

Transport-of-intensity phase imaging using Savitzky-Golay differentiation filter - theory and applications

Chao Zuo,^{1,2,3,*} Qian Chen,^{1,3} Yingjie Yu,⁴ and Anand Asundi²

¹*Jiangsu Key Laboratory of Spectral Imaging & Intelligence Sense, Nanjing University of Science and Technology, Nanjing, Jiangsu Province 210094, China*

²*Centre for Optical and Laser Engineering, School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore 639798, Singapore*

³*Key Laboratory of Photoelectronic Imaging Technology and System, Ministry of Education of China, Beijing Institute of Technology, Beijing 100081, China*

⁴*Department of Precision Mechanical Engineering, Shanghai University, Shanghai 200072, China*
surpasszuo@163.com

Abstract: Several existing strategies for estimating the axial intensity derivative in the transport-of-intensity equation (TIE) from multiple intensity measurements have been unified by the Savitzky-Golay differentiation filter - an equivalent convolution solution for differentiation estimation by least-squares polynomial fitting. The different viewpoint from the digital filter in signal processing not only provides great insight into the behaviors, the shortcomings, and the performance of these existing intensity derivative estimation algorithms, but more important, it also suggests a new way of improving solution strategies by extending the applications of Savitzky-Golay differentiation filter in TIE. Two novel methods for phase retrieval based on TIE are presented - the first by introducing adaptive-degree strategy in spatial domain and the second by selecting optimal spatial frequencies in Fourier domain. Numerical simulations and experiments verify that the second method outperforms the existing methods significantly, showing reliable retrieved phase with both overall contrast and fine phase variations well preserved.

©2013 Optical Society of America

OCIS codes: (100.5070) Phase retrieval; (100.3010) Image reconstruction techniques; (120.5050) Phase measurement.

References and links

1. E. D. Barone-Nugent, A. Barty, and K. A. Nugent, "Quantitative phase-amplitude microscopy I: optical microscopy," *J. Microsc.* **206**(3), 194–203 (2002).
2. S. Bajt, A. Barty, K. A. Nugent, M. McCartney, M. Wall, and D. Paganin, "Quantitative phase-sensitive imaging in a transmission electron microscope," *Ultramicroscopy* **83**(1-2), 67–73 (2000).
3. T. E. Gureyev and S. W. Wilkins, "On X-ray phase retrieval from polychromatic images," *Opt. Commun.* **147**(4-6), 229–232 (1998).
4. G. Popescu, T. Ikeda, R. R. Dasari, and M. S. Feld, "Diffraction phase microscopy for quantifying cell structure and dynamics," *Opt. Lett.* **31**(6), 775–777 (2006).
5. P. Marquet, B. Rappaz, P. J. Magistretti, E. Cuche, Y. Emery, T. Colomb, and C. Depeursinge, "Digital holographic microscopy: a noninvasive contrast imaging technique allowing quantitative visualization of living cells with subwavelength axial accuracy," *Opt. Lett.* **30**(5), 468–470 (2005).
6. M. Reed Teague, "Deterministic phase retrieval: a Green's function solution," *J. Opt. Soc. Am.* **73**(11), 1434–1441 (1983).
7. N. Streibl, "Phase imaging by the transport equation of intensity," *Opt. Commun.* **49**(1), 6–10 (1984).
8. A. Barty, K. A. Nugent, D. Paganin, and A. Roberts, "Quantitative optical phase microscopy," *Opt. Lett.* **23**(11), 817–819 (1998).
9. D. Paganin and K. A. Nugent, "Noninterferometric phase imaging with partially coherent light," *Phys. Rev. Lett.* **80**(12), 2586–2589 (1998).
10. 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).
11. S. S. Gorthi and E. Schonbrun, "Phase imaging flow cytometry using a focus-stack collecting microscope," *Opt. Lett.* **37**(4), 707–709 (2012).

12. L. Waller, S. S. Kou, C. J. R. Sheppard, and G. Barbastathis, "Phase from chromatic aberrations," *Opt. Express* **18**(22), 22817–22825 (2010).
13. M. Beleggia, M. A. Schofield, V. V. Volkov, and Y. Zhu, "On the transport of intensity technique for phase retrieval," *Ultramicroscopy* **102**(1), 37–49 (2004).
14. 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).
15. 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).
16. M. Soto and E. Acosta, "Improved phase imaging from intensity measurements in multiple planes," *Appl. Opt.* **46**(33), 7978–7981 (2007).
17. 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).
18. R. Bie, X.-H. Yuan, M. Zhao, and L. Zhang, "Method for estimating the axial intensity derivative in the TIE with higher order intensity derivatives and noise suppression," *Opt. Express* **20**(7), 8186–8191 (2012).
19. B. Xue, S. Zheng, L. Cui, X. Bai, and F. Zhou, "Transport of intensity phase imaging from multiple intensities measured in unequally-spaced planes," *Opt. Express* **19**(21), 20244–20250 (2011).
20. S. Zheng, B. Xue, W. Xue, X. Bai, and F. Zhou, "Transport of intensity phase imaging from multiple noisy intensities measured in unequally-spaced planes," *Opt. Express* **20**(2), 972–985 (2012).
21. A. V. Martin, F. R. Chen, W. K. Hsieh, J. J. Kai, S. D. Findlay, and L. J. Allen, "Spatial incoherence in phase retrieval based on focus variation," *Ultramicroscopy* **106**(10), 914–924 (2006).
22. L. N. Trefethen, Finite difference and spectral methods for ordinary and partial differential equations, unpublished text, available at <http://web.comlab.ox.ac.uk/oucl/work/nick.trefethen/pdetext.html>, 1996.
23. S. J. Orfanidis, *Introduction to Signal Processing* (Prentice-Hall, Inc., 1995).
24. A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares Procedures," *Anal. Chem.* **36**(8), 1627–1639 (1964).
25. P. A. Gorry, "General least-squares smoothing and differentiation of nonuniformly spaced data by the convolution method," *Anal. Chem.* **63**(5), 534–536 (1991).
26. J. Luo, K. Ying, P. He, and J. Bai, "Properties of Savitzky–Golay digital differentiators," *Digit. Signal Process.* **15**(2), 122–136 (2005).
27. 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).
28. 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).
29. P. Barak, "Smoothing and differentiation by an adaptive-degree polynomial filter," *Anal. Chem.* **67**(17), 2758–2762 (1995).
30. J. M. Cowley, *Diffraction Physics*, 2 ed. (North-Holland Pub. Co, 1993).
31. Q. Weijuan, C. O. Choo, Y. Yingjie, and A. Asundi, "Microlens characterization by digital holographic microscopy with physical spherical phase compensation," *Appl. Opt.* **49**(33), 6448–6454 (2010).
32. W. Qu, C. O. Choo, V. R. Singh, Y. Yingjie, and A. Asundi, "Quasi-physical phase compensation in digital holographic microscopy," *J. Opt. Soc. Am. A* **26**(9), 2005–2011 (2009).
33. S. S. Kou, L. Waller, G. Barbastathis, and C. J. R. Sheppard, "Transport-of-intensity approach to differential interference contrast (TI-DIC) microscopy for quantitative phase imaging," *Opt. Lett.* **35**(3), 447–449 (2010).
34. L. Tian, J. C. Petrucci, and G. Barbastathis, "Nonlinear diffusion regularization for transport of intensity phase imaging," *Opt. Lett.* **37**(19), 4131–4133 (2012).

1. Introduction

Phase measurement plays a prominent role in many fields of physics, such as optics [1], electron- and X-ray microscopy [2, 3], diffraction [4]. Many samples such as optical elements, biological soft tissues, and cells are phase objects with little intensity variation for conventional brightfield microscopic imaging. Digital holography, as a paradigm of interferometric techniques is the primary quantitative technique for phase reconstruction with sub-wavelength accuracy [5]. Recently, however, direct phase retrieval from intensity measurements using the Transport-of-intensity equation (TIE) [6, 7] derived from the free space Helmholtz wave equation in the paraxial wave approximation has gained increasing attention. The TIE-based method has the advantages of being non-interferometric [1], not needing phase unwrapping [8], and applicable with partially coherent beams [9, 10]. Thus phase reconstruction is now available over wide range of light- and electron-beam imaging systems without significant hardware modification and complicated computation [11, 12].

TIE has its basis in the relationship between phase and the first derivative of intensity along the optical axis. Therefore, an accurate estimate of intensity derivative should be obtained in order to solve the partial differential equation. However, the intensity derivative along the optic axis cannot be directly measured. Conventionally, the TIE approach involves recording two out-of-focus images, which usually being recorded symmetrically over- and

under-focused by the same focal step from the in-focus image [13–15]. These are then used to approximate the derivative of the image intensity in the direction of the beam propagation. However, the axial intensity derivative approximated by the finite difference method from two intensity measurements only achieves a maximum of second-order precision with respect to the separation of the adjacent intensity measurements under no noise condition [13, 15]. When noise exists which is natural, this method is rather vulnerable, necessitating large defocus distances in order to increase the signal-to-noise ratio (SNR) at the expense of phase resolution [14]. To overcome this difficulty, intensity measurements in multiple planes have been proposed to minimize the effect of the noise on the retrieved phase [16], reduce the impact of the higher order axial intensity derivatives [15, 17], or both [17, 18]. In addition, these methods have been extended to the situation in which the intensity is measured in unequally-spaced planes [19, 20]. While these strategies have been shown to work well in many situations, their performance, however, depends heavily on the noise level and the characteristics of the experimental data [18, 20]. With a given set of intensity measurements, making an appropriate choice of a suitable algorithm is difficult. Besides, a comprehensive framework to better understand and, potentially, to improve existing derivative estimation algorithms is still lacking.

The aim of this paper is to show that great insight can be gained into the properties, the shortcomings, and the performance of these existing intensity derivative estimation algorithms by unifying them into the category of Savitzky-Golay filter. Specifically, it is shown that all the above mentioned methods can be viewed as special cases of the Savitzky-Golay differentiation filter. Based on these findings, two novel phase retrieval algorithms that extend the applications of the Savitzky-Golay differentiation filters in TIE are introduced. The adaptive-degree Savitzky-Golay differentiation filter method makes possible the choice of the distinct filter degree in different part of the intensity image. The second algorithm, which we call it optimal frequency selection automatically chooses the best parts of the phases obtained from Savitzky-Golay differentiation filters with various degrees by frequency decomposition. Extensive numerical studies and experiments are carried out to confirm and validate the proposed approaches.

2. Problem formulation

2.1 Transport of intensity equation

With the paraxial approximation, the derivative of intensity in the light propagation direction contains phase information that can be retrieved via TIE [6]:

$$-k \frac{\partial I(\mathbf{r})}{\partial z} = \nabla \cdot [I(\mathbf{r}) \nabla \phi(\mathbf{r})], \quad (1)$$

Where k is the wave number $2\pi/\lambda$, \mathbf{r} is the position vector representing the spatial coordinates (x, y) . ∇ is the gradient operator over \mathbf{r} . z denotes the optical axis, perpendicular to the x - y plane. The intensity derivative $\partial I(\mathbf{r})/\partial z$ is commonly obtained by an intensity difference between two side planes, separated by Δz [13, 14]

$$\frac{\partial I(\mathbf{r})}{\partial z} \approx \frac{I(\mathbf{r}, \Delta z) - I(\mathbf{r}, -\Delta z)}{2\Delta z}, \quad (2)$$

This simple central difference formula gives the estimate of the derivative, where the error is the second order of the focus distance if the data are noise-free. Mathematically, the smaller the defocus distance is, the more accurate the Eq. (2) can be approximated to the ideal derivative. However, real measurements yield data with noise and discretization which does not allow Δz to be too small. Thus a *compromise* is made where Δz is chosen to balance the high-order (or non-linearity) error and the noise effect [14]. Specifically, the optimal Δz is dependent on both the maximum physically significant frequency of the object and the noise level [21]. A priori knowledge of these two aspects is difficult not be known in advance.

2.1 Multiple-plane schemes for derivative estimation

To improve the accuracy of derivative estimation, multiple-plane schemes are proposed by taking the non-linearity error or/and noise effect into account. The *high-order finite difference* method was first used by Ishizuka and Allman [15] and more recently generalized by Waller *et al.* [17]. For symmetric differencing, i.e., given a data set $I(\mathbf{r}, i\Delta z), i = -n, \dots, 0, \dots, n$. The high-order finite difference method using a Taylor expansion has the form

$$\frac{\partial I(\mathbf{r})}{\partial z} \approx \sum_{i=-n}^n \frac{a_i I(\mathbf{r}, i\Delta z)}{\Delta z}, \quad (3)$$

with coefficients [22]

$$a_i = \frac{i(-1)^{i+1} (n!)^2}{(n+i)!(n-i)!}. \quad (4)$$

It is also possible to vary the defocus distance Δz [19]. To do so, one replaces Δz by $\delta z = l\Delta z$ ($l \in \mathbb{N}$) and i by $i' = li$ in Eq. (3). By changing l , the defocus distance δz of different plane can be varied. In the following, we only consider the condition that the measured planes are spaced equally and symmetrically from the center plane.

Without considering the effect of noise, the estimation error of Eq. (3) is of order $O(\Delta z^{2n})$ by using $2n$ images. However, as with the two-plane method, one is faced with the unavoidable problem of noise and the quantitation error. To reduce the effect of noise, Soto and Acosta [16] proposed a *noise-reduction finite difference* formula which with coefficients

$$a_i = \frac{3i}{n(n+1)(2n+1)}. \quad (5)$$

The coefficients are derived by minimizing the noise effect (more specifically, the noise reduction ratio (NRR) [23] defined by $\sum_{i=-n}^n a_i^2$) while ignoring the effect of all high-order terms in the Taylor expansion. The high-order finite difference indicates that the planes close to the center have to assume much larger weight than the ones further away, while the noise-reduction method makes the coefficients to be distributed as evenly as possible. Therefore, these two methods are contradictory in a sense. To balance this contradiction, Bie *et al.* [18] presented a *higher order finite difference with noise-reduction* which blends these two methods together, taking both the higher-order terms and noise effect into account. For the data from $2n+1$ planes, the high-order terms are cancelled until m^{th} order ($m < 2n+1$), leaving some degree of freedom for noise suppression. In [18], the author did not give the explicit formula of these coefficients but showed the problem is well-posed when $m < 2n+1$.

Besides these explicit finite difference methods, Waller *et al.* [17] recommended using a *least-squares fitting* method to estimate the first order derivative. Because the least-squares fitting process weights all data equally, and the high-order polynomial fitting conserves the non-linear component of the original data, so both the higher order and noise effect can be treated simultaneously.

2.2 Main connections with Savitzky-Golay filters

After reviewing those multiple-plane schemes, one issue that arises is that it is not easy to determine which method should be used and is there any interrelation between these methods. Let us start with the Waller's least-squares fitting method [17]. Given $2n+1$ data points $\{I[i]\} = \{I_{i\Delta z} | -n \leq i \leq n\}$, the least-squares polynomial of degree m ($m < 2n+1$) has the form

$$P_m(i) = \sum_{k=0}^m b_k i^k. \quad (6)$$

In the normal least-squares fitting procedure, the coefficients b_k are uniquely obtained by solving the normal equations of the Vandermonde system which minimize the sum of square of the residual (SSR), defined as $SSR_m = \sum_{i=-n}^n \{P_m(i) - I[i]\}^2$. We then define a $2n+1$ by $m+1$ matrix $\mathbf{A} = \{\alpha_{i,j}\}$ as the matrix with elements $\alpha_{i,j} = i^j$, $-n \leq i \leq n$, $j = 0, 1, \dots, m$, the normal equations for the least-squares problem can be written in matrix form as

$$\mathbf{A}^T \mathbf{A} \mathbf{b} = \mathbf{A}^T \mathbf{I}, \quad (7)$$

Where vector $\mathbf{I} = \{I[-n], \dots, I[-1], I[0], I[1], \dots, I[n]\}^T$ and $\mathbf{b} = \{b_0, b_1, \dots, b_m\}^T$. The solution for the polynomial coefficients can thus be written as

$$\mathbf{b} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{I} = \mathbf{H} \mathbf{I}. \quad (8)$$

The above formula gives all the polynomial coefficients $\mathbf{b} = \{b_0, b_1, \dots, b_m\}^T$. But the term of interest here is only the polynomial at $z = 0$. Specifically, the first order derivative evaluated at $z = 0$ requires an expression for coefficient b_1 only. This allows Eq. (8) to be reduced to an expression of the form

$$b_1 = \sum_{i=-n}^n h_{1,i} I[i], \quad (9)$$

where $h_{1,i}$ is the element of the second row of \mathbf{H} . An important observation is that the matrix \mathbf{H} is independent of the input samples. Thus, we can think of the first order derivative estimation using least-squares fitting as a *shift-invariant discrete convolution* process, which share the *same format* with finite difference. Indeed as early as 1964, Savitzky and Golay [24] showed that fitting a polynomial to a set of input samples and then evaluating the resulting polynomial at a single point within the approximation interval is equivalent to discrete convolution with a fixed impulse response. The filters, obtained by this method are widely known (especially among chemists) as Savitzky-Golay filters [24]. More generally, to evaluate the s^{th} derivative at point t using a polynomial of degree m on $2n+1$ data points, the convolution weight at point i can be calculated as [25]

$$h_{s,i} = \sum_{k=0}^m \frac{(2k+1)(2n)^{\binom{k}{k}}}{(2n+k+1)^{\binom{k+1}{k+1}}} P_k^n(i) P_k^{n,s}(t), \quad (10)$$

where $(a)^{\binom{b}{b}}$ is a generalized factorial function $(a)(a-1)\dots(a-b+1)$, and $(a)^{\binom{0}{0}} = 0$, and the $P_k^n(t)$ is the Gram polynomials defined as

$$P_k^n(i) = \sum_{j=0}^k \frac{(-1)^{j+k} (j+k)^{\binom{2j}{2j}} (m+i)^{\binom{j}{j}}}{(j!)^2 (2m)^{\binom{j}{j}}}, \quad (11)$$

and

$$P_k^{n,s}(t) = \left(\frac{d^s}{dx^s} P_k^n(t) \right)_{x=t}. \quad (12)$$

We are now ready to establish the correspondence between the different finite difference methods and the Savitzky-Golay differentiation filter (SGDF), i.e. least-squares fitting method. For the case that the center point is the evaluated point, the variable t is set to zero. Let $s = 1$ for the 1st derivative, the weight becomes

$$h_{1,i} = \sum_{k=0}^m \frac{(2k+1)(2n)^{\binom{k}{k}}}{(2n+k+1)^{\binom{k+1}{k+1}}} P_k^n(i) P_k^{n,1}(0). \quad (13)$$

Then we let the degree $m = 2n$, through proper simplification. Equation (13) can be reduced to

$$h_{1,i} = \frac{i(-1)^{i+1} (n!)^2}{(n+i)!(n-i)!}. \quad (14)$$

Compare it with Eq. (4), we then have:

Observation 1: The high-order finite difference method corresponds to the SGDF with degree $2n$.

We now let the degree $m = 1$, Eq. (13) can be reduced to

$$h_{1,i} = \frac{3i}{n(n+1)(2n+1)}. \quad (15)$$

The next observation now appears evident; it establishes the correspondence between the Soto's finite difference formula with Savitzky-Golay filter.

Observation 2: The noise-reduction finite difference method corresponds to the SGDF with degree 1.

So far, we may instinctively perceive that there should be a connection between Bie's method with Savitzky-Golay filter, since the two parents it inherited from are both particular cases of SGDF. But we cannot just follow the above routine because there is no explicit formula for Bie's method. Here we give this observation first and detailed prove will be given in Appendix A.

Observation 3: The higher order finite difference with noise-reduction method corresponds to the SGDF with degree m ($m < 2n+1$).

Finally, we have to mention that the unequally spaced multi-plane methods [19, 20] fall in the class of SGDF without exception, since the Savitzky-Golay filter has also been generalized for unequally or non-uniformly spaced data as its offspring [23, 25]. Even the two plane method (Eq. (2)) can be viewed as a special case of SGDF.

2.3 Property of Savitzky-Golay differentiation filters

In last subsection, all of the above derivative estimation strategies used in TIE are unified by the SGDF - an equivalent convolution solution for differentiation of data by a least-squares polynomial fitting. Therefore, it is necessary to examine the property of SGDF first. The SGDF have many good properties [26]: (1) the convolution operation is quite straightforward and much easier to implement than the standard least-squares fitting; (2) the coefficients can be easily obtained using a look-up table generated from the explicit solution, or pre-calculated using existing routines; (3) most importantly, they are the optimal differentiation filters that minimize the NRR, subject to moments preservation constraints [23].

To quantify the frequency domain properties of SGDF, we plot the frequency response of different degree SGDF with data points $2n+1 = 31$ in Fig. 1(a). The SGDF yields equivalent results for central point differentiation using odd polynomial degree and the next highest even degree, e.g., 1/2, 3/4, etc. The coefficients (impulse response) are anti-symmetric with the central weight always null. Therefore, the SGDF belongs to the type III FIR filter [23] that has a linear phase. With frequency response decomposition, a SGDF can be regarded as a combination of a low-pass filter and an ideal derivative filter:

$$H_{SG}(e^{j\omega}) = j\omega \frac{H_{SG}(e^{j\omega})}{j\omega} = H_{ideal}(e^{j\omega})H_{LP}(e^{j\omega}). \quad (16)$$

Similarly, a differentiation filtering on the signal can be regarded as calculating the 1st derivative on a smoothed version of the data. From the frequency response (in log scale) of the low-pass filter plotted in Fig. 1(b), it can be seen the filters have very flat frequency

response in their pass bands with modest attenuation in their stop bands. The frequency response of the SGDF is flatter and closer to the ideal derivative filter at low frequencies as the polynomial degree increases. The increase of the flatness is coincident with a relatively higher cut-off frequency or a wider low-pass range. Conversely, as the degree of fitting polynomial decreases, the effect of low-pass filter becomes more evident.

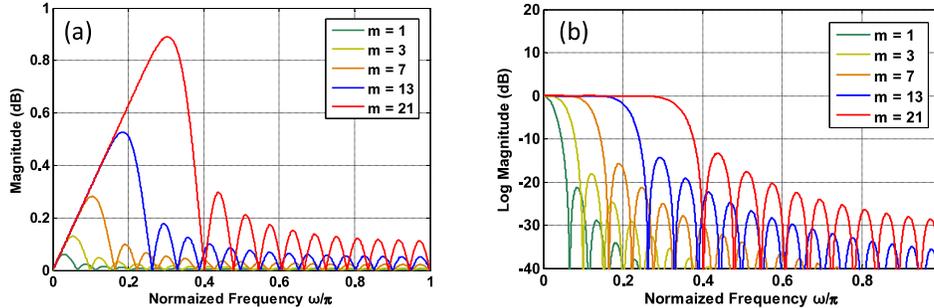


Fig. 1. Frequency characteristics of SGDF for $n = 15$ and various polynomial degrees; (a) frequency response of the SGDF filter; (b) the frequency response of the low-pass filter behind the SGDF filter in Log magnitude.

Another important issue is the accuracy of the derivative estimate using SGDF. For this prior information about the ideal signal is necessary to draw a reliable conclusion [16]. The Gauss–Markov theorem states that the m^{th} degree SGDF is an unbiased estimator of the derivative, and can achieve the Cramer-Rao lower bound if the signal can be perfectly modeled by an m -order polynomial and the observed data are with independent Gaussian white noise [26]. However, in practice, the signal order is unknown and hence using a non-optimal degree will inevitably affect the accuracy and the effect of noise reduction. So the problem now we facing is to choose a polynomial with proper degree which fit a data set best. This problem appears new but is in fact similar to one mentioned in subsection 2.2. Selecting a suitable multi-plane method is similar to determining the degree of the SGDF.

3. Applications of SGDF in phase retrieval by TIE

3.1 Derivative estimation using adaptive-degree SGDF

To find the optimal degree of polynomial, one choice is to apply SGDF using various polynomial degrees to the same data set, and then to choose one or some combination. However, it is difficult to decide which polynomial degree yields the optimal data accuracy. Consider the case when the tested sample is a pure-phase object, then Eq. (1) reduces to a standard 2D Poisson equation [17, 27, 28]:

$$-\frac{k}{I(\mathbf{r})} \frac{\partial I(\mathbf{r})}{\partial z} = \nabla^2 \phi(\mathbf{r}). \quad (17)$$

It can be seen the intensity derivative along z -axis is proportional to the second derivative of the phase distribution. Therefore, to best estimate the derivative, the degree of SGDF should be pixel-dependent, rather than fixed for the whole field. The filtering residual SSR_m cannot be used as criterion to choose the polynomial order because the polynomial of degree $k+1$ will always fit at least as well as than the polynomial of degree k , and it approaching a perfect fit for a polynomial order which is one less than the number of data points. So we need to introduce the concept of degrees of freedom (DOF). The DOF associated with each SSR is determined by the data size $N = 2n+1$ and the number of parameters m in the model; i.e., the DOF of SSR_m is $(N-m-1)$. In statistics, the analysis of variance (ANOVA) method states that the optimal degree of polynomial should be chosen for which the variance,

as computed by $SSR_m/(N-m-1)$, no significant decrease as the degree of polynomial is increased. To do this, a successive ANOVA F-tests procedure can be used [29]:

$$F = \frac{(SSR_{m_1} - SSR_{m_2})/(m_2 - m_1)}{SSR_{m_2}/(N - m_2 + 1)}, \quad (18)$$

where the polynomial degree $m_2 > m_1$. The resulting F value is tested against critical values of $F(m_2 - m_1, N - m_2 + 1)$, available in standard statistical tables, at a preset probability level, usually either 5% or 1%. If $F \geq F(m_2 - m_1, N - m_2 + 1)$, it means order m_2 gives a statistically significantly better fit to the data than order m_1 , otherwise we prefer the order m_1 . Since the SGDF yields equivalent results for the order 1/2, 3/4, 5/6, We can choose a starting polynomial degree at 1, and then the polynomial degree increases by 2 in each step of test. The test is terminated until the significance test fails or the highest degree $2n - 1$ reached. To reduce the computation-load, the SSR can be calculated using Gram polynomials based on previous values as [29]

$$SSR_m = SSR_{m-1} - \sum_{i=-n}^n P_m^n(i) I[i] / \sum_{i=-n}^n [P_m^n(i)]^2, m > 1. \quad (19)$$

The 1st order smoothing for SSR_1 can be performed as a convolution using Savitzky-Golay smoothing filter. All entries of filter coefficients and the value of Gram polynomials can be calculated and stored as look-up-tables beforehand.

3.2 Phase retrieval by optimal frequency selection

The adaptive-degree SGDF can give more accurate derivative estimates by finding a polynomial that fits the data of each pixel best. However, it does not guarantee a more accurate retrieved phase by applying TIE to this derivative estimate. In fact, the adaptive-degree SGDF usually gives an unsatisfactory phase result with severe patch effects - the result resembles a composite image that puts different pieces of estimated phases using different orders together (see Sections 4 and 5 for details), which is obviously not what we want. Because previously we concentrated our attention on the derivative estimation only, without considering the phase retrieval process using TIE. In this subsection, we present an alternative method which solves the optimal phase retrieval problem in spatial frequency domain.

Considering a thin object located at plane $z = 0$, illuminated with a monochromatic plane wave, the complex wave field in the focused plane can be written as (with no loss of generality, a single transverse dimension is used for notational simplicity)

$$U(x, 0) = |U(x, 0)| \exp[i\phi(x)] = \exp[i\phi(x) - \mu(x)], \quad (20)$$

where $\mu(x) = -0.5 \ln |U(x, 0)|^2$. By applying the weak phase object approximation, the intensity image captured at z can be represented by [14, 28, 30]

$$I(x, z) = |U(x, z)|^2 = 1 - 2\mu(x) * \mathcal{F} \cos(\pi\lambda z u^2) - 2\phi(x) * \mathcal{F} \sin(\pi\lambda z u^2). \quad (21)$$

The intensity derivative estimated by a fixed-degree SGDF with $2n + 1$ intensity measurements is

$$\frac{\partial I(x)}{\partial z} \approx \sum_{i=-n}^n \frac{a_i I(x, i\Delta z)}{\Delta z} = \frac{\sum_{i=-n}^n a_i - \sum_{i=-n}^n a_i \{2\mu(x) * \mathcal{F} \cos(i\Delta z \pi \lambda u^2) + 2\phi(x) * \mathcal{F} \sin(i\Delta z \pi \lambda u^2)\}}{\Delta z}. \quad (22)$$

As we mentioned in Sec. 2.3, the coefficients of SGDF $\{a_i | -n \leq i \leq n\}$ are odd-symmetric with the central weight always null. Using these properties, the Fourier representation of Eq. (22) can be reduced to

$$\mathcal{F} \frac{\partial I(x)}{\partial z} \approx -\frac{2}{\Delta z} \mathcal{F} \phi(x) \sum_{i=-n}^n [a_i \sin(i\Delta z \pi \lambda u^2)] = \frac{2}{j\Delta z} \mathcal{F} \phi(x) H_{SG}(e^{j\omega}) \Big|_{\omega=\Delta z \pi \lambda u^2}, \quad (23)$$

where $H_{SG}(e^{j\omega})$ is frequency response of the SGDF. Note the frequency response is simply the Fourier transform of the system's impulse response, and here we use the odd-symmetry property again. If we assume that the SGDF forms an ideal 1st derivative, i.e. $H_{SG}(e^{j\omega}) \Big|_{\omega=\Delta z \pi \lambda u^2} = j\Delta z \pi \lambda u^2$, this gives

$$\mathcal{F} \frac{\partial I(x)}{\partial z} = 2\pi \lambda u^2 \mathcal{F} \phi(x), \quad (24)$$

This equation coincides with the Fourier representation of TIE equation for the case of uniform unit intensity [17, 27, 28]. However, an ideal 1st derivative can never be achieved due to the discrete nature of SGDF. Therefore, the actual phase we obtain using SGDF is

$$\hat{\phi}(x) = \mathcal{F}^{-1} \left\{ \mathcal{F} \phi(x) \frac{H_{SG}(e^{j\omega}) \Big|_{\omega=\Delta z \pi \lambda u^2}}{j\Delta z \pi \lambda u^2} \right\} = \phi(x) * \mathcal{F}^{-1} \left\{ H_{LP}(e^{j\omega}) \Big|_{\omega=\Delta z \pi \lambda u^2} \right\}, \quad (25)$$

which is a low-pass filtered version of the ideal phase $\phi(x)$. Interestingly, the frequency response of the low-pass filter $H_{LP}(e^{j\omega})$ is exactly the same as Eq. (16). Equation (25) suggests that the frequency characteristic of the retrieved phase $\hat{\phi}(x)$ using SGDF is determined by the frequency response of the *low-pass filter that is implicit in the SGDF*. Now, reconsidering Fig. 1(b), which indicates the effect of low-pass filter decreases with increasing degree of SGDF, one may think that a higher-degree SGDF gives a more intact phase (with wider range of spatial frequencies can be accurately recovered) than a lower degree one. However, the higher degree SGDF results in larger NRR than the one assumes a smaller order. In another word, smaller order filters always give estimates of the low-frequency phase with higher SNR, but they suffer from the problem of not having enough response for high frequency phase variations. The above discussion reaffirmed that using a single degree of SGDF cannot offer the optimal solution and a tradeoff between phase information at high and low spatial frequencies is necessary.

Table 1. Normalized 0.3dB Cutoff Frequencies as a Function of m and n ($n = 1\sim 10$)

n	Filter Degree m									
	1/2	3/4	5/6	7/8	9/10	11/12	13/14	15/16	17/18	19/20
1	0.144	-								
2	0.078	0.330	-	-	-	-	-	-	-	-
3	0.055	0.209	0.439	-	-	-	-	-	-	-
4	0.042	0.156	0.302	0.510	-	-	-	-	-	-
5	0.034	0.126	0.236	0.370	0.560	-	-	-	-	-
6	0.029	0.105	0.196	0.298	0.422	0.597	-	-	-	-
7	0.025	0.091	0.167	0.252	0.348	0.462	0.627	-	-	-
8	0.022	0.080	0.146	0.219	0.298	0.388	0.496	0.651	-	-
9	0.020	0.071	0.130	0.194	0.262	0.337	0.422	0.524	0.671	-
10	0.018	0.064	0.117	0.174	0.235	0.299	0.370	0.451	0.548	0.688

Rather than choosing the optimal filter degree on a pixel-by-pixel basis, here we solve the optimal phase retrieval problem in the spatial frequency domain. The basic idea is to extract the best frequency components of the phase images obtained from SGDFs with various degrees and then recombine them into a composite phase. Considering the phase information lies within a certain bandwidth of spatial frequencies, such that it can be correctly reconstructed by both a smaller degree SGDF and a higher degree one. The smaller degree filter is preferred for its low NRR, or in another word, higher SNR. The frequency components of higher degree SGDF is only used when the smaller degree filter fails to

retrieve the phase information reliably. Thus, a rigorous definition of ‘*correct reconstruction*’ or ‘*reliable retrieval*’ is necessary.

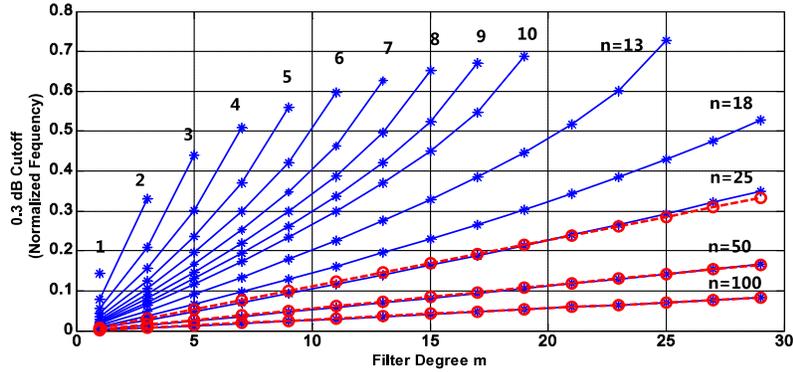


Fig. 2. Normalized 0.3 dB cutoff frequencies $f_c = \omega_c / \pi$ for different data half-length n and polynomial degree m .

Most frequently, a 3dB-point is chosen as a boundary in a filter's frequency response at which the signal through the system begins to be attenuated rather than passing through. In our case, we prefer a stricter boundary point at 0.3dB - amplitude of the filter falls to the 96.6% of the pass band. This means all spatial frequencies of the phase below 0.3dB cutoff frequency of the SGDF with one specific degree can be considered as *reliable retrieval*. The 0.3dB-points for filters with different n for odd polynomial orders m are displayed in Fig. 2 (remember the SGDF yields equivalent results for the order $1/2, 3/4, 5/6, \dots$). The points marked with * and connected by a blue line are the measured cutoff frequencies for a fixed value of half data length n . The values for short length filters are given more precisely in Table 1. For large n , it is impractical to list all the 0.3dB values due to the limited space here. Fortunately, we found in all cases in Fig. 2, f_c varies almost linearly with m when $m \ll 2n$ with the slope being dependent inversely on n , but the curves for $n < 25$ tend to deviate from a straight line as m increases. However, when m is large, e.g. 50 and 100, the linear region of the curve coincides with the range of usable values of m , so a nearly linear relation holds over a wide range of m . A reasonably accurate approximation to this behavior for the indicated range of parameters is

$$f_c = \frac{m - 0.3}{3.5n - 1}, \quad n \geq 25. \quad (26)$$

The values of f_c predicted by this equation are marked with a circle and connected by a red dash line. It shows that this simple formula fits the true values quite well even for the case $n = 25$ where the measurements deviate only slightly from the straight line over the entire range of m . The formula does not fit as well for values of m less than 25. However, the dependence of f_c on m is still linear except for small n . For $10 < n < 25$ and m suitably restricted, a formula similar to Eq. (26) with 1 in the denominator replaced by 2.5 gives more accurate predictions. While a more complicated functional form based on more measurements could provide more accurate predictions over a wider range of m and n , Eq. (26) should be adequate for most applications.

To decompose and reconstruct the retrieved phase in frequency domain according to the cutoff frequency obtained, a complementary filter bank can be designed with the frequency response $H_m(e^{j\omega})$, $m = 1, 3, \dots, 2n - 1$ and their amplitudes sum to unity. To achieve a flat response in the pass band and a fast roll off, the $H_m(e^{j\omega})$ can be in the form of high-order Butterworth filters or even ideal ones. For $m = 3, 5, \dots, 2n - 3$ the filters $H_k(e^{j\omega})$ are band-pass and have the pass band from f_c^m to f_c^{m-1} , where f_c^m represents the 0.3dB cutoff of m^{th}

degree SGDF. For $m=1$ the filter is a low-pass with the cutoff frequency f_c^1 and for $m=2n-1$ the filter is a high-pass with the cutoff frequency f_c^{2n-1} . The cutoff frequency of SGDF can be obtained by referring to Table 1, or using Eq. (26) for larger n . The phase retrieved by the m^{th} order SGDF $\hat{\phi}_m(x)$ is then filtered by $H_m(e^{j\omega})$

$$\phi'_m(x) = \hat{\phi}_m(x) * \mathcal{F}^{-1} \left[H_m(e^{j\omega}) \Big|_{\omega=\Delta z \pi \lambda u^2} \right] = \mathcal{F}^{-1} \left\{ \mathcal{F} \left[\frac{\partial I(x)}{\partial z} \right]_m \frac{H_m(e^{j\omega}) \Big|_{\omega=\Delta z \pi \lambda u^2}}{\Delta z \pi \lambda u^2} \right\}, \quad (27)$$

and the final reconstructed phase $\phi'(x)$ is their summation

$$\phi'(x) = \phi'_1(x) + \phi'_3(x) + \dots + \phi'_{2n-1}(x), \quad (28)$$

where the $[\partial I(x)/\partial z]_m$ is the intensity derivative estimate using m^{th} degree SGDF. The filter can either be applied to the phase obtained by TIE or together with the inverse Laplacian when solving the TIE in Fourier domain. A flowchart schematic of the whole procedure, which we call as optimal frequency selection (OFS) is shown in Fig. 3. Note usually the final reconstructed phase is not necessarily a composite of all possible orders because of the sampling effect and the limited physically significant frequency of the object. Therefore, we recommend the summation of Eq. (28) begins from $\phi'_1(x)$, $\phi'_3(x)$, ... and stops when the mean absolute of $\phi'_m(x)$ is below a given value, or the cutoff frequency with respect to u (i.e. $\sqrt{f_c^m/\Delta z \lambda}$) exceeds the maximal range of the frequency coordinate in the Fourier domain corresponding to x . Finally, to reduce the computation load when n is large; the filter order can be increased with a step of four rather than two without sacrificing the accuracy too much.

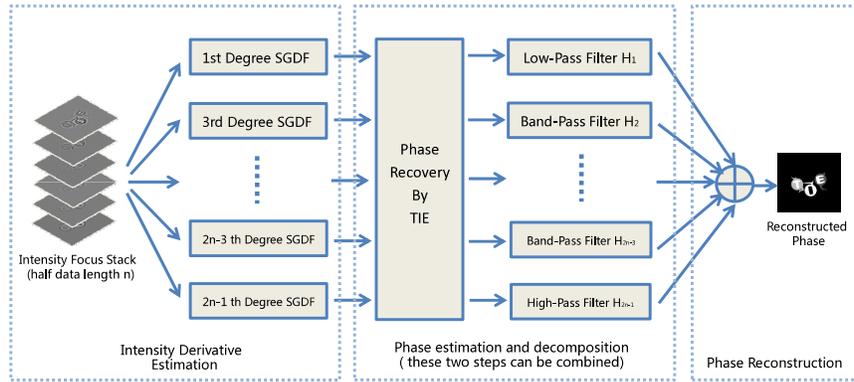


Fig. 3. A flowchart schematic of the OFS algorithm.

4. Simulations

To test the performance of the algorithm, we simulated the propagation of a complex field of a pure phase profile of the letters 'TIE' (shown in Fig. 4(a)) which was defined on a grid with 256×256 pixels with a pixel size of $2\mu\text{m} \times 2\mu\text{m}$. The wavelength is 632.8nm . Fifty one intensity images separated by $10\mu\text{m}$ were generated with defocused range between and $-250\mu\text{m} \sim 250\mu\text{m}$. To simulate noise effect, each defocused image was corrupted by Gaussian noise with standard deviation of 0.002. The Poisson noise case has been also tested on the same data. The results did not differ significantly from the case of Gaussian and hence not discussed further in this work.

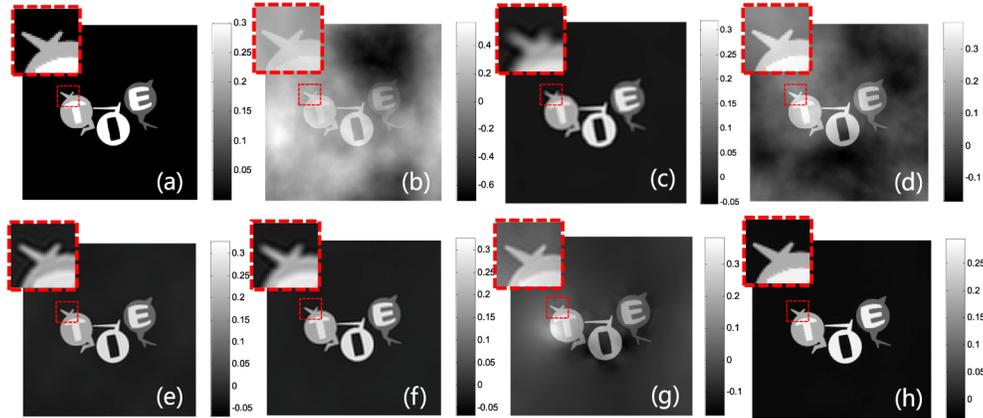


Fig. 4. Phase recovery comparison of the synthetic noisy test data; (a) true phase, (b) traditional TIE using two planes separated by $20\mu\text{m}$ (RMSE = 0.2552); (c) 1st fixed degree SGDF (RMSE = 0.0128); (d) 31st fixed degree SGDF (RMSE = 0.0651); (e) 7th fixed degree SGDF (RMSE = 0.0092); (f) traditional TIE using two planes separated by $100\mu\text{m}$ (RMSE = 0.0115); (g) adaptive-degree SGDF (RMSE = 0.0453); (h) OFS (RMSE = 0.0020); the square areas with red dot lines are magnified for clarity.

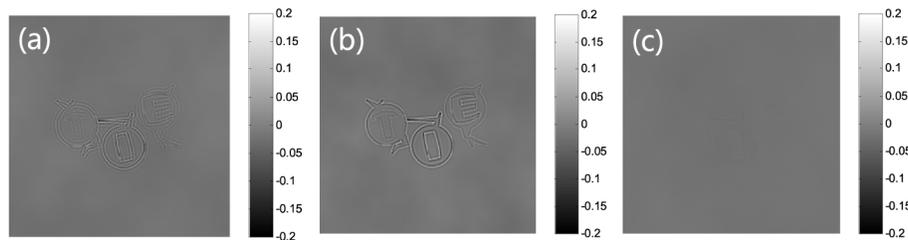


Fig. 5. Phase error images for (a) 7th fixed degree SGDF, (b) traditional TIE using two planes separated by $100\mu\text{m}$, (c) OFS.

The phase retrieval results of different methods are shown in Fig. 4. The metric used to measure the accuracy of phase retrieval is given by the root mean square error (RMSE), which quantify the overall difference between the ideal reference phase and the retrieved phase. The traditional TIE method estimated the derivative using two images separated by $10\mu\text{m}$ (Figs. 4(b)). The low-frequency artifacts superimposed on the acquired phase significantly obscures the object itself. The 1st-order SGDF (Soto's noise-reduction finite difference method) showed a much cleaner result with far fewer low-frequency artifacts. However, as is clearly shown in the magnified area, the fine structural details were blurred. It is expected because a lower polynomial degree is more effective in removing noise, but at the expense of distorting small details of the signal too much. Conversely, the 31st order SGDF showed a cloudy image but highlights the high frequency sharp object edges. The results presented in Figs. 4(c) and 4(d) indicate an improper choice of polynomial degree can lead to reduction in phase details if the polynomial degree is too low or sub-optimal removal of low-frequency noise if the polynomial degree chosen is unnecessarily high. To compromise between these opposing trends, we performed a brute-force search and found using a degree between 5 and 11 gave overall visually acceptable results both in resolution and contrast. The lowest RMSE with fixed order SGDF achieved is 0.0092, when the order was 7 (Figs. 4(e)). Similarly, we found that using two images separated by $100\mu\text{m}$ gave lowest RMSE (0.0115) for the two-plane method (Figs. 4(f)). But by inspecting their error images (shown in Figs. 5(a) and 5(b)), we can see that they still introduced some edge blurring and/or low-frequency artifacts. The adaptive-degree SGDF method gave a result with clean background and fine

phase details (Fig. 4(g)). However, as mentioned earlier, the patch-like artifacts deteriorated the result significantly, resulted in a high RMSE. However, the proposed OFS method provided a faithful result with uniform background and sharpest edges (Figs. 4(h)). The residual error was rather low so that we can hardly perceived in Fig. 5(c). These results are verified by the RMSE of each image.

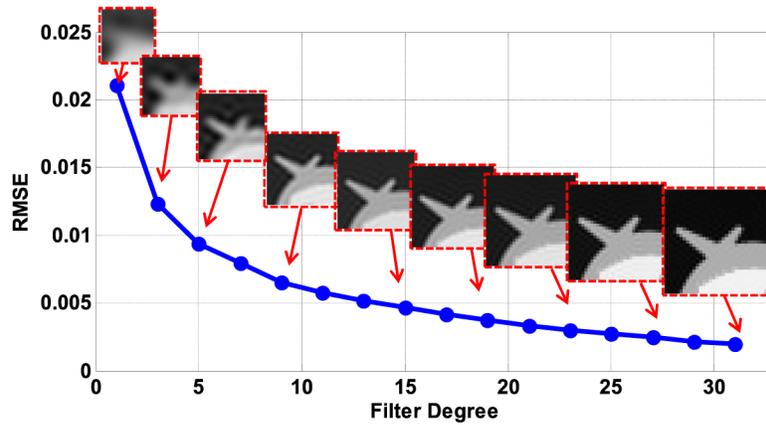


Fig. 6. (Media 1) The curve of RMSE evolution, shown with corresponding magnified areas.

To better explain the low error achieved by the OFS method, we show the RMSE curve versus the filter degree in Fig. 6. A video sequence (Media 1) is also presented to show the evolution of the phase reconstruction process. It can be seen the phase resolution gradually improved by adding high-frequency information corresponding to the optimal frequency bands in the higher order results. The reconstruction process converged when degree $m = 31$, so the final result was a composite of phases from the 1st to 31st orders. For simplicity, an ideal band-pass filter bank was used to decompose and extract the optimal spatial frequency component of each phase recovered. The Gibbs phenomenon combined with the diffraction ringing effect was clearly visible during the first few steps in reconstruction as ripples extending beyond the object region, but it faded out rapidly as the filter degree increases. The final reconstructed result shows a rather clean phase with fairly high resolution, which proves that our method automatically sieve the best parts of result from different orders without any over-fitting or under-fitting problems.

To test the noise adaptability of the proposed OFS method, we varied the standard deviation of the noise from 0 and 0.005, and the final RMSE curves versus standard deviation of the noise for different methods are shown in Fig. 7. For each noise level, the simulation was repeated 20 times and then the RMSE values were averaged to reduce data uncertainty. It can be seen when there is no noise, the higher the order of SGDF, the less the RMSE achieved because the derivative non-linearity is the only source of error. In the presence of noise, even though relatively small, the error of high-order SGDF (e.g. $m = 23, 31$) rise dramatically, indicating they are very sensitive to noise. The lower order SGDFs perform comparatively better, with varying degrees of noise resistibility. When the non-linearity error and the noise effect blend together, a compromise between the two opposing trends is more desirable. As the noise level further increases, the phase error is dominated by noise effect, so the lower degree SGDFs perform better. The 1st degree filter shows a rather good resistibility with an almost flat curve.

The OFS method always performs better than other methods using a fixed order because it chooses the frequency-bands optimally with highest SNR in the results of different degrees. Besides, the curve of OFS is also rather flat despite of a slight increase with the noise variance, showing its great adaptability in noisy situations. From these results, we can safely conclude that the OFS method removes the guesswork in the choice of polynomial degree and

significantly outperforms the fixed-order method even with the best selection of polynomial degree.

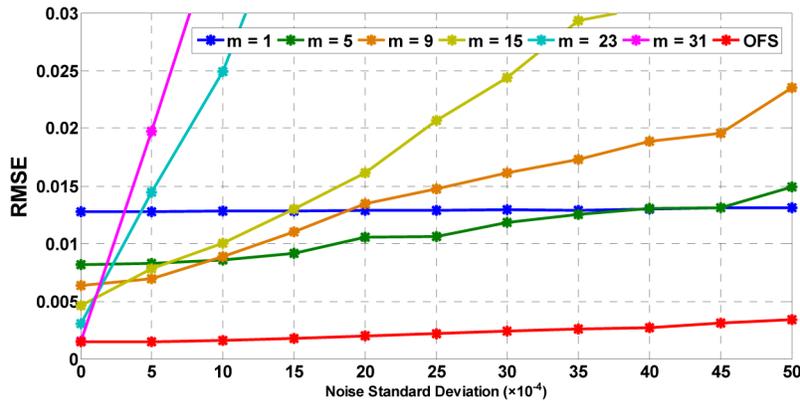


Fig. 7. Comparison of the RMSE for the estimated phase at different noise levels. The fixed degree SGDF uses $m = 1, 5, 9, 15, 23,$ and 31 orders.

5. Experiments

The experimental test setup is shown in Fig. 8(a). A He-Ne laser ($\lambda = 632.8\text{nm}$) source is expanded and collimated and then illuminates the object under test. The object is reimaged onto a CCD via a $4f$ system - two lenses of focal length $f = 25$ mm separated by the distance $2f$, and the distance from the object to the first lens is f . The camera is set on a translation stage in order to modify the defocus distance. The phase object under test is a geometry pattern etched on PMMA substrate, which is also shown in Fig. 8(a). Fifty-one images were captured with an equal separation of $50\mu\text{m}$ (Media 2). Some data samples from the intensity stack are shown in Fig. 8(b). All these images are recorded by a monochrome CCD imaging device (The Imaging Source DMK 41AU02, $4.65\mu\text{m}$ pixel size) and digitally processed using MATLAB.

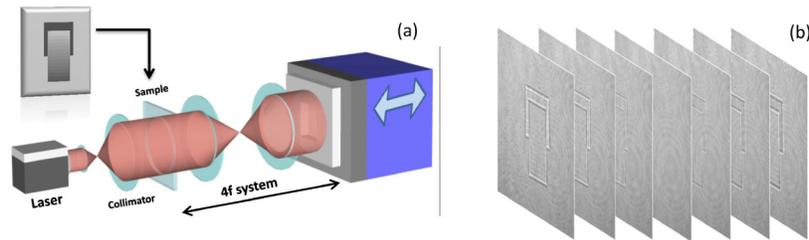


Fig. 8. (a) The experimental test setup, and (b) (Media 2) intensity images from the focus stack.

The phases recovered by different methods are shown in Fig. 9, which are similar with our simulation results. Smaller-order filters effectively suppress noise, at the expense of distorting smaller scale phase features. Higher-order filters preserve smaller scale features well, with the disadvantage of being less effective at suppressing low-frequency noise. By subjective evaluation of the reconstructed phase, it is noticed that the 5th degree SGDF produces best visual effect with regard to the contrast and resolution (Fig. 9(c)). Similarly, the optimum distance for traditional TIE (two-plane method) was found to be $700\mu\text{m}$ (Fig. 9(d)). However, either the background was too noisy or the phase resolution was sacrificed in their estimated phase (see Figs. 9(b)-(f)). The adaptive-degree SGDF method composited the phase in spatial domain, resulting in an unnatural result with contrast reversal (Figs. 9(g)). Figure 9(h) shows the reconstructed result by OFS, with the evolution process animated in Media 3. The final

phase of OFS combined the results of different degrees up to 27, indicating that not only the tiny details were detected but also the overall contrast of the phase was well preserved, despite slight non-uniform distribution in the background (which may be partially caused by the dust particles on the sample and/or the lens).

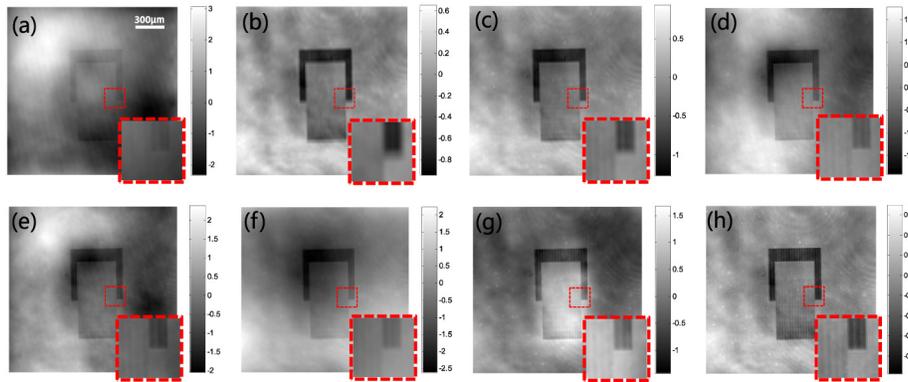


Fig. 9. Phase recovery comparison of a test phase object; (a) traditional TIE using two images separated by $100\mu\text{m}$; (b) 1st fixed degree SGDF; (c) 5th fixed degree SGDF; (d) 15th fixed degree SGDF; (e) 27th fixed degree SGDF; (f) traditional TIE using two images separated by $700\mu\text{m}$; (g) adaptive-degree SGDF; (h) OFS (Media 3); the bottom right red squares show the corresponding enlarged regions.

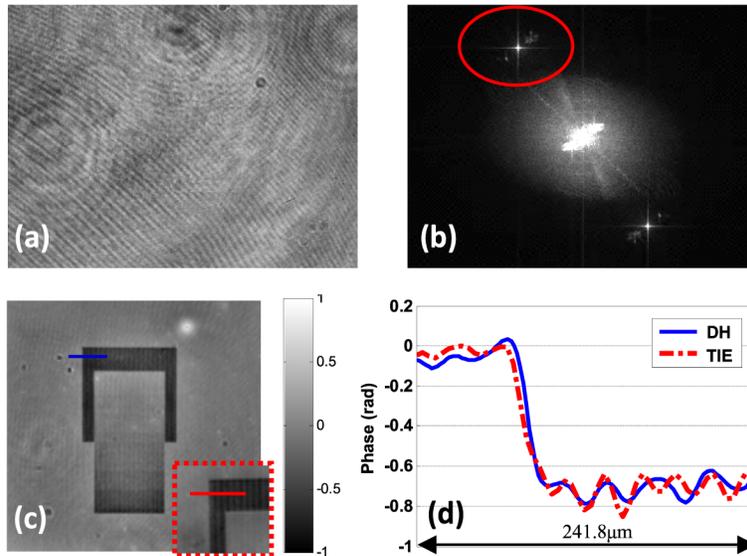


Fig. 10. Comparison between the phases retrieved from digital holography and TIE; (a) hologram of the same test sample; (b) hologram Spectrum; (c) reconstructed phase from digital holography (the red dot square at bottom right shows a corresponding area of Fig. 9(h)); (d) cross-sections of the phase marked with the red and blue lines in (c).

To better quantify the phase retrieved by the proposed method, we also measured the same sample using a digital holography (DH) microscope based on the Michelson interferometric configuration with a 4x objective ($\text{NA} = 0.13$) [31]. Figure 10 shows the hologram (a) and its spectrum (b). The phase demodulation was done by first filtering one of the first orders (the red circle area) from the hologram spectrum and then reconstructing the hologram in the traditional manner [31]. The phase curvature was physically compensated by introducing the same spherical phase curvature in the reference beam [32]. Figure 10(c)

shows the recovered phase. It can be seen that the reconstructed phase surface using the OFS method agrees well with the phase measurements made by DH. To clearly compare the phase profiles, a magnification of the corresponding area from Fig. 9(h) is also shown in Fig. 10(c), in the red dot line rectangle. Figure 10(d) shows a comparison of phase profiles taken across the sample edge showing the phase jump obtained with holography and with our TIE approach identified by blue continuous and red dashed lines, respectively. Both two methods were capable of following the large phase step and fine phase variations (e.g. the vertical line nicks results from the imperfection in fabrication) of the sample quite accurately.

6. Conclusions and discussions

The contribution of this paper is two-fold. First, several existing strategies for estimating the axial intensity derivative in TIE from multiple intensity measurements have been unified by the SGDF, which not only provides different views and insights on known derivative estimation methods but, more important, it also suggests a new way of improving solution strategies from an angle of digital filter design in signal processing. Second, two advanced applications of SGDF in TIE have been proposed. The adaptive-degree SGDF method based on the statistical testing allows the automatic choice of the proper degree of SGDF in different parts of the intensity image. The OFS method, more desirably, decomposes, processes, and then recombines the phases using the SGDF with various degrees in spatial frequency domain. The introduction of OFS provides a powerful tool which, in all of our numerical simulations and experiments, outperforms both optimally chosen and sub-optimally chosen fixed-degree methods that are commonly used.

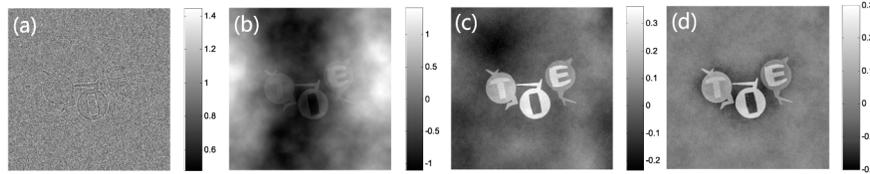


Fig. 11. Phase recovery results under significant noise (noise standard deviation 0.01); (a) intensity image with defocus distance $80\mu\text{m}$, (b) 7th fixed degree SGDF (RMSE = 0.6016); (c) OFS (RMSE = 0.0616); (d) OFS with Tikhonov-regularization (RMSE = 0.0332).

Although the OFS improves the noise resistibility and the phase resolution of TIE significantly, its performance also shares a common limitation of TIE based method - lower spatial frequencies are more sensitive to noise as compared with the higher spatial frequencies. Under *severe noisy conditions*, the low frequency noise is difficult to be entirely eliminated, because the inverse Laplacian operation is ill-conditioned at near zero-frequency. Figure 11 show the simulation results when the noise was increased to significant level (standard deviation 0.01). As shown in Fig. 11(a), the underlying intensity signal was submerged in the noise, so that even the 7th fixed degree SGDF can no longer render a visible result (Fig. 11(b)). The OFS reduced the noise significantly, but the residual low-frequency noise remained obvious (Fig. 11(c)). In this situation, proper regularization treatments [33, 34] can help to reduce the low frequency artifacts at the cost of attenuation in slow variations of the underlying true phase signal. By combining the Tikhonov-regularization with OFS, the low frequency noise further reduced, accompanied by some dark 'halos' appeared near edges (Fig. 11(d)). Another possible countermeasure is to increase the defocusing range of the intensity stack so that the cut-off frequency of 1st degree SGDF can be narrowed to near zero-frequency. Of course, this point is beyond the scope of the present work and needs more detailed investigation.

Appendix A: Proof of Observation 3

The optimization problem in higher order finite difference with noise-reduction method is to minimize NNR $\sum_{i=-n}^n a_i^2$, under constraints that (1) $\sum_{i=-n}^n a_i = 0$; (2) $\sum_{i=-n}^n a_i i = 1$; (3)

$\sum_{i=-n}^n a_i i^k = 0, k = 2, 3, \dots, m$. In [18], the authors solve this problem by matrix partition. Here we rewrite the problem as a quadratic programming problem:

$$\begin{cases} \min \frac{1}{2} \mathbf{a}^T \mathbf{a} \\ \text{s.t. } \mathbf{B}\mathbf{a} = \mathbf{c} \end{cases}, \quad (29)$$

where $\mathbf{B}\mathbf{a} = \mathbf{c}$ is written as

$$\begin{bmatrix} (-n)^0 & \cdots & (-1)^0 & 1 & 1^0 & \cdots & n^0 \\ (-n)^1 & \cdots & (-1)^1 & 0 & 1^1 & \cdots & n^1 \\ (-n)^2 & \cdots & (-1)^2 & 0 & 1^2 & \cdots & n^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ (-n)^m & \cdots & (-1)^m & 0 & 1^m & \cdots & n^m \end{bmatrix} \begin{bmatrix} a_{-n} \\ \vdots \\ a_{-1} \\ a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \quad (30)$$

Solving Eq. (29) using Legendre multiplier:

$$\begin{bmatrix} \mathbf{E} & -\mathbf{B}^T \\ -\mathbf{B} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \boldsymbol{\lambda}_L \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{c} \end{bmatrix}, \quad (31)$$

where $\boldsymbol{\lambda}_L$ is the Legendre vector and \mathbf{E} is identity matrix. The solution for the equations is:

$$\boldsymbol{\lambda}_L = (\mathbf{B}\mathbf{B}^T)^{-1} \mathbf{c} \text{ and } \mathbf{a} = \left((\mathbf{B}\mathbf{B}^T)^{-1} \mathbf{B} \right)^T \mathbf{c}. \quad (32)$$

Carefully examine the matrix \mathbf{B} , we can find it equals \mathbf{A}^T in Eq. (7). Therefore,

$$\mathbf{a} = \left((\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \right)^T \mathbf{c} = \mathbf{H}^T \mathbf{c}, \quad (33)$$

where \mathbf{H} is exactly the same one as in Eq. (8), and $\mathbf{H}^T \mathbf{c}$ extracts the second row from \mathbf{H} . Thus, the Observation 3 is established.

Acknowledgments

This project was supported by the Research Fund for the Doctoral Program of Ministry of Education of China (No. 20123219110016) and the Research and Innovation Plan for Graduate Students of Jiangsu Higher Education Institutions, China (No. CXZZ11_0237). The author Chao Zuo gratefully acknowledges the financial support from China Scholarship Council (No. 201206840009).