Phase Retrieval by Alternating Minimization With Random Initialization

Teng Zhang

Abstract— **We consider the phase retrieval problem, where the goal is to reconstruct an n-dimensional complex vector from its phaseless scalar products with m sensing vectors, independently sampled from complex normal distributions. We show that, if $m \geq Mn^{3/2}\log^{7/2} n$ for some $M > 0$, then the classical algorithm of alternating minimization with random initialization succeeds with high probability as $n, m \to \infty$. This is a step toward proving the conjecture in, which conjectures that the algorithm succeeds when $m = O(n)$. The analysis depends on an approach that enables the decoupling of the dependency between the algorithmic iterates and the sensing vectors.**

Index Terms— **Iterative algorithms, convergence of numerical methods.**

I. Introduction

THIS article concerns the phase retrieval problem as follows: let $\mathbf{z} \in \mathbb{C}^n$ be an unknown vector, and given m known sensing vectors $\{\mathbf{a}_i\}_{i=1}^m \in \mathbb{C}^n$, we have the observations

$$y_i = |\mathbf{a}_i^* \mathbf{z}|, i = 1, 2, \cdots, m.$$

Then can we reconstruct $\mathbf{z}$ from the observations $\{y_i\}_{i=1}^m$? In this work, we assume that the sensing vectors $\{\mathbf{a}_i\}_{i=1}^m$ are sampled from a complex normal distribution $CN(0, \mathbf{I})$. That is, their real component and imaginary components are independent and follow a real Gaussian distribution $N(0, \mathbf{I}/2)$.

This problem is motivated by the applications in imaging science, and we refer interested readers to [6], [12], [24] for more detailed discussions on the background in engineering and additional applications in other areas of sciences and engineering.

Because of the practical ubiquity of the phase retrieval problem, many algorithms and theoretical analyses have been developed for this problem. For example, an interesting recent approach is based on convex relaxation [7], [8], [29], that replaces the non-convex measurements by convex measurements through relaxation. Since the associated optimization problem is convex, it is possible to solve it in polynomial time, and it has been shown that under some assumptions on the sensing vectors, this method recovers the correct $\mathbf{z}$ [5], [17]. However, since these algorithms involve semidefinite programming for $n \times n$ positive semidefinite matrices, the computational cost is prohibitive when n is large. Recently, several works [1], [16], [18], [19], [23] proposed and analyzed an alternate convex method that uses linear programming instead of semidefinite programming, which is more computationally efficient, but the program itself requires an "anchor vector", which needs to be a good approximate estimation of $\mathbf{z}$.

Another line of works are based on Wirtinger flows, i.e., gradient flow in the complex setting [4], [6], [9], [10], [25], [30]–[32]. Some theoretical justifications are also provided [6], [25], and in particular, the geometric analysis in [26] allows random initialization to be used with this method. However, this method requires choosing step sizes, which makes the implementation slightly more complicated. Most existing theoretical analyses assume sufficiently small step sizes.

The most widely used method is perhaps the alternate minimization (Gerchberg-Saxton) algorithm and its variants [13]–[15], that is based on alternating projections onto nonconvex sets [2]. As a result, in some literature, it is also called the alternating projection method [28]. This method is very simple to implement and is parameter-free. However, since it is a nonconvex algorithm, its properties such as convergence are only partially known. Netrapalli *et al.* [22] studied a resampled version of this algorithm and established its convergence as the number of measurements m goes to infinity when the measurement vectors are independent standard complex normal vectors. Marchesini *et al.* [20] studied and demonstrated the necessary and sufficient conditions for the local convergence of this algorithm. Recently, Waldspurger [28] showed that when $m \geq Cn$ for sufficiently large C, the alternating minimization algorithm succeeds with high probability, provided that the algorithm is carefully initialized. This work also conjectured that the alternate minimizations algorithm with random initialization succeeds with $m \geq C'n$ for sufficiently large C'.

One particular difficulty in the analysis of the alternating minimization algorithm is the stationary points. Currently, most papers on nonconvex algorithms depend on the analysis showing that all (attractive) stationary points of the algorithm are well-behaved in the sense that it is the desired solution, or close to the desired solution, for example, [26]. Then standard algorithms such as gradient descent algorithm or trust-region method can be applied to the problem to obtain the stationary point. However, as pointed out in [28], in the regime $m = O(n)$, the alternating minimization algorithm has

Manuscript received December 4, 2018; revised January 21, 2020; accepted January 26, 2020. Date of publication February 7, 2020; date of current version June 18, 2020. This work was supported by the National Science Foundation (NSF) under Grant CNS-1739736.

The author is with the Department of Mathematics, University of Central Florida, Orlando, FL 32765 USA (e-mail: teng.zhang@ucf.edu).

Communicated by A. Rudra, Associate Editor for Complexity.

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2020.2971211

attractive stationary points that are not the desired solution. While empirically these undesired stationary points are not obstacles for the success of the algorithm since their attraction basins seem small, it prevents us from applying the common approach of analyzing stationary points.

Recently, [33] shows that the algorithm improves the correlation between the estimator and the truth in each iteration with high probability. Based on this observation, it shows that a resampled version of the alternating minimization algorithm converges to the solution with high probability when $m = O(n\log^5 n)$. However, this approach can not be applied to analyze the alternating minimization algorithm directly, since the estimator at the k-th iteration is correlated with the sensing vectors. As a result, to analyze the non-resampled version, one needs to find a way to decouple the estimator at the k-th iteration and the sensing vectors.

We remark that there are also algorithms based on Douglas-Rachford splitting [12], which is popular in practice, but we skip detailed introductions and comparisons as they lack strong theoretical guarantees as the other works reviewed here.

The contribution of this work is to show that the alternating minimization algorithm with random initialization succeeds with high probability when $m > Mn^{1.5}\log^{3.5} n$. While it does not match the conjecture of $m = O(n)$, it is an improvement over the result of $m > Cn^2$ in [28]. Compared with [33], which analyzes a resampled version of the alternating minimization algorithm, this work introduces an approach that decouples of the sensing vectors and the estimator at the k-th iteration, by fixing the first $k-1$ algorithmic iterates and analyzing the conditional distribution of the sensing vectors. This approach, inspired by the analysis of LASSO in [3], is the main technical contribution of this work. In spirit, this contribution is very similar to the leave-one-out approach that also enables decoupling in [10], where the authors show that an algorithm for the phase retrieval converges linearly based on the leave-one-out approach. However, the analyzed algorithm is very different and their work assumes that the sensing vectors and the signal $\mathbf{z}$ are real-valued. Besides, it seems more difficult to apply the leave-one-out approach for the alternating minimization algorithm, as the update formula is more complicated.

The paper is organized as follows. Section I-B presents the algorithm and the main result of the paper, Theorem 1. The proofs are given in Section II, where the proof of Theorem 1 is given in Section II-B, the proof of the main lemmas are given in Section II-C, and the auxiliary lemmas and their proofs are given in Section II-D.

A. Notations

For any $z \in \mathbb{C}$, $|z|$ represents the modulus of z and $\mathrm{phase}(z) = z/|z|$ represents the phase of z. We use $\mathrm{Sp}(\mathbf{a}_1, \cdots, \mathbf{a}_m)$ to represent the subspace spanned by $\mathbf{a}_1, \cdots, \mathbf{a}_n$, i.e., the set $\{\mathbf{x} \in \mathbb{C}^n : \mathbf{x} = \sum_{i=1}^m c_i\mathbf{a}_i,$ for some $c_1, \cdots, c_m \in \mathbb{C}\}$. Note that this subspace is slightly different from the standard subspace in $\mathbb{R}^n$, by allowing the coefficient of each vector to be a complex number. We use P_L to denote the projection onto the subspace L: $P_L(\mathbf{z})$ is the nearest point on L to $\mathbf{z}$.

For any vector $\mathbf{z} = (z_1, \cdots, z_m)$, $\mathrm{phase}(\mathbf{z})$ is the vector whose coordinates are the phases of the coordinates of $\mathbf{z}$:

$$\mathrm{phase}(\mathbf{z}) = (\mathrm{phase}(z_1), \cdots, \mathrm{phase}(z_m)).$$

We use $\odot$ to denote the pointwise product between the phase of the first vector and the modulus of the second vector. That is,

$$(\mathbf{w} \odot \mathbf{y})_i = \frac{w_i}{|w_i|}|y_i|.$$

For any vector $\mathbf{z} \in \mathbb{C}^m$, $\|\mathbf{z}\|$ represents its Euclidean norm: $\|\mathbf{z}\| = \sqrt{\sum_{i=1}^m |z_i|^2}$, and its 1-norm and ∞-norm are defined by $\|\mathbf{z}\|_1 = \sum_{i=1}^m |z_i|$ and $\|\mathbf{z}\|_\infty = \max_{1\leq i\leq m} |z_i|$.

B. Algorithm and Main Result

The alternating minimization method is one of the earliest methods that was introduced for phase retrieval problems [13]–[15], and it is based on alternating projections onto nonconvex sets [2]. Let $\mathbf{A} \in \mathbb{C}^{m\times n}$ be a matrix with columns given by $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_m$, then its goal is to find a vector in $\mathbb{C}^m$ such that it lies in both the subspace $L = \mathrm{range}(\mathbf{A}) \in \mathbb{C}^m$ and the set of correct amplitude $\mathcal{A} = \{\mathbf{w} \in \mathbb{C}^m : |\mathbf{w}_i| = y_i,$ for $i = 1, \cdots, m\}$. For this purpose, the algorithm picks an initial guess $\mathbf{x}^{(1)}$ in $\mathbb{C}^n$ and alternatively projects $\mathbf{A}\mathbf{x}^{(1)}$ to both sets. Define the projections $P_L, P_{\mathcal{A}} : \mathbb{C}^m \to \mathbb{C}^m$ by

$$P_L(\mathbf{w}) = \mathbf{A}(\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*\mathbf{w}, \;\; [P_{\mathcal{A}}(\mathbf{w})]_i = y_i\frac{\mathbf{w}_i}{|\mathbf{w}_i|},$$

and the alternating minimization algorithm is given by iteratively applying the operator $P_LP_{\mathcal{A}}$ to the vector $\mathbf{w}^{(1)} = \mathbf{A}\mathbf{x}^{(1)}$, i.e.,

$$\mathbf{w}^{(k+1)} = P_LP_{\mathcal{A}}\mathbf{w}^{(k)}. \tag{1}$$

Then the estimator of $\mathbf{x}$ at the k-th iteration is obtained by solving $\mathbf{w}^{(k)} = \mathbf{A}\mathbf{x}^{(k)}$.

This algorithm has been studied in [28] and Theorem 2 in [28] shows the convergence of the algorithm if $m > Mn$ and if $\mathbf{x}^{(1)}$ is a good initialization. Besides, it conjectures that random initialization also succeeds in this setting. In this article, we prove that this conjecture holds when $m > Mn^{1.5}\log^{3.5} n$ for some $M > 0$. The rigorous statement is as follows:

Theorem 1: Assuming that the sensing vectors $\{\mathbf{a}_i\}_{i=1}^m$ are i.i.d. sampled from the complex normal distribution $CN(0, \mathbf{I})$, there exists $M > 0$ such that if $m \geq Mn^{3/2}\log^{7/2} n$, then the alternating projection algorithm with random initialization (obtained from a uniform distribution on the sphere of $\mathbb{C}^n$) succeeds almost surely in the sense that

$$\Pr\left(\lim_{k\to\infty}\inf_{\psi\in\mathbb{R}} \|e^{i\psi}\mathbf{x}^{(k)} - \mathbf{z}\| = 0\right) \to 1, \quad \text{as } n, m \to \infty. \tag{2}$$

Specifically, there are two stages in the convergence of $\mathbf{x}^{(k)}$ to $\mathbf{z}$. In the first stage, the correlation between $\mathbf{w}^{(k)}$ and $\mathbf{A}\mathbf{z}$ increases linearly: with probability goes to 1 as $n, m \to \infty$, there exist constants $C > 1$ and $C'_d > 0$ such that C does not depend on n, C'_d depends on n, $C'_d \le O(\log n)$, and

$$\Pr\left(\frac{|\mathbf{w}^{(k+1)*}\mathbf{A}\mathbf{z}|}{\|\mathbf{w}^{(k+1)}\|\|\mathbf{A}\mathbf{z}\|} \ge C\frac{|\mathbf{w}^{(k)*}\mathbf{A}\mathbf{z}|}{\|\mathbf{w}^{(k)}\|\|\mathbf{A}\mathbf{z}\|}, \text{ for all } 1 \le k < C'_d\right) \to 1, \text{ as } n, m \to \infty. \tag{3}$$

Note that $\mathbf{w}^{(k)} = \mathbf{A}\mathbf{x}^{(k)}$, it implies that the correlation between $\mathbf{A}\mathbf{x}^{(k)}$ and $\mathbf{A}\mathbf{z}$ increases in the first C'_d iterations.

In the second stage, the distance between $\mathbf{x}^{(k)}$ and $\mathbf{z}$ decreases linearly: there exists $0 < \delta < 1$ such that

$$\Pr\left(\min_{0\le\psi\le 2\pi}\|e^{i\psi}\mathbf{z} - \mathbf{x}^{(k+1)}\| \le \delta \min_{0\le\psi\le 2\pi}\|e^{i\psi}\mathbf{z} - \mathbf{x}^{(k)}\|, \text{ for all } k \ge C'_d\right) \to 1, \text{ as } n, m \to \infty. \tag{4}$$

In the proof, for simplicity when we talk about a "random unit vector in $\mathbb{C}^m$/subspace L", we implicitly assume that it is sampled from the uniform distribution on the unit sphere in $\mathbb{C}^m$ or the unit sphere in subspace L. The constants c, C are used to represent a constant that is independent of m and n, and it is used to represent different constants in different equations. In addition, since the theorem focus on the setting when n and m are both large, we sometimes apply inequalities that hold only under this assumption. For example, we may write $\log^3 n < n$ even though it only holds for large n.

II. Proof of Theorem 1

In the proof, we will first present a reduced form of the statement of Theorem 1 in Section II-A, and then present the proof of this reduced statement in Section II-B. The proof of the main lemmas are given in Section II-C, and the auxiliary lemmas (which are mostly generic results on measure concentration) and their proofs are given in Section II-D.

A. An Equivalent Form of Theorem 1

In this section, we introduce some modifications to the algorithm, which do not impact the performance of the algorithm but will simplify the proof later.

First, we investigate the performance of the same algorithm if the sensing matrix $\mathbf{A}$, the underlying signal $\mathbf{z}$ and the initialization $\mathbf{x}^{(1)}$ are replaced by $\tilde{\mathbf{A}} = \mathbf{A}\mathbf{D}$, $\tilde{\mathbf{z}} = \mathbf{D}^{-1}\mathbf{z}$, and $\tilde{\mathbf{x}}^{(1)} = \mathbf{D}^{-1}\mathbf{x}^{(1)}$ respectively, for some $\mathbf{D} \in \mathbb{C}^{n\times n}$. Then $\mathbf{w}^{(1)}$ and $\mathbf{y}$ are unchanged, and $\mathrm{range}(\tilde{\mathbf{A}}) = \mathrm{range}(\mathbf{A})$, which means that the update in (1) is unchanged, and the estimators between these two settings have the connection of $\tilde{\mathbf{x}}^{(k)} = \mathbf{D}^{-1}\mathbf{x}^{(k)}$. As a result, $\|e^{i\psi}\tilde{\mathbf{x}}^{(k)} - \tilde{\mathbf{z}}\| \to 0$ if and only if $\|e^{i\psi}\mathbf{x}^{(k)} - \mathbf{z}\| \to 0$. For the rest of the proof, we will analyze the equivalent problem where $\mathbf{D} = (\mathbf{A}^*\mathbf{A})^{-1/2}$ and $\mathbf{A}$ is replaced with $\tilde{\mathbf{A}} = \mathbf{A}\mathbf{D} = \mathbf{A}(\mathbf{A}^*\mathbf{A})^{-1/2}$, an orthogonal matrix with columns being an orthonormal basis of L.

Second, WLOG we assume that $\|\mathbf{z}\| = 1$ (which implies that $\|\mathbf{y}\| = 1$ because $\mathbf{A}$ is an orthogonal matrix) and we normalize $\mathbf{w}$ in the update formula (1):

$$\mathbf{w}^{(k+1)} = \frac{P_L P_{\mathcal{A}}\mathbf{w}^{(k)}}{\|P_L P_{\mathcal{A}}\mathbf{w}^{(k)}\|}. \tag{5}$$

Compared with the original form (1), $\mathbf{w}^{(k)}$ is normalized to a unit vector in each iteration. Since the operator $P_{\mathcal{A}}$ is invariant to scaling ($P_{\mathcal{A}}(c\mathbf{x}) = P_{\mathcal{A}}(\mathbf{x})$), the alternating minimization algorithm with normalization (5) is equivalent to the standard version (1) with a "correct" scaling, and it is relatively straightforward to verify that Theorem 1 holds for (5) if and only if it holds for (1).

Since $\{\mathbf{a}_i\}_{i=1}^n$ are i.i.d. sampled from $CN(0, \mathbf{I}_{m\times m})$, L is a random n-dimensional subspace in $\mathbb{C}^m$. Combining the analysis above, to prove Theorem 1, we will address the following equivalent problem:

- Choose a unit vector $\mathbf{z} \in \mathbb{C}^n$ and a random n-dimensional subspace L in $\mathbb{C}^m$, and a random unit vector in L, denote it by $\mathbf{w}^{(1)}$. Let $\mathbf{y} = |\mathbf{\Pi}_L^*\mathbf{z}|$, where $\mathbf{\Pi}_L$ represents a matrix in $\mathbb{C}^{m\times n}$, whose columns form an orthonormal basis of L (there are many choices of $\mathbf{\Pi}_L$: for any unitary matrix $\mathbf{U} \in \mathbb{C}^{n\times n}$, $\mathbf{\Pi}_L\mathbf{U}$ also satisfies this property, and we randomly choose one).
- The iterative update formula is given by

$$\mathbf{w}^{(k+1)} = \frac{P_L[\mathbf{w}^{(k)} \odot \mathbf{y}]}{\|P_L[\mathbf{w}^{(k)} \odot \mathbf{y}]\|}, \tag{6}$$

  and $\mathbf{x}^{(k)} = \mathbf{\Pi}_L^*\mathbf{w}^{(k)}$.
- Goal: prove (2).

B. Main Proof

In the proof, we first define a set of orthogonal unit vectors in $\mathbb{C}^m$:

$$\mathbf{u}_0 = \mathbf{\Pi}_L\mathbf{z}, \text{ (note that } \|\mathbf{u}_0\| = 1 \text{ since } \|\mathbf{z}\| = 1)$$

$$\mathbf{u}_k = \frac{\mathbf{w}^{(k)} - \sum_{i=0}^{k-1}\mathbf{u}_i\mathbf{u}_i^*\mathbf{w}^{(k)}}{\|\mathbf{w}^{(k)} - \sum_{i=0}^{k-1}\mathbf{u}_i\mathbf{u}_i^*\mathbf{w}^{(k)}\|}, \text{ for all } 1 \le k \le d,$$

where $d = C_d \log n$ with constant $C_d = \frac{1}{2\log(\frac{C_f+3}{4})} + 1$, where C_f will be defined later in (30) and does not depend on n or m.

Since $d < m$, $\{\mathbf{u}_i\}_{i=0}^d$ is a set of $d+1$ orthogonal vectors in $\mathbb{C}^m$. By definition, $\mathbf{w}^{(k)} \in \mathrm{Sp}(\mathbf{u}_0, \mathbf{u}_1, \cdots, \mathbf{u}_k)$ for any $1 \le k \le d$ and $\mathbf{w}^{(k)}$ can be written as

$$\mathbf{w}^{(k)} = \sum_{i=0}^k c_i^{(k)}\mathbf{u}_i.$$

By writing $P_L[\mathbf{w}^{(k)} \odot \mathbf{y}]$ in the basis of $\mathbf{u}_0, \cdots, \mathbf{u}_{k+1}$ as $P_L[\mathbf{w}^{(k)} \odot \mathbf{y}] = \sum_{i=0}^{(k+1)}\tilde{c}_{k+1}^{(i)}\mathbf{u}_i$, the update formula (6) can then be rewritten as the update of $\{c_i^{(k)}\}_{i=0}^k$ as follows: first, $c_0^{(1)} = \mathbf{u}_0^*\mathbf{w}^{(1)}$ and $c_1^{(1)} = \sqrt{1 - |c_0^{(1)}|^2}$. Then, for $1 \le k \le d-1$, the update formula of $\mathbf{w}^{(k)}$ in (6) is equivalent to

$$\tilde{c}_i^{(k+1)} = \mathbf{u}_i^*\left[\left(\sum_{i=0}^k c_i^{(k)}\mathbf{u}_i\right) \odot \mathbf{u}_0\right], \quad 0 \le i \le k, \tag{7}$$

$$\tilde{c}_{k+1}^{(k+1)} = \left\|P_L\left[\left(\sum_{i=0}^k c_i^{(k)}\mathbf{u}_i\right) \odot \mathbf{u}_0 - \sum_{i=0}^k \tilde{c}_i^{(k+1)}\mathbf{u}_i\right]\right\|, \tag{8}$$

$$c_i^{(k+1)} = \frac{\tilde{c}_i^{(k+1)}}{\sqrt{\sum_{i=0}^{k+1}|\tilde{c}_i^{(k+1)}|^2}}, \quad 0 \le i \le k+1. \tag{9}$$

While (8) seems complicated, this explicit formula will not be used later in the proof. Instead, the estimations

$$0 \le \tilde{c}_{k+1}^{(k+1)} \le \sqrt{1-\sum_{i=0}^{k}|\tilde{c}_i^{(k+1)}|^2} \quad (10)$$

and (13) (will be presented later) are sufficient, where the second inequality of (10) follows from the fact that $\sum_{i=0}^{k+1}|\tilde{c}_i^{(k+1)}|^2 = \|P_L[\mathbf{w}^{(k)}\odot \mathbf{y}]\|^2 \le \|\mathbf{w}^{(k)}\odot \mathbf{y}\|^2 = \|\mathbf{y}\|^2 = 1$.

The outline of the proof is as follows: first, we show that $\mathbf{u}_i$ can be well approximated by random vectors $\mathbf{v}_i$ from $CN(0,\mathbf{I}/m)$ in Lemma 1. This step decouples the dependency between the sensing vectors and the estimations at the k-th iteration. Second, Lemma 3 and 4 investigate the approximate dynamics of $\{c_k^{(i)}\}_{i=0}^k$ defined in (7) - (9), by replacing $\mathbf{u}_i$ with $\mathbf{v}_i$. Third, we obtain the dynamics of $\{c_k^{(i)}\}_{i=0}^k$ from applying a perturbation result in Lemma 2 to the dynamics we obtained in the second step. The above steps describe the first stage in Theorem 1. Finally, we prove that at the d-th iteration, the estimation is already sufficiently good, and Lemma 5, which is a direct corollary of [28, Theorem 2], will be used to prove that the algorithm succeeds. This step describes the second stage in Theorem 1.

Lemma 1: There exists $\{\mathbf{v}_i\}_{i=0}^d$ such that $\mathbf{v}_i$ are i.i.d. sampled from $CN(0,\mathbf{I}/m)$, $\mathbf{u}_0 = \mathbf{v}_0/\|\mathbf{v}_0\|$, and

$$\Pr\left(\|\mathbf{u}_k-\mathbf{v}_k\| > \frac{\log m}{\sqrt{m}}\right) < C\exp(-C\log^2 m) \text{ for } k=0,1 \quad (11)$$

$$\Pr\left(\|\mathbf{u}_k-\mathbf{v}_k\| > 2\sqrt{\frac{n}{m}}\right) < C\exp(-Cn) \text{ for } 2\le k\le d \quad (12)$$

$$\Pr\left(|\tilde{c}_k^{(k)}| > 2\sqrt{\frac{n}{m}}\right) < C\exp(-Cn) \text{ for } 1\le k\le d. \quad (13)$$

In addition, the properties

$$\|\mathbf{v}_i\| \le 2, \|\mathbf{v}_i\|_\infty \le \frac{\log m}{\sqrt{m}}, \text{ for all } 0\le i\le d \quad (14)$$

hold with probability $1-2m(d+1)\exp(-\log^2 m) - 2m(d+1)\exp(-\log^2 m)$.

Lemma 2: For $\mathbf{x}\in\mathbb{C}^m$ sampled from $CN(0,\mathbf{I}_{m\times m}/m)$, with probability at least $1-m\exp(-n/6)$, the following statement holds for all $\mathbf{y}\in\mathbb{C}^m$:

$$\frac{1}{m}\|\mathrm{phase}(\mathbf{x}+\mathbf{y})-\mathrm{phase}(\mathbf{x})\|_1 \le C\log m \max\left(\|\mathbf{y}\|, \frac{n}{m}\right)$$

Lemma 3: Define $f,g:[-1,1]\to\mathbb{R}$ by

$$f(c) = \frac{1}{c}\mathbb{E}_{x_0,x_1\sim CN(0,1)}\frac{cx_0+\sqrt{1-c^2}x_1}{|cx_0+\sqrt{1-c^2}x_1|}|x_0|x_0^* \quad (15)$$

and

$$g(c) = \frac{1}{\sqrt{1-c^2}}\mathbb{E}_{x_0,x_1\sim CN(0,1)}\frac{cx_0+\sqrt{1-c^2}x_1}{|cx_0+\sqrt{1-c^2}x_1|}|x_0|x_1^*. \quad (16)$$

Then for $\{\mathbf{v}_i\}_{i=0}^d$ i.i.d. sampled from $CN(0,\mathbf{I}/m)$, any fixed $\{c_i\}_{i=0}^d$ such that $\sum_{i=0}^d c_i^2 = 1$, and $\mathbf{x} = \sum_{i=0}^d c_i\mathbf{v}_i$,

$$\Pr\left(|\mathbf{v}_0^*[\mathbf{x}\odot\mathbf{v}_0] - f(|c_0|)c_0| < \frac{\log^2 m}{\sqrt{m}}\right) > 1-\exp(-C\log^4 m), \quad (17)$$

and for any $1\le j\le d$,

$$\Pr\left(|\mathbf{v}_j^*[\mathbf{x}\odot\mathbf{v}_0] - g(|c_0|)c_j| < \sqrt{\frac{n}{m}}\right) > 1-\exp(-Cn). \quad (18)$$

Lemma 4: Given any $0<C_0<1$, we have

$$f(c)\ge 1, \text{ for all } 0<c<C_0. \quad (19)$$

In addition, there exists $0<C_{g,1}<C_{g,2}<1$ depending on C_0 such that

$$0<C_{g,1}<g(c)<C_{g,2}<1, \text{ for all } 0<c<C_0. \quad (20)$$

The following lemma follows from [28, Theorem 2]:

Lemma 5: There exists $0<C_0<1$, $C_1', C_2'>0$ such that if $|c_0^{(k_0)}| > C_0$ for some $k_0>0$, then the algorithm (6) converges to the solution with probability $1-\exp(-n/2) - C_1'\exp(-C_2'm)$, in the sense that there exists $\delta\in(0,1)$ such that

$$\Pr\Big(\min_{0\le\psi\le 2\pi}\|e^{i\psi}\mathbf{z}-\mathbf{x}^{(k+1)}\| \le \delta \min_{0\le\psi\le 2\pi}\|e^{i\psi}\mathbf{z}-\mathbf{x}^{(k)}\|,$$
$$\text{for all } k\ge k_0\Big) \to 1, \text{ as } n,m\to\infty.$$

Lemma 6: With probability at least $1-1/\log n - \exp(-Cn)$, $|c_0^{(1)}| \ge \frac{1}{2\sqrt{n\log n}}$.

For the rest of the proof, we first assume that for all $1\le k\le d$, $|c_0^{(k)}| < C_0$, since otherwise Lemma 5 already implies Theorem 1. Then we will show that $|c_0^{(d+1)}| > C_0$ and Lemma 5 implies Theorem 1.

Let $\mathbf{c} = \{c_i\}_{i=0}^d \in \mathbb{C}^{d+1}$, we choose a set of covering balls of radius n/m in the set $\mathcal{S} = \{\mathbf{c}\in\mathbb{C}^{d+1} : \|\mathbf{c}\| = 1, |c_0|\le C_0\}$. That is, we find a subset $\mathcal{S}_0\subset\mathcal{S}$ such that for any $\mathbf{c}\in\mathcal{S}$, there exists an element $\bar{\mathbf{c}} = \{\bar{c}_i\}_{i=0}^d \in \mathcal{S}_0$ such that $\|\mathbf{c}-\bar{\mathbf{c}}\| \le n/m$. Following [27, Lemma 5.2], $\mathcal{S}_0$ can be chosen such that $|\mathcal{S}_0| \le (1+\frac{2m}{n})^{2(d+1)}$. We assume that for all $\bar{\mathbf{c}}\in\mathcal{S}_0$, the property in Lemma 2 holds for $\mathbf{x} = \sum_{i=0}^d \bar{c}_i\mathbf{v}_i$, and the property in Lemma 3 also holds. Then for all $j=0,1,\cdots,d$, we have

$$\left|\mathbf{u}_j^*\left[\sum_{i=0}^d c_i\mathbf{u}_i\odot\mathbf{u}_0\right] - \mathbf{v}_j^*\left[\sum_{i=0}^d \bar{c}_i\mathbf{v}_i\odot\mathbf{v}_0\right]\right| \quad (21)$$
$$= \left|\frac{1}{\|\mathbf{v}_0\|}\mathbf{u}_j^*\left[\sum_{i=0}^d c_i\mathbf{u}_i\odot\mathbf{v}_0\right] - \mathbf{v}_j^*\left[\sum_{i=0}^d \bar{c}_i\mathbf{v}_i\odot\mathbf{v}_0\right]\right|$$
$$= \Bigg|\left(\frac{1}{\|\mathbf{v}_0\|}\mathbf{u}_j^* - \mathbf{v}_j^*\right)\left[\sum_{i=0}^d c_i\mathbf{u}_i\odot\mathbf{v}_0\right]$$
$$- \mathbf{v}_j^*\left[\left[\sum_{i=0}^d c_i\mathbf{u}_i\odot\mathbf{v}_0\right] - \left[\sum_{i=0}^d \bar{c}_i\mathbf{v}_i\odot\mathbf{v}_0\right]\right]\Bigg|$$

$$\leq \left\| \frac{1}{\|\mathbf{v}_0\|}\mathbf{u}_j^* - \mathbf{v}_j^* \right\| \|\mathbf{v}_0\| + \|\mathbf{v}_j\|_\infty \left\| \text{phase}(\sum_{i=0}^{d} c_i \mathbf{u}_i) - \text{phase}(\sum_{i=0}^{d} \bar{c}_i \mathbf{v}_i) \right\|_1 \|\mathbf{v}_0\|_\infty$$
$$\leq \|\mathbf{u}_j - \mathbf{v}_j\| + |\|\mathbf{v}_0\| - 1| \, \|\mathbf{v}_j\| + \|\mathbf{v}_j\|_\infty \|\mathbf{v}_0\|_\infty m \cdot \max\left(\sum_{i=0}^{d} |c_i| \|\mathbf{u}_i - \mathbf{v}_i\| + |c_i - \bar{c}_i| \|\mathbf{v}_i\|, \frac{n}{m}\right)$$
$$\leq 2\frac{\log m}{\sqrt{m}} + \log^2 m \left(2\frac{\log m}{\sqrt{m}} + 2d \max_{2\leq i\leq d} |c_i| \sqrt{\frac{n}{m}} + \frac{3n}{m}\right) + \|\mathbf{u}_j - \mathbf{v}_j\|.$$

In the last line, $|\|\mathbf{v}_0\| - 1| \, \|\mathbf{v}_j\|$ is bounded by $2\frac{\log m}{\sqrt{m}}$ (combining Lemma 16 and (14)), $\|\mathbf{v}_j\|_\infty \|\mathbf{v}_0\|_\infty m$ is bounded by $\log^2 m$ (applying (14)), and $\max(\sum_{i=0}^{d} |c_i| \|\mathbf{u}_i - \mathbf{v}_i\| + |c_i - \bar{c}_i| \|\mathbf{v}_i\|, \frac{n}{m})$ is bounded by applying (11) and (12):

$$\max\left(\sum_{i=0}^{d} |c_i| \|\mathbf{u}_i - \mathbf{v}_i\| + |c_i - \bar{c}_i| \|\mathbf{v}_i\|, \frac{n}{m}\right)$$
$$\leq \sum_{i=0}^{d} (|c_i| \|\mathbf{u}_i - \mathbf{v}_i\| + |c_i - \bar{c}_i| \|\mathbf{v}_i\|) + \frac{n}{m}$$
$$\leq \sum_{i=0}^{d} |c_i| \|\mathbf{u}_i - \mathbf{v}_i\| + \|\mathbf{c} - \bar{\mathbf{c}}\| \max_i \|\mathbf{v}_i\| + \frac{n}{m}$$
$$\leq \sum_{i=0}^{1} \|\mathbf{u}_i - \mathbf{v}_i\| + \sum_{i=2}^{d} |c_i| \|\mathbf{u}_i - \mathbf{v}_i\| + \|\mathbf{c} - \bar{\mathbf{c}}\| \max_{0\leq i\leq d} \|\mathbf{v}_i\| + \frac{n}{m}$$
$$\leq 2\frac{\log m}{\sqrt{m}} + d \max_{2\leq i\leq d} 2|c_i| \sqrt{\frac{n}{m}} + \frac{3n}{m}. \quad (22)$$

By the definition of $\mathcal{S}_0$, we have that for each $1 \leq k \leq d$, there exists $\bar{\mathbf{c}}^{(k)} = [\bar{c}_0^{(k)}, \cdots, \bar{c}_d^{(k)}] \in \mathcal{S}_0$ such that

$$\sum_{i=0}^{k} |\bar{c}_i^{(k)} - c_i^{(k)}|^2 \leq \left(\frac{n}{m}\right)^2, \quad |\bar{c}_0^{(k)}| < C_0.$$

Combining the analysis in (21) (with $\mathbf{c}, \bar{\mathbf{c}}$ replaced by $\mathbf{c}^{(k)}, \bar{\mathbf{c}}^{(k)}$), and applying (7) and Lemma 3, we have that for $1 \leq j \leq k$,

$$\left|\tilde{c}_j^{(k+1)} - g(|\bar{c}_0^{(k)}|)\bar{c}_j^{(k)}\right| \leq 4\sqrt{\frac{n}{m}} + \log^2 m \left(2d \max_{2\leq i\leq k} |c_i^{(k)}| \sqrt{\frac{n}{m}} + \frac{3n}{m} + 2\frac{\log m}{\sqrt{m}}\right) \quad (23)$$

and for $j = 0$,

$$\left|\tilde{c}_0^{(k+1)} - f(|\bar{c}_0^{(k)}|)\bar{c}_0^{(k)}\right| \leq 3\frac{\log^2 m}{\sqrt{m}} + \log^2 m \left(2d \max_{2\leq i\leq k} |c_i^{(k)}| \sqrt{\frac{n}{m}} + \frac{3n}{m} + 2\frac{\log m}{\sqrt{m}}\right). \quad (24)$$

Combining (23), (24), (20), and (10),

$$\sqrt{\sum_{i=0}^{k+1} |\tilde{c}_i^{(k+1)}|^2} \geq \sqrt{\sum_{i=0}^{k} |\tilde{c}_i^{(k+1)}|^2} \quad (25)$$
$$\geq \sqrt{f^2(|\bar{c}_0^{(k)}|)|\bar{c}_0^{(k)}|^2 + g^2(|\bar{c}_0^{(k)}|)(1 - |\bar{c}_0^{(k)}|^2)} - 4k\sqrt{\frac{n}{m}} - 3\frac{\log^2 m}{\sqrt{m}} - 2d(k+1)\log^2 m \left(\max_{2\leq i\leq k} |c_i^{(k)}| \sqrt{\frac{n}{m}} + \frac{n}{m} + \frac{\log m}{\sqrt{m}}\right)$$
$$\geq C_{g,1}|\bar{c}_0^{(k)}| - 4k\sqrt{\frac{n}{m}} - 3\frac{\log^2 m}{\sqrt{m}} - 2d(k+1)\log^2 m \left(\max_{2\leq i\leq k} |c_i^{(k)}| \sqrt{\frac{n}{m}} + \frac{n}{m} + \frac{\log m}{\sqrt{m}}\right),$$

Combining (23) and (25) with the update formula (9), using induction we can verify that for sufficiently large n, m, we have

$$\max_{2\leq j\leq k+1} |c_j^{(k+1)}| < \frac{4}{C_{g,1}} \frac{\max(\sqrt{n}, \log^3 m)}{\sqrt{m}} \quad (26)$$

for all $0 \leq k \leq d-1$. By the assumption $m \geq M n^{3/2} \log^{7/2} n$ and Lemma 6, we have that as $n, m \to \infty$,

$$|c_0^{(1)}| \geq \frac{1}{2\sqrt{n \log n}} > 3\frac{\log^2 m}{\sqrt{m}} + \log^2 m \left(2C_d \log n \frac{4}{C_{g,1}} \frac{\max(\sqrt{n}, \log^3 m)}{\sqrt{m}} \sqrt{\frac{n}{m}} + \frac{3n}{m} + 2\frac{\log m}{\sqrt{m}}\right). \quad (27)$$

Similarly to (25), and applying the estimation (13), we have

$$\sqrt{\sum_{i=0}^{k+1} |\tilde{c}_i^{(k+1)}|^2} \leq \sqrt{\sum_{i=0}^{k} |\tilde{c}_i^{(k+1)}|^2 + |\tilde{c}_{k+1}^{(k+1)}|} \quad (28)$$
$$\leq \sqrt{C_0^2 + C_{g,2}^2(1 - C_0^2)} + (4k+6)\sqrt{\frac{n}{m}} + 2d(k+1)\log^2 m \left(\max_{2\leq i\leq k} |c_i^{(k)}| \sqrt{\frac{n}{m}} + \frac{n}{m} + \frac{\log m}{\sqrt{m}}\right).$$

Combining (27), (24), (26), and (10), it can be verified by induction that when M is sufficiently large, for all $1 \leq k \leq d$ we have

$$|\tilde{c}_0^{(k+1)}| \geq |\bar{c}_0^{(k)}| - 3\frac{\log^2 m}{\sqrt{m}} - \log^2 m \left(2d \max_{2\leq i\leq k} |c_i^{(k)}| \sqrt{\frac{n}{m}} + \frac{3n}{m} + 2\frac{\log m}{\sqrt{m}}\right),$$

$$\sqrt{\sum_{i=0}^{k+1} |\tilde{c}_i^{(k+1)}|^2} \leq \sqrt{C_0^2 + C_{g,2}^2(1 - C_0^2)} + (4k+6)\sqrt{\frac{n}{m}} + 2d(k+1)\log^2 m \left(\max_{2\leq i\leq k} |c_i^{(k)}| \sqrt{\frac{n}{m}} + \frac{n}{m} + \frac{\log m}{\sqrt{m}}\right),$$

and

$$\begin{aligned}|c_0^{(k+1)}| &= \frac{|\tilde{c}_0^{(k+1)}|}{\sqrt{\sum_{i=0}^{k+1}|\tilde{c}_i^{(k+1)}|^2}}\\ &\geq \frac{C_f+1}{2}|\tilde{c}_0^{(k)}| \geq \frac{C_f+1}{2}(|c_0^{(k)}| - \frac{n}{m}) \geq \frac{C_f+3}{4}|c_0^{(k)}|,\end{aligned} \tag{29}$$

where

$$C_f = \frac{1}{\sqrt{C_0^2 + C_{g,2}^2(1-C_0^2)}} \tag{30}$$

and it can be verified from $C_{g,2} < 1$ that $C_f > 1$. Combining (29) with the definition of $c_k^{(0)}$, we obtain (3), the convergence in the first stage.

Combining (29) with Lemma 6 we have

$$|c_0^{(d+1)}| \geq \left(\frac{C_f+3}{4}\right)^{C_d \log n} \frac{1}{2\sqrt{n\log n}} > C_0 \tag{31}$$

In fact, the second inequality requires

$$C_d > \frac{\log(2C_0) + \frac{1}{2}\log n + \frac{1}{2}\log\log n}{\log(\frac{C_f+3}{4})\log n},$$

and for large n, our choice of $C_d = \frac{1}{2\log(\frac{C_f+3}{4})} + 1$ suffices. With (31), Lemma 5 implied (4), the convergence of the second stage. Then (2) is proved.

In the end, we summarize the probability that the above analysis holds: the proof requires the events in all lemmas, and in addition, the events in Lemma 2 and Lemma 3 should hold for all $\mathbf{x} = \sum_{i=0}^{d}\bar{c}_i\mathbf{v}_i$, where $\bar{\mathbf{c}} = [\bar{c}_0, \bar{c}_1, \cdots, \bar{c}_d]$ is an element in $\mathcal{S}_0$. As a result, the probability is at least

$$\begin{aligned}&1 - 2Cmd\exp(-\log^2 m) - \exp(-n/2) - C_1'\exp(-C_2'm)\\ &-\frac{1}{\log n} - \exp(Cn) - \left(1+\frac{2m}{n}\right)^{2d+2}\\ &\cdot\Big(m\exp(-n/6) - \exp(-C\log^4 m) - d\exp(-Cn)\Big),\end{aligned}$$

which can be verified to converge to 1 as $n, m \to \infty$.

C. Proof of the Main Lemmas

Proof of Lemma 1: Since L is a random n-dimensional subspace in $\mathbb{C}^m$, and $\mathbf{\Pi}_L$ is a random projection matrix to L, $\mathbf{u}_0$ is a random unit vector in $\mathbb{C}^m$ that is uniformly sampled from the sphere in $\mathbb{C}^m$. Therefore, it can be obtained through $\mathbf{v}_0 \sim CN(0, \mathbf{I}/m)$ by

$$\mathbf{u}_0 = \frac{\mathbf{v}_0}{\|\mathbf{v}_0\|}.$$

Applying Lemma 16 (with a scaling of $\sqrt{m}$) and $\|\mathbf{u}_0 - \mathbf{v}_0\| = |\|\mathbf{v}_0\| - 1|$, we proved (11) for $k = 0$.

Under the σ-algebra generated by $\mathbf{u}_0$, the conditional distribution of L is a random subspace generated by

$$\mathrm{Sp}(\mathbf{u}_0) \oplus L_0,$$

where L_0 is a random $n-1$-dimensional subspace in the $m-1$-dimensional hyperplane $\mathrm{Sp}(\mathbf{u}_0)^\perp$ (here $\oplus$ represents the direct sum of two subspaces). Since $\mathbf{w}^{(1)}$ is a random initialization on L and $\mathbf{u}^{(1)}$ is the projection of $\mathbf{w}^{(1)}$ onto $\mathrm{Sp}(\mathbf{u}_0)^\perp$, $\mathbf{u}_1$ is a random unit vector on L_0. Combining it with the conditional distribution of L_0, $\mathbf{u}_1$ is a random unit vector that is orthogonal to $\mathbf{u}_0$. As a result, it can be generated from $\mathbf{v}_1 \sim CN(0, \mathbf{I}/m)$ as follows:

$$\mathbf{u}_1 = \frac{\mathbf{v}_1 - \mathbf{u}_0\mathbf{u}_0^*\mathbf{v}_1}{\|\mathbf{v}_1 - \mathbf{u}_0\mathbf{u}_0^*\mathbf{v}_1\|}.$$

Applying Lemma 16, we have

$$Pr(|1 - \|\mathbf{v}_1\|| > \frac{\log m}{3\sqrt{m}}) < \exp(-C\log^2 m). \tag{32}$$

In addition, since $\mathbf{u}_0^*\mathbf{v}_1 \sim CN(0, 1/m)$, Lemma 17 with $m = 1$ implies that

$$\Pr\left(|\mathbf{u}_0^*\mathbf{v}_1| > \frac{\log m}{3\sqrt{m}}\right) < \exp(-C\log^2 m). \tag{33}$$

Applying Lemma 12, under the events of (32) and (33),

$$\begin{aligned}&\|\mathbf{u}_1 - \mathbf{v}_1\| \leq \|\mathbf{u}_1\|\mathbf{v}_1\| - \mathbf{u}_1\| + \|\mathbf{v}_1\|\left\|\mathbf{u}_1 - \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}\right\|\\ =&|\|\mathbf{v}_1\| - 1| + \|\mathbf{v}_1\|\left\|\frac{\mathbf{v}_1 - \mathbf{u}_0\mathbf{u}_0^*\mathbf{v}_1}{\|\mathbf{v}_1 - \mathbf{u}_0\mathbf{u}_0^*\mathbf{v}_1\|} - \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}\right\|\\ \leq&|\|\mathbf{v}_1\| - 1| + \|\mathbf{v}_1\|\cdot 2\frac{\|\mathbf{u}_0\mathbf{u}_0^*\mathbf{v}_1\|}{\|\mathbf{v}_1\|} = |\|\mathbf{v}_1\| - 1| + 2|\mathbf{u}_0^*\mathbf{v}_1|\\ \leq&\frac{\log m}{\sqrt{m}},\end{aligned}$$

which implies (11) with $k = 1$.

To prove (12), we first investigate the conditional distribution of L under the σ-algebra generated by the algorithm so far, that is, generated by $\{\mathbf{u}_i\}_{i=0}^{k-1}$ and $\{\mathbf{w}_i\}_{i=0}^{k-1}$. That is, what is the conditional distribution of L when $\{\mathbf{u}_i\}_{i=0}^{k-1}$ and $\{\mathbf{w}_i\}_{i=0}^{k-1}$ are fixed? Under this σ-algebra, L satisfies the following properties:

$$\begin{aligned}\mathbf{u}_i \in L, \quad 0 \leq i \leq k-1\\ [\mathbf{w}^{(i)} \odot \mathbf{y}] - \mathbf{w}^{(i+1)}\mathbf{w}^{(i+1)*}[\mathbf{w}^{(i)} \odot \mathbf{y}] \perp L,\ 1 \leq i \leq k-2.\end{aligned}$$

The second property above holds since $\mathbf{w}^{(i+1)}$ is the normalization projection of $\mathbf{w}^{(i)} \odot \mathbf{y}$ onto L, and as a result, $\mathbf{w}^{(i+1)}\mathbf{w}^{(i+1)*}[\mathbf{w}^{(i)} \odot \mathbf{y}] = P_L[\mathbf{w}^{(i)} \odot \mathbf{y}]$. Recall that L is a random n-dimensional subspace in $\mathbb{C}^m$, with this σ-algebra, its conditional distribution then can be written as

$$L = \mathrm{Sp}\{\mathbf{u}_i\}_{i=0}^{k-1} \oplus L_k,$$

where L_k is a random $n-k$-dimensional subspace in the $m-2k+2$-space R_k that is orthogonal to $\mathbf{u}_i$, $0 \leq i \leq k-1$ and $[\mathbf{w}^{(i)} \odot \mathbf{y}] - \mathbf{w}^{(i+1)}\mathbf{w}^{(i+1)*}[\mathbf{w}^{(i)} \odot \mathbf{y}]$, $1 \leq i \leq k-2$.

Since $\mathbf{w}^{(k)}$ is the projection of $\mathbf{w}^{(k-1)} \odot \mathbf{y}$ onto the subspace L and $\mathbf{u}_k$ is the unit vector of the projection of $\mathbf{w}^{(k)}$ to the subspace orthogonal to $\mathrm{Sp}\{\mathbf{u}_i\}_{i=0}^{k-1}$, in conclusion, $\mathbf{u}^{(k)}$ is the unit vector that corresponds to the projection of $P_{R_k}[\mathbf{w}^{(k-1)} \odot \mathbf{y}]$ onto L_k, a random $n-k$-dimensional subspace in R_k. Applying Lemma 7 (with $m, n, \mathbb{C}^m$ replaced by $m-2k+2$, $n-k$, R_k), $\mathbf{u}_k$ can be written as

$$\mathbf{u}_k = \sqrt{1-a^2}\mathbf{v}_k' + a\frac{P_{R_k}[\mathbf{w}^{(k-1)} \odot \mathbf{y}]}{\|P_{R_k}[\mathbf{w}^{(k-1)} \odot \mathbf{y}]\|}, \tag{34}$$

where $\mathbf{v}_k'$ is a unit vector on R_k', the $m-2k+1$-dimensional subspace inside R_k and orthogonal to $P_{R_k}[\mathbf{w}^{(k-1)} \odot \mathbf{y}]$, and a is the length of the projection of $P_{R_k}[\mathbf{w}^{(k-1)} \odot \mathbf{y}]/\|P_{R_k}[\mathbf{w}^{(k-1)} \odot \mathbf{y}]\|$ onto L_k.

Since L_k is a random subspace in R_k, $\mathbf{v}_k'$ is a random unit vector on R_k' and can be derived through $\mathbf{v}_k \sim CN(0, \mathbf{I}/m)$ by

$$\mathbf{v}_k' = \frac{P_{R_k'}\mathbf{v}_k}{\|P_{R_k'}\mathbf{v}_k\|}.$$

Again use the fact that

$$\Pr\left(|1-\|\mathbf{v}_k\|| > \frac{1}{6}\sqrt{\frac{n}{m}}\right) < \exp(-Cn)$$

and Lemma 17 implies

$$\Pr\left(\|\mathbf{v}_k - P_{R_k'}\mathbf{v}_k\| > \frac{1}{6}\sqrt{\frac{n}{m}}\right) < 2(2k-1)\exp(-Cn/(2k-1)),$$

and Lemma 8 implies

$$\Pr\left(a^2 > 2\cdot\frac{n-k}{m-2k+2}\right) < 4\exp\left(-C(n-k)\right). \quad (35)$$

Combining all estimations above with (34) and Lemma 12, (12) is proved as follows:

$$\begin{aligned}&\|\mathbf{u}_k - \mathbf{v}_k\| \le a + \sqrt{1-a^2}\|\mathbf{v}_k' - \mathbf{v}_k\|\\ \le& a + \sqrt{1-a^2}(|\|\mathbf{v}_k\|-1| + \|\mathbf{v}_k' - \frac{\mathbf{v}_k}{\|\mathbf{v}_k\|}\|)\\ \le& a + (|\|\mathbf{v}_k\|-1| + 2\|P_{R_k'}\mathbf{v}_k - \mathbf{v}_k\|/\|\mathbf{v}_k\|)\\ \le& \sqrt{2\cdot\frac{n-k}{m-2k+2}} + \left(\frac{1}{6}\sqrt{\frac{n}{m}} + 2\frac{\frac{1}{6}\sqrt{\frac{n}{m}}}{1-\frac{1}{6}\sqrt{\frac{n}{m}}}\right) \le 2\sqrt{\frac{n}{m}}.\end{aligned}$$

Note $\tilde{c}_k^{(k)}$ is the length of projection of $P_{R_k}[\mathbf{w}^{(k-1)} \odot \mathbf{y}]$ onto L_k, and $\|P_{R_k}[\mathbf{w}^{(k-1)} \odot \mathbf{y}]\| \le \|[\mathbf{w}^{(k-1)} \odot \mathbf{y}]\| = \|\mathbf{y}\| = 1$, by the definition of a we have $\tilde{c}_k^{(k)} \le a$. Then (35) implies (13).

At last, (14) is obtained by applying Lemma 17 (with union bound and $m=1$ for the $\|\cdot\|_\infty$ norm). □

Proof of Lemma 2: It is based on a combination of Lemma 10, 11, and 12. In particular, $t = \max(\|\mathbf{y}\|, n/m)$ in Lemma 10 (remark: there is a scaling factor of $\sqrt{m}$ between $\mathbf{x}$ in Lemma 2 and Lemma 10). □

Proof of Lemma 3: The proof is based on two components: first, we have

$$\mathbb{E}\,\mathbf{v}_0^*[\mathbf{x}\odot\mathbf{v}_0] = f(|c_0|)c_0, \quad (36)$$

and for any $1 \le j \le d$,

$$\mathbb{E}\,\mathbf{v}_j^*[\mathbf{x}\odot\mathbf{v}_0] = g(|c_0|)c_j. \quad (37)$$

To prove (37), we write $\mathbf{x} = c_0\mathbf{v}_0 + \sqrt{1-|c_0|^2}\mathbf{v}'$ with $\mathbf{v}' = \frac{1}{\sqrt{\sum_{i=1}^d |c_i|^2}}\sum_{i=1}^d c_i\mathbf{v}_i$, then $\mathbf{v}'$ and $\mathbf{v}_0$ are independently sampled from $CN(0,\mathbf{I}/m)$. The definition of g then implies

$$\mathbb{E}\,\mathbf{v}'^*[\mathbf{x}\odot\mathbf{v}_0] = \sqrt{1-|c_0|^2}g(|c_0|)$$

Noting that the correlation between $\mathbf{v}_j$ and $\mathbf{v}'$ is $\frac{c_j}{\sqrt{1-|c_0|^2}}$ and both are sampled from $\sim CN(0,\mathbf{I}/m)$, using the properties of Gaussian distributions, $\tilde{\mathbf{v}}' = \mathbf{v}_j - \frac{c_j}{\sqrt{1-|c_0|^2}}\mathbf{v}'$ follows a Gaussian distribution and is independent of $\mathbf{v}'$. Since $\mathbf{v}_0$ is independent of $\mathbf{v}'$, $\tilde{\mathbf{v}}'$ is also independent of $\mathbf{x} = c_0\mathbf{v}_0 + \sqrt{1-|c_0|^2}\mathbf{v}'$. As a result, $\mathbb{E}\,\tilde{\mathbf{v}}'^*[\mathbf{x}\odot\mathbf{v}_0] = 0$ and (37) is proved as follows:

$$\begin{aligned}&\mathbb{E}\,\mathbf{v}_j^*[\mathbf{x}\odot\mathbf{v}_0] = \frac{c_j}{\sqrt{1-|c_0|^2}}\mathbb{E}\,\mathbf{v}'^*[\mathbf{x}\odot\mathbf{v}_0] + \mathbb{E}\,\tilde{\mathbf{v}}'^*[\mathbf{x}\odot\mathbf{v}_0]\\ =&\frac{c_j}{\sqrt{1-|c_0|^2}}\mathbb{E}\,\mathbf{v}'^*[\mathbf{x}\odot\mathbf{v}_0] = g(|c_0|)c_j.\end{aligned}$$

Since each element of $\mathbf{v}_0$ is sampled from $CN(0,1/m)$, it can be verified that the real component and the imaginary component of each entry of $\mathbf{x}\odot\mathbf{v}_0$ is sub-Gaussian, with sub-Gaussian parameter bounded above by $C/\sqrt{m}$. Applying Lemma 13, (18) is proved.

The proof of (17) is based on the proof of (36), which is similar to the proof of (37). □

Proof of Lemma 4: The proof addresses (19), the upper bound in (20), and the lower bound in (20) separately.

We start with the proof of (19). Applying the property of Gaussian distribution, x_0 and x_1 are independently sampled from $CN(0,1)$ if and only if $z_0 = cx_0 + \sqrt{1-c^2}x_1$ and $z_1 = \sqrt{1-c^2}x_0 - cx_1$ are independently sampled from $CN(0,1)$. As a result,

$$\begin{aligned}&f(c) = \frac{1}{c}\mathbb{E}_{z_0,z_1\sim CN(0,1)}\frac{cz_0+\sqrt{1-c^2}z_1}{|cz_0+\sqrt{1-c^2}z_1|}|z_0|z_0^*\\ =&\frac{1}{c}\mathbb{E}_{x_0,x_1\sim CN(0,1)}\frac{|cx_0+\sqrt{1-c^2}x_1|}{|x_0|}x_0(cx_0+\sqrt{1-c^2}x_1)^*\\ =&\frac{1}{c}\mathbb{E}_{x_0,x_1}\frac{|cx_0+\sqrt{1-c^2}x_1|}{|x_0|}\mathrm{Re}(x_0(cx_0+\sqrt{1-c^2}x_1)^*),\end{aligned}$$

where the last step follows from the fact that $f(c)$ is real-valued (because of symmetry). Combining it with the original definition of $f(c)$ in (15), we have

$$\begin{aligned}&2f(c) \quad (38)\\ =&\frac{1}{c}\mathbb{E}_{x_0,x_1}\frac{|cx_0+\sqrt{1-c^2}x_1|}{|x_0|}\mathrm{Re}\big(x_0(cx_0+\sqrt{1-c^2}x_1)^*\big)\\ &+\frac{1}{c}\mathbb{E}_{x_0,x_1}\frac{|x_0|}{|cx_0+\sqrt{1-c^2}x_1|}\mathrm{Re}\big((cx_0+\sqrt{1-c^2}x_1)x_0^*\big)\\ \ge&\frac{2}{c}\mathbb{E}_{x_0,x_1}\mathrm{Re}\big(cx_0+\sqrt{1-c^2}x_1)x_0^*\big)\\ =&\frac{2}{c}\Big(c\,\mathbb{E}_{x_0,x_1}|x_0|^2 + \sqrt{1-c^2}\mathrm{Re}(\mathbb{E}_{x_0,x_1}x_1x_0^*)\Big)\\ =&\frac{2}{c}\Big(c+0\Big) = 2,\end{aligned}$$

where the last equality applies $\mathbb{E}_{x_0,x_1}x_1x_0^* = \big(\mathbb{E}_{x_0,x_1}x_1\big)\big(\mathbb{E}_{x_0,x_1}x_0^*\big) = 0\cdot 0 = 0$, and (19) is proved.

To prove the upper bound in (20), we first prove that for $0 \le c < 1$

$$\mathbb{E}_{z_0,z_1}\frac{|cz_0+\sqrt{1-c^2}z_1|}{|z_0|}\mathrm{Re}(z_0z_1^*) > 0. \quad (39)$$

In fact, if we fix z_0 and WLOG assume that it is real-valued and positive, and let $\tilde{z}_1 = -\mathrm{Re}(z_1) + i\mathrm{Im}(z_1)$, then when $\mathrm{Re}(z_1) > 0$,

$$\mathrm{Re}(z_0\tilde{z}_1^*) < 0, \;\; |cz_0+\sqrt{1-c^2}\tilde{z}_1| - |cz_0+\sqrt{1-c^2}z_1| < 0,$$

and when $\mathrm{Re}(z_1) < 0$,

$$\mathrm{Re}(z_0\tilde{z}_1^*) > 0, \ \ |cz_0 + \sqrt{1-c^2}\tilde{z}_1| - |cz_0 + \sqrt{1-c^2}z_1| > 0.$$

Combining these two cases, we have

$$\frac{|cz_0 + \sqrt{1-c^2}\tilde{z}_1|}{|z_0|}\mathrm{Re}(z_0\tilde{z}_1^*) - \frac{|cz_0 + \sqrt{1-c^2}z_1|}{|z_0|}\mathrm{Re}(z_0\tilde{z}_1^*)$$
$$= \left(\frac{|cz_0 + \sqrt{1-c^2}\tilde{z}_1| - |cz_0 + \sqrt{1-c^2}z_1|}{|z_0|}\right)\mathrm{Re}(z_0\tilde{z}_1^*) \geq 0. \tag{40}$$

As a result, for any z_0, we have

$$2\,\mathbb{E}_{z_1}\frac{|cz_0 + \sqrt{1-c^2}z_1|}{|z_0|}\mathrm{Re}(z_0 z_1^*) \tag{41}$$
$$=\mathbb{E}\Big(\frac{|cz_0+\sqrt{1-c^2}z_1|}{|z_0|}\mathrm{Re}(z_0z_1^*)+\frac{|cz_0+\sqrt{1-c^2}\tilde{z}_1|}{|z_0|}\mathrm{Re}(z_0\tilde{z}_1^*)\Big)$$
$$=\mathbb{E}\Big(-\frac{|cz_0+\sqrt{1-c^2}z_1|}{|z_0|}\mathrm{Re}(z_0\tilde{z}_1^*)+\frac{|cz_0+\sqrt{1-c^2}\tilde{z}_1|}{|z_0|}\mathrm{Re}(z_0\tilde{z}_1^*)\Big)$$
$$\geq 0.$$

Here the first equality follows from the fact that $\tilde{z}_1$ and z_1 have the same distribution, the second equality follows from $\mathrm{Re}(z_0z_1^*) = -\mathrm{Re}(z_0\tilde{z}_1^*)$, and the first inequality follows from (40). Then, (39) is proved by taking the expectation of (41) to z_0.

Applying (39), the upper bound of $g(c)$ can be estimated as follows:

$$g(c) = \frac{1}{\sqrt{1-c^2}}\,\mathbb{E}_{x_0,x_1}\frac{cx_0+\sqrt{1-c^2}x_1}{|cx_0+\sqrt{1-c^2}x_1|}|x_0|x_1^*$$
$$=\frac{1}{\sqrt{1-c^2}}\,\mathbb{E}_{x_0,x_1}\frac{z_0}{|z_0|}|x_0|(\sqrt{1-c^2}z_0 - cz_1)^*$$
$$=\mathbb{E}_{x_0,x_1}|z_0||x_0| - \frac{c}{\sqrt{1-c^2}}\frac{|x_0|}{|z_0|}\mathrm{Re}(z_0z_1^*)$$
$$\leq \mathbb{E}\frac{|z_0|^2+|x_0|^2}{2} - \mathbb{E}\frac{c}{\sqrt{1-c^2}}\frac{|cz_0+\sqrt{1-c^2}z_1|}{|z_0|}\mathrm{Re}(z_0z_1^*) < 1,$$

where the last inequality follows from (39). With the continuity of $g(c)$, it means that there exists $C_{g,2} < 1$ such that $g(c) < C_{g,2}$ for all $0 \leq c \leq C_0$, and the upper bound in (20) is proved.

To prove the lower bound in (20), we define $h : [0,\infty) \to \mathbb{R}$ as follows:

$$h(c) = \mathbb{E}_{x\in\mu}\frac{1+cx}{|1+cx|},$$

where μ is the uniform distribution on the unit circle in $\mathbb{C}$. Since $\mathrm{Re}(\frac{1+cx}{|1+cx|}) > \mathrm{Re}(\frac{cx}{|cx|}) = \mathrm{Re}(\frac{x}{|x|})$ and $\mathbb{E}\,\mathrm{Re}(\frac{x}{|x|}) = 0$, we have $\mathbb{E}\,\frac{1+cx}{|1+cx|} = \mathbb{E}\,\mathrm{Re}(\frac{1+cx}{|1+cx|}) > 0$ (the equality applies the symmetry of the distribution μ), and

$$g(c) = \frac{1}{\sqrt{1-c^2}}\,\mathbb{E}_{x_0,x_1\sim CN(0,1)}\frac{cx_1+\sqrt{1-c^2}x_0}{|cx_1+\sqrt{1-c^2}x_0|}|x_1|x_0^*$$
$$=\frac{1}{\sqrt{1-c^2}}\,\mathbb{E}_{x_0,x_1}\,h\left(\frac{c|x_1|}{\sqrt{1-c^2}|x_0|}\right)|x_0||x_1| > 0.$$

Since $g(c)$ is strictly positive for any $0 \leq c \leq C_0$ and is continuous, there exists $C_{g,1}$ such that $\min_{0<c<C_0} g(c) > C_{g,1}$. This proves the lower bound in (20).

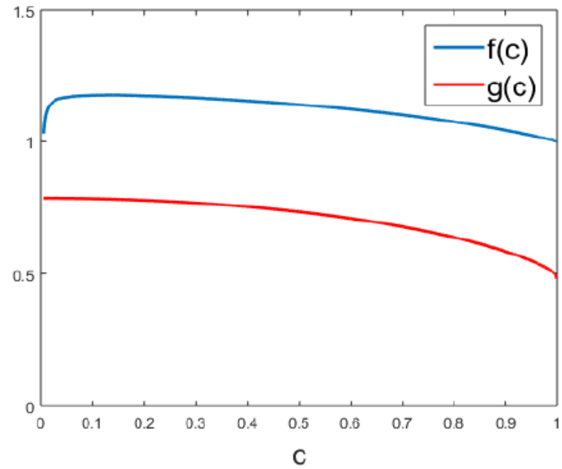

Fig. 1. The empirical values of f, g, and f/g.

We include the numerical values of $f(c)$ and $g(c)$ in Figure 1, to show that the inequalities (19) and (20) hold empirically. □

Proof of Lemma 5: For convenience, we first write down [28, Theorem 2] explicitly:

Theorem 2 ([28], Theorem 2): There exists $C_0', C_1', C_2', M > 0$ such that when $m > Mn$, then with probability at least $1 - C_1'\exp(-C_2'm)$, for any $\mathbf{x} \in \mathbb{C}^n$ such that

$$\inf_{\psi\in\mathbb{R}} \|e^{i\psi}\mathbf{z} - \mathbf{x}\| \leq C_0'\|\mathbf{z}\|,$$

then

$$\inf_{\psi\in\mathbb{R}} \|e^{i\psi}\mathbf{z} - \mathbf{x}^+\| \leq \delta \inf_{\psi\in\mathbb{R}} \|e^{i\psi}\mathbf{z} - \mathbf{x}\|,$$

where $\mathbf{x}^+$ is the vector obtained by applying one iteration of the standard alternating projection algorithm (without normalization) (1) to $\mathbf{x}$, and with $\{\mathbf{a}_i\}_{i=1}^m$ i.i.d. sampled from $CN(0, \mathbf{I})$.

By the analysis in Section II-A, (1) is equivalent to the algorithm that we are analyzing in (6) in terms of $\mathbf{w}^{(k)}$ as in $\mathbb{C}^m$. Therefore, [28, Theorem 2] implies that for $\mathbf{A}^+ = \mathbf{A}^*(\mathbf{A}^*\mathbf{A})^{-1}$, if

$$\inf_{\psi\in\mathbb{R}} \|\mathbf{A}^+(e^{i\psi}\mathbf{u}_0 - \mathbf{w}^{(k_0)})\| \leq C_0'\|\mathbf{A}^+\mathbf{u}_0\|,$$

then for any $k \geq k_0$,

$$\inf_{\psi\in\mathbb{R}} \|\mathbf{A}^+(e^{i\psi}\mathbf{u}_0 - \mathbf{w}^{(k+1)})\| \leq \delta \inf_{\psi\in\mathbb{R}} \|\mathbf{A}^+(e^{i\psi}\mathbf{u}_0 - \mathbf{w}^{(k)})\|.$$

Since $\mathbf{A}$ is a complex Gaussian, [11, Theorem 2.13] implies that the condition number of $\mathbf{A}$ is bounded with high probability:

$$\Pr\left(\frac{\sigma_{\max}(\mathbf{A})}{\sigma_{\min}(\mathbf{A})} \leq \frac{\sqrt{m}+\sqrt{n}+t}{\sqrt{m}-\sqrt{n}-t}\right) \leq 1 - 2\exp(-t^2/2).$$

Combining it (use $t=\sqrt{n}$) with

$$\inf_{\psi\in\mathbb{R}}\|e^{i\psi}\mathbf{u}_0-\mathbf{w}^{(k)}\|=\sqrt{(1-|c_0^{(k)}|)^2+\sum_{i=1}^{k}|c_i^{(k)}|^2}$$
$$=\sqrt{(1-|c_0^{(k)}|)^2+1-|c_0^{(k)}|^2},$$

there exists $0<C_0<1$ such that when $|c_0^{(k)}|>C_0$, $\inf_{\psi\in\mathbb{R}}\|\mathbf{A}^+(e^{i\psi}\mathbf{u}_0-\mathbf{w}^{(k)})\|<C_0'\|\mathbf{A}^+\mathbf{u}_0\|$ (note that $\mathbf{A}$ and $\mathbf{A}^+$ have the same condition numbers), and then [28, Theorem 2] implies $\lim_{k\to\infty}\inf_{\psi\in\mathbb{R}}\|\mathbf{A}^+(e^{i\psi}\mathbf{u}_0-\mathbf{w}^{(k)})\|=0$. Applying the fact that the condition number of $\mathbf{A}^+$ is bounded again, we have $\lim_{k\to\infty}\inf_{\psi\in\mathbb{R}}\|e^{i\psi}\mathbf{u}_0-\mathbf{w}^{(k)}\|=0$. □

Proof of Lemma 6: WLOG we may assume that $L=\mathrm{Sp}(\mathbf{e}_1,\cdots,\mathbf{e}_n)$ and $\mathbf{u}_0=\mathbf{e}_1$, and $\mathbf{w}^{(1)}\sim CN(0,P_L)$ (since $\mathbf{w}^{(1)}$ is a random vector on L). Then

$$|c_0^{(1)}|=\frac{|w_1^{(1)}|}{\|\mathbf{w}^{(1)}\|}.$$

Applying Lemma 16 and note that $\|\mathbf{w}^{(1)}\|^2$ is the sum of n unit complex gaussian squared, $\Pr(\|\mathbf{w}^{(1)}\|>2\sqrt{n})<2\exp(-Cn)$; and Lemma 9 implies that $\Pr(\|\mathbf{w}^{(1)}\|<1/\sqrt{\log n})<\log n$. □

D. Auxillary Lemmas

Lemma 7: Assuming that the projection of a unit vector $\mathbf{x}\in\mathbb{C}^m$ to a random n-dimensional subspace L has length a, i.e., $\|P_L\mathbf{x}\|=a$, then

$$\frac{P_L\mathbf{x}}{\|P_L\mathbf{x}\|}=a\mathbf{x}+\sqrt{1-a^2}\mathbf{v},$$

where $\mathbf{v}$ is a unit vector perpendicular to $\mathbf{x}$, that is, $\mathbf{v}^*\mathbf{x}=1$.

Proof: Since $\frac{P_L\mathbf{x}}{\|P_L\mathbf{x}\|}$ is a unit vector, we may assume that

$$\frac{P_L\mathbf{x}}{\|P_L\mathbf{x}\|}=b\mathbf{x}+\sqrt{1-b^2}\mathbf{v},\tag{42}$$

where $\|\mathbf{v}\|=1$ and $\mathbf{v}^*\mathbf{x}=0$. It remains to prove $a=b$.

By the definition of projection, we have $(\mathbf{x}-P_L\mathbf{x})\perp P_L\mathbf{x}$, i.e., $P_L\mathbf{x}^*(\mathbf{x}-P_L\mathbf{x})=0$. Applying the assumption (42) and $\|P_L\mathbf{x}\|=a$ we have

$$\left(b\mathbf{x}+\sqrt{1-b^2}\mathbf{v}\right)^*\left((1-ab)\mathbf{x}-a\sqrt{1-b^2}\mathbf{v}\right)=0.$$

With $\mathbf{v}^*\mathbf{x}=0$ and $\|\mathbf{x}\|=\|\mathbf{v}\|=1$, it implies $b(1-ab)=a(1-b^2)$ and $a=b$. □

Lemma 8: Given a vector $\mathbf{x}\in\mathbb{R}^m$ and a random n-dimensional subspace L, then

$$\Pr\left(\frac{1-\epsilon}{1+\epsilon}\le\frac{m\|P_L\mathbf{x}\|^2}{n\|\mathbf{x}\|^2}\le\frac{1+\epsilon}{1-\epsilon}\right)$$
$$\ge 1-4\exp\left(-cn\min\left(\frac{\epsilon^2}{C^2},\frac{\epsilon}{C}\right)\right),$$

Proof: WLOG we may assume that $\mathbf{x}\sim CN(0,\mathbf{I})$ and L is the subspace spanned by the first n standard basis $\mathbf{e}_1,\cdots,\mathbf{e}_n$. Then $\|\mathbf{x}\|^2=\sum_{i=1}^m|x_i|^2$ and $\|P_L\mathbf{x}\|^2=\sum_{i=1}^n|x_i|^2$. Applying Lemma 16, we have

$$\Pr\left((1-\epsilon)m\le\sum_{i=1}^m|x_i|^2\le(1+\epsilon)m\right)$$
$$\ge 1-2\exp\left(-cm\min\left(\frac{\epsilon^2}{C^2},\frac{\epsilon}{C}\right)\right)$$
$$\Pr\left((1-\epsilon)n\le\sum_{i=1}^n|x_i|^2\le(1+\epsilon)n\right)$$
$$\ge 1-2\exp\left(-cn\min\left(\frac{\epsilon^2}{C^2},\frac{\epsilon}{C}\right)\right)$$

Combining these two inequalities and $m\ge n$, the lemma is proved. □

Lemma 9: For $x\sim CN(0,1)$ and any $r>0$, $\Pr(|x|\le r)<r^2$.

Proof: By the definition of $CN(0,1)$, $\Pr(|x|\le r)$ is the equivalent to $\Pr(\|\mathbf{y}\|\le r)$ for $\mathbf{y}\in\mathbb{R}^2$ and sampled from $N(0,\mathbf{I}_{2\times2}/2)$, which has a probability density function of $\frac{1}{\pi}\exp(-\|\mathbf{y}\|^2)$. This function is maximized at $\mathbf{y}=\mathbf{0}$ with a value of $1/\pi$, and as a result, $\Pr(\|\mathbf{y}\|\le r)<\pi r^2\cdot\frac{1}{\pi}=r^2$. □

Lemma 10: Given a vector $\mathbf{x}\in\mathbb{R}^m$ sampled from $CN(0,\mathbf{I}_{m\times m})$, for all $\mathbf{y}\in\mathbb{R}^m$ satisfies $\frac{1}{m}\sum_{i=1}^m|y_i|^2\le t^2$ and $t\ge n/m$, with probability at least $1-m\exp(-n/6)$, we have

$$\frac{1}{m}\sum_{i=1}^m\max\left(\frac{|y_i|}{|x_i|},1\right)\le(4+\sqrt{2}l)t$$

for $l=\max(0,\lfloor-\log_2 t\rfloor)$.

Proof: WLOG we may rearrange the indices and assume that $|x_1|\le|x_2|\le\cdots\le|x_m|$. Then Lemma 11 implies that

$$\Pr(|x_j|>\sqrt{j/2m})\ge1-\exp(-j/6).\tag{43}$$

Applying a union bound,

$$\Pr\left(|x_j|>\sqrt{j/2m}\text{ for all }j\ge n\right)\ge1-m\exp(-n/6).\tag{44}$$

Denote the set of natural numbers by $\mathcal{N}$, and when the event in (44) holds, for all $j\ge n$ we have

$$\sum_{i\in\mathcal{N}:j\le i<2j}\frac{|y_i|}{|x_i|}\le\frac{1}{|x_j|}\sum_{i\in\mathcal{N}:j\le i<2j}|y_i|\tag{45}$$
$$\le\frac{1}{|x_j|}\sqrt{j\sum_{i\in\mathcal{N}:j\le i<2j}|y_i|^2}\le mt\sqrt{2}.$$

Combining (45) for $j=tm,2tm,4tm,\cdots,2^ltm$ (l is the largest integer such that $2^lt<1$) and $j=m/2$, we have

$$\sum_{i\in\mathcal{N}:tm\le i\le m}\frac{|y_i|}{|x_i|}\le(2+l)mt\sqrt{2}.\tag{46}$$

In addition, it is clear that

$$\sum_{i\in\mathcal{N}:1\le i<tm}\max\left(\frac{|y_i|}{|x_i|},1\right)\le tm.\tag{47}$$

Combining (46) and (47), the lemma is proved. □

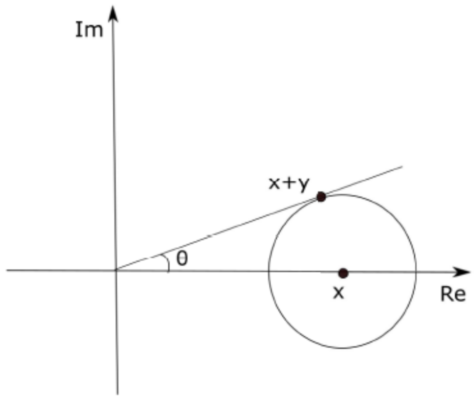

Fig. 2. Visualization of the proof of Lemma 12 when $r < 1$.

Lemma 11: For a random vector $\mathbf{x} \in \mathbf{C}^m$ sampled from $CN(0, \mathbf{I}_{m\times m})$, we have

$$\Pr\left(\sum_{i=1}^m I(|x_i| \le r) < 2r^2 m\right) > 1 - \exp(-r^2 m/3)$$

for all $r > 0$.

Proof: We apply Lemma 15 with $p = \Pr(|x_1| \le r)$ and $\delta = 2r^2/p - 1$. Applying Lemma 9,

$$\delta p = 2r^2 - \Pr(|x_1| \le r) > r^2,$$

which implies Lemma 11. □

Lemma 12: For any complex numbers $x, y \in \mathbb{C}$, $|\text{phase}(x+y) - \text{phase}(x)| \le \min(2\frac{|y|}{|x|}, 2)$. Similarly, for any vector $\mathbf{u}, \mathbf{v} \in \mathbb{C}^m$, $\|\frac{\mathbf{u}+\mathbf{v}}{\|\mathbf{u}+\mathbf{v}\|} - \frac{\mathbf{v}}{\|\mathbf{v}\|}\| \le \min(2\frac{\|\mathbf{u}\|}{\|\mathbf{v}\|}, 2)$.

Proof: WLOG we only need to prove the first sentence and we may assume that $x = 1$ and $|y| = r$. Then $\text{phase}(x) = e^{i0} = 1$, and on the complex plane, $x+y$ lies on a circle center at 1 with radius r.

When $r \ge 1$, $|\text{phase}(x+y) - \text{phase}(x)|$ is maximized when $y = -r$ and $\text{phase}(x+y) = -1$, then we have $|\text{phase}(x+y) - \text{phase}(x)| = 2$.

When $r < 1$, we would like to find a point on the circle such that its direction is as far from the direction of x-axis as possible. As visualized in Figure 2, $|\text{phase}(x+y) - \text{phase}(x)|$ is achieved when the line connecting $x+y$ and the origin is tangent to the circle. It implies that the maximal value is $|e^{i\theta} - 1|$, where $\theta = \sin^{-1} r$. Then we have the estimation $|e^{i\theta} - 1| = 2\sin(\theta/2) = \sin(\theta)/\cos(\theta/2) \le \sqrt{2}\sin(\theta) = \sqrt{2}r$ (the inequality uses the fact that $\theta \le \pi/2$).

Combining these two cases, Lemma 12 is proved. □

Lemma 13 (Sum of sub-gaussian variables, Proposition 5.10 in [27]): Given $X_1, \cdots, X_n$ i.i.d. from a distribution with zero mean and sub-gaussian norm defined by $\|X\|_{\psi_2} = \sup_{p\ge 1} p^{-1/2}(\mathbb{E}|X|^p)^{1/p}$, then

$$\Pr\left(\left|\frac{1}{n}\sum_{i=1}^n X_i\right| \ge t\right) \le \exp\left(-\frac{cnt^2}{\|X\|_{\psi_2}^2} + 1\right)$$

Lemma 14 (Sum of sub-exponential variables, Corollary 5.17 in [27]): Given $X_1, \cdots, X_n$ i.i.d. from a distribution with zero mean and sub-exponential norm defined by $\|X\|_{\psi_1} = \sup_{p\ge 1} p^{-1}(\mathbb{E}|X|^p)^{1/p}$, then

$$\Pr\left(\left|\frac{1}{n}\sum_{i=1}^n X_i\right| \ge t\right) \le 2\exp\left(-cn\min\left(\frac{t^2}{\|X\|_{\psi_1}^2}, \frac{t}{\|X\|_{\psi_1}}\right)\right)$$

Lemma 15: $X_1, X_2, \cdots$ are i.i.d. Bernoulli variables with expectation p, then for any $\delta > 1$,

$$\Pr(\sum_{i=1}^m X_i > (1+\delta)pm) \le \exp(-m\delta p/3).$$

Proof: It follows from [21, Theorem 4.4] and the observation that when $\delta > 1$, $(1+\delta)\log(1+\delta) > \frac{4}{3}\delta$. □

Lemma 16: For $\mathbf{v} \sim CN(0, \mathbf{I})$, $\Pr(|\frac{1}{m}\|\mathbf{v}\|^2 - 1| > t) < 2\exp\left(-cm\min\left(\frac{t^2}{C^2}, \frac{t}{C}\right)\right)$

Proof: We remark that $\|\mathbf{v}\|^2 - m = \sum_{i=1}^m((\Re(v_i)^2 - 1/2) + (\Im(v_i)^2 - 1/2))$, and both $\Re(v_i)$ and $\Im(v_i)$ are i.i.d. sampled from $N(0, \frac{1}{2})$. Since sub-gaussian squared is sub-exponential [27, Lemma 5.14] with mean $1/2m$, and after centering, a sub-exponential distribution is still sub-exponential [27, Remark 5.18], $\Re(v_i)^2 - \frac{1}{2}$ and $\Im(v_i)^2 - \frac{1}{2}$ are i.i.d. sampled from a sub-exponential distribution with sub-exponential norm smaller than a constant C. Applying Lemma 14, Lemma 16 is proved. □

Lemma 17: For any $\mathbf{x} \sim CN(0, \mathbf{I}_{m\times m})$, $\Pr(\|\mathbf{x}\| > t) \le 2m\exp(-t^2/m)$.

Proof: It follows from the classic tail bound: $\Pr(|N(0,1)| > t) \le \exp(-t^2/2)$ and the fact that $\text{Re}(x_i), \text{Im}(x_i) \sim N(0, 1/2)$ for all $1 \le i \le m$, we have

$$\begin{aligned}&\Pr(|\text{Re}(x_i)| > t/\sqrt{2m}) = \Pr(|\text{Im}(x_i)| > t/\sqrt{2m})\\ &= \Pr(|N(0,1)| > t/\sqrt{m}) \le \exp(-t^2/m).\end{aligned}$$

Applying a union bound of all real components of imaginary components of each element of $\mathbf{x}$ (there are $2m$ of them), Lemma 17 is proved:

$$\begin{aligned}&\Pr(\|\mathbf{x}\| > t)\\ &\le \sum_{i=1}^m \left(\Pr(|\text{Re}(x_i)| > t/\sqrt{2m}) + \Pr(|\text{Im}(x_i)| > t/\sqrt{2m})\right)\\ &\le 2m\exp(-t^2/m).\end{aligned}$$

□

III. Discussion

The current paper justifies the convergence of the alternating minimization algorithm with random initialization for phase retrieval. Specifically, we demonstrate that it succeeds with $m > Mn^{1.5}\log^{3.5} n$ for some $M > 0$. A future direction is to improve the sample complexity, possibly via more sophisticated arguments, so that it matches the empirical observation that the algorithm succeeds with $m > O(n)$. It would also be interesting to compare the decoupling approach in this work and the leave-one-out approach in [10], both in phase retrieval and in other problems.

References

[1] S. Bahmani and J. Romberg, "Phase retrieval meets statistical learning theory: A flexible convex relaxation," in *Proc. 20th Int. Conf. Artif. Intell. Statist. (AISTATS)*, Fort Lauderdale, FL, USA, vol. 54, Apr. 2017, pp. 252–260.

[2] H. H. Bauschke, P. L. Combettes, and D. R. Luke, "Hybrid projection–reflection method for phase retrieval," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 20, no. 6, pp. 1025–1034, Jun. 2003.

[3] M. Bayati and A. Montanari, "The LASSO risk for Gaussian matrices," *IEEE Trans. Inf. Theory*, vol. 58, no. 4, pp. 1997–2017, Apr. 2012.

[4] T. T. Cai, X. Li, and Z. Ma, "Optimal rates of convergence for noisy sparse phase retrieval via thresholded wirtinger flow," *Ann. Statist.*, vol. 44, no. 5, pp. 2221–2251, Oct. 2016.

[5] E. J. Candès and X. Li, "Solving quadratic equations via PhaseLift when there are about as many equations as unknowns," *Found. Comput. Math.*, vol. 14, no. 5, pp. 1017–1026, Jun. 2013.

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

[7] E. J. Candès, T. Strohmer, and V. Voroninski, "PhaseLift: Exact and stable signal recovery from magnitude measurements via convex programming," *Commun. Pure Appl. Math.*, vol. 66, no. 8, pp. 1241–1274, Nov. 2012.

[8] A. Chai, M. Moscoso, and G. Papanicolaou, "Array imaging using intensity-only measurements," *Inverse Probl.*, vol. 27, no. 1, Dec. 2010, Art. no. 015005.

[9] Y. Chen and E. Candès, "Solving random quadratic systems of equations is nearly as easy as solving linear systems," in *Advances in Neural Information Processing Systems*, C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, Eds. New York, NY, USA: Curran Associates, 2015, pp. 739–747.

[10] Y. Chen, Y. Chi, J. Fan, and C. Ma, "Gradient descent with random initialization: Fast global convergence for nonconvex phase retrieval," *Math. Program.*, vol. 176, nos. 1–2, pp. 5–37, Feb. 2019.

[11] K. R. Davidson and S. J. Szarek, "Local operator theory, random matrices and Banach spaces," in *Handbook of the Geometry of Banach Spaces*, vol. 1. Amsterdam, The Netherlands: Elsevier, 2001, pp. 317–366.

[12] V. Elser, T.-Y. Lan, and T. Bendory, "Benchmark problems for phase retrieval," *SIAM J. Imag. Sci.*, vol. 11, no. 4, pp. 2429–2455, Oct. 2018.

[13] J. R. Fienup, "Reconstruction of an object from the modulus of its Fourier transform," *Opt. Lett.*, vol. 3, no. 1, p. 27, Jul. 1978.

[14] J. R. Fienup, "Phase retrieval algorithms: A comparison," *Appl. Opt.*, vol. 21, no. 15, p. 2758, Aug. 1982.

[15] R. W. Gerchberg, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik*, vol. 35, pp. 237–246, 1972.

[16] T. Goldstein and C. Studer, "PhaseMax: Convex phase retrieval via basis pursuit," *IEEE Trans. Inf. Theory*, vol. 64, no. 4, pp. 2675–2689, Apr. 2018.

[17] D. Gross, F. Krahmer, and R. Kueng, "A partial derandomization of PhaseLift using spherical designs," *J. Fourier Anal. Appl.*, vol. 21, no. 2, pp. 229–266, Oct. 2014.

[18] P. Hand and V. Voroninski, "Corruption robust phase retrieval via linear programming, " Dec. 2016, *arXiv:1612.03547*. [Online]. Available: https://arxiv.org/abs/1612.03547

[19] P. Hand and V. Voroninski, "An elementary proof of convex phase retrieval in the natural parameter space via the linear program PhaseMax," *Commun. Math. Sci.*, vol. 16, no. 7, pp. 2047–2051, 2018.

[20] S. Marchesini, Y.-C. Tu, and H.-T. Wu, "Alternating projection, ptychographic imaging and phase synchronization," *Appl. Comput. Harmon. Anal.*, vol. 41, no. 3, pp. 815–851, Nov. 2016.

[21] M. Mitzenmacher and E. Upfal, *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[22] P. Netrapalli, P. Jain, and S. Sanghavi, "Phase retrieval using alternating minimization," *IEEE Trans. Signal Process.*, vol. 63, no. 18, pp. 4814–4826, Sep. 2015.

[23] F. Salehi, E. Abbasi, and B. Hassibi, "Learning without the phase: Regularized phasemax achieves optimal sample complexity," in *Advances in Neural Information Processing Systems*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds. New York, NY, USA: Curran Associates, 2018, pp. 8654–8665.

[24] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, "Phase retrieval with application to optical imaging: A contemporary overview," *IEEE Signal Process. Mag.*, vol. 32, no. 3, pp. 87–109, May 2015.

[25] M. Soltanolkotabi, "Structured signal recovery from quadratic measurements: Breaking sample complexity barriers via nonconvex optimization," *IEEE Trans. Inf. Theory*, vol. 65, no. 4, pp. 2374–2400, Apr. 2019.

[26] J. Sun, Q. Qu, and J. Wright, "A geometric analysis of phase retrieval," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 2379–2383.

[27] R. Vershynin, "Introduction to the non-asymptotic analysis of random matrices," in *Compressed Sensing*, Y. C. Eldar and G. Kutyniok, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2012, pp. 210–268.

[28] I. Waldspurger, "Phase retrieval with random Gaussian sensing vectors by alternating projections," *IEEE Trans. Inf. Theory*, vol. 64, no. 5, pp. 3301–3312, May 2018.

[29] I. Waldspurger, A. d'Aspremont, and S. Mallat, "Phase recovery, MaxCut and complex semidefinite programming," *Math. Program.*, vol. 149, nos. 1–2, pp. 47–81, Dec. 2013.

[30] G. Wang and G. Giannakis, "Solving random systems of quadratic equations via truncated generalized gradient flow," in *Advances in Neural Information Processing Systems*, D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, Eds. New York, NY, USA: Curran Associates, 2016, pp. 568–576.

[31] H. Zhang, Y. Chi, and Y. Liang, "Provable non-convex phase retrieval with outliers: Median truncated wirtinger flow," in *Proc. 33rd Int. Conf. Int. Conf. Mach. Learn. (ICML)*, vol. 48, 2016, pp. 1022–1031.

[32] H. Zhang and Y. Liang, "Reshaped wirtinger flow for solving quadratic system of equations," in *Advances in Neural Information Processing Systems*, D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, Eds. New York, NY, USA: Curran Associates, 2016, pp. 2622–2630.

[33] T. Zhang, "Phase retrieval using alternating minimization in a batch setting," *Appl. Comput. Harmon. Anal.*, to be published. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S1063520319300181