Phase discrepancy analysis and compensation for fast Fourier transform based solution of the transport of intensity equation

Chao Zuo,[1,2,4] Qian Chen,[1,*] Lei Huang,[3] and Anand Asundi[2]

[1]*Jiangsu Key Laboratory of Spectral Imaging & Intelligent Sense, Nanjing University of Science and Technology, Nanjing, Jiangsu Province 210094, China*
[2]*Centre for Optical and Laser Engineering, Nanyang Technological University, 639798 Singapore*
[3]*Brookhaven National Laboratory - NSLS II, 50 Rutherford Dr. Upton, New York 11973-5000, USA*
[4]*surpasszuo@163.com*
[*]*chenqian@njust.edu.cn*

Abstract: The transport of intensity equation (TIE) has long been recognized as a *quantitative* method for phase retrieval and phase contrast imaging. However, it is shown that the most widely accepted fast Fourier transform (FFT) based solutions do *not* provide an exact solution to the TIE in general. The root of the problem lies in the so-called "Teague's assumption" that the transverse flux is considered to be a conservative field, which cannot be satisfied for a general object. In this work, we present the theoretical analysis of the phase discrepancy owing to the Teague's assumption, and derive the necessary and sufficient conditions for the FFT-based solution to coincide with the exact phase. An iterative algorithm is then proposed aiming to compensate such phase discrepancy in a simple yet effective manner.

©2014 Optical Society of America

References and links

1. M. Reed Teague, "Deterministic phase retrieval: a Green's function solution," J. Opt. Soc. Am. **73**(11), 1434–1441 (1983).
2. D. Paganin and K. A. Nugent, "Noninterferometric phase imaging with partially coherent light," Phys. Rev. Lett. **80**(12), 2586–2589 (1998).
3. A. M. Zysk, R. W. Schoonover, P. S. Carney, and M. A. Anastasio, "Transport of intensity and spectrum for partially coherent fields," Opt. Lett. **35**(13), 2239–2241 (2010).
4. F. Roddier, "Curvature sensing and compensation: a new concept in adaptive optics," Appl. Opt. **27**(7), 1223–1225 (1988).
5. K. A. Nugent, T. E. Gureyev, D. J. Cookson, D. Paganin, and Z. Barnea, "Quantitative phase imaging using hard x rays," Phys. Rev. Lett. **77**(14), 2961–2964 (1996).
6. K. Ishizuka and B. Allman, "Phase measurement of atomic resolution image using transport of intensity equation," J. Electron Microsc. (Tokyo) **54**(3), 191–197 (2005).
7. A. Barty, K. A. Nugent, D. Paganin, and A. Roberts, "Quantitative optical phase microscopy," Opt. Lett. **23**(11), 817–819 (1998).
8. A. Orth and K. B. Crozier, "Light field moment imaging," Opt. Lett. **38**(15), 2666–2668 (2013).
9. C. Zuo, Q. Chen, and A. Asundi, "Light field moment imaging: comment," Opt. Lett. **39**(3), 654 (2014).
10. L. Waller, S. S. Kou, C. J. R. Sheppard, and G. Barbastathis, "Phase from chromatic aberrations," Opt. Express **18**(22), 22817–22825 (2010).
11. C. Zuo, Q. Chen, W. Qu, and A. Asundi, "High-speed transport-of-intensity phase microscopy with an electrically tunable lens," Opt. Express **21**(20), 24060–24075 (2013).
12. C. Zuo, Q. Chen, W. Qu, and A. Asundi, "Noninterferometric single-shot quantitative phase microscopy," Opt. Lett. **38**(18), 3538–3541 (2013).
13. T. E. Gureyev, A. Roberts, and K. A. Nugent, "Partially coherent fields, the transport-of-intensity equation, and phase uniqueness," J. Opt. Soc. Am. A **12**(9), 1942–1946 (1995).
14. L. J. Allen and M. P. Oxley, "Phase retrieval from series of images obtained by defocus variation," Opt. Commun. **199**(1-4), 65–75 (2001).

15. C. Zuo, Q. Chen, and A. Asundi, "Boundary-artifact-free phase retrieval with the transport of intensity equation: fast solution with use of discrete cosine transform," Opt. Express **22**(8), 9220–9244 (2014).
16. T. E. Gureyev and K. A. Nugent, "Rapid quantitative phase imaging using the transport of intensity equation," Opt. Commun. **133**(1-6), 339–346 (1997).
17. J. A. Schmalz, T. E. Gureyev, D. M. Paganin, and K. M. Pavlov, "Phase retrieval using radiation and matter-wave fields: Validity of Teague's method for solution of the transport-of-intensity equation," Phys. Rev. A **84**(2), 023808 (2011).
18. J. A. Ferrari, G. A. Ayubi, J. L. Flores, and C. D. Perciante, "Transport of intensity equation: Validity limits of the usually accepted solution," Opt. Commun. **318**, 133–136 (2014).
19. M. Testorf, B. Hennelly, and J. Ojeda-Castañeda, *Phase-Space Optics: Fundamentals and Applications: Fundamentals and Applications: Fundamentals and Applications* (McGraw Hill Professional, 2009).
20. R. Winston and W. T. Welford, "Geometrical vector flux and some new nonimaging concentrators," J. Opt. Soc. Am. **69**(4), 532–536 (1979).
21. M. J. Bastiaans, "Application of the Wigner distribution function to partially coherent light," J. Opt. Soc. Am. A **3**(8), 1227–1238 (1986).
22. B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal. I. Fundamentals," P IEEE **80**(4), 520–538 (1992).
23. H. Jeffreys and B. Jeffreys, *Methods of Mathematical Physics* (Cambridge University, 1999).
24. L. J. Allen, H. M. Faulkner, K. A. Nugent, M. P. Oxley, and D. Paganin, "Phase retrieval from images in the presence of first-order vortices," Phys. Rev. E Stat. Nonlin. Soft Matter Phys. **63**(3), 037602 (2001).
25. A. Lubk, G. Guzzinati, F. Börrnert, and J. Verbeeck, "Transport of Intensity Phase Retrieval of Arbitrary Wave Fields Including Vortices," Phys. Rev. Lett. **111**(17), 173902 (2013).
26. S. R. Crouch and J. D. Ingle, *Spectrochemical Analysis* (Prentice Hall, 1988).
27. V. Berinde, "Approximating fixed points of weak contractions using the Picard iteration," in *Nonlinear Analysis Forum* (2004), 43–54.
28. D. Paganin, A. Barty, P. J. McMahon, and K. A. Nugent, "Quantitative phase-amplitude microscopy. III. The effects of noise," J. Microsc. **214**(1), 51–61 (2004).
29. C. Zuo, Q. Chen, Y. Yu, and A. Asundi, "Transport-of-intensity phase imaging using Savitzky-Golay differentiation filter--theory and applications," Opt. Express **21**(5), 5346–5362 (2013).
30. A. J. Jerri, "The Shannon sampling theorem—Its various extensions and applications: A tutorial review," P IEEE **65**(11), 1565–1596 (1977).
31. J. Martinez-Carranza, K. Falaggis, and T. Kozacki, "Optimum measurement criteria for the axial derivative intensity used in transport of intensity-equation-based solvers," Opt. Lett. **39**(2), 182–185 (2014).
32. L. Waller, L. Tian, and G. Barbastathis, "Transport of Intensity phase-amplitude imaging with higher order intensity derivatives," Opt. Express **18**(12), 12552–12561 (2010).
33. T. Gureyev, C. Raven, A. Snigirev, I. Snigireva, and S. Wilkins, "Hard x-ray quantitative non-interferometric phase-contrast microscopy," J. Phys. D Appl. Phys. **32**(5), 563–567 (1999).
34. J. Martinez-Carranza, K. Falaggis, T. Kozacki, and M. Kujawinska, "Effect of imposed boundary conditions on the accuracy of transport of intensity equation based solvers," Proc. SPIE **8789**, 87890N (2013).

1. Introduction

Transport of intensity equation (TIE) is a powerful tool for phase retrieval due to its non-interferometric, deterministic nature [1]. This method has recently received revived interest, owing to its potential to provide wavefront metrology with simple and inexpensive setups, without requiring external interferometric reference beam. By solving the TIE, the quantitative phase can be directly recovered with only a minimum of two intensity measurements at distinct planes orthogonal to the optical axis. The TIE relaxes the stringent beam-coherence requirements for interferometry, making it possible to realize quantitative phase imaging over a wide range of light- and electron-beams [2, 3]. In the past few decades, the TIE has found a variety of applications in adaptive optics [4], X-ray diffraction [5], electron microscopy [6], optical quantitative phase imaging [7] and computational light field imaging [8, 9]. Several optical configurations for realizing dynamic TIE phase imaging have also been developed with use of the chromatic aberrations [10], tunable lens [11], and spatial light modulator [12].

Despite these great improvements and extensive applications, the literature survey conducted by the authors reveals that one important issue regarding the solution of the TIE has not yet been well addressed in the practice of the TIE-based methods. The well-posedness and the uniqueness of the TIE solution are well studied in the literature - with proper boundary conditions, one may find a unique (apart from an arbitrary additive constant) solution to the TIE if the intensity distribution is strictly positive [13]. Nevertheless, the standard procedure

for solving the TIE includes the introduction of the Teague's auxiliary function [1] that transforms the TIE into a classical two-dimensional (2D) Poisson equation [14, 15]. This treatment makes the TIE conveniently solvable by Fourier transform-based methods for periodic or some simplified homogeneous boundary conditions [2, 16], or the discrete cosine transform based method for inhomogeneous Neumann boundary conditions [15]. However, it is important to remark that the Teague's auxiliary function does not always exist in practical situations since the transverse energy flux may not be conservative, and consequently it would produce results that would not adequately match the exact solution [17, 18]. This discrepancy has not yet been recognized as a significant problem since previously the TIE was mainly applied for weak absorbing specimens, like optical components and biological samples. Although in such cases the methods using Teague's auxiliary function are not exactly accurate, they provide a very good approximation for the exact solution to the TIE because the transverse energy flux is nearly irrotational. However, as will be shown in this paper, the phase error is often nonnegligible and may even be relatively large when the measured sample exhibits a strong absorption. This obviously poses a significant problem in the practice of the TIE phase retrieval, which has been generally acknowledged to be quantitative. To this end, this work aims to present the theoretical analysis of the phase discrepancy owing to the Teague's auxiliary function as well as a simple iterative algorithm to compensate such phase errors. It addresses a long-lasting issue regarding the foundations of the widely applied TIE-based phase retrieval methods and extends the scope of their applicability.

2. Transport of intensity equation, partial coherence, and geometrical vector flux

The TIE was originally derived by Teague [1] from the stationary paraxial wave equation with the assumption of a purely coherent field. Its validity for partially coherent fields was then explored and confirmed by properly extending the physical meaning of phase, connecting it to the current density or the energy flow (Poynting vector) [2, 9]. Without loss of generality, we consider a quasi-monochromatic, stationary and ergodic scalar field $U(\mathbf{x})$, where $\mathbf{x}$ is the 2D spatial vector. To characterize such statistical field, we introduce the Wigner distribution function (WDF) in 4D phase space $(\mathbf{x}, \mathbf{u})$ [19], defined as

$$W(\mathbf{x}, \mathbf{u}) = \int \Gamma\left(\mathbf{x} + \frac{\mathbf{x}'}{2}, \mathbf{x} - \frac{\mathbf{x}'}{2}\right) \exp(-i2\pi\mathbf{u}\mathbf{x}') d\mathbf{x}', \tag{1}$$

where $\mathbf{u}$ is the spatial frequency vector corresponding to $\mathbf{x}$. $\Gamma(\mathbf{x}_1, \mathbf{x}_2) = \langle U(\mathbf{x}_1) U^*(\mathbf{x}_2) \rangle$ is the mutual intensity function of the statistical field. With use of the WDF, we can construct the 3D vector field $\boldsymbol{j} = [\boldsymbol{j}_\mathbf{x}, j_z]^T$, which is also known as the geometrical vector flux [20, 21]

$$\boldsymbol{j}_\mathbf{x}(\mathbf{x}) = \lambda \int \mathbf{u} W(\mathbf{x}, \mathbf{u}) d\mathbf{u}, \tag{2}$$

$$j_z(\mathbf{x}) = \frac{1}{k} \int \sqrt{k^2 - 4\pi^2 |\mathbf{u}|^2} W(\mathbf{x}, \mathbf{u}) d\mathbf{u}, \tag{3}$$

where $k = 2\pi/\lambda$ represents the wave number and λ is the wavelength. In quantum mechanics, $\boldsymbol{j}$ is the probability current, which is the expectation value of the local momentum operator. For an optical field well approximated by a scalar wave, which will be the case of interest in this work, $\boldsymbol{j}$ is the energy flux directed along the Poynting vector. The energy conservation in free space implies the geometrical vector flux has zero divergence, resulting to the following continuity equation

$$\frac{\partial j_z(\mathbf{x})}{\partial z} = -\nabla \cdot \boldsymbol{j}_\mathbf{x}(\mathbf{x}), \tag{4}$$

with ∇ only operating over the transverse plane. Under the paraxial approximation, $\sqrt{k^2 - 4\pi^2 |\mathbf{u}|^2}$ can be replaced by k, then the longitudinal energy flux reduces to the intensity, given by the space marginal of the WDF

$$j_z(\mathbf{x}) \approx \int W(\mathbf{x}, \mathbf{u}) d\mathbf{u} = I(\mathbf{x}). \qquad (5)$$

Substituting it into Eq. (1), we get the TIE for partially coherent fields

$$\frac{\partial I(\mathbf{x})}{\partial z} = -\nabla \cdot \boldsymbol{j}_\mathbf{x}(\mathbf{x}). \qquad (6)$$

Equation (6) is a generalized version of Teague's TIE since it explicitly considers the partial coherence and reduces to the conventional TIE in the case of perfect coherence:

$$\frac{\partial I(\mathbf{x})}{\partial z} = -\frac{1}{k}\nabla \cdot \left[I(\mathbf{x})\nabla\phi(\mathbf{x}) \right], \qquad (7)$$

where $\phi(\mathbf{x})$ is the phase of the completely coherent deterministic complex field, represented as $U(\mathbf{x}) = \sqrt{I(\mathbf{x})}\exp[i\phi(\mathbf{x})]$. Note here we employed the following relation between the first conditional frequency moment of the WDF (instantaneous frequency) and the transverse gradient of the phase of the complex field [22]:

$$\frac{\int \mathbf{u}W(\mathbf{x}, \mathbf{u}) d\mathbf{u}}{\int W(\mathbf{x}, \mathbf{u}) d\mathbf{u}} = I^{-1}\int \mathbf{u}W(\mathbf{x}, \mathbf{u}) d\mathbf{u} = \frac{1}{2\pi}\nabla\phi(\mathbf{x}). \qquad (8)$$

The TIE [Eq. (7) is a second order elliptic partial differential equation and the uniqueness of the solution seems well-acknowledged in the literature - with proper boundary conditions, one may find a unique (or unique up to an arbitrary additive constant) solution to TIE if the intensity distribution is strictly positive [13, 15]. Note Eq. (8) also provided a generalized definition of phase, which is very useful in optical phase retrieval when the object is illuminated by a partially coherent beam.

Though Eqs. (6) and (7) suggest that the transverse flux for a fully coherent field takes the form

$$\boldsymbol{j}_\mathbf{x}(\mathbf{x}) = -k^{-1}I(\mathbf{x})\nabla\phi(\mathbf{x}), \qquad (9)$$

one should always treat it with care because the $\nabla\phi(\mathbf{x})$ remains undefined where the amplitude goes to zero. Since the zeros of intensity are usually associated with the presence of phase vortices, there is no way to define a $\phi(\mathbf{x})$ so that it will be differentiable everywhere, which in turn leads to the fact that $\nabla\phi(\mathbf{x})$ in Eq. (9) cannot be considered to be simply the gradient of some scalar potential. According to the Helmholtz decomposition theorem for vector fields [23], it is therefore more properly to modify the $\nabla\phi(\mathbf{x})$ in Eq. (9) as the sum of the gradient of a scalar potential function $\phi_s(\mathbf{x})$ and the curl of a vector potential function $\phi_v(\mathbf{x})$

$$\boldsymbol{j}_\mathbf{x}(\mathbf{x}) = -k^{-1}I(\mathbf{x})\left[\nabla\phi_s(\mathbf{x}) + \nabla\times\phi_v(\mathbf{x}) \right]. \qquad (10)$$

This decomposition is unique (or unique up to a vectorial constant that may float between the two components) on bounded domains coupled with appropriate (Dirichlet, Neumann, or periodic) boundary conditions. Equation (10) implies that the scalar phase is a drive for the lateral translation of the intensity, while the vector phase twists the intensity as the wave moves along z. Therefore, as the wave propagates, both of the two kinds of phase will manifest themselves, contributing to the longitudinal intensity variations. However, the vortex cores are associated with intensity zeros that break the uniqueness theorems of the TIE [13], making conventional TIE solvers fail in such scenario [2]. Although some recent work extends the applicability of the TIE to arbitrary phase functions by explicitly taking into account the effect

of the vorticity [24, 25], using intensity zeros to identify singularities tends to be difficult since *very-near* zero and *at* zero may seem almost the same but can be entirely different *vis-a-vis* the presence of a phase singularity point. In the remainder of this paper, we exclude the consideration of the vector phase ($\phi(\mathbf{x}) = \phi_s(\mathbf{x})$) by assuming *a priori* that the phase is single-valued and continuous. This assumption is generally reasonable since the TIE phase measurements are habitually performed directly in (or very close to) the object plane (or its conjugated plane), where the phase function is single-valued and well-defined by the projected optical path length and the refractive index of the sample.

3. Theoretical analysis of the phase discrepancy

Putting the phase vorticity aside, we now turn to the problem of phase discrepancy in well-acknowledged TIE solvers. In the literature, the TIE is usually solved under the so-called "Teague's assumption" that the transverse flux is conservative so that an a scalar potential (auxiliary function) ψ exists that satisfying

$$\mathbf{j_x}(\mathbf{x}) = -k^{-1}I(\mathbf{x})\nabla\phi(\mathbf{x}) = -k^{-1}\nabla\psi(\mathbf{x}), \tag{11}$$

In this way, the TIE can be reduced to two following standard Poisson equations

$$\frac{\partial I(\mathbf{x})}{\partial z} = -\frac{1}{k}\nabla^2\psi(\mathbf{x}), \tag{12}$$

and

$$\nabla \cdot \left[I(\mathbf{x})^{-1} \nabla\psi(\mathbf{x}) \right] = \nabla^2\phi(\mathbf{x}), \tag{13}$$

which can be solved by means of the Green's function for the Laplacian operator. One typical method we mentioned here is derived by Paganin and Nugent [2], with the final solution of the phase taking the following form

$$\phi(\mathbf{x}) = -k\nabla^{-2}\nabla\left[\frac{1}{I(\mathbf{x})}\nabla\nabla^{-2}\frac{\partial I(\mathbf{x})}{\partial z} \right], \tag{14}$$

where ∇^{-2} is the inverse Laplacian operator, which is used as a simplified representation of the solution of the corresponding Poisson equation. Under simplified boundary conditions, for example, the periodic and the homogeneous Dirichlet/Neumann boundary conditions, the inverse Laplacian operator can be effectively implemented with use of fast Fourier transform (FFT) techniques [14]. Due to its simplicity and efficiency, this method has been recognized as a standard solution for the TIE and applied to a variety of matter- and radiation-wave fields in several other publications [6, 7, 10–12, 14]. Note for more general inhomogeneous Neumann boundary conditions, the Fourier transform should be replaced by the discrete cosine transform, and one additional hard-edged aperture is necessary to generate the required boundary signals [15]. In the following, we will assume simplified boundary conditions and limit our discussion to only the Paganin and Nugent's method, but please keep in mind the result of this work also applies to arbitrary boundary conditions.

Despite its widespread use, the "solution" given by Paganin and Nugent's method [Eq. (14)] is only an approximation to the exact solution, and there has been some research regarding its validity and accuracy [17, 18]. In this work, we refer this error as the phase discrepancy and take it to be equal to the difference between the true phase ϕ and the reconstructed phase ϕ_0 by Eq. (14) [corresponding to Eq. (13) of [17]]. It should be plain to see that the phase discrepancy originates from the Teague's assumption, since as we know from the Helmholtz decomposition theorem that the transverse flux should be decomposed in terms of the gradient of a scalar potential ψ and the curl of a vector potential η

$$\boldsymbol{j}_{\mathbf{x}}(\mathbf{x}) = -k^{-1}I(\mathbf{x})\nabla\phi(\mathbf{x}) = -k^{-1}\left[\nabla\psi(\mathbf{x}) + \nabla\times\eta(\mathbf{x})\right], \qquad (15)$$

Compared with Eq. (11), we see that the term $\nabla\times\eta(\mathbf{x})$ is ignored in Teague's assumption, making a silent hypothesis that the transverse flux is irrotational. One may argue that the vorticity of the transverse flux has been ruled out since we have already excluded consideration of the vector phase in Eq. (10). However, it must be noticed that in Eq. (15) the Helmholtz decomposition is performed directly based on $\boldsymbol{j}_{\mathbf{x}}$, instead of the $\nabla\phi$ as in Eq. (10). In general, the two decompositions will not be the same except for the trivial case that the intensity $I(\mathbf{x})$ is a constant (there would be neither phase vortices nor flux vorticity in this case). In another word, the rotational term $\nabla\times\eta$ is generally nonzero even when the phase is single-valued and continuous. Both the scalar and vector potentials can be respectively found by solutions of the following two Poisson equations, derived by taking the divergence/curl on both sides of Eq. (15) [17]

$$\nabla^2\psi(\mathbf{x}) = \nabla\cdot\left[I(\mathbf{x})\nabla\phi(\mathbf{x})\right], \qquad (16)$$

and

$$\nabla^2\eta(\mathbf{x}) = \nabla I(\mathbf{x})\times\nabla\phi(\mathbf{x}), \qquad (17)$$

where we used the identity $\nabla\times(I\nabla\phi) = \nabla I\times\nabla\phi$.

Next we examine how this missing rotational term will affect the phase reconstruction. When inserting the correct decomposition Eq. (15) into the TIE [Eq. (6)] we obtain the exact same equation as Eq. (12) because the rotational term $\nabla\times\eta(\mathbf{x})$ is divergence free. Consequently, the curl-free component of the transverse flux $\boldsymbol{j}_{\mathbf{x}}(\mathbf{x})$ can be reconstructed by solving this Poisson equation. The second Poisson equation, however, is different from Eq. (13), obtained by moving the $I(\mathbf{x})$ to the RHS of Eq. (15) and then taking the divergence on both sides

$$\nabla^2\phi(\mathbf{x}) = \nabla\cdot\left[I(\mathbf{x})^{-1}\nabla\psi(\mathbf{x})\right] + \nabla I(\mathbf{x})^{-1}\times\nabla\eta(\mathbf{x}). \qquad (18)$$

Here we used the identity $\nabla\cdot\left(I^{-1}\nabla\times\eta\right) = \nabla I^{-1}\times\nabla\eta$. Note that with certain (Dirichlet, Neumann, or periodic) boundary conditions, the uniqueness (or up to an arbitrary additive constant) of the solution for each Poisson equations in this work can be guaranteed. Consequently, it means the solutions to Eq. (13) and Eq. (18) are identical if and only if their source terms on the RHS are the same. Hence, we get the necessary and sufficient conditions for the Paganin and Nugent's solution [Eq. (14)] being an exact solution for the TIE:

$$\nabla I(\mathbf{x})^{-1}\times\nabla\eta(\mathbf{x}) = 0. \qquad (19)$$

The function $\eta(\mathbf{x})$ depends on both $I(\mathbf{x})$ and $\phi(\mathbf{x})$ as we shown in Eq. (17) and thus can be represented by $\eta(\mathbf{x}) = \nabla^{-2}\left[\nabla I(\mathbf{x})\times\nabla\phi(\mathbf{x})\right]$. Taking it into Eq. (19), we can further rewrite the necessary and sufficient conditions as

$$\nabla I(\mathbf{x})^{-1}\times\nabla^{-2}\left\{\nabla\cdot\left[\nabla I(\mathbf{x})\times\nabla\phi(\mathbf{x})\right]\right\} = 0. \qquad (20)$$

There are a few special cases with clear physical meanings when the above conditions are established. The first two trivial cases are uniform intensity (Teague's assumption is unnecessary) and constant phase (physically trivial). Another interesting special case is that $\nabla I(\mathbf{x})\times\nabla\phi(\mathbf{x}) = 0$, which is also the necessary and sufficient conditions for the validity of Teague's assumption [since $\eta(\mathbf{x})$ is defined by Eq. (17)]. This means $\nabla I(\mathbf{x})$ and $\nabla\phi(\mathbf{x})$ must be parallel, or equivalently, $\nabla\phi(\mathbf{x}) = c(\mathbf{x})\nabla I(\mathbf{x})$ for some scalar function $c(\mathbf{x})$. When $c(\mathbf{x})$ is taken as $\gamma I^{-1}(\mathbf{x})$ such that $\nabla\phi(\mathbf{x}) = \gamma\nabla\ln I(\mathbf{x})$, we have $\phi(\mathbf{x}) = \gamma\ln I(\mathbf{x})$. This is an equivalent definition of the Lambert–Beer's law of absorption [26] which states there is a logarithmic dependence between the transmission of light and the density or path length

through the material. Due to the inherent correlation between the intensity and phase, we could usually expect $\nabla I(\mathbf{x})$ and $\nabla\phi(\mathbf{x})$ to be nearly parallel for a typical transmissive object. More physical interpretations and examples regarding the validity of Teague's assumption can be found in [17]. However, generally speaking, the condition given by Eq. (20) cannot be satisfied for an arbitrary intensity and phase distribution, and therefore the Paganin and Nugent's method would not produce results that adequately match the exact solution of TIE.

The effects of the phase discrepancy have not yet been recognized as a significant problem since previously the TIE was mainly applied for weak absorbing specimens, like optical components, biological soft tissues, and cells. For these kind of objects, the phase discrepancy is negligible [second term on the RHS of Eq. (18) is close to zero] because the total intensity variation is small $\left\|\frac{\nabla I(\mathbf{x})}{I(\mathbf{x})}\right\|_2 \approx 0$, where $\|\cdot\|_2$ is the normalized L_2 norm, defined as

$$\|f\|_2 = \sqrt{\iint |f(\mathbf{x})|^2 \, d\mathbf{x}} \Big/ \sqrt{\iint d\mathbf{x}} \,. \qquad (21)$$

Actually in this case, the TIE can be directly reduced to a classic Poisson equation without the need of the Teague's axillary function

$$
\begin{aligned}
\frac{\partial I(\mathbf{x})}{\partial z} &= -\frac{1}{k}\nabla\cdot\left[I(\mathbf{x})\nabla\phi(\mathbf{x})\right] \\
&= -\frac{1}{k}\left[\nabla I(\mathbf{x})\cdot\nabla\phi(\mathbf{x}) + I(\mathbf{x})\nabla^2\phi(\mathbf{x})\right] \qquad (22) \\
&\approx -\frac{1}{k}I(\mathbf{x})\nabla^2\phi(\mathbf{x}).
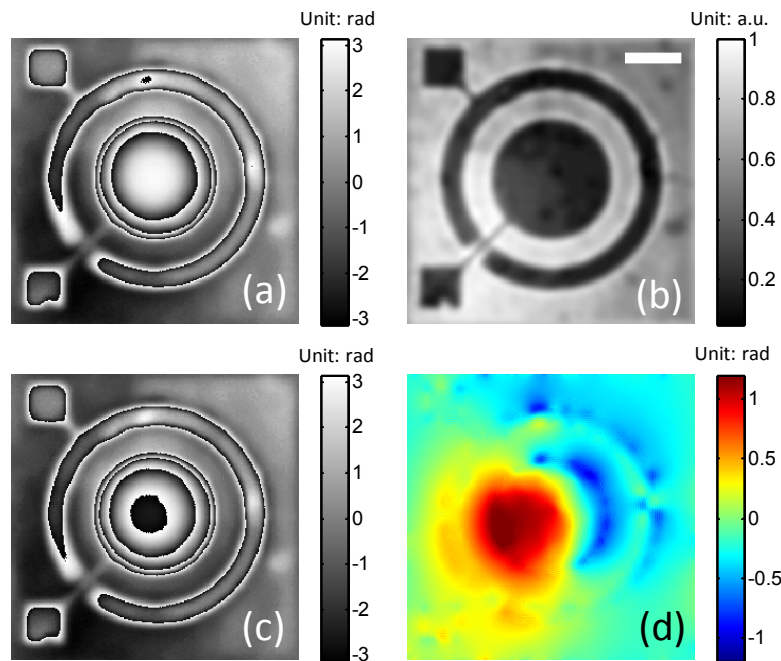\end{aligned}
$$

Fig. 1. Phase retrieval of a strongly absorbing object (MEMS diaphragm) using the Paganin and Nugent's solution. (a) Phase distribution. (b) Intensity distribution. (c) Phase retrieved by the Paganin and Nugent's solution (d) Phase discrepancy. Scale bar 500 μm

However, for a strongly absorbing specimen (like a reflective object, whose intensity and phase are generally uncorrelated), the effects of the phase discrepancy can be easily observed.

Figure 1 shows a realistic example of the error arising from a simulated reflective object (a piezo-attenuated circular micro-membrane diaphragm), reconstructed using the Paganin and Nugent's method. The phase and intensity of the object is shown in Figs. 1(a) and 1(b) respectively. The minimum value of the intensity is only 0.05 due to the strong absorption of the object. The simulation is designed to replicate the real measurement from the data obtained by a digital holographic microscope. Figure 1(c) shows the retrieved phase by the Paganin and Nugent's solution. Note the phase recovered by the TIE is continuous without 2π jumps, and the phase shown in Fig. 1(c) is digitally rewrapped into the interval $[-\pi, \pi)$ for better comparison. The phase discrepancy (the difference between the ground truth phase [Fig. 1(a)] with the reconstructed phase [Fig. 1(c)] is shown in Fig. 1(d). The relative root-mean-square error (RMSE) of the reconstruction is 6.76%. This example obviously indicates that the phase discrepancy can be a practical problem when dealing with absorbing objects with use of the TIE, which has been generally acknowledged to be quantitative.

4. Iterative algorithm for phase discrepancy compensation

We now proceed to develop a simple Picard-type iterative algorithm [27] to compensate the phase discrepancy owing to Teague's assumption. Assuming we have already got an inaccurate solution $\phi_0(\mathbf{x})$ by Paganin and Nugent's method, the Helmholtz decomposition theorem states that

$$I(\mathbf{x})^{-1}\nabla\psi(\mathbf{x}) = \nabla\phi_0(\mathbf{x}) + \nabla\times\eta_0(\mathbf{x}), \qquad (23)$$

where η_0 is a vector potential function. Equation (23) suggests that this inaccurate solution ϕ_0 is just the scalar potential for the function $I^{-1}\nabla\psi$. By properly rearranging Eq. (23), the transverse flux associated with this inaccurate phase solution can be written as

$$\mathbf{j}_\mathbf{x}^0(\mathbf{x}) = -k^{-1}I(\mathbf{x})\nabla\phi_0(\mathbf{x}) = -k^{-1}\left[\nabla\psi(\mathbf{x}) + I(\mathbf{x})\nabla\times\eta_0(\mathbf{x})\right]. \qquad (24)$$

Compared it with Eq. (15), it is clearly that the second terms on the RHS are different: the term $I(\mathbf{x})\nabla\times\eta_0(\mathbf{x})$ in Eq. (24) is not divergence free in general. Consequently, it will manifest itself by changing the longitudinal intensity during the wave propagation. In simple words, if we push this inaccurate solution $\phi_0(\mathbf{x})$ back to the RHS of the TIE, the resultant intensity derivative (we call it artificial intensity derivative) will be inconsistent with the real measurement. This inconsistency in intensity derivative can be regarded as the error signal which must be eliminated in order to obtain an accurate solution to the TIE. As a consequence of the linear nature of the TIE, the phase discrepancy can be obtained by solving the TIE once again with the error signal as the source term. However, due to the inaccuracy of the TIE solution owing to Teague's assumption, the phase discrepancy usually cannot be eliminated in one-step, but can be iteratively removed by repeating this process as a loop: the correction term $\Delta\phi_0(\mathbf{x})$ is added back to the previous estimate of the phase for the next round $\phi_1 = \phi_0 + \Delta\phi_0$. The iterative process stops when the RMS of the correction term $\Delta\phi_n$ is below a pre-assigned accuracy level ε. The block diagram of the overall algorithm is summarized in Fig. 2.

To verify the convergence of the proposed method, it is sufficient to prove that the sequence of functions $\Delta\phi_n$ converges uniformly to zero. In practice, the rigorous proof of the convergence is not possible without some *a priori* information. Here we will take a conservative stance and assume that the Paganin and Nugent's method is a *valid* (though not accurate) solver for the TIE, which means the phase discrepancy should *at least smaller than* the exact solution in norm: $\|\phi - \phi_0\|_2 \leq L\|\phi\|_2$, with constant $L < 1$. By analogy, we can derive $\|\phi - \phi_{n+1}\|_2 \leq L\|\phi - \phi_n\|_2$, which means the phase discrepancy can only decrease at each iteration. Besides, since $\|\phi - \phi_{n+1}\|_2 \leq L^n\|\phi - \phi_0\|_2$, and $L^n \to 0$ as $n \to \infty$, the compensated phase function is shown to be a Cauchy sequence and thus it converges to the exact solution. In the ideal noise-free condition, the above iterative algorithm usually converges very rapidly and the error decreases exponentially with the iterations. Within two to four iterations the phase discrepancy can be reduced to a negligible level and the exact solution to the TIE can be thus

obtained. However, in a real measurement the correction term can never reach a very low level because the noise, finite sampling effect, and the nonlinearity error in the finite difference estimation of the axial intensity derivative will introduce additional errors. Thus, in practice, the iteration should stop when there is no significant decrease in the RMS of the correction term $\Delta\phi_n$ as the iteration number increases. Besides, the maximum iteration number can also be set as an additional termination condition to prevent unnecessary ineffective iterations.
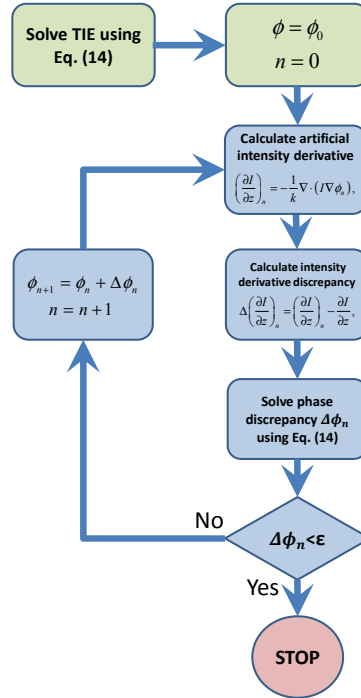
Fig. 2. Block diagram of the iterative compensation method.

5. Simulation results

5.1 Smooth wavefront

Several simulations were performed to test the proposed method. First, we adopted the "counter example" suggested in [17] to verify whether the proposed approach can effectively compensate the phase discrepancy even with sufficiently large scale. In the simulation, the complex field was defined on an 256 × 256 grid with 0.025μm × 0.025μm pixel size. The wavelength is 632.8nm (HeNe Laser). The intensity derivative $\partial I/\partial z$ is estimated by central finite difference between two defocused images I_+ and I_- with the distance of $\Delta z = \pm 10$ μm generated by the angular spectrum numerical propagation. As shown in Fig. 3, the phase and intensity functions are defined as

$$I(x, y) = \exp\left(-a_0 x^2 - b_0 y^2\right), \tag{25}$$

and

$$\phi(x, y) = a_0 x^2 - b_0 y^2 - a_1 x^8 + b_1 y^8, \tag{26}$$

where $a_0 = b_0 = 10^{-2}\mu m^{-2}$ and $a_1 = b_2 = 0.25^{-2} \times 10^{-8}\mu m^{-8}$. The metric RMSE was used to quantify phase retrieval accuracy. Since the TIE can only determine the phase uniquely up to

an additive constant, the piston term (the mean value) of each result was removed before the RMSE evaluation.
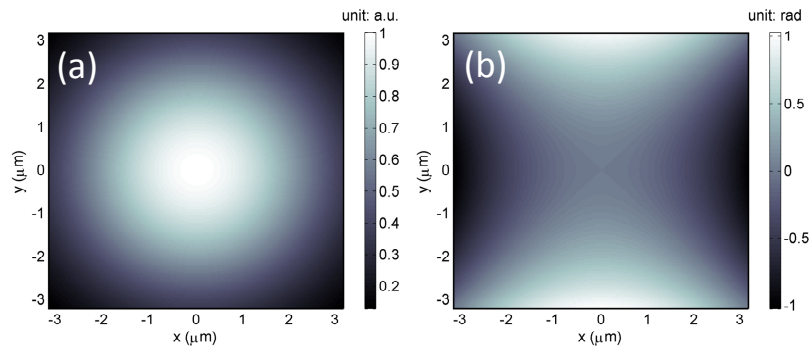

Fig. 3. Simulated intensity and phase distributions. (a) Intensity distribution defined by Eq. (25). (b) Phase distribution defined by Eq. (26).

To illustrated the effects of the Teague's assumption in the phase reconstruction, the Helmholtz decomposition was performed based on the transverse flux field $I\nabla\phi$ (neglecting the constant $-k$) and the results are shown in Fig. 4. The butterfly-like transverse flux field can be decomposed into a gradient field and a rotational field. The Teague's assumption simply neglects the rotational field, which obviously is comparable in norm to the gradient field. As a result, a large phase discrepancy (RMSE 9.53%) appeared when the Paganin and Nugent's solution was used [Eq. (14)], as demonstrated in Fig. 5(a). With the proposed compensation process after the first iteration, the phase RMSE reduced significantly to around 0.99%, as shown in Fig. 5(b). The error was even down to a negligible level - around 0.016% after the third iteration, as shown in Figs. 5(c). The evolution of the RMSE values versus the iteration times is shown as one-line profile in Fig. 5(d), which clearly illustrates the rapid convergence of the proposed approach.
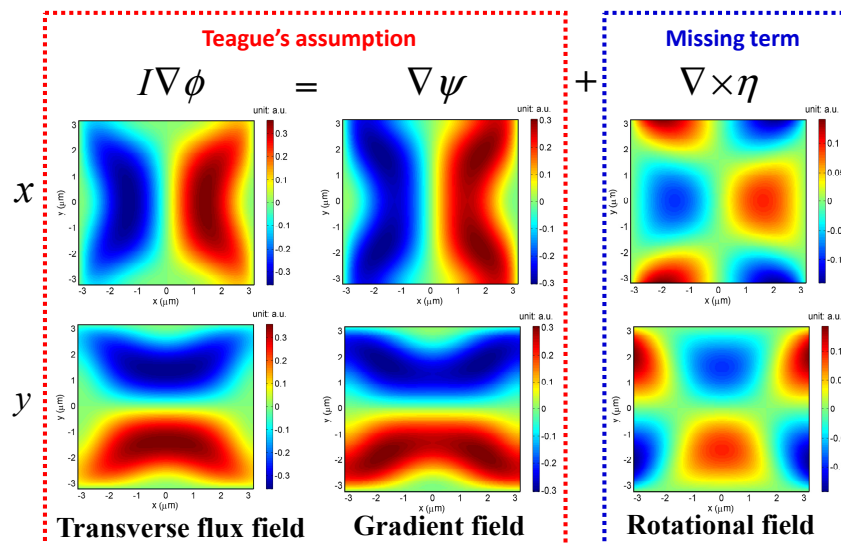

Fig. 4. Helmholtz decomposition of the transverse flux field. The x and y components of the vector fields are shown in the first row and the second row, respectively.
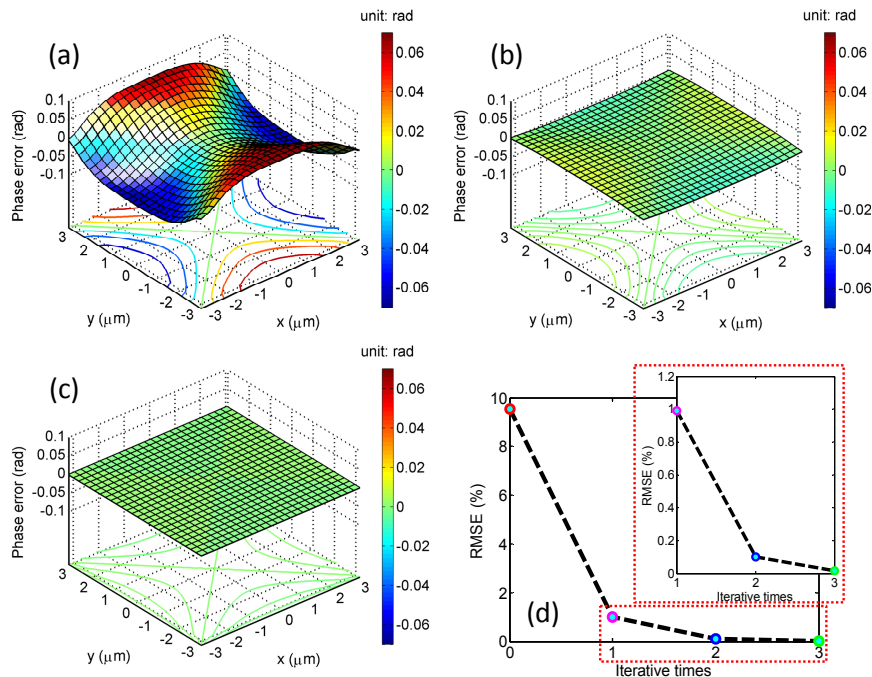
Fig. 5. Simulation results of the iterative compensation algorithm for a smooth wavefront. (a) Phase error of the Paganin and Nugent's solution before compensation (RMSE 9.53%). (b) Phase error after the first iteration of the compensation (RMSE 0.99%). (c) Phase error after the third iteration of the compensation (RMSE 0.016%). (d) One-line profile of the RMSE verse the iteration times of the proposed compensation approach.

In a practical TIE measurement system, noise exists on each captured intensity images. To test the performance of the proposed method in the presence of noise, different levels of pseudo-random Gaussian noise with standard deviations (STD) of 10^{-5}, 10^{-4}, and 10^{-3} was generated and superimposed on each intensity image. In such scenario, the phase error was not only resulted from the Teague's assumption but also from the influence of the noise. For the low-noise-level case, the underlying saddle-like structure in the phase error of the Paganin and Nugent's solution before compensation can be clearly perceived [Fig. 6(a)]. When our method was applied, the majority amount of phase discrepancy could be compensated after three iterations, with the residual phase error shown in Fig. 6(b). The iterative process successfully removed the large waviness due to the phase discrepancy, leaving only the cloud-like random errors [28, 29] from the intensity noise. However, with the increase of the noise level, the noise effect became dominant. The initial phase errors were almost submerged by the cloud-like artifacts, showing no evident sign of the phase discrepancy (Fig. 7). Their corresponding RMSE curves only decreased slightly and could never go down to very small values due to the nonnegligible noise.

To verify whether our method can effectively remove the phase discrepancy in the presence of significant noise, we also show the correction term by our method (which is the difference between the initial solution and the final refined solution after 4 iterations) in Fig. 7. For the medium noise level, the correction term closely resembles the phase discrepancy in the noise-free case. While for the high noise level, the correction term was distorted and no longer matches the noise-free phase discrepancy, which can be expected since the noise has already changed the basic structure of intensity distribution fundamentally. These results clearly confirm the validity and the effectiveness of the compensation algorithm in the presence of noise.
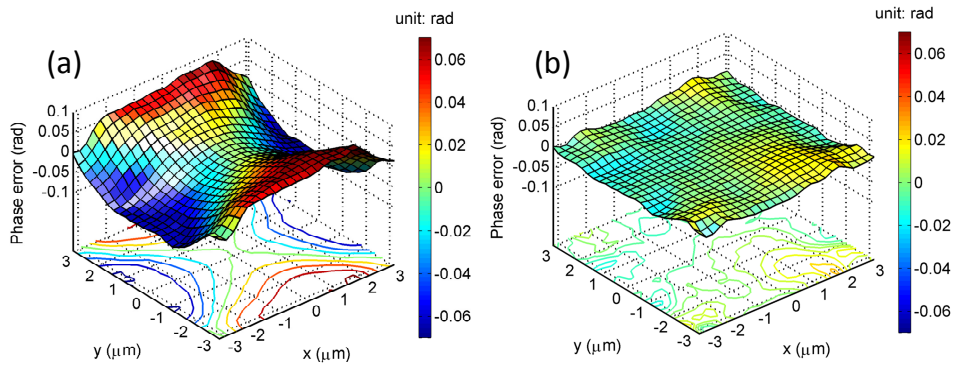
Fig. 6. Simulation results of the compensation algorithm with the existence of noise. (a) Phase error of the Paganin and Nugent's solution before compensation (RMSE 10.64%). (b) Phase error after three iterations of the compensation (RMSE 1.90%).
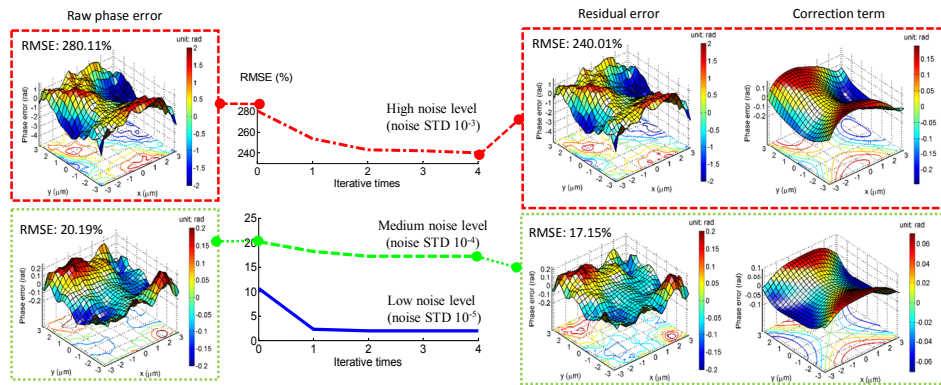

Fig. 7. RMSE versus iterative number under different levels of noise. The red/green boxes outline the raw phase errors, the residual errors after compensation, and the corresponding correction terms for the cases of high/medium noise levels, respectively.

5.2 Complex phase and intensity distribution

In the previous simulation, the complex field was constructed with very simple and smooth mathematical functions. Next, we focus on more complex phase and intensity distributions. We consider the complex intensity distribution of letters 'COLE' (Fig. 8(a)) and the phase profile of letters 'TIE' (Fig. 3(b)). The value range of the intensity is [0.05, 1]. For such complex object with a lot of high spatial frequency details, the effect of the finite sampling in both spatial (x-y) and axial (z) dimensions must be given additional attentions.
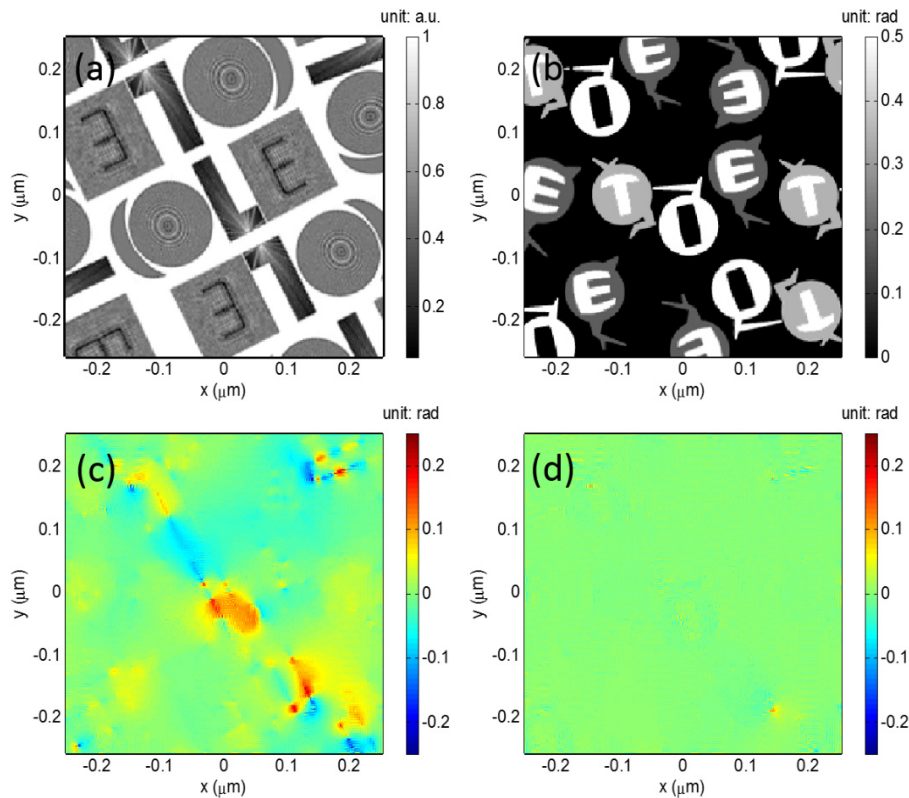
Fig. 8. Simulation result of complex intensity and phase distribution. (a) Intensity distribution. (b) Phase distribution. (c) Phase error of the Paganin and Nugent's solution before compensation (RMSE 15.91%). (d) Phase error after five iterations of the compensation (RMSE 3.66%).

For the *x-y* dimension, the Shannon theorem indicates that for the correct acquisition of a given 2D signal, the sample frequency must be at least twice as high as the highest signal frequency of interest [30]. For the axial direction, the situation is a bit more complicated since the axial intensity derivative $\partial I/\partial z$ in TIE is calculated via finite differences from two (or more) defocused intensities. When the separation between the two measurement planes was sufficiently small ($\Delta z = \pm 10$ μm), the effect of the finite sampling was not evident. Figure 8(c) shows the phase discrepancy of the Paganin and Nugent's solution before compensation (RMSE 15.91%). After five iterations of compensation, the majority amount of phase discrepancy was removed, with the residual phase error shown in Fig. 8(d). However, for relatively larger Δz (usually required for high signal-to-noise measurements [29, 31, 32]), the breakdown of the linear approximation that underlies the finite difference approximation induces nonlinear errors, and the estimate $\partial I/\partial z$ does not accurately represent the axial intensity derivative. Typically, this results in a loss of resolution of the retrieved phase map. It can be clearly seen from Fig. 9 that object edge information (high frequency details) appears in the phase error maps when a larger defocus distance is used ($\Delta z = \pm 100$ or ± 500 μm). When our compensation method was applied, phase discrepancies were effectively compensated, leaving only the edge error resulting from the nonlinearity errors. In all the three cases, the compensation procedure converges after 4 to 5 iterations and does not show significant improvement with more subsequent iterations. These results verify the stability of our algorithm under large defocus distances and nonlinearity errors.
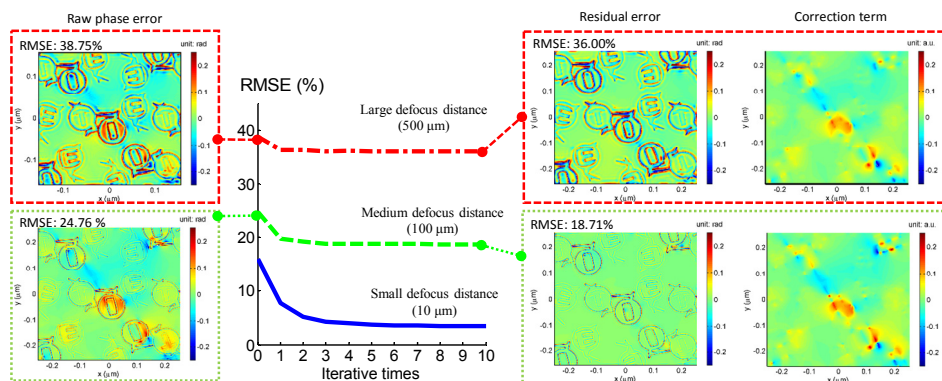
Fig. 9. RMSE versus iterative number with different defocus distance. The red/green boxes outlines the raw phase errors, the residual errors after compensation, and the correction terms for the cases of medium /large defocus distance, respectively.

5.3 A realistic example and some practical considerations

Finally, let us turn to the realistic MEMS sample shown in Fig. 1. The value range of the intensity is [0.05, 1] and the RMSE of the Paganin and Nugent's solution before compensation is 6.76%. Once again, our approach converges to the exact solution (blue curve) with very small residual error (RMSE 0.119%), and the final compensated phase closely matches the ground-truth image (Fig. 10).
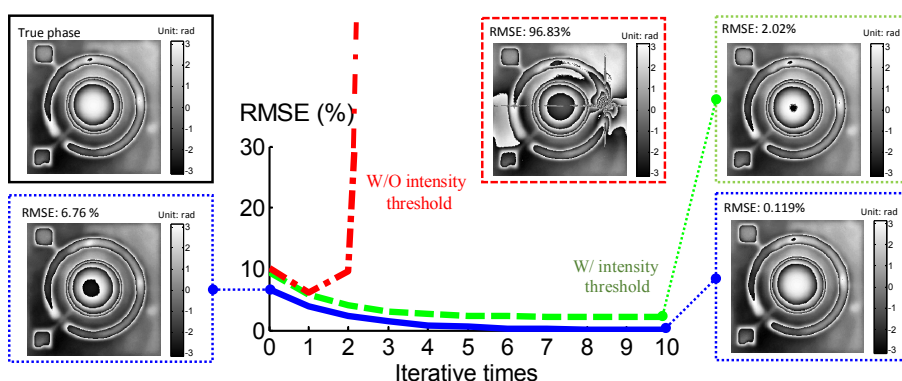

Fig. 10. RMSE versus iterative number for a realistic MEMS example. When the value range of the intensity is [0.05, 1], our approach can converge to the exact solution (blue curve). When the minimum intensity is close to zero (0.001), the iterative process will diverge (red curve). This problem can be addressed by setting an intensity threshold (green curve).

Before concluding, several important issues in the practice of our approach need to be clarified. Although no rigorous proof of convergence of our approach has been devised, the process did converge in all above cases. This seems reasonable because the iterative process is essentially self-correcting. In the presence of noise and nonlinearity error, our algorithm effectively extracts and removes the phase discrepancy without noticeable disturbance by other types of errors. However, as we mentioned at the end of Section 4, the stability and convergence of our method depend heavily on the accuracy of the Paganin and Nugent's solution. Since the intensity I appears in the denominator [Eq. (14)], the Paganin and Nugent's solution becomes quite unstable when the minimum intensity is close to zero. Therefore, caution is required when very small values appear in the intensity distribution. On the other hand, these small intensity values also increase the phase discrepancy accordingly since the total intensity variation $\left\|\frac{\nabla I(\mathbf{x})}{I(\mathbf{x})}\right\|_2$ can become much larger. For example, if we rescale the intensity range of the MEMS sample

to [0.001, 1], the iterative process will be out of control after two iterations (red curve). The singular point creates significant artifacts that degrade the whole phase reconstruction prevailingly. In this limiting case, we should avoid dividing too small values to guarantee the stability of solution by setting a certain intensity threshold (e.g. 1% of the maximum value of I) when evaluating the term in the square brackets in Eq. (14). When the intensity threshold (0.01) is used, our method becomes convergent again and finally reaches a reasonable accuracy (RMSE 2.02%).

5. Conclusions

To conclude, this work has presented the theoretical analysis of the phase discrepancy in the most widely applied FFT-based solution of the TIE owing to the Teague's assumption. The necessary and sufficient conditions for the solution to coincide with the exact phase have been derived. We also have proposed a simple and novel iterative algorithm to compensate the phase discrepancy due to the Teague's assumption when such conditions cannot be satisfied for a general object. Several numerical results have been presented showing the effectiveness of the proposed algorithm.

A clear advantage of the proposed method resides in the very rapid convergence of the iterative phase corrections. The phase error drops exponentially with iterations and shows no sign of stagnation for a wide variety of test cases. Moreover, the convergence is insensitive to the intensity noise and the separation between the input image planes. Probably the most limiting characteristic of the method is its instability to the intensity singularity, a feature inherited from the FFT-based TIE solvers. This problem can be countered by setting a certain intensity threshold. Since generally only 2 to 4 iterations are sufficient to compensate most phase discrepancy and thanks to the efficient FFT-based operations, the iterative nature of our method does not preclude its application to high-speed, real-time phase imaging. It should be noted that although the phase discrepancy problem can be avoided by using some other TIE solvers that do not involve the use of Teague's assumption, like the multigrid method [14, 33], our approach can be still considered as an attractive solution for its better spatial resolution (FFT-based solvers do not use finite difference to approximate the spatial derivatives) [34], simplified treatment for boundary conditions [15], and easy implementation.