# ON DC BASED METHODS FOR PHASE RETRIEVAL

MENG HUANG *, MING-JUN LAI †, ABRAHAM VARGHESE ‡, AND ZHIQIANG XU §

**Abstract.** In this paper, we develop a new computational approach which is based on minimizing the difference of two convex functionals (DC) to solve a broader class of phase retrieval problems. The approach splits a standard nonlinear least squares minimizing function associated with the phase retrieval problem into the difference of two convex functions and then solves a sequence of convex minimization sub-problems. For each subproblem, the Nesterov's accelerated gradient descent algorithm or the Barzilai-Borwein (BB) algorithm is used. In the setting of sparse phase retrieval, a standard $\ell_1$ norm term is added into the minimization mentioned above. The subproblem is approximated by a proximal gradient method which is solved by the shrinkage-threshold technique directly without iterations. In addition, a modified Attouch-Peypouquet technique is used to accelerate the iterative computation. These lead to more effective algorithms than the Wirtinger flow (WF) algorithm and the Gauss-Newton (GN) algorithm and etc.. A convergence analysis of both DC based algorithms shows that the iterative solutions is convergent linearly to a critical point and will be closer to a global minimizer than the given initial starting point. Our study is a deterministic analysis while the study for the Wirtinger flow (WF) algorithm and its variants, the Gauss-Newton (GN) algorithm, the trust region algorithm is based on the probability analysis. In particular, the DC based algorithms are able to retrieve solutions using a number $m$ of measurements which is about twice of the number $n$ of entries in the solution with high frequency of successes. When $m \approx n$, the $\ell_1$ DC based algorithm is able to retrieve sparse signals. Finally, the paper discusses the nonexistence of the solution to the exact recovery of the phase retrieval problem for arbitrary given measurement values. In addition, for a given set of measurement values, if there is a solution, an estimate of the upper bound of the number of distinct solutions is also given.

**Key words.** phase retrieval, DC method, nonlinear least squares, convex analysis

**AMS subject classifications.** 68Q25,52A41

## 1. Introduction.

**1.1. Phase retrieval.** The phase retrieval problem has been extensively studied in the last 40 years due to its numerous applications, such as X-ray diffraction, crystallography, electron microscopy, optical imaging and etc.. See, e.g. [11], [30], [31], [34], [39], [20], and [17]. In particular, see [27] for an explanation of the image recovery from the phaseless measurements and a survey of recent research results. Mathematically, the phase retrieval problem or simply called phase retrieval problem can be stated as follows. Given measurement vectors $\mathbf{a}_i \in \mathbb{R}^n$ (or $\in \mathbb{C}^n$), and the measurement values $b_i \geq 0$, we would like to recover an unknown signal $\mathbf{x} \in \mathbb{R}^n$ (or $\in \mathbb{C}^n$) through a set of quadratic equations:

$$b_1 = |\langle \mathbf{a}_1, \mathbf{x}\rangle|^2, \ldots, b_m = |\langle \mathbf{a}_m, \mathbf{x}\rangle|^2. \tag{1.1}$$

Note that for any $c \in \mathbb{R}^n$ (or $\in \mathbb{C}^n$) with $|c| = 1$, we have $b_i = |\langle \mathbf{a}_i, c\mathbf{x}\rangle|^2$ for all $i$. Thus we can only hope to recover $\mathbf{x}$ up to a unimodular constant. One fundamental problem

*Institute of Computational Mathematics, Academy of Mathematics and Systems of Science, Chinese Academic Sciences, Beijing 100190, China. (hm@lsec.cc.ac.cn).

†This author is partly supported by the National Science Foundation under grant DMS-1521537. Department of Mathematics, The University of Georgia, Athens, GA 30602. (mjlai@uga.edu).

‡Department of Mathematics, Shenandoah University, Winchester, VA 22601. (opticsabru@gmail.com).

§Zhiqiang Xu was supported by NSFC grant (11422113, 91630203, 11331012) and by National Basic Research Program of China (973 Program 2015CB856000). Institute of Computational Mathematics, Academy of Mathematics and Systems of Science, Chinese Academic Sciences, Beijing 100190, China. (xuzq@lsec.cc.ac.cn).

in phase retrieval is to give the minimal $m$ for which there exists $A = (\mathbf{a}_1, \ldots, \mathbf{a}_m)^\top$ which can recover $\mathbf{x}$ up to a unimodular constant. For the real case, it is well known that the minimal measurement number $m$ is $2n-1$ [4]. For the complex case $\mathbb{C}^n$, this question remains open. Conca, Edidin, Hering and Vinzant [14] proved $m \geq 4n-4$ generic measurement vectors $F = (\mathbf{f}_1, \ldots, \mathbf{f}_m)^\top$ have phase retrieval property for $\mathbb{C}^n$ and they furthermore show that $4n-4$ is sharp if $n$ is in the form of $2^k+1,\ k \in \mathbb{Z}_+$. In [41], for the case $n = 4$, Vinzant present $11 = 4n-5 < 4n-4$ measurement vectors which have phase retrieval property for $\mathbb{C}^4$ which implies that $4n-4$ is not sharp for some dimension $n$. There is a similar study for the sparse phase retrieval. See [42].

There are many computational algorithms available to find a true signal $\mathbf{x}$ up to a phase factor. It is common folklore that for given $\mathbf{a}_i, i = 1, \ldots, m$, we may not be able to find a solution $\mathbf{x}$ for any given vector $\mathbf{b} = (b_1, \cdots, b_m)^\top$, e.g. a perturbation of the exact observed value vector $\mathbf{b}^*$. We shall give this fact a mathematical explanation (see Theorem 2.3 in a later section). Thus, the phase retrieval problem is usually formulated as follows:

$$\min_{\mathbf{x}\in\mathbb{R}^n \text{ or } \mathbb{C}^n} \sum_{i=1}^m (|\langle \mathbf{a}_i, \mathbf{x}\rangle|^2 - b_i)^2. \tag{1.2}$$

Although it is not a convex minimization problem, the minimizing functional is a nice differentiable function and hence, many computational algorithms can be developed and they are very successful actually. A gradient descent method (called Wirtinger flow in the complex variable setting) is developed by Candès et al. in [12]. They show that the WF algorithm converges to the true signal up to a global phase factor with high probability provided the measurement vectors are $m = O(n \log n)$ Gaussian measurements. Many variants of Wirtinger flow algorithms were developed. See [11], [9], [13], [46], and [8] for Truncated WF, Thresholded WF, Reshaped WF, Accelerated WF, and etc.. In [22], Gao and Xu propose a Gauss-Newton (GN) algorithm to find a minimizer of (1.2). They proved that, for the real signal, the GN algorithm can converge to the global optimal solution quadratically with $O(n \log n)$ measurements starting from a good initial solution. Indeed, Gao and Xu provided a formula for the initialization which is much better than the initialization in [12] in numerical experiments. Another approach for the minimization (1.2) is called the true region method which was studied in [40] where a geometric description of the landscape function $f(\mathbf{x}) = \sum_{i=1}^m (|\langle \mathbf{a}_i, \mathbf{x}\rangle|^2 - b_i)^2$ is given. To recover sparse signals from the measurements (1.1), a standard approach is to add $\ell_1$ term $\lambda\|\mathbf{x}\|_1$ to (1.2) or use the projected gradient method as discussed in [38].

**1.2. Our contribution.** In this paper, we consider a broader class of phase retrieval problem which includes standard phase retrieval as a special case. We aim to recover $\mathbf{x} \in \mathbb{R}^n$ (or $\in \mathbb{C}^n$) from nonlinear measurements

$$b_i = f(\langle \mathbf{a}_i, \mathbf{x}\rangle), \quad i = 1, \ldots, m, \tag{1.3}$$

where $f : \mathbb{C} \to \mathbb{R}_+$ is a continuous convex function which satisfies the following coercive condition:

$$f(x) \to \infty \text{ when } |x| \to \infty.$$

If we take $f(x) = |x|^2$, then it reduces to the standard phase retrieval. To guarantee the unique recovery of $\mathbf{x}$, it has been proved that the measurement number satisfies

$m \geq n+1$ for the real case ($2n+1$ for the complex case, respectively) (see [26]). Recovering $\mathbf{x}$ from the nonlinear observation is also raised in many areas, such as neural network etc. (cf. [6, 36]).

To reconstruct $\mathbf{x}$ by solving (1.3), we can formulate it as

$$\min_{\mathbf{x}\in\mathbb{R}^n \text{ or } \mathbb{C}^n} \sum_{i=1}^{m} (f(\langle \mathbf{a}_i, \mathbf{x}\rangle) - b_i)^2. \tag{1.4}$$

We approach it by using the standard technique for a difference of convex minimizing functionals. Indeed, for the case $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{a}_i \in \mathbb{R}^n$, let $F(\mathbf{x}) = \sum_{i=1}^m (f(\langle \mathbf{a}_i, \mathbf{x}\rangle) - b_i)^2$ be the minimizing functional. As it is not convex, we can write it as

$$F(\mathbf{x}) = F_1(\mathbf{x}) - F_2(\mathbf{x}) := \sum_{i=1}^{m} \left(f^2(\langle \mathbf{a}_i, \mathbf{x}\rangle) + b_i^2\right) - \sum_{i=1}^{m} 2b_i f(\langle \mathbf{a}_i, \mathbf{x}\rangle).$$

Note that $f$ is a convex function with function value $f(x) \geq 0$ for all $x \in \mathbb{R}$. Then $F_1$ and $F_2$ are convex functions. The minimization (1.4) will be approximated by

$$\mathbf{x}^{(k+1)} := \arg\min_{\mathbf{x}} F_1(\mathbf{x}) - \nabla F_2(\mathbf{x})^\top (\mathbf{x} - \mathbf{x}^{(k)}) \tag{1.5}$$

for any given $\mathbf{x}^{(k)}$. We call this algorithm as DC based algorithm following from the ideas in [23], where the sparse solutions of underdetermined linear system were studied. Due to the nice properties of $F_1$ and $F_2$, we will be able to establish much better results than those in [23]. When $\mathbf{x} \in \mathbb{C}^n$ and $\mathbf{a}_j \in \mathbb{C}^n$, $j = 1, \ldots, m$, we have to write $\mathbf{x} = \mathbf{x}_R + \sqrt{-1}\mathbf{x}_I$ and similar for $\mathbf{a}_j$. Letting $\mathbf{c} = [\mathbf{x}_R^\top \; \mathbf{x}_I^\top]^\top \in \mathbb{R}^{2n}$, we view $F_1(\mathbf{x})$ as a functions in $G_1(\mathbf{c}) = F_1(\mathbf{x}_R + \sqrt{-1}\mathbf{x}_I)$. Then $G_1(\mathbf{c})$ is a convex function of variable $\mathbf{c}$. Similarly, $G_2(\mathbf{c}) = F_2(\mathbf{x}_R + \sqrt{-1}\mathbf{x}_I)$ is a convex function of $\mathbf{c}$. For convenience, we simply discuss the case when $\mathbf{x}$, $\mathbf{a}_j$, $j = 1, \ldots, m$ are real. The complex variable setting can be treated in the same fashion.

The above minimization (1.5) is a convex minimization problem with differentiable functional for each $k$. We can solve it by using the standard gradient descent method with Nestrov's acceleration (cf. [33]) or by using the Barzilai-Borwein (BB) method (cf. [5]). There are several nice properties of this DC based approach. We can show that

$$F(\mathbf{x}^{(k+1)}) \leq F(\mathbf{x}^{(k)}) - \ell\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|^2$$

for some constant $\ell > 0$. That is, $F(\mathbf{x}^{(k)}), k \geq 1$ is strictly decreasing sequence and hence, the sequence $\mathbf{x}^{(k)}$ will not converge to a local maximum. Also, we can prove the sequence $\{\mathbf{x}^{(k)}\}_{k=1}^\infty$ converges to a critical point $\mathbf{x}^*$. Using the Kurdyka-Łojasiewicz inequality, we can also show $\|\mathbf{x}^{(k)} - \mathbf{x}^*\| \leq C\theta^k$ for $\theta \in (0,1)$. If the function $F(\mathbf{x})$ has the property that any global minimizer $\mathbf{x}^\star$ is a local minimizer over a neighborhood $N(\mathbf{x}^\star)$ and the initial point $\mathbf{x}^{(1)}$ is within $N(\mathbf{x}^\star)$, then the DC based algorithm will converge to the global minimizer. Actually, the function $F(\mathbf{x})$ has such property for standard phase retrieval problem and the initial point is chosen by a careful initialization. Our numerical experiments show that our DC based algorithm can retrieve solutions when $m \approx 2n$.

Furthermore, we develop an $\ell_1$ DC based algorithm to reduce the number of measurements and recover sparse signals. That is, starting from $\mathbf{x}^{(k)}$, we solve

$$\mathbf{x}^{(k+1)} := \arg\min \lambda\|\mathbf{x}\|_1 + F_1(\mathbf{x}) - \nabla F_2(\mathbf{x})^\top (\mathbf{x} - \mathbf{x}^{(k)}) \tag{1.6}$$

using a proximal gradient method, where $\lambda > 0$ is a parameter. The convergence of the $\ell_1$ DC based algorithm can be established similar to the DC base algorithm. To accelerate the convergence of the $\ell_1$ DC based algorithm, we use Attouch-Peypouquet's acceleration (cf. [3]). To have a better initialization, we use the projection technique (cf. [19]). In addition, the hard thresholding operator is used to project each iteration onto the set of sparse vectors. With these updates, the algorithm works very well. The numerical experiments of the modified $\ell_1$ DC based algorithm can recover sparse signals as long as $m \approx n$.

**1.3. Organization.** The paper is organized as follows. First, using tools of algebraic geometry, we explain some existence for phase retrieval and give an estimate of how many distinct solutions in section 2. In section 3, we give the analysis of convergence for our DC based algorithm. Accelerated gradient descent methods, Nesterov's accelerated technique and the BB technqie for inner iterations will be discussed in section 4. Furthermore, we will study the $\ell_1$ based algorithm for retrieving sparse signals and discuss the convergence in section 5. Our numerical experimental results are collected in section 6, where we show the performance of our DC based algorithms and comparison with the Gauss-Newton algorithm for general signals and sparse signals. Mainly, we will show that the DC based algorithm is able to retrieve signals when $m \approx 2n$ with high frequency of successes. In addition, our $\ell_1$ DC based algorithm with the update techniques is able to retrieve sparse signals with high frequency of successes when $m \approx n$.

**2. On Existence and Number of Phase Retrieval Solutions.** In this section, we shall discuss the existence of phase retrieval solution and give an estimate of the number of distinct solutions. To do so, we first recall PhaseLift (cf. [10]) which shows the connection between phase retrieval and low-rank matrix recovery.

Letting $X = \mathbf{x}\mathbf{x}^\top$ and $A_j = \mathbf{a}_j\mathbf{a}_j^\top$, $j = 1, \ldots, m$, the constrains in (1.1) can be rewritten as

$$b_j = \mathrm{tr}(A_j X), \quad j = 1, \ldots, m,$$

where $\mathrm{tr}(\cdot)$ is the trace operator.

Note that a scaling of $\mathbf{x}$ by a unimodular constant $c$ would not change $X$. Indeed, $(c\mathbf{x})(c\mathbf{x})^\top = |c|^2\mathbf{x}\mathbf{x}^\top = \mathbf{x}\mathbf{x}^\top = X$. Conversely, given a positive semi definite matrix $X$ of rank 1, there exists a vector $\mathbf{x}$ such that $X = \mathbf{x}\mathbf{x}^\top$. So the phase retrieval problem can be recast as a matrix recover problem (cf. [10]): Find $X \in \mathcal{M}_1$ satisfying linear measurements: $\mathrm{tr}(A_j X) = b_j, j = 1, \ldots, m$, where $\mathcal{M}_r = \{X \in \mathbb{R}^{n\times n} : \mathrm{rank}(X) = r\}$. It also can be considered as a low rank matrix recovery problem:

$$\min\{\mathrm{rank}(X) : \mathrm{tr}(A_j X) = b_j, j = 1, \ldots, m, X \succeq 0\}. \tag{2.1}$$

As we have pointed out above, for any given $b_j \geq 0$, $j = 1, \ldots, m$, there may not have a matrix $X \in \mathcal{M}_r$ with $r < n$ satisfying the constraints exactly. Unless $b_j \geq 0$ are exactly the measurement values from a matrix $X$ so that we can use the minimization (2.1) to find the solution $X$, we have to reformulate the above problem otherwise:

$$\min\{\sum_{i=1}^m |\mathrm{tr}(A_j X) - b_j|^2, X \in \mathcal{M}_r, X \geq 0\}. \tag{2.2}$$

As $\mathcal{M}_r$ is a closed set, the above least squares problem will have a bounded solution if the following coercive condition holds:

$$\sum_{i=1}^{m} |\mathrm{tr}(A_j X) - b_j|^2 \to \infty \text{ when } \|X\|_F \to \infty.$$

In the case that the above coercive condition does not hold, one has to use other conditions to insure that the minimizer in (2.2) is bounded. For example, if there is a matrix $X_0$ which is orthogonal to $A_j$ in the sense that $\mathrm{tr}(A_j X_0) = 0$ for all $j = 1, \ldots, m$, then the coercive condition will not hold as one can let $X = \ell X_0$ with $\ell \to \infty$.

We are now ready to discuss the existence of the solution of phase retrieval problem. Let $\mathcal{M}_r$ be the set of matrices of size $n \times n$ with rank $r$ and $\overline{\mathcal{M}_r}$ be the set of all matrices with rank $\le r$. It is known that dimension of $\mathcal{M}_r$ is $2nr - r^2$ (cf. Proposition 12.2 in [24] for a proof). Since $\overline{\mathcal{M}_r}$ is the closure of $\mathcal{M}_r$ in the Zariski sense (cf. [45]) and hence the dimension of $\overline{\mathcal{M}_r}$ is also $2nr - r^2$. Furthermore, it is clear that $\overline{\mathcal{M}_r}$ is an algebraic variety. In fact, $\overline{\mathcal{M}_r}$ is an irreducible variety which is a standard result in algebraic geometry. To make the paper self-contain, we present a short proof.

Lemma 2.1. *$\overline{\mathcal{M}_r}$ is an irreducible variety.*

*Proof.* Denote by $GL(n)$ the set of invertible $n \times n$ matrices. Consider the action of $GL(n) \times GL(n)$ on $M_n(R)$ given by: $(G_1, G_2) \cdot M \mapsto G_1 M G_2^{-1}$, for all $G_1, G_2 \in GL(n)$. Fix a rank $r$ matrix $M$. Then the variety $\mathcal{M}_r$ is the orbit of $M$. Hence, we have a surjective morphism, a regular algebraic map described by polynomials, from $GL(n) \times GL(n)$ onto $\mathcal{M}_r$. Since $GL(n) \times GL(n)$ is an irreducible variety, so is $\mathcal{M}_r$. Hence, the closure $\overline{\mathcal{M}_{r_g}}$ of the irreducible set $\mathcal{M}_{r_g}$ is also irreducible *c.f* (cf. Example I.1.4 in [25]). ∎

Define a map

$$\mathcal{A} : \mathcal{M}_1 \to \mathbb{R}^m$$

by projecting any matrix $X \in \mathcal{M}_1$ to $(b_1, \cdots, b_m)^\top \in \mathbb{R}^m$ in the sense that

$$\mathcal{A}(X) = (\mathrm{tr}(A_1 X), \cdots, \mathrm{tr}(A_m X))^\top.$$

We define the range $\mathcal{R}_+ = \{\mathcal{A}(X) : X \in \mathcal{M}_1, X \succeq 0\}$ and the range $\mathcal{R} = \{\mathcal{A}(X) : X \in \mathcal{M}_1\}$ of the map $\mathcal{A}$. It is clear that the dimension of $\mathcal{R}_+$ is less than or equal to the dimension of $\mathcal{R}$. As the projection $\mathcal{A}$ is a regular map since each coordinate of the map $\mathcal{A}$ is a linear polynomial in entries of matrices, we expect that $\dim(\mathcal{R})$ is less than or equal to the dimension of the $\mathcal{M}_1$ which is equal to $2n-1$. If $m > 2n-1$, then $\mathcal{R}$ is not able to occupy the whole space $\mathbb{R}^m$. The Lebesgue measure of the range $\mathcal{R}$ is zero and hence, randomly choosing a vector $\mathbf{b} = (b_1, \cdots, b_m)^\top \in \mathbb{R}^m$, e.g. $\mathbf{b} \in \mathbb{R}^m_+$ will not be in $\mathcal{R}$ most likely and hence, not in $\mathcal{R}_+$. Thus, there will not be a solution $X \in \mathcal{M}_1$ such that $\mathcal{A}(X) = \mathbf{b}$.

Certainly, these intuitions should be made more precise. Recall the following result from Theorem 1.25 in Sec 6.3 of [37].

Lemma 2.2. *Let $f : X \to Y$ be a regular map between irreducible varieties. Suppose that $f$ is surjective: $f(X) = Y$ , and that* $\dim(X) = n$, $\dim(Y) = m$. *Then $m \le n$, and*

1. *for any $y \in Y$ and for any component $F$ of the fiber $f^{-1}(y)$,* $\dim(F) \ge n - m$;
2. *there exists a nonempty open subset $U \subset Y$ such that* $\dim(f^{-1}(y)) = n - m$ *for $y \in U$.*

We are now ready to prove

THEOREM 2.3. *If one chooses randomly a vector* $\mathbf{b} = (b_1, \cdots, b_m)^\top \in \mathbb{R}_+^m$ *with* $m > 2n-1$*, the probability of finding a solution* $X$ *to the minimization* (2.1) *is zero. In other words, for almost all the* $\mathbf{b} = (b_1, \cdots, b_m)^\top \in \mathbb{R}_+^m$ *the solution to* (2.1) *is a matrix with rank more than or equal to* 2.

*Proof.* We mainly use Lemma 2.2. Let $X = \overline{\mathcal{M}_1}$ which is an irreducible variety by Lemma 2.1. Let $Y = \{\mathcal{A}(M), M \in \overline{\mathcal{M}_1}\}$, i.e. $Y = \mathcal{R}$ which is also an irreducible variety as it is a continuous image of the irreducible variety $\overline{\mathcal{M}_1}$. Since $\mathcal{A}$ is a regular map, we have $\dim(\mathcal{R}) \leq \dim(\overline{\mathcal{M}_1}) = 2n-1 < m$. Thus, $\mathcal{R}$ is a proper lower dimensional closed subset in $\mathbb{R}^m$. For almost all points in $\mathbb{R}^m$, they do not belong to $\mathcal{R}$. In other words, for almost all points $\mathbf{b} = (b_1, \cdots, b_n) \in \mathbb{R}^m$, there is no matrix $M \in \overline{\mathcal{M}_1}$ such that $\mathcal{A}(M) = \mathbf{b}$ and hence, no matrix $M \in \overline{\mathcal{M}_1}$ with $M \geq 0$ such that $\mathcal{A}(M) = \mathbf{b}$ as the set $\mathcal{R}_+$ is a subset of $\mathcal{R}$. ◻

Note that the above discussion is still valid after replacing $\mathcal{M}_1$ by $\mathcal{M}_r$ with $r < n$. Under the assumption that $m > 2nr - r^2$, we can show that the generalized phase retrieval problem (2.1) may not have a solution for randomly chosen $\mathbf{b} = (b_1, \cdots, b_m) \in \mathbb{R}^m$.

Next define the subset $\chi_{\mathbf{b}} \subset \overline{\mathcal{M}_1}$ by

$$\chi_{\mathbf{b}} = \left\{ M \in \overline{\mathcal{M}_1} \mid \mathcal{A}(M) = \mathbf{b} \text{ and } \mathcal{A}^{-1}(\mathcal{A}(M)) \text{ is zero dimensional} \right\}.$$

As we are working over Noetherian fields like $\mathbb{R}$ or $\mathbb{C}$, it is worthwhile to keep in mind that all zero dimensional varieties over such fields will have only finitely many points. Let us consider those $\mathbf{b} \in \mathbb{R}_+^m$ such that the set $\chi_{\mathbf{b}}$ is nonempty. We are interested in an upper bound on number of solutions one can find via the minimization (2.1) when $\chi_{\mathbf{b}} \neq \emptyset$ (Theorem 2.8). To do so, we need more results from algebraic geometry.

LEMMA 2.4 ([24] Proposition 11.12.). *Let* $X$ *be a quasi-projective variety and* $\pi : X \to \mathbb{R}^m$ *a regular map; let* $Y$ *be closure of the image. For any* $p \in X$*, let* $X_p = \pi^{-1}\pi(p)) \subseteq X$ *be the fiber of* $\pi$ *through* $p$*, and let* $\mu(p) = \dim_p(X_p)$ *be the local dimension of* $X_p$ *at* $p$*. Then* $\mu(p)$ *is an upper-semi-continuous function of* $p$*, in the Zariski topology on* $X$*, i.e. for any* $m$ *the locus of points* $p \in X$ *such that* $\dim_p(X_p) > m$ *is closed in* $X$*. Moreover, if* $X_0 \subseteq X$ *is any irreducible component,* $Y_0 \subseteq Y$ *the closure of its image and* $\mu$ *the minimum value of* $\mu(p)$ *on* $X_0$*, then*

$$\dim(X_0) = \dim(Y_0) + \mu. \tag{2.3}$$

As we saw that $\dim(\mathcal{R}) \leq \dim(\overline{\mathcal{M}_r})$, we can be more precise about these dimensions as shown in the following

LEMMA 2.5. *Assume* $m > \dim(\overline{\mathcal{M}_r})$*. Then* $\dim(\overline{\mathcal{M}_r}) = \dim(\mathcal{R})$ *if and only if* $\chi_{\mathbf{b}} \neq \emptyset$ *for some* $\mathbf{b} \in \mathcal{R}$.

*Proof.* Assume $\dim(\overline{\mathcal{M}_r}) = \dim(\mathcal{R})$. Then using Lemma 2.2, there exists a nonempty open subset $U \subset \mathcal{R})$ such that $\dim(\mathcal{A}^{-1}(\mathbf{b})) = 0$ for all $\mathbf{b} \in U$. This implies that $\chi_{\mathbf{b}}$ has finitely many points. Hence $\chi_{\mathbf{b}} \neq \emptyset$.

We now prove the converse. Assume $\chi_{\mathbf{b}} \neq \emptyset$. We will apply Lemma 2.4 above by setting $X = \overline{\mathcal{M}_r}$, $Y = \mathcal{A}(\overline{\mathcal{M}_r})$ and $\pi = \mathcal{A}$. (As we apply lemma , please note that it does not matter whether we take the closure in $\mathbb{P}^m$ or in $\mathbb{C}^m$ since $\mathbb{C}^m$ is an open set in $\mathbb{P}^m$ and the Zariski topology of the affine space $\mathbb{C}^m$ is induced from the Zariski topology of $\mathbb{P}^m$. $\overline{\mathcal{M}_r}$ is an affine variety. In particular, it is a quasi-projective variety.)

By our assumption, $\chi_{\mathbf{b}}$ is not empty. It follows that there is a point $p \in Y$ such that $\pi^{-1}(p)$ is zero dimensional. Since zero is the least dimension possible, we have $\mu = 0$. Hence, using (2.3) above, we have $\dim(\overline{\mathcal{M}_1}) = \dim(\mathcal{R})$. But dimension does not change upon taking closure. So, $\dim(\mathcal{R}) = \dim(\overline{\mathcal{R}})$. ∎

Finally, we need the following

Definition 2.6. *The* degree *of an affine or projective variety of dimension $k$ is the number of intersection points of the variety with $k$ hyperplanes in general position.*

For example, the degree of the algebraic variety $\overline{\mathcal{M}_r}$ is known. See Example 14.4.11 in [21], i.e.

*Example* 2.7. Degree of the algebraic variety $\overline{\mathcal{M}_r}$ is

$$\prod_{i=0}^{n-r-1} \frac{\binom{n+i}{r}}{\binom{r+i}{r}}$$

In particular, the degree of $\mathcal{M}_1$ is

$$\prod_{i=0}^{n-2} \frac{n+i}{1+i}.$$

We are now ready to prove another main result in this section.

Theorem 2.8. *Assume that a given vector $\mathbf{b} \in \mathbb{R}_+^m$ lies in the range $\mathcal{R}_+$. Further assume that $\chi_{\mathbf{b}} \neq \emptyset$. Then, the number of distinct solutions in $\chi_{\mathbf{b}}$ will be less than or equal to* $\displaystyle\prod_{i=0}^{n-2} \frac{n+i}{1+i}$.

*Proof.* When we fix $m$ entries in $\mathbf{b}$, the set of matrices $M$ of rank 1 such that $\mathcal{A}(M) = \mathbf{b}$ are exactly the intersection points of the variety $\overline{\mathcal{M}_1}$ with $m$ hyperplanes, namely the hyperplanes defined by equations of the form $\langle A_i, M\rangle = b_i, i = 1, \cdots, m$. Since $m > \dim(\overline{\mathcal{M}_r}) = 2n-1$, the number of intersection points (matrices of rank 1) would be less than degree of $\overline{\mathcal{M}_1}$ generically. So, in particular, the number of positive semidefinite matrices $M$ of rank 1 such that $\mathcal{A}(M) = \mathbf{b}$ would be no more than the degree of $\overline{\mathcal{M}_1}$. Now using the exact formula for the degree from Example 2.7, the result follows. ∎

**3. The DC based Algorithm for Phase Retrieval.** Recall that we aim to recover $\mathbf{x}$ by minimizing

$$F(\mathbf{x}) = \sum_{i=1}^{m} (f(\langle \mathbf{a}_i, \mathbf{x}\rangle) - b_i)^2 := F_1(\mathbf{x}) - F_2(\mathbf{x}), \tag{3.1}$$

where $F_1(\mathbf{x}) = \sum_{i=1}^m f^2(\langle \mathbf{a}_i, \mathbf{x}\rangle) + b_i^2$ and $F_2(\mathbf{x}) = \sum_{i=1}^m (2b_i f(\langle \mathbf{a}_i, \mathbf{x}\rangle))$. It is easy to see that the minimization in (3.1) can happen in a bounded region $\mathcal{R}$ since the coercive condition $f(x) \to \infty$ when $x \to \infty$. The DC based computational method is as follows. Start from any iterative solution $\mathbf{x}^{(k)}$, we solve the following convex minimization problem:

$$\mathbf{x}^{(k+1)} = \arg\min_{\mathbf{x} \in \mathbb{R}^n} F_1(\mathbf{x}) - \nabla F_2(\mathbf{x}^{(k)})^\top (\mathbf{x} - \mathbf{x}^{(k)}) \tag{3.2}$$

for $k \geq 1$, where $\mathbf{x}^{(1)}$ is an initial guess which will be discussed how to choose later. Without of loss generality, we always assume $\mathbf{x}^{(1)}$ is located in a bounded region $\mathcal{R}$.

Our goal in this section is to show $\mathbf{x}^{(k)}, k \geq 1$ converges to a critical point. Later, we will discuss how to find a global minimization by choosing the initial guess $\mathbf{x}^{(1)}$ appropriately. Although it is standard to solve a convex minimization problem with differentiable minimizing functional, we have to solve (3.2) by using an iterative method. For example, we can use a gradient descent method with various acceleration techniques such as Nesterov's, BB's and other techniques or the Newton method. Hence, there will be two iterative procedures. The iterative procedure for solving (3.2) is an inner iteration which will be discussed in the next section. In this section, we mainly discuss the outer iteration.

We will instate the following assumptions on the function $F_1$ and $F_2$:

(1) The gradient function $\nabla F_1$ has Lipschitz constant $L_1$ in bounded region $\mathcal{R}$. That is, $\|\nabla F_1(\mathbf{x}) - \nabla F_1(\mathbf{y})\|_2 \leq L_1 \|\mathbf{x}-\mathbf{y}\|_2$ for all vectors $\mathbf{x}, \mathbf{y} \in \mathcal{R}$.

(2) $F_2$ is a strongly convex function with parameter $\ell$ in $\mathcal{R}$. That is, $F_2(\mathbf{y}) \geq F_2(\mathbf{x}) + \nabla F_2(\mathbf{x})^\top (\mathbf{y}-\mathbf{x}) + \frac{\ell}{2}\|\mathbf{y}-\mathbf{x}\|^2$ for all vectors $\mathbf{x}, \mathbf{y} \in \mathcal{R}$.

Note that $H = \sum_{i=1}^m 2b_i f^{''}(\mathbf{a}_i^\top \mathbf{x})\mathbf{a}_i\mathbf{a}_i^\top$ is the Hessian matrix of function $F_2 = 2\sum_{i=1}^m b_i f(\mathbf{a}_i^\top \mathbf{x})$, where $f^{''}(x) \geq 0$ since the convexity of $f$. Then the parameter of strong convexity is given by the minimal eigenvalue of $H$. We first introduce a standard result for DC based algorithm:

THEOREM 3.1. *Assume $F_2$ is a strongly convex function with parameter $\ell$. Starting from any initial guess $\mathbf{x}^{(1)}$, let $\mathbf{x}^{(k+1)}$ be the solution in (3.2) for all $k \geq 1$. Then*

$$F(\mathbf{x}^{(k+1)}) \leq F(\mathbf{x}^{(k)}) - \frac{\ell}{2}\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|^2, \quad \forall k \geq 1 \tag{3.3}$$

*and* $\nabla F_1(\mathbf{x}^{(k+1)}) - \nabla F_2(\mathbf{x}^{(k)}) = 0$.

*Proof.* By the strongly convexity of $F_2$, we have

$$F_2(\mathbf{x}^{(k+1)}) \geq F_2(\mathbf{x}^{(k)}) + \nabla F_2(\mathbf{x}^{(k)})^\top (\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}) + \frac{\ell}{2}\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|^2.$$

From (3.2), we see that

$$\begin{aligned} F(\mathbf{x}^{(k+1)}) &= F_1(\mathbf{x}^{(k+1)}) - F_2(\mathbf{x}^{(k+1)}) \\ &\leq F_1(\mathbf{x}^{(k+1)}) - \nabla F_2(\mathbf{x}^{(k)})^\top (\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}) - F_2(\mathbf{x}^{(k)}) - \frac{\ell}{2}\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|^2 \\ &\leq F_1(\mathbf{x}^{(k)}) - F_2(\mathbf{x}^{(k)}) - \frac{\ell}{2}\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|^2 = F(\mathbf{x}^{(k)}) - \frac{\ell}{2}\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|^2. \end{aligned}$$

The second property $\nabla F_1(\mathbf{x}^{(k+1)}) - \nabla F_2(\mathbf{x}^{(k)}) = 0$ follows from the minimization (3.2) directly. ∎

Next, we use the Kurdyka-Łojasiewicz (KL) inequality to establish the convergence rate of $\mathbf{x}^{(k)}$. We refer to [1], [2], and [44] for using the KL inequality for various minimization problems. The following is our major theorem in this section.

THEOREM 3.2. *Suppose that $F(\mathbf{x}) = F_1(\mathbf{x}) - F_2(\mathbf{x})$ is a real analytic function. Assume the gradient function $\nabla F_1$ has Lipschitz constant $L_1 > 0$ and $F_2$ is a strongly convex function with parameter $\ell > 0$ in bounded region $\mathcal{R}$. Starting from any initial guess $\mathbf{x}^{(1)}$, let $\mathbf{x}^{(k+1)}$ be the solution in (3.2) for all $k \geq 1$. Then $\mathbf{x}^{(k)}, k \geq 1$ converges to a critical point of $F$. Furthermore, if we let $\mathbf{x}^*$ be the unique limit, then*

$$\|\mathbf{x}^{(k+1)} - \mathbf{x}^*\| \leq C\tau^k \tag{3.4}$$

*for a positive constant $C$ and $\tau \in (0,1)$.*

To this end, we need the KL inequality which is central to the global convergence analysis.

Definition 3.3 (Łojasiewicz [29]). *We say a function $f(\mathbf{x})$ satisfies the Kurdyka-Lojasiewicz (KL) property at point $\bar{\mathbf{x}}$ if there exists $\theta \in [0,1)$ such that*

$$|f(\mathbf{x}) - f(\bar{\mathbf{x}})|^{\theta} \leq C\|\partial f(\mathbf{x})\|$$

*in a neighborhood $B(\bar{\mathbf{x}}, \delta)$ for some $\delta > 0$, where $C > 0$ is a constant independent of $\mathbf{x}$. In other words, there exists $\varphi(s) = cs^{1-\theta}$ with $\theta \in [0,1)$ such that the KL inequality holds:*

$$\varphi'(|f(\mathbf{x}) - f(\bar{\mathbf{x}})|)\|\partial f(\mathbf{x})\| \geq 1 \tag{3.5}$$

*for any $\mathbf{x} \in B(\bar{\mathbf{x}}, \delta)$ with $f(\mathbf{x}) \neq f(\bar{\mathbf{x}})$.*

This property is introduced by Lojasiewicz on the real analytic functions, for which (3.5) holds in any critical point with $\theta \in [1/2, 1)$. Later, many extensions of the above inequality are proposed. Typically, the extension in [28] for the setting of $o$-minimal structure. Recently, the KL inequality is extended to nonsmooth subanalytic functions. See [1, 2, 44], for application of the KL inequality for study in the optimization. In our setting, $\theta = 1/2$. We shall include an elementary proof to justify our choice of $\theta = 1/2$.

Proposition 3.4. *Suppose that $f : \mathbb{R}^n \mapsto \mathbb{R}$ is a continuously twice differentiable function whose Hessian $H(f)(\mathbf{x})$ is invertible at a critical point $\mathbf{x}^*$ of $f$. Then there exists a positive constant $C$, an exponent $\theta = 1/2$ and a positive $r$ such that*

$$|f(\mathbf{x}) - f(\mathbf{x}^*)|^{1/2} \leq C\|\nabla f(\mathbf{x})\|, \quad \forall \mathbf{x} \in B(\mathbf{x}^*, r), \tag{3.6}$$

*where $B(\mathbf{x}^*, r)$ is a ball at $\mathbf{x}^*$ with radius $r$.*

*Proof.* Since $f$ is continuously twice differentiable, using Taylor formula for f and noting $f(\mathbf{x}^*) = 0$, we have

$$|f(\mathbf{x}) - f(\mathbf{x}^*)| \leq c_1\|\mathbf{x} - \mathbf{x}^*\|^2, \quad \forall \mathbf{x} \in B(\mathbf{x}^*, r)$$

for some $r > 0$. On the other hand, we have $\|\nabla f(\mathbf{x})\| = \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{x}^*)\| \geq c_2\|\mathbf{x} - \mathbf{x}^*\|$ due to the fact the Hessian is invertible. Thus, (3.6) follows with $\theta = 1/2$ and $C = \sqrt{c_1}/c_2$. ∎

The importance of the Łajosiewicz inequality is the establishment of the above inequality in (3.6) when $f$ may not have an invertible Hessian at the critical point $\mathbf{x}^*$. The proof is based on knowledge from algebraic geometry, mainly the curve selecting lemma. See [28] for a more general setting.

Let us recall the geometric description of the landscape function $F(\mathbf{x})$ whose Hessian is restricted strong convex at the global minimizer (cf. [40]). In the real variable setting, we can even show that the Hessian is positive definite at a global minimizer. See the Appendix for a proof. We are now ready to establish Theorem 3.2.

*Proof of Theorem 3.2.* From Theorem 3.1, we have

$$\frac{\ell}{2}\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|^2 \leq F(\mathbf{x}^{(k)}) - F(\mathbf{x}^{(k+1)}). \tag{3.7}$$

That is, $F(\mathbf{x}^{(k)}), k \geq 1$ is strictly decreasing sequence. Without loss of generality, we assume

$$\mathcal{R} := \{\mathbf{x} \in \mathbb{R}^n, F(\mathbf{x}) \leq F(\mathbf{x}^{(1)})\}.$$

Then the sequence $\{\mathbf{x}^{(k)}\}_{k=1}^{\infty} \subset \mathcal{R}$ is a bounded sequence. Then there exists a cluster point $\mathbf{x}^*$ and a subsequence $\mathbf{x}^{(k_i)}$ such that $\mathbf{x}^{(k_i)} \to \mathbf{x}^*$. Note that $\{F(\mathbf{x}^{(k)})\}_{k=1}^{\infty}$ is a bounded monotonic descending sequence. Then $F(\mathbf{x}^{(k)}) \to F(\mathbf{x}^*)$ for all $k \geq 1$. We claim that there exists a a positive constant $C_1$ such that

$$C_1\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\| \leq \sqrt{F(\mathbf{x}^{(k)}) - F(\mathbf{x}^*)} - \sqrt{F(\mathbf{x}^{(k+1)}) - F(\mathbf{x}^*)} \tag{3.8}$$

holds for all $k \geq k_0$ where $k_0$ is large enough. To establish this claim, we need to use Proposition 3.4 which is the well-known Kurdyka-Lojasiewicz inequality. First, we prove that the condition $\|\nabla F(\mathbf{x}^*)\| = 0$ holds. Indeed, using one of the properties in Theorem 3.1, we have

$$\|\nabla F(\mathbf{x}^{(k)})\| = \|\nabla F_1(\mathbf{x}^{(k)}) - \nabla F_2(\mathbf{x}^{(k)})\| = \|\nabla F_1(\mathbf{x}^{(k)}) - \nabla F_1(\mathbf{x}^{(k+1)})\| \leq L_1\|\mathbf{x}^{(k)} - \mathbf{x}^{(k+1)}\|.$$

Combining with (3.7), it gives that $\|\nabla F(\mathbf{x}^{(k_i)})\| \to 0$. By the continuity of gradient function, we have $\|\nabla F(\mathbf{x}^*)\| = 0$ since $\mathbf{x}^{(k_i)} \to \mathbf{x}^*$. Next, consider $g(t) = \sqrt{t}$ which is concave over $[0, 1]$, we have $g(t) - g(s) \geq g'(t)(t - s)$. Then by the Kurdyka-Lojasiewicz inequality, there exists a positive constant $c_0 > 0$ and $\delta > 0$ such that

$$\|g'(F(\mathbf{x}) - F(\mathbf{x}^*))\nabla F(\mathbf{x}))\| \geq c_0 > 0 \tag{3.9}$$

for all $\mathbf{x}$ in the neighborhood $B(\mathbf{x}^*, \delta)$ of $\mathbf{x}^*$. As

$$F(\mathbf{x}^{(k)}) - F(\mathbf{x}^*) \to 0, \qquad k \to \infty,$$

then there is an integer $k_0$ such that for all $k \geq k_0$ it holds

$$\max\left(\sqrt{2/\ell}, L_1/(\ell c_0)\right) \cdot \sqrt{F(\mathbf{x}^{(k)}) - F(\mathbf{x}^*)} \leq \delta/2. \tag{3.10}$$

Also, $\mathbf{x}^{(k_i)} \to \mathbf{x}^*$ as $k_i \to \infty$. Without loss of generality, we may assume that $k_0 = 1$ and $\mathbf{x}^{(1)} \in B(\mathbf{x}^*, \delta/2)$. Let us show that $\mathbf{x}^{(k)}, k \geq 1$ will be in the neighborhood $B(\mathbf{x}^*, \delta)$. We shall use an induction to do so. By (3.10) we have

$$\|\mathbf{x}^{(2)} - \mathbf{x}^*\| \leq \|\mathbf{x}^{(2)} - \mathbf{x}^{(1)}\| + \|\mathbf{x}^{(1)} - \mathbf{x}^*\| \leq \sqrt{2(F(\mathbf{x}^{(1)}) - F(\mathbf{x}^*)/\ell} + \|\mathbf{x}^{(1)} - \mathbf{x}^*\| \leq \delta.$$

Assume that $\mathbf{x}^{(k)} \in B(\mathbf{x}^*, \delta)$ for $k \leq K$. Multiplying $g'(F(\mathbf{x}^{(k)}) - F(\mathbf{x}^*))$ to both sides of (3.7), we have

$$\begin{aligned}\frac{\ell}{2}\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|^2 g'(F(\mathbf{x}^{(k)}) - F(\mathbf{x}^*)) &\leq g'(F(\mathbf{x}^{(k)}) - F(\mathbf{x}^*))\left(F(\mathbf{x}^{(k)}) - F(\mathbf{x}^{(k+1)})\right) \\ &\leq \sqrt{F(\mathbf{x}^{(k)}) - F(\mathbf{x}^*)} - \sqrt{F(\mathbf{x}^{(k+1)}) - F(\mathbf{x}^*)}\end{aligned} \tag{3.11}$$

by using the concavity of $g$. However, the K-L inequality (3.9) and Theorem 3.1 gives that

$$\begin{aligned}|g'(F(\mathbf{x}^{(k)}) - F(\mathbf{x}^*))| &\geq \frac{c_0}{\|\nabla F(\mathbf{x}^{(k)})\|} = \frac{c_0}{\|\nabla F_1(\mathbf{x}^{(k)}) - \nabla F_2(\mathbf{x}^{(k)})\|} \\ &= \frac{c_0}{\|\nabla F_1(\mathbf{x}^{(k)}) - \nabla F_1(\mathbf{x}^{(k+1)})\|} \geq \frac{c_0}{L_1\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|}.\end{aligned} \tag{3.12}$$

Putting it in (3.11) gives that

$$\sqrt{F(\mathbf{x}^{(k)}) - F(\mathbf{x}^*)} - \sqrt{F(\mathbf{x}^{(k+1)}) - F(\mathbf{x}^*)} \geq \frac{\ell c_0}{2L_1}\|\mathbf{x}^{(k)} - \mathbf{x}^{(k+1)}\| \tag{3.13}$$

holds for all $2 \leq k \leq K$. It follows that

$$\frac{2L_1}{\ell c_0}\sqrt{F(\mathbf{x}^{(1)}) - F(\mathbf{x}^*)} \geq \sum_{j=1}^{K} \|\mathbf{x}^{(j+1)} - \mathbf{x}^{(j)}\|.$$

That is, we have

$$\begin{aligned}
\|\mathbf{x}^{(K+1)} - \mathbf{x}^*\| &\leq \|\mathbf{x}^{(K+1)} - \mathbf{x}^{(1)}\| + \|\mathbf{x}^{(1)} - \mathbf{x}^*\| \\
&\leq \sum_{j=1}^{K} \|\mathbf{x}^{(j+1)} - \mathbf{x}^{(j)}\| + \|\mathbf{x}^{(1)} - \mathbf{x}^*\| \\
&\leq \frac{2L_1}{\ell c_0}\sqrt{F(\mathbf{x}^{(1)}) - F(\mathbf{x}^*)} + \|\mathbf{x}^{(1)} - \mathbf{x}^*\| \leq \delta,
\end{aligned}$$

where the last inequality follows from (3.10). Thus, $\mathbf{x}^{(K+1)} \in B(\mathbf{x}^*, \delta)$. This shows that all $\mathbf{x}^{(k)}$ are in $B(\mathbf{x}^*, \delta)$ and inequality (3.13) holds for all $k$. Hence, we arrive at the claim (3.8) with $C_1 = \ell c_0/(2L_1)$. By summing the inequality in (3.8) above, it follows

$$\sum_{k \geq 1} \|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\| \leq \frac{1}{C_1}\sqrt{F(\mathbf{x}^{(1)}) - F(\mathbf{x}^*)}.$$

That is, $\mathbf{x}^{(k)}$ is a Cauchy sequence and hence, it is convergent with $\mathbf{x}^{(k)} \to \mathbf{x}^\star$. Note that $\nabla F(\mathbf{x}^\star) = 0$, which implies $\mathbf{x}^{(k)}$ converges to a critical point of $F$.

Next, we turn to prove the second part. Let $S_k = \sum_{i=k}^{\infty} \|\mathbf{x}^{(i+1)} - \mathbf{x}^{(i)}\|$. It follows from (3.13) that

$$\begin{aligned}
C_1 S_k &= \sum_{i=k}^{\infty} C_1\|\mathbf{x}^{(i+1)} - \mathbf{x}^{(i)}\| \\
&\leq \sum_{i=k}^{\infty}\left(\sqrt{F(\mathbf{x}^{(i)}) - F(\mathbf{x}^*)} - \sqrt{F(\mathbf{x}^{(i+1)}) - F(\mathbf{x}^*)}\right) \leq \sqrt{F(\mathbf{x}^{(k)}) - F(\mathbf{x}^*)}.
\end{aligned}$$

Recall from (3.12) that

$$\sqrt{F(\mathbf{x}^{(k)}) - F(\mathbf{x}^*)} \leq \frac{L_1}{2c_0}\|\mathbf{x}^{(k)} - \mathbf{x}^{(k+1)}\| = C_2(S_k - S_{k+1})$$

where $C_2 = L_1/(2c_0)$. Combining the two above inequality that

$$S_{k+1} \leq \frac{C_2 - C_1}{C_2} S_k \leq \cdots \leq \theta^k S_0$$

for $\tau = (C_2 - C_1)/(C_2)$. Since $\|\mathbf{x}^{(k)} - \mathbf{x}^*\| \leq S_k$, we complete the proof. □

*Remark* 3.5. We should point out that the conditions on $F, F_1$ and $F_2$ in Theorem 3.2 are not harsh. Notice that for standard phase retrieval, all these conditions are satisfied, especially when the region $\mathcal{R}$ is sufficiently small and near the global minimization by a technical initialization.

In summary, two obvious consequences are:

(1) For any given initial point $\mathbf{x}^{(1)}$, let $D = F(\mathbf{x}^{(1)}) - F(\mathbf{x}^\star) > 0$, where $\mathbf{x}^\star$ is one of the global minimizer of (3.1). Then

$$F(\mathbf{x}^{(k)}) - F(\mathbf{x}^\star) \le D - \frac{\ell}{2}\sum_{j=1}^{k-1} \|\mathbf{x}^{(j+1)} - \mathbf{x}^{(j)}\|^2.$$

That is, $\mathbf{x}^{(k)}$ is closer to one of global minimizer than the initial guess point.

(2) As our approach can find a critical point, if a global minimizer $\mathbf{x}^\star$ is a local minimizer over a neighborhood $N(\mathbf{x}^\star)$ and an initial vector $\mathbf{x}^{(1)}$ is in $N(\mathbf{x}^\star)$, then our approach finds $\mathbf{x}^\star$.

*Example* 3.6. In this example, we consider the standard phase retrieval problem where $f(x) = |x|^2$. Assume the measurements are Gaussian random vectors, it has been showed that one can use the initialization from [13, 22] to find an excellent initial vector. More specifically, to recover a vector $x \in \mathbb{R}^n$ (or $x \in \mathbb{C}^n$), if the number of measurement vectors $\mathbf{a}_i, i = 1, \ldots, m$ is $m = O(n)$, then with high probability we have

$$\|\mathbf{x}^{(1)} - \mathbf{x}^*\|_2 \le \delta \|\mathbf{x}^*\|_2,$$

where $\mathbf{x}^*$ is a global minimizer and $\delta$ is a positive constant. Furthermore, in a small neighborhood $N(\mathbf{x}^\star, \delta) := \{x : \|\mathbf{x} - \mathbf{x}^*\|_2 \le \delta\|\mathbf{x}^*\|_2\}$, the minimizing functional $F(x)$ is strongly convex. Thus, our algorithm can converge to the global minimizer by using a initialization.

**4. Computation of the Inner Minimization (3.2).** We now discuss how to compute the minimization in (3.2). For convenience, we rewrite the minimization in the following form

$$\min_{\mathbf{x}\in\mathbb{R}^n} G(\mathbf{x}) \tag{4.1}$$

for a differentiable convex function $G(\mathbf{x}) := F_1(\mathbf{x}) - \langle \nabla F_2(\mathbf{x}^{(k)}), \mathbf{x} - \mathbf{x}^{(k)}\rangle$. The first approach is to use the gradient descent method:

$$\mathbf{z}^{(j+1)} = \mathbf{z}^{(j)} - h\nabla G(\mathbf{z}^{(j)}) \tag{4.2}$$

for $j \in \mathbb{N}$ with $\mathbf{z}^{(1)} = \mathbf{x}^{(k)}$, where $h > 0$ is a fixed step size or variable step size. It is well-known that we need to choose $h \approx 1/(2L)$ for the Lipschitz differentiability constant $L$ of $G(\mathbf{x})$ and then the gradient descent method (4.2) will have a linear convergence. It is also known that we can choose $h = \nu/L$ for the Lipschitz differentiability constant $L$ of $G(\mathbf{x})$ and the $\nu$-strong convexity of $G$ and then use the Nestrov acceleration technique as explained in [33]. The convergence will be sped up. See the following result.

Lemma 4.1 (The Nesterov's Acceleration ([33])). *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a $\nu$-strong convex function and the gradient function has L-Lipschitz constant. Start at an arbitrary initial point $\mathbf{u}_1 = \mathbf{z}_1$, the following Nesterov's accelerated gradient descent*

$$\begin{aligned}\mathbf{z}^{j+1} :&= \mathbf{u}^{(j)} - \frac{\nu}{L}\nabla f(\mathbf{u}^{(j)}),\\ \mathbf{u}^{(j+1)} &= \mathbf{z}^{(j+1)} - q(\mathbf{z}^{(j+1)} - \mathbf{z}^{(j)})\end{aligned} \tag{4.3}$$

*satisfies*

$$f(\mathbf{z}^{j+1}) - f(\mathbf{z}^*) \le \frac{\nu + L}{2}\|\mathbf{z}^{(1)} - \mathbf{z}^*\|^2 \exp(-\frac{j}{\sqrt{L/\nu}}), \tag{4.4}$$

*where* $\mathbf{z}^*$ *is the optimal solution and* $q = (\sqrt{L/\nu} - 1)/(\sqrt{L/\nu} + 1)$ *is a constant.*

The significance of the Nestrov acceleration above is to reduce the number of iterations in (4.2) significantly. That is, for any tolerance $\epsilon$, we need $O(1/\epsilon)$ number of iterations for the gradient descent method due to the linear convergence, but $O(1/\sqrt{\epsilon})$ number of iterations if Nesterov's acceleration (4.3) is used.

Since $G$ is twice differentiable, we can certainly use the Newton method to solve (3.2) because it has quadratic convergence. However, we will not pursue it here due to the fact that when the number of variables of $\mathbf{z}$ is large, the Newton method will be extremely slow. Instead, another method to choose a good $h$ is to use the Barzilai-Borwein(BB) method which is an excellent approach for a large scale minimization problem (cf. [5]). The iteration of the BB method can be described as

$$\mathbf{z}^{(j+1)} = \mathbf{z}^{(j)} - \beta_j^{-1}\nabla G(\mathbf{z}^{(j)}), \tag{4.5}$$

where the step size

$$\beta_j = (\mathbf{z}^{(j)} - \mathbf{z}^{(j-1)})^\top(\nabla G(\mathbf{z}^{(j)}) - \nabla G(\mathbf{z}^{(j-1)}))/\|\mathbf{z}^{(j)} - \mathbf{z}^{(j-1)}\|^2. \tag{4.6}$$

We shall use the following Algorithm 4.1 to solve the minimization (3.2).

**Algorithm 4.1** The BB Algorithm for the Inner Minimization

Let $\mathbf{u}^{(1)} = \mathbf{z}^{(1)}$ be an initial guess.
For $j \ge 1$, we solve the minimization in (4.1) by computing $\beta_j$ according to (4.6).
Update

$$\begin{aligned}\mathbf{z}^{(j+1)} &:= \mathbf{u}^{(j)} - \beta_j^{-1}\nabla G(\mathbf{u}^{(j)})\\ \mathbf{u}^{(j+1)} &= \mathbf{z}^{(j+1)} - q(\mathbf{z}^{(j+1)} - \mathbf{z}^{(j)})\end{aligned} \tag{4.7}$$

until a maximum number $T$ of iteration is achieved.
**return** $\mathbf{u}^T$

Our computation of inner minimization is described in Algorithm 4.1, which is the combination of BB method with Nesterov's accelerated gradient descent. The intuition behind it based on the results in Lemma 4.1. Since BB method has a good performance in numerical experiment, we can hope our Algorithm 4.1 has better performance.

There are several modified versions of the BB-method available together their convergence analysis in the literature. See, e.g. [18, 47] and the references therein. A quick literature search shows that the convergence rate is still not established yet for general minimizing functional $F$ to the best of the authors knowledge. Next we give a necessary and sufficient condition for the algorithm (4.5) has a better convergence than linear rate. We say a algorithm is convergent superlinearly if

$$\sigma_k = \frac{\|\mathbf{x}^{(k+1)} - \mathbf{x}^*\|}{\|\mathbf{x}^{(k)} - \mathbf{x}^*\|} \to 0, \text{ when } k \to \infty.$$

To analyze the convergence of the BB method in our setting, let $\mathbf{s}_{k+1} = \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}$ and $\mathbf{y}_{k+1} = \nabla G(\mathbf{x}^{(k+1)}) - \nabla G(\mathbf{x}^{(k)})$.

LEMMA 4.2. *Suppose that the function $G(\mathbf{x})$ in (4.1) is $\alpha$-strongly convex and the gradient function has Lipschitz constant $L$ in a domain $D$. Assume $\mathbf{x}^* \in D$ and the sequence $\{\mathbf{x}^{(k)}, k \geq 1\}$ obtained from the BB method above remain in $D$. Then $\{\mathbf{x}^{(k)}, k \geq 1\}$ converges super linearly to $\mathbf{x}^*$ if and only if $(\beta_k - H_G(\mathbf{x}^*))\mathbf{s}_{k+1} = o(\|\mathbf{s}_{k+1}\|)$.*

*Proof.* From iteration (4.5), We have

$$\begin{aligned}(\beta_k - H_G(\mathbf{x}^*))\,\mathbf{s}_{k+1} &= -\nabla G(\mathbf{x}^{(k)}) - H_G(\mathbf{x}^*)\mathbf{s}_{k+1} \\ &= \nabla G(\mathbf{x}^{(k+1)}) - \nabla G(\mathbf{x}^{(k)}) - H_G(\mathbf{x}^*)\mathbf{s}_{k+1} - \nabla G(\mathbf{x}^{(k+1)}).\end{aligned} \tag{4.8}$$

Since the Hessian $H_G(\mathbf{x})$ is continuous at $\mathbf{x}^*$ and all $\mathbf{x}^{(k)} \in D$, we see

$$\nabla G(\mathbf{x}^{(k+1)}) - \nabla G(\mathbf{x}^{(k)}) - H_G(\mathbf{x}^*)\mathbf{s}_{k+1} \to 0, \quad k \to \infty.$$

By the assumption that $(\beta_k - H_G(\mathbf{x}^*))\mathbf{s}_{k+1} = o(\|\mathbf{s}_{k+1}\|)$, we have

$$\lim_{k\to\infty} \frac{||\nabla G(\mathbf{x}^{(k+1)})||}{||\mathbf{s}_{k+1}||} = 0. \tag{4.9}$$

Note that

$$\|\nabla G(\mathbf{x}^{(k+1)}) - G(\mathbf{x}^{(k)})\| \leq L\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|$$

and

$$||\nabla G(\mathbf{x}^{(k+1)})|| = ||\nabla G(\mathbf{x}^{(k+1)}) - \nabla G(\mathbf{x}^*)|| = ||H_G(\xi_k)(\mathbf{x}^{(k+1)} - \mathbf{x}^*)|| \geq \alpha\|\mathbf{x}^{(k+1)} - \mathbf{x}^*||$$

for $\mathbf{x}^{(k+1)} \in D$, where $\xi_k$ in $D$. Then, we have

$$\frac{||\nabla G(\mathbf{x}^{(k+1)})||}{||\mathbf{y}_{k+1}||} \geq \frac{\alpha||\mathbf{x}^{(k+1)} - \mathbf{x}^*||}{L||\mathbf{x}^{(k+1)} - \mathbf{x}^*|| + L||\mathbf{x}^{(k)} - \mathbf{x}^*||} = \frac{\alpha\sigma_k}{L(1+\sigma_k)},$$

where $\sigma_k = \dfrac{||\mathbf{x}^{(k+1)} - \mathbf{x}^*||}{||\mathbf{x}^{(k)} - \mathbf{x}^*||}$. It follows that $\dfrac{\sigma_k}{1+\sigma_k} \to 0$ and hence, $\sigma_k \to 0$. That is, the BB method converges super-linearly.

On the other hand, if $\sigma_k \to 0$, we can show that $(\beta_k - H_G(\mathbf{x}^*))\mathbf{s}_{k+1} = o(\|\mathbf{s}_{k+1}\|)$. In fact, it is known that when $\mathbf{x}^{(k)} \to \mathbf{x}^*$ super-linearly, then

$$\lim_{k\to+\infty} \frac{\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}||}{||\mathbf{x}^{(k)} - \mathbf{x}^*||} = 1. \tag{4.10}$$

Indeed, we have

$$\Big| ||\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}|| - ||\mathbf{x}^{(k)} - \mathbf{x}^*|| \Big| \leq ||\mathbf{x}^{(k+1)} - \mathbf{x}^*||.$$

It follows that

$$\left| \frac{||\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}||}{||\mathbf{x}^{(k)} - \mathbf{x}^*||} - 1 \right| \leq \frac{||\mathbf{x}^{(k+1)} - \mathbf{x}^*||}{||\mathbf{x}^{(k)} - \mathbf{x}^*||} \to 0.$$

Hence, we have

$$\begin{aligned}\frac{||\nabla G(\mathbf{x}^{(k+1)})||}{||\mathbf{s}_{k+1}||} &\leq \frac{||\nabla G(\mathbf{x}^{(k+1)}) - \nabla G(\mathbf{x}^*)||}{||\mathbf{s}_{k+1}||} \\ &\leq \frac{L\|\mathbf{x}^{(k+1)} - \mathbf{x}^*\|}{\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|} \\ &= \frac{\sigma_{k+1}}{\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|/\|\mathbf{x}^{(k)} - \mathbf{x}^*\|} \to 0\end{aligned}$$

because of the denominator is bounded by the property (4.10). Using the argument at the beginning of the proof, we can see that $(\beta_k - H_G(\mathbf{x}^*))\mathbf{s}_{k+1} = o(\|\mathbf{s}_{k+1}\|)$. These completes the proof. ∎

**5. Sparse Phase Retrieval.** In previous sections, several computational algorithms have been developed for the phase retrieval problem based on measurements in (1.1). We now extend the approaches to study the sparse phase retrieval. Suppose that $\mathbf{x_b}$ is a sparse solution to the given measurements (1.1). Let us use the DC based algorithm to explain how to do. First, we consider the following

$$\min_{\mathbf{x}\in\mathbb{R}^n \text{ or } \mathbb{C}^n} \lambda\|\mathbf{x}\|_1 + \sum_{i=1}^m (f(\langle \mathbf{a}_i, \mathbf{x}\rangle) - b_i)^2 \tag{5.1}$$

by adding $\lambda\|\mathbf{x}\|_1$ to (1.2) as a standard approach in compressive sensing. If we take $f(\langle \mathbf{a}_i, \mathbf{x}\rangle) = |\langle \mathbf{a}_i, \mathbf{x}\rangle|^2$, then (5.1) reduces to the sparse phase retrieval.

We now discuss how to solve it numerically. We approach it by using a similar method as in the previous section. Indeed, for the case $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{a}_i \in \mathbb{R}^n$, we rewrite $F(\mathbf{x}) = \sum_{i=1}^m (f(\langle \mathbf{a}_i, \mathbf{x}\rangle) - b_i)^2$ as

$$F(\mathbf{x}) = F_1(\mathbf{x}) - F_2(\mathbf{x}) := \sum_{i=1}^m f^2(\langle \mathbf{a}_i, \mathbf{x}\rangle) + b_i^2 - \sum_{i=1}^m 2b_i f(\langle \mathbf{a}_i, \mathbf{x}\rangle).$$

The minimization (5.1) will be approximated by

$$\mathbf{x}^{(k+1)} := \arg\min \lambda\|\mathbf{x}\|_1 + F_1(\mathbf{x}) - \nabla F_2(\mathbf{x}^{(k)})^\top(\mathbf{x} - \mathbf{x}^{(k)}) \tag{5.2}$$

for any given $\mathbf{x}^{(k)}$. We call this algorithm as sparse DC based method. When $\mathbf{x} \in \mathbb{C}^n$ and $\mathbf{a}_j \in \mathbb{C}^n$, $j = 1, \ldots, m$, we have to write $\mathbf{x} = \mathbf{x}_R + \sqrt{-1}\mathbf{x}_I$ and similar for $\mathbf{a}_j$. Letting $\mathbf{c} = [\mathbf{x}_R^\top \mathbf{x}_I^\top]^\top \in \mathbb{R}^{2n}$, we view $F_1(\mathbf{x})$ as a functions in $G_1(\mathbf{c}) = F_1(\mathbf{x}_R + \sqrt{-1}\mathbf{x}_I)$. Then $G_1(\mathbf{c})$ is a convex function of variable $\mathbf{c}$. Similarly, $G_2(\mathbf{c}) = F_2(\mathbf{x}_R + \sqrt{-1}\mathbf{x}_I)$ is a convex function of $\mathbf{c}$. We can formulate the same minimization problem as in (5.2). For convenience, we simply discuss the case when $\mathbf{x}$, $\mathbf{a}_j$, $j = 1, \ldots, m$ are real. The complex variable setting can be treated in the same fashion.

To solve (5.2), we use a proximal gradient method: for any given $\mathbf{y}^{(k)}$,

$$\mathbf{y}^{(k+1)} := \arg\min \lambda\|\mathbf{y}\|_1 + F_1(\mathbf{y}^{(k)}) + (\nabla F_1(\mathbf{y}^{(k)}) - \nabla F_2(\mathbf{y}^{(k)}))^\top(\mathbf{y} - \mathbf{y}^{(k)}) + \frac{L_1}{2}\|\mathbf{y} - \mathbf{y}^{(k)}\|^2 \tag{5.3}$$

for $k \geq 1$, where $L_1$ is the Lipschitz differentiability of $F_1$. The above minimization can be easily solved by using shrinkage-thresholding technique as in [7]. Note that Beck and Teboulle in [7] use a Nesterov's acceleration technique to speed up the iteration to form the well-known FISTA. However, we shall use the acceleration technique from

[3] which is slightly better than the Nestrov technique. The discussion above furnishes a computational method for sparse phase retrieval problem (5.1). Let us point out one significant difference from (5.3) is that one can find $\mathbf{y}^{(k+1)}$ by using a formula while the solution $\mathbf{x}^{(k+1)}$ in (3.2) has to be computed using an iterative method as explained before. Thus the sparse phase retrieval is more efficient in this sense.

Let us study the convergence of our sparse phase retrieval method. We again start with a standard result for a DC based algorithm:

THEOREM 5.1. *Assume $F_2$ is a strongly convex function with parameter $\ell$. Starting from any initial guess $\mathbf{y}^{(1)}$, let $\mathbf{y}^{(k+1)}$ be the solution in (5.3) for all $k \geq 1$. Then*

$$\lambda\|\mathbf{y}^{(k+1)}\|_1 + F(\mathbf{y}^{(k+1)}) \leq \lambda\|\mathbf{y}^{(k)}\|_1 + F(\mathbf{y}^{(k)}) - \frac{\ell}{2}\|\mathbf{y}^{(k+1)} - \mathbf{y}^{(k)}\|^2, \quad \forall k \geq 1 \tag{5.4}$$

*and* $\partial g(\mathbf{y}^{(k+1)}) + \nabla F_1(\mathbf{y}^{(k)}) - \nabla F_2(\mathbf{y}^{(k)}) + \dfrac{L_1}{2}(\mathbf{y}^{(k+1)} - \mathbf{y}^{(k)}) = 0$, *where* $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_1$ *and $\partial g$ stands for the subgradient of $g$.*

*Proof.* The Lipschitz differentiability of $F_1$ tells us

$$F_1(\mathbf{y}^{(k+1)}) \leq F_1(\mathbf{y}^{(k)}) + \nabla F_2(\mathbf{y}^{(k)})^\top(\mathbf{y}^{(k+1)} - \mathbf{y}^{(k)}) + \frac{L_1}{2}\|\mathbf{y}^{(k+1)} - \mathbf{y}^{(k)}\|^2,$$

where $L_1$ is the Lipschitz differentiability of $F_1$. By the strongly convexity of $F_2$, we have

$$F_2(\mathbf{y}^{(k+1)}) \geq F_2(\mathbf{y}^{(k)}) + \nabla F_2(\mathbf{y}^{(k)})^\top(\mathbf{y}^{(k+1)} - \mathbf{y}^{(k)}) + \frac{\ell}{2}\|\mathbf{y}^{(k+1)} - \mathbf{y}^{(k)}\|^2.$$

With the above two inequalities, we see that

$$\begin{aligned}
&\lambda\|\mathbf{y}^{(k+1)}\|_1 + F(\mathbf{x}^{(k+1)}) = \lambda\|\mathbf{y}^{(k+1)}\|_1 + F_1(\mathbf{y}^{(k+1)}) - F_2(\mathbf{y}^{(k+1)})\\
\leq\ &\lambda\|\mathbf{y}^{(k+1)}\|_1 + F_1(\mathbf{y}^{(k)}) + \nabla F_1(\mathbf{y}^{(k)})^\top(\mathbf{y}^{(k+1)} - \mathbf{y}^{(k)}) + \frac{L}{2}\|\mathbf{y}^{(k+1)} - \mathbf{y}^{(k)}\|^2\\
&\quad -F_2(\mathbf{y}^{(k)}) - \nabla F_2(\mathbf{y}^{(k)})^\top(\mathbf{y}^{(k+1)} - \mathbf{y}^{(k)}) - \frac{\ell}{2}\|\mathbf{y}^{(k+1)} - \mathbf{y}^{(k)}\|^2\\
=\ &F_1(\mathbf{y}^{(k)}) - F_2(\mathbf{y}^{(k)}) - \frac{\ell}{2}\|\mathbf{y}^{(k+1)} - \mathbf{y}^{(k)}\|^2\\
&\quad +\lambda\|\mathbf{y}^{(k+1)}\|_1 + (\nabla F_1(\mathbf{y}^{(k)}) - \nabla F_2(\mathbf{y}^{(k)})^\top(\mathbf{y}^{(k+1)} - \mathbf{y}^{(k)}) + \frac{L}{2}\|\mathbf{y}^{(k+1)} - \mathbf{y}^{(k)}\|^2\\
\leq\ &F_1(\mathbf{y}^{(k)}) - F_2(\mathbf{y}^{(k)}) - \frac{\ell}{2}\|\mathbf{y}^{(k+1)} - \mathbf{y}^{(k)}\|^2 + \lambda\|\mathbf{y}^{(k)}\|_1\\
=\ &\lambda\|\mathbf{y}^{(k)}\|_1 + F(\mathbf{y}^{(k)}) - \frac{\ell}{2}\|\mathbf{y}^{(k+1)} - \mathbf{y}^{(k)}\|^2,
\end{aligned}$$

where we have used the optimization condition in (5.3). Letting $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_1$, the second property $\partial g(\mathbf{y}^{(k+1)}) + \nabla F_1(\mathbf{y}^{(k)}) - \nabla F_2(\mathbf{y}^{(k)}) + \dfrac{L_1}{2}(\mathbf{y}^{(k+1)} - \mathbf{y}^{(k)}) = 0$ follows from the minimization (5.3). ∎

Next we show that the sequence $\mathbf{y}^{(k)}, k \geq 1$ from (5.3) converges to a critical point $\mathbf{y}^*$.

THEOREM 5.2. *Suppose that $f(\mathbf{x})$ is a real analytic function and the gradient function $\nabla f(\mathbf{x})$ has Lipschitz constant $L$. Let $\mathbf{y}^{(k)}, k \geq 1$ be the sequence obtained from (5.3). Then it converges to a critical point $\mathbf{y}^*$ of $F$.*

*Proof.* Recall $F(\mathbf{x}) = g(\mathbf{x}) + f(\mathbf{x})$. From Theorem 5.1, we have

$$\frac{\ell}{2}\|\mathbf{y}^{(k+1)} - \mathbf{y}^{(k)}\|^2 \le F(\mathbf{y}^{(k)}) - F(\mathbf{y}^{(k+1)}). \tag{5.5}$$

That is, $F(\mathbf{y}^{(k)}), k \ge 1$ is strictly decreasing sequence. Due to the coerciveness, we know that

$$\mathcal{R} := \{\mathbf{x} \in \mathbb{R}^n, F(\mathbf{y}) \le F(\mathbf{y}^{(1)})\}$$

is a bounded set. It follows that the sequence $\{\mathbf{y}^{(k)}\}_{k=1}^{\infty} \subset \mathcal{R}$ is a bounded sequence and there exists a cluster point $\mathbf{y}^*$ and a subsequence $\mathbf{y}^{(k_i)}$ such that $\mathbf{y}^{(k_i)} \to \mathbf{y}^*$. Note that $\{F(\mathbf{y}^{(k)})\}_{k=1}^{\infty}$ is a bounded monotonic descending sequence, then $F(\mathbf{y}^{(k)}) \to F(\mathbf{y}^*)$ for all $k \ge 1$. We claim that the sequence $\{\mathbf{y}^{(k)}\}_{k=1}^{\infty}$ has finite length, that is,

$$\sum_{k=1}^{\infty} \|\mathbf{y}^{(k+1)} - \mathbf{y}^{(k)}\| < \infty. \tag{5.6}$$

To establish the claim, we need to use the Kurdyka-Łojasiewicz inequality (cf. [28]). Note that the $\ell_1$ norm $\|\mathbf{x}\|_1$ is semialgebraic function and the function $f(\mathbf{x})$ is analytic, so the objective function $F(\mathbf{x})$ satisfies the KL property at any critical point (cf. [1], [2], [44]). Let us prove that $\|\nabla F(\mathbf{y}^*)\| = 0$ holds, that is, $\mathbf{y}^*$ is a critical point of $F$. Indeed, using one of the properties in Theorem 5.1, we have

$$\begin{aligned}\|\partial F(\mathbf{y}^{(k)})\| =& \|\partial g(\mathbf{y}^{(k)}) + \nabla F_1(\mathbf{y}^{(k)}) - \nabla F_2(\mathbf{y}^{(k)})\| \\ \le& \|\nabla F_1(\mathbf{y}^{(k)}) - \nabla F_1(\mathbf{y}^{(k-1)})\| + \|\nabla F_2(\mathbf{y}^{(k)}) - \nabla F_2(\mathbf{y}^{(k-1)})\| \\ &+ \frac{L}{2}\|\mathbf{y}^{(k)} - \mathbf{y}^{(k-1)}\|\end{aligned}$$

by using the second conclusion of Theorem 5.1. Combining with (5.5) and the Lipschitz differentiation of $F_1$ and $F_2$, it gives that $\|\partial F(\mathbf{y}^{(k_i)})\| \to 0$. By a property of subgradient of $g$ (cf. [35]) and the continuity of the gradients $F_1$ and $F_2$, we have $\|\partial F(\mathbf{y}^*)\| = 0$ when $\mathbf{y}^{(k_i)} \to \mathbf{y}^*$. Thus, $\mathbf{y}^* \in \text{domain}(\partial F)$, the set of all critical points of $F$.

Therefore, we can use KL inequality to obtain that

$$\varphi'(F(\mathbf{y}) - F(\mathbf{y}^*))\|\partial F(\mathbf{y})\| \ge 1 \tag{5.7}$$

for all $\mathbf{y}$ in the neighborhood $B(\mathbf{y}^*, \delta)$. As $F(\mathbf{y}^{(k)}) - F(\mathbf{y}^*) \to 0$, $k \to \infty$, there is an integer $k_0$ such that for all $k \ge k_0$ it holds

$$\max\left(\sqrt{2/\ell}\sqrt{F(\mathbf{y}^{(k)}) - F(\mathbf{y}^*)}, 2C/\ell \cdot \varphi(F(\mathbf{y}^{(k)}) - F(\mathbf{y}^*))\right) \le \delta/2. \tag{5.8}$$

Without loss of generality, we may assume that $k_0 = 1$ and $\mathbf{y}^{(1)} \in B(\mathbf{y}^*, \delta/2)$. Let us show that $\mathbf{y}^{(k)}, k \ge 1$ will be in the neighborhood $B(\mathbf{y}^*, \delta)$. We shall use an induction to do so. By (5.8) we have

$$\|\mathbf{y}^{(2)} - \mathbf{y}^*\| \le \|\mathbf{y}^{(2)} - \mathbf{y}^{(1)}\| + \|\mathbf{y}^{(1)} - \mathbf{y}^*\| \le \sqrt{2(F(\mathbf{y}^{(1)}) - F(\mathbf{y}^*)/\ell} + \|\mathbf{y}^{(1)} - \mathbf{y}^*\| \le \delta.$$

Assume that $\mathbf{y}^{(k)} \in B(\mathbf{y}^*, \delta)$ for $k \le K$. From Theorem 5.1, we have

$$\begin{aligned}\|\partial F(\mathbf{y}^{k+1})\| &= \|\partial g(\mathbf{y}^{k+1}) + \nabla f(\mathbf{y}^{k+1})\| \\ &= \|\nabla f(\mathbf{y}^{k+1}) - \nabla f(\mathbf{y}^{k}) - \frac{L_1}{2}(\mathbf{y}^{k+1} - \mathbf{y}^{k})\| \le C\|\mathbf{y}^{k+1} - \mathbf{y}^{k}\|,\end{aligned}$$

where constant $C := L + L_1/2$. Putting it into (5.7), it gives that

$$\varphi'(F(\mathbf{y}^k) - F(\mathbf{y}^*)) \geq \frac{1}{C\|\mathbf{y}^k - \mathbf{y}^{k-1}\|}. \tag{5.9}$$

On the other hand, from the concavity of $\varphi$ we get that

$$\varphi(F(\mathbf{y}^k) - F(\mathbf{y}^*)) - \varphi(F(\mathbf{y}^{k+1}) - F(\mathbf{y}^*)) \geq \varphi'(F(\mathbf{y}^k) - F(\mathbf{y}^*))(F(\mathbf{y}^k) - F(\mathbf{y}^{k+1})).$$

Combining with (5.5) and (5.9), we obtain

$$\varphi(F(\mathbf{y}^k) - F(\mathbf{y}^*)) - \varphi(F(\mathbf{y}^{k+1}) - F(\mathbf{y}^*)) \geq \frac{\ell}{2C} \cdot \frac{\|\mathbf{y}^{k+1} - \mathbf{y}^k\|^2}{\|\mathbf{y}^k - \mathbf{y}^{k-1}\|}.$$

Multiplying $\|\mathbf{y}^{(k)} - \mathbf{y}^{(k-1)}\|$ both sides of the above inequality, taking a square root both sides, and the using a standard inequality $2ab \leq a^2 + b^2$ on the left-hand side, we have

$$\|\mathbf{y}^{(k)} - \mathbf{y}^{(k-1)}\| + \frac{2C}{\ell}(\varphi(F(\mathbf{y}^k) - F(\mathbf{y}^*)) - \varphi(F(\mathbf{y}^{k+1}) - F(\mathbf{y}^*))) \geq 2\|\mathbf{y}^{(k)} - \mathbf{y}^{(k+1)}\|$$

for all $2 \leq k \leq K$. It follows that

$$\frac{2C}{\ell}\varphi(F(\mathbf{y}^{(1)}) - F(\mathbf{y}^*)) \geq \sum_{j=1}^{K} \|\mathbf{y}^{(j+1)} - \mathbf{y}^{(j)}\| + \|\mathbf{y}^{(K+1)} - \mathbf{y}^{(K)}\|. \tag{5.10}$$

That is, we have

$$\begin{aligned}\|\mathbf{y}^{(K+1)} - \mathbf{y}^*\| \leq \|\mathbf{y}^{(K+1)} - \mathbf{y}^{(1)}\| + \|\mathbf{y}^{(1)} - \mathbf{y}^*\| &\leq \sum_{j=1}^{K} \|\mathbf{y}^{(j+1)} - \mathbf{y}^{(j)}\| + \|\mathbf{y}^{(1)} - \mathbf{y}^*\| \\ &\leq \frac{2C}{\ell}\varphi(F(\mathbf{y}^{(1)}) - F(\mathbf{y}^*)) + \|\mathbf{y}^{(1)} - \mathbf{y}^*\| \leq \delta.\end{aligned}$$

That is, $\mathbf{y}^{(K+1)} \in B(\mathbf{y}^*, \delta)$, which implies that all $\mathbf{y}^{(k)}$ are in $B(\mathbf{y}^*, \delta)$. From above, we know that the inequality (5.10) holds for all $k$, which show the claim (5.6) holds. It is clear that (5.6) implies that $\{\mathbf{y}^{(k)}\}_{k=1}^{\infty}$ is a Cauchy sequence and hence, it is convergent with $\mathbf{y}^{(k)} \to \mathbf{y}^\star$. Note that $\nabla F(\mathbf{y}^\star) = 0$, which implies $\mathbf{y}^{(k)}$ converges to a critical point of $F$. ∎

Finally, let us show that the convergence is in a linear fashion. We begin with

LEMMA 5.3. *Let $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_1$ for $\lambda > 0$. Then for any $\mathbf{x}$, there exists a $\delta > 0$ such that for any $\mathbf{y} \in B(\mathbf{x}, \delta)$, the open ball of radius $\delta$ at $\mathbf{x}$, there exists a subgradient $\nabla g$ at $\mathbf{x}$,*

$$(\partial g(\mathbf{y}) - \partial g(\mathbf{x}))^\top(\mathbf{y} - \mathbf{x}) = 0. \tag{5.11}$$

*Proof.* For similicity, consider $\mathbf{x} \in \mathbb{R}^1$. Then if $\mathbf{x} \neq 0$, we can find $\delta = |\mathbf{x}| > 0$ such that when $\mathbf{y} \in B(\mathbf{x}, \delta)$, we have $\partial g(\mathbf{y}) = \partial g(\mathbf{x})$ and hence, we have (5.11). If $\mathbf{x} = 0$, for any $y \neq 0$, we choose $\partial g(0)$ according to $\mathbf{y}$, i.e. $\partial g(0) = 1$ if $\mathbf{y} > 0$ and $\partial g(0) = -1$ if $\mathbf{y} < 0$. Then we have (5.11). ∎

In the following lemma, we need to use the sparse set $\mathcal{R}_s$

$$\mathcal{R}_s = \{\mathbf{x} \in \mathbb{R}^n | \|\mathbf{x}\|_0 \le s\} = \bigcup_{\substack{I\subset\{1,\cdots,n\}\\ |I|=s}} \mathbb{R}^s_I \tag{5.12}$$

which is clearly the union of all canonical subspaces $\mathbb{R}^s_I = \text{span}\{\mathbf{e}_{i_1}, \cdots, \mathbf{e}_{i_s}\}$ if $I = \{i_1, i_2, \cdots, i_s\}$.

LEMMA 5.4. *Let $F(\mathbf{x}) = g(\mathbf{x}) + f(\mathbf{x})$. Suppose that $f$ is L-Lipschitz differentiable. Let $\mathbf{x}^*$ be a critical point of $F$ as explained in Theorem 5.2. Suppose that either none of entries of $\mathbf{x}^*$ is zero or suppose $\mathbf{x} \in \mathbb{R}^s_I$ if $\mathbf{x}^* \in \mathbb{R}^s_I$ for some $s \in \{1, \cdots, n\}$. Then there exists $\delta > 0$, e.g. $\delta = \min_{|x_i^*| \neq 0} |x_i^*|$ such that for all $\mathbf{x} \in B(\mathbf{x}^*, \delta)$,*

$$|F(\mathbf{x}) - F(\mathbf{x}^*)| \le C\|\mathbf{x} - \mathbf{x}^*\|^2. \tag{5.13}$$

*Proof.* At $\mathbf{x}^*$, we have $\partial F(\mathbf{x}^*) = \partial g(\mathbf{x}^*) + \nabla f(\mathbf{x}^*) = 0$. By using Lemma 5.3 and either one of the assumptions, we have

$$\begin{aligned}
F(\mathbf{x}) - F(\mathbf{x}^*) &= g(\mathbf{x}) - g(\mathbf{x}^*) + f(\mathbf{x}) - f(\mathbf{x}^*) \\
&\le \partial g(\mathbf{x})^\top (\mathbf{x} - \mathbf{x}^*) + \nabla f(\mathbf{x}*)(\mathbf{x} - \mathbf{x}^*) + \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^\top \nabla^2 f(\xi)(\mathbf{x} - \mathbf{x}^*) \\
&= (\partial g(\mathbf{x}) - \partial g(\mathbf{x}^*))^\top (\mathbf{x} - \mathbf{x}^*) + \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^\top \nabla^2 f(\xi)(\mathbf{x} - \mathbf{x}^*) \\
&= \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^\top \nabla^2 f(\xi)(\mathbf{x} - \mathbf{x}^*),
\end{aligned}$$

where $\xi$ is a point in between $\mathbf{x}^*$ and $\mathbf{x}$. That is, $|F(\mathbf{x}) - F(\mathbf{x}^*)| \le C\|\mathbf{x} - \mathbf{x}^*\|^2$ for a positive constant $C$. ∎

We are now ready to establish the following result on the rate of convergence

THEOREM 5.5. *Suppose that $F_2$ is strongly convex. Starting from any initial guess $\mathbf{x}^{(1)}$, let $\mathbf{x}^{(k+1)}$ be the solution in (3.2) for all $k \ge 1$. Without loss of generality, we assume that $\mathbf{x}^{(k)}, k \ge 1$ converge to a critical point $\mathbf{x}^*$ of $F$ by using Theorem 5.2. Then for any $\epsilon > 0$, either $\mathbf{x}^{(k+1)} \in B(\mathbf{x}^*, \epsilon)$ or*

$$\|\mathbf{x}^{(k+1)} - \mathbf{x}^*\| \le C_\epsilon \tau^k \tag{5.14}$$

*for a positive constant $C_\epsilon$ dependent on $\epsilon$ and $\tau \in (0, 1)$ independent of $\epsilon$.*

*Proof.* According to the results in Theorem 5.1 and Theorem 5.2, we have

$$C_0\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|^2 \le (F(\mathbf{x}^{(k)}) - F(\mathbf{x}^*)) - (F(\mathbf{x}^{(k+1)}) - F(\mathbf{x}^*)) \tag{5.15}$$

for a positive constant $C_0$. We now claim that

$$C_1\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\| \le \sqrt{F(\mathbf{x}^{(k)}) - F(\mathbf{x}^*)} - \sqrt{F(\mathbf{x}^{(k+1)}) - F(\mathbf{x}^*)} \tag{5.16}$$

for a positive constant $C_1$. To establish this claim, we need to use the result in Lemma 5.4. Let us rewrite the inequality in Lemma 5.4 as

$$\frac{1}{\sqrt{F(\mathbf{x}^{(k+1)}) - F(\mathbf{x}^*)}} \ge \frac{C}{\|\mathbf{x}^{(k+1)} - \mathbf{x}^*\|}.$$

Multiplying the above inequality to the inequality in (5.16), we have

$$C_0 C\frac{\|\mathbf{x}^{(k+1)}-\mathbf{x}^{(k)}\|^2}{\|\mathbf{x}^{(k+1)}-\mathbf{x}^*\|} \le \frac{(F(\mathbf{x}^{(k)})-F(\mathbf{x}^*))-(F(\mathbf{x}^{(k+1)})-F(\mathbf{x}^*))}{\sqrt{F(\mathbf{x}^{(k+1)})-F(\mathbf{x}^*)}} \tag{5.17}$$

Consider $g(t)=\sqrt{t}$ which is concave over $[0,1]$, we know $g(t)-g(s)\ge g'(t)(t-s)$. Thus, the right-hand side above is less than or equal to the right-hand side of (5.16). We now work on the left-hand side of the inequality above. Let us first note that $F$ is strongly convex outside the ball $B(\mathbf{x_b},\epsilon)$ (in the real variable setting). If $\mathbf{x}^{(k+1)}$ is within the $B(\mathbf{x_b},\epsilon)$, we do not need to do iterations further when $\epsilon>0$ is a tolerance. Otherwise, we use the strong convexity of $F$ to have

$$C_\epsilon\|\mathbf{x}^{(k+1)}-\mathbf{x}^*\| \le \|\nabla F(\mathbf{x}^{(k+1)})-\nabla F(\mathbf{x}^*)\|$$

for a positive constant dependent on $\epsilon$. The second property of Theorem 5.1 implies

$$\partial g(\mathbf{x}^{(k+1)})+\nabla F_1(\mathbf{x}^{(k)}-\nabla F_2(\mathbf{x}^{(k)})+\frac{L_1}{2}(\mathbf{x}^{(k+1)}-\mathbf{x}^{(k)})=0$$

and $\partial g(\mathbf{x}^*)+\nabla f(\mathbf{x}^*)=0$. By using Lemma 5.3, it follows that

$$\nabla F(\mathbf{x}^{(k+1)})-\nabla F(\mathbf{x}^*)=\nabla f(\mathbf{x}^{(k+1)})-\nabla f(\mathbf{x}^{(k)})-\frac{L_1}{2}(\mathbf{x}^{(k+1)}-\mathbf{x}^{(k)}).$$

In other works,

$$C_\epsilon\|\mathbf{x}^{(k+1)}-\mathbf{x}^*\| \le \|\nabla f(\mathbf{x}^{(k)})-\nabla f(\mathbf{x}^{(k+1)})\|+\frac{L_1}{2}\|\mathbf{x}^{(k+1)}-\mathbf{x}^{(k)}\|$$

Using the Lipschitz differentiability of $f$, we have

$$\|\mathbf{x}^{(k+1)}-\mathbf{x}^*\| \le \frac{L+L_1}{C_\epsilon}\|\mathbf{x}^{(k+1)}-\mathbf{x}^{(k)}\|.$$

The left-hand side of the equation in (5.17) can be simplified to be

$$C_0 C\frac{L_1+L}{C_\epsilon}\|\mathbf{x}^{(k+1)}-\mathbf{x}^{(k)}\|$$

which is the desired term on the left-hand of the inequality in (5.16). These establish the claim.

By summing the inequality in (5.16) above, it follows

$$\sum_{k\ge 1}\|\mathbf{x}^{(k+1)}-\mathbf{x}^{(k)}\| \le \frac{1}{C_1}\sqrt{f(\mathbf{x}^{(1)})-f(\mathbf{x}^*)}.$$

That is, $\mathbf{x}^{(k)}$ is a Cauchy sequence and hence, it is convergent.

The remaining part of the proof is to establish the convergence rate. The proof is similar to the one in a previous section. We leave the detail to the interested reader.□

**6. Numerical Results.** In this section, we report some computational results from our DC based algorithm and $\ell_1$ DC based algorithm. The significance of these results is to demonstrate that the DC based algorithm is able to retrieve real signals of size $n$ from $m$ measurements with high probability around 80% over 1000 repeated runs when $m\approx 2n$. As demonstrated in [13] (cf. Figures 8 and 9), the truncated

Wirtinger flow algorithm, and the original Wirtinger flow algorithm needs $m \approx 3n$ to be able to retrieve Gaussian random signals of size $n$. To retrieve sparse solution, the $\ell_1$ DC based algorithm together with thresholding technique needs only $m \approx n$ measurements. Thus this section is divided into two subsections. We first present how to use our DC based algorithm to retrieve general signals in the real and complex variable settings. As the Wirtinger flow (Wf) algorithm requires $m \geq 3n$ to be able to retrieve the real variable solution, we shall not show the performance of the Wf algorithm. Instead, we shall also present the Gauss-Newton algorithm from [22] to compare with our DC based algorithm. Next we shall present numerical experimental results to demonstrate our $\ell_1$ DC based algorithm is able to use $m \approx n$ measurements to retrieve sparse signals.

### 6.1. Phase Retrieval of General Signals.

*Example* 6.1. In this example, we recover the solution $\mathbf{x_b}$ from the given measurements (1.1) using Gaussian random measurement vectors $\mathbf{a}_j, j = 1, \cdots, m$. The number $m$ of measurements is around the twice of the size of the solution $\mathbf{x_b}$. In Table 1, we show the number of successes of retrieving $\mathbf{x_b}$ over 1000 repeated runs. We fix $n = 128$ and $m = k * n/16$ for $k = 12, 13, \cdots, 35$.

TABLE 1

*The numbers of successful retrieved solution over 1000 repeated runs based on numbers of measurements satisfying the relations $m/n$ listed above, where $n$ is the size of the solution*

| $m/n$ | 1.3750 | 1.4375 | 1.5000 | 1.5625 | 1.6250 | 1.6875 | 1.7500 |
|---|---|---|---|---|---|---|---|
| successes | 20 | 48 | 107 | 150 | 239 | 284 | 446 |
| $m/n$ | 1.8125 | 1.8750 | 1.9375 | 2 | 2.0625 | 2.1250 | 2.1875 |
| successes | 511 | 588 | 650 | 708 | 771 | 844 | 882 |

From Table 1, we can see that the DC based algorithm can retrieve the solutions using the number $m$ of measurements around $2n$ with high probabilities $\geq 70\%$.

*Example* 6.2. We next repeat Example 6.1 using more number of measurements. In this case, we are able to use the Gauss-Newton method to retrieve solutions as the Hessian can be inverted. Hence, we will compare the numbers of successes from the Gauss-Newton method and the DC based method in Table 2.

TABLE 2

*The numbers of successful retrieved solution over 1000 repeated runs based on numbers of measurements satisfying the relations $m/n$ listed above, where $n = 128$ is the size of the solution*

| $m/n$ | 2.3125 | 2.3750 | 2.4375 | 2.5000 | 2.5625 | 2.6250 | 2.6875 |
|---|---|---|---|---|---|---|---|
| GN successes | 560 | 672 | 708 | 776 | 831 | 882 | 912 |
| DC successes | 937 | 957 | 961 | 967 | 991 | 991 | 989 |
| $m/n$ | 2.7500 | 2.8125 | 2.8750 | 2.9375 | 3 | 3.0625 | 3.1250 |
| GN successes | 939 | 950 | 960 | 980 | 986 | 987 | 991 |
| DC successes | 995 | 994 | 995 | 1000 | 1000 | 998 | 998 |

*Example* 6.3. This example shows that robustness of the DC based algorithm. We repeat the computation in Example 6.1 by adding noises to the measurements. One way to generate a noisy input is to add noises $\epsilon_j$ to the clean measurements:

$$\hat{b}_j = |\langle \mathbf{a}_j, \mathbf{x_b} \rangle|^2 + \epsilon_j, \quad j = 1, \cdots, m \tag{6.1}$$

Another way to generate a noisy input is to add noises $\delta_j$ and $\epsilon_j$ to the clean measurements:

$$\tilde{b}_j = |\langle \mathbf{a}_j, \mathbf{x_b} \rangle + \delta_j|^2 + \epsilon_j, \quad j = 1, \cdots, m \tag{6.2}$$

For noisy measurements of model (6.1), we assume that $\epsilon_j$ are subject to uniform random distribution between $[-u, u]$ with mean zero, where $u = 1e-1$, $1e-3$ and $1e-5$. As long as the tolerance as a stopping criterion for the Gauss-Newton method and DC based method is the same as $e$ or bigger than $e$ both algorithms produce the same successes of retrieval as in Table 2. For noisy measurements of model (6.2), we assume that both $\epsilon_j$ and $\delta_j$ are subject to uniform distribution between $[-u, u]$ with mean zero. When the tolerance is the same as $e$ or larger than $e$, both algorithms can retrieve the solution from noisy measurements (6.2) just as the same as in the previous example.

In addition to the retrieve real variable solutions, we shall also repeat the same experiments for complex variable solutions.

*Example* 6.4. In this example, we use the DC based algorithm and the Gauss-Newton method to retrieve complex variable solutions. The number of measurements vs. the number of entries of complex variable solutions is around 3. That is, $m \approx 3n$ with $n = 128$. For Gaussian random measurements $\mathbf{a}_j = \mathbf{a}_{j,R} + \mathbf{i}\mathbf{a}_{j,I}$, $j = 1, \cdots, m$, we retrieve $\mathbf{z} \in \mathbb{R}^n$ with $\mathbf{z} = \mathbf{x} + \mathbf{i}\mathbf{y}$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ from $|\langle \mathbf{a}_j, \mathbf{z} \rangle|^2, j = 1, \cdots, m$. We will show the numbers of successes from the Gauss-Newton method and the DC based method over 100 repeated runs in Table 3. The Table 3 shows that the DC

TABLE 3
*The numbers of successful retrieved solution over 100 repeated runs based on numbers of measurements satisfying the relations $m/n$ listed above, where $n = 128$ is the size of the solution*

| $m/n$ | 2.562 | 2.625 | 2.687 | 2.750 | 2.812 | 2.875 | 2.937 | 3.000 | 3.062 | 3.125 |
|---|---|---|---|---|---|---|---|---|---|---|
| GN alg. | 7 | 21 | 21 | 37 | 56 | 60 | 75 | 72 | 84 | 88 |
| DC alg. | 0 | 0 | 0 | 3 | 19 | 52 | 78 | 80 | 94 | 99 |

based algorithm is able to retrieve complex variable solutions very well when $m \geq 3n$ while the Gauss-Newton method can find solutions even when $m \geq 2.5n$. However, the successful rate is lower than the DC based algorithms starting from $m \geq 2.9n$.

We now present some numerical results to demonstrate that the $\ell_1$ DC based Algorithm 6.1 works well.

**Algorithm 6.1** $\ell_1$ DC based Algorithm

We use the same initialization as in the previous examples.
**while** $k \geq 1$ **do**
  Solve (5.3) to get $\mathbf{y}^{(k+1)}$.
  Apply a modified Attouch-Peypouquet technique to get a new $\mathbf{y}^{(k+1)}$.
  Until the maximal number of iterations is reached.
**end while**
**return** $\mathbf{y}^T$

In Algorithm 6.1, the modified Attouch-Peypouquet technique is to use the Attouch-Peypouquet iteration (cf. [3]) in the first few $k$ iterations, say $k \leq K$ with variable step size $\beta_k = k/(k + \alpha)$ and then a fixed step size $\beta_K$ for the remaining iterations.

*Example* 6.5. We have experimented Algorithm 6.1 numerically for retrieving solutions in the real variable setting. We use $n = 100$ and $m = 1.1n, 1.2n, \ldots, 2.5n$. All measurement vectors $\mathbf{a}_j, j = 1, \ldots, m$ are Gaussian random vectors. So is $\mathbf{x_b}$. We use Algorithm 6.1 to recover $\mathbf{x_b}$ from $b_j = |\langle \mathbf{a}_j, \mathbf{x_b}\rangle|^2, j = 1, \ldots, m$ based on 5000 iterations when $n = 100$. To recover a general solution $\mathbf{x_b}$, we use a small value $\lambda = 1e - 5$. We repeat the experiment 100 times and summarize the frequency of retrievals listed in Table 4.

TABLE 4

*The numbers of successful retrieved solution over 100 repeated runs based on numbers of measurements satisfying the relations $m/n$ listed above, where $n = 100$ is the size of the solution*

| $m/n$ | 1.5 | 1.6 | 1.7 | 1.8 | 1.9 | 2 | 2.1 | 2.2 | 2.3 | 2.4 | 2.5 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\ell_1$ DC alg. | 0 | 0 | 2 | 8 | 28 | 57 | 72 | 91 | 93 | 93 | 99 |

**6.2. Phase Retrieval of Sparse Signals.** Next we explain how to use our $\ell_1$ DC based algorithm to retrieve sparse solutions. A key point is to use $m$ measurement values with $m$ smaller than $2n$ or even small than $n$. In such a setting, many existing algorithms will fail. As shown in the subsection above, when $m \approx 2n$, we are able to retrieve any solution, no matter sparse or not. However, when $m \approx 1.5n$, we are not able to retrieve general signals. The point of our numerical experiments is to see if we are able to retrieve sparse signals when $m \approx n$. The performance of (1.6) is not very good. We need to improve it. By using the sparsity, we will enhance the $\ell_1$ based algorithm by using the projection technique. That is, we project $\mathbf{y}^{(k+1)}$ from (1.6) to the set of all $s$-sparse vectors. That is, we use the hard thresholding technique to update $\mathbf{y}^{(k+1)}$. This leads to an $\ell_1$ DC based algorithm with hard thresholding technique given below.

**Algorithm 6.2** $\ell_1$ DC based Algorithm with Hard Thresholding

We use the same initialization as in the previous subsection.
**while** $k \geq 1$ **do**
Solve (5.3) using the shrinkage-thresholding technique to get $\mathbf{y}^{(k+1)}$.
Apply a modified Attouch-Peypouquet technique to get a new $\mathbf{y}^{(k+1)}$.
Project $\mathbf{y}^{(k+1)}$ into the collection $\mathcal{R}_s$, $s$-sparse set. That is, let $\mathbf{z}^{(k+1)}$ solve the following minimization problem:

$$\sigma_s(\mathbf{x}^k) = \min_{\mathbf{z} \in \mathcal{R}_s} \|\mathbf{y}^{(k+1)} - \mathbf{z}\|_1 \tag{6.3}$$

Let $\mathbf{y}^{(k+1)} = \mathbf{z}^{(k+1)}$
**end while**
**return** the maximal number of iterations $\mathbf{y}^T$

We now present some numerical results to demonstrate that Algorithm 6.2 works well.

*Example* 6.6. Fix $m \leq n$. Many existing computational algorithms fail as the number of measurements is too small. However, Algorithm 6.2 is able to retrieve sparse solutions from the phaseless measurements. Let us present our numerical findings in Table 5 and Table 6.

Table 5
*The numbers of successful retrieval of sparse solutions with sparsities $s = 1, 5, 10, 20, 30, 40$ over 100 repeated runs based on numbers of measurements satisfying the relations $m/n$ listed below, where $n = 100$ is the size of the solution*

| $m/n$ | | 1.1 | 1.2 | 1.3 | 1.4 | 1.5 | 1.6 | 1.7 | 1.8 | 1.9 | 2 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Alg. 6.2 | $s = 1$ | 75 | 83 | 86 | 80 | 88 | 82 | 83 | 87 | 86 | 94 |
| Alg. 6.2 | $s = 5$ | 54 | 61 | 72 | 63 | 80 | 80 | 74 | 85 | 83 | 87 |
| Alg. 6.2 | $s = 10$ | 40 | 37 | 54 | 46 | 54 | 70 | 64 | 72 | 79 | 80 |
| Alg. 6.2 | $s = 20$ | 15 | 16 | 22 | 27 | 36 | 45 | 42 | 47 | 59 | 57 |
| Alg. 6.2 | $s = 30$ | 0 | 3 | 3 | 10 | 15 | 22 | 33 | 36 | 39 | 44 |
| Alg. 6.2 | $s = 40$ | 0 | 0 | 0 | 0 | 6 | 10 | 11 | 17 | 30 | 39 |

Table 6
*The numbers of successful retrieval of sparse solutions with sparsities $s = 1, 2, ..., 10$ over 100 repeated runs based on numbers of measurements satisfying the relations $m/n$ listed above, where $n = 100$ is the size of the solution*

| $m/n$ | | 1 | 0.9 | 0.8 | 0.7 | 0.6 | 0.5 |
|---|---|---|---|---|---|---|---|
| Alg. 6.2 | $s = 1$ | 83 | 80 | 66 | 66 | 58 | 52 |
| Alg. 6.2 | $s = 2$ | 80 | 64 | 73 | 57 | 53 | 45 |
| Alg. 6.2 | $s = 4$ | 59 | 60 | 47 | 43 | 30 | 18 |
| Alg. 6.2 | $s = 5$ | 56 | 44 | 38 | 26 | 11 | 7 |
| Alg. 6.2 | $s = 10$ | 23 | 14 | 4 | 0 | 0 | 0 |

From Table 5 and Table 6, we can see that Algorithm 6.2 is able to recover sparse solutions with high frequency of success.

**7. Appendix.** In this section we give some deterministic description of the landscape function of the minimizing function $F$ in (1.2). We will show that $F$ has a positive definite Hessian at a global minimizer in the real value setting and $F$ is non-negative positive definite in the complex value setting. These results were used when applying the K-L inequality. For convenience, let $A_\ell = \mathbf{a}_\ell \bar{\mathbf{a}}_\ell^\top$ be the Hermitian matrix of rank one for $\ell = 1, \cdots, m$. We first need

Definition 7.1. *We say $\mathbf{a}_j, j = 1, \cdots, m$ are $a_0$-generic if they satisfy*

$$\|(\mathbf{a}_{j_1}^* \mathbf{y}, \ldots, \mathbf{a}_{j_n}^* \mathbf{y})\| \geq a_0 \|\mathbf{y}\|, \quad \forall \mathbf{y} \in \mathbb{C}^n$$

*for a positive $a_0 \in (0, 1)$ for any $1 \leq j_1 < j_2 < \cdots < j_n \leq m$.*

Theorem 7.2. *Consider the real variable setting. Let $H_f(\mathbf{x})$ be the Hessian of the minimizing function $f(\mathbf{z})$ and let $\mathbf{x}^\star$ be a global minimizer of (1.2). Suppose that $\mathbf{a}_j, j = 1, \cdots, m$ are in $a_0$-generic position. Then $H_f(\mathbf{x}^\star)$ is positive definite.*

*Proof.* Recall $A_\ell = \mathbf{a}_\ell \bar{\mathbf{a}}_\ell^\top$ for $\ell = 1, \cdots, m$. It is easy to see

$$\nabla f(\mathbf{x}) = 2 \sum_{\ell=1}^m (\mathbf{x}^\top A_\ell \mathbf{x} - b_\ell) A_\ell \mathbf{x}$$

and the entries $h_{ij}$ of the Hessian $H_f(\mathbf{x})$ is

$$h_{ij} = \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} f(\mathbf{x}) = 2 \sum_{\ell=1}^m (\mathbf{x}^\top A_\ell \mathbf{x} - b_\ell) a_{ij}(\ell) + 4 \sum_{p=1}^n a_{i,p}(\ell) x_p \sum_{q=1}^n a_{j,q}(\ell) x_q,$$

where $A_\ell = [a_{ij}(\ell)]_{ij=1}^n$. As we have $(\mathbf{x}^*)^\top A_\ell \mathbf{x}^* = b_\ell, \ell = 1, \cdots, m$, the first summation term of $h_{ij}$ above is zero at $\mathbf{x}^*$. Letting $M(\mathbf{y}) = \mathbf{y}^\top H_f(\mathbf{x}^*) \mathbf{y}$ be a quadratic function of $\mathbf{y}$, we have

$$\begin{aligned}M(\mathbf{y}) &= 4\sum_{\ell=1}^m (\mathbf{y}^\top A_\ell \mathbf{x}^*(\mathbf{x}^*)^\top A_\ell \mathbf{y} = 4\sum_{\ell=1}^m |\mathbf{y}^\top A_\ell \mathbf{x}^*|^2 \\ &= 4\sum_{\ell=1}^m |\mathbf{y}^\top \mathbf{a}_\ell|^2 |\bar{\mathbf{a}}_\ell^\top \mathbf{x}^*|^2 \geq 4a_0\|\mathbf{x}^*\|^2\|\mathbf{y}\|^2.\end{aligned}$$

by using the definition of $a_0$ as in a previous section. It follows that $H_f(\mathbf{x}^*)$ is positive definite. ∎

Next let us show that the global minimizer $\mathbf{x}^\star$ in the complex setting. In this case, the Hessian $H_F(\mathbf{x}^*)$ is no longer positive definite. Instead, it is nonnegative definite. To this end, let us fix some notations. Write $\mathbf{a}_\ell = a_\ell + \mathbf{i}c_\ell$ for $\ell = 1, \cdots, m$. For $\mathbf{z} = \mathbf{x} + \mathbf{i}\mathbf{y}$, we have $\mathbf{a}_\ell^\top \mathbf{z}^* = b_\ell$ for the global minimizer $\mathbf{z}^*$. Writing $f_\ell(\mathbf{x}, \mathbf{y}) = |\mathbf{a}_\ell^\top \mathbf{z}|^2 - b_\ell = (a_\ell^\top \mathbf{x} - c_\ell^\top \mathbf{y})^2 + (c_\ell^\top \mathbf{x} + a_\ell^\top \mathbf{y})^2 - b_\ell$, we consider

$$f(\mathbf{x}, \mathbf{y}) = \frac{1}{m}\sum_{\ell=1}^m f_\ell^2. \tag{7.1}$$

The gradient of $f$ can be easily found as follows: $\nabla f = [\nabla_\mathbf{x} f, \nabla_\mathbf{y} f]$ with

$$\nabla_\mathbf{x} f(\mathbf{x}, \mathbf{y}) = \frac{1}{m}\sum_{\ell=1}^m \nabla_\mathbf{x} f_\ell^2 = \frac{4}{m}\sum_{\ell=1}^m f_\ell(\mathbf{x}, \mathbf{y})[(a_\ell^\top \mathbf{x} - c_\ell^\top \mathbf{y})a_\ell + (c_\ell^\top \mathbf{x} + a_\ell^\top \mathbf{y})c_\ell] \tag{7.2}$$

and

$$\nabla_\mathbf{y} f(\mathbf{x}, \mathbf{y}) = \frac{1}{m}\sum_{\ell=1}^m \nabla_\mathbf{y} f_\ell^2 = \frac{4}{m}\sum_{\ell=1}^m f_\ell(\mathbf{x}, \mathbf{y})[(a_\ell^\top \mathbf{x} - c_\ell^\top \mathbf{y})(-c_\ell) + (c_\ell^\top \mathbf{x} + a_\ell^\top \mathbf{y})a_\ell]. \tag{7.3}$$

The Hessian of $f$ is more complicated:

$$H_f(\mathbf{x}, \mathbf{y}) = \begin{bmatrix} \nabla_\mathbf{x}\nabla_\mathbf{x} f(\mathbf{x}, \mathbf{y}) & \nabla_\mathbf{x}\nabla_\mathbf{y} f(\mathbf{x}, \mathbf{y}) \\ \nabla_\mathbf{y}\nabla_\mathbf{x} f(\mathbf{x}, \mathbf{y}) & \nabla_\mathbf{y}\nabla_\mathbf{y} f(\mathbf{x}, \mathbf{y}) \end{bmatrix} \tag{7.4}$$

with $\nabla_\mathbf{x}\nabla_\mathbf{x} f(\mathbf{x}, \mathbf{y}), \cdots, \nabla_\mathbf{y}\nabla_\mathbf{y} f(\mathbf{x}, \mathbf{y})$ given below.

$$\begin{aligned}\nabla_\mathbf{x}\nabla_\mathbf{x} f(\mathbf{x}, \mathbf{y}) &= \frac{4}{m}\sum_{\ell=1}^m f_\ell(\mathbf{x}, \mathbf{y})[a_\ell a_\ell^\top + c_\ell c_\ell^\top] \\ &+ \frac{8}{m}\sum_{\ell=1}^m [(a_\ell^\top \mathbf{x} - c_\ell^\top \mathbf{y})a_\ell + (c_\ell^\top \mathbf{x} + a_\ell^\top \mathbf{y})c_\ell][(a_\ell^\top \mathbf{x} - c_\ell^\top \mathbf{y})a_\ell^\top + (c_\ell^\top \mathbf{x} + a_\ell^\top \mathbf{y})c_\ell^\top]\end{aligned}$$

$$\begin{aligned}\nabla_\mathbf{x}\nabla_\mathbf{y} f(\mathbf{x}, \mathbf{y}) &= \frac{4}{m}\sum_{\ell=1}^m f_\ell(\mathbf{x}, \mathbf{y})[a_\ell(-c_\ell)^\top + c_\ell a_\ell^\top] \\ &+ \frac{8}{m}\sum_{\ell=1}^m [(a_\ell^\top \mathbf{x} - c_\ell^\top \mathbf{y})a_\ell + (c_\ell^\top \mathbf{x} + a_\ell^\top \mathbf{y})c_\ell][(a_\ell^\top \mathbf{x} - c_\ell^\top \mathbf{y})(-c_\ell)^\top + (c_\ell^\top \mathbf{x} + a_\ell^\top \mathbf{y})a_\ell^\top]\end{aligned}$$

and similar for $\nabla_\mathbf{y}\nabla_\mathbf{x} f(\mathbf{x}, \mathbf{y})$ abd $\nabla_\mathbf{y}\nabla_\mathbf{y} f(\mathbf{x}, \mathbf{y})$.

We are now ready to prove the nonnegativity of the Hessian at a global minimizer $\mathbf{z}^*$ in the complex value setting.

Theorem 7.3. *At any global minimizer* $\mathbf{z}^* = (\mathbf{x}^*, \mathbf{y}^*)$*, we have the Hessian* $H_f(\mathbf{x}^*, \mathbf{y}^*) \geq 0$*. In fact,* $H_f(\mathbf{x}^*, \mathbf{y}^*) = 0$ *along the direction* $[-(\mathbf{y}^*)^\top, (\mathbf{x}^*)^\top]^\top$.

*Proof.* At $\mathbf{z}^* = \mathbf{x}^* + \mathbf{i}\mathbf{y}^*$, we have

$$\nabla_{\mathbf{x}}\nabla_{\mathbf{x}} f(\mathbf{x}^*, \mathbf{y}^*) = \frac{8}{m}\sum_{\ell=1}^m [(a_\ell^\top \mathbf{x}^* - c_\ell^\top \mathbf{y}^*)a_\ell + (c_\ell^\top \mathbf{x}^* + a_\ell^\top \mathbf{y}^*)c_\ell]\,[(a_\ell^\top \mathbf{x}^* - c_\ell^\top \mathbf{y}^*)a_\ell^\top + (c_\ell^\top \mathbf{x}^* + a_\ell^\top \mathbf{y}^*)c_\ell^\top]$$

$$\nabla_{\mathbf{x}}\nabla_{\mathbf{y}} f(\mathbf{x}^*, \mathbf{y}^*) = \frac{8}{m}\sum_{\ell=1}^m [(a_\ell^\top \mathbf{x}^* - c_\ell^\top \mathbf{y}^*)a_\ell + (c_\ell^\top \mathbf{x}^* + a_\ell^\top \mathbf{y}^*)c_\ell]\,[(a_\ell^\top \mathbf{x}^* - c_\ell^\top \mathbf{y}^*)(-c_\ell)^\top + (c_\ell^\top \mathbf{x}^* + a_\ell^\top \mathbf{y}^*)a_\ell^\top]$$

and similar for the other two terms. It is easy to see that for any $\mathbf{w} = \mathbf{u} + \mathbf{i}\mathbf{v}$ with $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, we have

$$\begin{aligned}
&[\mathbf{u}^\top\ \mathbf{v}^\top]^\top H_f(\mathbf{x}^*, \mathbf{y}^*) \begin{bmatrix}\mathbf{u}\\ \mathbf{v}\end{bmatrix} \\
=&\frac{8}{m}\sum_{\ell=1}^m [(a_\ell^\top \mathbf{x}^* - c_\ell^\top \mathbf{y}^*)a_\ell^\top \mathbf{u} + (c_\ell^\top \mathbf{x}^* + a_\ell^\top \mathbf{y}^*)c_\ell^\top \mathbf{u}]^2 + \frac{8}{m}\sum_{\ell=1}^m [(a_\ell^\top \mathbf{x}^* - c_\ell^\top \mathbf{y}^*)(-c_\ell)^\top \mathbf{v} + (c_\ell^\top \mathbf{x}^* + a_\ell^\top \mathbf{y}^*)a_\ell^\top \mathbf{v}]^2 \\
&+ \frac{8}{m}\sum_{\ell=1}^m 2[(a_\ell^\top \mathbf{x}^* - c_\ell^\top \mathbf{y}^*)a_\ell^\top \mathbf{u} + (c_\ell^\top \mathbf{x}^* + a_\ell^\top \mathbf{y}^*)c_\ell^\top \mathbf{u}]\,[(a_\ell^\top \mathbf{x}^* - c_\ell^\top \mathbf{y}^*)(-c_\ell)^\top \mathbf{v} + (c_\ell^\top \mathbf{x}^* + a_\ell^\top \mathbf{y}^*)a_\ell^\top \mathbf{v}] \\
=&\frac{8}{m}\sum_{\ell=1}^m [(a_\ell^\top \mathbf{x}^* - c_\ell^\top \mathbf{y}^*)a_\ell^\top \mathbf{u} + (c_\ell^\top \mathbf{x}^* + a_\ell^\top \mathbf{y}^*)c_\ell^\top \mathbf{u} + (a_\ell^\top \mathbf{x}^* - c_\ell^\top \mathbf{y}^*)(-c_\ell)^\top \mathbf{v} + (c_\ell^\top \mathbf{x}^* + a_\ell^\top \mathbf{y}^*)a_\ell^\top \mathbf{v}]^2 \\
\geq&0.
\end{aligned}$$

∎

Furthermore, if we choose $\mathbf{u} = -\mathbf{y}^*$ and $\mathbf{v} = \mathbf{x}^*$, we can easily see that the Hessian $H_f$ along this direction is zero:

$$[-(\mathbf{y}^*)^\top\ (\mathbf{x}^*)^\top]^\top H_f(\mathbf{x}^*, \mathbf{y}^*) \begin{bmatrix}-\mathbf{y}^*\\ \mathbf{x}^*\end{bmatrix} = 0.$$

## REFERENCES

[1] Hedy Attouch and Jérôme Bolte. On the convergence of the proximal algorithm for nonsmooth functions involving analytic features. *Mathematical Programming*, 116(1-2):5–16, 2009.

[2] Hédy Attouch, Jérôme Bolte, Patrick Redont, and Antoine Soubeyran. Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality. *Mathematics of Operations Research*, 35(2):438–457, 2010.

[3] H. Attouch and J. Peypouquet, The rate of convergence of Nesterov's accelerated forward-backward method is actually faster than $1/k^2$, SIAM J. Optim. 26(2016), 1824–1834.

[4] Radu Balan, Pete Casazza, and Dan Edidin. On signal reconstruction without phase. *Applied and Computational Harmonic Analysis*, 20(3):345–356, 2006.

[5] Jonathan Barzilai and Jonathan M Borwein. Two-point step size gradient methods. *IMA journal of numerical analysis*, 8(1):141–148, 1988.

[6] A. Beck and Y. C. Eldar. Sparsity constrained nonlinear optimization: Optimality conditions and algorithms. *SIAM Journal on Optimization*, 23(3):1480C1509, 2013a.

[7] A. Beck and M. Teboulle, A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems, SIAM J. Imaging Sciences, Vol. 2 (2009), pp. 183–202.

[8] Emrah Bostan, Mahdi Soltanolkotabi, David Ren, and Laura Waller. Accelerated Wirtinger flow for multiplexed fourier ptychographic microscopy. *arXiv preprint arXiv:1803.03714*, 2018.

[9] T. Tony Cai, Xiaodong Li, and Zongming Ma, Optimal rates of convergence for noisy sparse phase retrieval via thresholded Wirtinger flow, The Annals of Statistics 2016, Vol. 44, No. 5, 2221-2251.
[10] Emmanuel J Candes, Yonina C Eldar, Thomas Strohmer, and Vladislav Voroninski. Phase retrieval via matrix completion. *SIAM review*, 57(2):225–251, 2015.
[11] Emmanuel J. Candes, Xiaodong Li,and Mahdi Soltanolkotabi, Phase retrieval from coded diffraction patterns, Applied and Computational Harmonic Analysis 39(2015), no.2, 277–299.
[12] Emmanuel J Candes, Xiaodong Li, and Mahdi Soltanolkotabi. Phase retrieval via wirtinger flow: Theory and algorithms. *IEEE Transactions on Information Theory*, 61(4):1985–2007, 2015.
[13] Yuxin Chen and Emmanuel Candes. Solving random quadratic systems of equations is nearly as easy as solving linear systems. Communications on Pure and Applied Mathematics, vol. 70, issue 5, pp. 822-883, May 2017.
[14] Aldo Conca, Dan Edidin, Milena Hering, and Cynthia Vinzant. An algebraic characterization of injectivity in phase retrieval. *Applied and Computational Harmonic Analysis*, 38(2):346–356, 2015.
[15] Yonina C. Eldar and Shahar Mendelson. Phase retrieval: Stability and recovery guarantees. *Applied and Computational Harmonic Analysis*,Volume 36, Issue 3, May 2014, Pages 473–494.
[16] Y. C. Eldar and S. Mendelson. Phase retrieval: Stability and recovery guarantees. Applied and Computational Harmonic Analysis, 36(3):473-494, 2014.
[17] Chris Dainty and James R. Fienup, Phase retrieval and image reconstruction for astronomy, Image Recovery: Theory and Application(1987),231–275.
[18] Y. -H. Dai and L. -Z. Liao, R-linear convergence of the Barzilai and Borwein gradient method, IMA Journal of Numerical Analysis (2002) 22, 1–10.
[19] Frank R Deutsch. *Best approximation in inner product spaces*. Springer, New York, 2001.
[20] James R. Fienup, Phase retrieval algorithms: a comparison, Applied Optics 21(1982),no.15, 2758–2769.
[21] William Fulton. *Intersection theory*, volume 2. Springer Science & Business Media, 2013.
[22] Bing Gao and Zhiqiang Xu. Phaseless recovery using the gauss–newton method. *IEEE Transactions on Signal Processing*, 65(22):5885–5896, 2017.
[23] P. Gong, C. Zhang, Z. Lu, J. Huang, J. Ye, A general iterative shrinkage and thresholding algorithm for non-convex regularized optimization problems, The 30th International Conference on Machine Learning (ICML), 37–45, 2013.
[24] Joe Harris. Algebraic geometry: a first course. 1992.
[25] Robin Hartshorne. *Algebraic geometry*, volume 52. Springer Science & Business Media, 2013.
[26] Meng Huang and Zhiqiang Xu. Phase retrieval from the magnitudes of affine measurements. *arXiv preprint arXiv:1805.07899*, 2018.
[27] Kishore Jaganathan, Yonina C. Eldar, Babak Hassibi, Phase Retrieval: An Overview of Recent Developments, arXiv:1510.07713v1, Oct. 26, 2015.
[28] Krzysztof Kurdyka. On gradients of functions definable in o-minimal structures. In *Annales de l'institut Fourier*, volume 48, pages 769–784. Chartres: L'Institut, 1950-, 1998.
[29] S. Łojasiewicz, Une propriété topologique des sous-ensembles analytiques réels, in Les Équations aux Dérivées Partielles, pp. 87–89, Éditions du centre National de la Recherche Scientique, Paris, 1963.
[30] Jianwei Miao, Tetsuya Ishikawa, Bart Johnson, Erik H. Anderson, Barry Lai, and Keith O. Hodgson, High resolution 3D X-Ray diffraction microscopy, Phys. Rev. Lett. 89(2002), no.8,088303.
[31] R.P. Millane, Phase retrieval in crystallography and optics, Journal of the Optical Society of America A7(1990), no.3, 394–411.
[32] David Mumford. The red book of varieties and schemes, volume 1358 of lecture notes in mathematics. 1999.
[33] Yurii Nesterov. *Introductory lectures on convex optimization: A basic course*, volume 87. Springer Science & Business Media, 2013.
[34] W. Harrison Robert, Phase problem in crystallography, Journal of the Optical Society of America A 10(1993),no.5,1046–1055.
[35] R. T. Rockafellar, Convex Analysis, Princeton University Press, 1970.
[36] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, Learning representations by back-propagating errors. Nature, 323(6088):533-536, 1986.
[37] I. R. Shafarevich. *Basic algebraic geometry 1*. Springer-Verlag Berlin Heidelberg, 2013.
[38] Y. Shechtman, A. Beck and Y. C. Eldar, GESPAR: Efficient phase retrieval of sparse signals,

IEEE Transactions on Signal Processing 62, no. 4 (2014): 928–938.

[39] Yoav Shechtman, Yonina C. Eldar, Oren Cohen, Henry N. Chapman, Jianwei Miao, and Mordechai Segev, Phase retrieval with application to optical imaging: A contemporary overview, Signal Processing Magazine, IEEE 32(2015),no.3,87–109.

[40] Ju Sun, Qing Qu, and John Wright, A Geometric Analysis of Phase Retrieval, Foundations of Computational Mathematics, (2017) pp 1–68.

[41] Cynthia Vinzant. A small frame and a certificate of its injectivity. In *Sampling Theory and Applications (SampTA), 2015 International Conference on*, pages 197–200. IEEE, 2015.

[42] Yang Wang, Zhiqiang Xu, Phase Retrieval for Sparse Signals, Applied Computational Harmonic Analysis, Vol 37, 2014, 531–544, 2014.

[43] Yang Wang and Zhiqiang Xu. Generalized phase retrieval: measurement number, matrix recovery and beyond. *Applied and Computational Harmonic Analysis*, 2017.

[44] Yangyang Xu and Wotao Yin. A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. *SIAM Journal on imaging sciences*, 6(3):1758–1789, 2013.

[45] Oscar Zariski. On the purity of the branch locus of algebraic functions. *Proceedings of the National Academy of Sciences*, 44(8):791–796, 1958.

[46] Huishuai Zhang and Yingbin Liang, Reshaped Wirtinger flow for solving quadratic systems of equations, arXiv preprint arXiv:1605.07719(2016).

[47] Yutao Zheng and Bing Zheng. A new modified barzilai–borwein gradient method for the quadratic minimization problem. *Journal of Optimization Theory and Applications*, 172(1):179–186, 2017.