

the object itself. Thus, we aim not for the absolute optimal set of measurements, but rather for the optimal coverage of sensitivity across all spatial frequencies that pass through the imaging system, treating each equally.

Consider the goal of selecting a set of z planes such that the maximum phase measurement sensitivity (set by the value of the transfer function) is at least α for the largest range of spatial frequencies possible. In the following, we demonstrate an exponential spacing measurement scheme that achieves this goal with the least number of measurement required. The highest frequency that can be recovered is set by the diffraction limit as NA/λ , where NA is the numerical aperture of the imaging system. This will define the minimum defocus distance from which we can capture relevant information. We define $f = \pi\lambda(u^2 + v^2)$ and $f \leq \pi\lambda(\frac{\text{NA}}{\lambda})^2$. First, we select the minimum defocus distance z_1 such that the sensitivity $g(u, v, z_1)$ is α at frequencies corresponding to the maximum frequency allowed $f_1 = \pi\lambda(\frac{\text{NA}}{\lambda})^2$, the solution of which is $z_1 = \frac{\pi - \arcsin(\alpha)}{\pi(\text{NA})^2} \lambda$. Defocus steps smaller than z_1 will not provide useful information and are thus unnecessary, though in practice the minimum z step size may be set by the axial motion stage. As shown by the blue curve in Fig. 3, the sensitivity remains above α until $f_2 = \frac{\arcsin(\alpha)\pi(\text{NA})^2}{\lambda[\pi - \arcsin(\alpha)]}$. Next, we select the second defocus distance $z_2 = \beta z_1$, where $\beta = \frac{\pi - \arcsin(\alpha)}{\arcsin(\alpha)} > 1$, such that the sensitivity $g(u, v, z_2)$ is α at f_2 , and will remain at least α for a range of frequencies, as illustrated by the green curve in Fig. 3. By induction, the optimal measurement scheme that satisfies a minimum phase measurement sensitivity α should satisfy the following exponential relation for the defocus distances

$$z_{n+1} = \beta z_n. \quad (5)$$

The exponential spacing implies that a large z can be reached with far fewer measurements as compared to the equal-spacing measurement schemes. Large z images are crucial for transferring low-frequency phase information, and the exponential spacing scheme enables us to reach this without taking an excessive number of measurements or trading off high-frequency information. The larger the maximum z , the better the low-frequency result, and we desire to keep the minimum z as small as possible in order to maintain diffraction-limited phase resolution, both extremes being limited by the motion stage axial range. Accuracy can be traded off against number of images through the choice of α . Larger values results in more accurate phase retrieval, at the cost of requiring more images.

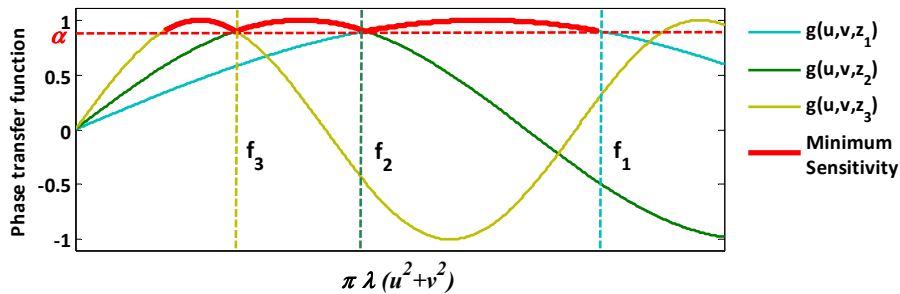


Fig. 3. Rationale for exponential spacing measurement scheme. Plot shows phase transfer functions for exponentially spaced z steps, which ensure a minimum sensitivity of α across a range of frequencies. $g(u, v, z_1)$, $g(u, v, z_2)$, and $g(u, v, z_3)$ are the phase transfer functions at z_1 , z_2 , and z_3 , respectively. The minimum sensitivity plot shows the frequencies which are transferred at the sensitivity higher than α by choosing z_1 , z_2 , and z_3 . Larger z brings more low-frequency sensitivity.

3. Simulations

We compare the performance of various phase recovery methods that use equal spacing and exponential spacing. In the simulation, the illumination wavelength is set as 632.8nm , and each image has 100×100 pixels (pixel size $2\mu\text{m} \times 2\mu\text{m}$). **Equally Spaced Stack** has 9 intensity images simulated with a constant defocus step size of $5\mu\text{m}$. **Exponentially Spaced Stack** also has 9 images; however, they are exponentially spaced, at z positions around focus of $\pm 5\mu\text{m}$, $\pm 20\mu\text{m}$, $\pm 80\mu\text{m}$ and $\pm 320\mu\text{m}$ (see Fig. 4). Although both data sets use the same number of images, the exponentially spaced data set contains more information about the low-frequency phase information because it has a higher z range.

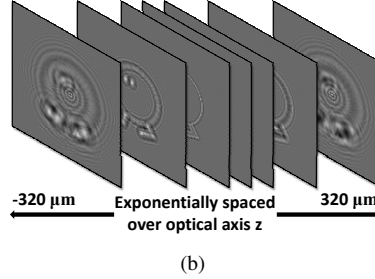


Fig. 4. Simulated data sets for equal and exponential spacing of z steps. (a) Equally Spaced Stack: images equally spaced by $5\mu\text{m}$. (b) Exponentially Spaced Stack: images exponentially spaced, with z of $\pm 5\mu\text{m}$, $\pm 20\mu\text{m}$, $\pm 80\mu\text{m}$, and $\pm 320\mu\text{m}$. The exponential spacing data contains more low-frequency phase information due to the larger range of z steps.

In order to assess the error performance, the intensity images of the equally and exponentially spaced stacks are corrupted by white noise with SNR ranging from 18.5 to 8 dB (noise variance from 0 to 0.02). Fig. 5 shows the average mean square error (MSE) of the recovered phase over 50 trials as SNR decreases. For the Equally Spaced Stack, higher order TIE performs significantly worse than SGDF TIE, and GP TIE is slightly better than SGDF TIE. For the Exponentially Spaced Stack, we show the results for two possible choices of Higher order TIE: $m = 9$, which performs better in low noise, and $m = 5$, which performs better in high noise. GP TIE with exponential spacing clearly exhibits the lowest MSE. This can be explained by the fact that the exponentially spaced data contains more low-frequency phase content in the measurements than the equally spaced data, and there is no trade-off between noise and nonlinearity. Figure 6 gives an example of the recovered phase at SNR of 11.1 dB .

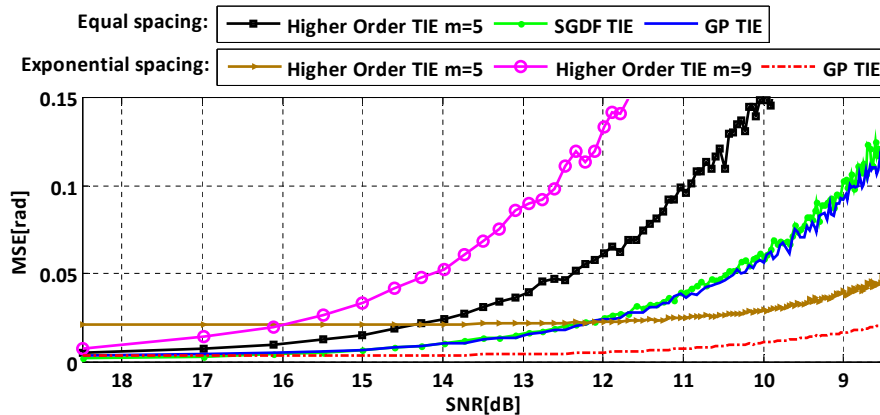


Fig. 5. Comparison of mean square error (MSE) in phase results for various methods as noise level increases. GP TIE with exponential spacing yields the best error performance.

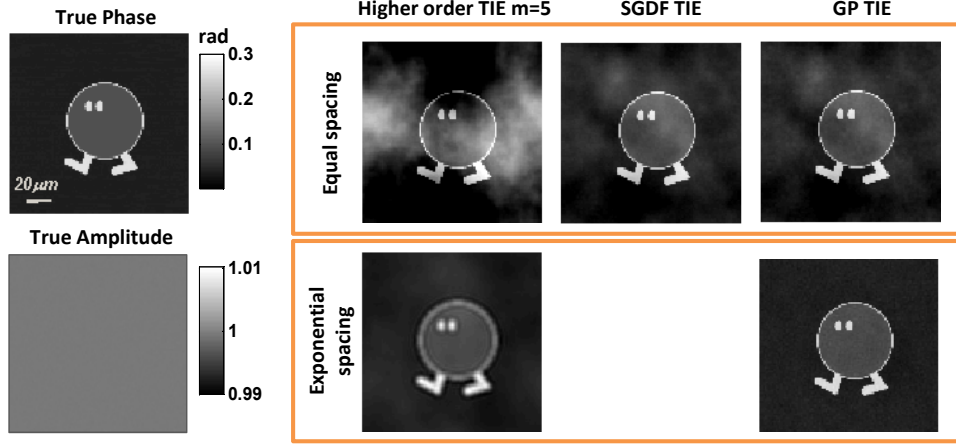


Fig. 6. Phase images recovered from simulated data with SNR of 11.1dB. (Top) Results of the Equally Spaced Stack, using higher order TIE ($m=5$, MSE 0.1194 in radian), SGDF TIE (0.0295), and the proposed GP TIE (0.0279). (Bottom) Results for Exponentially Spaced Stack, using Higher order TIE ($m=5$, MSE 0.0237) and GP TIE (0.0065). GP TIE using exponential spacing provides the best phase result.

4. Experimental results

We tested our method experimentally in a microscope (magnification 20x, NA=0.5) with filtered white light illumination (center wavelength 650nm, 10nm bandwidth). **Data Set 1** comprises 129 images of human cheek cells, equally spaced by a constant small step size $dz = 4\mu\text{m}$ over a large defocus range $[-252\mu\text{m} \text{ to } 252\mu\text{m}]$ (Fig. 8(a)). In Fig. 7, the GP fitted intensity spatial frequency variations along the propagation direction z for 3 different (u, v) values are shown. Both the measured and fitted curves follow nearly sinusoidal patterns, as predicted by Eq. (3). Note that the delay of each sinusoid is dependent on the absorption of the object at the corresponding spatial frequency. When the frequency (u, v) is high, the intensity spectrum variations diminish for large defocus distance (see plot of (u_{60}, v_{60}) in Fig. 7). This is due to the effect of partially coherent illumination, which will be the subject of future work.

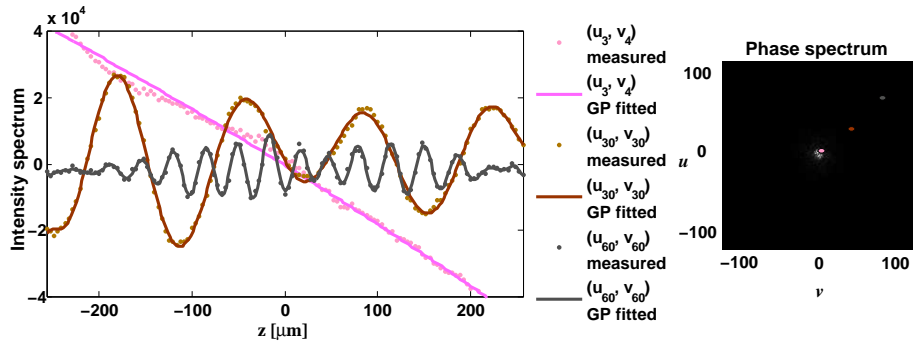


Fig. 7. (Left) GP regression of the intensity spectrum's real part over z for three sample frequency points (u, v) (Data Set 1). (Right) The frequency (u, v) of the three components depicted on real part of the recovered spectrum of phase (the image size is 945×888 but only the central part of the spectrum is shown for clarity). According to Eq. (3), the values for $\pi\lambda(u^2 + v^2)$ are $0.029 \times 10^4 m^{-1}$, $4.145 \times 10^4 m^{-1}$, and $17.299 \times 10^4 m^{-1}$, respectively.

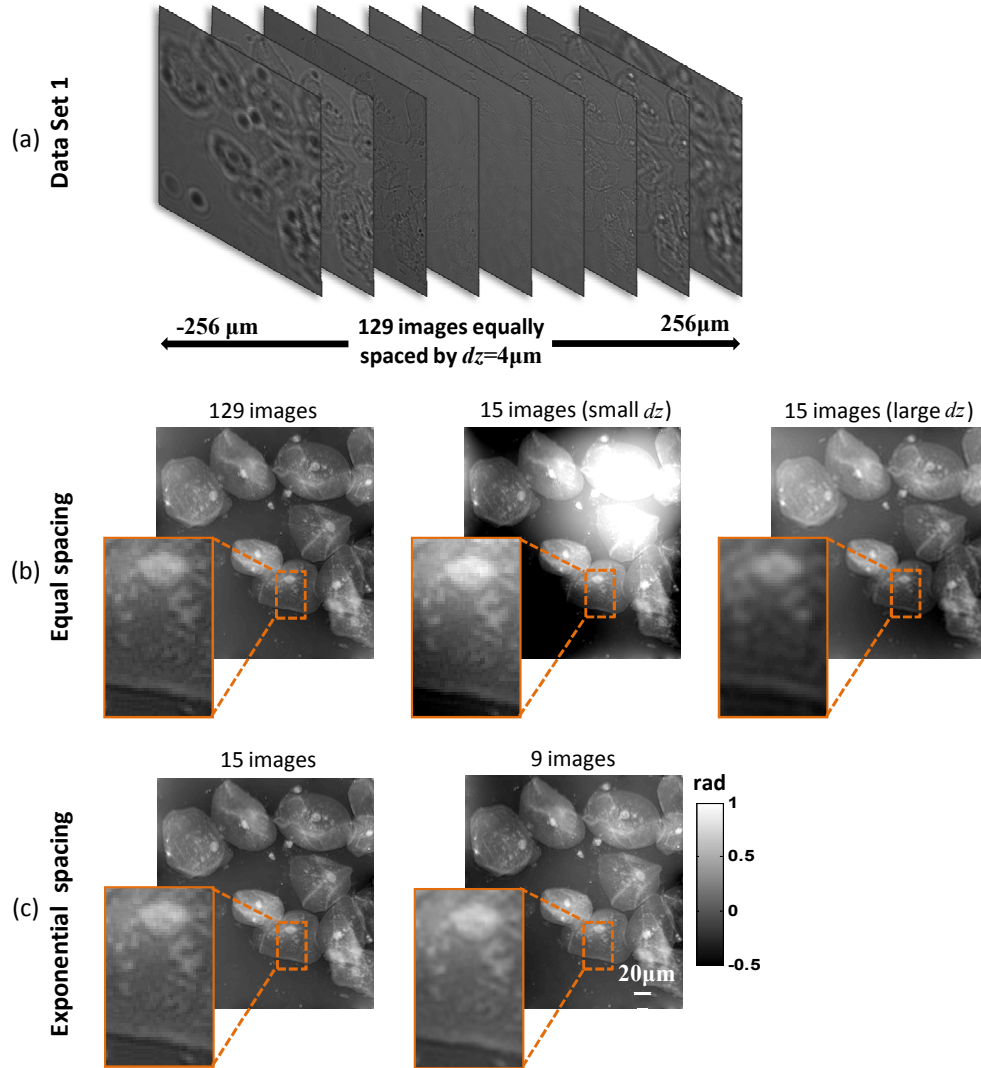


Fig. 8. Exponentially spaced defocus steps with GP TIE provide accurate phase results using less images. (a) Data Set 1: equally spaced $dz = 4\mu\text{m}$, from $-256\mu\text{m}$ to $256\mu\text{m}$. Each image has 945×888 pixels with effective size $0.31\mu\text{m} \times 0.31\mu\text{m}$. (b) Phase recovered with equally spaced z steps: (left) with all 129 images [$-256\mu\text{m}$ to $256\mu\text{m}$, $dz = 4\mu\text{m}$]; (middle) subset of 15 images using minimum z step size [$-28\mu\text{m}$ to $28\mu\text{m}$, $dz = 4\mu\text{m}$]; and (right) subset of 15 images using maximum z range [$-252\mu\text{m}$ to $252\mu\text{m}$, $dz = 36\mu\text{m}$]. With equal spacing, there is a forced trade-off between low-frequency noise and high-frequency blurring, such that many images are required for good quality phase results. (c) Phase recovered with exponentially spaced z steps: (left) subset of 15 images ($\beta = 2$), and (right) subset of 9 images ($\beta = 4$). The minimum and maximum defocus distances are fixed at $\pm 4\mu\text{m}$ and $\pm 256\mu\text{m}$, respectively.

With our exponentially spaced measurement scheme, GP TIE requires fewer images to be captured in order to obtain a high quality phase result. To demonstrate this, we compare the recovered phase by GP TIE from different subsets of Data Set 1 (Fig. 8) by sampling the image stack along z using various strategies. Figure 8(b) uses data subsets with equally spaced

z-planes. First, we show the best possible phase result, when all 129 images are used. With equally spaced planes, we have two possibilities for reducing the number of images: we can either reduce the defocus range (keeping the step size small) or increase the step size (reducing the defocus range). If we reduce the defocus range, the recovered phase becomes susceptible to low-frequency noise (see the middle of Fig. 8(b)). This is due to the fact that the low-frequency information of phase is not well captured at small defocus distances. If we instead increase the step size (keeping the defocus range large), high-frequency components are lost due to nonlinearity (see the right of Fig. 8(b)). In order to accurately capture both high and low-frequency information with the same reduced number of images, we need nonlinearly spaced measurements. Near the focus, the step size should be small, yet a large focus range can still be covered with only a few measurements in our exponentially spaced scheme. In Fig. 8(c), we extract images from Data Set 1 according to the exponential spacing scheme described in Section 2.3. As can be seen from Fig. 8(c), the phase results are free of low-frequency noise and also have high resolution. Even after further reducing the dataset to only 9 exponentially spaced images, we obtain a similar result to that with all 129 images. Thus, we have reduced the data capture requirement by more than an order of magnitude, without sacrificing quality.

Having shown that exponential spacing is advantageous, we now compare GP TIE with Higher order TIE, both of which are capable of using unequally spaced data. **Data Set 2** comprises 9 images of human cheek cells, exponentially spaced about focus at $\pm 5.7\mu\text{m}$, $\pm 11.4\mu\text{m}$, $\pm 22.8\mu\text{m}$, and $\pm 45.6\mu\text{m}$ (Fig. 9(a)). The imaging system has 10x magnification and NA of 0.5. Each image has 350×360 pixels of size $0.62\mu\text{m} \times 0.62\mu\text{m}$. Figure 9(b) shows the recovered phase of Data Set 2 obtained by Higher order TIE and GP TIE. We show the results for higher order TIE with the order of polynomial fitting m equal to 2, 3, and 4. The phase images of $m = 2$ and 3 have small low-frequency noise but appear blurred. The phase of $m = 4$ has strong contrast in some regions, but contains low-frequency noise. In contrast, GP TIE does not suffer from this tradeoff, so the phase recovered has less low-frequency noise and also exhibits high contrast. We can see details inside cells in the phase image recovered by GP TIE.

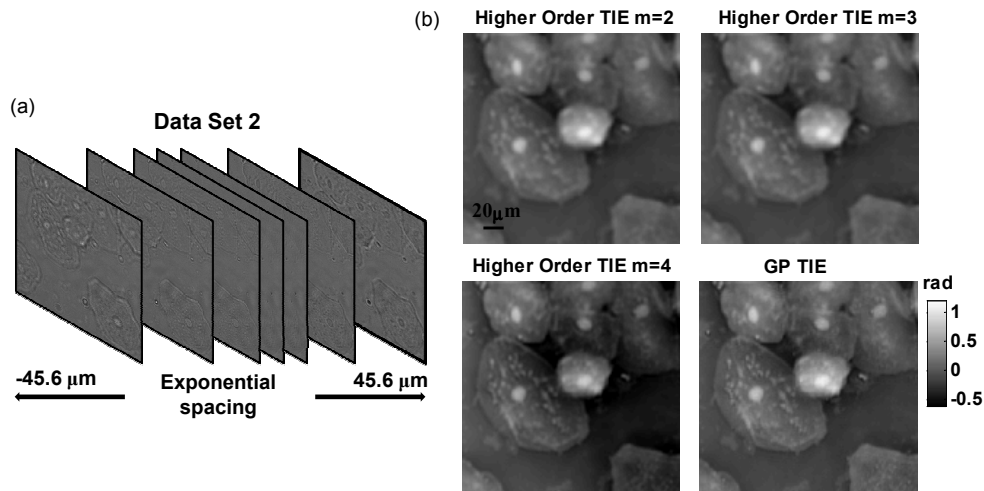


Fig. 9. (a) Data Set 2: exponentially spaced, with z of $\pm 5.7\mu\text{m}$, $\pm 11.4\mu\text{m}$, $\pm 22.8\mu\text{m}$, and $\pm 45.6\mu\text{m}$. Each image has 350×360 pixels of size $0.62\mu\text{m} \times 0.62\mu\text{m}$. (b) Phase images of Data Set 2 by Higher order TIE ($m = 2, 3$, and 4) and GP TIE.

5. Conclusions

In this paper, we proposed a TIE phase recovery method that exploits prior knowledge of spatial frequency variations of the intensity in order to yield more accurate phase images from exponentially spaced images through-focus. Our proposed method estimates the intensity derivative by means of GP regression. It is robust and stable with noise, while incorporating the a priori knowledge of the sinusoidal patterns predicted when fitting spectrum variations over z . We derived an exponentially spaced measurement scheme which guarantees a minimum phase measurement sensitivity across the full range of available spatial frequencies with the least possible number of images required. In future work, we will extend the current technique to add more a priori information, such as partial coherence of the illumination and large phase effects. With the freedom to choose nonlinearly spaced measurement positions, we can develop new strategies to optimize measurements if any prior knowledge of the phase spectrum is known.

It should be noted that improving the TIE result by incorporating prior knowledge has been previously considered in [30–32]. The difference is that the previous approach considers priors over the phase [30, 31] or refractive index distribution [32], whereas here we use a prior on intensity spectral evolution to improve the intensity derivative estimate. We expect the approach considered here should have wide application as it not only can be directly apply to the special cases in [30–32], but also to more general situations without constraints on the phase distribution. If any reader is interested in this algorithm, open source code can be obtained by emailing the authors or visiting the website either www.laurawaller.com or www.dauwels.com.

Appendix A: review of Gaussian process regression

We review the basics of Gaussian process regression [25]. Consider the problem of 2-D regression: given input/output pairs (z_n, f_n) , where $n = 1, \dots, N$, we would like to estimate $f(z)$ at arbitrary position z . Under the Gaussian process assumption, the outputs f_n are drawn from the zero-mean Gaussian distribution with the covariance as a function of z_n :

$$(f_1, f_2, \dots, f_N | z_1, z_2, \dots, z_N) \sim \mathcal{N}(\mathbf{0}, \mathbf{K}(\mathbf{Z}, \mathbf{Z}) + \sigma_n \mathbf{I}), \quad (6)$$

where $\mathbf{K}(\mathbf{Z}, \mathbf{Z})$ is the covariance matrix of the outputs given the input set \mathbf{Z} , and σ_n is the variance of additive Gaussian noise in outputs. Generally, the squared exponential covariance function is used to model the covariance matrix:

$$\mathbf{K}_{ij} = \sigma_f^2 \exp\left[-\frac{1}{2\ell^2}(z_i - z_j)^2\right], \quad (7)$$

where \mathbf{K}_{ij} is the element at coordinate (i, j) of the matrix $\mathbf{K}(\mathbf{Z}, \mathbf{Z})$. The parameters σ_f , ℓ , and σ_n in Eq. (6) are defined as the hyper-parameters of the GP model. We can write the joint distribution of the observed input/output pairs with the unknown value of $f(z)$ at z as:

$$\begin{bmatrix} \mathbf{f} \\ f(z) \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \mathbf{K}(\mathbf{Z}, \mathbf{Z}) + \sigma_n \mathbf{I} & \mathbf{K}(\mathbf{Z}, z) \\ \mathbf{K}(z, \mathbf{Z}) & \mathbf{K}(z, z) \end{bmatrix}\right), \quad (8)$$

where $\mathbf{f} = [f_1, f_2, \dots, f_N]^T$. The conditional distribution of the unknown output $f(z)$ at z is calculated as:

$$(f(z) | f_1, f_2, \dots, f_N, z_1, z_2, \dots, z_N, z) \sim \mathcal{N}(\bar{f}(z), \bar{\mathbf{K}}), \quad (9)$$

where

$$\bar{f}(z) = \mathbf{K}(z, \mathbf{Z})(\mathbf{K}(\mathbf{Z}, \mathbf{Z}) + \sigma_n \mathbf{I})^{-1} \mathbf{f}, \quad (10)$$

$$\bar{\mathbf{K}} = \mathbf{K}(z, z) - \mathbf{K}(z, \mathbf{Z})(\mathbf{K}(\mathbf{Z}, \mathbf{Z}) + \sigma_n \mathbf{I})^{-1} \mathbf{K}(\mathbf{Z}, z). \quad (11)$$

For convenience, we can define a vector of functions $\mathbf{h}(z)$:

$$\mathbf{h}(z)^T = \mathbf{K}(z, \mathbf{Z})(\mathbf{K}(\mathbf{Z}, \mathbf{Z}) + \sigma_n \mathbf{I})^{-1}. \quad (12)$$

Thus we have

$$\bar{f}(z) = \mathbf{h}(z)^T \mathbf{f}. \quad (13)$$

The predicted function $\bar{f}(z)$ can be understood as a weighted combination of the shifted equivalent kernel $h(z)$ [25, 33]. In [33], the Fourier transform of the equivalent kernel $h(z)$ for the squared exponential covariance function is given as:

$$\tilde{h}_{SE}(\mathbf{s}) = \frac{1}{1 + b \exp(2\pi^2 \ell^2 |\mathbf{s}|^2)}, \quad (14)$$

where $b = \sigma_n^2 / \rho (2\pi \ell^2)^{1/2}$ and ρ is the average number of observations per unit (for example length). When b is small, $\tilde{h}_{SE}(\mathbf{s})$ is approximated by a step function. The rapid change from 1 to 0 happens at the point when

$$s_c^2 = \log(1/b) / (2\pi^2 \ell^2). \quad (15)$$

Therefore, the frequency components above the threshold s_c in the fitted function $\bar{f}(z)$ are suppressed. From Eq. (15), we can set the desired threshold s_c by changing the hyper-parameters. The property of suppressing the unwanted high frequency components in the fitted function is applied in the regression of the intensity images in TIE.

Appendix B: derivation of GP TIE

We have a 3D stack $\mathcal{I}(u, v, z_1), \dots, \mathcal{I}(u, v, z_N)$, which are the 2D Fourier transforms of the measured intensity images. Our goal is to use regression to estimate the first derivative of the intensity spectrum at $z = 0$, $\left. \frac{\partial \mathcal{I}(u, v, z)}{\partial z} \right|_{z=0}$. Instead of doing 3D regression, we perform GP regression for each lateral spatial frequency (u_m, v_n) on N data points $\mathcal{I}(u_m, v_n, z_1), \mathcal{I}(u_m, v_n, z_2), \dots, \mathcal{I}(u_m, v_n, z_N)$. It is easy to observe that $\mathcal{I}(u_m, v_n, z)$ is taken from $\mathcal{I}(u, v, z)$ at the same frequency coordinates (u_m, v_n) . From Eq. (3), $\mathcal{I}(u_m, v_n, z)$ follows a sinusoidal pattern with frequency $\pi \lambda (u_m^2 + v_n^2)$. The sinusoids prior is incorporated into the GP regression by setting the frequency threshold s_c as $\pi \lambda (u_m^2 + v_n^2)$. This is realized by setting appropriate hyper-parameters σ_f, σ_n , and ℓ in the regression. The hyper-parameters σ_f and σ_n are initialized to keep b in Eq. (14) small. Next, the parameter ℓ is solved from Eq. (15) with s_c and σ_f, σ_n already known. The frequency threshold s_c can be larger than $\pi \lambda (u_m^2 + v_n^2)$ to allow trade-offs between accuracy and noise filtering.

Define $\left. \frac{\partial \mathcal{I}(u_m, v_n, z)}{\partial z} \right|_{z=0}$ as the element of $\left. \frac{\partial \mathcal{I}(u, v, z)}{\partial z} \right|_{z=0}$ at the coordinate (u_m, v_n) . From Eqs. (10)(12), the function $\bar{\mathcal{I}}(u_m, v_n, z)$ fitted by GP regression is expressed as:

$$\bar{\mathcal{I}}(u_m, v_n, z) = \mathbf{h}(u_m, v_n, z)^T \mathbf{I}_{mn}, \quad (16)$$

where $\mathbf{I}_{mn} = [\mathcal{I}(u_m, v_n, z_1), \mathcal{I}(u_m, v_n, z_2), \dots, \mathcal{I}(u_m, v_n, z_N)]^T$ and $\mathbf{h}(u_m, v_n, z)$ can be obtained from Eq. (12). The fitted function $\bar{\mathcal{I}}(u_m, v_n, z)$ is a function of the single variable z . Therefore, its first derivative at $z = 0$, $\left. \frac{\partial \bar{\mathcal{I}}(u_m, v_n, z)}{\partial z} \right|_{z=0}$, is approximated by:

$$\left. \frac{\partial \bar{\mathcal{I}}(u_m, v_n, z)}{\partial z} \right|_{z=0} = \left. \frac{\partial \mathbf{h}(u_m, v_n, z)^T}{\partial z} \right|_{z=0} \mathbf{I}_{mn}, \quad (17)$$

where

$$\frac{\partial \mathbf{h}(u_m, v_n, z)^T}{\partial z} = \left[-\frac{\sigma_f^2}{\ell^2} (z - z_1) \exp\left[-\frac{1}{2\ell^2} (z - z_1)^2\right], \dots, -\frac{\sigma_f^2}{\ell^2} (z - z_N) \exp\left[-\frac{1}{2\ell^2} (z - z_N)^2\right] \right] \cdot (\mathbf{K}(\mathbf{Z}, \mathbf{Z}) + \sigma_n \mathbf{I})^{-1}. \quad (18)$$

In Eq. (17), the vector \mathbf{I}_{mn} is known from the measurements, and the vector $\frac{\partial \mathbf{h}(u_m, v_n, z)^T}{\partial z} \Big|_{z=0}$ can be derived from the GP regression with hyper-parameters already known. Therefore, $\frac{\partial \mathcal{I}(u_m, v_n, z)}{\partial z} \Big|_{z=0}$ is estimated by performing GP regression over data points $\mathcal{I}(u_m, v_n, z_1), \mathcal{I}(u_m, v_n, z_2), \dots, \mathcal{I}(u_m, v_n, z_N)$.

By repeating the same process, we can obtain all of the frequency components for the first derivative of intensity spectrum at focus $\frac{\partial \mathcal{I}(u, v, z)}{\partial z} \Big|_{z=0}$, then the phase is recovered with the Laplacian inversion in Eq. (1)(2). Because the computational complexity of GP regression is proportional to the cubic of the number of measurements [25], the algorithm has a computational complexity of $\mathcal{O}(N(N_z)^3)$, where N is the number of pixels in one image, and N_z is the number of measurements. In order to save computational time, the frequency components which have similar $u_m^2 + v_n^2$ values can share the same hyper-parameters and hence the same $\frac{\partial \mathbf{h}(u_m, v_n, z)^T}{\partial z} \Big|_{z=0}$.

Acknowledgments

The authors would like to thank Sijia Liu and Jingyan Wang for helping with experiments, and Aamod Shanker, Daniel Shuldman, and Andrew R. Neureuther for helpful discussions.