To the program you wrote last week, add a calculation of the data resolution matrix, the model resolution matrix, the model covariance matrix and an estimate of the standard deviation of the estimated model parameters if the errors in the data are uncorrelated with uniform variance.

The following is a simple exercise that shows the steps involved in proposing a model and then estimating the parameters and assessing the validity of that model.

Use the newer file 'co2_mm_mlo.txt' from the class web site

The data in ~~MaunaLoaCO2.dat~~ are monthly averages of $CO_2$ concentration in the atmosphere (in ppm) between 1958 and ~~2015. There are 14 columns, the year, the CO2 concentration Jan to Dec, in twelve columns and the yearly average in the last column. You can load these data into matlab directly (*load MaunaLoaCO2.dat*), or if you are using C or fortran you will have to write a formatted read. Load and find the times for which there are no valid observations (these are flagged with values of -99.99. You will need to reshape the array into a single line or column, create a month scale, and then use the matlab *find* to find and eliminate the no data~~. Since the authors have used two significant figures after the decimal, you may assume that the errors of observation are about .01 ppm. You could also assume the errors are normally distributed with a standard deviation 0.01 ppm but you do not really know this.

By inspection, the data represent a quasi-annual signal, as well as a quasi-exponential increase with a time constant of about 55 years. We could therefore propose a model for these variations that has an annual variation, and an exponential variation with a time constant of 55 years. We could test that this model indeed "explains" the data by least squares fitting an offset and a coefficient for an exponential.

$$A + Be^{\frac{t-t(1)}{55}}$$

Let's start with an hypothesis that the data can be modeled solely as an exponential with a time constant of 55 years. Least squares fit an exponential with a time constant of 55 years to these data. Examine the residuals $d^{obs} - d^{pre}$. Is there a trend? If there is, then your model for the drift is inadequate and you would need to consider something else. By trial and error you could adjust the time constant for the exponential drift, or linearize the exponential, or move to a non-linear least squares, which we will do later. If there is no trend then your model for the drift is good, but there should still be an obvious annual variation. Now look at a histogram of these residuals. Do they look normally distributed? What is the standard deviation of the residuals? If this is much larger than the (guessed) standard deviation of the observations (0.01 ppm), then either the assumed standard deviation is incorrect, or, there is some variance that is not explained by the model. This is certainly true, because an annual variation is evident even in the raw data.

Now fit an annual signal, as well as an exponential, to the data and examine the residuals the same way. Again, if your proposed model and assumed errors are correct then the residuals will be stationary (no trend) and normally distributed with a standard deviation of 0.01 ppm.

$$A + Be^{\frac{t-t(1)}{55}} + Ccos(2\pi t) + Dsin(2\pi t)$$

You may find there is still some systematic residual, or even a histogram that does not look especially gaussian. Any idea of what to do next?