

# Technical Report

TR-2010-023

**Structured linear algebra problems in adaptive optics imaging**

by

Jonathan Bardsley, Sarah Knepper, James G Nagy

**MATHEMATICS AND COMPUTER SCIENCE**

**EMORY UNIVERSITY**

# Structured Linear Algebra Problems in Adaptive Optics Imaging

Johnathan M. Bardsley\*, Sarah Knepper, and James Nagy†

## Abstract

A main problem in adaptive optics is to reconstruct the phase spectrum given noisy phase differences. We present an efficient approach to solve the least-squares minimization problem resulting from this reconstruction, using either a truncated singular value decomposition (TSVD)-type or a Tikhonov-type regularization. Both of these approaches make use of Kronecker products and the generalized singular value decomposition. The TSVD-type regularization operates as a direct method whereas the Tikhonov-type regularization uses a preconditioned conjugate gradient type iterative algorithm to achieve fast convergence.

## 1 Introduction

When viewing objects in space using a ground-based telescope, the images are blurred by the atmosphere. To remove the blur, adaptive optics systems solve two inverse problems. First, quantitative information about the blur must be determined; second, this quantitative information is used to remove the blur by a combination of telescope mirror adjustments and computational postprocessing of the images. All of these computations and adjustments must be performed in real time. This paper focuses on the first problem, namely that of determining quantitative information about the blur.

Using the classic model of image formation relating an object  $f$  and a blurring kernel  $k$  to an image  $g$ , we can write:

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

The kernel is assumed to be spatially invariant; that is, we assume  $k(x, y; \xi, \eta) = k(x - \xi, y - \eta)$ . The Fourier optics model for  $k$  [10, 18] allows us to express it as a function

---

\*Research supported by the National Science Foundation under grant DMS-0915107.

†Research supported by the National Science Foundation under grant DMS-0811031, and the Air Force Office of Scientific Research under grant FA9550-09-1-0487.

of the phase error  $\phi$  of incoming wavefronts of light:

$$k(s, t) = \left| \mathcal{F}^{-1} \{ \mathcal{A}(s, t) e^{i\phi(s, t)} \} \right|^2.$$

Here  $\mathcal{F}^{-1}$  is the inverse Fourier transform,  $\mathcal{A}(s, t)$  is the mirror aperture (a characteristic) function,  $i = \sqrt{-1}$ , and  $\phi$  is the phase error, which is the deviation from planarity of the wavefront that reaches the telescope mirror.

The goal of the adaptive optics system on a telescope is to remove the effects of the phase error  $\phi$  on the kernel  $k$ . This is accomplished by first computing an estimate  $\hat{\phi}$  of  $\phi$  and then by constructing a counter wavefront  $-\hat{\phi}$  via a deformable mirror. The kernel  $k$  is then modified as follows:

$$\hat{k}(s, t) = \left| \mathcal{F}^{-1} \{ \mathcal{A}(s, t) e^{i(\phi(s, t) - \hat{\phi}(s, t))} \} \right|^2.$$

If  $\phi = \hat{\phi}$ ,  $\hat{k}$  is said to have diffraction limited form, where the resolution is dependent only on the telescope.

To obtain an estimate of  $\phi$ , a wave front sensor in the telescope collects discrete, noisy measurements of the gradient of incoming wavefronts of light emitted by an object. Since the gradient of the incoming wavefronts equals the gradient of the phase error  $\phi$  of those wavefronts, the data noise model takes the form

$$\begin{bmatrix} \mathbf{s}_x \\ \mathbf{s}_y \end{bmatrix} = \begin{bmatrix} P_x \\ P_y \end{bmatrix} \phi + \epsilon, \quad (1)$$

where  $\mathbf{s}_x$  and  $\mathbf{s}_y$  are discrete, noisy measurements of the horizontal and vertical derivatives of  $\phi$ ;  $P_x$  and  $P_y$  are discrete, horizontal and vertical derivative operators; and  $\epsilon$  represents independent and identically distributed (iid) Gaussian noise. This linear model can then be solved to compute an estimate of  $\phi$ .

Model (1) requires some assumptions that are sufficiently accurate, but not exact. For example, the mapping from the phase  $\phi$  to the gradient measurements contains nonlinearities [3]. Here we model it as linear. Additionally, the assumption of iid Gaussian noise, while common in the literature (see e.g., [3, 4, 5, 8]), in particular for simulations, is not precise as it ignores photon noise.

The operators  $P_x$  and  $P_y$  are determined by the wavefront sensor geometry that is used. Common wavefront sensor geometries discussed in the literature are those of Hudgin [11] and Fried [7]. In this paper, we focus on Fried geometry, both because it is more commonly used in operational adaptive optics systems, and because it is more interesting mathematically. We also assume a rectangular pupil for the telescope. This simplifies the mathematics, however the approach set forth here can be extended to circular or annular pupils in a preconditioned iterative framework, as in [16, 17]. This is saved for a later work.

We propose to solve the phase reconstruction problem (1) using a preconditioned iterative approach. The preconditioner is formed by using the generalized singular value

decomposition (GSVD) and Kronecker products to exploit the structure of the coefficient matrix in (1). The preconditioner also suggests a very efficient direct solver using TSVD-type regularization. Though this approach gives good results and is quite efficient, the preconditioned iterative scheme, which uses Tikhonov-type regularization, is preferred by the application people, and in this case statistical information for choosing a regularization parameter can be obtained from the data; this point is discussed further in Section 4. Real-time solutions may be obtained using either method.

The problem of reconstructing the shape of an object from measurements of its gradient appears in a variety of applications (see e.g., [6] and the references therein), while the more specific problem of wavefront sensing appears not only in adaptive optics, but also in phase imaging and visual optics [1].

Section 2 describes some of the relevant mathematical background information, including information on GSVD and Kronecker products, as well as the matrices that come from Hudgin or Fried geometries. Section 3 describes the direct TSVD-type regularization approach to solving the reconstruction problem. The preconditioned iterative approach is given in Section 4. Results about the approximation quality of the preconditioner are given in Section 5, while overall numerical results are given in Section 6.

## 2 Mathematical Background

We make significant use of various properties of the generalized singular value decomposition (GSVD) [9] and Kronecker products [13] in our algorithm. Thus, in this section we briefly describe a few of the relevant properties, as well as the matrices that are formed from both the Hudgin and Fried geometries.

### 2.1 Generalized Singular Value Decomposition

The generalized singular value decomposition (GSVD) of two matrices  $B$  and  $C$  is given as follows:

$$B = U\Sigma X^T, \quad C = V\Delta X^T \quad (2)$$

where

- $U$  and  $V$  are orthogonal matrices
- $X$  is nonsingular
- $\Sigma$  and  $\Delta$  are nonnegative diagonal matrices, though the entries may be on a super-diagonal

## 2.2 Kronecker Products

If  $B \in \mathfrak{R}^{m \times n}$  and  $C$  is another matrix, then the Kronecker product of  $B$  and  $C$  produces the block structured matrix

$$B \otimes C = \begin{bmatrix} b_{11}C & \cdots & b_{1n}C \\ \vdots & & \vdots \\ b_{m1}C & \cdots & b_{mn}C \end{bmatrix},$$

where  $b_{ij}$  is the  $ij$ th element of  $B$ . Assume that  $B = U_B \Sigma_B V_B^T$  is some decomposition for  $B$  (and similarly for  $C$ ). Kronecker products have a plethora of convenient properties, many of which are detailed in Van Loan's survey [13]. We present here but a few of these:

$$(B \otimes C)^T = B^T \otimes C^T, \quad (3)$$

$$(B \otimes C)^{-1} = B^{-1} \otimes C^{-1}, \quad (4)$$

$$B \otimes C = (U_B \Sigma_B V_B^T) \otimes (U_C \Sigma_C V_C^T) = (U_B \otimes U_C)(\Sigma_B \otimes \Sigma_C)(V_B \otimes V_C)^T. \quad (5)$$

One additional property we need relates matrix-matrix multiplication with Kronecker products. As is common, we define the  $\text{vec}(X)$  operation as stacking the columns of a matrix  $X$ ; that is, if  $X \in \mathfrak{R}^{m \times n}$ , then  $\text{vec}(X)$  is an  $nm \times 1$  vector. Given matrices  $C, X, B$  such that the product  $CXB^T$  is defined, then

$$Y = CXB^T \Leftrightarrow \mathbf{y} = (B \otimes C)\mathbf{x}, \quad (6)$$

where  $\mathbf{x} = \text{vec}(X)$  and  $\mathbf{y} = \text{vec}(Y)$ .

## 2.3 Matrix Formulation of Geometries

B. R. Hunt came up with a way to develop the wavefront reconstruction problem in a matrix-algebra framework [12]. We assume the two-dimensional phase spectrum exists over a grid of points indexed by  $i, j = 1, 2, \dots, n$ , so that  $\Phi_{ij}$  is the phase spectrum at the  $ij$ th coordinate. The wave-front sensor produces phase differences on each of the two coordinates; we denote by  $\mathbf{s}_x$  the phase differences on the  $i$  coordinates and by  $\mathbf{s}_y$  the phase differences on the  $j$  coordinates.

Using the  $\text{vec}$  operator so that  $\text{vec}(\Phi) = \phi$ , we then write the problem as

$$\mathbf{s} = P\phi + \epsilon, \quad (7)$$

where

$$\mathbf{s} = \begin{bmatrix} \mathbf{s}_x \\ \mathbf{s}_y \end{bmatrix} \quad \text{and} \quad P = \begin{bmatrix} P_x \\ P_y \end{bmatrix}.$$

The vector  $\epsilon$  represents noise and other errors in the measured data, and is typically assumed to be independent and identically distributed (iid) normal (Gaussian) random values.

In the case of Hudgin geometry [11],

$$P_x = I \otimes H \quad \text{and} \quad P_y = H \otimes I,$$

where

$$H = \begin{bmatrix} 1 & -1 & & & \\ & 1 & -1 & & \\ & & \ddots & \ddots & \\ & & & 1 & -1 \end{bmatrix} \quad (8)$$

is  $(n-1) \times n$ , and  $I$  is the  $n \times n$  identity.  $H$  is a 1-dimensional phase difference matrix.

In the case of Fried geometry [7],

$$P_x = F \otimes H \quad \text{and} \quad P_y = H \otimes F,$$

where

$$F = \frac{1}{2} \begin{bmatrix} 1 & 1 & & & \\ & 1 & 1 & & \\ & & \ddots & \ddots & \\ & & & 1 & 1 \end{bmatrix}$$

is  $(n-1) \times n$ , and  $H$  is as above.  $F$  is an averaging matrix. For the remainder of this paper, we consider only the Fried geometry because it is used most often in adaptive optics applications.

When least squares estimation is used to obtain a solution of (7), the coefficient matrix for the normal equations has the form  $P_x^T P_x + P_y^T P_y$ . When Hudgin geometry is used,  $P_x^T P_x + P_y^T P_y$  is the standard discrete Laplacian, with grid representation given by

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

whereas when Fried geometry is used,  $P_x^T P_x + P_y^T P_y$  is a discrete Laplacian with grid representation given by

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

### 3 Wave-Front Reconstruction Using TSVD-Type Regularization

We now discuss the mathematical problem at the heart of this reconstruction [16, 17]: solving equation (7) for  $\phi$ , which is a rank-deficient problem. Thus, we must solve for  $\phi$  in the least-squares sense.

In the case of Fried geometry, we have:

$$\min_{\phi} \left\| \begin{bmatrix} F \otimes H \\ H \otimes F \end{bmatrix} \phi - \mathbf{s} \right\|_2^2 \quad (9)$$

We compute the GSVD of  $F$  and  $H$ , such that  $F = U_1 \Sigma_1 X_1^T$  and  $H = V_1 \Delta_1 X_1^T$ . Then, using equations (3) and (5), we can see that:

$$\begin{bmatrix} F \otimes H \\ H \otimes F \end{bmatrix} = \begin{bmatrix} U_1 \otimes V_1 & \\ & V_1 \otimes U_1 \end{bmatrix} \begin{bmatrix} \Sigma_1 \otimes \Delta_1 \\ \Delta_1 \otimes \Sigma_1 \end{bmatrix} (X_1 \otimes X_1)^T \equiv U \Sigma X^T. \quad (10)$$

Since  $\Sigma_1$  is a superdiagonal  $(n-1) \times n$  matrix and  $\Delta_1$  is a diagonal  $(n-1) \times n$  matrix, the matrix  $\Sigma$  is a  $2(n-1)^2 \times n^2$  matrix, composed of two roughly diagonal blocks. This matrix is rank deficient; its first and last columns are all zero. By systematically applying Givens rotations, we can reduce this block diagonal matrix  $\Sigma$  to diagonal form; we denote this as  $\Sigma = QD$ , where the matrix  $Q$  contains the Givens rotations.

Using equation (9), we obtain

$$\begin{aligned} \min_{\phi} \left\| \begin{bmatrix} F \otimes H \\ H \otimes F \end{bmatrix} \phi - \mathbf{s} \right\|_2^2 &= \min_{\phi} \|U \Sigma X^T \phi - \mathbf{s}\|_2^2 \\ &= \min_{\phi} \|U Q D X^T \phi - \mathbf{s}\|_2^2 \\ &= \min_{\phi} \|D X^T \phi - Q^T U^T \mathbf{s}\|_2^2 \\ &= \min_{\hat{\phi}} \|D \hat{\phi} - \hat{\mathbf{s}}\|_2^2, \quad \text{where } \hat{\phi} = X^T \phi \text{ and } \hat{\mathbf{s}} = Q^T U^T \mathbf{s}. \end{aligned}$$

Since  $D$  is a diagonal matrix, it becomes trivial to solve this minimization problem. Givens rotations can be organized such that the last two diagonal elements of  $D$  are equal to 0, and we then have freedom in deciding what to use for the solution terms corresponding to these values. We effectively truncate these two elements by setting their reciprocals to 0 in the pseudo-inverse  $\hat{D}^{-1}$ . This is where the term TSVD-like regularization comes from. It is important to note that the remaining  $n^2 - 2$  values of  $D$  are relatively large; the problem is rank-deficient, but it does not have the classical properties of an ill-posed problem, which are that there are many small singular values, and no gap to indicate a numerical rank.

Because  $F$  and  $H$  are defined by the geometry of the telescope, the generalized singular value decomposition may be performed once and stored. Thus,  $U_1$ ,  $V_1$ ,  $X_1$ ,  $D$ , and the rotators defining  $Q$  can be precomputed off line. Additionally, a suitable factorization of  $X_1$  (such as the  $LU$  decomposition) can be precomputed off line. It is important to note that we never actually explicitly compute  $U$ ,  $\Sigma$ , or  $X$ , but instead we always work with

the smaller matrices that make up these larger ones, using the properties of Kronecker products. In particular, if we let  $\text{vec}(\begin{bmatrix} S_x & S_y \end{bmatrix}) = \mathbf{s}$  and  $\text{vec}(\Phi) = \phi$ , the work to be done in real time then becomes

1. Compute  $\hat{\mathbf{s}} = Q^T U^T \mathbf{s} \Leftrightarrow \hat{S} = \hat{Q}^T \left( \begin{bmatrix} V_1^T S_x U_1 \\ U_1^T S_y V_1 \end{bmatrix} \right)$ , where  $\hat{Q}^T(A)$  applies the Givens rotators described in  $Q^T$  to the matrix  $A$ , retaining matrix form throughout.
2. Solve  $\min_{\hat{\phi}} \left\| D\hat{\phi} - \hat{\mathbf{s}} \right\|_2^2 \Leftrightarrow \hat{\Phi} = \hat{S} \odot \hat{D}^{-1}$ , where  $\hat{D}^{-1}$  contains the pseudo-inverse of the diagonal elements of  $D$ , with two elements of  $\hat{D}^{-1}$  set to 0. The symbol  $\odot$  represents Hadamard (i.e., element wise) multiplication.
3. Solve  $X^T \phi = \hat{\phi} \Leftrightarrow \Phi = X_1^{-T} \hat{\Phi} X_1^{-1}$ , using our precomputed factorization of  $X_1$ .

## 4 Wave-Front Reconstruction Using Tikhonov-Type Regularization

While solving equation (9) using TSVD-type regularization works well, astronomers prefer a form of Tikhonov regularization motivated by *a priori* knowledge of the phase error statistics. The values of  $\phi$  are determined by turbulent temperature variations within certain layers of the earth's atmosphere. The turbulence is often modeled by a Gaussian random vector with mean  $\mathbf{0}$  and a covariance matrix  $C_\phi$  with spectrum given by the von Karman turbulence model [18]. In [5], it is shown that the inverse biharmonic (or squared Laplacian) matrix gives an accurate, sparse approximation of  $C_\phi$  (see also [2]). Using this approximation, the minimum variance estimator for the true phase error is obtained by solving

$$\min_{\phi} \left\| \begin{bmatrix} F \otimes H \\ H \otimes F \end{bmatrix} \phi - \begin{bmatrix} \mathbf{s}_x \\ \mathbf{s}_y \end{bmatrix} \right\|_2^2 + (\sigma^2/c_0) \|L\phi\|^2, \quad (11)$$

where  $L$  is a discretization of the Laplacian;  $\sigma^2$  is the variance of the iid Gaussian noise vector  $\epsilon$  in (7); and  $c_0$  is a scaling parameter chosen so that  $E(\phi^T C_\phi \phi) = c_0 E(\phi^T L^{-2} \phi)$ ; see [2, 8] for more details. As is the case in other AO literature, we assume that the user has an estimate of  $\sigma^2$  in hand.

In order to express  $L$  in Kronecker product form, we estimate it using the Laplacian arising from the Hudgin geometry. In particular,

$$\|L\phi\|^2 \stackrel{\text{def}}{=} \left\| \begin{bmatrix} (I \otimes H)\phi \\ (H \otimes I)\phi \end{bmatrix} \right\|^2,$$

where  $H$  is defined in (8); note then that  $L^T L = I \otimes (H^T H) + (H^T H) \otimes I$ .

Equating  $\sigma^2/c_0$  with the regularization parameter  $\alpha$ , problem (11) can be written

$$\min_{\phi} \left\| \begin{bmatrix} F \otimes H \\ H \otimes F \\ \alpha I \otimes H \\ H \otimes \alpha I \end{bmatrix} \phi - \begin{bmatrix} \mathbf{s}_x \\ \mathbf{s}_y \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \right\|_2^2. \quad (12)$$

An important observation may be made regarding the coefficient matrix in (12); a permutation matrix  $\Pi$  exists such that

$$\Pi \begin{bmatrix} F \otimes H \\ H \otimes F \\ \alpha I \otimes H \\ H \otimes \alpha I \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} F \\ \alpha I \end{bmatrix} \otimes H \\ H \otimes \begin{bmatrix} F \\ \alpha I \end{bmatrix} \end{bmatrix} \equiv \begin{bmatrix} F_\alpha \otimes H \\ H \otimes F_\alpha \end{bmatrix}.$$

Thus, if the value of  $\alpha$  did not change from one data set to another, we could apply the techniques of the previous section and solve this problem in the same manner. Namely, we could compute the GSVD of  $F_\alpha$  and  $H$ , then efficiently solve using matrix-matrix multiplications and Hadamard arithmetic.

Unfortunately, the value of  $\alpha$  does not remain the same between data sets. However, though the value of  $\alpha$  will typically change, the range of  $\alpha$  values is restricted. Thus, we can choose an ‘‘average’’  $\alpha$  value, denoted  $\alpha_0$ , that approximates the value of  $\alpha$  for any data set, and use this  $\alpha_0$  value to form a preconditioner. More details are given in the next section.

To demonstrate how we construct the preconditioner, we rewrite equation (12) as

$$\min_{\phi} \left\| \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \phi - \begin{bmatrix} \mathbf{s} \\ \mathbf{0} \end{bmatrix} \right\|_2^2, \quad \text{where } A_1 = F_{\alpha_0} \otimes H, A_2 = H \otimes F_{\alpha_0}.$$

Our goal is to find a preconditioner  $M$  such that  $M^T M \approx A_1^T A_1 + A_2^T A_2$ , while retaining the desirable Kronecker product structure. We begin by calculating the GSVD of  $F_{\alpha_0}$  and  $H$ :  $F_{\alpha_0} = U \Sigma X^T, H = V \Delta X^T$ . Observe that

$$\begin{aligned} & A_1^T A_1 + A_2^T A_2 \\ &= (X \Sigma^T U^T \otimes X \Delta^T V^T)(U \Sigma X^T \otimes V \Delta X^T) + (X \Delta^T V^T \otimes X \Sigma^T U^T)(V \Delta X^T \otimes U \Sigma X^T) \\ &= (X \Sigma^T U^T U \Sigma X^T) \otimes (X \Delta^T V^T V \Delta X^T) + (X \Delta^T V^T V \Delta X^T) \otimes (X \Sigma^T U^T U \Sigma X^T) \\ &= (X \Sigma^T \Sigma X^T) \otimes (X \Delta^T \Delta X^T) + (X \Delta^T \Delta X^T) \otimes (X \Sigma^T \Sigma X^T) \\ &= (X \otimes X)(\Sigma^T \Sigma \otimes \Delta^T \Delta)(X \otimes X)^T + (X \otimes X)(\Delta^T \Delta \otimes \Sigma^T \Sigma)(X \otimes X)^T \\ &= (X \otimes X)(\Sigma^T \Sigma \otimes \Delta^T \Delta + \Delta^T \Delta \otimes \Sigma^T \Sigma)(X \otimes X)^T \\ &= (X \otimes X)C(X \otimes X)^T, \end{aligned}$$

where  $C = \Sigma^T \Sigma \otimes \Delta^T \Delta + \Delta^T \Delta \otimes \Sigma^T \Sigma$  is a diagonal matrix.

We want  $M^T M \approx (X \otimes X)C(X \otimes X)^T$ , which we achieve (exactly) by setting  $M = QC^{1/2}(X \otimes X)^T$ .  $Q$  can be any orthogonal matrix; as long as it has an appropriate Kronecker structure, the computations using  $M$  will be efficient. For all the results in this paper, we set  $Q$  to be the identity matrix.

We remark that the preconditioner  $M$  is computed once using an appropriate average  $\alpha_0$  for all likely data sets. Thus, we do not need to consider the computational costs of computing the GSVD of  $F_{\alpha_0}$  and  $H$  as part of the real-time time constraints needed by the application. The appropriate  $Q$ ,  $X$  and  $C^{1/2}$  factors will be precomputed once and stored, or, rather, their inverses will be formed and stored. Since  $C$  contains one zero diagonal element, we set the corresponding element in  $C^{-1/2}$  to be 0.

Using the preconditioner, the problem now becomes

$$\min_{\phi} \left\| AM^{-1}M\phi - \begin{bmatrix} \mathbf{s} \\ \mathbf{0} \end{bmatrix} \right\|_2^2, \quad \text{where } A = \begin{bmatrix} F \otimes H \\ H \otimes F \\ \alpha I \otimes H \\ H \otimes \alpha I \end{bmatrix}. \quad (13)$$

We use preconditioned LSQR [14, 15] to solve this problem. Note that, again, all matrix-vector multiplications can be performed by exploiting the Kronecker product structure of this problem and using matrix-matrix multiplications. This exploitation occurs both when applying the preconditioner (or its transpose) or when multiplying a vector by  $A$  or  $A^T$ .

## 5 Approximation Quality of Preconditioner

It is important to consider how well a preconditioner approximates a matrix to determine the quality of the preconditioner. In this section, we consider the approximation quality of the preconditioner in two ways: first, by looking at the singular values of the original matrix and of the preconditioned system, and second, by looking at the relative errors of the matrix and its preconditioner.

To obtain realistic values for  $\alpha_0$ , we ran a series of Matlab simulations. We created an  $n^2 \times 1$  “true”  $\phi$  [2] and formed  $\mathbf{s}_x$  and  $\mathbf{s}_y$  from equation (7). We then repeated the following 1000 times: we added a realization of 10% noise to the gradients  $\mathbf{s}_x$  and  $\mathbf{s}_y$ , then determined an optimal  $\alpha$  value for that realization by solving equation (12) for multiple values of  $\alpha$ . An  $\alpha$  value was considered optimal if it produced a reconstructed  $\phi$  with minimal relative residual error. After finding the 1000 optimal  $\alpha$  values for a given size  $n$ , we determined the  $\alpha_0$  for that  $n$  by using the value at the peak of the histogram of the  $\alpha$  values. As  $n$  increases, the range of optimal  $\alpha$  values decreases, leading to a closer fit of  $\alpha_0$  for all  $\alpha$  values. Because the  $\alpha$  value can be found without knowledge of the true  $\phi$  (see [2]), the  $\alpha_0$  value for real data can be found by again using the peak value of many  $\alpha$  values.

We consider the singular values of the matrix  $A$  given in equation (13) as well as the singular values of the preconditioned matrix  $AM^{-1}$ . We found that the singular values of the preconditioned system cluster to 1, with tighter clustering the closer  $\alpha_0$  is to  $\alpha$ . Figure 1 shows the singular values for six preconditioned systems as well as the singular values for one non-preconditioned system. In this figure, we took  $n$  to be 64 and  $\alpha_0$  to be 0.058. We varied  $\alpha$  to be from 0.03 to 0.124, which were the extremes of the optimal  $\alpha$  values we found during the 1000 realizations of noise. The dots above the dashed line show the singular values for these six preconditioned systems. Note that each preconditioned system has one zero singular value, while the remaining singular values cluster toward one. In particular, for the case  $\alpha = \alpha_0$ , all of the nonzero singular values cluster at one. For comparison, we include the singular values of one non-preconditioned system, for the case  $\alpha = 0.058$ . These are displayed in the figure below the dashed line. The smallest nonzero singular value for this system is 0.0488 and the largest is 2.0022.

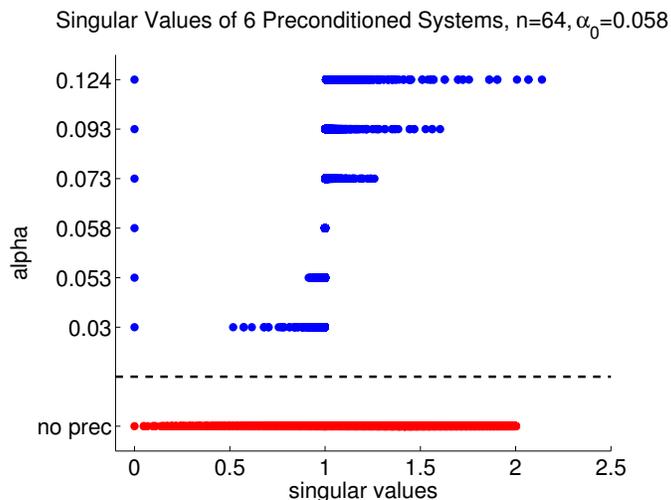


Figure 1: Singular values for six preconditioned systems as well as one non-preconditioned system.

The second measure of the quality of the preconditioner can be obtained by considering the relative errors of the matrix and its preconditioner. Let  $\alpha_0$  be the value used in the preconditioner  $M$  and suppose  $\alpha$  is the actual value specified by the data collection process. Then,

$$\begin{aligned} A^T A &= (F_\alpha^T F_\alpha) \otimes (H^T H) + (H^T H) \otimes (F_\alpha^T F_\alpha) \\ M^T M &= (F_{\alpha_0}^T F_{\alpha_0}) \otimes (H^T H) + (H^T H) \otimes (F_{\alpha_0}^T F_{\alpha_0}). \end{aligned}$$

We can compute the error between these matrices as

$$E = A^T A - M^T M = (\alpha^2 - \alpha_0^2)I \otimes H^T H + H^T H \otimes (\alpha^2 - \alpha_0^2)I.$$

Due to the simple nature of these matrices, we can explicitly compute the Frobenius norm of each:

$$\begin{aligned} \|A^T A\|_F &= \sqrt{4\alpha^4(5n^2 - 8n + 2) + 8\alpha^2(2n^2 - 5n + 3) + 5n^2 - 14n + 10} \\ \|M^T M\|_F &= \sqrt{4\alpha_0^4(5n^2 - 8n + 2) + 8\alpha_0^2(2n^2 - 5n + 3) + 5n^2 - 14n + 10} \\ \|E\|_F &= 2|\alpha^2 - \alpha_0^2|\sqrt{5n^2 - 8n + 2}. \end{aligned}$$

Thus,

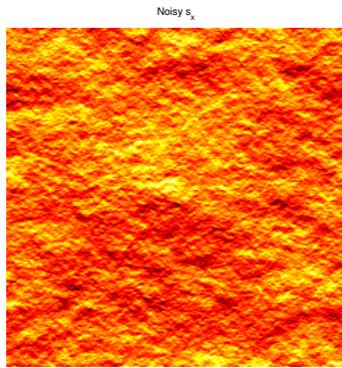
$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{\|E\|}{\|A^T A\|} &= \lim_{n \rightarrow \infty} \frac{2|\alpha^2 - \alpha_0^2|\sqrt{5 - \frac{8}{n} + \frac{2}{n^2}}}{\sqrt{4\alpha^4(5 - \frac{8}{n} + \frac{2}{n^2}) + 8\alpha^2(2 - \frac{5}{n} + \frac{3}{n^2}) + (5 - \frac{14}{n} + \frac{10}{n^2})}} \\ &= \frac{2\sqrt{5}|\alpha^2 - \alpha_0^2|}{\sqrt{20\alpha^4 + 16\alpha^2 + 5}} \\ &\leq 2|\alpha^2 - \alpha_0^2|. \end{aligned}$$

## 6 Results

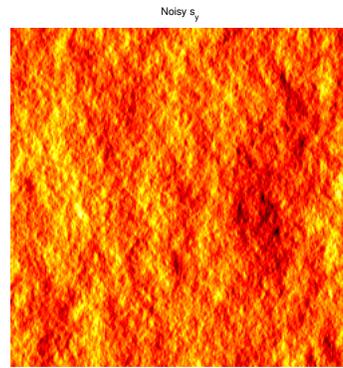
We now present some results showing the effectiveness of our preconditioning approach, even when  $\alpha_0$  differs from the  $\alpha$  value computed for a data set. We determined the optimal  $\alpha_0$  for  $n = 256$  using the procedure described in Section 5. First, though, we give an example of the noisy data and reconstructed solution.

Figure 2 shows a realization of the noisy phase differences  $S_x$  and  $S_y$  for the case  $n = 256$  with 10% noise. Figure 3 shows the reconstructed  $\Phi$  for the TSVD-type regularization on the left and the Tikhonov-type regularization on the right. Visually, the results appear indistinguishable. However, the relative errors differ slightly; the relative error for the TSVD solution is 1.112e-2 and for the Tikhonov solution is 9.856e-3 – a roughly 10% improvement. This result, that the Tikhonov approach produces a solution with slightly less relative error than the TSVD solution, is consistent for all the noise realizations, and explains why the regularized approach is often favored by astronomers [5]. Note that the cost of LSQR is more expensive than using the simple TSVD approach; the cost of using the TSVD approach to solve the problem is the same as one preconditioner solve, namely  $O(n^3)$  for images with  $n^2$  pixels. The storage requirements are essentially the same for both approaches.

The final set of results we present focus entirely on the Tikhonov-type regularization. We look at the number of preconditioned LSQR iterations required for different data sets, solving to the default tolerance of  $10^{-6}$ . Figure 4 compares the value of  $\alpha$  to the number of preconditioned LSQR iterations required for 1000 realizations of 10% noise. As expected, when  $\alpha$  exactly equals  $\alpha_0$ , only one preconditioned LSQR iteration is required. As  $\alpha$

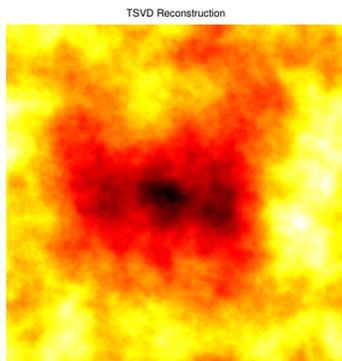


(a) Noisy  $S_x$

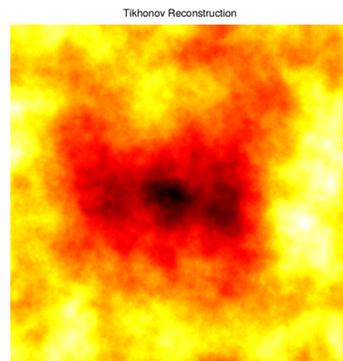


(b) Noisy  $S_y$

Figure 2: Noisy phase differences for  $n = 256$  with 10% noise.



(a) TSVD Reconstruction



(b) Tikhonov Reconstruction

Figure 3: Reconstructions using two types of regularization.

becomes further from the preconditioner value  $\alpha_0$  (either being smaller or larger than  $\alpha_0$ ), the number of LSQR iterations required increases. Figure 5 displays the results from this same set of simulations but in a different way. Here, we display the number of times a given number of preconditioned LSQR iterations is required during our 1000 simulations. For example, in nearly one-fourth of the experiments, exactly three LSQR iterations are required to reach a convergence of  $10^{-6}$ . For comparison, Table 1 gives the average number of LSQR iterations required both with and without preconditioning, when solved to different tolerance levels and with different amounts of noise.

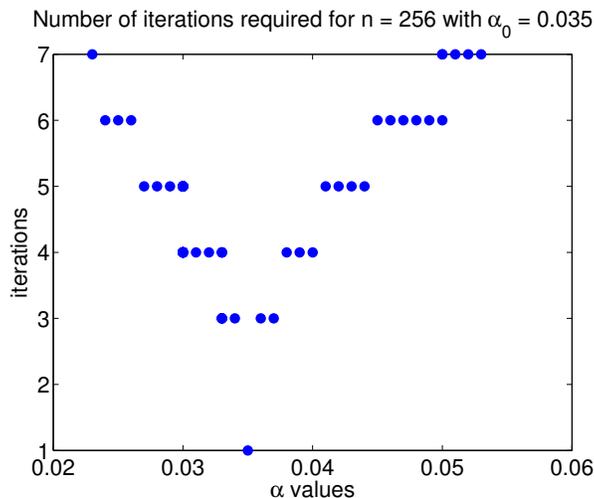


Figure 4: Comparison of  $\alpha$  to number of preconditioned LSQR iterations required.

Table 1: Number of LSQR iterations required, on average, for  $n = 256$  with four different amounts of noise for various tolerance levels.

Noise level	Preconditioning?	$1e-1$	$1e-2$	$1e-3$	$1e-4$	$1e-5$	$1e-6$
1%	No	5	156	346	490	554	611
1%	Yes	1	1	1	2	3	3
5%	No	5	112	220	408	531	578
5%	Yes	1	1	2	2	3	4
10%	No	5	98	193	372	515	564
10%	Yes	1	1	2	3	3	4
20%	No	5	81	175	345	488	554
20%	Yes	1	1	2	3	3	4

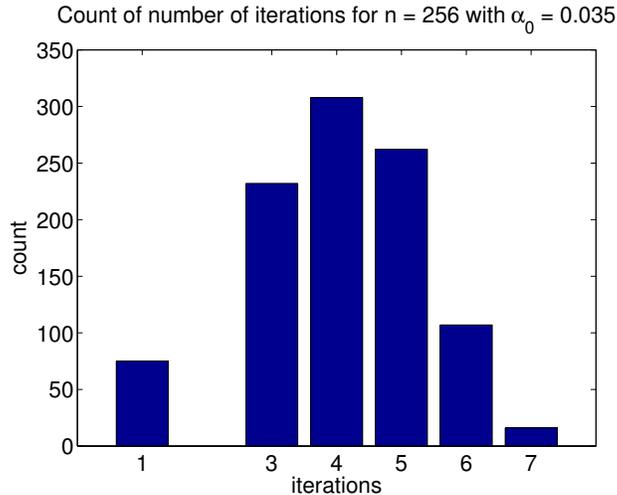


Figure 5: Count of number of preconditioned LSQR iterations required for 1000 realizations of noise.

## 7 Concluding Remarks

A Kronecker product structured, rank-deficient least-squares adaptive optics problem was described, for determining the phase values from noisy phase differences. Efficient techniques for solving on the Fried geometry with square aperture were described. Two types of regularization were considered: a TSVD-type and a preferred Tikhonov-type. Using properties of Kronecker products and the GSVD, a direct solution using TSVD regularization was described. Though this gives good results, a relatively better reconstruction may be obtained using Tikhonov-type regularization. By forming a preconditioner that will work well with any data set, fast convergence is obtained when using LSQR. Approximation quality results of the preconditioner were given, as well as numerical results on simulated data. It was shown that as long as the  $\alpha_0$  value used by the preconditioner is suitably chosen, the number of preconditioned LSQR iterations required is small. Exploiting the linear algebra structure of this problem has led to extremely efficient computational approaches that can be solved in real time.

## References

- [1] S. Barbero, J. Rubinstein, and L. N. Thibos. Wavefront sensing and reconstruction from gradient and Laplacian data measured with a Hartmann-Shack sensor. *Optics Letters*, 31(12):1845–1847, 2006.

- [2] J. M. Bardsley. Wavefront reconstruction methods for adaptive optics systems on ground-based telescopes. *SIAM J. on Matrix Analysis and Applications*, 30(1):67–83, 2008.
- [3] H. H. Barrett, C. Dainty, and D. Lara. Maximum-likelihood methods in wavefront sensing: stochastic models and likelihood functions. *J. Opt. Soc. Am. A. Opt Image Sci Vis.*, 24(2):391–414, 2007.
- [4] R. C. Cannon. Global wave-front reconstruction using Shack-Hartmann sensors. *J. Opt. Soc. Am. A.*, 12:2031–2039, 1995.
- [5] B. L. Ellerbroek. Efficient computation of minimum-variance wave-front reconstructors with sparse matrix techniques. *J. Opt. Soc. Am. A*, 19(9):1803–1816, 2002.
- [6] S. Ettl, J. Kaminski, M. C. Knauer, and G. Husler. Shape reconstruction from gradient data. *Applied Optics*, 47(12):2091–2097, 2008.
- [7] D. L. Fried. Least-square fitting a wave-front distortion estimate to an array of phase-difference measurements. *J. Opt. Soc. Am.*, 67(3):370–375, 1977.
- [8] L. Gilles. Order-n sparse minimum-variance open-loop reconstructor for extreme adaptive optics. *Optics Letters*, 28(20):1927–1929, 2003.
- [9] G. H. Golub and C. Van Loan. *Matrix computations, 3rd edition*. Johns Hopkins University Press, 1996.
- [10] J. W. Goodman. *Introduction to Fourier Optics, 2nd edition*. McGraw-Hill, 1996.
- [11] R. H. Hudgin. Wave-front reconstruction for compensated imaging. *J. Opt. Soc. Am.*, 67(3):375–378, 1977.
- [12] B. R. Hunt. Matrix formulation of the reconstruction of phase values from phase differences. *J. Opt. Soc. Am.*, 69(3):393–399, 1979.
- [13] C. Van Loan. The ubiquitous Kronecker product. *J. Comp. and Appl. Math.*, 123(1):85–100, 2000.
- [14] C. C. Paige and M. A. Saunders. Algorithm 583 LSQR: Sparse linear equations and least squares problems. *ACM Trans. Math. Soft.*, 8(2):195–209, 1982.
- [15] C. C. Paige and M. A. Saunders. LSQR: An algorithm for sparse linear equations and sparse least squares. *ACM Trans. Math. Soft.*, 8(1):43–71, 1982.
- [16] H. Ren and R. Dekany. Fast wave-front reconstruction by solving the Sylvester equation with the alternating direction implicit method. *Optics Express*, 12(14):3279–3296, 2004.

- [17] H. Ren, R. Dekany, and M. Britton. Large-scale wave-front reconstruction for adaptive optics systems by use of a recursive filtering algorithm. *Applied Optics*, 44(13):2626–2637, 2005.
- [18] M. Roggemann and B. Welsh. *Imaging Through Turbulence*. CRC Press, 1996.