arXiv:submit/0903412 [math.OC] 3 Feb 2014

ALTERNATING PROJECTION, PTYCHOGRAPHIC IMAGING AND PHASE SYNCHRONIZATION

STEFANO MARCHESINI[‡], YU-CHAO TU[†], AND HAU-TIENG WU[◊]

Abstract. We demonstrate the global convergence of the alternating projection algorithm to a unique solution up to a global phase factor. Additionally, for the ptychographic imaging problem, we discuss phase synchronization and connection graph Laplacian, and show how to construct an accurate initial guess to accelerate convergence speed to handle the big imaging data in the coming new light source era.

1. Introduction

The reconstruction of a scattering potential from measurements of scattered intensity in the far-field has occupied scientists and applied mathematicians for over a century, and arises in fields as varied as optics [32, 45], astronomy [33], X-ray crystallography [26], tomographic imaging [52], holography [21, 49], electron microscopy [38] and particle scattering generally. Although phase-less diffraction measurements using short wavelength (such as X-ray, neutron, or electron wavepackets) have been at the foundation of some of the most dramatic breakthrough in science - such as the first direct confirmation of the existence of atoms [10, 11], the structure of DNA [67], RNA [23] and over 70,000 proteins or drugs involved in human life [8, 43] - the solution to the scattering problem for a general object was generally thought to be impossible for many years. Nevertheless, numerous experimental techniques that employ forms of interferometric/holographic [21, 49] measurements, gratings [56], and other phase mechanisms like random phase masks, sparsity structure, etc [1, 4, 15, 14, 65, 30, 66, 3] to help overcome the problem of phase-less measurements have been proposed over the years [54, 29, 37].

More recently an experimental technique has emerged that enables to image what no-one was able to see before: macroscopic specimens in 3D at wavelength (i.e. potentially atomic) resolution, with chemical state specificity. Ptychography was proposed in 1969 [41, 40, 53, 17, 57] to improve the resolution in electron or x-ray microscopy by combining microscopy with scattering measurements. This technique enables one to build up very large images at wavelength resolution by combining the large field of view of a high precision scanning microscope system with the resolution enabled by diffraction measurements.

Initially, technological problems made ptychography impractical. Now thanks to advances in source brightness [18, 9] and detector speed [12, 25], research institutions around the world are rushing to develop hundreds of ptychographic microscopes to help scientists understand ever more complex nano-materials, self-assembled devices, or to study different length-scales involved in life, from macro-molecular machines to bones [24], and whenever observing the whole picture is as important as recovering local atomic arrangement of the components.

Experimentally, ptychography works by retrofitting a scanning microscope with a parallel detector. In a scanning microscope, a small beam is focused onto the sample via a lens, and the transmission is measured in a single- element detector. The image is built up by plotting the transmission as a function of the sample position as it is rastered across the beam. In such microscope, the resolution of the image is given by the beam size. In ptychography, one replaces the single element detector with a two-dimensional array detector such as a CCD and measures the intensity distribution at many scattering angles, much like a radar detector system for the microscopic world. Each recorded diffraction pattern contains short-spatial Fourier frequency information [17] about features that are smaller than the beam-size, enabling higher resolution. At short wavelengths however it is only possible to measure the intensity of the diffracted light. To reconstruct an image of the object, one needs to retrieve the phase. The phase retrieval problem is made tractable in ptychography by recording multiple diffraction patterns from the same region of the object, compensating phase-less information with a redundant set of measurements.

[‡] Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, CA 94720.
[†] Department of Mathematics, Princeton, NJ 08540.
[◊] Department of Mathematics, Stanford University, Stanford, CA 94305.

While reconstruction methods often work well in practice, fundamental mathematical questions concerning their convergence remain unresolved. The reader of an experimental paper is often left to wonder if the image and the resulting claims are valid, or one possibility among many solutions. Retractions of experimental results do happen (see [62] for a discussion of controversial results in the optical community), and the problem is exacerbated because reproducing an image a nanoscale object is often not practical. What are often referred to as convergence results for projection algorithms are far from what we need for global convergence [45].

A popular algorithm for solving the phase retrieval problem was proposed in 1972. In their famous paper, Gerchberg and Saxton [35], independently of previous mathematical results for projections onto convex sets, proposed a simple algorithm for solving phase retrieval problems in two dimensions. In [44] the algorithm was recognized as a projection algorithm that involves alternating projections between measurement space and object space. In 1982 Fienup [32] generalized the Gerchberg-Saxton algorithm and analyzed many of its properties, showing, in particular, that the directions of the projections in the generalized Gerchberg-Saxton algorithm are formally similar to directions of steepest descent for a distance metric. Projection algorithms for convex sets have been well understood since 1960s. The phase retrieval problem, however, involves nonconvex sets. For this reason, the convergence properties of the Gerchberg-Saxton algorithm and its variants is still an open question except in very special cases [45].

In this paper, based on the well established phase retrieval paper [2], we demonstrate the global convergence of the alternating projection (AP) algorithm to the unique solution up to a global phase factor, and apply it to the ptychographic imaging problem. Additionally, we survey the intimate relationship between the AP algorithm and the notion of *phase synchronization*, which motivates the application of the recently developed technique *connection graph Laplacian*. Phase synchronization and connection graph Laplacian are applied to quickly construct an accurate initial guess for the alternating projection algorithm to accelerate convergence speed for large scale diffraction data problems.

The paper is organized as following. In Section 2 we introduce the ptychography experimental setup and notation. In Section 3 we show the global convergence of AP. In Section 4 we discuss the relationship between the AP algorithm and optimization and show that the second derivative of the associated objective function is positive close to the solution. In Section 5 we discuss the relationship between the AP algorithm and the notion of phase synchronization, and propose methods based on connection graph Laplacian to obtain an accurate initial guess. In Section 6 we show numerical results of proposed methods, we also propose a new lens design and synchronization strategies that achieve over 40× faster convergence than the AP algorithm.

2. Background and notations

We start from summarizing notations we use in this paper. Denote $\mathbb{R}$ to be the real field, $\mathbb{C}$ to be the complex field and $\mathbb{R}_+ = \{x \geq 0, x \in \mathbb{R}\}$ to be the set of non-negative real numbers. We will use the boldface symbol to denote vectors in the column form and non-boldface symbol to denote scalars. We consider the Hilbert space $\mathbb{C}^L$, $L \in \mathbb{N}$ with the inner product $\langle \boldsymbol{u}, \boldsymbol{w} \rangle := \sum_{k=1}^{L} \boldsymbol{u}(k)^* \boldsymbol{w}(k)$, where $\boldsymbol{u}, \boldsymbol{w} \in \mathbb{C}^L$, $\boldsymbol{u}(i)$ is the i-th entry of $\boldsymbol{u}$. We also use the notation $u_i := \boldsymbol{u}(i)$ to denote the i-th entry of $\boldsymbol{u}$. Also, the complex conjugate $\overline{\boldsymbol{u}} \in \mathbb{C}^L$ is a column vector with its l-th entry $\boldsymbol{u}(l)^*$ and $\boldsymbol{w}^*$ is the complex conjugate transpose of $\boldsymbol{w}$, which is a $1 \times L$ row vector. With the inner product, define $\|\boldsymbol{u}\| := \sqrt{\sum_{i=1}^{m} |\boldsymbol{u}(i)|^2} = \sqrt{\boldsymbol{u}^* \boldsymbol{u}}$ to be the Euclidean norm of $\boldsymbol{u}$. Let $\boldsymbol{e}_l \in \mathbb{C}^L$ to be the unit vector with 1 in the l-th entry and $\mathbf{1}$ to be the vector with 1 in all entries.

Given a function $f : \mathbb{C} \to \mathbb{C}$, $f(\boldsymbol{u})$ is defined as the vector so that its i-th entry is $f(\boldsymbol{u}(i))$. For example, the vector $\frac{1}{\boldsymbol{u}}$ has its i-th entry as $\frac{1}{\boldsymbol{u}(i)}$; $|\boldsymbol{u}|$ is the entry-wise modulation of $\boldsymbol{u}$, that is, $|\boldsymbol{u}| \in \mathbb{R}_+^L$ and the j-th entry of $|\boldsymbol{u}|$ is $|\boldsymbol{u}(j)|$; $\boldsymbol{u}^p$, where $p \in \mathbb{R}$, is the p-th power of u entriwisely, that is, the i-th entry of $\boldsymbol{u}^p$ is just $\boldsymbol{u}(i)^p$. Also, we have an indicator vector for $\boldsymbol{u} \in \mathbb{C}^L$, denoted as $\chi_{\boldsymbol{u}} \in \mathbb{R}^L$, that is, $\chi_{\boldsymbol{u}}(i) = 1$ when $\boldsymbol{u}(i) \neq 0$ and $\chi_{\boldsymbol{u}}(j) = 0$ when $\boldsymbol{u}(i) = 0$. Given a function $g : \mathbb{C} \times \mathbb{C} \to \mathbb{C}$, $g(\boldsymbol{u}, \boldsymbol{v})$ is defined as the vector so that its i-th entry is $g(\boldsymbol{u}(i), \boldsymbol{v}(i))$, where $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{C}^L$. For example, the division $\frac{\boldsymbol{u}}{\boldsymbol{v}}$ and production $\boldsymbol{uv}$ are intended as element-wise operations, that is, the j-th entry of $\frac{\boldsymbol{u}}{\boldsymbol{v}}$ (resp. $\boldsymbol{uv}$) is defined as $\frac{\boldsymbol{u}(j)}{\boldsymbol{v}(j)}$ (resp. $\boldsymbol{u}(j)\boldsymbol{v}(j)$). Furthermore, we denote $\mathrm{diag}(\boldsymbol{u})$ to be a diagonal matrix so that its i-th diagonal entry is $\boldsymbol{u}(i)$. With this notation, we know that $\boldsymbol{uv} = \mathrm{diag}(\boldsymbol{u})\boldsymbol{v}$ when $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{C}^L$.

Given a matrix $A \in \mathbb{C}^{L \times L'}$, where $L, L' \in \mathbb{N}$, we denote A^T to be the transpose of A and A^* to be the conjugate transpose of A, that is, $A^* = \overline{A^T}$. Also we denote A_{ij} to be the (i,j)-th entry of A. For two matrices $A, B \in \mathbb{C}^{L \times L'}$, we define $A \circ B$ to be the Hadamard product, that is, $(A \circ B)_{ij} = A_{ij} B_{ij}$. Note

that if we view $\boldsymbol{u}, \boldsymbol{w} \in \mathbb{C}^L$ as $L \times 1$ matrices, $\boldsymbol{uw}$ in the vector form is actually the Hadamard product of these two matrices. To express the notation in a compact format, we stack the columns of a complex matrix $A \in \mathbb{C}^{L\times L'}$ representing "data" into a complex vector $A^\vee \in \mathbb{C}^{LL'}$, where $L, L' \in \mathbb{N}$, where the superscript $\vee$ means the *vector form*, that is, the $(l-1)L+1$-th to the lL-th entries in $A^\vee$ is the l-th column of A, where $l = 1, \dots L'$.

We will denote $\mathbb{T}_1$ to be the unit torus embedded in $\mathbb{C}$, that is, $\mathbb{T}_1 = \{e^{it}, t \in [0, 2\pi)\}$. Given $\mathbf{a} \in \mathbb{R}^m_+$, the notation $\mathbb{T}_{\mathbf{a}}$ means the real torus embedded in $\mathbb{C}^L$, that is, $\mathbb{T}_{\mathbf{a}} := \{\boldsymbol{u} \in \mathbb{C}^L : \boldsymbol{u}(j) = \mathbf{a}(j)e^{it_j}, t_j \in [0, 2\pi), \text{ for all } j = 1, \dots, L\}$. In particular, $\mathbb{T}_{\mathbf{1}} = \otimes^L \mathbb{T}_1$. For $R > 0$, $B_R(\boldsymbol{z}_0) \subset \mathbb{C}^L$ is the ball centered at $\boldsymbol{z}_0 \in \mathbb{C}^L$ with radius R, that is, $B_R(\boldsymbol{z}_0) := \{\boldsymbol{z} \in \mathbb{C}^L; \|\boldsymbol{z} - \boldsymbol{z}_0\| \le R\}$. We also define the set of grids with size $L \in \mathbb{N}$ and length scale $r > 0$ as

$$D^{L\times L} := \{r(\alpha, \beta)\}_{\alpha,\beta=0}^{L-1}.$$

2.1. **The mathematical framework of the ptychography experiment.** In a ptychography experiment, an object of interest is illuminated by a coherent beam, and the resulting diffraction pattern intensity is discretized by a pixellated camera. Numerically, the illuminated portion of the object is discretized to enable fast numerical methods, such approximation is a valid representation of the physical experiment when the illumination function is smaller than the maximum bandwidth allowed by detector. We refer to [51] to situations when these conditions are not strictly satisfied.

For the purpose of this paper, an object of interest is discretized as a $n \times n$ matrix and defined as $\psi : D^{n\times n} \to \mathbb{C}$, where $n \in \mathbb{N}$. For simplicity here we only consider the square matrix case and a uniform discretization rate in both axes. A more general setup is possible with a more heavy notation. Take a two dimensional small beam with known distribution is discretized as a $m \times m$ matrix, where $m < n$, denoted as ω. ω is the kernel function associated with the lens we use in the experiment. We can view the matrix ω as a complex valued function defined on $D^{m\times m}$ so that its value on $r(\alpha-1, \beta-1)$ is $\omega(\alpha, \beta)$, where $\alpha, \beta = 1, \dots, m$. Define

$$D_\omega^{m\times m} := \{r(\alpha-1, \beta-1) \in D^{m\times m} : \omega(\alpha, \beta) \neq 0\},$$

which is the *support of* ω and similarly the support of ψ, $D_\psi^{n\times n}$.

In the experiment, we move the lens around the sample, illuminate $K > 1$ subregions and obtain K diffraction images. Please see Figure 1 for reference. To express this experimental procedure in mathematical form, denote $\iota_{\boldsymbol{x}}$ to be the embedding of $D^{m\times m}$ onto $D^{n\times n}$ so that the left upper corner of $D^{m\times m}$ is located in $\boldsymbol{x} \in D^{n\times n}$, that is, $\iota_{\boldsymbol{x}}(\boldsymbol{r}) = \boldsymbol{x} + \boldsymbol{r}$, where $\boldsymbol{r} \in D^{m\times m}$. Also denote $\mathcal{F}$ to be the 2D DFT operator, that is, $(\mathcal{F}f)(\boldsymbol{q}) = \sum_{\boldsymbol{r}} e^{i\boldsymbol{q}\cdot\boldsymbol{r}} f(\boldsymbol{r})$ when $f \in \mathbb{C}^{L\times L}$ indexed by $\boldsymbol{r}$. Then, the chosen *raster points* are denoted as $\boldsymbol{x}_i \in D^{n\times n}$, where $i = 1, \dots, K$. With these raster points, the experimenter collects a sequence of K diffraction images $\mathbf{a}_{(i)}$ of size $m \times m$, $i = 1, \dots, K$ associated with ψ restricted to $\iota_{\boldsymbol{x}_i}(D^{m\times m})$ by

$$\begin{aligned} &\mathbf{a}_{(i)}(\boldsymbol{q}) = \left|\mathcal{F}(\omega \circ \psi_{(i)})(\boldsymbol{q})\right|, \\ &\boldsymbol{r} = r(\mu, \nu), \quad \boldsymbol{q} = \frac{2\pi}{r}(\mu, \nu), \quad \mu, \nu \in \{0, \dots, m-1\}. \end{aligned}$$

where $\psi_{(i)} : D^{m\times m} \to \mathbb{C}$ is the object over the subregion $\iota_{\boldsymbol{x}_i}(D^{m\times m})$ satisfying $\psi_{(i)}(\boldsymbol{r}) := \psi(\iota_{\boldsymbol{x}_i}(\boldsymbol{r}))$ for all $\boldsymbol{r} \in D^{m\times m}$. We call $\mathcal{X}_K := \{\boldsymbol{x}_i\}_{i=1}^K$ the *illumination scheme*. In this paper, K is assumed to be fixed. We make the following assumption about the illumination scheme:

Assumption 2.1. The chosen illumination scheme $\mathcal{X}_K$ satisfies the following two conditions

(1) $\boldsymbol{x}_i \neq \boldsymbol{x}_j$ for all $i \neq j$;
(2) $\mathcal{X}_K$ is ordered so that $\cup_{i=1}^l \iota_{x_i}(D_\omega^{m\times m}) \subsetneq \cup_{i=1}^{l+1} \iota_{x_i}(D_\omega^{m\times m})$, where $l = 1, \dots K-1$, $\cup_{i=1}^{K-1} \iota_{x_i}(D_\omega^{m\times m}) \subsetneq D^{n\times n}$ and $\cup_{i=1}^l \iota_{x_i}(D_\omega^{m\times m}) = D^{n\times n}$;
(3) For each i_1, there exists i_2 so that $\iota_{\boldsymbol{x}_{i_1}}(D_\omega^{m\times m}) \cap \iota_{\boldsymbol{x}_{i_2}}(D_\omega^{m\times m}) \neq \emptyset$.

The third assumption essentially says that each given pixel $\boldsymbol{x}$ is covered by at least two subregions so that there is a channel for these subregions "exchange information".

Build an undirected graph $\mathbb{G}_\psi$ so that its vertices are points in $D_\psi^{n\times n}$ and an edges is formed if the pair of vertices, $r(i, j), r(i-1, j) \in D^{n\times n}$, or $r(i, j), r(i, j-1) \in D^{n\times n}$, simultaneous exist in $D_\psi^{n\times n}$, for all $i, j = 1, \dots n-1$. We call ψ connected if $\mathbb{G}_\psi$ is connected. Suppose this graph is composed of $J \ge 1$ connected subgraphs so that the i-th subgraph has vertices $D_{\psi,i}^{n\times n}$, where $i = 1, \dots, J$. Viewing each subgraph as an object, with a given illumination scheme $\mathcal{X}_K$, we build a new graph $\mathbb{G}_{\mathcal{X}_K}$ on it by taking

these objects as vertices and putting an edge between $D^{n\times n}_{\psi,i}$ and $D^{n\times n}_{\psi,j}$ if there exist $\boldsymbol{x}_k \in \mathcal{X}_K$ so that $\iota_{\boldsymbol{x}_k}(D^{m\times m}_\omega) \cap D^{n\times n}_{\psi,i} \neq \emptyset$ and $\iota_{\boldsymbol{x}_k}(D^{m\times m}_\omega) \cap D^{n\times n}_{\psi,j} \neq \emptyset$. In other words, for two connected components, there exists an illumination window mounting on them so that the phase information of each connected component can be exchanged. We call the sample ψ *connected with respect to* $\mathcal{X}_K$ if $\mathbb{G}_{\mathcal{X}_K}$ is connected and for $k \in \mathcal{I}_{j,\psi} := \left\{i;\, \iota_{\boldsymbol{x}_i}(D^{m\times m}_\omega) \cap D^{n\times n}_{j,\psi} \neq \emptyset\right\}$, these exists $l \in \mathcal{I}_{j,\psi}$ so that $\iota_{\boldsymbol{x}_k}(D^{m\times m}_\omega) \cap \iota_{\boldsymbol{x}_l}(D^{m\times m}_\omega) \neq \emptyset$. That is, for each connected component $D^{n\times n}_{\psi,i}$, each illumination window in $\mathcal{I}_{j,\psi}$ has an overlapping with some other illumination window in $\mathcal{I}_{j,\psi}$ so that the phase information can be exchanged. Note that if $\mathbb{G}_{\mathcal{X}_K}$ is not connected, then we can view the ptychography imaging problem as two or more subproblems, and solve the problem one by one.

Assumption 2.2. Given $\mathcal{X}_K$, the object of interest ψ is connected with respect to $\mathcal{X}_K$.

To simplify the notation, we further assume that $D^{m\times m}_\omega = D^{m\times m}$ and all discussions in this paper can be easily adapted to $D^{m\times m}_\omega \subsetneqq D^{m\times m}$.

To express the experiment in a compact format, the unknown object $\psi \in \mathbb{C}^{n\times n}$ is also represented as a complex vector $\psi^\vee \in \mathbb{C}^{n^2}$. To take care of the relation between the indices of the matrix form and vector form, for $L > 0$, we define the bijective maps $\mathsf{r}_L : r\mathbb{Z}^2_L \to \mathbb{Z}_{L^2}$ and $\mathsf{q}_n : \frac{2\pi}{r}\mathbb{Z}^2_L \to \mathbb{Z}_{L^2}$ as

$$\mathsf{r}_L : r(\alpha,\beta) \mapsto \alpha L + \beta + 1$$
$$\mathsf{q}_L : \frac{2\pi}{r}(\alpha,\beta) \mapsto \alpha L + \beta + 1,$$

where $\alpha, \beta = 0, \ldots, L-1$. In other words, the mapping r_L is used to convert the index when we rewrite the spatial data, and q_L is used to convert the index when we work with the data in the Fourier domain. For example, for $\boldsymbol{r} = r(\mu,\nu)$, $\mathsf{r}_m(\boldsymbol{r}) = \mu m + \nu + 1$, $\psi^\vee(l) = \psi(\mathsf{r}_n^{-1}(l))$.

Define the *illumination operator* $\mathbf{Q}_{(i)} : \mathbb{C}^{n^2} \to \mathbb{C}^{m^2}$ associated with the raster point $\boldsymbol{x}_i$ as

$$\mathbf{Q}_{(i)}\psi^\vee := (w \circ \psi_{(i)})^\vee =: \boldsymbol{z}_{(i)} \in \mathbb{C}^{m^2}.$$

To a bit abuse the notation, in the following we use the same notation $\mathbf{a}_{(i)} := |\boldsymbol{z}_{(i)}| \in \mathbb{R}^{m^2}_+$.

To express $\mathbf{Q}_{(i)}$ is the matrix form, we need the following definitions. Define the *restriction matrix* $\mathbf{R}$, which is of size $m^2 \times n^2$ so that $\mathbf{R}(i,i) = 1$ for all $i = 1, \ldots, m^2$ and 0 otherwise. Also define the translation matrix $\mathbf{T}_{\boldsymbol{x}}$ which circularly translates $\boldsymbol{x} \in D^{n\times n}$ to $r(0,0) \in D^{n\times n}$. Thus, we have

$$\mathbf{Q}_{(i)} = \mathrm{diag}(w^\vee)\mathbf{R}\mathbf{T}_{\boldsymbol{x}_i} \in \mathbb{C}^{m^2\times n^2}. \tag{1}$$

With these notations, the relationship between the diffraction measurements collected in a ptychography experiment and ψ can be represented compactly as

$$\mathbf{a} = |\mathbf{F}\boldsymbol{z}|, \quad \boldsymbol{z} = \mathbf{Q}\psi^\vee, \tag{2}$$

or $\mathbf{a} = |\mathbf{F}\mathbf{Q}\psi^\vee|$, where

$$\mathbf{a} := \begin{bmatrix} \mathbf{a}_{(1)} \\ \vdots \\ \mathbf{a}_{(K)} \end{bmatrix} \in \mathbb{R}^{Km^2}, \quad \mathbf{F} := \begin{bmatrix} F & \ldots & 0 \\ \vdots & \vdots & \vdots \\ 0 & \ldots & F \end{bmatrix} \in \mathbb{C}^{Km^2\times Km^2}$$
$$\boldsymbol{z} := \begin{bmatrix} \boldsymbol{z}_{(1)} \\ \vdots \\ \boldsymbol{z}_{(K)} \end{bmatrix} \in \mathbb{C}^{Km^2}, \quad \mathbf{Q} := \begin{bmatrix} \mathbf{Q}_{(1)} \\ \vdots \\ \mathbf{Q}_{(K)} \end{bmatrix} \in \mathbb{C}^{Km^2\times n^2},$$

where F is the associated 2D DFT matrix when we write everything in the stacked form, that is, F is a $m^2 \times m^2$ matrix satisfying $F_{lk} = \boldsymbol{e}_l^T F \boldsymbol{e}_k = e^{i\mathsf{q}_m^{-1}(l-1)\cdot\mathsf{r}_m^{-1}(k-1)}$. The objective of the ptychographic reconstruction problem is to find ψ given $\mathbf{a}$ and the form (2).

3. The alternating projection algorithm and its global convergence

In this section, we study the convergence of the alternating projection (AP) algorithm. In general, given an object $\psi_0 \in \mathbb{C}^N$ and a frame $\{\boldsymbol{f}_i\}_{i=1}^M \subset \mathbb{C}^N$ so that $M \geq N$. Denote S to be a $M \times N$ matrix with the i-th row is $\boldsymbol{f}_i^*$. The *phase retrieval problem* we might ask in this setup is the following. Given

$$\mathbf{a} = |\mathsf{S}\psi_0|,$$

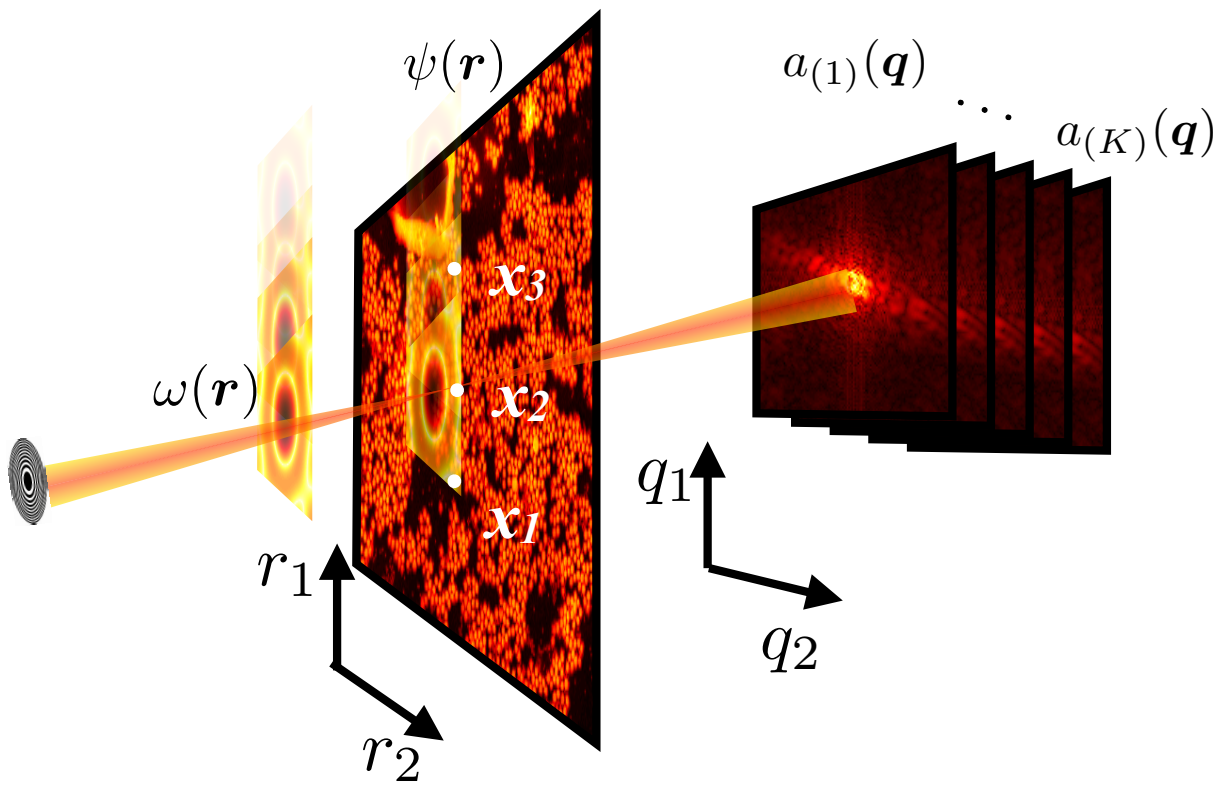


FIGURE 1. Experimental geometry in ptychography: an unknown sample with transmission $\psi(\boldsymbol{r})$ is rastered through an illuminating beam $\omega(\boldsymbol{r})$, and a sequence of diffraction measurements $|\mathbf{a}_{(i)}|^2$ are recorded on an area detector as the sample is rastered around. The point-wise product between illuminating function and sample, $\boldsymbol{z}_{(i)}(\boldsymbol{r}) := \omega(\boldsymbol{r})\psi(\boldsymbol{r}+\boldsymbol{x}_i)$, is related to the measurement by a Fourier magnitude relationship $\mathbf{a}_{(i)} = \left|F\boldsymbol{z}_{(i)}\right|$.

is it possible to recover ψ_0 from $\mathbf{a}$? Note that we have two pieces of information about the phase retrieval problem – the solution has the amplitude $\mathbf{a}$ and is located on the range of S, which is denoted as R_{S}. That is, the solution $\mathsf{S}\psi_0$ exists in $\mathbb{T}_{\mathbf{a}} \cap R_{\mathsf{S}}$. Thus, a popular approach to solve the ptychography problem by finding a vector $\widetilde{\boldsymbol{z}} \in \mathbb{C}^M$ such that

$$\left\{ \begin{array}{l} \|(I-P_{\mathsf{S}})\widetilde{\boldsymbol{z}}\| = 0 \\ \|(I-P_{\mathbf{a}})\widetilde{\boldsymbol{z}}\| = \||\widetilde{\boldsymbol{z}}| - \mathbf{a}\| = 0 \end{array} \right. \tag{3}$$

are both satisfied, where $P_{\mathsf{S}} : \mathbb{C}^M \to \mathbb{C}^M$, referred to as the *phase correction operator*, projects a complex vector to R_{S}, that is,

$$P_{\mathsf{S}} := \mathsf{S}(\mathsf{S}^*\mathsf{S})^{-1}\mathsf{S}^* \in \mathbb{C}^{M\times M},$$

and $P_{\mathbf{a}} : \mathbb{C}^M \to \mathbb{C}^M$, referred to as the *amplitude correction operator*, is

$$P_{\mathbf{a}}\boldsymbol{z} = \mathbf{a}\frac{\boldsymbol{z}}{|\boldsymbol{z}|}\chi_{\boldsymbol{z}} + \mathbf{a}(1-\chi_{\boldsymbol{z}}).$$

Note that $(\mathsf{S}^*\mathsf{S})^{-1}$ exists by the assumption of S and $P_{\mathbf{a}}$ substitutes the amplitude of $\boldsymbol{z}(j)$ by $\mathbf{a}(j)$ and preserve the phase information. Once we find the optimal solution, the object of interest ψ_0 is estimated by

$$\widetilde{\boldsymbol{\psi}}_0 := (\mathsf{S}^*\mathsf{S})^{-1}\mathsf{S}^*\widetilde{\boldsymbol{z}} \in \mathbb{C}^N.$$

We start from recalling the commonly applied *AP algorithm* aiming to solve (3), and then we show the convergence of the AP algorithm to the unique solution under some proper conditions. In the AP

algorithm, the optimization problem (3) is tackled by the following iterative scheme

$$\zeta^{(\ell+1/2)} := P_{\mathbf{a}}\zeta^{(l)}, \quad \zeta^{(\ell+1)} = P_{\mathsf{S}}\zeta^{(\ell+1/2)}.$$

It is easy to verify that $|\zeta^{(\ell+1/2)}| = \mathbf{a}$ for all $\ell \in \mathbb{N}$ and $P_{\mathbf{a}}$ is a projection onto $\mathbb{T}_{\mathbf{a}}$ in the sense that

$$P_{\mathbf{a}}\zeta = \operatorname*{argmin}_{\bar{\zeta}\in\mathbb{T}_{\mathbf{a}}} \|\bar{\zeta} - \zeta\|. \tag{4}$$

This optimization step is non-linear in nature. Indeed, the constrained space we are searching for the $\bar{\zeta} \in \mathbb{C}^M$ closest to ζ is characterized by its Fourier amplitude. Note that at the first glance, $P_{\mathbf{a}}$ behaves like a dilation operator in $\mathbb{C}^M$; however, since the dilation might be different from entry to entry, it is nonlinear in nature. On the other hand, P_{S} linearly projects $\zeta^{(\ell+1/2)}$ to the range of S, denoted as R_{S}. The algorithm can be illustrated in Figure 2.

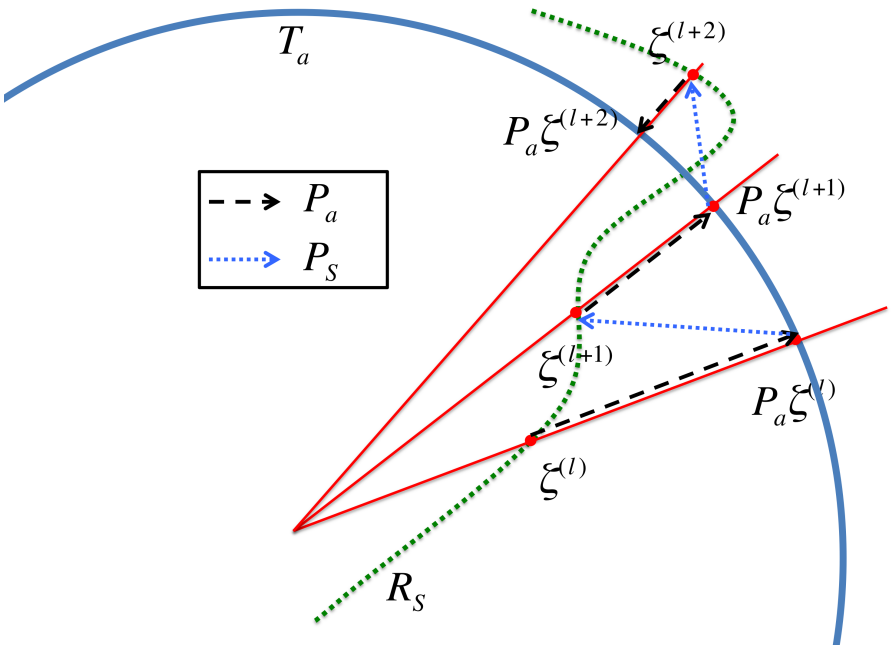


FIGURE 2. The summary of the Lemma 3.7. The lengths of the black dashed arrows associated with $P_{\mathbf{a}}$ decrease during the iteration and the lengths of the blue dashed arrows associated with P_{S} decrease, too. However, $\|\zeta^{(l)} - \zeta^{(l-1)}\|$ may not decrease. R_{S} is illustrated as a curve to emphasize the nonlinear nature of the $P_{\mathbf{a}}$ map.

The main purpose of the AP algorithm is finding the solution $\mathsf{S}\psi_0$, which is located on the set $R_{\mathsf{S}} \cap \mathbb{T}_{\mathbf{a}}$. In order to characterize this set, we introduce some notations and quote the theorems from [2]. Note that for the frame S, we have the following mapping:

$$\mathbb{M}^{\mathsf{S}} : \mathbb{C}^N \to \mathbb{C}^M, \qquad \mathbb{M}^{\mathsf{S}}(\boldsymbol{z}) = \mathsf{S}\boldsymbol{z},$$

where $\boldsymbol{z} \in \mathbb{C}^N$. We thus can view the range of the $\mathbb{M}^{\mathsf{S}}$ as a complex N-dimensional subspace of $\mathbb{C}^M$. Thus, from the frame theory view point [2], S determines a point of the fiber bundle $\mathsf{F}[N, M; \mathbb{C}]$, whose base manifold is the complex Grassmannian manifold $\mathsf{Gr}(N, M; \mathbb{C})$ with fiber $\mathsf{GL}(N, \mathbb{C})$. The phase retrieval problem is directly related to the following nonlinear map:

$$\mathbb{M}_a^{\mathsf{S}} : \mathbb{C}^N/\mathbb{T}_1 \to \mathbb{C}^M, \qquad \mathbb{M}_a^{\mathsf{S}}(\boldsymbol{z}) = |\mathbb{M}^{\mathsf{S}}(\boldsymbol{z})| = \sum_{k=1}^{M} |\boldsymbol{f}_k^* \boldsymbol{z}| \boldsymbol{e}_k, \tag{5}$$

where $\boldsymbol{z} \in \mathbb{C}^N$ and the subscript a means taking the absolute value. That is, we only have the amplitude information of the coordinates of the signal $\boldsymbol{z}$ with related to the frame but the phase information is lost.

In the following, by *generic* we mean that there is a Zariski open set in the real algebraic variety $\mathsf{Gr}(N, M, \mathbb{C})$ so that the result holds for all frames of the associated linear subspace. Note that we only discuss the genericity of $\mathsf{Gr}(N, M; \mathbb{C})$ since the following proposition

Proposition 3.1 (the complex version of Proposition 2.1 [2])**.** For any two frames S and $\widetilde{\mathsf{S}}$ that have the same range of coefficients, $\mathbb{M}_a^{\mathsf{S}}$ is injective if and only if $\mathbb{M}_a^{\widetilde{\mathsf{S}}}$ is injective.

The main theorem in [2] we count on is the following.

Theorem 3.2 (Theorem 3.3 [2])**.** If $M \geq 4N-2$, then $\mathbb{M}_a^{\mathsf{S}}$ is injective for a generic frame S.

In conclusion, we know that generically the solution to the phase retrival problem is unique when $M \geq 4N-2$, and thus solving the problem is possible. As useful as the Theorems, however, they do not answer the practical question – how does the phase optimization algorithm lead to the solution? In particular, the operator $(\mathbb{M}_a^{\mathsf{S}})^{-1}$ is unclear to us. In this section, we analyze the convergence behavior of the AP algorithm, which leads to $(\mathbb{M}_a^{\mathsf{S}})^{-1}$. We mention that the uniqueness result of the phase retrieval problem in a different setup, in particular, when the signal of interest is real-valued with dimension higher than 2 and the frame is the oversampling Fourier transform, has been reported in [13, 6, 39, 58]. In such set-up, the set of non-unique solutions is of measure zero. However, such structures do exist in nature [55].

Notice that while the operator $\mathbb{M}_a^{\mathsf{S}}$ is defined on $\mathbb{C}^N/\mathbb{T}_1$, where the global constant phase difference is moduled out, the inverse $(\mathbb{M}_a^{\mathsf{S}})^{-1}$ does not distinguish between the global constant phase difference. Thus, when $M \geq 4N-2$ and R_{S} generic, given $\psi_0 \in \mathbb{C}^N$ and $\mathbf{a} = |\mathsf{S}\psi_0|$, we define the *solution set* as

$$S_{\mathbf{a}} := \{e^{it}\mathsf{S}\psi_0 : t \in [0, 2\pi)\} = R_{\mathsf{S}} \cap \mathbb{T}_{\mathbf{a}} \cong \mathbb{T}_1,$$

where the second equality holds due to the above Theorem.

Before proceeding, we have some immediate consequences of the Theorem.

Lemma 3.3. When $M \geq 4N-2$ and R_{S} generic, for all $\boldsymbol{z}, \boldsymbol{w} \in R_{\mathsf{S}}$ and $\boldsymbol{z} \neq c\boldsymbol{w}$ for $c \in \mathbb{T}_1$, then $|\boldsymbol{z}| \neq |\boldsymbol{w}|$. Moreover, not all $\mathbb{T}_{\mathbf{a}}$, where $\mathbf{a} \in \mathbb{R}_+^M$, intersects R_{S}.

Proof. The first claim is immediate from Theorem 3.2. Note that when R_{S} and $\mathbb{T}_{\mathbf{a}}$ intersect, it means that $\mathbf{a}$ comes from $\mathbb{M}_a^{\mathsf{S}}$. Also note that the mapping $\mathbb{M}_a^{\mathsf{S}} : \mathbb{C}^N \to \mathbb{R}_+^M$ can be viewed as an embedding of $\mathbb{C}^N$ into $\mathbb{C}^M$ followed by a nonlinear mapping from R_{S} to $\mathbb{R}_+^M$. Here the nonlinear mapping is 1-1 when $M \geq 4N-2$ by Theorem 3.2. By counting the dimension, we know that the mapping $\mathbb{M}_a^{\mathsf{S}}$ can not be onto, and hence the first claim is proved. □

We conclude from this Lemma that for $\boldsymbol{z} \in R_{\mathsf{S}}$ with $\boldsymbol{b} = |\boldsymbol{z}|$, there exists a unique phase $\boldsymbol{\phi}^{\boldsymbol{b}} \in \mathbb{T}_{\mathbf{1}}$ so that $\boldsymbol{z} = \boldsymbol{b}e^{i(t+\boldsymbol{\phi}^{\boldsymbol{b}})}$ for some $t \in [0, 2\pi)$. Here the subscript $\boldsymbol{b}$ in $\boldsymbol{\phi}^{\boldsymbol{b}}$ indicates the dependence of the phase on the amplitude $\boldsymbol{b}$. We mention that if we replace the conditions in Lemma 3.3, we have the following corresponding result.

Lemma 3.4. When $M \geq 4N-2$ and R_{S} generic, for all $\boldsymbol{z}, \boldsymbol{w} \in R_{\mathsf{S}}$ and $\boldsymbol{z} \neq r\boldsymbol{w}$ for $r \in \mathbb{R}^1$, then $\boldsymbol{z} \neq c\boldsymbol{w}$ for any $c \in \mathbb{T}_1$.

Proof. This also comes from Theorem 3.1. The only difference is interchanging $\mathbb{T}_1$ and $\mathbb{R}^1$ in the proof of Theorem 3.1 and note that they both have real dimension 1. □

To show the convergence result, we start from introducing the following *stagnation set (or the fixed points) of the AP algorithm when the given data is* $\mathbf{a}$:

$$\Theta_{\mathbf{a}}^{\mathrm{AP}} := \{\zeta \in R_{\mathsf{S}} : P_{\mathsf{S}}P_{\mathbf{a}}\zeta = \zeta\} \subset R_{\mathsf{S}} \cap B_{\|\mathbf{a}\|}(0). \tag{6}$$

Note that $\Theta_{\mathbf{a}}^{\mathrm{AP}}$ is a subset of $B_{\|\mathbf{a}\|}(0)$ simply because P_{S} is a projection operator. Please see Figure 3 for illustration.

We can compare the definition of the stagnation set with the solution set of the phase retrieval problem. Clearly the solution set $S_{\mathbf{a}} \subset \Theta_{\mathbf{a}}^{\mathrm{AP}}$. The stagnation set reflects the fact that $P_{\mathbf{a}}\zeta - \zeta \neq 0$ does not imply $\zeta \neq P_{\mathsf{S}}P_{\mathbf{a}}\zeta$, that is, when $\zeta = P_{\mathsf{S}}P_{\mathbf{a}}\zeta$, ζ may or may not be the solution. We have the following quantification of the stagnation set, which says that when $M \geq 4N-2$, the stagnation set is precisely the solution set $S_{\mathbf{a}}$.

Theorem 3.5. $\Theta_{\mathbf{a}}^{\mathrm{AP}} = R_{\mathsf{S}} \cap \mathbb{T}_{\mathbf{a}}$. In particular, when $M \geq 4N-2$ and R_{S} generic, $\Theta_{\mathbf{a}}^{\mathrm{AP}} = S_{\mathbf{a}}$.

Proof. We first claim that $I - P_{\mathbf{a}}$ is onto. For any given $\zeta \in \mathbb{C}^M$ we are able to define $\eta \in \mathbb{C}^M$ so that

$$\eta(k) = \begin{cases} (|\zeta(k)| + \mathbf{a}(k))\frac{\zeta(k)}{|\zeta(k)|} & \text{when } \zeta(k) \neq 0 \\ \mathbf{a}(k)u & \text{when } \zeta(k) = 0 \end{cases},$$

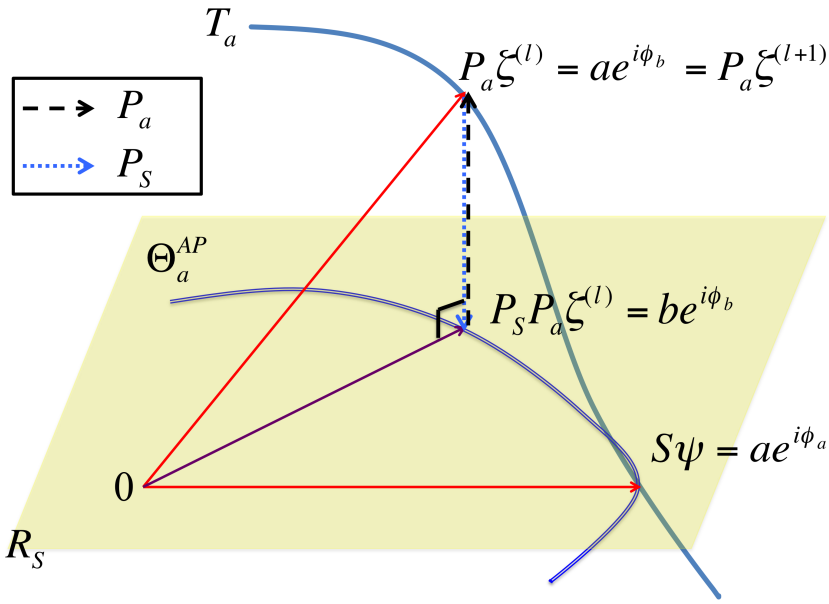


FIGURE 3. The stagnation set $\Theta_{\mathbf{a}}^{\mathrm{AP}}$. $T_{\mathbf{a}}$ is illustrated as a curve to emphasize the non-linear nature of the $P_{\mathbf{a}}$ map.

where u is randomly chosen from $\mathbb{T}_1$. By this definition, we know $\eta - P_{\mathbf{a}}\eta = \zeta$ and hence $I - P_{\mathbf{a}}$ is onto. Note that when there is an entry 0 in ζ, $I - P_{\mathbf{a}}$ is not one-to-one, and the more entries of ζ are zero, the larger the dimension of the set $(I - P_{\mathbf{a}})^{-1}\zeta$ is. Next, we claim that when all entries of ζ are non-zero, $(I - P_{\mathbf{a}})^{-1}(\zeta)$ is a point, that is, $I - P_{\mathbf{a}}$ is one to one on the set $\{(I - P_{\mathbf{a}})^{-1}(\zeta); \zeta(i) \neq 0, i = 1, \ldots, M\}$. Suppose $\eta_1 \neq \eta_2$ so that $\zeta = (I - P_{\mathbf{a}})\eta_1 = (I - P_{\mathbf{a}})\eta_2$. Since $I - P_{\mathbf{a}}$ operator acts on $\mathbb{C}^M$ entry-wisely, we may focus on the k-th entry so that $\eta_1(k) \neq \eta_2(k)$. Then, since $(|\eta_1(k)| - \mathbf{a}(k))\frac{\eta_1(k)}{|\eta_1(k)|} \neq 0$, we clearly have $(|\eta_1(k)| - \mathbf{a}(k))\frac{\eta_1(k)}{|\eta_1(k)|} \neq (|\eta_2(k)| - \mathbf{a}(k))\frac{\eta_2(k)}{|\eta_2(k)|}$, which is absurd. Indeed, since $\eta_1(k) \neq \eta_2(k)$, we know either $|\eta_1(k)| - \mathbf{a}(k) \neq |\eta_2(k)| - \mathbf{a}(k)$ or $\frac{\eta_1(k)}{|\eta_1(k)|} \neq \frac{\eta_2(k)}{|\eta_2(k)|}$. The claim is thus proved.

It is clear that $R_{\mathsf{S}} \cap \mathbb{T}_{\mathbf{a}} \subset \Theta_{\mathbf{a}}^{\mathrm{AP}}$ since for all $\boldsymbol{z} \in R_{\mathsf{S}} \cap \mathbb{T}_{\mathbf{a}}$, $\boldsymbol{z} = P_{\mathsf{S}}P_{\mathbf{a}}\boldsymbol{z}$. We now prove the other direction by contradiction. Suppose $\eta \in \Theta_{\mathbf{a}}^{\mathrm{AP}}$ is located on R_{S} but not on $\mathbb{T}_{\mathbf{a}}$. Then, denote $\zeta := \eta - P_{\mathbf{a}}\eta$, which is non-zero by the assumption. Thus, when all entries of ζ are non-zero, $\eta^*(\eta - P_{\mathbf{a}}\eta) = 0$ implies that $\sum_{k=1}^{M}(|\zeta(k)| + \mathbf{a}(k))|\zeta(k)| = 0$, which is only possible if $\zeta(k) = 0$ for all k. However, we know that by assumption $\zeta \neq 0$, so it is absurd. Similarly, when there are some 0 entries in ζ, say, $\zeta(k) = 0$, then the k-th entry does not play a role in $\eta^*(\eta - P_{\mathbf{a}}\eta) = 0$, while the non-zero entries in ζ leads to the same contradiction. □

We emphasize that this Theorem does not imply the convergence of the AP algorithm since there might be other kinds of stagnation in the alternating algorithm. We need more to show the convergence of the AP algorithm.

Lemma 3.6. (a) For $\zeta \neq 0$ and $w \in \mathbb{T}_{\mathbf{a}}$, we have

$$\|P_{\mathbf{a}}\zeta - \zeta\| \leq \|w - \zeta\|,$$

where the equality holds when $w = P_{\mathbf{a}}\zeta$.

(b) For $w \in \mathbb{T}_{\mathbf{a}}$ and $z \in R_{\mathsf{S}}$, we have

$$\|P_{\mathsf{S}}w - w\| \leq \|z - w\|,$$

where the equality holds when $z = P_{\mathsf{S}}w$.

(c) For all nonzero $\zeta \in R_{\mathsf{S}}$, $P_{\mathbf{a}}\zeta$ is not perpendicular to R_{S}.

(d) When $M \geq 4N - 2$ and R_{S} generic, given $\zeta \in R_{\mathsf{S}}$, $P_{\mathbf{a}}\zeta \in R_{\mathsf{S}}$ holds if and only if $\zeta \in S_{\mathbf{a}}$.

(e) All possible initial values $\zeta^{(0)}$ with non-zero entries can be parametrized by a $(2N-1)$-dim real sphere embedded in R_{S}. In particular, given $z \in R_{\mathsf{S}}$ so that all entries are not zero and $rz \notin S_{\mathbf{a}}$ for all $r \in \mathbb{R}^+$, the phase of z is different from the phase of all $w \in S_{\mathbf{a}}$.

Proof. To prove (a), denote $\zeta = (b_i e^{i\theta_i})_{i=1}^M \in \mathbb{C}^M$ and $w = (a_i e^{i\phi_i})_{i=1}^M \in \mathbb{C}^M$, where $b_i \geq 0$ and $\theta_i, \phi_i \in [0, 2\pi)$. Then by definition $P_{\mathbf{a}}\zeta = (a_i e^{i\theta_i})_{i=1}^M$. Thus, $\|P_{\mathbf{a}}\zeta - \zeta\| = \sqrt{\sum |a_i - b_i|^2}$ and $\|w - \zeta\| = \sqrt{\sum |a_i - b_i e^{i(\theta_i - \phi_i)}|^2}$, which leads to the result since $|a_i - b_i| < |a_i - b_i e^{i(\theta_i - \phi_i)}|$. Note that the equality holds when $\theta_i = \phi_i$ for all i.

The proof of (b) is directly from the fact the P_{S} is a projection operator.

For (c), denote $\zeta = (b_i e^{i\theta_i})_{i=1}^M \in R_{\mathsf{S}}$, where $b_i \in \mathbb{R}^+$ and $\theta_i \in [0, 2\pi)$. Then by definition $P_{\mathbf{a}}\zeta = (a_i e^{i\theta_i})_{i=1}^M$. Then it is clear that $\langle P_{\mathbf{a}}\zeta, \zeta\rangle > 0$, which shows the claim.

The statement (d) is direct from Theorem 3.2.

The show the statement (e), note that R_{S} can be viewed as a real vector space of dimension $2N$. If $z, w \in R_{\mathsf{S}}$ so that $w = rz$, where $r \in \mathbb{R}^+$, by definition we have $P_{\mathbf{a}}z = P_{\mathbf{a}}w$. In other words, each "real positive ray" is associated with an initial value. For the other part, if the phase of z is the same as the phase of $w \in S_{\mathbf{a}}$, we know $P_{\mathbf{a}}z = w$, which means that there exists $r > 0$ so that $rz = w$, which is absurd. □

Lemma 3.7. The following bounds hold:

$$\|(P_{\mathbf{a}} - I)\zeta^{(l)}\| \leq \|(P_{\mathbf{a}} - I)\zeta^{(l-1)}\|$$
$$\|(P_{\mathsf{S}} - I)\zeta^{(l+1/2)}\| \leq \|(P_{\mathsf{S}} - I)\zeta^{(l-1/2)}\|.$$

When $M \geq 4N-2$ and $\zeta^{(0)} \notin \Theta_{\mathbf{a}}^{\mathrm{AP}}$ and R_{S} generic, there exist $\alpha_l < 1$ and $\beta_l < 1$, $l \in \mathbb{N}$ so that

$$\|(P_{\mathbf{a}} - I)\zeta^{(l)}\| = \alpha_l \|(P_{\mathbf{a}} - I)\zeta^{(l-1)}\| \tag{7}$$

$$\|(P_{\mathsf{S}} - I)\zeta^{(l+1/2)}\| = \beta_l \|(P_{\mathsf{S}} - I)\zeta^{(l-1/2)}\|. \tag{8}$$

Here $\{\alpha_l, \beta_l\}$ depend on $\zeta^{(1)}$, S and $\mathbf{a}$. Moreover, if we denote $\zeta^{(l)} = (b_k^{(l)} e^{i\boldsymbol{\phi}_k^{(l)}})_{k=1}^M$, where $b_k^{(l)} \in \mathbb{R}_+$ and $\boldsymbol{\phi}_k^{(l)} \in [0, 2\pi)$, the following inequality holds:

$$2\sum_{k=1}^M a_k b_k^{(l)} \big(1 - \cos(\boldsymbol{\phi}_k^{(l)} - \boldsymbol{\phi}_k^{(l-1)})\big) < \sum_{k=1}^M (a_k - b_k^{(l-1)})^2 - \sum_{k=1}^M (a_k - b_k^{(l)})^2, \tag{9}$$

Proof. Based on Lemma 3.6, we have the following inequalities. First,

$$\|P_{\mathbf{a}}\zeta^{(l)} - \zeta^{(l)}\| \leq \|P_{\mathbf{a}}\zeta^{(l-1)} - \zeta^{(l)}\| = \|P_{\mathbf{a}}\zeta^{(l-1)} - P_{\mathsf{S}}P_{\mathbf{a}}\zeta^{(l-1)}\|$$

due to Lemma 3.6 (a). Note that when $M \geq 4N-2$, the equality can not hold since $\zeta^{(l)} \neq \zeta^{(l-1)}$ due to Theorem 3.5. Then, by Lemma 3.6 (b), we have

$$\|P_{\mathbf{a}}\zeta^{(l-1)} - P_{\mathsf{S}}P_{\mathbf{a}}\zeta^{(l-1)}\| \leq \|P_{\mathbf{a}}\zeta^{(l-1)} - \zeta^{(l-1)}\|.$$

Again, when $M \geq 4N-2$, the equality can not hold since $\zeta^{(l-1)} \notin \Theta_{\mathbf{a}}^{\mathrm{AP}}$ due to Theorem 3.5. Similarly, by Lemma 3.6 (a), the following inequality holds

$$\|P_{\mathbf{a}}\zeta^{(l-1)} - \zeta^{(l-1)}\| \leq \|P_{\mathbf{a}}\zeta^{(l-2)} - \zeta^{(l-1)}\| = \|\zeta^{(l-3/2)} - P_{\mathsf{S}}\zeta^{(l-3/2)}\|.$$

By iteratively evaluating the above equality, we have (7). Now, since $\zeta^{(l)} = (b_k^{(l)} e^{i\boldsymbol{\phi}_k^{(l)}})_{k=1}^M$, we have

$$\begin{aligned}\|P_{\mathbf{a}}\zeta^{(l-1)} - P_{\mathsf{S}}P_{\mathbf{a}}\zeta^{(l-1)}\|^2 &= \sum_{k=1}^M |a_k - b_k^{(l)} e^{i(\boldsymbol{\phi}_k^{(l)} - \boldsymbol{\phi}_k^{(l-1)})}|^2 \\ &= \sum_{k=1}^M (a_k - b_k^{(l)})^2 + 2\sum_{k=1}^M a_k b_k^{(l)} \big(1 - \cos(\boldsymbol{\phi}_k^{(l)} - \boldsymbol{\phi}_k^{(l-1)})\big)\end{aligned}$$

and

$$\|P_{\mathbf{a}}\zeta^{(l-1)} - \zeta^{(l-1)}\|^2 = \sum_{k=1}^M |a_k - b_k^{(l-1)}|^2.$$

Thus, we have

$$2\sum_{k=1}^{M} a_k b_k^{(l)}\big(1-\cos(\boldsymbol{\phi}_k^{(l)}-\boldsymbol{\phi}_k^{(l-1)})\big) < \sum_{k=1}^{M}(a_k-b_k^{(l-1)})^2 - \sum_{k=1}^{M}(a_k-b_k^{(l)})^2,$$

and hence the proof is done. □

Note that (7) and (8) does not imply

$$\|\zeta^{(l+1)}-\zeta^{(l)}\| \le \|\zeta^{(l)}-\zeta^{(l-1)}\|.$$

Indeed, note that $P_{\mathbf{a}}\zeta^{(l)}-\zeta^{(l+1)}$ is perpendicular to $\zeta^{(l+1)}-\zeta^{(l)}$. Thus we have

$$\begin{aligned}\|(P_{\mathbf{a}}-I)\zeta^{(l)}\|^2 &= \|\zeta^{(l+1)}-\zeta^{(l)}\|^2 + \|(P_{\mathsf{S}}-I)P_{\mathbf{a}}\zeta^{(l)}\|^2\\ \|(P_{\mathbf{a}}-I)\zeta^{(l-1)}\|^2 &= \|\zeta^{(l)}-\zeta^{(l-1)}\|^2 + \|(P_{\mathsf{S}}-I)P_{\mathbf{a}}\zeta^{(l-1)}\|^2,\end{aligned} \tag{10}$$

where when (7) and (8) hold, it is still possible that $\|(P_{\mathsf{S}}P_{\mathbf{a}}-I)\zeta^{(l)}\| > \|(P_{\mathsf{S}}P_{\mathbf{a}}-I)\zeta^{(l-1)}\|$. Please see the numerical section for an example.

Lemma 3.8. When $M \ge 4N-2$ and R_{S} generic, the following conditions are equivalent

(1) AP algorithm converges to the solution;
(2) $\|(P_{\mathbf{a}}-I)\zeta^{(l)}\| \to 0$;
(3) $\|(P_{\mathsf{S}}-I)\zeta^{(l+1/2)}\| \to 0$;
(4) $\|\zeta^{(l+1)}-\zeta^{(l)}\| \to 0$.

Proof. When $M \ge 4N-2$, (1) and (4) are equivalent by Theorem 3.5. When (1) holds, it is clear that (2) and (3) hold. On the other hand, note that when $\|(P_{\mathbf{a}}-I)\zeta^{(l)}\| \to 0$, we have $\|\zeta^{(l+1)}-\zeta^{(l)}\| \to 0$ and $\|(P_{\mathsf{S}}-I)P_{\mathbf{a}}\zeta^{(l)}\| \to 0$ by (10). That is, (2) implies (3) and (4). Finally, since $P_{\mathsf{S}}\zeta^{(l+1/2)} \in R_{\mathsf{S}}$, (3) means $\zeta^{(l+1/2)} \in P_{\mathbf{a}}$ converges to a point located on $R_{\mathsf{S}} \cap T_{\mathbf{a}}$, that is, $\zeta^{(l+1/2)}$ converges to the solution set. Hence we have the fact that (3) implies (1). □

Thus, to study the convergence behavior of the AP algorithm, we may focus on the convergence behavior of $\|(P_{\mathbf{a}}-I)\zeta^{(l)}\|$. We are now ready to show our main theorem – when $M \ge 4N-2$, the AP algorithm converges to the solution.

Theorem 3.9. When $M \ge 4N-2$ and R_{S} generic, the AP algorithm converges to the solution set.

Proof. We will assume that $\zeta^{(0)} \notin S_{\mathbf{a}}$. By Lemma 3.7, we know that α_l and β_l are both less than 1 unless the AP algorithm converges to the solution set in finite steps. So we suppose $\alpha_l < 1$ and $\beta_l < 1$ for all $l \in \mathbb{N}$. We show the proof by considering the following different situations.

It is clear that if $\limsup_{l\to\infty}\alpha_l < 1$, then the AP algorithm converges linearly globally. Indeed, since there exists $l_0 \in \mathbb{N}$ and $\alpha < 1$ so that $\alpha_l \le \alpha$ when $l > l_0$, we have

$$\|P_{\mathbf{a}}\zeta^{(l)}-\zeta^{(l)}\| \le \alpha^{l-l_0}\|P_{\mathbf{a}}\zeta^{(l_0)}-\zeta^{(l_0)}\| \to 0.$$

Note that in this case, $\Pi_{l=1}^{\infty}\beta_l$ is forced to diverge to 0 by (10).

If $\limsup_{l\to\infty}\alpha_l = 1$ so we can not find $\alpha < 1$, there are two possibilities. First, suppose $\liminf_{l\to\infty}\alpha_l \le 1-\epsilon$ for some $\epsilon > 0$, then there exists a subsequence of α_l, denoted as α_{l_k}, where $k \in \mathbb{N}$, so that $\alpha_{l_k} \le 1-\epsilon$. In this case, we still have

$$\|P_{\mathbf{a}}\zeta^{(l)}-\zeta^{(l)}\| \to 0$$

and hence the convergence.

Second, suppose $\liminf_{l\to\infty}\alpha_l = 1$, that is, $\lim_{l\to\infty}\alpha_l = 1$. Clearly the series $p_n := \Pi_{l=1}^{n}\alpha_l$ converges as $n \to \infty$ since $\alpha_l < 1$. If the infinite product $\Pi_{l=1}^{\infty}\alpha_l$ diverges to 0, the AP algorithm converges to the solution, but at a slow rate. However, $\Pi_{l=1}^{\infty}\alpha_l$ may not diverge to 0 since $\Pi_{l=1}^{\infty}\alpha_l$ converges if and only if the series $\sum_{l=1}^{\infty}(1-\alpha_l)$ converges.

We now show that in the AP algorithm, $\Pi_{l=1}^{\infty}\alpha_l$ will diverge to 0. If $\sum_{l=1}^{\infty}(1-\alpha_l)$ converges, that is, $0 < p_\infty = \Pi_{l=1}^{\infty}\alpha_l < 1$, we have

$$\|P_{\mathbf{a}}\zeta^{(l)}-\zeta^{(l)}\| \to p_\infty\|P_{\mathbf{a}}\zeta^{(1)}-\zeta^{(1)}\| =: p(\zeta^{(1)}) > 0, \tag{11}$$

as $l \to \infty$, that is, the AP algorithm will not converge by Lemma 3.8. In this case, $\Pi_{l=1}^{\infty}\beta_l = q_\infty > 0$ must hold by Lemma 3.8, and we have

$$\|(P_{\mathbf{FQ}}-I)P_{\mathbf{a}}\zeta^{(l)}\| \to q_\infty\|(P_{\mathbf{FQ}}-I)P_{\mathbf{a}}\zeta^{(1)}\| =: q(\zeta^{(1)}) > 0, \tag{12}$$

and hence

$$\lim_{l\to\infty}\|\zeta^{(l+1)}-\zeta^{(l)}\|^2 = p^2(\zeta^{(1)})-q^2(\zeta^{(1)})>0.$$

Denote $\zeta^{(l)}=(b_k^{(l)}e^{i\boldsymbol{\phi}_k^{(l)}})_{k=1}^M$. By assumption, (9) and (11) we have:

$$0\le 2\sum_{k=1}^M a_k b_k^{(l)}\big(1-\cos(\boldsymbol{\phi}_k^{(l)}-\boldsymbol{\phi}_k^{(l-1)})\big)<\sum_{k=1}^M(a_k-b_k^{(l-1)})^2-\sum_{k=1}^M(a_k-b_k^{(l)})^2\to 0$$

as $l\to\infty$ since $\sum_{k=1}^M(a_k-b_k^{(l-1)})^2=\|P_{\mathbf{a}}\zeta^{(l-1)}-\zeta^{(l-1)}\|^2\to p(\zeta^{(1)})^2$ and $\sum_{k=1}^M(a_k-b_k^{(l)})^2=\|P_{\mathbf{a}}\zeta^{(l)}-\zeta^{(l)}\|^2\to p(\zeta^{(1)})^2$. On the other hand, by (12) we know

$$\begin{aligned}q(\zeta^{(1)})^2 &= \lim_{l\to\infty}\|(P_{\mathbf{FQ}}-I)P_{\mathbf{a}}\zeta^{(l)}\|^2=\lim_{l\to\infty}\sum_{k=1}^M|a_k-b_k^{(l+1)}e^{i(\boldsymbol{\phi}_k^{(l+1)}-\boldsymbol{\phi}_k^{(l)})}|^2\\ &=\lim_{l\to\infty}\sum_{k=1}^M(a_k-b_k^{(l+1)})^2+\lim_{l\to\infty}2\sum_{k=1}^M a_k b_k^{(l+1)}\big(1-\cos(\boldsymbol{\phi}_k^{(l+1)}-\boldsymbol{\phi}_k^{(l)})\big)=p(\zeta^{(1)})^2.\end{aligned}$$

As a result, we have

$$\lim_{l\to\infty}\|\zeta^{(l+1)}-\zeta^{(l)}\|^2=0,$$

which contradicts Lemma 3.8. Thus, the AP algorithm converges to the solution set as is claimed. □

Note that although the AP algorithm converges, it might converge very slowly, as is shown above.

Remark (Non-convex optimization framework). We mention that the AP algorithm can be studied in the non-convex optimization framework [22]. Given a set of subsets S_i, $i=1,\dots,L$ of a metric space X so that $S:=\cap_{i=1}^L S_i\neq\emptyset$. To find S, we may consider the proposed sequence of successive projections (SOSP) scheme, which successively project the estimator to S_i. Under suitable conditions, the convergence of the SOSP scheme is provided in [22, Theorem 4.3]. Indeed, it says that when the initial value x_0 of the SOSP $\{x_n\}_{n\ge0}$ is a point of attraction [22, Definition 4.4] of an ordered collection of proximial sets in a metric space whose intersection S is not empty, then either $\{x_n\}_{n\ge0}$ converges to a point in S or the set of the cluster points of $\{x_n\}_{n\ge0}$ is a nontrivial continuum in S.

Note that in our AP algorithm setup, the metric space X is a finite dimensional Hilbert space $\mathbb{C}^M$, S_1 is our $\mathbb{T}_{\mathbf{a}}$, and S_2 is our $R_{\mathbf{S}}$. By Lemma 3.6, we know that S_1 and S_2 are Chebychev sets so that the SOSP is unique. When $M\ge 4N-2$, we consider only the generic frame so that the intersection set $S_1\cap S_2$ is a compact set $S_{\mathbf{a}}$ diffeomorphic to $\mathbb{T}_1$. Although it is claimed in [7] that the AP algorithm locally converges, the confirmation of the positive attractor radius and that the initial value is a point of attraction needs to be confirmed. The above argument provides a different approach to show the convergence property of the AP algorithm under the non-convex optimization framework when $M\ge 4N-2$.

4. The Relationship between the AP Algorithm and Optimization

To better understand the AP algorithm, we assume $M\ge 4N-2$ in this section. Define an objective function [69]

$$\rho(\boldsymbol{z}):=\frac{1}{2}\||\boldsymbol{z}|-\mathbf{a}\|^2:=\frac{1}{2}r(\boldsymbol{z})^T r(\boldsymbol{z}),$$

where $\boldsymbol{z}=(z_1,\dots,z_M)^T\in\mathbb{C}^M$, $r:\mathbb{C}^M\to\mathbb{R}^M$ is defined by

$$r(\boldsymbol{z}):=|\boldsymbol{z}|-\mathbf{a}.$$

Note that we take the transpose since $r(\boldsymbol{z})$ is a real vector. The objective function ρ, when restricted on $R_{\mathbf{FQ}}$, gauges how far we are to the solution. Recall that the solution, or the right phase, are located on $R_{\mathbf{FQ}}$. To evaluate the gradient and Hessian of ρ, we prepare the following calculations [42]. First, we evaluate the derivative of $r(\boldsymbol{z})$ with respect to $\boldsymbol{z}$ at $\boldsymbol{z}$:

$$\frac{\partial r}{\partial \boldsymbol{z}}|_{\boldsymbol{z}}=\frac{\partial|\boldsymbol{z}|}{\partial\boldsymbol{z}}|_{\boldsymbol{z}}=\frac{\partial}{\partial\boldsymbol{z}}\begin{pmatrix}|z_1|\\ \vdots\\ |z_M|\end{pmatrix}=\begin{pmatrix}\frac{\partial|z_1|}{\partial z_1} & & 0\\ & \ddots & \\ 0 & & \frac{\partial|z_M|}{\partial z_M}\end{pmatrix}=\frac{1}{2}\mathrm{diag}\frac{\boldsymbol{z}^*}{|\boldsymbol{z}|},$$

where we use the fact that $\frac{\partial |w|}{\partial w} = \frac{w^*}{2|w|}$ when $w \in \mathbb{C}$. Similarly, we evaluate the derivative of $r(\boldsymbol{z})$ with respect to $\overline{\boldsymbol{z}}$ at $\boldsymbol{z}$:

$$\frac{\partial r}{\partial \overline{\boldsymbol{z}}}|_{\boldsymbol{z}} = \frac{1}{2}\text{diag}\frac{\boldsymbol{z}}{|\boldsymbol{z}|}.$$

Thus, by the chain rule we obtain the derivative of $\rho(\boldsymbol{z})$ with respect to $\boldsymbol{z}$ and $\overline{\boldsymbol{z}}$ at $\boldsymbol{z}$:

$$\frac{\partial \rho}{\partial \boldsymbol{z}}|_{\boldsymbol{z}} = \frac{1}{2}\left(\frac{\partial r}{\partial \boldsymbol{z}}|_{\boldsymbol{z}}\right)^T r(\boldsymbol{z}) + \frac{1}{2} r(\boldsymbol{z})^T \frac{\partial r}{\partial \boldsymbol{z}}|_{\boldsymbol{z}} = r(\boldsymbol{z})^T \frac{\partial r}{\partial \boldsymbol{z}}|_{\boldsymbol{z}} = \frac{1}{2}(I - P_{\mathbf{a}})\boldsymbol{z}^* \tag{13}$$

$$\frac{\partial \rho}{\partial \overline{\boldsymbol{z}}}|_{\boldsymbol{z}} = \frac{1}{2}\left(\frac{\partial r}{\partial \overline{\boldsymbol{z}}}|_{\boldsymbol{z}}\right)^T r(\boldsymbol{z}) + \frac{1}{2} r(\boldsymbol{z})^T \frac{\partial r}{\partial \overline{\boldsymbol{z}}}|_{\boldsymbol{z}} = r(\boldsymbol{z})^T \frac{\partial r}{\partial \overline{\boldsymbol{z}}}|_{\boldsymbol{z}} = \frac{1}{2}(I - P_{\mathbf{a}})\boldsymbol{z}, \tag{14}$$

Next we evaluate the following quantities evaluated at $\boldsymbol{z}$:

$$\mathcal{H}_{\boldsymbol{zz}} := \frac{\partial}{\partial \boldsymbol{z}}|_{\boldsymbol{z}}\left(\frac{\partial \rho}{\partial \boldsymbol{z}}\right)^* \quad \mathcal{H}_{\overline{\boldsymbol{z}}\boldsymbol{z}} := \frac{\partial}{\partial \overline{\boldsymbol{z}}}|_{\boldsymbol{z}}\left(\frac{\partial \rho}{\partial \boldsymbol{z}}\right)^* \tag{15}$$

$$\mathcal{H}_{\boldsymbol{z}\overline{\boldsymbol{z}}} := \frac{\partial}{\partial \boldsymbol{z}}|_{\boldsymbol{z}}\left(\frac{\partial \rho}{\partial \overline{\boldsymbol{z}}}\right)^* \quad \mathcal{H}_{\overline{\boldsymbol{z}}\overline{\boldsymbol{z}}} := \frac{\partial}{\partial \overline{\boldsymbol{z}}}|_{\boldsymbol{z}}\left(\frac{\partial \rho}{\partial \overline{\boldsymbol{z}}}\right)^*. \tag{16}$$

Clearly, by (13) we have

$$\begin{aligned}
\mathcal{H}_{\boldsymbol{zz}} =& \frac{1}{2}\big(I - \frac{\partial}{\partial \boldsymbol{z}}|_{\boldsymbol{z}} P_{\mathbf{a}}\boldsymbol{z}\big) = \frac{1}{2}I - \frac{1}{2}\frac{\partial}{\partial \boldsymbol{z}}\begin{pmatrix} a_1 \frac{z_1^*}{|z_1|} \\ \vdots \\ a_M \frac{z_M^*}{|z_M|} \end{pmatrix} \\
=& \frac{1}{2}I - \frac{1}{2}\begin{pmatrix} a_1 \frac{\partial}{\partial z_1}|_{z_1}(\frac{z_1^*}{|z_1|}) & & 0 \\ & \ddots & \\ 0 & & a_M \frac{\partial}{\partial z_M}|_{z_M}(\frac{z_M^*}{|z_M|}) \end{pmatrix} \\
=& \frac{1}{2}\left(I - \frac{1}{2}\text{diag}\frac{\mathbf{a}}{|\boldsymbol{z}|}\right),
\end{aligned}$$

where we use the fact that $\frac{\partial}{\partial w}|_z\Big(\frac{w^*}{|w|}\Big) = \frac{1}{2|w|}$, where $w \in \mathbb{C}$. Similarly we have

$$\mathcal{H}_{\overline{\boldsymbol{z}}\overline{\boldsymbol{z}}} = \frac{1}{2}\left(I - \frac{1}{2}\text{diag}\frac{\mathbf{a}}{|\boldsymbol{z}|}\right).$$

By (14) we have

$$\mathcal{H}_{\boldsymbol{z}\overline{\boldsymbol{z}}} = \frac{1}{2}\frac{\partial}{\partial \overline{\boldsymbol{z}}}|_{\boldsymbol{z}} P_{\mathbf{a}}\boldsymbol{z} = \frac{1}{4}\text{diag}\left(\frac{\mathbf{a}\boldsymbol{z}^{*2}}{|\boldsymbol{z}|^3}\right).$$

With the above preparations, we can evaluate the gradient and Hessian of ρ at $\boldsymbol{z}$. Denote $\boldsymbol{z} = (c_i e^{i\phi_i})_{i=1}^M \in \mathbb{C}^M$, where $c_i \in \mathbb{R}_+$ and $\phi_i \in [0, 2\pi)$. By definition, the gradient of ρ at $\boldsymbol{z}$ is the dual vector of $\frac{\partial}{\partial \boldsymbol{z}}|_{\boldsymbol{z}}\rho$ associated with the canonical metric on $\mathbb{C}^M$, that is,

$$\nabla\rho|_{\boldsymbol{z}} := \left(\frac{\partial}{\partial \boldsymbol{z}}|_{\boldsymbol{z}}\rho\right)^* = \frac{1}{2}(I - P_{\mathbf{a}})\boldsymbol{z} = \frac{1}{2}[(c_l - a_l)e^{i\phi_l}]_{l=1}^M. \tag{17}$$

The Hessian of ρ at $\boldsymbol{z}$, denoted by $\nabla^2\rho|_{\boldsymbol{z}}$, by a direct calculation is given by

$$\nabla^2\rho|_{\boldsymbol{z}} := \begin{pmatrix} \mathcal{H}_{\boldsymbol{zz}} & \mathcal{H}_{\boldsymbol{z}\overline{\boldsymbol{z}}} \\ \mathcal{H}_{\overline{\boldsymbol{z}}\boldsymbol{z}} & \mathcal{H}_{\overline{\boldsymbol{z}}\overline{\boldsymbol{z}}} \end{pmatrix},$$

which leads to the following evaluation of the curvature of the ρ. Take $\boldsymbol{w} \in \mathbb{C}^M$. Denote $\boldsymbol{w} = (b_i e^{i\theta_i})_{i=1}^M$, where $b_i \in \mathbb{R}_+$ and $\theta_i \in [0, 2\pi)$. Then by a direct expansion, the second derivative of ρ in the direction

$\boldsymbol{w}$ at $\boldsymbol{z}$ is

$$\begin{aligned}\nabla^2\rho|_{\boldsymbol{z}}(\boldsymbol{w}) &:= (\boldsymbol{w}^* \quad \overline{\boldsymbol{w}}^*)\nabla^2\rho|_{\boldsymbol{z}}\begin{pmatrix}\boldsymbol{w}\\ \overline{\boldsymbol{w}}\end{pmatrix}\\ &= (\boldsymbol{w}^* \quad \overline{\boldsymbol{w}}^*)\begin{pmatrix}\frac{1}{2}\boldsymbol{w}-\frac{1}{4}\frac{\mathbf{a}}{|\boldsymbol{z}|}\boldsymbol{w}+\frac{1}{4}\frac{\mathbf{a}\boldsymbol{z}^2}{|\boldsymbol{z}|^3}\overline{w}\\ \frac{1}{2}\overline{\boldsymbol{w}}-\frac{1}{4}\frac{\mathbf{a}}{|\boldsymbol{z}|}\overline{\boldsymbol{w}}+\frac{1}{4}\frac{\mathbf{a}\boldsymbol{z}^{*2}}{|\boldsymbol{z}|^3}\boldsymbol{w}\end{pmatrix}\\ &= \sum_{j=1}^{M} b_j^2\left(1-\frac{a_j}{c_j}\sin^2(\theta_j-\phi_j)\right). \qquad (18)\end{aligned}$$

We have the following observations about the gradient and Hessian:

- Note that we can view the AP algorithm as the *projected gradient descent algorithm* related to the objective function ρ [68]. Indeed, we have
$$\zeta^{(l+1)} = \zeta^{(l)} - 2P_{\mathsf{S}}\nabla\rho|_{\zeta^{(l)}} = P_{\mathsf{S}}P_{\mathbf{a}}\zeta^{(l)}.$$
By (17), for $z = \zeta^{(l)} = \boldsymbol{b}^{(l)}e^{i\phi^{(l)}}$ we have
$$\nabla\rho|_{\zeta^{(l)}} = \frac{1}{2}\left[(b_k^{(l)}-a_k)e^{i\phi_k^{(l)}}\right]_{k=1}^{M}.$$
By Lemma 3.3, for a generic R_{S}, the gradient of ρ on R_{S} is zero only at $S_{\mathbf{a}}$ since the only points on R_{S} that have modulations $\mathbf{a}$ is the solution set. Also, by Theorem 3.5, $\nabla\rho|_{\zeta^{(l)}}$ is not perpendicular to R_{S}, that is, $P_{\mathsf{S}}\nabla\rho|_{\zeta^{(l)}}$ is nonzero outside $S_{\mathbf{a}}$. Furthermore, $\nabla\rho|_{\zeta^{(l)}}$ is not located on R_{S} since $\boldsymbol{b}^{(l)}$ and $\mathbf{a}$ are not related by a constant factor and we know that $\phi^{(l)}$ is related to $\boldsymbol{b}^{(l)}$ but not $\boldsymbol{b}^{(l)}-\mathbf{a}$.
- For $\boldsymbol{z} = e^{it}\mathsf{S}\psi_0 = \mathbf{a}e^{i(\boldsymbol{\phi}^{\mathbf{a}}+t)} \in S_{\mathbf{a}}$, where $t\in[0,2\pi)$, and $w = \boldsymbol{b}e^{i\boldsymbol{\theta}}$, by (18) we know
$$\nabla^2\rho|_{\boldsymbol{z}}(\boldsymbol{w}) = \sum_{j=1}^{M} b_j^2\left(1-\sin^2(\theta_j-\phi_j^{\mathbf{a}}-t)\right),$$
which is always non-negative since $\sin^2\le 1$. When $\boldsymbol{\theta} = \boldsymbol{\phi}^{\mathbf{a}}+t+\pi/2$, $\nabla^2\rho|_{\mathsf{S}\psi_0}(\boldsymbol{w}) = 0$. However, by Lemma 3.3, for a generic R_{S}, we know that if $\boldsymbol{b}\neq\mathbf{a}$ and $\boldsymbol{w}\in R_{\mathsf{S}}$, then $\boldsymbol{\theta}\neq\boldsymbol{\phi}^{\mathbf{a}}+t'$ for all $t'\in[0,2\pi)$. In particular, $\nabla^2\rho|_{\boldsymbol{z}}(\boldsymbol{w})$ is strictly positive when $\boldsymbol{w}$ is in a small enough ball around the origin.

5. The ptychographic imaging problem and phase synchronization

In this section, we focus ourselves on the ptychography problem. Given $\mathbf{a} = |\mathbf{FQ}\psi|$, we combine the essences of the AP algorithm and consider the following optimization problem:

$$\operatorname*{argmin}_{\zeta\in\mathbb{T}_1} \|(I-P_{\mathbf{FQ}})\mathbf{a}\zeta\|^2. \qquad (19)$$

It is clear that the phase of $e^{it}\mathbf{FQ}\psi$, where $t\in\mathbb{T}_1$ is a solution to (19). Note that under the constraint of ζ, $\zeta^*\mathrm{diag}(\mathbf{a})^2\zeta = \|\mathbf{a}\|_2^2$ is fixed. So, solving (19) is equivalent to solving

$$\operatorname*{argmax}_{\zeta\in\mathbb{T}_{\mathbf{a}}} \zeta^* P_{\mathbf{FQ}}\zeta, \qquad (20)$$

where $P_{\mathbf{FQ}}$ is clearly a Hermitian matrix. Intuitively, (20) says that the phases of Fourier modes associated with diffraction images should be related via the operator $P_{\mathbf{FQ}}$. However, the constraint regarding $\mathbb{T}_{\mathbf{a}}$ drives the optimization problem into a non-convex one.

The first possible relaxation is taking into account the fact that $\mathbb{T}_{\mathbf{a}}$ is a subset of the sphere of radius $\|\mathbf{a}\|$, that is, we directly evaluate

$$\operatorname*{argmax}_{\zeta\in\mathbb{C}^{Km^2},\,\zeta^*\zeta=\|\mathbf{a}\|_2^2} \zeta^* P_{\mathbf{FQ}}\zeta, \qquad (21)$$

which is equivalent to solving the eigenvalue problem of $P_{\mathbf{FQ}}$. Clearly, the solution exists as an eigenvector with eigenvalue 1. However, since $P_{\mathbf{FQ}}$ is a projection operator, the only eigenvalues are 0 and 1. Thus, although the solution exists in the top eigenspace, we cannot obtain it directly by solving (21).

Before proceeding, we study the geometric meaning of (20) a bit more. Define an index map $\ell : \mathcal{X}_K \times D^{m\times m} \to \{1,\dots,Km^2\}$ by

$$\ell(\boldsymbol{x}_k,\boldsymbol{r}_k) = (k-1)m^2 + \mathsf{r}_m(\boldsymbol{r}_k),$$

which is a 1 to 1 map providing the index of the entry $\boldsymbol{r}_k$ of the k-th illumination window $\iota_{\boldsymbol{x}_k}(D^{m\times m})$ in the long stack vector. For $j=1,\dots,K$ and $\boldsymbol{s}\in D^{m\times m}$, define a set

$$I_{\boldsymbol{x}_j,\boldsymbol{s}} := \{k : \boldsymbol{x}_j+\boldsymbol{s} \in \iota_{\boldsymbol{x}_k}(D^{m\times m})\} \subset \{1,\dots,K\},$$

which contains the indices of all illumination windows covering $\boldsymbol{x}_j+\boldsymbol{s}$. Also define a subset of $D^{m\times m}$

$$J_{\boldsymbol{x}_j,\boldsymbol{s}} := \{\boldsymbol{r}\in D^{m\times m} : \boldsymbol{x}_k+\boldsymbol{r} = \boldsymbol{x}_j+\boldsymbol{s},\ k\in I_{\boldsymbol{x}_j,\boldsymbol{s}}\},$$

which collects the indices of the pixels in all illumination windows which cover $\boldsymbol{x}_j+\boldsymbol{s}$. We choose to use this seeming complicated index since we would like to make clear the relationship between the illumination windows and their pixels. By Assumption 2.1 and a direct calculation, we know that $\mathbf{Q}^*\mathbf{Q}$ is a $n^2\times n^2$ non-degenerate diagonal matrix describing how many illumination windows cover a given pixel of the object of interest, where the $\mathsf{r}_n(\boldsymbol{x}_j+\boldsymbol{r}_j)$-th diagonal entry is $\sum_{\boldsymbol{r}\in J_{\boldsymbol{x}_j,\boldsymbol{r}_j}} |\omega(\boldsymbol{r})|^2$. So, the matrix $P_{\mathbf{Q}} := \mathbf{Q}(\mathbf{Q}^*\mathbf{Q})^{-1}\mathbf{Q}^*$ satisfies

$$P_{\mathbf{Q}}(\ell(\boldsymbol{x}_i,\boldsymbol{r}_i),\ell(\boldsymbol{x}_j,\boldsymbol{r}_j)) = \frac{\omega(\boldsymbol{r}_i)\omega^*(\boldsymbol{r}_j)}{\sum_{\boldsymbol{r}\in J_{\boldsymbol{x}_j,\boldsymbol{r}_j}}|\omega(\boldsymbol{r})|^2}\delta_{\boldsymbol{x}_i+\boldsymbol{r}_i,\,\boldsymbol{x}_j+\boldsymbol{r}_j},$$

where δ is the Kronecker's delta. Note that $P_{\mathbf{Q}}$ is not a diagonal matrix since by Assumption 2.1 there are more than two illumination windows covering a given pixel. Clearly, for all $\boldsymbol{x}_i\in\mathcal{X}_K$, and $\zeta\in\mathbb{C}^{Km^2}$, we have $\left[\mathbf{F}^*\zeta\right]_{(i)} = F^*\zeta_{(i)}$, where $\zeta_{(i)}(\boldsymbol{r}) := \zeta(\boldsymbol{x}_i+\boldsymbol{r})$ and $\boldsymbol{r}\in D^{m\times m}$. Also, $P_{\mathbf{FQ}} = \mathbf{F}P_{\mathbf{Q}}\mathbf{F}^*$. As result,

$$\begin{aligned}\zeta^*P_{\mathbf{FQ}}\zeta &= \sum_{i,\boldsymbol{r}_i}\sum_{j,\boldsymbol{r}_j}\frac{\omega(\boldsymbol{r}_i)\omega^*(\boldsymbol{r}_j)}{\sum_{\boldsymbol{r}\in J_{\boldsymbol{x}_j,\boldsymbol{r}_j}}|\omega(\boldsymbol{r})|^2}\left[F^*\zeta_{(i)}\right]^*(\boldsymbol{r}_i)\left[F^*\zeta_{(j)}\right](\boldsymbol{r}_j)\delta_{\boldsymbol{x}_i+\boldsymbol{r}_i,\,\boldsymbol{x}_j+\boldsymbol{r}_j}\\ &= \sum_{(i,\boldsymbol{r}_i)\sim(j,\boldsymbol{r}_j)}\frac{\omega(\boldsymbol{r}_i)\omega^*(\boldsymbol{r}_j)}{\sum_{\boldsymbol{r}\in J_{\boldsymbol{x}_j,\boldsymbol{r}_j}}|\omega(\boldsymbol{r})|^2}\left[F^*\zeta_{(i)}\right]^*(\boldsymbol{r}_i)\left[F^*\zeta_{(j)}\right](\boldsymbol{r}_j),\end{aligned}$$

where $(i,\boldsymbol{r}_i)\sim(j,\boldsymbol{r}_j)$ means all illumination windows covering the pixel $\iota_{\boldsymbol{x}_i}(\boldsymbol{r}_i)$. Geometrically, $P_{\mathbf{Q}}$ describes how two illumination windows in the spatial domain are intersected and how the overlapped pixels are related via the illuminating function ω. Note that when $\zeta_{(i)}$ contains the right amplitude and phase, $F^*\zeta_{(i)}$ is the correct image on $\iota_{\boldsymbol{x}_i}(D^{m\times m})$. Thus, maximizing $\zeta^*P_{\mathbf{FQ}}\zeta$ is equivalent to requiring that the images on a pair of overlapping illumination windows match in the overlapping region. In particular, by Assumption 2.1, phases on one illumination window will be synchronized with at least one different illumination window if we maximize $\zeta^*P_{\mathbf{FQ}}\zeta$. Also, by Assumption 2.2, the phases in different disconnected regions of ψ associated with $\mathcal{X}_K$ are guaranteed to interact with each other so that the phase can be synchronized in the end.

To better understand $P_{\mathbf{FQ}}$, we further consider the relationship between the phases when the illumination windows overlap. Consider the following *phase synchronization* problem:

$$\operatorname*{argmax}_{\zeta\in\mathbb{T}_{\mathbf{1}}} \zeta^*P_{\mathbf{FQ}}\zeta. \tag{22}$$

To study this non-convex optimization problem, we start from studying the Hermitian matrix $P_{\mathbf{FQ}}$ in (22). The amplitude information $\mathbf{a}$ will be taken into account later. Denote $\mathsf{O}_{ij} := \iota_{\boldsymbol{x}_i}(D^{m\times m})\cap\iota_{\boldsymbol{x}_j}(D^{m\times m})$ to be the overlap of two illumination windows. By the definition (1), a direct expansion of (22) leads to

$$\begin{aligned}&\zeta^*P_{\mathbf{FQ}}\zeta = \zeta^*\mathbf{FQ}(\mathbf{Q}^*\mathbf{Q})^{-1}\mathbf{Q}^*\mathbf{F}^*\zeta\\ &= \sum_{i,j=1}^{K}\zeta^*_{(i)}F\mathrm{diag}(w)\mathbf{R}\mathbf{T}_{\boldsymbol{x}_i}(\mathbf{Q}^*\mathbf{Q})^{-1}\mathbf{T}^*_{\boldsymbol{x}_j}\mathbf{R}^*\mathrm{diag}(w^*)F^*\zeta_{(j)}\\ &= \sum_{i,j:\,\mathsf{O}_{ij}\neq\emptyset}\zeta^*_{(i)}F\mathrm{diag}(w)\mathbf{R}\mathbf{T}_{\boldsymbol{x}_i}(\mathbf{Q}^*\mathbf{Q})^{-1}\mathbf{T}^*_{\boldsymbol{x}_i}\mathbf{T}_{\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j}}\mathbf{R}^*\mathrm{diag}(w^*)F^*\zeta_{(j)},\end{aligned}$$

where $\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j} := \boldsymbol{x}_i-\boldsymbol{x}_j$ and the last equality comes from the fact that $\mathbf{T}_{\boldsymbol{x}_i}\mathbf{T}^*_{\boldsymbol{x}_j} = \mathbf{T}_{\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j}}$ and $\mathbf{T}^*_{\boldsymbol{x}_i}\mathbf{T}_{\boldsymbol{x}_i} = I$. Clearly if $\mathsf{O}_{ij}=\emptyset$, $\mathbf{T}_{\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j}}\mathbf{R}^*$ is a zero matrix. Note that $\mathbf{T}_{\boldsymbol{x}_i}(\mathbf{Q}^*\mathbf{Q})^{-1}\mathbf{T}^*_{\boldsymbol{x}_i}$, as the conjugation of $(\mathbf{Q}^*\mathbf{Q})^{-1}$

by $\mathbf{T}_{\boldsymbol{x}_i}$, is diagonal. It actually translates the $\mathsf{r}_n(\boldsymbol{x}_i)$-th diagonal entry to the 1-st diagonal entry. Also note that the overlapping information about the i-th and j-th illumination windows is preserved in $\mathbf{T}_{\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j}}\mathbf{R}^*$.

Now we move $\mathbf{T}_{\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j}}$ out of $F\mathrm{diag}(w)\mathbf{R}\mathbf{T}_{\boldsymbol{x}_i}(\mathbf{Q}^*\mathbf{Q})^{-1}\mathbf{T}^*_{\boldsymbol{x}_i}\mathbf{T}_{\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j}}\mathbf{R}^*\mathrm{diag}(w^*)F^*$ by a direct expansion:

$$\begin{aligned}&F\mathrm{diag}(w)\mathbf{R}\mathbf{T}_{\boldsymbol{x}_i}(\mathbf{Q}^*\mathbf{Q})^{-1}\mathbf{T}^*_{\boldsymbol{x}_i}\mathbf{T}_{\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j}}\mathbf{R}^*\mathrm{diag}(w^*)F^*\\ =&F\mathbf{M}^{\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j}}F^*\mathrm{diag}([e^{-i\mathsf{q}_m^{-1}(1)\cdot\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j}},\dots,e^{-i\mathsf{q}_m^{-1}(m^2)\cdot\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j}}]),\end{aligned}$$

where $\mathbf{M}^{\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j}}$ is a $m^2\times m^2$ *masking matrix* which is diagonal and depends on $\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j}$:

$$\boldsymbol{e}^T_{\mathsf{r}_m(\boldsymbol{s})}\mathbf{M}^{\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j}}\boldsymbol{e}_{\mathsf{r}_m(\boldsymbol{s})}:=\begin{cases}\frac{\omega(\boldsymbol{s})\omega^*(\boldsymbol{s}-\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j})}{\sum_{\boldsymbol{r}\in J_{\boldsymbol{x}_i,\boldsymbol{s}}}|\omega(\boldsymbol{r})|^2} & \text{when } \boldsymbol{s}\in\mathsf{D}_{ij}\\ 0 & \text{otherwise,}\end{cases}$$

and $\mathsf{D}_{ij}:=\iota_{r(0,0)}D^{m\times m}\cap[\mathbf{T}_{\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j}}\iota_{r(0,0)}D^{m\times m}]$. This equality indicates the influence of the restriction matrix $\mathbf{R}$ – the non-overlapped parts of the two overlapping subregions cannot be eliminated. Next, for $\boldsymbol{r}_i,\boldsymbol{r}_j\in D^{m\times m}$, when $\mathsf{O}_{ij}\neq\emptyset$, the $m^2\times m^2$ matrix $F\mathbf{M}^{\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j}}F^*\mathrm{diag}([e^{-i\mathsf{q}_m^{-1}(1)\cdot\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j}},\dots,e^{-i\mathsf{q}_m^{-1}(m^2)\cdot\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j}}])$ satisfies

$$\begin{aligned}&\boldsymbol{e}^T_{\mathsf{r}_m(\boldsymbol{r}_i)}F\mathbf{M}^{\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j}}F^*\mathrm{diag}([e^{i\mathsf{q}_m^{-1}(1)\cdot\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j}},\dots,e^{i\mathsf{q}_m^{-1}(m^2)\cdot\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j}}])\boldsymbol{e}_{\mathsf{r}_m(\boldsymbol{r}_j)}\\ =&\boldsymbol{e}^T_{\mathsf{r}_m(\boldsymbol{r}_i)}F\mathbf{M}^{\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j}}F^*\boldsymbol{e}_{\mathsf{r}_m(\boldsymbol{r}_j)}e^{i\boldsymbol{r}_j\cdot\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j}}\\ =&e^{i\boldsymbol{r}_j\cdot\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j}}\sum_{\boldsymbol{s}\in D^{m\times m}}\boldsymbol{e}^T_{\mathsf{r}_m(\boldsymbol{s})}\mathbf{M}^{\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j}}\boldsymbol{e}_{\mathsf{r}_m(\boldsymbol{s})}e^{i(\boldsymbol{r}_i-\boldsymbol{r}_j)\cdot\boldsymbol{s}}\\ =&e^{i\boldsymbol{r}_j\cdot\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j}}\sum_{\boldsymbol{s}\in\mathsf{D}_{ij}}\frac{\omega(\boldsymbol{s})\omega^*(\boldsymbol{s}-\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j})}{\sum_{\boldsymbol{r}\in J_{\boldsymbol{x}_i,\boldsymbol{s}}}|\omega(\boldsymbol{r})|^2}e^{i(\boldsymbol{r}_i-\boldsymbol{r}_j)\cdot\boldsymbol{s}}\\ =&e^{i(\boldsymbol{r}_i+\boldsymbol{r}_j)\cdot\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j}/2}\sum_{\boldsymbol{s}\in\mathbf{T}_{\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j}/2}\mathsf{D}_{ij}}\frac{\omega(\boldsymbol{s}+\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j}/2)\omega^*(\boldsymbol{s}-\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j}/2)}{\sum_{\boldsymbol{r}\in J_{\boldsymbol{x}_j,\boldsymbol{s}}}|\omega(\boldsymbol{r})|^2}e^{i(\boldsymbol{r}_i-\boldsymbol{r}_j)\cdot\boldsymbol{s}}\\ =:&e^{i(\boldsymbol{r}_i+\boldsymbol{r}_j)\cdot\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j}/2}V_{\omega_{ij}}(\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j},\Phi_{\boldsymbol{r}_i\boldsymbol{r}_j}),\end{aligned}\tag{23}$$

where $\Phi_{\boldsymbol{r}_i\boldsymbol{r}_j}:=\boldsymbol{r}_i-\boldsymbol{r}_j$,

$$\omega_{ij}(\boldsymbol{r}):=\frac{\omega(\boldsymbol{r})}{\sqrt{\sum_{\boldsymbol{s}\in J_{(\boldsymbol{x}_i+\boldsymbol{x}_j)/2,\boldsymbol{r}}}|\omega(\boldsymbol{s})|^2}}\chi_{\mathsf{D}_{ij}\cup\mathbf{T}^*_{\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j}/2}\mathsf{D}_{ij}\cup\mathbf{T}^*_{\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j}}\mathsf{D}_{ij}}$$

and $V_{\omega_{ij}}$ is the *Fourier-Wigner transform* [34] of the function ω_{ij}. To sum up, the $(\ell(i,\boldsymbol{r}_i),\ell(j,\boldsymbol{r}_j))$-th entry of $P_{\mathbf{FQ}}$ is $V_{\omega_{ij}}(\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j},\Phi_{\boldsymbol{r}_i\boldsymbol{r}_j})e^{i(\boldsymbol{r}_i+\boldsymbol{r}_j)\cdot\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j}/2}$.

Recall that the Fourier-Wigner transform of ω_{ij} is also called the *ambiguity function* of ω_{ij}, which measures the spatial lag $\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j}$ and frequency shift $\Phi_{\boldsymbol{r}_i\boldsymbol{r}_j}$ between the two diffraction images when $\iota_{\boldsymbol{x}_i}(D^{m\times m})\cap\iota_{\boldsymbol{x}_j}(D^{m\times m})\neq\emptyset$. It is well-known that the absolute value of the ambiguity function gauges how difficult we can distinguish two objects, that is, how similar two objects are [34, p.33]. Thus $V_{\omega_{ij}}(\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j},\Phi_{\boldsymbol{r}_i\boldsymbol{r}_j})$ can be viewed as a sort of affinity measuring the relationship between two illumination windows. Also, from (23) we know that the phase information of $F\mathrm{diag}(\omega)$ gets involved in $V_{\omega_{ij}}$, in particular when $i=j$. Indeed, when we are working with the same patch, $\mathbf{M}^0$ is a diagonal matrix with real entries $|\omega|^2$, so $F\mathbf{M}^0F^*$ contains only the phase information of $F\mathrm{diag}(\omega)$, which influences the phase estimation.

As a result, we have

$$\zeta^*P_{\mathbf{FQ}}\zeta=\sum_{i,j:\,\mathsf{O}_{ij}\neq\emptyset}\sum_{\boldsymbol{r}_i,\boldsymbol{r}_j\in D^{m\times m}}\zeta^*_{(i)}(\boldsymbol{r}_i)V_{\omega_{ij}}(\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j},\Phi_{\boldsymbol{r}_i\boldsymbol{r}_j})e^{i(\boldsymbol{r}_i+\boldsymbol{r}_j)\cdot\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j}/2}\zeta_{(j)}(\boldsymbol{r}_j).\tag{24}$$

Consider a graph $\mathbb{G}=(\mathbb{V},\mathbb{E})$, where the vertices $\mathbb{V}=\mathcal{X}_K\times D^{m\times m}$ are constituted by the pixels of all illuminated images, and there an edge between $(i,\boldsymbol{r}_i)$ and $(j,\boldsymbol{r}_j)$ for all $\boldsymbol{r}_i,\boldsymbol{r}_j\in D^{m\times m}$ if $\iota_{\boldsymbol{x}_i}(D^{m\times m})\cap\iota_{\boldsymbol{x}_j}(D^{m\times m})\neq\emptyset$. Please see Figure 4 for an illustration. The equation (24) is thus built from the graph $\mathbb{G}$ with a "phase synchronization function" Ω defined on $\mathbb{E}$:

$$\Omega:((i,\boldsymbol{r}_i),(j,\boldsymbol{r}_j))\in\mathbb{E}\mapsto V_{\omega_{ij}}(\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j},\Phi_{\boldsymbol{r}_i\boldsymbol{r}_j})e^{i(\boldsymbol{r}_i+\boldsymbol{r}_j)\cdot\Delta_{\boldsymbol{x}_i\boldsymbol{x}_j}/2},$$

that is, we have

$$\zeta^* P_{\mathsf{FQ}} \zeta = \sum_{i,j:\, \mathsf{O}_{ij}\neq\emptyset} \sum_{\boldsymbol{r}_i,\boldsymbol{r}_j\in D^{m\times m}} \zeta^*_{(i)}(\boldsymbol{r}_i)\Omega((i,\boldsymbol{r}_i),(j,\boldsymbol{r}_j))\zeta_{(j)}(\boldsymbol{r}_j).$$

Note that Ω depends on the illumination scheme and the lens. Intuitively, if two illumination windows overlap, they have common information in the Fourier space up to some phase difference determined by the relative position of the illuminations, while this information is contaminated by the non-overlapping parts of the two illuminations.

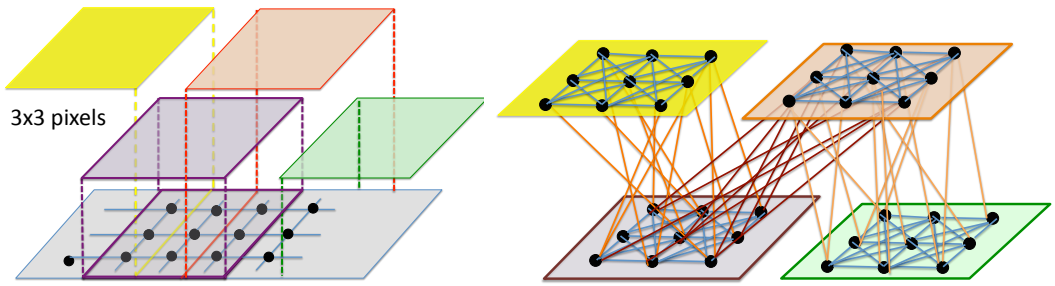


FIGURE 4. Left: the illuminative figure for the ptychographic problem. We assume that the unknown object of interest is covered by 4 illumination windows of size 3×3. Right: the graph $\mathbb{G}$ associated with the algorithm aiming to solve the ptychographic experiment. The block spots are vertices associated with pixels of each illumination image.

With the above understanding of the P_{FQ} matrix and the inspiration of [59], we link the ptychography imaging problem to the synchronization problem defined on the graph $\mathbb{G} = (\mathbb{V},\mathbb{E})$. Now we take the amplitude information $\mathbf{a}$ into account. It is well known that the larger the amplitude is, the more important its associated phase is if we want to "reconstruct the image". Thus, we would pay more attention on reconstructing the phase of pixels in the diffraction images with larger amplitudes, for example, we might want to maximizing the following functional with the constraint $\zeta\in\mathbb{T}_{\mathbf{1}}$:

$$\zeta^*\mathrm{diag}(\mathbf{a})P_{\mathsf{FQ}}\mathrm{diag}(\mathbf{a})\zeta = \sum_{i,j:\, \mathsf{O}_{ij}\neq\emptyset} \sum_{\boldsymbol{r}_i,\boldsymbol{r}_j\in D^{m\times m}} \zeta^*_{(i)}(\boldsymbol{r}_i)\mathbf{a}_{(i)}(\boldsymbol{r}_i)\Omega((i,\boldsymbol{r}_i),(j,\boldsymbol{r}_j))\mathbf{a}_{(i)}(\boldsymbol{r}_i)\zeta_{(j)}(\boldsymbol{r}_j). \tag{25}$$

5.1. **Spectral relaxation and phase synchronization.** Based on the above understanding regarding the P_{FQ} and the amplitude information, in this section we propose two relaxations of the non-convex optimization problems discussed above to estimate the phase, which lead to a better initial value of the AP algorithm.

The first algorithm is directly motivated by (25) where we take the affinity information among vertices and phase relationship into account. We have the following observations.

- the phase between vertices $(i,\boldsymbol{r}_i)$ and $(j,\boldsymbol{r}_j)$ are related by a non-unitary transform $\Omega((i,\boldsymbol{r}_i),(j,\boldsymbol{r}_j))$, which modulation indicated the affinity;
- the larger the amplitude $\mathbf{a}_{(i)}(\boldsymbol{r}_i)$ is, the more effort we should put in recovering the phase;
- the phase ramping effect introduced by

$$\widetilde{\omega} := \frac{F\mathrm{diag}(\omega^\vee)}{|F\mathrm{diag}(\omega^\vee)|}\chi_{|F\mathrm{diag}(\omega^\vee)|} + (1-\chi_{|F\mathrm{diag}(\omega^\vee)|}),$$

  as is shown in (23).

These observations suggest us to consider the following relaxation and its relationship with the recent developed data analysis framework *connection graph Laplacian (CGL)* [60, 61, 5, 20], which we discuss now. First, we define the *affinity function* (or weight function) $\mathfrak{w}:\mathbb{E}\to\mathbb{R}_+$ to encode the affinity information:

$$\mathfrak{w}((i,\boldsymbol{r}_i),(j,\boldsymbol{r}_j)) := \mathbf{a}_{(i)}(\boldsymbol{r}_i)|\Omega((i,\boldsymbol{r}_i),(j,\boldsymbol{r}_j))|\mathbf{a}_{(j)}(\boldsymbol{r}_j)$$

when $((i,\boldsymbol{r}_i),(j,\boldsymbol{r}_j))\in\mathbb{E}$, and the *relationship function* $\mathfrak{g}:\mathbb{E}\to U(1)$ so that

$$\mathfrak{g}((i,\boldsymbol{r}_i),(j,\boldsymbol{r}_j)) := \widetilde{\omega}(\boldsymbol{r}_i)\left(\frac{\Omega((i,\boldsymbol{r}_i),(j,\boldsymbol{r}_j))}{|\Omega((i,\boldsymbol{r}_i),(j,\boldsymbol{r}_j))|}\chi_{\Omega((i,\boldsymbol{r}_i),(j,\boldsymbol{r}_j))\neq 0} + (1-\chi_{\Omega((i,\boldsymbol{r}_i),(j,\boldsymbol{r}_j))=0})\right)\widetilde{\omega}^*(\boldsymbol{r}_j)$$

when $((i,\boldsymbol{r}_i),(j,\boldsymbol{r}_j)) \in \mathbb{E}$, which purely encodes the phase relationship among vertices. Given $\mathbb{G}$, $\mathfrak{w}$ and $\mathfrak{g}$, we further define a complex $Km^2 \times Km^2$ matrix $\mathbf{S}$ so that

$$\mathbf{S}(\ell(i,\boldsymbol{r}_i),\ell(j,\boldsymbol{r}_j)) = \begin{cases} \mathfrak{w}((i,\boldsymbol{r}_i),(j,\boldsymbol{r}_j))\mathfrak{g}((i,\boldsymbol{r}_i),(j,\boldsymbol{r}_j)) & \text{when } ((i,\boldsymbol{r}_i),(j,\boldsymbol{r}_j)) \in \mathbb{E} \\ 0 & \text{otherwise} \end{cases}$$

and a complex $Km^2 \times Km^2$ diagonal matrix $\mathbf{D}$ so that

$$\mathbf{D}(\ell(i,\boldsymbol{r}_i),\ell(i,\boldsymbol{r}_i)) = \sum_{((i,\boldsymbol{r}_i),(j,\boldsymbol{r}_j))\in\mathbb{E}} \mathfrak{w}((i,\boldsymbol{r}_i),(j,\boldsymbol{r}_j)).$$

Then, the CGL matrix is defined as $\mathbf{D}^{-1}\mathbf{S}$. Note that $\mathbf{D}$ is invertible by Assumption 2.1 and the positivity assumption of $\mathfrak{w}$. We thus propose our first phase estimator to be the phase of the top eigenvector of $\mathbf{D}^{-1}\mathbf{S}$, which we call *CGL-phase synchronization (CGL-PS)*.

We mention that the CGL is a generalization of the well known graph Laplacian in that it takes not only the affinity between vertices into account but also the relationship between verticex [60]. To be more precise, if we take a complex valued function $\theta : \mathbb{V} \to \mathbb{C}$, we have the following expansion

$$[\mathbf{D}^{-1}\mathbf{S}\theta](\ell(i,\boldsymbol{r}_i)) = \frac{\sum_{((i,\boldsymbol{r}_i),(j,\boldsymbol{r}_j))\in\mathbb{E}} \mathfrak{w}((i,\boldsymbol{r}_i),(j,\boldsymbol{r}_j))\mathfrak{g}((i,\boldsymbol{r}_i),(j,\boldsymbol{r}_j))\theta(j,\boldsymbol{r}_j)}{\sum_{((i,\boldsymbol{r}_i),(j,\boldsymbol{r}_j))\in\mathbb{E}} \mathfrak{w}((i,\boldsymbol{r}_i),(j,\boldsymbol{r}_j))}.$$

This formula can be viewed as a generalized random walk on the graph. Indeed, if we view the complex-valued function θ as the status of a particle defined on the vertices, when we move from one vertex to the other one, the status is modified according to the relationship between vertices encoded in $\mathfrak{g}$. Clearly, if the complex-valued status θ in all vertices are "synchronized" according to the described relationship $\mathfrak{g}$, that is, $\theta(i,\boldsymbol{r}_i) = \mathfrak{g}((i,\boldsymbol{r}_i),(j,\boldsymbol{r}_j))\theta(j,\boldsymbol{r}_j)$ for all $((i,\boldsymbol{r}_i),(j,\boldsymbol{r}_j)) \in \mathbb{E}$, then $[\mathbf{D}^{-1}\mathbf{S}\theta](\ell(i,\boldsymbol{r}_i))$ will the same as $\theta(i,\boldsymbol{r}_i)$, and hence $\theta^*\mathbf{D}^{-1}\mathbf{S}\theta$ is maximized. Thus, the top eigenvector of $\mathbf{D}^{-1}\mathbf{S}$ contains the "synchronized phase" we are after. We mention that $\mathbf{D}^{-1}\mathbf{S}$ is similar to the Hermitian matrix $\mathbf{D}^{-1/2}\mathbf{S}\mathbf{D}^{-1/2}$, so evaluating its eigenstructure can be numerically efficient. See Section 6 for the numerical performance of this approach.

The synchronization property of CGL has been studied in [5, 20] and its statistical property and robustness behavior in the block random matrix framework have been reported in [27, 28]. In addition, under the manifold setup [60, 61], it asymptotically converges to the heat kernel of the associated connection Laplacian, which top eigenvector-field is the most parallel vector field branded in the manifold structure. We refer the reader to the literature for detail mathematical statements.

The second algorithm we propose has the same flavor, but we consider the amplitude information in a different way compared with (25). Indeed, the amplitude is taken into consideration as a truncation threshold leading to the following relaxation of (22) to estimate the phase. Based on the amplitude, we define a thresholding matrix

$$T_{\mathbf{a}} := \mathrm{diag}(\chi_{\mathbf{a}>\epsilon_a}),$$

where $\epsilon_a \geq 0$ is the threshold chosen by the user, and evaluate the following functional

$$\underset{\zeta\in\mathbb{C}^{Km^2},\,\|\zeta\|=1}{\mathrm{argmax}} \quad \zeta^* T_{\mathbf{a}} P_{\mathbf{FQ}} T_{\mathbf{a}} \zeta,$$

which is equivalent to finding the top eigenvector of the Hermitian matrix $T_{\mathbf{a}} P_{\mathbf{FQ}} T_{\mathbf{a}}$. Our second proposed estimator of the phase to the ptychography problem is then the phase of the top eigenvector of $T_{\mathbf{a}} P_{\mathbf{FQ}} T_{\mathbf{a}}$. We call this approach to the *truncation phase synchronization (t-PS)* algorithm. See Section 6 for its numerical performance. This optimization problem is essentially different from (21) due to the thresholding, and this difference plays an essential role in the optimization. Its theoretical property is beyond the scope of this paper and will be reported in another paper.

6. Numerical results

We begin with describing the two lens we use. The first one is a typical illumination probe in an experimental system. The illuminating beam is formed by a small lens, with a dark "beam-stop" to sort-out harmonic contaminations formed by diffractive Fresnel lenses, represented by a circular aperture in the Fourier domain. The lens is denoted as ω_{s} and is illustrated in the top row of Figure 5. The second is a *band-limited random (BLR) lens*, denoted as ω_{BLR} which we describe now. Note that a small lens can only "connect" Fourier frequencies that are close together, while a wide lens produces a small illumination and the illumination scheme can only connect frames that are near each other. The intuition behind the

synchronization analysis of the ptychographic problem leads us to suggest a different lens that enables to connect pixels across the data space. Experimental observations confirm that diffuse probes [36, 48], and wide apertures [47] produce better results in ptychography. We design our second lens by setting the amplitude and a random phase of an annular aperture in the Fourier domain, then iteratively adjust the amplitude in real and Fourier domains to determine a lens with a circular focus and given amplitude. The motivation for the limited size of the focus is to reduce the requirements of the experimental detector response function (such as pixel size). Such lens can be fabricated using lithographic techniques [16]. The second lens is described in the bottom row of Figure 5.

We begin with a small problem – an object of size 256×256 pixels, that is $n = 256$, shown in Figure 6, using the lens $\omega_{\rm s}$. We collect $k = 32 \times 32$ frames, with 128×128 pixels, that is $m = 128$. The frames are distributed uniformly to cover the object: we start by setting the positions $\boldsymbol{x}_i = (x_i, y_j)$ on a square grid lattice, with $x_i - x_{i+1} = \Delta x$ and $y_i - y_{i+1} = \Delta y$. In this first experiment, we take $\Delta x = \Delta y = 8$. Then we shear odd rows, that is, x_i, by $\Delta x/2$ and perturb the position by a random perturbation randomly sampled uniformly from $[-1.5, +1.5]$ in both x_i and y_i. Fractional pixel shifts are accounted by interpolation of the illumination matrix. We use the following algorithms

AP

(1) start with random object: $\zeta^{(0)} = \mathbf{FQ}(\text{random})$;
(2) compute $\zeta^{(\ell)} = [P_{\mathbf{FQ}} P_{\mathbf{a}}]^{\ell} \zeta^{(0)}$, $\ell \geq 1$ chosen by the user;
(3) $\psi^{(\ell)}_{\rm AP} = (\mathbf{Q}^*\mathbf{Q})^{-1}\mathbf{Q}^*\mathbf{F}\zeta^{(\ell)}$.

CGL-PS

(1) find the largest eigenvalue v_0 of the CGL matrix $\mathbf{D}^{-1}\mathbf{S}$;
(2) $\psi_{\text{CGL-PS}} = (\mathbf{Q}^*\mathbf{Q})^{-1}\mathbf{Q}^*\mathbf{F}P_{\mathbf{a}}v_0$.

t-PS

(1) find the largest eigenvalue v_0 of the phase synchronization matrix $T_{\mathbf{a}}P_{\mathbf{FQ}}T_{\mathbf{a}}$, where $T_{\mathbf{a}} = \text{diag}(\chi_{\mathbf{a}>\epsilon_a})$;
(2) $\psi_{\text{t-PS}} = (\mathbf{Q}^*\mathbf{Q})^{-1}\mathbf{Q}^*\mathbf{F}P_{\mathbf{a}}v_0$.

CGL-PS+AP

(1) find the largest eigenvalue v_0 of $\mathbf{D}^{-1}\mathbf{S}$;
(2) compute $\zeta^{(\ell)} = [P_{\mathbf{FQ}}P_{\mathbf{a}}]^{\ell}v_0$, $\ell \geq 1$ chosen by the user;
(3) $\psi^{(\ell)}_{\text{CGL-PS+AP}} = (\mathbf{Q}^*\mathbf{Q})^{-1}\mathbf{Q}^*\mathbf{F}\zeta^{(\ell)}$.

t-PS+AP

(1) find the largest eigenvalue v_0 of $T_{\mathbf{a}}P_{\mathbf{FQ}}T_{\mathbf{a}}$;
(2) compute $\zeta^{(\ell)} = [P_{\mathbf{FQ}}P_{\mathbf{a}}]^{\ell}v_0$, $\ell \geq 1$ chosen by the user;
(3) $\psi^{(\ell)}_{\text{t-PS+AP}} = (\mathbf{Q}^*\mathbf{Q})^{-1}\mathbf{Q}^*\mathbf{F}\zeta^{(\ell)}$.

Also, the convergence is monitored by:

$\varepsilon^{(\ell)}_{\mathbf{a}} := \frac{1}{\|\mathbf{a}\|}\|[I - P_{\mathbf{a}}]\zeta^{\ell}\|$,
$\varepsilon^{(\ell)}_{\mathbf{FQ}} := \frac{1}{\|\mathbf{a}\|}\|[I - P_{\mathbf{FQ}}]\zeta^{\ell}\|$
$\varepsilon^{(\ell)}_{\mathbf{aFQ}} := \frac{1}{\|\mathbf{a}\|}\|[P_{\mathbf{a}} - P_{\mathbf{FQ}}]\zeta^{\ell}\|$.
$\varepsilon^{(\ell)}_{0} := \frac{1}{\|\mathbf{a}\|}\min_t \|\zeta^{\ell} - e^{it}\mathbf{FQ}\psi_0\|$,
$\varepsilon^{(\ell)}_{\Delta\ell} := \frac{1}{\|\mathbf{a}\|}\|\zeta^{\ell} - \zeta^{\ell+1}\|$,

The result of the first experiment is shown in Figure 6.

We repeat the same experiment with an image of a self-assembled cluster of 50 nm colloidal gold nanoparticles obtained by Scanning Electron Microscopy. To produce a complex image, the gray-scale value are projected onto a circle in the complex plane. The size is 256×256 pixels and we use the lens $\omega_{\rm s}$. The result of the second experiment is shown in Figure 7.

A few things to notice from Figures (6,7): the first is that $\|\zeta^{\ell} - \zeta^{\ell+1}\| = \varepsilon^{(\ell)}_{\Delta\ell}\|\mathbf{a}\|$ does not decrease monotonically, and the second is that the eigenvector with the largest eigenvalue of $T_{\mathbf{a}}P_{\mathbf{FQ}}T_{\mathbf{a}}$ is already quite a good image, and last, the convergence rate is similar but t-PS produces a better start. Also note that typically $\epsilon_{\mathbf{FQ}}, \epsilon_a, \epsilon_{aQ}$ are very similar and overlap.

We compare these two illumination functions, $\omega_{\rm s}$ and $\omega_{\rm BLR}$, with the same two objects with the same parameters as before. The results are shown in Figure 8 and Figure 9. Clearly t-PS produces a better start with the new illumination. In this example, such better start also leads to higher rate of convergence.

Yet next, we test the algorithm in a larger problem, an object of 512×512 pixels, that is $n = 512$, with the same lens size (128×128). We increase the field of view of the illumination scheme with increased spacing among frames $\Delta x = 16$ and $\Delta y = 16$. One of the issues of projection algorithms such as AP is that frames that are far apart communicate very weakly with each other, this leads to slower rate of convergence. This is an issue when we are limited by the number of iterations, due to high data rate and finite computational resources. In Figure 10 we show the result of 101 iterations of AP with holes in the scarf, while t-PS gives a good initial start that leads to improved SNR. Notice that the hole in the scarf and other defects produced by AP alone.

In our last numerical experiment, we introduce new algorithms that lead to over $20\times$ acceleration in the rate of convergence. First, we use the RAAR algorithm [46] described below which is popular among the optical community [19] (using RAAR in combination with a shrink-wrap algorithm [50] to enforce sparsity) because it often leads to improved convergence rate. Second we introduce a frame-wise synchronization technique to adjust the phase of every frame at every iteration based on existing frame-wide local information.

RAAR

(1) start with random object $\zeta^{(0)} = \mathbf{FQ}(\text{random})$
(2) compute $\zeta^{(\ell)} = [2\beta P_{\mathbf{FQ}} P_{\mathbf{a}} + (1-2\beta)\beta P_{\mathbf{a}} + \beta(P_{\mathbf{FQ}} - I)]^{\ell}\zeta^{(0)}$ where $\beta = 0.9$ and $\ell \geq 1$ is chosen by the user.
(3) $\psi^{(\ell)}_{\rm RAAR} = (\mathbf{Q}^*\mathbf{Q})^{-1}\mathbf{Q}^*\mathbf{F}\zeta^{(\ell)}$,

t-PS+RAAR

(1) find the largest eigenvalue v_0 of the kernel $T_{\mathbf{a}} P_{\mathbf{FQ}} T_{\mathbf{a}}$
(2) compute $\zeta^{(\ell)} = [2\beta P_{\mathbf{FQ}} P_{\mathbf{a}} + (1-2\beta)\beta P_{\mathbf{a}} + \beta(P_{\mathbf{FQ}} - I)]^{\ell} P_{\mathbf{a}} v_0$, where $\beta = 0.9$ and $\ell \geq 1$ is chosen by the user;
(3) $\psi^{(\ell)}_{\text{t-PS+RAAR}} = (\mathbf{Q}^*\mathbf{Q})^{-1}\mathbf{Q}^*\mathbf{F}\zeta^{(\ell)}$,

t-PS+synchro-RAAR

(1) t-PS:
find the largest eigenvalue v_0 of the kernel $T_{\mathbf{a}} P_{\mathbf{FQ}} T_{\mathbf{a}}$. Start
$$\zeta^{(0)} = P_{\mathbf{FQ}} P_{\mathbf{a}} v_0;$$
(2) frame-wise synchronization:
find the smallest eigenvalue $\xi^{(l)} \in \mathbb{C}^K$ of the matrix $H^{(l)}$ of size $K \times K$ where the (i,j)-th entry is
$$H^{(l)}_{i,j} := (P_{\mathbf{a}}\zeta^{(l)*})_{(i)} \left[\delta_{i,j} I_{m^2} - F\mathbf{Q}_{(i)}(\mathbf{Q}^*\mathbf{Q})^{-1}\mathbf{Q}^*_{(j)}F^*\right](P_{\mathbf{a}}\zeta^{(l)})_{(j)}$$
and replace $P_{\mathbf{FQ}}$ by
$$P^{(l)}_{\mathbf{FQ}} := \operatorname{diag}\left(\mathbf{B}\frac{\xi^{(l)}}{|\xi^{(l)}|}\right)^* P_{\mathbf{FQ}} \operatorname{diag}\left(\mathbf{B}\frac{\xi^{(l)}}{|\xi^{(l)}|}\right),$$
where $\mathbf{B}$ is a $K \times K$ diagonal block matrix with its diagonal the $m^2 \times 1$ row vector $\mathbf{1}^T$;
(3) RAAR with $P^{(l)}_{\mathbf{FQ}}$:
$$\zeta^{(l)} = [2\beta P^{(l-1)}_{\mathbf{FQ}} P_{\mathbf{a}} + (1-2\beta)\beta P_{\mathbf{a}} + \beta(P^{(l-1)}_{\mathbf{FQ}} - I)]\zeta^{(l-1)};$$
where $\beta = 0.9$;
(4) repeat (2)-(5) $\ell \geq 1$ steps until convergences or maximum iteration, where ℓ is determined by the user;
(5) $\psi^{(\ell)}_{\text{t-PS+synchro-RAAR}} = (\mathbf{Q}^*\mathbf{Q})^{-1}\mathbf{Q}^*\mathbf{F}\zeta^{(\ell)}$,

The frame-wise synchronization, step (2), is motivated by the augmented approach [51]. We estimate a phase factor for each frame based on the existing phase estimator of each frames, which leads to long-range

phase synchronization across the image. Indeed, we consider

$$\underset{\xi\in\mathbb{C}^K;\,|\xi|=\mathbf{1}}{\operatorname{argmin}} \|(I-P_{\mathbf{FQ}})\operatorname{diag}(P_{\mathbf{a}}\zeta^{(l)})\mathbf{B}\xi\|,$$

which is relaxed by finding the smallest eigenvector of H, that is, H comes from expanding the functional $\|(I-P_{\mathbf{FQ}})\operatorname{diag}(P_{\mathbf{a}}\zeta^{(l)})\mathbf{B}\xi\|^2$. The estimated frame-wise phase corrector $\xi^{(l)}$ are thus distributes to all the pixels by $\mathbf{B}$. We mention that this frame-wise synchronization can be justified by realizing that at each iteration, $H^{(l)}$ can be understood as the CGL built from the graph associated with the illumination windows so that the estimated frame-wise phases are synchronized according to the CGL $H^{(l)}$. Thus, this frame-wise phase estimation leads to the *long range* phase synchronization. The nomination of "synchro-RAAR" is to emphasize that we do not use $P_{\mathbf{FQ}}$ in the ordinary RAAR step but use the frame-wise synchronized $P_{\mathbf{FQ}}$, that is, $P_{\mathbf{FQ}}^{(l-1)}$ in the update. We tested these algorithms, as well as the AP and t-PS+AP algorithms, on the same data setup in Figure 10, and the convergence results of different algorithms are shown in Figure 11 for comparison. Notice the change of scale in the last plot, where convergence is over 40× faster than the AP algorithm.

7. Conclusion

In this paper, we demonstrate the global convergence of the alternating projection (AP) algorithm to the unique solution up to a global phase factor, and apply it to the ptychographic imaging problem. To be more precise, we have shown that the AP algorithm gives the inverse transform of the phase retrieval problem when the frame is generic. We also survey the intimate relationship between the AP algorithm and the notion of phase synchronization (PS) and propose two algorithm, CGL-PS and t-PS, to quickly construct an accurate initial guess for the AP algorithm for large scale diffraction data problems. In addition, by combining the RAAR algorithm with the frame-wise synchronization, the convergence is over 40× faster than the AP algorithm and is about 10× faster than the RAAR algorithm.

There are several problems left unanswered in this paper. We mention at least the following four directions. First, how to design the best lens and illumination scheme so that we can obtain an accurate reconstruction for the real samples; given a detector, with a limited rate, dynamic range and response function, what is the best scheme to encode more information per detector channel. Second, the noise influence on the convergence behavior needs further investigation. Experimental uncertainties include not only photon-counting statistics but also perturbations of the lens [64, 63, 31], illumination scheme (positions), incoherent measurements, detector response and discretization, time dependent fluctuations, etc. Third, spectral methods such as the proposed algorithms in this paper (CGL-PS and t-PS) have the potential to be scaled up on high-performance computing architectures to handle the big imaging data in the coming new light source era [18, 9]. Last, although RAAR, synchro-RAAR and other iterative schemes perform well in practice, their convergence behavior needs to be further studied. Can we design better iterative methods based on our findings that exploit phase synchronization schemes more efficiently?

8. Acknowledgements

This research was supported in part by the Applied Mathematical Sciences subprogram of the Office of Energy Research, U.S. Department of Energy, under contract DE-AC02-05CH11231 (SM) and by AFOSR grant FA9550-09-1-0643 (HT). H.-T. Wu thanks Professor Ingrid Daubechies for the discussion. We acknowledge NVIDIA for providing us with a Tesla K40 GPU for our tests.

References

[1] B. Alexeev, A. S. Bandeira, M. Fickus, and D. G. Mixon. Phase retrieval with polarization. *SIAM Journal on Imaging Sciences*, 2013.

[2] R. Balan, P. Casazza, and D. Edidin. On signal reconstruction without phase. *Appl. Comput. Harmon. Anal.*, 20:345–356, 2006.

[3] R. Balan and Y. Wang. Invertibility and robustness of phaseless reconstruction. *Appl. Comput. Harmon. Anal.*, abs/1308.4718, 2013.

[4] A. S. Bandeira, J. Cahili, D. G. Mixon, and A. A. Nelson. Saving phase: Injectivity and stability for phase retrieval. *Appl. Comput. Harmon. Anal.*, 2013.

[5] A. S. Bandeira, A. Singer, and D. A. Spielman. A Cheeger Inequality for the Graph Connection Laplacian. *SIAM Journal on Matrix Analysis and Applications*, to appear, 2013. arXiv:1204.3873 [math.SP].

[6] R.H.T. Bates. Uniqueness of solutions to two-dimensional fourier phase problems for localized and positive images. *Computer Vision, Graphics, and Image Processing*, 25(2):205 – 217, 1984.

[7] H. H. Bauschke, P. L. Combettes, and D. R. Luke. Phase retrieval, error reduction algorithm, and Fienup variants: a view from convex optimization. *Journal of the Optical Society of America. A, Optics, image science, and vision*, 19(7):1334–45, 2002.

[8] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. The protein data bank. *Nucleic Acids Research*, 28(1):235–242, 2000.

[9] M. Borland. Progress toward an ultimate storage ring light source. *Journal of Physics: Conference Series*, 425(4):042016, 2013.

[10] W. H. Bragg and W. L. Bragg. The Reflection of X-rays by Crystals. *Royal Society of London Proceedings Series A*, 88:428–438, 1913.

[11] W. L. Bragg. The Specular Reflection of X-rays. *Nature*, 90:410, December 1912.

[12] Ch. Broennimann, E. F. Eikenberry, B. Henrich, R. Horisberger, G. Huelsen, E. Pohl, B. Schmitt, C. Schulze-Briese, M. Suzuki, T. Tomizaki, H. Toyokawa, and A. Wagner. The pilatus 1m detector. *Journal of Synchrotron Radiation*, 13(2):120–130, 2006.

[13] Yu.M. Bruck and L.G. Sodin. On the ambiguity of the image reconstruction problem. *Optics Communications*, 30(3):304–308, 1979.

[14] E. J. Candes, Y. C. Eldar, T. Strohmer, and V. Voroninski. Phase retrieval via matrix completion. *SIAM Journal on Imaging Sciences*, 6(1):199–225, 2013.

[15] E. J. Candes, T. Strohmer, and V. Voroninski. Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. *Communications on Pure and Applied Mathematics*, 66(8):1241–1274, 2013.

[16] W. Chao, J. Kim, S. Rekawa, P. Fischer, and E.H. Anderson. Demonstration of 12 nm resolution fresnel zone plate lens based soft x-ray microscopy. *Opt Express*, 17:17669–77, 2009.

[17] H. N. Chapman. Phase-retrieval x-ray microscopy by wigner -distribution deconvolution. *Ultramicroscopy*, 66:153–172, 1996.

[18] H. N. Chapman. X-ray imaging beyond the limits. *Nat Mater*, 8(4):299–301, 2009.

[19] H. N. Chapman, A. Barty, S. Marchesini, A. Noy, S. P. Hau-Riege, C. Cui, M. R. Howells, R. Rosen, H. He, J. C.H. Spence, et al. High-resolution ab initio three-dimensional x-ray diffraction microscopy. *JOSA A*, 23(5):1179–1200, 2006.

[20] F. Chung, W. Zhao, and M. Kempton. Ranking and sparsifying a connection graph. In Anthony Bonato and Jeannette Janssen, editors, *Algorithms and Models for the Web Graph*, volume 7323 of *Lecture Notes in Computer Science*, pages 66–77. Springer Berlin Heidelberg, 2012.

[21] R.J. Collier, C.B. Burckhardt, and L.H. Lin. *Optical holography*. Student editions. Academic Press, 1971.

[22] P. L. Combettes and H. J. Trussell. Method of successive projections for finding a common point of sets in metric spaces. *Journal of Optimization Theory and Applications*, 67(3):487–507, December 1990.

[23] P. Cramer, D. A. Bushnell, and R. D. Kornberg. Structural basis of transcription: Rna polymerase ii at 2.8 ngstrom resolution. *Science*, 292(5523):1863–1876, 2001.

[24] M. Dierolf, A. Menzel, P. Thibault, P. Schneider, C. M. Kewish, R. Wepf, O. Bunk, and F. Pfeiffer. Ptychographic x-ray computed tomography at the nanoscale. *Nature*, 467(7314):436–439, 2010.

[25] D. Doering, Y.-D. Chuang, N. Andresen, K. Chow, D. Contarato, C. Cummings, E. Domning, J. Joseph, J. S. Pepper, B. Smith, G. Zizka, C. Ford, W. S. Lee, M. Weaver, L. Patthey, J. Weizeorick, Z. Hussain, and P. Denes. Development of a compact fast ccd camera and resonant soft x-ray scattering endstation for time-resolved pump-probe experiments. *Review of Scientific Instruments*, 82(7):073303, 2011.

[26] M. Eckert. Disputed discovery: the beginnings of X-ray diffraction in crystals in 1912 and its repercussions. *Acta Crystallographica Section A*, 68(1):30–39, 2012.

[27] N. El Karoui and H.-T. Wu. Vector diffusion maps and random matrices with random blocks. *ArXiv e-prints*, 2013. arXiv:1310.0188 [math.PR].

[28] N. El Karoui and H.-T. Wu. Vector diffusion maps is robust to noise. *In preparation*, 2013.

[29] R. Falcone, C. Jacobsen, J. Kirz, S. Marchesini, D. Shapiro, and J. Spence. New directions in x-ray microscopy. *Contemporary Physics*, 52(4):293–318, 2011.

[30] A. Fannjiang and W. Liao. Phase retrieval with random phase illumination. *J. Opt. Soc. Am. A*, 29(9):1847–1859, 2012.

[31] A. Fannjiang and W. Liao. Fourier phasing with phase-uncertain mask. *ArXiv e-prints*, abs/1212.3858, 2013.

[32] J. R. Fienup. Phase retrieval algorithms: a comparison. *Appl. Opt.*, 21:2758–2769, 1982.

[33] 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(10):1747–1767, 1993.

[34] G.B. Folland. *Harmonic Analysis in Phase Space*. Princeton, 1989.

[35] R.W. Gerchberg and W.O. Saxton. Phase determination for image and diffraction plane pictures in the electron microscope. *Optik*, 34(3):275–284, 1971.

[36] M. Guizar-Sicairos, M. Holler, A. Diaz, J. Vila-Comamala, O. Bunk, and A. Menzel. Role of the illumination spatial-frequency spectrum for ptychography. *Phys. Rev. B*, 86:100103, Sep 2012.

[37] J. Guo. *X-Rays in Nanoscience: Spectroscopy, Spectromicroscopy, and Scattering Techniques*. Wiley. com, 2010.

[38] P. W. Hawkes and J. C. H. Spence, editors. *Science of microscopy*. Springer, New York, 2007.

[39] M. Hayes. The reconstruction of a multidimensional sequence from the phase or magnitude of its Fourier transform. *Transactions on Acoustics, Speech and Signal Processing*, 30:294–302, 1982.

[40] R. Hegerl and W. Hoppe. Dynamic theory of crystalline structure analysis by electron diffraction in inhomogeneous primary wave field. *Berichte Der Bunsen-Gesellschaft Fur Physikalische Chemie*, 74:1148, 1970.

[41] W. Hoppe. Beugung im inhomogenen Primärstrahlwellenfeld. I. Prinzip einer Phasenmessung von Elektronenbeungungsinterferenzen. *Acta Crystallographica Section A*, 25(4):495–501, 1969.

[42] K. Kreutz-Delgado. The complex gradient operator and the CR-calculus. *UCSD*, 2003.

[43] C. L. Lawson, M. L. Baker, C. Best, C. Bi, M. Dougherty, P. Feng, G. van Ginkel, B. Devkota, I. Lagerstedt, S. J. Ludtke, R. H. Newman, T. J. Oldfield, I. Rees, G. Sahni, R. Sala, S. Velankar, J. Warren, J. D. Westbrook, K. Henrick, G. J. Kleywegt, H. M. Berman, and W. Chiu. Emdatabank.org: unified data resource for cryoem. *Nucleic Acids Research*, 39(suppl 1):D456–D464, 2011.

[44] A. Levi and H. Stark. Image restoration by the method of generalized projections with application to restoration from magnitude. *J. Opt. Soc. Am. A*, 1(9):932–943, 1984.

[45] D. R. Luke, J. V. Burke, and R. G. Lyon. Optical wavefront reconstruction: Theory and numerical methods. *SIAM review*, 44(2):169–224, 2002.

[46] R. Luke. Relaxed averaged alternating reflections for diffraction imaging. *Inverse Problems*, 21:37–50, 2005.

[47] A. M. Maiden, M. J. Humphry, F. Zhang, and J. M. Rodenburg. Superresolution imaging via ptychography. *JOSA A*, 28(4):604–612, 2011.

[48] A.M. Maiden, G.R. Morrison, B. Kaulich, A. Gianoncelli, and J.M. Rodenburg. Soft x-ray spectromicroscopy using ptychography with randomly phased illumination. *Nature communications*, 4:1669, 2013.

[49] S. Marchesini, S. Boutet, and et al. Sakdinawat. Massively parallel X-ray holography. *Nature Photonics*, 2, September 2008.

[50] S. Marchesini, H. He, H. N. Chapman, S. P. Hau-Riege, A. Noy, M. R. Howells, U. Weierstall, and J. C.H. Spence. X-ray image reconstruction from a diffraction pattern alone. *Physical Review B*, 68(14):140101, 2003.

[51] S. Marchesini, A. Schirotzek, C. Yang, H.-T. Wu, and F. Maia. Augmented projections for ptychographic imaging. *Inverse Problems*, 29(11):115009, 2013.

[52] A. Momose, T. Takeda, Y. Itai, and K. Hirano. Phase–contrast x–ray computed tomography for observing biological soft tissues. *Nature medicine*, 2(4):473–475, 1996.

[53] P. D. Nellist, B. C. McCallum, and J. M. Rodenburg. Resolution beyond the 'information limit' in transmission electronmicroscopy. *Nature*, 374:630–632, 1995.

[54] K. A. Nugent. Coherent methods in the x-ray sciences. *Advances in Physics*, 59(1):1–99, 2010.

[55] L. Pauling and M.D. Shappell. The crystal structure of bixbyite and the c-modification of the sesquioxides. *Z. Kristallogr*, 75(1-2):128–142, 1930.

[56] F. Pfeiffer, T. Weitkamp, O. Bunk, and C. David. Phase retrieval and differential phase-contrast imaging with low-brilliance x-ray sources. *Nature physics*, 2(4):258–261, 2006.

[57] J. M. Rodenburg. Ptychography and related diffractive imaging methods. volume 150 of *Advances in Imaging and Electron Physics*, chapter Ptychography and Related Diffractive Imaging Methods, pages 87–184. Elsevier, 2008.

[58] J.L.C. Sanz. Mathematical considerations for the problem of Fourier transform phase retrieval from magnitude. *SIAM Journal on Applied Mathematics*, 45(4):651–664, 1985.

[59] A. Singer. Angular Synchronization by Eigenvectors and Semidefinite Programming. *Appl. Comput. Harmon. Anal.*, 30(1):20–36, 2011.

[60] A. Singer and H.-T. Wu. Vector diffusion maps and the connection Laplacian. *Comm. Pure Appl. Math.*, 65(8):1067–1144, 2012.

[61] A. Singer and H.-T. Wu. Spectral convergence of the connection laplacian from random samples. *ArXiv e-prints*, 2013. arXiv:1306.1587v1 [math.NA].

[62] P. Thibault. *Algorithmic methods in diffraction microscopy*. PhD thesis, Cornell University, 2007.

[63] P. Thibault, M. Dierolf, O. Bunk, A. Menzel, and F. Pfeiffer. Probe retrieval in ptychographic coherent diffractive imaging. *Ultramicroscopy*, 109(4):338–343, 2009.

[64] P. Thibault, M. Dierolf, A. Menzel, O. Bunk, C. David, and F. Pfeiffer. High-Resolution scanning x-ray diffraction microscopy. *Science*, 321(5887):379–382, 2008.

[65] I. Waldspurger, A. dAspremont, and S. Mallat. Phase recovery, maxcut and complex semidefinite programming. *Mathematical Programming*, 2013.

[66] Y. Wang and Z. Xu. Phase retrieval for sparse signals. *ArXiv e-prints*, abs/1310.0873, 2013.

[67] J. D. Watson and F. H.C. Crick. Molecular structure of nucleic acids. *Nature*, 171(4356):737–738, 1953.

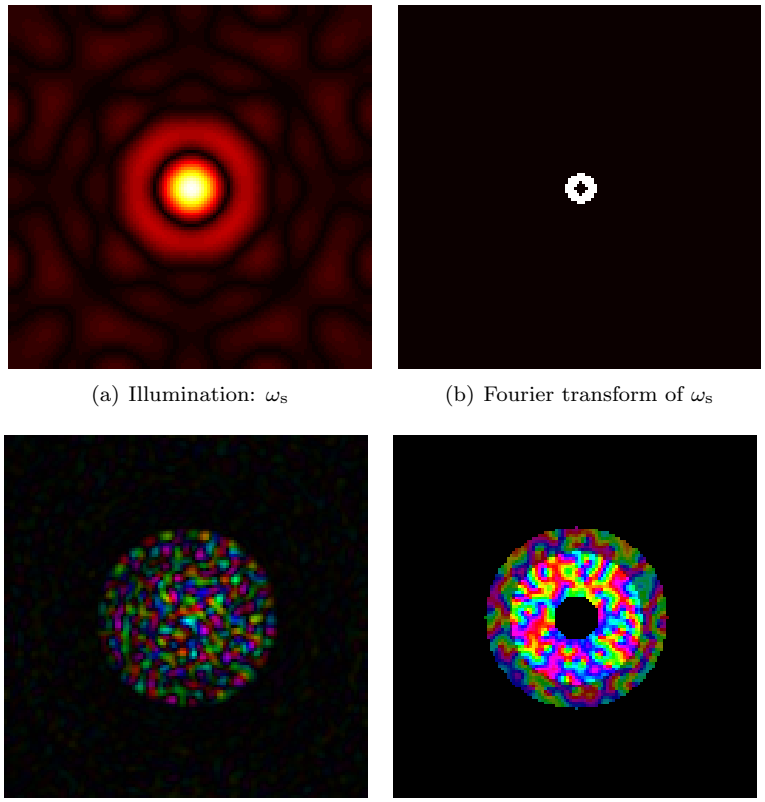
[68] Z. Wen, C. Yang, X Liu, and S. Marchesini. Alternating direction methods for classical and ptychographic phase retrieval. *Inverse Problems*, 28(11):115010, 2012.

[69] C. Yang, J. Qian, A. Schirotzek, F. Maia, and S. Marchesini. Iterative Algorithms for Ptychographic Phase Retrieval. *ArXiv e-prints*, 2011.

Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, CA 94720
E-mail address: smarchesini@lbl.gov

Department of Mathematics, Princeton University, Princeton, NJ 08540
E-mail address: hauwu@stanford.edu

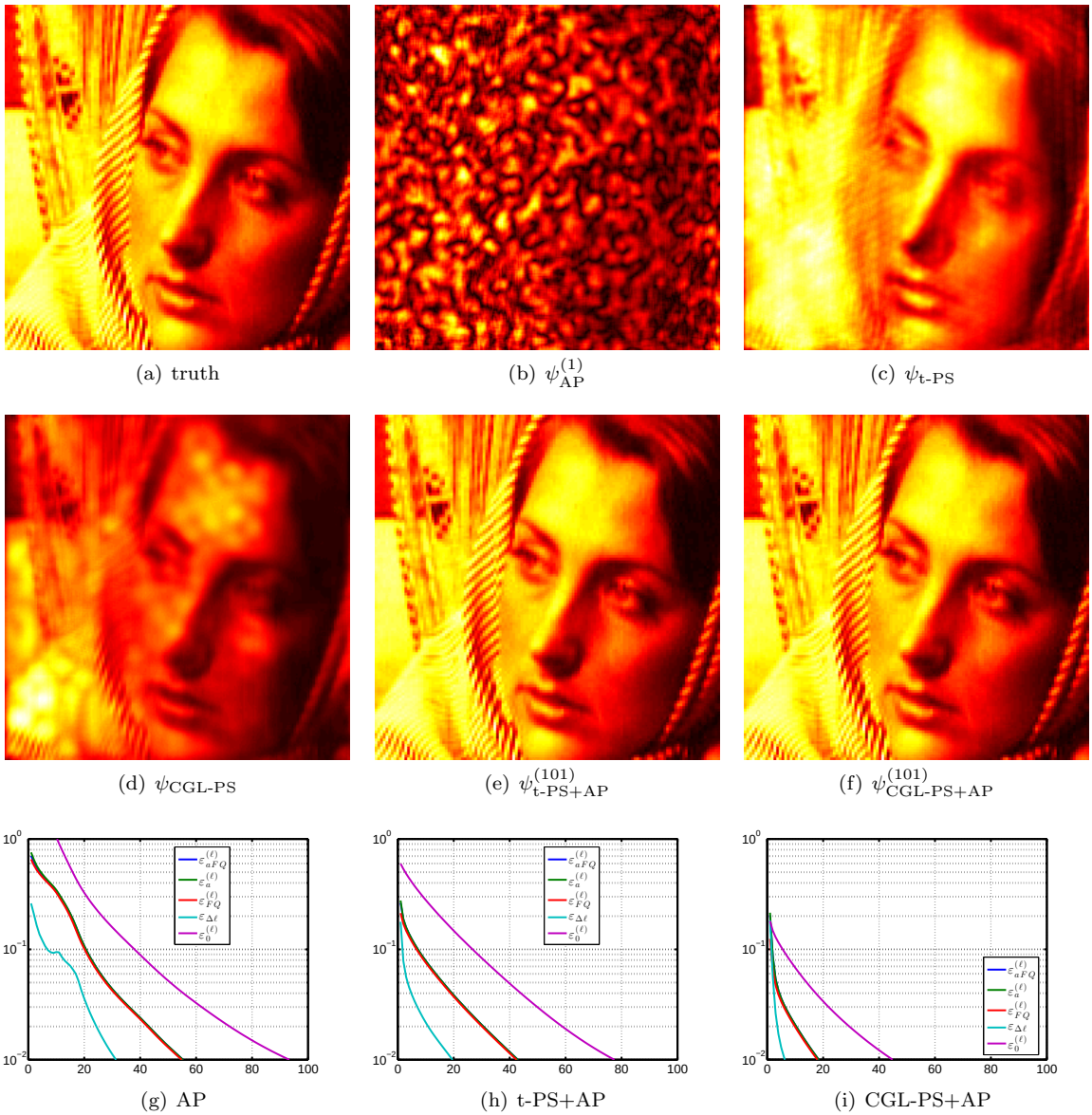
Department of Mathematics, Stanford University, Stanford, CA 94305
E-mail address: hauwu@stanford.edu



(a) Illumination: ω_{s} (b) Fourier transform of ω_{s}

(c) Illumination: ω_{BLR} (d) Fourier transform of ω_{BLR}

FIGURE 5. Illumination functions and their Fourier transform. The top row is the small lens ω_s and the bottom row is the band-limited random (BLR) lens ω_{BLR}. The phase of the complex illumination is represented in color.

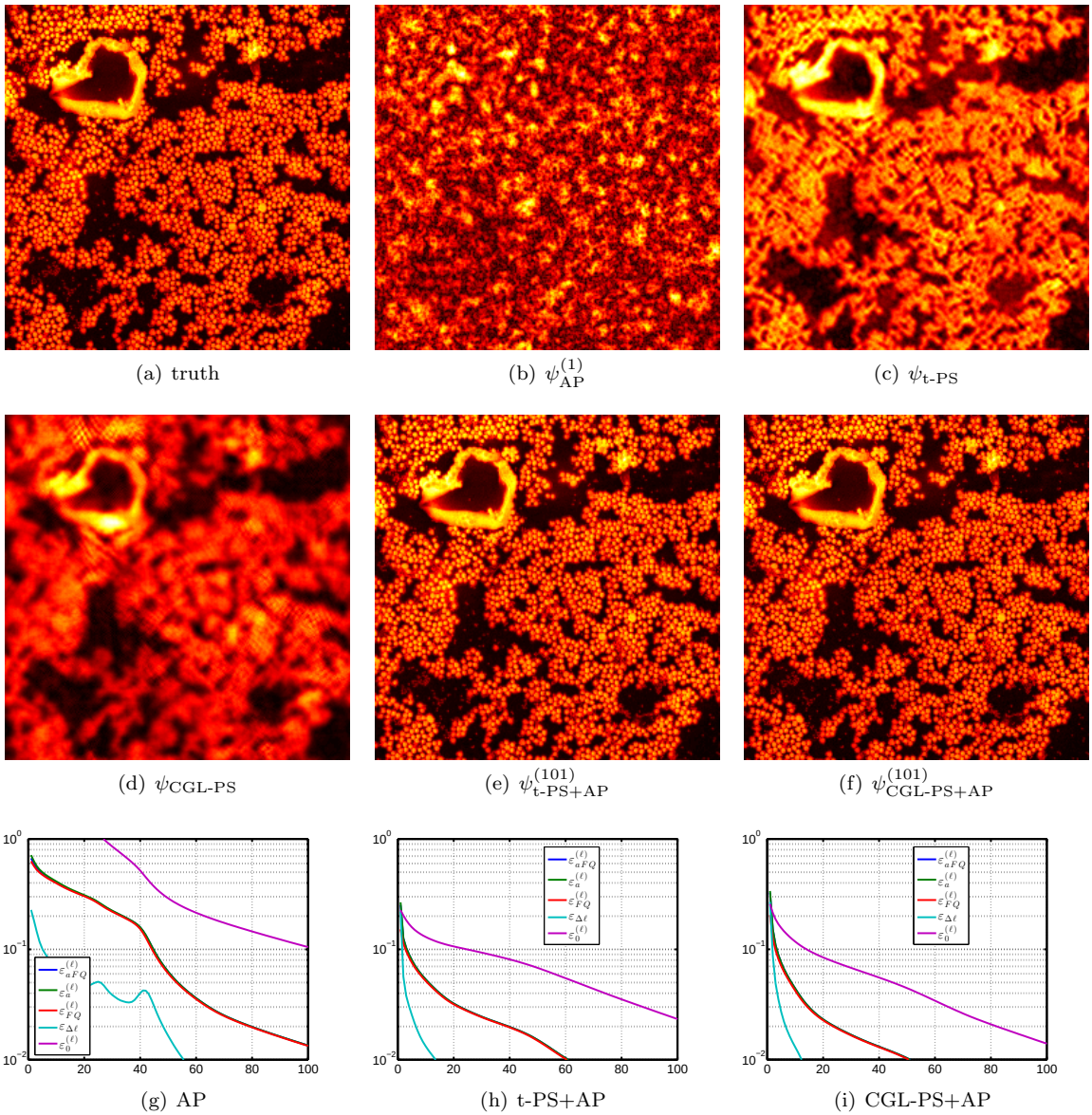


(a) truth (b) $\psi^{(1)}_{\text{AP}}$ (c) $\psi_{\text{t-PS}}$

(d) $\psi_{\text{CGL-PS}}$ (e) $\psi^{(101)}_{\text{t-PS+AP}}$ (f) $\psi^{(101)}_{\text{CGL-PS+AP}}$

(g) AP (h) t-PS+AP (i) CGL-PS+AP

FIGURE 6. Results on the Barbara image of size 256×256 with ω_{s} lens and the illumination scheme described in the content ($\Delta x = \Delta y = 8$ with perturbation). For the t-PS algorithm, we set ϵ_a so that it selects 98% of the highest values of $\mathbf{a}$. (a) the ground truth; (b) $\psi^{(1)}_{\text{AP}}$; (c) $\psi_{\text{t-PS}}$; (d) $\psi_{\text{CGL-PS}}$; (e) $\psi^{(101)}_{\text{t-PS+AP+AP}}$; (f) $\psi^{(101)}_{\text{GCL-PS+AP}}$; (g) convergence of AP with a random start; (h) convergence of AP with the t-PS start; (i) convergence of AP with the CGL-PS start.

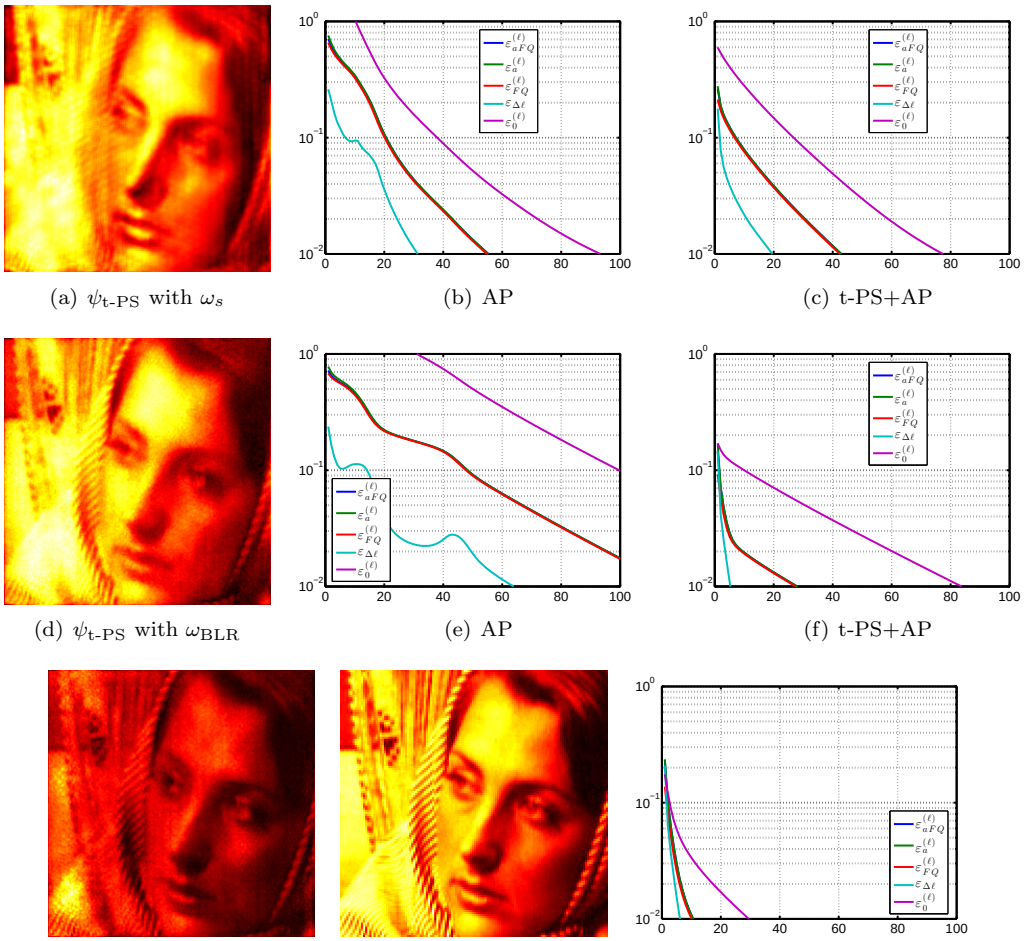


(a) truth (b) $\psi_{\mathrm{AP}}^{(1)}$ (c) $\psi_{\text{t-PS}}$

(d) $\psi_{\text{CGL-PS}}$ (e) $\psi_{\text{t-PS+AP}}^{(101)}$ (f) $\psi_{\text{CGL-PS+AP}}^{(101)}$

(g) AP (h) t-PS+AP (i) CGL-PS+AP

FIGURE 7. Results on the gold ball image of size 256×256 with ω_{s} lens and the illumination scheme described in the content ($\Delta x = \Delta y = 8$ with perturbation). For the t-PS algorithm, we set ϵ_a so that it selects 80% of the highest values of $\mathbf{a}$. (a) the ground truth; (b) $\psi_{\mathrm{AP}}^{(1)}$; (c) $\psi_{\text{t-PS}}$; (d) $\psi_{\text{CGL-PS}}$; (e) $\psi_{\text{t-PS+AP+AP}}^{(101)}$; (f) $\psi_{\text{GCL-PS+AP}}^{(101)}$; (g) convergence of AP with a random start; (h) convergence of AP with the t-PS start; (i) convergence of AP with the CGL-PS start.

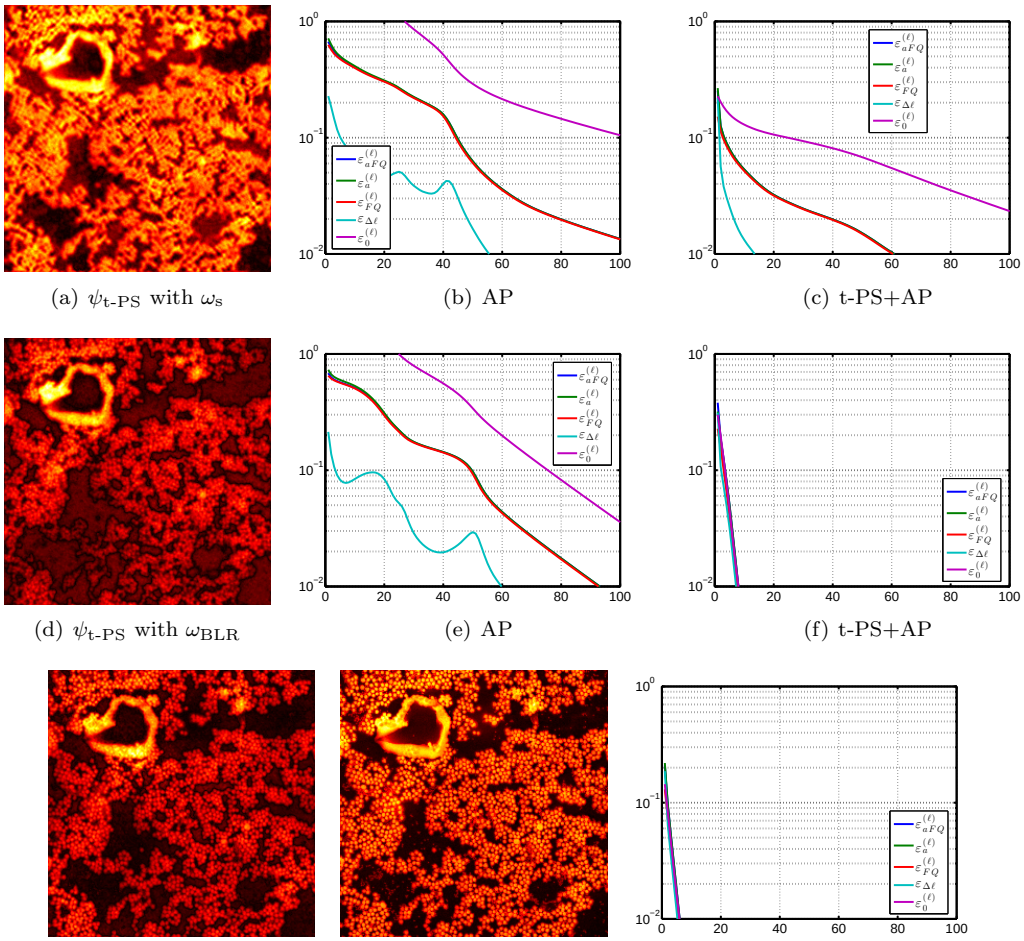


(a) $\psi_{\text{t-PS}}$ with ω_s (b) AP (c) t-PS+AP

(d) $\psi_{\text{t-PS}}$ with ω_{BLR} (e) AP (f) t-PS+AP

(g) $\psi_{\text{CGL-PS}}$ with ω_{BLR} (h) truth (i) CGL-PS+AP

FIGURE 8. Comparison of lens ω_{s} and lens ω_{BLR} on the Barbara image of size 256×256 with ω_{s} lens and the illumination scheme described in the content ($\Delta x = \Delta y = 8$ with perturbation). For the t-PS algorithm, we set ϵ_a so that it selects 98% of the highest values of $\mathbf{a}$. The top row is the result with lens ω_{s}; from left to right: the $\psi_{\text{t-PS}}$, the convergence of the AP algorithm, and the convergence of AP+t-PS algorithm. The middle row is the result with lens ω_{BLR}; from left to right: the $\psi_{\text{t-PS}}$, the convergence of AP with a random start, and the convergence of AP with the t-PS start. The bottom row, from left to right: the $\psi_{\text{CGL-PS}}$, the ground truth, and the convergence of AP with the CGL-PS start.

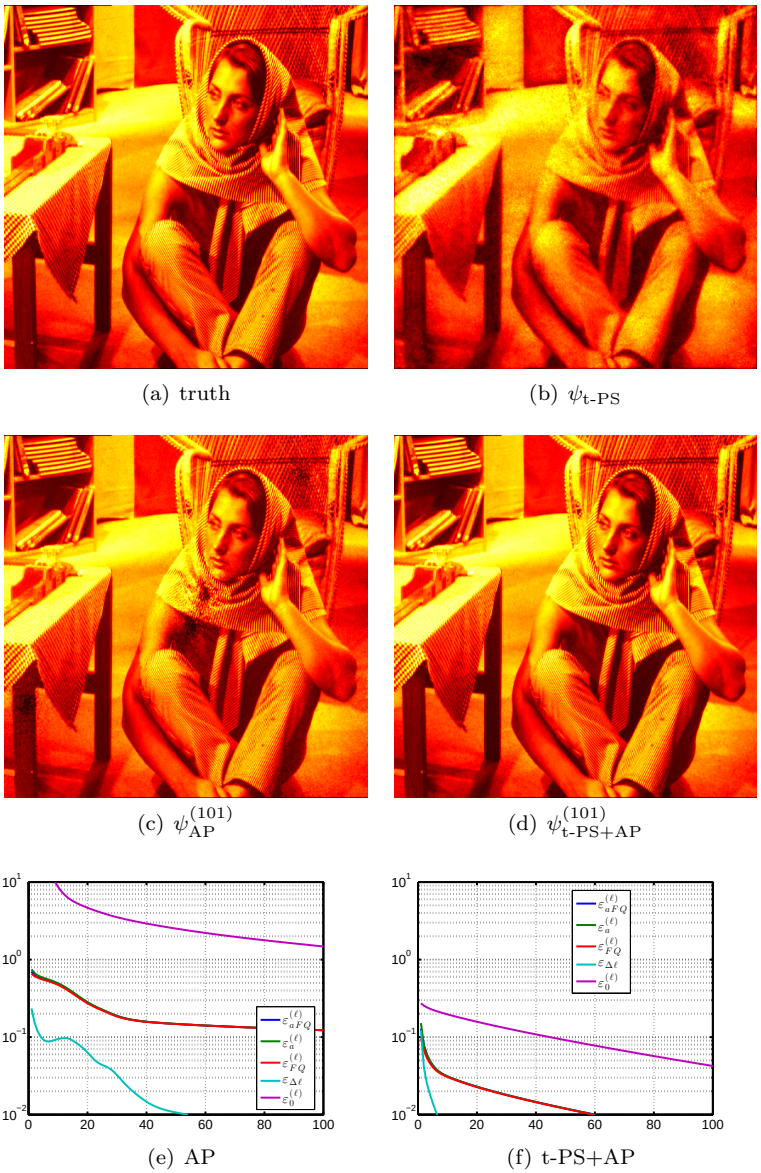


(a) $\psi_{\text{t-PS}}$ with ω_{s} (b) AP (c) t-PS+AP

(d) $\psi_{\text{t-PS}}$ with ω_{BLR} (e) AP (f) t-PS+AP

(g) $\psi_{\text{CGL-PS}}$ with ω_{BLR} (h) truth (i) CGL-PS+AP

FIGURE 9. Comparison of lens ω_{s} and lens ω_{BLR} on the gold ball image of size 256×256 with ω_{s} lens and the illumination scheme described in the content ($\Delta x = \Delta y = 8$ with perturbation). For the t-PS algorithm, we set ϵ_a so that it selects 80% of the highest values of $\mathbf{a}$. The top row is the result with lens ω_{s}; from left to right: the $\psi_{\text{t-PS}}$, the convergence of the AP algorithm, and the convergence of AP+t-PS algorithm. The middle row is the result with lens ω_{BLR}; from left to right: the $\psi_{\text{t-PS}}$, the convergence of AP with a random start, and the convergence of AP with the t-PS start. The bottom row, from left to right: the $\psi_{\text{CGL-PS}}$, the ground truth, and the convergence of AP with the CGL-PS start.

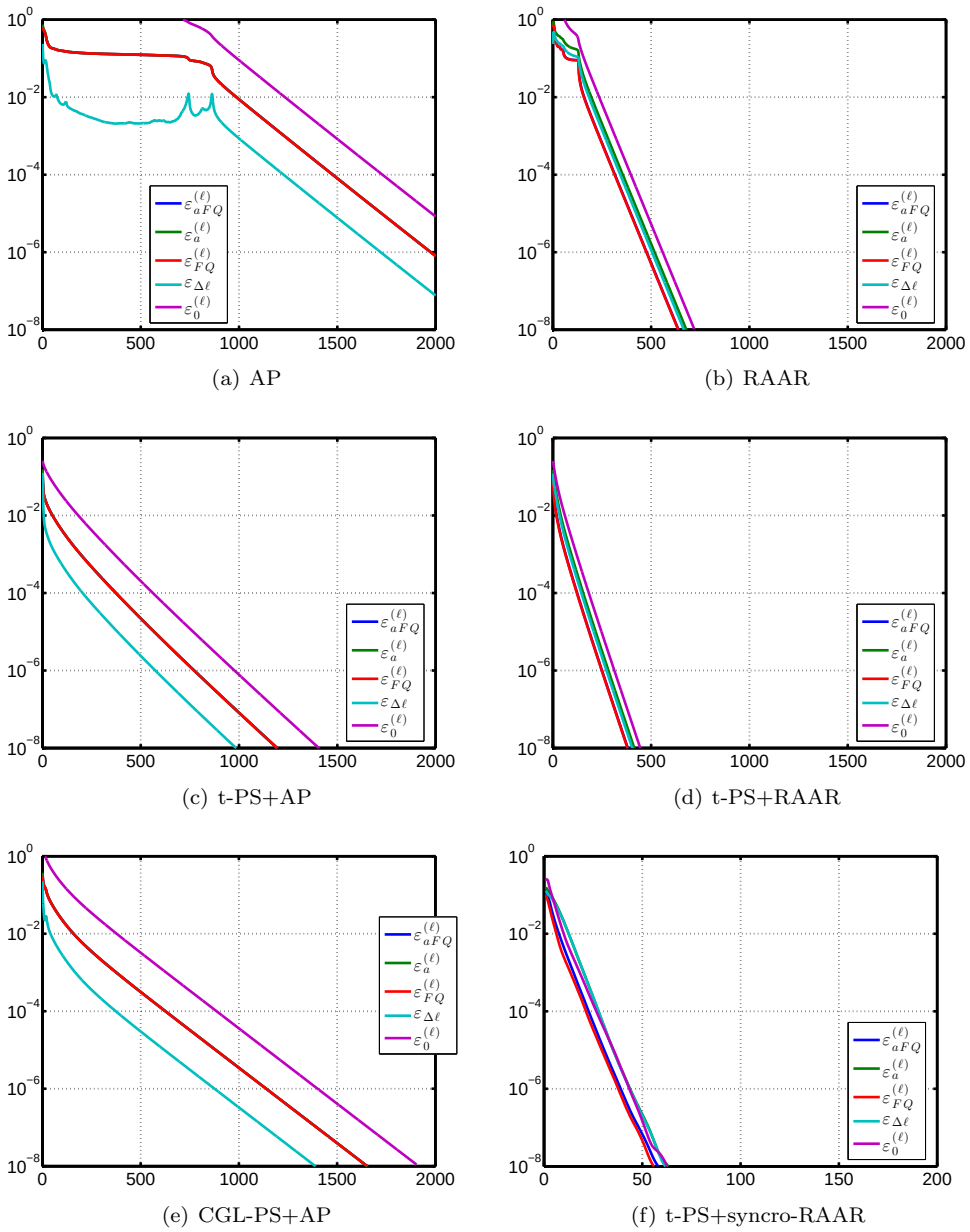


(a) truth (b) $\psi_{\text{t-PS}}$

(c) $\psi_{\text{AP}}^{(101)}$ (d) $\psi_{\text{t-PS+AP}}^{(101)}$

(e) AP (f) t-PS+AP

FIGURE 10. Results on a larger object. The object of interest is the Barbara image of size 512×512 with ω_{BLR} lens and the illumination scheme described in the content ($\Delta x = \Delta y = 16$ with perturbation). For the t-PS algorithm, we set ϵ_a so that it selects 80% of the highest values of $\mathbf{a}$. (a) ground truth; (b) $\psi_{\text{t-PS}}$; (c) $\psi_{\text{AP}}^{(101)}$; (d) $\psi_{\text{t-PS+AP}}^{(101)}$; (e) convergence of AP with random start; (f) convergence of AP with t-PS start. Notice that AP alone produces a hole in the scarf, which may lead the viewer to the wrong interpretation.



(a) AP (b) RAAR

(c) t-PS+AP (d) t-PS+RAAR

(e) CGL-PS+AP (f) t-PS+syncro-RAAR

FIGURE 11. Convergence rate for different algorithms applied to the Barbara image of size 512×512 with ω_{BLR} lens and the illumination scheme described in the content ($\Delta x = \Delta y = 16$ with perturbation). For the t-PS algorithm, we set ϵ_a so that it selects 80% of the highest values of $\mathbf{a}$. (a) the convergence of AP with a random start. Note that it is the zoom out figure of subfigure (e) in Figure 10; (b) the convergence of RAAR with a random start; (c) the convergence of AP with the t-PS start. Note that it is the zoom out figure of the subfigure (f) in Figure 10; (d) the convergence of RAAR with the t-PS start; (e) the convergence AP with the CGL-PS start; (f) the convergence t-PS+synchro-RAAR. Notice the change of scale in the last plot, where convergence is over $40\times$ faster than the AP algorithm and is about $10\times$ faster than the RAAR algorithm.

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.