k\\ &= \frac{2\sqrt{2}}{\beta}\sigma^2, \qquad (7.17)\end{aligned}$$

where the last equation follows from the fact that

$$\sum_{k=1}^{\infty}kx^{k-1}=\frac{1}{(1-x)^2}\quad \text{for all}\quad 0\le x<1.$$

Putting (7.16) and (7.17) into (7.15), we know that for $\lambda \ge 2.95$ it holds that

$$F(\lambda,\sigma)\ge\frac{3}{8}-\frac{1}{\beta}.$$

This completes the proof. ∎

REFERENCES

[1] R. Balan, *On signal reconstruction from its spectrogram*, in 44th Annual Conference on Information Sciences and Systems, 2010, pp. 1–4, https://doi.org/10.1109/CISS.2010.5464828.

[2] R. Balan, P. Casazza, and D. Edidin, *On signal reconstruction without phase*, Appl. Comput. Harmon. Anal., 20 (2006), pp. 345–356, https://doi.org/10.1016/j.acha.2005.07.001.

[3] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*, SIAM, Philadelphia, 2001, https://doi.org/10.1137/1.9780898718829.

[4] T. Bendory, Y. C. Eldar, and N. Boumal, *Non-convex phase retrieval from STFT measurements*, IEEE Trans. Inform. Theory, 64 (2018), pp. 467–484, https://doi.org/10.1109/TIT.2017.2745623.

[5] 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., 58 (2022), pp. 60–84, https://doi.org/10.1016/j.acha.2022.01.002.

[6] E. J. Candès and X. Li, *Solving quadratic equations via phaselift when there are about as many equations as unknowns*, Found. Comput. Math., 14 (2014), pp. 1017–1026, https://doi.org/10.1007/s10208-013-9162-z.

[7] E. J. Candès, X. Li, and M. Soltanolkotabi, *Phase retrieval via Wirtinger flow: Theory and algorithms*, IEEE Trans. Inform. Theory, 61 (2015), pp. 1985–2007, https://doi.org/10.1109/TIT.2015.2399924.

[8] E. J. Candès, T. Strohmer, and V. Voroninski, *PhaseLift: Exact and stable signal recovery from magnitude measurements via convex programming*, Comm. Pure Appl. Math., 66 (2013), pp. 1241–1274, https://doi.org/10.1002/cpa.21432.

[9] A. Chai, M. Moscoso, and G. Papanicolaou, *Array imaging using intensity-only measurements*, Inverse Problems, 27 (2010), 015005, https://doi.org/10.1088/0266-5611/27/1/015005.

[10] Y. Chen and E. J. Candès, *Solving random quadratic systems of equations is nearly as easy as solving linear systems*, Comm. Pure Appl. Math., 70 (2017), pp. 822–883.

[11] A. Conca, D. Edidin, M. Hering, and C. Vinzant, *An algebraic characterization of injectivity in phase retrieval*, Appl. Comput. Harmon. Anal., 38 (2015), pp. 346–356, https://doi.org/10.1016/j.acha.2014.06.005.

[12] J. Dainty and C. Fienup, *Phase retrieval and image reconstruction for astronomy*, in Image Recovery: Theory and Application, Academic Press, New York, 1987, pp. 231–275.

[13] F. Deutsch, *Rate of convergence of the method of alternating projections*, in Parametric Optimization and Approximation, Birkhäuser, Basel, 1985, pp. 96–107.

[14] F. Deutsch and H. Hundal, *The rate of convergence for the method of alternating projections.* II, J. Math. Anal. Appl., 205 (1997), pp. 381–405, https://doi.org/10.1006/jmaa.1997.5202.

[15] A. Galántai, *On the rate of convergence of the alternating projection method in finite dimensional spaces*, J. Math. Anal. Appl., 310 (2005), pp. 30–44, https://doi.org/10.1016/j.jmaa.2004.12.050.

[16] B. Gao, X. Sun, Y. Wang, and Z. Xu, *Perturbed amplitude flow for phase retrieval*, IEEE Trans. Signal Process., 68 (2020), pp. 5427–5440, https://doi.org/10.1109/TSP.2020.3022817.

[17] B. Gao and Z. Xu, *Phaseless recovery using the Gauss-Newton method*, IEEE Trans. Signal Process., 65 (2017), pp. 5885–5896, https://doi.org/10.1109/TSP.2017.2742981.

[18] R. W. Harrison, *Phase problem in crystallography*, J. Opt. Soc. Am. A, 10 (1993), pp. 1046–1055, https://doi.org/10.1364/JOSAA.10.001046.

[19] M. Huang and Z. Xu, *Solving systems of quadratic equations via exponential-type gradient descent algorithm*, J. Comput. Math., 38 (2020), pp. 638–660, https://doi.org/10.4208/jcm.1902-m2018-0109.

[20] K. Jaganathan, Y. C. Eldar, and B. Hassibi, *Phase retrieval: An overview of recent developments*, in Optical Compressive Imaging, CRC Press, Boca Raton, FL, 2017, pp. 263–296.

[21] H. Jeong and C. S. Güntürk, *Convergence of the Randomized Kaczmarz Method for Phase Retrieval*, preprint, https://arxiv.org/abs/1706.10291, 2017.

[22] S. Kaczmarz, *Angenäherte Auflösung von Systemen linearer Gleichungen*, Bull. Int. Acad. Pol. Sci. Lett. A, 35 (1937), pp. 355–357.

[23] N. D. Kazarinoff, *Analytic Inequalities*, Courier Corporation, North Chelmsford, MA, 2014.

[24] G. Li, Y. Gu, and Y. M. Lu, *Phase retrieval using iterative projections: Dynamics in the large systems limit*, in 53rd Annual Allerton Conference on Communication, Control, and Computing, 2015, pp. 1114–1118, https://doi.org/10.1109/ALLERTON.2015.7447133.

[25] W. Luo, W. Alghamdi, and Y. M. Lu, *Optimal spectral initialization for signal recovery with applications to phase retrieval*, IEEE Trans. Signal Process., 67 (2019), pp. 2347–2356, https://doi.org/10.1109/TSP.2019.2904918.

[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., 59 (2008), pp. 387–410, https://doi.org/10.1146/annurev.physchem.59.032607.093642.

[27] R. P. Millane, *Phase retrieval in crystallography and optics*, J. Opt. Soc. Am. A, 7 (1990), pp. 394–411, https://doi.org/10.1364/JOSAA.7.000394.

[28] M. Mondelli and A. Montanari, *Fundamental limits of weak recovery with applications to phase retrieval*, Found. Comput. Math., 19 (2019), pp. 703–773.

[29] P. Netrapalli, P. Jain, and S. Sanghavi, *Phase retrieval using alternating minimization*, IEEE Trans. Signal Process., 63 (2015), pp. 4814–4826, https://doi.org/10.1109/TSP.2015.2448516.

[30] H. Sahinoglou and S. Cabrera, *On phase retrieval of finite-length sequences using the initial time sample*, IEEE Trans. Circuits Syst., 38 (1991), pp. 954–958, https://doi.org/10.1109/31.85639.

[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., 32 (2015), pp. 87–109, https://doi.org/10.1109/MSP.2014.2352673.

[32] T. Strohmer and R. Vershynin, *A randomized Kaczmarz algorithm with exponential convergence*, J. Fourier Anal. Appl., 15 (2009), pp. 262–278, https://doi.org/10.1007/s00041-008-9030-4.

[33] J. Sun, Q. Qu, and J. Wright, *A geometric analysis of phase retrieval*, Found. Comput. Math., 18 (2018), pp. 1131–1198, https://doi.org/10.1007/s10208-017-9365-9.

[34] Y. S. Tan and R. Vershynin, *Phase retrieval via randomized Kaczmarz: Theoretical guarantees*, Inf. Inference, 8 (2019), pp. 97–123, https://doi.org/10.1093/imaiai/iay005.

[35] R. Vershynin, *High-Dimensional Probability: An Introduction with Applications in Data Science*, Camb. Ser. Stat. Probab. Math. 47, Cambridge University Press, Cambridge, UK, 2018.

[36] I. Waldspurger, *Phase retrieval with random Gaussian sensing vectors by alternating projections*, IEEE Trans. Inform. Theory, 64 (2018), pp. 3301–3312, https://doi.org/10.1109/TIT.2018.2800663.

[37] I. Waldspurger, A. d'Aspremont, and S. Mallat, *Phase recovery, MaxCut and complex semidefinite programming*, Math. Program., 149 (2015), pp. 47–81, https://doi.org/10.1007/s10107-013-0738-9.

[38] A. Walther, *The question of phase retrieval in optics*, Optica Acta, 10 (1963), pp. 41–49.

[39] G. Wang, G. B. Giannakis, and Y. C. Eldar, *Solving systems of random quadratic equations via truncated amplitude flow*, IEEE Trans. Inform. Theory, 64 (2018), pp. 773–794, https://doi.org/10.1109/TIT.2017.2756858.

[40] Y. Wang and Z. Xu, *Generalized phase retrieval: Measurement number, matrix recovery and beyond*, Appl. Comput. Harmon. Anal., 47 (2019), pp. 423–446, https://doi.org/10.1016/j.acha.2017.09.003.

[41] K. Wei, *Solving systems of phaseless equations via Kaczmarz methods: A proof of concept study*, Inverse Problems, 31 (2015), 125008, https://doi.org/10.1088/0266-5611/31/12/125008.

[42] 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., 18 (2017), 141.