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

Instructor: Thomas Shores
Department of Mathematics

Lecture 18, March 21, 2006
AvH 10
<table>
<thead>
<tr>
<th>Chapter 5: Tikhonov Regularization</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.2: SVD Implementation of Tikhonov Regularization</td>
</tr>
<tr>
<td>5.3: Resolution, Bias and Uncertainty in the Tikhonov Solution</td>
</tr>
<tr>
<td>5.4: Higher Order Tikhonov Regularization</td>
</tr>
<tr>
<td>TGSVD and GCV</td>
</tr>
<tr>
<td>Error Bounds</td>
</tr>
</tbody>
</table>
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}
\begin{bmatrix}
m \\

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

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

\[
\begin{bmatrix}
\frac{\sigma_1}{\sigma_1^2 + \alpha^2} \\
\vdots \\
\frac{\sigma_p}{\sigma_p^2 + \alpha^2}
\end{bmatrix}
\begin{bmatrix}
V \\
0
\end{bmatrix}
U^T
\]
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
  \[
  (G^T G + \alpha^2 I)^{-1} G^T =
  \begin{bmatrix}
  \frac{\sigma_1}{\sigma_1^2 + \alpha^2} \\
  \vdots \\
  \frac{\sigma_p}{\sigma_p^2 + \alpha^2}
  \end{bmatrix}
  \begin{bmatrix}
  V \\
  \vdots \\
  U^T
  \end{bmatrix}
  \]
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

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

To solve, calculate

\[
(G^T G + \alpha^2 I)^{-1} G^T = 
\begin{bmatrix}
\frac{\sigma_1}{\sigma_1^2 + \alpha^2} \\
\vdots \\
\frac{\sigma_p}{\sigma_p^2 + \alpha^2}
\end{bmatrix} 
\begin{bmatrix}
\sigma_1 \\
\vdots \\
\sigma_p
\end{bmatrix}
\begin{bmatrix}
U^T
\end{bmatrix}
\]
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
  \[
  \left( G^T G + \alpha^2 I \right)^{-1} G^T = 
  \begin{bmatrix}
  \sigma_1/\sigma_1^2 + \alpha^2 \\
  \sigma_2/\sigma_2^2 + \alpha^2 \\
  \vvdots \\
  \sigma_p/\sigma_p^2 + \alpha^2
  \end{bmatrix} \begin{bmatrix}
  U^T \\
  0 \\
  \vvdots
  \end{bmatrix}
  \]
From the previous equation we obtain that the Moore-Penrose inverse and solution to the regularized problem are given by

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

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

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 \( h = U^T d \).)
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.
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 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 $\|\delta d\|_2$. 

Instructor: Thomas Shores Department of Mathematics
Math 4/896: Seminar in Mathematics Topic: Inverse Theory
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 $\|\delta d\|_2$. 

Instructor: Thomas Shores Department of Mathematics

Math 4/896: Seminar in Mathematics Topic: Inverse Theory
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 $\|\delta d\|_2$. 

Instructor: Thomas Shores Department of Mathematics
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 \( d = K m \) 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 \textit{adjoint operator} \( K^* : H_2 \to H_1 \) (analogous to transpose of matrix operator.)
- Least squares solutions to \( \min \| K m - d \| \) are just solutions to the \textit{normal equation} \( K^* K m = K^* d \) (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 \( K : H_1 \to H_2 \) has an **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 **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 \( d = K m \) 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 \( K : H_1 \to H_2 \) has an **adjoint operator** \( K^* : H_2 \to H_1 \) (analogous to transpose of matrix operator.)

- Least squares solutions to \( \min \| K m - d \| \) are just solutions to the **normal** equation \( K^* K m = K^* d \) (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 \textbf{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 \(m_\alpha^\delta\) Tikhonov defines a regular algorithm to be a choice \(\alpha = \alpha(\delta)\) such that

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

- 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 \(m_\alpha^\delta\) Tikhonov defines a **regular algorithm** to be a choice \(\alpha = \alpha(\delta)\) such that

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

- 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.
Chapter 5: Tikhonov Regularization

5.2: SVD Implementation of Tikhonov Regularization
5.3: Resolution, Bias and Uncertainty in the Tikhonov Solution
5.4: Higher Order Tikhonov Regularization
TGSVD and GCV
Error Bounds
Chapter 5: Tikhonov Regularization

5.2: SVD Implementation of Tikhonov Regularization
5.3: Resolution, Bias and Uncertainty in the Tikhonov Solution
5.4: Higher Order Tikhonov Regularization
TGSVD and GCV
Error Bounds
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 \( m_\alpha = G^\dagger d = \sum_{j=1}^{p} f_j \frac{U_j^T d}{\sigma_j} V_j = VFS^\dagger U^T d. \)
- Model resolution matrix: \( R_{m,\alpha} = G^\dagger G = VFV^T \)
- Data resolution matrix: \( R_{d,\alpha} = GG^\dagger = UFU^T \)

The Example 5.1 file constructs the model resolution matrix of the Shaw problem and shows poor resolution in this case.
Resolution Matrix

Definition:

Resolution matrix for a regularized problem starts with this observation:

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

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

- Model resolution matrix: \( R_{m,\alpha} = G^\dagger G = VFV^T \)

- Data resolution matrix: \( R_{d,\alpha} = GG^\dagger = UFU^T \)

The Example 5.1 file constructs the model resolution matrix of the Shaw problem and shows poor resolution in this case.
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 \( m_\alpha = G^\dagger d = \sum_{j=1}^{p} \frac{f_j}{\sigma_j} U_j^T d \).
- Model resolution matrix: \( R_{m,\alpha} = G^\dagger G = VFV^T \)
- Data resolution matrix: \( R_{d,\alpha} = GG^\dagger = UFU^T \)

The Example 5.1 file constructs the model resolution matrix of the Shaw problem and shows poor resolution in this case.
Resolution Matrix

Definition:

Resolution matrix for a regularized problem starts with this observation:

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

The Example 5.1 file constructs the model resolution matrix of the Shaw problem and shows poor resolution in this case.
Resolution Matrix

**Definition:**

Resolution matrix for a regularized problem starts with this observation:

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

The Example 5.1 file constructs the model resolution matrix of the Shaw problem and shows poor resolution in this case.
Chapter 5: Tikhonov Regularization

5.2: SVD Implementation of Tikhonov Regularization
5.3: Resolution, Bias and Uncertainty in the Tikhonov Solution
5.4: Higher Order Tikhonov Regularization

TGSVD and GCV
Error Bounds
Basic Idea

We can think of the regularization term $\alpha^2 \|m\|_2^2$ as favoring minimizing the 0-th order derivative of a function $m(x)$ under the hood. Alternatives:

- Minimize a matrix approximation to $m'(x)$. This is a first order method.
- Minimize a matrix approximation to $m''(x)$. This is a second order method.
- These lead to new minimization problems: to minimize

$$\| Gm - d \|_2^2 + \alpha^2 \| Lm \|_2^2.$$ 

- How do we resolve this problem as we did with $L = I$?
Higher Order Regularization

Basic Idea

We can think of the regularization term $\alpha^2 \|m\|_2^2$ as favoring minimizing the 0-th order derivative of a function $m(x)$ under the hood. Alternatives:

- Minimize a matrix approximation to $m'(x)$. This is a first order method.
- Minimize a matrix approximation to $m''(x)$. This is a second order method.
- These lead to new minimization problems: to minimize

$$\|Gm - d\|_2^2 + \alpha^2 \|Lm\|_2^2.$$

- How do we resolve this problem as we did with $L = I$?
Higher Order Regularization

Basic Idea

We can think of the regularization term $\alpha^2 \|m\|^2_2$ as favoring minimizing the 0-th order derivative of a function $m(x)$ under the hood. Alternatives:

- Minimize a matrix approximation to $m'(x)$. This is a first order method.
- Minimize a matrix approximation to $m''(x)$. This is a second order method.

These lead to new minimization problems: to minimize

$$\|Gm - d\|^2_2 + \alpha^2 \|Lm\|^2_2.$$ 

- How do we resolve this problem as we did with $L = I$?
Higher Order Regularization

Basic Idea

We can think of the regularization term \( \alpha^2 \|m\|_2^2 \) as favoring minimizing the 0-th order derivative of a function \( m(x) \) under the hood. Alternatives:

- Minimize a matrix approximation to \( m'(x) \). This is a first order method.
- Minimize a matrix approximation to \( m''(x) \). This is a second order method.
- These lead to new minimization problems: to minimize

\[
\|Gm - d\|_2^2 + \alpha^2 \|Lm\|_2^2.
\]

How do we resolve this problem as we did with \( L = I \)?
Higher Order Regularization

Basic Idea

We can think of the regularization term $\alpha^2 \| \mathbf{m} \|^2_2$ as favoring minimizing the 0-th order derivative of a function $m(x)$ under the hood. Alternatives:

- Minimize a matrix approximation to $m'(x)$. This is a first order method.
- Minimize a matrix approximation to $m''(x)$. This is a second order method.
- These lead to new minimization problems: to minimize

$$\| G\mathbf{m} - \mathbf{d} \|^2_2 + \alpha^2 \| L\mathbf{m} \|^2_2.$$ 

- How do we resolve this problem as we did with $L = I$?
We will explore approximations to first and second derivatives at the board.
Theorem

Let $G$ be an $m \times n$ matrix and $L$ a $p \times n$ matrix. Then there exist $m \times m$ orthogonal $U$, $p \times p$ orthogonal $V$ and $n \times n$ nonsingular matrix $X$ with $m \geq n \geq \min\{p, n\} = q$ such that

$$U^T GX = \text{diag}\{c_1, c_2, \ldots, c_n\}$$
$$V^T LX = \text{diag}\{s_1, s_2, \ldots, s_q\}$$
$$C^T C + S^T S = 1$$

$$0 \leq c_1 \leq c_2 \cdots \leq c_n \leq 1$$
$$1 \geq s_1 \geq s_2 \cdots \geq s_n \geq 0$$

The numbers $\gamma_i = c_i/s_i$, $i = 1, \ldots, q$ are called the generalized singular values of $G$ and $L$ and $0 \leq \gamma_1 \leq \gamma_2 \cdots \leq \gamma_q$.

Notes: If rank $(L) = q$, then the singular values are finite.
The minimization problem is shown, just as we did earlier, to be equivalent to the problem

\[
\left( G^T G + \alpha^2 L^T L \right) \mathbf{m} = G^T \mathbf{d}
\]

which has solution

\[
\mathbf{m}_{\alpha,L} = \left( G^T G + \alpha^2 L^T L \right) G^T \mathbf{d} \equiv G^b \mathbf{d}.
\]

With some work:

\[
\mathbf{m}_{\alpha,L} = \sum_{j=1}^{p} \frac{\gamma_j^2}{\gamma_j^2 + \alpha^2} \frac{\left( U_j^T \mathbf{d} \right)}{c_j} \mathbf{X}_j + \sum_{j=p+1}^{n} \left( U_j^T \mathbf{d} \right) \mathbf{X}_j
\]
Chapter 5: Tikhonov Regularization

5.2: SVD Implementation of Tikhonov Regularization
5.3: Resolution, Bias and Uncertainty in the Tikhonov Solution
5.4: Higher Order Tikhonov Regularization

TGSVD and GCV
Error Bounds
TGSVD:

We have seen this idea before. Simply apply it to formula above, remembering that the generalized singular values are reverse ordered.

- Formula becomes

\[
\mathbf{m}_{\alpha,L} = \sum_{j=k}^{p} \frac{\gamma_j^2}{\gamma_j^2 + \alpha^2} \left( \mathbf{U}_j^T \mathbf{d} \right) \mathbf{X}_j + \sum_{j=p+1}^{n} \left( \mathbf{U}_j^T \mathbf{d} \right) \mathbf{X}_j
\]

- Key question: where to start \( k \).

Example 5.6 gives a nice illustration of the ideas. We’ll use the CD script to explore it.
Basic Idea:

Comes from statistical “leave-one-out” cross validation.

- Leave out one data point and use model to predict it.
- Sum these up and choose regularization parameter $\alpha$ that minimizes the sum of the squares of the predictive errors

$$V_0(\alpha) = \frac{1}{m} \sum_{k=1}^{m} \left( \left( Gm^k_{\alpha,L} \right)_k - d_k \right)^2.$$ 

- One can show a good approximation is

$$V_0(\alpha) = \frac{m \| Gm_\alpha - d \|_2}{\text{Tr}(I - GG^h)^2}.$$
Basic Idea:

Comes from statistical “leave-one-out” cross validation.

- Leave out one data point and use model to predict it.
- Sum these up and choose regularization parameter $\alpha$ that minimizes the sum of the squares of the predictive errors

$$V_0(\alpha) = \frac{1}{m} \sum_{k=1}^{m} \left( \left( Gm_{\alpha,L}^k \right)_k - d_k \right)^2.$$  

- One can show a good approximation is

$$V_0(\alpha) = \frac{m \|Gm_{\alpha} - d\|_2^2}{\text{Tr} \left( I - GG^\dagger \right)^2}$$
GCV

Basic Idea:

Comes from statistical “leave-one-out” cross validation.

- Leave out one data point and use model to predict it.
- Sum these up and choose regularization parameter $\alpha$ that minimizes the sum of the squares of the predictive errors:

$$V_0 (\alpha) = \frac{1}{m} \sum_{k=1}^{m} \left( \left( Gm_{\alpha,L}^{[k]} \right)_k - d_k \right)^2.$$ 

- One can show a good approximation is:

$$V_0 (\alpha) = \frac{m}{\text{Tr} (I - GG^H)^2} \| Gm_{\alpha} - d \|^2_2$$
Chapter 5: Tikhonov Regularization

5.2: SVD Implementation of Tikhonov Regularization

5.3: Resolution, Bias and Uncertainty in the Tikhonov Solution

5.4: Higher Order Tikhonov Regularization

TGSVD and GCV

Error Bounds

GCV

Basic Idea:

Comes from statistical “leave-one-out” cross validation.

- Leave out one data point and use model to predict it.
- Sum these up and choose regularization parameter \( \alpha \) that minimizes the sum of the squares of the predictive errors

\[
V_0 (\alpha) = \frac{1}{m} \sum_{k=1}^{m} \left( (Gm_{\alpha,L}^{[k]} - d_k)^2 .
\]

- One can show a good approximation is

\[
V_0 (\alpha) = \frac{m \|Gm_\alpha - d\|_2}{\text{Tr} (I - GG^b)^2}
\]
Chapter 5: Tikhonov Regularization

5.2: SVD Implementation of Tikhonov Regularization
5.3: Resolution, Bias and Uncertainty in the Tikhonov Solution
5.4: Higher Order Tikhonov Regularization
TGSVD and GCV
Error Bounds
Error Estimates:

They exist, even in the hard cases where there is error in both $G$ and $d$.

- In the simpler case, $G$ known exactly, they take the form

$$\frac{\|m_\alpha - \tilde{m}_\alpha\|_2}{\|m_\alpha\|_2} \leq \kappa_\alpha \frac{\|d - \tilde{d}\|_2}{\|Gm_\alpha\|_2}$$

where $\kappa_\alpha$ is inversely proportional to $\alpha$. 

Error Estimates:

They exist, even in the hard cases where there is error in both $G$ and $d$.

- In the simpler case, $G$ known exactly, they take the form

$$\frac{\|m_\alpha - \tilde{m}_\alpha\|_2}{\|m_\alpha\|_2} \leq \kappa_\alpha \frac{\|d - \tilde{d}\|_2}{\|Gm_\alpha\|_2}$$

where $\kappa_\alpha$ is inversely proportional to $\alpha$. 

Error Bounds

Error Estimates:

They exist, even in the hard cases where there is error in both \( G \) and \( d \).

- In the simpler case, \( G \) known exactly, they take the form

\[
\frac{\|m_\alpha - \tilde{m}_\alpha\|_2}{\|m_\alpha\|_2} \leq \kappa_\alpha \frac{\|d - \tilde{d}\|_2}{\|Gm_\alpha\|_2}
\]

where \( \kappa_\alpha \) is inversely proportional to \( \alpha \).