



**MATEMATICKO-FYZIKÁLNÍ  
FAKULTA**  
Univerzita Karlova

**BAKALÁŘSKÁ PRÁCE**

Šimon Vojtek

**Řešení soustav lineárních rovnic v  
nestandardních podmínkách**

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: prof. RNDr. Ivan Mizera, CSc

Studijní program: Obecná matematika

Studijní obor: Matematika

Praha 2024

Prohlašuji, že jsem tuto bakalářskou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů. Tato práce nebyla využita k získání jiného nebo stejného titulu.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V ..... dne .....

Podpis autora

Chtěl bych poděkovat profesoru Mizerovi za vedení mé práce a dávání cenných rad.

Taky bych chtěl poděkovat panu Žežulovi za trpělivost a jeho nadměrné schopnosti s Overleafem.

Název práce: Řešení soustav lineárních rovnic v nestandardních podmínkách

Autor: Šimon Vojtek

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: prof. RNDr. Ivan Mizera, CSc, Katedra pravděpodobnosti a matematické statistiky

Abstrakt: V práci jsme se zabírali soustavou lineárních rovnic a její souvislostí s reprezentací dat. Problém jsme rozdělili na přeурčené a podurčené soustavy. U přeурčených jsme zmínili MNČ s QR rozkladem, dále metodu Theila a Sena a minimalizaci  $\ell_1$ -normy za pomoci lineárního programování. U podurčené soustavy jsme aplikovali MNČ. Pro způsob výběru jednoho řešení jsme použili metodu Lasso nebo hřebenovou regresi. Alternativním způsobem jsou krokové metody nebo nejlepší výběrová regrese.

Klíčová slova: Metoda nejmenších čtverců (MNČ), Unitární matice, Norma, Reziduální součet čtverců (RSC)

Title: Solutions of the systems of linear equations in nonstandard conditions

Author: Šimon Vojtek

Department: Department of Probability and Mathematical Statistics

Supervisor: prof. RNDr. Ivan Mizera, CSc, Department of Probability and Mathematical Statistics

Abstract: In this thesis, we were considering the similarities between data representation and system of linear equation. We divided the problem into overdetermined and underdetermined system. For overdetermined systems, We mentioned OLS method with QR decomposition, Theil-Sen estimator and minimizing the  $\ell_1$ -norm using linear programming. For underdetermined systems, we applied OLS method. To select one unique solution, also we use Lasso method or ridge regression. An alternative approach includes stepwise methods or best subset regression.

Keywords: Ordinary least square (OLS), Unitary matrix, Norm, Residual sum square (RSS)

# Obsah

Úvod	2
1 Předmluva	3
2 Řešení přeúřčených soustav lineárních rovnic	4
2.1 Metoda nejmenších čtverců (MNČ)	4
2.2 MNČ za pomoci QR rozkladu	4
2.2.1 MNČ pro singulární rozklad	5
2.2.2 QR rozklad s permutací sloupců	6
2.3 Lineární regrese	7
2.3.1 MNČ a nejlepší nestranný lineární odhad $x_0$	8
2.3.2 Příklady lineární regrese	10
2.4 Nelineární regrese	11
2.4.1 Linearizace problému	12
2.4.2 Příklad nelineární regrese	13
2.5 Minimalizace $\ell_1$ -normy a metoda Theila a Sena	14
2.5.1 Jednoduchá lineární regrese podle Theila a Sena	14
2.5.2 Lineární regrese pro $\ell_1$ -normu	14
3 Podurčené/Nedourčené soustavy rovnic	16
3.1 Minimalizace $\ell_2$ -normy	16
3.2 Selekcce lineárních modelů	17
3.2.1 Nejlepší výběrová regrese	18
3.2.2 Kroková regrese	19
3.2.3 Zpětná kroková regrese	19
3.3 Shrinkage metody	19
3.3.1 Hřebenová regrese	20
3.3.2 Lasso (Least Absolute Shrinkage and Selection Operator)	21
3.3.3 Geometrická interpretace Lassa a hřebenové regrese	22
Závěr	24
Seznam použité literatury	25

# Úvod

V této bakalářské práci nejdříve v předmluvě objasníme základní pojmy s triviálními příklady.

Pak si ukážeme viz (Saad a Schultz, 1986) různé metody (MNČ později minimalizace  $\ell_1$ -normy), jak z obsáhlých souborů dat získat nějakou funkci, která bude dobře predikovat výsledky. Následně přejdeme k aplikaci QR rozkladu v MNČ.

Získané vědomosti pak využijeme v podkapitole lineární regrese viz (Yan a Su, 2009), kde budeme prokládat body lineární kombinací funkcí a ukážeme, že za určitých předpokladů je odhad získaný z MNČ nejlepší nestranný lineární odhad, kde jsem velmi důkladně rozepsal nejasné části viz (Baum, 2004) a doplnil veškeré předpoklady viz (Hill a kol., 2011) a nejasnosti v definicích doplněné z (Bhattacharya a Burman, 2016). Poté aplikujeme metody na vytvořené příklady.

Následně nahlédneme do nelineární regrese s využitím iterování Newtonovou metodou viz (Fletcher, 1987). Ukážeme, že různé nelineární problémy lze převést na lineární, což velmi zjednodušuje výpočet.

Dále přejdeme k jiné metodě pro jednoduchou regresi, přesněji k Theilově a Senově metodě viz (Birkes a Dodge, 1993), zmíníme výhody a nevýhody oproti  $\ell_2$ -normě. Ke konci kapitoly zmíníme spojitost mezi lineární regresí pro  $\ell_1$ -normou a lineárním programováním viz (Arenales a kol., 2015).

V poslední kapitole se budeme zabývat případy, kde počet parametrů je větší než počet naměřených dat viz (James a kol., 2013). Opět využijeme MNČ a nastíníme problémy, které mohou nastat v takové situaci a jejich řešení. Řešením může být jeden ze způsobů podle  $\ell_2$ -normy, kde dostaneme nekonečně mnoho řešení a zde budu dokazovat, že existuje právě jedno řešení, které má nejmenší  $\ell_2$ -normu. Nadále se budeme věnovat selekcím lineárních modelů, zmíníme tři metody a vysvětlím v jakých situacích určitou metodu upřednostnit. V závěru této kapitoly se budeme zabírat eliminací parametrů, na kterých naměřené hodnoty nezávisí vůbec nebo velmi málo. Hlavní dvě metody, zmíněné v kapitole jsou hřebenová regrese a Lasso, kde si nejdřív ukážeme, že je lze ekvivalentně přeformulovat, což nám usnadní vizualizaci problému a to předvedu na vytvořených obrázcích. Ve zbytku kapitoly budeme řešit rozdíly a neintuitivní důvod, proč je Lasso lepší, pro eliminaci parametrů.

# 1. Předmluva

V této bakalářské práci se budeme zabývat, hledáním ideálního řešení soustav lineárních rovnic. Budeme rozlišovat tři případy určené, přeурčené a podurčené.

**Definice 1.** *Určeným nebo-li determinovaným případem nazveme soustavu s  $n$  rovnic a  $n$  neznámých, pro kterou existuje právě jedno řešení.*

**Příklad 1.** Ukážeme jednoduchou soustavu:

$$\begin{aligned}x + y &= 1 \\2x + y &= 1.\end{aligned}$$

Gaussovou eliminací dostáváme, že  $x = 0$   $y = 1$  je jediné řešení. Proto určené případy nejsou nijak zajímavé a tedy nadále se tím nebudeme zabývat.

**Definice 2.** *Soustavu nazýváme přeурčenou, pokud počet neznámých je menší než počet rovnic. Rozeznáváme dva případy.*

1. *Soustava je konzistentní, když existuje řešení, neboli hodnota matice reprezentující soustavu je stejná jako hodnota rozšířené (Frobeniova věta).*
2. *Soustava je nekonzistentní, pokud žádné řešení nemá, neboli hodnota matice reprezentující soustavu je různá od hodnoty rozšířené.*

Ukážeme příklady dvou soustav:

**Příklad 2. (konzistentní)**

**Příklad 3. (nekonzistentní)**

$$\begin{aligned}x + y &= 1 \\2x + y &= 1 \\2x + 2y &= 2,\end{aligned}$$

$$\begin{aligned}x + y &= 1 \\2x + y &= 1 \\2x + 2y &= 0.\end{aligned}$$

Řešení v příkladu 2. je stejné jako pro příklad 1. a pro příklad 3. žádné řešení neexistuje.

**Definice 3.** *Podurčenou soustavu nazveme takovou, kde počet neznámých je větší než počet rovnic. Opět rozeznáváme konzistentní a nekonzistentní případy.*

**Příklad 4. (konzistentní)** Ukážeme jednoduchou soustavu:

$$\begin{aligned}x + y + z &= 1 \\2x + 2y + 2z &= 2,\end{aligned}$$

která má bázi množiny řešení ve tvaru  $((1, -1, 1)^T, (1, 1, -1)^T)$ .

**Příklad 5. (nekonzistentní)** Ukážeme soustavu:

$$\begin{aligned}x + y + z &= 1 \\2x + 2y + 2z &= 0,\end{aligned}$$

která žádné řešení nemá.

Pro případy bez řešení se dostáváme k otázce najít takové řešení, které uspokojí soustavu "nejvíc, jak je to možné" a pro případy, kde je těchto řešení více než jedno se pokusíme množinu řešení zmenšit.

## 2. Řešení přeurlčených soustav lineárních rovnic

Soustavu rovnic budeme prezentovat pomocí maticového zápisu  $Ax = b$ , kde  $A \in \mathbb{R}^{m \times n}$ ,  $m > n$ ,  $n, m \in \mathbb{N}$ ,  $b \in \mathbb{R}^m$  a  $x = (x_1, \dots, x_n)$ , pokud jde o konzistentní soustavu, tak existuje  $x \in \mathbb{R}^n$ , které je řešením a v opačném případě hledáme nějaké  $x \in \mathbb{R}^n$ , pro které platí  $Ax \approx b$  ve smyslu minimalizace normy  $\|Ax - b\|$ .

### 2.1 Metoda nejmenších čtverců (MNČ)

Pro MNČ se snažíme minimalizovat nezáporný výraz

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2 \iff \min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2. \quad (2.1)$$

Tyto dva výroky jsou ekvivalentní, protože norma je nezáporné zobrazení.

**Tvrzení 1.** *Mějme přeurlčenou soustavu lineárních rovnic s reprezentující maticí  $A$  s **nezávislými sloupci**, pak  $x_{MNČ}$  řešící (2.1) je ve tvaru*

$$x_{MNČ} = (A^T A)^{-1} A^T b.$$

*Důkaz.* Chceme najít minimální  $x$ , které jednoduše najdeme pomocí derivace podle  $x$

$$\begin{aligned} (\|Ax - b\|_2^2)' &= ((Ax - b)^T (Ax - b))' \\ &= (x^T A^T Ax - x^T A^T b - b^T Ax + b^T b)' \\ &= 2A^T Ax - 2A^T b. \end{aligned}$$

Položíme derivaci rovno nule

$$A^T Ax - A^T b = 0,$$

kde ekvivalentní úpravou dostaneme

$$x = (A^T A)^{-1} A^T b. \quad \square$$

Problémem v praxi je, že z numerického hlediska hledat inverzní matici k  $A^T A$  je mnohdy velmi náročné a vzniká tak chyba v přesnosti řešení, a proto se využívají jiné metody i přesto, že teorie vychází jasně a jednoduše.

### 2.2 MNČ za pomoci QR rozkladu

V této kapitole využíváme poznatky z práce (Saad a Schultz, 1986) týkající se QR rozkladu.



**Definice 4.** Necht  $A \in \mathbb{R}^{m \times n}$ , **QR-rozklad** nazveme

$$A = QR,$$

kde  $Q \in \mathbb{R}^{m \times m}$  je ortonormální matice a  $R \in \mathbb{R}^{m \times n}$  je matice v horním trojúhelníkovém tvaru.

Pomocí QR rozkladu si můžeme (2.1) velmi zjednodušit.

**Tvrzení 2.** Mějme přeuročenu soustavu lineárních rovnic s reprezentující maticí  $A \in \mathbb{R}^{m \times n}$  s **lineárně nezávislými sloupci** a  $b \in \mathbb{R}^m$ , pak

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2 = \min_{x \in \mathbb{R}^n} \|\bar{R}x - \bar{Q}b\|_2 + \|\tilde{Q}b\|_2,$$

kde  $\bar{R} \in \mathbb{R}^{n \times n}$  je **regulární horní trojúhelníková** matice a  $Q = (\bar{Q}, \tilde{Q})^T$ , kde  $R$  a  $Q$  jsou z definice výše.

*Důkaz.* Nejdříve si ukážeme

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \|Ax - b\|_2 &= \min_{x \in \mathbb{R}^n} \|Q^T(QRx - b)\|_2 = \min_{x \in \mathbb{R}^n} \|Rx - Q^Tb\|_2 = \\ &= \min_{x \in \mathbb{R}^n} \left\| \begin{bmatrix} \bar{R} \\ 0 \end{bmatrix} x - \begin{bmatrix} \bar{Q} \\ \tilde{Q} \end{bmatrix}^T b \right\|_2 = \min_{x \in \mathbb{R}^n} \|\bar{R}x - \bar{Q}b\|_2 + \|\tilde{Q}b\|_2. \end{aligned}$$

Vzhledem k tomu, že  $\|\tilde{Q}b\|_2$  nezávisí na hodnotě  $x$ , tak můžeme minimalizaci (2.1) převést na úlohu

$$\min_{x \in \mathbb{R}^n} \|\bar{R}x - \bar{Q}b\|_2. \quad \square$$

Daný QR-rozklad můžeme získat pomocí (modifikovaného) Gram-Schmitova procesu, Householderovy metody nebo Givensovy metody.

## 2.2.1 MNČ pro singulární rozklad

V předchozí kapitole jsme předpokládali lineární nezávislost sloupců matice  $A$ , což nelze vždycky zaručit, a tedy mějme  $A$  s **lineárně závislými sloupci**. Důvodem, proč nelze použít normální QR rozklad je, že neexistuje trojúhelníková matice  $R$ , proto využijeme SVD rozklad definovaný následovně.

**Definice 5.** Mějme matici  $A \in \mathbb{R}^{m \times n}$ , pak jejím **singulárním rozkladem** nazveme

$$A = U\Sigma V^T = \begin{bmatrix} \bar{U} & \tilde{U} \\ n \times r & n \times (n-r) \end{bmatrix} \begin{bmatrix} \bar{\Sigma} & 0 \\ r \times r & r \times (n-r) \\ 0 & 0 \\ (m-r) \times r & (m-r) \times (n-r) \end{bmatrix} \begin{bmatrix} \bar{V} & \tilde{V} \\ n \times r & n \times (n-r) \end{bmatrix}^T,$$

kde  $U \in \mathbb{R}^{m \times m}$  je **unitární matice**,  $\Sigma \in \mathbb{R}^{m \times n}$  je nulová až na hlavní diagonálu s nezápornými prvky a  $V \in \mathbb{R}^{n \times n}$  je **unitární matice**.

Rozpis v definici se bude hodit v důkazu následující věty.

**Tvrzení 3.** Mějme přeурčenu soustavu lineárních rovnic s reprezentující maticí  $A \in \mathbb{R}^{m \times n}$  s  $\text{rank}(A) = r < n$  a  $b \in \mathbb{R}^m$ , pak

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2 = \min_{x \in \mathbb{R}^n} \|\bar{\Sigma} \bar{V}^T x - \bar{U}b\|_2 + \|\tilde{U}b\|_2,$$

kde  $\bar{\Sigma}, \bar{V}, \bar{U}, \tilde{U}$  jsou z definice 5,  $x$ , které řeší úlohu (2.1) je nekonečně mnoho.

*Důkaz.* První části důkazu se budeme věnovat

$$\begin{aligned} \|Ax - b\|_2 &= \|U\Sigma V^T x - b\|_2 = \|\Sigma V^T x - U^T b\|_2 \\ &= \|\Sigma \begin{bmatrix} \bar{V}^T x \\ \tilde{V}^T x \end{bmatrix} - \begin{bmatrix} \bar{U}b \\ \tilde{U}b \end{bmatrix}\|_2 = \|\bar{\Sigma} \bar{V}^T x - \bar{U}b\|_2 + \|\tilde{U}b\|_2. \end{aligned}$$

Zase problém minimalizace  $\|Ax - b\|_2$  můžeme převést na minimalizaci našeho výrazu  $\|\bar{\Sigma} \bar{V}^T x - \bar{U}b\|_2$ . Lze si všimnout, že takových  $x$  bude nekonečně mnoho, neboť

$$\begin{bmatrix} \bar{V}^T x_{MN\check{C}} \\ \tilde{V}^T x_{MN\check{C}} \end{bmatrix} = \begin{bmatrix} \bar{y} \\ \tilde{y} \end{bmatrix}.$$

Z toho dostáváme vzorec pro  $x$

$$x_{MN\check{C}} = \begin{bmatrix} \bar{V} & \tilde{V} \end{bmatrix} \begin{bmatrix} \bar{y} \\ \tilde{y} \end{bmatrix} = \bar{V}\bar{y} + \tilde{V}\tilde{y}.$$

Vzhledem k tomu, že hodnota  $\|Ax - b\|_2$ , nijak nezávisí na  $\tilde{y}$ , tak můžeme volit  $\tilde{y}$  libovolně, což bude měnit hodnoty  $x_{MN\check{C}}$ .  $\square$

Na problém s maticí  $A$  s lineárně závislými sloupci můžeme pohlížet více způsoby.

## 2.2.2 QR rozklad s permutací sloupců

Mějme maticí  $A$  stejnou jako z 2.2.1 z ní uděláme neredukovaný QR-rozklad, dostaneme  $A = QR$ , pak lze volit maticí  $\Pi$ , která bude **permutovat sloupce** matice tak, že po rozepsání rovnice

$$A\Pi = QR\Pi = Q \begin{bmatrix} \bar{R} & \tilde{R} \\ 0 & 0 \end{bmatrix}$$

bude  $\bar{R}$  **horní trojúhelníková** matice. Vzhledem k tomu, že  $\Pi$  je ortogonální matice, můžeme vyjádřit  $A$  ve tvaru

$$A = Q \begin{bmatrix} \bar{R} & \tilde{R} \\ 0 & 0 \end{bmatrix} \Pi^T,$$

s pomocí tohoto zápisu formulujeme následující větu.

**Tvrzení 4.** Mějme přeурčenu soustavu lineárních rovnic s reprezentující maticí  $A \in \mathbb{R}^{m \times n}$  s  $\text{rank}(A) = r < n$  a  $b \in \mathbb{R}^m$ , pak

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2 = \min_{\tilde{y} \in \mathbb{R}^r} \|\bar{R}\bar{y} + \tilde{R}\tilde{y} - \bar{z}\|_2 + \|\tilde{z}\|_2,$$

kde  $\bar{R}, \tilde{R}, \Pi, Q$  jsou definovány výše,  $\Pi^T x = [\bar{y} \ \tilde{y}]^T$  a  $Q^T b = [\bar{z} \ \tilde{z}]^T$ , pak řešením  $x_{MN\check{C}}$  musí být ve tvaru

$$x_{MN\check{C}} = \Pi \begin{bmatrix} \bar{y}_{MN\check{C}} \\ \tilde{y} \end{bmatrix} = \Pi \begin{bmatrix} \bar{R}^{-1}(\bar{z} - \tilde{R}\tilde{y}) \\ \tilde{y} \end{bmatrix}.$$

*Důkaz.* Pro minimalizaci normy (2.1) použijeme

$$\|Ax - b\|_2 = \|Q \begin{bmatrix} \bar{R} & \tilde{R} \\ 0 & 0 \end{bmatrix} \Pi^T x - b\|_2 = \left\| \begin{bmatrix} \bar{R} & \tilde{R} \\ 0 & 0 \end{bmatrix} \Pi^T x - Q^T b \right\|_2.$$

Po substituci  $\Pi^T x = [\bar{y} \ \tilde{y}]^T$  a  $Q^T b = [\bar{z} \ \tilde{z}]^T$  dostáváme

$$\left\| \begin{bmatrix} \bar{R} & \tilde{R} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \bar{y} \\ \tilde{y} \end{bmatrix} - \begin{bmatrix} \bar{z} \\ \tilde{z} \end{bmatrix} \right\|_2 = \|\bar{R}\bar{y} + \tilde{R}\tilde{y} - \bar{z}\|_2 + \|\tilde{z}\|_2.$$

Vyjádříme  $\bar{y}_{MN\check{C}} = \bar{R}^{-1}(\bar{z} - \tilde{R}\tilde{y})$ . Z toho rovnou vyjádříme dosazením do substituce  $x_{MN\check{C}}$ , pak dostaneme

$$\begin{aligned} \Pi^T x_{MN\check{C}} &= \begin{bmatrix} \bar{y}_{MN\check{C}} \\ \tilde{y} \end{bmatrix} \\ x_{MN\check{C}} &= \Pi \begin{bmatrix} \bar{y}_{MN\check{C}} \\ \tilde{y} \end{bmatrix} = \Pi \begin{bmatrix} \bar{R}^{-1}(\bar{z} - \tilde{R}\tilde{y}) \\ \tilde{y} \end{bmatrix}, \end{aligned}$$

kde  $\tilde{y}$  můžeme volit zcela libovolně, abychom minimalizovali  $\|x_{MN\check{C}}\|_2$ .  $\square$

Soustavy můžeme reprezentovat, jako množinu bodů v prostoru  $\mathbb{R}$ . Pro představu mějme naměřená data, pro která za pomoci MNČ můžeme najít nějakou funkci, která bude predikovat výstupní hodnoty. Toto zkoumáme v další kapitole.

## 2.3 Lineární regrese

Jde o metodu, která aproximuje body křivkou viz (Yan a Su, 2009), a lze ji vyjádřit jako **lineární kombinaci**  $n$  zadaných funkcí. Tedy hledanou neznámou funkci budeme hledat různými metodami, v konkrétním určitém tvaru  $\hat{y} = f(z, \xi_1, \xi_2, \dots, \xi_n) = \xi_1 f_1(z) + \xi_2 f_2(z) + \dots + \xi_n f_n(z)$ , kde reálný vektor neznámých ve tvaru  $[\xi_1, \xi_2, \dots, \xi_n] \in \mathbb{R}^n$  a  $f_1, f_2 \dots f_n$  jsou zadané funkce  $z$ -tu (často se za  $f_1$  bere konstantní funkce) a  $\hat{y}$  vyjadřuje predikovanou hodnotu pro dané  $z$ .

Mějme sadu dvojic  $[z_i, y_i]$  pro každé  $i \in \{1, 2, \dots, m\}$ , kde  $y_i$  je skutečná naměřená hodnota v  $z_i$ . Pak definujeme

$$A = \begin{bmatrix} f_1(z_1) & f_2(z_1) & \dots & f_n(z_1) \\ f_1(z_2) & f_2(z_2) & \dots & f_n(z_2) \\ \vdots & \ddots & \ddots & \vdots \\ f_1(z_m) & f_2(z_m) & \dots & f_n(z_m) \end{bmatrix} \quad x = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{bmatrix} \quad b = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}.$$

Potřebujeme zjistit použitím MNČ, hledaný vektor  $x_{MN\check{C}}$ . Takže budeme řešit

$$\|Ax - b\|_2^2 = \sum_{i=1}^m (\hat{y}_i - y_i)^2 = \sum_{i=1}^m (\xi_1 f_1(z_i) + \xi_2 f_2(z_i) + \dots + \xi_n f_n(z_i) - y_i)^2$$

Nadále se budeme snažit najít minimum za pomoci parciální derivace  $\frac{\partial \|Ax-b\|_2^2}{\partial \xi_j} = 0$  pro každé  $j \in \{1, 2, \dots, n\}$ , takže

$$\begin{aligned} \frac{\partial \|Ax-b\|_2^2}{\partial \xi_j} &= \sum_{i=1}^m 2(\xi_1 f_1(z_i) + \xi_2 f_2(z_i) + \dots + \xi_n f_n(z_i) - y_i) f_j(z_i) = 0, \\ \sum_{i=1}^m (\xi_1 f_1(z_i) + \xi_2 f_2(z_i) + \dots + \xi_n f_n(z_i)) f_j(z_i) &= \sum_{i=1}^m y_i f_j(z_i), \\ \xi_1 \sum_{i=1}^m f_1(z_i) f_j(z_i) + \xi_2 \sum_{i=1}^m f_2(z_i) f_j(z_i) + \dots + \xi_n \sum_{i=1}^m f_n(z_i) f_j(z_i) &= \sum_{i=1}^m y_i f_j(z_i). \end{aligned} \quad (2.2)$$

Opakované využití (2.2) dostaneme soustavu

$$\begin{array}{rcl} \xi_1 \sum_{i=1}^m f_1(z_i) f_1(z_i) + \dots + \xi_n \sum_{i=1}^m f_n(z_i) f_1(z_i) & = & \sum_{i=1}^m y_i f_1(z_i) \\ \vdots & & \vdots \\ \xi_1 \sum_{i=1}^m f_1(z_i) f_n(z_i) + \dots + \xi_n \sum_{i=1}^m f_n(z_i) f_n(z_i) & = & \sum_{i=1}^m y_i f_n(z_i), \end{array}$$

která má  $n$  rovnic o  $n$  neznámých a její řešení vektor  $x_0 = (\xi_{1_0}, \dots, \xi_{n_0})^T$ , což je nejlepší nestranný lineární odhad koeficientů naší hledané funkce  $\hat{y}$ . V následující části dokážeme, že tomu tak doopravdy je.

### 2.3.1 MNČ a nejlepší nestranný lineární odhad $x_0$

V důkazu využijeme Gauss-Markovovy předpoklady (G-M) viz (Hill a kol., 2011).  
**G-M Předpoklady**

1. Existuje **lineární vztah** mezi  $b$  a  $A$ , takže to lze vyjádřit jako

$$b = Ax_x + \varepsilon,$$

kde  $x_x$  je parametrický vektor se skutečnými hodnotami vyjadřující nejlepší kombinační koeficienty vytvářející skutečné  $y_x$  a  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_m)^T$  je vektor náhodných veličin vyjadřující neodstranitelnou chybu v datech.

2.  $\text{rank}(A) = n$
3.  $E[\varepsilon|A] = 0$
4.  $E[\varepsilon\varepsilon^T|A] = \sigma^2 I$ , kde  $\sigma^2$  je rozptyl chyby  $\varepsilon$ -u a  $I$  je identická matice s rozměry  $m \times m$ .

S využitím (Baum, 2004) předvedeme, že  $x_0$  je nestranný odhad  $x_x$ , což okamžitě plyne z

$$\begin{aligned} E[x_0|A] &= E[(A^T A)^{-1} A^T b|A] = E[(A^T A)^{-1} A^T (Ax_x + \varepsilon)|A] \\ &= E[x_x|A] + E[(A^T A)^{-1} A^T \varepsilon|A] = E[x_x|A] + (A^T A)^{-1} A^T E[\varepsilon|A] = x_x, \end{aligned}$$

kde v poslední rovnosti využíváme 3. (G-M). Pak následně s pomocí další definice ukážeme, že je  $x_{MNČ}$  **nejlepší nestranný lineární odhad** v následujícím smyslu viz (Bhattacharya a Burman, 2016).

**Definice 6.** Necht  $x_0$  je lineární funkce proměnné  $b$ , kterou nazveme **nejlepší nestranný lineární odhad**, když splňuje

(i)  $x_0$  je **nestranný odhad**  $x_x$ .

(ii) Pro každý **jiný nestranný lineární odhad**  $\hat{x}$  musí platit, že pro každé  $v \in \mathbb{R}^n$  splňuje  $\text{var}(v^T x_0) \leq \text{var}(v^T \hat{x})$

Nejdříve spočteme rozptyl  $x_0$

$$\begin{aligned}
 \text{Var}(x_0) &= \text{E}[(x_0 - x_x)(x_0 - x_x)^T | A] \\
 &= \text{E}[(A^T A)^{-1} A^T b - x_x)((A^T A)^{-1} A^T b - x_x)^T | A] \\
 &= \text{E}[(A^T A)^{-1} A^T (Ax_x + \varepsilon) - x_x)((A^T A)^{-1} A^T (Ax_x + \varepsilon) - x_x)^T | A] \\
 &= \text{E}[(A^T A)^{-1} A^T \varepsilon ((A^T A)^{-1} A^T \varepsilon)^T | A] = \text{E}[(A^T A)^{-1} A^T \varepsilon \varepsilon^T A (A^T A)^{-1} | A] \\
 &= (A^T A)^{-1} A^T \text{E}[\varepsilon \varepsilon^T | A] A (A^T A)^{-1} = (A^T A)^{-1} A^T \sigma^2 I A (A^T A)^{-1} \\
 &= \sigma^2 (A^T A)^{-1}.
 \end{aligned} \tag{2.3}$$

Nyní si vezmeme nějakou jinou nestrannou lineární statistiku s předpisem  $\hat{x} = Cb$ , kde  $C$  je matice s vhodnými rozměry, tudíž z nestrannosti splňuje

$$\text{E}[Cb | A] = \text{E}[C(Ax_x + \varepsilon) | A] = \text{E}[CAx_x | A] + \text{E}[C\varepsilon | A] = \text{E}[CAx_x | A] = x_x,$$

kde v předposlední nerovnosti využíváme 3. předpoklad (G-M) a poslední rovnost předpokládáme.

Z toho plyne, že

$$CA = I. \tag{2.4}$$

Dostaneme  $\text{Var}(\hat{x}) = \sigma^2 (C^T C)^{-1}$  postup stejně jako (2.3), k tomu když si  $C$  vyjádříme za pomoci  $(A^T A)^{-1} A^T$  jako

$$C = B + (A^T A)^{-1} A^T,$$

kde  $B$  je taková matice, aby rovnost seděla.

Dosazením do (2.5) dostaneme

$$I = CA = BA + (A^T A)^{-1} A^T A = BA + I,$$

to je ekvivalentní s tvrzením  $BA = \bar{0}$ , kde  $\bar{0}$  reprezentuje nulovou matici.

Pak rozepsáním  $\text{Var}(\hat{x})$

$$\begin{aligned}
 \text{Var}(\hat{x}) &= \sigma^2 (C^T C)^{-1} = \sigma^2 (B + (A^T A)^{-1} A^T)^T (B + (A^T A)^{-1} A^T) \\
 &= \sigma^2 (B^T B) + \sigma^2 (A^T A)^{-1},
 \end{aligned}$$

kde  $(B^T B)$  je reálná Gramova matice, která je vždy **pozitivně semidefinitní**, a tedy ukážeme, že splňuje (ii) z Definice 6

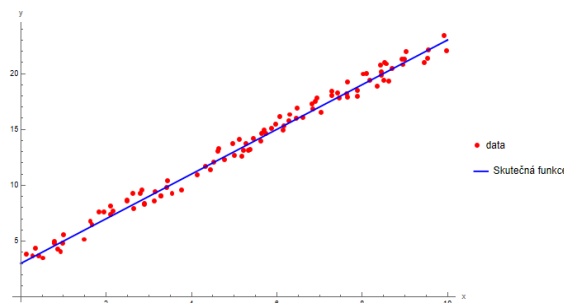
$$v^T \text{Var}(x_0) v = \sigma^2 v^T (A^T A)^{-1} v \leq \sigma^2 v^T (A^T A)^{-1} v + \sigma^2 v^T (B^T B) v = v^T \text{Var}(\hat{x}) v,$$

kde  $v \in \mathbb{R}^m$  a kde  $0 \leq v^T (B^T B) v$  z pozitivní semidefinitnosti. Pak když za  $B$  zvolíme nulovou matici, tak dostaneme přesně  $x_0$ , z toho plyne, že jde o nejlepší nestranný odhad.

## 2.3.2 Příklady lineární regrese

**Příklad 6.** Mějme skutečnou funkci  $g(x) = 2\xi_1 + \xi_2 = 2x + 3$  ( $n = 2$ ) s jejíž pomocí vygeneruji sto náhodných bodů ( $m = 100$ ) s výchylkou  $\varepsilon \sim N(0,1)$  od  $g(x)$  a dostanu body v tabulce níže.

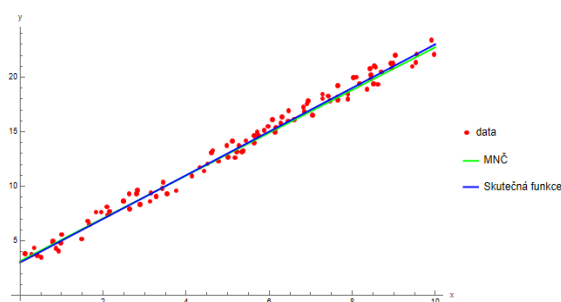
hod. $z_i$	hod. $y_i$	hod. $\varepsilon_i$
2.00508	8.35097	5.73396
5.72017	19.1888	-6.33727
1.31178	-6.09637	-9.29077
$\vdots$	$\vdots$	$\vdots$
3.17797	-8.10982	-8.97686
0.967483	-4.28836	-7.34318



Obrázek 2.1: Tabulka s hodnotami a příslušný graf

Použitím QR rozkladu a MNČ jsme dostali za vteřinu odhady  $\xi_{10} = 1.97656$  a  $\xi_{20} = 3.1448$ , oproti tomu když použijeme minimalizační funkci ve Wolfram Mathematica, tak čekací doba je v řádech desítek minut a řešení se liší až na dvanáctém desetinném místě.

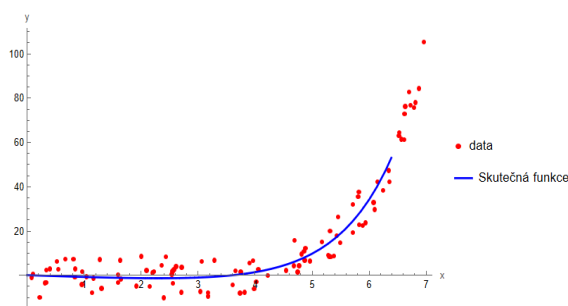
Výsledný graf vypadá následovně



Obrázek 2.2: Rozdíl mezi skutečnou funkcí a odhadem

**Příklad 7.** Mějme složitější funkci  $g(x) = 2e^{(x-3)} - x$  a s její pomocí vygenerujme sto náhodných bodů ( $m = 100$ ) s výchylkou  $\varepsilon \sim N(0,10)$  od  $g(x)$  a dostaneme body v tabulce níže.

hod. $z_i$	hod. $y_i$	hod. $\varepsilon_i$
2.00508	8.35097	5.73396
5.72017	19.1888	-6.33727
1.31178	-6.09637	-9.29077
$\vdots$	$\vdots$	$\vdots$
3.17797	-8.10982	-8.97686
0.96748	-4.28836	-7.34318

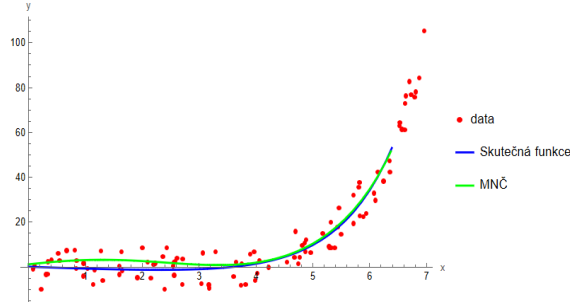


Obrázek 2.3: Tabulka s hodnotami a příslušný graf

Na rozdíl od minulého příkladu se tady budeme snažit najít ideální funkci ve tvaru  $f(x) = ae^x + b \sin x + c \cos x$ , jde o lineární kombinací funkcí  $\{e^x, \sin x, \cos x\}$ .

Použitím QR rozkladu a MNČ jsme během vteřiny dostali všechny hodnoty vektoru  $(a_0, b_0, c_0) = (0.08540, 2.68904, 1.08532)$  a jako v minulém příkladě jsem využil i minimalizační funkci ve Wolfram Mathematica s dobou čekání přes deset minut a liší se až na dvanáctém desetinném místě.

Výsledný graf vypadá následovně



Obrázek 2.4: Rozdíl mezi skutečnou funkcí a odhadem

## 2.4 Nelineární regrese

V této části se budeme snažit aproximovat křivkou, která **nelze** vyjádřit jako **lineární kombinace funkcí** ve stejném smyslu jak v kapitole 2.3. Ku příkladu

$$f(z_i, \xi_1, \xi_2) = \frac{z_i - \xi_1}{2z_i + \xi_2},$$

kde budeme hledat ideální hodnoty  $\xi_1, \xi_2$  pro zadané vektory  $z \in \mathbb{R}^m$  a  $b \in \mathbb{R}^m$  a  $f(z, \xi_1, \xi_2) = (f(z_1, \xi_1, \xi_2), \dots, f(z_m, \xi_1, \xi_2))^T$ . Obecně máme

$$b = f(z, \xi) + \varepsilon,$$

kde  $b \in \mathbb{R}^m$ ,  $z \in \mathbb{R}^m$ ,  $\xi \in \mathbb{R}^n$  a  $\varepsilon \in \mathbb{R}^m$ , kde  $\varepsilon$  je neměřitelná chyba.

Označme si reziduální vektor jako  $r = b - f(z, \xi)$  a budeme chtít zase minimalizovat jeho dvojkovou normu, kterou si označíme  $L$ , takže chceme

$$\min_{\xi \in \mathbb{R}^n} L = \min_{\xi \in \mathbb{R}^n} \|r\|_2^2 = \min_{\xi \in \mathbb{R}^n} \|b - f(z, \xi)\|_2^2,$$

což podobně jako v 2.3 najdeme řešením soustav

$$\nabla_{\xi} L = \bar{0}, \tag{2.5}$$

kde  $\nabla_{\xi} L$  je vektor parciálních derivací  $L$  vzhledem ke složkám v  $\xi$  a  $\bar{0}$  je nulový vektor.

Na rozdíl od minulé kapitoly nemusíme dané řešení (2.5) najít v konečně mnoha krocích, a proto se používají iterační metody.

Budeme využívat **Newtonovu metodu**, která je plně rozvedená viz (Fletcher, 1987), jenž má předpis

$$\xi^{k+1} = \xi^k - \mathbb{J}_g(\xi^k)^{-1} g(\xi^k),$$

kde  $\mathbb{J}_g^{-1}$  je **Jacobiho matice** nějaké funkce  $g$  a předpokládáme, že  $\xi^k$  postupně konvergují k hodnotě  $\xi^0$ , pro kterou platí  $g(\xi^0) = 0$ .

V našem případě je  $g = \nabla_{\xi^k} L$  z toho

$$\xi^{k+1} = \xi^k - \nabla_{\xi^k}^2 L^{-1} \nabla_{\xi^k} L, \quad (2.6)$$

kde  $\nabla_{\xi^k}^2 L$  je **Hessova matice**  $k$ -té iterace, takže musíme předpokládat, že je regulární.

Vypočítat přesně Hessovu matici může být problém, a proto se ji budeme snažit aproximovat.

Nejdřív si vyjádříme  $j$ -tý prvek z  $\nabla_{\xi} L$

$$\nabla_{\xi_j} L = \sum_{i=1}^m 2r_i \frac{\partial r_i}{\partial \xi_j} = \sum_{i=1}^m 2r_i \frac{\partial (y_i - f(z, \xi)_i)}{\partial \xi_j} = -2 \sum_{i=1}^m r_i \frac{\partial f(z, \xi)_i}{\partial \xi_j} = -2 \mathbb{J}_{f_j}^T r,$$

kde  $\mathbb{J}_{f_j}$  je  $j$ -tý sloupec Jacobiho matice funkce  $f$  a  $f(z, \xi)_i$  je  $i$ -tý prvek ve vektoru  $f(z, \xi)$ . Proto gradient vyjádříme jako

$$\nabla_{\xi} L = -2 \mathbb{J}_f^T r.$$

Prvky z Hessovy matice si vyjádříme podobným způsobem

$$\nabla_{\xi_j \xi_k}^2 L = \frac{(-2 \sum_{i=1}^m r_i \frac{\partial f(z, \xi)_i}{\partial \xi_j})}{\partial \xi_k} = 2 \sum_{i=1}^m \left( \frac{\partial f(z, \xi)_i}{\partial \xi_k} \frac{\partial f(z, \xi)_i}{\partial \xi_j} - r_i \frac{\partial^2 f(z, \xi)_i}{\partial \xi_k \partial \xi_j} \right).$$

Najít druhou parciální derivaci je mnohdy velmi náročné, a proto budeme aproximovat Hessovu matici jenom

$$\nabla_{\xi_j \xi_k}^2 L \approx 2 \sum_{i=1}^m \left( \frac{\partial f(z, \xi)_i}{\partial \xi_k} \frac{\partial f(z, \xi)_i}{\partial \xi_j} \right),$$

protože druhý člen na pravé straně rovnosti závisí na  $r_i$ , tak musíme předpokládat, že  $r_i$  jsou dost malé a druhé parciální derivace nejsou velké, což si můžeme vyjádřit pomocí Jacobiho matice, jako

$$\nabla^2 L \approx 2 \mathbb{J}_f^T \mathbb{J}_f.$$

Dosazením do (2.6) dostaneme

$$\begin{aligned} \xi^{k+1} &= \xi^k - \nabla^2 L(\xi^k)^{-1} \nabla L(\xi^k) \approx \xi^k - (2 \mathbb{J}_f(\xi^k)^T \mathbb{J}_f(\xi^k))^{-1} (-2 \mathbb{J}_f(\xi^k)^T r(\xi^k)) = \\ &= \xi^k + (\mathbb{J}_f(\xi^k)^T \mathbb{J}_f(\xi^k))^{-1} (\mathbb{J}_f(\xi^k)^T r(\xi^k)), \end{aligned}$$

kde  $r(\xi^k) = b - f(z, \xi^k)$ .

Nevýhoda Newtonovy metody je, že se může zacyklit, takže hodnoty  $\xi^k = \xi^h$  pro  $k \neq h$  nebo může nastat, že  $\mathbb{J}_f(\xi^k)$  bude nulová matice a tedy  $\xi^k = \xi^{k+1}$ .

### 2.4.1 Linearizace problému

V některých případech se můžeme vyhnout iterování a to tím způsobem, že nelineární funkci si převedeme **ekvivalentními úpravami** na lineární kombinaci funkcí.

Ku příkladu mějme

$$y_i = f(z_i, a, b, c) = a e^{bz_i + cz_i^2},$$



když zlogaritmujeme obě strany dostaneme

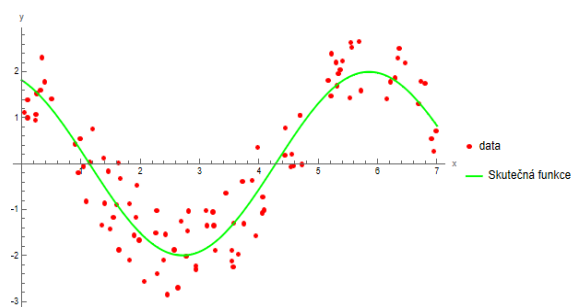
$$\ln(y_i) = \ln(a) + bz_i + cz_i^2 = g_1(z_i)\ln(a) + bg_2(z_i) + cg_3(z_i).$$

Uděláme substituci za  $\ln(a) = \hat{a}$  a dostaneme funkci  $g(g_1(z_i), g_2(z_i), g_3(z_i), \hat{a}, b, c)$ , která dělá lineární kombinaci funkcí  $(g_1, g_2, g_3)$ . Pak použijeme postup z 2.3, dostaneme nejlepší odhady  $(\ln(a_0), b_0, c_0)$ , kde  $(a_0, b_0, c_0)$  jsou odhady získané MNČ.

## 2.4.2 Příklad nelineární regrese

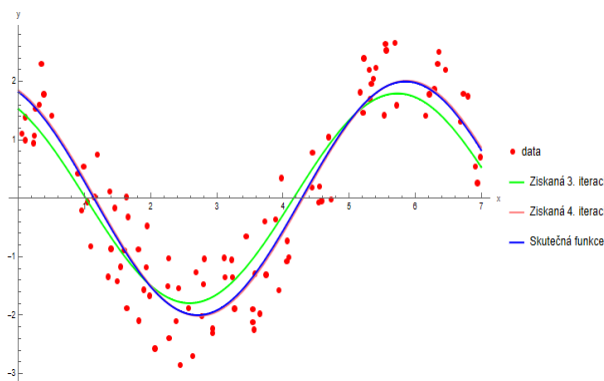
**Příklad 8.** Mějme funkci  $g(x) = a \sin b + x = 2 \sin 2 + x$  ( $a = 2, b = 2$ ) a s její pomocí vygeneruji sto náhodných bodů ( $m = 100$ ) s výchytkou  $\varepsilon \sim N(0,1)$  od  $g(x)$  a získám body v tabulce níže.

hod. $z_i$	hod. $y_i$	hod. $\varepsilon_i$
4.44534	0.177173	-0.14572
6.91651	0.537205	-0.436118
5.23251	2.38689	0.760846
$\vdots$	$\vdots$	$\vdots$
6.94943	0.254428	-0.660864
6.16575	1.40538	-0.498213



Obrázek 2.5: Tabulka s hodnotami a příslušný graf

Iterováním pomocí Newtonovy metody z počátečního bodu  $\xi^0 = (1,1)^T = (a_0, a_1)$  již v prvních iteracích dostáváme dobrý odhad a od čtvrté iterace dál nelze na obrázku skoro rozeznat mezi původní funkcí a odhadem.



Obrázek 2.6: Rozdíl mezi skutečnou funkcí a iterovanými odhady

Odhady se od 6. iterace na prvních pěti desetinných místech nemění, z čehož lze usoudit, že je tato metoda velmi efektivní a přesná s  $\xi^6 = (2.0289, 1.99442)^T$ .

## 2.5 Minimalizace $\ell_1$ -normy a metoda Theila a Sena

V této části se zabýváme stejným problémem s rozdílem v zápisu, kde místo  $\ell_2$ -normy (2.1) budeme brát  $\ell_1$ -normu

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_1.$$

Jedna ze známých alternativní metod pro jednoduchou lineární regresi je Theilova a Senova metoda.

### 2.5.1 Jednoduchá lineární regrese podle Theila a Sena

Regrese připisovaná autorům Theilovi a Senovi (tzv. Theil-Sen regression) viz (Birkes a Dodge, 1993) dává řešení s podobnými vlastnostmi jako  $\ell_1$ -norma, v některých případech je pro jednoduchou lineární regresi lepší než MNČ; například nedává takový důraz na data, která se hodně odchyľují od přímky, nebo také nevyžaduje pro neasymptotické testy předpoklad o chybě, neboli  $\varepsilon \sim N(0, \sigma^2)$ .

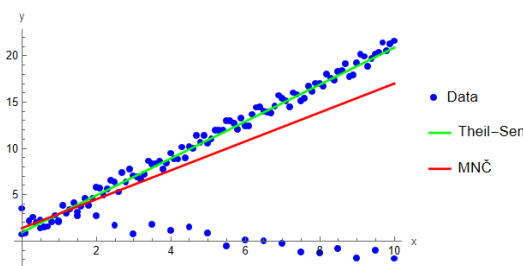
V **jednoduché lineární regresi** chceme najít odhady koeficientů  $a$  a  $b$ , kde parametry jsou ve funkci  $f(x) = ax + b$ . Nejdříve touto metodou vypočítáme odhad  $a_0$  tak, že

$$a_0 = \text{median} \left\{ \frac{y_i - y_j}{x_i - x_j} \right\}_{i,j=1}^{m,m};$$

pokud se hodnoty rovnají,  $x_k = x_l$  pro  $k, l \in \mathbb{N}$ , pak se člen z množiny vynechává. Tato regrese tedy pracuje s mediánem sklonů přímek mezi všemi body, které mají různou hodnotu v  $x$ . Odhad hodnoty  $b_0$  pak dostáváme jako

$$b_0 = \text{median} \{y_i - a_0 x_i\}_{i=1}^m.$$

Rozdíly mezi metodami předvedu na obrázku.



Obrázek 2.7: Rozdíl mezi metodou Theila a Sena a MNČ. Je na první pohled zřejmé, že v tomto speciálním případě se metoda Theila a Sena téměř nenechá ovlivnit vzdálenými body, na rozdíl od MNČ.

### 2.5.2 Lineární regrese pro $\ell_1$ -normu

Navraťme se tedy k původnímu problému

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_1,$$

můžeme ekvivalentně přepsat

$$\min_{x \in \mathbb{R}^n, t \in \mathbb{R}} \|t\|_1 = \min_{x \in \mathbb{R}^n, t \in \mathbb{R}} \mathbf{1}^T t, \quad (2.7)$$

s podmínkou

$$-t \preceq Ax - b \preceq t, \quad (2.8)$$

kde  $\preceq$  má v tomto kontextu význam, že je větší nebo roven po prvcích ve vektorech stejné velikosti a  $\mathbf{1}$  je  $n$  dimenzionální vektor jedniček.

Vidíme, že minimalizace výrazu (2.7) se vyskytuje pouze  $t$  a  $x$  jen v podmínce (2.8), proto úlohu přepíšeme za pomoci  $q = \begin{bmatrix} t^T & x^T \end{bmatrix}^T$ , což je náš nový cíl pro minimalizaci, podle toho přepíšeme do tvaru

$$\min_{q \in \mathbb{R}^{2n}} \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{1} \times n & \mathbf{1} \times n \end{bmatrix} q,$$

ještě se musíme vrátit k podmínce (2.8), ze které získáme,

$$-t \preceq Ax - b \implies -t - Ax \preceq -b \implies \begin{bmatrix} -I & -A \end{bmatrix} q \preceq -b,$$

kde  $I$  je identická matice vhodné velikosti. V následném kroku provedeme totéž pro druhou stranu podmínky. Spojením dostaneme

$$\begin{bmatrix} -I & -A \\ -I & A \end{bmatrix} q \preceq \begin{bmatrix} -b \\ b \end{bmatrix}. \quad (2.9)$$

Rovnice (2.9) je ve vhodném tvaru pro neznámý vektor  $q$ , takže stačí použít pouze metodu lineárního programování viz (Arenales a kol., 2015) a dostaneme ideální řešení.

# 3. Podurčené/Nedourčené soustavy rovnic

V této kapitole se budeme zabývat situací

$$Ax = b,$$

kde na rozdíl od předešlé kapitoly  $m < n$ . Dále nás budou zajímat konzistentní případy, tedy když existuje nějaké řešení  $x_0 \in \mathbb{R}^n$ . Potom jich existuje nekonečně mnoho, neboť pro každé  $z \in Ker(A)$  platí

$$A(x_0 + z) = Ax_0 + Az = b + 0,$$

jelikož  $0 < n - m \leq dim(Ker(A))$ , tak  $Ker(A)$  je izomorfní s  $\mathbb{R}^{dim(Ker(A))}$ , což je nekonečná množina.

To nás vede k otázce vybrat z množiny  $\{x + z : Ax = b, z \in Ker(A)\} = S_{A,b}$  "nejlepší  $x$ ".

## 3.1 Minimalizace $\ell_2$ -normy

Budeme předpokládat, že  $rank(A) = m$ , kvůli invertovatelnosti matice  $(AA^T)$ . Vezmeme pak  $x$ , které dostaneme z MNČ, takže

$$x_{MNČ} = A^T(AA^T)^{-1}b. \quad (3.1)$$

**Tvrzení 5.**  $x_{MNČ}$  řeší  $Ax = b$  a má nejmenší  $\ell_2$ -normu z množiny  $\{z : Az = b\}$ .

*Důkaz.* Mějme  $z \in X_{A,b}$ , takže  $A(z - x_{MNČ}) = 0$ , potom

$$\begin{aligned} \|z\|_2^2 &= \|x_{MNČ} + z - x_{MNČ}\|_2^2 \\ &= \|x_{MNČ}\|_2^2 + \|z\|_2^2 + 2\langle x_{MNČ}, z - x_{MNČ} \rangle \\ &= \|x_{MNČ}\|_2^2 + \|z\|_2^2 \geq \|x_{MNČ}\|_2^2, \end{aligned}$$

což nám dává, že  $x_{MNČ}$  má nejmenší  $\ell_2$ -normu ze všech  $z$ , která řeší  $Ax = b$ . Jen nám zbývá ukázat, že poslední rovnost platí.

Tedy chceme ukázat, že  $\langle z - x_{MNČ}, x_{MNČ} \rangle = 0$ , což plyne z

$$\begin{aligned} \langle z - x_{MNČ}, x_{MNČ} \rangle &= (z - x_{MNČ})^T x_{MNČ} = (z - x_{MNČ})^T A^T (AA^T)^{-1} y = \\ &= (A(z - x_{MNČ}))^T (AA^T)^{-1} y = 0. \end{aligned}$$

Alternativním způsobem, jak ukázat, že  $x_{MNČ}$  minimalizuje  $\ell_2$ -normu je použití Lagrangeových multiplikátorů. Chceme splnit dvě podmínky zároveň a to

$$Ax = b \quad \text{a} \quad \min_{x \in \mathbb{R}^n} \|x\|_2 = \min_{x \in \mathbb{R}^n} x^T x.$$

Takže si definujeme funkci  $L(x, \lambda) = x^T x + \lambda^T (Ax - b)$ , kde  $\lambda \in \mathbb{R}^m$  a chceme najít její minimum pomocí gradientu.

Parciální derivací dostaneme

$$\nabla_x L = 2x + A^T \lambda = 0 \quad \& \quad \nabla_\lambda L = Ax - b = 0,$$

kde z první rovnice vyjádříme  $x$  a dostaneme tvar

$$x = -\frac{1}{2}A^T \lambda. \quad (3.2)$$

Vynásobením zleva maticí  $A$  dostaneme

$$b = Ax = -\frac{1}{2}AA^T \lambda.$$

Z toho vyjádříme  $\lambda$

$$\lambda = -2(AA^T)^{-1}b.$$

Dosazením  $\lambda$  do (3.2) dostaneme naše hledané  $x_{MN\check{C}} = A^T(AA^T)^{-1}b$ . Z toho dostáváme takové řešení MNČ, které minimalizuje  $\ell_2$ -normu.  $\square$

Problém s MNČ je, že rozptyl odhadu je velmi závislý na velikosti  $m$ . Avšak v této kapitole bereme  $m < n$  (v náročnějším případě  $m \ll n$ ), tak ani řešení  $Ax = b$  není jednoznačně určené. Proto musíme zvolit způsob, jako například (minimalizace  $\ell_2$ -normy) pro získání jednoho řešení, o kterém ani nevíme, zdali je způsob zvolený správně.

To nás vede k otázce, jestli existuje efektivní způsob, který bude vyřazovat/zmenšovat nedůležité parametry. A nebudeme při jeho použití riskovat, že ztratíme ideální řešení na základě špatně zvolené podmínky pro jednoznačnost.

## 3.2 Selekce lineárních modelů

V této část vycházející z (James a kol., 2013) se budeme zabývat jinou metodou než MNČ, kde už nebudeme **nevyžadovat konzistenci** soustavy. Ke hledání standardních lineárních modelů o  $n$  parametrech ve tvaru

$$y_i = \xi_0 + \xi_1 a_{i,1} + \dots + \xi_{n-1} a_{i,n-1} + \varepsilon, \quad (3.3)$$

kde  $y_i$  je naměřená hodnota pro hodnoty řádku matice  $A$  vypadající následovně jako  $a_i = (a_{i,1}, \dots, a_{i,n-1})^T$ , je vektor parametrů  $x = (\xi_0, \dots, \xi_{n-1})^T$  a  $\varepsilon$  je neměřitelná chyba. **Model** o velikosti  $k$  budeme chápat jako množinu  $\{\xi_{i_1}, \xi_{i_2}, \dots, \xi_{i_k}\}$ , kde  $i_1, i_2, \dots, i_k \in \{0, 1, \dots, n-1\}$  a  $i_j \neq i_l$  pro  $j \neq l$ . Odlišné metody od MNČ mohou mít lepší vlastnosti ve smyslu přesnějšího odhadu  $x$  a schopnosti interpretace modelu.

Zpřesnit neboli zmenšovat rozptyl odhadů můžeme pomocí použití méně parametrů, které vyhodnotíme jako "nepodstatné" v nějakém smyslu.

Lepší interpretaci ve smyslu, že mnohdy při určování  $x$ -u, můžeme zvolit  $\xi_i$ , na kterých nezávisí vektor  $b$ , proto se takové hodnoty budeme snažit vyliminoval/minimalizovat.

Ukážeme si metody, které budou zlepšovat výše zmíněné vlastnosti. V následující kapitolách budeme zmiňovat metody, které lze použít, jak pro  $m \leq n$ , tak pro  $m \geq n$  pokud neřekneme jinak. Protože jsme v kapitole podurčených lineárních soustav, tak se budeme zabývat převážně případem  $m \leq n$ .

### 3.2.1 Nejlepší výběrová regrese

Metoda nejlepší výběrové regrese nám zaručeně vrátí nejlepší model, který lze získat z naměřených dat. Její nevýhodou zůstává, že musí projít všechny možné modely. To pro velké  $n$  může být nedosažitelný cíl z časového hlediska, protože počet různých modelů je  $2^n$  (pro konkrétní případ buď jednotlivý parametr necháme nebo ho vynecháme), což je pro velké  $n$  velmi nepraktické. Dalším problémem může být zvolení nějakého parametru, který ve skutečnosti nemá s naměřenými daty nic společného, čemuž se chceme vyhnout.

Způsob, jak takový model najít se dá vyřešit algoritmicky

#### Algoritmus:

1. Najít  $\mathcal{M}_0$  nulový model, který nemá žádný parametr bude odhadovat pouze výběrový průměr.
2. Pro každé  $k = \{0, 1, 2, \dots, n\}$  provedeme MNČ, pro každou kombinaci  $k$  parametrů z toho dostaneme potenciačních  $\binom{n}{k}$  modelů, pro každé  $k$  a pro  $k$  určíme  $\mathcal{M}_k$ , takové které bude mít nejmenší RSČ (residuální součet čtverců), takže vyhodnotíme pomocí MNČ.
3. Chceme vybrat nejlepší model z množiny  $\{\mathcal{M}_0, \mathcal{M}_1, \dots, \mathcal{M}_n\}$ . Nemůžeme brát nejlepší ve smyslu minimalizace RSČ nebo maximalizace  $R^2$  (koeficient determinace), protože s rostoucím  $k$  se zmenšuje hodnota RSČ a tedy bychom volili  $\mathcal{M}_n$ . To není ideální, protože bychom měli zohledňovat množství použitých parametrů, jelikož se chceme vyhnout zbytečně velkému množství parametrů, které nám nezlepšuje odhad. Kvůli tomu použijeme například  $\bar{R}^2$  upravený koeficient determinace, který chceme maximalizovat,

kde koeficient determinace je definovaný jako

$$R^2 = \frac{\sum_{i=1}^m (f(a_i) - \bar{y}_m)}{\sum_{i=1}^m (y_i - \bar{y}_m)},$$

kde  $f(a_i)$  je predikovaná hodnota naším modelem pro  $a_i$  a  $\bar{y}_m$  je výběrový průměr  $m$  naměřených hodnot a upravený koeficient determinace je definovaný jako

$$\bar{R}^2 = 1 - \frac{\sum_{i=1}^m (f(a_i) - \bar{y}_i)^2}{\sum_{i=1}^m (y_i - \bar{y}_m)^2} \frac{m-1}{m-d-1},$$

kde  $d$  je počet použitých parametrů.

V kroku 2. musíme projít všechny modely až na  $\mathcal{M}_0$  a  $\mathcal{M}_n$ , poté to zredukujeme na  $n+1$  modelů, na kterých provedeme krok 3.

Pokud z důvodu moc velkého  $n$  není možné použít metodu nejlepší výběrové regrese, použijeme rychlejší metodu i přesto, že bude šance, že nenajdeme úplně nejlepší model.

### 3.2.2 Kroková regrese

Výhodou této metody je, že prochází mnohem menší počet modelů a to přesně  $m+1$  (Důvodem, proč je množina jen do  $m$ , je MNČ, která dá jednoznačné řešení jen pro  $k \leq m$ ). Metoda využívá následující algoritmus.

**Algoritmus:**

1. Najít  $\mathcal{M}_0$  nulový model, který nemá žádný parametr, bude odhadovat pouze výběrový průměr.
2. Pro  $k = \{0, 1, \dots, m\}$  (Důvodem, proč je množina jen do  $m$ , je MNČ, která dá jednoznačné řešení jen pro  $k \leq m$ .) máme nejlepší odhad  $\mathcal{M}_k$ , pak daný model rozšíříme o jeden parametr, z toho dostaneme  $n - k$  modelů, za pomoci MNČ vybereme nejlepší. Poté za  $\mathcal{M}_{k+1}$  vybereme model, který má nejmenší hodnotu RSC.
3. Vybereme z vytvořené množiny  $\{\mathcal{M}_0, \mathcal{M}_1, \dots, \mathcal{M}_m\}$  úplně stejně jako v minulém algoritmu. K vyhodnocení nejlepšího modelu využijeme například  $\bar{R}^2$ , který chceme maximalizovat.

V kroku 2. tento postup kontroluje pouze

$$1 + \sum_{k=0}^{m-1} (n - k) = 1 + \frac{(n - m)(2n - m + 1)}{2},$$

kde 1 je počáteční  $\mathcal{M}_0$ , což je v praxi mnohem přívětivější než  $2^n$  metodou nejlepšího výběru.

Nevýhoda tohoto postupu je, že postupné určování modelů závisí na těch předchozích, pokud na začátku špatně zvolíme parametr, na kterém vůbec nezávisí měřené hodnoty, tak všechny další modely budou také neideální. Proto si ukážeme zpětně krokovou metodu, která tomuto problému předchází.

### 3.2.3 Zpětná kroková regrese

Rozdíl mezi krokovou metodou a zpětnou krokovou metodou je, že zpětná kroková metoda začíná s plným modelem  $\mathcal{M}_n$  a tedy vyžaduje, aby  $m \geq n$ , což je nevýhodou. Čímž se liší od krokové metody, kde parametry odebíráme (proto zpětná kroková metoda), což je jediná odlišnost mezi algoritmy.

Pravděpodobnost vyřazení důležitých parametrů je mnohem menší než těch nedůležitých, a proto preferujeme pro případ ( $m \geq n$ ) používat tuto metodu. I tak nemáme garantováno, že vrátí ideální model.

Vrácené modely zpětnou krokovou regresi a krokovou regresi jsou podobné, ale mnohdy nemusí být stejné.

V praxi se používá hybridní regrese, která kombinuje tyto dvě krokové metody. Tento metoda vrací model dost podobný jako nejlepší výběrová regrese.

## 3.3 Shrinkage metody

Shrinkage metody viz (James a kol., 2013) jsou takové, co zmenšují (shrink) odhady parametrů, které na sadu dat nemají žádný dopad nebo jen velmi malý. Takové parametry vybíráme, za pomoci minimalizace upraveného RSC. V této kapitole si ukážeme dvě nejčastěji používané metody.

### 3.3.1 Hřebenová regrese

Mimo krokových metod a metody nejlepšího výběru existuje další metoda, která téměř nikdy nevrátí model s méně parametry než je hodnota  $n$ , ale nevýznamné parametry výrazně zmenšuje. Tuto metodu hřebenové regrese definujeme následovně:

**Definice 7.** *Nechť  $A \in \mathbb{R}^{m \times n}$ , kde  $m, n \in \mathbb{N}$ , vektor  $b \in \mathbb{R}^m$  a parametrický vektor  $x = (\xi_0, \xi_1, \dots, \xi_{n-1})^T$ . Pak řešením úlohy*

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^m (y_i - \xi_0 - \sum_{j=1}^{n-1} \xi_j a_{i,j})^2 + \lambda \sum_{j=1}^{n-1} \xi_j^2,$$

kde  $\lambda \geq 0$  nazýváme ladícím parametrem,  $\sum_{i=1}^m (y_i - \xi_0 - \sum_{j=1}^{n-1} \xi_j a_{i,j})^2 = RS\check{C}$  a poslední  $\lambda \sum_{j=1}^{n-1} \xi_j^2$  nazveme penalizační funkcí, označíme funkci  $x_L(\lambda)$  s parametrem  $\lambda$  jako řešení dané úlohy pro zadané  $\lambda$ .

Definici můžeme ekvivalentně přepsat jako

**Definice 8.** *Nechť  $A \in \mathbb{R}^{m \times n}$ , kde  $m, n \in \mathbb{N}$ , vektor  $b \in \mathbb{R}^m$  a parametrický vektor  $x = (\xi_0, \xi_1, \dots, \xi_{n-1})^T$ . Pak řešením úlohy*

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^m (y_i - \xi_0 - \sum_{j=1}^{n-1} \xi_j a_{i,j})^2 \quad \text{s podmínkou} \quad \sum_{j=1}^{n-1} \xi_j^2 \leq t,$$

kde  $0 \leq t$  nám určuje poloměr koule (vzhledem k  $\ell_2$  normě), ze které vybíráme  $x$ ,  $\sum_{i=1}^m (y_i - \xi_0 - \sum_{j=1}^{n-1} \xi_j a_{i,j})^2 = RS\check{C}$  a, označíme funkci  $x_L(t)$  s parametrem  $t$  jako řešení dané úlohy pro zadané  $t$ .

Ukážeme tvrzení ekvivalence s důkazem inspirované z (Klouda a kol., 2016)

**Tvrzení 6.** *Definice 7 a 8 jsou ekvivalentní ve smyslu, že pro každé  $\lambda$  existuje  $t$  a pro každé  $t$  existuje  $\lambda$ , že existuje  $x_H$ , které je řešením obou úloh.*

*Důkaz.*  $7 \implies 8$

Mějme  $\hat{x} = (\hat{\xi}_0, \hat{\xi}_1, \dots, \hat{\xi}_{n-1})^T$ , které je řešením 7. Definujme  $t = \sum_{i=1}^{n-1} (\hat{\xi}_i)^2$ , pak použitím  $t$  v 8 dostaneme  $\bar{x} = (\bar{\xi}_0, \bar{\xi}_1, \dots, \bar{\xi}_{n-1})^T$ .

Pro spor předpokládejme, že  $\sum_{i=1}^{n-1} (\bar{\xi}_i)^2 < t = \sum_{i=1}^{n-1} (\hat{\xi}_i)^2$  a  $\hat{x}$  splňuje podmínku z 8, pak musí platit  $RS\check{C}(\hat{x}) \geq RS\check{C}(\bar{x})$ , kde  $RS\check{C}$  bereme jako funkci odhadu parametru  $x$ . Dostáváme

$$RS\check{C}(\hat{x}) + \lambda \sum_{i=1}^{n-1} (\hat{\xi}_i)^2 > RS\check{C}(\bar{x}) + \lambda \sum_{i=1}^{n-1} (\bar{\xi}_i)^2,$$

což je spor s minimalitou  $\bar{x}$ . Z toho plyne, že  $\bar{x}$  musí splňovat  $\sum_{i=1}^{n-1} (\bar{\xi}_i)^2 = t$ . Takže pro tenhle případ 7 a 8 lze chápat, že z množiny, ze které vybíráme musí splňovat



$\sum_{i=1}^{n-1} (\bar{\xi}_i)^2 = t$  a minimalizovat RSČ, což má jedno řešení, ale jak  $\bar{x}$ , tak  $\hat{x}$  to splňují, takže se musí rovnat.

7  $\Leftarrow$  8 Mějme řešení 8, pro nějaké  $t \geq 0$ ,  $\bar{x}$  pokud  $\sum_{i=1}^{n-1} (\bar{\xi}_i)^2 < t$ , pak se řešení  $x_{MN\check{C}}$  nachází v kružnici o s poloměrem  $t$  viz 3.1 a tedy volíme  $\lambda = 0$ . Pro případ, že  $\sum_{i=1}^{n-1} (\bar{\xi}_i)^2 = t$  s touto podmínkou budeme řešit pomocí Lagrangeového multiplikátoru. Definujme si funkci

$$L(x, \tau) = RS\check{C}(x) + \tau \left( \sum_{i=1}^{n-1} (\xi_i)^2 - t \right).$$

Pomocí derivací najdeme minimum funkce  $L(x, \tau)$  a zvolíme jako naše hledané  $\lambda$  pro 7. Pak řešení 7  $\hat{x} = \bar{x}$ . □

Můžeme si všimnout, že když  $\lambda = 0$ , tak nemá žádnou penaltu a tedy minimalizujeme RSČ, což vlastně provádíme MNČ, kde máme nekonečně mnoho řešení, čemuž se chceme vyhnout (proto se v praxi volí  $\lambda \rightarrow 0$ ) a výsledný odhad, co dostaneme, má velmi malou šanci, že nějaké  $\xi_i = 0$ , ukážeme si v části 3.3.3. Pro  $\lambda \rightarrow \infty$  naopak vynuluje všechny  $\xi_i$  až na  $\xi_0$ , které se v sumě čtverců u ladícího parametru nenachází.  $\xi_0$  nechceme zmenšovat, protože ideální odhad pro  $\xi_0$  je průměr naměřených dat s  $a_i = (0, 0, \dots, 0)^T$ , kde  $a_i$  je řádek matice  $A$ .

Odhad z MNČ je nestranný, ale má veliký rozptyl, který často roste se zvětšujícím  $n$ . Na rozdíl od hřebenové regrese, kde s rostoucí  $\lambda$  ztrácíme nestrannost odhadu, ale zmenšuje se rozptyl. Proto chceme dobře volit  $\lambda$ , abychom neztratili nestrannost úplně, ale zmenšili rozptyl co nejvíce.

Jedna z dalších výhod hřebenové regrese je, že pro pevné  $\lambda$  zkoumá pouze jeden model stejně jako MNČ, takže je mnohem rychlejší řešení než krokové metody a mnohem lepší než nejlepší výběrová regrese v tomto smyslu.

### 3.3.2 Lasso (Least Absolute Shrinkage and Selection Operator)

**Definice 9.** Necht  $A \in \mathbb{R}^{m \times n}$ , kde  $m, n \in \mathbb{N}$ , vektor  $b \in \mathbb{R}^m$  a parametrický vektor  $x = (\xi_0, \xi_1, \dots, \xi_{n-1})^T$ . Pak řešením úlohy

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^m (y_i - \xi_0 - \sum_{j=1}^{n-1} \xi_j a_{i,j})^2 + \lambda \sum_{j=1}^{n-1} |\xi_j|,$$

kde  $\lambda \geq 0$  nazýváme ladícím parametrem,  $\sum_{i=1}^m (y_i - \xi_0 - \sum_{j=1}^{n-1} \xi_j a_{i,j})^2 = RS\check{C}$  a poslední

$\lambda \sum_{j=1}^{n-1} |\xi_j|$  nazveme penalizační funkcí, označíme funkci  $x_L(\lambda)$  s parametrem  $\lambda$  jako řešení dané úlohy pro zadané  $\lambda$ .

Rozdíl mezi Lassem a hřebenovou regresí je v penalizační funkci, kde Lasso používá  $\ell_1$ -normu a hrubá regrese  $\ell_2$ -normu. To je hlavní důvod, proč Lasso je považovaný za efektivnější způsob zmenšování, či dokonce nulování nepodstatných  $\xi_i$ , protože  $\ell_1$ -norma dává důraz stejně velký na všechny  $\xi_i$ , ať je velké či

malé oproti  $\ell_2$ -normě, která zdůrazňuje velké hodnoty a ignoruje malé hodnoty. To je základní intuice, proč je Lasso preferované.

K tomu, abychom ukázali budeme potřebovat ekvivalentní zápis Lassa:

**Definice 10.** *Nechť  $A \in \mathbb{R}^{m \times n}$ , kde  $m, n \in \mathbb{N}$ , vektor  $b \in \mathbb{R}^m$  a parametrický vektor  $x = (\xi_0, \xi_1, \dots, \xi_{n-1})^T$ . Pak řešením úlohy*

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^m (y_i - \xi_0 - \sum_{j=1}^{n-1} \xi_j a_{i,j})^2 \quad \text{s podmínkou} \quad \sum_{j=1}^{n-1} |\xi_j| \leq t,$$

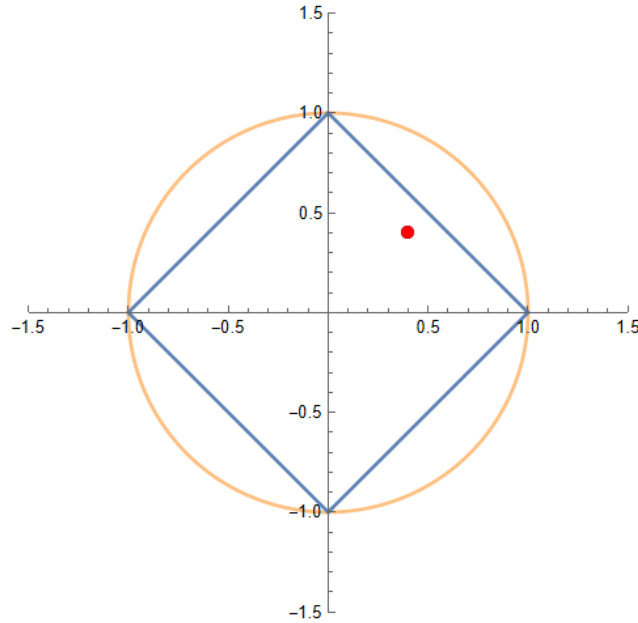
kde  $0 \leq t$  nám určuje poloměr koule (vzhledem k  $\ell_1$  normě), ze které vybíráme  $x$ ,  $\sum_{i=1}^m (y_i - \xi_0 - \sum_{j=1}^{n-1} \xi_j a_{i,j})^2 = RSC$  a, označíme funkci  $x_L(t)$  s parametrem  $t$  jako řešení dané úlohy pro zadané  $t$ .

Tato definice je vhodná pro pochopení geometrické interpretace. Ekvivalence by se dokazovala stejně jako pro hřebenovou regresi.

### 3.3.3 Geometrická interpretace Lassa a hřebenové regrese

Využitím definic 8 a 10 máme omezený prostor ve tvaru  $(n-1)$ -dimenzionální koule vzhledem k normám  $\ell_1, \ell_2$ , což jsou  $(n-1)$ -dimenzionální kvádr a koule.

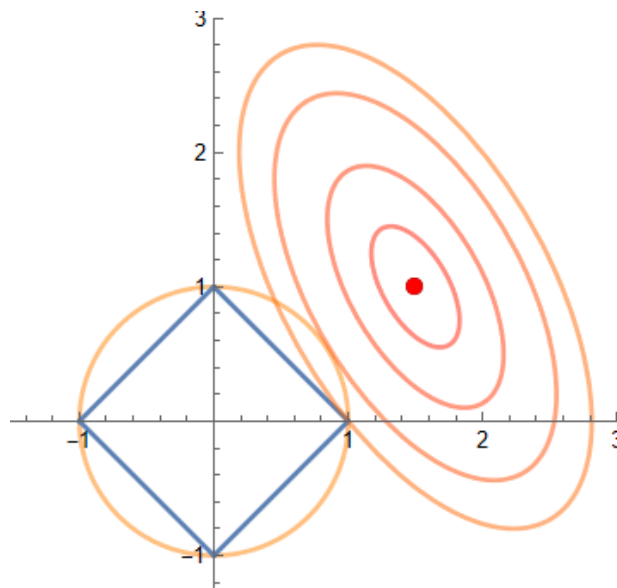
Pro představu budeme pracovat s  $n = 3$  (tedy jde o 2-dimenzionální prostor s osy  $\xi_1, \xi_2$ ). Pokud zvolíme  $t_L$  ( $t$  pro Lasso) a  $t_H$  ( $t$  pro hřebenovou regresi) dostatečně velké, pak  $x_{MN\check{C}} = x_L = x_H$ . Ukážeme na obrázku, proč tomu tak je.



Obrázek 3.1: Oranžová kružnice reprezentuje hranici množiny uspokojující podmínku pro hřebenovou regresi s  $t_H = 1$ , modrý rotovaný čtverec určuje hranici pro Lasso s  $t_L = 1$  a červený bod vyjadřuje hodnotu  $x_{MN\check{C}}$ .

Důvod, proč se  $x_{MN\check{C}} = x_L = x_H$ , je zřejmě viditelný, protože  $x_{MN\check{C}}$  minimalizuje RSC a splňuje podmínku z definice 8 nebo 10, pak díky ekvivalenci s definicí 7 a 9 musí být jejím řešením.

Zajímavý případ nastává, když  $x_{MNČ}$  leží mimo hranice sestrojené na základě podmínek, tím pádem je nesplňuje a tedy je nemůže řešit. V tomhle případě nás zajímá, jak takové řešení interpretovat geometricky.



Obrázek 3.2: Oranžová kružnice, rotovaný modrý čtverec a červený bod reprezentují to stejné jako obrázek 3.1 a elipsy se středem v  $x_{MNČ}$  jsou  $x$ - $y$ , které mají stejnou RSČ. Vidíme, že existuje taková elipsa, která se pouze dotýká daných útvarů. Bod doteku je přesně vyprodukované řešení dané metody.

Lze si všimnout, že když se pro metodu Lasso bude elipsa dotýkat v rohu, tak to znamená, že jeden z parametrů se volí za nulový. Šance, že to nastane, s klesajícím  $t_L$  se zvyšuje, když se ale podíváme na hřebenovou regresi, tak dochází k přesnému vynulování pouze ve velmi speciálních případech. Tyto úvahy platí i pro  $n$  větší než tři. Z toho se dá usoudit, že Lasso je mnohem lepší pro eliminaci parametrů než hřebenová regrese.

# Závěr

V této bakalářské práci jsme čtenáře seznámili s přeurlčenými a podurčenými systémy lineárních rovnic. U přeurlčených soustav jsme ukázali MNČ a praktické urychlení a zjednodušení použitím QR rozkladu. Ukázali jsme využití MNČ v lineární regresi a podpořili vypočtenými příklady. V jednom z mých příspěvků bylo úplné podrobné rozepsání a doplněním definic v důkazu, že MNČ generuje nejlepší nestranný lineární odhad. Nastínili jsme lehkou spojitost mezi lineární a nelineární regresi a vypočetli příklad pomocí Newtonovi metody s iterováním. Zmínili jsme i alternativní metody pro lineární regresi a to Theilovu a Senovu metodu nebo  $\ell_1$ -normou. Pro podurčenou soustavu jsme předvedli, jakými způsoby lze vybrat jednoznačné řešení z nekonečné množiny. Ukázali jsme, že MNČ minimalizuje  $\ell_2$ -normu, kde jsem přispěl dokázání, že  $x_{MNČ}$  je kolmé na množinu řešení. Nadále jsme zmínili krokové metody a jejich výhody a nevýhody. Nakonec jsem se zaobíral hřebenovou regresi a Lasso, vysvětlil rozdíly mezi metodami a vytvořil ilustrační obrázky, ze kterých čtenář získá vizuální intuici řešení problému.

# Seznam použité literatury

- ARENALES, M., MORABITO, R., ARMENTANO, V. a YANASSE, H. (2015). *Pesquisa Operacional: Para cursos de engenharia*. Elsevier Brasil. ISBN 9788535281835. URL <https://books.google.cz/books?id=aZbpCgAAQBAJ>.
- BAUM, C. F. (2004). EC771: Econometrics, Spring 2004. <http://fmwww.bc.edu/ec-c/S2004/771/EC771.S2004.nn4.pdf>.
- BHATTACHARYA, P. a BURMAN, P. (2016). *Theory and Methods of Statistics*. Elsevier Science. ISBN 9780128024409. URL <https://books.google.cz/books?id=XB7JsgEACAAJ>.
- BIRKES, D. a DODGE, Y. (1993). *Alternative methods of regression*. Wiley series in probability and mathematical statistics. Wiley, New York [u.a.], [2. print.] edition. ISBN 0471568813. URL [http://gso.gbv.de/DB=2.1/CMD?ACT=SRCHA&SRT=YOP&IKT=1016&TRM=ppn+121845656&sourceid=fbw\\_bibsonomy](http://gso.gbv.de/DB=2.1/CMD?ACT=SRCHA&SRT=YOP&IKT=1016&TRM=ppn+121845656&sourceid=fbw_bibsonomy).
- FLETCHER, R. (1987). *Practical Methods of Optimization*. John Wiley & Sons, New York, NY, USA, second edition.
- HILL, R. C., GRIFFITHS, W. E. a LIM, G. C. (2011). *Principles of Econometrics*. Čtvrté vydání. Wiley, USA. ISBN 978-0-470-62673-3.
- JAMES, G., WITTEN, D., HASTIE, T. a TIBSHIRANI, R. (2013). *An Introduction to Statistical Learning: with Applications in R*. Springer. URL <https://faculty.marshall.usc.edu/gareth-james/ISL/>.
- KLOUDA, K., STAROSTA, S. a VAŠATA, D. (2016). Matematika pro znalostní inženýrství. [https://kam.fit.cvut.cz/deploy/mi-mzi/MI-MZI\\_skripta.pdf](https://kam.fit.cvut.cz/deploy/mi-mzi/MI-MZI_skripta.pdf).
- SAAD, Y. a SCHULTZ, M. H. (1986). GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM Journal on Scientific and Statistical Computing*, **7**(3), 856–869. doi: 10.1137/0907058. URL <https://doi.org/10.1137/0907058>.
- YAN, X. a SU, X. (2009). *Linear regression analysis: theory and computing*. world scientific.