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

Instructor: Thomas Shores
Department of Mathematics

Lecture 22, April 4, 2006
AvH 10

# Outline

# 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 $L = I$?

# 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 $L = I$?

# 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 $L = I$?

# 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 $L = I$?

# 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 $L = I$?

# Example Matrices

We will explore approximations to first and second derivatives at the board.

# Key Idea: Generalized SVD (GSVD)

**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*

$$\begin{aligned} U^T G X &= \operatorname{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_n\} = \Lambda = \Lambda_{m,n} \\ V^T L X &= \operatorname{diag}\{\mu_1, \mu_2, \ldots, \mu_q\} = M = M_{p,n} \\ \Lambda^T \Lambda + M^T M &= 1. \end{aligned}$$

*Also $0 \leq \lambda_1 \leq \lambda_2 \cdots \leq \lambda_n \leq 1$ and $1 \geq \mu_1 \geq \mu_2 \cdots \geq \mu_q \geq 0$. The numbers $\gamma_i = \lambda_i / \mu_i$, $i = 1, \ldots, \operatorname{rank}(L) \equiv r$ are called the* ***generalized singular values*** *of $G$ and $L$ and $0 \leq \gamma_1 \leq \gamma_2 \cdots \leq \gamma_r$.*

# Application to Higher Order Regularization

The minimization problem is equivalent to the problem

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

which has solution forms

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

Filter factors: $f_j = \dfrac{\gamma_j^2}{\gamma_j^2 + \alpha^2}$, $j = 1, \ldots, p$, $f_j = 1$, $j = p + 1, \ldots, n$.

Thus

$$\mathbf{m}_{\alpha,L} = \sum_{j=1}^{n} f_j \frac{\left(\mathbf{U}_j^T \mathbf{d}\right)}{\lambda_j} \mathbf{X}_j.$$

# Vertical Seismic Profiling Example

## The Experiment:

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

- Generate a seizmic 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?

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 seizmic 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?

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 seizmic 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?

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 seizmic 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?

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 seizmic 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?

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 seizmic 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?

Do Example 4-5 from the CD.

# Model Resolution

## Model Resolution Matrix:

As usual, $R_{\mathbf{m},\alpha,L} = G^{\natural}G$.

- Comment 1.
- Comment 2.
- Comment 3.

# Model Resolution

## Model Resolution Matrix:

As usual, $R_{\mathbf{m},\alpha,L} = G^{\natural}G$.

- Comment 1.
- Comment 2.
- Comment 3.

# Model Resolution

## Model Resolution Matrix:

As usual, $R_{\mathbf{m},\alpha,L} = G^{\natural}G$.

- Comment 1.
- Comment 2.
- Comment 3.

# Model Resolution

**Model Resolution Matrix:**

As usual, $R_{\mathbf{m},\alpha,L} = G^{\natural} G$.

- Comment 1.
- Comment 2.
- Comment 3.

# Outline

## TGSVD:

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

- Formula becomes

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

- Key question: where to start $k$.

# 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

$$V_0(\alpha) = \frac{m\left\|G\mathbf{m}_\alpha - \mathbf{d}\right\|_2}{\mathrm{Tr}\left(I - GG^\natural\right)^2}$$

Example 5.6-7 gives a nice illustration of the ideas. Use the CD

# 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

$$V_0(\alpha) = \frac{m\left\|G\mathbf{m}_\alpha - \mathbf{d}\right\|_2}{\mathrm{Tr}\left(I - GG^\natural\right)^2}$$

Example 5.6-7 gives a nice illustration of the ideas. Use the CD

# 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

$$V_0(\alpha) = \frac{m \left\| G\mathbf{m}_\alpha - \mathbf{d} \right\|_2}{\operatorname{Tr}\left(I - GG^\natural\right)^2}$$

Example 5.6-7 gives a nice illustration of the ideas. Use the CD

# 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

$$V_0(\alpha) = \frac{m\left\|G\mathbf{m}_\alpha - \mathbf{d}\right\|_2}{\operatorname{Tr}\left(I - GG^\natural\right)^2}$$

Example 5.6-7 gives a nice illustration of the ideas. Use the CD

# Outline

# 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

$$\frac{\|\mathbf{m}_\alpha - \widetilde{\mathbf{m}}_\alpha\|_2}{\|\mathbf{m}_\alpha\|_2} \leq \kappa_\alpha \frac{\left\|\mathbf{d} - \widetilde{\mathbf{d}}\right\|_2}{\|G\mathbf{m}_\alpha\|_2}$$

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

$$\frac{\|\mathbf{m}_\alpha - \widetilde{\mathbf{m}}_\alpha\|_2}{\|\mathbf{m}_\alpha\|_2} \leq \kappa_\alpha \frac{\left\|\mathbf{d} - \widetilde{\mathbf{d}}\right\|_2}{\|G\mathbf{m}_\alpha\|_2}$$

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

# Error Bounds

## More Estimates:

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

$$\left\|\mathbf{m}_{true} - G^{\natural}\mathbf{d}\right\|_2 = \gamma\,(p+1)\,\widehat{\alpha}^p = \mathcal{O}\left(\Delta^{\frac{p}{p+1}}\right).$$

- 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.

As time permits, do Example 5.8 from CD.

# Error Bounds

## More Estimates:

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

$$\left\|\mathbf{m}_{true} - G^{\natural}\mathbf{d}\right\|_2 = \gamma\,(p+1)\,\widehat{\alpha}^p = \mathcal{O}\left(\Delta^{\frac{p}{p+1}}\right).$$

- 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.

As time permits, do Example 5.8 from CD.

# Error Bounds

## More Estimates:

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

$$\left\|\mathbf{m}_{true} - G^{\natural}\mathbf{d}\right\|_2 = \gamma\,(p+1)\,\widehat{\alpha}^p = \mathcal{O}\left(\Delta^{\frac{p}{p+1}}\right).$$

- 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.

As time permits, do Example 5.8 from CD.

# Image Recovery

## Problem:

An image is blurred and we want to sharpen it. Let intensity function $I_{true}(x, y)$define the true image and $I_{blurred}(x, y)$ define the blurred image.

- A typical model results from convolving true image with Gaussian point spread function

$$I_{blurred}(x, y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I_{true}(x - u, y - v)\,\Psi(u, v)\;du\,dv$$

  where $\Psi(u, v) = e^{-\left(u^2+v^2\right)/\left(2\sigma^2\right)}$.
- Think about discretizing this over an SVGA image $(1024 \times 768)$.
- But the discretized matrix should be sparse!

# Image Recovery

## Problem:

An image is blurred and we want to sharpen it. Let intensity function $I_{true}(x, y)$define the true image and $I_{blurred}(x, y)$ define the blurred image.

- A typical model results from convolving true image with Gaussian point spread function

$$I_{blurred}(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I_{true}(x - u, y - v) \Psi(u, v)\, du\, dv$$

where $\Psi(u, v) = e^{-\left(u^2+v^2\right)/\left(2\sigma^2\right)}$.

- Think about discretizing this over an SVGA image $(1024 \times 768)$.
- But the discretized matrix should be sparse!

# Image Recovery

## Problem:

An image is blurred and we want to sharpen it. Let intensity function $I_{true}(x, y)$define the true image and $I_{blurred}(x, y)$ define the blurred image.

- A typical model results from convolving true image with Gaussian point spread function

$$I_{blurred}(x, y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I_{true}(x - u, y - v)\,\Psi(u, v)\,du\,dv$$

  where $\Psi(u, v) = e^{-(u^2+v^2)/(2\sigma^2)}$.

- Think about discretizing this over an SVGA image $(1024 \times 768)$.
- But the discretized matrix should be sparse!

# Image Recovery

## Problem:

An image is blurred and we want to sharpen it. Let intensity function $I_{true}(x, y)$define the true image and $I_{blurred}(x, y)$ define the blurred image.

- A typical model results from convolving true image with Gaussian point spread function

$$I_{blurred}(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I_{true}(x - u, y - v) \Psi(u, v) \, du \, dv$$

  where $\Psi(u, v) = e^{-\left(u^2+v^2\right)/\left(2\sigma^2\right)}$.
- Think about discretizing this over an SVGA image $(1024 \times 768)$.
- But the discretized matrix should be sparse!

# Sparse Matrices and Iterative Methods

## Sparse Matrix:

A matrix with sufficiently many zeros that we should pay attention to them.

- There are efficient ways of storing such matrices and doing linear algebra on them.
- Given a problem $A\mathbf{x} = \mathbf{b}$ with $A$ sparse, iterative methods become attractive because they usually only require storage of $A$, $\mathbf{x}$ and some auxillary vectors, and saxpy, gaxpy, dot algorithms – ("scalar a*x+y" , "general A*x+y", "dot product")
- Classical methods: Jacobi, Gauss-Seidel, Gauss-Seidel SOR and conjugate gradient.
- Methods especially useful for tomographic problems: Kaczmarz's method, ART (algebraic reconstruction technique).

# Sparse Matrices and Iterative Methods

## Sparse Matrix:

A matrix with sufficiently many zeros that we should pay attention to them.

- There are efficient ways of storing such matrices and doing linear algebra on them.
- Given a problem $A\mathbf{x} = \mathbf{b}$ with $A$ sparse, iterative methods become attractive because they usually only require storage of $A$, $\mathbf{x}$ and some auxillary vectors, and saxpy, gaxpy, dot algorithms – ("scalar a*x+y" , "general A*x+y", "dot product")
- Classical methods: Jacobi, Gauss-Seidel, Gauss-Seidel SOR and conjugate gradient.
- Methods especially useful for tomographic problems: Kaczmarz's method, ART (algebraic reconstruction technique).

# Sparse Matrices and Iterative Methods

## Sparse Matrix:

A matrix with sufficiently many zeros that we should pay attention to them.

- There are efficient ways of storing such matrices and doing linear algebra on them.
- Given a problem $A\mathbf{x} = \mathbf{b}$ with $A$ sparse, iterative methods become attractive because they usually only require storage of $A$, $\mathbf{x}$ and some auxillary vectors, and saxpy, gaxpy, dot algorithms – ("scalar a*x+y" , "general A*x+y", "dot product")
- Classical methods: Jacobi, Gauss-Seidel, Gauss-Seidel SOR and conjugate gradient.
- Methods especially useful for tomographic problems: Kaczmarz's method, ART (algebraic reconstruction technique).

# Sparse Matrices and Iterative Methods

## Sparse Matrix:

A matrix with sufficiently many zeros that we should pay attention to them.

- There are efficient ways of storing such matrices and doing linear algebra on them.
- Given a problem $A\mathbf{x} = \mathbf{b}$ with $A$ sparse, iterative methods become attractive because they usually only require storage of $A$, $\mathbf{x}$ and some auxillary vectors, and saxpy, gaxpy, dot algorithms – ("scalar a*x+y" , "general A*x+y", "dot product")
- Classical methods: Jacobi, Gauss-Seidel, Gauss-Seidel SOR and conjugate gradient.
- Methods especially useful for tomographic problems: Kaczmarz's method, ART (algebraic reconstruction technique).

# Sparse Matrices and Iterative Methods

## Sparse Matrix:

A matrix with sufficiently many zeros that we should pay attention to them.

- There are efficient ways of storing such matrices and doing linear algebra on them.
- Given a problem $A\mathbf{x} = \mathbf{b}$ with $A$ sparse, iterative methods become attractive because they usually only require storage of $A$, $\mathbf{x}$ and some auxillary vectors, and saxpy, gaxpy, dot algorithms – ("scalar a*x+y" , "general A*x+y", "dot product")
- Classical methods: Jacobi, Gauss-Seidel, Gauss-Seidel SOR and conjugate gradient.
- Methods especially useful for tomographic problems: Kaczmarz's method, ART (algebraic reconstruction technique).

# Yet Another Regularization Idea

## To regularize in face of iteration:

Use the number of iteration steps taken as a regularization parameter.

- Conjugate gradient methods are designed to work with SPD coefficient matrices $A$ in the equation $A\mathbf{x} = \mathbf{b}$.
- So in the unregularized least squares problem $G^T G\mathbf{m} = G^T\mathbf{d}$ take $A = G^T G$ and $\mathbf{b} = G^T\mathbf{d}$, resulting in the CGLS method, in which we avoid explicitly computing $G^T G$.
- Key fact: in exact arithmetic, if we start at $\mathbf{m}^{(0)} = \mathbf{0}$, then $\left\|\mathbf{m}^{(k)}\right\|$ is monotone increasing in $k$ and $\left\|G\mathbf{m}^{(k)} - \mathbf{d}\right\|$ is monotonically decreasing in $k$. So we can make an L-curve in terms of $k$.

Do Example 6.3 from the CD.

# Yet Another Regularization Idea

## To regularize in face of iteration:

Use the number of iteration steps taken as a regularization parameter.

- Conjugate gradient methods are designed to work with SPD coefficient matrices $A$ in the equation $A\mathbf{x} = \mathbf{b}$.
- So in the unregularized least squares problem $G^T G\mathbf{m} = G^T\mathbf{d}$ take $A = G^T G$ and $\mathbf{b} = G^T\mathbf{d}$, resulting in the CGLS method, in which we avoid explicitly computing $G^T G$.
- Key fact: in exact arithmetic, if we start at $\mathbf{m}^{(0)} = \mathbf{0}$, then $\left\|\mathbf{m}^{(k)}\right\|$ is monotone increasing in $k$ and $\left\|G\mathbf{m}^{(k)} - \mathbf{d}\right\|$ is monotonically decreasing in $k$. So we can make an L-curve in terms of $k$.

Do Example 6.3 from the CD.

# Yet Another Regularization Idea

**To regularize in face of iteration:**

Use the number of iteration steps taken as a regularization parameter.

- Conjugate gradient methods are designed to work with SPD coefficient matrices $A$ in the equation $A\mathbf{x} = \mathbf{b}$.
- So in the unregularized least squares problem $G^T G\mathbf{m} = G^T\mathbf{d}$ take $A = G^T G$ and $\mathbf{b} = G^T\mathbf{d}$, resulting in the CGLS method, in which we avoid explicitly computing $G^T G$.
- Key fact: in exact arithmetic, if we start at $\mathbf{m}^{(0)} = \mathbf{0}$, then $\left\|\mathbf{m}^{(k)}\right\|$ is monotone increasing in $k$ and $\left\|G\mathbf{m}^{(k)} - \mathbf{d}\right\|$ is monotonically decreasing in $k$. So we can make an L-curve in terms of $k$.

Do Example 6.3 from the CD.

# Yet Another Regularization Idea

## To regularize in face of iteration:

Use the number of iteration steps taken as a regularization parameter.

- Conjugate gradient methods are designed to work with SPD coefficient matrices $A$ in the equation $A\mathbf{x} = \mathbf{b}$.
- So in the unregularized least squares problem $G^T G\mathbf{m} = G^T\mathbf{d}$ take $A = G^T G$ and $\mathbf{b} = G^T\mathbf{d}$, resulting in the CGLS method, in which we avoid explicitly computing $G^T G$.
- Key fact: in exact arithmetic, if we start at $\mathbf{m}^{(0)} = \mathbf{0}$, then $\left\|\mathbf{m}^{(k)}\right\|$ is monotone increasing in $k$ and $\left\|G\mathbf{m}^{(k)} - \mathbf{d}\right\|$ is monotonically decreasing in $k$. So we can make an L-curve in terms of $k$.

Do Example 6.3 from the CD.