Day 7 Notes

Mann, J

September 17, 2015

1 Problem Set 1

1. Be careful with $\frac{\Delta f}{\Delta x} \approx f'$. Numerical derivatives of data can be really bad.
2. Think about Γ as $\Gamma[=]\frac{\text{molecules}}{(\text{nanometer})^2}$ or $\frac{1}{\Gamma} = A[=]\frac{(\text{nanometer})^2}{\text{molecules}}$, (in order to do this, you should use k_B instead of R (Boltzmann's constant instead of the ideal gas constant)
3. Inverting Power Series.
4. QUESTIONS: "Some Details on least squares?"

2 Some Details on Least Squares

2.1 General Least squares analysis

$\Lambda_n \equiv$ Weight factor. This is used to place an amount of confidence on your data.
$y_n \equiv$ Measured dependent variable data point
$x_n \equiv$ Measured independent variable data point

$$\text{Sum}(\vec{\alpha}) = \sum_{n=1}^{N} \Lambda_n (y_n - f(\vec{\alpha}, x_n))^2$$

This was developed by Gauss, and the data is assumed to follow a Gaussian distribution function it its errors.

$$\epsilon(x) \equiv \text{A Gaussian distribution with mean: } \mu = 0$$
$$y = f(\vec{\alpha}, x) + \epsilon(x)$$
$$\text{now } \Lambda_n = \left(\frac{\sigma}{\sigma_n}\right)^2$$

Where σ is the standard deviation, and σ_n is the standard deviation of y_n.

The fit is done so that the variation $\delta\text{sum} = 0$. Compute on $\alpha_1, \alpha_2, \ldots$.

$$0 = \delta(\text{Sum}(\alpha)) = \sum_{n=1} \Lambda_n \delta(y_n - f(\vec{\alpha}, x_n))^2$$

This generates a matrix problem. Expand $f(\vec{\alpha}, x_n)$ in a Taylor's series keeping only the 1st order term in. $\Delta\vec{\alpha} = \vec{\alpha} - \alpha_0$. (Exchange $\sum_n$ and sum over $\alpha_1, \alpha_2, \ldots$)

2.2 Suppose you plot Sum vs $\vec{\alpha}$

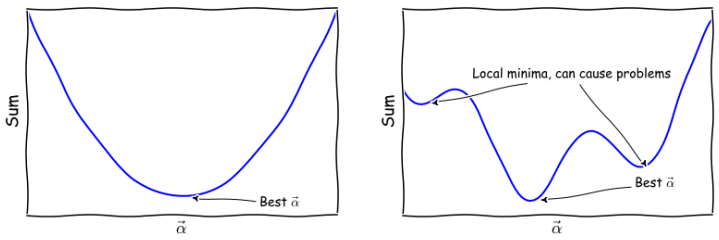

Figure 1: A plot of two different error functions with respect to the sum. Note that this is oversimplified. Because $\vec{\alpha}$ is a vector, this will most often be an n-dimensional plot.

If you are not careful, you may find the wrong minimum in some cases. For example, consider Figure 1

2.3 Matrix approach

The result of $\delta\text{Sum} = 0$ is a matrix problem:

$$\begin{aligned} M_{ij} &= \sum_{n=1}^{N} \Lambda_n \frac{\partial f}{\partial \alpha^i} \frac{\partial f}{\partial \alpha^j}|_{x_n, \vec{\alpha_0}} \\ \Delta\vec{\alpha} &= \vec{\alpha} - \alpha_0 \\ \boldsymbol{M}\Delta\vec{\alpha} &= \vec{R} \end{aligned}$$

This is the product matrix, some people call it the Hessian matrix. ***Note:*** A couple of properties to remember:

$$\begin{aligned} M_{ij} &= M_{ji} \\ \det(\boldsymbol{M}) &\neq 0 \end{aligned}$$

This will be true when the ratio of $\frac{\text{maximum Eigenvalue}}{\text{minimum Eigenvalue}}$ is small enough.

$$R_i = \sum_{n=1}^{N} \Lambda_n (y_n - f(\vec{\alpha}, x_n)) \frac{\partial f}{\partial \alpha_i}\Big|_{x,\alpha_0}$$
$$\boldsymbol{M}\vec{\alpha} = \vec{R}$$
$$\vec{\alpha} = \boldsymbol{M}^{-1}\vec{R}$$

This is a matrix problem solved iteratively until $\delta\vec{\alpha} \to 0$, so we need a good $\vec{\alpha}_0$.

For converged $\boldsymbol{M}$,

$$(\boldsymbol{M}^{-1})_{ii} \to \sigma_{ii}^2$$
$$\text{e.g.}(\boldsymbol{M}^{-1})_{11} = \sigma_{\alpha_1}^2$$

Consider the Eigenvalue problem with $\boldsymbol{M}$, this gives you a set of Eigenvalues, one of which is a maximum, you divide that by the minimum eigenvalue, and you get the condition number, c_N.

$c_N \approx 1$, good fit
≈ 10 ok
≈ 100 ok
$\approx 10^4$ start being nervous
$> 10^5$ be very careful

Plot the data! Plot the residuals $(y_n - f(\vec{\alpha}, x_n))$! The residuals plot should look random and centered around zero.

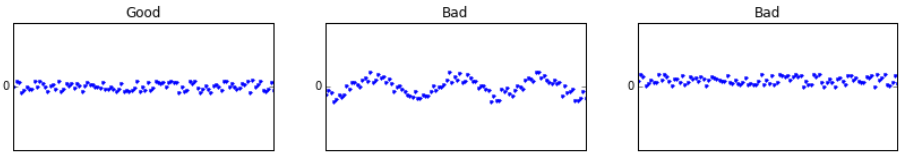

Figure 2: Three plots of the residuals. The first is good, the second shows correlation, and the third shows constant offset.

2.4 Summary

1. The fit converges and the condition number is reasonable.
2. Look at the residuals: Are they Gaussian? Plot them.
3. Are the uncertainties of the parameters reasonable?

4. The correlations are insignificant: ρ_{ij} is the correlation of parameter α_i with α_j.

$$\rho_{ij} = \frac{\boldsymbol{M}^{-1}}{\sqrt{(\boldsymbol{M}^{-1})_{ii}}\sqrt{(\boldsymbol{M}^{-1})_{jj}}}, -1 \leq \rho_{ij} \leq 1$$

   - $\rho_{ij} \approx 1$, α_i is a positive correlation with α_j
   - $\rho_{ij} \approx -1$, α_i is a negative correlation with α_j
   - $\rho_{ij} \in [-0.5, 0.5]$, probably okay.
   - $\rho_{ij} \notin [-0.5, 0.5]$, be very careful.

Note: If the fit is very good, then

$$\frac{1}{N-P}\sum_{n=1}^{N}\Lambda_n(y_n - f(\vec{\alpha}, x_n))^2 \to 1$$

as the number of items, N, is large; $N \to \infty$. P is the number of parameters. $N - P$ is the number of degrees of freedom for the fit.

Note: If the condition number is between 10 and 1000, your numerical inversion of $\boldsymbol{M}$ will be accurate. Start to be concerned if $c_N > 1000$, but $\boldsymbol{M}^{-1}$ may still be okay. $c_N > 10000$ shows that you should be concerned about the result. (This is from experience)

3 Notes about the homework

Temperature was in Kelvin. Using origin, there is a built-in algorithm for differentiation that is very nice.

Analysis→Mathematice→Differentiate→Open Dialog

Choose the two columns that you will be using. There is a smoothing operation (it's pretty good). Look carefully at the pattern of your data, and make sure that it's smooth if it is supposed to be smooth. $\gamma = \gamma_\circ - kT\Gamma_{\max}\ln(1 + ac)$ Things to look for in a good fit:

- The inverse hessian matrix (Look up the difference between Inverse Hessian and correlation matrix)

4 XRC, X-Ray Crystallography

$1/\Gamma_{\text{Max}}$ is the surface area of the molecule. NIH Has a crystallographic database that is useful in getting cross-sectional area of unit cells of a molecule, in case you need to use biological cells.

Suppose you need to know the concentration you need to put in a spray pesticide that is surface-active.

How would you get that?

Know what the molecule is, do an estimation of what the cross-sectional area of a molecule is as it goes into the interface, if it's long, get a piece of the molecule, then get geometric information about how it's going to pack.

5 Inverting power series