

Amplitude metrics for field retrieval with hard-edged and uniformly illuminated apertures

Samuel T. Thurman,* Ryan T. DeRosa, and James R. Fienup

The Institute of Optics, University of Rochester, Rochester, New York 14627, USA

*Corresponding author: fienup@optics.rochester.edu

Received September 17, 2008; revised January 5, 2009; accepted January 19, 2009;
posted January 26, 2009 (Doc. ID 101701); published February 27, 2009

In field retrieval, the amplitude and phase of the generalized pupil function for an optical system are estimated from multiple defocused measurements of the system point-spread function. A baseline field reconstruction algorithm optimizing a data consistency metric is described. Additionally, two metrics specifically designed to incorporate *a priori* knowledge about pupil amplitude for hard-edged and uniformly illuminated aperture systems are given. Experimental results demonstrate the benefit of using these amplitude metrics in addition to the baseline metric. © 2009 Optical Society of America
OCIS codes: 100.3190, 100.5070, 110.1220.

1. INTRODUCTION

In wavefront sensing by phase retrieval, the phase of the generalized pupil function for an optical system is estimated from a measurement of the system point-spread function (PSF) and knowledge of the pupil amplitude [1,2]. For special aperture shapes, both the pupil amplitude and phase can be retrieved from one PSF measurement and full [3] or partial [4] knowledge of the aperture shape. In phase-diverse phase retrieval, multiple PSF measurements with diverse amounts of defocus (or another known phase function) are used to avoid algorithm convergence problems associated with local minima and improve the fidelity of the retrieved phase [5–7].

Although it is computationally intensive, phase retrieval has a number of practical advantages over other wavefront sensing techniques. Interferometric methods require either a reference wavefront or an autocollimation flat having the same dimensions as the system entrance pupil, while phase retrieval does not. Shack–Hartman wavefront sensors cannot work with discontinuous wavefronts from segmented- or sparse-aperture telescopes, while phase retrieval can. Because of these and other considerations, phase retrieval is the planned approach for wavefront sensing on the James Webb Space Telescope [8]. Additionally, multiple PSF measurements can be used to jointly estimate both the pupil phase and amplitude [9–14]. Here, this approach is referred to as field retrieval. In conventional phase diversity (PD), multiple focus-diverse images of an extended, incoherent object are used to jointly reconstruct the object and estimate the pupil phase [5]. Like field retrieval, however, PD can be used to additionally estimate the pupil amplitude [15,16].

There are a number of scenarios in which knowledge of the pupil amplitude may be incomplete, requiring some level of pupil amplitude estimation. In [4], the orientation of obscuring secondary-mirror support struts and primary-mirror mounting pads and the location of the relay lens obscurations for the Hubble Space Telescope were

determined from one PSF measurement and an initial guess of an annular aperture for the pupil amplitude. When the pupil amplitude was unknown because of scintillation caused by imaging through atmospheric turbulence, higher-quality pupil phase estimates from PD were obtained by simultaneously estimating the pupil amplitude in [15]. In [14,16], computer simulations were used to investigate the use of field retrieval in determining the plate scale and the pupil geometry of sparse-aperture optical systems. The thesis of this paper is that the quality of pupil amplitude estimates from field retrieval can be improved through the use of amplitude metrics that incorporate *a priori* knowledge about hard-edged or uniformly illuminated apertures.

In Section 2, a baseline field retrieval algorithm is described. In Section 3, two amplitude metrics designed to incorporate *a priori* knowledge for hard-edged and uniformly illuminated pupils are proposed as enhancements to the baseline algorithm. Section 4 describes an experiment in which PSF measurements were made for an optical system with various pupil masks. In Section 5, field retrieval results obtained from these measurements are presented. These results demonstrate the benefits of using the amplitude metrics. Section 6 is a summary. Appendix A contains equations useful for implementing this field retrieval approach.

2. BASELINE ALGORITHM

In this section, a physical model for the PSF measurements as a function of the pupil amplitude and phase, the defocus distances, and the transverse detector shifts is outlined. Also, a data consistency metric based on the normalized mean-squared error between the physical model and the actual measurements is formulated. Additionally, a sieve method for regularizing a baseline algorithm based on optimization of the data consistency metric is described.

A. Physical Model

Our physical model is based on the 4F system shown in Fig. 1. Given estimates for the amplitude $\hat{A}(m, n)$ and the phase $\hat{\phi}(m, n)$ for a generalized pupil function, the optical field in the pupil plane can be written as

$$\hat{E}_p(m, n) = \hat{A}(m, n) \exp[i\hat{\phi}(m, n)], \quad (1)$$

where $m \in \{-M/2, (2-M)/2, \dots, (M-2)/2\}$ and $n \in \{-N/2, (2-N)/2, \dots, (N-2)/2\}$ are pupil plane sample indices, and M and N are the number of samples along the two Cartesian directions. Nonnegativity and normalization of the pupil amplitude are ensured by parameterizing $\hat{A}(m, n)$ in terms of a dummy function $\hat{B}(m, n)$:

$$\hat{A}(m, n) = \frac{MN|\hat{B}(m, n)|}{\sum_{(m', n')} |\hat{B}(m', n')|}. \quad (2)$$

The optical field in the nominal focal plane $\hat{E}_f(p, q)$ is given by the discrete Fourier transform (DFT) of $\hat{E}_p(m, n)$, i.e.,

$$\begin{aligned} \hat{E}_f(p, q) &= \text{DFT}[\hat{E}_p] = \frac{1}{\sqrt{MN}} \sum_{(m, n)} \hat{E}_p(m, n) \\ &\times \exp\left[-i2\pi\left(\frac{mp}{M} + \frac{nq}{N}\right)\right], \end{aligned} \quad (3)$$

where $p \in \{-M/2, (2-M)/2, \dots, (M-2)/2\}$ and $q \in \{-N/2, (2-N)/2, \dots, (N-2)/2\}$ are focal plane sample indices. Equation (3) is a discrete approximation of the standard Fourier transform-based equation for propagation between the pupil and the focal plane of a 4F optical system [17]. The focal plane sample spacings Δ_p and Δ_q are chosen to be equal to the detector pixel pitch Δ_d , i.e., $\Delta_p = \Delta_q = \Delta_d$. Thus, the pupil plane sample spacings are given by $\Delta_m = 1/(M\Delta_p) = 1/(M\Delta_d)$ and $\Delta_n = 1/(N\Delta_q) = 1/(N\Delta_d)$ in units of spatial frequency and $\lambda f \Delta_m$ and $\lambda f \Delta_n$ in units of physical length, where λ is the optical wavelength and f is the lens focal length.

An angular spectrum propagator can be used to propagate $\hat{E}_f(p, q)$ from the back focal plane of the final lens to the various defocus measurement planes. $\hat{U}_f(m, n)$, the angular spectrum of $\hat{E}_f(p, q)$, is given by

$$\hat{U}_f(m, n) = \frac{1}{\sqrt{MN}} \sum_{(p, q)} \hat{E}_f(p, q) \exp\left[-i2\pi\left(\frac{pm}{M} + \frac{qn}{N}\right)\right]. \quad (4)$$

The angular spectrum propagated to the k th defocus/ measurement plane $\hat{U}_k(m, n)$ is given by

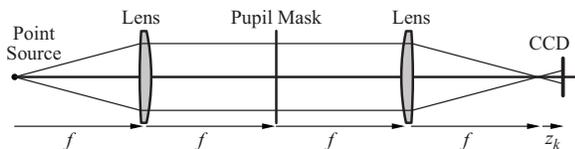


Fig. 1. Diagram of 4F system used for experiment.

$$\hat{U}_k(m, n) = \hat{U}_f(m, n) \exp\left(i2\pi\hat{z}_k \sqrt{\frac{1}{\lambda^2} - m^2\Delta_m^2 - n^2\Delta_n^2}\right), \quad (5)$$

where \hat{z}_k is the distance between the nominal focal plane and the k th defocus plane, and phase constants are ignored. The optical field in the k th defocus plane is given by $\hat{E}_k(p, q) = \text{IDFT}[\hat{U}_k]$. The computed intensity in the k th defocus plane is given by

$$\hat{I}_k(p, q) = |\hat{E}_k(p, q)|^2. \quad (6)$$

The detector impulse response and possible misregistrations of each frame of data are modeled in the Fourier domain. The detector transfer function is modeled as

$$H_d(m, n) = \text{sinc}\left(\frac{\sqrt{f_d}}{N}n\right) \text{sinc}\left(\frac{\sqrt{f_d}}{M}m\right), \quad (7)$$

where f_d is the area fill factor of the detector and it is assumed that the detector pixels are square.

The transfer function for a coordinate shift is defined as

$$H_{s,k}(m, n) = \exp\left[-i2\pi\left(\frac{m\hat{p}_{s,k}}{M} + \frac{n\hat{q}_{s,k}}{N}\right)\right], \quad (8)$$

where $\hat{p}_{s,k}$ and $\hat{q}_{s,k}$ are the transverse shifts along the Cartesian axes in units of pixels. $H_d(m, n)$ and $H_{s,k}(m, n)$ are included in the physical model by first computing $\hat{f}_k(m, n)$, the DFT of $\hat{I}_k(p, q)$, multiplying by the transfer functions,

$$\hat{g}_k(m, n) = H_{s,k}(m, n) H_d(m, n) \hat{f}_k(m, n), \quad (9)$$

and computing the inverse DFT to arrive at the modeled PSF $\hat{G}_k(p, q)$.

B. Data Consistency Metric

The agreement between $\hat{G}_k(p, q)$ and a set of actual PSF measurements $G_k(p, q)$ can be quantified using a weighted normalized mean-squared error (NMSE) metric [2,18], defined as

$$\Phi_d = \frac{1}{K} \sum_{k=1}^K \left\{ \frac{\sum_{(p, q)} W_k(p, q) [\alpha_k \hat{G}_k(p, q) - G_k(p, q)]^2}{\sum_{(p, q)} W_k(p, q) G_k^2(p, q)} \right\}, \quad (10)$$

where the coefficients α_k , which minimize the value of Φ_d for any given $\hat{G}_k(p, q)$, are given by

$$\alpha_k = \frac{\sum_{(p, q)} W_k(p, q) \hat{G}_k(p, q) G_k(p, q)}{\sum_{(p, q)} W_k(p, q) \hat{G}_k^2(p, q)}, \quad (11)$$

and $W_k(p, q)$ is a weighting function. Inserting Eq. (11) into Eq. (10) and simplifying yields

$$\Phi_d = 1 - \frac{1}{K} \sum_{k=1}^K \frac{\left[\sum_{(p,q)} W_k(p,q) \hat{G}_k(p,q) G_k(p,q) \right]^2}{\left[\sum_{(p,q)} W_k(p,q) G_k^2(p,q) \right] \left[\sum_{(p,q)} W_k(p,q) \hat{G}_k^2(p,q) \right]}. \quad (12)$$

The value of Φ_d is interpreted as the square of the fractional error between $\hat{G}_k(p,q)$ and $G_k(p,q)$, i.e., $\Phi_d=1$ corresponds to complete disagreement, $\Phi_d=0$ corresponds to exact agreement, and $\Phi_d=0.0025$ corresponds to an average root-mean-square (RMS) error of 5%. The baseline field retrieval approach is to use a conjugate-gradient (CG) nonlinear optimization routine to minimize Φ_d with respect to $\hat{B}(m,n)$, $\hat{\phi}(m,n)$, \hat{z}_k , $\hat{p}_{s,k}$, and $\hat{q}_{s,k}$.

C. Regularization

In many inverse problems, the incorporation of some sort of regularization against noise and artifacts is desirable. In conventional phase retrieval and PD, parameterization of the pupil phase as an expansion over a set of basis functions, e.g., Zernike polynomials, is a convenient and effective method for doing this [5]. In this approach, regularization is achieved by effectively reducing the solution space for $\hat{\phi}(m,n)$ to some submanifold that is spanned by the basis functions within MN -dimensional space. The choice of an appropriate set of basis functions, however, is uncertain when the pupil itself is uncertain in Eq. (1).

Another regularization approach, which will be used here, is the method of sieves [15,19,20]. In the CG routine, Φ_d is minimized by iteratively picking a direction within the solution space and performing a line search. The progress of the algorithm through the solution space to a final solution is thus determined in part by the rule for picking the search direction on each iteration. Normally, the search direction in a CG algorithm is a linear combination of the gradient of Φ_d for the current and previous iterations. The method of sieves involves modifying this rule by replacing the gradient components $\partial\Phi_d/\partial\hat{B}(m,n)$ and $\partial\Phi_d/\partial\hat{\phi}(m,n)$ with spatially smoothed versions of these quantities, i.e., $\partial\Phi_d/\partial\hat{B}(m,n)$ is replaced by

$$\sum_{(m',n')} \frac{\partial\Phi_d}{\partial\hat{B}(m',n')} s(m-m', n-n'), \quad (13)$$

where $s(m,n)$ is a smoothing kernel, and the gradient component $\partial\Phi_d/\partial\hat{\phi}(m,n)$ is replaced with an analogously smoothed quantity. If $s(m,n)$ is a low-pass smoothing kernel, e.g., a 2-D Gaussian, this approach causes the CG routine to converge on the coarse spatial features of $\hat{B}(m,n)$ and $\hat{\phi}(m,n)$ more quickly than on the fine spatial features, which helps the algorithm avoid problems with local minima and reduces high-spatial-frequency noise.

The results obtained in Section 5 were obtained using a Gaussian smoothing kernel with a FWHM of three pixels for $s(m,n)$ applied to both $\partial\Phi_d/\partial\hat{B}(m,n)$ and $\partial\Phi_d/\partial\hat{\phi}(m,n)$ for the first 100 iterations and to $\partial\Phi_d/\partial\hat{\phi}(m,n)$ thereafter. There were approximately 60 pupil samples across the 25.4 mm diameter circular aperture in the retrieval re-

sults. A FWHM value of three pixels was chosen to limit the retrieved pupil amplitude and phase initially to spatial frequencies less than or equal to ≈ 10 cycles per aperture. After the first 100 iterations, the smoothing was applied to $\partial\Phi_d/\partial\hat{\phi}(m,n)$, because we expected the pupil phase to be smooth, but not applied to $\partial\Phi_d/\partial\hat{B}(m,n)$ to allow the algorithm to retrieve the sharp aperture edges of the pupil amplitude.

3. AMPLITUDE METRICS

The two amplitude metrics described here are meant to incorporate specific knowledge about hard-edged or uniformly illuminated apertures into the field retrieval algorithm. The metrics are defined as

$$\Phi_1(\kappa_1) = \frac{1}{MN} \sum_{(m,n)} \Gamma[\hat{A}(m,n), \kappa_1] \quad (14)$$

and

$$\Phi_2(\kappa_2) = \sum_{(\mu,\eta) \in D} \left\{ \frac{1}{MN} \sum_{(m,n)} \Gamma[\hat{A}(m,n) - \hat{A}(m+\mu, n+\eta), \kappa_2] \right\}, \quad (15)$$

where

$$\Gamma(x, \kappa) = \begin{cases} \frac{2|x|^2}{3\kappa^2} - \frac{8|x|^3}{27\kappa^3} + \frac{|x|^4}{27\kappa^4}, & |x| \leq 3\kappa \\ 1, & |x| > 3\kappa \end{cases}, \quad (16)$$

(μ, η) are sample shift indices belonging to $D = \{(0,1), (1,0), (1,1), (1,-1)\}$, and κ_1 and κ_2 are adjustable parameters. Figure 2 shows a plot of $\Gamma(x, \kappa)$ along with $\Gamma'(x, \kappa)$ and $\Gamma''(x, \kappa)$, the first and second partial derivatives of $\Gamma(x, \kappa)$ with respect to x .

Since the sum of $\hat{A}(m,n)$ is a conserved quantity due to Eq. (2), the effect of using Φ_1 can be partially understood using the second derivative rule explained in [21]. Minimizing Φ_1 will tend to compress the histogram of $\hat{A}(m,n)$ for values $\hat{A}(m,n) < \kappa_1$, since $\Gamma''(x, \kappa) > 0$ for $|x| < \kappa$, and stretch the histogram for values $\kappa_1 < \hat{A}(m,n) < 3\kappa_1$, since $\Gamma''(x, \kappa) < 0$ for $\kappa < x < 3\kappa$. The first derivative $\Gamma'(x, \kappa)$ also

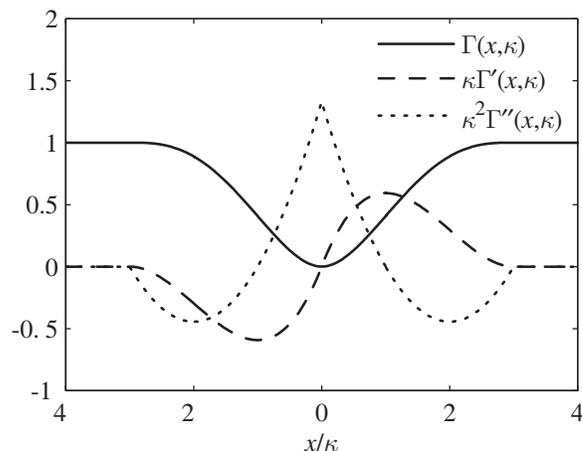


Fig. 2. Plot of $\Gamma(x, \kappa)$, $\Gamma'(x, \kappa)$, and $\Gamma''(x, \kappa)$.

plays a role in determining the effect of Φ_1 . Since $\Gamma'(x, \kappa) > 0$ for $|x| < 3\kappa$, use of Φ_1 always tends to reduce the values of $\hat{A}(m, n) < 3\kappa_1$, as long as there are some values of $\hat{A}(m, n) > 3\kappa_1$ that can be increased to conserve the sum of $\hat{A}(m, n)$. Note that the values of Φ_1 are insensitive to changes in values of $\hat{A}(m, n) > 3\kappa_1$, since both $\Gamma'(x, \kappa)$ and $\Gamma''(x, \kappa) = 0$ for $|x| > 3\kappa$. For a hard-edged aperture, $\hat{A}(m, n)$ should equal zero for points (m, n) outside the true support of the pupil. While use of Φ_d may yield small values of $\hat{A}(m, n)$ in these regions of the pupil plane, $\hat{A}(m, n)$ will often be nonzero there due to noise in the data $G_k(p, q)$, even with regularization. Additional iterations with both Φ_d and Φ_1 , using an appropriately chosen value of κ_1 , can further reduce the already small but nonzero values of $\hat{A}(m, n)$, hopefully leading to a better estimate $\hat{A}(m, n)$.

We explored a number of different functions $\Gamma(x, \kappa)$, and the results shown here are for the form given by Eq. (16). Previously, we used $\Gamma(x, \kappa) = x^2 / (\kappa^2 + x^2)$ [22], which is very similar to the form in Eq. (16) with the exception that $\Gamma'(x, \kappa) = 2\kappa^2 x / (x^2 + \kappa^2)^2$ and $\Gamma''(x, \kappa) = 2\kappa^2(\kappa^2 - 3x^2) / (x^2 + \kappa^2)^3$ do not equal zero for $|x| \gg \kappa$. Because of this, minimizing Φ_1 with this form of $\Gamma(x, \kappa)$ will not drive values of $\hat{A}(m, n) < \kappa_1$ to zero. Instead the histogram of $\hat{A}(m, n) < \kappa_1$ will be compressed about some small, nonzero value $< \kappa_1$ that is in equilibrium with the small penalty associated with increasing values of $\hat{A}(m, n) \gg \kappa_1$. The form of $\Gamma(x, \kappa)$ in Eq. (16), with $\Gamma'(x, \kappa) = \Gamma''(x, \kappa) = 0$ for $x > 3\kappa$, is such that values of $\hat{A}(m, n) < \kappa_1$ can be driven to zero through use of Φ_1 by increasing values of values of $\hat{A}(m, n) > 3\kappa_1$ without penalty. Equation (16) also has the handy feature of a continuous second derivative for all x except $x = 0$.

Since $|\hat{A}(m, n) - \hat{A}(m + \mu, n + \eta)|$ is not a conserved quantity, the effect of minimizing Φ_2 can be understood by considering $\Gamma'(x, \kappa)$. Φ_2 is minimized by reducing the magnitude of the differences between neighboring samples of $\hat{A}(m, n)$, with the value of Φ_2 being most sensitive to changes in values of $|\hat{A}(m, n) - \hat{A}(m + \mu, n + \eta)| = \kappa_2$. Similar to Φ_1 , the value of Φ_2 is insensitive to values of $|\hat{A}(m, n) - \hat{A}(m + \mu, n + \eta)| > 3\kappa_2$, since $\Gamma'(x, \kappa) = 0$ for $|x| > 3\kappa$. For a uniformly illuminated aperture, $\hat{A}(m, n)$ should be piecewise constant. Use of Φ_d alone generally will not yield a piecewise constant $\hat{A}(m, n)$, again due to noise if nothing else. Additional iterations with Φ_d and Φ_2 , with an appropriately chosen κ_2 , can yield a more piecewise constant

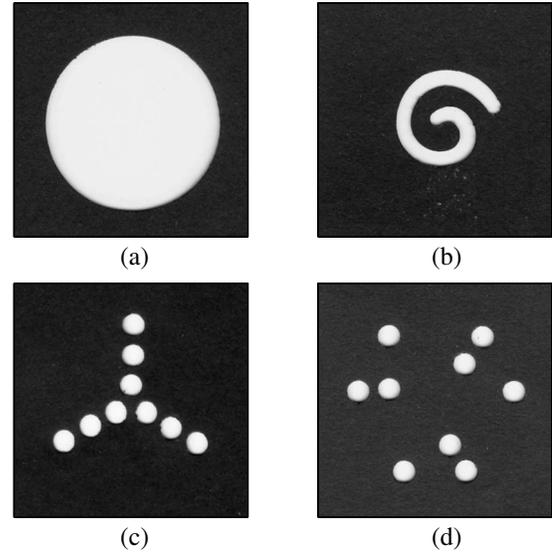


Fig. 3. Digital scans of the pupil amplitude masks used in the experiment: (a) circle, (b) spiral, (c) nine-aperture triarm, and (d) nine-aperture Golay.

$\hat{A}(m, n)$ by reducing small differences between neighboring samples while preserving sharp edges for which $|\hat{A}(m, n) - \hat{A}(m + \mu, n + \eta)| \gg \kappa_2$. Section 5 provides more details on choosing the values of κ_1 and κ_2 .

4. EXPERIMENT

Figure 1 shows the layout of the 4F optical system that was used for the experiment. A 5 μm diameter pinhole illuminated by a focused HeNe ($\lambda = 632.8$ nm) laser beam was used as a point source. The two identical lenses (Newport NPAC 091) had a focal length of $f = 500$ mm. The pupil plane contained a slide mount in which various amplitude masks were placed to define the aperture stop of the system. The amplitude masks were made by using a hole-punch or die-cutting tool to cut out various patterns in black cardstock. Figure 3 shows digital scans of each amplitude mask used in the experiment. For each amplitude mask, a number of PSF measurements were recorded with an 8-bit CCD camera (Imaging Source DMK21BF04). The camera was mounted on a manual translation stage to allow PSF measurements to be made in various defocus planes with nominal defocus distances of $z_k = \{-4, -2, 0, 2, 4\}$ mm. The detector pixel pitch was $\Delta_d = 5.6$ μm .

The encircled diameter of the amplitude masks was limited to no more than $D = 25.4$ mm, such that the mini-

Table 1. Details of Various Field Retrieval Estimates

Estimate	Starting Guess	Metric	Iterations	Result
1	$\hat{A}(m, n) = 1$ and $\hat{\phi}(m, n) = 0$	Φ_d	500	$\hat{A}_1(m, n)$ and $\hat{\phi}_1(m, n)$
2	$\hat{A}_1(m, n)$ and $\hat{\phi}_1(m, n)$	Φ_d	250	$\hat{A}_2(m, n)$ and $\hat{\phi}_2(m, n)$
3	$\hat{A}_1(m, n)$ and $\hat{\phi}_1(m, n)$	$\Phi_d + \lambda_1 \Phi_1(\kappa_1)$	250	$\hat{A}_3(m, n)$ and $\hat{\phi}_3(m, n)$
4	$\hat{A}_1(m, n)$ and $\hat{\phi}_1(m, n)$	$\Phi_d + \lambda_2 \Phi_2(\kappa_2)$	250	$\hat{A}_4(m, n)$ and $\hat{\phi}_4(m, n)$
5	$\hat{A}_1(m, n)$ and $\hat{\phi}_1(m, n)$	$\Phi_d + \lambda_1 \Phi_1(\kappa_1) + \lambda_2 \Phi_2(\kappa_2)$	250	$\hat{A}_5(m, n)$ and $\hat{\phi}_5(m, n)$

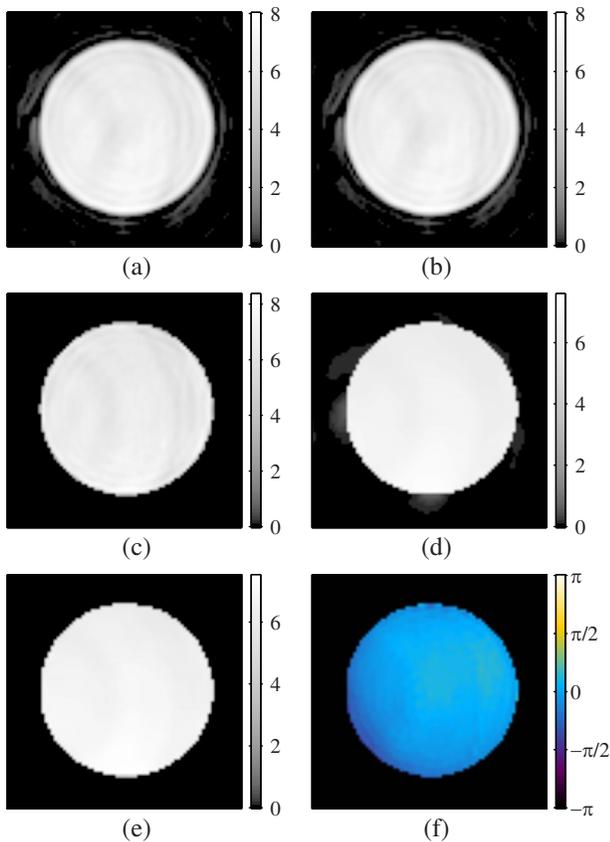


Fig. 4. (Color online) Field retrieval results for the circular pupil mask: (a) $\hat{A}_1(m,n)$, (b) $\hat{A}_2(m,n)$, (c) $\hat{A}_3(m,n)$, (d) $\hat{A}_4(m,n)$, (e) $\hat{A}_5(m,n)$, and (f) $\hat{\phi}_5(m,n)$ (in units of radians) with piston tip, tilt, and focus terms removed. Note that $\hat{\phi}_5(m,n)$ is shown only within the aperture at points where $\hat{A}_5(m,n) > \kappa_1$.

Table 2. Metric Values for Each Field Retrieval Result with the Circular Pupil Mask^a

Estimate	Φ_d	$\Phi_1(\kappa_1)$	$\Phi_2(\kappa_2)$
1	0.0013	0.1596	0.0995
2	0.0013	0.1596	0.0995
3	0.0013	0.1521	0.0660
4	0.0016	0.1520	0.0340
5	0.0017	0.1516	0.0351

^aUsing $\lambda_1=1$, $\kappa_1=1$, $\lambda_2=0.25$, and $\kappa_2=1$.

imum detector sampling ratio $Q=\lambda f/D\Delta_d=2.22$ was large enough to ensure that the PSF measurements were sampled above the Nyquist limit [23]. In addition to the PSF measurements, a number of dark and flat-field frames were recorded for detector calibration. The field retrieval data $G_k(p,q)$ were obtained by averaging 10 PSF measurements from each defocus plane and applying a dark subtraction and flat-field correction obtained from the detector calibration. Additionally, a constant term was subtracted from each PSF to account for an unknown detector bias. The weighting function $W_k(p,q)$ was not used, i.e., $W_k(p,q)=1$.

5. FIELD RETRIEVAL RESULTS

Results are presented for the five different estimation approaches listed in Table 1. The initial guess for the first

estimate was $\hat{A}(m,n)=1$, $\hat{\phi}(m,n)=0$, and the nominal values for \hat{z}_k . Starting guesses for $\hat{p}_{s,k}$ and $\hat{q}_{s,k}$ were obtained from the centroid of each measured PSF. Point-by-point estimates $\hat{A}_1(m,n)$ and $\hat{\phi}_1(m,n)$ were obtained after 500 CG iterations with just Φ_d . The second estimate was obtained by running an additional 250 CG iterations. Estimates 3, 4, and 5 were obtained by starting with $\hat{A}_1(m,n)$ and $\hat{\phi}_1(m,n)$ and running 250 CG iterations with Φ_1 , Φ_2 , and both Φ_1 and Φ_2 , respectively, in addition to Φ_d . The parameters κ_1 and κ_2 were picked for each pupil mask by inspection of $\hat{A}_1(m,n)$, based on knowledge of how Φ_1 influences the histogram of $\hat{A}(m,n)$ and how the effect of Φ_2 depends on the differences between neighboring $\hat{A}(m,n)$ samples (see Section 3). The values of the weighting parameters λ_1 and λ_2 , however, were chosen by trial and error to balance the effect of each amplitude metric with the data consistency metric. When the weighting parameter is too small, the amplitude metrics have only a minor influence on pupil retrieval results, yielding no benefit from Φ_1 or Φ_2 . If the weighting parameter is too large, the amplitude metrics dominate, yielding retrieval results with poor data consistency. Several values of λ_1 and λ_2 were tried for each pupil mask to determine the appropriate values between these two extremes.

Field retrieval results for the circular pupil mask obtained using just three defocus planes ($z_k=\{-4,0,4\}$ mm) are given in Fig. 4 and Table 2. Figure 4(a) shows that the pupil amplitude estimate obtained after 500 iterations with Φ_d agrees fairly well with the circular pupil mask shown in Fig. 3(a), but the nonzero values for $\hat{A}_1(m,n)$ outside the support of the circular aperture and the spatial structure within the aperture are not representative of the true pupil amplitude. Figure 4(b) shows that these features remain after an additional 250 iterations with Φ_d .

From Fig. 4(a), it appears that the maximum of $\hat{A}_1(m,n)$ outside the support of the circular aperture ≈ 1 , while the average value within the aperture appears to be ≈ 6 . Based on these observations and the properties of $\Phi_1(\kappa_1)$ discussed in Section 3, use of $\Phi_1(\kappa_1)$ with $\kappa_1=1$ (in addition to Φ_d) should reduce the amplitude of $\hat{A}(m,n)$ outside the aperture support and have only a minor influence inside the aperture, where $\hat{A}_1(m,n) \gg \kappa_1$, thus preserving the hard edge of the aperture. Figure 4(c) shows that this result is achieved for $\hat{A}_3(m,n)$. A value of $\lambda_1=1$, determined by trial and error, was used to obtain this result. For $\lambda_1 \ll 1$, the relative weighting of Φ_1 to Φ_d in the combined objective function was too small to yield the desired result. For $\lambda_1 \gg 1$, the relative weighting of Φ_1 was too large, resulting in an amplitude estimate with near-zero values inside the aperture.

Figure 5 shows histograms of $\hat{A}_1(m,n)$ and $\hat{A}_3(m,n)$ to better illustrate the effect of using $\Phi_1(\kappa_1)$, which compressed the histogram of retrieved amplitude values less than $\kappa_1=1$, driving them to zero. Use of $\Phi_1(\kappa_1)$ also stretched or spread out the histogram for amplitude values between κ_1 and $3\kappa_1$, resulting in an $\hat{A}_3(m,n)$ with only two samples in this range. While $\Phi_1(\kappa_1)$ basically ignores amplitude values greater than $3\kappa_1$, the values of

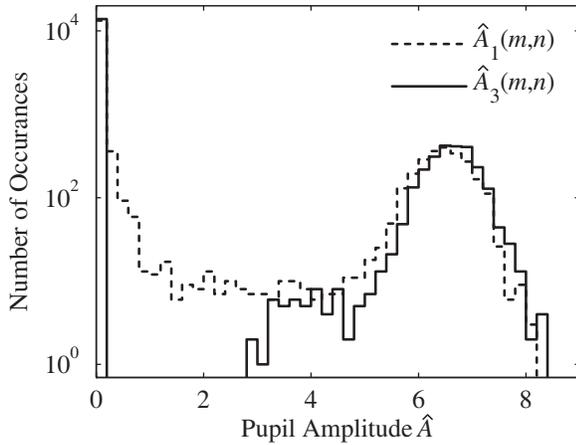


Fig. 5. Histograms of the retrieved pupil amplitude for the circular pupil mask: dashed curve, $\hat{A}_1(m,n)$ and solid curve $\hat{A}_3(m,n)$. The scale for the vertical axis is logarithmic.

$\hat{A}_3(m,n)$ in this range are shifted to slightly larger values to conserve the sum of the retrieved amplitude values.

From Fig. 4(a) it also appears that the amplitude fluctuations of $\hat{A}_1(m,n)$ within the aperture are ≈ 1 . Based on this observation and the properties of $\Phi_2(\kappa_2)$ described in Section 3, use of $\Phi_2(\kappa_2)$ with $\kappa_2=1$ (in addition to Φ_d)

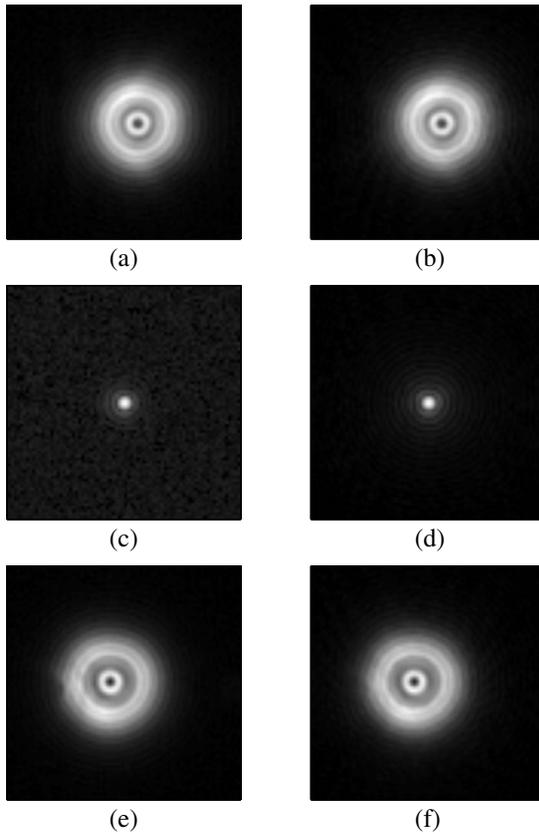


Fig. 6. Comparison between measured and modeled PSFs using $\hat{A}_5(m,n)$ and $\hat{\phi}_5(m,n)$ for the circular pupil mask. Measured PSFs $G_k(p,q)$ are shown in the left-hand column and modeled PSFs $\hat{G}_k(p,q)$ are shown in the right-hand column. The defocus distances for each PSF are (a), (b) -4 mm; (c), (d) 0 mm; and (e), (f) 4 mm.

should reduce these amplitude fluctuations while maintaining a sharp aperture edge, yielding an $\hat{A}(m,n)$ that is more representative of a hard-edged, uniformly illuminated pupil. Figure 4(d) shows that this is the result for $\hat{A}_4(m,n)$, obtained using $\lambda_2=0.25$. Also, note in Fig. 4(d) that the variation of $\hat{A}_4(m,n)$ outside the support of the circular aperture has been reduced by the use of Φ_2 , but there remain regions where $\hat{A}_4(m,n)$ is nonzero outside the aperture. Table 2 indicates that while the value of Φ_2 is reduced by this procedure, the value of Φ_d increases. This is not entirely unexpected, since use of Φ_d alone has a greater ability to fit noise in the data. Figure 4(e) shows that use of both Φ_1 and Φ_2 yields a piecewise uniform amplitude estimate with near-zero amplitude outside the aperture support for $\hat{A}_5(m,n)$. Figure 4(f) shows the pupil phase estimate $\hat{\phi}_5(m,n)$ for this case with piston, tip, tilt, and focus terms removed. For comparison, Fig. 6 shows both the measured PSFs $G_k(p,q)$ and the modeled PSFs

Table 3. Metric Values for Each Field Retrieval Result with the Spiral Pupil Mask^a

Estimate	Φ_d	$\Phi_1(\kappa_1)$	$\Phi_2(\kappa_2)$
1	0.0066	0.0742	0.1282
2	0.0051	0.0711	0.1408
3	0.0044	0.0280	0.0723
4	0.0069	0.0613	0.0147
5	0.0070	0.0251	0.0375

^aUsing $\lambda_1=2$, $\kappa_1=4$, $\lambda_2=0.3$, and $\kappa_2=4$.

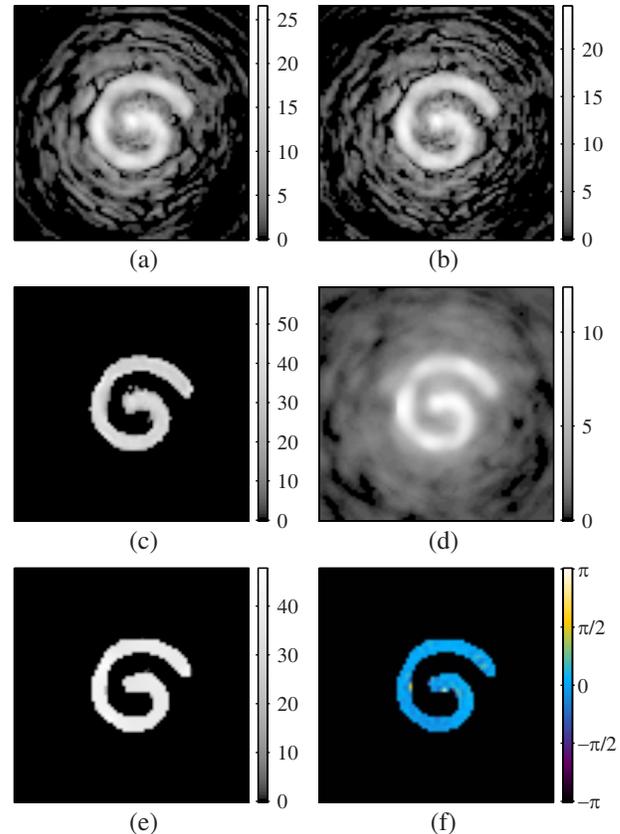


Fig. 7. (Color online) Same as Fig. 4, except for the spiral pupil mask.

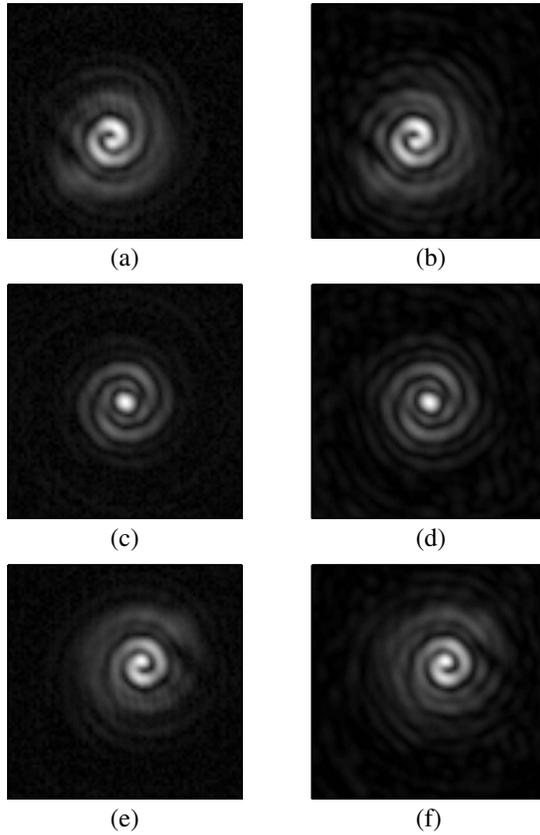


Fig. 8. Same as Fig. 6, except for the spiral pupil mask.

$\hat{G}_k(p, q)$ based on $\hat{A}_5(m, n)$ and $\hat{\phi}_5(m, n)$. Visually, the PSFs appear to agree well. Based on the value of Φ_d given in Table 2 for estimate 5, the RMS difference between the measured and modeled data is about 5%.

Table 3 and Figs. 7 and 8 show field retrieval results for the spiral pupil mask obtained from three PSF measurements with defocus amounts of -4 , 0 , and 4 mm. For this case, the pupil amplitude estimates obtained with just Φ_d , shown in Figs. 7(a) and 7(b), are rather noisy with many nonzero samples outside the support of the true aperture. Visually, the amplitude estimates obtained with use of Φ_1 and both Φ_1 and Φ_2 , shown in Figs. 7(c) and 7(e), respectively, better match the true pupil amplitude distribution shown in Fig. 3(b). Use of Φ_2 alone did not improve the amplitude estimate much, as is shown in Fig. 7(d). The retrieved pupil phase shown in Fig. 7(f) is nearly constant within the aperture, as expected.

Field retrieval results for the triarm and Golay pupil masks are given in Tables 4 and 5 and Figs. 9–12. These results were obtained using five PSF measurements with defocus amounts of -4 , -2 , 0 , 2 , and 4 mm. For both cases, the visual agreement between $\hat{A}(m, n)$ and the digital scan of each pupil mask is improved by the use of the amplitude metrics versus use of just Φ_d . While use of the amplitude metrics generally yields an $\hat{A}(m, n)$ that appears better visually, the value of Φ_d typically increases, indicating a loss in data consistency. As mentioned above, this is not unexpected, as Φ_d can more easily fit noise in the data when used alone than when used in conjunction with Φ_1 and/or Φ_2 . For the spiral and Golay apertures, how-

ever, use of Φ_1 resulted in a lower Φ_d than use of just Φ_d . This suggests that use of the amplitude metrics has the additional benefit of avoiding convergence problems associated with local minima of Φ_d in some cases.

While only three defocus positions were needed to obtain good results for the circular and spiral pupil masks, five defocus positions were needed for the sparse triarm and circular pupil masks. This may be due to a combination of effects associated with a limited capture range of the field retrieval algorithm and the initial guess for the pupil function. This claim is supported by the fact that we could obtain good results for the triarm and Golay pupil masks using five defocus positions for the first 100 iterations and only three defocus positions for the remaining iterations, while we did not obtain good results starting with only three defocus positions.

6. SUMMARY

Two metrics for incorporating *a priori* knowledge of hard-edged and uniformly illuminated pupil functions were implemented into a field retrieval algorithm. Experimental results indicate that use of these metrics in addition to a baseline data consistency metric yield amplitude estimates that appear to be more representative of the true pupil amplitude than does use of just the data consistency metric. The results also suggest that the amplitude metrics have the additional benefit of reducing convergence problems associated with local minima of the data consistency metric.

APPENDIX A: METRIC DERIVATIVES

For the nonlinear optimization algorithm, it is useful to have expressions for the partial derivatives of Φ_d , $\Phi_1(\kappa_1)$, and $\Phi_2(\kappa_2)$, with respect to $\hat{B}(m, n)$, $\hat{\phi}(m, n)$, \hat{z}_k , $\hat{p}_{s,k}$, and $\hat{q}_{s,k}$. The derivatives of Φ_d are obtained by first taking the partial derivative of Φ_d , given by Eq. (12), with respect to $\hat{G}_k(p, q)$

$$\frac{\partial \Phi_d}{\partial \hat{G}_k(p, q)} = \frac{2}{K} \frac{W_k(p, q) \left[\sum_{(p', q')} W_k(p', q') \hat{G}_k(p', q') G_k(p', q') \right]}{\left[\sum_{(p', q')} W_k(p', q') G_k^2(p', q') \right] \left[\sum_{(p', q')} W_k(p', q') \hat{G}_k^2(p', q') \right]^2} \times \left\{ \hat{G}_k(p, q) \left[\sum_{(p', q')} W_k(p', q') \hat{G}_k(p', q') G_k(p', q') \right] - G_k(p, q) \left[\sum_{(p, q)} W_k(p', q') \hat{G}_k^2(p', q') \right] \right\}. \quad (\text{A1})$$

To simplify later expressions, we define

$$\begin{aligned} \hat{g}_k^\dagger(m,n) &= \frac{\partial\Phi_d}{\partial \text{Re}[\hat{g}_k(m,n)]} + i \frac{\partial\Phi_d}{\partial \text{Im}[\hat{g}_k(m,n)]} \\ &= \frac{1}{\sqrt{MN}} \sum_{(p,q)} \frac{\partial\Phi_d}{\partial \hat{G}_k(p,q)} \exp\left[-i2\pi\left(\frac{mp}{M} + \frac{nq}{N}\right)\right]. \end{aligned} \tag{A2}$$

Using this expression along with Eqs. (8) and (9), two of

Table 4. Metric Values for Each Field Retrieval Result with the Triarm Pupil Mask^a

Estimate	Φ_d	$\Phi_1(\kappa_1)$	$\Phi_2(\kappa_2)$
1	0.0052	0.0407	0.1012
2	0.0042	0.0401	0.1045
3	0.0047	0.0265	0.0869
4	0.0087	0.0374	0.0581
5	0.0059	0.0300	0.0624

^aUsing $\lambda_1=5$, $\kappa_1=5$, $\lambda_2=0.6$, and $\kappa_2=5$.

Table 5. Metric Values for Each Field Retrieval Result with the Golay Pupil Mask^a

Estimate	Φ_d	$\Phi_1(\kappa_1)$	$\Phi_2(\kappa_2)$
1	0.0072	0.0398	0.1060
2	0.0056	0.0391	0.1082
3	0.0051	0.0288	0.0855
4	0.0094	0.0366	0.0689
5	0.0067	0.0306	0.0649

^aUsing $\lambda_1=5$, $\kappa_1=5$, $\lambda_2=0.6$, and $\kappa_2=5$.

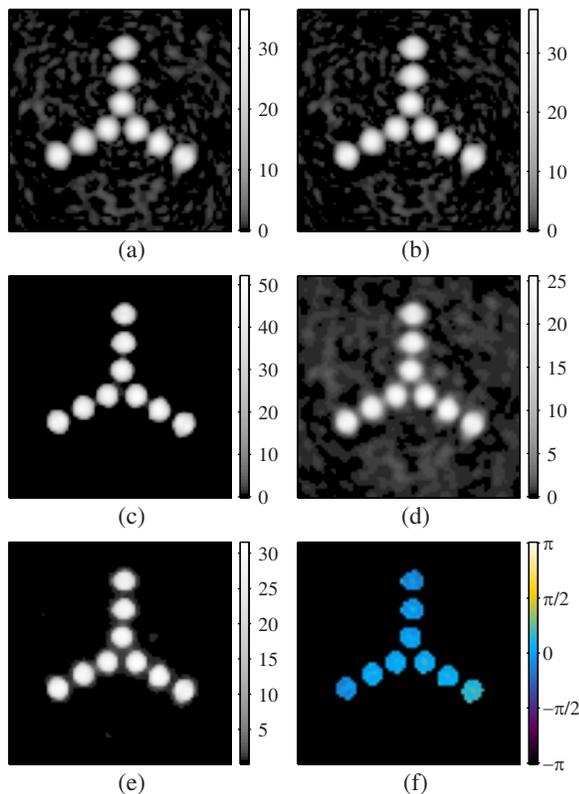


Fig. 9. (Color online) Same as Fig. 4, except for the triarm pupil mask.

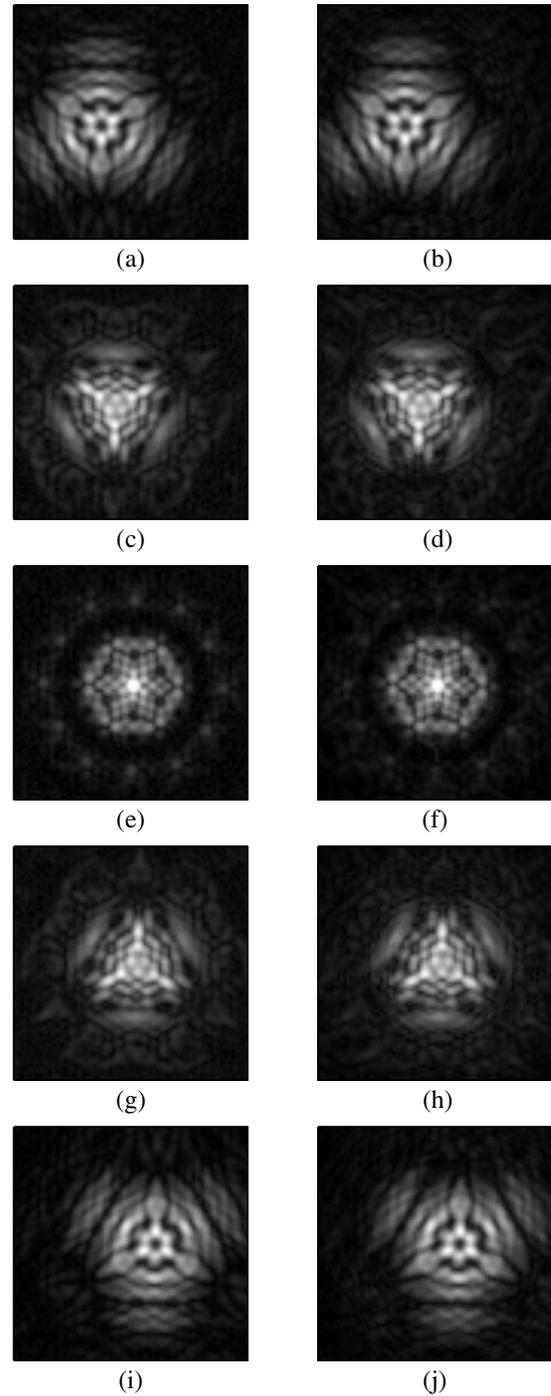


Fig. 10. Same as Fig. 6, except for the triarm pupil mask. The defocus distances for each PSF are (a), (b) -4 mm; (c), (d) -2 mm; (e), (f) 0 mm; (g), (h) 2 mm; and (i), (j) 4 mm.

the desired partial derivatives can be obtained, i.e.,

$$\frac{\partial\Phi_d}{\partial \hat{p}_{s,k}} = -\text{Im} \left[\sum_{(m,n)} \frac{2\pi m}{M} \hat{g}_k^\dagger(m,n) \hat{g}_k^*(m,n) \right], \tag{A3}$$

$$\frac{\partial\Phi_d}{\partial \hat{q}_{s,k}} = -\text{Im} \left[\sum_{(m,n)} \frac{2\pi n}{N} \hat{g}_k^\dagger(m,n) \hat{g}_k^*(m,n) \right]. \tag{A4}$$

Using Eqs. (6) and (9), we can write

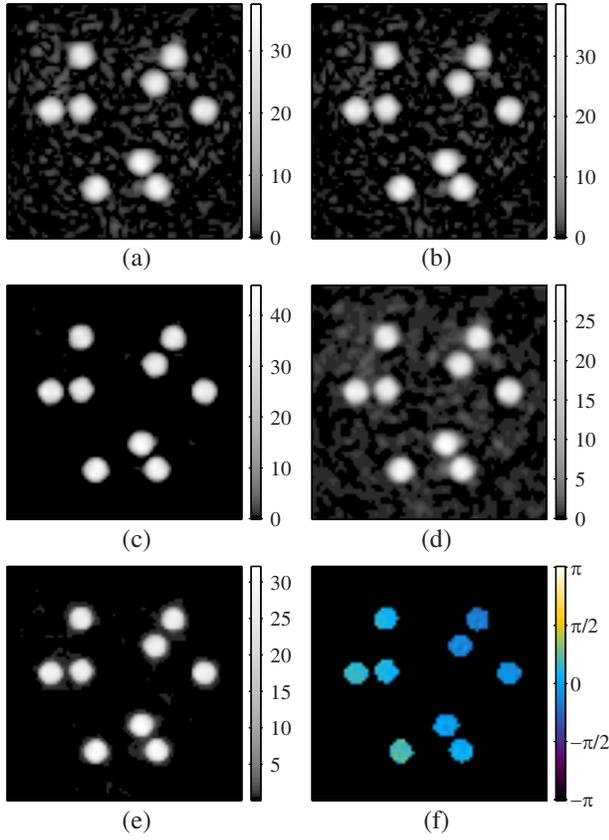


Fig. 11. (Color online) Same as Fig. 4, except for the Golay pupil mask.

$$\frac{\partial \Phi_d}{\partial \hat{I}_k(p,q)} = \frac{1}{\sqrt{MN}} \sum_{(m,n)} \hat{f}_k^\dagger(m,n) \exp \left[i2\pi \left(\frac{pm}{M} + \frac{qn}{N} \right) \right], \quad (\text{A5})$$

where

$$\hat{f}_k^\dagger(m,n) = \hat{H}_{s,k}^*(m,n) H_d^*(m,n) \hat{g}_k^\dagger(m,n). \quad (\text{A6})$$

Again, the following terms are defined to simplify notation:

$$\hat{E}_k^\dagger(p,q) = 2\hat{E}_k(p,q) \frac{\partial \Phi_d}{\partial \hat{I}_k(p,q)}, \quad (\text{A7})$$

$$\hat{U}_k^\dagger(m,n) = \frac{1}{\sqrt{MN}} \sum_{(p,q)} \hat{E}_k^\dagger(p,q) \exp \left[-i2\pi \left(\frac{mp}{M} + \frac{nq}{N} \right) \right]. \quad (\text{A8})$$

Using Eqs. (5) and (A8), we can write one more of the desired partial derivatives,

$$\frac{\partial \Phi_d}{\partial \hat{z}_k} = \text{Im} \left[\sum_{(m,n)} 2\pi \sqrt{\frac{1}{\lambda^2} - m^2 \Delta_m^2 - n^2 \Delta_n^2} \hat{U}_k^\dagger(m,n) \hat{U}_k^*(m,n) \right]. \quad (\text{A9})$$

To continue, we define

$$\hat{U}_f^\dagger(m,n) = \sum_k \hat{U}_k^\dagger(m,n) \exp \left(-i2\pi \hat{z}_k \sqrt{\frac{1}{\lambda^2} - m^2 \Delta_m^2 - n^2 \Delta_n^2} \right), \quad (\text{A10})$$

$$\hat{E}_f^\dagger(p,q) = \frac{1}{\sqrt{MN}} \sum_{(m,n)} \hat{U}_f^\dagger(m,n) \exp \left[i2\pi \left(\frac{pm}{M} + \frac{qn}{N} \right) \right], \quad (\text{A11})$$

$$\hat{E}_p^\dagger(m,n) = \frac{1}{\sqrt{MN}} \sum_{(p,q)} \hat{E}_f^\dagger(p,q) \exp \left[i2\pi \left(\frac{mp}{M} + \frac{nq}{N} \right) \right]. \quad (\text{A12})$$

Using Eqs. (1) and (A12), we can write the following partial derivatives of Φ_d , i.e.,

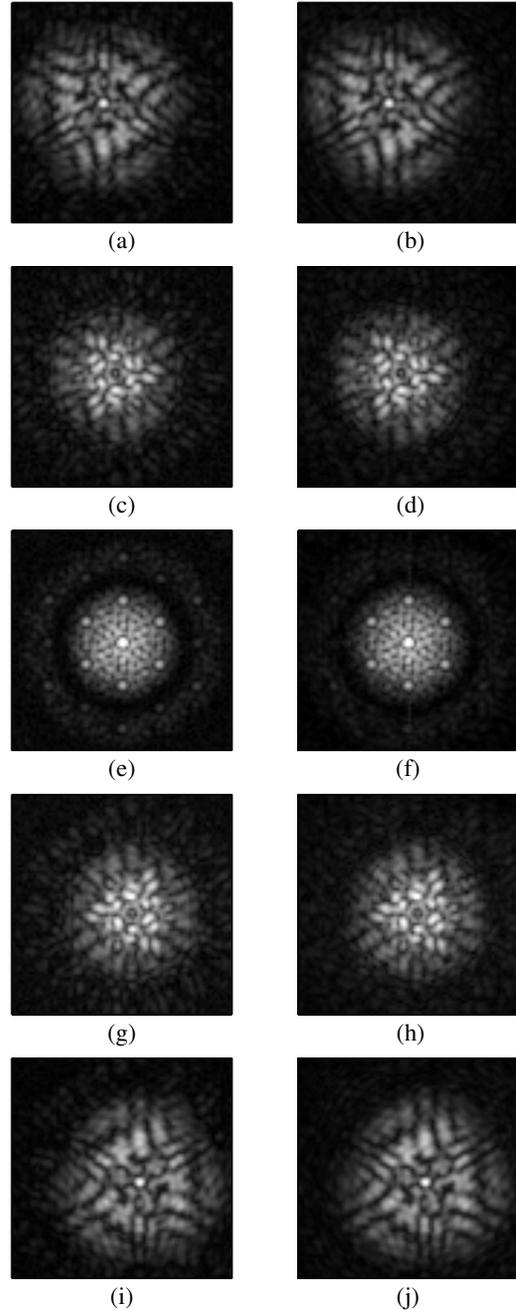


Fig. 12. Same as Fig. 10, except for the Golay pupil mask.

$$\frac{\partial \Phi_d}{\partial \hat{\phi}(m,n)} = \text{Im}[\hat{E}_p^\dagger(m,n)\hat{E}_p^*(m,n)], \quad (\text{A13})$$

$$\frac{\partial \Phi_d}{\partial \hat{A}(m,n)} = \text{Re}\{\hat{E}_p^\dagger(m,n)\exp[-i\hat{\phi}(m,n)]\}. \quad (\text{A14})$$

By Eq. (2) and the chain rule, the corresponding partial derivative with respect to $\hat{B}(m,n)$ is

$$\frac{\partial \Phi_d}{\partial \hat{B}(m,n)} = \frac{\text{sgn}[\hat{B}(m,n)]}{\sum_{(m',n')} |\hat{B}(m,n)|} \left[MN \frac{\partial \Phi_d}{\partial \hat{A}(m,n)} - \sum_{(m',n')} \frac{\partial \Phi_d}{\partial \hat{A}(m',n')} \hat{A}(m',n') \right]. \quad (\text{A15})$$

The partial derivatives with respect to $\hat{B}(m,n)$ are all that are needed for $\Phi_1(\kappa_1)$ and $\Phi_2(\kappa_2)$. Differentiating Eqs. (14)–(16) with respect to $\hat{A}(m,n)$ yields

$$\frac{\partial \Phi_1(\kappa_1)}{\partial \hat{A}(m,n)} = \frac{1}{MN} \Gamma'[\hat{A}(m,n), \kappa_1], \quad (\text{A16})$$

$$\begin{aligned} \frac{\partial \Phi_2(\kappa_2)}{\partial \hat{A}(m,n)} &= \sum_{(\mu,\eta) \in D} \frac{1}{MN} \{\Gamma'[\hat{A}(m,n) - \hat{A}(m+\mu, n+\eta), \kappa_2] \\ &\quad - \Gamma'[\hat{A}(m-\mu, n-\eta) - \hat{A}(m,n), \kappa_2]\}, \end{aligned} \quad (\text{A17})$$

where

$$\Gamma'(x, \kappa) = \begin{cases} \text{sgn}(x) \left(\frac{4|x|}{3\kappa^2} - \frac{8|x|^2}{9\kappa^3} + \frac{4|x|^3}{27\kappa^4} \right), & |x| \leq 3\kappa \\ 0, & |x| > 3\kappa \end{cases}. \quad (\text{A18})$$

The corresponding partial derivatives with respect to $\hat{B}(m,n)$ are given by Eq. (A15) with Φ_d replaced by $\Phi_1(\kappa_1)$ and $\Phi_2(\kappa_2)$.

ACKNOWLEDGMENTS

This work was funded in part by NASA Goddard Space Flight Center (GSF), Lockheed Martin, and the National Science Foundation (NSF) through the Research Experience for Undergraduates Program. Portions of this work were presented in [22].

REFERENCES

- J. R. Fienup, "Phase retrieval algorithms: a comparison," *Appl. Opt.* **21**, 2758–2769 (1982).
- J. R. Fienup, "Phase-retrieval algorithms for a complicated optical system," *Appl. Opt.* **32**, 1737–1746 (1993).
- J. N. Cederquist, J. R. Fienup, C. C. Wackerman, S. R. Robinson, and D. Kryskowski, "Wave-front phase estimation from Fourier intensity measurements," *J. Opt. Soc. Am. A* **6**, 1020–1026 (1989).
- J. R. Fienup, J. C. Marron, T. J. Schulz, and J. H. Seldin, "Hubble Space Telescope characterized by using phase-retrieval algorithms," *Appl. Opt.* **32**, 1747–1767 (1993).
- R. A. Gonsalves, "Phase retrieval and diversity in adaptive optics," *Opt. Eng. (Bellingham)* **21**, 829–832 (1982).
- J. R. Fienup, "Phase retrieval for undersampled broadband images," *J. Opt. Soc. Am. A* **16**, 1831–1839 (1999).
- B. H. Dean and C. W. Bowers, "Diversity selection for phase-diverse phase retrieval," *J. Opt. Soc. Am. A* **20**, 1490–1504 (2003).
- J. J. Green, B. H. Dean, C. M. Ohara, and D. C. Redding, "Target selection and imaging requirements for JWST fine phasing," *Proc. SPIE* **5487**, 994–1003 (2004).
- D. L. Misell, "A method for the solution of the phase problem in electron microscopy," *J. Phys. D* **6**, L6–L9 (1973).
- C. Roddier and F. Roddier, "Combined approach to Hubble Space Telescope wave-front distortion analysis," *Appl. Opt.* **32**, 2992–3008 (1993).
- Y. Zhang, G. Pedrini, W. Osten, and H. Tiziani, "Whole optical wave field reconstruction from double or multi in-line holograms by phase retrieval algorithm," *Opt. Express* **11**, 3234–3241 (2003).
- G. Pedrini, W. Osten, and Y. Zhang, "Wave-front reconstruction from a sequence of interferograms recorded at different planes," *Opt. Lett.* **30**, 833–835 (2005).
- P. Almero, G. Pedrini, and W. Osten, "Complete wavefront reconstruction using sequential intensity measurements of a volume speckle field," *Appl. Opt.* **45**, 8596–8605 (2006).
- G. R. Brady and J. R. Fienup, "Nonlinear optimization algorithm for retrieving the full complex pupil function," *Opt. Express* **14**, 474–486 (2006).
- S. M. Jefferies, M. Lloyd-Hart, E. K. Hege, and J. Georges, "Sensing wave-front amplitude and phase with phase diversity," *Appl. Opt.* **41**, 2095–2102 (2002).
- J. H. Seldin and R. G. Paxman, "Joint estimation of amplitude and phase from phase-diversity data," in *Adaptive Optics: Analysis and Methods/Computational Optical Sensing and Imaging/Information Photonics/Signal Recovery and Synthesis Topical Meetings* on CD-ROM, Technical Digest (Optical Society of America, 2005), paper JTub4.
- J. W. Goodman, *Introduction to Fourier Optics*, 3rd ed. (Roberts, 2005).
- J. R. Fienup, "Invariant error metrics for image reconstruction," *Appl. Opt.* **36**, 8352–8357 (1997).
- D. L. Snyder and M. I. Miller, "The use of sieves to stabilize images produced with the EM algorithm for emission tomography," *IEEE Trans. Nucl. Sci.* **32**, 3864–3871 (1985).
- D. R. Gerwe, M. Jain, B. Calef, and C. Luna, "Regularization for non-linear image restoration using a prior on the object power spectrum," *Proc. SPIE* **5896**, 21–36 (2005).
- J. R. Fienup and J. J. Miller, "Aberration correction by maximizing generalized sharpness metrics," *J. Opt. Soc. Am. A* **20**, 609–620 (2003).
- S. T. Thurman, R. T. DeRosa, and J. R. Fienup, "Amplitude metrics for field retrieval," presented at the OSA Frontiers in Optics Meeting, Rochester, NY, October 2006.
- R. D. Fiete, "Image quality and $\lambda FN/p$ for remote sensing systems," *Opt. Eng. (Bellingham)* **38**, 1229–1240 (1999).