6.7 Practical Problems with Curve Fitting

- simple conceptual problems
- more complex conceptual problems
- unequal coefficient contributions to chi-square space
- false minima in chi-square space
- "bumpy" chi-square space
- false minima with identical functional forms
- false minima with similar functions
- nearly impossible fits due to noise blurring false minima
Simple Conceptual Problems

- curve fitting will not work if the function does not represent the data (look at residuals!).
- having too few data points
- having too many parameters
- minimization of chi-square requires that the noise on $y$ be adequately represented by a normal pdf
- poor coefficient estimates when searching chi-square space
- not checking the adequacy of the fit by comparing a regression line to the data, or not comparing the standard deviation of the fit to the expected error of the $y$ measurement
- using a normal least-squares approach when the $x$-axis has error [see Mandel, "The Statistical Analysis of Experimental Data," Wiley Interscience, 1964, Chapter 12, p. 288]
- fitting a straight line to a log-log graph, when a slope different from unity makes no physical sense (instead fit the data to an equation of the form, $y = a_0 + x$)
- using curve fitting to enhance the signal-to-noise ratio of poorly designed experiments
More Complex Conceptual Problems

- using a function that worked with one range of values to describe a different range of values. *As an example, fluorescence is linear with low concentrations, but is non-linear at high concentrations.*
- using an x-axis spacing so coarse that the measured points miss higher resolution features
- forgetting that assigning data to x-axis bins convolves the original function with the bin width. *This makes a fitting function, that would be correct in the absence of binning, incorrect in the presence of binning.*
- not recognizing that the data have an offset (in either x or y). *The presence of an offset must be explicitly taken into account by the use of an additional equation coefficient.*
- not recognizing that the instrument response is convolved with the measured data. *As an example, an RC electronic time constant can modify the shape of a Gaussian peak by skewing it.*
- trying to use advanced fitting of non-linear functions when there are too many equation coefficients for the quality of the data
Unequal Contributions to $\chi^2$-Space

For an exponentially damped cosine (FID) an error in the cosine period has a far larger effect on chi-square than an equal-sized error in the decay constant. This makes mathematical sense since an error in period causes the trial cosine to "walk off" the data cosine.

$$y = \exp\left(-\frac{t}{\tau}\right) \cos\left(2\pi \frac{t}{t^0}\right)$$

Solution: Normalize all parameters by the guess values such that the minima will be close to unity for all.
False Minima in Chi-Square Space

Some functional forms create false minima far from the true minimum. This puts even more emphasis on the need for exceedingly good initial estimates of the coefficients.

In this particular contour graph, violet is the minimum. All of the green "islands" are false minima.

\[ y = \sin \left( 2\pi \frac{x}{a_0} \right) + \cos \left( 2\pi \frac{x}{a_1} \right) \]
Chi-square space for two exponential functions with the same amplitude has two equally-deep minima. These minima differ only by the interchange of the parameters. Here $\mu_0 = 3$ and $\mu_1 = 6$.

Without a good guess, programs that search chi-square space can become confused.

\[
y = 0.5\exp(-t/a_0) + 0.5\exp(-t/a_1)
\]
\[
y = 0.5\exp(-t/a_1) + 0.5\exp(-t/a_0)
\]
False Minima with Similar Functions

With similar functions there will be one true and one or more false minima. As the functions become less similar, the false minima become less deep.

With this case, it is important which function is associated with which guess. The false minimum will trap ANY method of searching chi-square space.

\[
y = 0.4\exp\left(-t/a_0\right) + 0.6\exp\left(-t/a_1\right)
\]

\[
y = 0.3\exp\left(-t/a_0\right) + 0.7\exp\left(-t/a_1\right)
\]
Nearly Impossible Fits

In the presence of noise, it becomes extremely difficult to distinguish between the true minimum and the false minimum. Result: High coefficient uncertainties.

The contour graphs are for the following function with two levels of noise added. $a_0 = 4$ and $a_1 = 5$

$$y = 0.4\exp(-t/a_0) + 0.6\exp(-t/a_1)$$

$$\sigma = 0.001$$

$$\sigma = 0.01$$