Title

Partially coherent ptychography by gradient decomposition of the probe

Journal

Acta Crystallographica Section A: Foundations and advances, 74(3)

ISSN

0108-7673

Authors

Chang, Huibin
Enfedaque, Pablo
Lou, Yifei
et al.

Publication Date

2018-05-01

DOI

10.1107/s2053273318001924

Peer reviewed

Partially Coherent Ptychography by Gradient Decomposition of the Probe

Huibin Chang,[1,2] Pablo Enfedaque,[2] Yifei Lou,[3] and Stefano Marchesini[2,*]
[1] *School of Math. Sci., Tianjin Normal University*
[2] *Computational Research Division, Lawrence Berkeley National Laboratory*
[3] *Dept. of Math, University of Texas, Dallas*
(Dated: October 23, 2017)

Coherent ptychographic imaging experiments often discard over 99.9 % of the flux from a light source to define the coherence of an illumination. Even when coherent flux is sufficient, the stability required during an exposure is another important limiting factor. Partial coherence analysis can considerably reduce these limitations. A partially coherent illumination can often be written as the superposition of a single coherent illumination convolved with a separable translational kernel. In this paper we propose the Gradient Decomposition of the Probe (*GDP*), a model that exploits translational kernel separability, coupling the variances of the kernel with the transverse coherence. We describe an efficient first-order splitting algorithm *GDP-ADMM* to solve the proposed nonlinear optimization problem. Numerical experiments demonstrate the effectiveness of the proposed method with Gaussian and binary kernel functions in *fly-scan* measurements. Remarkably, *GDP-ADMM* produces satisfactory results even when the ratio between kernel width and beam size is more than one, or when the distance between successive acquisitions is twice as large as the beam width.

I. INTRODUCTION

Ptychography is a popular imaging technique in scientific fields as diverse as condensed matter physics [25], cell biology [11], materials science [24] and electronics [15], among others. In a coherent ptychography experiment (Figure 1), a localized coherent X-ray probe (or illumination) ω scans through an specimen u, while the detector collects a sequence J of phaseless intensities f in the far field. The goal is to obtain a high resolution reconstruction of the specimen u from the sequence of intensity measurements. In a discrete setting, $u \in \mathbb{C}^n$ is a 2D image with $\sqrt{n} \times \sqrt{n}$ pixels, $\omega \in \mathbb{C}^{\bar{m}}$ is a localized 2D probe with $\sqrt{\bar{m}} \times \sqrt{\bar{m}}$ pixels (u and ω are both written as a vector by a lexicographical order), and $f_j = |\mathcal{F}(\omega \circ \mathcal{S}_j u)|^2$ is a stack of phaseless measurements $f_j \in \mathbb{R}_+^{\bar{m}}$ $\forall 0 \le j \le J-1$. Here $|\cdot|$ represents the elementwise absolute value of a vector, $\circ$ denotes the elementwise multiplication, and $\mathcal{F}$ denotes the normalized 2 dimensional discrete Fourier transform. Each $\mathcal{S}_j \in \mathbb{R}^{\bar{m}\times n}$ is a binary matrix that crops a region j of size $\bar{m}$ from the image u.

In practice, as the probe is almost never completely known, one has to solve a blind ptychographic phase retrieval problem or probe retrieval [26], as follows:

$$\text{To find } \ \omega \in \mathbb{C}^{\bar{m}} \text{ and } u \in \mathbb{C}^n, \quad s.t. \ |\mathcal{A}(\omega, u)|^2 = f, \tag{1}$$

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 follows:

$$\begin{aligned}\mathcal{A}(\omega, u) :=& (\mathcal{A}_0^T(\omega, u), \mathcal{A}_1^T(\omega, u), \cdots, \mathcal{A}_{J-1}^T(\omega, u))^T,\\ \mathcal{A}_j(\omega, u) :=& \mathcal{F}(\omega \circ \mathcal{S}_j u),\end{aligned}$$

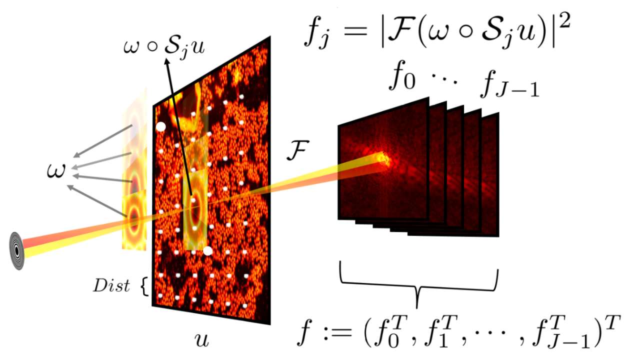


FIG. 1. Ptychographic experiment (Far-field): A stack of intensities $f_j = |\mathcal{F}(\omega \circ \mathcal{S}_j u)|^2$ are collected, with ω being the localized coherent probe and u being the image of interest (or specimen). The white dots on the image represent the center of probe or scanning lattice points with *Dist* denoting the sliding distance between centers of two successive frames.

and $f := (f_0^T, f_1^T, \cdots, f_{J-1}^T)^T \in \mathbb{R}_+^m$.

Coherent ptychographic imaging experiments often rely on apertures to define a coherent illumination. Research institutions around the world are investing considerable resources to produce brighter x-ray sources in order to overcome this limitation. Meanwhile, most of the x-ray photons generated are currently discarded by secondary apertures. Even when there is enough coherent flux, the stability required during an exposure is often another limiting factor. Both flux and stability limitations can be reduced using partial coherence analysis [7, 16, 21].

First, we briefly review the existing algorithmic work for partial coherence. In [10] the authors applied a gradient descent phase retrieval algorithm from incoherently averaged illuminations to compute the aberration of the Hubble space telescope. In [5], the authors considered a constant quasi-monochromatic illumination on the sam-

* smarchesini@lbl.gov

ple (beam much larger than the sample), and derived a convolution based model using the mutual optical intensity as:

$$f_{pc} = f * \kappa, \tag{2}$$

where f_{pc} is the measured partial coherent intensity, f is the coherent intensity, $*$ denotes the convolution operator, and κ is the kernel function (Fourier transform of the complex coherence function). Physically, the convolution kernel represents the combination of the detector response function and the angular spread of the illumination. It was solved by the alternating projection algorithm in [5], when κ is a Gaussian kernel function with free horizontal and vertical coherence length parameters. Burdet et al. [2] applied the above model to ptychographic imaging and proposed the Douglas-Rachford algorithm with a known Gaussian kernel function.

A more general description of a partially coherent wave field illuminating a specimen in a ptychography experiment was considered in [27], where the authors employed an orthogonal decomposition of the mutual coherence function [30] to describe partially coherent measurement as follows:

$$f_{pc,j} = \sum\nolimits_{l=0}^{L-1} |\mathcal{A}_j(\omega_l, u)|^2,\ \forall 0 \le j \le J-1, \tag{3}$$

with L orthogonal probes $\{\omega_l\}$, where $f_{pc} := (f_{pc,0}^T, \cdots, f_{pc,J-1}^T)^T$ is the measured intensity. The extended ptychographic iterative engine [18] was adopted to solve such model [3]. Experimental '*fly-scan*' data with translational blur was successfully reconstructed using (3) by [7, 16, 21]. However, it is important to note that such blur is a special case of a model that has many more degrees of freedom. Moreover, the physical interpretation of the multiple modes is unclear, and relationship with the coherence function is indirect.

Other existing studies focus mostly on (2), which only addressed blurring coherent intensities at the detector. In practice, the blur is dominated by the source dimensions and by the translation of the probe with respect to the image during an exposure [7, 21]. To the best of our knowledge, there is no algorithm in the literature to jointly recover the unknown image, probe, and coherence kernel function that exploits such property.

In this paper, we propose *G*radient *D*ecomposition of the *P*robe (*GDP*), a new forward model to characterize the partially coherent ptychography problem. The new model is based on coupling the experimental coherence widths with the variances of the kernel functions using the second order Taylor expansion to translate the probe. In the second part of this work we present *GDP-ADMM*: a novel fast iterative solver that jointly optimizes the image, the probe and the variance of the kernel function. The main benefits of the algorithm are listed below:

1. The approximation accuracy for a general partial coherent source is increased, while providing its coherent properties.
2. Subproblems can be solved using low computational and memory resources, and usually have close form.
3. It is insensitive to the coherence kernel and scanning step sizes, and achieves high SNR even when the data is contaminated by Poisson noise.

Numerical experiments show that satisfactory results with *GDP-ADMM* can be obtained even when the ratio between kernel and beam widths is more than one, or when the distance between successive acquisitions is almost twice as large as the beam width (Full Width Half Maximum-FWHM).

II. APPROXIMATE FORWARD MODEL FOR PARTIAL COHERENCE

Partially coherent illumination from standard microscopes can be written as the superposition of a single quasi-monochromatic coherent illumination convolved with separable angular and translational kernels [6, 13, 17, 22]. Translational convolution of the illumination is equivalent to translational motion during a scan while exposing the detector [7]. In other words, an extended incoherent source upstream of the lens can be viewed as the superposition, photon-by-photon, of a rapidly moving source demagnified by the lens onto the image plane. The demagnified source defines the degree of coherence of the probe, or the blurring kernel. Coherence and vibrations kernels can be combined into one, such that partially coherent ptychography imaging with coherence kernel function κ in a continuous setting (same notations as in the discrete setting) is formulated as:

$$\int |\mathcal{F}_{x\to q}\left(\mathcal{S}_j u\left(x\right)\omega\left(x-y\right)\right)|^2 \kappa(y)\mathrm{d}y = f_{pc,j}(q), \tag{4}$$

where f_{pc} is the measured partial coherent intensity and $\mathcal{F}_{x\to q}$ is the normalized Fourier transform. When setting κ to a binary function, the above model is exactly the same as *fly-scan* ptychography [7, 16, 21]; Setting κ to the Dirac delta function reduces it to the coherent model (1). We remark that (4) is quite different from (2), since (4) illustrates the effects of blurring of images with respect to the probe, while (2) can be interpreted as blurring or binning multiple pixels at the detector.

Generally speaking, solving Eq. (4) is a non-linear ill-posed problem with an unknown kernel and there is no fast method to even compute the integral on the right hand side with known kernel, probe and images. In [20, 28], the authors considered L probes translated with weights $\{w_l\}$:

$$f_{pc,j} = \sum_{l=0}^{L-1} w_l |\mathcal{A}_j(\mathcal{T}_l\omega, u)|^2,\ \forall 0 \le j \le J-1, \tag{5}$$

with the translation operators $\{\mathcal{T}_l\}$, and partially coherent intensity f_{pc}. However, those methods in [20, 28] cannot be directly applied to (4), since either the weights

[28] or the probe [20] are assumed to be known in advance. Moreover, as coherence degrades, or the number of probes increases, their computation and memory costs would increase dramatically. Hence instead of solving (4) directly as in [20], in the following sections we will reformulate the above model (4), and solve the related nonlinear optimization problem.

A. Gradient Decomposition of the Probe (*GDP*)

Following (4) and using the Taylor expansion of ω, one has:

$$f_{pc,j}(q) = \int \left|\mathcal{F}_{x\to q}\left(\mathcal{S}_j u(x)\left(\omega(x) - y^T\nabla\omega(x) + \tfrac{1}{2}y^T\nabla^2\omega(x)y + \mathcal{O}(|y|^3)\right)\right)\right|^2 \kappa(y)\mathrm{d}y,$$

where we assume that the third order derivatives are uniformly bounded, e.g.

$$\max_{\{\alpha\in\mathbb{N}^2:\ |\alpha|=3\}} \max_{x\in\mathbb{C}^2} |\partial^\alpha\omega(x)| \le C, \tag{6}$$

with a positive constant C. It is easy to verify that this condition is satisfied when the illumination is generated by a small lens (with small aperture).

Consider a kernel function characterized by its moment expansion $m_{i_1,i_1} = \int \kappa(y)y_1^{i_1}y_2^{i_2}\mathrm{d}y$, normalized with $m_{0,0} = 1$, center of mass $(m_{1,0}, m_{0,1}) = (0,0)$, and second order moments σ_1^2, σ_2^2, and σ_{12} ($\sigma_1^2\sigma_2^2 - \sigma_{12}^2 \ge 0$). We further assume that:

$$m_{i_1,i_2} = 0, \text{if } \mathrm{mod}(i_1+i_2, 2) = 1. \tag{7}$$

The above relation holds if the kernel function is centrosymmetric with respect to the origin, i.e. $\kappa(-y_1,-y_2) = \kappa(y_1,y_2)$. For higher order moments, we also assume that

$$\int |y|^k\kappa(y) = \mathcal{O}(\max\{\sigma_1,\sigma_2\}^k),\ k\ge 3. \tag{8}$$

Therefore, one has:

$$\begin{aligned} f_{pc,j}(q) =& |\mathcal{A}_j(\omega + \tfrac{1}{2}(\sigma_1^2\nabla_{11}\omega + \sigma_2^2\nabla_{22}\omega + 2\sigma_{12}\nabla_{12}\omega), u)|^2 \\ &+ \sigma_1^2|\mathcal{A}_j(\nabla_1\omega, u)|^2 + \sigma_2^2|\mathcal{A}_j(\nabla_2\omega, u)|^2 \\ &+ \mathcal{O}(\int |y|^3\kappa(y)\mathrm{d}y). \end{aligned} \tag{9}$$

More details of the above derivation can be found in Appendix B. In order to further simplify the partial coherence approximation we introduce a new variable $\tilde{\omega}$ (variance adjusted probe):

$$\tilde{\omega} := \omega + \tfrac{1}{2}(\sigma_1^2\nabla_{11}\omega + \sigma_2^2\nabla_{22}\omega + 2\sigma_{12}\nabla_{12}\omega),$$

and nonlinear operator $\mathcal{G}_j : \mathbb{C}^{\bar{m}} \times \mathbb{C}^n \times \mathbb{R}^2 \to \mathbb{R}_+^{\bar{m}}$:

$$\begin{aligned} \mathcal{G}_j(\omega, u, \boldsymbol{\sigma}) :=& |\mathcal{A}_j(\omega, u)|^2 \\ &+ \sigma_1^2|\mathcal{A}_j(\nabla_1\omega, u)|^2 + \sigma_2^2|\mathcal{A}_j(\nabla_2\omega, u)|^2, \end{aligned} \tag{10}$$

where $\boldsymbol{\sigma} := (\sigma_1, \sigma_2)$. Neglecting high order terms $\mathcal{O}(\|\boldsymbol{\sigma}\|^k)$, $k \ge 3$, following (9), we obtain the approximate forward model:

$$\text{GDP}: \quad f_{pc,j} \approx \mathcal{G}_j(\tilde{\omega}, u, \boldsymbol{\sigma}). \tag{11}$$

a. Remarks:

- *The above assumptions hold if κ is a Gaussian kernel function, which well approximates standard light sources such as synchrotrons [6, 17, 22], SASE FELs [23], and others. When $\boldsymbol{\sigma} = (0,0)$, the above formula* (11) *reduces to the coherent model in* (1).
- *The variance adjusted probe $\tilde{\omega}$ incorporates second order effects (such as the covariance σ_{12}) into the illumination, while reducing the cost of computing the intensities using the forward model* (11) *compared to* (5).
- *If the probe ω has a support by a lens of finite size, $(\mathcal{F}\omega)(q) = 0\ \forall\ q \in \Omega_0 \subset \mathbb{R}^2$, then the same support applies to the variance adjusted probe $\forall\ q \in \Omega_0$, $(\mathcal{F}\tilde{\omega})(q) = 0$ due to the Fourier transform relationship:*

$$\begin{aligned} (\mathcal{F}\tilde{\omega})(q) &= \mathcal{F}((1 + \tfrac{1}{2}\sigma_1^2\nabla_{11} + \tfrac{1}{2}\sigma_2^2\nabla_{22} + \sigma_{12}\nabla_{12})\omega)(q) \\ &= \left(1 - \tfrac{1}{2}\sigma_1^2q_1^2 - \tfrac{1}{2}\sigma_2^2q_2^2 - \sigma_{12}q_1q_2\right)(\mathcal{F}\omega)(q), \\ &\forall\ q = (q_1, q_2) \in \mathbb{R}^2. \end{aligned}$$

- *Similarly to the decomposition model* (3), GDP *has three different modes $\{\tilde{\omega}, \nabla_1\tilde{\omega}, \nabla_2\tilde{\omega}\}$. The difference between them is that in the* GDP *model, two modes $\nabla_l\omega$ can be expressed by the first mode $\tilde{\omega}$ explicitly, while the multiple modes in the decomposition model* (3) *are only assumed to be orthogonal to each other [27, 30]. Such orthogonal constraint is essentially nonconvex, which is more difficult to handle and leads to more local minima as well.*

III. FAST ITERATIVE ALGORITHM: *GDP-ADMM*

The amplitude based nonlinear optimization model can be established as:

$$\min_{\tilde{\omega},u,\boldsymbol{\sigma}} \tfrac{1}{2}\sum_{j=0}^{J-1} \left\|\sqrt{f_{pc,j}} - \sqrt{\mathcal{G}_j\left(\tilde{\omega}, u, \boldsymbol{\sigma}\right)}\right\|^2, \tag{12}$$

where $\|\cdot\|$ denotes the standard L^2 norm in Euclidean space. We remark that by introducing the new variable $\tilde{\omega}$, it is much easier to solve the subproblems of the nonlinear optimization model (12) with respect to $\boldsymbol{\sigma}$ than using the original formula (9).

The *GDP* based nonlinear optimization model (12) is nonconvex and non-differentiable. We are interested in designing a fast first order algorithm, whose subproblems can be easily implemented. The Alternating Direction Method of Multipliers (ADMM) [12] has been successfully applied to large scale nonlinear and non-differential optimization problems arising from machine learning or computer vision, among other areas. The connection between ADMM, the Douglas-Rachford algorithm, and the popular Hybrid Input Output [9] for classical phase retrieval was discussed in [1, 29]. By introducing auxiliary variables $z_{l,j} \in \mathbb{C}^{\tilde{m}}$, one has:

$$\min_{\tilde{\omega},u,\boldsymbol{\sigma},z_{l,j}} \frac{1}{2}\sum_{j=0}^{J}\left\|\sqrt{f_{pc,j}}-\sqrt{\sum_{l=0}^{2}|z_{l,j}|^2}\right\|^2, \tag{13}$$
$$s.t. \quad z_{0,j}=\mathcal{A}_j(\tilde{\omega},u), z_{1,j}=\sigma_1\mathcal{A}_j(\nabla_1\tilde{\omega},u), z_{2,j}=\sigma_2\mathcal{A}_j(\nabla_2\tilde{\omega},u),\ 0\le j\le J-1.$$

The corresponding augmented Lagrangian reads as:

$$\begin{aligned}\mathcal{L}_r(\tilde{\omega},u,\{z_l\}_{l=0}^2,\boldsymbol{\sigma},\{\Lambda_l\}_{l=0}^2) &:= \frac{1}{2}\sum_j\left\|\sqrt{f_{pc,j}}-\sqrt{\sum_l|z_{l,j}|^2}\right\|^2\\ &+r\Re(\langle z_0-\mathcal{A}(\tilde{\omega},u),\Lambda_0\rangle)+\frac{r}{2}\|z_0-\mathcal{A}(\tilde{\omega},u)\|^2\\ &+r\Re(\langle z_1-\mathcal{A}(\sigma_1\nabla_1\tilde{\omega},u),\Lambda_1\rangle)+\frac{r}{2}\|z_1-\mathcal{A}(\sigma_1\nabla_1\tilde{\omega},u)\|^2\\ &+r\Re(\langle z_2-\mathcal{A}(\sigma_2\nabla_2\tilde{\omega},u),\Lambda_2\rangle)+\frac{r}{2}\|z_2-\mathcal{A}(\sigma_2\nabla_2\tilde{\omega},u)\|^2,\end{aligned} \tag{14}$$

with multipliers $\Lambda_l = (\Lambda_{l,0}^T,\cdots,\Lambda_{l,J-1}^T)^T \in \mathbb{C}^m$, $z_l = (z_{l,0}^T,\cdots,z_{l,J-1}^T)^T \in \mathbb{C}^m$, and the positive parameter r, where $\langle\cdot,\cdot\rangle$ denotes the standard inner product in complex Euclidean space, and $\Re(\cdot)$ denotes the real part of a complex number. Briefly, the *GDP-ADMM* algorithm alternates minimization of the above augmented Lagrangian with respect to the variables $\omega, u, \sigma_1, \sigma_2$, updates the multipliers $\Lambda_l, l = 0, 1, 2$, and repeats. In our previous work, ADMM was applied to coherent ptychographic imaging in [4, 29]. Here we propose a new, more robust variant of ADMM. It employs an additional proximal term added to the augmented Lagrangian: $\frac{\tau}{2}\|u - u^k\|^2_{M^k}$ to avoid division by zeros when minimizing with respect to u. The detailed description of *GDP-ADMM* can be found in Algorithm 1 (further details in Appendix A). The numerical results can be found in the following section. For simplicity, the gradient operator ∇_1, ∇_2 in the numerical section is considered in a discrete setting, using the forward and finite difference operator with respect to x, y directions.

Theoretical analysis about the convergence properties of this blind algorithm (to refine the probe, the image, and the coherence function variances during iterations) is likely to be very challenging. However, if the variances are known or fixed, convergence to the stationary point of the nonlinear optimization model (12) can be guaranteed by assuming that the iterative sequence is bounded and the parameter r is sufficiently large [4].

Algorithm 1: *GDP-ADMM*

Initialization: Set $\tilde{\omega}^0 = \mathcal{F}^*(\frac{1}{J}\sum\sqrt{f_{pc,j}})$, $u^0 = 1, z_0^0 = \mathcal{A}(\tilde{\omega}^0,u^0)$, $\sigma_1^0 = \sigma_2^0 = 1$, $z_0^0 = \mathcal{A}(\tilde{\omega}^0,u^0)$, $z_1^0 = \sigma_1^0\mathcal{A}(\nabla_1\tilde{\omega}^0,u^0)$, $z_2^0 = \sigma_2^0\mathcal{A}(\nabla_2\tilde{\omega}^0,u^0)$, $\Lambda_l^0 = 0, k := 0$, maximum iteration number $Iter_{Max}$, parameters r and τ.

output:x $u^\star := u^{Iter_{Max}}$ and $\tilde{\omega}^\star := \tilde{\omega}^{Iter_{Max}}$.

for $k = 0$ to $Iter_{Max} - 1$

1. Compute $\tilde{\omega}^{k+1}$ by solving the linear system:
$$\tilde{\omega}^{k+1} = \left(N_1^k\right)^{-1}\times c^k, \tag{15}$$
using conjugate gradient method, where N_1^k, and c^k are given in Eq. (A3) (Appendix A).
2. Compute u^{k+1} as
$$u^{k+1} = \left(N_2^k + \frac{\tau}{r}M^k\right)^{-1}\times\left(b^k + \frac{\tau}{r}M^k u^k\right). \tag{16}$$
with diagonal matrices N_2^k, M^k and vector b^k defined in Eq. (A5), and (A6) (see Appendix A).
3. Compute z_l^{k+1} as:
$$z_{l,j}^{k+1} = \frac{\sqrt{f_{pc,j}}+r\sqrt{\sum_{d=0}^2|\tilde{z}_{d,j}^{k+1}|^2}}{1+r}\circ\operatorname{sign}(\tilde{z}_{l,j}^{k+1}), \tag{17}$$
with
$$\begin{aligned}\operatorname{sign}(z_{l,j}) &:= \frac{z_{l,j}}{\sqrt{\sum_{d=0}^2|z_{d,j}|^2}},\\ \tilde{z}_{0,j}^{k+1} &= \mathcal{A}(\tilde{\omega}^{k+1},u^{k+1})-\Lambda_0^k,\\ \tilde{z}_{1,j}^{k+1} &= \sigma_1^k\mathcal{A}(\nabla_1\tilde{\omega}^{k+1},u^{k+1})-\Lambda_1^k,\\ \tilde{z}_{2,j}^{k+1} &= \sigma_2^k\mathcal{A}(\nabla_2\tilde{\omega}^{k+1},u^{k+1})-\Lambda_2^{k+1}.\end{aligned}$$
4. Compute σ_l^{k+1} as:
$$\sigma_l^{k+1} = \frac{\Re(\langle z_l^{k+1}+\Lambda_l^k,\mathcal{A}(\nabla_l\tilde{\omega}^{k+1},u^{k+1})\rangle)}{\|\mathcal{A}(\nabla_l\tilde{\omega}^{k+1},u^{k+1})\|^2},\ l=1,2.$$
5. Update the multipliers $\{\Lambda_l^{k+1}\}$ as
$$\begin{aligned}\Lambda_0^{k+1} &= \Lambda_0^k + (z_0^{k+1}-\mathcal{A}(\tilde{\omega}^{k+1},u^{k+1}));\\ \Lambda_1^{k+1} &= \Lambda_1^k + (z_1^{k+1}-\mathcal{A}(\sigma_1^{k+1}\nabla_1\tilde{\omega}^{k+1},u^{k+1}));\\ \Lambda_2^{k+1} &= \Lambda_2^k + (z_2^{k+1}-\mathcal{A}(\sigma_2^{k+1}\nabla_2\tilde{\omega}^{k+1},u^{k+1})).\end{aligned}$$

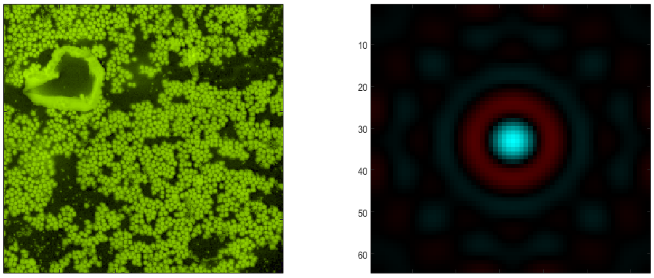


FIG. 2. True image "Goldballs" [19] 249×249 (left) and true probe (right) 64×64, with "beam diameter" $D=7$ pixels (FWHM).

IV. NUMERICAL EXPERIMENTS

The experimental setup of this section is introduced next. The reference specimen used is a complex valued image "Goldballs" [19], of size 249×249 pixels (Figure 2). The probe is an Airy disk of size 64×64 pixels (Figure 2).

We compare the proposed *GDP-ADMM* (Appendix A) with the "fully-coherent" model by ADMM (*FC-ADMM* [4]). Both algorithms are executed with a maximum iteration number of 300. We measure the quality of recovery images by relative residuals as:

$$e_{fc} = \frac{\left\|\sqrt{f} - |\mathcal{A}(\omega^k, u^k)|\right\|_1}{\left\|\sqrt{f}\right\|_1},$$

$$e_{pc} = \frac{\left\|\sqrt{f} - \sqrt{\mathcal{G}(\tilde{\omega}^k, u^k, \boldsymbol{\sigma}^k)}\right\|_1}{\left\|\sqrt{f}\right\|_1},$$

for *FC-ADMM* and *GDP-ADMM* respectively, where $\omega^k, \tilde{\omega}^k, u^k, \boldsymbol{\sigma}^k$ are the iterative solutions, and f is the measured intensity. The signal-to-noise ratio (SNR) is also measured:

$$\mathrm{SNR}(u^k, u_{true}) = -20 \min_{c\in\mathbb{C}} \log \frac{\|u^k - c\, u_{true}\|}{\|u^k\|},$$

where u_{true} is the ground truth of the image.

We show the performance of *GDP-ADMM* with the following Gaussian kernel function:

$$\kappa(y) := \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{y_1^2}{2\sigma_1^2} - \frac{y_2^2}{2\sigma_2^2}\right). \tag{18}$$

The discrete truncated Gaussian matrix is generated by *fspecial* in MATLAB with variances σ_1, σ_2 and support sizes $(4\sigma_1+1, 4\sigma_2+1)$. We remark that the kernel width is $2\sqrt{2\log(2)}\times(\sigma_1,\sigma_2) \approx 2.35\times(\sigma_1,\sigma_2)$.

Simulated data is generated using raster grid scanning with sliding distance Dist $=8$ pixels (slightly bigger than the beam width) as a default, and we incorporate the support constraint of the lens as in [19]. The parameter r is selected manually with default value 0.2, and $\tau = 1.01r$.

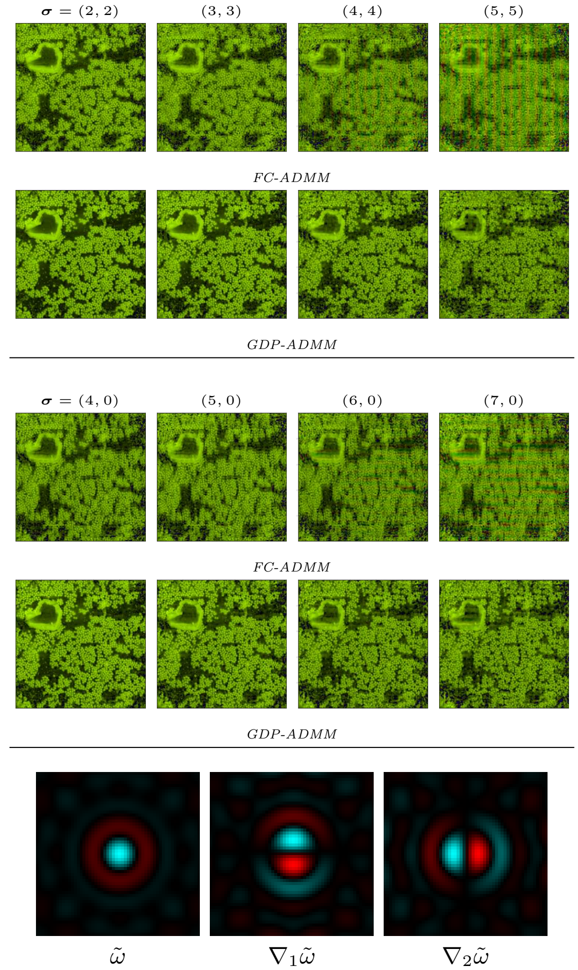


FIG. 3. (First and third rows) reconstructed images of *FC-ADMM*, and (second and fourth rows) *GDP-ADMM* while varying $\boldsymbol{\sigma}$. (Bottom row) Probe with its three-modes $\tilde{\omega}, \nabla_1\tilde{\omega}, \nabla_2\tilde{\omega}$ from left to right recovered from the second row with $\boldsymbol{\sigma}=(4,4)$.

a. Noiseless data. Following (4), the partially coherent intensity in a discrete setting is generated as

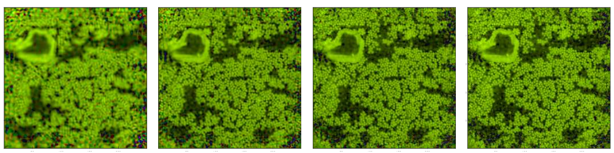
$$f_{pc,j} = \sum_i \kappa_i \left|\mathcal{F}((\mathcal{T}_i\omega)\circ\mathcal{S}_j u)\right|^2, \tag{19}$$

with translation operator $\mathcal{T}_i$, discrete Gaussian weights $\{\kappa_i\}$, and periodical boundary condition for the probe. We conduct the first numerical experiment to explore the performance of the proposed algorithm with respect to different degrees of partial coherence by varying the variances σ_1, σ_2, while keeping the beam width constant (the smaller the variances, the more coherent the data).

The reconstructed images can be seen in Figure 3, and the accuracy of the reconstructed results can be seen in Table I. Figure 3 shows significant improvements when using *GDP-ADMM* compared to *FC-ADMM*. When coherence is very low (4th column), the visual quality of *FC-ADMM* drops, and small-scale features are com-

TABLE I. Performance of *GDP-ADMM* and *FC-ADMM*. e_{pc} and e_{fc} are the residuals of *GDP-ADMM* and *FC-ADMM* respectively (the smaller, the better); SNR_{pc}, SNR_{fc} are the SNRs for *GDP-ADMM* and *FC-ADMM*, respectively (the larger, the better).

$\boldsymbol{\sigma}$	(2,2)	(3,3)	(4,4)	(5,5)
e_{fc}	1.36E-1	1.88E-1	2.11E-1	2.20E-1
e_{pc}	2.60E-2	5.68E-2	8.95E-2	1.14E-1
SNR_{fc}	14.78	9.53	6.14	4.37
SNR_{pc}	23.73	18.83	13.02	8.37
$\boldsymbol{\sigma}$	(4,0)	(5,0)	(6,0)	(7,0)
e_{fc}	1.70E-1	1.89E-1	2.04E-1	2.14E-1
e_{pc}	5.62E-2	7.87E-2	9.76E-2	1.12E-1
SNR_{fc}	11.47	9.42	7.61	5.64
SNR_{pc}	19.83	16.39	13.46	10.94



(a) SNR=7.08 (b)10.17 (c) 11.30 (d) 12.35

FIG. 4. Reconstructed images of *GDP-ADMM* while varying the level of Poisson noise, and the SNR with respect to the ground truth. From left to right: Noise level $\gamma = 0.25, 1, 4, 16$, (the larger γ, the weaker the noise). The corresponding SNRs of the data, defined by $\mathrm{SNR}(f^{noise}, f)$, are 23.89, 29.87, 35.91, and 41.07.

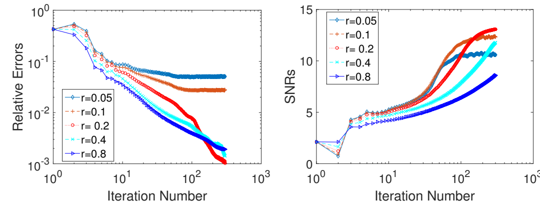
pletely lost. In comparison, with the same configuration, *GDP-ADMM* can still produce images containing significantly sharper large- and small-scale features. However the results by *GDP-ADMM* seem blurry when the kernel variance $\|\boldsymbol{\sigma}\|$ becomes too large, as can be seen from the fourth column of Figure (3). Table (I) shows the enhanced accuracy of GDP-ADMM compared to the coherent model: residuals are at least 50% percent smaller, and images with SNRs about twice as high. We also include the results of the recovered probe with its modes in Figure 3.

b. Noisy data. We consider Poisson noisy measurements as

$$f_{pc,j}^{\text{noise}} = \tfrac{1}{\gamma}\mathrm{Poisson}(\gamma f_{pc,j}),$$

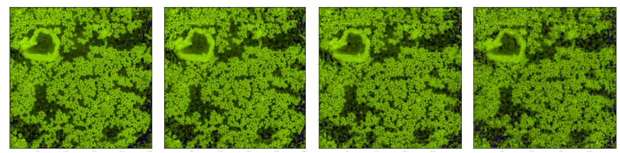
where $\gamma > 0$ is used to control the noise level, and $f_{pc,j}$ is the clean partial coherent intensity data defined by (19). Figure 4 shows the recovered images at different noise levels. It can be seen that, as the noise level decreases (increasing γ), the quality of the results increases. We remark that when $\gamma = 0.25$, the image in Figure 4 (a) is very blurry, which implies that it is more challenging to recover high quality images from partially coherent data contaminated by strong noise.

c. Parameter r and sliding distances. The following experiment is conducted to study the influence of parameter r by varying $r \in \{0.05, 0.1, 0.2, 0.4, 0.8\}$. See the



(a) (b)

FIG. 5. Performance influenced by parameter $r \in \{0.05, 0.1, 0.2, 0.4, 0.8\}$ of *GDP-ADMM*. (a) Relative errors $\frac{\|u^k - u^{k-1}\|}{\|u^k\|}$ of the iterative solution with respect to the iteration number; (b) SNRs of iterative solutions with respect to the iteration number.



Dist=4	6	8	12
SNR=15.93	14.54	13.02	8.96

FIG. 6. Reconstructed images using *GDP-ADMM* with different sliding distances Dist, Gaussian kernel function ($\boldsymbol{\sigma} = (4,4)$), and beam width $D = 7$.

convergence curve in Figure 5, where the change histories of the relative errors and SNRs with respect to iteration numbers are reported. One can readily see that with smaller parameter r, the algorithm tends to be unstable (see Figure 5 (b)). It is consistent with the condition for convergence guarantee in [4], where the parameter r should be sufficiently large to make the augmented Lagrangian monotonically decrease. If the parameter is too big, the iterative solution could be trapped into the unsatisfactory local minima. Therefore, to gain optimal performance, a moderate r should be used. We remark that although the parameter r is selected manually, it can be applied to diverse cases with different degrees of partial coherence and sliding distances when fixing the kernel function.

In Figure 6 we conduct the tests with different sliding distances Dist $\in \{4, 6, 8, 12\}$, which determine the redundancy levels of the data. On one hand, inferred from the reported results, scanning with a smaller step size or Dist will help to increase the quality of the images. On the other hand, when Dist $= 12$ (almost twice as large as the beam width), *GDP-ADMM* can still produce satisfactory results with clear small-scale features, showing the robustness of *GDP-ADMM* with respect to the redundancy of the measured data.

d. Proximal term. In the following experiment we show the effect of the proximal term. We conduct tests comparing *GDP-ADMM* baseline with a variant that replaces the proximal term with a penalization term $\frac{\tau}{2}\|u\|^2$,

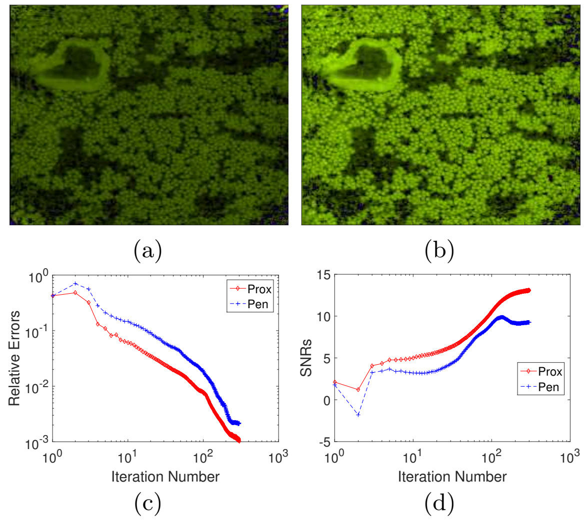


FIG. 7. GDP-ADMM variant with penalization ("Pen", (a)), versus *GDP-ADMM* baseline ("Prox", (b)). The relative errors and SNRs are reported in (c) and (d), respectively. Dist = 8, $\boldsymbol{\sigma} = (4, 4)$, beam width $D = 7$.

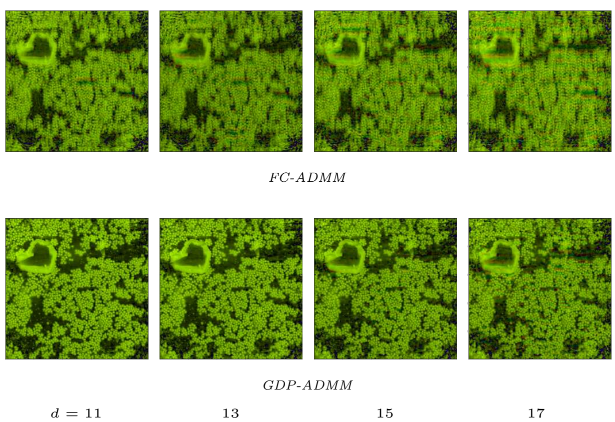


FIG. 8. Reconstructed images by *FC-ADMM* (top row) and *GDP-ADMM* (bottom) for *fly-scan* ptychography with varying kernel width d, beam width $D = 7$, sliding distance Dist = 8.

i.e.

$$u^{k+1} = (N_2^k + \tfrac{\tau}{r}\mathbf{I})^{-1} b^k,$$

with identity matrix $\mathbf{I}$ and $\tau = 1\times10^{-5}\times\|\mathrm{diag}(N_2^k)\|_\infty\times r$ when solving Eq. (A5), Appendix A. See the results in Figure 7. One can readily observe that with the help of the proximal term, the recovered images have cleaner boundaries (Figure 7 (b)) and higher SNR values (Figure 7 (d)). Moreover, *GDP-ADMM* is speeded up when using the proximal term, as shown by the convergence of the relative errors in Figure 7 (c).

e. Different kernels. Previous experiments reported in this section are based on the Gaussian kernel function. Here we assess the performance of *GDP-ADMM* when using a binary kernel function for *fly-scan* ptychography [7]. Figure 8 shows the reconstructed images obtained by varying the kernel size d. Similarly to the case of Gaussian kernel function, *GDP-ADMM* displays a significant increase in visual quality. In the first row of Figure 8, the results produced by *FC-ADMM* are completely blurry. On the other hand, *GDP-ADMM* (second row of Figure 8) achieves much sharper overall results. One can also see the obvious decrease of the residual and increase of SNRs by *GDP-ADMM* in Table II.

TABLE II. *Fly-scan*: Performance of *GDP-ADMM* and *FC-ADMM*. Here d is the size of the binary kernel, and the beam width is $D = 7$ pixels.

d	11	13	15	17
e_{fc}	2.00E-1	2.16E-1	2.29E-1	2.37E-1
e_{pc}	4.47E-2	6.46E-2	9.04E-2	1.16E-1
SNR_{fc}	6.19	5.60	7.15	6.71
SNR_{pc}	21.15	18.07	14.63	11.58

Similar improvements by *GDP-ADMM* can be obtained with other motion blur type kernel functions, however, due to page limitations, we do not provide further results. These results show that *GDP-ADMM* can be applied to partial coherence problems with more general kernel function.

f. Runtime and memory performance Finally we report the the computational performance of the *GDP-ADMM* algorithm on a machine with an Intel i7-5600U CPU and 16G RAM using MATLAB. *GDP-ADMM* requires two additional modes and four additional variables z_1, z_2 and Λ_1, Λ_2 compared with *FC-ADMM* . Because of this, the memory cost and runtime of *GDP-ADMM* are in theory about three times as large as those of *FC-ADMM* per iteration. When computing the image in Figure 2, *GDP-ADMM* requires an average of 873MB of RAM and takes 655 seconds to compute, whereas *FC-ADMM* requires 344MB ad takes 218 seconds. These results are consistent with the previous theoretical estimate.

We further investigate the change histories of the SNRs for *FC-ADMM* and *GDP-ADMM* with respect to elapsed the time, and report the results in Figure 9. The SNR histories show that, *FC-ADMM* can recover better images in the first 50 seconds, but *GDP-ADMM* improves the image further after that. Hence, in order to accelerate the *GDP-ADMM* algorithm, we could use the iterative solution of *FC-ADMM* as the initialization for *GDP-ADMM*. We also emphasize that the runtime and memory requirements *GDP-ADMM* are insensitive to the variances and the support sizes of the kernel functions. It is important to note that, if we solve the problem directly following [20], the probe and the weights in (5) in the case of the setting in Figure 3 (a)-(b), at most $(4\sigma_1 + 1) \times (4\sigma_1 + 1) = 441$ translated probes for $\sigma_1 = \sigma_2 = 5$ should be introduced, which requires much more memory and computation costs.

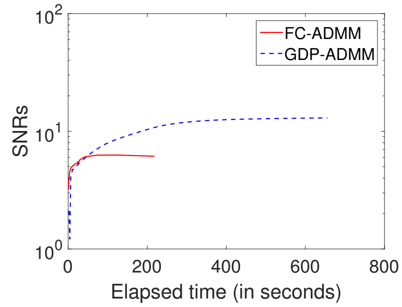


FIG. 9. Histories of the SNRs changes for *FC-ADMM* and *GDP-ADMM* with respect to the elapsed time in seconds. $\sigma_1 = \sigma_2 = 4$, Dist=8.

V. CONCLUSIONS

In this paper, we propose Gradient Decomposition of the Probe (*GDP*), an efficient model that exploits translational kernel separability. The *GDP* model increases the approximation accuracy compared with the coherent model, and it holds for a general partial coherent source. We derive an optimization model coupling the variances of the kernel with the transverse coherence widths, and a fast and memory efficient proximal first order *GDP-ADMM* algorithm to solve the nonlinear optimization problem based on *GDP*. Numerical experiments demonstrate the effectiveness of such approximation in the case of Gaussian kernel function, and binary functions in fly-scanning schemes, showing that the proposed methods are suitable for general partially coherent sources.

However the results by *GDP-ADMM* seem blurry when inferred from the fourth column of Figure 3, and further improvement may be achieved by considering sparse prior, or using piecewise Taylor expansion series or other integration schemes. This will be subject of future work. Additional research directions to extend this work are: (i) consider partial coherence with framewise different variances $\boldsymbol{\sigma}_j$ when using scan geometries on irregular grids or spirals, (ii) combine translational blur at the sample with detector blur in Eq. (2) for near field ptychography. (ii) incorporate longitudinal coherence when using broadband illumination with chromatic aberrations of the probe, dispersion of the diffraction pattern [8], and spectral Kramers-Kronig dispersion across resonances [14].

VI. ACKNOWLEDGMENTS

This work was 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, and by the Advanced Light Source, which is a DOE Office of Science User Facility under contract no. DE-AC02-05CH11231.

[1] H. H. Bauschke, P. L. Combettes, and D. R. Luke, *Hybrid projection reflection method for phase retrieval*, J. Opt. Soc. Amer. A, 20 (2003), pp. 1025–1034.
[2] N. Burdet, X. Shi, D. Parks, J. N. Clark, X. Huang, S. D. Kevan, and I. K. Robinson, *Evaluation of partial coherence correction in x-ray ptychography*, Optics express, 23 (2015), pp. 5452–5467.
[3] S. Cao, P. Kok, P. Li, A. M. Maiden, and J. M. Rodenburg, *Modal decomposition of a propagating matter wave via electron ptychography*, Phys. Rev. A, 94 (2016), p. 063621.
[4] H. Chang and S. Marchesini, *Blind ptychographic phase retrieval by globally convergent ADMM*, submitted, (2017).
[5] J. N. Clark and A. G. Peele, *Simultaneous sample and spatial coherence characterisation using diffractive imaging*, Applied Physics Letters, 99 (2011), p. 154103.
[6] R. Coisson and S. Marchesini, *Gauss–schell sources as models for synchrotron radiation*, Journal of synchrotron radiation, 4 (1997), pp. 263–266.
[7] J. Deng, Y. S. Nashed, S. Chen, N. W. Phillips, T. Peterka, R. Ross, S. Vogt, C. Jacobsen, and D. J. Vine, *Continuous motion scan ptychography: characterization for increased speed in coherent x-ray imaging*, Optics express, 23 (2015), pp. 5438–5451.
[8] B. Enders, M. Dierolf, P. Cloetens, M. Stockmar, F. Pfeiffer, and P. Thibault, *Ptychography with broad-bandwidth radiation*, Applied Physics Letters, 104 (2014), p. 171104.
[9] J. R. Fienup, *Phase retrieval algorithms: a comparison*, Appl. Opt., 21 (1982), pp. 2758–2769.
[10] J. R. Fienup, J. C. Marron, T. J. Schulz, and J. H. Seldin, *Hubble space telescope characterized by using phase-retrieval algorithms*, Appl. Opt., 32 (1993), pp. 1747–1767.
[11] 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 (2010), pp. 529–534.
[12] R. Glowinski and P. L. Tallec, *Augmented Lagrangian and operator-splitting methods in nonlinear mechanics*, SIAM Studies in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1989.
[13] F. Gori and C. Palma, *Partially coherent sources which give rise to highly directional light beams*, Optics Communications, 27 (1978), pp. 185 – 188.
[14] M. Hirose, K. Shimomura, N. Burdet, and Y. Takahashi, *Use of kramers–kronig relation in phase retrieval calculation in x-ray spectro-ptychography*, Optics Ex-

press, 25 (2017), pp. 8593–8603.

[15] 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 (2017), pp. 402–406.

[16] X. Huang, K. Lauer, J. N. Clark, W. Xu, E. Nazaretski, R. Harder, I. K. Robinson, and Y. S. Chu, *Fly-scan ptychography*, Scientific reports, 5 (2015).

[17] K.-J. Kim, *Brightness, coherence and propagation characteristics of synchrotron radiation*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 246 (1986), pp. 71–76.

[18] A. M. Maiden and J. M. Rodenburg, *An improved ptychographical phase retrieval algorithm for diffractive imaging*, Ultramicroscopy, 109 (2009), pp. 1256–1262.

[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 (2016), pp. 1245–1252.

[20] S. Marchesini, A. Schirotzek, C. Yang, H.-t. Wu, and F. Maia, *Augmented projections for ptychographic imaging*, Inverse Problems, 29 (2013), p. 115009.

[21] P. M. Pelz, M. Guizar-Sicairos, P. Thibault, I. Johnson, M. Holler, and A. Menzel, *On-the-fly scans for x-ray ptychography*, Applied Physics Letters, 105 (2014), p. 251101.

[22] R. P. W. R. Coisson, *Phase space distribution of brilliance of undulator sources*, Proc.SPIE, 0582 (1986), pp. 0582 – 0582 – 8.

[23] E. Saldin, E. Schneidmiller, and M. Yurkov, *Coherence properties of the radiation from x-ray free electron laser*, Optics Communications, 281 (2008), pp. 1179–1188.

[24] D. A. Shapiro, Y.-S. Yu, T. Tyliszczak, J. Cabana, R. Celestre, W. Chao, K. Kaznatcheev, A. D. Kilcoyne, F. Maia, S. Marchesini, et al., *Chemical composition mapping with nanometre resolution by soft x-ray microscopy*, Nature Photonics, 8 (2014), pp. 765–769.

[25] X. Shi, P. Fischer, V. Neu, D. Elefant, J. Lee, D. Shapiro, M. Farmand, T. Tyliszczak, H.-W. Shiu, S. Marchesini, et al., *Soft x-ray ptychography studies of nanoscale magnetic and structural correlations in thin smco5 films*, Applied Physics Letters, 108 (2016), p. 094103.

[26] P. Thibault, M. Dierolf, O. Bunk, A. Menzel, and F. Pfeiffer, *Probe retrieval in ptychographic coherent diffractive imaging*, Ultramicroscopy, 109 (2009), pp. 338–343.

[27] P. Thibault and A. Menzel, *Reconstructing state mixtures from diffraction measurements*, Nature, 494 (2013), p. 68.

[28] L. Tian, X. Li, K. Ramchandran, and L. Waller, *Multiplexed coded illumination for fourier ptychography with an led array microscope*, Biomedical optics express, 5 (2014), pp. 2376–2389.

[29] Z. Wen, C. Yang, X. Liu, and S. Marchesini, *Alternating direction methods for classical and ptychographic phase retrieval*, Inverse Probl., 28 (2012), p. 115010.

[30] E. Wolf, *New theory of partial coherence in the space–frequency domain. part i: spectra and cross spectra of steady-state sources*, JOSA, 72 (1982), pp. 343–351.

Appendix A: *GDP-ADMM*

Once the variables $\omega, u, z, \Lambda, \sigma$ are set, in the $(k+1)^{th}$ iteration the augmented Lagrangian is modified by adding a proximal term as follows:

$$\tilde{\mathcal{L}}_r^k(\tilde{\omega}, u, z, \boldsymbol{\sigma}, \Lambda) = \mathcal{L}_r(\tilde{\omega}, u, z, \boldsymbol{\sigma}, \Lambda) + \tfrac{\tau}{2}\|u - u^k\|^2_{M^k} ,$$

where for simplicity we removed superscripts and subscripts, and $\|u\|^2_{M^k} = \langle M^k u, u\rangle$ is a positive semi-definite matrix M^k, defined by the approximated solution $\left(\tilde{\omega}^{k+1}, u^k, \{z_l^k\}, \{\sigma_l^k\}, \{\Lambda_l^k\}\right)$ in the k^{th} iteration. We remark that with the help of an additional proximal term $\frac{\tau}{2}\|u - u^k\|^2_{M^k}$, the subproblem in Step 2 admits a unique solution. Below we demonstrate how to solve each subproblem.

$$\begin{aligned}
&\text{Step 1: } \tilde{\omega}^{k+1} = \arg\min_{\tilde{\omega}} \mathcal{L}_r(\tilde{\omega}, u^k, \{z_l^k\}, \boldsymbol{\sigma}^k, \{\Lambda_l^k\});\\
&\text{Step 2: } u^{k+1} = \arg\min_{u} \tilde{\mathcal{L}}_r^k\left(\tilde{\omega}^{k+1}, u, \{z_l^k\}, \boldsymbol{\sigma}^k, \{\Lambda_l^k\}\right);\\
&\text{Step 4: } \{z_l^{k+1}\}_{l=0}^2 = \arg\min_{\{z_l\}} \mathcal{L}_r(\tilde{\omega}^{k+1}, u^{k+1}, \{z_l\}, \boldsymbol{\sigma}^k, \{\Lambda_l^k\});\\
&\text{Step 5: } \boldsymbol{\sigma}^{k+1} = \arg\min_{\boldsymbol{\sigma}} \mathcal{L}_r(\tilde{\omega}^{k+1}, u^{k+1}, \{z_l^{k+1}\}, \boldsymbol{\sigma}, \{\Lambda_l^k\});\\
&\text{Step 6: } \Lambda_0^{k+1} = \Lambda_0^k + (z_0^{k+1} - \mathcal{A}(\tilde{\omega}^{k+1}, u^{k+1}));\\
&\qquad\quad \Lambda_1^{k+1} = \Lambda_1^k + (z_1^{k+1} - \mathcal{A}(\sigma_1^{k+1}\nabla_1\tilde{\omega}^{k+1}, u^{k+1}));\\
&\qquad\quad \Lambda_2^{k+1} = \Lambda_2^k + (z_2^{k+1} - \mathcal{A}(\sigma_2^{k+1}\nabla_2\tilde{\omega}^{k+1}, u^{k+1}));
\end{aligned} \tag{A1}$$

Step 1 involves finding the first order stationary point ω^{k+1} of the augmented Lagrangian which has the following form:

$$N_1^k \tilde{\omega}^{k+1} - c^k = 0, \tag{A2}$$

with symmetric sparse matrix N_1^k defined as:

$$\begin{aligned}
N_1^k :=& \operatorname{diag}\left(\sum_j \left|\mathcal{S}_j u^k\right|^2\right)\\
&+ \sum_{l=1}^2 (\sigma_l^k)^2 \nabla_l^T \left(\operatorname{diag}\left(\sum_j |\mathcal{S}_j u^k|^2\right)\nabla_l\right),\\
c^k :=& \sum_j \left(\mathcal{S}_j (u^k)^* \circ \bar{z}_{0,j}^k + \sum_{l=1}^2 \sigma_l^k \nabla_l^T \left(\mathcal{S}_j (u^k)^* \circ \bar{z}_{l,j}^k\right)\right),
\end{aligned} \tag{A3}$$

and $\bar{z}_{l,j}^k := \mathcal{F}^*(z_{l,j}^k + \Lambda_{l,j}^k)$. In this section, the gradient operator is given in a discrete setting for simplicity, where ∇_1, ∇_2 represent the forward finite difference operator with respect to x, y directions.

Step 2 is similar to Step 1, the closed form solution can be expressed as:

$$\left(N_2^k + \tfrac{\tau}{r} M^k\right) u^{k+1} - \left(b^k + \tfrac{\tau}{r} M^k u^k\right) = 0, \tag{A4}$$

with diagonal matrix N_2^k:

$$
\begin{aligned}
N_2^k :=&\mathrm{diag}\left(\sum_j \mathcal{S}_j^T\left(|\tilde{\omega}^{k+1}|^2+\sum_{l=1}^{2}\sigma_l^2|\nabla_l\tilde{\omega}^{k+1}|^2\right)\right),\\
b^k :=&\sum\nolimits_j \mathcal{S}_j^T\left((\tilde{\omega}^{k+1})^*\circ\tilde{z}_{0,j}^k+\sum_{l=1}^{2}\sigma_l^k\nabla_l(\tilde{\omega}^{k+1})^*\circ\tilde{z}_{l,j}^k\right).
\end{aligned}
\tag{A5}
$$

Empirically, in order to avoid divisions by 0 when solving (A4), set the matrix M^k to a diagonal matrix with diagonal elements:

$$
M^k(l,l) :=\begin{cases}0, & \text{if } N_2^k(l,l)\geq \frac{1}{10}\|\mathrm{diag}(N_2^k)\|_\infty;\\ \|\mathrm{diag}(N_2^k)\|_\infty-\frac{r}{\tau}N_2^k(l,l), & \text{otherwise.}\end{cases}
\tag{A6}
$$

Step 3 can be expressed as the following problem:

$$
\begin{aligned}
\min_{\{z_{l,j}\}_{0\leq l\leq 2}^{0\leq j\leq J-1}} &\sum\nolimits_j \tfrac{1}{2}\left\|\sqrt{f_{pc,j}}-\sqrt{\sum\nolimits_l |z_{l,j}|^2}\right\|^2\\
&+\tfrac{r}{2}(\|z_0+\Lambda_0-\mathcal{A}(\tilde{\omega},u)\|^2\\
&+\|z_1+\Lambda_1-\mathcal{A}(\sigma_1\nabla_1\tilde{\omega},u)\|^2\\
&+\|z_2+\Lambda_2-\mathcal{A}(\sigma_2\nabla_2\tilde{\omega},u)\|^2),
\end{aligned}
$$

which can be solved independently with respect to each frame, hence one needs to consider:

$$
\begin{aligned}
\min_{\{z_{l,j}\}_{l=0}^{2}} &\tfrac{1}{2}\left\|\sqrt{f_{pc,j}}-\sqrt{\sum\nolimits_l |z_{l,j}|^2}\right\|^2\\
&+\tfrac{r}{2}\sum\nolimits_l \|z_{l,j}-\tilde{z}_{l,j}\|^2 \;\forall 0\leq j\leq J-1,
\end{aligned}
$$

where $\tilde{z}_{0,j}=\mathcal{A}_j(\tilde{\omega},u)-\Lambda_{0,j}, \tilde{z}_{1,j}=\sigma_1\mathcal{A}_j(\nabla_1\tilde{\omega},u)-\Lambda_{1,j}, \tilde{z}_{2,j}=\sigma_2\mathcal{A}_j(\nabla_2\tilde{\omega},u)-\Lambda_{2,j}$. Similarly, it also has a closed form solution, see Eq. (17), Section III.

Step 4 requires solving a linear least square problem whose closed form solution is:

$$
\sigma_l=\frac{\Re(\langle z_l+\Lambda_l,\mathcal{A}(\nabla_l\tilde{\omega},u)\rangle)}{\|\mathcal{A}(\nabla_l\tilde{\omega},u)\|^2}.
$$

Appendix B: Derivation of (9)

$$
\begin{aligned}
f_{pc,j}(q)=&\int\left|\mathcal{F}_{x\to q}\left(\mathcal{S}_j u(x)\left(\omega(x)-y^T\nabla\omega(x)+\tfrac{1}{2}y^T\nabla^2\omega(x)y\right)\right)\right|^2\kappa(y)\mathrm{d}y+\mathcal{O}\left(\int|y|^3\kappa(y)\mathrm{d}y\right)\\
=&\int|\mathcal{F}_{x\to q}(\mathcal{S}_j u(x)(\omega(x)+\tfrac{1}{2}y^T\nabla^2\omega(x)y))|^2\kappa(y)\mathrm{d}y+\int|\mathcal{F}_{x\to q}(\mathcal{S}_j u(x)y^T\nabla\omega(x))|^2\kappa(y)\mathrm{d}y\\
&+\mathcal{O}\left(\int|y|^3\kappa(y)\mathrm{d}y\right)\\
=&\int|\mathcal{A}_j(\omega,u)+\tfrac{1}{2}y^T\mathcal{A}_j(\nabla^2\omega,u)y|^2\kappa(y)\mathrm{d}y+\int|y^T\mathcal{A}_j(\nabla\omega,u)|^2\kappa(y)\mathrm{d}y+\mathcal{O}(\int|y|^3\kappa(y)\mathrm{d}y)\\
=&|\mathcal{A}_j(\omega,u)|^2+2\Re\left(\mathcal{A}_j^*(\omega,u)\left(\tfrac{1}{2}\sigma_1^2\mathcal{A}_j(\nabla_{11}\omega,u)+\tfrac{1}{2}\sigma_2^2\mathcal{A}_j(\nabla_{22}\omega,u)+\sigma_{12}\mathcal{A}_j(\nabla_{12}\omega,u)\right)\right)\\
&+\mathcal{O}\left(\int|y|^4\kappa(y)\mathrm{d}y\right)+\sigma_1^2|\mathcal{A}_j(\nabla_1\omega,u)|^2+\sigma_2^2|\mathcal{A}_j(\nabla_2\omega,u)|^2+\mathcal{O}\left(\int|y|^3\kappa(y)\mathrm{d}y\right)\\
=&|\mathcal{A}_j(\omega,u)|^2+2\Re\left(\mathcal{A}_j^*(\omega,u)\mathcal{A}_j(\tfrac{1}{2}(\sigma_1^2\nabla_{11}\omega+\sigma_2^2\nabla_{22}\omega+2\sigma_{12}\nabla_{12}\omega),u)\right)\\
&+\sigma_1^2|\mathcal{A}_j(\nabla_1\omega),u)|^2+\sigma_2^2|\mathcal{A}_j(\nabla_2\omega,u)|^2+\mathcal{O}\left(\int|y|^4\kappa(y)\mathrm{d}y\right)+\mathcal{O}\left(\int|y|^3\kappa(y)\mathrm{d}y\right)\\
=&|\mathcal{A}_j(\omega+\tfrac{1}{2}(\sigma_1^2\nabla_{11}\omega+\sigma_2^2\nabla_{22}\omega+2\sigma_{12}\nabla_{12}\omega),u)|^2\\
&+\sigma_1^2|\mathcal{A}_j(\nabla_1\omega,u)|^2+\sigma_2^2|\mathcal{A}_j(\nabla_2\omega,u)|^2+\mathcal{O}(\int|y|^3\kappa(y)\mathrm{d}y),
\end{aligned}
\tag{B1}
$$

where the first equality is based on:

$$
\begin{aligned}
&\int \left|\mathcal{F}_{x\to q}(\mathcal{S}_j u(x)(\omega(x) + \tfrac{1}{2}y^T\nabla^2\omega(x)y)) - \mathcal{F}_{x\to q}(\mathcal{S}_j u(x) y^T \nabla\omega(x))\right|^2 \kappa(y)\mathrm{d}y \\
=&\int \left(\left|\mathcal{F}_{x\to q}(\mathcal{S}_j u(x)(\omega(x) + \tfrac{1}{2}y^T\nabla^2\omega(x)y))\right|^2 + \left|\mathcal{F}_{x\to q}(\mathcal{S}_j u(x) y^T \nabla\omega(x))\right|^2\right)\kappa(y)\mathrm{d}y \\
&-2\int_y \Re\left((\mathcal{F}_{x\to q}(\mathcal{S}_j u(x)(\omega(x) + \tfrac{1}{2}y^T\nabla^2\omega(x)y)))^* \mathcal{F}_{x\to q}(\mathcal{S}_j u(x) y^T \nabla\omega(x))\right)\kappa(y)\mathrm{d}y) \\
=&\int \left(|\mathcal{F}_{x\to q}(\mathcal{S}_j u(x)(\omega(x) + \tfrac{1}{2}y^T\nabla^2\omega(x)y))|^2 + |\mathcal{F}_{x\to q}(\mathcal{S}_j u(x) y^T \nabla\omega(x))|^2\right)\kappa(y)\mathrm{d}y \\
&-2\Re\left(\int_y (\mathcal{A}_j(\omega,u) + y^T\mathcal{A}_j(\nabla^2\omega,u)y)^*(y^T\mathcal{A}_j(\nabla\omega,u))\kappa(y)\mathrm{d}y\right) \\
=&\int |\mathcal{F}_{x\to q}(\mathcal{S}_j u(x)(\omega(x) + \tfrac{1}{2}y^T\nabla^2\omega(x)y))|^2\mathrm{d}y + |\mathcal{F}_{x\to q}(\mathcal{S}_j u(x) y^T \nabla\omega(x))|^2\kappa(y)\mathrm{d}y \\
&-2\Re((\mathcal{A}_j(\omega,u))^*\mathcal{A}_j(\nabla\omega,u)\overbrace{\int_y y^T\kappa(y)\mathrm{d}y}^{=0\,(7)} \\
&+\sum_{1\le i_1,i_2,i_3\le 2}\mathcal{A}_j^*(\nabla_{i_1,i_2}\omega,u)\mathcal{A}_j(\nabla_{i_3}\omega,u)\overbrace{\int_y y_{i_1}y_{i_2}y_{i_3}\kappa(y)\mathrm{d}y)}^{=0\,(7)} \\
=&\int \left(|\mathcal{F}_{x\to q}(\mathcal{S}_j u(x)(\omega(x) + \tfrac{1}{2}y^T\nabla^2\omega(x)y))|^2 + |\mathcal{F}_{x\to q}(\mathcal{S}_j u(x) y^T \nabla\omega(x))|^2\right)\kappa(y)\mathrm{d}y,
\end{aligned}
$$

similarly, the first two terms of the fourth equality are derived by:

$$
\begin{aligned}
&\int |\mathcal{A}_j(\omega,u) + \tfrac{1}{2}y^T\mathcal{A}_j(\nabla^2\omega,u)y|^2\kappa(y)\mathrm{d}y \\
=&\int |\mathcal{A}_j(\omega,u)|^2\kappa(y)\mathrm{d}y + \Re\left(\mathcal{A}_j^*(\omega,u)\int y^T\mathcal{A}_j(\nabla^2\omega,u)y\kappa(y)\mathrm{d}y\right) \\
&\quad+\int |\tfrac{1}{2}y^T\mathcal{A}_j(\nabla^2\omega,u)y|^2\kappa(y)\mathrm{d}y \\
=&\int |\mathcal{A}_j(\omega,u)|^2\kappa(y)\mathrm{d}y + \Re(\mathcal{A}_j^*(\omega,u)(\sigma_1^2\mathcal{A}_j(\nabla_{11}\omega,u) \\
&\quad+\sigma_2^2\mathcal{A}_j(\nabla_{22}\omega,u) + 2\sigma_{12}\mathcal{A}_j(\nabla_{12}\omega,u))) + \mathcal{O}\left(\int |y|^4\kappa(y)\mathrm{d}y\right),
\end{aligned}
$$

and the third term of the fourth equality is derived by:

$$
\int |y^T\mathcal{A}_j(\nabla\omega,u)|^2\kappa(y)\mathrm{d}y \overset{(7)}{=} \sigma_1^2|\mathcal{A}_j(\nabla_1\omega,u)|^2 + \sigma_2^2|\mathcal{A}_j(\nabla_2\omega,u)|^2.
$$