Adatmodellezés, függvényillesztés

A fizika numerikus módszerei I.

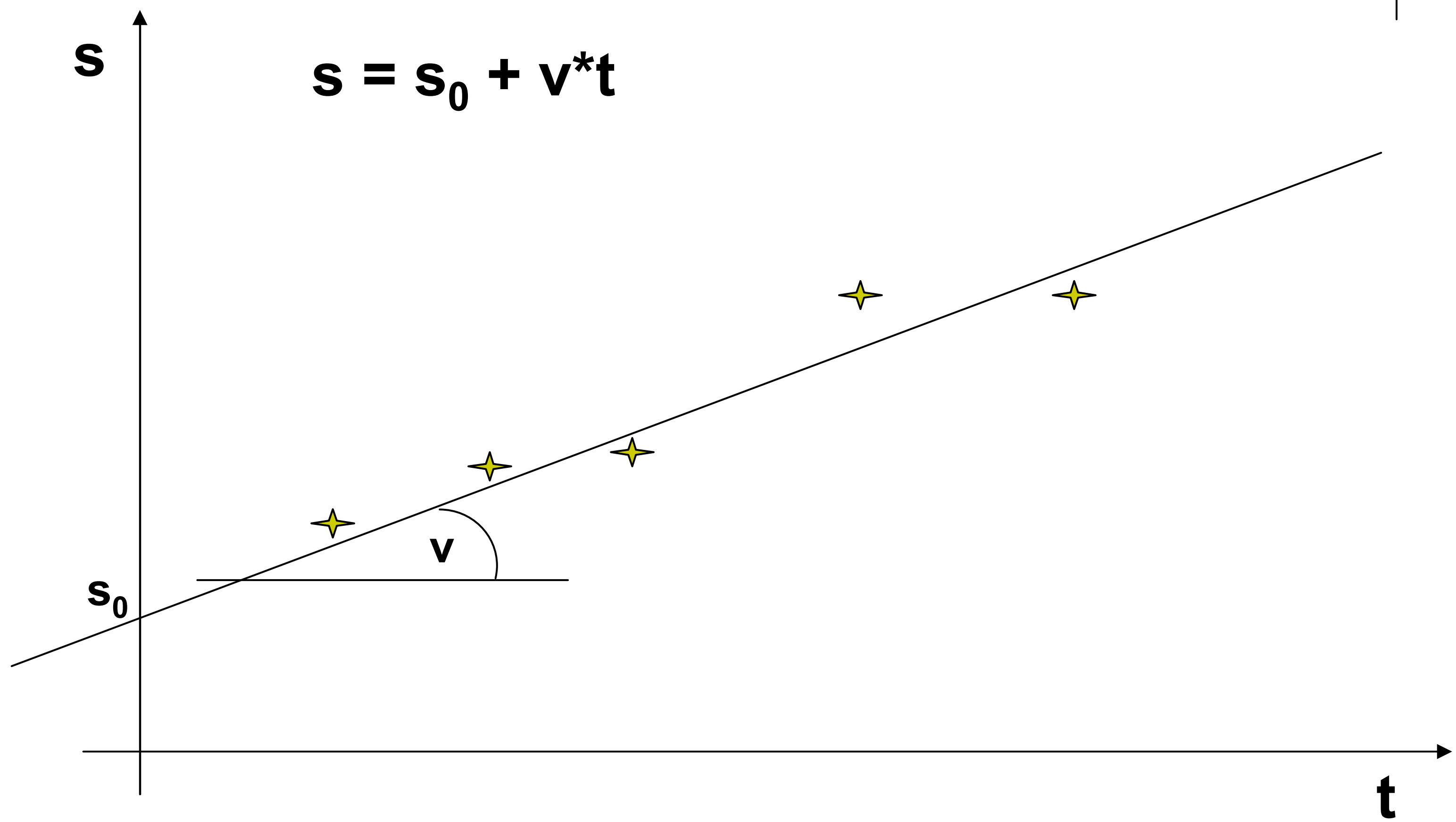
mf1n1a06- mf1n2a06

Csabai István

Bevezetés

- Mérési/szimulációs adatokhoz keresünk egy modellt, függvényt:
  - Összhangban van az elmélettel
  - Jól visszaadja a mérési eredményeket (jóslás)
  - Paraméterezhető
- A paramétereknek gyakran fizikai jelentősége van
  - út/idő diagram – egyenesillesztés - sebesség

Példa: sebesség

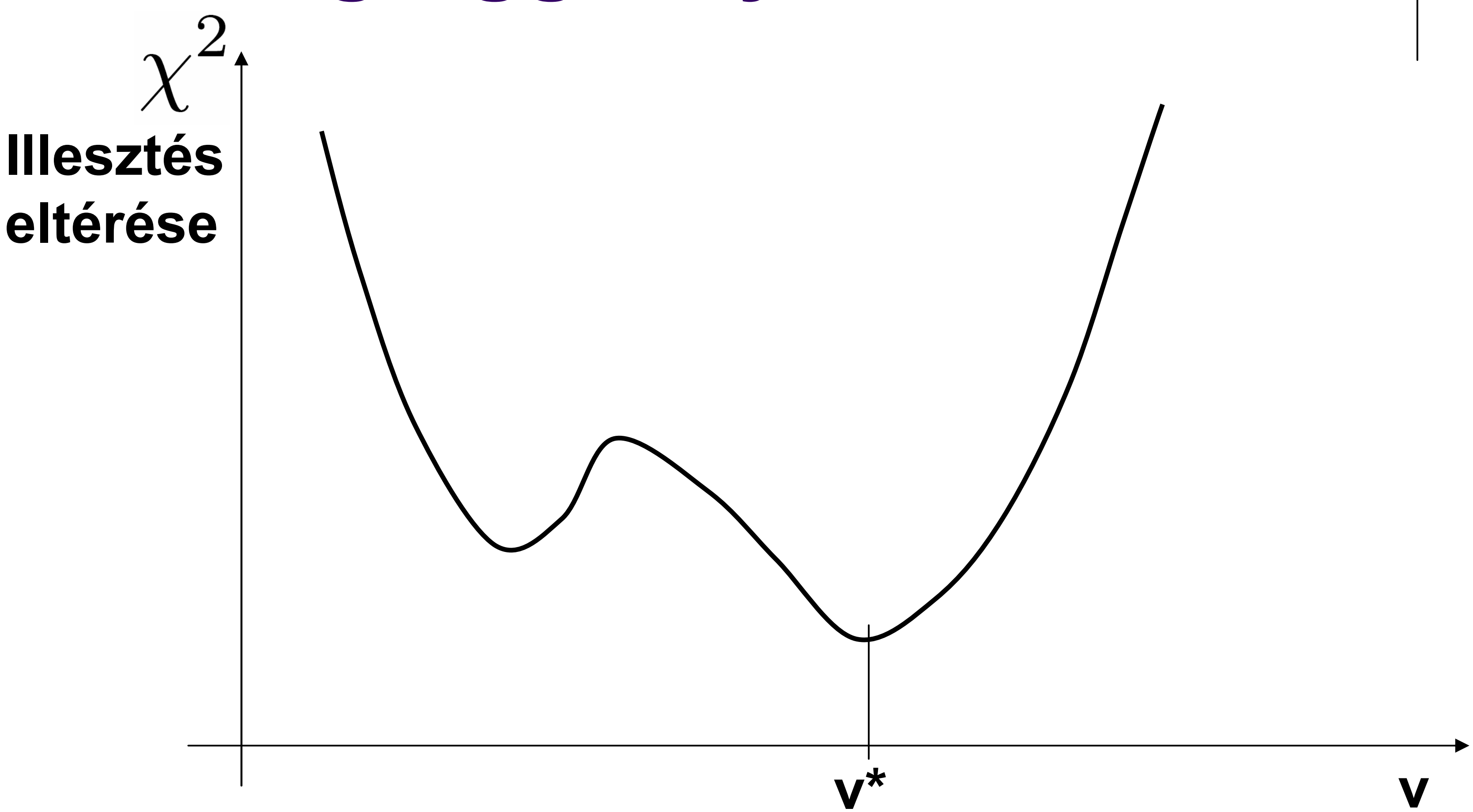


Költségfüggvény

Paramétereket változtatva mérhető
az illesztés jósága

Keressük a paraméterek optimális értékeit, amelyek a méréseket legjobban visszaadják

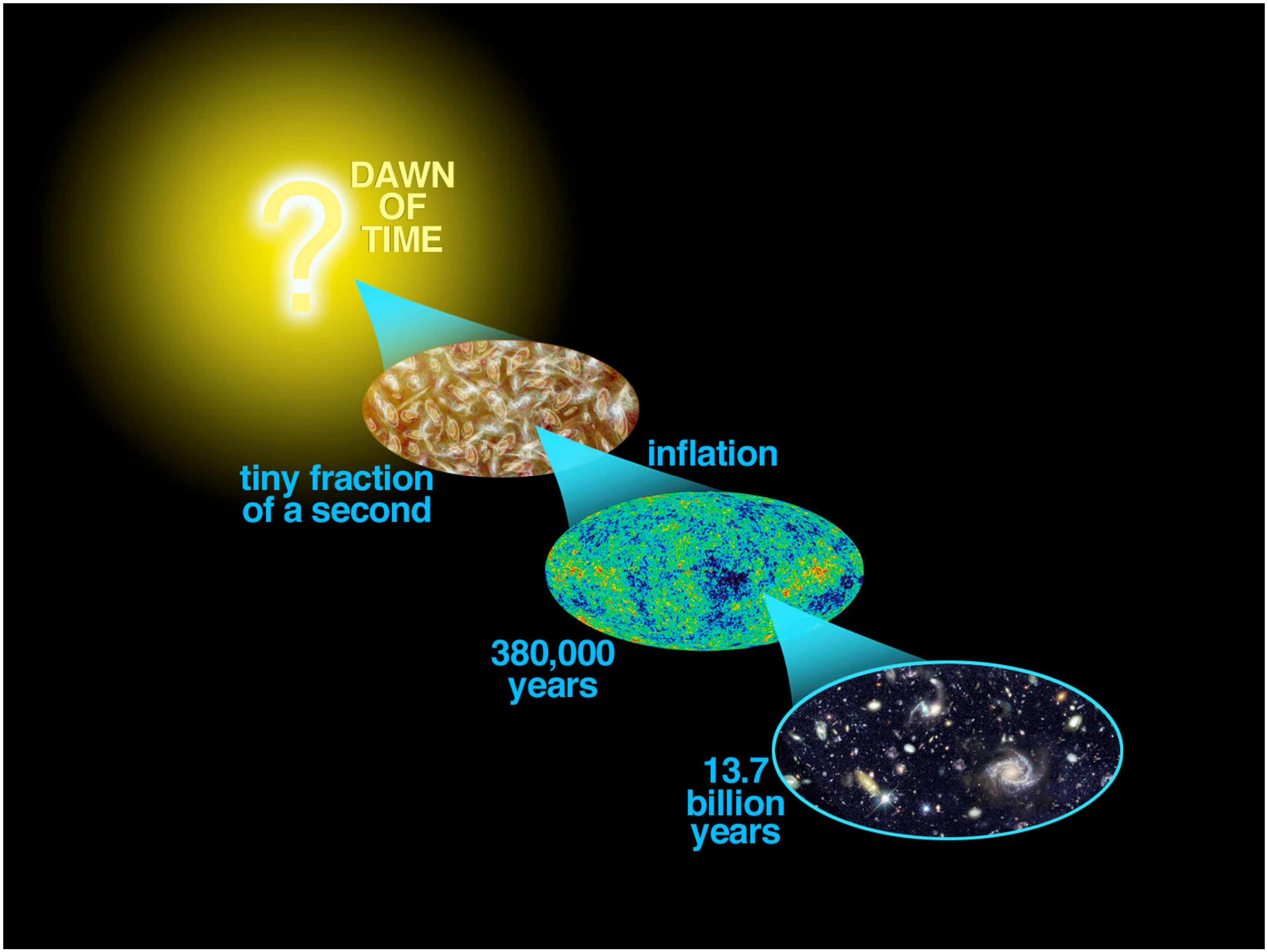
Költségfüggvény



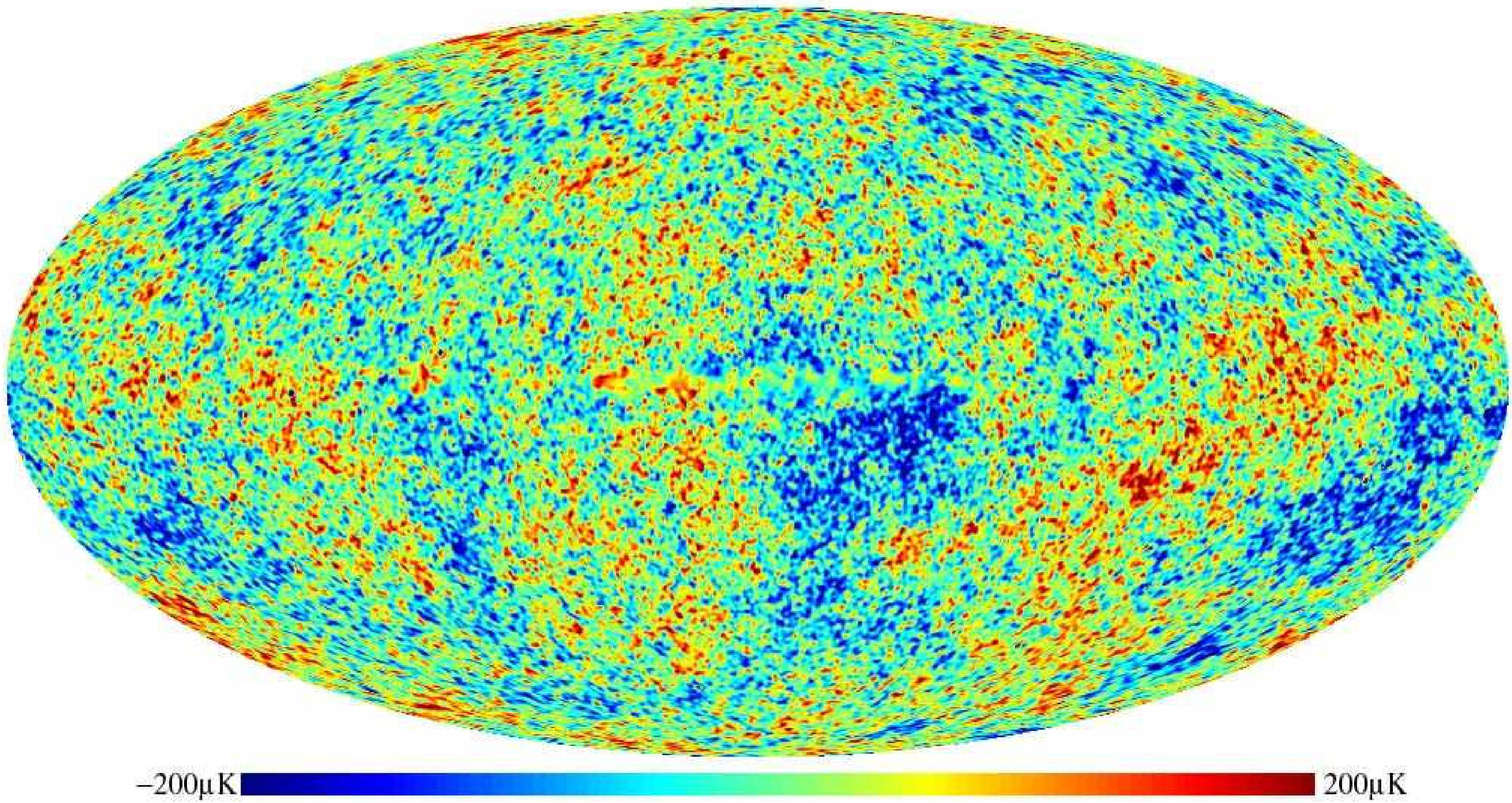
Minimumát keressük: általánosan ez egy nemlineáris optimalizációs probléma

Speciális eset: lineáris egyenlet

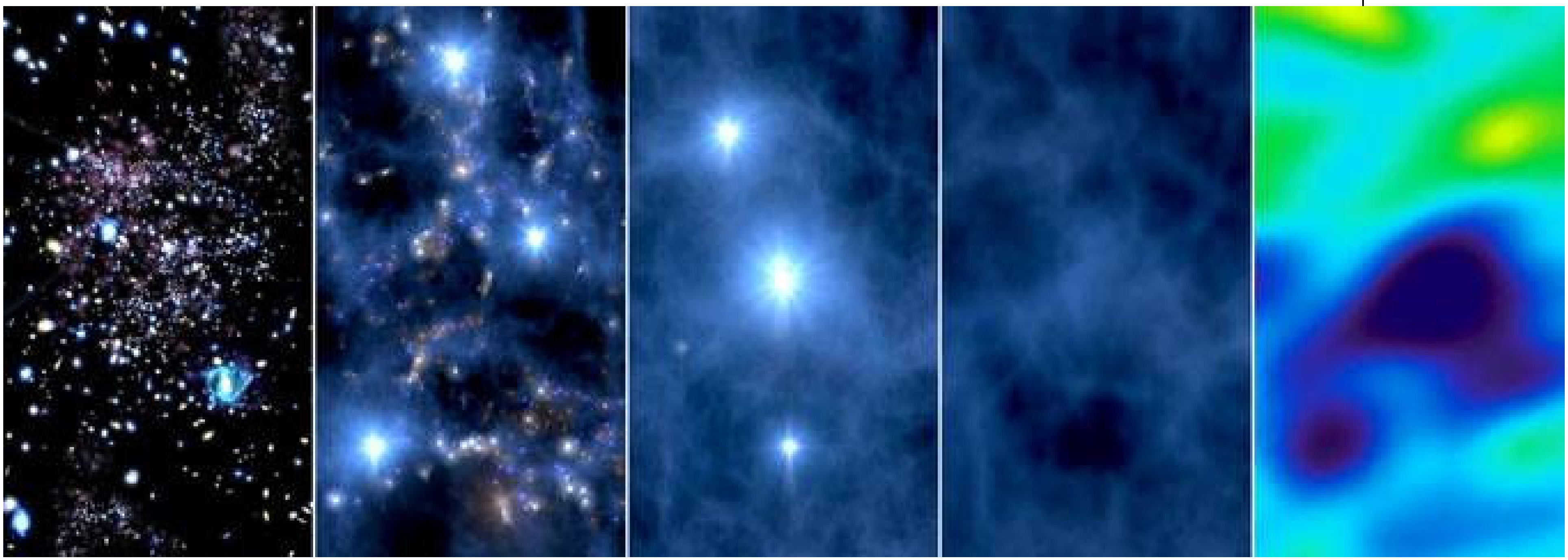
Példa: a világegyetem



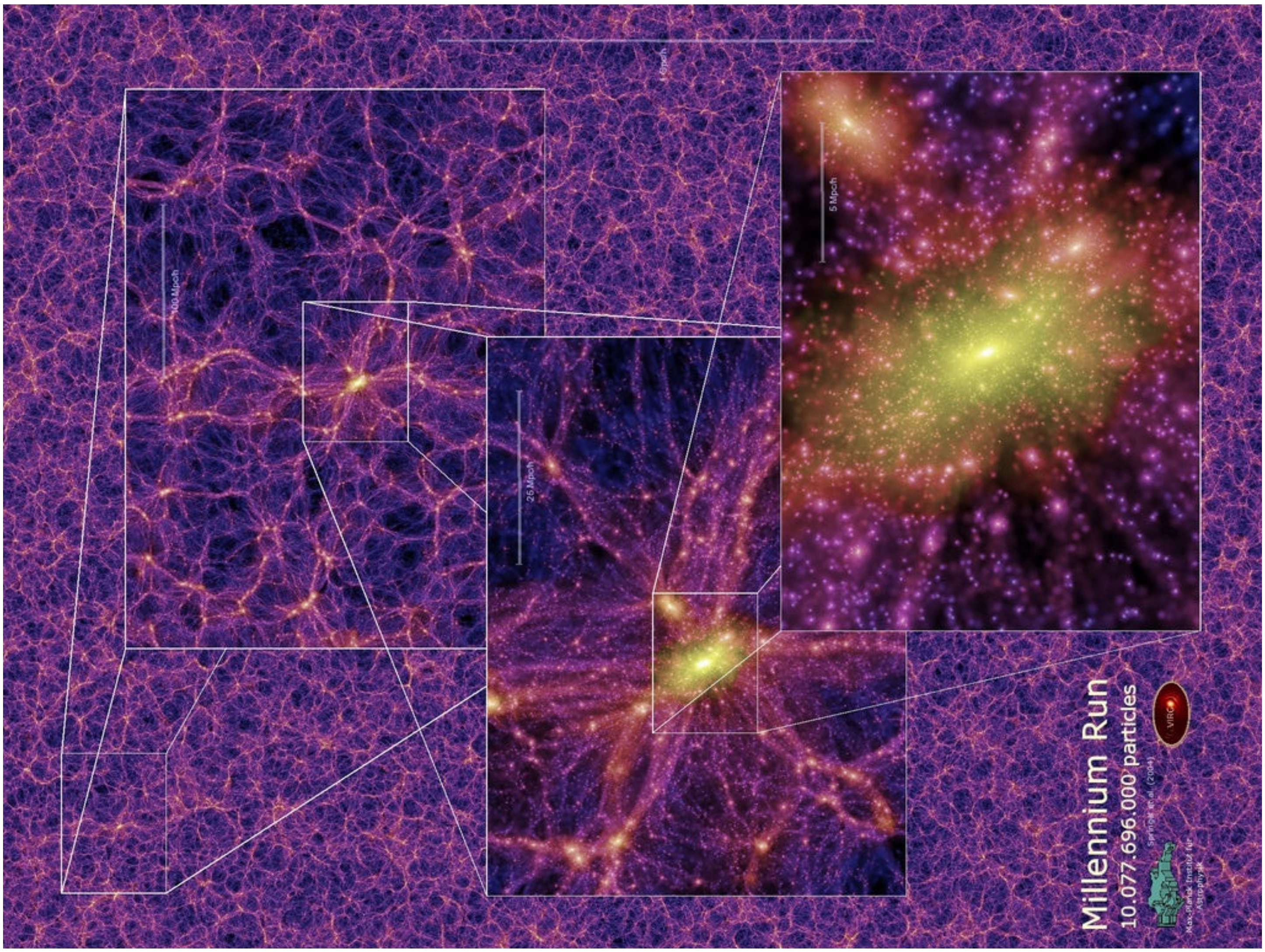
A mérés: WMAP



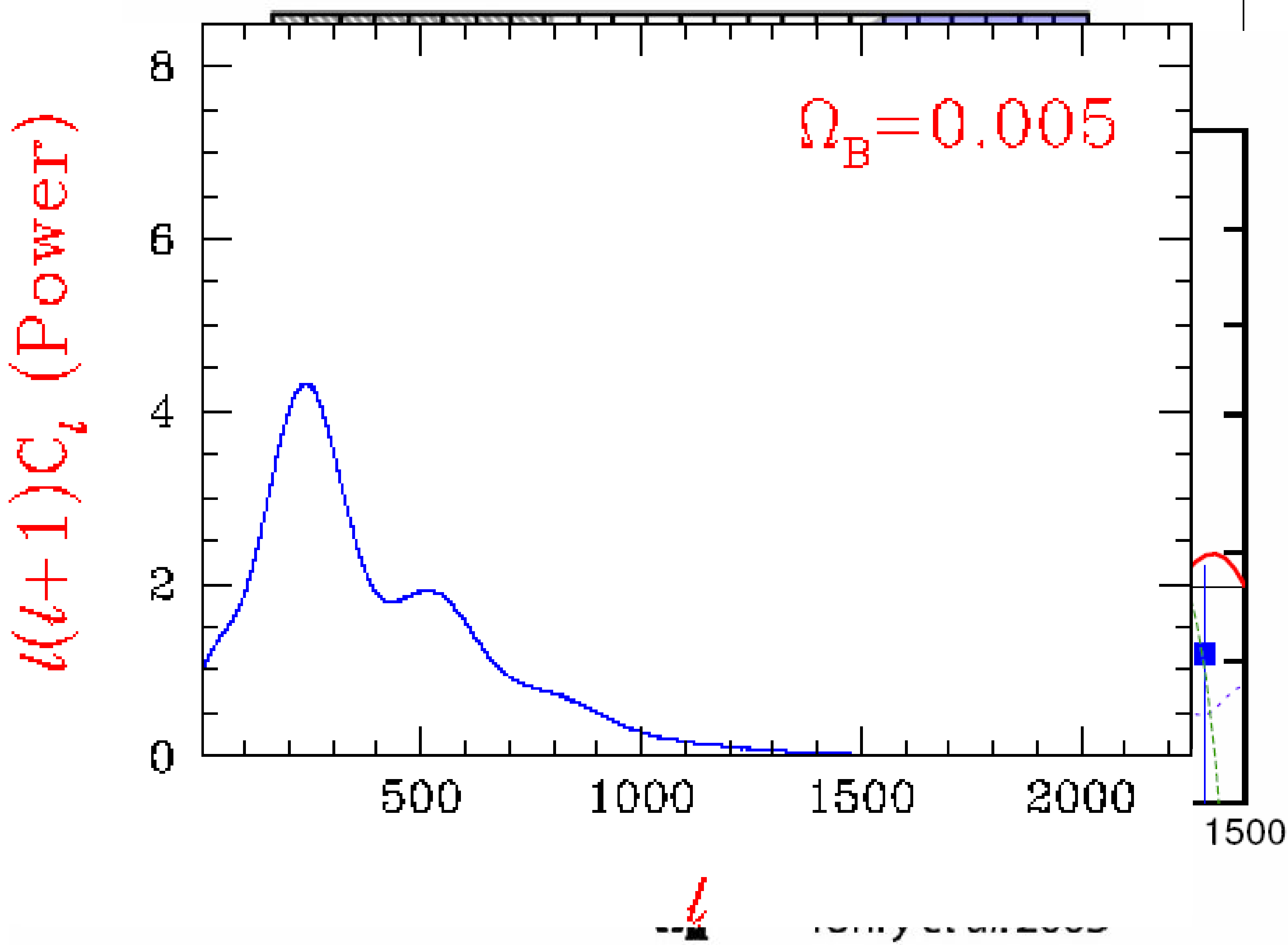
Szimuláció: Millennium Run



The Virgo Consortium
Dozen research groups, 10 milliárd “részecske”, 2 milliárd fényév
1500 processzor, 30 nap, 25TB kimenet


Millennium Run
10.077.696.000 particles

Példa: kozmológiai állandók



Hibák

- A mérési adatok **hibákat** tartalmaznak (mindegyik koordináta!)
  - Zaj, pontatlanság
  - Szisztematikus hibák
  - Kiugró pontok
- A kiszámolt paramétereknek is van bizonytalansága

A legkisebb négyzetek módszere

Adott: (x_i, y_i) ; $i=1..N$ mérési pontokra illesztünk egy függvényt :

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

a_j paraméterek, $j=1..M$

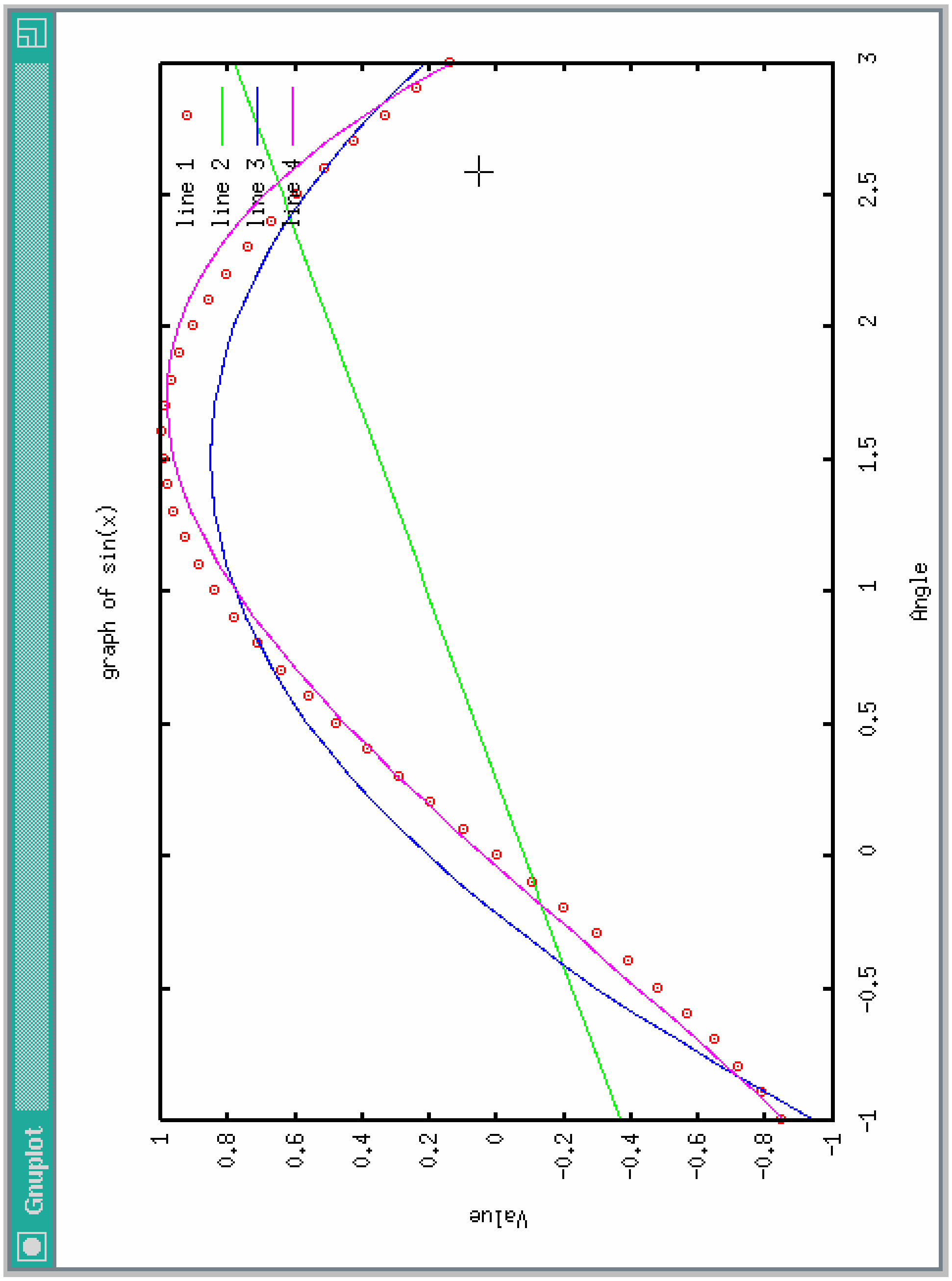
Legkisebb négyzetek:

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

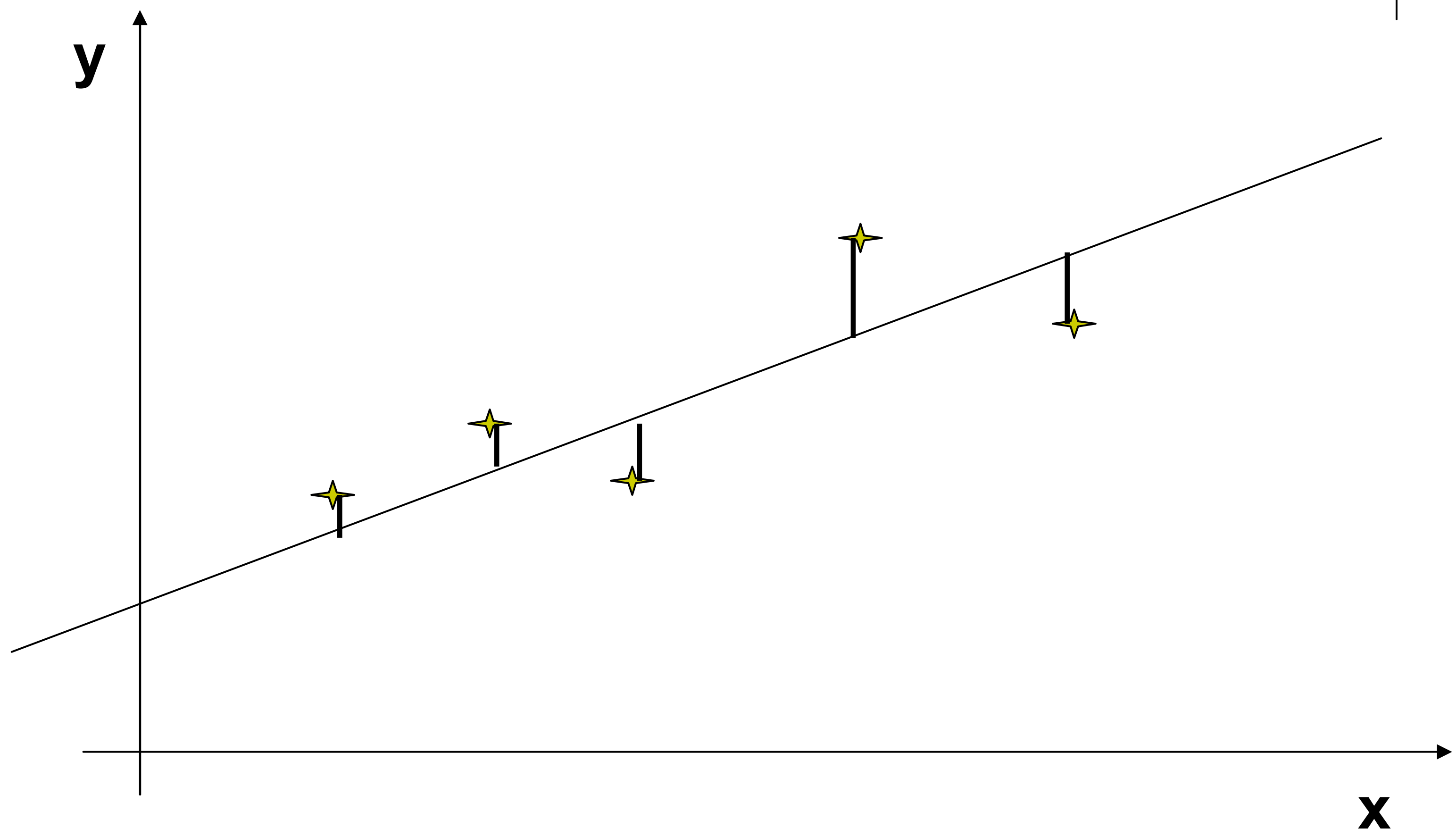
Függvényillesztés: octave-ban

```
octave:##> x=[-1,0,1,2,3];
octave:##> y=sin(x);
octave:##> p1=polyfit(x,y,1)
p1 =
    0.287448
   -0.077364
octave:##> p2=polyfit(x,y,2)
p2 =
   -0.28521
    0.85787
    0.20785
```

```
octave:##> p3=polyfit(x,y,3)
p3 =
  -0.069667
  -0.076209
   0.885735
   0.040645
octave:##> xx=-1:0.1:3;
octave:##> p1Fit=polyval(p1,xx);
octave:##> p2Fit=polyval(p2,xx);
octave:##> p3Fit=polyval(p3,xx);
octave:##> plot(xx,sin(xx), '*',
                xx,p1Fit,
                xx,p2Fit,
                xx,p3Fit)
```


Gnuplot
graph of sin(x)
line 1
line 2
line 3
line 4
Value
Angle

Legkisebb négyzetek



Legkisebb abszolútértékek?

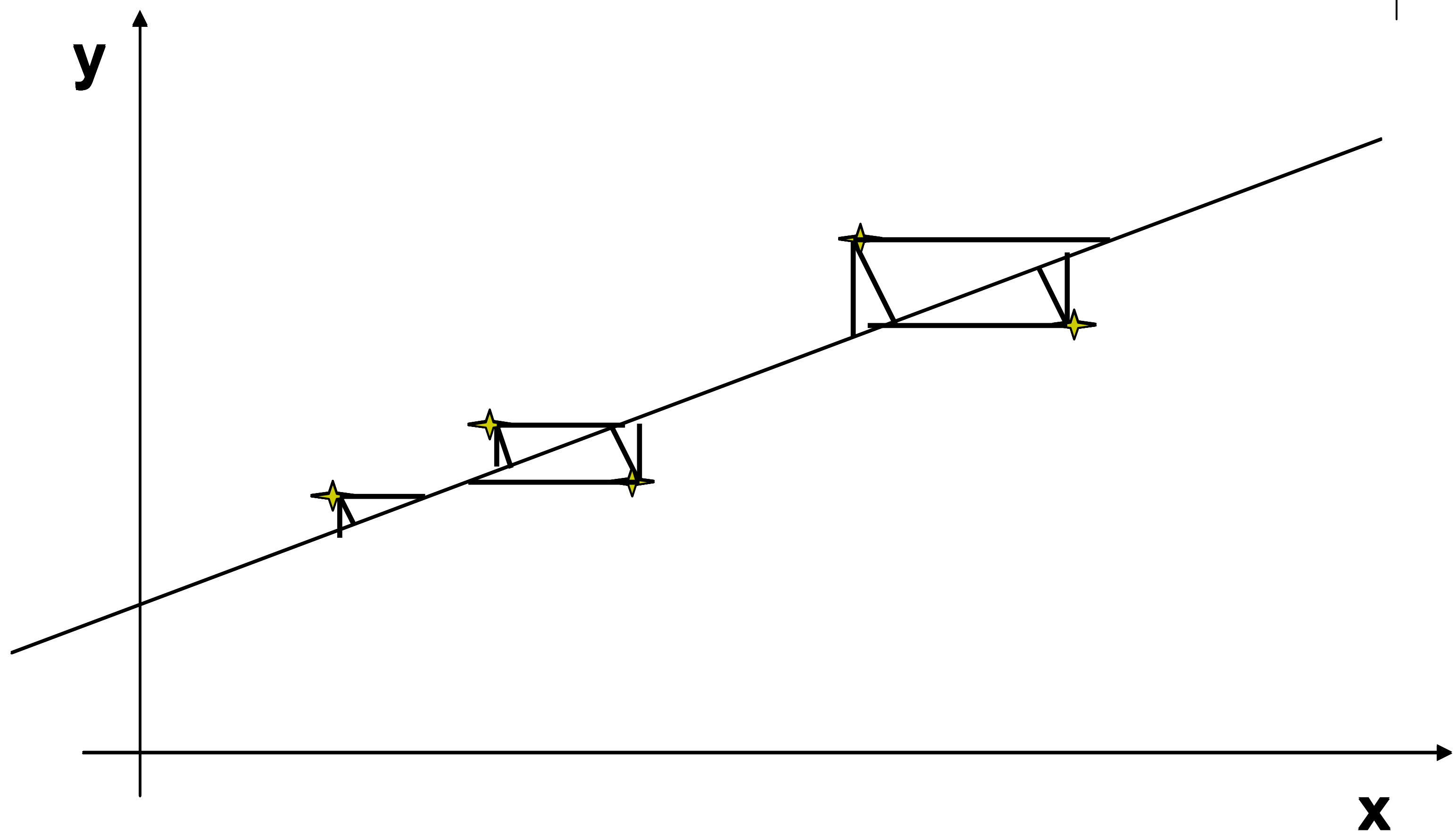
$$\min_{a_1 \ldots a_M} \left(\sum_{i=1}^{N} \left[y_i - y(x_i; a_1 \ldots a_M) \right]^2 \right)$$

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

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

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

Legkisebb távolságok?



Miért a legkisebb négyzetek?

- Maximális valószínűségű paraméterbecslés
  Maximum likelihood estimation
- A paraméterek legvalószínűbb értékét kapjuk, ha:
  - Csak y_i adatok hibáit vesszük figyelembe
  - A hibák Gauss eloszlásúak
  - A hibák függetlenek

Maximum likelihood

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

Azokat a paramétereket keressük ahol P maximális.

A log() fv. monoton, ezért kereshetjük log(P) maximumát, vagy –log(P) minimumát

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

Megkaptuk a legkisebb négyzeteket!

Interpoláció, extrapoláció

- Szintén alkalmas függvény közelítésre, de lényegesen más a cél:
  - nem fontos a „modell”
  - a görbe menjen át minden ponton
  - át-mintavételezés, sima görbék
  - globális: (n-1)-ed rendű polinom
  - lokális: szakaszonként simán találkozó interpoláció : spline
  - extrapoláció: kiterjesztés a tartományon kívülre
  - Octave: lásd kiadott anyag