Phase-retrieval exact solution based on window modulation

Jianhui Huang\textsuperscript{1,2}, Thomas. T. Gaylord\textsuperscript{2}, An Pan\textsuperscript{3}, Huiliang Jin\textsuperscript{1}, Guoxiang Meng\textsuperscript{1}, and Qian Ye\textsuperscript{1,*}

\textsuperscript{1}School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China, 200240.  
\textsuperscript{2}School of Electric and Computer Engineering, Georgia Institute of Technology, Atlanta, US, 30308.  
\textsuperscript{3}Pioneering Interdiscipline Center (PIC), State Key Laboratory of Transient Optics and Photonics, Xi’an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi’an 710119, China  
\textsuperscript{*}yeqian@sjtu.edu.cn.

Abstract

Quantitative phase imaging (QPI) is a promising tool for imaging complex objects. By combining self-reference interferometry and phase retrieval, this paper proposes a general exact QPI for arbitrary complex objects as well as a one-shot exact QPI for transparent objects with small phase range (i.e., weak scattering object). The optical configuration is similar to the Zernike phase contrast, while the phase-shifting area is a little smaller. With three frames of phase-shift, the algebraic relationship between the phase and the measured intensity is built, providing an excellent approximate phase recovery. Then, an efficient iterative optimization strategy is reported to turn that approximate solution into an exact one. The iteration exhibits linear convergence property. Thus exact phase map is achieved with a sufficient number of iterations. When the object is a pure phase sample with a small phase ranges of $[0, \pi]$, a single intensity measurement will be enough for exact phase recovery by selecting a tiny phase shifting value, based on a similar linear-convergence iterative algorithm. The feasibility and accuracy of our method are verified by both numerical simulations and experiments on diverse phase objects.

\textit{Key words:} quantitative phase imaging, interferometry, phase retrieval, phase contrast
1. Introduction

Quantitative phase imaging (QPI) has a wide range of applications in many fields, such as X-ray crystallography [1], antenna measurement [2] and adaptive optics [3], and biomedical imaging [4]. Especially in the field of biomedical imaging, it has been attracting increasing attention and making continuous progress [4-6]. It is a promising tool in label-free and stain-free cell imaging [7], three-dimensional tissue tomography [8-10], neuroimaging [11], etc. In these fields, the structure information of the tested object is carried mainly by the phase, while the phase cannot be measured directly by current technology. Phase retrieval is developed to solve this problem, where the hidden phase information is recovered by digitally computing from the coded intensity patterns, based on a specific optical configuration and the corresponding relationship (decoding method) between the unknown phase and the measurable intensity. There are mainly four types of QPI techniques: external-reference interferometry (ERI), such as digital holography [12], phase-shift interferometry [13], and wavefront-divisional digital holography [14] and its variations based on diverse techniques [15, 16]; Self-reference interferometry (SRI), such as Zernike phase contrast (ZPC) [17], generalized phase contrast (GPC) [18] differential interference contrast (DIC) [4], diffraction phase microscopy [4], quadriwave lateral shearing interferometry [20] and phase-shifting phase contrast [21]; Transport of intensity equation (TIE) based deterministic near-field phase retrieval [22] and its generalized forms such as speckle contrast phase imaging [23-25]; Phase retrieval from multi-intensity measurements based on diffraction principle [26-29]. In particular, if the object has a small phase range, then based on first-order Taylor linearization, there obtain some closed-form analytical solutions [30-32].

In the view of accuracy, strictly speaking, most of the above QPI methods have more or fewer difficulties in achieving exact phase reconstruction. Since the defocus of
TIE cannot be zero, the difference approximation inevitably brings in error. For the iterative phase retrieval, the iteration falls stagnant quickly as the problem is severely nonlinear and non-convex, and therefore we can obtain a near-optimal solution at most. For SRI, the estimation on the reference wave cannot be completely accurate as it depends on the unknown object, leading to an approximate phase recovery. Comparatively, the phase extraction by ERI can be exact theoretically. Still, it requires an invariant reference wave as well as a stable environment, which becomes the main drawback of this method. This study aims to develop an exact phase retrieval solution to SRI so that an exact QPI can be achieved under general environments. Generally, the phase recovery by SRI always has a high accuracy where the error lies in that the reference wave is not well evaluated. Note that in the iterative phase retrieval, the phase recovery is improved continuously during the iteration before falling stagnant. The stagnation is a result of a poor initial phase guess. Considering the phase recovery of SRI could be an excellent initial phase for the iterative phase retrieval, perhaps the exact QPI could be obtained if we apply iterative optimization to the direct phase recovery of SRI. Based on the concept, we proposed an efficient phase retrieval exact solution by combining SRI and iterative optimization, so that an exact QPI is obtainable without reference wave.

Taking a typical 4f phase-contrast imaging system shown in Fig. 1 as the example, perform a phase shift on the low-frequency portion of the object’s spectrum on the focal plane three times. Meanwhile, three intensity patterns are detected correspondingly at the output plane, each of which reflects the phase at varying degrees. The flat output wave formed by the object spectrum within this small phase-shifting window serves as the reference wave. Since it is related to the unknown object, the reference wave is also unable to be determined. There are many means to give a good estimation. A primary
way is considering it as zero, which is taken in most self-reference interferometry such as PDI and DPC. Gluckstad has studied this problem profoundly and proposed a general high-precision self-referenced interferometry, i.e., GPC [33]. Based on an excellent approximation to the object’s central spectrum under ideal diffraction hypothesis, the reference wave amplitude is fitted in high accuracy. Therefore, a near-optimal QPI is achieved [18]. In this study, according to the trigonometric identity carried by the measured intensities, an analytical solution of the reference wave amplitude is derived which is more general compared to GPC as the illumination style is out of consideration. After the object is reconstructed, the approximate reference wave can further be corrected by alternative projection, which is a loop calculation formed by continuous forward propagation and backward propagation of the light field. In the case of a small phase-shifting window, the iteration optimization exhibits a linear converges, which means the reference wave can be solved precisely; meanwhile, the exact QPI of the object is also achieved. On this basis, when the object is moreover phase-only with a small phase range, the above exact QPI technique using three phase-shifts can be simplified into a one-shot one. In this way, a single phase-shift with a small phase shifting value as well as a single intensity measurement is sufficient to obtain the exact object reconstruction.

In this paper, the rigorous theoretical derivation and algorithm are given in Section 2, the performance and characteristics of the method are studied through simulation in Section 3, the phase recovery performance of the method on diverse phase objects is investigated by experiments in Section 4. Besides, the research is an extension of one of our previous study [34], which proposes an efficient QPI method for coherent diffraction imaging. They share many theoretical similarities, especially on the establishment of the algebraic phase retrieval solution; hence a detailed comparison is
Fig. 1. Theoretical and experimental schematic diagram of the exact phase retrieval technique. The window-area for phase-shifting should be small enough. A commercial laser of 532-nm wavelength acts as the light source. Two SLMs (labelled as SLM #1 and SLM #2) are applied to load phase object and phase modulation, respectively. A grey-scale CMOS is placed at the end of the 4f imaging system to record the intensity image.

2. Method

A. Problem setup and definition

As illustrated in Fig. 1, this study discusses a kind of non-reference exact phase imaging method taking a 4f phase contrast imaging as the platform. It involves three devices, a laser source, two SLMs and a CMOS image sensor. Produced by the laser, the light beam is filtered by a pinhole to be smooth and slightly phase-distorted, collimated by a lens into a parallel beam. It is finally shaped by a diaphragm pupil into a circle. On the object plane, a 2D phase object is loaded on SLM #1 in the format of a grayscale image. Another SLM, labelled as SLM #2, is placed on the focal plane and plays the role of shifting the phase of the object’s Fourier spectrum within a small central window. At the end of the 4f imaging system, a CMOS image sensor is arranged to record the intensity distribution. In general, the phase object can be defined as a complex function
as $Ce^{iE}$, where $C$ is the amplitude, and $E$ is the phase. Similarly, its Fourier spectrum is defined by $he^{i\alpha}$, the phase modulation is defined by $e^{itw}$, where $w$ indicates the phase-shifting window area, and $t$ is the phase-shifting value in radian. The optical field on the image plane is defined by $Ue^{i\nu}$, whose intensity is defined by $I$, $I = U^2$. For the illumination laser beam, it is assumed that there is no phase aberration, so it can be written as a real function $B$, indicating the amplitude distribution. In this paper, the functions on the object plane and output plane are defined on the space domain $(x, y)$, and the functions on the focal plane are defined on the spatial frequency domain $(k_x, k_y)$. For a clear distinction, functions in the spatial domain are written in capital, and functions in the spatial frequency domain are written in lowercase. Besides, the coordinates are omitted for simplicity. It is reasonable to regard the background amplitude and the object amplitude as a whole, which is called the new object amplitude, which is defined by $\tilde{C} = BC$. Suppose the lens acts as an ideal Fourier transform, the captured image by the CMOS can be modelled as

$$Ue^{i\nu} = \mathcal{F}\left(\mathcal{F}\left(\tilde{C}e^{iE}\right) \cdot e^{itw}\right) = \mathcal{F}\left(hhe^{i\alpha} \cdot e^{itw}\right), \tag{1}$$

where the symbol $\mathcal{F}$ and $\mathcal{F}^{-1}$ refers to 2D Fourier transform and 2D inverse Fourier transform operations, respectively. Expand the phase modulation term $e^{itw}$ as $e^{itw} = e^{itw} + (1 - w)$, and hence Eq. (1) is simplified into $Ue^{i\nu} = \tilde{C}e^{iE} - Ke^{ip} + e^{it}Ke^{ip}$, where the complex function $Ke^{ip}$ is defined as the Fourier transform of $he^{i\alpha}w$, which is, in fact, the output field when blocks the focal plane spectrum with the window $w$. Calculate the intensity on both sides, and we obtain

$$I (t) = \tilde{C}^2 + (2\cos t - 2)\left[\tilde{C}K \cos (E-P) - K^2\right] + (2\sin t)\tilde{C}K \sin (E-P). \tag{2}$$
This equation establishes the algebraic relationship between the output intensity pattern $I(t)$ and the phase-shifting values $t$. For a particular case that $t = 0$, it means no phase modulation is applied on the object’s Fourier spectrum so that the measured intensity pattern is precisely the object’s intensity $I = \bar{C}^2$. Hence, it is usually easy to recover the amplitude. In contrast, it seems that no general solution exists to recover the phase from one-shot intensity measurement. A widely adopted strategy is taking more phase shifting values. In this paper, we focus on establishing a kind of phase retrieval exact solution based on multi-value small window phase shifting.

**B. Algebraic phase retrieval**

Similar to the establishment of far-field phase retrieval algebraic solution described in our previous research [34], here we also conduct three phase-shifts with different values of $t$, and hence there forms a determined system of linear equations

$$\begin{bmatrix} I(t_1) \\ I(t_2) \\ I(t_3) \end{bmatrix} = \begin{bmatrix} 1 & 2\cos t_1 - 2 & 2\sin t_1 \\ 1 & 2\cos t_2 - 2 & 2\sin t_2 \\ 1 & 2\cos t_3 - 2 & 2\sin t_3 \end{bmatrix} \begin{bmatrix} \bar{C}^2 \\ \bar{C}K \cos (E - P) - K^2 \\ \bar{C}K \sin (E - P) \end{bmatrix}. \quad (3)$$

In this relationship, the intensity measurements and the phase-shifting values are known, while the functions on the right column vector, $\bar{C}, K, E$ and $P$, are remaining to be determined. Record the coefficient matrix as $T$. It can be proved that, in the case of inputting three $t$ whose values are all different from each other, $T$ is always invertible. Therefore, the column vector on the right side can be determined uniquely through matrix inversion. Record the element functions on the first, second and third row of the solved column vector as $R_1, R_2$ and $R_3$, respectively. As $R_1$ is just the object intensity, the object amplitude $\bar{C}$ is recovered directly. For the object phase $E$, it is seen the
solution should be

\[ E = P + \arctan \frac{R_3}{R_2 + K^2}. \]  \tag{4}

To reconstruct the phase exactly, one should obtain the solution of \( P \) and \( K \) at first. According to the trigonometric identity \( \sin^2 \theta + \cos^2 \theta = 1 \), an indefinite solution to \( K \) is obtained, recorded as \( K_1 \), i.e.,

\[ K_1 = \frac{1}{2} \left( R_1 - 2R_2 \pm \sqrt{(R_1 - 2R_2)^2 - 4(R_2^2 + R_3^2)} \right). \]  \tag{5}

However, we do not know whether it should be either ‘+’ or ‘−’ for the symbol ‘±’ in the equation. It should be ‘+’ at some coordinates and ‘−’ elsewhere. As it depends not only on the window \( w \), but also on the unknown object itself, Eq. (5) cannot be used directly as the solution of \( K \). Then, we discuss the production of \( K \) and \( P \) in the view of signal processing. On our technical settings, the small window \( w \) is located in the middle (or near middle) of the focal plane, allowing the pass of only the object’s spectrum with low spatial frequency and blocking the high spatial frequency part that lies outside the window. In other words, only the low spatial frequency signal remains to participate in the formation of \( Ke^{iP} \). Therefore, it is predictable that the generated amplitude \( K \) will be very smooth, and the generated phase \( P \) will be very flat. This inference reveals that Eq. (5) is not suitable for solving \( K \), because \( K_1 \) is surely unsmooth, contrary to the fact that \( K \) has a smooth shape. Moreover, the signal inside the window comes from the central diffraction spot, its smooth shape and small phase distortion makes \( K \) smoother, and \( P \) flatter (even close to a constant). Following this inference, the shape of \( K \) can be approximated by the following formula, recorded as \( K_2 \), i.e.,
\[ K = \left| \mathcal{F}(h e^{i\eta w}) \right| \propto \left| \mathcal{F}(w) \right| = K_2. \]  (6)

Considering that \( K_1 \) is exact in the scale though it is indefinite while \( K_2 \) is very precise in shape, this study provides a combination use of \( K_1 \) and \( K_2 \) to achieve a new approximation to \( K \) with higher accuracy, i.e., \( K \approx K_2 \cdot K_1^{\text{max}} / K_2^{\text{max}} \), where the superscript ‘max’ means taking the maximum. In our previous research [34], it is found \( K \) does not need to be very accurate, because it has few effects on the phase recovery, especially when there is subsequent iteration optimization. For example, in practical experiments, we could even consider it as 0 to simplify the solution process. This operation is also suitable for this study. Now that \( K \) has been well approached and \( P \) is intrinsically flat and is considered as 0, an algebraic solution to the object phase \( E \) can be achieved through Eq. (4).

The above is the theoretical establishment of the algebraic phase retrieval solution. The core of this method is phase modulations on the object’s spectrum within a small window. The window should be sufficiently small to ensure that the reference phase \( P \) is close to 0 and hence become a known condition. Generally, a smaller window aids in achieving a more precise phase imaging. Based on these characteristics, it is reasonable to consider the proposed method is also a kind of common-path interferometry where the reference wave is produced by itself. On the other hand, this method shares some similarities in the theoretical basis with the previous research [34]. To investigate their relationship, a detailed comparison between this spectrum-modulated algebraic phase retrieval and the previous object-modulated algebraic phase retrieval solution is made and summarized in the discussion section.

The validity of the algebraic phase retrieval solution is verified by experiments on detecting both smooth phase object and rough phase object. Sometimes there is non-
negligible phase aberration of the background illumination which will be superposed into the final object’s phase recovery. A practical solution is to test the background using the same technique before the experiment. Then in the reconstruction to the object, remove the background by dividing by the amplitude and subtracting the phase to achieve a more accurate result. This approach is proved feasible by an experiment in detecting a vortex phase object.

C. Exact phase retrieval

In previous perceptions, it is widely considered that the accurate phase reconstruction of an unknown object under test can be achieved only when there is an external stable reference wave, which is known as classical interferometry. In this study, this stereotype will be broken by the proposed exact phase retrieval solution. At present, the algebraic phase retrieval solution expressed by Eq. (4) gives us an approximate phase detection. By applying a powerful iterative optimization strategy to that approximate solution, it is found the iteration exhibits a fast and stable convergence. Moreover, it shows a linear convergence (first-order convergence). That is, we could always obtain an exact phase reconstruction as long as the iteration is enough. In practical, ten iterations are adequate to get a good phase recovery generally, with the error being primarily reduced.

The iteration optimization algorithm is described in the table below. It involves two sections. The first section is the acquisition of a good approximate phase retrieval solution, which is obtained already by developed algebraic phase retrieval method. The second section is the improvement to the approximate solution with a high-efficiency iteration optimization algorithm. The optimization strategy is based on the alternative projection [26]. By propagating the optical field forward and backward ceaselessly and replacing the calculated intensity with the measured one; meanwhile, the phase
becomes accurate gradually. For clarity and completeness, the phase retrieval exact solution is described in detail with a series of formulas, some of which can be found in the above contents. In this study, we use the convergence order to describe the optimization performance of this algorithm, which is defined as

$$\lim_{k \to \infty} \left( \frac{\varepsilon_{k+1}}{\varepsilon_k^p} \right) = r.$$  \hspace{1cm} (7)

In this equation, $\varepsilon_k$ and $\varepsilon_{k+1}$ are the $k^{th}$ and $(k+1)^{th}$ root-mean-squared errors (RMS) of the phase recovery during the iteration, respectively. The parameter $p$ is a constant that defines the convergence order, and $r$ is also a constant that defines the convergence rate. The phase RMS is defined as follows, where $E_R$ and $E$ are the recovered and inputted object phase, respectively, $N$ is the number of pixels, and the summation operates over all the pixels.

$$\varepsilon = \sqrt{\sum (E_R - E)^2 / N},$$  \hspace{1cm} (8)

**ALGORITHM 1  GENERAL PHASE RETRIEVAL EXACT SOLUTION**

**INPUT**  three phase-shifts and the measured intensities $w$, \{ $t_1$, $t_2$, $t_3$ \}, \{ $I_1$, $I_2$, $I_3$ \}

**CONDITION**  available for arbitrary complex objects, $w$ must be small

**Section I**  an approximate solution obtained by phase retrieval algebraic solution

1) Inverse linear transform:

\[
\begin{bmatrix}
R_1 \\
R_2 \\
R_3 \\
\end{bmatrix} = \begin{bmatrix}
1 & 2 \cos t_1 - 2 & 2 \sin t_1 \\
1 & 2 \cos t_2 - 2 & 2 \sin t_2 \\
1 & 2 \cos t_3 - 2 & 2 \sin t_3 \\
\end{bmatrix}^{-1} \begin{bmatrix}
I_1 \\
I_2 \\
I_3 \\
\end{bmatrix}
\]

2) Approximation of $Ke^p$:

\[
\tilde{K} = |\mathcal{F}(w)| \cdot \frac{R_1 - 2R_2 + \sqrt{(R_1 - 2R_2)^2 - 4(R_2^2 - R_3^2)}}{2|\mathcal{F}(w)|}\max \quad \text{or} \quad 0
\]

\[
\tilde{P} = 0
\]

3) Algebraic phase recovery:

$$C_i = \sqrt{R_i}, \quad E_i = \tilde{P} + \arctan \left( \frac{R_3}{R_2 + \tilde{K}} \right)$$

**OUTPUT**  a generally accurate phase recovery $\{C_i, E_i\}$
**Section II** iterative optimization to the object phase (efficiency: linear convergence)

**Initialization**  \( \bar{C}_2 = \bar{C}_1, \ E_2 = E_1 \)

**General Step**

1) Propagating to focal plane:  \( h_2e^{i\alpha_y} = \mathcal{F}(\bar{C}_2e^{i\theta_y}) \)

2) Update \( K \) and \( P \) meanwhile:  \( K_2e^{i\beta} = \mathcal{F}(h_2e^{i\alpha_y\cdot w}) \)

3) New phase recovery:  \( E_3 = P_2 + \arctan\left(R_3/[R_2 + K_2^2]\right) \)

**Executing steps (1~3)** in a loop for several times

**OUTPUT** an extremely accurate phase recovery \( \{\bar{C}_3, E_3\} \)

In this algorithm, both \( K \) and \( P \) are updated with the iteration. Since \( K \) is updated continuously, the phase difference \( (E - P) \) that is solved by \( \arctan(R_3/[R_2 + K_2^2]) \) will be increasingly precise. And because \( P \) is being improved meanwhile, the object’s phase recovery \( E_R \) becomes accurate quickly. So this is a dual-drive optimization. It should also be mentioned that the initial input of \( K \) can be arbitrary, either is set as a reasonable approximation given by Eq. (5, 6) or 0. It affects only the iteration progress but does not influence the convergence rate. Because of the linear convergence property, the final phase recovery can always be adequately accurate after some iterations. This issue will be discussed later in the simulation section.

**D. One-shot exact phase retrieval**

The above algebraic phase retrieval solution as well as the exact phase retrieval solution, are developed for reconstructing a complex object with arbitrary phase range. In their establishments, there are no linear approximations applied to the object phase. It requires three intensity measurements in general. However, when the object is a pure phase object, there is a particular case where two intensities are sufficient for exact phase recovery. The strategy is setting one of the phase-shifting values as zero, such as \( t_1 = 0 \). Then, the intensity \( I_1 \) will be independent of the object, but be a fixed function
\( I_1 = \bar{C}^2 = B^2 \), which is precisely the background illumination intensity and is hence written as \( I_B \). Obviously, \( I_B \) can be measured and calibrated in before the test, and thus we only need to apply two phase shifts \( \{t_2, t_3\} \) and measure corresponding intensities \( \{I_2, I_3\} \) in practical experiments. In this way, the algebraic relationship between the intensity and the phase difference is simplified as

\[
\begin{bmatrix}
I_2 - I_B \\
I_3 - I_B
\end{bmatrix} = \begin{bmatrix}
2\cos t_2 - 2 & 2\sin t_2 \\
2\cos t_3 - 2 & 2\sin t_3
\end{bmatrix} \begin{bmatrix}
BK \cos (E - P) - K^2 \\
BK \sin (E - P)
\end{bmatrix}.
\]  \tag{9}

The final phase retrieval solution is almost the same as that of Eq. (3), where \( \bar{C} \) and \( I_1 \) should be replaced by \( B \) and \( I_B \), respectively. Furthermore, the iterative optimization algorithm is also the same as Algorithm 1, but needs to replace \( C_2 \) with \( B \). It can be found that even the iterative performance is analogous, with almost the same \( p \) and \( r \). In a word, the exact phase retrieval solution of pure phase objects shares no difference from that of complex objects, except that there is a fixed intensity that can be calibrated in advance. However, when the pure phase object has a small phase range moreover, things will be different. For the fundamental relationship as expressed in Eq. (3), a tiny \( t \) would make the coefficient \( [2\cos(t) - 2] \) much smaller than another coefficient \( 2\sin(t) \). Thus only the sine term will be left if considers \( [2\cos(t) - 2] \) as 0. Applying a higher-order approximation to \( \cos(t) \), the relationship is finally built as

\[
I_t \approx \left( \sqrt{I_B} + \sin t \cdot K \cdot \sin (E - P) \right)^2, \tag{10}
\]

where \( I_t \) is the measured output intensity. This approximate formula shows that the phase difference can be extracted with only one intensity measurement, that is
\[ E = P + \arcsin \left( \frac{\sqrt{I_t} - \sqrt{I_B}}{\sin t \cdot K} \right) \xrightarrow{P = 0} \arcsin \left( \frac{\sqrt{I_t} - \sqrt{I_B}}{\sin t \cdot K} \right). \quad (11) \]

In this equation, the real solution cannot be determined since the function ‘arcsin’ is multivalued. If there is a precondition that the range of \((E - P)\) is less than \(\pi\), however, the solution ambiguity will disappear, and the phase is therefore recovered uniquely. That is the reason why here the object is required to be with a small phase range. Considering the reference phase \(P\) is very close to 0, the permitted maximum object phase range is around \(\pi\). Similar to Eq. (4), the one-shot algebraic phase retrieval solution expressed in Eq. (11) is also achieved based on an approximation of \(K\) and \(P\), where \(P\) is also regarded as 0, but \(K\) has a different initialization. Because of the incompleteness of the established relationship Eq. (10), we cannot obtain a rigorous equation about \(K\) as before. Nevertheless, note that \(\tilde{C}\) is pre-measured as \(B\) and \(E\) is supposed to be small that is regarded as 0. In this way, the following approximation solution of \(K\) would be feasible.

\[ K = \left| \mathcal{F} \left( \mathcal{F} \left( Ce^{i\tilde{E}} \right) \cdot w \right) \right| \approx \left| \mathcal{F} \left( \mathcal{F} \left( B \right) \cdot w \right) \right|. \quad (12) \]

Since \(K\) and \(P\) have been approximated well, the object phase now can be reconstructed precisely by Eq. (11). Concluding the proposed one-shot phase retrieval solution, it involves two intensity measurements, one of which is the background that is determined in before the test, the other is the output intensity measured under a small window phase-shifting applied to the object’s spectrum. In this method, the phase-shifting window \(w\) must be within a tiny region of the focal plane, typically should be inside the Airy pattern, and the phase shifting value must be small, typically should be smaller than 0.6 rad. Most importantly, there is a necessary precondition that the object
must be a pure phase object with the phase range no more than $\pi$.

Similar to the iteration optimization shown in Algorithm 1, the one-shot algebraic phase recovery obtained by Eq. (11) can also be improved into an exact one by iterative optimization. The iterative formula must base on an accurate relationship to be able to approach the exact solution continuously. However, the relationship built in Eq. (10) is just an approximation and therefore cannot be applied for the iteration. By reorganizing the accurate relationship Eq. (2), a feasible iterative formula should be

$$E = P + \arcsin \frac{I_t - I_B - 4\sin^2 \frac{t}{2} \left( \sqrt{I_B} K - \cos(E - P) - K^2 \right)}{2 \sin t \cdot \sqrt{I_B} K}. \quad (13)$$

The one-shot phase retrieval exact solution is described in the following table. It consists of two sections as well. First, a relatively accurate phase recovery $E_3$ is obtained using the one-shot algebraic phase retrieval solution. Second, $E_3$ is improved to be the exact one gradually with a designed iterative optimization strategy with high efficiency. In the iteration procedure, both $K$ and $P$ are updated in the same way as it is in Algorithm 1. Therefore, this is also a dual-drive optimization, speeding up the convergence. Furthermore, it is found the convergence performance highly depends on the value of $t$. When $t$ is adequately small, the convergence order will always be $p = 1$ (linear convergence), and the convergence rate will be the same as that of Algorithm 1 in the case of using the same phase shifting window. In contrast, when $t$ has a large value, the iteration will fall stagnation quickly, and the exact phase recovery will no longer be achieved. This issue will be discussed later in the simulation part.

**ALGORITHM 2  ONE-SHOT PHASE RETRIEVAL EXACT SOLUTION**

**INPUT**  background intensity, one small phase modulation and output intensity $\{I_B, w, t, I_t\}$

**CONDITION**  pure phase object with $E < \pi$, $w$ must be small, $t$ must be small
Section I an approximate solution obtained by phase retrieval algebraic solution

1) Initialization: \(E_0 = 0\)

2) Approximation of \(K e^{i\phi}\): \[
\begin{align*}
\tilde{K} &= \mathcal{F}\left(\mathcal{F}\left(\sqrt{I_b} e^{i\phi_0}\right) \cdot w\right) \\
\tilde{p} &= 0
\end{align*}
\]

3) Algebraic phase recovery: \(E_3 = \tilde{p} + \arcsin\left(\frac{\sqrt{I_t} - \sqrt{I_b}}{\sin t \cdot \tilde{K}}\right)\).

**OUTPUT** a relatively accurate phase recovery \(\{E_3\}\)

Section II iterative optimization to the object phase (efficiency: linear convergence)

**Initialization** \(E_2 = E_3\)

**General Step**

1) Propagating to focal plane: \(h_4 e^{i\phi_4} = \mathcal{F}\left(\sqrt{I_b} e^{i\phi_4}\right)\)

2) Update \(K\) and \(P\) meanwhile: \(K_4 e^{i\phi_4} = \mathcal{F}\left(h_4 e^{i\phi_4} \cdot w\right)\)

3) New phase recovery: \(E_4 = P_4 + \arcsin\left(\frac{I_t - I_y - 4\sin^2 t}{2\sin t \cdot \sqrt{I_b} K_4}\right)\)

**Executing steps (1–3)** in a loop for several times

**OUTPUT** an extremely accurate phase recovery \(\{E_4\}\)

3. Simulation

A. General exact phase retrieval

This section investigates the performances of the general exact phase retrieval solution. For a brief presentation, it is assumed the illumination has a uniform distribution, i.e., \(B \equiv 1\). The complex object is designed to be with large aberrations on both amplitude and phase, as shown in Fig. 2. The phase has a large range as \(E \in [0, 4]\). The minimum amplitude is set to 0.01 rather than 0 to ensure the phase has a meaningful value. In this way, the maximum amplitude is 100 times its minimum. The simulation is conducted on MATLAB 2015b platform. All the data are discretized with a size of 256×256. The 2D Fourier transform is applied to generate the object’s spectrum and the output intensities based on Eq. (1). The inputs \(\{B, C, E\}\) are extended into 1536×1536 by
zero-padding, so that obtains a higher spatial frequency resolution. As shown in Fig. 2(c), the phase on the focal plane is shifted within the central square area. This window area has a very small size of $7 \times 7$ pixels, occupying only $2 \times 10^{-5}$ the area of whole, but carrying more than 21% the intensity of the whole.

![Fig. 2. The complex object and the phase-shifting window used in the numerical simulation. The phase-shifting window locates on the centre of the Fourier spectrum, occupying $7 \times 7$ pixels among $1536 \times 1536$ pixels of the whole region.](image)

The phase shift values are set as $t = \{0, \pi/2, \pi\}$, and the corresponding output intensities are simulated using Fourier transform by considering the optical numerical aperture is unlimited. As shown in the left-hand figure of Fig. 3(a), the output amplitude without phase shifting is exactly the object amplitude, where the phase information is completely hidden. When there are phase shifts $t = \{\pi/2, \pi\}$ applied to the object’s spectrum, the phase structure emerges gradually but suffers severe interference by the amplitude image. Then according to the algebraic phase retrieval solution described in Algorithm 1 Section I, the object phase is reconstructed successfully, as shown in Fig. 3(b). It seems a good phase recovery result, while there is still significant residual error, seeing Fig. 3(c). This error can be reduced continuously by the iterative optimization strategy expressed in Algorithm 1 Section II. As depicted in Fig. 3(d), after applying 25 iterations, the improved recovered phase appears almost the same with the inputted one. Quantitatively, the phase RMS value drops from 0.2103 rad to 0.0003 rad, about 1‰
that of Fig. 3(b). As seen in the right-hand figure of Fig. 3(d), the residual error still has a global distribution as similar to Fig. 3(c), while the scale becomes negligible. Therefore, the developed iterative optimization algorithm is a powerful tool that brings significant improvement to the algebraic phase recovery.

Fig. 3. The performance of the general exact phase retrieval solution in recovering a complex object. (a) measured output intensity pattern under different phase-shifting values. (b) algebraic phase recovery by Algorithm 1 Section I. (c) residual error of algebraic phase recovery. (d) exact phase recovery by Algorithm 1 section II, 25 iterations are applied.

The convergence performance of the iteration is illustrated in the left-hand figure of Fig. 4. It is seen the RMS value in the logarithmic scale goes down with a nearly constant speed. According to the definition in Eq. (8), the convergence order is determined exactly, i.e., $p = 1$, also called as linear convergence. From the tendency, it is predictable that the phase recovery error will decrease until it approaches zero unlimitedly, ensuring us obtain the exact phase recovery. In this linear convergence, the convergence rate is defined by the parameter $r$, which is around 0.778, according to the phase RMS ratio curve. Note that the above convergence features are summarized
based on the current specific simulation. Further, a series of diverse simulations are carried on and have proved that the linear convergence is an intrinsic property of Algorithm 1. First, the convergence performance on different object phase ranges is studied. This simulation takes the same settings as given in Fig. 2 but resets the phase range. As shown in the disk-marked curve in the right-hand figure of Fig. 4, all the iterations are of linear convergence property but have different convergence rates. A basic conclusion is drawn that the convergence becomes slower with the increase of the phase range, which could be explained by the distribution change of the object’s Fourier spectrum. A complex object with a larger phase range would have its spectrum more dispersed and speckled. Hence, the reference phase $P$ will not be so much close to 0, leading to a less precise phase recovery ultimately. Second, the convergence speed for pure phase objects is investigated. The inputted object phase is the same as that in Fig. 2, but the object amplitude now is $C \equiv 1$. As shown in the square-marked curve of Fig. 4, it is also a linear convergence, and even the convergence rate follows a quite similar trace with that of complex objects. Third, the effect the phase-shifting window size has on the convergence rate is investigated, where there are a series of phase shifting window width (referred to as $W_d$), i.e., $W_d = \{3, 5, 7, 9\}$. As seen in the right-hand figure of Fig. 4, all the iterations have the features of linear convergence and the convergence rate $r$ under different $W_d$ appears a very similar tendency with each other as the object phase range increases. Besides, it is found the smaller the window size is, the faster the iteration converges. This is an undoubted phenomenon. As shown in Fig. 4(a) embedded in the right-hand figure of Fig. 4, the object’s spectrum inside a smaller central window appears flatter than that of a larger window, so that the produced reference phase $P$ becomes smoother, leading to a more precise phase recovery. However, a small phase-shifting window tends to make the output intensities
indistinguishable and thus increases the measurement difficulty. Therefore, there is a trade-off between the phase recovery accuracy and the experimental feasibility. Generally, it is recommended that the phase-shifting window occupies about half the Airy pattern.

![Fig. 4. The convergence performance of the iterative optimization strategy in Algorithm 1. The left figure shows the variations of phase RMS and phase RMS ratio along with the iteration. The right figure shows the convergence rates \( r \) of the linear convergence (\( p = 1 \)) under diverse situations. In this figure, \( C = \text{cat} \) means a complex object whose amplitude is the cat image of Fig. 2(1), \( C = 1 \) means a pure phase object. Due to the inaccurate RMS calculation caused by phase wrapping, simulation with phase range over 1.6\( \pi \) were not done.]

In summary, the linear convergence is an intrinsic characteristic of the developed iterative optimization strategy: Algorithm 1, Section II. This property ensures the exact phase recovery is achievable. To accelerate the convergence, we should choose a phase-shifting window as small as possible. On the contrary, if both the object phase range and the window area are large, then the iteration may fall stagnant quickly or even be divergent instead of a linear convergence.

**B. One-shot exact phase retrieval**

Different from the general exact phase retrieval solution, the one-shot exact phase retrieval solution has a necessary condition: pure phase object with phase range no more than \( \pi \). There require two intensity measurements to reconstruct the phase object. One
is the background intensity $I_B$ that is measurable in advance, and the other is the intensity $I_t$ which is measured with phase modulation applied. This methodology shares some similarity with a classical qualitative phase retrieval technology, i.e., the Zernike phase contrast (ZPC). In ZPC, the phase-shifting on the object’s spectrum happens on the Airy pattern, and the value is $\pi/2$. The core of ZPC is the rough linearization approximation applied to the object phase, i.e., $e^{iE} \approx 1 + iE$. Suppose the phase-shifting acts on only the background wave, and then we will obtain a phase recovery solution $I_Z \approx I_B (1 + E)^2$, where $I_Z$ is the measured output intensity, and $I_B$ is the background. In this way, the phase is reconstructed by the formula

$$E_Z = \sqrt{I_Z} / \sqrt{I_B} - 1.$$  \hfill (14)

For simplicity, $U$ is applied to indicate the square root of the intensity, $U = \sqrt{I}$. The simulation is arranged as follows. The pure phase object is the same as Fig. 2(b), but its phase range is reset to 2 rad, meeting the requirement of small phase range. Two background intensity types are considered, i.e., a uniform background and an unsmooth one. The phase-shifting window size $W_d$ is an important factor in both ZPC and the proposed one-shot quantitative phase imaging method (referred to as QPC). For ZPC, as shown in the embedded figure of Fig. 5(a), $W_d$ should be 11 to cover the Airy pattern exactly. While for QPC, it is selected as a smaller value, $W_d = 7$, to ensure the object’s spectrum within the window is rather smooth. The phase-shifting value $t$ is also a key parameter. For ZPC it is surely $\pi/2$. While for QPC, it is set as $t = 0.5$ to satisfy the requirement that $t$ should be as small as possible.

At first, we discuss the uniform background illumination case, that is, $B \equiv 1$. The output intensities of ZPC and QPC are calculated with Eq. (1) and are shown in Fig. 5(a, c), respectively. It is found the difference between the background amplitude and
the measured one reflects the phase clearly, especially for that of the QPC method. Hence, they can be regarded as a good qualitative phase imaging to the pure phase object. Then apply Eq. (16) and Algorithm 2 to reconstruct the phase quantitatively. The result of ZPC is shown in Fig. 5(b), which appears to be a good phase recovery but the error is still obvious. In contrast, the phase recovery of QPC is very accurate and even can be completely exact. As shown in Fig. 5(d), the iteration optimization exhibits to be of a linear convergence, with the RMS value dropping from 0.103 to 0.00008 within only ten iterations. It is estimated the convergence rate is $r = 0.472$. From the square-marked curve in the right-hand figure of Fig. 4, the convergence rate for the general exact phase retrieval solution under the same phase range is around 0.464, which is close to the convergence rate here.
Fig. 5. The performance of one-shot exact phase retrieval solution under uniform illumination. (a, e) the output amplitude and the qualitative phase imaging of ZPC. (b, f) the quantitative phase imaging and residual error of ZPC. (c, g) the output amplitude and the qualitative phase imaging of QPC. (d, h) the quantitative phase imaging of QPC and the convergence properties of Algorithm 2.

Second, the phase recovery performance of QPC in unsmooth illumination case is investigated. Here the unsmooth illumination is defined by a grayscale image, ‘cat’, as shown in Fig. 2(a). Other simulation settings, including the phase object and the phase-shifting window, are the same as the above simulation. The intensities on the output plane for ZPC and QPC are simulated and shown in Fig. 5(e) and Fig. 5(g), respectively. For ZPC, the measured amplitude reflects the phase directly; however, the subsequent
quantitative phase recovery solved by Eq. (16), as shown in Fig. 5(f), still contains an obvious error. Comparing the RMS values of ZPC under uniform illumination (RMS = 0.2567) and unsmooth illumination (RMS = 0.3988), it is concluded that ZPC is more suitable to be applied in uniform background situations. However, the phase recovery accuracy is also poor. For QPC, the measured amplitude is the background amplitude itself, in which the phase is almost hidden. However, the phase emerges clearly by subtracting the background. This is a good qualitative phase image that is much better than that of ZPC. Then apply Algorithm 2 to solve the phase quantitatively. As seen in Fig. 5(h), the phase recovery result at 10th-iteration as well as the convergence process, are given. A near-optimal phase recovery is achieved, and the residual error is negligible. Moreover, the iteration optimization process appears to be with a linear convergence property too. It is estimated the convergence rate is $r = 0.470$, a number close to that of Fig. 5(d).

In conclusion, the one-shot exact phase retrieval solution is feasible to recover the pure phase object with small phase range under arbitrary illumination patterns. The convergence property is almost the same as the general exact phase retrieval solution. This is because the iterative mechanism of Algorithm 1 and Algorithm 2 are the same. The convergence features depend mainly on the phase-shifting window and the object phase range but are almost irrelevant to the tested object itself and the background illumination pattern. Thanks to the linear convergence property, the exact solution is always ensured as long as the iteration is enough. To maintain the linear convergence, the phase shifting value of QPC must be tiny, typically should be smaller than 1 rad. A smaller $t$ and a smaller object phase range contributes to faster convergence. Overall, the above simulations demonstrate the outstanding advantage of the proposed QPC over the classical ZPC. Using similar measurement methodology but applying a smaller
phase-shifting window and a tiny phase shifting value, QPC can reconstruct the pure small-phase object at very high accuracy under arbitrary illuminations. This is an unachievable target for typical ZPC or other similar one-shot phase imaging methods, as their approximation model, \( e^{i\theta} \approx 1 + iE \), is too rough.

4. Experiment

A. Experimental platform

The experimental optical setup is arranged, as illustrated in Fig. 1. Both the unknown phase object and the small-window phase modulation are generated by reflecting phase-only SLMs (UPOLabs, HDSLM63R, 1280×720, 8 bits, 6.37μm pixel width), where SLM #1 is placed at the object plane to load the phase object and SLM #2 at the focal plane to load the phase mask. The incident beam is produced from a commercial single-mode laser (Changchun New Industries, MGL-III-532, 532±1 nm, TEM\(_{00}\) mode). A pinhole with 10-μm diameter is placed behind to filter the original laser beam to be a smooth and aberration-less one and is then collimated by a condenser lens. To match the SLM screen, the parallel beam is truncated by a 3mm-diameter diaphragm, leaving only the central quite-smooth part. A polarizer is added before the SLM to ensure the SLM operating in phase-only mode. On the focal plane, the object’s Fourier spectrum is shifted in phase within a small window area. Subsequently, a black and white CMOS image sensor (Microview, HK-A5100-GM17, 2560×2160, 16 bit, 6.5μm pixel width) are arranged on the image plane to record the output intensity pattern.

It was found the SLM would be inaccurate when loads a phase modulation with too small phase shifting value; therefore, the present experimental platform is only suitable for validating the phase retrieval exact solution on three shots. The experimental proof for the one-shot phase retrieval exact solution will be given in the
short future after a better SLM is available. In the experiments, three phase-shifting values \{0, \pi/2, \pi\} are loaded on SLM #2 successively, and consequently, three intensity images are measured by the CMOS. Then the phase recovery performance of the proposed general exact phase retrieval solution (expressed in Algorithm 1) is investigated and evaluated. It should be pointed out then there is always an unmodulated light beam that merges into the modulated beam due to the pixel gap of the SLM and hence causes a blend. To eliminate aliasing, an additional suitable oblique phase mask is added into SLM #1 and SLM #2 to separate the modulated light beam into the measurement.

B. Smooth phase object

At first, we test the performance of the proposed method for detecting smooth phase objects. The light beam is shaped by a circular pupil whose diameter is 3 mm. On the object plane, as shown in Fig. 6(a), the inputted phase object is a simulated lens with a focal length of 1000 mm, covering the whole screen of SLM #1. On the focal plane, the phase on the central square area of side length 70 \mu m (11\times11 pixels on SLM #2) is shifted by \{0, \pi/2, \pi\} successively. As shown in Fig. 6(b), three intensity images are measured, which contains some systematic errors. The ring artifacts are caused by the Fresnel diffraction of the diaphragm, while the dark speckles result from the dead pixels or specks of dust on the CMOS screen. It is seen from the intensity variation that the phase structure (a lens) emerges when applies phase modulation on the spectrum. Following the phase retrieval solution expressed in Algorithm 1, the phase object is reconstructed. Seeing Fig. 6(c), the reconstructed phase appears obvious paraboloid features, and intuitively it matches the inputted one very well. To quantify the recovery accuracy, simply we focus on the marked cross line in Fig. 6(a). The phase recovery on
this line is given in Fig. 6(d), where the fitting parabola has a curvature value as $c = 5.998$. Thus the recovered focal-length is $\pi/\lambda c = 982$ mm. Compared with 1000 mm of the given one, the phase detection exhibits high accuracy, with a 2% relative error.

Fig. 6. Phase imaging experimental results of smooth phase object. (a) inputted phase on SLM #1, a lens-phase with focal length as 1000 mm. (b) recorded intensity patterns on the CMOS. (c) algebraic phase recovery. (d) phase recovery fitting parabola, which has a curvature about 6 and indicates there is a lens-phase with 982 mm focal length.

C. Rough phase object

Then, we test the performance of the method in detecting rough phase objects, and meanwhile, the spatial resolution is investigated. Theoretically, from the given beam width ($D = 3$ mm), imaging distance ($f = 150$ mm), and the wavelength ($\lambda = 532$ nm), the spatial resolution of current optical imaging platform is determined as $d = \lambda f / D = 26.65$ μm. As shown in Fig. 7(a), a complex grayscale image “balls” that has the range $[0, 2\pi]$ rad is inputted in SLM #1 as the rough phase object. The phase modulation is the same as that of Fig. 6. The recorded intensity images are listed in Fig. 7(b), shown in log-scale to present a clearer distinction. Due to the existing system errors and
experimental defects such as the SLM not operating on phase-only mode, a limited numerical aperture and a loose 4f-imaging-system configuration, the phase information emerges even when there is no phase modulation. However, that cannot provide us with quantitative phase detection. With two additional intensity images under phase modulation of \( t = \{\pi/2, \pi\} \), the phase object is well recovered with only one algebraic calculation, as shown in Fig. 7(c). It appears a high-accuracy phase reconstruction and exhibits many details. From the comparison on the section line of Fig. 7(d), it is estimated that the phase recovery error in average is about 0.1 rad, and the spatial resolution is better than 40 \( \mu \)m, matching the theoretical expectation.

In addition, there are two noteworthy phenomena in practical experiments. First, the phase recovery solution is point-to-point, so it is workable for any selected area of the intensity data into the calculation. For example, here, the central data of 501×501 pixels are selected from the whole region of 736×736 pixels for the phase object reconstruction, and that works well. Second, the phase recovery can avoid suffering the influence of the meaningless intensity data caused by the dead pixels. For instance, there is a ‘branch’ indicated by an arrow in the intensity maps, as shown in Fig. 7(b), but it is missing in the recovered phase map, showing a self-healing ability of the method.
Fig. 7. Phase imaging experimental results of rough phase object. (a) inputted phase on SLM #1, a bunch of balls with the phase range as $[0, 2\pi]$ rad. (b) recorded intensity patterns on the CMOS. (c) algebraic phase recovery. (d) comparison between the inputted phase and the recovered one.

D. Vortex phase object

Finally, we test the performance of the method for detecting vortex phase objects. In addition to the performance in detecting topological number, meanwhile, the accuracy improvement brought by background elimination is also investigated. In some fields of optical communication, phase retrieval is a significant tool to restore the phase distribution carried by the vortex light beam to decode the digital information. So it is important to test the performance on detecting a vortex phase. Besides, there are some obvious ring artifacts caused by near-field diffraction in both the recorded intensity patterns and the recovered phase image, as seen in Fig. 7(b, c). This reduces phase recovery accuracy. One effective means to remove this kind of artifact could be the background elimination strategy. Following that method, there are two phase-only objects to be loaded on the SLM #1 screen in this experiment. One is the unknown vortex phase object to be reconstructed, and another is a zero-phase input which is used to detect the background phase distortions.
Fig. 8. Phase imaging experimental results of vortex phase object. (a) inputted phase on SLM #1, a vortex phase plate with a topological number as 16. (b) recoded intensity patterns on the CMOS. (c) algebraic phase recovery. (d) inputted zeros-phase object. (e) recorded background intensity patterns. (f) algebraic phase recovery of the background. (g) new amplitude recovery by dividing by the background amplitude. (h) new phase recovery after subtracting the background phase. (i) phase recovery errors of (c) and (h).

As listed in Fig. 8, the whole process and results of this experiment are shown in detail. Six intensity patterns are being measured, as given by Fig. 8(b, e). The phase modulation is the same as that of Fig. 6. Based on Algorithm 1, the phase objects are solved from their intensity patterns, respectively, as shown in Fig. 8(c) and Fig. 8(f).

For the vortex phase, it is seen the reconstructed phase appears obvious vortex structure and has a topological number of 16. It exhibits the same vortex structure and even a very similar phase shape. However, there is an obvious phase error on edge. Besides, there are also many ring artifacts on the intensities as well as the recovered phase. By inputting a zero-phase on SLM #1, the background amplitude and phase of the laser beam are detected. It is seen in Fig. 8(f) the background phase aberration looks like an oblique plane with a small range of variation. After removing the background, a more
accurate phase retrieval result is obtained. As shown in Fig. 8(g, h), now the ring artifacts, caused by Fresnel diffraction, are eliminated. Also, the laser beam distribution is indicated clearly. It is seen in the new amplitude recovery that the SLM has a nearly homogenous modulation on the light intensity, but blocks the light significantly at the vortex angles. And in the new phase recovery, the errors on edge gets eliminated basically, resulting in a much-improved phase reconstruction. The optimization of phase recovery accuracy can also be quantified by Fig. 8(i), where the phase RMS has decreased from 0.81 rad to 0.39 rad after the background elimination. Similar to the above experiments, the small speckles on the CMOS screen affected only the intensity measurement but has little influence on the phase recovery. Overall, this experiment verifies the potential of the method for detecting the vortex phase, meanwhile validates the accuracy improvement brought by the background elimination. Note that such background calibration needs only to be done once before the experiment, so it brings little extra work.
5. Discussion

A discussion about the difference in the phase recovery performance of applying small-window phase shifting on the object and the object spectrum is presented. This study proposes an exact phase recovery solution based on shifting the phase of the object’s Fourier spectrum. In our previous work [34], a similar method is developed for solving the phase retrieval problem that occurs in coherent diffraction imaging, where the object is reconstructed by measuring only the diffraction intensity patterns. In that method, the small-window phase modulation is applied to the object rather than its Fourier spectrum. So it is called object-modulated algebraic phase retrieval. The algebraic relationship is similar to Eq. (3) but happens on the Fourier domain, that is, it connects the diffractive intensity and the diffractive phase difference. Since the relationship is built on far-field, an additional inverse Fourier transform is required to recover the object after the inverse matrix operation. Therefore, unlike the spectrum-modulated algebraic phase retrieval, the object-modulated algebraic phase retrieval does not have such a good property of point-by-point calculation. Also, since the object’s far-field phase ranges naturally from 0 to 2π, there always requires three phase-shifts and three diffracted intensity patterns to reconstruct the objects, regardless of whether the object is a general complex object, a phase-only object or even a small phase object. Another significant difference lies in the iterative optimization. First, the reference amplitude can be solved accurately. Second, it is found only the reference phase can be updated, or the iteration will not be of a stable convergence. The result is that the exact phase recovery is unobtainable in object-modulated algebraic phase retrieval because the iterative optimization cannot avoid falling stagnant. For an intuitive comparison, the differences between the object-modulated algebraic phase retrieval solution and the spectrum-modulated algebraic phase retrieval solution are summarized below.
### TABLE 3. CHARACTERISTICS OF PHASE RETRIEVAL ALGEBRAIC SOLUTIONS

|                        | Object-modulated algebraic phase retrieval solution [34] | Spectrum-modulated algebraic phase retrieval solution |
|------------------------|---------------------------------------------------------|------------------------------------------------------|
| **Modulation on**      | Object phase                                            | Far-field phase                                      |
| **Modulation Location**| Any place                                               | The central                                         |
| **Modulation Area**    | Less than 4% of the object area                         | Inside the Airy pattern                              |
| **Intensity Measurements** | Three, for any kinds of object                          | One, for small phase object                          |
|                        |                                                          | Two, for general phase object                        |
|                        |                                                          | Three, for general complex object                    |
| **Algebraic Relationship** | Built on the focal plane                               | Built on the object plane                            |
| **Algebraic Result**   | The phase difference between the object's far-field phase and reference | The phase difference between the object's phase and reference |
| **Reference Amplitude**| Uniquely defined and almost exact                       | Undefined and low-accuracy                           |
| **Reference Phase**    | Approximated to zero, low-accuracy                      | Approximated to zero, high-accuracy                  |
| **First Phase Recovery** | Low-accuracy                                           | High-accuracy                                        |
| **Iterative Optimization** | Single-drive (updating the reference phase only)       | Dual-drive (updating both reference amplitude and phase) |
|                        | Rapid convergence but falling stagnant quickly          | Rapid convergence and exhibiting linear convergence  |
| **Iteration Times**    | Several, typically less than 10                        | Some, typically no more than 100                     |
| **Final Phase Recovery** | High-accuracy                                          | Exact                                                |

### 6. Conclusion

A general non-reference phase retrieval exact solution is proposed, which has much potential to be applied in accurate QPI. The optical configuration is similar to a typical 4f phase contrast platform, to which we have made a significant modification, including the phase-shifting area, phase shifting value, and even the phase reconstruction algorithm. Without any approximation and limitation to the tested object as well as the illumination style, this method is able to achieve an exact reconstruction for arbitrary complex-objects with three intensity measurements. Particularly, for phase-only objects and small phase objects, two intensities and a single intensity will be sufficient,
respectively. The exact phase recovery is solved by two steps, i.e., algebraic phase recovery and then a high-efficiency iterative optimization. Based on a rigorous algebraic relationship built between the object phase and the measured intensities, a good approximate phase recovery is obtained. Afterwards, an effective iterative optimization algorithm is developed to improve that approximate solution. Because of the linear convergence property of the iteration, the error caused by regarding the reference phase as zero could be eliminated completely, and finally, an exact phase recovery is ensured.

In conclusion, this study proposed a methodology of obtaining accurate QPI by combining self-reference interferometry and iterative phase retrieval. This method is significant progress of previous works, including generalized phase contrast [33] and object-modulated algebraic phase retrieval [34]. In addition to the general exact QPI used for arbitrary complex-objects, an exact one-shot QPI is also developed, which is applicable for phase imaging to transparent objects immersed in surrounding materials. In brief, this study enables a completely-accurate QPI of general complex objects in the absence of direct reference wave. The future work will focus on extending this 2D exact QPI to 3D application cases.

CRediT

Jianhui Huang: Conceptualization, Methodology, Software, Visualization, Experiment, Writing- Original draft preparation Qian Ye: Supervision, Reviewing and Editing, Funding acquisition, Discussion Huiliang Jin: Writing- Project administration, Discussion
Acknowledgments

This work is supported by the National Science Foundation of China (NSFC), U1931137 and the National Basic Research Program of China (973 Program), 2015CB857100. The support provided by China Scholarship Council (CSC) during a visit of Jianhui Huang to Georgia Tech is acknowledged. Besides, the authors sincerely appreciate the help of UPO Labs (http://www.realic.cn) on the optical experimental setup and test.

References

[1] M. Maiden, G. R. Morrison, B. Kaulich, A. Gianoncelli, J. M. Rodenburg, Soft X-ray spectromicroscopy using ptychography with randomly phased illumination. Nature Communications 4, 1669 (2013).

[2] G. Hislop, L. Li, A. Hellicar, Phase Retrieval for Millimeter- and Submillimeter-Wave Imaging. IEEE Trans. Antennas Propag. 57, 286-290 (2009).

[3] Y. Zhang, S. Poonja, A. Roorda, MEMS-based adaptive optics scanning laser ophthalmoscopy. Opt. Lett. 31, 1268-1270 (2006).

[4] L. Tian, L. Waller, Quantitative differential phase contrast imaging in an LED array microscope. Opt. Express 23, 11394-11403 (2015).

[5] Y. Park, C. Depeursinge, G. Popescu, Quantitative phase imaging in biomedicine. Nat. Photonics 12, 578-589 (2018).

[6] Lee et al., Single-shot phase retrieval via Fourier ptychographic microscopy. Optica 5, 976-983 (2018).

[7] Zuo, Q. Chen, W. Qu, A. Asundi, Noninterferometric single-shot quantitative phase microscopy. Opt. Lett. 38, 3538-3541 (2013).

[8] S. Chowdhury, W. J. Eldridge, A. Wax, J. Izatt, Refractive index tomography with structured illumination. Optica 4, 537-545 (2017).

[9] Y. Bao, T. K. Gaylord, Iterative optimization in tomographic deconvolution phase microscopy. J. Opt. Soc. Am. A 35, 652-660 (2018).

[10] J. Li et al., Three-dimensional tomographic microscopy technique with multi-frequency combination with partially coherent illuminations. Biomed. Opt. Express 9, 2526-2542 (2018).

[11] P. Marquet, C. Depeursinge, P. J. Magistretti, Exploring neural cell dynamics with digital holographic microscopy. Annu. Rev. Biomed. Eng. 15, 407-431 (2013).

[12] M. Paturzo et al., Digital Holography, a metrological tool for quantitative analysis: Trends and
future applications. *Opt. Laser Eng.* **104**, 32-47 (2018).

[13] T. Kakue *et al.*, High-speed phase imaging by parallel phase-shifting digital holography. *Opt. Lett.* **36**, 4131-4133 (2011).

[14] N. Patel *et al.*, Wavefront division digital holographic microscopy. *Biomed. Opt. Express* **9**, 2779-2784 (2018).

[15] S. Ebrahimi, M. Dashtdar, E. Sánchez-Ortiga, M. Martínez-Corral, B. Javidi, Stable and simple quantitative phase-contrast imaging by Fresnel biprism. *Appl. Phys. Lett.* **112**, 113701 (2018).

[16] S. H. S. Yaghoubi, S. Ebrahimi, M. Dashtdar, A. Doblas, B. Javidi, Common-path, single-shot phase-shifting digital holographic microscopy using a Ronchi ruling. *Appl. Phys. Lett.* **114**, 183701 (2019).

[17] F. Zernike, Phase contrast, a new method for the microscopic observation of transparent objects Part II. *Physica* **9**, 974-986 (1942).

[18] P. J. Rodrigo, D. Palima, J. Gluckstad, Accurate quantitative phase imaging using generalized phase contrast. *Opt. Express* **16**, 2740-2751 (2008).

[19] G. Popescu, T. Ikeda, R. R. Dasari, M. S. Feld, Diffraction phase microscopy for quantifying cell structure and dynamics. *Opt. Lett.* **31**, 775-777 (2006).

[20] P. Bon, G. Maucort, B. Wattellier, S. Monneret, Quadriwave lateral shearing interferometry for quantitative phase microscopy of living cells. *Opt. Express* **17**, 13080-13094 (2009).

[21] P. Gao, I. Harder, V. Nercissian, K. Mantel, B. Yao, Phase-shifting point-diffraction interferometry with common-path and in-line configuration for microscopy. *Opt. Lett.* **35**, 712-714 (2010).

[22] N. Streibl, Phase imaging by the transport equation of intensity. *Opt. Commun.* **49**, 6-10 (1984).

[23] Wang, Q. Fu, X. Dun, W. Heidrich, Quantitative Phase and Intensity Microscopy Using Snapshot White Light Wavefront Sensing. *Sci. Rep.* **9**, 13795 (2019).

[24] L. Lu *et al.*, Quantitative Phase Imaging Camera with a Weak Diffuser. *Frontiers in Physics* **7**, (2019).

[25] J. Huang, H. Jin, Q. Ye, G. Meng, Surface deformation recovery algorithm for reflector antennas based on geometric optics. *Opt. Express* **25**, 24346-24361 (2017).

[26] Y. Shechtman *et al.*, Phase retrieval with application to optical imaging: A contemporary overview. *IEEE Signal Processing Magazine* **32**, 87-109 (2015).

[27] L. J. Allen, M. P. Oxley, Phase retrieval from series of images obtained by defocus variation. *Opt. Commun.* **199**, 65-75 (2001).

[28] J. Huang, H. Jin, Q. Ye, G. Meng, Iterative phase retrieval by combining modulus constraints and angle relationships. *Inverse Probl.* **35**, 014002 (014023 pp.) (2018).

[29] X. Ou, R. Horstmeyer, C. Yang, G. Zheng, Quantitative phase imaging via Fourier ptychographic microscopy. *Opt. Lett.* **38**, 4845-4848 (2013).

[30] N. Nakajima, Noniterative phase retrieval from a single diffraction intensity pattern by use of
an aperture array. *Phys. Rev. Lett.* **98**, 223901 (223904 pp.) (2007).

[31] R. A. Gonsalves, Small-phase solution to the phase-retrieval problem. *Opt. Lett.* **26**, 684-685 (2001).

[32] M. Li, X.-Y. Li, W.-H. Jiang, Small-phase retrieval with a single far-field image. *Opt. Express* **16**, 8190-8197 (2008).

[33] Palima, J. Glückstad, Generalized phase contrast: microscopy, manipulation and more. *Contemporary Physics* **51**, 249-265 (2010).

[34] J. Huang, H. Jin, Q. Ye, G. Meng, Phase-retrieval algebraic solution based on window modulation. *Annalen der Physik* **530**, 1800063 (1800069 pp.) (2018).
Fig. 1. Theoretical and experimental schematic diagram of the exact phase retrieval technique. The window-area for phase-shifting should be small enough. A commercial laser of 532-nm wavelength acts as the light source. Two SLMs (labelled as SLM #1 and SLM #2) are applied to load phase object and phase modulation, respectively. A grey-scale CMOS is placed at the end of the 4f imaging system to record the intensity image.
**Figure 2**

Fig. 2. The complex object and the phase-shifting window used in the numerical simulation. The phase-shifting window locates on the centre of the Fourier spectrum, occupying 7×7 pixels among 1536×1536 pixels of the whole region.
Fig. 3. The performance of the general exact phase retrieval solution in recovering a complex object. (a) measured output intensity pattern under different phase-shifting values. (b) algebraic phase recovery by Algorithm 1 Section I. (c) residual error of algebraic phase recovery. (d) exact phase recovery by Algorithm 1 section II, 25 iterations are applied.
Fig. 4. The convergence performance of the iterative optimization strategy in Algorithm 1. The left figure shows the variations of RMS and RMS ratio along with the iteration. The right figure shows the convergence rates $r$ of the linear convergence ($p = 1$) under diverse situations. In this figure, $C = \text{cat}$ means a complex object whose amplitude is the cat image of Fig. 2(1). $C = 1$ means a pure phase object. Due to the inaccurate RMS calculation caused by phase wrapping, simulation with phase range over $1.6\pi$ were not done.
Fig. 5. The performance of one-shot exact phase retrieval solution under uniform illumination. (a, e) the output amplitude and the qualitative phase imaging of ZPC. (b, f) the quantitative phase imaging and residual error of ZPC. (c, g) the output amplitude and the qualitative phase imaging of QPC. (d, h) the quantitative phase imaging of QPC and the convergence properties of Algorithm 2.
Figure 6

Fig. 6. Phase imaging experimental results of smooth phase object. (a) inputted phase on SLM #1, a lens-phase with focal length as 1000 mm. (b) recorded intensity patterns on the CMOS. (c) algebraic phase recovery. (d) phase recovery fitting parabola, which has a curvature about 6 and indicates there is a lens-phase with 982 mm focal length.
Fig. 7. Phase imaging experimental results of rough phase object. (a) inputted phase on SLM #1, a bunch of balls with the phase range as [0, 2π] rad. (b) recorded intensity patterns on the CMOS. (c) algebraic phase recovery. (d) comparison between the inputted phase and the recovered one.
Fig. 8. Phase imaging experimental results of vortex phase object. (a) inputted phase on SLM #1, a vortex phase plate with a topological number as 16. (b) recoded intensity patterns on the CMOS. (c) algebraic phase recovery. (d) inputted zeros-phase object. (e) recorded background intensity patterns. (f) algebraic phase recovery of the background. (g) new amplitude recovery by dividing by the background amplitude. (h) new phase recovery after subtracting the background phase. (i) phase recovery errors of (c) and (h).
Algorithm 1

**ALGORITHM 1  GENERAL PHASE RETRIEVAL EXACT SOLUTION**

**INPUT**  three phase-shifts and the measured intensities $w, \{t_1, t_2, t_3\}, \{I_1, I_2, I_3\}$

**CONDITION** available for arbitrary complex objects, $w$ must be small

**Section I** an approximate solution obtained by phase retrieval algebraic solution

1) Inverse linear transform:

$$
\begin{bmatrix}
R_1 \\
R_2 \\
R_3
\end{bmatrix} = \begin{bmatrix}
1 & 2 \cos t_1 - 2 \sin t_1 \\
1 & 2 \cos t_2 - 2 \sin t_2 \\
1 & 2 \cos t_3 - 2 \sin t_3
\end{bmatrix}^{-1} \begin{bmatrix}
I_1 \\
I_2 \\
I_3
\end{bmatrix}
$$

2) Approximation of $K e^{i\theta}$:

$$
\begin{cases}
\tilde{K} = |\mathcal{F}(w)| \cdot \left( \frac{R_1 - 2R_2 + \sqrt{(R_1 - 2R_2)^2 - 4(R_2^2 + R_3^2)}}{2 |\mathcal{F}(w)|_{\text{max}}} \right) \\
\tilde{P} = 0
\end{cases}
$$

3) Algebraic phase recovery:

$$
\tilde{C}_1 = \sqrt{R_1}, \quad \tilde{E}_1 = \tilde{P} + \arctan\left( \frac{R_3}{R_2 + \tilde{K}} \right)
$$

**OUTPUT** a generally accurate phase recovery $\{\tilde{C}_1, \tilde{E}_1\}$

**Section II** iterative optimization to the object phase (efficiency: linear convergence)

**Initialization** $\tilde{C}_2 = \tilde{C}_1, \quad \tilde{E}_2 = \tilde{E}_1$

**General Step**

1) Propagating to focal plane:

$$h_2 e^{i\theta} = \mathcal{F}(\tilde{C}_2 e^{i\tilde{K}_2})$$

2) Update $K$ and $P$ meanwhile:

$$K_2 e^{i\theta} = \mathcal{F}(h_2 e^{i\theta} \cdot w)$$

3) New phase recovery:

$$\tilde{E}_2 = P_2 + \arctan\left( \frac{R_3}{R_2 + K_2^2} \right)$$

**Executing steps (1–3)** in a loop for several times

**OUTPUT** an extremely accurate phase recovery $\{\tilde{C}_2, \tilde{E}_2\}$
Algorithm 2

**Algorithm 2**  
**ONE-SHOT PHASE RETRIEVAL EXACT SOLUTION**

**INPUT**  
background intensity, one small phase modulation and output intensity \( \{I_B, w, t, I_t\} \)

**CONDITION**  
pure phase object with \( E < \pi \), \( w \) must be small, \( t \) must be small

**Section I**  
an approximate solution obtained by phase retrieval algebraic solution

1) Initialization:  
\( E_3 = 0 \)

2) Approximation of \( Ke^{ip} \):  
\[
\begin{align*}
\bar{K} &= |\mathcal{F}(\sqrt{I_B} e^{i\phi})| \\
\bar{P} &= 0
\end{align*}
\]

3) Algebraic phase recovery:  
\[
E_3 = \bar{P} + \arcsin \left( \frac{\sqrt{I_t} - \sqrt{I_B}}{\sin t \cdot \bar{K}} \right).
\]

**OUTPUT**  
a relatively accurate phase recovery \( \{E_3\} \)

**Section II**  
iterative optimization to the object phase (efficiency: linear convergence)

**Initialization**  
\( E_4 = E_3 \)

**General Step**

1) Propagating to focal plane:  
\( h_t e^{i\phi} = \mathcal{F}(\sqrt{I_B} e^{i\phi}) \)

2) Update \( K \) and \( P \) meanwhile:  
\( K_t e^{i\phi} = \mathcal{F}(h_t e^{i\phi} \cdot w) \)

3) New phase recovery:  
\[
E_4 = P_t + \arcsin \frac{I_t - I_B - 4\sin^2 \frac{t}{2} \left( \sqrt{I_B} K_t - \cos(E_t - P_t) - K_t^2 \right)}{2\sin t \sqrt{I_B} K_t}
\]

**Executing steps (1-3)** in a loop for several times

**OUTPUT**  
an extremely accurate phase recovery \( \{E_4\} \)
|                                | Object-modulated algebraic phase retrieval solution [34] | Spectrum-modulated algebraic phase retrieval solution |
|--------------------------------|----------------------------------------------------------|------------------------------------------------------|
| Modulation on                  | Object phase                                             | Far-field phase                                      |
| Modulation Location            | Any place                                                | The central                                          |
| Modulation Area                | Less than 4% of the object area                           | Inside the Airy pattern                              |
| Intensity Measurements         | Three, for any kinds of object                            | One, for small phase object                          |
|                                |                                                          | Two, for general phase object                         |
|                                |                                                          | Three, for general complex object                     |
| Algebraic Relationship         | Built on the focal plane                                 | Built on the object plane                             |
| Algebraic Result               | The phase difference between the object's far-field phase and reference | The phase difference between the object's phase and reference |
| Reference Amplitude            | Uniquely defined and almost exact                         | Undefined and low-accuracy                           |
| Reference Phase                | Approximated to zero, low-accuracy                        | Approximated to zero, high-accuracy                   |
| First Phase Recovery           | Low-accuracy                                             | High-accuracy                                        |
| Iterative Optimization         | Single-drive (updating the reference phase only)          | Dual-drive (updating both reference amplitude and phase) |
|                                | Rapid convergence but falling stagnant quickly            | Rapid convergence and exhibiting linear convergence   |
| Iteration Times                | Several, typically less than 10                          | Some, typically no more than 10                       |
| Final Phase Recovery           | High-accuracy                                            | Exact                                                |