GLOBALLY CONVERGENT LEVENBERG-MARQUARDT METHOD FOR PHASE RETRIEVAL

CHAO MA*, XIN LIU†, AND ZAIWEN WEN‡

Abstract. In this paper, we consider a nonlinear least squares model for the phase retrieval problem. Since the Hessian matrix may not be positive definite and the Gauss-Newton (GN) matrix is singular at any optimal solution, we propose a modified Levenberg-Marquardt (LM) method, where the Hessian is substituted by a summation of the GN matrix and a regularization term. Similar to the well-known Wirtinger flow (WF) algorithm under certain assumptions, we start from an initial point provably close to the set of the global optimal solutions. Global linear convergence and local quadratic convergence to the global solution set are proved by estimating the smallest nonzero eigenvalues of the GN matrix, establishing local error bound properties and constructing a modified regularization condition. The computational cost becomes tractable if a preconditioned conjugate gradient (PCG) method is applied to solve the LM equation inexactly. Specifically, the pre-conditioner is constructed from the expectation of the LM coefficient matrix by assuming the independence between the measurements and iteration point. Preliminary numerical experiments show that our algorithm is robust and it is often faster than the WF method on both random examples and natural image recovery.

Key words. Non-convex optimization, phase retrieval, Levenberg-Marquardt method, convergence to global optimum

AMS subject classification. 49N30, 49N45, 90C26, 90C30, 94A20

1. Introduction.

One popular formulation of the phase retrieval problem is solving a system of quadratic equations in the form

$$y_r = |\langle a_r, z\rangle|^2, \quad r = 1, 2, ..., m, \tag{1.1}$$

where $z \in \mathbb{C}^n$ is the decision variable, $a_r \in \mathbb{C}^n$ are known sampling vectors, $\langle a_r, z\rangle$ is the inner product between a_r and z in $\mathbb{C}^n$, $|a|$ is the magnitude of $a \in \mathbb{C}$, and $y_r \in \mathbb{R}$ are the observed measurements. This problem arises from many areas of science and engineering such as X-ray crystallography [25, 35], microscopy [34], astronomy [19], diffraction and array imaging [8, 10], and optics [43]. It also appears in a few other important fields, including acoustics [2, 3], blind channel estimation in wireless communications [1, 20], interferometry [13], quantum mechanics [11, 39] and quantum information [26].

Many algorithms have been developed to solve (1.1). One of the most widely used method is the error reduction algorithm derived by Gerchberg and Saxton [24] and Fienup [20, 21]. This approach has been extended as the hybrid input-output (HIO) algorithm proposed by Fienup [21]. Bauschke et. al. established a few connections between the ER and HIO algorithms and classical convex optimization methods in [4]. Based on these connections, they proposed the hybrid projection-refection (HPR) method in [5]. Luke further developed in [31] the relaxed averaged alternating reflection (RAAR) method which can often be more efficient and reliable than the HIO and HPR methods. The quadratic system (1.1) can be formulated as the following nonlinear least squares (NLS) problem:

$$\min_{z\in\mathbb{C}^n} \; f(z) = \frac{1}{2m}\sum_{r=1}^{m}\left(|a_r^* z|^2 - y_r\right)^2. \tag{1.2}$$

‡School of Mathematical Sciences, Peking University, Beijing, CHINA (machao_pku@pku.edu.cn).

†State Key Laboratory of Scientific and Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, CHINA (liuxin@lsec.cc.ac.cn). Research supported in part by NSFC grants 11101409, 11331012 and 91330115, and the National Center for Mathematics and Interdisciplinary Sciences, CAS.

‡Beijing International Center for Mathematical Research, Peking University, Beijing, CHINA (wenzw@pku.edu.cn). Research supported in part by NSFC grant 11322109 and by the National Basic Research Project under the grant 2015CB856002.

Wen et. al. introduced an alternating direction method of multipliers (ADMM) to solve (1.2) in [44] and showed that the ADMM is usually comparable to many existing methods for both classical and ptychographic phase retrieval problems. In [40], Yoav Shechtman et. al. proposed a damped Gauss-Newton scheme. Other approaches include the difference map (DF) algorithm developed by Elser [15] and the so-called saddle-point optimization algorithms developed by Machesini [32]. Netrapalli et. al designed alternating minimization methods in [37]. Although the methods mentioned above often perform well numerically, their convergence to the global optimal solutions is not clear, yet.

Recently, there are a few important progress on achieving the global optimality for solving nonconvex optimization problems. Minimizing a composite function with nonconvex sparse regularization term is studied in [46, 30, 17]. Sun and Luo proved in [42] that a first-order method converges to global optimality on a matrix completion problem. Candes et. al. proposed a so-called Wirtinger flow (WF) algorithm for solving the model (1.2) in [9]. The WF algorithm is consisted of two parts. An initial point z_0 is obtained from the leading eigenvector of a certain matrix, and the point is refined by a gradient descent scheme in the sense of Wirtinger calculus iteratively. When there are no noise involved in the measurements of (1.1), it is proved in [9] that the initialization step yields an initial point z_0 very close to the set of global optimal solutions with a high probability. Then it is showed that the WF algorithm converges to the global minimizer in a global linear rate. Since the computational cost of the each step of the WF algorithm is cheap, the numerical results seem to be practically useful.

In this paper, we propose a modified LM approach for solving the NLS model (1.2). In fact, numerical methods for the general NLS problems $\min \|r(z)\|^2$, where $r(z)$ are the residual functions, have been well studied for decades. The Gauss-Newton (GN) method calculates a search direction determined by a so-called GN matrix through the first-order information. Global convergence to a stationary point can be guaranteed after combining certain line search techniques. If the NLS has a zero residual at the global optimal solutions, the GN matrix equals to the Hessian at these points, which ensures the quadratic local convergence rate of the GN method. However, when the Hessian is singular at the solutions, the GN method may fail. Another widely used approach is the LM method [29, 33] by adding a regularization term to the GN matrix. The regularization parameter is usually updated adaptively in a fashion similar to the trust-region scheme [22]. The regularization term makes the LM method to conquer the singularity issue. Yamashita and Fukushima established quadratic convergence for singular problems satisfying certain error bound conditions when the regularization parameter is chosen to be $||r(z)||^2$ in [45]. Fan and Yuan [18] provided a more general analysis and extended the applicable regularization parameters to be a family $\mu_k = ||r(z_k)||^\delta$ with $\delta \in [1, 2]$. The readers are referred to [14, 28, 23, 36, 12, 7] and the reference therein, for other algorithms for NLS, including the structured quasi-Newton method.

Our main contribution is a practical linearly convergent LM method with a provable second-order local convergence rate. Our approach is divided into two stages. The first stage is an initialization procedure exactly the same as the WF method in [9]. The second stage is to update the iterate by an LM method where the regularization parameter is based on the residual norm, i.e., the objective function value $f(z)$ in (1.2). Since the Hessian is indefinite and calculating a positive definite correction to the Hessian may be expensive, it is reasonable to use the LM method rather than the modified Newton method. By estimating the smallest nonzero eigenvalues of the GN matrix, and establishing local error bound properties and a modified regularization condition, we are able to prove that our approach can achieve a globally linear convergence to the global solution set and attain a locally quadratic convergence rate with high probability. In particular, the region of quadratic convergence is estimated explicitly. In order to reduce the computational cost, the LM equation is solved

inexactly by the PCG method. The globally linear convergence to the global solution set is still ensured if the accuracy is proportional to the residual. We further construct a simple practical pre-conditioner using the expectation of the LM coefficient matrix by assuming that the measurements and iteration point are independent. Although the LM coefficient matrix tends to be singular close to the optimal solution, the PCG method still runs smoothly since all iterations are performed in an invariant subspace. Because the condition number of the preconditioned coefficient matrix in this subspace is small, the number of iterations of the PCG method can be controlled reasonably small. Consequently, the total computational cost becomes at least competitive to the WF method. Our numerical experiments illustrate that the inexact LM method indeed outperforms the WF method on both random examples and natural image recovery.

We notice that the authors of [16] show local quadratic convergence rate of the modified LM method under certain deterministic local error bound conditions. However, it is not clear how to verify if the original NLS problem (1.2) satisfies these local error bound conditions, and how to estimate an explicit neighborhood around the solution set where these local error bound holds. The difference is that we can prove the existence of certain local error bound condition in a neighborhood close to the solution set with high probability. Although this theoretically neighborhood may be quite small when the dimension n is large, our analysis is still meaningful for a second-type algorithm.

In the rest of this paper, we first give a brief description of the WF approach and its convergence properties in Section 2. Our proposed LM approach for the Gaussian model is introduced in Section 3. The theoretical analysis on the exact LM method is presented in Section 4. In Section 5, we establish the convergence of the inexact LM framework and construct a preconditioner for computing the LM direction. The algorithm is extended to the coded diffraction model and is analyzed in Section 6. Numerical experiments are reported in Section 7 to demonstrate the effectiveness and efficiency of our LM method.

2. Preliminary.

2.1. Problem Statement. We first introduce the Gaussian model for the choices of the sampling vectors.

ASSUMPTION 2.1. *A problem is called the Gaussian model if the sample vectors* $a_r \in \mathbb{C}^n \sim \mathcal{N}(0, I/2) + i\mathcal{N}(0, I/2)$*, where* $\mathcal{N}(\mu, \Sigma)$ *denotes a Gaussian distribution with mean* μ *and covariance* Σ*. It holds* $\|a_r\| \le \sqrt{6n}$ *for* $r = 1, 2, ..., m$*. There is no noise in the observation measurements. Namely, the global minimum of* (1.2) *is zero.*

Similar to the analysis in [9], the event $\|a_r\| \le \sqrt{6n}$ holds with probability no less than $1 - me^{-1.5n}$ in Assumption 2.1. During the theoretical analysis in this paper, we always make this assumption. Hence, $-me^{-1.5n}$ will always be a term of the probabilities in the main theorems.

Since the decision variable z of (1.2) is complex, we use the Wirtinger derivatives [38] to calculate the derivatives of the objective function. For any $z \in \mathbb{C}^n$, the complex conjugate of z is written as $\bar{z}$. For ease of notation, we define two augmented vectors in bold face as

$$\mathbf{z} = \begin{bmatrix} z \\ \bar{z} \end{bmatrix} \text{ and } \tilde{\mathbf{z}} = \begin{bmatrix} z \\ -\bar{z} \end{bmatrix}. \tag{2.1}$$

Then the objective function of (1.2) can be viewed as a function with respect to the variable $\mathbf{z}$, i.e.,

$$f(\mathbf{z}) = \frac{1}{2m} \sum_{r=1}^{m} \left(z^*(a_r a_r^*)\bar{z} - y_r\right)^2.$$

It follows from the calculation rules of the Wirtinger derivatives that the gradient is

$$g(\mathbf{z}) := \nabla f(\mathbf{z}) = \frac{1}{m}\sum_{r=1}^{m}\left(|a_r^* z|^2 - y_r\right)\begin{bmatrix} (a_r a_r^*)z \\ (\bar{a}_r a_r^\top)\bar{z} \end{bmatrix}. \tag{2.2}$$

For convenience, we denote $\nabla f(z) := \frac{1}{m}\sum_{r=1}^{m}\left(|a_r^* z|^2 - y_r\right)(a_r a_r^*)z$.

2.2. The WF Algorithm. We briefly review the WF algorithm in [9] in this subsection. The initial point is constructed from the eigenvector corresponding to the largest eigenvalue of a matrix $Y = \frac{1}{m}\sum_{r=1}^{m} y_r a_r a_r^*$. The detailed procedure is outlined in Algorithm 1. It is shown in [9] that the initialization procedure can generate a good approximation to the set of optimal solutions. In fact, let $x \in \mathbb{C}^n$ be an optimal solution to (1.2) and assume that x is independent of a_r. The expectation of Y is $\mathbb{E}Y = xx^* + \|x\|^2 I$, whose leading eigenvector is parallel to x. When m is sufficiently large, Y is close to its expectation so that the angle between x and the leading eigenvector of Y is small, and $n\frac{\sum_r y_r}{\sum_r \|a_r\|^2}$ is close to $\|x\|^2$.

Algorithm 1: Initialization in the WF method

1 Input measurements $\{a_r\}$ and observations $\{y_r\}$ $(r = 1, 2, \ldots, m)$.

2 Calculate z_0 to be the leading eigenvector of $Y = \frac{1}{m}\sum_{r=1}^{m} y_r a_r a_r^*$.

3 Normalize z_0 such that $\|z_0\|^2 = n\frac{\sum_r y_r}{\sum_r \|a_r\|^2}$.

Once an initial point z_0 is obtained, the WF method executes gradient descent steps via Wirtinger derivative using a restricted step size $\frac{\mu_k}{||z_0||^2}$:

$$z_{k+1} = z_k - \frac{\mu_k}{||z_0||^2}\nabla f(z_k). \tag{2.3}$$

The update of the conjugates $\{\bar{z}_k\}$ is omitted since it is equivalent to the calculation of $\{z_k\}$.

Let $x \in \mathbb{C}^n$ be an optimal solution to (1.2). For each $z \in \mathbb{C}^n$, the distance between x and z is measured as

$$\mathbf{dist}(z, x) = \min_{\phi\in[0,2\pi]} \|z - e^{i\phi}x\| = \|z\|^2 + \|x\|^2 - 2|z^* x|.$$

The next theorem shows the property of the initialization Algorithm 1 and the global linear convergence of the WF algorithm (2.3). When the number of measurements is sufficiently large, the spectral initialization can produce a good initial point. Consequently, by initiating from this point, a linear convergence can be achieved with high probability.

THEOREM 2.2. ***(Theorem 3.3 of [9])*** *Suppose that Assumption 2.1 holds. Let $x \in \mathbb{C}^n$ be any solution of* (1.2)*, $m \geq c_0 n \log n$, where c_0 is a sufficiently large constant. Then the initial estimate z_0 normalized to have a squared Euclidean norm equal to $m^{-1}\sum_r y_r$, obeys*

$$\mathbf{dist}(z_0, x) \leq \frac{1}{8}\|x\| \tag{2.4}$$

with probability at least $1 - 10e^{-\gamma n} - 8/n^2$ (γ is a fixed positive constant). Let $\{z_k\}$ be a sequence generated by (2.3) *starting from any initial solution z_0 obeying* (2.4) *with $\mu_k = \mu \leq c_1/n$ for all k and some fixed constant c_1. Then there is an event of probability at least $1 - 13e^{-\gamma n} - me^{-1.5m} - 8/n^2$, such that on this event, we have*

$$\mathbf{dist}(z_k, x) \leq \frac{1}{8}\left(1 - \frac{\mu}{4}\right)^{k/2}||x||. \tag{2.5}$$

3. A Modified LM Method. The WF algorithm is essentially a gradient descent method with a restricted step size. Since the model (1.2) is a NLS problem, it is natural to consider the LM method for a faster local convergence rate than the WF method. Using the calculation rules of the Wirtinger derivatives, we obtain the Jacobian and GN matrix of $f(\mathbf{z})$:

$$J(\mathbf{z}) := \frac{1}{\sqrt{m}} \begin{bmatrix} |a_1^* z| a_1, & |a_2^* z| a_2, & \cdots, & |a_m^* z| a_m \\ |a_1^* z| \bar{a}_1, & |a_2^* z| \bar{a}_2, & \cdots, & |a_m^* z| \bar{a}_m \end{bmatrix}^*, \tag{3.1}$$

$$\Psi(\mathbf{z}) := J(\mathbf{z})^* J(\mathbf{z}) = \frac{1}{m} \sum_{r=1}^{m} \begin{bmatrix} |a_r^* z|^2 a_r a_r^* & (a_r^* z)^2 a_r a_r^\top \\ (a_r^* z)^2 \bar{a}_r a_r^* & |a_r^* z|^2 \bar{a}_r a_r^\top \end{bmatrix}. \tag{3.2}$$

The LM direction $\mathbf{s}_k$ is calculated by solving the following linear system

$$\Psi_{z_k}^{\mu_k} \mathbf{s}_k = -g(\mathbf{z}_k), \tag{3.3}$$

where $\mu_k \geq 0$ and $\Psi_z^\mu = \Psi(\mathbf{z}) + \mu I$. Then the iteration scheme of the LM algorithm is

$$\mathbf{z}_{k+1} = \mathbf{z}_k + \mathbf{s}_k. \tag{3.4}$$

The role of the parameter μ_k is important. It can be updated similar as the strategies in the classic trust-region type algorithms. For the sake of theoretical analysis, we propose the following updating rules for the Gaussian model:

$$\mu_k = \begin{cases} 70000n\sqrt{nf(z_k)}, & \text{if } f(z_k) \geq \frac{1}{900n}\|z_k\|^2; \\ \sqrt{f(z_k)}, & \text{otherwise.} \end{cases} \tag{3.5}$$

Roughly speaking, when the residual is large and the iteration is far away from the optimal solution set, the larger parameter $\mu_k = 70000n\sqrt{nf(z_k)}$ can guarantee a global linear convergence. As long as the residual becomes small enough, the choice of $\mu_k = \sqrt{f(z_k)}$ adapted from [45, 18] ensures a fast local convergence rate.

To further improve the efficiency of the LM algorithm in practice, the equation (3.3) can be solved inexactly after reaching certain criterion, such as

$$\|\Psi_{z_k}^{\mu_k} \mathbf{s}_k + g(\mathbf{z}_k)\| \leq \eta_k \|g(\mathbf{z}_k)\| \tag{3.6}$$

for some constant $\eta_k \geq 0$. With a suitably chosen parameter η_k, a global linear convergence rate of the LM method can be guaranteed while a better numerical performance than the exact LM method can be achieved.

The framework of the exact and inexact LM method are unified in Algorithm 2.

Algorithm 2: An Modified LM method for Phase Retrieval

1 **Input:** Measurements $\{a_r\}$, observations $\{y_r\}$. Set $\epsilon \geq 0$.
2 Construct an initial guess z_0 using Algorithm 1. Set $k := 0$.
3 **while** $||g(\mathbf{z}_k)|| \geq \epsilon$ **do**
4 Compute $\mathbf{s}_k$ by solving (3.3) with μ_k specified in (3.5) until (3.6) is satisfied.
5 Set $\mathbf{z}_{k+1}$ by (3.4) and $k := k + 1$.
6 **Output:** z_k.

Similar to the WF method, the calculation involving the conjugates of $\{z_k\}$ is not necessary. As we will describe later in Section 5.2, the LM equation (3.3) can be solved by the PCG method which only consists of a series of vector summations and matrix-vector

multiplications. It allows us to calculate s_k without considering its conjugate. Therefore, the computational cost and storage are reduced. Nevertheless, for convenience of theoretical analysis, we still deal with matrices in $\mathbb{C}^{2n\times 2n}$ and treat variables in $\mathbb{C}^{2n}$.

We should mention that the GN and Newton methods are not used because of singularity issues. Note that the NLS (1.2) admits a zero residual at an optimal solution under Assumption 2.1. The GN matrix $\Psi(\mathbf{z})$ equals to the Hessian at this solution and they are mostly singular. Consequently, Newton and GN methods cannot be employed directly. The modified Newton method is not practical either because the Hessian is indefinite and it is often intractable to calculate a suitable regularization parameter. Our modified LM method whose parameter μ_k tending to zero conquers the singularity issue and ensures a local quadratic convergence.

4. Analysis of the Exact LM Method. In this section, we analyze the convergence of our LM algorithm with $\eta_k = 0$ in (3.6). The main result consists of two parts. When $f(z_k) \geq \frac{1}{900n}\|z_k\|^2$ holds, our modified LM algorithm can achieve a globally linear convergence with high probability. Otherwise, it implies $\mathbf{dist}(z_k, x) \leq \frac{1}{14\sqrt{n}}||x||$ and guarantees a quadratic convergence rate with high probability.

Our main result on Gaussian model is stated as follows.

THEOREM 4.1. *Suppose that Assumption 2.1 holds. Let $x \in \mathbb{C}^n$ be any solution of (1.2) and $m \geq c_0 n \log n$, where c_0 is a sufficiently large constant. Let $\{z_k\}$ be a sequence generated by Algorithm 2 where the LM equation exactly solved. Then, starting from any initial solution z_0 satisfying $\mathbf{dist}(z_0, x) \leq \frac{1}{8}\|x\|$, there is an event of probability at least $1 - 15e^{-\gamma n} - 8/n^2 - me^{-1.5n}$ (γ is a fixed positive constant), such that on this event,*

$$\mathbf{dist}(z_{k+1}, x) < c_1 \mathbf{dist}(z_k, x), \qquad \text{for all } k = 0, 1, ... \tag{4.1}$$

where

$$c_1 := \begin{cases} \left(1 - \frac{||x||}{4\mu_k}\right), & \text{if } f(z_k) \geq \frac{1}{900n}\|z_k\|^2; \\ \frac{4.28+5.56\sqrt{n}}{9.89\sqrt{n}}, & \text{otherwise.} \end{cases} \tag{4.2}$$

Furthermore, there exists a sufficiently large integer l satisfying $f(z_l) < \frac{1}{900n}\|z_l\|^2$. Consequently, it holds for all $k \geq l$ that

$$\mathbf{dist}(z_{k+1}, x) < c_2 \mathbf{dist}(z_k, x)^2, \tag{4.3}$$

where

$$c_2 := \frac{4.28 + 5.56\sqrt{n}}{\|x\|}. \tag{4.4}$$

The lower bound of the probability of convergence in Theorem 4.1 is of the same order as that of Theorem 2.2 although the constant γ is different. When n is sufficiently large, $e^{-\gamma n}$ and $me^{-1.5n}$ becomes negligible compared to the term $1/n^2$. Then the probabilities in Theorems 2.2 and 4.1 tend to be equal. Since the method is monotone, and according to the selection of the parameter μ_k, the coefficient c_1 is uniformly bounded above by $1 - \frac{||x||}{4\mu_0} = 1 - \frac{||x||}{2.8\times 10^5 n\sqrt{nf(z_0)}}$ and tends to $\frac{4.28+5.56\sqrt{n}}{9.89\sqrt{n}}$, which is a constant less than 1. In this sense, our linear convergence rate is no worse than the WF method.

One advantage of our modified LM method is its locally quadratic convergence property. It cannot be derived directly from the analysis for the deterministic problems in [18] from two main perspectives: i) we admit a more relaxed region where the local error bound properties hold; ii) the neighborhood of provable quadratic convergence can be estimated specifically.

4.1. Lemmas for the Proof. Let $X^* \subset \mathbb{C}^n$ be the set of optimal solutions of (1.2) and the letter $x \in \mathbb{C}^n$ be reserved for a solution of (1.2). We first prove Theorem 4.1 in the case $\|x\| = 1$. In the end, we complete the proof by showing that the case $\|x\| \neq 1$ can be reduced to the case $\|x\| = 1$.

When z is independent to $\{a_r\}$, it is easily verified that $\mathbb{E}\Psi(\mathbf{z}) = \Phi(\mathbf{z})$, where

$$\Phi(\mathbf{z}) = \begin{bmatrix} zz^* + z^*zI & 2zz^\top \\ 2\bar{z}z^* & \bar{z}z^\top + z^\top \bar{z}I \end{bmatrix}. \tag{4.5}$$

Although the LM iterates $\{z_k\}$ are not independent to the measurements $\{a_r\}$, the relationship between $\Psi(\mathbf{z})$ and $\Phi(\mathbf{z})$ still plays an important role in our theoretical analysis. For convenience of notation, we also use $\Phi_z^\mu = \Phi(\mathbf{z}) + \mu I$ hereafter.

The first lemma describes the concentration of the GN matrix at a solution x.

LEMMA 4.2. *For any $z \in \mathbb{C}^n$ and $\delta > 0$, there exists a sufficiently large number $c = c(\delta)$. If $m > cn \log n$, then*

$$\|\Psi(\mathbf{z}) - \Phi(\mathbf{z})\| \leq \delta \|z\|^2$$

holds with probability at least $1 - 10e^{-\gamma n} - 8/n^2$.

Lemma 4.2 can be verified in the same manner of Lemma 4.7 in [9]. The next lemma is on the sample covariance matrix which can be proved in a similar fashion.

LEMMA 4.3. *Assume $\|\Psi(\mathbf{x}) - \Phi(\mathbf{x})\| \leq \delta$, then*

$$\left\| I_n - \frac{1}{m} \sum_{r=1}^{m} a_r a_r^* \right\| \leq \delta$$

with probability no less than $1 - 2e^{-\gamma m}$. On this event, it holds

$$(1-\delta)\|u\|^2 \leq \frac{1}{m} \sum_{r=1}^{m} |a_r^* u|^2 \leq (1+\delta)\|u\|^2, \ \ \forall u \in \mathbb{C}^n. \tag{4.6}$$

The next lemma reveals the distribution of the eigenvalues of $\Psi(\mathbf{x})$.

LEMMA 4.4. *Suppose that $\|x\| = 1$. Then $\Phi(\mathbf{x})$ has one eigenvalue of 4, one eigenvalue of 0, and all other eigenvalues are 1. If $\|\Psi(\mathbf{x}) - \Phi(\mathbf{x})\| \leq \delta$, then the largest eigenvalue of $\Psi(\mathbf{x})$ is less than $4 + \delta$.*

The above lemma is straightforward and hence its proof is omitted. Our proof also uses the following lemma from [6].

LEMMA 4.5. *Suppose $X_1, X_2, ..., X_m$ are $i.i.d.$ real-valued random variables obeying $X_r \leq b$ for some nonrandom $b > 0$, $\mathbb{E}X_r = 0$, and $\mathbb{E}X_r^2 = v^2$. Set $\sigma^2 = m \max(b^2, v^2)$, then*

$$\mathbb{P}(X_1 + ... + X_m \geq y) \leq \exp\left(-\frac{y^2}{2\sigma^2}\right).$$

For any $z \in \mathbb{C}^n$, we define x_z to be the vector in X^* nearest to z, i.e.,

$$x_z = \arg\min_{x \in X^*} \|z - x\|.$$

Then, we denote $h_z = z - x_z$. We now describe a few essential characteristics of $\Psi(\mathbf{z})$, $f(\mathbf{z})$, and $g(\mathbf{z})$ near the global solution. The so-called "local error bound property" is an instinctive

property of the objective function. Since its proof is different from that of [9], the detailed analysis is included. The other two properties highly depend on our modified LM method. These three properties are the foundation of our analysis. We emphasize that the bold face letters $\mathbf{z}, \mathbf{u}, \mathbf{v}, \mathbf{h}$ are the augmented vectors defined as (2.1) for z, u, v, h, respectively.

LEMMA 4.6. *Suppose that Assumption 2.1 holds, $m \geq cn\log n$ where c is sufficiently large, and $\|\Psi(\mathbf{x}) - \Phi(\mathbf{x})\| \leq \delta$ holds with $\delta = 0.01$. Let μ be determined by* (3.5)*. Then, with probability at least $1 - e^{-3\gamma n}$, we have the following properties.*

1. *Estimate of the smallest nonzero eigenvalues:*

$$\mathbf{v}^*\Psi(\mathbf{u})\mathbf{v} \geq \|u\|^2\|v\|^2 \tag{4.7}$$

   holds for all $u, v \in \mathbb{C}^n$, such that $\|u\| = \|v\| = 1$ and $Im(u^*v) = 0$;
2. *Local error bound property:*

$$\frac{1}{4}\mathbf{dist}(z,x)^2 \leq f(z) \leq 8.04\mathbf{dist}(z,x)^2 + 6.06n\mathbf{dist}(z,x)^4, \tag{4.8}$$

   holds for any z satisfying $\mathbf{dist}(z,x) \leq \frac{1}{8}$;
3. *Regularization condition:*

$$\mu(z)\mathbf{h}^*\left(\Psi_z^\mu\right)^{-1} g(\mathbf{z}) \geq \frac{1}{16}\|\mathbf{h}\|^2 + \frac{1}{66000n\|h\|}\|g(\mathbf{z})\|^2 \tag{4.9}$$

   holds for any $z = x + h$, $\|h\| \leq \frac{1}{8}$, and $f(z) \geq \frac{\|z\|^2}{900n}$.

Proof. 1) To prove (4.7), we first prove that for any $u, v \in \mathbb{C}^n$,

$$\mathbf{v}^*\Psi(\mathbf{u})\mathbf{v} \geq \mathbf{v}^*\Phi(\mathbf{u})\mathbf{v} - \|u\|\|v\|, \tag{4.10}$$

by employing Lemma 4.5. Then, by the condition $Im(u^*v) = 0$, we have $\mathbf{v}^*\Phi(\mathbf{u})\mathbf{v} \geq 2\|u\|\|v\|$, which completes the proof.

We first consider the case when u and v are fixed. Define $X_r(u,v) = 2|a_r^*u|^2|a_r^*v|^2 + 2Re((a_r^*u)^2(a_r^\top\bar{v})^2)$, $E_r(u,v) = \mathbb{E}X_r(u,v)$, then

$$\mathbf{v}^*\Psi(\mathbf{u})\mathbf{v} = \frac{1}{m}\sum_{r=1}^m X_r(u,v), \quad \mathbf{v}^*\Phi(\mathbf{u})\mathbf{v} = \frac{1}{m}\sum_{r=1}^m E_r(u,v).$$

Let $Y_r(u,v) = E_r(u,v) - X_r(u,v)$, we obtain

$$\mathbf{v}^*\Phi(\mathbf{u})\mathbf{v} - \mathbf{v}^*\Psi(\mathbf{u})\mathbf{v} = \frac{1}{m}\sum_{r=1}^m Y_r(u,v).$$

Since $Re((a_r^*u)^2(a_r^\top\bar{v})^2) \leq |a_r^*u|^2|a_r^*v|^2$, we have $X_r(u,v) \geq 0$. In addition, considering $\|u\| = \|v\| = 1$, it is easy to know $E_r(u,v) = 2|u^*v|^2 + 2 + 4Re(v^*u)^2 \leq 8$. Hence, we know $Y_r(u,v) \leq 8$. Meanwhile, it follows from the inequalities $\mathbb{E}((\mathbb{E}(X) - X)^2) \leq \mathbb{E}(X^2)$ and $|X_r| \leq 4|a_r^*u|^2|a_r^*v|^2$ that

$$\mathbb{E}Y_r(u,v)^2 \leq \mathbb{E}X_r(u,v)^2 \leq 16\mathbb{E}|a_r^*u|^4|a_r^*v|^4 \leq 16\sqrt{\mathbb{E}|a_r^*u|^8\mathbb{E}|a_r^*v|^8} = 384.$$

By choosing $\sigma^2 = 384m$ and $y = m/2$, Lemma 4.5 implies

$$\mathbb{P}(\mathbf{v}^*\Phi(\mathbf{u})\mathbf{v} - \mathbf{v}^*\Psi(\mathbf{u})\mathbf{v} \geq 0.5) \leq e^{-\frac{m}{1536}}.$$

Choosing γ to be a sufficiently small positive number, such as $\frac{1}{3072}$, we obtain

$$\mathbb{P}(\mathbf{v}^*\Psi(\mathbf{u})\mathbf{v} - \mathbf{v}^*\Phi(\mathbf{u})\mathbf{v} \geq -0.5) \geq 1 - e^{-2\gamma m}.$$

We have verified (4.10) when u and v are a fixed pair of vectors. We next extend the result to any pair of u and v. To achieve this goal, we prove that $\frac{1}{m}Y_r$ will not change too much when the variation of u and v are small, and use a net on $S^n \times S^n$ to complete the extension. We next define

$$g(u,v) = \frac{1}{m}Y_r(u,v) = \mathbf{v}^*\Phi(\mathbf{u})\mathbf{v} - \mathbf{v}^*\Psi(\mathbf{u})\mathbf{v}.$$

Then for any $u, v, v' \in \mathbb{C}^n$, we have

$$\begin{aligned}
&|g(u,v) - g(u,v')| \\
&\leq \left|\frac{1}{m}E_r(u,v) - \frac{1}{m}E_r(u,v')\right| + \left|\frac{1}{m}X_r(u,v) - \frac{1}{m}X_r(u,v')\right| \\
&\leq 2\left||u^*v|^2 - |u^*v'|^2\right| + 4|Re\left((v^*u)^2 - (v'^*u)^2\right)| \\
&\quad + \frac{1}{m}\sum_{r=1}^{m} 2|a_r^*u|^2\left||a_r^*v|^2 - |a_r^*v'|^2\right| + 2\left|Re\left((a_r^\top\bar{u})^2\left((a_r^*v)^2 - (a_r^*v')^2\right)\right)\right|. \qquad (4.11)
\end{aligned}$$

For the first two parts of (4.11), we have

$$\left||u^*v|^2 - |u^*v'|^2\right| = |(v-v')^*uu^*v + v'^*uu*(v-v')| \leq 2\|v-v'\|, \qquad (4.12)$$

$$|Re\left((v^*u)^2 - (v'^*u)^2\right)| \leq 2\|v-v'\|. \qquad (4.13)$$

For the third part of (4.11), we can derive

$$\begin{aligned}
\frac{1}{m}\sum_{r=1}^{m}|a_r^*u|^2\left||a_r^*v|^2 - |a_r^*v'|^2\right| &\leq \frac{1}{m}\sum_{r=1}^{m}|a_r^*u|^2|(v-v')^*a_ra_r^*v + v'^*a_ra_r^*(v-v')| \\
&\leq 2\cdot 6n\|v-v'\|\cdot\frac{1}{m}\sum_{r=1}^{m}|a_r^*u|^2 \\
&\leq 12(1+\delta)n\|v-v'\|. \qquad (4.14)
\end{aligned}$$

A similar derivation on the fourth part of (4.11) gives

$$\frac{1}{m}\sum_{r=1}^{m}\left|Re\left((a_r^\top\bar{u})^2\left((a_r^*v)^2 - (a_r^*v')^2\right)\right)\right| \leq 12(1+\delta)n\|v-v'\|. \qquad (4.15)$$

Substituting (4.12)-(4.15) into (4.11) yields

$$|g(u,v) - g(u,v')| \leq (12 + 48(1+\delta)n)\,\|v-v'\|.$$

Similarly, for any $u, u', v \in \mathbb{C}^n$, we obtain

$$|g(u,v) - g(u',v)| \leq (12 + 48(1+\delta)n)\,\|u-u'\|.$$

Hence, for any $u, u', v, v' \in \mathbb{C}^n$, it holds

$$|g(u,v) - g(u',v')| \leq (12 + 48(1+\delta)n)\left(\|u-u'\| + \|v-v'\|\right).$$

Choose $\epsilon \leq \frac{1}{48+192(1+\delta)n}$, such as $\epsilon = \frac{1}{250n}$, and let $\mathcal{N}_\epsilon$ be an ϵ-net of S^n. Then for any $(u,v) \in S^n \times S^n$, we can find $u', v' \in \mathcal{N}_\epsilon \times \mathcal{N}_\epsilon$, satisfying $\|u-u'\|+\|v-v'\| \leq \frac{1}{24+96(1+\delta)n}$. Hence,

$$|g(u,v) - g(u',v')| \leq \frac{1}{2}.$$

We can choose an $\mathcal{N}_\epsilon$ obeying $|\mathcal{N}_\epsilon| \leq (1+\frac{2}{\epsilon})^{2n}$. Therefore, with probability larger than

$$1 - \left(1+\frac{2}{\epsilon}\right)^{4n} e^{-2\gamma m},$$

we have for any $(u',v') \in \mathcal{N}_\epsilon \times \mathcal{N}_\epsilon$, $g(u',v') \leq 0.5$. In this occasion, for any $u, v \in S^n$, we have

$$g(u,v) \leq g(u',v') + |g(u,v) - g(u',v')| \leq 1,$$

which means

$$\mathbf{v}^* \Psi(\mathbf{u}) \mathbf{v} \geq \mathbf{v}^* \Phi(\mathbf{u}) \mathbf{v} - 1 = 2|u^* v|^2 + 4Re((v^* u)^2) + 1.$$

This completes the proof of (4.10). When $Im(u^*v) = 0$, we have $Re((v^*u)^2) \geq 0$. Therefore, $\mathbf{v}^* \Psi(\mathbf{u}) \mathbf{v} \geq 1$. In addition, when $m \geq cn \log n$ and c is sufficiently large, we have

$$1 - \left(1+\frac{2}{\epsilon}\right)^{4n} e^{-2\gamma m} = 1 - (1+500n)^{4n}\, n^{-c\gamma n} e^{-\gamma m} \leq 1 - e^{-\gamma m}.$$

This completes the proof of (4.7).

2) We now prove the left hand side of (4.8). Recalling that $z = x + h$, what we want to prove is that with high probability,

$$\frac{1}{2m} \sum_{r=1}^{m} \left(|a_r^*(x+h)|^2 - |a_r^* x|^2\right)^2 \geq \frac{1}{4}\|h\|^2 \tag{4.16}$$

holds for any $\|h\| \leq \frac{1}{8}$. Note that $|a_r^*(x+h)|^2 - |a_r^* x|^2 = 2Re(x^* a_r a_r^* h) + |a_r^* h|^2$. Let $h = sy$, where $s = \|h\| \in \mathbb{R}$, $y \in \mathbb{C}^n$ and $\|y\| = 1$. Then, it suffices to prove

$$\frac{1}{m} \sum_{r=1}^{m} \left(2Re(x^* a_r a_r^* y) + s|a_r^* y|^2\right)^2 \geq \frac{1}{2}, \tag{4.17}$$

for $0 \leq s \leq \frac{1}{8}$. We first prove the inequality for a fixed y, then extend the result to any y by using a covering argument. Since the technique is nearly the same as what is done in (VII.F) of [9], we only summarize the main steps here.

Let $X_r(y,s) := \left(2Re(x^* a_r a_r^* y) + s|a_r^* y|^2\right)^2$ and $Y_r(y,s) := \mathbb{E}X_r(y,s) - X_r(y,s)$. Then, by (VII.5-7) of [9] and the fact that $Im(x^*y) = 0$, we can easily calculate:

$$\mathbb{E}X_r(y,s) = 2s^2 + 8sRe(x^*y) + 6Re(x^*y)^2 + 2. \tag{4.18}$$

Using $0 \le s \le \frac{1}{8}$ and $X_r(y,s) \ge 0$, we obtain the following estimations

$$\begin{aligned}
Y_r(y,s) &\le \mathbb{E}X_r(y,s) \le 2s^2 + 8s + 8 < 10, \\
\mathbb{E}Y_r(y,s)^2 &\le \mathbb{E}X_r(y,s)^2 \\
&= s^4\mathbb{E}|a_r^*y|^8 + 8s^3\mathbb{E}|a_r^*y|^6 Re(x^*a_ra_r^*y) + 24s^2\mathbb{E}|a_r^*y|^4 Re(x^*a_ra_r^*y)^2 \\
&\quad +32s\mathbb{E}|a_r^*y|^2 Re(x^*a_ra_r^*y)^3 + 16\mathbb{E}Re(x^*a_ra_r^*y)^4 \\
&\le s^4\mathbb{E}|a_r^*y|^8 + 8s^3\sqrt{\mathbb{E}|a_r^*x|^2\mathbb{E}|a_r^*y|^{14}} + 24s^2\sqrt{\mathbb{E}|a_r^*x|^4\mathbb{E}|a_r^*y|^{12}} \\
&\quad +32s\sqrt{\mathbb{E}|a_r^*x|^6\mathbb{E}|a_r^*y|^{10}} + 16\sqrt{\mathbb{E}|a_r^*x|^8\mathbb{E}|a_r^*y|^8} \\
&\le 24s^4 + 568s^3 + 911s^2 + 859s + 384 \\
&< 510.
\end{aligned}$$

Applying Lemma 4.5 with $\sigma^2 = m\max(510, 10^2) = 510m$ and $y = m/4$ yields

$$\mathbb{P}\left(\sum_{r=1}^m Y_r(y,s) \ge \frac{m}{4}\right) \le e^{-2\gamma m}$$

with $\gamma = 1/8160$. It further implies

$$\mathbb{P}\left(\frac{1}{m}\sum_{r=1}^m X_r(y,s) \le \mathbb{E}X_r(y,s) - \frac{1}{4}\right) \le e^{-2\gamma m}.$$

Since $\mathbb{E}X_r(y,s) = 2s^2 + 8sRe(x^*y) + 6Re(x^*y)^2 + 2 \ge 2 - 8s\|x\|\|y\| \ge 1$, we have

$$\mathbb{P}\left(\frac{1}{m}\sum_{r=1}^m X_r(y,s) \le \frac{3}{4}\right) \le e^{-2\gamma m}. \tag{4.19}$$

This completes the proof of (4.19) for a fixed y.

In order to extend the result to all $y \in \mathbb{C}^n$, we only need to estimate

$$\left|\frac{1}{m}\sum_{r=1}^m X_r(y,s) - \frac{1}{m}\sum_{r=1}^m X_r(y',s)\right| \le L\|y - y'\|,$$

for any $y, y' \in \mathbb{C}^n$, and find an ϵ-net $\mathcal{N}_\epsilon$ with $\epsilon \le 1/(4L)$. Then, with probability no less than $1 - e^{-\gamma m}$, for all $y \in \mathcal{N}_\epsilon$, (4.19) holds for $0 \le s \le 1/8$. Under this circumstance, for any $y \in \mathbb{C}^n$, we can find a $y' \in \mathcal{N}_\epsilon$, and have

$$\begin{aligned}
\frac{1}{m}\sum_{r=1}^m X_r(y,s) &\ge \frac{1}{m}\sum_{r=1}^m X_r(y',s) - \left|\frac{1}{m}\sum_{r=1}^m X_r(y,s) - \frac{1}{m}\sum_{r=1}^m X_r(y',s)\right| \\
&\ge \frac{3}{4} - \frac{1}{4} = \frac{1}{2}.
\end{aligned}$$

This completes the proof of (4.17) and thus the left side of (4.8).

We next prove the right hand side of (4.8). By some simple calculation, we have

$$\begin{aligned}
f &= \frac{1}{2m}\sum_{r=1}^m \left(2Re(h^*a_ra_r^*x) + |a_r^*h|^2\right)^2 \\
&\le \frac{4}{m}\sum_{r=1}^m Re(h^*a_ra_r^*x)^2 + \frac{1}{m}\sum_{r=1}^m |a_r^*h|^4 \le \frac{4}{m}\sum_{r=1}^m |a_r^*h|^2|a_r^*x|^2 + \frac{1}{m}\sum_{r=1}^m |a_r^*h|^4.
\end{aligned}$$

Together with the inequalities (4.6), Corollary 7.6 of [9] and (VII.19) of [9], we can further obtain

$$f \leq 4(2+\delta)\|h\|^2 + 6(1+\delta)n\|h\|^4.$$

Recall the fact $\delta = 0.01$, f can be bounded as

$$f \leq 8.04\|h\|^2 + 6.06n\|h\|^4.$$

This completes the proof of (4.8).

3) Finally, we verify (4.9). The right side of (4.8) implies $\|h\| \geq \frac{1}{100\sqrt{n}}$ when $f(z) \geq \frac{\|z\|^2}{900n}$ holds. We notice that

$$\mu \mathbf{h}^* \left(\Psi_z^\mu\right)^{-1} g = \mathbf{h}^* g - \mathbf{h}^* \left(\Psi_z^\mu\right)^{-1} \Psi_z g. \tag{4.20}$$

Therefore, we estimate the two terms in the right hand side of (4.20), respectively. Similar to what is done in (VII.G) of [9], we obtain

$$\mathbf{h}^* g \geq \frac{1}{8}\|\mathbf{h}\|^2 + \frac{1}{11000n\|h\|}\|g\|^2. \tag{4.21}$$

Let λ_i and w_i, $i = 1, ..., 2n$ be the i-th smallest eigenvalue and associated eigenvector of Ψ_z, respectively. Suppose that g has the following decomposition $g = \sum_{s=1}^{2n} c_s w_s$, where c_s are complex numbers. Then, we obtain

$$\left(\Psi_z^\mu\right)^{-1} \Psi g = \sum_{s=1}^{2n} \frac{\lambda_s}{\lambda_s + \mu} c_s x_s,$$

which gives

$$\|\mathbf{h}^* \left(\Psi_z^\mu\right)^{-1} \Psi g\| \leq \||\mathbf{h}\|\| \left(\Psi_z^\mu\right)^{-1} \Psi g\| \leq \frac{\lambda_{2n}}{\lambda_{2n} + \mu}\|\mathbf{h}\|\|g\|. \tag{4.22}$$

On the other hand, for any $y \in \mathbb{C}^n$, $\|y\| = 1$, we have

$$\begin{aligned}
\mathbf{y}^* \left(\Psi(\mathbf{z}) - \Psi(\mathbf{x})\right) \mathbf{y} &= \frac{2}{m}\sum_{r=1}^{m} \left(|a_r^* z|^2 - |a_r^* x|^2\right) |a_r^* y|^2 + \frac{2}{m}\sum_{r=1}^{m} Re\left(\left((a_r^* z)^2 - (a_r^* x)^2\right)(a_r^* y)^2\right) \\
&\leq \frac{2}{m}\sum_{r=1}^{m} \left(|a_r^* z|^2 - |a_r^* x|^2\right) |a_r^* y|^2 + \frac{2}{m}\sum_{r=1}^{m} \left|\left((a_r^* z)^2 - (a_r^* x)^2\right)\right| |a_r^* y|^2 \\
&\leq 2\sqrt{\left(\frac{2}{m}\sum_{r=1}^{m} |a_r^* y|^4\right)\left(\frac{1}{m}\sum_{r=1}^{m} \left||a_r^* z|^2 - |a_r^* x|^2\right|^2 + \left|(a_r^* z)^2 - (a_r^* x)^2\right|^2\right)}.
\end{aligned} \tag{4.23}$$

Similar to the proof of the right side of (4.8), we can get

$$\frac{1}{m}\sum_{r=1}^{m} \left||a_r^* z|^2 - |a_r^* x|^2\right|^2 + \left|(a_r^* z)^2 - (a_r^* x)^2\right|^2 \leq 16.08\|h\|^2 + 12.12n\|h\|^4,$$

which gives

$$\mathbf{y}^* \left(\Psi(\mathbf{z}) - \Psi(\mathbf{x})\right) \mathbf{y} \leq 2\sqrt{12.12n} \cdot \sqrt{16.08\|h\|^2 + 12.12n\|h\|^4} \leq 28.09n\|h\|, \tag{4.24}$$

where the last inequality uses $\|h\| \leq 1/8$. Together with Lemma 4.4, we obtain

$$\lambda_{2n} \leq 4.01 + 28.09n\|h\|.$$

Substituting the above eigenvalue evaluation to (4.22) and together with $\mu = 70000n\sqrt{nf(z)} \geq 35000n\sqrt{n}\|h\|$, we have

$$\begin{aligned}
\|\mathbf{h}^* \left(\Psi_z^\mu\right)^{-1} \Psi g\| &\leq \frac{4.01 + 28.09n\|h\|}{4.01 + 28.09n\|h\| + 35000n\sqrt{n}\|h\|}\|\mathbf{h}\|\|g\| \\
&\leq \frac{4.01 + 28.09\sqrt{n}/100}{4.01 + 28.09\sqrt{n}/100 + 35000n/100}\|\mathbf{h}\|\|g\| \\
&\leq \frac{859}{70000\sqrt{n}}\|\mathbf{h}\|\|g\| \\
&\leq \frac{1}{16}\|\mathbf{h}\|^2 + \frac{1}{13200n\|h\|}\|g\|^2, \qquad (4.25)
\end{aligned}$$

where the second inequality uses that $\|h\| \geq \frac{1}{100\sqrt{n}}$, and $\frac{a+bz}{a+cz}$ decreases on z when $b < c$, and the last inequality uses the relationship $\|h\| \leq \frac{1}{8}$. Substituting (4.21) and (4.25) into (4.20), we immediately obtain (4.9). This completes the proof of Lemma 4.6. □

4.2. Proof of Theorem 4.1. By abuse of the notation, we simply denote z as the current iterate and z_+ determined by $\mathbf{z}_+ = \mathbf{z} - (\Psi_z^\mu)^{-1} g(\mathbf{z})$. Subtracting $\mathbf{x}$ from this equation, we have

$$\|\mathbf{h}_{z^+}\| \leq \|\mathbf{z}_+ - \mathbf{x}_z\| = \|\mathbf{h}_z - (\Psi_z^\mu)^{-1} g(\mathbf{z})\|. \qquad (4.26)$$

For the sake of simplicity, we omit the letter z in $f(z)$, $\mu(z)$, h_z and h_{z^+}, and omit the letter $\mathbf{z}$ in $g(\mathbf{z})$, $\Psi(\mathbf{z})$ and $\Phi(\mathbf{z})$, when it causes no ambiguity.

We divide the proof of Theorem 4.1 into two parts. 1) We verify (4.1) and (4.3) under the condition $f(z) < \frac{\|z\|^2}{900n}$. The updating formula for μ gives $\mu = \sqrt{f(z)}$. Using the left hand side of (4.8) of Lemma 4.6, we have

$$\|h_z\| \leq 2\sqrt{f} = 2\sqrt{\frac{\|z\|^2}{900n}} \leq \frac{1 + ||h_z||}{15\sqrt{n}},$$

which implies

$$\|h\| \leq \frac{1}{14\sqrt{n}}.$$

By the definition of h, we know that $Im(z^* h) = 0$, which means $\tilde{\mathbf{z}}^* \mathbf{h} = 0$. On the other hand, it is easy to verify that

$$\begin{aligned}
\mathbf{h}_z - (\Psi_z^\mu)^{-1} g(\mathbf{z}) &= (\Psi_z^\mu)^{-1} \left(\Psi_z^\mu \mathbf{h}_z - g(\mathbf{z})\right) \\
&= (\Psi_z^\mu)^{-1} \left(\mu\mathbf{h} + \frac{1}{m}\sum_{r=1}^m h^* a_r a_r^* h \begin{bmatrix} (a_r a_r^*) z \\ (\bar{a}_r a_r^\top)\bar{z} \end{bmatrix}\right),
\end{aligned}$$

and

$$\tilde{\mathbf{z}}^* \left(\mu\mathbf{h} + \frac{1}{m}\sum_{r=1}^m h^* a_r a_r^* h \begin{bmatrix} (a_r a_r^*) z \\ (\bar{a}_r a_r^\top)\bar{z} \end{bmatrix}\right) = 0,$$

which further gives $Im\left(z^*(h+\frac{1}{m}\sum_{r=1}^{m}h^*a_ra_r^*h(a_ra_r^*)z)\right)=0$. Hence, the eigenvalue estimate (4.7) implies that the smallest eigenvalue of Ψ restricted in the subspace $\mathcal{S}:=\{v \mid Im(z^*v)=0\}$ is $\frac{1}{2}||z||^2$. Therefore, the largest eigenvalue of $(\Psi_z^\mu)^{-1}$ restricted in $\mathcal{S}$ is $\frac{2}{||z||^2+2\mu}$. Then, we can obtain

$$\|\mathbf{h}_+\| \le \left\|(\Psi_z^\mu)^{-1}\left(\mu\mathbf{h}+\frac{1}{m}\sum_{r=1}^{m}h^*a_ra_r^*h\begin{bmatrix}(a_ra_r^*)z\\(\bar{a}_ra_r^\top)\bar{z}\end{bmatrix}\right)\right\| \tag{4.27}$$

$$\le \frac{2}{\|z\|^2+2\mu}\left\|\left(\mu\mathbf{h}+\frac{1}{m}\sum_{r=1}^{m}h^*a_ra_r^*h\begin{bmatrix}(a_ra_r^*)z\\(\bar{a}_ra_r^\top)\bar{z}\end{bmatrix}\right)\right\|. \tag{4.28}$$

Denote $v:=\frac{1}{m}\sum_{r=1}^{m}h^*a_ra_r^*h(a_ra_r^*)z$, we obtain

$$\|\mathbf{h}_+\| \le \frac{2}{\|z\|^2+2\mu}\|\mu\mathbf{h}+\mathbf{v}\| \le \frac{2\mu}{\|z\|^2+2\mu}\|\mathbf{h}\|+\frac{2}{\|z\|^2+2\mu}\|\mathbf{v}\|. \tag{4.29}$$

By using the definition of μ and the local error bound condition (4.8), it holds that

$$\|z\|^2+2\mu=\|x-h\|^2+2\sqrt{f}\ge 1-2\|h\|+\|h\|^2+\|h\|\ge 1, \tag{4.30}$$

$$\mu=\sqrt{f}\le\sqrt{8.04}||h||+\sqrt{6.06}\sqrt{n}||h||^2\le 2.01||\mathbf{h}||+1.24\sqrt{n}||\mathbf{h}||^2 \tag{4.31}$$

We next estimate $\|v\|$. For any $u\in\mathbb{C}^n$ and $\|u\|=1$, using (4.6), Corollary 7.6 of [9] and (VII.19) of [9], we have

$$\begin{aligned}|u^*v| &\le \frac{1}{m}\sum_{r=1}^{m}|a_r^*h|^2|a_r^*z||a_r^*u| = \frac{1}{m}\sum_{r=1}^{m}|a_r^*h|^2|a_r^*(x+h)||a_r^*u|\\ &\le \frac{1}{m}\sum_{r=1}^{m}|a_r^*h|^3|a_r^*u|+\frac{1}{m}\sum_{r=1}^{m}|a_r^*h|^2|a_r^*x||a_r^*u|\\ &\le 6n\|h\|\frac{1}{m}\sum_{r=1}^{m}|a_r^*h|^2+\sqrt{(2+\delta)\frac{1}{m}\sum_{r=1}^{m}|a_r^*h|^4}\\ &\le 6n(1+\delta)\|h\|^3+\sqrt{6n(1+\delta)(2+\delta)}\|h\|^2.\end{aligned}$$

Therefore, the norm of v can be bounded by

$$\begin{aligned}\|\mathbf{v}\| &= \sqrt{2}\|v\| \le \sqrt{2}\left(6n(1+\delta)\|h\|^3+\sqrt{6n(1+\delta)(2+\delta)}\|h\|^2\right)\\ &= 3.03n\|\mathbf{h}\|^3+2.47\sqrt{n}\|\mathbf{h}\|^2.\end{aligned} \tag{4.32}$$

Substituting (4.30)-(4.32) into (4.29) yields

$$\begin{aligned}\|\mathbf{h}_+\| &\le 2\mu\|\mathbf{h}\|+2\|\mathbf{v}\|\\ &\le 4.02\|\mathbf{h}\|^2+2.48\sqrt{n}\|\mathbf{h}\|^3+6.06n\|\mathbf{h}\|^3+4.94\sqrt{n}\|\mathbf{h}\|^2.\end{aligned}$$

Using the fact that $|\mathbf{h}|=\sqrt{2}\|h\|\le\frac{1}{7\sqrt{2n}}$, we further have

$$\|\mathbf{h}_+\|<(4.28+5.56\sqrt{n})\|\mathbf{h}\|^2<\frac{4.28+5.56\sqrt{n}}{9.89\sqrt{n}}\|\mathbf{h}\|,$$

which guarantees the inequalities (4.1) and (4.3) under the situation that $f(\mathbf{z}_k) < \frac{\|z\|^2}{900n}$.

2) We next consider the case under the conditions $f(z) \geq \frac{\|z\|^2}{900n}$ and $\|h\| \leq \frac{1}{8}$. Recalling the inequality (4.8) and (4.9) in Lemma 4.6, and the positive definiteness of Ψ_z, we obtain

$$\begin{aligned}\|\mathbf{h}_+\|^2 &\leq \|\mathbf{h} - (\Psi_z^\mu)^{-1}g\|^2 \leq \|\mathbf{h}\|^2 - 2\mathbf{h}^*(\Psi_z^\mu)^{-1}g + \frac{1}{\mu^2}\|g\|^2 \\ &\leq \left(1 - \frac{1}{8\mu}\right)\|\mathbf{h}\|^2 + \left(\frac{1}{\mu^2} - \frac{2}{66000n\|h\|\mu}\right)\|g\|^2 \\ &\leq \left(1 - \frac{1}{8\mu}\right)\|\mathbf{h}\|^2 + \frac{1}{\mu}\left(\frac{1}{35000n\sqrt{n}\|h\|} - \frac{2}{66000n\|h\|}\right)\|g\|^2 \leq \left(1 - \frac{1}{8\mu}\right)\|\mathbf{h}\|^2,\end{aligned}$$

which implies

$$\|\mathbf{h}_+\| \leq \left(1 - \frac{1}{4\mu}\right)\|\mathbf{h}\|. \tag{4.33}$$

Therefore, we finish the proof for the special case $||x|| = 1$.

3) Finally, we consider $\|x\| \neq 1$. By observing the iteration scheme (3.4), it is not difficult to verify that starting from $\frac{z_0}{\|x\|}$, the kth LM iteration for a problem with a solution $\frac{x}{\|x\|}$ is

$$\frac{\mathbf{z}_{k+1}}{\|x\|} = \frac{\mathbf{z}_k}{\|x\|} + \frac{\mathbf{s}_k}{\|x\|}.$$

Therefore, by the previous proof for the case $\|x\| = 1$, we have

$$\begin{aligned}&\mathbf{dist}\left(\frac{z_{k+1}}{\|x\|}, \frac{x}{\|x\|}\right) < \left(1 - \frac{1}{4\left(\frac{\mu_k}{\|x\|}\right)}\right)\mathbf{dist}\left(\frac{z_k}{\|x\|}, \frac{x}{\|x\|}\right), \\ &\frac{\mathbf{dist}\left(\frac{z_{k+1}}{\|x\|}, \frac{x}{\|x\|}\right)}{\mathbf{dist}\left(\frac{z_k}{\|x\|}, \frac{x}{\|x\|}\right)^2} < 4.28 + 5.56\sqrt{n},\end{aligned}$$

which yields (4.1) and (4.3). This completes the proof for all x.

5. Analysis of the Inexact LM Method. In this section, we first establish the convergence result for the inexact LM framework, then present a pre-conditioned conjugate gradient (PCG) method for solving the LM system inexactly and provide a practically useful choice of the pre-conditioner.

5.1. Convergence of the Inexact LM Method. The following theorem describes the global linear convergence of the inexact LM method.

THEOREM 5.1. *Suppose that Assumption 2.1 holds. Let $x \in \mathbb{C}^n$ be any solution of* (1.2)*, and $m \geq c_0 n \log n$, where c_0 is a sufficiently large constant. Assume that $\{z_k\}$ is a sequence generated by Algorithm 2 with the parameter*

$$\eta_k := \begin{cases} \frac{(1-c_1)}{22.35n} \cdot \frac{\mu_k}{||x||}, & \text{if } f(z_k) \geq \frac{1}{900n}\|z_k\|^2; \\ \frac{9.89\sqrt{n}-c_2||x||}{249.57n^2||x||} \cdot \frac{\mu_k\|g(\mathbf{z}_k)\|}{||x||^2}, & \text{otherwise.} \end{cases} \tag{5.1}$$

Then, starting from any initial solution z_0 satisfying $\mathbf{dist}(z_0, x) \leq \|x\|/8$, there is an event of probability at least $1 - 15e^{-\gamma n} - 8/n^2 - me^{-1.5n}$, such that on this event, it holds that

$$\mathbf{dist}(z_{k+1}, x) < \frac{1 + c_1}{2}\mathbf{dist}(z_k, x), \qquad \textit{for all } k = 0, 1, ... \tag{5.2}$$
$$\mathbf{dist}(z_{k+1}, x) < \frac{9.89\sqrt{n} + c_2||x||}{2||x||}\mathbf{dist}(z_k, x)^2, \qquad \textit{for all } k \geq l,$$

where c_1, c_2 are defined by (4.2) *and* (4.4)*, respectively, and l satisfying $f(z_l) < \frac{1}{900n}\|z_l\|^2$.*

Proof. We only prove the result when $\|x\| = 1$ and $f(z_k) \geq \frac{1}{900n}\|z_k\|^2$. The other part can be proved in the same manner and hence omitted. Let

$$\mathbf{z}'_{k+1} := \mathbf{z}_k - (\Psi_{z_k}^{\mu_k})^{-1} g(\mathbf{z}_k)$$

be the exactly LM step at the k-th iteration. By using Theorem 4.1, we have

$$\mathbf{dist}(z'_{k+1}, x) < c_1\mathbf{dist}(z_k, x), \tag{5.3}$$

and

$$\begin{aligned}
\mathbf{dist}(z_{k+1}, x) &\leq \mathbf{dist}(z'_{k+1}, x) + \|z'_{k+1} - z_{k+1}\| \\
&= \mathbf{dist}(z'_{k+1}, x) + \frac{1}{\sqrt{2}}\|\mathbf{s}_k + \left(\Psi_{z_k}^{\mu_k}\right)^{-1} g(\mathbf{z}_k)\| \\
&= \mathbf{dist}(z'_{k+1}, x) + \frac{1}{\sqrt{2}}\|\left(\Psi_{z_k}^{\mu_k}\right)^{-1}\left(\Psi_{z_k}^{\mu_k}\mathbf{s}_k + g(\mathbf{z}_k)\right)\| \\
&\leq \mathbf{dist}(z'_{k+1}, x) + \frac{\eta}{\sqrt{2}\mu_k}\|g(\mathbf{z}_k)\|. \qquad (5.4)
\end{aligned}$$

By using Lemma 4.3, Lemma 4.6 and Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}
\|g(\mathbf{z}_k)\| &= \sqrt{2}\left\|\frac{1}{m}\sum_{r=1}^{m}\left(|a_r^* z_k|^2 - |a_r^* x|^2\right) a_r a_r^* z_k\right\| \\
&\leq \sqrt{24n\|z_k\|^2(1+\delta)\frac{1}{m}\sum_{r=1}^{2m}\left(|a_r^* z|^2 - |a_r^* x|^2\right)^2} \\
&\leq \frac{\sqrt{2}}{2}\cdot 22.35n\|h_k\|. \qquad (5.5)
\end{aligned}$$

Substituting (5.3), (5.5) and the updating formula (5.1) into (5.4), we immediately obtain the relationship (5.2). □

Theorem 5.1 tells us that if η_k takes the order of $\|f(z_k)\|$, the inexactly LM method guarantees a global linear convergence to a global solution. When η_k takes the order of $\|f(z_k)\|^{\frac{3}{2}}$, the inexactly LM method achieves a local quadratic convergence rate.

5.2. Solve the LM Equation by PCG. In this subsection, we discuss the PCG method for solving the LM equation (3.3). The CG methods admits a global linear convergence rate which depends on the condition number of the coefficient matrix (see [41, 27]). However, the linear system matrix Ψ_z^μ tends to be singular as the parameter μ_k decreases, which takes place when the iteration is close enough to the solution set. Our recipe is using the PCG method with a suitable pre-conditioner M. Therefore, the original linear system (3.3) is replaced by

$$M^{-1}\Psi_z^\mu \mathbf{s} = M^{-1} g(\mathbf{z}). \tag{5.6}$$

Since $\mathbb{E}\Psi(\mathbf{z}) = \Phi(\mathbf{z})$ if z is independent to $\{a_r\}$, we suggest to use a pre-conditioner

$$\Phi_z^\mu := \Phi_z + \mu_z I_{2n} \tag{5.7}$$

and $-(\Phi_z^\mu)^{-1} g(\mathbf{z})$ as the initial point of the PCG method. A simple verification shows that Φ_z^μ is positive definite and its inverse has an explicit formulation:

$$(\Phi_z^\mu)^{-1} = aI_{2n} + b\mathbf{z}\mathbf{z}^* + c\tilde{\mathbf{z}}\tilde{\mathbf{z}}^*, \tag{5.8}$$

where

$$a = \frac{1}{\|z\|^2 + \mu}, \quad b = -\frac{3}{2(\|z\|^2 + \mu)(4\|z\|^2 + \mu)}, \quad c = \frac{1}{2(\|z\|^2 + \mu)\mu}.$$

Hence, the linear system $(\Phi_z^\mu)^{-1}\mathbf{s} = b$ can be calculated in $\mathcal{O}(n)$ arithmetic operations.

The remaining task is to analyze the condition number of $(\Phi_z^\mu)^{-1}\Psi_z^\mu$. Similar to Ψ_z^μ, Φ_z^μ is also nearly singular once μ is small. Therefore, the condition number of $(\Phi_z^\mu)^{-1}\Psi_z^\mu$ is likely to be huge. Fortunately, the subspace

$$V := \left\{ \mathbf{x} \mid \mathbf{x} = \begin{bmatrix} s \\ \bar{s} \end{bmatrix}, \forall\, s \in \mathbb{C}^n \right\}$$

is a common range space of Φ_z and Ψ_z. It can be easily verified that any iteration $\mathbf{z}$ is in V and $(\Phi_z^\mu)^{-1}\Psi_z^\mu \mathbf{z} \in V$ if $\mathbf{z} \in V$. It is easy to establish the following convergence property of the CG method.

LEMMA 5.2. *Assume that A is a positive semidefinite matrix and V is its range space. Denote $A^\mu := A + \mu I$. Let $y_* \in V$ be the solution of the linear system*

$$A^\mu y = b,$$

and $\{y_k\}$ be the sequence generated by the CG method from an initial point $y_0 \in V$. Then, for any $k \geq 1$, it holds

$$\|y_k - y_*\|_{A^\mu} \leq 2\left(\frac{\sqrt{\kappa_V(A^\mu)} - 1}{\sqrt{\kappa_V(A^\mu)} + 1}\right)^k \|y_0 - y_*\|_{A^\mu},$$

where $\|y\|_{A^\mu} = (y^* A^\mu y)^{1/2}$ and $\kappa_V(A^\mu)$ refers to the restricted condition number

$$\kappa_V(A^\mu) := \frac{\max\limits_{y \in V, ||y||_2 = 1} y^* A^\mu y}{\min\limits_{y \in V, ||y||_2 = 1} y^* A^\mu y}.$$

Lemma 5.2 shows that one only need to evaluate the restricted condition number of $(\Phi_z^\mu)^{-1}\Psi_z^\mu$. Without loss of generality, we assume $||x|| = 1$. Let λ be an eigenvalue of $(\Phi_z^\mu)^{-1}\Psi_z^\mu$, and $\mathbf{y}_\lambda$ be the corresponding eigenvector. Firstly, we have

$$||(\Phi_z^\mu)^{-1}\Psi_z^\mu \mathbf{y}_\lambda - \mathbf{y}_\lambda|| = |\lambda - 1|. \tag{5.9}$$

Using the relationship (4.7) and (4.24) and the fact that $\|h\| \leq \frac{1}{8}$, we obtain

$$\begin{aligned}
||(\Phi_z^\mu)^{-1}\Psi_z^\mu \mathbf{y}_\lambda - \mathbf{y}_\lambda|| &= ||((\Phi_z^\mu)^{-1})(\Psi_z^\mu - \Phi_z^\mu)\mathbf{y}_\lambda|| \\
&\leq \frac{2}{||z|| + 2\mu} ||\Psi_z^\mu - \Phi_z^\mu|| ||\mathbf{y}_\lambda|| \\
&\leq \frac{2}{\left(\frac{7}{8}\right)^2 + 2\mu} \left(||\Psi_z - \Psi_x|| + ||\Psi_x - \Phi_x|| + ||\Phi_x - \Phi_z||\right) \\
&\leq \frac{2}{\left(\frac{7}{8}\right)^2 + 2\mu} (37.09n\|h\| + 0.01).
\end{aligned}$$

Assume $\mu = Kn\sqrt{nf(z)}$, then

$$|\lambda - 1| \leq \frac{74.18n\|h\| + 0.02}{\left(\frac{7}{8}\right)^2 + 2Kn\sqrt{nf(z)}} < \frac{75}{K\sqrt{n}} + 0.03.$$

Hence, we have

$$\kappa_V((\Phi_z^\mu)^{-1}\Psi_z^\mu) \leq \frac{1.03K\sqrt{n} + 75}{0.97K\sqrt{n} - 75},$$

which means the condition number is close to 1 if either K or n is large.

In each PCG iteration, the computational cost of the gradient evaluation is $\mathcal{O}(mn)$, and the cost of the matrix-vector multiplications for calculating $(\Phi_z^\mu)^{-1}\mathbf{s}$ is also $\mathcal{O}(mn)$. Lemma 5.2 shows that the upper bound of the number of iterations is related to the restricted condition number and the distance between the initial guess and the solution set. Since the restricted condition number of $(\Phi_z^\mu)^{-1}\Psi_z^\mu$ is small, the PCG method often takes just a few iterations to achieve a good accuracy. Therefore, the computational cost at a single iteration of our PCG method is not too expensive than that of the WF method.

6. Extensions to the Coded Diffraction (CD) Model. We make the following assumption in this section.

ASSUMPTION 6.1. *A problem is called the CD model if*

$$y_r = \left|\sum_{t=0}^{n-1} x(t)\bar{d}_l(t)e^{-i2\pi kt/n}\right|^2, \quad r = (l,k), \ \ 0 \leq k \leq n-1, \ \ 1 \leq l \leq L, \tag{6.1}$$

where $x(t)$ and $d_l(t)$ denote the t-th element of x and d_l, respectively. Assume that $L \geq c(\log n)^4$, where c is a sufficiently large numerical constant, and d_l are i.i.d sampled from a distribution d, which is symmetric and satisfies $|d| \leq M$ and

$$\mathbb{E}d = 0, \ \ \mathbb{E}d^2 = 0, \ \ \mathbb{E}|d|^4 = 2(\mathbb{E}|d|^2)^2.$$

For the CD model, an initialization via resampled Wirtinger Flow is introduced in [9] as Algorithm 3. By conducting a resampled gradient descent steps, this initialization scheme can provide a better initial guess than that of Algorithm 1.

Algorithm 3: Initialization via the resampled WF method

1 Input measurements $\{a_r\}$ and observations $\{y_r\}$ $(r = 1, 2, \ldots, m)$.

2 Divide the measurements and observations into $B+1$ groups of size $m' = \lfloor m/(B+1) \rfloor$. The measurements and observations in group b are denoted as $a_r^{(b)}$ and $y_r^{(b)}$ for $b = 0, 1, ..., B$.

3 Obtain u_0 by conducting Algorithm 1 on group 0.

4 For $b = 0$ to $B-1$, perform the following update:

$$u_{b+1} = u_b - \frac{\mu}{\|u_0\|^2}\left(\frac{1}{m'}\sum_{r=1}^{m'}\left(|z^* a_r^{(b+1)}|^2 - y_r^{(b+1)}\right)(a_r^{(b+1)}a_r^{(b+1)^*})z\right)$$

5 Set $z_0 = u_B$.

By employing Algorithm 3, the distance between the initial guess z_0 and a solution x can be improved to

$$\mathbf{dist}(z_0, x) \leq \frac{1}{8\sqrt{n}}\|x\|. \tag{6.2}$$

Then the WF method can achieve

$$\mathbf{dist}(z_k, x) \leq \frac{1}{8\sqrt{n}}\left(1 - \frac{\mu}{3}\right)^{k/2}\|x\|.$$

Readers who are interested in the WF algorithm can refer to section V and VII of [9] for the detailed information.

The modified LM Algorithm 2 can be extended to solve the CD model directly. For the sake of theoretical analysis, the regularization parameter μ_k is updated as

$$\mu_k = \begin{cases} 35000n\sqrt{f(z_k)}, & \text{if } f(z_k) \geq \frac{1}{3260n}\|z_k\|^2; \\ \sqrt{f(z_k)}, & \text{otherwise.} \end{cases} \tag{6.3}$$

If $L \geq c(\log n)^3$, then a counterpart of Lemma 4.2 holds with probability at least $1 - (2L+1)/n^3$. The first equality in Lemma 4.3 also holds with probability no less than $1-1/n^2$. Finally, we extend Lemma 4.6 to the CD model.

LEMMA 6.2. *Suppose that Assumption 6.1 holds, $\|\Psi(\mathbf{x}) - \Phi(\mathbf{x})\| \leq \delta$ holds with $\delta = 0.01$ and μ_k is updated by* (6.3)*. Then, with probability at least $1 - 3/n^2$, we have*

1. *Estimate of the smallest nonzero eigenvalue:*

$$\mathbf{v}^*\Psi(\mathbf{u})\mathbf{v} \geq \frac{1}{2}\|u\|^2\|v\|^2 \tag{6.4}$$

holds for all $u, v, z \in \mathbb{C}^n$, such that $Im(u^*v) = 0$, and $dist(u, x) \leq 1/(50\sqrt{n})$;

2. *Local error bound property:*

$$\frac{4}{5}\mathbf{dist}(z, x)^2 \leq f(z) \leq 8.04\mathbf{dist}(z, x)^2 + 6.06n\mathbf{dist}(z, x)^4, \tag{6.5}$$

holds for any z satisfying $\mathbf{dist}(z, x) \leq \frac{1}{8\sqrt{n}}$;

3. *Regularization condition:*

$$\mu(z)\mathbf{h}^*\left(\Psi_z^\mu\right)^{-1} g(\mathbf{z}) \geq \frac{1}{16}\|\mathbf{h}\|^2 + \frac{1}{33000n\|h\|}\|g(\mathbf{z})\|^2 \tag{6.6}$$

holds for any $z = x + h$, $\|h\| \leq \frac{1}{8\sqrt{n}}$, and $f(z) \geq \frac{\|z\|^2}{3260n}$.

Proof. Since the measurements are not independent from each other in the CD model, Lemma 4.5 cannot be applied. We first prove (6.4). Note that for $u, v, z \in \mathbb{C}^n$ satisfying $Im(u^*v) = 0$,

$$\begin{aligned} \mathbf{v}^*\Psi(\mathbf{u})\mathbf{v} &= \mathbf{v}^*\Phi(\mathbf{u})\mathbf{v} + \mathbf{v}^*(\Psi(\mathbf{u}) - \Phi(\mathbf{u}))\mathbf{v} \\ &\geq 2\|u\|^2\|v\|^2 + 2|u^*v|^2 + 4Re\left((v * u)^2\right) - 2\|v\|^2\|\Psi(\mathbf{u}) - \Phi(\mathbf{u})\| \\ &\geq 2\|u\|^2\|v\|^2 - 2\|v\|^2\|\Psi(\mathbf{u}) - \Phi(\mathbf{u})\|. \end{aligned}$$

Hence, (6.4) holds if $\|\Psi(\mathbf{u}) - \Phi(\mathbf{u})\| \leq \frac{3}{4}\|u\|^2$ for all u obeying $dist(u, x) \leq \frac{1}{50\sqrt{n}}$. Because $\Psi(\mathbf{u}) - \Phi(\mathbf{u})$ is homogeneous for u when x is fixed, we assume $\|u\| = 1$ without loss of generality. Then, we have to prove $\|\Psi(\mathbf{u}) - \Phi(\mathbf{u})\| \leq \frac{3}{4}$. It holds that

$$\|\Psi(\mathbf{u}) - \Phi(\mathbf{u})\| \leq \|\Psi(\mathbf{u}) - \Psi(\mathbf{x})\| + \|\Psi(\mathbf{x}) - \Phi(\mathbf{x})\| + \|\Phi(\mathbf{x}) - \Phi(\mathbf{u})\|. \tag{6.7}$$

Taking $h = u - x$ leads to $\|h\| \leq 1/(50\sqrt{n})$. By using (4.24), we have

$$\begin{aligned}\mathbf{y}^* \left(\Psi(\mathbf{u}) - \Psi(\mathbf{x})\right) \mathbf{y} &\leq 2\sqrt{12.12n(16.08\|h\|^2 + 12.12n\|h\|^4)} \\ &\leq 0.57,\end{aligned}$$

for all $y \in \mathbb{C}^n, \|y\| = 1$. Therefore, $\|\Psi(\mathbf{u}) - \Psi(\mathbf{x})\| \leq 0.57$ is an estimation of the first term of the right side of (6.7). The second term of (6.7) satisfies

$$\|\Psi(\mathbf{x}) - \Phi(\mathbf{x})\| \leq \delta = 0.01.$$

Similar to the first term, the third term of the right side of (6.7) can be estimated as

$$\|\Phi(\mathbf{x}) - \Phi(\mathbf{u})\| \leq 8(1 + \|h\|)\|h\| \leq 0.17.$$

Hence, we have

$$\|\Psi(\mathbf{u}) - \Phi(\mathbf{u})\| \leq 0.57 + 0.01 + 0.17 = 0.75. \tag{6.8}$$

This completes the proof of (6.4).

We next prove the left side of (6.5). By following $(a-b)^2 \geq \frac{a^2}{2} - b^2$, we have

$$\begin{aligned}f(z) &= \frac{1}{m}\sum_{r=1}^{m}\left(|a_r^* z|^2 - |a_r^* x|^2\right)^2 \\ &= \frac{1}{m}\sum_{r=1}^{m}\left(2Re(h^* a_r^* a_r x^*) + |a_r^* h|^2\right)^2 \\ &\geq \frac{2}{m}\sum_{r=1}^{m} Re(h^* a_r^* a_r x^*)^2 - \frac{1}{m}\sum_{r=1}^{m}|a_r^* h|^4.\end{aligned}$$

Using Corollary 7.5 of [9] and $Im(h^* x) = 0$, we know

$$\frac{1}{m}\sum_{r=1}^{m} Re(h^* a_r^* a_r x^*)^2 \geq \frac{1-\delta}{2}\|h\|^2.$$

Together with $\frac{1}{m}\sum\limits_{r=1}^{m}|a_r^* h|^4 \leq 6n(1+\delta)\|h\|^4$ and $\|h\| \leq 1/(8\sqrt{n})$, we obtain

$$f(z) \geq \left(1 - \delta - \frac{6(1+\delta)}{64}\right)\|h\|^2 > \frac{4}{5}\|h\|^2.$$

The right side of (6.5) can be proven in the same way as (4.8). Hence, the detailed proof is omitted. Finally, we prove (6.6). Similar to (4.23) and using $\|h\| \leq \frac{1}{8\sqrt{n}}$, we can estimate the largest eigenvalue of $\Psi(\mathbf{z})$:

$$\lambda_{2n} \leq 5.78 + 2\sqrt{12.12nf(z)}.$$

Therefore, using a same derivation as (4.25), we obtain

$$\|\mathbf{h}^* \left(\Psi_z^{\mu}\right)^{-1} \Psi g\| \leq \frac{1}{16}\|\mathbf{h}\|^2 + \frac{1}{16500n\|h\|}\|g\|^2,$$

which together with (4.21) gives (6.6). □

Consequently, both global linear convergence rate and local quadratic rate can be established for the CD model based on the above lemmas.

THEOREM 6.3. *Suppose that Assumption 6.1 holds. Let $x \in \mathbb{C}^n$ be any solution of* (1.2) *and $\{z_k\}$ be a sequence generated by Algorithm 2 where the LM equation exactly solved and μ_k is chosen as* (6.3)*. Then, starting from any initial solution z_0 obeying $\mathbf{dist}(z_0, x) \leq \frac{\|x\|}{8\sqrt{n}}$, there is an event of probability at least $1 - (2L+1)/n^3 - 1/n^2$ such that on this event,*

$$\begin{aligned} &\mathbf{dist}(z_{k+1}, x) < c_1 \mathbf{dist}(z_k, x), \qquad \textit{for all } k = 0, 1, ... \\ &\mathbf{dist}(z_{k+1}, x) < c_2 \mathbf{dist}(z_k, x)^2, \qquad \textit{for all } k \geq l, \end{aligned}$$

where s satisfies $f(z_s) < \frac{1}{3260n}\|z_s\|^2$, and

$$c_1 := \begin{cases} \left(1 - \frac{||x||}{4\mu_k}\right), & \textit{if } f(z_k) \geq \frac{1}{3260n}\|z_k\|^2; \\ \frac{8.19+10.23\sqrt{n}}{35.35\sqrt{n}}, & \textit{otherwise;} \end{cases} \tag{6.9}$$

$$c_2 := \frac{8.19 + 10.23\sqrt{n}}{\|x\|}. \tag{6.10}$$

Similar theoretical results on the inexact LM method can also be derived. The proof of the theorem follows the same procedure and shares the same inequalities, although the calculation is different. We omit them for conciseness.

7. Numerical Experiments. In this section, we present some numerical results to demonstrate the performance of the LM method using the parameter $\mu_k = \sqrt{f(z_k)}$, and compare it with the WF method in [9].

7.1. Recovery of 1D signals. We begin our numerical experiments on 1-D random signals under Gaussian and CD model. In order to make comparison with the WF method, we choose the same type of signals as that in [9]:

- Random low-pass signals, where x is given by

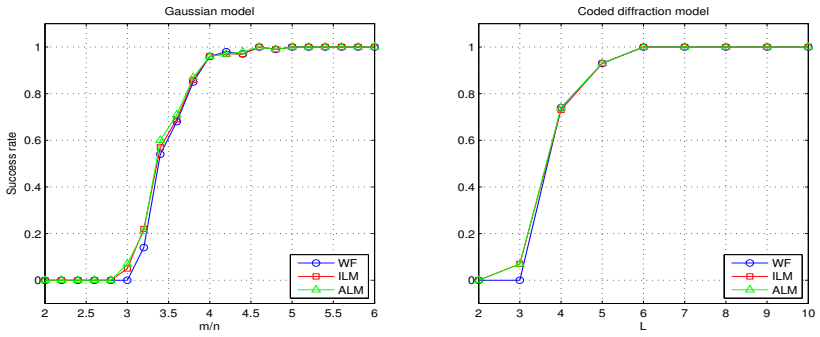
$$x[t] = \sum_{k=-(M/2-1)}^{M/2} (X_k + iY_k) e^{2\pi i(k-1)(t-1)/n},$$

  with $M = n/8$ and X_k and Y_k are i.i.d. $\mathcal{N}(0, 1)$.
- Random Guassian signals, where $x \in \mathbb{C}^n$ is a random complex Gaussian vector with i.i.d. entries of the form

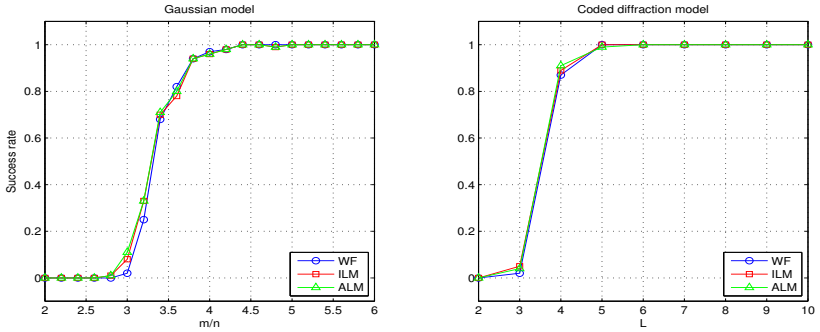
$$X[t] = X + iY,$$

  with X and Y distributed as $\mathcal{N}(0, 1)$.

In the initialization step, 50 iterations of power method are run to calculate the eigenvector needed in Algorithm 1. For the LM method, we solve the LM system accurately and inaccurately, and these two versions are denoted by "ALM" and "ILM", respectively. For the ALM method, we set $\eta_k = 10^{-16}$ in (3.6). For the ILM method, we set the maximum iteration number of PCG to be $\min(k + 2, 5)$, where k is the iteration number in Algorithm 2. We stop the LM algorithm after 200 iterations. For the WF method, we use the step length $\mu_k = \min(1 - exp(-k/\tau_0), 0.2)$, where $\tau_0 \approx 330$, and stop after 2500 iterations. Notice that in each ILM iteration, there are at most 5 PCG iterations. Consequently, the computational



(a) Success rate for Gaussian signals



(b) Success rate for low-pass signals

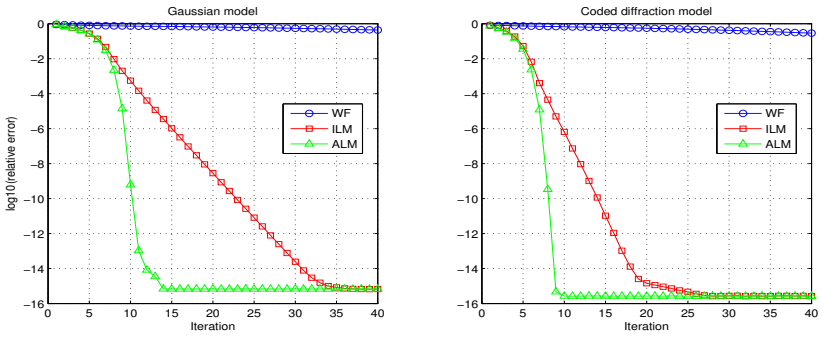
FIG. 7.1. *Empirical probability of success based on 100 random trials.*

cost of every PCG iteration is about two times of a WF iteration. Hence, considering the calculation of gradient in each ILM iterations, the computational cost of one ILM iteration is no more than that of 11 WF iterations. Therefore, excuting 200 ILM iterations is not more expensive than 2500 WF iterations.

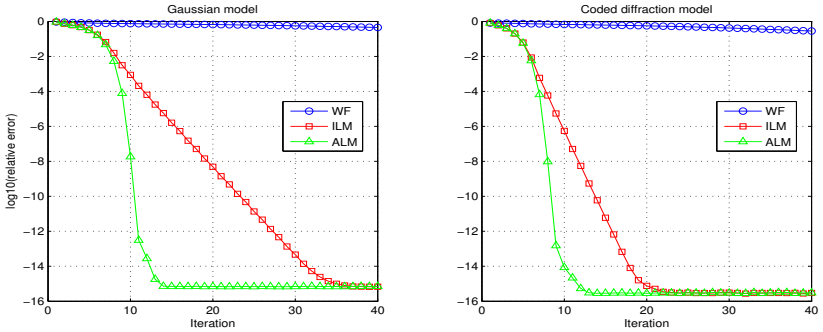
In this experiment, we set $n = 512$ and compare the empirical success rate and the CPU time of the LM and WF methods. The empirical probability of success is an average over 100 trials, where in each instance, new random sampling vectors are generated according to the Gaussian or CD models. For coded diffraction model, we use octanary patterns as the masks in [9]. We declare a trial successful if the relative error of the reconstruction $\mathbf{dist}(z_{final}, X^*)/\|x\|$ falls below 10^{-5} before the iteration process is stopped. For a successful trial, we define the CPU time of this trial to be the time used until the first iteration after which the relative error is smaller than 10^{-5}.

Figure 7.1 shows that around $4.5n$ Gaussian phaseless measurements or 6 octanary patterns are enough for an exact recovery with high probability for all algorithms. For all tested signals and models, the success rate of the LM methods rises a little bit earlier than the WF method as the number of measurements increases. The three algorithms perform similarly in terms of the success rates. Furthermore, this figure shows that solving the LM equations inexactly does not exert significant impact on success rate of the LM method.

We next examine the order of convergence of the WF and LM methods. Figure 7.2 shows the relationship between the relative error in logarithm scale and the number of iterations for the three algorithms. To better illustrate the performance of the LM methods, we only show errors of the first 40 iterations. We can see from Figure 7.2 that the ALM method does show



(a)Gaussian signals



(b)Low-pass signals

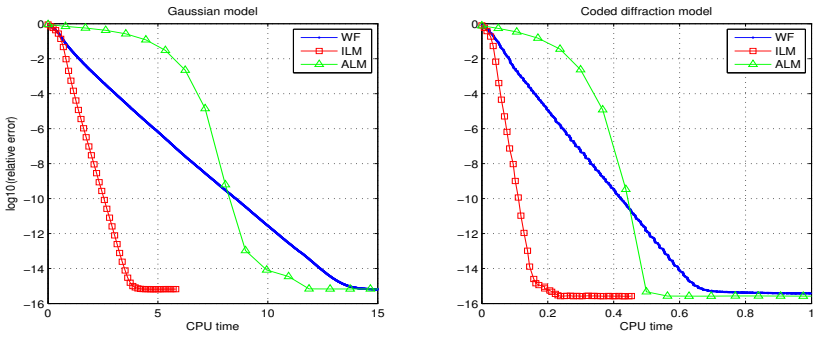
FIG. 7.2. *Relationship between the relative errors and the number of iterations.* $m/n = 6$ *for Gaussian model and* $L = 10$ *for CD model.*

Signal	Gaussian signal				Low-pass signal			
model	Gaussian		Coded diffraction		Gaussian		Coded diffraction	
	iter	CPU	iter	CPU	iter	CPU	iter	CPU
ALM	8.39	20.40s	6.41	0.53s	7.95	30.02s	5.64	0.79s
ILM	10.07	2.47s	6.58	0.06s	8.70	2.88s	6.04	0.07s
WF	300.79	7.14s	171.79	0.21s	229.08	8.94s	134.62	0.27s

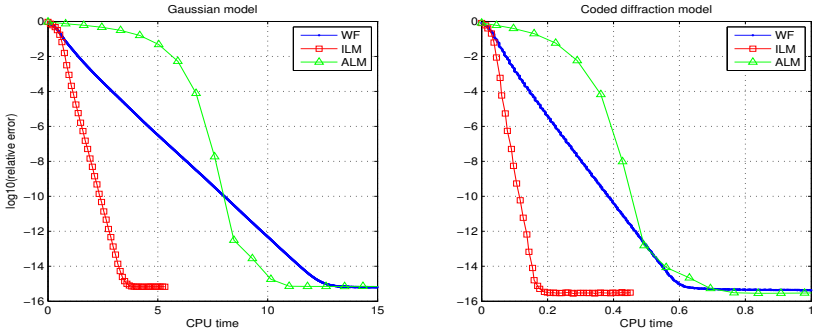
TABLE 7.1
Computational results on random examples

quadratic convergence. As it is expected, the ILM method shows linear convergence after the first several iterations. However, its convergence rate is much faster than that of the WF method. Table 7.1 presents the averaged number of iterations of the LM and WF methods used to achieve an accuracy of 10^{-5} under a fixed m/n or L. In the table, the statistics of the ALM method is approximately proportional to the logarithm of that of the WF method, which shows the quadratic convergence of the ALM method. Although the ILM method converges linearly, it converges fast and does not take many more iterations than the accurate algorithm.

We should point out that Figure 7.2 is not very fair to the WF method, since the gradient method tends to takes a large number of iterations. Therefore, we show the relationship between the relative error and CPU time in Figure 7.3. From this figure, we can see that, although the ALM method converges quadratically, it consumes much more CPU time than the other two methods because solving the LM equation accurately needs a lot of PCG iter-



(a)Gaussian signals



(b)Low-pass signals

FIG. 7.3. *Relationship between the relative errors and CPU time.* $m/n = 6$ *for Gaussian model and* $L = 10$ *for CD model.*

ations. On the other side, the ILM method consumes the smallest CPU time. Table 7.1 also shows the averaged CPU time (it does not include the CPU time of the initialization step) of the three methods to make a successful recovery. We still can see that the ALM method is the most time-consuming method, while the ILM method tends to take much less CPU time (about 1/3 or less) than the WF method. Obviously, solving the LM equation is expensive although a promising PCG is employed, and making a suitable truncation to the PCG method can efficiently reduce the computational cost.

7.2. Performance on natural image. We next perform a few numerical experiments on recovering natural images, similar to Section IV.C of [9]. The two images that we use are colored photographs of the Turret of Palace Museum ("turret") and the Milky Way Galaxy ("galaxy"). The colored images are viewed as $n_1 \times n_2 \times 3$ arrays, where the first two indices encode the pixel location, and the last is the color band. We run the LM and WF methods on each of the three RGB images.

We generate $L = 20$ random octanary patterns and gather the CD patterns for each color band using these 20 samples. We run 50 iterations of the power method in the initialization step. For the ALM method, the stopping tolerance of the PCG is set to $\eta_k = 10^{-16}$. For the ILM method, the maximum iteration number of PCG is set to 5. For the WF method, the step length is $\mu_k = \min(1 - \exp(-k/\tau_0), 0.4)$, where $\tau_0 \approx 330$. We perform 25 LM iterations and 300 WF iterations. The relative error is calculated by $\|\bar{x} - x\|_F / \|x\|_F$ where $\bar{x}$ and x are the recovered and original image, respectively. The CPU time is an average of the CPU time from three RGB images.

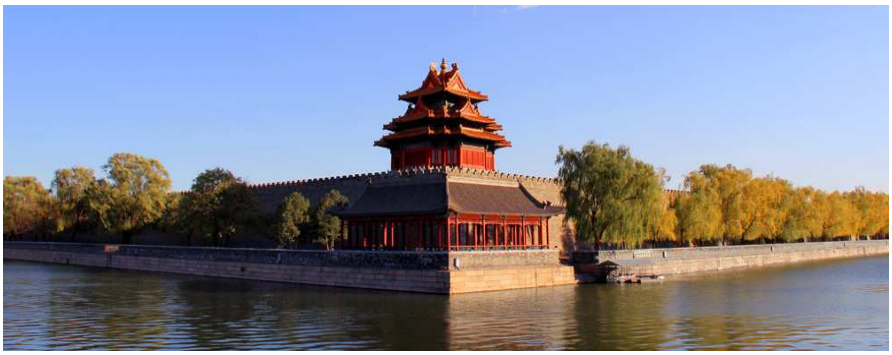


FIG. 7.4. *Turret of Palace Museum. The image size is 352×1000 pixels.*

FIG. 7.5. *The Milky Way Galaxy. The image size is 1080×1920 pixels.*

Figure 7.4 and 7.5 show the image "turret" and "galaxy" recovered by the ILM, respectively. The images recovered by the other two algorithms are not reported because they are similar. Table 7.2 shows the average number of iterations and average CPU time used by the three algorithms to reduce the relative error to 10^{-5} and 10^{-10} for the three color band. We can see an obvious advantage of the ILM method over the WF method. However, the CPU time of the ALM method is much larger than the other two methods.

Figure 7.6 shows the relationship between the relative errors and the number of iterations. It obviously demonstrates quadratic convergence of the ALM method and fast linear convergence of the ILM method. In particular, the ALM method takes about one iteration to reduce the accuracy from 10^{-5} to 10^{-10}. Figure 7.7 shows the relationship between the CPU time and the relative errors of the three methods. Due to the inexactness in solving the LM equation, the ILM method takes much less than than the ALM method.

7.3. Phase retrieval with noise. We now evaluate the numerical performance of the LM methods when there exists noises in the observation. We add different level of noises to $\{y_r\}$ and explore the relationship between the signal-to-noise rate (SNR) of the observation and the mean square error (MSE) of the recovered solution. Specifically, SNR and MSE are

Image	Turret of Palace Musuem				The Milky Way Galaxy			
Criterion	10^{-5}		10^{-10}		10^{-5}		10^{-10}	
	iter	CPU	iter	CPU	iter	CPU	iter	CPU
ALM	5.3	2244.80s	6.3	2664.72s	5.0	3605.58s	6.0	4316.84s
ILM	6.3	349.28s	9.3	519.69s	5.7	1071.86s	9.0	1701.89s
WF	136.3	685.46s	216.3	1117.15s	136.0	2396.80s	216.0	3798.73s

TABLE 7.2
Computational results in natural image recovering.

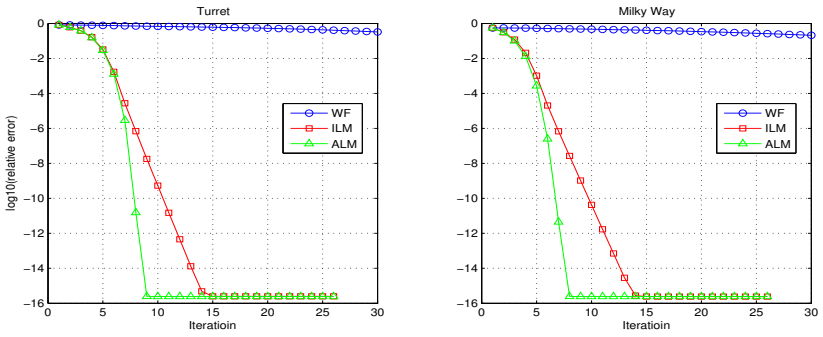


FIG. 7.6. *Relationship between the relative errors and the number of iterations for natural images recovery.*

calculated by

$$\text{MSE} := \frac{\mathbf{dist}^2(\tilde{x}, x)}{\|x\|^2}, \quad \textit{and} \quad \text{SNR} := \frac{\sum_{i=1}^m |a_r^* x|^4}{\|w\|^2}, \tag{7.1}$$

where $\tilde{x}$ is the output of the LM methods after 50 iterations or of the WF method after 2500 iterations, and w the added noise. The dB-scale of MSE and SNR is calculated by $20\log$ MSE and $10\log$ SNR, respectively. We construct random signals with $n = 512$, set $m = 6n$ for Gaussian model and $L = 12$ for the CD model. The SNR is varied from 10db to 60db. For each case, 100 Monte Carlo trials are repeated. Figure 7.8 shows the results on the change of MSE versus SNR. It shows that both algorithms achieve a similar order of accuracy. In fact, both algorithms can converge to the same minimum $\mathbf{x}$ with high probability.

8. Conclusion and future work. In this paper, we develop a modified LM method via Wirtinger derivative to solve the phase retrieval problem. Starting from the same spectral initialization step as the WF method, our method converges to the global solution linearly under the same assumption as the WF method. The convergence rate is further improved to be quadratic in a predictable neighborhood of the solution. Similar theoretical analysis holds even if the LM equation is solved inexactly. In particular, a simple yet useful preconditioner is constructed based on the expectation of the LM coefficient matrix by assuming the independence between measurements and the LM iteration. Since the restricted condition number of this preconditioned coefficient matrix is small, it enables a fast convergence of the PCG method for solving the LM equation.

In our numerical experiments, we verify that the proposed LM method indeed converges quadratically in recovering both random examples and natural images if the LM equation is solved sufficiently accurate. Our inexact LM method is comparable to the WF method in terms of the success rate and it has advantage in terms of the CPU time.

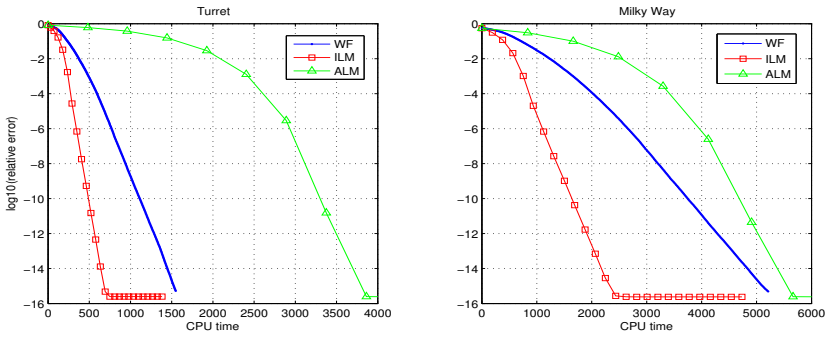


FIG. 7.7. *Relationship between the relative errors and CPU time for natural images recovery.*

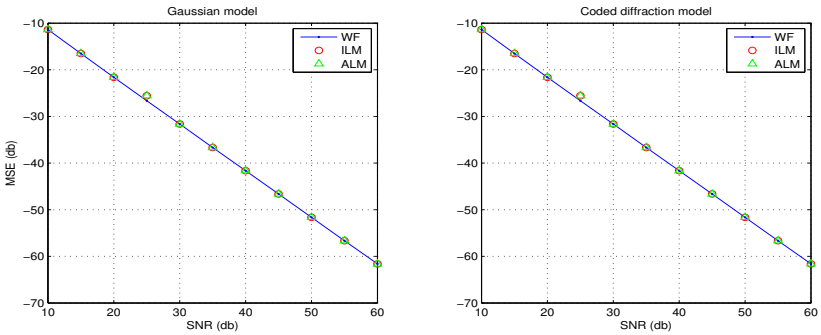


FIG. 7.8. *SNR versus relative MSE on a dB-scale for the LM and WF methods*

Our algorithm and analysis can be improved in several aspects. Theoretically, it is worth studying whether the convergence still holds globally if the LM regularization parameters μ_k are chosen simpler than the current ones since our numerical experiments show that $\mu_k = \sqrt{f(z_k)}$ performs well. Practically, it is meaningful to improve the success rate and reduce the computational complexity.

Acknowledgements. The authors would like to thank Prof. Xiaodong Li and Prof. Zhiqiang Xu for the insightful discussions on phase retrieval.

REFERENCES

[1] A. AHMED, B. RECHT, AND J. ROMBERG, *Blind deconvolution using convex programming*, Information Theory IEEE Transactions on, 60 (2012), pp. 1711 – 1732.

[2] R. BALAN, *On signal reconstruction from its spectrogram*, in Information Sciences and Systems (CISS), 2010 44th Annual Conference on, IEEE, 2010, pp. 1–4.

[3] R. BALAN, P. CASAZZA, AND E. DAN, *On signal reconstruction without phase*, Applied & Computational Harmonic Analysis, 20 (2006).

[4] H. H. BAUSCHKE, P. L. COMBETTES, AND D. R. LUKE, *Phase retrieval, error reduction algorithm, and fienup variants: a view from convex optimization*, J.opt.soc.amer.a, 19 (2002), pp. 1334–1345.

[5] ———, *Hybrid projection-reflection method for phase retrieval.*, Journal of the Optical Society of America A, 20 (2003), pp. 1025–1034.

[6] V. BENTKUS, *An inequality for tail probabilities of martingales with differences bounded from one side*, Journal of Theoretical Probability, 16 (2003), pp. 161–173.

[7] J. T. BETTS, *Solving the nonlinear least square problem: Application of a general method*, Journal of Optimization Theory and Applications, 18 (1976), pp. 469–483.

[8] O. BUNK, A. DIAZ, F. PFEIFFER, C. DAVID, B. SCHMITT, D. K. SATAPATHY, AND V. D. V. JF, *Diffractive imaging for periodic samples: retrieving one-dimensional concentration profiles across microfluidic channels*, Acta Crystallographica, 63 (2007), pp. 306–314.
[9] E. J. CANDES, X. LI, AND M. SOLTANOLKOTABI, *Phase retrieval via wirtinger flow: Theory and algorithms*, Information Theory IEEE Transactions on, 61 (2014), pp. 1985–2007.
[10] A. CHAI, M. MOSCOSO, AND G. PAPANICOLAOU, *Array imaging using intensity-only measurements*, Inverse Problems, 27 (2011), p. 015005.
[11] J. CORBETT, *The pauli problem, state reconstruction and quantum-real numbers*, Reports on Mathematical Physics, 57 (2006), pp. 53–68.
[12] W. C. DAVIDON, *New least-square algorithms*, Journal of Optimization Theory & Applications, 18 (1976), pp. 187–197.
[13] L. DEMANET AND V. JUGNON, *Convex recovery from interferometric measurements*, arXiv preprint arXiv:1307.6864, (2013).
[14] J. E. DENNIS, D. M. GAY, AND R. E. WELSCH, *Algorithm 573nl2sol—an adaptive nonlinear least-squares algorithm [e4]*, Acm Transactions on Mathematical Software, 7 (1981), pp. 369–383.
[15] V. ELSER, *Phase retrieval by iterated projections.*, Journal of the Optical Society of America A Optics Imageence & Vision, 20 (2003), pp. 40–55.
[16] J. FAN AND J. PAN, *Convergence properties of a self-adaptive levenberg-marquadt algorithm under local error bound condition*, Computational Optimization and Applications, 34 (2006), pp. 47–62.
[17] J. FAN, L. XUE, AND H. ZOU, *Strong oracle optimality of folded concave penalized estimation*, Annals of statistics, 42 (2014), p. 819.
[18] J. Y. FAN AND Y. X. YUAN, *On the quadratic convergence of the levenberg-marquardt method without nonsingularity assumption*, Computing, 74 (2005), pp. 23–39.
[19] C. FIENUP AND J. DAINTY, *Phase retrieval and image reconstruction for astronomy*, Image Recovery: Theory and Application, (1987), pp. 231–275.
[20] J. R. FIENUP, *Reconstruction of an object from the modulus of its fourier transform*, Optics Letters, 3 (1978), pp. 27–29.
[21] ———, *Phase retrieval algorithms: a comparison*, Applied Optics, 21 (1982), pp. 2758–2769.
[22] R. FLETCHER, *Practical methods of optimization*, John Wiley & Sons, 2013.
[23] R. FLETCHER AND C. XU, *Hybrid methods for nonlinear least squares*, IMA Journal of Numerical Analysis, 7 (1987), pp. 371–389.
[24] R. W. GERCHBERG, *A practical algorithm for the determination of phase from image and diffraction plane pictures*, Optik, 35 (1972), pp. 237–250.
[25] R. W. HARRISON, *Phase problem in crystallography*, Journal of the Optical Society of America, Part A: Optics and Image Science; (United States), 10:5 (1993), pp. 1046–1055.
[26] T. HEINOSAARI, L. MAZZARELLA, AND M. M. WOLF, *Quantum tomography under prior information*, Communications in Mathematical Physics, 318 (2013), pp. 355–374.
[27] M. R. HESTENES AND E. STIEFEL, *Methods of conjugate gradients for solving linear systems*, J.res.nat.bur.stand, 49 (1952), pp. 409–436.
[28] J. E. D. JR, H. J. MARTINEZ, AND R. A. TAPIA, *Convergence theory for the structured bfgs secant method with an application to nonlinear least squares*, Journal of Optimization Theory & Applications, 61 (1989), pp. 161–178.
[29] K. LEVENBERG, *A method for the solution of certain non-linear problems in least squares*, Quart.appl.math, 2 (1944), pp. 164–168.
[30] P. L. LOH AND M. J. WAINWRIGHT, *Regularized m-estimators with nonconvexity: Statistical and algorithmic theory for local optima*, Eprint Arxiv, 16 (2013), pp. 476–484.
[31] D. R. LUKE, *Relaxed averaged alternating reflections for diffraction imaging*, Inverse Problems, 21 (2005), p. 37.
[32] S. MARCHESINI, *Phase retrieval and saddle-point optimization.*, Journal of the Optical Society of America A Optics Image Science & Vision, 24 (2007), pp. 3289–3296.
[33] D. W. MARQUARDT, *An algorithm for least-squares estimation of nonlinear parameters*, J. Soc. Indust. Appl. Math., 11 (1963), pp. 431–441.
[34] J. MIAO, T. ISHIKAWA, Q. SHEN, AND T. EARNEST, *Extending x-ray crystallography to allow the imaging of noncrystalline materials, cells, and single protein complexes.*, Annual Review of Physical Chemistry, 59 (2008), pp. 387–410.
[35] R. P. MILLANE, *Phase retrieval in crystallography and optics*, Journal of the Optical Society of America A, 7 (1990), pp. 394–411.
[36] L. NAZARETH, *Some recent approaches to solving large residual nonlinear least squares problems*, Siam Review, 22 (1980), pp. 1–11.
[37] P. NETRAPALLI, P. JAIN, AND S. SANGHAVI, *Phase retrieval using alternating minimization*, IEEE, (2015).
[38] POINCARÉ, *Sur les propriétés du potentiel et sur les fonctions abéliennes"*, Acta Mathematica (in French), 22 (1899), pp. 89–178.

[39] H. REICHENBACH, *Philosophic foundations of quantum mechanics*, Courier Corporation, 1944.
[40] Y. SHECHTMAN, A. BECK, AND Y. C. ELDAR, *Gespar: Efficient phase retrieval of sparse signals*, IEEE Transactions on Signal Processing, 62 (2013), pp. 928 – 938.
[41] E. STIEFEL, 3 (1951), pp. 1–33.
[42] R. SUN AND Z. Q. LUO, *Guaranteed matrix completion via non-convex factorization*, Eprint Arxiv, (2014).
[43] A. WALTHER, *The question of phase retrieval in optics*, Optica Acta, 10 (1963), pp. 41–49.
[44] Z. WEN, C. YANG, X. LIU, AND S. MARCHESINI, *Alternating direction methods for classical and ptychographic phase retrieval*, Inverse Problems, 28 (2012), pp. 4666–4669.
[45] N. YAMASHITA AND M. FUKUSHIMA, *On the Rate of Convergence of the Levenberg-Marquardt Method*, Springer Vienna, 2001.
[46] C.-H. ZHANG, T. ZHANG, ET AL., *A general theory of concave regularization for high-dimensional sparse estimation problems*, Statistical Science, 27 (2012), pp. 576–593.