

12.215 Modern Navigation

Thomas Herring

Summary of last class

- Basic Statistics
 - Statistical description and parameters
 - Probability distributions
 - Descriptions: expectations, variances, moments
 - Covariances
 - Estimates of statistical parameters
- Propagation of variances
 - Methods for determining the statistical parameters of quantities derived other statistical variables

Today's class

- Estimation methods
 - Restrict to basically linear estimation problems (also non-linear problems that are nearly linear)
 - Restrict to parametric, over determined estimation
 - Concepts in estimation:
 - Mathematical models
 - Statistical models
 - Least squares and Maximum likelihood estimation
 - Covariance matrix of estimates parameters
 - Covariance matrix of post-fit residual

Concepts in estimation

- Given multiple observations of a quantity or related to a set of quantities how do you obtain a “best” estimate.
- What do we mean by “best”
- How do you quantify of quality of observations and the relationship between errors in observations.
- The complete answers to the above questions are complex
- We will limit our discussion to parametric estimation mostly for Gaussian distributed random errors in measurements.
- In parametric estimation, mathematical relationships between observations and parameters that can be used to model the observations is used (e.g., GPS measures pseudoranges to satellites: These observations can be related to the positions of the ground station and satellites plus other parameters that we discuss later).

Basics of parametric estimation

- All parametric estimation methods can be broken into a few main steps:
 - Observation equations: equations that relate the parameters to be estimated to the observed quantities (observables). Mathematical model.
 - Example: Relationship between pseudorange, receiver position, satellite position (implicit in ρ), clocks, atmospheric and ionosphere delays
 - Stochastic model: Statistical description that describes the random fluctuations in the measurements and maybe the parameters. In some forms the stochastic model is not explicit.
 - Inversion that determines the parameters values from the mathematical model consistent with the statistical model.

Observation model

- Observation model are equations relating observables to parameters of model:
 - Observable = function (parameters)
 - Observables should not appear on right-hand-side of equation
 - The observed values are the observable plus noise of some stochastic nature
- Often function is non-linear and most common method is linearization of function using Taylor series expansion.
- Sometimes log linearization for $f=a.b.c$ ie. Products fo parameters

Taylor series expansion

- In most common Taylor series approach:

$$y = f(x_1, x_2, x_3, x_4)$$

$$y_0 + \Delta y = f(\mathbf{x})|_{\mathbf{x}_0} + \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} \Delta \mathbf{x} \quad \mathbf{x} = (x_1, x_2, x_3, x_4)$$

- The estimation is made using the difference between the observations and the expected values based on a priori values for the parameters.
- The estimation returns adjustments to a priori parameter values
- The observations are $y + \text{noise}$

Linearization

- Since the linearization is only an approximation, the estimation should be iterated until the adjustments to the parameter values are zero.
- For GPS estimation: Convergence rate is 100-1000:1 typically (ie., a 1 meter error in a priori coordinates could result in 1-10 mm of non-linearity error).
- To assess the level of non-linear contribution, the Taylor series expansion is compared to the non-linear evaluation. If the differences are similar in size to the noise in the measurements, then a new Taylor series expansion, about the better estimates of the parameters, is needed.

Estimation

- Most common estimation method is “least-squares” in which the parameter estimates are the values that minimize the sum of the squares of the differences between the observations and modeled values based on parameter estimates.
- For linear estimation problems, direct matrix formulation for solution
- For non-linear problems: Linearization or search technique where parameter space is searched for minimum value
- Care with search methods that local minimum is not found (will not treat in this course)

Least squares estimation

- Originally formulated by Gauss.
- Basic equations: $\Delta\mathbf{y}$ is vector of observations; \mathbf{A} is linear matrix relating parameters to observables; $\Delta\mathbf{x}$ is vector of parameters; \mathbf{v} is residual

$$\Delta\mathbf{y} = \mathbf{A}\Delta\mathbf{x} + \mathbf{v}$$

minimize $(\mathbf{v}^T \mathbf{v})$; superscript T means transpose

$$\Delta\mathbf{x} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \Delta\mathbf{y}$$

Weighted Least Squares

- In standard least squares, nothing is assumed about the residuals v except that they are zero expectation.
- One often sees weight-least-squares in which a weight matrix is assigned to the residuals. Residuals with larger elements in W are given more weight.

$$\text{minimize } (\mathbf{v}^T \mathbf{W} \mathbf{v})$$

$$\Delta \mathbf{x} = (\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W} \Delta \mathbf{y}$$

Statistical approach to least squares

- If the weight matrix used in weighted least squares is the inverse of the covariance matrix of the residuals, then weighted least squares is a maximum likelihood estimator for Gaussian distributed random errors.
- This choice maximizes the probability density (called a maximum likelihood estimate, MLE)

$$f(\mathbf{v}) = \frac{1}{\sqrt{(2\pi)^n |\mathbf{V}|}} e^{-\frac{1}{2} \mathbf{v}^T \mathbf{V}^{-1} \mathbf{v}}$$

- This latter form of least-squares is most statistically rigorous version.
- Sometimes weights are chosen empirically

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
- For sextant measurements:
 - Index error measurements
 - Individual observers
- Issues to be considered for GPS specifically:
 - 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 $8 \times 2880 = 23000$ phase measurements. Since the inverse of the covariance matrix is required, fully accounting for correlations requires the inverse of 23000×23000 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 (However, remember LU decomposition in linear algebra lecture)

Covariance matrix of parameter estimates

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

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

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

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

- Notice that the covariance matrix of parameter estimates is a natural output of the estimator if $\mathbf{A}^T \mathbf{V}^{-1} \mathbf{A}$ 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: \mathbf{e} is the vector of true errors, and \mathbf{v} is vector of residuals

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

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

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

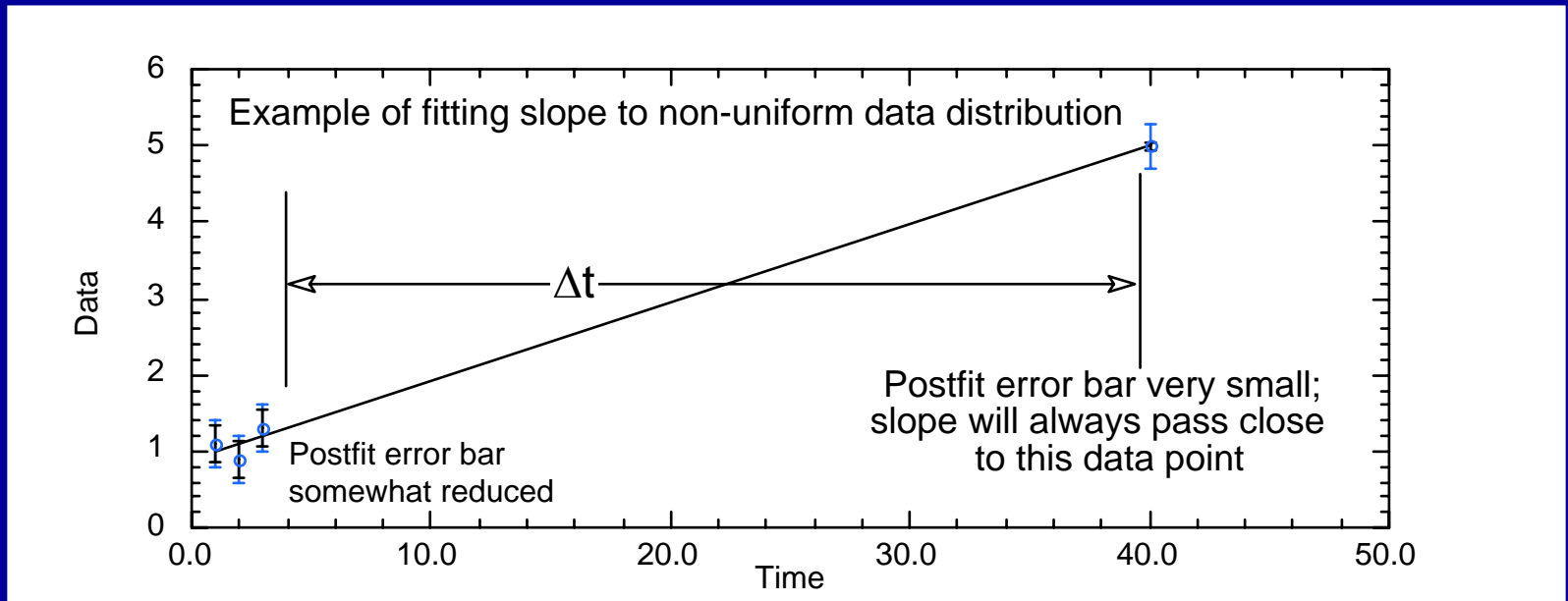
$$\mathbf{V}_{vv} = \langle \mathbf{v}\mathbf{v}^T \rangle = \mathbf{V}_{yy} - \mathbf{A}(\mathbf{A}^T \mathbf{V}_{yy}^{-1} \mathbf{A})^{-1} \mathbf{A}^T$$

Post-fit residuals

- Notice that we can compute the covariance matrix of the post-fit residuals (a large matrix in general)
- 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)



Summary

- Estimation methods
 - Restrict to basically linear estimation problems (also non-linear problems that are nearly linear)
 - Restrict to parametric, over determined estimation
 - Concepts in estimation:
 - Mathematical models
 - Statistical models
 - Least squares and Maximum likelihood estimation
 - Covariance matrix of estimated parameters
 - Statistical properties of post-fit residuals