Computational Methods (PHYS 2030)

Instructors: Prof. Christopher Bergevin (cberge@yorku.ca)

Schedule: Lecture: MWF 11:30-12:30 (CLH M)

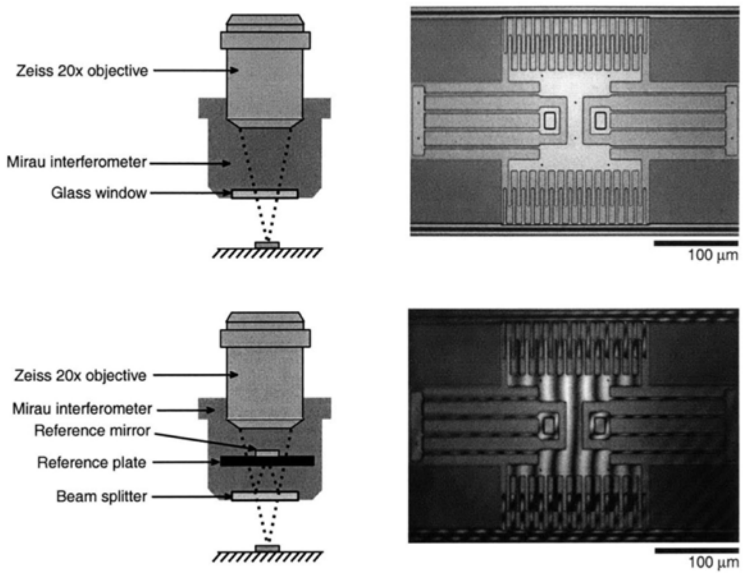
Website: http://www.yorku.ca/cberge/2030W2018.html

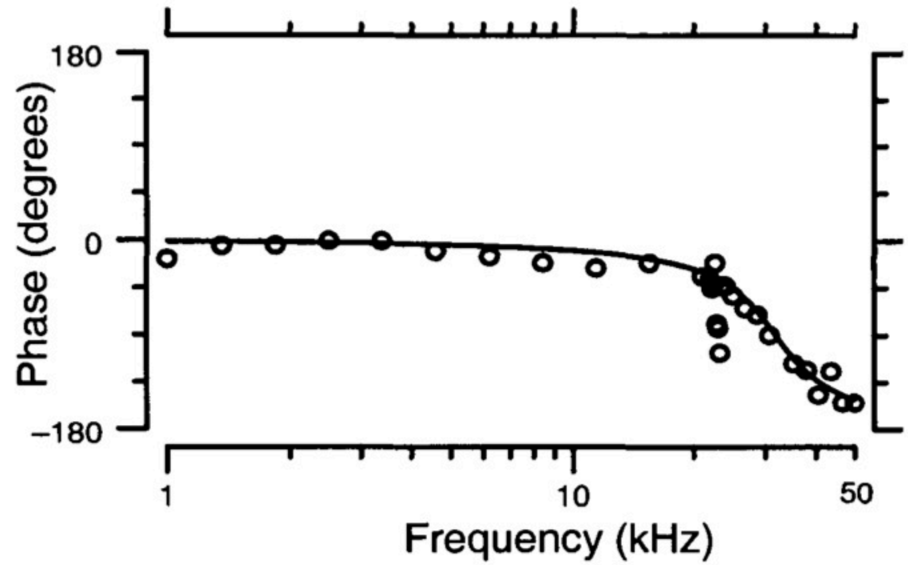
York University
Winter 2018
Lecture 13

see previous notes

Motivation: Fitting a curve to data

Micro-mechanical resonator

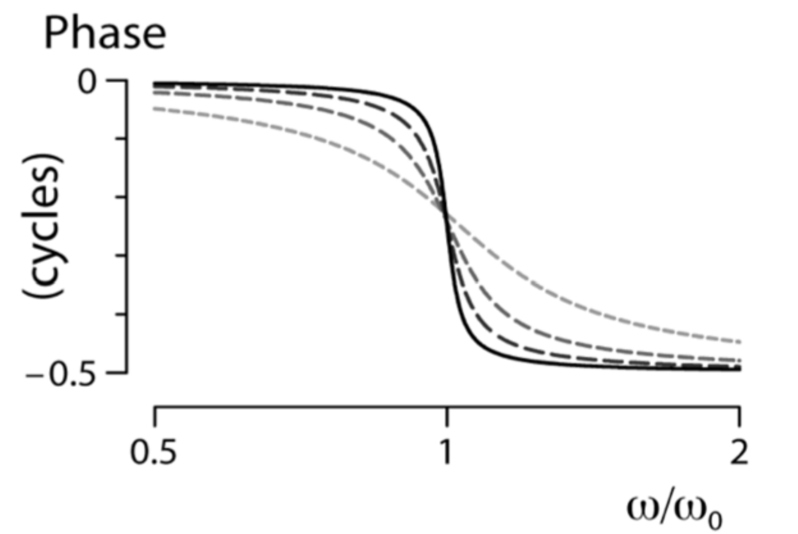




⇒ Characterizing phase slope near resonance provides measure of damping

$$\ddot{x} + \gamma \dot{x} + \omega_o^2 x = \frac{F_o}{m} e^{i\omega t}$$

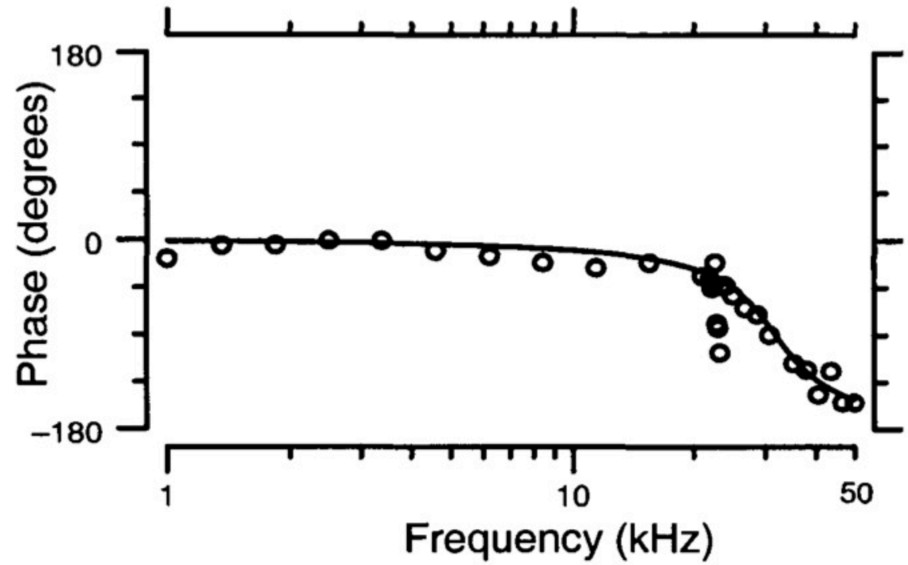
$$\delta(\omega) = \arctan\left(\frac{\gamma\omega}{\omega^2 - \omega_o^2}\right)$$



http://dspace.mit.edu/handle/1721.1/87312#files-area

Regression: Fitting a curve to data

- Present focus is how to fit a function (or some curve) to the data



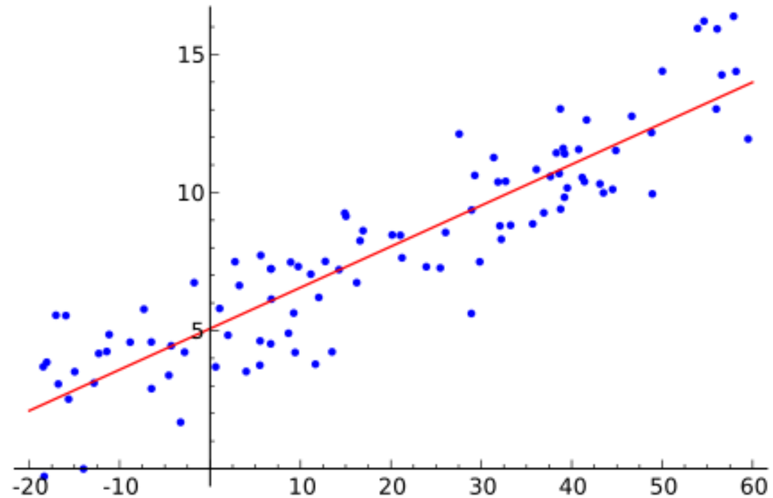
- Two basic ingredients:
  - data points
  - 'model' (i.e., the function to fit)

- Here we had some function as determined from theory

$$\delta(\omega) = \arctan\left(\frac{\gamma\omega}{\omega^2 - \omega_o^2}\right)$$

Basic idea: If we knew the optimal values of γ and ω_o, then we would know the ***best fitting*** function $\delta(\omega)$ (which may be useful in numerous ways)

Regression



- Etymological roots stem from Francis Galton and the biological notion to regress down towards an average value ('regression towards the mean')

(Very) common application:
Fitting a straight line to data (linear regression)

- Very general/powerful concept, manifests in many scientific and engineering applications

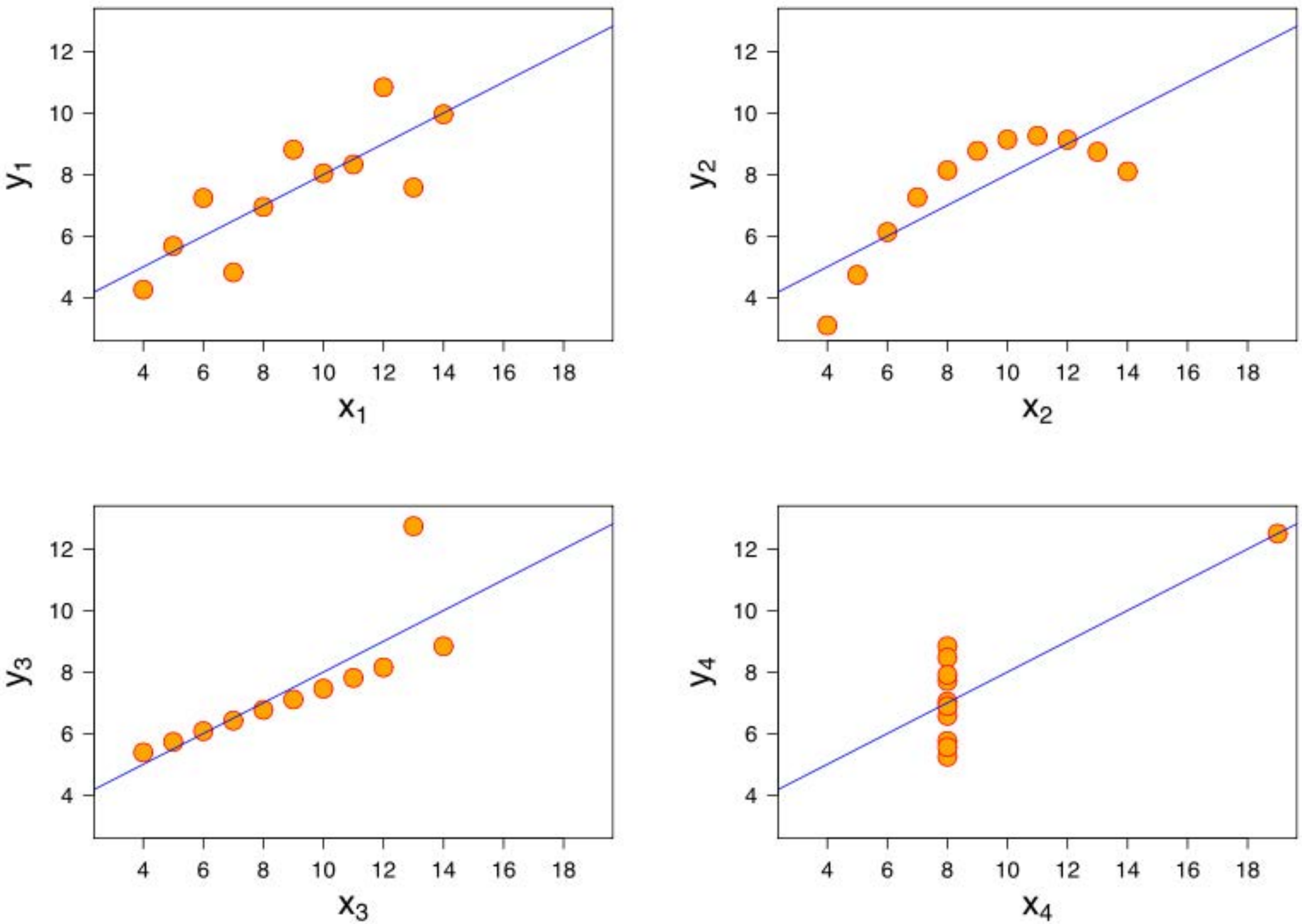
- We will initially focus on a *parametric* method known at *least-squares analysis*

Important point #1: Regression analysis typically involves 'modeling' in that one commonly has assumed a model they are trying to fit to the data

Important point #2: Essentially an optimization problem

wikipedia (regression analysis)

Aside: Anscombe's quartet



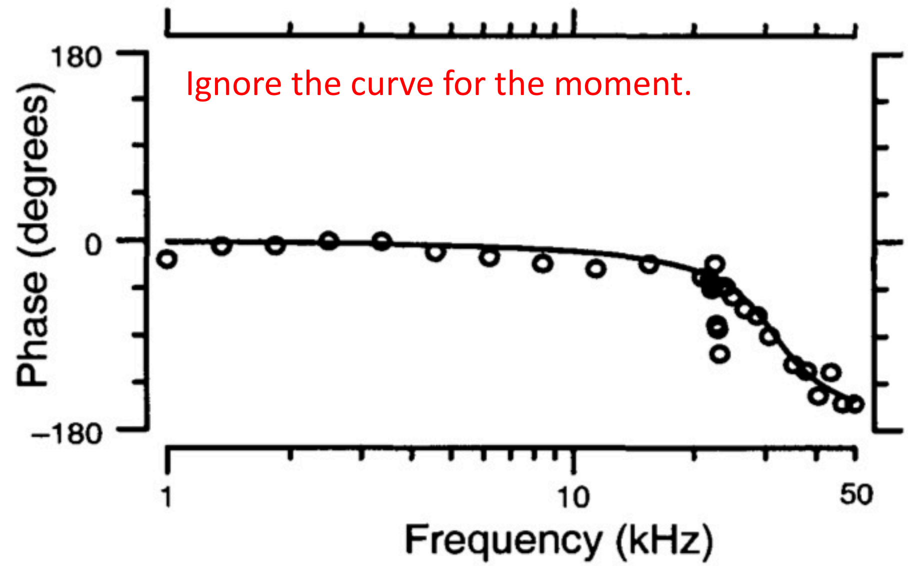
Important point #3: Be smart about how you handle data and make analysis decisions!

wikipedia (Anscomb's quartet)

<u>Aside</u>: Extracting data from figures

- How might we extract the ‘data’ from the graph itself?

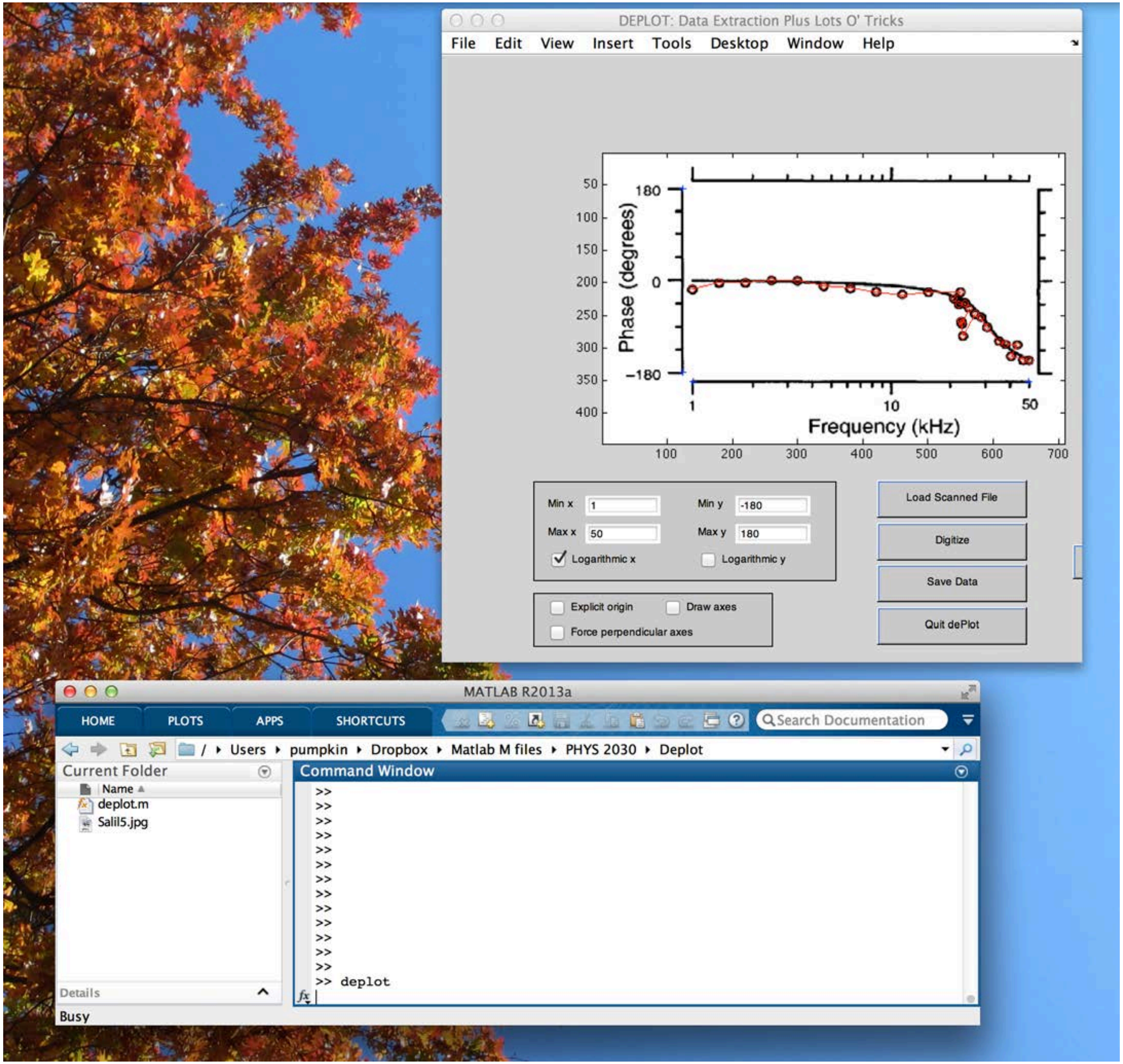
⇒ Such would allow us to do some analysis (e.g., determine the curve of best fit)



- Many strategies possible, we will focus on use of one function called deplot.m
  (hacked together by Christopher Shera; http://web.mit.edu/apg/)

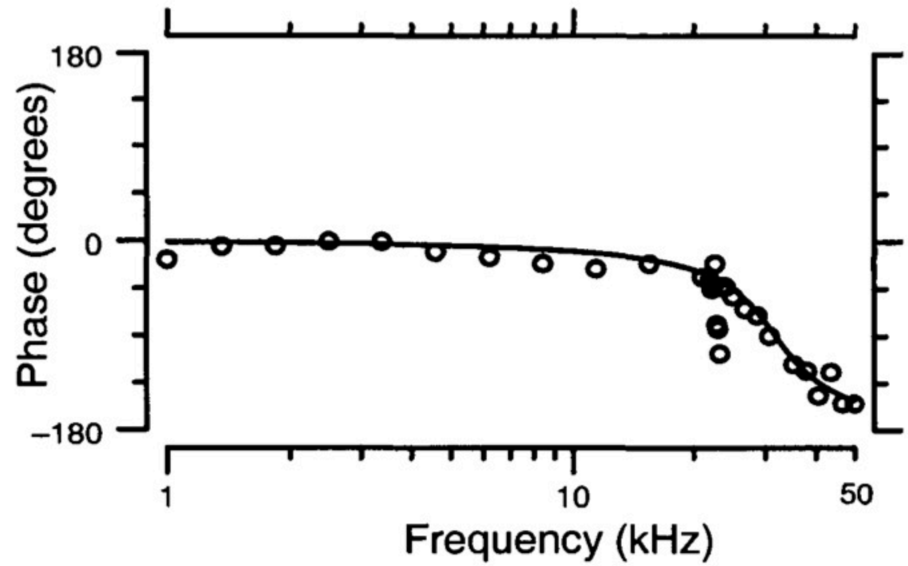
<u>Basic idea</u>: Graphical-user interface (GUI) that allows user to semi-manually extract points from the graph and store away)

Aside: Extracting data from figures

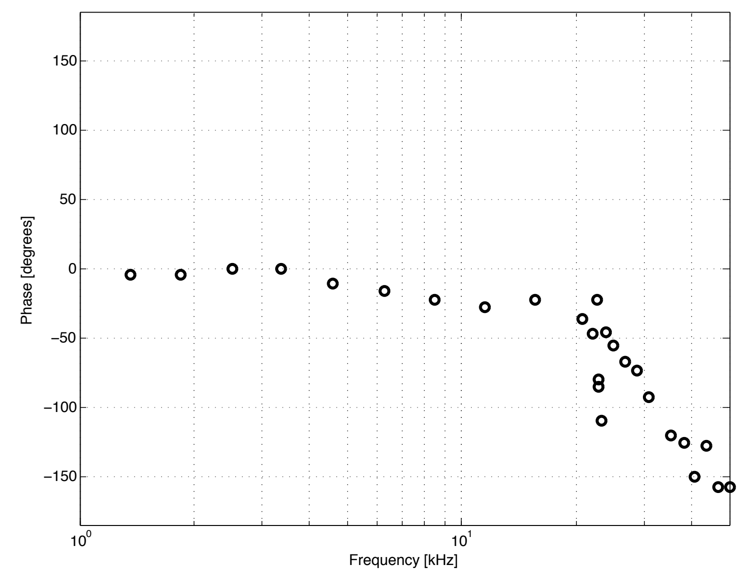


Aside: Extracting data from figures

- Use deplot.m to extract data and save to file



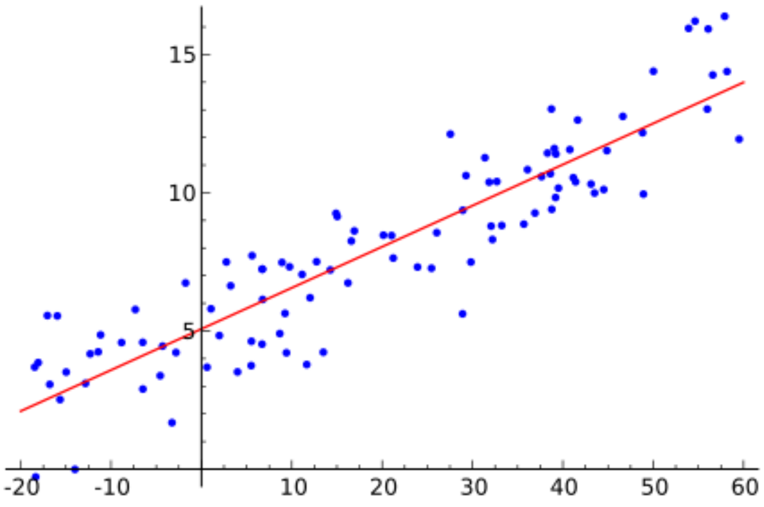

```
clear
A= load('extractedData.txt');
A= A(:,1:2);     % ignore last column
semilogx(A(:,1),A(:,2),'ko','LineWidth',2);
axis([1 50 -185 185]);
xlabel('Frequency [kHz]'); ylabel('Phase
[degrees]');
hold on; grid on;
```



⇒ Might not be exact, but now we have a means to deal with the numbers directly (e.g., nonlinear regression fit of arctan)

Linear regression

- Useful starting point:
  - intuitive
  - has an exact solution
  - easy to implement numerically



→ Natural foundation for more advanced topics (e.g., nonlinear regression, non-parametric regression, bootstrapping)

- In this case, we have some (2-D) 'data' and our 'model' is simply a linear function

$$y = a + bx$$

- the data form x_i (independent var.) and y_i (dependent var.)
- the model is described by $y(x)$
- goal is to determine the best values of a and b

- A key quantity in 'least squares' analysis is χ^2 ("chi-squared")

We'll return to this shortly

Linear regression: Basic ideas

Example 6.1. A student is studying electrical currents and potential differences. He has been provided with a 1-m nickel-silver wire mounted on a board, a lead-acid battery, and an analog voltmeter. He connects cells of the battery across the wire and measures the potential difference or voltage between the negative end and various positions along the wire. From examination of the meter, he estimates the uncertainty in each potential measurement to be 0.05 V. The uncertainty in the position of the probe is less than 1 mm and is considered to be negligible.

→ What assumptions are made here?

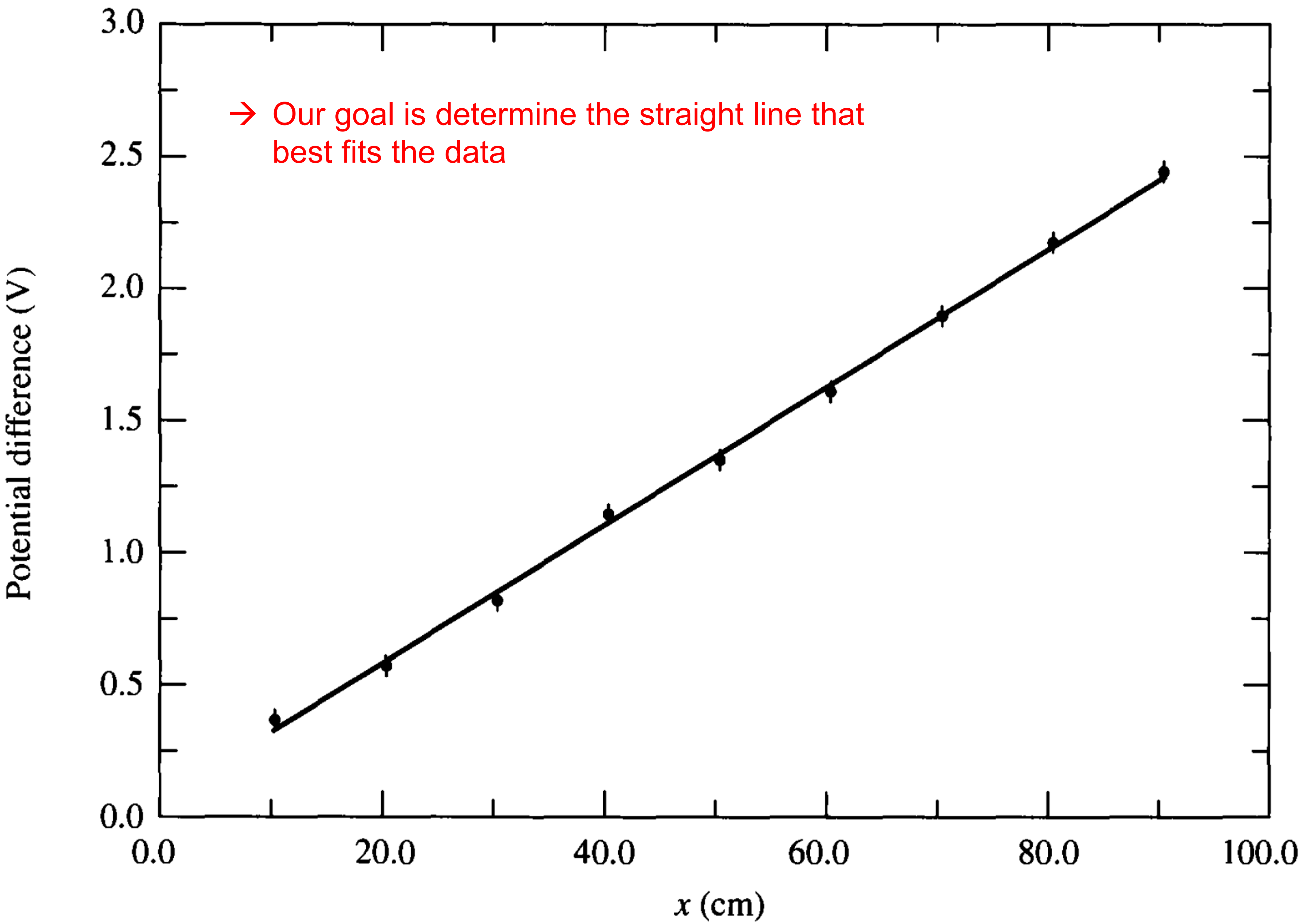
Bevington (2003)

Potential difference V as a function of position along a current-carrying nickel-silver wire

Point number	Postition x_i (cm)	Potential difference V_i (V)	x_i^2	$x_i V_i$	Fitted potential difference $a + bx$
1	10.0	0.37	100	3.70	0.33
2	20.0	0.58	400	11.60	0.60
3	30.0	0.83	900	24.90	0.86
4	40.0	1.15	1,600	46.00	1.12
5	50.0	1.36	2,500	68.00	1.38
6	60.0	1.62	3,600	97.20	1.64
7	70.0	1.90	4,900	133.00	1.91
8	80.0	2.18	6,400	174.40	2.17
9	90.0	2.45	8,100	220.50	2.43
Sums	450.0	12.44	28,500	779.30	

Bevington (2003)

Linear regression: Basic ideas



Bevington (2003)

Basic statistical considerations (we'll need these later)

Note: There is a deeper mathematical/statistical theory here (e.g., probability distributions, Method of Maximum Likelihood) we are only scratching the surface of

- Assume we have some set of data points X:

$$x_i \in X$$

$$i = 1, 2, \cdots, N$$

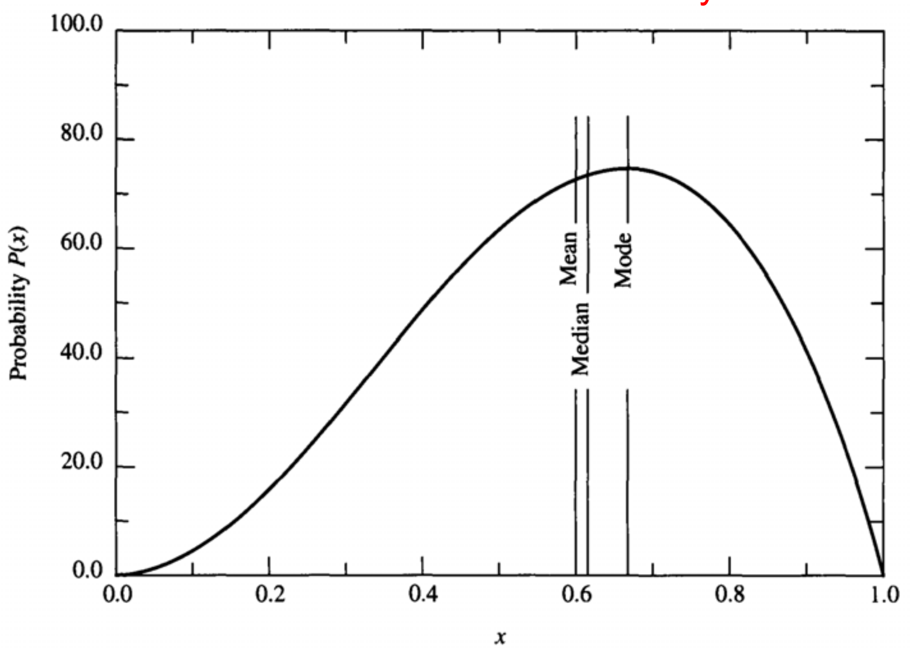
- Sigma notation: $$\sum x_i \equiv \sum_{i=1}^{N} x_i$$

- Then the **mean** of X is: $$\bar{x} \equiv \frac{1}{N} \sum x_i$$

- In the limit of large numbers, the notion of a parent distribution emerges:

$$\mu \equiv \lim_{N \to \infty} \left(\frac{1}{N} \sum x_i \right)$$

Median and Mode also commonly used



Bevington (2003)

Basic statistical considerations (we'll need these later)

- Note that in the real world, we can only have a finite number of points, so the notion of a parent distribution is an ideal one (we deal with the 'sample distribution')

$$\mu \equiv \lim_{N\to\infty}\left(\frac{1}{N}\sum x_i\right)$$

$$(\text{parent parameter}) = \lim_{N\to\infty}(\text{experimental parameter})$$

Basic rules of thumb – Try to make sure:
- Samples are random
- N is large

- Standard deviation – Tells us how much a given point 'deviates' from the average

$$\sigma^2 \equiv \lim_{N\to\infty}\left[\frac{1}{N}\sum(x_i-\mu)^2\right] = \lim_{N\to\infty}\left(\frac{1}{N}\sum x_i^2\right)-\mu^2$$

When N is finite:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2}$$

Basic statistical considerations (we'll need these later)

- Assume we have some set of data points X:

$$x_i \in X$$
$$i = 1, 2, \cdots, N$$

- Each data point (i.e., x_i) can have its own standard deviation (σ_i)

Note: σ_i is something that you typically measure (i.e., directly tied to the empirical nature of x_i)

Example 6.2. In another experiment, a student is provided with a radioactive source enclosed in a small 8-mm-diameter plastic disk and a Geiger counter with a 1-cm-diameter end window. Her object is to investigate the $1/r^2$ law by recording Geiger counter measurements over a fixed period of time at various distances from the source between 20 and 100 cm. Because the counting rate is not expected to vary from measurement to measurement, except for statistical fluctuations, the student can record data long enough to obtain good statistics over the entire range of the experiment. She uses an automatic recording system and records counts for thirty 15-s intervals at each position. For analysis in this experiment, she sums the counts from each set of 30 measurements to obtain the number of counts in 7.5 m intervals. The separate 15-s interval measurements at each position can be used in other statistical studies.

Bevington (2003)

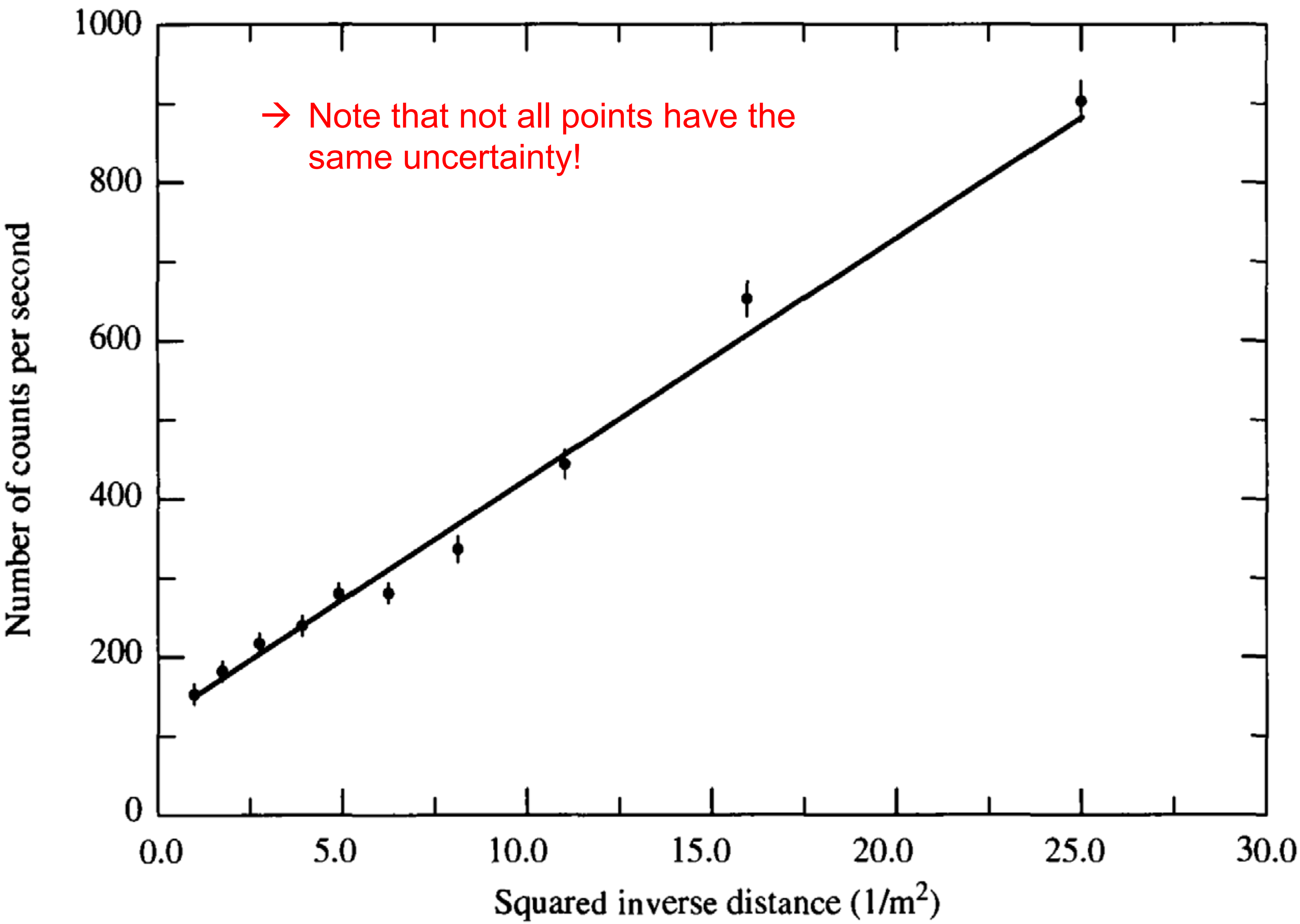
Linear regression: Basic ideas

Number of counts detected in 7½-min intervals as a function of distance from the source

i	Distance d_i (m)	$x_i = 1/d_i^2$ (m^{-2})	Counts C_i	σ_{C_i}	Weight ($1/C_i^2$) w_i	$w_i x_i$	$w_i C_i$	$w_i x_i^2$	$w_i x_i C_i$	Fitted counts $a + bx_i$
1	0.20	25.00	901	30.0	0.00111	0.0278	1	0.694	25.0	887
2	0.25	16.00	652	25.5	0.00153	0.0254	1	0.393	16.0	610
3	0.30	11.11	443	21.0	0.00226	0.0251	1	0.279	11.1	461
4	0.35	8.16	339	18.4	0.00295	0.0241	1	0.197	8.2	370
5	0.40	6.25	283	16.8	0.00353	0.0221	1	0.138	6.3	311
6	0.45	4.94	281	16.8	0.00356	0.0176	1	0.087	4.9	271
7	0.50	4.00	240	15.5	0.00417	0.0167	1	0.067	4.0	242
8	0.60	2.78	220	14.8	0.00455	0.0126	1	0.035	2.8	205
9	0.75	1.78	180	13.4	0.00556	0.0099	1	0.018	1.8	174
10	1.00	1.00	154	12.4	0.00649	0.0065	1	0.007	1.0	150
Sums					0.03570	0.1868	10	1.912	81.0	

Bevington (2003)

Linear regression: Basic ideas



→ Any method we develop to determine the best fit should take into account that certain points might be 'weighted' differently

Bevington (2003)

Least squares

- We measure y_i and want to determine a function $y(x)$ such that we have a predicted value $y(x_i)$

- Deviations between observed value [y_i] and predicted value [$y(x_i)$] is Δy_i. For a linear function to fit, this is then

$$\Delta y_i = y_i - y(x_i) = y_i - a - bx_i$$

→ Goal is to determine the best values of a and b so to minimize Δy_i

- Assume underlying probability distribution is Gaussian:

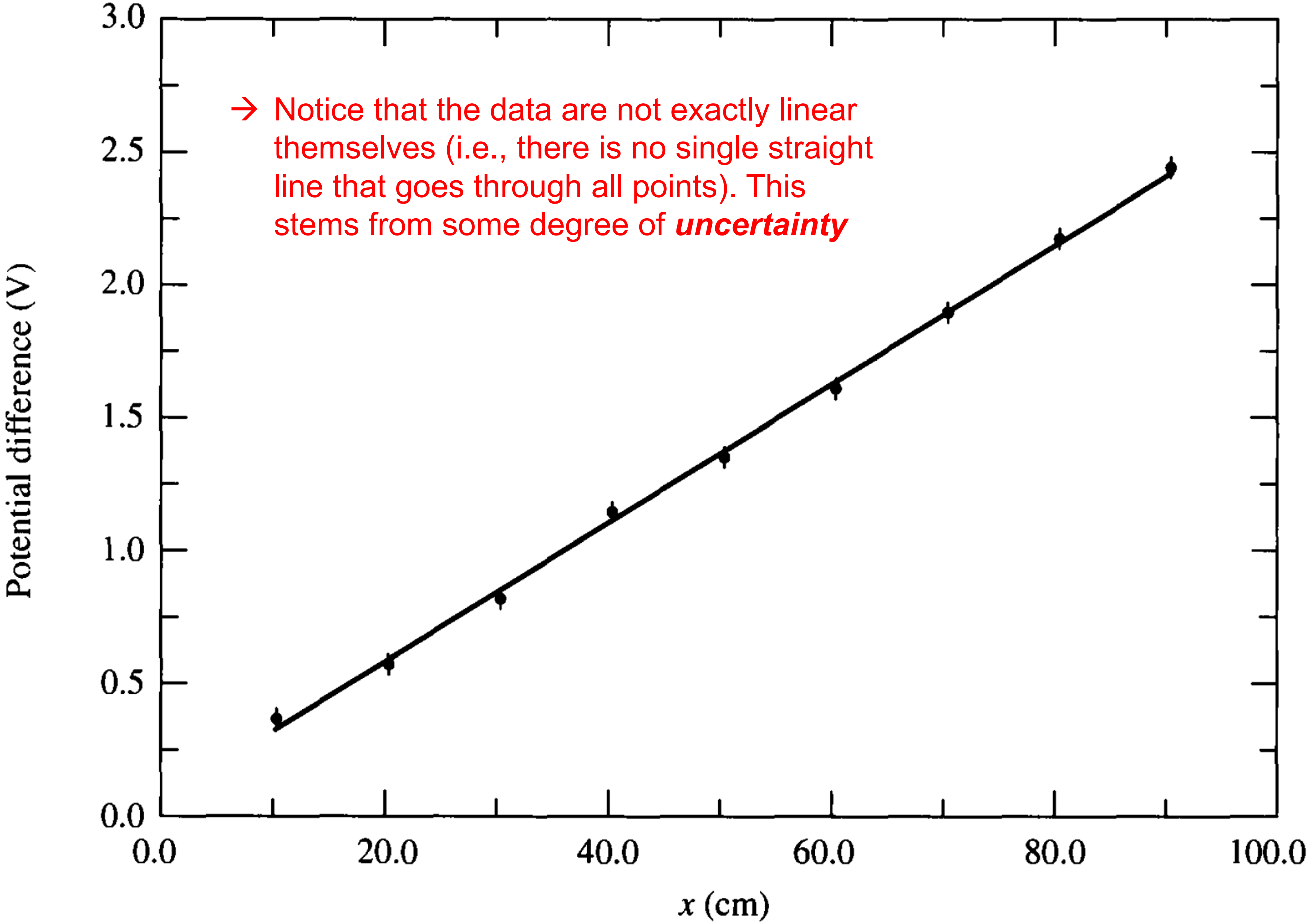
$$P_i = \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left[\frac{y_i - y_0(x_i)}{\sigma_i}\right]^2\right\}$$

Gaussian parent distributions are very common/important in physics!

That is

We shall assume that each individual measured value of y_i is itself drawn from a Gaussian distribution with mean $y_0(x_i)$ and standard deviation σ_i.

Bevington (2003)



→ Notice that the data are not exactly linear themselves (i.e., there is no single straight line that goes through all points). This stems from some degree of ***uncertainty***

Bevington (2003)

Least squares

Various error measurements can be minimized when approximating with a given function $f(x)$. Three standard possibilities are given as follows

I. MaximumError :

$$E_\infty(f) = \max_{1<k<n} |f(x_k) - y_k|. \quad (3.1.4a)$$

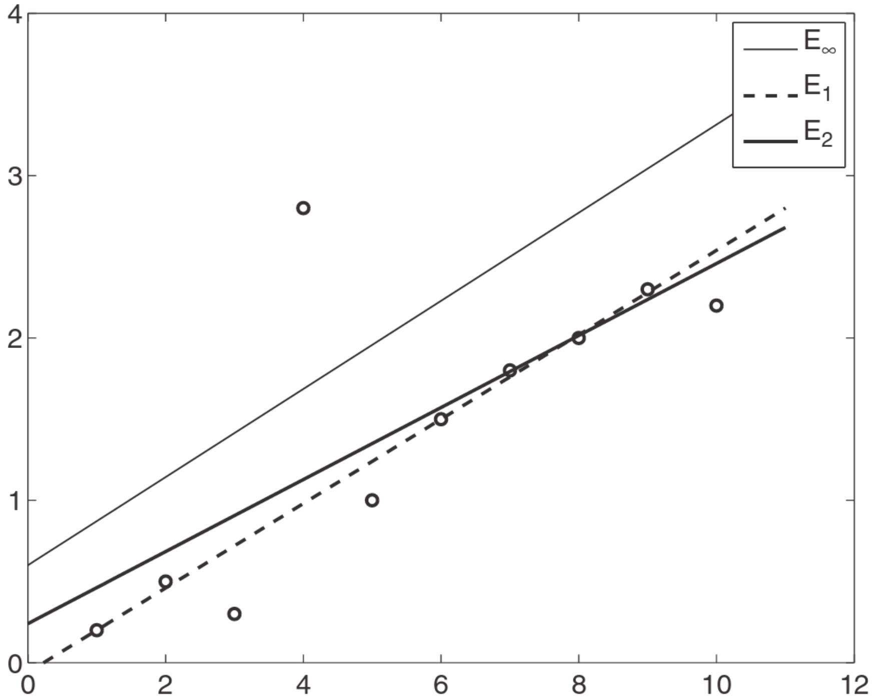
II. AverageError :

$$E_1(f) = \frac{1}{n}\sum_{k=1}^{n} |f(x_k) - y_k|. \quad (3.1.4b)$$

III. Root-meanSquare :

$$E_2(f) = \left(\frac{1}{n}\sum_{k=1}^{n} |f(x_k) - y_k|^2\right)^{1/2}. \quad (3.1.4c)$$

Basic idea: Squaring eliminates bias due cancellations



→ Numerous strategies could be employed, but a root-mean square is the most popular/common

→ Minimizing such leads to the name 'least squares'

Kutz (2013)

Least squares

- We call our 'goodness-of-fit' parameter χ^2 ("chi-squared")

Weighted sum of the squares of the deviations

$$\chi^2 = \sum \left[\frac{y_i - y(x_i)}{\sigma_i} \right]^2 = \sum \left[\frac{1}{\sigma_i} (y_i - a - bx_i) \right]^2$$

for a linear fit

- To determine the smallest value of χ^2, the following factors should be kept in mind:

1. Fluctuations in the measured values of the variables y_i, which are random samples from a parent population with expectation values $y_0(x_i)$.

2. The values assigned to the uncertainties σ_i in the measured variables y_i. Incorrect assignment of the uncertainties σ_i will lead to incorrect values of χ^2.

3. The selection of the analytical function $y(x)$ as an approximation to the "true" function $y_0(x)$. It might be necessary to fit several different functions in order to find the appropriate function for a particular set of data.

4. The values of the parameters of the function $y(x)$. Our objective is to find the "best values" of these parameters.

Bevington (2003)

Linear least squares

- To minimize χ^2, we differentiate and find the associated zeros (such will always be a minimum here)

  Note: Mathematically, this is equivalent to finding the equilibria for a set of PDEs

$$\frac{\partial}{\partial a}\chi^2 = \frac{\partial}{\partial a}\sum\left[\frac{1}{\sigma_i^2}(y_i - a - bx)^2\right]$$
$$= -2\sum\left[\frac{1}{\sigma_i^2}(y_i - a - bx_i)\right] = 0$$
$$\frac{\partial}{\partial b}\chi^2 = \frac{\partial}{\partial b}\sum\left[\frac{x_i}{\sigma_i^2}(y_i - a - bx_i)^2\right]$$
$$= -2\sum\left[\frac{x_i}{\sigma_i^2}(y_i - a - bx_i)\right] = 0$$

Rearrange as:

$$\sum\frac{y_i}{\sigma_i^2} = a\sum\frac{1}{\sigma_i^2} + b\sum\frac{x_i}{\sigma_i^2}$$
$$\sum\frac{x_i y_i}{\sigma_i^2} = a\sum\frac{x_i}{\sigma_i^2} + b\sum\frac{x_i^2}{\sigma_i^2}$$

→ This is just a linear system of equations (i.e., two equations, two unknowns)! Here we are solving for a and b

Bevington (2003)

<u>Least squares</u>

- Determinant solution:

$$a = \frac{1}{\Delta}\begin{vmatrix} \sum \frac{y_i}{\sigma_i^2} & \sum \frac{x_i}{\sigma_i^2} \\ \sum \frac{x_i y_i}{\sigma_i^2} & \sum \frac{x_i^2}{\sigma_i^2} \end{vmatrix} = \frac{1}{\Delta}\left(\sum \frac{x_i^2}{\sigma_i^2} \sum \frac{y_i}{\sigma_i^2} - \sum \frac{x_i}{\sigma_i^2} \sum \frac{x_i y_i}{\sigma_i^2}\right)$$

$$b = \frac{1}{\Delta}\begin{vmatrix} \sum \frac{1}{\sigma_i^2} & \sum \frac{y_i}{\sigma_i^2} \\ \sum \frac{x_i}{\sigma_i^2} & \sum \frac{x_i y_i}{\sigma_i^2} \end{vmatrix} = \frac{1}{\Delta}\left(\sum \frac{1}{\sigma_i^2} \sum \frac{x_i y_i}{\sigma_i^2} - \sum \frac{x_i}{\sigma_i^2} \sum \frac{y_i}{\sigma_i^2}\right)$$

$$\Delta = \begin{vmatrix} \sum \frac{1}{\sigma_i^2} & \sum \frac{x_i}{\sigma_i^2} \\ \sum \frac{x_i}{\sigma_i^2} & \sum \frac{x_i^2}{\sigma_i^2} \end{vmatrix} = \sum \frac{1}{\sigma_i^2} \sum \frac{x_i^2}{\sigma_i^2} - \left(\sum \frac{x_i}{\sigma_i^2}\right)^2$$

→ Direct recipe to solve for a and b

Bevington (2003)

Least squares

- These expression simplify further when all uncertainties are equal ($\sigma=\sigma_i$):

$$a = \frac{1}{\Delta'}\begin{vmatrix}\Sigma y_i & \Sigma x_i \\ \Sigma x_i y_i & \Sigma x_i^2\end{vmatrix} = \frac{1}{\Delta'}(\Sigma x_i^2 \Sigma y_i - \Sigma x_i \Sigma x_i y_i)$$

$$b = \frac{1}{\Delta'}\begin{vmatrix}N & \Sigma y_i \\ \Sigma x_i & \Sigma x_i y_i\end{vmatrix} = \frac{1}{\Delta'}(N \Sigma x_i y_i - \Sigma x_i \Sigma y_i)$$

$$\Delta' = \begin{vmatrix}N & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2\end{vmatrix} = N \Sigma x_i^2 - (\Sigma x_i)^2$$

- Kutz's book pitches it slightly differently:

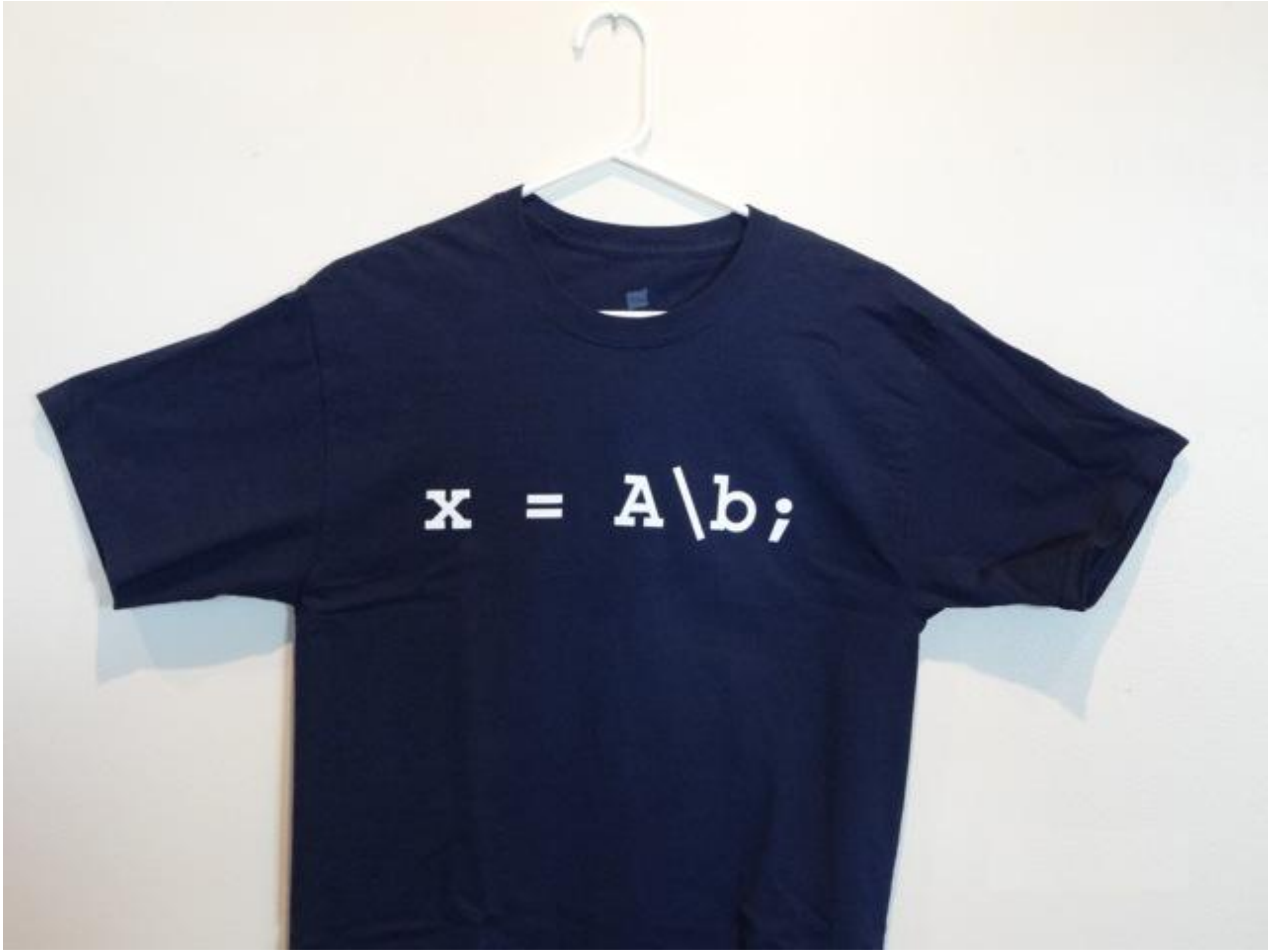
Upon rearranging, the 2×2 system of linear equations is found for A and B:

$$\begin{pmatrix}\sum_{k=1}^n x_k^2 & \sum_{k=1}^n x_k \\ \sum_{k=1}^n x_k & n\end{pmatrix}\begin{pmatrix}A \\ B\end{pmatrix} = \begin{pmatrix}\sum_{k=1}^n x_k y_k \\ \sum_{k=1}^n y_k\end{pmatrix}.$$

This equation can be easily solved using the backslash command in MATLAB.

Bevington (2003)
Kutz (2013)

Do you speak Matlab?



http://blogs.mathworks.com/cleve/2013/08/19/backslash/

Computational methods to minimize χ^2

- Matlab has numerous built-in functions. For a linear fit (and other polynomials), one can use `polyfit.m` (see example code EXregression1.m for syntax)

- For the linear case, the preceding formulae for the exact solution provide an explicit recipe (see example code EXregression1.m for syntax)

- Brute force estimate χ^2 for a range of parameter values and see which ones provide the smallest value. Refine your search and repeat. (this is called the grid-search method)

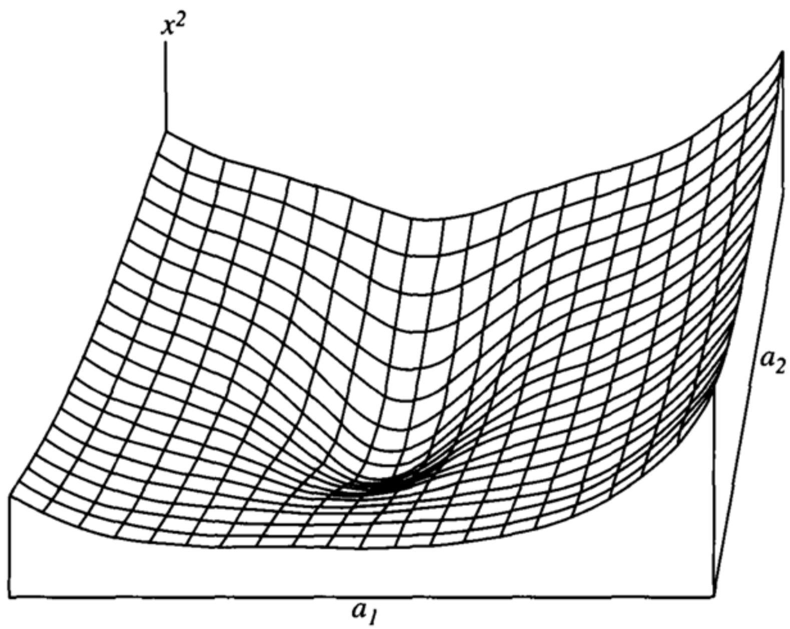


FIGURE 8.2
Chi-square hypersurface as a function of two parameters.

Bevington (2003)