10.34 Fall 2006
Homework: Models vs. Data
Due Monday November 13, 2005, 9 am.

Your task is to analyze the dataset Signal(λ,t) presented in the SVD homework. Each point in this dataset is an average of 100 repeated experiments, so you expect the uncertainties in these data to be approximately given by a Gaussian distribution. Unfortunately, the individual points from each repeat were not stored, only the average value, so you do not know the standard deviation corresponding to each point, and you do not know the expected variance in the averaged data.

Before beginning, you are 100% certain that the data is of the form

$$\text{Signal}(\lambda,t) = \Sigma\, c_i(t)\, A_i(\lambda) + \text{noise}$$

Also because of their physical meanings, concentrations c and absorption strengths A can never be negative. You are 99.9% sure the volume change during the reaction is negligible, i.e. $dc_i/dt = (1/V)\, dn_i/dt$ is accurate to as many digits as you can measure, and you are 99.9% sure that the noise level in the instrument is the same at every wavelength and time point to the accuracy with which you can determine it.

Based on the SVD analysis (see solution to homework 5), you believe that there are only two absorbing species contributing to the signal, call them B (which decays away, related to species 2 in homework 5 posted solution) and C (which is formed, related to species 1 in homework 5 posted solution).

The SVD analysis appears to show two different time constants. This suggests two different very simple kinetic models, either sequential reaction (where C is formed from B) or independent reaction (where C is not formed from B):

Sequential reaction model: $B \rightarrow X$ rate constant k_1
$X \rightarrow C$ rate constant k_2
initially no X or C in the mixture, just B.
According to this model, B drops off as a single exponential, time constant $1/k_1$. C rises as a double exponential, with time constants $1/k_1$ and $1/k_2$.

Independent reaction model: $B \rightarrow X$ rate constant k_1
$Y \rightarrow C$ rate constant k_3
Initially no X or C in the mixture, just B and Y.
According to this model, B drops off as a single exponential time constant $1/k_1$, and C rises as a single exponential, time constant $1/k_3$.

The SVD analysis also suggests that the absorption spectra of B and C can be modeled as a sum of one or two Gaussian Peaks, i.e.

$$A_B(\lambda) = \Sigma\, A_{peakB,n} \exp((\lambda-\lambda_{B,n})/w_{B,n})^2)$$
$$A_C(\lambda) = \Sigma\, A_{peakC,n} \exp((\lambda-\lambda_{C,n})/w_{C,n})^2)$$

Cite as: William Green, Jr., course materials for 10.34 Numerical Methods Applied to Chemical Engineering, Fall 2006. MIT OpenCourseWare (http://ocw.mit.edu), Massachusetts Institute of Technology. Downloaded on [DD Month YYYY].

Your assignment is to:

1) Write out the analytical expression for $S_{model}(\lambda,t,\theta)$ for each of the competing models. Observe that some of the physically-meaningful parameters in the models are not separable, e.g. $c_B(t_o)$ always appears multiplied by an Absorbance, so you might be able to determine the product of $c_B(t_o)$ and an Absorbance, but you will not be able to determine $c_B(t_o)$ separately. (You could get an equally good fit by cutting cB(to) in half and doubling all the A_B's.) Rewrite each $S_{model}(\lambda,t,\theta)$ in terms of a smaller set of parameters that you think you can determine from the data.

2) Based on what you know from the beginning and from the SVD analysis, come up with an initial estimate of the noise level σ, and a prior distribution $p(\theta)$ expressing what you think you know about the likely values of the parameters and their uncertainty ranges.

3) Determine a set of best-fit parameters for each of the two kinetic models. Compare the deviations between the data and the best-fit model with your expected noise level. Are the deviations centered about zero? Does their distribution look normal (i.e. Gaussian)?

Note: The parameters you would like to determine are the A_n's, $\lambda_{B,n}$'s, and w_n's in the Gaussian model for the absorbance, the initial conditions in the kinetic models, and the rate constants. Since you will definitely not be able to determine all these parameters independently, you will probably have to fix some parameters or impose some constraint, or in some other way reduce the number of parameters θ being adjusted. Explain clearly how you do this, and whether or not you think it reduces the ability of the fit to match the data.

Hint: Do some simple fits first, holding some parameters fixed at values estimated from the SVD analysis. Then use these refined values as initial guesses when you allow other parameters to float. You will probably need quite a good initial guess to solve the nonlinear least-squares problem if you allow more than 10 parameters to float simultaneously!

3) Determine if either of the kinetic models is plausibly consistent with the experimental data, based on a quantitative χ^2 test.

4) For the independent-reaction model, make a contour plot of χ^2 vs. (k_1,k_3) holding all the other parameters at their best-fit values. From this plot estimate the uncertainties in k_1 and k_3, and explain how you made the estimate. Compare this to the numbers you obtain from the covariance matrix (including all the parameters). These are both conventional approaches to estimate uncertainties, and to see correlations in uncertainties between two parameters.

5) Make a contour plot of $p_{Model}(Y_{data}|k_1,k_3)*p_{Prior}(k_1,k_3)$ vs. (k_1,k_3) holding all the other parameters at their best-fit values. From this Bayesian estimate of the shape of the new probability distribution $p(k_1,k_3)$ make an estimate of the uncertainties in k_1 and k_3. Explain.

Write a summary paragraph giving your conclusions about which model(s) can be made to fit the data, the corresponding best-fit values of the parameters, and your best estimate of the uncertainties in the parameters you obtained.

Cite as: William Green, Jr., course materials for 10.34 Numerical Methods Applied to Chemical Engineering, Fall 2006. MIT OpenCourseWare (http://ocw.mit.edu), Massachusetts Institute of Technology. Downloaded on [DD Month YYYY].