VARIATIONAL PHASE RETRIEVAL WITH GLOBALLY CONVERGENT PRECONDITIONED PROXIMAL ALGORITHM

HUIBIN CHANG*, STEFANO MARCHESINI†, YIFEI LOU‡, AND TIEYONG ZENG§

Abstract. We reformulate the original phase retrieval problem into two variational models (with and without regularization), both containing a globally Lipschitz differentiable term. These two models can be efficiently solved via the proposed Partially Preconditioned Proximal Alternating Linearized Minimization (P^{3}ALM) for masked Fourier measurements. Thanks to the Lipschitz differentiable term, we prove the global convergence of P^{3}ALM for solving the nonconvex phase retrieval problems. Extensive experiments are conducted to show the effectiveness of the proposed methods.

Key words. Phase Retrieval; Poisson/Gaussian Noise; Regularization; Total Variation; Variational model; Partially Preconditioned Proximal Alternating Linearized Minimization; Global convergence.

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

1. Introduction.

Given the collected phaseless measurements $f \in \mathbb{R}^m_+$, a general phase retrieval problem can be expressed as

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

where $\mathcal{A} : \mathbb{C}^n \to \mathbb{C}^m$ is a linear operator in complex Euclidean space and $|\cdot|$ denotes pointwise absolute values. A classical phase retrieval problem refers to the case that $\mathcal{A}$ is the discrete Fourier transform. It is well-known that classical phase retrieval has trivial ambiguities, such as global phase shift, conjugate inversion, and spatial shift; and non-trivial solutions also exist for one-dimensional signals [41]. Therefore, retrieving the phase from its Fourier magnitude alone is not unique. Hayes [25] proved that at least $2^d \times n$ magnitude measurements should be collected to guarantee the unique recovery of $d-$dimensional ($d \geq 2$) real-valued, non-reducible signals[1]. If the transform $\mathcal{A}$ is generated by a *generic* frame[2], then the injectivity of its quadratic operator $|\mathcal{A}(\cdot)|^2$ is guaranteed by collecting at least $2n-1$ [2] and $4n-4$ [19] measurements for any real and complex signal, respectively. The uniqueness of the phase retrieval can also be proved with additional information, such as collecting oversampled measurements by a coded diffraction pattern [10], and a holographic pattern [9, 13]. Other than exact recovery, researchers are also interested in stable phase retrieval [21, 3]. Please refer to a review paper [44] for more discussions about the uniqueness and stability analysis of phase retrieval.

Appropriate regularizations are often introduced to increase the likelihood of having the unique solution and to improve the robustness of phase retrieval algorithms.

*School of Mathematical Sciences, Tianjin Normal University, Tianjin, 300387, China, E-mail: changhuibin@gmail.com. He is currently a visiting researcher of Computational Research Division at Lawrence Berkeley National Laboratory.

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

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

§Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong, Email: zeng@hkbu.edu.hk

[1] A signal is non-reducible if its z-transform is a non-reducible polynomial.

[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 $\mathbb{C}^n$ [2]. Here a frame is a generalization of a basis of a vector space to sets, which is possibly linearly dependent.

Two types of regularizations are popular in the literature. The first one is associated with sparsity. Driven by compressive sensing, there is a burst of research on sparse or compressive phase retrieval. To enforce the signal to be sparse, it is natural to incorporate the L^0 norm in the phase retrieval models such as *SparseFienup* [37], GrEedy Sparse PhAse Retrieval (GESPAR) [43], and some recent works [50, 20, 27]. As the L^0 norm is often relaxed by the convex L^1 norm in the paradigm of compressive sensing, the L_1 approaches are considered in phase retrieval [35, 56, 38, 30, 8]. The second type of regularizations is inherent from image priors, *e.g.*, Tikhonov [46], shearlet [31], total variation [13], dictionary learning [47, 14] and general sparse prior [15], *etc.*

One major difficulty in phase retrieval comes from the nonconvex constraint (1.1). It is standard to construct a multivariable quadratic system, which is generally NP-complete [4]. A classical phase retrieval approach is an alternating projection algorithm [24, 23] and its variants; please refer to a review paper [33]. Although simple and effective in practice, these alternating projection methods lack of convergence analysis. Notably, this gap has been bridged by convex relaxation via semi-definite programming such as PhaseLift [9] and PhaseCut [49]. Although having guaranteed convergence, these convex methods are computationally expensive due to the lifted dimension, especially for image phase retrieval problems. Alternatively, a gradient descent approach with dynamic stepsizes, called Wirtinger flow [11], together with its variants [18, 8], rigorously demonstrates the exact retrieval of phase information from a set of random measurements. However, convergence analysis of such methods for general phase retrieval measurements is still an open question. In addition, operator-splitting based algorithms [52, 56, 13, 12, 20] exhibit effectiveness and flexibility in phase retrieval. However, this type of approaches suffer from two main drawbacks: (1) one can only prove the local convergence [28, 52, 16]; and (2) convergence analysis often requires strong assumptions, *e.g.*, the convergence of successive differences of the multipliers [13], boundedness of iterative sequences [14], and some special conditions of the linear mapping $\mathcal{A}$ [34]. Recently, global convergence of a Proximal Alternating Linearized Minimization (PALM) algorithm for phase retrieval [26] was derived based on the Kurdyka-Łojasiewicz property [1, 6], but their proposed model with equality constraints was established based on noiseless measurements.

We aim to develop a phase retrieval algorithm that is effective, flexible (easy to incorporate regularization terms), and with guaranteed convergence. In particular, we reformulate the quadratic system of the phase retrieval by a splitting technique and introduce an L^2 fitting term, which is Lipschitz differentiable. Although the idea is straightforward, the existence of such Lipschitz differentiable term in the objective functional plays a critical role in the theoretical guarantee of global convergence, which is not explicitly present in the existing variational models [52, 56, 13] for phase retrieval. Algorithmically, PALM is a generic and popular algorithm for nonlinear optimization problems, which can solve both the convex and nonconvex problems with various setting and constraints very efficiently and flexibly. However, a direct call may be not optimal for phase retrieval. We propose to speed up the plain PALM via pre-conditioning. Especially, for the masked Fourier transform based phase retrieval problems, the corresponding pre-conditioning matrix can be explicitly expressed such that the proposed pre-conditioning algorithm can be fast implemented. In summary, the main contributions of this paper are listed as follows:

1) We reformulate the original phase retrieval problem (1.1) by penalizing a quadratic term $\|z - \mathcal{A}u\|^2$ with an auxiliary variable z, referred to as least

square phase retrieval (LSPR). We show that LSPR is exactly equivalent to (1.1) in the noise-free case. More importantly, the reformulation has the flexibility of accommodating any regularization to deal with noisy data, referred to as Reg-LSPR.

2) We propose to incorporate an efficient pre-conditioning technique into the celebrated PALM algorithm, called Partially Pre-conditioned PALM (P^{3}ALM), to improve the convergence of the plain PALM and the robustness *w.r.t.* noise and the number of measurements. Thanks to the presence of Lipschitz differentiable terms in the objective functionals, the global convergence of P^{3}ALM for both LSPR and Reg-LSPR is theoretically guaranteed with proper stepsizes. To our best knowledge, it is the first time to show the global convergence for the regularized phase retrieval model under very mild conditions.
3) We conduct extensive experiments to demonstrate the effectiveness of proposed methods. Particularly with two coded diffraction patterns, our proposed algorithm is at least twice as fast as the state of the art first-order[3] algorithms. For compressive phase retrieval, the proposed P^{3}ALM for Reg-LSPR with L^0 regularization outperforms SparseFienup [37] and the probabilistic approach (PRGAMP) [42] with higher successful recovery rates. For the noisy cases, P^{3}ALM produces comparable denoising results as TVPoiPR in [12] with faster convergence.

The rest of this paper is organized in the following way. We describe both least square phase retrieval model and a regularized version in Section 2. The numerical algorithms for these two models are discussed in Section 3 with the global convergence established in Section 4. Section 5 is devoted to extensive experiments for both noiseless and noisy cases to demonstrate the efficiency of the proposed methods for phase retrieval from classical Fourier transform, coded diffraction, and ptychographic patterns. We give conclusions and future works in Section 6.

2. Proposed Models.

We introduce a quadratic term for phase retrieval, which plays an important role in analyzing global convergence of first-order operator-splitting based algorithms. In particular, we propose two phase retrieval models: without regularization in Section 2.1 and with regularization in Section 2.2.

2.1. Proposed model without regularization.

In addition to linear operator $\mathcal{A}$ in (1.1), we assume that the measured intensity data are further corrupted by noise, i.e.,

$$f(t) \overset{\text{ind.}}{\sim} \text{Noise}(|\mathcal{A}u|^2(t)),\ \forall 0 \leq t \leq m-1, \tag{2.1}$$

where "Noise" represents the degradation of the intensity data $|\mathcal{A}u|^2$ due to noise. We consider both Gaussian and Poisson distributions, which are typical in phase retrieval measurements.

To determine the underlying image u from noisy measurements f, one standard approach involves an optimization problem

$$\min_u \mathcal{B}(|\mathcal{A}u|^2, f), \tag{2.2}$$

[3]The first-order algorithm refers to the method that only involves the first-order gradient of the objective functional.

where $\mathcal{B}(\cdot,\cdot)$ measures the distance between the unknown intensity $|\mathcal{A}u|^2$ and collected phaseless data f. Various metrics have been proposed to deal with different noise settings, such as amplitude based metric for Gaussian measurements (AGM) [52, 13], intensity based metric for Poisson measurements (IPM) [46, 18, 12, 15], and intensity based metric for Gaussian measurements (IGM) [39, 11, 8, 45, 15]. We consider all the aforementioned data fitting terms, which can be mathematically expressed as

$$\mathcal{B}(g,f) := \begin{cases} \frac{1}{2}\left\|\sqrt{g}-\sqrt{f}\right\|^2; & \text{(AGM)} \\ \frac{1}{2}\langle g - f\circ\log(g), \mathbf{1}\rangle; & \text{(IPM)} \\ \frac{1}{2}\|g-f\|^2; & \text{(IGM)} \end{cases} \tag{2.3}$$

where $\circ$ denotes the pointwise multiplication[4], $\mathbf{1}$ denotes a vector whose entries all equals to ones, and $\|\cdot\|$ denotes the L^2 norm in Euclidean space. Note that the Alternating Direction Method of Multipliers (ADMM) was adopted in [52, 13, 12] to solve the variational phase retrieval model in (2.2). However, due to the lack of the globally Lipschitz differentiable terms in the objective functionals, it seems difficult to guarantee its convergence.

In this paper, we consider to reformulate the model (2.2) as a least square phase retrieval (LSPR) problem:

$$\text{LSPR:} \qquad \min_{u,z} \mathcal{F}_\sigma(u,z) := \frac{\sigma}{2}\|z-\mathcal{A}u\|^2 + \mathcal{B}(|z|^2, f), \tag{2.4}$$

with the parameter $\sigma > 0$. It can be derived by penalizing the constraint $z = \mathcal{A}u$ to (2.2) with the weight σ. We demonstrate in Theorem 1 that if the measured intensity f is noise free, then finding a point in the solution set of phase retrieval problem (1.1), defined as

$$\mathcal{S}(f) = \{u \in \mathbb{C}^n : \ |\mathcal{A}u|^2 = f\}, \tag{2.5}$$

is equivalent to solving the LSPR model (2.4).

Theorem 1. *Assume that $f \geq 0$, $S(f)$ is nonempty and σ is an arbitrary positive constant.*

1). If $(u^\star, z^\star) \in \arg\min_{u,z} \mathcal{F}_\sigma(u,z)$, then we have $u^\star \in \mathcal{S}(f)$ and $z^\star = \mathcal{A}u^\star$.

2). If $u^\star \in \mathcal{S}(f)$, then we have $(u^\star, z^\star) \in \arg\min_{u,z} \mathcal{F}_\sigma(u,z)$ and $z^\star = \mathcal{A}u^\star$.

Please refer to Appendix A.1 for the proof. Theorem 1 establishes the equivalence of the proposed LSPR model (2.4) to the original phase retrieval problem (1.1) in the case of noiseless measurements. The proof also implies that (1.1) and (2.2) are equivalent, and hence (2.4) is also equivalent to (2.2) under the same conditions as Theorem 1.

We want to further analyze the LSPR model (2.4) in terms of optimality condition, which only addresses critical points instead of global solutions. To make the paper self-contained, we provide basic definitions for the generalized subdifferential [36] as follows.

Definition 1. *Let $J : \mathbb{R}^d \to (-\infty, +\infty]$ be proper and lower semicontinuous (l.s.c.).*

[4] The operation such as $\sqrt{\cdot}, \log(\cdot), |\cdot|$ are all defined pointwisely in this paper.

1). Fréchet subdifferential:

$$\hat{\partial} J(u) = \left\{ v \in \mathbb{R}^d : \liminf_{w \neq u, w \to u} \frac{J(w) - J(u) - \langle v, w-u \rangle}{\|w-u\|} \geq 0 \right\}, \forall u \in \mathrm{dom} J.$$

2). Limiting subdifferential:

$$\partial J(u) = \Big\{ v \in \mathbb{R}^d : \exists u^k \to u \text{ with } J(u^k) \to J(u), \\ \text{and } v^k \to v \text{ with } v^k \in \hat{\partial} J(u^k) \text{ as } k \to \infty \Big\}.$$

We remark that if the standard derivative of a function exists, denoted by ∇, the Fréchet subderivative and limiting subderivative are coincident with this derivative. Immediately it holds

$$\partial J = \nabla J_1 + \partial J_2, \text{ if } J = J_1 + J_2, \tag{2.6}$$

if J_1 is continuously differentiable;

$$\partial J(u_1, u_2) = (\nabla_{u_1} J(u_1, u_2), \partial_{u_2} J(u_1, u_2)), \tag{2.7}$$

if $J(u_1, u_2)$ is continuously differentiable w.r.t. u_1.

We separate the real and imaginary components and rewrite the objective functional as

$$\mathcal{B}(|z|^2, f) = \mathcal{B}(|z_1|^2 + |z_2|^2, f),$$
$$\|z - \mathcal{A}u\|^2 = \|z_1 - \mathcal{A}_1 u_1 + \mathcal{A}_2 u_2\|^2 + \|z_2 - \mathcal{A}_1 u_2 - \mathcal{A}_2 u_1\|^2,$$

with

$$\mathcal{A} := \mathcal{A}_1 + \mathbf{i}\mathcal{A}_2, \; u = u_1 + \mathbf{i} u_2, \text{ and } z = z_1 + \mathbf{i} z_2,$$

and $\mathbf{i} = \sqrt{-1}$. Then we are ready to generalize the above limiting subdifferential to complex-valued variables (subdifferential w.r.t. real and imaginary parts) as

$$\partial J(u) := \partial_{u_1} J(u) + \mathbf{i} \partial_{u_2} J(u).$$

Definition 2. *For a proper and l.s.c. functional J, $\check{u}$ is a critical point of J if $0 \in \partial J(\check{u})$.*

For any critical point of the proposed LSPR model (2.4), denoted as $(\check{u}, \check{z})$, the first order optimality condition can be expressed as

$$\begin{cases} 0 = \mathcal{A}^*(\check{z} - \mathcal{A}\check{u}), \\ 0 \in \sigma(\check{z} - \mathcal{A}\check{u}) + 2\partial_g \mathcal{B}(g, f)\big|_{g = |\check{z}|^2} \circ \check{z}, \end{cases} \tag{2.8}$$

where the second equality is derived by (2.6). After multiplying the second equation by $\mathcal{A}^*$, and eliminating the first term, we have

$$\begin{cases} 0 = \mathcal{A}^*(\check{z} - \mathcal{A}\check{u}), \\ 0 \in \mathcal{A}^* \left(\partial_g \mathcal{B}(|\check{z}|^2, f) \circ \check{z} \right). \end{cases} \tag{2.9}$$

Theorem 2. *Each element $\check{u} \in \mathcal{S}(f)$ is a critical point of* (2.4), i.e.

$$0 \in \partial \mathcal{F}_\sigma(\check{u}, \mathcal{A}\check{u}),$$

with $\partial \mathcal{F}_\sigma(u,z) = (\nabla_u \mathcal{F}_\sigma(u,z), \partial_z \mathcal{F}_\sigma(u,z))$.

See the proof in Appendix A.2. Note that there exists a gap between the global minimizers and the critical points. It would be interesting to further classify critical points as global minimizer, local minimizer, or saddle points, which will be our future work; please refer to some discussions in the conclusion section. In summary, we have the following relationship among (1.1), (2.2), (2.4): if $f \geq 0$ and $\mathcal{S}(f) \neq \emptyset$, then

$$\mathcal{S}(f) = \left\{ u^\star : \ (u^\star, \mathcal{A}u^\star) \in \arg\min_{u,z} \mathcal{F}_\sigma(u,z) \right\} = \arg\min_u \mathcal{B}(|\mathcal{A}u|^2, f),$$

and

$$\mathcal{S}(f) \subseteq \{\check{u} : \ 0 \in \partial \mathcal{B}(|\mathcal{A}\check{u}|^2, f)\} \subseteq \{\check{u} : \ \exists \check{z}, \ s.t. \ 0 \in \partial \mathcal{F}_\sigma(\check{u}, \check{z})\}.$$

Note that the set of critical points of (2.2) is a subset of the one of the proposed model (2.4), which can be readily obtained by (2.9) and following the proof of Theorem 2. If $(\check{u}, \check{z})$ is a critical point of (2.4), the error $\check{z} - \mathcal{A}\check{u} \in \mathrm{Null}(\mathcal{A}^*)$. Therefore, if $\mathcal{A}$ is a nonsingular square matrix, then the sets of critical points for (2.2) and (2.4) are identical.

For noisy measurements f, it is possible that the solution set $\mathcal{S}(f)$ is empty. Nevertheless, LSPR (2.4) always admits a solution u_σ depending on σ if $\mathcal{A}^*\mathcal{A}$ is nonsingular, which is the case for most real applications. For example, masked Fourier transform has full column rank. Indeed, in the noisy case, the parameter σ balances the projections onto two constraint sets, *i.e.*

$$\{z \in \mathbb{C}^m : \ z = \mathcal{A}u, \ \forall u \in \mathbb{C}^n\} \quad \text{and} \quad \{z \in \mathbb{C}^m : \ |z|^2 \approx f\}.$$

If $\sigma \to +\infty$, the solution to (2.4) approaches to a minimizer to (2.2), which is indicated by the following theorem. It is a direct extension of [53, Theorem 17.1] to the complex-valued case; hence the proof is omitted.

Theorem 3. *Let each* (u^l, z^l) *be the exact global minimizer of* $\mathcal{F}_{\sigma^l}(u,z)$ *defined by* (2.4)*, and* $\sigma^l \to +\infty$ *as* $l \to +\infty$*. Then every limit point of the sequence* u^l *is a global solution of the problem* (2.2).

In practice, we do not need to select a very large σ or update its value gradually to infinity. We study the impact by σ in Section 5.4, particularly Figure 14 (a), (d).

2.2. Proposed model with regularization. With further assumptions of the underlying image, variational regularized methods are adopted to suppress the noise. Inspired by compressive sensing, Moravec *et al.* [35], Ohlsson *et al.* [38], and Yang *et al.* [56] proposed the standard L_1 minimization,

$$\min \|u\|_1, \quad s.t. \ |\mathbf{F}u|^2 = f,$$

or the unconstrained formulation,

$$\min \lambda \|u\|_1 + \frac{1}{2} \||\mathbf{F}u| - \sqrt{f}\|^2,$$

with $\mathbf{F} \in \mathbb{C}^{n \times n}$ denoting the normalized discrete Fourier transform. Duan *et al.* [20] proposed the L^0 regularization for phase retrieval,

$$\min \lambda \|u\|_0 + \frac{1}{s} \||\mathcal{A}u| - \sqrt{f}\|_s^s,$$

with $s = 1, 2$. In coherent diffractive imaging, Thibault and Guizar-Sicairos [46] proposed Tikhonov regularization of the transform domain for Poisson noise removal,

$$\min \lambda\|\nabla u\|^2 + \frac{1}{2}\langle|\mathcal{A}u|^2 - f \circ \log(|\mathcal{A}u|^2), \mathbf{1}\rangle,$$

with gradient operator ∇. In order to further preserve the edges of recovered images, Chang *et al.* [12, 13] established total variation regularized model for general phase retrieval

$$\min_u \lambda \mathrm{TV}(u) + \mathcal{B}(|\mathcal{A}u|^2, f),$$

where TV stands for total variation regularization [40]. In summary, the regularized model can be rewritten in a unified form as

$$\min_u \lambda\mathcal{R}(u) + \mathcal{B}(|\mathcal{A}u|^2, f), \tag{2.10}$$

with regularization $\mathcal{R}(\cdot)$ and data fitting term $\mathcal{B}(\cdot,\cdot)$.

Similarly to LSPR, we propose a modified version of (2.10), referred to as "Reg-LSPR",

$$\text{Reg-LSPR:} \qquad \min_{u,z} \mathcal{G}(u, z) := \lambda\mathcal{R}(u) + \mathcal{F}_\sigma(u, z), \tag{2.11}$$

where $\mathcal{F}_\sigma$ is defined in (2.4) and $\lambda > 0$ is a parameter to balance the regularization and data fitting terms. We show the existence of solutions to (2.11) under mild conditions in the following theorem. The proof is standard, thus omitted.

Theorem 4. *Assume that (i) $\mathcal{R}(\cdot)$ is proper, lower semi-continuous and (ii) $\mathcal{A}^*\mathcal{A}$ is nonsingular, then there exists a minimizer $(u^\star, z^\star)$ for (2.11),* i.e.

$$(u^\star, z^\star) \in \arg\min_{u,z} \mathcal{G}(u, z).$$

We remark that $\mathcal{A}$ is usually nonsingular for different Fourier masked measurements patterns, such as classical phase retrieval pattern involving with Fourier transform, coded diffraction pattern, holographic pattern, and ptychographic pattern [12]. In general, it is difficult to show that a nonconvex minimization problem has a unique solution. If no regularization, *i.e.* $\lambda = 0$, we establish in Theorem 1 that the minimizer of (2.11) is unique up to global phase factor if the dimension of $\mathcal{S}(f)$ is not greater than one. Please refer to [13] for more discussions about the uniqueness of phase retrieval solutions.

3. Numerical algorithms. We first review the plain PALM in Section 3.1. We then describe the partially preconditioned version for LSPR in Section 3.2, and present the relationship of the proposed algorithm to alternating projection algorithms in Section 3.3. Finally, we describe the proposed partially preconditioned algorithm for Reg-LSPR in Section 3.4.

3.1. Proximal alternating linearized minimization(PALM). We present a general optimization problem in the form of

$$\min_v \varepsilon_1(v) + \varepsilon_2(v),$$

with C^1 functional $\varepsilon_1(\cdot)$ and nonsmooth functional $\varepsilon_2(\cdot)$. The PALM algorithm [6] is based on a proximal operator defined as follows,

Definition 3. *Given function $f : \mathbb{C}^N \to \mathbb{R}\bigcup\{+\infty\}$ and a positive definite matrix M, the proximal operator $\mathrm{Prox}_\mu^f : \mathbb{C}^N \to \mathbb{C}^N$ of f is defined by*

$$\mathrm{Prox}_\mu^f(v; M) = \arg\min_x \Big(f(x) + \frac{\mu}{2}\|x - v\|_M^2\Big), \tag{3.1}$$

where $\|v\|_M^2 := \langle v, v\rangle_M$ with $\langle v, u\rangle_M = \langle Mv, u\rangle$ and the standard L^2 inner product $\langle\cdot,\cdot\rangle$.

Given an approximated solution v^k, the forward and backward splitting algorithm [22] with a preconditioning matrix M and stepsize α^k can be formulated as follows,

$$\begin{aligned}
v^{k+1} &\in \arg\min_v \underbrace{\varepsilon_1(v^k) + \langle v - v^k, \nabla\varepsilon_1(v^k)\rangle}_{1^{st}\text{-order expansion of } \varepsilon_1(v) \text{ at } v^k} \underbrace{+\varepsilon_2(v) + \frac{\alpha^k}{2}\|v - \hat{v}^k\|_M^2}_{\text{Prox-Regularization}} \\
&\in \arg\min_v \overbrace{\varepsilon_2(v) + \frac{\alpha^k}{2}\|v - \underbrace{(v^k - \frac{1}{\alpha^k}M^{-1}\nabla\varepsilon_1(v^k))}_{Forward}\|_M^2}^{Backward} \\
&:= \mathrm{Prox}_{\alpha^k}^{\varepsilon_2}\Big(v^k - \frac{1}{\alpha^k}M^{-1}\nabla\varepsilon_1(v^k); M\Big),
\end{aligned}$$

where ∇ denotes the gradient for the C^1 smooth function. Hereafter if M is identity operator $\mathbf{I}$, we denote $\mathrm{Prox}_{\alpha^k}^{\varepsilon_2}(u) := \mathrm{Prox}_{\alpha^k}^{\varepsilon_2}(\hat{u}^k; \mathbf{I})$ for simplicity.

If $\varepsilon_2(\cdot)$ has a separable structure as

$$\varepsilon_2(v) := \psi_1(v_1) + \psi_2(v_2),$$

with $v = (v_1^T, v_2^T)^T$, we can consider the following optimization problem

$$\min_{v_1, v_2} \pi(v_1, v_2) + \psi_1(v_1) + \psi_2(v_2),$$

where $\pi(v_1, v_2) = \varepsilon_1(v)$. Then PALM based on the forward and backward splitting (without preconditioning) can be given with stepsizes α_1^k, α_2^k as

$$\begin{cases}
v_1^{k+1} \in \mathrm{Prox}_{\alpha_1^k}^{\psi_1}\Big(v_1^k - \frac{1}{\alpha_1^k}\nabla_{v_1}\pi(v_1^k, v_2^k)\Big), \\
v_2^{k+1} \in \mathrm{Prox}_{\alpha_2^k}^{\psi_2}\Big(v_2^k - \frac{1}{\alpha_2^k}\nabla_{v_2}\pi(v_1^{k+1}, v_2^k)\Big),
\end{cases} \tag{3.2}$$

with approximated solutions v_1^k, v_2^k.

3.2. Partially preconditioned PALM (P^{3}ALM) for LSPR. We rewrite (2.4) as

$$\min_{u,z} \mathcal{F}_\sigma(u, z) := \mathcal{H}(u, z) + \mathcal{B}(|z|^2, f), \tag{3.3}$$

by denoting $\mathcal{H}(u, z) := \frac{\sigma}{2}\|z - \mathcal{A}u\|^2$. We can compute the derivative of the functional with respect to complex-valued variables by taking the derivative with respect to the real and imaginary parts of a complex-valued variable separately. Following [13], one can readily get the partial derivative of $\mathcal{H}$ with respect to variables u and z as

$$\begin{cases} \nabla_u \mathcal{H}(u,z) = \sigma \mathcal{A}^*(\mathcal{A}u - z), \\ \nabla_z \mathcal{H}(u,z) = \sigma(z - \mathcal{A}u). \end{cases} \tag{3.4}$$

Since $\|\nabla\mathcal{H}(u_1,z_1) - \nabla\mathcal{H}(u_1,z_2)\| \le \sigma \max\{1, \|\mathcal{A}\|\}(\|\mathcal{A}^*\|+1)\|(u_1^T, z_1^T) - (u_2^T, z_2^T)\|$, for all $u_1, u_2 \in \mathbb{C}^n$, and $z_1, z_2 \in \mathbb{C}^m$, we have $\nabla\mathcal{H}(\cdot,\cdot)$ is Lipschitz continuous.

Applying (3.2) to the LSPR model (2.4), we obtain an iterative scheme

$$u^{k+1} = \check{u}^k, \qquad z^{k+1} = \mathrm{Prox}_{d^k}^{\mathcal{B}(|\cdot|^2, f)}(\hat{z}^k), \tag{3.5}$$

with positive descent stepsizes c^k, d^k, where

$$\begin{cases} \check{u}^k := u^k - \dfrac{1}{c^k}\nabla_u \mathcal{H}(u^k, z^k) & = \left(\mathbf{I} - \dfrac{\sigma}{c^k}\mathcal{A}^*\mathcal{A}\right) u^k + \dfrac{\sigma}{c^k}\mathcal{A}^* z^k, \\ \hat{z}^k := z^k - \dfrac{1}{d^k}\nabla_z \mathcal{H}(u^{k+1}, z^k) & = \left(1 - \dfrac{\sigma}{d^k}\right) z^k + \dfrac{\sigma}{d^k}\mathcal{A}u^{k+1}, \end{cases} \tag{3.6}$$

by (3.4). For each data fitting term $\mathcal{B}(|\cdot|^2, f)$ defined in (2.3), there is a closed-form formula for the proximal operator. We list them all below,

$$\mathrm{Prox}_{\beta}^{\mathcal{B}(|\cdot|^2, f)}(z) = \begin{cases} \dfrac{\sqrt{f} + \beta|z|}{1+\beta} \circ \mathrm{sign}(z), & \text{for AGM [52, 13];} \\ \dfrac{\beta|z| + \sqrt{(\beta|z|)^2 + 4(1+\beta)f}}{2(1+\beta)} \circ \mathrm{sign}(z), & \text{for IPM [55, 12];} \\ \varpi_\beta(|z|) \circ \mathrm{sign}(z), & \text{for IGM [15];} \end{cases} \tag{3.7}$$

The expression of $\varpi_\beta(\cdot)$ is complicated, which is given in Appendix B.

Remark 3.1. *We want to point out that the computational bottleneck for quadratic inverse problems lies in the subproblem of z. Thanks to the proximal operators (3.7), some operator-splitting based algorithms such as PALM and ADMM [52, 13, 20] can be adopted, and therefore computational cost of the quadratic inverse problem per iteration is in the same order to the one of linear inverse problems.*

We observe empirically that PALM for LSPR converges very slowly; see numerical examples in Figure 1. To speed up, we propose a Partially[5] Preconditioned PALM ($\mathrm{P^3ALM}$),

$$u^{k+1} = \hat{u}^k, \qquad z^{k+1} = \mathrm{Prox}_{d^k}^{\mathcal{B}(|\cdot|^2, f)}(\hat{z}^k), \tag{3.8}$$

with

$$\hat{u}^k := u^k - \frac{1}{c^k} M^{-1}\partial_u \mathcal{H}(u^k, z^k) = \left(1 - \frac{\sigma}{c^k}\right) u^k + \frac{\sigma}{c^k} M^{-1}\mathcal{A}^* z^k.$$

Heuristically, we choose $M := \mathcal{A}^*\mathcal{A}$, if the linear mapping $\mathcal{A}$ is masked Fourier transform, leading to a diagonal matrix $\mathcal{A}^*\mathcal{A}$. Therefore, the inverse of M is simply taking reciprocal of diagonal elements of $\mathcal{A}^*\mathcal{A}$. We show in experimental section that this preconditioning significantly speeds up the convergence of the plain PALM and also

[5] Here "Partially" means only preconditioning *w.r.t.* variable u.

gives reasonable phase retrieval results. If the mapping $\mathcal{A}$ does not have such structure, *e.g.,* a random dense matrix $\mathcal{A}$ with each element following Gaussian distribution [11], this preconditioning matrix may lead to higher computational cost.

We summarize the proposed P^{3}ALM for LSPR (2.4) in Algorithm 1. Note that we cannot choose $u^0 = z^0 = 0$ as initial conditions; otherwise, the iterative sequences would converge to zero, i.e., $(\check{u}, \check{z}) = (0, 0)$, which is a critical point of LSPR, but not a solution to (1.1) if $\|f\| \neq 0$. In experiments, we choose $u^0 \neq 0$ randomly.

Algorithm 1 P^{3}ALM for LSPR (2.4)

Initialization: $u^0 \neq 0, z^0 = \mathcal{A}u^0, k := 0$, maximum iteration number MAX_{out} and parameters σ, and stepsizes $\{c^k\}, \{d^k\}$. Set precondition matrix $M = \mathcal{A}^*\mathcal{A}$ for masked Fourier measurements.

Output: $u^\star = u^{MAX_{out}-1}$.

1: **for** $k = 0$ to $MAX_{out} - 1$ **do**

2: $u^{k+1} = \left(1 - \frac{\sigma}{c^k}\right) u^k + \frac{\sigma}{c^k} M^{-1}\mathcal{A}^* z^k$.

3: $z^{k+1} = \mathrm{Prox}_{d^k}^{\mathcal{B}(|\cdot|^2,f)}\left(\left(1 - \frac{\sigma}{d^k}\right) z^k + \frac{\sigma}{d^k}\mathcal{A}u^{k+1}\right)$, where the closed-form solution of the proximal mapping is derived by (3.7) with different metrics.

4: **end for**

3.3. Connections to alternating projection algorithms. In this subsection, we want to draw connections of Algorithm 1 to other alternating projection algorithms that are popular in optic community.

If $M = \mathcal{A}^*\mathcal{A}$ is invertible, two projection operators [12] can be introduced, *i.e.*,

$$\mathcal{P}_1(z) = \sqrt{f} \circ \mathrm{sign}(z) \quad \text{and} \quad \mathcal{P}_2(z) = \mathcal{A}M^{-1}\mathcal{A}^* z.$$

Therefore, the error reduction (ER) [24] algorithm can be rewritten as

$$z^{k+1} = \mathcal{P}_2\mathcal{P}_1(z^k), \text{ for } k = 0, 1, \cdots,$$

and u^{k+1} can be computed as

$$u^{k+1} = M^{-1}\mathcal{A}^* z^{k+1}. \tag{3.9}$$

Similarly, one can obtain algorithms for HIO, DF, and RAAR with relaxed parameters δ, γ_1 and γ_2 as follows,

$$\begin{cases} \text{HIO:} & z^{k+1} = \left((1+\delta)\mathcal{P}_2\mathcal{P}_1 + \mathbf{I} - \mathcal{P}_2 - \delta\mathcal{P}_1\right)(z^k), \\ \text{DF:} & z^{k+1} = \left(\mathbf{I} + \delta\big(\mathcal{P}_2((1+\gamma_2)\mathcal{P}_1 - \gamma_2\mathbf{I}) - \mathcal{P}_1((1+\gamma_1)\mathcal{P}_2 - \gamma_1\mathbf{I})\big)\right)(z^k), \\ \text{RAAR:} & z^{k+1} = \left(2\delta\mathcal{P}_2\mathcal{P}_1 + \delta\mathbf{I} - \delta\mathcal{P}_2 + (1-2\delta)\mathcal{P}_1\right)(z^k). \end{cases} \tag{3.10}$$

Note that alternating projection algorithms can work well in practice, as the matrix M is usually diagonal . Our proposed algorithm (Algorithm 1) can be expressed as

$$\begin{cases} \tilde{z}^{k+1} = (1-\delta)\tilde{z}^k + \delta\mathcal{P}_2(z^k), \\ z^{k+1} = \mathrm{Prox}_{d^k}^{\mathcal{B}(|\cdot|^2,f)}\left((1-\gamma)z^k + \gamma(1-\delta)\tilde{z}^k + \gamma\delta\mathcal{P}_2(z^k)\right), \end{cases}$$

where $\tilde{z}^k := \mathcal{A}u^k$, and $\delta := \frac{\sigma}{c}$, $\gamma := \frac{\sigma}{d}$ with fixed stepsizes $c \equiv c^k, d \equiv d^k$.

Remark 3.2. *It is interesting to consider two special cases of $\delta = 1$ and $\gamma = 1$, since they are equivalent to solving $u-$subproblem and $z-$subproblem of the proposed*

LSPR directly without forward-backward procedure, which is essentially the alternating minimization with respect to u or v. For $\delta = 1$, one obtains

$$z^{k+1} = \mathrm{Prox}_{d^k}^{\mathcal{B}(|\cdot|^2,f)}\Big(\big((1-\gamma)\mathbf{I} + \gamma\mathcal{P}_2\big)\big(z^k\big)\Big). \tag{3.11}$$

For $\gamma = 1$, one has

$$\begin{cases} \tilde{z}^{k+1} = (1-\delta)\tilde{z}^k + \delta\mathcal{P}_2(z^k), \\ z^{k+1} = \mathrm{Prox}_{d^k}^{\mathcal{B}(|\cdot|^2,f)}\left((1-\delta)\tilde{z}^k + \delta\mathcal{P}_2(z^k)\right). \end{cases}$$

We will provide numerical tests for these special cases; please refer to Figure 14(b), (c), (e), and (f), which show that the algorithm converges much slower without forward-backward procedure.

We show heuristically that the proximal operator plays a similar role to the projection operator $\mathcal{P}_1$. The analysis is based on a strong assumption that the proposed algorithm produces the iterative sequence (u^k, z^k) converging to the global minimizer. Taking the data fitting term of AGM for an example, we have

$$\lim_{k\to+\infty} |z^k| = \sqrt{f}, \quad \lim_{k\to+\infty} \tilde{z}^k = \lim_{k\to+\infty} \mathcal{A}u^k = \lim_{k\to+\infty} z^k, \quad \lim_{k\to+\infty} \mathcal{P}_2(z^k) = \lim_{k\to+\infty} z^k,$$

which immediately gives

$$\lim_{k\to+\infty} |\hat{z}^k| = \lim_{k\to+\infty} |(1-\gamma)z^k + \gamma(1-\delta)\tilde{z}^k + \gamma\delta\mathcal{P}_2(z^k)| = \lim_{k\to+\infty} |z^k| = \sqrt{f}.$$

Hence one has

$$\lim_{k\to+\infty} \left|\mathrm{Prox}_{d^k}^{\mathcal{B}(|\cdot|^2,f)}(\hat{z}^k)\right| = \lim_{k\to+\infty} \frac{\sqrt{f} + d^k|\hat{z}^k|}{1+d^k} = \sqrt{f}, \tag{3.12}$$

which shows that the proximal mapping provides an approximation of the magnitude constraint. For other data fitting terms, one can get the similar results to (3.12). Replacing the projection operator $\mathcal{P}_1$ with the proximal operator could yield a new algorithm, but convergence analysis of such scheme is not available.

3.4. P^{3}ALM for Reg-LSPR. In this section, we give a similar P^{3}ALM for the Reg-LSPR model (2.11) as follows:

$$\begin{cases} u^{k+1} = \mathrm{Prox}_{c^k}^{\lambda\mathcal{R}(u)}(\hat{u}^k; M); \\ z^{k+1} = \mathrm{Prox}_{d^k}^{\mathcal{B}(|\cdot|^2,f)}(\hat{z}^k), \end{cases} \tag{3.13}$$

with $\hat{u}^k, \hat{z}^k$ defined in (3.6). The iteration (3.13) can be regarded as a general framework for phase retrieval by a two-step scheme [15], in which the u-update is referred to as denoising and z-update is the generalized least squares problem. Particularly for the denoising step, we consider two types of regularization terms and discuss how to solve the corresponding regularized linear reconstruction problem in detail.

3.4.1. L^s regularization for $s = 0, 1$ for compressive phase retrieval. Motivated from compressive sensing, we apply the L^s regularization, i.e., $\mathcal{R}(u) = \|u\|_s$ for $s = 0, 1$, to recover sparse or compressive signals. The subproblem *w.r.t* u reads

$$\min_{u\in\mathbb{C}^n} \lambda\|u\|_s + \frac{c^k}{2}\|u - \hat{u}\|_M^2. \tag{3.14}$$

For diagonal preconditioned matrix M, the closed-form solutions of (3.14) are expressed as

$$u^{k+1} = \begin{cases} \text{Thresh}_{hard}\left(\hat{u}^k; \frac{\lambda}{c^k}\text{diag}(M^{-1})\right), \text{ if } s=0, \\ \text{Thresh}_{soft}\left(\hat{u}^k; \frac{\lambda}{c^k}\text{diag}(M^{-1})\right), \text{ if } s=1, \end{cases} \tag{3.15}$$

where

$$\text{Thresh}_{soft}(u;\beta) = \max\{0, |u|-\beta\} \circ \text{sign}(u),$$

$$\text{Thresh}_{hard}(u;\beta)(t) = \begin{cases} u(t), \text{ if } |u(t)| \geq \sqrt{2\beta}, \\ 0, \text{ otherwise}, \end{cases} \quad \forall 0 \leq t \leq n-1. \tag{3.16}$$

If M is not diagonal, one needs inner iterations such as the forward and backward splitting to solve (3.14). For real-valued images, additional constraint can be incorporated, such as $u \geq 0$, and one can readily obtain the closed-form solution by directly projection onto the nonnegative set.

3.4.2. Total variation regularization. The TV regularization was applied in phase retrieval by Chang *et al.* [12, 13]. For $\mathcal{R}(u) = \text{TV}(u)$, the u-subproblem reads

$$\min_{u\in\mathbb{C}^n} \lambda\text{TV}(u) + \frac{d^k}{2}\|u-\hat{u}\|_M^2.$$

In order to tackle the nondifferential term, an auxiliary variable $\boldsymbol{p} = \nabla u$ is introduced, which yields an equivalent minimization problem

$$\min_{u,\boldsymbol{p}} \lambda\|\boldsymbol{p}\| + \frac{c^k}{2}\|u-\hat{u}^k\|_M^2, \qquad s.t.\ \boldsymbol{p} - \nabla u = 0. \tag{3.17}$$

An equivalent augmented Lagrangian is established as follows

$$\mathcal{L}_r(u,\boldsymbol{p};\Lambda) := \lambda\|\boldsymbol{p}\| + \frac{c^k}{2}\|u-\hat{u}^k\|_M^2 + \Re(\langle \boldsymbol{p}-\nabla u, \Lambda\rangle) + \frac{r}{2}\|\boldsymbol{p}-\nabla u\|^2. \tag{3.18}$$

Consequently, one needs to solve

$$\max_{\Lambda}\min_{u,\boldsymbol{p}} \mathcal{L}_r(u,\boldsymbol{p};\Lambda).$$

The ADMM [54, 7] is adopted to solve the above saddle point problem. In particular, the iterations go as follows,

$$\begin{cases} u_{j+1} = (c^k M + r\Delta)^{-1}\left(c^k M\hat{u}^k - r(\boldsymbol{p}_j + \frac{\Lambda_j}{r})\right), \\ \boldsymbol{p}_{j+1} = \text{Thresh}_{soft}\left(\nabla u_{j+1} - \frac{\Lambda_j}{r}; \frac{\lambda}{r}\right), \\ \Lambda_{j+1} = \Lambda_j + r(\boldsymbol{p}_{j+1} - u_{j+1}). \end{cases} \tag{3.19}$$

Please refer to Appendix C for more details. Note that since $c^k M + r\Delta$ is a sparse positive definite matrix, we can employ the conjugate gradient method to solve the u-subproblem. The pseudo-code of P^{3}ALM for Reg-LSPR is listed in Algorithm 2.

Algorithm 2 P^{3}ALM for Reg-LSPR (2.11)

Initialization: $u^0, z^0 = \mathcal{A}u^0, k := 0, M := A^*A$, maximum iteration number MAX_{out}, Max_{in}, and parameters $\sigma, \{c^k\}, \{d^k\}, \lambda, r$.

Output: $u^\star = u^{MAX_{out}-1}$.

for $k = 0$ to $MAX_{out} - 1$ **do**
  $\hat{u}^k := \left(1 - \frac{\sigma}{c^k}\right) u^k + \frac{\sigma}{c^k} M^{-1}\mathcal{A}^* z^k$.
  if Using L^s as regularization term with $s = 0, 1$ **then**
    Solve u^{k+1} by (3.15).
  else
    if Using TV as regularization term **then**
      Set $\boldsymbol{p}_0 = 0, \Lambda_0 = 0$.
      for $j = 0$ to $Max_{in} - 1$ **do**
        $u_{j+1} = (c^k M - r\Delta)^{-1}\left(c^k M\hat{u}^k - r\nabla\cdot(\boldsymbol{p}_j + \frac{\Lambda_j}{r})\right)$,
        $\boldsymbol{p}_{j+1} = \mathrm{Thresh}_{soft}\left(\nabla u_{j+1} - \frac{\Lambda_j}{r}; \frac{\lambda}{r}\right)$,
        $\Lambda_{j+1} = \Lambda_j + r(\boldsymbol{p}_{j+1} - \nabla u_{j+1})$,
      end for
      $u^{k+1} := u_{Max_{in}-1}$.
    end if
  end if
  $z^{k+1} = \mathrm{Prox}_{d^k}^{\mathcal{B}(|\cdot|^2,f)}\left(\left(1 - \frac{\sigma}{d^k}\right) z^k + \frac{\sigma}{d^k}\mathcal{A}u^{k+1}\right)$, where the closed-form solution of the proximal mapping is derived by (3.7) with different metrics.
end for

4. Convergence analysis.

The existing variational methods [35, 56, 46, 20, 12, 13] have focused on efficient computational tools for phase retrieval. However, it is challenging to describe the global convergence for the first-order operator-splitting based algorithms such as PALM and ADMM, due to the lack of Lipschitz differentiable terms. Both LSPR and Reg-LSPR models contain a quadratic term, which is obviously Lipschitz differentiable. Therefore, the global convergence of the proposed algorithm (P^{3}ALM) can be achieved. We remark that our analysis applies to the preconditioned version of the algorithms with a positive definite matrix M as in Definition 3.

4.1. Convergence analysis of P^{3}ALM for LSPR.

The convergence analysis follows the work of Bolte *et al.* [6]. The proof consists three steps. First, we prove the sufficient decrease of the iterative sequences. Second, we show that the subgradient is bounded by successive errors of iterative sequences. Finally, we obtain the global convergence due to the Kurdyka-Łojasiewicz property of the data fitting term $\mathcal{B}(|\cdot|^2, f)$.

We need the following lemma to show that the iterative sequence $\{(u^k, z^k)\}$ generated by (3.8) has monotonically decreasing objective values.

LEMMA 4.1 (Sufficient decreasing).

$$\mathcal{F}_\sigma(u^k, z^k) - \mathcal{F}_\sigma(u^{k+1}, z^{k+1}) \geq (c^k - \tfrac{\sigma}{2})\|u^{k+1} - u^k\|_M^2 + \tfrac{1}{2}(d^k - \sigma)\|z^{k+1} - z^k\|^2. \tag{4.1}$$

REMARK 4.1. *In order to guarantee the monotonic decrease of the objective* *functional, the stepsizes should satisfy*

$$c^k > \frac{\sigma}{2}, \text{ and } \; d^k > \sigma. \tag{4.2}$$

Through experiments, we find choosing $d^k < \sigma$ can further improve the convergence of P^3ALM. We will investigate how to sharpen the lower bounds of the stepsizes in the future. An interesting feature of the proposed algorithm is that the convergence conditions with respect to stepsizes do not rely on the linear mapping $\mathcal{A}$, which seems different to the original PALM algorithm. Although we cannot characterize exactly the relationship between the convergence rate and the stepsizes, we observe from numerical experiments that the convergence behavior of P^3ALM is very robust to the type of $\mathcal{A}$ and the number of measurements.

The proof of Lemma 4.1 is given in Appendix A.3. The lemma gives the following two corollaries.

Corollary 4.1. *If the descent stepsizes $\{c^k, d^k\}$ satisfy (4.2), we have*

1). The sequence $\{\mathcal{F}_\sigma(u^k, z^k)\}$ is monotonically decreasing, and there exists a positive constant $F^\star$, s.t.

$$\lim_{k\to\infty} \mathcal{F}_\sigma(u^k, z^k) = F^\star.$$

Moreover, the iterative sequence $\{(u^k, z^k)\}$ is bounded.

2). The successive error for the iterative sequence $\{(u^k, z^k)\}$ satisfies

$$\sum_{k=0}^{\infty} \left(\|u^{k+1} - u^k\|_M^2 + \|z^{k+1} - z^k\|^2 \right) \leq C, \tag{4.3}$$

with a positive constant C independent with $\{(u^k, z^k)\}$.

Corollary 4.2. *Any accumulated point $(\bar{u}, \bar{z})$ of the iterative sequence $\{(u^k, z^k)\}$ of P^3ALM is a critical point of (2.4).*

See the proof in Appendix A.4.

Lemma 4.2 (Subgradient bounded by successive error). *Assume that the iterative sequence $\{(u^k, z^k)\}$ are generated by P^3ALM, and two variables are introduced as*

$$\begin{aligned} E_u^k &:= (c^{k-1} + \sigma) M (u^{k-1} - u^k) + \sigma \mathcal{A}^* (z^{k-1} - z^k), \\ E_z^k &:= (d^{k-1} + \sigma)(z^{k-1} - z^k). \end{aligned} \tag{4.4}$$

Then we have

$$(E_u^k, E_z^k) \in \partial \mathcal{F}_\sigma(u^k, z^k), \tag{4.5}$$

and it is bounded by the successive error, i.e.

$$\|E_u^k\| + \|E_z^k\| \leq (c^{k-1} + \sigma)\|M^{\frac{1}{2}}\| \|u^k - u^{k-1}\|_M + (\sigma\|M^{\frac{1}{2}}\| + d^{k-1} + \sigma)\|z^k - z^{k-1}\|. \tag{4.6}$$

See the proof in Appendix A.5.

Finally, the convergence analysis is based on the Kurdyka-Łojasiewicz property (Definition 4) [6] for general nonconvex optimization problems.

Definition 4. *Assume that the function $J:\mathbb{R}^d\to(-\infty,+\infty]$ is proper and l.s.c. For $u^\star\in\operatorname{dom}(\partial J)$, if there exists a constant $\bar\eta\in(0,+\infty]$, a neighborhood U of $u^\star$, and a continuous concave function $\phi:[0,\bar\eta)\to\mathbb{R}_+$ satisfying (i) $\phi(0)=0,\phi\in C^1((0,\bar\eta))$, and (ii) $\phi'(s)>0,\ \forall s\in(0,\bar\eta)$, such that*

$$\phi'(J(u)-J(u^\star))\operatorname{dist}(0,\partial J(u))\geq 1\ \forall u\in U\cup\{u:\ J(u^\star)<J(u)<J(u^\star)+\bar\eta\},$$

then we call this function J has the Kurdyka-Łojasiewicz property at $u^\star$ with desingularizing function ϕ. Further the function J is a Kurdyka-Łojasiewicz function if it has Kurdyka-Łojasiewicz property at each point of $\operatorname{dom}(\partial J)$.

Following [12], by lifting the dimension after separating the real and imaginary parts of a complex-valued variable, one readily obtains that $\mathcal{F}_\sigma(u,z)$ is a semi-algebraic function [1] in the cases of IGM and AGM, and it is real-analytic in the case of IPM. Since the quadratic term $\|z-\mathcal{A}u\|^2$ is a semi-algebraic and real-analytic functional, $\mathcal{F}_\sigma$ is a Kurdyka-Łojasiewicz function for all three cases. The boundeness of $\{(u^k,z^k)\}$ can be easily obtained, since $\{\mathcal{F}_\sigma(u^k,z^k)\}$ is bounded due to Lemma 4.1, the function $\mathcal{B}(|\cdot|^2,f)$ is coercive, and M is a positive definite matrix. Further By Lemmas 4.1-4.2, can can readily obtain the convergence theorem following Theorem 1 in [6].

Theorem 5. *If $\{c^k,d^k\}$ satisfy (4.2), the iterative sequence $\{(u^k,z^k)\}$ generated by P^{3}ALM for LSPR (2.4) with any positive definite preconditioning matrix M converge to a critical point of (2.4).*

The convergence rate relies on desingularizing functions of Kurdyka-Łojasiewicz property: $\phi(s)=s^{1-\tau}$ [6] in the cases of IGM and AGM (semi-algebraic functions). Following Remark 6 in [6], we directly have the following corollary.

Corollary 4.3. *Let $\{(u^k,z^k)\}$ be generated by P^{3}ALM for LSPR (2.4) with $\mathcal{B}(\cdot,\cdot)$ being IGM and AGM defined in (2.3), and $\phi(s)=s^{1-\tau}$ be the desingularizing function of $\mathcal{F}_\sigma$. There exists a critical point $(\check u,\check z)$,* i.e. $0\in\partial F_\sigma(\check u,\check z)$, *and two constants $\rho\in[0,1)$ and $\hat C>0$, such that*

- *(i) If $\tau=0$, the sequence $\{(u^k,z^k)\}$ converges to $(\check u,\check z)$ in a finite number of steps.*
- *(ii) If $\tau\in(0,\frac12]$, then $\|u^k-\check u\|+\|z^k-\check z\|\leq\hat C\rho^k$.*
- *(iii) If $\tau\in(\frac12,1)$, then $\|u^k-\check u\|+\|z^k-\check z\|\leq\hat C k^{-\frac{1-\tau}{2\tau-1}}$.*

In the noiseless case, we observe empirically that the iterative sequences generated by P^{3}ALM linearly converges to the critical point of (2.4) (see numerical experiments for details), which is consistent with the case (ii) of the above corollary. Generally speaking, it is difficult to compute the exponent τ for a general nonconvex optimization problem. Only for some special semi-algebraic functions [29], the exponent can be estimated. Unfortunately the analysis cannot be applied to the proposed LSPR and Reg-LSPR.

4.2. Convergence analysis of P^{3}ALM for Reg-LSPR. The convergence analysis for Reg-LSPR is similar to that of P^{3}ALM for LSPR. The only difference is that the constant in front of $\|u^{k+1}-u^k\|_M^2$ is $c^k-\sigma$ for nonconvex L^0 regularization, while it is $c^k-\frac{\sigma}{2}$ for convex models including LSPR, L^1, and TV. Therefore, we present the following lemmas without proof.

Lemma 4.3 (Sufficient decreasing). *Let $\{(u^k,z^k)\}$ be generated by P^{3}ALM of Algorithm 2 for Reg-LSPR, and $\mathcal{R}(u)$ be proper and l.s.c. We have*

$$\mathcal{G}(u^k,z^k)-\mathcal{G}(u^{k+1},z^{k+1})\geq(c^k-\sigma)\|u^{k+1}-u^k\|_M^2+\frac12(d^k-\sigma)\|z^{k+1}-z^k\|^2.$$

Furthermore, if $\mathcal{R}(u)$ is convex,

$$\mathcal{G}(u^k, z^k) - \mathcal{G}(u^{k+1}, z^{k+1}) \geq (c^k - \frac{\sigma}{2})\|u^{k+1} - u^k\|_M^2 + \frac{1}{2}(d^k - \sigma)\|z^{k+1} - z^k\|^2.$$

LEMMA 4.4 (Subgradient bounded by successive error). *Letting the iterative sequences $\{(u^k, z^k)\}$ be generated by P^{3}ALM of Algorithm 2 for Reg-LSPR, for (E_u^k, E_z^k) denoted in (4.4),*

$$(E_u^k, E_z^k) \in \partial\mathcal{G}(u^k, z^k),$$

which is bounded by the successive error of iterative sequences, i.e.

$$\|E_u^k\| + \|E_z^k\| \leq (c^{k-1} + \sigma)\|M^{\frac{1}{2}}\|\|u^k - u^{k-1}\|_M + (\sigma\|M^{\frac{1}{2}}\| + d^{k-1} + \sigma)\|z^k - z^{k-1}\|.$$

Readily we know that $\lambda\mathcal{R}(u) + \mathcal{B}(|z|^2, f)$ with L^s or TV regularizations is a Kurdyka-Łojasiewicz function. Based on Lemmas 4.3-4.4, we conclude the convergence theorem.

Theorem 6. *If $\{c^k, d^k\}$ satisfy (4.2) or $0 < c^k, d^k < \sigma$, then the iterative sequence $\{(u^k, z^k)\}$ generated by P^{3}ALM for Reg-LSPR with any positive definite preconditioning matrix M converge to a critical point of (2.11).*

5. Numerical experiments. We conduct experiments to demonstrate the performance of the proposed methods. The codes are implemented in MATLAB and performed on a laptop with Intel I7-5600U2.6G/16GB RAM. We focus on phase retrieval from Fourier measurements with three kinds of diffraction patterns:

(i) A classical pattern with $\mathcal{A}$ being $\mathbf{F} \in \mathbb{C}^{n\times n}$. Since it is impossible to faithfully recovery the phase without oversampling, we only consider this pattern for compressive phase retrieval by using L^s regularized model ($s = 0, 1$) to enforce sparsity in image domain with additional positivity constraint. The preconditioned matrix $M = \mathbf{F}^*\mathbf{F} = \mathbf{I}$[6].

(ii) Coded diffraction pattern (CDP) [10] with $\mathcal{A}$ defined as,

$$\mathcal{A}u = \left[\mathbf{F}(I_0 \circ u)^T, \mathbf{F}(I_1 \circ u)^T, \cdots, \mathbf{F}(I_{K-1} \circ u)^T\right]^T, \tag{5.1}$$

where $\{I_j\}$ are random masks with $I_j \in \mathbb{C}^n$. We particularly consider the octanary CDP [11] meaning that each element of I_j in (5.1) randomly takes a value among the eight candidates of $\{\pm\sqrt{2}/2, \pm\sqrt{2}\mathbf{i}/2, \pm\sqrt{3}, \pm\sqrt{3}\mathbf{i}\}$. Readily one obtains the preconditioned matrix $M = \mathcal{A}^*\mathcal{A} = \mathrm{diag}\big(\sum_j |I_j|^2\big)$.

(iii) Ptychographic phase retrieval (PtychoPR) pattern [52] with $\mathcal{A}$ defined as

$$\mathcal{A}u = \left[\hat{\mathbf{F}}(\omega \circ u_0)^T, \hat{\mathbf{F}}(\omega \circ u_1)^T, \cdots, \hat{\mathbf{F}}(\omega \circ u_{K-1})^T\right]^T \text{ with } u_j := R_j u, \tag{5.2}$$

where $\hat{\mathbf{F}} \in \mathbb{C}^{\hat{n}\times\hat{n}}$ denotes the normalized discrete Fourier transform over a smaller patch with size $\sqrt{\hat{n}} \times \sqrt{\hat{n}}$, $\omega \in \mathbb{C}^{\hat{n}}$ denotes the illumination function and $R_j \in \mathbb{R}^{n\times\hat{n}}$ is a binary matrix selecting smaller patches with size $\hat{n} = \sqrt{\hat{n}} \times \sqrt{\hat{n}}$, by which the subimage u_j are generated. In our experiments, we set $n = 256 \times 256, \hat{n} = 64 \times 64, K = 16 \times 16$, by setting a sliding distance to 16 pixels, and therefore $m = K \times \hat{n} = 16n$. The preconditioned matrix $M = \mathcal{A}^*\mathcal{A} = \mathrm{diag}\big(\sum_j |R_j^T\omega|^2\big)$.

[6]In this case, P^{3}ALM is exactly the PALM algorithm.

We collect in total m measurements; and we further clarify that $m = n$ for discrete Fourier transform, $m = Kn$ ($K = 2, 3, 4$) for CDP, and $m = 16n$ for PtychoPR. Note that in all three cases, preconditioned matrix $M = \mathcal{A}^*\mathcal{A}$ is diagonal.

Two types of noise distributions are considered. One is additive white Gaussian noise, i.e., $f(t) = |(\mathcal{A}u_g)(t)|^2 + \mathbf{n}(t), \forall\ t \in \Omega$, where u_g is the ground truth image and $\mathbf{n}(t)$ are *i.i.d.* random variables. The other is Poisson noise, $f(t) = \mathrm{Poisson}(|(\mathcal{A}u_\zeta)(t)|^2)$, with the ground truth $u_\zeta = \zeta u$[7].

For the proposed P^{3}ALM (Algorithm 1 and Algorithm 2), the initial value u^0 is randomly generated and we then set $z^0 := \mathcal{A}u^0$. The parameters σ and λ in the model (2.4) and (2.11) are selected manually, and the parameters c^k, d^k are set to $c^k = \sigma/1.5, d^k = \sigma/1.7$ as default values. All the other parameters are problem-dependent and specified in the following subsections. We stop Algorithm 1 and Algorithm 2 if the maximum iteration number reaches MAX_{out}, which will be specified for different problems. In addition to maximum iteration, we stop Algorithm 1 for noiseless cases when the errors between the iteration solution and ground truth[8] reach the given tolerance $TOL = 2 \times 10^{-14}$. We remark that we employ real-valued constraint when recovering the real-valued images. Specifically for Algorithms 1-2, the update of $u-$subproblem becomes $u^{k+1} = \Re(\hat{u}^k)$, which is derived based on

$$u^{k+1} = \arg \min_{\Im(u)=0} \|u - \hat{u}^k\|_M^2.$$

The real-valued constraint is also used for other competing algorithms.

We use the absolute error (AE) and signal-to-noise ratio (SNR) to measure the quality of iterative solution u^k *w.r.t.* the ground truth image u_g, defined as

$$AE := \min_{|\varrho|=1} \|u^k - \varrho u_g\|, \quad \text{and} \qquad \mathrm{SNR} := -20 \log_{10} \min_{|\varrho|=1} \frac{\|u^k - \varrho u_g\|}{\|u^k\|},$$

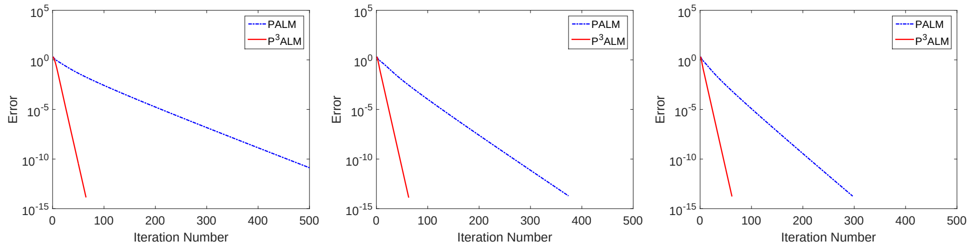
where the global phase shift ϱ is taken into account as a trivial ambiguity. The relative error between the successive iteration, defined as $SE := \frac{\|u^k - u^{k-1}\|}{\|u^k\|}$, is also used to observe the convergence. All the plots regarding above two errors are in a logarithmic scale.

5.1. P^{3}ALM versus PALM. We first demonstrate the efficiency of the preconditioning by comparing P^{3}ALM and PALM using noiseless CDP measurements with $K = 2, 3, 4$. The test image is "Cameraman" of resolution 256×256, as shown in Figure 7 (d). We set $\sigma = 0.01$ and $MAX_{out} = 500$ for both methods. The parameters for PALM in (3.5) are chosen manually to gain best performances, and we set $c^k = \sigma\|\mathcal{A}^*\mathcal{A}\|/2.5, \sigma\|\mathcal{A}^*\mathcal{A}\|/2.3, \sigma\|\mathcal{A}^*\mathcal{A}\|/2.2$, and $d^k = \sigma/1.8, \sigma/1.8, \sigma/1.7$ for $K = 2, 3, 4$, respectively. Other parameters in P^{3}ALM are set as default values. We plot the absolute error in Figure 1, which demonstrates a significant speed-up of P^{3}ALM over the PALM. Therefore, we use P^{3}ALM for the rest of the experimental section.

5.2. P^{3}ALM for LSPR. We test the performance of the proposed model without regularization (2.4) via P^{3}ALM (Algorithm 1) on two diffraction patterns: CDP and PtychoPR. In particular, we examine the noiseless cases in Section 5.2.1, noisy measurements in Section 5.2.2, and different data fitting terms in Section 5.2.3.

[7]Noise level is higher if ζ is smaller.

[8]It is only for the purpose of convergence analysis that we use the error between the recovered image and ground-truth as a stopping condition. We could adopt relative error, but it does not yield a fair comparison, since an algorithm with slower convergence rate may give smaller relative errors.



(a) $K = 2$ (b) $K = 3$ (c) $K = 4$

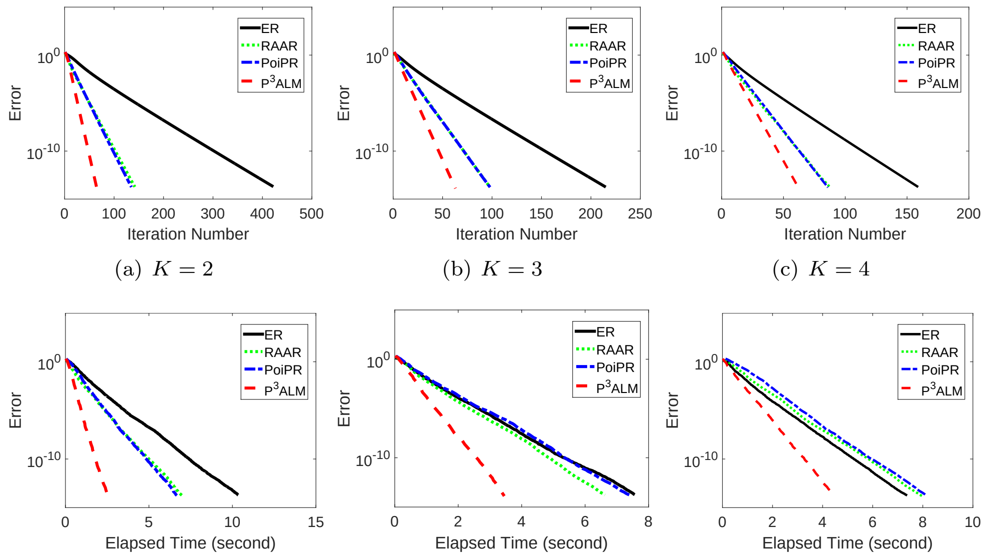
FIG. 1. *Comparison of P^3ALM and PALM for noiseless CDP measurements. All the plots are AE v.s. iteration number in a logarithmic scale.*

5.2.1. Noiseless measurements. We analyze the convergence of the proposed method (Algorithm 1) to retrieve the phase from CDP noiseless measurements with $K = 2, 3, 4$ in (5.1). We include the comparison to ER [24], RAAR [32], and PoiPR [12] with the same random initial guess and the same stopping condition, that is, the error is smaller than 2.0×10^{-14}. For P^{3}ALM, we set $\sigma = 0.01$. Hereafter, parameters in other competing algorithms are tuned to achieve the optimal results. In Figure 2, we plot the absolute error v.s. iteration number in the first row and the absolute error v.s. elapsed time in the second row. We observe that all the methods converge in the noiseless case and the proposed P^{3}ALM is the fastest, ER the slowest, and PoiPR/RAAR have almost identical convergence speed. Specifically in Figure 2 (a), (d) with a limited number of measurements ($K = 2$), we show that P^{3}ALM only requires nearly half of iterations and one-third of computational time to reach the same accuracy compared to RAAR/PoiPR. In addition, ER/RAAR/PoiPR require more iterations to reach the accuracy as the number of measurements decreases, while the iteration number of P^{3}ALM is almost unaffected (one or two additional iterations) by different K values, that implies the robustness of P^{3}ALM *w.r.t.* number of measurements.

We compared with more algorithms, *e.g.* Wirtinger Flow (WF) [11] and Truncated Wirtinger Flow (TWF) [18]. It was reported in [12] that WF/TWF do not work for very few CDP measurements. Therefore, we increase the number of CDP measurements to $K = 12$, sufficiently enough for WF/TWF to work. Please see the compared results in Figure 3. Obviously, Figure 3 (a) shows that P^{3}ALM/RAAR/PoiPR have similar convergent behaviors, while ER/WF/TWF converge relatively slower. According to Figure 3 (b), ER is the fastest in terms of elapsed time, since it only involves two simple projections. P^{3}ALM is slightly slower than ER and is comparable to RAAR/PoiPR; these methods are much faster than WF/TWF. In summary, P^{3}ALM is the most efficient for a limited number of CDP measurements, and as good as PoiPR/RAAR when more measurements are available.

We also conduct the experiment on complex-valued image “Goldballs” with resolution 256×256 for noiseless PtychoPR measurements. We set $\sigma = 0.2, MAX_{out} = 1000, c^k = 0.1, d^k = 0.115$ for P^{3}ALM. The convergence curves are given in Figure 4, which illustrates that P^{3}ALM converges the fastest; specifically it reduces nearly 40% computational time than RAAR/PoiPR. Furthermore, we observe linear convergence of P^{3}ALM, and non-monotonic decreasing behavior of RAAR/PoiPR.

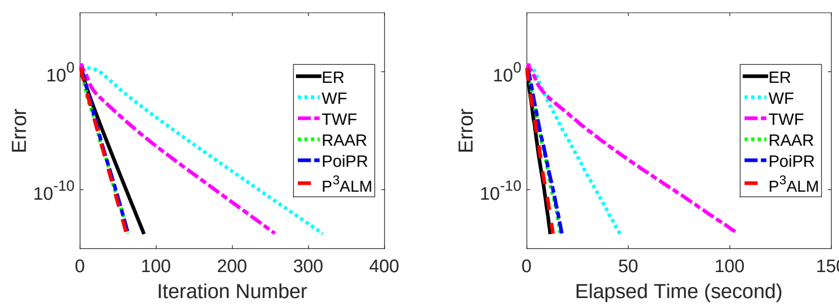
Note that in the noise-free cases, P^{3}ALM converges linearly for solving the LSPR model (2.4) by observing the convergence curves in Figures 2-4, which is consistent with our analysis in Corollary 4.3.



(a) $K = 2$ (b) $K = 3$ (c) $K = 4$

(d) $K = 2$ (e) $K = 3$ (f) $K = 4$

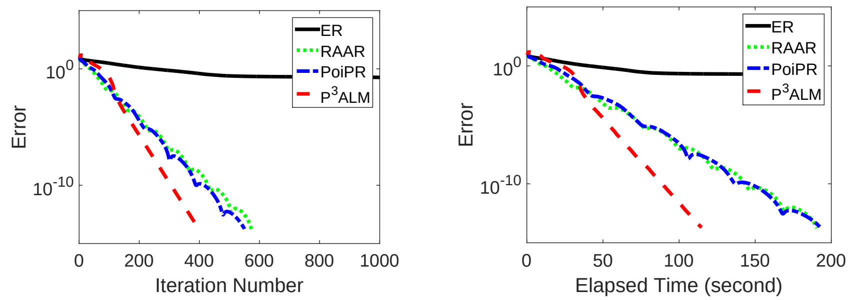
FIG. 2. *Convergence analysis of P^3ALM compared with ER, RAAR and PoiPR using noiseless CDP measurements: AE v.s. iteration number (top) and AE v.s. elapsed time (bottom).*



(a) (b)

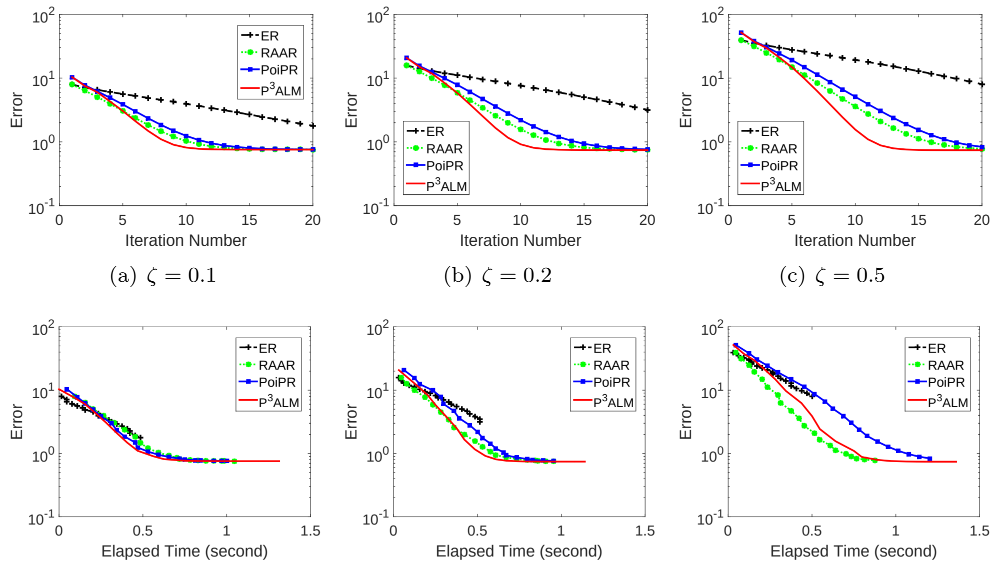
FIG. 3. *Convergence analysis of P^3ALM compared with ER, WF, TWF,RAAR and PoiPR using noiseless CDP measurements with $K = 12$ (WF and TWF do not work for smaller K as in Figure 2). (a) AE v.s. iteration number and (b) AE v.s. elapsed time.*

5.2.2. Noisy measurements. Here we conduct the experiment on phase retrieval from noisy CDP measurements with $K = 2$, in which the noise is assumed to follow Poisson distribution and the noisy level is controlled by the parameter ζ in (5). We examine three noise levels at $\zeta = 0.1, 0.2$ and 0.5. Due to the nature of the Poisson distribution, we use the IPM as the data fitting term and compare the proposed method with ER, RAAR and PoiPR. In Figure 5, we plot the convergence curves within 20 iterations. We observe that all the algorithms except ER reach almost the same accuracy. Moreover, P^{3}ALM requires the smallest number of iterations to converge. As for the computational time, ER is the slowest, and P^{3}ALM/PoiPR/RAAR have similar speed at relatively higher noise level $\zeta = 0.1$. When $\zeta = 0.2, 0.5$, RAAR converges the fastest in the first few iterations and coincides with P^{3}ALM eventually.



(a) (b)

FIG. 4. *Convergence analysis of* P^3*ALM compared with ER, RAAR and PoiPR using noiseless PtychoPR measurements. (a): AE v.s. iteration number and (b): AE v.s. elapsed time.*



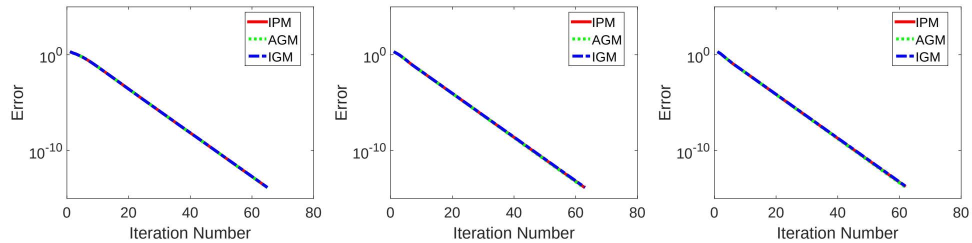
(a) $\zeta = 0.1$ (b) $\zeta = 0.2$ (c) $\zeta = 0.5$

(d) $\zeta = 0.1$ (e) $\zeta = 0.2$ (f) $\zeta = 0.5$

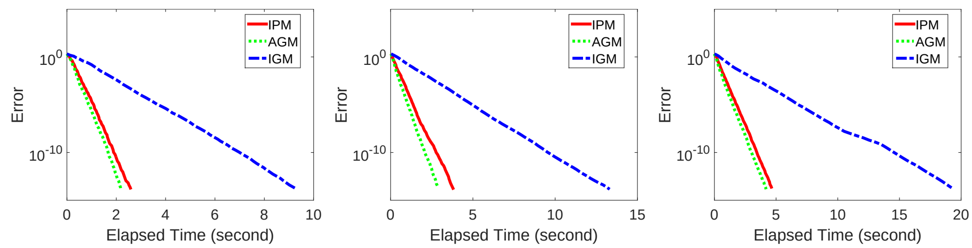
FIG. 5. *Convergence analysis for* P^3*ALM compared with ER, RAAR and PoiPR using Poisson noisy CDP measurements with* $K = 2$*: AE v.s. iteration number (top) and AE v.s. elapsed time (bottom).*

5.2.3. Performances with different data fitting terms. We compare three data fitting terms: AGM, IPM, and IGM in the model (2.4), and show their effects on the convergence of P^{3}ALM . We start by the noiseless CDP measurements with $K = 2, 3, 4$ and set $\sigma = 0.01$ for all the cases. As shown in Figure 6, the proposed algorithm (P^{3}ALM) exhibits almost the same convergent behavior in terms of iteration numbers, while Qian *et al.* [39] claimed that disparate data fitting terms lead to different recovery performances. As for computational time, AGM and IPM reach the desired accuracy faster than IGM, due to the lower cost of computing their proximal operator in (3.7).

We also conduct the experiments for Poisson noisy case; the results are given in



(a) $K = 2$ (b) $K = 3$ (c) $K = 4$



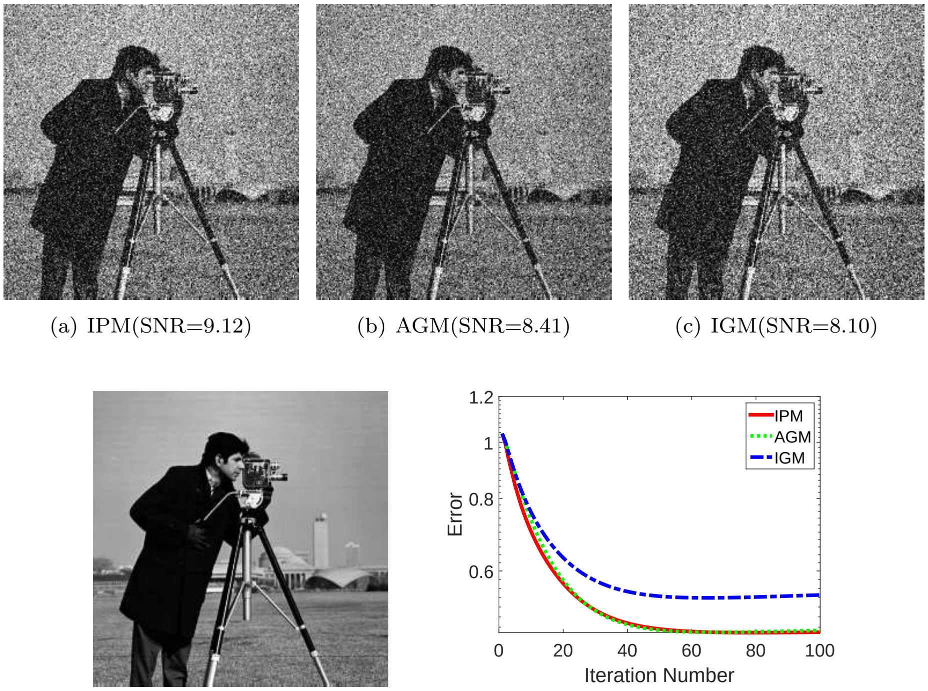
(d) $K = 2$ (e) $K = 3$ (f) $K = 4$

FIG. 6. *Convergence analysis of P^{3}ALM with different data fitting terms (IPM, AGM, and IGM) using noiseless CDP measurements: AE v.s. iteration number (top) and AE v.s. elapsed time (bottom).*

Figure 7. Since it is Poisson noise, IPM gives the best recovery results both visually and in terms of SNR. In Figure 7(e), the error for P^{3}ALM with IGM is larger than the one with IPM/AGM (although AGM surprisingly works here), which demonstrates that one shall compute the minimization problem by utilizing the corresponding MLEs of noise. We will show in Section 5.3 that the noise can be further removed by incorporating regularization terms (2.11).

5.3. P^{3}ALM for Reg-LSPR. We demonstrate the performance of phase retrieval with regularization, i.e., Reg-LSPR (2.11), which can be solved via P^{3}ALM (Algorithm 2). In particular, we consider two types of regularization terms: L^s for sparse signals in Section 5.3.1 and TV for natural images in Section 5.3.2.

5.3.1. L^s regularized (compressive) phase retrieval. It is more interesting and challenging to consider the classical phase retrieval problem, i.e., $\mathcal{A}$ is the discrete Fourier transform. We consider a ground-truth image that can be generated by thresholding a 2D projection of caffeine's electron density map [5] by 0.2 and 0.6 to produce a sparse positive image with only approximately 5% or 0.8% nonzero pixels. Two examples with different sparsity levels are given in Figure 8. In this case, there exist trivial ambiguities such as translation and time reverse, which are compensated to compute the errors. In addition, we incorporate the positivity constraint, $u \geq 0$, in Algorithm 2. We adopt warm-start from [42] to robustify our algorithm, i.e, restarting the iteration with random initializations at most 10 times if the relative residual $\frac{\|\sqrt{f}-|\mathcal{A}u^k|\|}{\|\sqrt{f}\|}$ of recovery image u^k is greater than 0.05. In Figure 8, we present results of L^1 and L^0 regularization terms, both of which give visually satisfactory images; specifically the L^0 regularization can yield exact recovery. We also plot the convergence curves in Figure 9, which shows that P^{3}ALM for L^1 regularized model



(a) IPM(SNR=9.12) (b) AGM(SNR=8.41) (c) IGM(SNR=8.10)

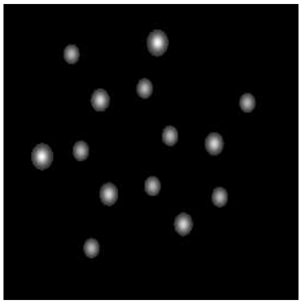
(d) Ground Truth (e)

Fig. 7. *Recovery results from Poisson noisy measurements with $\zeta = 0.01, K = 4$ using P^3ALM with three different data fitting terms: (a) IPM, (b) AGM, and (c) IGM respectively. (d) Ground Truth; (e) AE v.s. iteration number.*

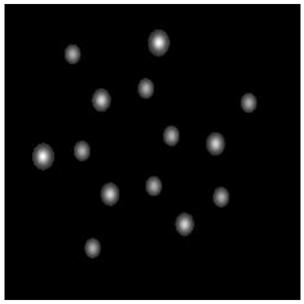
converges faster in the first few iterations and then traps into local minima and L^0 model is more likely to find the global minima.

We also compare both L^0 and L^1 approaches with other compressive phase retrieval algorithms, such as SparseFienup [37] and PRGAMP [42][9]. For a fair comparison, we test these algorithms based on 100 random initializations without warm-restart. PRGAMP assumes that noise exists in a linear transform of the underlying images, not directly in the phaseless measurements, which is different from our setting in (2.1). Therefore we only conduct experiments using the noiseless discrete Fourier transform based measurements. We use the default parameters in PRGAMP and set the total number of iterations of SparseFienup to 1000. For Reg-LSPR with L^1 regularization, we set $\sigma = 0.1, 0.06$, and $\lambda = 1.5 \times 10^{-3}, 5.0 \times 10^{-4}$ at sparsity levels $5\%, 0.8\%$ respectively; For Reg-LSPR with L^0 regularization, set $\sigma = 1$, and $\lambda = 1.5 \times 10^{-3}, 1.0 \times 10^{-3}$ at sparsity levels $5\%, 0.8\%$ respectively. Figure 10 (a)-(b) shows that the L^0 model is the most likely to find the global minima. Specifically, the probability of exact recovery (SNR$>$ 200) using the L^0 model is about 50%, which is much higher than 20% by SparseFienup; and PRGAMP cannot produce results with such a high accuracy. We also look at the probability of recovered SNR larger than 20db, which heuristically gives visually satisfactory results. We observe 90% and 60% for L^0 and L^1 model, respectively; and it is less than 25% for the other two algorithms. We present the results of sparser signals in Figure 10 (c)-(d), which show

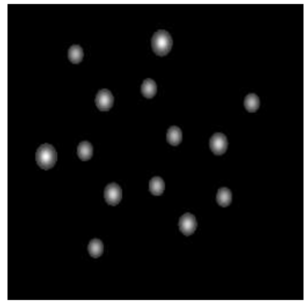
[9] https://sourceforge.net/projects/gampmatlab/



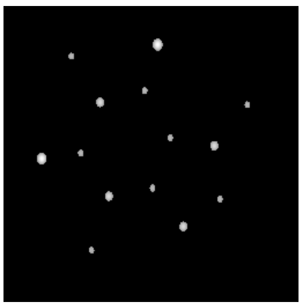
(a) Ground Truth



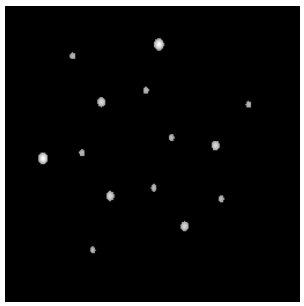
(b) L^1(SNR=26.97)



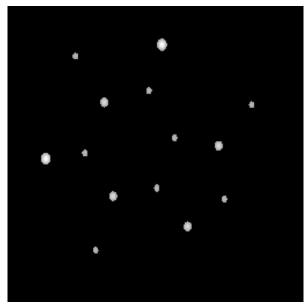
(c) L^0(SNR=294.17)



(d) Ground Truth

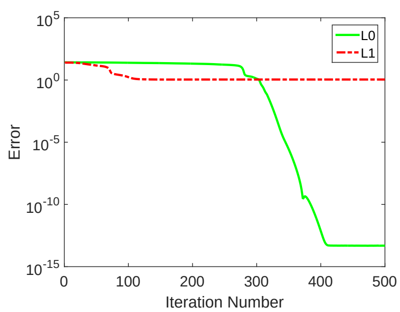


(e) L^1(SNR=38.25)

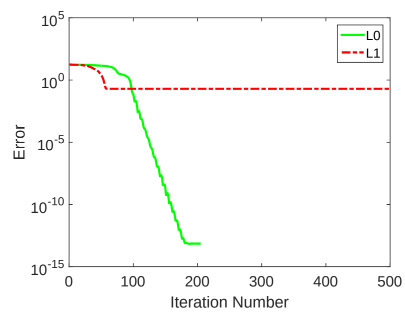


(f) L^0(SNR=287.32)

FIG. 8. *Performances of P^3ALM for both L^0 and L^1 regularized models for noiseless measurements. Two sparsity levels of the grount-truth images are considered: 5% nonzero pixels (top) and 0.8% nonzero pixels (bottom).*



(a)

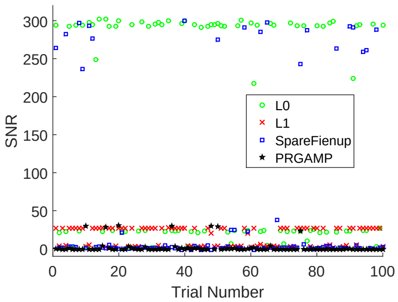


(b)

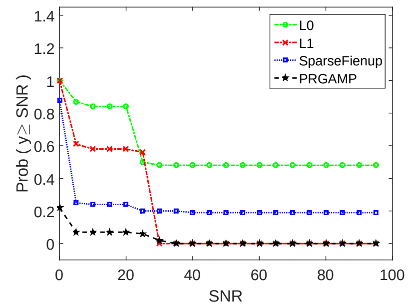
FIG. 9. *Convergence curves (AE w.r.t. iteration number) for L^s regularized phase retrieval via P^3ALM from noiseless measurements: (a) 5% nonzero pixels and (b) 0.8% nonzero pixels.*

that the L^0 model gives almost 100% exact recovery and the L^1 approach produces visual acceptable results (SNR$\geq$ 20) at about 70%. Notice that PRGAMP is more sensitive to the sparsity level than the other competing algorithms.

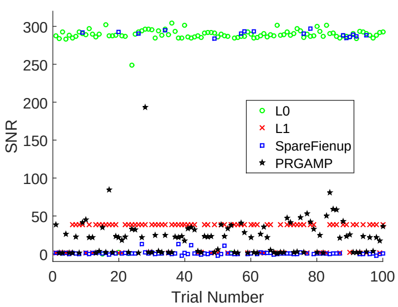
5.3.2. TV regularized phase retrieval. We consider noisy CDP measurements, which are corrupted by either Gaussian or Poisson noises. We start with Gaussian noisy measurements, where the SNRs of the noisy data are set to 5, 10, and 20. We consider TV regularized model with IPM data fitting term solved by Algorithm 2.



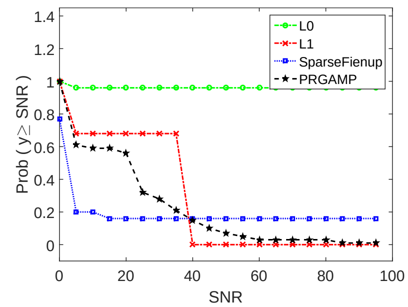
(a)



(b)



(c)



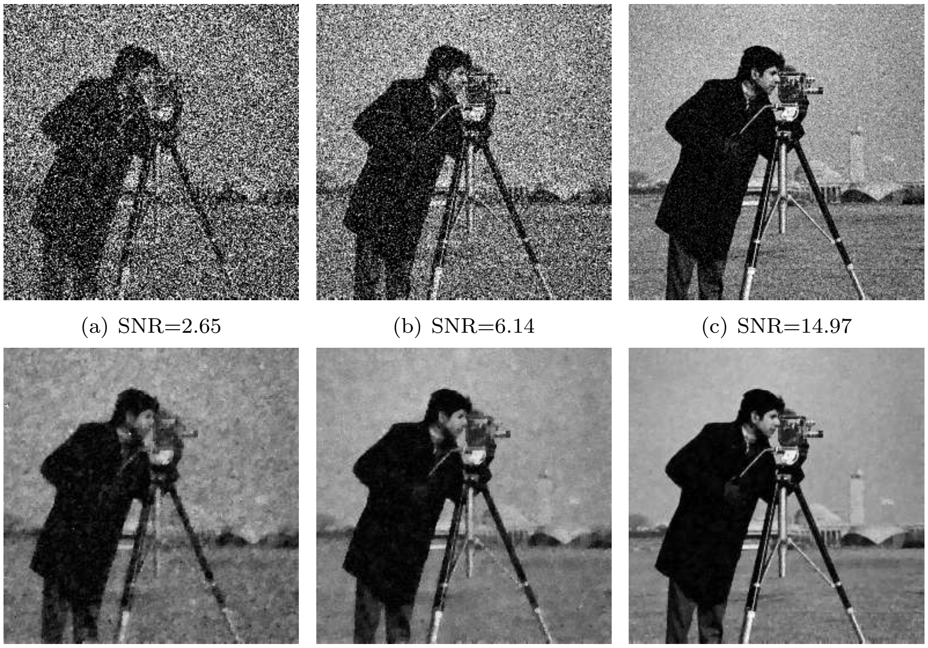
(d)

FIG. 10. *Performances with different random initializations (100 trials). (a)-(b):* 5% *nonzero pixels; (c)-(d):* 0.8% *nonzero pixels. (a),(c): SNR w.r.t. trial number; (b),(d): the probability (y−axis) of the resulting SNR larger than the given SNR (x−axis) based on 100 trials.*

We set $MAX_{out} = 50, MAX_{in} = 2$, and $\sigma = 6, 8, 10$, $\lambda = 5.0\times 10^4, 5.0\times 10^4, 2.0\times 10^4$ for different noise levels. In Figure 11, we include the comparison to P^{3}ALM without regularization (Algorithm 1) by setting set $\sigma = 5$. P^{3}ALM without regularization fails to remove noises, while P^{3}ALM with TV regularization yields clean background and sharp edges, with about 10dB increase in SNR.

Numerical experiments are performed for Poisson noisy measurements as well. For TV regularized model, we set $\sigma = 30, 20, 20$ and $\lambda = 200, 180, 170$ for noisy data with different levels $\zeta = 0.005, 0.1, 0.2$, respectively, and $\sigma = 500, 100, 5$ for P^{3}ALM without regularization (Algorithm 1). In Figure 12, we include the comparison of TVPoiPR [12]. Both P^{3}ALM with regularization and TVPoiPR greatly improve the recovery quality compared to P^{3}ALM without regularization, and at least 10dB increase is gained by regularization. Since both P^{3}ALM and TVPoiPR involve total variation regularization, there is no obvious difference between their recovery results. We also plot the convergence curves in Figure 13 for $\zeta = 0.005$. In particular, Figure 13 (a) shows that P^{3}ALM produces results with a bit higher SNR than TVPoiPR as iteration goes, and Figure 13 (b) illustrates that P^{3}ALM converges much faster than TVPoiPR, since its successive errors decreases more rapidly than TVPoiPR due to global convergence guarantee.

5.4. Impact by Parameters. In order to show the impact by different parameters, we conduct the experiments of P^{3}ALM by changing one parameter and fixing the others. We first consider the LSPR (2.4) via P^{3}ALM (Algorithm 1) using CDP measurements. The parameters are chosen as $\sigma \in \{1.0\times 10^{-5}, 1.0\times 10^{-4}, 1.0\times 10^{-3}, 1.0\times$



(a) SNR=2.65 (b) SNR=6.14 (c) SNR=14.97

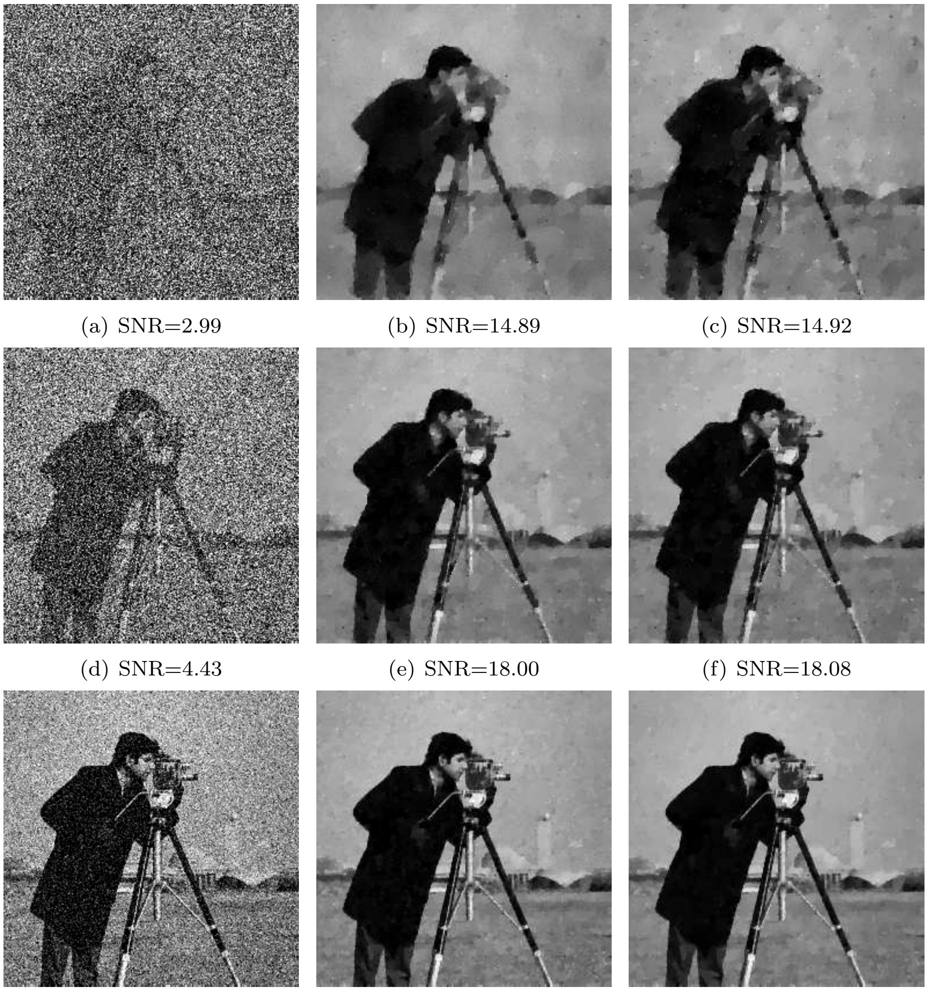
(d) SNR=12.26 (e) SNR=16.83 (f) SNR=23.11

FIG. 11. *Comparison of P^3ALM without any regularization (top) and with the TV regularization (bottom) for Gaussian noisy measurements with SNR=5, 10, 20 from left to right.*

$10^{-2}, 1.0 \times 10^{-1}, 1.0, 1.0 \times 10^{-1}\}$ and $c^k, d^k \in \{0.5, 0.55, 0.6, 0.65, 0.7, 0.8, 1\} \times \sigma$. All algorithms stop if the error is smaller than 2.0×10^{-14} or maximum iteration numbers reach 500 in the noiseless case, and stop after 50 iterations for Poisson noisy measurements. It is shown in Figure 14 (a) and (d) for both the noiseless and noisy cases that Algorithm 1 is insensitive to σ, except that a relatively large value ($\sigma = 10$) slows down the convergence speed. The other plots in Figure 14 suggest that a moderate value of stepsize c^k and d^k is better to gain fast convergence. Note that if we choose $c^k = \sigma$ or $d^k = \sigma$ then P^{3}ALM is equivalent to the alternating minimization method, which converges slowly, as discussed in subsection 3.3 and Remark 3.2. How to choose the optimal stepsizes will be our future work. In addition, although we prove in Theorem 5 that the stepsize d^k should less than σ, we observe empirically that the algorithm with $d^k > \sigma/2$ also converges. It will be our future work to analyze how to choose stepsizes and get a sharper bound in convergence analysis.

Further tests are performed to analyze the parameters in P^{3}ALM (Algorithm 2) to solve TV-regularized model (2.11). Since the impact by descent stepsizes is similar to Figure 14, we illustrate the impact by parameters λ and σ in Figure 15, where $\sigma \in \{10, 20, 40, 80, 160\}$ and $\lambda \in \{50, 100, 200, 400, 800\}$. Figure 15 (a) shows that a large σ tends to slow down the algorithm, while Figure 15 (b) shows that a moderate value of $\lambda \sim 200$ should be chosen to achieve high recovery accuracy.

5.5. Comparison to block-Kaczmarz method. One reviewer brings our attention to a "block-Kaczmarz method" for phase retrieval proposed by Wei [51]. Here is a brief review of the method. If objective functional in (3.3) has a block-



(a) SNR=2.99 (b) SNR=14.89 (c) SNR=14.92

(d) SNR=4.43 (e) SNR=18.00 (f) SNR=18.08

(g) SNR=10.26 (h) SNR=21.09 (i) SNR=21.12

FIG. 12. *Comparison of LSPR without regularization (left), TVPoiPR (middle), and TV regularized model (right) for phase retrieval from noisy Poisson measurements with* $\zeta = 0.005, 0.01, 0.02$, *from top to bottom.*

wise structure with $\mathcal{A} := (\mathcal{A}_0^T, \mathcal{A}_1^T, \cdots, \mathcal{A}_{K-1}^T)^T$, $z = (z_0^T, z_1^T, \cdots, z_{K-1}^T)$ and $f = (f_0^T, f_1^T, \cdots, f_{K-1}^T)^T$, which satisfy

$$|\mathcal{A}_j u|^2 = f_j. \tag{5.3}$$

By assuming that $\mathcal{A}_j$ has full column rank[10], the block-Kaczmarz method for (1.1) is given below:

$$\begin{aligned} u^{k+1} &= u^k + (\mathcal{A}_{n_k}^* \mathcal{A}_{n_k})^{-1} \mathcal{A}_{n_k}^* (\sqrt{f_{n_k}} \circ \mathrm{sign}(\mathcal{A}_{n_k} u^k) - \mathcal{A}_{n_k} u^k) \\ &= (\mathcal{A}_{n_k}^* \mathcal{A}_{n_k})^{-1} \mathcal{A}_{n_k}^* \left(\sqrt{f_{n_k}} \circ \mathrm{sign}(\mathcal{A}_{n_k} u^k) \right), \end{aligned} \tag{5.4}$$

[10]If $\mathcal{A}_j$ does not have full column rank, its Moore-Penrose pseudo-inverse matrix can be considered [51].

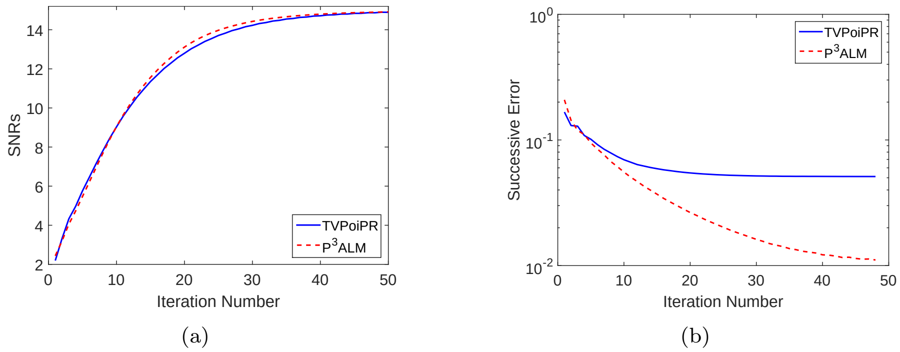


FIG. 13. *Comparison between the TV-regularized model via P^{3}ALM compared and TVPoiPR: (a) SNRs v.s. iteration number, and (b) The relative error between the successive iteration (SE) v.s. iteration number.*

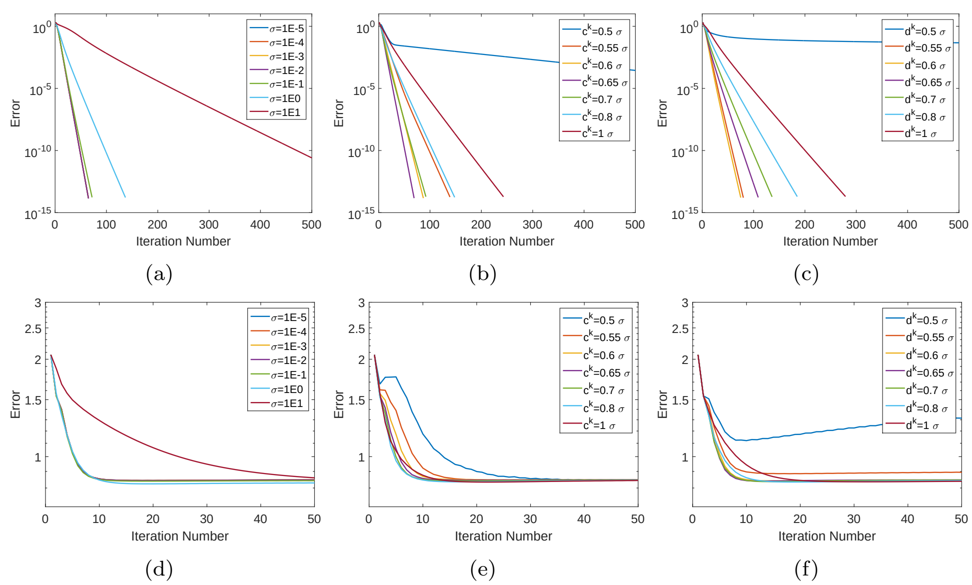
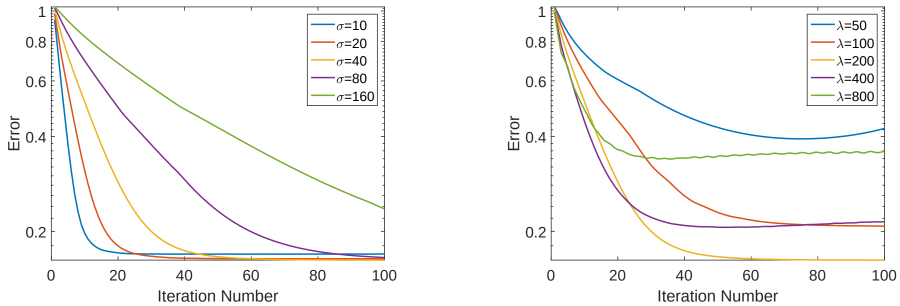


FIG. 14. *Impact by parameters σ (left), c^k (middle), d^k (right) of P^{3}ALM (Algorithm 1) for noiseless CDP measurements (top) and Poisson noisy measurements with $\zeta = 0.02$ (bottom).*

with the random index $n_k \in \{0, 1, \cdots, K-1\}$, which can also be interpreted as ER algorithm for solving the subproblem (5.3). Recently, Chen *et al.* [17] proposed a serial alternating projection algorithm in the case of isometric matrices as $\mathcal{A}_j^* \mathcal{A}_j = \mathbf{I}$, which is a special case of block-Kaczmarz method (5.4).

We compare the proposed P^{3}ALM (Algorithm 1) with the block-Kaczmarz method for both noiseless and noisy CDP measurements in terms of convergent behaviors in Figure 16. In order to use fast Fourier transform for the subproblem of each block for the block-Kaczmarz method, we denote $\mathcal{A}_j$ as $\mathcal{A}_j u := \mathbf{F}(I_j \circ u)$. In the noiseless case, one can readily see that with very few measurements as $K = 2$, P^{3}ALM is much faster than block-Kaczmarz method, while it is the other way around if the number of measurements increases as $K = 4$. For noisy measurements, both two algorithms con-



(a) (b)

FIG. 15. *Impact by parameters σ (left), λ (right) of P^3ALM (Algorithm 2) for Poisson noisy CDP measurements with $\zeta = 0.01$.*

TABLE 1
Absolution errors (AEs) and signal-to-noise ratio (SNRs) of recovered images by block-Kaczmarz method (BKM [51]) and the proposed P^3ALM from noisy CDP measurements with $K = 2, 3, 4$.

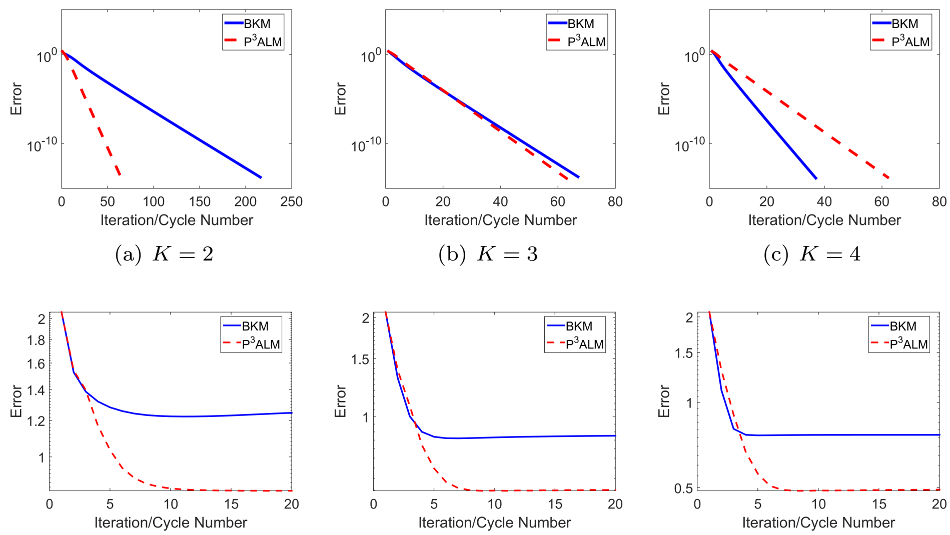
K		2	3	4
AEs	BKM	1.19	0.85	0.74
	P^3ALM	0.82	0.58	0.48
SNRs (dB)	BKM	7.17	9.93	11.04
	P^3ALM	10.30	13.06	14.74

verge in almost the same speed within the first few iterations while block-Karczmarz method is stuck with less accurate solutions; this phenomenon was also reported in [17]. We report the SNRs of recovered images by two compared algorithms in Table 1, which shows that recovery results by P^3ALM are at least 3dB higher than by the block-Kaczmarz method.

During the course of this experiment, we observe that the block type algorithms [51, 17] increase the convergence speed when more noiseless measurements are used. However, block based acceleration is not trivial. In the future, we should explore how to further accelerate the proposed P^3ALM by employing the block structures for both noiseless and noisy measurements while maintaining recovery accuracy.

6. Conclusions and future works. In this paper, we established two new models "LSPR" and "Reg-LSPR" for phase retrieval. Especially, for noiseless case, the equivalence between LSPR and original phase retrieval problem (1.1) was derived. Computationally, we designed P^3ALM algorithms to solve these two models, which significantly speed up the plain PALM method thanks to the preconditioning technique and the diagonal structure of $\mathcal{A}^*\mathcal{A}$ for the masked Fourier transform. The existence of a quadratic term in the proposed models helped to establish the global convergence to the critical point, based on the Kurdyka-Łojasiewicz property. Numerous experiments demonstrated the convergence and efficiency of the proposed algorithms.

We are interested in the geometric structures of critical points for the proposed models to further classify critical points into global minimizers or saddle points. Along this line of research, Sun, Qu, and Wright [45] analyzed the geometric properties of



(a) $K = 2$ (b) $K = 3$ (c) $K = 4$

(d) $K = 2$ (e) $K = 3$ (f) $K = 4$

FIG. 16. *Convergence histories of P^3ALM and block-Kaczmarz method (BKM [51]) with block size* 256×256*: AE v.s. iteration/cycle number for noiseless measurements (top) and Poisson noisy measurements with* $\zeta = 0.02$ *(bottom). For noiseless case (top), the two algorithms stop if the maximum iteration number reaches 1000 or the errors between the iteration solution and ground truth reach the given tolerance* $TOL = 2 \times 10^{-14}$*; For noisy case (bottom), the two algorithms stop if the maximum iteration number reaches 20.*

critical points for a general phase retrieval problem[11]. They revealed that with high probability, there are no spurious local minimizers, and the global minimizer is unique up to global phase factor. In this paper, we focused on masked Fourier transform, which is quite different from the setting in [45]. Our future work involves geometric analysis of critical points for phase retrieval from masked Fourier measurements with or without noise.

Since our proposed Algorithm 2 for Reg-LSPR model with TV regularization requires inner iterations, we want to improve the algorithm by removing the nested optimization. One approach is to adopt the primal-dual scheme [48]. We expect it is possible to explore the global convergence of such primal-dual scheme under the condition that the iterates stay bounded.

Acknowledgement. We thank the two reviewers for their valuable comments, and bringing our attentions to some important references, which helps to improve our paper greatly. Dr. H. Chang was partially supported by NSFC Nos. 11501413 and 11426165, China Scholarship Council (CSC), and Innovation Project No. 52XC1605 of Tianjin Normal University. Dr. Y. Lou was partially supported by NSF grant DMS-1522786. Dr. T. Zeng was partially supported by NSFC 11271049, RGC 12302714 and RFGs of HKBU. This work was also partially funded 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-DE-AC03-76SF00098.

[11] $\mathcal{B}$ by IGM and $\mathcal{A}$ following *i.d.d.* complex Gaussian distribution

REFERENCES

[1] Hédy Attouch, Jérôme Bolte, Patrick Redont, and Antoine Soubeyran, *Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the kurdyka-łojasiewicz inequality*, Mathematics of Operations Research, 35 (2010), pp. 438–457.

[2] Radu Balan, Pete Casazza, and Dan Edidin, *On signal reconstruction without phase*, Applied and Computational Harmonic Analysis, 20 (2006), pp. 345–356.

[3] Afonso S Bandeira, Jameson Cahill, Dustin G Mixon, and Aaron A Nelson, *Saving phase: Injectivity and stability for phase retrieval*, Applied and Computational Harmonic Analysis, 37 (2014), pp. 106–125.

[4] Ahron Ben-Tal and Arkadi Nemirovski, *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*, vol. 2, Siam, 2001.

[5] Helen M Berman, Tammy Battistuz, TN Bhat, Wolfgang F Bluhm, Philip E Bourne, Kyle Burkhardt, Zukang Feng, Gary L Gilliland, Lisa Iype, Shri Jain, et al., *The protein data bank*, Acta Crystallographica Section D: Biological Crystallography, 58 (2002), pp. 899–907.

[6] Jérôme Bolte, Shoham Sabach, and Marc Teboulle, *Proximal alternating linearized minimization for nonconvex and nonsmooth problems*, Mathematical Programming, 146 (2014), pp. 459–494.

[7] 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.

[8] T Tony Cai, Xiaodong Li, and Zongming Ma, *Optimal rates of convergence for noisy sparse phase retrieval via thresholded wirtinger flow*, The Annals of Statistics, 44 (2016), pp. 2221–2251.

[9] Emmanuel J Candés, Yonina C Eldar, Thomas Strohmer, and Vladislav Voroninski, *Phase retrieval via matrix completion*, SIAM Journal on Imaging Sciences, 6 (2013), pp. 199–225.

[10] Emmanuel J Candes, Xiaodong Li, and Mahdi Soltanolkotabi, *Phase retrieval from coded diffraction patterns*, Applied and Computational Harmonic Analysis, 39 (2015), pp. 277–299.

[11] ———, *Phase retrieval via wirtinger flow: Theory and algorithms*, IEEE Transactions on Information Theory, 61 (2015), pp. 1985–2007.

[12] Huibin Chang, Yifei Lou, Yuping Duan, and Stefano Marchesini, *Total variation based phase retrieval for poisson noise removal*, preprint, CAM report 16-76, (2016).

[13] 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.

[14] Huibin Chang and Stefano Marchesini, *Denoising poisson phaseless measurements via orthogonal dictionary learning*, arXiv preprint arXiv:1612.08656, (2016).

[15] ———, *A general framework for denoising phaseless diffraction measurements*, arXiv preprint arXiv:1611.01417, (2016).

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

[17] 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).

[18] Yuxin Chen and Emmanuel 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.

[19] Aldo Conca, Dan Edidin, Milena Hering, and Cynthia Vinzant, *An algebraic characterization of injectivity in phase retrieval*, Applied and Computational Harmonic Analysis, 38 (2015), pp. 346–356.

[20] Yuping Duan, Chunlin Wu, Zhi-Feng Pang, and Huibin Chang, *L0-regularized variational methods for sparse phase retrieval*, arXiv preprint arXiv:1612.02538, (2016).

[21] Yonina C Eldar and Shahar Mendelson, *Phase retrieval: Stability and recovery guarantees*, Applied and Computational Harmonic Analysis, 36 (2014), pp. 473–494.

[22] Ernie Esser, Xiaoqun Zhang, and Tony F Chan, *A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science*, SIAM Journal on Imaging Sciences, 3 (2010), pp. 1015–1046.

[23] James R Fienup, *Phase retrieval algorithms: a comparison*, Applied optics, 21 (1982),

pp. 2758–2769.
[24] Ralph W Gerchberg, *A practical algorithm for the determination of phase from image and diffraction plane pictures*, Optik, 35 (1972), p. 237.
[25] MHMH Hayes, *The reconstruction of a multidimensional sequence from the phase or magnitude of its fourier transform*, IEEE Transactions on Acoustics, Speech, and Signal Processing, 30 (1982), pp. 140–154.
[26] Robert Hesse, D Russell Luke, Shoham Sabach, and Matthew 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.
[27] Mark Iwen, Aditya Viswanathan, and Yang Wang, *Robust sparse phase retrieval made easy*, Applied and Computational Harmonic Analysis, 42 (2017), pp. 135–142.
[28] Adrian S Lewis, David Russel Luke, and Jérôme Malick, *Local linear convergence for alternating and averaged nonconvex projections*, Foundations of Computational Mathematics, 9 (2009), pp. 485–513.
[29] Guoyin Li and Ting Kei Pong, *Calculus of the exponent of kurdyka-{\L} ojasiewicz inequality and its applications to linear convergence of first-order methods*, arXiv preprint arXiv:1602.02915, (2016).
[30] Xiaodong Li and Vladislav Voroninski, *Sparse signal recovery from quadratic measurements via convex programming*, SIAM Journal on Mathematical Analysis, 45 (2013), pp. 3019–3033.
[31] Stefan Loock and Gerlind Plonka, *Phase retrieval for fresnel measurements using a shearlet sparsity constraint*, Inverse Problems, 30 (2014), p. 055005.
[32] D Russell Luke, *Relaxed averaged alternating reflections for diffraction imaging*, Inverse Problems, 21 (2005), p. 37.
[33] Stefano Marchesini, *Invited article: A unified evaluation of iterative projection algorithms for phase retrieval*, Review of scientific instruments, 78 (2007), p. 011301.
[34] S. Marchesini, Y.-C. Tu, and H.-T. Wu, *Alternating projection, ptychographic imaging and phase synchronization*, Applied and Computational Harmonic Analysis, (2015).
[35] Matthew L Moravec, Justin K Romberg, and Richard G Baraniuk, *Compressive phase retrieval*, in Optical Engineering+ Applications, International Society for Optics and Photonics, 2007, pp. 670120–670120.
[36] Boris S Mordukhovich, *Variational analysis and generalized differentiation I: Basic theory*, vol. 330, Springer Science & Business Media, 2006.
[37] Subhadip Mukherjee and Chandra Sekhar Seelamantula, *Fienup algorithm with sparsity constraints: application to frequency-domain optical-coherence tomography*, IEEE Transactions on Signal Processing, 62 (2014), pp. 4659–4672.
[38] Henrik Ohlsson, Allen Yang, Roy Dong, and Shankar Sastry, *Cprl–an extension of compressive sensing to the phase retrieval problem*, in Advances in Neural Information Processing Systems, 2012, pp. 1367–1375.
[39] Jianliang Qian, Chao 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.
[40] Leonid I Rudin, Stanley Osher, and Emad Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D: Nonlinear Phenomena, 60 (1992), pp. 259–268.
[41] Jorge LC Sanz, *Mathematical considerations for the problem of fourier transform phase retrieval from magnitude*, SIAM Journal on Applied Mathematics, 45 (1985), pp. 651–664.
[42] Philip Schniter and Sundeep Rangan, *Compressive phase retrieval via generalized approximate message passing*, IEEE Transactions on Signal Processing, 63 (2015), pp. 1043–1055.
[43] Yoav Shechtman, Amir Beck, and Yonina C Eldar, *Gespar: Efficient phase retrieval of sparse signals*, IEEE transactions on signal processing, 62 (2014), pp. 928–938.
[44] Yoav Shechtman, Yonina C Eldar, Oren Cohen, Henry Nicholas Chapman, Jianwei Miao, and Mordechai Segev, *Phase retrieval with application to optical imaging: a contemporary overview*, IEEE Signal Processing Magazine, 32 (2015), pp. 87–109.
[45] Ju Sun, Qing Qu, and John Wright, *A geometric analysis of phase retrieval*, in Information Theory (ISIT), 2016 IEEE International Symposium on, IEEE, 2016, pp. 2379–2383.
[46] P Thibault and M Guizar-Sicairos, *Maximum-likelihood refinement for coherent diffractive imaging*, New Journal of Physics, 14 (2012), p. 063004.
[47] Andreas M Tillmann, Yonina C Eldar, and Julien Mairal, *Dolphin-dictionary learning for phase retrieval*, arXiv preprint arXiv:1602.02263, (2016).
[48] Tuomo Valkonen, *A primal–dual hybrid gradient method for nonlinear operators with applications to mri*, Inverse Problems, 30 (2014), p. 055012.
[49] I. Waldspurger, A. Aspremont, and S. Mallat, *Phase recovery, maxcut and complex*

semidefinite programming, Math. Program., Ser. A, (DOI 10.1007/s10107-013-0738-9, 2012), pp. 1–35.
[50] Gang Wang, Liang Zhang, Georgios B Giannakis, Mehmet Akçakaya, and Jie Chen, *Sparse phase retrieval via truncated amplitude flow*, arXiv preprint arXiv:1611.07641, (2016).
[51] Ke Wei, *Solving systems of phaseless equations via kaczmarz methods: A proof of concept study*, Inverse Problems, 31 (2015), p. 125008.
[52] Zaiwen Wen, Chao Yang, Xin Liu, and Stefano Marchesini, *Alternating direction methods for classical and ptychographic phase retrieval*, Inverse Probl., 28 (2012), p. 115010.
[53] Stephen Wright and Jorge Nocedal, *Numerical optimization*, Springer Science, 35 (1999), pp. 67–68.
[54] Chunlin Wu and Xue-Cheng Tai, *Augmented lagrangian method, dual methods, and split bregman iteration for rof, vectorial tv, and high order models*, SIAM Journal on Imaging Sciences, 3 (2010), pp. 300–339.
[55] Chunlin Wu, Juyong Zhang, and Xue-Cheng Tai, *Augmented lagrangian method for total variation restoration with non-quadratic fidelity*, Inverse problems and imaging, 5 (2011), pp. 237–261.
[56] Zai Yang, Cishen Zhang, and Lihua Xie, *Robust compressive phase retrieval via l1 minimization with application to image reconstruction*, arXiv preprint arXiv:1302.0081, (2013).

Appendix A. Proofs.

A.1. Proof of Theorem 1. *Proof.* Since $f \geq 0$, it is straightforward that

$$f = \arg\min_g \mathcal{B}(g, f), \text{[12]} \tag{A.1}$$

with $\mathcal{B}(\cdot,\cdot)$ defined in (2.3), which also implies that the set $\arg\min\limits_g \mathcal{B}(g,f)$ has a unique element based on simple calculation. We first prove that if $\mathcal{S}(f)$ is not empty, then $\min\limits_{u,z} \mathcal{F}_\sigma(u,z) = \mathcal{B}(f,f)$. Suppose $\check{u} \in \mathcal{S}(f)$ and let $\check{z} = \mathcal{A}\check{u}$, then we have $|\check{z}|^2 = f$ and $\mathcal{F}_\sigma(\check{u},\check{z}) = \mathcal{B}(f,f)$. On the other hand, simple calculations give that

$$\begin{aligned} \min_{u,z} \mathcal{F}_\sigma(u,z) &= \min_{u,z} \left(\tfrac{\sigma}{2}\|z - \mathcal{A}u\|^2 + \min_z \mathcal{B}(|z|^2, f) \right) \\ &\geq \tfrac{\sigma}{2} \min_{u,z} \|z - \mathcal{A}u\|^2 + \min_z \mathcal{B}(|z|^2, f) = \mathcal{B}(f,f). \end{aligned} \tag{A.2}$$

Therefore, we have $\min\limits_{u,z} \mathcal{F}_\sigma(u,z) = \mathcal{B}(f,f)$. Now we are ready to prove these two cases in this theorem.

1) Since $(u^\star, z^\star) \in \arg\min\limits_{u,z} \mathcal{F}_\sigma(u,z)$, and $\mathcal{S}(f)$ is nonempty, we have $\mathcal{F}_\sigma(u^\star, z^\star) = \mathcal{B}(f,f)$. Further we have

$$0 = \tfrac{\sigma}{2}\|z^\star - \mathcal{A}u^\star\|^2 + \mathcal{B}(|z^\star|^2, f) - \mathcal{B}(f,f) \overset{(A.1)}{\geq} \tfrac{\sigma}{2}\|z^\star - \mathcal{A}u^\star\|^2.$$

Hence $z^\star = \mathcal{A}u^\star$, and $\mathcal{B}(|z^\star|^2, f) = \mathcal{B}(f,f)$. Immediately we have $u^\star \in \mathcal{S}(f)$ based on (A.1).

2) $\forall\ u^\star \in \mathcal{S}(f)$, we have $\mathcal{F}_\sigma(u^\star, \mathcal{A}u^\star) = \mathcal{B}(f,f)$, which implies that $(u^\star, \mathcal{A}u^\star)$ is a minimizer of $\mathcal{F}_\sigma$ based on (A.2). It also demonstrates that $\arg\min_{u,z} \mathcal{F}_\sigma(u,z) \cap \{(u^\star, z) :\ z \in \mathbb{C}^m\}$ is nonempty. Hence, for any variable $z^\star$ satisfying $(u^\star, z^\star) \in \arg\min \mathcal{F}_\sigma(u,z)$, we have $\mathcal{F}_\sigma(u^\star, z^\star) = \mathcal{F}_\sigma(u^\star, \mathcal{A}u^\star) = \mathcal{B}(f,f)$. It immediately gives

$$\tfrac{\sigma}{2}\|z^\star - \mathcal{A}u^\star\|^2 = \mathcal{B}(f,f) - \mathcal{B}(|z^\star|^2, f) \overset{(A.1)}{\leq} 0,$$

such that $z^\star = \mathcal{A}u^\star$. ☐

[12] Recall that $\arg\min\limits_w \mathcal{G}(w) := \{w^\star : \mathcal{G}(w) \geq \mathcal{G}(w^\star),\ \forall\ w\}$.

A.2. Proof of Theorem 2. *Proof.* For any $\check{u} \in \mathcal{S}(f)$, let $\check{z} := \mathcal{A}\check{u}$, which immediately implies that $\mathcal{A}^*(\mathcal{A}\check{u} - \check{z}) = 0$, i.e., $0 \in \nabla_u \mathcal{F}_\sigma(\check{u}, \check{z})$. On the other hand, we have $|\check{z}|^2 = f$. It follows from the proof of Theorem 1 that $f = \arg\min_g \mathcal{B}(g, f)$, thus leading to $0 \in \partial_g \mathcal{B}(f, f)$. Therefore, we have

$$\partial_z \mathcal{F}(\check{u}, \check{z}) \overset{(2.6)}{=} \sigma(\check{z} - A\check{u}) + \partial_g \mathcal{B}(g, f)\big|_{g=|\check{z}|^2} \circ \check{z} = \partial_g \mathcal{B}(f, f) \circ \check{z} \ni 0.$$

Based on (2.7), we conclude this theorem. ∎

A.3. Proof of Lemma 4.1. *Proof.* Simple calculations show that

$$\begin{aligned}
&\mathcal{F}_\sigma(u^{k+1}, z^k) - \mathcal{F}_\sigma(u^k, z^k) \\
=&\tfrac{\sigma}{2}\left(\|\mathcal{A}u^{k+1} - z^k\|^2 - \|\mathcal{A}u^k - z^k\|^2\right) \\
=&\tfrac{\sigma}{2}\left(\|\mathcal{A}(u^{k+1} - u^k)\|^2 + 2\Re(\langle u^{k+1} - u^k, \mathcal{A}^*(\mathcal{A}u^k - z^k)\rangle)\right), \\
=&\tfrac{\sigma}{2}\left(\|u^{k+1} - u^k\|_M^2 + 2\Re(\langle u^{k+1} - u^k, MM^{-1}\mathcal{A}^*(\mathcal{A}u^k - z^k)\rangle)\right) \\
=&\tfrac{\sigma}{2}\|u^{k+1} - u^k\|_M^2 + \Re(\langle u^{k+1} - u^k, M^{-1}\nabla_u \mathcal{F}_\sigma(u^k, z^k)\rangle_M).
\end{aligned} \tag{A.3}$$

Plugging $u^{k+1} = u^k - \frac{1}{c^k} M^{-1}\nabla_u \mathcal{F}_\sigma(u^k, z^k)$, into above equation, we obtain

$$\mathcal{F}_\sigma(u^k, z^k) - \mathcal{F}_\sigma(u^{k+1}, z^k) = (c^k - \tfrac{\sigma}{2})\|u^{k+1} - u^k\|_M^2. \tag{A.4}$$

It follows from [6, Lemma 2] that one readily derives

$$\mathcal{F}_\sigma(u^{k+1}, z^k) - \mathcal{F}_\sigma(u^{k+1}, z^{k+1}) \ge \tfrac{1}{2}(d^k - \sigma)\|z^{k+1} - z^k\|^2. \tag{A.5}$$

Combining (A.4) and (A.5) yields the inequality (4.1). ∎

A.4. Proof of Corollary 4.2. *Proof.* Corollary 4.1 guarantees the boundedness of the iterative sequences generated by P^{3}ALM. As a result, for any accumulated point $(\bar{u}, \bar{z})$, there exist two subsequences $\{u^{n_k}\} \subseteq \{u^k\}, \{z^{n^k}\} \subseteq \{z^k\}$ such that $(u^{n_k}, z^{n_k}) \to (\bar{u}, \bar{z}),$ as $k \to +\infty$. Again by Corollary 4.1, we have $\lim\limits_{k\to\infty} u^{k+1} - u^k = 0$, $\lim\limits_{k\to\infty} z^{k+1} - z^k = 0$, and therefore $(u^{n_k+1}, z^{n_k+1}) \to (\bar{u}, \bar{z}),$ as $k \to +\infty$. Additionally, it is straightforward that the proximal mapping for any data fitting terms defined in (2.3) is continuous, due to their closed-form solution expressed in (3.7). Therefore, by taking limit of both sides of equations in Step 2 and Step 3 of Algorithm 1 for the subsequences, one has

$$\begin{cases} \bar{u} - M^{-1}\mathcal{A}^*\bar{z} = 0, \\ \bar{z} = \mathrm{Prox}_{d^\star}^{\mathcal{B}(|\cdot|^2, f)}\Big((1 - \frac{\sigma}{d^\star})\bar{z} + \frac{\sigma}{d^\star}\mathcal{A}\bar{u}; M\Big), \end{cases}$$

by letting $c^k \equiv c^\star$, and $d^k \equiv d^\star$. We conclude that $\bar{z}$ is a fixed point. Therefore, any accumulated point $(\bar{u}, \bar{z})$ of the iterative sequences is a critical point of (2.4).
∎

A.5. Proof of Lemma 4.2. *Proof.* First one readily get an equivalent form of E_u^k, E_z^k as $E_u^k = c^{k-1}M(u^{k-1} - u^k) + \partial_u \mathcal{H}(u^k, z^k) - \partial_u \mathcal{H}(u^{k-1}, z^{k-1})$, and $E_z^k := d^{k-1}(z^{k-1} - z^k) + \partial_z \mathcal{H}(u^k, z^k) - \partial_z \mathcal{H}(u^k, z^{k-1})$. By the first order optimality condition for the first updating scheme in (3.8), one has $\nabla_u \mathcal{H}(u^{k-1}, z^{k-1}) + c^{k-1}M(u^k - u^{k-1}) = 0$. The definition of $\mathcal{F}_\sigma$ yields $\partial_u \mathcal{F}_\sigma(u^k, z^k) = \nabla_u H(u^k, z^k)$. Combining the above two equations one obtains $E_u^k \in \partial_u \mathcal{F}_\sigma(u^k, z^k)$. In a similar manner, one can obtain $E_z^k \in \partial_z \mathcal{F}_\sigma(u^k, z^k)$, and therefore (4.5) is proved.

We then show how to estimate the upper bound. Based on (4.4), one readily obtains

$$\begin{aligned} \|E_u^k\| &\le (c^{k-1}+\sigma)\|M(u^k-u^{k-1})\| + \sigma\|\mathcal{A}^*(z^k-z^{k-1})\| \\ &\le (c^{k-1}+\sigma)\|M^{\frac{1}{2}}\|\|u^k-u^{k-1}\|_M + \sigma\|\mathcal{A}^*\|\|z^k-z^{k-1}\|. \end{aligned}$$

Similarly, one obtains $\|E_z^k\| \le (d^{k-1}+\sigma)\|z^k-z^{k-1}\|$. Summing up the two estimates concludes to this lemma. □

Appendix B. Closed-form Solution for Proximal Mapping for IGM. By the definition of proximal mapping, we have

$$\varpi_\beta(|z|) = \arg\min_{\varsigma\in\mathbb{R}_+^m} \mathcal{B}(|\varsigma|^2, f) + \tfrac{\beta}{2}\left\|\varsigma - |z|\right\|^2. \tag{B.1}$$

As derived in [15], the minimizer to (B.1) is

$$\varpi_\beta(|z|)(t) = \begin{cases} \sqrt[3]{\frac{\beta|z(t)|}{4} + \sqrt{D(t)}} + \sqrt[3]{\frac{\beta|z(t)|}{4} - \sqrt{D(t)}}, & \text{if } D(t)\ge 0; \\ 2\sqrt{\frac{f(t)-\frac{\beta}{2}}{3}}\cos\left(\arccos\frac{\theta(t)}{3}\right), & \text{otherwise}, \end{cases} \tag{B.2}$$

for $0 \le t \le m-1$, with $D(t) = \frac{(\frac{\beta}{2}-f(t))^3}{27} + \dfrac{\beta^2|z(t)|^2}{16}$, and $\theta(t) = \dfrac{\beta|z(t)|}{4\sqrt{\frac{(f(t)-\frac{\beta}{2})^3}{27}}}$.

Appendix C. ADMM for (3.17) . Given the augmented Lagrange $\mathcal{L}_r(u, \boldsymbol{p}; \Lambda)$ in (3.18), the standard ADMM algorithm goes as follows,

$$\begin{cases} u_{j+1} = \arg\min\limits_u \mathcal{L}_r(u, \boldsymbol{p}_j; \Lambda_j), \\ \boldsymbol{p}_{j+1} = \arg\min\limits_{\boldsymbol{p}} \mathcal{L}_r(u_{j+1}, \boldsymbol{p}; \Lambda_j), \\ \Lambda_{j+1} = \Lambda_j + r(\boldsymbol{p}_{j+1} - \nabla u_{j+1}). \end{cases} \tag{C.1}$$

For the u-subproblem, we need to compute the first order optimal condition to

$$\min_u \tfrac{c^k}{2}\|u-\hat{u}^k\|_M^2 + \tfrac{r}{2}\|\boldsymbol{p}_j + \frac{\Lambda_j}{r} - \nabla u\|^2.$$

Taking the gradient of the objective functional w.r.t. u [12] gives

$$(c^k M - r\Delta)u = c^k M\hat{u}^k - r\nabla\cdot(\boldsymbol{p}_j + \frac{\Lambda_j}{r}),$$

with $\Delta u = \nabla\cdot(\nabla u)$ and divergence operator $\nabla\cdot$ under some boundary conditions.

The $\boldsymbol{p}$-subproblem is expressed as

$$\boldsymbol{p}_{j+1} = \arg\min_{\boldsymbol{p}} \lambda\|\boldsymbol{p}\|_1 + \tfrac{1}{2}\|\boldsymbol{p} - (\nabla u - \Lambda_j/r)\|.$$

It has closed-form solutions via the soft thresholding *i.e.*

$$\boldsymbol{p}_{j+1} = \mathrm{Thresh}_{soft}((\nabla u - \Lambda_j/r); \lambda),$$

where Thresh_{soft} is defined in (3.16).