Subspace Regularization
for Large Linear Discrete Ill-Posed Problems

8th GAMM Workshop Applied and Numerical Linear Algebra
TU Hamburg-Harburg, September 11-12, 2008

Angelika Bunse-Gerstner
Zentrum für Technomathematik, Fachbereich 3 - Mathematik und Informatik
Universität Bremen
and
Valia Guerra
Instituto de Cibernetica, Matematica y Fisica
Havanna

Overview

Linear Discrete Ill-Posed Problems

$$x_{ex} = \arg\min_x \|A_{ex}x - b_{ex}\|_2 \qquad \hat{x} = \arg\min_x \|Ax - b\|_2$$

with $\|A_{ex} - A\|_2 / \|A_{ex}\|_2, \quad \|b_{ex} - b\|_2 / \|b_{ex}\|_2$ small

Reconstruct x_{ex} using the perturbed data A, b!

Problem: $\|x_{ex} - \hat{x}\|_2 / \|x_{ex}\|_2$ can be huge !

Standard techniques to solve $\|Ax - b\|_2 = min!$ cannot be applied !!

Characteristics of Linear Discrete Ill-Posed Problems

Typical properties of the problem $\|Ax - b_{ex}\|_2 = min!$ (inherited from source problem):
(Perturbations in A_{ex} neglected,i.e. $A = A_{ex}$)

- A dense, often very large
- $A = U\Sigma V^T$ with $\sigma_i \overset{"i\to\infty"}{\longrightarrow}$ 0, rapidly decreasing, no distinct gap,
  Often but not always: u_i, v_i '*smooth*' for small i, more '*oscillations*' in u_i, v_i with decreasing σ_i
- solution $x_{ex} = \sum_{i=1}^n \frac{u_i^T b_{ex}}{\sigma_i} v_i$
- Discrete Picard Condition: $\frac{u_i^T b_{ex}}{\sigma_i} \overset{"i\to\infty"}{\longrightarrow} 0,$

Therefore

$$x_{ex} = \sum_{\sigma_i \text{ not small}} \frac{u_i^T b_{ex}}{\sigma_i} v_i + \underbrace{\sum_{\sigma_i \text{ small}} \frac{u_i^T b_{ex}}{\sigma_i} v_i}_{\text{harmless}}$$

'*smooth*', entries not large

Characteristics of Linear Discrete Ill-Posed Problems

In practice: measurement errors. Instead of b_{ex} we have $b = b_{ex} + \eta$, and we can only solve $\|Ax - b\| = min$!

- A dense, often very large
- $A = U\Sigma V^T$ with $\sigma_i \overset{"i\to\infty"}{\longrightarrow}$ 0, rapidly decreasing, no distinct gap,
  Often but not always: u_i, v_i '*smooth*' for small i, more '*oscillations*' in u_i, v_i with decreasing σ_i
- solution $x_{ex} = \sum_{i=1}^{n} \frac{u_i^T b_{ex}}{\sigma_i} v_i$
- Discrete Picard Condition: $\frac{u_i^T b_{ex}}{\sigma_i} \overset{"i\to\infty"}{\longrightarrow} 0$,

Therefore

$$x_{ex} = \sum_{\sigma_i \text{ not small}} \frac{u_i^T b_{ex}}{\sigma_i} v_i + \underbrace{\sum_{\sigma_i \text{ small}} \frac{u_i^T b_{ex}}{\sigma_i} v_i}_{\text{harmless}}$$

'*smooth*', entries not large

Characteristics of Linear Discrete Ill-Posed Problems

In practice: measurement errors. Instead of b_{ex} we have $b = b_{ex} + \eta$, and we can only solve $\|Ax - b\| = min$!

- A dense, often very large
- $A = U\Sigma V^T$ with $\sigma_i \overset{"i\rightarrow\infty"}{\longrightarrow}$ 0, rapidly decreasing, no distinct gap,
  Often but not always: u_i, v_i '*smooth*' for small i, more '*oscillations*' in u_i, v_i with decreasing σ_i
- solution $\hat{x} = \sum_{i=1}^n \frac{u_i^T b}{\sigma_i} v_i = \sum_{i=1}^n \frac{u_i^T b_{ex}}{\sigma_i} v_i + \sum_{i=1}^n \frac{u_i^T \eta}{\sigma_i} v_i$
- Discrete Picard Condition: $\frac{u_i^T b_{ex}}{\sigma_i} \overset{"i\rightarrow\infty"}{\longrightarrow} 0$,

Therefore

$$x_{ex} = \sum_{\sigma_i \text{ not small}} \frac{u_i^T b_{ex}}{\sigma_i} v_i + \underbrace{\sum_{\sigma_i \text{ small}} \frac{u_i^T b_{ex}}{\sigma_i} v_i}_{\text{harmless}}$$

'*smooth*', entries not large

Characteristics of Linear Discrete Ill-Posed Problems

In practice: measurement errors. Instead of b_{ex} we have $b = b_{ex} + \eta$, and we can only solve $\|Ax - b\| = min$!

- A dense, often very large
- $A = U\Sigma V^T$ with $\sigma_i \overset{"i\to\infty"}{\longrightarrow}$ 0, rapidly decreasing, no distinct gap,
  Often but not always: u_i, v_i '*smooth*' for small i, more '*oscillations*' in u_i, v_i with decreasing σ_i
- solution $\hat{x} = \sum_{i=1}^n \frac{u_i^T b}{\sigma_i} v_i = \sum_{i=1}^n \frac{u_i^T b_{ex}}{\sigma_i} v_i + \sum_{i=1}^n \frac{u_i^T \eta}{\sigma_i} v_i$
- No Picard Condition: $\frac{u_i^T b}{\sigma_i}$ growing with decreasing σ_i

Therefore

$$x_{ex} = \sum_{\sigma_i \text{ not small}} \frac{u_i^T b_{ex}}{\sigma_i} v_i + \underbrace{\sum_{\sigma_i \text{ small}} \frac{u_i^T b_{ex}}{\sigma_i} v_i}_{\text{harmless}}$$

'*smooth*', entries not large

Characteristics of Linear Discrete Ill-Posed Problems

In practice: measurement errors. Instead of b_{ex} we have $b = b_{ex} + \eta$, and we can only solve $\|Ax - b\| = min$!

- A dense, often very large
- $A = U\Sigma V^T$ with $\sigma_i \overset{"i \to \infty"}{\longrightarrow} 0$, rapidly decreasing, no distinct gap,
  Often but not always: u_i, v_i '*smooth*' for small i, more '*oscillations*' in u_i, v_i with decreasing σ_i
- solution $\hat{x} = \sum_{i=1}^{n} \frac{u_i^T b}{\sigma_i} v_i = \sum_{i=1}^{n} \frac{u_i^T b_{ex}}{\sigma_i} v_i + \sum_{i=1}^{n} \frac{u_i^T \eta}{\sigma_i} v_i$
- No Picard Condition: $\frac{u_i^T b}{\sigma_i}$ growing with decreasing σ_i

Therefore

$$\hat{x} = \sum_{\sigma_i \text{ not small}} \frac{u_i^T b}{\sigma_i} v_i + \underbrace{\sum_{\sigma_i \text{ small}} \frac{u_i^T b}{\sigma_i} v_i}_{\text{extremely harmful}}$$

'*highly oscillating*', entries extremely large, dominated by perturbations

Regularization

Regularization methods use knowledge about the exact solution to find a good approximation.
Example: Tychonov regularization.

Instead of $\|Ax - b\|_2 = min!$ solve

$$\boxed{\|Ax - b\|_2^2 + \lambda^2 \|Lx\|_2^2 = min!, \quad \lambda \text{ small}}$$

equivalently

$$\boxed{\left\| \begin{bmatrix} A \\ \lambda L \end{bmatrix} x - \begin{bmatrix} b \\ 0 \end{bmatrix} \right\|_2 = \text{min!}}$$

e. g. $L = I$ or L discretized differential operator

Need a good choice for the regularization paramter λ (and for L)

Numerical Methods for the Large Dimensional Case

Iterative method needed (A is dense!), e.g. CG or LSQR

Advantage: Krylov methods have regularization properties
(regularization parameter: iteration index)

Problems:
Good regularization parameter not really known
Convergence maybe slow

Additional regularization in general still needed.

Work Done in This Field

- Ake Björck, Lars Eldén, Per Christian Hansen, Tommy Elfving
- Martin Hanke, Robert Plemmons
- Misha E. Kilmer, Dianne P., Oleary, James Nagy
- Daniela Calvetti, Lothar Reichel, Brian Lewis
- Gene H. Golub, Urs von Matt
- Jerry Eriksson, Marten Gullikson, Per-Åke Wedin
- Uri Asher, Eldad Haber, Douglas Oldenburg
- Marielba Rojas, Trond Steihaug
- Tony Chan, Stanley Osher, Curtis R. Vogel
- Matlab Software (Toolboxen):
  - Restore Tools (Nagy, 2002)
  - MOORe Tools (Jacobsen,2005)

Preconditioning

Conventional preconditioning is not an option!

Only very few approaches exist for the general case.
[e.g. Calvetti und Reichel 2002]: Smoothing of the solution

Ideal: Process only the parts in the "right" (smooth) subspace V.
Leave the remaining parts unchanged.

- Recall x_k essentially determined by the dominant SVD-parts
- The leading singular vectors $v_1, v_2, \ldots$ are often "smooth", i.e. V "is smooth".
- for special structure (Toeplitz etc. image processing): Fourier basis vectors
- general A: use Fourier- or Wavelet basis vectors or approximations of the singular vectors

[[Hanke and Vogel 1999, Hansen and Jacobsen 2000, Hansen,Jacobsen and Saunders 2003 (Two Level Preconditioning PreLSQR)]

What about non-smooth solutions?

Modified Two Level Preconditioning

[ABG, Guerra, Madrid 2005]
Split the problem into a harmless and a critical portion.

Let $V \in \mathbb{R}^{n \times k}$, k not too large, and $W \in \mathbb{R}^{n \times (n-k)}$ with $V \perp W$, orthonormal columns

Let $AV = Q \begin{bmatrix} R \\ 0 \end{bmatrix} = [YZ] \begin{bmatrix} R \\ 0 \end{bmatrix} = YR \qquad QR-$ decomposition ("cheap")

Then $Z^T AV = 0$ and $R = Y^T AV$ and with $x = V\xi + W\eta$ we get

$$\begin{aligned} \| Ax - b \|_2 &= \| \begin{bmatrix} Y^T \\ Z^T \end{bmatrix} A \begin{bmatrix} V & W \end{bmatrix} \begin{bmatrix} \xi \\ \eta \end{bmatrix} - \begin{bmatrix} Y^T \\ Z^T \end{bmatrix} b \|_2 \\ &= \| \begin{bmatrix} R & Y^T AW \\ 0 & Z^T AW \end{bmatrix} \begin{bmatrix} \xi \\ \eta \end{bmatrix} - \begin{bmatrix} Y^T \\ Z^T \end{bmatrix} b \|_2 \end{aligned}$$

Choose V such that $\mathrm{cond}(R)$ is moderate.

Solve $\min_\eta \| Z^T AW\eta - Z^T b \|_2$ with regularization.

Modified Two Level Preconditioning ctd.

$$\| Ax - b \|_2 = \| \begin{bmatrix} R & Y^T AW \\ 0 & Z^T AW \end{bmatrix} \begin{bmatrix} \xi \\ \eta \end{bmatrix} - \begin{bmatrix} Y^T \\ Z^T \end{bmatrix} b \|_2$$

- Solve $\min_\eta \| Z^T AW\eta - Z^T b \|_2$ with regularization
- Solve $R\xi = Y^T(b - AW\eta)$
- Compute $x = V\xi + W\eta$

Remark

- W is not needed for the computation
- $\min_\eta \| Z^T AW\eta - Z^T b \|_2$ is still large, ill-posed, but $\mathrm{cond}(Z^T AW) \leq \mathrm{cond}(A)$
  Observation: In Tychonov regularization, regularization parameter easier to detect.
- principle: first projection, then regularization
  [O'Leary and Simmons 1991, Hanke and Hansen 1993, Kilmer and O'Leary 2000 for unpreconditoned Krylov subspace methods]

Subspace Regularization

Theorem: The solution x_λ of the minimization problem

$$\min_x(\|Ax - b\|_2^2 + \lambda\|(W^TW)^{-1}W^Tx\|_2^2)$$

satisfies $x_\lambda = V\xi + W\eta_\lambda$ where η_λ is the solution of the regularized problem

$$\min_w(\|Z^TAW\eta - Z^Tb\|_2^2 + \lambda\|\eta\|_2^2)$$

and ξ is the solution of the triangular system

$$R\xi = Y^T(b - AW\eta_\lambda)$$

Strategies to choose V

Make use of information about the solution if available.

Proposition: $\mathrm{cond}(R)$ is monotonically increasing with k, the number of columns in V.

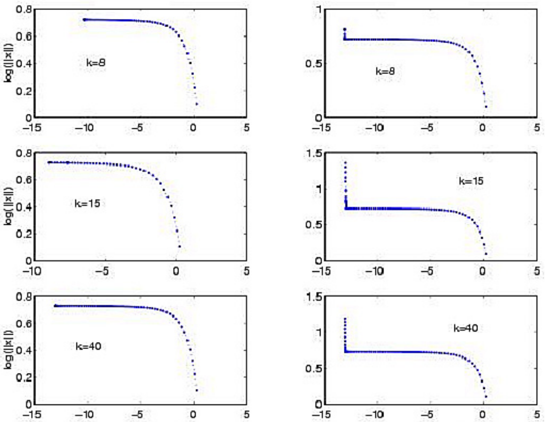
Allows monitoring of the condition number in the computational process !!

Crucial for the good sequential choice of V's columns.

- The first vectors of a wavelet basis etc. (for smooth solutions)
- Approximations of the dominant right singular vectors, e.g calculated by a restarted block-Lanczos method (here: irbleigs)

L-curves with and without Preconditioner

For iteration step k: $L^k := \{\log \|x_\lambda^k\|, \log \|Ax_\lambda^k - b\|\}$

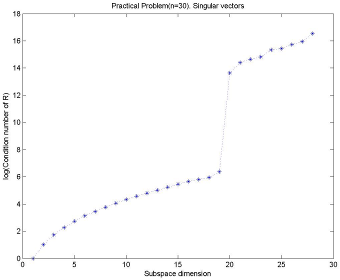


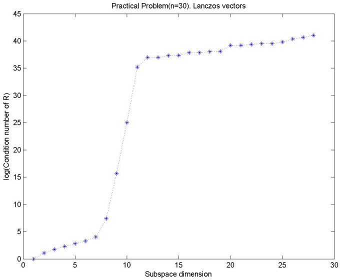
Even for large k no corner in the L-curve in CG

Subspace dimension versus Condition number of R

Three different bases: Lanczos vectors, Wavelet and Singular vectors.

Logarithmic scale for the condition number

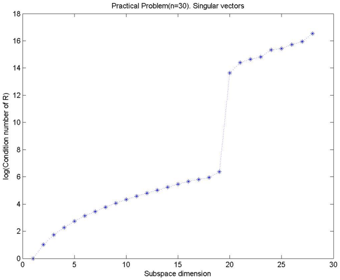


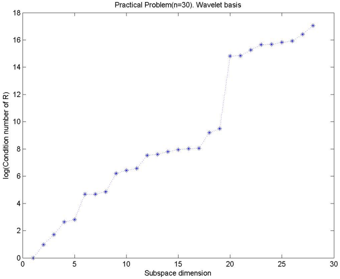


Subspace dimension versus Condition number of R

Three different bases: Lanczos vectors, Wavelet and Singular vectors.

Logarithmic scale for the condition number





Non-Smooth Solution

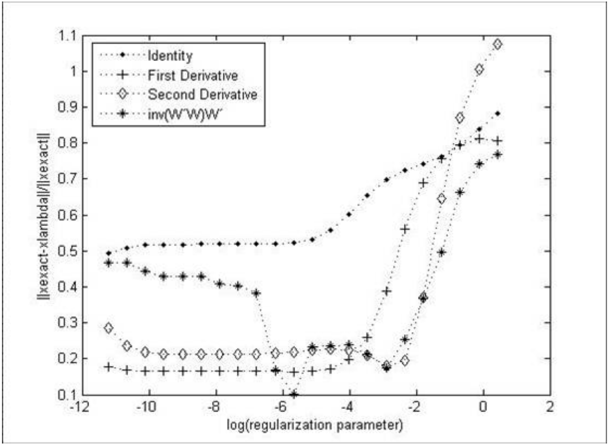
Linear system stemming from the discretization of a Fredholm integral equation of first-kind associated with the Fraunhofer difraction equation.

Level of noise: 1.e-04

Here: Relative errors between the non-smooth exact solution and the regularized solutions for different values of λ

$$\mathrm{Err}(\lambda) = \|x_{ex} - x_\lambda\|_2 / \|x_{ex}\|_2$$

using different subspaces


Identity
First Derivative
Second Derivative
inv(W'W)W'
||xexact-xlambda||/||xexact||
log(regularization parameter)
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
-12
-10
-8
-6
-4
-2
0
2

Conclusion

- Tychnov subspace regularization can be performed as a modified two level regularization.
- A "save" subspace can be computed in a monitored process.
- The reconstruction of non-smooth solutions for ill-posed problems is a difficult problem. Subspace regularization combined with a careful computation of the subspace may be helpful.