Perspectives on Least Squares

GEOS 655 Tectonic Geodesy
Jeff Freymueller

Least Squares Solution

- Least squares is a general approach to solve linear systems of equations:
- General form is:

$$\begin{pmatrix} \Delta P^{(1)} \\ \Delta P^{(2)} \\ \Delta P^{(3)} \\ \Delta P^{(4)} \end{pmatrix} = \begin{pmatrix} \frac{\partial P^{(1)}}{\partial x} & \frac{\partial P^{(1)}}{\partial y} & \frac{\partial P^{(1)}}{\partial z} & \frac{\partial P^{(1)}}{\partial \tau} \\ \frac{\partial P^{(2)}}{\partial x} & \frac{\partial P^{(2)}}{\partial y} & \frac{\partial P^{(2)}}{\partial z} & \frac{\partial P^{(2)}}{\partial \tau} \\ \frac{\partial P^{(3)}}{\partial x} & \frac{\partial P^{(3)}}{\partial y} & \frac{\partial P^{(3)}}{\partial z} & \frac{\partial P^{(3)}}{\partial \tau} \\ \frac{\partial P^{(4)}}{\partial x} & \frac{\partial P^{(4)}}{\partial y} & \frac{\partial P^{(4)}}{\partial z} & \frac{\partial P^{(4)}}{\partial \tau} \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \\ \Delta \tau \end{pmatrix} + \begin{pmatrix} v^{(1)} \\ v^{(2)} \\ v^{(3)} \\ v^{(4)} \end{pmatrix}$$

  - **d** = A**x** + **v**
    - **d** = data
    - A = design matrix or model matrix
    - **x** = model parameters
    - **v** = residuals or measurement errors
  - No unique "best" way to solve this kind of equation
  - The least squares solution is $\mathbf{x}_{est} = (A^T A)^{-1} A^T \mathbf{d}$
    - Assuming that $(A^T A)^{-1}$ exists!
  - Another notation: **d** = G**m**

Several Ways to Get There

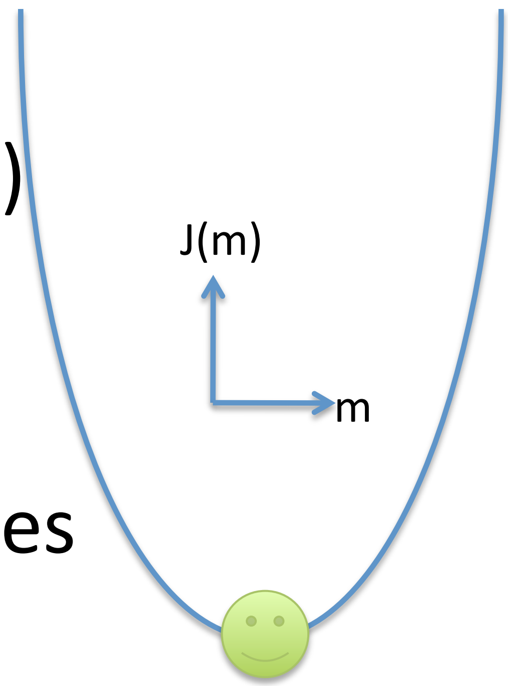
- Variational approach
  - Start from principle that the optimal solution is the one that has minimum length of the residuals $J = \mathbf{v}^T\mathbf{v}$
- Probabilistic approach
  - Start from the principle that the optimal solution is the most probable one (maximum likelihood), derived from probability density function of observing the measured data.
- Geometric (projection approach)
- For many problems, all of these approaches lead to the same least squares solution

Variational or Minimum Length

- Choose the solution where the residual vector v has minimum length
- Most common measure of length is the standard geometric length, called the L_2 norm:
  - Length = $(v_1^2 + v_2^2 + v_3^2 + v_4^2 + \ldots)^{1/2}$
- This is not the only way one could describe the length of a vector. Another example is the L_1 norm:
  - Length = $(|v_1| + |v_2| + |v_3| + |v_4| + \ldots)$
  - The L_1 norm gives a solution that is less sensitive to bias when you have a single bad data point, but if no data are bad, it does not give the maximum likelihood solution.

Variational Approach

- $\mathbf{d} = G\mathbf{m} + \mathbf{v}$
- Find the values of **m** that give the smallest residuals **v**, call this set of values $\mathbf{m}_{est}$
  - Residuals for $\mathbf{m}_{est}$ are $\mathbf{v}_{est} = \mathbf{d} - G\mathbf{m}_{est}$
- Define $J(\mathbf{m}) = \mathbf{v}^T\mathbf{v} = (\mathbf{d} - G\mathbf{m})^T(\mathbf{d} - G\mathbf{m})$
- At the minimum of J, $\delta J(\mathbf{m}_{est}) = 0$
  - Because slope is zero locally
- Can also do this by taking derivatives

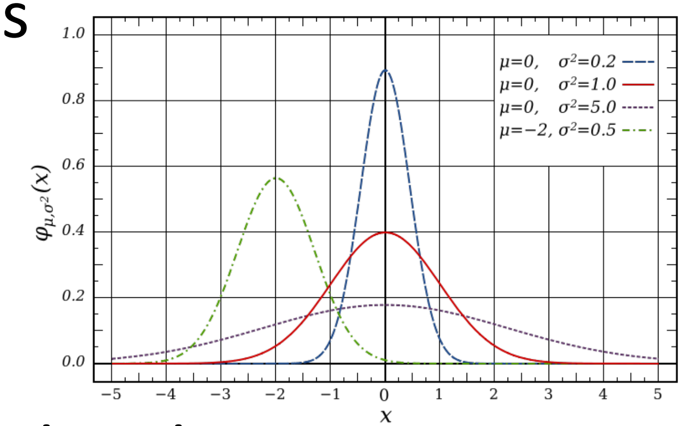


Variational Approach 2

- Solve the equations
  - $\delta J(\mathbf{m}_{est}) = 0$
  - $\delta[\ (\mathbf{d} - G\mathbf{m}_{est})^T(\mathbf{d} - G\mathbf{m}_{est})\ ] = 0$
- Expand the product inside the square brackets and then apply the δ operator to each term
  - The **d** are data and do not change, so $\delta\mathbf{d} = 0$
  - Design matrix G is constant, so $\delta(G\mathbf{m}_{est}) = G\delta\mathbf{m}_{est}$
- Collecting terms gives
  - $(\delta\mathbf{m}_{est}^T G^T)(\mathbf{d} - G\mathbf{m}_{est}) = 0$
  - $\delta\mathbf{m}_{est}^T(G^T\mathbf{d} - G^TG\mathbf{m}_{est}) = 0$
- This is true for any variation $\delta\mathbf{m}_{est}$ only if
  - $(G^T\mathbf{d} - G^TG\mathbf{m}_{est}) = 0 \;\Rightarrow\; G^TG\mathbf{m}_{est} = G^T\mathbf{d} \;\Rightarrow\; \mathbf{m}_{est} = (G^TG)^{-1}G^T\mathbf{d}$

Probabilistic Approach (in brief)

- We usually assume that measurement errors are random and follow a Gaussian or normal distribution. The probability distribution for a random variable d with mean <d> and variance σ is



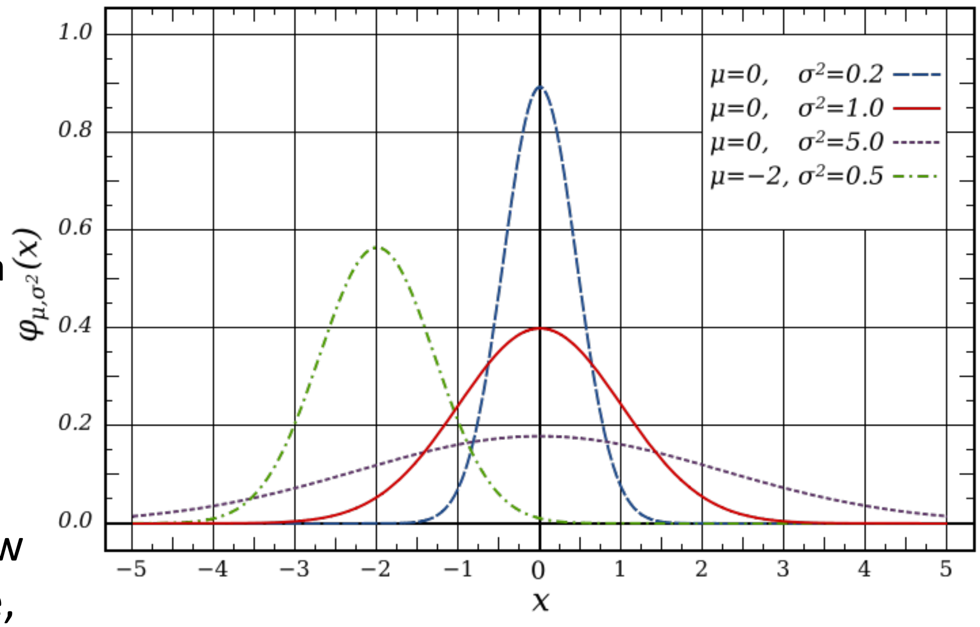
$$P(d) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{\left(d - \langle d \rangle\right)^2}{2\sigma^2}\right]$$

- In a probabilistic approach to estimation, we assert that <d> = Gm (or that the residuals v have zero mean). Expanding to multiple variables, this gives us:

$$P(d) \propto \exp\left[-(d - Gm)^T (d - Gm)\right]$$

Biases and Errors

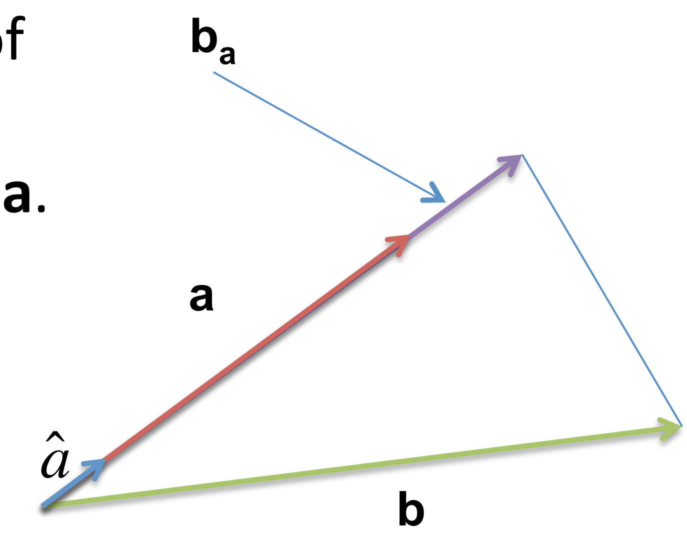
- Suppose we know what the measurement errors are. How would these known errors bias our estimates of position (and clock bias)?
  - $\mathbf{v}_x = (A^TA)^{-1}A^T\mathbf{v}$
  - You might use this to determine whether a newly discovered error has a big impact on your estimated parameters, but in general you would simply want to correct the data!
- In reality, you don't know the measurement errors, but you may know their statistical properties. For example, you may know that the mean measurement error is 0, with uncertainty σ, and that the measurement errors follow a Gaussian or ***normal distribution***.



In this case:
Expectation: $E(v) = 0$
Covariance: $C = E(vvT) = \sigma^2 I$

Projection Approach

- Given two vectors **a** and **b**, what is the projection of **b** in the direction of **a**?
- Define unit vector in the direction of **a**, **a**-"hat" = **a**/|**a**|, and b_a as the component of **b** in the direction of **a**.



$$b_a = \left(\frac{(a \cdot b)}{\|a\|}\right)\left(\frac{a}{\|a\|}\right)$$

$$b_a = \left[(a \cdot b)/\|a\|^2\right]a = \left[\|a\|^{-2}(a^T b)\right]a$$

$$\|a\|^{-2} = \left(a^T a\right)^{-1}$$

$$b_a = \left[\left(a^T a\right)^{-1} a^T b\right]a$$

← Looks familiar?

Projections 2