Multilevel methods for ill-posed problems

Malena Español

Misha Kilmer

Mathematics Department
Tufts University
Outline

- Ill-posed Problem
- Multilevel Approach
- Numerical Results
- Extension to 2D
- Total Least Squares Problem
- Conclusion and Future Work
Ill-posed Problems

A problem is ill-posed if

- its solution is not unique, or
- its solution does not depend continuously on the data

Example: Fredholm integral equation of first kind

$$\int K(s,t) f(t) dt = g(s)$$
Linear Discrete Ill-Posed problem

Find $x$, given $A, b$ and the model

$$Ax = b$$

where $b = b_{true} + e = Ax_{true} + e$.

Properties:

- Decaying singular values without gap
- $e$ is unknown (white) noise
- Singular vectors become more oscillatory
- Discrete Picard condition holds
Need for regularization

Let $A = U\Sigma V$ the SVD of $A$.

Then the least square solutions are:

$$x_{true} = A^\dagger b_{true} = \sum_{i=1}^{\text{rank}(A)} \frac{u_i^T b_{true}}{\sigma_i} \nu_i$$

$$x = A^\dagger (b_{true} + e) = \sum_{i=1}^{\text{rank}(A)} \frac{u_i^T (b_{true} + e)}{\sigma_i} \nu_i$$

$$= \sum_{i=1}^{\text{rank}(A)} \frac{u_i^T b_{true}}{\sigma_i} \nu_i + \sum_{i=1}^{\text{rank}(A)} \frac{u_i^T e}{\sigma_i} \nu_i = x_{true} + error$$
Example: Signal restoration

\[ A = U \quad V^T \quad x = \quad b = \]

\[ b + e = \]

\[ V = [v_1 v_2 \ldots v_n] = \]
The Least Squares solution

\[ x = V(\Sigma^{-1}U^Tb) = \frac{u_1^Tb}{\sigma_1}v_1 + \frac{u_2^Tb}{\sigma_2}v_2 + \ldots + \frac{u_{10}^Tb}{\sigma_{10}}v_{10} + \ldots + \frac{u_{20}^Tb}{\sigma_{20}}v_{20} + \ldots + \frac{u_{32}^Tb}{\sigma_{32}}v_{32} \]

\[ = \frac{u_1^Tb}{\sigma_1} + \frac{u_2^Tb}{\sigma_2} + \ldots + \frac{u_{10}^Tb}{\sigma_{10}} + \ldots + \frac{u_{20}^Tb}{\sigma_{20}} + \ldots + \frac{u_{32}^Tb}{\sigma_{32}} \]

\[ = -7.3 + 3.6 + \ldots + 0.23 + \ldots + 63.95 + \ldots + 9179500 \]
Regularization methods

Truncated SVD (TSVD) : \( x_{TSVD} = \sum_{i=1}^{k} \frac{u_i^T b}{\sigma_i} \nu_i \)

Tikhonov Regularization : \( x_{Tik} = \sum_{i=1}^{n} \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2} \frac{u_i^T b}{\sigma_i} \nu_i \)

\( x_{Tik} = \arg \min_x \left\{ \| Ax - b \|_2^2 + \lambda^2 \| Lx \|_2^2 \right\} \)

Iterative Methods : \( k \) iterations of Krylov methods, eg. LSQR, or CGLS.

MATLAB
Multilevel: Introduction

\[ A_h x_h = b_h \]

\[ A_{2h} x_{2h} = b_{2h} \]

\[ A_{4h} x_{4h} = b_{4h} \]
Downsampling ("going down")

\[ P_D v_{2h} = \frac{1}{4} \begin{bmatrix} 1 & 2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \end{bmatrix}_h = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}_{2h} \]

\[ v_{2h} = \frac{1}{4} (v_{2j-1}^h + 2v_{2j}^h + v_{2j+1}^h) \]
Interpolation ("going up")

\[ v_{2j}^h = v_j^{2h} \]
\[ v_{2j+1}^h = \frac{1}{2}(v_j^{2h} + v_{j+1}^{2h}) \]

\[
P_I v_{2h} = \frac{1}{2} \begin{bmatrix} 1 & 0 & 0 \\ 2 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \end{bmatrix}_h = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \end{bmatrix}_h
\]
Two-level algorithm

\[ A_h x_h = b_h \]
\[ x = P_I x_{2h} \]

\[ P_D A_h P_I x_{2h} = P_D b_h \]
Some previous work

- Tikhonov model: King (1992), Hanke-Vogel (1999)
- Cascadic multilevel: Reichel and Shyshkov (2008)
Two-Level Method

"Solve" $A^i x^i = b^i$

$r^i = b^i - A^i x^i$

$r^{i+1} = R^i r^i$

$A^{i+1} = R^i A^i P^i$

"Solve" $A^{i+1} x^{i+1} = r^{i+1}$

$x^i = x^i + P^i x^{i+1}$

$r^i = b^i - A^i x^i$

"Solve" $A^i x_c = r^i$

$x^i = x^i + x_c$
Multilevel Method

function $x^i = MGM(A^i, b^i)$

If coarsest grid

Solve $A^i x^i = b^i$

Else

Solve $A^i x^i = b^i$

$r^i = b^i - A^i x^i$

$r^{i+1} = R^i r^i, A^{i+1} = R^i A^i P^i$

$x^{i+1} = MGM(A^{i+1}, b^{i+1})$

$x^i = x^i + P^i x^{i+1}$

$r^i = b^i - A^i x^i$

Solve $A^i x_c = r^i$

$x^i = x^i + x_c$

End If

November 12, 2008

Temple University
Restriction and Prolongation Operators

Haar wavelet transform

\[ W^T = \frac{\sqrt{2}}{2} \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix} \]
Wavelet Domain

In wavelet domain $Ax = b$ becomes

$$\hat{A}\hat{x} = \hat{b}$$

where $\hat{A} = W^T A W$, $\hat{x} = W^T x$ and $\hat{b} = W^T b$.

\[
\begin{bmatrix}
\hat{A}_1 & \hat{A}_2 \\
\hat{A}_3 & \hat{A}_4
\end{bmatrix}
\begin{bmatrix}
\hat{x}_1 \\
\hat{x}_2
\end{bmatrix} =
\begin{bmatrix}
\hat{b}_1 \\
\hat{b}_2
\end{bmatrix}
\]
Coarse-scale equation

\[
\begin{bmatrix}
\hat{A}_1 & \hat{A}_2 \\
\hat{A}_3 & \hat{A}_4
\end{bmatrix}
\begin{bmatrix}
\hat{x}_1 \\
\hat{x}_2
\end{bmatrix}
= 
\begin{bmatrix}
\hat{b}_1 \\
\hat{b}_2
\end{bmatrix}
\]

\[
\hat{A}_1 \hat{x}_1 = \hat{b}_1 - \hat{A}_2 \hat{x}_2
\]

\[
\hat{A}_1 \hat{x}_1 = \hat{b}_1
\]

\[
x = 
\]

\[
\hat{x}_1 = 
\]
Coarse-scale equation

\[
\begin{align*}
\hat{A}_1 \hat{x}_1 &= \hat{b}_1 - \hat{A}_2 \hat{x}_2 = \hat{b}_1^{\text{true}} + \hat{e}_1 - \hat{A}_2 \hat{x}_2 \\
\hat{x}_1 &= \sum_{i=1}^{\text{rank}(\hat{A}_1)} \left[ \frac{\hat{u}_i^T \hat{b}_1^{\text{true}}}{\hat{\sigma}_i} + \frac{\hat{u}_i^T \hat{e}_1}{\hat{\sigma}_i} - \frac{\hat{u}_i^T \hat{A}_2 \hat{x}_2}{\hat{\sigma}_i} \right] \hat{\nu}_i
\end{align*}
\]

Then,

\[\hat{A}_1 \hat{x}_1 = \hat{b}_1.\]
To get a regularized solution of
\[ \hat{A}_1 \hat{x}_1 = \hat{b}_1 \]
we solve

\[ \hat{x}_1 = \text{LSQR}(\hat{A}_1, \hat{b}_1, 2 \text{ or } 3) \]

\[
\min_{\hat{x}_1} \left\{ \left\| \hat{A}_1 \hat{x}_1 - \hat{b}_1 \right\|_2^2 + \lambda^p \left\| L \hat{x}_1 \right\|_p^p \right\}
\]
Post-Smoothing

\[
\begin{bmatrix}
\hat{A}_2 \\
\hat{A}_4
\end{bmatrix} \hat{x}_2 = \begin{bmatrix}
\hat{b}_1 \\
\hat{b}_2
\end{bmatrix} - \begin{bmatrix}
\hat{A}_1 \\
\hat{A}_3
\end{bmatrix} \hat{x}_1
\]

\[
\begin{aligned}
\min_{\hat{x}_2} & \left\{ \left\| \begin{bmatrix}
\hat{A}_2 \\
\hat{A}_4
\end{bmatrix} \hat{x}_2 - \begin{bmatrix}
\hat{b}_1 \\
\hat{b}_2
\end{bmatrix} \right\|^2_2 + \lambda^p \left\| L(x_{\text{pre}} + W_1 \hat{x}_1 + W_2 \hat{x}_2) \right\|^p_p \right\}
\end{aligned}
\]
Multilevel Method

function $x^i = MGM(A^i, b^i)$

If coarsest grid

$$x^i = \text{direct solve or Newton } (A^i, b^i)$$

Else

If non finest grid

$$x^i = LSQR (A^i, b^i, 3)$$

End If

$$r^i = b^i - A^i x^i$$

$$b_{i+1} = W_1^T r_i; \quad A^{i+1} = W_1^T A^i W_1$$

$$\hat{x}_{1}^{i+1} = MGM(A^{i+1}, b^{i+1})$$

$$x_{new}^i = x^i + W_1 \hat{x}_{1}^{i+1}$$

$$r_{new}^i = b^i - A^i x_{new}^i$$

$$\hat{x}_{2}^{i+1} = \text{Newton } (WA^i W_2^T, W r_{new}^i, L, x_{new}^i)$$

$$x^i = x_{new}^i + W_2 \hat{x}_{2}^{i+1}$$

End If
Haar Wavelet Advantages

- Signal main features
- Matrix spectral characteristics
- Toeplitz structure (Fast Matrix-vector multiplication)
- Bandwidth reduction

\[
A = \begin{pmatrix}
    t_0 & t_1 & t_2 & \cdots & t_{-(m-1)} \\
    t_1 & t_0 & t_1 & \cdots & t_{-(m-2)} \\
    t_2 & t_1 & t_0 & \cdots & t_{-(m-3)} \\
    \vdots & \vdots & \vdots & \ddots & \vdots \\
    t_{m-1} & t_{m-2} & t_{m-3} & \cdots & t_0
\end{pmatrix}
\]
Numerical Example

\[ A : \text{Gaussian, Toeplitz, symmetric matrix} \]
\[ x_{true} : \text{edgy solution} \]
\[ b = Ax_{true} + e \text{ with } e \approx N(0,0.05\|b\|) \]
\[ L = \text{derivative operator, } p = 1.1 \]
Numerical Example

\[ x_{true} = \]

\[ b = \]

\[ x_{LSQR} = \]

\[ x_{MGM} = \]
The 2D case: Image restoration
Total Least-Squares Method

To solve the system $Ax \approx b$, we solve

$$\min_{E,r,x} \left\{ \|E\|_F^2 + \|r\|_2^2 \right\} \quad \text{s.t.} \quad (A + E)x = b + r$$

A Regularized TLS formulation is

$$\min_{E,r,x} \left\{ \|E\|_F^2 + \|r\|_2^2 + \lambda^p \|Lx\|_p^p \right\} \quad \text{s.t.} \quad (A + E)x = b + r$$

*Regularization by Truncated TLS* by Fierro, Golub, Hansen and O’Leary (1997)
Pre-Smoothing

\[(A + E)\hat{x} = \hat{b} + \hat{r}\]

\[
\begin{bmatrix}
\hat{A}_1 & \hat{A}_2 \\
\hat{A}_3 & \hat{A}_4 \\
\end{bmatrix}
+ \begin{bmatrix}
\hat{E}_1 & \hat{E}_2 \\
\hat{E}_3 & \hat{E}_4 \\
\end{bmatrix}
\begin{bmatrix}
\hat{x}_1 \\
\hat{x}_2 \\
\end{bmatrix}
= \begin{bmatrix}
\hat{b}_1 \\
\hat{b}_2 \\
\end{bmatrix}
+ \begin{bmatrix}
\hat{r}_1 \\
\hat{r}_2 \\
\end{bmatrix}
\]

\[(A_1 + E_1)\hat{x}_1 = \hat{b}_1 + \hat{r}_1 - (A_2 + E_2)\hat{x}_2\]

\[
\min_{\hat{x}_1, \hat{E}_1} \left\{\|\hat{E}_1\|_F^2 + \| (A_1 + E_1)\hat{x}_1 - \hat{b}_1 \|_2^2 + \lambda^p \|L\hat{x}_1\|_p^p \right\}
\]
Post-Smoothing

\[
\begin{align*}
\left( \begin{bmatrix} \hat{A}_2 \\ \hat{A}_4 \end{bmatrix} + \begin{bmatrix} \hat{E}_2 \\ \hat{E}_4 \end{bmatrix} \right) \hat{x}_2 + \hat{E}_3 \hat{x}_1 &= \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \end{bmatrix} - \left( \begin{bmatrix} \hat{A}_1 + \hat{E}_1 \\ \hat{A}_3 \end{bmatrix} \right) \hat{x}_1 + \begin{bmatrix} \hat{r}_1 \\ \hat{r}_2 \end{bmatrix} \\
\rho
\end{align*}
\]

\[
\min_{\hat{x}_2, \hat{E}_2, \hat{E}_3, \hat{E}_4} \left\{ \| \hat{E}_2 \|_F^2 + \| \hat{E}_4 \|_F^2 + \| \hat{E}_3 \|_F^2 + \left\| \left( \begin{bmatrix} \hat{A}_2 \\ \hat{A}_4 \end{bmatrix} + \begin{bmatrix} \hat{E}_2 \\ \hat{E}_4 \end{bmatrix} \right) \hat{x}_2 + \hat{E}_3 \hat{x}_1 - \rho \right\}^2_2 + \lambda^p \left\| LW^T \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} \right\|^p_p
\]
Two-Level Method

\[ x = V_1^T \hat{x}_1, \quad E = V_1^T \hat{E}_1 V_1 \]
\[ \Rightarrow r = b - (A + E)x \]

\[ x = V_1^T \hat{x}_1 + W_1^T \hat{x}_2 \]

\[ E = W^T \begin{bmatrix} \hat{E}_1 & \hat{E}_2 \\ \hat{E}_3 & \hat{E}_4 \end{bmatrix} W \]

\[
\begin{align*}
\min_{\tilde{x}_1, \tilde{x}_1} \left\{ \| \hat{E}_1 \|_F^2 + \| (\tilde{A}_1 + \hat{E}_1) \tilde{x}_1 - \hat{b}_1 \|_2^2 + \lambda \| \hat{L} \|_F^p \right\} \\
\min_{\hat{x}_2, \hat{x}_2, \hat{x}_1, \tilde{x}_1} \left\{ \| \hat{E}_2 \|_F^2 + \| \hat{E}_3 \|_F^2 + \left\| \begin{bmatrix} \hat{A}_1 \\ \hat{A}_2 \end{bmatrix} \right\|_F^p \hat{x}_2 + \hat{E}_3 \hat{x}_1 - \hat{r} \|_2^2 + \lambda \| LW^T \hat{x}_1 \|_F^p \right\}
\end{align*}
\]
Future work

- More Structures and Non Structured Matrices
- Parameter Selection
- Adaptive p-norm
- Extension to 2D, 3D