New Journal of Physics

The open access journal at the forefront of physics

CrossMark

PAPER

Ptychographic phase retrieval by proximal algorithms

Hanfei Yan

National Synchrotron Light Source II, Brookhaven National Laboratory, Upton, NY 11973, United States of America

E-mail: hyan@bnl.gov

Abstract

We derive a set of ptychography phase-retrieval iterative engines based on proximal algorithms originally developed in convex optimization theory, and discuss their connections with existing ones. The use of proximal operator creates a simple frame work that allows us to incorporate the effect of noise from a maximum-likelihood (ML) principle. We focus on three particular algorithms, namely proximal minimization, alternating direction method of multiplier and accelerated proximal gradient (APG). We benchmark their performance with numerical simulations, and discuss their optimal conditions for convergence and accuracy. An experimental dataset is used to demonstrate their effectiveness as well, in which case an array of cubic Au nanoparticles with a size of 50 nm is imaged. We show that with the presence of Poisson noise, a dataset with photon counts up to 10^4 at one detector pixel already requires ML-based methods to achieve a stable convergence. Among the three algorithms derived in this work, APG method is reported first time for its application in ptychographic reconstruction and shows superior performance in terms of both accuracy and convergence rate with a noisy dataset.

1. Introduction

Ptychography is a powerful scanning imaging technique that utilizes advanced mathematical tools to retrieve the missing phase information of the wave-field from a sequence of intensity measurements [1–3]. The attraction of this technique comes from its capability of recovering both the complex-valued probe and object functions, a blind deconvolution, as well as its ability of breaking the resolution barrier set by the focusing optics. It gains increasing popularity in recent years for its robustness in practice and was used successfully for many imaging applications in different fields [4–9]. The major challenge of this technique resides in the fact that the mathematical problem is non-convex and ill posed. Particularly for real-world problems, the experimental data always contains noise and other types of errors, therefore finding a solution optimized globally is extremely difficult, if not completely impossible. A great deal of efforts have been devoted to developing robust ptychographic iterative engines, either based on alternating-projection or gradient-descent methods [3, 10–18]. More complex algorithms that can handle mixed states [19], positioning errors [20, 21], diffraction effect [22, 23] and instability of the probe [24] were developed as well.

For convex optimization problems, a class of algorithms called proximal algorithms have been studied extensively [25]. They turn out to be well-suited for constrained, large-scale and distributed optimization problems, and ptychography falls into this category. In these techniques, the solving process is divided into sub-problems involving the evaluation of the proximal operator, which usually has a closed-form solution. Inspired by these developments, here we propose to combine proximal algorithms and Wirtinger derivatives to create a simple frame work for solving ptychography problems, where either alternating-projection or gradient-descent algorithms can be derived straightforwardly.

We want to emphasize that proximal algorithms are originally developed for convex problems dealing with real-valued numbers, and ptychography problems are non-convex and involves complex-valued numbers. The Wirtinger derivative allows us to use the common rules for differentiation known from real-valued analysis [26], so the developed solving techniques in proximal algorithms can be readily applied. We show some previously

reported ptychography algorithms can be derived in this frame work. Here we do not attempt to provide a rigorous theoretical proof of the overall convergence, but rather offer a heuristic for future work by demonstrating the effectiveness of these algorithms with numerical simulation and real-world applications. As have been shown that many phase-retrieval algorithms can find their counterparts in convex optimization theory [27], so can ptychography. In this paper, we focus on ptychographic reconstruction with noise and round-off errors, a common problem encountered in all measurements. Three maximum-likelihood (ML) based algorithms, namely proximal minimization (PM), alternating direction method of multiplier (ADMM) and accelerated proximal gradient (APG) are derived and benchmarked with both simulation and experiment data. Among them, APG, which has not been reported before, exhibits a superior performance. Although current model only considers noise and round-off errors, it is not difficult to extend it to a more complex case.

This paper is organized as follows. In the model section, we first build the mathematical model for the optimization problem in ptychography, and then discuss how different solving algorithms can be derived from proximal operators. Connections with existing techniques are discussed. Two statistical models for the noise, intensity Poisson and amplitude Gaussian, are considered throughout the paper. In the numerical simulation section, we benchmark performance of the derived algorithms at different signal-to-noise ratios (SNRs), and discuss the optimal conditions for convergence and accuracy in respective cases. In the experimental data section, PM, ADMM and APG are tested with an experimental dataset taken with x-rays. The result confirms that APG outperforms the other two and achieves the state-of-the-art performance.

2. Model

In a ptychographic scan, a probe, p, impinges on different parts of an object, o. The transmitted wave is assumed to be simply a product of the probe and object function, and propagates to a farfield detector. The wave-field at the object plane is linked to the wave-field at detector plane by Fourier transform. Because a physical detector only measures intensity (or amplitude in other words), the phase of the wave-field is lost in an experiment. A phase-retrieval algorithm intends to recover this set of missing information based on the measured amplitude under certain conditions. Assuming that we collect farfield diffraction patterns at K different positions, We now have two constraints for the measured data to satisfy. One is in real space (sample plane). The wave-field has to be written as a product of a probe and an object function with known translations. The other is in reciprocal space (detector plane). The modulus of its Fourier transform has to agree with the measurement

$$y_i = |F[p(\mathbf{r})o(\mathbf{r} - \mathbf{r}_i)]|, \quad i = 1, 2, \ldots K, \tag{1}$$

where y_i is the measured amplitude of the ith image and F is the two-dimensional Fourier transform operator. In equation (1), another set of known information is the probed position, $\mathbf{r}_i$. The goal here is to find complex-valued functions p and o that satisfy both the amplitude and translation constraints. We discretize the problem and rewrite equation (1) into a vector form by concatenating an image along its column directions

$$\mathbf{y}_i = |\mathbf{x}_i|, \quad \mathbf{x}_i = \mathbf{FPS}_i\mathbf{o}. \tag{2}$$

Here the lower-case bold letter represents a column vector and a capital bold one corresponds to a matrix. The absolute operator is element-wise. In the above equation, $\mathbf{y}_i \in \mathbb{R}^{N \times 1}$ is the measured amplitude, $\mathbf{F} \in \mathbb{C}^{N \times N}$ is the Fourier transform matrix, $\mathbf{p} \in \mathbb{C}^{N \times 1}$ is the probe vector, $\mathbf{P} = \mathrm{diag}(\mathbf{p})$ is a diagonal matrix, $\mathbf{S}_i \in \mathbb{R}^{N \times M}$ is a sparse matrix containing only zeros and ones that selects the illuminated elements of the object vector, $\mathbf{o} \in \mathbb{C}^{M \times 1}$, at the ith probed position. The resulting vector is $\mathbf{o}_i \in \mathbb{C}^{N \times 1}$. Problem described by equation (2) can be turned into an unconstrained optimization problem

$$\min_{\mathbf{x}_i} \quad f(\mathbf{x}_i) + g(\mathbf{x}_i), \tag{3}$$

where *f* and *g* are indicator functions corresponding to the two constraints respectively

$$f(\mathbf{x}_i) = \begin{cases} 0 & \text{if } \mathbf{x}_i \in \mathbf{dom}f \\ \infty & \text{otherwise} \end{cases}, \quad \mathbf{dom}f = \{\mathbf{x}_i \in \mathbb{C}^{N \times 1} | |\mathbf{x}_i| = \mathbf{y}_i\},$$

$$g(\mathbf{x}_i) = \begin{cases} 0 & \text{if } \mathbf{x}_i \in \mathbf{dom}g \\ \infty & \text{otherwise} \end{cases}, \quad \mathbf{dom}g = \{\mathbf{x}_i \in \mathbb{C}^{N \times 1} | \mathbf{x}_i = \mathbf{FPS}_i\mathbf{o}\}. \tag{4}$$

2.1. Alternating projection (AP)
The classical method of solving the above problem is AP algorithm that projects the solution into two domains,

$$z_i^{t+1} = \Pi_f(x_i^t),$$
$$x_i^{t+1} = \Pi_g(z_i^{t+1}). \tag{5}$$

Here Π_f and Π_g are Euclidean projection operators. By iteratively projecting into these two domains, we hope that an initial guess can converge to a solution fulfilling both constraints.

A proximal operator of a function f is defined by

$$\mathbf{prox}_{\lambda f}(\mathbf{v}) = \underset{\mathbf{x}}{\mathrm{argmin}}\left(f(\mathbf{x}) + \frac{1}{\lambda}\|\mathbf{x} - \mathbf{v}\|_2^2\right). \tag{6}$$

Here we use $1/\lambda$ instead of $1/2\lambda$ for convenience because we need to deal with complex-valued variables. For the two indicator functions in equation (3) their corresponding proximal operators are simplified to

$$\mathbf{prox}_f(\mathbf{v}_i) = \underset{\mathbf{x}_i \in \mathbf{dom}f}{\mathrm{argmin}} \sum_{i=1}^{K}\|\mathbf{x}_i - \mathbf{v}_i\|_2^2$$
$$= \Pi_f(\mathbf{v}_i),$$
$$\mathbf{prox}_g(\mathbf{v}_i) = \underset{\mathbf{x}_i \in \mathbf{dom}g}{\mathrm{argmin}} \sum_{i=1}^{K}\|\mathbf{x}_i - \mathbf{v}_i\|_2^2$$
$$= \Pi_g(\mathbf{v}_i). \tag{7}$$

Based on the definition, these proximal operators are just Euclidean projections. Therefore, the classical AP method can also be written as

$$z_i^{t+1} = \mathbf{prox}_f(x_i^t),$$
$$x_i^{t+1} = \mathbf{prox}_g(z_i^{t+1}). \tag{8}$$

The solution to the first projection is simply to replace the amplitude of $\mathbf{v}_i$ with $\mathbf{y}_i$ while retain its phase

$$\Pi_f(\mathbf{v}_i) = \mathrm{diag}(\mathbf{y}_i)\vartheta(\mathbf{v}_i), \quad \vartheta(\mathbf{v}_i) = \begin{cases} \mathbf{v}_i/|\mathbf{v}_i| & \mathbf{v}_i \neq 0 \\ 0 & \text{otherwise} \end{cases}. \tag{9}$$

The divide is an element-wise operation. The solution to the second projection can be obtained by many different ways. The probe and object functions can be updated sequentially [16], collectively [3] or jointly [14]. Here we choose the usual collective update

$$\mathbf{o} = \left(\sum_{i=1}^{K}\mathbf{S}_i^H\mathbf{P}^H\mathbf{P}\mathbf{S}_i\right)^{-1}\left(\sum_{i=1}^{K}\mathbf{S}_i^H\mathbf{P}^H\mathbf{F}^H\mathbf{v}_i\right),$$
$$\mathbf{p} = \left(\sum_{i=1}^{K}\mathbf{O}_i^H\mathbf{O}_i\right)^{-1}\left(\sum_{i=1}^{K}\mathbf{O}_i^H\mathbf{F}^H\mathbf{v}_i\right),$$
$$\mathbf{O}_i = \mathrm{diag}(\mathbf{S}_i\mathbf{o}),$$
$$\Pi_g(\mathbf{v}_i) = \mathbf{F}\mathbf{P}\mathbf{S}_i\mathbf{o}, \tag{10}$$

where superscript H denotes conjugate and transpose operation. Π_g defines another projection satisfying the second constraint. Specifically, we back-propagate the wave-field to the sample plane using inverse Fourier transform. Then we update probe and object functions collectively based on the probed positions. One can run the iteration once or multiple times for high accuracy. Lastly, we replace the wave-field at the sample plane by the product of the probe and object functions and propagate it to the detector plane. This method is often referred as error reduction (ER) algorithm.

From a statistical point of view, if we assume a Gaussian likelihood function of the amplitude and use its negative log as our cost function, we arrive at

$$\mathcal{L} = \sum_{i=1}^{K}\|\mathbf{y}_i - |\mathbf{F}\mathbf{P}\mathbf{S}_i\mathbf{o}|\|_2^2. \tag{11}$$

Their Wirtinger derivatives with respect to the probe and object are

$$\frac{\partial\mathcal{L}}{\partial\mathbf{o}^*} = \sum_{i=1}^{K}(\mathbf{F}\mathbf{P}\mathbf{S}_i)^H[\mathbf{x}_i - \mathrm{diag}(\mathbf{y}_i)\vartheta(\mathbf{x}_i)],$$
$$\frac{\partial\mathcal{L}}{\partial\mathbf{p}^*} = \sum_{i=1}^{K}(\mathbf{F}\mathbf{O}_i)^H[\mathbf{x}_i - \mathrm{diag}(\mathbf{y}_i)\vartheta(\mathbf{x}_i)]. \tag{12}$$

At a stationary point the derivatives have to be zero

$$\sum_{i=1}^{K} (\mathbf{S}_i^H \mathbf{P}^H \mathbf{P} \mathbf{S}_i) \mathbf{o} - \sum_{i=1}^{K} (\mathbf{S}_i^H \mathbf{P}^H \mathbf{F}^H) \mathrm{diag}(\mathbf{y}_i) \vartheta(\mathbf{x}_i) = 0,$$

$$\sum_{i=1}^{K} (\mathbf{O}_i^H \mathbf{O}_i) \mathbf{p} - \sum_{i=1}^{K} \mathbf{O}_i^H \mathbf{F}^H \mathrm{diag}(\mathbf{y}_i) \vartheta(\mathbf{x}_i) = 0. \tag{13}$$

They can be solved iteratively by a fixed-point algorithm that seeks a fix point of the equation, $\mathbf{z} = q(\mathbf{z})$. In this case, the unknown variables are $\mathbf{o}$ and $\mathbf{p}$. If they are solved in sequence we arrive at

$$\mathbf{o}^{t+1} = \left[\sum_{i=1}^{K} (\mathbf{P}^t \mathbf{S}_i)^H \mathbf{P}^t \mathbf{S}_i \right]^{-1}$$

$$\left[\sum_{i=1}^{K} (\mathbf{P}^t \mathbf{S}_i)^H \mathbf{F}^H \mathrm{diag}(\mathbf{y}_i) \vartheta(\mathbf{x}_i^t) \right],$$

$$\mathbf{p}^{t+1} = \left(\sum_{i=1}^{K} \mathbf{O}_i^{t+1 H} \mathbf{O}_i^{t+1} \right)^{-1} \tag{14}$$

$$\left[\sum_{i=1}^{K} (\mathbf{O}_i^{t+1 H}) \mathbf{F}^H \mathrm{diag}(\mathbf{y}_i) \vartheta(\mathbf{x}_i^t) \right],$$

$$\mathbf{x}_i^{t+1} = \mathbf{F} \mathbf{P}^{t+1} \mathbf{S}_i \mathbf{o}^{t+1}.$$

This is no different from the classical ER algorithm shown in equation (10). Therefore, we can also interpret ER algorithm for ptychography as some sort of fix-point algorithm that seeks the stationary point of the amplitude Gaussian likelihood function with respect to $\mathbf{p}$ and $\mathbf{o}$.

Algorithm 1. ER

Data: $\mathbf{y}_i \in \mathbb{R}^{N \times 1}$, $\mathbf{S}_i \in \mathbb{R}^{N \times M}$, $i = 1, 2, \ldots K$
Result: probe function $\mathbf{p}$ and object function $\mathbf{o}$
initialization: $\mathbf{p}^0, \mathbf{o}^0, \mathbf{x}_i^0, t_{max}$;
Repeat
$\quad$ $\mathbf{z}_i^{t+1} = \mathrm{diag}(\mathbf{y}_i) \vartheta(\mathbf{x}_i)$;
$\quad$ $\mathbf{o}^{t+1} = \mathrm{update\_o}(\mathbf{p}^t, \mathbf{z}_i^{t+1})$;
$\quad$ $\mathbf{p}^{t+1} = \mathrm{update\_p}(\mathbf{o}^{t+1}, \mathbf{z}_i^{t+1})$;
$\quad$ $\mathbf{x}_i^{t+1} = \mathbf{F} \mathrm{diag}(\mathbf{p}^{t+1}) \mathbf{S}_i \mathbf{o}^{t+1}$;
until $t > t_{max}$;

2.2. Alternating direction method of multiplier

For real-world phase-retrieval problems, ER is known to suffer from slow convergence and stagnation issues. A far more robust and popular algorithm is ADMM. One special variant of its form is Douglas–Rachford splitting method, also known as difference map (DM), which is widely used for ptychography reconstructions. ADMM is usually derived from argumented Lagrangian method. In the framework of proximal algorithms, it can be written in a very concise form. An in-depth discussion of ADMM can be found in the monography by Boyd *et al* [28], and its application for ptychography were reported in previous publications [12, 18]. Recently it was applied for joint ptycho-tomography reconstruction [29]. Thus, here we skip the derivation process. We change our optimization problem (equation (3)) slightly

$$\min_{\mathbf{x}_i, \mathbf{z}_i} \quad f(\mathbf{x}_i) + g(\mathbf{z}_i),$$

$$\text{subject to} \quad \mathbf{x}_i = \mathbf{z}_i. \tag{15}$$

For two indicator functions defined in equation (4), the ADMM algorithm is,

$$\mathbf{x}_i^{t+1} = \mathbf{prox}_f(\mathbf{z}_i^t - \mathbf{u}_i^t),$$

$$\mathbf{z}_i^{t+1} = \mathbf{prox}_g(\mathbf{x}_i^{t+1} + \mathbf{u}_i^t),$$

$$\mathbf{u}_i^{t+1} = \mathbf{u}_i^t + \mathbf{x}_i^{t+1} - \mathbf{z}_i^{t+1}. \tag{16}$$

If we define a new variable $\mathbf{w}_i^t = \mathbf{z}_i^t + \mathbf{u}_i^t$ and substitute it into equation (16), we arrive at

$$\mathbf{x}_i^{t+1} = \mathbf{prox}_f(2\mathbf{z}_i^t - \mathbf{w}_i^t),$$

$$\mathbf{z}_i^{t+1} = \mathbf{prox}_g(\mathbf{x}_i^{t+1} + \mathbf{w}_i^t - \mathbf{z}_i^t),$$

$$\mathbf{w}_i^{t+1} = \mathbf{w}_i^t + \mathbf{x}_i^{t+1} - \mathbf{z}_i^t. \tag{17}$$

Variable $\mathbf{x}_i^{t+1}$ is not independent and can be replaced. Rearrange terms and use Euclidean projection operators derived in equations (9) and (10), we have,

$$\mathbf{z}_i^t = \Pi_g(\mathbf{w}_i^t),$$
$$\mathbf{w}_i^{t+1} = \mathbf{w}_i^t + \Pi_f(2\mathbf{z}_i^t - \mathbf{w}_i^t) - \mathbf{z}_i^t. \tag{18}$$

This is the well-known DM algorithm [3].

For a noisy dataset, DM is known to have stability problem because it attempts to find a solution with its amplitude exactly equal to the measured value. A simple remedy is to replace the indicator function, f, with a negative log likelihood function, $\mathcal{L}$. In such a case, $\mathbf{x}_i$ update in equation (16) is modified to

$$\mathbf{x}_i^{t+1} = \mathbf{prox}_{\lambda\mathcal{L}}(\mathbf{z}_i^{t+1} - \mathbf{u}_i^t),$$

$$\mathbf{prox}_{\lambda\mathcal{L}}(\mathbf{v}_i) = \frac{\mathrm{diag}(\vartheta(\mathbf{v}_i))}{2(1+\lambda)} \begin{cases} 2(\lambda\mathbf{y}_i + |\mathbf{v}_i|), & \text{G} \\ |\mathbf{v}_i| + (\sqrt{|\mathbf{v}_i|^2 + 4\lambda(1+\lambda)\mathbf{y}_i^2}), & \text{P} \end{cases}$$

$$= \mathrm{E}_{\mathrm{MAP}}(\lambda, \mathbf{y}_i, \mathbf{v}_i). \tag{19}$$

Here 'G' and 'P' refers to amplitude Gaussian and intensity Poisson, respectively. One may notice that the update scheme for $\mathbf{x}_i$ is now parameter-dependent. From Bayes' theorem, the proximal operator can be interpreted as maximum-a-posterior (MAP) probability estimate, where the prior probability follows a normal distribution. The parameter, λ, controls how close the new update should be to its prior value, and plays an important role in determining the performance of the algorithm. We will have a more detailed dissuasion in the following section.

Algorithm 2. mADMM

Data: $\mathbf{y}_i \in \mathbb{R}^{N \times 1}, \mathbf{S}_i \in \mathbb{R}^{N \times M}, i = 1, 2, \ldots K$
Result: probe function $\mathbf{p}$ and object function $\mathbf{o}$
initialization: $\mathbf{p}^0, \mathbf{o}^0, \mathbf{z}_i^0, \mathbf{u}_i^0, \lambda, \beta, \delta, t_{max}$;
Repeat
    $\mathbf{x}_i^{t+1} = \mathrm{E}_{\mathrm{MAP}}(\lambda, \mathbf{y}_i, \mathbf{z}_i^t - \mathbf{u}_i^t)$;
    $\mathbf{o}^{t+1} = \mathrm{update\_o}(\mathbf{p}^t, \mathbf{x}_i^{t+1} + \mathbf{u}_i^t)$;
    $\mathbf{p}^{t+1} = \mathrm{update\_p}(\mathbf{o}^{t+1}, \mathbf{x}_i^{t+1} + \mathbf{u}_i^t)$;
    $\mathbf{z}_i^{t+1} = \mathrm{Fdiag}(\mathbf{p}^{t+1})\mathbf{S}_i\mathbf{o}^{t+1}$;
    $\mathbf{u}_i^{t+1} = \mathbf{u}_i^t + \mathbf{x}_i^{t+1} - \mathbf{z}_i^{t+1}$;
    If $\|\mathbf{u}^{t+1} - \mathbf{u}^t\|_2 / \|\mathbf{u}^{t+1}\|_2 < \delta$ **then**
        $\lambda = \beta\lambda$;
    end
until $t > t_{max}$

2.3. Proximal minimization

The fixed point of a proximal operator is also the minimizer of the original function. This leads to a simple proximal iterative algorithm

$$\mathbf{x}_i^{t+1} = \mathbf{prox}_{\lambda\mathcal{L}+g}(\mathbf{x}_i^t). \tag{20}$$

For the ptychography problem considered in this paper, we define

$$\mathbf{prox}_{\lambda\mathcal{L}+g}(\mathbf{v}_i) = \underset{\mathbf{x}_i \in \mathbf{dom}\,g}{\mathrm{argmin}} \quad \mathcal{L}(\mathbf{x}_i) + g(\mathbf{x}_i) + \sum_{i=1}^{K} \frac{1}{\lambda}\|\mathbf{x}_i - \mathbf{v}_i\|_2^2$$

$$= \underset{\mathbf{p},\mathbf{o}}{\mathrm{argmin}} \quad \mathcal{L}(\mathbf{p}, \mathbf{o}) + \sum_{i=1}^{K} \frac{1}{\lambda}\|\mathbf{FPS}_i\mathbf{o} - \mathbf{v}_i\|_2^2. \tag{21}$$

Again, we can interpret the update as a MAP estimate. The difference from ADMM discussed above is that $\mathbf{x}_i$ and $\mathbf{z}_i$ are forced to be equal here. In ADMM, the splitted variables belong to their individual domains and are not necessary the same. Similar to the derivation of equation (14), we make two-step update

$$\mathbf{z}_i^{t+1} = \mathrm{E}_{\mathrm{MAP}}(\lambda, \mathbf{y}_i, \mathbf{x}_i^t),$$
$$\mathbf{x}_i^{t+1} = \Pi_g(\mathbf{z}_i^{t+1}). \tag{22}$$

Compared to ER, the only difference is that the measured amplitude is replaced with a MAP-estimated value. This method is first proposed by Katkovnik *et al* [13]. Here we show it is equivalent to PM algorithm and provide an alternative perspective.

Algorithm 3. PM

Data: $\mathbf{y}_i \in \mathbb{R}^{N \times 1}$, $\mathbf{S}_i \in \mathbb{R}^{N \times M}$, $i = 1, 2, \ldots K$
Result: probe function $\mathbf{p}$ and object function $\mathbf{o}$
initialization: $\mathbf{p}^0, \mathbf{o}^0, \mathbf{x}_i^0, \lambda, t_{max}$;
Repeat
$\quad \left| \begin{array}{l} \mathbf{z}_i^{t+1} = \mathrm{E_{MAP}}(\lambda, \mathbf{y}_i, \mathbf{x}_i^t); \\ \mathbf{o}^{t+1} = \mathrm{update\_o}(\mathbf{p}^t, \mathbf{z}_i^{t+1}); \\ \mathbf{p}^{t+1} = \mathrm{update\_p}(\mathbf{o}^{t+1}, \mathbf{z}_i^{t+1}); \\ \mathbf{x}_i^{t+1} = F\mathrm{diag}(\mathbf{p}^{t+1})\mathbf{S}_i\mathbf{o}^{t+1}; \end{array} \right.$
until $t > t_{max}$

2.4. Accelerated proximal gradient

Let's consider the optimization problem

$$\min_{\mathbf{x}_i} \quad \mathcal{L}(\mathbf{x}_i) + g(\mathbf{x}_i). \tag{23}$$

Wirtinger derivative allows us to derive the gradient of the real-valued likelihood function with respect to the complex-valued variable $\mathbf{x}_i$,

$$\nabla_{\mathbf{x}_i^*}\mathcal{L} = \begin{cases} \mathbf{x}_i - \mathrm{diag}(\mathbf{y}_i)\vartheta(\mathbf{x}_i), & \mathrm{G} \\ \mathbf{x}_i - \mathrm{diag}(\dfrac{\mathbf{x}_i}{|\mathbf{x}_i|^2 + \varepsilon})(\mathbf{y}_i^2 + \varepsilon), & \mathrm{P} \end{cases} \tag{24}$$

Here ε is a small real-valued constant introduced to avoid the discontinuity at zero, as suggested in [18]. The negative of the gradient is also the steepest descent direction of the function. The proximal gradient algorithm is

$$\mathbf{x}_i^{t+1} = \mathbf{prox}_g(\mathbf{x}_i^t - \lambda^t \nabla_{\mathbf{x}_i^*}\mathcal{L}(\mathbf{x}_i^t)). \tag{25}$$

Here λ^t is a positive step size that can vary at each iteration. We use a simple method to determine its value [30]. The step size remains the same unless the following condition is violated

$$\begin{aligned} \mathbf{z}_i &= \mathbf{prox}_g(\mathbf{x}_i^t - \lambda^t \nabla_{\mathbf{x}_i^*}\mathcal{L}(\mathbf{x}_i^t)), \\ \mathcal{L}(\mathbf{z}_i) &\leqslant \mathcal{Q}_{\lambda^t}(\mathbf{z}_i, \mathbf{x}_i^t), \\ \mathcal{Q}_{\lambda^t}(\mathbf{z}_i, \mathbf{x}_i^t) &= \mathcal{L}(\mathbf{x}_i^t) + \sum_{i=1}^{K} 2\mathrm{Re}(\nabla_{\mathbf{x}_i^*}\mathcal{L}(\mathbf{x}_i^t)^H(\mathbf{z}_i - \mathbf{x}_i^t)) + \frac{1}{\lambda^t}\|\mathbf{z}_i - \mathbf{x}_i^t\|_2^2. \end{aligned} \tag{26}$$

In such a case, we reject the update and multiply the step size by a factor $\beta \in (0, 1)$. The process is repeated until the above condition is satisfied and then the iteration is completed, $\mathbf{x}_i^{t+1} = \mathbf{z}_i$. The proximal gradient algorithm can be understood from a point of view of localized optimization. For completeness, we give an explanation due to Beck and Teboulle [30]. The function, $\mathcal{Q}_{\lambda^t}(\mathbf{z}_i, \mathbf{x}_i^t)$, is an upper bound to $\mathcal{L}(\mathbf{x}_i^t)$ that is tight at $\mathbf{x}_i^t$, i.e. $\mathcal{Q}_{\lambda^t}(\mathbf{x}_i^t, \mathbf{x}_i^t) = \mathcal{L}(\mathbf{x}_i^t)$ and $\mathcal{Q}_{\lambda^t}(\mathbf{z}_i, \mathbf{x}_i^t) \geqslant \mathcal{L}(\mathbf{z}_i)$, provided that $\lambda^t \in (0, L]$, where L is a Lipschitz constant of $\nabla_{\mathbf{x}_i^*}\mathcal{L}$. This function can be considered as an first-order approximation to $\mathcal{L}$ with a regularization term. We may rewrite it as

$$\mathcal{Q}_{\lambda^t}(\mathbf{z}_i, \mathbf{x}_i^t) = \mathcal{L}(\mathbf{x}_i^t) + \sum_{i=1}^{K} \frac{1}{\lambda^t}\|\mathbf{z}_i - (\mathbf{x}_i^t - \lambda^t \nabla_{\mathbf{x}_i^*}\mathcal{L})\|_2^2 - \lambda^t\|\nabla_{\mathbf{x}_i^*}\mathcal{L}\|_2^2. \tag{27}$$

In the vicinity of $\mathbf{x}_i^t$, we replace the original optimization problem with an approximate one

$$\min_{\mathbf{z}_i} \quad \mathcal{Q}_{\lambda^t}(\mathbf{z}_i, \mathbf{x}_i^t) + g(\mathbf{z}_i). \tag{28}$$

Dropping constant terms in equation (27) does not affect the solution to equation (28). As a result we arrive at

$$\begin{aligned} \underset{\mathbf{z}_i}{\mathrm{argmin}} \quad &\sum_{i=1}^{K} \frac{1}{\lambda^t}\|\mathbf{z}_i - (\mathbf{x}_i^t - \lambda^t \nabla_{\mathbf{x}_i^*}\mathcal{L})\|_2^2 + g(\mathbf{z}_i) \\ = &\mathbf{prox}_g(\mathbf{x}_i^t - \lambda^t \nabla_{\mathbf{x}_i^*}\mathcal{L}). \end{aligned} \tag{29}$$

Consequently, we can interpret each iteration as a proximal operator of g along the steepest decent direction of $\mathcal{L}$, as the name proximal gradient suggests. By definition

$$\mathcal{L}(\mathbf{x}_i^{t+1}) + g(\mathbf{x}_i^{t+1}) \leqslant \mathcal{Q}_{\lambda^t}(\mathbf{x}_i^{t+1}, \mathbf{x}_i^t) + g(\mathbf{x}_i^{t+1}) \leqslant \mathcal{Q}_{\lambda^t}(\mathbf{x}_i^t, \mathbf{x}_i^t) + g(\mathbf{x}_i^t). \tag{30}$$

In this particular case g is an indicator function, thus

$$g(\mathbf{x}_i^t) = g(\mathbf{x}_i^{t+1}) = 0. \tag{31}$$

We can simplify the inequality as

$$\mathcal{L}(\mathbf{x}_i^{t+1}) \leqslant \mathcal{Q}_{\lambda^t}(\mathbf{x}_i^{t+1}, \mathbf{x}_i^t) \leqslant \mathcal{Q}_{\lambda^t}(\mathbf{x}_i^t, \mathbf{x}_i^t) = \mathcal{L}(\mathbf{x}_i^t). \tag{32}$$

Therefore, each iteration decent the negative log likelihood function meanwhile satisfying the constraint g. For a faster convergence, the accelerated version of the proximal gradient method which includes an additional extrapolation step can be used

$$\mathbf{w}_i^t = \mathbf{x}_i^t + \omega^t(\mathbf{x}_i^t - \mathbf{x}_i^{t-1}),$$
$$\mathbf{x}_i^{t+1} = \mathbf{prox}_g(\mathbf{w}_i^t - \lambda^t \nabla_{\mathbf{w}_i^*}\mathcal{L}(\mathbf{w}_i^t)),$$
$$\omega^t = \frac{t}{t+3}. \tag{33}$$

We note that Xu *et al* [15] recently proposed accelerated Wirtinger flow (AWF) method for ptychography, which stems from the popular Wirtinger flow algorithm for phase-retrieval problems [31]. It shares some similarity with APG. However, they differ fundamentally in many aspects. AWF can be considered as a steepest decent method with a constant step size, while APG here is a projected gradient method with a varying step size. We may consider APG as a hybrid algorithm combining gradient descent and projection methods. We first decent the cost function in reciprocal space with a small step, and then project it to the domain in real space which satisfies the translation constraint. The update is completed only when such a move would make the cost function smaller. For a Gaussian amplitude likelihood function, if λ^t is equal to one, as can be seen the updating scheme is no different from ER algorithm. Therefore, we can choose one as the initial value of the step size.

Algorithm 4. APG

Data: $\mathbf{y}_i \in \mathbb{R}^{N \times 1}, \mathbf{S}_i \in \mathbb{R}^{N \times M}, i = 1, 2, \ldots K$
Result: probe function $\mathbf{p}$ and object function $\mathbf{o}$
initialization: $\mathbf{p}^0, \mathbf{o}^0, \mathbf{x}_i^0, \lambda^0 = 1, \beta, t_{max}$;
Repeat
$\quad$ $\omega^t = \frac{t}{t+3}$;
$\quad$ $\mathbf{w}_i^t = \mathbf{x}_i^t + \omega^t(\mathbf{x}_i^t - \mathbf{x}_i^{t-1})$;
$\quad$ **Repeat**
$\quad\quad$ $\mathbf{z}_i = \mathbf{w}_i^t - \lambda^t \nabla_{\mathbf{x}_i^*}\mathcal{L}(\mathbf{w}_i^t)$;
$\quad\quad$ $\mathbf{o}^{t+1} = \text{update\_o}(\mathbf{p}^t, \mathbf{z}_i)$;
$\quad\quad$ $\mathbf{p}^{t+1} = \text{update\_p}(\mathbf{o}^{t+1}, \mathbf{z}_i)$;
$\quad\quad$ $\mathbf{z}_i = \text{Fdiag}(\mathbf{p}^{t+1})\mathbf{S}_i\mathbf{o}^{t+1}$;
$\quad\quad$ **if** $\mathcal{L}(\mathbf{z}_i) \leqslant \mathcal{Q}_{\lambda^t}(\mathbf{z}_i, \mathbf{w}_i^t)$ **then**
$\quad\quad\quad$ **Return** $\mathbf{x}_i^{t+1} = \mathbf{z}_i$;
$\quad\quad$ **else**
$\quad\quad\quad$ $\lambda^t = \lambda^t \beta$;
$\quad\quad$ **end**
$\quad$ **until** λ^t *is sufficiently small*;
$\quad$ $\lambda^{t+1} = \lambda^t$;
$\quad$ **if** $\lambda^{t+1} < \delta$ **then**
$\quad\quad$ $\lambda^{t+1} = \lambda^0$;
$\quad$ **end**
until $t > t_{max}$

---

3. Numerical simulation

In this section we will compare the performance of different methods using simulation data. The test object function is shown in figure 1. The image 'Cameraman' is used as its amplitude and 'Barbara' as its phase. The pixel size of the image is assumed to be 5 nm. A probe of size 37 nm is produced by a Fresnel Zone plate and a special fermat scan that follows the equation (in polar coordinates) [32]

$$r_i = c\sqrt{i}, \quad \theta_i = 2.4i, \quad i = 1, 2 \ldots K \tag{34}$$
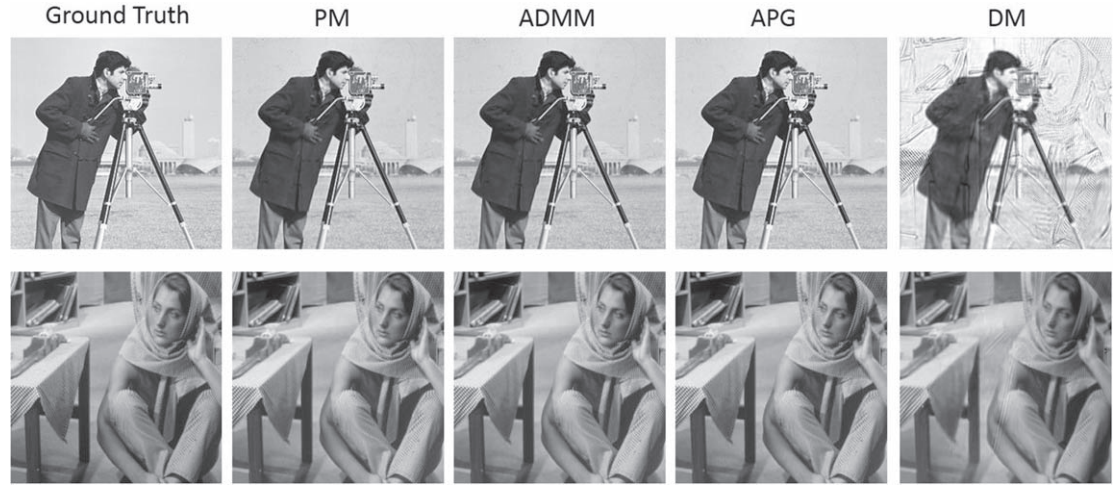
Figure 1. Comparison of reconstruction results obtained from different algorithms with a noisy dataset. The maximum number of photons received at one pixel of the detector is scaled to 10^4 counts, and a Poisson noise is added accordingly. This corresponds to SNR = 32.25 dB for the collected diffraction patterns. Top panel: amplitude image. Bottom panel: phase image.

is performed to illuminate the different parts of the sample. c is chosen to be 20 nm (4 pixels) and a total of 1261 far-field diffraction patterns are collected. The maximum detector intensity is scaled to the range of 10^2–10^6 and a Poisson noise is added to each pixel accordingly. The 'measured' intensity contains two different types of errors. One is the round-off error since the measured intensity are integers. This approximation reduces the dynamical range of the signal. Particularly in the detector region where the intensity drops below 0.5 count, they are all set to zero. Two is the Poisson noise, which adds background fluctuation to the signal. To be more quantitative, we calculate the SNR of the intensity as

$$\text{SNR} = 10 \log_{10}\left(\frac{\sum_{i=1}^{K} \|\hat{\mathbf{y}}_i^2\|_2^2}{\sum_{i=1}^{K} \|\mathbf{y}_i^2 - \hat{\mathbf{y}}_i^2\|_2^2} \right), \tag{35}$$

where $\hat{\mathbf{y}}_i$ are the ground-truth values.

We first study the case with SNR = 32.25 dB (max. detector pixel intensity = 10^4 counts). The reconstruction results obtained from different algorithms discussed in the preceding section are shown in figure 1. All of ML-based methods yield a high-quality reconstruction with nearly indistinguishable difference. If we pay close attention on the reconstructed amplitude, a very faint cross-talk from the phase image can be seen in the background. In contrast, the phase reconstruction does not show any visible artifacts. As a comparison, we also plot results from the non-statistical algorithm, DM. Strong cross-talk in the reconstructed amplitude image can be observed. In addition, the reconstructed phase image is less sharp than the others and contains some visible artifacts. This suggests that even at this level of signals, one may still need to use ML-based algorithm to achieve the best result.

To quantify the error for a systematic study, we use a root mean square error (RMSE) defined as

$$\text{RMSE} = \frac{\|\hat{\mathbf{o}} - a\mathbf{o}\|_2}{\|\hat{\mathbf{o}}\|_2}, \tag{36}$$

where $\hat{\mathbf{o}}$ is the ground-truth and a is a complex-valued constant to account for the ambiguity in ptychography reconstruction. Here we choose RMSE over other image quantification methods (such as the popular structural similarity index [33]) because the reconstructed quantity is a complex-valued number. RMSE allows the calculation of a single metric without evaluating the amplitude and phase images separately. The assessment under different conditions is presented in figure 2. For PM (figure 2(a)), its convergence rate is λ-dependent, while the resultant RMSE is not; they all converge to the same value. The Poisson model leads to a smaller RMSE, which is not a surprise since a Poisson noise is added. When the amplitude Gaussian is used, PM is no better than ER. As we discussed earlier, ER can also be considered as a ML-based method.

For ADMM (figure 2(b)), we observe that a larger value of λ leads to a faster convergence. In the limit of infinity, ML-based ADMM becomes DM, which shows the fastest convergence rate at the beginning. However, a large value of λ can cause a stability issue, which can be seen in the plot. In this case when $\lambda > 5$, they do not tend to converge to a stable solution after initial fast convergence, but rather fluctuate as iteration goes. Also, a large λ results in a higher RMSE. Therefore the solution is less accurate. This suggests a multi-stage strategy for ADMM to optimize both the convergence rate and accuracy. We can start with a large λ for fast convergence. When a stable
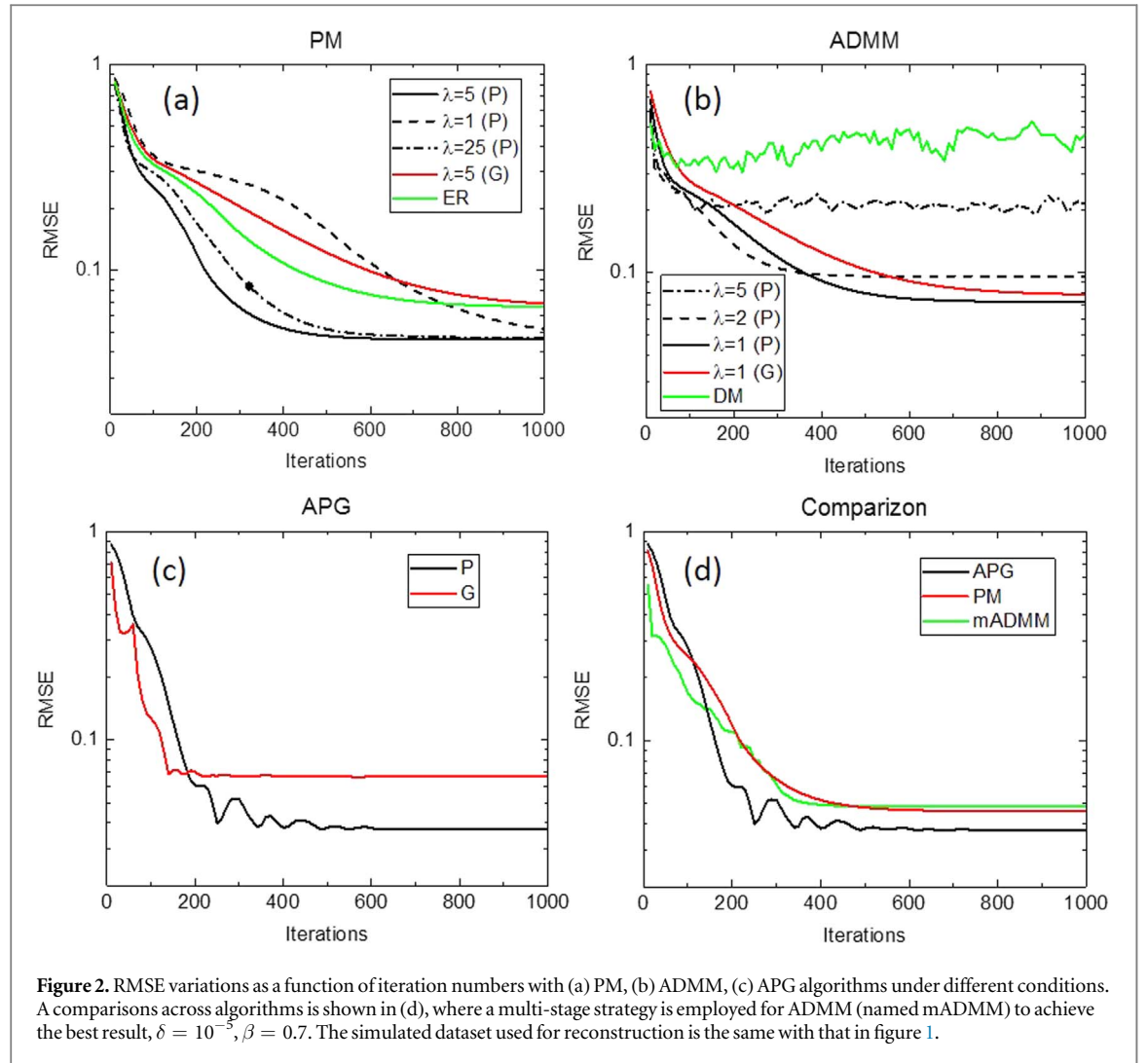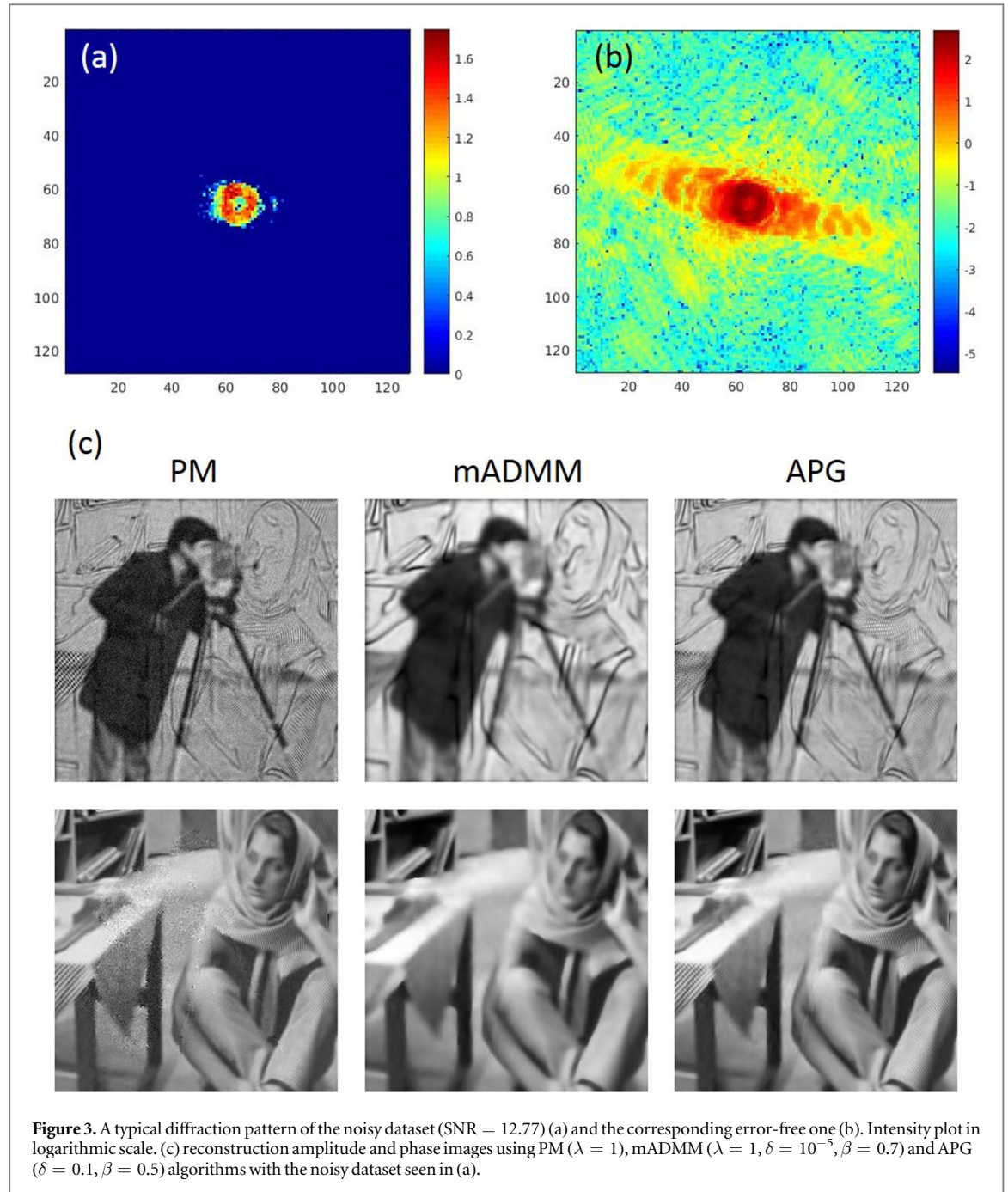
Figure 2. RMSE variations as a function of iteration numbers with (a) PM, (b) ADMM, (c) APG algorithms under different conditions. A comparisons across algorithms is shown in (d), where a multi-stage strategy is employed for ADMM (named mADMM) to achieve the best result, $\delta = 10^{-5}$, $\beta = 0.7$. The simulated dataset used for reconstruction is the same with that in figure 1.

solution is reached ($\|\mathbf{u}^t - \mathbf{u}^{t-1}\|/\|\mathbf{u}^t\| <$ threshold), we reduce the value of λ by multiplying it with a constant $\beta \in (0, 1)$, and then use the solution obtained from the last stage as the initial guess and continue the iteration. To distinguish it from a regular ADMM algorithm, we call it mADMM thereafter, where 'm' refer to multi-stage.

For APG (figure 2(c)), we always choose the initial value of λ^0 as one and it is adjusted automatically in the update. Similar to the cases in PM and ADMM, intensity Poisson model outperforms amplitude Gaussian. It is worth noting that when the initial guess of the probe is bad, λ^t can quickly becomes very small, particularly for intensity Poisson model. This will cause a stagnation problem. A remedy is to reset λ^t to its initial value when it becomes too small, but a price to pay is more computation time.

To assess performance across algorithms, in figure 2(d) we plot their RMSE variations as a function of iteration number. For a fair comparison, all start with a disk-like probe and a square object as the initial guess. Poisson model is used in all algorithms. Among them, APG has the best performance. Not only its overall convergences rate is higher, but also the resultant RMSE is smaller.

With current SNR, though quantitative analysis shows the difference in the reconstruction, visually the results look almost the same (figure 1). In a more extreme case, we consider a maximum detector intensity of 100 counts (SNR $= 12.77$). Compared to the previous case, the diffraction intensity is reduced by 100-fold. Figure 3(a) shows a typical diffraction pattern with limited counts in logarithmic scale, and a comparison with the ground-truth (figure 3(b)). The error-free data has a dynamical range over seven orders of magnitude, while that for the limited counts is reduced to less than two due to the round-off to integers. The added noise makes the situation even worse. Such a noisy dataset poses a significant challenge to ptychographic reconstruction. We present in figure 3(c) the reconstructed results obtained from PM, APG and mADMM algorithms. Because DM does not converge at all in this case, its result is not shown here. Again a disk-like probe and a square object as the initial guess are used. All three algorithms lead to a converged solution. As is clear, their results however differ considerably from algorithm to algorithm. PM recovers high-frequency features well, but the reconstructed images look more grainy. There are also strong cross-contamination between the amplitude and the phase

Figure 3. A typical diffraction pattern of the noisy dataset (SNR = 12.77) (a) and the corresponding error-free one (b). Intensity plot in logarithmic scale. (c) reconstruction amplitude and phase images using PM ($\lambda = 1$), mADMM ($\lambda = 1, \delta = 10^{-5}, \beta = 0.7$) and APG ($\delta = 0.1, \beta = 0.5$) algorithms with the noisy dataset seen in (a).

images. In contrast, mADMM yields smooth images, but the high-frequency details are lost. APG balances the two well and produces the best overall results. The quality of the obtained phase image is still very acceptable, without having visible artifacts and losing too much details.

Figure 4 depicts the achieved RMSE of the reconstructed complex image at various levels of signal. In general, the log–log plot shows a close-to-linear relationship, suggesting that they all follow a power law approximately. In most circumstances, APG outperforms the other two. Unlike PM and mADMM algorithms for which the value of λ has to be chosen accordingly with the noise level to achieve the best performance, there is no need to tune any parameter for APG when SNR changes.

4. Experimental data

The simulation data only take into account round-off errors and Poisson noise. The real-world problem can be much more complicated. In order to assess the robustness of these algorithms, we perform reconstruction on an experimental dataset that was taken at the hard x-ray nanoprobe beamline of National Synchrotron Light Source II, Brookhaven National Laboratory. The nanobeam with a size of ~13 nm^2 at 12 keV was produced by two
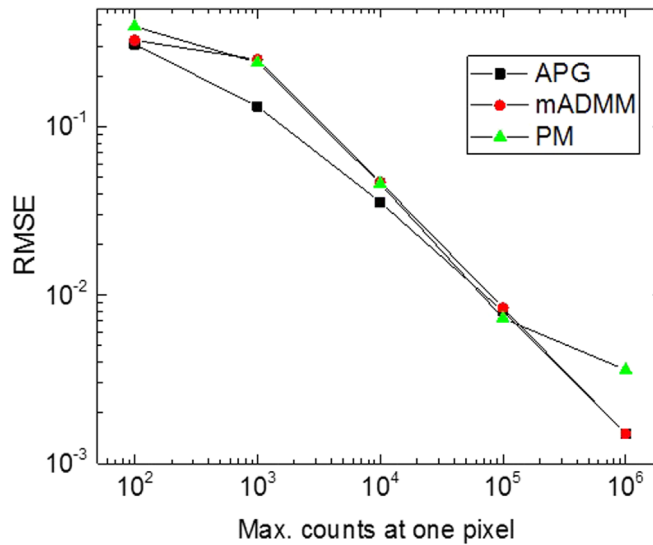
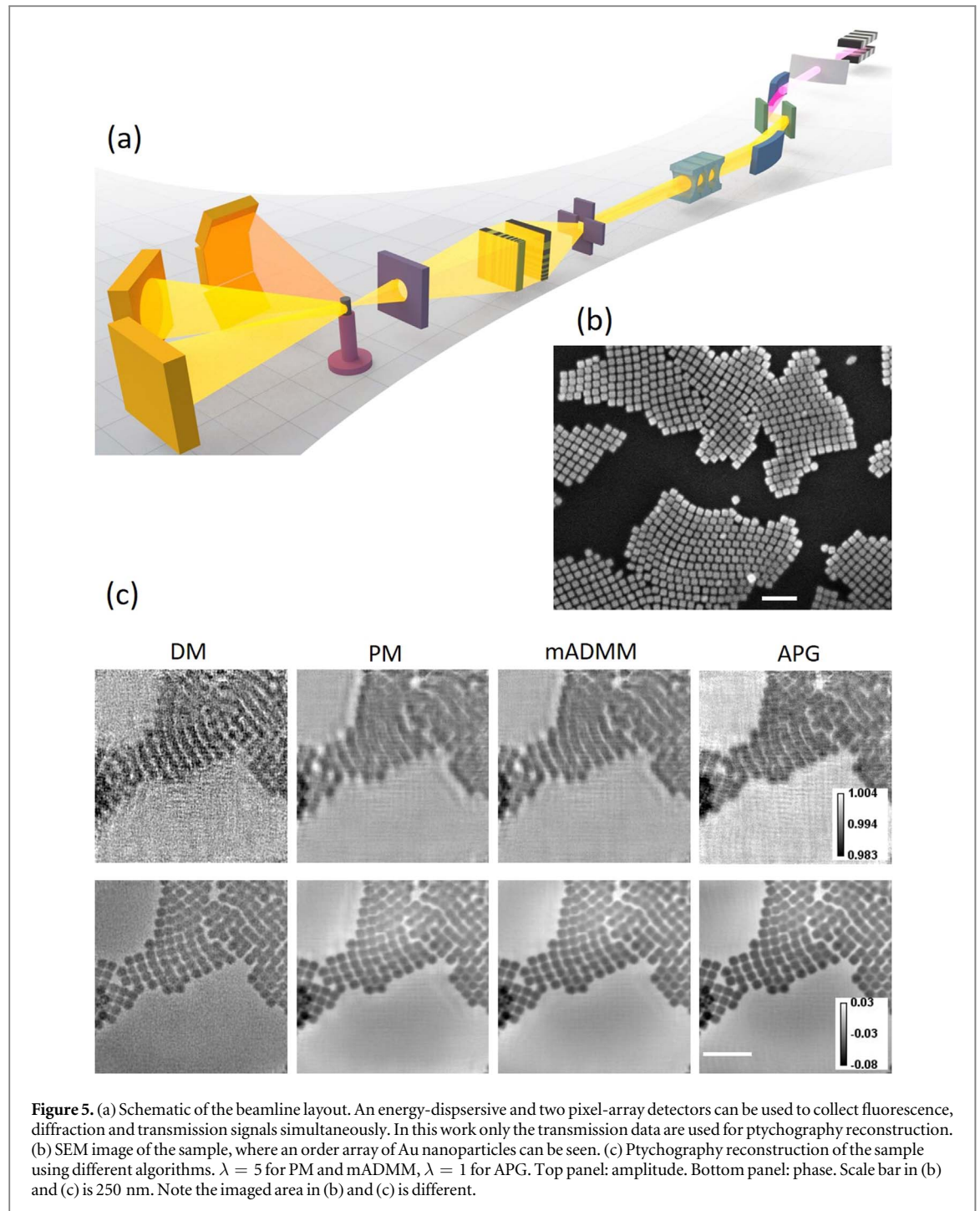**Figure 4.** RMSE variations at different levels of signal strength using PM, mADMM and APG algorithms.

crossed multilayer Laue lenses. The beamline layout is shown in figure 5(a). Details about the experimental setup can be found elsewhere [34]. The sample consists of cubic Au nanoparticles with a size of 50 nm deposited on a Si substrate [35]. They form an ordered array with gaps between them as small as 10 nm, as seen in the scanning electron microscopy image (figure 5(b)). The sample was placed at a downstream position with a distance of 25 μm to the focal plane. A fermat scan with $c = 20$ nm (equation (34)) was performed to avoid periodic aliasing effect, and a total of 792 frames were collected. The diffraction pattern collected on a far-field pixel-array detector (Merlin, Quantum Detectors) has over 4×10^4 maximum detector counts at one pixel.

In figure 5(c) reconstructed complex-valued images with different algorithms are presented. For a fair comparison, they all start with the same initial guess of the probe function, which is obtained by inverse Fourier transforming the measured far-field amplitude and then propagating 25 μm to the sample plane. In other words, the initial guess assumes a lens with no phase aberration. Because DM does not converge to a stable solution, we have to choose an intermediate reconstruction result that looks the best. Nevertheless, the phase image is a bit noisy, and the amplitude part is barely recognizable. PM yields a smoother result, but both the phase and amplitude exhibit some ghost image around the boundary of the array. mADMM produces a further improved result, but the ghost image can still be seen, particularly in the amplitude image. Not surprisingly, the best result is achieved with APG. It has no apparent artifacts seen in the reconstruction. We can clearly resolve the shape of individual 50 nm nanoparticles and their sharp edges in the phase image, even though the amount of the phase variation of one layer is in the order of ∼0.05 radian. Because the sample has a very low absorption contrast (∼1.7%), the amplitude image usually is too fuzzy to be useful. However in the one obtained from APG we can still recognize the particle array. For this dataset, we conclude that APG leads to a reconstruction result with overall image quality noticeably better than that of others.

There are a few remarks we'd like to point out. Among all the algorithms tested in this paper, DM takes the least number of iterations to arrive at a plausible solution, particularly when the initial guess of the probe is far from the ground truth. However, with the presence of noise DM can become unstable and diverge. On the contrary, ML-based algorithms converge slower, but are stable. The reason is that their iteration processes usually involves a MAP-based sub-optimization step which requires the updated solution to be close to its prior value. As a result, a good guess of the initial probe is more important for these ML-based algorithms. In addition, for the experimental data tested in this case we do not see visible difference between Poisson and Gaussian models. This suggests that the difference in reconstruction between two statistical models diminishes as intensity increases.

5. Conclusion

In summary, we presented several solving techniques for ptychographic imaging derived in the framework of proximal algorithms. The separable nature of the proximal operator makes it well-suited for dealing with large-scale ptychography reconstruction problems where its evaluation can be parallelized. The optimization problem

Figure 5. (a) Schematic of the beamline layout. An energy-dispsersive and two pixel-array detectors can be used to collect fluorescence, diffraction and transmission signals simultaneously. In this work only the transmission data are used for ptychography reconstruction. (b) SEM image of the sample, where an order array of Au nanoparticles can be seen. (c) Ptychography reconstruction of the sample using different algorithms. $\lambda = 5$ for PM and mADMM, $\lambda = 1$ for APG. Top panel: amplitude. Bottom panel: phase. Scale bar in (b) and (c) is 250 nm. Note the imaged area in (b) and (c) is different.

is usually divided into sub-optimization steps involving proximal operators for which often a closed-form solution can be found. Therefore, the problem becomes more tractable. We derived ER, PM, ADMM, DM and APG algorithms and benchmarked their performance with noisy datasets. Among them, APG depicted the best reconstruction result not only in numerical simulation but also in experiment.

In the current work, we only consider a noise model and round-off errors. In the same frame work, it is not difficult to enable modes to deal with partial coherence [19] and bluring effect in fly-scan [36, 37]. In such cases, contribution from different modes will add up incoherently and the likelihood function has to be modified accordingly. We can also consider to incorporate more constraints on probe or object function. For example, adding a regularization term with denoiser has shown suppressed noise in the reconstructed object function [13, 38]. With an additional constraint on the probe, it was demonstrated that the periodic aliasing effect seen in grid scans could be mitigated [18]. There is still a lot of room for improvement, and they will be the future work.

Acknowledgments

ORCID iDs

Hanfei Yan ⓘ https://orcid.org/0000-0001-6824-0367

References

[1] Rodenburg J M and Faulkner H M L 2004 A phase retrieval algorithm for shifting illumination *Appl. Phys. Lett.* **85** 4795–7
[2] Rodenburg J M, Hurst A C, Cullis A G, Dobson B R, Pfeiffer F, Bunk O, David C, Jefimovs K and Johnson I 2007 Hard-x-ray lensless imaging of extended objects *Phys. Rev. Lett.* **98** 4
[3] Thibault P, Dierolf M, Menzel A, Bunk O, David C and Pfeiffer F 2008 High-resolution scanning x-ray diffraction microscopy *Science* **321** 379–82
[4] Hruszkewycz S O, Highland M J, Holt M V, Kim D, Folkman C M, Thompson C, Tripathi A, Stephenson G B, Hong S and Fuoss P H 2013 Imaging local polarization in ferroelectric thin films by coherent x-ray Bragg projection ptychography *Phys. Rev. Lett.* **110** 177601
[5] Holler M, Guizar-Sicairos M, Tsai E H R, Dinapoli R, Muller E, Bunk O, Raabe J and Aeppli G 2017 High-resolution non-destructive three-dimensional imaging of integrated circuits *Nature* **543** 402
[6] Deng J J, Vine D J, Chen S, Jin Q L, Nashed Y S G, Peterka T, Vogt S and Jacobsen C 2017 X-ray ptychographic and fluorescence microscopy of frozen-hydrated cells using continuous scanning *Sci. Rep.* **7** 445
[7] Shapiro D A *et al* 2014 Chemical composition mapping with nanometre resolution by soft x-ray microscopy *Nat. Photon.* **8** 765–9
[8] Zheng G A, Horstmeyer R and Yang C H 2013 Wide-field, high-resolution fourier ptychographic microscopy *Nat. Photon.* **7** 739–45
[9] Jiang Y *et al* 2018 Electron ptychography of 2d materials to deep sub-Ångström resolution *Nature* **559** 343–9
[10] Maiden A M and Rodenburg J M 2009 An improved ptychographical phase retrieval algorithm for diffractive imaging *Ultramicroscopy* **109** 1256–62
[11] Thibault P and Guizar-Sicairos M 2012 Maximum-likelihood refinement for coherent diffractive imaging *New J. Phys.* **14** 20
[12] Wen Z W, Yang C, Liu X and Marchesini S 2012 Alternating direction methods for classical and ptychographic phase retrieval *Inverse Prob.* **28** 18
[13] Katkovnik V and Astola J 2013 Sparse ptychographical coherent diffractive imaging from noisy measurements *J. Opt. Soc. Am.* A **30** 367–79
[14] Odstrcil M, Menzel A and Guizar-Sicairos M 2018 Iterative least-squares solver for generalized maximum-likelihood ptychography *Opt. Express* **26** 3108–23
[15] Xu R, Soltanolkotabi M, Haldar J P, Unglaub W, Zusman J, Levi A F J and Leahy R M 2018 Accelerated Wirtinger flow: a fast algorithm for ptychography arXiv:1806.05546
[16] Maiden A, Johnson D and Li P 2017 Further improvements to the ptychographical iterative engine *Optica* **4** 736–45
[17] Hesse R, Luke D R, Sabach S and Tam M K 2015 Proximal heterogeneous block implicit-explicit method and application to blind ptychographic diffraction imaging *SIAM J. Imaging Sci.* **8** 426–57
[18] Chang H B, Enfedaque P and Marchesin S 2019 Blind ptychographic phase retrieval via convergent alternating direction method of multipliers *SIAM J. Imaging Sci.* **12** 153–85
[19] Thibault P and Menzel A 2013 Reconstructing state mixtures from diffraction measurements *Nature* **494** 68–71
[20] Maiden A M, Humphry M J, Sarahan M C, Kraus B and Rodenburg J M 2012 An annealing algorithm to correct positioning errors in ptychography *Ultramicroscopy* **120** 64–72
[21] Zhang F C, Peterson I, Vila-Comamala J, Berenguer A D F, Bean R, Chen B, Menzel A, Robinson I K and Rodenburg J M 2013 Translation position determination in ptychographic coherent diffraction imaging *Opt. Express* **21** 13592–606
[22] Maiden A M, Humphry M J and Rodenburg J M 2012 Ptychographic transmission microscopy in three dimensions using a multi-slice approach *J. Opt. Soc. Am.* A **29** 1606–14
[23] Tsai E H R, Usov I, Diaz A, Menzel A and Guizar-Sicairos M 2016 X-ray ptychography with extended depth of field *Opt. Express* **24** 29090–109
[24] Odstrcil M, Baksh P, Boden S A, Card R, Chad J E, Frey J G and Brocklesby W S 2016 Ptychographic coherent diffractive imaging with orthogonal probe relaxation *Opt. Express* **24** 8360–9
[25] Parikh N and Boyd S 2014 Proximal algorithms *Found. Trends Optim.* **1** 127–239
[26] Hunger R 2007 An introduction to complex differentials and complex differentiability *Technical Report* Technische Universitat Munchen
[27] Bauschke H H, Combettes P L and Luke D R 2002 Phase retrieval, error reduction algorithm, and fienup variants: a view from convex optimization *J. Opt. Soc. Am.* A **19** 1334–45
[28] Boyd S, Parikh N, Chu E, Peleato B and Eckstein J 2011 Distributed optimization and statistical learning via the alternating direction method of multipliers *Found. Trends® Mach. Learn.* **3** 1–122
[29] Aslan S, Nikitin V, Ching D J, Bicer T, Leyffer S and Gürsoy D 2019 Joint ptycho-tomography reconstruction through alternating direction method of multipliers *Opt. Express* **27** 9128–43
[30] Beck A and Teboulle M Gradient-based algorithms with applications to signal recovery problems *Convex Optimization in Signal Processing and Communications* ed D Palomar and Y Eldar (Cambridge: Cambridge University Press) pp 42–88
[31] Candes E J, Li X D and Soltanolkotabi M 2015 Phase retrieval via Wirtinger flow: theory and algorithms *IEEE Trans. Inf. Theory* **61** 1985–2007
[32] Huang X, Yan H, Harder R, Hwu Y, Robinson I K and Chu Y S 2014 Optimization of overlap uniformness for ptychography *Opt. Express* **22** 12634–44

[33] Zhou W, Bovik A C, Sheikh H R and Simoncelli E P 2004 Image quality assessment: from error visibility to structural similarity *IEEE Trans. Image Process.* **13** 600–12

[34] Yan H, Bouet N, Zhou J, Huang X, Nazaretski E, Xu W, Cocco A, Chiu W K S, Brinkman K and Chu Y S 2018 Multimodal hard x-ray imaging with resolution approaching 10 nm for studies in material science *Nano Futures* **2** 011001

[35] Lu F *et al* 2017 Surface proton transfer promotes four-electron oxygen reduction on gold nanocrystal surfaces in alkaline solution *J. Am. Chem. Soc.* **139** 7310–7

[36] Pelz P M, Guizar-Sicairos M, Thibault P, Johnson I, Holler M and Menzel A 2014 On-the-fly scans for x-ray ptychography *Appl. Phys. Lett.* **105** 251101

[37] Huang X J, Lauer K, Clark J N, Xu W H, Nazaretski E, Harder R, Robinson I K and Chu Y S 2015 Fly-scan ptychography *Sci. Rep.* **5** 9074

[38] Shi B S, Lian Q S, Chen S Z and Fan X Y 2018 Sbm3d: sparse regularization model induced by bm3d for weighted diffraction imaging *IEEE Access* **6** 46266–80