Graphical Analysis and 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 on the Computer

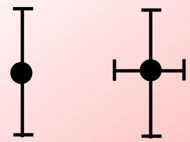
- Excel
- Gnuplot – http://www.gnuplot.info/
- PSI plot
- Mathcad
- Mathmatica
- Origin
- SciDavis
- Etc.

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.

Graphs in Your Reports

"…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."

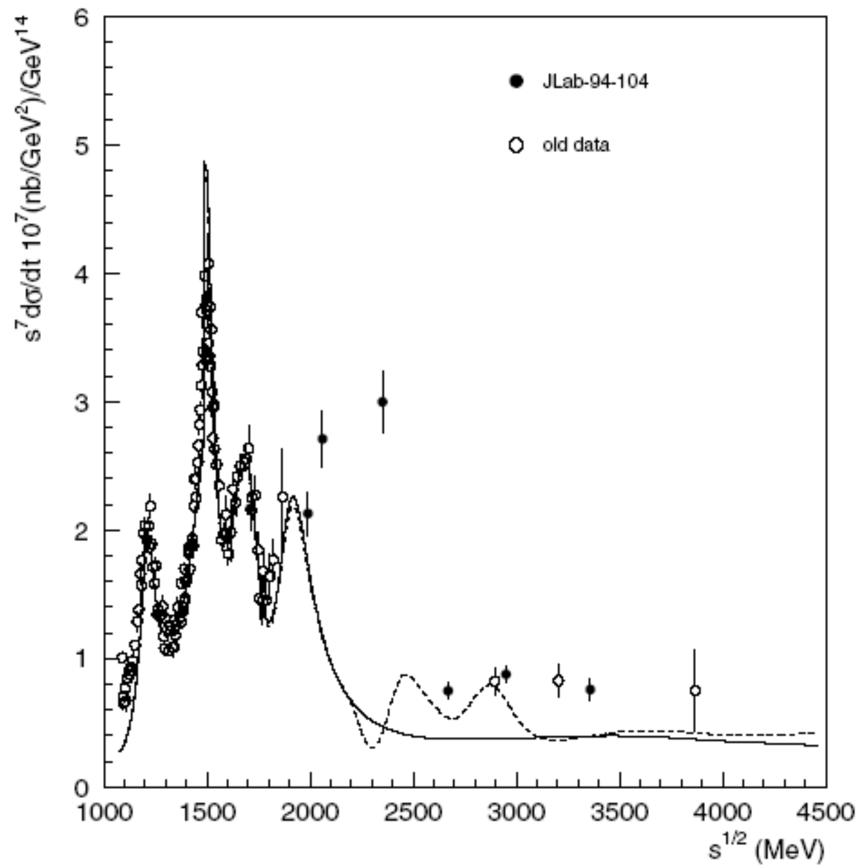


FIG. 1. Energy dependence of the differential cross section for π^+ photoproduction at $\theta = 90°$. The solid curve denotes degeneracy breaking for $n \leq 2$, while the dotted for $n \leq 4$. The empty circles are old data from Ref. [20], and the solid dots are new data from JLab [21].

"Locality of Quark-Hadron Duality and Deviations from Quark Counting Rules Above the Resonance Region", Qiang Zhao1 and Frank E. Close,Phys. Rev. Lett., 022004, **91**, 2003.

Examples

Figure 1 displays the data points along with the best fit model.

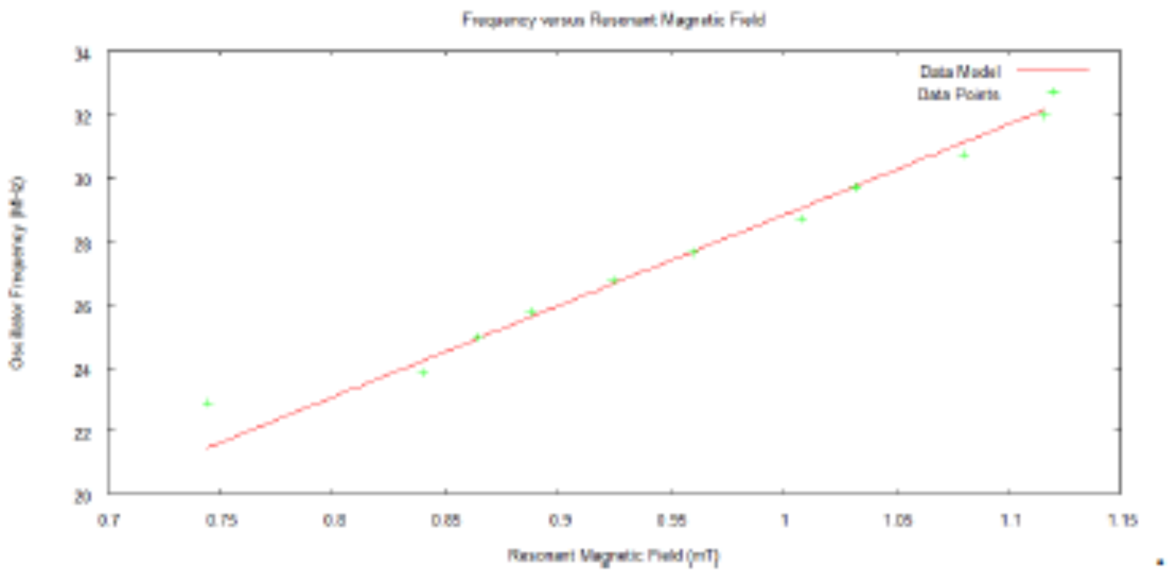


FIG. 1. Setting equation 8 equal to $h\upsilon$ and solving for υ gave rise to an equation suitable for finding μ_s using the least sum of squares method for a linear equation in GNUPlot.

<u>Good</u>

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

<u>Bad</u>

- All fonts too small
- No error bars
- Too much info in Caption.

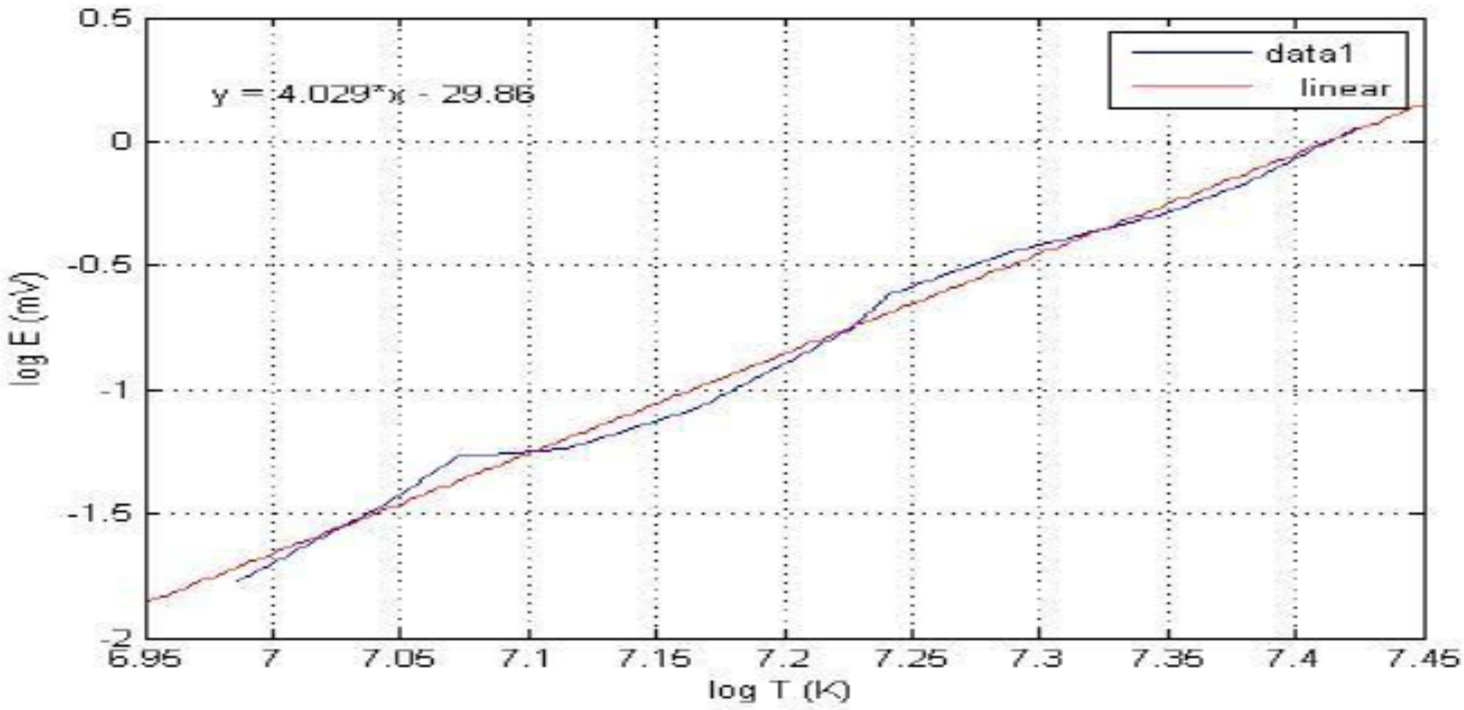


Figure 2: Displays the slope of log T vs. log E. A linear fit is placed on our data and we take a linear fit to get the slope equivalent to approximately four for the Stefan-Boltzmann experiment.

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.

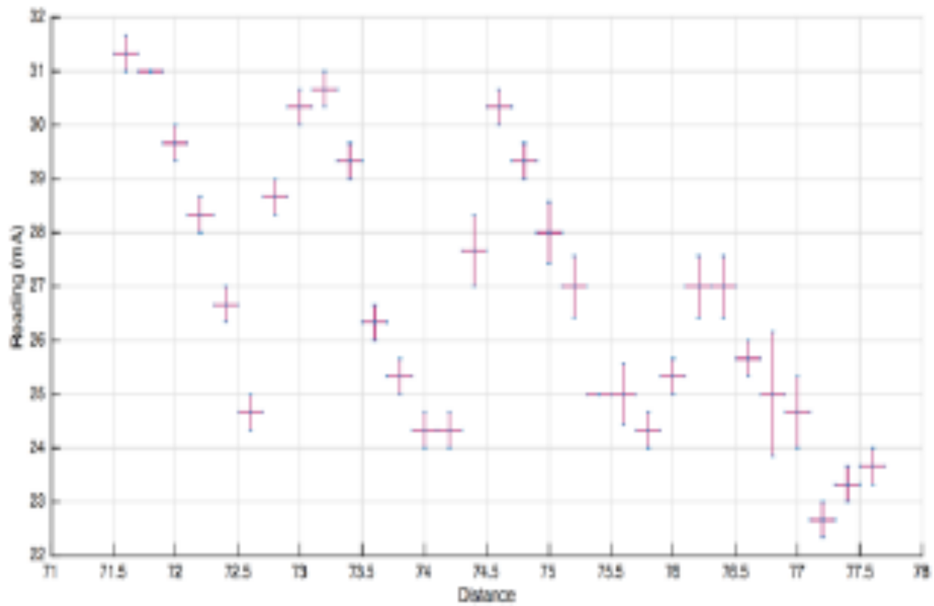


Fig. 1: Plot of distance versus meter reading for the first experiment

<u>Good</u>

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

<u>Bad</u>

- 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

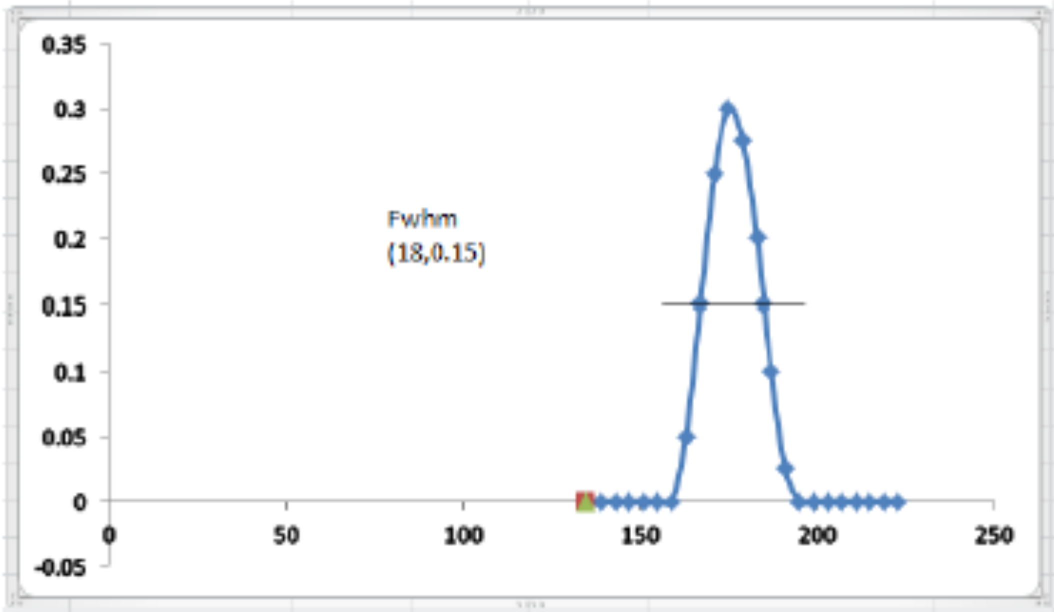


Figure 1: FWHM (18.0,0.15), a point was made to on point (185,0.15) to repersent the fwhm while the rest are data points.

Good

- Caption
- Data point symbols

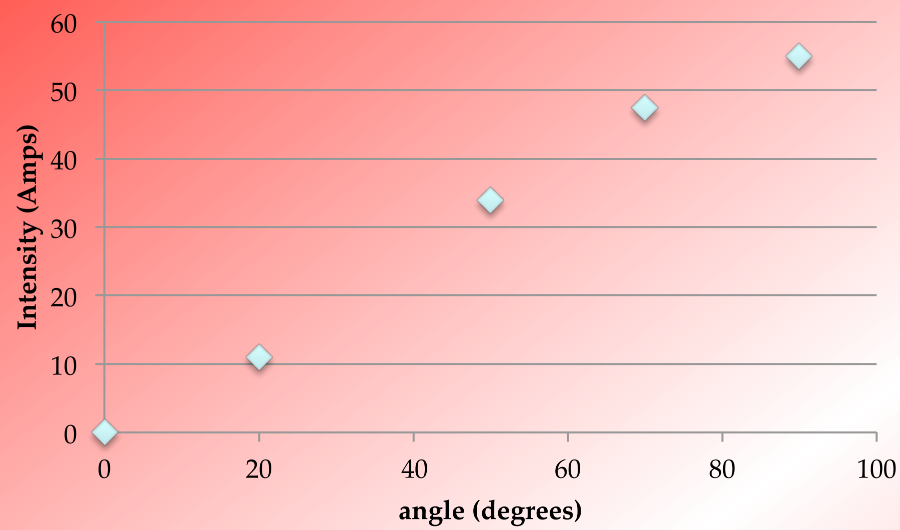
Bad

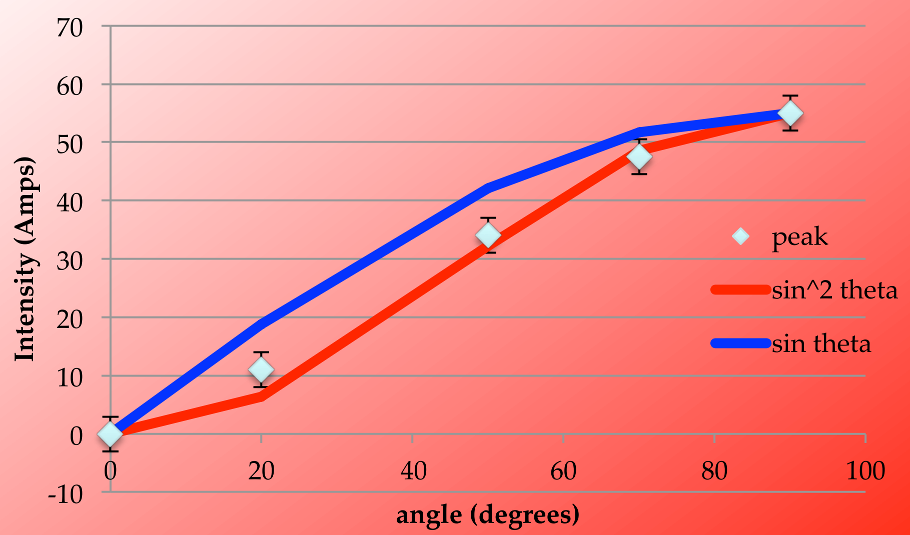
- Caption doesn't make sense
- No error bars
- No axis labels or units
- X-axis should focus on data of interest

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

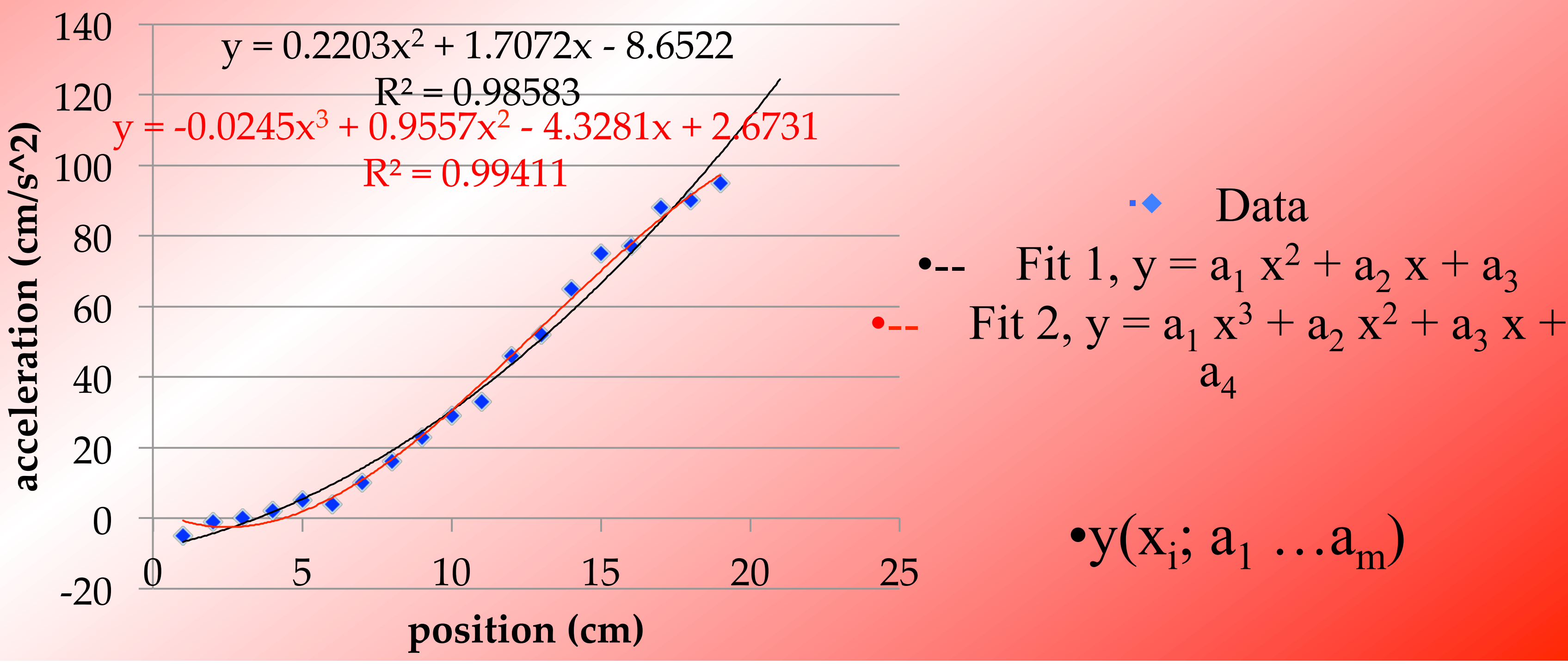
Trend Analysis





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



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

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.

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

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

• 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:

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

Minimize χ^2

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

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

Taking Derivatives

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

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

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

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

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

$$aS + bS_x = S_y$$

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

- Find a & b

Solution to Linear System

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

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

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

•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

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

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

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

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

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

- 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

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

- χ^2

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

  - 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 …

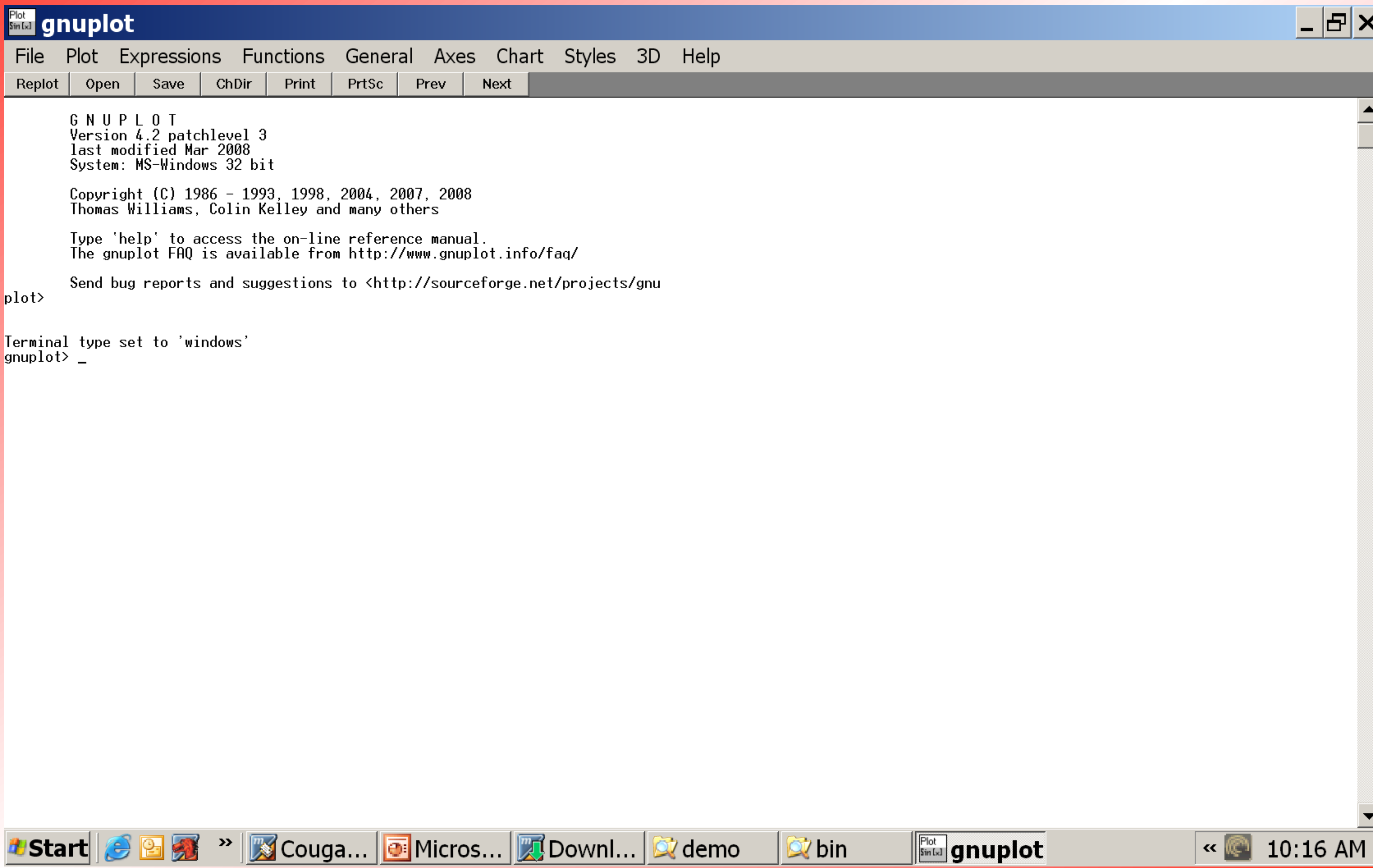
Other Popular Methods

- 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

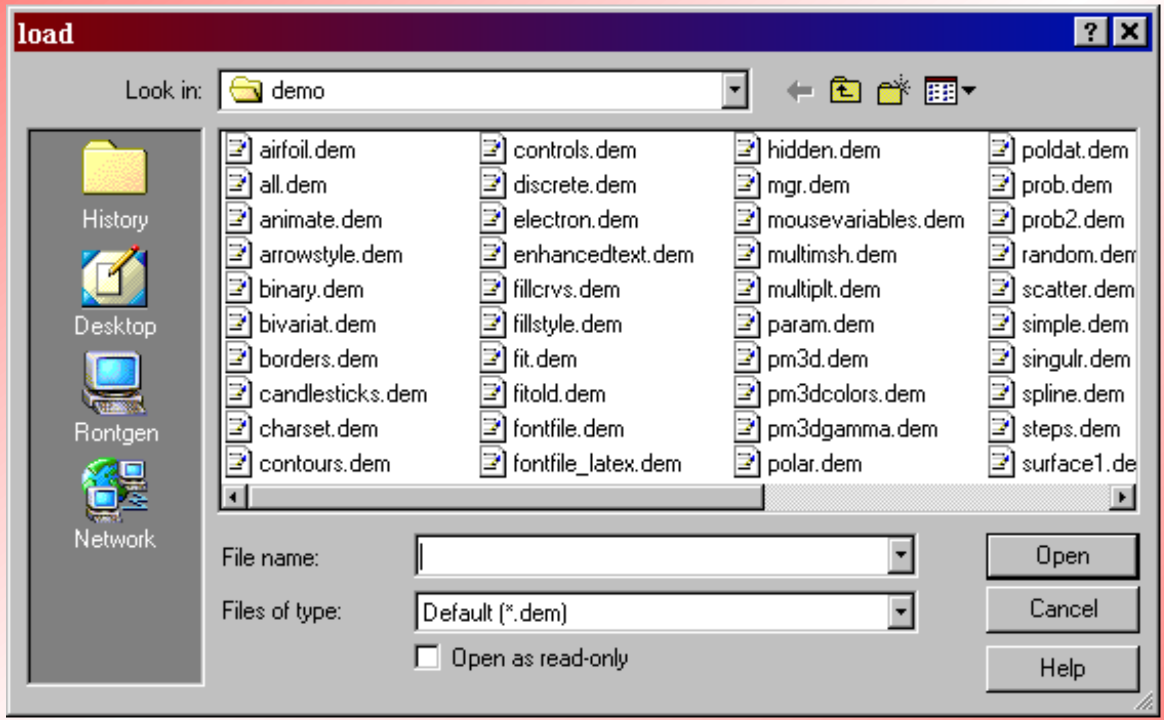
Using *Gnuplot* for Fitting

- 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

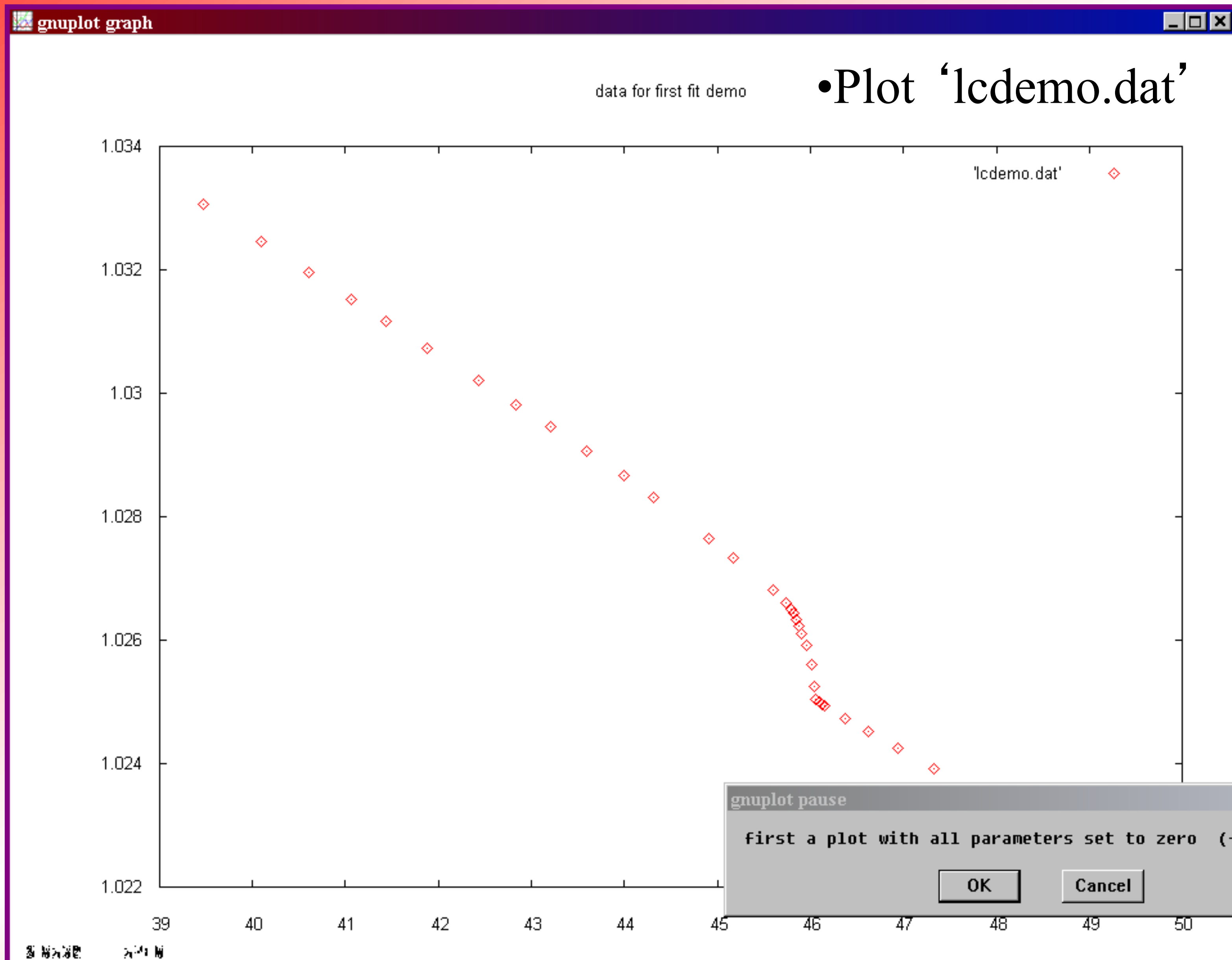

gnuplot
File Plot Expressions Functions General Axes Chart Styles 3D Help
Replot Open Save ChDir Print PrtSc Prev Next
G N U P L O T
Version 4.2 patchlevel 3
last modified Mar 2008
System: MS-Windows 32 bit
Copyright (C) 1986 - 1993, 1998, 2004, 2007, 2008
Thomas Williams, Colin Kelley and many others
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/gnu
plot>
Terminal type set to 'windows'
gnuplot> _
Start
Couga...
Micros...
Downl...
demo
bin
gnuplot
10:16 AM

File/Demos


load
Look in:
demo
History
Desktop
Rontgen
Network
airfoil.dem
all.dem
animate.dem
arrowstyle.dem
binary.dem
bivariat.dem
borders.dem
candlesticks.dem
charset.dem
contours.dem
controls.dem
discrete.dem
electron.dem
enhancedtext.dem
fillcrvs.dem
fillstyle.dem
fit.dem
fitold.dem
fontfile.dem
fontfile_latex.dem
hidden.dem
mgr.dem
mousevariables.dem
multimsh.dem
multiplt.dem
param.dem
pm3d.dem
pm3dcolors.dem
pm3dgamma.dem
polar.dem
poldat.dem
prob.dem
prob2.dem
scatter.dem
simple.dem
singulr.dem
spline.dem
steps.dem
File name:
Files of type:
Default (*.dem)
Open as read-only
Open
Cancel
Help

Fit Demo for Density data

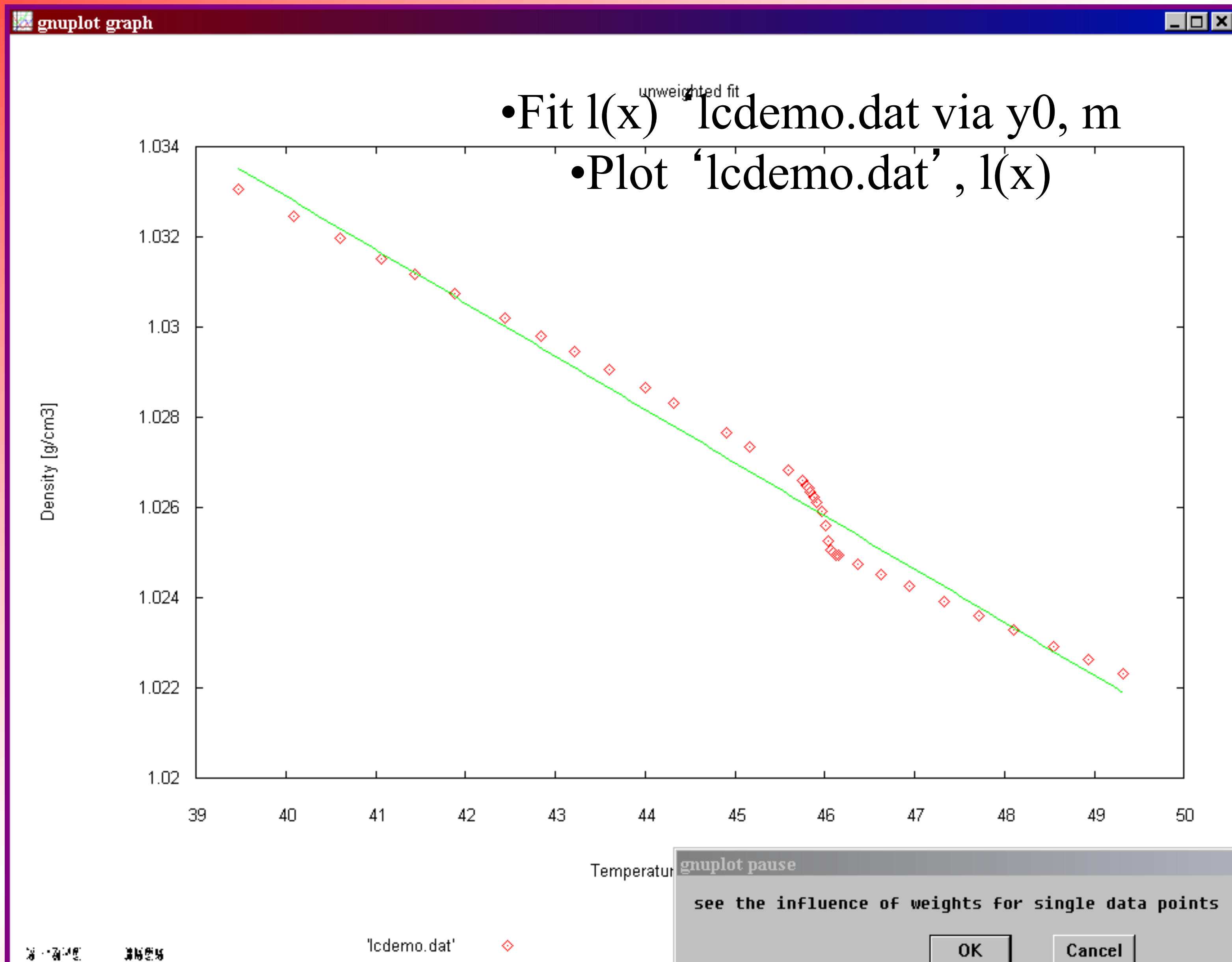
- Plot 'lcdemo.dat'



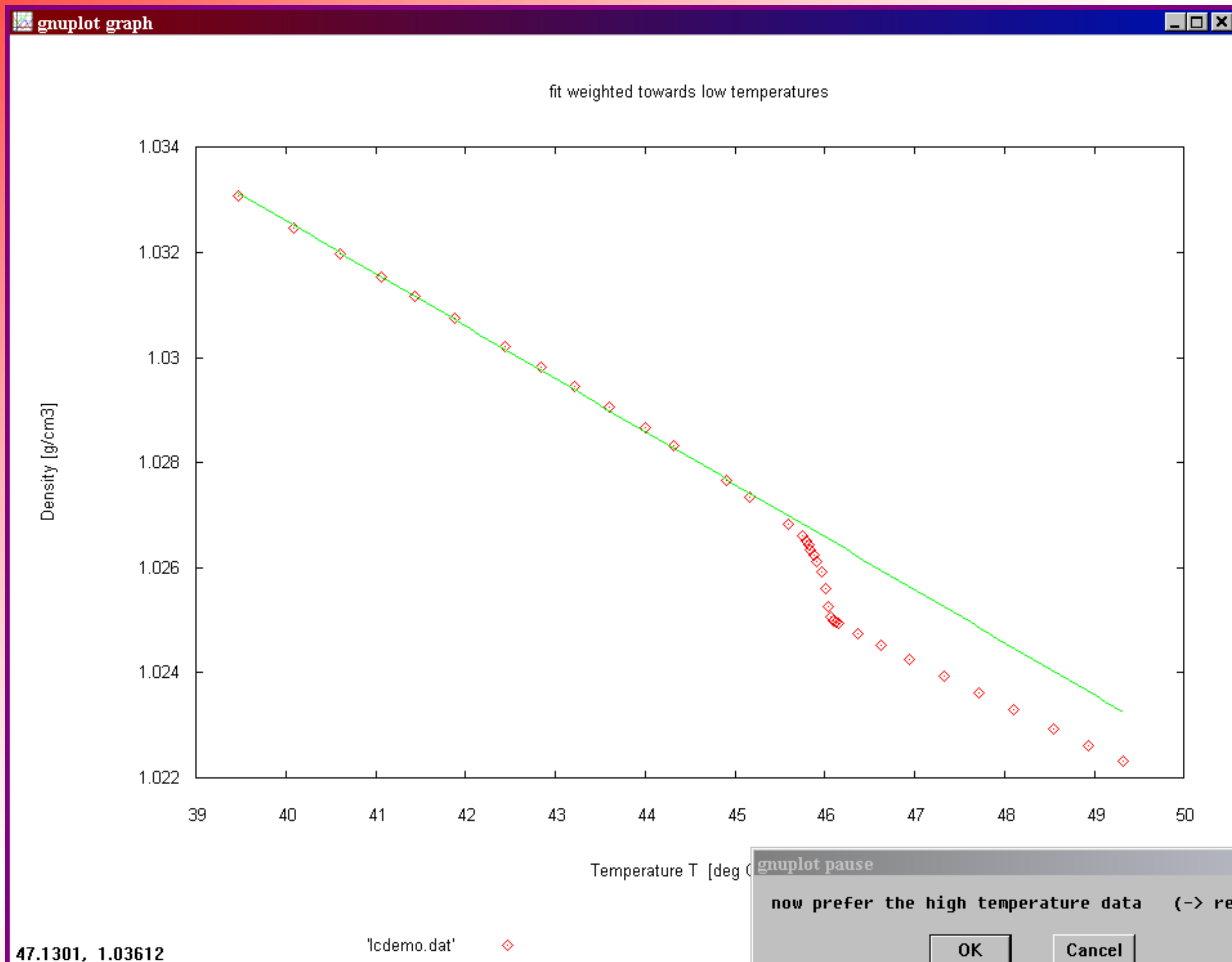
Unweighted Linear Fit

•Fit l(x) 'lcdemo.dat via y0, m

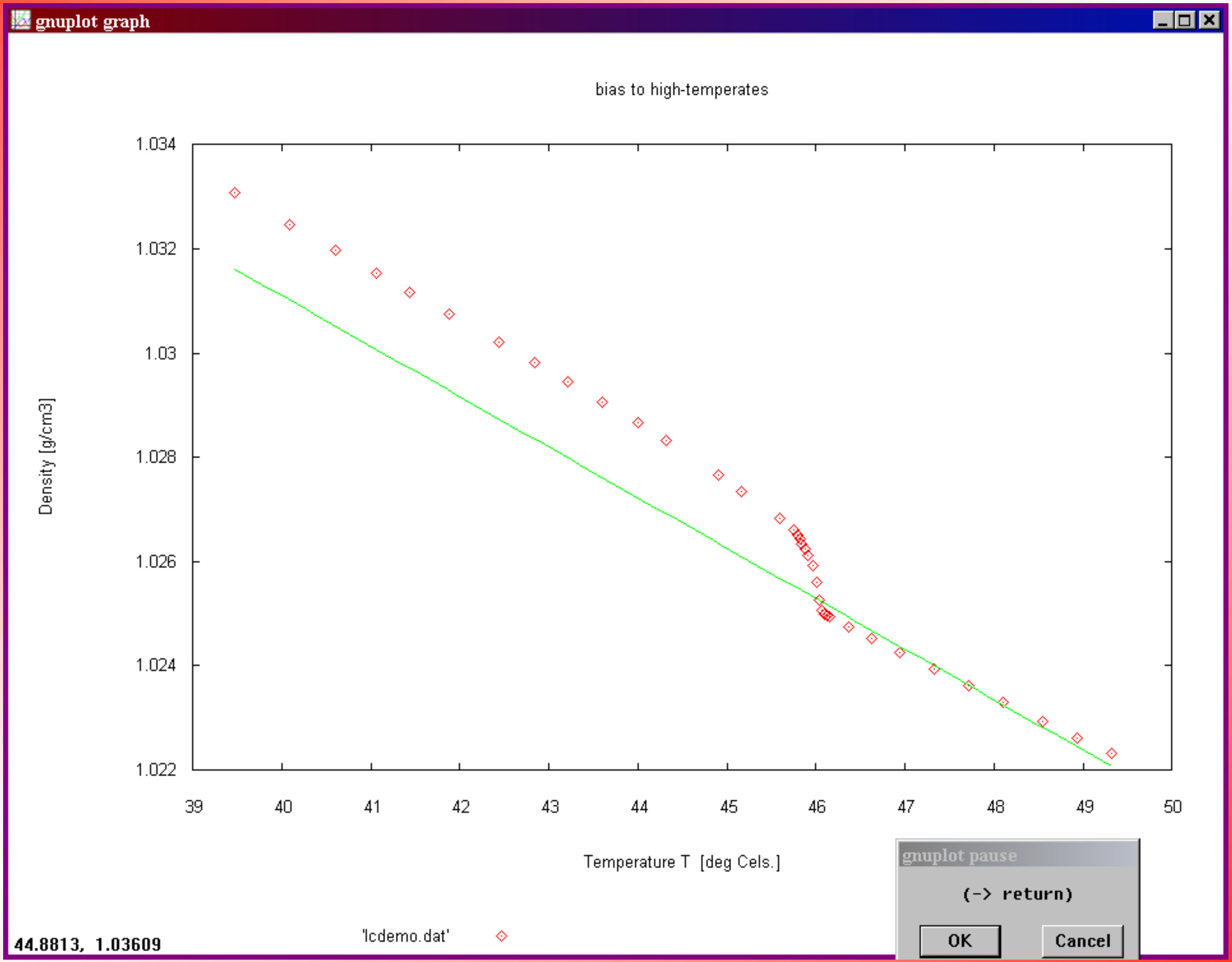
•Plot 'lcdemo.dat', l(x)



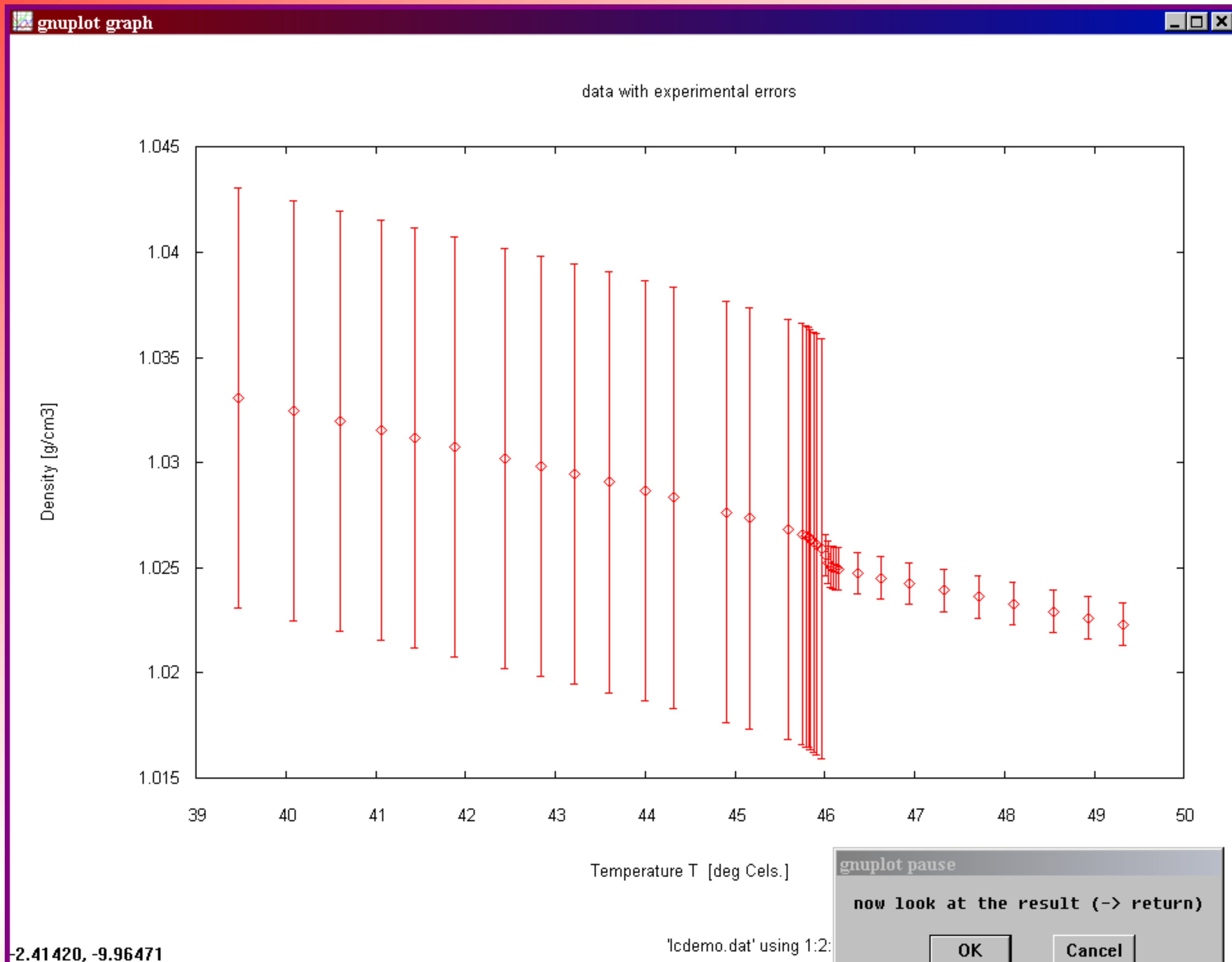
•fit l(x) 'lcdemo.dat' using 1:2:3 via y0, m

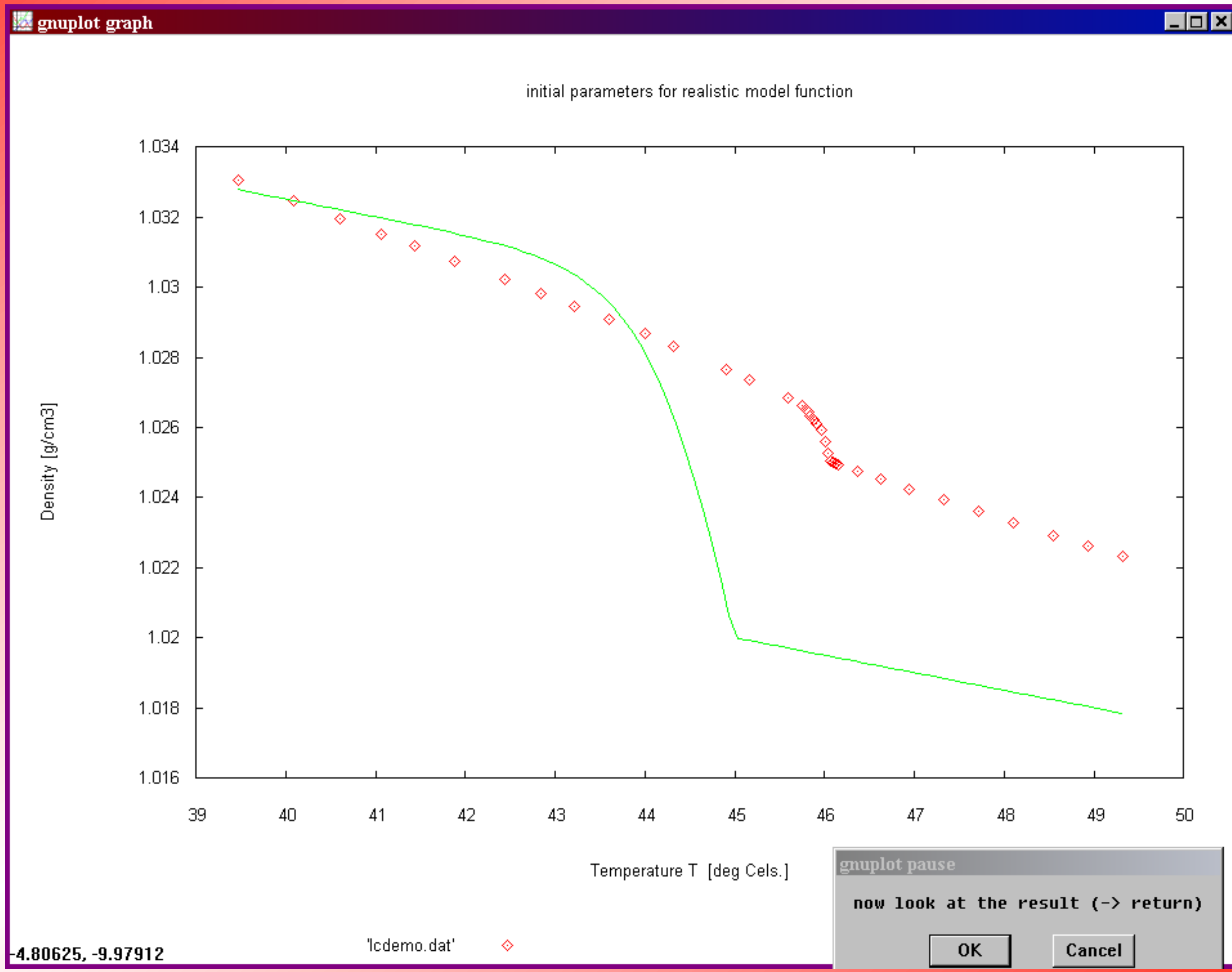


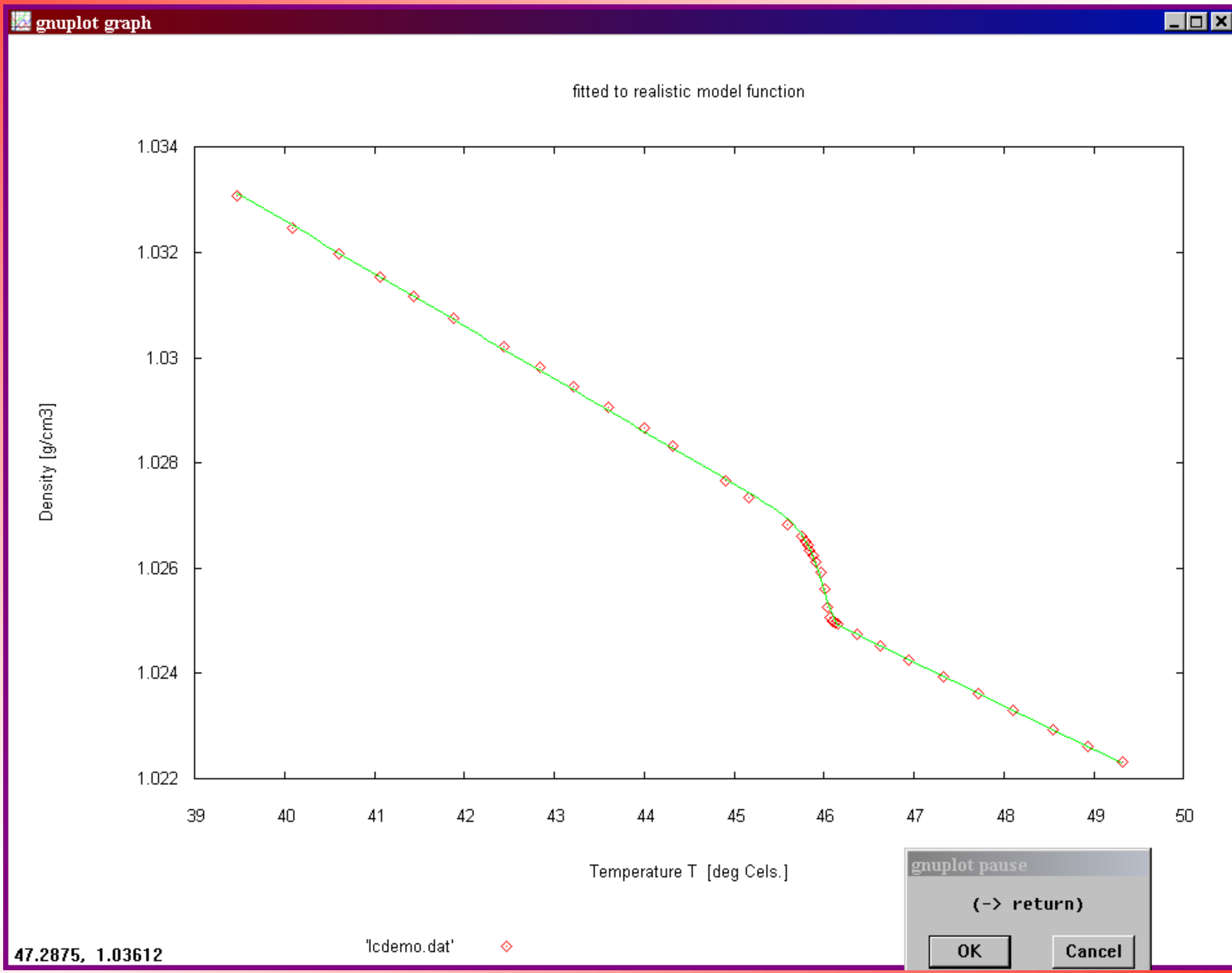
• fit l(x) 'lcdemo.dat' using 1:2:4 via y0, m



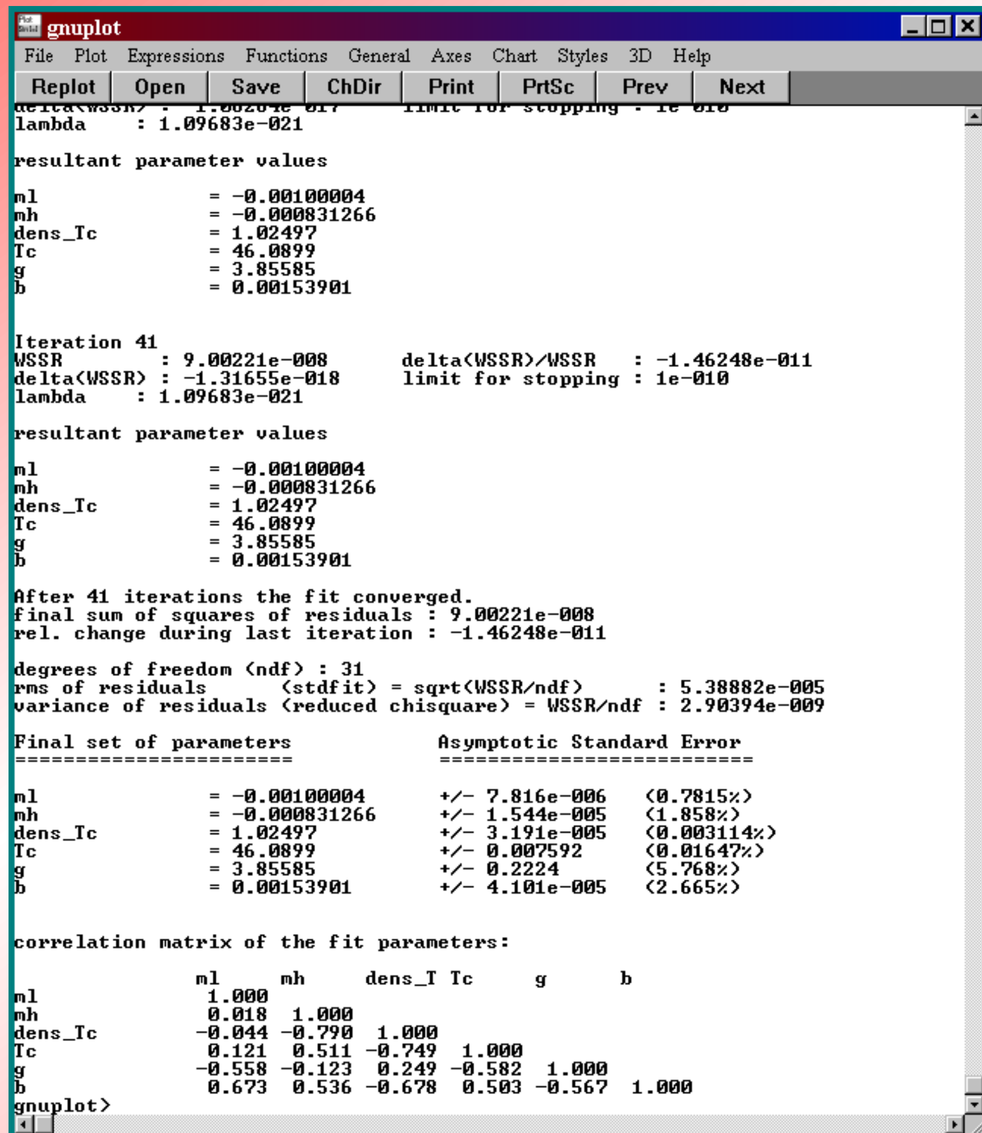
•plot 'lcdemo.dat' using 1:2:5 with errorbars




gnuplot graph
initial parameters for realistic model function
Density [g/cm3]
1.034
1.032
1.03
1.028
1.026
1.024
1.022
1.02
1.018
1.016
39
40
41
42
43
44
45
46
47
48
49
50
Temperature T [deg Cels.]
'lcdemo.dat'
-4.80625, -9.97912
gnuplot pause
now look at the result (-> return)
OK
Cancel


gnuplot graph
fitted to realistic model function
1.034
1.032
1.03
1.028
1.026
1.024
1.022
Density [g/cm3]
39
40
41
42
43
44
45
46
47
48
49
50
Temperature T [deg Cels.]
'lcdemo.dat'
47.2875, 1.03612
gnuplot pause
(-> return)
OK
Cancel

Final Parameters and Error



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,

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

- In computer syntax, this could be written as
N(x) = (A/(sqrt(2*3.1416)*w)) * exp(-(x-xave)**2/(2*(w**2)))
- Show a plot of the data with error bars and the fit, and report the fitted values of A, σ, and x_{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, σ, and x_{ave}, and an estimate of the goodness of the fit, i.e. reduced χ^2 or R^2.
- Tell me what program you used.