Lawrence Berkeley National Laboratory
Recent Work

Title
Partially coherent ptychography by gradient decomposition of the probe.

Permalink
https://escholarship.org/uc/item/9jm4f38s

Journal
Acta crystallographica. Section A, Foundations and advances, 74(Pt 3)

ISSN
2053-2733

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, Pablo Enfedaque, Yifei Lou, and Stefano Marchesini

School of Math. Sci., Tianjin Normal University
Computational Research Division, Lawrence Berkeley National Laboratory
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, cell biology, materials science, and electronics, among others. In a coherent ptychography experiment (Figure 1), a localized coherent X-ray probe (or illumination) \( \omega \) 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, \( \omega \in \mathbb{C}^m \) is a localized 2D probe with \( \sqrt{m} \times \sqrt{m} \) pixels (\( u \) and \( \omega \) are both written as a vector by a lexicographical order), and \( f_j = |\mathcal{F}(\omega \circ S_j u)|^2 \) is a stack of phaseless measurements \( f_j \in \mathbb{R}^n_{+} \forall 0 \leq j \leq J - 1 \). Here \( |cdot| \) represents the element-wise absolute value of a vector, \( \circ \) denotes the element-wise multiplication, and \( \mathcal{F} \) denotes the normalized 2-dimensional discrete Fourier transform. Each \( S_j \in \mathbb{R}^{m \times n} \) is a binary matrix that crops a region \( j \) of size \( m \) 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:

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

where bilinear operators \( A : \mathbb{C}^m \times \mathbb{C}^n \to \mathbb{C}^m \) and \( A_j : \mathbb{C}^m \times \mathbb{C}^n \to \mathbb{C}^m \forall 0 \leq j \leq J - 1 \), are denoted as follows:

\[
A(\omega, u) := (A^T_0(\omega, u), A^T_1(\omega, u), \cdots, A^T_{J-1}(\omega, u))^T,
\]

\[
A_j(\omega, u) := \mathcal{F}(\omega \circ S_j u),
\]

FIG. 1. Ptychographic experiment (Far-field): A stack of intensities \( f_j = |\mathcal{F}(\omega \circ S_j u)|^2 \) are collected, with \( \omega \) 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 \( f := (f_0^T, f_1^T, \cdots, f_{J-1}^T)^T \in \mathbb{R}^n_+ \).

Coherent ptychography 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][10][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 [6], the authors considered a constant quasi-monochromatic illumination on the sam-
ne (beam much larger than the sample), and derived a convolution based model using the mutual optical intensity as:

\[ f_{pc} = f * \kappa, \]  

(2)

where \( f_{pc} \) is the measured partial coherent intensity, \( f \) is the coherent intensity, \( * \) denotes the convolution operator, and \( \kappa \) 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 \( \kappa \) 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:

\[ f_{pc,j} = \sum_{l=0}^{L-1} |A_j(\omega_l, u)|^2, \forall 0 \leq j \leq J - 1, \]  

(3)

with \( L \) orthogonal probes \( \{\omega_l\} \), where \( f_{pc} := (f_{pc,0}, \ldots , f_{pc,J-1})^T \) is the measured intensity. The extended ptychographic iterative engine [18] was adopted to solve such model [9]. Experimental ‘fly-scan’ data with translational blur was successfully reconstructed using [3] by [7] [10] [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 Gradient Decomposition of the Probe (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 \( \kappa \) in a continuous setting (same notations as in the discrete setting) is formulated as:

\[ \int |F_{x \rightarrow q}(S_j u (x) \omega (x - y))|^2 \kappa(y) dy = f_{pc,j}(q), \]  

(4)

where \( f_{pc} \) is the measured partial coherent intensity and \( F_{x \rightarrow q} \) is the normalized Fourier transform. When setting \( \kappa \) to a binary function, the above model is exactly the same as fly-scan ptychography [7] [16] [21]. Setting \( \kappa \) to the Dirac delta function reduces it to the coherent model [1]. We remark that (4) is quite different from [2], since [1] 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. (1) 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 \( \{w_l\} \):

\[ f_{pc,j} = \sum_{l=0}^{L-1} w_l |A_j(\tau_l \omega, u)|^2, \forall 0 \leq j \leq J - 1, \]  

(5)

with the translation operators \( \{\tau_l\} \), and partially coherent intensity \( f_{pc} \). However, those methods in [20] [28] cannot be directly applied to [4], since either the weights
The above relation holds if the kernel function is centro-

dimensional. Hence instead of solving directly as in [20], in the following sections we will formulate 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:

\[
f_{pc,j}(q) = \int \mathcal{F}_x \left[ \mathcal{S}_j u(x)(\omega(x) - y^T \nabla \omega(x) + \frac{1}{2} y^T \nabla^2 \omega(x) y + \mathcal{O}(|y|^3)) \right] \kappa(y) dy,
\]

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

\[
\max_{\kappa \in \mathbb{R}^2} \max_{|\kappa| = 3} |\partial^3 \omega(x)| \leq C,
\]

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 \( m_{i_1,i_2} = \int \kappa(y) y_{i_1} y_{i_2}^2 dy \), normalized with \( m_{0,0} = 1 \), center of mass \( m_{1,0,0} = (0,0) \), and second order moments \( \sigma^2_1 \), \( \sigma^2_2 \), and \( \sigma_{12} \) \( \sigma^2_1 \sigma^2_2 - \sigma^2_{12} \geq 0 \). We further assume that:

\[
m_{1,i_2} = 0, \text{ if } \text{mod}(i_1 + i_2, 2) = 1.
\]

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

\[
\int |y|^k \kappa(y) = \mathcal{O}(\max\{\sigma_1, \sigma_2\}^k), \quad k \geq 3.
\]

Therefore, one has:

\[
f_{pc,j}(q) = |A_j(\omega + \frac{1}{2} (\sigma^2_1 \nabla_{11} \omega + \sigma^2_2 \nabla_{22} \omega + 2 \sigma_{12} \nabla_{12} \omega), u)|^2
+ \sigma^2_1 |A_j(\nabla_1 \omega, u)|^2 + \sigma^2_2 |A_j(\nabla_2 \omega, u)|^2
+ \mathcal{O}(\int |y|^3 \kappa(y) dy).
\]

(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):

\[
\tilde{\omega} := \omega + \frac{1}{2} (\sigma^2_1 \nabla_{11} \omega + \sigma^2_2 \nabla_{22} \omega + 2 \sigma_{12} \nabla_{12} \omega),
\]

and nonlinear operator \( G_j : \mathbb{C}^m \times \mathbb{C}^n \times \mathbb{R}^2 \rightarrow \mathbb{R}_+^m \):

\[
G_j(\omega, u, \sigma) := |A_j(\omega, u)|^2 + \sigma^2_1 |A_j(\nabla_1 \omega, u)|^2 + \sigma^2_2 |A_j(\nabla_2 \omega, u)|^2,
\]

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

\[
\text{GDP} : f_{pc,j} \approx G_j(\tilde{\omega}, u, \sigma).
\]

(11)

a. Remarks:

- The above assumptions hold if \( \kappa \) is a Gaussian kernel function, which well approximates standard light sources such as synchrotrons [21, 22], SASE FELs [23], and others. When \( \sigma = (0,0) \), the above formula reduces to the coherent model in [4].

- The variance adjusted probe \( \tilde{\omega} \) incorporates second order effects (such as the covariance \( \sigma_{12} \)) into the illumination, while reducing the cost of computing the intensities using the forward model [11] compared to [6].

- If the probe \( \omega \) has a support by a lens of finite size, \( (\omega)(q) = 0 \quad \forall \ q \in \Omega_0 \subset \mathbb{R}^2 \), then the same support applies to the variance adjusted probe \( \forall \ q \in \Omega_0 \), \( (\tilde{\omega})(q) = 0 \) due to the Fourier transform relationship:

\[
(\mathcal{F} \tilde{\omega})(q) = \mathcal{F}((1 + \frac{1}{2} \sigma^2_1 \nabla_{11} + \frac{1}{2} \sigma^2_2 \nabla_{22} + \sigma_{12} \nabla_{12}) \omega)(q)
= (1 - \frac{1}{2} \sigma^2_1 q^2_1 - \frac{1}{2} \sigma^2_2 q^2_2 - \sigma_{12} q_1 q_2) (\mathcal{F} \omega)(q),
\]

\( (\forall \ q = (q_1, q_2) \in \mathbb{R}^2). \)

- Similarly to the decomposition model [3], GDP has three different modes \( \{\omega, \nabla_1 \omega, \nabla_2 \omega\} \). The difference between them is that in the GDP model, two modes \( \nabla_i \omega \) can be expressed by the first mode \( \omega \) explicitly, while the multiple modes in the decomposition model [3] are only assumed to be orthogonal to each other [21, 22]. 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:

\[
\min_{\tilde{\omega}, u, \sigma} \frac{1}{2} \sum_{j=0}^{J-1} \left\| \sqrt{f_{pc,j}} - \sqrt{G_j(\tilde{\omega}, u, \sigma)} \right\|^2,
\]

(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 \( \sigma \) 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-differentiable
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 [11]. By introducing auxiliary variables $z_{l,j} \in \mathbb{C}^{m}$, one has:

$$\min_{\tilde{\omega}, u, \sigma, z_{l,j}} \frac{1}{2} \sum_{j=0}^{J} \left| \sqrt{f_{pc,j}} - \sqrt{\sum_{l=0}^{2} |z_{l,j}|^2} \right|^2,$$

s.t. $z_{0,j} = A_j(\tilde{\omega}, u), z_{1,j} = \sigma_1 A_j(\nabla_1 \tilde{\omega}, u), z_{2,j} = \sigma_2 A_j(\nabla_2 \tilde{\omega}, u), 0 \leq j \leq J - 1.$

The corresponding augmented Lagrangian reads as:

$$L_r(\tilde{\omega}, u, \{z_l\}_{l=0}^{2}, \sigma, \{\Lambda_l\}_{l=0}^{2}) := \frac{1}{2} \sum_{j} \left| \sqrt{f_{pc,j}} - \sqrt{\sum_{l} |z_{l,j}|^2} \right|^2 + r \Re(\langle z_0 - A(\tilde{\omega}, u), \Lambda_0 \rangle) + \frac{r}{2} \|z_0 - A(\tilde{\omega}, u)\|^2$$

$$+ r \Re(\langle z_1 - A(\sigma_1 \nabla_1 \tilde{\omega}, u), \Lambda_1 \rangle) + \frac{r}{2} \|z_1 - A(\sigma_1 \nabla_1 \tilde{\omega}, u)\|^2$$

$$+ r \Re(\langle z_2 - A(\sigma_2 \nabla_2 \tilde{\omega}, u), \Lambda_2 \rangle) + \frac{r}{2} \|z_2 - A(\sigma_2 \nabla_2 \tilde{\omega}, u)\|^2,$$

with multipliers $\Lambda_l = (\Lambda_{l,0}^{T}, \ldots, \Lambda_{l,J-1}^{T})^{T} \in \mathbb{C}^{m}$, $z_l = (z_{l,0}^{T}, \ldots, z_{l,J-1}^{T})^{T} \in \mathbb{C}^{m}$, 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 Ptychography in [11]. Here we propose a new, more robust variant of ADMM. It employs an additional proximal term added to the augmented Lagrangian: $\frac{r}{2} \|u - u^k\|^2_{MK}$ 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 $\nabla_1, \nabla_2$ 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 $\mathbb{R}$.

---

**Algorithm 1: GDP-ADMM**

**Initialization:** Set $\tilde{\omega}^0 = F^*(\frac{1}{2} \sum \sqrt{f_{pc,j}}), u^0 = 1, \sigma_0^1 = \sigma_0^2 = 1, z_0^0 = \mathbb{A}(\tilde{\omega}^0, u^0), z_1^0 = \sigma_0^2 A(\nabla_1 \tilde{\omega}^0, u^0), z_2^0 = \sigma_0^2 A(\nabla_2 \tilde{\omega}^0, u^0), \Lambda_1^0 = 0, k := 0, \text{maximum iteration number } \text{Iter}_{\text{Max}}$, parameters $r$ and $\tau$.

**output:** $u^* := u^{\text{Iter}_{\text{Max}}}$ and $\tilde{\omega}^* := \tilde{\omega}^{\text{Iter}_{\text{Max}}}$

for $k = 0$ to $\text{Iter}_{\text{Max}} - 1$

1. Compute $\tilde{\omega}^{k+1}$ by solving the linear system:

$$\tilde{\omega}^{k+1} = (N_1^k)^{-1} c^k,$$

using conjugate gradient method, where $N_1^k$ and $c^k$ are given in Eq. (A3) (Appendix A).

2. Compute $u^{k+1}$ as

$$u^{k+1} = (N_2^k + \frac{\tau}{r} M^k)^{-1} (b^k + \frac{\tau}{r} M^k u^k).$$

with diagonal matrices $N_2^k, M^k$ and vector $b^k$ defined in Eq. (A5), and (A6) (see Appendix A).

3. Compute $z_i^{k+1}$ as:

$$z_i^{k+1} = \sqrt{f_{pc,j}} + r \sqrt{\sum_{d=0}^{2} |z_{d,j}^{k+1}|^2} \circ \text{sign}(z_i^{k+1}),$$

with $\text{sign}(z_i,j) := \frac{z_i,j}{\sqrt{\sum_{d=0}^{2} |z_{d,j}^{k+1}|^2}}$, $z_{0,j}^{k+1} = A(\tilde{\omega}^{k+1}, u^{k+1}) - \Lambda_0^k,$ $z_{1,j}^{k+1} = \sigma_1^k A(\nabla_1 \tilde{\omega}^{k+1}, u^{k+1}) - \Lambda_1^k,$ $z_{2,j}^{k+1} = \sigma_2^k A(\nabla_2 \tilde{\omega}^{k+1}, u^{k+1}) - \Lambda_2^k$.

4. Compute $\sigma_i^{k+1}$ as:

$$\sigma_i^{k+1} = \frac{\Re(\langle z^{k+1} + \Lambda_i^k A(\nabla_i \tilde{\omega}^{k+1}, u^{k+1}) \rangle)}{||A(\nabla_i \tilde{\omega}^{k+1}, u^{k+1})||^2}, i = 1, 2.$$  

5. Update the multipliers $\{\Lambda_i^{k+1}\}$ as

$$\Lambda_0^{k+1} = \Lambda_0^k + (z_{0,j}^{k+1} - A(\tilde{\omega}^{k+1}, u^{k+1}));$$

$$\Lambda_1^{k+1} = \Lambda_1^k + (z_{1,j}^{k+1} - \sigma_1^k A(\nabla_1 \tilde{\omega}^{k+1}, u^{k+1}));$$

$$\Lambda_2^{k+1} = \Lambda_2^k + (z_{2,j}^{k+1} - \sigma_2^k A(\nabla_2 \tilde{\omega}^{k+1}, u^{k+1})).$$
IV. NUMERICAL EXPERIMENTS

The experimental setup of this section is introduced next. The reference specimen used is a complex valued image “Goldballs” \cite{19} of size $249 \times 249$ pixels (Figure 2). The probe is an Airy disk of size $64 \times 64$ pixels (Figure 2).

We compare the proposed GDP-ADMM (Appendix A) with the “fully-coherent” model by ADMM (FC-ADMM \cite{4}). Both algorithms are executed with a maximum iteration number of 300. We measure the quality of recovery images by relative residuals as:

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

$$e_{pc} = \frac{\left\| \sqrt{T} - \sqrt{\hat{\omega} \hat{\sigma}^k A(\omega^k, \sigma^k)} \right\|_1}{\| \sqrt{T} \|_1},$$

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

$$\text{SNR}(u^k, u_{true}) = -20 \min_{c \in \mathbb{C}} \log \frac{\left\| u^k - c u_{true} \right\|_2}{\left\| u^k \right\|_2},$$

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

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

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

(18)

The discrete truncated Gaussian matrix is generated by \texttt{fspecial} in MATLAB with variances $\sigma_1, \sigma_2$ 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 $\text{Dist} = 8$ pixels (slightly bigger than the beam width) as a default, and we incorporate the support constraint of the lens as in \cite{19}. The parameter $\tau$ is selected manually with default value 0.2, and $\tau = 1.01\tau$.

a. Noiseless data. Following \cite{4}, the partially coherent intensity in a discrete setting is generated as

$$f_{pc,j} = \sum_{i} \kappa_i |F((T_j\omega) \circ S_j u)|^2,$$

(19)

with translation operator $T_j$, 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 $\sigma_1, \sigma_2$, 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 1. 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).

<table>
<thead>
<tr>
<th>( \mathbf{\sigma} )</th>
<th>( e_{fc} )</th>
<th>( e_{pc} )</th>
<th>SNR(_{fc} )</th>
<th>SNR(_{pc} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(2.2)</td>
<td>1.36E-1 1.88E-1 2.11E-1 2.20E-1</td>
<td>2.60E-2 5.68E-2 8.95E-2 1.14E-1</td>
<td>14.78 9.53 6.14 4.37</td>
<td>23.73 18.83 13.02 8.37</td>
</tr>
<tr>
<td>(5.5)</td>
<td>(4.0)</td>
<td>(5.0)</td>
<td>(6.0)</td>
<td>(7.0)</td>
</tr>
</tbody>
</table>

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 \( \gamma \), the weaker the noise). The corresponding SNRs of the data, defined by SNR\((f^{noise}, f)\), are 23.89, 29.87, 35.91, and 41.07.

completely 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 \( \| \mathbf{\sigma} \| \) 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

\[
f_{pc,j}^{\text{noise}} = \frac{1}{\gamma} \text{Poisson} (\gamma f_{pc,j}),
\]

where \( \gamma > 0 \) is used to control the noise level, and \( f_{pc,j} \) is the clean partial coherent intensity data defined by 15). Figure 4 shows the recovered images at different noise levels. It can be seen that, as the noise level decreases (increasing \( \gamma \)), 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 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 \( \text{Dist} \in \{4, 6, 8, 12\} \), which determine the redundancy 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 \( \frac{1}{2} \| u \|^2 \).
Using a binary kernel function for fly-scan ptychography. Figure 7 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 7 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.

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 $A_1, A_2$ 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 [21], the probe and the weights in [7] in the case of the setting in Figure 3 (a)-(b), at most $(4\sigma_1 + 1) \times (4\sigma_2 + 1) = 441$ translated probes for $\sigma_1 = \sigma_2 = 5$ should be introduced, which requires much more memory and computation costs.

![Graph](image)

**Figure 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, $\sigma = (4, 4)$, beam width $D = 7$.

![Images](image)

**Figure 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.

![Table](image)

**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.

<table>
<thead>
<tr>
<th>$d$</th>
<th>11</th>
<th>13</th>
<th>15</th>
<th>17</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\varepsilon_{fc}$</td>
<td>2.00E-1</td>
<td>2.16E-1</td>
<td>2.29E-1</td>
<td>2.37E-1</td>
</tr>
<tr>
<td>$\varepsilon_{pc}$</td>
<td>4.47E-2</td>
<td>6.46E-2</td>
<td>9.04E-2</td>
<td>1.16E-1</td>
</tr>
<tr>
<td>SNR$_{fc}$</td>
<td>6.19</td>
<td>5.60</td>
<td>7.15</td>
<td>6.71</td>
</tr>
<tr>
<td>SNR$_{pc}$</td>
<td>21.15</td>
<td>18.07</td>
<td>14.63</td>
<td>11.38</td>
</tr>
</tbody>
</table>
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 $\sigma_j$ 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].

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 $\sigma_j$ 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.


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:

$$\tilde{\mathcal{L}}_F(\tilde{\omega}, u, z, \sigma, \Lambda) = \mathcal{L}_F(\tilde{\omega}, u, z, \sigma, \Lambda) + \frac{\rho}{2}\|u - u_k\|^2_{M_k},$$

where for simplicity we removed superscripts and subscripts, and $\|u\|^2_{M_k}$ is a positive semi-definite matrix $M_k$, defined by the approximated solution $(\tilde{\omega}^{k+1}, u^k, \{z^k_i\}, \{\sigma^k_i\}, \{\Lambda^k_i\})$ in the $k^{th}$ iteration. We remark that with the help of an additional proximal term $\frac{\rho}{2}\|u - u_k\|^2_{M_k}$, the subproblem in Step 2 admits a unique solution. Below we demonstrate how to solve each subproblem.

Step 1: $\tilde{\omega}^{k+1} = \arg\min_{\omega} \mathcal{L}_F(\tilde{\omega}, u^k, \{z^k_i\}, \sigma^k, \{\Lambda^k_i\})$.

Step 2: $u^{k+1} = \arg\min_u \mathcal{L}_F(\tilde{\omega}^{k+1}, u, \{z^k_i\}, \sigma^k, \{\Lambda^k_i\})$.

Step 4: $\{z^k_i\}_{i=0}^l = \arg\min_{\{z^k_i\}} \mathcal{L}_F(\tilde{\omega}^{k+1}, u^{k+1}, \{z^k_i\}, \sigma^k, \{\Lambda^k_i\})$.

Step 5: $\sigma^{k+1} = \arg\min_{\sigma} \mathcal{L}_F(\tilde{\omega}^{k+1}, u^{k+1}, \{z^k_i\}, \sigma^k, \{\Lambda^k_i\})$.

Step 6: $\Lambda^k_{l+1} = \Lambda^k_{l+1} - (\tilde{\omega}^{k+1} - \tilde{\omega}^{k+1}, u^{k+1})$.

(1A)

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

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

with symmetric sparse matrix $N_k^k$ defined as:

$$N_k := \text{diag} \left( \sum_j |S_j u_j|^2 \right) + \frac{2}{\sigma^k} \nabla^T \left( \text{diag} \left( \sum_j |S_j u_j|^2 \right) \nabla \right),$$

and $\tilde{z}^k_{l,j} := \mathcal{F}^*(\tilde{z}^k_{l,j} + \Lambda^k_{l,j})$. In this section, the gradient operator is given in a discrete setting for simplicity, where $V_1, V_2$ 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:

$$\left( N_2^k + z^k M^k \right) u^{k+1} - (b^k + z^k M^k u^k) = 0,$$
with diagonal matrix $N^k_2$:

$$N^k_2 := \text{diag} \left( \sum_j S^T_j \left( |\omega^{k+1}|^2 + \sum_{l=1}^2 \sigma^2_l |\nabla (\omega^{k+1})|^2 \right) \right).$$

$$b^k := \sum_j S^T_j \left( (\omega^{k+1})^* \circ z^0_{0,j} + \sum_{l=1}^2 \sigma^k_l \nabla (\omega^{k+1})^* \circ z^k_{l,j} \right).$$

(A5)

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

$$M^k(l, l) := \begin{cases} 0, & \text{if } N^k_2(l, l) \geq \frac{1}{10} \|\text{diag}(N^k_2)\|_\infty; \\ \|\text{diag}(N^k_2)\|_\infty - \frac{2}{5} N^k_2(l, l), & \text{otherwise.} \end{cases}$$

Step 3 can be expressed as the following problem:

$$\min_{\{z_{i,j}\}_{0 \leq i \leq J}} \frac{1}{2} \left\| \sqrt{f_{pc,j}} - \sqrt{\sum_i |z_{i,j}|^2} \right\|^2 \quad \begin{array}{ll}
+ \frac{\tau}{2} \|z_0 - A(\omega, u)\|^2 \\
+ \|z_1 + A_1(\sigma_1 \nabla \omega, u)\|^2 \\
+ \|z_2 + A_2(\sigma_2 \nabla \omega, u)\|^2
\end{array}$$

(A6)

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

$$\min_{\{z_{i,j}\}_{0 \leq i \leq J}} \left\{ \sqrt{f_{pc,j}} - \sqrt{\sum_i |z_{i,j}|^2} \right\|^2 \quad \begin{array}{ll}
+ \frac{\tau}{2} \|z_0 - A(\omega, u)\|^2 \\
+ \|z_1 + A_1(\sigma_1 \nabla \omega, u)\|^2 \\
+ \|z_2 + A_2(\sigma_2 \nabla \omega, u)\|^2
\end{array}$$

where $z_{0,j} = A_j(\omega, u) - \Lambda_{0,j}$, $z_{1,j} = \sigma_1 A_j(\nabla \omega, u) - \Lambda_{1,j}$, $z_{2,j} = \sigma_2 A_j(\nabla \omega, u) - \Lambda_{2,j}$. 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:

$$\sigma_l = \frac{\Re(\langle z_l + A_l, A(\nabla \omega, u) \rangle)}{\|A(\nabla \omega, u)\|^2}.$$

Appendix B: Derivation of \((\text{B})\)

$$f_{pc,j}(q) = \int |F_{x \rightarrow q}(S_j u(x)(\omega(x) - y^T \nabla \omega(x) + \frac{1}{2} y^T \nabla^2 \omega(x)y))|^2 \kappa(y) dy + \mathcal{O} \left( \int |y|^3 \kappa(y) dy \right)$$

$$= \int |F_{x \rightarrow q}(S_j u(x) y^T \nabla \omega(x))|^2 \kappa(y) dy + \mathcal{O} \left( \int |y|^3 \kappa(y) dy \right)$$

$$= |A_j(\omega, u)|^2 + 2 \Re \langle A^*_j(\omega, u) (\frac{1}{2} |\sigma_1^2 A_j(\nabla_{11} \omega, u) + \frac{1}{2} |\sigma_2^2 A_j(\nabla_{22} \omega, u) + \sigma_1 A_j(\nabla_{12} \omega, u)) \rangle$$

$$+ \mathcal{O} \left( \int |y|^3 \kappa(y) dy \right)$$

$$= |A_j(\omega, u)|^2 + 2 \Re \langle A^*_j(\omega, u) A_j(\frac{1}{2} (|\sigma_1^2 \nabla_{11} \omega + |\sigma_2^2 \nabla_{22} \omega + 2 \sigma_{12} \nabla_{12} \omega), u) \rangle$$

$$+ \sigma_1^2 |A_j(\nabla_{11} \omega, u)|^2 + \sigma_2^2 |A_j(\nabla_{22} \omega, u)|^2 + \mathcal{O} \left( \int |y|^4 \kappa(y) dy \right) + \mathcal{O} \left( \int |y|^3 \kappa(y) dy \right)$$

$$= |A_j(\omega + \frac{1}{2} (|\sigma_1^2 \nabla_{11} \omega + |\sigma_2^2 \nabla_{22} \omega + 2 \sigma_{12} \nabla_{12} \omega), u)|^2$$

$$+ \sigma_1^2 |A_j(\nabla_{11} \omega, u)|^2 + \sigma_2^2 |A_j(\nabla_{22} \omega, u)|^2 + \mathcal{O} \left( \int |y|^4 \kappa(y) dy \right) + \mathcal{O} \left( \int |y|^3 \kappa(y) dy \right).$$
where the first equality is based on:

\[
\int \left| \mathcal{F}_x \cdot q \left( S_j u(x) (\omega(x) + \frac{1}{2} y^T \nabla^2 \omega(x)y) \right) - \mathcal{F}_x \cdot q \left( S_j u(x) y^T \nabla \omega(x) \right) \right|^2 \kappa(y) dy
\]

\[
= \int \left( \left| \mathcal{F}_x \cdot q \left( S_j u(x) (\omega(x) + \frac{1}{2} y^T \nabla^2 \omega(x)y) \right) \right|^2 + \left| \mathcal{F}_x \cdot q \left( S_j u(x) y^T \nabla \omega(x) \right) \right|^2 \right) \kappa(y) dy
\]

\[
- 2 \int_y \Re \left( \mathcal{F}_x \cdot q \left( S_j u(x) (\omega(x) + \frac{1}{2} y^T \nabla^2 \omega(x)y) \right) \right) * \mathcal{F}_x \cdot q \left( S_j u(x) y^T \nabla \omega(x) \right) \kappa(y) dy
\]

\[
= \int \left( \left| \mathcal{F}_x \cdot q \left( S_j u(x) (\omega(x) + \frac{1}{2} y^T \nabla^2 \omega(x)y) \right) \right|^2 + \left| \mathcal{F}_x \cdot q \left( S_j u(x) y^T \nabla \omega(x) \right) \right|^2 \right) \kappa(y) dy
\]

\[
- 2 \Re \left( \left( A_j (\omega, u) + y^T A_j (\nabla^2 \omega, u)y \right) * \left( y^T A_j (\nabla \omega, u) \right) \kappa(y) dy \right)
\]

\[
= \int \left( \left| \mathcal{F}_x \cdot q \left( S_j u(x) (\omega(x) + \frac{1}{2} y^T \nabla^2 \omega(x)y) \right) \right|^2 + \left| \mathcal{F}_x \cdot q \left( S_j u(x) y^T \nabla \omega(x) \right) \right|^2 \right) \kappa(y) dy,
\]

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

\[
\int \left| A_j (\omega, u) + \frac{1}{2} y^T A_j (\nabla^2 \omega, u)y \right|^2 \kappa(y) dy
\]

\[
= \int \left| A_j (\omega, u) \right|^2 \kappa(y) dy + \Re \left( A_j^* (\omega, u) \int y^T A_j (\nabla^2 \omega, u)y \kappa(y) dy \right)
\]

\[
+ \int \left| \frac{1}{2} y^T A_j (\nabla^2 \omega, u)y \right|^2 \kappa(y) dy
\]

\[
= \int \left| A_j (\omega, u) \right|^2 \kappa(y) dy + \Re \left( A_j^* (\omega, u) \left( \sigma_1^2 A_j (\nabla_{11} \omega, u) + \sigma_2^2 A_j (\nabla_{12} \omega, u) \right) \right)
\]

\[
+ \sigma_2^2 A_j (\nabla_{22} \omega, u) + 2 \sigma_{12} A_j (\nabla_{12} \omega, u)) + O \left( \int |y|^4 \kappa(y) dy \right),
\]

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

\[
\int |y^T A_j (\nabla \omega, u)|^2 \kappa(y) dy \overset{\sigma_1 \sigma_2}{=} \sigma_1^2 |A_j (\nabla_{11} \omega, u)|^2 + \sigma_2^2 |A_j (\nabla_{12} \omega, u)|^2.
\]