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

- To solve, calculate \((G^T G + \alpha^2 I)^{-1} G^T = \frac{\sigma_1}{\sigma_1^2 + \alpha^2} \begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \vdots \\ \sigma_p \end{bmatrix} U^T \)
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

$$\left( G^T G + \alpha^2 I \right)^{-1} G^T = \begin{bmatrix} \frac{\sigma_1}{\sigma_1^2 + \alpha^2} \\ \frac{\sigma_2}{\sigma_2^2 + \alpha^2} \\ \vdots \\ \frac{\sigma_p}{\sigma_p^2 + \alpha^2} \\ 0 \\ \vdots \end{bmatrix} U^T$$
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 \((G^T G + \alpha^2 I)^{-1} G^T = \)

\[
\begin{bmatrix}
\frac{\sigma_1}{\sigma_1 + \alpha^2} & \ldots & 0 \\
\frac{\sigma_2}{\sigma_2 + \alpha^2} & \ldots & \ldots \\
\frac{\sigma_p}{\sigma_p + \alpha^2} & \ldots & 0
\end{bmatrix} U^T
\]
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:

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

From the previous equation we obtain that the Moore-Penrose inverse and solution to the regularized problem are given by

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

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

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.
- $f(\sigma) = \max\left\{\text{sgn}(\sigma - \epsilon), 0\right\}$ 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 $\|\delta d\|_2$. 
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 $\|\delta d\|_2$. 
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 $\|\delta d\|_2$. 
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 \( K : H_1 \rightarrow H_2 \) 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 \((\text{bounded} = \text{continuous})\) 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 \( d = Km \) 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.)
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

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

- 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.
Resolution Matrix

**Definition:**

Resolution matrix for a regularized problem starts with this observation:

- Let $G^\dagger \equiv \left( G^T G + \alpha^2 I \right)^{-1} G^T$ (generalized inverse)
- Then $m_\alpha = G^\dagger d = \sum_{j=1}^{p} f_j \frac{U_j^T d}{\sigma_j} V_j = V F S^\dagger U^T d$.
- Model resolution matrix: $R_{m,\alpha} = G^\dagger G = V F V^T$
- Data resolution matrix: $R_{d,\alpha} = G G^\dagger = U F U^T$

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 \( G^h \equiv \left( G^T G + \alpha^2 I \right)^{-1} G^T \) (generalized inverse)

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

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

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

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 \( G^\dagger \equiv \left( G^T G + \alpha^2 I \right)^{-1} G^T \) (generalized inverse)
- 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 $G^\dagger \equiv \left( G^T G + \alpha^2 I \right)^{-1} G^T$ (generalized inverse)
- 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 $G^\# \equiv \left(G^T G + \alpha^2 I\right)^{-1} G^T$ (generalized inverse)

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

- Model resolution matrix: $R_{m,\alpha} = G^\# G = VFV^T$

- Data resolution matrix: $R_{d,\alpha} = GG^\# = UFU^T$

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

Outline
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 $L = I$?
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

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

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

$$U^T GX = \text{diag} \{c_1, c_2, \ldots, c_n\} = C$$

$$V^T LX = \text{diag} \{s_1, s_2, \ldots, s_q\} = S$$

$$C^T C + S^T S = 1$$

$$0 \leq c_1 \leq c_2 \cdots \leq c_n \leq 1$$

$$1 \geq s_1 \geq s_2 \cdots \geq s_n \geq 0$$

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

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

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

which has solution

\[
m_{\alpha,L} = \left( G^T G + \alpha^2 L^T L \right)^{-1} G^T d \equiv G^h d.
\]

With some work:

\[
m_{\alpha,L} = \sum_{j=1}^{p} \gamma_j^2 \left( \frac{U_j^T d}{c_j} \right) X_j + \sum_{j=p+1}^{n} \left( U_j^T d \right) X_j
\]
The Experiment:

Place sensors at vertical depths $z_j$, $j = 1, \ldots, n$, in a borehole, then:

- Generate a seismic wave at ground level, $t = 0$.
- Measure arrival times $d_j = t(z_j)$, $j = 1, \ldots, n$.
- Now try to recover the slowness function $s(z)$, given

$$t(z) = \int_0^z s(\xi) \, d\xi = \int_0^\infty s(\xi) H(z - \xi) \, d\xi$$

- It should be easy: $s(z) = t'(z)$.
- Hmmm.....or is it?

Let’s do Example 4-5 from the CD.
Vertical Seismic Profiling Example

The Experiment:
Place sensors at vertical depths $z_j$, $j = 1, \ldots, n$, in a borehole, then:

- Generate a seismic wave at ground level, $t = 0$.
- Measure arrival times $d_j = t(z_j)$, $j = 1, \ldots, n$.
- Now try to recover the slowness function $s(z)$, given

$$t(z) = \int_0^z s(\xi) d\xi = \int_0^\infty s(\xi) H(z - \xi) d\xi$$

- It should be easy: $s(z) = t'(z)$.
- Hmmm…..or is it?

Let’s do Example 4-5 from the CD.
The Experiment:

Place sensors at vertical depths $z_j$, $j = 1, \ldots, n$, in a borehole, then:

- Generate a seismic wave at ground level, $t = 0$.
- Measure arrival times $d_j = t(z_j)$, $j = 1, \ldots, n$.
- Now try to recover the slowness function $s(z)$, given

$$t(z) = \int_0^z s(\xi) \, d\xi = \int_0^\infty s(\xi) \, H(z - \xi) \, d\xi$$

- It should be easy: $s(z) = t'(z)$.
- Hmmm.....or is it?

Let’s do Example 4-5 from the CD.
Vertical Seismic Profiling Example

The Experiment:
Place sensors at vertical depths $z_j$, $j = 1, \ldots, n$, in a borehole, then:

- Generate a seismic wave at ground level, $t = 0$.
- Measure arrival times $d_j = t(z_j)$, $j = 1, \ldots, n$.
- Now try to recover the slowness function $s(z)$, given

$$
t(z) = \int_0^z s(\xi) d\xi = \int_0^\infty s(\xi) H(z - \xi) d\xi
$$

- It should be easy: $s(z) = t'(z)$.
- Hmmm.....or is it?

Let’s do Example 4-5 from the CD.
The Experiment:

Place sensors at vertical depths $z_j$, $j = 1, \ldots, n$, in a borehole, then:

- Generate a seismic wave at ground level, $t = 0$.
- Measure arrival times $d_j = t(z_j)$, $j = 1, \ldots, n$.
- Now try to recover the slowness function $s(z)$, given

$$t(z) = \int_{0}^{z} s(\xi) \, d\xi = \int_{0}^{\infty} s(\xi) \, H(z - \xi) \, d\xi$$

- It should be easy: $s(z) = t'(z)$.
- Hmmm.....or is it?

Let’s do Example 4-5 from the CD.
The Experiment:

Place sensors at vertical depths $z_j$, $j = 1, \ldots, n$, in a borehole, then:

- Generate a seismic wave at ground level, $t = 0$.
- Measure arrival times $d_j = t(z_j)$, $j = 1, \ldots, n$.
- Now try to recover the slowness function $s(z)$, given

$$t(z) = \int_0^z s(\xi) \, d\xi = \int_0^\infty s(\xi) \, H(z - \xi) \, d\xi$$

- It should be easy: $s(z) = t'(z)$.
- Hmmm.....or is it?

Let’s do Example 4-5 from the CD.
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

$$ m_{\alpha, L} = \sum_{j=k}^{p} \frac{\gamma_j^2}{\gamma_j^2 + \alpha^2} \left( U_j^T d \right) X_j + \sum_{j=p+1}^{n} \left( U_j^T d \right) X_j. $$

- Key question: where to start $k$. 
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

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

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

- One can show a good approximation is

\[ V_0(\alpha) = \frac{m \| G m_\alpha - d \|^2}{\text{Tr} \ (I - G G^\dagger)^2}. \]
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

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

- One can show a good approximation is

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

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

- One can show a good approximation is

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

As time permits, do Example 5.8 from CD.
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 \( \alpha \).

As time permits, do Example 5.8 from CD.
More Estimates:

- Suppose that the true model $\mathbf{m}_{true}$ is “smooth” in the sense that there exists vector $\mathbf{w}$ such that $(p = 1) \mathbf{m}_{true} = \mathbf{G}^T \mathbf{w}$ or $(p = 2) \mathbf{m}_{true} = \mathbf{G}^T \mathbf{G} \mathbf{w}$. Let $\Delta = \delta / \|\mathbf{w}\|$ and $\gamma = 1$ if $p = 1$ and $\gamma = 4$ if $p = 2$. Then the choice $\hat{\alpha} = (\Delta / \gamma)^{1/(p+1)}$ is optimal in the sense that we have the error bound

$$\|\mathbf{m}_{true} - \mathbf{G}^{\dagger} \mathbf{d}\|_2 = \gamma (p + 1) \hat{\alpha}^p = O \left( \Delta^{p/(p+1)} \right).$$

- This is about the best we can do. Its significance: the best we can hope for is about 1/2 or 2/3 of the significant digits in the data.
More Estimates:

- Suppose that the true model $m_{\text{true}}$ is “smooth” in the sense that there exists vector $w$ such that ($p=1$) $m_{\text{true}} = G^T w$ or ($p=2$) $m_{\text{true}} = G^T G w$. Let $\Delta = \delta / \|w\|$ and $\gamma = 1$ if $p = 1$ and $\gamma = 4$ if $p = 2$. Then the choice $\hat{\alpha} = (\Delta / \gamma)^{1/(p+1)}$ is optimal in the sense that we have the error bound

$$\left\| m_{\text{true}} - G^d d \right\|_2 = \gamma (p + 1) \hat{\alpha}^p = O \left( \Delta \frac{p}{p+1} \right).$$

- This is about the best we can do. Its significance: the best we can hope for is about 1/2 or 2/3 of the significant digits in the data.
More Estimates:

Suppose that the true model $m_{\text{true}}$ is “smooth” in the sense that there exists vector $w$ such that $(p = 1) \ m_{\text{true}} = G^T w$ or $(p = 2) \ m_{\text{true}} = G^T Gw$. Let $\Delta = \delta / \|w\|$ and $\gamma = 1$ if $p = 1$ and $\gamma = 4$ if $p = 2$. Then the choice $\hat{\alpha} = (\Delta / \gamma)^{1/(p+1)}$ is optimal in the sense that we have the error bound

$$\|m_{\text{true}} - G^{\hat{\alpha}} d\|_2 = \gamma (p + 1) \hat{\alpha}^p = O \left( \Delta^{\frac{p}{p+1}} \right).$$

This is about the best we can do. Its significance: the best we can hope for is about 1/2 or 2/3 of the significant digits in the data.