Graphical Analysis and Fitting

Physics 3110
Dr. Rebecca Forrest
University of Houston

The Art of Experimental Physics, D. Preston & E. Dietz,
Plotting on the Computer

• Excel
• Gnuplot – http://www.gnuplot.info/
• PSI plot
• Mathcad
• Mathematica
• Origin
• SciDavis
• Etc.
Graph Guidelines

- Graphs should be big and clear, with data symbols, numbers, and labels that can be easily read.
- Label axes, include units.
- Use symbols to indicate data points.
- Graphs, like all figures, must have a caption that explains their contents.
- Graphs, like all figures, must be referred to in the text by #, in order of appearance.
Graph Guidelines

• Use error bars to indicate errors in measurements.

• DO NOT connect points with straight lines (almost never)

• If you are trying to show that your data is described by a certain function, i.e. linear, sine, etc., you need to either show the function on the same plot, or fit the data.
“...Taking account of the nondegeneracy for $n \leq 2$ gives the solid curve in Fig. 1, which includes prominent well known resonances. Including nondegeneracy for $n \leq 4$ [26] gives the dotted curve in Fig. 1.”


FIG. 1. Energy dependence of the differential cross section for $\pi^+$ photoproduction at $\theta = 90^\circ$. The solid curve denotes degeneracy breaking for $n \leq 2$, while the dotted for $n \leq 4$. The empty circles are old data from Ref. [20], and the solid dots are new data from JLab [21].
Examples
Figure 1 displays the data points along with the best fit model.

**Good**
- Caption
- Fig. 1 is mentioned by name in text above.

**Bad**
- All fonts too small
- No error bars
- Too much info in Caption.
Figure 2: Displays the slope of $\log T$ vs. $\log E$. A linear fit is placed on our data and we take a linear fit to get the slope equivalent to approximately four for the Stefan-Boltzmann experiment.

**Good**
- Caption
- Axes are labeled and units are shown
- Legend

**Bad**
- No data symbols shown
- Instead, data points connected with lines
The first part of our experiment showed a distinct wave-like pattern that we would expect. This shape can be seen in Fig. 1.

---

**Good**
- Caption
- Fig. 1 is mentioned by name in text above.
- Axes are labeled
- Data symbols shown with error bars

**Bad**
- All fonts too small
- Axis units not labeled
- This is an exception about connecting data points. Without it, the data trend is not obvious.
- Meter reading vs. distance
Figure 1: FWHM (18.0,0.15), a point was made to on point (185,0.15) to represent the fwhm while the rest are data points.
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

Intensity (Amps)

angle (degrees)

Intensity (Amps)

angle (degrees)

peak

$\sin^2 \theta$

$\sin \theta$
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

\[ y = 0.2203x^2 + 1.7072x - 8.6522 \quad R^2 = 0.98583 \]

\[ y = -0.0245x^3 + 0.9557x^2 - 4.3281x + 2.6731 \quad R^2 = 0.99411 \]

\[ y(x_i; a_1 \ldots a_m) \]
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: \( \chi^2 = \sum_i |F_i - f_i|^2 \), sum of residuals
- Origin
  - Several Choices
- Gnuplot
- SciDavis
- ....
Least Squares Fitting

Adapted from:

Numerical Recipes
The Art of Scientific Computing
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 \((\sigma_x ' s = 0)\) and errors in \(y' s, \sigma_y ' s\), 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 $\sigma$
• Minimize $\chi^2$:

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

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

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

• To apply this, we need to know the function \( y(x_i; a_1 \ldots a_m) \)
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 \( \chi^2 \): Solve \( \frac{\partial(\chi^2)}{\partial a_i} = 0 \)
Taking Derivatives

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

\[ 0 = \frac{\partial \chi^2}{\partial a} = -2 \sum_{i=1}^{N} \frac{y_i - a - bx_i}{\sigma_i^2} \]

\[ = -2(S_y - aS - bS_x) \]

\[ 0 = \frac{\partial \chi^2}{\partial b} = -2 \sum_{i=1}^{N} \frac{x_i(y_i - a - bx_i)}{\sigma_i^2} \]

\[ = -2(S_{xy} - aS_x - bS_{xx}) \]

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

\[ S_{xx} \equiv \sum_{i=1}^{N} \frac{x_i^2}{\sigma_i^2} \quad S_{xy} \equiv \sum_{i=1}^{N} \frac{x_iy_i}{\sigma_i^2} \]

•Find a & b

\[ aS + bS_x = S_y \]

\[ aS_x + bS_{xx} = S_{xy} \]
Solution to Linear System

\[ \Delta \equiv SS_{xx} - (S_x)^2 \]
\[ a = \frac{S_{xx} S_y - S_x S_{xy}}{\Delta} \]
\[ b = \frac{SS_{xy} - S_x S_y}{\Delta} \]

• 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

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

\[ \frac{\partial a}{\partial y_i} = \frac{S_{xx} - S_x x_i}{\sigma_i^2 \Delta} \]
\[ \frac{\partial b}{\partial y_i} = \frac{S x_i - S_x}{\sigma_i^2 \Delta} \]

\[ a = \frac{S_{xx} S_y - S_x S_{xy}}{\Delta} \]
\[ b = \frac{SS_{xy} - S_x S_y}{\Delta} \]

• 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

• $\chi^2$
  – should be small, $\chi^2 \sim \nu$, where $\nu = \text{degrees of freedom} = \text{number of data points} - \text{number of parameters being fit}$

• Reduced $\chi^2 = \chi^2 / \nu$
  – $\chi^2/\nu \sim 1.0$ is good

• Others …
Other Popular Methods

• If 1\textsuperscript{st} and 2\textsuperscript{nd} derivatives are known: \textit{Levenberg-Marquard} method

• If derivatives are not known and have to be approximated numerically: \textit{Downhill-Simplex} or \textit{Powell} method; \textbf{in those cases, you can not get correlations or goodness of fit}
Using *Gnuplot* for Fitting

- [http://www.gnuplot.info/](http://www.gnuplot.info/)
- *Gnuplot* is a **portable** command-line driven interactive data and function plotting utility for UNIX, IBM OS/2, MS Windows, DOS, Macintosh, VMS, Atari and many other platforms. The software is copyrighted but freely distributed.
- For MS Windows, download the file with win32 in its name.
- *Gnuplot’s* fit uses the nonlinear least-squares Marquardt-Levenberg algorithm.
Windows GUI

GNUPLOT
Version 4.2 patchlevel 3
Last modified Mar 2008
System: MS-Windows 32 bit

Thomas Williams, Colin Kelley and many others

Type 'help' to access the on-line reference manual.
The gnuplot FAQ is available from http://www.gnuplot.info/faq/

Send bug reports and suggestions to <http://sourceforge.net/projects/gnuplot>

Terminal type set to 'windows'
gnuplot> _
File/Demos
Fit Demo for Density data

- Plot ‘lcdemo.dat’
Unweighted Linear Fit

- Fit \( l(x) \) from `lcdemo.dat` via \( y_0, m \)
- Plot `lcdemo.dat`, \( l(x) \)
fit l(x) 'lcdemo.dat' using 1:2:3 via y0, m
• fit l(x) 'lcdemo.dat' using 1:2:4 via y0, m
• plot 'lcdemo.dat' using 1:2:5 with errorbars
fitted to realistic model function

Density [g/cm³]

Temperature T [°C]

47.2875, 1.03612

Ice density

OK Cancel
Final Parameters and Error

---

**resultant parameter values**

\[\begin{align*}
\lambda_l &= -0.00190094 \\
\lambda_h &= -0.000031266 \\
\text{dens}_T &= 1.02497 \\
\text{Tc} &= 46.8899 \\
g &= 3.85585 \\
h &= 0.00153901
\end{align*}\]

**Iteration 41**

\[\begin{align*}
\text{USSR} &= 9.00221e-008 \\
\text{delta(USSR)} &= -1.16355e-010 \\
\lambda_l &= 1.07683e-021
\end{align*}\]

**resultant parameter values**

\[\begin{align*}
\lambda_l &= -0.00190094 \\
\lambda_h &= -0.000031266 \\
\text{dens}_T &= 1.02497 \\
\text{Tc} &= 46.8899 \\
g &= 3.85585 \\
h &= 0.00153901
\end{align*}\]

After 41 iterations the fit converged.

**Final sum of squares of residuals**: 9.00221e-008

**rel. change during last iteration**: -1.16355e-010

**degrees of freedom (df)**: 31

**variance of residuals (reduced chi-square)**: 5.38882e-005

**Final set of parameters**

---

\[\begin{align*}
\lambda_l &= -0.00190094 \pm 7.816e-006 \langle 0.9815\% \rangle \\
\lambda_h &= -0.000031266 \pm 1.544e-005 \langle 1.8515\% \rangle \\
\text{dens}_T &= 1.02497 \pm 3.171e-005 \langle 0.04114\% \rangle \\
\text{Tc} &= 46.8899 \pm 0.007592 \langle 0.01647\% \rangle \\
g &= 3.85585 \pm 6.2224 \langle 0.5782\% \rangle \\
h &= 0.00153901 \pm 4.101e-005 \langle 0.2655\% \rangle
\end{align*}\]

**correlation matrix of the fit parameters**:

\[\begin{pmatrix}
1.000 & & & & \\
\lambda_l & 0.000 & & & \\
\lambda_h & 0.000 & 1.000 & & \\
\text{dens}_T & 0.000 & 0.000 & 1.000 & \\
\text{Tc} & 0.000 & 0.000 & 0.000 & 1.000 \\
g & 0.000 & 0.000 & 0.000 & 0.000 & 1.000 \\
h & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.000 \\
\end{pmatrix}\]
Homework

1. (3 pts.) Plot the data in data.dat with error bars, found on Blackboard, using a computer program. The 3rd column is the error. The plot should have the appropriate format, but no caption is required.

2. (7 pts.) Fit the data in data.dat to this Gaussian distribution,

\[ N(x) = \frac{A}{\sqrt{2\pi}\sigma} e^{-\frac{(x-x_{\text{ave}})^2}{2\sigma^2}} \]

- In computer syntax, this could be written as
  \[ N(x) = \frac{A}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-x_{\text{ave}})^2}{2\sigma^2}\right) \]
- Show a plot of the data with error bars and the fit, and report the fitted values of A, \(\sigma\), and \(x_{\text{ave}}\). The plot should have the appropriate format, but no caption is required. (This may count as the plot for part 1.)
- If available, report the errors in A, \(\sigma\), and \(x_{\text{ave}}\), and an estimate of the goodness of the fit, i.e. reduced \(\chi^2\) or \(R^2\).
- Tell me what program you used.