

Theoretical performance of phase retrieval on a subdivided aperture

Marcos A. van Dam
 Richard G. Lane*
 University of Canterbury
 Department of Electrical and Electronic
 Engineering
 Private Bag 4800
 Christchurch, New Zealand
 E-mail: r.lane@elec.canterbury.ac.nz

Abstract. Phase retrieval is a nonlinear technique used to recover the phase in the Fourier domain using intensity measurements at the image plane and additional constraints. We describe a method to solve the phase retrieval problem using linear iterations near the solution, which provides both analytical insight into phase retrieval and numerical results. The algorithm finds the maximum *a posteriori* estimate of the phase using prior information about the statistics of the noise and the phase, and it was found to converge well in practice. When phase retrieval is performed on data from subdivided apertures, there is a loss of information regarding the relative piston terms of the subapertures. This error is quantified. We find that there is a smaller wavefront error when estimating the phase from a full aperture rather than from a subdivided aperture. Using a combination of intensity measurements from full and subdivided apertures results in a small improvement at very high photon levels only. © 2002 Society of Photo-Optical Instrumentation Engineers. [DOI: 10.1117/1.1475336]

Subject terms: phase retrieval; adaptive optics; wavefront sensing.

Paper 010099 received Mar. 16, 2001; revised manuscript received Oct. 30, 2001; accepted for publication Jan. 2, 2002.

1 Introduction

Light from astronomical objects undergoes phase aberrations when passing through atmospheric turbulence. These aberrated wavefronts propagate to the aperture of the telescope, where the object is imaged onto a detector at the focal plane. An important problem in astronomical imaging is to estimate the phase of a wavefront at the aperture of the telescope. It is not possible to measure the phase of the wavefronts directly, so the phase must be recovered from intensity measurements.

The wave is focused onto the focal plane of a telescope. We adopt the notation $\mathbf{x}=(x,y)$ to represent the Cartesian coordinates in the aperture, with the corresponding coordinates (ξ,η) used in the focal plane where the photons are detected. The relationship between the incident wave of amplitude $A(\mathbf{x})$ and phase $\phi(\mathbf{x})$, $A(\mathbf{x})\exp[i\phi(\mathbf{x})]$, and the complex amplitude at the focal plane, $w(\xi,\eta)$, is given by¹:

$$w(\xi,\eta) = \frac{\exp[i\pi(\xi^2 + \eta^2)/\lambda f]}{i\lambda f} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A(\mathbf{x}) \times \exp[i\phi(\mathbf{x})] \exp\left[-i\frac{2\pi}{\lambda f}(x\xi + y\eta)\right] dx dy, \quad (1)$$

where λ is the wavelength of the photons and f is the focal length of the lens or mirror. When we define $\mathbf{u}=(u,v) = (\xi,\eta)/\lambda f$ and ignoring the term before the integral sign,

the complex amplitudes at the aperture and the image plane form a two-dimensional Fourier transform pair:

$$w(u,v) = \mathcal{F}[A \exp[i\phi(x,y)]]\{u,v\}, \quad (2)$$

where \mathcal{F} denotes the Fourier transform,

$$\mathcal{F}[f(x,y)]\{u,v\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y) \times \exp[-i2\pi(xu + yv)] dx dy. \quad (3)$$

A detector at the focal plane measures the intensity $I(\mathbf{u})$, where

$$I(\mathbf{u}) = |w(\mathbf{u})|^2. \quad (4)$$

The phase retrieval problem here consists of determining the most likely phase screen $\phi(\mathbf{x})$ from a noisy intensity measurement $I(\mathbf{u})$, and prior information about the statistics of the phase and the noise.

We investigate how photon noise limits the performance of any phase retrieval algorithm in estimating the phase of the wavefront. Section 2 describes the phase retrieval problem, which we solve in Sec. 3 by linearizing the problem around the solution. In this way, we can obtain a maximum-likelihood estimate of the phase and quantify the error resulting from the addition of noise. We also use the Cramér-Rao lower bound to compute this performance limit.

Phase retrieval is nonlinear and consequently difficult to solve. The nonlinearity of the phase retrieval problem has often been avoided by using other wavefront sensing techniques, notably the Shack-Hartmann sensor, to estimate the

*Current affiliation: Observatoire de Lyon, 9 Avenue Charles André, 69 561 St.-Genis-Laval Cedex, France.

phase.² Conventional Shack–Hartmann sensing uses the linear relationship between the displacement of the speckle image and the mean wavefront slope in an aperture to estimate the slope of the phase. Subdividing the aperture produces a set of slope estimates, and the resulting linear equations are then solved to obtain the phase. However, each speckle image contains more information than just the slope, and a subdivided aperture poses a set of linked phase retrieval problems. The feasibility of doing phase retrieval using images from subdivided apertures has been proposed^{3–6} and is the subject of further investigation here. A difficulty that arises is that there is an irretrievable loss of information that occurs when the aperture is subdivided, and we quantify this analytically in Sec. 5. Simulation results are presented in Sec. 7 to confirm the theory.

2 The Phase Retrieval Problem

The Fourier transform model requires the intensity to be measured continuously and for an infinite extent. In practice, the Fourier transform is approximated by the discrete Fourier transform (DFT), so that $w(u,v)$ is given by

$$w(u,v) = \frac{1}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} A(x,y) \exp[i\phi(x,y)] \times \exp\left[-i \frac{2\pi}{N}(xu + yv)\right], \quad (5)$$

where N is the number of samples in each dimension of the phase screen. The intensity is now

$$I(u,v) = \left| \frac{1}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} A(x,y) \exp[i\phi(x,y)] \times \exp\left[-i \frac{2\pi}{N}(xu + yv)\right] \right|^2. \quad (6)$$

This approximation can lead to aliasing when the phase screen is undersampled, and to spectral leakage, due to the assumption of periodicity. The extent of the aperture is defined by an $N \times N$ rectangular array inserted in the middle of a $2N \times 2N$ array of zeros, because the intensity distribution has twice the bandwidth of the complex amplitude at the focal plane. At good astronomical sites, $A(\mathbf{x})$ is almost constant within the aperture,⁷ so we assume that the amplitude is a constant that forces $I(u,v)$ to sum to 1. Outside the aperture, $A(\mathbf{x})$ is zero. We then use the DFT to obtain a $2N \times 2N$ array of intensity measurements.

The objective of phase retrieval is to estimate the wavefront from an intensity measurement.^{8,9} If there is no noise, an exact semianalytical solution can be obtained.¹⁰ In the presence of noise, however, there exist many algorithms for obtaining an approximate solution, for example those devised by Fienup,¹¹ and Irwan and Lane.¹²

The solution obtained is, however, not unique, as there are three changes that can be made to the phase screen that do not affect the intensity distribution¹⁰:

Ambiguity 1: Addition of a constant to $\phi(\mathbf{x})$.

Ambiguity 2: Addition of a multiple of 2π to any point in $\phi(\mathbf{x})$ for monochromatic light.

Ambiguity 3: Replacing $\phi(\mathbf{x})$ by $-\phi(-\mathbf{x})$ if $A(\mathbf{x})$ is symmetrical.

The mean phase is called the piston term, and is set to zero to remove the first ambiguity. The other two are a problem for iterative phase retrieval algorithms, because there can be convergence to the wrong solution. This obstacle can be avoided by starting sufficiently close to the true solution, which is the case in the analysis here. We then obtain the solution in the presence of noise by adopting a linear model close to the true solution. In practice, this implies an algorithm capable of tracking phase variations, assuming that the rate of change is sufficiently slow so that a good starting phase estimate is available.

3 Linearization of Phase Retrieval

3.1 Motivation

We wish to consider the effect of photon noise on the solution and quantify the error induced by the noise. To find the error, we need to know the best estimate of the phase that can be made using the intensity measurements and prior information about the statistics of the noise and wavefront phase. One way to do this is to linearize the problem around the true phase screen. Linearization reduces a nonlinear problem to a linear one by discarding the higher order terms. The linear problem is easier to solve, and it is a good approximation to the full problem, provided that the effect of the higher order terms is small. This is the case when the deviations from the true phase are small.

Wilde¹³ has also proposed a linear phase retrieval scheme, but the scheme uses three or more intensity images with known phase offsets and only small wavefront perturbations. The algorithm linearizes the phase around a planar wavefront rather than the true phase.

3.2 Direct Solution

If we perturb a unit amplitude phase screen $\phi(\mathbf{x})$ by a small amount $\Delta\phi(\mathbf{x})$, we obtain a perturbed complex amplitude at the focal plane, $\tilde{w}(\mathbf{u})$:

$$\tilde{w}(\mathbf{u}) = \mathcal{F}[A(\mathbf{x}) \exp[i(\phi(\mathbf{x}) + \Delta\phi(\mathbf{x}))]] \approx \mathcal{F}[A(\mathbf{x}) \exp[i\phi(\mathbf{x})] \cdot (1 + i\Delta\phi(\mathbf{x}))], \quad (7)$$

using the approximation

$$\exp[i\Delta\phi(\mathbf{x})] \approx 1 + i\Delta\phi(\mathbf{x}). \quad (8)$$

The change in complex amplitude $\Delta w(\mathbf{u})$ can be approximated by

$$\Delta w(\mathbf{u}) = \tilde{w}(\mathbf{u}) - w(\mathbf{u}) = \mathcal{F}[i\Delta\phi(\mathbf{x}) \cdot A(\mathbf{x}) \exp[i\phi(\mathbf{x})]], \quad (9)$$

which is a linear function of $\Delta\phi(\mathbf{x})$. Using an overbar to denote complex conjugation, we can write the new intensity as

$$I(\mathbf{u}) + \Delta I(\mathbf{u}) = \overline{[w(\mathbf{u}) + \Delta w(\mathbf{u})]} \cdot [w(\mathbf{u}) + \Delta w(\mathbf{u})]. \quad (10)$$

Therefore, using $\text{Re}[\cdot]$ to denote the real part of an expression,

$$\begin{aligned} \Delta I(\mathbf{u}) &= \overline{w(\mathbf{u})} \cdot \Delta w(\mathbf{u}) + w(\mathbf{u}) \cdot \overline{\Delta w(\mathbf{u})} + \overline{\Delta w(\mathbf{u})} \cdot \Delta w(\mathbf{u}) \\ &\approx 2 \text{Re}[\overline{w(\mathbf{u})} \cdot \Delta w(\mathbf{u})], \end{aligned} \quad (11)$$

by making the approximation

$$|\Delta w(\mathbf{u})|^2 \approx 0. \quad (12)$$

The approximation is only valid when the intensity is much larger than the change in intensity. Combining Eqs. (9) and (11) yields

$$\begin{aligned} \Delta I(\mathbf{u}) &= 2 \text{Re}[\overline{\mathcal{F}[A(\mathbf{x}) \exp[i\phi(\mathbf{x})]]} \cdot \mathcal{F}[i\Delta\phi(\mathbf{x}) \\ &\quad \cdot A(\mathbf{x}) \exp[i\phi(\mathbf{x})]]], \end{aligned} \quad (13)$$

which can be rewritten as a system of linear equations,

$$\Delta \mathbf{I} = H \Delta \phi. \quad (14)$$

\mathbf{I} , $\Delta \mathbf{I}$, ϕ , and $\Delta \phi$ are vectors obtained by stacking the columns of the respective matrices. The columns of the $4N^2 \times N^2$ matrix H are found by setting one element in $\Delta \phi$ at a time to 1 and evaluating Eq. (13). We start with a phase screen ϕ_0 that gives rise to an intensity distribution \mathbf{I}_0 . The intensity is normalized so that $\Sigma \mathbf{I}_0 = 1$. From a noisy measurement, \mathbf{I} , normalized such that $\Sigma \mathbf{I} = 1$, we want to estimate the phase screen that most probably caused that distribution. If the noise is normally distributed, then the least-squares estimator gives the maximum-likelihood estimate. For an intensity perturbation of $\Delta \mathbf{I} = \mathbf{I} - \mathbf{I}_0$, the corresponding phase screen estimate is obtained by the least-squares solution to Eq. (14):

$$\Delta \phi = (H^T H)^+ H^T \Delta \mathbf{I}, \quad (15)$$

where $(H^T H)^+$ denotes the generalized inverse¹⁴ of $H^T H$. Matrix $H^T H$ is singular, because it is of rank $N^2 - 1$. One way to see this is that no matter what the intensity is, $\Sigma \Delta \phi = 0$.

3.3 Linear Iterative Algorithm

The problem at hand is to find the phase distribution $\hat{\phi}$, with a corresponding intensity $\hat{\mathbf{I}}$, that minimizes the error metric $\Sigma(\mathbf{I} - \hat{\mathbf{I}})^2$. When the linear approximations apply, Eq. (15) defines the solution. However, for the noise levels of interest, the errors in the linear approximations of Eqs. (8) and (12) play an important role. As the perturbation $\Delta \phi$ increases, so do the errors involved in the approximations. One way to combat this problem is to apply the linear model iteratively, so that at each step $\Delta \phi$ is small. At each iteration, the H matrix is updated. The linear iterative algorithm (LIA) is as follows:

Step 1: Initially $\hat{\phi} = \phi_0$, $\hat{\mathbf{I}} = \mathbf{I}_0$.

Step 2: Calculate H using Eq. (13).

Step 3: Let $\Delta \hat{\phi} = (H^T H)^+ H^T (\mathbf{I} - \hat{\mathbf{I}})$.

Step 4: Let $\Delta \hat{\phi}_{\max}$ be equal to the element of $\Delta \hat{\phi}$ with the largest absolute value.

Step 5: Set the step size t to $t = \min(\gamma / \Delta \hat{\phi}_{\max}, 1)$.

Step 6: Let $\hat{\phi} = \hat{\phi} + t \Delta \hat{\phi}$.

Step 7: Let $\hat{\mathbf{I}} = |\mathcal{F}[A \exp[i\hat{\phi}]]|^2$.

Step 8: If $\hat{\phi}$ has not converged, go to Step 2.

The parameter γ dictates the maximum change in phase at each iteration. In our simulations, γ was set to 0.001, which ensured convergence without being excessively slow.

The LIA is a modified version of Newton's method.¹⁵ The modification is in Step 5, where the step size is limited. There are many other algorithms that can also be used to minimize $\Sigma(\mathbf{I} - \hat{\mathbf{I}})^2$. For comparison, we also employed the steepest descent and the conjugate gradient methods,¹⁵ but found that the convergence was slower.

The question of convergence of the algorithms to a global minimum is tightly linked to the uniqueness of the solution. Sanz and Huang¹⁶ note that the problem appears overdetermined, since the number of equations $4N^2$ exceeds the number of unknowns N^2 . The solution is unique to within the three ambiguities mentioned in Sec. 2 if H is of rank $N^2 - 1$. While this was the case in the simulations, we have been unable to provide an analytic proof.

4 Prior Information

A better estimate can be obtained by adding prior information about the measurement noise and the phase screen to be estimated. Wavefront reconstruction from Shack-Hartmann data routinely employs prior information about the turbulence: it either implicitly assumes that the wavefront is smooth¹⁷ or it explicitly assumes that the wavefront obeys Kolmogorov statistics.¹⁸

The addition of prior information to the phase retrieval problem was shown by Irwan and Lane to result in a substantial reduction in the error, provided that the starting estimate is close to the solution.¹² Using $E[\cdot]$ to denote the expectation operator, the phase covariance matrix C_ϕ is defined as

$$C_\phi(i, j) = E[\phi(i)\phi(j)], \quad (16)$$

and is obtained by the method of Wallner.¹⁹ However, we use $\phi(i)$ here to denote the mean phase in that pixel, not the phase sampled at the center, as was the case in our previous work.²⁰ The mean phase has the desirable property that the mean phases over the entire aperture sum to zero, while the sum of the points at the center of the pixels do not. The mean phase is also more compatible with the DFT, which assumes that the phase is integrated over the pixel. In addition, we show in Sec. 5.3 that there is a smaller error in reconstructing the phase from subdivided aperture data when the mean phase is the estimated quantity. The statis-

tics of the phase screens are taken to be Kolmogorov,⁷ and the arrival of photons at each pixel of the detector is modeled as independent Poisson processes. Since a Gaussian distribution is a good approximation to the Poisson distribution when the average number of photons is high, we write the noise covariance matrix $C_{\Delta I}$ as

$$C_{\Delta I}(i,j) = E[\Delta I(i)\Delta I(j)] = \begin{cases} I(i)/P & \text{if } i=j \\ 0 & \text{otherwise,} \end{cases} \quad (17)$$

where P is the total number of photons detected at the focal plane.

We define the solution $\hat{\phi}$ to the phase retrieval problem as being the phase screen that minimizes an error metric composed of the measurements and prior information:

$$\hat{\phi} = \underset{\phi}{\min} \left\{ \phi^T C_{\phi}^+ \phi + (|\mathcal{F}[A \exp[i\phi]]|^2 - I)^T \times C_{\Delta I}^{-1} (|\mathcal{F}[A \exp[i\phi]]|^2 - I) \right\}. \quad (18)$$

This is the maximum *a posteriori* (MAP) estimate of ϕ . It must be emphasized that C_{ϕ} is singular, because it forces the sum of the phases to be zero. This means that the generalized inverse C_{ϕ}^+ must be used in place of the inverse.

If we now utilize the linear model and let $\phi = \phi_0 + \Delta\phi$, the quantity to be minimized is

$$\phi^T C_{\phi}^+ \phi + (H\Delta\phi - \Delta\mathbf{I})^T C_{\Delta I}^{-1} (H\Delta\phi - \Delta\mathbf{I}). \quad (19)$$

Equation (19) can be expanded to

$$\phi^T C_{\phi}^+ \phi + (H\phi - H\phi_0 - \Delta\mathbf{I})^T C_{\Delta I}^{-1} (H\phi - H\phi_0 - \Delta\mathbf{I}). \quad (20)$$

Assuming that the noise $\Delta\mathbf{I}$ is normally distributed, this expression is minimized by

$$\Delta\phi = (H^T C_{\Delta I}^{-1} H + C_{\phi}^+)^{-1} H^T C_{\Delta I}^{-1} (H\phi_0 + \Delta\mathbf{I}) - \phi_0. \quad (21)$$

4.1 Iterative Solution

As in Sec. 3.3, the phase corresponding to a noisy intensity measurement was computed by iterating the linear approximation. At each iteration $C_{\Delta I}^{-1}$ is calculated from $\hat{\mathbf{I}}$ and H is calculated using Eq. (13). The algorithm is as follows:

Step 1: Initially $\hat{\phi} = \phi_0$, $\hat{\mathbf{I}} = \mathbf{I}_0$.

Step 2: Calculate H and $C_{\Delta I}^{-1}$.

Step 3: Let $\Delta\hat{\phi} = (H^T C_{\Delta I}^{-1} H + C_{\phi}^+)^{-1} H^T C_{\Delta I}^{-1} (H\hat{\phi} + \mathbf{I} - \hat{\mathbf{I}}) - \hat{\phi}$

Step 4: Let $\Delta\hat{\phi}_{\max}$ be equal to the element of $\Delta\hat{\phi}$ with the largest absolute value.

Step 5: Set the step size t to $t = \min(\gamma/\Delta\hat{\phi}_{\max}, 1)$.

Step 6: Let $\hat{\phi} = \hat{\phi} + t\Delta\hat{\phi}$.

Step 7: Let $\hat{\mathbf{I}} = |\mathcal{F}[A \exp[i\hat{\phi}]]|^2$.

Step 8: If $\hat{\phi}$ has not converged, go to Step 2.

This algorithm does not have guaranteed convergence to the minimum value of Eq. (18). However, in practice the algorithm converged monotonically to the solution, except when the aperture was subdivided and $P \leq 10^5$, in which case the algorithm sometimes oscillated between two solutions.

4.2 Comparison with Phase Retrieval Algorithm

The performance of the linear iterative method was compared to another algorithm based on MAP estimation. Irwan and Lane¹² solved the phase retrieval problem using conjugate gradient minimization. The same prior information about the statistics of the phase was used, but the noise was explicitly set to be Poisson, leading to a different error metric to that of Eq. (18):

$$\hat{\phi} = \underset{\phi}{\max} \left\{ \sum (\mathbf{I} \cdot \ln(|\mathcal{F}[A \exp[i\phi]]|^2) - |\mathcal{F}[A \exp[i\phi]]|^2) + \phi^T C_{\phi}^+ \phi \right\}. \quad (22)$$

The error metric of Eq. (22) represents the natural logarithm of the likelihood of a phase screen. To compare the linear iterative algorithm to this phase retrieval algorithm, random phase screens were simulated with $D/r_0 = 1$ and $P = 10^8$. The starting point was set by displacing the true phase by another random phase screen with $D/r_0 = 0.1$. Note that the displacement is much greater than the error expected by the photon noise alone. It was found that the linear iterative algorithm had a mean-squared error of $(1.70 \pm 0.03) \times 10^{-6}$, which is the same as was obtained starting from the true phase. By contrast, the conjugate gradient algorithm based on Poisson statistics returned a mean-squared error of $(1.8 \pm 1.0) \times 10^{-5}$. The difference can be attributed to the different statistical model, which confirms previous experience that Gaussian statistics are more robust than Poisson statistics.²¹

5 Aperture Subdivision

5.1 Subdivided Aperture

The aperture can be subdivided, as if it were a Shack-Hartmann sensor, and each subdivision of the aperture is focused onto a different point on the focal plane. We consider the case of subdividing the aperture into a 2×2 array; any further subdivision can be considered to be a repetition of this process. Each of the four subapertures $A(x) \exp[i\phi_j(x)]$ is embedded in an $N \times N$ array of zeros and Fourier transformed using an $N \times N$ point DFT to obtain intensity vectors \mathbf{I}_j for each of the $j = \{1, 2, 3, 4\}$ subapertures. We can calculate the corresponding $N^2 \times N^2/4$ matrices H_j to obtain

$$\begin{bmatrix} \Delta \mathbf{I}_1 \\ \Delta \mathbf{I}_2 \\ \Delta \mathbf{I}_3 \\ \Delta \mathbf{I}_4 \end{bmatrix} = \begin{bmatrix} H_1 & 0 & 0 & 0 \\ 0 & H_2 & 0 & 0 \\ 0 & 0 & H_3 & 0 \\ 0 & 0 & 0 & H_4 \end{bmatrix} \begin{bmatrix} \Delta \phi_1 \\ \Delta \phi_2 \\ \Delta \phi_3 \\ \Delta \phi_4 \end{bmatrix}. \quad (23)$$

Note that the ordering of the phase coordinates in the ϕ vector has changed to reflect the geometry of the problem. The columns of each subaperture are stacked separately, one subaperture at a time. We cannot solve for $\Delta \phi_j$ directly through the generalized inverse of the matrix in Eq. (23), because the system of equations is underdetermined. This occurs because there is no way to determine the relative mean phases between the subapertures from the intensity measurements obtained. Consequently, prior information must be used to obtain a solution in this case. The phase covariance matrix for the full aperture C_ϕ imposes a smoothness constraint on the phase at the boundary of the subdivision that overcomes this difficulty. Inserting the C_ϕ matrix is equivalent to adding an equation for every free variable. The equation is obtained by assuming that the elements of ϕ are normally distributed with zero mean and with a covariance given by C_ϕ . This guarantees a solution to the problem. If we label the three quantities in Eq. (23) $\Delta \mathbf{I}$, H , and $\Delta \phi$, respectively, the solution for $\Delta \phi$ is again given by Eq. (21).

5.2 Full and Subdivided Aperture Data

The feasibility of using both full and subdivided aperture imaging to do phase retrieval has been experimentally confirmed by Roggemann, Schulz, and Ngai,⁵ Takahashi and Takajo⁶ also found that its convergence was much quicker than using only the image from the full aperture in the absence of noise.

We implemented it by combining the matrices from Secs. 4 and 5.1 to obtain the following system of equations:

$$\begin{bmatrix} \Delta \mathbf{I} \\ \Delta \mathbf{I}_1 \\ \Delta \mathbf{I}_2 \\ \Delta \mathbf{I}_3 \\ \Delta \mathbf{I}_4 \end{bmatrix} = \begin{bmatrix} & & & & H \\ H_1 & 0 & 0 & 0 & 0 \\ 0 & H_2 & 0 & 0 & 0 \\ 0 & 0 & H_3 & 0 & 0 \\ 0 & 0 & 0 & H_4 & 0 \end{bmatrix} \begin{bmatrix} \Delta \phi_1 \\ \Delta \phi_2 \\ \Delta \phi_3 \\ \Delta \phi_4 \end{bmatrix}. \quad (24)$$

Again, we rename the three expressions in Eq. (24) $\Delta \mathbf{I}$, H , and $\Delta \phi$ and solve using Eq. (21).

We first consider the case where the photons are evenly distributed between the full and subdivided apertures. Because there are only half as many photons at each detector, the signal-to-noise ratio of each measurement decreases by a factor of two. We then show by simulation that the reconstruction does not depend strongly on the exact partitioning of the photons between the full and subdivided apertures.

5.3 Loss of Information in Subdivision

When the aperture is subdivided into a 2×2 array, the four intensity measurements give no indication about the relative piston of each subaperture. The relative pistons can be estimated using prior knowledge of the phase covariance, since the phases in the different subapertures are statisti-

cally correlated. However, this correlation is not deterministic, so there is an unavoidable error in the piston estimation for a finite number of pixels.

Let ϕ_0 be the initial phase screen. In the absence of noise, the MAP estimate from subdivided aperture data is:

$$\hat{\phi} = \psi \{ (\phi_0 - L\psi)^T C_\phi^+ (\phi_0 - L\psi) \}. \quad (25)$$

Here, $\psi = [\psi_1, \psi_2, \psi_3, \psi_4]^T$ is the piston added to each subaperture and L is the $N^2 \times 4$ matrix that relates the phase at each pixel to the corresponding piston term ψ_j :

$$L = \begin{bmatrix} \mathbf{1} & 0 & 0 & 0 \\ 0 & \mathbf{1} & 0 & 0 \\ 0 & 0 & \mathbf{1} & 0 \\ 0 & 0 & 0 & \mathbf{1} \end{bmatrix}. \quad (26)$$

Equation (25) states that the estimator predicts the phase screen obtained by adding piston terms to the subapertures, so that the overall phase best matches the Kolmogorov statistics. The difference between this estimate and the original phase screen is the error that results from aperture subdivision. The solution to Eq. (25) is²²

$$\psi = (L^T C_\phi^+ L)^+ L^T C_\phi^+ \phi_0 = B \phi_0, \quad (27)$$

where B is a $4 \times N^2$ matrix. Random phase screens with Kolmogorov statistics can be obtained directly from the covariance matrix.²³ We find the eigenvalues λ_i and the corresponding eigenvectors \mathbf{u}_i of C_ϕ . Let

$$U = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{N^2}], \quad (28)$$

and

$$\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \dots, \lambda_{N^2}]^T. \quad (29)$$

A random vector \mathbf{x} is generated, where each element is an independent zero-mean Gaussian random variable with the variance given by the corresponding element in $\boldsymbol{\lambda}$. The phase screen is given by

$$\phi_0 = U\mathbf{x}. \quad (30)$$

Hence,

$$\psi = BU\mathbf{x}, \quad (31)$$

and

$$E[|\psi|^2] = E[|BU\mathbf{x}|^2] = E[|BU|^2] E[|\mathbf{x}|^2] = |BU|^2 \boldsymbol{\lambda}. \quad (32)$$

Here, $|\cdot|^2$ indicates element-by-element squaring rather than matrix multiplication. Equation (32) holds because the elements in \mathbf{x} are statistically independent. Note that the covariance matrix is proportional to $(D/r_0)^{5/3}$, so the squared error is also proportional to $(D/r_0)^{5/3}$, where D is the length of the aperture and r_0 is the Fried parameter.

Table 1 Mean-squared error due to aperture subdivision using the mean phase of the pixel.

N	4	8	16	32
MSE (rad ²) 2×2	0.04289	0.006671	0.001119	0.0001850
MSE (rad ²) 4×4		0.01940	0.003369	0.0005728
MSE (rad ²) 8×8			0.007438	0.001383

Consider an $N \times N$ square aperture that is imaged onto a $2N \times 2N$ pixel detector. We then subdivide the aperture into a 2×2 array and image each subaperture onto an $N \times N$ pixel detector. Table 1 shows the mean-squared error introduced by the subdivision process, found by evaluating Eq. (32) for $D/r_0 = 1$. We repeat this process for a 4×4 and 8×8 subdivision. Other values of D/r_0 can be found by multiplying by $(D/r_0)^{5/3}$. It can be seen from Table 1 that the error in the wavefront estimate associated with subdividing the aperture decreases as the number of detector pixels increases. The reason for this is that the smaller the spacing between adjacent pixels in neighboring apertures, the greater the phase correlation between them and the easier it is to match the phase at the boundary. Better performance can thus be obtained by estimating the phase at more pixels in the aperture. Intuitively, we expect the effect of subdividing an aperture directly into a 4×4 array to produce a similar error to subdividing it into a 2×2 array, and then further subdividing each of the four subapertures into a 2×2 array. This repeated subdivision results in a mean-squared error of $0.001119 + (1/2)^{5/3} \times 0.006671 = 0.003220$, which is comparable to 0.003369 obtained by directly subdividing into a 4×4 array.

If the apertures are matched using the covariance of the phase at the center of each pixel instead of the mean, the error associated with the subdivision more than doubles (Table 2). This indicates that it is advantageous to estimate the average phase over the pixel rather than its phase at the center.

6 Cramér-Rao Lower Bound

The Cramér-Rao lower bound (CRLB) is the most commonly used lower bound for estimation problems.^{24,25} It predicts the minimum variance of any unbiased estimate of one or more parameters, given the statistics of the measurement noise. For nonlinear estimation problems, the CRLB is not the tightest lower bound and it can only be approached in practice at high signal-to-noise ratios. References for the CRLB are van Trees²⁶ and Kay.²⁷

For the problem at hand of estimating N^2 phase parameters, ϕ_0 , from $4N^2$ noisy intensity measurements, \mathbf{I} , the CRLB states that

Table 2 Mean-squared error due to aperture subdivision using the phase at the center of the pixel.

N	4	8	16	32
MSE (rad ²) 2×2	0.1359	0.01783	0.002646	0.0004266

$$\text{Var}[\hat{\phi}_k] \geq [J^{-1}]_{kk}, \tag{33}$$

where J is the Fisher matrix. The elements of the Fisher matrix J_{kl} consist of

$$J_{kl} = \int_{-\infty}^{\infty} \frac{\partial^2 \ln f(\mathbf{I}|\phi_0)}{\partial \phi_k \partial \phi_l} f(\mathbf{I}|\phi_0) d\mathbf{I} = -\text{E} \left[\frac{\partial^2 \ln f(\mathbf{I}|\phi_0)}{\partial \phi_k \partial \phi_l} \right], \tag{34}$$

where $\text{E}[\cdot]$ represents the expectation operator. For measurements with additive Gaussian noise of variance $\sigma_{\Delta I}^2$, the probability density function $f(\mathbf{I}|\phi_0)$ is

$$f(\mathbf{I}|\phi_0) = \prod \frac{1}{\sigma_{\Delta I} \sqrt{2\pi}} \exp \left[-\frac{(\mathbf{I} - \mathbf{I}_0)^2}{2\sigma_{\Delta I}^2} \right], \tag{35}$$

where \mathbf{I}_0 is the intensity corresponding to ϕ_0 . Hence,

$$\ln f(\mathbf{I}|\phi_0) = \sum \left(-\frac{(\mathbf{I} - \mathbf{I}_0)^2}{2\sigma_{\Delta I}^2} + \frac{1}{\sigma_{\Delta I} \sqrt{2\pi}} \right), \tag{36}$$

and

$$\begin{aligned} J_{kl} &= \text{E} \left[-\frac{\partial}{\partial \phi_l} \frac{\partial \ln f(\mathbf{I}|\phi_0)}{\partial \mathbf{I}_0} \frac{\partial \mathbf{I}_0}{\partial \phi_k} \right] \\ &= \text{E} \left[-\frac{\partial}{\partial \phi_l} \frac{\sum (\mathbf{I} - \mathbf{I}_0)}{\sigma_{\Delta I}^2} \frac{\partial \mathbf{I}_0}{\partial \phi_k} \right] \\ &= \left[\frac{\partial \mathbf{I}_0}{\partial \phi_l} \right]^T C_{\Delta I}^{-1} \left[\frac{\partial \mathbf{I}_0}{\partial \phi_k} \right]. \end{aligned} \tag{37}$$

Note that the expectation operator is no longer needed, as there is no dependency on the measured intensity \mathbf{I} . By making use of the relationship $\partial \mathbf{I}_0 = H \partial \phi$, this leads to

$$J = H^T C_{\Delta I}^{-1} H. \tag{38}$$

The CRLB for the sum of the phase errors in estimating ϕ_0 is

$$\text{Var}[\hat{\phi}_0] \geq \text{Trace}[(H^T C_{\Delta I}^{-1} H)^+], \tag{39}$$

where the trace of a matrix is the sum of the leading diagonal entries. The error in estimating ϕ_0 depends on H , which is itself a function of ϕ_0 .²⁸

7 Simulations

There is a limit imposed as to how well we can estimate the phase screen in the presence of noise. This limitation was quantified by using the LIA to estimate how much the solution to Eq. (18) varies from the true phase screen ϕ_0 .

Due to the high computational load of the LIA, the analysis was limited to 16×16 square phase screens inside a 32×32 array of zeros. To limit the amount of aliasing, we restricted the turbulence to a maximum of $D/r_0 = 4$. Random phase screens were generated using the method outlined in Sec. 5.3. The intensity distribution, \mathbf{I}_0 , corresponding to the phase screen was computed using Eq. (6).

Table 3 Mean-squared error for a single phase screen with uncorrelated image-independent additive Gaussian noise.

$S_{\Delta I}$	10^{-10}	10^{-9}	10^{-8}	10^{-7}	10^{-6}	10^{-5}
CRLB (rad ²)	2.534×10^{-6}	2.534×10^{-5}	2.534×10^{-4}	2.534×10^{-3}	2.534×10^{-2}	2.534×10^{-1}
LIA (rad ²)	2.58×10^{-6}	2.61×10^{-5}	2.32×10^{-4}	1.15×10^{-3}	0.371×10^{-2}	0.100×10^{-1}
1σ error (rad ²)	0.04×10^{-6}	0.05×10^{-5}	0.05×10^{-4}	0.02×10^{-3}	0.004×10^{-2}	0.001×10^{-1}

7.1 Comparison with the Cramér-Rao Lower Bound

We compared the phase error for the uncorrelated, image-independent, additive Gaussian noise case using the CRLB and LIA. A single 16×16 phase screen with Kolmogorov statistics and $D/r_0 = 1$ was generated, but here we did not use prior information about the phase. The resulting image was 32×32 . Hence the covariance of the noise is a $32^2 \times 32^2$ matrix with entries of $\sigma_{\Delta I}^2$ along the leading diagonal. In this case, $\sigma_{\Delta I}^2$ is equal to $S_{\Delta I}/32^2$, where $S_{\Delta I}$ is the total noise power and varied between 10^{-10} and 10^{-5} . The results displayed in Table 3 show that the CRLB falls within the 2σ uncertainty bounds of the LIA at low noise.

It was observed in the reconstructions that $E[\hat{\phi}] \neq \phi_0$, implying that the estimates produced by the LIA are biased, and this bias was statistically significant. As the noise level increases, the bias in the phase estimate using the LIA also increases. The bias occurs because there is an implicit constraint enforced in the LIA, namely that the intensity has to be positive. In addition, when the noise exceeds the signal at a given pixel, the approximation made in Eq. (12) no longer holds and the linear model relating the change in phase to the change in intensity is inadequate. This results in a lower error but introduces bias, and also the convergence of the LIA becomes very slow. The CRLB, on the other hand, is based on an unconstrained estimate of ϕ using the H matrix at ϕ_0 . Since the intensity estimate is constrained to be positive by Eq. (6), for appreciable values of noise H changes significantly during the iterations. Consequently, the CRLB is only applicable at low noise levels.

To see what happens when the turbulence is more severe, we multiplied each phase value on the phase screen by $(D/r_0)^{5/6}$. For $D/r_0 = 4$, the CRLB mean-squared error was 0.02731 at $S_{\Delta I} = 10^{-5}$, a value nearly ten times lower than for $D/r_0 = 1$. This occurs because the observed intensity I is spread over a larger number of pixels, thereby improving the conditioning of the inversion and lowering the theoretically achievable mean-squared error.

The CRLB can also be computed for image-dependent noise. Poisson noise was approximated by an independent

Gaussian random variable, with the variance set equal to the expected value of the signal. The noise covariance is that given by Eq. (17), where P is set to $S_{\Delta I}^{-1}$. $S_{\Delta I}$ was varied between 10^{-8} and 10^{-4} . From Table 4, we see once again that the CRLB is attained at low noise levels, but the LIA returns values lower than the CRLB when the noise is large. Again for $D/r_0 = 4$, the CRLB mean-squared error was 1.598×10^{-3} at $S_{\Delta I} = 10^{-5}$, about 25% lower than for $D/r_0 = 1$. This indicates that phase retrieval is inherently better posed when the turbulence is high, independent of the noise model.

7.2 Full Aperture

The noise-free intensity distribution was formed using the full aperture. Poisson noise was then added and the noisy intensity distribution was renormalized to give \mathbf{I} . In all the remaining simulations, prior information about the statistics of the phase and of the noise were used in the reconstructions. We also use actual Poisson noise rather than the Gaussian approximation, with the number of photons varying between 10^3 and 10^{10} . As the noise level increases, so does the number of iterations needed to recover the phase.

The log mean-squared error of the phase is plotted against the log of the number of photons in Fig. 1. All the graphs have 1σ error bars.

At high photon counts, there is a linear relationship between the logarithms of the squared error and the number of photons when $P > 10^7$. Note that the gradient of the curve at high photon levels is -1.00 , so the relationship is $E[(\Delta\phi)^2] \propto P^{-1}$. Although it is difficult to see from the graph, the error for $D/r_0 = 4$ is about 5% lower than for $D/r_0 = 1$ when $P \geq 10^6$, again reflecting the improved conditioning of the phase retrieval problem at high turbulence levels. This is expected, as the information about the phase is spread over many intensity pixels. When the photon noise is high, the prior becomes significant and reduces the error, especially for low turbulence levels, where the prior is strongest. In the limiting case when there are no photons, the best phase estimate is zero, and the error is simply the

Table 4 Mean-squared error for a single phase screen with uncorrelated image-dependent additive Gaussian noise.

$S_{\Delta I}$	10^{-8}	10^{-7}	10^{-6}	10^{-5}	10^{-4}
CRLB (rad ²)	2.099×10^{-6}	2.099×10^{-5}	2.099×10^{-4}	2.099×10^{-3}	2.099×10^{-2}
LIA (rad ²)	2.15×10^{-6}	2.15×10^{-5}	2.20×10^{-4}	1.82×10^{-3}	1.01×10^{-2}
1σ error (rad ²)	0.03×10^{-6}	0.03×10^{-5}	0.04×10^{-4}	0.05×10^{-3}	0.02×10^{-2}

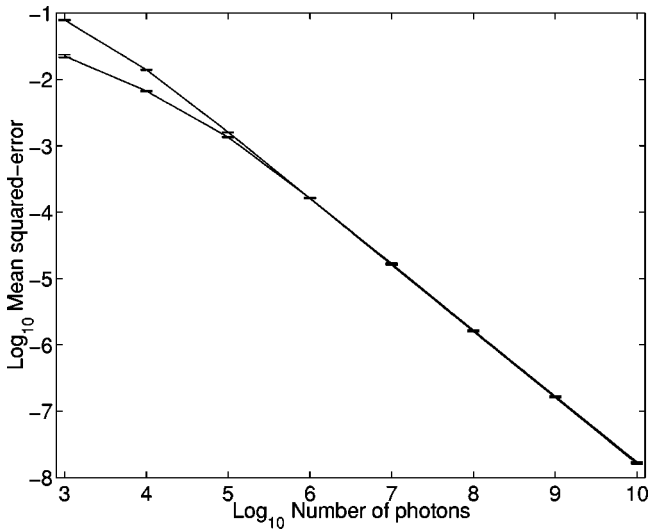


Fig. 1 Log mean-squared error for a full aperture. The graphs correspond to D/r_0 ratios of 4 (top) and 1 (bottom).

mean-squared wavefront. This is obtained by averaging the diagonal elements of the phase covariance matrix, and for a 16×16 phase screen, takes a value of $1.30(D/r_0)^{5/3} \text{rad}^2$.

7.3 Subdivided Aperture

The error obtained in phase retrieval by subdividing the aperture into a system of 2×2 subapertures is plotted in Fig. 2. The error due to the subdivision dominates for $P \geq 10^6$. The results at high photon counts agree well with the theory in Sec. 5.3, which predicts that the error due solely to the relative pistons of the subapertures is $0.001119(D/r_0)^{5/3}$. In all cases, the error is larger than for the full aperture.

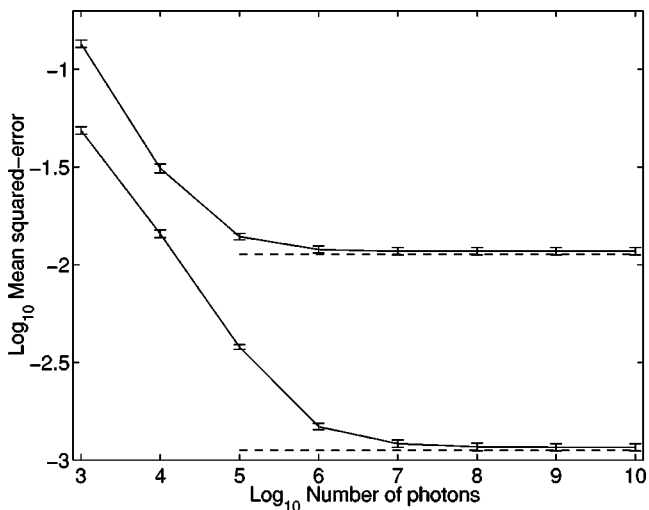


Fig. 2 Log mean-squared error for a 2×2 subdivided aperture and 16×16 pixels. The graphs correspond to D/r_0 ratios of 4 (top) and 1 (bottom). The dashed lines indicate the error due to subdivision alone (Table 1).

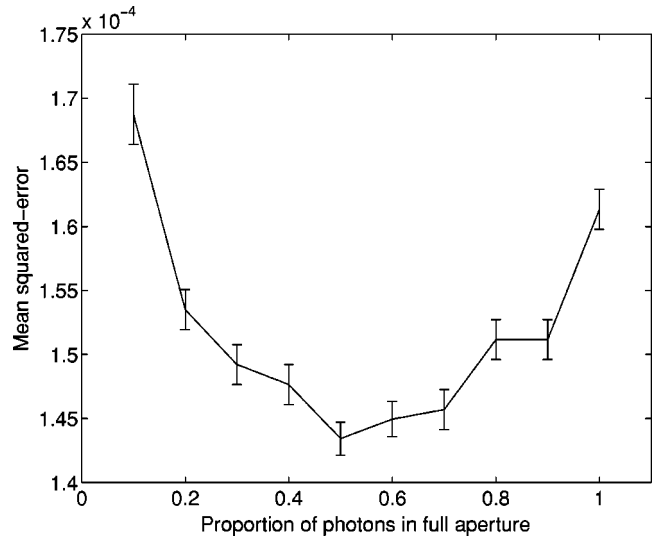


Fig. 3 Mean-squared error versus proportion of photons in single aperture for $D/r_0=4$ and $P=10^6$.

7.4 Full and Subdivided Aperture

The question arises as to whether there is any advantage in a hybrid system based on full and subdivided aperture data. The photons were initially split evenly between the two detectors, with the results being visually indistinguishable from Fig. 1. The results obtained are only a small improvement to the case where just a single aperture is used with a reduction in the error of up to 12% at high photon counts. However, for $D/r_0=1$ and $P=10^3$, the hybrid system with the 50:50 photon split had a higher error than having all the photons on the single aperture. The mean-squared error was again lower for $D/r_0=4$ than for $D/r_0=1$ when $P \geq 10^6$.

The proportion of photons at each detector was varied for $D/r_0=4$ and $P=10^6$, and the results are plotted in Fig. 3. It can be seen that the error is smallest when the photons are evenly shared between the two detectors, but the minimum is broad. However, when the number of photons is smaller, there is no gain to be made by taking the two images in the theoretical performance limit. It may, however, help avoid the phase reconstructor being trapped in a local minimum when the algorithm is started far from the solution.

8 Conclusion

A method to solve the phase retrieval problem using linear iterations near the solution has been developed. This method quantifies the asymptotic error in phase retrieval due to both the presence of noise and the effect of aperture subdivision.

It was found that prior information about the phase and noise improves the conditioning of the problem. When phase retrieval is performed on data from subdivided apertures, there is a loss of information about the relative piston terms of the subapertures, and this error has been quantified. There is a smaller wavefront error when estimating the phase from a full aperture than from a subdivided aperture. Using a combination of both sets of intensity measurements improves the performance when the number of photons is very high, but for moderate photon levels the limiting per-

formance is better for a single image using the unsubdivided aperture. In the absence of prior information, the phase retrieval problem is more well posed when the turbulence is strong. The effect is less significant in the presence of prior information, since the prior is stronger for weak turbulence.

Acknowledgments

The authors wish to thank the New Zealand government for financial aid in the form of a Bright Futures scholarship and a Marsden Fund research grant. Considerable thanks are due to anonymous reviewers for helpful suggestions.

References

1. J. Goodman, *Introduction to Fourier Optics*, pp. 59–100, McGraw-Hill, San Francisco (1968).
2. M. C. Roggemann and B. Welsh, *Imaging Through Turbulence*, pp. 182–183, CRC Press, Boca Raton, FL (1996).
3. R. G. Lane and R. Irwan, "Phase retrieval as a means of wavefront sensing," in *Proc. Intl. Conf. Image Process.*, B. R. Hunt and R. M. Gray, Eds., pp. 242–245, IEEE Signal Processing Society Press, Los Alamitos, CA (1997).
4. R. C. Cannon, "Global wave-front reconstruction using Shack-Hartmann sensors," *J. Opt. Soc. Am. A* **12**, 2031–2039 (1995).
5. M. C. Roggemann, T. J. Schulz, and C. W. Ngai, "Joint processing of Hartmann sensor and conventional image measurements to estimate large aberrations: theory and experimental results," *Appl. Opt.* **38**, 2249–2255 (1999).
6. T. Takahashi and H. Takajo, "Global wavefront reconstruction from its intensity distribution in the focal plane and Shack-Hartmann sensor images," *Opt. Eng.* **38**, 1960–1964 (1999).
7. F. Roddier, "The effect of atmospheric turbulence in optical astronomy," *Progress in Optics*, E. Wolf, Ed., pp. 283–376, North-Holland, Amsterdam (1981).
8. R. A. Gonsalves, "Phase retrieval from modulus data," *J. Opt. Soc. Am.* **66**, 961–964 (1976).
9. 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).
10. R. G. Lane, W. R. Fright, and R. H. T. Bates, "Direct phase retrieval," *IEEE Trans. Acoust., Speech, Signal Process.* **ASSP-35**, 520–526 (1987).
11. J. R. Fienup, "Phase retrieval algorithms: A comparison," *Appl. Opt.* **21**, 2758–2769 (1982).
12. R. Irwan and R. G. Lane, "Phase retrieval with prior information," *J. Opt. Soc. Am. A* **15**, 2302–2311 (1998).
13. W. J. Wilde, "Linear phase retrieval for wave-front sensing," *Opt. Lett.* **23**, 573–575 (1998).
14. Gilbert Strang, *Linear Algebra and Its Applications*, pp. 442–452, Harcourt Brace, FL (1988).
15. D. G. Luenberger, *Linear and Nonlinear Programming*, pp. 243–248, Addison-Wesley, Reading, MA (1984).
16. J. L. C. Sanz and T. S. Huang, "Polynomial system of equations and its applications to the study of the effect of noise on multidimensional Fourier transform phase retrieval from magnitude," *IEEE Trans. Acoust., Speech, Signal Process.* **ASSP-33**, 997–1004 (1985).
17. W. H. Southwell, "Wave-front estimation from wave-front slope measurements," *J. Opt. Soc. Am.* **70**, 998–1006 (1980).
18. N. F. Law and R. G. Lane, "Wavefront estimation at low light levels," *Opt. Commun.* **126**, 19–24 (1996).
19. E. P. Wallner, "Optimal wave-front correction using slope measurements," *J. Opt. Soc. Am.* **73**, 1771–1776 (1983).
20. M. A. van Dam and R. G. Lane, "Effect of aperture subdivision on wavefront sensing," *Proc. SPIE* **4125**, 53–64 (2000).
21. R. G. Lane, "Methods for maximum likelihood deconvolution," *J. Opt. Soc. Am. A* **13**, 1992–1998 (1996).
22. L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*, pp. 359–422, Addison-Wesley, Reading, MA (1991).
23. C. M. Harding, R. A. Johnston, and R. G. Lane, "Fast simulation of a Kolmogorov phase screen," *Appl. Opt.* **38**, 2161–2170 (1999).
24. J. N. Cederquist and C. C. Wackermann, "Phase-retrieval error: A lower bound," *J. Opt. Soc. Am. A* **4**, 1788–1792 (1987).
25. J. N. Cederquist, S. R. Robinson, D. Kryskowski, J. R. Fienup, and C. C. Wackermann, "Cramer-Rao lower bound on wavefront sensor error," *Opt. Eng.* **25**, 586–592 (1986).
26. H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part 1*, pp. 66–85, Wiley and Sons, New York (1968).
27. S. M. Kay, *Fundamentals of Statistical Signal Processing Estimation Theory*, pp. 27–82, Prentice-Hall, NJ (1993).
28. D. J. Lee, M. C. Roggemann, and B. M. Welsh, "Cramer-Rao analysis of phase-diverse wave-front sensing," *J. Opt. Soc. Am. A* **16**, 1005–1015 (1999).



Marcos A. van Dam received his BSc degree in physics and mathematics and the BE (Hons) degree in electrical and electronic engineering from the University of Canterbury, New Zealand, in 1999. In 2002, he expects to complete a PhD in electrical and electronic engineering at the same university. His research interests include adaptive optics, image processing, and electrical metrology.



Richard G. Lane received his PhD in electrical and electronic engineering from the University of Canterbury, New Zealand, in 1988. Since then he has worked at the University of Adelaide, Australia, Imperial College London, the University of Tasmania, Australia, and the Observatoire de Lyon. He is currently a Senior Lecturer at the University of Canterbury, New Zealand. His research interests include adaptive optics and image processing.