Local, hierarchic, and iterative reconstructors for adaptive optics

Douglas G. MacMartin

Department of Control and Dynamical Systems, California Institute of Technology, Pasadena, California 91125

Adaptive optics systems for future large optical telescopes may require thousands of sensors and actuators. Optimal reconstruction of phase errors using relative measurements requires feedback from every sensor to each actuator, resulting in computational scaling for n actuators of n^2. The optimum local reconstructor is investigated, wherein each actuator command depends only on sensor information in a neighboring region. The resulting performance degradation on "global" modes is quantified analytically, and two approaches are considered for recovering global performance. Combining local and global estimators in a two-layer hierarchic architecture yields computations scaling with $n^{4/3}$; extending this approach to multiple layers yields linear scaling. An alternative approach that maintains a local structure is to allow actuator commands to depend on both local sensors and prior local estimates. This iterative approach is equivalent to a temporal low-pass filter on global information and gives a scaling of $n^{3/2}$. The algorithms are simulated by using data from the Palomar Observatory adaptive optics system. The analysis is general enough to also be applicable to active optics or other systems with many sensors and actuators. © 2003 Optical Society of America

OCIS codes: 010.1080, 350.1260.

1. INTRODUCTION

Many large optical telescopes have adaptive optics (AO) systems that compensate for constantly changing atmospheric distortion by using wave-front information feedback to a deformable mirror (DM).[1] Current AO systems generally involve hundreds of actuators and sensors, while future telescopes may require many more actuators and sensors [e.g., California Extremely Large Telescope (CELT)].[2] The optimal control algorithm feeds back every sensor to every actuator, resulting in a potentially unacceptable computational burden for future systems. A similar issue also arises for active optics systems [active control system (ACS)] for future segmented extremely large telescopes.[3]

The AO and ACS problems motivate development of computationally efficient control algorithms for systems with many sensors and actuators. Both problems require global feedback because *relative* sensor measurements are used to reconstruct (estimate) *absolute* displacements. This paper analyzes the optimal local (hence sparse) reconstructor and develops hierarchic and iterative approaches for recovering the resulting loss of performance on "global" modes, focusing on AO to motivate and validate the design approaches.

In AO, the sensors typically do not directly measure the atmospheric phase error but instead measure the slope of the wave-front error, e.g., with a Shack–Hartmann sensor. The wave-front phase at the actuator location, and therefore the desired actuator command, are estimated from the sensor measurements through a weighted pseudoinverse of the sparse influence matrix from actuator displacements to sensor measurements (see, e.g., Hardy).[1] This reconstruction matrix is in general fully populated; hence the entire sensor vector is required to give a good estimate of global wave-front error at any

point. Control of segmented mirror telescopes similarly uses relative measurements to estimate segment displacement, resulting in a fully populated reconstruction matrix.

One approach to reducing computation is based on using spatial Fourier transforms[4–6]; with n actuators, the resulting computations are of order $n \log n$ rather than n^2 for the full-matrix computation. Similar approaches can be developed for any circulant system.

Several approaches exist to reduce computation by relying only on sparse-matrix operations. The exact solution to the least-squares estimate can be obtained more efficiently through Cholesky factorization and backsubstitution using the sparse Cholesky factors; this approach avoids ever directly computing the reconstructor matrix. This approach has been extended to the multiconjugate adaptive optics case and to include atmospheric statistics.[7] A conjugate gradient approach can also solve the problem by using only sparse-matrix operations. Convergence is obtained through iteration and can be improved by using preconditioning.[8] The multigrid preconditioning method used in Gilles *et al.*[8] is related to the hierarchic estimation approach developed herein. More general iterative approaches are briefly discussed in Hardy[1] and form the basis of the sparse-matrix approach of Wild *et al.*[9] Other sparse-matrix approaches have also been suggested[10]; however, the reconstructors are not in general optimal. Decentralized controllers have been studied for telescope active control systems,[11] and for general applications.[12] The emphasis in these papers is typically on designing decentralized controllers in the presence of dynamic coupling between subsystems, not on coupling through the measurement.

The sparse reconstructors developed herein are optimal given the restricted available information. Performance

on low-spatial-wave-number "global" modes is degraded, and two approaches are considered to recover global performance.

A global estimator can be combined with the local estimate in a hierarchic architecture; this approach is developed herein for the AO problem. The extent of the local control can be optimized to minimize the total computational burden of the combined hierarchic system. By maintaining geometric similarity in developing the global layer, the approach is readily extended to multiple layers, analogous to multigrid approaches, yielding computations linear in the number of actuators n.

An alternative, iterative approach to recover global performance with small computational burden is to allow each actuator command to depend not only on local sensor information but also on prior local displacement estimates. This approach maintains a purely local information flow and thus is amenable to a decentralized hardware implementation. With actuator commands based only on local sensor information, global displacement information propagates across the array at a rate dependent on the control bandwidth; including local displacement estimates allows this propagation speed to be dependent on the sample rate or the communication rate instead. For quasi-steady disturbances, this algorithm is equivalent to the iterative approaches in Young.[13] The dynamic performance is equivalent to that of a low-pass filter on global information whose corner frequency depends on the spatial extent of the local control.

This paper develops local, hierarchic, and iterative reconstruction approaches and compares them for AO. The AO problem assumes a Shack–Hartmann-type wave-front sensor with a Fried geometry. The analysis assumes a frozen turbulence model, resulting in global disturbances changing more slowly than high-wave-number disturbances. Simulations use the actuator/sensor configuration of the Palomar Observatory AO system; the local and hierarchic algorithms have also been experimentally validated at Palomar.[14] The approach is also applicable to a broader class of problems involving many sensors and actuators, including control of a segmented primary mirror.

2. CONTROL PROBLEM

A. Actuators and Sensors

For analysis, consider a single DM with n actuators in a grid of size $(D + 1) \times (D + 1)$ and a Shack–Hartmann wave-front sensor with $D \times D$ subapertures aligned with the DM actuators in a Fried geometry. Each element of these sensors provides a measure of the average atmospheric tip and tilt over the region bounded by the neighboring actuators. Only those subapertures within the boundaries defined by the included actuators are used; otherwise, the estimation process described in Subsection 2.B becomes ill conditioned. In the limit of many actuators, there are twice as many sensors as actuators; the approximation $n_s \simeq 2n$ will be used in comparing computations of different algorithms. The number of actuators is approximately $n \simeq (\pi/4)D^2$; however, the factor of $\pi/4$ will not always be included in approximating computations. Flexible modes are typically at much higher fre-

quencies than the control bandwidth and can therefore be ignored. Furthermore, assume that the motion at one actuator location does not depend on any other actuator command and that the actuator displacement equals the command u.

Assuming frozen turbulence, disturbances of large spatial scale change more slowly than those of small scale. With an infinite aperture and a single turbulent layer of velocity v, disturbances of wave number k then have temporal frequency $\omega = kv$. With finite aperture, a given mode includes all wave numbers and thus has energy at all frequencies, but the bandwidth required for a given mode is still proportional to the wave number.

Performance is simulated by using telemetry data obtained at Palomar.[15] This system has 241 actuators and a 256-element Shack–Hartmann wave-front sensor ($D = 16$); the actuator and sensor geometry is shown in Fig. 1. There are 192 subapertures (giving 384 sensor measurements) that are within the diameter of the mirror and not obscured by the telescope secondary mirror.

The system state is described by the displacements $x \in \Re^n$ at the actuator locations:

$$x = u + w \tag{1}$$

for some disturbance vector w and control u. The sensor measurements y are related to the displacements x through

$$y = Ax + \eta, \tag{2}$$

where η is sensor noise. The $n_s \times n$ influence matrix A is sparse, since it is assumed that each sensor is influenced only by neighboring actuators.

Denote M as the dimension of the null space of A. For any estimation problem using relative measurements, piston motion (same displacement at every location) is unobservable. For the Fried AO wave-front sensor geom-
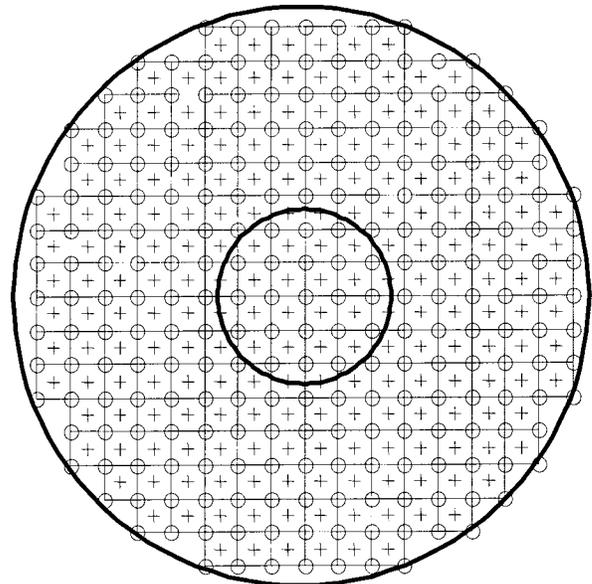

Fig. 1. Actuator ($\bigcirc$) and sensor ($+$) layout for Palomar adaptive optics (AO). The circles indicate the outer edge of the mirror and the region obscured by the telescope secondary mirror. Only those subapertures used in reconstruction are shown.
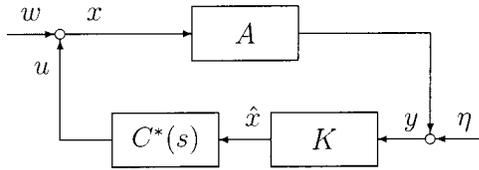
Fig. 2. Control block diagram. Displacement x is influenced by control u and disturbance w. Control consists of estimation $\hat{x} = Ky$ and control $u = C(s)\hat{x}$.

etry in Fig. 1, a "waffle" mode is also unobservable, wherein neighboring displacements have equal magnitude but alternating sign.

B. Baseline Control Algorithm

The baseline disturbance rejection control algorithm[1] can be divided into two steps, shown in Fig. 2. The first step is to estimate the displacements at the actuator locations given the sensor information, and the second step is to minimize this error by using a controller $C(s)$ (typically proportional-integral), assumed to be the same for each actuator. Actuator dynamics can be included in $C(s)$ for notational simplicity. The computations are dominated by the estimation, which is of the form $\hat{x} = Ky$ for a reconstructor matrix K.

In the absence of sensor noise, the optimal estimator is $\hat{x} = Ky$, $K = A^{\#}$, where the pseudoinverse is given by

$$A^{\#} = \lim_{\rho \to 0} (A^{T}A + \rho I)^{-1} A^{T}. \tag{3}$$

Despite A being sparse, $A^{\#}$ is in general fully populated. The physical interpretation of this result is that the absolute displacement can be estimated from relative measurements only by integrating those measurements to the boundary of the domain.

The error is given by

$$\hat{x} - x = (KA - I)x + K\eta \tag{4}$$

$$= -\sum_{m=1}^{M} \mu_m(\mu_m^{T}x) + K\eta, \tag{5}$$

where $\mu_m \in \Re^n$, $m = 1,...,M$, are an orthonormal basis for the null space of A. The error is the sum of all unobservable modes of x plus sensor noise; this result will be used in the subsequent sections.

With sensor noise covariance $\langle \eta\eta^{T}\rangle = R$ and disturbance covariance $\langle ww^{T}\rangle = Q$, the optimal reconstructor is given by the weighted pseudoinverse[4]

$$K = QA^{T}(AQA^{T} + R)^{-1} \tag{6}$$

$$= (A^{T}R^{-1}A + Q^{-1})^{-1}A^{T}R^{-1}. \tag{7}$$

A higher ratio of disturbance to sensor noise results in higher singular values of K. For simplicity, the remainder of this paper will focus on the form of the reconstructor in Eq. (3).

Tip and tilt are typically controlled separately by a fast-steering mirror and can be projected out of either the sensor measurement y or the estimate $\hat{x}$. Define ϕ_t as the tip or tilt mode shape; then the remaining estimate after projection is $\hat{x} = \hat{x} - \phi_t(\phi_t^{T}\hat{x})$. This requires $4n$ multiplies and thus can be ignored in the operation count for

large systems. The computations of this baseline algorithm are $\mathbf{C}(n) = nn_s \simeq 2n^2$.

Transforming into and out of a modal basis allows the loop gain to be tuned independently for different modes[16]; this requires two matrix multiplies with computations $n_z \times n_s$ to estimate n_z modes and $n \times n_z$ to compute control commands for total computations that scale with $3nn_z$. Thus there is also a small computational savings if the number of modes kept is less than 2/3 the number of actuators.

A significant reduction in the total computations can be obtained by using approaches based on the fast Fourier transform (FFT). Sensor measurements need to be extended from the circular to a square geometry. The estimate is then obtained through a two-dimensional spatial FFT of the sensor information, a spatial filter, and an inverse FFT. The operation count is roughly $5\bar{n}\log_2\bar{n}$, where $\bar{n} \simeq (4/\pi)n$. The multigrid preconditioned conjugate gradient approach[8] also has $n\log_2 n$ computational scaling.

C. Performance Assessment

The loop transfer function is $C(s)(KA)$. The matrix $L = KA$ determines the closed-loop behavior; with the optimal least-squares pseudoinverse, then, $L = I$. The loop gain along directions of interest gives an assessment of the relative performance of any reconstructor K. For an orthonormal basis set given by $\xi = \phi^{T}x$, compute the modal loop gain $L_\phi = \phi^{T}L\phi$; herein Zernike polynomials[17] are used. With the optimal reconstructor in Eq. (3), every mode is independent. With a general reconstructor K, there is coupling between modes, so noise in mode ξ_j can affect the residual energy in ξ_i. Thus an accurate assessment of performance requires a model of the disturbance energy in every mode and computation of the resulting closed-loop performance. For all of the reconstructors developed herein, the coupling between modes remains small, and the modal loop gains $(L_\phi)_{ii}$ can be used as a rough guide to the expected behavior. With an integral controller $C(s)$, a gain reduction for a given spatial mode corresponds directly to a bandwidth reduction.

The reconstructor also affects sensor noise propagation; for a reconstructor K, unit-covariance uncorrelated noise on each sensor yields rms noise on the estimate given by the noise multiplier[6] $[\text{Tr}(KK^{T})/n]^{1/2}$.

3. LOCAL CONTROL

We seek a local controller structure wherein each state estimate and hence control output depend only on information within a given region around the actuator location. Initially, consider estimators dependent only on local sensor measurements; this will be generalized in later sections. The best possible estimate of the displacement at each location given the sensor information available (or conditioned optimal estimate) results in a local least-squares problem. Note that truncating the optimal controller by zeroing elements of $K = A^{\#}$ corresponding either to distant sensors or to those with small gain is suboptimal and in general may give arbitrarily poor performance.

At each actuator location i, define Ω_i^s as the subset of sensors to be used in computing the estimate $\hat{x}_i$ and the corresponding actuator command u_i. Choose regions that include all sensors within distance $d/2$, i.e.,

$$y_j \in \Omega_i^s \;\Leftrightarrow\; \|\mathbf{x}(y_j) - \mathbf{x}(x_i)\| < d/2,$$

where $\mathbf{x}(\,\cdot\,)$ denotes spatial location (representing the two axes by a complex number) and $\|\cdot\|$ denotes distance. Herein $\|\mathbf{x}\| = \max(|\Re(\mathbf{x})|, |\Im(\mathbf{x})|)$, so the regions are square of size $d \times d$. Define Ω_i^a by $x_j \in \Omega_i^a$ if $\exists k$ such that $y_k \in \Omega_i^s$ and $A_{kj} \neq 0$; this is the set of states that influence any sensor in Ω_i^s. For actuators with a local influence, then, Ω^a has roughly the same spatial extent as that of Ω^s. Define $\tilde{A}_i = A_{\Omega_i^s, \Omega_i^a}$, where the notation indicates that rows in Ω_i^s and columns in Ω_i^a are kept.

Define $\delta = d/D$; the computations will scale as $\mathbf{C}(n) = 2\delta^2 n^2$. The computation time will not scale as efficiently as the number of multiplies because of the additional overhead required to implement the sparse-matrix–vector multiply. However, since the sparse matrices maintain some structure, the additional overhead is small.[18]

Given the subset Ω_i^s of available information, the optimal estimate of the overall state $\hat{x}|_{y \in \Omega_i^s}$ is obtained from the pseudoinverse as in Eq. (3), where only the rows of A corresponding to Ω_i^s are kept. State estimates not in Ω_i^a will be zero; hence the nonzero elements of $\hat{x}|_{y \in \Omega_i^s}$ can be computed from the pseudoinverse of the truncated matrix $\tilde{A}_i$. The ith row of the resulting sparse reconstructor has nonzero elements only for sensors in Ω_i^s; K_{i,Ω_i^s} is the row of $(\tilde{A}_i)^{\#}$ corresponding to the ith column of A. The optimal estimate given Ω_i^s, $i = 1,...,n$, is denoted $\hat{x}|_\Omega = Ky$. If A and Ω are spatially invariant away from the boundary, then K will also be spatially invariant.

This local, sparse estimator can be interpreted as a spatial high-pass filter. For each actuator location i, define a set of M_i vectors $\tilde{\mu}_m^i \in \Re^n$, $m = 1,...,M_i$, such that the elements corresponding to the actuators in Ω_i^a form an orthonormal basis for the null space of $\tilde{A}_i$. The estimation error for the ith element of $\hat{x}$ that results from using only the information in Ω_i^s can then be written, as in Eq. (5), as

$$\hat{x}_i|_{y \in \Omega_i} = \hat{x}_i|_y - \sum_{m=1}^{M_i} (\tilde{\mu}_m^i)_i [(\tilde{\mu}_m^i)^{\mathrm{T}} x] + K\eta. \qquad (8)$$

The local estimate is the global estimate minus the locally unobservable modes. Because of the geometric similarity, $M_i = M$ and the vectors $\tilde{\mu}_m^i$ correspond to μ_m of the full A matrix (piston and waffle). For $d = 2$, the dimension of Ω_i^a exceeds that of Ω_i^s, but by less than $M_i = M$, and thus the system is still effectively overdetermined (the number of degrees of freedom that are possible to determine is fewer than the number of sensor measurements).

Using relative measurements such as slope within a bounded domain Ω results in the average displacement over Ω being unobservable. Local waffle motion is also unobservable; this contribution to the error is not significant for low spatial frequencies but does affect the performance at high spatial frequencies. Thus the local estimate at low spatial frequencies is approximately equal to the global estimate minus the average displacement over Ω_i^a. For Ω square of size $d \times d$, the high-pass spatial-frequency response is given by

$$g(k_x, k_y) = 1 - \mathrm{sinc}(k_x d)\mathrm{sinc}(k_y d). \qquad (9)$$

With the use of a Fourier basis, the displacements can be represented by $k_x = m_x \pi/D$, $k_y = m_y \pi/D$, and thus the gain on the lowest-spatial-frequency mode is

$$g_{11} = 1 - \mathrm{sinc}^2(\pi\delta)$$
$$\simeq \pi^2 \delta^2/3. \qquad (10)$$

Even for modest δ, there is a significant reduction in low-frequency gain. For general problems, the spatial-frequency response can be derived from Eq. (8).

While the sparse reconstructor maintains good performance on high-spatial-wave-number displacements, the gain is reduced on the lowest $n_\xi = (D/d)^2$ wave-number global modes with half-wavelength between d and D. The following sections explore alternative approaches for recovering global performance without losing the computational savings of the local controllers. Both of the approaches considered are based on estimating the coefficients corresponding to $\tilde{\mu}_m^i$ in Eq. (8).

4. HIERARCHIC ESTIMATION

A. Two-Layer Hierarchy
Performance at low spatial frequencies can be improved by estimating the information $\tilde{\mu}_m^i$ that the local estimator does not, using global measurements. This consists of three steps: spatial filtering the sensor information and condensing it into a reduced set of data, estimating global parameters from these condensed data, and expanding these global parameters over the domain. Although it is not the only option, the global parameterization taken herein is chosen to be geometrically similar to that of the full problem, making the addition of further layers of hierarchy a simple extension. The "superelements" of the global layer have size equal to that of the local regions Ω.

With geometric similarity, the three steps are as follows:

1. Define $\tilde{y} = \Psi y$ as the average tip and tilt over superelements of size Ω.
2. Define a set of global variables ξ as the displacements at the corners of the super elements; so ξ and $\tilde{y} = A_\xi \xi$ are the displacements and the sensor measurements of a similar AO geometry with resolution larger by a factor of d, and $\hat{\xi} = A_\xi \tilde{y}$.
3. Define the global component of the state x as $x_g = \Phi\xi$ to estimate the piston mode $(\tilde{\mu}_1^i)_i [(\tilde{\mu}_1^i)^{\mathrm{T}} x]$ through interpolation and spatial averaging over Ω_i^a, as described below.

The superelements are illustrated schematically in Fig. 3. Note that with these superelements, variables of the global layer ξ are sampled values of the state x without the spatial filtering necessary to avoid (spatial) aliasing. Some filtering will be obtained through Φ; each component of x_g is obtained by first interpolating ξ to estimate the displacement at each location and then averaging the interpolated state over the region Ω_i^a. This region in general overlaps four superelements, and thus the nine
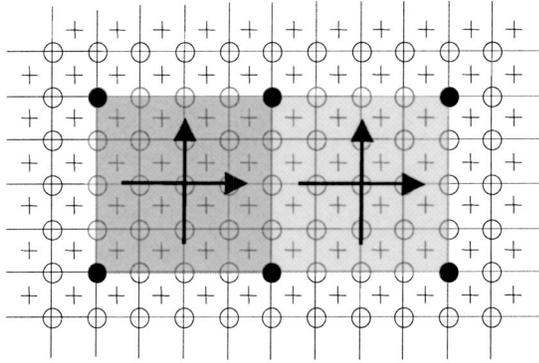
Fig. 3.   Hierarchy schematic for AO geometry with $d = 4$. Global superelements are shaded, and global displacements are indicated with solid circles.

global variables ξ at the corners of these superelements are used to construct each component of x_g. The spatial aliasing yields an estimation error that leads to some performance degradation; quantifying the magnitude of this error will be discussed in the context of the simulation results in Section 6.

The combined local and global estimator is therefore

$$\hat{x} = \hat{x}_\ell + \hat{x}_g = Ky + \Phi[A_\xi^{\#}(\Psi y)]. \tag{11}$$

A different control bandwidth can be obtained for different spatial scales (e.g., to compensate for different disturbance covariances) by computing the control as the sum of contributions from the different hierarchic layers. Thus, rather than $u = C(s)\hat{x}$, choose $u = C_\ell(s)\hat{x}_\ell + C_g(s)\hat{x}_g$.

With this global parameterization, both Φ and Ψ are sparse; each sensor is in a single superelement, and each global displacement estimate is interpolated from at most nine global variables $\hat{\xi}$; hence these calculations scale linearly with n. The hierarchic controller computations are the sum of contributions from the global and local layers; with the use of $n_\xi \simeq \delta^{-2}$, then, $\mathbf{C}(n) = 2\delta^{-4} + 2\delta^2 n^2 + O(n)$. Optimizing with respect to δ (or d) gives $d^2 = (2n)^{1/3}$ and computations $\mathbf{C}(n) = 6(n/\sqrt{2})^{4/3}$. The dependence on n is slow; requiring d even (Ω symmetric) gives $d = 4$ for roughly $200 < n < 6500$. If one also takes advantage of the lower bandwidth required for larger spatial scales, the global computations can be done at $1/d$ times the rate of the local computations. Including this factor in the optimization yields $d^2 = (5n/2)^{2/7}$ and computational scaling of $n^{9/7}$.

B.   Multiple Hierarchic Layers

The above approach for two hierarchic layers is readily extended to multiple layers. With $D \times D$ Shack–Hartmann subapertures, up to $\log_2 D$ layers of hierarchy can be included. Each layer, denoted $(\,\cdot\,)^{(\ell)}$, involves three steps similar to those described for the two-layer approach:

1.   Obtain the sparse local least-squares estimator $K^{(\ell)}$ to efficiently estimate the behavior appropriate to the ℓth scale by using aggregate information from the next-finer scale.

2.   Average the information $y^{(\ell)}$ available to the ℓth scale through an aggregation matrix $\Psi^{(\ell)}$.

3.   Interpolate and average the estimate from the next-coarser scale through $\Phi^{(\ell)}$ to correct the global behavior of the local estimate at the current scale.

Each of the matrices $K^{(\ell)}$, $\Psi^{(\ell)}$, and $\Phi^{(\ell)}$ is obtained as described in Subsection 4.A. This multiple-layer hierarchic approach is illustrated for a one-dimensional geometry in Fig. 4.

For the coarsest layer, then, the optimal estimate is the full least-squares solution:

$$\hat{x}^{(1)} = K^{(1)}y^{(1)}, \qquad K^{(1)} = (A^{(1)})^{\#}. \tag{12}$$

For each finer layer, the estimates are

$$\hat{x}^{(\ell)} = K^{(\ell)}y^{(\ell)} + \Phi^{(\ell)}\hat{x}^{(\ell-1)}, \tag{13}$$

$$y^{(\ell)} = \Psi^{(\ell+1)}y^{(\ell+1)}, \tag{14}$$

where, for the finest layer, $y^L = y$.

With the maximum layers of $L = \log_2 D$, this algorithm has *linear* computational scaling with n. The finest layer, with $n \sim (\pi/4)D^2$ subapertures, requires four neighboring subapertures (eight sensors) for the sparse calculation, the computations involving Ψ use each sensor only once, and the computations involving Φ use at most nine estimates from the next-coarser layer. Thus the total computations for this layer are $(8 + 2 + 9)n = 19n$. For each finer layer, the number of subapertures, the elements in $x^{(\ell)}$, and the number of computations all decrease by a factor of 4. If we note that

$$\sum_{k=0}^{\infty} \frac{1}{4^k} = \frac{4}{3}$$

then the total computations for this algorithm are $\mathbf{C}(n) < (4/3)19n \simeq 25n$. With multiple layers of $d = 4$, the local least-squares solution requires 16 subapertures (32 sensors). The computations now decrease by a factor of 16 for each coarser layer. The computational scaling increases to $\mathbf{C}(n) < (16/15)(32 + 2 + 9)n \simeq 46n$, but better performance should be obtained.
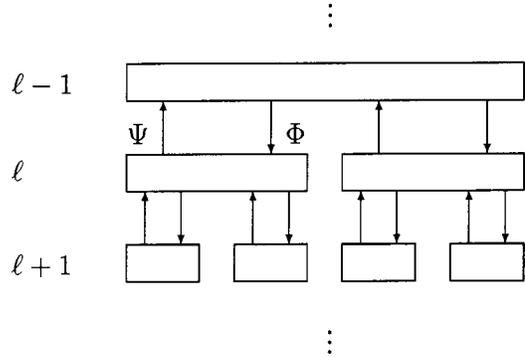
$$\vdots$$

Fig. 4.   Schematic of multiple-layer hierarchy in one spatial dimension. Each layer ℓ aggregates information $y^{(\ell-1)} = \Psi^{(\ell)}y^{(\ell)}$ to pass up to the next-coarser layer and uses interpolation and averaging of the estimate from the next-coarser layer, $\hat{x}_g^{(\ell)} = \Phi^{(\ell)}\hat{x}^{(\ell-1)}$, to correct the information missing from the local estimate $\hat{x}_l^{(\ell)} = K^{(\ell)}y^{(\ell)}$ at layer ℓ.

5. LOCAL CONTROL REVISITED

An alternative approach toward recovering the global performance of the optimal controller is to use prior state estimates to estimate the information $\tilde{\mu}_m^i$ that is missing from the local estimate. This results in an iterative approach and maintains the decentralized structure of the local controller. The estimate at a given location will depend not only on the sensor measurements Ω^s but also on the displacement estimates within Ω^a. If the hardware implementation were decentralized, then each local controller would communicate with its neighbors; similar architectures have appeared for distributed control systems.[12]

The dependence at time step k is of the form

$$\hat{x}_k = Ky_k + G\hat{x}_{k-1}, \tag{15}$$

where both K and G are sparse, containing nonzero entries only for sensor or actuator locations, respectively, within Ω^s or Ω^a. Multiple subiterations can be included between time steps k:

$$\zeta_1 = Ky_k + G\hat{x}_{k-1},$$

$$\xi_{m+1} = Ky_k + G\zeta_m, \qquad m = 1,...,N-2,$$

$$\hat{x}_k = Ky_k + G\zeta_{N-1}. \tag{16}$$

This dependence is the same as that of iterative algorithms for solving the matrix pseudoinverse problem[13]; these algorithms are discussed briefly for AO applications in Hardy.[1] Young[13] gives the conditions on the matrix G for $\hat{x}$ in the above equation to converge to $A^{\#}y$:

$$G = I - KA. \tag{17}$$

Note that $G_{ij} \neq 0 \Leftrightarrow \exists k$ such that $K_{ik} \neq 0$ and $A_{kj} \neq 0$; that is, G contains nonzero elements only for actuators in Ω^a defined earlier. With roughly twice as many sensors in Ω^s as actuators in Ω^a, the total computations per time step for N subiterations are $\mathbf{C}(n) = (2+N)\delta^2 n^2$.

For the local least-squares solutions considered in Section 3, the structure of G follows from Eq. (8); since KA has unity gain in directions where $\tilde{A}$ is nonsingular, G contains information only in directions where $\tilde{A}$ is singular:

$$(G\hat{x})_i = \sum_{m=1}^{M} (\tilde{\mu}_m^i)_i [(\tilde{\mu}_m^i)^{\mathrm{T}}\hat{x}]. \tag{18}$$

Thus, for AO, $G\hat{x}$ corresponds to the average of the estimates $\hat{x}$ within Ω^a plus an additional contribution due to local waffle.

The condition in Eq. (17) ensures that the estimate $\hat{x}$ in Eq. (15) converges to the correct value in steady state. Taking the z transform of Eq. (15) yields a temporal filter interpretation:

$$\hat{x}(z) = (I - Gz^{-1})^{-1}KAx(z) = H(z)x(z). \tag{19}$$

Substituting a singular-value decomposition for $KA = U\Sigma V^{\mathrm{T}}$ yields properties of $H(z)$ as a function of the modal loop gain σ_i along these singular directions:

$$H(z) = V \begin{bmatrix} \ddots & & \\ & \dfrac{\sigma_i z}{z - (1 - \sigma_i)} & \\ & & \ddots \end{bmatrix} V^{\mathrm{T}}.$$

The dc gain $\lim_{z \to 1} v_i^{\mathrm{T}} H(z)v_i = 1$ if $\sigma_i \neq 0$ and is zero if $\sigma_i = 0$. $H(z)$ is a low-pass filter with modally dependent corner frequency $z = 1 - \sigma_i$. For small σ_i, this gives a corner frequency $f_c/f_s = \sigma_i N/(2\pi)$ for sampling rate f_s. Substituting the gain in Eq. (10) for the sparse-matrix reconstructor yields a rough estimate for the corner frequency on the lowest-order mode:

$$\frac{f_c}{f_s} = \frac{\pi}{6}\delta^2 N. \tag{20}$$

The phase lag introduced by this filter limits the bandwidth that can be achieved in controlling global modes. With $N = 1, f_c/f_s \sim 0.1$ requires that $\delta \sim 0.44$. Increasing the number of subiterations improves the filter corner frequency linearly, and thus the number of sensor subapertures in the local control can decrease linearly for the same performance. Since there are roughly twice as many sensors as actuators in a given local region, it is always more efficient to increase the number of subiterations than the size of the local domain. Thus to achieve a specified ratio f_c/f_s, the optimum computations $(N \gg 2)$ give $\mathbf{C}(n) = (f_c/f_s)(6/\pi)n^2$. The iterative approach is useful if lower bandwidth is acceptable on the global modes relative to the local modes.

Assuming frozen turbulence, the bandwidth required drops linearly with wavelength, and thus the ratio of the bandwidth required between the largest and the smallest spatial scale is given by D. From Eq. (9), the sparse reconstructor gain, and therefore the bandwidth at which information is available by using this iterative approach, drop quadratically with wavelength. To ensure sufficient bandwidth on all modes, choose the minimum corner frequency to be a factor of D less than the Nyquist frequency, or $f_c/f_s > 1/(2D) \sim 1/(2n^{1/2})$. The optimum computations per iteration scale as $\mathbf{C}(n) = (3/\pi)n^{3/2} \simeq n^{3/2}$. This assumes that the sample rate f_s is chosen appropriately to control the smallest spatial scales. A different control gain can be used for different spatial scales by computing the control as a sum of contributions based on each iteration ζ_i; this may be necessary to optimize the gain at each scale according to the bandwidth of available information.

With $N = 1$, the noise multiplier for uncorrelated sensor noise is $[\mathrm{Tr}(X)/n]^{1/2}$, where X satisfies the Lyapunov equation $X = GXG^{\mathrm{T}} + KK^{\mathrm{T}}$. Equations (16) can be rewritten as

$$\hat{x}_k = \left(\sum_{i=0}^{N-1} G^i\right) Ky + G^N \hat{x}_{K-1}, \tag{21}$$

from which the noise multiplier can be similarly obtained for $N > 1$.

6. SIMULATION

The purpose of developing the algorithms in the previous sections is to yield a tractable computational burden for future large AO systems. However, simulating using real data from a current AO system both validates the algorithms and provides an understanding of the sources of performance loss that can be scaled to large systems. Simulation results can also be compared with experimental results to further validate the algorithms.[14]

Reconstructors were designed for the Palomar AO geometry described in Section 2, and simulated by using data collected at Palomar. Modal loop gain, computations, noise multipliers, and simulated performance are compared. Using these simulation results to give a completely general assessment of the expected performance on any AO system is difficult because algorithm performance depends on the relative amplitude of low- and high-wave-number residual closed-loop energy, which in turn depends on the nominal control bandwidth relative to the turbulence wind-crossing time. Assume a Kolmogorov spectrum that decays with wave number at $k^{-11/3}$, as well as frozen turbulence, implying higher temporal bandwidth for higher wave number. Then for most expected control bandwidths, the global, low-wave-number response still dominates the residual energy, while for sufficiently high bandwidth, there may be more energy in higher-wave-number modes; this behavior is relevant in estimating the expected performance degradation for large systems.

The least-squares reconstructors currently used in the Palomar AO system are computed without explicitly recognizing the central obscuration; the obscured sensor subapertures simply provide zero input to the reconstruction. Although this should be somewhat suboptimal, the same approach is taken herein for simplicity. Local least-squares controllers were derived for $d/2 = 2$ to 6, spanning $\delta = 0.25$ to 0.75. The modal gains are plotted in Fig. 5 with the use of Zernike basis functions. Global tip and tilt ($Z_{1,\pm 1}$) are projected out and are not shown. The loop gain on higher-order modes (small spatial wave-
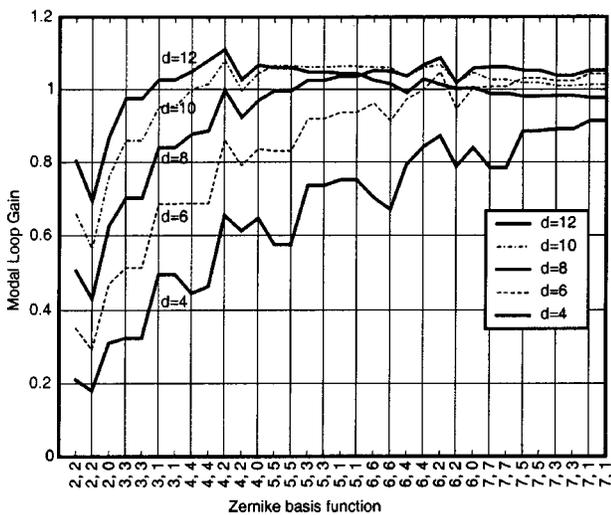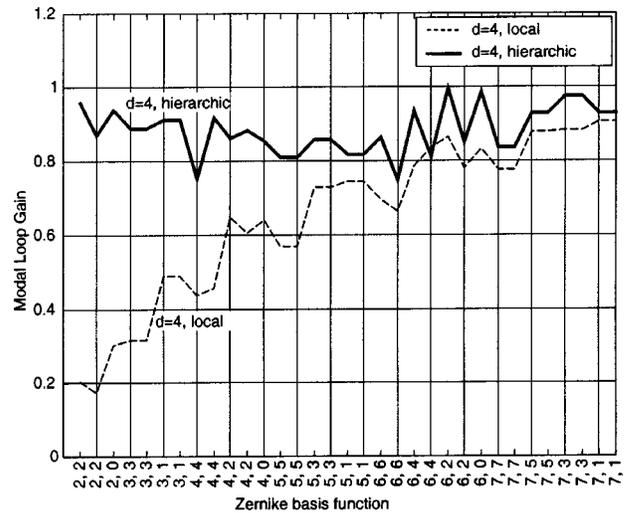

Fig. 6.   Modal loop gain comparing local control with $d = 4$ and two-layer hierarchic control with 21 global variables.

length deformations) is roughly constant for all of the local controllers. The gain on the lower-order, global modes (large spatial wavelength deformations) deteriorates as the local extent used in estimating the phase error is decreased. The lowest gain is consistent with the estimate given in Eq. (10).

For Ω^s of $d = 4$, $\delta = 0.25$, the hierarchic parameterization in Section 4 results in 21 global variables. The modal loop gains with and without hierarchic estimation are shown in Fig. 6; there is some aliasing from higher modes to lower that cannot be seen on this plot. A four-layer hierarchic reconstructor was also created with subaperture grids of 16×16, 8×8, 4×4, and 2×2 and $d = 2$ for each layer. The loop gains are similar to those of the two-layer hierarchy and are not plotted.

The iterative algorithm in Section 5 was evaluated for $d = 4$ and 6 and for $N = 1,...,4$. With this algorithm, the dc gain of every nonsingular mode is always unity, but the corner frequency at which the information is filtered decreases with decreasing δ and increases with N. The modal corner frequencies f_c/f_s can be estimated from Fig. 5 as the gain times $N/(2\pi)$.

A closed-loop simulation of the different algorithms was conducted by using telemetry data obtained at Palomar with the tip/tilt loop closed (to avoid saturating the Shack–Hartmann wave-front sensors). The performance metric from the simulation is the rms residual actuator position error and does not include errors from calibration, fitting, or the tip/tilt loop. The recorded wave-front sensor data include the response to both atmospheric wave-front disturbance information and sensor (photon and read) noise, while the simulation assumes that all of the data represent atmospheric turbulence. As a result, the simulation has more than double the disturbance energy in the highest-spatial-frequency modes than should be present. To avoid possibly misleading results, the performance is evaluated as the sum of contributions from the first 91 Zernike modes only (up to radial degree 12). Furthermore, while simulating without added sensor noise simplifies the identification of different error sources, it results in greater-than-realistic performance

Fig. 5.   Modal loop gain of different spatial extent of local control. Basis functions are Zernike modes ordered with increasing spatial wave number.

reductions for the baseline full least-squares reconstructor and therefore larger implied errors for any of the alternative algorithms.

A 400-Hz sample rate is used with a two-cycle delay simulating the actual delay at Palomar, and the gains are those that have been optimized at Palomar, leading to a loop crossover frequency of roughly 20 Hz. The telemetry data were recorded at 100 Hz and resampled with a noisy interpolation to ensure some disturbance content above the original Nyquist frequency of 50 Hz. This allows a more accurate assessment of the local reconstructors, which have increased loop gain for some disturbance shapes, leading to an increase in the sensitivity function above 50 Hz.

The modal performance of several designs is shown in Fig. 7, illustrating reduced performance for lower-order Zernike modes that is consistent with the decrease in model gain.

The performance of different approaches relative to the optimal provides a basis for comparing the trade-offs among performance, computation, and propagation of sensor noise. The total number of multiplies required for these designs, normalized by the multiplies required for the optimal global controller, is shown in Fig. 8. The vertical axis is the simulated rms performance on the first 91 Zernike modes normalized by that of the least-squares controller.

The local reconstructors introduce error due to both of the unobservable modes, local piston and local waffle; the former influences the global behavior, while the latter influences the local behavior. This paper has focused on the former, as the global modes dominate the open-loop atmospheric disturbances and typically dominate the closed-loop residuals as well. The contribution to the performance metric in Fig. 8 due to local waffle errors is approximately 5% for $d = 4$. (The relative contributions of the two can be identified by creating a "perfect" hierarchic controller that uses the full least-squares estimate to correct either the piston or waffle errors of the local reconstructor.) For a larger system, a similar global performance penalty would be expected for equal δ (not d), while the error due to local waffle would in general be smaller.

Both the hierarchic and iterative approaches yield performance improvements relative to the sparse controller while retaining computational savings relative to the full-matrix reconstructor.

The hierarchic controller does not correct local waffle errors and therefore has the same performance degradation as that of the local controller for high-wave-number modes. The primary error source on the global modes is spatial aliasing, which degrades global performance by approximately 10% (the remaining errors in Fig. 8 are due to uncorrected local waffle). In general, the effect of aliasing depends on the closed-loop residual energy as a function of wave number. A worst-case estimate can be obtained for a high control bandwidth where the low- and high-frequency wave-number energies are comparable; in this case, the aliasing noise could increase residual phase errors by as much as $\sqrt{2}$. In practice, the errors are likely to be much smaller; because of the excess high-wave-number noise, the Palomar simulation should be worse than typical, and the performance in Fig. 8 should be an upper bound on what is likely to occur in other systems of any size. The multiple-layer hierarchic controller compounds the aliasing problem and suffers a higher performance penalty. For larger systems, as more layers are used, the estimation errors are likely to increase.

The iterative controller by its structure corrects both the local waffle and local piston errors. However, for this small system, there is only a small computational benefit relative to simply increasing the size of the local regions of the local approach.

There are two critical observations that must be considered relative to the results shown in Fig. 8 for Palomar. First is that the relative computational improvement for these algorithms would be much more significant for larger problem sizes. Second, the simulation is conducted with equal time delay for all algorithms. However, reducing the computational burden would afford a decreased time delay and a correspondingly increased bandwidth. This in turn would lead to improved perfor-
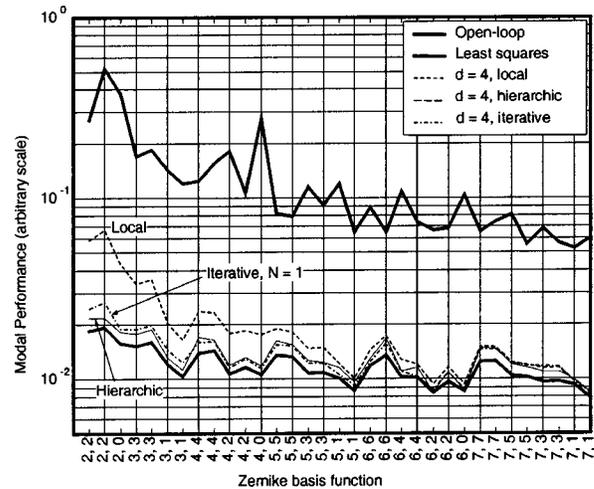

Fig. 7. Modal performance on Palomar AO simulation comparing the least-squares optimum with local control ($d = 4$), two-layer hierarchic control, and iterative control ($N = 1$).
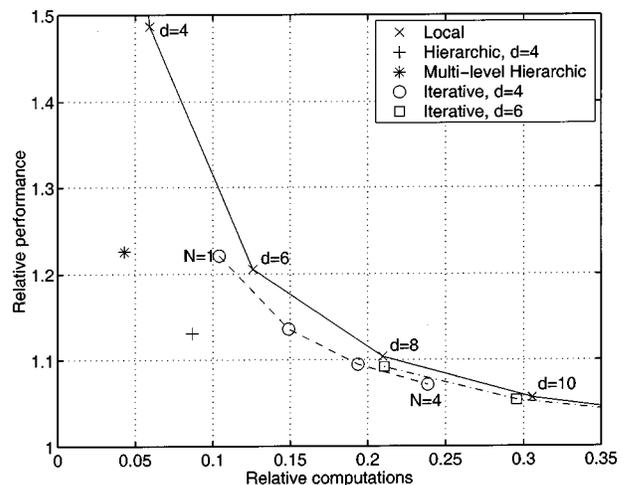

Fig. 8. Simulated performance versus computational improvement for Palomar AO with sparse, hierarchic, and iterative controllers. Performance is determined by the rms residual over the first 91 Zernike modes for a fixed sample rate.
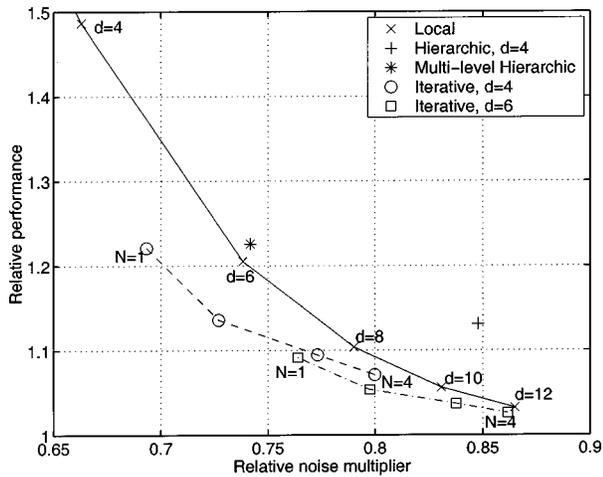
Fig. 9. Simulated performance versus noise multiplier for Palomar AO with sparse, hierarchic, and iterative controllers.

mance. In addition to the reconstruction calculations, the time delay includes detector integration, zero-order hold, and computation for centroiding calculations, and thus the total time delay reduction depends on the rest of the system. With the bandwidth improvement taken into account, the designs in Fig. 8 could result in a net improvement in performance, even for current-generation AO systems, and a considerable net improvement for future systems with many more actuators and sensors. The extent to which increased bandwidth offsets increased estimation errors associated with these algorithms clearly depends on the computational power available; however, for large AO systems with computers available today, the computations of the full least-squares algorithm are likely to severely limit achievable bandwidth.

The noise multiplier for the different designs is plotted in Fig. 9, scaled relative to the optimal controller and again plotted against the performance obtained in simulation. The noise multiplier for the sparse reconstructors decreases monotonically from that of the optimum least-squares solution as the number of sensors used for each actuator decreases. Both the hierarchic and iterative methods yield acceptable noise propagation.

7. CONCLUSIONS

Current active and adaptive optics systems for large optical telescopes typically employ an optimal control algorithm that uses all sensor information in computing each actuator command. For extremely large telescopes currently being designed, the associated computational burden of such algorithms can be significant. Local control algorithms are developed wherein each actuator command is computed based only on a subset of neighboring sensors. These decentralized algorithms perform close to optimally on localized disturbances, but the performance is degraded relative to that of the globally optimal controller on low-order, global disturbances. The degradation can be quantified in terms of a spatial high-pass filter.

Introducing an estimator of the global deformation and combining the local and global controllers with a hierachic control approach yields performance close to that of the globally optimal algorithm, with significantly lower computational burden that scales with $n^{4/3}$, while a multiple-layer approach yields linear scaling.

Based on iterative methods, an alternative approach to recovering global performance is also developed in which a decentralized structure is retained. The estimate at each location now depends on both the local sensor measurements and the prior local state estimates. This algorithm recovers global performance at dc but acts as a temporal low-pass filter on global information. With a constraint on filter bandwidth that is reasonable for a frozen turbulence disturbance model, the computational burden of the iterative algorithm scales with $n^{3/2}$.

Simulations have been performed by using the Palomar Observatory adaptive optics geometry. Both the hierarchic and iterative approaches recover most of the global performance while maintaining significant computational savings relative to the optimal least-squares solution. Improved sensor noise propagation is also obtained. Local and hierarchic reconstructors have been demonstrated at Palomar.[14]

Extensions of this work will include spatial coupling through the disturbance model to allow for knowledge of atmospheric turbulence statistics.

APPENDIX A: NOTATION

D	Size of domain ($D \times D$)
n (n_s)	Number of actuators (sensors)
$\mathbf{C}(n)$	Computational scaling with n
Ω^s (Ω^a)	Local set of sensors (actuators)
d	Size of Ω ($d \times d$)
δ	Relative size of Ω ($\delta = d/D$)
x	Vector of displacement at actuator locations
y	Measurement vector
A	Influence matrix ($y = Ax$)
$\{\mu_m\}_{m=1}^M$	Orthonormal basis for null space of A
$\hat{x}$	State estimate
$\hat{x}\vert_\Omega$	State estimate given $y \in \Omega$
K	Reconstructor matrix ($\hat{x} = Ky$)
G	Gain matrix on prior estimate
ξ	Vector of global states
Φ	Interpolation and averaging of global states ($x = \Phi\xi$)
$\tilde{y}$	Reduced set of measurements
Ψ	Measurement averaging ($\tilde{y} = \Psi y$)
$(\,\cdot\,)^{(\ell)}$	Variable ($\cdot$) at ℓth layer of hierarchy

Douglas G. MacMartin may be reached by e-mail at macmardg@cds.caltech.edu.

REFERENCES

1. J. W. Hardy, *Adaptive Optics for Astronomical Telescopes*, Oxford Series on Optical and Imaging Sciences 16 (Oxford U. Press, New York, 1998).
2. R. Dekany, J. E. Nelson, and B. Bauman, "Design considerations for CELT adaptive optics," in *Optical Design, Materials, Fabrication, and Maintenance*, P. Dierickx, ed., Proc. SPIE **4003**, 212–225 (2000).

3. J. Nelson and T. Mast, eds., "Conceptual design for a 30-meter telescope," CELT Rep. 34 (University of California and California Institute of Technology, Berkeley, Calif., 2002).

4. R. H. Hudgin, "Optimal wave-front estimation," J. Opt. Soc. Am. **67**, 378–382 (1977).

5. K. Freischlad and C. Zeiss, "Wavefront integration from difference data," in *Interferometry: Techniques and Analysis*, G. M. Brown, O. Y. Kwon, M. Kujawinska, and G. T. Reid, eds., Proc. SPIE **1755**, 212–218 (1992).

6. L. A. Poyneer, D. T. Gavel, and J. M. Brase, "Fast wavefront reconstruction in large adaptive optics systems with use of the Fourier transform," J. Opt. Soc. Am. A **19**, 2100–2111 (2002).

7. B. L. Ellerbroek, "Efficient computation of minimum-variance wave-front reconstructors with sparse matrix techniques," J. Opt. Soc. Am. A **19**, 1803–1816 (2002).

8. L. Gilles, C. R. Vogel, and B. L. Ellerbroek, "Multigrid preconditioned conjugate-gradient method for large-scale wave-front reconstruction," J. Opt. Soc. Am. A **19**, 1817–1822 (2002).

9. W. J. Wild, E. J. Kibblewhite, and R. Vuilleumier, "Sparse matrix wave-front estimators for adaptive-optics systems for large ground-based telescopes," Opt. Lett. **20**, 955–957 (1995).

10. T. P. Murphy, R. G. Lyon, J. E. Dorband, and J. M. Hollis, "Sparse matrix approximation method for an active optical control system," Appl. Opt. **40**, 6505–6514 (2001).

11. K. Li, E. B. Kosmatopoulos, P. A. Ioannou, and H. Ryaciotaki-Boussalis, "Large segmented telescopes: centralized, decentralized and overlapping control designs," IEEE Control Syst. Mag., October 2000, 59–72.

12. R. D'Andrea, C. Langbort, and R. Chandra, "A state space approach to control of interconnected systems," in *Mathematical Systems Theory in Biology, Communication, Computation and Finance*, IMA Vol. 134 in Mathematics and Its Application, J. Rosenthal and D. S. Gillian, eds. (Springer-Verlag, New York, 2003), pp. 157–182.

13. D. M. Young, *Iterative Solution of Large Linear Systems* (Academic, New York, 1971).

14. F. Shi, D. G. MacMartin, M. Troy, G. L. Brack, R. S. Burruss, and R. G. Dekany, "Sparse matrix wavefront reconstruction: simulations and experiments," in *Adaptive Optical System Technologies II*, P. L. Wizinowich and D. Bonaccini, eds., Proc. SPIE **4839**, 1035–1044 (2002).

15. M. Troy, R. Dekany, G. Brack, B. Oppenheimer, E. Bloemhof, T. Trinh, F. Dekens, F. Shi, T. Hayward, and B. Brandl, "Palomar adaptive optics project: status and performance," in *Adaptive Optical Systems Technology*, P. L. Wizinowich, ed., Proc. SPIE **4007**, 31–40 (2000).

16. J.-P. Gaffard and G. Ledanois, "Adaptive optics transfer function modeling," in *Active and Adaptive Optical Systems*, M. A. Ealey, ed., Proc. SPIE **1542**, 34–45 (1991).

17. R. J. Noll, "Zernike polynomials and atmospheric turbulence," J. Opt. Soc. Am. **66**, 207–211 (1976).

18. T. Truong, G. Brack, T. Trinh, M. Troy, F. Shi, and R. G. Dekany, "Real-time wavefront processors for the next generation of adaptive optics systems: a design and analysis," in *Adaptive Optical System Technologies II*, P. L. Wizinowich and D. Bonaccini, eds., Proc. SPIE **4839**, 911–922 (2002).