Phase Retrieval with Roughly Known Mask

Albert Fannjiang and Wenjing Liao

Department of Mathematics, University of California, Davis, CA 95616

Uniqueness, up to a global phase, is proved for phasing with a random roughly known mask (RKM) of high uncertainty. Phasing algorithms alternating between the object update and the mask update are systematically tested and demonstrated to have the capability of nearly perfect recovering the object and the mask (within the object support) simultaneously for up to 50% mask uncertainty. The uncertainty threshold for successful recovery can be further lifted to more than 70% under the additional sector condition on the object. The phasing algorithms with RKM are robust to low resolution masking as well as other types of noise.

1. Introduction

Phase retrieval is the problem of reconstructing an unknown image from its Fourier intensity data and is fundamental in many applications. Recent breakthroughs center around diffractive imaging of non-periodic objects, combining the penetration power of hard X-ray and the high sensitivity of lensless imaging [1–3].

Despite tremendous progresses many questions, fundamental as well as algorithmic, remain to be solved. For example, the standard phasing algorithms, based on alternating projections [4–6], are plagued by stagnation and spurious errors partly due to intrinsic non-uniqueness of the standard phasing problem.

We believe that the two problems, non-uniqueness and non-convergence, can be solved in one stroke. The present work is a sequel to our previous paper [8]. A centerpiece of our approach is the random-mask phasing method [7,8] which is shown to yield a unique solution,

up to at worst a global phase factor, as well as superior numerical performances, including rapid convergence, much reduced data and noise stability. In particular, the random mask method is robust to various types of noise, including Gaussian, Poisson and mask noises, with a noise amplification factor about 2.

A critique that can be leveled against the random mask approach is the use of exact knowledge of the mask which is not always available. Here we present an approach specifically to the problem of high mask uncertainty that goes well beyond the stability to mask errors demonstrated already in [8]. We will show that *nearly perfect* recovery can be achieved for more than 50% mask uncertainty.

Our approach is based on the stronger notion of uniqueness (i.e. *up to a global phase factor*), and we extend two uniqueness results of [7] to the present setting with roughly known mask (RKM). Instead of running phasing algorithms with a fixed erroneous mask, we design algorithms to recover the object and the mask simultaneously. At each iteration, the object and the mask are updated alternatively aiming at fitting the object constraint, the mask constraint as well as the Fourier intensity data. As shown below our numerical schemes can accurately recover the object with close to 50% mask errors and, under an additional sector condition, more than 70% mask errors.

The effect of a mask amounts to changing the original object f to the masked object

$$g(\mathbf{n}) = \lambda(\mathbf{n}) f(\mathbf{n}), \quad \mathbf{n} \in \mathbb{Z}^d \tag{1}$$

where λ is an array representing the mask and d is the dimension. In the standard phasing problem the uniform mask (UM) with $\lambda \equiv 1$ is used. In our approach the mask λ is random and only roughly known. For ease of discussion, we will focus on the case of phase masks

$$\lambda(\mathbf{n}) = \exp\left(i\phi(\mathbf{n})\right), \quad \phi(\mathbf{n}) \in [0, 2\pi], \ \mathbf{n} \in \mathbb{Z}^d \tag{2}$$

We assume that the true phase angles $\phi(\mathbf{n})$, in radian, of the *unknown* mask lie within $\delta\pi$ from the *given* initial estimates $\phi_0(\mathbf{n})$ for all $\mathbf{n}$, i.e.

$$\phi(\mathbf{n}) \in [\![\phi_0(\mathbf{n}) - \delta\pi, \phi_0(\mathbf{n}) + \delta\pi]\!]. \tag{3}$$

Here and below we adopt the following notation: $\theta \in [\![a, b]\!]$ means

$$\begin{cases} a(\mathrm{mod}\, 2\pi) \le \theta(\mathrm{mod}\, 2\pi) \le b(\mathrm{mod}\, 2\pi) & \text{if } a(\mathrm{mod}\, 2\pi) \le b(\mathrm{mod}\, 2\pi) \\ a(\mathrm{mod}\, 2\pi) \le \theta(\mathrm{mod}\, 2\pi) < 2\pi \text{ or } 0 \le \theta \le b(\mathrm{mod}\, 2\pi) & \text{else} \end{cases}.$$

For ease of notation, we shall write below the inequality (3) as $\phi(\mathbf{n}) \in [\![\phi_0(\mathbf{n}) \mp \delta\pi]\!]$.

The paper is organized as follows. We state the uniqueness theorems for phasing with a random RKM in Section 2 and give the proofs in Appendices A and B. We discuss the basic algorithm of Alternating Error Reduction (AER) and prove the residual reduction property in Section 3 and Appendix C. We discuss the Douglas-Rachford-Error-Reduction (DRER) algorithm in Section 4 and the algorithms with two sets of Fourier intensity data in Section 5. We present numerical results in Section 6 and conclude in Section 7.

2. Uniqueness

Let $\mathbf{n} = (n_1, \ldots, n_d) \in \mathbb{Z}^d$ and $\mathbf{z} = (z_1, \ldots, z_d) \in \mathbb{C}^d$. Define the multi-index notation $\mathbf{z}^{\mathbf{n}} = z_1^{n_1} z_2^{n_2} \ldots z_d^{n_d}$. Let $\mathcal{C}(\mathcal{N})$ denote the set of finite complex-valued functions on $\mathbb{Z}^d$ vanishing outside $\mathcal{N} = \{\mathbf{0} \leq \mathbf{n} \leq \mathbf{N}\}, \mathbf{N} = (N_1, N_2, \ldots, N_d)$. Here $\mathbf{m} \leq \mathbf{n}$ if $m_j \leq n_j, \forall j$. Set $|\mathcal{N}| = \prod_{j=1}^{d}(N_j + 1)$.

The $\mathbf{z}$-transform $F(\mathbf{z}) = \sum_{\mathbf{n}} f(\mathbf{n})\mathbf{z}^{-\mathbf{n}}$ of $f \in \mathcal{C}(\mathcal{N})$ is an analytic continuation, from the d-dimensional torus, of the the Fourier transform for $\mathbf{z} = (\exp(2\pi i\omega_1), \ldots, \exp(2\pi i\omega_d)), \omega_j \in [0, 1]$.

If the absolute value of the Fourier transform is sampled on the lattice

$$\mathcal{L} = \left\{\boldsymbol{\omega} = (\omega_1, ..., \omega_d) \mid \omega_j = 0, \frac{1}{2N_j + 1}, \frac{2}{2N_j + 1}, ..., \frac{2N_j}{2N_j + 1}\right\} \tag{4}$$

then the autocorrelation function $\mathcal{C}_f(\mathbf{n}) = \sum_{\mathbf{m}\in\mathcal{N}} f(\mathbf{m} + \mathbf{n})\overline{f(\mathbf{m})}$ is uniquely determined. Sampling on $\mathcal{L}$ produces the oversampling ratio $\sigma \approx 2^d$ where

$$\sigma \quad = \quad \frac{\text{Fourier intensity data number}}{\text{unknown image pixel number}}.$$

However, the uniqueness of the autocorrelation given the Fourier intensity data does not imply the uniqueness of the object.

First, there are three types of *global* ambiguities: (a) constant global phase, $f(\cdot) \longrightarrow \exp(i\theta)f(\cdot)$, for some $\theta \in [0, 2\pi]$, (b) spatial shift, $f(\cdot) \longrightarrow f(\cdot + \mathbf{m})$, for some $\mathbf{m} \in \mathbb{Z}^d$, (c) conjugate inversion, $f(\cdot) \longrightarrow \overline{f(\mathbf{N} - \cdot)}$. Conjugate inversion produces the so-called twin image.

Our basic tool is the following improved result of irreducibility [9, 10] with, however, a more practical and useful perspective.

Proposition 1. [7] *Let f be a finite complex-valued object of rank ≥ 2. Let $\{\lambda(\mathbf{n})\}$ be continuous random variables on nonzero real algebraic varieties $\{\mathcal{V}(\mathbf{n})\}$ in $\mathbb{C}(\simeq \mathbb{R}^2)$ with an* absolutely continuous *joint distribution with respect to the standard product measure on*

$\prod_{\mathbf{n}\in\Sigma}\mathcal{V}(\mathbf{n})$ *where* $\Sigma\subset\mathbb{N}^d$ *is the support set of* f*. Then the* $\mathbf{z}$*-transform of the masked object (1) is irreducible up to a power of* $\mathbf{z}^{-1}$ *with probability* 1*.*

The main point here is that while the classical result [9,10] works for generic (thus random) objects from a certain ensemble Proposition 1 deals with a *given, deterministic* object of rank ≥ 2. This improvement is achieved by endowing the probability measure on the ensemble of masks, which we can manipulate, instead of the space of objects, which we can not control, as in the standard setting.

As a consequence of Proposition 1, the global ambiguities are the sole ambiguities possible as far as the masked object (1) is concerned [10].

This is not all. We are able to remove all ambiguities, with the only exception of a global phase, for the original object f even when the random mask is only roughly known.

Theorem 1. *Let* f *be a real-valued object of rank* ≥ 2*. Suppose the exact mask phases* $\{\phi(\mathbf{n})\}$ *are independently and uniformly distributed on* $[0,2\pi]$*. Suppose the uncertainty of the mask estimate* $\lambda_0=\{\exp(i\phi_0(\mathbf{n}))\}$ *in (3) is* $\delta<0.5$*.*

Suppose that another non-negative image $\tilde{f}$ *and mask estimate* $\tilde{\lambda}=\{\exp(i\tilde{\phi}(\mathbf{n}))\}$ *satisfying*

$$\tilde{\phi}(\mathbf{n})\in[\![\phi_0(\mathbf{n})\mp\delta\pi]\!] \tag{5}$$

together produce the same Fourier intensity data on $\mathcal{L}$ *as do* f *and* λ*. Then, with probability no less than* $1-|\mathcal{N}|(2\delta)^{\lfloor S/2\rfloor}$*,* $\tilde{f}(\mathbf{n})=f(\mathbf{n})\ \forall\mathbf{n}$ *and furthermore* $\tilde{\phi}(\mathbf{n})=\theta+\phi(\mathbf{n})$ *for a constant* $\theta\in\mathbb{R}$ *wherever* $f(\mathbf{n})\neq 0$*. Here a positive number* x*,* $\lfloor x\rfloor$ *is the greatest integer at most* x*.*

Remark 1. *If the object is known to be non-negative, then* δ *can be any number in* $[0,1)$ *and uniqueness holds with probability no less than* $1-|\mathcal{N}|\delta^{\lfloor S/2\rfloor}$*.*

For complex-valued objects uniqueness requires two independent sets of Fourier intensity data.

Theorem 2. *Let* f *be a complex-valued object of* rank ≥ 2*. Let the first mask* $\lambda^{(1)}=\lambda$ *in Theorem 1 with the initial mask estimate* λ_0 *satisfying (5).*

Suppose the second mask $\lambda^{(2)}$ *is exactly known and the* $\mathbf{z}$*-transform of* $\lambda^{(2)}f$ *is irreducible up to a power of* $\mathbf{z}^{-1}$*. Moreover, assume the non-degeneracy condition that there is no* $\mathbf{m}\neq\mathbf{0}$ *such that* $\lambda^{(2)}(\mathbf{n}+\mathbf{m})f(\mathbf{n}+\mathbf{m})=\exp(i\xi)\exp(i\eta(\mathbf{n}))\lambda^{(2)}(\mathbf{n})f(\mathbf{n}),\forall\mathbf{n}$*, or no* $\mathbf{m}$ *such that* $\overline{\lambda^{(2)}(\mathbf{m}-\mathbf{n})f(\mathbf{m}-\mathbf{n})}=\exp(i\xi)\exp(i\eta(\mathbf{n}))\lambda^{(2)}(\mathbf{n})f(\mathbf{n}),\forall\mathbf{n}$*, for some* $\xi\in[0,2\pi],|\eta(\mathbf{n})|\leq\pi\delta$*.*

Suppose that for a phase mask $\tilde{\lambda}$ *with (5) and an object* $\tilde{f}$ *the two pairs of masked objects* λf *and* $\tilde{\lambda}\tilde{f}$*,* $\lambda^{(2)}f$ *and* $\lambda^{(2)}\tilde{f}$*, respectively, produce the same Fourier magnitudes*

on $\mathcal{L}$*. Then with probability no less than* $1 - |\mathcal{N}|\delta^{\lfloor S/2 \rfloor}$ $\tilde{f}(\mathbf{n}) = \exp(i\nu_1) f(\mathbf{n}), \forall \mathbf{n}$*, and* $\tilde{\lambda}(\mathbf{n}) = \exp(i\nu_2)\lambda(\mathbf{n})$ *if* $f(\mathbf{n}) \neq 0$*, where* $\nu_1, \nu_2 \in [0, 2\pi]$.

Remark 2. *Clearly, most objects and masks obey the non-degeneracy condition.*

The proofs of Theorems 1 and 2 are given in Appendices A and B, respectively.

Both theorems assert that not only the uniqueness of the object but also the uniqueness of the mask, up to a constant phase, inside the object support. This motivates the design of our numerical schemes that turn out to be capable of enforcing uniqueness and hence nearly perfect recovery even in the presence of high mask uncertainty.

3. Alternating Error Reduction (AER)

Let Λ be the diagonal matrix with diagonal elements $\{\lambda(\mathbf{n})\}$ and let Φ represent the discrete Fourier transform. Denote the Fourier intensity data vector by $Y = |\Phi\Lambda f|$ where $\Lambda f(\mathbf{n}) = \lambda(\mathbf{n}) f(\mathbf{n})$.

3.A. Object Update

Given the object estimate f_k and mask estimate λ_k at the k-th iteration, we use standard phasing algorithms to obtain f_{k+1}.

Let $\mathcal{O}$ denote the ensemble of objects $\tilde{f}$ satisfying various object constraints (see Section 6). Let $\mathcal{P}_{\mathrm{o}}$ be the orthogonal projection onto $\mathcal{O}$ (cf. [8] for details about numerical implementation of $\mathcal{P}_{\mathrm{o}}$) and $\mathcal{P}_{\mathrm{f},k} = \Lambda_k^{-1}\Phi^{-1}\mathcal{T}\Phi\Lambda_k$, where $\mathcal{T}$ is the intensity fitting operator

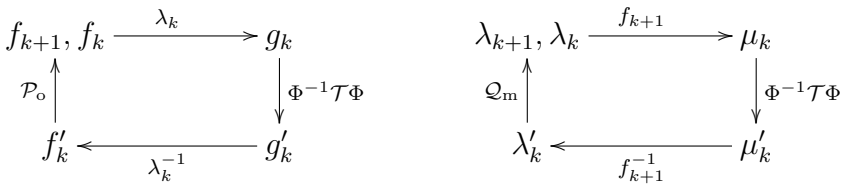
$$\mathcal{T}G(\boldsymbol{\omega}) = \begin{cases} Y(\boldsymbol{\omega}) \exp(i\measuredangle G(\boldsymbol{\omega})) & \text{if } |G(\boldsymbol{\omega})| > 0 \\ Y(\boldsymbol{\omega}) & \text{if } |G(\boldsymbol{\omega})| = 0 \end{cases}. \tag{6}$$

Here and below $\measuredangle z \in [0, 2\pi)$ denotes the wrapped phase angle of z. When $z = 0$, $\measuredangle z$ is taken to be 0 unless specified otherwise.

Error Reduction (ER) takes the form $f_{k+1} = \mathcal{P}_{\mathrm{o}}\mathcal{P}_{\mathrm{f},k} f_k$ which can be conveniently represented by the diagram in Figure 1(a).

Let $r(\tilde{f}, \tilde{\lambda}) = \| \, |\Phi\tilde{\Lambda}\tilde{f}| - Y \, \|$ denote the residual. With a phase mask, ER enjoys the residual reduction property [8]:

$$r(f_{k+1}, \lambda_k) \leq r(f_k, \lambda_k). \tag{7}$$



(a) object update (b) mask update

Fig. 1. Alternating Error Reduction (AER) between object and mask.

3.B. Mask Update

Based on the newly updated object estimate f_{k+1}, we define $\mathcal{Q}_{\mathrm{f},k}$ as

$$\lambda'_k = \mathcal{Q}_{\mathrm{f},k}\lambda_k(\mathbf{n}) = \begin{cases} \Phi^{-1}\mathcal{T}\Phi\Lambda_k f_{k+1}(\mathbf{n})/f_{k+1}(\mathbf{n}) & \text{if } f_{k+1}(\mathbf{n}) \neq 0 \\ \\ \lambda_k(\mathbf{n}) & \text{else} \end{cases} . \tag{8}$$

Let $\mathcal{M}$ be the ensemble of phase masks satisfying the phase uncertainty constraint (5):

$$\mathcal{M} \;=\; \{\tilde{\lambda} \mid \forall \mathbf{n}, \measuredangle\tilde{\lambda}(\mathbf{n}) \in [\![\phi_0(\mathbf{n}) \mp \delta\pi]\!]\}. \tag{9}$$

Let $\mathcal{Q}_{\mathrm{m}}$ be the orthogonal projection onto $\mathcal{M}$. Note that $\mathcal{Q}_{\mathrm{m}}$ can be computed pixel by pixel as follows.

Let $a = (\phi_0(\mathbf{n}) - \delta\pi)(\mathrm{mod}\ 2\pi)$, $b = (\phi_0(\mathbf{n}) + \delta\pi)(\mathrm{mod}\ 2\pi)$ and

$$c = \begin{cases} \pi + (a+b)/2\,(\mathrm{mod}\,2\pi), & \text{if } \ a \leq b \\ (a+b)/2\,(\mathrm{mod}\,2\pi), & \text{else.} \end{cases}$$

Then $\mathcal{Q}_{\mathrm{m}}$ can be expressed as

$$\mathcal{Q}_{\mathrm{m}}\lambda'_k(\mathbf{n}) = \begin{cases} \exp\left(i\measuredangle\lambda'_k(\mathbf{n})\right) & \text{if } \measuredangle\lambda'_k(\mathbf{n}) \in [\![a, b]\!] \\ \exp\left(ib\right) & \text{if } \measuredangle\lambda'_k(\mathbf{n}) \in [\![b, c]\!] \\ \exp\left(ia\right) & \text{if } \measuredangle\lambda'_k(\mathbf{n}) \in [\![c, a]\!]. \end{cases} \tag{10}$$

Since the the object and the mask have interchangeable roles, we set $\lambda_{k+1} = \mathcal{Q}_{\mathrm{m}}\mathcal{Q}_{\mathrm{f},k}\lambda_k$ in the spirit of ER (see Fig. 1(b)). Note the differences between the mask update rule here and that of the extended ptychographical engine (ePIE) ((4) in Ref. [11]): First, (8) uses the newly updated object f_{k+1} while ePIE uses the previous one. Second, more importantly, the rough prior knowledge about the mask is enforced by $\mathcal{Q}_{\mathrm{m}}$ here while ePIE does not consider this aspect.

Now we prove the following residual reduction property.

Lemma 1. *With $\mathcal{Q}_{\rm m}$ we have*

$$r(f_{k+1},\lambda_{k+1}) \le r(f_{k+1},\lambda_k)$$

The proof of Lemma 1 is given in Appendix C.

Define the Alternating Error Reduction (AER) as

$$(f_{k+1},\lambda_{k+1}) \;=\; (\mathcal{P}_{\rm o}\mathcal{P}_{{\rm f},k}f_k, \mathcal{Q}_{\rm m}\mathcal{Q}_{{\rm f},k}\lambda_k). \tag{11}$$

In words, AER alternates between updating the object and the mask estimates until the iteration converges.

Lemma 1 and (7) together yield the following residual reduction property for AER.

Theorem 3. *AER (11) has the residual reduction property: $r(f_{k+1},\lambda_{k+1}) \le r(f_k,\lambda_k)$.*

In our numerical tests, we find that while $\mathcal{Q}_{\rm m}$ works well for real-valued objects, for complex-valued objects the following alternative rule works better

$$\hat{\mathcal{Q}}_{\rm m}\lambda'_k(\mathbf{n}) = \begin{cases} \exp\left(i\measuredangle\lambda'_k(\mathbf{n})\right) & \text{if } \measuredangle\lambda'_k(\mathbf{n}) \in [\![a,b]\!] \\ \lambda_0(\mathbf{n}), & \text{else.} \end{cases} \tag{12}$$

With $\hat{\mathcal{Q}}_{\rm m}$ we have the alternative version of AER

$$(f_{k+1},\lambda_{k+1}) \;=\; (\mathcal{P}_{\rm o}\mathcal{P}_{{\rm f},k}f_k, \hat{\mathcal{Q}}_{\rm m}\mathcal{Q}_{{\rm f},k}\lambda_k). \tag{13}$$

4. DRER

AER (either version) by itself converges slowly, typically taking up to several thousands steps for accurate recovery in our numerical tests. To speed up convergence we consider the Douglas Rachford (DR) algorithm [12], also called the averged successive reflections (ASR), which coincides with the hybrid input-output (HIO) algorithm for the parameter $\beta = 1$ when object value constraints are not imposed [13]

$$f_{k+1} = \frac{I + \mathcal{R}_{\rm o}\mathcal{R}_{{\rm f},k}}{2} f_k \tag{14}$$

where $\mathcal{R}_{\rm o} = 2\mathcal{P}_{\rm o} - I, \quad \mathcal{R}_{{\rm f},k} = 2\mathcal{P}_{{\rm f},k} - I$ are reflection operators.

Define the DRER iteration as

$$(f_{k+1},\lambda_{k+1}) \;=\; \left(\frac{1}{2}(I + \mathcal{R}_{\rm o}\mathcal{R}_{{\rm f},k})f_k, \mathcal{Q}_{\rm m}\mathcal{Q}_{{\rm f},k}\lambda_k\right) \tag{15}$$

and the alternative version as

$$(f_{k+1},\lambda_{k+1}) \;=\; \left(\frac{1}{2}(I + \mathcal{R}_{\rm o}\mathcal{R}_{{\rm f},k})f_k, \hat{\mathcal{Q}}_{\rm m}\mathcal{Q}_{{\rm f},k}\lambda_k\right). \tag{16}$$

To strictly enforce the mask constraint, we do not use DR for mask update.

5. AER/DRER with two sets of data

Let $\lambda^{(1)} = \lambda$ and $\lambda^{(2)}$ be two masks with which two sets of Fourier intensity data $Y = |\Phi\Lambda f|$ and $Y^{(2)} = |\Phi\Lambda^{(2)} f|$ are measured on $\mathcal{L}$. Let $\mathcal{T}$ and $\mathcal{T}^{(2)}$ be the intensity fitting operators corresponding to Y and $Y^{(2)}$, respectively.

For simplicity of presentation we assume the second mask (random or deterministic) is exactly known and independent from the first mask which is random. In this case, there is no need for the second mask update.

Suppose f_k and λ_k are the image and the mask recovered at the end of the k-th iteration. At the $(k+1)$-st iteration, the image is first updated from f_k to f_{k+1} based on λ_k and $\lambda^{(2)}$. Let $\mathcal{P}_k = (\Lambda_k)^{-1}\Phi^{-1}\mathcal{T}\Phi\Lambda_k$ and $\mathcal{P}^{(2)} = (\Lambda^{(2)})^{-1}\Phi^{-1}\mathcal{T}^{(2)}\Phi\Lambda^{(2)}$.

The AER and DRER algorithms with two masks are respectively

$$(f_{k+1}, \lambda_{k+1}) = \left(\mathcal{P}_{\rm o}\mathcal{P}^{(2)}\mathcal{P}_k f_k, \mathcal{Q}_m\mathcal{Q}_{{\rm f},k}\lambda_k\right), \quad k = 0, 1, \cdots \tag{17}$$

$$(f_{k+1}, \lambda_{k+1}) = \left(\frac{1}{2}(I + \mathcal{R}_{\rm o}(2\mathcal{P}^{(2)}\mathcal{P}_k - I))f_k, \mathcal{Q}_m\mathcal{Q}_{{\rm f},k}\lambda_k\right). \tag{18}$$

The following alternative versions work better for complex-valued objects:

$$(f_{k+1}, \lambda_{k+1}) = \left(\mathcal{P}_{\rm o}\mathcal{P}^{(2)}\mathcal{P}_k f_k, \hat{\mathcal{Q}}_{\rm m}\mathcal{Q}_{{\rm f},k}\lambda_k\right), \quad k = 0, 1, \cdots \tag{19}$$

$$(f_{k+1}, \lambda_{k+1}) = \left(\frac{1}{2}(I + \mathcal{R}_{\rm o}(2\mathcal{P}^{(2)}\mathcal{P}_k - I))f_k, \hat{\mathcal{Q}}_{\rm m}\mathcal{Q}_{{\rm f},k}\lambda_k\right) \tag{20}$$

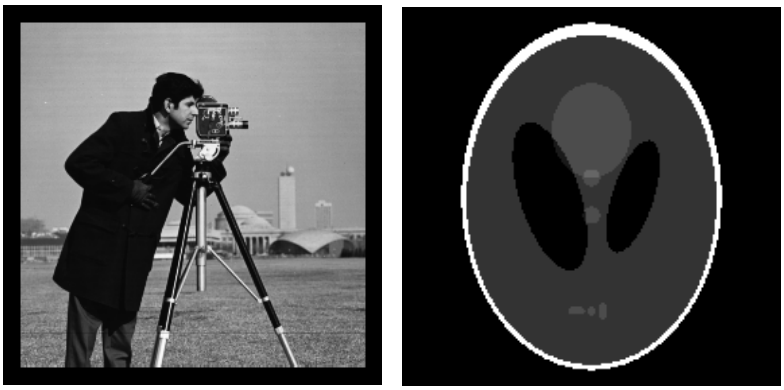
6. Numerical Simulations

In this section, we test our numerical schemes by performing phasing with a RKM.

The original images are the 256×256 cameraman and the 138×184 phantom. We surround both images by dark (i.e. zero-valued) borders to create images of loose support. The resulting images to be recovered are 282×282 cameraman and 200×200 phantom of loose supports (Fig. 2). Objects of loose support are harder to recover than objects of tight support. Everything else being the same, the larger the dark borders the harder it is to recover the objects.

We set the true phases of the RKM to be $\phi(\mathbf{n}) = \phi_0(\mathbf{n}) + t(\delta, \mathbf{n})$ where $t(\delta, \mathbf{n})$ are independent, uniform random variables in $[-\pi\delta, \pi\delta]$ with uncertainty $\delta \in [0, 1)$.

The mask is called a *high resolution* mask (HRM) if $\phi_0(\mathbf{n})$ are independently and uniformly distributed in $[0, 2\pi]$ (assumed in Theorems 1 and 2) and a *low resolution* mask (LRM) if $\phi_0(\mathbf{n})$ are not independently distributed but grouped in independently distributed 10×10-blocks, each of which is a (random) constant. For our examples, a LRM has less than 900



(a) 282×282 cameraman (b) 200×200 phantom

Fig. 2. Test images of loose support.

and 400 degrees of freedom, respectively, for the cameraman and the phantom. A HRM, however, always performs better than a LRM as HRM has many more (100 times) degrees of freedom (cf. Fig. 8). For 200×200 HRM and 20×20 LRM used for Fig. 8, see Fig. 3 where the gray level represents the mask phases in $[0, 2\pi]$.

When a second set of Fourier data is used (for complex-valued objects), it is generated with a UM (i.e. $\lambda^{(2)} = 1$).

A standard way to utilize the oversampled data for $\sigma > 1$ is to enlarge the original image by adding corresponding number of zero pixels which is then enforced as an additional object constraint. This procedure is called the oversampling method [14] and implemented in all our simulations with $\sigma = 4$.

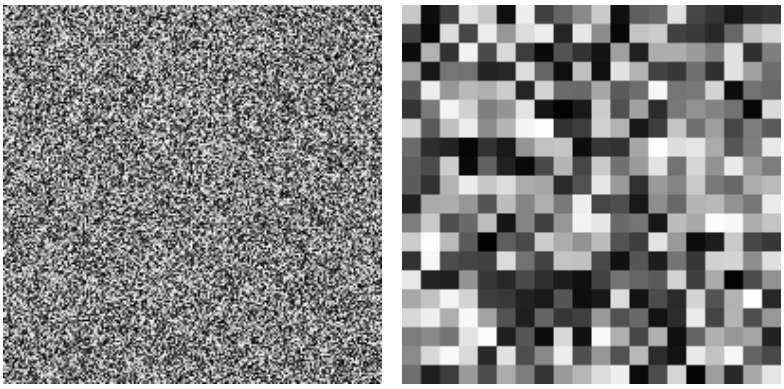
6.A. Error and Residual

To estimate the recovery, we define relative error and relative residual as follows. Let $\hat{f}$ and $\hat{\lambda}$ be the recovered image and mask respectively. The relative error of reconstruction is defined as

$$e(\hat{f}) = \begin{cases} \|f - \hat{f}\|/\|f\| & \text{if absolute uniqueness holds} \\ \min\limits_{\nu \in [0,2\pi)} \|f - \exp(i\nu)\hat{f}\|/\|f\| & \text{if uniqueness holds only up to a global phase} \end{cases}.$$

Let $\hat{\Lambda}$ be the diagonal matrix whose diagonal elements are $\hat{\lambda}(\mathbf{n})$. The relative residual is defined as

$$\rho(\hat{f}, \hat{\lambda}) = \frac{\| \, I - |\Phi \hat{\Lambda} \mathcal{P}_{\mathrm{o}} \hat{f}| \, \|}{\|I\|}$$



(a) 200×200 HRM (b) 20×20 LRM

Fig. 3. High (a) and low (b) resolution masks

where $\mathcal{P}_{\text{o}}$ is introduced to enforce the object constraints in the case of DRER.

6.B. Performance study of AER and DRER

First we use AER (11) to recover the non-negative images with the stopping rule $\|f_{k+1} - f_k\|/\|f_k\| < 0.05\%$ and one LRM of uncertainty $\delta = 0.2$. The results, shown in Fig. 4, are noisy and inaccurate with 33.75% error for the cameraman and 44.74% error for the phantom. Consistent with the residual reduction property (Theorem 3), the residual curves in Fig. 4 are monotonically decreasing.

Much improvement can be gained by running DRER, followed by AER. For real-valued objects, we use the version of DRER (15). DRER (15) is stopped when $\|f_{k+1} - f_k\|/\|f_k\| < 1\%$, with the maximum of 400 steps, and AER (11) is terminated when $\|f_{k+1} - f_k\|/\|f_k\| < 0.05\%$, with the maximum of 400 steps. As shown in Figure 5, for the cameraman, 131 DRER and 6 AER steps took place producing 1.43% error while for the phantom, 75 DRER and 5 AER steps took place with 0.33% error. Consistent with Theorem 1, the mask errors occur only outside the object supports.

Next we consider the case of the generic complex-valued objects with one UM and one LRM of uncertainty $\delta = 0.2$. We apply the alternative versions of DRER (20) and AER (19) which tend to produce better results than (15) and (11) for complex-valued objects. DRER (20) is stopped when $\|f_{k+1} - f_k\|/\|f_k\| < 1\%$, with the maximum of 400 steps, and AER (19) is terminated when $\|f_{k+1} - f_k\|/\|f_k\| < 0.05\%$, with the maximum of 400 steps. Fig. 6 shows the results for object phases randomly distributed on $[0, 2\pi]$. Both algorithms ran

their full course of 400 steps with 2.39% error for the cameraman and 1.37% error for the phantom. The mask errors occur only outside the object supports, consistent with Theorem 2.

Still next, we let the object phases be randomly distributed on $[0, \pi/2]$ (the sector constraint) and include this prior knowledge into $\mathcal{P}_o$. With the additional sector constraint, we found that the following stopping rule can significantly reduce the number of iterations: DRER (20) is stopped if the residual increases in five consecutive steps, with the maximum of 400 steps, and AER (19) is terminated when $\|f_{k+1} - f_k\|/\|f_k\| < 0.05\%$, with the maximum of 400 steps. Fig. 7 shows the results with one UM and one LRM of uncertainty $\delta = 0.2$. With the new stopping rule and the sector constraint, 24 DRER and 400 AER steps took place with 1.95% error for the cameraman and 24 DRER and 365 AER steps with 0.61 % error for the phantom.

Finally Fig. 8 shows the averaged relative error $e(\hat{f})$, after 5 runs of independently chosen initial guesses for the object, with or without mask update, as a function of the mask uncertainty of HRM or LRM for (a) non-negative phantom, (b) generic complex-valued phantom and (c) phantom under the sector condition. We use the same stopping rules and updating rules as above for each case, except that the maximum number of steps is changed to $200 + \delta \cdot 1000$ for DRER and AER separately to deal with variable uncertainty.

Without mask update the error curves are roughly linear with the noise amplification factor roughly 2 (blue and black curves), consistent with our previous results reported in [8]. With mask update, the results (pink and red curves) are drastically improved for uncertainty up to 50% (Fig. 8 (a) and (b)) and, under the sector constraint, nearly 70% uncertainty in the mask phases (Fig. 8 (c)).

7. Conclusion

We have proved the uniqueness, up to a global phase, for phasing with RKM with probability exponentially close to one, depending on the object sparsity and mask uncertainty. We have designed algorithms that achieve nearly perfect recovery for mask uncertainty up to half of that promised by the uniqueness results. Under the additional sector condition, the algorithms can achieve nearly perfect recovery for up to 70% uncertainty in the mask. As a by-product of object recovery the unknown mask can be recovered accurately within the object support. The numerical performance is robust to low resolution mask as well as other types of noises as shown in the previous study [8].

Our method can be easily extended to general masks as long as they modulate object *phases* randomly (i.e. they are not just amplitude masks). This can be done by first re-

constructing not the original object but the object multiplied with the intensities (assumed nonzero) of the mask by using the normalized (phase) mask. In the second step, the original object is recovered by dividing by the intensities of the mask.

A. Proof of Theorem 1

Proof. In view of Proposition 1, there exist some $\mathbf{m}$ and $\theta \in [0, 2\pi]$ such that either

$$\tilde{\lambda}(\mathbf{n})\tilde{f}(\mathbf{n}) = \exp(i\theta)\lambda(\mathbf{m}+\mathbf{n})f(\mathbf{m}+\mathbf{n}) \tag{21}$$

or

$$\tilde{\lambda}(\mathbf{n})\tilde{f}(\mathbf{n}) = \exp(i\theta)\overline{\lambda(\mathbf{m}-\mathbf{n})f(\mathbf{m}-\mathbf{n})}. \tag{22}$$

In the case of (21) with any $\mathbf{m} \neq \mathbf{0}$ and any $\theta \in [0, 2\pi]$,

$$\tilde{f}(\mathbf{n}) = \exp(i\theta)\frac{|\lambda(\mathbf{m}+\mathbf{n})|\exp(i\phi(\mathbf{n}+\mathbf{m}))}{|\tilde{\lambda}(\mathbf{n})|\exp(i\measuredangle\tilde{\lambda}(\mathbf{n}))}f(\mathbf{n}+\mathbf{m}).$$

Consider the $\lfloor S/2 \rfloor$ independently distributed r.v.s. of $\lambda(\mathbf{n}+\mathbf{m})$ where $f(\mathbf{n}+\mathbf{m}) \neq 0$ corresponding to $\lfloor S/2 \rfloor$ nonoverlapping pairs of points $\{\mathbf{n}, \mathbf{n}+\mathbf{m}\}$. For every $\mathbf{n}$ where $f(\mathbf{n}+\mathbf{m}) \neq \mathbf{0}$, a proper choice of $\measuredangle\tilde{\lambda}(\mathbf{n})$ makes $\tilde{f}(\mathbf{n})$ real-valued if and only if either

$$\phi(\mathbf{n}+\mathbf{m}) \in [\![\phi_0(\mathbf{n}) - \theta - \measuredangle f(\mathbf{n}+\mathbf{m}) \mp \delta\pi]\!]$$

or

$$\phi(\mathbf{n}+\mathbf{m}) \in [\![(\phi_0(\mathbf{n}) - \theta - \measuredangle f(\mathbf{n}+\mathbf{m}) + \pi) \mp \delta\pi]\!]$$

However, $\phi(\mathbf{n}+\mathbf{m})$ is independently and uniformly distributed in $[0, 2\pi]$, so it falls in these two regions with probability 2δ. The probability for every such $\tilde{f}(\mathbf{n})$ to be real-valued is 2δ and hence the probability for all $\tilde{f}(\mathbf{n})$ with $\mathbf{m} \neq 0$ to be real-valued is at most $(2\delta)^{\lfloor S/2 \rfloor}$.

The union over $\mathbf{m} \neq 0$ of these events has probability at most $|\mathcal{N}|(2\delta)^{\lfloor S/2 \rfloor}$. Therefore, with probability at least $1 - |\mathcal{N}|(2\delta)^{\lfloor S/2 \rfloor}$, $\mathbf{m} = \mathbf{0}$ and $\exp(i\theta)\lambda(\mathbf{n})f(\mathbf{n}) = \tilde{\lambda}(\mathbf{n})\tilde{f}(\mathbf{n})$ $\forall \mathbf{n}$ which further implies that $\tilde{f}(\mathbf{n}) = \pm f(\mathbf{n})$ $\forall \mathbf{n}$ and $\tilde{\lambda}(\mathbf{n}) = \pm\exp(i\theta)\lambda(\mathbf{n})$ on $\mathbf{n}$ where $f(\mathbf{n}) \neq 0$.

Likewise the probability for all $\tilde{f}(\mathbf{n})$ given by (22) to be real-valued for any $\mathbf{m}$ is at most $|\mathcal{N}|(2\delta)^{\lfloor S/2 \rfloor}$. □

B. Proof of Theorem 2

Proof. In view of Proposition 1, for some $\mathbf{m}_1, \mathbf{m}_2$ and $\theta_1, \theta_2 \in [0, 2\pi]$ either

$$\exp(i\theta_1)\lambda(\mathbf{n}+\mathbf{m}_1)f(\mathbf{n}+\mathbf{m}_1) = \tilde{\lambda}(\mathbf{n})\tilde{f}(\mathbf{n}) \tag{23}$$

or

$$\exp(i\theta_1)\overline{\lambda(\mathbf{m}_1-\mathbf{n})f(\mathbf{m}_1-\mathbf{n})}=\tilde{\lambda}(\mathbf{n})\tilde{f}(\mathbf{n}) \tag{24}$$

as well as

$$\exp(i\theta_2)\lambda^{(2)}(\mathbf{n}+\mathbf{m}_2)f(\mathbf{n}+\mathbf{m}_2)=\lambda^{(2)}(\mathbf{n})\tilde{f}(\mathbf{n}) \tag{25}$$

or

$$\exp(i\theta_2)\overline{\lambda^{(2)}(\mathbf{m}_2-\mathbf{n})f(\mathbf{m}_2-\mathbf{n})}=\lambda^{(2)}(\mathbf{n})\tilde{f}(\mathbf{n}). \tag{26}$$

There are four possible combinations of (23), (24), (25) and (26).

In the case of (23)&(25), we have

$$\exp(i\theta_1)\lambda(\mathbf{n}+\mathbf{m}_1)\lambda^{(2)}(\mathbf{n})f(\mathbf{n}+\mathbf{m}_1)=\exp(i\theta_2)\lambda^{(2)}(\mathbf{n}+\mathbf{m}_2)\tilde{\lambda}(\mathbf{n})f(\mathbf{n}+\mathbf{m}_2). \tag{27}$$

For any $\mathbf{m}_1\neq\mathbf{0}$ and any $\theta_1,\theta_2\in[0,2\pi]$, consider the $\lfloor S/2\rfloor$ pairs of independently distributed r.v.s. of $\lambda(\mathbf{n}+\mathbf{m}_1)$ where $f(\mathbf{n}+\mathbf{m}_1)\neq 0$ corresponding to $\lfloor S/2\rfloor$ non overlapping sets of points $\{\mathbf{n},\mathbf{n}+\mathbf{m}_1\}$. For every $\mathbf{n}$, a proper choice of $\tilde{\lambda}(\mathbf{n})$ makes (27) true if and only if

$$\begin{aligned}&\phi(\mathbf{n}+\mathbf{m}_1)\\ &\quad\in\ [\![(\phi_0(\mathbf{n})+\theta_2-\theta_1+\phi^{(2)}(\mathbf{n}+\mathbf{m}_2)-\phi^{(2)}(\mathbf{n})+\measuredangle f(\mathbf{n}+\mathbf{m}_2)-\measuredangle f(\mathbf{n}+\mathbf{m}_1))\mp\delta\pi]\!]\end{aligned} \tag{28}$$

where $\phi^{(2)}(\mathbf{n})=\measuredangle\lambda^{(2)}(\mathbf{n})$.

Since $\phi(\mathbf{n}+\mathbf{m}_1)$ are independently and uniformly distributed in $[0,2\pi]$, (28) holds for each $\mathbf{n}$ with probability at most δ and hence (27) holds for all $\mathbf{n}$ at once with probability at most $\delta^{\lfloor S/2\rfloor}$.

The union over $\mathbf{m}_1\neq\mathbf{0}$ of these events has probability at most $|\mathcal{N}|\delta^{\lfloor S/2\rfloor}$. Therefore, with probability at least $1-|\mathcal{N}|\delta^{\lfloor S/2\rfloor}$, $\mathbf{m}_1=\mathbf{0}$ and (27) becomes

$$\frac{\lambda(\mathbf{n})}{\tilde{\lambda}(\mathbf{n})}=\exp(i\theta_2-i\theta_1)\frac{\lambda^{(2)}(\mathbf{n}+\mathbf{m}_2)f(\mathbf{n}+\mathbf{m}_2)}{\lambda^{(2)}(\mathbf{n})f(\mathbf{n})}.$$

Moreover, if $\lambda_2 f$ satisfies the non-degeneracy condition, then $\mathbf{m}_2=\mathbf{0}$, $\tilde{f}(\mathbf{n})=\exp(i\theta_2)f(\mathbf{n})$, $\forall\mathbf{n}$, and $\tilde{\lambda}(\mathbf{n})=\exp(i\theta_1-i\theta_2)\lambda(\mathbf{n})$, if $f(\mathbf{n})\neq 0$, with probability at least $1-|\mathcal{N}|\delta^{\lfloor S/2\rfloor}$.

In the case of (23)&(26), we have

$$\exp(i\theta_1)\lambda(\mathbf{n}+\mathbf{m}_1)\lambda^{(2)}(\mathbf{n})f(\mathbf{n}+\mathbf{m}_1)=\exp(i\theta_2)\tilde{\lambda}(\mathbf{n})\overline{\lambda^{(2)}(\mathbf{m}_2-\mathbf{n})f(\mathbf{m}_2-\mathbf{n})}. \tag{29}$$

The same argument applies and $\mathbf{m}_1=0$ with probability at least $1-|\mathcal{N}|\delta^{\lfloor S/2\rfloor}$, and (29) becomes

$$\frac{\lambda(\mathbf{n})}{\tilde{\lambda}(\mathbf{n})}=\exp(i\theta_2-i\theta_1)\frac{\overline{\lambda^{(2)}(\mathbf{m}_2-\mathbf{n})f(\mathbf{m}_2-\mathbf{n})}}{\lambda^{(2)}(\mathbf{n})f(\mathbf{n})},$$

which violates the non-degeneracy condition. In other words, (23)&(26) holds with probability at most $|\mathcal{N}|\delta^{\lfloor S/2 \rfloor}$.

Similar conclusions follow in the case of (24)&(25) and (24)&(26). □

C. Proof of Lemma 1

Proof. Since the operator $\mathcal{T}$ enforces the measured Fourier intensities Y

$$\begin{aligned} r(f_{k+1}, \lambda_{k+1}) &= \| \, |\Phi\Lambda_{k+1} f_{k+1}| - Y \, \| \\ &\leq \|\Phi\Lambda_{k+1} f_{k+1} - \mathcal{T}\Phi\Lambda_k f_{k+1}\| = \|\Lambda_{k+1} f_{k+1} - \Phi^{-1}\mathcal{T}\Phi\Lambda_k f_{k+1}\| \end{aligned}$$

by the unitarity of the Fourier transform. By splitting the summation and using the definition (8), the rightmost term becomes

$$\left(\sum_{f_{k+1}(\mathbf{n})\neq 0} |f_{k+1}|^2(\mathbf{n}) \Big| \lambda_{k+1}(\mathbf{n}) - \lambda'_k(\mathbf{n}) \Big|^2 + \sum_{f_{k+1}(\mathbf{n})=0} \Big| \Phi^{-1}\mathcal{T}\Phi\Lambda_k f_{k+1}(\mathbf{n}) \Big|^2 \right)^{1/2} . \quad (30)$$

Now since $\lambda_{k+1}(\mathbf{n}) = \mathcal{Q}_{\mathrm{m}}\lambda'_k(\mathbf{n})$ is a pixel-wise projection of $\lambda'_k(\mathbf{n})$, $|\lambda_{k+1}(\mathbf{n}) - \lambda'_k(\mathbf{n})| \leq |\lambda_k(\mathbf{n}) - \lambda'_k(\mathbf{n})|$ and hence (30) is less than or equal to

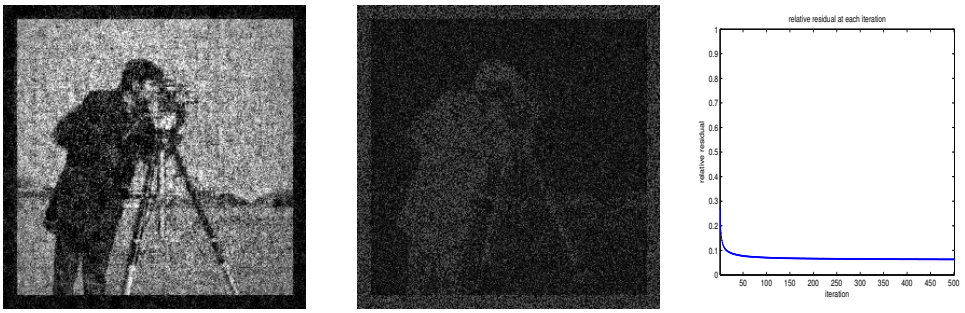
$$\begin{aligned} &\leq \left(\sum_{f_{k+1}(\mathbf{n})\neq 0} |f_{k+1}|^2(\mathbf{n}) \Big| \lambda_k(\mathbf{n}) - \lambda'_k(\mathbf{n}) \Big|^2 + \sum_{f_{k+1}(\mathbf{n})=0} \Big| \Phi^{-1}\mathcal{T}\Phi\Lambda_k f_{k+1}(\mathbf{n}) \Big|^2 \right)^{1/2} \\ &= \|\Lambda_k f_{k+1} - \Phi^{-1}\mathcal{T}\Phi\Lambda_k f_{k+1}\| = \|\Phi\Lambda_k f_{k+1} - \mathcal{T}\Phi\Lambda_k f_{k+1}\| \\ &= \| \, |\Phi\Lambda_k f_{k+1}| - Y \, \| = r(f_{k+1}, \lambda_k) \end{aligned}$$

which is the desired result. □

References

1. J. Miao, P. Charalambous, J. Kirz and D. Sayre, "Extending the methodology of X-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens," Nature **400**, 342–344 (1999).
2. P. Thibault, M. Dierolf, A. Menzel, O. Bunk, C. David, F. Pfeiffer, "High-resolution scanning X-ray diffraction microscopy", Science **321**, 379-382 (2008).
3. M. Dierolf, A. Menzel, P. Thibault, P. Schneider, C. M. Kewish, R. Wepf, O. Bunk, and F. Pfeiffer, " Ptychographic x-ray computed tomography at the nanoscale," Nature **467**, 436-439 (2010).
4. R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of the phase from image and diffraction plane pictures," Optik **35**, 237-246, 1972.

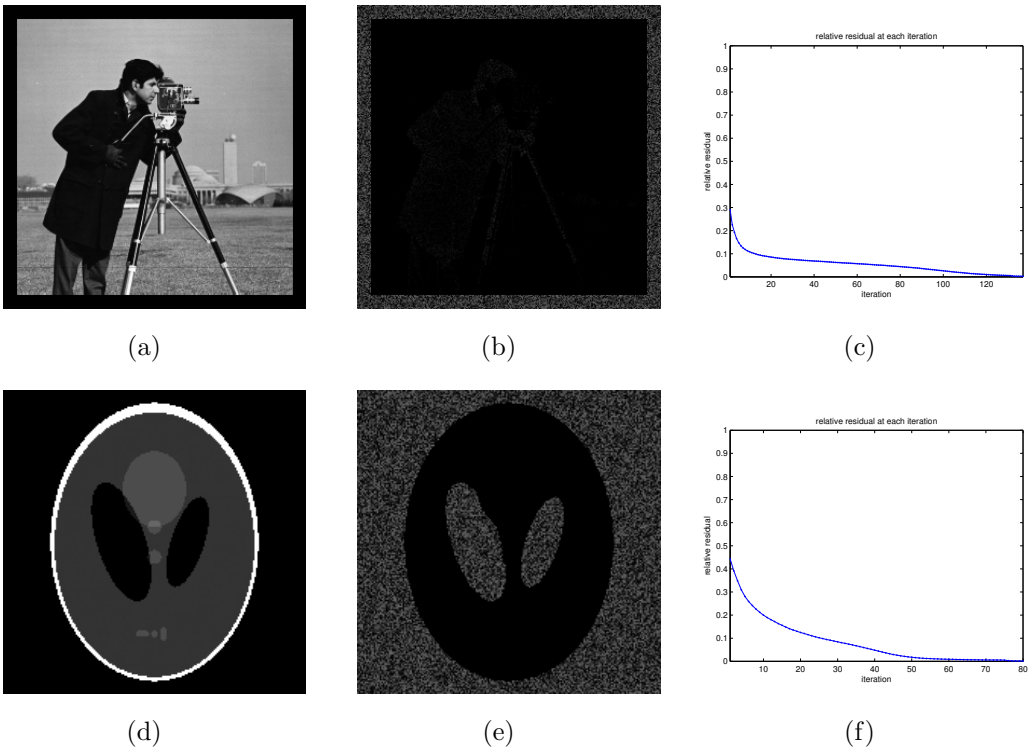
5. J. R. Fienup, "Phase retrieval algorithms: a comparison," Appl. Opt. **21**, 2758-2769 (1982).
6. S. Marchesini, "A unified evaluation of iterative projection algorithms for phase retrieval," Rev. Sci. Instr. **78**, 011301 (2007).
7. A. Fannjiang, "Absolute uniqueness of phase retrieval with random illumination," Inverse Problems **28**, 075008(2012).
8. A. Fannjiang and W. Liao, "Phase retrieval with random phase illumination," J. Opt. Soc. A, **29**, 1847-1859(2012).
9. M.H. Hayes and J.H. McClellan. "Reducible Polynomials in More Than One Variable." Proc. IEEE **70**(2):197 198, (1982).
10. M. Hayes, "The reconstruction of a multidimensional sequence from the phase or magnitude of its Fourier transform," IEEE Trans. Acoust. Speech Sign. Proc. **30** 140- 154 (1982).
11. A. M. Maiden, M. J. Humphry, F. Zhang and J. M. Rodenburg, "Superresolution imaging via ptychography," J. Opt. Soc. Am. A **28**, 604-612 (2011).
12. P.-L. Lions and B. Mercier,"Splitting algorithms for the sum of two nonlinear operators," SIAM J. Num. Anal. **16**, 964-979 (1979).
13. 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. Am. A **19**, 13341-1345 (2002).
14. J. Miao, J. Kirz and D. Sayre, "The oversampling phasing method," Acta Cryst. D **56**, 1312–1315 (2000).



(a) error = 33.75% (b) (c) res = 6.33%

(d) error = 44.74% (e) (f) res = 9.82%

Fig. 4. Recovery of non-negative images by AER with one LRM of $\delta = 0.2$. The middle column shows the absolute phase differences between λ and $\hat{\lambda}$. The right column shows the relative residual at each iteration.



(a) (b) (c)

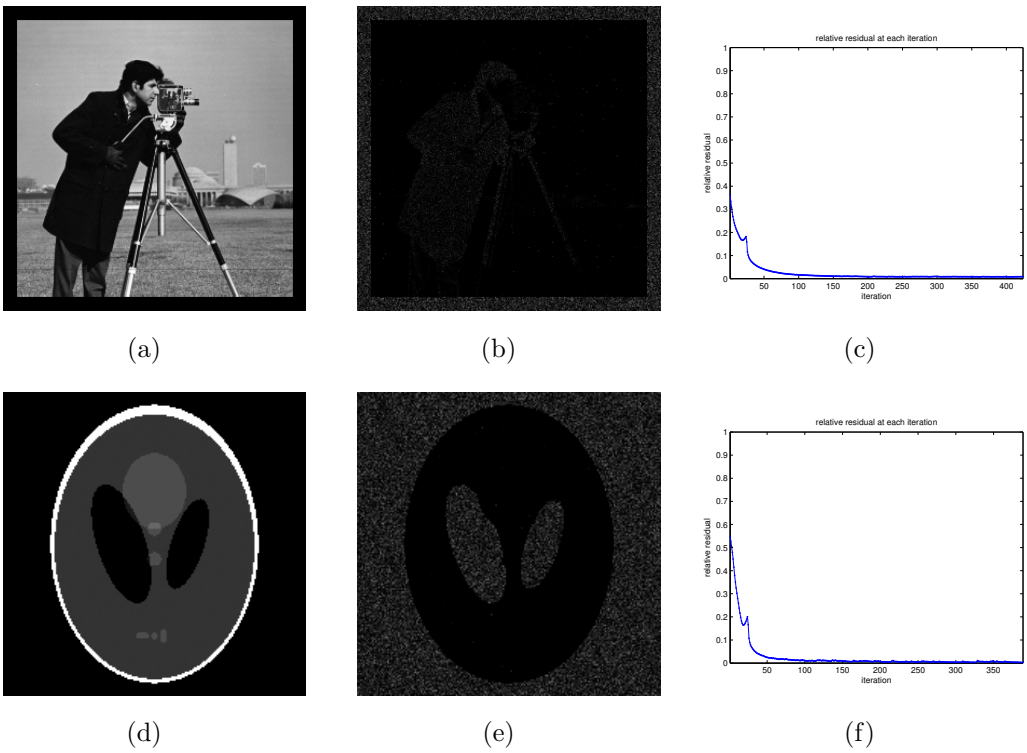
(d) (e) (f)

Fig. 5. Recovery of non-negative images with one LRM of $\delta = 0.2$. (a) the recovered cameraman $\hat{f}$ by 131 DRER + 6 AER steps. $e(\hat{f}) \approx 1.43\%$ and $\rho(\hat{f}, \hat{\lambda}) \approx 0.25\%$. (d) the recovered phantom $\hat{f}$ by 75 DRER + 5 AER steps. $e(\hat{f}) \approx 0.33\%$ and $\rho(\hat{f}, \hat{\lambda}) \approx 0.12\%$. The middle column shows the absolute phase differences between λ and $\hat{\lambda}$. The right column shows the relative residual at each iteration.

(a) (b) (c)

(d) (e) (f)

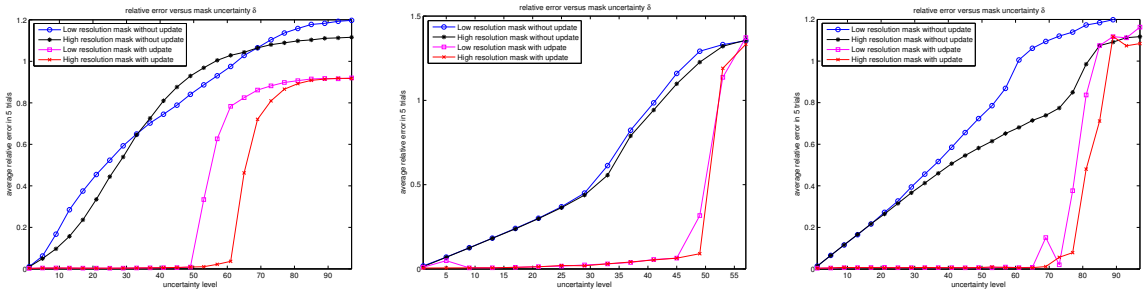
Fig. 6. Recovery of generic complex-valued images with one UM and one LRM of $\delta = 0.2$. (a) absolute values of the recovered cameraman $\hat{f}$ by 400 DRER + 400 AER steps. $e(\hat{f}) \approx 2.39\%$ and $\rho(\hat{f}, \hat{\lambda}) \approx 1.03\%$. (d) absolute values of the recovered phantom $\hat{f}$ by 400 DRER + 400 AER steps. $e(\hat{f}) \approx 1.37\%$ and $\rho(\hat{f}, \hat{\lambda}) \approx 0.65\%$. The middle column shows the absolute phase differences between λ and $\hat{\lambda}$. The right column shows the relative residual at each iteration.



(a) (b) (c)

(d) (e) (f)

Fig. 7. Recovery of complex-valued objects under the sector constraint with one UM and one LRM of $\delta = 0.2$. (a) absolute values of the recovered cameraman $\hat{f}$ by 24 DRER + 400 AER steps. $e(\hat{f}) \approx 1.95\%$ and $\rho(\hat{f}, \hat{\lambda}) \approx 0.94\%$. (d) absolute values of the recovered phantom $\hat{f}$ by 24 DRER + 365 AER steps. $e(\hat{f}) \approx 0.61\%$ and $\rho(\hat{f}, \hat{\lambda}) \approx 0.37\%$. The middle column shows the absolute phase differences between λ and $\hat{\lambda}$. The right column shows the relative residual at each iteration.



(a) (b) (c)

Fig. 8. Averaged relative error $e(\hat{f})$ of 5 independent runs versus the percentage of mask uncertainty for (a) non-negative phantom (b) generic complex phantom and (c) phantom under the sector condition. The stopping rules and mask updating rules are the same as described in the main text for each case with the maximum of $200 + 1000 \cdot \delta$ steps for DRER and AER separately.