



Phase retrieval with a dual recursive scheme

LI SONG AND EDMUND Y. LAM*

Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam, Hong Kong SAR, China

*elam@eee.hku.hk

Abstract: Since optical sensors cannot detect the phase information of the light wave, recovering the missing phase from the intensity measurements, called phase retrieval (PR), is a natural and important problem in many imaging applications. In this paper, we propose a learning-based recursive dual alternating direction method of multipliers, called RD-ADMM, for phase retrieval with a dual and recursive scheme. This method tackles the PR problem by solving the primal and dual problems separately. We design a dual structure to take advantage of the information embedded in the dual problem that can help with solving the PR problem, and we show that it is feasible to use the same operator for both the primal and dual problems for regularization. To demonstrate the efficiency of this scheme, we propose a learning-based coded holographic coherent diffractive imaging system to generate the reference pattern automatically according to the intensity information of the latent complex-valued wavefront. Experiments on different kinds of images with a high noise level indicate that our method is effective and robust, and can provide higher-quality results than other commonly-used PR methods for this setup.

© 2023 Optica Publishing Group under the terms of the [Optica Open Access Publishing Agreement](#)

1. Introduction

In many imaging applications, such as crystallography [1], optical imaging [2], microscopy [3] and digital holography [4,5], phase information reveals important properties of an object. For example, in biological imaging, if the sample is transparent, the intensity of the transmitted light may change slightly, but its phase can vary significantly, giving a better representation of this sample. However, conventional photodetectors are only sensitive to the number of photons instead of phase information. The missing phase of the measured image plays a critical role in the whole image reconstruction process. As shown in [6], if only the Fourier phase of an image is changed, the reconstructed image after inverse Fourier transform can be quite different. This is a recurrent problem in optical imaging. We need to estimate the phase accurately in order to achieve a good image reconstruction. This problem is called phase retrieval (PR), since we try to recover the phase information from intensity-only measurements.

PR is an inverse imaging problem, which aims at recovering an unknown signal from the intensity measurements. The general phase retrieval problem tries to find a latent image $\mathbf{x}$ from the intensity measurement $\mathbf{y}$, given by

$$\mathbf{y} = |\mathbf{A}\mathbf{x}| + \boldsymbol{\epsilon}, \quad (1)$$

where $\mathbf{y} \in \mathbb{R}^m$ is the vectorized intensity-only measurement [7], $|\cdot|$ is the element-wise magnitude operation, $\mathbf{A}$ is the known forward imaging operator, $\boldsymbol{\epsilon}$ is the additive noise, and $\mathbf{x} \in \mathbb{C}^n$ is the vectorized complex-valued latent image. Due to the existence of the magnitude operation, this is typically a non-linear problem. If the phase information of $\mathbf{A}\mathbf{x}$ is obtained, then $\mathbf{x}$ can be calculated by inverting the forward imaging process. PR is also highly under-determined, since if $\tilde{\mathbf{x}}$ is a possible solution, then after any phase shift, it also complies with the measurement constraint. Hence, it is necessary to choose a proper solution from these possibilities according to some prior knowledge.

Depending on the applications, the forward imaging models may often be different. Nevertheless, it is often the case that $\mathbf{A}$ represents the discrete Fourier transform (DFT). According to the

optical diffraction theory [8], the far-field image is given by the Fourier transform of its near-field image with a known phase factor. That makes it possible to recover the phase information from its far-field image with numerical algorithms. This is the basis of coherent diffractive imaging (CDI). In CDI, the oversampled far-field image of an object illuminated by a coherent wave is recorded, and then PR algorithms are used to recover the object.

We propose here an approach to tackle the PR problem that is simultaneously model-based and with data-driven components. The main contributions of this paper are summarized as follows:

- We solve the complex-valued PR optimization problem directly based on a dual alternating direction method of multipliers (D-ADMM) scheme [9]. We design a dual structure to take advantage of the information embedded in the dual problem which can provide help to the primal problem.
- Since it is challenging to design an explicit regularization function to the complex-valued images, we instead use a data-driven approach by means of a U-net. We show with mathematical derivations that it can act as a regularization function and compute its Fenchel conjugate using the same network.
- The proposed recursive dual alternating direction method of multipliers (RD-ADMM) scheme is applicable for various PR problems. To show its effectiveness, we apply it to the holographic coherent diffractive imaging (HCDI) system. Since the reference pattern is undetermined, we use a learning-based approach to find proper reference patterns for different objects according to their intensity information.

The rest of this paper is organized as follows. Section 2 gives a brief review of the PR problem and CDI. Section 3 develops the general mathematical model for PR and shows how it is combined with D-ADMM and the use of neural networks. Section 4 presents the idea and mathematical model for coded holographic coherent diffractive imaging (CHCDI). Numerical experimental results for different kinds of images are also shown in this section. Finally, section 5 draws some conclusions.

2. Related work

2.1. Phase retrieval

PR is a non-convex problem and therefore it is not easy to find a proper solution directly. Alternating projection methods are first used to solve this problem. In 1972, Gerchberg and Saxton presented a technique that attempts to optimize the reconstruction results in the measurement domain and real image domain alternately [10]. An initial random phase is added to the measured image and then inverse transformed to the real image domain. After considering its support, the revised image is then transformed to the measurement domain to update the phase attached to the measurement image. The whole process is repeated until convergence. Fienup made a significant improvement to the update step in the real domain, and the new method is called hybrid input-output algorithm (HIO) [11]. This method is widely adopted in many PR applications. Generally, the initialization step is important for these algorithms. A variety of initializations are required to improve the reconstruction results.

Optimization methods can also be used to handle the PR problem. Since the intensity measurement is related to the rank-one matrix $X = \mathbf{x}\mathbf{x}^H$, a rank minimization problem can be solved to obtain X and then $\mathbf{x}$ by singular value decomposition (SVD). This is called PhaseLift [12]. A main advantage of this approach is that it can be relaxed as a convex problem. However, since it considers the higher-dimensional optimization problem, it usually needs more computation than alternating projection methods.

Another approach is to solve the non-convex PR problem directly. Wirtinger Flow (WF) [13] adopts the gradient descent method to solve this problem. Recently, alternating projection

methods related to optimization are also used to tackle the PR problem, such as using alternating direction method of multipliers (ADMM) [14] and proximal algorithms [15]. The complex-valued image is separated into the real part and the imaginary part, and then these methods can handle the two sub-problems respectively in the real domain.

With the development of deep learning [16,17], some end-to-end neural networks are trained to solve PR problems, such as convolutional neural network (CNN) [18], U-net [19], etc. These methods need large datasets for training, and often it is not easy to obtain sufficient images. To generalize the application of learning-based methods, neural networks are combined with iterative conventional methods [20,21]. For example, prDeep [22] puts the imaging model and external denoiser in the loss function and a DnCNN network is trained based on the regularization by denoising (RED) [23]. Plug-and-play ADMM (PnP-ADMM) [24] directly combines a trained denoiser with the conventional ADMM iterations, and this idea is also commonly used in other inverse imaging problems [25,26]. These model-based networks are easily trained and available for different kinds of PR problems, or even other inverse imaging problems [27]. They also require fewer training images than end-to-end networks.

2.2. Coherent diffractive imaging

CDI is a lensless imaging system that can infer the phase information from the far-field intensity measurements [28]. A typical CDI system is depicted in [2]. Coherent light is used to illuminate the target object. A photodetector, such as a charge-coupled device (CCD), can capture its far-field intensity pattern. The captured image is the squared intensity of the latent image after an oversampled DFT. The oversampling rate has a great impact on the uniqueness of the final reconstructed image. At least twice the Nyquist frequency is required for CDI, and the sampling rate should be higher when considering the effect of noise [29].

According to Fourier theory, the center region in a CDI diffraction pattern is related to the low-frequency part in the latent image, but the high-frequency part contains rich texture information. In real CDI applications, a beam stop is adopted to remove the overexposed center region and enhance the capture of other regions.

Sometimes, a far-field image is not easy to obtain due to the requirement on distance. Near-field diffraction images can also be used in CDI, in a setup known as Fresnel CDI [30]. This is more complex, since the diffraction image is also related to the actual distance between the object and the imaging plane.

CDI was first used in crystallography [31], since periodic objects would lead to peaks in the diffraction images. With the development of X-ray sources, X-ray free electron laser (XFEL) is now widely-used in CDI, since it requires no additional optical elements for imaging. Miao et al. [32] first extended CDI to non-periodic objects in biological imaging. Recently, some researches focus on the improvement of light sources [33] and imaging setups [34] to provide high-resolution images.

In conventional CDI setup, no prior knowledge of the object plane is available. The supports required by the PR algorithms are usually user-defined. To simplify the PR process and increase its efficiency, masks with known coding patterns are used in CDI [35], such as a random mask [36], pinhole array [37], coded aperture [38], etc. Such setups are regarded as a combination of in-line holography and CDI, since the coded element is set between the object and the image sensor.

Another possibility is to put a known object, regarded as a reference pattern, in the object plane, and then use the conventional CDI for imaging. A typical setup is presented in [39]. This is known as holographic CDI (HCDI), as the object light interferes with another known reference light. The known reference can be regarded as a support for the PR algorithms [40]. Note that the imaging setup for HCDI mentioned above is different from [41], which is actually a fusion of a hologram and a CDI image.

3. Phase retrieval via D-ADMM with deep learning

3.1. Phase retrieval problem formulation

As shown in Eq. (1), solving the PR problem means finding a latent image $\mathbf{x}$ whose magnitude, after a known imaging process A , produces the measurement $\mathbf{y}$. If the phase information of $\mathbf{y}$, denoted as ω , is known, then $\mathbf{x}$ is obtained after an inverse step. The relationship between $\mathbf{x}$ and ω is

$$A\mathbf{x} = \mathbf{y} \circ e^{j\omega} + \mathbf{e}, \quad (2)$$

where j is the imaginary number, ω is the unknown phase, $\circ$ is the element-wise product and $\mathbf{e}$ is the additive noise. Due to noise, a real-valued regularization function $\mathcal{R}(\cdot)$ is required to find a proper solution that agrees with the prior knowledge [42]. Generally, the regularization function can be either a manually-designed function or replaced by a neural network. The PR problem with regularization can be stated as

$$\begin{aligned} &\text{minimize} \quad \mathcal{R}(\mathbf{x}) + \frac{\gamma}{2} \|\mathbf{e}\|_2^2 \\ &\text{subject to} \quad A\mathbf{x} = \mathbf{y} \circ e^{j\omega} + \mathbf{e}, \end{aligned} \quad (3)$$

where γ is a scale parameter. This is not a linearly constrained optimization problem due to the existence of an exponential function. If ω is restricted on the region $[0, 2\pi)$, then ω is unique given $e^{j\omega}$, and the term $e^{j\omega}$ can be regarded as a new variable ϕ . Now the PR problem is a linearly constrained optimization problem

$$\begin{aligned} &\text{minimize} \quad \mathcal{R}(\mathbf{x}) + \frac{\gamma}{2} \|\mathbf{e}\|_2^2 + \mathcal{I}(\phi) \\ &\text{subject to} \quad A\mathbf{x} - \mathbf{y} \circ \phi = \mathbf{e}, \end{aligned} \quad (4)$$

where $\mathcal{I}(\phi)$ is the unit-circle indicator function

$$\mathcal{I}(\phi) = \begin{cases} 0 & , \quad |\phi| = \mathbf{1} \\ \infty & , \quad \text{otherwise} \end{cases}, \quad (5)$$

and $\mathbf{1}$ is a column vector with all the elements equal to 1. Since $\mathcal{I}(\phi)$ is a non-convex function, the whole optimization problem is also non-convex.

Following the same idea as in the conventional Lagrange multiplier method [14], we can obtain the unaugmented Lagrangian as [43]

$$L(\mathbf{x}, \phi, \mathbf{e}, \lambda) = \mathcal{R}(\mathbf{x}) + \mathcal{I}(\phi) + \frac{\gamma}{2} \|\mathbf{e}\|_2^2 + \text{Re}[\lambda^H (A\mathbf{x} - \mathbf{y} \circ \phi - \mathbf{e})], \quad (6)$$

where $\text{Re}(\cdot)$ is the real part of a complex number, λ is the Lagrange multiplier and $(\cdot)^H$ is the Hermitian operator. Now the problem is a min-max problem [44]

$$\{\lambda^*, \mathbf{x}^*, \phi^*, \mathbf{e}^*\} = \arg \max_{\lambda} \min_{\{\mathbf{x}, \phi, \mathbf{e}\} \in \mathbb{D}} L(\mathbf{x}, \phi, \mathbf{e}, \lambda), \quad (7)$$

where $\mathbb{D}$ is the domain of $\{\mathbf{x}, \phi, \mathbf{e}\}$. The general problem we consider in this paper is such a generic saddle-point problem which has a typical corresponding dual problem [45].

3.2. Dual ADMM solution

In the dual theory, λ is also known as dual variable. Most algorithms deal with Eq. (6) directly and the dual variable λ is updated by a simple dual ascent step based on the values of the primal

variables. In this paper, following the idea of dual ADMM [9], we try to find the value of λ based on the dual problem, which can give better results than the simple gradient ascent step.

To tackle the min-max problem above, we divide it into a minimization step and a maximization step, and then solve them alternately. The former is called the primal problem, because $\mathbf{x}$ is the primal variable we are most concerned with. The latter is called the dual problem, because it is only related to the dual variable λ . These two problems can be solved respectively by ADMM. Taking advantage of the duality between these two problems, the approach, known as dual ADMM, can be used to solve the whole min-max problem [9].

First we consider the primal problem. In this sub-problem, if we consider the form in Eq. (6) directly, then there are only linear terms considering the variable ϕ and it is not easy to find an optimal value of ϕ in each step since the gradient of ϕ is not related to ϕ , so we add a quadratic term $\|A\mathbf{x} - \mathbf{y} \circ \phi - \mathbf{e}\|_2^2$ here. According to the property of the Lagrangian method, if an optimal solution is found, then this quadratic term is also 0. Hence it will not change the solution of this primal problem. To separate the unknown $\mathcal{R}(\mathbf{x})$ from the other known operators, an auxiliary variable $\mathbf{u}$ is used. The primal problem is

$$\begin{aligned} \text{minimize} \quad & \mathcal{R}(\mathbf{u}) + \mathcal{I}(\phi) + \frac{\gamma}{2} \|\mathbf{e}\|_2^2 + \text{Re}[\lambda^H (A\mathbf{x} - \mathbf{y} \circ \phi - \mathbf{e})] + \frac{\rho}{2} \|A\mathbf{x} - \mathbf{y} \circ \phi - \mathbf{e}\|_2^2 \\ \text{subject to} \quad & \mathbf{u} - \mathbf{x} = 0, \end{aligned} \quad (8)$$

where ρ is the penalty parameter.

For convenience, we use a regularization function $\mathcal{R}(\cdot)$ where $\nabla \mathcal{R}(\cdot)$ exists and ∇ is the Wirtinger derivative [46]. Accordingly, we can obtain the following iterations

$$\mathbf{u}^{(k+1)} = \mathbf{x}^{(k)} - \frac{1}{\rho_1} \nabla \mathcal{R}(\mathbf{u}^{(k+1)}) - \boldsymbol{\mu}_1^{(k)} \quad (9)$$

$$\mathbf{x}^{(k+1)} = (\rho_1 I + \rho A^H A)^{-1} [A^H (\rho \mathbf{y} \circ \phi^{(k)} + \rho \mathbf{e}^{(k)} - \lambda_1) + \rho_1 (\mathbf{u}^{(k+1)} + \boldsymbol{\mu}_1^{(k)})] \quad (10)$$

$$\phi^{(k+1)} = \mathcal{H}_y \left(\frac{1}{\rho} \lambda_1 + A\mathbf{x}^{(k+1)} - \mathbf{e}^{(k)} \right) \quad (11)$$

$$\mathbf{e}^{(k+1)} = \frac{1}{\gamma + \rho} [\lambda_1 + \rho (A\mathbf{x}^{(k+1)} - \mathbf{y} \circ \phi^{(k+1)})] \quad (12)$$

$$\boldsymbol{\mu}_1^{(k+1)} = \boldsymbol{\mu}_1^{(k)} + \mathbf{u}^{(k+1)} - \mathbf{x}^{(k+1)}, \quad (13)$$

where I is the identity matrix, $\lambda_1 = \frac{1}{2} \lambda$, $\boldsymbol{\mu}_1$ is the scaled Lagrange multiplier, ρ_1 is the penalty parameter, and $\mathcal{H}_y(\cdot)$ is an element-wise normalization operator defined as

$$\mathcal{H}_y(s)_{(i)} = \begin{cases} \frac{s_{(i)}/y_{(i)}}{|s_{(i)}/y_{(i)}|}, & y_{(i)} \neq 0, s_{(i)} \neq 0 \\ 1, & \text{otherwise} \end{cases}, \quad (14)$$

and i refers to the i^{th} element.

Now we consider the dual problem. The Fenchel conjugate of a real-valued function with complex variable $\mathcal{R}(\cdot)$ is defined as

$$\mathcal{R}^*(\mathbf{r}) = \sup_{\mathbf{q}} G(\mathbf{r}, \mathbf{q}), \quad (15)$$

where $G(\mathbf{r}, \mathbf{q})$ is

$$G(\mathbf{r}, \mathbf{q}) = \mathbf{r}^H \mathbf{q} + \mathbf{q}^H \mathbf{r} - \mathcal{R}(\mathbf{q}). \quad (16)$$

The regularization function must be a real-valued function to evaluate the reconstruction results. We first deal with the indicator function $\mathcal{I}(\phi)$. Its Fenchel conjugate is

$$\begin{aligned}\mathcal{I}^*(\mathbf{w}) &= \sup_{|\phi|=1} \{\operatorname{Re}\{2\mathbf{w}^H \phi\}\} \\ &= \sup_{\theta} \{\operatorname{Re}\{2\mathbf{w}^H e^{j\theta}\}\} = 2\|\mathbf{w}\|_1.\end{aligned}\quad (17)$$

According to the definition of the conjugate function, the dual problem can be formulated as

$$\begin{aligned}\text{minimize} \quad & \mathcal{R}^*(\mathbf{z}_1) + 2\|\mathbf{z}_2\|_1 + \frac{1}{2\gamma}\|\lambda_1\|_2^2 \\ \text{subject to} \quad & M\mathbf{z}_1 + \lambda_1 = 0, \\ & \mathbf{z}_2 - \mathbf{y} \circ \lambda_1 = 0,\end{aligned}\quad (18)$$

where $\mathbf{z}_1$ and $\mathbf{z}_2$ are auxiliary variables, $M = (AA^H)^{-1}A$ since AA^H is invertible under mild conditions [47].

Following the same method and requirement that $\nabla \mathcal{R}^*(\cdot)$ exists as in the primal problem, we have

$$M^H M \mathbf{z}_1^{(k+1)} = -M^H (\lambda_1^{(k)} + \mu_2^{(k)}) - \frac{1}{\rho_2} \nabla \mathcal{R}^*(\mathbf{z}_1^{(k+1)}) \quad (19)$$

$$\mathbf{z}_2^{(k+1)} = \mathcal{S}_{\frac{2}{\rho_3}}(\mathbf{y} \circ \lambda_1^{(k)} - \mu_3^{(k)}) \quad (20)$$

$$\lambda_1^{(k+1)} = \mathbf{1} \oslash (\rho_3 \mathbf{y} \circ \mathbf{y} + \rho_2 \mathbf{1} + \mathbf{1}/\gamma) \oslash [\rho_3 \mathbf{y} \circ (\mathbf{z}_2^{(k+1)} + \mu_3^{(k)}) - \rho_2 (M \mathbf{z}_1^{(k+1)} + \mu_2^{(k)})] \quad (21)$$

$$\mu_2^{(k+1)} = \mu_2^{(k)} + M \mathbf{z}_1^{(k+1)} + \lambda_1^{(k+1)}, \quad (22)$$

$$\mu_3^{(k+1)} = \mu_3^{(k)} + \mathbf{z}_2^{(k+1)} - \mathbf{y} \circ \lambda_1^{(k+1)}, \quad (23)$$

where $\oslash$ is the element-wise division, $\mathcal{S}$ is the soft-thresholding operator [48], μ_2, μ_3 are the scaled Lagrange multipliers and ρ_2, ρ_3 are the penalty parameters. We refer to it as the D-ADMM scheme for phase retrieval.

3.3. Learning-based approach

In this section, we develop a learning-based approach to compute the undetermined part in the D-ADMM scheme.

Although the regularization function is constrained to satisfy the condition that $\nabla \mathcal{R}(\cdot)$ exists, there is room for designing such functions. Learning-based methods can help to arrive at a solution that is potentially better than a handcrafted one [26].

To take advantage of the relationship between $\mathcal{R}(\cdot)$ and $\mathcal{R}^*(\cdot)$, we first find the connection between the $\mathbf{u}$ -update step and the $\mathbf{z}$ -update step. According to the definition of the conjugate function, the necessary condition of the optimization problem in Eq. (15) is

$$\mathbf{r} = \nabla \mathcal{R}(\mathbf{q}). \quad (24)$$

Now the conjugate function $\mathcal{R}^*(\mathbf{r})$ can also be regarded as a function of the variable $\mathbf{q}$ with reference to Eq. (24) as $\mathcal{R}^*(\nabla \mathcal{R}(\mathbf{q}))$. Using the chain rule of differentiation, we have

$$\frac{\partial \mathcal{R}^*}{\partial \mathbf{q}} = \left[\frac{\partial \mathbf{r}}{\partial \mathbf{q}} \right]^H \frac{\partial G}{\partial \mathbf{r}} + \frac{\partial G}{\partial \mathbf{q}} \quad (25)$$

$$= \left[\frac{\partial \mathbf{r}}{\partial \mathbf{q}} \right]^H \mathbf{q} + \mathbf{r} - \nabla \mathcal{R}(\mathbf{q}) = \left[\frac{\partial \mathbf{r}}{\partial \mathbf{q}} \right]^H \mathbf{q} \quad (26)$$

$$\nabla \mathcal{R}^*(\mathbf{r}) = \left(\frac{\partial \mathbf{q}}{\partial \mathbf{r}} \right)^H \frac{\partial \mathcal{R}^*}{\partial \mathbf{q}} = \mathbf{q}, \quad (27)$$

then

$$\mathbf{r} = \nabla \mathcal{R}(\nabla \mathcal{R}^*(\mathbf{r})). \quad (28)$$

Substituting Eq. (28) into the $\mathbf{z}_1$ -update step, we have

$$\mathbf{s} + \rho_2 \mathbf{M}^H \mathbf{M} \nabla \mathcal{R}(\mathbf{s}) = -\rho_2 \mathbf{M}^H (\boldsymbol{\lambda}_1^{(k)} + \boldsymbol{\mu}_2^{(k)}), \quad (29)$$

where $\mathbf{s} = -\rho_2 (\mathbf{M}^H \mathbf{M} \mathbf{z}_1^{(k+1)} + \mathbf{M}^H \boldsymbol{\lambda}_1^{(k)} + \mathbf{M}^H \boldsymbol{\mu}_2^{(k)})$.

Now we use a neural network $\mathbf{x}_n = f(\mathbf{y}_n, \mathbf{S}\mathbf{1})$ to find the solution of such a problem

$$\mathbf{S} \nabla \mathcal{R}(\mathbf{x}_n) + \mathbf{x}_n - \mathbf{y}_n = 0, \quad (30)$$

where $\mathbf{S}$ is a matrix with a compatible dimension. If f is determined, then the regularization function is defined indirectly by Eq. (30). The left side of Eq. (30) is actually the derivative of a proximal problem

$$\nabla_{\mathbf{x}_n} [\mathbf{S} \nabla \mathcal{R}(\mathbf{x}_n) + \frac{1}{2} \|\mathbf{x}_n - \mathbf{y}_n\|_2^2] = \mathbf{S} \nabla \mathcal{R}(\mathbf{x}_n) + \mathbf{x}_n - \mathbf{y}_n \quad (31)$$

and the neural network $\mathbf{x}_n = f(\mathbf{y}_n, \mathbf{S}\mathbf{1}) = \text{prox}_{\mathbf{S} \nabla \mathcal{R}}(\mathbf{y}_n)$ is actually a proximal operator [49].

Comparing Eq. (29) and Eq. (30), we can find $\mathbf{x}_n = \mathbf{s}$, $\mathbf{S} = \rho_2 \mathbf{M}^H \mathbf{M}$ and $\mathbf{y}_n = -\rho_2 \mathbf{M}^H (\boldsymbol{\lambda}_1^{(k)} + \boldsymbol{\mu}_2^{(k)})$. Similarly, Eq. (9) can also be converted to Eq. (30) with $\mathbf{x}_n = \mathbf{s}$, $\mathbf{S} = 1/\rho_1 \mathbf{I}$ and $\mathbf{y}_n = \mathbf{x}^{(k)} - \boldsymbol{\mu}_1^{(k)}$.

Then the two update steps are

$$\mathbf{u}^{(k+1)} = f\left(\mathbf{x}^{(k)} - \boldsymbol{\mu}_1^{(k)}, \frac{1}{\rho_1} \mathbf{I} \mathbf{1}\right) \quad (32)$$

$$\mathbf{z}_1^{(k+1)} = (-\rho_2 \mathbf{M}^H \mathbf{M})^{-1} [f(-\rho_2 \mathbf{M}^H (\boldsymbol{\lambda}_1^{(k)} + \boldsymbol{\mu}_2^{(k)}), \rho_2 \mathbf{M}^H \mathbf{M} \mathbf{1}) + \rho_2 \mathbf{M}^H (\boldsymbol{\lambda}_1^{(k)} + \boldsymbol{\mu}_2^{(k)})]. \quad (33)$$

The input of the neural network is separated into two channels according to its real and imaginary parts.

Finally, the whole algorithm is summarized in Algorithm 1. A more intuitive description is shown in Fig. 1. We call the proposed method RD-ADMM, due to the recursive iterative structure. Since the input and output of the network have the same dimension, we refer to U-net [50] as the network structure, which is the same as the unrolled network in [51]. The complete network structure is presented in Supplement 1. Compared with the plug-and-play ADMM [52], RD-ADMM relies on end-to-end training with shared weights across iterations. All the undetermined parameters, including networks (as well as the reference generation network) and the parameters in the optimization steps, are trainable in the training process based on the differences between the estimated image and the ground truth, instead of plugging in a trained denoiser directly. All the weights are trained with the known forward model, which is more suitable to the problem at hand than the pre-trained denoisers.

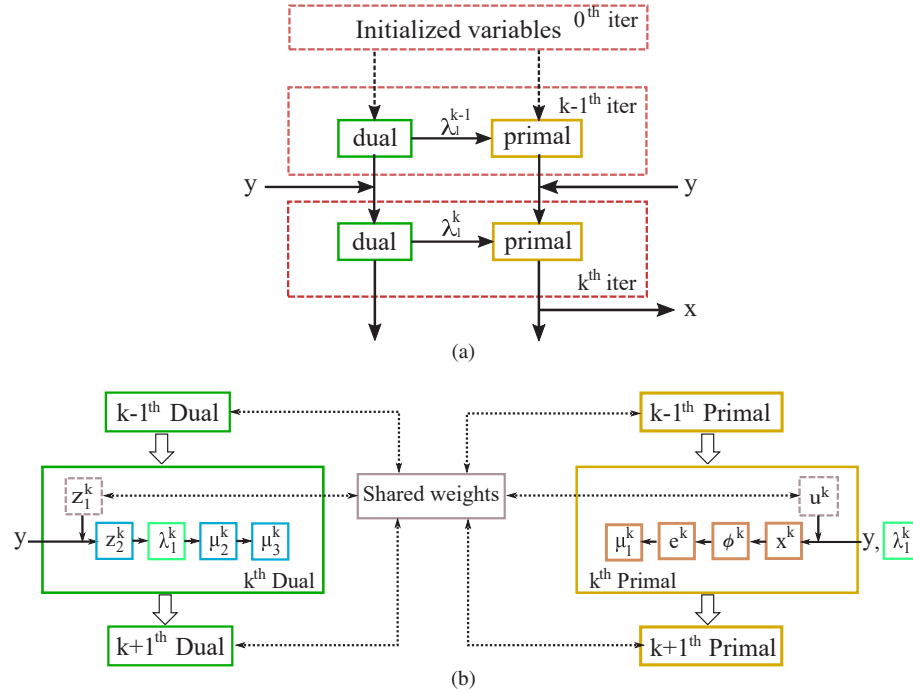


Fig. 1. Structure of the RD-ADMM scheme. (a) The whole recursive structure, the variables $u^{(0)}, x^{(0)}, e^{(0)}, z_1^{(0)}, z_2^{(0)}, \lambda_1^{(0)}, \mu_1^{(0)}, \mu_2^{(0)}, \mu_3^{(0)}$ are set as zero vectors, and $\phi^{(0)}$ is set as 1. (b) Structures of the dual and primal iterations. The network structure is related to z_1 and u , and its parameters remain the same across all the iterations. The value of $\lambda_1^{(k)}$ as the input of the k^{th} Primal comes from the result of the k^{th} Dual on the left.

Algorithm 1. Recursive neural network structure related to D-ADMM for phase retrieval (RD-ADMM)

Input: y : measurement image; $\rho, \rho_1, \rho_2, \rho_3$: penalty parameters; γ : scale parameter, $f(\cdot)$: U-net; A : system matrix for the measurement process; $iter$: number of iterations.

Output: x : reconstructed complex-valued image.

- 1: **for** $k = 1 : iter$ **do**
- 2: **Dual problem solution:**
- 3: $z_1^{(k+1)} = (-\rho_2 M^H M)^{-1} [f(-\rho_2 M^H (\lambda_1^{(k)} + \mu_2^{(k)}), \rho_2 M^H M \mathbf{1}) + \rho_2 M^H (\lambda_1^{(k)} + \mu_2^{(k)})],$
- 4: $z_2^{(k+1)} = \mathcal{S}_{\frac{\rho_2}{\rho_3}} (y \circ \lambda_1^{(k)} - \mu_3^{(k)}),$
- 5: $\lambda_1^{(k+1)} = \mathbf{1} \oslash (\rho_3 y \oslash y + \rho_2 \mathbf{1} + 1/\gamma) \oslash [\rho_3 y \oslash (z_2^{(k+1)} + \mu_3^{(k)}) - \rho_2 (M z_1^{(k+1)} + \mu_2^{(k)})],$
- 6: $\mu_2^{(k+1)} = \mu_2^{(k)} + M z_1^{(k+1)} + \lambda_1^{(k+1)},$
- 7: $\mu_3^{(k+1)} = \mu_3^{(k)} + z_2^{(k+1)} - y \oslash \lambda_1^{(k+1)},$
- 8: **Primal problem solution:**
- 9: $u^{(k+1)} = f(x^{(k)} - \mu_1^{(k)}, \frac{1}{\rho_1} \mathbf{1}),$
- 10: $x^{(k+1)} = (\rho_1 I + \rho A^H A)^{-1} [A^H (\rho y \oslash \phi^{(k)} + \rho e^{(k)} - \lambda_1^{(k+1)}) + \rho_1 (u^{(k+1)} + \mu_1^{(k)})],$
- 11: $\phi^{(k+1)} = \mathcal{H}_y (\frac{1}{\rho} \lambda_1^{(k+1)} + A x^{(k+1)} - e^{(k)}),$
- 12: $e^{(k+1)} = \frac{1}{\gamma + \rho} [\lambda_1 + \rho (A x^{(k+1)} - y \oslash \phi^{(k+1)})],$
- 13: $\mu_1^{(k+1)} = \mu_1^{(k)} + u^{(k+1)} - x^{(k+1)}.$
- 14: **end for**

4. Numerical experiments

4.1. Coded holographic coherent diffractive imaging

Up to this point, the RD-ADMM scheme is applicable to different kinds of PR problems. To show the effectiveness of this method, we refer to the CDI setup. In conventional CDI, only the location of the target object is used as the constraint in the object plane. In this paper, we focus on a modification called HCDI, where a known reference pattern is used as a hard constraint in the phase retrieval algorithms.

Nevertheless, how the reference pattern is selected still remains an open problem. In this paper, we refer to the learning-based approach to find an optimal reference pattern. We assume that the reference pattern is intensity-only and has the same size as the latent image. Since the intensity of light is easier to be detected and modulated than the phase, a trained network is expected to give a reference pattern according to the noisy intensity information of the latent image. The network we use here is also a U-net [50] with two downsampling steps, which is shown in [Supplement 1](#). The whole imaging system is denoted as coded HCDI (CHCDI) [53].

4.2. Mathematical model and datasets

Following the CDI setup, the system matrix A is specified as an oversampled DFT matrix F , leading to the conventional Fourier phase retrieval problem, then the matrix M in Eq. (18) makes $M\mathbf{v}$ the oversampled DFT of $\mathbf{v}$ and $M^H\mathbf{v}$ the downsampled DFT of $\mathbf{v}$.

Training images are generated from the PatchCamelyon benchmark [54]. 3000 images are randomly selected and resized to 128×128 as the input images. The corresponding phase images are randomly generated. These two parts are combined together as the latent image $\mathbf{x}$. The reference pattern is provided by a U-net with reference to the intensity information of $\mathbf{x}$. A $4\times$ oversampled DFT is used to generate the intensity image obtained by the detector and the additional Gaussian noise with a 10% noise level is added to the generated images.

The test set is generated using another 1000 clear images from the PatchCamelyon benchmark. It contains 4 kinds of ground truth images: intensity-only image, intensity image with random phase, phase-only image, and phase image with random intensity. The inputs are the simulated diffraction pattern and the reference mask.

The loss function is designed as

$$\mathcal{L}(\mathbf{x}, \mathbf{x}_0) = \|\text{Re}(\mathbf{x}) - \text{Re}(\mathbf{x}_0)\|_1 + \|\text{Im}(\mathbf{x}) - \text{Im}(\mathbf{x}_0)\|_1, \quad (34)$$

where $\text{Im}(\cdot)$ is the imaginary part of a complex number and $\mathbf{x}_0$ is the ground truth image.

The number of iterations is set as 10 and the initial learning rate is 0.0001. The optimizer is Adam [55] and 100 epochs are required. All the experiments are conducted on a computer with 2.2 GHz Intel Xeon 4210 CPU and NVIDIA Tesla V100 (32Gb memory).

4.3. Different kinds of images

In a complex-valued image, meaningful information may be embedded in the intensity part or the phase part, or both. In this section, we consider four kinds of images according to their intensity and phase features.

4.3.1. Intensity-only image

An intensity-only image means the intensity part contains all useful information and the phase is always zero. Since the phase information of the latent image is unnecessary, these images can be captured by conventional cameras directly.

If the latent image is intensity-only then it can be expressed as

$$\mathbf{x}_0 = \mathbf{x}_{\text{clear}} \circ e^{j\mathbf{0}} = \mathbf{x}_{\text{clear}}, \quad (35)$$

where $\mathbf{0}$ is a column vector with all the elements equal to 0 and $\mathbf{x}_{\text{clear}}$ is the clear real-valued image in the original dataset.

4.3.2. Intensity image with random phase

A more general situation is an intensity image with a noisy phase. The latent image can be expressed as

$$\mathbf{x}_0 = \mathbf{x}_{\text{clear}} \circ e^{j\boldsymbol{\eta}}, \quad (36)$$

where $\boldsymbol{\eta}$ is a noisy column vector. Due to the noisy phase image, the real and imaginary images are also noisy. It is not easy to solve this kind of PR problem by separating it into the real and imaginary sub-problems, since the real and imaginary images do not comply with commonly-used image priors.

4.3.3. Phase-only image

Now we consider the phase-only images, which are different from those in the training process. The information is embedded in the phase and the intensity is a constant. The latent image is

$$\mathbf{x}_0 = c \cdot e^{j\mathbf{x}_{\text{clear}}}, \quad (37)$$

where c is a constant. Since the intensity image contains no information, the reference-generation network has the same output for all these images.

4.3.4. Phase image with random intensity

As for the phase image with random intensity, the latent image can be described as

$$\mathbf{x}_0 = \boldsymbol{\eta} \circ e^{j\mathbf{x}_{\text{clear}}}. \quad (38)$$

In this case, the mask is generated according to the random intensity information only, and no information related to the clear image is available for the mask generation network.

To show the performance of RD-ADMM on the whole dataset, we calculate the average peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) [56] of all the reconstructed images in the testing dataset. When calculating these metrics, the real and imaginary parts of the complex numbers are regarded as two channels. Results are shown in Table 1. For all kinds of images, our method can achieve phase retrieval results with good quantitative evaluations.

Table 1. PSNR and SSIM of the RD-ADMM PR results for 4× CHCDI

Types	PSNR (dB)	SSIM
Intensity-only image	36.76	0.949
Intensity image with random phase	33.88	0.969
Phase-only image	34.62	0.819
Phase image with random intensity	35.65	0.936

The error curves of the RD-ADMM method are also shown in Fig. 2. The relative error evaluates the differences of $\mathbf{x}$ between two successive iterations, and data error shows the distance of the retrieved image to the captured image in the measurement domain. As is shown in Fig. 2, both of them decrease gradually, but the data error is finally not close to 0 due to the existence of the additive noise.

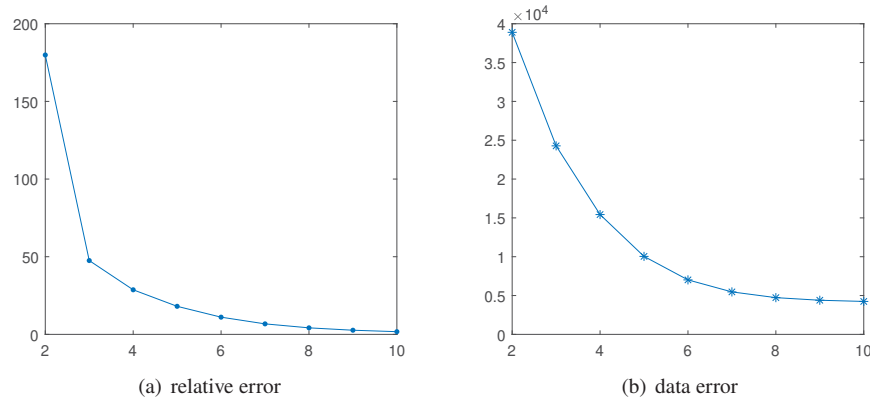


Fig. 2. Error curves of the RD-ADMM method. x -axis: the number of iterations; y -axis: (a) relative error: $\|\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\|_2$, (b) data error: $\|F\mathbf{x}^{(k)} - \mathbf{y}\|_2$.

4.4. Comparison with other methods

This section shows the performances of other PR methods on the CHCDI system for comparison. The oversampling smoothness (OSS) [57] exploits the correlation information of the pixels in the real space and then spatial frequency filters are used in the iterations. HIO is a classical method which can optimize the reconstructed images in the real and Fourier domains [11]. Both BM3D-prGAMP [58] and prDeep [22] adopt trained denoisers as prior knowledge in the iterations. REF-CDI [40] is a deconvolution process based on the correlation between the latent image and the reference pattern. Conventional PnP-ADMM is derived in the real domain. To make it consistent with the CHCDI system, we extend it to the complex domain for comparison.

Since other PR methods mainly consider intensity-only object, for fairness, we only consider the intensity-only image as the latent image. Besides, the reference pattern is randomly selected as a binary mask, i.e. the intensity information is unavailable to all of these methods, but the random mask is known to all of them. For RD-ADMM*, we also use a randomly-generated mask instead of a mask generated by the trained network for fair comparison.

We train an unrolled recursive ADMM network (R-ADMM) based on the original ADMM iterations as well. R-ADMM has a mask generation network and a regularization network with the same network structures as RD-ADMM. Both R-ADMM and RD-ADMM are trained with the dataset for intensity images with random phase.

The reconstruction results of the $4\times$ CHCDI system for intensity-only object are shown in Fig. 3. It shows that OSS, BM3D-prGAMP, HIO and prDeep algorithms fail to give proper reconstructed images with a random reference pattern. REF-CDI method is sensitive to high-level noise. PnP-ADMM needs 30 iterations to provide a reasonable image.

The PSNR and SSIM of these reconstructed results, as well as their processing time for each diffraction pattern, are shown in Table 2. Among these methods, RD-ADMM achieves the best performance. Comparison between RD-ADMM and R-ADMM also shows the advantage of the dual structure. Besides, we can see that the results of RD-ADMM have a much higher PSNR and SSIM than RD-ADMM*, which shows the power of the reference generation network. The network can help to design different reference patterns for different objects, which is helpful for the phase retrieval process. For the processing time, learning-based methods need less time due to parallel computing. Conventional methods, such as HIO, usually require hundreds of iterations to converge, and hence they need more time.

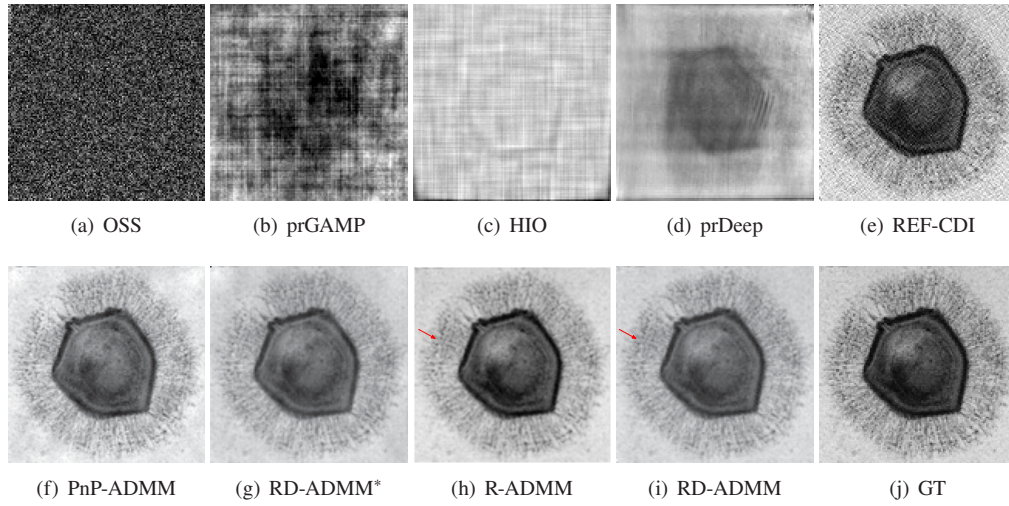


Fig. 3. PR results of mimivirus CHCDI image using different algorithms (PSNR(unit:dB), SSIM). OSS (7.61,0.053), prGAMP (12.81,0.226), HIO (9.97, 0.337), prDeep (17.36,0.421), REF-CDI (24.25,0.551), PnP-ADMM (27.65,0.967), RD-ADMM* (28.18,0.899), R-ADMM (37.17,0.969) and RD-ADMM (**37.85,0.984**). RD-ADMM provides the best visual result, the corner of the RD-ADMM* result has clear artifacts, and some textures of the virus are missing for the R-ADMM result, as indicated by the arrow in the figure.

Table 2. Quantitative evaluations of the PR results with different algorithms

Methods	PSNR (dB)	SSIM	Time (s)
OSS	7.45	0.034	18.415
BM3D-prGAMP	12.68	0.113	160.43
HIO	9.95	0.178	13.111
prDeep	17.14	0.241	34.82
REF-CDI	22.63	0.728	0.726
PnP-ADMM	23.95	0.809	0.221
RD-ADMM*	25.03	0.824	0.101
R-ADMM	35.96	0.941	0.073
RD-ADMM	36.76	0.949	0.102

4.5. CHCDI with different upsampling rates

Now we consider the influence of upsampling rates. In the training process, the imaging system is a $4\times$ oversampling DFT. In this section, we evaluate the proposed method on different sampling rates without retraining the whole network. Specifically, the reconstruction performances on the $3\times$ and $6\times$ imaging systems are evaluated.

PSNR and SSIM of the reconstructed images are summarized in Table 3. When the sampling rate is $6\times$, the reconstructed images have higher or equal SSIM than the $4\times$ DFT system for all kinds of images. Since the resolution of the captured intensity image is higher, it is more difficult for the denoising task, and the PSNR is lower than $4\times$ CHCDI system. When the sampling rate is lower, the quality of the reconstructed images decreases, but the structures can still be preserved for intensity images with SSIM >0.8 .

Table 3. PSNR and SSIM of the RD-ADMM PR results for CHCDI with different oversampling rates

Types	6×		3×	
	PSNR (dB)	SSIM	PSNR (dB)	SSIM
Intensity-only image	32.82	0.952	25.21	0.856
Intensity image with random phase	30.06	0.969	25.18	0.868
Phase-only image	31.71	0.866	25.59	0.571
Phase image with random intensity	33.97	0.964	24.14	0.662

5. Conclusions

In this paper, we propose a new recursive scheme based on the D-ADMM for phase retrieval, called RD-ADMM. It deals with complex-valued images directly based on the Wirtinger derivatives. A U-net is adopted to handle the regularization problem for complex-valued images. To show the performance of the RD-ADMM scheme, we focus on the HCDI system. Due to the undetermined reference pattern in HCDI, we use a learning-based method to find an optimal pattern based on the intensity information of the object light. The whole imaging system is called CHCDI. In this setup, the reference light varies with different objects, which can provide better encoding results in the far-field imaging plane. The optimal pattern can also make it easier for the decoder, RD-ADMM, to finish the decoding task and provide high-quality imaging results.

Funding. Research Grants Council of Hong Kong, (GRF 17200321); University Research Committee, University of Hong Kong (104006536).

Acknowledgments. The authors would like to thank Mr. Pei Zhang for useful discussions about this paper.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

Supplemental document. See [Supplement 1](#) for supporting content.

References

1. R. P. Millane, "Phase retrieval in crystallography and optics," *J. Opt. Soc. Am. A* **7**(3), 394–411 (1990).
2. Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, "Phase retrieval with application to optical imaging: a contemporary overview," *IEEE Signal Process. Mag.* **32**(3), 87–109 (2015).
3. G. Zheng, R. Horstmeyer, and C. Yang, "Wide-field, high-resolution Fourier ptychographic microscopy," *Nat. Photonics* **7**(9), 739–745 (2013).
4. Z. Ren, Z. Xu, and E. Y. Lam, "Learning-based nonparametric autofocusing for digital holography," *Optica* **5**(4), 337–344 (2018).
5. N. Chen, E. Y. Lam, T.-C. Poon, and B. Lee, "Sectional hologram reconstruction through complex deconvolution," *Opt. Lasers Eng.* **127**, 105945 (2020).
6. A. V. Oppenheim and J. S. Lim, "The importance of phase in signals," *Proc. IEEE* **69**(5), 529–541 (1981).
7. X. Wu, T. Fühner, A. Erdmann, and E. Y. Lam, "Level-set-based inverse lithography under random field shape uncertainty in a vector Hopkins imaging model," in *OSA Topical Meeting in Computational Optical Sensing and Imaging*, (2017), p. CW1B.4.
8. J. W. Goodman, *Introduction to Fourier Optics (4th Edition)* (Roberts and Company Publishers, 2017).
9. L. Song, Z. Ge, and E. Y. Lam, "Dual alternating direction method of multipliers for inverse imaging," *IEEE Trans. on Image Process.* **31**, 3295–3308 (2022).
10. R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik* **35**(2), 237–246 (1972).
11. J. R. Fienup, "Phase retrieval algorithms: a comparison," *Appl. Opt.* **21**(15), 2758–2769 (1982).
12. E. J. Candès, Y. C. Eldar, T. Strohmer, and V. Voroninski, "Phase retrieval via matrix completion," *SIAM Rev.* **57**(2), 225–251 (2015).
13. E. J. Candès, X. Li, and M. Soltanolkotabi, "Phase retrieval via Wirtinger flow: Theory and algorithms," *IEEE Trans. Inf. Theory* **61**(4), 1985–2007 (2015).
14. S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers* (Now Publishers Inc, 2011).

15. N. Parikh and S. Boyd, "Proximal algorithms," *FNT in Optimization* **1**(3), 127–239 (2014).
16. E. Y. Lam, N. Meng, and H. K. So, "Deep convolutional neural network for single-cell image analysis," *Proc. SPIE* **10505**, 105050K (2018).
17. Z. Ren, Z. Xu, and E. Y. Lam, "End-to-end deep learning framework for digital holographic reconstruction," *Adv. Photonics* **1**(01), 1 (2019).
18. A. Kappeler, S. Ghosh, J. Holloway, O. Cossairt, and A. Katsaggelos, "Ptychnet: CNN based Fourier ptychography," in *IEEE International Conference on Image Processing*, (2017), pp. 1712–1716.
19. G. Zhang, T. Guan, Z. Shen, X. Wang, T. Hu, D. Wang, Y. He, and N. Xie, "Fast phase retrieval in off-axis digital holographic microscopy through deep learning," *Opt. Express* **26**(15), 19388–19405 (2018).
20. T. Zeng, H. K.-H. So, and E. Y. Lam, "Computational image speckle suppression using block matching and machine learning," *Appl. Opt.* **58**(7), B39–B45 (2019).
21. L. Song and E. Y. Lam, "Iterative phase retrieval with a sensor mask," *Opt. Express* **30**(14), 25788–25802 (2022).
22. C. A. Metzler, P. Schniter, A. Veeraraghavan, and R. G. Baraniuk, "prDeep: Robust phase retrieval with a flexible deep network," in *International Conference on Machine Learning*, (2018), pp. 3501–3510.
23. E. T. Reehorst and P. Schniter, "Regularization by denoising: Clarifications and new interpretations," *IEEE Trans. Comput. Imaging* **5**(1), 52–67 (2019).
24. S. H. Chan, "Performance analysis of plug-and-play ADMM: A graph signal processing perspective," *IEEE Trans. Comput. Imaging* **5**(2), 274–286 (2019).
25. S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *IEEE Global Conference on Signal and Information Processing*, (2013), pp. 945–948.
26. K. Zhang, Y. Li, W. Zuo, L. Zhang, L. Van Gool, and R. Timofte, "Plug-and-play image restoration with deep denoiser prior," *IEEE Transactions on Pattern Analysis Mach. Intell.* **44**, 6360–6376 (2021).
27. T. Zeng and E. Y. Lam, "Robust reconstruction with deep learning to handle model mismatch in lensless imaging," *IEEE Trans. Comput. Imaging* **7**, 1080–1092 (2021).
28. P. Godard, M. Allain, V. Chamard, and J. Rodenburg, "Noise models for low counting rate coherent diffraction imaging," *Opt. Express* **20**(23), 25914–25934 (2012).
29. R. P. Millane, "Multidimensional phase problems," *J. Opt. Soc. Am. A* **13**(4), 725–734 (1996).
30. G. J. Williams, H. M. Quiney, A. G. Peele, and K. A. Nugent, "Fresnel coherent diffractive imaging: treatment and analysis of data," *New J. Phys.* **12**(3), 035020 (2010).
31. D. Sayre, "Some implications of a theorem due to Shannon," *Acta Crystallogr.* **5**(6), 843 (1952).
32. 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**(6742), 342–344 (1999).
33. A. Rana, J. Zhang, M. Pham, A. Yuan, Y. H. Lo, H. Jiang, S. J. Osher, and J. Miao, "Potential of attosecond coherent diffractive imaging," *Phys. Rev. Lett.* **125**(8), 086101 (2020).
34. Y. H. Lo, L. Zhao, M. Gallagher-Jones, A. Rana, J. J. Lodico, W. Xiao, B. Regan, and J. Miao, "In situ coherent diffractive imaging," *Nat. Commun.* **9**(1), 1826 (2018).
35. L. Song and E. Y. Lam, "Fast and robust phase retrieval for masked coherent diffractive imaging," *Photonics Res.* **10**(3), 758–768 (2022).
36. M. H. Seaberg, A. d'Aspremont, and J. J. Turner, "Coherent diffractive imaging using randomly coded masks," *Appl. Phys. Lett.* **107**(23), 231103 (2015).
37. X. Lu, Y. Shao, C. Zhao, S. Konijnenberg, X. Zhu, Y. Tang, Y. Cai, and H. P. Urbach, "Noniterative spatially partially coherent diffractive imaging using pinhole array mask," *Adv. Photonics* **1**(01), 1 (2019).
38. R. Horisaki, R. Egami, and J. Tanida, "Experimental demonstration of single-shot phase imaging with a coded aperture," *Opt. Express* **23**(22), 28691–28697 (2015).
39. S. Marchesini, S. Boutet, A. E. Sakdinawat, M. J. Bogan, S. Bajt, A. Barty, H. N. Chapman, M. Frank, S. P. Hau-Riege, A. Szöke, C. Cui, D. A. Shapiro, M. R. Howells, J. C. H. Spence, J. W. Shaevitz, J. Y. Lee, J. Hajdu, and M. M. Seibert, "Massively parallel X-ray holography," *Nat. Photonics* **2**(9), 560–563 (2008).
40. D. A. Barmherzig, J. Sun, P.-N. Li, T. J. Lane, and E. J. Candès, "Holographic phase retrieval and reference design," *Inverse Problems* **35**(9), 094001 (2019).
41. T. Latychevskaia, J.-N. Longchamp, and H.-W. Fink, "When holography meets coherent diffraction imaging," *Opt. Express* **20**(27), 28871–28892 (2012).
42. Z. Xu and E. Y. Lam, "Maximum a posteriori blind image deconvolution with Huber-Markov random-field regularization," *Opt. Lett.* **34**(9), 1453–1455 (2009).
43. K. Kreutz-Delgado, "The complex gradient operator and the CR-calculus," *arXiv preprint arXiv:0906.4835* (2009).
44. L. Song and E. Y. Lam, "MBD-GAN: Model-based image deblurring with a generative adversarial network," in *International Conference on Pattern Recognition*, (2021), pp. 7306–7313.
45. A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imaging Vis.* **40**(1), 120–145 (2011).
46. R. Remmert, *Theory of complex functions* (Springer Science & Business Media, 1991).
47. J. Liang, P. Stoica, Y. Jing, and J. Li, "Phase retrieval via the alternating direction method of multipliers," *IEEE Signal Process. Lett.* **25**(1), 5–9 (2018).
48. D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inf. Theory* **41**(3), 613–627 (1995).

49. N. G. Polson, J. G. Scott, and B. T. Willard, "Proximal algorithms in statistics and machine learning," *Statist. Sci.* **30**(4), 559–581 (2015).
50. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015*, (Springer, 2015), pp. 234–241.
51. K. Monakhova, J. Yurtsever, G. Kuo, N. Antipa, K. Yanny, and L. Waller, "Learned reconstructions for practical mask-based lensless imaging," *Opt. Express* **27**(20), 28075–28090 (2019).
52. S. H. Chan, X. Wang, and O. A. Elgendy, "Plug-and-play ADMM for image restoration: Fixed-point convergence and applications," *IEEE Trans. Comput. Imaging* **3**(1), 84–98 (2017).
53. L. Song and E. Y. Lam, "Phase retrieval with data-driven dual alternating direction method of multipliers for coherent diffraction imaging," in *Novel Techniques in Microscopy*, (Optica Publishing Group, 2021), pp. NTu1C–2.
54. B. S. Veeling, J. Linmans, J. Winkens, T. Cohen, and M. Welling, "Rotation equivariant CNNs for digital pathology," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (2018), pp. 210–218.
55. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980* (2014).
56. A. Hore and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *International Conference on Pattern Recognition*, (IEEE, 2010), pp. 2366–2369.
57. J. A. Rodriguez, R. Xu, C.-C. Chen, Y. Zou, and J. Miao, "Oversampling smoothness: an effective algorithm for phase retrieval of noisy diffraction intensities," *J. Appl. Crystallogr.* **46**(2), 312–318 (2013).
58. C. A. Metzler, A. Maleki, and R. G. Baraniuk, "BM3D-PRGAMP: Compressive phase retrieval based on BM3D denoising," in *IEEE International Conference on Image Processing*, (2016), pp. 2504–2508.