Lawrence Berkeley National Laboratory

Title
Advanced denoising for X-ray ptychography.

Authors
Chang, Huibin
Enfedaque, Pablo
Zhang, Jie
et al.

Advanced denoising for X-ray ptychography

HUIBIN CHANG,[1,2,10] PABLO ENFEDAQUE,[2] JIE ZHANG, [1] JULIANE REINHARDT, [3,4] BJOERN ENDERS, [5, 6] YOUNG-SANG YU, [5] DAVID SHAPIRO, [5] CHRISTIAN G. SCHROER, [7,8] TIEYONG ZENG [9], STEFANO MARCHESINI [2,11]

[1] School of Mathematical Sciences, Tianjin Normal University, Tianjin, China
[2] Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[3] ARC Centre of Excellence for Advanced Molecular Imaging, Department of Chemistry and Physics, La Trobe Institute for Molecular Sciences, La Trobe University, Bundoora, Australia
[4] The Australian Synchrotron, Clayton, Australia
[5] Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[6] Physics Department, University of California at Berkeley, Berkeley, CA, USA
[7] Deutsches Elektronen-Synchrotron DESY, D-22607 Hamburg, Germany
[8] Department Physik, Universität Hamburg, Luruper Chaussee 149, 22761 Hamburg, Germany
[9] Department of Mathematics, The Chinese University of Hong Kong, Hong Kong
[10] *changhuibin@gmail.com
[11] *smarchesini@lbl.gov

Abstract: The success of ptychographic imaging experiments strongly depends on achieving high signal-to-noise ratio. This is particularly important in nanoscale imaging experiments when diffraction signals are very weak and the experiments are accompanied by significant parasitic scattering (background), outliers or correlated noise sources. It is also critical when rare events such as cosmic rays, or bad frames caused by electronic glitches or shutter timing malfunction take place. In this paper, we propose a novel iterative algorithm with rigorous analysis that exploits the direct forward model for parasitic noise and sample smoothness to achieve a thorough characterization and removal of structured and random noise. We present a formal description of the proposed algorithm and prove its convergence under mild conditions. Numerical experiments from simulations and real data (both soft and hard X-ray beamlines) demonstrate that the proposed algorithms produce better results when compared to state-of-the-art methods.

1. Introduction

Ptychography [1–4] has become an increasingly popular imaging technique, and it is used nowadays in scientific fields as diverse as condensed matter physics [5], cell biology [6], materials science [7, 8], and electronics [4, 9], among other areas. Compared to standard lens-based microscopy, the resolution in ptychography is not limited by the size of the illumination probe, but by the wavelength and number of photons used in an experiment. Unfortunately, multiple experimental challenges have to be tackled to achieve a high-quality reconstruction from a ptychographic experiment in practice. Complex experimental systems require nanometer stability while collecting large amounts of diffraction data, which can require several hours depending on the experimental setup. Data can be then contaminated by structured parasitic scattering [10–12], detector read-out noise, Poisson noise derived from the photon counting, and different types of outliers. Additionally, the characterization of the illumination source is also commonly considered as a joint problem to the object reconstruction, known as Blind Ptychography (BP) [3, 13–15].

During the last decades, researchers have developed several schemes to solve the BP problem. Arguably, the most popular ones are extended Ptychographic Iterative Engine (ePIE) [3], Difference Map [13, 16], Maximum Likelihood (ML) method [17], Proximal Splitting algorithm [14],

Relaxed Averaged Alternating Reflections (RAAR [18]) based algorithms [19], and generalized Alternating Direction Method of Multiplies (ADMM) [20–22]) based BP [15]. Although some implementations of those methods present *ad hoc* solutions for structural noise removal, there is no formal analysis or in-depth characterization of such experimental problems, even though they can be critical to achieve robust high-resolution images, specially from weakly scattering or low contrast specimens.

Experimental noise overview   Direct reconstructions based on raw experimental data often contain visible artifacts. They commonly derive from outliers and structured and randomly distributed uncorrelated noise sources. Below there is a characterization of the main causes of such anomalies.

Structured noise

(1) Parasitic scattering: Derives from the scattering from any element along the beam path other than the sample and the optical elements desired harmonic order e.g.: slits, pinholes, lens imperfections, harmonic contamination, air, gas, etc. We refer to this as the *background*.

(2) Saturation: Occurs when the flux exceeds the detector capacity, the range of the Analog to Digital Converter (ADC) or the maximum photon counting rate.

(3) Dark noise: Produced by the detector dark current, light from position encoders and interferometers, or light from the environment. It differs from the parasitic scattering as it occurs also when the x-ray beam is off and can be measured in advance by turning off the beam and (commonly referred to as *dark frame*). The constant component can be subtracted from the measured signal before applying a reconstruction solver or alternatively it can be incorporated in the generalized *background*.

Random noise

(4) Photon-counting noise: Caused by the quantization of the wavefront at the detector.

(5) Read-out and detector noise: Derives from electronic interference when the ADC converts the charge distribution to a digital signal and the random component of the dark current. The interference and parallel ADC read-out electronics can produce ringing and correlation in the digital signal.

Outliers

(6) Bad frames: Glitches on the functioning of mechanical components inside an instrument such as shutter timing can cause some measured frames to be corrupted.

(7) Cosmic rays: High energy cosmic particles that penetrate the atmosphere, the building and the instrument enclosure and hit the detector. Although they are rare, those types of signals normally present very high charge.

(8) Bad pixels: Result from fabrication errors may cause dead pixels on a detector, these can be typically measured in advanced and masked out during the reconstruction process.

In this paper we focus on a solution to address *(1)*, *(6)* and *(7)*, regarding structured noise and outliers, and also propose a technique to deal with a combination of different random noise distributions coming from *(4)* and *(5)*. The noise from *(3)* and *(8)* can be measured in advance, and it is commonly subtracted. Saturation noise *(2)* can be equally identified and addressed by masking out the saturated signal areas and it is outside the focus of this paper. Random noise sources *(4)* and *(5)* have been considered in the literature [15, 17, 23, 24], but there is no solution that can address the combination of two or more random noise distributions. In this paper, we

propose a mixed-noise term for ptychography to effectively deal with any combination of Poisson and Gaussian noise sources.

Some of the previous pathologies can be identified inspecting the measured data. An illustrative example of parasitic noise *(1)* is reported in Fig. 1. A single measured diffraction pattern (Fig. 1(a)) presents a distinct reverse "L-Shape" artifact, which consistently appears on all other diffraction measurements from the same dataset. If parasitic scattering is not considered, a standard ptychography reconstruction produces very noisy and low contrast images (Figs. 1(d) and 1(e)). The phase part is lost in some areas, concealed by the biased background of the recovered phase, and the resulted intensities (Fig. 1(b)) contain a slight alteration of the parasitic noise from the measured data. (A more detailed comparison between different methods, with and without structured noise correction, is provided in section 3)

(a) Diffraction pattern   (b) RAAR   (c) Proposed alg.

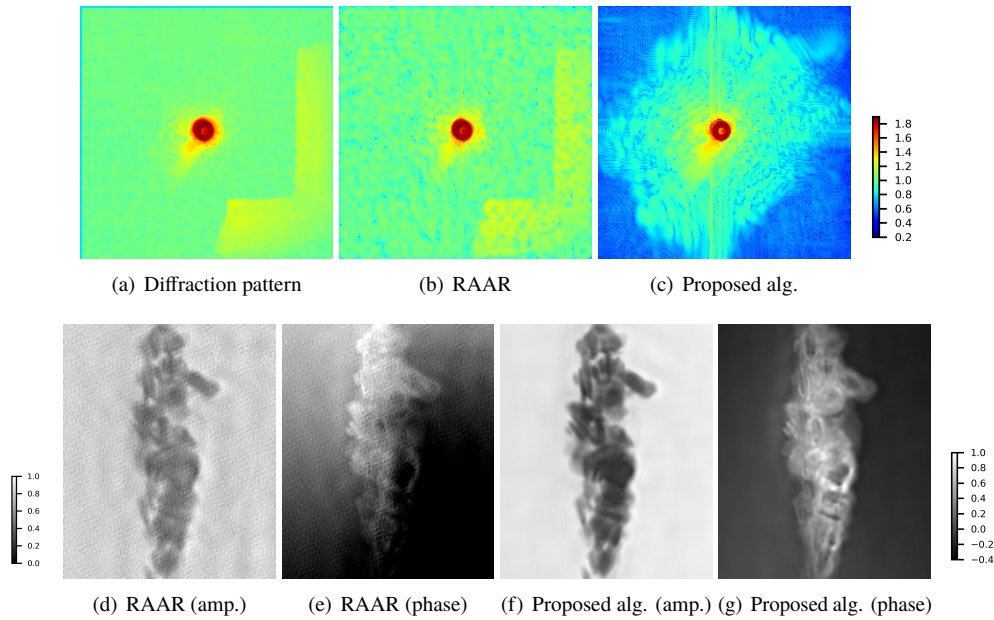(d) RAAR (amp.)   (e) RAAR (phase)   (f) Proposed alg. (amp.) (g) Proposed alg. (phase)

Fig. 1. Reconstruction results using RAAR (implemented in SHARP [19]) and the proposed algorithm. The dataset corresponds to one of the 3D ptychography experiments from the results presented in [8]. (a) Measured intensities from a single diffraction pattern. Recovered intensities (b), object amplitude (amp.) (d) and phase (e), respectively, using RAAR without structured noise correction. Recovered intensities (c), object amplitude (f) and phase (g), respectively, using the proposed algorithm (alg.). The figures in the first row are shown in scale $(\cdot)^{0.05}$ following the colorbar shown on the top right corner; the figures in the second row are shown in linear scale following the colorbars on the left corner (amplitude) and right corner (phase).

Related work   A beam stop scheme was employed in [11] to reduce the parasitic noise using intensities from two raster scans (i.e. scanning on a Cartesian 2D grid) with and/or without beamstop. However, the artifacts caused by the structured noise are still noticeable in Fig. 2 of [11], and some fine features still appear undefined. A multiple-mode approach in [25] was proposed to reduce the background. In [12], the data was pre-processed by subtracting an adjustable scalar value multiplied by the *dark frame*. A method to remove the parasitic background automatically employed a gradient descent method with adaptive stepsizes was

proposed in [26] and implemented using a distributed multi-GPU implementation [19] in a real-time streaming production code at a user facility [27].

Several researchers have also studied techniques to remove the random noise from photon-counting or read-out noises by employing a sparse regularization technique to further improve the image quality for conventional ptychography, .e.g. Tikhonov regularization with nonlinear conjugate gradient method [17], total variation regularization with ADMM [28] and dictionary learning method with proximal algorithm [29]. Specially, for the more practical noises, like a mix of Poisson and Gaussian noises, several variational methods have been proposed for linear inverse problems, e.g. generalized Anscombe transform [30], total variation based Shift-Poisson method [31] and weighted least squares method [32]. However, such methods have not been applied to the conventional ptychography yet, which is essentially a nonlinear inverse problem. Beyond algorithmic methods, there are also several contributions from beamline scientists attenuating the background problem via hardware solutions [10, 33].

Motivations and contributions   In a ptychography experimental setup it is impossible to isolate parasitic noise from the measured data. Furthermore, the data is often contaminated by outliers and an additional mix of random noise types. In this context, an automatic denoising algorithm with convergence guarantee to jointly remove both structured and random noise can be very valuable. In this paper, we propose ADMM Denoising for Phase retrieval (ADP), a new algorithm that addresses all the main sources of experimental noise in a blind X-ray ptychography experimental setting. The proposed algorithm achieves similar convergence speed as state-of-the-art blind ptychography methods and it is constructed on the framework presented in [15]. Even if additional hardware or pre-processing techniques are in place, ADP can be used to further enhance the reconstruction quality of the retrieved object. The main contributions of this work are listed below:

    i. A new algorithm for automatic structured and random mixed noise removal with outliers correction. The algorithm is based on the forward physical model of the parasitic scattering noise (additive frame-invariant background) and on the maximum a posteriori (MAP) estimation combined with the shift-Poisson method [31] for mixed types of noise. By assuming the piecewise smoothness of sample patches, We propose an additional framewise regularization term to further enhance the denoising process and improve the image quality, instead of using a simple sparsity term of the sample [28]. To the best of our knowledge, it is the first time an iterative algorithm with rigorous analysis for background removal in ptychographic imaging is proposed.

    ii. By formulating the parasitic noise variable with positive constraint as a quadratic term, fast ADMM algorithms are designed for the proposed model, with and without regularization. The convergence guarantee of the method is proved under very mild conditions. Computationally, each subproblem can be solved very efficiently, using pointwise operations or simple linear algebra solvers, requiring few arithmetic operations per iteration and exposing high parallelism.

    iii. Multiple numerical experiments on both simulated and experimental data from different light sources demonstrate the enhanced quality of the proposed algorithm. Reconstruction results using the proposed algorithm can be found in Fig. 1 and in the experimental results section.

2.   Proposed iterative algorithm

Mathematical formula for standard ptychography   In a standard ptychography experiment, a localized coherent X-ray probe (or illumination) ω scans through an image (or sample) u, while

the detector collects a sequence of phaseless intensities in the far field. Throughout this paper, we consider the following discrete setting:

The variable $u \in \mathbb{C}^n$ ($\mathbb{C}^n$ is the complex Euclidean space) corresponds to a 2D sample with $\sqrt{n} \times \sqrt{n}$ pixels, and $\omega \in \mathbb{C}^{\bar{m}}$ is a localized 2D probe with $\sqrt{\bar{m}} \times \sqrt{\bar{m}}$ pixels, both u and ω written as a vector by a lexicographical order. A stack of phaseless measurements $\tilde{I}_j \in \mathbb{R}_+^{\bar{m}}$ $\forall 0 \le j \le J-1$ ($\mathbb{R}_+^{\bar{m}}$ is the Euclidean space with non-negative elements) is collected with $\tilde{I}_j = |\mathcal{F}(\omega \circ \mathcal{S}_j u)|^2$, where $|\cdot|$ represents the element-wise absolute value of a vector, $\circ$ denotes the element-wise multiplication, and $\mathcal{F}$ denotes the normalized discrete Fourier transformation. $\mathcal{S}_j \in \mathbb{R}^{\bar{m} \times n}$ is a binary matrix that specifies a small window with the index j and size $\bar{m}$ over the entire sample u. The BP problem can then be expressed as follows:

$$\text{To find } \omega \in \mathbb{C}^{\bar{m}} \text{ and } u \in \mathbb{C}^n, \text{ such that } |\mathcal{A}(\omega, u)|^2 = \tilde{I}, \tag{1}$$

where bilinear operators $\mathcal{A} : \mathbb{C}^{\bar{m}} \times \mathbb{C}^n \to \mathbb{C}^m$ and $\mathcal{A}_j : \mathbb{C}^{\bar{m}} \times \mathbb{C}^n \to \mathbb{C}^{\bar{m}}$ $\forall 0 \le j \le J-1$, are denoted as: $\mathcal{A}(\omega, u) := (\mathcal{A}_0^T(\omega, u), \mathcal{A}_1^T(\omega, u), \cdots, \mathcal{A}_{J-1}^T(\omega, u))^T$, $\mathcal{A}_j(\omega, u) := \mathcal{F}(\omega \circ \mathcal{S}_j u)$, and $\tilde{I} := (\tilde{I}_0^T, \tilde{I}_1^T, \cdots, \tilde{I}_{J-1}^T)^T \in \mathbb{R}_+^m$ ($m = J \times \bar{m}$).

2.1. Proposed model

Experimentally, not all of the parasitic scattering (background) is the same on each frame (here, "frame" means one image from the CCD detector). However, the frame-invariant component is one of the key sources of the artifacts in standard reconstruction results due to the redundancy of ptychography [11, 12]. Essentially, it is an additive components of signals, apart from those originating from the sample and the beam hitting it. Hence, we start by assuming that the background of each frame is approximately unchanged [26], i.e., it is frame-invariant. It is important to note that parasitic noise can also be interpreted as an independent structured noise source per frame. The frame-invariant approximation is critical for proper modeling and the problem is actually not well defined if this approximation is not considered. This is due to the fact that the independent noise per frame alternative has a trivial solution where the background is equal to the residual between the measured and reconstructed intensities from the iterative solution of the sample and probe. The frame-invariant approximation is, besides simple, very efficient at removing structural noise to produce high-contrast recovery images.

Formally, the true intensity data $\tilde{I}$ is contaminated by non-negative parasitic structured noise $\hat{\phi} \in \mathbb{R}_+^{\bar{m}}$, such that the measured data $\hat{I} = (\hat{I}_0^T, \hat{I}_1^T, \cdots, \hat{I}_{J-1}^T)^T \in \mathbb{R}_+^m$ with $\hat{I}_j \in \mathbb{R}_+^{\bar{m}}, \forall 0 \le j \le J-1$ is recorded for each frame as follows:

$$0 \le \hat{I}_j = \tilde{I}_j + \hat{\phi} \ \forall 0 \le j \le J-1. \tag{2}$$

In practice, the data is contaminated by noise as

$$\hat{I}_j = \text{Noi}(\tilde{I}_j + \hat{\phi}) \ \forall 0 \le j \le J-1, \tag{3}$$

where Noi denotes the degrading operator caused by different noise sources. Without loss of generality, we consider the Poisson and Gaussian mixed noise model with parasitic noise as:

$$\hat{I}_j(t) \overset{\text{i.i.d}}{\sim} \text{Poisson}(|\mathcal{A}_j(\omega, u)(t)|^2 + \hat{\phi}(t)) + \text{n}_j(t), \tag{4}$$

where i.i.d. is short for "Independent and Identically Distributed", n_j denotes the white Gaussian noise with zero mean and variance $\sigma > 0$. It is not difficult to verify that $\text{var}(\hat{I}_j(t) + \sigma^2) = \text{var}(\hat{I}_j(t)) = |\mathcal{A}_j(\omega, u)(t)|^2 + \hat{\phi} + \sigma^2$, and $\text{mean}(\hat{I}_j(t) + \sigma^2) = |\mathcal{A}_j(\omega, u)(t)|^2 + \hat{\phi} + \sigma^2$. Hence, $I_j(t) := \hat{I}_j(t) + \sigma^2$ has the same variance and mean. By further estimating I_j by a Poisson distribution with the shift-Poisson technique [31] and using KL divergence derived by maximum

likelihood estimation of Poisson noise, the following optimization problem is derived:

$$\min_{\omega,u,\phi} \frac{1}{2} \sum_j \left\langle \mathbf{1}_{\bar{m}}, |\mathcal{A}_j(\omega,u)|^2 + \phi - I_j \circ \log(|\mathcal{A}_j(\omega,u)|^2 + \phi) \right\rangle + \mathbb{I}_{\mathscr{E}}(\phi), \tag{5}$$

where $\mathbf{1}_{\bar{m}}$ is a vector of all ones, $\phi := \hat{\phi} + \sigma^2$, $\langle \cdot, \cdot \rangle$ denotes the inner product as $\langle v_1, v_2 \rangle := \sum_{j=0}^{\bar{m}} v_1(t)v_2(t) \; \forall v_1, v_2 \in \mathbb{R}^{\bar{m}}$, and positivity constraint set is defined as $\mathscr{E} := \left\{ \phi \in \mathbb{R}_+^{\bar{m}} : \phi \geq 0 \right\}$. Here, the indicator function $\mathbb{I}_{\mathscr{E}}$ is defined as: $\mathbb{I}_{\mathscr{E}}(\phi) = 0$, if $\phi \in \mathscr{E}$; $\mathbb{I}_{\mathscr{E}}(\phi) = +\infty$, otherwise.

Remark 2.1. *In order to jointly estimate the background and illumination (double-blind), based on Eq.* (2)*, a variant optimization model with least squares fitting can be established regardless of the specific noise mechanism as*

$$\min_{\omega,u,\phi} \frac{1}{2} \sum_{j=0}^{J-1} \left\| |\mathcal{A}_j(\omega,u)|^2 + \phi - I_j \right\|^2 + \mathbb{I}_{\mathscr{E}}(\phi). \tag{6}$$

However, when this metric is used to measure the distance between the collected and recovered intensities, the corresponding first-order iterative algorithm is significantly slow [15]. This is because the residual slowly decreases as the iterations go, and the subproblem with respect to the auxiliary variable is related with quartic equations, which are solved with high computational cost [34].

In order to deal with the misfit caused by experimental outliers, a weight vector $C := (C_0^T, C_1^T, \cdots, C_{J-1}^T)^T \in \mathbb{R}^m$ is introduced to improve the fitting residual. To further handle the noise or other artifacts, we consider the total variation regularization on the split patches of the sample, instead of the regularization of the entire sample in [28]. As a result, we propose the following Kullback-Leibler (KL) divergence-based problem with framewise sparsity and parasitic noise retrieval:

$$\min_{\omega,u,\phi} \frac{1}{2} \sum_j \left\langle C_j, |\mathcal{A}_j(\omega,u)|^2 + \phi + \varepsilon^2 \mathbf{1}_{\bar{m}} - (I_j + \varepsilon^2 \mathbf{1}_{\bar{m}}) \circ \log(|\mathcal{A}_j(\omega,u)|^2 + \phi + \varepsilon^2 \mathbf{1}_{\bar{m}}) \right\rangle$$
$$+ \lambda \sum_j \mathrm{TV}(\mathcal{S}_j u) + \mathbb{I}_{\mathscr{E}}(\phi), \tag{7}$$

with a penalization parameter $\varepsilon > 0$, where TV denotes the total variation and λ is a positive parameter to balance the regularization and data fitting terms.

2.2. ADMM denoising for phase retrieval (ADP)

Generally speaking, the ADMM is a variant of the augmented Lagrangian method [20, 21] that uses partial updates for the dual variable, and each subproblem can be easily solved compared with the original optimization problem. Due to its scalability and flexibility, it has been a popular algorithm for large-scale optimization problems arising in computer vision, statistics, machine learning, and other related areas, and see more details in the review paper [22] and references therein. Since the proposed model is non-convex and non-smooth, ADMM will be adopted to solve it, which allows for bigger stepsizes by avoiding directly calculating the gradient of the objective functional and therefore has fast convergence with low computation cost per iteration.

We formulate below the proposed ADP algorithm to solve Eq. (7). First, we consider the problem without regularization by setting $\lambda = 0$. If directly following [35, 36], the subproblem with respect to the variable ϕ does not have a closed form solution:

$$\min_{\omega,u,\phi} \frac{1}{2} \sum_j \left\langle C_j, |\mathcal{A}_j(\omega,u)|^2 + \phi + \varepsilon^2 \mathbf{1}_{\bar{m}} - (I_j + \varepsilon^2 \mathbf{1}_{\bar{m}}) \circ \log(|\mathcal{A}_j(\omega,u)|^2 + \phi + \varepsilon^2 \mathbf{1}_{\bar{m}}) \right\rangle + \mathbb{I}_{\mathscr{E}}(\phi).$$
$$\tag{8}$$

A gradient descent or proximal type algorithm reported in [15] to directly solve this problem presents very slow convergence. In order to develop a more efficient algorithm for each subproblem, we first rewrite Eq. (7) with $\lambda = 0$ as

$$\min_{\omega, u, \tilde{\phi}} \tfrac{1}{2} \sum_j \left\langle C_j, |\mathcal{A}_j(\omega, u)|^2 + \tilde{\phi}^2 + \varepsilon^2 \mathbf{1}_{\bar{m}} - (I_j + \varepsilon^2 \mathbf{1}_{\bar{m}}) \circ \log(|\mathcal{A}_j(\omega, u)|^2 + \tilde{\phi}^2 + \varepsilon^2 \mathbf{1}_{\bar{m}}) \right\rangle, \quad (9)$$

where $\tilde{\phi} := \sqrt{\phi}$. Note that the constraint for ϕ is automatically removed in Eq. (9). Also, the variables $\mathcal{A}_j(\omega, u)$ and $\tilde{\phi}$ play now the same role in the objective function and adding auxiliary variables can derive a more efficient algorithm than directly solving the previous problem.

Introducing the auxiliary variables $z_j = \mathcal{A}_j(\omega, u)$ and $\mu_j = \tilde{\phi} \;\; \forall 0 \le j \le J - 1$, produces the following equivalent constraint optimization problem:

$$\min_{\omega, u, z, \mu, \tilde{\phi}} \mathcal{G}_\varepsilon(z, \mu), \quad \text{such that } z_j - \mathcal{A}_j(\omega, u) = 0, \;\; \mu_j - \tilde{\phi} = 0 \;\; \forall 0 \le j \le J - 1, \quad (10)$$

with

$$\mathcal{G}_\varepsilon(z, \mu) := \tfrac{1}{2} \sum_j \left\langle C_j, |z_j|^2 + \mu_j^2 + \varepsilon^2 \mathbf{1}_{\bar{m}} - (I_j + \varepsilon^2 \mathbf{1}_{\bar{m}}) \circ \log(|z_j|^2 + \mu_j^2 + \varepsilon^2 \mathbf{1}_{\bar{m}}) \right\rangle. \quad (11)$$

The corresponding augmented Lagrangian for Eq. (10) reads:

$$
\begin{aligned}
\mathcal{L}(\omega, u, z, \mu, \tilde{\phi}, \Lambda_1, \Lambda_2) &:= \mathcal{G}_\varepsilon(z, \mu) + r \sum_j \mathfrak{R}\langle \Lambda_{1,j}, z_j - \mathcal{A}_j(\omega, u) \rangle \\
&+ \tfrac{r}{2} \sum_j \|z_j - \mathcal{A}_j(\omega, u)\|^2 + r \sum_j \langle \Lambda_{2,j}, \mu_j - \tilde{\phi} \rangle + \tfrac{r}{2} \sum_j \|\mu_j - \tilde{\phi}\|^2,
\end{aligned}
\quad (12)
$$

where $\mathfrak{R}$ denotes the real part of the complex-valued number. The proposed ADP algorithm is formulated as follows:

$$
\left\{
\begin{aligned}
\omega^{k+1} &= \arg\min_{\omega} \mathcal{L}(\omega, u^k, z^k, \mu^k, \tilde{\phi}^k, \Lambda_1^k, \Lambda_2^k) + \tfrac{\alpha_1}{2} \|\omega - \omega^k\|^2; \\
u^{k+1} &= \arg\min_{u} \mathcal{L}(\omega^{k+1}, u, z^k, \mu^k, \tilde{\phi}^k, \Lambda_1^k, \Lambda_2^k) + \tfrac{\alpha_2}{2} \|u - u^k\|^2; \\
(z^{k+1}, \mu^{k+1}) &= \arg\min_{z,\mu} \mathcal{L}(\omega^{k+1}, u^{k+1}, z, \mu, \tilde{\phi}^k, \Lambda_1^k, \Lambda_2^k); \\
\tilde{\phi}^{k+1} &= \arg\min_{\tilde{\phi}} \mathcal{L}(\omega^{k+1}, u^{k+1}, z^{k+1}, \mu^{k+1}, \tilde{\phi}, \Lambda_1^k, \Lambda_2^k); \\
\Lambda_1^{k+1} &= \Lambda_1^k + z^{k+1} - \mathcal{A}(\omega^{k+1}, u^{k+1}); \\
\Lambda_{2,j}^{k+1} &= \Lambda_{2,j}^k + \mu_j^{k+1} - \tilde{\phi}^{k+1} \;\; \forall 0 \le j \le J - 1,
\end{aligned}
\right.
\quad (13)
$$

with $\Lambda_1 := (\Lambda_{1,0}^T, \Lambda_{1,1}^T, \cdots, \Lambda_{1,J-1}^T)^T \in \mathbb{C}^m$, and $\Lambda_2 := (\Lambda_{2,0}^T, \Lambda_{2,1}^T, \cdots, \Lambda_{2,J-1}^T)^T \in \mathbb{R}^m$.

The solution to each subproblems above are reported in Algorithm 1, and further details can be found in the Appendix B. When setting $\mu_j^0 = \tilde{\phi}^0 = 0$, and $\Lambda_2^0 = 0$, following the above algorithm we have that $\mu_j^k = 0 \;\; \forall k = 1, 2, \cdots$, since $\mu^\star = 0$ is a stationary point of the proposed model Eq. (7). Therefore, we use the following *warm-start* scheme to initialize μ: first solve a problem without structured noise correction by setting $\mu_j^0 = 0$, and after J_0 iterations reset μ_j^k using $\mu_j^{k+1} := \sqrt{\max\{0, \tfrac{1}{J} \sum_j (I_j - |\mathcal{A}(\omega^{k+1}, u^{k+1})|^2)\}}$. The weight function $\{C_j\}$ can be simply fixed to be one. It can also be dynamically changed if the measured data contains outliers, such that the ADP algorithm is slightly modified with additional update of the weight function between Step 3 and Step 4, and please see more details for update scheme of this function in Remark 2.2.

Algorithm 1: ADP

Step 0. Initialization: $u^0 := \mathbf{1}_n$, $\omega^0 := \frac{1}{J} \sum_j \mathcal{F}^* \sqrt{I_j}$ ($\mathcal{F}^*$ denotes the normalized inverse discrete Fourier Transform), $z_j^0 := \mathcal{A}_j(\omega^0, u^0)$, $\mu^0 = 0$, $\Lambda_1 = 0$, $\Lambda_2 = 0$.

Step 1. Refine the illumination by

$$\omega^{k+1} = \frac{\sum_j \mathcal{S}_j(u^k)^* \circ \mathcal{F}^*(z_j^k + \Lambda_{1,j}^k) + \alpha_1 \omega^k}{\sum_j |\mathcal{S}_j u^k|^2 + \alpha_1 \mathbf{1}_{\bar{m}}}. \tag{14}$$

Step 2. Refine the object by

$$u^{k+1} = \frac{\sum_j \mathcal{S}_j^T ((\omega^{k+1})^* \circ \mathcal{F}^*(z_j^k + \Lambda_{1,j}^k)) + \alpha_2 u^k}{\sum_j \mathcal{S}_j^T |\omega^{k+1}|^2 + \alpha_2 \mathbf{1}_n}. \tag{15}$$

Step 3. Update z_j^{k+1}, μ_j^{k+1} by

$$((z_j^{k+1})^T, (\mu_j^{k+1})^T)^T = ((\rho_j^k)^T, (\rho_j^k)^T)^T \circ \mathrm{sign}(\bar{X}_j^k), \tag{16}$$

with

$$(\bar{X}_j^k)^T := (\mathcal{A}_j^T(\omega^{k+1}, u^{k+1}) - (\Lambda_{1,j}^k)^T, (\tfrac{1}{J} \sum_j \mu_j^k)^T - (\Lambda_{2,j}^k)^T) \tag{17}$$

where ρ_j^k is computed as:

If $\varepsilon = 0$:

$$\rho_j^k := \frac{r|\bar{X}_j^k|_* + \sqrt{r^2|\bar{X}_j^k|_*^2 + 4(C_j + r\mathbf{1}_{\bar{m}}) \circ C_j \circ I_j}}{2(C_j + r\mathbf{1}_{\bar{m}})}; \tag{18}$$

Else if $\varepsilon > 0$: it is iteratively determined by

$$\rho_{j,l+1} = \max\left\{0, \left((1 - \tau r)\mathbf{1}_{\bar{m}} - \tau C_j + \tau C_j \circ \frac{I_j + \varepsilon^2 \mathbf{1}_{\bar{m}}}{\rho_{j,l}^2 + \varepsilon^2 \mathbf{1}_{\bar{m}}}\right) \circ \rho_{j,l} + \tau r|\bar{X}_j^k|_*\right\} \tag{19}$$

$l = 0, 1, \cdots$, with $\rho_{j,0} := \rho_j^{k-1}$

Step 4.

$$\begin{aligned}
\Lambda_{1,j}^{k+1} &= \Lambda_{1,j}^k + z_j^{k+1} - \mathcal{A}_j(\omega^{k+1}, u^{k+1}); \\
\Lambda_{2,j}^{k+1} &= \Lambda_{2,j}^k + \mu_j^{k+1} - \tfrac{1}{J} \sum_j \mu_j^{k+1}.
\end{aligned} \tag{20}$$

Step 5. If $k = J_0$: reinitialize the average of μ^{k+1} by

$$\tfrac{1}{J} \sum \mu_j^{k+1} := \sqrt{\max\{0, \tfrac{1}{J} \sum_j (I_j - |\mathcal{A}(\omega^{k+1}, u^{k+1})|^2)\}}; \tag{21}$$

Else if (*satisfying some stopping condition*): output u^{k+1} as the final reconstructed result
Else: set $k := k + 1$, and goto Step 1

Convergence analysis We assess the convergence of ADP for the proposed model with $\lambda = 0$ in the following theorem:

Theorem 1. *By denoting $Y^k := (\omega^k, u^k, z^k, \mu^k, \tilde{\phi}^k, \Lambda_1^k, \Lambda_2^k)$, any limit point of $\{Y^k\}$ produced by ADP is the stationary point of Eq. (7) with $\lambda = 0$ if the stepsize r is sufficiently large and $\varepsilon > 0$.*

The proof of the above theorem can be found in Appendix C. The referred proof demonstrates that any limit point of the iterative sequence is the stationary point of the proposed model. In order to prove the convergence of the whole sequence we need more constraints for the object and illumination, and also a careful selection scheme of the proximal parameters α_1, α_2 [15]. The introduction of ε is also needed for the convergence analysis. Numerically, such penalization does not produce any obvious improvement on either convergence speed or reconstruction quality. Hence, for simplicity, we only show the performance of proposed algorithm with $\varepsilon = 0$.

2.3. ADP with regularization (ADPr)

In this subsection we consider the proposed model Eq. (7) with $\lambda > 0$. Introducing the auxiliary variables $p_j = \nabla \mathcal{S}_j u$, $z_j = \mathcal{A}_j(\omega, u)$ and $\mu_j = \tilde{\phi}$ $\forall 0 \leq j \leq J-1$ produces the following equivalent constraint optimization problem:

$$\min_{\omega, u, z, \mu, \tilde{\phi}} \lambda \sum_j \left\| |p_j| \right\| + \mathcal{G}_\varepsilon(z, \mu), \quad \text{such that } z_j - \mathcal{A}_j(\omega, u) = 0, \ \mu_j - \tilde{\phi} = 0, \ p_j - \nabla \mathcal{S}_j u = 0.$$

$$(22)$$

The corresponding augmented Lagrangian reads

$$\mathcal{L}_{reg}(\omega, u, p, z, \mu, \tilde{\phi}, \Lambda_1, \Lambda_2, \Lambda_3)$$

$$:= \mathcal{G}_\varepsilon(z, \mu) + \sum_j \left(r\mathfrak{R}\langle \Lambda_{1,j}, z_j - \mathcal{A}_j(\omega, u)\rangle + \frac{r}{2}\|z_j - \mathcal{A}_j(\omega, u)\|^2 + r\langle \Lambda_{2,j}, \mu_j - \tilde{\phi}\rangle \right.$$

$$\left. + \frac{r}{2}\|\mu_j - \tilde{\phi}\|^2 + \lambda\left\||p_j|\right\| + \beta\mathfrak{R}\langle p_j - \nabla \mathcal{S}_j u, \Lambda_{3,j}\rangle + \frac{\beta}{2}\|p_j - \nabla \mathcal{S}_j u\|^2 \right).$$

$$(23)$$

We propose the following generalization of ADP to solve the problem above, referred to as ADP with regularization (ADPr):

$$\begin{cases} \omega^{k+1} = \arg\min_\omega \mathcal{L}_{reg}(\omega, u^k, p^k, z^k, \mu^k, \tilde{\phi}^k, \Lambda_1^k, \Lambda_2^k, \Lambda_3^k) + \frac{\alpha_1}{2}\|\omega - \omega^k\|^2; \\[4pt] u^{k+1} = \arg\min_u \mathcal{L}_{reg}(\omega^{k+1}, u, p^k, z^k, \mu^k, \tilde{\phi}^k, \Lambda_1^k, \Lambda_2^k, \Lambda_3^k) + \frac{\alpha_2}{2}\|u - u^k\|^2; \\[4pt] p^{k+1} = \arg\min_p \mathcal{L}_{reg}(\omega^{k+1}, u^{k+1}, p, z^k, \mu^k, \tilde{\phi}^k, \Lambda_1^k, \Lambda_2^k, \Lambda_3^k); \\[4pt] (z^{k+1}, \mu^{k+1}) = \arg\min_{z,\mu} \mathcal{L}_{reg}(\omega^{k+1}, u^{k+1}, p^{k+1}, z, \mu, \tilde{\phi}^k, \Lambda_1^k, \Lambda_2^k, \Lambda_3^k); \\[4pt] \tilde{\phi}^{k+1} = \arg\min_{\tilde{\phi}} \mathcal{L}_{reg}(\omega^{k+1}, u^{k+1}, p^{k+1}, z^{k+1}, \mu^{k+1}, \tilde{\phi}, \Lambda_1^k, \Lambda_2^k, \Lambda_3^k); \\[4pt] \Lambda_1^{k+1} = \Lambda_1^k + z^{k+1} - \mathcal{A}(\omega^{k+1}, u^{k+1}); \\[4pt] \Lambda_{2,j}^{k+1} = \Lambda_{2,j}^k + \mu_j^{k+1} - \tilde{\phi}^{k+1} \qquad \forall 0 \leq j \leq J-1; \\[4pt] \Lambda_{3,j}^{k+1} = \Lambda_{3,j}^{k+1} + p_j^{k+1} - \nabla \mathcal{S}_j u^{k+1} \ \forall 0 \leq j \leq J-1. \end{cases}$$

$$(24)$$

Details of solvers for these subproblems are reported in Appendix B. We only summarize the overall algorithm in Algorithm 2 below.

Algorithm 2: ADP with regularization (ADPr)

Step 0. Initialization: $u^0 := \mathbf{1}_n$, $\omega^0 := \frac{1}{J}\sum_j \mathcal{F}^* \sqrt{I_j}$, $z_j^0 := \mathcal{A}_j(\omega^0, u^0)$, $\Lambda_1 = 0, \Lambda_2 = 0, \Lambda_3 = 0, p^0 = 0$, and $\mu_j^0 := 0$

Step 1. Refine the illumination as Step 1 of Algorithm 1

Step 2. Refine the object by solving the following equation using the conjugate gradient method

$$
\begin{aligned}
&\sum_j \left(\mathrm{diag}(r\mathcal{S}_j^T |\omega^{k+1}|^2 + \alpha_2 \mathbf{1}_n) - \beta \mathcal{S}_j^T \Delta \mathcal{S}_j \right) u^{k+1} \\
&= \sum_j \left(r\mathcal{S}_j^T ((\omega^{k+1})^* \circ \mathcal{F}^*(z_j^k + \Lambda_{1,j}^k)) - \beta \mathcal{S}_j^T \mathrm{div}(p_j^k + \Lambda_{3,j}^k) \right) + \alpha_2 u^k,
\end{aligned}
\tag{25}
$$

where $\mathrm{diag}(\cdot)$ gives a square diagonal matrix with the elements of a vector on the main diagonal, and $\mathrm{div}(\cdot)$ denotes the discrete divergence operator.

Step 3. Refine the gradient of the object by

$$
p_j^{k+1} = \max\{0, |\nabla \mathcal{S}_j u^{k+1} - \Lambda_{3,j}^k| - \tfrac{\lambda}{\beta}\mathbf{1}_{\bar{m}}\} \circ \frac{\nabla \mathcal{S}_j u^{k+1} - \Lambda_{3,j}^k}{|\nabla \mathcal{S}_j u^{k+1} - \Lambda_{3,j}^k|}.
\tag{26}
$$

Step 4. Update z_j^{k+1}, μ_j^{k+1} as Step 3 of Algorithm 1

Step 5. Update $\Lambda_1^{k+1}, \Lambda_2^{k+1}$ as Step 4 Algorithm 1, and update Λ_3^k by

$$
\Lambda_{3,j}^{k+1} = \Lambda_{3,j}^k + p_j^{k+1} - \nabla \mathcal{S}_j u^{k+1}.
\tag{27}
$$

Step 6. If $k = J_0$: reinitialize the average of $\{\mu_j^{k+1}\}$ as Step 5 of Algorithm 1;

Else if (*satisfying some stopping condition*): output u^{k+1} as the final reconstructed result

Else: set $k := k + 1$, and goto Step 1.

In a similar manner as with ADP, one can derive the convergence of ADPr. We provide the following theorem omitting the proof details in this case:

Theorem 2. *Any limit point of the iterative sequence generated by ADPr is the stationary point of Eq.* (7) *if the stepsizes r, β are sufficiently large and $\varepsilon > 0$.*

Remark 2.2. *In order to handle outliers, we propose the following adaptive weights $\{C_j\}$. They are designed to produce bigger values when the residual between the measured data and recovered intensities are smaller. The weight function C_j is given based on the inverse of the residual as follows:*

$$
C_j \leftarrow \frac{2}{\gamma} \left| \frac{1}{\sqrt{|z_j^{k+1}|^2 + |\mu_j^{k+1}|^2 + \varepsilon^2 \mathbf{1}_{\bar{m}}} - \sqrt{I_j + \varepsilon^2 \mathbf{1}_{\bar{m}}}} \right|^{2-\gamma},
\tag{28}
$$

with $\gamma \le 2$. The weight function can be updated between Step 3 and Step 4 in ADP algorithm, or Step 4 and Step 5 in ADPr algorithm.

Remark 2.3. *Raster grid scanning can cause visible periodical artifacts, and therefore a constraint of the lens can be enforced as [19]. Furthermore, a detector mask can be used as well.*

3. Experimental results

The experimental results of this section are generated using both simulated and experimental ptychography data. In most experiments, the performance of the proposed algorithm is compared with that of SHARP (using RAAR) [19], a ptychography software solution that implements state-of-the-art blind ptychography algorithms and background retrieval techniques [26] (further details can be found in Appendix A). In some test we will evaluate SHARP without and with background retrieval, the later will be referred to as SHARP-B. All experimental results in this section employ raster scans. Regarding the parameter initialization of the proposed method, we set $J_0 = 5$ (after J_0 iterations reinitialize the background) for both ADP and ADPr. We use 5 iterations for Step 2 of ADPr.

We introduce two different criteria to evaluate performance. For simulation data, the ground truth can be used to compute the signal-to-noise ratio (SNR) of u^k as:

$$\text{SNR}(u^k, u_g) = -10 \log_{10} \sum_{t=0}^{n-1} |\zeta^* u^k (t + T^*) - u_g(t)|^2 / \|\zeta^* u^k\|^2, \qquad (29)$$

where u_g corresponds to the ground truth image. The error is computed up to the translation T^*, and the phase shift and scaling factor ζ^* are determined by:

$$(\zeta^*, T^*) := \arg \min_{\zeta \in \mathbb{C}, T \in \mathbb{Z}} \sum_t |\zeta u^k (t + T) - u_g(t)|^2. \qquad (30)$$

For experiment data, the R-factor for ADP and ADPr is used, defined as:

$$\text{R-factor}^k := \frac{\sum_j \left\| \sqrt{|\mathcal{A}_j(\omega^k, u^k)|^2 + (\tilde{\phi}^k)^2} - \sqrt{I_j} \right\|_1}{\|\sqrt{I}\|_1}. \qquad (31)$$

The R-factor of SHARP is defined by setting $\tilde{\phi} = 0$, whereas the R-factor of SHARP-B is defined in the same way as with ADP and ADPr.

3.1. Synthetic data

Given a true illumination and sample $\omega^\star$ and $u^\star$, respectively, and $\phi^\star$ as the parasitic noise component, the simulated intensities are generated as follows:

$$\hat{I}_j(t) \overset{\text{i.i.d}}{\sim} \text{Poisson}\left(|\mathcal{A}_j(\omega^\star, u^\star)|^2(t) + \phi^\star(t) \right) + \text{n}_j(t), \ \forall \ 0 \leq t \leq \bar{m} - 1, \qquad (32)$$

where n_j denotes the white Gaussian noise (the variance is set to $1.0 \times 10^{-6} \times \frac{1}{m} \sum_t I_{poi}(t)$, with $I_{poi}(t) := \text{Poisson}\left(|\mathcal{A}_j(\omega^\star, u^\star)|^2(t) + \phi^\star(t) \right)$). We also consider a further corruption of the intensities by outliers. They are simulated as patches of bad measurements appearing in 10% of the frames, with size $\sim 50\%$ of the frame size. The corrupted intensity values are set to 0. We use raster grid scan with stepsizes of 16 pixels, so an additional constraint of the illumination (a.k.a. support of the lens or illumination Fourier mask) is used to prevent potential periodical artifacts. In the experiments below, the frame size employed is 64×64 pixels and all algorithms stop after 1000 iterations.

Fig. 2 presents reconstruction results employing two different simulation datasets. The first experiment (second and fourth rows) uses data contaminated only by parasitic noise ($\phi^\star$). The second experiment (third and fifth rows) reconstructs data with a mix of white Gaussian and Poisson noises as in Eq. (32) and the outliers described above as well. The ground truth images for the sample, illumination and background are depicted in Figs. 2(c), 2(k), Figs. 2(a)-2(b), and Fig. 3(a), respectively. For each of the previous reconstruction experiments, we report the

retrieved background in Fig. 3 for SHARP-B, ADP and ADPr. When no mix of noises and outliers are considered, ADP produces much higher quality reconstructions (Figs. 2(f) and 2(n)) than SHARP and SHARP-B (Figs. 2(d)-2(e) and Figs. 2(l)-2(m)). The retrieved background, for both SHARP-B and ADP (Figs. 3(b)-3(c)), is consistent with those results. When additional outliers are introduced, ADP still generates better result with a dramatic reduction in outliers-related artifacts (Figs. 2(i) and 2(q)). In this case, ADPr (Figs. 2(j) and 2(r)) helps further removing the noise from the baseline ADP reconstruction. For the second experiment, the proposed algorithms also retrieve the background (Figs. 3(e)-3(f)) more precisely than SHARP-B (Fig. 3(d)). The SNRs of recovery results are reported: SNR = 13.2, 15.7, 76.7 by SHARP, SHARP-B, and ADP for the case of single parasitic noise; SNR = 9.8, 10.2, 18.5, 27.9 by SHARP, SHARP-B, ADP and ADPr for the case with additional mixture noises and outliers. Again, one can readily see the advantages of proposed ADP and ADPr.

3.2. Experimental data

Soft X-ray dataset from the ALS, Lawrence Berkeley Lab   The data used in the following experiment was published in [8]. It corresponds to a 3D ptychography imaging experiment of battery cells at 708 eV, obtained from different projection angles.

The recovered amplitude and phase are reported in Fig. 4 and zoom-in view in Fig. 5 when using SHARP, SHARP-B, ADP and ADPr. The recovered intensities of a single frame and the recovered background for the same experiment are shown in Fig. 6. When no background retrieval is used, the recovery results from SHARP (Figs. 4(a) and 4(e), and zoom-in view in Figs. 5(a) and 5(e)) are significantly noisy, with some areas presenting vague or inappreciable features. SHARP-B was specially designed for structured noise from ALS soft X-ray sources and the results in Figs. 4(b) and 4(f), and zoom-in view in Figs. 5(b) and 5(f), show a considerable improvement with respect to SHARP baseline. Visually, ADP produces even sharper results and cleaner background than SHARP-B, specially when employing regularization (Figs. 4-5 last two columns). This can also be appreciated by inspecting the recovered intensities and backgrounds from Fig 6. Artifacts due to some bad frames and other source of noise can be clearly appreciated in Figs. 4(a)-4(b), 4(e)-4(f), and Figs. 5(a)-5(b), 5(e)-5(f) in the experiments with SHARP and SHARP-B. When using the proposed algorithms, such artifacts are greatly attenuated (in Figs. 4(c)-4(d), and Figs. 5(c)-5(d)).

The R-factors of the previous experiment are 0.2399, 0.1112, 0.0957, and 0.0969, when using SHARP, SHARP-B, ADP and ADPr, respectively. This demonstrates how the proposed algorithm achieves smaller residuals thus producing higher accuracy results. We also provide cutline values results from the previous experiment in Fig. 8, selecting a line from the center of the reconstructed image. It can be seen from the figure how the proposed algorithm generates a higher contrast reconstruction than SHARP-B, specially in the retrieved phase. Convergence results are presented in Fig. 7, showing the R-factor as the iterations go. Those results illustrate how the R-factors of the proposed algorithm steadily decrease, providing additional evidence on the stability of the method.

Hard X-ray dataset from PETRA III at DESY   We also report experimental results using a dataset measured at beamline P06 at PETRA III, DESY [11] at a photon energy of 11919 eV. This dataset consists of 2 ptychographic measurements, one without and one with beamstop. The following results are generated using (1) the data without beamstop (noBS), (2) the data with beamstop (BS), and (3) the merged data from BS and noBS (the low frequencies from noBS plus the high frequency from BS). The datasets are reconstructed using ePIE (extended Ptychographic Iterative Engine [3]), which corresponds to the results presented in [11], and with the proposed ADP algorithm. For a fair comparison, all results are produced without position retrieval. The amount of noise of this dataset is significantly higher than the one reported in the

(a) True illum.(amp.)

(b) True illum.(phase)

(c) True image $(u^\star)$(amp.)

(d) SHARP *(amp.)*

(e) SHARP-B *(amp.)*

(f) ADP *(amp.)*

(g) SHARP *(amp.)*

(h) SHARP-B *(amp.)*

(i) ADP *(amp.)*

(j) ADPr *(amp.)*

(k) True image $(u^\star)$(phase)

(l) SHARP *(phase)*

(m) SHARP-B *(phase)*

(n) ADP *(phase)*

(o) SHARP *(phase)*

(p) SHARP-B *(phase)*

(q) ADP *(phase)*

(r) ADPr *(phase)*

Fig. 2. Simulation experiment. First row: (a)-(b) Amplitude and phase for true illumination (illu., 64×64 pixels), with Full Width at Half Max (FWHM) = 7 pixels. Second row: Amplitude of true image (c) with 496×496 pixels, and amplitude of recovered images with data contaminated by only parasitic noise, using SHARP (d), SHARP-B (e) and ADP (f). Third row: Amplitude of recovered images with data contaminated as in Eq. (32) and also by additional outliers, using SHARP (g), SHARP-B (h), ADP (i), and ADPr (j). The fourth, and fifth rows show the corresponding phase parts of the images in the second and third rows.
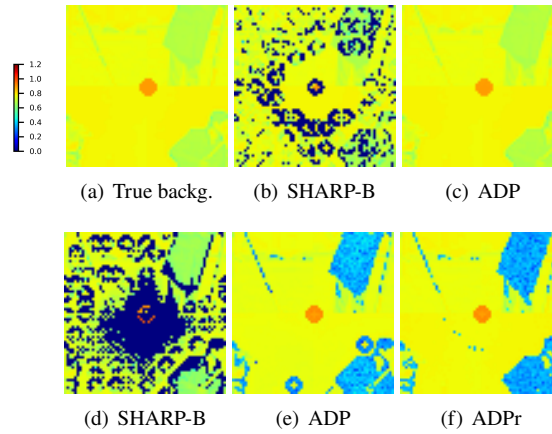
Fig. 3. Retrieved backgrounds (backg.) for the simulation experiment of Fig. 2. First row: True background (parasitic noise) (a) and results retrieved in a simulation only with parasitic noise for SHARP-B (b) and ADP (c). Second row: Background retrieved in a simulation as in Eq. (32) with additional outliers for SHARP-B (d), ADP (e) and ADPr (f).The figures are shown in scale $(\cdot)^{0.05}$ with the colorbar shown on the top left corner
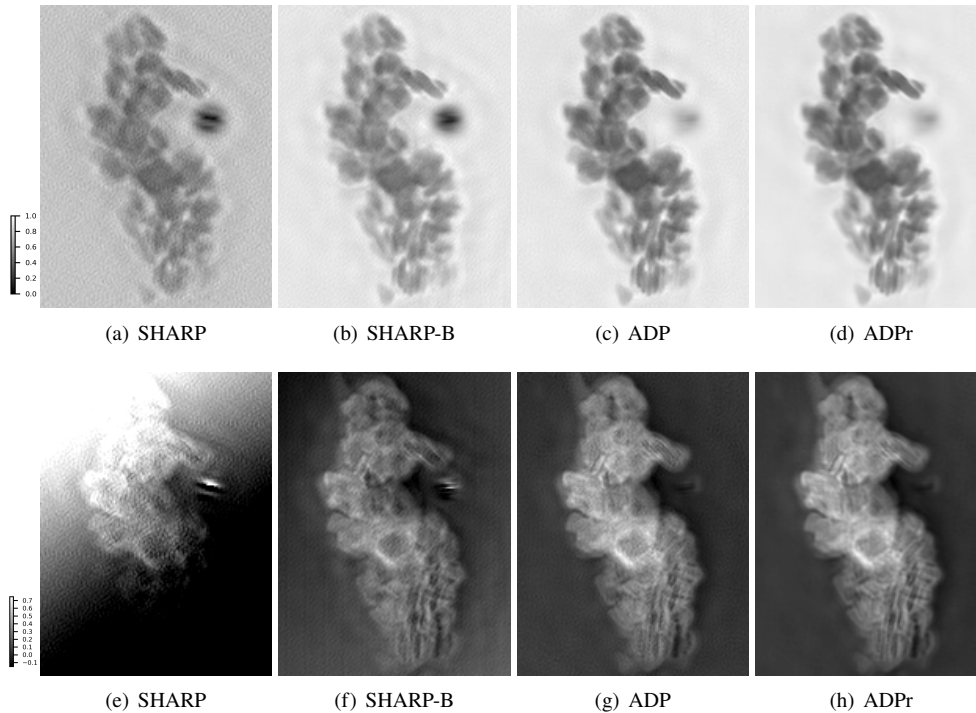
Fig. 4. Soft X-ray experimental results from the data presented in [8]. First row: reconstructed amplitude using SHARP (a), SHARP-B (b), ADP (c) and ADPr (d). Second row: reconstructed phase using SHARP (e), SHARP-B (f), ADP (g) and ADPr (h).
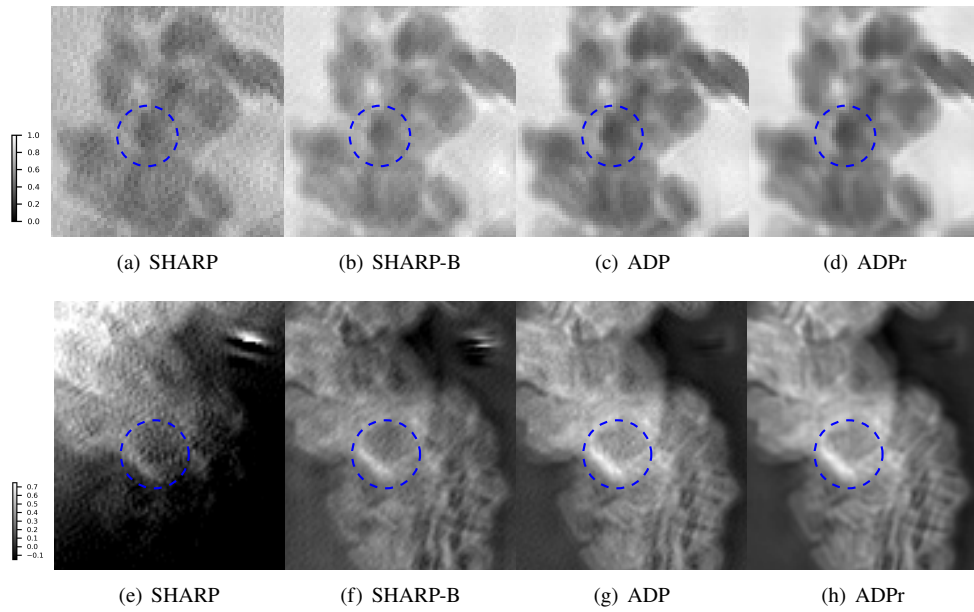
Fig. 5. Zoom-in view of parts of Fig. 4. First row: reconstructed amplitude using SHARP (a), SHARP-B (b), ADP (c) and ADPr (d). Second row: reconstructed phase using SHARP (e), SHARP-B (f), ADP (g) and ADPr (h).
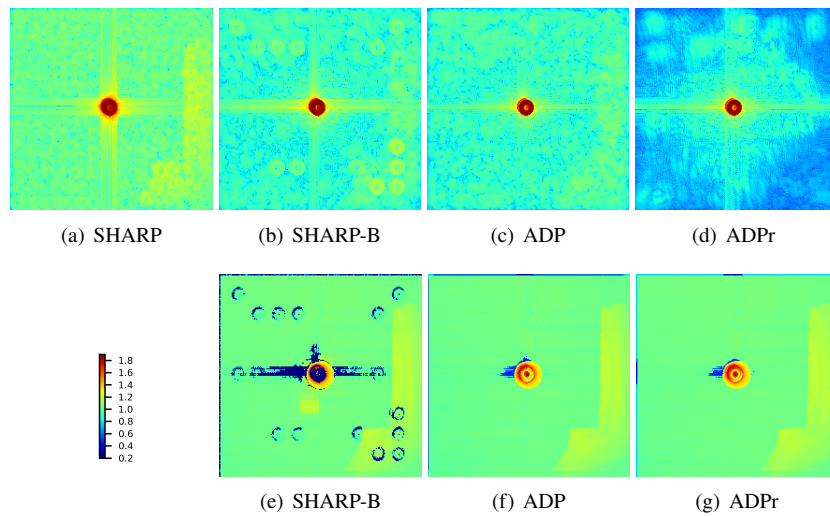
Fig. 6. Recovered intensities and background from the experiment presented in Fig. 4. First row: Recovered intensities by SHARP (a) , SHARP-B (b), ADP (c), and ADPr (d). Second row: recovered backgrounds by SHARP-B (e), ADP (f), and ADPr (g). The figures are shown in scale $(\cdot)^{0.05}$ with the colorbar shown on the down left corner.
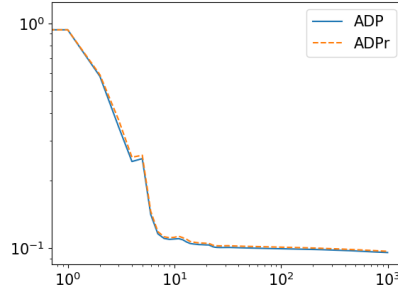
(a) R-factor

Fig. 7. Convergence histories of the R-factor for the proposed algorithm (both ADP and ADPr) when performing the experiment reported in Fig. 4, using the dataset from [8].
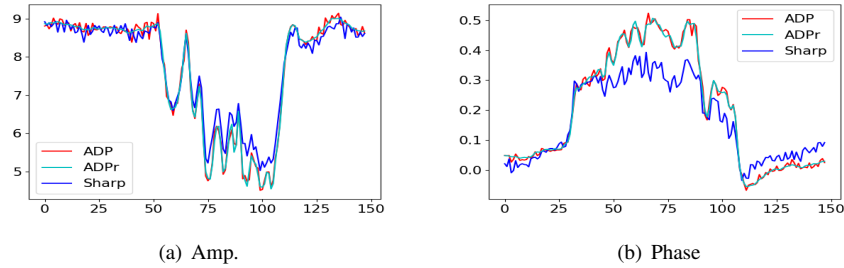
(a) Amp.

(b) Phase

Fig. 8. Cutlines of the retrieved amplitude (amp.) (a) and phase part (b) from the center of the reconstruction reported in Fig. 4, using the dataset from [8].

previous experiments: in this case, the phase contrast produced by the hard X-ray beam is much weaker than in the soft X-ray dataset. Because of this, more iterations are required to achieve reasonable reconstructions. The following results use at most 2000 iterations (we use early-stop for ePIE since it will blow up finally) for both algorithms.

The reconstruction results are shown in Fig. 9. The reconstructed phase parts using ePIE is significantly noisy, with lots of blurred out features (Fig. 9 (a) and (b)). Visually, ADP generates a much cleaner reconstructed phase parts with much better defined features (especially noticeable in the features around the color circles). ADPr in this case produces similar quality results as ADP with no regularization.

We also report an additional reconstruction result using only the BS dataset with ADP (Fig. 9 (e)). It is important to note that for this dataset low-frequency information is almost completely lost. We can see how very sharp features are well recovered in the reconstructed phase, while producing a very clean background. This illustrates the robustness of the proposed algorithm even when the measured data is incomplete or contaminated by heavy noise. Some features from the BS experiment with ADP seem to be enhanced, compared with the merged data (specially noticeable again in the color circle areas). We remark that to make our algorithm work well, a good initialization by the illumination from the non-beamstop dataset is adopted. As a future work, we will explore more efficient strategy for the initialization.

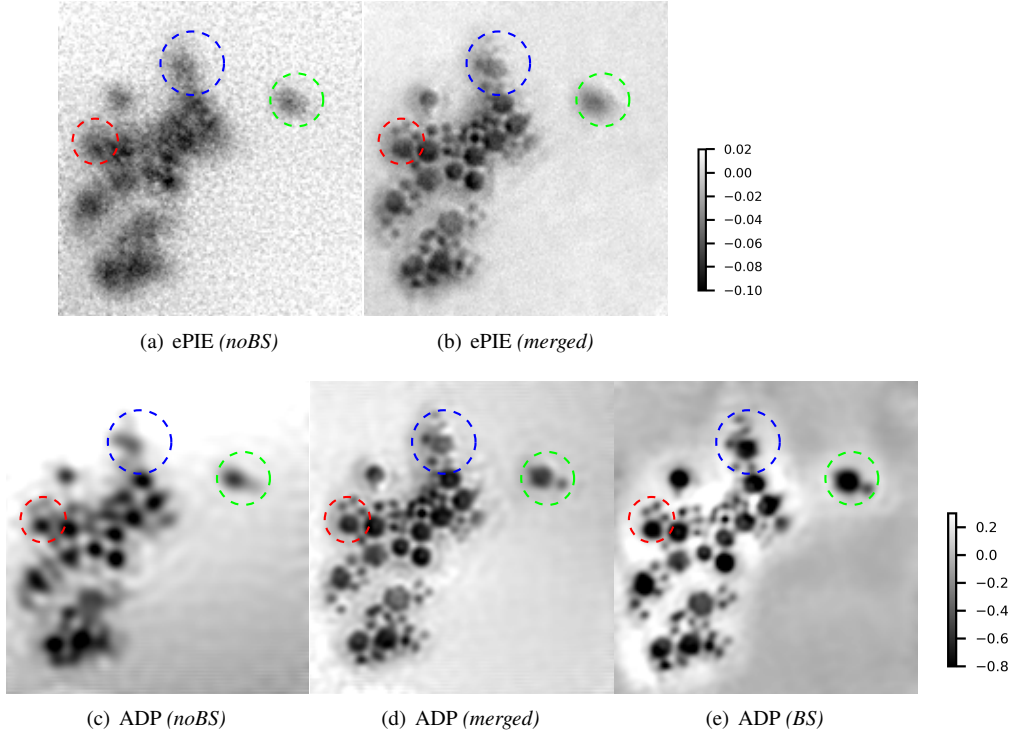Since our proposed algorithms are implemented by python, which is not specially optimized

Fig. 9. Hard X-ray experimental results from the data presented in [11]. First row: recovered phase using ePIE, with the noBS dataset (a), and with the merged dataset (b). Second row: recovered phase using ADP, with noBS data (c), merged data (d), and BS data (e).Results from (a) to (d) are displayed in the range [-0.1,0.02] (colorbar shown in the top right corner) while (e) is depicted in the range [-0.8,0.3] (colorbar shown in the down right corner).

compared with SHARP (implemented by highly optimized CUDA code). Therefore, we only give the estimate of computational complexity of proposed ADP as about $63m + 2J \times$ FFT per iteration, while about $78m + 2J \times$ FFT for SHARP-B, that demonstrates the proposed ADP needs a bit less computational cost than SHARP-B, where FFT means the computational complexity of 2D fast Fourier Transform for the matrix with size of $\sqrt{m} \times \sqrt{m}$ (same size of the illumination).

4. Conclusion

This paper proposes a novel algorithm for (blind) ptychography reconstruction with integral experimental denoising. All main experimental sources of noise are addressed and characterized in the proposed solution as a mixture of (1) structured parasitic noise, (2) a mix of random noise sources (both Gaussian or Poison), and (3) data outliers. The algorithm is formulated to be robust and efficient, requiring few arithmetic operations and being inherently parallel in most of the steps. The convergence of the method is also proved under mild conditions. Experimental results analyze a variety of datasets with different experimental noise sources and demonstrate how the proposed algorithm achieves superior reconstruction results and denoising than state-of-the-art solutions. As a future work, we will consider partial coherence [37] and position retrieval [26] in the model, and also explore additional sparsity techniques [38], deep-learning [39], and dictionary-learning methods [29] to further enhance the reconstruction results.

Appendix

A. Revisit of SHARP background retrieval algorithm

The following description was proposed in [26]. Given an intensity I_j and for the kth iteration, the background ϕ is retrieved as follows:

$$\min_{\phi, \eta \geq \eta_-} \mathcal{H}(\eta, \phi) := \tfrac{1}{2} \sum_j \|\eta \circ (I_j - \phi) - |z_j^k|^2\|^2, \tag{33}$$

given z^k, where a weight function $1 \geq \eta \in \mathbb{R}_+^{\bar{m}}$ crossing all frames is introduced to be optimized. By alternating minimization:

$$\eta^k = \arg\min_{\eta \geq \eta_-} \sum_j \|\eta \circ (I_j - \phi^{k-1}) - |z_j^k|^2\|^2 = \max\left\{\eta_-, \frac{\sum_j \langle |z_j^k|^2, I_j - \phi^{k-1}\rangle}{\sum_j |I_j - \phi^{k-1}|^2}\right\}. \tag{34}$$

Then, by the preconditioned gradient descent for the background ϕ:

$$\begin{aligned}
\phi^k &= \phi^{k-1} - \frac{1}{J|\eta^k|^2}\nabla_\phi \mathcal{H}(\eta^k, \phi) \\
&= \tfrac{1}{J} \sum_j (I_j - |z_j^k|^2) + (1 - \tfrac{1}{\eta^k}) \circ \left(\tfrac{1}{J} \sum_j |z_j^k|^2\right),
\end{aligned} \tag{35}$$

where the first term is essentially the closed form solution for the least squares problem Eq. (33) with respect to ϕ. When the algorithm converges, $\eta^k \to 1$ if $k \to +\infty$, and Eq. (35) is a gradient descent scheme with additional stabilization term $(1 - \tfrac{1}{\eta^k}) \circ \left(\tfrac{1}{J} \sum_j |z_j^k|^2\right)$, and tends to zeros as $k \to 0$, which helps to speed up the evolution of background retrieval, heuristically.

B. Derivations for ADP (Algorithm 1) and ADPr (Algorithm 2)

ADP The computation of each subproblem is presented here. For simplicity, we omit the superscripts. For the ω and $u-$subproblems, with additional proximal terms, we have:

$$
\begin{aligned}
\omega^{opt} &:= \arg\min_{\omega} \mathcal{L}(\omega, u, z, \mu, \tilde{\phi}, \Lambda_1, \Lambda_2) + \tfrac{\alpha_1}{2}\|\omega - \hat{\omega}\|^2 \\
&= \arg\min_{\omega} \tfrac{1}{2}\sum_j \|\omega \circ \mathcal{S}_j u - \mathcal{F}^*(z_j + \Lambda_{1,j})\|^2 + \tfrac{\alpha_1}{2}\|\omega - \hat{\omega}\|^2,
\end{aligned}
\tag{36}
$$

and

$$
\begin{aligned}
u^{opt} &:= \arg\min_{u} \mathcal{L}(\omega, u, z, \mu, \tilde{\phi}; \Lambda_1, \Lambda_2) + \tfrac{\alpha_2}{2}\|u - \hat{u}\|^2 \\
&= \arg\min_{u} \tfrac{1}{2}\sum_j \|\omega \circ \mathcal{S}_j u - \mathcal{F}^*(z_j + \Lambda_{1,j})\|^2 + \tfrac{\alpha_2}{2}\|u - \hat{u}\|^2,
\end{aligned}
\tag{37}
$$

where $\hat{\omega}$ and $\hat{u}$ are the approximate solutions in the previous iterations. By solving a least squares problem, we have:

$$
\begin{cases}
\omega^{opt} = \dfrac{\sum_j \mathcal{S}_j u^* \circ \mathcal{F}^*(z_j + \Lambda_{1,j}) + \alpha_1 \hat{\omega}}{\sum_j |\mathcal{S}_j u|^2 + \alpha_1 \mathbf{1}_{\bar{m}}} \\[2ex]
u^{opt} = \dfrac{\sum_j \mathcal{S}_j^T (\omega^* \circ \mathcal{F}^*(z_j + \Lambda_{1,j})) + \alpha_2 \hat{u}}{\sum_j \mathcal{S}_j^T |\omega|^2 + \alpha_2 \mathbf{1}_n}.
\end{cases}
\tag{38}
$$

The variables z_j and μ_j can be determined jointly by solving:

$$
\min_{z_j, \mu_j} \mathcal{G}_{\varepsilon,j}(z_j, \mu_j) + \tfrac{r}{2}\|z_j - (\mathcal{A}_j(\omega, u) - \Lambda_{1,j})\|^2 + \tfrac{r}{2}\|\mu_j - (\tilde{\phi} - \Lambda_{2,j})\|^2,
\tag{39}
$$

with

$$
\mathcal{G}_{\varepsilon,j}(z_j, \mu_j) := \tfrac{1}{2}\left\| \sqrt{C_j} \circ \left(\sqrt{|z_j|^2 + \mu_j^2 + \varepsilon^2 \mathbf{1}_{\bar{m}}} - \sqrt{I_j + \varepsilon^2 \mathbf{1}_{\bar{m}}} \right) \right\|^2.
\tag{40}
$$

By denoting $X_j = (z_j^T, \mu_j^T)^T \in \mathbb{C}^{2\bar{m}}$, we have:

$$
X_j^{opt} := \arg\min \mathcal{G}_{\varepsilon,j}(X_j) + \tfrac{r}{2}\|X_j - X_j^0\|^2,
\tag{41}
$$

with $X_j^0 := (\mathcal{A}_j^T(\omega, u) - \Lambda_{1,j}^T, \tilde{\phi}^T - \Lambda_{2,j}^T)^T$. The solution to the above problem has the following forms:

$$
X_j^{opt} = ((\rho_j^{opt})^T, (\rho_j^{opt})^T)^T \circ \mathrm{sign}(X_j^0),
\tag{42}
$$

with

$$
\mathrm{sign}(X_j^0) := \left(\frac{\mathcal{A}_j^T(\omega, u) - \Lambda_{1,j}^T}{|X_j^0|_*^T}, \frac{\tilde{\phi}^T - \Lambda_{2,j}^T}{|X_j^0|_*^T} \right)^T,
\tag{43}
$$

$$
|X_j^0|_* := \sqrt{|\mathcal{A}_j(\omega, u) - \Lambda_{1,j}|^2 + |\tilde{\phi} - \Lambda_{2,j}|^2},
\tag{44}
$$

where ρ_j^{opt} is determined by:

$$
\rho_j^{opt} = \arg\min_{\rho_j \in \mathbb{R}_+^{\bar{m}}} \mathcal{H}_\varepsilon(\rho_j) := \tfrac{1}{2}\langle C_j, \rho_j^2 - (I_j + \varepsilon^2 \mathbf{1}_{\bar{m}}) \circ \log(\rho_j^2 + \varepsilon^2 \mathbf{1}_{\bar{m}}) \rangle + \tfrac{r}{2}\left\| \rho_j - |X_j^0|_* \right\|^2.
\tag{45}
$$

If $\varepsilon = 0$, the closed form solution [36] to Eq. (45) is given by:

$$
\rho_j^{opt} = \frac{r|X_j^0|_* + \sqrt{r^2|X_j^0|_*^2 + 4(C_j + r\mathbf{1}_{\bar{m}}) \circ C_j \circ I_j}}{2(C_j + r\mathbf{1}_{\bar{m}})},
\tag{46}
$$

such that

$$X_j^{opt} = \frac{r|X_j^0|_* + \sqrt{r^2|X_j^0|_*^2 + 4(C_j + r\mathbf{1}_{\bar{m}}) \circ C_j \circ I_j}}{2(C_j + r\mathbf{1}_{\bar{m}})} \circ \text{sign}(X_j^0). \tag{47}$$

Otherwise, we use the projection gradient algorithm to solve it, which gives:

$$\begin{aligned}
\rho_{j,l+1} &= \max\{0, \rho_{j,l} - \tau \nabla \mathcal{H}_\epsilon(\rho_{j,l})\} \\
&= \max\left\{0, \left((1-\tau r)\mathbf{1}_{\bar{m}} - \tau C_j + \tau C_j \circ \frac{I_j + \varepsilon^2 \mathbf{1}_{\bar{m}}}{\rho_{j,l}^2 + \varepsilon^2 \mathbf{1}_{\bar{m}}}\right) \circ \rho_{j,l} + \tau r|X_j^0|_*\right\},
\end{aligned} \tag{48}$$

$l = 0, 1, \cdots$, with stepsize τ, since

$$\nabla \mathcal{H}_\epsilon(\rho_j) = \left(C_j + r\mathbf{1}_{\bar{m}} - C_j \circ \frac{I_j + \varepsilon^2 \mathbf{1}_{\bar{m}}}{\rho_j^2 + \varepsilon^2 \mathbf{1}_{\bar{m}}}\right) \circ \rho_j - r|X_j^0|_*. \tag{49}$$

For the $\tilde{\phi}$−subproblem, the solution is given directly as

$$\tilde{\phi}^{opt} = \arg\min_{\tilde{\phi}} \mathcal{L}(\omega, u, z, \mu, \tilde{\phi}, \Lambda_1, \Lambda_2) = \tfrac{1}{J} \textstyle\sum(\mu_j + \Lambda_{2,j}). \tag{50}$$

We can now further simplify the proposed algorithm. Since $0 = \tilde{\phi}^{k+1} - \frac{1}{J}\sum_j(\mu_j^{k+1} + \Lambda_{2,j}^k)$, and $\Lambda_{2,j}^{k+1} = \Lambda_{2,j}^k + \mu_j^{k+1} - \tilde{\phi}^{k+1}$, we have $\sum_j \Lambda_{2,j}^{k+1} = 0$, such that $\tilde{\phi}^{k+1} = \frac{1}{J}\sum_j \mu_j^{k+1}$. That is to say, one can replace the variable $\tilde{\phi}^k$ by the pointwise average $\frac{1}{J}\sum_j \mu_j^k$ and remove the subproblem with respect to $\tilde{\phi}$.

ADPr Similarly, we omit the superscripts for simplicity. We consider the u−subproblem, which is expressed below:

$$\min \tfrac{r}{2}\textstyle\sum_j \|\omega \circ \mathcal{S}_j u - \mathcal{F}^*(z_j + \Lambda_{1,j})\|^2 + \tfrac{\beta}{2}\sum_j \|\Lambda_{3,j} + p_j - \nabla \mathcal{S}_j u\|^2 + \tfrac{\alpha_2}{2}\|u - \hat{u}\|^2, \tag{51}$$

where $\hat{u}$ is the previous iterative solution. The minimizer to the above optimization problem satisfies the following equation:

$$\begin{aligned}
&\textstyle\sum_j \left(\text{diag}(r\mathcal{S}_j^T |\omega|^2 + \alpha_2 \mathbf{1}_n) - \beta \mathcal{S}_j^T \Delta \mathcal{S}_j\right) u \\
&= \textstyle\sum_j \left(r\mathcal{S}_j^T(\omega^* \circ \mathcal{F}^*(z_j + \Lambda_{1,j})) - \beta \mathcal{S}_j^T \text{div}(p_j + \Lambda_{3,j})\right) + \alpha_2 \hat{u},
\end{aligned} \tag{52}$$

where div denotes the divergence operator satisfying $\text{div} = -\nabla^T$, and Δ denotes the Laplacian operators satifying $\Delta = \text{div}(\nabla)$. One can readily verify that the coefficient matrix is Hermitian, and positive definite, i.e.:

$$\begin{aligned}
&\langle (\text{diag}(r\mathcal{S}_j^T |\omega|^2 + \alpha_2 \mathbf{1}_n) - \beta \mathcal{S}_j^T \Delta \mathcal{S}_j) v, v \rangle \\
&= r\langle \mathcal{S}_j^T \omega \circ v, \mathcal{S}_j^T \omega \circ v \rangle + \alpha_2 \|v\|^2 + \beta \langle \nabla \mathcal{S}_j v, \nabla \mathcal{S}_j v \rangle \\
&= r\|\mathcal{S}_j^T \omega \circ v\|^2 + \alpha_2 \|v\|^2 + \beta \|\nabla \mathcal{S}_j v\|^2 > 0, \ \forall 0 \neq v \in \mathbb{C}^n,
\end{aligned} \tag{53}$$

such that we adopt conjugate gradient (CG) algorithm to solve Eq. (52).

The p_j−subproblem has the following closed form solution:

$$p_j^{opt} = \max\{0, |\nabla \mathcal{S}_j u - \Lambda_{3,j}| - \tfrac{\lambda}{\beta}\mathbf{1}_{\bar{m}}\} \circ \frac{\nabla \mathcal{S}_j u - \Lambda_{3,j}}{|\nabla \mathcal{S}_j u - \Lambda_{3,j}|}. \tag{54}$$

The subproblems with respect to other variables are mainly the same as in the previous subsection, and we omit the details.

C. Proof of Theorem 1

The relation of the stationary points is given in the following lemma:

Lemma C.1. *Iterative solutions* $\{(\omega^k, u^k, z^k, \mu^k, \Lambda_1^k, \Lambda_2^k)\}_k$ *satisfy the following relations:*

$$
\begin{cases}
0 = \omega^{k+1} \circ \sum_j |\mathcal{S}_j u^k|^2 - \sum_j (\mathcal{S}_j u^k)^* \circ \mathcal{F}^*(z_j^k + \Lambda_{1,j}^k) + \alpha_1(\omega^{k+1} - \omega^k); \\
0 = \sum_j \mathcal{S}_j^T |\omega^{k+1}|^2 \circ u^{k+1} - \sum_j \mathcal{S}_j^T ((\omega^{k+1})^* \circ \mathcal{F}^*(z_j^k + \Lambda_{1,j}^k)) + \alpha_2(u^{k+1} - u^k); \\
0 = \nabla \mathcal{G}_\varepsilon(z^{k+1}, \mu^{k+1}) + r((\Lambda_1^{k+1})^T, (\Lambda_2^{k+1})^T)^T; \\
0 = -\Lambda_{1,j}^{k+1} + \Lambda_{1,j}^k + z_j^{k+1} - \mathcal{A}_j(\omega^{k+1}, u^{k+1}); \\
0 = -\Lambda_{2,j}^{k+1} + \Lambda_{2,j}^k + \mu_j^{k+1} - \frac{1}{J} \sum_j \mu_j^{k+1}.
\end{cases}
\tag{55}
$$

Proof. By Eq. (47), one has:

$$
\sqrt{|z_j^{k+1}|^2 + |\mu_j^{k+1}|^2} = \frac{\sqrt{I_j} + r|Y_j^k|}{1+r} \geq \frac{\sqrt{I_j}}{1+r} > 0,
\tag{56}
$$

such that the objective function in Eq. (41) is smooth at $X_j^{k+1} := (z_j^{k+1}, \mu_j^{k+1})$. Then one can readily derive Eq. (55) by calculating the first order derivative for each subproblem. ☐

Lemma C.2.

$$
\|\nabla \mathcal{G}_\varepsilon(z^1, \mu^1) - \nabla \mathcal{G}_\varepsilon(z^2, \mu^2)\| \leq L_\varepsilon(\|z^1 - z^2\| + \|\mu^1 - \mu^2\|),
\tag{57}
$$

where L_ϵ is a positive constant independent with z^1, μ^1, z^2, μ^2.

Proof. It can be readily proved following [15], and we omit the details here. ☐

Lemma C.3. *Denoting $Y^k := (\omega^k, u^k, z^k, \mu^k, \tilde{\phi}^k, \Lambda_1^k, \Lambda_2^k)$, if $r > 4L_\varepsilon$,*

$$
\mathcal{L}(Y^k) - \mathcal{L}(Y^{k+1}) \geq c_{\varepsilon,r} \|Y^{k+1} - Y^k\|^2,
\tag{58}
$$

where $c_{\varepsilon,r}$ is a positive parameter independent from $\{Y^k\}$.

Proof. Regarding the ω and $u-$subproblems:

$$
\begin{aligned}
& \mathcal{L}(Y^k) - \mathcal{L}(\omega^{k+1}, u^{k+1}, z^k, \mu^k, \tilde{\phi}^k, \Lambda_1^k, \Lambda_2^k) \\
& \geq \tfrac{1}{2}\|\mathcal{A}(\omega^{k+1} - \omega^k, u^k)\|^2 + \tfrac{\alpha_1}{2}\|\omega^{k+1} - \omega^k\|^2 \\
& + \tfrac{1}{2}\|\mathcal{A}(\omega^{k+1}, u^{k+1} - u^k)\|^2 + \tfrac{\alpha_2}{2}\|u^{k+1} - u^k\|^2.
\end{aligned}
\tag{59}
$$

For the z and μ subproblems, by Eq. (39) and the Lemma C.2, we have:

$$
\begin{aligned}
& \mathcal{L}(\omega^{k+1}, u^{k+1}, z^k, \mu^k, \tilde{\phi}^k, \Lambda_1^k, \Lambda_2^k) - \mathcal{L}(\omega^{k+1}, u^{k+1}, z^{k+1}, \mu^{k+1}, \tilde{\phi}^k, \Lambda_1^k, \Lambda_2^k) \\
& \geq \tfrac{r-3L_\varepsilon}{2}(\|z_j^{k+1} - z_j^k\|^2 + \|\mu_j^{k+1} - \mu_j^k\|^2).
\end{aligned}
\tag{60}
$$

For the subproblems of $\tilde{\phi}$, we have:

$$
\begin{aligned}
& \mathcal{L}(\omega^{k+1}, u^{k+1}, z^{k+1}, \mu^{k+1}, \tilde{\phi}^k, \Lambda_1^k, \Lambda_2^k) - \mathcal{L}(\omega^{k+1}, u^{k+1}, z^{k+1}, \mu^{k+1}, \tilde{\phi}^{k+1}, \Lambda_1^k, \Lambda_2^k) \\
& \geq \tfrac{rJ}{2}\|\tilde{\phi}^{k+1} - \tilde{\phi}^k\|^2.
\end{aligned}
\tag{61}
$$

Further we have:

$$\mathcal{L}(\omega^{k+1}, u^{k+1}, z^{k+1}, \mu^{k+1}, \tilde{\phi}^{k+1}, \Lambda_1^k, \Lambda_2^k) - \mathcal{L}(Y^{k+1})$$

$$\overset{\text{Eq.(20)}}{=} -r \sum_j (\|\Lambda_{1,j}^{k+1} - \Lambda_{1,j}^k\|^2 + \|\Lambda_{2,j}^{k+1} - \Lambda_{2,j}^k\|^2) \tag{62}$$

$$\overset{\text{Eq.(55)}}{\geq} -\frac{2L_\varepsilon^2}{r}(\|z^{k+1} - z^k\|^2 + \|\omega^{k+1} - \omega^k\|^2),$$

where the last relation is derived by Lemma C.2.

Finally, by summing up Eqs. (59)-(62) together, we get:

$$\mathcal{L}(Y^k) - \mathcal{L}(Y^{k+1}) \geq \min\left\{\frac{\alpha_1}{2}, \frac{\alpha_2}{2}, \frac{r-3L_\varepsilon}{2} - \frac{2L_\varepsilon^2}{r}, \frac{rJ}{2}\right\} \|Y^{k+1} - Y^k\|^2. \tag{63}$$

Setting $r > 4L_\varepsilon$ concludes this lemma. $\qquad\square$

Proof of Theorem 1. We first show that $\{\mathcal{L}(Y^k)\}$ is lower bounded:

$$\begin{aligned}
\mathcal{L}(Y^{k+1}) &= \mathcal{G}_\varepsilon(z^{k+1}, \mu^{k+1}) - \Re\langle\nabla_z\mathcal{G}_{\varepsilon,j}(z_j^{k+1}, \mu_j^{k+1}), z_j^{k+1} - \mathcal{A}_j(\omega^{k+1}, u^{k+1})\rangle \\
&\quad - \langle\nabla_\mu\mathcal{G}_{\varepsilon,j}(z_j^{k+1}, \mu_j^{k+1}), \mu_j^{k+1} - \tilde{\phi}^{k+1}\rangle + \frac{r}{2}\|z_j^{k+1} - \mathcal{A}_j(\omega^{k+1}, u^{k+1})\|^2 \\
&\quad + \frac{r}{2}\|\mu_j^{k+1} - \tilde{\phi}^{k+1}\|^2 \\
&= \mathcal{G}_\varepsilon(z^{k+1}, \mu^{k+1}) - \mathcal{G}_\varepsilon(\mathcal{A}_j(\omega^{k+1}, u^{k+1}), \tilde{\phi}^{k+1}) \\
&\quad - \Re\langle\nabla\mathcal{G}_{\varepsilon,j}(z_j^{k+1}, \mu_j^{k+1}), ((z_j^{k+1} - \mathcal{A}_j(\omega^{k+1}, u^{k+1}))^T, (\mu_j^{k+1} - \tilde{\phi}^{k+1})^T)^T\rangle \\
&\quad + \frac{r}{2}\|z_j^{k+1} - \mathcal{A}_j(\omega^{k+1}, u^{k+1})\|^2 + \frac{r}{2}\|\mu_j^{k+1} - \tilde{\phi}^{k+1}\|^2 + \mathcal{G}_\varepsilon(\mathcal{A}_j(\omega^{k+1}, u^{k+1}), \tilde{\phi}^{k+1}) \\
&\geq \frac{r-L_\varepsilon}{2}\left(\|z_j^{k+1} - \mathcal{A}_j(\omega^{k+1}, u^{k+1})\|^2 + \|\mu_j^{k+1} - \tilde{\phi}^{k+1}\|^2\right) + \mathcal{G}_\varepsilon(\mathcal{A}_j(\omega^{k+1}, u^{k+1}), \tilde{\phi}^{k+1}),
\end{aligned} \tag{64}$$

where the last relation is derived from the Lemma C.2. Therefore, $\mathcal{L}(Y^k) > -\infty$ with $r \geq 4L_\varepsilon$. Furthermore, by the Lemma C.3:

$$+\infty > \mathcal{L}(Y^0) - \mathcal{L}(Y^k) = \sum_{l=0}^k (\mathcal{L}(Y^l) - \mathcal{L}(Y^{l+1})) \geq c_{\varepsilon,r} \sum_{l=0}^k \|Y^{l+1} - Y^l\|^2. \tag{65}$$

Accordingly we get:

$$\lim_{l\to+\infty} \|Y^{l+1} - Y^l\| = 0. \tag{66}$$

For any limit point $Y^\star := (\omega^\star, u^\star, z^\star, \mu^\star, \tilde{\phi}^\star, \Lambda_1^\star, \Lambda_2^\star)$ of the iterative sequence $\{Y^k\}$, there exists a subsequence $Y^{n_k} := (\omega^{n_k}, u^{n_k}, z^{n_k}, \mu^{n_k}, \tilde{\phi}^{n_k}, \Lambda_1^{n_k}, \Lambda_2^{n_k}) \subset Y^k$, such that

$$\lim_{k\to+\infty} Y^{n_k} = Y^\star. \tag{67}$$

We can see that $\{Y^{n_k}\}$ is bounded such that one can derive: $\lim_{k\to+\infty}\mathcal{A}(\omega^{n_k}, u^{n_k}) = \mathcal{A}(\omega^\star, u^\star)$, and $\lim_{k\to+\infty}\sum_j \mathcal{S}_j^T |\omega^{n_k}|^2 \circ u^{n_k} = \sum_j \mathcal{S}_j^T |\omega^\star|^2 \circ u^\star$. By Eq. (66), $\{Y^{n_k-1}\}$ is bounded as well. Hence

$$\lim_{k\to+\infty} \omega^{n_k} \circ \sum_j |\mathcal{S}_j u^{n_k-1}|^2 = \omega^\star \circ \sum_j |\mathcal{S}_j u^\star|^2, \tag{68}$$

$$\lim_{k\to+\infty} \sum_j (\mathcal{S}_j u^{n_k-1})^* \circ \mathcal{F}^*(z_j^{n_k-1} + \Lambda_{1,j}^{n_k-1}) = \sum_j (\mathcal{S}_j u^\star)^* \circ \mathcal{F}^*(z_j^\star + \Lambda_{1,j}^\star), \tag{69}$$

and

$$\lim_{k \to +\infty} \sum_j \mathcal{S}_j^T((\omega^{n_k})^* \circ \mathcal{F}^*(z_j^{n_k-1} + \Lambda_{1,j}^{n_k-1})) = \sum_j \mathcal{S}_j^T((\omega^\star)^* \circ \mathcal{F}^*(z_j^\star + \Lambda_{1,j}^\star)). \tag{70}$$

By Lemma C.2:

$$\lim_{k \to +\infty} \nabla \mathcal{G}_\varepsilon(z^{n_k}, \mu^{n_k}) = \nabla \mathcal{G}_\varepsilon(z^\star, \mu^\star). \tag{71}$$

Finally, since

$$
\begin{aligned}
0 &= \omega^{n_k} \circ \sum_j |\mathcal{S}_j u^{n_k-1}|^2 - \sum_j (\mathcal{S}_j u^{n_k-1})^* \circ \mathcal{F}^*(z_j^{n_k-1} + \Lambda_{1,j}^{n_k-1}) + \alpha_1(\omega^{n_k} - \omega^{n_k-1}); \\
0 &= \sum_j \mathcal{S}_j^T |\omega^{n_k}|^2 \circ u^{n_k} - \sum_j \mathcal{S}_j^T((\omega^{n_k})^* \circ \mathcal{F}^*(z_j^{n_k-1} + \Lambda_{1,j}^{n_k-1})) + \alpha_2(u^{n_k} - u^{n_k-1}); \\
0 &= \nabla \mathcal{G}_\varepsilon(z^{n_k}, \mu^{n_k}) + r((\Lambda_1^{n_k})^T, (\Lambda_2^{n_k})^T)^T; \\
0 &= -\Lambda_{1,j}^{n_k} + \Lambda_{1,j}^{n_k-1} + z_j^{n_k} - \mathcal{A}_j(\omega^{n_k}, u^{n_k}); \\
0 &= -\Lambda_{2,j}^{n_k} + \Lambda_{2,j}^{n_k-1} + \mu_j^{n_k} - \frac{1}{J} \sum_j \mu_j^{n_k},
\end{aligned} \tag{72}
$$

following the above calculations of these limits, we can get:

$$
\begin{cases}
0 = \omega^\star \circ \sum_j |\mathcal{S}_j u^\star|^2 - \sum_j (\mathcal{S}_j u^\star)^* \circ \mathcal{F}^*(z_j^\star + \Lambda_{1,j}^*); \\
0 = \sum_j \mathcal{S}_j^T |\omega^\star|^2 \circ u^\star - \sum_j \mathcal{S}_j^T((\omega^\star)^* \circ \mathcal{F}^*(z_j^\star + \Lambda_{1,j}^\star)); \\
0 = \nabla \mathcal{G}_\varepsilon(z^\star, \mu^\star) + r((\Lambda_1^\star)^T, (\Lambda_2^\star)^T)^T; \\
0 = z_j^\star - \mathcal{A}_j(\omega^\star, u^\star); \\
0 = \mu_j^\star - \frac{1}{J} \sum_j \mu_j^\star,
\end{cases} \tag{73}
$$

which immediately implies that $Y^\star$ is the stationary point of Eq. (7) and concludes this theorem. $\qquad\square$

References

1. P. Nellist, B. McCallum, and J. Rodenburg, "Resolution beyond the 'information limit' in transmission electron microscopy," Nature **374**, 630 (1995).
2. J. M. Rodenburg and H. M. Faulkner, "A phase retrieval algorithm for shifting illumination," Applied physics letters **85**, 4795–4797 (2004).
3. A. M. Maiden and J. M. Rodenburg, "An improved ptychographical phase retrieval algorithm for diffractive imaging," Ultramicroscopy **109**, 1256–1262 (2009).
4. Y. Jiang, Z. Chen, Y. Han, P. Deb, H. Gao, S. Xie, P. Purohit, M. W. Tate, J. Park, S. M. Gruner, V. Elser, and D. A. Muller, "Electron ptychography of 2d materials to deep sub-Ångström resolution," Nature **559**, 343 (2018).
5. X. Shi, P. Fischer, V. Neu, D. Elefant, J. Lee, D. Shapiro, M. Farmand, T. Tyliszczak, H.-W. Shiu, S. Marchesini, and S. Roy, "Soft x-ray ptychography studies of nanoscale magnetic and structural correlations in thin SmCo$_5$ films," Applied Physics Letters **108**, 094103 (2016).
6. K. Giewekemeyer, P. Thibault, S. Kalbfleisch, A. Beerlink, C. M. Kewish, M. Dierolf, F. Pfeiffer, and T. Salditt, "Quantitative biological imaging by ptychographic x-ray diffraction microscopy," Proceedings of the National Academy of Sciences **107**, 529–534 (2010).
7. D. A. Shapiro, Y.-S. Yu, T. Tyliszczak, J. Cabana, R. Celestre, W. Chao, K. Kaznatcheev, A. D. Kilcoyne, F. Maia, S. Marchesini, Y. S. Meng, T. Warwick, L. L. Yang, and H. Padmore, "Chemical composition mapping with nanometre resolution by soft x-ray microscopy," Nature Photonics **8**, 765–769 (2014).
8. Y.-S. Yu, M. Farmand, C. Kim, Y. Liu, C. P. Grey, F. C. Strobridge, T. Tyliszczak, R. Celestre, P. Denes, J. Joseph, H. Krishnan, F. R. N. C. Maia, A. L. D. Kilcoyne, S. Marchesini, T. P. C. Leite, T. Warwick, H. Padmore, J. Cabana, and D. A. Shapiro, "Three-dimensional localization of nanoscale battery reactions using soft x-ray tomography," Nature communications **9**, 921 (2018).
9. M. Holler, M. Guizar-Sicairos, E. H. Tsai, R. Dinapoli, E. Müller, O. Bunk, J. Raabe, and G. Aeppli, "High-resolution non-destructive three-dimensional imaging of integrated circuits," Nature **543**, 402–406 (2017).
10. M. O. Wiedorn, S. Awel, A. J. Morgan, M. Barthelmess, R. Bean, K. R. Beyerlein, L. M. G. Chavas, N. Eckerskorn, H. Fleckenstein, M. Heymann, D. A. Horke, J. Knoška, V. Mariani, D. Oberthür, N. Roth, O. Yefanov, A. Barty, S. Bajt, J. Küpper, A. V. Rode, R. A. Kirian, and H. N. Chapman, "Post-sample aperture for low background diffraction experiments at x-ray free-electron lasers," Journal of Synchrotron Radiation **24**, 1296–1298 (2017).

11. J. Reinhardt, R. Hoppe, G. Hofmann, C. D. Damsgaard, J. Patommel, C. Baumbach, S. Baier, A. Rochet, J.-D. Grunwaldt, G. Falkenberg, and C. G. Schroer, "Beamstop-based low-background ptychography to image weakly scattering objects," Ultramicroscopy **173**, 52–57 (2017).
12. C. Wang, Z. Xu, H. Liu, Y. Wang, J. Wang, and R. Tai, "Background noise removal in x-ray ptychography," Applied optics **56**, 2099–2111 (2017).
13. P. Thibault, M. Dierolf, O. Bunk, A. Menzel, and F. Pfeiffer, "Probe retrieval in ptychographic coherent diffractive imaging," Ultramicroscopy **109**, 338–343 (2009).
14. R. Hesse, D. R. Luke, S. Sabach, and M. K. Tam, "Proximal heterogeneous block implicit-explicit method and application to blind ptychographic diffraction imaging," SIAM Journal on Imaging Sciences **8**, 426–457 (2015).
15. H. Chang, P. Enfedaque, and S. Marchesini, "Blind ptychographic phase retrieval via convergent alternating direction method of multipliers," SIAM Journal on Imaging Sciences **12**, 153–185 (2019).
16. V. Elser, "Phase retrieval by iterated projections," J. Opt. Soc. Am. A **20**, 40–55 (2003).
17. P. Thibault and M. Guizar-Sicairos, "Maximum-likelihood refinement for coherent diffractive imaging," New Journal of Physics **14**, 063004 (2012).
18. D. R. Luke, "Relaxed averaged alternating reflections for diffraction imaging," Inverse Probl. **21**, 37–50 (2005).
19. S. Marchesini, H. Krishnan, D. A. Shapiro, T. Perciano, J. A. Sethian, B. J. Daurer, and F. R. Maia, "SHARP: a distributed, GPU-based ptychographic solver," Journal of Applied Crystallography **49**, 1245–1252 (2016).
20. R. Glowinski and P. Le Tallec, *Augmented Lagrangian and operator-splitting methods in nonlinear mechanics* (Philadelphia, PA: SIAM, 1989).
21. C. Wu and X.-C. Tai, "Augmented Lagrangian method, dual methods and split-Bregman iterations for ROF, vectorial TV and higher order models," SIAM J. Imaging Sci. **3**, 300–339 (2010).
22. 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**, 1–122 (2011).
23. P. Godard, M. Allain, V. Chamard, and J. Rodenburg, "Noise models for low counting rate coherent diffraction imaging," Optics express **20**, 25914–25934 (2012).
24. M. Odstrčil, A. Menzel, and M. Guizar-Sicairos, "Iterative least-squares solver for generalized maximum-likelihood ptychography," Optics express **26**, 3108–3123 (2018).
25. B. Enders, M. Dierolf, P. Cloetens, M. Stockmar, F. Pfeiffer, and P. Thibault, "Ptychography with broad-bandwidth radiation," Applied Physics Letters **104**, 171104 (2014).
26. S. Marchesini, A. Schirotzek, C. Yang, H.-T. Wu, and F. Maia, "Augmented projections for ptychographic imaging," Inverse Problems **29**, 115009 (2013).
27. B. J. Daurer, H. Krishnan, T. Perciano, F. R. Maia, D. A. Shapiro, J. A. Sethian, and S. Marchesini, "Nanosurveyor: a framework for real-time data processing," Advanced structural and chemical imaging **3**, 7 (2017).
28. H. Chang, Y. Lou, Y. Duan, and S. Marchesini, "Total variation–based phase retrieval for Poisson noise removal," SIAM Journal on Imaging Sciences **11**, 24–55 (2018).
29. H. Chang and S. Marchesini, "Denoising Poisson phaseless measurements via orthogonal dictionary learning," Optics Express **26**, 19773–19796 (2018).
30. F. Murtagh, J.-L. Starck, and A. Bijaoui, "Image restoration with noise suppression using a multiresolution support." Astronomy and Astrophysics Supplement Series **112**, 179 (1995).
31. A. Chakrabarti and T. Zickler, "Image restoration with signal-dependent camera noise," arXiv:1204.2994 (2012).
32. J. Li, Z. Shen, R. Yin, and X. Zhang, "A reweighted l^2 method for image restoration with poisson and mixed poisson-gaussian noise," Inverse Probl. Imaging **9**, 875–894 (2015).
33. N. M. Kirby, S. T. Mudie, A. M. Hawley, D. J. Cookson, H. D. Mertens, N. Cowieson, and V. Samardzic-Boban, "A low-background-intensity focusing small-angle x-ray scattering undulator beamline," Journal of Applied Crystallography **46**, 1670–1680 (2013).
34. H. Chang, S. Marchesini, Y. Lou, and T. Zeng, "Variational phase retrieval with globally convergent preconditioned proximal algorithm," SIAM Journal on Imaging Sciences **11**, 56–93 (2018).
35. Z. Wen, C. Yang, X. Liu, and S. Marchesini, "Alternating direction methods for classical and ptychographic phase retrieval," Inverse Probl. **28**, 115010 (2012).
36. H. Chang, Y. Lou, M. K. Ng, and T. Zeng, "Phase retrieval from incomplete magnitude information via total variation regularization," SIAM Journal on Scientific Computing **38**, A3672–A3695 (2016).
37. H. Chang, P. Enfedaque, Y. Lou, and S. Marchesini, "Partially coherent ptychography by gradient decomposition of the probe," Acta Crystallographica Section A: Foundations and Advances **74**, 157–169 (2018).
38. H. Chang and S. Marchesini, "A general framework for denoising phaseless diffraction measurements," arXiv preprint arXiv:1611.01417 (2016).
39. M. J. Cherukara, Y. S. Nashed, and R. J. Harder, "Real-time coherent diffraction inversion using deep generative networks," arXiv preprint arXiv:1806.03992 (2018).