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

SVD Implementation

To solve the Tikhonov regularized problem, first recall:

- $\nabla \left(\|G\mathbf{m} - \mathbf{d}\|_2^2 + \alpha^2 \|\mathbf{m}\|_2^2 \right) = \left(G^T G\mathbf{m} - G^T\mathbf{d} \right) + \alpha^2 \mathbf{m}$
- Equate to zero and these are the normal equations for the system $\begin{bmatrix} G \\ \alpha I \end{bmatrix} \mathbf{m} = \begin{bmatrix} \mathbf{d} \\ \mathbf{0} \end{bmatrix}$, or $\left(G^T G + \alpha^2 I \right) \mathbf{m} = G^T \mathbf{d}$
- To solve, calculate $\left(G^T G + \alpha^2 I \right)^{-1} G^T =$

$$V \begin{bmatrix} \frac{\sigma_1}{\sigma_1^2+\alpha^2} & & & & \\ & \ddots & & & \\ & & \frac{\sigma_p}{\sigma_p^2+\alpha^2} & & \\ & & & 0 & \\ & & & & \ddots \end{bmatrix} U^T$$

SVD Implementation

To solve the Tikhonov regularized problem, first recall:

- $\nabla\left(\|G\mathbf{m}-\mathbf{d}\|_2^2+\alpha^2\|\mathbf{m}\|_2^2\right)=\left(G^TG\mathbf{m}-G^T\mathbf{d}\right)+\alpha^2\mathbf{m}$
- Equate to zero and these are the normal equations for the system $\begin{bmatrix} G \\ \alpha I \end{bmatrix}\mathbf{m}=\begin{bmatrix} \mathbf{d} \\ \mathbf{0} \end{bmatrix}$, or $\left(G^TG+\alpha^2 I\right)\mathbf{m}=G^T\mathbf{d}$
- To solve, calculate $\left(G^TG+\alpha^2 I\right)^{-1}G^T=$

$$V\begin{bmatrix} \frac{\sigma_1}{\sigma_1^2+\alpha^2} & & & & \\ & \ddots & & & \\ & & \frac{\sigma_p}{\sigma_p^2+\alpha^2} & & \\ & & & 0 & \\ & & & & \ddots \end{bmatrix}U^T$$

SVD Implementation

To solve the Tikhonov regularized problem, first recall:

- $\nabla \left(\|G\mathbf{m} - \mathbf{d}\|_2^2 + \alpha^2 \|\mathbf{m}\|_2^2 \right) = \left(G^T G\mathbf{m} - G^T\mathbf{d} \right) + \alpha^2 \mathbf{m}$
- Equate to zero and these are the normal equations for the system $\begin{bmatrix} G \\ \alpha I \end{bmatrix} \mathbf{m} = \begin{bmatrix} \mathbf{d} \\ \mathbf{0} \end{bmatrix}$, or $\left(G^T G + \alpha^2 I \right) \mathbf{m} = G^T \mathbf{d}$
- To solve, calculate $\left(G^T G + \alpha^2 I \right)^{-1} G^T =$

$$V \begin{bmatrix} \frac{\sigma_1}{\sigma_1^2+\alpha^2} & & & & \\ & \ddots & & & \\ & & \frac{\sigma_p}{\sigma_p^2+\alpha^2} & & \\ & & & 0 & \\ & & & & \ddots \end{bmatrix} U^T$$

SVD Implementation

To solve the Tikhonov regularized problem, first recall:

- $\nabla \left(\|G\mathbf{m} - \mathbf{d}\|_2^2 + \alpha^2 \|\mathbf{m}\|_2^2 \right) = \left(G^T G\mathbf{m} - G^T\mathbf{d} \right) + \alpha^2 \mathbf{m}$
- Equate to zero and these are the normal equations for the system $\begin{bmatrix} G \\ \alpha I \end{bmatrix} \mathbf{m} = \begin{bmatrix} \mathbf{d} \\ \mathbf{0} \end{bmatrix}$, or $\left(G^T G + \alpha^2 I\right) \mathbf{m} = G^T \mathbf{d}$
- To solve, calculate $\left(G^T G + \alpha^2 I\right)^{-1} G^T =$

$$V \begin{bmatrix} \frac{\sigma_1}{\sigma_1^2+\alpha^2} & & & & \\ & \ddots & & & \\ & & \frac{\sigma_p}{\sigma_p^2+\alpha^2} & & \\ & & & 0 & \\ & & & & \ddots \end{bmatrix} U^T$$

SVD Implementation

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

$$G_\alpha^\dagger = \sum_{j=1}^{p} \frac{\sigma_j}{\sigma_j^2 + \alpha^2} \mathbf{V}_j \mathbf{U}_j^T$$

$$\mathbf{m}_\alpha = G^\dagger \mathbf{d} = \sum_{j=1}^{p} \frac{\sigma_j^2}{\sigma_j^2 + \alpha^2} \frac{\left(\mathbf{U}_j^T \mathbf{d}\right)}{\sigma_j} \mathbf{V}_j$$

which specializes to the generalized inverse solution we have seen in the case that G is full column rank and $\alpha = 0$. (Remember $\mathbf{d} = U\mathbf{h}$ so that $\mathbf{h} = U^T\mathbf{d}$.)

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

The L-curve

L-curves are one tool for choosing the regularization paramter α:

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

The L-curve

L-curves are one tool for choosing the regularization paramter α:

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

The L-curve

L-curves are one tool for choosing the regularization paramter α:

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

The L-curve

L-curves are one tool for choosing the regularization paramter α:

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

The L-curve

L-curves are one tool for choosing the regularization paramter α:

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

The L-curve

L-curves are one tool for choosing the regularization paramter α:

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

Historical Notes

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.)

Historical Notes

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 $d = Km$ 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.)

Historical Notes

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_α.
- 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_α^δ 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.

Historical Notes

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_α.
- 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_α^δ 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_α.
- 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_α^δTikhonov 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_α.
- 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_α^δTikhonov 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.

Outline

Resolution Matrix

Definition:

Resolution matrix for a regularized problem starts with this observation:

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

Resolution Matrix

Definition:

Resolution matrix for a regularized problem starts with this observation:

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

Resolution Matrix

Definition:

Resolution matrix for a regularized problem starts with this observation:

- Let $G^\natural \equiv \left(G^TG + \alpha^2 I\right)^{-1} G^T$.
- Then $\mathbf{m}_\alpha = G^\natural \mathbf{d} = \sum_{j=1}^{p} f_j \frac{\left(\mathbf{U}_j^T \mathbf{d}\right)}{\sigma_j} \mathbf{V}_j = VFS^\dagger U^T \mathbf{d}$.
- Resolution matrix: $R_{\mathbf{m},\alpha} = G^\natural G = VFV^T$

Resolution Matrix

Definition:

Resolution matrix for a regularized problem starts with this observation:

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