Data Fitting

Physics 3110
Dr. Rebecca Forrest
University of Houston

The Art of Experimental Physics, D. Preston & E. Dietz, NY, John Wiley, (1991), pp. 18 - 22

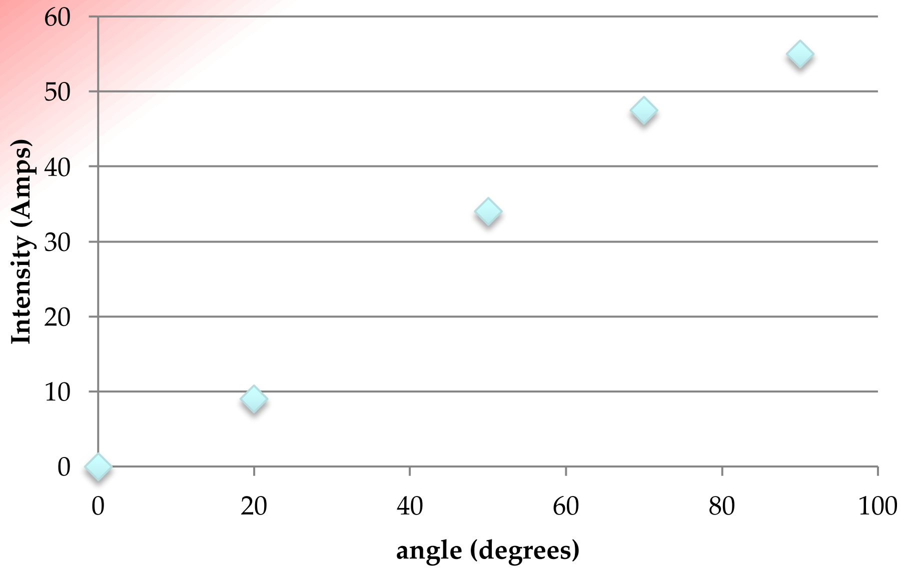
Plotting & Fitting on the Computer

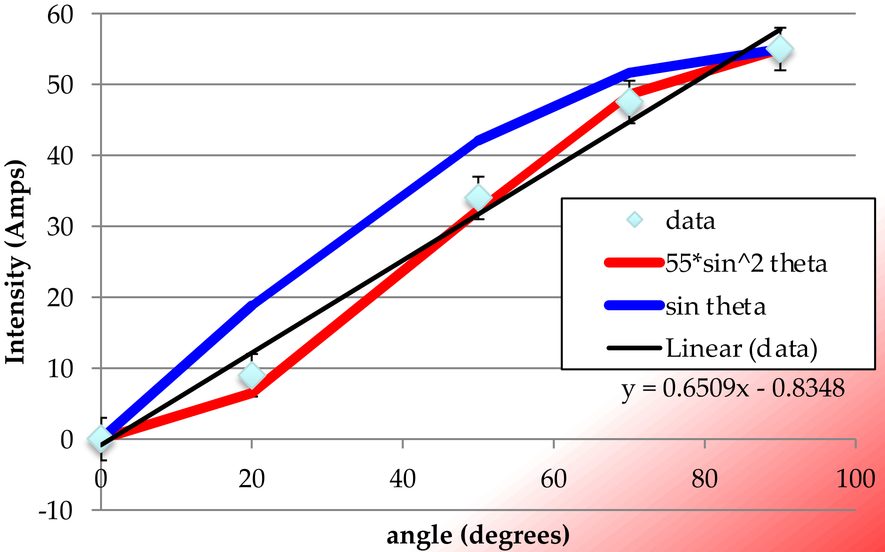
- Excel
- Gnuplot – http://www.gnuplot.info/
- PSI plot
- Mathcad
- Mathmatica
- SciDavis
- Origin
- Etc.

Trend Analysis & Fitting

- Trying to show that the data follows some formula, i.e. linear, sine, x^{-1}…
- Fitting your data to get a numerical result from the fit

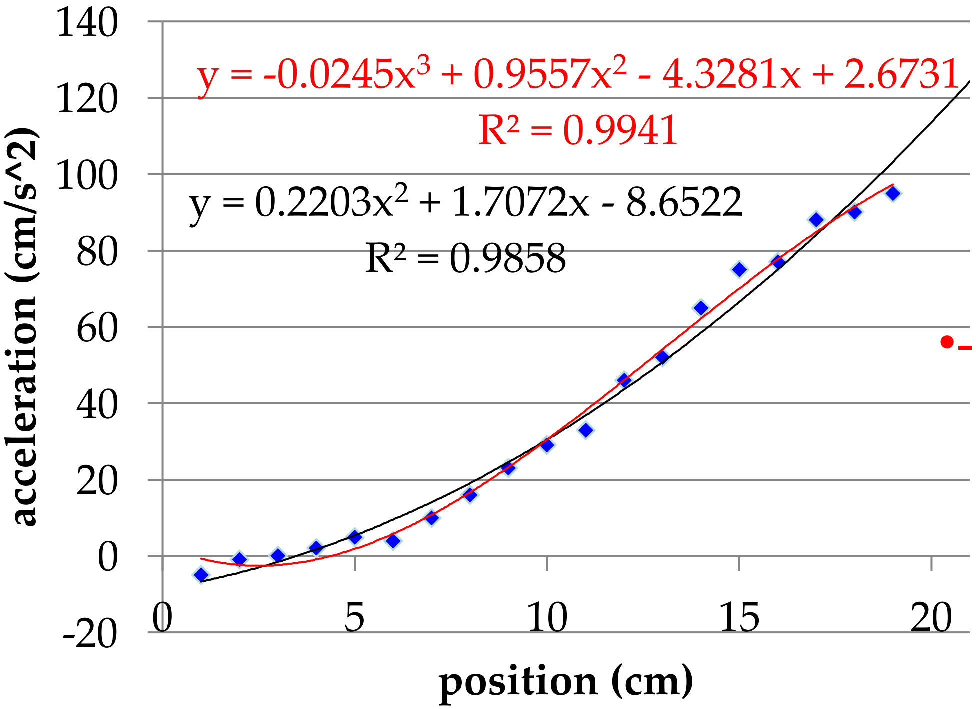
Trend Analysis





Data Fitting

- A set of observations/data are given
- You want to fit a “model” function to the data
- Figure-of-merit function measures agreement between the data and the model



Data

Fit 1, $y = a_1 x^2 + a_2 x + a_3$

Fit 2, $y = a_1 x^3 + a_2 x^2 + a_3 x + a_4$

- $y(x_i; a_1 \ldots a_m)$

Fitting with Computer Software

- Most common approach is Least Squares Fitting
- Excel
  - Chart: Add Trendline
  - Limited function choices
  - Goodness of fit: R-squared
- Mathematica
  - Fit[data,funs, vars]
  - Goodness of fit: "χ^2" = $\Sigma_i |F_i – f_i|^2$, sum of residuals
- Origin
  - Several Choices
- Gnuplot
- SciDavis
- ….

Least Squares Fitting

Adapted from:

Numerical Recipes

The Art of Scientific Computing

W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery

Cambridge University Press 1992

New (and free Older versions) at

www.nr.com

Least Squares Fitting

•You want to fit a function to a set of data (x_i, y_i). Assume no error in independent variables (σ_x' s = 0) and errors in y' s, σ_y' s, are known. a_i's are parameters in function.

$$y(x) = y(x; a_1 \ldots a_M)$$

$$\sum_{i=1}^{N} [y_i - y(x_i; a_1 \ldots a_M)]^2$$

• sum of the residuals should be small

Central Limit Theorem

- For large enough N, the measurement errors follow a Gaussian distribution with standard deviations σ
- Minimize χ^2:

$$\chi^2 \equiv \sum_{i=1}^{N} \left(\frac{y_i - y(x_i; a_1 \ldots a_M)}{\sigma_i} \right)^2$$

Minimize χ^2

$$\chi^2 \equiv \sum_{i=1}^{N} \left(\frac{y_i - y(x_i; a_1 \dots a_M)}{\sigma_i} \right)^2$$

Solve $\dfrac{\partial(\chi^2)}{\partial a_i} = 0$

- To apply this, we need to know the function $y(x_i; a_1 \dots a_m)$

Example:
Least Squares Fitting to a Straight Line

- Also called linear regression

$$y(x) = y(x; a, b) = a + bx$$

$$\chi^2(a,b) = \sum_{i=1}^{N} \left(\frac{y_i - a - bx_i}{\sigma_i} \right)^2$$

- Minimize χ^2: Solve $\dfrac{\partial(\chi^2)}{\partial a_i} = 0$

Taking Derivatives

$$\chi^2(a,b) = \sum_{i=1}^{N} \left(\frac{y_i - a - bx_i}{\sigma_i} \right)^2$$

$$0 = \frac{\partial \chi^2}{\partial a} = -2 \sum_{i=1}^{N} \frac{y_i - a - bx_i}{\sigma_i^2} \quad = -2(S_y - aS - bS_x)$$

$$0 = \frac{\partial \chi^2}{\partial b} = -2 \sum_{i=1}^{N} \frac{x_i(y_i - a - bx_i)}{\sigma_i^2} \quad = -2(S_{xy} - aS_x - bS_{xx})$$

$$S \equiv \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \quad S_x \equiv \sum_{i=1}^{N} \frac{x_i}{\sigma_i^2} \quad S_y \equiv \sum_{i=1}^{N} \frac{y_i}{\sigma_i^2}$$

$$S_{xx} \equiv \sum_{i=1}^{N} \frac{x_i^2}{\sigma_i^2} \quad S_{xy} \equiv \sum_{i=1}^{N} \frac{x_i y_i}{\sigma_i^2}$$

$$aS + bS_x = S_y$$

$$aS_x + bS_{xx} = S_{xy}$$

- Find a & b

Solution to Linear System

$$\Delta \equiv SS_{xx} - (S_x)^2$$

$$a = \frac{S_{xx}S_y - S_xS_{xy}}{\Delta}$$

$$b = \frac{SS_{xy} - S_xS_y}{\Delta}$$

•Now you have a & b that give the best fit to your data. What are the errors in a & b?

Propagation of Errors
Errors in a & b

$$\delta w^2 = \sum_i \left(\frac{\partial w}{\partial x_i}\delta x_i\right)^2, \qquad \sigma_f^2 = \sum_{i=1}^{N} \sigma_i^2 \left(\frac{\partial f}{\partial y_i}\right)^2$$

$$a = \frac{S_{xx}S_y - S_x S_{xy}}{\Delta}$$

$$b = \frac{S S_{xy} - S_x S_y}{\Delta}$$

$$\frac{\partial a}{\partial y_i} = \frac{S_{xx} - S_x x_i}{\sigma_i^2 \Delta}$$

$$\frac{\partial b}{\partial y_i} = \frac{S x_i - S_x}{\sigma_i^2 \Delta}$$

- Variances in the Estimates

$$\sigma_a^2 = S_{xx}/\Delta$$

$$\sigma_b^2 = S/\Delta$$

Goodness of Fit

- Sum of residuals
  - should be close to 0

$$\sum_{i=1}^{N} [y_i - y(x_i; a_1 \ldots a_M)]^2$$

- χ^2

$$\chi^2 \equiv \sum_{i=1}^{N} \left(\frac{y_i - y(x_i; a_1 \ldots a_M)}{\sigma_i} \right)^2$$

  - should be small, $\chi^2 \sim \nu$, where ν = degrees of freedom = number of data points minus the number of parameters being fit
- Reduced $\chi^2 = \chi^2 / \nu$
  - $\chi^2/\nu \sim 1.0$ is good
- Others …

Other Popular Methods

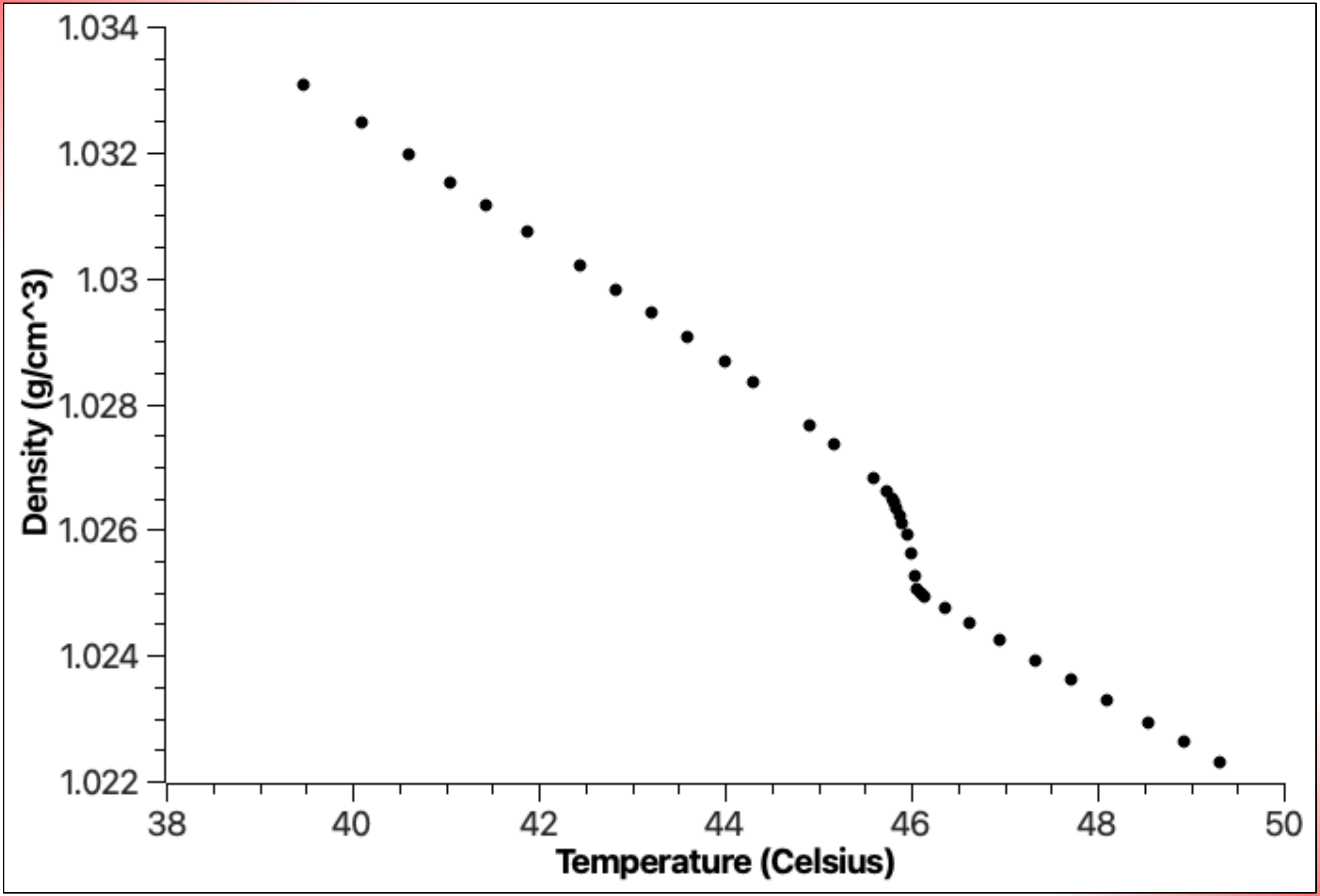
- If 1st and 2nd derivatives are known: *Levenberg-Marquard* method
- If derivatives are not known and have to be approximated numerically: *Downhill-Simplex* or *Powell* method; in those cases, you can not get correlations or goodness of fit

Using *SciDAVis* for Fitting

- Search for SciDAVis. Download it.
- We'll fit a datafile called lcdemo.dat

SciDAVis

- Import data. If it doesn’t work, you can create the table and copy and paste the data.
- Plot the data (may include error bars)
- Select graph, choose Analysis/Fit Wizard
- Analysis/Quick Fit
- Add error bars/Fit
- Analysis/Fit Wizard


1.034
1.032
1.03
1.028
1.026
1.024
1.022
Density (g/cm^3)
38
40
42
44
46
48
50
Temperature (Celsius)

Homework

1. Using Excel, fit the data in linedata.dat. Show the data and the fit in one plot.
2. In a different computer program of your choice, fit the following:
   - linedata.dat to a linear function, taking into account the error bars in the 3rd column.
   - Gauss2.dat to this Gaussian distribution, taking into account the error bars in the 3rd column.

$$N(x) = \frac{A}{\sqrt{2\pi\sigma}} e^{\frac{-(x-\overline{x})^2}{2\sigma^2}}$$

See handout for details.