On DC based Methods for Phase Retrieval

Meng Huang[1], Ming-Jun Lai[2], Abraham Varghese[2], and Zhiqiang Xu[3]

[1] Department of Mathematics, Hong Kong University of Science and Technology, Hong Kong, China. (menghuang@ust.hk)

[2] 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)

[3] 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).

Abstract. In this paper, we develop a new computational approach which is based on minimizing the difference of two convex functions (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 subproblems. For each subproblem, the Nesterov's accelerated gradient descent algorithm or the Barzilai-Borwein (BB) algorithm is adoped. In addition, we apply the alternating projection method to improve the initial guess in [21] and makes it much more closer to the true solution. In the setting of sparse phase retrieval, a standard ℓ_1 norm term is added to guarantee the sparsity, and the subproblem is solved approximately by a proximal gradient method with the shrinkage-threshold technique directly. Furthermore, a modified Attouch-Peypouquet technique is used to accelerate the iterative computation, which leads to more effective algorithms than the Wirtinger flow (WF) algorithm and the Gauss-Newton (GN) algorithm and etc.. Indeed, DC based algorithms are able to recover the solution with high probability when the measurement number $m \approx 2n$ in the real case and $m \approx 3n$ in the complex case, where n is the dimension of the true solution. When $m \approx n$, the ℓ_1-DC based algorithm is able to recover the sparse signals with high probability. Our main results show that the DC based methods converge to a critical point linearly. 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. Finally, the paper discusses the existence and the number of distinct solutions for phase retrieval problem.

Keywords: phase retrieval; sparse signal recovery; DC methods; nonlinear least squares; non-convex analysis

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, 29, 30, 32, 36, 19, 16]. In particular, an explanation of the image recovery from the phaseless measurements and a survey of recent research results can be found in [26]. Mathematically, the phaseless retrieval problem or simply called phase retrieval problem can be stated as follows. Given measurement vectors $\mathbf{a}_i \in \mathbb{R}^n$ (or $\in \mathbb{C}^n$), $i = 1, \ldots, m$ and the measurement values $b_i \geq 0, i = 1, \cdots, m$, 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:

$$b_1 = |\langle \mathbf{a}_1, \mathbf{x} \rangle|^2, \ldots, b_m = |\langle \mathbf{a}_m, \mathbf{x} \rangle|^2. \tag{1}$$

Noting that for any constant $c \in \mathbb{R}^n$ (or $\in \mathbb{C}^n$) with $|c| = 1$, it holds $|\langle \mathbf{a}_i, c\mathbf{x} \rangle|^2 = |\langle \mathbf{a}_i, \mathbf{x} \rangle|^2$ for all i. Thus we can only hope to recover $\mathbf{x}$ up to a unimodular constant. We say the measurements $\mathbf{a}_1, \cdots, \mathbf{a}_m$ are generic if $A = (\mathbf{a}_1, \ldots, \mathbf{a}_m)$ corresponds to a point in a non-empty Zariski open subset of $\mathbb{R}^{n\times m}$ (or $\mathbb{C}^{n\times m}$). Also, $b_1, \cdots, b_m$ are essential if there exist n values $b_{j_1}, \cdots, b_{j_n}$ are all positive. One fundamental problem in phase retrieval is to give the minimal m for which there exists $A = (\mathbf{a}_1, \ldots, \mathbf{a}_m)$ can recover $\mathbf{x}$ up to a unimodular constant. For the real case, it is well known that the minimal measurement number m is $2n-1$ (cf. [4]). For the complex case $\mathbb{C}^n$, this question remains open. Conca, Edidin, Hering and Vinzant [14] proved $m \geq 4n-4$ generic measurements $\mathbf{a}_1, \ldots, \mathbf{a}_m$ 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 [39], for the case $n = 4$, Vinzant present $11 = 4n - 5 < 4n - 4$ measurement vectors which have phase retrieval property for $\mathbb{C}^4$. It implies that $4n-4$ is not sharp for some dimension n. Similar results about the minimal measurement number for sparse phase retrieval can be found in [40].

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 $\mathbf{x}$ from any given vector $\mathbf{b} = (b_1, \cdots, b_m)^\top$, e.g. a perturbation of the exact observations $\mathbf{b}^*$. We shall give this fact a mathematical explanation (see Theorem 1 in the next section). Thus, the phase retrieval problem is usually formulated as follows:

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

Although it is not a convex minimization problem, the objective function is differentiable. Hence, many computational algorithms can be developed and they are very successful actually. A gradient descent method (called Wirtinger flow in the complex case) is developed by Candès et al. in [12]. They show that the Wirtinger flow algorithm converges to the true signal up to a global phase factor with high probability provided $m \geq O(n \log n)$ Gaussian measurements. Lately, many variants of Wirtinger flow algorithms were developed, such as Thresholded WF[9], Truncated WF [13], Reshaped WF [46], and Accelerated WF [8] etc.. In [21], Gao and Xu propose a Gauss-Newton (GN) algorithm to find a minimizer of (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 guess. Indeed, Gao and Xu also provide a initialization procedure which is much better than the initialization algorithm given in [12] numerically. Another approach to minimize (2) is called the trust region method which was studied in [37], and the geometric analysis of the landscape function $f(\mathbf{x}) = \sum_{i=1}^m (|\langle \mathbf{a}_i, \mathbf{x}\rangle|^2 - b_i)^2$ is also given. To recover sparse signals from the measurements (1), a standard approach is adding the ℓ_1 term $\lambda\|\mathbf{x}\|_1$ to (2) or using the proximal gradient method as discussed in [35].

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

$$b_i = f(\langle \mathbf{a}_i, \mathbf{x}\rangle), \quad i = 1, \ldots, m, \tag{3}$$

where $f : \mathbb{C} \to \mathbb{R}_+$ is a twice differentiable convex function and 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. For another example, we can take $f(x) = |x|^4$ and etc.. To guarantee the unique recovery of $\mathbf{x}$, it has been proved that the number of measurements satisfies $m \geq n+1$ for the real case ($2n+1$ for the complex case, respectively) (see Theorem 2.1 in [25]). Recovering $\mathbf{x}$ from the nonlinear observation is also raised in many areas, such as neural networks (cf. [6], [33]).

To reconstruct $\mathbf{x}$ from (3), it is standard to formulate it as

$$\min_{\mathbf{x}\in\mathbb{R}^n \text{ or } \mathbb{C}^n} \sum_{i=1}^m (f(\langle \mathbf{a}_i, \mathbf{x}\rangle) - b_i)^2. \tag{4}$$

We approach it by using the standard technique for a difference of convex minimizing functions. 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 function. As it is not convex, we then write it as

$$F(\mathbf{x}) = \sum_{i=1}^m (f(\langle \mathbf{a}_i, \mathbf{x}\rangle) - b_i)^2 := F_1(\mathbf{x}) - F_2(\mathbf{x}), \tag{5}$$

where $F_1(\mathbf{x}) = \sum_{i=1}^m f^2(\langle \mathbf{a}_i, \mathbf{x}\rangle) + b_i^2$ and $F_2(\mathbf{x}) = \sum_{i=1}^m (2b_i f(\langle \mathbf{a}_i, \mathbf{x}\rangle))$. Note that f is a convex function with $f(x) \geq 0$ for all $x \in \mathbb{R}$. Then F_1 and F_2 are convex functions. The minimization (4) will be approximated by

$$\mathbf{x}^{(k+1)} := \arg\min_{\mathbf{x}} F_1(\mathbf{x}) - \nabla F_2(\mathbf{x}^{(k)})^\top(\mathbf{x} - \mathbf{x}^{(k)}) \tag{6}$$

for any given $\mathbf{x}^{(k)}$. We call this algorithm as DC based algorithm following from the ideas in [22], where the sparse solutions of under-determined linear system were studied. Although DC based algorithms have been studied for a long time (see e.g. [38], [43], [42] and the references therein), this is the first time to use a DC based algorithm to solve phase retrieval problem and achieve the best numerical performance compared to others methods from the knowledge of the authors.

The above minimization (6) is a convex problem with differentiable function for each k. We solve it by using the standard gradient descent method with Nestrov's acceleration (cf. [31]) or the Barzilai-Borwein (BB) method (cf. [5]). There are several nice properties of this DC based approach. We can show that

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

for some constant $\ell > 0$. That is, $F(\mathbf{x}^{(k)}), k \ge 1$ is strictly decreasing sequence. Furthermore, 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 show $\|\mathbf{x}^{(k)} - \mathbf{x}^*\| \le C\tau^k$ for $\tau \in (0,1)$. If the function $F(\mathbf{x})$ has the property that any local minimizer $\mathbf{x}^\star$ is a global 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 linearly. Actually, the function $F(\mathbf{x})$ indeed has such a property for real phase retrieval problem and such initial point can be obtained based on the initialization scheme discussed in [21]. Our numerical experiments show that the DC based algorithm can recover the true solutions when $m \approx 2n$ in the real case and $m \approx 3n$ in the complex case. See §6 for our numerical simulations.

Furthermore, we develop an ℓ_1-DC based algorithm to recover sparse signals. That is, starting from $\mathbf{x}^{(k)}$, we solve

$$\mathbf{x}^{(k+1)} := \arg\min \lambda\|\mathbf{x}\|_1 + F_1(\mathbf{x}) - \nabla F_2(\mathbf{x}^{(k)})^\top(\mathbf{x} - \mathbf{x}^{(k)}) \tag{7}$$

using a proximal gradient method, where $\lambda > 0$ is a parameter. The convergence of the ℓ_1-DC based algorithm can be established similar to the DC based algorithm. To accelerate the convergence of the ℓ_1-DC based algorithm, we use Attouch-Peypouquet's acceleration method (cf. [3]). To have a better initialization, we use the projection technique (cf. [17]). 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 show that the modified ℓ_1-DC based algorithm can recover sparse signals as long as $m \approx n$ provided $s \ll n$, where s and n are the sparsity and dimension of signals.

In summary, to establish the convergence of the DC based algorithms, we follow the well-known approach based on the Kurdyka-Łojasiewicz inequality (cf. [1], [2], [44], and [42]). Due to the nice properties of F_1 and F_2 in the setting of phase retrieval, we are able to specify the exponent θ in the Kurdyka-Łojasiewicz function and hence, the rate of convergence is precisely given, see Theorem 4. More precisely, for phase retrieval in the real case, we will show that F is strongly convex at the minimizer due to the positive definiteness of the Hessian in Appendix. In the complex case, the Hessian is no long positive definite but nonnegative definite near the minimizers. For sparse phase retrieval, we are no longer able to determine θ. Thus in order to establish the convergence rate in this settings, we break the neighborhood of the minimizers into two parts: within the ball or outside the ball of the given tolerance. Note that it is easy to check if an iterative point $\mathbf{x}^{(k+1)}$ is within the ball or not by checking the minimal value $F(\mathbf{x}^{(k+1)}) \leq \epsilon$ or not. For the iterative points outside the ϵ-ball, we establish the convergence rate; for the iterative points within the ball, we no longer need to consider it, see Theorem 7.

1.3 Organization

The paper is organized as follows. Firstly, using tools of algebraic geometry, we give the existence of solutions in phase retrieval problem 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 algorithms. Accelerated gradient descent methods including Nesterov's and Attouch-Peypouquet's accelerated techniques as well as the BB technique for inner iterations will be discussed in Section 4. Furthermore, we will study the ℓ_1-DC based algorithm for recover sparse signals and discuss the convergence in Section 5. Our numerical experiments are collected in Section 6, where we give the performance of our DC based algorithms and compare it with the Gauss-Newton algorithm for general signals and sparse signals. Particularly, we show that the DC based algorithm is able to recover signals when $m \approx 2n$. In addition, our ℓ_1-DC based algorithm with the update techniques is able to recover sparse signals when $m \approx n$.

2 On Existence and Number of Phase Retrieval Solutions

In this section, we shall discuss the existence of solution for phase retrieval and give an estimate for the number of distinct solutions. To beginning, we first recall PhaseLift (cf. [10]) which shows the connection between phase retrieval and low-rank matrix recovery.

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

$$b_j = \mathrm{tr}(A_j X), \quad j = 1, \ldots, m, \tag{8}$$

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

Note that the 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 with 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 $\mathcal{M}_r = \{X \in \mathbb{R}^{n\times n} : \text{rank}(X) = r\}$. In mathematical formulation, it aims to solve the following low rank matrix recovery problem:

$$\min \ \text{rank}(X) \quad \text{s.t.} \quad \text{tr}(A_jX) = b_j,\ j = 1,\dots,m \quad \text{and} \quad X \succeq 0. \tag{9}$$

As we will show in Theorem 1, for given $b_j \geq 0$, $j = 1,\dots,m$ there may not exist a matrix $X \in \mathcal{M}_r$ with $r < n$ satisfying the constraint conditions exactly unless b_j are exactly the measurement values from a matrix X. Thus to find the solution X, we reformulate the above problem as follows:

$$\min \sum_{i=1}^{m} |\text{tr}(A_jX) - b_j|^2 \quad \text{s.t.} \quad X \in \mathcal{M}_r \quad \text{and} \quad X \succeq 0. \tag{10}$$

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

$$\sum_{i=1}^{m} |\text{tr}(A_jX) - b_j|^2 \to \infty \text{ when } \|X\|_F \to \infty. \tag{11}$$

In the case that the above coercive condition does not hold, one has to use other conditions to ensure that the minimizer of (10) is bounded. For example, if there is a matrix X_0 which is orthogonal to A_j in the sense that $\text{tr}(A_jX_0) = 0$ for all $j = 1,\dots,m$, 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 solution for phase retrieval problem. Let $\mathcal{M}_r$ be the set of $n \times n$ matrices with rank r and $\overline{\mathcal{M}_r}$ be the set of all matrices with rank no more than r. It is known that dimension of $\mathcal{M}_r$ is $2nr - r^2$ (cf. Proposition 12.2 in [23]). Since $\overline{\mathcal{M}_r}$ is the closure of $\mathcal{M}_r$ in the Zariski topology (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 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. (cf. Example I.1.4 in [24]). □

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

$$\mathcal{A}(X) = (\text{tr}(A_1X), \cdots, \text{tr}(A_mX))^\top.$$

Given the map $\mathcal{A}$, we define the range $\mathcal{R}_+ = \{\mathcal{A}(X) : X \in \overline{\mathcal{M}_1}, X \succeq 0\}$ and the range $\mathcal{R} = \{\mathcal{A}(X) : X \in \overline{\mathcal{M}_1}\}$. It is clear that the dimension of $\mathcal{R}_+$ is less than or equal to the dimension of $\mathcal{R}$. Since each entry $\mathrm{tr}(A_j X)$ of the map $\mathcal{A}$ is a linear polynomial about the entries of X, then the map $\mathcal{A}$ is a regular. 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, a randomly choosing 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 $\mathcal{R}$ with probability one 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. To this end, we first recall the following result from Theorem 1.25 in Sec 6.3 of [34].

Lemma 2. *Let $f : X \to Y$ be a regular map and X, Y are irreducible varieties with $\dim(X) = n$ and $\dim(Y) = m$. If f is surjective, then $m \le n$. Furthermore, it holds:*

(a) for any $y \in Y$ and for any component F of the fiber $f^{-1}(y)$, $\dim(F) \ge n-m$;
(b) 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 1. *If one randomly chooses 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 satisfying the minimization (9) is zero. In other words, for almost all vectors $\mathbf{b} = (b_1, \cdots, b_m)^\top \in \mathbb{R}^m_+$ the solution of (9) is a matrix with rank more than or equal to 2.*

Proof. Let $X = \overline{\mathcal{M}_1}$ and $Y = \{\mathcal{A}(M), M \in \overline{\mathcal{M}_1}\}$. From Lemma 1, we know X is an irreducible variety. Since Y is the continuous image of the irreducible variety $\overline{\mathcal{M}_1}$, it is also an irreducible variety. Note that $\mathcal{A}$ is a regular map. By Lemma 2, we have $\dim(Y) \le \dim(\overline{\mathcal{M}_1}) = 2n-1 < m$. Thus, Y is a proper lower dimensional closed subset in $\mathbb{R}^m$. For almost all points in $\mathbb{R}^m$, they do not belong to Y. 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 $M \succeq 0$ such that $\mathcal{A}(M) = \mathbf{b}$. □

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 [41] does not have a solution for randomly chosen $\mathbf{b} = (b_1, \cdots, b_m) \in \mathbb{R}^m$ with probability one.

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

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

Here, a set S is said to be zero dimensional if the dimension of real points is zero for $S \subset \mathbb{R}^n$ or the dimension of complex points is zero for $S \subset \mathbb{C}^n$. As we are working over the fields like $\mathbb{R}$ or $\mathbb{C}$, it means that if the fiber is 0-dimensional, then it has only finite number of real or complex points. For those $\mathbf{b} \in \mathbb{R}_+^m$ with $\chi_{\mathbf{b}} \neq \emptyset$, we are interested in the upper bound of the number of solutions which satisfy the minimization (9). To do so, we need more results from algebraic geometry.

Lemma 3 ([23] Proposition 11.12.). *Let X be a quasi-projective variety and $\pi : X \to \mathbb{R}^m$ be 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 π 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 μ the minimum value of $\mu(p)$ on X_0, then*

$$\dim(X_0) = \dim(Y_0) + \mu. \tag{12}$$

As we have shown in the proof of Theorem 1, we have $\dim(\mathcal{R}) \leq \dim(\overline{\mathcal{M}_r})$. Next, we give a more precise characterization about these dimensions.

Lemma 4. *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 $\dim(\overline{\mathcal{M}_r}) = \dim(\mathcal{R})$. From Lemma 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 3 by setting $X = \overline{\mathcal{M}_r}$, $Y = \mathcal{A}(\overline{\mathcal{M}_r})$ and $\pi = \mathcal{A}$. (To apply this 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 (12) above, we have $\dim(\overline{\mathcal{M}_1}) = \dim(\mathcal{R})$. ☐

Finally, we need the following definition.

Definition 1. *The* degree *of an affine or projective variety with dimension k is the number of intersection points with k hyperplanes in general position.*

It has been shown [20, Example 14.4.11] that the 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

$$\prod_{i=0}^{n-2} \frac{n+i}{1+i}. \tag{13}$$

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

Theorem 2. *Given a vector* $\mathbf{b} \in \mathbb{R}_+^m$ *lies in the range* $\mathcal{R}_+$. *Assume that* $\chi_{\mathbf{b}} \neq \emptyset$. *Then the number of distinct solutions in* $\chi_{\mathbf{b}}$ *is less than or equals to* $\prod_{i=0}^{n-2} \frac{n+i}{1+i}$.

Proof. For any fixed $\mathbf{b}$, the matrices M which satisfy $\mathcal{A}(M) = \mathbf{b}$ and $\mathrm{rank}(M) = 1$ are exactly the intersection points of the variety $\overline{\mathcal{M}_1}$ with m hyperplanes, namely the hyperplanes defined by equations $\langle A_i, M\rangle = b_i, i = 1, \cdots, m$. Since $m > \dim(\overline{\mathcal{M}_r}) = 2n - 1$, the number of intersection points would be less than degree of $\overline{\mathcal{M}_1}$ generically. So, the number of positive semidefinite matrices M which satisfy $\mathcal{A}(M) = \mathbf{b}$ and $\mathrm{rank}(M) = 1$ would be no more than the degree of $\overline{\mathcal{M}_1}$. Finally, using the exact formula for the degree from (13), the result follows.

3 A DC based Algorithm for Phase Retrieval

For convenience, we simply discuss the case where $\mathbf{x}$ and $\mathbf{a}_j,\ j = 1, \ldots, m$ are real. The complex case can be treated in the same way from the algorithmic perspective. Indeed, when $\mathbf{x} \in \mathbb{C}^n$ and $\mathbf{a}_j \in \mathbb{C}^n$, $j = 1, \ldots, m$, we only need to write $\mathbf{x} = \mathbf{x}_R + \sqrt{-1}\mathbf{x}_I$ and similar for $\mathbf{a}_j$. Letting $\mathbf{u} = [\mathbf{x}_R^\top\ \mathbf{x}_I^\top]^\top \in \mathbb{R}^{2n}$, we view $F_1(\mathbf{x})$ as the function $G_1(\mathbf{u}) = F_1(\mathbf{x}_R + \sqrt{-1}\mathbf{x}_I)$. Then $G_1(\mathbf{u})$ is a convex function of real variable $\mathbf{u}$. Similarly, $G_2(\mathbf{u}) = F_2(\mathbf{x}_R + \sqrt{-1}\mathbf{x}_I)$ is a convex function of $\mathbf{u}$. The difference is that we can on longer recover $\mathbf{u}$ up to a sign but up to a orthogonal matrix. That is, for any orthogonal matrix $O \in \mathbb{R}^{2n\times 2n}$, the vector $O\mathbf{u}$ is also the true solution.

Recall that we aim to recover $\mathbf{x}$ by minimizing $F(\mathbf{x}) = F_1(\mathbf{x}) - F_2(\mathbf{x})$ in (4). It is easy to see that the minimization can happen in a bounded region $\mathcal{R}$ due to the coercive condition $f(x) \to \infty$ when $x \to \infty$. Our DC based method is given as follows. Start from any iterative solution $\mathbf{x}^{(k)}$, we solve the following convex minimization problem:

$$\mathbf{x}^{(k+1)} = \arg\min_{\mathbf{x}\in\mathbb{R}^n} F_1(\mathbf{x}) - \nabla F_2(\mathbf{x}^{(k)})^\top(\mathbf{x} - \mathbf{x}^{(k)}) \tag{14}$$

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

Our goal in this section is to show the sequence $\{\mathbf{x}^{(k)}\}_{k=1}^\infty$ converges to a critical point. Later, we will discuss how to find a global minimization by choosing a good initial guess $\mathbf{x}^{(1)}$ appropriately. For a fixed $\mathbf{x}^{(k)}$, there are many standard iterative methods to solve the convex minimization problem (14), such as the gradient descent method with various acceleration techniques. After getting $\mathbf{x}^{(k+1)}$ from solving (14), we update $\mathbf{x}^{(k)}$ with $\mathbf{x}^{(k+1)}$ and then solve (14) again. Hence, there are two iterative procedures. The iterative procedure for solving (14) is an inner iteration which will be discussed in the next section. In this section, we mainly discuss the outer iteration assuming $\mathbf{x}^{(k+1)}$ has been found.

We will state the following assumptions on functions F_1 and F_2:

(1) The gradient function ∇F_1 has Lipschitz constant L_1 in bounded region $\mathcal{R}$. That is, $\|\nabla F_1(\mathbf{x}) - \nabla F_1(\mathbf{y})\|_2 \le L_1\|\mathbf{x}-\mathbf{y}\|_2$ for all vectors $\mathbf{x},\mathbf{y} \in \mathcal{R}$.
(2) F_2 is a strongly convex function with parameter ℓ in $\mathcal{R}$. That is, $F_2(\mathbf{y}) \ge F_2(\mathbf{x}) + \nabla F_2(\mathbf{x})^\top(\mathbf{y}-\mathbf{x}) + \frac{\ell}{2}\|\mathbf{y}-\mathbf{x}\|^2$ for all vectors $\mathbf{x},\mathbf{y} \in \mathcal{R}$.

Observe that the function $F_2 = 2\sum_{i=1}^m b_i f(\mathbf{a}_i^\top \mathbf{x})$. Through a simple calculation, the Hessian matrix of function F_2 is

$$H_{F_2} = \sum_{i=1}^m 2b_i f^{''}(\mathbf{a}_i^\top \mathbf{x})\mathbf{a}_i\mathbf{a}_i^\top,$$

where $f^{''}(x) \ge 0$ since the convexity of f. Thus the parameter ℓ of strong convexity corresponds to the minimal eigenvalue of H_{F_2}. In fact, we can prove that in the case of the standard phase retrieval problem with $f(x) = x^2$, the function F_2 is strongly convex under a standard assumption that the measurement vectors are generic and measurement values are essential, see Theorem 8 in the Appendix. Similar result also holds for the general phase retrieval problem with $f(x) = |x|^4$.

We first start with a standard result for our DC based algorithm:

Theorem 3. *Assume F_2 is a strongly convex function with parameter ℓ. Starting from any initial guess $\mathbf{x}^{(1)}$, let $\mathbf{x}^{(k+1)}$ be the solution of (14) for all $k \ge 1$. Then*

$$F(\mathbf{x}^{(k+1)}) \le F(\mathbf{x}^{(k)}) - \frac{\ell}{2}\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|^2, \quad \forall k \ge 1. \tag{15}$$

Furthermore, it holds $\nabla F_1(\mathbf{x}^{(k+1)}) - \nabla F_2(\mathbf{x}^{(k)}) = 0$.

Proof. By the strongly convexity of F_2, we have

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

Recall that $F(\mathbf{x}) = F_1(\mathbf{x}) - F_2(\mathbf{x})$. Combining with (14), we obtain that

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

Since $\mathbf{x}^{(k+1)}$ is the minima of (14), the property $\nabla F_1(\mathbf{x}^{(k+1)}) - \nabla F_2(\mathbf{x}^{(k)}) = 0$ follows from the first order optimality condition directly. □

Next, we use the Kurdyka-Łojasiewicz (KL) inequality to establish the convergence rate of $\mathbf{x}^{(k)}$. The applications which use the KL inequality to solve various minimization problems can be found in [1, 2, 44, 42]. The following is our major theorem in this section.

Theorem 4. *Suppose that* $F(\mathbf{x}) = F_1(\mathbf{x}) - F_2(\mathbf{x})$ *is a real analytic function. Assume the gradient function* ∇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 (14) 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 limit, then*

$$\|\mathbf{x}^{(k+1)} - \mathbf{x}^*\| \leq C\tau^k \tag{16}$$

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

To prove the theorem, we need the following KL inequality which is central to the global convergence analysis.

Definition 2 (Łojasiewicz [28]). *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*

$$|f(\mathbf{x}) - f(\bar{\mathbf{x}})|^\theta \leq C\,dist(0, \partial f(\mathbf{x}))$$

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 a function* $\varphi(s) = cs^{1-\theta}$ *with* $\theta \in [0, 1)$ *such that it holds*

$$\varphi'(|f(\mathbf{x}) - f(\bar{\mathbf{x}})|)\,dist(0, \partial f(\mathbf{x})) \geq 1 \tag{17}$$

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, which the inequality (17) holds in a critical point with $\theta \in [1/2, 1)$. Later, many extensions of the above inequality are proposed. Typically, the extension to the setting of *o*-minimal structure in [27] is a general version. Recently, the KL inequality is extended to nonsmooth subanalytic functions. For our proof in the setting of phase retrieval, we need to specify $\theta = 1/2$. Indeed, we shall include an elementary proof to justify that our choice of $\theta = 1/2$ can be achieved. To show this, we need the following proposition.

Proposition 1. *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 number* δ *such that*

$$|f(\mathbf{x}) - f(\mathbf{x}^*)|^\theta \leq C\|\nabla f(\mathbf{x})\|, \quad \forall \mathbf{x} \in B(\mathbf{x}^*, \delta), \tag{18}$$

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

Proof. Since f is continuous and twice differentiable, using Taylor formula 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, due to he fact the Hessian is invertible, we have

$$\|\nabla f(\mathbf{x})\| = \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{x}^*)\| \geq c_2\|\mathbf{x} - \mathbf{x}^*\|.$$

Combining the above two estimates, we obtain (18) with $\theta = 1/2$ and $C = \sqrt{c_1}/c_2$. □

The importance of the Łajosiewicz inequality is to establish the inequality (18) under the case where f may not have an invertible Hessian at the critical point $\mathbf{x}^*$. However, for our phase retrieval problem, the Hessian matrix is restricted strong convex at the global minimizer (cf. [37]). In the real case, we can even show that the Hessian is positive definite at a global minimizer, see Theorem 9 in the Appendix. We are now ready to establish Theorem 4.

Proof of Theorem 4. As we have shown in Theorem 3,

$$\frac{\ell}{2}\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|^2 \leq F(\mathbf{x}^{(k)}) - F(\mathbf{x}^{(k+1)}). \tag{19}$$

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. It means that 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 positive constant C_1 such that

$$C_1\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\| \leq \sqrt{F(\mathbf{x}^{(k)}) - F(\mathbf{x}^*)} - \sqrt{F(\mathbf{x}^{(k+1)}) - F(\mathbf{x}^*)} \tag{20}$$

holds for all $k \geq k_0$ where k_0 is large enough. To establish this claim, we shall use the Proposition 1. Firstly, we prove that the condition $\nabla F(\mathbf{x}^*) = 0$ holds. Indeed, from Theorem 3 we have

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

Combining with (19), 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, Theorem 9 shows that F has a positive definite Hessian near $\mathbf{x}^*$. Thus the Kurdyka-Lojasiewicz inequality holds for $\theta = 1/2$ by Proposition 1. Consider the function $g(t) = \sqrt{t}$ which is concave over $[0, 1]$ and hence, $g(t) - g(s) \geq g'(t)(t - s)$. From the Kurdyka-Lojasiewicz inequality, there exists a positive constant $c_0 > 0$ and $\delta > 0$ such that

$$\|g'(F(\mathbf{x}) - F(\mathbf{x}^*))\nabla F(\mathbf{x}))\| \geq c_0 > 0 \tag{21}$$

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

$$\max\left(\sqrt{2/\ell}, L_1/(\ell c_0)\right) \cdot \sqrt{F(\mathbf{x}^{(k)}) - F(\mathbf{x}^*)} \leq \delta/2. \tag{22}$$

Recall that $\mathbf{x}^{(k_i)} \to \mathbf{x}^*$ as $k_i \to \infty$. Without loss of generality, we may assume that $k_0 = 1$ and $\mathbf{x}^{(1)} \in B(\mathbf{x}^*, \delta/2)$. We next show that $\mathbf{x}^{(k)} \in B(\mathbf{x}^*, \delta)$ for all $k \geq 1$ and prove it by induction. By (22) 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 $g'(F(\mathbf{x}^{(k)}) - F(\mathbf{x}^*))$ on both sides of (19), we have

$$\begin{aligned} &\frac{\ell}{2}\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|^2 g'(F(\mathbf{x}^{(k)}) - F(\mathbf{x}^*)) \\ &\leq g'(F(\mathbf{x}^{(k)}) - F(\mathbf{x}^*)) \left(F(\mathbf{x}^{(k)}) - F(\mathbf{x}^{(k+1)})\right) \\ &\leq \sqrt{F(\mathbf{x}^{(k)}) - F(\mathbf{x}^*)} - \sqrt{F(\mathbf{x}^{(k+1)}) - F(\mathbf{x}^*)}, \end{aligned} \tag{23}$$

where the last inequality follows from the concavity of g. On the other hand, combining the KL inequality (21) with Theorem 3, we have

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

Putting it in (23), we obtain

$$\sqrt{F(\mathbf{x}^{(k)}) - F(\mathbf{x}^*)} - \sqrt{F(\mathbf{x}^{(k+1)}) - F(\mathbf{x}^*)} \geq \frac{\ell c_0}{2L_1}\|\mathbf{x}^{(k)} - \mathbf{x}^{(k+1)}\| \tag{24}$$

for all $2 \leq k \leq K$. Taking the sum, it follows that

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

Finally, observe that

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

where the last inequality follows from (22). Thus $\mathbf{x}^{(K+1)} \in B(\mathbf{x}^*, \delta)$, which means that all $\mathbf{x}^{(k)}$ are in $B(\mathbf{x}^*, \delta)$ and inequality (24) holds for all k. Hence, we arrive at the claim (20) with $C_1 = \ell c_0/(2L_1)$. By summing the inequality (20) 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.

We next turn to prove the second part. Let $S_k = \sum_{i=k}^{\infty} \|\mathbf{x}^{(i+1)} - \mathbf{x}^{(i)}\|$. It follows from (24) that

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

Recall from (24) 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, we obtain

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

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

Remark 1. We should point out that the assumptions on F, F_1 and F_2 in Theorem 4 are easy to satisfy. For example, in standard phase retrieval all these assumptions are satisfied, especially when the region $\mathcal{R}$ is sufficiently small and near the global minimization by a technical initialization. More details can be found in Theorems 8 and 9.

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 (5). Then

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

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 1. 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 shown that one can use the initialization from [13, 21] 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 measurements $m \geq O(n)$, then with high probability we have

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

where $\mathbf{x}^*$ is a global minimizer and δ is a sufficient positive constant. Furthermore, in a small neighborhood $N(\mathbf{x}^\star, \delta) := \{\mathbf{x} : \|\mathbf{x} - \mathbf{x}^*\|_2 \leq \delta \|\mathbf{x}^*\|_2\}$, the minimizing function $F(x)$ is strongly convex [37]. Thus, our algorithm can converge to the global minimizer by using a good initialization.

4 Computation of the Inner Minimization (14)

We now discuss how to compute the minimization (14). For convenience, we rewrite the minimization in the following form

$$\min_{\mathbf{x}\in\mathbb{R}^n} G(\mathbf{x}) \tag{25}$$

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:

$$\mathbf{z}^{(j+1)} = \mathbf{z}^{(j)} - h\nabla G(\mathbf{z}^{(j)}) \tag{26}$$

for $j \in \mathbb{N}$ with $\mathbf{z}^{(1)} = \mathbf{x}^{(k)}$, where $h > 0$ is the step size. It is well-known that if we choose $h \approx 1/(2L)$ where L is the Lipschitz constant of $G(\mathbf{x})$, the gradient descent method (26) will have a linear convergence. It has also been shown that if we choose $h = \nu/L$ with Lipschitz constant L and the strong convex parameter ν, the Nestrov's acceleration technique will speed up the convergence rate. Some results are given as follows.

Lemma 5 (The Nesterov's Acceleration ([31])). *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a ν-strong convex function and the gradient function has L-Lipschitz constant. Start from an arbitrary initial point $\mathbf{u}_1 = \mathbf{z}_1$, the following Nesterov's accelerated gradient descent*

$$\begin{aligned}\mathbf{z}^{(j+1)} &:= \mathbf{u}^{(j)} - \frac{\nu}{L}\nabla f(\mathbf{u}^{(j)}),\\ \mathbf{u}^{(j+1)} &= \mathbf{z}^{(j+1)} - q(\mathbf{z}^{(j+1)} - \mathbf{z}^{(j)})\end{aligned} \tag{27}$$

satisfies

$$f(\mathbf{z}^{(j+1)}) - f(\mathbf{z}^*) \le \frac{\nu + L}{2}\|\mathbf{z}^{(1)} - \mathbf{z}^*\|^2 \exp\left(-\frac{j}{\sqrt{L/\nu}}\right), \tag{28}$$

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

The role of the Nestrov's acceleration is to reduce the number of iterations in (26) significantly. That is, for any tolerance ϵ, 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 (27) is used.

Since G is twice differentiable, we can certainly use the Newton method to solve (14) because of its quadratic convergence. However, we will not pursue it here due to the fact that when the dimension of $\mathbf{z}$ is large, the Newton method will be extremely slow. Instead, we apply the Barzilai-Borwein(BB) method to choose a good h, which is an excellent approach for the large scale minimization problem (cf. [5]). The iteration of the BB method can be described as

$$\mathbf{z}^{(j+1)} = \mathbf{z}^{(j)} - \beta_j^{-1}\nabla G(\mathbf{z}^{(j)}), \tag{29}$$

where the step size

$$\beta_j = (\mathbf{z}^{(j)} - \mathbf{z}^{(j-1)})^\top(\nabla G(\mathbf{z}^{(j)}) - \nabla G(\mathbf{z}^{(j-1)}))/\|\mathbf{z}^{(j)} - \mathbf{z}^{(j-1)}\|^2. \tag{30}$$

Algorithm 1 The BB Algorithm for the Inner Minimization

Let $\mathbf{u}^{(1)} = \mathbf{z}^{(1)}$ be an initial guess.
For $j \geq 1$, we solve the minimization of (25) by computing β_j according to (30).
Update

$$\begin{aligned} \mathbf{z}^{(j+1)} &:= \mathbf{u}^{(j)} - \beta_j^{-1} \nabla G(\mathbf{u}^{(j)}) \\ \mathbf{u}^{(j+1)} &= \mathbf{z}^{(j+1)} - q(\mathbf{z}^{(j+1)} - \mathbf{z}^{(j)}) \end{aligned} \tag{31}$$

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

Our computation of inner minimization is described in Algorithm 1, which is a combination of the BB technique with Nesterov's acceleration technique. The intuition behind it based on Lemma 5. Since BB method has a good performance in numerical experiment, we can hope our Algorithm 1 has better performance.

There are several modified versions of the BB method available with their convergence analysis in the literature, see, e.g. [15, 47] and the references therein. Although a large number of numerical experiments show that the BB method has excellent performance, however, the convergence rate is still not established yet for general minimizing function F. We next give some necessary and sufficient conditions to show why the Algorithm 1 has a better convergence rate. To this end, we say a algorithm is convergent superlinearly if

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

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

Lemma 6. *Suppose that the function $G(\mathbf{x})$ in (25) is α-strongly convex and the gradient has Lipschitz constant L in a domain D. Let $\mathbf{u}^* \in D$. Assume the sequence $\{\mathbf{u}^{(k)}, k \geq 1\}$ is obtained from the BB method and remains in D. Then $\{\mathbf{u}^{(k)}, k \geq 1\}$ converges super-linearly to $\mathbf{u}^*$ if and only if $(\beta_k - H_G(\mathbf{u}^*))\mathbf{s}_{k+1} = o(\|\mathbf{s}_{k+1}\|)$.*

Proof. From BB update rule (29), we have

$$\begin{aligned} (\beta_k - H_G(\mathbf{x}^*))\, \mathbf{s}_{k+1} &= -\nabla G(\mathbf{u}^{(k)}) - H_G(\mathbf{u}^*)\mathbf{s}_{k+1} \\ &= \nabla G(\mathbf{u}^{(k+1)}) - \nabla G(\mathbf{u}^{(k)}) - H_G(\mathbf{u}^*)\mathbf{s}_{k+1} - \nabla G(\mathbf{u}^{(k+1)}). \end{aligned} \tag{32}$$

Since the Hessian matrix $H_G(\mathbf{u})$ is continuous at $\mathbf{u}^*$ and all $\mathbf{u}^{(k)} \in D$, then we have

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

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

$$\lim_{k\to\infty} \frac{||\nabla G(\mathbf{u}^{(k+1)})||}{||\mathbf{s}_{k+1}||} = 0. \tag{33}$$

Note that

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

and

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

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

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

where $\sigma_k = \dfrac{||\mathbf{u}^{(k+1)} - \mathbf{u}^*||}{||\mathbf{u}^{(k)} - \mathbf{u}^*||}$. 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 $(\beta_k - H_G(\mathbf{u}^*))\mathbf{s}_{k+1} = o(\|\mathbf{s}_{k+1}\|)$. In fact, if $\mathbf{u}^{(k)} \to \mathbf{u}^*$ super-linearly, then we have

$$\lim_{k\to+\infty} \frac{\|\mathbf{u}^{(k+1)} - \mathbf{u}^{(k)}||}{||\mathbf{u}^{(k)} - \mathbf{u}^*||} = 1. \tag{34}$$

Indeed, since

$$\Big|||\mathbf{u}^{(k+1)} - \mathbf{u}^{(k)}|| - ||\mathbf{u}^{(k)} - \mathbf{u}^*||\Big| \le ||\mathbf{u}^{(k+1)} - \mathbf{u}^*||,$$

it is clear that

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

Hence, from (34) it follows

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

because of the denominator is bounded by the property (34). Using the argument at the beginning of the proof, we can see that $(\beta_k - H_G(\mathbf{u}^*))\mathbf{s}_{k+1} = o(\|\mathbf{s}_{k+1}\|)$. 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 (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). We want to recover $\mathbf{x_b}$ using the DC based algorithm. Firstly, we consider the following optimization

$$\min_{\mathbf{x}\in\mathbb{R}^n \text{ or } \mathbb{C}^n} \lambda\|\mathbf{x}\|_1 + \sum_{i=1}^m (f(\langle \mathbf{a}_i, \mathbf{x}\rangle) - b_i)^2, \tag{35}$$

which is a standard approach in compressive sensing by adding $\lambda\|\mathbf{x}\|_1$ to (2). If we take $f(\langle \mathbf{a}_i, \mathbf{x}\rangle) = |\langle \mathbf{a}_i, \mathbf{x}\rangle|^2$, then (35) becomes as the sparse phase retrieval. See [6] and [35] for recent literature on sparse phase retrieval problem.

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 $\sum_{i=1}^m (f(\langle \mathbf{a}_i, \mathbf{x}\rangle) - b_i)^2$ to be the difference of $F_1(\mathbf{x}) - F_2(\mathbf{x})$ as in (5). The minimization (35) will be approximated by

$$\mathbf{x}^{(k+1)} := \arg\min \lambda\|\mathbf{x}\|_1 + F_1(\mathbf{x}) - \nabla F_2(\mathbf{x}^{(k)})^\top (\mathbf{x} - \mathbf{x}^{(k)}) \tag{36}$$

for any given $\mathbf{x}^{(k)}$. We call this algorithm as sparse DC based method. For the general convex function f, we can also obtain the minimization problem as in (36) with the similar formulation. For convenience, we only consider the case when $\mathbf{x},\ \mathbf{a}_j,\ j = 1, \ldots, m$ are real. The complex case can be treated in the same way.

To solve (36), we use the proximal gradient method: for any given $\mathbf{y}^{(k)}$, we update it by

$$\begin{aligned}\mathbf{y}^{(k+1)} := \operatorname{argmin} \lambda\|\mathbf{y}\|_1 + F_1(\mathbf{y}^{(k)}) + (\nabla F_1(\mathbf{y}^{(k)}) - \nabla F_2(\mathbf{y}^{(k)}))^\top (\mathbf{y} - \mathbf{y}^{(k)}) \\ + \tfrac{L_1}{2}\|\mathbf{y} - \mathbf{y}^{(k)}\|^2\end{aligned} \tag{37}$$

for $k \geq 1$, where L_1 is the Lipschitz differentiability of F_1. This is a typical DC algorithm discussed in [42]. 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's technique. The discussion above furnishes a computational method for sparse phase retrieval problem (35). Let us point out one significant difference between update rule (37) and (14) is that one can find $\mathbf{y}^{(k+1)}$ by using a explicit formula while the solution $\mathbf{x}^{(k+1)}$ of (14) 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. To the best of the authors' knowledge, the convergence is not available in the literature so far. We first start with a standard result for the ℓ_1-DC based algorithm.

Theorem 5. *Assume F_2 is a strongly convex function with parameter ℓ. Starting from any initial guess $\mathbf{y}^{(1)}$, let $\mathbf{y}^{(k+1)}$ be the solution of (37) for all $k \geq 1$. Then it holds*

$$\lambda\|\mathbf{y}^{(k+1)}\|_1 + F(\mathbf{y}^{(k+1)}) \leq \lambda\|\mathbf{y}^{(k)}\|_1 + F(\mathbf{y}^{(k)}) - \frac{\ell}{2}\|\mathbf{y}^{(k+1)} - \mathbf{y}^{(k)}\|^2, \quad \forall k \geq 1 \tag{38}$$

and

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

where $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_1$ *and* ∂g *denotes the subgradient of* g.

Proof. The Lipschitz differentiability of F_1 gives

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

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

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

Combing the above two inequalities and using used the first order optimality condition for (37), we obtain that

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

Letting $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_1$, the second property $\partial g(\mathbf{y}^{(k+1)}) + \nabla F_1(\mathbf{y}^{(k)}) - \nabla F_2(\mathbf{y}^{(k)}) + L_1(\mathbf{y}^{(k+1)} - \mathbf{y}^{(k)}) = 0$ follows from the minimization (37).

Next we show that the sequence $\mathbf{y}^{(k)}, k \ge 1$ from (37) converges to a critical point $\mathbf{y}^*$.

Theorem 6. *Write* $\mathcal{F}(\mathbf{x}) = \lambda\|\mathbf{x}\|_1 + \sum_{i=1}^m (f(\langle \mathbf{a}_i, \mathbf{x}\rangle) - b_i)^2$. *Suppose that* $f(\mathbf{x})$ *is a real analytic function and the gradient* $\nabla f(\mathbf{x})$ *has Lipschitz constant* L. *Let* $\mathbf{y}^{(k)}, k \ge 1$ *be the sequence obtained from (37). Then it converges to a critical point* $\mathbf{y}^*$ *of* F.

Proof. From Theorem 5, we have

$$\frac{\ell}{2}\|\mathbf{y}^{(k+1)} - \mathbf{y}^{(k)}\|^2 \le \mathcal{F}(\mathbf{y}^{(k)}) - \mathcal{F}(\mathbf{y}^{(k+1)}). \tag{39}$$

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

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

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 $\{\mathcal{F}(\mathbf{y}^{(k)})\}_{k=1}^{\infty}$ is a bounded monotonic descending sequence, and hence $\mathcal{F}(\mathbf{y}^{(k)}) \to \mathcal{F}(\mathbf{y}^*)$ for all $k \geq 1$. We claim that the sequence $\{\mathbf{y}^{(k)}\}_{k=1}^{\infty}$ has finite length, that is,

$$\sum_{k=1}^{\infty} \|\mathbf{y}^{(k+1)} - \mathbf{y}^{(k)}\| < \infty. \tag{40}$$

To establish the claim, we need to use the Kurdyka-Łojasiewicz inequality (cf. [27]). Note that the ℓ_1 norm $\|\mathbf{x}\|_1$ is semialgebraic function and the function $f(\mathbf{x})$ is analytic, so the objective function $\mathcal{F}(\mathbf{x})$ satisfies the KL property at any critical point (cf. [1], [2], [44]). Let us prove that $\|\nabla\mathcal{F}(\mathbf{y}^*)\| = 0$ holds, that is, $\mathbf{y}^*$ is a critical point of $\mathcal{F}$. Indeed, using the second property of (5), we have

$$\begin{aligned}\|\partial\mathcal{F}(\mathbf{y}^{(k)})\| &= \|\partial g(\mathbf{y}^{(k)}) + \nabla F_1(\mathbf{y}^{(k)}) - \nabla F_2(\mathbf{y}^{(k)})\| \\ &\leq \|\nabla F(\mathbf{y}^{(k)}) - \nabla F(\mathbf{y}^{(k-1)})\| + L_1\|\mathbf{y}^{(k)} - \mathbf{y}^{(k-1)}\|.\end{aligned} \tag{41}$$

Combining (39) with (41) and using the Lipschitz differentiation of F_1 and F_2, we obtain that $\|\partial F(\mathbf{y}^{(k_i)})\| \to 0$. By the property of subgradient of g and the continuity of the gradients F_1 and F_2, we have $\|\partial\mathcal{F}(\mathbf{y}^*)\| = 0$ when $\mathbf{y}^{(k_i)} \to \mathbf{y}^*$. Thus, $\mathbf{y}^* \in \text{domain}(\partial F)$, the set of all critical points of $\mathcal{F}$.

Therefore, we can use KL inequality to obtain that

$$\varphi'(\mathcal{F}(\mathbf{y}) - \mathcal{F}(\mathbf{y}^*))\|\partial\mathcal{F}(\mathbf{y})\| \geq 1 \tag{42}$$

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

$$\max\left(\sqrt{2/\ell}\sqrt{\mathcal{F}(\mathbf{y}^{(k)}) - \mathcal{F}(\mathbf{y}^*)}, 2C/\ell \cdot \varphi(\mathcal{F}(\mathbf{y}^{(k)}) - \mathcal{F}(\mathbf{y}^*))\right) \leq \delta/2. \tag{43}$$

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 $\mathbf{y}^{(k)}, k \geq 1$ will be in the neighborhood $B(\mathbf{y}^*, \delta)$. We shall use an induction to do so. By (43) we have

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

Assume that $\mathbf{y}^{(k)} \in B(\mathbf{y}^*, \delta)$ for $k \leq K$. From (5), we have

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

where constant $C := L + L_1/2$. Putting it into (42), it gives that

$$\varphi'(\mathcal{F}(\mathbf{y}^k) - \mathcal{F}(\mathbf{y}^*)) \geq \frac{1}{C\|\mathbf{y}^k - \mathbf{y}^{k-1}\|}. \tag{44}$$

On the other hand, from the concavity of φ we get that

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

Combining (39) with (44), we obtain

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

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

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

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

$$\frac{2C}{\ell}\varphi(\mathcal{F}(\mathbf{y}^{(1)}) - \mathcal{F}(\mathbf{y}^*)) \geq \sum_{j=1}^{K} \|\mathbf{y}^{(j+1)} - \mathbf{y}^{(j)}\| + \|\mathbf{y}^{(K+1)} - \mathbf{y}^{(K)}\|. \tag{45}$$

That is, we have

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

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 the above arguments, we know that the inequality (45) holds for all k, which implies the claim (40) holds. It is clear that (40) 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, we show that the convergence rate is linear. To beginning, we give the following technique lemma.

Lemma 7. *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 $\mathbf{y} \in B(\mathbf{x}, \delta)$, there exists a subgradient ∂g at $\mathbf{y}$ such that*

$$(\partial g(\mathbf{y}) - \partial g(\mathbf{x}))^{\top}(\mathbf{y} - \mathbf{x}) = 0. \tag{46}$$

Proof. For simplicity, we only 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 (46). 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 (46). □

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

$$\mathcal{R}_s := \{\mathbf{x} \in \mathbb{R}^n : \ \|\mathbf{x}\|_0 \leq s\} = \bigcup_{\substack{I \subset \{1, \cdots, n\} \\ |I| = s}} \mathbb{R}_I^s, \tag{47}$$

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

Lemma 8. *Let $\mathcal{F}(\mathbf{x}) = g(\mathbf{x}) + F(\mathbf{x})$ with $g(\mathbf{x}) = \lambda\|\mathbf{x}\|_1$. Suppose that F is L-Lipschitz differentiable. Let $\mathbf{x}^*$ be a critical point of $\mathcal{F}$ as explained in (6). Suppose that either all entries of $\mathbf{x}^*$ are nonzero or $\mathbf{x}^* \in \mathbb{R}^s_I$ for some $s \in \{1, \cdots, n\}$. Then there exists $\delta > 0$ such that for all $\mathbf{x} \in B(\mathbf{x}^*, \delta)$,*

$$|\mathcal{F}(\mathbf{x}) - \mathcal{F}(\mathbf{x}^*)| \leq C\|\mathbf{x} - \mathbf{x}^*\|^2. \tag{48}$$

Proof. Under the assumption that either all entries of $\mathbf{x}^*$ are nonzero or $\mathbf{x}^* \in \mathbb{R}^s_I$ for some $s \in \{1, \cdots, n\}$, we know that $\mathcal{F}(\mathbf{x})$ is differentiable at $\mathbf{x}^*$. Since $\mathbf{x}^*$ is a critical point, we have

$$\partial\mathcal{F}(\mathbf{x}^*) = \partial g(\mathbf{x}^*) + \nabla F(\mathbf{x}^*) = 0.$$

Combing it with (36), we obtain

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

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

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

Theorem 7. *Suppose that F_2 is strongly convex. Starting from any initial guess $\mathbf{x}^{(1)}$, let $\mathbf{x}^{(k+1)}$ be the solution of (14) for all $k \geq 1$. Then for any $\epsilon > 0$, there exists a point $\mathbf{x}^*$ such that either $\mathbf{x}^{(k+1)} \in B(\mathbf{x}^*, \epsilon)$ or*

$$\|\mathbf{x}^{(k+1)} - \mathbf{x}^*\| \leq C_\epsilon \tau^k \tag{49}$$

for a positive constant C_ϵ dependent on ϵ and $\tau \in (0, 1)$.

Proof. As we have shown in Theorem 5 and Theorem 6, the sequence $\mathbf{x}^{(k)}, k \geq 1$ converges to a critical point $\mathbf{x}^*$ of $\mathcal{F}$. Furthermore, we have

$$C_0\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|^2 \leq (\mathcal{F}(\mathbf{x}^{(k)}) - \mathcal{F}(\mathbf{x}^*)) - (\mathcal{F}(\mathbf{x}^{(k+1)}) - \mathcal{F}(\mathbf{x}^*)) \tag{50}$$

for a positive constant C_0. We now claim that

$$C_1\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\| \leq \sqrt{\mathcal{F}(\mathbf{x}^{(k)}) - \mathcal{F}(\mathbf{x}^*)} - \sqrt{\mathcal{F}(\mathbf{x}^{(k+1)}) - \mathcal{F}(\mathbf{x}^*)} \tag{51}$$

holds for a positive constant C_1. To establish this claim, we first note that Lemma 8 gives

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

Multiplying the above inequality to (50), we have

$$C_0 C \frac{\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|^2}{\|\mathbf{x}^{(k)} - \mathbf{x}^*\|} \le \frac{(\mathcal{F}(\mathbf{x}^{(k)}) - \mathcal{F}(\mathbf{x}^*)) - (\mathcal{F}(\mathbf{x}^{(k+1)}) - \mathcal{F}(\mathbf{x}^*))}{\sqrt{\mathcal{F}(\mathbf{x}^{(k)}) - \mathcal{F}(\mathbf{x}^*)}}. \tag{52}$$

Consider $h(t) = \sqrt{t}$ which is concave over $[0,1]$ and we know $h(t) - h(s) \ge h'(t)(t-s)$. Thus, the right-hand side above is less than or equal to the right-hand side of (51). We next show that the left-hand side of the inequality satisfies

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

Note that F is strongly convex outside the ball $B(\mathbf{x}^*, \epsilon)$. If $\mathbf{x}^{(k+1)}$ is within the $B(\mathbf{x}^*, \epsilon)$, then we complete the proof. Otherwise, the strong convexity of F outside $B(\mathbf{x}^*, \epsilon)$ (see Theorem 9 for the real case and Theorem 10 for the complex case) gives

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

for a positive constant dependent on ϵ. As $\partial g(\mathbf{x}^*) + \nabla F(\mathbf{x}^*) = 0$, Lemma 7 implies that $-\nabla F(\mathbf{x}^*) = \partial g(\mathbf{x}^*) = \partial g(\mathbf{x}^{(k+1)})$ when $\mathbf{x}^{(k+1)}$ is close to $\mathbf{x}^*$ enough (i.e., the support of $\mathbf{x}^{(k+1)}$ is the same as the support of $\mathbf{x}^*$ and the sign of each entry in $\mathbf{x}^{(k+1)}$ is the same to $\mathbf{x}^*$). By Theorem 5, we have

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

In other words, using the Lipschitz differentiability of F, it holds

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

and

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

The left-hand side of (52) can be simplified to be

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

which implies the claim (51) holds. By summing the inequality (51), it follows

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

Then the remaining part of the proof is similar to the proof of Theorem 4 and we leave the details to the interested readers.

6 Numerical Results

In this section, we give some numerical experiments for our DC based algorithm and the ℓ_1-DC based algorithm. We compare the empirical success rate of our DC based algorithms with WF [9] and Gauss-Newton [21] methods. All experiments are carried out with 1000 repeated trials. The results show that the DC based algorithm is able to recover real signals with probability around 80% with $m \approx 2n$ measurements, where n is the dimension of signals. As demonstrated in [13], more precisely the Figures 8 and 9 in [13], it needs $m \approx 3n$ measurements to recover real signals for truncated Wirtinger flow and Wirtinger flow algorithms. Similarly, the DC based algorithm can recover the complex signals with $m \approx 3n$ measurements (cf. Table 2) while WF requires $m \approx 4n$. Finally, for the sparse signals, the ℓ_1-DC based algorithm with thresholding technique needs only $m \approx n$ measurements.

6.1 Phase Retrieval for Real and Complex Signals

Example 2. In this example, we consider to recover a real signal $\mathbf{x_b}$ from the given measurements (1) using Gaussian random measurement vectors $\mathbf{a}_j, j = 1, \cdots, m$. We fix $n = 128$ and consider the number of measurements m is around the twice of the dimension of $\mathbf{x_b}$, i.e., $m = kn/16$ for $k = 24, 25, \cdots, 35$. For the initialization, we first obtain a initial guess by the initialization algorithm in [21] and then improve the initial guess by applying alternating projection method discussed in Algorithm 3. We say a trail is successful if the relative error is less than 10^{-5}. Table 1 gives the empirical success rate of recovering $\mathbf{x_b}$ for DC, WF and Gauss-Newton methods. From Table 1, we can see that the DC based algorithm can recover the solutions with probability large than 60% under $m \geq 2n$. According to the result in [4], one needs $m \geq 2n - 1$ measurements to guarantee the recovery of all real signals. Thus the DC based algorithm almost reaches the theoretical low bound.

Table 1. The numbers of successes over 1000 repeated trials versus the number of measurements m/n listed above.

m/n	1.5	1.5625	1.6250	1.6875	1.75	1.8125	1.875	1.9375	2	2.0625	2.125	2.1875
WF successes	0	0	0	0	0	1	11	10	24	27	41	64
GN successes	0	0	0	18	13	36	71	114	167	251	315	415
DC successes	50	78	119	182	266	318	406	542	600	681	744	807

Example 3. We next repeat Example 2 using a litter more number of measurements. The numbers of successes for the Wirtinger Flow algorithm, Gauss-Newton algorithm and the DC based algorithm are shown in Table 2. One can

Table 2. The numbers of successes over 1000 repeated trails versus the number of measurements m/n listed above, where $n = 128$.

m/n	2.4375	2.5	2.5625	2.625	2.6875	2.75	2.8125	2.875	2.937	3
WF successes	168	220	254	352	372	459	513	612	641	706
GN successes	728	749	844	886	908	934	931	963	968	982
DC successes	944	952	975	982	984	989	994	993	995	998

see that the performance of the DC based algorithm is the best and can achieve 95% success rate with $m = 2.5n$.

Example 4. This example is to show the robustness of the DC based algorithm. We repeat the computation in Example 2 for noisy measurements. There are two ways to generate the noisy measurements. One way is to add the noises η_j to b_j directly and obtain

$$\hat{b}_j = |\langle \mathbf{a}_j, \mathbf{x_b} \rangle|^2 + \eta_j, \quad j = 1, \cdots, m. \tag{53}$$

Another way is

$$\tilde{b}_j = |\langle \mathbf{a}_j, \mathbf{x_b} \rangle + \delta_j|^2 + \eta_j, \quad j = 1, \cdots, m, \tag{54}$$

where δ_j and η_j are noises. For noisy measurements (53), we assume that η_j are subject to uniform random distribution between $[-\beta, \beta]$ with mean zero, where $\beta = 10^{-1}$, 10^{-3} and 10^{-5}. If the stopping tolerance ϵ satisfies $\epsilon \geq \beta$, then the Gauss-Newton method and DC based method produce the same empirical success rate as in Table 2. For noisy measurements (54), we assume that both ϵ_j and δ_j are subject to uniform distribution between $[-\beta, \beta]$ with mean zero. Similarly, if the stopping tolerance ϵ satisfies $\epsilon \geq \beta$, then both algorithms can recover the true solution.

Example 5. In this example, we use the DC based algorithm and the Gauss-Newton method to recover the complex signals. We choose $n = 128$ and the number of measurements m is around $3n$, i.e., $m \approx 3n$. For Gaussian random measurements $\mathbf{a}_j = \mathbf{a}_{j,R} + \mathbf{i}\mathbf{a}_{j,I}$, $j = 1, \cdots, m$, we aim to recover $\mathbf{z} \in \mathbb{C}^n$ with $\mathbf{z} = \mathbf{x} + \mathbf{i}\mathbf{y}$ from $|\langle \mathbf{a}_j, \mathbf{z} \rangle|^2, j = 1, \cdots, m$. The maximum iteration numbers for WF, GN, DC are $3,000$, 100 and 1000, respectively. We say a trial is successful if the relative error is less than 10^{-5}. Table 3 gives the numbers of successes for WF, Gauss-Newton and the DC based methods with 1000 repeated trials. From the Table 3, we can see that the DC based algorithm can recover the complex signals very well when $m \geq 3n$, which is slightly better than the GN algorithm and much better than the WF algorithm.

We next present some numerical experiments to demonstrate that the ℓ_1-DC based algorithm works well. The procedure is presented in Algorithm 2, where a modified Attouch-Peypouquet technique [3] is used. We use the step size $\beta_k = k/(k + \alpha)$ for the first few k iterations, say $k \leq K$, and then a fixed step size β_K for the remaining iterations.

Table 3. The numbers of successes over 1000 repeated trials based on m/n listed above for complex case.

m/n	2.938	3	3.062	3.125	3.187	3.25	3.312	3.375	3.437	3.5	3.562	3.625	3.687	3.75
WF	0	0	0	0	0	0	0	0	0	56	192	204	322	401
GN	191	338	304	416	452	536	594	739	744	762	801	815	910	912
DC	422	563	537	565	623	730	829	887	881	894	954	946	981	986

Algorithm 2 ℓ_1-DC based Algorithm

We use the same initialization as in the previous examples.
while $k \geq 1$ **do**
  1° Solve (37) to get $\mathbf{y}^{(k+1)}$.
  2° 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 $\mathbf{y}^T$

Example 6. In this example, we test the performance of Algorithm 2 for recovering the real signals without sparsity. We choose $n = 128$ and the number of measurements $m = 1.1n, 1.2n, \ldots, 2.5n$. The target signal $\mathbf{x_b}$ and the measurement vectors $\mathbf{a}_j, j = 1, \ldots, m$ are Gaussian random vectors. We choose the parameter $\lambda = 10^{-5}$ in Algorithm 2. The numbers of successes are summarized in Table 4. The results show that the ℓ_1-DC based algorithm can recover the real signals with high probability if $m \geq 2.2n$.

Table 4. The numbers of successes over 1000 repeated trials based on Algorithm 2.

m/n	1.5	1.6	1.7	1.8	1.9	2	2.1	2.2	2.3	2.4	2.5
ℓ_1-DC alg.	0	0	206	317	352	557	724	913	938	947	994

6.2 Phase Retrieval of Sparse Signals

We next turn to consider how to use our ℓ_1-DC based algorithm to recover the sparse signals. We know that if the number of measurements $m \leq 2n$, then many existing algorithms will fail to recover the signals. For our DC based algorithm, it can recover any signal when $m \approx 2n$, no matter sparse or not. However, when $m \approx 1.5n$, we are not able to recover the general signals. The purpose of our numerical experiments is to see if we are able to recover the sparse signals when $m \approx n$. By using the sparsity, we will enhance the ℓ_1-DC based algorithm with the projection technique. More specifically, we use the hard thresholding technique to project $\mathbf{y}^{(k+1)}$ from (7) into the set of all s-sparse vectors. This leads to an ℓ_1-DC based algorithm with hard thresholding which given below.

Algorithm 3 ℓ_1-DC based Algorithm with Hard Thresholding

Obtain a initial guess with the initialization in [9].
while $k \geq 1$ **do**
  Solve (37) 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 of s-sparse set $\mathcal{R}_s$. That is, let $\mathbf{z}^{(k+1)}$ be the solution of the following minimization problem:

$$\sigma_s(\mathbf{x}^k) = \min_{\mathbf{z} \in \mathcal{R}_s} \|\mathbf{y}^{(k+1)} - \mathbf{z}\|_1. \tag{55}$$

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

Example 7. In this example, we show that our ℓ_1-DC based algorithm with hard thresholding works well. We choose $n = 128$ and the number of measurements $m = 0.5n, 0.6n, \ldots, 2n$. For each m, we test the performance of Algorithm 3 with sparsities $s = 2, 4, 5, 10, 20$. The experiments are implemented under 1000 repeated trials. The results on the numbers of successes are presented in Table 5. From Table 5, we can see that Algorithm 3 is able to recover sparse solutions with high probability.

Table 5. The numbers of successes with sparsities $s = 2, 4, 5, 10, 20$ over 1000 repeated trials.

m/n		0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2
Alg. 3	$s = 2$	422	517	566	619	662	710	745	829	846	837	833	849	862	870	877	946
Alg. 3	$s = 4$	187	333	429	474	590	673	693	747	778	811	819	832	840	862	873	919
Alg. 3	$s = 5$	71	116	264	383	452	594	618	674	726	771	802	821	837	858	869	894
Alg. 3	$s = 10$	0	0	0	52	151	247	416	385	537	482	590	680	701	737	796	812
Alg. 3	$s = 20$	0	0	0	0	0	55	156	180	227	271	368	422	451	527	574	599

7 Appendix

In this section we give some deterministic description of the minimizing function F as well as strong convexity of F_2. We will show that at any global minimizer, the Hessian matrix of F is positive definite in the real case and is nonnegative positive definite in the complex case. These results are used when we apply the KL inequality. For convenience, let $A_\ell = \mathbf{a}_\ell \bar{\mathbf{a}}_\ell^\top$ be the Hermitian matrix of rank one for $\ell = 1, \cdots, m$.

Definition 3. *We say* $\mathbf{a}_j, j = 1, \cdots, m$ *are* a_0*-generic if there exists a positive constant* $a_0 \in (0,1)$ *such that*

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

holds for any $1 \leq j_1 < j_2 < \cdots < j_n \leq m$.

Theorem 8. *Let* $m \geq n$. *Assume* $\mathbf{a}_j, j = 1, \cdots, m$ *are* a_0*-generic for some constant* a_0. *If there exist* n *nonzero elements among the measurements* $b_j, j = 1, \cdots, m$, *then for the phase retrieval problem with* $f(x) = |x|^2$, F_2 *is positive definite.*

Proof. Recall that $F_2 = 2\sum_{i=1}^m b_i f(\mathbf{a}_k^\top \mathbf{x})$. Then the Hessian matrix of F_2 is

$$H_{F_2}(\mathbf{x}) = 2\sum_{i=1}^m b_i f''(\mathbf{a}_k^\top \mathbf{x}) \mathbf{a}_i \mathbf{a}_i^\top.$$

Note that $f''(x) = 2$. Thus we have

$$H_{F_2}(\mathbf{x}) = 4\sum_{\ell=1}^m b_\ell A_\ell.$$

Let $b_0 = \min\{b_\ell \neq 0\}$. Then

$$\mathbf{y}^\top H_{F_2}(\mathbf{x})\mathbf{y} \geq 4b_0 \|(\mathbf{a}_{j_1}^* \mathbf{y}, \ldots, \mathbf{a}_{j_n}^* \mathbf{y})\|^2 \geq 4b_0 a_0^2 \|\mathbf{y}\|^2.$$

Thus, F_2 is strongly convex. □

Theorem 9. *Let* $H_F(\mathbf{x})$ *be the Hessian matrix of the function* $F(\mathbf{x})$ *and let* $\mathbf{x}^\star$ *be a global minimizer of (2). Suppose that* $\mathbf{a}_j, j = 1, \cdots, m$ *are* a_0*-generic. Then* $H_F(\mathbf{x}^\star)$ *is positive definite in a neighborhood of* $\mathbf{x}^*$.

Proof. Recall that $A_\ell = \mathbf{a}_\ell \bar{\mathbf{a}}_\ell^\top$ 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 $A_\ell = [a_{ij}(\ell)]_{ij=1}^n$. Since $(\mathbf{x}^*)^\top A_\ell \mathbf{x}^* = b_\ell$ for all $\ell = 1, \cdots, m$, the first summation term of h_{ij} above is zero at $\mathbf{x}^*$. Let $M(\mathbf{y}) = \mathbf{y}^\top H_f(\mathbf{x}^*)\mathbf{y}$ be a quadratic function of $\mathbf{y}$. Then we have

$$M(\mathbf{y}) = 4\sum_{\ell=1}^m (\mathbf{y}^\top A_\ell \mathbf{x}^* (\mathbf{x}^*)^\top A_\ell \mathbf{y} = 4\sum_{\ell=1}^m |\mathbf{y}^\top A_\ell \mathbf{x}^*|^2$$

$$= 4\sum_{\ell=1}^{m} |\mathbf{y}^\top \mathbf{a}_\ell|^2 |\bar{\mathbf{a}}_\ell^\top \mathbf{x}^*|^2 \geq 4a_0\|\mathbf{x}^*\|^2\|\mathbf{y}\|^2,$$

where the inequality follows from the fact that $\mathbf{a}_j, j = 1, \cdots, m$ are a_0-generic. It implies that $H_F(\mathbf{x}^*)$ is positive definite. □

Next, we show that the Hessian $H_F(\mathbf{x}^*)$ is nonnegative definite at the global minimizer $\mathbf{x}^\star$ in the complex case. To this end, we first introduce 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

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

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

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

and

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

Furthermore, the Hessian of F is given by

$$H_F(\mathbf{x}, \mathbf{y}) = [\nabla_{\mathbf{x}}\nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{y})\ \nabla_{\mathbf{x}}\nabla_{\mathbf{y}} f(\mathbf{x}, \mathbf{y});\ \nabla_{\mathbf{y}}\nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{y})\ \nabla_{\mathbf{y}}\nabla_{\mathbf{y}} f(\mathbf{x}, \mathbf{y})],$$

where the terms $\nabla_{\mathbf{x}}\nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{y}), \cdots, \nabla_{\mathbf{y}}\nabla_{\mathbf{y}} f(\mathbf{x}, \mathbf{y})$ are given below.

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

and

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

The terms $\nabla_{\mathbf{y}}\nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{y})$ and $\nabla_{\mathbf{y}}\nabla_{\mathbf{y}} f(\mathbf{x}, \mathbf{y})$ can be obtained similarly.

Theorem 10. *For phase retrieval problem in the complex case, the Hessian matrix $H_f(\mathbf{x}^*, \mathbf{y}^*)$ at any global minimizer $\mathbf{z}^* := (\mathbf{x}^*, \mathbf{y}^*)$ satisfies $H_f(\mathbf{x}^*, \mathbf{y}^*) \geq 0$. Furthermore, $H_f(\mathbf{x}^*, \mathbf{y}^*) = 0$ along the direction $[-(\mathbf{y}^*)^\top, (\mathbf{x}^*)^\top]^\top$.*

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

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

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

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

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

It means that $H_f(\mathbf{x}^*, \mathbf{y}^*) \geq 0$. Furthermore, if we choose $\mathbf{u} = -\mathbf{y}^*$ and $\mathbf{v} = \mathbf{x}^*$, then it is easy to show that

$$[-(\mathbf{y}^*)^\top \, (\mathbf{x}^*)^\top]^\top H_f(\mathbf{x}^*, \mathbf{y}^*)\, [-\mathbf{y}^*\,; \mathbf{x}^*] = 0,$$

which gives that the Hessian H_F along this direction is zero. □

References

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

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

[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] R. Balan, P. Casazza, and D. Edidin, On signal reconstruction without phase. *Applied and Computational Harmonic Analysis*, 20(3) (2006) 345–356.

[5] J. Barzilai and J. M Borwein, Two-point step size gradient methods, *IMA Journal of Numerical Analysis*, 8(1) (1988) 141–148.

[6] A. Beck and Y. C. Eldar, Sparsity constrained nonlinear optimization: Optimality conditions and algorithms, *SIAM Journal on Optimization*, 23(3) (2013) 1480–1509.

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

[8] E. Bostan, M. Soltanolkotabi, D. Ren, and L. Waller, Accelerated Wirtinger flow for multiplexed Fourier ptychographic microscopy, *2018 25th IEEE International Conference on Image Processing*, (2018) 3823–3827.

[9] T. T. Cai, X. Li, and Z. Ma, Optimal rates of convergence for noisy sparse phase retrieval via thresholded Wirtinger flow, *The Annals of Statistics*, 44(5) (2016) 2221-2251.

[10] E. J. Candes, Y. C. Eldar, T. Strohmer, and V. Voroninski, Phase retrieval via matrix completion. *SIAM Review*, 57(2) (2015) 225–251.

[11] E. J. Candes, X. Li, and M. Soltanolkotabi, Phase retrieval from coded diffraction patterns, *Applied and Computational Harmonic Analysis*, 39(2) (2015) 277–299.

[12] E. J. Candes, X. Li, and M. Soltanolkotabi, Phase retrieval via wirtinger flow: Theory and algorithms, *IEEE Transactions on Information Theory*, 61(4) (2015) 1985–2007.

[13] Y. Chen and E. J. Candes, Solving random quadratic systems of equations is nearly as easy as solving linear systems, *Communications on Pure and Applied Mathematics*, 70(5) (2017) 822-883.

[14] A. Conca, D. Edidin, M. Hering, and C. Vinzant, An algebraic characterization of injectivity in phase retrieval, *Applied and Computational Harmonic Analysis*, 38(2) (2015) 346–356.

[15] Y. H. Dai and L. Z. Liao, R-linear convergence of the Barzilai and Borwein gradient method, *IMA Journal of Numerical Analysis*, 22 (2002) 1–10.

[16] C. Dainty and J. R. Fienup, Phase retrieval and image reconstruction for astronomy, *Image Recovery: Theory and Application*, (1987), 231–275.

[17] F. R. Deutsch, Best approximation in inner product spaces, *Springer Science and Business Media*, (2012).

[18] Y. C. Eldar and S. Mendelson, Phase retrieval: Stability and recovery guarantees, *Applied and Computational Harmonic Analysis*, 36(3) (2014) 473–494.

[19] J. R. Fienup, Phase retrieval algorithms: a comparison, *Applied Optics*, 21(15) (1982) 2758–2769.

[20] W. Fulton, Intersection theory, *Springer Science and Business Media*, 2 2013.

[21] B. Gao and Z. Xu, Phaseless recovery using the gauss–newton method, *IEEE Transactions on Signal Processing*, 65(22) (2017) 5885–5896.

[22] 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*, (2013) 37–45.

[23] J. Harris, Algebraic geometry: a first course, *Springer Science and Business Media*, 133 (2013).

[24] R. Hartshorne, Algebraic geometry, *Springer Science and Business Media*, 52 (2013).

[25] M. Huang and Z. Xu, Phase retrieval from the norms of affine transformations, *arXiv:1805.07899*, (2018).

[26] K. Jaganathan, Y. C. Eldar, B. Hassibi, Phase Retrieval: An Overview of Recent Developments, *arXiv:1510.07713v1*, (2015).

[27] K. Kurdyka, On gradients of functions definable in o-minimal structures, *Annales de l'institut Fourier*, 48(3) (1998) 769–784.

[28] S. Łojasiewicz, Une propriété topologique des sous-ensembles analytiques réels, *Les équations aux dérivées partielles*, 117 (1963) 87–89.

[29] J. Miao, T. Ishikawa, B. Johnson, E. H. Anderson, B. Lai, and K. O. Hodgson, High resolution 3D X-Ray diffraction microscopy, *Phys. Rev. Lett.*, 89(8) (2002).

[30] R.P. Millane, Phase retrieval in crystallography and optics, *JOSA A*, 7(3) (1990) 394–411.

[31] Y. Nesterov, Introductory lectures on convex optimization: A basic course, *Springer Science and Business Media*, 87 (2013).

[32] W. H. Robert, Phase problem in crystallography, *Journal of the Optical Society of America A*, 10(5) (1993) 1046–1055.

[33] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, Learning representations by back-propagating errors, *Nature*, 323(6088) (1986) 533-536.

[34] I. R. Shafarevich, Basic algebraic geometry 1, *Springer-Verlag Berlin Heidelberg*, (2013).

[35] Y. Shechtman, A. Beck, and Y. C. Eldar, GESPAR: Efficient phase retrieval of sparse signals, *IEEE Transactions on Signal Processing*, 62(4) (2014) 928–938.

[36] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, Phase retrieval with application to optical imaging: A contemporary overview, *IEEE Signal Processing Magazine*, 32(3) (2015) 87–109.

[37] J. Sun, Q. Qu, and J. Wright, A Geometric Analysis of Phase Retrieval, *Foundations of Computational Mathematics*, 18(5) (2018) 1131–1198.

[38] P. D. Tao, The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems, *Annals of Operations Research*, 133 (2005) 23–46.

[39] C. Vinzant, A small frame and a certificate of its injectivity, *2015 International Conference on Sampling Theory and Applications*, (2015) 197–200.

[40] Y. Wang, Z. Xu, Phase Retrieval for Sparse Signals, *Applied Computational Harmonic Analysis*, 37 (2014) 531–544.

[41] Y. Wang and Z. Xu, Generalized phase retrieval: measurement number, matrix recovery and beyond, *Applied and Computational Harmonic Analysis*, 47(2) (2019) 423–446.

[42] B. Wen, X. Chen, and T. K. Pong, A proximal difference-of-convex algorithm with extrapolation, *Comput. Optim. Appl.*, 69(2) (2018) 297–324.

[43] C. Wu, C. Li, and Q. Long, A DC programming approach for sensor network localization with uncertainties in anchor positions, *J. Ind. Manag. Optim.*, 10(3) (2014) 817–826 .

[44] Y. Xu and W. 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) (2013) 1758–1789.

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

[46] H. Zhang and Y. Liang, Reshaped Wirtinger flow for solving quadratic systems of equations, *Advances in Neural Information Processing Systems*, (2016) 2622–2630.

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