

Statistics and Estimation

Question 1: Using the data from problem set 2, estimate using your own estimator (not polyfit) the coefficients of the quadratic polynomial that best fit the data in a least squares sense.

- (a) Write the equation for the polynomial. What are the observations and unknown parameters in the polynomial (5-points).
- (b) Write the above set of equations in matrix form (5-points).
- (c) Form the least squares estimator and solve for the coefficients. You may use the Matlab matrix inversion routine inv and/or a calculator matrix inversion to solve the system of equations (10 points).

Solution:

(a) Denoting the measured 2*(Sun Elevation angle) as DE(t) (for "double elevation") where t is UT, we can write the polynomial as

$$DE(t) = a_0 + a_1 t + a_2 t^2$$

The observations are the 14 measured values of DE (actual values are in degrees and minutes and we will need to convert these to decimal degrees). The unknown parameters we seek to estimate are a_0 , a_1 , a_2 .

(b) In matrix form we can write these equations as:

$$\underbrace{\begin{bmatrix} DE(t_1) \\ DE(t_2) \\ \dots \\ DE(t_{n-1}) \\ DE(t_n) \end{bmatrix}}_{\mathbf{DE}} = \underbrace{\begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ \dots & \dots & \dots \\ 1 & t_{n-1} & t_{n-1}^2 \\ 1 & t_n & t_n^2 \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}}_{\mathbf{X}}$$

where there are 14 rows in the left hand side vector. In standard matrix form the above equation can be written as

$$\mathbf{DE} = \mathbf{A X}$$

(c) The [HW03_09.m](#) Matlab program that goes with the solution generates the following results.

```
Solution for polynomial coefficients
Offset          -488.95640
Linear           90.86377
Quadratic       -3.63352
```

Question 2: Estimate the standards deviation of the errors in the measurements using the differences between the observed values and the polynomial fit. (20-points)

Solution:

Assuming the same noise process acts on all of the measurements, we can generate the differences between the observed values and the estimates from the model fit with

$$\mathbf{R} = \mathbf{DE} - \mathbf{A} \hat{\mathbf{X}}$$

The estimate of the standard deviation of the noise process is then given by:

$$\sigma^2 = \mathbf{R}^T \mathbf{R} / (n - 3)$$

where n is the number of observations and 3 is the number of parameters estimated.

From the MatLab program we obtain

Q2: Standard Deviation of Fit to polynomial
Sigma (RMS fit) 0.0993 degs

Question 3: *What is the probability that the 12th measurement (time 16hr 18m 21s Measurement 78° 37.2') differs from the polynomial fit due to random error assuming that the noise in the measurements is Gaussian distributed and the data standard deviation computed in question 2 (20-points).*

The 12th measurement has a residual of 0.25 degrees, and given the above sigma estimate of 0.099 degrees, the residual deviates by 2.53σ . From a table of Standard Normal Distributions or $(1 - (\text{CDF}[\text{NormalDistribution}[0, 1], 2.53] - 0.5) * 2)$ in Mathematica, we find the probability of deviating by the absolute value of this amount or more due to random error is 1.1%. Hence there is about 1% chance we would have a residual of this size. Had we made 100 measurements, then one value of this sign and magnitude should have been expected. We made 26 measurements. For each measurement there is 98.90% chance the residual will have an absolute value $< 2.53\sigma$. Assuming all the measurements are uncorrelated, then there is 0.989^{26} chance that all values will be with 2.53σ (75%). Therefore only 25% of the time would we expect all residuals to be less than 2.53σ .

Question 4: *Estimate the standard deviation of the peak in the polynomial (i.e., its maximum value and the time at which the maximum occurs) based on the least squares estimate in Question 1. From these results infer how well the latitude and longitude were determined (20-points).*

Solution:

To solve this problem we use propagation of variances and we need to write a linear equations between the time the maximum "double elevation" (DE) (variable obs in Matlab code) is reached, denoted by T_m , and the value of DE at the time, denoted DE_m . From HW3 Solution, we have an equation for T_m and substituting this value into the equation for DE_m yield non-linear equations for T_m and DE_m :

$$T_m = \frac{-a_1}{2a_2}$$

$$DE_m = a_o + a_1 T_m + a_2 T_m^2$$

$$\therefore DE_m = a_o - \frac{a_1^2}{4a_2^2}$$

We can linearize these equations which generates:

$$\underbrace{\begin{bmatrix} \Delta T_m \\ \Delta DE_m \end{bmatrix}}_{\mathbf{M}} \approx \underbrace{\begin{bmatrix} 0 & \frac{-1}{2a_2} & \frac{a_1}{2a_2^2} \\ 1 & \frac{-2a_1}{4a_2} & \frac{a_1^2}{4a_2^2} \end{bmatrix}}_{\mathbf{B}} \underbrace{\begin{bmatrix} \Delta a_0 \\ \Delta a_1 \\ \Delta a_2 \end{bmatrix}}_{\Delta \mathbf{X}}$$

The variance covariance matrix of the parameter estimates is obtained from by scaling the inverse of $\mathbf{A}^T \mathbf{A}$ by the observation variance computed in Question 2. We denote this matrix by \mathbf{C}_{xx} . The covariance matrix of the maximum values is then given by

$$\mathbf{C}_m = \mathbf{B} \mathbf{C}_{xx} \mathbf{B}^T$$

These equations are solved in the Matlab program and the results converted in the errors in longitude and latitude (for the latter we divide the sigma by 2 since we use divide DE_m by 2 to obtain the latitude).

The results from the Matlab program are:

```
Q4: Time of Maximum obs is      12.52438 hrs
Maximum value of obs is      78.53923 deg
Estimate of sigma of Tm (minutes of arc)      11.00 min 0.1834 deg
Estimate of sigma of Latitude (minutes of arc)  0.69 min 0.0115 deg
Correlation between values is  0.176
```

Question 5: Using the non-linear model for the observed double elevation angle as a function of the site latitude and longitude (and the Sun's declination and Greenwich hour angle which you can assume are known), rigorously estimate the latitude and longitude of campus. Approach the problem with the following steps:

- Write the equation for the double-elevation as a function of latitude and longitude (5-pts)
- Find the partial derivative of the double angle with respect to latitude and longitude (10-pts)
- Find the differences between the measured double elevation angles and the values predicted from an a priori estimate of the latitude and longitude of 42.0 and 71.0 (5-pts)
- Using the partial derivatives from (b) and the prefit residuals from (c), form the least squares solution for the estimates of the adjustments to the latitude and longitude. Since this is a non-linear problem, iterate your solution until the adjustments are small compared to the standard deviations of the estimates i.e., compute new residuals from the non-linear model with the new estimates of the latitude and longitude. Use the estimates if the standard deviations of sextant measurements from Q2 above. (10-pts)

(e) Give the final estimates of the latitude and longitude and their standard deviations and correlation between the estimates. (10-pts)

Solution:

The basic steps here follow the pattern of nearly all least squares problems (see Lecture 13, slide 5 on):

- Develop the non-linear model between the measurements and the parameters of interest
- Develop the stochastic model of the measurement noise
- Apply the standard least squares inversion techniques to determine the estimates of the parameters that minimizes the sum squares of the differences between observations and model based on the parameter estimates.

As explained on slide 6: For non-linear models, the usual method is to linearize the system using a first-order Taylor series expansion.

Non-linear model:

The measurements made with the sextant were twice the elevation angle to the sun, which in turn is twice 90-zenith distance. The equation for the zenith distance as a function of latitude and longitude were given in lecture 10. This equation also depends on the Greenwich Hour Angle of the Sun, the declination of the Sun, the atmospheric bending and the index error. These later parameters can be assumed known since only the latitude and longitude are to be estimated (the other parameters can not be separately estimated except maybe the index error).

From lecture 10: We have

$$\cos Z_d = \sin \phi \cos(90 - \delta) + \cos \phi \sin(90 - \delta) \cos \Delta GHA \quad (1)$$

where ΔGHA is the difference between the hour angle to the sun at some specific time and the time of meridian crossing at the longitude of the site.

The ΔGHA value for a specific time is computed by finding the time for the Almanac when the GHA of the Sun equals the longitude and the time difference from this time is multiplied by approximately 15 to obtain degrees. In the Matlab solution to this homework, a few hours of the Almanac data are used to obtain a linear model of the motion of the Sun over the duration of our measurements. By assuming a guess of the latitude and longitude (apriori values), we can use equation (1) to compute the zenith distances to the Sun at the times measurements were made. The full model then of the observations is:

$$DE = 2(90 - Z_d + A) + I \quad (2)$$

where DE is double elevation angle, Z_d is zenith distance, A is the atmospheric bending (computed from the zenith distance) and I is the index error. The combination of equations 2 and 1 is the theoretical model of the measurements as a function of latitude and longitude.

Linearization

To use these equations in a least squares estimator, we need to linearize them by Taylor series expansion (i.e., find the partial derivatives of DE with respect to latitude and longitude. This is easiest done with the chain rule and differentials.

$$\frac{\partial DE}{\partial \phi} = \frac{\partial DE}{\partial Zd} \frac{\partial Zd}{\partial \phi} \quad \text{and} \quad \frac{\partial DE}{\partial \lambda} = \frac{\partial DE}{\partial Zd} \frac{\partial Zd}{\partial \Delta GHA} \frac{\partial \Delta GHA}{\partial \lambda} \quad (3)$$

The only complicated derivatives here are for ZD with respect to latitude and DGHA. The derivatives are given by:

$$\begin{aligned} \frac{\partial Zd}{\partial \phi} &= - \frac{\cos \phi \sin \delta - \sin \phi \cos \delta \cos \Delta GHA}{\sin Zd} \\ \frac{\partial Zd}{\partial \Delta GHA} &= \frac{\cos \phi \sin \delta \sin \Delta GHA}{\sin Zd} \end{aligned} \quad (4)$$

These expressions for the derivatives allow the partial derivative matrix to be constructed (a pair of values at each time measurements were made).

Stochastic Model

In general this model can be difficult to get correct, but in this case we assume all measurements have the same standard deviation and therefore the covariance matrix of the data is a unit matrix with a constant variance multiplier, $\mathbf{V} = \sigma^2 \mathbf{I}$

Estimation:

This part is now very straightforward: Denote the vector of differences between the measurements and the theoretical values computed from assumed values for latitude and longitude, by y ; the $n \times 2$ matrix of partial derivatives (n is the number of measurements) by \mathbf{A} ; and the 2-element vector of adjustments to the latitude and longitude by Δx ; the least squares solution of x is

$$\Delta x = \underbrace{(\mathbf{A}^T \mathbf{V}^{-1} \mathbf{A})^{-1}}_{\mathbf{N}} \mathbf{A}^T y \quad (5)$$

Since this is a non linear problem, the values of Δx are added to the a priori values for the latitude and longitude and the whole process starting by evaluating the non-linear model is iterated until the magnitude adjustments Δx are acceptably small (usually some small fraction of the standard deviation with which they are determined from the data. The center part of the matrix in equation 5 are often called the normal equations and their inverse is the variance-covariance matrix of the parameter estimates. (Since the data covariance matrix is a unit matrix by a constant, the homework solution, uses a unit matrix and multiplies the inverse of the normal by σ^2 .

The Homework solution is in [HW03_09.m](#). The results from running the program are:

```
12.215 Homework 3 Starting Least squares iteration
Iteration 1, Adjustment magnitude 0.07180; RMS fit 5.025476 min
Latitude 42.2989 deg, Longitude 71.1276 deg
Iteration 2, Adjustment magnitude 0.00006; RMS fit 5.025467 min
Latitude 42.2989 deg, Longitude 71.1275 deg
```

Final Latitude and longitude estimates: RMS Fit 5.03 min
 Sextant GPS Pol dGPS dPol
Latitude 42.299 ± 0.010 42.360 42.326 -0.061 -0.027 deg
Longitude 71.128 ± 0.065 71.089 71.383 0.038 -0.256 deg
Correlation between lat and long estimates -0.4886

(Some intermediate results from the program: To implement a model of the GHA and Sun declination as a function of time, a linear model fit to the tabular points in the almanac. Simply taking two points would have worked just as well. The fit of the almanac values to linear trends in shown below:

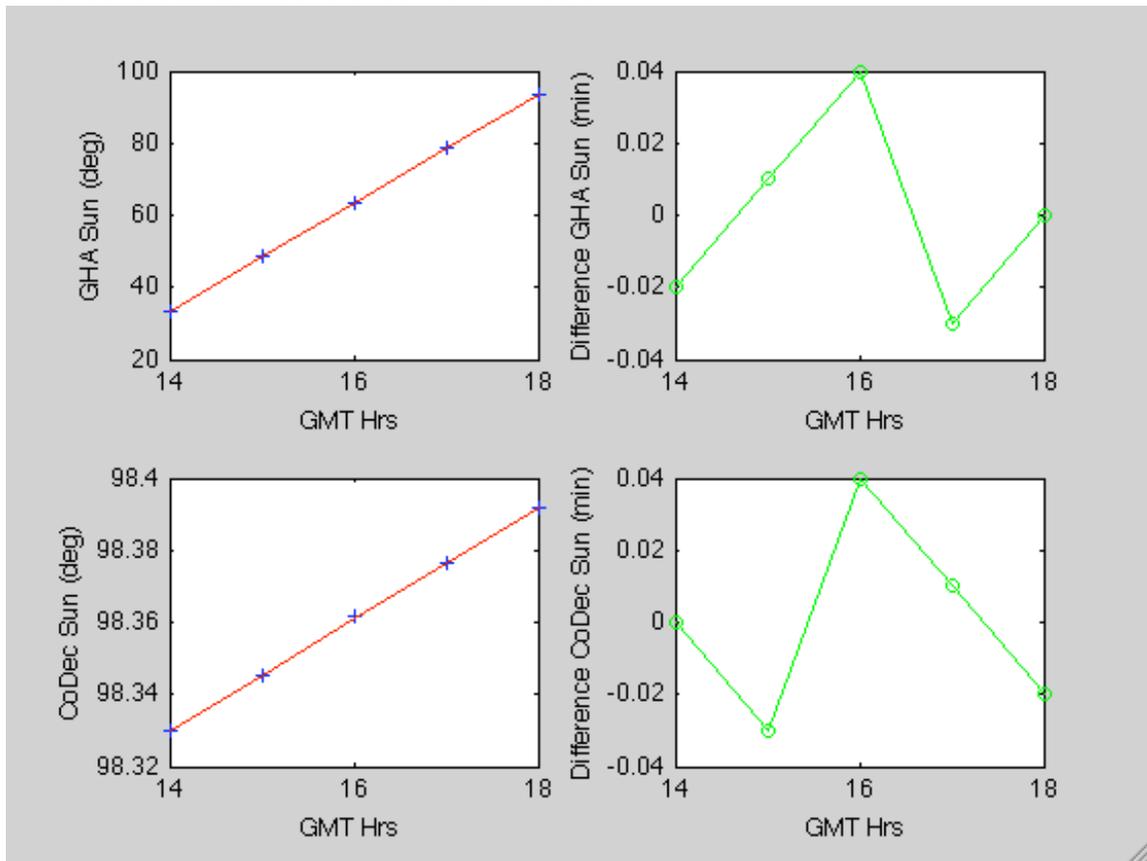


Figure 1: Fits to the values from Almanac.

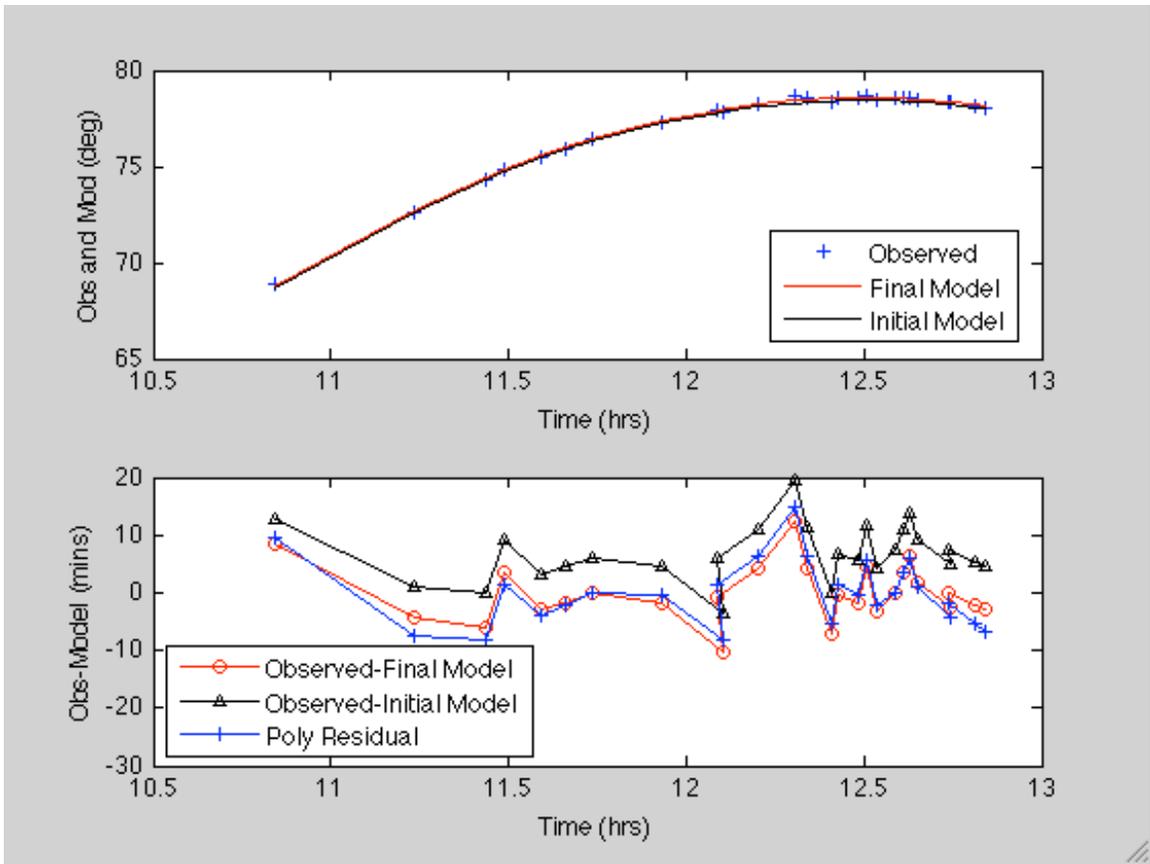


Figure 2: Comparison of observed and model values.