Joint estimation of object and aberrations by using phase diversity

Richard G. Paxman, Timothy J. Schulz, and James R. Fienup

Optical and Infrared Science Laboratory, Environmental Research Institute of Michigan,
P. O. Box 134001, Ann Arbor, Michigan 48113-4001

The joint estimation of an object and the aberrations of an incoherent imaging system from multiple images incorporating phase diversity is investigated. Maximum-likelihood estimation is considered under additive Gaussian and Poisson noise models. Expressions for an aberration-only objective function that accommodates an arbitrary number of diversity images and its gradient are derived for the case of a Gaussian noise model. Expressions for the log-likelihood function and its gradient are presented for the case of Poisson noise. An expectation-maximization algorithm that enforces a nonnegativity constraint in a natural fashion is constructed for use in the Poisson noise case.

1. INTRODUCTION

The resolution of an incoherent imaging system is often limited by phase aberrations. Phase aberrations arise from a variety of sources including atmospheric turbulence, misaligned optics in phased-array systems, and improper mirror figure. Knowledge of phase aberrations affords either their correction by using adaptive optics or postdetection deblurring of the imagery. Phase aberrations can be measured directly. For example, the Hartmann–Shack wave-front sensor is often used for measuring atmospheric turbulence. In addition, laser interferometers and lateral-effect detectors have been used to measure piston and tilt misalignments in phased-array telescopes.[1] These techniques require considerable additional optical hardware that could also be subject to misalignments.

Phase aberrations may also be inferred directly from the image data. For example, phase-retrieval methods have been used in conjunction with knowledge of the pupil function to estimate phase aberrations from point-spread function (PSF) data.[2] This approach was recently applied to images of stars from the Hubble Space Telescope to diagnose errors in the mirror figure.[3,4] In many imaging scenarios, however, a point object is not available. Even for astronomical applications, the assumption of a point object involves some risk owing to the abundance of binary stars.

A technique known as phase diversity can also be used to infer phase aberrations from image data while accommodating extended objects or even scenes. The technique requires the collection of two or more images. One of these images is the conventional focal-plane image that has been degraded by the unknown aberrations. Additional images of the same object are formed by perturbing these unknown aberrations in some known fashion. For example, a simple translation of the detector array along the optical axis further degrades the imagery with a known amount of defocus. The quadratic phase error introduced by this intentional defocus is an example of one

type of phase diversity. The goal is to identify a combination of object and aberrations that is consistent with all the collected images, given the known phase diversities.

The phase-diversity technique offers several advantages over other aberration-sensing methods. The optical hardware required is modest. For example, a simple beam splitter and a second detector array will permit the simultaneous collection of the two images, as illustrated in Fig. 1. In addition, the method relies heavily on an external reference: the object being imaged. Therefore the method is less susceptible to systematic errors introduced by optical hardware. The technique also works well for extended objects. Finally, each photon is used for both imaging and aberration estimation. This may be preferable to the strategy of diverting valuable photons from the imagery to a separate wave-front sensor, where they are used solely for aberration estimation.

The use of phase diversity to infer aberrations when imaging incoherently illuminated, extended objects was proposed by Gonsalves.[5,6] Gonsalves derived an objective function for the estimation of aberration parameters that requires two collected images but does not explicitly depend on an object estimate. Conventional nonlinear optimization has been applied to this objective function in simulation experiments that resulted in the accurate estimation of aberration parameters.[6–8] Once these estimates are known, an estimate of the system optical transfer function (OTF) can be constructed, and the object can be restored by using, for example, Wiener filtering. Simulation studies have also been performed to investigate the sensitivity of the phase-diversity method to additive noise and to certain systematic errors.[8] The ability to operate on image subframes successfully, which is a fundamental requirement in the application of phase diversity to scenes, has also been demonstrated.[8]

In this paper we examine the phase-diversity technique from the viewpoint of maximum-likelihood estimation. We also generalize the technique to accommodate an arbitrary number of diversity measurements. In Section 2 we pose the problem of jointly estimating the object and aber-
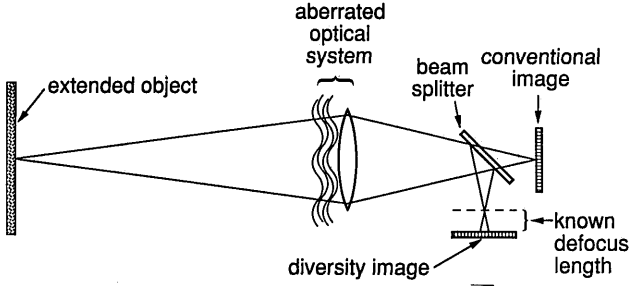
Fig. 1. Optical layout of a phase-diversity system. The conventional image is degraded by aberrations in the optical system. The diversity image is degraded by the combination of the same aberrations and a known amount of defocus.

rations from diversity data. This problem is considered under an additive Gaussian noise model in Section 3. We derive an objective function that does not depend explicitly on the object and that yields the maximum-likelihood estimate (MLE) for the aberration parameters when maximized. This objective function is shown to be a generalization of the Gonsalves objective function. An expression for the gradient of this objective function is also derived. The computational burden incurred in evaluating this gradient expression is surprisingly light. In Section 4 the same problem is considered under a Poisson (photon-limited) noise model. In this case an objective function that does not depend explicitly on the object has not been found. However, an expression for the gradient of the log-likelihood function is derived and found to be computationally tractable. This suggests the use of a gradient-search technique to maximize the log-likelihood function over the set of object pixels and aberration parameters. The gradient of the log likelihood is also given for the special case in which the object is known *a priori* to be pointlike. Finally, we construct an expectation-maximization (EM) algorithm for the joint estimation of an object and the aberrations. One advantage of this iterative algorithm is that object estimates are constrained to be nonnegative in a natural fashion.

2. STATEMENT OF PROBLEM

Frequently, space-invariant incoherent imagery is formed continuously through a convolution operation and detected discretely with a detector array. The analysis for our problem simplifies, however, if we model the object and its Fourier transform as discrete arrays. Accordingly, the incoherent image-formation process is approximated by the following discrete and cyclic convolution:

$$g_k(x) = \sum_{x' \in \chi} f(x')s_k(x - x') \tag{1}$$

$$\equiv f * s_k(x), \tag{2}$$

where f is the object array, s_k is a PSF having diversity k, g_k is the kth diversity image, and x is a two-dimensional coordinate. We treat the object, the PSF's, and the images as periodic arrays with a period cell of size $N \times N$. These arrays are completely specified by their functional values on the set χ, where

$$\chi = \{0, 1, \ldots, N - 1\} \times \{0, 1, \ldots, N - 1\}. \tag{3}$$

Diversity is introduced by including a known phase function in the generalized pupil function of the system:

$$H_k(u) = |H_k(u)|\exp\{i[\phi(u) + \theta_k(u)]\}, \tag{4}$$

where ϕ is the unknown phase-aberration function that we would like to estimate, θ_k is a known phase function associated with the kth diversity image, and $u \in \chi$. Phase diversity can be created, for example, by intentionally defocusing the system by known amounts. Note that in general each binary pupil mask $|H_k|$ could also change with the differing diversity measurements. It is often convenient to parameterize the unknown phase-aberration function:

$$\phi(u) = \sum_{j=1}^{J} \alpha_j \phi_j(u), \tag{5}$$

where J coefficients in the set $\{\alpha_j\}$ serve as parameters and $\{\phi_j\}$ is a convenient set of basis functions, such as discretized Zernike polynomials for a monolithic aperture or piston and tilt basis functions, used to represent misalignments in a phased-array system. The form and the number of basis functions employed will depend on the nature of the aberrations involved. However, there is no loss of generality in parameterizing the phase function since the parameters could represent the point-by-point phase values, in which case the basis functions will be Kronecker delta functions:

$$\phi_j(u) = \delta_{uu_j} \equiv \begin{cases} 1 & u = u_j \\ 0 & u \neq u_j \end{cases}. \tag{6}$$

For notational convenience we construct a J-dimensional aberration parameter vector α with elements $\alpha_1, \alpha_2, \ldots, \alpha_J$.

The inverse discrete Fourier transform of the generalized pupil function gives the impulse-response function for a coherently illuminated object:

$$h_k(x) = \frac{1}{N^2} \sum_{u \in \chi} H_k(u)\exp(i2\pi\langle u, x\rangle/N), \tag{7}$$

where $\langle \cdot, \cdot \rangle$ represents an inner product. The incoherent PSF is just the squared modulus of the coherent impulse-response function:

$$s_k(x) = |h_k(x)|^2. \tag{8}$$

The PSF depends on the aberration parameters through Eqs. (4), (5), (7), and (8). The noiseless image g_k depends, in turn, on the PSF through Eq. (1). Of course any detected imagery will contain noise. The relationship between a noiseless image and the actual detected image d_k will depend on the specific noise mechanisms.

The problem that we address may now be stated. Given the set of K detected diversity images $\{d_k\}$, the corresponding set of phase-diversity functions $\{\theta_k\}$, and the binary pupil functions $\{|H_k|\}$, estimate the object f and the aberration parameters α. We consider this problem in the cases for which the noise is modeled as additive Gaussian and Poisson.

3. ADDITIVE GAUSSIAN NOISE CASE

We now consider the case in which the noise at each detector element is modeled as an additive, independent, and

identically distributed random variable with a zero-mean Gaussian probability density having a variance $\sigma_n{}^2$. Such a model would be appropriate, for example, if the dominant noise were thermal noise. In this case each detected diversity image d_k is related to the corresponding noiseless diversity image g_k as follows:

$$d_k(x) = g_k(x) + n_k(x) \tag{9}$$

$$= f * s_k(x) + n_k(x), \tag{10}$$

where n_k represents the additive noise. Note that, because of the noise component, $d_k(x)$ will be a random variable with a normal probability density:

$$p[d_k(x); f, \alpha] = \frac{1}{(2\pi\sigma_n{}^2)^{1/2}} \exp\left\{ -\frac{[d_k(x) - f * s_k(x)]^2}{2\sigma_n{}^2} \right\}. \tag{11}$$

The probability density for realizing an entire data set $\{d_k\}$, consisting of all the pixels in each detected diversity image, is given by

$$p(\{d_k\}; f, \alpha) = \prod_{k=1}^{K} \prod_{x \in \chi} \frac{1}{(2\pi\sigma_n{}^2)^{1/2}}$$
$$\times \exp\left\{ -\frac{[d_k(x) - f * s_k(x)]^2}{2\sigma_n{}^2} \right\}. \tag{12}$$

The MLE is the estimate that is most likely to have produced a specific measurement.[9] It is found by maximizing the likelihood function [Eq. (12) evaluated with a specific measurement] with respect to f and α. The maximization is more easily carried out on a modified log-likelihood function:

$$L(f, \alpha) = -\sum_{k=1}^{K} \sum_{x \in \chi} [d_k(x) - f * s_k(x)]^2, \tag{13}$$

which is obtained by taking the natural logarithm of the likelihood function and dropping an inconsequential constant term and scale factor. For convenience we refer to L as the log-likelihood function. Applying discrete versions of both Parseval's theorem and the convolution theorem[10] to Eq. (13), we have that

$$L(f, \alpha) = -\frac{1}{N^2} \sum_{k=1}^{K} \sum_{u \in \chi} |D_k(u) - F(u)S_k(u)|^2, \tag{14}$$

where D_k, F, and S_k are discrete Fourier transforms of d_k, f, and s_k, respectively. Recall that S_k has the role of an unnormalized OTF.[11] Note that maximizing the log-likelihood function is equivalent to minimizing the sum of squared differences, which is the error metric used by Gonsalves.

The application of nonlinear optimization techniques to the log-likelihood function could be used to estimate the object and aberration functions simultaneously. This would require searching over a parameter space of large dimension, with object pixel values and aberration parameters serving as the axes of the space. Gonsalves showed, for the case of $K = 2$, that the aberration parameters can be estimated by optimizing an objective function that

depends explicitly on the aberration parameters but only implicitly on the object pixel values.[5] The result is a significant reduction in the dimension of the parameter space over which a numerical optimization is performed. This approach is made possible by the existence of a closed-form expression for the object that maximizes the log-likelihood function, given a fixed aberration function:

$$F_M(u) = \frac{D_1(u)S_1{}^*(u) + D_2(u)S_2{}^*(u)}{|S_1(u)|^2 + |S_2(u)|^2}, \tag{15}$$

where an asterisk used as a superscript implies complex conjugation. Substituting this expression into the log-likelihood function [Eq. (14)] and dropping the $1/N^2$ scale factor yields a new objective function that does not depend explicitly on an object estimate:

$$L_M(\alpha) = -\sum_{u \in \chi} \frac{|D_1(u)S_2(u) - D_2(u)S_1(u)|^2}{|S_1(u)|^2 + |S_2(u)|^2}, \tag{16}$$

where it has been assumed that $S_1(u)$ and $S_2(u)$ do not simultaneously go to zero for $u \in \chi$. It is important to recognize that maximizing the Gonsalves objective function L_M yields the MLE for the aberration parameters explicitly and the object parameters implicitly, under the additive Gaussian noise model. Note that the dimension of the parameter space over which a numerical optimization is performed is dramatically reduced, since the $N \times N$ object parameters have been eliminated.

Figure 2 illustrates how a two-dimensional log-likelihood function would be mapped into a one-dimensional objective function by this method. Let the X axis represent an aberration parameter and the Y axis represent an object parameter. The contour plot represents the log-likelihood function L, and the goal is to find the coordinates of the peak value. The locus of points (X, Y_M), for which Y_M maximizes L given X, traces out a curved line that corre-
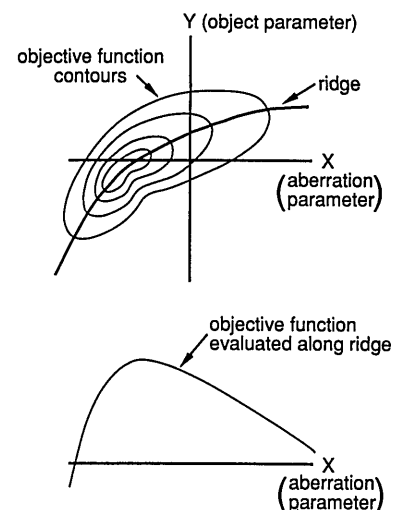

Fig. 2. Pictorial representation of the construction of the aberration-only objective function. The two-dimensional contour plot represents the log-likelihood function. A ridge of the contour plot is defined by the locus of points (X, Y_M) for which Y_M maximizes L for each value of X. The projection of the ridge values in the Y direction yields the one-dimensional, aberration-only objective function.

sponds to a ridge in the contour plot. Substitution of a closed-form expression for Y_M into the expression for L maps the objective-function values along the ridge into a function of X alone. The resulting one-dimensional function is analogous to the Gonsalves objective function. The maximum of this objective function, corresponding to the MLE of the aberration parameter, can then be independently pursued. Once this MLE for the aberration parameter is found, it can be used in the closed-form expression for Y_M to find the MLE for the object parameter.

We now proceed to generalize the Gonsalves objective function to accommodate an arbitrary number K of diversity measurements. We seek an expression for the particular F that maximizes L, given in Eq. (14). The real and imaginary parts of any such F_M will satisfy the following equations:

$$\frac{\partial}{\partial F_r(u)}L = 0,\qquad (17)$$

$$\frac{\partial}{\partial F_i(u)}L = 0,\qquad u \in \chi,\qquad (18)$$

where $F_r(u)$ and $F_i(u)$ are the real and the imaginary parts of $F(u)$, respectively. The solution for Eqs. (17) and (18) is derived in Appendix A and is given by

$$F_M(u) = \begin{cases} \sum_{k=1}^{K} D_k(u)S_k{}^*(u)\Big/\sum_{l=1}^{K}|S_l(u)|^2 & u \in \chi_1 \\ F_M{}^*(-u) & u \in \chi_0 \end{cases},\qquad (19)$$

where the set of spatial frequencies χ has been partitioned into the subset χ_0, the set of spatial frequencies at which all the OTF's are zero valued, and its complement, χ_1. Equation (19) indicates that, when $u \in \chi_0$, F_M can take on any complex value so long as the Hermitian property of F or, equivalently, the real-valued property of f is satisfied. Any of the set of functions defined by Eq. (19) is a maximum-likelihood object estimate for a fixed aberration function. When a single solution is required, additional constraints must be imposed. For example, the minimum-norm solution is found by setting $F_M = 0$ for $u \in \chi_0$. This estimate has been proposed for use in self-referenced speckle holography.[12-14] In Appendix B we show that substituting any one of the estimates defined by Eq. (19) into Eq. (14) and dropping the $1/N^2$ scale factor yields a generalized objective function,

$$L_M(\alpha) = -\sum_{u\in\chi_1}\frac{\sum_{j=1}^{K-1}\sum_{k=j+1}^{K}|D_j(u)S_k(u) - D_k(u)S_j(u)|^2}{\sum_{l=1}^{K}|S_l(u)|^2}$$

$$-\sum_{u\in\chi_0}\sum_{k=1}^{K}|D_k(u)|^2.\qquad (20)$$

Note that this expression agrees with Eq. (16) when $K = 2$. Furthermore, L_M is now well defined even if the OTF's

simultaneously go to zero at selected spatial frequencies. An alternative form for Eq. (20), also derived in Appendix B, is

$$L_M(\alpha) = \sum_{u\in\chi_1}\frac{\left|\sum_{j=1}^{K} D_j(u)S_j{}^*(u)\right|^2}{\sum_{l=1}^{K}|S_l(u)|^2} - \sum_{u\in\chi}\sum_{k=1}^{K}|D_k(u)|^2.\qquad (21)$$

This alternative expression is preferable to Eq. (20) since its evaluation requires fewer operations. The second term in Eq. (21) is just a constant, independent of the estimated aberration parameters. As a result, only the first term needs to be computed during the optimization sequence. Furthermore, the numerator of the first term contains a single summation, whereas the numerator in the first term of Eq. (20) is an iterated summation.

We may employ nonlinear optimization techniques[15] to find aberration parameters that maximize Eq. (21). Many of these techniques repeatedly compute the gradient of the objective function. An analytic expression for the partial derivative of the objective function with respect to the aberration parameters, given a mild assumption on the OTF's, is derived in Appendix C and is given by

$$\frac{\partial}{\partial\alpha_n}L_M = -\frac{4}{N^2}\sum_{u\in\chi}\phi_n(u)\text{Im}\left[\sum_{k=1}^{K} H_k(u)(Z_k * H_k{}^*)(u)\right],\qquad (22)$$

where

$$Z_k(u) \equiv \begin{cases} \left[\left[\sum_l|S_l|^2\left(\sum_j D_j S_j{}^*\right)D_k{}^* - \left|\sum_j D_j S_j{}^*\right|^2 S_k{}^*\right]\Big/\left(\sum_l|S_l|^2\right)^2\right] & u \in \chi_1 \\ 0 & u \in \chi_0 \end{cases}.\qquad (23)$$

In these equations the operator $\text{Im}[\cdot]$ takes the imaginary part of the argument, and the summations over j, k, and l run from 1 to K. Equation (22) is exact and is therefore preferable to approximations provided by the method of finite differences. Moreover, using this analytic gradient can provide a significant computational savings over such approximations. Note that the argument of the operator taking the imaginary part is the same for all aberration parameters. Therefore this argument needs to be computed only once for a full gradient computation. The computation of the gradient, through the use of Eq. (22), will be dominated by the $4K$ fast Fourier transforms (FFT's) required ($2K$ to construct the S_k and $2K$ for the convolutions), where K is the number of diversity measurements. By comparison, $2K(J + 1)$ FFT's are required to approximate the gradient by using the method of one-sided finite differences, where J is the number of aberration parameters.

We now consider the form of the gradient in two special cases. For the important case of $K = 2$, Eq. (22) reduces to

$$\frac{\partial}{\partial\alpha_n}L_M = \frac{4}{N^2}\sum_{u\in\chi}\phi_n(u)\text{Im}[H_1(u)(ZS_2 * H_1{}^*)(u)$$

$$- H_2(u)(ZS_1 * H_2{}^*)(u)],\qquad (24)$$

where

$$Z(u) \equiv \begin{cases} \dfrac{(D_1 S_1{}^* + D_2 S_2{}^*)(D_2{}^* S_1{}^* - D_1{}^* S_2{}^*)}{(|S_1|^2 + |S_2|^2)^2} & u \in \chi_1 \\ 0 & u \in \chi_0 \end{cases}$$
(25)

A second special case arises when we are trying to estimate the point-by-point phase values in the generalized pupil function. In this case the aberration basis functions will be Kronecker delta functions δ_{uu_n}, and Eq. (22) reduces to

$$\frac{\partial}{\partial \alpha_n} L_M = -\frac{4}{N^2} \text{Im}\left[\sum_{k=1}^{K} H_k(u_n)(Z_k * H_k{}^*)(u_n) \right].$$
(26)

Note that the number of FFT's required for computing the gradient in this case is the same as that when the aberration basis functions are polynomials. However, the projection onto all J basis functions is not needed here. Hence the point-by-point phase gradient is somewhat easier to compute.

Using simulation experiments, we have observed that, when Gaussian noise is added to the imagery, the objective function L_M takes on a rough texture containing many local maxima.[7] The naïve application of gradient-search methods to this objective function will result in rapid entrapment by a local maximum. One strategy for handling this problem is to use an algorithm, such as simulated annealing, that is designed to find a global maximum in the presence of local maxima. Alternatively, regularization techniques can be used to smooth the objective function so that a gradient-search algorithm would yield a regularized MLE. We have had preliminary experience with two regularization strategies, and in both cases appropriately modifying the closed-form expression for the gradient has been straightforward.

Once an aberration function that maximizes L_M is found, the corresponding OTF's can be constructed, and the MLE of the object is easily computed with Eq. (19).

4. POISSON NOISE CASE

We now consider the case in which the data are limited by photon noise. In this case the number of photoconversions that occur at each detector element will be a Poisson-distributed random variable with a mean value prescribed by the noiseless image g_k given in units of mean detected photons per pixel. Therefore the probability of detecting d_k photoevents at location x is

$$\Pr[d_k(x)] = \frac{g_k(x)^{d_k(x)} \exp[-g_k(x)]}{d_k(x)!}.$$
(27)

We assume that the number of photoevents realized will be statistically independent for each pixel. Therefore the probability of realizing an entire data set $\{d_k\}$ will be

$$\Pr(\{d_k\}) = \prod_{k=1}^{K} \prod_{x \in \chi} \frac{g_k(x)^{d_k(x)} \exp[-g_k(x)]}{d_k(x)!}.$$
(28)

A modified log-likelihood function (henceforth referred to as the log-likelihood function) for the joint estimation of

the object and aberration parameters is

$$L(f, \alpha) = \sum_{k=1}^{K} \sum_{x \in \chi} [d_k(x) \ln g_k(x) - g_k(x)],$$
(29)

where an inconsequential constant has been dropped. Consider the second term in Eq. (29):

$$\sum_{k=1}^{K} \sum_{x \in \chi} g_k(x) = \sum_{k=1}^{K} \sum_{x \in \chi} \sum_{x' \in \chi} f(x') s_k(x - x')$$
(30)

$$= \sum_{k=1}^{K} \sum_{x' \in \chi} f(x') \sum_{x \in \chi} s_k(x - x')$$
(31)

$$= \sum_{k=1}^{K} \sum_{x' \in \chi} f(x') \sum_{x \in \chi} |h_k(x - x')|^2$$
(32)

$$= \sum_{k=1}^{K} \sum_{x' \in \chi} f(x') \frac{1}{N^2}$$
$$\times \sum_{u \in \chi} |H_k(u) \exp(-i2\pi\langle u, x'\rangle/N)|^2$$
(33)

$$= \sum_{x' \in \chi} f(x') \frac{1}{N^2} \sum_{k=1}^{K} \sum_{u \in \chi} |H_k(u)|^2,$$
(34)

where we have used discrete versions of the Fourier shift theorem and Parseval's theorem.[10] Note that the double sum over the squared moduli of the pupil functions is just a constant, independent of the object, the aberration parameters, and the phase diversity. Let

$$C \equiv \frac{1}{N^2} \sum_{k=1}^{K} \sum_{u \in \chi} |H_k(u)|^2.$$
(35)

Equation (34) then becomes

$$\sum_{k=1}^{K} \sum_{x \in \chi} g_k(x) = C \sum_{x \in \chi} f(x),$$
(36)

and the log-likelihood function simplifies to

$$L(f, \alpha) = \sum_{k=1}^{K} \sum_{x \in \chi} d_k(x) \ln g_k(x) - C \sum_{x \in \chi} f(x)$$
(37)

$$= \sum_{k=1}^{K} \sum_{x \in \chi} d_k(x) \ln\left[\sum_{x' \in \chi} f(x') s_k(x - x') \right] - C \sum_{x \in \chi} f(x).$$
(38)

At this point we could logically follow the same strategy that we pursued in the Gaussian noise case in an effort to reduce the dimension of the parameter space over which a numerical optimization is performed. The desired sequence is to solve for the MLE of the object, given fixed aberration parameters, and substitute this result into the expression for the log-likelihood function to create an objective function that does not depend explicitly on the object estimate. We begin by computing the partial derivative of the log-likelihood function with respect to the ith pixel value of the object estimate:

$$\frac{\partial}{\partial f(x_i)} L = \sum_{k=1}^{K} \sum_{x \in \chi} d_k(x) \frac{\partial}{\partial f(x_i)} \ln\left[\sum_{x' \in \chi} f(x') s_k(x - x') \right]$$
$$- C \sum_{x \in \chi} \frac{\partial}{\partial f(x_i)} f(x)$$
(39)

$$= \sum_{k=1}^{K} \sum_{x \in \chi} \frac{d_k(x) s_k(x - x_i)}{\sum\limits_{x' \in \chi} f(x') s_k(x - x')} - C.$$
(40)

To find the MLE of the object for fixed aberration parameters we set this partial derivative equal to zero and attempt to solve for f:

$$0 = \sum_{k=1}^{K} \sum_{x \in \chi} \frac{d_k(x) s_k(x - x_i)}{\sum_{x' \in \chi} f(x') s_k(x - x')} - C. \qquad (41)$$

Solving Eq. (41) would provide the MLE of the object for the multiple-kernel deconvolution problem given Poisson noise. The same problem is encountered in time-of-flight positron emission topography.[16] Unfortunately, a closed-form solution for f has not been found, and hope for the derivation of an aberration-only objective function is diminished. Note, however, that the partial derivative with respect to each object pixel [Eq. (40)] can be found for all object pixels with a reasonable number of FFT's ($3K + 1$ FFT's, assuming that the OTF's are precomputed). This computation would be unmanageable if the method of finite differences were used.

Gradient-Search Algorithms

The tractability of the computation implied by Eq. (40) suggests the use of a gradient-search technique in maximizing the log-likelihood function over object pixels and aberration parameters simultaneously. Recall that the complete gradient also requires partial derivatives of the log-likelihood function with respect to the aberration parameters. The expression for these partial derivatives is derived in Appendix D and is given here as

$$\frac{\partial}{\partial \alpha_n} L = -2 \sum_{u \in \chi} \phi_n(u) \mathrm{Im} \left[\sum_{k=1}^{K} H_k(u) \frac{1}{N^2} \sum_{x' \in \chi} h_k^*(x') \right.$$
$$\left. \times \exp(i 2\pi \langle u, x' \rangle / N) \sum_{x \in \chi} \frac{d_k(x) f(x - x')}{\sum_{x'' \in \chi} f(x'') s_k(x - x'')} \right]. \qquad (42)$$

Careful scrutiny of Eq. (42) shows that the number of FFT's ($6K + 1$) required for computing the entire set of partial derivatives $\{\partial L / \partial \alpha_n\}$ is reasonable and does not depend on the number of parameters J. Because there is significant overlap in the computation of $\partial L / \partial f(x_i)$ and $\partial L / \partial \alpha_n$, the total number of FFT's required for computation of the gradient of the log-likelihood function is only $7K + 1$.

Having derived expressions for the partial derivatives of the log-likelihood function with respect to object pixels and aberration parameters, we are now in a position to suggest a strategy for the joint estimation of object and aberration parameters. These partial derivatives constitute a gradient of the log-likelihood function in a parameter space of dimension $N \times N \times J$. The computation of this analytic gradient requires multiple FFT's but is manageable, unlike a finite-difference computation. Any of a variety of unconstrained and constrained nonlinear optimization algorithms that utilize the gradient can be employed to search iteratively for a global maximum.[15] Constrained optimization methods allow us to enforce the nonnegativity property of incoherently illuminated objects in our estimate. Perhaps a good initial estimate would be the solution under the Gaussian noise model. If it can be found, the global maximum will provide the MLE for both the object pixel values and the aberration parameters.

The log-likelihood function can be modified to accommodate faulty detector elements:

$$L_w(f, \phi) = \sum_{k=1}^{K} \sum_{x \in \chi} [d_k(x) \ln g_k(x) - g_k(x)] w_k(x), \qquad (43)$$

where w_k is a binary window function with a value of unity for the detector elements that are functioning and zero for those that are faulty, in the kth detector array. This consideration would be important, for example, when phase-diversity methods are applied to the Hubble phase-estimation problem, for which selected detector elements are known to have failed. The partial derivative of this log-likelihood function with respect to the ith pixel value is readily shown to be

$$\frac{\partial}{\partial f(x_i)} L_w = \sum_{k=1}^{K} \sum_{x \in \chi} \left[\frac{d_k(x)}{g_k(x)} - 1 \right] w_k(x) s_k(x - x_i). \qquad (44)$$

Similarly, the partial derivative with respect to the nth aberration parameter is found to be

$$\frac{\partial}{\partial \alpha_n} L_w = -2 \sum_{u \in \chi} \phi_n(u) \mathrm{Im} \left\{ \sum_{k=1}^{K} H_k(u) \frac{1}{N^2} \sum_{x' \in \chi} h_k^*(x') \right.$$
$$\left. \times \exp(i 2\pi \langle u, x' \rangle / N) \sum_{x \in \chi} \left[\frac{d_k(x)}{g_k(x)} - 1 \right] w_k(x) f(x - x') \right\}. \qquad (45)$$

It is also worth considering the special case in which the object is known *a priori* to be unresolved or pointlike. This case has been important to researchers seeking to estimate the aberrations present in the Hubble Space Telescope from Hubble images of stars.[3,4] When the point-object assumption is valid, the object is known, and there remain only aberration parameters to estimate. It is straightforward to show that the log-likelihood function is insensitive to shifts in object position. Therefore there is no loss of generality in modeling the point object as a weighted Kronecker delta located at the origin:

$$f(x) = a \delta(x), \qquad (46)$$

where

$$\delta(x) \equiv \begin{cases} 1 & x = (0, 0) \\ 0 & \text{otherwise} \end{cases}. \qquad (47)$$

Given this model, Eq. (42) quickly simplifies to

$$\frac{\partial}{\partial \alpha_n} L = -2 \sum_{u \in \chi} \phi_n(u) \mathrm{Im} \left[\sum_{k=1}^{K} H_k(u) \frac{1}{N^2} \sum_{x' \in \chi} \frac{h_k^*(x') d_k(x')}{s_k(x')} \right.$$
$$\left. \times \exp(i 2\pi \langle u, x' \rangle / N) \right]. \qquad (48)$$

If we define

$$\mathscr{F}_u^{-1} \left[\frac{h_k^*(x) d_k(x)}{s_k(x)} \right] \equiv \frac{1}{N^2} \sum_{x \in \chi} \frac{h_k^*(x) d_k(x)}{s_k(x)} \exp(i 2\pi \langle u, x \rangle / N), \qquad (49)$$

then Eq. (48) assumes the succinct form

$$\frac{\partial}{\partial \alpha_n} L = -2 \sum_{u \in \chi} \phi_n(u) \mathrm{Im} \left\{ \sum_{k=1}^{K} H_k(u) \mathscr{F}_u^{-1} \left[\frac{h_k^*(x) d_k(x)}{s_k(x)} \right] \right\}. \qquad (50)$$

This expression gives the partial derivative of the log-likelihood function with respect to aberration parameters for a point object given multiple diversity images corrupted by Poisson noise. A gradient evaluation requires only $2K$ FFT's, again suggesting the use of a gradient-search algorithm. Care must be used when invoking the point-object assumption. For example, if this assumption were unwittingly applied to images of a binary star, the resulting aberration estimates could be significantly biased.

Expectation-Maximization Algorithm

In this section we construct an EM algorithm as an alternative means to find the MLE of the object and aberrations in the case of Poisson noise. The use of an EM algorithm is suggested by similarities between our problem and time-of-flight positron emission tomography, in which EM algorithms have been applied with considerable success.[16,17] Kaufman has shown that there is a close relationship between the EM algorithm used in positron emission tomography and gradient-search methods.[18] One potential advantage of the proposed EM algorithm over a constrained gradient-search algorithm is that the nonnegativity constraint is enforced in a natural fashion. The EM algorithm is iterative, producing a sequence of estimates of the object and aberration parameters with the property that the log likelihood of these estimates is a nondecreasing function of the iteration number. General descriptions of the EM algorithm and its convergence properties are found in Refs. 19 and 20.

The derivation of our EM algorithm begins with the construction of an abstract statistical model. Recall that the measured data $d_k(x)$ indexed by k and x consist of independent Poisson random variables, each with the expected value

$$g_k(x) = \sum_{x' \in \chi} f(x')s_k(x - x'). \tag{51}$$

Now consider a collection of random variables $\tilde{d}_k(x \mid x')$ indexed by k, x, and x', and let them be independent Poisson-distributed random variables, each with the expected value

$$\tilde{g}_k(x \mid x') = f(x')s_k(x - x'). \tag{52}$$

This expected value can be interpreted as the mean number of photons detected at pixel x, originating from the object at pixel x'. Any sum of the random variables $\tilde{d}_k(x \mid x')$ will produce a new random variable that is Poisson distributed with an expected value that is obtained by appropriately summing the $\tilde{g}_k(x \mid x')$, as can be readily verified by using characteristic functions. For example, the random variables indexed by k and x,

$$\sum_{x' \in \chi} \tilde{d}_k(x \mid x'), \tag{53}$$

will be independent and Poisson distributed, with expected values equal to

$$\sum_{x' \in \chi} \tilde{g}_k(x \mid x') = \sum_{x' \in \chi} f(x')s_k(x - x') \tag{54}$$

$$= g_k(x). \tag{55}$$

Therefore a model that is statistically consistent with our measured data results from treating the collected data as

$$d_k(x) = \sum_{x' \in \chi} \tilde{d}_k(x \mid x'). \tag{56}$$

Using terminology associated with the EM algorithm, we refer to the random variables $\tilde{d}_k(x \mid x')$ as the complete data and the actual collected data $d_k(x)$ as the incomplete data. The choice of the abstract data set $\tilde{d}_k(x \mid x')$ as the complete data is somewhat arbitrary, and other choices could lead to alternative algorithms.

The EM algorithm is a prescription for updating the set of parameter estimates $\{f^{(r)}, \alpha^{(r)}\}$ to the new set of parameter estimates $\{f^{(r+1)}, \alpha^{(r+1)}\}$ having the property that

$$L[f^{(r+1)}, \alpha^{(r+1)}] \geq L[f^{(r)}, \alpha^{(r)}], \tag{57}$$

where the superscripts in parentheses represent the iteration numbers. The prescription requires an expression for the (modified) log-likelihood function associated with the complete data L_{cd}. For our problem this expression is found [by analogy with Eq. (29)] to be

$$L_{cd}(f, \alpha) = \sum_{k=1}^{K} \sum_{x \in \chi} \sum_{x' \in \chi} \{\tilde{d}_k(x \mid x')$$
$$\times \ln[f(x')s_k(x - x')] - f(x')s_k(x - x')\} \tag{58}$$

$$= \sum_{x' \in \chi} \left[\sum_{k=1}^{K} \sum_{x \in \chi} \tilde{d}_k(x \mid x') \right] \ln f(x')$$
$$+ \sum_{k=1}^{K} \sum_{x \in \chi} \sum_{x' \in \chi} \tilde{d}_k(x \mid x') \ln s_k(x - x')$$
$$- \sum_{x' \in \chi} f(x') \sum_{k=1}^{K} \sum_{x \in \chi} s_k(x - x') \tag{59}$$

$$= \sum_{x' \in \chi} \left[\sum_{k=1}^{K} \sum_{x \in \chi} \tilde{d}_k(x \mid x') \right] \ln f(x')$$
$$+ \sum_{k=1}^{K} \sum_{x' \in \chi} \left[\sum_{x \in \chi} \tilde{d}_k(x \mid x - x') \right] \ln s_k(x')$$
$$- \sum_{x' \in \chi} f(x') \sum_{k=1}^{K} \sum_{x \in \chi} s_k(x), \tag{60}$$

where inconsequential terms have been dropped. The expectation step (or E step) consists of evaluating the conditional expectation of L_{cd}, given the measured data d_k and assuming that the parameters governing the complete data are equal to $\{f^{(r)}, \alpha^{(r)}\}$. This conditional expectation is defined as

$$Q[f, \alpha \mid f^{(r)}, \alpha^{(r)}] \equiv E^{(r)}[L_{cd}(f, \alpha) \mid \{d_k\}], \tag{61}$$

where $E^{(r)}[\cdot]$ denotes an expectation in which the parameters associated with the complete data take on the values $\{f^{(r)}, \alpha^{(r)}\}$. The maximization step (or M step) of the EM algorithm consists of assigning to $\{f^{(r+1)}, \alpha^{(r+1)}\}$ the values of $\{f, \alpha\}$ that maximize Q subject to any constraints that may exist. As derived in Appendix E, the resulting iterations are

$$f^{(r+1)}(x') = f^{(r)}(x') \frac{1}{\frac{1}{K} \sum_{k=1}^{K} S_k(0)} \sum_{k=1}^{K} \sum_{x \in \chi} s_k^{(r)}(x - x') \frac{d_k(x)}{g_k^{(r)}(x)}, \tag{62}$$

$$\alpha^{(r+1)} = \arg\max_{\alpha} \left\{ \sum_{k=1}^{K} \sum_{x' \in \chi} \right.$$
$$\left. \times \left[\sum_{x \in \chi} \frac{f^{(r)}(x - x')s_k^{(r)}(x')}{g_k^{(r)}(x)} d_k(x) \right] \ln s_k(x') \right\}, \tag{63}$$

where

$$g_k^{(r)}(x) \equiv \sum_{x' \in \chi} f_k^{(r)}(x) s_k^{(r)}(x - x'), \qquad (64)$$

$$s_k^{(r)} \equiv k\text{th diversity PSF evaluated for } \alpha^{(r)}, \qquad (65)$$

$$S_k(0) = \sum_{x \in \chi} s_k(x). \qquad (66)$$

Note that $S_k(0)$ is independent of the particular aberration parameters. In the iterations zero divided by zero is defined to be zero. Since the likelihood is a nondecreasing function of r, when it is properly initialized the algorithm will never produce $g_k^{(r)}(x) = 0$ when $d_k(x) \neq 0$.

Whereas Eq. (62) presents a closed-form expression for updating the object parameter estimates, Eq. (63) requires the solution to an optimization problem in order to update the aberration-parameter estimates. Fortunately, this optimization occurs in a space with a relatively small dimension, the space of aberration parameters. Moreover, a closed-form expression for the gradient of the argument to be maximized in Eq. (63) is available for use in the optimization sequence. This gradient expression is analogous to that derived for the point-object case, as is discussed in Appendix E.

We conclude this discussion of the EM algorithm by considering its application in two special cases. One special case arises when the phase-aberration parameters are known *a priori*. In this case the EM algorithm for estimating the object intensity is

$$f^{(r+1)}(x') = f^{(r)}(x') \frac{1}{\frac{1}{K}\sum_{k=1}^{K} S_k(0)} \sum_{k=1}^{K} \sum_{x \in \chi} s_k(x - x') \frac{d_k(x)}{g_k^{(r)}(x)}, \quad (67)$$

which is essentially the same as the EM algorithm derived by Snyder and Politte[16] and Politte[17] for time-of-flight positron emission tomography. For this case, arguments similar to those used by Vardi *et al.*[21] can be used to show that $f^{(r)}$ converges to a global maximum of the log likelihood. Under certain assumptions about the PSF's, $\{s_k\}$, this solution can be shown to be unique. With the additional restriction that $K = 1$, the iterative rule in Eq. (67) is the same as that independently derived by Richardson[22] and Lucy[23] using a Bayesian viewpoint. The second special case arises when the object is known to be a point source. In this case the EM algorithm for estimating the unknown aberration parameters will converge in one iteration to the maximum-likelihood solution.

5. SUMMARY

We have formally stated the problem of the joint estimation of the object and aberrations by using phase diversity. Maximum-likelihood estimation has been considered under Gaussian and Poisson noise models. In the Gaussian noise case the estimation of the aberration parameters can be accomplished without explicitly estimating the object pixel values. An aberration-only objective function, first derived by Gonsalves, was shown to yield a maximum-likelihood estimate, and this objective function was generalized to accommodate an arbitrary number of diversity measurements. Moreover, an analytic expression for the gradient of this aberration-only objective function that will allow for an efficient gradient-search phase-estimation algorithm was derived.

An expression for the log-likelihood function has been presented for the Poisson noise case as well. An attempt to derive an aberration-only objective function by analogy with the Gaussian noise case failed. However, an analytic expression for the gradient of the log-likelihood function was derived and shown to be computationally tractable, suggesting the use of a gradient-search algorithm. An EM algorithm that incorporates a nonnegativity constraint in a natural fashion was also presented.

The viability of the phase-diversity concept was previously demonstrated through computer simulation for the case of Gaussian noise with $K = 2$.[6–8] In the case of Poisson noise, the gradient-search and EM algorithms described herein need to be exercised to demonstrate their utility. In every case additional simulations and theory are needed to quantify the performance of the estimates as a function of the number of diversity images, the types of diversity, and the level of noise. A particularly intriguing issue is the optimum distribution of a fixed number of photons among diversity images.

APPENDIX A: MAXIMUM-LIKELIHOOD OBJECT ESTIMATE FOR THE GAUSSIAN CASE

In this appendix we derive the expression for the MLE of the object under the additive Gaussian noise model and for fixed aberration parameters. Recall that the log-likelihood function is given by

$$L(f, \alpha) = -\frac{1}{N^2} \sum_{k=1}^{K} \sum_{u' \in \chi} |D_k(u') - F(u')S_k(u')|^2. \qquad (A1)$$

We seek an expression for $F(u)$ that maximizes L. Stationary points of L are achieved when

$$\frac{\partial}{\partial F_r(u)} L = 0, \qquad (A2)$$

$$\frac{\partial}{\partial F_i(u)} L = 0, \qquad u \in \chi, \qquad (A3)$$

where $F_r(u)$ and $F_i(u)$ are the real and the imaginary parts of $F(u)$, respectively. Consider the partial derivative of L with respect to $F_r(u)$:

$$\frac{\partial}{\partial F_r} L = -\frac{1}{N^2} \sum_{k=1}^{K} \frac{\partial}{\partial F_r} (|D_k|^2 + |FS_k|^2$$
$$- D_k F^* S_k^* - D_k^* F S_k) \qquad (A4)$$

$$= -\frac{1}{N^2} \sum_{k=1}^{K} \left[|S_k|^2 \frac{\partial}{\partial F_r}(F_r^2 + F_i^2) - D_k S_k^* - D_k^* S_k \right] \qquad (A5)$$

$$= -\frac{1}{N^2} \sum_{k=1}^{K} (2|S_k|^2 F_r - D_k S_k^* - D_k^* S_k), \qquad (A6)$$

where we have suppressed the dependence on u for brevity. Similarly,

$$\frac{\partial}{\partial F_i} L = -\frac{1}{N^2} \sum_{k=1}^{K} (2|S_k|^2 F_i + i D_k S_k^* - i D_k^* S_k). \qquad (A7)$$

We now set the right-hand side of Eq. (A6) equal to zero and solve for F_r:

$$2F_r \sum_{k=1}^{K} |S_k|^2 = \sum_{k=1}^{K} (D_k S_k{}^* + D_k{}^* S_k), \qquad (A8)$$

$$F_r(u) = \begin{cases} \sum_{k=1}^{K} (D_k S_k{}^* + D_k{}^* S_k) \Big/ 2 \sum_{l=1}^{K} |S_l|^2 & u \in \chi_1 \\ F_r(-u) & u \in \chi_0 \end{cases}, \qquad (A9)$$

where we have partitioned the set of spatial frequencies χ into the subset χ_0, the set of spatial frequencies at which all the OTF's are zero valued, and its complement, χ_1. Formally,

$$\chi_0 = \{u : S_k(u) = 0, \ k = 1, \ldots, K\}, \qquad (A10)$$

$$\chi_1 = \{u : u \in \chi, \ u \notin \chi_0\}. \qquad (A11)$$

The set χ_0 includes all spatial frequencies that fall outside the union of the supports of all K diffraction-limited OTF's, that is, all the spatial frequencies beyond the system diffraction-limited cutoff. Note that it also includes any spatial frequencies interior to this union of supports for which the OFT's simultaneously go to zero owing to aberrations. Equation (A9) indicates that, when $u \in \chi_0$, $F_r(u)$ can take on any complex value so long as $F(u) = F(-u)$, which is consistent with the Hermitian property of F or, equivalently, the real-valued property of f. Setting the right-hand side of Eq. (A7) equal to zero, we get

$$F_i(u) = \begin{cases} \sum_{k=1}^{K} (-D_k S_k{}^* + D_k{}^* S_k)i \Big/ 2 \sum_{l=1}^{K} |S_l|^2 & u \in \chi_1 \\ -F_i(-u) & u \in \chi_0 \end{cases}. \qquad (A12)$$

The set of stationary points of L is found by combining Eqs. (A9) and (A12) as follows:

By examining the second-order partial derivatives of L with respect to the real and imaginary parts of F at each spatial frequency, one can show that L is a concave function.[15] As a consequence, Eq. (A15) represents the set of objects that globally maximize L and therefore defines the set of maximum-likelihood object estimates.

APPENDIX B: ABERRATION-ONLY OBJECTIVE FUNCTION FOR THE GAUSSIAN CASE

Substitution of any of the maximum-likelihood object estimates into the log-likelihood function and eliminating the $1/N^2$ scale factor yields a new objective function $L_M(\alpha)$ for the estimation of the aberrations. The derivation of this new objective function is presented here:

$$L_M(\alpha) = -\sum_{u \in \chi} \sum_{k=1}^{K} |D_k(u) - F_M(u) S_k(u)|^2 \qquad (B1)$$

$$= -\sum_{u \in \chi_1} \sum_{k} \left| D_k - S_k \frac{\sum_j D_j S_j{}^*}{\sum_l |S_l|^2} \right|^2 - \sum_{u \in \chi_0} \sum_{k} |D_k|^2, \qquad (B2)$$

where the u dependence has been suppressed and the summations over j, k, and l run from 1 to K. Equation (B2) consists of two major terms distinguished by the summations over χ_1 and χ_0. Recall that χ_0 is the set of spatial frequencies at which all the OTF's are zero valued and χ_1 is the complement of χ_0. Note that the values of F_M at spatial frequencies in χ_0 are irrelevant since $S_k(u)$ is zero there. Also note that the contribution to the second term in Eq. (B2) from data at spatial frequencies outside the union of diffraction-limited OTF supports will be constant and will not affect the contour of the objective function. Strictly speaking, this term must be retained, however, in the event that the OTF's simultaneously go to zero at spatial frequencies within the union of diffraction-limited OTF supports. We now focus our attention on the first term in Eq. (B2), which we designate by L_{M1}. Expanding this term, we have that

$$F_M(u) = F_r(u) + iF_i(u) \qquad (A13)$$

$$= \begin{cases} \sum_{k=1}^{K} [D_k S_k{}^* + D_k{}^* S_k + i(-D_k S_k{}^* i + D_k{}^* S_k i)] \Big/ 2 \sum_{l=1}^{K} |S_l|^2 & u \in \chi_1 \\ F_M{}^*(-u) & u \in \chi_0 \end{cases}, \qquad (A14)$$

$$= \begin{cases} \sum_{k=1}^{K} D_k S_k{}^* \Big/ \sum_{l=1}^{K} |S_l|^2 & u \in \chi_1 \\ F_M{}^*(-u) & u \in \chi_0 \end{cases} \qquad (A15)$$

$$L_{M1} = -\sum_{u \in \chi_1} \sum_{k} \left| \frac{D_k \sum_l |S_l|^2 - S_k \sum_j D_j S_j{}^*}{\sum_l |S_l|^2} \right|^2 \qquad (B3)$$

$$= -\sum_{u \in \chi_1} \sum_{k} \frac{\left| D_k \sum_l |S_l|^2 \right|^2 + \left| S_k \sum_j D_j S_j{}^* \right|^2 - D_k{}^* \sum_l |S_l|^2 S_k \sum_j D_j S_j{}^* - D_k \sum_l |S_l|^2 S_k{}^* \sum_j D_j{}^* S_j}{\left(\sum_l |S_l|^2 \right)^2}. \qquad (B4)$$

Note that the summation over k affects only the numerator in Eq. (B4). We now distribute the k summation over each term in the numerator to get

$$
\text{numerator} = \sum_k |D_k|^2 \left(\sum_l |S_l|^2 \right)^2 + \sum_k |S_k|^2 \left| \sum_j D_j S_j^* \right|^2
$$

$$
- \sum_k D_k^* S_k \sum_l |S_l|^2 \sum_j D_j S_j^*
$$

$$
- \sum_k D_k S_k^* \sum_l |S_l|^2 \sum_j D_j^* S_j \tag{B5}
$$

$$
= \sum_k |D_k|^2 \left(\sum_l |S_l|^2 \right)^2 - \sum_l |S_l|^2 \left| \sum_j D_j S_j^* \right|^2, \tag{B6}
$$

where the second and third terms in Eq. (B5) have been canceled. Factoring out the common factor $\sum_l |S_l|^2$ in the numerator and canceling this with the same factor in the denominator, we get

$$
L_{M1} = -\sum_{u \in \chi_1} \frac{\sum_k |D_k|^2 \sum_j |S_j|^2 - \left| \sum_j D_j S_j^* \right|^2}{\sum_l |S_l|^2} \tag{B7}
$$

$$
= -\sum_{u \in \chi_1} \frac{\sum_k \sum_j |D_k S_j|^2 - \sum_j \sum_k D_j S_j^* D_k^* S_k}{\sum_l |S_l|^2} \tag{B8}
$$

$$
= -\sum_{u \in \chi_1} \frac{(1/2) \sum_k \sum_j |D_k S_j - D_j S_k|^2}{\sum_l |S_l|^2}, \tag{B9}
$$

where we have used the fact that k and j are dummy indices. Note that the terms in the numerator for which $j = k$ vanish. Rearranging the summation limits to retain half the remaining terms in the numerator, we have that

$$
L_{M1} = -\sum_{u \in \chi_1} \frac{\sum_{j=1}^{K-1} \sum_{k=j+1}^{K} |D_j(u) S_k(u) - D_k(u) S_j(u)|^2}{\sum_{l=1}^{K} |S_l(u)|^2}. \tag{B10}
$$

Reintroducing the second term in the objective function gives

$$
L_M(\alpha) = -\sum_{u \in \chi_1} \frac{\sum_{j=1}^{K-1} \sum_{k=j+1}^{K} |D_j(u) S_k(u) - D_k(u) S_j(u)|^2}{\sum_{l=1}^{K} |S_l(u)|^2}
$$

$$
- \sum_{u \in \chi_0} \sum_{k=1}^{K} |D_k(u)|^2. \tag{B11}
$$

This expression retains the form of the Gonsalves objective function ($K = 2$) for direct comparison. An alternative form derives directly from Eq. (B7) and is given by

$$
L_M(\alpha) = -\sum_{u \in \chi_1} \left[\sum_{k=1}^{K} |D_k(u)|^2 - \frac{\left| \sum_{j=1}^{K} D_j(u) S_j^*(u) \right|^2}{\sum_{l=1}^{K} |S_l(u)|^2} \right]
$$

$$
- \sum_{u \in \chi_0} \sum_{k=1}^{K} |D_k(u)|^2 \tag{B12}
$$

$$
= \sum_{u \in \chi_1} \frac{\left| \sum_{j=1}^{K} D_j(u) S_j^*(u) \right|^2}{\sum_{l=1}^{K} |S_l(u)|^2} - \sum_{u \in \chi} \sum_{k=1}^{K} |D_k(u)|^2. \tag{B13}
$$

APPENDIX C: PARTIAL DERIVATIVE OF THE ABERRATION-ONLY OBJECTIVE FUNCTION FOR THE GAUSSIAN CASE

In this appendix we derive an expression for the partial derivative of the objective function L_M with respect to an aberration parameter α_n for the case of additive Gaussian noise. Using the form of the objective function expressed in Eq. (21), we have that

$$
\frac{\partial}{\partial \alpha_n} L_M = \frac{\partial}{\partial \alpha_n} \sum_{u \in \chi_1} \frac{\left| \sum_{j=1}^{K} D_j(u) S_j^*(u) \right|^2}{\sum_{l=1}^{K} |S_l(u)|^2} - \frac{\partial}{\partial \alpha_n} \sum_{u \in \chi} \sum_{k=1}^{K} |D_k(u)|^2. \tag{C1}
$$

The second term in Eq. (C1) vanishes because the data are independent of the parameter estimates. The partial derivative in the first term is difficult to evaluate since the set χ_1 that defines the limits of the summation can actually depend on the aberration parameters. Indeed, the derivative may not exist at coordinates in the space of the aberration parameters for which the limits of summation are changing. To circumvent this problem we place a mild restriction on the set of OTF's. Consider the set of pixels χ_s defined by the union of the diffraction-limited OTF supports. Spatial frequencies in the complement of χ_s are permanent elements of χ_0 and never enter into the limits of summation in the first term in Eq. (C1). It is only when the OTF's simultaneously go to zero at spatial frequencies within χ_s that the limits of summation can vary as the aberrations change. Recall that the value of an OTF at a spatial frequency $u \in \chi_s$ is the sum of unit-length phasors. Such a sum is unlikely to be exactly zero, particularly with random phases.[24] The prospect of all K OTF's summing to zero at the same spatial frequency is even more remote. Therefore the limits of summation in Eq. (C1) will be constant almost everywhere in the space of the aberration parameters. We assume that the partial derivative is evaluated at the aberration parameters about which the OTF's do not simultaneously go to zero at spatial frequencies within χ_s. When this is true, the partial derivative can be taken inside the summation as

$$\frac{\partial}{\partial \alpha_n} L_M = \sum_{u \in \chi_1} \frac{\partial}{\partial \alpha_n} \frac{\left| \sum_{j=1}^{K} D_j(u) S_j^*(u) \right|^2}{\sum_{l=1}^{K} |S_l(u)|^2} \tag{C2}$$

$$= \sum_{u \in \chi_1} \frac{\sum_l |S_l|^2 \left(\sum_j D_j S_j^* \sum_k D_k^* S_k' + \text{c.c.} \right) - \left| \sum_j D_j S_j^* \right|^2 \left(\sum_l S_l^* S_l' + \text{c.c.} \right)}{\left(\sum_l |S_l|^2 \right)^2} \tag{C3}$$

$$= \sum_{u \in \chi_1} \frac{\sum_l |S_l|^2 \sum_j D_j S_j^* \sum_k D_k^* S_k' - \left| \sum_j D_j S_j^* \right|^2 \sum_l S_l^* S_l'}{\left(\sum_l |S_l|^2 \right)^2} + \text{c.c.} \tag{C4}$$

$$= \sum_{u \in \chi} \sum_k Z_k S_k' + \text{c.c.}, \tag{C5}$$

where the u dependence has been suppressed; the summations over j, k, and l run from 1 to K; the prime applied to a function signifies a partial derivative; c.c. represents a term that is the complex conjugate of the preceding term; and we have defined

$$Z_k(u) \equiv \begin{cases} \left[\sum_l |S_l|^2 \left(\sum_j D_j S_j^* \right) D_k^* - \left| \sum_j D_j S_j^* \right|^2 S_k^* \right] \Big/ \left(\sum_l |S_l|^2 \right)^2 & u \in \chi_1 \\ 0 & u \in \chi_0 \end{cases} \tag{C6}$$

Note that defining Z_k in this way extends the summation in Eq. (C5) over all of χ.

At this point we digress to derive an expression for the partial derivative of an OTF with respect to an aberration parameter. Recall that the unnormalized OTF for the kth diversity system is the discrete Fourier transform of the corresponding PSF:

$$S_k(u) = \sum_{x \in \chi} s_k(x) \exp(-i2\pi \langle u, x \rangle / N) \tag{C7}$$

$$= \sum_{x \in \chi} |h_k(x)|^2 \exp(-i2\pi \langle u, x \rangle / N). \tag{C8}$$

Equation (C8) can be rewritten in terms of the generalized pupil function H_k, using the discrete version of the autocorrelation theorem:

$$S_k(u) = \frac{1}{N^2} \sum_{u' \in \chi} H_k(u') H_k^*(u' - u), \tag{C9}$$

where

$$H_k(u) \equiv |H_k(u)| \exp \left\{ i \left[\theta_k + \sum_{j=1}^{J} \alpha_j \phi_j(u) \right] \right\}. \tag{C10}$$

The partial derivative of the OTF is

$$\frac{\partial}{\partial \alpha_n} S_k(u) = \frac{1}{N^2} \sum_{u' \in \chi} \left[H_k^*(u' - u) \frac{\partial}{\partial \alpha_n} H_k(u') + H_k(u') \frac{\partial}{\partial \alpha_n} H_k^*(u' - u) \right] \tag{C11}$$

$$= (1/N^2) \sum_{u' \in \chi} [H_k^*(u' - u) i\phi_n(u') H_k(u') - H_k(u') i\phi_n(u' - u) H_k^*(u' - u)] \tag{C12}$$

$$= \frac{1}{N^2} \left[\sum_{u' \in \chi} i\phi_n(u') H_k(u') H_k^*(u' - u) - \sum_{u'' \in \chi} i\phi_n(u'') H_k(u'' + u) H_k^*(u'') \right] \tag{C13}$$

$$= (i/N^2) \sum_{u' \in \chi} \phi_n(u') [H_k(u') H_k^*(u' - u) - H_k^*(u') H_k(u' + u)], \tag{C14}$$

where we have used the change of variables $u'' = u' - u$ in Eq. (C13).

Returning to the primary derivation, we substitute Eq. (C14) into Eq. (C5):

$$\frac{\partial}{\partial \alpha_n} L_M = \sum_{u \in \chi} \sum_{k=1}^{K} \left\{ Z_k(u) \frac{i}{N^2} \sum_{u' \in \chi} \phi_n(u') \right.$$
$$\left. \times [H_k(u') H_k^*(u' - u) - H_k^*(u') H_k(u' + u)] \right\} + \text{c.c.} \tag{C15}$$

$$= \left[\frac{i}{N^2} \sum_{u' \in \chi} \phi_n(u') \sum_{k=1}^{K} H_k(u') \sum_{u \in \chi} Z_k(u) H_k^*(u' - u) \right] + \text{c.c.}$$
$$- \left[\frac{i}{N^2} \sum_{u' \in \chi} \phi_n(u') \sum_{k=1}^{K} H_k^*(u') \sum_{u \in \chi} Z_k(u) H_k(u' + u) \right] + \text{c.c.} \tag{C16}$$

Using the Hermitian property of Z_k and the change of variables $u'' = -u'$, we rewrite the second term above as

$$\text{second term} = -\left[\frac{i}{N^2} \sum_{u' \in \chi} \phi_n(u') \sum_{k=1}^{K} H_k^*(u') \right.$$
$$\left. \times \sum_{u \in \chi} Z_k^*(-u) H_k(u' + u) \right] + \text{c.c.} \tag{C17}$$

$$= -\left[\frac{i}{N^2} \sum_{u' \in \chi} \phi_n(u') \sum_{k=1}^{K} H_k^*(u') \right.$$
$$\left. \times \sum_{u'' \in \chi} Z_k^*(u'') H_k(u' - u'') \right] + \text{c.c.}, \tag{C18}$$

from which it is apparent that the second term is just the complex conjugate of the first term.   Therefore

$$\frac{\partial}{\partial \alpha_n} L_M = \left[\frac{2i}{N^2} \sum_{u' \in \chi} \phi_n(u') \sum_{k=1}^{K} H_k(u') \right.$$

$$\left. \times \sum_{u \in \chi} Z_k(u) H_k^*(u' - u) \right] + \text{c.c.} \qquad (C19)$$

$$= -\frac{4}{N^2} \sum_{u' \in \chi} \phi_n(u') \text{Im}\left[\sum_{k=1}^{K} H_k(u')(Z_k * H_k^*)(u') \right], \qquad (C20)$$

where the operator Im[·] takes the imaginary part of the argument.

APPENDIX D: PARTIAL DERIVATIVE WITH RESPECT TO ABERRATION PARAMETERS FOR THE POISSON CASE

In this appendix we derive the expression for the partial derivative of the log-likelihood function with respect to

Eq. (D2):

$$\frac{\partial}{\partial \alpha_n} s_k(x) = h_k^*(x) \frac{\partial h_k(x)}{\partial \alpha_n} + \text{c.c.} \qquad (D3)$$

$$= \frac{h_k^*(x)}{N^2} \frac{\partial}{\partial \alpha_n} \sum_{u \in \chi} H_k(u) \exp(i2\pi\langle u, x\rangle/N) + \text{c.c.} \qquad (D4)$$

$$= \frac{h_k^*(x)}{N^2} \frac{\partial}{\partial \alpha_n} \sum_{u \in \chi} |H_k(u)|$$

$$\times \exp\left\{ i\left[\theta_k(u) + \sum_{j=1}^{J} \alpha_j \phi_j(u) \right] \right\}$$

$$\times \exp(i2\pi\langle u, x\rangle/N) + \text{c.c.} \qquad (D5)$$

$$= \frac{h_k^*(x)}{N^2} \sum_{u \in \chi} i\phi_n(u) H_k(u) \exp(i2\pi\langle u, x\rangle/N) + \text{c.c.}, \qquad (D6)$$

where c.c. represents a term that is the complex conjugate of the preceding term.   Equation (D2) then becomes

$$\frac{\partial}{\partial \alpha_n} L = \sum_{k=1}^{K} \sum_{x \in \chi} \frac{d_k(x)}{g_k(x)} \sum_{x' \in \chi} f(x') \left[\frac{h_k^*(x - x')}{N^2} \sum_{u \in \chi} i\phi_n(u) H_k(u) \exp(i2\pi\langle u, x - x'\rangle/N) + \text{c.c.} \right] \qquad (D7)$$

$$= \left[\sum_{k=1}^{K} \sum_{x \in \chi} \frac{d_k(x)}{g_k(x)} \sum_{x' \in \chi} f(x') \frac{h_k^*(x - x')}{N^2} \sum_{u \in \chi} i\phi_n(u) H_k(u) \exp(i2\pi\langle u, x - x'\rangle/N) \right] + \text{c.c.} \qquad (D8)$$

$$= \left[\sum_{k=1}^{K} \sum_{u \in \chi} i\phi_n(u) H_k(u) \frac{1}{N^2} \sum_{x \in \chi} \frac{d_k(x)}{g_k(x)} \sum_{x' \in \chi} f(x') h_k^*(x - x') \exp(i2\pi\langle u, x - x'\rangle/N) \right] + \text{c.c.} \qquad (D9)$$

$$= \left[\sum_{k=1}^{K} \sum_{u \in \chi} i\phi_n(u) H_k(u) \frac{1}{N^2} \sum_{x \in \chi} \frac{d_k(x)}{g_k(x)} \sum_{x' \in \chi} f(x - x') h_k^*(x') \exp(i2\pi\langle u, x'\rangle/N) \right] + \text{c.c.} \qquad (D10)$$

$$= \left[\sum_{k=1}^{K} \sum_{u \in \chi} i\phi_n(u) H_k(u) \frac{1}{N^2} \sum_{x' \in \chi} h_k^*(x') \exp(i2\pi\langle u, x'\rangle/N) \sum_{x \in \chi} \frac{d_k(x) f(x - x')}{g_k(x)} \right] + \text{c.c.} \qquad (D11)$$

$$= -2 \, \text{Im}\left[\sum_{k=1}^{K} \sum_{u \in \chi} \phi_n(u) H_k(u) \frac{1}{N^2} \sum_{x' \in \chi} h_k^*(x') \exp(i2\pi\langle u, x'\rangle/N) \sum_{x \in \chi} \frac{d_k(x) f(x - x')}{g_k(x)} \right] \qquad (D12)$$

$$= -2 \sum_{u \in \chi} \phi_n(u) \text{Im}\left[\sum_{k=1}^{K} H_k(u) \frac{1}{N^2} \sum_{x' \in \chi} h_k^*(x') \exp(i2\pi\langle u, x'\rangle/N) \sum_{x \in \chi} \frac{d_k(x) f(x - x')}{\sum_{x''} f(x'') s_k(x - x'')} \right], \qquad (D13)$$

aberration parameters for the case of Poisson noise.   Recall that the log-likelihood function is given in Eq. (38). We now take the partial derivative,

$$\frac{\partial}{\partial \alpha_n} L = \sum_{k=1}^{K} \sum_{x \in \chi} \frac{d_k(x)}{\sum_{x''} f(x'') s_k(x - x'')} \sum_{x' \in \chi} f(x') \frac{\partial}{\partial \alpha_n} s_k(x - x')$$

$$(D1)$$

$$= \sum_{k=1}^{K} \sum_{x \in \chi} \frac{d_k(x)}{g_k(x)} \sum_{x' \in \chi} f(x') \frac{\partial}{\partial \alpha_n} s_k(x - x'), \qquad (D2)$$

where α_n is the nth parameter of the phase function ϕ.   Consider the partial derivative of the PSF, found in

where we have assumed that f is real valued in order to step from Eq. (D7) to Eq. (D8).   Equation (D13) is the desired expression for the partial derivative of the log-likelihood function with respect to aberration parameters.

APPENDIX E: DERIVATION OF THE EXPECTATION-MAXIMIZATION ALGORITHM FOR THE POISSON CASE

In this appendix we derive an EM algorithm for the joint estimation of the object and aberrations in the case of Poisson noise.   Recall that the E step of the EM algorithm requires the evaluation of the conditional expectation of the complete-data log likelihood:

$$Q[f, \alpha \mid f^{(r)}, \alpha^{(r)}] = E^{(r)}[L_{cd}(f, \alpha) \mid \{d_k\}], \qquad (E1)$$

where $E^{(r)}[\cdot]$ denotes an expectation assuming that the parameters governing the complete data take on the values $\{f^{(r)}, \alpha^{(r)}\}$. The log likelihood of the complete data was shown in Eq. (60) to be

$$
\begin{aligned}
L_{cd}(f, \alpha) = &\sum_{x' \in \chi} \left[\sum_{k=1}^{K} \sum_{x \in \chi} \tilde{d}_k(x \mid x') \right] \ln f(x') \\
&+ \sum_{k=1}^{K} \sum_{x' \in \chi} \left[\sum_{x \in \chi} \tilde{d}_k(x \mid x - x') \right] \ln s_k(x') \\
&- \sum_{x' \in \chi} f(x') \sum_{k=1}^{K} \sum_{x \in \chi} s_k(x).
\end{aligned}
\tag{E2}
$$

Examining this expression, we see that the only random variables are the $\{\tilde{d}_k\}$. Therefore for the E step we need to evaluate the conditional expectation

$$
E^{(r)}[\tilde{d}_k(x \mid x') \mid \{d_k\}] = E^{(r)}[\tilde{d}_k(x \mid x') \mid d_k(x)],
\tag{E3}
$$

where we remind the reader that

$$
d_k(x) = \sum_{x' \in \chi} \tilde{d}_k(x \mid x').
\tag{E4}
$$

From Eq. (E4) we see that the expectation in Eq. (E3) is conditioned on the requirement that all the $\tilde{d}_k(x \mid x')$ sum to a known value. As noted by Shepp and Vardi,[25] if z_1 and z_2 are independent Poisson random variables with means λ_1 and λ_2, respectively, then

$$
E[z_1 \mid z_1 + z_2] = \frac{\lambda_1}{\lambda_1 + \lambda_2}(z_1 + z_2).
\tag{E5}
$$

Through the use of this rule and Eq. (52), the conditional expectation in Eq. (E3) is readily evaluated as

$$
E^{(r)}[\tilde{d}_k(x \mid x') \mid d_k(x)] = \frac{f^{(r)}(x')s_k^{(r)}(x - x')}{g_k^{(r)}(x)} d_k(x),
\tag{E6}
$$

where

$$
g_k^{(r)}(x) \equiv \sum_{x' \in \chi} f_k^{(r)}(x)s_k^{(r)}(x - x'),
\tag{E7}
$$

$$
s_k^{(r)} \equiv k\text{th diversity PSF evaluated for } \alpha^{(r)}.
\tag{E8}
$$

Substituting Eqs. (E2) and (E6) into Eq. (E1), we have that

$$
\begin{aligned}
Q[f, \alpha \mid f^{(r)}, \alpha^{(r)}] \\
= &\sum_{x' \in \chi} \left[\sum_{k=1}^{K} \sum_{x \in \chi} \frac{f^{(r)}(x')s_k^{(r)}(x - x')}{g_k^{(r)}(x)} d_k(x) \right] \ln f(x') \\
&+ \sum_{k=1}^{K} \sum_{x' \in \chi} \left[\sum_{x \in \chi} \frac{f^{(r)}(x - x')s_k^{(r)}(x')}{g_k^{(r)}(x)} d_k(x) \right] \ln s_k(x') \\
&- \sum_{x' \in \chi} f(x') \sum_{k=1}^{K} S_k(0),
\end{aligned}
\tag{E9}
$$

where $S_k(0) = \Sigma_{x \in \chi} s_k(x)$ is independent of the aberration parameters. For the M step we determine the parameters $\{f, \alpha\}$ that maximize Q subject to any constraints that may exist and then assign the maximizing values to $\{f^{(r+1)}, \alpha^{(r+1)}\}$. Fortunately, this maximization may be performed separately over f and α. We may rewrite Q as

$$
Q[f, \alpha \mid f^{(r)}, \alpha^{(r)}] = Q_f + Q_\alpha,
\tag{E10}
$$

where we have defined

$$
\begin{aligned}
Q_f \equiv &\sum_{x' \in \chi} \left[\sum_{k=1}^{K} \sum_{x \in \chi} \frac{f^{(r)}(x')s_k^{(r)}(x - x')}{g_k^{(r)}(x)} d_k(x) \right] \ln f(x') \\
&- \sum_{x' \in \chi} f(x') \sum_{k=1}^{K} S_k(0),
\end{aligned}
\tag{E11}
$$

$$
Q_\alpha \equiv \sum_{k=1}^{K} \sum_{x' \in \chi} \left[\sum_{x \in \chi} \frac{f^{(r)}(x - x')s_k^{(r)}(x')}{g_k^{(r)}(x)} d_k(x) \right] \ln s_k(x').
\tag{E12}
$$

Then $f^{(r+1)}$ is obtained by maximizing Q_f, and $\alpha^{(r+1)}$ is obtained by maximizing Q_α. To maximize Q_f, observe that

$$
\begin{aligned}
\frac{\partial Q_f}{\partial f(x')} = &\frac{1}{f(x')} \sum_{k=1}^{K} \sum_{x \in \chi} \frac{f^{(r)}(x')s_k^{(r)}(x - x')}{g_k^{(r)}(x)} d_k(x) \\
&- \sum_{k=1}^{K} S_k(0),
\end{aligned}
\tag{E13}
$$

$$
\begin{aligned}
\frac{\partial^2 Q_f}{\partial f(x')\partial f(x'')} = &-\delta(x' - x'') \frac{1}{[f(x')]^2} \\
&\times \sum_{k=1}^{K} \sum_{x \in \chi} \frac{f^{(r)}(x')s_k^{(r)}(x - x')}{g_k^{(r)}(x)} d_k(x) \leq 0,
\end{aligned}
\tag{E14}
$$

where $\delta(x)$ represents a Kronecker delta. Therefore a global maximum of Q_f is found by setting Eq. (E13) to zero. Consequently, we assign to $f^{(r+1)}$ the values

$$
f^{(r+1)}(x') = f^{(r)}(x') \frac{1}{\sum_{k=1}^{K} S_k(0)} \sum_{k=1}^{K} \sum_{x \in \chi} s_k^{(r)}(x - x') \frac{d_k(x)}{g_k^{(r)}(x)}.
\tag{E15}
$$

Unfortunately, no closed-form expression is obvious for the maximizer of Q_α, so we write

$$
\begin{aligned}
\alpha^{(r+1)} = \arg\max_\alpha \Bigg\{ &\sum_{k=1}^{K} \sum_{x' \in \chi} \left[\sum_{x \in \chi} \frac{f^{(r)}(x - x')s_k^{(r)}(x')}{g_k^{(r)}(x)} d_k(x) \right] \\
&\times \ln s_k(x') \Bigg\}.
\end{aligned}
\tag{E16}
$$

The optimization in Eq. (E16) bears a striking resemblance to the optimization of $L(f, \alpha)$ when the object is known to be a point. In this special case Eq. (38) reduces to

$$
L(\alpha) = \sum_{k=1}^{K} \sum_{x \in \chi} d_k(x) \ln s_k(x) + C.
\tag{E17}
$$

Therefore we can use the closed-form expression for the gradient of the log-likelihood function in the point-object case in the optimization of Eq. (E16) by substituting

$$
d_k(x) \rightarrow \sum_{x' \in \chi} \frac{f^{(r)}(x' - x)s_k^{(r)}(x)}{g_k^{(r)}(x')} d_k(x')
\tag{E18}
$$

into Eq. (50).

ACKNOWLEDGMENTS

REFERENCES

1. J. D. Gonglewski, C. R. DeHainaut, C. M. Lampkin, and R. C. Dymale, "System design of a wavefront sensing package for a wide field of view optical phased array," Opt. Eng. **27,** 785–792 (1988).
2. 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).
3. R. G. Lyon, "HST phase retrieval: a parameter estimation," in *Applications of Digital Image Processing XIV*, A. G. Tescher, ed., Proc. Soc. Photo-Opt. Instrum. Eng. **1567,** 317–326 (1991).
4. J. R. Fienup, "Phase retrieval for the Hubble Space Telescope using iterative propagation algorithms," in *Applications of Digital Image Processing XIV*, A. G. Tescher, ed., Proc. Soc. Photo-Opt. Instrum. Eng. **1567,** 327–332 (1991).
5. R. A. Gonsalves and R. Childlaw, "Wavefront sensing by phase retrieval," in *Applications of Digital Image Processing III*, A. G. Tescher, ed., Proc. Soc. Photo-Opt. Instrum. Eng. **207,** 32–39 (1979).
6. R. A. Gonsalves, "Phase retrieval and diversity in adaptive optics," Opt. Eng. **21,** 829–832 (1982).
7. R. G. Paxman and J. R. Fienup, "Optical misalignment sensing and image reconstruction using phase diversity," J. Opt. Soc. Am. A **5,** 914–923 (1988).
8. R. G. Paxman and S. L. Crippen, "Aberration correction for phased-array telescopes using phase diversity," in *Digital Image Synthesis and Inverse Optics*, A. F. Gmitro, P. S. Idell, and I. J. LaHaie, eds., Proc. Soc. Photo-Opt. Instrum. Eng. **1351,** 787–797 (1990).
9. J. M. Mendel, *Lessons in Digital Estimation Theory* (Prentice-Hall, Englewood Cliffs, N.J., 1987).
10. A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing* (Prentice-Hall, Englewood Cliffs, N.J., 1989).
11. J. W. Goodman, *Introduction to Fourier Optics* (McGraw-Hill, San Francisco, Calif., 1968).
12. F. Roddier, "Passive versus active methods in optical interferometry," in *Proceedings of the ESO/NOAO Conference on High Resolution Imaging by Interferometry*, F. Merkle, ed. (European Southern Observatory, Garching, Federal Republic of Germany, 1988), pp. 565–574.
13. J. Primot, G. Rousset, and J. C. Fontanella, "Deconvolution from wave-front sensing: a new technique for compensating turbulence-degraded images," J. Opt. Soc. Am. A **7,** 1598–1608 (1990).
14. J. D. Gonglewski, D. G. Voelz, J. S. Fender, D. C. Dayton, B. K. Spielbusch, and R. E. Pierson, "First astronomical application of postdetection turbulence compensation: images of α Aurigae, ν Ursae Majoris, and α Geminorum using self-referenced speckle holography," Appl. Opt. **29,** 4527–4529 (1990).
15. D. C. Luenberger, *Linear and Nonlinear Programming* (Addison-Wesley, Reading, Mass., 1984).
16. D. L. Snyder and D. G. Politte, "Image reconstruction from list-mode data in an emission tomography system having time-of-flight measurements," IEEE Trans. Nucl. Sci. **NS-30,** 1843–1849 (1983).
17. D. G. Politte, "Reconstruction algorithms for time-of-flight assisted positron-emission tomographs," M.S. thesis (Sever Institute of Technology, Washington University, St. Louis, Mo., 1983).
18. L. Kaufman, "Implementing and accelerating the EM algorithm for positron emission tomography," IEEE Trans. Med. Imag. **MI-6,** 37–51 (1987).
19. A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," J. R. Stat. Soc. Ser. B **39,** 1–38 (1977).
20. C. F. J. Wu, "On the convergence properties of the EM algorithm," Ann. Stat. **11,** 95–103 (1983).
21. Y. Vardi, L. A. Shepp, and L. Kaufman, "A statistical model for positron emission tomography," J. Am. Stat. Assoc. **80,** 8–38 (1985).
22. W. H. Richardson, "Bayesian-based iterative method of image restoration," J. Opt. Soc. Am. **62,** 55–59 (1972).
23. L. B. Lucy, "An iterative technique for the rectification of observed distributions," Astron. J. **79,** 745–754 (1974).
24. N. B. Baranova and B. Y. Zel'dovich, "Dislocations of the wave-front surface and zeros of the amplitude," Sov. Phys. JETP **53,** 925–929 (1981).
25. L. A. Shepp and Y. Vardi, "Maximum-likelihood reconstruction for emission tomography," IEEE Trans. Med. Imag. **MI-1,** 113–121 (1982).