Statistical ORBIT DETERMINATION, ASEN5070

Lecture 8

Fundamentals of Orbit Determination

Least Squares & Minimum Norm

Fall 2011, 9/9/2011

Supplemental Reading:

Chapter 4

Colorado Center for Astrodynamics Research
The University of Colorado

Relating the Observations to an Epoch State

$$\mathbf{x}(t) = \Phi(t, t_k)\mathbf{x}_k \qquad (4.2.7)$$

$$\mathbf{y}_i = \widetilde{H}_i \mathbf{x}_i + \boldsymbol{\epsilon}_i \qquad (i = 1, \ldots, \ell)$$

$$\begin{aligned}
\mathbf{y}_1 &= \widetilde{H}_1 \Phi(t_1, t_k)\mathbf{x}_k + \boldsymbol{\epsilon}_1 \\
\mathbf{y}_2 &= \widetilde{H}_2 \Phi(t_2, t_k)\mathbf{x}_k + \boldsymbol{\epsilon}_2 \\
&\vdots \\
\mathbf{y}_\ell &= \widetilde{H}_\ell \Phi(t_\ell, t_k)\mathbf{x}_k + \boldsymbol{\epsilon}_\ell .
\end{aligned} \qquad (4.2.37)$$

Relating the Observations to an Epoch State

Using the following definition:

$$\mathbf{y} \equiv \begin{bmatrix} y_1 \\ \vdots \\ y_\ell \end{bmatrix}; \quad H \equiv \begin{bmatrix} \widetilde{H}_1 \Phi(t_1, t_k) \\ \vdots \\ \widetilde{H}_\ell \Phi(t_\ell, t_k) \end{bmatrix}; \quad \boldsymbol{\epsilon} \equiv \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_\ell \end{bmatrix} \tag{4.2.38}$$

and if the subscript on $\mathbf{x}_k$ is dropped for convenience, then Eq. (4.2.37) can be expressed as follows:

$$\mathbf{y} = H\mathbf{x} + \boldsymbol{\epsilon} \tag{4.2.39}$$

Relating the Observations to an Epoch State

$$y = Hx + \epsilon \tag{4.2.39}$$

$\boldsymbol{y}$ is an $m \times 1$ vector

$\boldsymbol{x}$ is an $n \times 1$ vector

ϵ is an $m \times 1$ vector

H is an $m \times n$ mapping matrix

$m = p \times l$ is the total number of observations

$m > n$ is an essential condition

Have m unknown observation errors

Relating the Observations to an Epoch State

$$\mathbf{y} = H\mathbf{x} + \boldsymbol{\epsilon} \tag{4.2.39}$$

Results in:

m unknown observation errors

$m + n$ total unknowns

m equations (Observations)

The least squares criterion provides us with conditions on the m observation errors that allow a solution for the n state variables, $\boldsymbol{X}_k$, at the epoch time t_k

Least Squares

Least Squares Solution

$$\mathbf{y} = H\mathbf{x} + \boldsymbol{\epsilon} \qquad (4.2.39)$$

$$J(\mathbf{x}) = 1/2\boldsymbol{\epsilon}^T\boldsymbol{\epsilon}. \qquad (4.3.1)$$

$$J(\mathbf{x}) = 1/2\boldsymbol{\epsilon}^T\boldsymbol{\epsilon} = \sum_{i=1}^{\ell} 1/2\boldsymbol{\epsilon}_i^T\boldsymbol{\epsilon}_i$$

(4.3.2)

$$\sum_{i=1}^{\ell} 1/2\boldsymbol{\epsilon}_i^T\boldsymbol{\epsilon}_i = 1/2(\mathbf{y} - H\mathbf{x})^T(\mathbf{y} - H\mathbf{x})$$

Least Squares Solution

The quadratic function of x

$$1/2(\mathbf{y} - H\mathbf{x})^T(\mathbf{y} - H\mathbf{x})$$

Will have a unique minima when

$$\frac{\partial J}{\partial \mathbf{x}} = 0, \quad \text{and} \quad \delta\mathbf{x}^T \frac{\partial^2 J}{\partial \mathbf{x}^2} \delta\mathbf{x} > 0$$

for all $\delta\mathbf{x} \neq 0$. Where it is implied that $\frac{\partial^2 J}{\partial \mathbf{x}^2}$ is positive definite

Least Squares Solution

Using eqn B.7.3 yields:

$$\frac{\partial J}{\partial \mathbf{x}} = 0 = -(\mathbf{y} - H\mathbf{x})^T H = -H^T(\mathbf{y} - H\mathbf{x}) \qquad (4.3.3)$$

Solving for x will give the best estimate $\hat{x}$

$$(H^T H)\hat{\mathbf{x}} = H^T \mathbf{y} \qquad (4.3.4)$$

Least Squares Solution

$$\frac{\partial J}{\partial \mathbf{x}} = 0 = -(\mathbf{y} - H\mathbf{x})^T H = -H^T(\mathbf{y} - H\mathbf{x})$$

Evaluating eqn 4.3.3 also provides:

$$\frac{\partial^2 J}{\partial \mathbf{x}^2} = H^T H$$

Which will be positive definite if H is full-rank

Least Squares Solution

If the *normal matrix* is positive definite then:

$$\hat{\mathbf{x}} = (H^T H)^{-1} H^T \mathbf{y}. \qquad (4.3.6)$$

Is our solution for the best estimate of x given the linear observation-state relationship expressed by:

$$\mathbf{y} = H\mathbf{x} + \epsilon \qquad (4.2.39)$$

Least Squares Solution

Given:

$$\mathbf{y} = H\mathbf{x} + \boldsymbol{\epsilon} \qquad (4.2.39)$$

Can solve for the best estimate of the *observation errors*:

$$\hat{\boldsymbol{\epsilon}} = \mathbf{y} - H\hat{\mathbf{x}}. \qquad (4.3.7)$$

Geometric Least Squares

The cost function J is defined by

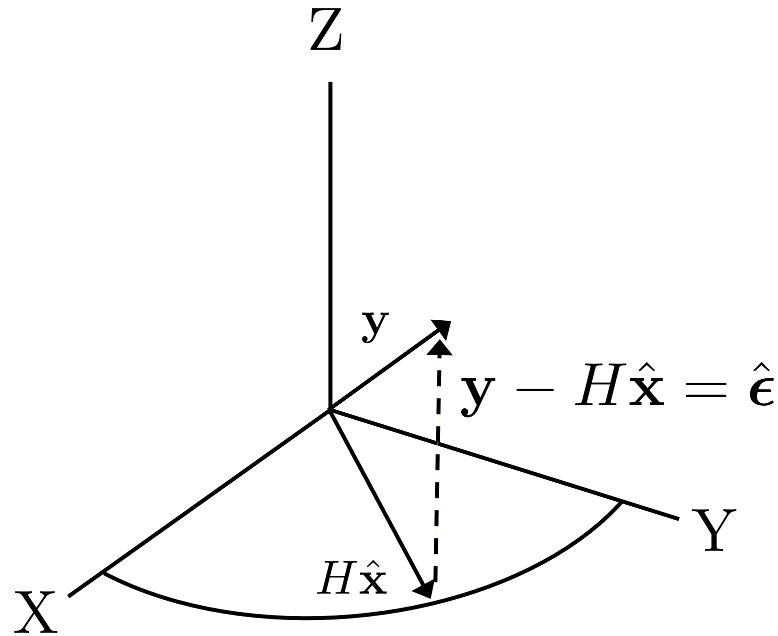
$$J = (\mathbf{y} - H\mathbf{x})^T(\mathbf{y} - H\mathbf{x})$$

In order to minimize J, we want $H\mathbf{x}$ to be as close to $\mathbf{y}$ as possible. We want to choose $\hat{\mathbf{x}}$ such that $(\mathbf{y} - H\mathbf{x})$ is minimized.

Assuming that $[\mathbf{y}]_{m\times 1}$, $[\mathbf{x}]_{n\times 1}$, then $[H]_{m\times n}$. Assume also that H is of rank n. Thus the vector $[H\mathbf{x}]_{m\times 1}$ has n<m DOF and thus cannot span the space occupied by $[\mathbf{y}]_{m\times 1}$.

Geometric Least Squares Example

Assume that y has 3 DOF and $H\mathbf{x}$ has 2 (i.e. m=3, n=2). Graphically, y can lie anywhere in the XYZ space but $H\mathbf{x}$ is constrained to the XY plane.



To minimize $\hat{\epsilon}$, choose $\hat{\mathbf{x}}$ so that $H\hat{\mathbf{x}}$ is the projection of y on the XY plane (i.e. $(\mathbf{y} - H\hat{\mathbf{x}})$ is perpendicular or normal to $H\hat{\mathbf{x}}$).

Geometric Least Squares

Since $H\hat{\mathbf{x}}$ is orthogonal to $(\mathbf{y} - H\hat{\mathbf{x}})$,

$$(H\hat{\mathbf{x}})^T(\mathbf{y} - H\hat{\mathbf{x}}) = 0$$

$$\hat{\mathbf{x}}^T H^T \mathbf{y} - \hat{\mathbf{x}}^T H^T H\hat{\mathbf{x}} = 0$$

$$\hat{\mathbf{x}}^T (H^T \mathbf{y} - H^T H\hat{\mathbf{x}}) = 0$$

The trivial solution is $\hat{\mathbf{x}} = 0$, thus

$$H^T \mathbf{y} - H^T H\hat{\mathbf{x}} = 0$$

and the normal equations become

$$H^T H\hat{\mathbf{x}} = H^T \mathbf{y}$$

Least Squares Example

Example of least squares (assume $w = I$ and we have no apriori)

Let

$$Y_i = \alpha + \beta t_i + \varepsilon_i$$

(Note that this is a linear system)

Assume we wish to estimate

$$X = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \longrightarrow Y_i = \begin{bmatrix} 1 & t_i \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \varepsilon_i$$

$$Y_i = H_i \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \varepsilon_i$$

Least Squares Example

$$\hat{X} = \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = \left(H^T H\right)^{-1} H^T Y$$

where

$$H = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \\ 1 & t_\ell \end{bmatrix}, \quad Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_\ell \end{bmatrix}$$

$$H^T H = \begin{bmatrix} \ell & \sum_{i=1}^{\ell} t_i \\ \sum_{i=1}^{\ell} t_i & \sum_{i=1}^{\ell} t_i^2 \end{bmatrix}, \quad H^T Y = \begin{bmatrix} \sum_{i=1}^{\ell} Y_i \\ \sum_{i=1}^{\ell} t_i Y_i \end{bmatrix}$$

Note that $H^T H$ will always a symmetric matrix

Least Squares Example

Assume:

$$\ell = 3, \quad t = 1, 2, 3 \;\; \& \;\; Y = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$$

Then

$$\hat{\mathrm{X}} = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}^{-1} \begin{bmatrix} 15 \\ 32 \end{bmatrix} = \begin{bmatrix} 7/3 & -1 \\ -1 & 1/2 \end{bmatrix} \begin{bmatrix} 15 \\ 32 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

$$\longrightarrow \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix} \qquad \hat{\varepsilon} = Y - H\hat{x} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} - \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

i.e. we have chosen perfect observations

The Minimum Norm Solution

Minimum Norm

For the least squares solution:

$$\hat{\mathbf{x}} = (H^T H)^{-1} H^T \mathbf{y}.$$

To exist $m \geq n$ and H be of rank n

Consider a case with $m \leq n$ and rank $H < n$

There are more unknowns than linearly independent observations

Minimum Norm

Option 1: specify any $n - m$ of the n components of x and solve for remaining m components of x using observation equations with $\epsilon = 0$

Result: an infinite number of solutions for $\hat{x}$

Option 2: use the minimum norm criterion to uniquely determine $\hat{x}$

Using the generally available nominal/initial guess for x the minimum norm criterion chooses x to minimize the sum of the squares of the difference between X and X* with the constraint that $\epsilon = 0$

Minimum Norm

Recall: $$\boldsymbol{x} = \boldsymbol{X} - \boldsymbol{X}^*$$

Want to minimize the sum of the squares of the difference given $\boldsymbol{\epsilon} = 0$

That is

$$\boldsymbol{y} = H\boldsymbol{x}$$

Minimum Norm

Hence the performance index becomes:

$$J(\mathbf{x}, \boldsymbol{\lambda}) = 1/2\mathbf{x}^T\mathbf{x} + \boldsymbol{\lambda}^T(\mathbf{y} - H\mathbf{x}) \qquad (4.3.8)$$

$$\hat{\mathbf{x}} = H^T(HH^T)^{-1}\mathbf{y} \qquad (4.3.13)$$

$(HH^T)^{-1}$ is known as the pseudo inverse

Three Possible Cases for m and n

$$
\begin{aligned}
\hat{\mathbf{x}} &= (H^T H)^{-1} H^T \mathbf{y}, && \text{if } m > n \\
\hat{\mathbf{x}} &= H^{-1} \mathbf{y}, && \text{if } m = n \\
\hat{\mathbf{x}} &= H^T (H H^T)^{-1} \mathbf{y}, && \text{if } m < n.
\end{aligned}
\tag{4.3.14}
$$

4.3.2 Least Squares Shortcomings

Three major shortcomings of simple least squares solution:

1. Each observation error is weighted equally even though the accuracy of observations may differ
2. The observation errors may be correlated (not independent), and the simple least squares solution makes no allowance for this.
3. The method does not consider that the errors are samples from a random process and makes no attempt to utilize statistical information

Addressing shortcomings

Weighted Least Squares

-includes weighting matrix for observations

Minimum Variance

-considers statistical characteristics of measurement errors

Minimum Variance w/ A Priori Information

-batch/sequential evaluation

4.3.3 Weighted Least Squares Solution

$$
\begin{aligned}
\mathbf{y}_1 &= H_1\mathbf{x}_k + \boldsymbol{\epsilon}_1; \qquad w_1 \\
\mathbf{y}_2 &= H_2\mathbf{x}_k + \boldsymbol{\epsilon}_2; \qquad w_2 \\
&\vdots \quad \vdots \quad \vdots \\
\mathbf{y}_\ell &= H_\ell\mathbf{x}_k + \boldsymbol{\epsilon}_\ell; \qquad w_\ell
\end{aligned}
\tag{4.3.15}
$$

$$
H_i = \widetilde{H}_i\Phi(t_i, t_k).
$$

4.3.3 Weighted Least Squares Solution

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_\ell \end{bmatrix}; \quad H = \begin{bmatrix} H_1 \\ H_2 \\ \vdots \\ H_\ell \end{bmatrix}; \tag{4.3.16}$$

$$\epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_\ell \end{bmatrix}; \quad W = \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \cdots & w_\ell \end{bmatrix}$$

4.3.3 Weighted Least Squares Solution

$$\mathbf{y} = H\mathbf{x}_k + \boldsymbol{\epsilon}; \quad W. \tag{4.3.17}$$

$$J(\mathbf{x}_k) = 1/2\boldsymbol{\epsilon}^T W \boldsymbol{\epsilon} = \sum_{i=1}^{\ell} 1/2\boldsymbol{\epsilon}_i^T w_i \boldsymbol{\epsilon}_i . \tag{4.3.18}$$

$$J(\mathbf{x}_k) = 1/2(\mathbf{y} - H\mathbf{x}_k)^T W(\mathbf{y} - H\mathbf{x}_k). \tag{4.3.19}$$

4.3.3 Weighted Least Squares Solution

$$\frac{\partial J}{\partial \mathbf{x}_k} = 0 = -(y - H\mathbf{x}_k)^T W H = -H^T W(\mathbf{y} - H\mathbf{x}_k)$$

(4.3.21)

$$(H^T W H)\mathbf{x}_k = H^T W \mathbf{y}. \qquad (4.3.22)$$

$$\hat{\mathbf{x}}_k = (H^T W H)^{-1} H^T W \mathbf{y}. \qquad (4.3.23)$$

Geometric Least Squares with weight matrix

The weighting matrix W is symmetric,

$$
\begin{aligned}
J &= (\mathbf{y} - H\mathbf{x})^T \; W \; (\mathbf{y} - H\mathbf{x}) \\
&= (\mathbf{y} - H\mathbf{x})^T \; W^{\frac{1}{2}} W^{\frac{1}{2}} \; (\mathbf{y} - H\mathbf{x}) \\
&= (\mathbf{y}^T W^{\frac{1}{2}} - \mathbf{x}^T H^T W^{\frac{1}{2}})(W^{\frac{1}{2}}\mathbf{y} - W^{\frac{1}{2}} H\mathbf{x})
\end{aligned}
$$

Defining

$$\mathbf{y}' \equiv W^{\frac{1}{2}}\mathbf{y} \qquad \mathbf{y}'^T = \mathbf{y}^T W^{\frac{1}{2}} \qquad H' = W^{\frac{1}{2}} H$$

Using the same procedure as before,

$$J = (\mathbf{y}' - H'\mathbf{x})^T (\mathbf{y}' - H'\mathbf{x})$$

or

$$H'^T H'\hat{\mathbf{x}} = H'^T \mathbf{y}' \qquad H^T W H\hat{\mathbf{x}} = H^T W \mathbf{y}$$

Choosing the Weighting Matrix

The Weighting Matrix may be chosen by using the RMS of the observation residuals.

$$\hat{\varepsilon} = y - H\hat{x}$$

Compute the RMS of the observation errors for each type of observation

$$[RMS]_i = \sqrt{\frac{\hat{\varepsilon}_1^2 + \hat{\varepsilon}_2^2 + \ldots \hat{\varepsilon}_\ell^2}{\ell}}$$

Choosing the Weighting Matrix

Let *i* represent the observation type-say

$$i = 1 \Rightarrow \text{range}$$

$$i = 2 \Rightarrow \text{range rate}$$

so for two observation types let

$$W = \begin{bmatrix} \dfrac{1}{(RMS)_1^2} & 0 \\ 0 & \dfrac{1}{(RMS)_2^2} \end{bmatrix}$$

We use the mean square (MS) so that $J(x) = (y - Hx)^T W (y - Hx)$ will be dimensionless. This can enhance numerical stability of the normal equations.