Square regularization matrices for large linear discrete ill-posed problems

Marco Donatelli

Dipartimento di Fisica e Matematica
Università dell'Insubria - Como

Collaborator: *L. Reichel*

"Algebra Lineare Numerica nei Problemi Inversi"
Como 3–4 March, 2011

Main Issues

1. Tikhonov regularization
2. Range restricted Arnoldi and GMRES
3. Square regularization matrices
4. Boundary conditions
5. Square regularization matrices by boundary conditions
6. Numerical results

Large discrete ill-posed problems

$$\mathbf{b} = A\mathbf{x} + \mathbf{e}$$

- $A \in \mathbb{R}^{n \times n}$ large and severely ill-conditioned
- $\mathbf{b} \in \mathbb{R}^n$ known, measured data
- $\mathbf{e} \in \mathbb{R}^n$ noise,

Goal: compute approximation of the noise free solution $\mathbf{x}$

Tikhonov regularization

$$\min_{\mathbf{x}\in\mathbb{R}^n}\{\|A\mathbf{x}-\mathbf{b}\|^2+\mu\|L\mathbf{x}\|^2\} \tag{1}$$

with solution

$$\mathbf{x}_\mu=(A^TA+\mu L^TL)^{-1}A^T\mathbf{b},$$

where we assume that

$$\mathcal{N}(A)\cap\mathcal{N}(L)=\{\mathbf{0}\}.$$

- $\mu>0$ regularization parameter
- L regularization matrix

Regularization matrices

Common choices of regularization matrices $L \in \mathbb{R}^{k \times n}$, $1 \leq k \leq n$, are the identity matrix and scaled finite difference matrices, such as

$$L = \frac{1}{4} \begin{bmatrix} -1 & 2 & -1 & & & & 0 \\ & -1 & 2 & -1 & & & \\ & & \ddots & \ddots & \ddots & & \\ 0 & & & & -1 & 2 & -1 \end{bmatrix} \in \mathbb{R}^{(n-2) \times n}, \qquad (2)$$

$$\mathcal{N}(L) = \mathrm{span}\{[1, 1, \ldots, 1]^T, [1, 2, \ldots, n]^T\}.$$

Null space of regularization matrices is important!

Vectors in the $\mathcal{N}(L)$ are not damped in Tikhonov regularization problem (1).

Standard form

Tikhonov regularization problems (1) are said to be

- $L = I \rightarrow$ in standard form
- $L \neq I \rightarrow$ in general form

Large-scale Tikhonov regularization problems in standard form

First reducing them to a small problem by a few steps of Lanczos bidiagonalization, then determining an approximate solution $\mathbf{y}_\mu$ of the small problem, and finally computing an approximation $\mathbf{x}_\mu$ of (1) (see O'Leary and Simmons '81, Björck '88, Golub and von Matt '97 for several implementations).

Reduction in standard form

The A-weighted pseudoinverse of L, Eldén '82, is

$$L_A^\dagger = \left(I - (A(I - L^\dagger L))^\dagger A\right) L^\dagger \in \mathbb{R}^{n \times k}.$$

The Tikhonov regularization problem (1) is equivalent to

$$\min_{\bar{\mathbf{x}} \in \mathbb{R}^k} \{\|AL_A^\dagger \bar{\mathbf{x}} - \bar{\mathbf{b}}\|^2 + \mu \|\bar{\mathbf{x}}\|^2\}. \quad (3)$$

Matrix-vector products with the matrices $L_A^\dagger$ and $AL_A^\dagger$ can be evaluated quite inexpensively provided that

1. $\mathcal{N}(L)$ is of small dimension and has an explicitly known basis,
2. matrix-vector products with $L^\dagger$ are inexpensive to compute,
3. the restriction of A to $\mathcal{N}(L)$ is not very ill-conditioned.

Range restricted Arnoldi process

When A is square, reduction to a small problem also can be achieved by few steps of a (range restricted) Arnoldi-type process instead of Lanczos bidiagonalization:

- usually this reduction requires fewer matrix-vector product evaluations than Lanczos bidiagonalization,
- for some problems A^T may not be available.

L has to be square!

Range restricted GMRES (RR-GMRES)

The solution can be approximated by truncated iteration of RR-GMRES where the ℓth iterated, $\mathbf{x}_\ell$, satisfies

$$\mathbf{x}_\ell = \underset{\mathbf{x} \in \mathcal{K}_\ell(A, A\mathbf{b})}{\operatorname{argmin}} \|A\mathbf{x} - \mathbf{b}\|,$$

with $\mathbf{x}_0 = \mathbf{0}$ and

$$\mathcal{K}_\ell(A, A\mathbf{b}) = \operatorname{span}\{A\mathbf{b}, A^2\mathbf{b}, \ldots, A^\ell\mathbf{b}\}.$$

Preconditioning

- The A-weighted pseudoinverse $L_A^\dagger$ is a popular preconditioner for discrete ill-posed problems (Hanke '92).
- If L is inventible than $L_A^\dagger = L^{-1}$ and it can be used as right preconditioner for RR-GMRES (Calvetti et al. '04).

Our Goal

L square such that:

1. matrix-vector products with $L^\dagger$ are inexpensive to compute,
2. the main components of the signal have to belong to $\mathcal{N}(L)$ or, if we look for an invertible L, to the subspace generated from the eigenvectors associated to small eigenvalues.

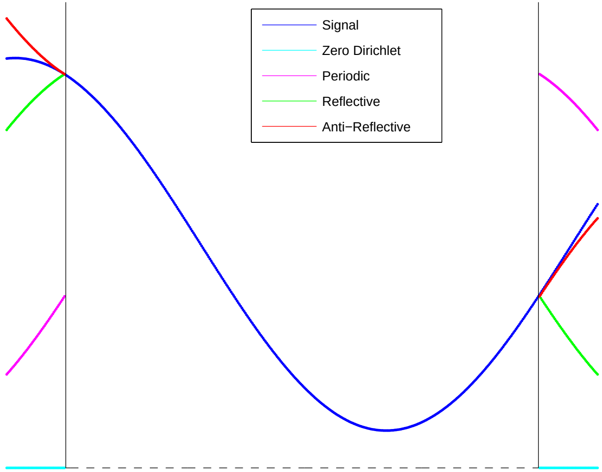
Square regularization matrices in literature

1. Appending rows to the differential operator, Calvetti et al. '04

$$\begin{bmatrix} 1 & -1 & 0 & \dots & 0 \\ & 1 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & & 1 & -1 \\ 0 & \dots & 0 & 0 & 10^{-8} \end{bmatrix}, \quad \begin{bmatrix} 1 & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ 0 & \dots & 0 & -1 & 2 \end{bmatrix}$$

2. Orthogonal projectors: $L = I - WW^T$, Morigi et al. '07.
3. $L = (\hat{C}_2^\dagger D_\delta^{-1})^\dagger$, where $\hat{C}_2$ is the Laplacian circulant matrix with the two smallest eigenvalues changed in zero and $D_\delta = \mathrm{diag}([\delta, 1, \dots, 1, \delta]^T)$, $\delta = 10^{-8}$, Reichel and Ye '09.

Boundary conditions



Boundary condition for Laplacian

G. Strang '99:

"The discrete case has a new level of variety and complexity, often appearing in the boundary conditions".

- The boundary conditions affect only the first and the last row.
- Consider a vector $\mathbf{u} \in \mathbb{R}^n$, $\mathbf{u} = (u_0, \ldots, u_{n-1})^T$, corresponding to the approximation of the solution u.

Dirichlet boundary conditions

$$u_{-1} = 0 \qquad \text{and} \qquad u_n = 0,$$

$$\Downarrow$$

$$L_D = \begin{bmatrix} 2 & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ 0 & \dots & 0 & -1 & 2 \end{bmatrix},$$

$$\mathcal{N}(L_D) = \{\mathbf{0}\}.$$

Neumann boundary conditions

Midpoint symmetry $u_{-1} = u_0$ and $u_n = u_{n-1}$,

$$\Downarrow$$

$$L_R = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ 0 & \dots & 0 & -1 & 1 \end{bmatrix}.$$

$$\mathcal{N}(L_R) = \mathrm{span}\{[1, 1, \dots, 1]^T\}$$

Mixed boundary conditions $\Rightarrow \mathcal{N}(L) = \{\mathbf{0}\}$

Antireflective boundary conditions

Serra Capizzano '03

$$u_{-1} - u_0 = u_0 - u_1 \qquad \text{and} \qquad u_n - u_{n-1} = u_{n-1} - u_{n-2},$$

$$\Downarrow$$

$$L_A = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ 0 & \dots & 0 & 0 & 0 \end{bmatrix}. \qquad (4)$$

$$\mathcal{N}(L_A) = \operatorname{span}\{[1, 1, \dots, 1]^T, [1, 2, \dots, n]^T\}.$$

AR Spectral decomposition

Let h be a cosine real-valued polynomial of degree at most n-3. Then

$$A = AR_n(h) = T_n \mathrm{diag}(h(\hat{\mathbf{x}}))\, T_n^{-1},$$

where $\hat{\mathbf{x}} = [0,\, \tilde{\mathbf{x}}^T,\, 0]^T$, $\tilde{\mathbf{x}} = [\frac{j\pi}{n-1}]_{j=1}^{n-2}$ and

$$T_n = \left[1 - \frac{\mathbf{x}}{\pi}, \sin(\mathbf{x}),\, \ldots,\, \sin((n-2)\mathbf{x}),\, \frac{\mathbf{x}}{\pi}\right],$$

with $\mathbf{x} = [0,\, \tilde{\mathbf{x}}^T,\, \pi]^T$.

[Aricò et al. '11]

Algebraic interpretation

Eigenvalues and eigenvectors

Let $\mathbf{x}$ be a uniform sampling in $[0, \pi]$:

- $h(0) = 0$ with multiplicity 2 (Laplacian: $h(z) = 2 - 2\cos(z)$)
- $\mathcal{N}(L_A) = \mathrm{span}\{\pi - \mathbf{x}, \mathbf{x}\} \Rightarrow$ preserve linear functions.
- eigenvector $\sin(k\mathbf{x})$ with eigenvalue $\sin\left(\frac{k\pi}{n-1}\right)$, $k = 1, \ldots, n-2$.

G. Strang '99:

> *"We hope that the eigenvector approach will suggest more new transforms, and that one of them will be fast and visually attractive."*.

High order BC with cosine basis

Let h be a cosine real-valued polynomial of degree at most n-3. Then

$$A_C(h) = C_n \mathrm{diag}(h(\hat{\mathbf{x}})) C_n^{-1},$$

where $\hat{\mathbf{x}} = [0,\, \tilde{\mathbf{x}}^T,\, 0]^T$, $\tilde{\mathbf{x}} = \left[\frac{(j-1)\pi}{n-2}\right]_{j=1}^{n-2}$ and

$$C_n = [\mathbf{h_1},\, \cos(0\mathbf{x}),\, \ldots,\, \cos((n-3)\mathbf{x}),\, \mathbf{h_2}]\,,$$

with $\mathbf{x} = \left[\frac{(2i-1)\pi}{2n-4}\right]_{j=0}^{n-1}$.

[D., '10]

Laplacian with high order BC

$$L_H = C_n \mathrm{diag}(2 - 2\cos(\hat{\mathbf{x}}))C_n^{-1},$$

there is not a geometrical interpretation!

Eigenvalues and eigenvectors

- $h(0) = 0$ with multiplicity 3
- $\mathcal{N}(L_H) = \mathrm{span}\{\mathbf{1}, \mathbf{h_1}, \mathbf{h_2}\}$
- eigenvector $\cos(k\mathbf{x})$ with eigenvalue $\cos\left(\frac{k\pi}{n-2}\right)$, $k = 1, \dots, n-3$, $\mathbf{x}$ is a uniform sampling in $[0, \pi]$.

Invertible regularization matrices

A proposal in Calvetti et al. '04:

$$F_{FR} = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 \\ & 1 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & & 1 & -1 \\ 0 & \dots & 0 & 0 & \alpha \end{bmatrix},$$

- For $\alpha = 0$, F_{FR} corresponds to impose Neumann boundary conditions.
- It is equivalent to replace the zero eigenvalue with $\alpha > 0$.

Invertible antireflective and high order BC

- Antireflective boundary conditions:

$$L_A(\alpha) = \begin{bmatrix} \alpha & 0 & 0 & \dots & 0 \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ 0 & \dots & 0 & 0 & \alpha \end{bmatrix},$$

- High order boundary conditions $\Rightarrow \lambda = \alpha$ for the eigenvectors $\{\mathbf{1}, \mathbf{h_1}, \mathbf{h_2}\}$.
- $0 < \alpha < 2 - 2\cos(\pi/n)$, the smallest eigenvalue greater than zero.

$L^\dagger$ for antireflective and high order BC

- The antireflective and the high order transforms are "close" to unitary transforms.
- Fix $X \in \{A, H\}$ and replace L_X with

$$\tilde{L}_X := (Y_n\, D^\dagger\, Y_n^{-1})^\dagger, \qquad D^\dagger_{i,j} = \begin{cases} 0 & \text{if } \; d_{i,j} = 0, \\ \frac{1}{d_{i,j}} & \text{otherwise,} \end{cases} \quad .$$

  where $Y_n = T_n$ for antireflective and $Y_n = C_n$ for high order boundary conditions.
- Only $\tilde{L}^\dagger_X = Y_n\, D^\dagger\, Y_n^{-1}$ is needed in the computations.

The Null space of $\tilde{L}_X^\dagger$

Roughly speaking $\mathcal{N}(\tilde{L}_X) \approx \mathcal{N}(L_X)$ for n large enough.

Theorem

$$\frac{\|\tilde{L}_X \mathbf{y}\|}{\|\mathbf{y}\|} = O\left(\frac{1}{\sqrt{n}}\right), \qquad \forall\, \mathbf{y} \in \mathcal{N}(L_X).$$

Numerical Experiments

- Troncated iteration preconditioned RR-GMRES
- The minimal error

$$\|\hat{\mathbf{x}} - \mathbf{x}_{k^*}\| = \min_{k \geq 0} \|\hat{\mathbf{x}} - \mathbf{x}_k\|$$

  and the index k^*.
- Examples from Hansen's Regularization Tools.
- Comparisons only with the best choice in literature.

Example 1 (deriv2)

Fredholm integral equation of the first kind

$$\int_0^1 k(s,t)x(t)\mathrm{d}t = g(s), \qquad 0 \le s \le 1,$$

where

$$k(s,t) = \begin{cases} s(t-1), & s < t \\ t(s-1), & s \ge t. \end{cases}$$

$$g(s) = \exp(s) + (1-\mathrm{e})s - 1$$

the solution is

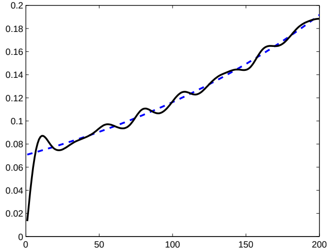
$$x(t) = \exp(t).$$

Discretization

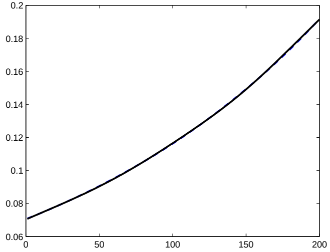
- Discretize the integral equation by a Galerkin method with orthonormal box functions
- symmetric indefinite matrix $A \in \mathbb{R}^{200\times 200}$
- we add white noise with level $\delta = 1 \cdot 10^{-3}$

Preconditioner	$\|\hat{\mathbf{x}} - \mathbf{x}_{k^*}\|$	k^*
L_{DR}	9.7e-2	34
$L_A(\alpha)$	1.6e-3	13
$L_H(\alpha)$	6.1e-4	17

Restorations



L_{DR}



$L_A(\alpha)$

Example 2

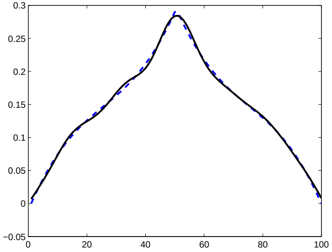
$$g(s) = \begin{cases} (4s^3 - 3s)/24, & s < \frac{1}{2} \\ (-4s^3 + 12s^2 - 9s + 1) & s \geq \frac{1}{2} \end{cases},$$

and the solution is

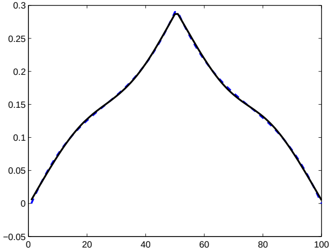
$$x(t) = \begin{cases} t, & t < \frac{1}{2} \\ 1 - t & t \geq \frac{1}{2} \end{cases}.$$

We define a new solution $\hat{\mathbf{x}} = 10\tilde{\mathbf{x}} - \cos(\pi(2\mathbf{t} - 1)) - c$, where $\mathbf{t}$ is a uniform grid in $[0, 1]$ and c is a constant such that $\hat{\mathbf{x}}$ is zero at the boundary.

Restorations

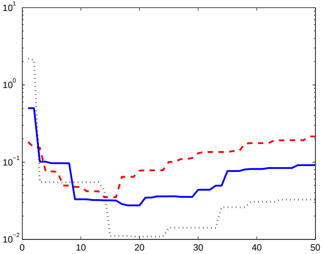


L_D



$L_H(\alpha)$

Error norm



Precond.	$\|\hat{\mathbf{x}} - \mathbf{x}_{k^*}\|$	k^*
L_D	3.5e-2	13
$L_A(\alpha)$	2.7e-2	17
$L_H(\alpha)$	1.0e-2	18

$\|\hat{\mathbf{x}} - \mathbf{x}_k\|$ varying k. The dashed curve is L_D, the solid curve is $L_A(\alpha)$, and the dotted curve is $L_H(\alpha)$.

Example 3 (phillips)

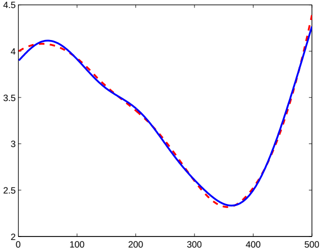
$$\int_{-6}^{6} x(s-t)x(t)\mathrm{d}t = g(s), \qquad -6 \le s \le 6,$$

with

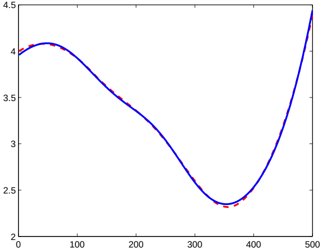
$$x(t) := \begin{cases} 1 + \cos(\frac{\pi}{3}t), & \text{if } |t| < 3, \\ 0, & \text{otherwise.} \end{cases}$$

- We define a new solution $\hat{\mathbf{x}} = \tilde{\mathbf{x}} + 3\cos(\pi \mathbf{s}) + \exp(\mathbf{s})^2$, where $\mathbf{s}$ is a uniform grid in $[0, 1]$
- noise-level $1 \cdot 10^{-2}$
- we compare $(\hat{C}_2^{\dagger} D_{\delta}^{-1})^{\dagger}$ for $\delta = 1 \cdot 10^{-8}$, $\hat{L}_{A,2}$, and $\tilde{L}_H$ where $\mathbf{h_1} = \mathbf{s}$ and $\mathbf{h_2} = \exp(\mathbf{s})$.

Restorations



$(\hat{C}_2^\dagger D_\delta^{-1})^\dagger$



$\tilde{L}_H$

Preconditioner	$\|\hat{\mathbf{x}} - \mathbf{x}_{k^*}\|$	k^*
$(\hat{C}_2^\dagger D_\delta^{-1})^\dagger$	9.3e-3	4
$\hat{L}_{A,2}$	1.0e-2	4
$\tilde{L}_H$	4..8e-3	4

Conclusions

- The use of boundary conditions is a simple and general framework to combine fast trigonometric transforms with the selection of a specific subspace that we want to preserve (the signal space).
- If no prior information about the solution are available:
  1. compute a first restoration with antireflective boundary conditions,
  2. if this is not satisfactory, an approximation of the signal space can be deduced from the shape of the computed solution and such approximation can be used to define the eigenvectors $\mathbf{h_1}$ and $\mathbf{h_2}$ for computing a second improved restoration by high order boundary conditions.