Optical wavefront measurement using phase retrieval with transverse translation diversity

Gregory R. Brady[*], Manuel Guizar-Sicairos, and James R. Fienup

The Institute of Optics, University of Rochester, Rochester, New York, USA, 14627-0186
[*]Corresponding author: grbrady@gmail.com
http://www.optics.rochester.edu/workgroups/fienup

Abstract: We demonstrate the use of transverse translation-diverse phase retrieval as a method for the measurement of wavefronts in situations where the detected intensity patterns would be otherwise undersampled. This technique involves using a smaller moving subaperture to produce a number of adequately sampled intensity patterns. The wavefront is then retrieved using an optimization jointly constrained by them. Expressions for the gradient of an error metric with respect to the optimization parameters are given. An experimental arrangement used to measure the transmitted wavefront of a plano-convex singlet using this technique is described. The results of these measurements were repeatable to within approximately $\lambda/100$ RMS.

References and Links

1. G.R. Brady and J.R. Fienup, "Range of Phase Retrieval in Optical Metrology," in Frontiers in Optics 2005 / Laser Science XXI (Optical Society of America, Washington DC, 2005), paper FTuS3.
2. J.R. Fienup, "Phase Retrieval for Undersampled Broadband Images," J. Opt. Soc. Am. A, **16**, 1831-1839 (1999).
3. P. Dumas, J. Fleig, G. Forbes, and P.E. Murphy, "Extending the range of interferometry through subaperture stitching," Proc. SPIE **TD02**, 134-7 (2003).
4. M. Bray, "Stitching interferometer for large optics using a standard interferometer: description of an automated system [for ICF optics]," Proc. SPIE **3047**, 911-18 (1997).
5. M. Guizar-Sicairos and J.R. Fienup, "Phase retrieval with transverse translation diversity: a nonlinear optimization approach," Opt. Express **16**, 7264-78 (2008).
6. H. M. L. Faulkner and J. M. Rodenburg, "Movable aperture lensless transmission microscopy: a novel phase retrieval algorithm," Phys. Rev. Lett. **93**, 023903 (2004).
7. J. M. Rodenburg and H. M. L. Faulkner, "A phase retrieval algorithm for shifting illumination," Appl. Phys. Lett. **85**, 4795–4797 (2004).
8. J. M. Rodenburg, A. C. Hurst and A. G. Cullis, "Transmission microscopy without lenses for objects of unlimited size," Ultramicroscopy **107**, 227–231 (2007).
9. J. M. Rodenburg, A. C. Hurst, A. G. Cullis, B. R. Dobson, F. Pfeiffer, O. Bunk, C. David, K. Jefimovs and I. Johnson, "Hard-x-ray lensless imaging of extended objects," Phys. Rev. Lett. **98**, 034801 (2007).
10. H. M. L. Faulkner and J. M. Rodenburg, "Error tolerance of an iterative phase retrieval algorithm for moveable illumination microscopy," Ultramicroscopy **103**, 153–164 (2005).
11. M. Guizar-Sicairos and J. R. Fienup, "Measurement of coherent x-ray focused beams by phase retrieval with transverse translation diversity," manuscript submitted to Opt. Express.
12. J.R. Fienup, "Phase Retrieval Algorithms: A Comparison," Appl. Opt. **21**, 2758-2769 (1982).
13. J.R. Fienup, "Phase-Retrieval Algorithms for a Complicated Optical System," Appl. Opt. **32**, 1737-1746 (1993).
14. W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes: The Art of Scientific Computing* (Cambridge University Press, 1986), Chap. 10.
15. M. Guizar-Sicairos, S.T. Thurman and J.R. Fienup, "Efficient Subpixel Image Registration Algorithms," Opt. Lett. **33**, 156-158 (2008).

1. Introduction

Phase retrieval is a useful technique for wavefront determination that requires only simple measurements of the intensity distributions produced by the field of interest. Phase retrieval distinguishes itself from other wavefront sensing methods by often not requiring additional optics to perform the measurement. The intensity measurements are typically made in a plane removed from the plane where the wavefront determination is desired. Often the plane where the wavefront is desired is the exit pupil of a focusing optical system. In this case the intensity measurements for phase retrieval are typically made at or near the image plane where the pattern of light can be collected onto a detector array such as a CCD.

If this intensity pattern is to be adequately sampled by the detector, the f-number of the beam is limited by

$$f/\# \geq \frac{Qd_u}{\lambda} \tag{1}$$

or in terms of the numerical aperture (NA)

$$NA \leq \frac{\lambda}{2Qd_u}, \tag{2}$$

where d_u is the detector sample spacing (pixel pitch), λ is the wavelength, and $Q = (\lambda f/\#)/d_u$ is a parameter describing the sampling rate [1]. If $Q \geq 2$ the intensity pattern is adequately (Nyquist) sampled and if $Q \geq 1$ the field is adequately sampled. If a detector with a 5 μm sample spacing is used with 632.8 nm illumination and the field is adequately sampled, the f-number must be greater than 7.9 (or the NA less than 0.063). This limits the use of most phase retrieval algorithms to high $f/\#$ wavefronts [2].

Here we describe an approach that relieves this limitation. As illustrated in Fig. 1, we place a mask (termed a subaperture) in a plane at or near the plane of the field or aperture of interest, thereby reducing its NA and forming an adequately sampled intensity pattern, which we measure. We can estimate the phase across the entire aperture by using multiple intensity measurements, transversely translating the subaperture between each of them. Note that there are no additional optical components other than the lens being measured and the movable subaperture. Using this approach the wavefront over an arbitrarily large aperture can be measured, limited only by data collection time, computation time and data storage. In certain situations it may also be necessary to transversely translate the detector array to capture the light passing through the subaperture, depending on the plane in which the detector is placed and the nominal shape of the wavefront to be measured.

Subaperture "stitching" approaches for determining wavefronts across a larger area have been demonstrated using interferometry [3, 4]. Combining the subaperture measurements involves registering overlapping regions of the interferometry measurements, and is frequently complicated by imaging distortions in the interferometer optics. Similar methods could be employed to combine multiple subaperture phase retrieval wavefront measurements. Intensity data, gathered as shown in Fig. 1(a) and (c), could be processed individually to arrive at a wavefront estimate of that particular subaperture. The individual subaperture phases would then be stitched together to form a composite wavefront measurement. This task may be simpler for phase retrieval data than for interferometry because interferometer imaging errors would not have to be corrected. When independently retrieving the phase in each subaperture in this way, often multiple axially separate (i.e. in different focus planes, for phase-diverse phase retrieval) intensity patterns are collected to increase the robustness of the algorithm and resolve certain phase ambiguities.
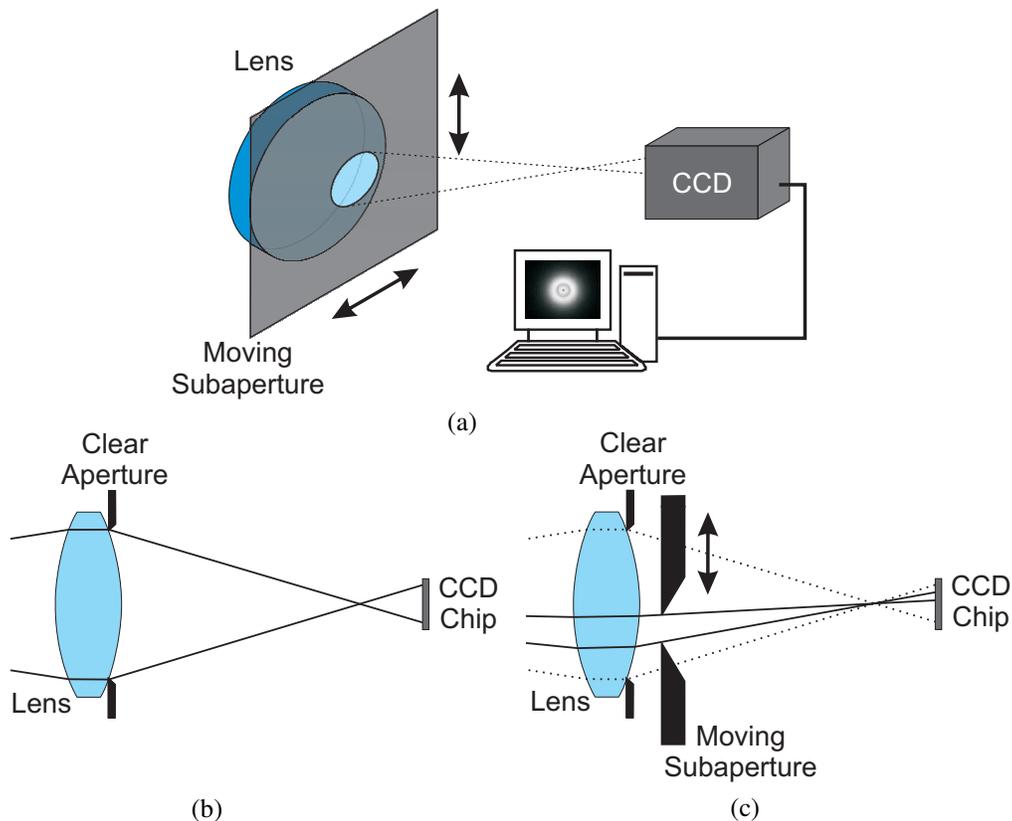
Fig. 1. Experimental arrangement for phase retrieval with subaberture tranverse translation diversity. (a) Perspective view of the arrangement. (b) Without tranverse translation diversity, if the numerical aperture of the lens is too large the intensity pattern at the CCD will be undersampled. (c) With translation diversity, the numerical aperture of the beam is reduced by truncating it with a smaller subaperture. A larger area is mapped out by making multiple subaperture measurements, translating the subaperture between each.

In contrast to conventional stitching techniques, our method estimates a single underlying wavefront that is consistent with all of the intensity measurements produced from multiple, overlapping positions of the subaperture, which is more robust than a conventional stitching approach. Overlapping of the subaperture positions is necessary to resolve inherent ambiguities in the phase estimate and improves convergence. This also generally eliminates the necessity of collected data in multiple focus planes.

Related work has been done in the fields of x-ray imaging and x-ray beam characterization. In these situations intensity measurements are typically taken in the far-field regime, where a longitudinal displacement of the detector does not provide suitable diversity for robust phase reconstructions. In this case, transverse translation diversity has been used primarily to improve the robustness of image reconstruction by phase retrieval [5-10], while in our work we use it primarily to increase the measurement range of the technique for optical metrology. A technique for measuring 1D and 2D focused x-ray beams with an analogous method is described in Ref. [11], assuming a simple FT relation between the plane where the moveable structure is placed and the measurement plane. Although for phase retrieval at x-ray wavelengths a Fourier transform relationship is typically appropriate, a more general propagation is usually required to describe propagation of optical wavefronts. Here we provide analytical expressions for the gradient of an error metric with respect to the parameters that describe the wavefront and movable subaperture for a generalized

propagation. The method described here does not require exact knowledge of the translations of the subaperture for the multiple measurements. This distinguishes it from earlier methods based on the iterative transform algorithm which are limited by the positioning accuracy of the motion control equipment [6-10].

In Section 2 and the Appendix of this paper we describe the phase retrieval algorithm in detail. In Section 3 we show phase reconstructions from laboratory data over a composite aperture that could not be performed using conventional phase retrieval because of the pixel pitch limitation. We conclude in Section 4.

2. Nonlinear optimization over multiple transverse translations of a subaperture

Our algorithms for phase retrieval with transverse translation diversity are based on an explicit nonlinear optimization of parameters describing the field in the pupil plane so that the intensity distributions, computed when the field is numerically propagated, agree with the measured intensity distributions [5, 11-13].

The first necessary component for this kind of algorithm is a model of the field in the plane of interest. As in Ref. [5], we write the field transmitted by the subaperture in the nth position as

$$\tilde{g}_n(x, y) = h(x, y) a(x - x_n, y - y_n),$$

(3)

where $h(x, y)$ is the field impinging on the subaperture, $a(x, y)$ is the complex amplitude transmittance of the subaperture, and the subaperture is translated by (x_n, y_n) for the nth position. Depending on the propagation method used, the phase of $h(x, y)$ may be the deviation from a plane or spherical wavefront. This can equivalently be written as

$$g_n(x, y) = \tilde{g}_n(x + x_n, y + y_n) = h(x + x_n, y + y_n) a(x, y),$$

(4)

where we have now taken the subaperture to be stationary and the incident field moving, which was numerically more convenient in our software. The sizes of the arrays representing $h(x, y)$ and $a(x, y)$ do not need to be the same, with the larger being truncated to the size of the smaller during the calculation of $g_n(x, y)$. Usually the unknown field, $h(x, y)$, will be the larger array. Subpixel shifted versions of $h(x, y)$ or $a(x, y)$ are calculated making use of the discrete Fourier transform shift theorem. Since a field $h(x, y)$ is usually band-limited, this sinc-like interpolation can be more accurate for it than for an aperture function $a(x, y)$, thereby making it advantageous to model the field as translating.

The fields in the measurements plane are computed from $g_n(x, y)$ using a generalized Fourier optics propagation,

$$G_n(u, v) = \mathbf{P}\big[g_n(x, y) \big],$$

(5)

where $\mathbf{P}[\ ]$ is a linear propagation operator, which may be as simple as a Fourier transform, used in Ref. [5], or may be a propagation through a more general optical system [13]. For converging wavefronts we typically use a two-step propagation, where Fresnel diffraction is used to propagate to the paraxial image plane of the system under test, and angular spectrum is used to propagate that resulting field to a defocused plane a small distance Δz away. This propagator can be written as

$$G_n^f(u',v') = \exp\left\{\frac{i\pi}{\lambda z}\left[(u'd_u)^2 + (v'd_v)^2\right]\right\} DFT\left[g_n(x,y)\right],$$

$$G_n(u,v) = IDFT\left[DFT\left[G_n^f(u',v')\right]\exp\left\{i2\pi\Delta z\left[\frac{1}{\lambda^2} - \left(\frac{r}{Nd_u}\right)^2 - \left(\frac{s}{Nd_v}\right)^2\right]^{\frac{1}{2}}\right\}\right], \tag{6}$$

where $DFT[\ ]$ and $IDFT[\ ]$ represent the forward and inverse discrete Fourier transform, z is the distance to the paraxial image plane, Δz is the small defocus distance, r and s are spatial frequency indices, u' and v' are coordinates indices in the paraxial image plane, and constant factors are ignored. The relation between the sample spacings (d_x, d_y) in the input plane and (d_u, d_v) in the focal plane and in defocused measurement planes is

$$d_{u,v} = \frac{\lambda z}{Nd_{x,y}}. \tag{7}$$

We represent the terms $a(x,y)$, $g_n(x,y)$, and $G_n(u,v)$ by arrays of size $N \times N$, and $h(x,y)$ by an array of size $M \times M$. Usually M is larger than N.

We compared the amplitudes of the propagated field $\left|G_n(u,v)\right|$ to the field magnitudes $\left|F_n(u,v)\right| = \sqrt{I_n(u,v)}$, where $I_n(u,v)$ is the measured intensity pattern, using a squared-difference error metric

$$E = \sum_{n=1}^{q}\sum_{u,v} W_n(u,v)\left[\left|F_n(u,v)\right| - \left|G_n(u,v)\right|\right]^2, \tag{8}$$

where q is the number of subaperture positions and $W_n(u,v)$ are weighting terms that allow us to ignore regions of poor signal to noise ratio or known bad detector pixels.

The error metric is minimized by varying the parameters describing $h(x,y)$, $a(x,y)$ and the subaperture translations (x_n, y_n). For this purpose we typically use gradient search algorithms, e.g., the conjugate gradient search [14]. These algorithms require the gradient of the error metric with respect to the optimization parameters. The large number of free parameters in this problem makes calculation of the gradient using finite difference methods prohibitively expensive, so we use computationally efficient analytic expressions for the gradients. Depending on the particular application and *a priori* knowledge of the system, different parameterizations of $h(x,y)$ and $a(x,y)$ may be advantageous. The gradients for a number of parameterizations are given in the Appendix.

3. Experiment

3.1. Experimental Arrangement

The experimental arrangement shown in Figs. 2 and 3 was constructed to collect data sets taken at multiple subaperture transversely translated positions. A 10 mW helium-neon (HeNe) laser beam passed through a shutter, which was used to block the beam under computer control so that dark frames can be taken automatically. A range of neutral density (ND) filters were mounted in a computer-controlled filter wheel so that the light level could be controlled automatically. Typically optical densities in the range of 3 to 4 were required to avoid saturating the CCD camera. The attenuated beam was focused by a 40× microscope

objective onto a 5 μm diameter pinhole, which was a sub-resolution point source in this arrangement. The point source was placed 1453 mm in front of a 2" diameter plano-convex lens with a focal length of 250 mm forming an image of the point source 302 mm behind the lens. The orientation of the lens (planar surface toward longer conjugate) was chosen so that significant spherical aberration would be produced, giving an interesting, yet well-behaved wavefront to measure. The moving aperture was placed as close to the lens under test as was practical given the constraints of the mounts used. The moving subaperture was mounted on a two-axis computer-controlled stage which produced the required transverse translations. The CCD camera, a Q-Imaging Retiga 2000R, was mounted on a third axis of motion control along the direction of propagation to allow for the focus to be adjusted so that data could be taken at a selected plane. The CCD is fixed axially during collection of the data. This is a 12 bit camera, with reasonably good noise characteristics (40,000 photoelectron full well capacity, 16 photoelectrons of read noise), with a 1600×1200 array of 7.4 μm pixels, but is by no means a state-of-the-art scientific camera.
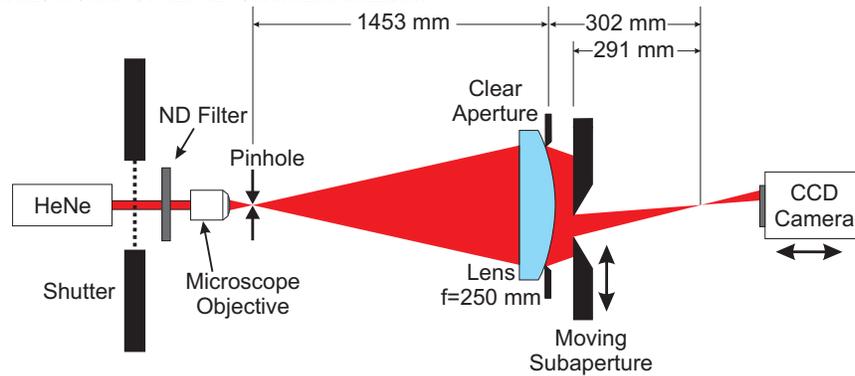

Fig. 2. Diagram of the experimental arrangement used to collect subaperture intensity pattern data (not to scale). The CCD camera is placed near the focus. The subaperture and CCD are mounted on computer-controlled translation stages.
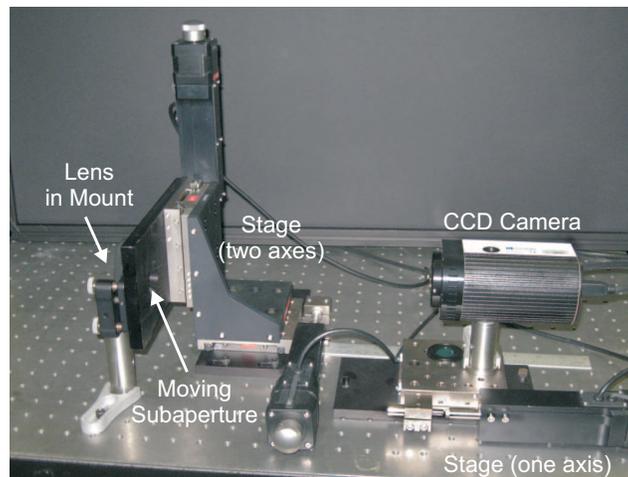

Fig. 3. A portion of the experimental setup including the lens that forms the wavefront of interest, the moving subaperture, the CCD camera, and the associated motion control equipment. The illumination point source is out of the frame to the left.
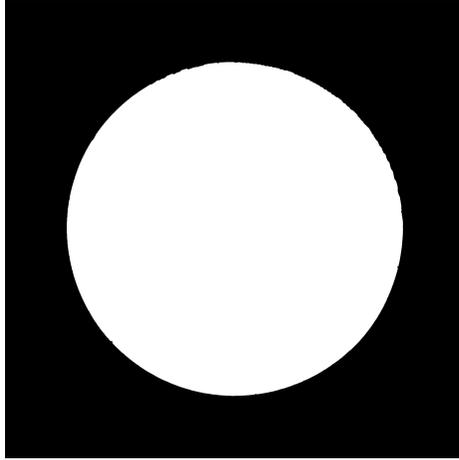
Fig. 4. An image of the subaperture from a flatbed scanner. The diameter is 9.85 mm.

The subaperture was machined in a 0.5" thick aluminum plate which was anodized black to reduce reflections. The final machined aperture was found to be 9.85 mm in diameter. The shape of the subaperture, shown in Fig. 4, was measured by scanning the plate using a 6400 dpi flatbed scanner with a backlighting attachment typically used for scanning photographic negatives. The machined subaperture proved to be close to a perfect circle, although small imperfections are visible.

3.2. Measurement of Subaperture to Camera Distance

The distance between the subaperture and the camera was determined using a double pinhole placed in the subaperture and performing Young's experiment, giving a pattern as shown in Fig. 5. The distance between the pinhole plane and the camera plane is given by

$$L = \frac{s}{2\left\{\tan\left[\arcsin\left(\frac{\lambda v}{2}\right)\right]\right\}} \tag{9}$$

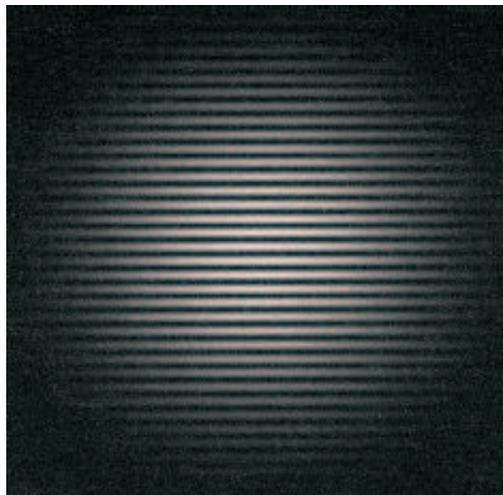where s is the pinhole separation, and v is the fringe frequency.

Fig. 5. Double pinhole interference pattern that was analyzed to determine the distance between the pinhole (subaperture) plane and the CCD camera plane.

The fringe pattern was analyzed by Fourier transforming it and finding, with sub-pixel accuracy [15], the location of the peak corresponding to the fringe frequency. This allowed us to measure the distance, L, to a resolution of about 5 μm.

The distance from the subaperture (double pinhole) plane to the nominal focal plane was measured to be 291 mm. When this measurement was complete the double pinhole was removed from the arrangement. As this distance measurement was no longer possible, we relied on the accuracy of the motion-control equipment to track the offset from the nominal focal plane when the camera was moved axially. This may cause small, relative positioning errors resulting in small focus terms that are easily removed.

3.3. Results

In order to verify the usefulness of this technique for measuring faster beams than allowed by Eqs. (1) or (2), we measured the wavefront over a composite aperture with an effective diameter of 33.32 mm. A full aperture measurement over this area would be undersampled, having a sampling parameter $Q = 0.748$. The 9.85 mm diameter subaperture results in a $Q = 2.53$, which is oversampled for both intensity and field. To cover the area of the composite aperture with significant subaperture overlap, we measured intensity patterns at 43 different subaperture positions illustrated in Fig. 6. A point in the composite aperture could be sampled by as few as one (near the edge) or as many as eight subaperture measurements. Example intensity patterns produced by various subaperture positions are shown in the middle column of Fig. 7, and the corresponding subaperture position is indicated in the left column. For processing in our algorithm, each intensity pattern was centered in a 256×256 array of data. This resulted in a subaperture plane sample spacing of 97.4 μm.

The raw intensity data were processed by first subtracting an average dark frame from each frame. Any remaining bias was subtracted using the average value of a small patch of pixels well away from the intensity pattern of interest. The intensity patterns are normalized so that each has unity power integrated over the array, as are the numerically computed intensity patterns used in the optimization algorithm.

The dark-subtracted intensity data was used as the input to our phase retrieval algorithm using the following procedure. First, since the phase was expected to be slowly varying, we optimized the phase of $h(x, y)$ using Zernike polynomial coefficients up to 8th order, assuming a uniform amplitude distribution and using the gradient expression in Eq. (22). At the same time we optimized tip, tilt, and focus for each subaperture, $a_n(x, y)$, assuming the amplitude distribution shown in Fig. 4, to correct for focus errors and small shifts of the detector, using the gradient expression in Eq. (36). We also optimized for the subaperture positions, using the expressions in Eqs. (47) and (48). This optimization ran for 50 iterations.

Second, we used the result from the first step as the initial guess for an optimization of 100 iterations of a pixel-by-pixel representation of $h(x, y)$, making use of the gradient in Eq. (19). This captures high frequency variations of the phase not well expressed by Zernike polynomials.

Third, we used the result from the second step as the initial guess for an optimization of 1000 iterations over the complex values of $h(x, y)$ at each point in the composite aperture using Eq. (29), at which point we do not rely on the assumption that the amplitude of $h(x, y)$ is uniform.

Finally, we used this result as the initial guess for an optimization of 100 iterations where we allowed the complex numbers at each point in the moving subaperture to vary using Eq. (43). These complex values were kept the same for each subaperture position. This allows for correction of errors in our assumption of the aperture shape. In addition, we optimized for different tip, tilt and focus values for each subaperture and the complex values of $h(x, y)$ at each pixel in the same way that we did in the previous optimization. The resulting estimate of

the field in the composite pupil is shown in Fig. 8. The phase is quite smooth, exhibiting mostly spherical aberration. The amplitude is quite uniform, but does contain artifacts at the edges of the subapertures. It is interesting that amplitude variations due to digs on the lens surface were well reconstructed on the left side of the subaperture. These appear as amplitude variations because the large slope of the dig scatters light away from the detector array. The digs were visible to the naked eye. The reconstructed intensity patterns in the measurement plane are shown in the right column of Fig. 7. The agreement between the measured and retrieved intensity patterns is excellent.

As a check on our results we additionally performed a reconstruction using the same algorithm with a smaller, independent data set of 17 subaperture measurements which cover the central region of our previous result, as shown in Fig. 9. Notice that different subaperture positions are used than in the earlier result. The results of this reconstruction are shown in Fig. 10(b). The reconstruction compares well to the data from the same region of the larger aperture data shown in Fig. 8. A cropped version of this data is shown in Fig. 10(a). The difference between the two phase distributions is shown in Fig. 10(c). The RMS difference between the two patterns is 0.0107 waves. Since the largest differences appear to be due to high frequency variations, we fit the two data sets to 36 Zernike polynomials and compared those, as shown in Fig. 11. The RMS difference between the two polynomial fits is 0.0079 waves.
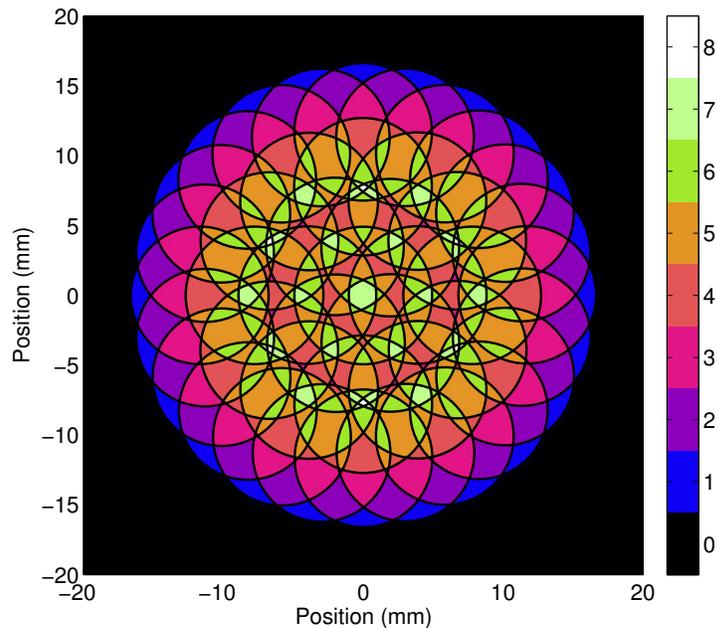

Fig. 6. A 9.85 mm subaperture is moved to 43 positions to measure a composite aperture with a diameter of 33.32 mm. The aperture postions are shown above, with the grey level indicating how many times a given point in the composite aperture was sampled by a subaperture.
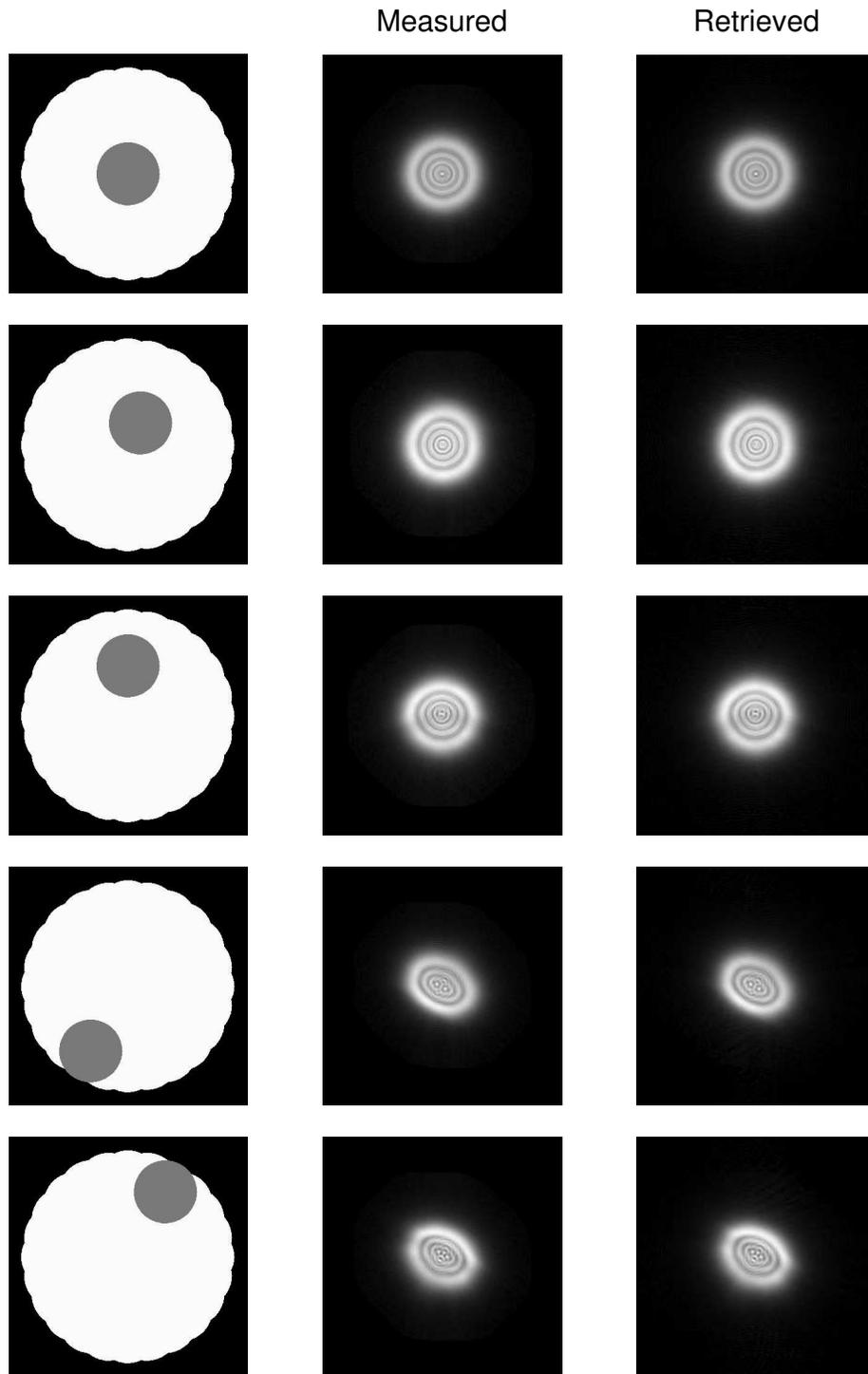
Measured   Retrieved

Fig. 7. The left column indicates the subaperture position of the measured intensity patterns (center column) and retrieved intensity patterns (right column). Intensities are shown raised to the 0.4 power.
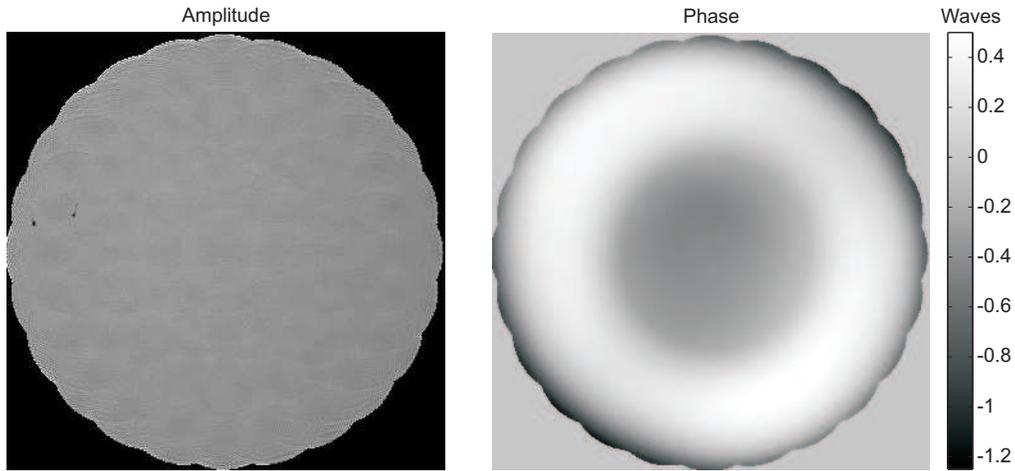
Fig. 8. Amplitude (left) and phase (right) retrieved over the composite aperture using the intensity patterns measured at the 43 subaperture positions shown in Fig. 6. The phase shown has tip, tilt and focus removed.
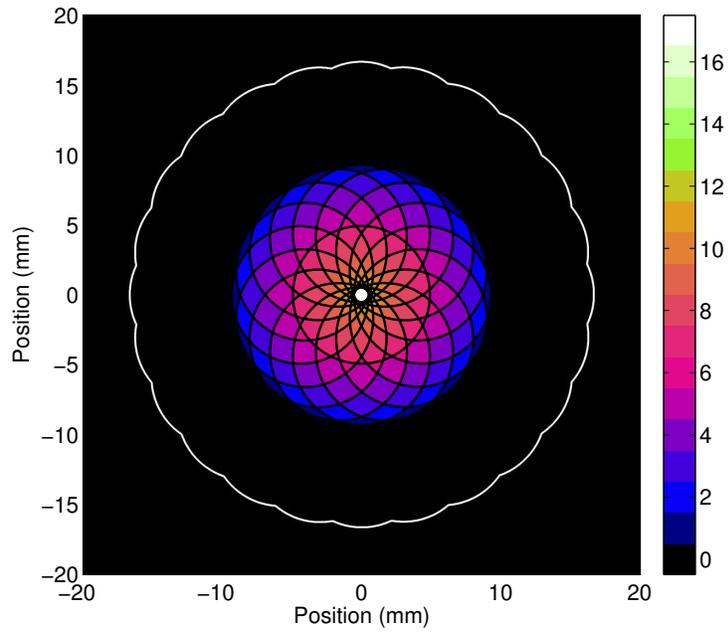

Fig. 9. Pattern of subaperture positions for a smaller data set taken to check the phase retrieval results. The color scale shows the number of times a particular point is sampled by a subaperture. The outer white line indicates the edge of the larger data set.
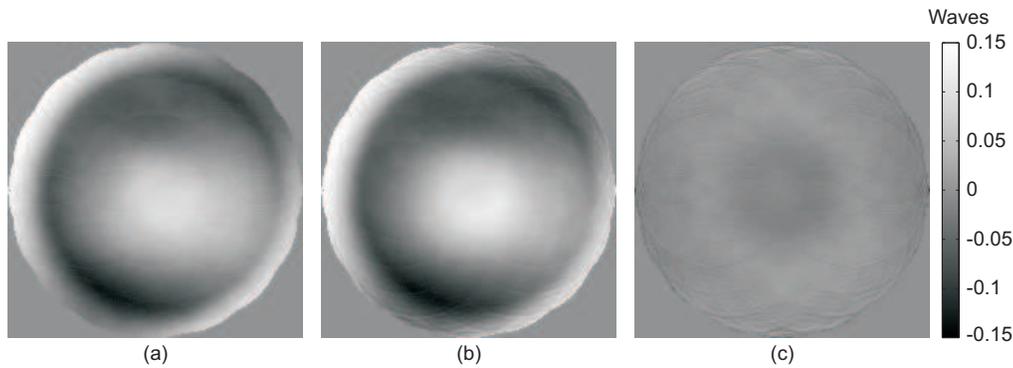
Fig. 10. (a) Wavefront of Fig. 8 cropped to the size of the small composite aperture pattern. Note that some artifacts are visible at the edge of the subapertures on the scale of this figure as compared to that of Fig. 8. (b) Wavefont retrieved using independent data collected over the small composite aperture shown in Fig. 9. (c) Difference between (a) and (b). Tip, tilt and piston have been removed from each, and all are on the same color scale.
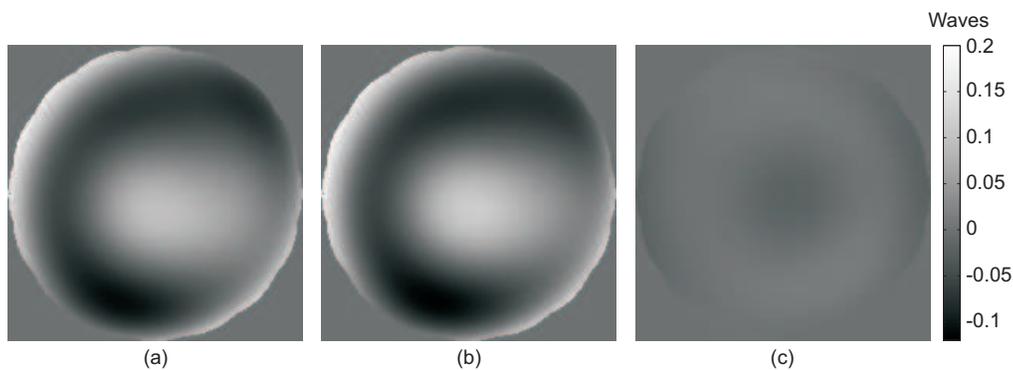

Fig. 11. Data of Fig. 10 fit to 36 Zernike polynomials. (a) Cropped and Zernike fit version of the wavefront in Fig. 8. (b) Zernike fit version of the phase shown in Fig. 10 (b). (c) Difference between (a) and (b). Tip, tilt and piston have been removed from each, and all are on the same color scale.

4. Conclusion

We have described and demonstrated a new method of phase retrieval that allows for measurement of faster converging beams than was possible with the conventional method. This method involves making intensity measurements for multiple overlapping transversely translated subaperture positions and jointly optimizing parameters of the entire field of interest so that modeled intensity patterns are consistent with the measurements. The use of overlapping subaperture data increases the robustness of the method in a similar way as taking data in multiple axially separated planes.

We have derived expressions for the gradient of a squared-difference of magnitudes error metric with respect to a number of different parameterizations of field or phase. The expressions include derivatives with respect to the positions of the subapertures used to form the intensity measurement, making the method very robust to uncertainties in these positions. The gradient expressions allow for the use of a generalized propagation calculation.

We described an experimental arrangement to make measurements of a wavefront transmitted through a plano-convex singlet, which was used at conjugates to produce significant spherical aberration. As many as 43 intensity measurements were jointly used to constrain a multiple-stage optimization over Zernike coefficients, point-by-point phase values and finally the complex field values of the field of interest. To verify the method, we measured different intensity patterns over a smaller region of the same field and performed

another optimization. The result of this second optimization agreed with the earlier result to within 0.0107 waves RMS. The good repeatability shown here is an indication that the algorithm is recovering an accurate wavefront, although a comparison with an independent measurement method is needed to definitely establish the accuracy.

This method promises to make phase retrieval a more practical measurement method for many problems, especially with faster beams, including those used in the metrology of optical surfaces and wavefronts.

Appendix: Analytic gradient calculations for different parameterizations

Following the derivation in Ref. [13], the gradient of the error metric in Eq. (8) with respect to a real-valued parameter α can be written as

$$\frac{\partial E}{\partial \alpha} = -\sum_{n=1}^{q} \sum_{u,v} G_n^{w*}(u,v) \frac{\partial G_n(u,v)}{\partial \alpha} + \text{c.c.} \qquad (10)$$

where

$$G_n^w(u,v) = W_n(u,v) \left[|F_n(u,v)| \frac{G_n(u,v)}{|G_n(u,v)|} - G_n(u,v) \right], \qquad (11)$$

and c.c. stands for complex conjugate. Inserting Eqs. (5) and (4) into Eq. (10) gives

$$\frac{\partial E}{\partial \alpha} = -\sum_{n=1}^{q} \sum_{u,v} G_n^{w*}(u,v) \mathbf{P} \left[\frac{\partial}{\partial \alpha} \{h(x+x_n, y+y_n)a(x,y)\} \right] + \text{c.c.} \qquad (12)$$

As in Ref. [13], the order of summations can be manipulated so that

$$\frac{\partial E}{\partial \alpha} = -\sum_{n=1}^{q} \sum_{x,y} \left\{ \mathbf{P}^\dagger \left[G_n^w(u,v) \right] \right\}^* \frac{\partial}{\partial \alpha} \left[h(x+x_n, y+y_n)a(x,y) \right] + \text{c.c.}, \qquad (13)$$

where $\mathbf{P}^\dagger [\ ]$ is the inverse operation to $\mathbf{P}[\ ]$. Defining

$$g_n^w(x,y) = \mathbf{P}^\dagger \left[G_n^w(u,v) \right] \qquad (14)$$

allows us to write Eq. (13) as

$$\frac{\partial E}{\partial \alpha} = -\sum_{n=1}^{q} \sum_{x,y} g_n^{w*}(x,y) \frac{\partial}{\partial \alpha} \left[h(x+x_n, y+y_n)a(x,y) \right] + \text{c.c.} \qquad (15)$$

For derivatives with respect to parameters describing $h(x,y)$, we employ a change of coordinates for (x,y) and write

$$\frac{\partial E}{\partial \alpha} = -\sum_{n=1}^{q} \sum_{x,y} g_n^{w*}(x-x_n, y-y_n) \frac{\partial}{\partial \alpha} \left[h(x,y)a(x-x_n, y-y_n) \right] + \text{c.c.} \qquad (16)$$

Expressing $h(x,y)$ in terms of its amplitude and phase, we have

$$h(x,y) = |h(x,y)| \exp\left[i\theta_h(x,y) \right]. \qquad (17)$$

The derivative with respect to the value of the phase at a particular pixel (x', y') is given by

$$\frac{\partial}{\partial \theta_h(x',y')} \left[h(x,y)a(x-x_n, y-y_n) \right] = ia(x-x_n, y-y_n)h(x,y)\delta(x,x';y,y'). \qquad (18)$$

We use this with Eq. (16) to write

$$\frac{\partial E}{\partial \theta_h(x',y')} = -i \sum_{n=1}^{q} g_n^{w*}(x'-x_n,y'-y_n)h(x',y')a(x'-x_n,y'-y_n) + \text{c.c.}$$

$$= 2\,\text{Im}\left[\sum_{n=1}^{q} g_n^{w*}(x'-x_n,y'-y_n)h(x',y')a(x'-x_n,y'-y_n)\right]. \tag{19}$$

If the phase is known to be relatively smooth it is advantageous to model $h(x,y)$ using a set of polynomial basis functions, such as the Zernike polynomials, with coefficients c_k^h,

$$h(x,y) = |h(x,y)|\exp\left[i\sum_k c_k^h Z_k(x,y)\right], \tag{20}$$

and we can write

$$\frac{\partial}{\partial c_{k'}^h}\left[h(x,y)a(x-x_n,y-y_n)\right] = ia(x-x_n,y-y_n)h(x,y)Z_{k'}(x,y). \tag{21}$$

Using this with Eq. (16) gives

$$\frac{\partial E}{\partial c_{k'}^h} = 2\,\text{Im}\left[\sum_{n=1}^{q}\sum_{x,y} g_n^{w*}(x-x_n,y-y_n)a(x-x_n,y-y_n)h(x,y)Z_{k'}(x,y)\right]. \tag{22}$$

We can optimize over $|h(x,y)|$, the amplitude of $h(x,y)$ in Eqs. (17) and (20). Then

$$\frac{\partial}{\partial|h(x',y')|}\left[h(x,y)a(x-x_n,y-y_n)\right] = a(x-x_n,y-y_n)\exp\left[i\theta_h(x,y)\right]\delta(x,x';y,y') \tag{23}$$

which, when used with Eq. (16), gives

$$\frac{\partial E}{\partial|h(x',y')|} = -2\,\text{Re}\left[\sum_{n=1}^{q} g_n^{w*}(x'-x_n,y'-y_n)a(x'-x_n,y'-y_n)\exp\left[i\theta_h(x',y')\right]\right] \tag{24}$$

We can alternatively model $h(x,y)$ as the real and imaginary parts (complex value) of the field at each pixel,

$$h(x,y) = h_R(x,y) + ih_I(x,y). \tag{25}$$

Since

$$\frac{\partial}{\partial h_R(x',y')}\left[h(x,y)a(x-x_n,y-y_n)\right] = a(x-x_n,y-y_n)\delta(x,x';y,y'), \tag{26}$$

we have

$$\frac{\partial E}{\partial h_R(x',y')} = -\sum_{n=1}^{q} g_n^{w*}(x'-x_n,y'-y_n)a(x'-x_n,y'-y_n) + \text{c.c.} \tag{27}$$

Similarly we have

$$\frac{\partial E}{\partial h_I(x',y')} = -i\sum_{n=1}^{q} g_n^{w*}(x'-x_n,y'-y_n)a(x'-x_n,y'-y_n) + \text{c.c.} \tag{28}$$

Combining the real and imaginary parts gives

$$\frac{\partial E}{\partial h_R(x',y')} + i\frac{\partial E}{\partial h_I(x',y')} = -2\sum_{n=1}^{q} g_n^w(x'-x_n, y'-y_n)a^*(x'-x_n, y'-y_n). \quad (29)$$

Similarly expressing the subaperture transmittance $a(x,y)$ in terms of its amplitude and phase,

$$a(x,y) = |a(x,y)|\exp[i\theta_a(x,y)] \quad (30)$$

we have

$$\frac{\partial E}{\partial \theta_a(x',y')} = 2\,\text{Im}\left[\sum_{n=1}^{q} g_n^{w*}(x',y')h(x'+x_n, y'+y_n)a(x',y')\right]. \quad (31)$$

We can model $a(x,y)$ in terms of the coefficients of basis functions as well. In this situation it is useful to have two sets of coefficients, one set that is constant over all subaperture measurements, c_k^a, and one that varies for each translation of the subaperture, $c_{k,n}^a$ [2]. The latter coefficients allow us to include parameters in the optimization to account for focus variations or small lateral shifts of the CCD;

$$a_n(x,y) = |a(x,y)|\exp\left[i\sum_k \left(c_k^a + c_{k,n}^a\right)Z_k(x,y)\right]. \quad (32)$$

For the part that is constant across subaperture positions we have

$$\frac{\partial}{\partial c_{k'}^a}\left[h(x+x_n, y+y_n)a(x,y)\right] = ih(x+x_n, y+y_n)a_n(x,y)Z_{k'}(x,y). \quad (33)$$

Using this with Eq. (15) gives

$$\frac{\partial E}{\partial c_{k'}^a} = 2\,\text{Im}\left[\sum_{n=1}^{q}\sum_{x,y} g_n^{w*}(x,y)h(x+x_n, y+y_n)a_n(x,y)Z_{k'}(x,y)\right]. \quad (34)$$

For the part that varies across subaperture measurements we have

$$\frac{\partial}{\partial c_{k',n'}^a}\left[h(x+x_n, y+y_n)a(x,y)\right] = ih(x+x_n, y+y_n)a_n(x,y)Z_{k'}(x,y)\delta(n,n'). \quad (35)$$

Using this with Eq. (15) gives

$$\frac{\partial E}{\partial c_{k',n'}^a} = 2\,\text{Im}\left[\sum_{x,y} g_{n'}^{w*}(x,y)h(x+x_{n'}, y+y_{n'})a_{n'}(x,y)Z_{k'}(x,y)\right]. \quad (36)$$

We can optimize over the amplitude of $a(x,y)$, $|a(x,y)|$ in Eqs. (30) and (32). Then

$$\frac{\partial}{\partial|a(x',y')|}\left[h(x+x_n, y-y_n)a(x,y)\right] = h(x+x_n, y+y_n)\exp[i\theta_a(x,y)]\delta(x,x';y,y') \quad (37)$$

which when used with Eq. (15) gives

$$\frac{\partial E}{\partial|a(x',y')|} = -2\,\text{Re}\left[\sum_{n=1}^{q} g_n^{w*}(x',y')h(x'+x_n, y'+y_n)\exp[i\theta_a(x',y')]\right] \quad (38)$$

We can also use the complex values of $a(x,y)$ at each pixel,

$$a(x,y) = a_R(x,y) + ia_I(x,y). \tag{39}$$

The derivative with respect to the real part is

$$\frac{\partial}{\partial a_R(x',y')}\left[h(x+x_n,y+y_n)a(x,y)\right] = h(x+x_n,y+y_n)\delta(x,x';y,y'), \tag{40}$$

which gives

$$\frac{\partial E}{\partial a_R(x',y')} = -\sum_{n=1}^{q} g_n^{w*}(x',y')h(x'+x_n,y'+y_n) + \text{c.c.} \tag{41}$$

Similarly for the imaginary part,

$$\frac{\partial E}{\partial a_I(x',y')} = -i\sum_{n=1}^{q} g_n^{w*}(x',y')h(x'+x_n,y'+y_n) + \text{c.c.} \tag{42}$$

Combining the real and imaginary parts gives

$$\frac{\partial E}{\partial a_R(x',y')} + i\frac{\partial E}{\partial a_I(x',y')} = -2\sum_{n=1}^{q} g_n^{w}(x',y')h^*(x'+x_n,y'+y_n). \tag{43}$$

The final derivative that we have found to be useful is the derivative with respect to the subaperture locations (x_n,y_n). We are assuming here that the function that is shifted is $h(x,y)$. To compute the derivative

$$\frac{\partial}{\partial x_n}\left[h(x+x_n,y+y_n)a(x,y)\right] = ia(x,y)\frac{\partial}{\partial x_n}h(x+x_n,y+y_n), \tag{44}$$

we use the shift theorem,

$$h(x+x_n,y+y_n) = IDFT\left[\exp\left[i2\pi\left(\frac{x_n u'}{M}+\frac{y_n v'}{M}\right)\right]DFT[h(x,y)]\right]. \tag{45}$$

Evaluating this gives

$$\frac{\partial}{\partial x_{n'}}h(x+x_n,y+y_n) = IDFT\left[\frac{i2\pi u'}{M}\exp\left[i2\pi\left(\frac{x_n u'}{M}+\frac{y_n v'}{M}\right)\right]DFT[h(x,y)]\right]\delta(n,n'). \tag{46}$$

Substituting this into Eq. (15),

$$\frac{\partial E}{\partial x_{n'}} = \frac{4\pi}{M}\text{Im}\left\{\sum_{x,y} g_n^{w*}(x,y)a(x,y)IDFT\left[u'\exp\left[i2\pi\left(\frac{x_{n'}u'}{M}+\frac{y_{n'}v'}{M}\right)\right]DFT[h(x,y)]\right]\right\}. \tag{47}$$

Similarly for the y-shift,

$$\frac{\partial E}{\partial y_{n'}} = \frac{4\pi}{M}\text{Im}\left\{\sum_{x,y} g_n^{w*}(x,y)a(x,y)IDFT\left[v'\exp\left[i2\pi\left(\frac{x_{n'}u'}{M}+\frac{y_{n'}v'}{M}\right)\right]DFT[h(x,y)]\right]\right\}. \tag{48}$$

These expressions are similar to those in Ref. [5] for the error metric being the squared difference of amplitudes.