Bayesian Scientific Computing
Day 4

Daniela Calvetti, Erkki Somersalo

Case Western Reserve University
Department of Mathematics, Applied Mathematics and Statistics

Lyngby, December 2019

Introduction

Iterative linear systems solvers = a different way of solving $Ax = b$.

- Direct method
- Get a factorization of matrix, e.g., LU via Gaussian elimination, Cholesky,;
- Solve two triangular systems.
- Hit or miss: if you stop partway all is lost!

- Iterative method
- Start with an initial approximate solution x_0, e.g., 0;
- At each step correct the current approximation to be closer to the true solution x_*.
- Can stop at any time and have an approximate solution.

Iterative linear solver: specs and wish list

The key steps defining an iterative method are

- How the current solution is corrected;
- How the iterative process is terminated.

In the context of Bayesian Inverse Problem we want iterative linear solvers

- to approximate the MAP solution in few iterations, possibly working on a small system if data are few;
- retain the nice properties of classical regularization schemes, at a fraction of the computational cost.

Iterative linear solvers: zooming in

If the matrix $A \in \mathbb{R}^{m \times n}$, is very large, sparse, or does not exist, it is attractive to solve $Ax = b$ by

1. Starting from an arbitrary initial approximate solution, e.g., $x_0 = 0$;
2. Building a sequence of approximate solutions computing only products with A or A^T;
3. Stopping the iteration when $\|b - Ax_k\|$ is small enough (or the max number of steps is reached).

Iterative solvers

The desirable traits of an iterative linear system solver are that

- The method is simple to implement.
- The method is rapidly convergent.
- The method is stable with respect to roundoff errors.
- At each step, the new approximate solution is a *better* approximation of the solution than the current one.
- At each step the quality of the approximation can be assessed at negligible additional cost.

The Conjugate Gradient Method

Needed: $A \in \mathbb{R}^{m \times m}$ symmetric positive definite, $b \in \mathbb{R}^m$.

The Conjugate Gradient Method (1952)

Initialize: Given x_0, $i = 0$, compute $p_0 = r_0 = b - Ax_0$

Iterate: Until stopping criterion is satisfied:

$$\alpha_i = \frac{\|r_i\|^2}{p_i^T A p_i}$$
$$x_{i+1} = x_i + \alpha_i p_i$$
$$r_{i+1} = r_i - \alpha_i A p_i$$
$$\beta_i = \frac{\|r_{i+1}\|^2}{\|r_i\|^2}$$
$$p_{i+1} = r_{i+1} + \beta_i p_i$$
$$i = i + 1$$

end

Cost: one product with matrix A per iteration.

Minimization property of CG

The CG iterate x_j minimizes the error function

$$f(x) = (x - x^*)^T A(x - x^*)$$

over $\text{span}\{b, Ab, \ldots, A^{j-1}\}$, i.e., $x_j \in \text{span}\{b, Ab, \ldots, A^{j-1}\}$.

- Since A spd, the linear system $Ax = b$ has a unique solution x^*;
- The spaces where the iterates live are nested

$$\text{span}\{b\} \subset \text{span}\{b, Ab\} \subset \ldots \subset \text{span}\{b, Ab, \ldots, A^{j-1}b\}$$

- The minimum gets smaller at each iteration;
- In at most n steps the iterates converge to the unique unconstrained minimizer of $f(x)$, x^*.

Generalization of CGLS

If $A \in \mathbb{R}^{m \times n}$ has linearly independent columns, then

- The matrix $A^T A$ is symmetric positive definite;
- The least squares solution to $Ax = b$ solves the normal equations

$$A^T A x = A^T b;$$

- Can apply CG to solve the normal equations.

There is a variant of the CG method for solves the latter system without forming $A^T A$.

CG method for Least Squares (CGLS)

Initialize: Given x_0, compute $d_0 = b - Ax_0$, $\quad p_0 = r_0 = A^T r_0$; $y = Ap_0$

Iteration: For $i = 0$ until a stopping criterion is satisfied:

1. $\alpha_i = \frac{\| r_i^2}{\|Ap_i|^2}$
2. $x_{i+1} = x_i + \alpha_i p_i$
3. $d_{i+1} = d_i - \alpha_i A p_i$;
4. $r_{i+1} = A^T d_i$;
5. $\beta_i = \frac{\|r_{i+1}\|^2}{\|r_i\|^2}$
6. $p_{i+1} = r_{i+1} + \beta_i p_i$
7. $i = i + 1$

Cost: one product with matrix A and one with A^T per iteration.

CGLS minimization properties

At the kth iteration step the approximate solution x_k computed by the CGLS method minimizes the error function

$$f(x) = (x - x^*)^{\mathsf{T}} \mathsf{A}^{\mathsf{T}} \mathsf{A} (x - x^*)$$

over the kth Krylov subspace associated with $\mathsf{A}^{\mathsf{T}} b$ and $\mathsf{A}^{\mathsf{T}}\mathsf{A}\mathsf{A}$

$$\mathscr{K}_k(\mathsf{A}^{\mathsf{T}} b, \mathsf{A}^{\mathsf{T}}\mathsf{A}) = \mathrm{span}\{\mathsf{A}^{\mathsf{T}} b, (\mathsf{A}^{\mathsf{T}}\mathsf{A})\mathsf{A}^{\mathsf{T}} b, \ldots, (\mathsf{A}^{\mathsf{T}}\mathsf{A})^k \mathsf{A}^{\mathsf{T}} b\}.$$

Equivalently, x_k satisfies

$$x_k = \mathrm{argmin}\{\|b - \mathsf{A}x\| \mid x \in \mathscr{K}_k\}.$$

CGLS iterates and discrepancies

By construction $x_k \in \mathscr{K}_k(\mathsf{A}^\mathsf{T} b, \mathsf{A}^\mathsf{T}\mathsf{A})$, thus $x_k \in \mathcal{R}(\mathsf{A}^\mathsf{T})$, and for $k = 1, 2, \ldots,$

$$x_k \perp \mathcal{N}(\mathsf{A}).$$

Moreover, since

$$\mathscr{K}_1 \subset \mathscr{K}_2 \subset \ldots \subset \mathscr{K}_k \subset$$

the norm of the discrepancy forms a decreasing sequence, and the norm of the iterates an increasing sequence,

$$\|b\| \geq \|b - \mathsf{A}x_1\| \geq \|b - \mathsf{A}x_2\| \geq \ldots \geq \|b - \mathsf{A}x_k\| \geq$$

$$\|x_0\| \leq \|x_1\| \leq \ldots \leq \|x_k\| \leq$$

so we decrease the norm of the discrepancy at the cost of increasing the norm of the solution.

Regularization by truncated CGLS

The properties of the CGLS iterates can be used to solve linear discrete inverse problems.

If we know the norm of the additive noise, i.e. $\|\eta\| \approx \delta$,

- start solving the linear system $Ax = b$ with CGLS
- stop right before the amplification of the noise start to dominates, i.e. at x_k where

$$\|b - Ax_k\| \geq \delta > \|b - Ax_{k+1}\|.$$

This is known as Morozov Discrepancy Principle (MDP).

CGLS stopped on MDPt

If the noise is additive, zero-mean white Gaussian, then

$$E\left\{\|\epsilon\|^2\right\} = m;$$

we stop iterating as soon as

$$\|Ax - b\|^2 < \tau m,$$

where $\tau > 1$ is a safeguard factor. Typically, $k_{\text{last}} \ll m$.

MAP via CGLS.

In the Gaussian-Gaussian-Linear case, the Maximum a Posteriori (MAP) estimate of x, x_{MAP} is

- the value of highest posterior probability, or
- the minimizer of the negative log-posterior $\mathsf{G}(x)$, or
- the *least squares* solution of the linear system

$$\begin{bmatrix} \mathsf{A} \\ \mathsf{L} \end{bmatrix} x = \begin{bmatrix} b \\ 0 \end{bmatrix},$$

  that can be computed via CGLS, or
- the solution of the square linear system

$$(\mathsf{A}^{\mathsf{T}}\mathsf{A} + \mathsf{L}^{\mathsf{T}}\mathsf{L})x = \mathsf{A}^{\mathsf{T}}b.$$

  that can be computed by CG method.

Joy and pain of L

- The operator L brings into the solution additional information about x
- When the matrix A is underdetermined, and L is invertible, L makes the two-story matrix full rank, thus "filling" the null space .
- When n is large the presence of L may lead to a very large linear system

Can we retain the benefits of L keeping the computational cost down?

Whitening and preconditioning

- Letting $w = Lx$ and substituting $x = \mathrm{L}^{-1} w$ we obtain

$$\begin{bmatrix} \mathrm{A} \\ \mathrm{L} \end{bmatrix} x = \begin{bmatrix} b \\ 0 \end{bmatrix} \Rightarrow \begin{bmatrix} \mathrm{A}\mathrm{L}^{-1} \\ \mathrm{I} \end{bmatrix} w = \begin{bmatrix} b \\ 0 \end{bmatrix},$$

Whitening and preconditioning

- Idea 1: Ignore the lower half of matrix and solve $Ax = b$ via CGLS plus MDP. This solves much smaller system but ignore benefits of L.
- Idea 2: Solve $AL^{-1}w = b$ via CGLS plus MDP, then set $x_k = L^{-1}w_k$. $\|w_j\|$ increase monotonically and only as needed to explain signal. In the transformation

$$AL^{-1}w = b,\ Lx = w$$

  L whitenes the unknown and acts as a $\mathrm{right\ preconditioner}$ or *priorconditioner*.
- Idea 3: Proceed as in Idea 2 but stopping either on MDP or as soon as $\|AL^{-1}w_k - b\|^2 + \|w_k\|^2$ stops decreasing.

Priorconditioning and the null space of A

The jth CGLS iterate of the whitened problem $\widetilde{\mathsf{A}}w = b$, $\widetilde{\mathsf{A}} = \mathsf{A}\mathsf{L}^{-1}$ satisfies

$$w_j = \operatorname{argmin}\{\|\widetilde{\mathsf{A}}w - b\| \mid w \in \mathcal{K}_j(\widetilde{\mathsf{A}}^{\mathsf{T}} b, \widetilde{\mathsf{A}}^{\mathsf{T}}\widetilde{\mathsf{A}})\}.$$

The corresponding jth priorconditioned CGLS solution $\widetilde{x}_j = \mathsf{L}^{-1} w_j$ satisfies

$$\widetilde{x}_j \in \operatorname{span}\{\mathsf{L}^{-1}(\widetilde{A}^{\mathsf{T}}\widetilde{A})^{\ell}\widetilde{A}^{\mathsf{T}} b \mid 0 \le \ell \le j-1\}.$$

It follows from

$$\mathsf{L}^{-1}\widetilde{A}^{\mathsf{T}} = \mathsf{L}^{-1}\mathsf{L}^{-\mathsf{T}}\mathsf{A}^{\mathsf{T}} = \mathsf{C}\mathsf{A}^{\mathsf{T}},$$

that

$$\mathsf{L}^{-1}\big(\widetilde{A}^{\mathsf{T}}\widetilde{\mathsf{A}}\big)^{\ell}\widetilde{\mathsf{A}}^{\mathsf{T}} = \big(\mathsf{C}\mathsf{A}^{\mathsf{T}}\mathsf{A}\big)^{\ell}\mathsf{C}\mathsf{A}^{\mathsf{T}}, \quad 0 \le \ell \le j-1.$$

Therefore

$$\widetilde{x}_j \in \mathsf{C}\big(\mathcal{N}(\mathsf{A})^{\perp}\big),$$

hence $\widetilde{x}_j$ is not necessarily orthogonal to the null space of A.

Analyzing the Krylov subspaces with the GSVD

Theorem

Given (A, L) with $A \in \mathbb{R}^{m\times n}$, $L \in \mathbb{R}^{n\times n}$, $m < n$, there is a factorization of the form

$$A = U \begin{bmatrix} O_{m,n-m} & \Sigma_A \end{bmatrix} X^{-1}, \quad L = V \begin{bmatrix} I_{n-m} & \\ & \Sigma_L \end{bmatrix} X^{-1},$$

called the *generalized singular value decomposition*, where $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ are orthogonal matrices, $X \in \mathbb{R}^{n\times n}$ is invertible, and $\Sigma_A \in \mathbb{R}^{m\times m}$ and $\Sigma_L \in \mathbb{R}^{m\times m}$ are diagonal matrices.

The diagonal entries $s_1^{(A)}, \ldots, s_m^{(A)}$ and $s_1^{(L)}, \ldots, s_m^{(L)}$ of the matrices Σ_A and Σ_L are real, nonnegative and satisfy

$$\begin{aligned} s_1^{(A)} &\leq s_2^{(A)} \leq \ldots \leq s_m^{(A)} \\ s_1^{(L)} &\geq s_2^{(L)} \geq \ldots \geq s_m^{(L)} \\ (s_j^{(A)})^2 &+ (s_j^{(L)})^2 = 1, \qquad 1 \leq j \leq m. \end{aligned} \tag{1}$$

thus $0 < s_j^{(A)} \leq 1$ and $0 < s_j^{(L)} \leq 1$. The ratios $s_j^{(A)}/s_j^{(L)}$ for $1 \leq j \leq m$ are the generalized singular values of (A, L).
If A has full rank, the diagonal entries of Σ_A are positive.

C-orthogonality

Theorem

If we partition the matrix $X \in \mathbb{R}^{n \times n}$ in GSVD above as

$$X = \begin{bmatrix} X' & X'' \end{bmatrix}, \quad X' \in \mathbb{R}^{n \times (n-m)}, \; X'' \in \mathbb{R}^{n \times m},$$

it follows that

$$\text{span}\{X'\} = \mathcal{N}(A),$$

and we can express $\mathbb{R}^n$ as a C-orthogonal direct sum,

$$\mathbb{R}^n = \text{span}\{X'\} \oplus_C \text{span}\{X''\} = \mathcal{N}(A) \oplus_C \text{span}\{X''\}.$$

Orthogonality and not

Corollary 1

$$\mathcal{N}(\mathsf{A})^{\perp} = \mathcal{R}(\mathsf{A}^{\mathsf{T}}), \quad \mathcal{N}(\mathsf{A})^{\perp_{\mathsf{C}}} = \operatorname{span}\{\mathsf{X}''\}.$$

Corollary 2
If $\mathcal{R}(\mathsf{A}^{\mathsf{T}})$ is an invariant subspace of the covariance matrix C, then the iterates $\widetilde{x}_j$ are orthogonal to the null space of A.

Corollary 3
When $\mathsf{C}\big(\mathcal{R}(\mathsf{A}^{\mathsf{T}})\big)$ is not C-orthogonal to $\mathcal{N}(\mathsf{A})$, $\widetilde{x}_j$ may have a component in the null space of A. This component is invisible to the data.

CGLS and Lanczos

To understand how preconditioning:

- changes the spectral properties of $M = A^T A$;
- changes the order in which eigendirections are included in the computed solution

we relate the CGLS method and the Lanczos process.

Lanczos process

Lanczos process computes an orthonormal basis of $\mathcal{K}_k(v, \mathrm{M})$ using the three term recurrence relation from the Gram-Schmidt orthogonalization for the spanning vectors.

Lanczos process for $\mathrm{M} = \mathrm{A}^{\mathsf{T}}\mathrm{A}$

Initialization: $v_1 = v/\|v\|;$

$\gamma_1 = v_1^{\mathsf{T}} \mathrm{M} v_1;$

$\eta_0 = 0;$

$w = \mathrm{M}v_1 - \gamma_1 v_1$

For $k = 2, \ldots$

$\eta_{k-1} = \|w\|$. If $\eta_k = 0$ stop.

$v_k = w/\eta_{k-1}$

$\gamma_k = v_k^{\mathsf{T}} \mathrm{M} v_k$

$w = \mathrm{M}v_k - \gamma_k v_k - \eta_{k-1} v_{k-1}$

end

The Lanczos tridiagonal matrix

After k steps of the Lanczos algorithm, the symmetric tridiagonal matrix

$$\mathsf{T}_k = \begin{bmatrix} \gamma_1 & \eta_1 & & & \\ \eta_1 & \gamma_2 & \eta_2 & & \\ & \ddots & \ddots & \ddots & \\ & & & & \eta_{k-1} \\ & & & \eta_{k-1} & \gamma_k \end{bmatrix}$$

satisfies

$$\mathsf{M}\mathsf{V}_k = \mathsf{V}_k \mathsf{T}_k + \eta_k v_{k+1} e_k^{\mathsf{T}}$$

where the columns of the matrix V_k are the orthonormal vectors v_j.

From CGLS to Lanczos

It has been shown that

- The orthonormal Lanczos vectors v_{k+1} are proportional to the CGLS residual vectors r_k;
- If α_j, β_j are the coefficients in the CGLS iterations,

$$\gamma_k = \frac{1}{\alpha_{k-1}} + \frac{\beta_{k-1}}{\alpha_{k-2}}, \qquad \eta_k = \frac{\sqrt{\beta_k}}{\alpha_{k-1}}, \quad \beta_0 = 0,\ \alpha_{-1} = 1$$

are the entries of Lanczos matrix T_k for $\mathsf{M} = \mathsf{A}^\mathsf{T}\mathsf{A}$ and $r_0 = \mathsf{A}^\mathsf{T} d_0$.

Priorconditionting and the Lanczos process

The first k residual vectors computed by CGLS normalized to have unit length $v_0, v_1, \ldots, v_{k-1}$ form an orthonormal basis for the Krylov subspace $\mathcal{K}_k(\mathrm{A}^\mathsf{T} b, \mathrm{A}^\mathsf{T}\mathrm{A})$.

It can be shown that

$$\mathrm{A}^\mathsf{T}\mathrm{A} V_k = V_k \mathrm{T}_k - \frac{\sqrt{\beta_{k-1}}}{\alpha_{k-1}} v_k e_k^\mathsf{T}., \qquad \mathrm{V}_k = [v_0, v_1, \ldots, v_{k-1}].$$

It follows from the orthogonality of the v_j that the tridiagonal matrix T_k is the projection of $\mathrm{A}^\mathsf{T}\mathrm{A}$ onto the Krylov subspace $\mathcal{K}_k(\mathrm{A}^\mathsf{T} b, \mathrm{A}^\mathsf{T}\mathrm{A})$.

$$\mathrm{V}_k^\mathsf{T}(\mathrm{A}^\mathsf{T}\mathrm{A}) V_k = \mathrm{T}_k.$$

The Lanczos tridiagonal matrix

The kth CGLS iterate $x_k \in \mathcal{K}_k(\mathsf{A}^\mathsf{T} b, \mathsf{A}^\mathsf{T}\mathsf{A})$ can be expressed as

$$x_k = \mathsf{V}_k y_k,$$

where y_k solves the $k \times k$ linear system

$$\mathsf{T}_k y = \|r_0\| e_1.$$

Thus the kth CGLS iterate x_k is the lifting of y_k via V_k.

The eigenvalues of T_k are the Ritz values of $\mathsf{A}^\mathsf{T}\mathsf{A}$ and approximate of the eigenvalues of $\mathsf{A}^\mathsf{T}\mathsf{A}$.

Ritz values and convergence rate

Theorem
For all k, $1 \leq k \leq r$, where r is the rank of A, there exists $\xi_k, \lambda_1 \leq \xi_k \leq \lambda_r$ such that the norm of the residual vector satisfies

$$\|r_k\|^2 = \frac{1}{\xi_k^{2k+1}} \sum_{i=1}^{n} \left[\prod_{j=1}^{k} \left(\lambda_i - \theta_j^{(k)} \right)^2 \right] \left(r_0^{\mathsf{T}} q_i \right)^2,$$

where q_i is the eigenvector of $\mathsf{A}^{\mathsf{T}}\mathsf{A}$ corresponding to the eigenvalue λ_i, and $\theta_j^{(k)}$ is the jth eigenvalue of the tridiagonal matrix T_k.

The quality of the eigenvalues approximations in the projected problem affects the number of iterations needed to meet the stopping rule.

A simple deconvolution problem

Forward model: Deconvolution problem with few data,

$$g(t) = \int_0^1 a(t-s)f(s)ds, \quad a(t) = \left(\frac{J_1(\kappa t)}{\kappa t}\right)^2,$$

Discretize:

$$g(t) \approx \frac{1}{n}\sum_{k=1}^{n} a(t-s_k)f(s_k), \quad 1 \le j \le n,$$

Discrete noisy observations at $t_1, \ldots, t_m$, $m \ll n$.

$$b_\ell = g(t_\ell) + \varepsilon_\ell, \quad 1 \le \ell \le m,$$

or, in matrix notation, $\mathsf{A} \in \mathbb{R}^{m\times n}$,

$$b = \mathsf{A}x + \varepsilon, \quad x_k = f(s_k).$$

Computed examples: Deconvolution

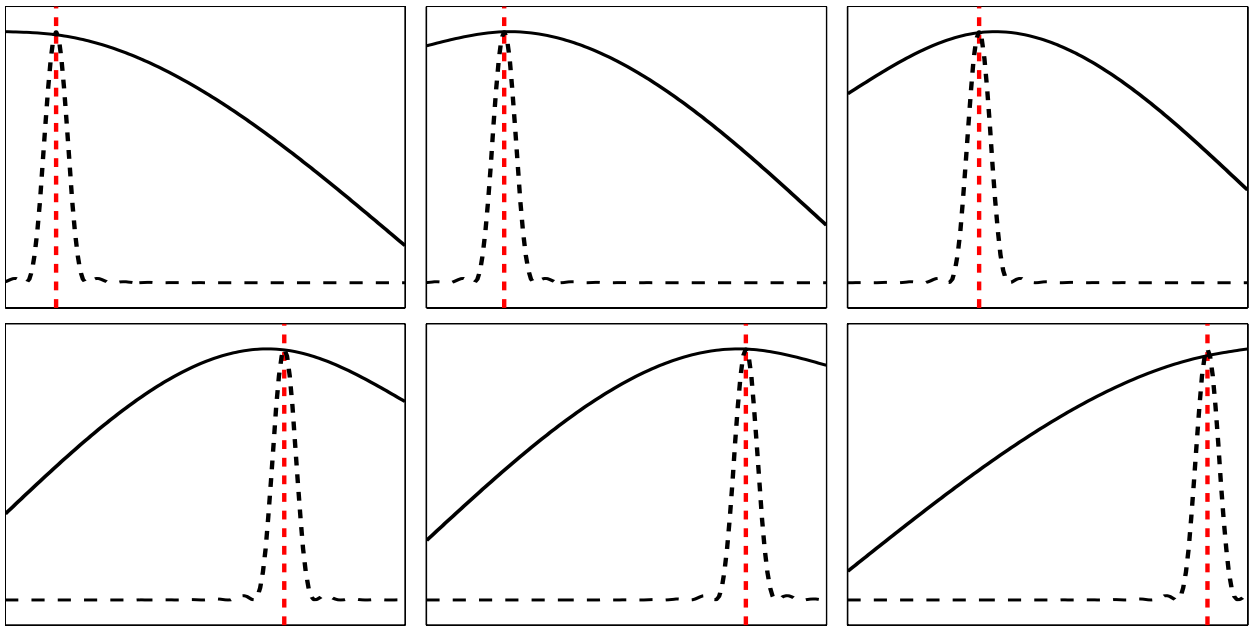
Prior: Define the precision matrix C^{-1} as

$$\mathsf{C}^{-1} = \mathsf{L}^{\mathsf{T}}\mathsf{L}, \quad \mathsf{L} = \beta \begin{bmatrix} \alpha & & & & \\ -1 & 2 & -1 & & \\ & & \ddots & & \\ & & -1 & 2 & -1 \\ & & & & \alpha \end{bmatrix},$$

where $\alpha > 0$ is chosen so that prior variance is as uniform as possible over the interval.

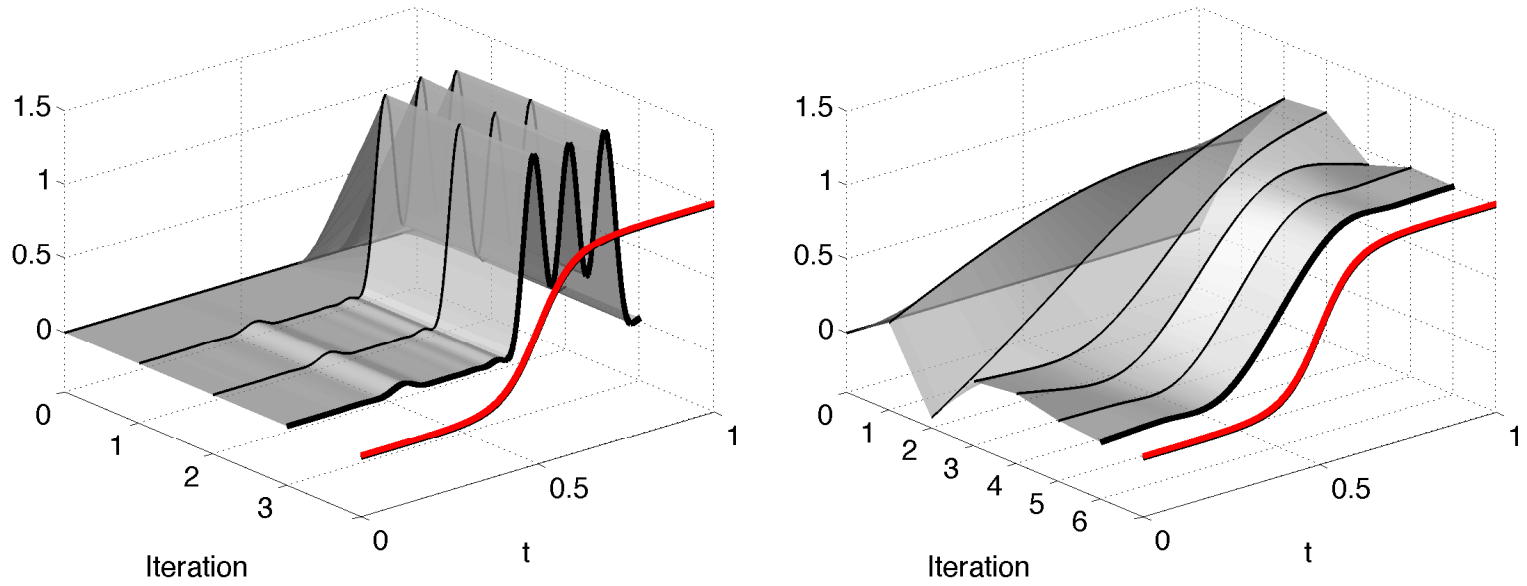
Parameters: Set $n = 150$, $m = 6$.

Basis vectors



The six basis vectors that span $\mathscr{R}(A^T)$ (dashed line), and the vectors that span $C\left(\mathscr{R}(A^T)\right)$ (solid line).

Approximate solutions

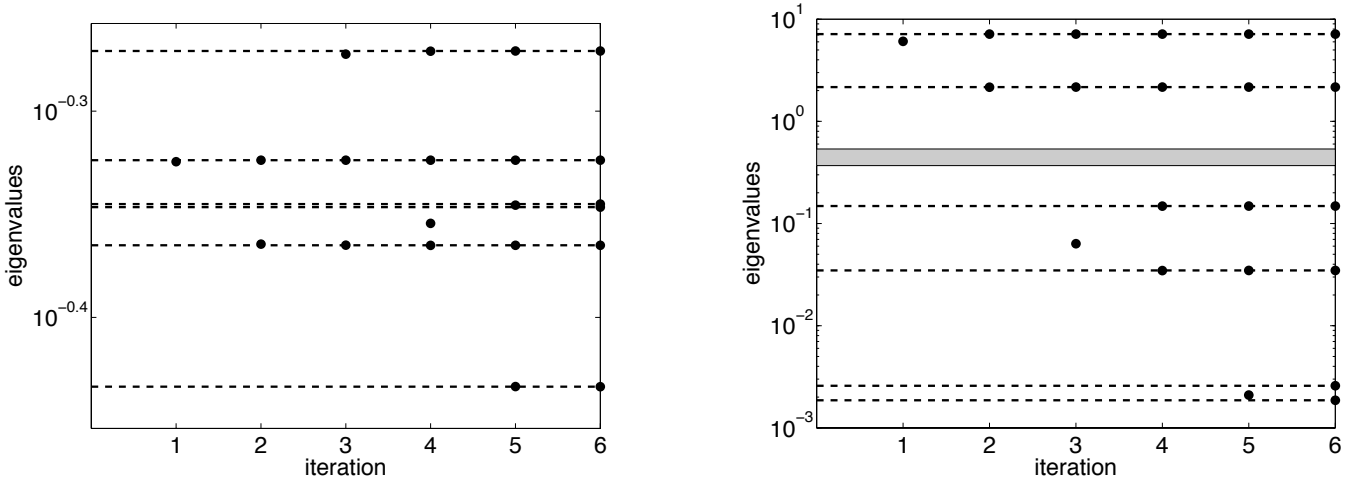


Iterations with low additive noise ($\sigma = 5 \times 10^{-5}$) without prior conditioner (left) and with preconditioner.

Observations

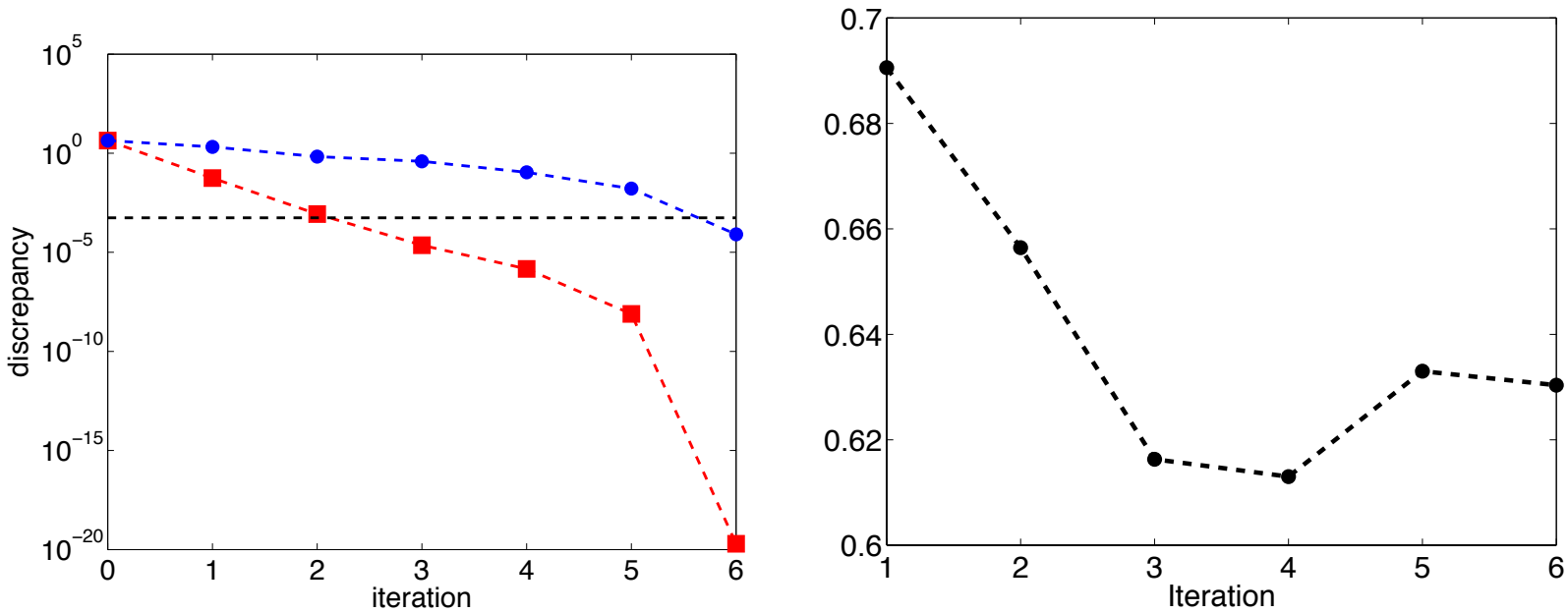
- Every vector whose support consists of points where all six basis functions of $\mathscr{R}(\mathsf{A}^{\mathsf{T}})$ vanish is in the null space $\mathscr{N}(\mathsf{A})$
- Consequently, plain CGLS produces approximate solutions that are zero at those points
- The basis functions of $\mathsf{C}\big(\mathscr{R}(\mathsf{A}^{\mathsf{T}})\big)$ are non-zero everywhere
- Consequently, priorconditioned CGLS has no blind spots
- The price to pay is that priorconditioned CGLS requires more iterations

Spectral approximation



Spectral approximation: Plain CGLS (left) and preconditioned CGLS (right). The grey band on the right is the spectral interval of the non-preconditioned matrix A^TA.

Convergence history and null space contributions



Left: Convergence rates of the two algorithms. The dashed line marks the stopping criterion. Right: Component of the computed solution in the null space measured as $\nu_k = \frac{\|\mathrm{P}\widetilde{x}_k\|}{\|\widetilde{x}_k\|}, \quad \mathrm{P} : \mathbb{R}^n \longrightarrow^{\perp} \mathcal{N}(\mathrm{A})$.

Computed examples: X-ray tomography

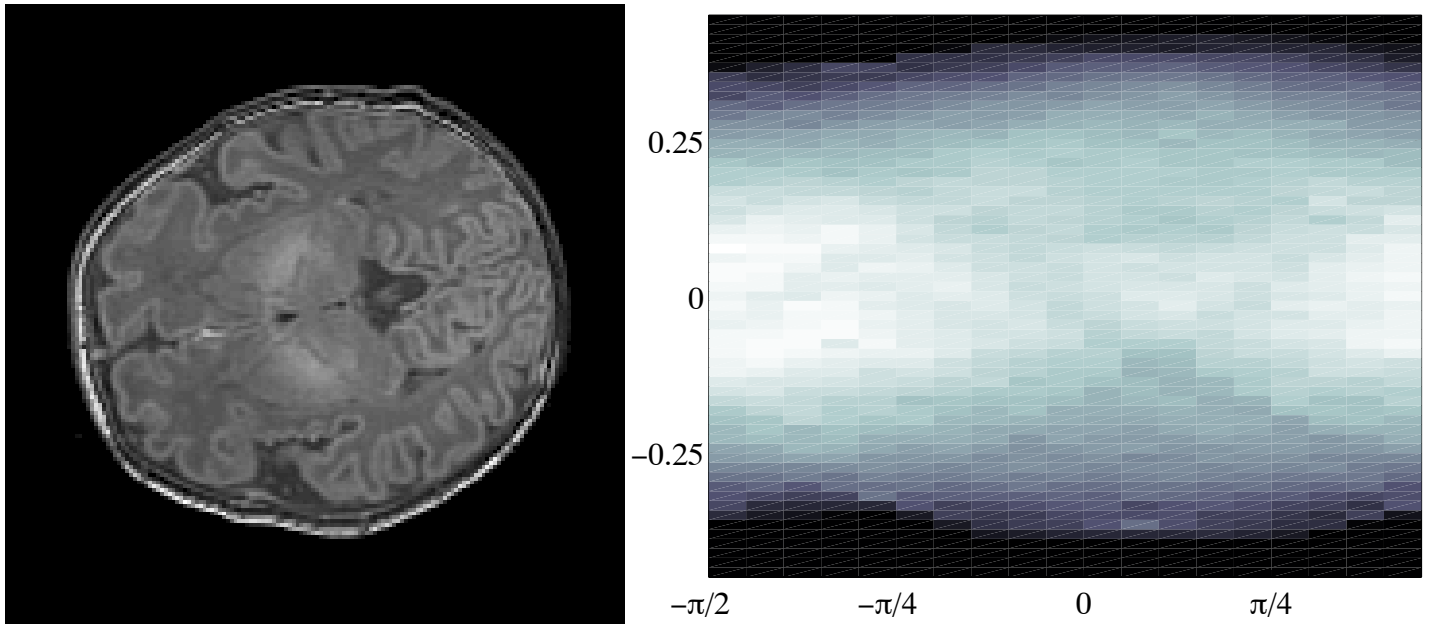


Image size: $N = 160 \times 160$ pixels. 20 illumination angles, 60 parallel beams per illumination angle.

Correlation priors

Matèrn-Whittle correlation priors: Define the precision matrix as

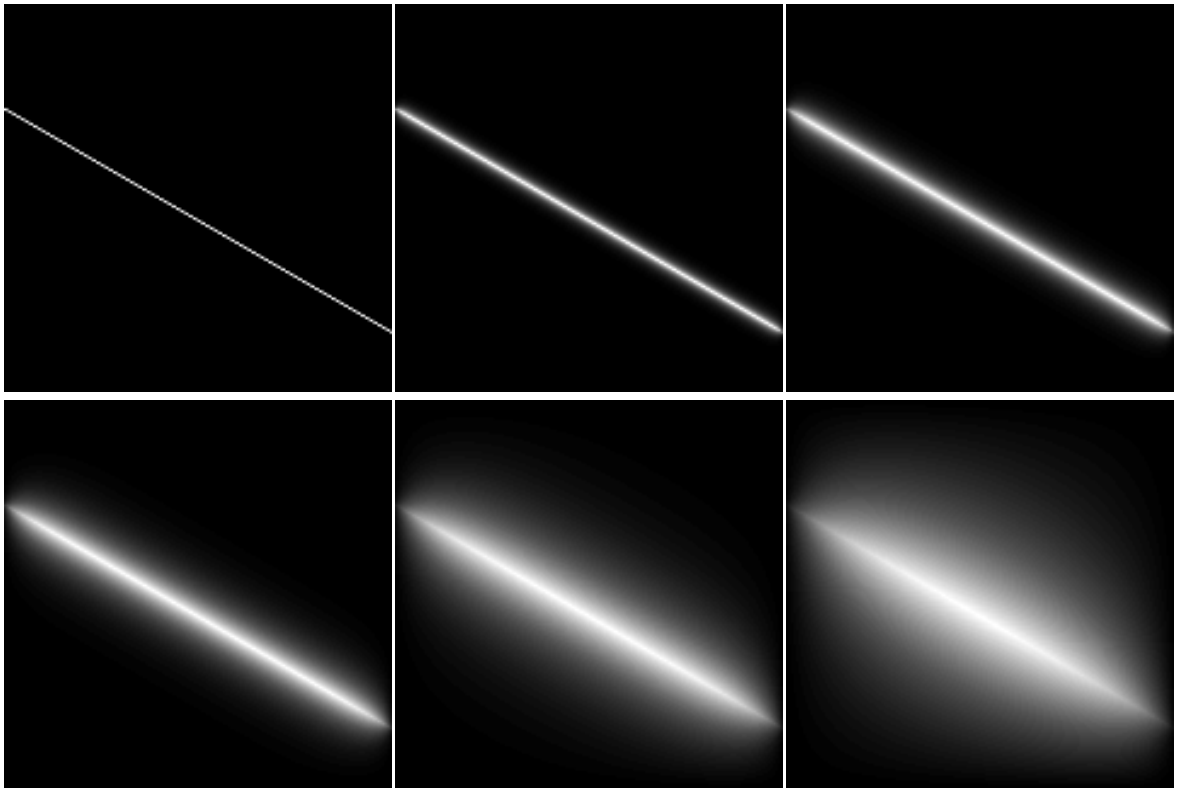
$$C^{-1} = -I_n \otimes D - D \otimes I_n + \frac{1}{\lambda^2} I_N,$$

where $D \in \mathbb{R}^{n \times n}$ is the three-point finite difference approximation of the one-dimensional Laplacian with Dirichlet boundary conditions,

$$D = \frac{1}{n^2} \begin{bmatrix} -2 & 1 & & \\ 1 & -2 & \ddots & \\ & \ddots & & 1 \\ & & 1 & -2 \end{bmatrix},$$

and $\lambda > 0$ is the correlation length.

Basis functions

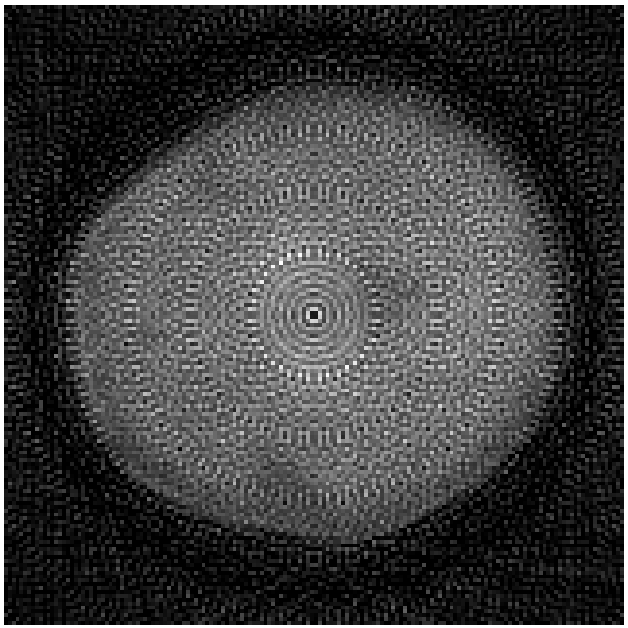


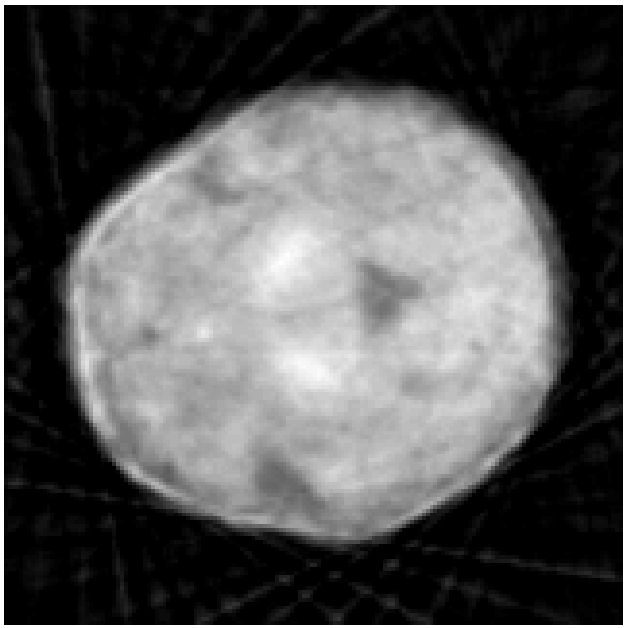
Basis vector with no priorconditioning (upper left) and with priorconditioning. Correlation length 2, 4, 8, 16 and 32 pixels.

Observations

- Every image whose support is on pixels not touched by a ray is in the null space of A
- $\Rightarrow$ plain CGLS iterates will be zero at those pixels
- Preconditioning makes the rays fuzzy, illuminating the dark pixels
- Reconstruction will be slightly blurred, but has fewer geometric artifacts
- Number of iterations needed will increase.

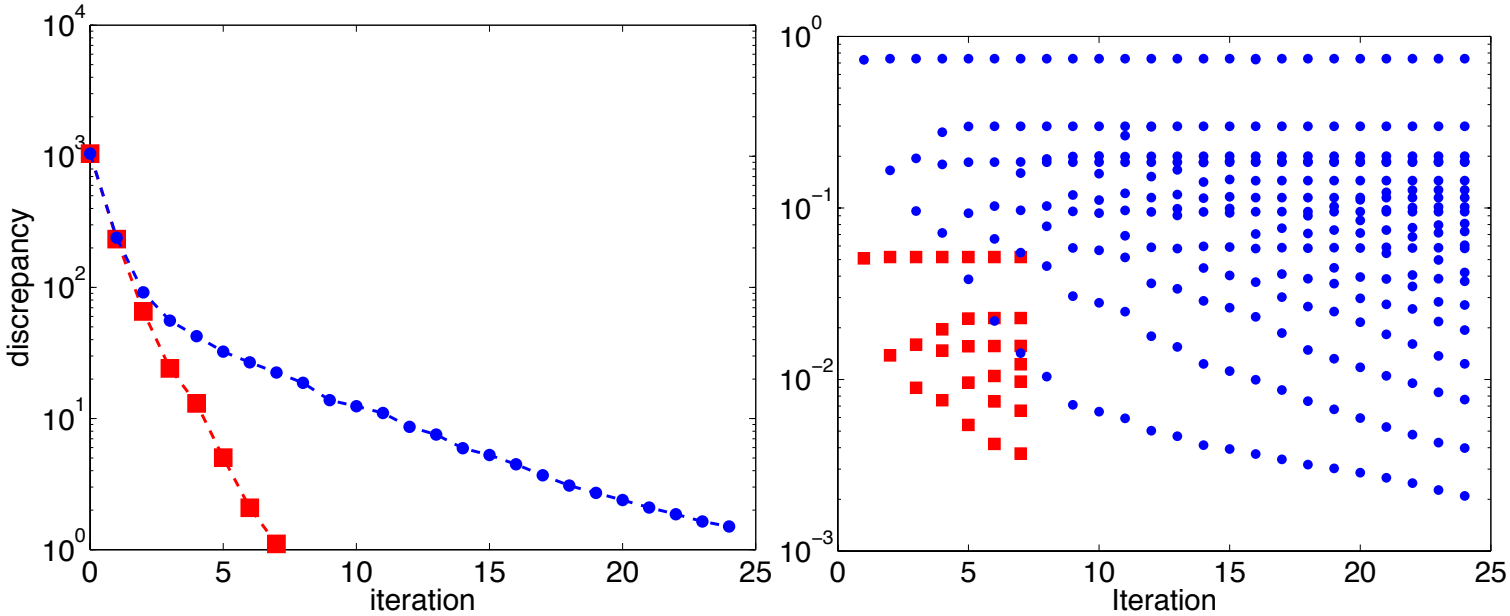
Computed solutions





Reconstructions with plain CGLS (left) and priorconditioned CGLS (right).

Converge history and spectral approximation



Convergence and spectral approximation.