

WAVEFRONT RECONSTRUCTION METHODS FOR ADAPTIVE OPTICS SYSTEMS ON GROUND-BASED TELESCOPES

JOHNATHAN M. BARDSLEY^{†‡}

Abstract. The earth's atmosphere is not a perfect media through which to view objects in outer-space; turbulence in the atmospheric temperature distribution results in refractive index variations that interfere with the propagation of light. As a result, wavefronts are non-planar when they reach the ground. The deviation from planarity of a wavefront is known as phase error, and it is phase error that causes the refractive blurring of images. Adaptive optics systems seek to remove phase error from incoming wavefronts. In ground-based astronomy, an estimate of the phase error in a wavefront is typically obtained from wavefront gradient measurements collected by a Shack-Hartmann sensor. The estimate is then used to create a counter wavefront, e.g. using a deformable mirror, that (approximately) removes the phase error from the incoming wavefronts. The problem of reconstructing the phase error from Shack-Hartmann gradient measurements requires the solution of a large linear system whose form is defined by the configuration of the sensor. We derive this system and present both the regular least squares and minimum variance approaches to its solution. The most effective existing approaches are then presented alongside new computational methods, and comparisons are made.

Keywords: adaptive optics, wavefront reconstruction, minimum variance estimation.

1. Introduction. The standard mathematical model for image formation in ground-based astronomy is

$$(1.1) \quad d(x, y) = \int_{\mathbb{R}^2} k(x, y; \xi, \eta) f(\xi, \eta) d\xi d\eta.$$

Here f is the object being viewed; d is the image of f seen by the telescope; and k is the *point spread function* (PSF), which characterizes the blurring effects of the imaging system. In traditional approaches, the PSF characterizes the diffractive blur of the telescope as well as the refractive blur of the atmosphere. Adaptive optics systems, however, seek to remove the refractive effects of the atmosphere prior to image formation. If this is done exactly, so-called *diffraction limited imaging* is obtained, which is the goal of the astronomer.

The idea behind adaptive optics can be illustrate using the spatially invariant PSF model

$$(1.2) \quad k[\phi](x, y) = \left| \mathcal{F}^{-1} \left\{ P(x, y) e^{i\phi(x, y)} \right\} \right|^2,$$

which is obtained using techniques from Fourier optics [5]. Here $P(x, y)$ is the telescope's pupil indicator function, i.e. is 1 inside the pupil and 0 otherwise; and $\phi(x, y)$ denotes the phase error, or simply the phase, and is defined to be the deviation from planarity of the wavefront at the point (x, y) . Adaptive optics systems seek to remove the phase error ϕ from the incoming wavefronts. If done exactly, the resulting PSF then has the form

$$(1.3) \quad k[0](x, y) = \left| \mathcal{F}^{-1} \{ P(x, y) \} \right|^2,$$

[†]Department of Mathematical Sciences, University of Montana, Missoula, MT. Email: bardsleyj@mso.umt.edu.

[‡]This work was done during the author's visit to the University of Helsinki, Finland in 2006-07 under the University of Montana Faculty Exchange Program.

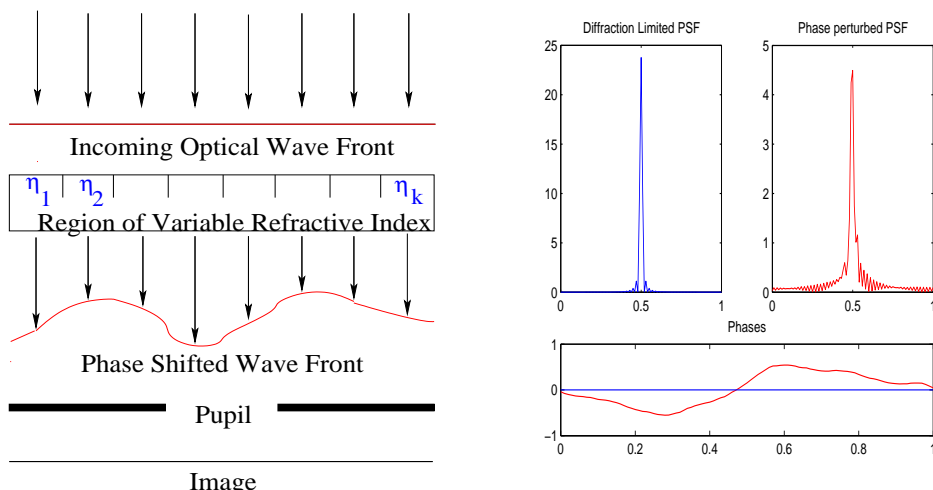


FIG. 1.1. *One-Dimensional Phase Error Schematic.* On the left, the effects of refractive index variations in the earth's atmosphere is illustrated. On the right, where both diffraction limited and phase perturbed PSFs are plotted, the effects of phase errors on the corresponding PSFs is demonstrated.

in which case the *diffraction limited image*

$$(1.4) \quad d_{\text{DL}}(x, y) \stackrel{\text{def}}{=} \int_{\mathbb{R}^2} k[0](x - \xi, y - \eta) f(\xi, \eta) d\xi d\eta,$$

is what is seen by the telescope.

Phase errors arise due to index of refraction variations in the atmosphere. A one-dimensional schematic of this process is given on the left in Figure 1.1. Since the refractive index - denoted by η_i in the schematic - determines the speed of propagation of the wavefront, variations in the refractive index result in wavefront perturbations, or phase errors. To see the effects of phase errors on the PSF, on the right-hand side in Figure 1.1 we plot two PSFs: one when phase errors characteristic of atmospheric turbulence are present, and one when the phase error is zero.

The phase can be estimated in a number of ways [2]. The most common approach in ground-based astronomy is to use the Shack-Hartmann sensor, which collects measurements of the gradient of the incoming wavefronts and then seeks to reconstruct the phase from those measurements. The phase estimate is then used to create a counter wavefront distortion ϕ_{DM} via the deformation of an optical component known as a deformable mirror (DM). If the PSF has the form (1.2) then the phase corrected PSF will have the form

$$(1.5) \quad k[\phi + \phi_{\text{DM}}](x, y) = \left| \mathcal{F}^{-1} \left\{ P(x, y) e^{i(\phi + \phi_{\text{DM}})(x, y)} \right\} \right|^2,$$

Ideally, the DM created counter wavefront satisfies $\phi_{\text{DM}} = -\phi$, so that the resulting PSF has diffraction limited form (1.3). In practice, however, an accurate approximation of ϕ suffices.

In this paper, our focus is on the problem of estimating the phase from measurements of the wavefront gradient. We assume that the gradient data \mathbf{g} is collected by a Shack-Hartmann sensor. The corresponding discrete phase ϕ will then satisfy the

stochastic linear equation

$$(1.6) \quad \mathbf{g} = \mathbf{\Gamma}\boldsymbol{\phi} + \mathbf{n},$$

where $\mathbf{\Gamma}$ is a discrete gradient matrix, whose form is determined by the configuration of the Shack-Hartmann sensor, and \mathbf{n} is the noise vector. Early and existing approaches to solving this problem involve minimizing the least squares function $\|\mathbf{\Gamma}\boldsymbol{\phi} - \mathbf{g}\|^2$. However, for large-scale adaptive optics systems, least squares solutions can be unstable, and the minimum variance estimator is preferred [6]. Minimum variance estimation is a Bayesian statistical approach in which a prior probability density is assumed on the phase. In our case, it can be accurately assumed that $\boldsymbol{\phi}$ is a realization of a Gaussian random vector with mean $\mathbf{0}$ and known covariance matrix \mathbf{C}_ϕ . This, together with (1.6), and the assumption that the noise vector \mathbf{n} is Gaussian with mean $\mathbf{0}$ and covariance matrix $\sigma^2\mathbf{I}$ yields a linear system of the form

$$(1.7) \quad (\mathbf{\Gamma}^T\mathbf{\Gamma} + \sigma^2\mathbf{C}_\phi^{-1})\boldsymbol{\phi} = \mathbf{\Gamma}^T\mathbf{g}.$$

The problem of efficiently solving (1.7), or, equivalently, of minimizing the penalized least squares function $\|\mathbf{\Gamma}\boldsymbol{\phi} - \mathbf{g}\|^2 + \sigma^2\boldsymbol{\phi}^T\mathbf{C}_\phi^{-1}\boldsymbol{\phi}$, has seen much recent attention. An efficient direct method for the solution of (1.7) using sparse matrix techniques is explored in [6]. However, the most computationally efficient approaches have involved the use of multigrid to precondition conjugate gradient iterations [9, 10]. In this paper, we introduce two new approaches for approximately solving (1.7). The first involves the use of a symmetric positive definite approximation of $\mathbf{\Gamma}^T\mathbf{\Gamma}$ as a preconditioner for conjugate gradient iterations. The second approach is completely different, and involves first computing the least squares solution of minimum norm $\boldsymbol{\phi}_{\text{MNLS}} = \mathbf{\Gamma}^\dagger\mathbf{g}$, where “ \dagger ” denotes pseudo-inverse. The minimum norm solution is then denoised and stabilized via the solution of a linear system motivated by (1.7).

The paper is organized as follows. In Section 2, we present the linear system that arises from the use of Shack-Hartmann sensor; we derive the minimum variance linear system (1.7); and we discuss approximations of the covariance matrix \mathbf{C}_ϕ . Computational methods are presented in Section 3 and tested in Section 4. We end with conclusions in Section 5.

2. Wavefront Reconstruction from Discrete Gradient Measurements.

In this section, we present the wavefront reconstruction problem that arises when the Shack-Hartmann wavefront sensor is used. The Shack-Hartmann sensor collects measurements of the gradient of incoming wavefronts of light emitted by the object being viewed by the telescope. It consists of an array of lenslets, each of which focuses the light within its aperture, and a charge coupled device (CCD) camera that records the position of the focal point of the light within each lenslet. A measurement of the average gradient of the wavefront over the lenslet aperture is then given by the position of the focal point. A schematic of the Shack-Hartmann sensor in one-dimension is given in Figure 2.1. A more detailed description with further references can be found in [2].

The standard computational methodology for reconstructing the phase from Shack-Hartmann gradient data was introduced by Fried in [4], where from the gradient measurements, which are assumed to be centered within each lenslet array, the value of the phase at the corners is computed. This is the so-called Fried geometry and is illustrated in Figure 2.2. The gradient of the phase at (x_i, y_j) , denoted $\nabla\phi(x_i, y_j) = (\phi_{x,i,j}, \phi_{y,i,j})$, can be approximated given values of ϕ at the half grid points via the

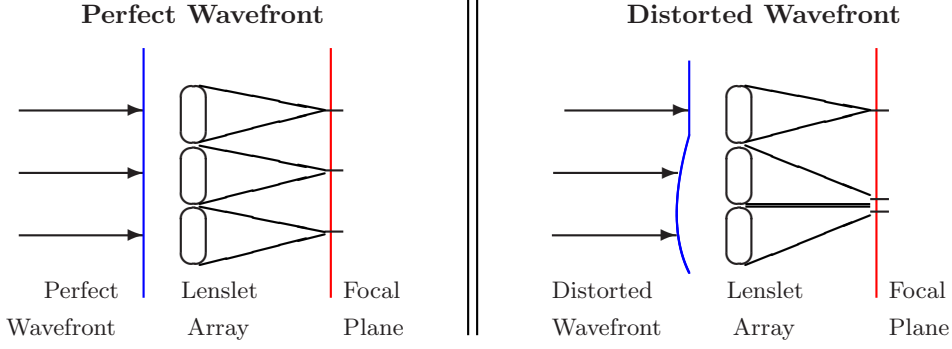


FIG. 2.1. *One-Dimensional Shack-Hartman Wavefront Sensor Schematic.* The position of the focal points determines the average derivative of the wavefront, and hence of the phase, over each lenslet aperture.

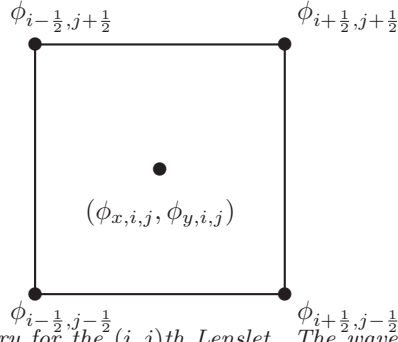


FIG. 2.2. *Fried Geometry for the (i, j) th Lenslet.* The wavefront sensor yields a measurement of the the gradient of the phase at (x_i, y_j) . The values of the phase at the half grid points, i.e. at the corners of the apertures, are then sought.

following two formulas:

$$(2.1) \quad \phi_{x,i,j} \approx \frac{1}{2} \left[(\phi_{i+\frac{1}{2},j-\frac{1}{2}} - \phi_{i-\frac{1}{2},j-\frac{1}{2}}) + (\phi_{i+\frac{1}{2},j+\frac{1}{2}} - \phi_{i-\frac{1}{2},j+\frac{1}{2}}) \right],$$

$$(2.2) \quad \phi_{y,i,j} \approx \frac{1}{2} \left[(\phi_{i-\frac{1}{2},j+\frac{1}{2}} - \phi_{i-\frac{1}{2},j-\frac{1}{2}}) + (\phi_{i+\frac{1}{2},j+\frac{1}{2}} - \phi_{i+\frac{1}{2},j-\frac{1}{2}}) \right],$$

where we have assumed a grid spacing $\Delta x = \Delta y = 1$. We note that in practice, the corners of the lenslet aperture correspond to the points on the deformable mirror of the adaptive optics system where the deformations are actuated.

By lexicographically ordering (column stacking) the $n \times n$ array of the values of ϕ at the half grid points, one obtains an $n^2 \times 1$ vector ϕ . The equations on the right-hand side of (2.1) and (2.2) can then be written in matrix-vector forms $\Gamma_x \phi$ and $\Gamma_y \phi$ respectively. If the $n \times n \times 2$ array of gradient measurements is also lexicographically ordered, and an $n^2 \times 2$ array \mathbf{g} results. The unknown phase can then be reconstructed by solving the stochastic linear system

$$(2.3) \quad \mathbf{g} = \Gamma \phi + \mathbf{n},$$

where $\Gamma = [\Gamma_x, \Gamma_y]^T$ and $\mathbf{n} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$. Throughout the paper, the notation $\mathbf{y} \sim N(\boldsymbol{\mu}, \mathbf{C})$ will mean that \mathbf{y} is a random draw from a Gaussian random vector with mean $\boldsymbol{\mu} \in \mathbb{R}^{n^2}$ and covariance matrix $\mathbf{C} \in \mathbb{R}^{n^2 \times n^2}$.

$$\begin{bmatrix} -1 & 0 & -1 \\ 0 & 4 & 0 \\ -1 & 0 & -1 \end{bmatrix} \quad \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix}$$

FIG. 2.3. Grid Representations of the Fried Laplacian (on the left) and the Hudgin Laplacian (on the right).

2.1. Least Squares Wavefront Reconstruction. For small scale wavefront reconstruction problems, the least squares solution of (2.3), given by minimizing

$$\min_{\phi} \|\mathbf{\Gamma}\phi - \mathbf{g}\|_2^2,$$

is known to provide excellent results [6]. One can equivalently solve the normal equations

$$(2.4) \quad \mathbf{\Gamma}^T \mathbf{\Gamma} \phi = \mathbf{\Gamma}^T \mathbf{g}.$$

The matrix $\mathbf{\Gamma}^T \mathbf{\Gamma}$ corresponds to a nonstandard discretization of the Laplacian operator Δ , with the grid representation given on the left in Figure 2.3 and homogeneous Neumann boundary conditions. We will call this the Fried discrete Laplacian. The standard discretization of the Laplacian has the grid representation given on the right in Figure 2.3. We note that this is the discrete Laplacian that results from what is known in the adaptive optics community as Hudgin geometry [12], and is therefore referred to as the Hudgin discrete Laplacian [18]. We will use this terminology in what follows.

Fried geometry yields more robust phase estimates than does Hudgin geometry [19]. However, the null-space of the Fried Laplacian is larger than that of the Hudgin Laplacian. In particular, it contains what is known as the waffle mode, which is the $n \times n$ array with entries

$$(2.5) \quad [\phi_{\text{WM}}]_{ij} = (-1)^{i+j}.$$

In the sequel, we will also use ϕ_{WM} to denote the corresponding $n^2 \times 1$ lexicographically ordered vector. We say that a vector ϕ contains waffle mode if

$$\phi^T \phi_{\text{WM}} \neq 0.$$

Waffle mode has been observed in operational adaptive optics systems [13], and hence, its presence in the null-space of the Fried Laplacian is not of only academic interest.

Before continuing, we prove an interesting relationship between ϕ_{WM} and the Hudgin discrete Laplacian with homogeneous Dirichlet, homogeneous Neumann, and periodic boundary conditions.

THEOREM 2.1. *If homogeneous Neumann or periodic boundary conditions are used to build the $n^2 \times n^2$ Hudgin discrete Laplacian \mathbf{L} , then the waffle mode ϕ_{WM} solves*

$$(2.6) \quad \max_{\phi} \frac{\phi^T \mathbf{L} \phi}{\|\phi\|^2}.$$

If homogeneous Dirichlet boundary conditions are used, then ϕ_{WM} converges to the solution of (2.6) as $n \rightarrow \infty$.

Proof. First, by Gerschgorin's circle theorem, the eigenvalues of \mathbf{L} satisfy $0 \leq \lambda \leq 8$ for each of the three types of boundary conditions. Thus $0 \leq \phi^T \mathbf{L} \phi / \|\phi\|^2 \leq 8$.

In the case of both periodic and homogeneous Neumann boundary conditions $\phi_{\text{WM}}^T \mathbf{L} \phi_{\text{WM}} / \|\phi_{\text{WM}}\|^2 = 8$, and hence ϕ_{WM} solves (2.6).

In the case of homogeneous Dirichlet boundary conditions, we note that \mathbf{L} differs from the discrete Laplacian with periodic boundary conditions in that it has 0 in place of -1 in $4n$ locations. A straightforward calculation together with the fact that $\|\phi_{\text{WM}}\|^2 = n^2$ then yields $\phi_{\text{WM}}^T \mathbf{L} \phi_{\text{WM}} / \|\phi_{\text{WM}}\|^2 = 8 - 4/n \rightarrow 8$, and hence ϕ_{WM} converges to the solution of (2.6) as $n \rightarrow \infty$. \square

We note that for large-scale problems $8 - 4/n \approx 8$, and hence, ϕ_{WM} maximizes, or nearly maximizes, (2.6) in all three cases. This suggests the use of regularization by the Laplacian to remove waffle mode and other high frequency errors in the phase estimates. As we will see in the next section, such an approach can be motivated statistically when the minimum variance approach is taken.

2.2. Minimum Variance Wavefront Reconstruction.. The preferred approach for stabilizing least squares phase estimation is to compute a minimum variance estimate for ϕ (c.f., [17]). Minimum variance estimation can be viewed as the analogue of least squares estimation in the Bayesian setting. As we will see, the resulting equations are similar. In the minimum variance framework, we assume that the phase satisfies

$$(2.7) \quad \phi \sim N(\mathbf{0}, \mathbf{C}_\phi),$$

where the covariance \mathbf{C}_ϕ is specified *a priori*. The minimum variance estimator can then be defined as follows.

DEFINITION 2.2. *The minimum variance linear estimator of ϕ from \mathbf{g} is given by*

$$\phi_{\text{MV}} = \hat{\mathbf{B}} \mathbf{g},$$

where

$$\hat{\mathbf{B}} = \arg \min_{\mathbf{B} \in \mathbb{R}^{n \times m}} E(\|\mathbf{B} \mathbf{g} - \phi\|^2).$$

Here E denotes the expected value function.

In our case, the minimum variance estimator has an elegant close form, which we state in the next theorem. Standard proofs of this theorem use notation from probability theory. Here we present a proof from a matrix analysis viewpoint.

THEOREM 2.3. *Let \mathbf{g} be data arising from model (2.3) with $\mathbf{n} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$ and $\phi \sim N(\mathbf{0}, \mathbf{C}_\phi)$. Suppose in addition that \mathbf{C}_ϕ is nonsingular and that \mathbf{n} and ϕ are independent. Then the minimum variance linear estimator of ϕ from \mathbf{g} is given by*

$$(2.8) \quad \phi_{\text{MV}}^\sigma = \left(\mathbf{\Gamma}^T \mathbf{\Gamma} + \sigma^2 \mathbf{C}_\phi^{-1} \right)^{-1} \mathbf{\Gamma}^T \mathbf{g}.$$

Proof. First, we note that

$$\begin{aligned} E(\|\mathbf{B} \mathbf{g} - \phi\|^2) &= \text{trace} \left(E[(\mathbf{B} \mathbf{g} - \phi)(\mathbf{B} \mathbf{g} - \phi)^T] \right), \\ &= \text{trace} \left(\mathbf{B} E[\mathbf{g} \mathbf{g}^T] \mathbf{B}^T - \mathbf{B} E[\mathbf{g} \phi^T] - E[\phi \mathbf{g}^T] \mathbf{B}^T + E[\phi \phi^T] \right). \end{aligned}$$

Then, using the distributive property of the trace function and the identity

$$\frac{d}{d\mathbf{B}}\text{trace}(\mathbf{B}^T\mathbf{C}) = \left(\frac{d}{d\mathbf{B}}\text{trace}(\mathbf{B}\mathbf{C})\right)^T = \mathbf{C},$$

we see that $dE(\|\mathbf{B}\mathbf{g} - \phi\|^2)/d\mathbf{B} = \mathbf{0}$ when

$$\hat{\mathbf{B}} = E[\phi\mathbf{g}^T]E[\mathbf{g}\mathbf{g}^T]^{-1}.$$

Now, since ϕ and \mathbf{n} are independent, (2.3) implies $E[\phi\mathbf{g}^T] = E[\phi\phi^T]\mathbf{\Gamma}^T$ and $E[\mathbf{g}\mathbf{g}^T] = (\mathbf{\Gamma}^T E[\phi\phi^T]\mathbf{\Gamma} + E[\mathbf{n}\mathbf{n}^T])$. Thus

$$\begin{aligned}\hat{\mathbf{B}}\mathbf{g} &= \mathbf{C}_\phi\mathbf{\Gamma}^T \left(\mathbf{\Gamma}^T\mathbf{C}_\phi\mathbf{\Gamma} + \sigma^2\mathbf{I}\right)^{-1} \mathbf{g}, \\ &= \left(\mathbf{\Gamma}^T\mathbf{\Gamma} + \sigma^2\mathbf{C}_\phi^{-1}\right)^{-1} \mathbf{\Gamma}^T\mathbf{g}.\end{aligned}$$

The last equality follows from straightforward algebraic manipulation. \square

Thus the minimum variance wavefront estimate can be obtained by solving the linear system

$$(2.9) \quad \left(\mathbf{\Gamma}^T\mathbf{\Gamma} + \sigma^2\mathbf{C}_\phi^{-1}\right)\phi = \mathbf{\Gamma}^T\mathbf{g},$$

or, equivalently, by minimizing the penalized least squares function

$$(2.10) \quad \|\mathbf{\Gamma}\phi - \mathbf{g}\|_2^2 + \sigma^2\phi^T\mathbf{C}_\phi^{-1}\phi.$$

Note, then, that the minimum variance estimator can be viewed as a Tikhonov estimator, with quadratic regularization term $\sigma^2\phi^T\mathbf{C}_\phi^{-1}\phi$.

2.3. Incorporating the Telescope's Pupil. For simplicity of implementation, computations are typically done on a square computational grid, even though the telescope pupil geometry is usually either circular or annular. This information is incorporated into the problem formulation using the pupil mask matrix \mathbf{M} defined by

$$[\mathbf{M}]_{ii} = \begin{cases} 1, & i \text{ inside the pupil,} \\ 0 & \text{otherwise.} \end{cases}$$

We then modify the linear stochastic model (2.3) as follows:

$$(2.11) \quad \mathbf{M}\mathbf{g} = \mathbf{M}\mathbf{\Gamma}\phi + \mathbf{n},$$

where $\mathbf{M}\mathbf{\Gamma} \stackrel{\text{def}}{=} [\mathbf{M}\mathbf{\Gamma}_x, \mathbf{M}\mathbf{\Gamma}_y]^T$. Following the minimum variance approach outlined above, the penalized least squares function (2.10) takes the form

$$(2.12) \quad \|\mathbf{M}(\mathbf{\Gamma}\phi - \mathbf{g})\|_2^2 + \sigma^2\phi^T\mathbf{C}_\phi^{-1}\phi,$$

and the linear system (2.9) is reexpressed as

$$(2.13) \quad \left(\mathbf{\Gamma}^T\mathbf{M}\mathbf{\Gamma} + \sigma^2\mathbf{C}_\phi^{-1}\right)\phi = \mathbf{\Gamma}^T\mathbf{M}\mathbf{g}.$$

The corresponding regular least squares normal equations are given by

$$(2.14) \quad \mathbf{\Gamma}^T\mathbf{M}\mathbf{\Gamma}\phi = \mathbf{\Gamma}^T\mathbf{M}\mathbf{g}.$$

2.4. Approximating the Phase Covariance. The phase covariance \mathbf{C}_ϕ must be chosen so that it is both physically realistic and amenable to fast computational methods. We call attention to the fact that for an actual adaptive optics system, algorithms for solving (2.9) must do so in real time.

Perhaps the most standard approximation for \mathbf{C}_ϕ is to assume that it has the form

$$(2.15) \quad \mathbf{C}_\phi^{\text{VK}} = \mathbf{F}^* \mathbf{\Lambda} \mathbf{F},$$

where \mathbf{F} is the 2D discrete Fourier transform matrix, “ $*$ ” denotes conjugate transpose, and the matrix $\mathbf{\Lambda}$ is diagonal with entries coming from the von Karman spatial power spectral density of the atmospheric refractive index fluctuations, with universal $-11/3$ inverse power law:

$$(2.16) \quad [\mathbf{\Lambda}]_{k,k} = \frac{c^2}{[|k|^2 + 1/L_0^2]^{11/6}}.$$

Here k denotes spatial frequency, L_0 is the turbulence outer-scale, which prevents an unphysically infinite amount of energy at the origin, and c is the phase screen strength (c.f. [15]).

However given the desire for real time computations for large-scale lenslet arrays on so-called extreme adaptive optics systems, a sparse covariance approximation is desirable. Such an approximation was introduced in [6]. Here the Kolmogorov power spectral density $c_0 |k|^{-11/3}$ (set $L_0 = \infty$ and $c = c_0$ in (2.16)) is approximated as follows:

$$c_0 |k|^{-11/3} \approx c_0 |k|^{-4}.$$

Then, using the fact that the biharmonic, or squared Laplacian, operator “ Δ^2 ” has spectrum $|k|^4$, the following discrete approximation of the covariance should be accurate:

$$(2.17) \quad \mathbf{C}_\phi^{\text{BH}} = c_0 \mathbf{L}^{-2},$$

where \mathbf{L} is a discrete Laplacian matrix. The constant c_0 in (2.17) can be physically interpreted as the strength of the turbulence, and is chosen in our simulations so that

$$(2.18) \quad E[\phi^T (\mathbf{C}_\phi^{\text{VK}})^{-1} \phi] = E[\phi^T (\mathbf{C}_\phi^{\text{BH}})^{-1} \phi]$$

holds.

In order to visually compare the two covariance approximations (2.15) and (2.17), we plot random draws from the zero mean Gaussian random vectors with these as the covariance matrices in Figure 2.4. They clearly exhibit similar characteristics.

Remark: Using covariance approximation (2.17) in (2.12) corresponds to regularization by the ℓ^2 norm of the Laplacian, i.e. to adding the regularization function $(\sigma^2/c_0) \|\mathbf{L}\phi\|^2$ to the regular least squares function $\|\mathbf{M}(\mathbf{\Gamma}\phi - \mathbf{g})\|^2$. Recall that this is what was suggested in the remarks following Theorem 2.1.

3. Numerical Methods. In the early papers of Fried [4] and Hudgin [12], the Gauss-Seidel iteration was used for numerically solving the normal equations (2.14). In [16], symmetric Gauss-Seidel is implemented. However, it is well-known that iterative methods such these converge slowly in practice. Such methods are effective,

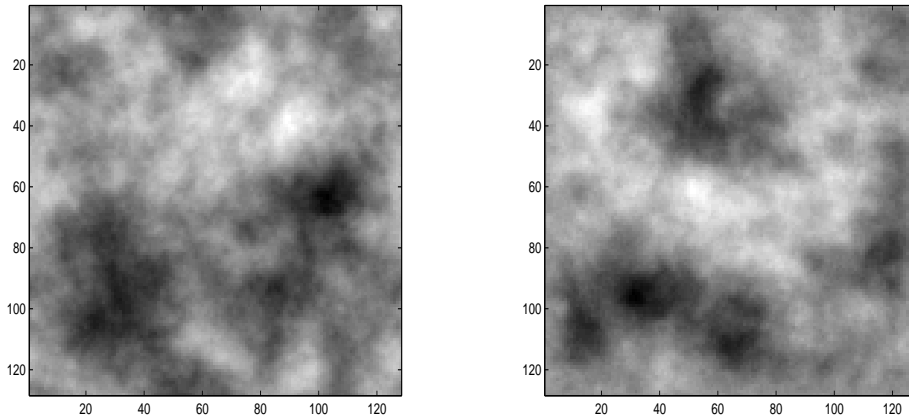


FIG. 2.4. Random draws from zero mean Gaussian random vectors with (on the left) von Karman covariance defined by (2.15) and (on the right) inverse of squared-Laplacian covariance defined in (2.17).

however, if used within a multi-grid framework or as preconditioners (given that they are symmetric) for conjugate gradient iterations. The use of multi-grid for solving (2.14) is explored in [1, 14].

Direct methods for (2.14) are made feasible by the fact that $\mathbf{\Gamma}^T \mathbf{M} \mathbf{\Gamma}$ is very sparse and is fixed for a specific telescope. In [11], the least squares solution of minimum norm solution is computed via the pseudo-inverse of $\mathbf{M} \mathbf{\Gamma}$, which we denote $(\mathbf{M} \mathbf{\Gamma})^\dagger$. The pseudo-inverse can be efficiently approximated using the Cholesky factorization. In particular, noting that

$$(\mathbf{M} \mathbf{\Gamma})^\dagger = \lim_{\epsilon \rightarrow 0^+} (\mathbf{\Gamma}^T \mathbf{M} \mathbf{\Gamma} + \epsilon \mathbf{I})^{-1} \mathbf{\Gamma}^T \mathbf{M},$$

one can compute a Cholesky factorization of $\mathbf{\Gamma}^T \mathbf{M} \mathbf{\Gamma} + \epsilon \mathbf{I}$ for ϵ small, e.g., the square root of machine epsilon ($\approx 10^{-8}$). Because it will be useful to us later, we give a detailed description of this approach now. After performing a reordering of indices using MATLAB's `symamd` function, $\mathbf{\Gamma}^T \mathbf{M} \mathbf{\Gamma} + \epsilon \mathbf{I}$ has the form

$$(3.1) \quad \tilde{\mathbf{A}} \stackrel{\text{def}}{=} \begin{bmatrix} \mathbf{A} + \epsilon \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \epsilon \mathbf{I} \end{bmatrix},$$

where \mathbf{A} is sparse and symmetric positive semi-definite. We can then compute the Cholesky factorization $\tilde{\mathbf{A}} = \mathbf{C}^T \mathbf{C}$. Assuming, without loss of generality, that $\tilde{\mathbf{A}} = \mathbf{\Gamma}^T \mathbf{M} \mathbf{\Gamma} + \epsilon \mathbf{I}$, the minimum norm least squares solution can be efficiently and accurately approximated via

$$(3.2) \quad \phi_{\text{MNLs}} = (\mathbf{M} \mathbf{\Gamma})^\dagger \mathbf{M} \mathbf{g} \approx \mathbf{C}^{-1} \mathbf{C}^{-T} \mathbf{\Gamma}^T \mathbf{M} \mathbf{g}.$$

Finally, since $\mathbf{\Gamma}^T \mathbf{M} \mathbf{\Gamma}$ depends only on the inherent structure of the telescope, the above Cholesky factorization can be computed offline, and hence, the cost of computing minimum norm least squares solutions using this approach is restricted to the computation of $\mathbf{\Gamma}^T \mathbf{M} \mathbf{g}$ and to the applications of \mathbf{C}^{-1} and \mathbf{C}^{-T} . Also, due to the presence of the pupil mask matrix \mathbf{M} and the use of a sparse reordering of indices,

the Cholesky factorization is very sparse, resulting in very efficient computations of ϕ_{MNLS} .

For adaptive optics systems with extremely large lenslet arrays - so called extreme adaptive optics - the minimum variance estimate ϕ_{MV}^σ obtained by solving (2.13) is preferable to ϕ_{MNLS} . An added difficulty arises, however, due to the presence of \mathbf{C}_ϕ . Until recently, the von Karman covariance approximation (2.15), (2.16) was standard. The fact that this is a full matrix made direct solutions of (2.13) infeasible. However, the sparse biharmonic approximation (2.17) presented in [6] allowed for a direct approach using a Cholesky factorization of

$$(3.3) \quad \mathbf{\Gamma}^T \mathbf{M} \mathbf{\Gamma} + (\sigma^2/c_0) \mathbf{L}^2,$$

where \mathbf{L} is a discretized Laplacian matrix.

3.1. Preconditioned Conjugate Gradient Methods. The preconditioned conjugate gradient method (PCG) is an iterative method for minimizing quadratic functions with symmetric positive semi-definite Hessian matrices [17], such as is the case for (2.12). The implementation of PCG requires the solution of a linear system of the form

$$(3.4) \quad \mathbf{P} \mathbf{z} = \mathbf{v}$$

at each iteration, where \mathbf{P} is the symmetric positive definite preconditioning matrix, or, simply, the preconditioner. For the resulting implementation of PCG to be efficient, solutions of (3.4) must be efficiently computable.

Thus far, the most computationally efficient approach for minimizing (2.12) is to let \mathbf{z} in (3.4) be what results following the application of one multigrid v-cycle [3] applied to the linear system

$$\left(\mathbf{\Gamma}^T \mathbf{M} \mathbf{\Gamma} + \sigma^2 \mathbf{C}_\phi^{-1} \right) \mathbf{z} = \mathbf{v}.$$

The corresponding preconditioning matrix \mathbf{P} is made symmetric by using either a symmetric smoother such as Jacobi or symmetric Gauss-Seidel, or by using Gauss-Seidel with forward substitution for the pre-smoothing iterations and an equal number of Gauss-Seidel iterations with backward substitution for the post-smoothing step. The resulting algorithm, which we denote MGPCG, was applied for the von Karman covariance approximation (2.15), (2.16) in [10] and for the biharmonic covariance approximation (2.17) in [9], with further analysis in [18]. The sparsity of the discrete biharmonic makes the latter implementation the more efficient of the two. The effectiveness of multigrid in this setting is not surprising when one considers that $\mathbf{\Gamma}^T \mathbf{M} \mathbf{\Gamma}$ is a discrete Laplacian matrix, and that \mathbf{C}_ϕ^{-1} either is, or is well-approximated by, a discrete biharmonic matrix. Multigrid is known to be very effective for solving linear systems involving both the discrete Laplacian and the discrete biharmonic matrices.

A preconditioner that has not to have been used for PCG applied to the problem of minimizing (2.12) is

$$(3.5) \quad \mathbf{P} \stackrel{\text{def}}{=} \begin{bmatrix} \mathbf{A} + \epsilon \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix},$$

where \mathbf{A} is as defined in (3.1). As in that case, we compute a Cholesky factorization $\mathbf{C}^T \mathbf{C}$ of \mathbf{P} . Since \mathbf{P} is noise independent, this can be done off-line. Thus the application of \mathbf{P}^{-1} requires only one forward and one backward substitution. We will call \mathbf{P} in (3.5) the least squares preconditioner, since in this case \mathbf{P} is the coefficient matrix for the regular least squares normal equations, and the resulting method LSPCG.

3.2. The Denoised Least Squares Method. In this subsection, we introduce a new approach for obtaining approximate solutions of (2.13). First, we denote the Hudgin discrete Laplacian with periodic boundary conditions by \mathbf{L} . Then, we have

$$(3.6) \quad \mathbf{L} = \mathbf{F}^* \text{diag}(\lambda_1, \dots, \lambda_{n^2}) \mathbf{F},$$

where $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_{n^2}$ are the eigenvalues of \mathbf{L} , and \mathbf{F} denotes the discrete Fourier transform matrix with \mathbf{F}^* its conjugate transpose. We note that then $\lambda_1 = 0$ corresponds to the constant eigenvector of \mathbf{L} , and $\lambda_{n^2} = 8$ corresponds to the waffle mode eigenvector ϕ_{WM} (recall Theorem 2.1).

Multiplying both sides of equation (2.13) by the pseudo-inverse of $\mathbf{\Gamma}^T \mathbf{M} \mathbf{\Gamma}$, we obtain

$$(3.7) \quad ((\mathbf{\Gamma}^T \mathbf{M} \mathbf{\Gamma})^\dagger \mathbf{\Gamma}^T \mathbf{M} \mathbf{\Gamma} + (\sigma^2/c_0)(\mathbf{\Gamma}^T \mathbf{M} \mathbf{\Gamma})^\dagger \mathbf{L}^2) \phi = \phi_{\text{MNLS}},$$

which can in turn be solved to obtain a smoothed, or denoised, approximation of ϕ_{MNLS} .

We ignore for the moment the pupil mask matrix \mathbf{M} and consider the matrix product $(\mathbf{\Gamma}^T \mathbf{\Gamma})^\dagger \mathbf{L}$. Recall that the Fried discrete Laplacian $\mathbf{\Gamma}^T \mathbf{\Gamma}$ has the grid representation given in Figure 2.2. Note that after a rotation of the computational grid by $\pi/4$ radians, the grid representation of $\mathbf{\Gamma}^T \mathbf{\Gamma}$ will match that of a Hudgin discrete Laplacian, but with a grid spacing that is larger by a factor of $\sqrt{2}$. Taking this into account, we approximate the Fried discrete Laplacian as follows:

$$(3.8) \quad \mathbf{\Gamma}^T \mathbf{\Gamma} \approx \mathbf{F}^* \text{diag}(\lambda_1, \dots, \lambda_r, 0, \dots, 0) \mathbf{F},$$

where the λ_i 's are as in (3.6), and r chosen so that

$$(3.9) \quad \lambda_r < 8 - \frac{1}{4\sqrt{2}} \leq \lambda_{r+1}.$$

Note that ϕ_{WM} has a period of 4 on the computational grid and of $4\sqrt{2}$ on the rotated grid, hence our choice of truncation rule. Furthermore, our own computational experiments indicate that (3.9) is optimal in the sense that it minimizes the error in the phase reconstructions that result when the method we now present is used.

From (3.8), we know that within the telescope's pupil,

$$\mathbf{\Gamma}^T \mathbf{M} \mathbf{\Gamma} \approx \mathbf{F}^* \text{diag}(\lambda_1, \dots, \lambda_r, 0, \dots, 0) \mathbf{F}.$$

Hence,

$$(3.10) \quad (\mathbf{\Gamma}^T \mathbf{M} \mathbf{\Gamma})^\dagger \mathbf{\Gamma}^T \mathbf{M} \mathbf{\Gamma} \approx \mathbf{F}^* \text{diag}(1, \dots, 1, 0, \dots, 0) \mathbf{F},$$

$$(3.11) \quad (\mathbf{\Gamma}^T \mathbf{M} \mathbf{\Gamma})^\dagger \mathbf{L}^2 \approx \mathbf{F}^* \text{diag}(\lambda_1, \dots, \lambda_r, 0, \dots, 0) \mathbf{F}.$$

This leads to the following approximation of (3.7):

$$(3.12) \quad \mathbf{F}^* \mathbf{D} \mathbf{F} \phi = \phi_{\text{MNLS}},$$

where \mathbf{D} is diagonal with elements

$$(3.13) \quad [\mathbf{D}]_{ii} = \begin{cases} 1 + (\sigma^2/c_0)\lambda_i & \lambda_i < r, \\ 0 & \lambda_i \geq r. \end{cases}$$

The denoised minimum norm least squares solution can then be efficiently computed via

$$(3.14) \quad \phi_{\text{DMNLS}} = \mathbf{F}^* \mathbf{D}^\dagger \mathbf{F} \phi_{\text{MNLS}},$$

where \mathbf{D}^\dagger is the pseudo-inverse of \mathbf{D} .

As we will see, this approach is effective in practice. However, it also has the benefit of being simple to implement and very computationally efficient. Furthermore, it can be easily incorporated into adaptive optics systems that compute regular least squares solutions other than ϕ_{MNLS} . In particular, a general least squares solution ϕ_{LS} can be denoised via

$$(3.15) \quad \phi_{\text{DLS}} = \mathbf{F}^* \mathbf{D}^\dagger \mathbf{F} \phi_{\text{LS}}.$$

In the next theorem, we show that ϕ_{DLS} will not contain waffle mode even if ϕ_{LS} does. This suggests that (3.15) should be considered as a method for removing waffle mode from least squares solutions in operational adaptive optics systems [13].

We end the section with a proof that the least squares solution of minimum norm, the minimum variance solution, and the denoised least squares solution do not contain waffle mode.

THEOREM 3.1. *The least squares solution of minimum norm $\phi_{\text{MNLS}} = (\mathbf{M}\mathbf{\Gamma})^\dagger \mathbf{M}\mathbf{g}$, the minimum variance solution ϕ_{MV}^σ defined in (2.8), and the DLS solution obtained by computing (3.15) for any least squares solution ϕ_{LS} do not contain the waffle mode.*

Proof. The null-space of $\mathbf{M}\mathbf{\Gamma}$ contains ϕ_{WM} . The result for ϕ_{MNLS} then follows from the fact that the range of $(\mathbf{M}\mathbf{\Gamma})^\dagger$ and the null-space $\mathbf{M}\mathbf{\Gamma}$ only trivially intersect.

For the minimum variance solution, we note that $\phi_{\text{WM}}^T \mathbf{C}_\phi^{-1} \phi_{\text{WM}} > 0$ for either choice of \mathbf{C}_ϕ . Noting that $\mathbf{M}\mathbf{\Gamma} \phi_{\text{WM}} = \mathbf{0}$, it follows immediately that the minimizer of (2.12) does not contain waffle mode.

From (3.15), (3.13), (3.9) and the fact that ϕ_{WM} is an eigenvector of \mathbf{L} with eigenvalue 8, it follows that $\mathbf{F}^* \mathbf{D}^\dagger \mathbf{F} \phi_{\text{WM}} = \mathbf{0}$. Thus ϕ_{DLS} will not contain waffle mode. \square

4. Numerical Experiments. We now test the effectiveness of the above approaches on simulated Shack-Hartman sensor data. We simulate a phase profile by taking a random draw from the Gaussian random vector $N(\mathbf{0}, \mathbf{C}_\phi^{\mathbf{KV}})$ with physically realistic values for the parameters in (2.16). The phase screen used for our simulations is the 128×128 array plotted on the left in Figure 2.4. The noisy gradient data shown in Figure 4.1 was obtained using (2.11) with \mathbf{n} an i.i.d Gaussian random vector with variance chosen so that the signal-to-noise ratio is 20, which is what was used in [18].

In our first comparison, we apply preconditioned conjugate gradient (PCG) to problem (2.12) with covariance (2.17) and c_0 given by (2.18). Convergence results are plotted in Figure 4.2, where standard conjugate gradient (CG), multigrid preconditioned PCG (MGPCG), and PCG with least squares preconditioner (3.5) (LSPCG) are compared. Our implementation of MGPCG used two iterations of Gauss-Seidel with forward substitution for the pre-smoothing iterations and two iterations of Gauss-Seidel with backward substitution for the post-smoothing iterations. The grid transfer operator used full-weighting with homogeneous Dirichlet boundary conditions for the restriction and a constant times its transpose for interpolation. See [3] for details on

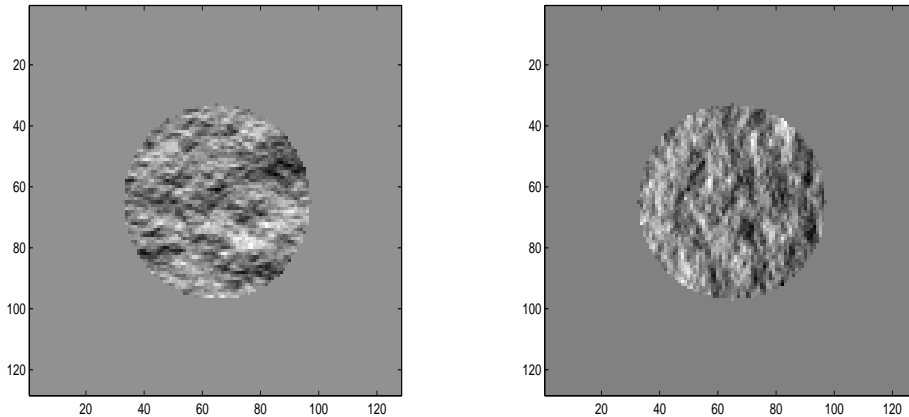


FIG. 4.1. *Noisy Gradient Data. The x and y components of the gradient data computed via (2.3) are seen on the left and right respectively.*

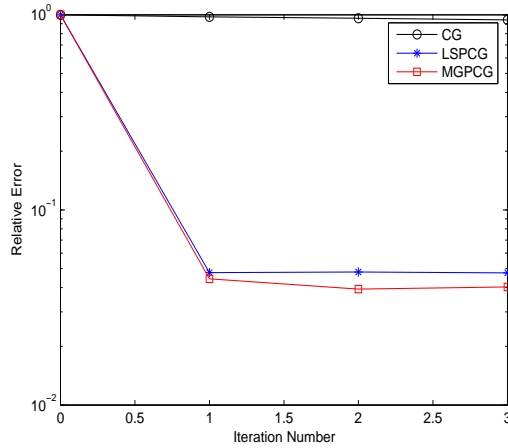


FIG. 4.2. *Relative Error versus PCG iterations, Signal-To-Noise Ratio = 20.*

this implementation. Since it is accuracy in the approximation of the true phase ϕ_{true} that we are concerned with, we plot the relative error

$$\frac{\|\phi_k - \phi_{\text{true}}\|}{\|\phi_{\text{true}}\|},$$

where k denotes PCG iterations. For this problem, MGPCG is the most effective at minimizing the relative error. In fact, tests at a number of noise levels indicate that LSPCG does not noticeably reduce the relative error after the first PCG iteration. However, the reconstruction obtained after the first iteration of LSPCG is very similar to that obtained by MGPCG. This can be seen in Figure 4.3, where the true phase and the reconstructions obtained using a single iteration of LSPCG and three iterations of MGPCG are given. Furthermore, analyzing CPU times over a number of runs

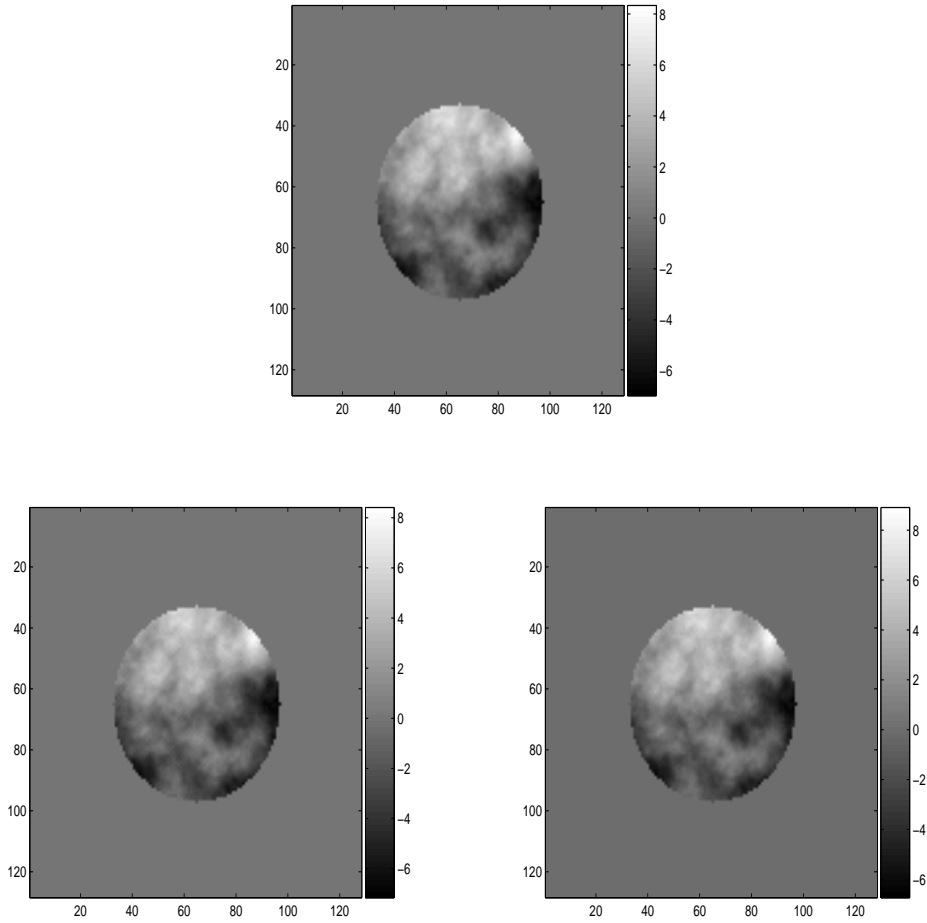


FIG. 4.3. *True phase on top, reconstruction obtained with one iteration of LSPCG in the lower left; and reconstruction obtained with three iterations of MGPCG iterations on the lower right.*

for this problem, LSPCG appears to be about 7-10 times faster per PCG iteration than MGPCG. Recall that multiplication by the inverse of the preconditioner (3.5) is computed at a cost of only two sparse linear system backsolves.

We now analyze the denoised least squares (DLS) approach presented above. In particular, we compute ϕ_{DMNLS} via (3.14), (3.2). In order to compare our results with those obtained by MGPCG and LSPCG, we list the relative errors obtained by each method for gradient data generated with signal-to-noise ratios of 50, 20, 10 and 5 in Table 4.1. The reconstructions obtained from one and two MGPCG iterations and one LSPCG iteration are included. We note that more iterations of either method results in, effectively, no reduction in relative error. In the last row of Table 4.1, we give the average CPU time for each method. These values are given only to provide a general notion of the computational efficiency of the respective methods. It appears however, that while two iterations of MGPCG yields the smallest relative error for all signal-to-noise ratios, the computation of ϕ_{DMNLS} yields nearly a two orders of

SNR	MGPCG 1	MGPCG 2	LSPCG	GDLS
50	0.0316	0.0250	0.0294	0.0260
20	0.0443	0.0393	0.0476	0.0415
10	0.0593	0.0550	0.0698	0.0596
5	0.0805	0.0766	0.0953	0.0854
CPU	0.77 s	1.16 s	0.10 s	0.016 s

TABLE 4.1

Relative error table with the signal-to-noise ratio in the first column and CPU times for each method in the last row.

magnitude improvement in computation time. Furthermore, ϕ_{DMNLS} has a lower relative error than does the LSPCG solution at all noise levels. Given the similarity of the LSPCG and MGPCG reconstructions shown in Figure 4.3, this suggests that GDLS is an approach that deserves attention.

5. Conclusions. We have presented a detailed discussion of the problem of wavefront, or phase, reconstruction from Shack-Hartmann wavefront gradient data. This included a derivation of a discrete, stochastic linear system relating the gradient measurements to the underlying discrete phase ϕ ; a derivation of the minimum variance estimator for ϕ given the prior probability density $\phi \sim N(\mathbf{0}, \mathbf{C}_\phi)$; and a discussion of estimates for the covariance \mathbf{C}_ϕ .

Computational methods for the phase reconstruction problem were then presented. First, an efficient method for computing an accurate approximation of the minimum norm least squares solution was given; it used a Cholesky factorization with sparse reordering. Then the current gold standard for accuracy and efficiency, the multigrid preconditioned conjugate gradient method (MGPCG), was compared with two new approaches presented in this paper: the so-called least squares preconditioned conjugate gradient method (LSPCG), and the denoised least squares method (DLS). The results indicate that though MGPCG yields lower values for the relative error error, the methods introduced here, and in particular the DLS method, are more efficient and yield comparable results. The implementation of DLS used in our comparisons utilized the efficient method for computing the minimum norm least squares solution ϕ_{MNLS} mentioned above, at a cost of two sparse linear system backsolves. The denoised phase estimate ϕ_{DMNLS} was then computed at a computational cost of only two discrete Fourier transforms, making DLS a very efficient approach.

We also presented results relating what is known in the adaptive optics community as waffle mode to standard discretizations of the Laplacian operator, and showed that the minimum norm least squares solution, the minimum variance solution, and the denoised least squares solutions do not contain waffle mode

6. Acknowledgements. The author would like to acknowledge the help that he received from Curt Vogel in the form of discussion and for the MATLAB code that was used for data generation in our experiments. Also, this work would not have been possible without the support of the University of Montana through its International Faculty Exchange Program and the University of Helsinki - the host university for the exchange.

7. References.

REFERENCES

- [1] K. L. Baker, *Least-Squares wave-front reconstruction of Shack-Hartmann sensors and shearing interferometers using multigrid techniques*, Review of Scientific Instruments, 76, 2005, pp. 053502.1-6.
- [2] Jacques M. Beckers, *Adaptive Optics for Astronomy: Principles, Performance, and Applications*, Annu. Rev. Astron. Astrophys., 1993, 31, pp. 13-62.
- [3] W. L. Briggs, V. E. Henson, and S. F. McCormick, *A Multigrid Tutorial*, 2nd ed., SIAM 2000.
- [4] David L. Fried, *Least-squares fitting a wave-front distortion estimate to an array of phase-difference measurements*, J. Opt. Soc. Am., 67(3), 1977.
- [5] Joseph W. Goodman, *Introduction to Fourier Optics*, McGraw-Hill, 1996.
- [6] Brent L. Ellerbroek, *Efficient computation of minimum-variance wave-front reconstructors with sparse matrix techniques*, J. Opt. Soc. Am. A, 19(9), 2002.
- [7] Lawrence Evans, *Partial Differential Equations*, American Mathematical Society, 1998.
- [8] R. C. Flicker, *Efficient first-order performance estimation for high-order adaptive optics systems*, Astronomy and Astrophysics, 405, 2003, pp. 1177-1189.
- [9] Luc Gilles, *Order-N sparse minimum-variance open-loop reconstructor for extreme adaptive optics*, Optics Letters, 28(20), 2003, pp. 1927-1929.
- [10] Luc Gilles, C.R. Vogel, and Brent Ellerbroek, *A Multigrid Preconditioned Conjugate Gradient Method for Large Scale Wavefront Reconstruction*, Journal of the Optical Society of America A, 19 (2002), pp. 1817-1822.
- [11] Jan Herrmann, *Least-squares wave front errors of minimum norm*, J. Opt. Soc. Am., 70(1), 1980.
- [12] Richard H. Hudgin, *Wave-front reconstruction for compensated imaging*, J. Opt. Soc. Am., 67(3), 1977.
- [13] R. B. Makidon, A. Sivaramakrishnan, M. D. Perrin, L. C. Roberts, B. R. Oppenheimer, R. Soummer, J. Graham, *An Analysis of Fundamental Waffle Mode in Early AEOS Adaptive Optics Images*, The Publications of the Astronomical Society of the Pacific, 117(834), 2005, pp. 831-846.
- [14] Mark D. Pritt, *Phase Unwrapping by Means of Multigrid Techniques for Interferometric SAR*, IEEE Trans. on Geoscience and Remote Sensing, 34(3), 1996, pp. 728-738.
- [15] M. Roggemann and B. Welsh, *Imaging Through Turbulence*, CRC Press, 1996.
- [16] W. H. Southwell, *Wave-front estimation from wave-front slope measurements*, Journal of the Optical Society of America, 70(8), 1980, pp. 998-1006.
- [17] C. R. Vogel, *Computational Methods for Inverse Problems*, SIAM, Philadelphia, 2002.
- [18] C. R. Vogel and Q. Yang, *Multigrid algorithm for least-squares wavefront reconstruction*, Applied Optics, 45(4), 2006, pp. 705-715.
- [19] B. M. Welsh, B. L. Ellerbroek, M. C. Roggemann, and T. L. Pennington, *Fundamental performance comparison of a Hartmann and a shearing interferometer wave-front sensor*, Applied Optics, 34(21), 1995, pp. 4186-4195.