12.540 Principles of the Global Positioning System
Lecture 11

Prof. Thomas Herring

http://geoweb.mit.edu/~tah/12.540

Statistical approach to estimation

- Summary
  - Look at estimation from statistical point of view
  - Propagation of covariance matrices
  - Sequential estimation

Statistical approach to estimation

- Examine the multivariate Gaussian distribution:

$$\text{Multivariant} \quad f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n |V|}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^T \mathbf{V}^{-1}(\mathbf{x}-\mu)}$$

$$\text{Minimize} \quad (\mathbf{x}-\mu)^T \mathbf{V}^{-1}(\mathbf{x}-\mu) \quad \text{gives largest probability density}$$

- By minimizing the argument of the exponential in the probability density function, we maximize the likelihood of the estimates (MLE).
- This is just weighted least squares where the weight matrix is chosen to be the inverse of the covariance matrix of data noise

Data covariance matrix

- If we use the inverse of the covariance matrix of the noise in the data, we obtain a MLE if data noise is Gaussian distribution.
- How do you obtain data covariance matrix?
- Difficult question to answer completely
- Issues to be considered:
  - Thermal noise in receiver gives on component
  - Multipath could be treated as a noise-like quantity
  - Signal-to-noise ratio of measurements allows an estimate of the noise (discussed later in course).
  - In-complete mathematical model of observables can sometimes be treated as noise-like.
  - Gain of GPS antenna will generate lower SNR at low elevation angles

Data covariance matrix

- In practice in GPS (as well as many other fields), the data covariance matrix is somewhat arbitrarily chosen.
- Largest problem is temporal correlations in the measurements. Typical GPS data set size for 24-hours of data at 30 second sampling is 8x2880=23000 phase measurements. Since the inverse of the covariance matrix is required, fully accounting for correlations requires the inverse of 23000x23000 matrix.
- To store the matrix would require, 4Gbytes of memory
- Even if original covariance matrix is banded (ie., correlations over a time short compared to 24-hours), the inverse of banded matrix is usually a full matrix

Data covariance matrix

- Methods on handling temporal correlations:
  - If measurements correlated over say 5-minute period, then use samples every 5-minutes (JPL method)
  - Use full rate data, but artificially inflate the noise on each measurement so that equivalent to say 5-minute sampling (ie., sqrt(10) higher noise on the 30-second sampled values (GAMIT method)
  - When looking a GPS results, always check the data noise assumptions (discussed more near end of course).

- Assuming a valid data noise model can be developed, what can we say about noise in parameter estimates?

Propagation of covariances

- Given a data noise covariance matrix, the characteristics of expected values can be used to determine the covariance matrix of any linear combination of the measurements.

Given linear operation : $\mathbf{y} = \mathbf{Ax}$ with $\mathbf{V_{xx}}$ as covariance matrix of $\mathbf{x}$

$$\mathbf{V_{yy}} = <\mathbf{yy^T}> = <\mathbf{Axx^T A^T}> = \mathbf{A} <\mathbf{xx^T}> \mathbf{A^T}$$

$$\mathbf{V_{yy}} = \mathbf{A V_{xx} A^T}$$

Propagation of covariance

- Propagation of covariance can be used for any linear operator applied to random variables whose covariance matrix is already known.

- Specific examples:
  - Covariance matrix of parameter estimates from least squares
  - Covariance matrix for post-fit residuals from least squares
  - Covariance matrix of derived quantities such as latitude, longitude and height from XYZ coordinate estimates.

Covariance matrix of parameter estimates

- Propagation of covariance can be applied to the weighted least squares problem:

$$\hat{\mathbf{x}} = (\mathbf{A}^{\mathbf{T}}\mathbf{V}_{yy}^{-1}\mathbf{A})^{-1}\mathbf{A}^{\mathbf{T}}\mathbf{V}_{yy}^{-1}\mathbf{y}$$

$$<\hat{\mathbf{x}}\hat{\mathbf{x}}^{T}> = (\mathbf{A}^{\mathbf{T}}\mathbf{V}_{yy}^{-1}\mathbf{A})^{-1}\mathbf{A}^{\mathbf{T}}\mathbf{V}_{yy}^{-1}<\mathbf{y}\mathbf{y}^{T}>\mathbf{V}_{\mathbf{yy}}^{-1}\mathbf{A}(\mathbf{A}^{\mathbf{T}}\mathbf{V}_{\mathbf{yy}}^{-1}\mathbf{A})^{-1}$$

$$\mathbf{V}_{\hat{\mathbf{x}}\hat{\mathbf{x}}} = (\mathbf{A}^{\mathbf{T}}\mathbf{V}_{yy}^{-1}\mathbf{A})^{-1}$$

- Notice that the covariance matrix of parameter estimates is a natural output of the estimator if $\mathbf{A^{T}V^{-1}A}$ is inverted (does not need to be)

Covariance matrix of estimated parameters

- Notice that for the rigorous estimation, the inverse of the data covariance is needed (time consuming if non-diagonal)
- To compute to parameter estimate covariance, only the covariance matrix of the data is needed (not the inverse)
- In some cases, a non-rigorous inverse can be done with say a diagonal covariance matrix, but the parameter covariance matrix is rigorously computed using the full covariance matrix. This is a non-MLE but the covariance matrix of the parameters should be correct (just not the best estimates that can found).
- This techniques could be used if storage of the full covariance matrix is possible, but inversion of the matrix is not because it would take too long or inverse can not be performed in place.

Covariance matrix of post-fit residuals

- Post-fit residuals are the differences between the observations and the values computed from the estimated parameters
- Because some of the noise in the data are absorbed into the parameter estimates, in general, the post-fit residuals are not the same as the errors in the data.
- In some cases, they can be considerably smaller.
- The covariance matrix of the post-fit residuals can be computed using propagation of covariances.

Covariance matrix of post-fit residuals

- This can be computed using propagation on covariances: **e** is the vector of true errors, and **v** is vector of residuals

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$$

$$\hat{\mathbf{x}} = (\mathbf{A}^{\mathbf{T}}\mathbf{V}_{\mathbf{yy}}^{-1}\mathbf{A})^{-1}\mathbf{A}^{\mathbf{T}}\mathbf{V}_{\mathbf{yy}}^{-1}\mathbf{y}$$

$$\mathbf{v} = \mathbf{y} - \mathbf{A}\hat{\mathbf{x}} = \left[\mathbf{I} - \underbrace{\mathbf{A}(\mathbf{A}^{\mathbf{T}}\mathbf{V}_{\mathbf{yy}}^{-1}\mathbf{A})^{-1}\mathbf{A}^{\mathbf{T}}\mathbf{V}_{\mathbf{yy}}^{-1}}_{\text{Amount error reduced}} \right] \mathbf{e} \quad \text{Eqn 1}$$

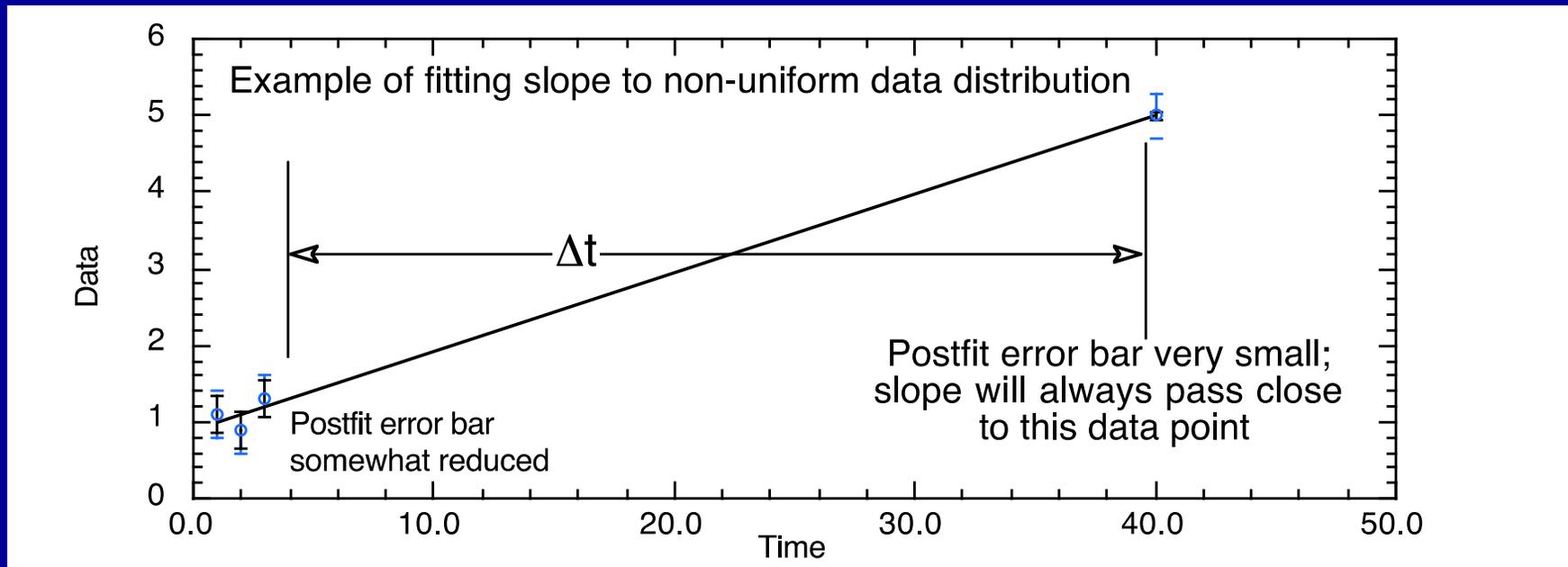$$\mathbf{V}_{\mathbf{vv}} = <\mathbf{v}\mathbf{v}^{T}> = \mathbf{V}_{\mathbf{yy}} - \mathbf{A}(\mathbf{A}^{\mathbf{T}}\mathbf{V}_{\mathbf{yy}}^{-1}\mathbf{A})^{-1}\mathbf{A}^{\mathbf{T}}$$

Post-fit residuals

- Notice that we can compute the compute the covariance matrix of the post-fit residuals (a large matrix in generate)
- Eqn 1 on previous slide gives an equation of the form **v=Be**; why can we not compute the actual errors with **e=B^{-1}v**?
- B is a singular matrix which has no unique inverse (there is in fact one inverse which would generate the true errors)
- Note: In this case, singularity does not mean that there is no inverse, it means there are an infinite number of inverses.

Example

- Consider the case shown below: When a rate of change is estimated, the slope estimate will absorb error in the last data point particularly as Δt increases. (Try this case yourself)

Example of fitting slope to non-uniform data distribution

Δt

Postfit error bar somewhat reduced

Postfit error bar very small; slope will always pass close to this data point

Covariance of derived quantities

- Propagation of covariances can be used to determine the covariance of derived quantities. Example latitude, longitude and radius. θ is co-latitude, λ is longitude, R is radius. ΔN, ΔE and ΔU are north, east and radial changes (all in distance units).

Geocentric Case :

$$\begin{bmatrix} \Delta N \\ \Delta E \\ \Delta U \end{bmatrix} = \underbrace{\begin{bmatrix} -\cos(\theta)\cos(\lambda) & -\cos(\theta)\sin(\lambda) & \sin(\theta) \\ -\sin(\lambda) & \cos(\lambda) & 0 \\ X/R & Y/R & Z/R \end{bmatrix}}_{\textbf{A}\text{ matrix for use in propagation from }\textbf{Vxx}} \begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{bmatrix}$$

Estimation in parts/Sequential estimation

- A very powerful method for handling large data sets, takes advantage of the structure of the data covariance matrix if parts of it are uncorrelated (or assumed to be uncorrelated).

$$\begin{bmatrix} V_1 & 0 & 0 \\ 0 & V_2 & 0 \\ 0 & 0 & V_3 \end{bmatrix}^{-1} = \begin{bmatrix} V_1^{-1} & 0 & 0 \\ 0 & V_2^{-1} & 0 \\ 0 & 0 & V_3^{-1} \end{bmatrix}$$

Sequential estimation

- Since the blocks of the data covariance matrix can be separately inverted, the blocks of the estimation ($\mathbf{A^T V^{-1} A}$) can be formed separately can combined later.

- Also since the parameters to be estimated can be often divided into those that effect all data (such as station coordinates) and those that effect data a one time or over a limited period of time (clocks and atmospheric delays) it is possible to separate these estimations (shown next page).

Sequential estimation

- Sequential estimation with division of global and local parameters. **V** is covariance matrix of new data (uncorrelated with priori parameter estimates), $\mathbf{V_{xg}}$ is covariance matrix of prior parameter estimates with estimates $\mathbf{x_g}$ and $\mathbf{x_l}$ are local parameter estimates, $\mathbf{x_g^+}$ are new global parameter estimates.

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{x_g} \end{bmatrix} = \begin{bmatrix} \mathbf{A_g} & \mathbf{A_l} \\ \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x_g} \\ \mathbf{x_l} \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{x_g^+} \\ \mathbf{x_l} \end{bmatrix} = \begin{bmatrix} \left(\mathbf{A_g}^T \mathbf{V}^{-1} \mathbf{A_g} + \mathbf{V_{xg}^{-1}}\right) & \mathbf{A_g}^T \mathbf{V}^{-1} \mathbf{A_l} \\ \mathbf{A_l}^T \mathbf{V}^{-1} \mathbf{A_g} & \mathbf{A_l}^T \mathbf{V}^{-1} \mathbf{A_l} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{A_g}^T \mathbf{V}^{-1} \mathbf{y} + \mathbf{V_{xg}^{-1}} \mathbf{x_g} \\ \mathbf{A_l}^T \mathbf{V}^{-1} \mathbf{y} \end{bmatrix}$$

Sequential estimation

- As each block of data is processed, the local parameters, x_l, can be dropped and the covariance matrix of the global parameters x_g passed to the next estimation stage.

- Total size of adjustment is at maximum the number of global parameters plus local parameters needed for the data being processed at the moment, rather than all of the local parameters.

Summary

- We examined the way covariance matrices and be manipulated

- Estimation from a statistical point of view

- Sequential estimation.

- Next class continue with sequential estimation in terms of Kalman Filtering.

12.540 Principles of the Global Positioning System

Spring 2012