

# Metoda nejmenších čtverců a její aplikace

Zdeněk Mikulášek, Ústav teoretické fyziky a astrofyziky, PřF Masarykovy Univerzity,  
Brno

Příloha - Metoda nejmenších čtverců a její aplikace

## 1 Úvod

Objekty s proměnnými charakteristikami jsou předmětem soustředěného zájmu astrofyziků, protože svou proměnností toho o sobě prozrazují mnohem více, než objekty neproměnné. Zjištění a matematické vyjádření povahy časové proměnnosti měřených veličin (magnetické pole, jasnost, intenzita spektrálních čar, polarizace apod.), hledání trendů, cyklických změn, periodicit apod. - to jsou nejčastější úkoly, které praktická astrofyzika řeší. Nejdůležitějším nástrojem pro zpracování těchto závislostí je tzv. *metoda nejmenších čtverců* (MNČ - LSM).

Dříve než přistoupíte ke zpracování pomocí MNČ, doporučuji abyste si celou situaci nejprve zevrubně obhlédli, což mj. znamená, že si do nejrůznějších grafů či schémat vynešete vzájemné závislosti všech možných veličin dotyčného objektu, ať už vámi naměřených nebo převzatých z literatury. Věřte, že tyto „obrázky“ vám o povaze vzájemných souvislostí mezi jednotlivými charakteristikami povědí více než sebedokonalejší číselné rozbory. Zjistíte-li, že zobrazené výsledky měření  $\{y_i\}$  jeví jistou časovou závislost, zřejmě též pocítíte neodolatelné nutkání tuto závislost proložit (fit) nějakou elegantní hladkou křivkou. Dříve, než se do toho pustíte, byste ale měli zvážit, zda je to skutečně nezbytné! Chceme-li totiž jen dokumentovat, že tu ona závislost existuje, pak je poctivější do grafu žádnou křivku nevkreslovat, stačí jen zvolit vhodná měřítka na osách a obrázek prezentovat v jeho originální podobě. Pouze tehdy, chceme-li s výsledky proložením dále pracovat a něco z nich vyvozovat, je čas pustit se do matematického zpracování.

### 1.1. Regresní model

Vyšetřujeme nejprve časovou závislost vybrané měřené veličiny  $y$  na základě souboru  $n$  dvojic  $\{t_i, y_i\}$ . Předpokládejme přitom, že čas měření  $t$  známe naprosto přesně, lze jej tedy pokládat za *nezávislou veličinu*, zatímco jednotlivá měření *závisle proměnné veličiny*  $y, y_i$ , jsou zatížena určitou nejistotou, řekněme  $\delta y_i$ .

Naším záměrem nyní bude najít takovou skalární funkci času  $t$ ,  $f(t)$ , která optimálně prochází mezi mezi naměřenými body a co nejlépe vystihuje reálnou časovou závislost pozorované veličiny.

Triviálním řešením této úlohy v případě časové závislosti je pospojování všech po časově sobě následujících bodů lomenou čarou  $\{t_i, y_i\}$ , případně nějakou sice hladkou, ale dostatečně zvlhčenou čarou (např. polynomem  $n-1$  stupně), která by procházela důsledně všemi naměřenými body. Takovýto postup by měl své opodstatnění pouze tehdy, pokud bychom jak čas, tak závisle proměnnou veličinu znali absolutně přesně, což je nereálné. Mnohem hodnověrnější výsledky dává prostá grafická metoda, kdy mezi body vyneseny do grafu táhneme od ruky hladkou křivku, která dle našeho přesvědčení co nejlépe vyjadřuje pozorovanou závislost. Nevýhodou je však to,

že tento způsob proložení není obecně reprodukovatelný (i vy sami ji podruhé proložíte trochu jinak), navíc se s tímto grafickým řešením potom dosti špatně pracuje.

Běžně se dává přednost takovým metodám, které vedou k analytickému vyjádření prokládané funkce a k objektivnímu, reprodukovatelnému stanovení kritéria nejlepší shody. Obvykle si hned na počátku definujeme tzv. *regresní model* (regression model). Regresním modelem si z nekonečného množství funkcí, jimiž by bylo možno pozorovanou závislost proložit, vybereme jen jistou omezenou množinu funkcí, přičemž každá z funkcí této zvolené množiny modelových funkcí bude plně definována  $g$  předem neznámými volnými parametry, které si pracovně označíme  $\beta_1, \beta_2, \beta_3, \dots, \beta_g$ . Veličina  $g$  pak vyjadřuje *počet stupňů volnosti* (degree of freedom) zvoleného modelu. Na tom, jak dobře si umíte vytipovat optimální regresní model, který v sobě obsahuje funkce co nejpodobnější tušené závislosti  $y(t)$ , závisí úspěch nebo neúspěch celého našeho dalšího počínání.

Pokud nevíme o fyzikální podstatě závislosti jedné z pozorovaných veličin na druhé vůbec nic, pak jako regresní model volíme soubor těch co nejjednodušších funkcí - polynomy, harmonické funkce - s nimiž lze snadno pracovat. Nicméně vždy bychom se měli snažit najít takový model, který by skutečnost popisoval s co nejmenším počtem volných parametrů. Pokud však již předem víme, jakou modelovou funkcí by měla být pozorovaná závislost popsána, měli bychom jí dát přednost, protože jinak si způsobíme zbytečné problémy při interpretaci zjištěné závislosti.

Volba odpovídajícího regresního modelu s optimálním stupněm volnosti je tím nejdůležitějším momentem při zpracování, momentem na němž ve značné míře závisí i výsledky a jejich hodnocení. Právě zde se uplatní znalosti, zkušenosti a všeobecný rozhled zpracovávatele, právě tu se projeví jeho vztah k povaze naměřených dat. Správnou a citlivou volbou regresního modelu lze ze souboru dat vytěžit spoustu informací, naopak zvolením neadekvátního modelu, lze snadno dospět i ke zcela mylným a falešným vývodům. Chcete-li mít v tomto oboru dobré výsledky, pak se musíte obrnit značnou dávkou trpělivosti a již předem počítat s tím, že jen zřídka se vám podaří najít ten správný regresní model hned napoprvé. Z vlastní zkušenosti vím, že k některým modelům se člověk propracuje až po letech marných pokusů.

Regresní model představuje množinu podobných funkcí, které se od sebe liší jen pouze jinými hodnotami volných parametrů  $\beta_1, \beta_2, \dots, \beta_g$ :  $f(t) = f(\beta_1, \beta_2, \dots, \beta_g, t)$ . Uspořádanou  $g$ -ticí parametrů  $\beta_j$  je výhodné zapisovat jako  $g$ -rozměrný vektor nebo sloupcovou matici  $\boldsymbol{\beta}$  o rozměrech  $g \times 1$  ( $g$  řádků a 1 sloupec):  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_g)^T$ .

Předpokládejme nyní, že jsme v rámci regresního modelu zvolili nějakou konkrétní hodnotu vektoru parametrů pro  $i$ -té měření  $\{t_i, y_i\}$  pak lze vyjádřit odchylku tohoto měření od dané závislosti  $e_i$  vztahem:

$$y_i = f(t_i, \boldsymbol{\beta}) + e_i. \quad (1)$$

Je zjevné, že čím menší budou odchylky, tím lepší bude proložení pozorované závislosti mezi veličinami  $y$  a  $t$ .

Naším úkolem nyní bude vybrat z množiny funkcí, které připouští zvolený regresní model,  $f(t, \boldsymbol{\beta})$  popsaných vektorem  $\boldsymbol{\beta}$ , najít takový vektor  $\boldsymbol{\beta} = \mathbf{b}$ , pro nějž budou odchylky  $\{e_i\}$  minimální. Onu podmínku minimálnosti je ovšem třeba nejprve matematicky precizovat. Nejčastěji používanou, a z mnoha důvodů nejoblíbenější (nikoli však jedinou<sup>1</sup>), je

---

<sup>1</sup>Jinou takovou podmínkou může být minimálnost součtu absolutních hodnot odchylek nebo jejich čtvrtých mocnin. Nicméně takto definované podmínky se používají jen zřídka, a ve zcela odůvodněných případech. Naopak často se používají jisté modifikace MNC, které dokáží eliminovat hrubé chyby. Těmto modifikacím se pak říká *robustní regrese*.

podmínka, aby součet kvadrátů odchylek pro všechny body měření byl minimální. Z této podmínky pak vychází tzv. metoda nejmenších čtverců<sup>2</sup>, které se budeme nadále věnovat.

## 1.2. Nástin metody nejmenších čtverců

Zaveďme si nejprve skalární veličinu  $S(\boldsymbol{\beta})$ , zvanou též součet čtverců odchylek:

$$S(\boldsymbol{\beta}) = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n [y_i - f(t_i, \boldsymbol{\beta})]^2. \quad (2)$$

Je zřejmé, že suma čtverců odchylek  $S(\boldsymbol{\beta})$  je vždy veličinou nezápornou<sup>3</sup>.

Nyní hledáme takový vektor  $\boldsymbol{\beta}$ , ( $\boldsymbol{\beta} = \mathbf{b}$ ) pro nějž je tato suma  $S(\boldsymbol{\beta} = \mathbf{b})$  minimální. Funkci  $S(\boldsymbol{\beta})$  si můžete představit jako zprohýbanou plochu v  $(g+1)$  rozměrném prostoru, kde  $g$  rozměrů je vyhrazeno pro složky vektoru  $\boldsymbol{\beta}$  a  $(g+1)$ -tý rozměr je rezervován pro funkční hodnotu  $S(\boldsymbol{\beta})$ . Obecně může mít taková plocha dosti komplikovaný vzhled. Nicméně většinou na ní můžeme najít jedno nebo i více lokálních minim, z nichž ovšem jen některá budou mít nějaký dobrý fyzikální smysl.

Pro vyhledávání minim v průběhu funkce několika proměnných je vypracována řada metod, vesměs numerických. V omezeném počtu případů však lze k výsledku dospět i postupy analytické matematiky. Fyzikálně reálné minimum se vyznačuje tím, že funkce  $S(\boldsymbol{\beta})$  v něm je spojitá a spojitě jsou i všechny parciální derivace, které jsou v bodu minima rovny nule. Platí tedy:

$$\left. \frac{\partial S(\boldsymbol{\beta})}{\partial \beta_k} \right|_{\boldsymbol{\beta}=\mathbf{b}} = 0, \text{ pro všechna } k = 1, 2, \dots, g. \quad (3)$$

Dosadíme-li za  $S(\boldsymbol{\beta})$  z rovnic (2 a 3) dostaneme  $g$  rovnic ve tvaru:

$$\sum_{i=1}^n \frac{\partial f(t_i, \mathbf{b})}{\partial \beta_k} f(t_i, \mathbf{b}) = \sum_{i=1}^n \frac{\partial f(t_i, \mathbf{b})}{\partial \beta_k} y_i; \quad (4)$$

Nastane-li pak v bodě  $\mathbf{b}$ , ( $\boldsymbol{\beta} = \mathbf{b}$ ) minimum, pak je splněno všech  $g$  podmínkových rovnic daných výše uvedeným vztahem. Funkce  $f(t, \mathbf{b})$ , nazývaná též *regresní funkce*, je pak onou hledanou funkcí, která představuje nejlepší přiblížení (nebo je jedním z nich) k průběhu funkční závislosti v rámci zadaného modelu závislosti  $y(t)$ .

Ještě poznámku: při hledání extrémů (minima nebo maxima) skalárních funkce je vhodné si zavést pojem *gradient funkce*. Gradient v daném bodě je vektor orientovaný v opačném směru než spádnice, přičemž délka vektoru je tím větší, čím strměji v daném

---

<sup>2</sup>Zde je třeba upozornit, že metodu nejmenších čtverců lze definovat a pojímat mnohem obecněji, než v následujícím textu, který je jenom úvodem do celé problematiky. V následujících kapitolách MNC poněkud zobecníme, tak aby se jí daly řešit i komplikovanější astrofyzikální úlohy.

<sup>3</sup>V reálných případech je to navíc veličina kladná, a to ze dvou důvodů: 1) jen zřídka se nám podaří regresní model vybrat natolik dobře, aby pozorovanou závislost popisoval realisticky v celém rozsahu i v detailech, 2) i kdyby se nám to podařilo, pak je nutno počítat s tím, že závislou veličinu  $y$  neměříme nikdy absolutně přesně. Každé měření je zatíženo chybou měření, u nichž budeme předpokládat, že odchylky jimi způsobené mají náhodné rozložení.

bodě funkce probíhá. Číselně jsou složky vektoru gradientu funkce  $S$ , která je funkcí  $g$  proměnných parametrů, rovny parciálním derivacím podle těchto parametrů:

$$\vec{\nabla} S(\boldsymbol{\beta}) = \left( \frac{\partial S}{\partial \beta_1}, \frac{\partial S}{\partial \beta_2}, \dots, \frac{\partial S}{\partial \beta_g} \right). \quad (5)$$

Gradient lze takto podle potřeby chápat jako buď jako vektor o  $g$  složkách nebo řádkovou matici s  $g$  sloupci. Pomocí gradientu součtu čtverců odchylek lze podmínku pro nalezení minima funkce lze pak elegantně zapsat:

$$\vec{\nabla} S(\mathbf{b}) = \mathbf{0}, \quad (6)$$

kde  $\mathbf{0}$  je řádkový vektor o  $g$  složkách, jež jsou všechny rovny nule. Podmínka tak říká, že minimum skalární funkce nastává v tom bodě, kdy všechny složky gradientu funkce jsou rovny nule. Velikost vektoru gradientu je v tomto bodě nulová, jsme na dně - hlouběji se již dostat nelze. Popisované metodě hledání minima skalární funkce se proto říká též *gradientní metoda* (gradient method).

Dosadíme-li do vektorové rovnice (6) výraz pro sumu odchylek, lze po triviálních úpravách dospět do tvaru:

$$\sum_{i=1}^n \vec{\nabla} f_i [y_i - f_i(\mathbf{b})] = \sum_{i=1}^n \mathbf{x}_i [y_i - f_i(\mathbf{b})] = \mathbf{0}; \quad \mathbf{x}_i = \vec{\nabla} f_i = \left( \frac{\partial f_i}{\partial \beta_1}, \frac{\partial f_i}{\partial \beta_2}, \dots, \frac{\partial f_i}{\partial \beta_g} \right). \quad (7)$$

Vektor příslušný k  $i$ -tému měření  $\mathbf{x}_i$  s  $g$  složkami je tedy gradientem podle složek parametrů prokládané funkce v daném bodě. Složky tohoto vektoru tak lze pokládat za nezávislé proměnné.

Příklad. Prostým příkladem regrese řešené pomocí metody nejmenších čtverců je nalezení střední hodnoty  $n$  naměřených hodnot  $\{y_i\}$  se stejnou (jednotkovou) vahou. Model regresní funkce  $f(t) = \beta$ ,  $\mathbf{x}_i = \vec{\nabla} f_i = (\partial f_i / \partial \beta) = 1$ ,  $S(\beta) = \sum_{i=1}^n (y_i - \beta)^2$ . Minimum funkce  $S(\beta)$  nastává v bodě  $\beta = b$ , v němž platí, že  $\partial S(\beta) / \partial \beta = -2 \sum_{i=1}^n (y_i - b) = 0$ , tedy  $b = \frac{1}{n} \sum_{i=1}^n y_i = \bar{y}$  hledaným středem je aritmetický průměr. Suma kvadrátů odchylek pro  $b = \bar{y}$ ,  $R = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - 2 y_i \bar{y} + \bar{y}^2 = n(\bar{y}^2 - \bar{y}^2)$ . Poučný je i průběh funkce  $S(\beta) = \sum_{i=1}^n (y_i - \beta)^2 = \sum_{i=1}^n y_i^2 - 2 \beta y_i + \beta^2 = R + n(\beta - \bar{y})^2$  - jde o parabolu, křivku s jediným minimem v  $\beta = b = \bar{y}$  s minimální hodnotou  $S(\beta)_{\min} = R = n(\bar{y}^2 - \bar{y}^2)$ .

### 1.3. Nezbytné předpoklady pro použití MNČ. Zavedení vah

Dříve, než se pustíme do hledání regrese, je třeba uvést několik podmínek, které je třeba splnit, abychom mohli metodu nejmenších čtverců použít. Vesměs se týkají souboru odchylek  $\{e_i\}$ . Předně tyto odchylky mají mít statistiku náhodných veličin s centrem  $\mu = 0$ , tedy  $\sum e_i = 0$  a rozptylem  $\sigma$ . Tak tomu ovšem bude jen tehdy, bude-li platit, že

- regresní model je stanoven správně, to znamená, že bude pozorovanou časovou závislost vystihovat jak globálně, tak i v detailech<sup>4</sup>;

<sup>4</sup>To, zda jsme se strefili při volbě regresního modelu nejlépe poznáme, vyneseme-li si závislost jednotlivých odchylek  $e_i$  na čase  $t_i$  (O-C diagram). V případě, že tyto odchylky vykazují systematickou závislost nikoli jenom prostý rozptyl okolo nulové hodnoty, pak je jasné, že jsme model nestanovili správně a je nutné daný model buď modifikovat nebo nemilosrdně opustit.

- soubor nesmí obsahovat žádné odlehlé body (hodnoty s hrubými chyby) ani nesmí být předem neoprávněně upravován vypouštěním bodů, které se jenom zdají být;
- odchylky  $\{e_i\}$  nesmějí být vzájemně korelovány - musí být nezávislé, tedy není přípustné nějaké předběžné vyhlazování<sup>5</sup>.
- všechny naměřené hodnoty musejí mít stejnou nejistotu svého určení;

Zvláště ta poslední z podmínek by dosti omezovala použití metody nejmenších čtverců, neboť rozdílná kvalita měřených veličin je zcela běžná, zejména pokud sledujeme dlouhodobé změny a využíváme k tomu pozorování získaná různými pozorovateli a různými metodami. Naštěstí tuto skutečnost lze elegantně ošetřit zavedením tzv. *váh*  $\{w_i\}$ , které jednotlivým pozorováním přisoudíme.

Předpokládejme nyní, že předem známe nejistotu  $\delta y_i$ , s níž byla daná konkrétní veličina  $y_i$  naměřena. Přejdeme-li teď k modifikovaným odchylkám  $e'_i = e_i/\delta y_i$ , tak by měly mít všechny měřené veličiny stejný rozptyl. Suma kvadrátů modifikovaných odchylek  $S(\boldsymbol{\beta})$  se pak dá zapsat takto:

$$S(\boldsymbol{\beta}) = k \sum_{i=1}^n e_i'^2 = k \sum_{i=1}^n e_i^2 \delta y_i^{-2} = \sum_{i=1}^n e_i^2 w_i = \sum_{i=1}^n [y_i - f(t_i, \boldsymbol{\beta})]^2 w_i, \quad (8)$$

kde je  $w_i$  ona váha určitého pozorování a  $k$  je kladná konstanta, volená zpravidla tak, aby platilo  $\bar{w} = 1$ , takže:

$$w_i = k \delta y_i^{-2} = \frac{n \delta y_i^{-2}}{\sum \delta y_i^{-2}}. \quad (9)$$

Dosadíme-li nyní výraz pro váhovanou sumu čtverců odchylek do (6) a po krátkých úpravách dojdeme k jediné vektorové podmínce:

$$\sum_{i=1}^n \mathbf{x}(t_i, \mathbf{b}) f(t_i, \mathbf{b}) w_i = \sum_{i=1}^n \mathbf{x}(t_i, \mathbf{b}) y_i w_i. \quad (10)$$

Vzhledem k tomu, že váhy by měly být v metodě nejmenších čtverců používány zcela standardně, bude celý další výklad veden v obecnější verzi váhované metody nejmenších čtverců.

Nyní je ovšem třeba vyřešit otázku, jak zjistit nejistotu nalezení měřené veličiny  $\delta y$ , která váhu jednotlivého měření určuje? Je třeba se smířit se skutečností, že onu nejistotu individuálního měření nikdy nedokážeme určit přesně: každé měření je jedinečné, neopakovatelné a nikdy zpětně nebudeme znát všechny okolnosti, které v tu chvíli mohly vlastní měření ovlivnit. Je pravda, že jistým vodítkem může být udávaná vnitřní nejistota (chyba), která ovšem zpravidla představuje jen dolní odhad skutečné nejistoty. Zde je třeba si uvědomit, že ona nejistota by se měla vztahovat k právě použitému regresnímu modelu, který nemusí realitu popisovat ideálně. Jistý odhad ale lze učinit za předpokladu, že přesnost měření v rámci určité relativně homogenní podskupiny dat bude zhruba stejná

---

<sup>5</sup>Nedoporučujeme ani předběžné slučování bodů např. kvůli zpřehlednění probíhajících změn. Vytváření tzv. normálních bodů by mělo následovat regresi.

(např. měření z určité noci v určitém filtru atp.). Tuto přesnost  $\delta y_j$ , danou rozptylem měření podskupiny vzhledem k modelové předpovědi. Z ní lze upřesnit relativní váhy všech měření ve zpracovávaném souboru a celou regresi zopakovat. Po několika iteracích dojdeme k ustálenému stavu, kdy se již výsledky nebudou dále měnit.

## 2 Lineární regrese

Řešení soustavy rovnic daných výše uvedenými vztahy často bývá dosti komplikované a není proto divu, že se již předem hledají takové regresní modely, s nimiž by se dalo zacházet jednodušeji než s obecnými funkcemi. Příjemná práce je s tzv. *lineárními regresními funkcemi*  $f(t, \boldsymbol{\beta})$ , které je možné vyjádřit jako lineární kombinací  $g$  funkcí času  $\{x_1(t), x_2(t), \dots, x_g(t)\}$ , které tvoří vektorovou funkci  $\mathbf{x}(t) = (x_1, x_2, \dots, x_g)$ . Hovoříme pak o lineární regresní funkci nebo o lineárním regresním modelu. Platí tedy:

$$f(t, \boldsymbol{\beta}) = \beta_1 x_1(t) + \beta_2 x_2(t) + \dots + \beta_g x_g(t) = \sum_{j=1}^g \beta_j x_j(t) = \boldsymbol{\beta} \mathbf{x}(t), \quad \Rightarrow \quad (11)$$

$$\vec{\nabla} f(t, \boldsymbol{\beta}) = \left( \frac{\partial f}{\partial \beta_1}, \frac{\partial f}{\partial \beta_2}, \dots, \frac{\partial f}{\partial \beta_g} \right) = \mathbf{x}(t). \quad (12)$$

Dosadíme-li nyní do rovnice (10) za  $f(t, \boldsymbol{\beta})$  dostane:

$$\sum_{i=1}^n \mathbf{x}(t_i) w_i \sum_{j=1}^g b_j x_j(t_i) = \sum_{i=1}^n \mathbf{x}(t_i) y_i w_i, \quad (13)$$

$k$ -tou složku předchozí soustavy rovnic lze po roznásobení sum přepsat do tvaru:

$$\sum_{j=1}^g b_j \sum_{i=1}^n x_k(t_i) x_j(t_i) w_i = \sum_{i=1}^n y_i x_k(t_i) w_i. \quad (14)$$

Celou soustavu  $g$  lineárních rovnic o  $g$  neznámých, jimiž jsou složky hledaného vektoru  $\mathbf{b}$  lze zapsat takto:

$$\begin{aligned} V_{11}b_1 + V_{12}b_2 + \dots + V_{1g}b_g &= U_1 \\ V_{21}b_1 + V_{22}b_2 + \dots + V_{2g}b_g &= U_2 \\ &\vdots \\ V_{g1}b_1 + V_{g2}b_2 + \dots + V_{gg}b_g &= U_g, \end{aligned} \quad (15)$$

kde

$$V_{kj} = V_{jk} = \sum_{i=1}^n x_k(t_i) x_j(t_i) w_i; \quad U_k = \sum_{i=1}^n y_i x_k(t_i) w_i. \quad (16)$$

Soustavu  $g$  rovnic o  $g$  neznámých ( $b_j$ ) pak lze standardním způsobem řešit. Nalezením všech hledaných koeficientů je pak nalezena i regresní funkce, kde  $\boldsymbol{\beta} = \mathbf{b}$ . Pokud nás dále nezajímá přesnost měření, hodnověrnost proložení, chyby parametrů a neurčitost předpovědi, pak jsme hotovi.

## 2.1. Lineární regrese užitím maticového počtu

Lineární regresi lze elegantně řešit použitím maticového počtu. Ten budeme přednostně používat i v následujícím textu.

Pozorovaný vztah mezi závisle proměnnou (nepřesně měřenou veličinou, nejčastěji hvězdnou velikostí, ale i třeba radiální rychlostí, teplotou aj.)  $y$  a nezávislou proměnnou (přesně měřenou veličinou – typicky časem)  $t$  může být proložen vhodnou funkcí **modelovou funkcí**  $f$ . Matematický model závislosti nechť je určen uspořádanou  $g$ -ticí volných parametrů  $\beta_j$ , ve formě sloupcového vektoru  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_g)^T$ . Horní index  $T$  znamená transpozici matice. Pokud je možné modelovou funkcí  $f$  zapsat jako lineární kombinaci  $g$  různých funkcí času  $x_k(t)$  tak hovoříme o tzv. lineární modelové funkci. Pak lze psát:

$$\mathbf{x} = (x_1, x_2, \dots, x_g), \quad f(\mathbf{x}, \boldsymbol{\beta}) = \sum_{k=1}^g \beta_k x_k = \mathbf{x} \boldsymbol{\beta}. \quad (17)$$

Zaveďme sloupcový vektor závislé veličiny  $\mathbf{y}$  s délkou  $n$ , matici  $\mathbf{X}$  s rozměrem  $n \times g$  a sloupcový vektor  $\mathbf{y}$  o délce  $n$ :

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}; \quad \mathbf{X} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1g} \\ x_{21} & x_{22} & \cdots & x_{2g} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{ng} \end{pmatrix}, \quad (18)$$

kde  $y_i$  je hodnota  $i$ -tého pozorování,  $x_{ik}$  je funkční hodnota  $k$ -té funkce pro  $i$ -té pozorování,  $\mathbf{f}(t_i)$  je hodnota vektoru řádkového vektoru definovaného v (17)<sup>6</sup>.

$$\mathbf{W} = \begin{pmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_n \end{pmatrix}; \quad \mathbf{f}(\mathbf{X}, \boldsymbol{\beta}) = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{pmatrix} \boldsymbol{\beta} = \mathbf{X} \boldsymbol{\beta}, \quad (19)$$

kde  $\mathbf{W}$  je diagonální matice  $n \times n$  s vahami jednotlivých měření v diagonále,  $\mathbf{f}(\boldsymbol{\beta})$  je sloupcový vektor s jednotlivými hodnotami modelové funkce  $f_i(\mathbf{x}_i)$  pro  $i$ -té pozorování pro zadané  $\boldsymbol{\beta}$ .

Jako objektivní míru úspěšnosti proložení modelovou funkcí s parametry  $\boldsymbol{\beta}$  použijeme součet váhovaných čtverců odchylek pozorovaných hodnot od předpověděných  $S(\boldsymbol{\beta})$ :

$$S(\boldsymbol{\beta}) = [\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})]^T \mathbf{W} [\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})] = (\mathbf{y}^T - \boldsymbol{\beta}^T \mathbf{X}^T) \mathbf{W} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}) = \mathbf{y}^T \mathbf{W} \mathbf{y} - \boldsymbol{\beta}^T \mathbf{U} - \mathbf{U}^T \boldsymbol{\beta} + \boldsymbol{\beta}^T \mathbf{V} \boldsymbol{\beta} = \mathbf{y}^T \mathbf{W} \mathbf{y} - 2 \boldsymbol{\beta}^T \mathbf{U} + \boldsymbol{\beta}^T \mathbf{V} \boldsymbol{\beta}, \quad (20)$$

<sup>6</sup>Standardně používanými modely lineárních regresních funkcí jsou běžné nebo trigonometrické polynomy vhodných stupňů. Jako příklad lze zvolit parabolický model, jenž je nejjednodušším modelem části světelné křivky s extrémem. Parabolický model lze předpokládat ve formě:  $f(t) = \beta_1 t^2 + \beta_2 t + \beta_3$ ,  $\mathbf{f}(t) = [t^2, t, 1]$ ,  $\mathbf{X} = [\{t_i^2\} \{t_i\} \{1\}]$ .

$\mathbf{U}$  je řádkový vektor s délkou  $g$ ,  $\mathbf{V}$  je čtvercová matice  $g \times g$ , a  $\mathbf{H}$  je její inverzní matice.

$$\mathbf{U} = \mathbf{X}^T \mathbf{W} \mathbf{y}; \quad \mathbf{V} = \mathbf{X}^T \mathbf{W} \mathbf{X}; \quad \mathbf{H} = \mathbf{V}^{-1} = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1}. \quad (21)$$

Metoda nejmenších čtverců (MNČ) modelovou funkcí  $f(t, \boldsymbol{\beta})$  bere za optimální takové proložení, pro něž je suma  $R = S(\boldsymbol{\beta} = \mathbf{b})$  minimální. V případě lineární modelové funkce  $f(t, \boldsymbol{\beta})$  platí, že takové minimum je jen jediné. Pro řešení v podobě sady parametrů  $\mathbf{b}$  a sumu kvadrátů odchylek  $R$  platí:

$$\left. \frac{\partial S}{\partial \boldsymbol{\beta}} \right|_{\boldsymbol{\beta}=\mathbf{b}} = \mathbf{0} = -2\mathbf{U} + 2\mathbf{V}\mathbf{b}, \quad \Rightarrow \quad \mathbf{b} = \mathbf{H}\mathbf{U} = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{y}. \quad (22)$$

Předpověď hodnot modelové lineární funkce pro  $\boldsymbol{\beta} = \mathbf{b}$ ,  $\mathbf{y}_p$  je dána jednoduchým vztahem:

$$\mathbf{y}_p = \mathbf{X}\mathbf{b} = [\mathbf{X}(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}] \mathbf{y} = \boldsymbol{\Xi} \mathbf{y}. \quad (23)$$

Výraz v hranaté závorce:  $\boldsymbol{\Xi}$  je symetrická matice  $n \times n$ , která zde vystupuje jako operátor, který každé hodnotě pozorování přiřadí její „vyhlazenou“ hodnotu. Toto zobrazení je tím věrnější, čím více se matice  $\boldsymbol{\Xi}$  blíží jednotkové matici  $\mathbf{E}(n, n)$ .

Minimální sumu kvadrátů odchylek  $R = S(\boldsymbol{\beta} = \mathbf{b})$  lze pro lineární regresi zapsat různým způsobem:

$$R = (\mathbf{y} - \mathbf{X}\mathbf{b})^T \mathbf{W} (\mathbf{y} - \mathbf{X}\mathbf{b}) = \mathbf{y}^T \mathbf{W} \mathbf{y} - \mathbf{b}^T \mathbf{U} = \mathbf{y}^T \mathbf{W} \mathbf{y} - \mathbf{y}_p^T \mathbf{W} \mathbf{y}_p \quad (24)$$

V posledních dvou variantách vystupuje i váhovaná suma čtverců funkčních hodnot, což je veličina vstupní, vyplývající z pozorování, tudíž zcela nezávislá na modelování. V posledním vztahu se vyskytuje suma váhovaných čtverců předpovědí daných modelem. Metodu nejmenších čtverců tak lze alternativně chápat i jako metodu největších čtverců modelových předpovědí. Tento pohled lze s výhodou využít např. při hledání nejlepších period, tedy při tvorbě LSM periodogramů.

Sumu čtverců odchylek  $S(\boldsymbol{\beta})$  pro lineární regresní model lze po určitých úpravách zapsat v následujícím instruktivním tvaru:

$$S(\boldsymbol{\beta}) = R + \sum_{k=1}^g (\beta_k - b_k)^2 \sum_{i=1}^n w_i x_{ki}^2. \quad (25)$$

Ze zápisu je okamžitě patrné, že funkce  $S(\boldsymbol{\beta})$  má tvar paraboloidu s minimem o hodnotě v bodu  $\boldsymbol{\beta} = \mathbf{b}$ . Má tedy jediné, a tudíž absolutní minimum.

## 2.2. Směrodatná odchylka. Nejistoty

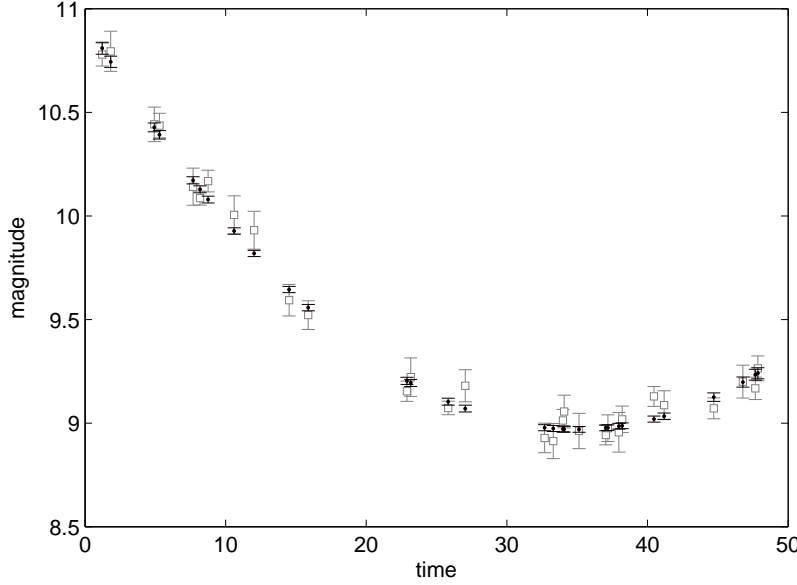
V této kapitole budeme používat normalizované váhy, tedy takové, že  $\bar{w} = 1$ . *Směrodatnou odchylku* (standardní deviace)  $\sigma$  lze odhadnout pomocí vztahů a odhad nejistoty předpovědi jednotlivých vstupních dat  $\delta \mathbf{y}_p$

$$\sigma = \sqrt{\frac{R}{\bar{w}(n-g)}}; \quad \delta \mathbf{y}_p = \sigma \sqrt{\text{diag}(\mathbf{X} \mathbf{H} \mathbf{X}^T)}; \quad \delta \mathbf{b} = \sigma \sqrt{\text{diag}(\mathbf{H})}. \quad (26)$$



Složky sloupcového vektoru  $\delta\mathbf{b}$  se často uvádějí jako rigorózní odhad nejistot jednotlivých parametrů modelu. Bohužel, tento význam mají jen výjimečně, nicméně na nich občas trvají recenzenti odborných článků a oponenti diplomových prací. Naproti tomu velmi cenný je následující odhad předpovědi modelu  $\delta f$

$$\delta f(t, \mathbf{b}) = \sigma \sqrt{\mathbf{x} \mathbf{H} \mathbf{x}^T} = \sigma \sqrt{\vec{\nabla} f \mathbf{H} (\vec{\nabla} f)^T}. \quad (27)$$



**Fig. 1.** Na obrázku jsou kolečky znázorněna simulovaná pozorování proměnné hvězdy v okolí jejího minima jasnosti. Vnitřní přesnost jednotlivých měření je znázorněna šedými chybovými úsečkami. Proložená parabola je naznačena černými tečkami s chybovými úsečkami odpovídajícími nejistotě předpovědi pomocí zvoleného parabolického lineárního modelu.

Odhady nejistoty jednotlivých parametrů obsažených ve vektoru řešení  $\mathbf{b}$ ,  $\delta\mathbf{b}$  se zdají být důležité, neboť pomocí nich lze odhadnout i nejistotu libovolného výrazu  $Q(\boldsymbol{\beta}, t)$ , podle zákona o šíření chyb:  $\delta Q(\boldsymbol{\beta}, t) \simeq \sqrt{\sum_{k=1}^g \left( \frac{\partial Q}{\partial \beta_k} \delta b_k \right)^2}$ . Tak jednoduché to však není. Zmíněný výraz platí jen tehdy, nejsou-li jednotlivé parametry korelované, v opačném případě takto dostaneme jen horní hranici nejistoty. Pokud chcete vzít v potaz možné korelace mezi parametry, pak určitě použijte následující vztah:

$$\delta Q = \sigma \sqrt{\vec{\nabla} Q \mathbf{H} (\vec{\nabla} Q)^T}. \quad (28)$$

Tou funkcí může být i první nebo druhá derivace modelové funkce podle času  $\dot{f}$ ,  $\ddot{f}$ , což jsou veličiny nezbytné např. k výpočtu nejistoty určení okamžiku extrému světelné křivky:

$$\delta \dot{f}(t, \mathbf{b}) = \sigma \sqrt{\vec{\nabla} \dot{f} \mathbf{H} (\vec{\nabla} \dot{f})^T} = \sigma \sqrt{\dot{\mathbf{x}} \mathbf{H} \dot{\mathbf{x}}^T}; \quad (29)$$

$$\delta \ddot{f}(t, \mathbf{b}) = \sigma \sqrt{\vec{\nabla} \ddot{f} \mathbf{H} (\vec{\nabla} \ddot{f})^T} = \sigma \sqrt{\ddot{\mathbf{x}} \mathbf{H} \ddot{\mathbf{x}}^T}, \quad (30)$$

kde  $\dot{\mathbf{x}}(t) = (\dot{x}_1(t), \dot{x}_2(t), \dots, \dot{x}_g(t))$  a  $\ddot{\mathbf{x}}(t) = (\ddot{x}_1(t), \ddot{x}_2(t), \dots, \ddot{x}_g(t))$ .

### 2.3. Základní regresní modely - aplikace lineární regrese

Následuje několik praktických příkladů aplikace lineární regrese metody nejmenších čtverců, které mají ilustrovat způsob, jak lze metodu lineární regrese v maticové podobě používat. Pokud tyto příklady někomu někomu připadnou jako triviální, pak se nemýlí, neboť jde o záměr. Pokud ovšem zvládnete toto, můžete si troufnout na složitější modely.

V řadě příkladů budou s výhodou použity některé střední veličiny, nezávislých i závislých veličin  $t$  a  $y$ :

$$\overline{t^m y^l} = \sum_{i=1}^n t_i^m y_i^l w_i / \sum_{i=1}^n w_i, \quad (31)$$

$$u_{tt} = \overline{t^2} - \bar{t}^2, \quad s_t = \sqrt{u_{tt}}, \quad u_{tt} = \overline{y^2} - \bar{y}^2, \quad s_y = \sqrt{u_{yy}}, \quad u_{ty} = \overline{ty} - \bar{t}\bar{y}, \quad (32)$$

$$r = \frac{\overline{ty} - \bar{t}\bar{y}}{s_t s_y} = \sqrt{\frac{u_{ty}^2}{u_{tt} u_{yy}}} = \frac{u_{ty}}{s_t s_y} \quad (33)$$

Korelační koeficient  $r$  je bezrozměrná veličina nabývající hodnotu mezi -1 a 1, přičemž 0 je roven tehdy, kdy mezi veličinami  $t$  a  $y$  neexistuje žádná lineární korelace,  $\pm 1$  je roven tehdy, kdy jsou všechny hodnoty  $\{t_i, y_i\}$  vyskládaný na jediné přímce.

#### 2.4. Střední hodnota veličin se stejnou vahou

V případě, že mezi  $n$  dvojicemi  $t$  a  $y$ ,  $\{t_i, y_i\}$  neexistuje žádná závislost (korelační koeficient je blízký nule), bude hodnota  $y(t)$  v mezích chyb nejspíš konstantní, postavíme můžeme sestavit regresní model takto:  $y_i = \beta + e_i$ ,  $F(\beta) = \beta$ . Optimální hodnotu  $\beta$ , při níž je suma kvadrátů odchylek  $e_i$  minimální,  $b$ , nazveme střední hodnotou. Můžeme ji najít přímo minimalizací funkcionálu  $S(\beta)$ :

$$S(\beta) = \sum_{i=1}^n (y_i - \beta)^2 = \sum_{i=1}^n y_i^2 - 2\beta \sum_{i=1}^n y_i + n\beta^2 = n(\overline{y^2} - 2\beta\bar{y} + \beta^2). \quad (34)$$

$$\frac{\partial S(\beta)}{\partial \beta} = n(-2\bar{y} + 2b) = 0 \Rightarrow b = \bar{y}; \quad R = \sum_{i=1}^n (y_i - \bar{y})^2 = n(\overline{y^2} - \bar{y}^2) = n u_{yy}, \quad (35)$$

$$S(\beta) = n[\overline{y^2} - \bar{y}^2 + (\beta - b)^2] = R + n(\beta - \bar{y})^2. \quad (36)$$

Grafem funkce  $S(\beta)$  je parabola se středem v  $\beta = \bar{y}$  (viz rov. (36)), přičemž dno paraboly má výšku minimální sumy čtverců odchylek  $R$ .

I když minimalizací funkce  $S(\beta)$  lze střední hodnotu vypočítat přímo, zkusme si nyní ze cvičných důvodů všechny potřebné vztahy odvodit pomocí maticových vztahů.

$$\mathbf{X} = [1, 1, \dots, 1]^T, \quad \mathbf{Y} = [y_1, y_2, \dots, y_n]^T, \quad \mathbf{V} = \mathbf{X}^T \mathbf{X} = n, \quad \mathbf{H} = \mathbf{V}^{-1} = \frac{1}{n}, \quad (37)$$

$$\mathbf{U} = \mathbf{X}^T \mathbf{Y} = \sum_{i=1}^n y_i, \quad b = \mathbf{H}\mathbf{U} = \frac{1}{n} \sum_{i=1}^n y_i = \bar{y}, \quad R = \mathbf{Y}^T \mathbf{Y} - \mathbf{b}^T \mathbf{U} = n(\overline{y^2} - \bar{y}^2); \quad (38)$$

$$\sigma = \sqrt{\frac{R}{n-1}} = s_y \sqrt{\frac{n}{n-1}}, \quad \delta b = \sigma \sqrt{\text{diag}(\mathbf{H})} = \frac{\sigma}{\sqrt{n}} \quad (39)$$

$$\delta \mathbf{Y}_p = \sigma \sqrt{\text{diag}(\mathbf{X} \mathbf{H} \mathbf{X}^T)} = \frac{\sigma}{\sqrt{n}} \mathbf{E}(n, 1), \quad (40)$$

### 2.5. Střední hodnota veličin s nestejnou vahou

Zde lze převzít z předchozí podkapitoly matice  $\mathbf{X}$  a  $\mathbf{Y}$  (viz rov. 37), nově je třeba zavést diagonální matici  $\mathbf{W}$  s vahami  $\{w_i\}$  na diagonále:

$$\mathbf{W} = \text{diag}[w_1, w_2, \dots, w_n]; \quad \mathbf{V} = \mathbf{X}^T \mathbf{W} \mathbf{X} = n \bar{w}; \quad \mathbf{H} = \mathbf{V}^{-1} = \frac{1}{n \bar{w}}, \quad (41)$$

$$\mathbf{U} = \mathbf{X}^T \mathbf{Y} = \sum_{i=1}^n y_i w_i = n \bar{w} \bar{y}, \quad b = \mathbf{H} \mathbf{U} = \frac{1}{n \bar{w}} \sum_{i=1}^n y_i w_i = \bar{y}, \quad (42)$$

$$R = \mathbf{Y}^T \mathbf{Y} - \mathbf{b}^T \mathbf{U} = n \bar{w} u_{yy}; \quad \sigma = \sqrt{\frac{R}{\bar{w}(n-1)}} = s_y \sqrt{\frac{n}{n-1}}, \quad (43)$$

$$\delta b = \sigma \sqrt{\text{diag}(\bar{w} \mathbf{H})} = \frac{\sigma}{\sqrt{n}}, \quad \delta \mathbf{Y}_p = \sigma \sqrt{\bar{w} \text{diag}(\mathbf{X} \mathbf{H} \mathbf{X}^T)} = \frac{\sigma}{\sqrt{n}} \mathbf{E}(n, 1) \quad (44)$$

Za povšimnutí jistě stojí, že vztahy pro  $b, \sigma, \delta b$  a  $\delta \mathbf{Y}_p$  jsou formálně stejné jako v případě bez vah. Rozdíl ovšem je v tom, jak jsou definovány střední veličiny, z nichž se při výpočtu vychází.

### 2.6. Přímka jdoucí počátkem - PMNČ

Občas se můžeme setkat se situací, kdy je jeden nebo více bodů závislosti pevně fixováno. Z této skutečnosti musíme při volbě regresního modelu vycházet. Nejjednodušším příkladem toho druhu je naše očekávání, že  $n$  bodů o souřadnicích  $[t_i, y_i]$  se stejnými vahami lze proložit přímkou jdoucí bodem o souřadnicích  $[0, 0]$ , neboli počátkem. Regresní model je pak:  $y_i = \beta t_i + e_i$ ,  $f(\beta, t) = \beta t$ . Optimální hodnotu  $\beta = b$ , při níž je suma kvadrátů odchylek  $e_i$  minimální, nazveme tentokrát koeficientem úměrnosti.

$$\mathbf{X} = [t_1, t_2, \dots, t_n]^T, \quad \mathbf{y} = [y_1, y_2, \dots, y_n]^T, \quad \mathbf{V} = \mathbf{X}^T \mathbf{W} \mathbf{X} = \sum_{i=1}^n t_i^2 = n \bar{t}^2, \quad (45)$$

$$\mathbf{H} = \mathbf{V}^{-1} = \frac{1}{n \bar{t}^2}, \quad \mathbf{U} = \mathbf{X}^T \mathbf{y} = \sum_{i=1}^n y_i t_i = n \bar{t} \bar{y}, \quad b = \mathbf{H} \mathbf{U} = \frac{\sum_{i=1}^n t_i y_i}{\sum_{i=1}^n t_i^2} = \frac{\bar{t} \bar{y}}{\bar{t}^2}, \quad (46)$$

$$y_p = b t, \quad R = \mathbf{y}^T \mathbf{y} - \mathbf{b}^T \mathbf{U} = n \left( \bar{y}^2 - b \bar{t} \bar{y} \right) = n \left[ y^2 - \frac{(\bar{t} \bar{y})^2}{\bar{t}^2} \right], \quad (47)$$

$$\sigma = \sqrt{\frac{R}{n-1}} = \sqrt{\frac{n \left[ t^2 \bar{y}^2 - (\bar{t} \bar{y})^2 \right]}{(n-1) \bar{t}^2}}, \quad \delta b = \sigma \sqrt{H} = \frac{\sigma}{\sqrt{n \bar{t}^2}}, \quad (48)$$

$$\mathbf{x} = \frac{\partial f}{\partial \beta} = t; \quad \delta y_p = \sigma \sqrt{\mathbf{x}(t) \mathbf{H} \mathbf{x}(t)^T} = \sigma \sqrt{\frac{t^2}{n \bar{t}^2}} \quad (49)$$

Poznámka. Mohlo by vás napadnout vypočítat koeficient úměrnosti jinak, názorněji. Uvažme, že pomocí každé z dvojic  $t_i, y_i$  lze vypočítat „individuální“ koeficient úměrnosti  $b_i : b_i = y_i/t_i$  a střední koeficient úměrnosti, který si nyní označíme  $b'$ , by pak logicky měl být roven aritmetickému průměru jednotlivých  $b_i$ :

$$b' = \frac{1}{n} \sum_{i=1}^n b_i = \frac{1}{n} \sum_{i=1}^n \frac{y_i}{x_i} \quad (50)$$

Jak patrně, tento vztah se od výše uvedeného vztahu (46) pro hodnotu středního koeficientu úměrnosti liší a žádnou z dovolených matematických operací nelze tyto dva vztahy ztotožnit. Vysvětlení diskrepance plyne z předpokladu, na němž je prostá (bez vah) MNČ postavena: rozptyl měření  $y$  od reálného průběhu daného funkční závislosti  $y(x)$  má povahu náhodné veličiny a

zejména nijak nezávisí na hodnotě další měřené veličiny  $x$ . Je-li tato podmínka splněna pro množinu měření  $\{x_i, y_i\}$ , pak ovšem nemůže být splněna pro veličinu  $y_i/x_i$ , jejíž očekávaná nepřesnost je nepřímo úměrná hodnotě  $x_i$ .

Různě velkou očekávanou nepřesnost je třeba v obecné metodě nejmenších čtverců ocenit různým ováhováním bodů, kdy váha jednotlivého bodu -  $w_i$  - bude nepřímo úměrná kvadrátu očekávaného rozptylu měření, čili v tomto případě by měla být úměrná  $x_i^2$ . Položme proto přímo, že  $w_i = x_i^2$ . Aritmetický průměr z množiny  $b_i$  se započítáním jejich individuálních vah se vypočte podle vztahu (42):

$$\bar{b} = \frac{\sum_{i=1}^n b_i w_i}{\sum_{i=1}^n w_i} = \frac{\sum_{i=1}^n t_i y_i}{\sum_{i=1}^n t_i^2} = \frac{\bar{t}y}{\bar{t}^2} = b. \quad (51)$$

### 2.7. Příímka jdoucí počátkem (OMNČ)

Oproti předchozímu případu budeme navíc předpokládat, že každému z bodů měření o souřadnicích  $\{t_i, y_i\}$  bude přisouzena určitá individuální váha  $w_i$ . Přesvědčete se, že pokud budeme počítat se středními váženými veličinami a jejich součiny, pak vztahy pro střední koeficient úměrnosti  $b$ , střední váženou kvadratickou odchylku jednoho měření  $\sigma$ , nejistota koeficientu  $\delta b$  a chyba předpovědi  $\delta y_p$ , budou formálně stejné jako v podkapitole 3.6.

### 2.8. Proložení obecnou příímkou (OMNČ)

Při zpracování časově proměnných pozorovacích dat můžeme často setkat s úlohou nalezení parametrů časové trendu, přičemž se v prvním přiblížení nejčastěji předpokládá, že mezi veličinami závislou veličinou  $y$  a nezávislou veličinou  $t$  (standardně časem měření) existuje lineární závislost. Jinými slovy body v grafu lze proložit příímku. Regresní model pro takovou situaci je zřejmý:  $y_i = \beta_1 + \beta_2 t_i + e_i$ .

Příímka nechť je prokládána  $n$  body o souřadnicích  $[t_i, y_i]$ , přičemž každému z bodů je přisouzena jeho individuální váha  $w_i$ . Řešením úlohy je nalezení vektoru  $\mathbf{b}$  se složkami  $b_1, b_2$ , pro něž je suma váhovaných čtverců odchylek  $S(\beta_1, \beta_2)$  minimální:

$$S(\beta_1, \beta_2) = \sum_{i=1}^n w_i (y_i - \beta_1 - \beta_2 t_i)^2, \quad (52)$$

$$\frac{\partial S}{\partial \beta_1} = -2 \sum_{i=1}^n w_i (y_i - b_1 - b_2 t_i) = 0, \quad \frac{\partial S}{\partial \beta_2} = -2 \sum_{i=1}^n w_i (y_i - b_1 - b_2 t_i) t_i = 0. \quad (53)$$

Soustavu dvou rovnic o dvou neznámých (53) řešíme prostředky maticového počtu:

$$\mathbf{X} = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_n \end{bmatrix}; \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}; \quad \mathbf{W} = \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_n \end{bmatrix}; \quad (54)$$

$$\mathbf{V} = \mathbf{X}^T \mathbf{W} \mathbf{X} = n\bar{w} \begin{bmatrix} 1 & \bar{t} \\ \bar{t} & \bar{t}^2 \end{bmatrix}; \quad \mathbf{U} = \mathbf{X}^T \mathbf{W} \mathbf{y} = n\bar{w} \begin{bmatrix} \bar{y} \\ \bar{t}y \end{bmatrix}; \quad (55)$$

$$\mathbf{H} = \mathbf{V}^{-1} = \frac{1}{n\bar{w}u_{tt}} \begin{bmatrix} \bar{t}^2 & -\bar{t} \\ -\bar{t} & 1 \end{bmatrix}; \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \mathbf{H}\mathbf{U} = \frac{1}{u_{tt}} \begin{bmatrix} \bar{t}^2 \bar{y} - \bar{t} \bar{t}y \\ -\bar{t} \bar{y} + \bar{t}y \end{bmatrix}, \quad (56)$$

$$(57)$$

Přesvědčte se, že platí:  $y_p = \bar{y}$ , tedy že regresní přímka prochází těžištěm.

$$R = \mathbf{y}^T \mathbf{W} \mathbf{y} - \mathbf{b}^T \mathbf{U} = n \bar{w} \left( \bar{y}^2 - b_1 \bar{y} - b_2 \bar{t} \bar{y} \right), \quad \sigma = \sqrt{\frac{R}{\bar{w} (n-2)}}, \quad (58)$$

$$\mathbf{y}_p = \mathbf{X} \mathbf{b}, \quad \delta \mathbf{y}_p = \sigma \sqrt{\bar{w} \text{diag}[\mathbf{X} \mathbf{H} \mathbf{X}^T]}, \quad (59)$$

$$\delta b_2 = \sigma \sqrt{\bar{w} H_{22}} = \frac{\sigma}{s_t \sqrt{n}}, \quad \delta b_1 = \sigma \sqrt{\bar{w} H_{11}} = \frac{\sigma}{s_t} \sqrt{\frac{\bar{t}^2}{n}} = \delta b_2 \sqrt{\bar{t}^2}. \quad (60)$$

Nejistota směrnice přímky  $\delta b_2$  tedy nezávisí na umístění počátku, zatímco chyba absolutního členu  $\delta b_1$  ano. Minimální je tato chyba v případě, kdy počátek souřadnic ztotožníme s těžištěm. Nejistota pak bude  $\delta b_1 = \sigma/\sqrt{n}$ .

Chyba předpovědi funkční hodnoty  $\delta y_p$  v čase  $t$  je dána vztahem:

$$\mathbf{x}(t) = [1, t]; \quad \delta y_p = \sigma \sqrt{\bar{w} [\mathbf{x} \mathbf{H} \mathbf{x}^T]} = \frac{\sigma}{\sqrt{n}} \sqrt{1 + \frac{(t - \bar{t})^2}{s_t^2}}; \quad (61)$$

Vidíme, že chyba předpovědi je minimální v oblasti v těsné blízkosti těžiště, ve velkých vzdálenostech od něj je asymptoticky přímo úměrná této vzdálenosti. Absolutní člen  $b_1$  lze geometricky interpretovat jako úsek na ose  $y$ , který na ní vytíná regresní přímka. Neurčitost polohy tohoto průsečíku udává chyba předpovědi  $\delta y_p(t = 0)$  v bodě 0. Číselně tato chyba je rovna chybě absolutního členu  $\delta b_1$ , tak jak jej uvedeno v (60).

Korelační koeficient  $r$  je dobrou mírou toho, jak dobře právě přímka vystihuje pozorovanou časovou závislost.

$$r = \frac{\overline{ty} - \bar{t}\bar{y}}{s_t s_y} = \frac{u_{ty}}{s_t s_y}. \quad (62)$$

### 2.9. Proložení časových řad polynomem (OMNČ)

Při zpracování delších časových řad pozorovacích dat často aproximujeme pozorovaný vývoj pozorované veličiny  $y$  polynomem  $g - 1$  řádu. Lineární regresní model předpokládáme ve tvaru:  $y_i = \beta_1 + \beta_2 t_i + \dots + \beta_g t_i^{g-1} + e_i$ .

Polynomiální závislost nechť je prokládána  $n$  body o souřadnicích  $[t_i, y_i]$ , přičemž každému z bodů je přisouzena jeho individuální váha  $w_i$ . Řešením úlohy je nalezení sloupcového vektoru  $\mathbf{b}$  se  $g$  složkami  $b_1, b_2, \dots, b_g$ , pro něž je suma váhovaných čtverců odchylek  $S(\beta_1, \beta_2, \dots, \beta_g) = S(\boldsymbol{\beta})$  minimální. Řešíme pomocí maticového počtu. Definice matic  $\mathbf{W}$  a  $\mathbf{y}$  je táž jako v (54), jediný rozdíl je v matici  $\mathbf{X}$ :

$$\mathbf{X} = \begin{bmatrix} 1 & t_1 & t_1^2 & \dots & t_1^{g-1} \\ 1 & t_2 & t_2^2 & \dots & t_2^{g-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_n & t_n^2 & \dots & t_n^{g-1} \end{bmatrix}. \quad (63)$$

Ostatní kroky jsou popsány v textu.

### 2.10. Proložení časových řad harmonickým polynomem (OMNČ)

Řada astrofyzikálních dějů probíhá více či méně periodicky. Známe-li z dřívějšíka parametry periodicity, lze si zavést tzv. *fázovou funkci*  $\vartheta$ , kterou dostanete jako součet běžné fáze  $\varphi$  a epochy

E. Pokud je perioda  $P$  konstantní, lze si fázovou funkci vypočítat jednoduchým vztahem:

$$\vartheta = \frac{t - M_0}{P}, \quad (64)$$

Kde  $t$  je juliánské datum pozorování,  $M_0$  je juliánské datum počátku počítání fázové funkce,  $P$  je fixní perioda ve dnech.

Pozorované periodicky se měnící veličiny (jasnosti, radiální rychlosti, intenzity spektrálních čar, indukce magnetického pole aj.)  $y$  vytvářejí *fázovou křivku*, kterou nejčastěji znázorňujeme jako závislost proměnné veličiny na fázi  $\varphi = \text{frac}(\vartheta)$ . Fázové křivky zpravidla prokládáme harmonickým polynomem stupně  $q = 1/2(g - 1)$ , kde  $g$  je počet stupňů volnosti. Matematický model s harmonickým polynomem stupně  $q$  lze zapsat:  $y_i = \beta_1 + \sum_{k=1}^q \beta_{2k} \cos(2k\pi\vartheta_i) + \beta_{2k+1} \sin(2k\pi\vartheta_i) + e_i$ .<sup>7</sup> Odpovídající matice  $\mathbf{X}$ :

$$\mathbf{X} = \begin{bmatrix} 1 & \cos(2\pi\vartheta_1) & \sin(2\pi\vartheta_1) & \cos(4\pi\vartheta_1) & \sin(4\pi\vartheta_1) & \cdots & \cos(2q\pi\vartheta_1) & \sin(2q\pi\vartheta_1) \\ 1 & \cos(2\pi\vartheta_2) & \sin(2\pi\vartheta_2) & \cos(4\pi\vartheta_2) & \sin(4\pi\vartheta_2) & \cdots & \cos(2q\pi\vartheta_2) & \sin(2q\pi\vartheta_2) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & \cos(2\pi\vartheta_n) & \sin(2\pi\vartheta_n) & \cos(4\pi\vartheta_n) & \sin(4\pi\vartheta_n) & \cdots & \cos(2q\pi\vartheta_n) & \sin(2q\pi\vartheta_n) \end{bmatrix}. \quad (65)$$

### 2.11. Zobecnění lineární regrese I - vektorová závislá proměnná

Občas se stane, že měřená veličina není skalár, ale  $m$ -rozměrný řádkový vektor nebo uspořádaná  $m$ -tice několika veličin:  $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{im})$ , veškerá měření bude představovat matice  $\mathbf{Y}$  s rozměrem  $n \times m$ . Řešením regrese bude

$$\mathbf{f}(t, \mathbf{B}) = \beta_1 x_1(t) + \beta_2 x_2(t) + \dots + \beta_g x_g(t) = \sum_{j=1}^g \beta_j x_j(t) = \mathbf{x}(t) \mathbf{B}, \quad (66)$$

$$\mathbf{Y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_n \end{pmatrix} = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1m} \\ y_{21} & y_{22} & \cdots & y_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nm} \end{pmatrix}; \quad \mathbf{F}(\mathbf{B}) = \begin{pmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \vdots \\ \mathbf{f}_n \end{pmatrix} = \mathbf{X} \mathbf{B} \quad (67)$$

$$\mathbf{U} = \mathbf{X}^T \mathbf{W} \mathbf{Y}; \quad \mathbf{B} = \mathbf{H} \mathbf{U} = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{Y}; \quad \mathbf{y}_p = \mathbf{x}(t) \mathbf{B}. \quad (68)$$

### 2.12. Zobecnění lineární regrese II - více nezávisle proměnných

Až doposud jsme jako jedinou nezávislou proměnnou brali čas a vše jsme nahlíželi z prizmatu časové proměnnosti. Složky vektoru  $\mathbf{x} = (x_1, x_2, \dots, x_g)$  pak byly funkcemi času. To však vůbec není nutné. Jednotlivé položky mohou být sice funkcemi času, ale nemusejí, mohou

<sup>7</sup>Zde je třeba mít na paměti skutečnost, že fázová funkce je funkcí periody, která se může v průběhu času měnit. Úlohu, kde bychom kromě tvaru světelné křivky řešili i časový vývoj periody, lze zvládnout až prostředky nelineární regrese.

třeba funkcemi prostorových souřadnic, rychlosti nebo to mohou být jen indikace popisující povahu měření (zda šlo třeba o fotometrické měření, či měření radiálních rychlostí nebo intenzity spektrálních čar). Vše to jsou nezávislé, nenáhodné veličiny charakterizující konkrétní měření v rámci zvoleného komplexního modelu. Proto má smysl dívat se na celý soubor veličin obsažených ve vektoru  $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ig})$  přímo jako na soubor  $g$  nezávislých veličin, které mohou nabývat různých hodnot. Pro určitý typ měření mohou být některé z nezávislých proměnných rovny 0, pro jiný typ měření mohou být nulové jiné nezávislé proměnné. Ve vektoru  $\mathbf{y}_i = (y_1, y_2, \dots, y_n)^T$  s naměřenými veličinami jsou pak jednotlivé položky řazeny často v pořadí, v jakém byly naměřeny.

**Příklad:** Takovým lineárním modelem může být funkce se dvěma stupni volnosti popisující měření šířky a délky nějakého obdélníku. V případě, že v  $i$ -tém měření měříme šířku, je  $\mathbf{x}_i = (0, 1)$ , jde-li naopak o měření délky, pak je  $\mathbf{x}_i = (1, 0)$ ,  $y_i$  je ona naměřená veličina. Modelová funkce pro  $i$ -té měření pro  $f_i = \beta_1 x_{i1} + \beta_2 x_{i2} = \mathbf{x}_i \boldsymbol{\beta}$ ,  $\beta_1$  je délka,  $\beta_2$  je šířka. Cílem zpracování je najít střední velikost těchto parametrů  $\mathbf{b}$  na základě  $n$  měření. Při výpočtu budeme předpokládat, že váhy všech měření jsou jednotkové - tedy že je měříme se stejnou chybou.

$$\begin{array}{l}
 3.16 \\
 2.15 \\
 2.18 \\
 3.13 \\
 2.15 \\
 2.19 \\
 3.13
 \end{array}
 \left| \begin{array}{l}
 s \\
 d \\
 d \\
 s \\
 d \\
 d \\
 s
 \end{array} \right.
 \mathbf{y} = \begin{pmatrix} 3.16 \\ 2.15 \\ 2.18 \\ 3.13 \\ 2.15 \\ 2.19 \\ 3.13 \end{pmatrix} ; \mathbf{X} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} ; \mathbf{H} = \begin{pmatrix} \frac{1}{4} & 0 \\ 0 & \frac{1}{3} \end{pmatrix} ; \mathbf{b} = \begin{array}{l} 2.168 \pm 0.009 \\ 3.140 \pm 0.010 \end{array} . \quad (69)$$

Výhodou tohoto přístupu je, že můžeme solidně odhadnout směrodatnou odchylku a tedy i nejistotu určení hledané délky a šířky. Vzhledem k tomuto zobecnění se takto mohou pod sebe dostat i velmi odlišné typy měření s velmi odlišným rozsahem měřených veličin. Proto je důležité, aby byly jednotlivé typy měření správně oceněny svou vahou  $w_i$  nepřímo úměrnou své disperzi.

### 2.13. Nalezení okamžiku minima ze dvou sad pozorování - domácí úloha

Cílem této domácí úlohy je aplikace zobecněné lineární regrese na problému, který simuluje situaci, do níž se pozorovatele proměnných hvězd často dostávají.

Představme si, že dva pozorovatelé v odlišných časových pásmech spolupracovali při pozorování minima jasnosti určité dlouperiodické proměnné hvězdy, přičemž spolupracujícímu Číňanovi ( $q = 1$ ) se podařilo provést celkem 15 pozorování, vesměs na sestupné větvi. Český pozorovatel ( $q = 2$ ) zachytil až výstup světelné křivky z minima v 30 pozorováních ovšem s poněkud horší kvalitou. Samotné minimum žádný z pozorovatelů nezachytil.

V obou případech se pozorování vedla ve filtru  $V$ , hvězdné velikosti se vztahovala k vybrané srovnávací hvězdě, pozorovatelé se však neshodli na její volbě, takže světelné křivky na sebe nenavazovaly. Světelné křivky byla simulována parabolou:

$$\Delta m(t) = a_1 (t - t_{\min})^2 + a_5 \delta_{i1} + a_6 \delta_{i2} = a_1 t^2 + a_2 t + a_3 \delta_{i1} + a_4 \delta_{i2}, \quad t_{\min} = -\frac{a_2}{2a_1}, \quad (70)$$

kde  $a_1$  je koeficient parabolického členu (pro simulaci zvoleno  $a_1 = 1$ ),  $t_{\min}$  je okamžik minima (zvoleno  $t_{\min} = 0,350$ ),  $a_5, a_6$  jsou rozdíly hvězdné velikosti v minimu jasnosti pro čínského a

českého pozorovatele (zvoleno  $a_5 = 0,000$ ,  $a_6 = 0,400$ ). Funkce  $\delta_{i1} = 1$ , pokud jde o pozorování Číňana, jinak  $\delta_{i1} = 0$ , naproti tomu  $\delta_{i2} = 1$ , pokud jde o pozorování Čecha, jinak  $\delta_{i2} = 0$ .  $a_2$  je lineární člen,  $a_3$ ,  $a_4$  jsou hodnoty  $\Delta m(t = 0)$  pro jednotlivé pozorovatele. Okamžiky pozorování jsou udávány ve dnech od začátku určitého juliánského dne. Jednotlivé okamžiky  $t_i$  byly voleny náhodně v intervalu 0 až 0,3 ( $q = 1$ ) a 0,4 až 0,8 ( $q = 2$ ). K simulovaným hodnotám rozdílu hvězdné velikosti  $\Delta m(t_i)$  určeným vztahem (70) pro dané hodnoty časů  $t_i$  byl přičten náhodný gaussovský šum o standardních odchylkách postupně:  $s_1 = 0.005$  mag a  $s_2 = 0.007$  mag. Tabulka s takto nasimulovanými časy  $t_i$  a hodnotami  $\Delta m(t_i)$  včetně příznaku  $q$  následuje.

$t_i$	$\Delta m_i$	$q$	$t_i$	$\Delta m_i$	$q$	$t_i$	$\Delta m_i$	$q$
0,013	0,117	1	0,428	-0,037	2	0,596	0,014	2
0,039	0,093	1	0,455	-0,035	2	0,609	0,015	2
0,053	0,086	1	0,473	-0,042	2	0,623	0,026	2
0,100	0,058	1	0,486	-0,036	2	0,623	0,002	2
0,112	0,054	1	0,488	-0,031	2	0,634	0,033	2
0,114	0,055	1	0,489	-0,024	2	0,672	0,049	2
0,120	0,056	1	0,502	-0,035	2	0,672	0,056	2
0,131	0,041	1	0,502	-0,032	2	0,681	0,063	2
0,132	0,051	1	0,543	-0,017	2	0,697	0,086	2
0,206	0,014	1	0,549	-0,005	2	0,739	0,102	2
0,220	0,020	1	0,561	0,005	2	0,740	0,095	2
0,248	0,019	1	0,568	-0,005	2	0,743	0,097	2
0,252	0,006	1	0,572	0,006	2	0,743	0,101	2
0,264	0,005	1	0,573	0,005	2	0,761	0,123	2
0,294	-0,006	1	0,587	0,007	2	0,772	0,133	2

Vaším úkolem bude:

- Nakreslit graf pozorovaných světelných křivek.
- Pomocí prosté lineární regrese bez vah vypočítat zvlášť pro 1. a 2. sadu pozorování hodnotu koeficientů  $a_1$ ,  $a_2$ ,  $a_3$ , případně  $a_4$ , včetně odhadu jejich nejistot, hodnoty standardní odchylky. Výsledné hodnoty mezi sebou porovnejte a srovnajte je se zadanými parametry simulace.
- Vypočítat okamžiky  $t_{\min}$ , včetně nejistoty jejich určení, přičemž využijete vztah uvedený v (70) a vztah pro výpočet odhadu chyby funkce koeficientů (28) a funkční hodnotu v minimu proložené paraboly  $a_5$  a  $a_6$ , včetně nejistoty. Výsledné hodnoty mezi sebou porovnejte a srovnajte je se zadanými parametry simulace.
- Spojte obě pozorování dohromady a předpokládejte, že absolutní členy lineární regrese jsou různé. Předpokládejte nejprve, že váhy všech pozorování jsou identické, rovné 1. Vypočtete koeficienty  $a_1$ ,  $a_2$ ,  $a_3$ ,  $a_4$ , včetně odhadu jejich nejistot, hodnotu standardní odchylky. Výsledné hodnoty mezi sebou porovnejte a srovnajte je se zadanými parametry simulace.



- Vypočítejte okamžik  $t_{\min}$ , včetně nejistoty jeho určení, přičemž využijete vztah uvedený v (70), vztah pro výpočet odhadu chyby funkce koeficientů (28) a funkční hodnotu v minimu proložené paraboly  $a_5$  a  $a_6$ , včetně nejistoty. Výsledné hodnoty mezi sebou porovnejte a srovnajte je se zadanými parametry simulace.
- Vypočítejte standardní odchylky vzhledem k předpovědi vůči tomuto modelu zvlášť pro čínské a české pozorování. Pomocí nich vypočtete normalizovanou váhu jednotlivých čínských a českých pozorování. S těmito vahami pak opakujte výpočet parametrů  $a_1$ ,  $a_2$ ,  $a_3$ ,  $a_4$ , včetně odhadu jejich nejistot, hodnotu standardní odchylky. Výsledné hodnoty mezi sebou porovnejte a srovnajte je se zadanými parametry simulace.
- Vypočítejte okamžik  $t_{\min}$ , včetně nejistoty jeho určení, a funkční hodnotu v minimu proložené paraboly  $a_5$  a  $a_6$ , včetně nejistoty. Výsledné hodnoty mezi sebou porovnejte a srovnajte je se zadanými parametry simulace.
- Pro spojené sady pozorování předpovězte funkční hodnoty a jejich nejistoty pro obě sady pozorování. Diskutujte, vynesete do grafu.

### 3 Nelineární regrese

Je třeba se smířit se skutečností, že naprostá většina regresních modelů dobře popisujících reálné astrofyzikální situace prostě není lineární a ani ji nelze na lineární převést. Řešení nelineární regrese už není tak přímočaré, mj. i proto, že taková regrese může mít i více řešení, z nichž jen některá jsou fyzikálně přijatelná. Nicméně, lze ukázat, že ve většině fyzikálně akceptovatelných řešení lze v okolí minima plochu sumy kvadrátů  $S(\boldsymbol{\beta})$  nahradit paraboloidem - lze tedy nelineární model v okolí minima nahradit jeho linearizovanou aproximací. K tomu ovšem musíme mít dobrý odhad řešení, v jehož okolí budeme skutečně řešení hledat. K odhadu se lze dopracovat třeba použitím údajů z literatury spolu se zjednodušením modelu, tak abyste pomocí něj k odhadu řešení dospěli.

#### 3.1. Linearizace nelineárních regresních modelů

Připusťme nyní, že se nám podařilo se k takovému odhadu v podobě výchozího vektoru parametrů  $\mathbf{b}_0$  dopídit. Minimum pak bude hledat v bezprostředním okolí tohoto startovního odhadu. Jakmile se nám podaří výchozí regresní model linearizovat, hned se můžeme začít těšit se z vymožeností poskytovaných lineární regresí.

Při linearizaci modelu zpravidla používáme jeho Taylorův rozvoj prvního řádu podle parametrů, v nichž je model nelineární.

$$f(\mathbf{b}_0, \Delta\boldsymbol{\beta}) \cong f(\mathbf{b}_0) + \sum_{k=1}^g \Delta\beta_k \frac{\partial f(\mathbf{b}_0)}{\partial \beta_k} = f(\mathbf{b}_0) + \Delta\boldsymbol{\beta} \mathbf{x}, \quad (71)$$

kde Takto přepsaná modelová funkce je pak lineární vzhledem k nově zavedeným parametrům  $\Delta\boldsymbol{\beta}$ , přičemž vektor nezávisle proměnných  $\mathbf{x}$  je dán vztahem:

$$\mathbf{x} = \vec{\nabla} f = \left( \frac{\partial f}{\partial \beta_1}, \frac{\partial f}{\partial \beta_2}, \dots, \frac{\partial f}{\partial \beta_g} \right). \quad (72)$$

Vidíme, že situace je velmi podobná té, co známe u lineární regrese, rozdíly tu ale jsou, a to významné. Vektor  $\mathbf{x}_i$  příslušející  $i$ -tému měření je opět gradientem, a proto se této metodě řešení nelineární regrese také říká metoda gradientní. Z vektorů  $\mathbf{x}_i$  si vytvoříme matici  $\mathbf{X}$  (viz rovnice (18)),  $\Delta\mathbf{y} = \mathbf{y} - \mathbf{f}$ ,  $\mathbf{W}$  a vypočítat řešení v podobě diferenčního vektoru  $\Delta\mathbf{b}$ .

$$\mathbf{V} = \mathbf{X}^T \mathbf{W} \mathbf{X}; \quad \mathbf{U} = \mathbf{X}^T \mathbf{W} \Delta\mathbf{y}; \quad \mathbf{H} = \mathbf{V}^{-1}; \quad \Delta\mathbf{b} = \mathbf{H} \mathbf{U}. \quad (73)$$

Tento korigující vektor přičteme k počátečnímu odhadu vektoru volných parametrů  $\mathbf{a}_0$  modelové funkce a obdržíme tak další, zlepšený odhad uspořádané  $g$ -tice parametrů  $\mathbf{b}_1 = \mathbf{b}_0 + \Delta\mathbf{b}$ . S novou hodnotou parametrů pak můžeme celý popsany postup znovu opakovat. V průběhu iterativního procesu se absolutní velikost vektoru  $\Delta\mathbf{b}$  zpravidla rychle zmenšuje a již po několika krocích se přiblíží nule, což znamená, že jsme již našli hledané řešení celé úlohy.

A ještě poznámka: není třeba linearizovat všechny parametry, některé z nich bývají lineární a lze je tak počítat přímo. Lze to ukázat na následujícím příkladu.

### 3.2. Příklad – linearizace parabolického modelu

Nyní můžeme linearizovat náš kvadratický model podle rovnice (71), vycházejíce z našeho počátečního odhadu parametrů  $\mathbf{a}_0$

$$f = [a_{02} (t - a_{01})^2 + a_{03}] + \Delta a_1 2 a_{02} (t - a_{01}) + \Delta a_2 (t - a_{01})^2 + \Delta a_3. \quad (74)$$

Je zjevné, že modelová funkce je v parametrech  $a_2, a_3$  lineární. Znamená to, že tyto dva parametry není třeba linearizovat, ale počítat přímo, pouze první parametr,  $a_1$  je třeba hledat iterativně, tedy:

$$f(t, \mathbf{a}) = \Delta a_1 2 a_2 (t - a_{01}) + a_2 (t - a_{01})^2 + a_3. \quad (75)$$

$$\mathbf{X} = \begin{pmatrix} 2 a_2 (t_1 - a_{01}) & (t_1 - a_{01})^2 & 1 \\ 2 a_2 (t_2 - a_{01}) & (t_2 - a_{01})^2 & 1 \\ \vdots & \vdots & \vdots \\ 2 a_2 (t_n - a_{01}) & (t_n - a_{01})^2 & 1 \end{pmatrix} \quad (76)$$

Parametry  $a_2, a_3$  popisující vzhled světelné křivky neiterujeme.

### 3.3. Odhad nejistoty okamžiků extrémů

Nejistota volných parametrů včetně okamžiků extrémů je dána vztahem

$$\Delta\mathbf{y} = \mathbf{y} - \mathbf{f}(t, \mathbf{b}); \quad R = \Delta\mathbf{y}^T \mathbf{W} \Delta\mathbf{y}; \quad \sigma = \sqrt{R/(n - g)}; \quad \delta\mathbf{b} = \sigma \sqrt{\text{diag}(\mathbf{H})}. \quad (77)$$

Jakkoli je výsledný soubor korekčních parametrů téměř čistě nulový, jejich nejistota nulová není a odpovídá nejistotě jednotlivých parametrů. To nám umožňuje učinit spolehlivý odhad neurčitost, s níž známe okamžik, kdy proložená funkce nabývá svého extrému.

## Reference