# Phase-retrieval algorithms for a complicated optical system

J. R. Fienup

Phase-retrieval algorithms have been developed that handle a complicated optical system that requires multiple Fresnellike transforms to propagate from one end of the system to the other including the absorption by apertures in more than one plane and allowance for bad detector pixels. Gradient-search algorithms and generalizations of the iterative-transform phase-retrieval algorithms are derived. Analytic expressions for the gradient of an error metric, with respect to polynomial coefficients and with respect to point-by-point phase descriptions, are given. The entire gradient can be computed with the number of transforms required to propagate a wave front from one end of the optical system to the other and back again, independent of the number of coefficients or phase points. This greatly speeds the computation. The reconstruction of pupil amplitude is also given. A convergence proof of the generalized iterative transform algorithm is given. These improved algorithms permit a more accurate characterization of complicated optical systems from their point spread functions.

## 1. Introduction

Phase retrieval is the determination of the phase of a complex-valued function from the magnitude of the function by using some *a priori* information about the function or its transform. The problem occurs in several fields including wave-front sensing,[1] astronomical imaging by interferometry,[2] and x-ray crystallography.[3] We discuss the problem of retrieving the phase of a wave front that has passed through a complicated optical system.

The motivation for the study described here was the desire to characterize accurately the Hubble Space Telescope (HST), which soon after launch was found to be severely aberrated. This characterization is needed (1) to design correction optics to be included in replacement instruments, (2) to align the secondary mirror of the HST's Optical Telescope Assembly, and (3) to form an analytic model of the system's point spread function (PSF) that can be used for image deblurring until the telescope is repaired. Unfortunately the conventional forms of interferometry and wave-front sensing that could be performed if the telescope were readily accessible on earth cannot be performed remotely in space. For this reason we rely on phase-retrieval algorithms that determine the aberrations (phase errors) from the PSF's measured with an onboard CCD detector array and transmitted to earth. The PSF's are blurred images of pointlike stars taken through narrow-band spectral filters. These algorithms should also be useful in the testing of general complicated optical systems.

Phase retrieval for the HST consists of finding an aberrated wave front (optical field), which, when digitally propagated through the optical system, gives rise to a wave front in the plane of the CCD array detector whose intensity, the modeled PSF, matches the measured PSF, the image of a star.

As is the case with most phase-retrieval problems, the relationship between the optical field in the entrance pupil and the optical field at the detector plane of the HST can be fairly accurately modeled as a Fourier (or Fresnel) transform. However, in the Wide-Field/Planetary camera (WF/PC) mode of the HST, the image formed by the main telescope, the Optical Telescope Assembly (OTA), is reimaged onto a CCD detector array by a relay telescope.[4,5] The WF/PC relay telescopes contain obscurations that are not in a plane conjugate to the entrance pupil of the OTA. Consequently, for the most accurate computation of a modeled PSF from an estimate of the aberrations and a model of the optical system, it is necessary to propagate digitally an aberrated wave front from the entrance pupil to the detector plane by using multiple propagation transforms and by multi-

The author is with the Optical and IR Science Laboratory, Environmental Research Institute of Michigan, P.O. Box 134001, Ann Arbor, Michigan 48113-4001.

Received 22 April 1992.
0003-6935/93/101737-10$05.00/0.

plying by appropriate masks representing the obscurations in the planes where they occur. Such detailed modeling is required to design correction optics to be used to fix the telescope with the desired accuracy. Plans are to accomplish this optical correction by replacing the present WF/PC relay optics with new optics that would include correction optics consisting of a mirror with aberrations that are opposite to those of the OTA. A second reason for such high accuracy is to compute analytically the PSF's that could be used for image deconvolution, which is sensitive to errors in the estimate of the PSF. Phase retrieval is also important as an aid in the alignment of the secondary mirror of the OTA.

Two popular phase-retrieval approaches for this application are gradient-search algorithms[6–8] and the iterative transform algorithm.[7–10] In Sections 2–4, expanding on previous results, we derive generalizations of these algorithms including the effects of multiple-plane propagation and of a weighting function that permits one to ignore bad CCD pixels. In particular, in Section 2 analytic expressions for the gradient of an error metric are given that allow for a fast computation of the gradient that is used by a gradient-search method. Reconstruction algorithms for a variety of parameters, including amplitude as well as phase, are derived. In Section 3 an alternative error metric is discussed. In Section 4 the iterative transform algorithm is derived, a convergence proof is given for the error reduction and Gerchberg–Saxton versions of the algorithm, and the necessity for combining algorithms is discussed. Conclusions are drawn in Section 5. Results for the HST are reserved for another paper.

## 2. Gradient-Search Phase-Retrieval Algorithms

Gradient-search techniques have been used in the past for solving the phase-retrieval problem.[6–8] As described in Ref. 8 the first three steps of the iterative transform algorithm accomplish quite efficiently most of the work of computing the gradient. Here we extend these results to include (1) multiple-plane diffraction, (2) a weighting function, and (3) differentiation with respect to a variety of parameters. Previously Guozhen *et al.*[11] derived the case of multiple-plane diffraction in a more limited circumstance.

Let $|F(u)|$ be the magnitude of the complex optical field at the detector array (which is estimated by taking the square root of the measured intensity of an image of a pointlike star taken through a narrow-band spectral filter), where $u$ is a two-dimensional (2-D) coordinate in the detector plane. (Although light leaving a star is incoherent, after it propagates to the telescope it appears to have originated from a point, and so it becomes spatially coherent, and after it passes through a narrow-band filter it becomes temporally coherent, which makes it fully coherent.) In what follows, coordinates are given as the integer pixel number within a digitized array in the computer.

We wish to minimize a weighted error metric:

$$E = \sum_u W(u)[|G(u)| - |F(u)|]^2, \tag{1}$$

where $W(u)$ is a weighting function and $|G(u)|$ is the magnitude of an optical wave front computed from a model of the optical system including an estimate of the phase errors. The summation is over all values of $u$ in the digitized array. $W(u)$ is zero at the locations of bad or saturated CCD pixels or where no data were measured. It can also be set to zero at pixels where the signal-to-noise ratio is poor. Several other error metrics could be used as well. Section 3 gives an example of another error metric and shows that the error metric above has some desirable properties.

$G(u)$, the wave front at the detector, is computed as follows. Let an estimate of the aberrations in the input plane (often the entrance pupil) of the optical system be $\theta(x_1)$, and let the pupil function (obscurations) be represented in that plane by a transmittance function $m_1(x_1)$, where $x_1$ is the 2-D spatial coordinate in that plane. The aberrations can be caused either by the optical system or by atmospheric turbulence immediately in front of the optical system. Then the optical wave front immediately after the input plane is

$$g(x_1) = U_1(x_1) = m_1(x_1)\exp[i\theta(x_1)], \tag{2}$$

where $i = \sqrt{-1}$. We denote the wave front in the $n$th plane by $U_n(x_n)$, but for compactness of notation we define $U_1(x_1) = g(x_1)$ in the input plane and $U_d(x_d) = G(u)$ in the plane of the detector. To propagate $g(x_1)$ to the detector plane, first digitally propagate it from the input plane to the first plane in which an obscuration (or an additional aberration) occurs, multiply the wave front by a transmittance function representing the obscuration (or aberration), then propagate it to the next obscuration, multiply it by the appropriate transmittance function, and continue to propagate and multiply until it reaches the $d$th plane, the detector plane, where $G(u) = U_d(x_d)$ results. Any one of the digital propagations between the planes of obscurations is computed by using the paraxial digital transform of the optical field:

$$\begin{aligned} U_{n+1}(x_{n+1}) &= \exp(i\alpha_n x_{n+1}^2)\frac{1}{N}\sum_{x_n=0}^{N-1} U_n(x_n) \\ &\quad \times \exp(i\beta_n x_n^2)\exp(-i2\pi x_n x_{n+1}/N) \\ &= \exp(i\alpha_n x_{n+1}^2)\mathscr{F}[U_n(x_n)\exp(i\beta_n x_n^2)], \end{aligned} \tag{3}$$

where the discrete Fourier transform is given by

$$\mathscr{F}[U_n(x_n)] = \frac{1}{N}\sum_{x_n=0}^{N-1} U_n(x_n)\exp(-i2\pi x_n x_{n+1}/N). \tag{4}$$

Here $x_n x_{n+1}$ means the dot (inner) product of the 2-D vector coordinates $x_n$ and $x_{n+1}$. The coefficients $\alpha_n$ and $\beta_n$ and the relative scaling represented by the spatial coordinates $x_n$ and $x_{n+1}$ are determined by the ABCD method.[12] The major computational burden for one propagation is a single discrete Fourier transform, which is computed by using the fast Fourier transform (FFT) method. Note that the normalizing factor in front of the 2-D discrete Fourier transform is $1/N$, which makes it unitary (power conserving). This propagation is similar to a digital version of the Fresnel transform,[13] except that the quadratic phase factors, $\alpha_n$ and $\beta_n$, differ from each other and from the coefficient of the Fourier kernel. In some cases it is necessary to propagate to another plane prior to propagating to the next plane of interest so that the array size for the FFT's can be kept small. Representing the obscuration in the $n$th plane by $m_n(x_n)$ and letting the last (the $d$th) plane be $u = x_d$, we calculate the optical field by the propagation

$$
\begin{aligned}
G(u) &= \mathbf{P}[g(x_1)] \\
&= \exp(i\alpha_{d-1}u^2)\frac{1}{N}\sum_{x_{d-1}} m_{d-1}(x_{d-1})\exp(i\beta_{d-1}{x_{d-1}}^2) \\
&\quad \times \exp(-i2\pi x_{d-1}u/N)\exp(i\alpha_{d-2}{x_{d-1}}^2) \\
&\quad \times \frac{1}{N}\sum_{x_{d-2}} m_{d-2}(x_{d-2})\exp(i\beta_{d-2}{x_{d-2}}^2) \\
&\quad \times \exp(-i2\pi x_{d-2}x_{d-1}/N)\exp(i\alpha_{d-3}{x_{d-2}}^2) \\
&\quad \times \frac{1}{N}\sum_{x_{d-3}} \cdots \exp(i\alpha_1 {x_2}^2) \\
&\quad \times \frac{1}{N}\sum_{x_1} g(x_1)\exp(-i2\pi x_1 x_2/N) \\
&= \exp(i\alpha_{d-1}u^2)\frac{1}{N}\sum_{x_{d-1}} \exp(-i2\pi x_{d-1}u/N) \\
&\quad \times M_{d-1}(x_{d-1})\frac{1}{N}\sum_{x_{d-2}} \exp(-i2\pi x_{d-2}x_{d-1}/N) \\
&\quad \times M_{d-2}(x_{d-2})\cdots M_2(x_2)\frac{1}{N} \\
&\quad \times \sum_{x_1} \exp(-i2\pi x_1 x_2/N) g(x_1) \\
&= \exp(i\alpha_{d-1}u^2)\mathscr{F}\{M_{d-1}(x_{d-1}) \\
&\quad \times \mathscr{F}\{M_{d-2}(x_{d-2})\cdots M_2(x_2)\mathscr{F}[g(x_1)]\}\ldots\}, \qquad (5)
\end{aligned}
$$

where

$$M_n(x_n) = \exp[i(\beta_n + \alpha_{n-1}){x_n}^2]m_n(x_n) \qquad (6)$$

is a generalized transmittance function that includes the amplitude transmittance caused by obscurations and the quadratic phase factors associated with the propagations. We can also include within $m_n(x_n)$ a phase error associated with the plane $x_n$ if we desire. For simplicity we included the phase $\beta_1 {x_1}^2$ within $\theta(x_1)$.

### A. First Derivatives with Respect to Input-Plane Parameters

One approach to minimizing the error metric $E$ as a function of unknown aberrations or system parameters is to use a gradient-search algorithm. The gradient of $E$ is composed of the set of partial derivatives of $E$ with respect to each unknown parameter. We derive the partial derivative of $E$ with respect to a parameter $p$ of the input plane in a way that is similar to what we did in Ref. 8:

$$
\begin{aligned}
\frac{\partial E}{\partial p} &= 2\sum_u W(u)[|G(u)| - |F(u)|]\frac{\partial |G(u)|}{\partial p} \\
&= \sum_u W(u)\left[G^*(u) - \frac{|F(u)|}{|G(u)|}G^*(u)\right]\frac{\partial G(u)}{\partial p} + \text{c.c.} \\
&= -\sum_u G^{w*}(u)\frac{\partial G(u)}{\partial p} + \text{c.c.}, \qquad (7)
\end{aligned}
$$

where the asterisk denotes the complex conjugate, c.c. denotes the complex conjugate of the term that precedes it, and

$$G^w(u) = W(u)\left[|F(u)|\frac{G(u)}{|G(u)|} - G(u)\right]. \qquad (8)$$

For the case of a parameter $p$ in the input plane, we have, because of the linearity of the propagation $P$,

$$\frac{\partial G(u)}{\partial p} = \frac{\partial}{\partial p}[\mathbf{P}[g(x_1)]] = \mathbf{P}\left[\frac{\partial g_1(x_1)}{\partial p}\right]. \qquad (9)$$

Inserting this into Eq. (7) yields

$$\frac{\partial E}{\partial p} = -\sum_u G^{w*}(u)P\left[\frac{\partial g(x_1)}{\partial p}\right] + \text{c.c.} \qquad (10)$$

If the gradient is evaluated in this form, its computation would require a propagation (which requires $d$ FFT's) for each of the unknown parameters $p$. Depending on the problem, there could be dozens of parameters for a polynomial-coefficient parameterization of the phase $\theta(x_1)$ and tens of thousands of parameters for a point-by-point map of $\theta(x_1)$. This evaluation would entail a quite demanding computational load. The required computations can be drastically reduced by using analytic expressions for the gradient, which we derive as follows. With the explicit form of the propagation and reversing the order of summation, we have

$$\frac{\partial E}{\partial p} = -\sum_{u} G^{w*}(u)\exp(i\alpha_{d-1}u^2)\frac{1}{N}$$
$$\times \sum_{x_{d-1}} \exp(-i2\pi u x_{d-1}/N) M_{d-1}(x_{d-1})$$
$$\times \frac{1}{N}\sum_{x_{d-2}} \exp(-i2\pi x_{d-2}x_{d-1}/N)$$
$$\times M_{d-2}(x_{d-2})\cdots$$
$$\times \frac{1}{N}\sum_{x_1} \exp(-i2\pi x_1 x_2/N)\frac{\partial g(x_1)}{\partial p} + \text{c.c.}$$
$$= -\sum_{x_1}\frac{\partial g(x_1)}{\partial p}$$
$$\times \frac{1}{N}\sum_{x_2}\exp(-i2\pi x_1 x_2/N) M_2(x_2)\cdots M_{d-2}(x_{d-2})$$
$$\times \frac{1}{N}\sum_{x_{d-1}}\exp(-i2\pi x_{d-2}x_{d-1}/N) M_{d-1}(x_{d-1})$$
$$\times \frac{1}{N}\sum_{u}\exp(-i2\pi u x_{d-1}/N)\exp(i\alpha_{d-1}u^2) G^{w*}(u)$$
$$+ \text{c.c.} \tag{11}$$

Note that $G^{w*}(u)$ is proportional to $G^*(u)$, which includes the phasor term $\exp(-i\alpha_{d-1}u^2)$. This term cancels the term $\exp(i\alpha_{d-1}u^2)$ in the equation above; thus without loss of generality we henceforth set $\alpha_{d-1} = 0$. This equation can also be expressed as

$$\frac{\partial E}{\partial p} = -\sum_{x_1}\frac{\partial g(x_1)}{\partial p}\mathbf{P}^{-1}[G^{w*}(u)] + \text{c.c.}$$
$$= -2\,\text{Re}\left\{\sum_{x_1}\frac{\partial g(x_1)}{\partial p}\mathbf{P}^{-1}[G^{w*}(u)]\right\}, \tag{12}$$

where we define a backward propagation operation as

$$\mathbf{P}^{-1}[G^{w*}(u)] = \mathscr{F}\{M_2(x_2)\cdots\mathscr{F}\{M_{d-2}(x_{d-2})$$
$$\times \mathscr{F}\{M_{d-1}(x_{d-1}(\mathscr{F}[G^{w*}(u)]\}\}\ldots\}. \tag{13}$$

Reversing the roles of the first term in Eq. (12) and its complex conjugate, we can also write

$$\frac{\partial E}{\partial p} = -\sum_{x_1}\left[\frac{\partial g(x_1)}{\partial p}\right]^* \mathbf{P}^{\dagger}[G^{w}(u)] + \text{c.c.}$$
$$= -2\,\text{Re}\left\{\sum_{x_1}\left[\frac{\partial g(x_1)}{\partial p}\right]^* \mathbf{P}^{\dagger}[G^{w}(u)]\right\}$$
$$= -2\,\text{Re}\left(\sum_{x_1}\frac{\partial g(x_1)}{\partial p}\{\mathbf{P}^{\dagger}[G^{w}(u)]\}^*\right), \tag{14}$$

where we define an inverse propagation operator as

$$\mathbf{P}^{\dagger}[G^{w}(u)] = \mathscr{F}^{-1}\{M_2{}^*(x_2)\cdots\mathscr{F}^{-1}\{M_{d-2}{}^*(x_{d-2})$$
$$\times \mathscr{F}^{-1}\{M_{d-1}{}^*(x_{d-1})\mathscr{F}^{-1}[G^{w}(u)]\}\}\ldots\}. \tag{15}$$

Equation (12) constitutes propagating $g(x_1)$ to form $G(u)$, computing $G^w(u)$ from $G(u)$, complex conjugating, propagating through the optical system backward, and then projecting the result with $\partial g(x_1)/\partial p$. This requires only $2d$ 2-D FFT's, independent of the number of parameters, which makes this form computationally much less demanding than when finite differences are employed to compute the gradient. Equation (14) constitutes the same steps for computing $G^w(u)$; then (without complex conjugating) we propagate $G^w(u)$ backward through the system by using inverse Fourier transforms and complex-conjugated phasors, complex conjugating, and projecting the result with $\partial g(x_1)/\partial p$. If we define the inverse propagated wave front in the $x_1$ plane to be

$$g^w(x_1) = \mathbf{P}^{\dagger}[G^w(u)], \tag{16}$$

we see that

$$\mathbf{P}^{-1}[G^{w*}(u)] = g^{w*}(x_1) = \{\mathbf{P}^{\dagger}[G^w(u)]\}^*. \tag{17}$$

Now we consider specific examples of parameters $p$.

### 1. *Polynomial Phase Coefficients*

Let the parameter $p$ be $a_j$, the $j$th coefficient of a polynomial expansion of $\theta(x_1)$:

$$\theta(x_1) = \sum_{j=1}^{J} a_j Z_j(x_1). \tag{18}$$

where $Z_j(x_1)$ is the $j$th polynomial. Differentiating Eq. (2) with respect to $a_j$ yields

$$\frac{\partial g(x_1)}{\partial a_j} = g(x_1) i Z_j(x_1). \tag{19}$$

Inserting this into Eq. (14) yields

$$\frac{\partial E}{\partial a_j} = -2\,\text{Re}\left[\sum_{x_1} i g(x_1) Z_j(x_1) g^{w*}(x_1)\right]$$
$$= 2\,\text{Im}\left[\sum_{x_1} g(x_1) Z_j(x_1) g^{w*}(x_1)\right]. \tag{20}$$

### 2. *Point-by-Point Phase*

Let the parameter $p$ be the value of the phase $\theta(x_1)$ at the point $x_1 = x$. We have

$$\frac{\partial g(x_1)}{\partial \theta(x)} = i g(x)\delta(x, x_1), \tag{21}$$

where

$$\delta(x, x_1) = \begin{cases} 1, & x_1 = x, \\ 0, & \text{otherwise}, \end{cases} \tag{22}$$

is the Kronecker δ function. Inserting this into Eq. (14) yields

$$\frac{\partial E}{\partial \theta(x)} = -2\,\mathrm{Re}\left[\sum_{x_1} ig(x)\delta(x, x_1)g^{w*}(x_1)\right]$$
$$= 2\,\mathrm{Im}[g(x)g^{w*}(x)], \tag{23}$$

where we now take the output of $\mathbf{P}^\dagger$ to be in the $x_1 = x$ plane.

Note that this is a generalization of the gradients computed in earlier works.[8,11]

### 3. *Point-by-Point Magnitude*

Let the parameter $p$ be the value of the magnitude $m_1(x_1)$ at point $x_1 = x$. We have

$$\frac{\partial g(x_1)}{\partial m_1(x)} = \exp[i\theta(x)]\delta(x, x_1). \tag{24}$$

Inserting this into Eq. (14) yields

$$\frac{\partial E}{\partial m_1(x)} = 2\,\mathrm{Re}\left\{\sum_{x_1} \exp[i\theta(x)]\delta(x, x_1)g^{w*}(x_1)\right\}$$
$$= -2\,\mathrm{Re}\{\exp[i\theta(x)]g^{w*}(x)\}. \tag{25}$$

### 4. *Point-by-Point Complex Values*

Let the parameter $p$ be the value of the real or imaginary part of $g(x_1)$ at point $x_1 = x$. Letting

$$g(x_1) = g_R(x_1) + ig_I(x_1), \tag{26}$$

we have

$$\frac{\partial g(x_1)}{\partial g_R(x)} = \delta(x, x_1), \tag{27}$$

$$\frac{\partial g(x_1)}{\partial g_I(x_1)} = i\delta(x, x_1). \tag{28}$$

Inserting each of these into Eq. (14) yields

$$\frac{\partial E}{\partial g_R(x)} = -2\,\mathrm{Re}\left[\sum_{x_1} \delta(x, x_1)g^{w*}(x_1)\right]$$
$$= -2\,\mathrm{Re}[g^{w*}(x)] = -2\,\mathrm{Re}[g^{w}(x)], \tag{29}$$

$$\frac{\partial E}{\partial g_I(x)} = -2\,\mathrm{Re}\left[\sum_{x_1} i\delta(x, x_1)g^{w*}(x_1)\right]$$
$$= 2\,\mathrm{Im}[g^{w*}(x)] = -2\,\mathrm{Im}[g^{w}(x)]. \tag{30}$$

These two partial derivatives can be combined by using the following notation for a derivative with respect to a complex number:

$$\frac{\partial E}{\partial g(x)} \equiv \frac{\partial E}{\partial g_R(x)} + i\,\frac{\partial E}{\partial g_I(x)} = -2g^{w}(x), \tag{31}$$

which reveals the nature of $g^w(x)$.

These forms of the gradient, which include both weighting functions in the error metric and multiple-plane propagation effects, are generalizations of the results obtained earlier.[8] Here $g^w(x)$ takes on the role of $g'(x) - g(x)$ in that earlier work.

These gradients can be used with any gradient-search algorithm. For example, consider gradients with respect to the real and imaginary values of $g(x_1)$ when the steepest descent algorithm is used. Starting with an estimate of the unknown parameters, we would compute the gradient. Then we would perform a line search, computing $E$ as a function of $g(x_1; s)$, which depends on the step size $s$ according to

$$g(x_1; s) = g(x_1) - s\,\frac{\partial E}{\partial g(x_1)}$$
$$= g(x_1) + 2sg^{w}(x_1). \tag{32}$$

The minimum of $E$ as a function of $s$ would be found to arrive at a new estimate of the unknown parameters. This process is iterated until no further progress is made. The steepest descent is usually the slowest-converging gradient-search algorithm; conjugate gradient and others are much better.[14,15]

## B. First Derivatives with Respect to Intermediate-Plane Parameters

Let $q$ be a parameter of $m_n(x_n)$. It could be an unknown phase (aberration) or magnitude (transmittance) parameter. Then we have, just like Eq. (7),

$$\frac{\partial E}{\partial q} = -\sum_u G^{w*}(u)\,\frac{\partial G(u)}{\partial q} + \text{c.c.}, \tag{33}$$

where $G^w(u)$ is as defined in Eq. (8). Differentiating Eq. (5) with respect to $q$, we have (recalling that $\alpha_{d-1} = 0$)

$$\frac{\partial G(u)}{\partial q} = \frac{1}{N}\sum_{x_{d-1}} \cdots \sum_{x_n} \exp(-i2\pi x_n x_{n-1}/N)\,\frac{\partial M_n(x_n)}{\partial q}$$
$$\cdots \frac{1}{N}\sum_{x_1} \exp(-i2\pi x_1 x_2/N)g(x_1). \tag{34}$$

Inserting this into Eq. (33) and rearranging the order of summation yield

$$\frac{\partial E}{\partial q} = -\sum_{x_n} \frac{\partial m_n(x_n)}{\partial q}\,\mathbf{P}_{1\to n}[g(x_1)]\mathbf{P}_{d\to n}^{-1}[G^{w*}(u)] + \text{c.c.}$$
$$= -\sum_{x_n} \frac{\partial m_n(x_n)}{\partial q}\,\mathbf{P}_{1\to n}[g(x_1)]\{\mathbf{P}_{d\to n}^{\dagger}[G^{w}(u)]\}^* + \text{c.c.}, \tag{35}$$

where

$$\mathbf{P}_{1\to n}[g(x_1)] = \exp(i\alpha_{n-1}{x_n}^2)\frac{1}{N}$$
$$\times \sum_{x_{n-1}} \exp(-i2\pi x_{n-1}x_n/N)M_{n-1}(x_{n-1})$$
$$\cdots \frac{1}{N}\sum_{x_1} \exp(-i2\pi x_1 x_2/N)g(x_1), \tag{36}$$

$$\mathbf{P}_{d\to n}^{-1}[G^{w}{}^{*}(u)] = \exp(i\beta_n x_n{}^2)\frac{1}{N} \times \sum_{x_{n+1}} \exp(-i2\pi x_n x_{n+1}/N) M_{n+1}(x_{n+1}) \cdots \frac{1}{N}\sum_{u} \exp(-i2\pi x_{d-1}u/N) G^{w}{}^{*}(u). \tag{37}$$

In the expressions above $\mathbf{P}_{1\to n}[g(x_1)]$ constitutes propagating $g(x_1)$ to the $x_n$ plane [after multiplication by the quadratic phasor but prior to multiplication by $m_n(x_n)$], and $\mathrm{P}_{d\to n}^{-1}[G^{w}{}^{*}(u)]$ constitutes propagating $G^{w}{}^{*}(u)$ backward to the $x_n$ plane. In this case only $d + (d - n) = 2d - n$ FFT's are required since $\mathbf{P}_{1\to n}[g(x_1)]$ is an intermediate result in the process of computing $G^{w}(u)$. The notation $\mathbf{P}_{d\to n}^{\dagger}(\cdot)$ refers to inverse propagation (as described earlier) from plane $u$ to plane $x_n$.

Consider a plane in which both magnitude and phase uncertainties are present:

$$m_n(x_n) = |m_n(x_n)|\exp[i\theta_n(x_n)] \tag{38}$$

for $1 < n < d$.

When parameter $q$ is a polynomial phase coefficient, with

$$\theta_n(x_n) = \sum_{j=1} b_j Z_j(x_n), \tag{39}$$

then

$$\frac{\partial m_n(x_n)}{\partial b_j} = m_n(x_n) i Z_j(x_n), \tag{40}$$

$$\frac{\partial E}{\partial b_j} = 2\,\mathrm{Im}\left(\sum_{x_n} m_n(x_n) Z_j(x_n) \mathbf{P}_{1\to n}[g(x_1)] \times \{\mathbf{P}_{d\to n}^{\dagger}[G^{w}(u)]\}^{*}\right). \tag{41}$$

Similarly, when $q$ is a point-by-point phase value,

$$\frac{\partial m_n(x_n)}{\partial \theta_n(x)} = i m_n(x)\delta(x, x_n), \tag{42}$$

$$\frac{\partial E}{\partial \theta_n(x)} = 2\,\mathrm{Im}(m_n(x)\mathbf{P}_{1\to n}[g(x_1)]\{\mathbf{P}_{d\to n}^{\dagger}[G^{w}(u)]\}^{*}), \tag{43}$$

where the propagated wave fronts are evaluated at $x_n = x$.

When parameter $q$ is the magnitude $|m_n(x_n)|$ of the transmittance in the $x_n$ plane,

$$\frac{\partial m_n(x_n)}{\partial |m_n(x)|} = \exp[i\theta_n(x_n)]\delta(x, x_n), \tag{44}$$

$$\frac{\partial E}{\partial |m_n(x)|} = -2\,\mathrm{Re}(\exp[i\theta_n(x)]\mathbf{P}_{1\to n}[g(x_1)] \times \{\mathbf{P}_{d\to n}^{\dagger}[G^{w}(u)]\}^{*}), \tag{45}$$

where the propagated wave fronts are evaluated at $x_n = x$.

For the HST there was a poorly known translation of obscurations arising from the secondary mirror and spiders of the WF/PC. For this we can assume that $|m_n(x_n)|$ representing that obscuration is a binary function that is known except for its position. Therefore we need to determine only the relative position of a known mask. This can be done without performing an iterative optimization by using the gradient as defined above. When we inspect the negative of the gradient, it becomes apparent where the mask should be located. The obscurations of the mask $|m_n(x_n)|$ should be moved to the area of the minima of the negative of the gradient. The same is true for determining $m_1(x_1)$. If the size (scale) of the obscurations is uncertain, this might also be inferred directly from the gradient.

With the gradients derived above a gradient-search nonlinear optimization algorithm, such as the conjugate-gradient algorithm, can be used to minimize the error metric and thereby estimate the unknown phase errors and other parameters.

### C. Second Derivatives with Respect to Input-Plane Parameters

Methods employing second partial derivatives, such as Newton–Raphson, generally converge with far fewer iterations than gradient-search methods. Analytic expressions can be found for the second partial derivatives in the same manner as the first partial derivatives. When this is done for the case of $J$ polynomial coefficients, we find that the total number of propagations required is proportional to $J^2$. For a small $J$ this would be practical; but for a large $J$ it is more efficient to employ a greater number of iterations of a gradient-search method than the fewer number of iterations of a method using all the second partial derivatives. Alternatively an efficient strategy might be to use a method that employs only the diagonal terms of the matrix of second partial derivatives. Since for the case of point-by-point (phase, magnitude, or complex) values there are typically $N^2 \simeq$ thousands of points, the inversion of an $N^2$ by $N^2$ matrix would be a formidable task, and so nondiagonal second partial derivatives in this case would be of limited interest.

## 3. Derivatives for a Second Error Metric

Several other error metrics, besides the squared difference in magnitudes given in Eq. (1), may be used. Consider, for example, the squared difference in intensities (squared magnitudes) given by

$$E_2 = \sum_{u} W(u)[|G(u)|^2 - |F(u)|^2]^2. \tag{46}$$

The derivative with respect to a parameter $p$ in the input plane is

$$\frac{\partial E_2}{\partial p} = 2\sum_u W(u)[|G(u)|^2 - |F(u)|^2]G^*(u)\frac{\partial G(u)}{\partial p} + \text{c.c.}$$

$$= 4\,\text{Re}\left\{\sum_u [|G(u)|^2 - |F(u)|^2]G^*(u)\mathbf{P}\left[\frac{\partial g(x_1)}{\partial p}\right]\right\}$$

$$= 4\,\text{Re}\left(\sum_{x_1} \frac{\partial g(x_1)}{\partial p}\{\mathbf{P}^\dagger[|G(u)|^2G^*(u) - |F(u)|^2G^*(u)]\}^*\right), \tag{47}$$

which is derived with steps that are similar to those leading to Eq. (14). The expressions for $\partial g(x_1)/\partial p$ for various parameters $p$ are given in Subsection 2.A. The products of three fields, $|G(u)|^2G^*(u) = G(u)G^*(u)G^*(u)$ and similarly for $|F(u)|^2G^*(u)$, would require a sampling rate that is 3 times that which would be required to propagate the one field $G(u)$ without aliasing. For many practical circumstances this degree of oversampling would not be done, and aliasing would occur. (This would be true for the HST application.) Some aliasing would also occur in the computation of $\partial E/\partial p$ by Eq. (14), but we believe that typically it would be considerably less. This is one reason why we favor the error metric $E$ given in Eq. (1) over $E_2$ given in Eq. (46). We also favor $E$ given in Eq. (1), because, for photon-limited measurements, it can be shown that minimizing it is approximately the same as maximizing a log-likelihood metric; i.e., it is close to a maximum-likelihood solution.

## 4. Iterative Propagation Algorithms

For the iterative transform algorithm (ITA) approaches, we begin by describing projection-onto-sets types of algorithm, e.g., error reduction[8] and Gerchberg–Saxton,[9] and then move onto an input–output type of algorithm to speed convergence and avoid stagnation at local minima.[8] In this case of multiple-plane propagation the algorithm should be renamed the iterative propagation algorithm (IPA). The error-reduction version of the algorithm is derived as follows.

### A. Algorithm

In the $u$ domain (detector plane) the ordinary projection operation (which makes the wave front consistent with the measured data) is given by[8]

$$G'(u) = |F(u)|G(u)/|G(u)|. \tag{48}$$

Since our error metric contains a weighting function $W(u)$, we must consider a weighted projection. With a binary weighting function the projection operation is given by

$$G^p(u) = W(u)\frac{|F(u)|G(u)}{|G(u)|} + [1 - W(u)]G(u) \tag{49a}$$

$$= G(u)\left\{W(u)\left[\frac{|F(u)|}{|G(u)|} - 1\right] + 1\right\}$$

$$= W(u)G(u)\left[\frac{|F(u)|}{|G(u)|} - 1\right] + G(u)$$

$$= G^w(u) + G(u), \tag{49b}$$

where $G^w(u)$ is the same function given by Eq. (8) defined for the gradient-search algorithms. From Eq. (49a) we see that $G^p(u)$ has a magnitude of $|F(u)|$ where $W(u) = 1$ and is equal to $G(u)$ where $W(u) = 0$. That is, where we have no confidence in the data $|F(u)|$, we leave the wave front unaltered. The algorithm will retrieve the phase of $F(u)$ where $W(u) = 1$ and interpolate the wave front (retrieve both magnitude and phase) where $W(u) = 0$. A weighting function taking on any value between 0 and 1 could also be used in Eqs. (49a) and (49b), although the concept of a projection (in a Hilbert space) is less well defined in that case.

Suppose that in the $x_1$ domain the magnitude is replaced by the measured magnitude (i.e., Gerchberg–Saxton or error reduction with a magnitude constraint) or the magnitude is set to zero outside the known region of support (outside the pupil function) (i.e., error reduction with a support constraint). Then, for the case of propagation consisting of a single Fourier transform, it has been shown[8] that this (error-reduction) algorithm converges in the weak sense that the error metric cannot increase with increasing iterations. This is a result of the algorithm being a projection-onto-sets algorithm (nonconvex sets in this case). However, for the case of multiple-plane propagation this is not true. Note that $\mathbf{P}^\dagger\{\mathbf{P}[g(x_1)]\} = g(x_1)$ only for propagations that do not involve obscurations. The lack of a convergence proof results from the fact that light is absorbed at the $x_n$ plane by obscurations $|m_n(x_n)|$ during the forward propagation, and the corresponding light is not regenerated during an inverse propagation. When the inverse propagated wave front arrives at the $x_1$ plane, missing from it is the wave front from the obscured region(s) that are necessary to make it match the input wave front.

One solution to this problem of the missing wave front is as follows. Let $\mathbf{P}_{1\to n}[g(x_1)]$ denote the forward propagation of the wave front $g(x_1)$ to the $x_n$ plane and let $\mathbf{P}^\dagger_{d\to n}[G^p(u)]$ represent the inverse propagation of the wave front $G^p(u)$ to the $x_n$ plane. During the forward propagation, save the wave fronts $P_{1\to n}[g(x_1)]$ and then on the inverse propagation add them back in. For example, if the last plane with an obscuration is the $n$th, then, after inverse propagating to the $x_n$ plane and multiplying by the mask

$|m_n(x_n)|$, replace $|m_n(x_n)|\mathbf{P}^{\dagger}_{d\to n}[G^p(u)]$ by

$$\mathbf{P}^{\dagger +}_{d\to n^+}[G^p(u)] = |m_n(x_n)|\mathbf{P}^{\dagger}_{d\to n}[G^p(u)] + [1 - |m_n(x_n)|]\mathbf{P}_{1\to n}[g(x_1)], \quad (50)$$

which restores the portion of the wave front absorbed during the forward propagation. Then complete the inverse propagation in this fashion to arrive at a wave front, which we will call $g^p(x_1)$, in the input plane.

The first three steps of the iteration would be the propagation of $g(x_1)$ to $G(u)$, projecting to form $G^p(u)$, and inverse propagating, adding back the obscured wave fronts to obtain $g^p(x_1)$. This is mathematically identical to a simpler form, which we now show. Inserting Eqs. (49) into Eq. (50) yields, at the first obscuration in the inverse propagation,

$$\begin{aligned}\mathbf{P}^{\dagger +}_{d\to n^+}&[G^p(u)] \\ &= |m_n(x_n)|\mathbf{P}^{\dagger}_{d\to n}[G^w(u)] + |m_n(x_n)|\mathbf{P}^{\dagger}_{d\to n}[G(u)] \\ &\quad + [1 - |m_n(x_n)|]\mathbf{P}_{1\to n}[g(x_1)] \\ &= |m_n(x_n)|\mathbf{P}^{\dagger}_{d\to n}[G^w(u)] + \mathbf{P}_{1\to n}[g(x_1)], \quad (51)\end{aligned}$$

since

$$|m_n(x_n)|\mathbf{P}^{\dagger}_{d\to n}[G(u)] = |m_n(x_n)|\mathbf{P}_{1\to n}[g(x_1)] \quad (52)$$

at that obscuration. Completing the inverse propagation of this wave front back to the $x_1$ plane, adding back the obscured wave fronts as we go along, yields

$$g^p(x_1) = \mathbf{P}^{\dagger}[G^w(u)] + g(x_1) = g^w(x_1) + g(x_1). \quad (53)$$

That is, by inverse propagating $G^w(u)$ (without adding back obscured wave fronts) to obtain $g^w(x_1)$, and adding to it $g(x_1)$, we obtain the same result as the more complicated procedure of inverse propagating $G^p(u)$ when obscured wave fronts are added back.

The IPA consists of the following four steps (when the simpler method of inverse propagation is used) for the $k$th iteration:

(1) Propagate an input wave front $g_k(x_1)$ to the measurement plane: $G_k(u) = \mathbf{P}[g_k(x_1)]$ by using Eq. (5).

(2) Compute $G_k{}^w(u)$ from $G_k(u)$ by using Eq. (8).

(3) Inverse propagate $G_k{}^w(u)$ back to the input plane (without adding back the obscured wave fronts) giving $g_k{}^w(x_1)$. Then compute $g_k{}^p(x_1) = g_k{}^w(x_1) + g_k(x_1)$.

(4) Form the new input wave front $g_{k+1}(x_1)$ from $g_k{}^p(x_1)$ and $g_k(x_1)$ by using any version of the ITA.[8]

### B. Convergence Proof for Error-Reduction Version

A convergence proof for the error reduction and Gerchberg–Saxton versions of the algorithm above is as follows. At the $k$th iteration the $u$-domain error metric is (for binary $W$ and binary $|m_n|$)

$$\begin{aligned}E_{Fk} &= \sum_u W(u)[|G_k(u)| - |F(u)|]^2 \\ &= \sum_u W(u)\left|\frac{G_k(u)}{|G_k(u)|}|G_k(u)| - \frac{G_k(u)}{|G_k(u)|}|F(u)|\right|^2 \\ &= \sum_u |-G_k{}^w(u)|^2 = \sum_u |G_k{}^w(u)|^2. \quad (54)\end{aligned}$$

When it is inverse propagated back to the input plane, $G^w(u)$ will lose some energy as a result of the multiplication by the masks $|m_n(x)|$, and thus we have

$$E_{Fk} = \sum_u |G_k{}^w(u)|^2 = \sum_{x_1} |g^w(x_1)|^2 + C_1, \quad (55)$$

where

$$C_1 = \sum_{n=2}^{d-1}\sum_{x_n}[1 - |m_n(x_n)|]|\mathbf{P}^{\dagger}_{d\to n}[G_k{}^w(u)]|^2 \quad (56)$$

is the energy from $G_k{}^w(u)$ that is lost as a result of the obscurations on inverse propagation to the input plane. According to Eq. (53) the computed wave front in the $x_1$ domain is given by $g_k{}^p(x_1) = g_k{}^w(x_1) + g_k(x_1)$. For the error reduction and Gerchberg–Saxton algorithms the new input wave front $g_{k+1}(x_1)$ is formed by projecting $g_k{}^p(x_1)$ onto the $x_1$-domain constraint space (for example, by setting it to zero outside the support constraint). Then

$$\begin{aligned}E_{Fk} &= \sum_{x_1} |g_k{}^w(x_1)|^2 + C_1 \\ &= \sum_{x_1} |g_k{}^p(x_1) - g_k(x_1)|^2 + C_1 \\ &\geq \sum_{x_1} |g_k{}^p(x_1) - g_{k+1}(x_1)|^2 + C_1 \equiv E_{ok} + C_1, \quad (57)\end{aligned}$$

since both $g_k(x_1)$ and $g_{k+1}(x_1)$ are within the $x_1$-domain constraint space and by definition $g_{k+1}(x_1)$ is the wave front in that constraint space closest to $g_k{}^p(x_1)$. $E_{ok}$ is the $x_1$-domain error metric. Since $C_1 \geq 0$, $E_{Fk} \geq E_{ok}$.

Using $g_k{}^p(x_1) = g_k{}^w(x_1) + g_k(x_1)$ and propagating $[g_k(x_1) - g_{k+1}(x_1)]$ back to the $u$ plane, we have

$$\begin{aligned}E_{Fk} &\geq \sum_{x_1} |g_k{}^w(x_1) + g_k(x_1) - g_{k+1}(x_1)|^2 + C_1 \\ &= \sum_u |G_k{}^w(u) + G_k(u) - G_{k+1}(u)|^2 + C_2, \quad (58)\end{aligned}$$

where

$$C_2 = \sum_{n=2}^{d-1}\sum_{x_n}[1 - |m_n(x_n)|]|\mathbf{P}_{1\to n}[g_k(x_1) - g_{k+1}(x_1)]|^2 \quad (59)$$

is the energy lost to obscurations in the propagation of $[g_k(x_1) - g_{k+1}(x_1)]$ through the system. To obtain this result, we had to add the optical fields with energy $C_1$, lost by inverse propagation of $G^w(u)$ to $g^w(x_1)$, back to $g_k{}^w(x_1)$ at each obscuration as it forward propagated through the system in order to arrive at $G^w(u)$. This is made possible by the fact that the field being added back in each plane is nonzero only where $|m_n(x_n)| = 0$. Therefore we have

$$\begin{aligned} E_{Fk} &\geq \sum_u |G_k{}^p(u) - G_{k+1}(u)|^2 + C_2 \\ &\geq \sum_u |G_{k+1}{}^p(u) - G_{k+1}(u)|^2 + C_2 \\ &= \sum_u \left| W(u)G_{k+1}(u)\left[\frac{|F(u)|}{|G_{k+1}(u)|} - 1\right]\right|^2 + C_2 \\ &= \sum_u W(u)[|G_{k+1}(u)| - |F(u)|]^2 + C_2 \\ &= E_{F(k+1)} + C_2 \geq E_{F(k+1)}, \end{aligned} \tag{60}$$

where we used the fact that since both $G_k{}^p(u)$ and $G^p_{k+1}(u)$ are within the $u$-domain constraint space, and since by definition $G^p_{k+1}(u)$ is the wave front in the constraint space closest to $G_{k+1}(u)$, $G^p_{k+1}(u)$ is closer to $G_{k+1}(u)$ than $G_k{}^p(u)$ is. Since $E_{Fk} \geq E_{F(k+1)}$ we have proved convergence of the error-reduction (Gerchberg–Saxton) algorithms in the usual weak sense that the error metric is nonincreasing with the increasing iteration number. This has now been shown to be true in the case of multiple-plane propagation including diffraction (absorption) and with a weighting function for ignoring bad pixels.

The more advanced forms of the ITA's[8] are expected to converge faster than the error-reduction algorithm, which has the convergence proof discussed above. For the case where $m_1(x_1)$ is used in the aperture plane as a support constraint, the fourth step of the error-reduction algorithm would be

$$g_{k+1}(x_1) = m_1(x_1)g_k{}^p(x_1), \tag{61}$$

whereas for the hybrid input–output algorithm the fourth step would be

$$\begin{aligned} g_{k+1}(x_1) &= m_1(x_1)g_k{}^p(x_1) \\ &\quad + [1 - m_1(x_1)][g_k(x_1) - \beta g_k{}^p(x_1)], \end{aligned} \tag{62}$$

where $\beta$ is a feedback parameter.[8] The fourth step for the Gerchberg–Saxton algorithm, which uses $m_1(x_1)$ as a magnitude constraint, would be

$$g_{k+1}(x_1) = m_1(x_1)\frac{g_k{}^p(x_1)}{|g_k{}^p(x_1)|}\,. \tag{63}$$

### C. Comparison with Gradient Search

For the error-reduction algorithm with a support constraint, inserting Eq. (53) into Eq. (61) yields

$$g_{k+1}(x_1) = m_1(x_1)[g_k{}^w(x_1) + g_k(x_1)]. \tag{64}$$

Now compare this with the steepest-descent gradient search for the parameters that are the real and imaginary parts (or complex values) of the wave front in the input plane. For a step size $s = 1/2$ in Eq. (32), the steepest descent gives us

$$\begin{aligned} g_{k+1}(x_1) &= g_k(x_1) + 2(1/2)g_k{}^w(x_1) \\ &= g_k{}^w(x) + g_k(x_1) \end{aligned} \tag{65}$$

for an unconstrained gradient search. However, for the gradient search over the complex values to be effective, the wave front should be constrained to zero outside the clear aperture. That is, a constrained steepest-descent gradient search would be accomplished by

$$g_{k+1}(x_1) = m_1(x_1)[g_k{}^w(x_1) + g_k(x_1)]. \tag{66}$$

Since this is identical to Eq. (64) we see, exactly as before,[8] that the error-reduction version of the IPA is identical to a steepest-descent gradient-search algorithm with a particular step size.

As usual[8] other gradient-search algorithms and the hybrid input–output version of the ITA are expected to converge much faster than the error-reduction/steepest-descent algorithms, even though for the latter there is a convergence proof.

### D. Combination of Algorithms

For the HST problem the aberration to be reconstructed is dominated by spherical aberration, but it also has small amounts of other polynomial-type phase errors and some fine-scale phase errors associated with the microroughness of the mirror surfaces left by the polishing process. Consequently the phase error can be described as the sum of several polynomials (only the spherical aberration term of which is large) plus a point-by-point phase map (with small values only). In theory the IPA's can, with no help from other algorithms, determine all the phase-error terms. Indeed for image reconstruction applications the ITA, despite some tendencies to stagnate,[16] has been demonstrated to converge to the solution from random starting points. However, for the HST problem, we found that the IPA would stagnate on a nonphysical aperture-plane wave front (a wave front that goes through zeros in clear areas of the OTA pupil) unless it was started near the true solution with a smooth phase. Furthermore, the locations of obscurations within the WF/PC (i.e., our support constraints) were poorly known. Therefore for the HST problem a combined approach was necessary. The combined approach used on the HST data was as follows: (1) Starting with an initial estimate of the locations of the obscurations, we performed a Zernike polynomial fit to obtain the low-order aberration terms. (2) Using this estimate of the phase error, we reestimated the position of the obscurations, as described at the end of Subsection 2.B. (3) We repeated steps (1) and (2) until no further changes were made (once or twice was adequate). At this point an estimate of the polynomial phase errors was available. (4) Starting with the polynomial phase-error estimate, we used the IPA to estimate the fine structure in the phase error.

## 5. Conclusions

We have derived generalizations of both gradient-search and iterative-transform phase-retrieval algorithms that allow for multiple planes of diffraction (vignetting or obscurations) within a complicated optical system and for ignoring bad pixels when a binary weighting function is used. A proof of convergence (in a weak sense) for the error-reduction and Gerchberg–Saxton algorithms, which employ multiple-plane propagation, was given. An equivalence was shown between the error-reduction algorithm and a constrained steepest-descent gradient-search algorithm with a particular step size. However, other gradient-search methods and versions of the IPA are recommended. The use of the derived analytic gradients, with respect to either the values of a point-by-point phase map or the coefficients of a polynomial expansion of the phase, can greatly speed the computation compared with the finite-differences computation of the gradient. Other versions of these algorithms are also useful for reconstructing the obscurations (pupil function) of an optical system. A combination of algorithms may be required to reconstruct polynomial-type phase errors, fine-scale phase errors, and unknown pupil functions. These generalized algorithms were motivated by the need to characterize accurately the aberrations and alignment of the HST. Results for the given in Ref. 17. They should also be useful for characterizing other complicated optical systems from measured PSF's.

Portions of this work are presented in Refs. 18 and 19.

This work was supported by the Jet Propulsion Laboratory and by Environmental Research Institute of Michigan internal research and development funds.

## References

1. 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).
2. J. C. Dainty and J. R. Fienup, "Phase retrieval and image reconstruction for astronomy," in *Image Recovery: Theory and Application,* H. Stark, ed. (Academic, New York, 1987), Chap. 7, pp. 231–275.
3. G. H. Stout and L. H. Jensen, *X-Ray Structure Determination* (Macmillan, London, 1968).
4. C. Burrows, *Hubble Space Telescope Optical Telescope Assembly Handbook* (Space Telescope Science Institute, Baltimore, Md., 1990).
5. R. Griffiths, *Hubble Space Telescope Wide Field and Planetary Camera Instrument Handbook* (Space Telescope Science Institute, Baltimore, Md., 1990).
6. R. A. Gonsalves, "Phase retrieval from modulus data," J. Opt. Soc. Am. A **66,** 961–964 (1976).
7. W. O. Saxton, *Computer Techniques for Image Processing in Electron Microscopy* (Academic, New York, 1978).
8. J. R. Fienup, "Phase retrieval algorithms: a comparison," Appl. Opt. **21,** 2758–2769 (1982).
9. R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," Optik (Stuttgart) **35,** 237–246 (1972).
10. J. R. Fienup, "Reconstruction and synthesis applications of an iterative algorithm," in *Transformations in Optical Signal Processing,* W. T. Rhodes, J. R. Fienup, and B. E. A. Saleh, eds., Proc. Soc. Photo-Opt. Instrum. Eng. **373,** 147–160 (1981).
11. Y. Guozhen, W. Li, D. Bizhen, and G. Benyuan, "On the amplitude-phase retrieval problem in an optical system involved non-unitary transformation," Optik (Stuttgart) **75,** 68–74 (1987).
12. A. E. Siegman, *Lasers* (University Science, Mill Valley, Calif., 1986), Chaps. 15 and 20.
13. J. W. Goodman, *Introduction to Fourier Optics* (McGraw-Hill, San Francisco, Calif., 1968).
14. D. P. Feder, "Automatic optical design," Appl. Opt. **2,** 1209–1226 (1963).
15. A. K. Rigler and R. J. Pegis, "Optimization methods in optics," in B. R. Frieden, ed., *The Computer in Optical Research* (Springer-Verlag, New York, 1980), Chap. 4, pp. 211–268.
16. J. R. Fienup and C. C. Wackerman, "Phase retrieval stagnation problems and solutions," J. Opt. Soc. Am. A **3,** 1897–1907 (1986).
17. J. R. Fienup, J. C. Marron, T. J. Schulz, and J. H. Seldin, "Hubble Space Telescope characterized by using phase-retrieval algorithms," Appl. Opt. **31,** 1747–1767 (1993).
18. J. R. Fienup, "Phase retrieval for the Hubble Space Telescope using iterative propagation algorithms," in *Applications of Digital Image Processing XIV,* Proc. Soc. Photo-Opt. Instrum. Eng. **1567,** 33 (1991).
19. J. R. Fienup, "HST aberrations and alignment determined by phase retrieval algorithms," in *Space Optics for Astrophysics and Earth and Planetary Remote Sensing,* Vol. 19 of 1991 OSA Technical Digest Series (Optical Society of America, Washington, D.C., 1991), pp. 19–21.