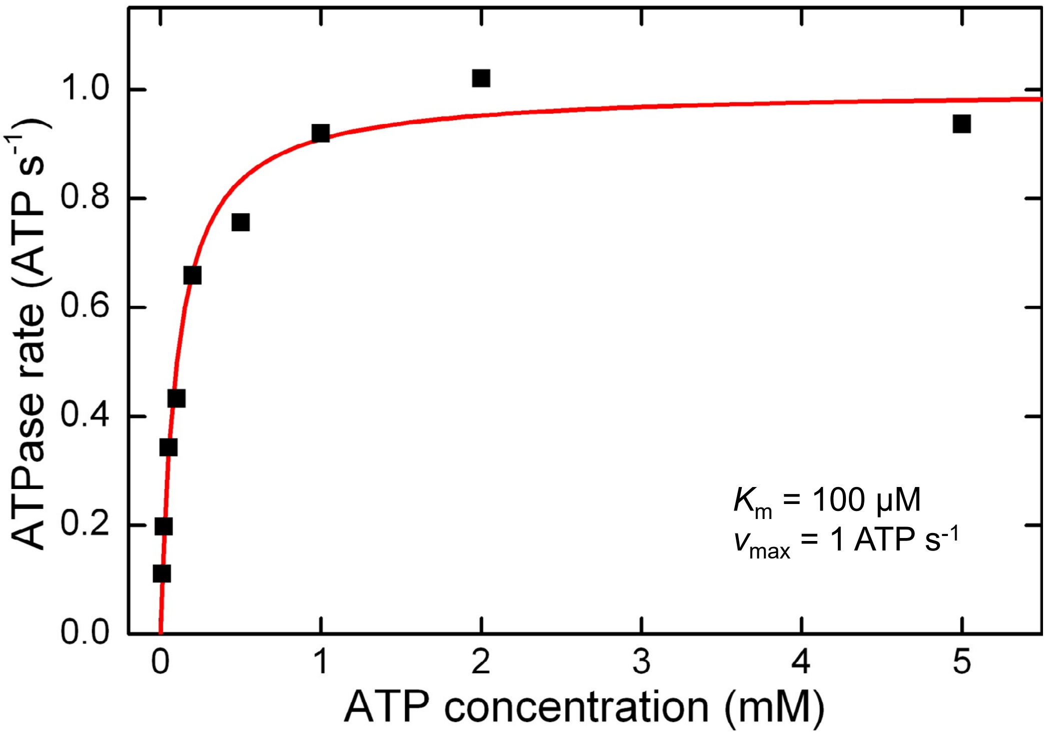
Curve fitting – Least squares

Curve fitting



Excurse error of multiple measurements

Starting point: Measure *N* times the same parameter
Obtained values are Gaussian distributed around mean with standard deviation σ

What is the error of the mean of all measurements?

Sum = $x_1 + x_2 + \ldots + x_N$

Variance of sum $= N * \sigma^2$ **(Central limit theorem)**

Standard deviation of sum $= \sqrt{N} \cdot \sigma$

Mean = $(x_1 + x_2 + \ldots + x_N)/N$

Standard deviation of mean (called **standard error of mean**) $= \dfrac{\sqrt{N} \cdot \sigma}{N} = \dfrac{\sigma}{\sqrt{N}}$

$$\sigma_N = \frac{\sigma_1}{\sqrt{N}}$$

Excurse error propagation

What is error for *f*(*x,y,z,…*) if we know errors of *x,y,z,…* (σ_x, σ_y, σ_z, …) for purely statistical errors?

-> Individual variances add scaled by squared partial derivatives (if parameters are uncorrelated)

$$V(f) = \sigma_f^2 = \left(\frac{df}{dx}\right)^2 \sigma_x^2 + \left(\frac{df}{dy}\right)^2 \sigma_y^2 + \left(\frac{df}{dz}\right)^2 \sigma_z^2$$

Examples :

Addition/substraction:

$$\begin{aligned} f &= x \pm y \\ V(f) &= V(x) + V(y) \quad \text{or} \\ \sigma_f^2 &= \sigma_x^2 + \sigma_y^2 \end{aligned}$$

squared errors add up

Product:

$$\begin{aligned} f &= xy \\ V(f) &= y^2 V(x) + x^2 V(y) \quad \text{or} \\ \left(\frac{\sigma_f}{f}\right)^2 &= \left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2 \end{aligned}$$

squared relative errors add up

Excurse error propagation

Ratios:

$$f = x/y$$
$$V(f) = \left(\frac{1}{y}\right)^2 V(x) + \left(-\frac{x}{y^2}\right)^2 V(y) \quad \text{or}$$
$$\left(\frac{\sigma_f}{f}\right)^2 = \left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2$$

squared relative errors add up

Powers:

$$f = x^2$$
$$V(f) = (2x)^2 V(x)$$
$$\sigma_f = 2x\sigma_x \quad \text{or}$$
$$\frac{\sigma_f}{f} = 2\frac{\sigma_x}{x}$$

relative error times power

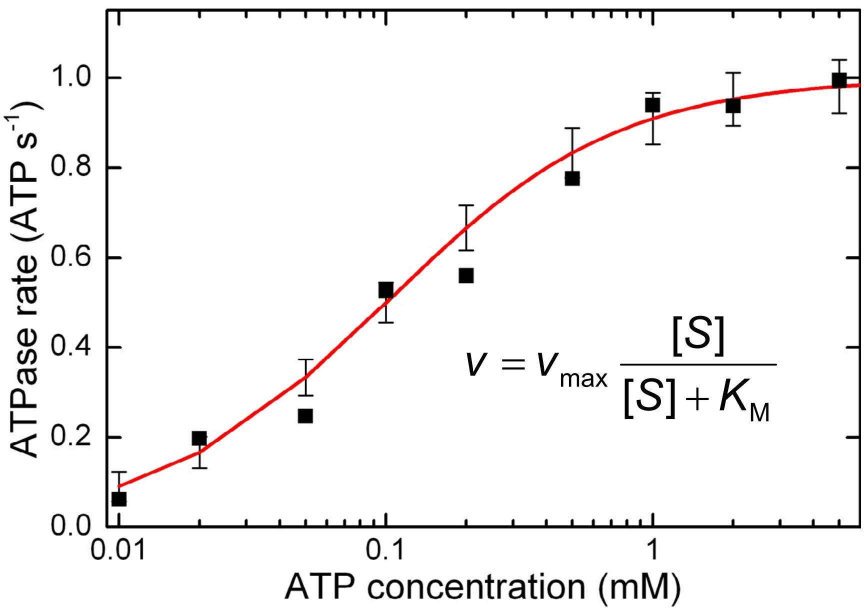
Logarithms:

$$f = \ln x$$
$$V(f) = \left(\frac{1}{x}\right)^2 V(x)$$
$$\sigma_f = \frac{\sigma_x}{x}$$

error is relative error

Curve fitting – Least squares

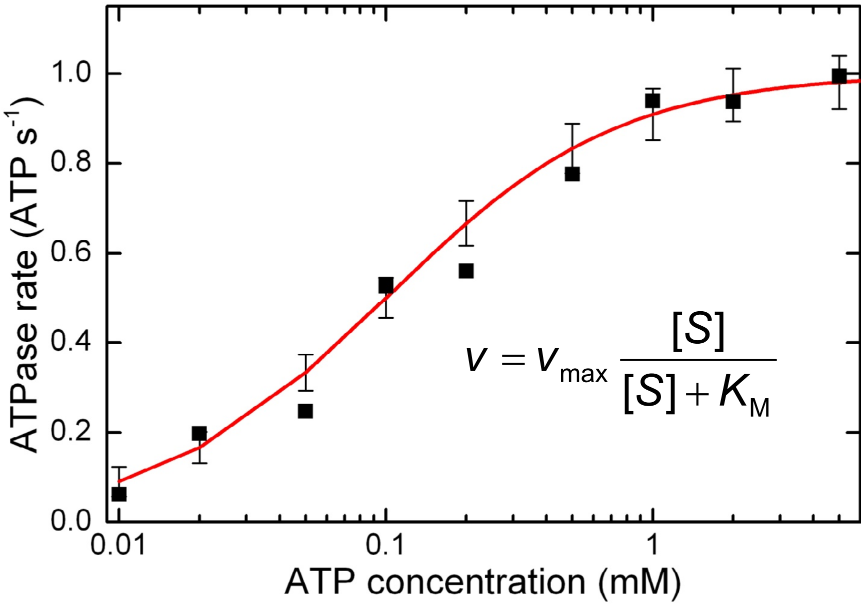
Starting point: - data set with N pairs of (x_i, y_i)
- x_i known exactly,
- y_i Gaussian distributed around true value with error σ_i
- errors uncorrelated
- function $f(x)$ which shall describe the values y ($y = f(x)$)
- $f(x)$ depends on one or more parameters a



Curve fitting – Least squares

Probability to get y_i for given x_i

$$P(y_i;a)=\frac{1}{\sqrt{2\pi}\sigma_i}e^{-[y_i-f(x_i;a)]^2/2\sigma_i^2]}$$

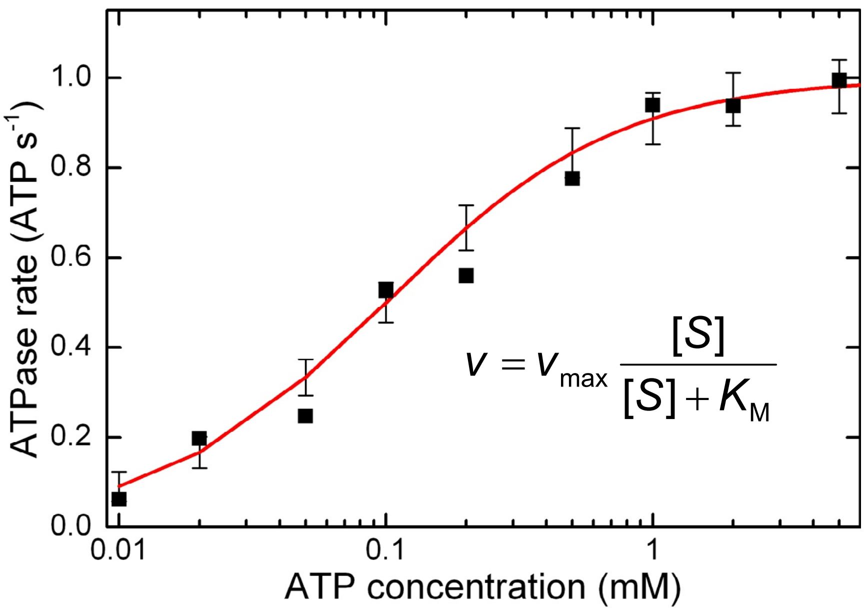


Curve fitting – Least squares

Prob. to get whole set y_i for set of x_i

$$P(y_1,..,y_N;a)=\prod_{i=1}^{N}\frac{1}{\sqrt{2\pi}\sigma_i}e^{-[y_i-f(x_i;a)]^2/2\sigma_i^2]}$$

For best fitting theory curve (red curve) $P(y_1,..y_N;a)$ becomes maximum!



Curve fitting – Least squares

Prob. to get whole set y_i for set of x_i $P(y_1,..,y_N;a) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_i} e^{-[y_i - f(x_i;a)]^2/2\sigma_i^2]}$

For best fitting theory curve (red curve) $P(y_1,..y_N;a)$ becomes maximum!

Use logarithm of product, get a sum and maximize sum:

$$\ln[P(y_1,..,y_N;a)] = -\frac{1}{2}\sum_1^N \left[\frac{y_i - f(x_i;a)}{\sigma_i}\right]^2 - \sum_1^N \ln\left(\sigma_i\sqrt{2\pi}\right)$$

OR minimize χ^2 with:

$$\chi^2 = \sum_{i=1}^{N} \left[\frac{y_i - f(x_i;a)}{\sigma_i}\right]^2$$

Principle of least squares!!!

Curve fitting – Least squares

$$\chi^2 = \sum_{i=1}^{N} \left[\frac{y_i - f(x_i;a)}{\sigma_i}\right]^2$$

Principle of least squares!!!
(X^2 minimization)

Solve:

$$\frac{d\chi^2}{da} = 0$$

$$\sum_i \frac{1}{\sigma_i^2} \frac{df(x_i;a)}{da} [y_i - f(x_i;a)] = 0$$

Solve equation(s) either analytically (only simple functions)
or numerically (specialized software, different algorithms)
χ^2 value indicates goodness of fit

Errors available: **USE THEM!** → so called **weighted fit**
Errors not available: σ_i's are set as constant → conventional fit

Reduced χ²

Expectation value of **χ^2 for weighted fit:**

$$\left\langle \chi^2 \right\rangle = \sum_{i=1}^{N} \frac{\left\langle (y_i - f(x_i;a))^2 \right\rangle}{\sigma_i^2} = \sum_{i=1}^{N} \frac{\sigma_i^2}{\sigma_i^2} = N$$

Define reduced χ^2:

$$\chi^2_{red} = \frac{\chi^2_{red}}{N - M} \qquad \text{with} \qquad \left\langle \chi^2_{red} \right\rangle = 1$$

number of fit paramters

For weighted fit the reduced χ^2 should become 1, if errors are properly chosen

Expectation value of **χ^2 for unweighted fit:**

$$\left\langle \chi^2 \right\rangle = \frac{1}{N-M} \sum_{i=1}^{N} \left\langle (y_i - f(x_i;a))^2 \right\rangle = \frac{1}{N-M} \sum_{i=1}^{N} \sigma^2 \approx \sigma^2$$

Should approach the variance of a single data point

Simple proportion

$$f(y) = mx.$$

$$\chi^2 = \sum_i \frac{(y_i - mx_i)^2}{\sigma_i^2}$$

Differentiation: $$\frac{d\chi^2}{dm} = \sum_i -2x_i \frac{y_i - mx_i}{\sigma_i^2}$$

For σ_i = const = σ $$= -\frac{2}{\sigma^2} \sum_i x_i y_i - m x_i^2$$

Solve: $$\sum_i (x_i y_i - \widehat{m} x_i^2) = 0$$

Get: $$\widehat{m} = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\overline{xy}}{\overline{x^2}}$$

Simple proportion - Errors

$$\widehat{m} = \frac{\sum x_i y_i}{\sum x_i^2}$$

Rewrite: $$\widehat{m} = \sum_i \frac{x_i}{N\overline{x^2}} y_i$$

-> Error of m given by errors of y_i
-> Use rules for error propagation

$$V(\widehat{m}) = \sum \left(\frac{x_i}{N\overline{x^2}}\right)^2 \sigma^2 = \frac{\sigma^2}{N\overline{x^2}}$$

-> Standard error of determination of m (single confidence interval) is then square-root of $V(m)$

-> σ^2 is estimated as mean square deviation of fitted function from the *y*-values if now errors are available

$$\sigma^2 = \frac{1}{N}\sum_1^N [y_i - f(x_i;a)]^2$$

Straight line fit (2 parameters)

$$f(x;m,c) = mx + c$$

$$\chi^2 = \sum_i (y_i - mx_i - c)^2$$ (or σ_i = const = σ)

Differentiation w.r.t *c*: $-2\sum_i (y_i - \widehat{m}x_i - \widehat{c}) = 0$ or $\overline{y} - \widehat{m}\overline{x} - \widehat{c} = 0$

Differentiation w.r.t *m*: $-2\sum_i x_i(y_i - \widehat{m}x_i - \widehat{c}) = 0$ or $\overline{xy} - \widehat{m}\overline{x^2} - \widehat{c}\overline{x} = 0$

Get:
(solve eqn. array)

$$\widehat{m} = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2} - (\overline{x})^2} \qquad \widehat{c} = \frac{\overline{x^2}\overline{y} - \overline{x}\,\overline{xy}}{\overline{x^2} - (\overline{x})^2} = \overline{y} - \widehat{m}\overline{x}$$

Errors:

$$V(\widehat{m}) = \frac{\sigma^2}{N(\overline{x^2} - (\overline{x})^2)} \qquad V(\widehat{c}) = \frac{\sigma^2\overline{x^2}}{N(\overline{x^2} - (\overline{x})^2)}$$

For all relations which are linear with respect to the fit parameters, analytical solutions possible!

Straight line fit (2 parameters)

Additional quantities for multiple parameters: **Covariances**
-> describes interdependency/correlation between the obtained parameters

$$\begin{aligned} cov(x_{(i)}, x_{(j)}) &= \overline{(x_{(i)} - \overline{x_{(i)}})(x_{(j)} - \overline{x_{(j)}})} \\ &= \overline{x_{(i)} x_{(j)}} - \overline{x_{(i)}}\,\overline{x_{(j)}} \end{aligned}$$

Covariance matrix $$V_{ij} = cov(x_{(i)}, x_{(j)})$$ $V_{ii} = Var(x_{(i)})$

$$\mathrm{cov}(p_i, p_j) = \begin{pmatrix} V_{p1p1} & V_{p1p2} & V_{p1p3} \\ V_{p1p2} & V_{p2p2} & V_{p2p3} \\ V_{p1p3} & V_{p2p3} & V_{p3p3} \end{pmatrix} = \begin{pmatrix} Var(p_1) & V_{p1p2} & V_{p1p3} \\ V_{p1p2} & Var(p_2) & V_{p2p3} \\ V_{p1p3} & V_{p2p3} & Var(p_3) \end{pmatrix}$$

Squared errors of fit parameters!

Straight line fit: $$cov(\widehat{m}, \widehat{c}) = \frac{\sigma^2 \overline{x}}{N(\overline{x^2} - (\overline{x})^2)}$$

More in depth

Online lecture: Statistical Methods of Data Analysis by Ian C. Brock

http://www-zeus.physik.uni-bonn.de/~brock/teaching/stat_ws0001/

Numerical recipes in C++, Cambridge university press