Math 4/896: Seminar in Mathematics
Topic: Inverse Theory

Instructor: Thomas Shores
Department of Mathematics

Lecture 19, March 21, 2006
AvH 10
Outline

1 Chapter 5: Tikhonov Regularization
   • 5.2: SVD Implementation of Tikhonov Regularization
   • 5.3: Resolution, Bias and Uncertainty in the Tikhonov Solution
To solve the Tikhonov regularized problem, first recall:

\[
\nabla \left( \| Gm - d \|_2^2 + \alpha^2 \| m \|_2^2 \right) = (G^T Gm - G^T d) + \alpha^2 m
\]

- Equate to zero and these are the normal equations for the system

\[
\begin{bmatrix}
G \\
\alpha l
\end{bmatrix} m = \begin{bmatrix}
d \\
0
\end{bmatrix}, \text{ or } (G^T G + \alpha^2 I) m = G^T d
\]

- To solve, calculate \((G^T G + \alpha^2 I)^{-1} G^T = U^T V \frac{\sigma_1}{\sigma_1 + \alpha^2}
\]

\[
\begin{bmatrix}
\frac{\sigma_1}{\sigma_1 + \alpha^2} & \cdots & \frac{\sigma_p}{\sigma_p + \alpha^2} \\
\sigma_1 & \cdots & \sigma_p \\
\sigma_1 + \alpha^2 & \cdots & \sigma_p + \alpha^2
\end{bmatrix}
\]
To solve the Tikhonov regularized problem, first recall:

\[ \nabla \left( \| Gm - d \|^2_2 + \alpha^2 \| m \|^2_2 \right) = (G^T Gm - G^T d) + \alpha^2 m \]

- Equate to zero and these are the normal equations for the system

\[ \begin{bmatrix} G \\ \alpha I \end{bmatrix} m = \begin{bmatrix} d \\ 0 \end{bmatrix}, \text{ or } (G^T G + \alpha^2 I) m = G^T d \]

- To solve, calculate \( (G^T G + \alpha^2 I)^{-1} G^T = \)

\[
\begin{bmatrix}
\frac{\sigma_1}{\sigma_1^2 + \alpha^2} \\
\sigma_2 \\
\vdots \\
\frac{\sigma_p}{\sigma_p^2 + \alpha^2}
\end{bmatrix}
\begin{bmatrix}
V \\
0 \\
\vdots
\end{bmatrix}
\]

\[ U^T \]
To solve the Tikhonov regularized problem, first recall:

\[ \nabla \left( \| Gm - d \|^2_2 + \alpha^2 \| m \|^2_2 \right) = (G^T Gm - G^T d) + \alpha^2 m \]

- Equate to zero and these are the normal equations for the system

\[ \begin{bmatrix} G & \alpha I \end{bmatrix} m = \begin{bmatrix} d \\ 0 \end{bmatrix}, \text{ or } (G^T G + \alpha^2 I) m = G^T d \]

- To solve, calculate

\[ (G^T G + \alpha^2 I)^{-1} G^T = \begin{bmatrix} \frac{\sigma_1}{\sigma_1^2 + \alpha^2} & \cdots & \frac{\sigma_p}{\sigma_p^2 + \alpha^2} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\sigma_p}{\sigma_p^2 + \alpha^2} \end{bmatrix} U^T \]
To solve the Tikhonov regularized problem, first recall:

\[ \nabla \left( \| Gm - d \|_2^2 + \alpha^2 \| m \|_2^2 \right) = (G^T Gm - G^T d) + \alpha^2 m \]

Equate to zero and these are the normal equations for the system

\[
\begin{bmatrix}
G \\
\alpha I
\end{bmatrix} m = \begin{bmatrix} d \\
0
\end{bmatrix}, \text{ or } (G^T G + \alpha^2 I) m = G^T d
\]

To solve, calculate \((G^T G + \alpha^2 I)^{-1} G^T = \)

\[
V \begin{bmatrix}
\frac{\sigma_1}{\sigma_1^2 + \alpha^2} \\
\vdots \\
\frac{\sigma_p}{\sigma_p^2 + \alpha^2} \\
\alpha I
\end{bmatrix} U^T
\]
From the previous equation we obtain that the Moore-Penrose inverse and solution to the regularized problem are given by

$$G^\dagger_\alpha = \sum_{j=1}^{p} \frac{\sigma_j}{\sigma_j^2 + \alpha^2} V_j U_j^T$$

and

$$m_\alpha = G^\dagger d = \sum_{j=1}^{p} \frac{\sigma_j^2}{\sigma_j^2 + \alpha^2} \left( U_j^T d \right) \sigma_j V_j$$

which specializes to the generalized inverse solution we have seen in the case that $G$ is full column rank and $\alpha = 0$. (Remember $d =Uh$ so that $h = U^Td.$)
The Filter Idea

About Filtering:

The idea is simply to “filter” the singular values of our problem so that (hopefully) only “good” ones are used.

- We replace the $\sigma_i$ by $f(\sigma_i)$. The function $f$ is called a filter.
- $f(\sigma) = 1$ simply uses the original singular values.
- $f(\sigma) = \frac{\sigma^2}{\sigma^2 + \alpha^2}$ is the Tikhonov filter we have just developed.
- $f(\sigma) = \max\{\text{sgn}(\sigma - \epsilon), 0\}$ is the TSVD filter with singular values smaller than $\epsilon$ truncated to zero.
The Filter Idea

About Filtering:

The idea is simply to “filter” the singular values of our problem so that (hopefully) only “good” ones are used.

- We replace the $\sigma_i$ by $f(\sigma_i)$. The function $f$ is called a filter.
  - $f(\sigma) = 1$ simply uses the original singular values.
  - $f(\sigma) = \frac{\sigma^2}{\sigma^2 + \alpha^2}$ is the Tikhonov filter we have just developed.
  - $f(\sigma) = \max\{\text{sgn}(\sigma - \epsilon), 0\}$ is the TSVD filter with singular values smaller than $\epsilon$ truncated to zero.
The Filter Idea

About Filtering:

The idea is simply to “filter” the singular values of our problem so that (hopefully) only “good” ones are used.

- We replace the $\sigma_i$ by $f(\sigma_i)$. The function $f$ is called a filter.
- $f(\sigma) = 1$ simply uses the original singular values.
- $f(\sigma) = \frac{\sigma^2}{\sigma^2 + \alpha^2}$ is the Tikhonov filter we have just developed.
- $f(\sigma) = \max\{\text{sgn}(\sigma - \epsilon), 0\}$ is the TSVD filter with singular values smaller than $\epsilon$ truncated to zero.
The Filter Idea

About Filtering:

The idea is simply to “filter” the singular values of our problem so that (hopefully) only “good” ones are used.

- We replace the $\sigma_i$ by $f(\sigma_i)$. The function $f$ is called a filter.
- $f(\sigma) = 1$ simply uses the original singular values.
- $f(\sigma) = \frac{\sigma^2}{\sigma^2 + \alpha^2}$ is the Tikhonov filter we have just developed.
- $f(\sigma) = \max\{\text{sgn} (\sigma - \epsilon), 0\}$ is the TSVD filter with singular values smaller than $\epsilon$ truncated to zero.
About Filtering:
The idea is simply to “filter” the singular values of our problem so that (hopefully) only “good” ones are used.

- We replace the \( \sigma_i \) by \( f(\sigma_i) \). The function \( f \) is called a **filter**.
- \( f(\sigma) = 1 \) simply uses the original singular values.
- \( f(\sigma) = \frac{\sigma^2}{\sigma^2 + \alpha^2} \) is the Tikhonov filter we have just developed.
- \( f(\sigma) = \max\{\text{sgn}(\sigma - \epsilon), 0\} \) is the TSVD filter with singular values smaller than \( \epsilon \) truncated to zero.
The L-curve

L-curves are one tool for choosing the regularization parameter $\alpha$:

- Make a plot of the curve $(\|m_\alpha\|_2, \|Gm_\alpha - d\|_2)$.
- Typically, this curve looks to be asymptotic to the axes.
- Choose the value of $\alpha$ closest to the corner.
- Caution: L-curves are NOT guaranteed to work as a regularization strategy.
- An alternative: (Morozov's discrepancy principle) Choose $\alpha$ so that the misfit $\|Gm_\alpha - d\|_2$ is the same size as the data noise $\|\delta d\|_2$. 
The L-curve

L-curves are one tool for choosing the regularization parameter $\alpha$:

- Make a plot of the curve $(\|m_\alpha\|_2, \|Gm_\alpha - d\|_2)$
- Typically, this curve looks to be asymptotic to the axes.
- Choose the value of $\alpha$ closest to the corner.
- Caution: L-curves are NOT guaranteed to work as a regularization strategy.
- An alternative: (Morozov’s discrepancy principle) Choose $\alpha$ so that the misfit $\|Gm_\alpha - d\|_2$ is the same size as the data noise $\|\delta d\|_2$.
The L-curve

L-curves are one tool for choosing the regularization parameter $\alpha$:

- Make a plot of the curve $(\|m_\alpha\|_2, \|Gm_\alpha - d\|_2)$
- Typically, this curve looks to be asymptotic to the axes.
- Choose the value of $\alpha$ closest to the corner.
- Caution: L-curves are NOT guaranteed to work as a regularization strategy.
- An alternative: (Morozov’s discrepancy principle) Choose $\alpha$ so that the misfit $\|Gm_\alpha - d\|_2$ is the same size as the data noise $\|\delta d\|_2$. 
L-curves are one tool for choosing the regularization parameter $\alpha$:

- Make a plot of the curve $(\|m_\alpha\|_2, \|Gm_\alpha - d\|_2)$
- Typically, this curve looks to be asymptotic to the axes.
- Choose the value of $\alpha$ closest to the corner.
- Caution: L-curves are NOT guaranteed to work as a regularization strategy.
- An alternative: (Morozov’s discrepancy principle) Choose $\alpha$ so that the misfit $\|Gm_\alpha - d\|_2$ is the same size as the data noise $\|\delta d\|_2$. 
The L-curve

L-curves are one tool for choosing the regularization parameter $\alpha$:

- Make a plot of the curve $(\|m_\alpha\|_2, \|Gm_\alpha - d\|_2)$
- Typically, this curve looks to be asymptotic to the axes.
- Choose the value of $\alpha$ closest to the corner.
- Caution: L-curves are NOT guaranteed to work as a regularization strategy.

- An alternative: (Morozov’s discrepancy principle) Choose $\alpha$ so that the misfit $\|Gm_\alpha - d\|_2$ is the same size as the data noise $\|\delta d\|_2$. 
The L-curve

L-curves are one tool for choosing the regularization parameter $\alpha$:

- Make a plot of the curve $(\| m_\alpha \|_2, \| Gm_\alpha - d \|_2)$
- Typically, this curve looks to be asymptotic to the axes.
- Choose the value of $\alpha$ closest to the corner.
- Caution: L-curves are NOT guaranteed to work as a regularization strategy.
- An alternative: (Morozov’s discrepancy principle) Choose $\alpha$ so that the misfit $\| Gm_\alpha - d \|_2$ is the same size as the data noise $\| \delta d \|_2$. 
Tikhonov’s original interest was in operator equations

\[ d(s) = \int_a^b k(s, t) m(t) \, dt \]

or \( d = Km \) where \( K \) is a compact (bounded = continuous) linear operator from one Hilbert space \( H_1 \) into another \( H_2 \). In this situation:

- Such an operator \( K : H_1 \rightarrow H_2 \) has an adjoint operator \( K^* : H_2 \rightarrow H_1 \) (analogous to transpose of matrix operator.)
- Least squares solutions to \( \min \|Km - d\| \) are just solutions to the normal equation \( K^*Km = K^*d \) (and exist.)
- There is a Moore-Penrose inverse operator \( K^\dagger \) such that \( m = K^\dagger d \) is the least squares solution of least 2-norm. But this operator is generally unbounded (not continuous.)
Tikhonov’s original interest was in operator equations

\[ d(s) = \int_a^b k(s, t) m(t) \, dt \]

or \( d = Km \) where \( K \) is a compact (\textit{bounded} = \textit{continuous}) linear operator from one Hilbert space \( H_1 \) into another \( H_2 \). In this situation:

- Such an operator \( K : H_1 \to H_2 \) has an \textbf{adjoint operator} \( K^* : H_2 \to H_1 \) (analogous to transpose of matrix operator.)

- Least squares solutions to \( \min \|Km - d\| \) are just solutions to the \textbf{normal equation} \( K^*Km = K^*d \) (and exist.)

- There is a Moore-Penrose inverse operator \( K^\dagger \) such that \( m = K^\dagger d \) is the least squares solution of least 2-norm. But this operator is generally \textit{unbounded} (not continuous.)
Tikhonov's original interest was in operator equations

\[ d(s) = \int_a^b k(s, t) m(t) \, dt \]

or \( d = Km \) where \( K \) is a compact (bounded = continuous) linear operator from one Hilbert space \( H_1 \) into another \( H_2 \). In this situation:

- Such an operator \( K : H_1 \rightarrow H_2 \) has an **adjoint operator** \( K^* : H_2 \rightarrow H_1 \) (analogous to transpose of matrix operator.)
- Least squares solutions to \( \min \|Km - d\| \) are just solutions to the **normal** equation \( K^*Km = K^*d \) (and exist.)
- There is a Moore-Penrose inverse operator \( K^\dagger \) such that \( m = K^\dagger d \) is the least squares solution of least 2-norm. But this operator is generally **unbounded** (not continuous.)
Tikhonov’s original interest was in operator equations

\[ d(s) = \int_a^b k(s, t) m(t) \, dt \]

or \( d = Km \) where \( K \) is a compact (bounded = continuous) linear operator from one Hilbert space \( H_1 \) into another \( H_2 \). In this situation:

- Such an operator \( K : H_1 \rightarrow H_2 \) has an adjoint operator \( K^* : H_2 \rightarrow H_1 \) (analogous to transpose of matrix operator.)

- Least squares solutions to \( \min \|Km - d\| \) are just solutions to the normal equation \( K^*Km = K^*d \) (and exist.)

- There is a Moore-Penrose inverse operator \( K^\dagger \) such that \( m = K^\dagger d \) is the least squares solution of least 2-norm. But this operator is generally unbounded (not continuous.)
More on Tikhonov’s operator equation:

- The operator \((K^*K + \alpha I)\) is bounded with bounded inverse and the regularized problem \((K^*K + \alpha I) m = K^*d\) has a unique solution \(m_\alpha\).

- Given that \(\delta = \|\delta d\|\) is the noise level and that the problem actually solved is \((K^*K + \alpha I) m = K^*d^\delta\) with \(d^\delta = d + \delta d\) yielding \(m^\delta_\alpha\). Tikhonov defines a regular algorithm to be a choice \(\alpha = \alpha(\delta)\) such that

\[
\alpha(\delta) \to 0 \text{ and } m^\delta_{\alpha(\delta)} \to K^\dagger d \text{ as } \delta \to 0.
\]

- Morozov’s discrepancy principle is a regular algorithm.

Finish Section 5.2 by exploring the Example 5.1 file, which constructs the L-curve of the Shaw problem using tools from the Regularization Toolbox.
More on Tikhonov’s operator equation:

- The operator \((K^*K + \alpha I)\) is bounded with bounded inverse and the regularized problem \((K^*K + \alpha I) m = K^*d\) has a unique solution \(m_\alpha\).

- Given that \(\delta = \|\delta d\|\) is the noise level and that the problem actually solved is \((K^*K + \alpha I) m = K^*d^\delta\) with \(d^\delta = d + \delta d\) yielding \(m_\alpha^\delta\). Tikhonov defines a regular algorithm to be a choice \(\alpha = \alpha(\delta)\) such that

\[
\alpha(\delta) \to 0 \text{ and } m_{\alpha(\delta)}^\delta \to K^\dagger d \text{ as } \delta \to 0.
\]

- Morozov’s discrepancy principle is a regular algorithm.

Finish Section 5.2 by exploring the Example 5.1 file, which constructs the L-curve of the Shaw problem using tools from the Regularization Toolbox.
More on Tikhonov’s operator equation:

- The operator \((K^*K + \alpha I)\) is bounded with bounded inverse and the regularized problem \((K^*K + \alpha I) m = K^*d\) has a unique solution \(m_\alpha\).

- Given that \(\delta = \|\delta d\|\) is the noise level and that the problem actually solved is \((K^*K + \alpha I) m = K^*d^\delta\) with \(d^\delta = d + \delta d\) yielding \(m^\delta_\alpha\), Tikhonov defines a regular algorithm to be a choice \(\alpha = \alpha(\delta)\) such that

  \[
  \alpha(\delta) \to 0 \text{ and } m^\delta_{\alpha(\delta)} \to K^\dagger d \text{ as } \delta \to 0.
  \]

- Morozov’s discrepancy principle is a regular algorithm.

Finish Section 5.2 by exploring the Example 5.1 file, which constructs the L-curve of the Shaw problem using tools from the Regularization Toolbox.
More on Tikhonov’s operator equation:

- The operator \((K^*K + \alpha I)\) is bounded with bounded inverse and the **regularized problem** \((K^*K + \alpha I) m = K^*d\) has a unique solution \(m_\alpha\).

- Given that \(\delta = \|\delta d\|\) is the noise level and that the problem actually solved is \((K^*K + \alpha I) m = K^*d^\delta\) with \(d^\delta = d + \delta d\) yielding \(m^\delta\), Tikhonov defines a **regular algorithm** to be a choice \(\alpha = \alpha(\delta)\) such that

  \[
  \alpha(\delta) \to 0 \text{ and } m^\delta_{\alpha(\delta)} \to K^\dagger d \text{ as } \delta \to 0.
  \]

- Morozov’s discrepancy principle is a regular algorithm.

Finish Section 5.2 by exploring the Example 5.1 file, which constructs the L-curve of the Shaw problem using tools from the Regularization Toolbox.
Chapter 5: Tikhonov Regularization

5.2: SVD Implementation of Tikhonov Regularization

5.3: Resolution, Bias and Uncertainty in the Tikhonov Solution
Chapter 5: Tikhonov Regularization

5.2: SVD Implementation of Tikhonov Regularization

5.3: Resolution, Bias and Uncertainty in the Tikhonov Solution
Resolution Matrix

Definition:
Resolution matrix for a regularized problem starts with this observation:

- Let $G^\ddag \equiv \left(G^T G + \alpha^2 I\right)^{-1} G^T$. 
- Then $m_\alpha = G^\ddag d = \sum_{j=1}^{p} f_j \frac{U_j^T d}{\sigma_j} V_j = VFS^\dagger U^T d$. 
- Resolution matrix: $R_{m,\alpha} = G^\ddag G = VFV^T$
Definition:

Resolution matrix for a regularized problem starts with this observation:

1. Let \( G^\dagger \equiv \left( G^T G + \alpha^2 I \right)^{-1} G^T \).

2. Then \( m_\alpha = G^\dagger d = \sum_{j=1}^{p} f_j \frac{U_j^T d}{\sigma_j} V_j = V F S^\dagger U^T d \).

3. Resolution matrix: \( R_{m,\alpha} = G^\dagger G = V F V^T \).
Definition:

Resolution matrix for a regularized problem starts with this observation:

- Let \( G^\ddagger \equiv \left( G^T G + \alpha^2 I \right)^{-1} G^T \).

- Then \( m_\alpha = G^\ddagger d = \sum_{j=1}^{p} f_j \frac{U_j^T d}{\sigma_j} V_j = VFS^T U^T d \).

Resolution matrix: \( R_{m,\alpha} = G^\ddagger G = VFV^T \)
Resolution Matrix

**Definition:**

Resolution matrix for a regularized problem starts with this observation:

- Let \( G^\dagger \equiv \left( G^T G + \alpha^2 I \right)^{-1} G^T. \)
- Then \( m_\alpha = G^\dagger d = \sum_{j=1}^{p} f_j \frac{U_j^T d}{\sigma_j} V_j = VFS^\dagger U^T d. \)
- Resolution matrix: \( R_{m,\alpha} = G^\dagger G = VFV^T. \)