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

Instructor: Thomas Shores
Department of Mathematics

Lecture 22, April 4, 2006
AvH 10

Total Variation

Key Property:

- TV doesn't smooth discontinuities as much as Tikhonov regularization.

Change startupfile path to Examples/chap7/examp3 execute it and examp.

Total Variation

Key Property:

- TV doesn't smooth discontinuities as much as Tikhonov regularization.

Change startupfile path to Examples/chap7/examp3 execute it and examp.

Outline

Statistics

Problem is $G(\mathbf{m}) = \mathbf{d}$ with least squares solution $\mathbf{m}^*$:

Now what? What statistics can we bring to bear on the problem?

- We minimize $\|\mathbf{F}(\mathbf{m})\|^2 = \sum_{i=1}^{n} \frac{(G(\mathbf{m}) - d_i)^2}{\sigma_i^2}$
- Treat the linear model as locally accurate, so misfit is $\nabla\mathbf{F} = \mathbf{F}(\mathbf{m} + \Delta\mathbf{m}) - \mathbf{F}(\mathbf{m}^*) \approx \nabla\mathbf{F}(\mathbf{m}^*)\nabla\mathbf{m}$
- Obtain covariance matrix $\mathrm{Cov}(\mathbf{m}^*) = \left(\nabla\mathbf{F}(\mathbf{m}^*)^T \nabla\mathbf{F}(\mathbf{m}^*)\right)^{-1}$
- If σ is unknown but constant across measurements, take $\sigma_i = 1$ above and use for σ in $\frac{1}{\sigma^2}\left(\nabla\mathbf{F}(\mathbf{m}^*)^T \nabla\mathbf{F}(\mathbf{m}^*)\right)^{-1}$ the estimate
$$s^2 = \frac{1}{m-n}\sum_{i=1}^{m}(G(\mathbf{m}) - d_i)^2 .$$
- Do confidence intervals, χ^2 statistic and p-value as in Chapter 2.

Statistics

Problem is $G(\mathbf{m}) = \mathbf{d}$ with least squares solution $\mathbf{m}^*$:

Now what? What statistics can we bring to bear on the problem?

- We minimize $\|\mathbf{F}(\mathbf{m})\|^2 = \sum_{i=1}^{n} \frac{(G(\mathbf{m}) - d_i)^2}{\sigma_i^2}$
- Treat the linear model as locally accurate, so misfit is $\nabla \mathbf{F} = \mathbf{F}(\mathbf{m} + \Delta \mathbf{m}) - \mathbf{F}(\mathbf{m}^*) \approx \nabla \mathbf{F}(\mathbf{m}^*) \nabla \mathbf{m}$
- Obtain covariance matrix $\mathrm{Cov}(\mathbf{m}^*) = \left(\nabla \mathbf{F}(\mathbf{m}^*)^T \nabla \mathbf{F}(\mathbf{m}^*)\right)^{-1}$
- If σ is unknown but constant across measurements, take $\sigma_i = 1$ above and use for σ in $\frac{1}{\sigma^2}\left(\nabla \mathbf{F}(\mathbf{m}^*)^T \nabla \mathbf{F}(\mathbf{m}^*)\right)^{-1}$ the estimate
$$s^2 = \frac{1}{m-n} \sum_{i=1}^{m} (G(\mathbf{m}) - d_i)^2 .$$
- Do confidence intervals, χ^2 statistic and p-value as in Chapter 2.

Statistics

Problem is $G(\mathbf{m}) = \mathbf{d}$ with least squares solution $\mathbf{m}^*$:

Now what? What statistics can we bring to bear on the problem?

- We minimize $\|\mathbf{F}(\mathbf{m})\|^2 = \sum_{i=1}^{n} \frac{(G(\mathbf{m}) - d_i)^2}{\sigma_i^2}$
- Treat the linear model as locally accurate, so misfit is $\nabla\mathbf{F} = \mathbf{F}(\mathbf{m} + \Delta\mathbf{m}) - \mathbf{F}(\mathbf{m}^*) \approx \nabla\mathbf{F}(\mathbf{m}^*)\nabla\mathbf{m}$
- Obtain covariance matrix $\mathrm{Cov}(\mathbf{m}^*) = \left(\nabla\mathbf{F}(\mathbf{m}^*)^T \nabla\mathbf{F}(\mathbf{m}^*)\right)^{-1}$
- If σ is unknown but constant across measurements, take $\sigma_i = 1$ above and use for σ in $\frac{1}{\sigma^2}\left(\nabla\mathbf{F}(\mathbf{m}^*)^T \nabla\mathbf{F}(\mathbf{m}^*)\right)^{-1}$ the estimate
$$s^2 = \frac{1}{m-n}\sum_{i=1}^{m}(G(\mathbf{m}) - d_i)^2.$$
- Do confidence intervals, χ^2 statistic and p-value as in Chapter 2.

Problem is $G(\mathbf{m}) = \mathbf{d}$ with least squares solution $\mathbf{m}^*$:

Now what? What statistics can we bring to bear on the problem?

- We minimize $\|\mathbf{F}(\mathbf{m})\|^2 = \sum_{i=1}^{n} \frac{(G(\mathbf{m}) - d_i)^2}{\sigma_i^2}$
- Treat the linear model as locally accurate, so misfit is $\nabla \mathbf{F} = \mathbf{F}(\mathbf{m} + \Delta \mathbf{m}) - \mathbf{F}(\mathbf{m}^*) \approx \nabla \mathbf{F}(\mathbf{m}^*) \nabla \mathbf{m}$
- Obtain covariance matrix $\mathrm{Cov}(\mathbf{m}^*) = \left(\nabla \mathbf{F}(\mathbf{m}^*)^T \nabla \mathbf{F}(\mathbf{m}^*)\right)^{-1}$
- If σ is unknown but constant across measurements, take $\sigma_i = 1$ above and use for σ in $\frac{1}{\sigma^2}\left(\nabla \mathbf{F}(\mathbf{m}^*)^T \nabla \mathbf{F}(\mathbf{m}^*)\right)^{-1}$ the estimate
$$s^2 = \frac{1}{m-n} \sum_{i=1}^{m} (G(\mathbf{m}) - d_i)^2 .$$
- Do confidence intervals, χ^2 statistic and p-value as in Chapter 2.

Statistics

Problem is $G(\mathbf{m}) = \mathbf{d}$ with least squares solution $\mathbf{m}^*$:

Now what? What statistics can we bring to bear on the problem?

- We minimize $\|\mathbf{F}(\mathbf{m})\|^2 = \sum_{i=1}^{n} \frac{(G(\mathbf{m}) - d_i)^2}{\sigma_i^2}$
- Treat the linear model as locally accurate, so misfit is $\nabla \mathbf{F} = \mathbf{F}(\mathbf{m} + \Delta\mathbf{m}) - \mathbf{F}(\mathbf{m}^*) \approx \nabla \mathbf{F}(\mathbf{m}^*) \nabla \mathbf{m}$
- Obtain covariance matrix $\operatorname{Cov}(\mathbf{m}^*) = \left(\nabla \mathbf{F}(\mathbf{m}^*)^T \nabla \mathbf{F}(\mathbf{m}^*)\right)^{-1}$
- If σ is unknown but constant across measurements, take $\sigma_i = 1$ above and use for σ in $\frac{1}{\sigma^2}\left(\nabla \mathbf{F}(\mathbf{m}^*)^T \nabla \mathbf{F}(\mathbf{m}^*)\right)^{-1}$ the estimate
$$s^2 = \frac{1}{m-n} \sum_{i=1}^{m} (G(\mathbf{m}) - d_i)^2 .$$
- Do confidence intervals, χ^2 statistic and p-value as in Chapter 2.

Statistics

Problem is $G(\mathbf{m}) = \mathbf{d}$ with least squares solution $\mathbf{m}^*$:

Now what? What statistics can we bring to bear on the problem?

- We minimize $\|\mathbf{F}(\mathbf{m})\|^2 = \sum_{i=1}^{n} \frac{(G(\mathbf{m}) - d_i)^2}{\sigma_i^2}$
- Treat the linear model as locally accurate, so misfit is $\nabla\mathbf{F} = \mathbf{F}(\mathbf{m} + \Delta\mathbf{m}) - \mathbf{F}(\mathbf{m}^*) \approx \nabla\mathbf{F}(\mathbf{m}^*)\nabla\mathbf{m}$
- Obtain covariance matrix $\mathrm{Cov}(\mathbf{m}^*) = \left(\nabla\mathbf{F}(\mathbf{m}^*)^T \nabla\mathbf{F}(\mathbf{m}^*)\right)^{-1}$
- If σ is unknown but constant across measurements, take $\sigma_i = 1$ above and use for σ in $\frac{1}{\sigma^2}\left(\nabla\mathbf{F}(\mathbf{m}^*)^T \nabla\mathbf{F}(\mathbf{m}^*)\right)^{-1}$ the estimate
$$s^2 = \frac{1}{m-n}\sum_{i=1}^{m}(G(\mathbf{m}) - d_i)^2 .$$
- Do confidence intervals, χ^2 statistic and p-value as in Chapter 2.

Outline

Implementation Issues

What could go wrong?

- Problem may have many local minima.
- Even if it has a unique solution, it might lie in a long flat basin.
- Analytical derivatives may not be available. This presents an interesting regularization issue not discussed by the authors. We do so at the board.
- One remedy for first problem: use many starting points and statistics to choose best local minimum.
- One remedy for second problem: use a better technique than GN or LM.
- Do Example 9.2 from the CD to illustrate some of these ideas.
- If time permits, do data fiting from Great Britian population data.

Implementation Issues

What could go wrong?

- Problem may have many local minima.
- Even if it has a unique solution, it might lie in a long flat basin.
- Analytical derivatives may not be available. This presents an interesting regularization issue not discussed by the authors. We do so at the board.
- One remedy for first problem: use many starting points and statistics to choose best local minimum.
- One remedy for second problem: use a better technique than GN or LM.
- Do Example 9.2 from the CD to illustrate some of these ideas.
- If time permits, do data fiting from Great Britian population data.

Implementation Issues

What could go wrong?

- Problem may have many local minima.
- Even if it has a unique solution, it might lie in a long flat basin.
- Analytical derivatives may not be available. This presents an interesting regularization issue not discussed by the authors. We do so at the board.
- One remedy for first problem: use many starting points and statistics to choose best local minimum.
- One remedy for second problem: use a better technique than GN or LM.
- Do Example 9.2 from the CD to illustrate some of these ideas.
- If time permits, do data fiting from Great Britian population data.

Implementation Issues

What could go wrong?

- Problem may have many local minima.
- Even if it has a unique solution, it might lie in a long flat basin.
- Analytical derivatives may not be available. This presents an interesting regularization issue not discussed by the authors. We do so at the board.
- One remedy for first problem: use many starting points and statistics to choose best local minimum.
- One remedy for second problem: use a better technique than GN or LM.
- Do Example 9.2 from the CD to illustrate some of these ideas.
- If time permits, do data fiting from Great Britian population data.

Implementation Issues

What could go wrong?

- Problem may have many local minima.
- Even if it has a unique solution, it might lie in a long flat basin.
- Analytical derivatives may not be available. This presents an interesting regularization issue not discussed by the authors. We do so at the board.
- One remedy for first problem: use many starting points and statistics to choose best local minimum.
- One remedy for second problem: use a better technique than GN or LM.
- Do Example 9.2 from the CD to illustrate some of these ideas.
- If time permits, do data fiting from Great Britian population data.

Implementation Issues

What could go wrong?

- Problem may have many local minima.
- Even if it has a unique solution, it might lie in a long flat basin.
- Analytical derivatives may not be available. This presents an interesting regularization issue not discussed by the authors. We do so at the board.
- One remedy for first problem: use many starting points and statistics to choose best local minimum.
- One remedy for second problem: use a better technique than GN or LM.
- Do Example 9.2 from the CD to illustrate some of these ideas.
- If time permits, do data fiting from Great Britian population data.

Implementation Issues

What could go wrong?

- Problem may have many local minima.
- Even if it has a unique solution, it might lie in a long flat basin.
- Analytical derivatives may not be available. This presents an interesting regularization issue not discussed by the authors. We do so at the board.
- One remedy for first problem: use many starting points and statistics to choose best local minimum.
- One remedy for second problem: use a better technique than GN or LM.
- Do Example 9.2 from the CD to illustrate some of these ideas.
- If time permits, do data fiting from Great Britian population data.

Implementation Issues

What could go wrong?

- Problem may have many local minima.
- Even if it has a unique solution, it might lie in a long flat basin.
- Analytical derivatives may not be available. This presents an interesting regularization issue not discussed by the authors. We do so at the board.
- One remedy for first problem: use many starting points and statistics to choose best local minimum.
- One remedy for second problem: use a better technique than GN or LM.
- Do Example 9.2 from the CD to illustrate some of these ideas.
- If time permits, do data fiting from Great Britian population data.

Outline

Penalized (Damped) Least Squares

Basic Problem:

Solve $G(\mathbf{m}) = \mathbf{d}$, where G is a nonlinear function. As usual, $\mathbf{d}$ will have error and this may not be a well-posed problem. Assume variables are scaled, so standard deviations of measurements are incorporated. So we follow the same paths as in Chapter 5.

- Recast: minimize $\|\mathbf{Gm} - \mathbf{d}\|_2$ – unconstrained least squares.
- Recast: minimize $\|G\mathbf{m} - \mathbf{d}\|_2$ subject to $\|L\mathbf{m}\|_2 \leq \epsilon$, where L is a damping matrix (e.g., $L = I$.)
- Recast: minimize $\|L\mathbf{m}\|_2$ subject to $\|\mathbf{Gm} - \mathbf{d}\|_2 \leq \delta$.
- Recast: (**damped least squares**) minimize $\|G\mathbf{m} - \mathbf{d}\|_2^2 + \alpha^2 \|L\mathbf{m}\|_2^2$. This is also a **Tikhonov regularization** of the original problem, possibly higher order.
- Method of Lagrange multipliers doesn't care if G is nonlinear, so we can apply it as in Chapter 5 to show that these problems are essentially equivalent.
- A big difference: no linear normal equations for least squares problem

Penalized (Damped) Least Squares

Basic Problem:

Solve $G(\mathbf{m}) = \mathbf{d}$, where G is a nonlinear function. As usual, $\mathbf{d}$ will have error and this may not be a well-posed problem. Assume variables are scaled, so standard deviations of measurements are incorporated. So we follow the same paths as in Chapter 5.

- Recast: minimize $\|\mathbf{Gm} - \mathbf{d}\|_2$ – unconstrained least squares.
- Recast: minimize $\|G\mathbf{m} - \mathbf{d}\|_2$ subject to $\|L\mathbf{m}\|_2 \leq \epsilon$, where L is a damping matrix (e.g., $L = I$.)
- Recast: minimize $\|L\mathbf{m}\|_2$ subject to $\|\mathbf{Gm} - \mathbf{d}\|_2 \leq \delta$.
- Recast: (**damped least squares**) minimize $\|G\mathbf{m} - \mathbf{d}\|_2^2 + \alpha^2 \|L\mathbf{m}\|_2^2$. This is also a **Tikhonov regularization** of the original problem, possibly higher order.
- Method of Lagrange multipliers doesn't care if G is nonlinear, so we can apply it as in Chapter 5 to show that these problems are essentially equivalent.
- A big difference: no linear normal equations for least squares problem

Penalized (Damped) Least Squares

Basic Problem:

Solve $G(\mathbf{m}) = \mathbf{d}$, where G is a nonlinear function. As usual, $\mathbf{d}$ will have error and this may not be a well-posed problem. Assume variables are scaled, so standard deviations of measurements are incorporated. So we follow the same paths as in Chapter 5.

- Recast: minimize $\|\mathbf{Gm} - \mathbf{d}\|_2$ – unconstrained least squares.
- Recast: minimize $\|G\mathbf{m} - \mathbf{d}\|_2$ subject to $\|L\mathbf{m}\|_2 \le \epsilon$, where L is a damping matrix (e.g., $L = I$.)
- Recast: minimize $\|L\mathbf{m}\|_2$ subject to $\|\mathbf{Gm} - \mathbf{d}\|_2 \le \delta$.
- Recast: (**damped least squares**) minimize $\|G\mathbf{m} - \mathbf{d}\|_2^2 + \alpha^2 \|L\mathbf{m}\|_2^2$. This is also a **Tikhonov regularization** of the original problem, possibly higher order.
- Method of Lagrange multipliers doesn't care if G is nonlinear, so we can apply it as in Chapter 5 to show that these problems are essentially equivalent.
- A big difference: no linear normal equations for least squares problem

Penalized (Damped) Least Squares

Basic Problem:

Solve $G(\mathbf{m}) = \mathbf{d}$, where G is a nonlinear function. As usual, $\mathbf{d}$ will have error and this may not be a well-posed problem. Assume variables are scaled, so standard deviations of measurements are incorporated. So we follow the same paths as in Chapter 5.

- Recast: minimize $\|\mathbf{Gm} - \mathbf{d}\|_2$ – unconstrained least squares.
- Recast: minimize $\|G\mathbf{m} - \mathbf{d}\|_2$ subject to $\|L\mathbf{m}\|_2 \leq \epsilon$, where L is a damping matrix (e.g., $L = I$.)
- Recast: minimize $\|L\mathbf{m}\|_2$ subject to $\|\mathbf{Gm} - \mathbf{d}\|_2 \leq \delta$.
- Recast: (**damped least squares**) minimize $\|G\mathbf{m} - \mathbf{d}\|_2^2 + \alpha^2 \|L\mathbf{m}\|_2^2$. This is also a **Tikhonov regularization** of the original problem, possibly higher order.
- Method of Lagrange multipliers doesn't care if G is nonlinear, so we can apply it as in Chapter 5 to show that these problems are essentially equivalent.
- A big difference: no linear normal equations for least squares problem

Penalized (Damped) Least Squares

Basic Problem:

Solve $G(\mathbf{m}) = \mathbf{d}$, where G is a nonlinear function. As usual, $\mathbf{d}$ will have error and this may not be a well-posed problem. Assume variables are scaled, so standard deviations of measurements are incorporated. So we follow the same paths as in Chapter 5.

- Recast: minimize $\|\mathbf{Gm} - \mathbf{d}\|_2$ – unconstrained least squares.
- Recast: minimize $\|G\mathbf{m} - \mathbf{d}\|_2$ subject to $\|L\mathbf{m}\|_2 \leq \epsilon$, where L is a damping matrix (e.g., $L = I$.)
- Recast: minimize $\|L\mathbf{m}\|_2$ subject to $\|\mathbf{Gm} - \mathbf{d}\|_2 \leq \delta$.
- Recast: **(damped least squares)** minimize $\|G\mathbf{m} - \mathbf{d}\|_2^2 + \alpha^2 \|L\mathbf{m}\|_2^2$. This is also a **Tikhonov regularization** of the original problem, possibly higher order.
- Method of Lagrange multipliers doesn't care if G is nonlinear, so we can apply it as in Chapter 5 to show that these problems are essentially equivalent.
- A big difference: no linear normal equations for least squares problem...

Penalized (Damped) Least Squares

Basic Problem:

Solve $G(\mathbf{m}) = \mathbf{d}$, where G is a nonlinear function. As usual, $\mathbf{d}$ will have error and this may not be a well-posed problem. Assume variables are scaled, so standard deviations of measurements are incorporated. So we follow the same paths as in Chapter 5.

- Recast: minimize $\|\mathbf{Gm} - \mathbf{d}\|_2$ – unconstrained least squares.
- Recast: minimize $\|G\mathbf{m} - \mathbf{d}\|_2$ subject to $\|L\mathbf{m}\|_2 \le \epsilon$, where L is a damping matrix (e.g., $L = I$.)
- Recast: minimize $\|L\mathbf{m}\|_2$ subject to $\|\mathbf{Gm} - \mathbf{d}\|_2 \le \delta$.
- Recast: **(damped least squares)** minimize $\|G\mathbf{m} - \mathbf{d}\|_2^2 + \alpha^2 \|L\mathbf{m}\|_2^2$. This is also a **Tikhonov regularization** of the original problem, possibly higher order.
- Method of Lagrange multipliers doesn't care if G is nonlinear, so we can apply it as in Chapter 5 to show that these problems are essentially equivalent.
- A big difference: no linear normal equations for least squares problem

Penalized (Damped) Least Squares

Basic Problem:

Solve $G(\mathbf{m}) = \mathbf{d}$, where G is a nonlinear function. As usual, $\mathbf{d}$ will have error and this may not be a well-posed problem. Assume variables are scaled, so standard deviations of measurements are incorporated. So we follow the same paths as in Chapter 5.

- Recast: minimize $\|\mathbf{Gm} - \mathbf{d}\|_2$ – unconstrained least squares.
- Recast: minimize $\|G\mathbf{m} - \mathbf{d}\|_2$ subject to $\|L\mathbf{m}\|_2 \leq \epsilon$, where L is a damping matrix (e.g., $L = I$.)
- Recast: minimize $\|L\mathbf{m}\|_2$ subject to $\|\mathbf{Gm} - \mathbf{d}\|_2 \leq \delta$.
- Recast: (**damped least squares**) minimize $\|G\mathbf{m} - \mathbf{d}\|_2^2 + \alpha^2 \|L\mathbf{m}\|_2^2$. This is also a **Tikhonov regularization** of the original problem, possibly higher order.
- Method of Lagrange multipliers doesn't care if G is nonlinear, so we can apply it as in Chapter 5 to show that these problems are essentially equivalent.
- A big difference: no linear normal equations for least squares problem.

Solution Methodology: Penalized Least Squares

Basic Idea:

Regularize, then linearize.

- Regularize: $\|G(\mathbf{m}) - \mathbf{d}\|_2^2 + \alpha^2 \|L\mathbf{m}\|_2^2$.
- Equivalently: minimize $\left\| \begin{bmatrix} G(\mathbf{m}) - \mathbf{d} \\ \alpha L\mathbf{m} \end{bmatrix} \right\|_2^2 \equiv \|H(\mathbf{m})\|_2^2$.
- Linearize: Compute the Jacobian of this vector function: $\nabla H(\mathbf{m}) = \begin{bmatrix} \nabla G(\mathbf{m}) \\ \alpha L \end{bmatrix}$.
- The linear model of G near current guesstimate $\mathbf{m}^k$, with $\Delta\mathbf{m} = \mathbf{m} - \mathbf{m}^k$: $G(\mathbf{m}) \approx G(\mathbf{m}) + \nabla G(\mathbf{m}) \Delta\mathbf{m}$.
- This leads to the system

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

Solution Methodology: Penalized Least Squares

Basic Idea:

Regularize, then linearize.

- Regularize: $\|G(\mathbf{m}) - \mathbf{d}\|_2^2 + \alpha^2 \|L\mathbf{m}\|_2^2$.
- Equivalently: minimize $\left\| \begin{bmatrix} G(\mathbf{m}) - \mathbf{d} \\ \alpha L\mathbf{m} \end{bmatrix} \right\|_2^2 \equiv \|H(\mathbf{m})\|_2^2$.
- Linearize: Compute the Jacobian of this vector function: $\nabla H(\mathbf{m}) = \begin{bmatrix} \nabla G(\mathbf{m}) \\ \alpha L \end{bmatrix}$.
- The linear model of G near current guesstimate $\mathbf{m}^k$, with $\Delta\mathbf{m} = \mathbf{m} - \mathbf{m}^k$: $G(\mathbf{m}) \approx G(\mathbf{m}) + \nabla G(\mathbf{m}) \Delta\mathbf{m}$.
- This leads to the system

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

Solution Methodology: Penalized Least Squares

Basic Idea:

Regularize, then linearize.

- Regularize: $\|G(\mathbf{m}) - \mathbf{d}\|_2^2 + \alpha^2 \|L\mathbf{m}\|_2^2$.
- Equivalently: minimize $\left\| \begin{bmatrix} G(\mathbf{m}) - \mathbf{d} \\ \alpha L\mathbf{m} \end{bmatrix} \right\|_2^2 \equiv \|H(\mathbf{m})\|_2^2$.
- Linearize: Compute the Jacobian of this vector function: $\nabla H(\mathbf{m}) = \begin{bmatrix} \nabla G(\mathbf{m}) \\ \alpha L \end{bmatrix}$.
- The linear model of G near current guesstimate $\mathbf{m}^k$, with $\Delta\mathbf{m} = \mathbf{m} - \mathbf{m}^k$: $G(\mathbf{m}) \approx G(\mathbf{m}) + \nabla G(\mathbf{m})\Delta\mathbf{m}$.
- This leads to the system

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

Solution Methodology: Penalized Least Squares

Basic Idea:

Regularize, then linearize.

- Regularize: $\|G(\mathbf{m}) - \mathbf{d}\|_2^2 + \alpha^2 \|L\mathbf{m}\|_2^2$.
- Equivalently: minimize $\left\| \begin{bmatrix} G(\mathbf{m}) - \mathbf{d} \\ \alpha L\mathbf{m} \end{bmatrix} \right\|_2^2 \equiv \|H(\mathbf{m})\|_2^2$.
- Linearize: Compute the Jacobian of this vector function: $\nabla H(\mathbf{m}) = \begin{bmatrix} \nabla G(\mathbf{m}) \\ \alpha L \end{bmatrix}$.
- The linear model of G near current guesstimate $\mathbf{m}^k$, with $\Delta\mathbf{m} = \mathbf{m} - \mathbf{m}^k$: $G(\mathbf{m}) \approx G(\mathbf{m}) + \nabla G(\mathbf{m})\Delta\mathbf{m}$.
- This leads to the system

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

Solution Methodology: Penalized Least Squares

Basic Idea:

Regularize, then linearize.

- Regularize: $\|G(\mathbf{m}) - \mathbf{d}\|_2^2 + \alpha^2 \|L\mathbf{m}\|_2^2$.
- Equivalently: minimize $\left\| \begin{bmatrix} G(\mathbf{m}) - \mathbf{d} \\ \alpha L\mathbf{m} \end{bmatrix} \right\|_2^2 \equiv \|H(\mathbf{m})\|_2^2$.
- Linearize: Compute the Jacobian of this vector function:
$\nabla H(\mathbf{m}) = \begin{bmatrix} \nabla G(\mathbf{m}) \\ \alpha L \end{bmatrix}$.
- The linear model of G near current guesstimate $\mathbf{m}^k$, with $\Delta\mathbf{m} = \mathbf{m} - \mathbf{m}^k$: $G(\mathbf{m}) \approx G(\mathbf{m}) + \nabla G(\mathbf{m})\,\Delta\mathbf{m}$.
- This leads to the system

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

Solution Methodology: Penalized Least Squares

Basic Idea:

Regularize, then linearize.

- Regularize: $\|G(\mathbf{m}) - \mathbf{d}\|_2^2 + \alpha^2 \|L\mathbf{m}\|_2^2$.
- Equivalently: minimize $\left\| \begin{bmatrix} G(\mathbf{m}) - \mathbf{d} \\ \alpha L\mathbf{m} \end{bmatrix} \right\|_2^2 \equiv \|H(\mathbf{m})\|_2^2$.
- Linearize: Compute the Jacobian of this vector function: $\nabla H(\mathbf{m}) = \begin{bmatrix} \nabla G(\mathbf{m}) \\ \alpha L \end{bmatrix}$.
- The linear model of G near current guesstimate $\mathbf{m}^k$, with $\Delta\mathbf{m} = \mathbf{m} - \mathbf{m}^k$: $G(\mathbf{m}) \approx G(\mathbf{m}) + \nabla G(\mathbf{m})\Delta\mathbf{m}$.
- This leads to the system

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

Outline

Solution Methodology: An Output Least Squares

Basic Idea:

Linearize, then regularize. Authors call this method "Occam's inversion" – it is a special type of output least squares.

- Develop the linear model of $G(\mathbf{m})$ near $\mathbf{m}^k$:
  $G(\mathbf{m}) \approx G\left(\mathbf{m}^k\right) + \nabla G\left(\mathbf{m}^k\right)\left(\mathbf{m} - \mathbf{m}^k\right)$
- Linearize $\|G(\mathbf{m}) - \mathbf{d}\|_2^2 + \alpha^2 \|L\mathbf{m}\|_2^2$ by making the above replacement for $G(\mathbf{m})$. Call the solution $\mathbf{m}^{k+1}$.
- This leads to the system $m^{k+1} =$
  $\left(\nabla G\left(\mathbf{m}^k\right)^T \nabla G\left(\mathbf{m}^k\right) + \alpha^2 L^T L\right)^{-1} \nabla G\left(\mathbf{m}^k\right)^T \widehat{\mathbf{d}}\left(\mathbf{m}^k\right)$,
  where $\widehat{\mathbf{d}}\left(\mathbf{m}^k\right) = d - G\left(\mathbf{m}^k\right) + \nabla G\left(\mathbf{m}^k\right)^T \mathbf{m}^k$.
- The algorithm is to solve this equation with initial guess $\mathbf{m}^0$, but at each iteration choose the largest value of α such that $\chi^2\left(\mathbf{m}^{k+1}\right) \leq \delta^2$. If none, pick value of α that minimizes χ^2. Stop if/when sequence converges to solution with $\chi^2 \leq \delta^2$.

Solution Methodology: An Output Least Squares

Basic Idea:

Linearize, then regularize. Authors call this method "Occam's inversion" – it is a special type of output least squares.

- Develop the linear model of $G(\mathbf{m})$ near $\mathbf{m}^k$:
  $G(\mathbf{m}) \approx G\left(\mathbf{m}^k\right) + \nabla G\left(\mathbf{m}^k\right)(\mathbf{m} - \mathbf{m}^k)$
- Linearize $\|G(\mathbf{m}) - \mathbf{d}\|_2^2 + \alpha^2 \|L\mathbf{m}\|_2^2$ by making the above replacement for $G(\mathbf{m})$. Call the solution $\mathbf{m}^{k+1}$.
- This leads to the system $m^{k+1} =$
  $\left(\nabla G\left(\mathbf{m}^k\right)^T \nabla G\left(\mathbf{m}^k\right) + \alpha^2 L^T L\right)^{-1} \nabla G\left(\mathbf{m}^k\right)^T \widehat{\mathbf{d}}\left(\mathbf{m}^k\right),$
  where $\widehat{\mathbf{d}}\left(\mathbf{m}^k\right) = d - G\left(\mathbf{m}^k\right) + \nabla G\left(\mathbf{m}^k\right)^T \mathbf{m}^k.$
- The algorithm is to solve this equation with initial guess $\mathbf{m}^0$, but at each iteration choose the largest value of α such that $\chi^2\left(\mathbf{m}^{k+1}\right) \leq \delta^2$. If none, pick value of α that minimizes χ^2. Stop if/when sequence converges to solution with $\chi^2 \leq \delta^2$.

Solution Methodology: An Output Least Squares

Basic Idea:

Linearize, then regularize. Authors call this method "Occam's inversion" – it is a special type of output least squares.

- Develop the linear model of $G(\mathbf{m})$ near $\mathbf{m}^k$:
  $G(\mathbf{m}) \approx G\left(\mathbf{m}^k\right) + \nabla G\left(\mathbf{m}^k\right)(\mathbf{m} - \mathbf{m}^k)$
- Linearize $\|G(\mathbf{m}) - \mathbf{d}\|_2^2 + \alpha^2 \|L\mathbf{m}\|_2^2$ by making the above replacement for $G(\mathbf{m})$. Call the solution $\mathbf{m}^{k+1}$.
- This leads to the system $m^{k+1} = \left(\nabla G\left(\mathbf{m}^k\right)^T \nabla G\left(\mathbf{m}^k\right) + \alpha^2 L^T L\right)^{-1} \nabla G\left(\mathbf{m}^k\right)^T \widehat{\mathbf{d}}\left(\mathbf{m}^k\right)$, where $\widehat{\mathbf{d}}\left(\mathbf{m}^k\right) = d - G\left(\mathbf{m}^k\right) + \nabla G\left(\mathbf{m}^k\right)^T \mathbf{m}^k$.
- The algorithm is to solve this equation with initial guess $\mathbf{m}^0$, but at each iteration choose the largest value of α such that $\chi^2\left(\mathbf{m}^{k+1}\right) \leq \delta^2$. If none, pick value of α that minimizes χ^2. Stop if/when sequence converges to solution with $\chi^2 \leq \delta^2$.

Solution Methodology: An Output Least Squares

Basic Idea:

Linearize, then regularize. Authors call this method "Occam's inversion" – it is a special type of output least squares.

- Develop the linear model of $G(\mathbf{m})$ near $\mathbf{m}^k$:
  $G(\mathbf{m}) \approx G\left(\mathbf{m}^k\right) + \nabla G\left(\mathbf{m}^k\right)\left(\mathbf{m} - \mathbf{m}^k\right)$
- Linearize $\|G(\mathbf{m}) - \mathbf{d}\|_2^2 + \alpha^2 \|L\mathbf{m}\|_2^2$ by making the above replacement for $G(\mathbf{m})$. Call the solution $\mathbf{m}^{k+1}$.
- This leads to the system $m^{k+1} =$
  $\left(\nabla G\left(\mathbf{m}^k\right)^T \nabla G\left(\mathbf{m}^k\right) + \alpha^2 L^T L\right)^{-1} \nabla G\left(\mathbf{m}^k\right)^T \widehat{\mathbf{d}}\left(\mathbf{m}^k\right),$
  where $\widehat{\mathbf{d}}\left(\mathbf{m}^k\right) = d - G\left(\mathbf{m}^k\right) + \nabla G\left(\mathbf{m}^k\right)^T \mathbf{m}^k$.
- The algorithm is to solve this equation with initial guess $\mathbf{m}^0$, but at each iteration choose the largest value of α such that $\chi^2\left(\mathbf{m}^{k+1}\right) \leq \delta^2$. If none, pick value of α that minimizes χ^2. Stop if/when sequence converges to solution with $\chi^2 \leq \delta^2$.

Solution Methodology: An Output Least Squares

Basic Idea:

Linearize, then regularize. Authors call this method "Occam's inversion" – it is a special type of output least squares.

- Develop the linear model of $G(\mathbf{m})$ near $\mathbf{m}^k$:
  $G(\mathbf{m}) \approx G\left(\mathbf{m}^k\right) + \nabla G\left(\mathbf{m}^k\right)\left(\mathbf{m} - \mathbf{m}^k\right)$
- Linearize $\|G(\mathbf{m}) - \mathbf{d}\|_2^2 + \alpha^2 \|L\mathbf{m}\|_2^2$ by making the above replacement for $G(\mathbf{m})$. Call the solution $\mathbf{m}^{k+1}$.
- This leads to the system $m^{k+1} =$
  $\left(\nabla G\left(\mathbf{m}^k\right)^T \nabla G\left(\mathbf{m}^k\right) + \alpha^2 L^T L\right)^{-1} \nabla G\left(\mathbf{m}^k\right)^T \widehat{\mathbf{d}}\left(\mathbf{m}^k\right)$,
  where $\widehat{\mathbf{d}}\left(\mathbf{m}^k\right) = d - G\left(\mathbf{m}^k\right) + \nabla G\left(\mathbf{m}^k\right)^T \mathbf{m}^k$.
- The algorithm is to solve this equation with initial guess $\mathbf{m}^0$, but at each iteration choose the largest value of α such that $\chi^2\left(\mathbf{m}^{k+1}\right) \leq \delta^2$. If none, pick value of α that minimizes χ^2. Stop if/when sequence converges to solution with $\chi^2 \leq \delta^2$.