Iterative Algorithms for Ptychographic Phase Retrieval

Chao Yang [*]   Jianliang Qian [†]   Andre Schirotzek [‡]   Filipe Maia [§]   Stefano Marchesini [¶]

May 26, 2011

Abstract

Ptychography promises diffraction limited resolution without the need for high resolution lenses. To achieve high resolution one has to solve the phase problem for many partially overlapping frames. Here we review some of the existing methods for solving ptychographic phase retrieval problem from a numerical analysis point of view, and propose alternative methods based on numerical optimization.

1   Introduction

An emerging imaging technique in X-ray science is to use a localized moving probe to collect multiple diffraction measurements of an unknown object [1, 2, 3, 4, 5, 6, 7, 8]. This technique is called "ptychography". In a ptychography experiment, one collects a sequence of diffraction images of dimension $m \times m$. Each image frame $y_{\mathbf{x}}(\mathbf{r}')$ represents the magnitude of the Fourier transform of $a(\mathbf{r})\hat{\psi}(\mathbf{r} + \mathbf{x})$, where $a(\mathbf{r})$ is a localized illumination (window) function or a probe, $\hat{\psi}(\mathbf{r})$ is the unknown object of interest, and $\mathbf{x}$ is a translational vector. We can express $y_{\mathbf{x}}$ as

$$y_{\mathbf{x}}(\mathbf{r}') = |\mathcal{F}\{a(\mathbf{r})\hat{\psi}(\mathbf{r} + \mathbf{x})\}|, \tag{1}$$

[*]Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720. Email: cyang@lbl.gov

[†]Department of Mathematics, Michigan State University, East Lansing, MI. Email: qian@math.msu.edu

[‡]Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, CA 94720. Email: aschirotzek@lbl.gov

[§]NERSC, Lawrence Berkeley National Laboratory, Berkeley, CA 94720. Email: frmaia@lbl.gov

[¶]Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, CA 94720. Email: smarchesini@lbl.gov

where $\mathcal{F}\{f\}$ denotes the Fourier transform of f with respect to $\mathbf{r}$.

In order to reconstruct the unknown object, we must retrieve the phases of the measured images. A number of methods have been proposed to recover $\hat{\psi}(\mathbf{r})$ from ptychographic measurements $y_{\mathbf{x}}(\mathbf{r}')$ [9, 4, 5, 7, 8]. The connection among these methods is not entirely clear from the existing literature. Furthermore, little detail is provided on the convergence rate or computational efficiency of these methods.

In this paper, we review some of the existing methods for solving ptychographic phase retrieval problem from a numerical analysis point of view, and propose to solve the problem by alternative methods that are standard in the numerical optimization community. In particular, we formulate the ptychographic phase retrieval problem as an unconstrained nonlinear minimization problem in section 2, and compare the convergence of several well known iterative methods for solving this type of problem in section 6. We discuss computational details such as line search and preconditioning that are important for achieving optimal performance in these methods in section 2. We also describe the connection between optimization based algorithms and projection algorithms that are often discussed in the phase retrieval literature in section 4.

We point out that ptychographic minimization problem is not globally convex, which means that iterative methods can be trapped at a local minimizer if a poor starting guess is chosen. We show by a numerical example that one way to escape from a local minimizer is to switch to a different objective function in section 6.

We observed that the convergence of the optimization based iterative algorithms used to perform ptychographic phase retrieval is accelerated when the amount of overlap between two adjacent image frames increases. We provide an intuitive explaination on why the amount of overlap between adjacent frames affects the convergence of iterative optimization algorithms in section 6.

An alternative approach for performing ptychographic phase retrieval is a method known as Wigner deconvolution. We review this approach in section 5 and point out its connection to iterative algorithms and its limitations.

We use standard linear algebra notation whenever possible to describe various quantities evaluated in the iterative algorithms we present. To simplify notation we use a/b to denote an element-wise division between two vectors a and b. Similarly, we use $a \cdot b$ to denote an element-wise multiplication of a and b. We also use a^2 and $a^{1/2}$ occasionally to denote the element-wise square and square root of a respectively. The conjugate of a complex variable a is denoted by $\bar{a}$. The real part of a is denoted by $\mathrm{Re}(a)$. The conjugate transpose of a matrix (or a vector) A is denoted by A^*. The $|x|$ symbol is reserved for the magnitude (or absolute value) of x. The Euclidean norm of x is denoted by $\|x\| = \sqrt{x^*x}$. We use $\mathrm{Diag}\,(x)$ to represent a diagonal matrix with the vector x on its diagonal.

2 Ptychographic reconstruction formulated as an optimization problem

The problem we would like to solve is to recover $\hat{\psi}$ from a number of intensity measurements represented by (1). For a finite set of translational vectors $\mathbf{x}_i$, we will denote each measurement by

$$b_i = |FQ_i\hat{\psi}|, \quad i = 1, 2, ..., k,$$

where $\hat{\psi}$ is the sampled unknown object that contains n pixels, b_i is a sampled measurement that contains m pixels, F is the matrix representation of a discrete Fourier transform, and Q_i

is an $m \times n$ "illumination matrix" that extracts a frame containing m pixels out of an image containing n pixels. Each row of Q_i contains at most one nonzero element. The nonzero values in Q_i are determined by the illumination function $a(\mathbf{r})$.
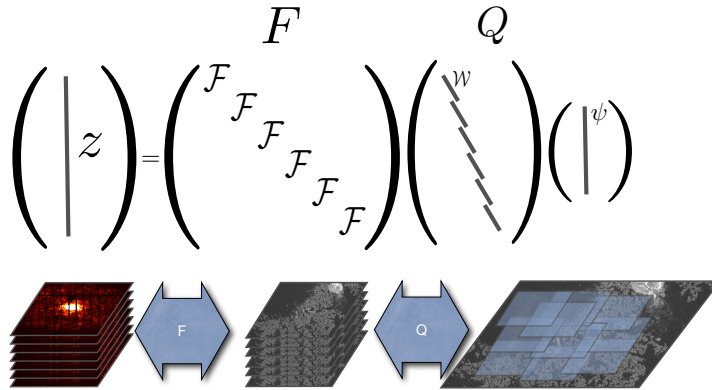

Figure 1: An unknown object of interest $\hat{\psi}$, and the measured amplitudes z_x related by matrix operations

Given a set of measurements, b_1, b_2, ..., b_k, we may attempt to recover $\hat{\psi}$ by solving the least squares problem

$$\min_{\psi} \frac{1}{2} \sum_{i=1}^{k} \| |z_i| - b_i \|^2, \tag{2}$$

where $z_i \equiv FQ_i\psi$, and the factor of $1/2$ is included here merely for convenience.

An alternative objective function we may use to recover $\hat{\psi}$ is

$$\epsilon = \frac{1}{2} \sum_{i=1}^{k} \| |z_i|^2 - b_i^2 \|^2, \tag{3}$$

where $|z_i|^2$ and b_i^2 denote vectors obtained from squaring each component of $|z_i|$ and b_i respectively. The advantage of using (3) is that it is more differentiable, hence more amenable to analysis. In practice, we found the objective function in (2) to be a better choice in terms of computational efficiency in most cases.

To obtain the minimizers of (2) or (3) using numerical optimization techniques, we often need to evaluate the gradient and possibly the Hessian of these objective functions. Because both (2) and (3) are real-valued functions of a (potentially) complex vector ψ, derivative calculations must be done with care. One can either take the derivative of (2) and (3) with respect to the real and imaginary parts of ψ independently or follow the $\mathbb{CR}$-calculus formalism established in [10, 11] by treating ψ and $\bar{\psi}$ as two independent variables. The latter approach is what we use throughout this paper.

2.1 Gradient

If we let $r_i \equiv |z_i|^2 - b_i^2$, and define

$$r \equiv \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_k \end{pmatrix},$$

we can rewrite (3) as $\epsilon(\psi) = r^T r/2$. Let the matrix $J_i = \partial r_i/\partial \psi$ be the Jacobian of r_i with respect to ψ. It follows from the chain rule that the gradient of ϵ in vector form is

$$\nabla \epsilon(\psi) = \left(\frac{\partial \epsilon}{\partial \psi}\right)^* = \left(\frac{\partial \epsilon}{\partial r}\frac{\partial r}{\partial \psi}\right)^* = J^* r, \tag{4}$$

where

$$J \equiv \begin{pmatrix} J_1 \\ J_2 \\ \vdots \\ J_k \end{pmatrix}.$$

Note that we may rewrite $|z_i|^2$ as $\mathrm{Diag}(z_i)^* z_i$, where $\mathrm{Diag}(x)$ denotes a diagonal matrix that contains the vector x on its diagonal and $z_i \equiv FQ_i\psi$. Using this observation, we can show that

$$J_i = \frac{\partial |z_i|^2}{\partial \psi} = \frac{\partial \mathrm{Diag}\,(\bar{z}_i)\,z_i}{\partial z_i}\frac{\partial z_i}{\partial \psi} + \frac{\partial \mathrm{Diag}\,(z_i)\,\bar{z}_i}{\partial \bar{z}_i}\frac{\partial \bar{z}_i}{\partial \psi} = \mathrm{Diag}(\bar{z}_i)FQ_i = \mathrm{Diag}(FQ_i\psi)^* FQ_i. \tag{5}$$

It follows from (4) and the above expression that

$$\nabla \epsilon = \sum_{i=1}^{k} Q_i^* F^* \mathrm{Diag}(z_i)[|z_i|^2 - b_i^2]. \tag{6}$$

The gradient of the objective function in (2), which we will denote by $\rho(\psi)$, is slightly more complicated. By rewriting $|z_i|$ as $(|z_i|^2)^{1/2}$, with the understanding that the square root is taken component-wise, and by using the chain rule and replacing $\partial |z_i|^2/\partial \psi$ with the expression given in (5), we obtain

$$J_i = \frac{\partial |z_i|}{\partial \psi} = \frac{\partial (|z_i|^2)^{1/2}}{\partial |z_i|^2} \cdot \frac{\partial |z_i|^2}{\partial \psi} = \frac{1}{2}\mathrm{Diag}\,(\bar{z}_i/|z_i|)\,FQ_i,$$

if $|z_i|$ does not contain any zero element for all $i = 1, 2, ..., m$.

Consequently, we may now express $\nabla \rho(\psi)$ as

$$\begin{aligned}
\nabla \rho(\psi) = J^* r &= \sum_{i=1}^{k} J_i^* r_i \\
&= \frac{1}{2}\sum_{i=1}^{k} Q_i^* F^* \mathrm{Diag}\left(\frac{z_i}{|z_i|}\right)(|z_i| - b_i) \\
&= \frac{1}{2}\sum_{i=1}^{k} Q_i^* F^* \left[z_i - \mathrm{Diag}\left(\frac{z_i}{|z_i|}\right)b_i\right] \\
&= \frac{1}{2}\sum_{i=1}^{k} Q_i^* F^* \left[FQ_i\psi - \mathrm{Diag}\left(\frac{z_i}{|z_i|}\right)b_i\right] \\
&= \frac{1}{2}\sum_{i=1}^{k} \left[Q_i^* Q_i\psi - Q_i^* F^* \mathrm{Diag}\left(\frac{z_i}{|z_i|}\right)b_i\right]. \tag{7}
\end{aligned}$$

Recall that $z_i = FQ_i\psi$. Thus, the expression $\mathrm{Diag}(FQ_i\psi)\mathrm{Diag}(|z_i|)^{-1}b_i$ simply represents projecting ψ onto the Fourier magnitude constraint imposed by the data b_i. Note that the

4

expression given above for the gradient of $\rho(\psi)$ is only valid when $|z_i|$ does not contain any zero for all $i = 1, 2, ..., m$. If $|z_i|$ contains a zero component for some i, and if the corresponding component in b_i is nonzero, $\nabla\rho$ is not well defined, i.e., $\nabla\psi$ has singularities at ψ's where $FQ_i\psi$ contains a zero element for some i.

Note that both (6) and (7) remain real when ψ is real and when b_i is obtained from a discrete Fourier transform of a real image (so that conjugate symmetry is preserved in $\mathrm{Diag}\,(z_i/|z_i|)\,b_i$.)

The directional derivatives of ϵ and ρ along a direction ϕ are defined by

$$\frac{\partial\epsilon}{\partial\psi}\phi + \frac{\partial\epsilon}{\partial\bar\psi}\bar\phi = 2\sum_{i=1}^{k}\mathrm{Re}\left[(|z_i|^2 - b_i^2)^T\mathrm{Diag}\,(z_i)^*\,FQ_i\phi\right] \tag{8}$$

and

$$\frac{\partial\rho}{\partial\psi}\phi + \frac{\partial\rho}{\partial\bar\psi}\bar\phi = \sum_{i=1}^{k}\mathrm{Re}\left[\phi^*Q_i^*Q_i\psi - \phi^*Q_i^*F^*\mathrm{Diag}\left(\frac{z_i}{|z_i|}\right)b_i\right] \tag{9}$$

respectively

2.2   Hessian

The Hessian of $\epsilon(\psi)$ and $\rho(\psi)$ provides information on the convexity of these objective functions. A globally convex function has a unique minimizer. Such a minimizer can be obtained by standard optimization techniques that we will describe in the next section. If the objective function is not convex everywhere, a standard optimization algorithm may produce a local minimizer that is not the true solution to ptychographic reconstruction problem.

Again, because both $\epsilon(\psi)$ and $\rho(\psi)$ are real valued functions of a potentially complex vector ψ, their Hessians are defined as

$$H^o = \left(\begin{array}{cc} H^o_{\psi\psi} & H^o_{\psi\bar\psi} \\ H^o_{\bar\psi\psi} & H^o_{\bar\psi\bar\psi} \end{array}\right),$$

where

$$H^o_{\psi\psi} \equiv \frac{\partial}{\partial\psi}\left(\frac{\partial f}{\partial\psi}\right)^*, \quad H^o_{\bar\psi\psi} \equiv \frac{\partial}{\partial\bar\psi}\left(\frac{\partial f}{\partial\psi}\right)^*, \quad H^o_{\psi\bar\psi} \equiv \frac{\partial}{\partial\psi}\left(\frac{\partial f}{\partial\bar\psi}\right)^*, \quad H^o_{\bar\psi\bar\psi} \equiv \frac{\partial}{\partial\bar\psi}\left(\frac{\partial f}{\partial\bar\psi}\right)^*,$$

and o is either ϵ or ρ.

It is not difficult to show that

$$H^\epsilon_{\psi\psi} = \sum_i Q_i^*F^*\mathrm{Diag}\left(2|z_i|^2 - b_i^2\right)FQ_i, \tag{10}$$

$$H^\epsilon_{\bar\psi\bar\psi} = \sum_i Q_i^TF^T\mathrm{Diag}\left(2|z_i|^2 - b_i^2\right)\bar F\bar Q_i, \tag{11}$$

$$H^\epsilon_{\psi\bar\psi} = \sum_i Q_i^*F^*\mathrm{Diag}\left(z_i\right)^2\bar F\bar Q_i, \tag{12}$$

$$H^\epsilon_{\bar\psi\psi} = \left(H^\epsilon_{\psi\bar\psi}\right)^* = \sum_i Q_i^TF^T\mathrm{Diag}\left(\bar z_i\right)^2 FQ_i. \tag{13}$$

If we let $t_{ji} \equiv |t_{ji}|e^{i\mu_{ji}}$, $\zeta_{ji} \equiv |\zeta_{ji}|e^{i\theta_{ji}}$ and β_{ji} be the jth component of $t_i = FQ_i\phi$, $z_i = FQ_i\psi$ and b_i respectively, then the curvature $\tau_\epsilon(\psi, \phi)$ at ψ along any direction ϕ can be

calculated as follows

$$
\begin{aligned}
\tau_\epsilon(\psi, \phi) &= (\phi^* \ \phi^T) \begin{pmatrix} H_{\psi\psi}^\epsilon & H_{\psi\bar\psi}^\epsilon \\ H_{\bar\psi\psi}^\epsilon & H_{\bar\psi\bar\psi}^\epsilon \end{pmatrix} \begin{pmatrix} \phi \\ \bar\phi \end{pmatrix} \\
&= \sum_i (t_i^* \ t_i^T) \begin{pmatrix} \mathrm{Diag}\left(2|z_i|^2 - b_i^2\right) & \mathrm{Diag}\left(z_i\right)^2 \\ \mathrm{Diag}\left(\bar z_i\right)^2 & \mathrm{Diag}\left(2|z_i|^2 - b_i^2\right) \end{pmatrix} \begin{pmatrix} t_i \\ \bar t_i \end{pmatrix} \\
&= \sum_i 2 t_i^* \mathrm{Diag}\left(2|z_i|^2 - b_i^2\right) t_i + 2\mathrm{Re}[t_i^T \mathrm{Diag}\left(\bar z_i\right)^2 t_i] \\
&= \sum_i 2 t_i^* \mathrm{Diag}\left(|z_i|^2 - b_i^2\right) t_i + 2\left(t_i^* \mathrm{Diag}\left(|z_i|\right)^2 t_i + \mathrm{Re}[t_i^T \mathrm{Diag}\left(\bar z_i\right)^2 t_i] \right) \\
&= 2\sum_{i=1}^k \sum_{j=1}^n |t_{ji}|^2 (|z_{ji}|^2 - \beta_{ji}^2) + \left(|t_{ji}|^2 |z_{ji}|^2 + \mathrm{Re}\left[(t_{ji}\bar z_{ji})^2 \right] \right) \\
&= 2\sum_{i=1}^k \sum_{j=1}^n |t_{ji}|^2 (|z_{ji}|^2 - \beta_{ji}^2) + 2|t_{ji}|^2 |z_{ji}|^2 \cos^2(\mu_{ji} - \theta_{ji}). \quad (14)
\end{aligned}
$$

At the minimizer of $\epsilon(\psi)$, $|z_i| = b_i$. So the first term of (14) is zero. Because the second term of (14) is nonnegative, $\tau \geq 0$, i.e., ϵ is convex at the solution. Moreover, the convexity of ϵ is preserved in the area where $|z_{ji}| \geq \beta_{ji}$.

A similar observation can be made from the curvature of ρ. It is not difficult to show that

$$
H_{\psi\psi}^\rho = \frac{1}{2}\left(\sum_{i=1}^k Q_i^* Q_i - \frac{1}{2} Q_i^* F^* \mathrm{Diag}\left(\frac{b_i}{|z_i|}\right) F Q_i \right), \quad (15)
$$

$$
H_{\bar\psi\bar\psi}^\rho = \frac{1}{2}\left(\sum_{i=1}^k Q_i^T \bar Q_i - \frac{1}{2} Q_i^T F^T \mathrm{Diag}\left(\frac{b_i}{|z_i|}\right) \bar F \bar Q_i \right), \quad (16)
$$

$$
H_{\psi\bar\psi}^\rho = \frac{1}{4} \sum_{i=1}^k Q_i^* F^* \mathrm{Diag}\left(\frac{b_i \cdot z_i^2}{|z_i|^3}\right) \bar F \bar Q_i, \quad (17)
$$

$$
H_{\bar\psi\psi}^\rho = \frac{1}{4} \sum_{i=1}^k Q_i^T F^T \mathrm{Diag}\left(\frac{b_i \cdot \bar z_i}{|z_i|}\right) F Q_i. \quad (18)
$$

It follows that

$$
\begin{aligned}
\tau_\rho(\psi,\phi) &= (\phi^* \ \phi^T) \begin{pmatrix} H_{\psi\psi}^\rho & H_{\psi\bar\psi}^\rho \\ H_{\bar\psi\psi}^\rho & H_{\bar\psi\bar\psi}^\rho \end{pmatrix} \begin{pmatrix} \phi \\ \bar\phi \end{pmatrix} \qquad (19) \\[2mm]
&= \frac{1}{2}\sum_i (t_i^* \ \bar t_i^T) \begin{pmatrix} I - \frac{1}{2}\mathrm{Diag}\left(\frac{b_i}{|z_i|}\right) & \frac{1}{2}\mathrm{Diag}\left(\frac{b_i}{|z_i|}\cdot\frac{z_i^2}{|z_i|^2}\right) \\ \frac{1}{2}\mathrm{Diag}\left(\frac{b_i}{|z_i|}\cdot\frac{\bar z_i^2}{|z_i|^2}\right) & I - \frac{1}{2}\mathrm{Diag}\left(\frac{b_i}{|z_i|}\right) \end{pmatrix} \begin{pmatrix} t_i \\ \bar t_i \end{pmatrix} \\[2mm]
&= \frac{1}{2}\sum_{i=1}^k \sum_{j=1}^n \left(2|t_{ji}|^2 - |t_{ji}|^2 \frac{\beta_{ji}}{|\zeta_{ji}|} + \mathrm{Re}\left[\bar t_{ji}^2 \frac{\beta_{ji}\zeta_{ji}^2}{|\zeta_{ji}|^3} \right] \right) \\[2mm]
&= \frac{1}{2}\sum_{i=1}^k \sum_{j=1}^n |t_{ji}|^2 \left(2 - \frac{\beta_{ji}}{|\zeta_{ji}|} + \frac{\beta_{ji}}{|\zeta_{ji}|}\mathrm{Re}\left[\frac{\bar t_{ji}^2}{|t_{ji}|^2} \frac{\zeta_{ji}^2}{|\zeta_{ji}|^2} \right] \right) \\[2mm]
&= \sum_{i=1}^k \sum_{j=1}^n |t_{ji}|^2 \left(1 - \frac{\beta_{ji}}{|\zeta_{ji}|}\sin^2(\mu_{ji} - \theta_{ji}) \right). \qquad (20)
\end{aligned}
$$

Thus, $\tau_\rho \geq 0$ when $|\zeta_{ji}| \geq \beta_{ji}$ for all $j = 1, 2, ..., n$ and $i = 1, 2, ..., k$. Even if $|\zeta_{ji}|$ is slightly less than β_{ji} for some j and i, τ_ρ may remain positive when the corresponding $\sin^2(\mu_{ji} - \theta_{ji})$ is sufficiently small and other terms in the summation in (20) are sufficiently large and positive.

A typical problem encountered in optics is when $k = 1$. When only one diffraction image is recorded, experience shows that local minima are common. Regions of negative curvature separate local minima from the global solution.

3 Iterative Algorithms based on Nonlinear Optimization

Because the gradient and Hessian of (2) and (3) are relatively easy to evaluate, we may use standard minimization algorithms such as the steepest descent method, the Newton's method and the nonlinear conjugate gradient method to find the solution to the ptychographic reconstruction problem. We will review some of these methods in section 3.1 and discuss some techniques for improving the performance of these algorithms in the rest of this section.

3.1 Basic Algorithms

In many standard numerical optimization algorithms, we construct a sequence of approximations to $\hat\psi$ by

$$ \psi^{(\ell+1)} = \psi^{(\ell)} + \beta p^{(\ell)}, \qquad (21) $$

where $p^{(\ell)}$ is a *search direction* along which the objective function (2) or (3) decreases, and $\beta > 0$ is an appropriate step length.

The simplest type of search direction is the *steepest descent* direction

$$ p_{\mathrm{sd}}^{(\ell)} = -\nabla_\psi o(\psi^{(\ell)}, \bar\psi^{(\ell)}), $$

where o is either ϵ or ρ. When the Hessian of ρ or ϵ is positive definite at $\psi^{(\ell)}$, the Newton's direction $p_{\mathrm{nt}}^{(\ell)}$, which is the solution of

$$ \begin{pmatrix} H_{\psi\psi}^o & H_{\psi\bar\psi}^o \\ H_{\bar\psi\psi}^o & H_{\bar\psi\bar\psi}^o \end{pmatrix} \begin{pmatrix} p_{\mathrm{nt}}^{(\ell)} \\ \bar p_{\mathrm{nt}}^{(\ell)} \end{pmatrix} = \begin{pmatrix} p_{\mathrm{sd}}^{(\ell)} \\ \bar p_{\mathrm{sd}}^{(\ell)} \end{pmatrix}, \qquad (22) $$

is also a descent direction.

Due to the nonlinear least squares nature of the objective functions (2) and (3), we may replace the true Hessian in (22) by a simpler approximation constructed from the Jacobian of the residual functions $|z_i| - b_i$ or $|z_i|^2 - b_i^2$ for $i = 1, 2, ..., k$. This technique yields the *Gauss-Newton* (GN) search directions.

Both Newton's method and the Gauss-Newton method require solving a system of linear equations at each step in order to obtain a search direction. Because the dimension of these linear systems is $n \times n$, where n is the number of pixels in the image to be reconstructed, constructing the Hessian or Jacobian and solving these equations by matrix factorization based methods will be prohibitively expensive. Iterative methods that make use of matrix vector multiplications without forming the Hessian or the J matrix explicitly are more appropriate. However, several iterations may be required to reach a desired accuracy needed to produce a good search direction. Hence methods based on Newton or Gauss-Newton search directions tend to be more expensive.

The Hessian in (22) can also be replaced by approximations constructed from changes in the gradient computed at each iteration. Such approximation yields *Quasi-Newton* search directions.

Another commonly used search direction is the conjugate gradient direction defined by

$$p_{\text{cg}}^{(\ell)} = -g^{(\ell)} + \alpha p_{\text{cg}}^{(\ell-1)},$$

where $g^{(\ell)}$ is the gradient of (2) or (3) at $\psi^{(\ell)}$ and α is often chosen to be

$$\alpha = \frac{\text{Re}\left[(g^{(\ell)})^*(g^{(\ell)} - g^{(\ell-1)})\right]}{\|g^{(\ell-1)}\|^2}.$$

This choice of α yields what is known as the *Polak-Ribiere* conjugate gradient method.

There are a variety of ways to choose an appropriate step length β in (21). They are often referred to as *line search* methods. The purpose of line search is to ensure that the objective function decreases as we move from $\psi^{(\ell)}$ to $\psi^{(\ell+1)}$ so that $\psi^{(\ell)}$ will converge to at least a local minimizer as ℓ increases. Such type of convergence is often called *global convergence.*

Another way to achieve global convergence in an iterative optimization procedure is to use the *trust region* technique to determine a search direction and step length simultaneously. Under the trust region framework, we minimize the true objective function by minimizing a sequence of simpler "surrogate" functions that mimic the behavior of the true objective function within a small neighborhood of the current approximations. That is, in each step of this iterative optimization procedure, we solve what is called a trust region subproblem

$$\min_{\|\phi\| \le \Delta} q(\psi^{(\ell)} + \phi), \tag{23}$$

where $q(\psi)$ is the surrogate function, and the parameter Δ is known as a trust region radius that defines the region in which $q(\psi)$ approximates $\rho(\psi)$ or $\epsilon(\rho)$ well. Such a radius must be chosen carefully. It may need to be adjusted iteratively based on the ratio of the reduction in the surrogate and the reduction in the true objective function achieved by the solution to (23).

A commonly used surrogate function is the second order Taylor expansion of the true objective function. The minimizer of this particular surrogate gives a full step Newton direction. However, the Newton step may not satisfy the trust region constraint, thus may not be the solution to (23).

8

The trust region subproblem can be solved either exactly or approximately depending on the cost of evaluating $q(\psi)$ and its derivatives. If the second order Taylor expansion is chosen as the surrogate, most methods need to solve the Newton equation

$$\nabla^2 q(\phi)s = -\nabla q(\phi),$$

where $\nabla^2 q$ is the Hessian of the true objective at the current iterate ψ. This equation can be solved approximately by the (linear) conjugate gradient algorithm when ∇^2 is positive definite. When $\nabla^2 q$ is not positive definite, (23) can also be solved by following a negative curvature direction to the boundary of the trust region. These techniques are used in an efficient iterative procedure for solving a large-scale trust region subproblem developed by Steihaug [12]. The method requires compute the matrix vector product $\nabla^2 qv$ for some vector v. This product can be approximated by a finite difference approximation

$$(\nabla^2 q)v \approx \frac{\nabla q(\phi + \eta v) - \nabla(\phi)}{\eta},$$

for some small η. Therefore, it is not necessary to explicitly construct the Hessian of the objective function in Steihaug's method.

3.2 Weighted Objective and Precondition

The least squares objective function in (2) and (3) can be expressed as

$$\rho(\psi) = \frac{1}{2} \sum_{i=1}^{k} \langle |z_i| - b_i, |z_i| - b_i \rangle,$$

and

$$\epsilon(\psi) = \frac{1}{2} \sum_{i=1}^{k} \langle |z_i|^2 - b_i^2, |z_i|^2 - b_i^2 \rangle$$

respectively, where $\langle x, y \rangle = x^* y$ denotes the standard Euclidean inner product. This inner product can be replaced by a weighted inner product $\langle x, y \rangle_B = x^* By$, where B is a symmetric positive definite matrix, to accelerate the convergence of iterative methods used to recover the unknown image ψ. As we will show in section 6, the choice of $B = \text{Diag}(b_i)^{-1}$ is particularly useful for accelerating the convergence of all iterative methods we have looked at. To maintain numerical stability and reduce noise amplification, it is often necessary to add a small constant to the diagonal of B to prevent it from becoming singular or ill-conditioned.

Another useful technique for accelerating iterative methods for solving unconstrained minimization problem is preconditioning. Instead of minimizing $\rho(\psi)$ or $\epsilon(\psi)$, we make a change of variable and minimize $\hat{\rho}(\phi)$ and $\hat{\epsilon}(\phi)$, where $\phi = K\psi$, and K is a preconditioner that is usually required to be Hermitian and positive definite. The purpose of introducing the preconditioner K is to reduce the condition number of the Hessian of the objective function. A highly ill-conditioned Hessian often leads to slow convergence of an iterative method. A well-known example is the zig-zag behavior of the steepest descent algorithm when it is applied to the Rosenbrock function.

It follows from the chain rule and (7) that the gradient of $\hat{\rho}(\psi)$ is simply

$$\nabla \hat{\rho}(\psi) = \frac{1}{2} K^{-1} \sum_{i=1}^{k} [Q_i^* Q_i \psi - Q_i^* F^* \text{Diag} \left(\frac{z_i}{|z_i|} \right) b_i],$$

9

where $z_i = FQ_i\psi$.

If we take the preconditioner to be the constant term on the diagonal blocks of $H_{\psi\psi}^{\rho}$, i.e.,

$$K = \sum_{i=1}^{k} Q_i^* Q_i, \tag{24}$$

which is a diagonal matrix, the gradient of $\hat{\rho}$ simply becomes

$$\nabla\hat{\rho}(\psi) = \frac{1}{2}\left[\psi - \left(\sum_{i=1}^{k} Q_i^* Q_i\right)^{-1}\left(\sum_{i=1}^{k} Q_i^* F^* \text{Diag}\left(\frac{z_i}{|z_i|}\right) b_i\right)\right],$$

and the corresponding preconditioned steepest descent algorithm with a constant step length of 2 yields the following updating formula:

$$\psi^{(\ell+1)} = \left(\sum_{i=1}^{k} Q_i^* Q_i\right)^{-1}\left(\sum_{i=1}^{k} Q_i^* F^* \text{Diag}\left(\frac{z_i^{(\ell)}}{|z_i^{(\ell)}|}\right) b_i\right),$$

where $z_i^{(\ell)} = FQ_i\psi^{(\ell)}$. This updating formula is identical to that used in the *error reduction* algorithm or *alternate projection* algorithm mentioned in the standard phase retrieval literature [18], which is guaranteed to converge to at least a local minimizer as shown in section 4.

3.3 Line Search

The global convergence of an unconstrained optimization algorithm depends on effective line search strategies. Assuming that ϕ is a descent direction for $\rho(\psi)$ at ψ, i.e., $\nabla\rho(\psi)^T\phi < 0$, we would like to seek an appropriate step length α so that

$$\rho(\psi + \alpha\phi) < \rho(\psi).$$

One way to obtain such a step length is to minimize the scalar function $\xi(\alpha) = \rho(\psi + \alpha\phi)$ with respect to α. This can be done by applying the Newton's method to generate a sequence of α's that satisfy

$$\alpha_{i+1} = \alpha_{i+1} - \frac{\xi'(\alpha_i)}{\xi''(\alpha_i)}, \tag{25}$$

and accepting an α_i that satisfies

$$\xi(\alpha_i) < c_1\xi(0), \quad \text{and} \quad |\xi'(\alpha_i)| < c_2|\xi'(\alpha_{i-1})|,$$

for some small constants $0 < c_1, c_2 < 1$. In order to obtain the values of $\xi'(\alpha_i)$ and $\xi''(\alpha_i)$ required in (25), we need to evaluate the directional derivative and curvature of ρ at $\psi + \alpha_i\phi$ along the search direction ϕ. That is,

$$\begin{aligned}
\xi'(\alpha_i) &= 2\text{Re}(\phi^* \nabla\rho(\psi + \alpha_i\phi)) \\
\xi''(\alpha_i) &= \tau_\rho(\psi + \alpha_i\phi, \phi).
\end{aligned}$$

Although these derivative calculations will incur additional computation, the cost of these computation can be kept at a minimal by keeping $FQ_i\phi$ in memory as we will discuss at the end of this section.

We should note that the Newton's method may not always succeed in finding an appropriate α due to the fact that $\xi(\alpha)$ is generally not globally convex. The convergence of the Newton's method will depend on the choice of the starting guess. When a good starting guess is chosen, we typically need to run only a few Newton iterations to reach a reasonable α value. Because the purpose of line search is to identify a step length that would lead to a sufficient reduction in the objective function, it is not necessary to find the actual minimizer of $\xi(\alpha)$.

However, an exact line search may not satisfy what is known as the second Wolfe condition

$$\nabla\rho(\psi + \alpha\phi)^* p \geq c_2 \nabla\rho(\psi)^T \phi,$$

where $0 < c_2 < 1$ is typically chosen to be 0.9. This condition on the change of the curvature of the objective function and the first Wolfe condition

$$\rho(\psi + \alpha\phi) \leq \rho(\psi) + c_1 \alpha \nabla\rho(\psi)^* \phi,$$

for some constant c_1 typically chosen to be 10^{-3}, which is a condition that guarantees a sufficient decrease in the objective function, are required to prove the global convergence of the sequence $\{\psi^{(\ell)}\}$ generated by (21) in many optimization algorithms. Line search techniques that satisfy both Wolfe conditions can be found in [13] and many other standard optimization textbooks [14]. We should note that these techniques may also be sensitive to the choice of the initial guesses to the step length as well as the choice of c_1 and c_2 parameters. When a poor initial guess is chosen, these techniques can yield α values that are too small. Strategies for choosing a good starting guess of α can be found in [14] also.

Regardless which line search technique one uses, one typically needs to evaluate the objective function $\epsilon(\psi + \alpha\phi)$ or $\rho(\psi + \alpha\phi)$ and its directional derivatives for a number of different α values. If we compute $\tilde{\psi} = \psi + \alpha\phi$ first and use the formulae given in (2), (3), (8) and (9) to evaluate the objective function and directional derivative (by replacing ψ with $\tilde{\psi}$), each evaluation will perform k FFTs. To reduce the cost of line search, we may evaluate $t_i = FQ_i\phi$ in advance so that no FFT is required in the the line search procedure itself. For example, to evaluate (2), we can simply compute

$$\rho(\tilde{\psi}) = \sum_{i=1}^{k} \left\| |z_i + \alpha t_i| - b_i \right\|^2,$$

where $z_i = FQ_i\psi$ and t_i have been computed already. Similarly, the direction derivative of ρ at $\psi + \alpha\phi$ can be obtained from

$$\sum_{i=1}^{k} \mathrm{Re}\left[t_i^*(z_i + \alpha t_i) - t_i^* \mathrm{Diag}\left(\frac{z_i + \alpha t_i}{|z_i + \alpha t_i|} \right) b_i \right].$$

Also, notice that no FFT is required in the curvature calculation (20) once t_i's are available.

4   Fixed-Point Iteration and Projection Algorithms

An alternative approach to finding a minimizer of (2) is to set its gradient to zero and seek ψ that satisfies the first order necessary condition of the minimization problem. If $\sum_{i=1}^{k} Q_i^* Q_i$ is nonsingular, by setting $\nabla\rho(\psi) = \frac{1}{2} \sum_{i=1}^{k} \left[Q_i^* Q_i \psi - Q_i^* F^* \mathrm{Diag}\left(\frac{z_i}{|z_i|} \right) b_i \right]$ to 0, we obtain

$$\psi = f(\psi) \tag{26}$$

where

$$f(\psi) = \left(\sum_{i=1}^{k} Q_i^* Q_i \right)^{-1} \left[\sum_{i=1}^{k} Q_i^* F^* \text{Diag}\left(\frac{z_i}{|z_i|} \right) b_i \right]. \tag{27}$$

Recall that $z_i \equiv FQ_i\psi$. Clearly, ψ is a fixed point of the function f.

A simple iterative technique one may use to find the solution to (27) is the fixed point iteration that has the form

$$\psi^{(\ell+1)} = f(\psi^{(\ell)}).$$

Replacing f with the right hand size of (27) yields

$$\psi^{(\ell+1)} = \left(\sum_{i=1}^{k} Q_i^* Q_i \right)^{-1} \left[\sum_{i=1}^{k} Q_i^* F^* \text{Diag}\left(\frac{z_i^{(\ell)}}{|z_i|^{(\ell)}} \right) b_i \right], \tag{28}$$

where $z_i^{(\ell)} \equiv FQ_i\psi^{(\ell)}$. This is the same sequence of iterates produced in what is known as the *error reduction* algorithm in standard phase retrieval literature [18]. This method is also known as the *alternate projection* algorithm for reasons to be discussed below.

It is easy to verify that the updating formula in (28) is identical to that produced by a preconditioned steepest descent algorithm in which the preconditioner K is chosen to be $K = \sum_{i=1}^{k} Q_i^* Q_i$, and a constant step length of 2 is taken at each iteration, i.e.,

$$\psi^{(\ell+1)} = \psi^{(\ell)} - 2\nabla\rho(\psi^{(\ell)}).$$

The sequence of iterates $\{\psi^{(\ell)}\}$ produced by (28) is guaranteed to converge to the fixed point of $f(\psi)$ from any starting point $\{\psi^{(0)}\}$, if the spectral radius (i.e., the largest eigenvalue) of the Jacobian of f (with respect to ψ) is strictly less than 1. Because the function f in (26) can be viewed as a function of ψ and $\bar{\psi}$, we should examine the Jacobian matrix of the system

$$\psi = \left(\sum_{i=1}^{k} Q_i^* Q_i \right)^{-1} \left[\sum_{i=1}^{k} Q_i^* F^* \text{Diag}\left(\frac{z_i}{|z_i|} \right) b_i \right], \tag{29}$$

$$\bar{\psi} = \left(\sum_{i=1}^{k} Q_i^T \bar{Q}_i \right)^{-1} \left[\sum_{i=1}^{k} Q_i^T F^T \text{Diag}\left(\frac{\bar{z}_i}{|z_i|} \right) b_i, \right] \tag{30}$$

where (30) is simply the conjugate of (29). It is not difficult to show that this Jacobian matrix has the form

$$J = \begin{pmatrix} K^{-1} & 0 \\ 0 & \bar{K}^{-1} \end{pmatrix} \begin{pmatrix} K - 2H_{\psi\psi}^{\rho} & -2H_{\psi\bar{\psi}}^{\rho} \\ -2H_{\bar{\psi}\psi}^{\rho} & \bar{K} - 2H_{\bar{\psi}\bar{\psi}}^{\rho} \end{pmatrix}, \tag{31}$$

where $H_{\psi\psi}^{\rho}$, $H_{\psi\bar{\psi}}^{\rho}$, $H_{\bar{\psi}\psi}^{\rho}$ and $H_{\bar{\psi}\bar{\psi}}^{\rho}$ are as defined in (15), (17), (18) and (16) respectively.

If (λ, ϕ) is an eigenpair of J, we can easily show that

$$2 \begin{pmatrix} H_{\psi\psi}^{\rho} & H_{\psi\bar{\psi}}^{\rho} \\ H_{\bar{\psi}\psi}^{\rho} & H_{\bar{\psi}\bar{\psi}}^{\rho} \end{pmatrix} \begin{pmatrix} \phi \\ \bar{\phi} \end{pmatrix} = (1-\lambda) \begin{pmatrix} K & 0 \\ 0 & \bar{K} \end{pmatrix} \begin{pmatrix} \phi \\ \bar{\phi} \end{pmatrix}.$$

If we again let $t_{ji} \equiv |t_{ji}|e^{i\mu_{ji}}$, $\zeta_{ji} \equiv |\zeta_{ji}|e^{i\theta_{ji}}$ and β_{ji} be the jth component of the vectors $t_i = FQ_i\phi$, $z_i = FQ_i\psi$ and b_i respectively, we can easily show that

$$\lambda = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n} \sin^2(\mu_{ji} - \theta_{ji})|t_{ji}|^2 \beta_{ji}/|\zeta_{ji}|}{\sum_{i=1}^{k} \sum_{j=1}^{n} |t_{ji}|^2}. \tag{32}$$

Clearly, when $\beta_{ji} \leq |\zeta_{ji}|$ for all $j = 1, 2, ..., m$ and $i = 1, 2, ...n$, $|\lambda| \leq 1$, and the fixed point iteration is guaranteed to converge to at least a local minimizer of ρ.

The fixed point of f may also be obtained by applying Newton's algorithm to seek the root of $r(\psi) = 0$, where $r(\psi) = \psi - f(\psi)$. The Newton's method produces a sequences of iterates $\{\psi^{(\ell)}\}$ that satisfy

$$\psi^{(\ell+1)} = \psi^{(\ell)} - J(\psi^{(\ell)})^{-1} r(\psi^{(\ell)}),$$

where the J matrix here is the Jacobian of r with respect to ψ. This approach is equivalent to applying Newton's algorithm (with appropriate line search and trust region strategies) to minimize $\rho(\psi)$.

Successive approximations to J can be constructed from $\psi^{(\ell)}$ and $r(\psi^{(\ell)})$ using Broyden's technique. This is similar to the Quasi-Newton algorithm discussed in the previous section. As a special case, replacing J with the crudest approximation, the identity matrix I, yields the standard error reduction algorithm.

If we multiply (29) from the left by Q_i for $i = 1, 2, ..., k$, and let $y^{(\ell)} = Q\psi^{(\ell)}$, where $Q = (Q_1^* \ Q_2^* \ ... \ Q_k^*)^*$, we obtain

$$y^{(\ell+1)} = P_Q P_F(y^{(\ell)}), \tag{33}$$

where $P_Q = Q(Q^*Q)^{-1}Q^*$, and

$$P_F(y) = \hat{F}^* \frac{y}{|y|} \cdot b,$$

where $\hat{F} = \text{Diag}\,(F, F, ..., F)$ and $b = (b_1^T \ b_2^T \ ... \ b_k^T)^T$.

Because a fixed point y of $P_Q P_F$ is in the range of Q, which is typically full rank when $mk > n$, we may recover the corresponding fixed point of f from y via the least squares solution $\psi^{(\ell)} = (Q^*Q)^{-1}Q^*y^{(\ell)}$.

This nonlinear map is the composition of a (linear) orthogonal projector P_Q and a (non-linear) Fourier magnitude projector P_F. A fixed point iteration based on (33) is also called *alternating projection* (AP) algorithm in the phase retrieval literature because the approximation to the solution of (33) is obtained by applying P_Q and P_F in an alternating fashion.

It is easy to verify that P_F is indeed a projection operator in the sense that

$$\|P_F(y) - y\| \leq \|w - y\| \quad \text{for all} \ \ w \in \{w|w = P_F(w)\}. \tag{34}$$

This property of P_F, together with the fact that P_Q is an orthogonal projection operator, i.e. $\|P_Q y - y\| \leq \|w - y\|$ for all $w \in \text{Range}(Q)$, allows us to show that the residual error $\|P_Q P_F(y^{(\ell)}) - y^{(\ell)}\|$ decreases monotonically in the AP algorithm. The proof of this observation was shown by Fienup in [15], which we summarize below.

Let $y^{(\ell)}$ be the vector produced in the ℓ-th AP iterate. Clearly, $y^{(\ell)} \in \text{Range}(Q)$. Because P_Q is an orthogonal projector, we have

$$\|P_Q P_F(y^{(\ell)}) - P_F(y^{(\ell)})\| \leq \|P_Q P_F(y^{(\ell)}) - y^{(\ell)}\| = \|y^{(\ell+1)} - y^{(\ell)}\|. \tag{35}$$

Because $P_F(y^{(\ell)}) \in \{w|w = P_F(w)\}$, it follows from (34) that

$$\|P_F(y^{(\ell+1)}) - y^{(\ell+1)}\| = \|P_F(P_Q P_F(y^{(\ell)})) - P_Q P_F(y^{(\ell)})\| \leq \|P_Q P_F(y^{(\ell)}) - P_F(y^{(\ell)})\|. \tag{36}$$

Consequently, we can deduce from (35) and (36) that

$$\|P_F(y^{(\ell+1)}) - y^{(\ell+1)}\| \leq \|y^{(\ell+1)} - y^{(\ell)}\|.$$

Finally, it follows from the following inequality

$$\|P_Q(P_F(y^{(\ell+1)}) - y^{(\ell+1)})\| \le \|P_F(y^{(\ell+1)}) - y^{(\ell+1)}\|,$$

and the fact that $y^{(\ell+1)} \in \mathrm{Range}(Q)$ that

$$\|y^{(\ell+2)} - y^{(\ell+1)}\| \le \|y^{(\ell+1)} - y^{(\ell)}\|. \tag{37}$$

The equality in (37) holds only when $P_F(y^{(\ell)}) = y^{(\ell)}$, i.e., when convergence is reached.

The inequality (37) shows that the AP algorithm converges to a stationary point. However, the convergence can be extremely slow because

$$\|z^{(\ell+1)}\| = \|Fy^{(\ell+1)}\| = \|y^{(\ell+1)}\| = \|P_Q P_F(y^{(\ell)})\| \le \|P_F(y^{(\ell)})\| = \|b\|,$$

and many of the terms β_{ji}/ζ_{ji}, $i = 1, 2, ..., k$ and $j = 1, 2, ..., m$, in (32) may be great than 1. Only when $y^{(\ell)}$ is very close to the fixed point of $P_Q P_F$, the spectral radius of the Jacobian of (27) may become much smaller than 1 in (32) due to the reduction effect of the $\sin^2(\mu_{ji} - \theta_{ji})$ terms.

The simple alternating projection algorithm has been extended to the hybrid input-output (HIO) algorithm [15], the relaxed averaged alternating reflection (RAAR) algorithm [16], and many other variants [17, 18] in the phase retrieval literature. Just to give a few examples, in the HIO and RAAR algorithms, the approximation to the solutions of (30) and (33) are updated by

$$
\begin{aligned}
y^{(\ell+1)} &= [P_Q P_F + (I - P_Q)(I - \beta P_F)]\, y^{(\ell)}, \text{ HIO,} \\
y^{(\ell+1)} &= [2\beta P_Q P_F + (1 - 2\beta)P_F + \beta(P_Q - I)]\, y^{(\ell)}, \text{ RAAR.} \\
\psi^{(\ell+1)} &= (Q^*Q)^{-1}Q^*y^{(\ell)},
\end{aligned}
$$

where β is a constant often chosen to be between 0 and 1.

Although these algorithms tend to accelerate the convergence of $y^{(\ell)}$, their convergence behavior is less predictable and not well understood.

5 Wigner Deconvolution

Long before iterative methods were applied to solve the ptychography problem, Rodenburg and his colleagues suggested that the problem can be solved via what they called Wigner deconvolution [9].

To explain the basic idea behind Wigner deconvolution, we need to state a continuum version of the ptychography problem. If the set of translation vectors $\{\mathbf{x}\}$ forms a continuum in 2D, then it can be shown [19] that the Fourier transform of $y_{\mathbf{x}}^2 \equiv |\mathcal{F}\{a(\mathbf{r})\hat{\psi}(\mathbf{r} + \mathbf{x})\}|^2$ with respect to $\mathbf{x}$, which we denote by $\mathcal{F}_{\mathbf{x}}\{y_{\mathbf{x}}^2\}$, can be written as the convolution of two functions with respect to $\mathbf{r}'$, i.e.,

$$\mathcal{F}_{\mathbf{x}}\{y_{\mathbf{x}}^2(\mathbf{r})\} = [A(\mathbf{r}')\bar{A}(\mathbf{r}' + \mathbf{x}')] \star_{\mathbf{r}'} [\Psi(\mathbf{r}')\bar{\Psi}(\mathbf{r}' - \mathbf{x}')], \tag{38}$$

where $A(\mathbf{r}') = \mathcal{F}\{a(\mathbf{r})\}$, $\Psi(\mathbf{r}') = \mathcal{F}\{\psi(\mathbf{r})\}$, $\bar{A}$ denotes the conjugate of A, $\bar{\Psi}$ denotes the conjugate of Ψ, and $\star_{\mathbf{r}'}$ denotes a convolution operation with respect to $\mathbf{r}'$. Note that $\mathcal{F}_{\mathbf{x}}\{y_{\mathbf{x}}^2(\mathbf{r})\}$ is a function of $\mathbf{x}'$. The Fourier transform of $A(\mathbf{r}')\bar{A}(\mathbf{r}'+\mathbf{x}')$ or $\Psi(\mathbf{r}')\bar{\Psi}(\mathbf{r}'-\mathbf{x}')$ is called a *Wigner distribution* in [9].

The Fourier transforms used in the definition of $A(\mathbf{r}')$ and $\Psi(\mathbf{r}')$ can be replaced by discrete Fourier transforms (DFT) if both $a(\mathbf{r})$ and $\psi(\mathbf{r})$ are band-limited and they are sampled at or beyond the Nyquist frequency. The Fourier transform of $y_{\mathbf{x}}^2$ with respect to $\mathbf{x}$ can be replaced by a DFT only if the translation vector $\mathbf{x}$ is sampled at or beyond the Nyquist frequency of $\psi(\mathbf{r})$.

We will define a fully sampled $\Psi(\mathbf{r}')$ by a column vector

$$f = (f_1 \; f_2 \; \cdots \; f_n)^T,$$

where $f_i = \Psi(\mathbf{r}_i')$. Note that, when appeared by itself in Ψ, the variable $\mathbf{x}'$ and $\mathbf{r}'$ can be used interchangeably, i.e., $f_i = \Psi(\mathbf{x}_i')$ holds also.

There are at least two ways to represent $\Psi(\mathbf{r}')\bar{\Psi}(\mathbf{r}' - \mathbf{x}')$ systematically in a vector form. We choose to write it as

$$u(f) = \begin{pmatrix} \mathrm{Diag}\,(f)\,P_1^T\bar{f} \\ \mathrm{Diag}\,(f)\,P_2^T\bar{f} \\ \vdots \\ \mathrm{Diag}\,(f)\,P_n^T\bar{f} \end{pmatrix},$$

where P_i is a permutation matrix that shifts $\bar{f}$ cyclically by $i - 1$ pixels, and $\bar{f}$ denotes the conjugate of f. This representation corresponds to writing down $\Psi(\mathbf{r}')\bar{\Psi}(\mathbf{r}' - \mathbf{x}')$ by having $\mathbf{r}'$ as the fastest changing index. By enumerating $\mathbf{x}'$ first, we can represent $\Psi(\mathbf{r}')\bar{\Psi}(\mathbf{r}' - \mathbf{x}')$ in an alternative form

$$\Pi u(f) = \begin{pmatrix} f_1 P_1^T\bar{f} \\ f_2 P_2^T\bar{f} \\ \vdots \\ f_n P_n^T\bar{f} \end{pmatrix}, \tag{39}$$

where Π is an $n^2 \times n^2$ permutation matrix that reorders $\mathbf{x}'$ and $\mathbf{r}'$.

Employing the same ordering we use to represent the fully sampled $\Psi(\mathbf{r}')\bar{\Psi}(\mathbf{r}' - \mathbf{x}')$, we can express the convolution kernel $A(\mathbf{r}')A(\mathbf{r}' + \mathbf{x}')$ by a matrix W. This matrix has a block diagonal form, i.e.,

$$W = \begin{pmatrix} W_1 & & & \\ & W_2 & & \\ & & \ddots & \\ & & & W_n \end{pmatrix},$$

where W_i is a block cyclic matrix with cyclic blocks (BCCB). This type of BCCB structure allow the convolution $W_i \mathrm{Diag}\,(f)\,P_i\bar{f}$ to be carried out efficiently by using FFTs.

Using the notation established above, we can now express the sampled version of (38) as

$$\Pi W u(f) = \bar{b}^2,$$

where

$$\bar{b}^2 = \hat{F}\Pi \begin{pmatrix} b_1^2 \\ b_2^2 \\ \vdots \\ b_m^2 \end{pmatrix}, \quad \hat{F} = \begin{pmatrix} F & & & \\ & F & & \\ & & \ddots & \\ & & & F \end{pmatrix},$$

and F is the matrix representation of a 2D discrete Fourier transform of an image with n pixels.

If W is nonsingular, i.e., W_i is nonsingular for all $i = 1, 2, ..., n$, we can recover $u(f))$ by simply inverting W, i.e.

$$u(f) = W^{-1}\Pi^T \bar{b}^2, \tag{40}$$

Equation (40) represents a deconvolution process, and is known as *Wigner deconvolution* [9]. The application of W^{-1} to the vector $\Pi^T \bar{b}^2$ can be achieved through an FFT based fast deconvolution or an iterative solver such as the conjugate gradient algorithm. We do not need to explicitly invert the matrix W. If W is singular or ill-conditioned, we may add a small constant to the diagonal of W to regularize the deconvolution.

Applying the permutation Π to $u(f)$ allows us to rewrite the solution of the deconvolution problem in the form of (39). If $f_i \neq 0$ for $i = 1, 2, ..., n$, we define $c_i = 1/f_i$. Furthermore, let us define $\hat{g}^2 = \Pi W^{-1}\Pi^T \bar{f}^2$, which can be partitioned as

$$\hat{g}^2 = \begin{pmatrix} \hat{g}_1^2 \\ \hat{g}_2^2 \\ \vdots \\ \hat{g}_n^2 \end{pmatrix}.$$

where $\hat{g}_i^2 \in \mathbb{C}^{n \times 1}$.

By treating c_i as a separate set of unknowns, with the exception of of c_1, which we will set to an arbitrary constant, e.g., 1, we can turn (40) into a linear least squares problem by minimizing the norm of

$$r = \begin{pmatrix} P_1^T & & & \\ P_2^T & -\text{Diag}\left(\hat{g}_2^2\right) & & \\ \vdots & & \ddots & \\ P_n^T & & & -\text{Diag}\left(\hat{g}_n^2\right) \end{pmatrix} \begin{pmatrix} \hat{f} \\ c_2 \\ \vdots \\ c_n \end{pmatrix} - \begin{pmatrix} c_1 \hat{g}_1^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{41}$$

The minimization of $\|r\|$ can be easily solved by back substitution. This is essentially the "stepping out" procedure described in [9]. The reason that we can set c_1 to an arbitrary constant is that we are often interested in the relative amplitudes and phases of $\hat{\psi}(\mathbf{r})$, multiplying the entire image $\psi(\mathbf{r})$ or $\Psi(\mathbf{r'})$ by a constant does not change the quality of the image.

It may seem that the use of iterative method is not necessary if we can solve the ptychography problem by Wigner deconvolution, which can be viewed as a linear inversion scheme. However, as we will show below, the Wigner deconvolution problem cannot be solved directly (using an FFT based deconvolution scheme) if $\mathbf{x}$ is sampled below the Nyquist frequency, i.e. when the amount of probe translation is larger than the resolution of the image to be reconstructed.

When $\mathbf{x}$ is sampled below the Nyquist frequency, which can occur in an experiment, we must modify (38) by introducing an *aliasing operator* $S_{\mathbf{x'}}$. Because $a(\mathbf{r})$ is a localized window in practice, $A(\mathbf{r'})$ is subsampled in the reciprocal space. Therefore a subsampling operator $S_{\mathbf{r'}}$ must be included in a finite-dimensional analog of (38) to account for this effect.

With these additional operators, the sampled version of equation (38) can be expressed as

$$S_{\mathbf{x'}}\Pi S_{\mathbf{r'}} W u(f) = \bar{b}^2, \tag{42}$$

where the dimensions of Π, W, $u(f)$ and $\bar{b}^2$ need to be adjusted to reflect fewer pixel samples per diffraction frame and fewer frames resulting from increased distance $\mathbf{x}$ between two adjacent frames. For simplicity, let us assume that f and each frame b_i^2 are square images with n and m

pixels respectively, and the distance between two adjacent frames is d_x (in either the horizontal or the vertical direction). Then, the aliasing operator $S_{\mathbf{x}'}$ in (42) is a block diagonal matrix consisting of n_f diagonal blocks of dimension $m \times n$, where $n_f = \lfloor \sqrt{n}/\sqrt{m} \rfloor$. The subsampling operator $S_{\mathbf{r}'}$ is a block diagonal matrix consisting of m diagonal blocks of dimension $n_f \times n$, and Π is an $n_f m \times n_f m$ row permutation matrix that reshuffles the rows of $S_{\mathbf{r}'} W u(f)$ so that $\mathbf{x}'$ is the fastest changing index. For 1D signals, a diagonal block of $S_{\mathbf{x}'}$ can be represented by

$$(I_m \quad I_m \quad \cdots \quad I_m),$$

where I_m is an $m \times m$ identity matrix. Similarly, a typical diagonal block of $S_{\mathbf{r}'}$ has the form

$$(\cdots \quad 0 \quad I_{n_f} \quad 0 \quad \cdots),$$

where I_{n_f} is an $n_f \times n_f$ identity matrix.

Because $S_{\mathbf{x}'}$, Π and $S_{\mathbf{r}'}$ are not square matrices, we cannot obtain $u(f)$ by simply applying the inverse of these matrices and W^{-1} to $\bar{b}^2$.

Instead, we must recover f, hence the fully sampled $\hat{\psi}(r)$, by solving the following nonlinear least squares problem

$$\min_f \|S_{\mathbf{x}'}\Pi S_{\mathbf{r}'} W u(f) - \bar{b}^2\|^2. \tag{43}$$

It is not difficult to see that the objective function in the nonlinear least squares problem (43) is equivalent to (3). Therefore, iterative optimization techniques applied to minimize (3) can be used to solve (43) also. However, the evaluation of the objective function in (43) and its derivatives, which we will not show here, are more costly because evaluating $u(f)$ requires at least $\mathcal{O}(n^2)$ operations, and multiplying W with $u(f)$ requires an additional $\mathcal{O}(n^2 \log(n))$ operations. This operation count is much higher than that associated with evaluating (3), which is $\mathcal{O}(mn_f \log(n_f) + mn_f)$.

We should mention that, if one is interested a reconstruction of limited resolution, $\tilde{f}$, which is a cropped version of $\Psi(-\mathbf{x}')$, the objective function in (43) can be modified to become

$$\|S_{\mathbf{x}'}\Pi(S_{\mathbf{r}'} W S_{\mathbf{r}'}^T)\tilde{u}(\tilde{f}) - \bar{b}^2\|^2,$$

where $\tilde{u}(\tilde{f}) \in \mathbb{C}^{m \times 1}$. Furthermore, if the translation of the frame $\mathbf{x}'$ is chosen to be commensurate with the size of each frame, e.g., $x' = \sqrt{n/m}$, then $S_{\mathbf{x}'}$ becomes an identity matrix. Consequently, one may obtain $\tilde{u}(\tilde{f})$ (and subsequently $\tilde{f}$) by performing a Wigner deconvolution.

6   Numerical Examples

In this section, we demonstrate and compare the convergence of iterative algorithms for ptychographic reconstruction using two test images. The first test image is a 256×256 real-valued cameraman image shown in Figure 2. The image is often used in the image processing community to test image reconstruction and restoration algorithms. The second test image is a complex valued image. It also contains 256×256 pixels that correspond to the complex transmission coefficients of a collection of gold balls embedded in some medium. The amplitude and phase angles of these pixels are shown in Figure 3.

All numerical examples presented in this paper are performed in MATLAB.

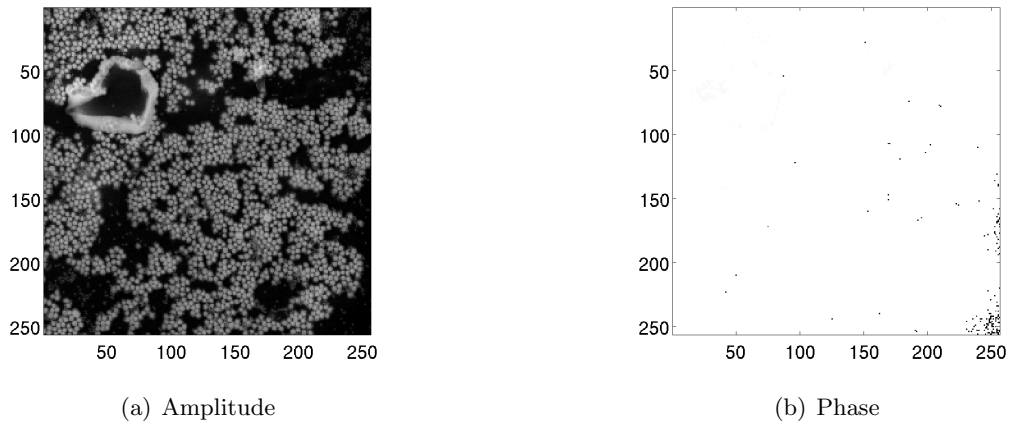Figure 2: The cameraman test image.

(a) Amplitude

(b) Phase

Figure 3: The amplitude and phase of the transmission coefficient of a collection of gold balls.

6.1 Comparison of Convergence Rate

In this section, we show the convergence behavior of different iterative algorithms we discussed in section 3 by numerical experiments. In the cameraman image reconstruction experiment, we choose the illuminating probe $a(\mathbf{r})$ to be a 64×64 binary probe shown in Figure 4(a). The pixels within the 32×32 square at the center of the probe assume the value of 1. All other pixels take the value of 0. The zero padding of the inner 32×32 square ensures that the diffraction pattern of a 64×64 frame associated with this probe is oversampled in the reciprocal space. In the gold ball image reconstruction experiment, the illuminating probe is chosen to be the amplitude of the Fourier transform of an annular ring with inner radius of $r_1 \approx 5.4$ and outer radius of $r_2 \approx 19.4$. This probe mimics the true illumination used in a physical experiment.
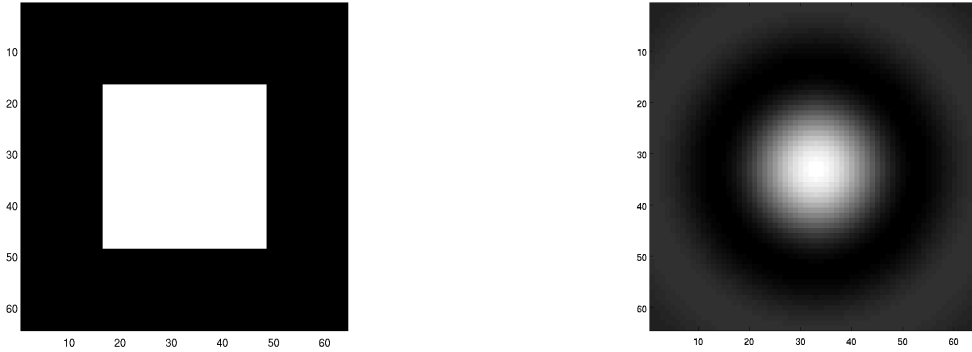
(a) The binary probe used in the reconstruction of the cameraman image.

(b) The probe used in the reconstruction of the gold ball image.

Figure 4: The illuminating probes $a(\mathbf{r})$ used in ptychographic reconstructions of the cameraman and gold ball images.

In the cameraman experiment, the probe is translated by 8 pixels at a time in either horizontal or vertical direction. To prepare a stack of k diffraction images b_i, $i = 1, 2, ..., k$, we start from the upper left corner of the true image, extract a 64×64 frame, and multiply it with the probe, and then apply a 2D FFT to the product. The magnitude of transform is recorded and saved before we move either horizontally or vertically to obtain the next frame. If the lower right corner of the frame goes outside of the image (which does not happen in this particular case), we simply "wrap the probe around" the image as if the image is periodically extended. As a result, the total number of diffraction frames we use for each reconstruction is

$$k = \frac{256}{8} \cdot \frac{256}{8} = 1024.$$

As we will show in section 6.4, the size of translation, which determines the amount of overlap between adjacent frames, has a noticeable effect on the convergence of the iterative reconstruction algorithms.

Figure 5 shows the convergence history of several iterative algorithms discussed in section 3 when they are applied to the diffraction frames extracted from the cameraman image. We plot
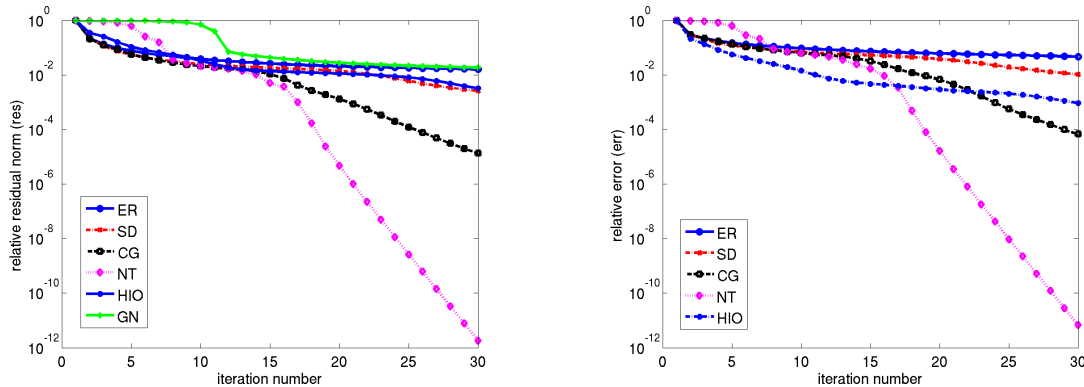
both the relative residual norm defined by

$$res = \frac{\sqrt{\sum_{i=1}^{k} \||z_i|^{(\ell)} - b_i\|^2}}{\sqrt{\sum_{i=1}^{k} \|b_i\|^2}}, \tag{44}$$

where $|z_i|^{(\ell)} = |FQ_i\psi^{(\ell)}|$ and ℓ is the iteration number, and the relative error of the reconstructed image defined by

$$err = \frac{\|\psi^{(\ell)} - \hat{\psi}\|}{\|\hat{\psi}\|}.$$

In these runs, an exact line search is used in the steepest descent (SD), nonlinear conju-

(a) Change of the relative residual norm (res) for the reconstruction of the cameraman image.

(b) Change of the relative error (err) for the reconstruction of the cameraman image.

Figure 5: A comparison of the convergence behavior of different iterative ptychographic reconstruction algorithms for the cameraman image.

gate gradient (CG). The Steihaug's trust region technique implemented in [20] is used in the Newton's method (NT). We set the starting guess of the solution $\hat{\psi}$ to

$$\psi^{(0)} = \left(\sum_{i=1}^{k} Q_i^* Q_i \right)^{-1} \sum_{i=1}^{k} Q_i^* b_i.$$

It is clear from Figure 5 that NT converges much faster than the other algorithms. Its performance is followed by the CG algorithm which is much faster than the error reduction (ER), SD, Gauss-Newton (GN) and the hybrid input-output (HIO) algorithms. Similar convergence behavior is observed when other random starting guesses are used, although occasionally, a random starting guess can lead to stagnation or convergence to a local minimizer. We will discuss this issue in section 6.3. We set the maximum number of iterations allowed in all runs to 30. This is somewhat excessive for both NT and CG algorithms. Typically, when the relative error of the reconstructed image falls below 10^{-3}, it is nearly impossible to visually distinguish the reconstruction from the true image. When the relative error is larger, the reconstructed cameraman images may contain visible artifacts such as those shown in Figures 6(a) and 6(b) which are produced at the end of the 30th ER and SD iterations respectively.

It is somewhat surprising that GN performs poorly on this problem. We believe the problem is that we used the MATLAB implementation of the large-scale Gauss-Newton algorithm,

20

(a) ER reconstruction   (b) SD reconstruction

Figure 6: The reconstructed cameraman images by ER and SD algorithms contain visible ringing artifacts.

i.e., the function `lsqnonlin` in the MATLAB's Optimization Toolbox, which does not handle functions of complex variable very well. Moreover, it is not easy to obtain the relative error associated with the approximate reconstruction produced at each iteration from this function.

For the reconstruction of the gold ball image, we choose the starting guess to be

$$\psi^{(0)} = \left(\sum_{i=1}^{k} Q_i^* Q_i \right)^{-1} \sum_{i=1}^{k} Q_i^* \mathrm{Diag}\,(b_i)\,\mathrm{Diag}\,(|u_i|)^{-1}\,u_i,$$

where u_i is a complex random vector, and the real and imaginary part of each component has a uniform distribution within $[-1, 1]$.

In this experiment, the probe is translated by a larger amount (16 pixels) in either horizontal or vertical direction. Figure 7 shows the convergence history of ER, SD, CG, HIO, and NT. From Figure 7(a), it appears that CG is the best among all the methods we tried. The HIO algorithm performs well in the first 60 iterations, but then stagnates. As we can see from Figure 7 that the neither the residual norm nor the relative error associated with HIO changes monotonically. This is not completely surprising because HIO does not try to minimize either objective functions. For this example, the performance of NT lags behind CG by a large margin although both algorithms exhibit monotonic convergence with a more predictable error reduction. We should mention that to measure the relative error associated with a reconstructed gold ball image $\psi^{(\ell)}$, we need to multiply it by a constant phase factor γ first, i.e., the relative error is defined as

$$err = \frac{\|\gamma \psi^{(\ell)} - \hat{\psi}\|}{\|\hat{\psi}\|}.$$

In Figure 8, we can clearly see that the magnitude of the reconstructed images produced by CG (Figure 8(a)) and HIO (Figure 8(c)) are nearly indistinguishable from the magnitude of the true image. However, the phase angles of the reconstructed image produced by CG (Figure 8(d)) appear to be better than those produced by HIO, which is indicated by the magnitude of the absolute errors $|\gamma \psi^{(\ell)} - \hat{\psi}|$ shown in Figures 8(b) and 8(d).
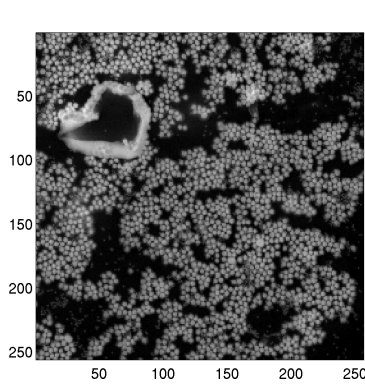
21
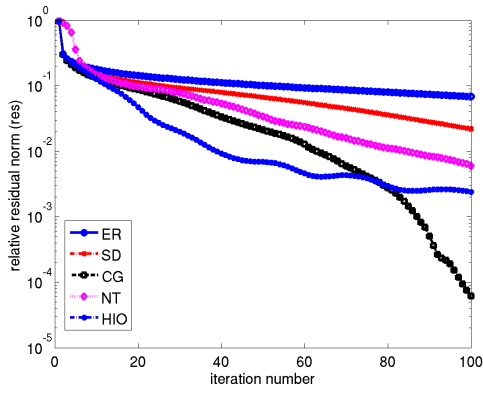
(a) Change of the relative residual norm (res) for the reconstruction of the gold ball image.
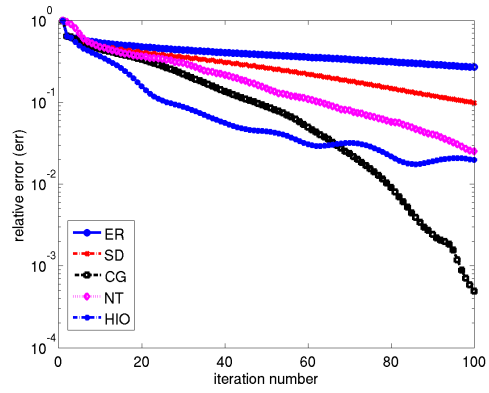
(b) Change of the relative error (err) for the reconstruction of the gold ball image.

Figure 7: A comparison of the convergence behavior of different iterative ptychographic reconstruction algorithms for the gold ball image.
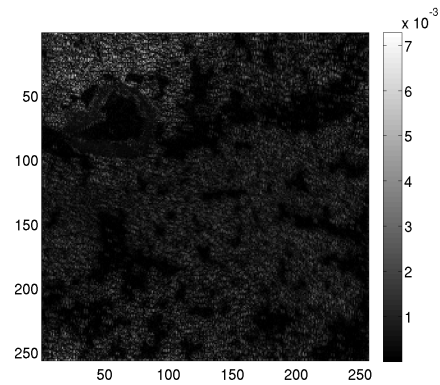

(a) The magnitude of the reconstructed gold ball image produced by the CG algorithm.
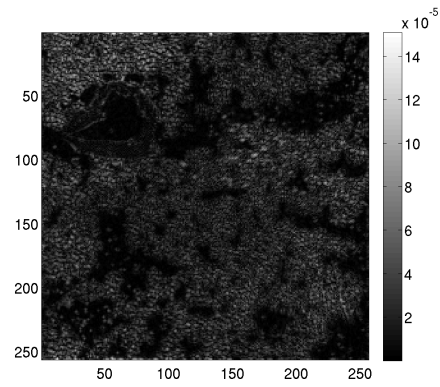
(b) The magnitude of the error associated with the reconstructed gold ball image produced by the CG algorithm.
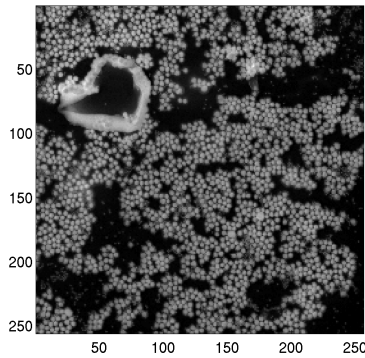
(c) The magnitude of the reconstructed gold ball image produced by the HIO algorithm.

(d) The magnitude of the error associated with the reconstructed gold ball image produced by the HIO algorithm.

Figure 8: The reconstructed cameraman images produced by CG and HIO.

6.2 The Effect of Preconditioning

As we indicated in Section 3.2, the use of a preconditioner can enhance the convergence of SD and CG. A natural preconditioner that is easy to construct is (24). However, this preconditioner is only effective, when the condition number of K is relatively large. For the binary probe used in the reconstruction of the cameraman image, $K = 4I$. The condition number of this matrix is 1. Hence, using this preconditioner has no effect on the convergence of the CG iteration, as we can clearly see in Figure 9(a). The condition number associated with the probe used in the gold ball image reconstruction is around 4.5. Hence the effect of the preconditioner is negligible as we can see from Figure 9(b).
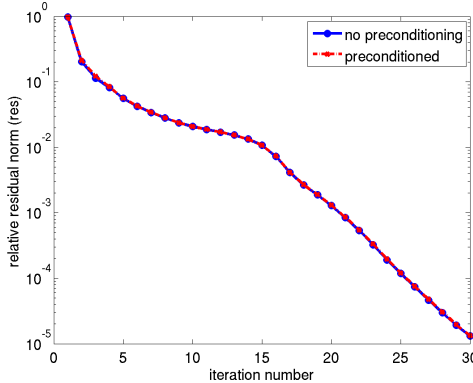
(a) The effect of the preconditioner on the convergence of SD.
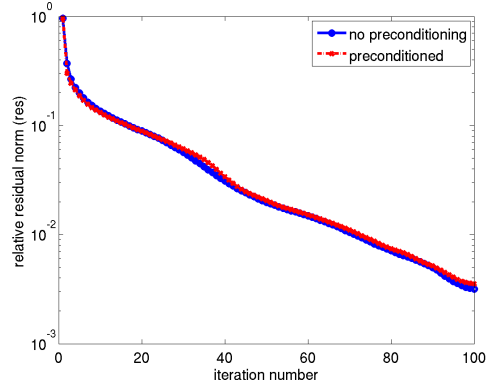
(b) The effect of the preconditioner on the convergence of CG.

Figure 9: The effect of a preconditioner on the convergence of the CG algorithms applied to cameraman and gold ball image reconstruction.

6.3 Local Minimizer and the Choice of the Objective Function

As we indicated in section 2.2, based on the analytic Hessian and curvature expression, that neither $\epsilon(\psi)$ nor $\rho(\psi)$ is globally convex. This observation suggests that all iterative optimization algorithm discussed above may converge to a local minimizer. Although we found that in practice, local minimizers are not easy to find, they do exist as the following example show.

In order to find a local minimizer, we construct many random starting guesses using the MATLAB `rand` function. To save time, we chose to reconstruct a 64×64 subimage of the cameraman image shown in Figure 2. This subimage is shown in Figure 12(a). A 16×16 binary probe that has a value 1 in the 8×8 center of the probe and 0 elsewhere is used. The diffraction stack consisting of 64 diffraction images is obtained by translating the probe 4 pixels a time in either the horizontal and vertical direction.

Figure 10 shows that one of the random starting guesses lead to the convergence of the CG algorithm to a local minimizer. In particular, the relative residual (44) which is proportional to the objective function ρ stagnates around 0.9 after the first 15 iterations (Figure 10(a)), whereas the relative gradient $\|\nabla\rho(\psi^{(\ell)})\|/\|\hat{\psi}\|$ decreases to 10^{-8} after 40 iterations.

Figure 12(b) shows how the reconstructed image compares with the true image for this particular starting guess used. In this case, the local minimizer appears to contain visible artifacts in a small region near top of the tripod. The amplitude of this localized error is also

revealed in the relative error plot shown in Figure 11(a). The phase error associated with a particular frame of the reconstruction obtained from

$$\frac{\overline{Q_i\psi}}{|Q_i\psi|} \cdot \frac{Q_i\hat{\psi}}{|Q_i\hat{\psi}|},$$
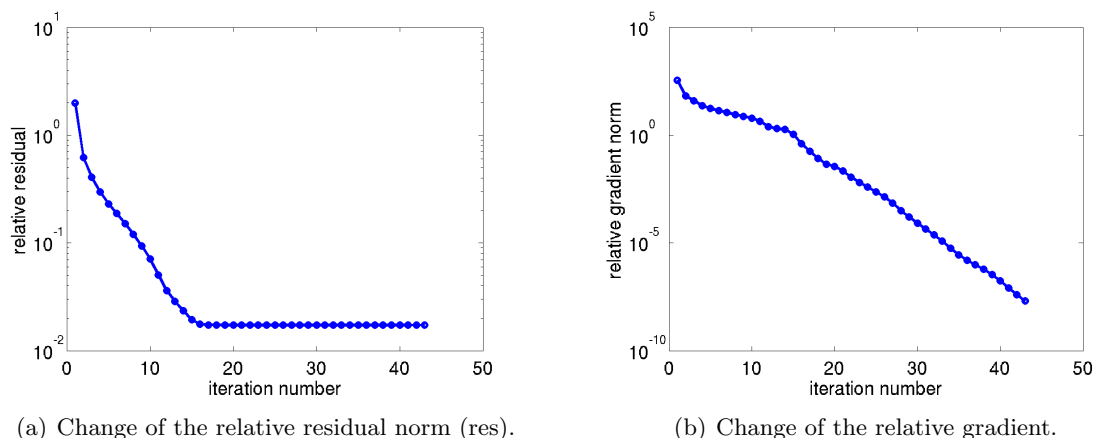
for some particular Q_i is shown in Figure 11(b).

(a) Change of the relative residual norm (res).

(b) Change of the relative gradient.

Figure 10: The convergence of CG to a local minimizer.

(a) Amplitude error in the reconstruct image

(b) Phase error in degrees associated with a particular frame
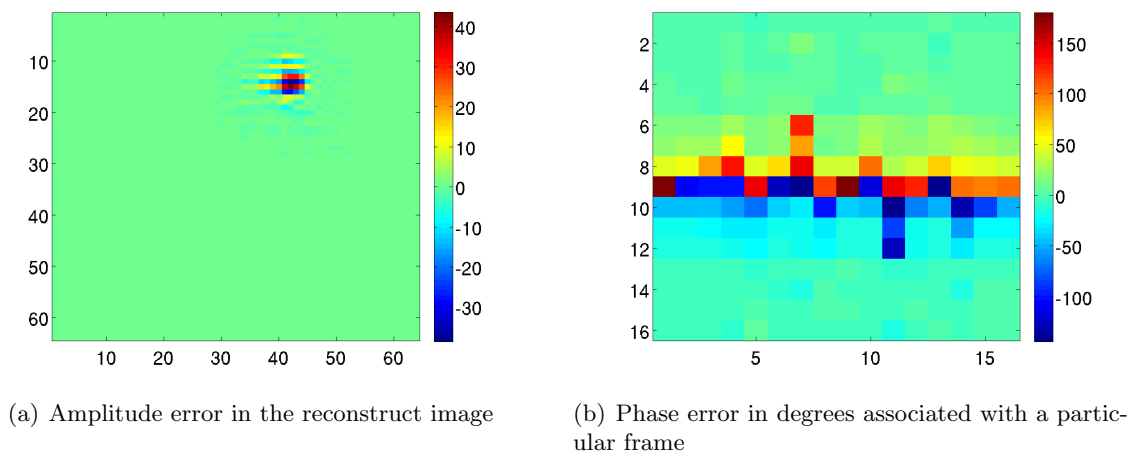
Figure 11: The error associated with a local minimizer.

We should also note that for this particular starting guess, all methods we tried converged to the same local minmizer. This is not all that surprising. It simply shows (empirically) that local minimizers of (2) exists, and our starting guess is sufficiently close to it.

However, what is interesting is that if we choose to minimize (3) by using any one of the iterative methods discussed above from the same starting guess, we are able to obtain the correct solution. For examples, Figure 13(a) shows that when the NT applied to the weighted
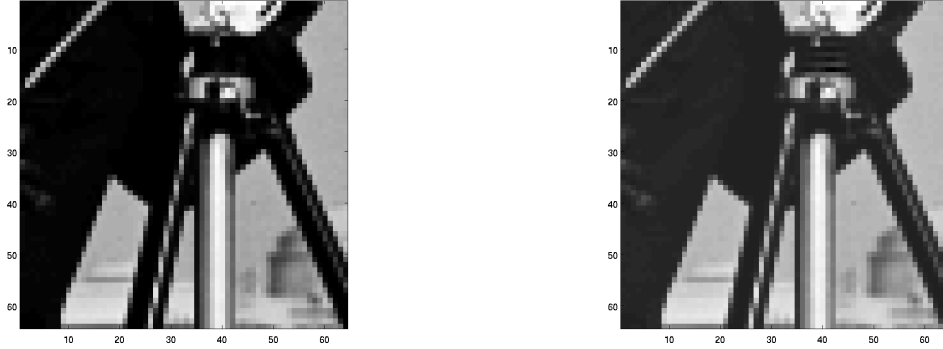
(a) True image.

(b) The reconstructed image (a local minimizer).

Figure 12: The artifacts produced by a local minimizer of ρ.

(scaled) objective function

$$\tilde{\epsilon}_{(}\psi) = \frac{1}{2}\sum_{i=1}^{k}(|z_i|^2 - b_i^2)^T \text{Diag}\,(b_i)^{-1}\,(|z_i|^2 - b_i^2), \tag{45}$$

where $|z_i| = |FQ_i\psi|$ and $b_i = |FQ_i\hat{\psi}|$, an accurate reconstruction can be obtained in roughly 350 iterations. Admittedly, the convergence rate is much slower in this case when compared to the convergence of NT when it's applied to (2) from a different starting point. The convergence is even slower if no weighting (or scaling) is used, i.e. when (3) is used as the objective function. However, the fact that convergence can be reached for (45) but not (2) from the same starting point is quite interesting. Furthermore, Figure 13(b) shows that if we take the local minimizer returned from an iterative minimization of (2) as the starting guess for minimizing (45), convergence can be reached in 12 iterations. This experiment suggests that it may be useful to have a hybrid optimization scheme in which (2) is minimized first. If a local minimizer of (2) is identified, one can then try to minimize (45) starting from the local minimizer of (2).

6.4 The Effect of Overlapping on the Convergence of Iterative Algorithm

As we alluded to earlier, the amount of overlap between two adjacent diffraction frames has a noticeable effect on the convergence of optimization based iteration algorithms (e.g., CG, NT, SD etc.) used to reconstruct the true image. Although we currently do not have a clear way to quantify such an effect mathematically, the following examples demonstrate this effect.

In the first example, we try to reconstruct the gold ball image from four different diffraction stacks. Each stack contains a set of 64×64 diffraction frames. These frames are generated by translating the probe shown in Figure 4(b) by different amount in horizontal and vertical directions. The larger the translation, the smaller the overlap is between two adjacent images. Figure 15(a) shows that CG converges very slowly when the diffraction stack contains diffraction frames obtained by translating the probe 20 pixels at a time (the black curve). Faster convergence is observed when the amount of translation is decreased to $\Delta x = 16, 12, 8$. It is interesting to see from Figure 15(b) that the amount of overlap does not affect the convergence of the HIO algorithm.
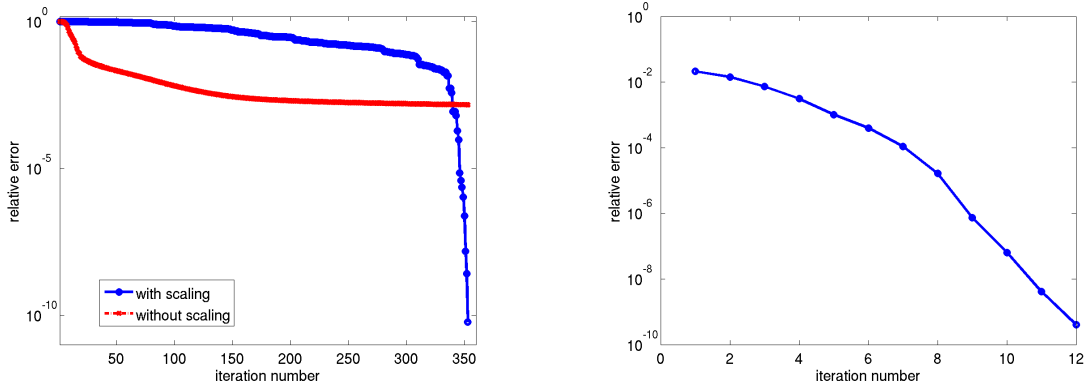
(a) The convergence of the NT algorithm when it is applied to (3) (red) and (45). The starting guess chosen in these runs is the same one used in the minimization of (2).

(b) The convergence of the NT algorithm when the starting guess is chosen to be the local minimizer shown in Figure 12(b)

Figure 13: The convergence of the NT algorithm when applied to (3) (red) and (45) (blue).

In the second example, we try to reconstruct the gold ball image from 1024 diffraction frames of 128×128 pixels. The illumination function is similar to that used in Figure 4. It is scaled by a factor of 2 to 128×128 pixels. The probe FWHM (full width at half maximum) is 30 pixels. We choose to fix the number of frames. So the reconstructed area increases with step size. When probe is near the edge of the image, we "wrap it around the edge" as if the image itself is periodically extended. The overlap is varied by changing the step size Δx. The larger the Δx, the smaller the amount of overlap.

The starting point is produced from a random number generator for each test. A range of step sizes between 6 and 30 pixels have been tried. For a fixed step size, the test is repeated 100 times. We observe that the step size Δx does not influence the convergence rate up to $\Delta x \simeq 20$. Figures 14(a) and 14(c) show that the conjugate gradient method converges in less than 400 iterations, while the RAAR algorithm requires almost 1500 iterations. Figures 14(b) and 14(d) illustrate the percentage of successful runs started from a random guess for each of the step sizes $0 \leq \Delta x \leq 30$. The percentage of successful runs (shown in color) is plotted against the maximum number of allowed iterations. When $\Delta x \leq 20$, both CG and RAAR converge nearly 100% of the time when a relatively small number of iterations are used in these methods. However, when $20 \leq \Delta x \leq 25$, more iterations are required to ensure the convergence of CG and RAAR. When $25 \leq \Delta x \leq 30$, CG appears to stagnate for all random starting guesses we tried, whereas RAAR can still converge when a very large number of iterations are taken.

To explain the effect of overlapping on the convergence of optimization based iterative algorithms such as the nonlinear CG, we examine the structure of the Hessian of the objective function ρ in (2). It follows from (15)-(16) that the H^ρ can be written as

$$H^\rho = \left((\hat{F}Q)^* \ (\hat{F}Q)^T \right) \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \begin{pmatrix} \hat{F}Q \\ \hat{\bar{F}}\bar{Q} \end{pmatrix}, \tag{46}$$

where $B_{11} = \overline{B_{22}}$ and $B_{12} = B_{21}^*$ are all diagonal, $\hat{F}$ is a block diagonal matrix of discrete Fourier transforms, i.e. $\hat{F} = \text{Diag}(F, F, ..., F)$, and $Q = (Q_1^* Q_2^* ... Q_k^*)^*$. The diagonal elements of B_{11} and B_{12} are simply $1 - \beta_{ji}/(2\zeta_{ji})$ and $\beta_{ji}\zeta_{ji}^2/(2|\zeta_{ji}|^3)$ respectively for $i = 1, 2, ..., k$ and $j = 1, 2, ..., m$.

26
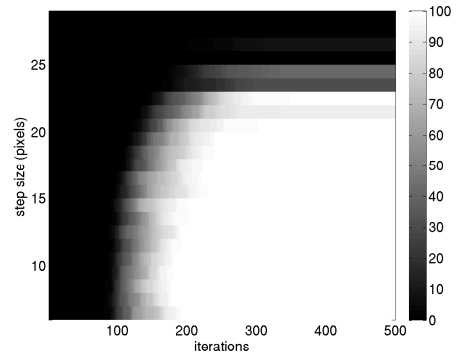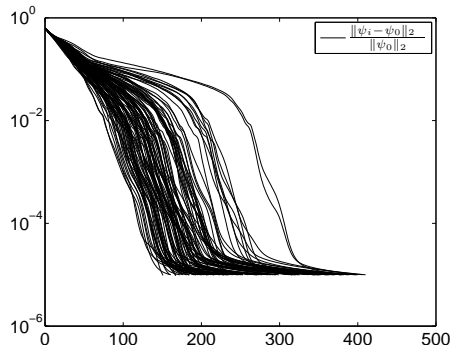
(a) convergence of CG from 100 random starts $\Delta x = 20$

(b) percentage tests that converge to $err \leq 10^{-4}$
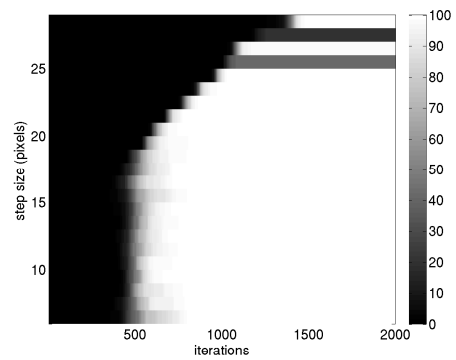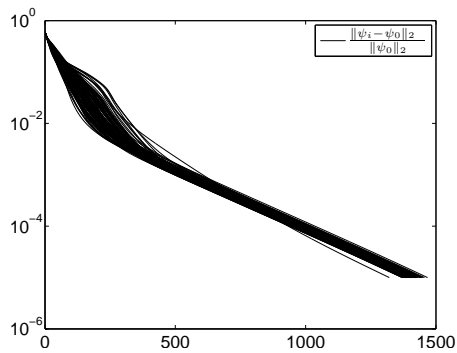
(c) convergence of RAAR from 100 random starts $\Delta x = 20$

(d) percentage of RAAR iterations that converge to an error of 10^{-4}

Figure 14: The convergence rate of the CG and RAAR methods from different random starting points.

We will show that H^ρ is diagonal dominant when there is a sufficient amount of overlap between adjacent diffraction frames. To simplify our discussion, let us assume for the moment that b_i is a 1D diffraction pattern obtained from a binary probe that illuminates three pixels at a time, and the probe is translated one pixel at a time so that the image frame that produces b_i overlaps with that produces b_{i-1} by two pixels. In this case, the $\hat{F}Q$ term in (46) has the form

$$
\begin{pmatrix}
f_1 & f_2 & f_3 & \cdots & 0 \\
0 & f_2 & f_3 & \ddots & \vdots \\
0 & 0 & f_3 & \ddots & f_k \\
f_1 & 0 & 0 & \ddots & f_k \\
f_1 & f_2 & 0 & \cdots & f_k
\end{pmatrix},
$$

where f_i is the ith column of F.

As a result, a typical diagonal term of H^ρ has the form

$$
H_{i,i}^\rho = f_i^* D_{i-2} f_i + f_i^* D_{i-1} f_i + f_i^* D_i f_i = \text{trace}(D_{i-2} + D_{i-1} + D_i), \tag{47}
$$

where D_i is a diagonal matrix that contains elements $1 - \beta_{ji}/(2\zeta_{ji})$ for $j = 1, 2, 3$.

When ψ is near the solution, z_i is close to b_i. Hence, D_i is likely to contain positive entries only. Therefore, the diagonal elements of H^ρ are likely to be much larger compared to the nonzero off-diagonal elements which contain terms in the form of either $f_j^* D_i f_\ell$ and its conjugate, where $j \neq \ell$, or $f_j^T E_i f_\ell$ and its conjugate, where E_i is a diagonal matrix (and part of B_{12}) that contains elements $\beta_{ji}\zeta_{ji}^2/(2|\zeta_{ji}|^3)$ for $j = 1, 2, 3$. Due to the phase difference between f_j and f_ℓ, D_i's do not add up "coherently" on the off-diagonal of H^ρ as they do on the diagonal. Neither do nonzero entries in E_i's add up coherently on the off-diagonal blocks of H^ρ either. Hence, the matrix H^ρ becomes diagonal dominant when there is larger amount of overlap between two adjacent frames. In fact, the diagonal of H^ρ may become so dominant that the spectral property of H^ρ is determined largely by the diagonal part of the matrix, which is typically well conditioned due to the averaging of D_i in (47). This observation provides an intuitive explaination on why increasing the amount of overlap between adjacent frames tends to improve the convergence rate of CG and other optimization based iterative ptychographical phase retrieval algorithms. Although this is not a precise analysis of the spectral property of H^ρ, the analysis does match with observations made in our numerical expriments. Moreover, this type of analysis can be extended to the 2D case in which F is represented as a tensor product of two 1D discrete Fourier transforms.

7 Conclusion

We formulated the ptychographic phase retrieval problem as a nonlinear optimization problem and discussed how standard iterative optimization algorithms can be applied to solve this problem.

We showed that the optimization problems we solve are not globally convex. Hence standard optimization algorithms can produce local minimizers. However, the Hessian of the objective functions we minimize do have special structures that may be exploited.

We compared the performance of several optimization algorithms and found that Newton's method with Steihaug's trust region technique gave the best performance on a real valued image. For a complex valued image, the nonlinear conjugate gradient algorithm appears to perform better.
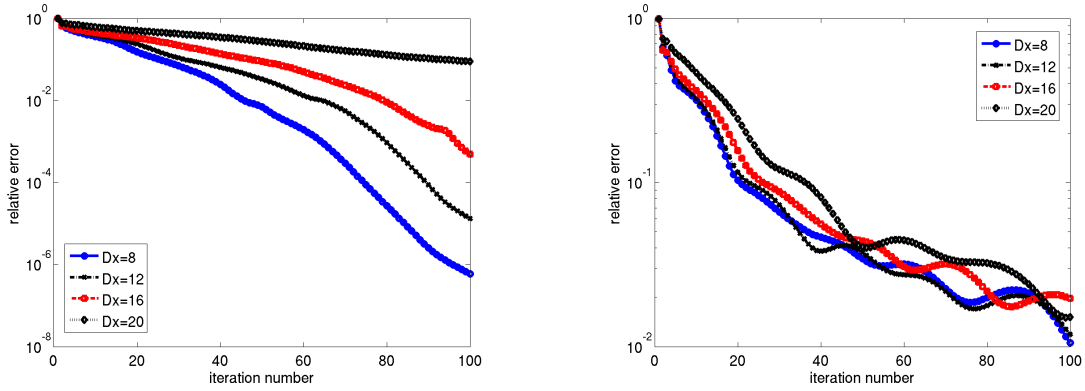
(a) The effect of overlapping on the convergence of CG for the gold ball image reconstruction.

(b) The effect of overlapping on the convergence of HIO for the gold ball image reconstruction.

Figure 15: The effect of overlapping on the convergence of CG and HIO algorithms.

We discussed the effect of preconditioning on convergence of the CG algorithm. We also demonstrated it is possible for an optimization algorithm to converge to a local minimizer although in practice such type of convergence failure is rare, especially when the amount of overlap between two adjacent diffraction frames is large.

We demonstrated by a numerical example that the convergence rate of an optimization algorithm depends on the amount of overlapping between two adjacent diffraction frames. We provided an intuitive analysis on why this occurs. More research is needed to provide a more precise analysis on this phenonmenon.

We identified the connection between the optimization based approach with both Wigner deconvolution and projection algorithms often used in phase retrieval literatures. We pointed out the limitation of Wigner deconvolution and showed that the optimization based algorithm tend to perform better than projection algorithms such as HIO when the amount of overlap between adjacent images is sufficiently large.

Acknowledgment

References

[1] R Hegerl and W Hoppe. Dynamic theory of crystalline structure analysis by electron diffraction in inhomogeneous primary wave field. *Berichte Der Bunsen-Gesellschaft Fur Physikalische Chemie*, 74:1148, 1970.

[2] P. D. Nellist, B. C. McCallum, and J. M. Rodenburg. Resolution beyond the 'information limit' in transmission electronmicroscopy. 374:630–632, 04 1995.

[3] H N Chapman. Phase-retrieval x-ray microscopy by wigner-distribution deconvolution. *Ultramicroscopy*, 66:153–172, 1996.

[4] H. M. L. Faulkner and J. M. Rodenburg. Movable aperture lensless transmission microscopy: a novel phase retrieval algorithm. *Phy. Rev. Lett.*, 93:023903, 2004.

[5] J. M. Rodenburg and H. M. L. Faulkner. A phase retrieval algorithm for shifting illumination. *Appl. Phy. Lett.*, 85:4795–4797, 2004.

[6] J M Rodenburg. Ptychography and related diffractive imaging methods. volume 150 of *Advances in Imaging and Electron Physics*, chapter Ptychography and Related Diffractive Imaging Methods, pages 87–184. Elsevier, 2008.

[7] M. Guizar-Sicairos and J. R. Fineup. Phase retrieval with transverse translation diversity: a nonlinear optimization approach. *Opt. Express*, 16:7264–7278, 2008.

[8] P. Thibault, M. Dierolf, O. Bunk, A. Menzel, and F. Pfeiffer. Probe retrieval in ptychographic coherent diffractive imaging. *Ultramicroscopy*, 109:338–343, 2009.

[9] J. M. Rodenburg and R. H. T. Bates. The theory of super-resolution electron microscopy via wigner-distribution deconvolution. *Phil. Trans. R. Soc. Lond. A*, 339:521–553, 1992.

[10] R. Remmert. *Theory of Complex Functions*. Springer-Verlag, 1991.

[11] K. Kreutz-Delgado. *The Complex Gradient Operator and the $\mathbb{CR}$-Calculus*. UCSD, 2003.

[12] T. Steihaug. The conjugate gradient method and trust regions in large scale optimization. *SIAM J. Num. Anal.*, 20:626–637, 1983.

[13] J. Moré and D. J. Thuente. Line search algorithms with guaranteed sufficient decrease. *ACM Trans. Math. Software*, 20(3):286–307, 1994.

[14] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer-Verlag, New York, 1999.

[15] J. R. Fienup. Phase retrieval algorithms: a comparison. *Appl. Opt.*, 21:2758–2769, 1982.

[16] R. Luke. Relaxed averaged alternating reflections for diffraction imaging. *Inverse Problems*, 21:37–50, 2005.

[17] S. Marchesini. Phase-retrieval and saddle-point optimization. *arXiv:physics*, 0611233v5:1–8, 2007.

[18] S Marchesini. A unified evaluation of iterative projection algorithms for phase retrieval. *Rev Sci Instrum*, 78:011301, Jan 2007.

[19] J. M. Rodenburg. Ptychography and related diffractive imaging methods. *Advances in Imaging and Electron Physics*, 150, 2008.

[20] C. T. Kelley. *Iterative Methods for Optimization*. SIAM, Philadelphia, 1999.