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

Instructor: Thomas Shores
Department of Mathematics

Lecture 15, February 28, 2006
AvH 10

Outline

Linear Tomography Models

Note: Rank deficient problems are *automatically* ill-posed.

Basic Idea:

A ray emanates from one known point to another along a known path ℓ, with a detectable property which is observable data. These data are used to estimate a travel property of the medium. For example, let the property be travel time, so that:

- Travel time is given by $t = \int_\ell \frac{dt}{dx} dx = \int_\ell \frac{1}{v(x)} dx$
- We can linearize by making paths straight lines.
- Discretize by embedding the medium in a square (cube) and subdividing it into regular subsquares (cubes) in which we assume "slowness" (parameter of the problem) is constant.
- Transmit the ray along specified paths and collect temporal data to be used in estimating "slowness".

Linear Tomography Models

Note: Rank deficient problems are *automatically* ill-posed.

Basic Idea:

A ray emanates from one known point to another along a known path ℓ, with a detectable property which is observable data. These data are used to estimate a travel property of the medium. For example, let the property be travel time, so that:

- Travel time is given by $t = \int_\ell \frac{dt}{dx} dx = \int_\ell \frac{1}{v(x)} dx$
- We can linearize by making paths straight lines.
- Discretize by embedding the medium in a square (cube) and subdividing it into regular subsquares (cubes) in which we assume "slowness" (parameter of the problem) is constant.
- Transmit the ray along specified paths and collect temporal data to be used in estimating "slowness".

Linear Tomography Models

Note: Rank deficient problems are *automatically* ill-posed.

Basic Idea:

A ray emanates from one known point to another along a known path ℓ, with a detectable property which is observable data. These data are used to estimate a travel property of the medium. For example, let the property be travel time, so that:

- Travel time is given by $t = \int_\ell \frac{dt}{dx} dx = \int_\ell \frac{1}{v(x)} dx$
- We can linearize by making paths straight lines.
- Discretize by embedding the medium in a square (cube) and subdividing it into regular subsquares (cubes) in which we assume "slowness" (parameter of the problem) is constant.
- Transmit the ray along specified paths and collect temporal data to be used in estimating "slowness".

Linear Tomography Models

Note: Rank deficient problems are *automatically* ill-posed.

Basic Idea:

A ray emanates from one known point to another along a known path ℓ, with a detectable property which is observable data. These data are used to estimate a travel property of the medium. For example, let the property be travel time, so that:

- Travel time is given by $t = \int_\ell \frac{dt}{dx} dx = \int_\ell \frac{1}{v(x)} dx$
- We can linearize by making paths straight lines.
- Discretize by embedding the medium in a square (cube) and subdividing it into regular subsquares (cubes) in which we assume "slowness" (parameter of the problem) is constant.
- Transmit the ray along specified paths and collect temporal data to be used in estimating "slowness".

Linear Tomography Models

Note: Rank deficient problems are *automatically* ill-posed.

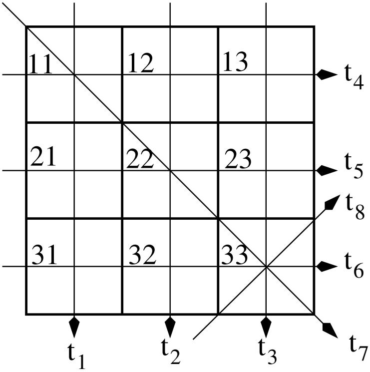
Basic Idea:

A ray emanates from one known point to another along a known path ℓ, with a detectable property which is observable data. These data are used to estimate a travel property of the medium. For example, let the property be travel time, so that:

- Travel time is given by $t = \int_\ell \frac{dt}{dx} dx = \int_\ell \frac{1}{v(x)} dx$
- We can linearize by making paths straight lines.
- Discretize by embedding the medium in a square (cube) and subdividing it into regular subsquares (cubes) in which we assume "slowness" (parameter of the problem) is constant.
- Transmit the ray along specified paths and collect temporal data to be used in estimating "slowness".

Example 1.6 and 4.1

The figure for this experiment (assume each subsquare has sides of length 1, so the size of the large square is 3×3):



Example 1.6 and 4.1

Corresponding matrix of distances G (rows of G represent distances along corresponding path, columns the ray distances across each subblock) and resulting system:

$$G\mathbf{m} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ \sqrt{2} & 0 & 0 & 0 & \sqrt{2} & 0 & 0 & 0 & \sqrt{2} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sqrt{2} \end{bmatrix} \begin{bmatrix} s_{11} \\ s_{12} \\ s_{13} \\ s_{21} \\ s_{22} \\ s_{23} \\ s_{31} \\ s_{32} \\ s_{33} \end{bmatrix} = \begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \\ t_6 \\ t_7 \\ t_8 \end{bmatrix} = \mathbf{d}$$

Observe: in this Example $m = 8$ and $n = 9$, so this is rank deficient. Now run the example file for this example. We need to fix the path. Assuming we are in the directory MatlabTools, do the following:

```
>addpath('Examples/chap4/examp1')
>path
```

Outline

What Are They?

These problems arise due to ill-conditioning of G, *as opposed to a rank deficiency problem. Theoretically*, they are not ill-posed, like the Hilbert matrix. But practically speaking, they behave like ill-posed problems. Authors present a hierarchy of sorts for a problem with system $G\mathbf{m} = \mathbf{d}$. These order expressions are valid as $j \to \infty$.

- $\mathcal{O}\left(\frac{1}{j^\alpha}\right)$ with $0 < \alpha \le 1$, the problem is **mildly** ill-posed.
- $\mathcal{O}\left(\frac{1}{j^\alpha}\right)$ with $\alpha > 1$, the problem is **moderately** ill-posed.
- $\mathcal{O}\left(e^{-\alpha j}\right)$ with $0 < \alpha$, the problem is **severely** ill-posed.

What Are They?

These problems arise due to ill-conditioning of G, *as opposed to a rank deficiency problem. Theoretically*, they are not ill-posed, like the Hilbert matrix. But practically speaking, they behave like ill-posed problems. Authors present a hierarchy of sorts for a problem with system $G\mathbf{m} = \mathbf{d}$. These order expressions are valid as $j \to \infty$.

- $\mathcal{O}\left(\frac{1}{j^\alpha}\right)$ with $0 < \alpha \leq 1$, the problem is **mildly** ill-posed.
- $\mathcal{O}\left(\frac{1}{j^\alpha}\right)$ with $\alpha > 1$, the problem is **moderately** ill-posed.
- $\mathcal{O}\left(e^{-\alpha j}\right)$ with $0 < \alpha$, the problem is **severely** ill-posed.

What Are They?

These problems arise due to ill-conditioning of G, *as opposed to a rank deficiency problem. Theoretically*, they are not ill-posed, like the Hilbert matrix. But practically speaking, they behave like ill-posed problems. Authors present a hierarchy of sorts for a problem with system $G\mathbf{m} = \mathbf{d}$. These order expressions are valid as $j \to \infty$.

- $\mathcal{O}\left(\frac{1}{j^\alpha}\right)$ with $0 < \alpha \leq 1$, the problem is **mildly** ill-posed.
- $\mathcal{O}\left(\frac{1}{j^\alpha}\right)$ with $\alpha > 1$, the problem is **moderately** ill-posed.
- $\mathcal{O}\left(e^{-\alpha j}\right)$ with $0 < \alpha$, the problem is **severely** ill-posed.

What Are They?

These problems arise due to ill-conditioning of G, *as opposed to a rank deficiency problem. Theoretically*, they are not ill-posed, like the Hilbert matrix. But practically speaking, they behave like ill-posed problems. Authors present a hierarchy of sorts for a problem with system $G\mathbf{m} = \mathbf{d}$. These order expressions are valid as $j \to \infty$.

- $\mathcal{O}\left(\frac{1}{j^\alpha}\right)$ with $0 < \alpha \leq 1$, the problem is **mildly** ill-posed.
- $\mathcal{O}\left(\frac{1}{j^\alpha}\right)$ with $\alpha > 1$, the problem is **moderately** ill-posed.
- $\mathcal{O}\left(e^{-\alpha j}\right)$ with $0 < \alpha$, the problem is **severely** ill-posed.

A Severly Ill-Posed Problem

The Shaw Problem:

An optics experiment is performed by dividing a circle using a vertical transversal with a slit in the middle. A variable intensity light source is placed around the left half of the circle and rays pass through the slit, where they are measured at points on the right half of the circle.

- Measure angles counterclockwise from the x-axis, using $-\pi/2 \leq \theta \leq \pi/2$ for the source intensity $m(\theta)$, and $-\pi/2 \leq s \leq \pi/2$ for destination intensity $d(s)$.
- The model for this problem comes from diffraction theory: $d(s) =$
$$\int_{-\pi/2}^{\pi/2} (\cos(s) + \cos(\theta))^2 \left(\frac{\sin(\pi(\sin(s) + \sin(\theta)))}{\pi(\sin(s) + \sin(\theta))} \right)^2 m(\theta)\, d\theta.$$

A Severly Ill-Posed Problem

The Shaw Problem:

An optics experiment is performed by dividing a circle using a vertical transversal with a slit in the middle. A variable intensity light source is placed around the left half of the circle and rays pass through the slit, where they are measured at points on the right half of the circle.

- Measure angles counterclockwise from the x-axis, using $-\pi/2 \le \theta \le \pi/2$ for the source intensity $m(\theta)$, and $-\pi/2 \le s \le \pi/2$ for destination intensity $d(s)$.
- The model for this problem comes from diffraction theory: $d(s) =$
$$\int_{-\pi/2}^{\pi/2} (\cos(s) + \cos(\theta))^2 \left(\frac{\sin(\pi(\sin(s) + \sin(\theta)))}{\pi(\sin(s) + \sin(\theta))} \right)^2 m(\theta)\, d\theta.$$

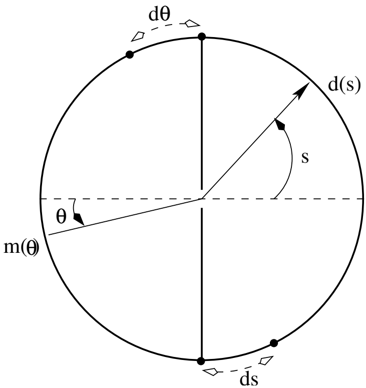
A Severly Ill-Posed Problem

The Shaw Problem:

An optics experiment is performed by dividing a circle using a vertical transversal with a slit in the middle. A variable intensity light source is placed around the left half of the circle and rays pass through the slit, where they are measured at points on the right half of the circle.

- Measure angles counterclockwise from the x-axis, using $-\pi/2 \leq \theta \leq \pi/2$ for the source intensity $m(\theta)$, and $-\pi/2 \leq s \leq \pi/2$ for destination intensity $d(s)$.
- The model for this problem comes from diffraction theory: $d(s) =$
$$\int_{-\pi/2}^{\pi/2} (\cos(s) + \cos(\theta))^2 \left(\frac{\sin(\pi(\sin(s) + \sin(\theta)))}{\pi(\sin(s) + \sin(\theta))} \right)^2 m(\theta)\, d\theta.$$

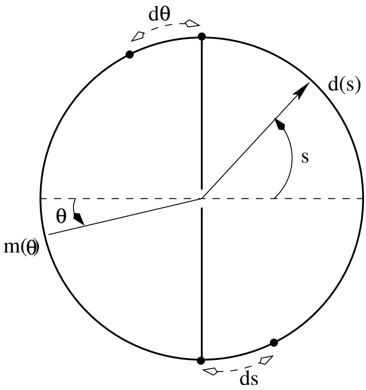
The Shaw Problem



Two Problems:

- The forward problem: given source intensity $m(\theta)$, compute the destination intensity $d(s)$.
- The inverse problem: given destination intensity $d(s)$, compute the source intensity $m(\theta)$.
- It can be shown that the inverse problem is severly ill-posed.

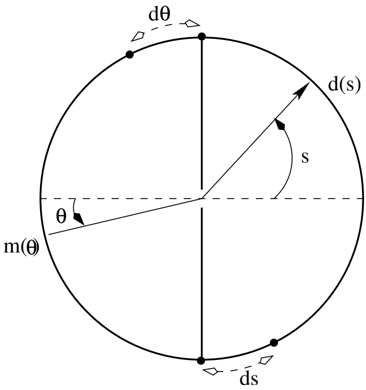
The Shaw Problem



Two Problems:

- The forward problem: given source intensity $m(\theta)$, compute the destination intensity $d(s)$.
- The inverse problem: given destination intensity $d(s)$, compute the source intensity $m(\theta)$.
- It can be shown that the inverse problem is severly ill-posed.

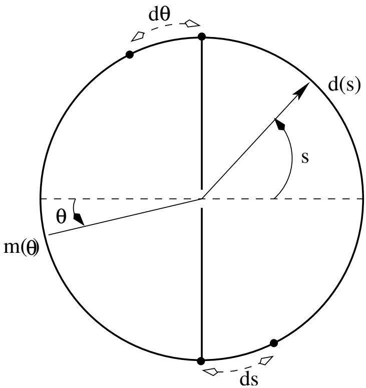
The Shaw Problem



Two Problems:

- The forward problem: given source intensity $m(\theta)$, compute the destination intensity $d(s)$.
- The inverse problem: given destination intensity $d(s)$, compute the source intensity $m(\theta)$.
- It can be shown that the inverse problem is severly ill-posed.

The Shaw Problem



Two Problems:

- The forward problem: given source intensity $m(\theta)$, compute the destination intensity $d(s)$.
- The inverse problem: given destination intensity $d(s)$, compute the source intensity $m(\theta)$.
- It can be shown that the inverse problem is severly ill-posed.

The Shaw Problem

How To Discretize The Problem:

- Discretize the parameter domain $-\pi/2 \leq \theta \leq \pi/2$ and the data domain $-\pi/2 \leq s \leq \pi/2$ into n subintervals of equal size $\Delta s = \Delta\theta = \pi/n$.
- Therefore, and let s_i, θ_i be the midpoints of the i-th subintervals:

$$s_i = \theta_i = -\frac{\pi}{2} + \frac{(i - 0.5)\,\pi}{n}, \; i = 1, 2, \ldots, n.$$

- Define

$$G_{i,j} = (\cos(s_i) + \cos(\theta_j))^2 \left(\frac{\sin(\pi(\sin(s_i) + \sin(\theta_j)))}{\pi(\sin(s_i) + \sin(\theta_j))} \right)^2 \Delta\theta$$

- Thus if $m_j \approx m(\theta_j)$, $d_i \approx d(s_i)$, $\mathbf{m} = (m_1, m_2, \ldots, m_n)$ and $\mathbf{d} = (d_1, d_2, \ldots, d_n)$, then discretization and the midpoint rule give $G\mathbf{m} = \mathbf{d}$, as in Chapter 3.

How To Discretize The Problem:

- Discretize the parameter domain $-\pi/2 \leq \theta \leq \pi/2$ and the data domain $-\pi/2 \leq s \leq \pi/2$ into n subintervals of equal size $\Delta s = \Delta\theta = \pi/n$.
- Therefore, and let s_i, θ_i be the midpoints of the i-th subintervals:

$$s_i = \theta_i = -\frac{\pi}{2} + \frac{(i - 0.5)\,\pi}{n},\ i = 1, 2, \ldots, n.$$

- Define

$$G_{i,j} = (\cos(s_i) + \cos(\theta_j))^2 \left(\frac{\sin(\pi(\sin(s_i) + \sin(\theta_j)))}{\pi(\sin(s_i) + \sin(\theta_j))} \right)^2 \Delta\theta$$

- Thus if $m_j \approx m(\theta_j)$, $d_i \approx d(s_i)$, $\mathbf{m} = (m_1, m_2, \ldots, m_n)$ and $\mathbf{d} = (d_1, d_2, \ldots, d_n)$, then discretization and the midpoint rule give $G\mathbf{m} = \mathbf{d}$, as in Chapter 3.

The Shaw Problem

How To Discretize The Problem:

- Discretize the parameter domain $-\pi/2 \leq \theta \leq \pi/2$ and the data domain $-\pi/2 \leq s \leq \pi/2$ into n subintervals of equal size $\Delta s = \Delta\theta = \pi/n$.
- Therefore, and let s_i, θ_i be the midpoints of the i-th subintervals:

$$s_i = \theta_i = -\frac{\pi}{2} + \frac{(i-0.5)\,\pi}{n}, \; i = 1, 2, \ldots, n.$$

- Define

$$G_{i,j} = (\cos(s_i) + \cos(\theta_j))^2 \left(\frac{\sin(\pi(\sin(s_i) + \sin(\theta_j)))}{\pi(\sin(s_i) + \sin(\theta_j))}\right)^2 \Delta\theta$$

- Thus if $m_j \approx m(\theta_j)$, $d_i \approx d(s_i)$, $\mathbf{m} = (m_1, m_2, \ldots, m_n)$ and $\mathbf{d} = (d_1, d_2, \ldots, d_n)$, then discretization and the midpoint rule give $G\mathbf{m} = \mathbf{d}$, as in Chapter 3.

How To Discretize The Problem:

- Discretize the parameter domain $-\pi/2 \leq \theta \leq \pi/2$ and the data domain $-\pi/2 \leq s \leq \pi/2$ into n subintervals of equal size $\Delta s = \Delta\theta = \pi/n$.
- Therefore, and let s_i, θ_i be the midpoints of the i-th subintervals:

$$s_i = \theta_i = -\frac{\pi}{2} + \frac{(i-0.5)\,\pi}{n}, \; i = 1, 2, \ldots, n.$$

- Define

$$G_{i,j} = \left(\cos\left(s_i\right) + \cos\left(\theta_j\right)\right)^2 \left(\frac{\sin\left(\pi\left(\sin\left(s_i\right) + \sin\left(\theta_j\right)\right)\right)}{\pi\left(\sin\left(s_i\right) + \sin\left(\theta_j\right)\right)}\right)^2 \Delta\theta$$

- Thus if $m_j \approx m\left(\theta_j\right)$, $d_i \approx d\left(s_i\right)$, $\mathbf{m} = \left(m_1, m_2, \ldots, m_n\right)$ and $\mathbf{d} = \left(d_1, d_2, \ldots, d_n\right)$, then discretization and the midpoint rule give $G\mathbf{m} = \mathbf{d}$, as in Chapter 3.

How To Discretize The Problem:

- Discretize the parameter domain $-\pi/2 \leq \theta \leq \pi/2$ and the data domain $-\pi/2 \leq s \leq \pi/2$ into n subintervals of equal size $\Delta s = \Delta\theta = \pi/n$.
- Therefore, and let s_i, θ_i be the midpoints of the i-th subintervals:

$$s_i = \theta_i = -\frac{\pi}{2} + \frac{(i - 0.5)\,\pi}{n},\; i = 1, 2, \ldots, n.$$

- Define

$$G_{i,j} = (\cos(s_i) + \cos(\theta_j))^2 \left(\frac{\sin(\pi(\sin(s_i) + \sin(\theta_j)))}{\pi(\sin(s_i) + \sin(\theta_j))}\right)^2 \Delta\theta$$

- Thus if $m_j \approx m(\theta_j)$, $d_i \approx d(s_i)$, $\mathbf{m} = (m_1, m_2, \ldots, m_n)$ and $\mathbf{d} = (d_1, d_2, \ldots, d_n)$, then discretization and the midpoint rule give $G\mathbf{m} = \mathbf{d}$, as in Chapter 3.

The Shaw Problem

Now we can examine the example files on the text CD for this problem. This file lives in 'MatlabTools/Examples/chap4/examp1'. First add the correctd path, then open the example file `examp.m` for editing. However, here's an easy way to build the matrix G without loops. Basically, these tools were designed to help with 3-D plotting.

```
> n = 20
> ds = pi/n
> s = linspace(ds/2, pi - ds/2,n)
> theta = s;
> [S, Theta] = meshgrid(s,theta);
>G = (cos(S) + cos(Theta)).^2 .* (sin(pi*(sin(S) + ...
sin(Theta)))./(pi*(sin(S) + sin(Theta))).^2*ds;
> % want to see G(s,θ)?
> mesh(S,Theta,G)
> cond(G)
> svd(G)
> rank(G)
```

Outline

Basics

Regularization:

This means "turn an ill-posed problem into a well-posed 'near by' problem". Most common method is Tikhonov regularization, which is motivated in context of our possibly ill-posed $G\mathbf{m} = \mathbf{d}$, i.e., minimize $\|G\mathbf{m} - \mathbf{d}\|_2$, problem by:

- Problem: minimize $\|\mathbf{m}\|_2$ subject to $\|\mathbf{Gm} - \mathbf{d}\|_2 \le \delta$
- Problem: minimize $\|G\mathbf{m} - \mathbf{d}\|_2$ subject to $\|\mathbf{m}\|_2 \le \epsilon$
- Problem: (**damped least squares**) minimize $\|G\mathbf{m} - \mathbf{d}\|_2^2 + \alpha^2 \|\mathbf{m}\|_2^2$. This is the **Tikhonov regularization** of the original problem.
- Problem: find minima of $f(\mathbf{x})$ subject to constraint $g(\mathbf{x}) \le c$.e function $L = f(\mathbf{x}) + \lambda g(\mathbf{x})$, for some $\lambda \ge 0$.

Basics

Regularization:

This means "turn an ill-posed problem into a well-posed 'near by' problem". Most common method is Tikhonov regularization, which is motivated in context of our possibly ill-posed $G\mathbf{m} = \mathbf{d}$, i.e., minimize $\|G\mathbf{m} - \mathbf{d}\|_2$, problem by:

- Problem: minimize $\|\mathbf{m}\|_2$ subject to $\|\mathbf{Gm} - \mathbf{d}\|_2 \leq \delta$
- Problem: minimize $\|G\mathbf{m} - \mathbf{d}\|_2$ subject to $\|\mathbf{m}\|_2 \leq \epsilon$
- Problem: (**damped least squares**) minimize $\|G\mathbf{m} - \mathbf{d}\|_2^2 + \alpha^2 \|\mathbf{m}\|_2^2$. This is the **Tikhonov regularization** of the original problem.
- Problem: find minima of $f(\mathbf{x})$ subject to constraint $g(\mathbf{x}) \leq c$.e function $L = f(\mathbf{x}) + \lambda g(\mathbf{x})$, for some $\lambda \geq 0$.

Basics

Regularization:

This means "turn an ill-posed problem into a well-posed 'near by' problem". Most common method is Tikhonov regularization, which is motivated in context of our possibly ill-posed $G\mathbf{m} = \mathbf{d}$, i.e., minimize $\|G\mathbf{m} - \mathbf{d}\|_2$, problem by:

- Problem: minimize $\|\mathbf{m}\|_2$ subject to $\|\mathbf{Gm} - \mathbf{d}\|_2 \leq \delta$
- Problem: minimize $\|G\mathbf{m} - \mathbf{d}\|_2$ subject to $\|\mathbf{m}\|_2 \leq \epsilon$
- Problem: (**damped least squares**) minimize $\|G\mathbf{m} - \mathbf{d}\|_2^2 + \alpha^2 \|\mathbf{m}\|_2^2$. This is the **Tikhonov regularization** of the original problem.
- Problem: find minima of $f(\mathbf{x})$ subject to constraint $g(\mathbf{x}) \leq c$.e function $L = f(\mathbf{x}) + \lambda g(\mathbf{x})$, for some $\lambda \geq 0$.

Basics

Regularization:

This means "turn an ill-posed problem into a well-posed 'near by' problem". Most common method is Tikhonov regularization, which is motivated in context of our possibly ill-posed $G\mathbf{m} = \mathbf{d}$, i.e., minimize $\|G\mathbf{m} - \mathbf{d}\|_2$, problem by:

- Problem: minimize $\|\mathbf{m}\|_2$ subject to $\|\mathbf{Gm} - \mathbf{d}\|_\mathbf{2} \leq \delta$
- Problem: minimize $\|G\mathbf{m} - \mathbf{d}\|_2$ subject to $\|\mathbf{m}\|_2 \leq \epsilon$
- Problem: **(damped least squares)** minimize $\|G\mathbf{m} - \mathbf{d}\|_2^2 + \alpha^2 \|\mathbf{m}\|_2^2$. This is the **Tikhonov regularization** of the original problem.
- Problem: find minima of $f(\mathbf{x})$ subject to constraint $g(\mathbf{x}) \leq c$.e function $L = f(\mathbf{x}) + \lambda g(\mathbf{x})$, for some $\lambda \geq 0$.

Basics

Regularization:

This means "turn an ill-posed problem into a well-posed 'near by' problem". Most common method is Tikhonov regularization, which is motivated in context of our possibly ill-posed $G\mathbf{m} = \mathbf{d}$, i.e., minimize $\|G\mathbf{m} - \mathbf{d}\|_2$, problem by:

- Problem: minimize $\|\mathbf{m}\|_2$ subject to $\|\mathbf{Gm} - \mathbf{d}\|_\mathbf{2} \leq \delta$
- Problem: minimize $\|G\mathbf{m} - \mathbf{d}\|_2$ subject to $\|\mathbf{m}\|_2 \leq \epsilon$
- Problem: (**damped least squares**) minimize $\|G\mathbf{m} - \mathbf{d}\|_2^2 + \alpha^2 \|\mathbf{m}\|_2^2$. This is the **Tikhonov regularization** of the original problem.
- Problem: find minima of $f(\mathbf{x})$ subject to constraint $g(\mathbf{x}) \leq c$.e function $L = f(\mathbf{x}) + \lambda g(\mathbf{x})$, for some $\lambda \geq 0$.

Regularization:

All of the above problems are equivalent under mild restrictions thanks to the principle of Lagrange multipliers:

- The minima of $f(\mathbf{x})$ subject to constraint $g(\mathbf{x}) \leq c$ must occur at the stationary points of function $L = f(\mathbf{x}) + \lambda g(\mathbf{x})$, for some $\lambda \geq 0$ (so we could write $\lambda = \alpha^2$ to emphasize non-negativity.)
- We can see why this is true in the case of a two dimensional $\mathbf{x}$ by examining contour curves.
- Square the terms in the first two problems and we see that the associated Lagrangians are related if we take reciprocals of α.
- Various values of α give a trade-off between the instability of the unmodified least squares problem and loss of accuracy of the smoothed problem. This can be understood by tracking the value of the minimized function in the form of a path depending on δ, ϵ or α.

Regularization:

All of the above problems are equivalent under mild restrictions thanks to the principle of Lagrange multipliers:

- The minima of $f(\mathbf{x})$ subject to constraint $g(\mathbf{x}) \leq c$ must occur at the stationary points of function $L = f(\mathbf{x}) + \lambda g(\mathbf{x})$, for some $\lambda \geq 0$ (so we could write $\lambda = \alpha^2$ to emphasize non-negativity.)
- We can see why this is true in the case of a two dimensional $\mathbf{x}$ by examining contour curves.
- Square the terms in the first two problems and we see that the associated Lagrangians are related if we take reciprocals of α.
- Various values of α give a trade-off between the instability of the unmodified least squares problem and loss of accuracy of the smoothed problem. This can be understood by tracking the value of the minimized function in the form of a path depending on δ, ϵ or α.

Basics

Regularization:

All of the above problems are equivalent under mild restrictions thanks to the principle of Lagrange multipliers:

- The minima of $f(\mathbf{x})$ subject to constraint $g(\mathbf{x}) \leq c$ must occur at the stationary points of function $L = f(\mathbf{x}) + \lambda g(\mathbf{x})$, for some $\lambda \geq 0$ (so we could write $\lambda = \alpha^2$ to emphasize non-negativity.)
- We can see why this is true in the case of a two dimensional $\mathbf{x}$ by examining contour curves.
- Square the terms in the first two problems and we see that the associated Lagrangians are related if we take reciprocals of α.
- Various values of α give a trade-off between the instability of the unmodified least squares problem and loss of accuracy of the smoothed problem. This can be understood by tracking the value of the minimized function in the form of a path depending on δ, ϵ or α.

Basics

Regularization:

All of the above problems are equivalent under mild restrictions thanks to the principle of Lagrange multipliers:

- The minima of $f(\mathbf{x})$ subject to constraint $g(\mathbf{x}) \leq c$ must occur at the stationary points of function $L = f(\mathbf{x}) + \lambda g(\mathbf{x})$, for some $\lambda \geq 0$ (so we could write $\lambda = \alpha^2$ to emphasize non-negativity.)
- We can see why this is true in the case of a two dimensional $\mathbf{x}$ by examining contour curves.
- Square the terms in the first two problems and we see that the associated Lagrangians are related if we take reciprocals of α.
- Various values of α give a trade-off between the instability of the unmodified least squares problem and loss of accuracy of the smoothed problem. This can be understood by tracking the value of the minimized function in the form of a path depending on δ, ϵ or α.

Basics

Regularization:

All of the above problems are equivalent under mild restrictions thanks to the principle of Lagrange multipliers:

- The minima of $f(\mathbf{x})$ subject to constraint $g(\mathbf{x}) \leq c$ must occur at the stationary points of function $L = f(\mathbf{x}) + \lambda g(\mathbf{x})$, for some $\lambda \geq 0$ (so we could write $\lambda = \alpha^2$ to emphasize non-negativity.)
- We can see why this is true in the case of a two dimensional $\mathbf{x}$ by examining contour curves.
- Square the terms in the first two problems and we see that the associated Lagrangians are related if we take reciprocals of α.
- Various values of α give a trade-off between the instability of the unmodified least squares problem and loss of accuracy of the smoothed problem. This can be understood by tracking the value of the minimized function in the form of a path depending on δ, ϵ or α.

Outline

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:

- $\nabla\left(\|G\mathbf{m}-\mathbf{d}\|_2^2+\alpha^2\|\mathbf{m}\|_2^2\right)=\left(G^TG\mathbf{m}-G^T\mathbf{d}\right)+\alpha^2\mathbf{m}$
- 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:

- $\nabla\left(\|G\mathbf{m}-\mathbf{d}\|_2^2+\alpha^2\|\mathbf{m}\|_2^2\right)=\left(G^TG\mathbf{m}-G^T\mathbf{d}\right)+\alpha^2\mathbf{m}$
- 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

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

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

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

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**.
- $f(\sigma) = \sigma$ simply uses the original singular values.
- $f(\sigma) = \dfrac{\sigma}{\sigma^2 + \alpha^2}$ is the Tikhonov filter we have just developed.
- $f(\sigma) = \max\{\operatorname{sgn}(\sigma - \epsilon)\,\sigma, 0\}$ 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**.
- $f(\sigma) = \sigma$ simply uses the original singular values.
- $f(\sigma) = \dfrac{\sigma}{\sigma^2 + \alpha^2}$ is the Tikhonov filter we have just developed.
- $f(\sigma) = \max\{\operatorname{sgn}(\sigma - \epsilon)\sigma, 0\}$ 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**.
- $f(\sigma) = \sigma$ simply uses the original singular values.
- $f(\sigma) = \dfrac{\sigma}{\sigma^2 + \alpha^2}$ is the Tikhonov filter we have just developed.
- $f(\sigma) = \max\{\operatorname{sgn}(\sigma - \epsilon)\,\sigma, 0\}$ 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**.
- $f(\sigma) = \sigma$ simply uses the original singular values.
- $f(\sigma) = \dfrac{\sigma}{\sigma^2 + \alpha^2}$ is the Tikhonov filter we have just developed.
- $f(\sigma) = \max\{\operatorname{sgn}(\sigma - \epsilon)\sigma, 0\}$ 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**.
- $f(\sigma) = \sigma$ simply uses the original singular values.
- $f(\sigma) = \dfrac{\sigma}{\sigma^2 + \alpha^2}$ is the Tikhonov filter we have just developed.
- $f(\sigma) = \max\{\operatorname{sgn}(\sigma - \epsilon)\,\sigma, 0\}$ 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 $(\|\mathbf{m}_\alpha\|_2, \|G\mathbf{m}_\alpha - \mathbf{d}\|_2)$
- 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 $\|\delta\mathbf{d}\|_2$.

Explore the Example 5.1 file, which constructs the L-curve of the Shaw problem using tools from the Regularization Toolbox.

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 $\|\delta\mathbf{d}\|_2$.

Explore the Example 5.1 file, which constructs the L-curve of the Shaw problem using tools from the Regularization Toolbox.

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 $\|\delta\mathbf{d}\|_2$.

Explore the Example 5.1 file, which constructs the L-curve of the Shaw problem using tools from the Regularization Toolbox.

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 $\|\delta\mathbf{d}\|_2$.

Explore the Example 5.1 file, which constructs the L-curve of the Shaw problem using tools from the Regularization Toolbox.

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 $\|\delta\mathbf{d}\|_2$.

Explore the Example 5.1 file, which constructs the L-curve of the Shaw problem using tools from the Regularization Toolbox.

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 $\|\delta\mathbf{d}\|_2$.

Explore the Example 5.1 file, which constructs the L-curve of the Shaw problem using tools from the Regularization Toolbox.

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.)

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 $K : H_1 \rightarrow H_2$ 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_α.
- Given that $\delta = \|\delta d\|$ is the noise level, Tikhonov defines a **regular algorithm** to be a choice $\alpha = \alpha(\delta)$ such that

$$\alpha(\delta) \to 0 \text{ and } m_{\alpha(\delta)} \to 0 \text{ as } \delta \to 0.$$

- Morozov's discrepancy principle is a regular algorithm.

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, Tikhonov defines a **regular algorithm** to be a choice $\alpha = \alpha(\delta)$ such that

$$\alpha(\delta) \to 0 \text{ and } m_{\alpha(\delta)} \to 0 \text{ as } \delta \to 0.$$

- Morozov's discrepancy principle is a regular algorithm.

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, Tikhonov defines a **regular algorithm** to be a choice $\alpha = \alpha(\delta)$ such that

$$\alpha(\delta) \rightarrow 0 \text{ and } m_{\alpha(\delta)} \rightarrow 0 \text{ as } \delta \rightarrow 0.$$

- Morozov's discrepancy principle is a regular algorithm.

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, Tikhonov defines a **regular algorithm** to be a choice $\alpha = \alpha(\delta)$ such that

$$\alpha(\delta) \to 0 \text{ and } m_{\alpha(\delta)} \to 0 \text{ as } \delta \to 0.$$

- Morozov's discrepancy principle is a regular algorithm.