ERTH 455 / GEOP 555
Geodetic Methods

– Lecture 18: Modeling - Parameter Estimation –

Ronni Grapenthin
rg@nmt.edu
MSEC 356
x5924

October 23, 2017

“Guess the Process”

This more of a “different angles on the same process:”

`http://topex.ucsd.edu/Ecuador/`

Parameter Estimation

- We have measurements and an idea about the process - how do we get **best estimate** for parameters? E.g.,

  $$d = a + b * x$$

  where
  - d are the measurements (column vector)
  - x are the “coordinates” of the measurements (column vector)
  - a, b describe the process (scalars)
- What is a **best estimate**?
- Yes, inference of parameters from measurements is an **estimation**! WHY?

Matrix Notation

. . . on board . . .

Parameter Estimation

Let's look at an example (`least_squares.py`) ...

Least Squares Solution

- least squares is general approach to solve **linear** systems of equations
- linear systems obey superposition and scaling
- assume m_i are model parameters, which of these are linear?

$$\begin{aligned} d &= m_1 + m_2 x - (1/2) m_3 x^2 \\ d &= (m_1 - m_2 x)^{1/2} - m_3^2 x \end{aligned}$$

- General form: $\mathbf{d} = \mathbf{G}\mathbf{m} + \epsilon$

Least Squares Solution

- least squares is general approach to solve **linear** systems of equations
- linear systems obey superposition and scaling
- assume m_i are model parameters, which of these are linear?

$$\begin{aligned} d &= m_1 + m_2 x - (1/2) m_3 x^2 \\ d &= (m_1 - m_2 x)^{1/2} - m_3^2 x \end{aligned}$$

- General form: $\mathbf{d} = \mathbf{G}\mathbf{m} + \epsilon$
  - **d** is data vector
  - **G** design/model/system matrix || Green's functions
  - **m** model parameters that "tweak" **G**
  - ϵ residuals / measurement errors

Least Squares Solution

- least squares is general approach to solve **linear** systems of equations
- linear systems obey superposition and scaling
- assume m_i are model parameters, which of these are linear?

$$\begin{aligned} d &= m_1 + m_2 x - (1/2) m_3 x^2 \\ d &= (m_1 - m_2 x)^{1/2} - m_3^2 x \end{aligned}$$

- General form: $\mathbf{d} = \mathbf{G}\mathbf{m} + \epsilon$
  - **d** is data vector
  - **G** design/model/system matrix || Green's functions
  - **m** model parameters that "tweak" **G**
  - ϵ residuals / measurement errors
- Solve for **m**!

Least Squares Solution

- General form: $\mathbf{d} = \mathbf{Gm} + \epsilon$
- Least squares solution: $\mathbf{m_{est}} = (\mathbf{G^T G})^{-1}\mathbf{G^T d}$

How to get there?

Most problems result in same least squares solution

Least Squares Solution

- General form: $\mathbf{d} = \mathbf{Gm} + \epsilon$
- Least squares solution: $\mathbf{m_{est}} = (\mathbf{G^T G})^{-1} \mathbf{G^T d}$

How to get there?

- **Variational approach**:
- Probabilistic approach:
- Geometric approach:

Most problems result in same least squares solution

Least Squares Solution

- General form: $\mathbf{d} = \mathbf{Gm} + \epsilon$
- Least squares solution: $\mathbf{m_{est}} = (\mathbf{G^T G})^{-\mathbf{1}} \mathbf{G^T d}$

How to get there?

- **Variational approach**:
  - assume optimal solution minimizes length, j of the residual vector r: $j = r^T r$
- Probabilistic approach:
- Geometric approach:

Most problems result in same least squares solution

Least Squares Solution

- General form: $\mathbf{d} = \mathbf{Gm} + \epsilon$
- Least squares solution: $\mathbf{m_{est}} = (\mathbf{G^T G})^{-1}\mathbf{G^T d}$

How to get there?

- **Variational approach**:
  - assume optimal solution minimizes length, j of the residual vector r: $j = r^T r$
- Probabilistic approach:
  - assume optimal solution is most probable one (maximum likelihood), derived from probability density function of observing measurements
- Geometric approach:

Most problems result in same least squares solution

Least Squares Solution

- General form: $\mathbf{d} = \mathbf{Gm} + \epsilon$
- Least squares solution: $\mathbf{m_{est}} = (\mathbf{G^T G})^{-\mathbf{1}} \mathbf{G^T d}$

How to get there?

- **Variational approach**:
  - assume optimal solution minimizes length, j of the residual vector r: $j = r^T r$
- Probabilistic approach:
  - assume optimal solution is most probable one (maximum likelihood), derived from probability density function of observing measurements
- Geometric approach:
  - solution is a projection from data space into model space, what is projection of vector b in direction of vector a

Most problems result in same least squares solution

Variational Approach

- choose solution where residual vector **r** has minimum length
- most common is standard geometric / Euclidean length / L_2 - norm:

$$L_2 = (r_1^2 + r_2^2 + r_3^2 + r_4^2 \dots)^{-1/2} = \sqrt{\sum_{i=1}^{N} r_i^2}$$

- L_1 - norm less sensitive to bias from single bad points:

$$L_1 = (|r_1| + |r_2| + |r_3| + |r_4| \dots)^{-1/2} = \sum_{i=1}^{N} |r_i|$$

Variational Approach

- choose solution where residual vector **r** has minimum length
- most common is standard geometric / Euclidean length / L_2 - norm:

$$L_2 = (r_1^2 + r_2^2 + r_3^2 + r_4^2 \dots)^{-1/2} = \sqrt{\sum_{i=1}^{N} r_i^2}$$

- L_1 - norm less sensitive to bias from single bad points:

$$L_1 = (|r_1| + |r_2| + |r_3| + |r_4| \dots)^{-1/2} = \sum_{i=1}^{N} |r_i|$$

Solutions:

- Least squares solution: $\mathbf{m_{est}} = (\mathbf{G^T G})^{-\mathbf{1}} \mathbf{G^T d}$

Variational Approach

- choose solution where residual vector **r** has minimum length
- most common is standard geometric / Euclidean length / L_2 - norm:

$$L_2 = (r_1^2 + r_2^2 + r_3^2 + r_4^2 \dots)^{-1/2} = \sqrt{\sum_{i=1}^{N} r_i^2}$$

- L_1 - norm less sensitive to bias from single bad points:

$$L_1 = (|r_1| + |r_2| + |r_3| + |r_4| \dots)^{-1/2} = \sum_{i=1}^{N} |r_i|$$

Solutions:

- Least squares solution: $\mathbf{m_{est}} = (\mathbf{G^T G})^{-\mathbf{1}} \mathbf{G^T d}$
- L_1 solution: $\mathbf{G^T R G m_{est}} = \mathbf{G^T R d}$
  - R: diagonal weighting matrix : $R_{i,i} = 1/|r_i|$
  - nonlinear, need iterative alorithm (IRLS) to solve
  - IRLS starts with $m_{est}^0 = m_{est,L_2}$ solution, construct R^0 using residuals
  - iterate until some threshold reached

Variational Approach

- $\mathbf{d} = \mathbf{Gm} + \epsilon$
- calculate $\mathbf{m_{est}} = (\mathbf{G^T G})^{-1} \mathbf{G^T d}$
- get residuals $\mathbf{r_{est}} = \mathbf{d} - \mathbf{Gm_{est}}$
- define $j(\mathbf{m}) = \mathbf{r^T r} = (\mathbf{d} - \mathbf{Gm})^\mathbf{T}(\mathbf{d} - \mathbf{Gm})$
- find minimum j: $\delta j(\mathbf{m_{est}}) = 0$

Confidence Intervals

- if independent and normally distributed data errors:
- $COV(m_{L_2}) = \sigma^2(G^T G)^{-1}$
- get 95% confidence intervals:
  - each model parameter m_i has normal distribution
  - mean given by corresponding $m_{i,true}$

Confidence Intervals

- if independent and normally distributed data errors:
- $COV(m_{L_2}) = \sigma^2(G^TG)^{-1}$
- get 95% confidence intervals:
  - each model parameter m_i has normal distribution
  - mean given by corresponding $m_{i,true}$
  - variance $COV(m_{L_2})_{i,i}$

$$m_{L_2} \pm 1.96(diag(COV(m_{L_2})))^{1/2}$$

Confidence Intervals

- if independent and normally distributed data errors:
- $COV(m_{L_2}) = \sigma^2(G^TG)^{-1}$
- get 95% confidence intervals:
    - each model parameter m_i has normal distribution
    - mean given by corresponding $m_{i,true}$
    - variance $COV(m_{L_2})_{i,i}$

$$m_{L_2} \pm 1.96(diag(COV(m_{L_2})))^{1/2}$$

- 1.96 comes from:

$$\frac{1}{\sigma\sqrt{2\pi}}\int_{-1.96\sigma}^{1.96\sigma} e^{-\frac{x^2}{2\sigma^2}}dx \approx 0.95$$

Parameter Estimation // Inverse Problems are hard . . .

Parameter Estimation // Inverse Problems are hard . . .

- model existence

- model uniqueness

- instability

Parameter Estimation // Inverse Problems are hard . . .

- model existence
  - There may be no model that fits data (exactly)
  - physics are approximate (or wrong)
  - data contain noise
- model uniqueness
- instability

Parameter Estimation // Inverse Problems are hard . . .

- model existence
  - There may be no model that fits data (exactly)
  - physics are approximate (or wrong)
  - data contain noise
- model uniqueness
  - There may be other models than m_{true} that satisfy data
  - e.g., non-trivial null space $\mathbf{G}\mathbf{m}_\mathbf{0} = \mathbf{0}$
  - smoothing or other biases may affect solution
  - model resolution analysis is critical!
- instability

Parameter Estimation // Inverse Problems are hard ...

- model existence
  - There may be no model that fits data (exactly)
  - physics are approximate (or wrong)
  - data contain noise
- model uniqueness
  - There may be other models than m_{true} that satisfy data
  - e.g., non-trivial null space $\mathbf{G}\mathbf{m_0} = \mathbf{0}$
  - smoothing or other biases may affect solution
  - model resolution analysis is critical!
- instability
  - small change in measurement results in enormous change in parameter estimates
  - possibly stabilize such problems regularization (smoothing)