# TOTAL VARIATION BASED PHASE RETRIEVAL FOR POISSON NOISE REMOVAL

HUIBIN CHANG*, YIFEI LOU†, YUPING DUAN‡, AND STEFANO MARCHESINI§

**Abstract.** Phase retrieval plays an important role in vast industrial and scientific applications. We consider a noisy phase retrieval problem in which the magnitudes of the Fourier transform (or a general linear transform) of an underling object are corrupted by Poisson noise, since any optical sensors detect photons, and the number of detected photons follows the Poisson distribution. We propose a variational model for phase retrieval based on a total variation regularization as an image prior and maximum a posteriori estimation of Poisson noise model, which is referred to as "TV-PoiPR". We also propose an efficient numerical algorithm based on alternating direction method of multipliers and establish its convergence. Extensive experiments for coded diffraction, holographic, and ptychographic patterns are conducted using both real and complex-valued images to demonstrate the effectiveness of our proposed methods.

**Key words.** Phase Retrieval; Poisson Noise; Total Variation; Kullback-Leibler divergence; Alternating direction method of multipliers.

**AMS subject classifications.** 46N10, 49N30, 49N45, 65F22, 65N21

## 1. Introduction.

Phase retrieval (PR) plays a very important role in vast industrial and scientific applications, such as in astronomical imaging [21, 43], crystallography [31, 49], and optics [58, 56], *etc.* The goal is to reconstruct an object where pointwise magnitudes of the Fourier transform (FT) of the object are available. Since the phase of the FT is missing, this procedure is referred to as "phase retrieval".

Throughout the paper, we consider PR in a discrete setting, *i.e.*, an underlying object $u : \Omega = \{0, 1, \cdots, n-1\} \to \mathbb{C}$ is of size $n$ with $n = n_1 \times n_2$, in which we represent a 2-dimensional (2D) object with $n_1 \times n_2$ pixels in terms of a vector of size $n$ by a lexicographical order. The measured data are magnitudes of the Fourier transform of $u$, *i.e.*, $|\mathcal{F}u|^2$, where $|\cdot|^2$ denotes the pointwise square of the absolute value of a vector, $\mathcal{F} : \mathbb{C}^n \to \mathbb{C}^n$ denotes the discrete Fourier transform (DFT)

$$(\mathcal{F}u)(\omega_1 + \omega_2 n_1) := \frac{1}{\sqrt{n_1 n_2}} \sum_{\substack{0 \le t_j \le n_j - 1, \\ j=1,2}} u(t_1 + t_2 n_1) \exp(-2\pi \mathbf{i}(\omega_1 t_1/n_1 + \omega_2 t_2/n_2)),$$

$\forall\ 0 \le \omega_j \le n_j - 1,$ for $j = 1, 2$ and $\mathbf{i} = \sqrt{-1}$. In fact, DFT can be replaced with an arbitrary linear operator, thus leading to a general phase retrieval problem [3, 57, 36],

$$\text{To find } \ u \in \mathbb{C}^n, \quad s.t. \ \ |\mathcal{A}u|^2 = b, \tag{1.1}$$

where $\mathcal{A} : \mathbb{C}^n \to \mathbb{C}^m$ is a linear operator in the complex Euclidean space and $b : \Lambda = \{0, 1, \cdots, m-1\} \to \mathbb{R}_+$. In general, phase retrieval is ill-posed and yet challenging, since it does not have a unique solution without additional information.

*School of Mathematical Sciences, Tianjin Normal University, Tianjin, 300387, China, E-mail: changhuibin@gmail.com

†Department of Mathematical Sciences, University of Texas at Dallas, Dallas, TX 75080, USA, E-mail: yifei.lou@utdallas.edu

‡Center for Applied Mathematics, Tianjin University, 300072, China, E-mail: yuping.duan@tju.edu.cn

§Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720,USA, Email: smarchesini@lbl.gov

The computational tools for phase retrieval can be classified into two categories. The first category is based on alternating projection since a pioneer work of error reduction (ER) by Gerchberg and Saxton [28], and its variants such as hybrid input-output (HIO) [25], hybrid projection-reflection [4, 5], iterated difference-map (DF) [23], relaxed averaged alternation reflection (RAAR) [42], and a saddle point optimization method [47]; please refer to [46, 62] and the reference therein. However, these methods are lack of convergent guarantees due to nonconvex constraint sets for alternating projection. Recently, Netrapalli *et al.* [50] analyzed the global convergence for Gaussian measurements, while the convergence under general settings was established by Marchesini *et al.* [48]. Chen and Fannjiang [17] provided local and geometric convergence to a unique fixed point for a Douglas-Rachford splitting algorithm. In addition, gradient-type methods [12, 19] become popular. For example, Candés *et al.* [12] proposed Wirtinger flow (WF), which is comprised of a careful initialization by spectral method and adaptive steps in gradient descent. The WF approach was further improved by truncated Wirtinger flow (TWF) [19]. More general spectral initialization methods for PR were rigorously analyzed by Lu and Li [41], who gave a precise asymptotic characterization of spectral methods for Gaussian measurements. Gradient-based approaches often have first-order convergence, while a higher-order method was proposed by Qian *et al.* [51] to accelerate the convergence in a ptychographic PR problem. The second category is convex methods based on semi-definite programming (SDP) or convex relaxation of the quadratic equations in (1.1). For example, Candés *et al.* proposed PhaseLift [13] that formulates a convex trace (nuclear) norm minimization by a lift technique of SDP. PhaseCut by Waldspurger *et al.* [57] convexified the PR problem by separating phases and magnitudes. PhaseLiftOff, a nonconvex variant of PhaseLift by subtracting off Frobenius norm from the trace norm, was proposed by Yin and Xin [64] to retrieve the phase with less measurements than PhaseLift. It is well-known that the SDP-based methods are hardly scalable. In order to overcome this drawback, Friedlander and Macdo [27] considered a dual formulation based on the fact that the dimension of the dual problem grows much more slowly than the one of the primal problem. Another convex approach was addressed by Bahmani-Romberg [2] and independently by Goldstein-Studer [30] to relax quadratic equations of PR and meanwhile maximize the inner product between an "anchor" vector and the unknown.

In addition to computational tools, researchers also devote to theoretical analysis on phase retrieval, particularly focusing on the uniqueness of solution. In general, there exist trivial ambiguities [56], such as global phase shift, conjugate inversion, and spatial shift. Furthermore, it is well known [34, 7, 54] that there exist many non-trivial solutions for one-dimensional signals. In this paper, we focus on 2D image and $4n$ measurements are required to uniquely recover real-valued images as indicated in [32] for irreducible signals. Following the idea of holography, Candés *et al.* [10] proved exact phase retrieval from $3n$ Fourier measurements, in which the linear operator $\mathcal{A}$ can be expressed as

$$\mathcal{A}u = \begin{bmatrix} \mathcal{F}u \\ \mathcal{F}(u + \mathscr{D}^{s_1,s_2}u) \\ \mathcal{F}(u - \mathbf{i}\mathscr{D}^{s_1,s_2}u) \end{bmatrix}, \tag{1.2}$$

where

$$(\mathscr{D}^{s_1,s_2}u)(t_1 + t_2 n_1) = \exp\left(\tfrac{2\pi\mathbf{i}s_1 t_1}{n_1} + \tfrac{2\pi\mathbf{i}s_2 t_2}{n_2}\right) u(t_1 + t_2 n_1), \quad 0 \le t_i \le n_i - 1,$$

with $i = 1, 2$ and integers $s_1, s_2$ coprime to $n_1, n_2$ respectively. However, they found this amount of measurements is practically insufficient to recover $u$ exactly and stably, and $7n$ measurements are suggested instead. It is demonstrated that $3n$ measurements with $s_1 = s_2 = 1/2$ can recover phase information both theoretically and empirically in [16]. Furthermore, additional sparse prior information [37, 13, 59, 55] is helpful to establish the uniqueness of solution as well as to design efficient PR algorithms. For example, Jaganathan *et al.* [37] proved that signals of aperiodic support can be uniquely recovered with high probability if the DFT dimension is no less than $2n$, while Wang and Xu [59] focused on the minimal number of measurements required to deal with sparse signals in both real and complex cases. The study of minimal number of measurements for unique recovery extends to general PR (not limited to Fourier measurements), mostly discussing about the injectivity of the quadratic operator $|\mathcal{A}(\cdot)|^2$. Denote a nonlinear mapping $\mathcal{M} : (\mathbb{C}^n/S^1) \to \mathbb{R}^m_+$ as

$$\mathcal{M}(u) = |\mathcal{A}u|^2, \tag{1.3}$$

where $(\mathbb{C}^n/S^1)$ can be obtained by identifying $u \in \mathbb{C}^n$ with $\exp(\mathbf{i}\theta)$ for $\theta \in (0, 2\pi]$[1]. Generally speaking, the injectivity is guaranteed by collecting $m \geq 2n - 1$ [3] and $m \geq 4n - 4$ [20] measurements for real $u \in \mathbb{R}^n$ and complex $u \in \mathbb{C}^n$ signals respectively, provided that the transform $\mathcal{A}$ is generated by a generic frame[2]. In particular, Shechtman *et al.* [56] showed that the lower bound $2n - 1$ can be achieved with high probability by collecting full-spark random measurements for real-valued signals. Although Fourier type of measurements is emphatically not generic, the uniqueness of the phase retrieval can be proved under additional information, such as collecting multiple measurements by coded diffraction pattern (CDP) [11, 24] and holographic pattern [13, 16].

In this paper, we consider the phase retrieval problem from the measurements that are contaminated by the Poisson noise. It is very useful, since any optical sensor detects photons, and the number of measured photons varies following the Poisson distribution in the sense that the noise level depends on the ground-truth intensity, *i.e.*, stronger noise appears at lower intensity. Furthermore, when the intensity value is high enough, the Poisson noise at this pixel behaves like a "Gaussian" noise and therefore Poisson noise can be approximated by a Gaussian distribution via Anscombe transformation [1, 45]. In order to denoise the data from such measurements, *prior* information is important in the reconstruction procedure, and please refer to [39, 65, 53, 29, 38] for various ways of imposing prior knowledge.

We formulate a variational model by introducing a total variation (TV) regularization to enforce sparsity, which is widely used in image processing since the seminal work of [52]. We extend our previous work [16] (focusing on holographic pattern for real-valued images) to more general PR setting (1.1) and further deal with the Poisson noise in measurements for both real and complex-valued images. We prove the existence and uniqueness of the minimizer to our proposed model and design an efficient alternating direction method of multipliers (ADMM) [63, 6] with convergence guarantees. Numerical experiments are based on Fourier measurements generated by CDP [11], holographic pattern [13, 16] in (1.2), and ptychographic pattern [62]. We show that satisfactory PR results can be obtained from noisy measurements of these three patterns for both real and complex-valued images. Furthermore, the proposed

[1] The quotient space is needed to study the injectivity up to a global phase shift

[2] Generic frame means a $K-$element frame belongs to an open dense subset of the set of all $K-$element frames in $\mathbb{R}^n$ or $\mathbb{C}^n$ [3]

method can deal with a large amount of downsampling; especially $0.4n$ measurements for CDP are shown to be sufficient for natural images with simple structures.

The rest of this paper is organized as follows. In Section 2, a TV regularized model for Poisson noise removal in a PR problem, referred to as TV-PoiPR, is established, where the existence of the minimizer to the proposed model is obtained. Section 3 discusses an ADMM algorithm for TV-PoiPR, referred to as "Algorithm I", with convergence analysis. Section 4 devotes to a special case, in which TV is not present, referred to as PoiPR. In this case, we further investigate the uniqueness of the solution. We also present an efficient ADMM referred to as "Algorithm II" for PoiPR and an accelerated algorithm, referred to as "Algorithm III", for a large number of CDP measurements. Numerical experiments are performed in Section 5 to demonstrate the effectiveness and robustness of the proposed methods for image recovery from noisy and incomplete phaseless data. Conclusions and future works are given in Section 6.

## 2. Proposed Model.

**2.1. Maximum a Posteriori (MAP) Estimation.** Poisson noise is one of the most common types of noise that occurs for photon-counting. Its name is stemmed from Poisson distribution, defined as follows

$$\Pr_\mu(n) = \frac{e^{-\mu}\mu^n}{n!}, \qquad n \geq 0,$$

where $\mu$ is mean and standard deviation. The number of photons measured at each pixel, denoted as $f(i)$, follows i.i.d. Poisson distributions with $\mu$ being the ground-truth value, $g(i)$, for $i \in \Omega$, denoted as

$$f(i) \overset{\text{ind.}}{\sim} \text{Poisson}(g(i)),\ \forall i \in \Omega. \tag{2.1}$$

Given the measured data $f$, the denoising problem is then formulated via MAP of a clean image $g$, which can be expressed as $\max \Pr(g(i)|f(i))$. By Bayes' Law, we have

$$\Pr(g(i)|f(i)) = \frac{\Pr(f(i)|g(i))\Pr(g(i))}{\Pr(f(i))}. \tag{2.2}$$

Therefore, $\max \Pr(g(i)|f(i))$ is equivalent to $\max \Pr(f(i)|g(i))\Pr(g(i))$. It follows from the definition of Poisson distribution that

$$\Pr(f(i)|g(i)) = \Pr_{g(i)}(f(i)) = \frac{e^{-g(i)}g(i)^{f(i)}}{(f(i))!}, \tag{2.3}$$

which suggests minimizing the logarithm of the $\Pr(f(i)|g(i))\Pr(g(i))$ instead, *i.e.*

$$\begin{aligned} &\min_{g\geq 0} \sum\nolimits_{i\in\Omega} -\log \Pr(f(i)|g(i)) - \log \Pr(g(i)) \\ &= \min_{g\geq 0} \sum\nolimits_{i\in\Omega} (g(i) - f(i)\log g(i)) - \log \Pr(g(i)), \end{aligned} \tag{2.4}$$

where the first term is related with the famous Kullback-Leibler (KL) divergence, and define $0\log 0 = 0$ and $\log 0 = -\infty$.

Le *et al.* [39] further incorporated a TV regularization term to improve the image reconstruction from Poisson noise, i.e.,

$$\min_{g\geq 0} \lambda \text{TV}(g) + \sum\nolimits_{i\in\Omega} (g(i) - f(i)\log g(i)), \tag{2.5}$$

where $\mathrm{TV}(g) = \|\nabla g\|_1 = \sum_j \sqrt{|(\nabla_x g)(j)|^2 + |(\nabla_y g)(j)|^2}$, and $\nabla_x$ and $\nabla_y$ define the $x-$direction and $y-$direction forward difference operators respectively. The model (2.5) can be regarded as incorporating a Gibbs prior distribution $\Pr(g) = \exp\left(-\lambda \mathrm{TV}(g)\right)$, into the MAP for Poisson noise removal. As indicated in [39], Poisson maximum likelihood (KL divergence) gives better recovery results than standard Gaussian maximum likelihood in that low contrast features are better preserved and the reconstructed images have higher contrast. More efficient methods were proposed to solve the TV regularized model, such as gradient descent [39], expectation-maximization (EM) [8], multigrid [15], and splitting methods [26, 9, 14, 61].

**2.2. TV-PoiPR.** In this paper, we consider a phase retrieval problem in which the measurements $b = |\mathcal{A}u|^2$ in (1.1) are corrupted by Poisson noise, that is,

$$f(i) \overset{\text{ind.}}{\sim} \text{Poisson}(b(i)),\ \forall i \in D,$$

with an undersampling set $D \subseteq \Lambda$. In the light of (2.5), we establish a minimization problem, referred to as "TV-PoiPR",

$$\min_{u\in\mathbb{C}^n} \mathcal{G}(u) := \lambda \mathrm{TV}(u) + \frac{1}{2}\sum\nolimits_{i\in D}(|(\mathcal{A}u)(i)|^2 - f(i)\log|(\mathcal{A}u)(i)|^2). \tag{2.6}$$

where $u \in \mathbb{C}^n$ is an underlying image that we want to reconstruct from intensity data.

The rest of the section devotes to theoretical analysis of TV-PoiPR. In particular, we prove existence of solutions to (2.6) under mild conditions. The uniqueness of solutions to (2.6) with $\lambda = 0$ is studied in Section 4. Note that we only focus on the discrete setting, while it is straightforward to extend the analysis to a continuous setting using the compactness property of bounded variation (BV) space and lower semi-continuity of the objective functional $\mathcal{G}$.

**Theorem 1.** *Assume that there exists a positive number $\beta$ for the operator $\mathcal{A}$, such that*

$$\beta\|u\|_2 \le \|\mathcal{A}u\|_{2,D},\ \forall u \in \mathbb{C}^n, \tag{2.7}$$

*with $\|z\|_{2,D} = \sqrt{\sum_{i\in D}|z(i)|^2}$, $z \in \mathbb{C}^m$, then there exists a minimizer $u^*$ for (2.6),* i.e. $u^* = \arg\min_{u\in\mathbb{C}^n} \mathcal{G}(u)$.

*Proof.* Define a data fidelity for Poisson noise,

$$\mathcal{J}(u) = \sum\nolimits_{i\in D}(|(\mathcal{A}u)(i)|^2 - f_i\log|(\mathcal{A}u)(i)|^2). \tag{2.8}$$

Since $x - f(i)\log x \ge f(i) - f(i)\log f(i), \forall x \ge 0$, and $i \in \{i \in D :\ f(i) > 0\}$, we have

$$\mathcal{J}(u) \ge \sum\nolimits_{\{i\in D:\ f(i)>0\}}(f(i) - f(i)\log|f(i)|),\ \forall u \in \mathbb{C}^n, \tag{2.9}$$

such that $\mathcal{G}(u)$ is bounded below. Therefore, we can choose a minimizing sequence $\{u_n\}$, such that $\mathcal{G}(u_0) \ge \mathcal{G}(u_1) \ge \cdots \ge \mathcal{G}(u_j) \ge \cdots$. We have

$$\begin{aligned}
&\mathcal{G}(u_0) \ge \lambda \mathrm{TV}(u_j) + \sum\nolimits_{i\in D}(|z_j(i)|^2 - f(i)\log|z_j(i)|^2)\\
&\ge \sum\nolimits_{\{i\in D:\ |z_j(i)|\ge 1\}}|z_j(i)|^2 - \|f\|_\infty \sum\nolimits_{\{i\in D:\ |z_j(i)|\ge 1\}}\log|z_j(i)|^2\\
&= \sum\nolimits_{\{i\in D:\ |z_j(i)|\ge 1\}}|z_j(i)|^2 - \|f\|_\infty \log\prod\nolimits_{\{i\in D:\ |z_j(i)|\ge 1\}}|z_j(i)|^2\\
&\ge \sum\nolimits_{\{i\in D:\ |z_j(i)|\ge 1\}}|z_j(i)|^2 - m_0\|f\|_\infty \log\left(\tfrac{1}{m_0}\sum\nolimits_{\{i\in D:\ |z_j(i)|\ge 1\}}|z_j(i)|^2\right)
\end{aligned} \tag{2.10}$$

where $z_j = \mathcal{A}u_j$, and $m_0 = \#\{i \in D : |z_j(i)| \geq 1\} \leq m$. Therefore, one readily obtains that $\{\|\mathcal{A}u_j\|_{2,D}\}$ is a bounded sequence such that $\{\|u_j\|\}$ is also bounded by (2.7). By the compactness of discrete $L^2$ space, one can readily select a convergent subsequence $\{u_{j_k}\} \subseteq \{u_j\}$, such that $\lim_{k\to+\infty} u_{j_k} = u^*$. By the continuity of the objective functional $\mathcal{G}$, we have $\lim_{k\to+\infty} \mathcal{G}(u_{j_k}) = \mathcal{G}(u^*)$. That concludes the proof. □

Note that we can relax the assumption (2.7) as

$$\|\mathcal{A}\mathbf{1}\|_{2,D} \neq 0, \tag{2.11}$$

where $\mathbf{1} \in \mathbb{C}^m$ whose elements are all equal to one, and obtain a similar results.

**Proposition 1.** *If the assumption* (2.11) *holds, there exists a minimizer* $u^*$ *for* (2.6), i.e. $u^* = \arg\min_{u\in\mathbb{C}^n} \mathcal{G}(u)$.

*Proof.* Since $\mathcal{G}(u)$ is bounded below by (2.9), we can choose a minimizing sequence $\{u_n\}$, such that $\mathcal{G}(u_0) \geq \mathcal{G}(u_1) \geq \cdots \geq \mathcal{G}(u_j) \geq \cdots$. We will show the minimizing sequence is bounded. Rewrite

$$u_j = \hat{u}_j + c_j\mathbf{1}, \tag{2.12}$$

where the constant $c_j = \sum_{i\in\Lambda} u_j(i)/m$. Since it exists a positive constant $C$, such that

$$\|z - \frac{1}{m}\sum\nolimits_{i\in\Lambda} z(i)\mathbf{1}\| \leq C\ \mathrm{TV}(z),\ \forall\ z \in \mathbb{C}^m,$$

we have the boundedness of $\{\hat{u}_j\}$. By (2.10), we obtain the boundedness of $\|\mathcal{A}u_j\|_{2,D}$. Since

$$\begin{aligned} |c_j|\|\mathcal{A}\mathbf{1}\|_{2,D} &= \|\mathcal{A}\hat{u}_j - \mathcal{A}u_j\|_{2,D} \leq \|\mathcal{A}\hat{u}_j\|_{2,D} + \|\mathcal{A}u_j\|_{2,D} \\ &\leq \|\mathcal{A}\hat{u}_j\| + \|\mathcal{A}u_j\|_{2,D} \leq \|\mathcal{A}\|\|\hat{u}_j\| + \|\mathcal{A}u_j\|_{2,D}, \end{aligned} \tag{2.13}$$

with the help of assumption (2.11), one can see that $\{c_j\}$ is bouneded. It follows from (2.12) that $\{u_j\}$ is bounded. Due to the lower semi-continuity of $\mathcal{G}$, we can complete the proof by similar analysis as in Theorem 1. □

Remark 2.1. *We can extend the existence proof for any lower semi-continuous regularization term, not limited to the TV. In other words, we can prove that there exists a minimizer of* $\min_{u\in\mathbb{C}^n} \mathcal{R}(u) + \mathcal{J}(u)$, *for any lower semi-continuous functional* $R(u)$ *and* $\mathcal{J}(u)$ *defined in* (2.8).

Remark 2.2. *The assumptions of* (2.7) *and* (2.11) *are satisfied for some special patterns such as CDP and holographic patterns without undersampling, i.e.,* $D = \Lambda$.

In general, it is difficult to show that a nonconvex minimization problem has a unique solution. We will investigate the uniqueness of the minimizer (2.6) without regularization term *i.e.* $\lambda = 0$ in Section 4.

**3. Numerical Algorithms.** We apply the ADMM algorithm [6] to solve the proposed model (2.6), which is equivalent to

$$\min_{u\in\mathbb{C}^n} \lambda\|\boldsymbol{p}\|_1 + \frac{1}{2}\sum\nolimits_{i\in D}(|z(i)|^2 - f(i)\log|z(i)|^2), \quad s.t.\ \ z = \mathcal{A}u,\ \boldsymbol{p} = \nabla u, \tag{3.1}$$

where $\nabla u = (\nabla_x u, \nabla_y u)$ denotes the gradient operator and we assume zero boundary conditions for these two gradient operators. One can readily construct the augmented

Lagrangian as

$$\begin{aligned}\mathcal{L}_{r_1,r_2}(u,z,\boldsymbol{p};v,\boldsymbol{q}) &:= \lambda\|\boldsymbol{p}\|_1 + \frac{1}{2}\sum\nolimits_{i\in D}(|z(i)|^2 - f(i)\log|z(i)|^2) \\ &+ \mathrm{Re}(\langle z-\mathcal{A}u, v\rangle) + \mathrm{Re}(\langle \boldsymbol{p}-\nabla u, \boldsymbol{q}\rangle) + \frac{r_1}{2}\|z-\mathcal{A}u\|^2 + \frac{r_2}{2}\|\boldsymbol{p}-\nabla u\|^2,\end{aligned} \tag{3.2}$$

where $\langle\cdot,\cdot\rangle$ denotes the complex inner product of two vectors, and $v, \boldsymbol{q}$ are called Lagrange multipliers or dual variables. Alternating minimization for the above Lagrangian consists of solving three subproblems *w.r.t.* $u, z, \boldsymbol{p}$, followed by updating dual variables. Below we elaborate on solving each subproblem.

**3.1. $u$-subproblem.** The $u$-subproblem is

$$\min_{u\in\mathbb{C}^n} \mathrm{Re}(\langle z-\mathcal{A}u, v\rangle) + \mathrm{Re}(\langle \boldsymbol{p}-\nabla u, \boldsymbol{q}\rangle) + \frac{r_1}{2}\|z-\mathcal{A}u\|^2 + \frac{r_2}{2}\|\boldsymbol{p}-\nabla u\|^2, \tag{3.3}$$

which can be simplified as

$$\min_{u\in\mathbb{C}^n} \frac{r_1}{2}\|z+v/r_1-\mathcal{A}u\|^2 + \frac{r_2}{2}\|\boldsymbol{p}+\boldsymbol{q}/r_2-\nabla u\|^2. \tag{3.4}$$

The operator $\mathcal{A}$ can be rewritten by the summation of two real linear operators $\mathcal{A} = \mathcal{A}_1 + \mathbf{i}\mathcal{A}_2$. In a similar way, $u = u_1 + \mathbf{i}u_2$ with $u_i \in \mathbb{R}^n$. One can obtain $\mathcal{A}u = \mathcal{A}_1u_1 - \mathcal{A}_2u_2 + \mathbf{i}(\mathcal{A}_1u_2 + \mathcal{A}_2u_1)$. Letting $v_0 = z + v/r_1, \boldsymbol{p}_0 = \boldsymbol{p} + \boldsymbol{q}/r_2$, (3.4) is equivalent to

$$\begin{aligned}&\frac{r_1}{2}\|v_0-\mathcal{A}u\|^2 + \frac{r_2}{2}\|\boldsymbol{p}_0-\nabla u\|^2 \\ =&\frac{r_1}{2}\|\mathcal{A}_1u_1-\mathcal{A}_2u_2-\mathrm{Re}(v_0)\|^2 + \frac{r_1}{2}\|\mathcal{A}_1u_2+\mathcal{A}_2u_1-\mathrm{Im}(v_0)\|^2 \\ &+ \frac{r_2}{2}\|\mathrm{Re}(\boldsymbol{p}_0)-\nabla u_1\|^2 + \frac{r_2}{2}\|\mathrm{Im}(\boldsymbol{p}_0)-\nabla u_2\|^2.\end{aligned} \tag{3.5}$$

By computing the derivative *w.r.t.* $u_i$, one obtains

$$\begin{aligned}&\begin{bmatrix} r_1(A_1^TA_1+A_2^TA_2)-r_2\Delta & -r_1(A_1^TA_2-A_2^TA_1) \\ r_1(A_1^TA_2-A_2^TA_1) & r_1(A_1^TA_1+A_2^TA_2)-r_2\Delta \end{bmatrix}\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \\ =&\begin{bmatrix} r_1(A_1^T\mathrm{Re}(v_0)+A_2^T\mathrm{Im}(v_0)) - r_2(\nabla\cdot\mathrm{Re}(\boldsymbol{p}_0)) \\ r_1(-A_2^T\mathrm{Re}(v_0)+A_1^T\mathrm{Im}(v_0)) - r_2(\nabla\cdot\mathrm{Im}(\boldsymbol{p}_0)) \end{bmatrix},\end{aligned} \tag{3.6}$$

with suitable boundary condition for $u_i$, where $\Delta u = \nabla\cdot(\nabla u)$, the divergence operator $\nabla(\cdot)$ denotes the conjugate operator of gradient $\nabla$. Readily we have

$$\begin{aligned}&\begin{bmatrix} r_1(\mathrm{Re}(\mathcal{A}^*\mathcal{A}))-r_2\Delta & -r_1(\mathrm{Im}(\mathcal{A}^*\mathcal{A})) \\ r_1(\mathrm{Im}(\mathcal{A}^*\mathcal{A})) & r_1(\mathrm{Re}(\mathcal{A}^*\mathcal{A}))-r_2\Delta \end{bmatrix}\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \\ =&\begin{bmatrix} r_1(\mathrm{Re}(\mathcal{A}^*v_0)) - r_2(\nabla\cdot\mathrm{Re}(\boldsymbol{p}_0)) \\ r_1(\mathrm{Im}(\mathcal{A}^*v_0)) - r_2(\nabla\cdot\mathrm{Im}(\boldsymbol{p}_0)) \end{bmatrix},\end{aligned} \tag{3.7}$$

since

$$A_1^TA_1 + A_2^TA_2 = \mathrm{Re}(\mathcal{A}^*\mathcal{A}), \qquad A_1^TA_2 - A_2^TA_1 = \mathrm{Im}(\mathcal{A}^*\mathcal{A}). \tag{3.8}$$

Note that there is a unique solution to the $u-$subproblem, as characterized in Proposition 2.

**Proposition 2.** *The linear equations (3.7) admit a unique solution.*

*Proof.* Let

$$\mathcal{B} := \begin{bmatrix} r_1(\mathrm{Re}(\mathcal{A}^*\mathcal{A})) - r_2\Delta & -r_1(\mathrm{Im}(\mathcal{A}^*\mathcal{A})) \\ r_1(\mathrm{Im}(\mathcal{A}^*\mathcal{A})) & r_1(\mathrm{Re}(\mathcal{A}^*\mathcal{A})) - r_2\Delta \end{bmatrix}. \tag{3.9}$$

We show that the linear operator $\mathcal{B}$ is nonsingular. For this purpose, we calculate

$$\begin{aligned} \langle \mathcal{B}(u_1,u_2)^T, (u_1,u_2)^T\rangle &= \langle r_1\mathrm{Re}(\mathcal{A}^*\mathcal{A})u_1 - r_2\Delta u_1, u_1\rangle - \langle r_1\mathrm{Im}(\mathcal{A}^*\mathcal{A})u_2, u_1\rangle \\ &\quad + \langle r_1\mathrm{Im}(\mathcal{A}^*\mathcal{A})u_1, u_2\rangle + \langle r_1\mathrm{Re}(\mathcal{A}^*\mathcal{A})u_2 - r_2\Delta u_2, u_2\rangle \\ &= \langle r_1\mathrm{Re}(\mathcal{A}^*\mathcal{A})u_1 - r_2\Delta u_1, u_1\rangle + \langle r_1\mathrm{Re}(\mathcal{A}^*\mathcal{A})u_2 - r_2\Delta u_2, u_2\rangle \\ &= r_1\left(\langle \mathrm{Re}(\mathcal{A}^*\mathcal{A})u_1, u_1\rangle + \langle \mathrm{Re}(\mathcal{A}^*\mathcal{A})u_2, u_2\rangle\right) + r_2\left(\langle -\Delta u_1, u_1\rangle + \langle -\Delta u_2, u_2\rangle\right). \end{aligned}$$

Since the operator $\Delta$ is negative definite with the difference scheme of zero boundary condition and $\mathrm{Re}(\mathcal{A}^*\mathcal{A})$ is semi-positive by (3.8), we have $\langle \mathcal{B}(u_1,u_2)^T, (u_1,u_2)^T\rangle > 0$, for $(u_1,u_2) \neq 0$, which implies that $\mathcal{B}$ is nonsingular and hence the solution to $u$-subproblem (3.7) is unique. ∎

Remark 3.1. *We can simplify the solution of $u$-subproblem if the matrix $\mathcal{A}$ involves Fourier measurements with masks $\{I_j\}_{j=1}^J$ as*

$$\mathcal{A}u = \begin{bmatrix} \mathcal{F}(I_1 \circ u) \\ \mathcal{F}(I_2 \circ u) \\ \vdots \\ \mathcal{F}(I_J \circ u) \end{bmatrix}, \tag{3.10}$$

*where $\circ$ denotes the pointwise multiplication, $I_j$ is a (mask) matrix indexed by $j$, each of which is represented by a vector in $\mathbb{C}^n$ in a lexicographical order. Therefore we have $\mathcal{A}^*\mathcal{A} = \mathrm{diag}(\sum_j I_j^* \circ I_j)$, which is a real-valued matrix. In this case, we can obtain $u = u_1 + \mathbf{i}u_2$ by*

$$\begin{cases} u_1 = \left(r_1(\mathrm{Re}(\mathcal{A}^*\mathcal{A})) - r_2\Delta\right)^{-1}\left(r_1\mathrm{Re}(\mathcal{A}^*v_0) - r_2\nabla\cdot\mathrm{Re}(\boldsymbol{p}_0)\right), \\ u_2 = \left(r_1(\mathrm{Re}(\mathcal{A}^*\mathcal{A})) - r_2\Delta\right)^{-1}\left(r_1\mathrm{Im}(\mathcal{A}^*v_0) - r_2\nabla\cdot\mathrm{Im}(\boldsymbol{p}_0)\right), \end{cases} \tag{3.11}$$

*which solves the following block-diagonal equations*

$$\begin{bmatrix} r_1\mathrm{Re}(\mathcal{A}^*\mathcal{A}) - r_2\Delta & \mathbf{0} \\ \mathbf{0} & r_1\mathrm{Re}(\mathcal{A}^*\mathcal{A}) - r_2\Delta \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} r_1\mathrm{Re}(\mathcal{A}^*v_0) - r_2\nabla\cdot\mathrm{Re}(\boldsymbol{p}_0) \\ r_1\mathrm{Im}(\mathcal{A}^*v_0) - r_2\nabla\cdot\mathrm{Im}(\boldsymbol{p}_0) \end{bmatrix}, \tag{3.12}$$

*with the zero matrix $\mathbf{0}$. A simplified form can be written as*

$$u = \left(r_1\mathrm{Re}(\mathcal{A}^*\mathcal{A}) - r_2\Delta\right)^{-1}\left(r_1\mathcal{A}^*v_0 - r_2\nabla\cdot\boldsymbol{p}_0\right).$$

*In our experiments, we consider CDP, holographic and ptychographic patterns, leading to a diagonal matrix $\mathcal{A}^*\mathcal{A}$ with different values for diagonal entries. As a result, FFT can not be directly used. Since the coefficient matrix $\mathcal{B}$ is sparse and symmetric, one can use the conjugate gradients (CG) method or bicojugate gradient*

*(BICG) to solve (3.12). We observe empirically that one or two CG/BICG iterations are sufficient to yield satisfactory results.*

REMARK 3.2. *If we consider real-valued image $u \in \mathbb{R}^n$, then the auxiliary variables $\boldsymbol{p}, \boldsymbol{q} \in \mathbb{R}^{2n}$ are also real-valued. In other words, we confine the minimization problem (3.3)* w.r.t. *$u$ on a real-valued space and the corresponding Euler equation* w.r.t. *real-valued $u$ is given as*

$$(r_1 \mathrm{Re}(\mathcal{A}^*\mathcal{A}) - r_2\Delta)\, u = r_1 \mathrm{Re}(\mathcal{A}^* v_0) - r_2 \nabla \cdot (\boldsymbol{p} + \boldsymbol{q}/r_2).$$

**3.2. $z$-subproblem.** We rewrite the $z$-subproblem as

$$\min_{z\in\mathbb{C}^m} \frac{1}{2}\sum\nolimits_{i\in D}(|z(i)|^2 - f(i)\log|z(i)|^2) + \frac{r_1}{2}\|z - \mathcal{A}u + v/r_1\|^2. \tag{3.13}$$

It is straightforward that the minimization *w.r.t.* $z$ is equivalent to minimizing *w.r.t.* each entry $z(i)$ independently and for $i \in \Lambda \setminus D$, an optimal solution is $z^*(i) = (\mathcal{A}u)(i) - \frac{v(i)}{r_1}$.

As for all $i \in D$, we can decompose the minimization problem *w.r.t.* $z(i)$,

$$z^*(i) = \arg\min_{z(i)\in\mathbb{C}} \frac{1}{2}(|z(i)|^2 - f(i)\log|z(i)|^2) + \frac{r_1}{2}|z(i) - (\mathcal{A}u)(i) + v(i)/r_1|^2. \tag{3.14}$$

into two subproblems, *i.e.*, $|z(i)|$ and $\mathrm{sign}(z(i))$ ($\mathrm{sign}(z(i)) = \frac{z(i)}{|z(i)|}$ if $z(i) \neq 0$; otherwise $\mathrm{sign}(0) = c$ with an arbitrary constant $c \in \mathbb{C}$ with unity length). One can readily obtain $\mathrm{sign}(z^*(i)) = \mathrm{sign}((\mathcal{A}u)(i) - v(i)/r_1)$. To minimize the subproblem *w.r.t.* $|z^*(i)|$, we have

$$|z^*(i)| = \arg\min_{\rho\in\mathbb{R}_+} \frac{1}{2}(\rho^2 - f(i)\log\rho^2) + \frac{r_1}{2}\big(\rho - |(\mathcal{A}u)(i) - v(i)/r_1|\big)^2,$$

which has a closed-form solution,

$$|z^*(i)| = \frac{r_1|(\mathcal{A}u)(i) - v(i)/r_1| + \sqrt{r_1^2|(\mathcal{A}u)(i) - v(i)/r_1|^2 + 4(1+r_1)f(i)}}{2(1+r_1)}.$$

Letting $w = \mathcal{A}u - v/r_1$, we arrive at a simplified expression of

$$z^*(i) = \frac{r_1|w(i)| + \sqrt{r_1^2|w(i)|^2 + 4(1+r_1)f(i)}}{2(1+r_1)}\mathrm{sign}(w(i)),\ \forall\, i \in D.$$

**3.3. $\boldsymbol{p}$-subproblem and overall algorithm.** At last, we consider the $\boldsymbol{p}$-subproblem

$$\min_{\boldsymbol{p}\in\mathbb{C}^{2n}} \lambda\|\boldsymbol{p}\|_1 + \frac{r_2}{2}\|\boldsymbol{p} - \nabla u + \boldsymbol{q}/r_2\|^2, \tag{3.15}$$

The solution is a soft shrinkage of variable $\nabla u - \boldsymbol{q}/r_2$ as

$$\boldsymbol{p}^* = \mathrm{Thresh}_{\lambda/r_2}(\nabla u - \boldsymbol{q}/r_2),$$

with $\mathrm{Thresh}_\eta(\boldsymbol{q}) = \max\{0, |\boldsymbol{q}| - \eta\}\,\mathrm{sign}(\boldsymbol{q})$.

In summary, a pseudo code of ADMM for solving TV-PoiPR is provided in Algorithm I.

Algorithm I: ADMM for TV-PoiPR (2.6)

1. Initialization: Set $\boldsymbol{q}^0 = 0, v^0 = 0,\ u^0, z^0 = \mathcal{A}u^0, \boldsymbol{p}^0 = \nabla u^0.\ j = 0.$
2. Solve $u^{j+1} = u_1^{j+1} + \mathbf{i}u_2^{j+1}$ by solving the following equations as

$$\mathcal{B}\begin{bmatrix} u_1^{j+1} \\ \\ u_2^{j+1} \end{bmatrix} = \begin{bmatrix} r_1 \mathrm{Re}(\mathcal{A}^* v_0^j) - r_2 \nabla \cdot \mathrm{Re}(\boldsymbol{p}_0^j) \\ \\ r_1 \mathrm{Im}(\mathcal{A}^* v_0^j) - r_2 \nabla \cdot \mathrm{Im}(\boldsymbol{p}_0^j) \end{bmatrix},$$

   with $v_0^j = z^j + v^j/r_1, \boldsymbol{p}_0^j = \boldsymbol{p}^j + \boldsymbol{q}^j/r_2$, and the operator $\mathcal{B}$ defined in (3.9).
3. Solve $z^{j+1}$ and $\boldsymbol{p}^{k+1}$ in parallel by

(3.16)
$$z^{j+1}(i) = \begin{cases} \dfrac{r_1|w^j(i)| + \sqrt{r_1^2|w^j(i)|^2 + 4(1+r_1)f(i)}}{2(1+r_1)} \mathrm{sign}(w^j(i)), \forall\ i \in D, \\ w^j(i), \qquad\qquad \forall i \in \Lambda \setminus D, \end{cases}$$

   with $w^j = \mathcal{A}u^{j+1} - v^j/r_1$, and

(3.17)
$$\boldsymbol{p}^{j+1} = \mathrm{Thresh}_{\lambda/r_2}(\nabla u^{j+1} - \boldsymbol{q}^j/r_2).$$

4. Update multipliers as

(3.18)
$$\begin{aligned} v^{j+1} &= v^j + r_1(z^{j+1} - \mathcal{A}u^{j+1}), \\ \boldsymbol{q}^{j+1} &= \boldsymbol{q}^j + r_2(\boldsymbol{p}^{j+1} - \nabla u^{j+1}). \end{aligned}$$

5. If some stopping condition is satisfied, stop the iterations and output the iterative solution; else set $j = j + 1$, and goto Step 2.

**3.4. Convergence analysis.** We then discuss a convergent behavior of the proposed algorithm. We show that the algorithm converges to a saddle point by satisfying Karush-Kuhn-Tucker (KKT) conditions, which is a typical situation for nonconvex problems. The KKT conditions for the Lagrangian $\mathcal{L}_{r_1,r_2}(u, z, \boldsymbol{p}; \boldsymbol{q}, v)$ in (3.2) are defined as follows,

(3.19)
$$\begin{cases} \partial_u \mathcal{L}_{r_1,r_2}(\tilde{u}, \tilde{z}, \tilde{\boldsymbol{p}}; \tilde{v}, \tilde{\boldsymbol{q}}) = 0, \\ \partial_z \mathcal{L}_{r_1,r_2}(\tilde{u}, \tilde{z}, \tilde{\boldsymbol{p}}; \tilde{v}, \tilde{\boldsymbol{q}}) = 0, \\ \partial_{\boldsymbol{p}} \mathcal{L}_{r_1,r_2}(\tilde{u}, \tilde{z}, \tilde{\boldsymbol{p}}; \tilde{v}, \tilde{\boldsymbol{q}}) \ni 0, \\ \partial_v \mathcal{L}_{r_1,r_2}(\tilde{u}, \tilde{z}, \tilde{\boldsymbol{p}}; \tilde{v}, \tilde{\boldsymbol{q}}) = 0, \\ \partial_{\boldsymbol{q}} \mathcal{L}_{r_1,r_2}(\tilde{u}, \tilde{z}, \tilde{\boldsymbol{p}}; \tilde{v}, \tilde{\boldsymbol{q}}) = 0. \end{cases}$$

for any saddle point $(\tilde{u}, \tilde{z}, \tilde{\boldsymbol{p}}, \tilde{v}, \tilde{\boldsymbol{q}})$. Since the Lagrangian $\mathcal{L}_{r_1,r_2}(u, z, \boldsymbol{p}; \boldsymbol{q}, v)$ is nonconvex *w.r.t.* $u, z, \boldsymbol{p}$, we detail the KKT conditions corresponding to these three

variables:

$$\mathcal{A}^*\tilde{v} = \nabla \cdot \tilde{\boldsymbol{q}}, \tag{3.20}$$

$$\begin{cases} \tilde{z}(i) = \dfrac{|\tilde{v}(i)| + \sqrt{|\tilde{v}(i)|^2 + 4f(i)}}{2}\text{sign}(\tilde{v}(i)), \forall\ i \in D, \\ 0 = \tilde{v}(i), \forall\ i \in \Lambda \setminus D, \end{cases} \tag{3.21}$$

$$\begin{cases} 0 \ni \lambda\partial_{\boldsymbol{p}_1}\|\tilde{\boldsymbol{p}}_1\|_1 + \tilde{\boldsymbol{q}}_1, \\ 0 \ni \lambda\partial_{\boldsymbol{p}_2}\|\tilde{\boldsymbol{p}}_2\|_1 + \tilde{\boldsymbol{q}}_2, \end{cases} \tag{3.22}$$

$$\tilde{z} = \mathcal{A}\tilde{u}, \tag{3.23}$$

$$\tilde{\boldsymbol{p}} = \nabla\tilde{u}, \tag{3.24}$$

where $\tilde{u} = \tilde{u}_1 + \mathbf{i}\tilde{u}_2$, $\tilde{\boldsymbol{p}} = \tilde{\boldsymbol{p}}_1 + \mathbf{i}\tilde{\boldsymbol{p}}_2$ and $\tilde{\boldsymbol{q}} = \tilde{\boldsymbol{q}}_1 + \mathbf{i}\tilde{\boldsymbol{q}}_2$.

**Theorem 2.** *Assume that the successive differences of the two multiplies* $\{v^j - v^{j-1}\}, \{\boldsymbol{q}^j - \boldsymbol{q}^{j-1}\}$ *converge to zero and* $\{u^j\}$ *is bounded, then there exists a subsequence of iterative sequences of Algorithm I converging to an accumulation point that satisfies KKT conditions of the saddle point problem* (3.2).

*Proof.* We complete the proof in two steps. First, we show the boundedness of all the variables. Due to the update of two multipliers (3.18) and the assumption that their successive differences converge, one can derive that

$$\lim_{j\to+\infty} z^j - \mathcal{A}u^j = 0, \ \lim_{j\to+\infty} \boldsymbol{p}^j - \nabla u^j = 0, \tag{3.25}$$

which implies the boundedness of $\{z^j\}$ and $\{\boldsymbol{p}^j\}$. By (3.16), we have

$$|z^{j+1}(i)| = \begin{cases} \dfrac{r_1|w^j(i)| + \sqrt{r_1^2|w^j(i)|^2 + 4(1+r_1)f(i)}}{2(1+r_1)}, & \forall\ i \in D, \\ |w^j(i)|, & \forall\ i \in \Lambda \setminus D, \end{cases}$$

which demonstrates that $\{w^j\}$ is bounded and so is $\{v^j\}$ since $w^j = \mathcal{A}u^{j+1} - v^j/r_1$. By (3.17), we have

$$\begin{aligned} &|\boldsymbol{p}^{j+1}| = \max\left\{0, |\nabla u^{j+1} - \boldsymbol{q}^j/r_2| - \lambda/r_2\right\} \\ &\geq |\nabla u^{j+1} - \boldsymbol{q}^j/r_2| - \lambda/r_2 \geq |\boldsymbol{q}^j|/r_2 - |\nabla u^{j+1}| - \lambda/r_2, \end{aligned} \tag{3.26}$$

which gives the boundedness of $\{\boldsymbol{q}^j\}$ due to the boundedness of $\{u^j\}$ and $\{\boldsymbol{p}^j\}$.

The boundedness of all variables guarantees that there exists a subsequence $\{(u^{j_l}, \boldsymbol{p}^{j_l}, z^{j_l}, v^{j_l}, \boldsymbol{q}^{j_l})\} \subset \{(u^j, \boldsymbol{p}^j, z^j, v^j, \boldsymbol{q}^j)\}$ and $(\tilde{u}, \tilde{\boldsymbol{p}}, \tilde{z}, \tilde{v}, \tilde{\boldsymbol{q}})$, such that

$$\lim_{l\to+\infty} (u^{j_l}, \boldsymbol{p}^{j_l}, z^{j_l}, v^{j_l}, \boldsymbol{q}^{j_l}) = (\tilde{u}, \tilde{\boldsymbol{p}}, \tilde{z}, \tilde{v}, \tilde{\boldsymbol{q}}).$$

We then prove that the accumulation point $(\tilde{u}, \tilde{\boldsymbol{p}}, \tilde{z}, \tilde{v}, \tilde{\boldsymbol{q}})$ satisfies the KKT conditions. It follows from (3.25) that the KKT conditions *w.r.t.* $z$ and $\boldsymbol{p}$, *i.e.*, (3.23) and (3.24), are satisfied. Since $\mathcal{B}$ is a linear operator in a finite dimensional space, (3.20) is straightforward. By the continuity of (3.16), (3.21) is obtained. Finally one can obtain that $\tilde{\boldsymbol{p}} = \text{Thresh}_{\lambda/r_2}(\nabla\tilde{u} - \tilde{\boldsymbol{q}}/r_2)$, which implies (3.22). Hence the proof is completed. □

Although the assumption for the convergence of the successive difference of multipliers in Theorem 2 seems strong, we observe through our numerical results that

the proposed algorithm is always convergent; see Figure 11. Theoretically speaking, it is challenging to prove the convergence of ADMM for a nonconvex nondifferential optimization problem without any Lipschitz differentiable term. As for the boundedness assumption in Theorem 2, it can be removed for image reconstruction as its lower/upper bounds are often known *a priori* and a box constraint can be enforced in the model [33, 16].

**4. A special case.** In the case of noiseless data, we consider a special case of TV-PoiPR by setting $\lambda = 0$ in (2.6), referred to as "PoiPR",

$$\min_{u\in\mathbb{C}^n} \mathcal{H}(u) := \frac{1}{2}\sum\nolimits_{i\in D}(|(\mathcal{A}u)(i)|^2 - f(i)\log|(\mathcal{A}u)(i)|^2). \tag{4.1}$$

We remark that this model can also be used for PR with noisy measurements, which produces noisy reconstructed results. The existence of the solution has been given in Theorem 1 and we can further prove that the solution to (4.1) with noiseless data is unique under some conditions.

**Theorem 3.** *For $f \in \mathrm{Range}(\mathcal{M})$ and $D = \Lambda$, if the phase retrieval problem has a unique solution up to global phase shifts, then the minimizer to (4.1) is also unique up to global phase shifts.*

*Proof.* We consider a minimization problem *w.r.t.* $w$ as follows,

$$w^* = \min_{w(i)\in\mathbb{R}_+} \frac{1}{2}\sum\nolimits_{i\in\Lambda}(w(i) - f(i)\log w(i)), \tag{4.2}$$

which can be solved pointwisely. Particularly for each entry $w(i)$, the minimization problem has a unique solution, *i.e.*, $w^*(i) = \min_{\rho\in\mathbb{R}_+}\left\{\rho - f(i)\log\rho\right\} = f(i)$. Therefore $f$ is the unique minimizer to (4.2). Furthermore, since $f \in \mathrm{Range}(\mathcal{M})$, there exists a vector $u \in \mathbb{C}^n$ such that $f = \mathcal{M}(u)$. Since $\mathcal{M}(u)$ admits a unique solution up to global phase shifts, the solution to (4.1) is also unique up to global phase shifts. ∎

Generally speaking, it is a strong assumption that the phase retrieval problem has a unique solution up to global shift. It follows from the work of Conca *et al.* [20] that if one defines a linear operator $\mathcal{A}_\Phi$ generated by an $m-$element complex *frame* $\Phi = \{\phi_0, \phi_1, \cdots, \phi_{m-1}\}$ in the sense of $(\mathcal{A}_\Phi u)_j = \langle\phi_j, u\rangle$, then the mapping $\mathcal{M}(u)$ is injective for a generic frame $\Phi$ with $m \geq 4n - 4$. However, it is usually difficult to verify that the frame $\Phi$ is generic. In our experiments, we consider Fourier type of measurements, which are often not generic. Fortunately, the uniqueness of Fourier measurements can be proven regardless of generic frame. For example, in the case of oversampled[3] CDP, Fannjiang [24] proved that the solution to one pattern for real-valued images (two for complex-valued) is unique with overwhelming probability; in standard CDP defined in (3.10) with the *admissible modulation*, Candés *et al* [11] proved that with probability at least $1 - 1/n$, the solution for PR is unique up to global phase shift using $Jn$ measurements, if the masks number $J \geq c\log^4 n$ for some constant $c$. For deterministic holographic pattern in (1.2), $u$ can be recovered up to global phase if the DFT of $u$ is nonvanishing, $s_i$ is prime with $n_i$, and $n_1$ is prime with $n_2$. In [16], by setting $s_1 = s_2 = 1/2$ in (1.2), the uniqueness can be obtained for real-valued images if the DFT of $D^s u$ and $u$ are nonvanishing and $n_1$ is prime with $n_2$.

[3] "Oversampled" means using the oversampled DFT on an enlarged grid.

Computationally, one can construct the augmented Lagrangian for PoiPR, similarly to (3.2),

$$\max_{v}\min_{u,z}\frac{1}{2}\sum\nolimits_{i\in D}(|z(i)|^2 - f(i)\log|z(i)|^2) + \mathrm{Re}(\langle z-\mathcal{A}u, v\rangle) + \frac{r}{2}\|z-\mathcal{A}u\|^2. \tag{4.3}$$

Following a similar procedure of solving TV-PoiPR, we obtain an ADMM-based algorithm for PoiPR. The pseudo-code is summarized in Algorithm II, which has much simpler form compared to Algorithm I (for solving TV-PoiPR).

---

Algorithm II: ADMM for PoiPR (4.1)

---

1. Initialization: Set $v^0 = 0$, $u^0, z^0 = \mathcal{A}u^0$ and $j = 0$.
2. Solve $u^{j+1} = u_1^{j+1} + \mathbf{i}u_2^{j+1}$ by solving the following equations as

$$\begin{bmatrix} \mathrm{Re}(\mathcal{A}^*\mathcal{A}) & -\mathrm{Im}(\mathcal{A}^*\mathcal{A}) \\ \mathrm{Im}(\mathcal{A}^*\mathcal{A}) & \mathrm{Re}(\mathcal{A}^*\mathcal{A}) \end{bmatrix} \begin{bmatrix} u_1^{j+1} \\ u_2^{j+1} \end{bmatrix} = \begin{bmatrix} \mathrm{Re}(\mathcal{A}^*v_0^j) \\ \mathrm{Im}(\mathcal{A}^*v_0^j) \end{bmatrix}, \tag{4.4}$$

   with $v_0^j = z^j + v^j/r$.
3. Solve $z^{j+1}$ by (3.16).
4. Update the multiplier as

$$v^{j+1} = v^j + r(z^{j+1} - \mathcal{A}u^{j+1}). \tag{4.5}$$

5. If the some stopping condition is satisfied, stop the iterations and output the iterative solution; else set $j = j + 1$, and goto Step 2.

---

Remark 4.1. *We want to reveal the equivalence between ADMM and Douglas-Rachford splitting algorithm (DRS) [22] applied to PoiPR. Assuming that $D = \Lambda$, we reformulate the proposed PoiPR (4.1) to an equivalent form,*

$$\min\nolimits_{z\in\mathbb{C}^m} \Phi(z) + \mathbb{I}_{\mathscr{K}}(z). \tag{4.6}$$

*where $\Phi(z) := \frac{1}{2}\sum\limits_i(|z(i)|^2 - f(i)\log|z(i)|^2)$ and $\mathbb{I}_{\mathscr{K}}$ denotes the indicator function with $\mathscr{K} = \{z \in \mathbb{C}^m : \exists u \in \mathbb{C}^n, s.t.\ \ z = \mathcal{A}u\}$. We introduce a proximal operator $\mathrm{Prox}_\Phi$ as $\mathrm{Prox}_\Phi(z) = \arg\min_{\bar{z}} \Phi(\bar{z}) + \frac{1}{2}\|\bar{z} - z\|^2$ and assume $\mathrm{Im}(\mathcal{A}^*\mathcal{A}) = 0$. Then it follows from Remark 3.1 and (3.16) that applying DRS to (4.6) can be expressed as*

$$z^{k+1} = z^k - \bar{z}^k + \mathrm{Prox}_{\mathbb{I}_{\mathscr{K}}}(2\bar{z}^k - z^k), \quad \bar{z}^k = \mathrm{Prox}_{\Phi/r_3}(z^k), \tag{4.7}$$

*with a positive parameter $r_3$ and*

$$\mathrm{Prox}_{\mathbb{I}_{\mathscr{K}}}(z) = \mathcal{A}(\mathcal{A}^*\mathcal{A})^{-1}(\mathcal{A}^*z),$$

$$(\mathrm{Prox}_{\Phi/r_3}(z))(i) = \frac{r_3|z(i)| + \sqrt{r_3^2|z(i)|^2 + 4(1+r_3)f(i)}}{2(1+r_3)}\mathrm{sign}(z(i)),\ \forall i \in \Lambda.$$

*By introducing new variables and switching orders of subproblems*[4]*, we can get an equivalent form of* (4.7) *as*

(4.8)
$$z^{k+1} = \mathrm{Prox}_{\mathbb{I}_{\mathscr{K}}}(x^k + \bar{v}^k),\ x^{k+1} = \mathrm{Prox}_{\Phi/r_3}(z^{k+1} - \bar{v}^k),\ \bar{v}^{k+1} = \bar{v}^k + z^{k+1} - x^{k+1},$$

*which is exactly the ADMM for solving an equivalent problem of* (4.6) *as*

$$\min_{x,z} \Phi(x) + \mathbb{I}_{\mathscr{K}}(z), \quad s.t.\ x = z. \tag{4.9}$$

*Obviously, Eqn.* (4.8) *by replacing* $u$ *by* $z = \mathcal{A}u$ *has the same form as Algorithm II, which implies the equivalence between DRS and ADMM applied to PoiPR.*

Similar theoretical results of PoiPR can be obtained from the analysis of TV-PoiPR. In particular, we can prove that there exists a unique solution of $u-$subproblem (4.4) and Algorithm II converges to a stationary point, as characterized in Proposition 3 and Theorem 4 respectively. The proofs are omitted here.

**Proposition 3.** *Assume that the operator* $\mathrm{Re}(\mathcal{A}^*\mathcal{A})$ *is positive, then the linear equations* (4.4) *admit a unique solution.*

Remark 4.2. *Note that the assumption in Proposition 3 holds for the Fourier measurements* (3.10) *without undersampling, since each entry of* $\sum_j I_j^* \circ I_j$ *is nonzero. Note that it also holds for holographic pattern* (1.2)*. For ptychographic pattern, if one measures the data with sufficient number of frames under the periodical boundary condition, the operator* $\mathrm{Re}(\mathcal{A}^*\mathcal{A})$ *is also positive.*

**Theorem 4.** *Assume that the successive differences of the multiplier* $\{v^j - v^{j-1}\}$ *converge to zero and* $\{u^j\}$ *is bounded, then there exists a subsequence of iterative sequences of Algorithm I converging to an accumulation point that satisfies KKT conditions of the saddle point problem* (4.3)*.*

In the rest of the section, we consider to accelerate Algorithm II in the case of a sufficiently large number of CDP measurements. Denote individual masked measurement matrices $\mathcal{A}_j$ and data $f_j$ such that $\mathcal{A} = (\mathcal{A}_1^T, \mathcal{A}_2^T, \cdots, \mathcal{A}_J^T)^T$, and $f = (f_1^T, f_2^T, \cdots, f_J^T)^T$ with $\mathcal{A}_j \in \mathbb{C}^{n\times n}, f_j \in \mathbb{R}^n$. By introducing the operators $\mathcal{T}_j^{\mathrm{AP}}$ as $\mathcal{T}_j^{\mathrm{AP}}(u) := \mathcal{A}_j^*(\sqrt{f_j} \circ \mathrm{sign}(\mathcal{A}_j u))$, serial alternating projection (SAP) [18] is given as

$$u^{k+1} = \left(\prod_{j=1}^{J} \mathcal{T}_j^{\mathrm{AP}}\right) u^k, \tag{4.10}$$

where $\mathcal{A}_j$ is assumed to be isometric, *i.e.* $\mathcal{A}_j^*\mathcal{A}_j = \mathbf{I}$ with identity operator $\mathbf{I}$. SAP is essentially an alternating projection algorithm to find the intersection $u^\star$ of multiple sets, *i.e.* $u^\star \in \bigcap_{j=1}^J \{u \in \mathbb{C}^n :\ |\mathcal{A}_j u| = f_j\}$.

For simplicity, we assume that $\mathcal{D} = \Lambda$ and $\mathcal{A}^*\mathcal{A}$ is invertible (not necessarily to be isometric). Motivated by the SAP and discussions in Remark 4.1, we heuristically establish a variant of Algorithm II, referred to as "Algorithm III", as follows:

$$\text{Algorithm III}: \qquad u^{k+1} = \left(\prod_{j=1}^{J} \mathcal{T}_j\right) u^k, \tag{4.11}$$

where

$$\mathcal{T}_j u := (\mathcal{A}_j^*\mathcal{A}_j)^{-1}\mathcal{A}_j^*(\mathcal{A}_j u - \bar{z}_j + \mathrm{Prox}_{\mathbb{I}_{\mathscr{K}_j}}(2\bar{z}_j - \mathcal{A}_j u)), \quad \bar{z}_j = \mathrm{Prox}_{\Phi_j/r_3}(\mathcal{A}_j u),$$

[4] http://www.seas.ucla.edu/~vandenbe/236C/lectures/dr.pdf

and $\mathscr{K}_j = \{z : \ \exists u \in \mathbb{C}^n, \ s.t. \ z = \mathcal{A}_j u\}$ and $\Phi_j(z) = \frac{1}{2}\sum_{i\in\Omega}(|z(i)|^2 - f_j(i)\log|z(i)|^2)$. One readily knows that Algorithm III is derived based on DRS (4.7) (or equivalently ADMM in Algorithm II) and SAP. As demonstrated in Section 5, not only does Algorithm III accelerate Algorithm II when a large number of noiseless CDP measurements are available, it also yields better recovery quality than the SAP for noisy CDP measurements.

**5. Numerical experiments.** Although the proposed approaches are applicable to phase retrieval in general, we only focus on Fourier measurements in the experimental section. In particular, we consider three different types of linear operators $\mathcal{A}$: coded diffraction pattern (CDP) with random masks, holographic pattern [13] with deterministic masks, and ptychgraphic pattern [62, 51], the last of which is currently a very popular and powerful imaging tool to generate high resolution images with large field of view. For coded diffraction pattern, we use octanary CDPs; specifically each element of $I_j$ in (3.10) takes a value randomly among the eight candidates, *i.e.*, $\{\pm\sqrt{2}/2, \pm\sqrt{2}\mathbf{i}/2, \pm\sqrt{3}, \pm\sqrt{3}\mathbf{i}\}$. For holographic pattern, the linear operator $\mathcal{A}$ is given in (1.2) and we choose $s_1 = s_2 = 1/2$, which is shown in our previous work [16] to give better results with less measurements than $s_1, s_2$ taking integer values as in [10]. For ptychographic pattern, the operator $\mathcal{A}$ is defined as follows,

$$\mathcal{A}u = \begin{bmatrix} \mathcal{F}(\omega \circ R_1 u) \\ \cdots \\ \mathcal{F}(\omega \circ R_j u) \\ \cdots \\ \mathcal{F}(\omega \circ R_J u) \end{bmatrix}, \tag{5.1}$$

where $R_j$ is a (binary) restriction matrix that selects a window for an index $j \in \{1, \cdots, J\}$ with $J = 16 \times 16$ frames in total and $\omega$ denotes the illumination mask generated by a zone plate len with size of $64 \times 64$. We assume that $R_j$ and $\omega$ are known in this paper; for unknown $\omega$ please refer to a recent work [33].

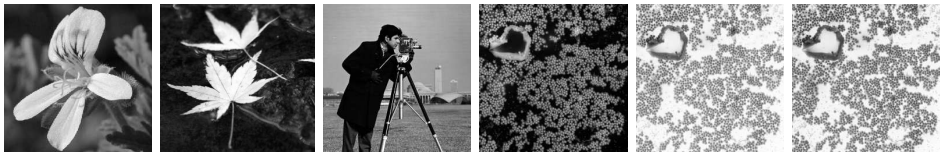
The testing images shown in Figure 1 include three real-valued images: "Flower" with resolution $256 \times 256$, "Leaf" with resolution $363 \times 378$, and "Cameraman" with resolution $256 \times 256$, as well as a complex-valued image "Goldballs"with resolution $256 \times 256$. As discussed in Section 2, the amount of Poisson noise at each pixel depends on its intensity value. Therefore, we introduce a scale factor $\eta \in (0, \infty)$ to control the scale of the image intensities (or the number of photons), which is inversely proportional to the amount of noise added to the data. Let $u_\eta = \eta u$, and the measured data is expressed as $f(i) \overset{\text{ind.}}{\sim} \text{Poisson}(|(\mathcal{A}u_\eta)(i)|^2), \ \forall i \in D$.

Quantitatively, we use Signal-Noise-Ratio (SNR) to measure the reconstruction quality, defined as

$$\text{SNR}(u, u_g) = -10\log_{10}\frac{\sum_{i\in\Omega}|u(i) - c^* u_g(i)|^2}{\sum_{i\in\Omega}|u(i)|^2},$$

where $u_g$ is the ground truth image of size $n$, $u$ is the reconstructed image, and $c^*$ denotes the global phase factor as $c^* = \arg\min_{\{c\in\mathbb{C}: \ |c|=1\}}\|u - cu_g\|$. We also use "SNR-Intensity" to measure the noise level in the measurement domain. Particularly, we look at SNR-Intensity of the measurement data, defined as $\text{SNR}(f, |\mathcal{A}u_g|^2)$, and SNR-Intensity of the recovered image $u$, defined as $\text{SNR}(|\mathcal{A}u|^2, |\mathcal{A}u_g|^2)$.

**5.1. Effectiveness of TV regularization.** We first show the effectiveness of TV regularization by comparing Algorithm I (with TV) and Algorithm II (without



(a) (b) (c) (d) (e) (f)

FIG. 1. *Testing images. Real-valued images: (a) "Flower" with resolution* $256 \times 256$*, (b) "Maple" with resolution* $363 \times 378$*, and (c) "Cameraman" with resolution* $256 \times 256$*. A complex-valued image "Goldballs" with resolution* $256 \times 256$*: (d) its amplitude, (e) real components, and (f) imaginary components.*

TABLE 1

*SNRs of image reconstructed by different algorithms. The data format is SNR/SNR-Intensity, "NULL" means the raw data (hence no SNR values), and "I", "II" are short for Algorithm I and Algorithm II.*

| $\eta$ | Algorithms | Flower | Maple | Cameraman |
|---|---|---|---|---|
| 0.01 | NULL | ——/8.34 | ——/7.19 | ——/6.57 |
| | II | 4.41/8.02 | 3.66/8.10 | 4.79/8.33 |
| | I (TV) | 19.94/19.98 | 18.93/18.74 | 17.99/18.41 |
| 0.02 | NULL | ——/11.57 | ——/12.53 | ——/11.82 |
| | II | 10.07/13.89 | 11.09/14.92 | 10.37/14.17 |
| | I (TV) | 22.95/23.48 | 22.75/23.08 | 20.74/21.62 |
| 0.05 | NULL | ——/19.20 | ——/20.32 | ——/19.57 |
| | II | 18.58/22.08 | 19.64/23.19 | 18.88/22.45 |
| | I (TV) | 27.18/28.27 | 27.67/28.72 | 24.96/26.55 |

TV) for phase retrieval from three types of Fourier measurements corrupted by Poisson noise. For both algorithms, we choose the initial condition $u^0$ by assuming zero phases and solving it from $\mathcal{A}u^0 = \sqrt{f}$, i.e.,

$$u^0 = (\mathrm{Re}(\mathcal{A}^*\mathcal{A}))^{-1}(\mathcal{A}^*\sqrt{f}), \tag{5.2}$$

and terminate the algorithms after 50 iterations as a default stopping condition. The choices of parameters $(\lambda, r, r_1, r_2)$ are provided in the caption of resulting figures. Please refer to Section 5.3 for detailed discussions about the influence of parameters on the performance of the proposed algorithms.

We start with phase retrieval of real-valued images from CDP measurements with $J = 2$ in (3.10). In this setting, the operator $\mathcal{M}$ takes $2n$ measurements. We consider $\eta = 0.01, 0.02, 0.05$ to generate three different scales of ground truth images in the sense that the smaller $\eta$ is, the more noisy the image exhibits. We record the SNR/SNR-Intensity values for all the testing images in Table 1, which shows at least 8dB/4dB improvement of Algorithm I (with TV) over Algorithm II (without TV). Due to page limit, we only present the PR results of "Cameraman" in Figure 2. Readily, one can infer the TV's effectiveness from the reconstructed images having sharper edges, cleaner background, and higher contrast than the non-TV version. Improvement of TV over non-TV is more obvious for smaller $\eta$, which corresponds to larger amount of noise in the data or lower amount of photon counts.

We then consider phase retrieval of the complex-valued image "Goldballs" (in Figure 1 (d)-(f)). We observe empirically that $J = 4$ in (3.10) often gives the robust recovery for complex data, and hence the experiments are based on $J = 4$. Figure

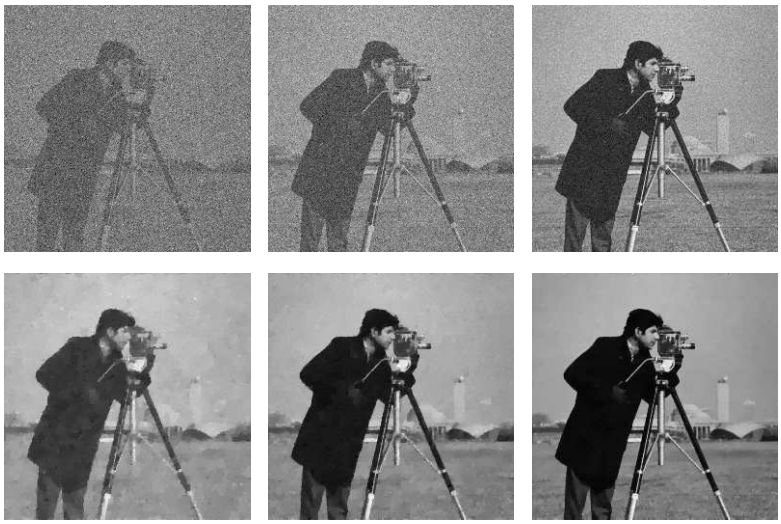


FIG. 2. *Comparison of Algorithm I (with TV, top row) and Algorithm II (without TV, bottom row) to recover real-valued "Cameraman" from noisy CDP measurements. From left to right:* $\eta = 0.01, 0.02, 0.05$ *(the smaller* $\eta$ *is, the more noisy the image exhibits.) We fix the parameters* $r_2 = 5 \times 10^4, \lambda = 200$ *for different* $\eta$*; set* $r = r_1 = 5$ *for* $\eta = 0.01$ *and* $r = r_1 = 2$ *for* $\eta = 0.02, 0.05$.

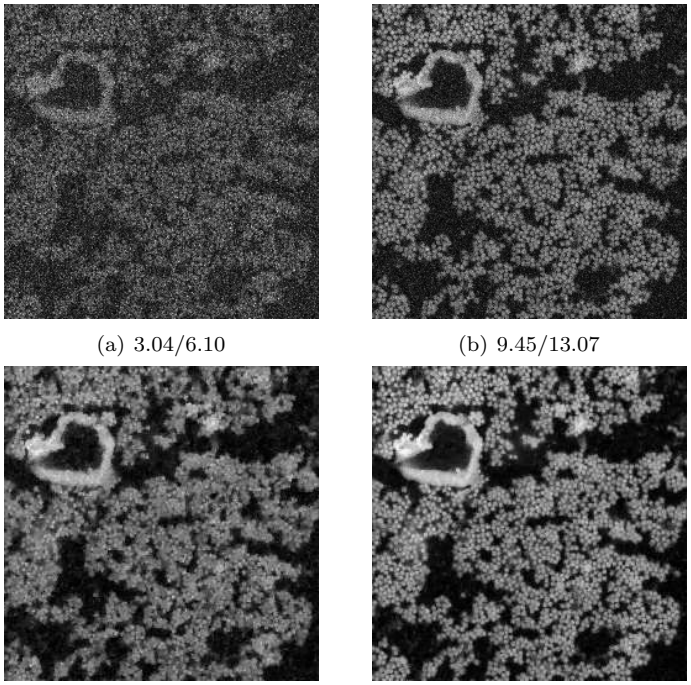
3 shows the phase retrieval results, which again illustrates the improvement of using TV regularization over non-TV by at least 6dB/3dB increase of SNR/SNR-Intensity values.

For holographic patterns with $s_1 = s_2 = 1/2$ in (1.2), it seems to require more iterations for such deterministic masks than CDP's random masks; and therefore we stop Algorithms I-II after 500 iterations. The results are given in Figure 4. Without TV, Algorithm II produces a lot of ringing artifacts and distortions, which are different to homogeneous noise in the reconstructions from CDP measurements as shown in Figures 2-3. With the help of TV, these severe artifacts inherited from noisy measurements are removed with at least 4dB increase of SNR compared to the non-TV, although no obvious increase in the SNR-Intensity values.

Finally, we perform the experiments for ptychographic phase retrieval using the complex-valued image "Goldballs." As shown in Figure 5, ptychographic pattern is more challenging than CDP in that the non-TV algorithm gives very blurry results for both small scale features and the edges. By using the TV regularization, the edges of large scale features are well recovered in Figure 5 (b), although small scale structures still looks blurry. When the noise level decreases, the TV can recover the images with sharp edges and repetitive small scale structures as shown in Figure 5 (d). The SNR/SNR-Intensity of recovery images are increased at least 4dB/5dB by using the TV regularization.

**5.2. Oversampling and undersampling.** We examine the performance of the proposed algorithms *w.r.t.* oversampling in the sense that one collects noisy data with different number of masks, *i.e.* $J = 1, 3, 5$, which acts as an oversampling factor. Figure 6 shows that the more data we have (larger $J$), the better the phase retrieval results by either Algorithm I or Algorithm II. When comparing results across different $J$ values, there is a diminishing gain of using the TV regularization in terms of SNR as $J$ increases.

We also study the effects of undersampling, the amount of which can be controlled



(a) 3.04/6.10 (b) 9.45/13.07

(c) 10.66/10.91 (d) 15.64/16.76

FIG. 3. *Comparison of Algorithm I (with TV, top row) and Algorithm II (without TV, bottom row) to recover a complex-valued image "Goldballs" from noisy CDP measurements. From left to right:* $\eta = 0.08, 0.2$ *with the corresponding SNR-Intensity values of noisy measurements as 4.64 and 11.07 respectively. Set* $\lambda = 200, r_1 = r = 5, r_2 = 5.0 \times 10^4$ *for different noise levels. The data format is SNR/SNR-Intensity.*

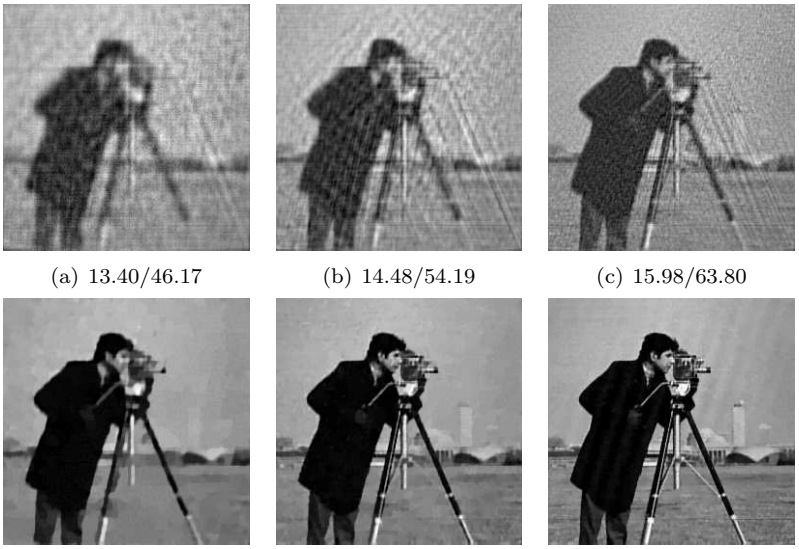
by the subset $D$. In particular, we define the undersampling ratio as $r_s = |D|/m$. The sampling subset $D$ is randomly generated according to a probability density function with polynomial variable density sampling[5]; please refer to [44] for details about various sampling schemes. We show one realization of sampling masks corresponding to $r_s = 0.1, 0.2, 0.4$ in Figure 7. Due to page limits, we only provide the PR results of "Cameraman" with $J = 2, \eta = 0.05$ in Figure 8, while SNRs of all the testing images are given in Table 2. Both Figure 8 and Table 2 show large improvements of Algorithm I over Algorithm II both visually and in terms of SNRs. Figure 8 also demonstrates that TV can give satisfactory results at 20% undersampling ratio. As $J = 2$, the number of measurements at $r_s = 20\%$ is $0.4n$, which is below the theoretical limit capped at $2n$ for real signal recovery.

**5.3. Discussion on Algorithms I-II.** We want to illustrate some properties of the proposed algorithms, specifically impact by parameters and convergent behaviors. All the results in this section are based on numerical simulation with Poisson noise level at $\eta = 0.02$.

**Impact by parameters.** The parameters $\lambda, r_1, r_2$ in Algorithm I and $r$ in Algorithm II are chosen by hand in order to obtain visually satisfactory results.

We first discuss the impact of $r$ for Algorithm II by choosing $r$ from $\{r^0 \times 2^{-l}, r^0 \times 2^{-l+1}, \cdots, r^0 \times 2^{l-1}, r^0 \times 2^l\}$ with $l = 7$, $r_0 = 2$. We plot the corresponding SNRs of the

[5]SparseMRI V0.2, http://www.eecs.berkeley.edu/~mlustig/Software.html



(a) 13.40/46.17 (b) 14.48/54.19 (c) 15.98/63.80

(d) 17.44/47.40 (e) 20.25/55.12 (f) 23.19/63.92

FIG. 4. *Comparison of Algorithm I (with TV, top row) and Algorithm II (without TV, bottom row) to recover "Cameraman" from noisy measurements of holographic patterns. Left column:* $\eta = 0.01, r = 0.3, r_1 = 5, r_2 = 2 \times 10^5, \lambda = 200$*; middle column:* $\eta = 0.02, r = 0.4, r_1 = 3, r_2 = 8 \times 10^4, \lambda = 140$*; right column:* $\eta = 0.05, r = 0.1, r_1 = 1, r_2 = 5 \times 10^4, \lambda = 100$*; the corresponding SNR-Intensity values for* $\eta = 0.01, 0.02, 0.05$ *are 44.92, 53.57, and 58.00 respectively. The data format is SNR/SNR-Intensity.*

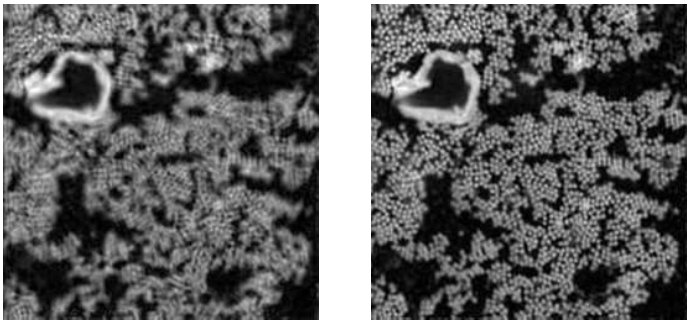
TABLE 2

*Downsampling: SNRs for the recovery images and the corresponding intensities by different algorithms. "NULL" means the raw data with no operations. The data format is SNR/SNR-Intensity, and "I", "II" are short for Algorithm I and Algorithm II.*

| $r_s$ | Algorithms | Flower | Maple | Cameraman |
|---|---|---|---|---|
| 10% | NULL | ——/-9.43 | ——/-9.54 | ——/-9.47 |
| | II | -1.94/-5.69 | -2.83/-6.39 | -1.88/-5.75 |
| | I (TV) | 15.21/15.00 | 7.72/6.08 | 13.73/13.46 |
| 20% | NULL | ——/-6.07 | ——/-6.01 | ——/-5.94 |
| | II | 0.81/0.57 | -0.56/-0.54 | 0.94/0.65 |
| | I (TV) | 21.14/21.42 | 11.62/11.17 | 18.44/18.99 |
| 40% | NULL | ——/-1.81 | ——/-1.73 | ——/-1.77 |
| | II | 3.45/6.31 | 1.43/3.96 | 3.78/6.42 |
| | I (TV) | 24.19/24.88 | 24.41/25.13 | 21.36/22.37 |

reconstructed images in Figure 9, where we include the results of running Algorithm II 1000 iterations (red plus) versus default 50 iterations (blue dots). It seems that the parameter $r$ only affects the convergence rate and Algorithm II is rather insensitive to $r$ in the range of $[1, 100]$.

As for Algorithm I, the impact of $r_1, r_2$, and $\lambda$ is illustrated in Figure 10, in which we fix $\lambda \in \{100, 200, 400\}$ and vary the parameters $(r_1, r_2) \in \{r_1^0 \times 2^{-l_1}, r_1^0 \times 2^{-l_1+1}, \cdots, r_1^0 \times 2^{l_1-1}, r_1^0 \times 2^{l_1}\} \times \{r_2^0 \times 2^{-l_2}, r_2^0 \times 2^{-l_2+1}, \cdots, r_2^0 \times 2^{l_2-1}, r_2^0 \times 2^{l_2}\}$ with $r_1^0 = 2, r_2^0 = 5 \times 10^4$, and $l_1 = l_2 = 7$. For the sake of better visualization, we raise the negative SNR values to zeros in Figure 10, which shows that Algorithm I is less sensitive to $\lambda$ than to $r_1, r_2$. The first two rows of Figure 10 are about different iter-



(a) 8.37/31.00 (b) 15.12/36.90

(c) 11.22/36.22 (d) 17.06/41.55

FIG. 5. *Comparison of Algorithm I (with TV, top row) and Algorithm II (without TV, bottom row) to recover 'Goldballs" from noisy measurements of ptychographic patterns. The resulting magnitudes are shown here and the data format is SNR/SNR-Intensity. From left to right:* $\eta = 0.5, 0.8$ *with corresponding SNR-Intensity values of the noisy measurements as 29.48, 33.59. Set* $\lambda = 10, r_1 = 0.5, r_2 = 100$ *for Algorithm I and* $r = 0.08$ *for Algorithm II.*

ations numbers, which implies that the parameters $r_1, r_2$ affect the convergence rates of Algorithm I (some combinations make Algorithm I divergent.) We also show the best results among various combinations of $r_1, r_2$ for each $\lambda = 100, 200, 400$ in Figure 10. One can see that large $\lambda$ leads to over-smoothed image recovery, and hence a moderate $\lambda$ should be chosen for the best results. In addition, we observe heuristically that Algorithm II is less sensitive to parameters than Algorithm I, as the TV regularization in Algorithm I introduces a non-differential term, which mysteriously interacts with the non-convex fidelity term.

In summary, for TV-PoiPR model, $\lambda$ is the model parameter to control the quality of recovery results, while $r_1$ and $r_2$ are algorithmic parameters that determine whether the algorithm converges and/or how fast it converges. We observe through various experiments, such as Figures 2-3 for CDP and Figure 5 for ptychographic pattern, Figure 6 for different number of measurements of CDP, that Algorithm I is relatively insensitive to different images and problem sizes for a specific pattern; and we nearly use the same $\lambda$ for different images and even for different noise level. As for downsampling of CDP (Figure 8), we find a bigger value of $\lambda$ is desired when noise level increases or number of measurements decreases. As shown in Figure 2 and Figure 4, the parameter $r_1$ should be smaller when noise level decreases, and it can be fixed for different images or different number of measurements. The parameter $r_2$ should be smaller as the noise level increase for holographic pattern as shown in

(a) 4.38/11.98 (b) 14.79/15.96 (c) 17.60/18.19

(d) 18.94/19.16 (e) 22.47/22.55 (f) 23.75/23.76

FIG. 6. *Performances* w.r.t. *oversampling factor* $J = 1, 3, 5$ *(from left to right) with corresponding SNR-Intensity values of noisy measurements as* $11.80, 11.78, 11.81$ *(* $\eta = 0.02$*). First row: results by Algorithm II for PoiPR; Second row: results by Algorithm I for TV-PoiPR. The data format is SNR/SNR-Intensity. Set* $r_1 = 2, r_2 = 5 \times 10^4, \lambda = 200$.

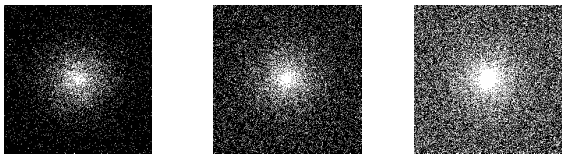


FIG. 7. *Sampling masks* $D$ w.r.t. *undersampling ratios* $r_s = 10\%, 20\%, 40\%$ *from left to right.*

Figure 4, while be kept fixed for other cases *w.r.t.* different images and problem sizes. It is helpful to determine optimal parameters automatically, which will be left as a future work.

**Convergence.** We want to demonstrate the convergence of the proposed algorithms by plotting objective functional values $\mathcal{G}(u^j)$ and $\mathcal{H}(u^j)$ as well as successive errors[6] $\frac{\|u^j - u^{j-1}\|}{\|u^j\|}$ *w.r.t.* the iteration number $j$ in Figure 11, based on "Cameraman" and noise level at $\eta = 0.1$. We consider both CDP and holographic measurements. As illustrated in Figure 11, all the curves are monotonically decreasing, which empirically validate the convergence. In addition, we observe that 30∼50 iterations are sufficient to obtain a satisfactory recovery result from noisy CDP measurements, while more iterations are required for holographic PR to converge. Note that the convergence of Algorithm II for noise free measurements is further examined in Figure 12, in comparison with the other state-of-the-art methods in Section 5.4.

**5.4. Comparison to other PR methods.** We compare our proposed algorithms with six PR methods: error reduction (ER) [28], hybrid projection-reflection

[6]When ground-truth solution for the regularized problem is not available, it is standard to examine the convergence by plotting the objective functional values and the successive errors.

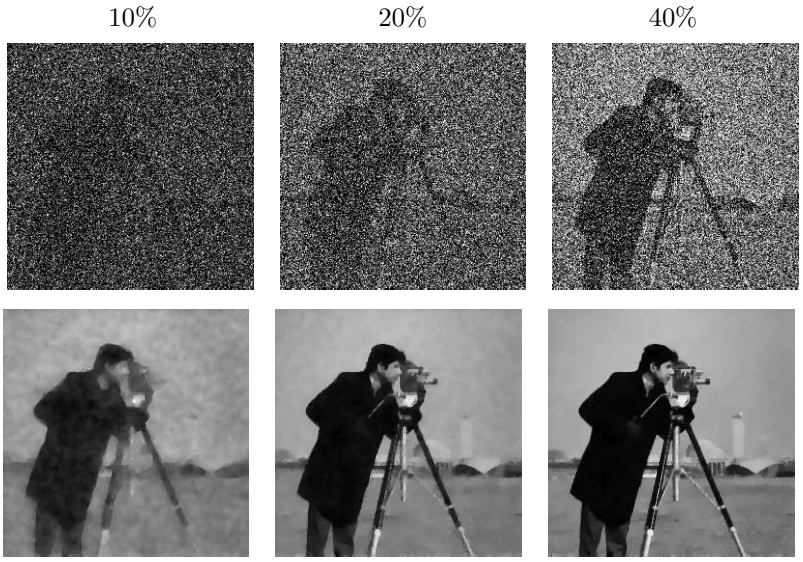



FIG. 8. *Performances* w.r.t. *different sampling masks D shown in Figure 7 (using the same examples in Figure 1 and* $\eta = 0.05$*). 1st row: results by Algorithm II for PoiPR; 2nd rows: results by Algorithm I for TV-PoiPR. From the left to right:* $r_s = 10\%, 20\%, 40\%$*. Set* $r = r_1 = 2, r_2 = 2.0 \times 10^4$*,* $\lambda = 300$ *(*$r_s = 0.1$*) or* $\lambda = 200$ *(*$r_s = 0.2, 0.4$*).*

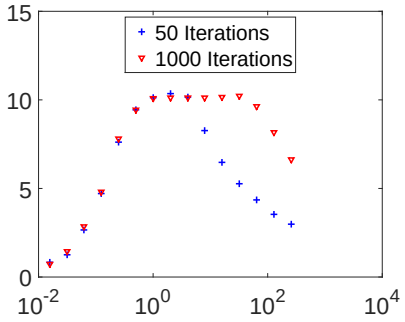



FIG. 9. *The performance of Algorithm II* w.r.t. *r stopped after 50 iterations in blue dots and after 1000 iterations in red plus. The y-axis gives the corresponding SNR values.*

(HIO) [25], difference map (DF) [23], relaxed averaged alternating reflection (RAAR) [42], truncated Wirtinger flow (TWF) [19], and SAP [18]. The Matlab implementation of TWF can be found on authors' website[7], while we implement the other projection algorithms by ourselves. Assuming that $\mathcal{A}\mathcal{A}^*$ is invertible and $\mathrm{Im}(\mathcal{A}\mathcal{A}^*) = 0$, we introduce two projection operators as $\mathcal{P}_M(z) = \sqrt{f} \circ \mathrm{sign}(z)$, and $\mathcal{P}_S(z) = \mathcal{A}(\mathcal{A}^*\mathcal{A})^{-1}\mathcal{A}^* z$, with measurements $f$. The ER algorithm can be written as $z^{k+1} = \mathcal{P}_S\mathcal{P}_M(z^k)$, for $k = 0, 1, \cdots$, and $u^{k+1}$ can be computed as

$$u^{k+1} = (\mathcal{A}^*\mathcal{A})^{-1}\mathcal{A}^* z^{k+1}. \tag{5.3}$$

[7] http://web.stanford.edu/~yxchen/TWF/code.html

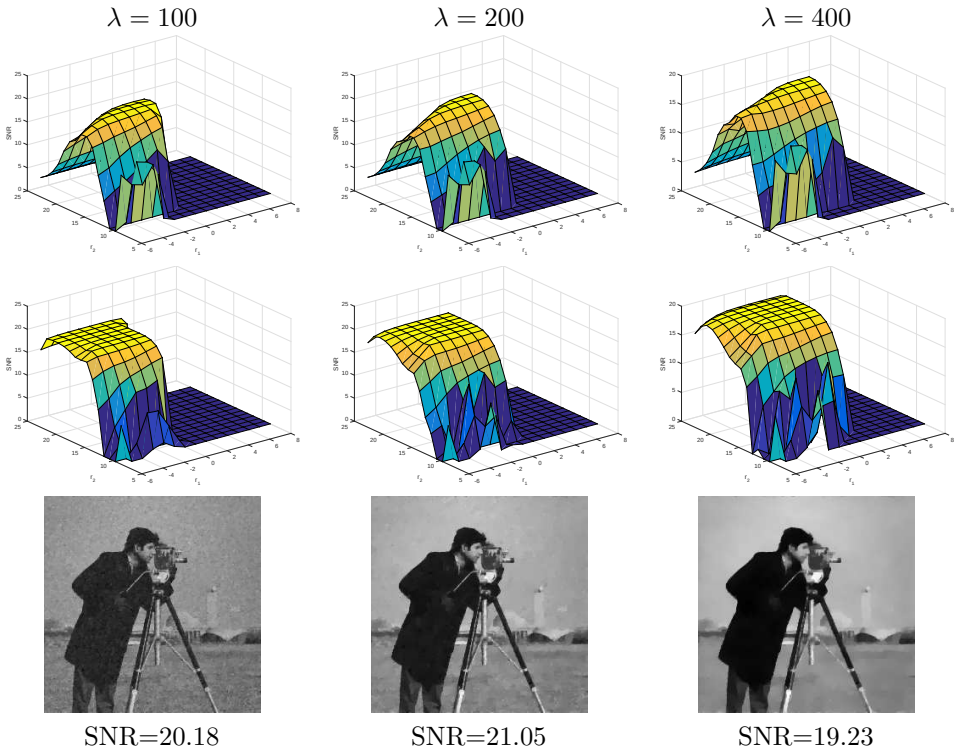



FIG. 10. *The performance of Algorithm I* w.r.t. $r_1, r_2$ *for* $\lambda = 100, 200$, *and 400 (from left to right). 1st-2nd row: the SNR values* w.r.t. $r_1, r_2$ *for each* $\lambda$ *stopped after 50 iterations in 1st row and 1000 iterations in 2nd row and bottom row: the image recovery results corresponding to the highest SNR.*

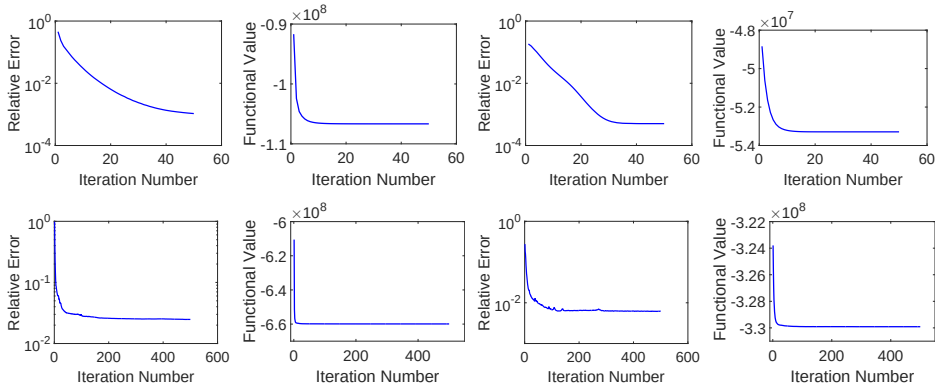



FIG. 11. *Convergence histories of Algorithm I (3rd-4th columns) and Algorithm II (1st-2nd columns) for CDP (top) and holographic pattern (bottom) by plotting objective functional values and relative errors to the ground-truth image* w.r.t. *the iteration number.*

Similarly, we can obtain algorithms for HIO, DF, and RAAR with relaxed parameters $\delta$, $\gamma_1$ and $\gamma_2$ as follows,

(5.4)
$$\begin{cases} \text{HIO:} & z^{k+1} = \left((1+\delta)\mathcal{P}_s\mathcal{P}_M + \mathbf{I} - \mathcal{P}_S - \delta\mathcal{P}_M\right)(z^k), \\ \text{DF:} & z^{k+1} = \left(\mathbf{I} + \delta\big(\mathcal{P}_S((1+\gamma_2)\mathcal{P}_M - \gamma_2\mathbf{I}) - \mathcal{P}_M((1+\gamma_1)\mathcal{P}_S - \gamma_1\mathbf{I})\big)\right)(z^k), \\ \text{RAAR:} & z^{k+1} = \left(2\delta\mathcal{P}_S\mathcal{P}_M + \delta\mathbf{I} - \delta\mathcal{P}_S + (1-2\delta)\mathcal{P}_M\right)(z^k), \end{cases}$$

and $u^{k+1}$ can be immediately solved by (5.3). We remark that for real-valued images, we use the modified projection operator $\widetilde{\mathcal{P}}_S$ as $\widetilde{\mathcal{P}}_S(z) = \mathcal{A}\ \mathrm{Re}((\mathcal{A}^*\mathcal{A})^{-1}\mathcal{A}^*z)$ instead of $\mathcal{P}_S$ and update $u$ as $u^{k+1} = \mathrm{Re}((\mathcal{A}^*\mathcal{A})^{-1}\mathcal{A}^*z^{k+1})$. Since SAP (4.10) requires that each $\mathcal{A}_j$ is isometric and TWF only works for CDP patterns, the following discussion is based on the CDP measurements with each illumination $I_j$ replaced by $\mathrm{sign}(I_j)$ to yield isometric $\mathcal{A}_j$.

Since all the competing methods are not designed to deal with any types of noise, we only compare them with the non-TV version (*i.e.*, PoiPR solved by Algorithm II) in the noise free case. As all the methods can find the ground-truth image from noiseless measurements, we compare them in terms of convergence speed. We use the same initial value (5.2) for all the algorithms except for TWF, which uses a truncated spectral initialization as default and set the stopping condition as

$$\text{Error} := \min_{|c|=1} \|cu^k - u^*\| \leq 1.0 \times 10^{-14},$$

or iteration number reaches 200, where $u^*$ denotes the ground truth image, and $u^k$ is the solution after the $k^{th}$ iteration. We conduct the experiments for CDP measurements of real-valued image "Cameraman" with $J = 2, 12$ and complex-valued image "Goldballs" with $J = 3^8, 12$. We remark that TWF requires more measurements, and we only report its performance with $J = 12$. We plot the convergence curves in Figure 12. For fewer measurements ($J = 2$), Algorithm II converges the fastest for real-valued image, while Algorithm III and SAP converge the fastest for complex-value image. When more measurements are collected with $J = 12$, Algorithm III and SAP converge much faster than others for both real and complex images. We also observe very close convergence speed between Algorithm II and RAAR as well as between Algorithm III and SAP. The relations between these algorithms should be investigated in the future. It seems that the iterative errors of HIO and DF algorithms decrease fast within first ten iterations and then stagnate. In addition, a linear convergence rate is observed practically, which motivates the future study of theoretical analysis on convergence rate.

We then test on noisy CDP measurements with $J = 12$ and $\eta = 0.005$. It seems the iterations of HIO and DF algorithms are unstable, and we only show the best recovery results (with highest SNR) among 50 iterations. We show the recovery results of "Cameraman" in Figure 13 and the SNR/SNR-Intensity values of all the testing images in Table 3. The recovered image by TWF is very noisy, so we exclude it in Figure 13. One can see that Algorithms II-III without TV produce better results visually, which also have higher SNRs than other competing algorithms; and the TV regularization (Algorithm I) significantly improves the results as in Table 3. We also plot the convergence curves in Figure 14, which implies that errors of ER/RAAR decrease fastest within the first ten iterations, followed by stagnation at a higher error level than that by Algorithms I-III. Also the error by Algorithm III decreases more quickly than that by Algorithm II. Based on Figures 13-14 and Table 3, we remark that the proposed Algorithms II-III using Poisson maximum likelihood estimate produce recovered results with higher contrast visually and higher accuracy compared with other non-TV algorithms such as ER/RAAR that do not consider the noise distribution.

**6. Conclusion.** In this paper, we proposed a total variation regularization model "TV-PoiPR" to recover an image (taking real or complex values) from its partial

[8] With isometric masks, $J = 3$ masks can empirically produce exact recovery

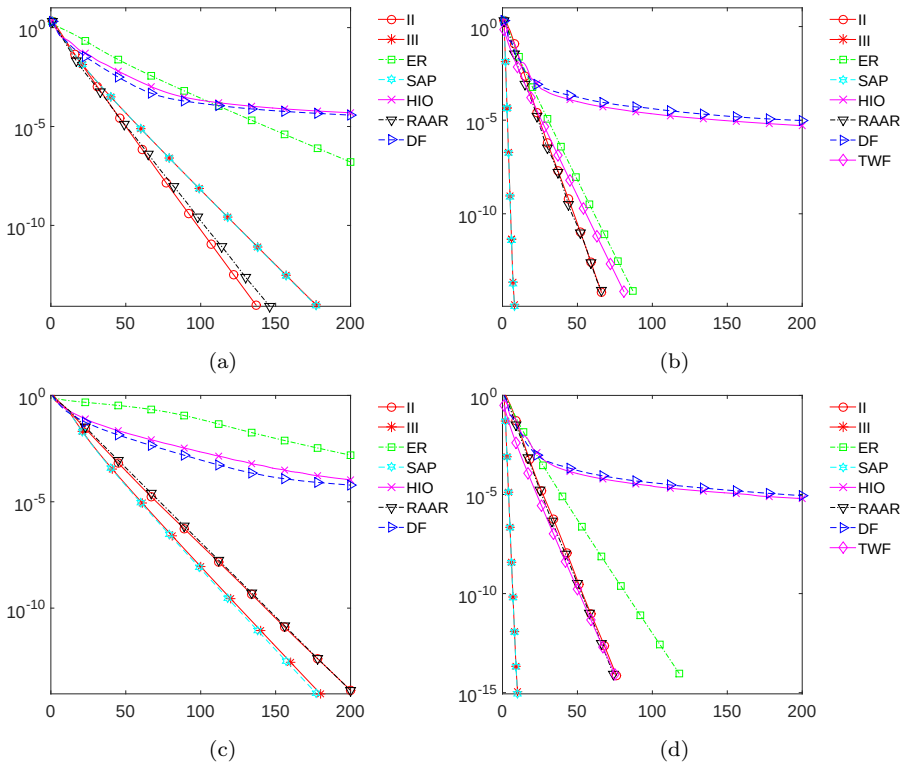



FIG. 12. *Convergence curves (Error v.s. Iteration Number) of Algorithm II ("II" in short) and Algorithm III ("III" in short) in comparison to ER, HIO, DF, RAAR, and TWF for noise free CDP measurements of real-valued "Cameraman" (top) with $J = 2$ (left), $J = 12$ (right) masks and complex-valued "Goldballs" (bottom) with $J = 3$ (left), $J = 12$ (right) masks. Set $r_3 = 0.01$ for Algorithm III, $\delta = 0.6$ for HIO and $\gamma_1 = -\gamma_2 = -1$ for DF. Set $r_1 = 1$ for Algorithm II and $\delta = 0.7$ for RAAR in (a); Set $r = 1.8$ for Algorithm II and $\delta = 0.55$ for RAAR in (b); Set $r = 0.65$ for Algorithm II, and $\delta = 0.75$ for RAAR in (c); Set $r = 1.6$ for Algorithm II and $\delta = 0.55$ for RAAR in (d).*

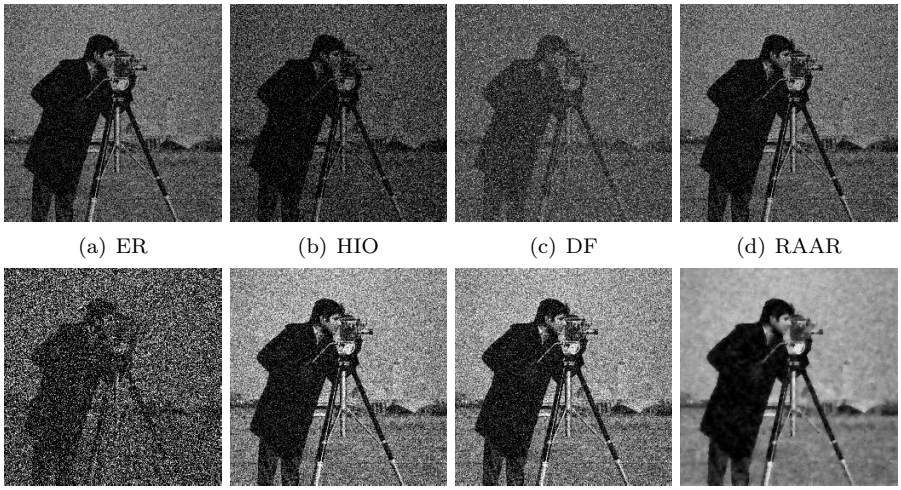
TABLE 3

*SNR/SNR-Intensity values of the recovered images by different algorithms. The two best SNRs are marked in bold. For HIO and DF algorithms, we only show the best recovery results among 50 iterations with highest SNRs. $\eta = 0.005, J = 12$.*

| Algorithms | NULL | ER | HIO | DF | RAAR | TWF | SAP | III | II | I |
|---|---|---|---|---|---|---|---|---|---|---|
| SNR | —— | 4.53 | -2.15 | 1.81 | 4.54 | -2.45 | -0.39 | 9.26 | **9.48** | **17.85** |
| SNR-Intensity | 2.76 | -0.53 | -8.71 | -2.92 | -0.53 | 1.05 | -0.68 | 9.61 | **9.80** | **17.46** |

and noisy magnitude measurements. Numerically, an efficient ADMM was designed with guaranteed convergence, which was validated by numerical experiments. TV regularization shows great potential for robust phase retrieval with very heavy noise, and our proposed methods can recover the images with sharp edges and clean background.

As a future work, we aim to analyze the global convergence of ADMM for the nonconvex optimization problem with non-Lipschitz continuous gradient. Recently, the global convergence for ADMM in [35, 60, 40] was provided for noncovex optimization problems, whose objective functionals contain a Lipschitz differentiable term. We al-



(a) ER (b) HIO (c) DF (d) RAAR

(e) SAP (f) III (g) II (h) I

FIG. 13. *Comparison of PR methods from noisy CDP measurements* (3.10) *with* $J = 12$ *and* $\eta = 0.005$. *The proposed methods are labeled by "I", "II" and "III", short for Algorithm I (with TV) and Algorithms II/III (without TV) respectively. For HIO and DF algorithms, we only show the best recovery results (with highest SNR) among 50 iterations.*

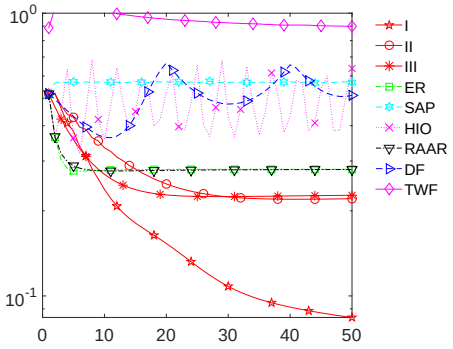



FIG. 14. *Convergence curves (Error v.s. Iteration Number) of Algorithms I/II/III in comparison to ER, HIO, DF, RAAR, WF and TWF for noisy data "Cameraman" with* $J = 12$ *for CDP.* $\eta = 0.005$. *Set* $r_1 = 10, r_2 = 5 \times 10^4, \lambda = 500$ *for Algorithm I,* $r = 10$ *for Algorithm II,* $r_3 = 60$ *for Algorithm III,* $\delta = 0.6, 0.45, 0.1$ *for HIO, RAAR, DF respectively; Set* $\gamma_1 = -\gamma_2 = -1$ *for DF.*

so notice that a proximal heterogeneous block implicit-explicit method was proposed in [33] to solve a blind ptychographic phase retrieval problem, where the objective functional also has such term. Since our proposed models do not contain any Lipschitz differentiable term to guarantee the sufficient decrease of the functional values of augmented Lagrangian, we need to assume that the successive difference of multiplier is vanishing. It is our intention to get rid of this assumption in the future. We are interested in applying the proposed methods to real ptychography data, which is also a future work.

**Acknowledgment.** The authors would like to thank insightful comments of the associated editor and two anonymous referees. Especially, the two referees pointed out some important and very recent references, which greatly help us to improve this

manuscript, and stimulated us to present an accelerated algorithm "Algorithm III" for the non-TV case.

Dr. H. Chang was partially supported by National Natural Science Foundation of China (No. 11501413 and 11426165), China Scholarship Council (CSC) and 2017-Outstanding Young Innovation Team Cultivation Program No. 135202TD1703, Innovation Project No. 52XC1605 of Tianjin Normal University. Dr. Y. Lou was partially supported by NSF DMS-1522786. Dr. S. Marchesini was partially supported by the Center for Applied Mathematics for Energy Research Applications, a joint ASCR-BES funded project within the Office of Science, US Department of Energy, under contract number DOE-DEAC03-76SF00098.

## REFERENCES

[1] F. J. Anscombe, *The transformation of poisson, binomial and negative-binomial data*, Biometrika, 35 (1948), pp. 246–254.

[2] S. Bahmani and J. Romberg, *Phase retrieval meets statistical learning theory: A flexible convex relaxation*, in Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, vol. 54 of Proceedings of Machine Learning Research, PMLR, 20–22 Apr 2017, pp. 252–260.

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

[4] H. H. Bauschke, P. L. Combettes, and D. R. Luke, *Phase retrieval, error reduction algorithm, and fienup variants: a view from convex optimization*, J. Opt. Soc. Amer. A, 19 (2002), pp. 1334–1345.

[5] ———, *Hybrid projectioncreflection method for phase retrieval*, J. Opt. Soc. Amer. A, 20 (2003), pp. 1025–1034.

[6] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Foundations and Trends® in Machine Learning, 3 (2011), pp. 1–122.

[7] Y. M. Bruck and L. G. Sodin, *On the ambiguity of the image reconstruction problem*, Optics Communications, 30 (1979), pp. 304–308.

[8] C. Brune, A. Sawatzky, and M. Burger, *Bregman-em-tv methods with application to optical nanoscopy*, in Scale Space and Variational Methods in Computer Vision, Springer, 2009, pp. 235–246.

[9] ———, *Primal and dual bregman methods with application to optical nanoscopy*, International Journal of Computer Vision, 92 (2011), pp. 211–229.

[10] E. J. Candés, Y. C. Eldar, T. Strohmer, and V. Voroninski, *Phase retrieval via matrix completion*, SIAM J. Imaging Sci., 6 (2013), pp. 199–225.

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

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

[13] E. J. Candes, T. Strohmer, and V. Voroninski, *Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming*, Commu. Pure Applied Math., 66 (2013), pp. 1241–1274.

[14] R. Chan, H. Yang, and T. Zeng, *A two-stage image segmentation method for blurry images with poisson or multiplicative gamma noise*, SIAM Journal on Imaging Sciences, 7 (2014), pp. 98–127.

[15] R. H. Chan and K. Chen, *Multilevel algorithm for a poisson noise removal model with total-variation regularization*, International Journal of Computer Mathematics, 84 (2007), pp. 1183–1198.

[16] Huibin Chang, Yifei Lou, Michael K Ng, and Tieyong Zeng, *Phase retrieval from incomplete magnitude information via total variation regularization*, SIAM Journal on Scientific Computing, 38 (2016), pp. A3672–A3695.

[17] Pengwen Chen and Albert Fannjiang, *Fourier phase retrieval with a single mask by douglas–rachford algorithms*, Applied and Computational Harmonic Analysis, (2016).

[18] P. Chen, A. Fannjiang, and G. R. Liu, *Phase retrieval with one or two diffraction patterns by alternating projections of the null vector*, Journal of Fourier Analysis and Applications, (2017).

[19] Y. Chen and E. Candes, *Solving random quadratic systems of equations is nearly as easy as solving linear systems*, in Advances in Neural Information Processing Systems, 2015, pp. 739–747.
[20] A. Conca, D. Edidin, M. Hering, and C. Vinzant, *An algebraic characterization of injectivity in phase retrieval*, Applied and Computational Harmonic Analysis, 38 (2015), pp. 346–356.
[21] J. C. Dainty and J. R. Fienup, *Phase retrieval and image reconstruction for astronomy*, Image Recovery: Theory and Application, ed. by H. Stark, Academic Press, (1987), pp. 231–275.
[22] Jim Douglas and Henry H Rachford, *On the numerical solution of heat conduction problems in two and three space variables*, Transactions of the American mathematical Society, 82 (1956), pp. 421–439.
[23] V. Elser, *Phase retrieval by iterated projections*, J. Opt. Soc. Am. A, 20 (2003), pp. 40–55.
[24] A. Fannjiang, *Absolute uniqueness in phase retrieval with random illumination*, Inverse Probl., 28 (2012), p. 075008.
[25] J. R. Fienup, *Phase retrieval algorithms: a comparison*, Appl. Opt., 21 (1982), pp. 2758–2769.
[26] M. Figueiredo and J. Bioucas-Dias, *Restoration of poissonian images using alternating direction optimization*, Image Processing, IEEE Transactions on, 19 (2010), pp. 3133–3145.
[27] M. Friedlander and I. Macdo, *Low-Rank Spectral Optimization via Gauge Duality*, SIAM J. Sci. Comput., 38 (2016).
[28] R. W. Gerchberg and W. O. Saxton, *A practical algorithm for the determination of the phase from image and diffraction plane pictures*, Optik, 35 (1972), pp. 237–246.
[29] R. Giryes and M. Elad, *Sparsity-based poisson denoising with dictionary learning*, IEEE Transactions on Image Processing, 23 (2014), pp. 5057–5069.
[30] T. Goldstein and C. Studer, *PhaseMax: Convex Phase Retrieval via Basis Pursuit*, arXiv:1610.07531, (2016).
[31] R. W. Harrison, *Phase problem in crystallography*, J. Opt. Soc. Am. A, 10 (1993), pp. 1046–1055.
[32] M. Hayes, *The reconstruction of a multidimensional sequence from the phase or magnitude of its fourier transform*, IEEE Trans. Acoust. Speech Signal Process., 30 (1982), pp. 140–154.
[33] R. Hesse, D. R. Luke, S. Sabach, and M. K. Tam, *Proximal heterogeneous block implicit-explicit method and application to blind ptychographic diffraction imaging*, SIAM Journal on Imaging Sciences, 8 (2015), pp. 426–457.
[34] E Hofstetter, *Construction of time-limited functions with specified autocorrelation functions*, IEEE Transactions on Information Theory, 10 (1964), pp. 119–126.
[35] M. Hong, Z-Q Luo, and M. Razaviyayn, *Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems*, SIAM Journal on Optimization, 26 (2016), pp. 337–364.
[36] K. Jaganathan, Y. C. Eldar, and B. Hassibi, *Phase retrieval: An overview of recent developments*, arXiv preprint arXiv:1510.07713, (2015).
[37] K. Jaganathan, S. Oymak, and B. Hassibi, *Sparse phase retrieval: Uniqueness guarantees and recovery algorithms*, arXiv preprint arXiv:1311.2745, (2013).
[38] X. Jiang, P. Reynaud-Bouret, V. Rivoirard, L. Sansonnet, and R. Willett, *A data-dependent weighted lasso under poisson noise*, arXiv preprint arXiv:1509.08892, (2015).
[39] T. Le, R. Chartrand, and T.J. Asaki, *A variational approach to reconstructing images corrupted by poisson noise*, Journal of mathematical imaging and vision, 27 (2007), pp. 257–263.
[40] Y. Lou and M. Yan, *Fast l_1-l_2 minimization via a proximal operator*, (to appear) J. Sci. Comput., (2017).
[41] Yue M Lu and Gen Li, *Phase transitions of spectral initialization for high-dimensional nonconvex estimation*, arXiv preprint arXiv:1702.06435, (2017).
[42] D. R. Luke, *Relaxed averaged alternating reflections for diffraction imaging*, Inverse Probl., 21 (2005), pp. 37–50.
[43] D. R. Luke, J. V. Burke, and R. G. Lyon, *Optical wavefront reconstruction: Theory and numerical methods*, SIAM review, 44 (2002), pp. 169–224.
[44] M. Lustig, D.L. Donoho, J. M. Santos, and J. M. Pauly, *Compressed sensing mri*, IEEE Signal Processing Magazine, 25 (2008), pp. 72–82.
[45] M. Mäkitalo and A. Foi, *Optimal inversion of the anscombe transformation in low-count poisson image denoising*, IEEE Transactions on Image Processing, 20 (2011), pp. 99–109.
[46] S. Marchesini, *Invited article: A unified evaluation of iterative projection algorithms for phase retrieval*, Review of scientific instruments, 78 (2007), p. 011301.
[47] S. Marchesini, *Phase retrieval and saddle-point optimization*, J. Opt. Soc. Am. A, 24 (2007), pp. 3289–3296.

[48] S. Marchesini, Y.-C. Tu, and H.-T. Wu, *Alternating projection, ptychographic imaging and phase synchronization*, Applied and Computational Harmonic Analysis, (2015).
[49] J. Miao, T. Ishikawa, Q. Shen, and T. Earnest, *Extending x-ray crystallography to allow the imaging of noncrystalline materials, cells, and single protein complexes*, Annu Rev Phys Chem., 59 (2008), pp. 387–410.
[50] P. Netrapalli, P. Jain, and S. Sanghavi, *Phase retrieval using alternating minimization*, Signal Processing, IEEE Transactions on, 63 (2015), pp. 4814–4826.
[51] J. Qian, C. Yang, A. Schirotzek, F. Maia, and S Marchesini, *Efficient algorithms for ptychographic phase retrieval*, Inverse Problems and Applications, Contemp. Math, 615 (2014), pp. 261–280.
[52] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation noise removal algorithm*, Phys. D, 60 (1992), pp. 259–268.
[53] J. Salmon, Z. Harmany, C-A Deledalle, and R. Willett, *Poisson noise reduction with non-local pca*, Journal of mathematical imaging and vision, 48 (2014), pp. 279–294.
[54] J. L. C. Sanz, *Mathematical considerations for the problem of fourier transform phase retrieval from magnitude*, SIAM J. Appl. Math., 45 (1985), pp. 651–664.
[55] Y. Shechtman, A. Beck, and Y. C Eldar, *Gespar: Efficient phase retrieval of sparse signals*, Signal Processing, IEEE Transactions on, 62 (2014), pp. 928–938.
[56] 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*, Signal Processing Magazine, IEEE, 32 (2015), pp. 87–109.
[57] Irène Waldspurger, Alexandre dAspremont, and Stéphane Mallat, *Phase recovery, maxcut and complex semidefinite programming*, Mathematical Programming, 149 (2015), pp. 47–81.
[58] A. Walther, *The question of phase retrieval in optics*, Journal of Modern Optics, 10 (1963), pp. 41–49.
[59] Y. Wang and Z. Xu, *Phase retrieval for sparse signals*, Applied and Computational Harmonic Analysis, 37 (2014), pp. 531–544.
[60] Y. Wang, W. Yin, and J. Zeng, *Global convergence of admm in nonconvex nonsmooth optimization*, arXiv preprint arXiv:1511.06324, (2015).
[61] Y. Wen, R. H. Chan, and T. Zeng, *Primal-dual algorithms for total variation based image restoration under poisson noise*, Science China Mathematics, 59 (2016), pp. 141–160.
[62] Z. Wen, C. Yang, X. Liu, and S. Marchesini, *Alternating direction methods for classical and ptychographic phase retrieval*, Inverse Probl., 28 (2012), p. 115010.
[63] C. Wu and X.-C. Tai, *Augmented Lagrangian method, dual methods and split-Bregman iterations for ROF, vectorial TV and higher order models*, SIAM J. Imaging Sci., 3 (2010), pp. 300–339.
[64] P. Yin and J. Xin, *Phaseliftoff: an accurate and stable phase retrieval method based on difference of trace and frobenius norms*, Comm. Math. Sci., 13 (2015), pp. 1033–1049.
[65] X. Zhang, Y. Lu, and T. Chan, *A novel sparsity reconstruction method from poisson data for 3d bioluminescence tomography*, Journal of scientific computing, 50 (2012), pp. 519–535.