计算物理

Lecture 9

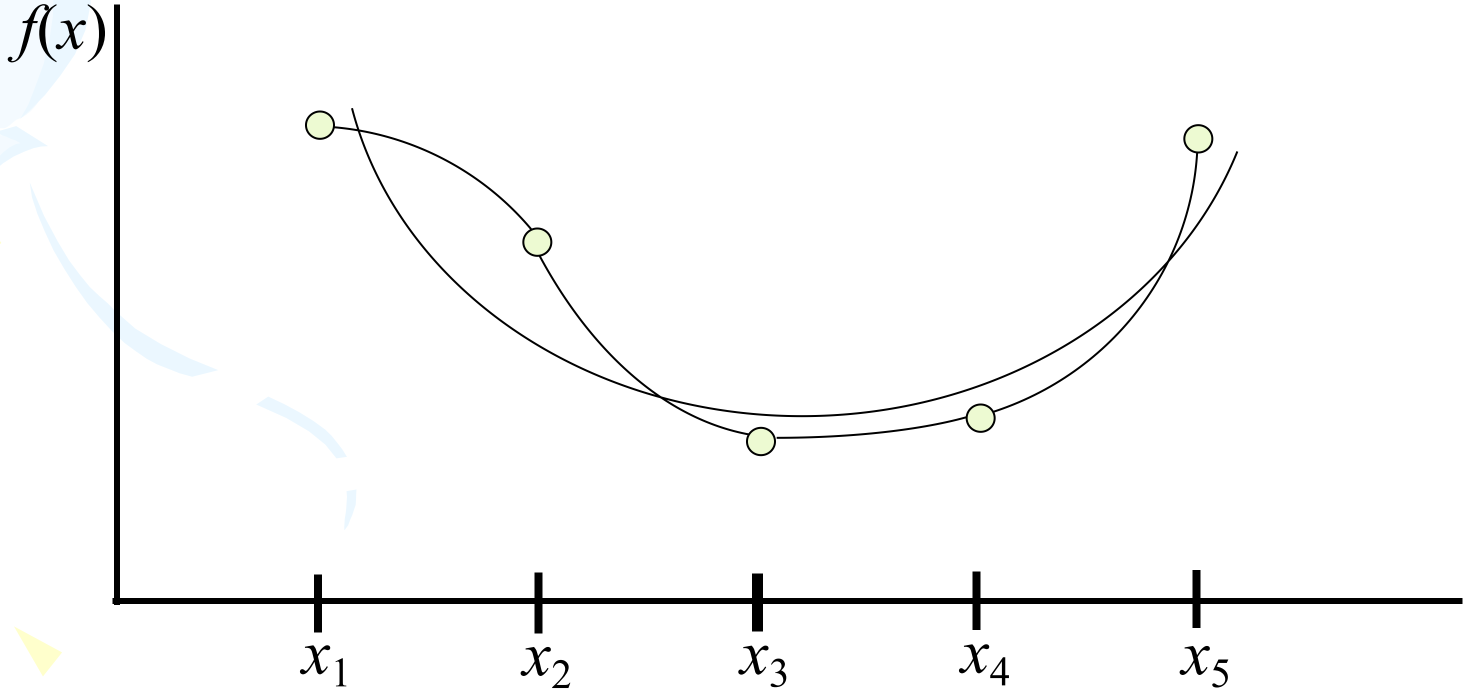
傅子文
fuziwen@scu.edu.cn

Today's Lecture

- The Maximum Likelihood Method
- Least squares polynomial fitting
  - Fitting to data
  - Fitting to functions – the Hilbert Matrix
  - Use of orthogonal polynomials to simplify the fitting process

Difference between interpolation&fitting

- Interpolation passes through each datum
- Fitting seeks to produce a curve that approximates the data



The Maximum Likelihood Method

- Suppose we try to measure the true value of some quantity (x_T).
  - We make repeated measurements of this quantity $\{x_1, x_2, \ldots x_n\}$.
  - The standard way to estimate x_T from our measurements is to calculate the mean value:

$$\mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i$$

and set $x_T = \mu_x$.

DOES THIS PROCEDURE MAKE SENSE???

The maximum likelihood method (MLM) answers this question and provides a general method for estimating parameters of interest from data.

The Maximum Likelihood Method

- Statement of the Maximum Likelihood Method
  - Assume we have made N measurements of x $\{x_1, x_2, \ldots x_n\}$.
  - Assume we know the probability distribution function that describes x: $f(x, \alpha)$.
  - Assume we want to determine the parameter α.

MLM: pick α to maximize the probability of getting the measurements (the x_i's) that we did

The Maximum Likelihood Method

- How do we use the MLM?
  - The probability of measuring x_1 is $f(x_1, \alpha)$dx
  - The probability of measuring x_2 is $f(x_2, \alpha)$dx
  - The probability of measuring x_n is $f(x_n, \alpha)$dx
  - If the measurements are independent, the probability of getting the measurements we did is:

$$L = f(x_1,\alpha)dx \cdot f(x_2,\alpha)dx \cdots f(x_n,\alpha)dx = f(x_1,\alpha)\cdot f(x_2,\alpha)\cdots f(x_n,\alpha)[dx^n]$$

We can drop the dx^n term as it is only a proportionality constant.

L is called the Likelihood Function : $L = \prod_{i=1}^{N} f(x_i,\alpha)$

The Maximum Likelihood Method

- We want to pick the α that maximizes L:

$$\left.\frac{\partial L}{\partial \alpha}\right|_{\alpha=\alpha^*} = 0$$

- Often easier to maximize ln$L$. L and ln$L$ are both maximum at the same location.

we maximize ln$L$ rather than L itself because ln$L$ converts the product into a summation.

$$\ln L = \sum_{i=1}^{N} \ln f(x_i, \alpha) \qquad \left.\frac{\partial \ln L}{\partial \alpha}\right|_{\alpha=\alpha^*} = \sum_{i=1}^{N} \frac{\partial}{\partial \alpha} \ln f(x_i, \alpha)\Big|_{\alpha=\alpha^*} = 0$$

The new maximization condition is:

– α could be an array of parameters (e.g. slope and intercept) or just a single variable.

– equations to determine α range from simple linear equations to coupled non-linear equations.

MLM Example:

- Let $f(x, \alpha)$ be given by a Gaussian distribution function.
- Let $\alpha = \mu$ be the mean of the Gaussian. We want to use our data+MLM to find the mean, μ.
- We want the best estimate of α from our set of n measurements $\{x_1, x_2, \ldots x_n\}$.
- Let's assume that σ is the same for each measurement.

$$f(x_i,\alpha) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\alpha)^2}{2\sigma^2}}$$

gaussian PDF

- The likelihood function for this problem is:

$$L = \prod_{i=1}^{n} f(x_i,\alpha) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\alpha)^2}{2\sigma^2}}$$

$$= \left[\frac{1}{\sigma\sqrt{2\pi}}\right]^n e^{-\frac{(x_1-\alpha)^2}{2\sigma^2}} e^{-\frac{(x_2-\alpha)^2}{2\sigma^2}} \cdots e^{-\frac{(x_n-\alpha)^2}{2\sigma^2}} = \left[\frac{1}{\sigma\sqrt{2\pi}}\right]^n e^{-\sum_{i=1}^{n}\frac{(x_i-\alpha)^2}{2\sigma^2}}$$

$$\ln L = \ln\prod_{i=1}^{n} f(x_i,\alpha) = \ln\left([\frac{1}{\sigma\sqrt{2\pi}}]^n e^{-\sum_{i=1}^{n}\frac{(x_i-\alpha)^2}{2\sigma^2}} \right) = n\ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \sum_{i=1}^{n}\frac{(x_i-\alpha)^2}{2\sigma^2}$$

We want to find the α that maximizes the log likelihood function:

$$\frac{\partial \ln L}{\partial \alpha} = \frac{\partial}{\partial \alpha}\left[n \ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \sum_{i=1}^{n} \frac{(x_i-\alpha)^2}{2\sigma^2}\right] = 0$$

Factor out σ since it is a constant

$$\frac{\partial}{\partial \alpha}\sum_{i=1}^{n}(x_i-\alpha)^2 = 0$$

$$\sum_{i=1}^{n} 2(x_i-\alpha)(-1) = 0$$

$$\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} {}_i\alpha = 0$$

Don't forget the factor of *n*

$$\sum_{i=1}^{n} x_i = n\alpha$$

$$\alpha = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Average

If σ are different for each data point then α is just the weighted average:

$$\alpha = \frac{\sum_{i=1}^{n} \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}}$$

Weighted Average

Example: Poisson Distribution

- Let $f(x, \alpha)$ be given by a Poisson distribution. $f(x,\alpha) = \dfrac{e^{-\alpha}\alpha^{x}}{x!}$
- Let $\alpha = \mu$ be the mean of the Poisson.
- We want the best estimate of α from our set of n measurements $\{x_1, x_2, \ldots x_n\}$.
- The likelihood function for this problem is:

$$L = \prod_{i=1}^{n} f(x_i,\alpha) = \prod_{i=1}^{n} \frac{e^{-\alpha}\alpha^{x_i}}{x_i!} = \frac{e^{-\alpha}\alpha^{x_1}}{x_1!}\frac{e^{-\alpha}\alpha^{x_2}}{x_2!}\cdots\frac{e^{-\alpha}\alpha^{x_n}}{x_n!} = \frac{e^{-n\alpha}\alpha^{\sum_{i=1}^{n} x_i}}{x_1!x_2!..x_n!}$$

- Find α that maximizes the log likelihood function:

$$\frac{d\ln L}{d\alpha} = \frac{d}{d\alpha}\left(-n\alpha + \ln\alpha\cdot\sum_{i=1}^{n} x_i - \ln(x_1!x_2!..x_n!)\right) = -n + \frac{1}{\alpha}\sum_{i=1}^{n} x_i = 0$$

$$\alpha = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Average

Some general properties of the MLM

- For large data samples (large *n*) the likelihood function, *L*, approaches a Gaussian distribution.
- Maximum likelihood estimates are usually *consistent.*
  For large *n* the estimates converge to the true value of the parameters we wish to determine.
- Maximum likelihood estimates are usually *unbiased.*
  For all sample sizes the parameter of interest is calculated correctly.
- Maximum likelihood estimate is *efficient*: the estimate has the smallest variance.
- Maximum likelihood estimate is *sufficient*: it uses all the information in the observations (the x_i's).
- The solution from MLM is unique.

- **Bad news: we must know the correct probability distribution for the problem at hand!**

Maximum Likelihood Fit of Data to a Function

- Suppose we have a set of n measurements:

$$x_1, y_1 \pm \sigma_1$$

$$x_2, y_2 \pm \sigma_2$$

...

$$x_n, y_n \pm \sigma_n$$

- Assume each measurement error (σ) is a standard deviation from a Gaussian pdf.
- Assume that for each measured value y, there's an x which is known exactly.
- Suppose we know the functional relationship between the y's and the x's:

$$y = q(x, \alpha, \beta, \ldots)$$

– α, β...are parameters.

MLM gives us a method to determine α, β... from our data.

Example: Fitting data points to a straight line.

We want to determine the slope (β) and intercept (α).

$$q(x,\alpha,\beta,\ldots) = \alpha + \beta x$$

$$L = \prod_{i=1}^{n} f(x_i,\alpha,\beta) = \prod_{i=1}^{n} \frac{1}{\sigma_i\sqrt{2\pi}} e^{-\frac{(y_i - q(x_i,\alpha,\beta))^2}{2\sigma_i^2}} = \prod_{i=1}^{n} \frac{1}{\sigma_i\sqrt{2\pi}} e^{-\frac{(y_i-\alpha-\beta x_i)^2}{2\sigma_i^2}}$$

Find α and β by maximizing the log of the likelihood function L:

$$\frac{\partial \ln L}{\partial \alpha} = \frac{\partial}{\partial \alpha}\sum_{i=1}^{n}\left[\ln\left(\frac{1}{\sigma_i\sqrt{2\pi}}\right) - \frac{(y_i-\alpha-\beta x_i)^2}{2\sigma_i^2}\right]$$

$$= \sum_{i=1}^{n}\left[-\frac{2(y_i-\alpha-\beta x_i)(-1)}{2\sigma_i^2}\right] = 0$$

$$\frac{\partial \ln L}{\partial \beta} = \frac{\partial}{\partial \beta}\sum_{i=1}^{n}\left[\ln\left(\frac{1}{\sigma_i\sqrt{2\pi}}\right) - \frac{(y_i-\alpha-\beta x_i)^2}{2\sigma_i^2}\right]$$

$$= \sum_{i=1}^{n}\left[-\frac{2(y_i-\alpha-\beta x_i)(-x_i)}{2\sigma_i^2}\right] = 0$$

Two linear equations with two unknowns

For the sake of simplicity assume that all σ's are the same ($\sigma_1 = \sigma_2 = \sigma$):

$$\frac{\partial \ln L}{\partial \alpha} = \sum_{i=1}^{n} \left[-\frac{2(y_i - \alpha - \beta x_i)(-1)}{2\sigma_i^2} \right] = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \alpha - \sum_{i=1}^{n} \beta x_i = 0$$

$$\frac{\partial \ln L}{\partial \beta} = \sum_{i=1}^{n} \left[-\frac{2(y_i - \alpha - \beta x_i)(-x_i)}{2\sigma_i^2} \right] = \sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} \alpha x_i - \sum_{i=1}^{n} \beta x_i^2 = 0$$

We now have two equations that are linear in the unknowns α, β:

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \alpha + \sum_{i=1}^{n} \beta x_i = n\alpha + \beta \sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n} y_i x_i = \alpha \sum_{i=1}^{n} x_i - \beta \sum_{i=1}^{n} x_i^2 = 0$$

in matrix form

$$\begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} y_i x_i \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

$$\alpha = \frac{\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} y_i x_i \sum_{i=1}^{n} x_i}{n \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}$$ and $$\beta = \frac{n \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i}{n \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}$$

- EXAMPLE: A trolley moves along a track at constant speed.

Suppose the following measurements of the time vs. distance were made. From the data find the best value for the speed (v) of the trolley.

Time t (seconds)	1.0	2.0	3.0	4.0	5.0	6.0
Distance d (mm)	11	19	33	40	49	61

- Our model of the motion of the trolley tells us that: $d = d_0 + vt$
  - We want to find v, the slope (β) of the straight line describing the motion of the trolley.
  - We need to evaluate the sums listed in the above formula:

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{6} t_i = 21s \qquad \sum_{i=1}^{n} y_i = \sum_{i=1}^{6} d_i = 213mm$$

$$\sum_{i=1}^{n} x_i y_i = \sum_{i=1}^{6} t_i d_i = 919s \cdot mm \qquad \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{6} t_i^2 = 91s^2$$

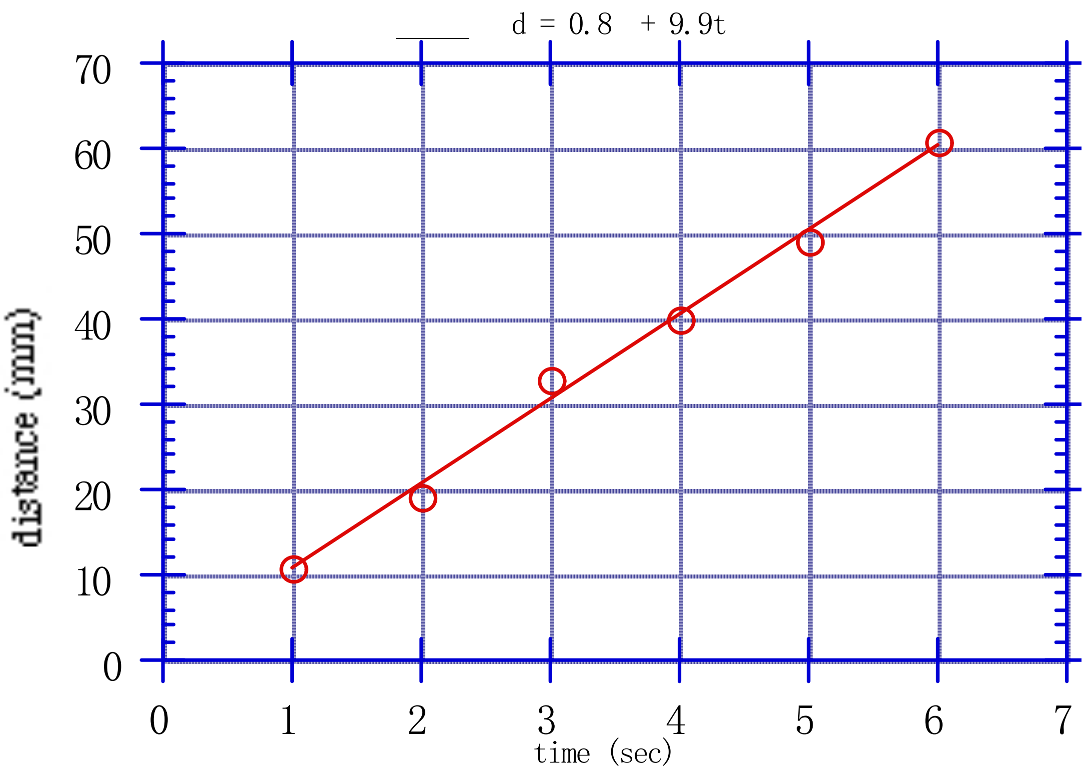
$$v = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} = \frac{6\times 919 - 21\times 213}{6\times 91 - 21^2} = 9.9\, mm/s$$

best estimate of the speed

$$d_0 = 0.8\ mm$$

best estimate of the starting point

The MLM fit to the data for $d=d_0+vt$



The line in the above plot "best" represents our data.

Not all the data points are "on" this line.

The line minimizes the sum of squares of deviations (δ) between a line and our data (d_i):

δ_i=data-prediction= $d_i-(d_0+vt_i)$ $\rightarrow$ minimize: $\Sigma[\delta_i^2]$ $\rightarrow$ same as MLM!

Often we call this technique "Least Squares" (LSQ).

LSQ is more general than MLM but not always justified.

Least Squares Fitting

Suppose we have n data points (x_i, y_i, σ_i).

- Assume that we know a functional relationship between the points,

$$y = f(x, a, b \ldots)$$

  - Assume that for each y_i we know x_i exactly.
  - The parameters a, b, … are constants that we wish to determine from out data.

- A procedure to obtain a and b is to minimize the following χ^2 with respect to a and b.

$$\chi^2 = \sum_{i=1}^{n} \frac{[y_i - f(x_i, a, b)]^2}{\sigma_i^2}$$

  - This is very similar to the Maximum Likelihood Method.

For the Gaussian case MLM and LS are identical.

Technically this is a χ^2 distribution only if the y's are from a Gaussian distribution.
Since most of the time the y's are not from a Gaussian we call it "least squares" rather than χ^2.

Least Squares Fitting is sometimes called χ^2 fitting

Example: We have a function with one unknown parameter:

Find b using the least squares technique.

- We need to minimize the following:

$$f(x,b) = 1 + bx$$

- To find the b that minimizes the above function, we do the following:

$$\chi^2 = \sum_{i=1}^{n} \frac{[y_i - f(x_i,a,b)]^2}{\sigma_i^2} = \sum_{i=1}^{n} \frac{[y_i - 1 - bx_i]^2}{\sigma_i^2}$$

$$\frac{\partial \chi^2}{\partial b} = \frac{\partial}{\partial b} \sum_{i=1}^{n} \frac{[y_i - 1 - bx_i]^2}{\sigma_i^2} = \sum_{i=1}^{n} \frac{-2[y_i - 1 - bx_i]x_i}{\sigma_i^2} = 0$$

$$\sum_{i=1}^{n} \frac{y_i x_i}{\sigma_i^2} - \sum_{i=1}^{n} \frac{x_i}{\sigma_i^2} - \sum_{i=1}^{n} \frac{bx_i^2}{\sigma_i^2} = 0$$

Solving for b we find:

$$b = \frac{\sum_{i=1}^{n} \frac{y_i x_i}{\sigma_i^2} - \sum_{i=1}^{n} \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^{n} \frac{x_i^2}{\sigma_i^2}}$$

Note: in this problem we can find a closed form solution for b. However, in general $f(x,a,b)$ will be of the form that does not yield a closed form solution. egs:

$f(x,a,b)=a*\sin(bx)$

$f(x,a,b)=a+e^{-b}$

In these cases we must minimize χ^2 w.r.t a & b using a numerical procedure.

Here each measured data point (y_i) is allowed to have a different standard deviation (σ_i).

The LS technique can be generalized to two or more parameters for simple and complicated (e.g. non-linear) functions.

One especially nice case is a polynomial function that is linear in the unknowns (a_i):

$$f(x, a_1 \dots a_n) = a_1 + a_2 x + a_3 x^2 + a_n x^{n-1}$$

We can always recast problem in terms of solving n simultaneous linear equations.

We use the techniques from linear algebra and invert an nxn matrix to find the a_i's!

Least squares fit of a polynomial

- We've already seen how to fit a n-th degree polynomial to n+1 points (x_i,y_i) i=1,…,n
  - Lagrange interpolation polynomial
- For noisy data we would not want a fit to pass through all the points
  - We want to find a best fit polynomial or *order m*

$$p(x) = c_0 + c_1 x + c_2 x^2 + \ldots + c_m x^m$$

- The sum of the squares of the residuals is

$$S = \sum_{j=1}^{n} (p(x_j) - y_j)^2$$

Minimize the "error"

- We next minimize S by taking partial derivatives *w.r.t.* the coefficients c_n
  - *e.g.* for c_0:

$$0=\frac{\partial S}{\partial c_0}=\frac{\partial}{\partial c_0}\sum_{j=1}^{n}(c_0+c_1x_j+c_2x_j^2+\ldots+c_mx_j^m-y_j)^2$$

$$0=2\sum_{j=1}^{n}(c_0+c_1x_j+c_2x_j^2+\ldots+c_mx_j^m-y_j)$$

$$\Rightarrow nc_0+c_1\sum_{j=1}^{n}x_j+c_2\sum_{j=1}^{n}x_j^2+\ldots+c_m\sum_{j=1}^{n}x_j^m=\sum_{j=1}^{n}y_j$$

More examples

- For c_1 we find:

$$0=\frac{\partial S}{\partial c_1}=\frac{\partial}{\partial c_1}\sum_{j=1}^{n}(c_0+c_1x_j+c_2x_j^2+\ldots+c_mx_j^m-y_j)^2$$

$$0=\sum_{j=1}^{n}2x_j(c_0+c_1x_j+c_2x_j^2+\ldots+c_mx_j^m-y_j)$$

$$\Rightarrow c_0\sum_{j=1}^{n}x_j+c_1\sum_{j=1}^{n}x_j^2+c_2\sum_{j=1}^{n}x_j^3+\ldots+c_m\sum_{j=1}^{n}x_j^{m+1}=\sum_{j=1}^{n}x_jy_j$$

- and so on… Taking partials *w.r.t.* to m+1 coefficients & equating to zero gives m+1 equations in m+1 unknowns (the c_j j=0,…,m)

Matrix form

$$\begin{bmatrix} n & \sum_{j=1}^{n} x_j & \sum_{j=1}^{n} x_j^2 & \cdots & \sum_{j=1}^{n} x_j^m \\ \sum_{j=1}^{n} x_j & \sum_{j=1}^{n} x_j^2 & \sum_{j=1}^{n} x_j^3 & \cdots & \sum_{j=1}^{n} x_j^{m+1} \\ \sum_{j=1}^{n} x_j^2 & \sum_{j=1}^{n} x_j^3 & \sum_{j=1}^{n} x_j^4 & \cdots & \sum_{j=1}^{n} x_j^{m+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{j=1}^{n} x_j^m & \sum_{j=1}^{n} x_j^{m+1} & \sum_{j=1}^{n} x_j^{m+2} & \cdots & \sum_{j=1}^{n} x_j^{2m} \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{n} y_j \\ \sum_{j=1}^{n} x_j y_j \\ \sum_{j=1}^{n} x_j^2 y_j \\ \vdots \\ \sum_{j=1}^{n} x_j^m y_j \end{bmatrix}$$

- These are the so-called "normal equations" in analogy with normals to the level set of a function
- Can be solved using the LU decomposition method we just discussed to find the coefficients c_j

How do we choose *m*?

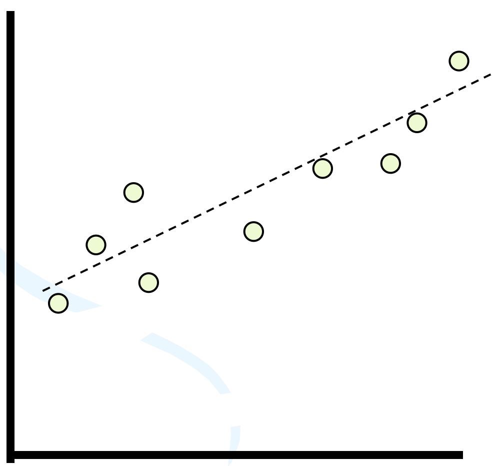
- Higher order fits are not always better

- Many dependent variables are inherently following some underlying “law.”
  - If the underlying relationship were quadratic, fitting a cubic would be largely pointless

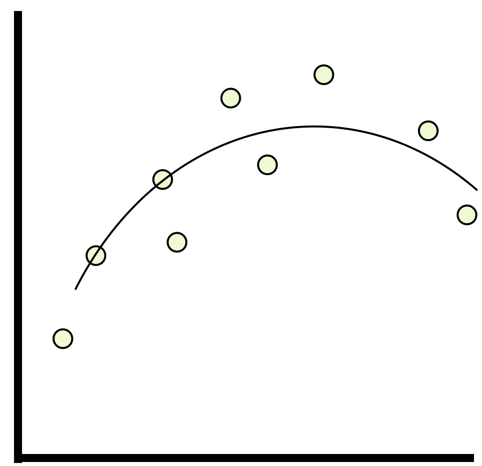
- However, we don’t always know the underlying relationship
  - So we need to look at the data…

Strategy

- Plot the data
  - If linear looks OK then start with m=1
  - If curve is visible try m=2 – if higher order then try log-log graph to estimate m



Linear fit looks reasonable
Start with m=1



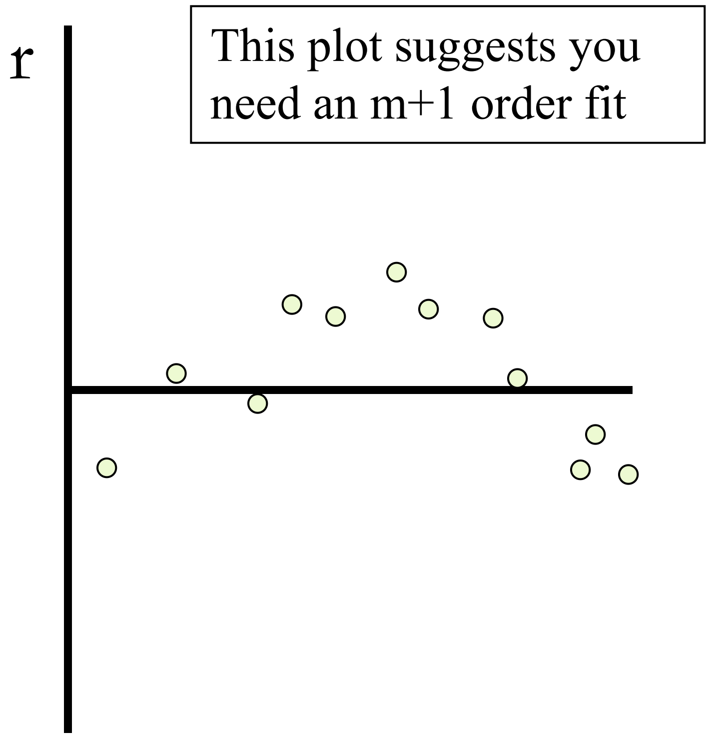
Strong suggestion of underlying curve
Start with m=2

Strategy cont.

- After each fit evaluate S_m and plot the residuals $r_i = p(x_i) - y_i$
- If $S_m \sim S_{m\text{-}1}$ then this indicates that the mth (or m-1 th) order is about as good as you can do
- Plotting the residuals will also give intuition to the quality of fit
  - Residuals should scatter around zero with no implicit “bias”

Residuals

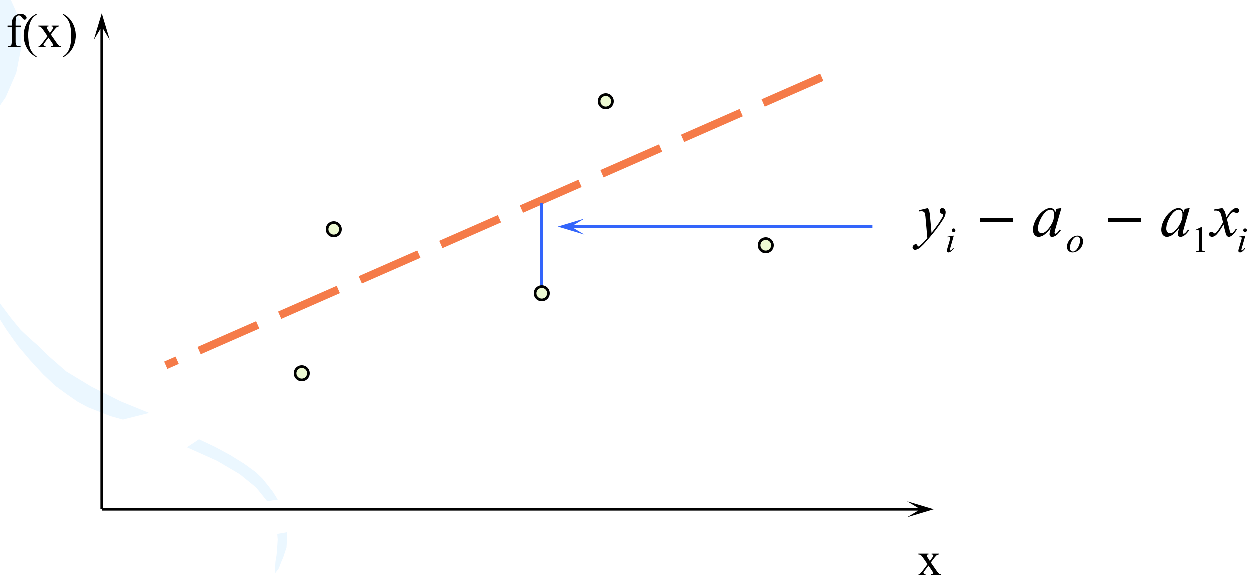
Residuals scattered about zero with
A large number of sign changes



Residuals clearly have a positive
bias with fewer sign changes

Sum of the Residual Error, S_r

$$S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(y_i - a_o - a_1 x_i \right)^2$$

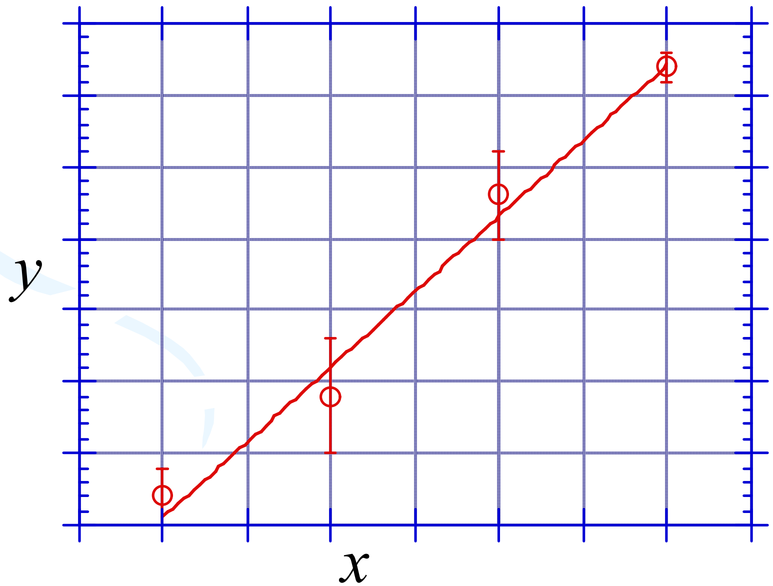


Example: Given the following data perform a least squares fit to find the value of b

x	1.0	2.0	3.0	4.0
y	2.2	2.9	4.3	5.2
σ	0.2	0.4	0.3	0.1

$$f(x,b) = 1 + bx$$

- Using the above expression for b we calculate: $\boldsymbol{b = 1.05}$
- Next is a plot of the data points and the line from the least squares fit:



Error Band?

If we assume that the data points are from a Gaussian distribution, we can calculate a χ^2 and the probability associated with the fit.

$$\chi^2 = \sum_{i=1}^{n} \frac{[y_i - 1 - 1.05x_i]^2}{\sigma_i^2}$$

$$= \left(\frac{2.2-2.05}{0.2}\right)^2 + \left(\frac{2.9-3.1}{0.4}\right)^2 + \left(\frac{4.3-4.16}{0.3}\right)^2 + \left(\frac{5.2-5.2}{0.1}\right)^2 = 1.04$$

From Table (some book?)

The probability to get $\chi^2 \geq 1.04$ for 3 degrees of freedom $\approx$ 80%.
We call this a **"good"** fit since the probability is close to 100%.

If however the χ^2 was large (e.g. 15),
the probability would be small ($\approx$ 0.2% for 3 dof).
We would say this was a **"bad"** fit.

RULE OF THUMB: A "good" fit has χ^2 /dof $\leq$ 1

More on Least Squares Fit (LSQF)

We discussed how we can fit our data points to a linear function (straight line) and get the "best" estimate of the slope and intercept.

However, we did not discuss two important issues:

I) How to estimate the uncertainties on our slope and intercept obtained from a LSQF?

II) How to apply the LSQF when we have a non-linear function?

More on Least Squares Fit (LSQF)

Estimation of Errors on parameters determined from a LSQF

Assume we have data points that lie on a straight line:

$$y = \alpha + \beta x$$

Assume we have n measurements of the x's and y's.

For simplicity, assume that each y measurement has same error, σ.

Assume that x is known much more accurately than y.

$\Rightarrow$ ignore any uncertainty associated with x.

Previously we showed that the solution for the intercept α and slope β is:

$$\alpha = \frac{\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i y_i \sum_{i=1}^{n} x_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \text{ and } \beta = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$\alpha = \frac{\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i y_i \sum_{i=1}^{n} x_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \text{ and } \beta = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

Since α and β are functions of the measurements (y_i's) **we can** use the Propagation of Errors technique to estimate σ_α and σ_β.

$$\sigma_Q^2 = \sigma_x^2 \left(\frac{\partial Q}{\partial x}\right)^2 + \sigma_y^2 \left(\frac{\partial Q}{\partial y}\right)^2 + 2\sigma_{xy}\left(\frac{\partial Q}{\partial x}\right)\left(\frac{\partial Q}{\partial y}\right)$$

Assume that:

a) Each measurement is independent of each other (σ_{xy}=0).

b) We can neglect any error (σ) associated with x.

c) Each y measurement has the same σ, i.e. $\sigma_i=\sigma$.

Assume that: a) Each measurement is independent of each other ($\sigma_{xy}=0$).
b) We can neglect any error (σ) associated with x.
c) Each y measurement has the same σ, i.e. $\sigma_i=\sigma$.

$$\sigma_Q^2=\sigma_x^2\left(\frac{\partial Q}{\partial x}\right)^2+\sigma_y^2\left(\frac{\partial Q}{\partial y}\right)^2+2\sigma_{xy}\left(\frac{\partial Q}{\partial x}\right)\left(\frac{\partial Q}{\partial y}\right)$$

$$\sigma_Q^2=\sigma_x^2\left(\frac{\partial Q}{\partial x}\right)^2+\sigma_y^2\left(\frac{\partial Q}{\partial y}\right)^2$$ **assumption a)**

$$\sigma_Q^2=\sigma_y^2\left(\frac{\partial Q}{\partial y}\right)^2$$ **assumption b)**

$$\sigma_\alpha^2=\sum_{i=1}^{n}\sigma_{y_i}^2\left(\frac{\partial \alpha}{\partial y_i}\right)^2=\sigma^2\sum_{i=1}^{n}\left(\frac{\partial \alpha}{\partial y_i}\right)^2$$ **assumption c)**

$$\frac{\partial \alpha}{\partial y_i}=\frac{\partial}{\partial y_i}\frac{\sum_{i=1}^{n}y_i\sum_{j=1}^{n}x_j^2-\sum_{i=1}^{n}x_iy_i\sum_{j=1}^{n}x_j}{n\sum_{i=1}^{n}x_i^2-(\sum_{i=1}^{n}x_i)^2}=\frac{\sum_{j=1}^{n}x_j^2-x_i\sum_{j=1}^{n}x_j}{n\sum_{i=1}^{n}x_i^2-(\sum_{i=1}^{n}x_i)^2}$$ **plugging our formula for α**

$$\sigma_\alpha^2=\sigma^2\sum_{i=1}^{n}\left(\frac{\sum_{j=1}^{n}x_j^2-x_i\sum_{j=1}^{n}x_j}{n\sum_{i=1}^{n}x_i^2-(\sum_{i=1}^{n}x_i)^2}\right)^2=\sigma^2\sum_{i=1}^{n}\left(\frac{(\sum_{j=1}^{n}x_j^2)^2+x_i^2(\sum_{j=1}^{n}x_j)^2-2x_i\sum_{j=1}^{n}x_j\sum_{j=1}^{n}x_j^2}{(n\sum_{i=1}^{n}x_i^2-(\sum_{i=1}^{n}x_i)^2)^2}\right)$$

$$\sigma_\alpha^2 = \sigma^2 \frac{n(\sum_{j=1}^{n} x_j^2)^2 + \sum_{i=1}^{n} x_i^2 (\sum_{j=1}^{n} x_j)^2 - 2(\sum_{j=1}^{n} x_j)^2 \sum_{j=1}^{n} x_j^2}{(n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2)^2} = \sigma^2 \frac{n(\sum_{j=1}^{n} x_j^2)^2 - \sum_{i=1}^{n} x_i^2 (\sum_{j=1}^{n} x_j)^2}{(n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2)^2}$$

$$= \sigma^2 \sum_{j=1}^{n} x_j^2 \frac{n\sum_{j=1}^{n} x_j^2 - (\sum_{j=1}^{n} x_j)^2}{(n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2)^2}$$

$$\boxed{\sigma_\alpha^2 = \sigma^2 \frac{\sum_{j=1}^{n} x_j^2}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}}$$

variance in the intercept

We can find the variance in the slope (β) using exactly the same procedure:

$$\sigma_\beta^2=\sum_{i=1}^{n}\sigma_{y_i}^2\left(\frac{\partial\beta}{\partial y_i}\right)^2=\sigma^2\sum_{i=1}^{n}\left(\frac{\partial\beta}{\partial y_i}\right)^2=\sigma^2\sum_{i=1}^{n}\left(\frac{nx_i-\sum_{j=1}^{n}x_j}{n\sum_{i=1}^{n}x_i^2-(\sum_{i=1}^{n}x_i)^2}\right)^2$$

$$=\sigma^2\frac{n^2\sum_{j=1}^{n}x_j^2+n(\sum_{j=1}^{n}x_j)^2-2n\sum_{i=1}^{n}x_i\sum_{j=1}^{n}x_j}{(n\sum_{i=1}^{n}x_i^2-(\sum_{i=1}^{n}x_i)^2)^2}$$

We can find the variance in the slope (β) using exactly the same procedure:

$$\boxed{\sigma_\beta^2=\frac{n\sigma^2}{n\sum_{i=1}^{n}x_i^2-(\sum_{i=1}^{n}x_i)^2}}$$

variance in the slope

If we don't know the true value of σ, we can estimate the variance using the spread between the measurements (y_i's) and the fitted values of y:

$$\sigma^2 \approx \frac{1}{n-2}\sum_{i=1}^{n}(y_i - y_i^{fit})^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2$$

$n - 2$ = number of degree of freedom

= number of data points - number of parameters (α, β) extracted from the data

$$\sigma_\alpha^2 = \frac{1}{D}\sum_{i=1}^{n}\frac{x_i^2}{\sigma_i^2}$$

$$\sigma_\beta^2 = \frac{1}{D}\sum_{i=1}^{n}\frac{1}{\sigma_i^2}$$

$$D = \sum_{i=1}^{n}\frac{1}{\sigma_i^2}\sum_{i=1}^{n}\frac{x_i^2}{\sigma_i^2} - (\sum_{i=1}^{n}\frac{x_i}{\sigma_i^2})^2$$

weighted slope and intercept

The above expressions simplify to the "equal variance" case.

Don't forget to keep track of the "n's" when factoring out equal σ's. For example:

$$\sum_{i=1}^{n}\frac{1}{\sigma_i^2} = \frac{n}{\sigma^2} \quad not \quad \frac{1}{\sigma^2}$$

Example: Fit to straight line

x	1.0	2.0	3.0	4.0
y	2.2	2.9	4.3	5.2
σ	0.2	0.4	0.3	0.1

(same data)

Previous fit to: $f(x,b)=1+bx$, $b=1.05$

New fit to: $f(x,b)=a+bx$,

$a=1.158\pm0.253$ $b=1.011\pm0.072$

a and b calculated using weighted fit

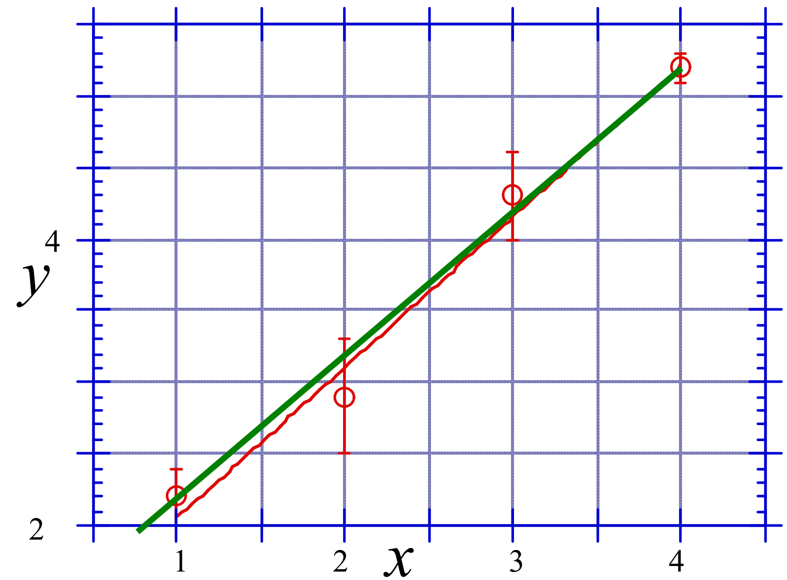
$$a=\frac{\sum_{i=1}^{n}\frac{x_i^2}{\sigma_i^2}\sum_{i=1}^{n}\frac{y_i}{\sigma_i^2}-\sum_{i=1}^{n}\frac{x_i}{\sigma_i^2}\sum_{i=1}^{n}\frac{x_iy_i}{\sigma_i^2}}{D}$$

$$b=\frac{\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\sum_{i=1}^{n}\frac{x_iy_i}{\sigma_i^2}-\sum_{i=1}^{n}\frac{x_i}{\sigma_i^2}\sum_{i=1}^{n}\frac{y_i}{\sigma_i^2}}{D}$$

$$D=\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\sum_{i=1}^{n}\frac{x_i^2}{\sigma_i^2}-\left(\sum_{i=1}^{n}\frac{x_i}{\sigma_i^2}\right)^2$$

Errors on a and b calculated using:

$$\sigma_\alpha^2=\frac{1}{D}\sum_{i=1}^{n}\frac{x_i^2}{\sigma_i^2} \qquad \sigma_\beta^2=\frac{1}{D}\sum_{i=1}^{n}\frac{1}{\sigma_i^2}$$



$\chi^2=0.65$ for 2 degrees of freedom

P($\chi^2>0.65$ for 2 dof) = 72%

The Error on the Mean

Question: If we have a set of measurements of the same quantity:

$$x_1 \pm \sigma_1 \quad x_2 \pm \sigma_2 \ldots x_n \pm \sigma_n$$

What's the best way to combine these measurements?

How to calculate the variance once we combine the measurements?

Assuming Gaussian statistics, the Maximum Likelihood Method says combine the measurements as:

$$x = \frac{\sum_{i=1}^{n} x_i / \sigma_i^2}{\sum_{i=1}^{n} 1/\sigma_i^2}$$

weighted average

If all the variances ($\sigma^2_1 = \sigma^2_2 = \sigma^2_3 ..$) are the same:

$$x = \frac{1}{n} \sum_{i=1}^{n} x_i$$

unweighted average

The variance of the weighted average can be calculated using propagation of errors:

$$\sigma_x^2 = \sum_{i=1}^{n}\left[\frac{\partial}{\partial x_i}x\right]^2 \sigma_i^2$$

And the derivatives are:

$$\frac{\partial}{\partial x_i}x = \frac{1/\sigma_i^2}{\sum_{i=1}^{n} 1/\sigma_i^2}$$

$$\sigma_x^2 = \sum_{i=1}^{n}\left[\frac{\partial}{\partial x_i}x\right]^2 \sigma_i^2 = \sum_{i=1}^{n}\frac{1/\sigma_i^4}{\left[\sum_{i=1}^{n} 1/\sigma_i^2\right]^2}\sigma_i^2 = \frac{1}{\left[\sum_{i=1}^{n} 1/\sigma_i^2\right]^2}\sum_{i=1}^{n} 1/\sigma_i^2 = \frac{1}{\sum_{i=1}^{n}\frac{1}{\sigma_i^2}}$$

$$\boxed{\sigma_x^2 = \frac{1}{\sum_{i=1}^{n}\frac{1}{\sigma_i^2}}}$$

σ_x=error in the weighted mean

If all the variances are the same

$$\sigma_x^2 = \frac{1}{\sum_{i=1}^{n} 1/\sigma_i^2} = \frac{\sigma^2}{n}$$

The error in the mean (σ_x) gets smaller as the number of measurements (n) increases.

Don't confuse the error in the mean (σ_x) with the standard deviation of the distribution (σ)!

If we make more measurements:

The standard deviation (σ) of the (gaussian) distribution remains the same,

BUT the error in the mean (σ_x) decreases

Example: Two experiments measure the mass of the proton:

Experiment 1 measures m_p=950±25 MeV

Experiment 2 measures m_p=936±5 MeV

Using just the average of the two experiments we find:

$$m_p=(950+936)/2=943\text{MeV}$$

Using the weighted average of the two experiments we find:

$$m_p=(950/25^2+936/5^2)/(1/25^2+1/5^2)=936.5\text{MeV}$$

and the variance:

$$\sigma^2=1/(1/25^2+1/5^2)=24\ \text{MeV}^2$$

$$\sigma=4.9\ \text{MeV}$$

Since experiment 2 was more precise than experiment 1 we would expect the final result to be closer to experiment 2's value.

$$m_p=(936.5\pm4.9)\ \text{MeV}$$

Ⓛ LSQF with non-linear functions:

– For our purposes, a non-linear function is a function where one or more of the parameters that we are trying to determine (e.g. α, β from the straight line fit) is raised to a power other than 1.

- Example: functions that are *non-linear* in the parameter τ.

$$y = A + x/\tau$$

$$y = A + x\tau^2$$

$$y = Ae^{-x/\tau}$$

However, these functions are *linear* in the parameters A.

– The problem with most non-linear functions is that we cannot write down a solution for the parameters in a closed form using, for example, the techniques of linear algebra (i.e. matrices).

- Usually non-linear problems are solved numerically using a computer or calculator.

Levenberg-Marquardt Method?

- Sometimes by a change of variable(s) we can turn a non-linear problem into a linear one.

  Example: take the natural log of both sides of the above exponential equation:

$$\ln y = \ln A - x/\tau = C + Dx$$

  Now a linear problem in the parameters *C* and *D*! $\boxed{\tau = -1/D}$

  In fact it's just a straight line!

  To measure the lifetime τ (Lab 7) we first fit for *D* and then transform *D* into τ.

- Example: Decay of a radioactive substance. Fit the following data to find N_0 and τ:

$$N(t) = N_0 e^{-t/\tau}$$

  - *N* represents the amount of the substance present at time *t*.
  - N_0 is the amount of the substance at the beginning of the experiment ($t = 0$).
  - τ is the lifetime of the substance.

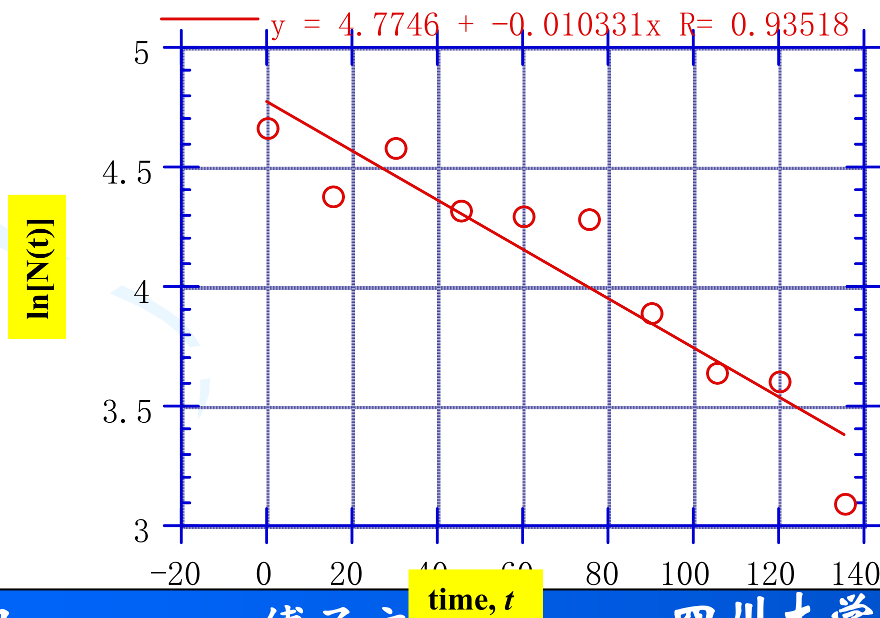
Lifetime Data

i	1	2	3	4	5	6	7	8	9	10
t_i	0	15	30	45	60	75	90	105	120	135
N_i	106	80	98	75	74	73	49	38	37	22
$y_i = \ln N_i$	4.663	4.382	4.585	4.317	4.304	4.290	3.892	3.638	3.611	3.091

$$D = \frac{n\sum_{i=1}^{n} t_i y_i - \sum_{i=1}^{n} y_i \sum_{i=1}^{n} t_i}{n\sum_{i=1}^{n} t_i^2 - (\sum_{i=1}^{n} t_i)^2} = \frac{10\times 2560.41 - 40.773\times 675}{10\times 64125-(675)^2} = -0.01033$$

$$\tau = -1/D = 96.80 \text{ seconds}$$

- The intercept is given by: $C = 4.77 = \ln A$ or $A = 117.9$



Example: Find the values A and τ taking into account the uncertainties in the data points.

Weighted formulas are found in some books ($w_i = 1/\sigma_i^2$)

The uncertainty in the number of radioactive decays is governed by Poisson statistics.

The number of counts N_i in a bin is assumed to be the average (μ) of a Poisson distribution:

$\mu = N_i$ = Variance

The variance of y_i(= ln N_i) can be calculated using propagation of errors:

$$\sigma_y^2 = \sigma_N^2 (\partial y / \partial N)^2 = (N)(\partial \ln N / \partial N)^2 = (N)(1/N)^2 = 1/N$$

The slope and intercept from a straight line fit that includes uncertainties in the data points:

$$\alpha = \frac{\sum_{i=1}^{n} \frac{y_i}{\sigma_i^2} \sum_{i=1}^{n} \frac{t_i^2}{\sigma_i^2} - \sum_{i=1}^{n} \frac{t_i y_i}{\sigma_i^2} \sum_{i=1}^{n} \frac{t_i}{\sigma_i^2}}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2} \sum_{i=1}^{n} \frac{t_i^2}{\sigma_i^2} - (\sum_{i=1}^{n} \frac{t_i}{\sigma_i^2})^2} \qquad \beta = \frac{\sum_{i=1}^{n} \frac{1}{\sigma_i^2} \sum_{i=1}^{n} \frac{t_i y_i}{\sigma_i^2} - \sum_{i=1}^{n} \frac{t_i}{\sigma_i^2} \sum_{i=1}^{n} \frac{y_i}{\sigma_i^2}}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2} \sum_{i=1}^{n} \frac{t_i^2}{\sigma_i^2} - (\sum_{i=1}^{n} \frac{t_i}{\sigma_i^2})^2} = \frac{\sum_{i=1}^{n} N_i \sum_{i=1}^{n} N_i t_i \ln N_i - \sum_{i=1}^{n} N_i t_i \sum_{i=1}^{n} N_i \ln N_i}{\sum_{i=1}^{n} N_i \sum_{i=1}^{n} N_i t_i^2 - (\sum_{i=1}^{n} N_i t_i)^2}$$

If all the σ's are the same then the expressions are identical to the unweighted case.

α=4.725 and β = -0.00903

$\tau = -1/\beta = -1/0.00903 = 110.7$ sec

To calculate the error on the lifetime (τ), we first must calculate the error on β:

$$\sigma_\beta^2 = \frac{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2} \sum_{i=1}^{n} \frac{t_i^2}{\sigma_i^2} - (\sum_{i=1}^{n} \frac{t_i}{\sigma_i^2})^2} = \frac{\sum_{i=1}^{n} N_i}{\sum_{i=1}^{n} N_i \sum_{i=1}^{n} N_i t_i^2 - (\sum_{i=1}^{n} N_i t_i)^2} = \frac{652}{652 \times 2684700 - (33240)^2} = 1.01 \times 10^{-6}$$

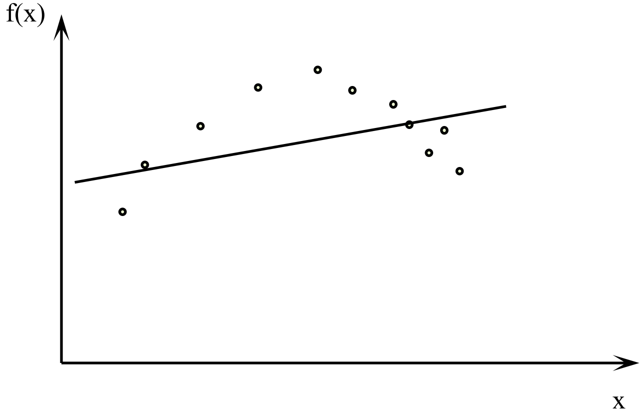
$$\sigma_\tau^2 = \sigma_\beta^2 (\partial \tau / \partial \beta)^2 \Rightarrow \sigma_\tau = \sigma_\beta (1/\beta^2) = \frac{1.005 \times 10^{-3}}{(9.03 \times 10^{-3})^2} = 12.3$$

The experimentally determined lifetime is: $\tau = 110.7 \pm 12.3$ sec.

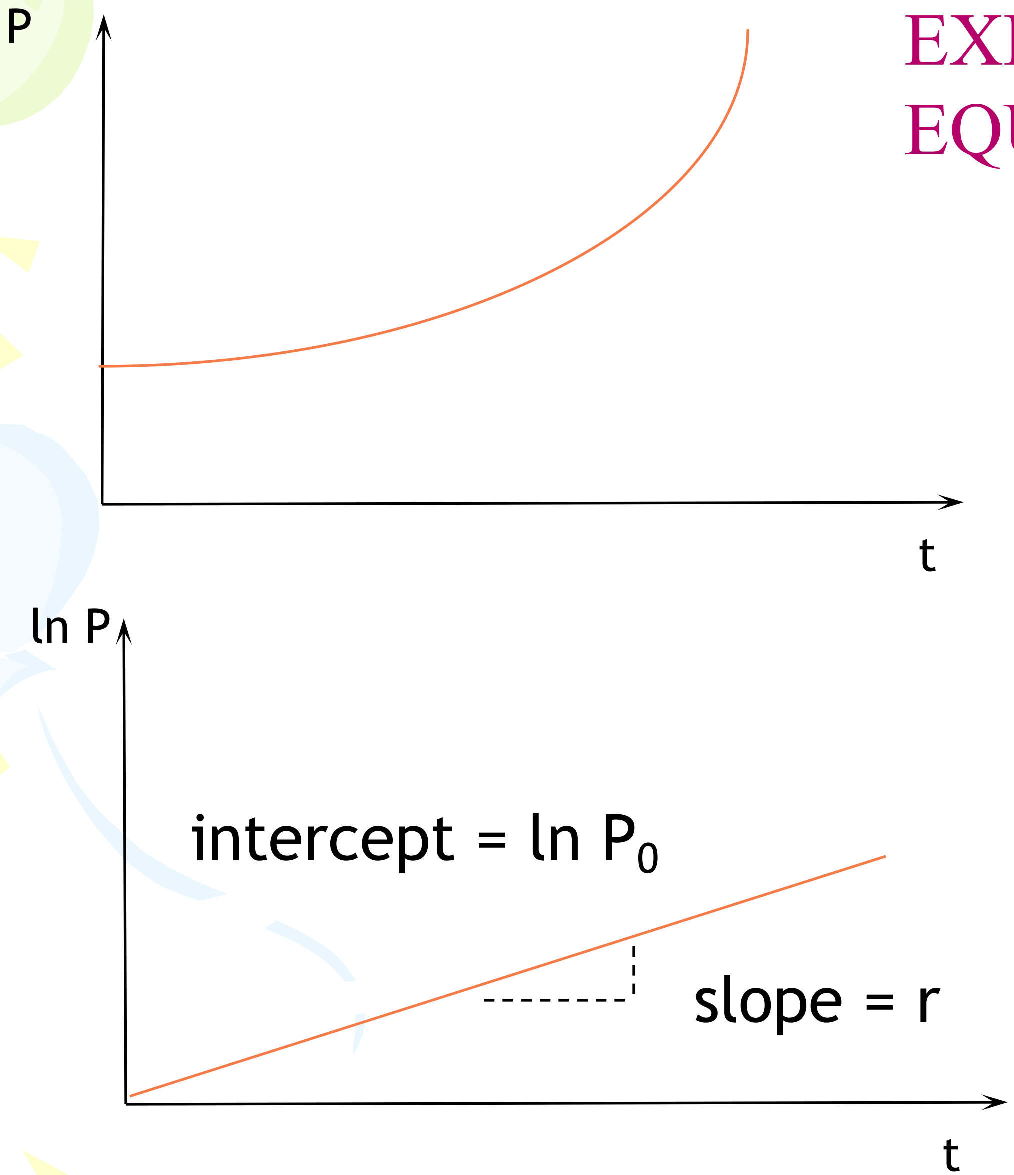
Linearization of non-linear relationships

Some data is simply ill-suited for linear least squares regression....

or so it appears.



EXPONENTIAL EQUATIONS



$$P = P_o e^{rt}$$

Linearize

why?

$$P = P_0 e^{rt}$$

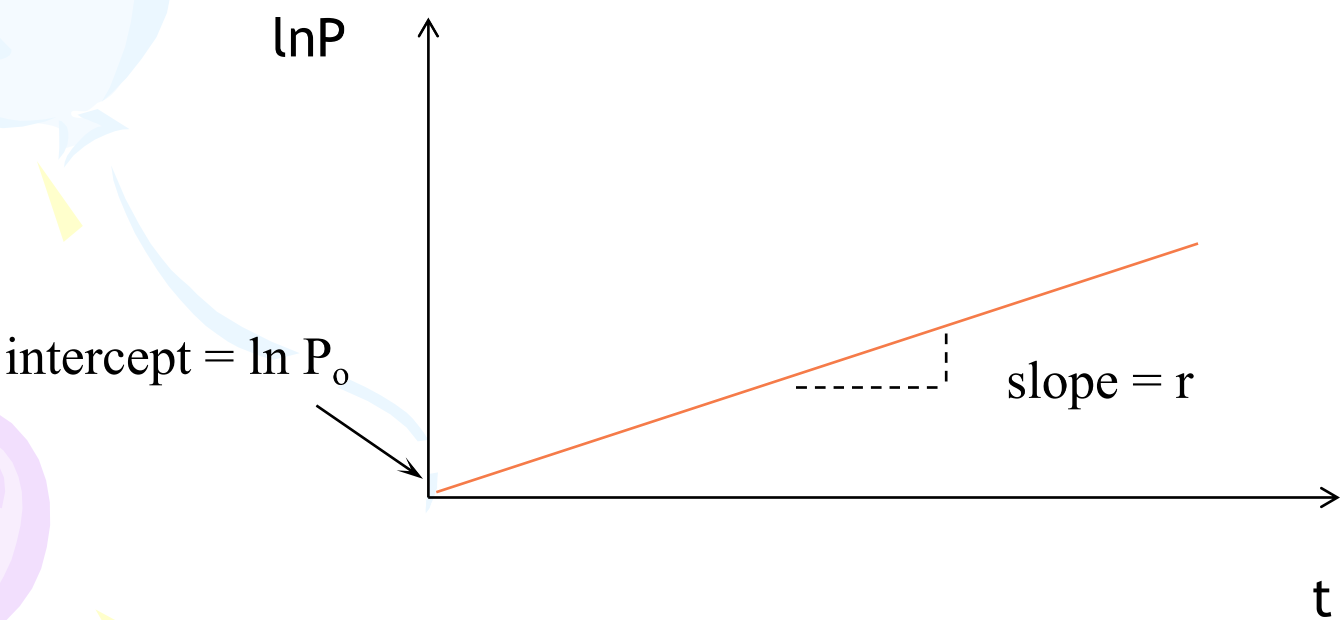
$$\ln P = \ln\left(P_0 e^{rt}\right)$$

$$= \ln\left(P_0\right) + \ln\left(e^{rt}\right)$$

$$= \ln\left(P_0\right) + rt$$

Can you see the similarity with the equation for a line:

$y = a_o + a_1 x$



$$P = P_0 e^{rt}$$

$$\ln P = \ln\left(P_0 e^{rt}\right)$$

$$= \ln\left(P_0\right) + \ln\left(e^{rt}\right)$$

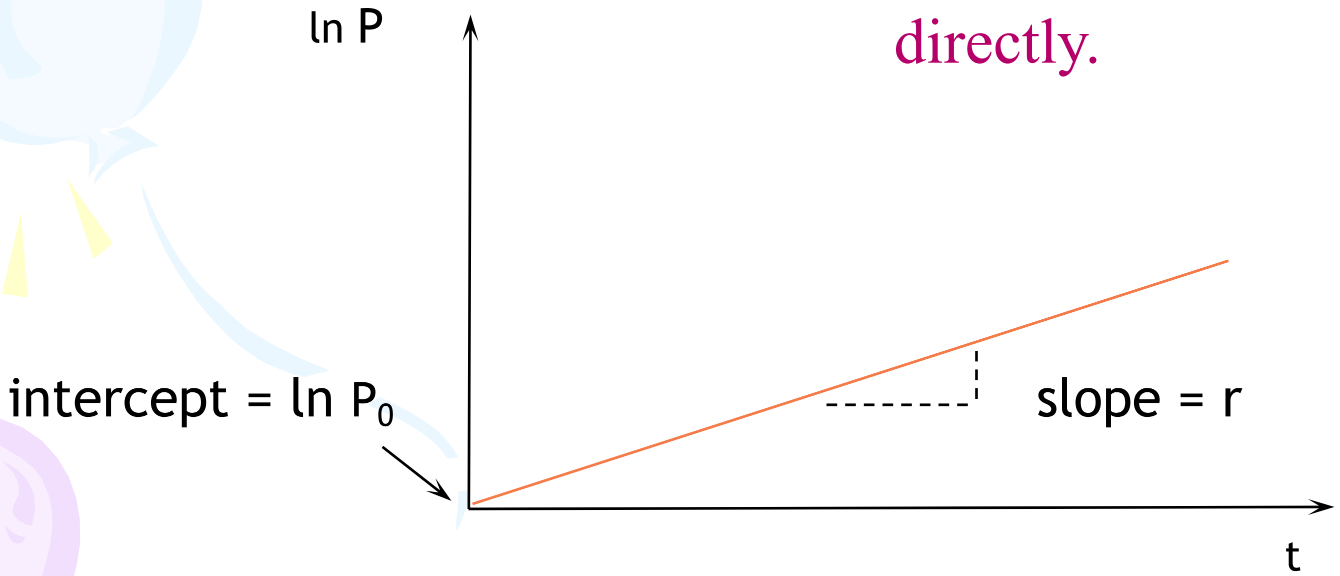
$$= \ln\left(P_0\right) + rt$$

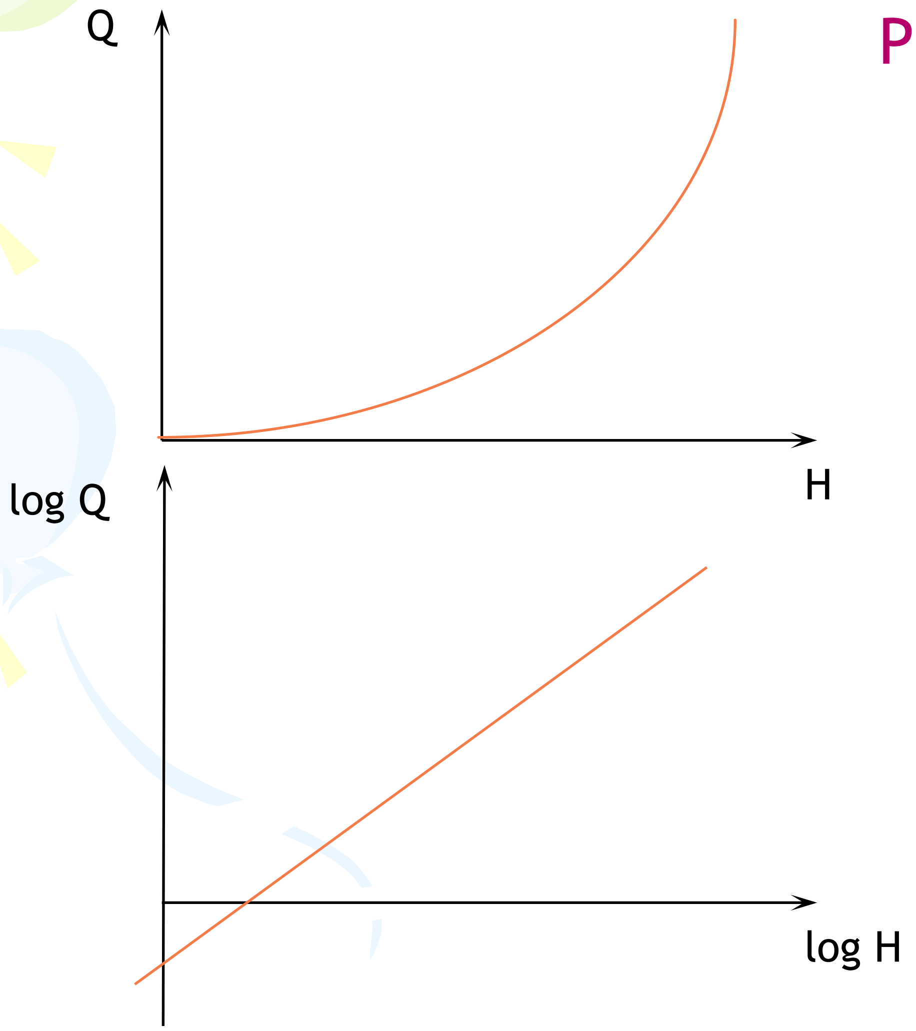
After taking the natural log of the y-data, perform linear regression.

From this regression:

The value of a_o will give us ln (P_0). Hence, $P_0 = e^{ao}$

The value of a_1 will give us r directly.





POWER EQUATIONS

$$Q = cH^{a}$$

(Flow over a weir)

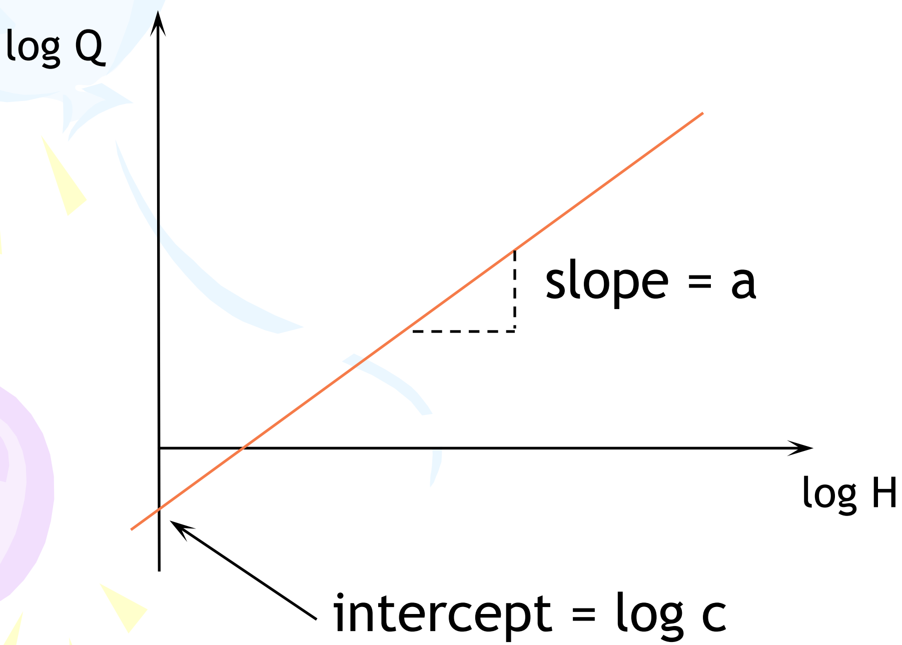
Here we linearize the equation by taking the log of *H* and *Q* data. What is the resulting intercept and slope?

$$Q = cH^a$$

$$\log Q = \log\left(cH^a\right)$$

$$= \log c + \log H^a$$

$$= \log c + a \log H$$



So how do we get *c* and *a* from performing regression on the log H vs log Q data?

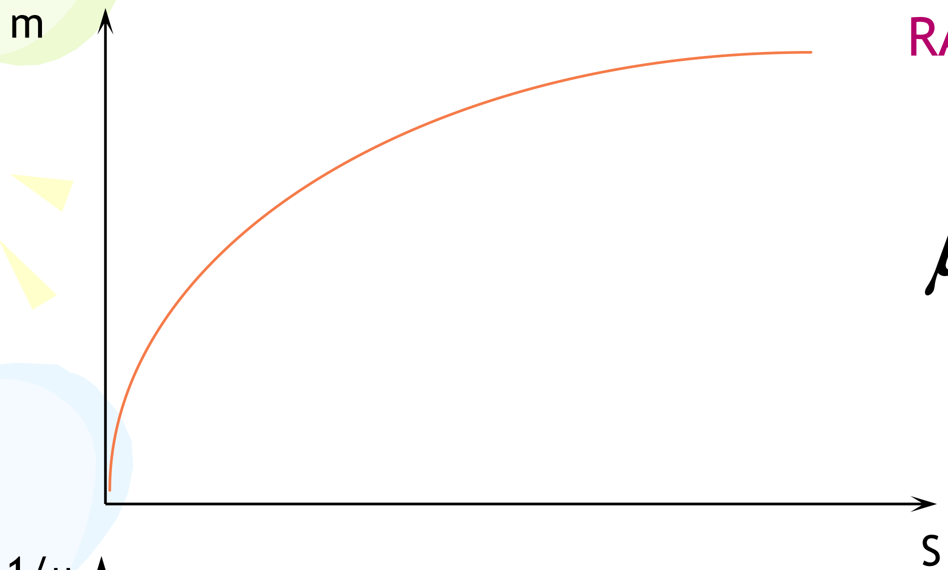
From : $y = a_o + a_1 x$

$a_o = \log c$

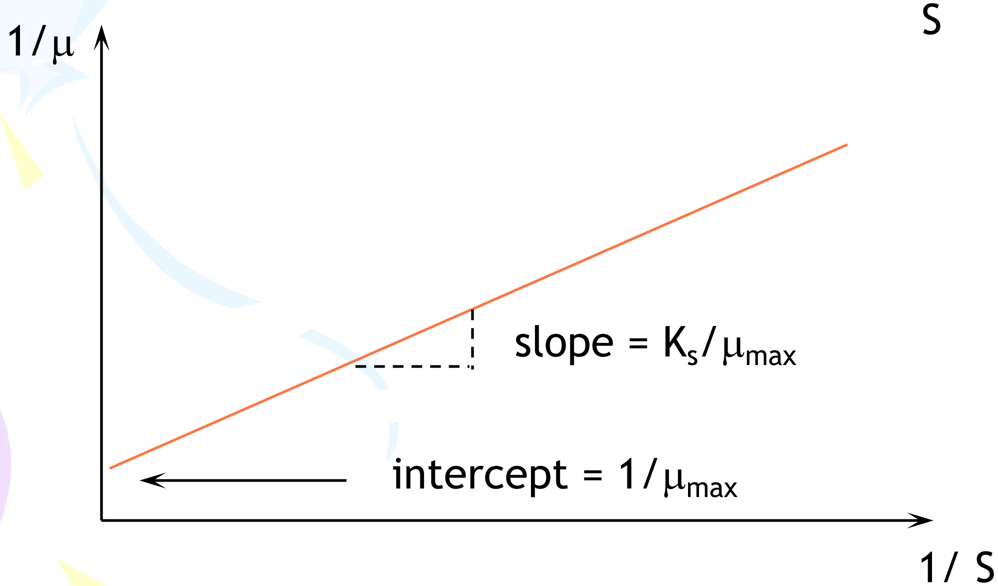
$c = 10^{ao}$

$a_1 = a$

SATURATION-GROWTH RATE EQUATION



$$\mu = \mu_{max} \frac{S}{K_s + S}$$



Here, m is the growth rate of a microbial population, m_{max} is the maximum growth rate, S is the substrate or food concentration, K_s is the substrate concentration at a value of $m = m_{max}/2$

Example

X	Y	Y_{model}
1	0.7	0.8
2	1.7	1.6
3	3.3	3.1
4	7.3	6.2
5	10.9	12.1
6	22.7	23.9

$$y = 0.407e^{0.679x}$$

Extending the approach to fitting functions

- We can use the method developed to fit a continuous function $f(x)$ in an interval $[a, b]$

- Rather than a sum of squares we now have an integral that we need to minimize

$$S = \sum_{j=1}^{n} (p(x_j) - y_j)^2 \to \int_a^b (f(x) - p(x))^2 \, dx$$

- As an example, let $f(x)$=sin(px) and the limits be [0,1], find the best quadratic ($c_0+c_1x+c_2x^2$) to fit
  - –We need to minimize

$$S = \int_a^b (\sin(\pi x) - (c_0 + c_1 x + c_2 x^2))^2 \, dx$$

Compute derivatives

- We work in the same way as the earlier fit:

$$\frac{\partial S}{\partial c_0}=0 \quad \Rightarrow \quad c_0\int_0^1 dx + c_1\int_0^1 x dx + c_2\int_0^1 x^2 dx = \int_0^1 \sin \pi x dx$$

$$\frac{\partial S}{\partial c_1}=0 \quad \Rightarrow \quad c_0\int_0^1 x dx + c_1\int_0^1 x^2 dx + c_2\int_0^1 x^3 dx = \int_0^1 x\sin \pi x dx$$

$$\frac{\partial S}{\partial c_2}=0 \quad \Rightarrow \quad c_0\int_0^1 x^2 dx + c_1\int_0^1 x^3 dx + c_2\int_0^1 x^4 dx = \int_0^1 x^2\sin \pi x dx$$

- All integrals are tractable, and in matrix form we get:

$$\begin{bmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{bmatrix}\begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 2/\pi \\ 1/\pi \\ (\pi^2-4)/\pi^3 \end{bmatrix}$$

Comparing to Taylor expansion

- If we solve the system on the previous page we derive a polynomial

$$p(x) = -0.0505 + 4.1225\,x - 4.1225\,x^2 \qquad \text{where } 0 \le \text{x} \le 1$$

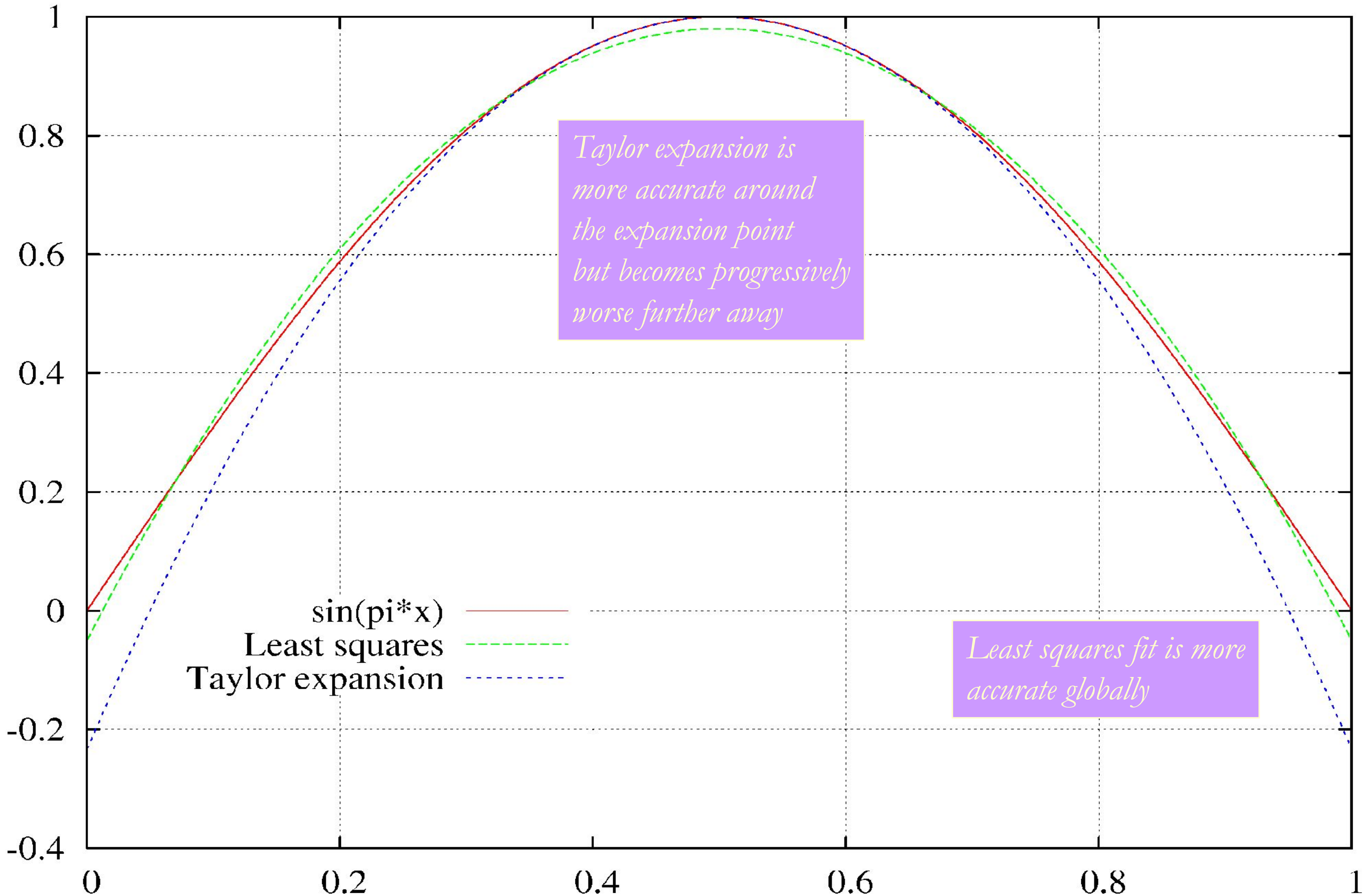
- We can compare this to the Taylor expansion to second order about x=1/2

$$t(x) = \underbrace{f(1/2)}_{=1} + \left(x - \frac{1}{2}\right)\underbrace{f'(1/2)}_{=0} + \frac{1}{2}\left(x - \frac{1}{2}\right)^2 \underbrace{f''(1/2)}_{=-\pi^2}$$

$$= 1 - \frac{\pi^2}{2}\left(x - \frac{1}{2}\right)^2 = 1 - \frac{\pi^2}{8} + \frac{\pi^2}{2}x - \frac{\pi^2}{2}x^2$$

$$= -0.2337 + 4.9348\,x - 4.9348x^2$$

Notice that the least squares fit and Taylor expansion are different


Taylor expansion is more accurate around the expansion point but becomes progressively worse further away
Least squares fit is more accurate globally
sin(pi*x)
Least squares
Taylor expansion
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
0
0.2
0.4
0.6
0.8
1

Hilbert Matrices

- The specific form of the coefficient matrix is called a "Hilbert Matrix" and arises because of the polynomial fit
  - *Key point*: $f(x)$ only affects the right hand side

$$\begin{bmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 2/\pi \\ 1/\pi \\ (\pi^2 - 4)/\pi^3 \end{bmatrix}$$

- For a polynomial fit of degree m the Hilbert Matrix has a dimension of $m+1$
  - As m increases, the determinant $\rightarrow 0$ for $m \rightarrow \infty$
  - This creates significant numerical difficulties

Hilbert Matrices:problems & solutions

- The Hilbert Matrix is classic example of an *ill conditioned matrix*

- By m=4 (*i.e.* a 5×5 Hilbert matrix) the small determinant causes difficulty in single precision

- By m=8 (*i.e.* a 9×9 Hilbert matrix) problems at double precision (real*8)

- We can improve upon the fitting method by choosing to fit a sum of polynomials:

  Instead of $p(x) = c_0 + c_1 x + c_2 x_2 + \ldots + c_m x^m$

  we use $p(x) = \alpha_0 p_0(x) + \alpha_1 p_1(x) + \alpha_2 p_2(x) + \ldots + \alpha_m p_m(x)$

  p_i are i-th degree polynomials

Fitting over basis polynomials

- We'll postpone the nature of the $p_i(x)$ for a moment…
- We now need to find the a_i by minimizing the least square function S

$$S = \int_a^b \left(f(x) - \sum_{j=0}^{m} \alpha_j p_j(x) \right)^2 dx$$

$$\text{Find } \frac{\partial S}{\partial \alpha_i} = 0 \Rightarrow \text{normal equations:}$$

$$\int_a^b p_i(x) \left(\sum_{j=0}^{m} \alpha_j p_j(x) \right) dx = \int_a^b p_i(x) f(x) dx \quad (1), \text{ expanding the sum,}$$

$$\alpha_0 \int_a^b p_i(x) p_0(x) dx + \alpha_1 \int_a^b p_i(x) p_1(x) dx + \ldots + \alpha_i \int_a^b p_i^2(x) dx + \ldots$$

$$+ \alpha_m \int_a^b p_i(x) p_m(x) dx = \int_a^b p_i(x) f(x) dx$$

Choosing the $p_i(x)$

- In general, choosing polynomials appropriately will ensure that this is not a Hilbert Matrix

- We can again use an LU solve to find the a_i and hence determine the fitting polynomial

- However, if we had a system of polynomials for which

$$\int_a^b p_i(x)p_j(x)dx = 0 \quad \text{if} \quad i \neq j \qquad (2)$$

  The system would be trivial to solve!
  - Such polynomials are said to be *orthogonal*

- Note that the integral is directly analogous to a dot product of vectors of rank N vectors (the integral has "infinite rank")!

$$\vec{P}^i.\vec{P}^j = \sum_{k=1}^{N} P_k^i P_k^j \text{ is analogous to } \int_a^b p_i(x)p_j(x)dx$$

Simplifying the normal equations

- If we have orthogonal polynomials, then from (1)

$$\int_a^b p_i(x)\left(\sum_{j=0}^{m}\alpha_j p_j(x)\right)dx = \sum_{j=0}^{m}\alpha_j \int_a^b p_i(x)p_j(x)dx$$

$$= \int_a^b p_i(x)f(x)dx$$

$$\Rightarrow \alpha_i \int_a^b p_i^2(x)dx = \int_a^b p_i(x)f(x)dx$$

$$\Rightarrow \alpha_i = \frac{\int_a^b p_i(x)f(x)dx}{\int_a^b p_i^2(x)dx} \qquad (3)$$

Further simplification

- In the case that the polynomials are orthonormal so that

$$\int_a^b p_i(x) p_i(x) dx = 1$$

- Then eqn (3) reduces to

$$\alpha_i = \int_a^b p_i(x) f(x) dx$$

- There is no need to solve a matrix in this case!

Orthogonality definition: weighting functions

- In practice it is often useful to include a weighting function, w(x), in the definition of the orthogonality requirement

$$\int_a^b p_i(x)p_j(x)w(x)dx = 0 \quad \text{if } i \neq j$$

- Orthonomality is then

$$\int_a^b p_i(x)p_j(x)w(x)dx = \delta_{ij}$$

- We can include the weighting function in the definition of the least squares formula for S

Examples of Orthogonal (Orthonormal) functions

$[a,b]$	$w(x)$	symbol	Name
[-1,1]	1	$P_n(x)$	Legendre polynomials
[-1,1]	$(1-x^2)^{-1/2}$	$T_n(x)$	Chebyshev I
[-1,1]	$(1-x^2)^{-1/2}$	$U_n(x)$	Chebyshev II
$[0,\infty)$	$\exp(-x)$	$L_n(x)$	Laguerre
$(-\infty,\infty)$	$\exp(-x^2)$	$H_n(x)$	Hermite

Levenberg-Marquardt Method

Summary

- Least squares fitting requires setting up the normal equations to minimize the sum of the squares
- Fitting functions with a simple polynomial will results in ill conditioned Hilbert matrices
- More general fitting using a sum over orthogonal/orthonormal polynomials can reduce the fitting problem to a series of integrals

Homework 9：05/08/2019

Problem 1: Consider the following four sets of data

Data 1		Data 2		Data 3		Data 4	
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.10	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.10	4	5.39	19	12.50
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

1) Please fit to straight line for each of the four sets of data and discuss it is good or bad fit?
2) Plot these fits

Problem 2: The life date of the Decay of a radioactive substance is

i	1	2	3	4	1	6	7	8	9	10	11	12
t_i	0	15	30	45	60	75	90	105	120	135	150	165
N_i	106	80	98	75	74	73	49	38	37	22	20	19

1)Find the values A and t taking into account the uncertainties in the data points.

2)Plot these fitting results

Next Lecture

- Introduction to numerical integration