

CBE 20258 Numerical and Statistical Analysis

Algorithm Problem Set 4

due in class Thursday, 3/23

Error Propagation and Analysis in Non-Linear Regression

While linear regression is very convenient to use, sometimes (as in this week's tutorial) you just can't linearize a problem, or if you do there are some serious distortions to the data weighting. You can still define a residual, however, and can get fitting parameters which minimize the 2-norm, it's just that now you have to solve it via a non-linear optimization routine. We will study the theory behind such schemes later this semester, but for now just remember "fminsearch", which is Matlab's go-to routine for solving such problems.

This week you will employ the techniques you have learned in class to solve for the reaction kinetics of a phenolphthalein fading reaction, an experiment in senior lab. You are probably all aware that phenolphthalein is a base indicator (it turns a pretty purple), but in the presence of a strong base it will fade over time. This is a reversible reaction: pseudo first order in the forward (fading) direction (it's actually second order, but there is so much $[OH^-]$ that it can be simplified as pseudo first order) and first order in the reverse direction. The rate expression is:

$$\frac{d[P]}{dt} = -k_1[P][OH^-] + k_2[POH]$$

where $k_1' = k_1 [OH^-]$ is the pseudo first order rate. $[POH]$ is only produced from this reaction, thus we have the mass balance:

$$[POH] = [POH]_{t=0} + ([P]_0 - [P])$$

If we have the initial conditions:

$$[P]_{t=0} = [P]_0 \quad ; \quad [POH]_{t=0} = 0$$

then we can solve this equation to yield:

$$[P] = [P]_0 \left(\frac{k_2}{k_1' + k_2} + \frac{k_1}{k_1' + k_2} e^{-(k_1' + k_2)t} \right)$$

This is usually written as:

$$[P] = [P]_{eq} + ([P]_0 - [P]_{eq}) e^{-\lambda t}$$

where the two rate constants can be calculated from the ratio of the initial and equilibrium concentrations and the exponential decay rate.

The problem in senior lab (and really any time you do this sort of experiment) is that you don't know the equilibrium concentration: you simply can't wait long enough to get it, or the experimental conditions might not be stable that long. Without knowing this, you can't use linear regression – and even if you did and could, the weights tend to be pretty massively distorted as the concentration gets close to this equilibrium value, which has a significant amount of error in it anyway.

OK, now for the problem. Students introduce a dilute amount of phenolphthalein into a 0.1M NaOH solution. Using a colorimeter, students measure the following absorbances as a function of time:

Time(min)	Absorbance
0	1.1546
10	0.9456
20	0.8257
30	0.7442
40	0.631
50	0.5536
60	0.4738
70	0.4507
80	0.3671
90	0.3826
100	0.3754
110	0.2926
120	0.3105

Because the solution is so dilute, the absorbance obeys Beer's Law, such that the absorbance is proportional to the concentration. Since the problem is first order in [P] we don't even need the calibration factor: very convenient!

Your job is to determine the fitted values and matrix of covariance of the two reaction rates k_1' and k_2 and the pseudo equilibrium constant $K_e = k_1' / k_2$. Report the 95% confidence interval of K_e .

There are a lot of different ways to approach this problem, varying from straight up non-linear regression to a combination of linear and non-linear regression. Error calculations can be done using the straight up grad f approach, via MC, or via the jackknife. Your choice, just get the answer... Assume that all concentrations are measured to fixed absolute error (constant standard deviation in [P]). Don't forget to plot your residuals!