White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/217776/

Version: Published Version

Article:
Maiden, A.M., Mei, W. and Li, P. (2024) WASP: weighted average of sequential projections for ptychographic phase retrieval. Optics Express, 32 (12). pp. 21327-21344. ISSN: 1094-4087

https://doi.org/10.1364/oe.516946

Reuse
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

Research Article Vol. 32, No. 12 / 3 Jun 2024 / Optics Express 21327
Optics EXPRESS

WASP: weighted average of sequential projections for ptychographic phase retrieval

ANDREW M MAIDEN,[1,2,*] WENJIE MEI,[1] AND PENG LI[2]

[1]EEE Dept. University of Sheffield, Sheffield, S1 3JD, UK
[2]Diamond Light Source, Harwell, Oxfordshire, OX11 0DE, UK
[*]a.maiden@sheffield.ac.uk

Abstract: We introduce the weighted average of sequential projections, or WASP, an algorithm for ptychography. Using both simulations and real-world experiments, we test this new approach and compare performance against several alternative algorithms. These tests indicate that WASP effectively combines the benefits of its competitors, with a rapid initial convergence rate, robustness to noise and poor initial conditions, a small memory footprint, easy tuning, and the ability to reach a global minimum when provided with noiseless data. We also show how WASP can be parallelised to split operation across several different computation nodes.

Published by Optica Publishing Group under the terms of the Creative Commons Attribution 4.0 License. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

1. Introduction

Over the past decade, ptychography has emerged as a breakthrough extension to Coherent Diffractive Imaging (CDI), deployed as an effective tool at wavelengths ranging from the picometre to the micron [1]. It has become widely practiced at X-ray synchrotrons [2], with significant promise in reserve for fourth-generation X-ray sources, and super-fast 4DSTEM detectors in the electron microscope have stoked considerable interest in this once-forgotten electron technique [3–5]. Conventional CDI reconstructs an image of a sample from a recording of a single interference pattern, generated by the sample's diffraction of a coherent beam of illumination; ptychography naturally extends this idea to use multiple patterns recorded at different sample positions. Overlap between the illuminated regions of the sample introduces "useful redundancy" into the data, conditioning the inverse problem of forming an image and making it amenable to many different optimization approaches. A growing number of such schemes have been proposed and tested. First was non-iterative, analytic inversion of ptychographic data [6], which required a focused beam of illumination and a diffraction pattern recording for every pixel in the reconstructed image. Interest in these analytic methods has returned in recent years, especially for electron microscopy [7], but twenty years ago their data requirements were prohibitive, leading Rodenburg and Faulkner to reimagine ptychography as a form of iterative phase retrieval. The first iterative optimization algorithm, the ptychographical iterative engine (PIE) [8], eliminated the need to record diffraction patterns for every reconstructed pixel. Instead, a large patch, or probe, of illumination was used and the sample scanned through a coarse grid of positions. This drastically reduced data requirements, but PIE suffered a key limitation: the need for an accurate model of the illuminating probe wavefront. Without it image quality degraded substantially, but determining the wavefront's phase was difficult and prone to error. This changed when Guizar-Sicairos and Fienup exploited the redundancy in the ptychographic data set to simultaneously recover an image of both sample and probe using a conjugate gradient algorithm [9], with work at the Swiss light source (SLS) soon showing that this process could spectacularly improve image resolution and clarity [10]. The SLS used a set projection algorithm, a version of the difference map [11], aligning ptychography with the broader CDI field. Shortly after, PIE was also extended to recover the probe, resulting in the ePIE algorithm [12]. Many more

#516946 https://doi.org/10.1364/OE.516946
Journal © 2024 Received 2 Jan 2024; revised 12 Apr 2024; accepted 15 May 2024; published 28 May 2024

Optics EXPRESS

methods have been demonstrated at X-ray synchrotrons since this early work, including various set projection methods [13,14] (particularly Luke's Relaxed Averaged Alternating Reflections (RAAR) algorithm [15]), the Alternating Directions Method of Multipliers (ADMM) [16], proximal algorithms [17,18], and maximum likelihood [19]. Meanwhile, Fourier ptychography at optical wavelengths has brought new ideas [20–24], Machine Learning has appeared on the scene – either as a source of optimisation algorithms and automatic differentiation routines [25,26] or as a direct solution method [27] – and in electron ptychography a flexible least-squares optimizer [28] was used recently to great effect to realise record-breaking picoscale resolutions [29].

Given this history, why another algorithm? Our work suggests the WASP approach described in this paper moulds together the benefits of its various forerunners. Like ePIE, it has a small memory footprint and a rapid initial rate of convergence; like RAAR or ADMM, it often converges to a global minimum when given perfect data (although in common with these approaches there are, as yet, no convergence guarantees); and importantly, it also parallelises in a natural way. In this paper we detail the WASP algorithm, provide pseudocode for its implementation, and assess its performance using simulated and real-world data. The paper is intended for those familiar with the basic ideas of ptychography who are interested in implementing and running the algorithms we describe on their own data, to which end we provide code and example datasets at Ref. [30]. WASP is also available as an Engine in PtyPy [31].

2. Forward model for ptychography

A ptychographic experiment proceeds as follows:

1. A region of a transmissive sample is illuminated by a probe beam of coherent radiation.
2. The wavefront exiting the sample propagates some distance through free-space to form a diffraction pattern, whose intensity is recorded by a pixelated detector.
3. The sample is translated laterally to illuminate a different region, and another diffraction pattern is recorded. The newly-illuminated region should overlap with previous regions by 70% or more.
4. Step 3 is repeated to cover a region of interest on the sample with an overlapping grid of illumination positions.

Figure 1 illustrates a forward model for this experiment. The probe is represented by the complex-valued matrix $\mathbf{P}_\mathrm{r}$, of dimension $[M, N]$ corresponding to the pixel dimensions of the detector and indexed by the pair of integers $\mathrm{r} = [m, n]$. The specimen is represented by a larger object matrix $\mathbf{O}_\mathrm{x}$ of dimension $[X, Y]$, addressed by the index $\mathrm{x} = [x, y]$. The $j = 1 \ldots J$ different lateral positions of the sample during the experiment map to a set of J "top left" indices into the object matrix, denoted $\mathrm{R}_j^\mathrm{tl} = [x_j^\mathrm{tl}, y_j^\mathrm{tl}]$.

To model the wavefront incident on the detector at the j^{th} sample position, first an $[M, N]$ sub-region of the object matrix is extracted, beginning at entry R_j^tl and ending at the bottom right entry, $\mathrm{R}_j^\mathrm{br} = \mathrm{R}_j^\mathrm{tl} + [M, N] - 1$. We will refer to this region as the j^{th} object box, $\mathbf{o}_{j\mathrm{r}}$, where:

$$\mathbf{o}_{j\mathrm{r}} = \mathbf{O}_{\mathrm{r}+\mathrm{R}_j^\mathrm{tl}}. \tag{1}$$

For optically thin samples [32], the wavefront exiting the sample, $\psi_{j\mathrm{r}}$, is well approximated by a multiplication of this object box with the probe:

$$\psi_{j\mathrm{r}} = \mathbf{P}_\mathrm{r} \cdot \mathbf{o}_{j\mathrm{r}}, \tag{2}$$

where the dot notation indicates elementwise multiplication of two matrices. Propagation of the exit wavefront to the detector plane produces $\mathbf{\Psi}_{j\mathrm{u}}$, a model of the wavefront incident on the

Optics EXPRESS

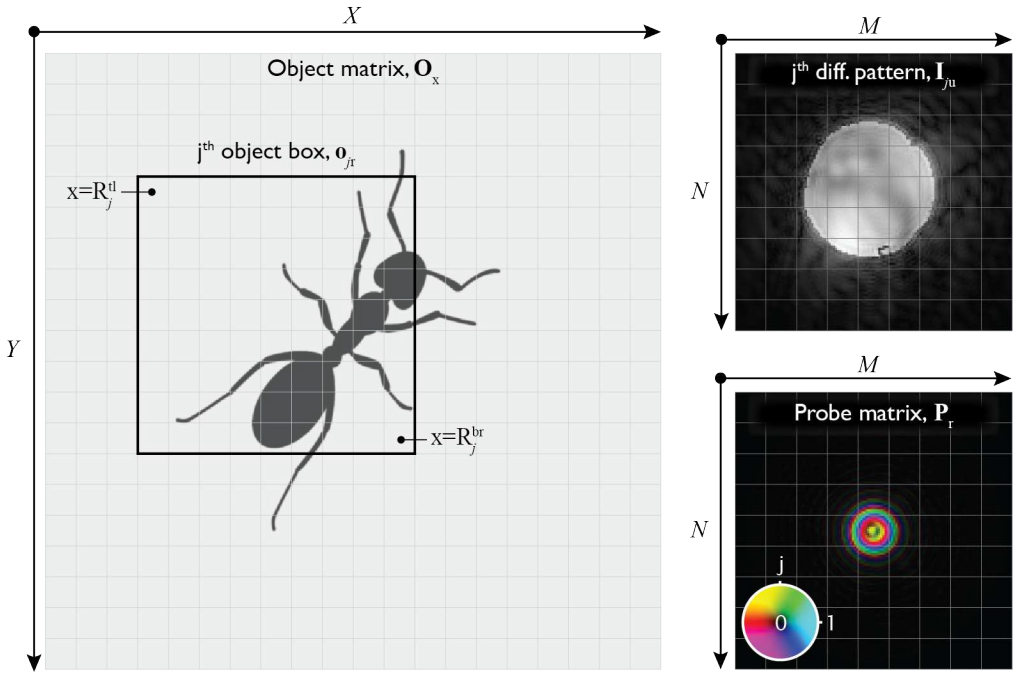

Fig. 1. An illustration of the variables used to model a ptychographic experiment. The j^{th} exit wave leaving the sample is modelled as an elementwise multiplication of $\mathbf{P}_\mathrm{r}$ and $\mathbf{o}_{j\mathrm{r}}$. A Fourier transform propagates this exit wave to the plane of a (far-field) detector. Ultimately, our goal is to find an object matrix and probe matrix such that the modulus of this modelled wave is equal to the square-root of the measured data, $\sqrt{\mathbf{I}_{j\mathrm{u}}}$, for every j (c.f. Eq. 7).

detector in the j^{th} sample position:

$$\mathbf{\Psi}_{j\mathrm{u}} = \mathcal{F}_z\left[\psi_{j\mathrm{r}}\right]. \tag{3}$$

Here the index $\mathrm{u} = [u, v]$ addresses the pixels of the propagated wavefront, z is the distance between the sample and detector, and free-space propagation over this distance is via the operator $\mathcal{F}_z$. For experiments where the detector is sufficiently distant from the sample to meet the far-field condition $\mathcal{F}_z$ is a fast Fourier transform (FFT), otherwise an angular spectrum or Fresnel propagator is used [33].

In the far-field the correct sample spacing for the model, relating entries of the probe and object matrices to physical points in space, is:

$$\delta_{xy} = \left[\frac{\lambda z}{M\delta_\mathrm{cam}}, \frac{\lambda z}{N\delta_\mathrm{cam}}\right], \tag{4}$$

where δ_cam is the pixel spacing of the detector. In the Fresnel or near-field region, the optical geometry for ptychography usually has associated with it a magnification, $\mathcal{M}$, so that the correct sample spacing is simply $\delta_{xy} = \delta_\mathrm{cam}/\mathcal{M}$.

Once the appropriate pixel spacing is determined, Eq. (5) converts the grid of x/y positions through which the sample is translated, D_j (in metres), to the corresponding offsets into the object matrix:

$$\mathrm{R}_j^\mathrm{tl} = \frac{\mathrm{D}_j - \min\left(\mathrm{D}_j\right)}{\delta_{xy}}. \tag{5}$$

Here the raw offsets are either rounded to the nearest integer value or fractional pixel shifts are incorporated through the Fourier shift theorem [34].

Optics EXPRESS

Equations (1) to (4) fully describe the forward model for ptychography, but one further modelling tool will prove useful in describing the various algorithms in the next Section. This is an "embedding" operator that copies the entries of an $[M, N]$ matrix into a larger $[X, Y]$ matrix of zeros, at the offset corresponding to the j^{th} position in the scan grid, $\mathrm{R}_j^{\mathrm{tl}}$. Taking the probe matrix as an example, the j^{th} *embedded* probe matrix is:

$$\mathbf{P}_{j\mathrm{x}} = \begin{cases} \mathbf{P}_{\mathrm{x}-\mathrm{R}_j^{\mathrm{tl}}} & \mathrm{R}_j^{\mathrm{tl}} \leq \mathrm{x} \leq \mathrm{R}_j^{\mathrm{br}} \\ 0 & o.w. \end{cases} \tag{6}$$

An example usage of this operation is to calculate the accumulated illumination intensity hitting each pixel of the object matrix, by summing up the probe intensities corresponding to each sample position. Using the operator of Eq. (6) this is simply written as: $\sum_j \left|\mathbf{P}_{j\mathrm{x}}\right|^2$.

3. Sequential projections and error reduction

In its most basic framing, the purpose of ptychographic algorithms is to find probe and object matrices that when fed through the model described above produce wavefronts, $\mathbf{\Psi}_{j\mathrm{u}}$, whose amplitudes match the diffraction patterns recorded during the experiment:

$$\left|\mathbf{\Psi}_{j\mathrm{u}}\right| = \sqrt{\mathbf{I}_{j\mathrm{u}}}, \quad \forall j, \tag{7}$$

where $\mathbf{I}_{j\mathrm{u}}$ are the recorded data. Optimising in terms of the modulus rather than the intensity of the data exploits the variance-stabilizing property of the square-root to better handle Gaussian-distributed noise [35], and so in a great many cases solving this optimisation problem suffices to reconstruct accurate, clean phase images from even quite noisy data – although for the lowest dose experiments, the highest levels of noise, or to accommodate experimental conditions such as partial coherence, the forward model and the framing of the optimisation can be adjusted [19,36–38].

3.1. Sequential projections

Equation 7 is a set of J equality constraints, linked more or less strongly with each other through the overlap between different probe positions. One way to find a probe and object that satisfy these constraints is to step through them one-by-one, usually in a random order, and at each step minimize the j^{th} regularized cost functions in Eq. (8):

$$\begin{aligned} \mathcal{L}_j^O &= \sum_{\mathrm{u}} \left(\left| \mathcal{F}_z \left[\mathbf{P}_{\mathrm{r}} \cdot \mathbf{o}_{j\mathrm{r}}^{\mathrm{new}} \right] \right| - \sqrt{\mathbf{I}_{j\mathrm{u}}} \right)^2 + \sum_{\mathrm{r}} \mathbf{A}_{\mathrm{r}} \cdot \left(\mathbf{o}_{j\mathrm{r}}^{\mathrm{new}} - \mathbf{o}_{j\mathrm{r}} \right)^2, \\ \mathcal{L}_j^P &= \sum_{\mathrm{u}} \left(\left| \mathcal{F}_z \left[\mathbf{P}_{\mathrm{r}}^{\mathrm{new}} \cdot \mathbf{o}_{j\mathrm{r}} \right] \right| - \sqrt{\mathbf{I}_{j\mathrm{u}}} \right)^2 + \sum_{\mathrm{r}} \mathbf{B}_{j\mathrm{r}} \cdot \left(\mathbf{P}_{\mathrm{r}}^{\mathrm{new}} - \mathbf{P}_{\mathrm{r}} \right)^2, \end{aligned} \tag{8}$$

where $\mathbf{o}_{j\mathrm{r}}^{\mathrm{new}}$ and $\mathbf{P}_{\mathrm{r}}^{\mathrm{new}}$ are the improved estimates sought at each step. Equation 8 captures the balance between the two priors that condition the ptychography problem: agreement with the data (the first terms in each equation) and consistency of the probe and object at each sample position (the second terms). The balance is fine-tuned by the regularization functions $\mathbf{A}_{\mathrm{r}}$ and $\mathbf{B}_{j\mathrm{r}}$, which dictate how strongly the revised object and probe estimates are anchored to their previous values. More or less any function of probe intensity that is small relative to the brightest probe pixels and large relative to the dimmer pixels can be used for $\mathbf{A}_{\mathrm{r}}$, and likewise any function of the object transmission that is small relative to highly transmissive object regions and large relative

Optics EXPRESS

to absorptive areas works for $\mathbf{B}_{j\mathrm{r}}$. For example, rPIE uses:

$$\begin{aligned}\mathbf{A}_\mathrm{r} &= \alpha\left(|\mathbf{P}_\mathrm{r}|^2_\mathrm{max} - |\mathbf{P}_\mathrm{r}|^2\right),\\ \mathbf{B}_{j\mathrm{r}} &= \beta\left(|\mathbf{o}_{j\mathrm{r}}|^2_\mathrm{max} - |\mathbf{o}_{j\mathrm{r}}|^2\right),\end{aligned} \tag{9}$$

where α and β are tuning constants, which when both set to unity result in ePIE; we will set out an alternative pair of regularizers later. Minimising the j^{th} cost functions in Eq. (8) can be thought of as finding a regularized projection onto a constraint set that represents the corresponding diffraction pattern data [17]. These projections can be found by taking Wirtinger derivatives of the costs, setting the result to zero and rearranged to solve for $\mathbf{o}_{j\mathrm{r}}^\mathrm{new}$ and $\mathbf{P}_\mathrm{r}^\mathrm{new}$. For the j^{th} cost function this leads to the update rules of Eq. (10):

$$\begin{aligned}\mathbf{o}_{j\mathrm{r}}^\mathrm{new} &= \mathbf{o}_{j\mathrm{r}} + \frac{\mathbf{P}_\mathrm{r}^* \cdot \left(\psi_{j\mathrm{r}}^\mathrm{new} - \psi_{j\mathrm{r}}\right)}{|\mathbf{P}_\mathrm{r}|^2 + \mathbf{A}_\mathrm{r}},\\ \mathbf{P}_\mathrm{r}^\mathrm{new} &= \mathbf{P}_\mathrm{r} + \frac{\mathbf{o}_{j\mathrm{r}}^* \cdot \left(\psi_{j\mathrm{r}}^\mathrm{new} - \psi_{j\mathrm{r}}\right)}{|\mathbf{o}_{j\mathrm{r}}|^2 + \mathbf{B}_{j\mathrm{r}}},\end{aligned} \tag{10}$$

where the '*' superscript indicates the entrywise complex conjugate of a matrix and $\psi_{j\mathrm{r}}^\mathrm{new}$ is the result of the ubiquitous phase retrieval 'replace modulus' operation, which employs the operator $\mathcal{F}_{-z}$ (simply an inverse FFT in the far-field geometry) to propagate a wavefront from the detector back to the sample plane:

$$\psi_{j\mathrm{r}}^\mathrm{new} = \mathcal{F}_{-z}\left[\frac{\mathbf{\Psi}_{j\mathrm{u}}}{|\mathbf{\Psi}_{j\mathrm{u}}|} \cdot \sqrt{\mathbf{I}_{j\mathrm{u}}}\right]. \tag{11}$$

To avoid zero-division in Eq. (11) a small constant can be added to its denominator.

For the object update, the final step of the projection is to paste the new object box back into the larger object matrix, so that:

$$\mathbf{O}_{\mathrm{r+R}_j} = \mathbf{o}_{j\mathrm{r}}^\mathrm{new}. \tag{12}$$

A single iteration of what we can call the sequential projections (SP) algorithm – of which ePIE and rPIE are particular instances characterised by the form of $\mathbf{A}_\mathrm{r}$ and $\mathbf{B}_{j\mathrm{r}}$ – shuffles the cost functions of Eq. (8) into a random order, then carries out the projections of Eq. (10) one-by-one, with the output of one pair of object/probe updates feeding the next until all J projections have been carried out. Although separate SP-type reconstructions can be tiled together [39], because the individual projections feed into one another these algorithms do not naturally parallelise; we will consider now an alternative framing that does.

3.2. Error reduction

Another way to solve Eq. (7) is to consider the entirety of the data in a single batch, by minimizing the two cost functions in Eq. (13):

$$\begin{aligned}\mathcal{L}^O &= \sum_j \sum_\mathrm{u}\left(\left|\mathcal{F}_z\left[\mathbf{P}_\mathrm{r} \cdot \mathbf{o}_{j\mathrm{r}}^\mathrm{new}\right]\right| - \sqrt{\mathbf{I}_{j\mathrm{u}}}\right)^2,\\ \mathcal{L}^P &= \sum_j \sum_\mathrm{u}\left(\left|\mathcal{F}_z\left[\mathbf{P}_\mathrm{r}^\mathrm{new} \cdot \mathbf{o}_{j\mathrm{r}}\right]\right| - \sqrt{\mathbf{I}_{j\mathrm{u}}}\right)^2.\end{aligned} \tag{13}$$

Minimising these cost functions forms the basis of set-based algorithms such as DM and RAAR, as well as implementations of conjugate gradient and maximum likelihood optimization

Optics EXPRESS

for ptychography [19,28]. Setting to zero the derivatives of Eqs. (13) w.r.t the new object and probe estimates results in Eq. (14):

$$\begin{aligned}\mathbf{P}_{\mathrm{r}}^{\mathrm{new}} &= \frac{\sum_j \mathbf{o}_{j\mathrm{r}}^{*} \cdot \psi_{j\mathrm{r}}^{\mathrm{new}}}{\sum_j |\mathbf{o}_{j\mathrm{r}}|^2} = \frac{\sum_j |\mathbf{o}_{j\mathrm{r}}|^2 \cdot (\psi_{j\mathrm{r}}^{\mathrm{new}} / \mathbf{o}_{j\mathrm{r}})}{\sum_j |\mathbf{o}_{j\mathrm{r}}|^2}, \\ \mathbf{O}_{\mathrm{x}}^{\mathrm{new}} &= \frac{\sum_j \mathbf{P}_{j\mathrm{x}}^{*} \cdot \psi_{j\mathrm{x}}^{\mathrm{new}}}{\sum_j |\mathbf{P}_{j\mathrm{x}}|^2} = \frac{\sum_j |\mathbf{P}_{j\mathrm{x}}|^2 \cdot (\psi_{j\mathrm{x}}^{\mathrm{new}} / \mathbf{P}_{j\mathrm{x}})}{\sum_j |\mathbf{P}_{j\mathrm{x}}|^2}.\end{aligned} \tag{14}$$

(Note the subscripts to the updated exit waves here: the j^{th} updated exit wave, $\psi_{j\mathrm{r}}^{\mathrm{new}}$, is the result of Eq. (11), whilst $\psi_{j\mathrm{x}}^{\mathrm{new}}$ uses Eq. (6) to embed this revised exit wave within a larger matrix of zeros at the offset corresponding to the j^{th} sample position.)

The second equalities in Eq. (14) reveal some intuition behind the mathematics. The term $\psi_{j\mathrm{r}}^{\mathrm{new}}/\mathbf{P}_{j\mathrm{x}}$ in the object update is an estimate of the j^{th} object box derived from the j^{th} diffraction pattern. The update equation is a weighted average of all J of these estimates, with a weighting for each estimate equal to the probe intensity that illuminated the object in the corresponding sample position. Bright areas of the probe contribute more to the sum than dark areas. Similarly, the term $\psi_{j\mathrm{r}}^{\mathrm{new}}/\mathbf{o}_{j\mathrm{r}}$ in the probe update is an estimate of the probe derived from the j^{th} diffraction pattern, which is averaged with a weighting corresponding to the opacity of the corresponding object box, so that transmissive areas of the object contribute more to the update than absorptive areas. The "WA" part of the name WASP comes from this weighted average description of the update functions.

The simplest way to deploy these equations as an iterative algorithm mirrors the error reduction algorithms developed in early work on phase retrieval by Gerchberg and Saxton [40] and Fienup [41], so we will refer to it as ptychographic error reduction, or ER. Ptychographic ER involves first calculating a full set of revised exit waves, via application of Eq. (11) for every value of j, feeding the result into Eqs. (14) to update the object and probe, then repeating. ER has the considerable advantage over SP that each instance of Eq. (11) can be calculated in parallel, before the resulting set of revised exit waves are fed into the probe and object update equations. It works best, however, when the updated probe from the first row of Eq. (14) feeds into the object update in the second row, meaning the probe and object updates should be carried out sequentially, not in parallel.

4. WASP: combining error reduction and sequential projections

The principle behind WASP is to use the output of the SP algorithm to feed calculation of the numerator and denominator sums in Eqs. (14), thereby accelerating convergence of the ER algorithm. This hybrid nature is captured in the pseudocode of Algorithm 1, which encompasses all three approaches. The pseudocode describes the full WASP algorithm, but by removing lines 9 to 12, 14 and 15 it becomes the SP algorithm, whilst by instead removing lines 7 and 8 it becomes an inline version of the ER algorithm, where the object and probe updates are calculated in parallel with the modulus projections.

In this pseudocode, the dot notation indicates elementwise multiplication of two matrices or multiplication of every element in a matrix by a scalar. The slash is elementwise division. **abs**()^2 returns the elementwise square of the absolute values in a complex matrix, and **conj**() returns the elementwise complex conjugate. Lines 4 to 7 implement Eqs. (1), (2) and the **ConstrainModulus** function of Eq. (11); lines 7 and 8 implement Eqs. (10) and (12); and lines 14 and 15 calculate the sums in Eq. (14).

The SP part of WASP is regulated by the functions $\mathbf{A}_{\mathrm{r}}$ and $\mathbf{B}_{j\mathrm{r}}$. The rPIE versions of these functions from Eq. (9) work well, as do many others; tolerance to different regularizers is high because the "WA" part of WASP has a damping effect on any divergent behaviour. Nevertheless,

Optics EXPRESS

we have found the following constant values for $\mathbf{A}_\mathrm{r}$ and $\mathbf{B}_{j\mathrm{r}}$ both highly reliable and computationally efficient:

$$\begin{aligned} \mathbf{A}_\mathrm{r} &= \alpha\langle|\mathbf{P}_\mathrm{r}|^2\rangle, \\ \mathbf{B}_{j\mathrm{r}} &= \beta, \end{aligned} \tag{15}$$

where <.> denotes the average over all entries in a matrix. For the object update, the average probe intensity is preferable to the maximum probe intensity, used in e- and rPIE, for two reasons: first the average is computationally cheaper than the maximum and second the average probe intensity remains fairly constant, whereas the maximum can vary widely and cause instability, especially in early iterations. The probe update is regularized by a constant value, again avoiding computation of the maximum. Normalisation of the constant is not required here since the amplitude of the object is known *a priori* to lie between 0 and 1 and this is enforced by additional constraints, as we will now briefly discuss.

Algorithm 1. WASP, the weighted average of sequential projections

Data: *position grid offsets* (R), *measured diffraction patterns* (I), *iterations (iters), initial probe* (probe), *initial object* (obj), *object and probe tuning parameters* (α, β).

for *($k = 1$ **to** iters)* **do**
  /* initialise numerator and denominator sums */
  numP = 0·probe, denP = 0·probe, numO = 0·obj, denO = 0·obj
  for *(a random order of each j in $1 \cdots J$)* **do**
    /* update exit wave to conform with diffraction data */
    objBox = obj($\mathrm{R}_j^{\mathrm{tl}} \rightarrow \mathrm{R}_j^{\mathrm{br}}$)
    exitWave = objBox·probe
    newExitWave = **ConstrainModulus** (exitWave,I_j)
    /* sequential projection update of object and probe */
    obj($\mathrm{R}_j^{\mathrm{tl}} \rightarrow \mathrm{R}_j^{\mathrm{br}}$) += **conj** (probe)· (newExitWave - exitWave)/(**abs** (probe)^2 + $\mathbf{A}_\mathrm{r}$)
    probe += **conj** (objBox)· (newExitWave - exitWave)/(**abs** (objBox)^2 + $\mathbf{B}_{j\mathrm{r}}$)
    /* update numerator and denominator sums */
    numO($\mathrm{R}_j^{\mathrm{tl}} \rightarrow \mathrm{R}_j^{\mathrm{br}}$) += **conj** (probe)·newExitWave
    denO($\mathrm{R}_j^{\mathrm{tl}} \rightarrow \mathrm{R}_j^{\mathrm{br}}$) += **abs** (probe)^2
    numP += **conj** (objBox)·newExitWave
    denP += **abs** (objBox)^2
  end
  /* weighted average update of object and probe */
  obj = numO/denO
  probe = numP/denP
  Apply additional constraints
end

Additional constraints further refine the object and probe estimates at the end of each WASP iteration. Examples of such constraints include suppressing hot pixels in the object by thresholding pixel amplitudes at some upper limit (whilst retaining the pixel phases) [10]; masking the probe extremities, so pixels outside a support area are set to zero [42]; forcing the probe to agree with a recording of an empty-space diffraction pattern (i.e. a diffraction pattern recorded without the sample present) [43]; correcting the probe power to a known value [44]; and recentring the probe and object to combat "probe wander", where the bright region of the probe appears to slowly move across the reconstruction window over the course of several iterations [14]. In the results we present next, recentering and hot pixel constraints are included. Although the more stringent constraints listed above are perhaps useful when data is extremely noisy or when overlap

Optics EXPRESS

redundancy is low due to a relatively large position grid step size, they have not been required in the algorithm tests we present next.

5. Results

In this Section, we test WASP through simulation and optical bench experiments. We compare it with the SP algorithms ePIE, rPIE, and a version of rPIE using Eqs. (15) instead of Eqs. (9) to update the object and probe. We also compare against the batch algorithms DM [10] and RAAR [13], and the ER algorithm as described in Section 3.2. rPIE and RAAR include tuning parameters. rPIE has two parameters, α_{rPIE} and β_{rPIE} that govern the update speed of the object and probe; we use values of $\alpha_{\mathrm{rPIE}} = 0.1, \beta_{\mathrm{rPIE}} = 1$, which give good general performance and align with the original paper [44]. RAAR has a single tuning parameter, β_{RAAR}, variable between $0 - 1$, with $\beta_{\mathrm{RAAR}} = 1$ corresponding to the original DM algorithm [10]. The literature [13] and our own experience give a reasonable nominal value of $\beta_{\mathrm{RAAR}} = 0.85$, although we will see that noisier data benefits from a lower value than this. For WASP (and the version of rPIE using Eq. (15)) we always use $\alpha = 2$, $\beta = 1$, which nicely balances convergence rate with robustness to initial conditions and noise.

5.1. Performance with simulated data

Our simulation scenarios are shown in Fig. 2. The object in 2(a, b) is a $[X, Y] = [920, 920]$ pixel complex-valued image of red blood cells derived from a real-world optical bench ptychography experiment. The $[M, N] = [128, 128]$ pixel probe in Fig. 2(c) is a model of the focal spot from a limited aperture lens, at a slight defocus so that the central bright region is ≈ 30 pixels in diameter. The shaded box and circle in Fig. 2(b) indicate the extent of the probe relative to the object. The scan pattern for this smaller probe comprises $J = 6400$ positions arranged in an 80×80 grid with an average step size of 6 pixels and ±20% random offsets from perfect uniformity, to eliminate the possibility of periodic artefacts in the reconstructions [16]. The solid red boundary trace in Fig. 2(b) indicates the extent of the scan pattern. The larger 512×512 pixel probe in Fig. 2(d) is modelled using a larger convergence angle and defocus. The scan pattern for this probe comprises $J = 400$ positions arranged in a 20×20 grid with an average step size of 36 pixels and ±20% random offsets, which covers the same region of interest (solid red boundary trace) as the scan grid of the smaller probe. Algorithm performance is measured by comparison of reconstructions with the known ground truth via a direct real-space simulation error metric, E_{sim}. This metric must account for the basic ambiguities that can affect the object reconstruction in ptychography: a phase ramp, phase offset, amplitude scaling, and – when the propagator is a Fourier transform – a global shift of both probe and object. We have described in previous work how these ambiguities can be accounted for when calculating E_{sim} [44]. To avoid any skewing of the metric by edge effects, which can pollute the extremities of the reconstruction window when using some algorithms, the error metric is calculated over the central region of the object matrix shown by the dashed red trace in Fig. 2(b). It is of interest to note that the edge areas do still influence the E_{sim} calculation since they affect the extremities of the probe, which transmits any resulting artefacts across the object. Regularization of Eq. (14) can be used to limit update speed in edge regions, with a smoothing effect in these areas, for both batch-type algorithms and WASP [16].

Our first simulations assessed raw convergence rate and the convergence floor for the seven different algorithms described above. Diffraction patterns were computed to double precision, without noise, using the smaller probe of Fig. 2. Reconstructions began with an initial object matrix of 1s, modelling free-space, and an initial Airy disc probe matrix of approximately the correct diameter, modelling a perfect focussed beam. Both probe recentering and hot pixel limiting were enabled for all of the algorithms, with the hot-pixel limit resetting to 2 the amplitudes of any object pixels whose amplitude exceeded 2 at the end of each iteration. Figure 3 shows the

Optics EXPRESS

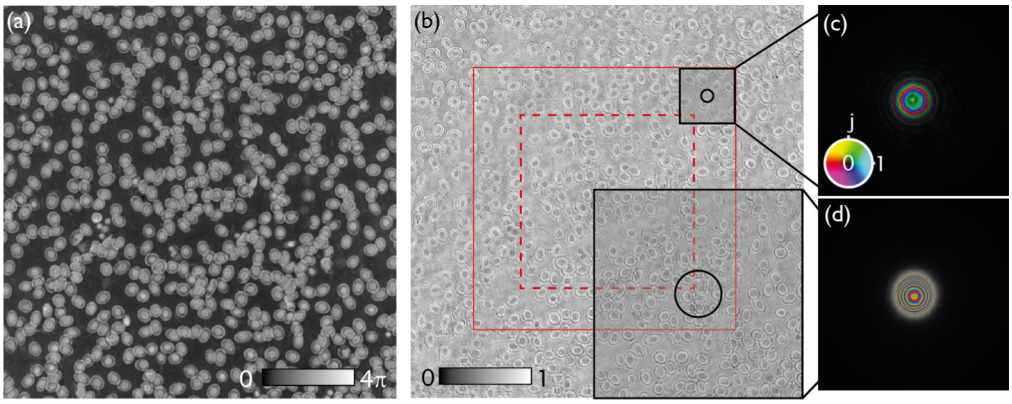

Fig. 2. The object and probes for the simulation scenarios. (a) and (b) show the object phase and modulus, (c) shows the small, 128 × 128 pixel probe on a colorwheel scale and (d) shows the larger, 512 × 512 pixel probe on the same scale. The solid red boundary in (b) indicates the extent of the scan patterns in the simulations, the dashed red boundary indicates the region of the reconstruction used in error calculations, and the black boundaries show the extent of the probes relative to the object.

progress of E_{sim} for the seven algorithms. All of the algorithms except ER reached a threshold of $E_{\text{sim}} = 10^{-5}$, at which level the central region of the reconstructed object amplitude and phase appear visually indistinguishable from the true object. WASP alone reached a global minimum at the working precision of our computer, which was determined by seeding the algorithms with the ground truth object and probe as initial estimates and measuring the resulting error after a few iterations. Maximum absolute phase errors in the central regions of the object reconstructions reached 13 mrad for DM, which had the highest final error of the convergent algorithms, and 9×10^{-11} mrad for WASP, corresponding to the lowest final error. To give an idea of how errors manifest across the field of view, the images in Fig. 3(b) show the difference between the amplitude of the ground truth object and the reconstructed amplitudes from ER, DM, ePIE and WASP. Of interest here is that SP algorithms produced a more uniformly distributed, white-noise error than did batch algorithms, which exhibited more systematic errors especially in the higher spatial frequencies.

Our second simulation explored the robustness of the same seven algorithms to initial errors in probe defocus. The object shown in Fig. 2 was reused, but this time with the larger 512 × 512 defocused probe and 400-point scan grid. The probe modelled the stopped-down beam from a soft X-ray source of 515 eV with an 8 mrad convergence semi-angle and a defocus of 750μm. The initial probe defocii in the reconstructions varied from 500 to 1000μm, equating to defocus errors from −33 to 33%. The results from reconstructions seeded by probes with these different defocus errors are shown in Fig. 4 with example initial probes shown for the defocii at 0 and ±33. The chart shows the number of iterations required for each algorithm to converge to $E_{\text{sim}} < 1 \times 10^{-4}$, for all instances where this level of error was reached; greyed-out entries denote instances where the error did not reach the threshold within 2000 iterations. Notable here is the relative robustness of the SP-type algorithms and WASP, relative to the batch algorithms, particularly RAAR and DM. These algorithms rapidly diverged in the first few iterations when the initial defocus error was larger than ≈ 10% and never fully recovered – although interestingly the basic ER approach did converge eventually for quite a wide range of defocus errors. In general, positive defocus errors (i.e. an initial probe that is too large) appear easier to recover from than cases where the probe is too small – perhaps because the overlap between probe positions is increased in the

Optics EXPRESS

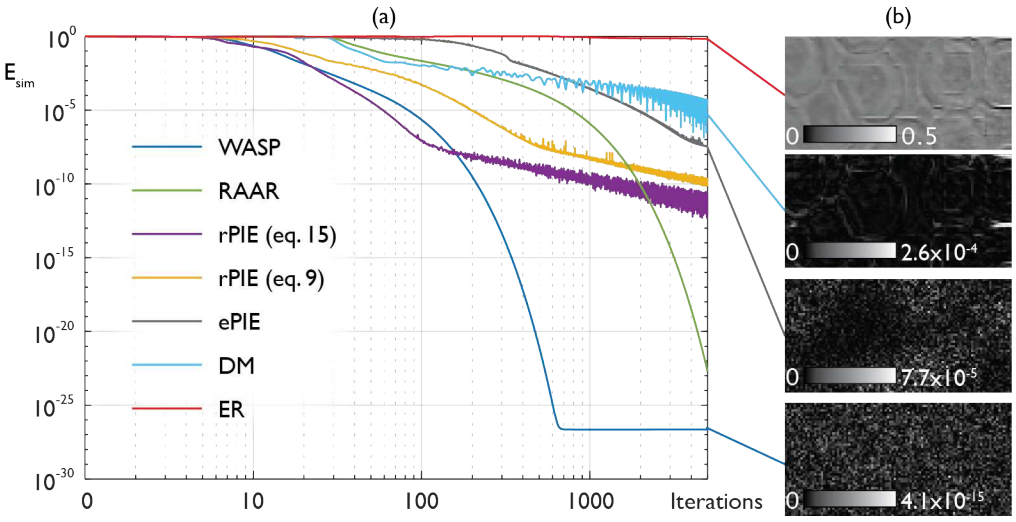

Fig. 3. Results from reconstructions based on perfect, double-precision, noise-free data. (a) The simulation error metric, E_{sim}, over 5000 iterations of seven ptychographic algorithms. The plots are shown on a log-log scale to highlight the initial convergence rate and the final error level for the different algorithms. Below an error of $E_{\text{sim}} = 10^{-5}$ the images reconstructed by the algorithms are visually very similar to the ground truth. (b) Examples of noise in the final reconstructions from four of the algorithms: these cutouts show the difference between the amplitude of the central 100×50 pixel regions of the reconstructions vs the central region of the ground truth.

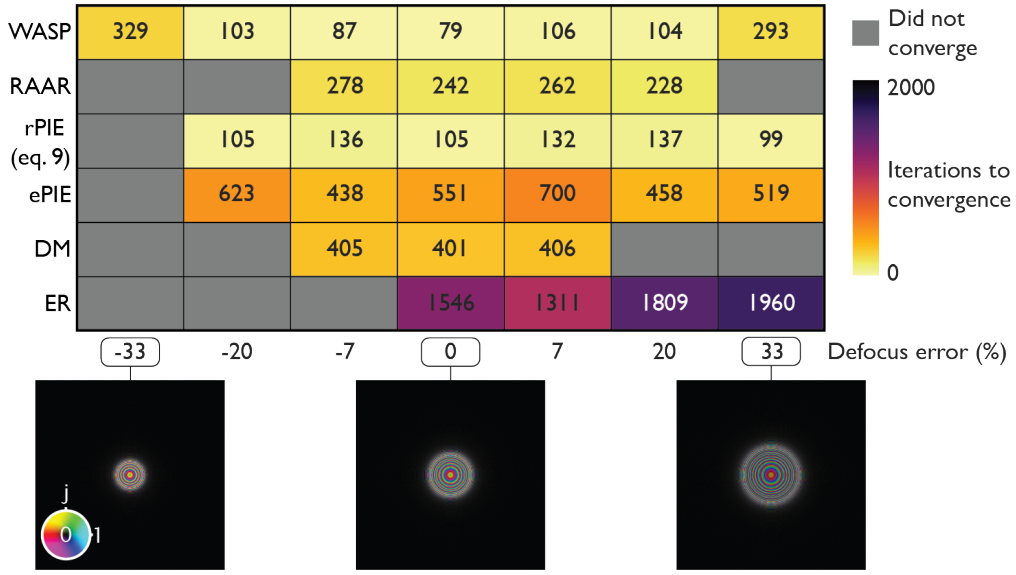

Fig. 4. A chart showing the convergence speed of different algorithms as a function of defocus error in the initial probe estimate. Shown below the chart are three examples of these probes that illustrate the extent of the initial error.

Optics EXPRESS

former case, leading to a stronger constraint (conversely, a too-small initial probe could result in zero overlap between probe positions, which would provide no constraint at all).

In a final simulation, Poisson-distributed noise was introduced into the data. Because minimising the cost functions of Eq. (8) corresponds to the maximum likelihood probe and object reconstructions for Gaussian-distributed noise, we can expect the algorithms to continue performing reasonably well for all but the very lowest noise levels, when statistical forms of the modulus constraint become more appropriate [19,45]. In our simulation, noise was calibrated such that the diffraction pattern of the probe generated 500, 000 counts, resulting in diffraction patterns having, on average, a 100 count maximum for any pixel. We used the larger probe (and smaller 20×20 scan grid) of Fig. 2(d), beginning the reconstructions with an initial probe that had a 7% defocus error. For noisy data, RAAR in particular benefits from a re-tuning of the parameter β_{RAAR} – noisier data appears to converge more reliably for lower values of this parameter, at the expense of convergence speed. To test this we trialled three different values of $\beta_{\mathrm{RAAR}} = 0.85, 0.75, 0.6$. Changing the WASP parameters has an impact on convergence only with very large values of α and β, when WASP becomes, essentially, the inline version of ER discussed earlier. Figure 5(a) shows the convergence of E_{sim} over 1000 iterations of WASP, rPIE, ER, and the differently tuned versions of RAAR. Extracts from the amplitudes of the final reconstructions are shown in Fig. 5(b), where the border colour corresponds to the legend in the accompanying graph. The amplitude is shown here as it inevitably suffers more from noisy data than does the phase. Surprisingly, ER performed very well in this particular simulation, and in very low-count experiments where the initial probe can be characterised accurately it is worth considering ER for the reconstruction; alternatively the tuning parameters for WASP can be increased as iterations proceed, so that it becomes, eventually, equivalent to ER.

5.2. Real-world results

Simulations are no substitute for experiments, but the huge range of wavelengths and sample types with which ptychography is conducted make any absolute measure of real-world performance difficult. With this caveat noted, our own limited experience does suggest WASP works very well in practice across many different experimental scenarios. The image mosaic in Fig. 6 gives three examples from experiments with far-field optical, near-field X-ray, and near-field optical ptychography [46–48], demonstrating algorithm performance on a strong phase sample (the ant in Fig. 6(a)), a noisy data set (the X-ray reconstruction in Fig. 6(b)) and a data set with a highly structured and poorly characterised probe (Fig. 6(c)). The ant experiment was conducted in a cone beam geometry [33] with a geometric magnification of 6.7×, using laser illumination at a wavelength of 635 nm. The X-ray experiment also used a cone beam geometry, this time with a magnification of 14× and at a beam energy of 9.6 keV. The far-field experiment used as a sample a prepared microscope slide holding a plant leaf structure, with a data collection NA of 0.25, a probe NA of 0.16 and laser illumination at a wavelength of 675 nm. All algorithm parameters remained identical to those used for the noise-free simulations detailed above.

Of the algorithms tested, only WASP and RAAR converged after 500 iterations. ePIE and rPIE both tended to excessively reduce the reconstructed object amplitude as iterations proceeded, which gave rise to the dark splotching appearing in those algorithm's results for the Siemens star and plant data sets. (The original rPIE paper described this tendency and showed how it could be combatted by periodically adjusting the probe power, something not included in these results.) Although RAAR takes considerably longer than WASP to converge (on average 257 iterations in these three examples *vs* an average of 42 iterations for WASP), both algorithms produce quite similar results in these tests, with the notable exception of the Siemens Star, where resolution is noticeably better in the WASP image.

Good performance on real-world data, such as that presented above, is clearly crucial to the success of any ptychographic algorithm. As data sets grow ever larger, however, an increasingly

Optics EXPRESS

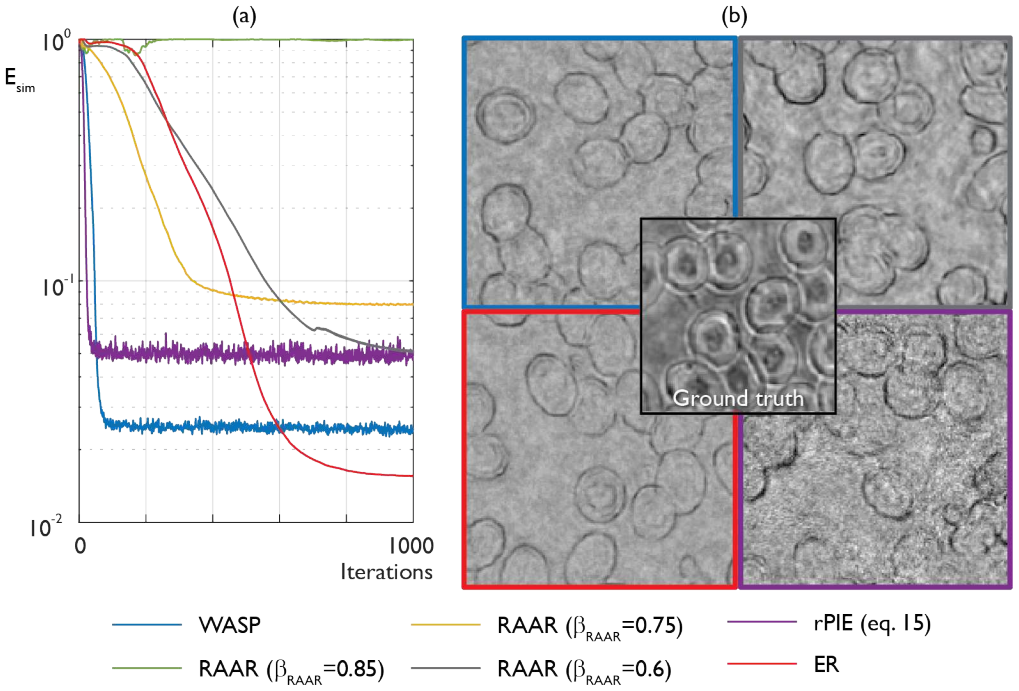

Fig. 5. Results from simulations using noisy data. Poisson-distributed shot noise was added to the forward model, such that the total counts of the probe was 500, 000. (a) convergence of different algorithms over the course of 1000 iterations. (b) cutouts of the reconstructed amplitudes from a selection of four algorithms, compared to the ground truth simulated object (shown in the centre). The amplitude is displayed as it suffers more acutely the impact of noisy data.

important additional consideration is efficient memory usage. Table 1 summarises the general memory requirements of the different algorithms tested here and gives example absolute numbers for the dataset and reconstruction of Fig. 6(c). Fully sequential algorithms are unbeatable in terms of memory footprint and store very little beyond the bare minimum object and probe matrices and diffraction data. WASP achieves a good balance between memory requirements and scalability, since it adds to sequential methods a natural way to distribute the reconstruction process over several nodes, as we will see in the next Section.

Table 1. The basic memory requirements for different algorithms, with example values derived from the data used in Fig. 6(c).

Algorithm	Number of stored digits	Example values ($\times 10^6$)	Total megabytes (double precision, 8 bytes per digit)
ePIE and rPIE	$2XY + 2MN + JMN$	5.58 + 0.52 + 104.9	888 MB
WASP	$5XY + 5MN + JMN$	13.9 + 1.31 + 104.9	961 MB
RAAR, DM and ER	$2XY + 2MN + 3JMN$	5.58 + 0.52 + 314.6	2565 MB

Optics EXPRESS

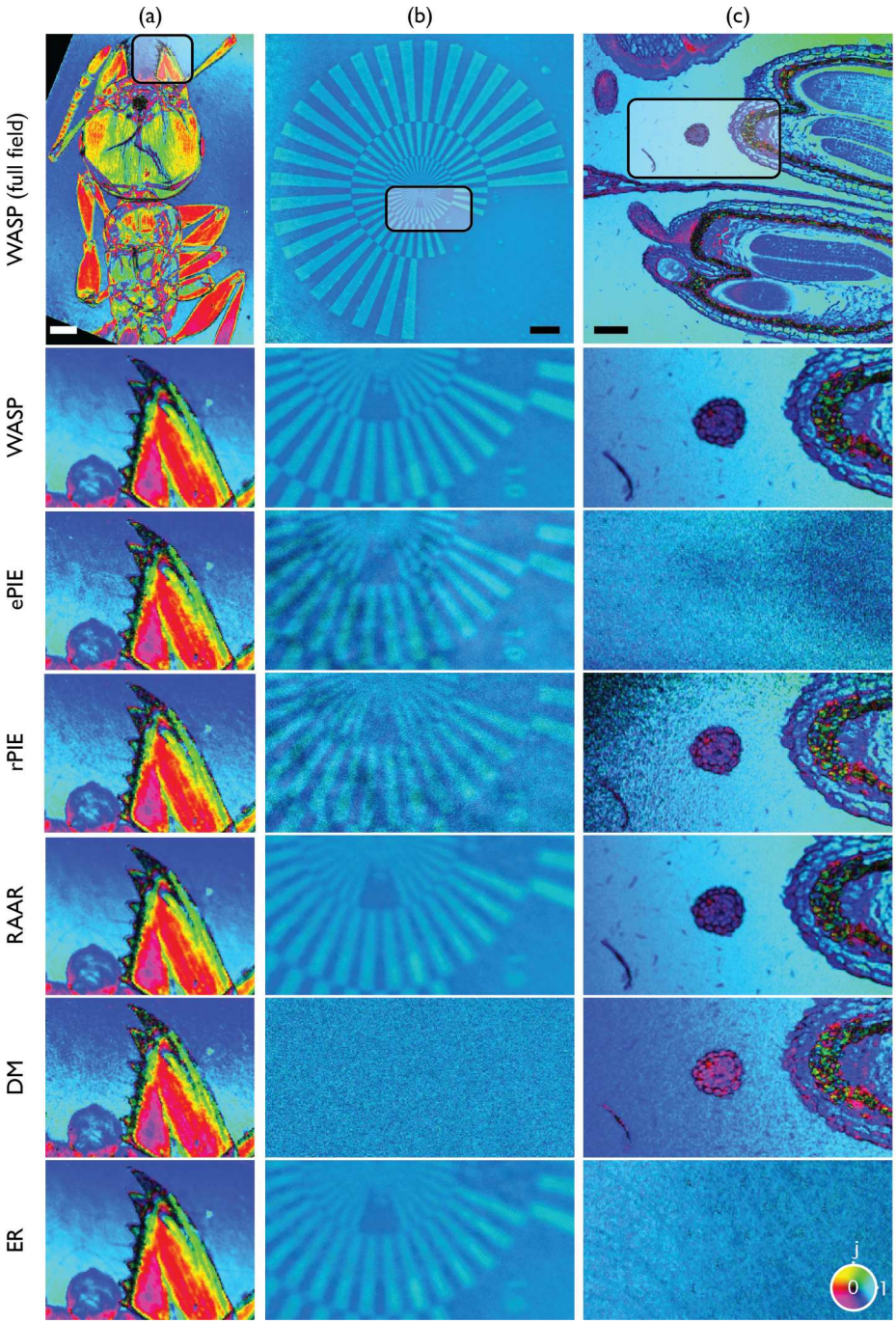

Fig. 6. A comparison of algorithm performance on three real-world data sets. The top row shows the full field of view from WASP reconstructions, with subsequent rows showing zoom-ins from the highlighted regions for the algorithms listed on the left. (a) An ant, reconstructed from cone beam near-field optical data [48]; (b) A Siemens Star, reconstructed from near-field X-ray data [47]; (c) a plant structure, reconstructed from far-field data with a diffuse probe [46]. Images are displayed on the colourwheel scale shown bottom right, scale bars in a,b and c are 200 μm, 10 μm, 100 μm.

Optics EXPRESS

6. WASP in parallel

Unlike batch algorithms such as DM and RAAR, SP algorithms cannot be fully parallelised, since the revised object and probe from each projection must feed into the next. One solution to this takes a leaf from neural network training algorithms and groups the cost functions of Eq. (8) into mini-batches [28]. Each mini-batch has associated with it a sub-set of projections, which are carried out in parallel using a version of Eq. (14) where the numerator sums are limited to the mini-batch's sub-set of indices. The mini-batches are processed in this way one by one, in a random order, with the output of one mini-batch feeding the next: parallel computation of the batches feed serially into the object and probe updates. Here we suggest an alternative where serial feeds parallel. The full data is again divided into random mini-batches, but this time (serial) sequential projections – running through the pseudo code of Algorithm 1 lines 3 to 13 – are carried out in parallel on the mini-batches. The outputs from each of these calculations are four partially-filled matrices: a numerator and denominator matrix for the object and the same for the probe. Once all of these partially-filled matrices are returned, they are summed and divided to implement Eq. (14), a process resembling the consensus algorithms discussed by Parikh and Boyd in their influential chapter on proximal algorithms [49].

The two pieces of pseudo-code below illustrate how this idea is implemented. To stretch an analogy too far, we will call the first piece of code, Algorithm 2, "the Hive" and the second piece, Algorithm 3, the "worker WASP". The workers' job is to feed the Hive with pieces of the numerator and denominator sums, which the Hive combines into new object and probe estimates that it then feeds back to the workers. Each worker can run for several iterations before returning the partial numerator and denominator sums, and we will assess the effect of the number of such sub-iterations and the number of workers in the Hive shortly.

Algorithm 2. The WASP hive, which coordinates the activities of the worker WASPs

Data: *iterations* (iters), *initial probe* (probe), *initial object* (obj), *number of workers* (numWorkers).

1 **for** ($k = 1$ **to** *iters*) **do**
  /* parallel for-loop */
2 **for** ($l = 1$ **to** *numWorkers*) **do**
    /* collect partial numerators and denominators from each worker */
3   [numO$_l$,numP$_l$, denO$_l$, denP$_l$] = **workerWASP** (obj, probe)
4 **end**
  /* weighted average update of object and probe */
5 obj = $\sum_l$numO$_l$/$\sum_l$denO$_l$
6 probe = $\sum_l$numP$_l$/$\sum_l$denP$_l$
7 *Apply additional constraints*
8 **end**

The nomenclature here is as for Algorithm 1. The sums and divisions at lines 5 and 6 of Algorithm 2 are elementwise, combining the worker numerators and denominators into single matrices equal in size to the object and probe then dividing pixel-by-pixel to implement Eq. (14). The *additional constraints* used in our tests are again suppression of hot pixels and probe/object recentering. The **workerWASP** call at line 3 of Algorithm 2 initiates Algorithm 3. Each worker is pre-loaded with a sub-set of the diffraction patterns, associated position data and update parameters – only the object and probe must be passed to the worker at each iteration, and it only need return the updated, partially-filled numerator and denominator matrices. The loop (lines 2 to 13) within each worker implements *subIters* iterations of the WASP algorithm on the worker's subset of data. Basic MATLAB code for the parallel version of WASP is included at Ref. [30] and the parallelisation is fully implemented in PtyPy [31].

Optics EXPRESS

Algorithm 3. A single worker WASP. Several workers run in parallel for a number of sub-iterations before feeding back to the hive.

Data: pre-loaded to each worker – *position grid offsets* (R), *allocation of diffraction patterns* (I), *object and probe update parameters* (α, β), *number of sub-iterations* (subIters).
Data: Passed each time the worker is called – *probe* (probe), *object* (obj)
/* initialise worker's numerator and denominator sums */
numP = 0·probe, denP = 0·probe, numO = 0·obj, denO = 0·obj
for *kk* = 1 **to** *subIters* **do**
  for *(a random order of each j in the worker's data allocation)* **do**
    /* update exit wave to conform with diffraction data */
    objBox = obj($R_j^{tl} \rightarrow R_j^{br}$)
    exitWave = objBox·probe
    newExitWave = **ConstrainModulus** (exitWave,I_j)
    /* sequential projection update of object and probe */
    obj($R_j^{tl} \rightarrow R_j^{br}$) += **conj** (probe)· (newExitWave - exitWave)/(**abs** (probe)^2 + $\mathbf{A}_r$)
    probe += **conj** (objBox)· (newExitWave - exitWave)/(**abs** (objBox)^2 + $\mathbf{B}_{jr}$
    /* update numerator and denominator sums */
    numO($R_j^{tl} \rightarrow R_j^{br}$) += **conj** (probe)·newExitWave
    denO($R_j^{tl} \rightarrow R_j^{br}$) += **abs** (probe)^2
    numP += **conj** (objBox)·newExitWave
    denP += **abs** (objBox)^2
  end
end
return *numO, numP, denO, denP*

We demonstrate here some features of the parallel implementation of WASP and compare the results to the full-batch version (Algorithm 1). Our tests used the blood cell object, smaller probe and 6400 position grid of Fig. 2(a–c). The first simulation assessed the effect of different numbers of worker WASPs on convergence, in terms of number of iterations, with a single sub-iteration computed by each worker (subIters = 1). The actual time per iteration was not considered, as we have not optimised the code sufficiently to make any test indicative. A second test fixed the number of workers at 4 and examined performance with different numbers of sub-iterations carried out within each worker.

Figure 7(a) shows the convergence of E_{sim} over 2500 iterations of WASP with a single worker and parallel versions with 2, 4, 8 and 16 workers. Convergence speed, measured as the point where the error hits 1×10^{-5}, dropped roughly linearly with the number of workers. For 16 workers, the convergence rate actually improved in later iterations compared to the 8-worker test, indicating there may be some diffraction pattern orderings and allocations that work better than others, something explored to some extent in Ref. [28] but worthy of further investigation. Figure 7(b) shows how convergence is affected by the number of sub-iterations conducted by each worker. As the number of sub-iterations increases, the number of full iterations needed to reach an error of 1×10^{-5} falls by a somewhat less predictable amount: from 500 for 1 sub-iteration through 370, 150, 50 and 30 for 2, 4, 8 and 16 sub-iterations. (We ascribe the odd behaviour at the very lowest error levels for the 16 worker result to the extremely small values involved, which were at the limits of the double precision accuracy of our machine.) To compare the different variables, the dashed trace in Fig. 7(a) is a copy of the 4 worker, 4 sub-iteration trace in Fig. 7(b). This shows that the increase in iterations required when splitting the data between several workers can be compensated quite effectively by allowing each worker to run several sub-iterations – this

Optics EXPRESS

is not a zero-sum game, since the sub-iterations are conducted in parallel on reduced-size data sets within each worker and so provide a net gain in computation time, as well as a reduction in overheads associated with transfer of data to and from each worker.

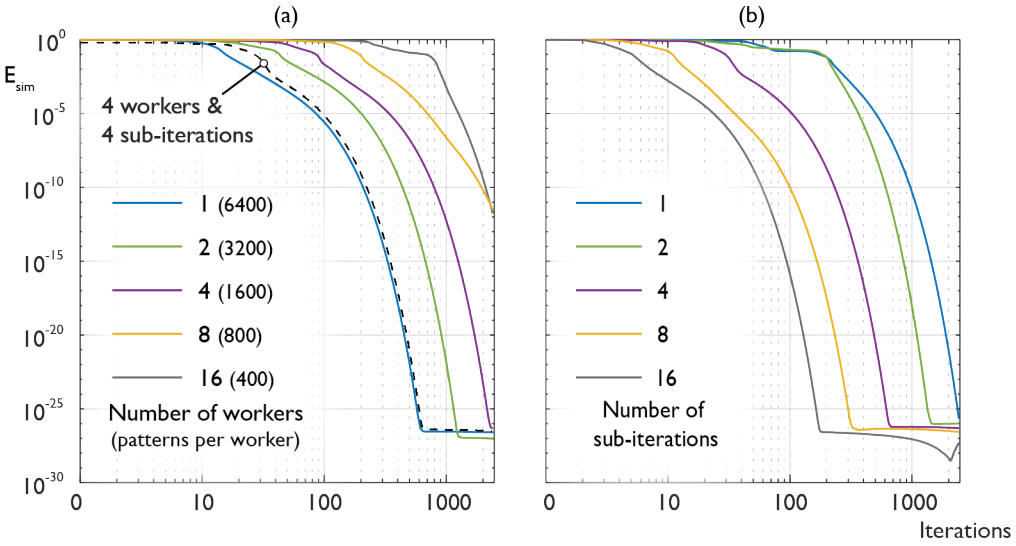

Fig. 7. Results from tests of the parallel version of WASP, both displayed on the E_{sim} scale shown on the left. (a) For a fixed number of sub-iterations (set to one here), the convergence rate reduces with the number of workers, although the random allocation of workers does influence this trend. (b) Conversely, the convergence rate improves with the number of sub-iterations carried out by each worker, here with the number of workers fixed at 4. For comparison, the dashed trace in (a) corresponds to the 4-worker, 4-sub-iteration trace in (b).

7. Conclusion

To date, ptychography has been implemented across wavelengths spanning approximately six orders of magnitude. Experimental configurations, noise sources, sample scattering strength and instrument instabilities are all similarly wide-ranging, and have led to algorithmic adaptations such as multi-slice [36] and modal decomposition [37] that can be implemented within any of the frameworks described above. This vast scope makes any definitive comparison of reconstruction algorithms difficult. That being said, our experience, limited as it is to only a corner of the possible implementations of ptychography, does suggest some general observations. First that SP algorithms handle initial conditions (especially a poorly-characterised initial probe) with a higher degree of robustness than batch algorithms; second, that SP algorithms converge more quickly in the initial iterations; third, that batch algorithms often reach a lower final error level and are less effected by high spatial frequency noise; and finally, of course, that the WASP algorithm we present here does a reasonable job of combining the benefits of both flavours of algorithm. Whether or not these observations are universal is difficult to say, but luckily the recent emergence of open source ptychographic code means experimenters are free to carry out further comparisons and to draw conclusions from their own data and analysis.

Whilst ptychographic data continues to present a unique and satisfying inverse problem, soluble using many different approaches, as the method matures the beauty and cleverness of what is "under the hood" matters less than practical considerations; ease of use and user experience still constrain more wide-spread uptake of the technique. Key to fully unlocking ptychography's potential are therefore robust solutions that run quickly and do not require extensive parameter

Optics EXPRESS

tuning or black-art algorithmic tricks. WASP goes some way toward meeting these criteria, and its basic framework suggests several variants and extensions that may further improve performance in the future.

Acknowledgements. We would like to thank Freddie Allars for useful discussions, Benedikt Daurer and Tim Poon for implementing WASP in PtyPy, Ziyang Hu and Yiqian Zhang for collecting optical bench data, and Darren Batey for his help collecting X-ray data at the I13-1 Coherence branch of the Diamond Light Source.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data and code underlying the results presented in this paper are available in Ref. [30].

References

1. J. Rodenburg and A. Maiden, *Ptychography* (Springer, 2019), pp. 819–904.
2. F. Pfeiffer, "X-ray ptychography," Nat. Photonics **12**(1), 9–17 (2018).
3. C. Ophus, "Four-dimensional scanning transmission electron microscopy (4d-stem): From scanning nanodiffraction to ptychography and beyond," Microsc. Microanal. **25**(3), 563–582 (2019).
4. C. Huang, J. S. Kim, and A. I. Kirkland, "Cryo-electron ptychography: Applications and potential in biological characterisation," Curr. Opin. Struct. Biol. **83**, 102730 (2023).
5. Y. Jiang, Z. Chen, Y. Han, *et al.*, "Electron ptychography of 2d materials to deep sub-ångström resolution," Nature **559**(7714), 343–349 (2018).
6. P. D. Nellist, B. C. McCallum, and J. M. Rodenburg, "Resolution beyond the 'information limit' in transmission electron microscopy," Nature **374**(6523), 630–632 (1995).
7. T. J. Pennycook, A. R. Lupini, H. Yang, *et al.*, "Efficient phase contrast imaging in stem using a pixelated detector. part 1: Experimental demonstration at atomic resolution," Ultramicroscopy **151**, 160–167 (2015).
8. J. M. Rodenburg and H. M. L. Faulkner, "A phase retrieval algorithm for shifting illumination," Appl. Phys. Lett. **85**(20), 4795–4797 (2004).
9. M. Guizar-Sicairos and J. R. Fienup, "Phase retrieval with transverse translation diversity: a nonlinear optimization approach," Opt. Express **16**(10), 7264–7278 (2008).
10. P. Thibault, M. Dierolf, A. Menzel, *et al.*, "High-resolution scanning x-ray diffraction microscopy," Science **321**(5887), 379–382 (2008).
11. V. Elser, I. Rankenburg, and P. Thibault, "Searching with iterated maps," Proc. Natl. Acad. Sci. **104**(2), 418–423 (2007).
12. A. M. Maiden and J. M. Rodenburg, "An improved ptychographical phase retrieval algorithm for diffractive imaging," Ultramicroscopy **109**(10), 1256–1262 (2009).
13. S. Marchesini, Y. Tu, and H. Wu, "Alternating projection, ptychographic imaging and phase synchronization," Appl. Comput. Harmon. Analysis **41**(3), 815–851 (2016).
14. S. Marchesini, H. Krishnan, B. J. Daurer, *et al.*, "Sharp: a distributed gpu-based ptychographic solver," J. Appl. Crystallogr. **49**(4), 1245–1252 (2016).
15. D. R. Luke, "Relaxed averaged alternating reflections for diffraction imaging," Inverse Probl. **21**(1), 37–50 (2005).
16. H. Chang, P. Enfedaque, and S. Marchesini, "Blind ptychographic phase retrieval via convergent alternating direction method of multipliers," SIAM J. Imaging Sci. **12**(1), 153–185 (2019).
17. H. Yan, "Ptychographic phase retrieval by proximal algorithms," New J. Phys. **22**(2), 023035 (2020).
18. R. Hesse, D. R. Luke, S. Sabach, *et al.*, "Proximal heterogeneous block implicit-explicit method and application to blind ptychographic diffraction imaging," SIAM J. Imaging Sci. **8**(1), 426–457 (2015).
19. P. Thibault and M. Guizar-Sicairos, "Maximum-likelihood refinement for coherent diffractive imaging," New J. Phys. **14**(6), 063004 (2012).
20. R. Horstmeyer, R. Y. Chen, X. Ou, *et al.*, "Solving ptychography with a convex relaxation," New J. Phys. **17**(5), 053044 (2015).
21. E. Bostan, M. Soltanolkotabi, D. Ren, *et al.*, "Accelerated wirtinger flow for multiplexed fourier ptychographic microscopy," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, (IEEE, 2018), pp. 3823–3827.
22. X. Chang, S. Jiang, G. Zheng, *et al.*, "Deep distributed optimization for blind diffuser-modulation ptychography," Opt. Lett. **47**(12), 3015–3018 (2022).
23. R. Wu, Z. Luo, M. Liu, *et al.*, "Fast fourier ptychographic quantitative phase microscopy for in vitro label-free imaging," Biomed. Opt. Express **15**(1), 95–113 (2024).
24. X. Chang, L. Bian, and J. Zhang, "Large-scale phase retrieval," eLight **1**(1), 4 (2021).
25. S. Jiang, K. Guo, J. Liao, *et al.*, "Solving fourier ptychographic imaging problems via neural network modeling and tensorflow," Biomed. Opt. Express **9**(7), 3306–3319 (2018).
26. J. Seifert, D. Bouchet, L. Loetgering, *et al.*, "Efficient and flexible approach to ptychography using an optimization framework based on automatic differentiation," OSA Continuum **4**(1), 121–128 (2021).
27. D. J. Chang, C. M. O'Leary, C. Su, *et al.*, "Deep-learning electron diffractive imaging," Phys. Rev. Lett. **130**(1), 016101 (2023).
28. M. Odstrcil, A. Menzel, and M. Guizar-Sicairos, "Iterative least-squares solver for generalized maximum-likelihood ptychography," Opt. Express **26**(3), 3108–3123 (2018).

Optics EXPRESS

29. Z. Chen, M. Odstrcil, Y. Jiang, *et al.*, "Mixed-state electron ptychography enables sub-angstrom resolution imaging with picometer precision at low dose," Nat. Commun. **11**(1), 1–10 (2020).
30. A. M. Maiden, "Ptychography algorithms from Sheffield University," 1.0, Github (2024). https://github.com/andyMaiden/SheffieldPtycho
31. B. Enders and P. Thibault, "A computational framework for ptychographic reconstructions," Proc. R. Soc. A **472**(2196), 20160640 (2016).
32. E. H. R. Tsai, I. Usov, A. Diaz, *et al.*, "X-ray ptychography with extended depth of field," Opt. Express **24**(25), 29089–29108 (2016).
33. M. Stockmar, P. Cloetens, I. Zanette, *et al.*, "Near-field ptychography: phase retrieval for inline holography using a structured illumination," Sci. Rep. **3**(1), 1927 (2013).
34. A. M. Maiden, M. J. Humphry, F. Zhang, *et al.*, "Superresolution imaging via ptychography," J. Opt. Soc. Am. A **28**(4), 604–612 (2011).
35. A. Konijnenberg, W. Coene, and H. Urbach, "Model-independent noise-robust extension of ptychography," Opt. Express **26**(5), 5857–5874 (2018).
36. 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**(8), 1606–1614 (2012).
37. P. Thibault and A. Menzel, "Reconstructing state mixtures from diffraction measurements," Nature **494**(7435), 68–71 (2013).
38. S. Marchesini, A. Schirotzek, C. Yang, *et al.*, "Augmented projections for ptychographic imaging," Inverse Probl. **29**(11), 115009 (2013).
39. X. Yu, V. Nikitin, D. J. Ching, *et al.*, "Scalable and accurate multi-gpu-based image reconstruction of large-scale ptychography data," Sci. Rep. **12**(1), 5334 (2022).
40. R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of the phase from image and diffraction plane pictures," Optik **35**(2), 237–246 (1972).
41. J. R. Fienup, "Phase retrieval algorithms: a comparison," Appl. Opt. **21**(15), 2758–2769 (1982).
42. P. D. Baksh, M. Ostrcil, M. Miszczak, *et al.*, "Quantitative and correlative extreme ultraviolet coherent imaging of mouse hippocampal neurons at high resolution," Sci. Adv. **6**(18), eaaz3025 (2020).
43. D. F. Gardner, M. Tanksalvala, E. R. Shanblatt, *et al.*, "Subwavelength coherent imaging of periodic samples using a 13.5nm tabletop high-harmonic light source," Nat. Photonics **11**(4), 259–263 (2017).
44. A. M. Maiden, D. Johnson, and P. Li, "Further improvements to the ptychographical iterative engine," Optica **4**(7), 736–745 (2017).
45. P. Godard, M. Allain, V. Chamard, *et al.*, "Noise models for low counting rate coherent diffraction imaging," Opt. Express **20**(23), 25914–25934 (2012).
46. A. M. Maiden, J. M. Rodenburg, and M. J. Humphry, "Optical ptychography: a practical implementation with useful resolution," Opt. Lett. **35**(15), 2585–2587 (2010).
47. Z. Hu, Y. Zhang, P. Li, *et al.*, "Near-field multi-slice ptychography: quantitative phase imaging of optically thick samples with visible light and x-rays," Opt. Express **31**(10), 15791–15809 (2023).
48. S. McDermott and A. Maiden, "Near-field ptychographic microscope for quantitative phase imaging," Opt. Express **26**(19), 25471–25480 (2018).
49. N. Parikh and S. Boyd, "Proximal algorithms," Foundations Trends Optim. **1**(3), 127–239 (2014).