Fundamentals of Computational Linear Algebra for Inverse Problems

Curt Vogel

Department of Mathematical Sciences

Montana State University

http://www.math.montana/∼vogel

vogel@math.montana.edu

Acknowledgments

Collaborators

Brent Ellerbroek, NOAO, Tuscon AZ

Luc Gilles, Michigan Tech

Mike Flanagan, TREX, Maui HI

Funding

AFOSR-DEPSCoR Computational Mathematics Program

CfAO Modeling and Simulation Group

Outline

- Fundamentals of Inverse Problems
  - Ill-Posedness
  - Regularization
- Fundamentals of Computational Linear Algebra
  - Direct vs Iterative Methods
  - Computational Complexity
  - Special Structure $\longleftrightarrow$ Special Algorithms
    - Sparse, banded direct solvers
    - Iterative methods for structured systems
    - Preconditioning
- Computational Linear Algebra for AO Wavefront Reconstruction (open loop)
  - Conventional AO
  - MCAO

Forward Model, or "Direct Problem", in Adaptive Optics

Model measurements of wavefront slope $\nabla\phi = (\frac{\partial\phi}{\partial x}, \frac{\partial\phi}{\partial y})$, given

- Sensor subapertures Ω_i, for $i = 1$, ..., no. of subapertures.
- View angles $\boldsymbol{\theta}_j$, for $j = 1$, ..., no. of view angles.
- Atmospheric turbulence (refractive index) profile $\rho = \rho(x, y, z)$.
- Idealized "point source" guide star at infinity.

Mathematical model for Shack-Hartmann wavefront sensor:

$$s_{i,j} = (s_x, s_y)_{i,j} = \frac{1}{\text{Area}(\Omega_i)} \int_{\Omega_i} \nabla\phi_j \stackrel{\text{def}}{=} \mathcal{G}_i \phi_j$$

where the aperture-plane phase in direction $\boldsymbol{\theta}_j = (\theta_{x,j}, \theta_{y,j})$ is

$$\phi_j(x, y) = \int_0^H \rho(x + z\,\theta_{x,j}, y + z\,\theta_{y,j},\ z)\, dz \stackrel{\text{def}}{=} \mathcal{P}_j \rho$$

Relevant Inverse Problems

Conventional AO Wavefront Sensing.
Estimate aperture-plane phase $\phi(x, y)$, given

$$s_i = \mathcal{G}_i \phi + \text{noise}, \quad i = 1, ..., \text{ no. of subapertures.}$$

Assumes ϕ is independent of view angle θ.

Operator Notation: $\mathbf{s} = \mathcal{G}\phi$.

Turbulence Profile Estimation, or"Tomography", in MCAO:
Estimate turbulence profile $\rho(x, y, z)$, given

$$s_{i,j} = \mathcal{G}_i \mathcal{P}_j \rho + \text{noise}, \quad i = 1, ..., \text{ no. subaps}, \quad j = 1, ..., \text{ no. view angles},$$

Operator Notation: $\mathbf{s} = \mathcal{G}\mathcal{P}\rho$.

Ill-Posedness

Operator equation $\mathbf{s} = \mathcal{G}\phi$ is called ill-posed if any of the following conditions hold:

1. Nonexistence of a solution: There are measurement vectors $\mathbf{s}$ which do not correspond to any solution ϕ.
2. Nonuniqueness: There are measurement vectors $\mathbf{s}$ which correspond to several different solutions ϕ.
3. Instability: Certain small changes in the data $\mathbf{s}$ give rise to large changes in the solution ϕ.

Abstract Mathematical Definition Has Conceptual Difficulties ...

- Given finite data $\mathbf{s}$ and a "distributed parameter" ϕ which is a function of continuous variables $(x, y, z, \ldots)$, the solution is always nonunique, and hence, ill-posed. Mathematicians examine ill-posedness of idealized problems with infinite-dimensional data. In wavefront sensing, this corresponds to an arbitrarily large number of sensors with arbitrarily small subapertures.
- In practice, solution ϕ must be discretized. Need to solve matrix-vector equation, $\mathbf{s} = G\boldsymbol{\phi}$.
- Structure of discretized problem reflects structure of underlying distributed parameter problem.

Discretization

Distributed parameters can be well-represented by finite-dimensional approximations.

Example. Kolmogorov model for aperture-plane phase:

$$\phi(x,y) = \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \beta_{m,n} e^{2\pi i (x\kappa_{x,m} + y\kappa_{y,n})}$$

with independent, random coefficients $\beta_{m,n} \sim \mathrm{Normal}(0, |(\kappa_{x,m}, \kappa_{y,n})|^{-11/6})$. Infinite series is well-represented by truncated finite sum.

Example. Spline (i.e., smooth piecewise polynomial) tensor product representation used in finite elements,

$$\phi(x,y) \approx \sum_{m=1}^{M} \sum_{n=1}^{N} \gamma_{m,n} B_m(x) B_n(y)$$

If ϕ is smooth, approximation becomes more accurate as M, N increase.

Singular Value Decomposition (SVD)

Tool to analyze discrete linear equations $G\boldsymbol{\phi} = \mathbf{s}$. (Has continuous analogue.)

The SVD is a bi-orthonormal diagonalization of $m \times n$ matrix

$$G = \underbrace{[\mathbf{u}_1 \ldots \mathbf{u}_n]}_{m\times n\ U} \underbrace{\mathrm{diag}(\mu_1, \ldots, \mu_n)}_{n\times n\ D} \underbrace{[\mathbf{v}_1 \ldots \mathbf{v}_n]}_{n\times n\ V}$$

- Singular values μ_i are nonnegative.
- Right singular vectors $\mathbf{v}_i \in \mathbb{R}^n$ are orthonormal, $V^T V = I_{n\times n}$.
- Left singular vectors $\mathbf{u}_i \in \mathbb{R}^m$ are orthonormal, $U^T U = I_{n\times n}$.

Closely related to eigendecomposition of symmetric matrix $G^T G$:

$$G^T G = V\ D^2\ V^{-1}.$$

Eigenvalues of $G^T G$ are squared singular values of G; eigenvectors of $G^T G$ are right singular vectors.

SVD, Least Squares, and the Pseudo-Inverse

Finite dimensional linear systems $G\boldsymbol{\phi} = \mathbf{s}$ always have a least squares solution

$$\boldsymbol{\phi}_{\mathrm{LS}} = \arg \min_{\boldsymbol{\phi} \in \mathbb{R}^n} ||G\phi - \mathbf{s}||^2.$$

If singular values are all positive, have unique least squares solution

$$\begin{aligned} \boldsymbol{\phi}_{\mathrm{LS}} &= (G^T G)^{-1} G^T \ \mathbf{s} \\ &= \underbrace{V \ \mathrm{diag}(1/\mu_i) \ U^T}_{\mathrm{pseudo-inverse} \ G^\dagger} \ \mathbf{s} \end{aligned}$$

Otherwise, of all possible least squares solutions, the one of minimum Euclidean norm is

$$\boldsymbol{\phi}_{\mathrm{LSMN}} = \underbrace{V \ \mathrm{diag}(\mu_i^+) \ U^T}_{G^\dagger} \ \mathbf{s}$$

where

$$\mu_i^+ = \begin{cases} 1/\mu_i & \text{if } \mu_i > 0, \\ 0 & \text{if } \mu_i = 0. \end{cases}$$

Information from the SVD

Provided that β well-represents the distributed parameter ϕ,

- Zero singular values $\Longrightarrow$ nonuniqueness.
    - Corresponding singular vectors are unsensed modes.
- Relatively small singular values $\Longrightarrow$ instability.
    - Corresponding right singular vectors are unstable modes.

The condition number is a measure of instability.

$$\kappa(G) \stackrel{\text{def}}{=} \frac{\text{largest singular value of } G}{\text{smallest nonzero singular value of } G}$$

Matrix G is called ill-conditioned if $\kappa(G)$ is relatively large.

Regularization

- In mathematics, “regularity” means smoothness.
- Historically, accurate approximations to ill-posed inverse problems $\mathcal{G}\phi = s$ were obtained by imposing smoothness constraints on the solution ϕ.
- Regularization has evolved to mean any technique that yields accuarate approximate solutions to $\mathcal{G}\phi = s$.

Very brief sketch of Mathematical Theory of regularization.
Problem with noisy data:

$$s = \mathcal{G}\phi_{\mathrm{true}} + \eta.$$

Regularized solution $\phi_{\mathrm{reg}} = \phi_{\mathrm{reg}}(s, ...)$, depends on data s, prior information, regularization parameters, ..., in a manner for which

$$\phi_{\mathrm{reg}} \to \phi_{\mathrm{true}} \quad \text{as} \quad \eta \to 0.$$

Truncated Singular Value Decomposition (TSVD)

Gives approximation $\boldsymbol{\phi}_\alpha$ to least squares minimimum norm solution to $G\boldsymbol{\phi} = \mathbf{s}$.

$$
\begin{aligned}
\boldsymbol{\phi}_{LSMN} &= G^\dagger \mathbf{s} = V \,\mathrm{diag}(\mu_i^+)\, U^T \mathbf{s} \\
&= \sum_{\mu_i > 0} \frac{\mathbf{u}_i^T \mathbf{s}}{\mu_i}\, \mathbf{v}_i \\
&\approx \underbrace{\sum_{\mu_i^2 > \alpha} \frac{\mathbf{u}_i^T \mathbf{s}}{\mu_i}\, \mathbf{v}_i}_{\boldsymbol{\phi}_\alpha}
\end{aligned}
$$

Can be rewritten as

$$
\boldsymbol{\phi}_\alpha = \sum_i w_\alpha(\mu_i^2) \frac{\mathbf{u}_i^T \mathbf{s}}{\mu_i}\, \mathbf{v}_i
$$

where filter function zeros components corresponding to small singular values.

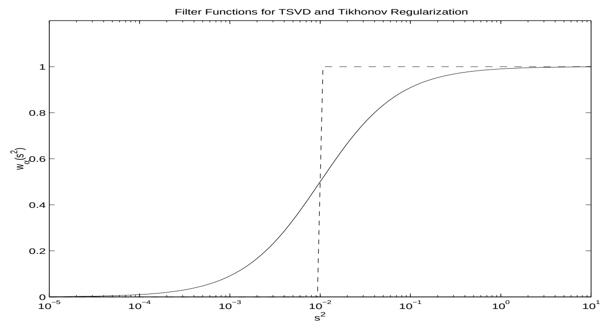
$$
w_\alpha(\lambda) = \begin{cases} 1, & \lambda > \alpha, \\ 0, & \lambda \le \alpha. \end{cases}
$$

Tikhonov Regularization

Yields penalized least squares approximation to $G\boldsymbol{\phi} = \mathbf{s}$.

$$\begin{aligned}
\boldsymbol{\phi}_\alpha &= \arg\min_{\boldsymbol{\phi}} ||G\boldsymbol{\phi} - \mathbf{s}||^2 + \alpha||\boldsymbol{\phi}||^2 \\
&= (G^TG + \alpha I)^{-1} G^T \mathbf{s} \\
&= \sum_i \frac{\mu_i \ \mathbf{u}_i^T \mathbf{s}}{\mu_i^2 + \alpha} \mathbf{v}_i \\
&= \sum_{\mu_i > 0} \underbrace{\frac{\mu_i^2}{\mu_i^2 + \alpha}}_{w_\alpha^{\mathrm{Tikh}}(\mu_i^2)} \ \frac{\mathbf{u}_i^T \mathbf{s}}{\mu_i} \ \mathbf{v}_i
\end{aligned}$$

Tikhonov filter function is smoothed version of TSVD filter.



Incorporating Prior Information

Illustrative Example. For simplicity suppose $\phi(x) \approx \sum_{i=1}^{N} \phi_i B_i(x), \; 0 \leq x \leq 1$. Measure roughness of ϕ by squared L^2 norm of derivative,

$$
\begin{aligned}
J(\phi) &= \int_0^1 \left(\frac{d\phi}{dx} \right)^2 \, dx \\
&\approx \int_0^1 \left(\sum_i \phi_i \frac{dB_i}{dx} \right) \left(\sum_j \phi_j \frac{dB_i}{dx} \right) \, dx \\
&= \sum_i \sum_j \phi_i \underbrace{\left(\int_0^1 \frac{dB_i}{dx} \frac{dB_j}{dx} dx \right)}_{L_{ij}} \phi_j \\
&= \boldsymbol{\phi}^T L \boldsymbol{\phi}.
\end{aligned}
$$

Tikhonov regularization with roughness penalty:

$$
\begin{aligned}
\boldsymbol{\phi}_\alpha &= \arg\min_{\boldsymbol{\phi}} ||G\boldsymbol{\phi} - \mathbf{s}||^2 + \alpha \boldsymbol{\phi}^T L \boldsymbol{\phi} \\
&= (G^T G + \alpha L)^{-1} G^T \mathbf{s}.
\end{aligned}
$$

Tikhonov Regularization–Minimum Variance Connection

Stochastic model for conventional AO sensor measurements:

$$\mathbf{s} = \mathcal{G}\phi + \boldsymbol{\eta}$$

with noise $\boldsymbol{\eta} \sim \mathrm{N}(\mathbf{0}, C_\eta)$, independent of phase $\phi \sim \mathrm{N}(0, C_\phi)$.
Assume deformable mirror figure ϕ_{DM} depends linearly on actuator vector $\mathbf{a}$:

$$\phi_{\mathrm{DM}}(x,y) = \sum_j a_j h_j(x,y) \stackrel{\mathrm{def}}{=} \mathcal{H}\mathbf{a}.$$

Assume (open loop) actuator vector depend linearly on sensor measurements:

$$\mathbf{a} = R\mathbf{s}.$$

Minimum variance reconstruct matrix is

$$\begin{aligned} R_{\mathrm{MV}} &= \arg\min_{\mathbf{a}=R\mathbf{s}} \langle ||\phi - \phi_{\mathrm{DM}}||^2 \rangle = \arg\min_R \langle ||\phi - \mathcal{H}R\mathbf{s}||^2 \rangle \\ &= \arg\min_R \langle ||(I - \mathcal{H}R\mathcal{G})\phi - \mathcal{H}R\boldsymbol{\eta}||^2 \rangle \end{aligned}$$

Tikhonov-MV Connection, Continued

$$
\begin{aligned}
R_{\mathrm{MV}} &= \arg\min_R \mathrm{Trace}\{[(I - \mathcal{H}R\mathcal{G})\phi - \mathcal{H}R\boldsymbol{\eta}][(I - \mathcal{H}R\mathcal{G})\phi - \mathcal{H}R\boldsymbol{\eta}]^T\} \\
&= \underbrace{(\mathcal{H}^T\mathcal{H})^{-1}\mathcal{H}^T}_{\mathcal{F}=\text{Fitting Operator}} \; \underbrace{C_\phi\mathcal{G}^T(\mathcal{G}C_\phi\mathcal{G}^T + C_\eta)^{-1}}_{\mathcal{E}=\text{Estimation Operator}}
\end{aligned}
$$

Fitting operator $\mathcal{F} = (\mathcal{H}^T\mathcal{H})^{-1}\mathcal{H}^T$ maps phase estimate $\hat{\phi}$ to actuator command $\mathbf{a}$;

$$
\begin{aligned}
\hat{\phi} &= \mathcal{E}\mathbf{s} = C_\phi\mathcal{G}^T\,(\mathcal{G}C_\phi\mathcal{G}^T + C_\eta)^{-1}\mathbf{s} \\
&= (\mathcal{G}^T C_\eta^{-1}\mathcal{G} + C_\phi^{-1})^{-1}\mathcal{G}^T C_\eta^{-1}\mathbf{s} \\
&= (\mathcal{G}^T\mathcal{G} + \sigma_\eta^2 C_\phi^{-1})^{-1}\mathcal{G}^T\mathbf{s}, \quad \text{provided that } C_\eta = \sigma_\eta^2 I \\
&= \arg\min_\phi \{||\mathcal{G}\phi - \mathbf{s}||^2 + \sigma_\eta^2\,\phi^T C_\phi^{-1}\phi\}
\end{aligned}
$$

Minimum variance phase estimation is equivalent to Tikhonov regularization applied to equation $\mathcal{G}\phi = \mathbf{s}$ with penalty parameter $\alpha = \sigma_\eta^2$ and penalty operator $L = C_\phi^{-1}$.

Approaches to Reconstructor Computation

Poke Matrix Inversion. $R = P^{\dagger}$, where "poke matrix" $P = \mathcal{G}\mathcal{H}$ maps actuators $\mathbf{a}$ to sensors $\mathbf{s}$.

- Unsensed modes cannot be recovered.
- Can be unstable if P has small singular values.
- Doesn't incorporate prior information; not adaptive.

Minimum Variance Reconstructor (Walner Decomposition).

$$R_{\mathrm{MV}} = (\underbrace{\mathcal{H}^T\mathcal{H}}_{R_W})^{-1}\,\underbrace{\mathcal{H}^T C_\phi \mathcal{G}^T}_{A_W}(\underbrace{\mathcal{G} C_\phi \mathcal{G}^T + C_\eta}_{S_W})^{-1}$$

- Requires inversion of $R_W = \mathcal{H}^T\mathcal{H}$ (easy); additional regularization may be needed if R_W has small singular values.
- Requires inversion of S_W (hard for large matrices).
- Inversion of S_W is stable due to C_η term.

MV Reconstruction via F-E Decomposition

Assume ϕ is discretized; replace $\mathcal{H} \leftarrow H$ and $\mathcal{G} \leftarrow G$.

$$R_{\mathrm{MV}} = \underbrace{(H^T H)^{-1} H^T}_{F} \underbrace{(G^T C_\eta^{-1} G + C_\phi^{-1})^{-1} G^T C_\eta^{-1}}_{E}$$

- Fitting step requires inversion of $H^T H$, perhaps with regularization (easy).
- Estimation step is stable, due to regularization.
- Estimation step requires inversion of C_η (easy).
- Estimation step requires inversion of C_ϕ and $A = G^T C_\eta^{-1} G + C_\phi^{-1}$ (hard, but good approximations exist to make this much easier).

Matrix Inversion

Canonical Problem: Solve linear system $A\mathbf{x} = \mathbf{b}$, where A is nonsingular.

- In fitting step, $A = H^T H$.
- In estimation step, $A = G^T C_\eta^{-1} G + L$, where $L \approx C_\phi^{-1}$.

Gaussian Elimination. General-purpose algorithm to solve canonical problem.

- Works (at least in principle) for any nonsingular matrix A.
- Complexity, or computational cost, is $N^3/3 + \mathcal{O}(N^2)$ when A is $N \times N$. Storage requirements are $\sim N^2$. Not practical when N is large.

More efficient algorithms to solve canonical problem must take advantage of special structure of A.

- Sparsity
- Spectral structure (eigenvalues and eigenvectors).

Sparse, Banded Matrices

Matrix A is called sparse is most of its entries are zeros.

More precisely, let $\mathrm{nz}(A)$ denote the number of nonzero entries in A, and let A be $N \times N$. Then A is sparse if

$$\mathrm{nz}(A) \ll N^2.$$

Sparse matrix A is called banded with bandwidth w if

$$a_{ij} = 0 \ \text{ whenever } \ |i - j| > w.$$

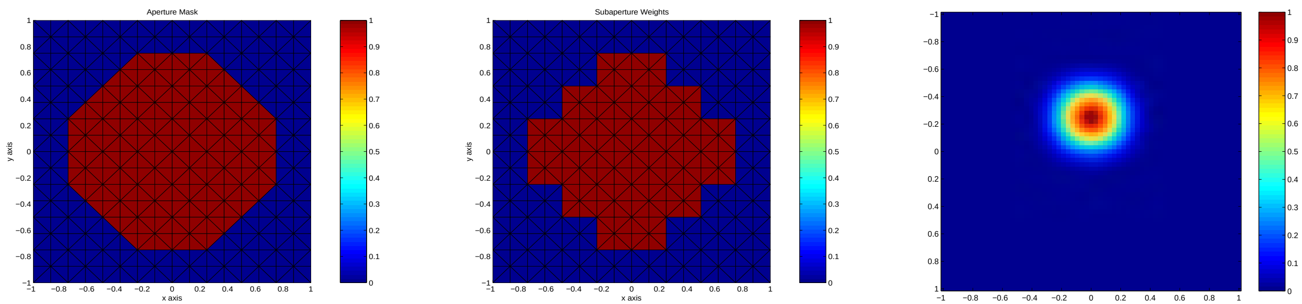
If A has bandwidth $w \ll N$, can modify Gaussian elimination (LU factorization) so that

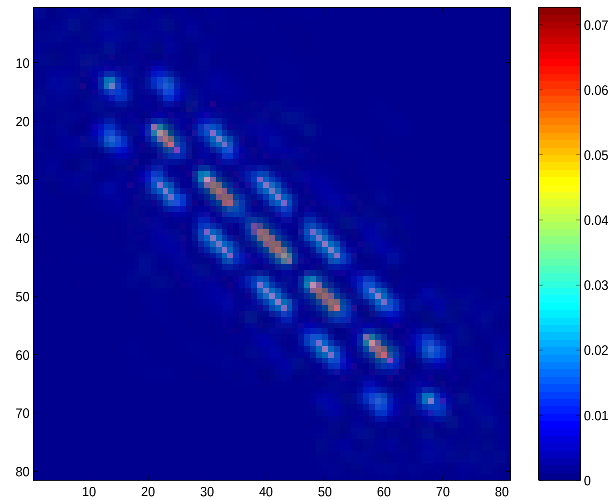
- Storage is $N \times w$.
- Complexity is $N \times w + \mathcal{O}(N)$.

Resulting method is called a sparse, banded direct solver.

Application: Fitting Step in AO Wavefront Reconstruction

Requires inversion of matrix $A = H^T H$, where ith column of H is (discretized) response to ith actuator. For DM's with piezo-electric stack actuators, $N \times N$ matrix A is sparse and banded with bandwidth $w \sim \sqrt{N}$.





Storage and computational cost are both $\sim N^{3/2}$.

Iterative Methods for Linear Systems

Typically fall into 2 classes:

- Classical stationary fixed-point iterations based on matrix splittings.
- Krylov subspace methods.

Can combine 2 approaches, e.g., use splitting-based iteration as a preconditioner for a Krylov method.

Gauss-Seidel Iteration

Classical stationary fixed-point iterations based on splitting

$$A = L + D + U \quad \text{(lower triangular + diagonal + upper triangular)}$$

Derivation of Method: $A\mathbf{x} = \mathbf{b} \;\Leftrightarrow\; (L+D)\mathbf{x} = \mathbf{b} - U\mathbf{x} \;\Leftrightarrow\; \mathbf{x} = (L+D)^{-1}(\mathbf{b} - U\mathbf{x})$.

Iteration is $\mathbf{x}_{k+1} = (L+D)^{-1}(\mathbf{b} - U\mathbf{x}_k),\; k = 0, 1, \ldots$

- $L + D$ is inverted using forward elimination (analogous to back substitution).
- Cost per iteration is $\sim \mathrm{nz}(A)$. This is often $\sim N$.
- Block variants are useful for estimation step in MCAO. Cost per iteration is dominated by inversion of diagonal blocks.
- Asymptotic convergence rate is usually slow, unless diagonal (or block diagonal) terms are relatively large.

Krylov Methods

If initial guess $\mathbf{x}_0 = \mathbf{0}$, these generate sequence of polynomial approximations to A^{-1}:

$$\mathbf{x}_{k+1} = \underbrace{(c_0 I + c_1 A + \ldots + c_k A^k)}_{p_k(A)} \mathbf{b} \;\approx\; A^{-1}\mathbf{b} \stackrel{\text{def}}{=} \mathbf{x}_*.$$

Best-known Krylov method is conjugate gradient iteration (CG).

- CG requires that A is symmetric and positive definite (SPD).
- CG is optimal in sense that "energy"

  $$E(\mathbf{x}_{k+1}) \stackrel{\text{def}}{=} (\mathbf{x}_{k+1} - \mathbf{x}_*)^T A (\mathbf{x}_{k+1} - \mathbf{x}_*)$$

  is minimized over all $\mathbf{x}_{k+1} = p_k(A)\mathbf{b}$, where the degree of polynomial p_k is $\leq k$.
- Requires one vector-matrix multiply per iteration, so cost per iteration is $\sim \mathrm{nz}(A)$.
- Convergence is fast if condition number is small or if eigenvalues of A are "clustered" (i.e., they have relatively little spread). Error bound:

$$E(\mathbf{x}_k) \leq 4 \left(\frac{\sqrt{\mathrm{cond}(A)} - 1}{\sqrt{\mathrm{cond}(A)} + 1} \right)^{2k} E(\mathbf{x}_0)$$

Preconditioning

Transformation from $A\mathbf{x} = \mathbf{b}$ to $M^{-1}A\mathbf{x} = M^{-1}\mathbf{b}$. Nonsingular matrix M is called the preconditioner. Krylov iterative method is then applied to the transformed system.

To be effective

- $\tilde{A} = M^{-1}A$ must have a smaller condition number (better eigenvalue distribution) than does A. This implies faster convergence.
- Vector-matrix multiplication $M^{-1}\mathbf{v}$ must be cheap. This implies low cost per iteration.

Preconditioned conjugate gradient iteration (PCG) is standard for SPD systems.

- Preconditioner M must be SPD.
- Cost per iteration dominated by vector-matrix multiplications $A\mathbf{v}$ and $M^{-1}\mathbf{v}$.

Choice of Preconditioners

No single preconditioner works well for all problems. Choice of preconditioner should be based on problem structure.

- ILU Preconditioner (stands for Incomplete Lower – Upper matrix factorization). Works well for some sparse systems.
- Multigrid Preconditioners. Work well for certain discretizations of strongly elliptic partial differential equations, e.g., Laplace's equation. Can have complexity $\sim N$!
- (Block) Circulant Preconditioners. Work well for certain (block) Toeplitz systems. These rely on the FFT and have complexity $\sim N \log N$. Variants rely on the fast cosine transform.

What Works for AO Wavefront Reconstruction?

Existing fast methods use fitting-estimation decomposition of reconstructor $R_{\mathrm{MV}} = FE$, with $F = (H^T H)^{-1} H^T$ and $E = (G^T C_\eta^{-1} G + C_\phi^{-1})^{-1}\, G^T C_\eta^{-1}$.

Fitting Step. Invert $A = H^T H$, where the “influence matrix” H maps actuators to mirror deformations. For piezo-electric stack actuators, H is sparse and banded due to local support of the influence functions.

- Sparse, banded direct solvers work well. Complexity is $\sim N^{3/2}$.
- CG works OK. PCG with ILU preconditioning may be better.

Estimation Step. Invert $A = G^T C_\eta^{-1} G + C_\phi^{-1}$. Note: A is SPD.

Special structure for conventional AO with Shack-Hartmann sensors:

- C_η is sparse. If sensor interactions are negligible, C_η is diagonal.
- G is sparse.
- G looks like a discrete gradient, so $G^T G$ behaves like a discrete Laplacian.

Phase Covariance Approximations

Assume Kolmogorov statistics, conventional AO.

- C_ϕ is positive semidefinite (piston mode is in null space).
- On a regular rectangular grid, C_ϕ is block Toeplitz. This follows from stationarity.
- C_ϕ is not sparse; entries decay as $r^{-5/3}$. Sparse wavelet representation???

Block Circulant Approximation.

- Imbed aperture in larger rectangular computational domain. Use block circulant approximation to C_ϕ.
- C_ϕ^{-1} also block circulant; can be computed using 2-D FFTs with cost $\sim M \log M$.
- $M \geq 4N$ to prevent periodic artifacts; need more storage.
- Use Fast Cosine Transform?

Ellerbroek's sparse appoximation to inverse covariance.

- Motivation: Eigenvalues of C_ϕ decay as $\kappa^{-11/3}$; eigenvalues of the biharmonic (squared Laplacian) grow as κ^4; approximate C_ϕ^{-1} by the biharmonic.
- With standard finite difference or finite element approximations to the biharmonic, C_ϕ^{-1} is sparse and banded with bandwith $w = \sqrt{N}$ and $nz(A) \sim N$.
- Natural boundary conditions?

Implementation Details for Block Circulant Approximation

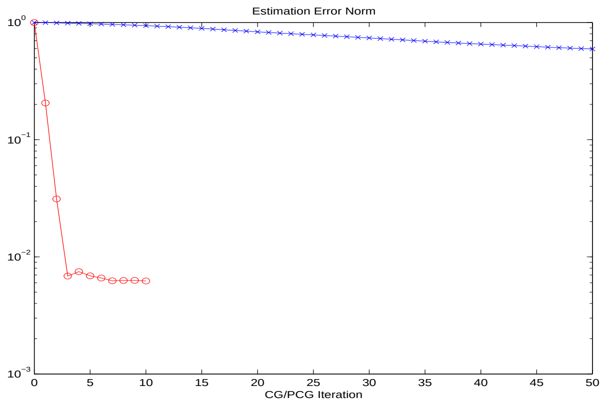
- With circular aperture and rectangular computational domain, $G^T C_\eta^{-1} G$ is not Toeplitz. Can't us FFTs to directly solve

$$(G^T C_\eta^{-1} G + C_\phi^{-1})\boldsymbol{\phi} = G^T C_\eta^{-1} \mathbf{s} \stackrel{\text{def}}{=} \mathbf{b}.$$

- Instead use PCG with preconditioner based on splitting

$$\underbrace{(\omega I + C_\phi^{-1})}_{\text{block circulant}}\ \boldsymbol{\phi} = \mathbf{b} - \underbrace{(\omega I - G^T C_\eta^{-1} G)}_{\text{sparse}}\ \boldsymbol{\phi}, \quad \omega > 0.$$

Cost per iteration is $\sim M \log M$; PCG convergence is fast.



Implementation Details for Sparse Covariance Approx

Let L denote discrete biharmonic. Approximate C_ϕ^{-1} by γL; can pick scaling factor

$$\begin{aligned} \gamma &= \arg\min \langle ||\hat{\phi} - \phi||^2 \rangle, \quad \hat{\phi} = \gamma L^{-1/2} w, \;\; \phi = C^{1/2} w \\ &= \arg\min \operatorname{Trace}\{(\gamma L^{-1/2} - C^{1/2})(\gamma L^{-1/2} - C^{1/2})^T\} \\ &= \operatorname{Trace}(L^{-1/2} C^{1/2}) \;/\; \operatorname{Trace}(L). \end{aligned}$$

Need to invert $A = (G^T G + \gamma L)$.

Direct Approach.

- A is sparse and banded with bandwidth $w = N^{1/2}$. Sparse, banded direct solvers have complexity and storage $\sim N^{3/2}$.

Iterative Approach.

- A is SPD and $\mathrm{nz}(A) \sim N$, so cost of vector-matrix multiplication with A is $\sim N$.
- As preconditioner, use multigrid with symmetric Gauss-Seidel smoother. Works well provided noise level is not too high (Laplacian term $G^T C_\eta^{-1} G$ dominates biharmonic term γL).
- Only 2 or 3 PCG iterations needed. Total cost is $\sim N$!

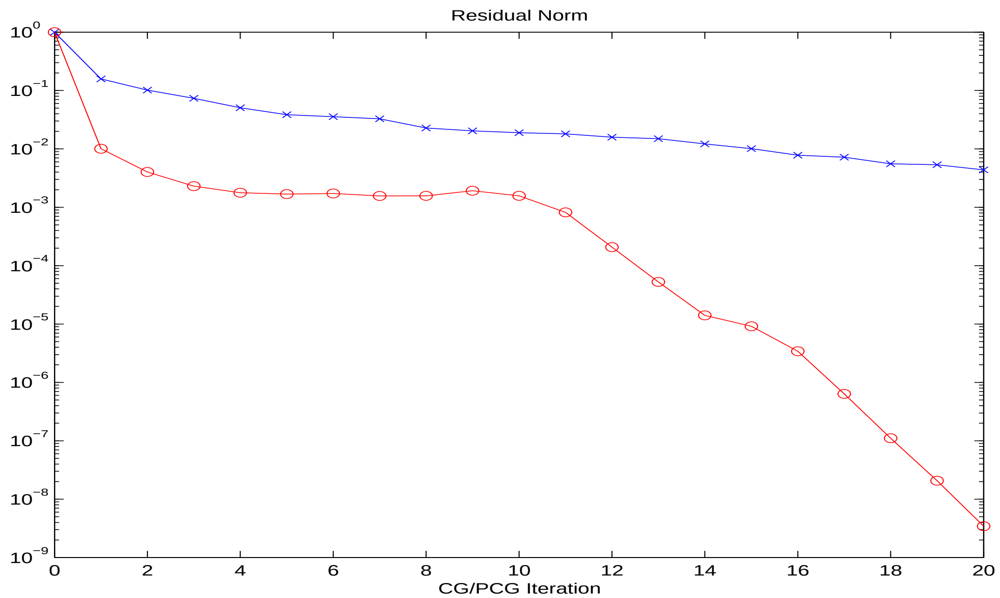
Computational Linear Algebra for MCAO Fitting Step

Requires inversion of matrix $A = \tilde{H}^T \tilde{H}$, where $\tilde{H}$ is a row-block matrix with blocks $\tilde{H}_j = P_j H_j$.

- P_j represents propagation from turbulence layer at height z_j to the ground.
- H_j represents actuator influence on DM conjugate to layer z_j.

Direct Approach: A is sparse, but bandwidth is relatively large; reordering of unknowns can reduce bandwidth before application of sparse, banded direct solver.

Iterative Approach: PCG with ILU preconditioning is effective. Cost per iteration is $\sim N$.



Estimation (Tomography) Step in MCAO

Wavefront sensor measurements can be represented (after discretization) as

$$\mathbf{s}_k = G\left(\sum_{\ell=1}^{n_L} P_{k\ell}\rho_\ell\right) + \boldsymbol{\eta}_k,$$

where $P_{k\ell}$ represents propagation in direction k from layer ℓ to the ground. Need to invert block matrix $\tilde{A}$ with blocks

$$\tilde{A}_{ij} = \sum_k P_{ki}^T G^T C_{\eta_k}^{-1} G P_{kj} + \delta_{ij}\alpha_i L_i.$$

Here $\delta_{ij} = 1$ if $i = j$ and 0 otherwise, and L_i represents regularization for layer i.

- Take each L_i to be a discrete biharmonic. Then $\mathrm{nz}(\tilde{A}) \sim N$.
- PCG with multigrid preconditioner, block symmetric Gauss Seidel smoother is effective. Inversion of diagonal blocks is dominant cost. With sparse, banded direct method, this cost is $\sim N_\ell^{3/2}$, where N_ℓ denotes size of diagonal blocks.

Selected Comp. Math & Inverse Problems References

1. O. Axelsson and V. A. Barker, *Finite Element Solution of Boundary Value Problems*, SIAM, Philadelphia, 2001.
2. Y. Saad, *Iterative Methods for Sparse Linear Systems*, (PWS Publishing Company, Boston, 1996).
3. U. Trottenberg, C. W. Oosterlee and A. Schüller, *Multigrid*, (Academic Press, London, 2001).
4. C. Vogel, *Computational Methods for Inverse Problems*, SIAM, 2002.

Selected AO References

1. B. L. Ellerbroek, “Efficient computation of minimum variance wavefront reconstructors using sparse matrix techniques,” Journal of the Optical Society of America A, **19** (2002), pp. 1817-1822.
2. B. L. Ellerbroek, Luc Gilles, and C. R. Vogel, “A Computationally efficient wavefront reconstructor for simulation of multi-conjugate adaptive optics on giant telescopes”, Proc. SPIE 4839-116, Adaptive Optics System Technologies II (2002).
3. B. L. Ellerbroek, Luc Gilles, and C.R. Vogel, "Numerical simulations of multiconjugate adaptive optics wavefront reconstuction on giant telescopes", Applied Optics (2003).
4. B. L. Ellerbroek and C.R. Vogel, "Simulations of closed-loop wavefront reconstruction for multiconjugate adaptive optics on giant telescopes", Proc. SPIE 5169-23, Adaptive Optics System Technologies II (2003).
5. Luc Gilles, “Order N Sparse Minimum-Variance Open-Loop Reconstructor for Extreme Adaptive Optics”, Optics Letters **28**, pp. 1927-1929 (2003)
6. Luc Gilles, C.R. Vogel, and B. L. Ellerbroek, “A Multigrid Preconditioned Conjugate Gradient Method for Large Scale Wavefront Reconstruction”, Journal of the Optical Society of America A, **19** (2002), pp. 1817-1822.

Selected AO References, Continued

7 Luc Gilles, Brent Ellerbroek, and C.R. Vogel, “Layer-oriented multigrid wavefront reconstruction algorithms for multi-conjugate adaptive optics", Proc. SPIE 4839-118, Adaptive Optics System Technologies II (2002).

8 E. P. Walner, “Optimal wavefront correction using slope measurements,” Journal of the Optical Society of America **73**, pp. 1771–1776 (1983).