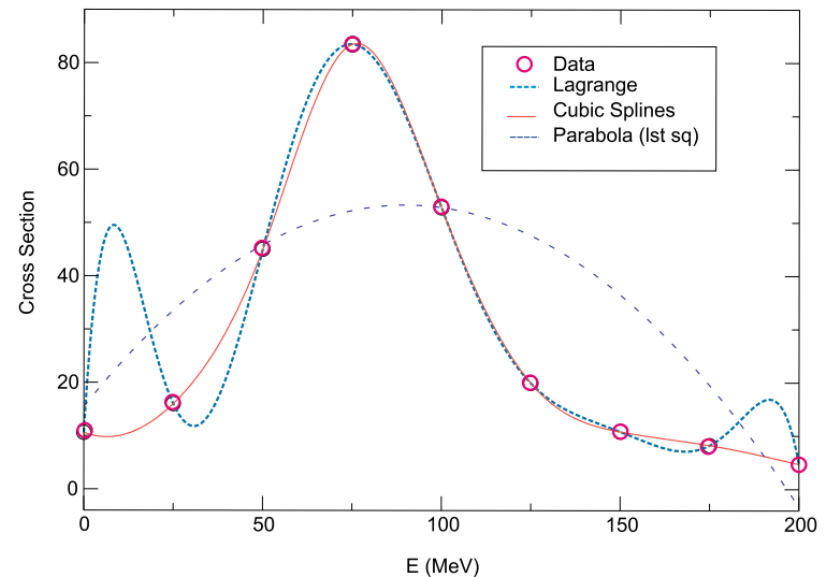


Interpolation vs regression

- Interpolation: find a continuous function that passes through all the data points (“join the dots”)
 - Typical use: change sampling, resolution
 - Formula of the fitting function is not interesting
- Fitting, regression: find the best parameters of a known model function, that is closest to data points
 - Typical use: find parameters of a physical model
 - Sometimes the model is just approximation



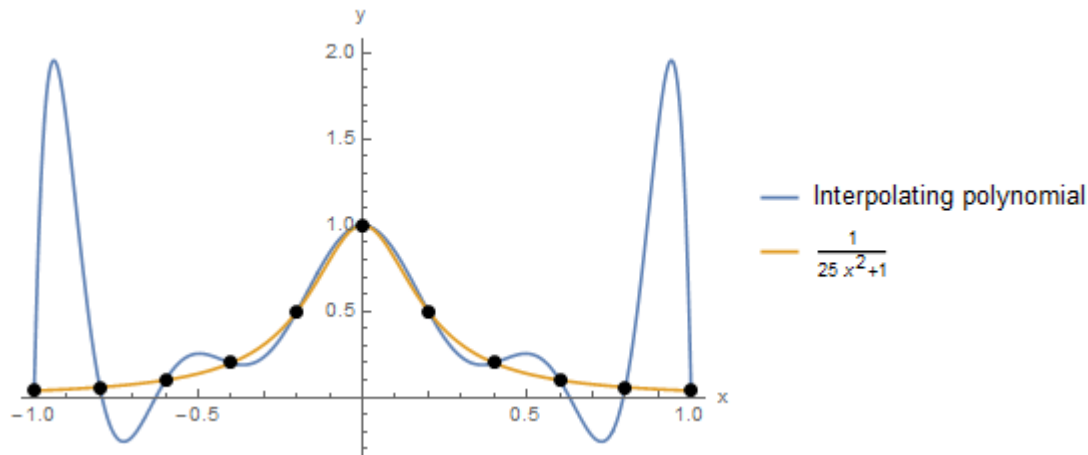
Interpolation

- Lagrange's formula
- Polynomial goes through **all** points

Through any two points there is a unique line. Through any three points, a unique quadratic. Et cetera. The interpolating polynomial of degree $N - 1$ through the N points $y_1 = f(x_1), y_2 = f(x_2), \dots, y_N = f(x_N)$ is given explicitly by Lagrange's classical formula,

$$P(x) = \frac{(x - x_2)(x - x_3) \dots (x - x_N)}{(x_1 - x_2)(x_1 - x_3) \dots (x_1 - x_N)} y_1 + \frac{(x - x_1)(x - x_3) \dots (x - x_N)}{(x_2 - x_1)(x_2 - x_3) \dots (x_2 - x_N)} y_2 \\ + \dots + \frac{(x - x_1)(x - x_2) \dots (x - x_{N-1})}{(x_N - x_1)(x_N - x_2) \dots (x_N - x_{N-1})} y_N \quad (3.1.1)$$

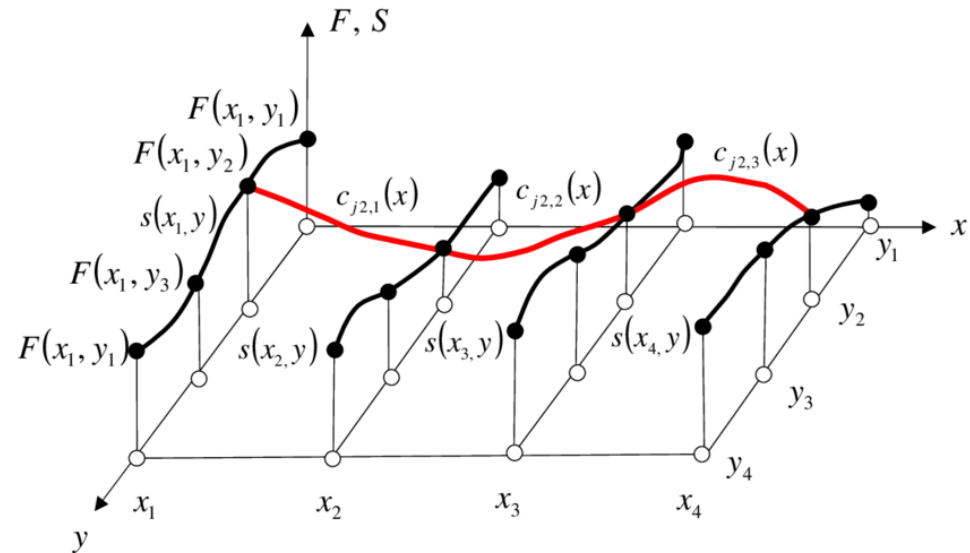
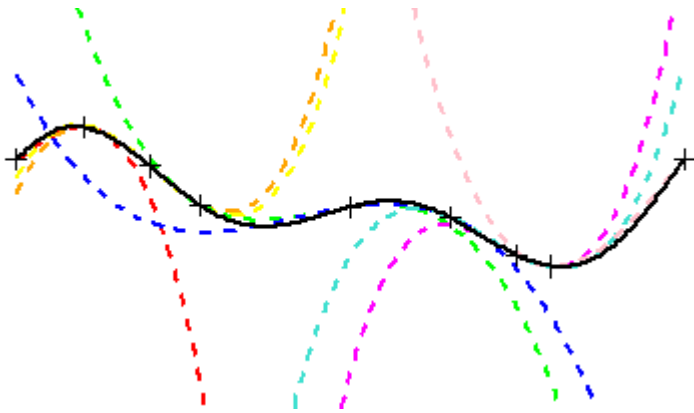
There are N terms, each a polynomial of degree $N - 1$ and each constructed to be zero at all of the x_i except one, at which it is constructed to be y_i .



Interpolation

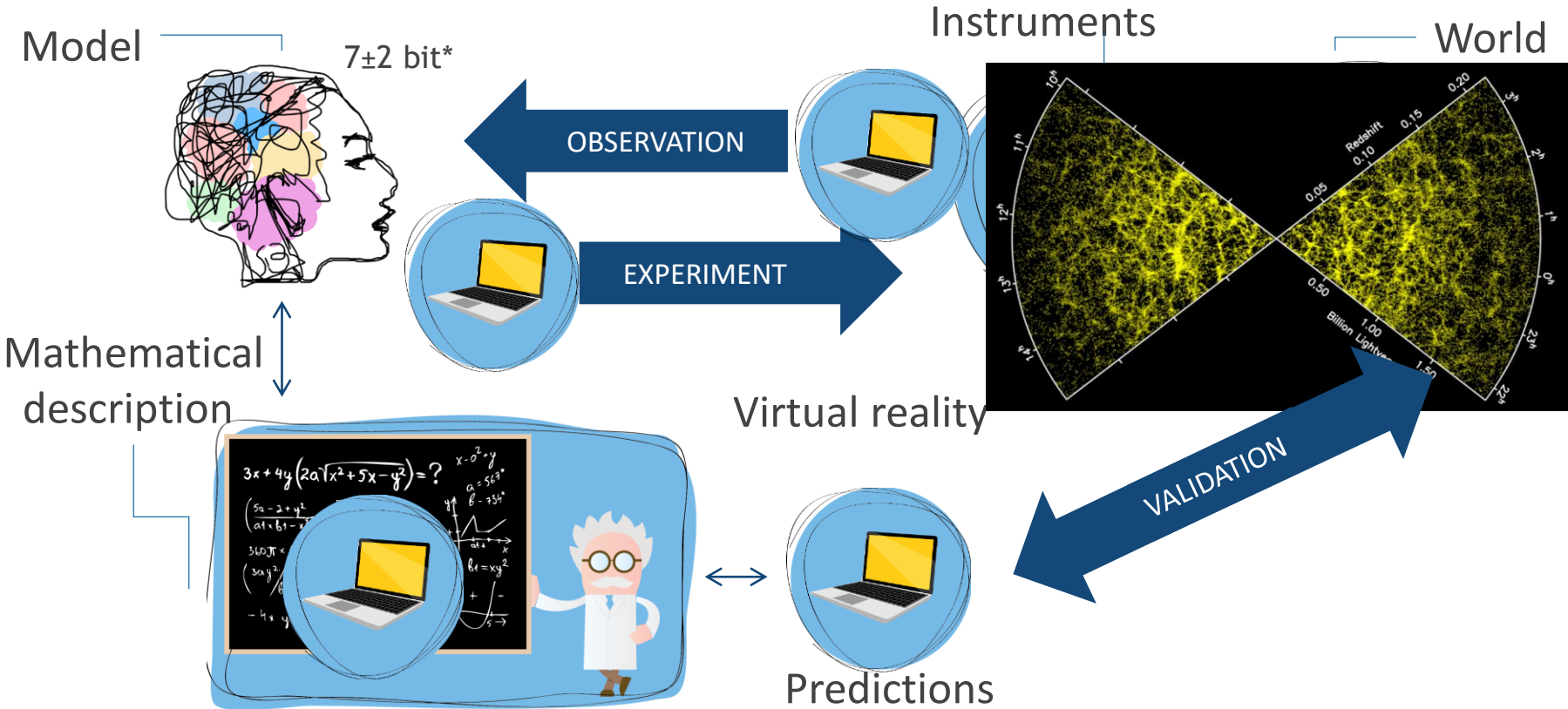
- Cubic spline:

- Local: low order polynomial through **neighboring** points
- Local interpolators join smoothly to each other
- Bi-cubic spline for 2D



Least Squares and Maximum Likelihood

Modern data science



Initial values

$\Lambda=0.7$
 $\Omega_m=0.3$

“laws”, equations

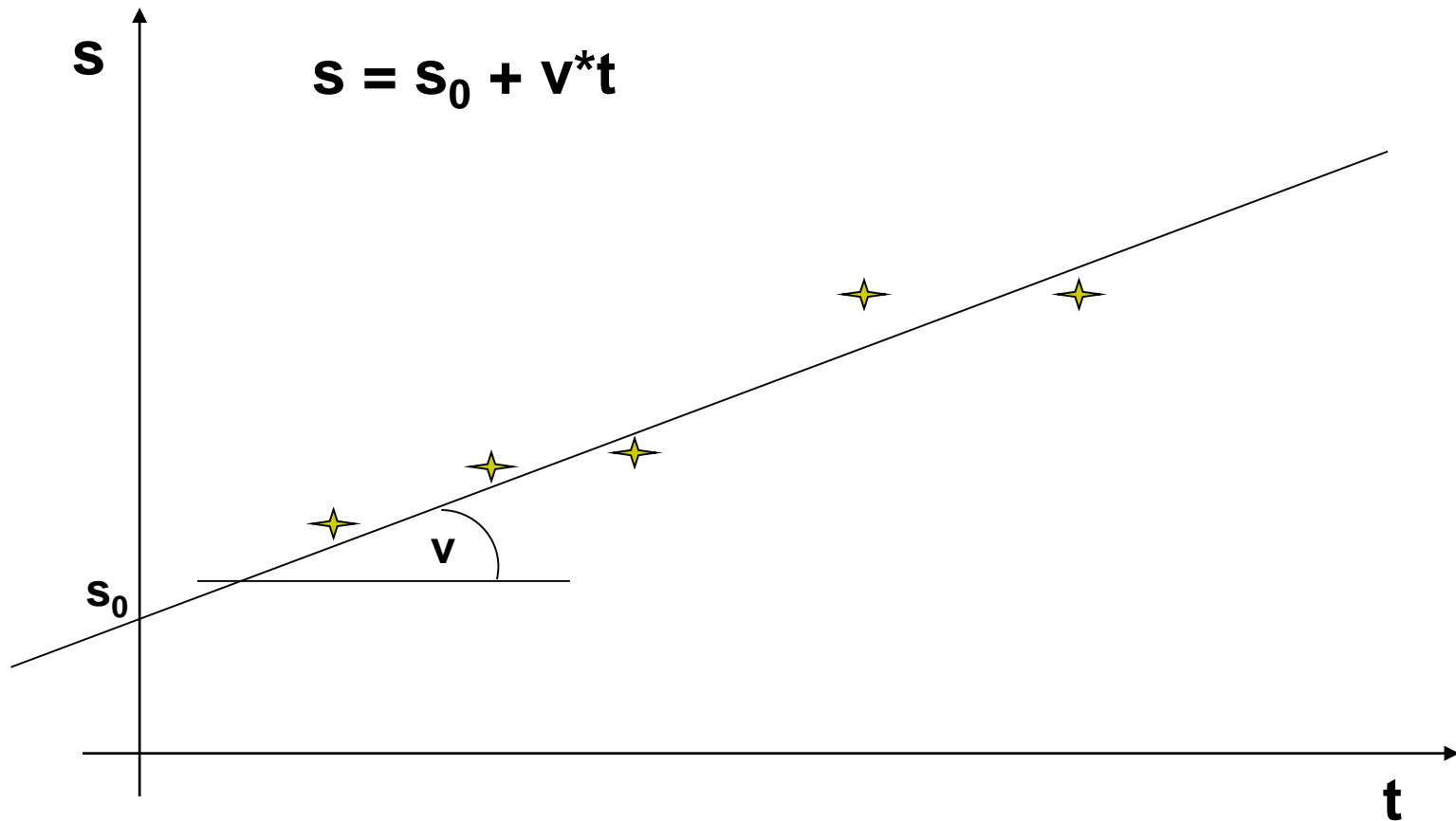
$$F = G \frac{m_1 m_2}{r^2}$$
$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}$$

Simulated reality

Goal

- Find a MODEL for measurement/simulation data, that:
 - Fits some theory
 - Reproduce measurement data
 - Can be parametrized
- Parameters may have „physical” meaning
 - Displacement/time diagram – linear fit – slope: speed

Example: linear motion: speed



Errors

- Measurements contain **errors** (in each variables!)
 - noise, inaccuracy
 - Systematic errors
 - outliers
- Hence: the calculated parameters will have errors, too

Least square fit

Given: $(x_i, y_i) ; i=1 \dots N$ measurement points. Fit some function :

$$y(x) = y(x; a_1, a_2, \dots, a_M)$$

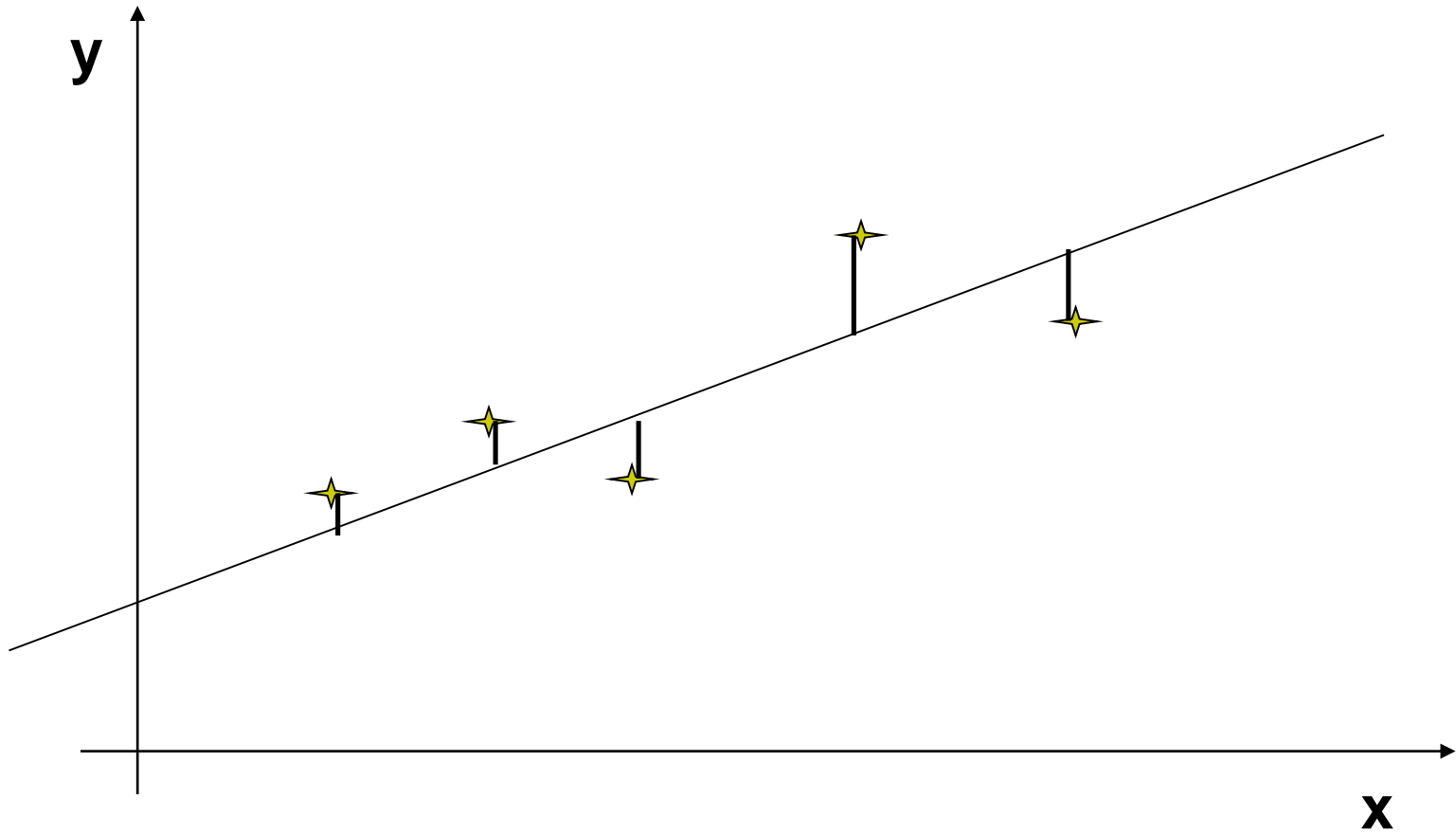
a_j parameters, $j=1 \dots M$

Least square fit:

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

Take derivatives by $a_1, a_2, \dots, a_M \rightarrow$ set of linear equations (!! Also for some nonlin func.)

Least squares



Why not least absolute values, fourth powers, etc?

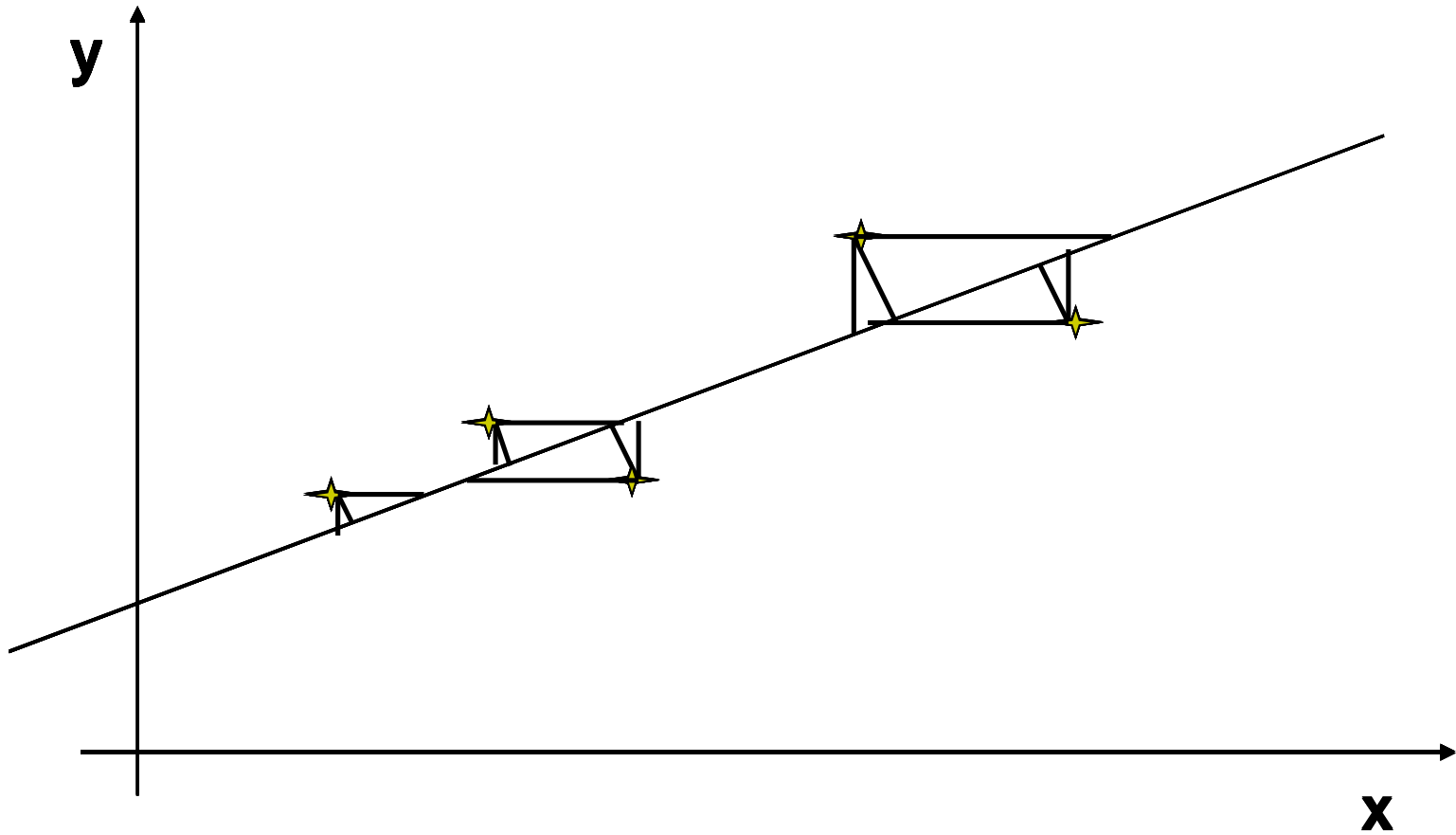
$$\min_{a_1 \dots a_M} \left(\sum_{i=1}^N [y_i - y(x_i; a_1 \dots a_M)]^2 \right)$$

$$\min_{a_1 \dots a_M} \left(\sum_{i=1}^N |y_i - y(x_i; a_1 \dots a_M)| \right)$$

$$\min_{a_1 \dots a_M} \left(\sum_{i=1}^N [y_i - y(x_i; a_1 \dots a_M)]^4 \right)$$

$$\min_{a_1 \dots a_M} \left(\sum_{i=1}^N \sqrt[8]{|y_i - y(x_i; a_1 \dots a_M)|} \right)$$

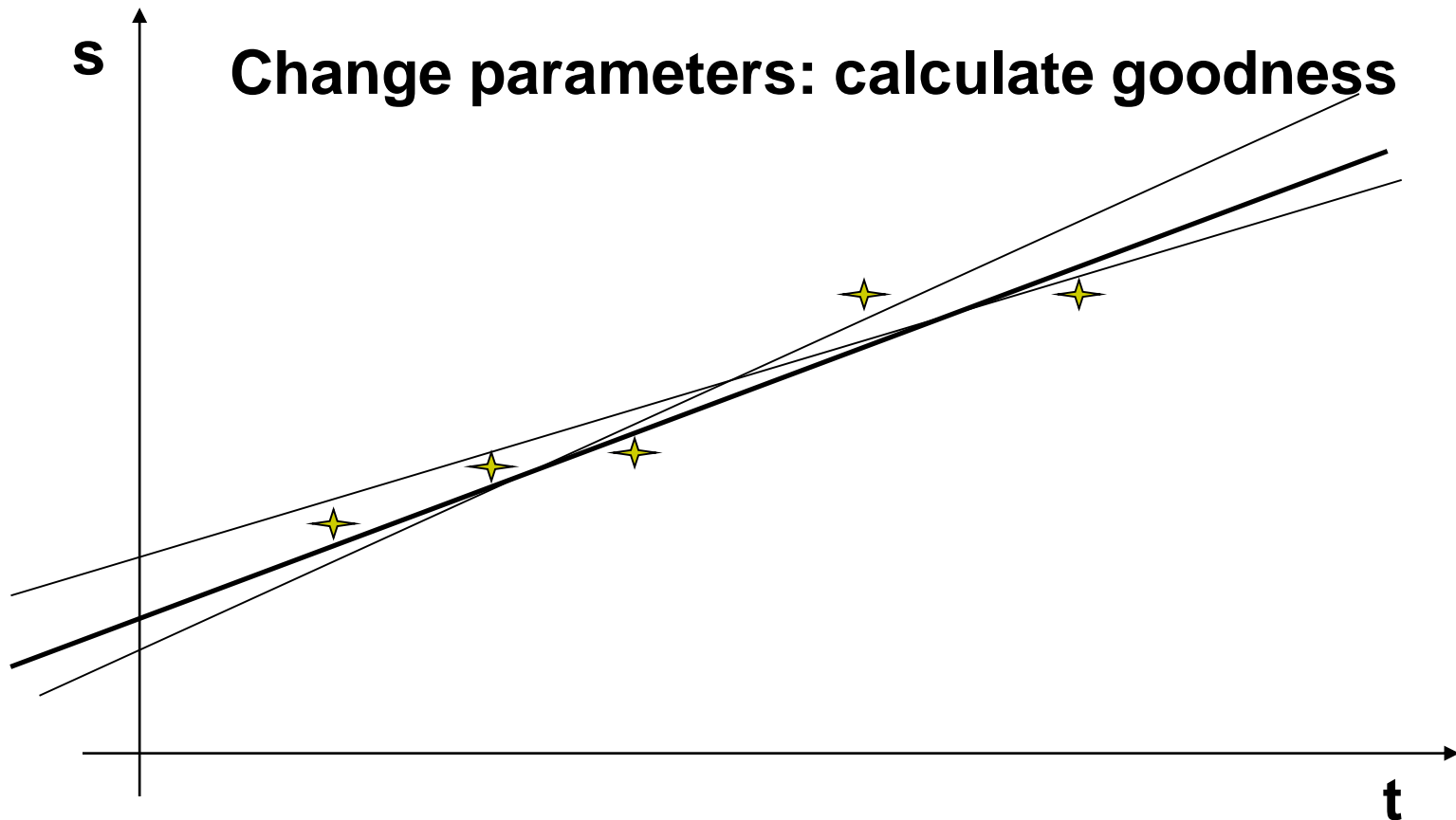
Least distances?



Why least squares?

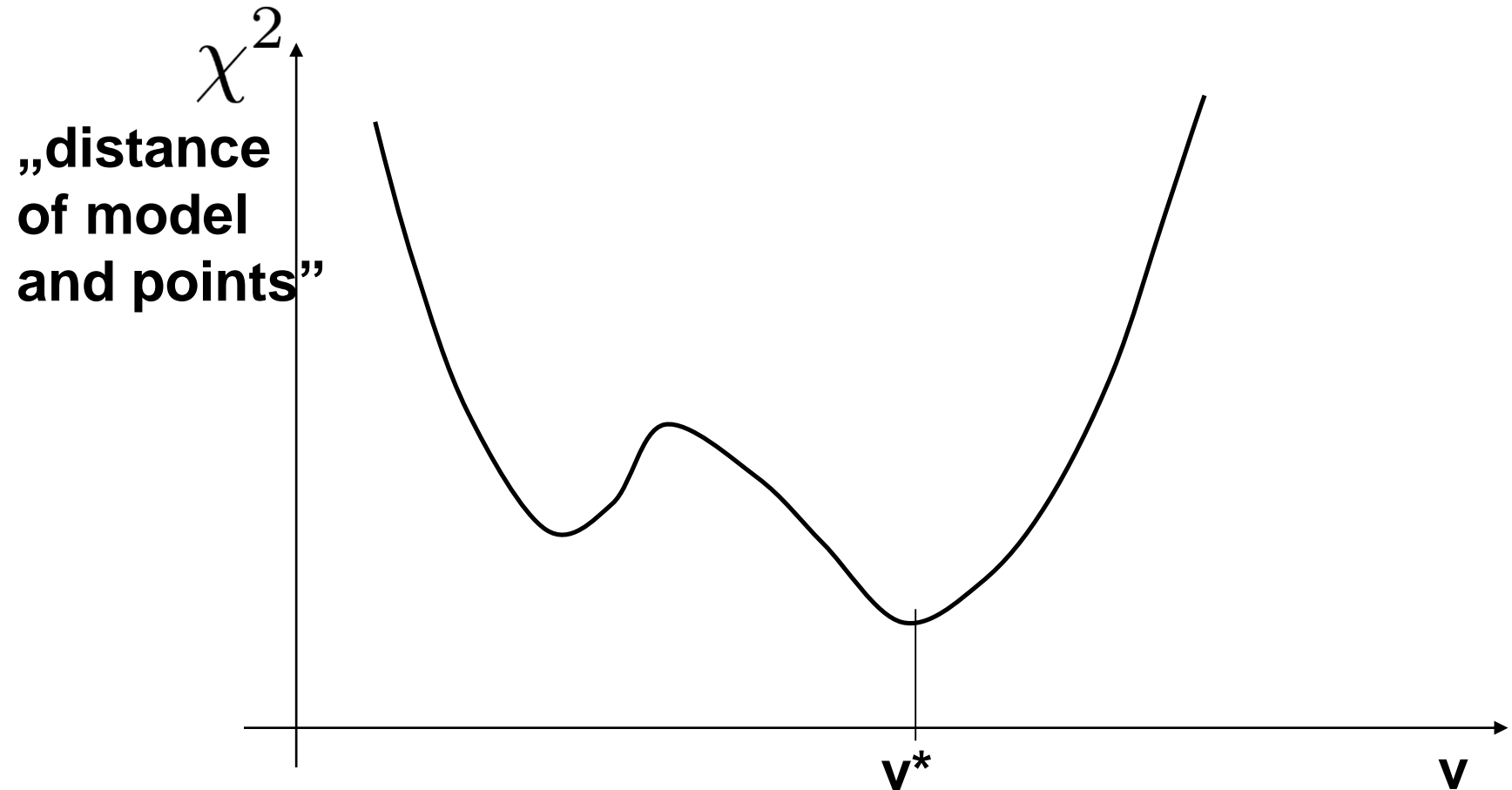
- Maximum likelihood estimation
- Least square gives best parameters, if :
 - Errors only in y_i
 - Errors are normal (Gauss) distributed
 - Errors are independent

Cost function



**Find optimal parameters,
with largest goodness / smallest „cost/energy”**

Cost function



Finding minimum: in general leads to a nonlinear optimization problem

In special case:
linear set of
equations

Maximum likelihood

$$P \propto \prod_{i=1}^N \left\{ \exp \left[-\frac{1}{2} \left(\frac{y_i - y(x_i)}{\sigma} \right)^2 \right] \Delta y \right\}$$

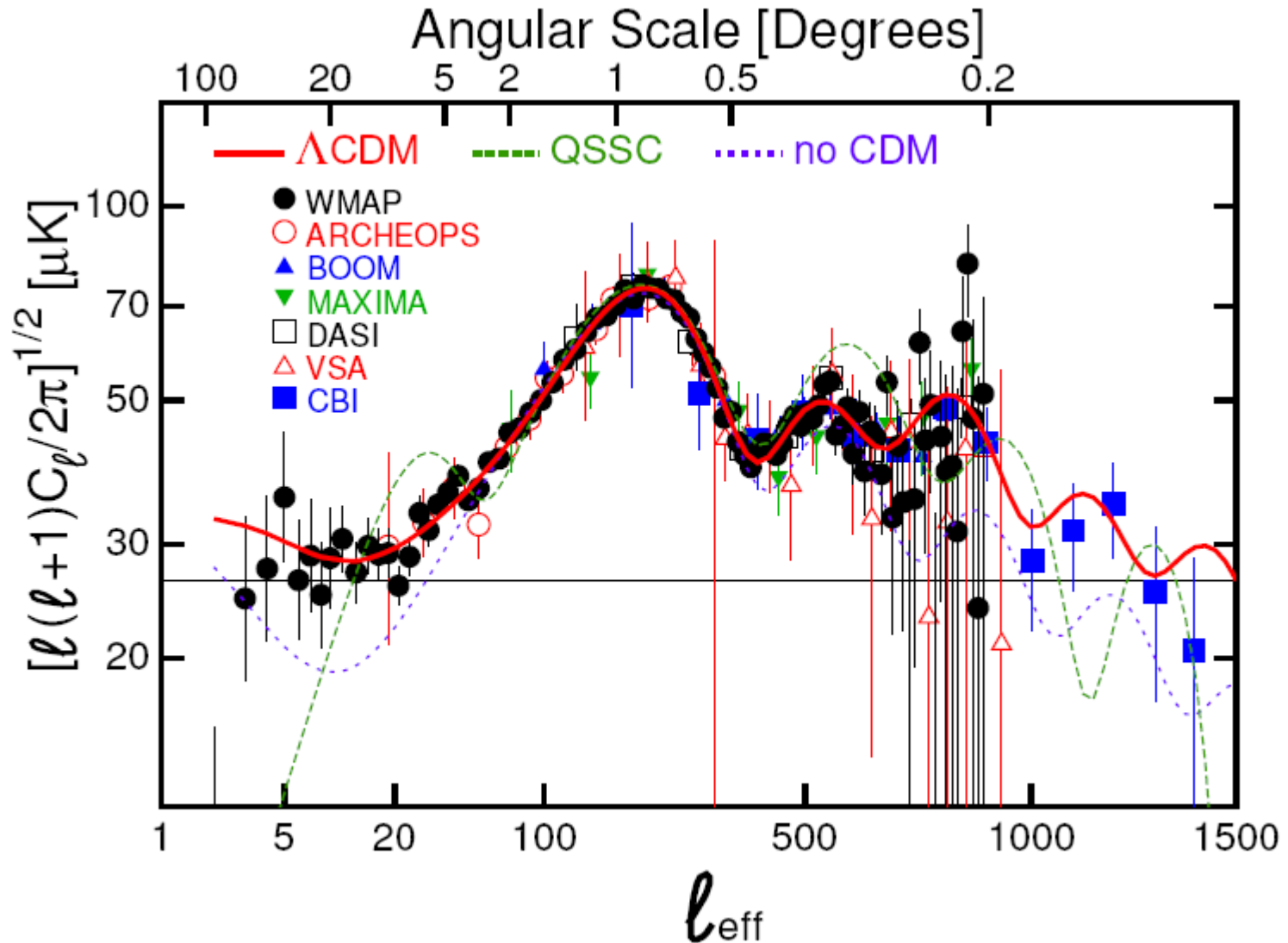
Find parameters where P is maximal.

The $\log()$ func. is monotonuious, so find maximum of $\log(P)$,
or minimum of $-\log(P)$

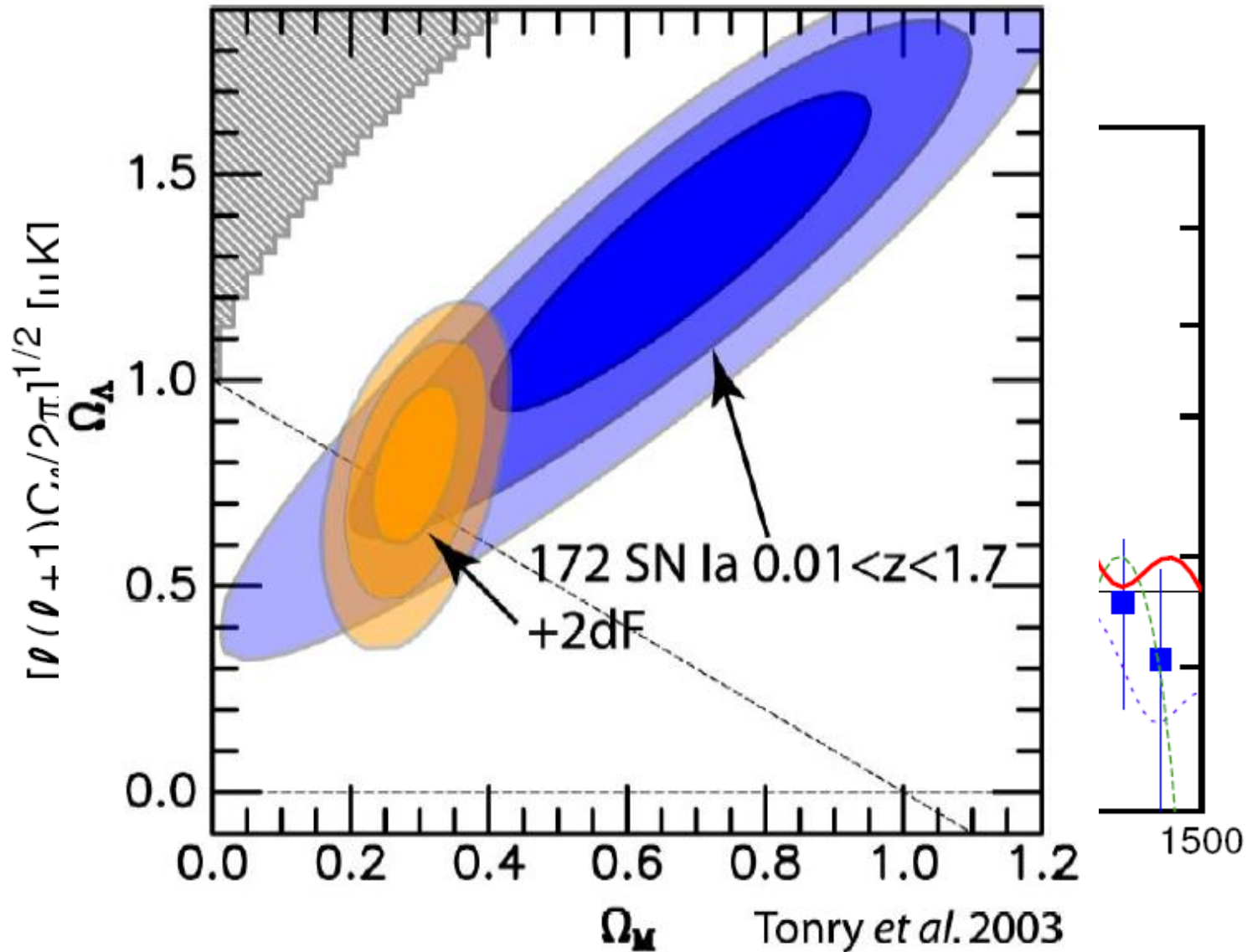
$$-\log(P) = \left[\sum_{i=1}^N \frac{[y_i - y(x_i)]^2}{2\sigma^2} \right] - N \log \Delta y$$

This is Least Squares!

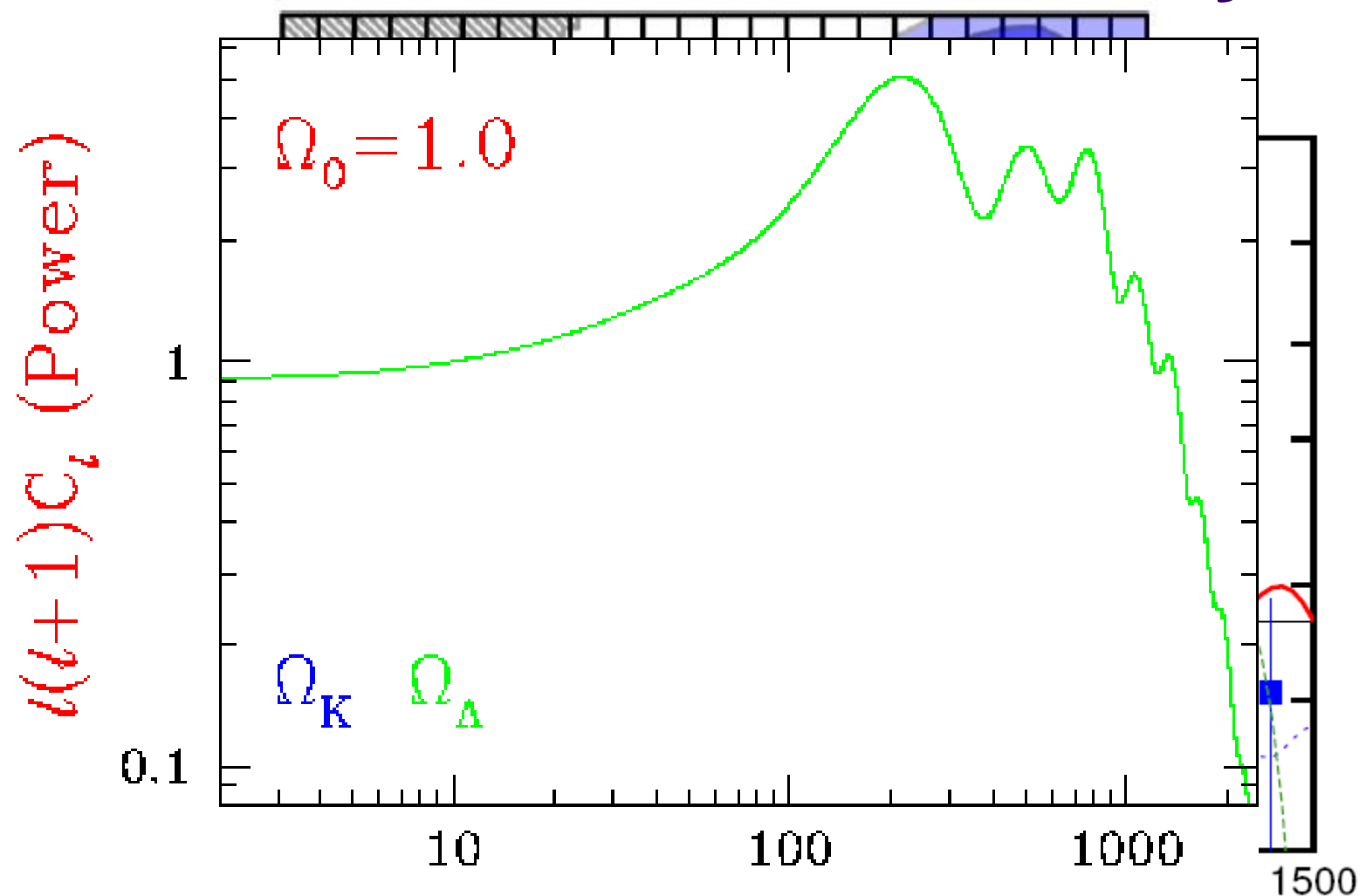
Complex example: cosmology



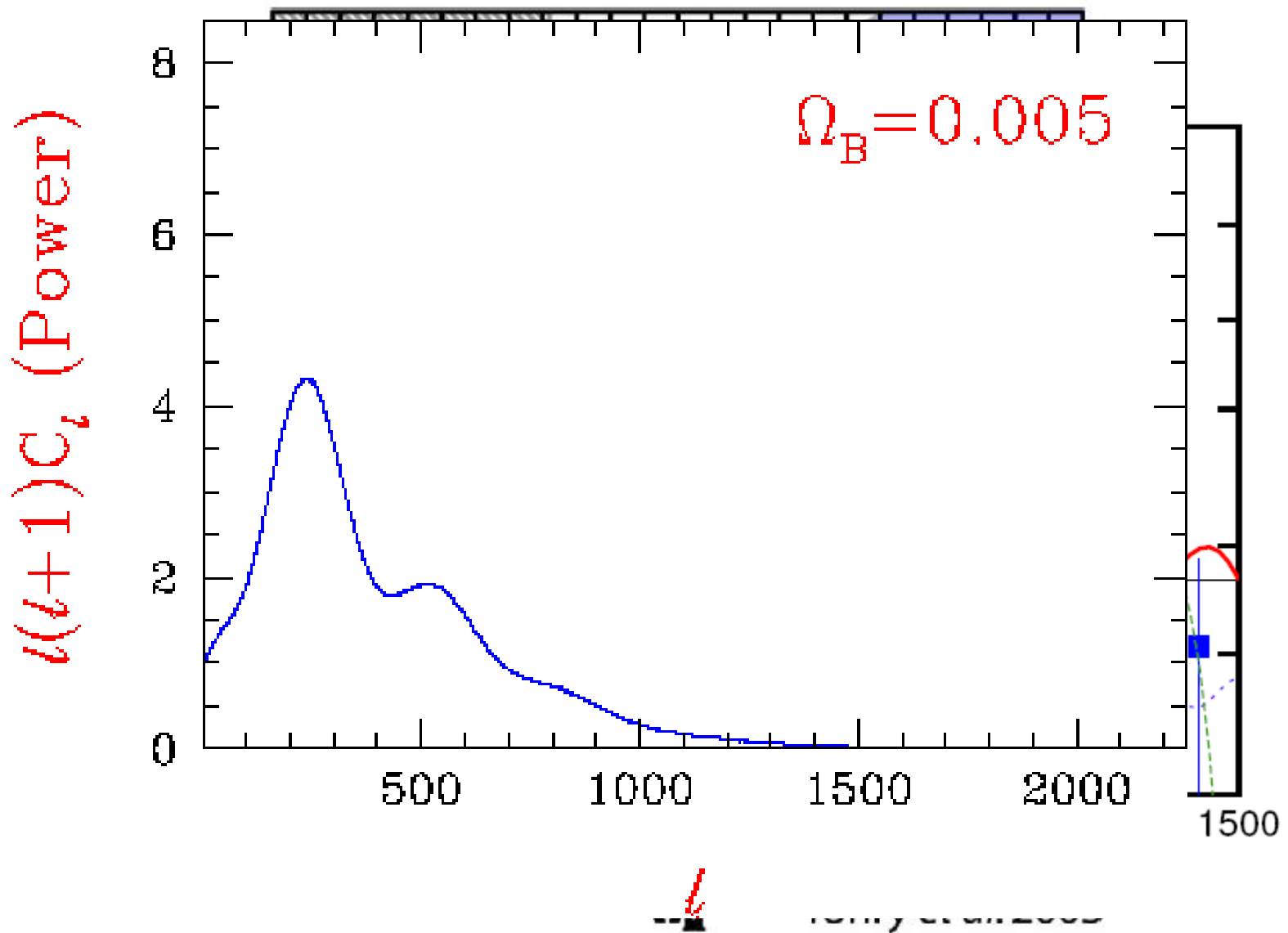
Complex example: cosmology



Complex example: cosmology



Complex example: cosmology



Complex example: cosmology

