Efficient computation of minimum-variance wave-front reconstructors with sparse matrix techniques

Brent L. Ellerbroek

Gemini Observatory, 670 North A'Ohuku Place, Hilo, Hawaii 96720

Received January 7, 2002; revised manuscript received April 1, 2002; accepted April 9, 2002

The complexity of computing conventional matrix multiply wave-front reconstructors scales as $O(n^3)$ for most adaptive optical (AO) systems, where n is the number of deformable mirror (DM) actuators. This is impractical for proposed systems with extremely large n. It is known that sparse matrix methods improve this scaling for least-squares reconstructors, but sparse techniques are not immediately applicable to the minimum-variance reconstructors now favored for multiconjugate adaptive optical (MCAO) systems with multiple wave-front sensors (WFSs) and DMs. Complications arise from the nonsparse statistics of atmospheric turbulence, and the global tip/tilt WFS measurement errors associated with laser guide star (LGS) position uncertainty. A description is given of how sparse matrix methods can still be applied by use of a sparse approximation for turbulence statistics and by recognizing that the nonsparse matrix terms arising from LGS position uncertainty are low-rank adjustments that can be evaluated by using the matrix inversion lemma. Sample numerical results for AO and MCAO systems illustrate that the approximation made to turbulence statistics has negligible effect on estimation accuracy, the time to compute the sparse minimum-variance reconstructor for a conventional natural guide star AO system scales as $O(n^{3/2})$ and is only a few seconds for $n = 3500$, and sparse techniques reduce the reconstructor computations by a factor of 8 for sample MCAO systems with 2417 DM actuators and 4280 WFS subapertures. With extrapolation to 9700 actuators and 17,120 subapertures, a reduction by a factor of approximately 30 or 40 to 1 is predicted.

OCIS code: 010.1080.

1. INTRODUCTION

Adaptive optical (AO) systems are used in ground-based optical and near-IR astronomical telescopes to correct for the phase aberrations induced by atmospheric turbulence.[1–3] These aberrations are compensated by adjusting the figure of a deformable mirror (DM) to null the residual phase errors as measured by a wave-front sensor (WFS). The control algorithm used to determine the DM commands from the WFS measurements is frequently referred to as the wave-front reconstructor. Analytical and numerical methods for optimizing and evaluating reconstructors are well developed, both for existing conventional AO systems,[4–7] and for proposed AO systems with multiple laser guide stars (LGSs)[8,9] and proposed multiconjugate AO (MCAO) systems that would employ multiple DMs and WFSs to compensate for atmospheric turbulence across an extended field of view.[10–14] In comparison, work on efficiently computing and implementing reconstruction algorithms is at a less advanced state. Since wave-front reconstruction is a linear estimation problem for the case of astronomical AO systems, the DM actuator command vector a can be determined from the WFS measurement vector s by using a matrix multiply. The complexity of computing the reconstruction matrix by using standard methods scales as $O(n^3)$, where n is the dimensionality of a. This will become impractical for proposed MCAO and "extreme" AO (ExAO) systems on extremely large telescopes (ELT's), where n may be on the order of 10,000 or even 100,000.

Fortunately, it has been known for some time that sparse matrix techniques may be used to significantly reduce the numerical complexity of classical least-squares reconstruction algorithms, which determine a by minimizing the rms value (or L^2 norm) of the residual measurement of s.[15] This method exploits the fact that the DM-to-WFS influence matrix is sparse for many AO component technologies, so that an adjustment to a single DM actuator effectively influences WFS measurements from only a few neighboring subapertures. The essential reason that this reduces computational complexity is the fact that a sparse, banded matrix M may be factored as $M = LL^{\mathrm{T}}$ with L sparse and lower triangular, so that the system $x = M^{-1}y$ can be solved for x in terms of y by first solving $Lv = y$ and then $L^{\mathrm{T}}x = v$ by backsubstitution. Sparse matrix reconstructors and other efficient approaches such as Fourier transform reconstructors, multigrid methods,[16] and preconditioned conjugate gradient algorithms[17] are promising approaches for future ExAO systems, since theory and simulation indicate that conventional least-squares reconstruction is near optimal for this class of AO systems.[18] Sparse matrix methods have also been proposed for efficiently deconvolving a known, but spatially varying, point-spread function from a blurry image.[19]

However, sparse matrix methods are not immediately applicable to reconstruction algorithms for MCAO systems for several reasons. First, conventional least-squares reconstructors generally perform poorly for MCAO systems, and a regularized, or modally filtered, algorithm is necessary to obtain optimal or near-optimal

performance.[20] Two approaches that are more satisfactory are singular-value decomposition filtering of the reconstructor to suppress poorly sensed modes and minimum-variance wave-front reconstruction,[7] but both techniques yield full matrices that are incompatible with the direct application of sparse matrix techniques. Second, the position uncertainty problem for LGSs introduces high levels of full aperture tip/tilt measurement noise in LGS WFS measurements. The global tip/tilt terms must be filtered from the LGS measurements, and auxiliary tip/tilt natural guide star (NGS) WFSs must be included in the MCAO system to measure the full-aperture wave-front tilt. Both of these effects complicate the structure of the DM-to-WFS influence matrix, with the result that sparse matrix methods once again cannot be immediately applied.

This paper describes techniques for dealing with these difficulties to obtain a representation for the minimum-variance wave-front reconstructor that can be efficiently evaluated by using sparse matrix methods, even for the case of a MCAO system. First, the nonsparse regularization term appearing in the reconstructor is replaced by a sparse approximation. In a heuristic sense this approximation is equivalent to replacing the Kolmogorov $\kappa^{-11/3}$ spatial power spectrum for atmospheric turbulence (where κ is a spatial-frequency variable) with κ^{-4}. It may be a surprise that this approximation has only a negligible impact upon the performance of the reconstructor, increasing the mean square residual phase error by from 0.1% to no more than 1.5% for all of the cases evaluated here. Second, the nonsparse matrix terms appearing in the minimum-variance reconstructor due to LGS position uncertainty and the inclusion of NGS tip/tilt WFSs in the AO system are of low rank, which allows sparse matrix methods to still be applied with the aid of the matrix inversion lemma. Briefly, this lemma implies that if U and V are matrices with only a few columns and are dimensioned such that $M + UV^{\mathrm{T}}$ is defined, then $(M + UV^{\mathrm{T}})^{-1}$ is equal to $M^{-1} + U'V'^{\mathrm{T}}$, where the matrices U' and V' have the same dimensions as those of U and V. It follows that if M has a structure that allows the system $x = M^{-1}y$ to be solved efficiently (e.g., a sparse matrix), the system $x = (M + UV^{\mathrm{T}})^{-1}y$ can be solved efficiently as well.

For sample cases involving NGS and LGS MCAO systems with approximately 2400 total DM actuators and 4280 WFS subapertures on a 16-m telescope aperture, the above methods provide approximately a factor-of-8 improvement in the time needed to compute the minimum-variance reconstructor. Upon extrapolation to a 32-m MCAO system with approximately 9700 actuators and 17,120 subapertures, an improvement by a factor of approximately 30 or 40 is predicted, although at present the author's computer does not have sufficient memory to evaluate this case. For ExAO systems the computation requirements for this sparse implementation of the minimum-variance reconstructor are comparable with results obtained previously for conventional least-squares reconstruction, allowing reconstructors to be computed in only a few seconds for very high-order AO systems.

The remainder of this paper is organized as follows. Section 2 reviews prior results on the use of sparse matrix techniques for least-squares reconstructors. Section 3 derives one of the standard formulas for the minimum-variance reconstructor, which decomposes into two parts: estimating the full turbulence profile and then finding the DM actuator commands that best correct for this profile over the desired field of view.[13] The sparse matrix methods for efficiently evaluating these two operators are presented in Sections 4 and 5. Section 6 summarizes sample simulation results obtained on reconstructor performance and computation requirements for a range of conventional AO, ExAO, and MCAO system configurations.

2. LEAST-SQUARES WAVE-FRONT RECONSTRUCTION

In this section we review the application of sparse matrix techniques to classical least-squares wave-front reconstruction, as first presented in Ref. 15. The improvements that can be obtained in computational efficiency are impressive, at least for the case of a conventional AO system with a single DM and a single NGS WFS.

Classical least-squares wave-front reconstruction is based upon the WFS measurement model

$$s = G_a a + n. \tag{1}$$

Here s is the WFS measurement vector, a is the DM actuator vector to be estimated from s, $G_a = \partial s/\partial a$ is the DM-to-WFS influence matrix, and n is additive, zero-mean measurement noise. All components of n are assumed to be uncorrelated and of equal variance. Note that atmospheric turbulence is not included in this measurement model, and the goal of least-squares wave-front reconstruction is simply to determine an estimate $\hat{a}$ of a that yields the best mean square fit to s in Eq. (1). The value of $\hat{a}$ is given formally by the expression

$$\hat{a} = \arg\min_a \|s - G_a a\|^2, \tag{2}$$

and this minimization problem can be solved by determining the value of a for which the partial derivatives $\partial\|s - G_a a\|^2/\partial a$ are identically zero. This is a standard linear least-squares problem, since the merit function to be minimized is quadratic in a. The minimum-norm solution is given by the formula[4–6]

$$\hat{a} = (G_a^{\mathrm{T}} G_a)^{\dagger} G_a^{\mathrm{T}} s, \tag{3}$$

where the superscript T denotes the transpose of a matrix or a vector and $M^{\dagger}$ is the pseudoinverse of the matrix M. Although this is nearly the most elementary reconstruction algorithm possible, least-squares estimation has been applied successfully in many hardware systems and remains perhaps the most commonly used AO control algorithm even today.

The classical least-squares reconstruction algorithm is of complexity $O(n_a^3)$ to compute and $O(n_a^2)$ to apply when implemented by using conventional matrix inversions and matrix/vector multiplies, where n_a is the dimension of the DM actuator command vector a.[21] This becomes a significant practical limitation for values of n_a in excess of approximately 1000, let alone the values of 10,000 or even 100,000 currently under consideration for so-called ExAO

systems. Fortunately, the matrix G_a is highly sparse for many WFS and DM technologies, since each individual DM actuator influence function couples into only a few (perhaps four or eight) elements of the wave-front sensor measurement vector s. It follows that the matrix $G_a^{\mathrm{T}}G_a$ is highly sparse as well, and this matrix can be decomposed into a Cholesky factorization

$$G_a^{\mathrm{T}}G_a = LL^{\mathrm{T}}, \tag{4}$$

where L is lower triangular and also sparse. In fact, the sparseness of L can be improved by reordering the columns of the influence matrix G_a, which corresponds to simply renumbering the elements of the DM actuator command vector.[22,23] Once the matrix $G_a^{\mathrm{T}}G_a$ has been factored, $\hat{a}$ may be computed from s in three steps by using the equations

$$v = G_a^{\mathrm{T}}s, \qquad Lw = v, \qquad L^{\mathrm{T}}\hat{a} = w. \tag{5}$$

The intermediate variable v is first computed from s by using the first equation, the second equation is next solved for w by using backsubstitution, and the third equation is then similarly solved for $\hat{a}$. Each of the matrices G_a^{T}, L, and L^{T} is sparse, leading to a very significant reduction in computational complexity. The reader should note that this technique still leads to the exact least-squares solution and is in fact less subject to round-off error than the conventional matrix multiply solution when n_a is very large.

The most dramatic reduction in computational requirements achieved through the use of sparse matrix techniques is obtained not in applying the least-squares wave-front reconstructor but rather in initially computing the matrix factorization $G_a^{\mathrm{T}}G_a = LL^{\mathrm{T}}$. Figure 1 plots the number of floating-point operations (FLOPs) required for this decomposition as a function of n_a for a square-aperture geometry and the so-called "Hudgins" or "shearing interferometer" WFS subaperture geometry.[5] The number of computations needed to explicitly compute $(G_a^{\mathrm{T}}G_a)^{\dagger}$ is also plotted for comparison. The computational requirements for the two approaches scale as $n_a^{3/2}$ and n_a^3, respectively, leading to a factor-of-3.35 $\times$ 10^5 advantage for the sparse matrix methods for the case of an extreme AO system with $n_a = 10^4$. This is a very nontrivial reduction in computational complexity, but, unfortunately, the classical least-squares reconstructor is not always an appropriate control algorithm for more sophisticated AO system configurations. A more general reconstruction algorithm that can be optimized for these applications is described below.

3. MINIMUM-VARIANCE WAVE-FRONT RECONSTRUCTION

The standard least-squares wave-front reconstruction algorithm performs poorly for some AO applications because the WFS measurement model described by Eq. (1) is oversimplified. The WFS measurement vector s is in fact a function of the atmospheric turbulence profile, not a set of preexisting DM actuator command errors that must be nulled to obtain a perfect wave front. The DM actuator command vector a should be selected to compensate the turbulence-induced wave-front error, which is not in general the same problem as that of finding the best fit to the WFS measurement vector s. If the statistics of the

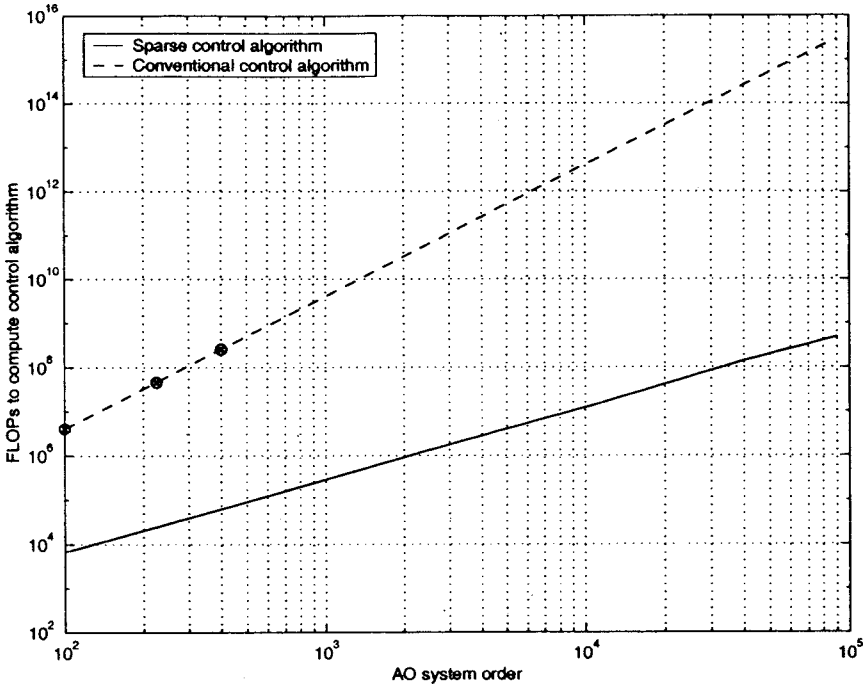

Fig. 1. Computation requirements for conventional and sparse calculations of classical least-squares reconstruction algorithms. These results are for a square-aperture geometry and the so-called "Hudgins" or "shearing interferometer" wave-front sensor geometry. The number of floating-point operations (FLOPs) needed to compute the control algorithm by using a conventional matrix inversion has been evaluated for systems of order 100, 225, and 400 and extrapolated by using the predicted third-order power law. The number of operations necessary for the sparse matrix factorization has been explicitly computed for AO systems of up to order 90,000 and scales with the three-halves power of the order of the system.

atmospheric turbulence profile and the WFS measurement noise are known, minimum-variance wave-front reconstruction[7] provides an optimal solution in the sense of (as the name implies) minimizing the variance of the residual wave-front error remaining after the DM actuator commands have been applied. This section briefly derives one of the standard representations for the minimum-variance reconstructor that can be, with some further work, efficiently evaluated by using sparse matrix techniques. This representation is general enough that it may be applied to AO applications involving one or several WFSs, DMs, atmospheric turbulence layers, and wave fronts to be corrected.[8,11]

A. Problem Formulation

The residual wave-front profile(s) remaining after commands have been applied to the DM(s) will be denoted ϕ and is defined by the equation

$$\phi = H_x x - H_a a. \tag{6}$$

The components of ϕ are the values of the residual phase profile(s) at a set of grid points in the telescope aperture plane, x is a vector of phase values on a grid of points on the atmospheric phase screen(s), a is the DM actuator command vector, and the columns of the matrices H_x and H_a are the "influence functions" associated with the discrete phase points and the DM actuators. The effects of x and a on the phase ϕ are assumed to be linear and are evaluated by tracing rays through the phase screens and the DM conjugate planes, as illustrated in Fig. 2(a). The vector x is a random variable with zero mean and finite second-order statistics. These statistics are typically modeled by using the Kolmogorov or von Kármán spectrum.

The mean square residual piston-removed wave-front error will be denoted σ^2 and is related to ϕ by the formula

$$\sigma^2 = \phi^{\mathrm{T}} W \phi, \tag{7}$$

where W is a symmetric, positive-semidefinite matrix. The coefficients of W may be defined, for example, so that the value of σ^2 is equal to the mean square piston-removed value of a continuous phase profile obtained by interpolating a smooth function through the values of ϕ specified on the discrete set of grid points.[12] The additional features of W needed to apply sparse matrix methods to this problem are outlined in Section 4.

The DM actuator command vector a is computed from the WFS measurement vector s by using a linear reconstruction algorithm of the form

$$a = Es, \tag{8}$$

where E is the wave-front reconstruction matrix. The WFS measurement s is modeled as

$$s = G_x x + n, \tag{9}$$

where G_x is the phase-to-WFS influence matrix and n is the WFS measurement noise. The elements of G_x are once again computed by tracing rays from the guide star(s) through the phase screen(s) to the wave-front sensing subapertures, as illustrated in Fig. 2(b). Equation (9) differs from Eq. (1) for the least-squares reconstruction algorithm in three important ways: s is now a function of x instead of a, the case of multiple WFSs is allowed, and n is a random vector with zero mean and finite second-order statistics. This more general noise model is required, for example, to model multiple NGSs of different brightness and the effects of LGS position uncertainty.

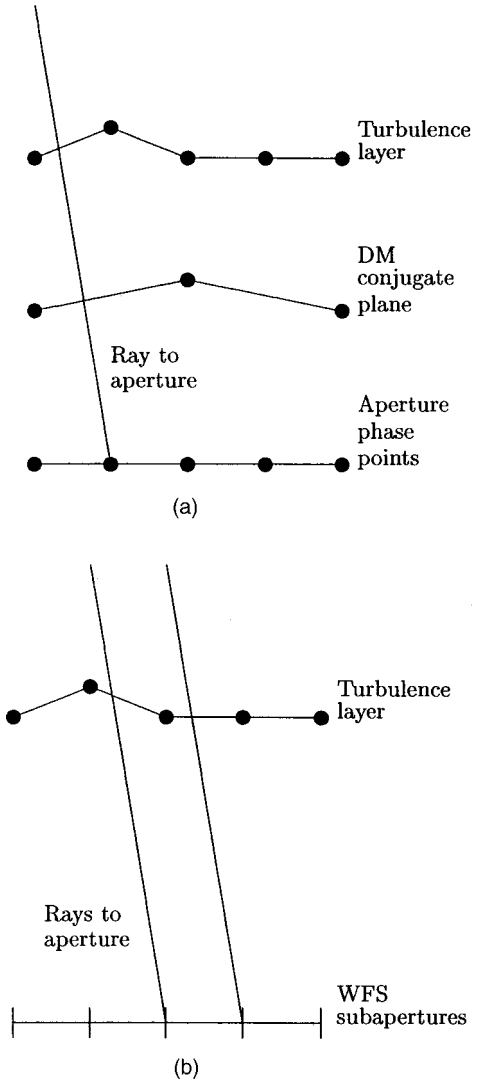

Fig. 2. Influence matrix models. Part (a) illustrates the relationship between the turbulence phase screen vector x, the DM actuator command vector a, and the residual phase error vector ϕ. These three vectors are defined as values on grids of points in the planes of the phase screens, the DM conjugate locations, and the telescope aperture, respectively. The influence matrices H_x and H_a are defined by tracing rays through the phase screens and mirrors as illustrated. Part (b) illustrates the similar relationship between the phase screen vector x and the WFS measurement vector s. In this case rays are traced from the guide star(s) through the phase screen(s) to obtain a wave front in the telescope aperture plane, and the WFS measurements are then computed as the average x and y wave-front gradients over each subaperture.

In this notation the minimum-variance reconstructor E_* is the value of E that minimizes the expected value of σ^2 averaged over the statistics of the phase profile x and the WFS measurement noise n. We generalize this definition slightly and write

$$E_* = \arg\min_E \langle \sigma^2 + k\|a\|^2 \rangle, \tag{10}$$

where the angle brackets $\langle \cdot \rangle$ denote ensemble averaging over the statistics of noise and turbulence. As shown in Subsection 3.B, the regularization term $k\|a\|^2 = ka^{\mathrm{T}}a$ must be included (with a very small value of k) to avoid singularities if the subspace of DM actuator commands having no effect on σ^2 is not *a priori* known.

B. Deriving the Reconstructor

Determining the minimum-variance reconstructor E_* that minimizes the mean square phase variance σ^2 is a least-squares minimization problem in the coefficients of E. Using the definitions for σ^2 and a given in Eqs. (7) and (8), we may write

$$\langle \sigma^2 + k\|a\|^2 \rangle = \langle (H_x x - H_a E s)^{\mathrm{T}} W (H_x x - H_a E s) + k s^{\mathrm{T}} E^{\mathrm{T}} E s \rangle \tag{11}$$

for the merit function to be minimized. The partial derivatives of this quantity with respect to the coefficients of E must vanish for the optimal value of the reconstructor; i.e.,

$$0 = \left. \frac{\partial^2 \langle \sigma^2 + k\|a\|^2 \rangle}{\partial E_{ij}} \right|_{E=E_*}. \tag{12}$$

Differentiating Eq. (11) with respect to the coefficients of E and rearranging yields the result

$$\langle s(H_a^{\mathrm{T}} W H_x x)^{\mathrm{T}} \rangle = \langle s(H_a^{\mathrm{T}} W H_a E_* s)^{\mathrm{T}} \rangle + k \langle s(E_* s)^{\mathrm{T}} \rangle. \tag{13}$$

It is convenient to introduce the notation

$$C_{vw} = \langle v w^{\mathrm{T}} \rangle \tag{14}$$

for the covariance of two zero-mean random variables v and w. By factoring the nonrandom matrices in Eq. (13) outside of the expected-value operations, we can rewrite this expression in the form

$$(H_a^{\mathrm{T}} W H_x) C_{xs} = (H_a^{\mathrm{T}} W H_a + kI) E_* C_{ss}. \tag{15}$$

Equation (15) can be solved immediately for the minimum-variance reconstructor E_* if the matrices C_{ss} and $H_a^{\mathrm{T}} W H_a + kI$ are invertible. The first of these two matrices will be invertible whenever the WFS measurement noise n is nonzero, and the remainder of the paper will restrict attention to this real-world case. The second matrix $H_a^{\mathrm{T}} W H_a$ will generally not be invertible for the case $k = 0$, i.e., the formally exact value of the minimum-variance reconstructor. This matrix will have a nontrivial null space, since there are some modes of DM actuator commands that have no effect on the piston-removed residual wave-front error—for example, overall piston adjustments to one or more of the DMs in the AO system. Clearly, there is no performance penalty if we require that the output of the wave-front reconstruction matrix be orthogonal to null $(H_a^{\mathrm{T}} W H_a)$, the null space of $H_a^{\mathrm{T}} W H_a$. This requirement is expressed by the constraint equation

$$N_w^{\mathrm{T}} E_* = 0, \tag{16}$$

where the columns of the matrix N_w are a linearly independent set of vectors from null $(H_a^{\mathrm{T}} W H_a)$. This constraint implies the condition $N_w N_w^{\mathrm{T}} E_* C_{ss} = 0$ as well, which may be summed with Eq. (15) to yield the result

$$(H_a^{\mathrm{T}} W H_x) C_{xs} = (H_a^{\mathrm{T}} W H_a + N_w N_w^{\mathrm{T}} + kI) E_* C_{ss}. \tag{17}$$

If we are certain that we know a full basis for null $(H_a^{\mathrm{T}} W H_a)$, it is now possible to set $k = 0$ and compute the precise value of the minimum-variance reconstructor. However, this assumption may not hold for MCAO systems because of the richer cross coupling between the actuators on the multiple DMs, and in this case k must be set to a (very small) nonzero value to avoid singularities.

(In passing, we note for future use that N_w will be a low-rank matrix with only a small number of columns, certainly much smaller than the dimension of the DM actuator command vector a.)

Solving Eq. (17) for the minimum-variance reconstructor E_* now yields the result

$$E_* = (H_a^{\mathrm{T}} W H_a + N_w N_w^{\mathrm{T}} + kI)^{-1} (H_a^{\mathrm{T}} W H_x) C_{xs} C_{ss}^{-1} = F_x E_x, \tag{18}$$

where the variables F_x and E_x are abbreviations for the terms

$$F_x = (H_a^{\mathrm{T}} W H_a + N_w N_w^{\mathrm{T}} + kI)^{-1} (H_a^{\mathrm{T}} W H_x), \tag{19}$$

$$E_x = C_{xs} C_{ss}^{-1}. \tag{20}$$

The matrix E_x estimates the turbulence profile x from the WFS measurement vector s, and the matrix F_x then fits a DM actuator command vector a to this estimated value. This decomposition of the minimum-variance wave-front reconstructor has been commented upon previously,[13] and it offers several useful insights into the reconstruction process. Note, for example, that the estimation matrix E_x is independent of the value of the weighting matrix W and that the fitting matrix F_x likewise does not depend upon the statistics of the WFS measurement noise and the atmospheric turbulence.

For relatively small-order AO systems the matrices F_x and E_x can be computed explicitly to obtain the minimum-variance reconstruction algorithm. For higher-order systems it is necessary to exploit the structure of the the matrices appearing in these definitions to obtain more efficient solutions, as described further in Section 4.

4. EFFICIENT SOLUTION FOR $u = F_x v$

Equation (19) for the matrix F_x is weakly similar to Eq. (3) for the least-squares wave-front reconstructor, which suggests the possibility of applying sparse matrix techniques. The matrices H_x and H_a are sparse, since [as is suggested by Fig. 2(a)], each element of the phase profile vector ϕ is a function of only a few elements of the atmospheric turbulence vector x and the DM actuator command vector a. However, the structure of the matrix F_x is complicated by the presence of the weighting matrix W and the extra term $N_w N_w^{\mathrm{T}}$. Subsections 4.A and 4.B develop the structure of W in greater detail and summarize

how sparse matrix methods may still be applied to find efficient solutions for $u = F_x v$.

A. Weighting Matrix Structure

As described in Subsection 3.A, the matrix W is a positive-semidefinite matrix chosen so that the quantity $\phi^{\mathrm{T}} W \phi$ equals the mean square piston-removed wave-front error for the phase profile(s) ϕ. In the wide-field-of-view case, the vector ϕ will be a concatenation of phase profiles $\phi^1, \ldots, \phi^n$ from n different evaluation directions:

$$\phi = \begin{pmatrix} \phi^1 \\ \vdots \\ \phi^n \end{pmatrix}, \tag{21}$$

where each profile ϕ^i is evaluated on the identical set of grid points in the telescope aperture plane. To account for aperture edge effects more precisely and obtain more accurate values for σ^2 with a limited number of grid points, we may associate each discrete phase profile ϕ^i with a continuous profile $\varphi^i(r)$ through the relationship

$$\varphi^i(r) = \sum_j \phi^i_j e_j(r), \tag{22}$$

where the terms $e_j(r)$ are localized influence functions associated with each point in the grid. The goal of this subsection is to determine a value for the weighting matrix W so that the value of $\phi^{\mathrm{T}} W \phi$ is actually equal to a weighted sum of the mean square, piston-removed values of the functions $\varphi^i(r)$, evaluated over the continuous telescope aperture.

The weighting matrix W is block diagonal, since there is no cross coupling between the distinct ϕ^i and ϕ^j in determining the mean square phase error. Each block is identical up to an overall scale factor, since the telescope aperture function is identical for each ϕ^i, but we may wish to assign different importance to each of the different phase profiles. The mean square phase error formula therefore takes the form

$$\phi^{\mathrm{T}} W \phi = \sum_{i=1}^{n} w_i (\phi^i)^{\mathrm{T}} V \phi^i, \tag{23}$$

where $w_1, \ldots, w_n$ are scalar weights and the matrix V is the same for each phase profile. The coefficients of V are defined so that the quantity $(\phi^i)^{\mathrm{T}} V \phi^i$ is equal to the mean square, piston-removed value of $\varphi^i(r)$ averaged over the continuous telescope aperture. With the telescope aperture function denoted as $A(r)$, the matrix V must satisfy the relationship

$$(\phi^i)^{\mathrm{T}} V \phi^i = \int \mathrm{d}r A(r) \left[\varphi^i(r) - \int \mathrm{d}r' A(r') \varphi^i(r') \right]^2, \tag{24}$$

where we have assumed for simplicity that the aperture function $A(r)$ has been normalized, so that

$$\int \mathrm{d}r A(r) = 1. \tag{25}$$

The coefficients of the matrix V may now be determined by substituting Eq. (22) into Eq. (24), expanding the square, and interchanging the order of summation and integration. The result is given by the formula

$$V = V_0 - V_1 V_1^{\mathrm{T}}, \tag{26}$$

where the coefficients of the matrix V_0 and the column vector V_1 are defined by the integrals

$$(V_0)_{ij} = \int \mathrm{d}r A(r) e_i(r) e_j(r), \tag{27}$$

$$(V_1)_i = \int \mathrm{d}r A(r) e_i(r). \tag{28}$$

The matrix V_0 is sparse, since it is assumed that the influence functions $e_i(r)$ are localized. On account of Eq. (23), the value of the overall weighting matrix W is now given by the formulas

$$W = W_0 - W_1 W_1^{\mathrm{T}}, \tag{29}$$

$$W_0 = \mathrm{diag}(w_1 V_0, \ldots, w_n V_0), \tag{30}$$

$$W_1 = \mathrm{diag}(\sqrt{w_1} V_1, \ldots, \sqrt{w_n} V_1). \tag{31}$$

(Note that the blocks appearing in the definition of the matrix W_1 are column vectors, not square matrices.) The key features of this representation are the facts that (1) the matrix W_0 is sparse and (2) the matrix W_1 is a low-rank matrix with a small number (n) of columns. This structure permits the efficient evaluation of $u = F_x v$, as described in Subsection 4.B.[24]

B. Solution Summary

According to Eq. (19) for the matrix F_x, solving the system $u = F_x v$ for u can be accomplished by first computing $u' = H_a^{\mathrm{T}} W H_x v$ and then solving $u = (H_a^{\mathrm{T}} W H_a + N_w N_w^{\mathrm{T}} + kI)^{-1} u'$. Both steps may be solved more efficiently by utilizing the facts that the matrices H_a, H_x, and W_0 are sparse and that the matrices W_1 and N_w are low-rank matrices with only a few columns.

The first of these two steps is conceptually simpler, since it does not involve a matrix inversion. To evaluate the value of u' defined by

$$u' = H_a^{\mathrm{T}} W H_x v, \tag{32}$$

we first rewrite the matrix $H_a^{\mathrm{T}} W H_x$ in the form

$$H_a^{\mathrm{T}} W H_x = M - U V^{\mathrm{T}}, \tag{33}$$

where by Eq. (29) the values of M, U, and V are defined by the expressions

$$M = H_a^{\mathrm{T}} W_0 H_x, \tag{34}$$

$$U = H_a^{\mathrm{T}} W_1, \tag{35}$$

$$V = H_x^{\mathrm{T}} W_1. \tag{36}$$

The matrix M is sparse because each of the matrices H_a, W_0, and H_x is sparse, and both U and V are low-rank matrices with only a few columns because the same is true of the matrix W_1. It follows that the vector u' can be computed efficiently by using the expression

$$u' = Mv - U(V^{\mathrm{T}} v), \tag{37}$$

where all of the matrix–vector operations involve sparse matrices or matrices with only a few rows or columns.

Solving the second step

$$u = (H_a^{\mathrm{T}} W H_a + N_w N_w^{\mathrm{T}} + kI)^{-1} u' \tag{38}$$

also depends upon a similar sparse-plus-low-rank representation

$$H_a^{\mathrm{T}} W H_a + N_w N_w^{\mathrm{T}} + kI = M - UV^{\mathrm{T}}, \tag{39}$$

where the values of the matrices M, U, and V are now given by

$$M = H_a^{\mathrm{T}} W_0 H_a + kI, \tag{40}$$

$$U = (H_a^{\mathrm{T}} W_1 \quad N_w), \tag{41}$$

$$V = (H_a^{\mathrm{T}} W_1 \quad -N_w). \tag{42}$$

The matrix M is sparse because H_a and W_0 are sparse, and the matrices U and V are once again low rank with only a small number of columns. Thanks to this representation, we may now apply the matrix inversion lemma

$$(M \mp UV^{\mathrm{T}})^{-1} = M^{-1} \pm M^{-1}U(I \mp V^{\mathrm{T}}M^{-1}U)^{-1}(M^{-1}V)^{\mathrm{T}}, \tag{43}$$

which may be verified by multiplying both sides of the equation by M and $M \mp UV^{\mathrm{T}}$ and then simplifying until an identity is obtained. Substituting Eqs. (39) and (43) back into Eq. (38) now yields the result

$$u = M^{-1}u' + ((M^{-1}U)\{(I - V^{\mathrm{T}}M^{-1}U)^{-1}[(M^{-1}V)^{\mathrm{T}}u']\}). \tag{44}$$

The matrices $M^{-1}U$ and $M^{-1}V$ can be efficiently precomputed, since M is sparse and the matrices U and V have only a small number of columns. The matrix $I - V^{\mathrm{T}}M^{-1}U$ can then be computed and inverted efficiently because both V and $M^{-1}U$ have only a small number of columns. Equation (44) can then be solved efficiently once these quantities are precomputed, since M is sparse and all operations following the addition symbol involve only matrices of low rank.

We have not formally evaluated the reduction in complexity that can be obtained by the above methods, but numerical results on actual computation times will be presented in Section 6. The greatest savings are actually in precomputing the matrices, since the number of operations necessary to compute $H_a^{\mathrm{T}}WH_x$ and $H_a^{\mathrm{T}}WH_a$ will scale as the *cube* of the dimension of a when utilizing conventional matrix multiplies.

5. EFFICIENT SOLUTION FOR $u = E_x v$

This part of the derivation begins by finding a more explicit representation for the phase estimation matrix E_x. Using Eq. (9), we may rewrite this matrix in the form

$$\begin{aligned} E_x &= C_{xs}C_{ss}^{-1} \\ &= C_{xx}G_x^{\mathrm{T}}(G_xC_{xx}G_x^{\mathrm{T}} + C_{nn})^{-1} \\ &= (G_x^{\mathrm{T}}C_{nn}^{-1}G_x + C_{xx}^{-1})^{-1}G_x^{\mathrm{T}}C_{nn}^{-1}, \end{aligned} \tag{45}$$

where the last equality may be verified by multiplying both sides of the expression by $G_xC_{xx}^{\mathrm{T}}G_x^{\mathrm{T}} + C_{nn}$ on the right and by $G_x^{\mathrm{T}}C_{nn}^{-1}G_x + C_{xx}^{-1}$ on the left and then further simplifying until an identity is obtained. Sparse-matrix methods are not immediately applicable to this representation for several reasons. Most important, the phase covariance matrix C_{xx} is in general nonsparse and in fact has infinite eigenvalues for the usual case of the Kolmogorov turbulence spectrum. An approximation to the *inverse* of C_{xx} must be chosen that is sparse and still fairly accurate. Second, the structures of the phase-to-WFS influence matrix G_x and the noise covariance matrix C_{nn} are complicated by the tilt uncertainty problem for LGSs and by the full-aperture tip/tilt NGS WFSs that must be included in LGS AO and MCAO systems as a consequence of this effect. Subsections 5.A and 5.B describe methods for coping with these complications, and Subsection 5.C then outlines how $u = E_x v$ can be efficiently solved for u.

A. Approximating the Regularizing Term C_{xx}^{-1}

1. Eliminating the Kolmogorov Singularity

For a pure Kolmogorov turbulence spectrum, the phase covariance matrix C_{xx} is not defined because the pure piston mode will have an infinite variance for each turbulence layer.[25] Even if we restrict attention to a von Kármán spectrum with a large but finite outer scale, the matrix C_{xx}^{-1} will be ill conditioned, with the pure piston mode of each turbulence layer approximating an eigenvector with a near-zero eigenvalue. The pure piston mode for each turbulence layer is also an eigenvector with a zero eigenvalue for the matrix $G_x^{\mathrm{T}}C_{nn}^{-1}G_x$, since these modes lie in the null space of G_x for any existing or postulated wave-front sensing concept. The purpose of this subsection is to describe how to add an additional sparse regularizing term to the sum $G_x^{\mathrm{T}}C_{nn}^{-1}G_x + C_{xx}^{-1}$ so that this matrix, which appears in the definitions of E_x, will be better conditioned.

To derive this regularizing term, we first define the matrices L and Z by the equations

$$L_{ij} = \begin{cases} 1 & \text{if phase point } i \text{ is located on screen } j \\ 0 & \text{otherwise} \end{cases}, \tag{46}$$

$$Z_{ij} = \begin{cases} 1 & \text{if phase point } i \text{ is the origin of screen } j \\ 0 & \text{otherwise} \end{cases}. \tag{47}$$

Each of these matrices has a number of rows equal to the total dimension of the phase error vector x and a number of columns equal to the number of turbulence layers. $G_xL = 0$, since each column of the matrix L is a pure piston mode for one of the turbulence layers, and we note also that $Z^{\mathrm{T}}L = I$. Consequently, we may write

$$\begin{aligned} (G_x^{\mathrm{T}}C_{nn}^{-1}G_x + C_{xx}^{-1} + ZZ^{\mathrm{T}})(I - LZ^{\mathrm{T}}) &= G_x^{\mathrm{T}}C_{nn}^{-1}G_x + C_{xx}^{-1} - C_{xx}^{-1}LZ^{\mathrm{T}} \\ &\approx G_x^{\mathrm{T}}C_{nn}^{-1}G_x + C_{xx}^{-1}, \end{aligned} \tag{48}$$

where the approximation follows because the pure piston modes for each turbulence layer are approximately eigenvalues for C_{xx}^{-1} with near-zero eigenvalues. Rearranging this expression yields the relationship

$$(G_x^{\mathrm{T}}C_{nn}^{-1}G_x + C_{xx}^{-1} + ZZ^{\mathrm{T}})^{-1} \approx (I - LZ^{\mathrm{T}})(G_x^{\mathrm{T}}C_{nn}^{-1}G_x + C_{xx}^{-1})^{-1}. \tag{49}$$

The right-hand side of this expression consists of the matrix inverse appearing in the expression for E_x on the last line of Eq. (45), followed by the term $I - LZ^{\mathrm{T}}$. The action of this second operator is to adjust the overall piston term of each phase screen layer to obtain a value of zero at the origin, which will have no impact upon the piston-removed accuracy of the phase estimate. The left-hand side of relation (49) may therefore be used instead of the term $(G_x^{\mathrm{T}} C_{nn}^{-1} G_x + C_{xx}^{-1})^{-1}$ to obtain a slightly modified version of the phase estimator E_x with improved conditioning. The possible loss of performance due to the approximation made in Eq. (48) will be investigated by means of simulations in Section 6.

2. Sparse-Covariance-Matrix Approximation

For the Kolmogorov turbulence spectrum, the covariance matrix C_{xx} and its inverse will be nonsparse and of full rank, so an approximation of some sort must be made to proceed by using the sparse techniques proposed by this paper. To justify our approximation, we consider the case of a single continuous turbulence layer of infinite extent. In this limit case, the bilinear functional defined by the matrix C_{xx}^{-1} may be approximated as

$$\begin{aligned} u^{\mathrm{T}} C_{xx}^{-1} v &= u^{\mathrm{T}} \langle x x^{\mathrm{T}} \rangle^{-1} v \\ &= \int\int \mathrm{d}r \mathrm{d}r' u(r) v(r') [\langle x(r) x^{\mathrm{T}}(r') \rangle^{-1}] \\ &= \int\int \mathrm{d}\kappa \mathrm{d}\kappa' \hat{u}(\kappa) \hat{v}^*(\kappa') [\langle \hat{x}(\kappa) \hat{x}^*(\kappa') \rangle^{-1}] \\ &\propto \int \mathrm{d}\kappa \hat{u}(\kappa) \hat{v}^*(\kappa) \kappa^{11/3} \\ &\approx \int \mathrm{d}\kappa [\kappa^2 \hat{u}(\kappa)] [\kappa^2 \hat{v}(\kappa)]^* \\ &\propto \int \mathrm{d}r \nabla^2 u(r) \nabla^2 v(r), \end{aligned} \tag{50}$$

where the spatial Fourier transform of a function f has been denoted as $\hat{f}$. The second line follows because we have assumed an infinite, continuous phase screen, the third follows from the Plancheral theorem, and the fourth follows from the definition of the Kolmogorov turbulence spectrum. The fifth line is based upon the approximation $11/3 \approx 4$, and the final line follows because $(\widehat{\mathrm{d}f/\mathrm{d}x}) = 2\pi\kappa_x i \hat{f}$. In other words, the quantity $u C_{xx}^{-1} v^{\mathrm{T}}$ is approximately proportional to the inner product between the Laplacians, or curvatures, of u and v.

Since this heuristic argument is valid for all choices of u and v, it suggests that we can approximate C_{xx}^{-1} in the case of a single discrete turbulence layer as

$$C_{xx}^{-1} \approx C^{\mathrm{T}} C, \tag{51}$$

where C is proportional to a discrete approximation of the Laplacian operator as illustrated in Fig. 3, with a constant of proportionality depending upon the strength of the turbulence. For a multilayer turbulence profile, the matrix C becomes a weighted sum of one such term per layer. The approximation to C_{xx}^{-1} is quite sparse, since the matrix C has no more than five nonzero elements per row. This computational simplification was first suggested and evaluated in Ref. 18 for the case of a conventional NGS AO system, but the motivation provided by Eq. (50) and the extension to more than one atmospheric turbulence layer were both overlooked. We note that the boundary conditions on C illustrated in Fig. 3 were chosen somewhat arbitrarily, but based upon the numerical results presented in Section 6, there is no motivation for optimizing them further.

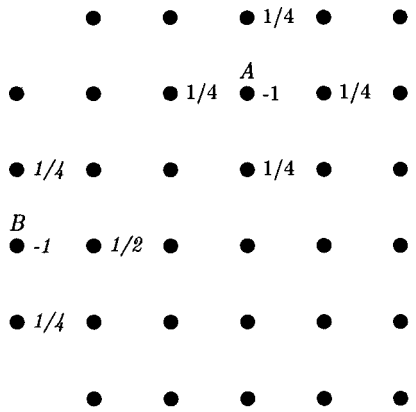

Fig. 3. Discrete Laplacian operator. This figure illustrates the coefficients for two rows of the discrete Laplacian, or curvature, operator C appearing in relation (51). The dots represent the grid points of the discrete turbulence layer. The values printed in a regular font are the five nonzero coefficients needed to compute the curvature of the phase profile at the interior grid point A. The italicized values are the nonzero coefficients needed to compute the curvature at the grid point B on the boundary of the phase profile. The coefficients that should be assigned to grid points lying outside the boundary have been "folded over" back into the grid so that the sum of the coefficients remains zero.

When we combine this approximation with the discussion in Subsection 5.A.1, the phase estimation matrix that we propose to evaluate is given by the expression

$$E_x' = (G_x^{\mathrm{T}} C_{nn}^{-1} G_x + C^{\mathrm{T}} C + Z Z^{\mathrm{T}})^{-1} G_x^{\mathrm{T}} C_{nn}^{-1}. \tag{52}$$

The remainder of this section describes how E_x' can be evaluated by using sparse matrix techniques, and Section 6 assesses the loss in phase estimation accuracy incurred by making these approximations.

B. Matrix Structure for Laser Guide Stars

As mentioned above, the noise covariance matrix C_{nn} is no longer diagonal for the case of LGS WFS measurements because of the effect of LGS position uncertainty. The exact position of a LGS projected into the sky is variable as a result of both fundamental effects of atmospheric turbulence and practical error sources in the laser system, and (at least for now) there is no independent means of measuring the actual position with any accuracy. This motion of the guide star is indistinguishable from overall wave-front tilt to the WFS, resulting in an additional source of measurement noise that is fully correlated between all of the subapertures of a particular WFS. The matrices C_{nn} and $G_x^{\mathrm{T}} C_{nn}^{-1} G_x$ are no longer sparse when this noise term is included, complicating the application of sparse methods to the evaluation of the phase estimation matrix E_x'.

More important from a practical perspective, one or more NGSs must always be included in the guide star

constellation of a LGS AO system to measure the overall tip/tilt mode of the wave-front error, since the tip/tilt measurement from the LGS WFS is very noisy because of the uncertainty of the guide star position. The overall tip/tilt measurement from a NGS WFS with only one or a few subapertures will depend upon phase values distributed across the entire telescope aperture, further degrading the sparseness of the term $G_x^{\mathrm{T}} C_{nn}^{-1} G_x$ appearing in the formula for E_x'. Fortunately, both of these effects are low-rank adjustments to the matrix, permitting the use of the matrix inversion lemma as used in Section 4. The remainder of this subsection introduces the notation required to discuss the NGS and LGS components of the WFS measurement vector s and describes the form of the noise covariance matrix C_{nn}^{-1} when LGS position uncertainty is present.

For the most general case of either a conventional AO or a MCAO system using a combination of NGSs and/or LGSs, it is still possible to decompose the WFS measurement vector s into a higher-order component s_h and a tip/tilt component s_t. The component s_h includes the measurements from all WFSs with two or more subapertures, while s_t consists of the tip/tilt measurements from NGS sensors with only a single subaperture. The dimensionality of s_t is consequently equal to twice the number of tip/tilt NGSs, typically much smaller than the dimensionality of the higher-order component s_h. With this definition the WFS measurement model defined by Eq. (9) can be rewritten as

$$s = \begin{pmatrix} s_h \\ s_t \end{pmatrix} = \begin{pmatrix} G_h \\ G_t \end{pmatrix} x + \begin{pmatrix} n_h \\ n_t \end{pmatrix}, \tag{53}$$

where the dimensionality of n_t and the number of rows of G_t are also equal to twice the number of tip/tilt NGSs. The matrix G_t is consequently of low rank relative to the overall size of the matrix G_x.

With the WFS measurement noise written as in Eq. (53), the form of the noise covariance matrix now becomes

$$C_{nn} = \begin{bmatrix} N_h + \sigma_t^2 T T^{\mathrm{T}} & 0 \\ 0 & N_t \end{bmatrix}. \tag{54}$$

The terms N_h and N_t describe the statistics of the noise within the higher-order and tip/tilt WFSs themselves and are diagonal matrices. The effect of LGS position uncertainty is captured in the term $\sigma_t^2 T T^{\mathrm{T}}$, where σ_t is the rms one-axis position uncertainty for each LGS, and the columns of the matrix T are the modes of WFS measurement noise induced by the LGS position errors. There are two such modes for each LGS, corresponding to the tip and the tilt (or x and y) position errors. Each column of T is $\{0, 1\}$ valued, with the 1s matching those elements of s_h that are the x or y measurements for a particular LGS. The point for the current discussion is that the matrix T is of low rank, since the number of LGS is very small compared with the total number of WFS measurements.

(The above representations of s and C_{nn} can also be applied to the case of a conventional AO or MCAO system with only higher-order NGS WFS measurements, simply by taking $\sigma_t = 0$ and setting G_t and N_t to be empty matrices.)

The inverse of the noise covariance matrix C_{nn} will be needed to evaluate the phase estimation matrix E_x'. As above, the nonsparse term in this inverse may be evaluated by using the matrix inversion lemma, yielding the result

$$\begin{aligned} (N_h + \sigma_t^2 T T^{\mathrm{T}})^{-1} &= N_h^{-1} - \sigma_t^2 N_h^{-1} T (I + \sigma_t^2 T^{\mathrm{T}} N_h^{-1} T)^{-1} (N_h^{-1} T)^{\mathrm{T}} \\ &\to N_h^{-1} - N_h^{-1} T (T^{\mathrm{T}} N_h^{-1} T)^{-1} (N_h^{-1} T)^{\mathrm{T}} \\ &\text{as } \sigma_t^2 \to \infty. \end{aligned} \tag{55}$$

The assumption that the rms LGS position uncertainty is effectively infinite is only slightly conservative in practical applications; σ_t will be no smaller than the actual rms value of the turbulence-induced tip/tilt errors unless the size of the laser launch telescope is actually larger than the aperture of the AO system. The measured position of the LGS is consequently of little or no use in estimating the tip/tilt error, which is the motivation for including one or more NGSs as part of the guide star constellation in the first place. The computational advantage gained by treating σ_t as infinite is that the second line of Eq. (55) is a sparse matrix with a low-rank adjustment, since N_h is diagonal and the matrix T has only a few columns.

C. Solution Summary

With the above preparations, we are finally ready to describe the efficient evaluation of $u = E_x' v$ using sparse-matrix methods. As in Section 4 for the actuator fitting operator F_x, the solution is given in two steps. The first step is the solution of the system

$$u' = G_x^{\mathrm{T}} C_{nn}^{-1} v. \tag{56}$$

Using the representations of G_x and C_{nn}^{-1} developed in Eqs. (53)–(55), we may rewrite the operator appearing in Eq. (56) in the form

$$G_x^{\mathrm{T}} C_{nn}^{-1} = [G_h^{\mathrm{T}} N_h^{-1} (I - T P_T) \quad G_t^{\mathrm{T}} N_t^{-1}], \tag{57}$$

where the noise-weighted LGS tilt projection operator P_T is defined as

$$P_T = (T^{\mathrm{T}} N_h^{-1} T)^{-1} T^{\mathrm{T}} N_h^{-1}. \tag{58}$$

Proceeding as in Section 4 for the actuator fitting problem, we now write the matrix $G_x^{\mathrm{T}} C_{nn}^{-1}$ in the form

$$G_x^{\mathrm{T}} C_{nn}^{-1} = M - U V^{\mathrm{T}}, \tag{59}$$

where the terms M, U, and V are defined by the formulas

$$M = [G_h^{\mathrm{T}} N_h^{-1} \quad G_t^{\mathrm{T}} N_t^{-1}], \tag{60}$$

$$U = G_h^{\mathrm{T}} N_h^{-1} T, \tag{61}$$

$$V = \begin{pmatrix} P_T^{\mathrm{T}} \\ 0 \end{pmatrix} = \begin{pmatrix} N_h^{-1} T (T^{\mathrm{T}} N_h^{-1} T)^{-1} \\ 0 \end{pmatrix}. \tag{62}$$

The matrix M is sparse because G_h is sparse, N_h is diagonal, and G_t is low rank (with only a few rows). The matrices U and V are low rank because the matrix T has only a few columns. This representation allows u' to be computed efficiently from v by using Eq. (37).

[In passing, it may be of interest to note that the operator $I - T P_T$ appearing in Eq. (57) has the effect of sub-

tracting off the noise-weighted overall wave-front tilt from the LGS WFS measurements as the first step of the wave-front reconstruction algorithm. This preprocessing step, derived here by using the matrix inversion lemma, may be considered intuitively obvious and is already found in the reconstruction algorithms used for many LGS AO systems and simulations.]

The second and final step in evaluating $u = E_x' v$ is to efficiently solve the system

$$u = (G_x^{\mathrm{T}} C_{nn}^{-1} G_x + C^{\mathrm{T}} C + ZZ^{\mathrm{T}})^{-1} u'. \tag{63}$$

Using Eqs. (53) and (57), we may write

$$G_x^{\mathrm{T}} C_{nn}^{-1} G_x + C^{\mathrm{T}} C + ZZ^{\mathrm{T}} = M - UV^{\mathrm{T}}, \tag{64}$$

where the matrices M, U, and V are this time defined as

$$M = G_h^{\mathrm{T}} N_h^{-1} G_h + C^{\mathrm{T}} C + ZZ^{\mathrm{T}}, \tag{65}$$

$$U = [G_h^{\mathrm{T}} N_h^{-1} T \quad -G_t^{\mathrm{T}} N_t^{-1}], \tag{66}$$

$$V = [G_h^{\mathrm{T}} P_T^{\mathrm{T}} \quad G_t^{\mathrm{T}}]. \tag{67}$$

The matrix M is sparse because all of the matrices appearing on the right-hand side of Eq. (65) are sparse. The matrices U and V are low rank with only a few columns because the same is true of the matrices T, G_t^{T}, and P_T^{T}. We may therefore proceed as in Section 4 for the actuator fitting operator, applying the matrix inversion lemma and Eq. (44) to solve for u in terms of u' by using operations that involve only low-rank matrices and back-substitutions through the sparse matrix M.

6. SAMPLE NUMERICAL RESULTS

This section summarizes initial numerical wave-front reconstruction results obtained by using the sparse computational methods developed in Sections 3–5. The cases considered have been chosen to address four particular questions:

- How do the approximations introduced in Section 5 with respect to atmospheric turbulence statistics degrade the accuracy of wave-front reconstruction?
- How large (or small) is the improvement in computational efficiency relative to the performance of the conventional matrix multiply reconstructor (CMMR)? We will concentrate on the time necessary to precompute the reconstructor, since this is generally the limiting factor in performing AO simulations.
- How large are the computer memory requirements to store the sparse reconstruction algorithm (SRA)? Memory requirements are a concern because the SRA computes an estimate of the full atmospheric turbulence profile x, which may have a significantly larger dimensionality than that of the DM actuator command vector a.
- What new results can be obtained about the performance of ExAO and MCAO systems for extremely large telescopes, assuming that the SRA permits simulations of high-order AO systems that were not previously possible?

The results obtained regarding these questions are summarized in Subsections 6.A–6.D. Subsection 6.A outlines the cases considered and some of the details of the simulation code. Subsection 6.B summarizes results for low-order NGS AO systems, which indicate that the loss in performance due to approximating atmospheric turbulence statistics is entirely negligible. Subsections 6.C and 6.D then present results for extreme NGS AO and MCAO systems. These results indicate that the reduction in computation times is very significant, the computer memory requirements are generally acceptable, and the trends in ExAO and MCAO performance with increasing AO system order are very gradual.

A. Cases Considered

Table 1 presents the atmospheric turbulence profile used for the simulations described in this section. This six-layer profile is based upon thermosonde and generalized scintillation detection and ranging (known as SCIDAR) measurements taken at Cerro Pachon, Chile, the site of the Gemini-South telescope.[26] The profile has been scaled to obtain an r_0 of 16 cm at a wavelength of 0.50 μm, corresponding roughly to median conditions at Cerro Pachon. The resulting value of θ_0 is 2.65 arc sec, or 12.85 μrad.

Table 2 summarizes the sample AO systems that have been evaluated. These include conventional NGS AO systems with telescope aperture diameters of 4, 8, 16, and 32 m and NGS and LGS MCAO systems with aperture diameters of 8 and 16 m. The WFS subaperture width and the DM actuator pitch for the conventional NGS AO systems were 0.5 m, with performance evaluated for a single evaluation direction and an on-axis guide star. For the MCAO system, performance was evaluated at the center, the edges, and the corners of a 1-arc min-square field of view, with the weights w_i assigned to the nine evaluation directions in Eq. (23) determined by using Simpson's rule. Five higher-order guide stars were located at the center and the corner of the 1-arc min field for both the NGS and LGS MCAO systems. For the LGS MCAO system, these guide stars were placed at a range of 90 km, and four tip/tilt NGSs were also located at the edges of the 1-arc min field. The subaperture width was also equal to 0.5 m for the higher-order WFSs in both MCAO systems. Both MCAO systems also included two DMs at conjugate ranges of 0.0 and 5.16 km with actuator pitches of 0.5 m and a third DM at a range of 10.31 km with a 1.0-m actuator pitch.

The code written to evaluate the sparse reconstruction algorithm (SRA) for these AO systems consisted of three basic parts. The first part computed the influence matrices H_x, H_a, and G_x, the phase error weighting matrix W,

Table 1. Atmospheric Turbulence Profile Used for Simulations[a]

Layer	Altitude (km)	Relative Layer Weight
1	0.00	0.652
2	2.58	0.172
3	5.16	0.055
4	7.73	0.025
5	12.89	0.074
6	15.46	0.022

[a] See the text for further details.

Table 2. Simulated AO System Parameters[a]

System	NGS AO	MCAO
Evaluation field	on axis	60-arc sec square
Aperture diameter (m)	4, 8, 16, 24, 32	8, 16
DM conjugate ranges (km)	0.0	0.0, 5.16, 10.31
DM interactuator spacing (m)	0.5	0.5, 0.5, 1.0
Higher-order guide stars		
Number	1	5
Subaperture spacing (m)	0.5	0.5
Directions (arc min)	(0, 0)	(0, 0) and ($\pm$0.5, $\pm$0.5)
Range (km)	∞	∞ (NGS) or 90 (LGS)
Tip/tilt guide stars		
Number	0	0 (NGS) or 4 (LGS)
Directions (arc min)	—	($\pm$0.5, 0) and (0, $\pm$0.5)
Range (km)	—	∞

[a] See the text for further details.

and the approximate regularization matrix $C^{\mathrm{T}}C$ for the given turbulence profile, aperture geometry, DM configuration, and guide star constellation. As illustrated in Fig. 2, the first four matrices were computed by assuming linear spline influence functions. The matrix elements corresponding to boundary actuators, phase points, and subapertures were computed with the circular telescope aperture taken into account. The exact turbulence covariance matrix C_{xx} was also computed (with the full-aperture piston mode removed to eliminate the singularity) for simulation cases with an aperture diameter of 8 m or less.

The second part of the code then computed the SRA, and for the 4- and 8-m cases, also the CMMR. The final step was to evaluate the SRA, and for small cases also the CMMR, through Monte Carlo simulation by using 100 turbulence profile realizations generated as the Fourier transforms of white noise filtered by the Kolmogorov spectrum. Identical profiles were used to evaluate the SRA and the CMMR for the 4- and 8-m cases.

The atmospheric turbulence profile sampling used to compute the reconstructors was at twice the spatial resolution of the DM actuators, i.e., grid point spacings of either 0.25 or 0.5 m for the six turbulence layers. The simulations themselves used grids with spatial resolutions 16 times finer than the DM actuator spacings, i.e., 0.03125 or 0.0625 m. A few preliminary simulation results indicated that further increasing the resolution used to compute the reconstructors had a negligible effect on estimation accuracy and that the "simulation fitting error" arising from unsampled high-spatial-frequency turbulence scaled as the two-thirds power of the grid spacing used in the simulation. The author believes that this simulation fitting error results from omitting the spatial aliasing of the highest-spatial-frequency turbulence from the WFS measurements, which is a source of measurement error. With this scaling law, the magnitude of this error with the chosen simulation grid spacings is estimated to be a mean square optical path difference of approximately 0.0028 μm^2, which is small compared with the effects studied in the following simulations.

Table 3. Reconstructor Performance versus WFS Measurement Noise for an Order 8 $\times$ 8 NGS AO System[a]

WFS Noise (arc sec)	CMMR σ^2 (μm^2)	SRA σ^2 (μm^2)
0.02	0.01304	0.01306
0.04	0.01734	0.01737
0.08	0.03218	0.03229
0.16	0.07834	0.07941

[a] This table compares the residual mean square wave-front error σ^2 due to the combined effects of fitting error and WFS measurement noise for the conventional matrix multiply reconstruction (CMMR) implementation of the minimum-variance estimator and the sparse reconstruction algorithm (SRA) described in this paper. The approximations made in modeling atmospheric turbulence statistics for the latter algorithm have only a very modest effect on the mean square wave-front estimation error σ^2.

B. Results for Low-Order Natural Guide Star Adaptive Optical Systems

For an initial performance comparison between the SRA and the CMMR, the first AO system studied was a conventional, narrow-field-of-view NGS AO system of order 8 $\times$ 8 with an on-axis guide star. The sources of wave-front error in these simulations were consequently WFS measurement noise and DM/WFS fitting error, and a range of WFS noise levels was considered such that the rms wave-front estimation error due to measurement noise varied from approximately one half to three times the rms fitting error. The results of these simulations are presented in Table 3. Note that four significant digits are necessary to detect the performance variations between the SRA and the CMMR at low noise levels, and at the largest noise level considered, the values of σ^2 for the two approaches still agree to within approximately 1%. At least in this case, the degradation in wave-front estimation accuracy due to the approximations made in deriving the SRA will be negligible for practical applications.

C. Results for Extreme Natural Guide Star Adaptive Optical Systems

Next, the order of the NGS AO system was varied for a fixed, small value of WFS measurement noise. The results obtained are summarized in Table 4. For an order 16×16 system, it is still possible to compute the CMMR without significant difficulty, and in this case the performance difference between the SRA and the CMMR remains very small. For the SRA the value of σ^2 grows by approximately 8% as the order of the NGS AO system increases from 16×16 to 64×64. This trend indicates that the algorithm remains numerically stable. It is also qualitatively consistent with the expected slow growth in reconstructor noise gain as a function of AO system order, although these values have not been compared explicitly with standard scaling laws.[12]

We note that the time required to compute the SRA increases somewhat more slowly than the three-halves power of the number of DM actuators and that the time needed to compute the reconstructor for an order 64×64 NGS AO system is approximately 9 s on a 1-GHz Pentium III using Matlab 6.[27] Significantly larger cases could easily be evaluated for truly ExAO simulations. The time necessary to apply the reconstructor has not been recorded, but this will be essentially proportional to the number of nonzero coefficients in the sparse representations of the estimation and fitting matrices E_x and F_x, respectively, or, equivalently, to the amount of memory needed to store these matrices. These storage requirements increased by a factor of approximately 26, while the number of DM actuators increased by a factor of $3461/257 = 13.5$. For the CMMR the corresponding increase would be approximately $13.5^2 = 181$.

D. Results for Natural Guide Star and Laser Guide Star Multicomponent Adaptive Optical Systems

Finally, Table 5 summarizes the results obtained on wave-front estimation error and computational efficiency for the NGS and LGS MCAO systems with parameters as described in Table 2. Systems of aperture diameters 8 and 16 m with a fixed WFS subaperture size of 0.5 m have been considered, yielding AO system orders of 16×16 and 32×32, respectively. For the order 16×16 case, it is possible to compute and simulate both the CMMR and SRA methods, and the mean square estimation errors for the two techniques are again virtually identical. We note that the simulated performance results for the SRA are actually very slightly superior to those for the CMMR in the LGS MCAO case, presumably because of the effects of finite numerical precision in evaluating the reconstructors and the finite number of simulation trials used to obtain the performance estimates. Also, the performance of the SRA is almost independent of telescope aperture diameter for both NGS AO and LGS MCAO systems, suggesting that results for a 32-m, order 64×64 system may likely be similar as well. The 32-m case could not be explicitly evaluated at this time because of computer memory limitations.

Table 4. SRA Wave-Front Fitting Error versus Order of Correction for a Conventional NGS AO System[a]

Aperture Diameter (m)	Order of Correction	DM Actuators	CMMR σ^2 (μm^2)	SRA σ^2 (μm^2)	SRA Computation Time (s)	SRA Memory Requirements (Mbytes)
8	16×16	257	0.01334	0.01336	0.25	0.72
16	32×32	921	—	0.01392	1.17	3.81
24	48×48	1981	—	0.01416	3.60	NA
32	64×64	3461	—	0.01440	8.94	19.30

[a] This table summarizes the performance of the SRA for a conventional NGS AO scenario where the only significant source of wave-front error is the finite spatial resolution of the DM actuators and the WFS subapertures. The NGS is coincident with the evaluation direction, and the noise equivalent angle for the WFS is an almost negligible 0.02 arc sec. The DM actuator spacing is held constant, so the AO order of correction is proportional to the telescope aperture diameter. The SRA estimation error is virtually identical with the minimum-variance CMMR for the case of an order 16×16 AO system and increases fairly gradually with increasing telescope aperture diameter. The computation times and the memory requirements for the SRA grow much less rapidly than the $O(n^3)$ and $O(n^2)$ scaling laws that apply for the case of the CMMR.

Table 5. Results and Scaling Law Predictions for CMMR and SRA Performance for MCAO Systems[a]

	NGS MCAO			LGS MCAO		
Aperture diameter (m)	8	16	32	8	16	32
System order	16×16	32×32	64×64	16×16	32×32	64×64
Total DM actuators	789	2417	(9700)	789	2417	(9700)
Total WFS subapertures	1020	4280	(17120)	1020	4280	(17120)
CMMR σ^2 (μm^2)	0.01904	—	—	0.02220	—	—
Time to compute CMMR (h)	0.51	(14.66)	(947.66)	0.51	(14.66)	(947.66)
Memory to store CMMR (Mbytes)	14	(129)	(2071)	13.8	(130)	(2086)
SRA σ^2 (μm^2)	0.01932	0.01877	—	0.02190	0.02220	—
Time to compute SRA (h)	0.19	1.92	(33.90)	0.21	1.68	(22.20)
Memory to store SRA (Mbytes)	112	680	(6380)	99	561	(4852)

[a] The simulated MCAO parameters are summarized in Table 2. The parenthesized values are extrapolations based upon the standard $O(n^2)$ and $O(n^3)$ power laws for the CMMR and two-point power-law curve fits for the SRA.

The computer memory requirements for the MCAO cases are very significantly greater than those for the ExAO case considered in Table 4 because (1) the entire three-dimensional turbulence profile must be evaluated for the MCAO case and (2) the DM-to-WFS influence matrix G_x and the matrices derived from it are much less sparse thanks to the presence of multiple guide stars distributed over an extended field of view. The memory requirements predicted for the 32-m case still appear to be feasible for existing computer systems but not for the personal computers available for this study.

The increased fill factor for these matrices has negative implications for the effort required to compute the SRA as well, but the time required for the 16-m, order 32×32 system with 2417 DM actuators and 4284 total WFS subapertures is still on the order of 1.7–1.9 h, representing a factor-of-8 improvement over that for the CMMR. Extrapolating these results by means of power laws to an order 64×64 MCAO system yields a predicted improvement factor of approximately 30 or 40 to 1. For the clock speed of the Pentium III used here, this would reduce the time needed to compute the reconstructor from approximately 1 month to the order of 1 day.

ACKNOWLEDGMENTS

This research was supported by the Gemini Observatory, which is operated by the Association of Universities for Research in Astronomy, Inc., under a cooperative agreement with the National Science Foundation (NSF) on behalf of the Gemini Partnership: the United States National Science Foundation, the United Kingdom Particle Physics and Astronomy Research Council (PPARC), the Canadian National Research Council (NRC), the Chilean Comisión Nacional de Investigación Cientifica y Tecnológica (CONICYT), the Australian Research Council (ARC), the Argentinean Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), and the Brazilian Conselho Nacional de Desenvolvimento Científico e Tecnoloógico (CNPq).

Brent L. Ellerbroek may be reached by e-mail at bellerbroek@gemini.edu.

REFERENCES AND NOTES

1. J. W. Hardy, J. E. Lefevbre, and C. L. Koliopoulous, "Real time atmospheric turbulence compensation," J. Opt. Soc. Am. **67**, 360–369 (1977).
2. G. Rousset, J.-C. Fontanella, P. Kern, P. Gigan, F. Rigaut, P. Lena, C. Boyer, P. Jagourel, J.-P. Gaffard, and F. Merkle, "First diffraction-limited astronomical images with adaptive optics," Astron. Astrophys. **230**, 29–32 (1990).
3. M. C. Roggemann and B. Welsh, *Imaging through Turbulence* (CRC Press, Boca Raton, Fla., 1996).
4. D. L. Fried, "Least-squares fitting a wave-front distortion estimate to an array of phase difference measurements," J. Opt. Soc. Am. **67**, 370–375 (1977).
5. R. Hudgin, "Wave-front reconstruction for compensated imaging," J. Opt. Soc. Am. **67**, 375–378 (1977).
6. J. Hermann, "Least-squares wave-front errors of minimum norm," J. Opt. Soc. Am. **70**, 28–35 (1980).
7. E. P. Wallner, "Optimal wave-front correction using slope measurement," J. Opt. Soc. Am. **73**, 1771–1776 (1983).
8. B. M. Walsh and C. S. Gardner, "Effects of turbulence-induced anisoplanatism on the imaging performance of adaptive-astronomical telescopes using laser guide stars," J. Opt. Soc. Am. A **8**, 69–80 (1991).
9. G. A. Tyler, "Merging: a new method for tomography through random media," J. Opt. Soc. Am. A **11**, 409–424 (1994).
10. J. M. Beckers, "Increasing the size of the isoplanatic patch with multi-conjugate adaptive optics," in *Proceedings of European Southern Observatory Conference and Workshop on Very Large Telescopes and Their Instrumentation*, M.-H. Ulrich, ed., Vol. 30 of ESO Conference and Workshop Proceedings (European Southern Observatory, Garching, Germany, 1988), pp. 693–703.
11. D. C. Johnston and B. M. Welsh, "Analysis of multi-conjugate adaptive optics," J. Opt. Soc. Am. A **11**, 394–408 (1994).
12. B. L. Ellerbroek, "First order performance evaluation of adaptive-optics systems for atmospheric turbulence compensation in extended field-of-view astronomical telescopes," J. Opt. Soc. Am. A **11**, 783–805 (1994).
13. T. Fusco, J.-M. Conan, G. Rousset, L. M. Mugnier, and V. Michau, "Optimal wave-front reconstruction strategies for multi-conjugate adaptive optics," J. Opt. Soc. Am. A **18**, 2527–2538 (2001).
14. B. L. Ellerbroek, "Methods for correcting tilt anisoplanatism in laser-guide-star-based multi-conjugate adaptive optics," J. Opt. Soc. Am. A **18**, 2539–2547 (2001).
15. G. M. Cochran, "Sparse matrix techniques in wavefront reconstruction," Rep. TR-668 (Optical Sciences Company, Anaheim, Calif., 1986).
16. L. Schmutz, B. M. Levine, A. Wirth, and C. Strandley, "Adaptive optics concepts for extremely large aperture telescopes," in *Bäckaskog Workshop on Extremely Large Telescopes*, T. Andersen, A. Ardeberg, and R. Gilmozzi, eds., Vol. 57 of ESO Conference and Workshop Proceedings (European Southern Observatory, Garching, Germany, 1999), pp. 217–223.
17. L. Gilles, C. R. Vogel, and B. L. Ellerbroek, "Iterative algorithms for large scale wave front reconstruction," in *Signal Recovery and Synthesis*, Vol. 67 of OSA Trends in Optics and Photonics (Optical Society of America, Washington D.C., 2001), pp. 100–101.
18. B. L. Ellerbroek, "Comparison of least squares and minimum variance wavefront reconstruction for atmospheric turbulence compensation in the presence of noise," Rep. TR721R (Optical Sciences Company, Anaheim, Calif., 1986).
19. G. M. Cochran, "Sparse matrix techniques applied to deconvolution," Comput. Elect. Eng. **18**, 499–505 (1992).
20. R. Flicker, F. J. Rigaut, and B. L. Ellerbroek, "Comparison of multi-conjugate adaptive optics configurations and control algorithms for the Gemini South 8-m telescope," in *Adaptive Optical Systems Technology*, P. L. Wizinowich, ed., Proc. SPIE **4007**, 1032–1043 (2000).
21. G. H. Golub and C. F. van Loan, *Matrix Computations* (John Hopkins U. Press, Baltimore, Md., 1996), p. 18.
22. A. George and J. Liu, *Computer Solutions of Large Symmetric Positive Definite Systems* (Prentice-Hall, Englewood Cliffs, N.J., 1981).
23. S. Pissantesky, *Sparse Matrix Technology* (Academic, Orlando, Fla., 1984), Chap. 4.
24. The weighting matrix W derived in this section includes the global tip and tilt modes in the calculation of the mean square wave-front error. This is appropriate for AO applications to long-exposure imaging, where random, time-varying tip and tilt errors will degrade image quality in the same fashion as higher-order wave-front errors. For short-exposure imaging it may be more desirable to consider tip/tilt and higher-order wave-front aberrations separately. In this case the weighting matrix W for the mean square,

higher-order wave-front error will still take the form defined by Eqs. (29)–(31), except that the matrix V_1 now has three columns instead of one. The computational methods developed in the remainder of this paper are consequently still applicable.

25. R. J. Noll, "Zernike polynomials and atmospheric turbulence," J. Opt. Soc. Am. **66**, 207–211 (1976).
26. J. Vernin, A. Agabi, R. Avila, M. Azouit, R. Conan, F. Martin, E. Masciadri, L. Sanchez, and A. Ziad, "1998 Gemini site testing campaign: Cerro Pachon and Cerro Tololo," Gemini Doc. RTP-AO-G0094 (Gemini Observatory, Hilo, Hawaii, 2000).
27. Matlab 6 no longer supports the FLOPS command included in previous versions to report the exact number of floating-point operations required by an algorithm. No other computationally significant tasks were running on this dual-processor system during the timing tests reported in this paper.