Adaptive step-size strategy for noise-robust Fourier ptychographic microscopy

CHAO ZUO,[1,2,*] JIASONG SUN,[1,2] AND QIAN CHEN[2,3]

[1]*Smart Computational Imaging (SCI) Laboratory, Nanjing University of Science and Technology, Nanjing, Jiangsu Province 210094, China*
[2]*Jiangsu Key Laboratory of Spectral Imaging and Intelligent Sense, Nanjing University of Science and Technology, Nanjing, Jiangsu Province 210094, China*
[3]*chenq@njust.edu.cn*
[*]*surpasszuo@163.com*

Abstract: The incremental gradient approaches, such as PIE and ePIE, are widely used in the field of ptychographic imaging due to their great flexibility and computational efficiency. Nevertheless, their stability and reconstruction quality may be significantly degraded when non-negligible noise is present in the image. Though this problem is often attributed to the non-convex nature of phase retrieval, we found the reason for this is more closely related to the choice of the step-size, which needs to be gradually diminishing for convergence even in the convex case. To this end, we introduce an adaptive step-size strategy that decreases the step-size whenever sufficient progress is not made. The synthetic and real experiments on Fourier ptychographic microscopy show that the adaptive step-size strategy significantly improves the stability and robustness of the reconstruction towards noise yet retains the fast initial convergence speed of PIE and ePIE. More importantly, the proposed approach is simple, nonparametric, and does not require any preknowledge about the noise statistics. The great performance and limited computational complexity make it a very attractive and promising technique for robust Fourier ptychographic microscopy under noisy conditions.

References and links

1. H. M. L. Faulkner and J. M. Rodenburg, "Movable aperture lensless transmission microscopy: A novel phase retrieval algorithm," Phys. Rev. Lett. **93**, 023903 (2004).
2. J. M. Rodenburg and H. M. L. Faulkner, "A phase retrieval algorithm for shifting illumination," Appl. Phy. Lett. **85**, 4795–4797 (2004).
3. G. Zheng, R. Horstmeyer, and C. Yang, "Wide-field, high-resolution fourier ptychographic microscopy," Nat. Photon. **7**, 739–745 (2013).
4. R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of the phase from image and diffraction plane pictures," Optik **35**, 237–249 (1972).
5. J. R. Fienup, "Phase retrieval algorithms: a comparison," Appl. Opt. **21**, 2758–2769 (1982).
6. V. Mico, Z. Zalevsky, P. García-Martínez, and J. García, "Synthetic aperture superresolution with multiple off-axis holograms," J. Opt. Soc. Am. A **23**, 3162–3170 (2006).
7. C. Yuan, H. Zhai, and H. Liu, "Angular multiplexing in pulsed digital holography for aperture synthesis," Opt. Lett. **33**, 2356–2358 (2008).
8. P. Gao, G. Pedrini, and W. Osten, "Structured illumination for resolution enhancement and autofocusing in digital holographic microscopy," Opt. Lett. **38**, 1328–1330 (2013).
9. A. M. Maiden and J. M. Rodenburg, "An improved ptychographical phase retrieval algorithm for diffractive imaging," Ultramicroscopy **109**, 1256 – 1262 (2009).
10. M. Guizar-Sicairos and J. R. Fienup, "Phase retrieval with transverse translation diversity: a nonlinear optimization approach," Opt. Express **16**, 7264–7278 (2008).
11. P. Thibault, M. Dierolf, O. Bunk, A. Menzel, and F. Pfeiffer, "Probe retrieval in ptychographic coherent diffractive imaging," Ultramicroscopy **109**, 338–343 (2009).
12. X. Ou, G. Zheng, and C. Yang, "Embedded pupil function recovery for fourier ptychographic microscopy," Opt. Express **22**, 4960–4972 (2014).
13. 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).

14. 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).

15. 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).

16. J. Sun, Q. Chen, Y. Zhang, and C. Zuo, "Efficient positional misalignment correction method for fourier ptychographic microscopy," Biomed. Opt. Express **7**, 1336–1350 (2016).

17. P. Thibault and A. Menzel, "Reconstructing state mixtures from diffraction measurements," Nature **494**, 68–71 (2013). 10.1038/nature11806.

18. D. J. Batey, D. Claus, and J. M. Rodenburg, "Information multiplexing in ptychography," Ultramicroscopy **138**, 13 – 21 (2014).

19. L. Tian, X. Li, K. Ramchandran, and L. Waller, "Multiplexed coded illumination for fourier ptychography with an led array microscope," Biomed. Opt. Express **5**, 2376–2389 (2014).

20. S. Dong, R. Shiradkar, P. Nanda, and G. Zheng, "Spectral multiplexing and coherent-state decomposition in fourier ptychographic imaging," Biomed. Opt. Express **5**, 1757–1767 (2014).

21. J. Sun, Q. Chen, Y. Zhang, and C. Zuo, "Sampling criteria for Fourier ptychographic microscopy in object space and frequency space," Opt. Express **24**, 15765–15781 (2016)

22. C. Yang, J. Qian, A. Schirotzek, F. Maia, and S. Marchesini, "Iterative algorithms for ptychographic phase retrieval," arXiv preprint arXiv:1105.5628 (2011).

23. R. Horstmeyer, R. Y. Chen, X. Ou, B. Ames, J. A. Tropp, and C. Yang, "Solving ptychography with a convex relaxation," New Journal of Physics **17**, 053044 (2015).

24. 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).

25. L. Tian and L. Waller, "3D intensity and phase imaging from light field measurements in an led array microscope," Optica **2**, 104–111 (2015).

26. 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).

27. D. R. Luke, "Relaxed averaged alternating reflections for diffraction imaging," Inverse Probl. **21**, 37–50 (2005).

28. S. Marchesini, H. He, H. N. Chapman, S. P. Hau-Riege, A. Noy, M. R. Howells, U. Weierstall, and J. C. H. Spence, "X-ray image reconstruction from a diffraction pattern alone," Phys. Rev. B **68**, 140101(R) (2003).

29. H. H. Bauschke, P. L. Combettes, and D. R. Luke, "Phase retrieval, error reduction algorithm, and Fienup variants: a view from convex optimization," J. Opt. Soc. Am. A **19**, 1334–1345 (2002)

30. P. S. Diniz, *Adaptive filtering*" (Springer, 1997).

31. P. Godard, M. Allain, V. Chamard, and J. Rodenburg, "Noise models for low counting rate coherent diffraction imaging," Opt. Express **20**, 25914–25934 (2012).

32. P. Thibault and M. Guizar-Sicairos, "Maximum-likelihood refinement for coherent diffractive imaging," New Journal of Physics **14**, 063004 (2012).

33. J. Rodenburg, A. Hurst, and A. Cullis, "Transmission microscopy without lenses for objects of unlimited size," Ultramicroscopy **107**, 227 – 231 (2007).

34. D. P. Bertsekas and D. P. Bertsekas, *Nonlinear Programming* (Athena Scientific, 1999), 2nd ed.

35. D. P. Bertsekas, "Incremental gradient, subgradient, and proximal methods for convex optimization: A survey," Optimization for Machine Learning **2010**, 1–38 (2011).

36. K. Guo, S. Dong, P. Nanda, and G. Zheng, "Optimization of sampling pattern and the design of Fourier ptychographic illuminator," Opt. Express **23**, 6171–6180 (2015).

37. X. Ou, R. Horstmeyer, G. Zheng, C. Yang, "High numerical aperture Fourier ptychography: principle, implementation and characterization," Opt. Express **23**, 3472–3491 (2015).

38. http://www.scilaboratory.com/h-col-123.html

39. L. Tian, Z. Liu, L. Yeh, M. Chen, J. Zhong, and L. Waller, "Computational illumination for high-speed in vitro Fourier ptychographic microscopy," Optica **2**, 904–911 (2016)

40. T. M. Godden, A. Muñiz-Piniella, J. D. Claverley, A. Yacoot, and M. J. Humphry, "Phase calibration target for quantitative phase imaging with ptychography," Opt. Express **24**, 7679–7692 (2016)

41. A. Nedic and D. P. Bertsekas, "Incremental subgradient methods for nondifferentiable optimization," SIAM J. Optim. **12**, 109–138 (2001).

42. R. Correa and C. Lemaréchal, "Convergence of some algorithms for convex minimization," Math. Program. **62**, 261–275 (1993).

1. Introduction

Ptychography is a computational imaging approach in which a limited-size illumination probe is moved sequentially across an extended object, while the resulting diffraction patterns are collected [1, 2]. With sufficient overlap between these illumination spots, the acquired images can be processed into an estimate of the object's complex function (i.e., its amplitude and phase) with

high sensitivity and a larger field of view than a single recorded diffraction pattern. Recently, the counterpart of (conventional) ptychography in the Fourier space, termed Fourier ptychography microscopy (FPM) has been introduced [3]. FPM shares its root with non-interferometric phase retrieval [4, 5] and interferometric synthetic aperture imaging [6–8]. Instead of spatially scanning the object with a narrow illumination beam, it uses variable-angle illuminations provided by a light-emitting diode (LED) array and does not involve any moving parts. The captured low-resolution images from different illumination angles can be synthesized in the Fourier space to surpass the diffraction limit of the objective lens and thus boost the space-bandwidth product of the microscope. Major advantages of the ptychographic approaches are linked to the redundancy of the acquired data: many inaccurately known parameters and that can be retrieved simultaneously with the object complex function, such as the probe (pupil) function [9–12], probe (LED) positions [13–16], and mixed coherent states of the incoming illuminations [17–21].

By adopting the concepts in phase retrieval [4,5], ptychography uses an iterative reconstruction process to recover the 'lost' phase information of the sample by updating functions back and forth between the real and Fourier spaces. The original solutions to ptychography used Gerchberg-Saxton [1, 4], or incremental gradient descent based techniques (termed PIE and ePIE) [2, 9], which belong to a type of alternating projections (AP). Ideally, the successive iterations should eventually reach a solution that the constraints resulting from the overlapping condition and the intensity measurements can be satisfied simultaneously. However, in the presence of non-negligible noise, such a reasonable solution cannot be achieved as different intensity patterns are not anymore consistent one with another, leading to degraded reconstruction quality. The situation is even worse for FPM because the experimental data contains mainly darkfield images from high-angle illuminations that larger than the objective numerical aperture allows. The high-angle darkfield illuminations usually do not produce images with sufficient intensity, leading to a low signal-to-noise ratio (SNR). The poor performance of AP strategies has often been attributed to the non-convex nature of phase recovery problem since few provable global convergence guarantees exists even under ideal (noise-free) conditions [5, 22–24]. Multiple algorithms employing the difference map [11], nonlinear optimization [10], Gauss-Newton search [15, 22], Wirtinger flows [24], or convex lifting [23] have been proposed to either achieve better convergence properties, improve the robustness towards noise, or both. However, those methods generally reside on expensive processing requirements, making them less appealing from a computational point of view.

Generally, the above mentioned solutions for solving the ptychographic phase retrieval problem can be classified into two categories: the incremental (sequential) algorithms (like the PIE and ePIE, which only update a small patch of the object function in each iteration) and the global algorithms (like the difference map, nonlinear optimization, Wirtinger flow etc., which attempt to update the entire object function in one iteration). The global algorithms have been found to be connected to the gradient descent methods (also known as steepest descent), and for which the convergence behavior (especially for convex sets) has been well studied. Various more advanced strategies in the optimization community, like the conjugate gradient, high-order derivative (Gauss-Newton), and line search have been introduced to improve their performance, which has been well summarized in [22] for ptychography and [15] for FPM. On the other hand, to the best of our knowledge, the incremental algorithms (PIE and ePIE) are more favored and widely used for their fast convergence, computational efficiency and low memory cost, especially for dealing with large dataset. Though their susceptibility to noise is well-noticed, the incremental approaches offer great flexibilities to be adapted to more complicate mathematical models of both object and imaging system for many advanced applications where great solution accuracy is not of paramount importance [20, 21, 25, 26]. By contrast, very few general results are known for convergence behavior of incremental approaches, and the gathered knowledge is problem-specific and often of an empirical nature.

In this paper, we focus on the convergence behavior of incremental ptychographic solutions and address precisely the problem of their performance degradation under noisy conditions. After reviewing the state-of-the-art methods for solving ptychographic phase retrieval problem from a numerical analysis point of view, we found that the PIE and ePIE algorithms can be viewed as *incremental gradient descent approaches* in the numerical optimization community. The difference between the incremental gradient descent approaches and the more well-known (global) gradient descent approaches is that the former use an "approximate gradient" computed from only a part of the data. Despite faster initial convergence rate, incremental gradient approaches are typically more sensitive to noise and frequently cannot converge to a reasonable solution. We prove that this is not only because they are first order gradient methods or the phase retrieval problem is non-convex, but more importantly, because they require *a diminishing step-size* for convergence even in the convex case (which is not so for global gradient methods). This important property remained unnoticed in the ptychography and FPM literature, where the step-size is often taken to be a large constant (e.g. 1) [2, 9, 12, 13]. The necessity of a diminishing step-size for convergence further motivates the adaptive step-size strategy we presented here. By simply introducing a variable step-size instead of a fixed one, the convergence behavior of the algorithm can be improved, and meanwhile, its tolerance to noise can be significantly enhanced. Furthermore, as a direct extension of the original PIE and ePIE, the proposed scheme is computationally simple and extremely easy to use. Numerical simulations and real experiments in the context of FPM are carried out to confirm and validate the performance of the proposed approach. It should also be emphasized that while we focus on specifically the FPM in this work, our approach can be directly extended to conventional ptychography via variable substitution.

There are two important issues about the proposed approach need to be clarified. First, we have recognized that the idea of adaptive step-size (sometimes is referred to 'feedback parameter' or 'relaxation parameter') is well known in the optimization and signal processing community. Even in the field of coherent diffractive imaging, many iterative phase retrieval algorithms were greatly improved through the introduction of adaptive feedback. For example, the Relaxed Averaged Alternating Reflection (RAAR) algorithm uses a relaxation parameter for steering the iteration and improving the phase retrieval quality [27]. In shrink-wrap algorithm, a feedback parameter is introduced in real space to dynamically tighten the support constraint [28]. And finally, a feedback parameter is used in ptychographic to control the refinement rate for solving position uncertainties [14]. However, in this work, we propose and analyze for the first time the adaptive step-size strategy for incremental solutions of ptychographic imaging. A convergence analysis of incremental gradient descent approaches for convex sets is also provided, which provides guidelines for the choice of step-size and makes our adaptive step-size rule mathematically tractable. Second, the convergence analysis of incremental gradient descent approaches described in this work is based on the framework of convex optimization, while the phase retrieval problem is inherently non-convex. For non-convex sets, the convergence behavior of incremental gradient descent approaches is much more difficult to predict. However, reformulating the phase retrieval and ptychography in the context of convex optimization does provide a powerful theoretical framework to gain insight and, potentially, to improve existing algorithms based on a sound mathematical theory [22, 27, 29]. Besides, in both ptychography and FPM, the measurement space modulus constraints from overlapping scan positions appear to be stringent enough to cause the optimization problem to be *almost convex*. As verified through extensive numerical simulations and experiments, the proposed adaptive step-size rule, along with those convergence properties described in this paper does not only has theoretical significance, but also has been proved to be pragmatic.

The remainder of this paper is organized as follows: In Section 2, we formulate the ptychographic imaging as an unconstrained nonlinear minimization problem. Some of the state-of-the-art solutions for ptychographic imaging are reviewed in Section 3, and they can be classified into

two categories: the global gradient approaches and incremental gradient solutions. By analyzing the convergence behaviors of incremental gradient solutions, an adaptive step-size strategy is introduced in Section 4 to improve the stability and robustness of the reconstruction under noisy conditions. Simulations and experimental results are presented in Section 5 and Section 6, respectively, and this paper ends with a discussion and the conclusions in Section 7.

2. Problem formulation

The experimental configuration and data acquisition process for FPM measurements can be found in the literature [3, 19, 23] and will not be detailed here. In FPM, the problem to be solved is to recover the high-resolution complex field (including both amplitude and phase) of the object from a number of low-resolution intensity measurements with different illumination angles. For a finite set of illumination vectors $\mathbf{u}_i = \left(u_x, u_y\right)$, $i = 1, 2, ..., N$, we denote each intensity image by:

$$I_i\left(\mathbf{r}\right) = \left|\mathscr{F}^{-1}\left\{P\left(\mathbf{u}\right)O\left(\mathbf{u} - \mathbf{u}_i\right)\right\}\right|^2, i = 1, 2, ..., N \tag{1}$$

where $\mathbf{r}$ is real-space coordinate vector and $\mathscr{F}$ is the 2D Fourier transform. The Fourier transform of the object function $O\left(\mathbf{u}\right)$, is firstly shifted by the illumination vector $\mathbf{u}_i$, then confined by the pupil function $P\left(\mathbf{u}\right)$, and finally inverse Fourier transformed back to the real-space to form a low-resolution image $I_i\left(\mathbf{r}\right)$ captured by the camera. To solve $O\left(\mathbf{u}\right)$, most of current algorithms can be derived by formulating FPM as the following optimization problem [15, 22]:

$$\min_{O(\mathbf{u})} \varepsilon = \min_{O(\mathbf{u})} \sum_i \sum_{\mathbf{r}} \left|\sqrt{I_i\left(\mathbf{r}\right)} - \left|\mathscr{F}^{-1}\left\{P\left(\mathbf{u}\right)O\left(\mathbf{u} - \mathbf{u}_i\right)\right\}\right|\right|^2 \tag{2}$$

In the following, the vectorial notion will be adopted as it corresponding to an appropriate discretization of the continuous problem. The image and the object function are raster-scanned into vectors $\mathbf{I}_i = \{I_{m,i}\}_{m=1}^{M}$ (with M pixels) and $\mathbf{O} = \{O_m\}_{m=1}^{L}$ (with L pixels). $\mathbf{P}_i$ is an $M \times L$ matrix determined by the pupil function $P\left(\mathbf{u}\right)$ that extracts a sub-spectrum containing M pixels out of the whole L pixels. We thus have the vectorized cost function:

$$\varepsilon = \sum_i \left\|\sqrt{\mathbf{I}_i} - \left|\mathbf{F}^{-1}\mathbf{P}_i\mathbf{O}\right|\right\|^2 \equiv \sum_i \left\|\sqrt{\mathbf{I}_i} - |\mathbf{g}_i|\right\|^2 \tag{3}$$

where $\|.\|$ is the Euclidean norm, $\mathbf{F}$ is the matrix representation of discrete Fourier transform, and $\mathbf{g}_i \equiv \mathbf{F}^{-1}\mathbf{P}_i\mathbf{O}$. The error metric given by Eq. (3) is also termed as the real-space error, which quantifies how close the current estimate fits the input data as a whole. Since the estimated high-resolution object complex field will be of higher space-bandwidth product than the raw image, we have $L > M$.

3. Global and incremental gradient solutions of Fourier ptychographic imaging

Most of the state-of-the-art solutions for ptychographic imaging are closely related to standard gradient-based iterative optimization algorithms in the numerical optimization community. These iterative algorithms can be split into two families: the global gradient approaches and incremental gradient solutions. We explore these algorithms by applying them to some two dimensional optimization problems in order to gain some intuition on their characteristics and behaviors.

3.1. Global gradient based solutions of Fourier ptychographic imaging

To minimize Eq. (3) using iterative optimization techniques, we often need to evaluate the gradient of the cost function. To facilitate the gradient calculation, we reformulate the error metric of Eq. (3) in Fourier space based on Parseval's theorem [5]

$$\varepsilon = \sum_i \|\mathbf{\Psi}_i - \mathbf{P}_i\mathbf{O}\|^2 \tag{4}$$

with

$$\mathbf{\Psi}_i = \Pi_{m \to \sqrt{\mathbf{I}_i}} (\mathbf{P}_i \mathbf{O}) \tag{5}$$

where the $\mathbf{\Psi}_i$ is updated sub-spectrum; $\Pi_{m \to \sqrt{\mathbf{I}}}$ operator is known as the real-space modulus projection ($\Pi_{m \to \sqrt{\mathbf{I}_i}} (\mathbf{P}_i \mathbf{O}) = \mathbf{F}\mathrm{Diag}\left(\mathbf{g}_i|\mathbf{g}_i|^{-1}\right) \sqrt{\mathbf{I}_i}$), which simply replaces the amplitude of $\mathbf{g}_i$ with the measured $\sqrt{\mathbf{I}_i}$ while conserving the phase. The gradient of ε with respect to $\mathbf{O}$ can be expressed as

$$\nabla \varepsilon = -\sum_i \left[\mathbf{P}_i^* (\mathbf{\Psi}_i - \mathbf{P}_i \mathbf{O}) \right] \tag{6}$$

To minimize the error of the metric given in Eq. (4), one can use fixed-point iteration by setting the gradient to zero ($\nabla \varepsilon = 0$), as in the difference map approach [11]), or alternatively, use gradient-based search algorithms, like the gradient descent [22], conjugate gradient [10], and Gauss-Newton methods [15, 22]. For the simplest gradient descent method, the object function is updated based on the following equation

$$\mathbf{O}^{k+1} = \mathbf{O}^k + \alpha \mathbf{W} \nabla \varepsilon(\mathbf{O}^k) = \mathbf{O}_i - \alpha \mathbf{W} \sum_i \left[\mathbf{P}_i^* \left(\mathbf{\Psi}_i^k - \mathbf{P}_i \mathbf{O}^k \right) \right] \tag{7}$$

where

$$\mathbf{W} = \sum_i \left(\mathbf{P}_i^* \mathbf{P}_i \right)^{-1} \tag{8}$$

can be interpreted as a preconditioner [22], or a normalization matrix [30], k is the iteration number, and α is the step-size. It should also be noticed that other than Eq. (3), there also exists other forms of cost function that evaluate the difference of the estimated and measured intensity function, or explicitly take the statistics of the noise into account [15, 31, 32].

3.2. Incremental gradient solutions of Fourier ptychographic imaging

In the above mentioned algorithms, the object function is updated globally with use of the full set of images for each iteration. However, when the number of intensity images is very large, those global methods may be ineffective because the size of the dataset makes each iteration very costly. Therefore, the incremental approaches [2, 9, 12, 19], which use only one intensity measurement for each update instead of the total data, may be more attractive. The incremental approach only operates on a single component of the cost function at each iteration

$$\varepsilon_i = \|\mathbf{\Psi}_i - \mathbf{P}_i \mathbf{O}\|^2 \tag{9}$$

The corresponding gradient can be expressed as

$$\nabla \varepsilon_i = -\mathbf{P}_i^* (\mathbf{\Psi}_i - \mathbf{P}_i \mathbf{O}) \tag{10}$$

Starting from an initial guess for $\mathbf{O}$, the optimization cycles through each intensity measurement $\mathbf{I}_i$ in sequence, based on the following gradient-based search algorithm

$$\mathbf{O}_{i+1}^k = \mathbf{O}_i^k + \alpha^k \mathbf{W}_i \nabla \varepsilon_i(\mathbf{O}_i^k) = \mathbf{O}_i^k - \alpha^k \mathbf{W}_i \mathbf{P}_i^* \left(\mathbf{\Psi}_i^k - \mathbf{P}_i \mathbf{O}_i^k \right) \tag{11}$$

where

$$\mathbf{W}_i = \mathrm{Diag}\left(\beta_{m,i} \left(\left| P_{m,i} \right|^2 + \gamma \right)^{-1} \right) \tag{12}$$

and α^k is the step-size, which is often set to $\alpha^k = \alpha \equiv 1$ in most literature [2, 9, 12, 13]. Note that one full iteration (so called one cycle, $k \leftarrow k + 1$) comprises N sub-iterations running over the whole dataset. A small constant γ is included in Eq. (12) to avoid divide-by-zero problem

when $\left|P_{m,i}\right|^2$ is too small. $\beta_{m,i}$ is a scaling factor that weights the well-transferred spectrum component while rejects the unreliable data falling outside the pupil

$$\beta_{m,i} = \begin{cases} \left|P_{m,i}\right| / \left|P_{m,i}\right|_{\max} & PIE-type \\ \left|P_{m,i}\right|^2 / \left|P_{m,i}\right|_{\max}^2 & ePIE-type \end{cases} \tag{13}$$

Besides the reduced computational cost, the incremental approach may attain a faster convergence rate than global ones, especially for a large dataset (an 'acceptable' estimate usually can be achieved after a single cycle [33]). However, it has been found that the convergence and reconstruction quality of the incremental approaches are quite susceptible to noise. This problem has often been considered to be the consequences of the non-convexity of the modulus projection [5,23,24]. But on the one hand, it is also well recognized that the real-space constraints from overlapping pupils are stringent enough to allow us to find not only the correct object function, but also the pupil aberration, LED misalignment, and beyond [12,16,19,20,25]. Thus, the reason behind the poor performance of incremental approaches with noisy data is still unclear.

4. Adaptive step-size scheme for incremental gradient solutions

Though the incremental gradient method is closely related to the conventional gradient descent method [which minimizes Eq. (4) instead of sequentially minimizing Eq. (9)], their convergence behaviors are quite different. For conventional gradient descent algorithm, it can be proven that the algorithm can finally converge to at least a local optimum of ε with a linear rate if the step-size α is less than 2 (in the convex setting, all local minima are also global minima, so in this case it can converge to the global optimum) [30,34]. However, for incremental approaches, we usually observe that it can find a relatively correct solution in early iterations, but then start to exhibit limit-cycle like behavior. The oscillatory behavior is even severer for noisy data since the intensity images are inconsistent with each other. To understand the convergence mechanism of incremental gradient methods, let us assume that the component functions ε_i are selected for iteration according to a cyclic order [i.e. $i \leftarrow (i \bmod N) + 1$] and α^k is constant within a cycle (i.e. for all $k = 0, 1, ..., \alpha_1^k = \alpha_2^k = ... = \alpha_N^k$). For each iteration, the incremental methods use an approximate gradient computed from only one component of the entire data. Thus, one incremental iteration ($i \leftarrow i + 1$) tends to make reasonable progress only in the "average" sense (viewed from one whole cycle instead of single sub-iteration). Viewing the iteration Eq. (11) in terms of cycles, we have for every k that marks the beginning of a cycle ($\mathbf{O}^k = \mathbf{O}_1^k = \mathbf{O}_{N+1}^{k-1}$)

$$\mathbf{O}^{k+1} = \mathbf{O}^k + \alpha^k \sum_{i=1}^{N} \mathbf{W}_i \nabla \varepsilon_i (\mathbf{O}_i^k) = \mathbf{O}^k + \alpha^k \mathbf{W} \left(\nabla \varepsilon (\mathbf{O}^k) + e^k \right), \tag{14}$$

where $\varepsilon = \sum_{i=1}^{N} \varepsilon_i$ is the cost function, or sum of the components given by Eq. (4), and e^k is

$$e^k = \sum_{i=1}^{N} \left(\nabla \varepsilon_i (\mathbf{O}^k) - \nabla \varepsilon_i (\mathbf{O}_i^k) \right), \tag{15}$$

which may be viewed as an error in the calculation of the (global) gradient $\nabla \varepsilon (\mathbf{O}^k)$. For Lipschitz continuous gradient functions $\nabla \varepsilon_i$, the error e^k is proportional to the step-size α^k [34,35], and this shows two fundamental convergence properties of incremental gradient approaches:

Property A: A constant step-size (e.g. $\alpha = 1$ as in most literature) typically cannot guarantee convergence, since then the size of the gradient error $\left\|e^k\right\|$ is typically bounded away from 0. Instead, a peculiar form of convergence takes place for constant but sufficiently small α (see Lemma 1 of Appendix A), whereby the iterates within cycles converge but to different points within a sequence of N points (i.e., the sequence of first points in the cycles converges to a different limit than the sequence of second points in the cycles, etc).

Property B: A diminishing step-size leads to diminishing error e^k, so it can result in convergence to a stationary point of ε. The basic idea is that with a constant step-size α we can get to within an $O(\alpha)$-neighborhood of the optimum (see Lemma 1 and Lemma 2 of Appendix A), so with a diminishing step-size α^k, we should be able to reach an arbitrarily small neighborhood of the optimum. Appendix A presents a convergence proof of the incremental gradient solution Eq. (11) with a diminishing step-size in the convex case. The algorithm can finally converge to a minimizer if $\lim_{k \to \infty} \alpha^k = 0$, $\sum_{k=0}^{\infty} \alpha^k = \infty$, and $\sum_{k=0}^{\infty} \left(\alpha^k \right)^2 < \infty$.

Note that the convergence properties described above (and the accompanying convergence proof in Appendix A) are restricted to certain assumptions on the function ε (ε and its each component ε_i are convex, $\nabla\varepsilon$ and $\nabla\varepsilon_i$ are Lipschitz continuous). For non-convex problems, gradient based approaches can only converge to a local minimum. Although the convex optimization theory cannot completely analyze the ptychographic phase retrieval problem, which is non-convex, it is in general much simpler and can indeed predict several important properties of the algorithm.

The two convergence properties reveal that the choice of the step-size α plays an important role for the success of incremental gradient methods, which is a key point that is often overlooked when using PIE and ePIE-type solutions for FPM reconstruction. According to Property A, though the routinely used unity step-size can achieve a much faster convergence rate than its global counterpart at early stage (when far from the solution, the evaluation of a single gradient component is sufficient for generating an approximate gradient direction [34]), it typically makes the algorithm strongly oscillatory in the neighborhood of a solution and thus may not converge to a stationary point (when getting close to the solution, a single component gradient can no longer well approximate the global gradient). In fact, the step-size α controls the amount of feedback from the previous estimate of the algorithm. When $\alpha = 1$, the incremental gradient method is actually identical to the alternating projection algorithm with multiple constraint sets defined by each sub-aperture image. In such case, at one given pupil position, the estimate is updated fully by the measured intensity at that illumination angle. However, for FPM, in most cases the object is under dark-field illuminations, the low-level of irradiance produces individual measurements that are heavily affected by noise. In such condition, one single image should not fully constrain the estimate at a given pupil position. When a unity step-size is used, it can be observed that the algorithm can quickly find a relatively correct (but noisy) estimate, but start to loop around after some iterations because the intensity dateset is inconsistent, as a consequence of noise [31]. At one given pupil position, the algorithm tends to 'undo' the progress made at the previous update, reintroducing fully the noise within the associated intensity image. In contrast, the global gradient descent methods consider the complete cost function Eq. (4) and evaluate the gradient with use of the full set of images for each iteration. Despite much slower converge in early iterations and more involved computation, they can finally converge to a more reasonably balanced and globally consistent solution.

Property A also suggests using a sufficiently small step-size to stabilize the incremental algorithm and ensure convergence, as the size of the oscillation is roughly proportional to α. The relaxation in the projections effectively prevents any single component from altering the current estimate too far towards an incorrect result. From another angle to explain, using a small step-size less than one tends to make the incremental solution more 'global', resulting in better "subset gradient balance": it is possible to allow the information from the overlapping pupil positions to blend together to reinforce the true signal while reduce the influence of noise. It is also interesting to note that the parameter γ (sometimes called regularization parameter) in Eq. (12) can also be tuned to achieve the similar noise-reduction effect, as demonstrated in [15, 19] (if γ is large and dominates the denominator, we have $\gamma \approx 1/\alpha$). Even though, the question of determining a proper α throughout still remains: a large α is insufficient to stabilize the resulting asymptotic oscillation,

while a too small α may significantly slow down the convergence, which compares unfavorably even with the linear rate of convergence associated with its non-incremental counterpart.

Instead of using a constant step-size, referring to Property B, an adaptive step-size (sequence) α^k is introduced in this work to improve the performance of incremental ptychographic solutions. Since the sufficient conditions on step-size α^k for global convergence to a minimizer in the convex setting are (i) $\lim_{k \to \infty} \alpha^k = 0$, (ii) $\sum_{k=0}^{\infty} \alpha^k = \infty$, and (iii) $\sum_{k=0}^{\infty} \left(\alpha^k \right)^2 < \infty$, the critical issue in practical will be how to determine a suitable sequence α^k to get close to a solution within fewer iterations. Among the three conditions, (i) and (iii) require that the α^k should diminishes to zero, which are easy to satisfy. However, shrinking the step-size to zero can only ensure that the algorithm converges to a point (as $\mathbf{O}_{i+1}^k = \mathbf{O}_i^k + \alpha^k \mathbf{W}_i \nabla \varepsilon_i (\mathbf{O}_i^k)$ goes to $\mathbf{O}_{i+1}^k = \mathbf{O}_i^k$ in the limit $\alpha^k \to 0$), which does not mean the algorithm found the desired optimum point. It is obvious that if the step-size shrinks too fast, $\mathbf{O}$ may converge to a point that isn't a minimum, especially when the initial point is sufficiently far from it. Thus, the condition (ii) $\sum_{k=0}^{\infty} \alpha^k = \infty$ should further be satisfied to ensure α^k not being reduced too fast so that the algorithm can travel infinitely far if necessary. One simple and straightforward choice satisfying those conditions is:

$$\alpha^k = \begin{cases} \alpha^{k-1} & \left(\varepsilon(\mathbf{O}^k) - \varepsilon(\mathbf{O}^{k-1}) \right) \big/ \varepsilon(\mathbf{O}^{k-1}) > \eta \\ \alpha^{k-1} \big/ 2 & otherwise \end{cases} \tag{16}$$

where η is a small constant which should be much less than 1. By comparing the (global) error function for $\mathbf{O}^k$ and $\mathbf{O}^{k-1}$ obtained in consecutive cycles, we keep the step-size fixed as long as non-trivial progress is made (determined by η), while halve its value if otherwise. Starting from $\alpha^0 = 1$ as in PIE and ePIE, the algorithm can make large progress to get close to a solution at early iterations. Then the inconsistent information from adjacent pupil regions resulting from noise and model mismatch can be gradually reconciled as α^k decays. Finally, the algorithm will converge to a stationary point quickly when the step-size reach a pre-specified minimum (a constant that small enough so that the asymptotic oscillation is of no essential concern). The small parameter η defines the threshold of a non-trivial 'progress', which guarantees that the step-size reduces only when necessary [the algorithm gets near a (local) minimum and starts to oscillate]. We further found that the performance of the adaptive size strategy is not quite sensitive to the choice of η. A reasonably good result can always be obtained by fixing $\eta = 0.01$, which makes our approach nonparametric and very easy to use in practical applications.

It should be emphasized again that the convergence proof provided here is restricted to convex sets only, while the phase retrieval problem is non-convex. For non-convex optimization problems, the exact convergence properties of the incremental gradient solutions remain elusive. However, in a practical setting of FPM, the convergence behaviors derived from the convex optimization seem still applicable, and the adaptive step-size approach indeed provides significant improvements over the constant step-size on convergence rate and reconstruction accuracy. More importantly, those improvements *do not trade with the computational efficiency* of the incremental gradient approaches. Note that the error metric ε^k in Eq. (16) is often used for monitoring the progress or the convergence of standard PIE and ePIE algorithms, which needs to be calculated after each full sub-iteration of the algorithm [2, 9]. While in our approach, after each full sub-iteration of the algorithm, ε^k can be further used to determine the step-size for the next cycle. Therefore, our adaptive step-size scheme can be view as an elegant extension of the original PIE and ePIE *at no extra computational cost*. Moreover, the (global) error metric ε^k can be further approximated by the sum of the incremental error metric $\varepsilon^k \approx \sum_i \varepsilon_i^k$, which can be accumulated during the incremental iterative process. Through this way, the computational cost can be further reduced (at the cost of lower accuracy in the error estimates), making it more suitable for time-critical applications where the execution time of algorithm should be minimized.

5.　Simulations

We first validate the adaptive step-size scheme using simulations. The simulation parameters were chosen to realistically model an FPM platform, with an incident illumination wavelength of 626nm, an image sensor with pixel size of $6.5\mu m$, and an 4X objective with NA_{obj} of 0.1. We simulated the use of the central 15×15 LEDs in the array placed 87.5mm beneath the sample, and the distance between adjacent LEDs is 4mm. The raw low-resolution data is limited to a small region with only 64×64 pixel resolution, and the final high-resolution complex field with 256×256 pixel is recovered by different approaches. The true high-resolution complex function is created from the images shown in Figs. 1(a) and 1(b) for its intensity and phase. Besides the idealized situation, each low resolution image were corrupted with Gaussian noise with different variances. Since the signal power of darkfield ones images from high-angle illuminations is much weaker than that of the brightfield images, they suffer more from the noise, which is in accord with the actual experimental conditions. To illustrate the level of noise more intuitively, six typical noisy darkfield intensity images with a high illumination angle [labelled by the blue circle in Fig. 1(c)] are given in Fig. 1(d)-1(i). The noise level is quantified by the average mean absolute error (MAE), defined as AMAE = $\langle|\mathbf{I}_n - \mathbf{I}|\rangle/\langle\mathbf{I}\rangle$, where $\langle\mathbf{I}\rangle$ is the mean value of all noise-free darkfield intensity images and $\langle|\mathbf{I}_n - \mathbf{I}|\rangle$ is the averaged mean absolute error of the corresponding noisy images. The metric MAE is also used to quantify amplitude and phase reconstruction accuracy.
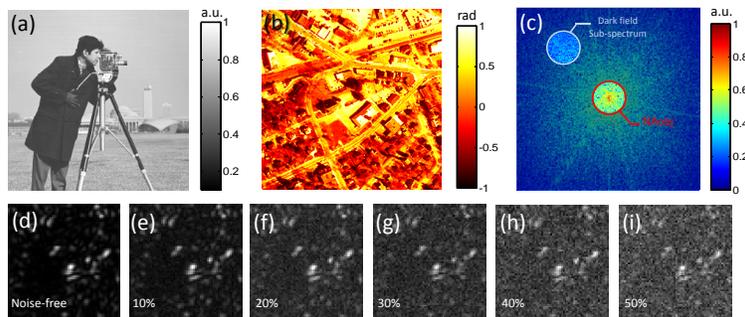
Fig. 1. Simulated amplitude and phase images for FPM reconstruction. (a) Amplitude; (b) Phase; (c) Fourier spectrum; the red and blue circles indicate the captured subspectrums under orthogonal and oblique illuminations; (d)-(i) low-resolution raw images with different levels of Gaussian noise.

There are two issues worth mentioning about the practical implementation of the FPM reconstruction algorithms. The first one is the initialization of the high-resolution complex field. Here, all of the algorithms use the same initial guess - the square root of the upsampled on-axis-illuminated intensity, which provides a good approximation to the high-resolution object function. We found that for incremental gradient approaches, the choice of initial value has little impact on the final reconstruction result. But a good initial value that closer to a solution does accelerate the convergence. Another one is the updating order of each sub-pupils in the Fourier domain. Three choices of updating sequence were suggested in [36]: (1) random order; (2) NA order (the sequence is ranked by the illumination NA); (3) energy order (the sequence is ranked by the total intensity value). In all simulations and experiments of this paper, the NA order was used, which starts at the center of the Fourier space and gradually expands outwards. It is also found that the NA order and the energy order produce almost the same results because the two sequences are quite similar because the energy of the sample in the Fourier domain typically decreases with the spatial frequency. Furthermore, the choice updating sequence does not fundamentally change the property and converge behavior of the incremental approach since

no specific requirements on the order of updating imposed in the converge analysis. But it is essential to include all components (images) in the fixed deterministic order. If the component images are selected in a randomized order, then the incremental gradient method turns into the stochastic gradient method and the convergence rate is found to be slower than the deterministic (NA or energy) orders, as demonstrated in [36].
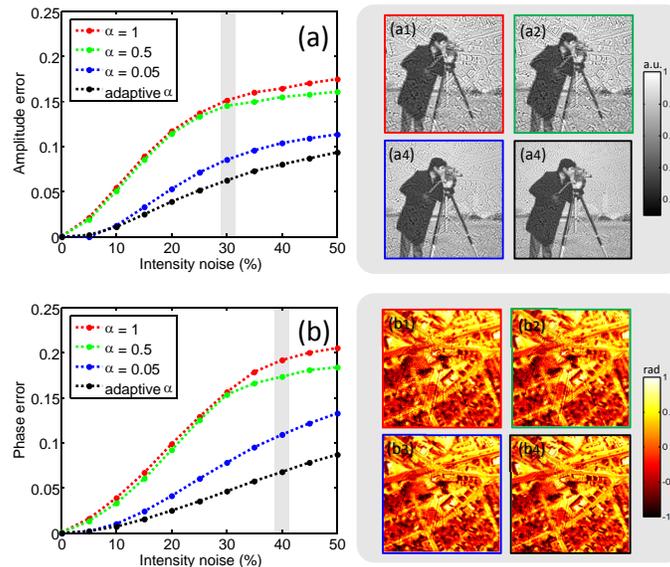
Fig. 2. Comparison of amplitude (a) and phase (b) reconstruction accuracy versus intensity noise for the adaptive step-size and fixed step-sizes α=1, 0.5, and 0.05. (a1)-(a4) and (b1)-(b4) are recovered amplitudes and phases when the noise level is 30% and 40%, respectively.

5.1. Performance comparison of fixed step-size and adaptive step-size approaches

First we compare the performance of incremental gradient approaches with different fixed step-sizes as well as our adaptive step-size scheme (note that the PIE and ePIE are exactly the same when there is no pupil estimation). Figure 2 plots the amplitude and phase reconstruction accuracy as a function of intensity noise. Both the adaptive step-size method and standard PIE method were implemented with several choices of step-size (α=1, 0.5, and 0.05, each algorithm was running for 100 iterations). It can be seen when there is no noise, perfect reconstruction can be achieved no matter the choice of the step-size. In the presence of noise, the red and green curves rise rapidly, indicating large α values (1, 0.5) are relatively sensitive to noise. Besides, the reconstructions suffer from a significant degree of cross-talk between the phase and amplitude part of the complex field. A smaller α preforms comparatively better at the expense of much slower convergence rate. The adaptive step-size method always performs better than other methods using a fixed step-size, and it reaches good solutions with minimum cross-talk between the phase and amplitude.

To better explain the small error achieved by the adaptive step-size method, we show the one typical group of phase error trajectories under 40% Gaussian noise in Fig. 3. Similar behavior can be observed for intensity error curves as well. It is seen that the adaptive step-size method decreases more rapidly than fixed step-size methods and converge in the early 20 iterations. For fixed step-size methods, the curves reach their respective minima in succession and then overshot. They finally 'converge' to a limit cycle with the oscillation amplitude roughly proportional to α, as predicted by the theory. From these results, we can safely conclude that the adaptive step-size

method outperforms the fixed step-size methods, with both faster convergence rate and lower mis-adjustment error simultaneously achieved.
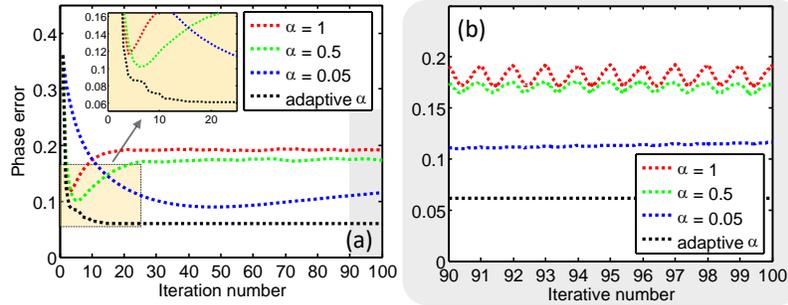
Fig. 3. Error trajectories for the adaptive step-size and fixed step-sizes α=1, 0.5, and 0.05. (a) Phase error curves under 40% Gaussian noise with last 10 iterations (shaded region) magnified in (b). Note each iteration comprises 225 sub-iterations carried on each low-resolution image.

Though the Gaussian noise model is assumed in the above simulations, the adaptive step-size approach can also be well applied to the case of Poisson noise. Figure 4 shows the final reconstruction accuracy for intensity MAE of 5% to 50%, which corresponds to an average of 10,000 to and 100 photons, per detector pixel. It can be seen that for the same noise level, the reconstruction errors under Poisson noise are smaller, however, the general trend of the error curves does not differ significantly to the case of Gaussian noise.
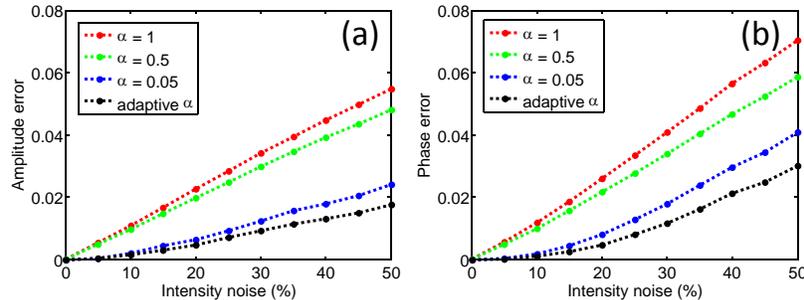
Fig. 4. Comparison of amplitude (a) and phase (b) reconstruction accuracy under Poisson noise for the adaptive step-size and fixed step-sizes α=1, 0.5, and 0.05.

5.2. *Comparison of the adaptive step-size approach with the state-of-the-arts under noisy conditions*

Next, we compare the proposed adaptive step-size approach with four state-of-the-art FPM reconstruction algorithms: (1) Gerchberg-Saxton [3, 12], (2) (sequential) Gauss-Newton [19], (3) Difference map [11], (4) Wirtinger flow [24]. The Gerchberg-Saxton and sequential Gauss-Newton are both incremental gradient approaches but the latter introduces an additional regularization parameter γ, which we use $\gamma = 5$ as suggested in [15]. The difference map and Wirtinger flow belong to global gradient approaches. The difference map approach is originally applied to conventional ptychography [11], and here we adapt it to the content of FPM and directly use the alternating projection instead of the hybrid input-output algorithm for Fourier-space updating. The Wirtinger flow algorithm optimizes the intensity-based cost function rather than

the amplitude-based one [Eq. (3)] with two adjustable parameters (k_0 and θ_{max}) in step-size to stabilize the convergence process. Here we use ($k_0 = 330$ and $\theta_{max} = 0.4$) as recommended in [24].
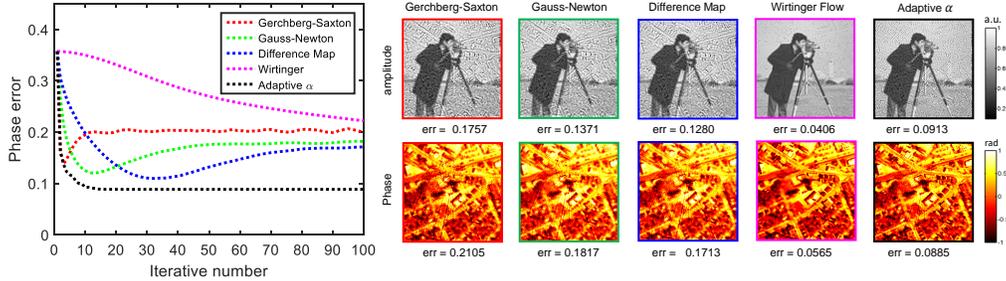
Fig. 5. Comparison of convergence speed and reconstruction accuracy of the adaptive step-size approach and four state-of-the-art FPM reconstruction algorithms under 50% Gaussian noise.

The four state-of-the-arts and the proposed adaptive step-size algorithm are compared under noisy conditions (50% Gaussian or Poisson noise). The convergence curves as well as the reconstructed amplitudes and phases are given in Fig. 5 for the case of Gaussian noise and in Fig. 6 for the case of Poisson noise. To study the convergence speed and computational complexity for each algorithm, their iteration numbers and total runtime required for convergence are compared in Table 1. These tests were made with an Intel i7 2.6 GHz processor and 16 Gbyte RAM in conjunction with MATLAB's cputime function. For different algorithms, the iterative process continues to run until it converges or stagnates. This allows to demonstrate their final reconstruction accuracy but not increase unnecessary running time for a fair comparison. Close inspection of the results from Figs. 5 and 6 reveals that the convergence behaviors of Gerchberg-Saxton, Gauss-Newton, and difference map approaches are quite similar to the standard PIE algorithm with a large, medium, and small step-size, respectively. This is understandable because when there is no pupil estimation, Gerchberg-Saxton algorithm is exactly the same as the standard PIE with a fixed step-size $\alpha = 1$. The Gauss-Newton algorithm improves the robustness of the algorithm towards noise with the addition of the regularization parameter γ. Strictly speaking, when the regularization parameter γ is much larger than 1 as in [15, 19] and in our case, it should not be termed as 'regularization parameter' (since typically regularization parameters should be much less than 1). Actually, it is equivalent to use a constant but reduced step-size ($\alpha = 1/6$) for the standard PIE updating process. If a very small step-size is used (like $\alpha = 0.05$), the convergence behavior of the incremental gradient approach approximates the global one (difference map), which can achieve lower reconstruction error at the expense of larger iteration number and longer runtime. Different from the other four approaches, the Wirtinger flow algorithm performs better for the case of Gaussian noise and achieves the lowest reconstruction error. This is because the intensity-based cost function used in Wirtinger flow algorithm assumes implicitly a Gaussian noise model [15], while the other four approaches minimize the amplitude-based cost function, which works better for the case of Poisson noise. Besides, the Wirtinger flow algorithm requires more than 1000 iterations to converge, and thus the total runtime is much longer than the other four approaches. In the case of the Gaussian noise, the proposed adaptive step-size method effectively prevents the overshooting and achieves the lowest error among the four amplitude-based approaches. While in the case of the Poisson noise, it takes less than half number of iterations to achieve a similar accuracy as the global approach (difference map). Though the Gercherg-Saxton is the fastest convergent in term of iteration number and runtime, it is the most sensitive to noise and tends to produce a noisy result

with strong artifacts. Therefore, the proposed adaptive step-size method can be regarded as the most efficient one if both the reconstruction accuracy and the convergence speed are considered.
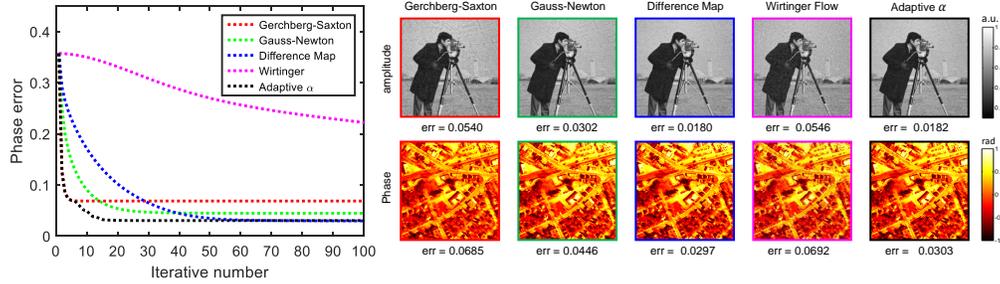
Fig. 6. Comparison of convergence speed and reconstruction accuracy of the adaptive step-size approach and four state-of-the-art FPM reconstruction algorithms under 50% Poisson noise.

Table 1. Runtime performance comparison of different algorithms

	Gaussian Noise		Poisson Noise	
	Iteration Number	Runtime (s)	Iteration Number	Runtime (s)
Gerchberg-Saxton	12	3.38	10	2.82
Gauss-Newton	51	15.52	47	14.20
Difference Map	82	21.99	67	18.05
Wirtinger Flow	1072	289.12	1743	478.86
Adaptive α	21	6.75	28	8.69

6. Experiments

We experimentally verify the adaptive step-size strategy based on a real FPM platform. The experimental setup generally followed our simulation parameters, except that a denser LED array is used. The lateral distance between two adjacent LEDs is 2.5mm, and the LED array is controlled by a self-developed Field-Programmable Gate Array (FPGA) circuit. For all the experiments reported in the following, the central 21×21 LEDs in the array are used to illumination the sample sequentially, resulting in a final synthetic NA_{syn} of ∼ 0.45 in both x and y direction. The synthetic NA equals to the sum of the NA_{obj} of the objective and the NA_{ill} of the illumination at the largest angle [37]. To achieve robust image reconstructions, a multi-exposure acquisition scheme is usually required to capture high dynamic range (HDR) images for both bright-field and dark-field illuminations [3]. However, in our experiments, all images are taken under the same exposure that accommodates the high intensity values of brightfield images to prevent sensor saturation and highlight the effect of noise.

6.1. Experimental results of an USAF resolution target

An USAF resolution target was first imaged to quantify the resolution improvement for each algorithm. The small portion of interest [128×128 pixels, shown in Fig. 7(b)], which contains the smallest group of feature (group 9 element 3, 1.56μm line pair spacing), is extracted from the full FoV image [Fig. 7(a)]. It can be seen from Fig. 7(b) that when we turn on the central LED only, the minimum resolvable feature is group 7 element 3 (6.2μm line pair spacing), which coincides with the well-known Rayleigh resolution limit of a conventional optical microscope under coherent illumination (λ/NA_{obj}). The maximum expected resolving power after FPM
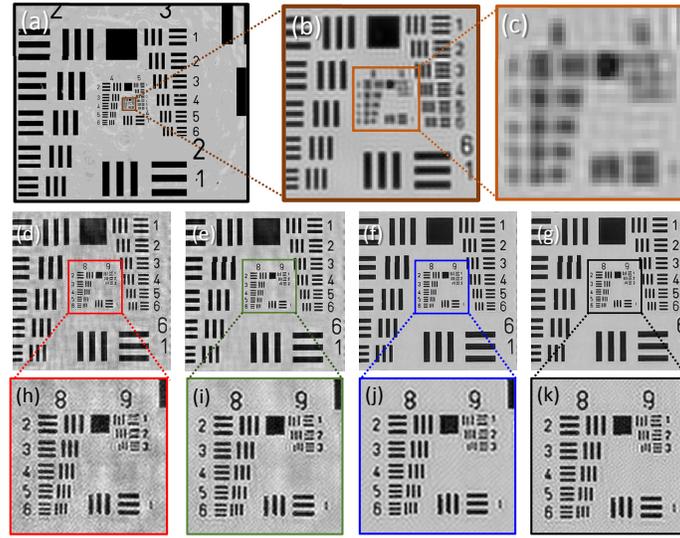
Fig. 7. Experimental results of a resolution target. (a) The raw full FoV image. (b) and (c) are the corresponding magnified areas. (d)-(g) are the reconstruction results with fixed step-sizes α=1, 0.5, and 0.05 (each runs for 30 iterations) and the adaptive step-size (converges in 17 iterations). (h)-(k) shows the corresponding zoom-ins of the smallest features.

reconstruction is $\lambda/NA_{syn} \approx 1.4\mu$m, which is sufficient for resolving the smallest feature in the USAF target. When dealing with experimental data, it is not always possible to obtain the true object complex function, so the MAE cannot be calculated. However, the normalized real-space error metric of the capture intensity images can be used as a measure or indicator of the reconstruction accuracy. The index is calculated as follows:

$$E = \frac{\varepsilon}{\sum_i \left\| \sqrt{\mathbf{I}_i} \right\|^2} \equiv \frac{\sum_i \left\| \sqrt{\mathbf{I}_i} - |\mathbf{g}_i| \right\|^2}{\sum_i \left\| \sqrt{\mathbf{I}_i} \right\|^2} \tag{17}$$

The low-resolution image set corresponding to the small tile was recovered with different strategies as in our simulation. The recovered high resolution images with fixed (α=1, 0.5, and 0.05) and adaptive step-sizes are shown in Figs. 7(d)-7(k), respectively. The corresponding error metric E [Eq. (17)] as a function of iteration number is plotted in Fig. 8. For a large fixed step-size α=1, strong artifacts are superimposed on the final reconstruction result, which not only smear the background but also distort several small-scale features [Figs. 7(d) and 7(h)]. Artifacts can be alleviated by reducing the step-size to 0.5, but the two smallest groups of feature (group 9 elements 2 and 3) are still not well resolved [Figs. 7(e) and 7(i)]. A small fixed step-size α=0.05 produced a much cleaner result with very little background noise at the expense of slower convergence rate [Fig. 7(f)]. Besides, as is shown in the magnified area [Fig. 7(j)], the high-spatial frequency details seem to be over-smoothed by the algorithm, resulting in blurred numbers and line structures at group 9 elements 2 and 3. These results indicate an improper choice of step-size can lead to blurring in high-frequency details if the step-size is too small or sub-optimal removal of noise artifacts if the step-size chosen is too large. Figures 7(g) and 7(k) shows the reconstructed result by the proposed adaptive step-size strategy. A video sequence (Visualization 1) is also presented to show the evolution of the iterative reconstruction process. It can be seen that a sharp but noisy reconstruction was quickly established during the first few iterations. Then the background noise gradually diminished and high-frequency
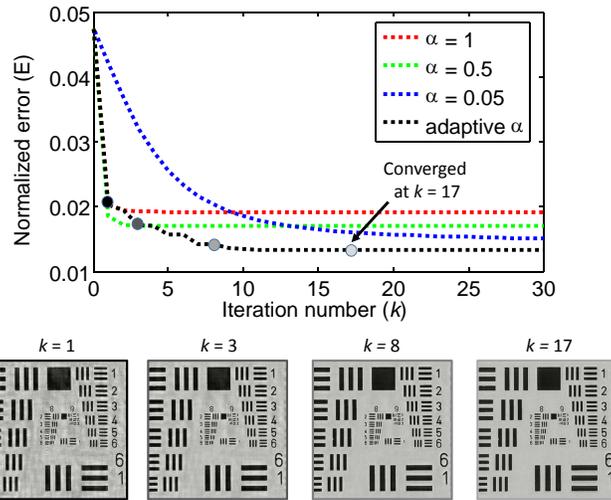
Fig. 8. The progress of the error metric E for the fixed (α=1, 0.5, and 0.05) and adaptive step-sizes. Visualization 1 and images in the second row illustrate the evolution of the amplitude reconstruction of the adaptive step-size approach.

inconsistencies were gradually refined as the step-size decreases. Finally, the iteration process converged smoothly after 17 iterations (α^k <0.001), resulting in a better reconstructed image with a uniform background and all groups of features clearly resolved.
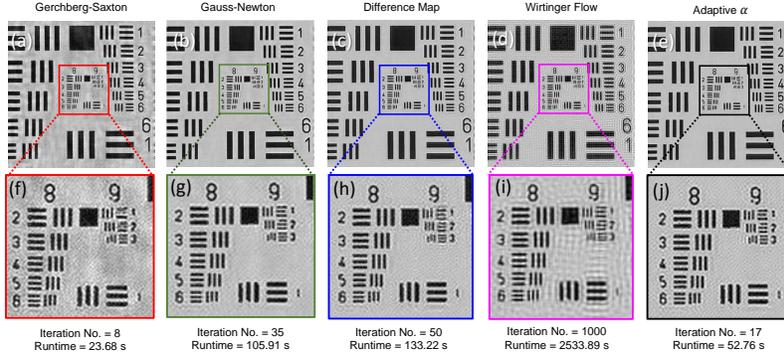
Fig. 9. Comparison of reconstruction results and runtime of the adaptive step-size approach and four state-of-the-art FPM reconstruction algorithms.

Besides its fixed step-size counterparts, the proposed adaptive step-size approach was also experimentally compared with four state-of-the-art FPM reconstruction algorithms, with the same parameter settings as in the simulation. Figure 9 shows the reconstructed amplitudes of different approaches, which generally mirror our simulation results. Note that for each algorithm, the runtime is longer than that of the simulation due to the raw dataset contains more images with four times pixel resolution, and thus the runtime per iteration for each algorithm is increased. As an equivalent of the PIE with a unity step-size, the Gerchberg-Saxton algorithm induced severe artifacts in the background, and several small-scale features were distorted [Figs. 9(a) and 9(f)]. By simply reducing the step-size (introducing regularization parameter γ), the background noise was greatly suppressed at the expense of more iterations required for convergence [Figs.

9(b) and 9(g)]. The difference map took 50 iterations to produce a clean image, but some fine line structures are not well reconstructed [Figs. 9(c) and 9(h)]. The result is quite similar to the result of the PIE with a small fixed step-size [Figs. 7(f) and 7(j)] because using a small α for incremental gradient approach tends to make the algorithm more 'global'. The Wirtinger flow algorithm has run for 1000 iterations, but finally produced an image with some ringing artifacts [Figs. 9(d) and 9(i)]. The explanation for this may be due to the fact that the noise of the realistic data is mainly Poisson, and the intensity-based cost function is more sensitive to model mismatch errors [15]. Comparatively, the proposed adaptive step-size approach achieves the best reconstruction quality with little computational overhead [Figs. 9(e) and 9(j)].
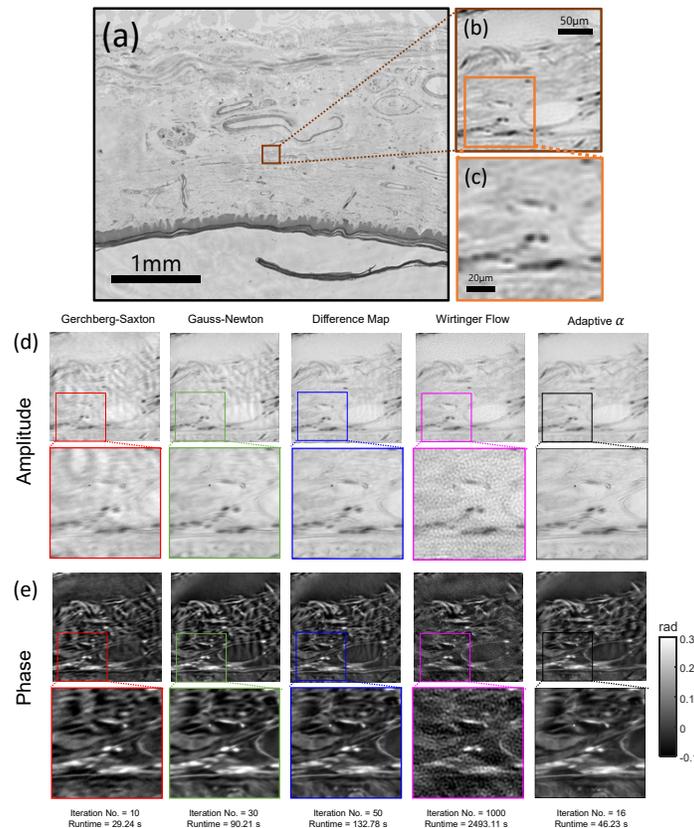
Fig. 10. Experimental results of a pathological section. (a) The full FoV image. (b-c) The corresponding region of interest. (d-e) The comparison of reconstructed amplitudes and phases of four state-of-the-arts and the adaptive step-size approach. Visualization 2 shows the evolution of amplitude and phase reconstructions with the adaptive step-size approach.

6.2. Experimental results of pathological section

Finally, the proposed adaptive step-size strategy was tested on a biological sample (a pathological section of human kidney), where the phase part of the complex field often contains valuable information on the studied specimen. Figure 10(a) presents the full FOV of the specimen and Figs. 10(b) and 10(c) show the corresponding magnified area of interest. Though the histological section is stained in red with hematoxylin and eosin (H. & E.), it is poorly visible under monochromatic red illumination. Figures 10(c)-10(d) show the reconstructed amplitude

and phase images of four state-of-the-arts and the adaptive step-size approach, respectively. Although the resolution improvements in the results of all the five reconstruction algorithms are obvious, the adaptive step-size approach, once again, outperforms the other methods with high reconstruction quality, low noise-induced artifacts, and fast convergence rate simultaneously achieved (Visualization 2).

7.   Conclusions and discussions

In this work, several state-of-the-art methods for solving ptychographic phase retrieval problem have been studied from a numerical analysis point of view. It has been found that the PIE and ePIE algorithms can be classified as incremental gradient descent approaches in the numerical optimization community, and their convergence behaviors are quite different from the global gradient descent approaches. Though incremental gradient approaches often offer reduced computational cost and faster convergence rate than global approaches, we prove that its convergence for convex sets requires the step-size to be gradually diminishing. Then an adaptive step-size strategy has been introduced in incremental solutions of ptychographic imaging and shown to achieve significant improvement in the stability and robustness of the reconstruction towards noise. Furthermore, the algorithm offers fast convergence rate with little computational overhead. As verified by simulations and experimental results in the context of FPM, it always produces good performance without manually tuning parameter or preknowledge about the noise statistics. In addition, we have compared our method with four state-of-the-art FPM reconstruction algorithms. It has been observed that in the early few iterations, it can approach a neighbourhood of the solution quickly just like the Gerchberg-Saxton algorithm, and then the vanishing step-size further allows to avoid stagnation at a poor local minimum and improve the quality of the reconstruction. The introduction of an adaptive step-size provides a simple tool which, in all of our numerical simulations and experiments, outperforms the current standards in terms of convergence speed and reconstruction quality without adding complexity to the reconstruction process. Our MATLAB source code for this algorithm has been made available on [38].

In this work, we only reconstruct the object function and assume that pupil function and LED positions are the accurately known a priori. It should be noted that several studies have demonstrate that with the help of inherent data redundancy of ptychography, it is possible to extract those inaccurately known system parameters along with the object complex function from the original FPM dataset [9, 12–14, 16]. As a result, the required setup stability and accuracy of system calibration can be relaxed. Recently, by incorporating the pixel sampling model and the multi-slice model into the original PIE framework, the data redundancy in FPM can further provide the potential of overcoming pixel aliasing problem [21] and depth-resolved imaging for thick specimens [25, 26]. Since the proposed adaptive step-size strategy is a simple extension of the original PIE and ePIE, we envisage it can be straightforwardly extended to those more sophisticated algorithms for pupil recovery [9, 12], misalignment correction [13, 16], multi-slice imaging [25, 26], and has the potential to be applied to label-free live cell imaging [39] and technical surface inspection [40]. In Appendix B, it is demonstrated that the same adaptive step-size rule can be well applied to the pupil updating (the adaptive α is used for both object and pupil updating). Nevertheless, based on the simulation results presented in Fig. 11, we recommend to use a smaller initial value ($\beta^0 = 1/\sqrt{N}$, N is the total number of low-resolution images in the dataset) for pupil step-size for getting the best performance from the adaptive step-size algorithm.

Appendix A: Convergence proof of incremental solution with diminishing step-sizes in the convex setting

In this Appendix, we prove the global convergence of the incremental gradient solution [Eq. (11)] with diminishing step-sizes in the convex setting. Main results of the proof are adapted

from [41] and [42]. The required assumptions for the proof are the following: ε_i is convex, $\nabla \varepsilon_i$ is bounded on $\mathbb{C}$. The $\mathbf{W}_i$ in Eq. (11) is some nonnegative definite diagonal matrix (function) that are independent with the subiteration index i ($\mathbf{W}_i \equiv \mathbf{W}$). Define a weighted norm $\|\cdot\|_{\mathbf{W}^{-1}}$ by $\|\mathbf{O}\|_{\mathbf{W}^{-1}} = \sqrt{\mathbf{O}^* \mathbf{W}^{-1} \mathbf{O}}$. Suppose that $\varepsilon^{\star} = \inf_{\mathbf{O} \in \mathbb{C}} \varepsilon(\mathbf{O})$. The following Lemmas 1 and 2 are firstly given and will be used in the convergence proof.

Lemma 1: Let $\{\mathbf{O}^k\}$ be a sequence generated by incremental gradient solution Eq. (11). Then for all $\mathbf{O} \in \mathbb{C}$ and all k and some $C > 0$, we have:

$$\left\|\mathbf{O}^{k+1} - \mathbf{O}\right\|^2_{\mathbf{W}^{-1}} \leq \left\|\mathbf{O}^k - \mathbf{O}\right\|^2_{\mathbf{W}^{-1}} - 2\alpha^k \left(\varepsilon\left(\mathbf{O}^k\right) - \varepsilon(\mathbf{O})\right) + \left(\alpha^k\right)^2 C. \tag{18}$$

Proof: One can verify that the algorithm Eq. (11) is equivalent to the following

$$\mathbf{Q}^k_{i+1} = \mathbf{Q}^k_i + \alpha^k \nabla \xi_i(\mathbf{O}_i{}^k). \tag{19}$$

where $\mathbf{Q}^k_i = \mathbf{W}_i^{-1/2} \mathbf{O}^k_i$, $\xi_i\left(\mathbf{O}^k_i\right) = \varepsilon_i\left(\mathbf{W}_i^{1/2} \mathbf{O}^k_i\right)$. Then use Lemma 2.1 of [41].

Lemma 1 guarantees that with a cyclic order, the incremental gradient solution Eq. (11) yields a point $\mathbf{O}^{k+1}$ at the end of the cycle will be closer to $\mathbf{O}$ than $\mathbf{O}^k$, provided the step-size α^k is less than $2\left(\varepsilon\left(\mathbf{O}^k\right) - \varepsilon(\mathbf{O})\right) C^{-1}$. In particular, for any $e > 0$ and assuming that there exists an optimal solution $\mathbf{O}^{\star}$, either we are within $\frac{1}{2}\alpha^k C + e$ of the optimal value

$$\varepsilon\left(\mathbf{O}^k\right) \leq \varepsilon\left(\mathbf{O}^{\star}\right) + \tfrac{1}{2}\alpha^k C + e, \tag{20}$$

or the squared distance to $\mathbf{O}^{\star}$ will be strictly decreased by at least $2\alpha^k e$.

$$\left\|\mathbf{O}^{k+1} - \mathbf{O}^{\star}\right\|^2_{\mathbf{W}^{-1}} \leq \left\|\mathbf{O}^k - \mathbf{O}^{\star}\right\|^2_{\mathbf{W}^{-1}} - 2\alpha^k e, \tag{21}$$

This suggests that for a constant step-size ($\alpha^k \equiv \alpha$), convergence can be established to a neighborhood of the optimum which shrinks to 0 as $\alpha \to 0$, as stated in Lemma 2.

Lemma 2: Let $\{\mathbf{O}^k\}$ be a sequence generated by incremental gradient solution Eq. (11) with the step-size α^k such that

$$\alpha^k > 0, \ \lim_{k \to \infty} \alpha^k = 0, \ and \ \sum_{k=0}^{\infty} \alpha_k = \infty. \tag{22}$$

Then $\liminf_{k \to \infty} \varepsilon\left(\mathbf{O}^k\right) = \varepsilon^{\star}$.

Proof: The proof uses Proposition 1.2 of [42]. Assume for contradiction that there are $\delta > 0$, $K \in \mathbb{N}$, $\upsilon \in \mathbb{C}$, and $\varepsilon(\upsilon) < \varepsilon\left(\mathbf{O}^k\right) - \delta$ for all $k \geq K$. Since $\lim_{k \to \infty} \alpha^k = 0$, we can assume that K is so large that $\alpha_k C < \delta$ where $C > 0$ is a constant from Lemma 1. Using Lemma 1, we obtain:

$$\begin{aligned} \left\|\mathbf{O}^{k+1} - \upsilon\right\|^2_{\mathbf{W}^{-1}} &\leq \left\|\mathbf{O}^k - \upsilon\right\|^2_{\mathbf{W}^{-1}} + \alpha^k \left(\alpha^k C - 2\delta\right) \\ &\leq \left\|\mathbf{O}^k - \upsilon\right\|^2_{\mathbf{W}^{-1}} - \alpha^k \delta \end{aligned} \tag{23}$$

for all $k \geq K$. Summing up, this gives

$$0 \leq \left\|\mathbf{O}^k - \upsilon\right\|^2_{\mathbf{W}^{-1}} \leq \left\|\mathbf{O}^K - \upsilon\right\|^2_{\mathbf{W}^{-1}} - \delta \sum_{n=K}^{k-1} \alpha^n. \tag{24}$$

for all $k > K$. Letting $k \to \infty$, $\sum_{k=0}^{\infty} \alpha_k = \infty$ is contradicted.

Convergence Theorem: Let $\{\mathbf{O}^k\}$ is a sequence generated by the algorithm Eq. (11) with the step-size α^k such that

$$\alpha^k > 0, \ \lim_{k \to \infty} \alpha^k = 0, \ \sum_{k=0}^{\infty} \alpha^k = \infty, and \ \sum_{k=0}^{\infty} \left(\alpha^k\right)^2 < \infty. \tag{25}$$

Then $\{\mathbf{O}^k\}$ converges to some $\mathbf{O}^\star \in \Lambda^\star = \{\upsilon \in \mathbb{C} : \varepsilon(\upsilon) \leq \varepsilon(\mathbf{O}), \forall \mathbf{O} \in \mathbb{C}\}$.

Proof: Using Lemma 1 with some $\upsilon \in \Lambda^\star$, we have

$$\left\|\mathbf{O}^{k+1} - \upsilon\right\|^2_{\mathbf{W}^{-1}} \leq \left\|\mathbf{O}^0 - \upsilon\right\|^2_{\mathbf{W}^{-1}} - 2\sum_{n=0}^{k} \alpha^n \left(\varepsilon(\mathbf{O}^n) - \varepsilon^\star\right) + \delta \sum_{n=0}^{k} (\alpha^n)^2 C. \quad (26)$$

for all k. Since $\sum_{k=0}^{\infty} \left(\alpha^k\right)^2 < \infty$, we have

$$2\sum_{n=0}^{k} \alpha^n \left(\varepsilon(\mathbf{O}^n) - \varepsilon^\star\right) \leq \left\|\mathbf{O}^0 - \upsilon\right\|^2_{\mathbf{W}^{-1}} + \sum_{n=0}^{k} (\alpha^n)^2 C < q < \infty, \quad (27)$$

for all n and some q where C is a constant from Lemma 1. This implies that $\sum_{n=0}^{\infty} \alpha^n \left(\varepsilon(\mathbf{O}^n) - \varepsilon^\star\right) < \infty$ since $\varepsilon(\mathbf{O}^n) - \varepsilon^\star \geq 0, \forall n$. Therefore, Eq. (26) implies that $\{\mathbf{O}^k\}$ is bounded. Furthermore, by Lemma 2, there exists a subsequence $\{\mathbf{O}^{k_n}\}$ of $\{\mathbf{O}^k\}$ such that $\lim_{n\to\infty} \varepsilon(\mathbf{O}^{k_n}) = \mathbf{O}^\star$. Since $\{\mathbf{O}^{k_n}\}$ is bounded, there exists a subsequence $\{\mathbf{O}^{k_{n_l}}\}$ of $\{\mathbf{O}^{k_n}\}$ such that $\{\mathbf{O}^{k_{n_l}}\}$ converge to some $\mathbf{O}^\star \in \mathbb{C}$. By the continuity of ε, we have $\varepsilon(\mathbf{O}^\star) = \varepsilon^\star$, that is, $\mathbf{O}^\star \in \Lambda^\star$. We have obtained a limit point $\mathbf{O}^\star \in \Lambda^\star$ of $\{\mathbf{O}^k\}$. Now we follow the line of the proof of Proposition 1.3 of [42]. For any $\delta > 0$, take $K \in \mathbb{N}$ such that $\left\|\mathbf{O}^k - \mathbf{O}^\star\right\|^2_{\mathbf{W}^{-1}} \leq \delta/2$ and $\sum_{n=K}^{\infty} \left(-2\alpha^n \left(\varepsilon(\mathbf{O}^n) - \varepsilon^\star\right) + (\alpha^n)^2 C\right) \leq \delta/2$. Using Lemma 1, one obtains:

$$\left\|\mathbf{O}^{k+1} - \mathbf{O}^\star\right\|^2_{\mathbf{W}^{-1}} \leq \left\|\mathbf{O}^k - \mathbf{O}^\star\right\|^2_{\mathbf{W}^{-1}} + \sum_{n=K}^{k} \left(-2\alpha^n \left(\varepsilon(\mathbf{O}^n) - \varepsilon^\star\right) + (\alpha^n)^2 C\right) \leq \delta. \quad (28)$$

Hence, $\{\mathbf{O}^k\}$ converges to an optimal solution $\mathbf{O}^\star \in \Lambda^\star$. Note that comparing Eq. (22) with Eq. (25), an additional mild restriction on the step-size sequence $\sum_{k=0}^{\infty} \left(\alpha^k\right)^2 < \infty$ is included to guarantee convergence of the entire sequence $\{\mathbf{O}^k\}$.

Appendix B: Performance analysis of adaptive step-size strategy for pupil recovery

In this Appendix, we evaluate the performance of adaptive step-size strategy for pupil recovery using simulated data in Section 5, but with additional pupil aberration error. The update of pupil function is same as the object function [Eq. (11)], with the roles of the object and pupil have been exchanged.

$$\mathbf{P}^k_{i+1} = \mathbf{P}^k_i - \beta^k \mathbf{W}_i \mathbf{O}^*_i \left(\mathbf{\Psi}^k_i - \mathbf{O}_i \mathbf{P}^k_i\right) \quad (29)$$

While the object and pupil cannot be decoupled analytically, the dataset redundancy allows the algorithm to update the object and pupil in turns to find their respective solutions. Note in the pupil update function, a new step-size β is introduced, which is often set to $\beta^k = \beta \equiv 1$ (same as the object step-size) as suggested in [9, 12, 13]. Here we wish to demonstrate that the same adaptive step-size rule can be well applied to the pupil updating (using the adaptive α is used for both object and pupil updating). Figure 11 shows the convergence curves as well as the reconstructed amplitudes and phases with various fixed step-sizes and adaptive step-size under aberrated and noisy data (the images were corrupted with 50% Poisson noise). As can be seen, using large fixed sizes $\alpha = \beta = 1$ leads to slow divergence and finally incurs relatively large errors after 100 iterations. Decreasing the step-size slows down but stabilizes the convergence process, and finally can reach points with lower errors. Comparatively, the adaptive step-size approach gives a better complex-field reconstruction as well as a more accurate pupil estimation with much fewer iterations. We further found that additional improvements can be obtained by reducing the initial value for pupil step-size to about 0.067 [$\beta^0 = 1/\sqrt{N} = 1/15$, N is the total number of LEDs (225)]. In such case, both α^k and β^k are halved when no apparent reduction
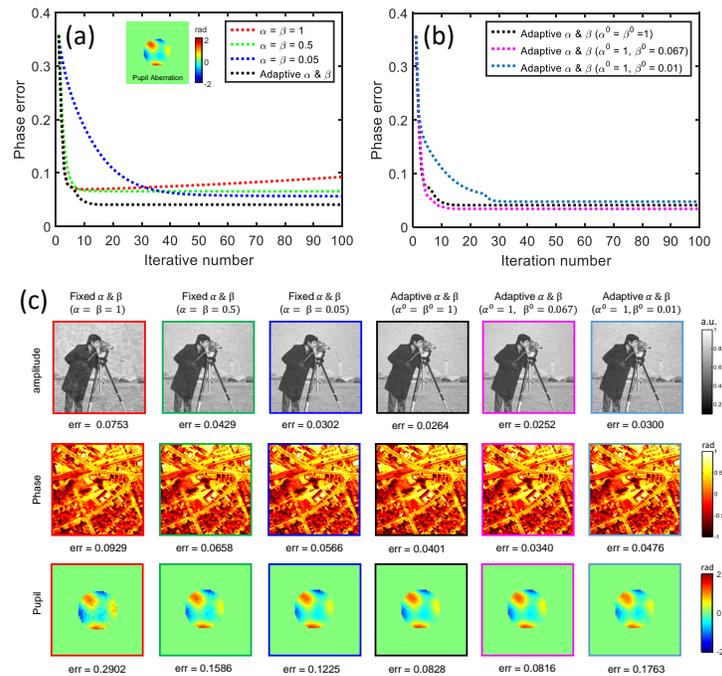
Fig. 11. Comparison of convergence speed and reconstruction accuracy for joint object and pupil recovery under 50% Poisson noise. (a) Convergence curves for different fixed and the adaptive step-size approaches. (b) Convergence curves of adaptive step-size approaches with different initial values for pupil step-size. (c) The reconstructed amplitudes, phases, and pupil aberration functions with different choices of step-size.

in the error metric is made by the iteration, and β^k is always $\sqrt{N}$ times smaller than α^k during the entire iteration process. The rationality for this lies in the fact that the pupil function is more frequently updated than the object function (as all of the sub-spectrum images share the same pupil function, the pupil function can be updated at each sub-pupil region; but the object function at a particular location in the Fourier space can only updated when it is covered by the sub-pupil). This intuitive modification helps to establish synchronous convergence of both object and pupil function, and thus allows for further improvement in algorithm performance. It can also be expected that a further reduction of initial pupil step-size to $\beta^0 = 0.01$ only slows down the convergence but offers no further improvement in reconstruction quality.

Funding

Acknowledgments