

Základy programovania pre fyzikov

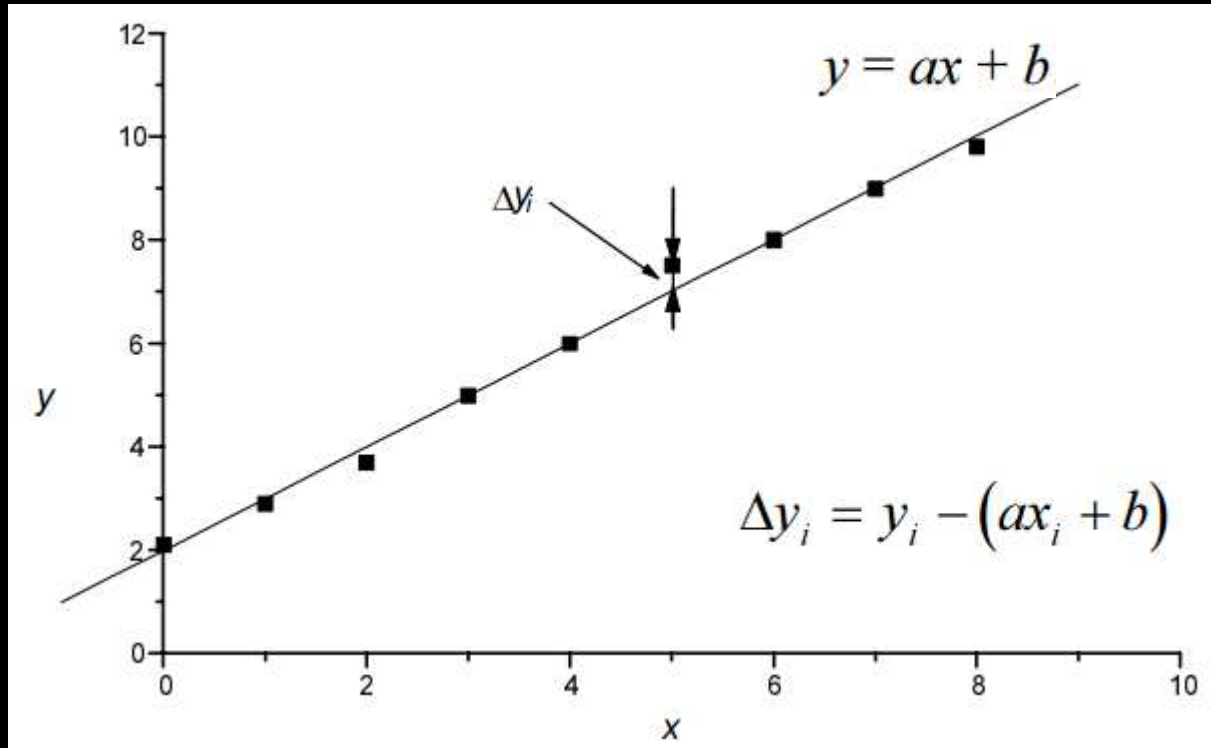
Erik Čižmár

Univerzita P.J. Šafárika

v Košiciach

Základy lineárnej regresie

Metóda najmenších štvorcov



$$S(a, b) = \sum_{i=1}^n [y_i - (ax_i + b)]^2 = \min$$

Základy lineárnej regresie

Metóda najmenších štvorcov

$$\frac{\partial S}{\partial a} = 0, \quad \frac{\partial S}{\partial b} = 0$$

$$\sum_{i=1}^n \{2[y_i - (ax_i + b)](-x_i)\} = 0$$

$$\sum_{i=1}^n \{2[y_i - (ax_i + b)](-1)\} = 0$$

$$a \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i = \sum_{i=1}^n x_i y_i$$

$$a \sum_{i=1}^n x_i + bn = \sum_{i=1}^n y_i$$

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

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

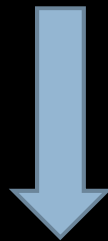
$$b = \frac{\sum_{i=1}^n y_i}{n} - a \frac{\sum_{i=1}^n x_i}{n} = \bar{y} - a\bar{x}$$

Základy lineárnej regresie

- Linearizácia funkcií $y(x)$, a , b $Y = AX + B$
- | | | |
|--------------------------------|--------------------------|--------------------------|
| □ $y = a e^{bx}$ | | |
| $\ln(y) = \ln(a) + b x$ | $Y = \ln(y), X = x$ | $A = \ln(a), B = b$ |
| □ $y = a x^b$ | | |
| $\ln(y) = \ln(a) + b \ln(x)$ | $Y = \ln(y), X = \ln(x)$ | $A = \ln(a), B = b$ |
| □ $y = a b^x, b > 0, b \neq 1$ | | |
| $\ln(y) = \ln(a) + \ln(b) x$ | $Y = \ln(y), X = x$ | $A = \ln(a), B = \ln(b)$ |
| □ $y = a + b/x$ | $Y = y, X = 1/x$ | $A = a, B = b$ |
| □ $y = 1/(a + bx)$ | $Y = 1/y, X = x$ | $A = a, B = b$ |

Kvadratická regresia

$$y_i = a_0 + a_1 x_i + a_2 x_i^2 \quad i = 1, \dots, n$$


$$\frac{\partial S}{\partial a_i} = 0, \quad i = 1, \dots, n$$

$$\sum_{i=1}^n y_i = a_2 \sum_{i=1}^n x_i^2 + a_1 \sum_{i=1}^n x_i + n a_0$$

$$\sum_{i=1}^n x_i y_i = a_2 \sum_{i=1}^n x_i^3 + a_1 \sum_{i=1}^n x_i^2 + a_0 \sum_{i=1}^n x_i$$

$$\sum_{i=1}^n x_i^2 y_i = a_2 \sum_{i=1}^n x_i^4 + a_1 \sum_{i=1}^n x_i^3 + a_0 \sum_{i=1}^n x_i^2$$

Všeobecný tvar lineárnej regresie

$$f(x) = c_1 f_1(x) + c_2 f_2(x) + \dots + c_k f_k(x)$$

$$y = ax + b$$

$$f_1(x) = x \text{ a } f_2(x) = 1$$

$$y = a/x + bx^2$$

$$f_1(x) = 1/x \text{ a } f_2(x) = x^2$$

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix}, \quad \mathbf{P} = \begin{pmatrix} p_1 & 0 & \dots & 0 \\ 0 & p_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & p_n \end{pmatrix}$$

$$\mathbf{X} = \begin{pmatrix} f_1(x_1) & f_2(x_1) & \dots & f_k(x_1) \\ f_1(x_2) & f_2(x_2) & \dots & f_k(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ f_1(x_n) & f_2(x_n) & \dots & f_k(x_n) \end{pmatrix}$$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} f_1(x_1) & f_2(x_1) & \dots & f_k(x_1) \\ f_1(x_2) & f_2(x_2) & \dots & f_k(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ f_1(x_n) & f_2(x_n) & \dots & f_k(x_n) \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix}$$

p^i - váha meraného bodu



$$\mathbf{y} = \mathbf{Xc}$$

Všeobecný tvar lineárnej regresie

$$S(c_1, c_2, \dots, c_k) = \sum_{i=1}^n \left\{ p_i \left[y_i - (x_{i1}c_1 + x_{i2}c_2 + \dots + x_{ik}c_k) \right]^2 \right\} = \min$$

$$\frac{\partial S}{\partial c_1} = 2 \sum_{i=1}^n \left\{ p_i [y_i - x_{i1}c_1 - x_{i2}c_2 - \dots - x_{ik}c_k]^2 (-x_{i1}) \right\} = 0$$

$$\frac{\partial S}{\partial c_2} = 2 \sum_{i=1}^n \left\{ p_i [y_i - x_{i1}c_1 - x_{i2}c_2 - \dots - x_{ik}c_k]^2 (-x_{i2}) \right\} = 0$$

⋮

$$\frac{\partial S}{\partial c_k} = 2 \sum_{i=1}^n \left\{ p_i [y_i - x_{i1}c_1 - x_{i2}c_2 - \dots - x_{ik}c_k]^2 (-x_{ik}) \right\} = 0$$

$$c_1 \sum_{i=1}^n x_{i1} p_i x_{i1} + c_2 \sum_{i=1}^n x_{i1} p_i x_{i2} + \dots + c_k \sum_{i=1}^n x_{i1} p_i x_{ik} = \sum_{i=1}^n x_{i1} p_i y_i$$

$$c_1 \sum_{i=1}^n x_{i2} p_i x_{i1} + c_2 \sum_{i=1}^n x_{i2} p_i x_{i2} + \dots + c_k \sum_{i=1}^n x_{i2} p_i x_{ik} = \sum_{i=1}^n x_{i2} p_i y_i$$

⋮

$$c_1 \sum_{i=1}^n x_{ik} p_i x_{i1} + c_2 \sum_{i=1}^n x_{ik} p_i x_{i2} + \dots + c_k \sum_{i=1}^n x_{ik} p_i x_{ik} = \sum_{i=1}^n x_{ik} p_i y_i$$

$$(X^T P X) c = X^T P y$$



$$c = (X^T P X)^{-1} X^T P y$$

Nelineárna regresia

- Ak neviem funkciu, ktorá popisuje moje dáta popísať ako lineárnu kombináciu bázových funkcií (ktoré nezávisia od hľadaných parametrov), neviem priamo spočítať hľadané parametre ako pri lineárnej regresii.
- Pre nájdenie minima sumy štvorcov následne používam iteratívne metódy – postupne sa mení sada hľadaných parametrov a na základe zmeny sumy štvorcov v ďalšom kroku zmením parametre
 - suma štvorcov je plocha v n -rozmernom priestore parametrov (ak mám n hľadaných parametrov) - podľa zmeny sklonu tejto plochy iterujem k parametrom zodpovedajúcim minimu
 - Levenberg-Marquardt
 - Simplex

Fitovanie v Originie

- **Linear Fit**
- **Polynomial Fit**
- **Multiple Linear Regression** – sada dát, ktorá je popísaná jedným modelom so spoločným parametrom – napr. poľná závislosť magnetizácie magnetickej látky zmeraná pri rôznych teplotách, kde spoločným parametrom je výmenná interakcia medzi magnetickými momentmi
- **Nonlinear Curve Fit** – nelineárna regresia funkciou jednej alebo viacerých premenných, fitovanie funkcií komplexných premenných

Fitovanie v Origin

The screenshot shows the Origin software interface for fitting an exponential decay function. The dialog box is titled "NLFit (ExpDec1)" and has tabs for "Settings", "Code", "Parameters", and "Bounds". The "Settings" tab is active, showing a "Function Selection" list on the left and a configuration panel on the right. The configuration panel includes:

- Category:** Origin Basic Functions
- Function:** ExpDec1
- Description:** Exponential Decay 1
- File Name(.fdf):** C:\Program Files\OriginLab\Origin85\fitfunc\expdec1.fdf



A red message bar at the bottom of the dialog reads: "Data for fitting not specified. Please update the Input Data in the Data Selection step on Settings tab." Below the message bar is a toolbar with icons for fit, data, save, zoom, and other functions, along with "Fit", "Done", and "Cancel" buttons.




At the bottom of the dialog, there are tabs for "Fit Curve", "Residual", "Formula", "Sample Curve", "Messages", "Function File", and "Hints". The "Sample Curve" tab is selected, displaying a graph of an exponential decay curve. The curve is shown in red, with a black tangent line at the point $(0, A+y_0)$. The horizontal asymptote is labeled $y=y_0$. The equation $y(1) = -A/t$ is also shown. A text box on the right contains the following parameters:


- offset: $y_0=1$
- amplitude: $A=2$
- decay constant: $t=1$


Fitovanie v Origin



	Edit Fitting Functions
	Create New Fitting Functions
	Save FDF File

	Calculate Chi-Square
	1 Iteration
	Fit until converged

	Initialize Parameters
---	-----------------------

	Simplex
---	---------

Editovanie funkcií

- Fitting Function Organizer

The screenshot shows the 'Fitting Function Organizer' dialog box. The left pane lists various functions, with 'ExpDec1' selected. The main area displays the following fields:

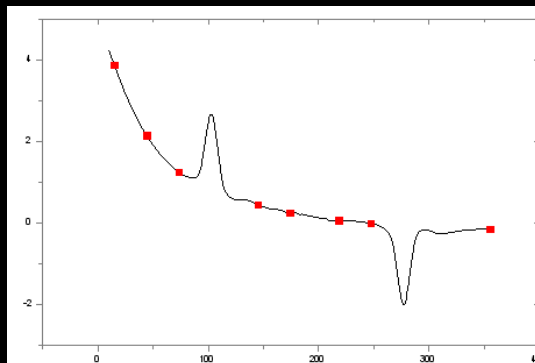
- Function Name:** ExpDec1
- File Name(.FDF):** C:\Program Files\OriginLab\Origin85\fitfunc\expdec1.fdf
- Brief Description:** Exponential Decay 1
- References:** (Empty list)
- Function Type:** Built-in
- Independent Variables:** x
- Dependent Variables:** y
- Parameter Names:** y0,A1,t1
- Function Form:** Expression
- Function:** $y = A1 \cdot \exp(-x/t1) + y0$
- Peak Function:**
- Parameter Settings:** NamingMethod = User-Defined; Meanings = offset,amplitude,decay constant; LowerBounds = -- 0.00X.00F1.0.00X.00F1

The bottom of the dialog shows the equation:

$$y = y_0 + Ae^{-x/t}$$

Fitovanie píkových závislostí

- Spektroskopické merania jeden alebo viac maxím - píkov
- Funkcie typu Gaussián, Lorentzián, Voigth, Laplace
- **Baseline** – signál pozadia



- **Určenie polohy píkov**
- **Plocha pod píkcom** – intenzita signálu – aké „množstvo“ daného fyzikálneho javu sme pozorovali – relatívne, absolútne
- **Polšírka píku** – ako veľmi je pozorovaný fyzikálny proces „rozmazaný“ - napr. pri magnetickej rezonancii nie všetky častice rezonujú pri rovnako veľkom magneticom poli
- Single Peak Fit alebo Multiple Peak Fit

Fitovanie píkových závislostí

- Peak Analyzer
 - Baseline – Constant, User Defined , Dataset
 - Create Baseline – interpolácia
 - Baseline Treatment – odpočítanie
 - Find Peaks – maximum, 1. derivácia, 2.derivácia
 - Intergrate peaks – intenzita, polšírka

