Further improvements to the ptychographical iterative engine

ANDREW MAIDEN,* DANIEL JOHNSON, AND PENG LI

Department of Electronic and Electrical Engineering, University of Sheffield, Sheffield S1 3JD, UK
Corresponding author: a.maiden@sheffield.ac.uk

Ptychography is a form of phase imaging that uses iterative algorithms to reconstruct an image of a specimen from a series of diffraction patterns. It is swiftly developing into a mainstream technique, with a growing list of applications across a range of imaging modalities. As the field has advanced, numerous reconstruction algorithms have been proposed, yet the early approaches have not seen major improvement and remain popular. In this paper, we revisit the first such algorithm, the ptychographical iterative engine (PIE), and show how a simple revision and powerful extension can deliver an order of magnitude speed increase and handle difficult data sets where the original version fails completely.

1. INTRODUCTION

Coherent diffractive imaging (CDI)—imaging an object based on the way it diffracts—is one of those appealing research areas that combines important applications with opportunities for experimental ingenuity, algorithmic innovation, and mathematical exploration. A good example of this is ptychography [1], a recent addition to the CDI family that has inherited these traits and is now establishing a reputation for innovation in its own right. The ptychographic concept is straightforward: illuminate a specimen (the "object") with a small "probe" beam, measure the resulting diffraction pattern, move the specimen laterally by a fraction of the probe diameter, and repeat. However, this simple idea has proven to be a powerful way to better condition the inversion algorithms that recover an image from diffraction data (by solving for the missing phase) and to simultaneously extend the field of view. Helped by ever-improving computing power and detector technology, the applications for ptychography are diversifying, from live cell tracking [2] to 3D x-ray imaging of microchips [3] and mapping the electron phase signal [4]. On the experimental side, innovations include Fourier ptychography, which adapts a standard visible-light microscope platform to realize high contrast, super-resolved gigapixel images of live cells [5], and Bragg ptychography, where diffraction patterns are recorded from the Bragg reflections of a crystal to determine structural properties [6].

Algorithms for ptychography have progressed apace, exploiting the redundancy in ptychographic data to relax the initially required stringent experimental conditions. Originally, the probe illumination was assumed to be well characterized and fully coherent, the specimen positions were assumed to be accurate, diffraction data was assumed to be noise-free, and the specimen was

assumed to be thin; ptychographic algorithms can now routinely handle partial coherence [7], solve for the probe [8–10], correct positioning errors [11–13], remove noise [14,15], and deal with multiple scattering [16,17].

Of all these advances, the ability to solve for the probe beam was the first to emerge and remains the most important. Although a number of variants have been suggested [18–21], especially in the area of Fourier ptychography (see Ref. [22] for a comprehensive review), most often researchers rely on one of three iterative algorithms for this purpose: the conjugate gradient (CG) [8], the difference map (DM) [9], or the extended ptychographic iterative engine (ePIE) [10]. (The relaxed averaged alternating reflections (RAAR) method has also proven effective in the work of Marchesini and colleagues, e.g., [23].) Among these algorithms, all but ePIE take a global approach to invert the ptychographic data set—that is, at each iteration they use the entire collection of diffraction patterns to perform a batch improvement to estimates of the probe and specimen, seeking to align them with the measured data. ePIE takes a different tack, using diffraction patterns one-by-one to iteratively revise the probe and object estimates in what has been described as a stochastic or incremental gradient approach [22]. We focus here exclusively on this incremental kind of algorithm, leaving a broader comparison between these and the numerous global/batch alternatives as future work.

Although, in general, ePIE converges reasonably robustly and at a reasonable rate, practical situations where it struggles to find a solution are not too difficult to come across. For instance, in ptychographic experiments using a focused beam to form the probe, the exact distance of the specimen from the beam focus is difficult to measure accurately, leading to a poor initial probe model with too much or too little curvature [24]. Ideally, a ptychographic reconstruction algorithm should be able to cope with this poor

first guess at the probe, but in this situation, ePIE often either fails completely or takes many iterations to converge. A second example arises from the common practice of using a diffuser in a ptychographic experiment to produce a randomly structured probe [25,26]. The diffuser provides diversity to the diffraction measurements, reduces the dynamic range, and can improve the resolution of the reconstructed specimen image, but guessing an initial probe in this case is difficult, and, depending on the degree of structure in the probe and the complexity of the object, ePIE may again take hundreds of iterations to converge or fail completely.

Despite these potential difficulties, both ePIE and its progenitor, the original PIE algorithm [1], have been widely used in a range of applications, including x-ray beam characterization [27], Fourier ptychography [5], optical gating [28], and reflection imaging with extreme UV light sources [29]. This is, perhaps, thanks to the benefits they offer in terms of easy coding, efficient memory use, a small number of tuning parameters, and a good speed boost from implementation using parallel computing resources such as graphics cards.

In this paper, we reexamine the PIE family of algorithms and show how two straightforward modifications can offer drastically improved robustness and convergence rate, so that problems such as those described above do not arise. Section 2 details the improvements, the first of which is a revision of the update functions that regulate changes to the probe and object estimates as the algorithms iterate, and the second of which borrows from the machine learning community the idea of *momentum*, a technique commonly used to accelerate training of weights in a neural network [30]. [The reader familiar with ptychography can skip forward to Eqs. (18)–(21) for a summary of these key ideas.]

We demonstrate our modifications using simulated and real data. For simulated data, we show in Section 3.C at least a twenty-fold improvement in convergence rate over the original ePIE scheme and a number of instances where ePIE fails but our new scheme successfully and quickly converges. We consider a real-world optical bench experiment in Section 3.D, where we demonstrate similar improvements to those from our simulations and obtain successful image reconstructions in cases where existing algorithms stagnate.

2. ePIE, PIE, rPIE, AND mPIE

A. Overview of Operation

The experimental apparatus used for ptychography comprises a specimen mounted on a linear x/y translation stage and a coherent illuminating beam of photons or electrons called the probe, which is localized to a small area on the specimen surface by a masking aperture or focusing optics. Completing the setup is a detector placed some distance downstream from the specimen to record diffraction patterns, which in some cases may first be magnified or projected by intermediary optics [4]. In the experiment itself, the translation stage is programmed to shift the specimen through a grid of positions, and at each position the detector records a diffraction pattern. A grid spacing of around 20%–30% of the probe diameter ensures sufficient redundancy in the data for the image reconstruction process.

In this paper, we label the set of $j = 1 \ldots J$ diffraction patterns recorded in the experiment as $I_{j\mathbf{u}}$, and the corresponding set of specimen x/y positions as $\mathbf{R}_j = (x_j, y_j)$. We index the $M \times N$

pixels of the diffraction patterns using the pair of integers $\mathbf{u} = (m_1, n_1)$. The distance from the specimen to the detector is z, the wavelength of the probe is λ and the pixel pitch of the detector/camera is Δ_c.

The PIE-style procedure to reconstruct a specimen image from ptychographic data is common to all the algorithms we discuss below and is detailed in Fig. 1 (the shading indicates operations where this paper improves over existing algorithms). The process begins with initial estimates of the probe, $P_{0\mathbf{r}}$, and the object, $O_{0\mathbf{x}}$. Here $\mathbf{r} = (m_2, n_2)$ indexes the $M \times N$ pixels of the probe, which has the same pixel dimensions as the diffraction patterns, and $\mathbf{x} = (k, l)$ indexes the $K \times L$ pixels of the object. (To accommodate the specimen shifts, K and L are chosen to be much larger than N and M.) The diffraction patterns are employed in a randomly shuffled order, s_j, to update the object and probe estimates, and a single PIE-type iteration comprises J passes through the flowchart of Fig. 1, after which each diffraction pattern in the shuffled "deck" will have been used to update the object and probe.

To model the diffracted wavefront that exits the specimen—the exit wave—the first step is to extract a box from the current object estimate with the same number of pixels as the probe. The
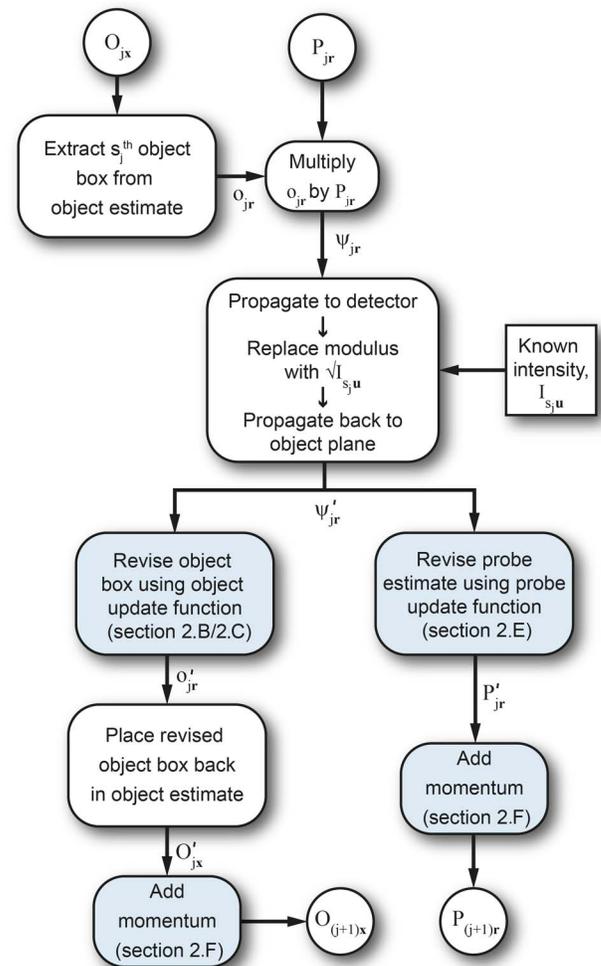
Fig. 1. Flowchart showing the operation of the PIE family of algorithms. One iteration of each algorithm consists of J loops through this process, so that each diffraction pattern is used, in the sequence s_j, to update the object and probe estimates.

jth object box, $o_{j\mathbf{r}}$, corresponds to the s_jth specimen position, and it is extracted from the full object estimate according to

$$o_{j\mathbf{r}} = O_{j\mathbf{x}_j}, \tag{1}$$

where the mapping from object to object box pixels is accomplished via Eq. (2):

$$\mathbf{x}_j = \frac{\mathbf{R}_{s(j)} - \mathbf{R}_{j(\min)}}{\Delta_{\mathbf{r}}} + \mathbf{r}. \tag{2}$$

Here, fractional pixel values can be accommodated as in Ref. [31], or the values of $\mathbf{x}_j$ can be rounded to the nearest integer. The pixel pitch in the object and probe estimates, $\Delta_{\mathbf{r}}$, depends on the method used to model propagation of the exit wave to the detector, which may be a Fourier transform if the detector is in the far field or a Fresnel or angular spectrum method for near-field ptychography. For far-field propagation via the Fourier transform,

$$\Delta_{\mathbf{r}} = \left(\frac{\lambda z}{M\Delta_c}, \frac{\lambda z}{N\Delta_c} \right), \tag{3}$$

and in the near-field, using the angular spectrum,

$$\Delta_{\mathbf{r}} = (\Delta_c, \Delta_c). \tag{4}$$

The exit wave model for the s_jth specimen position is then

$$\psi_{j\mathbf{r}} = P_{j\mathbf{r}} o_{j\mathbf{r}}, \tag{5}$$

where multiplication is elementwise. This exit wave is revised in the standard way to agree with the measured data—by replacing its propagated modulus with the square-root of the s_jth diffraction pattern, then propagating back to the specimen plane—giving the revised exit wave $\psi'_{j\mathbf{r}}$.

To this point, Fig. 1 follows the original PIE scheme. The first of our modifications, which we detail in Section 2.C, adapts the way the revised exit wave is used to update the object box and the probe estimate to give $o'_{j\mathbf{r}}$ and $P'_{j\mathbf{r}}$. Having implemented these updates, the revised object box is placed back in the full object estimate according to

$$O'_{j\mathbf{x}_j} = o'_{j\mathbf{r}}. \tag{6}$$

The second of our modifications, the momentum term, is added to the updated probe and object estimates to complete the journey through Fig. 1, after which the process repeats, with a freshly shuffled diffraction pattern order each time, until a fixed number of iterations are completed or a termination condition is met.

B. ePIE Object Update Function

We begin our discussion of update functions with a review of the object update function used by ePIE, which gives motivation to the revisions we suggest in Section 2.C. ePIE updates the object box according to Eq. (7):

$$o'_{j\mathbf{r}} = o_{j\mathbf{r}} + \alpha \frac{P^*_{j\mathbf{r}}}{|P_{j\mathbf{r}}|^2_{\max}} (\psi'_{j\mathbf{r}} - \psi_{j\mathbf{r}}), \tag{7}$$

where $P^*_{j\mathbf{r}}$ is the conjugate of $P_{j\mathbf{r}}$ and the other variables are defined in Fig. 1.

We will explain this equation in three ways, each of which gives some insight into its operation. First, a reasonable strategy to update the current object box, $o_{j\mathbf{r}}$, from the revised exit wave, $\psi'_{j\mathbf{r}}$, is simply to divide by the current estimate of the probe:

$$o'_{j\mathbf{r}} = \frac{\psi'_{j\mathbf{r}}}{P_{j\mathbf{r}}} = \frac{P^*_{j\mathbf{r}} \psi'_{j\mathbf{r}}}{|P_{j\mathbf{r}}|^2}. \tag{8}$$

However, it is clear that this update rule is poorly conditioned where the probe has a low intensity. Instead, we could assume that Eq. (8) gives a good update for the object in places where the probe is bright; we should accept the update in these areas. In areas where the probe is dim, Eq. (8) is likely to be inaccurate, so here we should retain the previous object estimate. Together these rules give

$$o'_{j\mathbf{r}} = (1 - w_{j\mathbf{r}}) o_{j\mathbf{r}} + w_{j\mathbf{r}} \frac{P^*_{j\mathbf{r}} \psi'_{j\mathbf{r}}}{|P_{j\mathbf{r}}|^2}$$
$$= o_{j\mathbf{r}} + w_{j\mathbf{r}} \frac{P^*_{j\mathbf{r}} (\psi'_{j\mathbf{r}} - \psi_{j\mathbf{r}})}{|P_{j\mathbf{r}}|^2}, \tag{9}$$

where $w_{j\mathbf{r}}$ is some spatially varying weight function that is close to unity where the probe is bright, and close to zero where it is dim. The ePIE update arises from a weighting equal to the normalized intensity of the probe, $w_{j\mathbf{r}} = \alpha |P_{j\mathbf{r}}|^2 / |P_{j\mathbf{r}}|^2_{\max}$, but we will see that this is far from the only sensible option.

For a second perspective on Eq. (7), consider the following error metric:

$$E_j^{(\text{obj})} = \sum_{\mathbf{r}} |P_{j\mathbf{r}} o_{j\mathbf{r}} - \psi'_{j\mathbf{r}}|^2. \tag{10}$$

We would like to find a revision to the object box that reduces this error and so brings the exit wave, $P_{j\mathbf{r}} o_{j\mathbf{r}}$, closer to the exit wave resulting from the diffraction pattern update steps of the algorithm (see Fig. 1). The gradient of this error with respect to the object box is

$$\nabla E_{j\mathbf{r}}^{(\text{obj})} = 2 P^*_{j\mathbf{r}} (P_{j\mathbf{r}} o_{j\mathbf{r}} - \psi'_{j\mathbf{r}}). \tag{11}$$

Since the error increases in the direction of this gradient, it can be reduced by moving the current object box by a small step, γ, in the negative gradient direction:

$$o'_{j\mathbf{r}} = o_{j\mathbf{r}} - \frac{\gamma}{2} \nabla E_{j\mathbf{r}}^{(\text{obj})} = o_{j\mathbf{r}} + \gamma P^*_{j\mathbf{r}} (\psi'_{j\mathbf{r}} - \psi_{j\mathbf{r}}). \tag{12}$$

Setting $\gamma = \alpha / |P_{j\mathbf{r}}|^2_{\max}$ gives the ePIE object update. This step size equates to a relatively well-known choice in other fields of optimization, for example, it is the Lipschitz constant of the gradient of $E_j^{(\text{obj})}$ [18] and the spectral radius used in a Landweber iteration [32]. Provided $\alpha \leq 1$, this choice of step is stable, and for convex problems convergence is guaranteed, but it is well known that gradient descent schemes converge slowly and stagnate at local minima [33]. That ePIE avoids the first of these undesirable properties is thanks to the incremental update steps, which greatly speed convergence; that it avoids the second is attributable to the structure of the probe, which causes each object pixel to move in a different gradient direction with each application of the update function. Nevertheless, the parallels with gradient descent are a strong indicator that improvements to ePIE are possible.

Thibault and Menzel gave a third view of the ePIE object update in the Supplement of their influential paper on ptychography with mixed states [7]. Their description considers a slightly different cost function that adds a regularization term to penalize large adjustments of the object box:

$$E_j^{(\text{reg})} = \sum_{\mathbf{r}} |P_{j\mathbf{r}} o'_{\mathbf{r}} - \psi'_{j\mathbf{r}}|^2 + \sum_{\mathbf{r}} u_{\mathbf{r}} |o'_{\mathbf{r}} - o_{\mathbf{r}}|^2. \tag{13}$$

The derivative of this cost function is

$$\nabla E_{j\mathbf{r}}^{(\text{reg})} = 2P_{j\mathbf{r}}^*(P_{j\mathbf{r}}o_{j\mathbf{r}}' - \psi_{j\mathbf{r}}') + 2u_{j\mathbf{r}}(o_{j\mathbf{r}}' - o_{j\mathbf{r}}). \quad (14)$$

Setting this gradient to zero and solving for the new object gives

$$o_{j\mathbf{r}}' = o_{j\mathbf{r}} + \frac{P_{j\mathbf{r}}^*(\psi_{j\mathbf{r}}' - \psi_{j\mathbf{r}})}{|P_{j\mathbf{r}}|^2 + u_{j\mathbf{r}}}. \quad (15)$$

Cost functions similar to Eq. (13) appear in the literature in various guises, for example within proximal algorithms [34], or as a "disappearing Tikhanov regularization" [35]. However, the term $u_{j\mathbf{r}}$ is somewhat unusual here as it is not a constant, but a spatially varying weighting of the degree of regularization, which, like the weighting function $w_{j\mathbf{r}}$, depends on the intensity profile of the probe. The rationale for selecting a suitable $u_{j\mathbf{r}}$ is that a large penalty should be added to the cost function for pixels in the object box where the probe illumination is dim, since in these regions the revised exit wave, $\psi_{j\mathbf{r}}'$ is highly susceptible to noise. Where the object box is strongly illuminated by the probe, the penalty term should be small, reflecting our greater confidence in $\psi_{j\mathbf{r}}'$ in these regions.

Looking back to Eqs. (9) and (12), it is clear that the weighting function idea, the gradient descent interpretation, and the regularized cost function can be linked by Eq. (16):

$$w_{j\mathbf{r}} = \gamma_{j\mathbf{r}}|P_{j\mathbf{r}}|^2 = \frac{|P_{j\mathbf{r}}|^2}{|P_{j\mathbf{r}}|^2 + u_{j\mathbf{r}}}. \quad (16)$$

For ePIE, this dictates that the regularization weighting should be

$$u_{j\mathbf{r}} = \frac{|P_{j\mathbf{r}}|_{\max}^2}{\alpha} - |P_{j\mathbf{r}}|^2. \quad (17)$$

C. Alternative Object Update Functions

For Rodenburg's original demonstrations of ptychography with simulated, optical, and x-ray data [1,36,37], a different object update function was used (at that time, the idea to solve for the probe had not been conceived). This original PIE scheme added a small regularization constant to the denominator of Eq. (9), and then weighted the function according to the normalized probe modulus, rather than the normalized intensity. Taken together, these two steps can be expressed in terms either of a new weighting function, $w_{j\mathbf{r}}$, or a new regularization weighting, $u_{j\mathbf{r}}$, as listed in Table 1. A number of authors have noted that the PIE update converges more rapidly than ePIE, which has been explained by casting PIE as a second-order gradient descent [22], although we will discuss shortly an alternative explanation based on the two weighting functions in Table 1. Note that in its original form, the PIE update with a fixed α value behaves differently for different maximum probe values; that is, the weighting $w_{j\mathbf{r}}$ cannot be expressed as a function of the normalized probe modulus, $|P_{j\mathbf{r}}|/|P_{j\mathbf{r}}|_{\max}$. This can be problematic,

since different experiments require retuning of the algorithm, so we suggest multiplying α in the update function by the maximum probe intensity, as shown in Table 1.

To complete our set of update functions, Eq. (18) gives a new form, which we suggest is superior to both PIE and ePIE:

$$o_{j\mathbf{r}}' = o_{j\mathbf{r}} + \frac{P_{j\mathbf{r}}^*(\psi_{j\mathbf{r}}' - \psi_{j\mathbf{r}})}{(1 - \alpha)|P_{j\mathbf{r}}|^2 + \alpha|P_{j\mathbf{r}}|_{\max}^2}. \quad (18)$$

We will refer to the reconstruction algorithm that uses Eq. (18) as the regularized PIE (rPIE), because its regularization weighting, $u_{j\mathbf{r}}$, appears as a particularly natural choice, although the original motivation was a convex combination of the denominator in Eq. (8) and the denominator in the ePIE update of Eq. (7). Table 1 details the two weighting functions corresponding to this new update function.

D. Comparing Update Functions

To give a better appreciation of the differences between the three object updates, Fig. 2 graphs the two weighting functions listed in

(a)

(b)
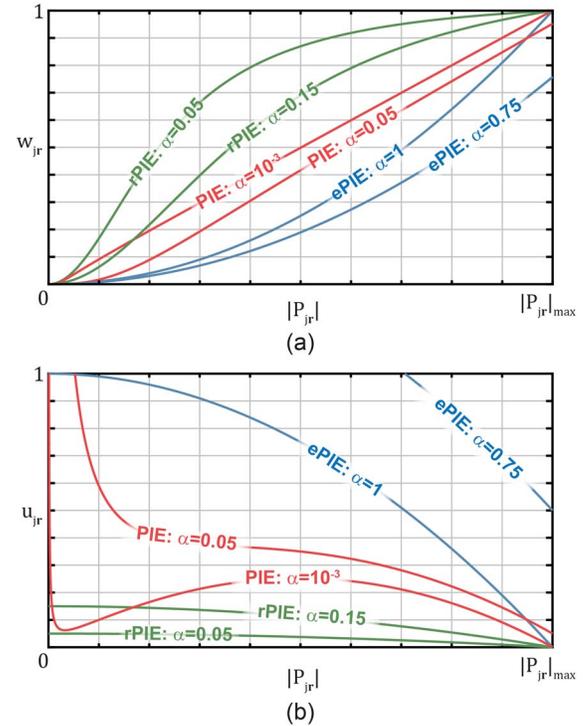
Fig. 2. Operation of the various object update functions. (a) Shows how the different algorithms weight the previous and new object estimates as a function of probe modulus [cf. Eq. (9)]. (b) Shows how they penalize changes to the object as a function of probe modulus [cf. Eq. (13)].

Table 1.   Three Different Interpretations of the Update Functions in PIE-Type Ptychographic Algorithms

Approach	$u_{j\mathbf{r}}$: Penalty Function Weighting of the Regularized Cost Function in Eq. (13)	$w_{j\mathbf{r}}$: Weighting of Previous and Revised Object Estimates in Eq. (9)	Update Function: Descent of the Gradient in Eq. (11) with a Spatially Varying Step Size
ePIE	$\frac{1}{\alpha}|P_{j\mathbf{r}}|_{\max}^2 - |P_{j\mathbf{r}}|^2$	$\alpha\frac{|P_{j\mathbf{r}}|^2}{|P_{j\mathbf{r}}|_{\max}^2}$	$o_{j\mathbf{r}}' = o_{j\mathbf{r}} + \alpha\frac{P_{j\mathbf{r}}^*}{|P_{j\mathbf{r}}|_{\max}^2}(\psi_{j\mathbf{r}}' - \psi_{j\mathbf{r}})$
PIE	$|P_{j\mathbf{r}}|_{\max}(|P_{j\mathbf{r}}| + \alpha|P_{j\mathbf{r}}|_{\max}^2/|P_{j\mathbf{r}}|) - |P_{j\mathbf{r}}|^2$	$\frac{|P_{j\mathbf{r}}|^3}{|P_{j\mathbf{r}}|_{\max}(|P_{j\mathbf{r}}|^2 + \alpha|P_{j\mathbf{r}}|_{\max}^2)}$	$o_{j\mathbf{r}}' = o_{j\mathbf{r}} + \frac{|P_{j\mathbf{r}}|P_{j\mathbf{r}}^*(\psi_{j\mathbf{r}}' - \psi_{j\mathbf{r}})}{|P_{j\mathbf{r}}|_{\max}(|P_{j\mathbf{r}}|^2 + \alpha|P_{j\mathbf{r}}|_{\max}^2)}$
rPIE	$\alpha(|P_{j\mathbf{r}}|_{\max}^2 - |P_{j\mathbf{r}}|^2)$	$\frac{|P_{j\mathbf{r}}|^2}{(1-\alpha)|P_{j\mathbf{r}}|^2 + \alpha|P_{j\mathbf{r}}|_{\max}^2}$	$o_{j\mathbf{r}}' = o_{j\mathbf{r}} + \frac{P_{j\mathbf{r}}^*(\psi_{j\mathbf{r}}' - \psi_{j\mathbf{r}})}{(1-\alpha)|P_{j\mathbf{r}}|^2 + \alpha|P_{j\mathbf{r}}|_{\max}^2}$

Table 1 as a function of the probe modulus for indicative values of α. Clear from Fig. 2(a) is the heavy drop-off of the ePIE weighting compared to the alternatives, which means a slow update rate for all but the most brightly illuminated pixels of the object. Adjusting the α of ePIE to values >1 increases the steps for less well-illuminated object pixels, but it also results in over-weighting of the well-illuminated regions—we will see in Section 2.F how momentum provides a much more reliable way to accelerate convergence than taking these large steps. Small α values for ePIE simply slow down convergence without noticeably improving stability.

Although the $w_{j\mathbf{r}}$ plot of PIE is a sensible, linear increase with probe modulus, the regularization weighting takes on a somewhat strange profile for small values of α [Fig. 2(b)], where it applies a stronger adjustment penalty to moderately illuminated object pixels than to those illuminated weakly. As α increases, the weighting then begins to flatten, and at $\alpha = 1/27 \approx 0.04$, the gradient no longer changes sign. It is therefore not suprising that α values around this level give a good trade-off between convergence rate and stability in our tests (see Section 3). (Note that the plots of Fig. 2 show the normalized weightings for PIE as discussed above and shown in Table 1.)

Regardless of the value of α, the PIE weighting function, $w_{j\mathbf{r}}$, is confined to the bottom righthand half of Fig. 2(a). This could be addressed with an extra multiplicative step-size parameter, but like ePIE, the result is over stepping for brightly illuminated pixels, and besides, introducing extra parameters into the algorithm compromises the simplicity we would like to retain. In contrast, our revised rPIE update can be tuned to occupy any region of Fig. 2(a). For $\alpha = 1$ it reverts to the conventional ePIE update (also with $\alpha = 1$), while smaller values give a much higher weighting to moderately well-illuminated object pixels than do PIE or ePIE. In terms of the regularization weighting, where different ePIE α values translate the quadratic curve up and down the y-axis, rPIE scales it in this direction instead, which from examination of the penalty term in Eq. (13), appears to be a more sensible way to control the step size in the update. Our tests show that rPIE remains stable even for α values as low as 0.05, and the curves for this α value shown in Fig. 2 give a good indication of the substantial improvement in convergence rate that this can provide.

Beyond the three methods discussed, a host of update functions can be derived by using the framework presented above to select sensible alternative weighting functions. Examples that we have used to successfully reconstruct ptychographic data include threshold weightings, power law or logarithmic functions, and piecewise linear functions. Some of these give good results in some cases, but none has proven more reliable or quicker than rPIE for the general case.

E. Probe Update Functions

The object update functions listed in Table 1 are adopted by the respective algorithms to update the probe simply by switching the roles of $o_{j\mathbf{r}}$ and $P_{j\mathbf{r}}$ in all the equations and replacing α with a second tuning parameter, β. Of course, this opens the possibility of using one kind of update for the probe and one for the object, for example, mixing the PIE probe update with the rPIE object update, etc., but this complication does not give any noticeable benefits. In fact, the form of the probe update function has far less influence on the success and convergence rate of the algorithms than does the object update. This can be explained by the dependence of the probe update functions on the normalized object modulus, which for weak phase objects varies only by a small amount from unity, and even for strong objects in our optical experiments, it is mostly above 0.5. The regions of the curves in Fig. 2 that are used in the probe update are therefore confined to the far righthand side, where they are broadly similar. The exception to this is Fourier ptychography, where the "object" exists in the Fourier domain and so has an extremely high dynamic range. Although we have not yet tested the different update functions on Fourier ptychographic data, Fig. 2 suggests adopting rPIE for the probe update in this case would be beneficial.

Our conclusion from the study we have undertaken for this work is that the ePIE probe update (or rPIE with $\beta = 1$) performs extremely well, and it is far better to introduce momentum, as detailed in the next section, than to spend time tuning the update function. There are also two minor extensions to the probe update that prove more useful than varying β: one controls the probe power and the other keeps the probe central to the reconstruction window. Both extensions are explained in Supplement 1.

F. Adding Momentum

Any of the object and probe updates described above can be enhanced by the powerful idea of momentum, which finds its principle use speeding up the training of neural network weights [30]. For problematic ptychographic datasets, such as those detailed in Section 3, adding momentum can realize an order of magnitude improvement in convergence speed—the price paid is an increase in the number of tuning parameters that control the new algorithm. The concept of momentum-based optimization is analogous to a cannon ball rolling down a hill, which becomes more and more difficult to divert from its course as it picks up speed. This translates into an ability to escape local minima (once the ball reaches the bottom of the hill it can climb up the other side) and to accelerate toward a minimum (the ball picks up speed even down a shallow slope); evidently, both properties are attractive features of optimization algorithms, and the overview by Ruder explains momentum in this context very nicely [38].

We implement ptychographic momentum in a slightly different way to that used for neural nets, as summarized by Fig. 3. Our approach allows the object (and probe) updates to progress without the addition of momentum for a fixed number of cycles, T, through the flowchart of Fig. 1. [In the diagram of Fig. 3(b), $T = 5$, so the first momentum update does not happen until the object and probe have been updated 5 times, by the first 5 diffraction patterns in the sequence s_j.] When momentum is to be applied, the first step is to update a *velocity map*, $v_{j\mathbf{x}}$, based on the current object estimate and the object estimate stored immediately after the $(j - T)^{\text{th}}$ update:

$$v_{j\mathbf{x}} = \eta_{\text{obj}}v_{(j-T)\mathbf{x}} + (O'_{j\mathbf{x}} - O_{(j+1-T)\mathbf{x}}), \qquad (19)$$

where $v_{0\mathbf{x}} = 0$, and $0 \leq \eta_{\text{obj}} < 1$ is a constant that can be interpreted as the "friction" of the ptychographic error landscape. Velocity will accumulate in the direction of any general trend in the object updates, while object pixels that oscillate will have a corresponding velocity map value that averages to zero.

To add momentum (or velocity) to the object, we have tested the two methods depicted by the diagram in Fig. 3(b). The first simply adds the velocity onto the previously stored object estimate:

$$O_{(j+1)\mathbf{x}} = O_{(j+1-T)\mathbf{x}} + v_{j\mathbf{x}}. \quad (20)$$

The second, inspired by the implementation of momentum suggested by Nesterov [38], we have found to be more stable and have used in all the results presented in Section 3. It adds a damped momentum term to the jth updated object estimate:

$$O_{(j+1)\mathbf{x}} = O'_{j\mathbf{x}} + \eta_{obj} v_{j\mathbf{x}}. \quad (21)$$

To best use this addition to the algorithm, it is necessary to reduce the step size in the update functions listed in Table 1. For ePIE, this is already included via the α parameter; for PIE and rPIE it means a further tuning parameter, γ_{obj}, in the update, which in the case of rPIE becomes

$$o'_{j\mathbf{r}} = o_{j\mathbf{r}} + \gamma_{obj} \frac{P^*_{j\mathbf{r}}(\psi'_{j\mathbf{r}} - \psi_{j\mathbf{r}})}{(1-\alpha)|P_{j\mathbf{r}}|^2 + \alpha|P_{j\mathbf{r}}|^2_{max}}. \quad (22)$$

Although momentum can be applied to any of the update styles, we will refer specifically to this enhancement of rPIE with momentum as the "momentum-accelerated PIE" (mPIE).

Momentum works exactly the same way for the probe update, which therefore requires two further parameters, η_{prb} and γ_{prb}. This set of new parameters are a somewhat unfortunate addition, violating as they do one of the benefits of incremental ptychography algorithms we listed in the introduction. In all of our trials to date, however, there has been a wide range of parameter values that produce excellent performance improvements; in Section 3,

we suggest a strategy to choose them and show that these extra complications can be worth the effort.

3. RESULTS

A. Description of Simulations and Optical Experiment

To investigate the performance of the four algorithms discussed in Section 2, we conducted two simulations and one optical bench experiment using the parameters shown in Table 2. The first simulation modeled an optical setup and tested the ability of the algorithms to converge successfully when both the probe and specimen are highly structured and have strong phase profiles. The second modeled a soft x-ray experiment with a weak phase specimen and a convergent beam probe, testing the ability to converge when the curvature of the initial probe estimate is inaccurate. Figures 4(a) and 4(b) show the amplitude and phase of the specimen used in both simulations, where the scalings were as listed in Table 2. A grid of 20×20 positions was used in each simulation, which gave the field of view indicated by the box in these figures; the circle indicates the approximate extent of the probes. Figure 4(c) shows the probe for the optical simulation. It was generated by modeling the wavefront from an aperture covered by a random phase mask, brought to focus by a 3 cm focal length lens, with the specimen plane positioned 100 µm beyond that focal point. The probe for the x-ray simulation [Fig. 4(d)] modeled a similar situation but with the random mask removed and the specimen positioned 750 µm beyond the focus of an ideal 10 mm focal length zone plate. Figures 4(e) and 4(f) show example diffraction patterns (with $M \times N = 512 \times 512$ pixels) from the optical and x-ray experiments, respectively. In the optical simulation, no noise was added to the diffraction patterns, while a small amount of Poisson-distributed shot noise was modeled in the x-ray simulation, equating to an average count in each diffraction pattern of 10^9.

The optical bench setup mimicked the optical simulation, with a microscope objective (NA = 0.25) used to focus the wavefront from an aperture covered by a plastic film diffuser. As a specimen, we used a microscope slide covered with lily pollen, which has a complicated structure at the micrometer scale and scatters strongly. The detector was an AVT Pike CCD, binned by a factor of 4, so that, as for the simulations, the diffraction patterns each contained $M \times N = 512 \times 512$ pixels. Three exposures, of 500, 5000, and 25,000 µs, were taken at each of the 30×30 specimen positions and stitched together to improve dynamic range.
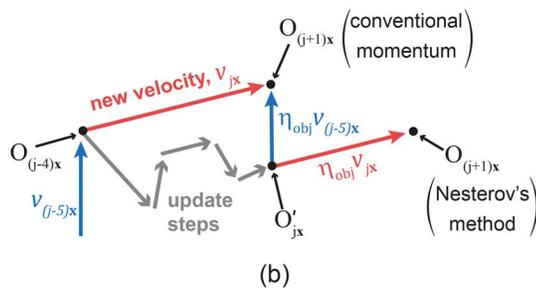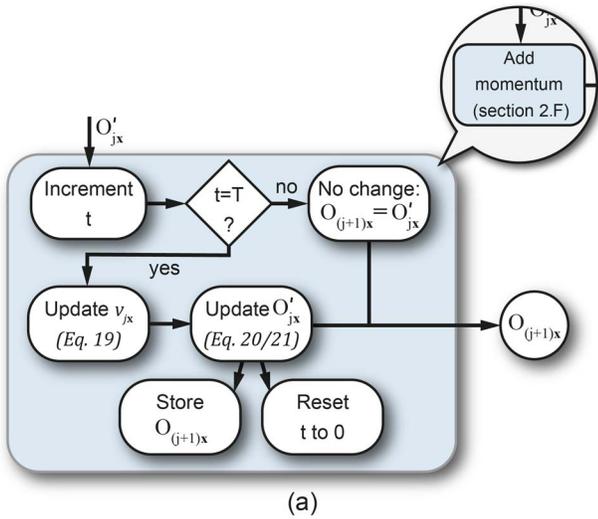
Fig. 3. Addition of momentum to the object update. (a) Shows how momentum is implemented within the flowchart of Fig. 1. (b) Visualizes two ways in which the momentum term can be added to the object estimate for the case $T = 5$: one is the conventional momentum scheme and the second is due to Nesterov.

Table 2. Parameters Used in Simulations and Experiment

Experiment Parameters	Optical Simulation	X-Ray Simulation	Optical Experiment
λ	635 nm	2.4 nm	635 nm
z	2.4 cm	50 cm	1.57 cm
Detector pitch	29.6 µm	27.0 µm	29.6 µm
Grid size	20×20	20×20	30×30
Step size	30 ± 10 µm	2.5 ± 0.75 µm	50 ± 15 µm
Object magnitude	0.25–1.0 a.u.	0.8–1.0 a.u.	0.20–1.0 a.u. (approx.)
Object phase	4π	$\pi/2$	3π (approx.)
Probe diameter	150 µm	10 µm	200 µm
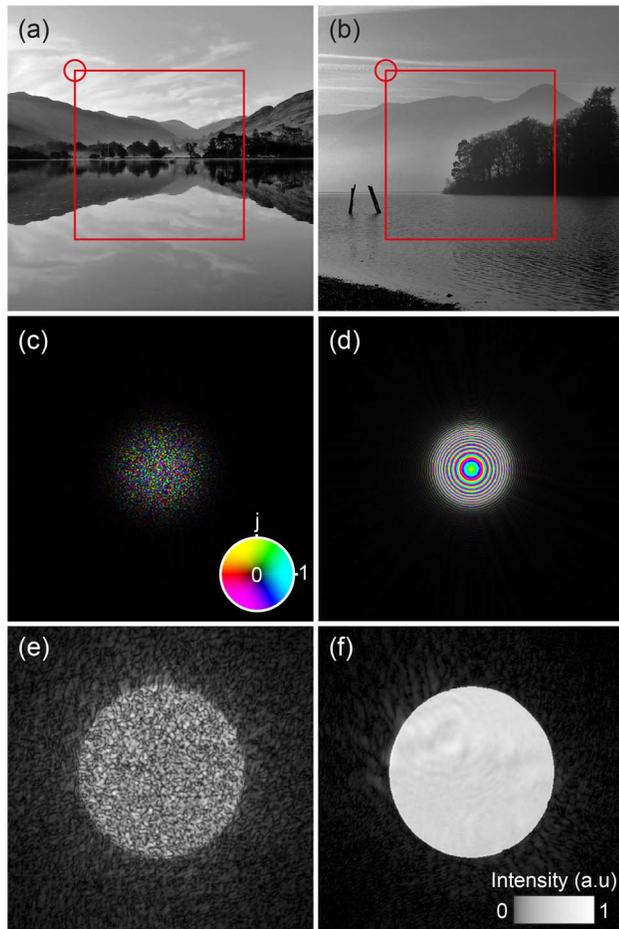
Fig. 4. Simulation setup. The specimen (a) modulus and (b) phase used in the two simulations; these template images were scaled as listed in Table 2. The red square indicates the field of view and the red circle indicates the size of the probes. (c) Shows the highly diffuse probe used in the optical simulation, and (d) shows the convergent beam probe used in the x-ray simulation. Both are displayed on the colorwheel scale shown in the inset to (c), which is also used in probe images of later figures. (e) and (f) give example diffraction patterns from the optical and x-ray simulations, respectively (contrast boosted).

B. Optical Simulation Results

Our first test directly compared the three update functions described in Section 2. For all the algorithms, an aperture of approximately the same size as the simulated probe was used as an initial probe estimate, and free-space was used as the initial object estimate. The random sequences in which diffraction patterns were addressed were kept consistent between the different algorithms.

We held $\alpha = \beta = 1$ for ePIE, both as a reference and because we have not observed that changes to these values confer any significant advantage. For PIE and rPIE, we ran a series of tests using different simulation images to determine α and β values that combined good convergence rates with stability. For both algorithms, setting α too low (less than $\sim 10^{-3}$ for PIE, less than $\sim 10^{-2}$ for rPIE) resulted in rapid destabilization of the reconstruction, while moderately low values (around 2×10^{-3} for PIE, 0.025 for rPIE) resulted in convergence properties that were sensitive to the random sequence, s_j, in which the diffraction patterns were employed. For PIE especially, setting α below around 10^{-2} also caused the mean intensity of the object reconstruction to slowly
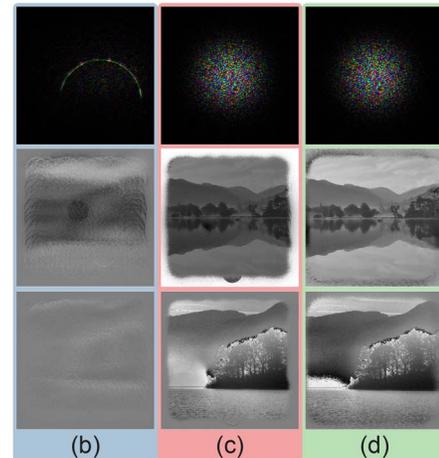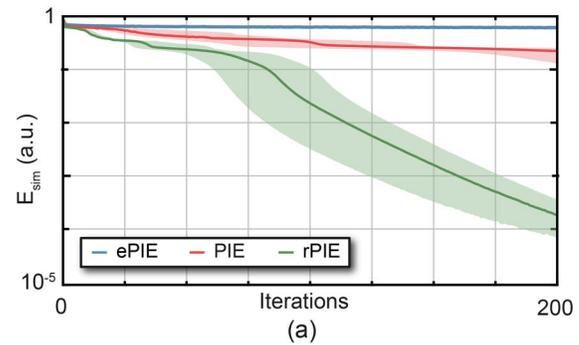
Fig. 5. Simulation results for the case of a strongly scattering specimen and a highly structured probe. (a) Convergence over 200 iterations of the ePIE, PIE, and rPIE algorithms. The solid traces indicate the median reconstruction from ten trials of each algorithm using different diffraction pattern orders, and the shaded regions indicate the range of convergence results over these ten trials. (b)–(d) The final probe (top) and specimen modulus (middle) and phase (bottom) for ePIE, PIE, and rPIE, respectively. Probes are shown on the colorwheel scale of Fig. 4.

reduce, eventually causing the reconstruction to destabilize; we suggest in Supplement 1 a solution to this problem. Both algorithms were far less sensitive to different β values. Our final parameters were $\alpha = \beta = 0.01$ for PIE and $\alpha = 0.05$ and $\beta = 1$ for rPIE.

We ran ten trials of the three algorithms, with a different set of s_j sequences in each trial. The progress of the trials are shown in Fig. 5(a), where the solid traces indicate the median run and the shaded regions enclose the convergence curves of the ten trials. The error metric plotted, E_{sim}, is a direct real-space comparison of the simulated and reconstructed objects as the algorithms progress. Often, ambiguities that are inherent to ptychographic reconstructions—a constant phase and amplitude offset, a linear phase ramp, and a global translation—are ignored in simulation results; here we compensate for these ambiguities in the error metric, as Supplement 1 describes.

As Fig. 5(b) shows, ePIE fails to converge, while PIE realizes a reasonable image in each trial but with a lack of accuracy reflected in the high final error level. rPIE converges quickly and repeatably to a lower final error.

In our next test, we implemented mPIE by adding momentum to rPIE, using the same α and β values as before. We then explored the influence of the momentum parameters on the convergence properties of the algorithm. The strategy we arrived at after this exploration was as follows:

– Fix $T = 30$, $\eta_{obj} = \eta_{prb} = 0.9$.
– Use the α and β values from rPIE.
– Tune γ_{obj} and γ_{prb} for performance, as highlighted in Fig. 6(a).

In this instance, γ_{obj} and γ_{prb} values of 0.2 and 1 gave excellent results compared to rPIE, halving the convergence time and realizing a slightly lower final error [see Fig. 6(a)].

Next, we repeated the simulation with an increasingly strong phase object, starting with an object phase range of 2π and increasing it to 18π radians; this increases the complexity of the phase, introducing phase wraps and gradients that the algorithms find difficult to recover. Again, ten trials at each object strength were carried out, each with a different set of diffraction pattern orders, and the parameters were left unchanged from the original simulations. Figure 6(b) shows the results, with the median final error value of each algorithm plotted with solid traces and the range of final error values over the ten trials shown shaded. The figure highlights the advantage momentum offers not only in convergence rate but also in the ability to converge robustly and escape local minima in highly taxing inversion problems.

C. X-Ray Simulation Results

The initial object for the soft x-ray simulation was again free space. The initial probe was modeled as for the true probe used to simulate the data, but with a slightly different aperture and a 1000 μm defocus (the true probe in the simulations had a 750 μm defocus). This represents a challenge to reconstruction algorithms, which even for smaller defocus errors are prone to induce phase vortices in the probe; a problem underlined in supplementary Visualization 1. Because noise was included in this simulation, we retuned the PIE and rPIE algorithms. This time, a good

compromise between convergence speed, reliability, and final error value was achieved with $\alpha = \beta = 0.004$ for PIE and $\alpha = 0.1$ and $\beta = 1$ for rPIE. Interestingly, in this weak phase scenario PIE could accommodate a less heavily regulated update than in the strong object simulations, despite the noise. For the momentum tuning, following the steps from Section 3.B gave $\gamma_{obj} = 0.15$ and $\gamma_{prbj} = 0.5$.

Figure 7 shows the results, with the solid lines giving the median reconstruction result from ten trials and the shaded regions outlining the range of error values over all the trials. Of note here are the large variations in convergence rate for PIE and rPIE—which can take anywhere between 120–200 iterations and 100–150 iterations, respectively, to reach an error of 10^{-3}—and the high reliability and fast convergence of mPIE. Although ePIE appears to have stagnated in this test, when it is allowed to continue it does eventually reach an error of 10^{-4} after 1300 iterations, or around 25 times slower than mPIE.

D. Optical Bench Results

Since the optical bench experiment quite closely matched the optical simulation, we retained the parameter values used in that simulation for the reconstructions here, with the exception of a reduction in γ_{prb} to 0.2 for mPIE. As in the optical simulation, we initialized the reconstruction with an aperture estimate of the probe and free-space for the object. In Fig. 8(a), we plot the evolution of the diffraction error over 100 iterations of each

Fig. 6. Including momentum significantly improves performance. (a) Tuning the momentum part of the update does require some effort. However, (b) shows how effective it can be for difficult data sets, in this case, an increasingly strong phase object and a highly diffuse probe. Solid traces indicate the median final error over ten trials of each algorithm, and shading indicates the range of final errors over these trials.

Fig. 7. Performance of the different algorithms in a simulated x-ray experiment. (a) Their convergence over 300 iterations—solid traces are the median and shading is the range of results from ten trials of each algorithm; (b)–(e) final probe (top row) and specimen modulus (middle) and phase (bottom) images of ePIE, PIE, rPIE, and mPIE, respectively. Probes are shown on the colorwheel scale of Fig. 4.
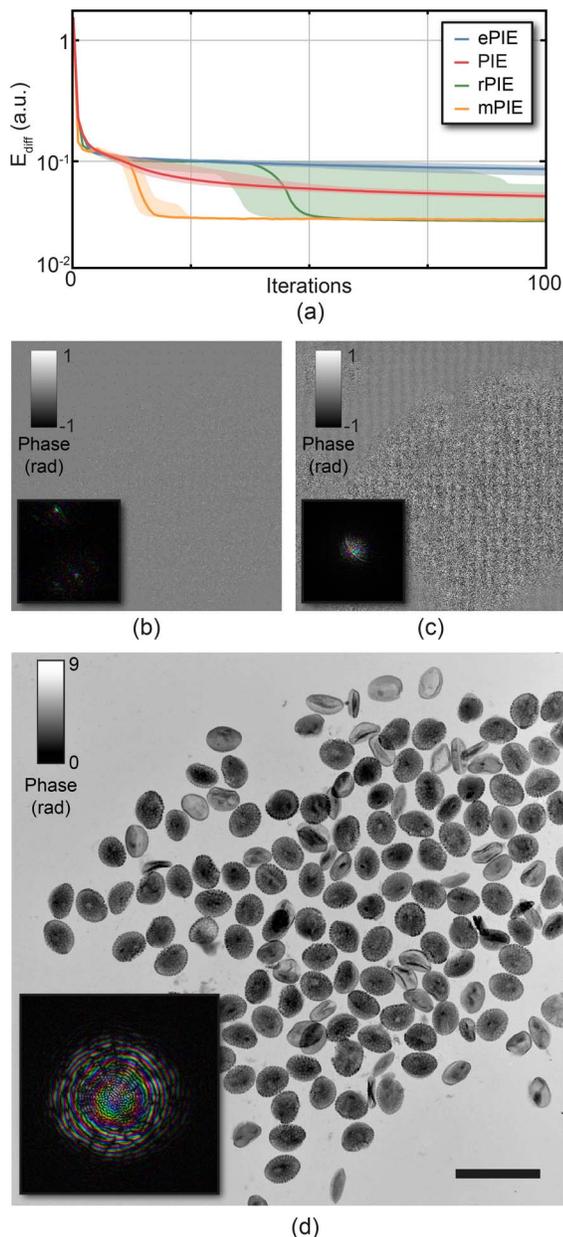
Fig. 8. Diffuse probe and an intricate specimen (lily pollen in this optical experiment) make for a difficult ptychographic reconstruction. (a) Progress of the diffraction error [Eq. (23)] over 100 iterations of four ptychographic algorithms. Ten runs of each algorithm were carried out with different position orders. The median run is shown as a solid trace, while all other outcomes were within the correspondingly colored boundaries. ePIE and PIE give the phase images shown in (b) and (c), respectively. rPIE give reasonable results in most cases, and mPIE converges reliably within 25 iterations. Both give excellent reconstructions, such as that shown in (d), where the phase has been unwrapped. Insets to the images show the reconstructed probe displayed using the colorwheel of Fig. 4. Scale bar 200 μm.

algorithm, where again solid traces give the median and shading gives the range of 10 trials using different diffraction pattern sequences, s_j. The diffraction error is given by Eq. (23):

$$E_{\text{diff}} = \frac{\sum_{\mathbf{u}} |I_{s_j \mathbf{u}} - |\Im[\psi_{jr}]|^2|^2}{\sum_{\mathbf{u}} I_{s_j \mathbf{u}}^2}, \qquad \textbf{(23)}$$

where $\Im$ is the propagator used in the reconstruction (a Fourier transform in all the work presented here).

　　Both ePIE and PIE quickly stagnated in this test, becoming trapped in local minima. With the parameters we used, neither algorithm could produce a reconstruction that resembled the specimen, nor could we noticeably improve the images in Figs. 8(b) and 8(c) by retuning these parameters. rPIE does converge in most cases, but as in the x-ray simulations, its performance is sensitive to the diffraction pattern order used in the reconstruction. The median reconstructions from the ten trials of rPIE and mPIE produced visually indistinguishable results; the images in Fig. 8(d) show the unwrapped object phase and the probe from the median mPIE reconstruction.

4. CONCLUSION

Given only the results above, it would be reasonable to conclude that ePIE performs uniformly poorly. In fact, the data we have used in this paper deliberately targets pathologically difficult cases, and ePIE works remarkably well most of the time (at least in our work on the optical bench and with the electron microscope). There are also important additions to ePIE—for instance multi-slice 3D imaging [16], position correction routines [11–13], and multi-mode reconstruction [39]—whose performance within the revised rPIE and mPIE algorithms has yet to be assessed. That said, there is no cost associated with adopting rPIE over ePIE, and its tuning parameter simply offers far better control over the update rate. We do also find that the original PIE scheme outperforms ePIE in most cases, especially when the power correction step detailed in Supplement 1 is included, but it cannot match rPIE in terms of robustness and stability over a range of different experimental geometries and specimen types. mPIE converges far more quickly and to a lower error value than any of the other algorithms, and it is indispensable when the data is extremely difficult to invert, but in its current form, it suffers from a large set of parameters. We are working to simplify the tuning process, but for now can offer Table 3 as a suggested range of effective values.

　　Our future work in this area will investigate adaptations to the diffraction pattern update step, for example incorporating the noise models described by Godard *et al.* [15]; testing the algorithms described in Section 2 for Fourier ptychography; comparing the algorithms to batch methods such as DM and RAAR; and investigating the use of automated parameter scheduling, along the lines of that described in Ref. [40].

See Supplement 1 and Visualization 1 for supporting content.

Table 3. Suggested Parameter Ranges for mPIE

Parameter	α	β	η_{obj}	η_{probe}	γ_{obj}	γ_{probe}	T
Suggested min.	0.05	0.5	0.5	0.75	0.1	0.2	10
Suggested max.	0.25	5	0.9	0.99	0.5	1	100

REFERENCES

1. J. M. Rodenburg and H. M. L. Faulkner, "A phase retrieval algorithm for shifting illumination," Appl. Phys. Lett. **85**, 4795–4797 (2004).
2. J. Marrison, L. Raty, P. Marriott, and P. O'Toole, "Ptychography—a label free, high-contrast imaging technique for live cells using quantitative phase information," Sci. Rep. **3**, 2369 (2013).
3. M. Holler, M. Guizar-Sicairos, E. H. R. Tsai, R. Dinapoli, E. Müller, O. Bunk, J. Raabe, and G. Aeppli, "High-resolution non-destructive three-dimensional imaging of integrated circuits," Nature **543**, 402–406 (2017).
4. A. M. Maiden, M. C. Sarahan, M. D. Stagg, S. M. Schramm, and M. J. Humphry, "Quantitative electron phase imaging with high sensitivity and an unlimited field of view," Sci. Rep. **5**, 14690 (2015).
5. G. Zheng, R. Horstmeyer, and C. Yang, "Wide-field, high-resolution Fourier ptychographic microscopy," Nat. Photonics **7**, 739–745 (2013).
6. S. O. Hruszkewycz, M. Allain, M. V. Holt, C. E. Murray, J. R. Holt, P. H. Fuoss, and V. Chamard, "High-resolution three-dimensional structural microscopy by single-angle Bragg ptychography," Nat. Mater. **16**, 244–251 (2017).
7. P. Thibault and A. Menzel, "Reconstructing state mixtures from diffraction measurements," Nature **494**, 68–71 (2013).
8. M. Guizar-Sicairos and J. R. Fienup, "Phase retrieval with transverse translation diversity: a nonlinear optimization approach," Opt. Express **16**, 7264–7278 (2008).
9. P. Thibault, M. Dierolf, A. Menzel, O. Bunk, C. David, and F. Pfeiffer, "High-resolution scanning X-ray diffraction microscopy," Science **321**, 379–382 (2008).
10. A. M. Maiden and J. M. Rodenburg, "An improved ptychographical phase retrieval algorithm for diffractive imaging," Ultramicroscopy **109**, 1256–1262 (2009).
11. A. Maiden, M. Humphry, M. Sarahan, B. Kraus, and J. Rodenburg, "An annealing algorithm to correct positioning errors in ptychography," Ultramicroscopy **120**, 64–72 (2012).
12. F. Zhang, I. Peterson, J. Vila-Comamala, A. Diaz, F. Berenguer, R. Bean, B. Chen, A. Menzel, I. K. Robinson, and J. M. Rodenburg, "Translation position determination in ptychographic coherent diffraction imaging," Opt. Express **21**, 13592–13606 (2013).
13. A. Tripathi, I. McNulty, and O. G. Shpyrko, "Ptychographic overlap constraint errors and the limits of their numerical recovery using conjugate gradient descent methods," Opt. Express **22**, 1452–1466 (2014).
14. P. Thibault and M. Guizar-Sicairos, "Maximum-likelihood refinement for coherent diffractive imaging," New J. Phys. **14**, 063004 (2012).
15. P. Godard, M. Allain, V. Chamard, and J. Rodenburg, "Noise models for low counting rate coherent diffraction imaging," Opt. Express **20**, 25914–25934 (2012).
16. A. M. Maiden, M. J. Humphry, and J. M. Rodenburg, "Ptychographic transmission microscopy in three dimensions using a multi-slice approach," J. Opt. Soc. Am. A **29**, 1606–1614 (2012).
17. E. H. R. Tsai, I. Usov, A. Diaz, A. Menzel, and M. Guizar-Sicairos, "X-ray ptychography with extended depth of field," Opt. Express **24**, 29089–29108 (2016).
18. R. Hesse, D. R. Luke, S. Sabach, and M. K. Tam, "Proximal heterogeneous block implicit-explicit method and application to blind ptychographic diffraction imaging," SIAM J. Imaging Sci. **8**, 426–457 (2015).
19. A. J. D'Alfonso, A. J. Morgan, A. W. C. Yan, P. Wang, H. Sawada, A. I. Kirkland, and L. J. Allen, "Deterministic electron ptychography at atomic resolution," Phys. Rev. B **89**, 064101 (2014).
20. R. Horstmeyer, R. Y. Chen, X. Ou, B. Ames, J. A. Tropp, and C. Yang, "Solving ptychography with a convex relaxation," New J. Phys. **17**, 053044 (2015).
21. L. Bian, J. Suo, G. Zheng, K. Guo, F. Chen, and Q. Dai, "Fourier ptychographic reconstruction using Wirtinger flow optimization," Opt. Express **23**, 4856–4866 (2015).
22. L.-H. Yeh, J. Dong, J. Zhong, L. Tian, M. Chen, G. Tang, M. Soltanolkotabi, and L. Waller, "Experimental robustness of Fourier ptychography phase retrieval algorithms," Opt. Express **23**, 33214–33240 (2015).
23. S. Marchesini, H. Krishnan, B. J. Daurer, D. A. Shapiro, T. Perciano, J. A. Sethian, and F. R. Maia, "SHARP: a distributed GPU-based ptychographic solver," J. Appl. Crystallogr. **49**, 1245–1252 (2016).
24. S. Wang, D. Shapiro, and K. Kaznatcheev, "X-ray ptychography with highly-curved wavefront," J. Phys. **463**, 012040 (2013).
25. A. Maiden, G. Morrison, B. Kaulich, A. Gianoncelli, and J. Rodenburg, "Soft X-ray spectromicroscopy using ptychography with randomly phased illumination," Nat. Commun. **4**, 1669 (2013).
26. M. Stockmar, P. Cloetens, I. Zanette, B. Enders, M. Dierolf, F. Pfeiffer, and P. Thibault, "Near-field ptychography: phase retrieval for inline holography using a structured illumination," Sci. Rep. **3**, 1927 (2013).
27. F. Seiboth, A. Schropp, M. Scholz, F. Wittwer, C. Rödel, M. Wünsche, T. Ullsperger, S. Nolte, J. Rahomäki, K. Parfeniukas, S. Giakoumidis, U. Vogt, U. Wagner, C. Rau, U. Boesenberg, J. Garrevoet, G. Falkenberg, E. C. Galtier, H. Ja Lee, B. Nagler, and C. G. Schroer, "Perfect X-ray focusing via fitting corrective glasses to aberrated optics," Nat. Commun. **8**, 14623 (2017).
28. P. Sidorenko, O. Lahav, Z. Avnat, and O. Cohen, "Ptychographic reconstruction algorithm for frequency-resolved optical gating: super-resolution and supreme robustness," Optica **3**, 1320–1330 (2016).
29. M. D. Seaberg, B. Zhang, D. F. Gardner, E. R. Shanblatt, M. M. Murnane, H. C. Kapteyn, and D. E. Adams, "Tabletop nanometer extreme ultraviolet imaging in an extended reflection mode using coherent Fresnel ptychography," Optica **1**, 39–44 (2014).
30. I. Sutskever, J. Martens, G. E. Dahl, and G. E. Hinton, "On the importance of initialization and momentum in deep learning," in *Proceedings of the 30th International Conference on International Conference on Machine Learning* (2013), Vol. **28**, pp. 1139–1147.
31. A. M. Maiden, M. J. Humphry, F. Zhang, and J. M. Rodenburg, "Superresolution imaging via ptychography," J. Opt. Soc. Am. A **28**, 604–612 (2011).
32. P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering* (Springer, 2011), pp. 185–212.
33. J. Qian, C. Yang, A. Schirotzek, F. Maia, and S. Marchesini, "Efficient algorithms for ptychographic phase retrieval," in *Inverse Problems and Applications* (Contemporary Mathematics, 2014), Vol. **615**, pp. 261–280.
34. D. P. Bertsekas, "Incremental gradient, subgradient, and proximal methods for convex optimization: a survey," in *Optimization for Machine Learning* (2011), Vol. **3**.
35. N. Parikh and S. Boyd, "Proximal algorithms," Found. Trends Optim. **1**, 127–239 (2014).
36. J. Rodenburg, A. Hurst, and A. Cullis, "Transmission microscopy without lenses for objects of unlimited size," Ultramicroscopy **107**, 227–231 (2007).
37. J. M. Rodenburg, A. C. Hurst, A. G. Cullis, B. R. Dobson, F. Pfeiffer, O. Bunk, C. David, K. Jefimovs, and I. Johnson, "Hard-X-ray lensless imaging of extended objects," Phys. Rev. Lett. **98**, 034801 (2007).
38. S. Ruder, "An overview of gradient descent optimization algorithms," arXiv:1609.04747 (2016).
39. P. Li, T. Edo, D. Batey, J. Rodenburg, and A. Maiden, "Breaking ambiguities in mixed state ptychography," Opt. Express **24**, 9038–9052 (2016).
40. C. Zuo, J. Sun, and Q. Chen, "Adaptive step-size strategy for noise-robust Fourier ptychographic microscopy," Opt. Express **24**, 20724–20744 (2016).