Modelska analiza 1

6. naloga

Tina Klobas

22. november 2018

1 Opis problema

Pri tej nalogi želimo danim nizom podatkov prilagoditi linearno funkcijo (*fit*). Imamo začeten linearni sistem

$$A\mathbf{x} = \mathbf{b} \tag{1}$$

kjer je A podana matrika parametrov (funkcij) sistema, $\mathbf{b}$ vektor rešitev, vektor $\mathbf{x}$ pa želimo poiskati tako, da bo napaka

$$\|A\mathbf{x} - \mathbf{b}\| = \min. \tag{2}$$

Singularni razcep

Za vsako matriko $A \in \mathbb{R}^{m\times n}$ obstaja singularni razcep:

$$A = U\Sigma V^T, \tag{3}$$

kjer sta $U \in \mathbb{R}^{m\times m}$ in $V \in \mathbb{R}^{n\times n}$ ortogonalni matriki in $\Sigma \in \mathbb{R}^{m\times n}$ diagonalna z elementi $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n$ (singularne vrednosti). Velja tudi, da je minimum (enačba 2) dosežen pri vektorju:

$$\mathbf{x} = \sum_{i=1}^{n} \frac{\mathbf{u}_i^T\mathbf{b}}{\alpha_i}\mathbf{v}_i, \tag{4}$$

kjer sta vektorja $\mathbf{u}_i$ in $\mathbf{v}_i$ stolpca matrik U in V. Napake elementov tega vektorja pa so:

$$\sigma_i = \sum_{j=1}^{n} \left(\frac{V_{ij}}{\alpha_j}\right)^2 \tag{5}$$

Test χ^2

Delamo minimizacijo kumulativne χ^2 funkcije:

$$\chi^2 = \sum_{i=1}^{N} \left(\frac{b_i - \mathbf{a}_i x_i}{\sigma_i} \right)^2, \tag{6}$$

kjer je $\mathbf{a}_i$ vrstica v matriki A.

2 Odziv tkiva

Opazujemo odziv tkiva na različne reagente;

$$\mathrm{Y} + \mathrm{X} \rightleftharpoons \mathrm{Y}^* \tag{7}$$

Če povišamo koncentracijo reagenta nad neko vrednost nasičenja, ga tkivo ne bo več sprejelo. V stacionarnem stanju lahko to zapišemo z enačbo:

$$y = \frac{y_0 x}{x + a}, \tag{8}$$

kjer je y_0 nasičeni odziv tkiva in a koncentracija, potrebna za odziv, ki je enak polovici nasičenega. V splošnem je ta zveza nelinearna, lahko pa jo preuredimo v obliko:

$$\frac{1}{y} = \frac{1}{y_0} + \frac{a}{y_0}\frac{1}{x}. \tag{9}$$

Z uvedbo $z = 1/y$, $z_0 = 1/y_0$, $t = 1/x$ ter $k = a/y_0$ pridemo do linearne enačbe oblike:

$$z = kt + z_0 \tag{10}$$

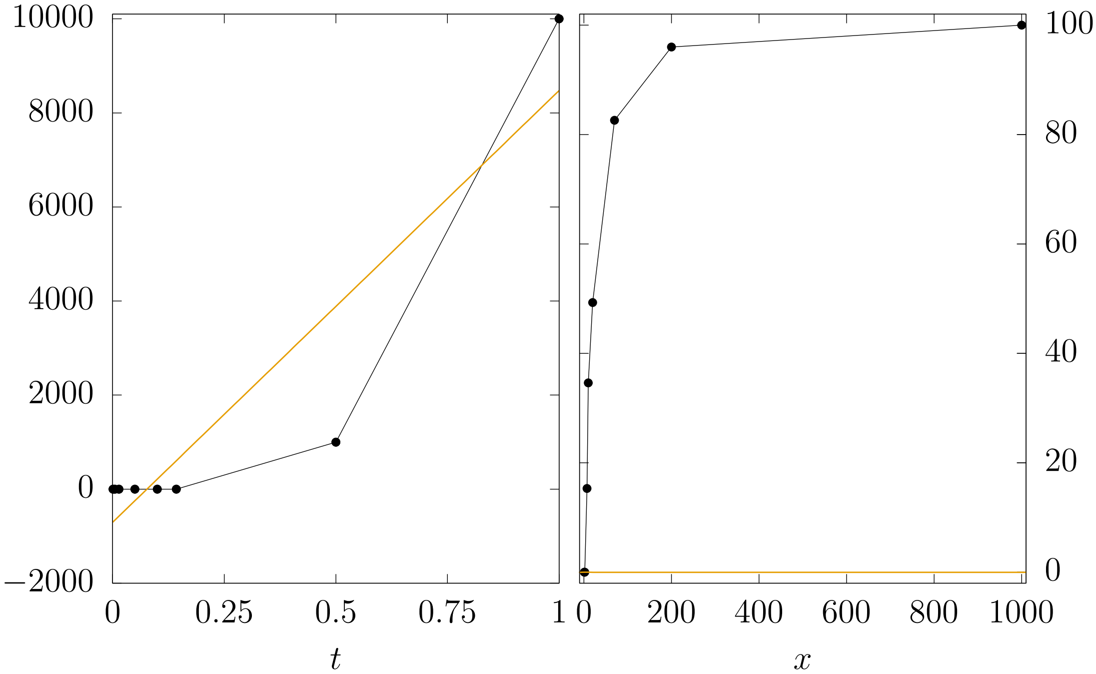
Napake začetnega problema 8 so σ_i, ki se z lineariziranjem enačbe tudi prepišejo:

$$\Delta z = \left| \frac{\mathrm{d}z}{\mathrm{d}y} \right| \Delta y = \frac{1}{y^2}\sigma. \tag{11}$$

Napake linearizirane enačbe 9 so:

$$\widetilde{\sigma}_i = \sigma_i / y_i^2 = \sigma_i z_i^2. \tag{12}$$

Na grafu 2.1 je prikazana linearna regresija v primeru, ko ne upoštevamo napak posameznih točk. Na levem grafu smo podatkom prilagajali enačbo 10 in vidimo, da precej odstopata zadnji dve točki (na desnem grafu prvi dve), ki imata v skladu z enačbo 12 tudi zelo veliko napako. Ker imajo vsi podatki



Slika 2.1: Na grafu je prikazana linearna regresija brez upoštevanja uteži napak na podatki y in obratni vrednosti z.

enako utež, oddaljene točke močno razpotegnejo premico in take oblike rešitev nima nobenega smisla.
Reševanja se lotimo tako, da minimiziramo χ^2 – poiščemo minimum enačbe 6 in pridemo do sistema enačb oblike $A\mathbf{a} = \mathbf{b}$, kjer definiramo elemente matrike A:

$$A_{ij} = \sum_{k_1}^{N} \frac{\phi_i(x_k)\phi_j(x_k)}{\widetilde{\sigma}_k}. \tag{13}$$

ϕ_i so testne funkcije. V našem primeru imamo samo dve funkciji; $\phi_1 = 1$ in $\phi_2 = t = 1/x$ in matrika A ima tako obliko:

$$A = \sum_{k=1}^{N} \frac{1}{\widetilde{\sigma}_k^2} \begin{bmatrix} 1 & \frac{1}{x_k} \\ \frac{1}{x_k} & \frac{1}{x_k^2} \end{bmatrix}. \tag{14}$$

Imamo še vektor $\mathbf{b}$ z elementi

$$b_i = \sum_{k=1}^{N} \frac{z_k \phi_k(x_k)}{\widetilde{\sigma}_k^2}, \tag{15}$$

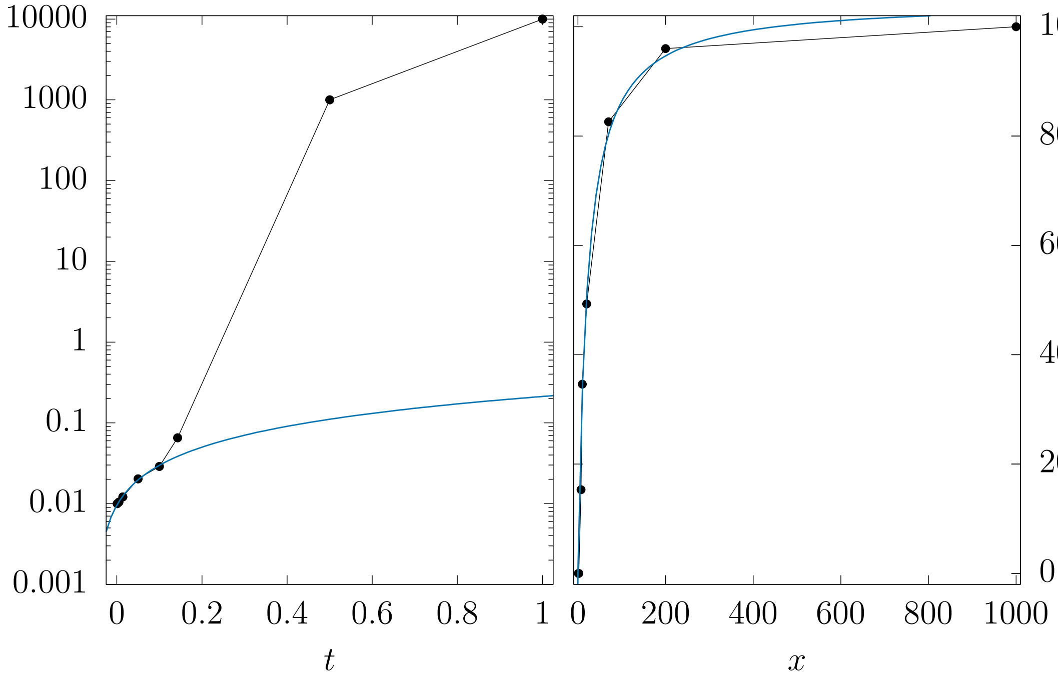
ki ima v tem primeru samo dva člena

$$\mathbf{b} = \sum_{k=1}^{N} \frac{z_k}{\widetilde{\sigma}_k^2} \begin{bmatrix} 1 \\ \frac{1}{x_k}. \end{bmatrix} \tag{16}$$

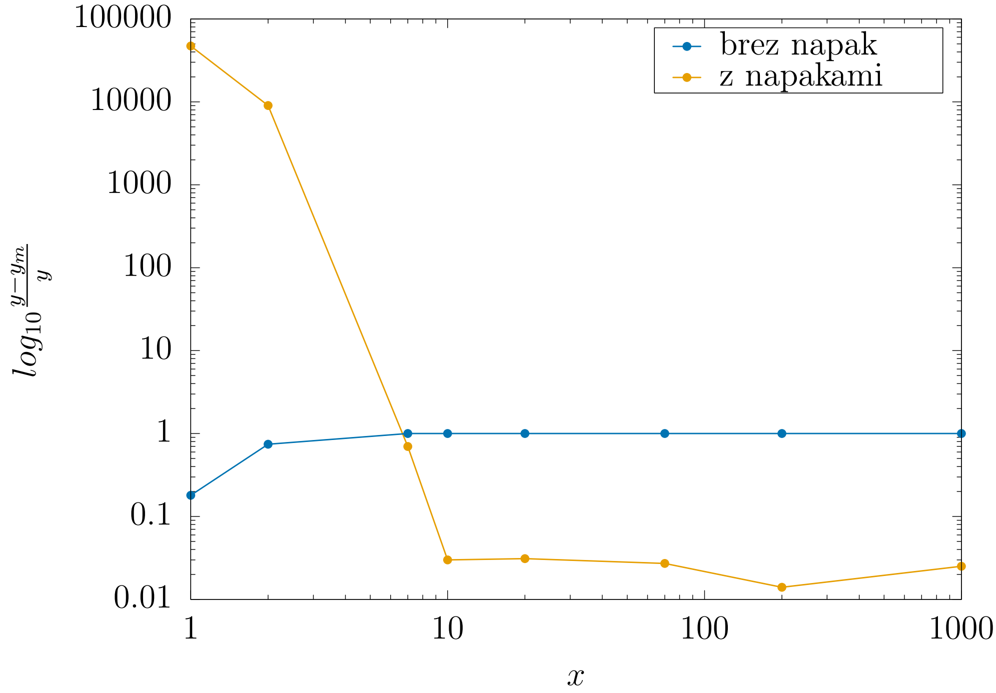
Matriki A naredimo inverz, da dobimo vektor

$$\mathbf{a} = A^{-1}\mathbf{b} \tag{17}$$

in rešitev narišemo na graf 2.2. Vidimo, da smo sedaj veliko bolj upoštevali prvih nekaj točk, ki imajo manjšo napako in tako je tudi krivulja na desnem grafu veliko bolj pravilne oblike kot na prejšnji sliki. Obe rešitvi lahko tudi primerjamo s pomočjo testa χ^2 kjer za prvi primer, prikazan na sliki 2.1 dobimo $\chi^2 = 3322,38$ za drugega, kjer smo upoštevali še napake, pa $\chi^2 = 26.08$. Sicer niti druga vrednost ni še enaka idealni $\chi^2 = N - M = 6$ (kjer je N število izmerkov, M pa število parametrov), vendar je veliko bolje kot v prvem primeru. Na grafu 2.3 tudi vidimo primerjavo napak obeh metod. Sicer je na začetku drugi rezultat, ko smo gledali regresijo z upoštevanjem uteži napak odstopanje od pravilne vrednosti veliko večje, vendar že po drugem izmerku pade pod drugo funkcijo in dosežemo natančnost na dve decimalni mesti.



Slika 2.2: Na grafu je prikazana linearna regresija z minimizacijo χ^2 porazdelitve.



Slika 2.3: Primerjava napak obeh metod.

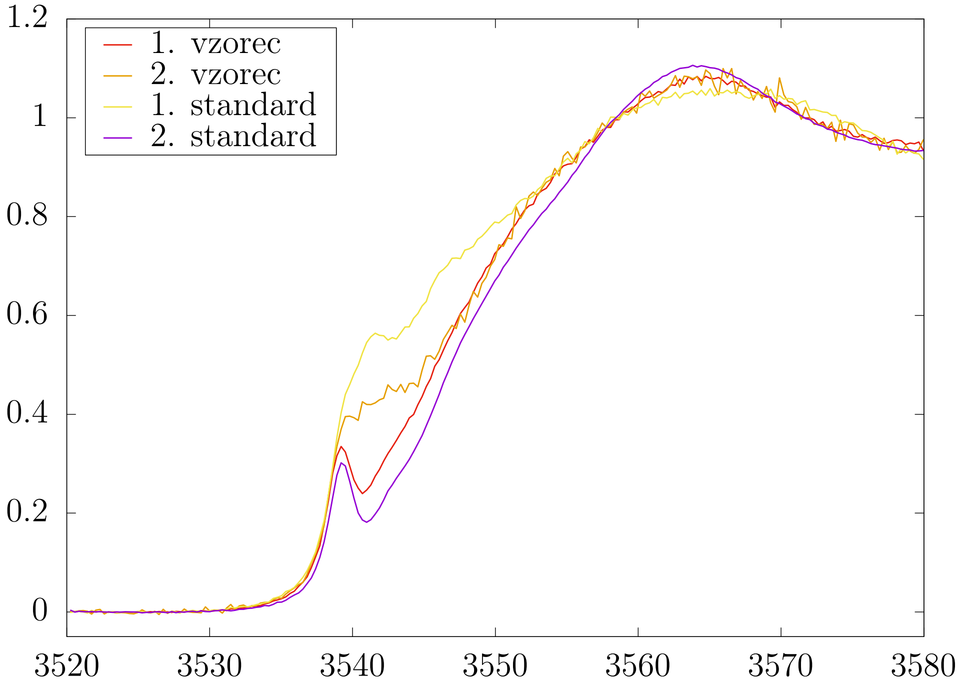
3 Spekter na vzorcih lista

Imamo zbrane štiri absorpcijske spektre kadmija na robu L_3 iz študije, kako ta kovina vpliva na rastline. V prvih dveh vzorcih so izolirane celične stene iz krovne plasti in iz sredice listov rastline C. *Thlaspi*, ki je znan hiperakumulator težkih kovin. Zadnja dva spektra sta dobljena na standardih, kompleksih Cd sulfata z glutationom (GSH) in pektinom: v prvem je Cd vezan izključno na žveplo, v drugem na kisik. V listnih vzorcih dopuščamo obe vrsti vezave, vemo pa, da sta prispevka obeh v spektru linearno sestavljena. Želimo dobiti odstotno razmerje vezi Cd−O in Cd−S v obeh listnih vzorcih. Vsi štirje spektri so prikazani na grafu 3.1. Linearno sestavljenost vzorcev lahko zapišemo z enačbama:

$$\text{vzorec}_1(E) = a \cdot \text{standard}_1(E) + b \cdot \text{standard}_2(E), \tag{18}$$

$$\text{vzorec}_2(E) = c \cdot \text{standard}_1(E) + d \cdot \text{standard}_2(E). \tag{19}$$

Problem se lahko lotimo na enak način kot prejšnjega ali pa z metodo singularnega razcepa kjer za optimalni vektor uporabimo enačbo 4 za njegovo napako pa 5. Z grafa 3.2 vidimo, da je naše prilagajanje kar uspešno. Pri drugem vzorcu vidimo, da je na nekaterih območjih precej bolj razgibana,



Slika 3.1: Vsi štirje podani sprektri.

kar tudi povzroči odstopanje od prave vrednosti. To se vidi tudi z grafa 3.3 kjer je napaka pri drugem spektru ponekod za tudi do več velikostnih redov večja, vendar sta obe še zmeraj zelo majhni, kar potrdi, da smo našli optimalno oceno.

Dobili smo parametre a, b, c, d, ki predstavljajo razmerja vezi Cd$-$O in Cd$-$S v obeh spektrih. Za prvi spekter je

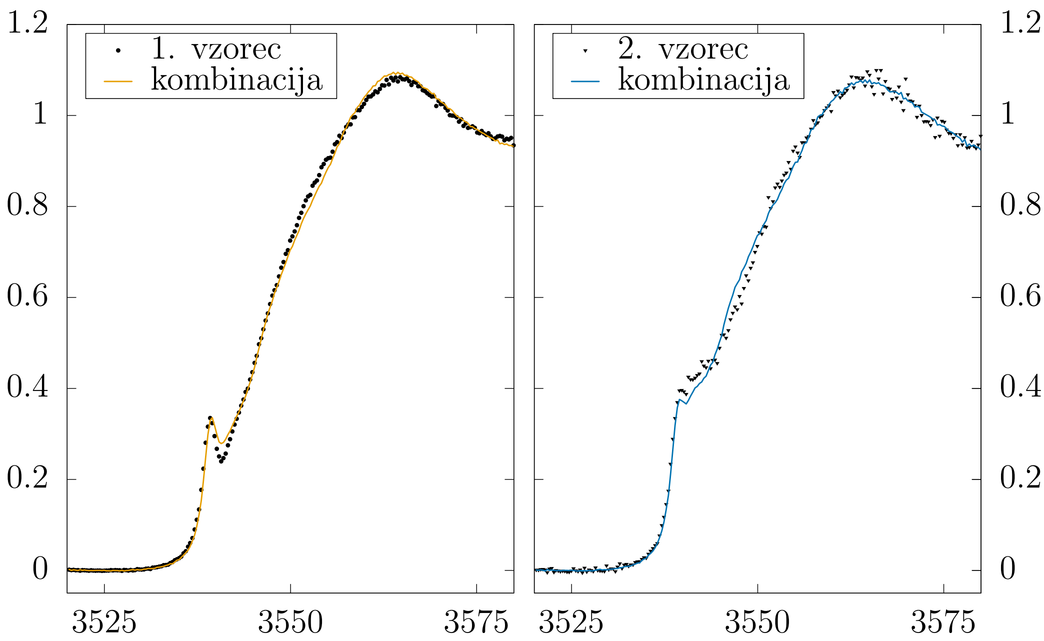
1. spekter: razmerje vezi Cd$-$O/Cd$-$S je

$$\frac{a}{b} = \frac{0.27}{0.73}$$

,

2. spekter: razmerje vezi Cd$-$O/Cd$-$S je

$$\frac{c}{d} = \frac{0.55}{0.45}.$$



Slika 3.2: Narisana sta oba vzorca in dobljena kombinacija standardov.

4 Magnetni spektrometer

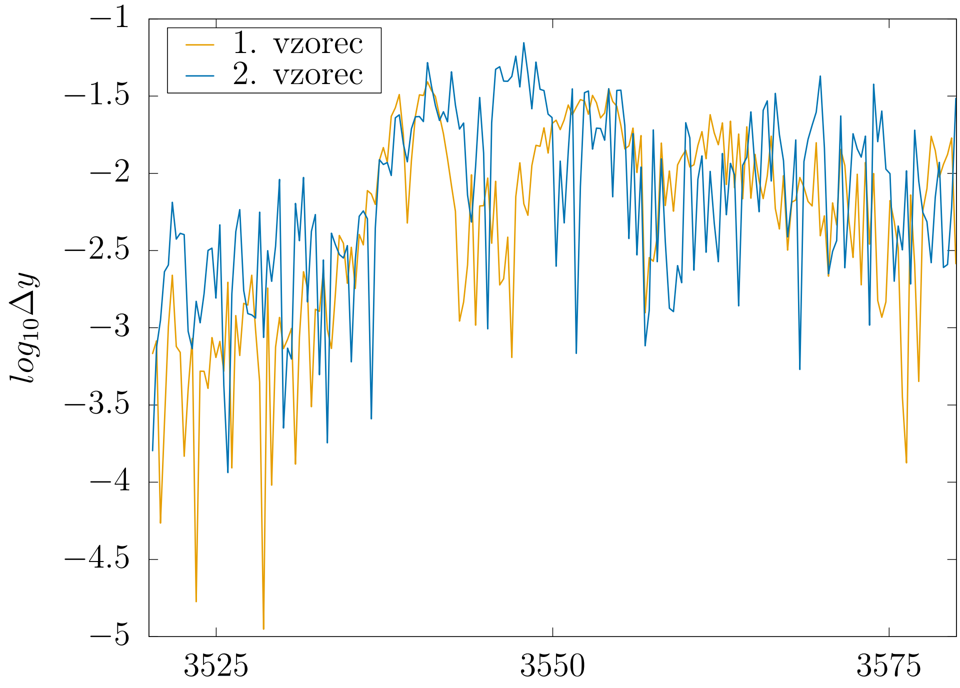
Z dobro rekonstrukcijo poti delcev v spektrometrih lahko potem izračunamo poljubne kinematične količine. Naša naloga bo tako poiskati to rekonstrukcijo. Podane imamo podatke o začetnih položajih x_{fp} v goriščni ravnini, odmik ϑ_{fp} od goriščne ravnine in disperzijski kot ϑ_{t} na tarči. Želimo poiskati preslikavo

$$(x_{\mathrm{fp}}, \vartheta_{\mathrm{fp}}) \to \vartheta_{\mathrm{t}}. \tag{20}$$

Za testni model vzamemo obliko:

$$f(\vartheta, x) = \left(\sum_{p=0}^{N} b_p \vartheta^p\right)\left(\sum_{q=0}^{M} c_q x^q\right). \tag{21}$$

Spet lahko rešujemo preko metode singularnega razcepa in po χ^2 testu preverimo kako dobro se prilagajajo podatkom. Grafi podatkov in *fita* niso najbolje pregledni, je pa toliko bolj zanimivo kako se χ^2 spreminja s spreminjanjem polinoma 21. S slike 4.1 vidimo, da smo najbolj ugoden moden preslikave dobili, ko smo vzeli polinom s dvajsetimi členi, kar bo takrat, ko bomo vzeli stopnjo $N = 4$ in $M = 5$ in upoštevali le produkte skupne stopnje

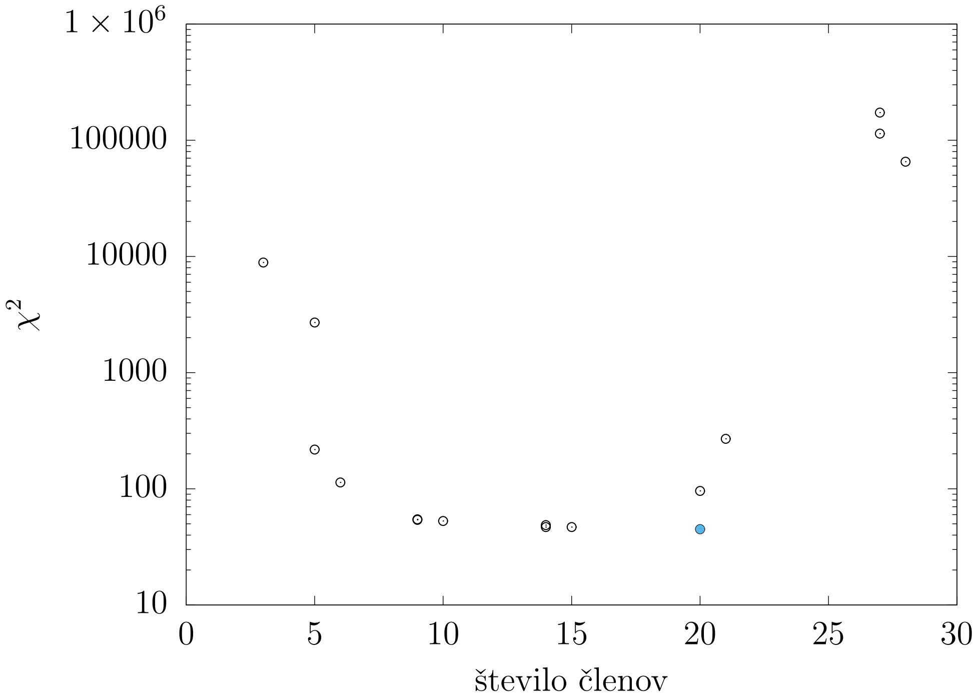


Slika 3.3: Prikazan je logaritem odmikov dobljenih vrednosti od izmerjenih.

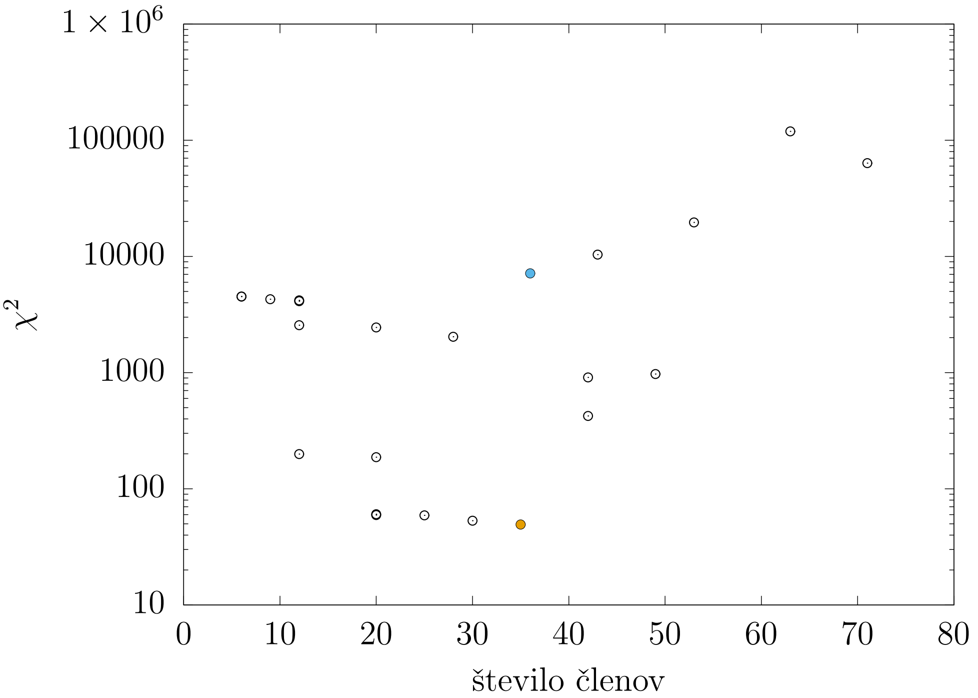
$p + q > 5$. Točko tik nad tem minimumom dobimo na enak način, le da je tam stopnja $N = 5$ in $N = 4$. Pri vsakem številu členov bi dobili več točk, če bi upoštevali tudi višje potence pri vsakem računanju, vendar pri večji razliki med maksimalnima stopnjama polinomov M in N dobivamo večja odstopanja kot tukaj kjer smo gledali le $|N - M| = 1$.

Negativne potence polinoma

Podobno lahko naredimo tudi, če vzamemo, da lahko p in q zavzameta tudi negativna števila in zopet narišemo graf 4.2 χ^2 v odvisnosti od števila členov polinoma. Tokratni minimum imamo pri 35 členih – narisan z oranžno, z modro (pri 36 členih) pa je narisan optimalni polinom s prejšnjega grafa, le da mu tokrat dovolimo da zavzame še člene na stopnjo $p = q = -1$. Vrednost minimuma je primerljiva s prejšnjim ($\chi_1^2 = 44.98$, $\chi_2^2 = 49.33$) vendar smo morali porabiti veliko več računanja, da smo dobili (nekoliko) slabše prilagajanje.



Slika 4.1: Primerjava odstopanja od prave vrednosti za različno število členov v polinomu.



Slika 4.2: Primerjava odstopanja od prave vrednosti za različno število členov v polinomu, ki vsebuje tudi negativne potence.