

UNIVERZITA PALACKÉHO V OLOMOUCI
PŘÍRODOVĚDECKÁ FAKULTA

DIPLOMOVÁ PRÁCE

Splajny v metodě nejmenších čtverců



Katedra matematické analýzy a aplikací matematiky

Vedoucí práce: **doc. RNDr. Jitka Machalová, Ph.D.**

Vypracovala: **Bc. Veronika Šmajserová**

Studijní program: N0541A170026 Aplikovaná matematika

Studijní obor: Aplikovaná matematika

Forma studia: prezenční

Rok odevzdání: 2024

BIBLIOGRAFICKÁ IDENTIFIKACE

Autor: Bc. Veronika Šmajserová

Název práce: Splajny v metodě nejmenších čtverců

Typ práce: Diplomová práce

Pracoviště: Katedra matematické analýzy a aplikací matematiky

Vedoucí práce: doc. RNDr. Jitka Machalová, Ph.D.

Rok obhajoby práce: 2024

Abstrakt: Cílem diplomové práce je ukázat použití splajnů ve standardní a penalizované metodě nejmenších čtverců. Důraz je kladen také na predikci v penalizované regresi. Teorie je doplněna praktickými příklady a vlastními kódy sestavenými v matematickém softwaru MATLAB.

Klíčová slova: splajny, metoda nejmenších čtverců, penalizovaná metoda nejmenších čtverců, P-splajny, penalizovaná regrese, přístup chybějící hodnoty, lineární smíšené modely, hladké modely smíšených efektů, dvoufázový přístup, jednofázový přístup

Počet stran: 97

Počet příloh: 1

Jazyk: český

BIBLIOGRAPHICAL IDENTIFICATION

Author: Bc. Veronika Šmajserová

Title: Least-squares spline approximation

Type of thesis: Master's

Department: Department of Mathematical Analysis and Application of Mathematics

Supervisor: doc. RNDr. Jitka Machalová, Ph.D.

The year of presentation: 2024

Abstract: The aim of the thesis is the application of splines in the standard and penalized least squares methods. Emphasis is also placed on prediction in penalized regression. The theory is complemented by practical examples and original codes, which are constructed in the mathematical software MATLAB.

Key words: splines, least squares method, penalized least squares method, P-splines, penalized regression, the missing value approach, linear mixed models, mixed effects smooth models, two-stage approach, one-stage approach

Number of pages: 97

Number of appendices: 1

Language: Czech

Prohlášení

Prohlašuji, že jsem diplomovou práci zpracovala samostatně pod vedením paní doc. RNDr. Jitky Machalové, Ph.D. a všechny použité zdroje jsem uvedla v seznamu literatury.

V Olomouci dne
.....
podpis

Obsah

Úvod	8
1 Základní teorie	10
1.1 Základní poznatky z lineární algebry	10
1.2 Polynomiální splajny	13
1.3 Splajny v metodě nejmenších čtverců	16
2 Splajny v penalizované metodě nejmenších čtverců	22
2.1 P -splajny	22
3 Predikování v penalizované regresi	28
3.1 Přístup chybějící hodnoty	28
4 Predikce pomocí hladkých modelů smíšených efektů	54
4.1 Lineární smíšený model	54
4.1.1 Odhad pevných efektů	55
4.1.2 Odhad náhodných efektů	57
4.1.3 Odhad parametrů rozptylů σ_ϵ^2 a σ_α^2	60
4.2 Hladké modely smíšených efektů	61
4.3 Dvoufázový přístup	63
4.3.1 Standardní metodologie pro predikce	63
4.3.2 Predikce založená na podmíněném rozdělení $\mathbf{y}_p \mathbf{y}$	68
4.4 Jednofázový přístup	69
5 Praktická část	80
5.1 Nalezení splajnu, který data aproximuje ve smyslu penalizované metody nejmenších čtverců	81
5.2 Doplnění chybějících hodnot pomocí přístupu chybějící hodnoty	84
5.3 Doplnění chybějících hodnot pomocí hladkých modelů smíšených efektů	86
5.4 Srovnání obou přístupů	89
Závěr	92
Literatura	94

Poděkování

Na tomto místě bych ráda poděkovala paní doc. RNDr. Jitce Machalové, Ph.D. za její cenné rady, trpělivost a čas, který mi během psaní diplomové práce věnovala. Také bych chtěla poděkovat svým blízkým za podporu během celého studia.

Seznam použitých symbolů

\mathbb{N}	množina všech přirozených čísel
\mathbb{N}_0	množina všech celých nezáporných čísel
\mathbb{R}^n	množina všech reálných vektorů délky n
$\mathbf{0}$	nulový vektor
$\mathcal{R}^{m,n}$	množina všech reálných matic typu $m \times n$
\mathbf{I}_n	jednotková matice typu $n \times n$
$\mathbf{1}_n$	sloupcový vektor jedniček délky n
$\mathbf{0}_{m,n}$	nulová matice typu $m \times n$
\mathbf{A}^-	zobecněná inverze matice \mathbf{A}
$\mathcal{R}(\mathbf{A})$	řádkový prostor matice \mathbf{A}
$\mathcal{C}(\mathbf{A})$	sloupcový prostor matice \mathbf{A}
$ \mathbf{A} $	determinant matice \mathbf{A}
$\mathbf{A} \otimes \mathbf{B}$	tenzorový součin matic \mathbf{A} a \mathbf{B}

Úvod

V dnešní době jsou data všude kolem nás a existuje mnoho metod, jak je analyzovat. Velmi často je cílem schopnost predikovat data pro budoucí pozorování, ale poměrně běžná je i situace, kdy data obsahují nějaká chybějící pozorování a tyto hodnoty je třeba nějakým způsobem doplnit. Hlavním cílem diplomové práce je popsat použití splajnů v klasické i penalizované metodě nejmenších čtverců a následně nastudovat některé algoritmy pro predikci. Popsané algoritmy jsou aplikovány na příkladech a v závěru práce také na konkrétní datové sadě. Kódy jsou psány v programovacím jazyku Matlab a předpokládá se, že čtenář zná základní příkazy a umí pracovat se softwarem Matlab.

První kapitola je věnovaná základní teorii, přičemž se předpokládá, že čtenář je seznámen se základními koncepty z oblasti lineární algebry. Některé podstatné definice a věty z této oblasti matematiky jsou zmíněny v odstavci 1.1. Následuje část práce, ve které jsou definovány a krátce popsány B -splajny. Poslední část první kapitoly se zabývá použitím splajnů v klasické metodě nejmenších čtverců, přičemž v celé práci předpokládáme, že splajny jsou dány jako lineární kombinace B -splajnů.

Ve druhé kapitole je popsáno použití splajnů v penalizované metodě nejmenších čtverců, konkrétně se ve zbývajících částech práce věnujeme tzv. P -splajnům. Následuje třetí kapitola, ve které je popsána predikce v penalizované regresi. Tzv. přístup chybějící hodnoty je teoreticky popsán především pro situaci, kdy chceme predikovat hodnoty pro budoucí pozorování. V příkladech je dále rozebrána i situace, kdy je třeba predikovat hodnoty pro pozorování z minulosti, nebo situace, kdy data obsahují nějaká chybějící pozorování.

Poslední teoretická kapitola se věnuje predikci pomocí hladkých modelů smíšených efektů, které představují propojení P -splajnů a lineárních smíšených modelů. Proto jsou v první části popsány právě lineární smíšené modely a důraz je kladen především na odhadování pevných a náhodných efektů a také parametrů rozptylů náhodné složky a náhodných efektů. Poté již následuje část, která se zaměřuje přímo na hladké modely smíšených efektů. Pro predikci hodnot jsou zde zmíněny dva přístupy. Nejprve je popsán tzv. dvoufázový přístup, který je založen na tom, že se nejprve vytvoří model a až poté se provádí predikce. Princip druhého přístupu, tzv. jednofázového, spočívá v tom, že tvorba modelu a predikce jsou prováděny současně.

Práce obsahuje také praktickou část. K dispozici máme data obsahující chybějící části, které je třeba doplnit. Nejprve je predikce provedena pomocí přístupu chybějící hodnoty a poté pomocí hladkých modelů smíšených efektů.

1. Základní teorie

V této kapitole uvedeme definice základních pojmů, které budeme v diplomové práci využívat.

1.1. Základní poznatky z lineární algebry

Nejprve si uvedeme základní definice z teorie matic, které jsou využity v dalších částech práce. Pokud není uvedeno jinak, vycházíme z [9]. Začneme definicemi ortogonální matice a zobecněné inverze matice.

Definice 1.1. Necht' $\mathbf{A} \in \mathcal{R}^{n,n}$. Matici \mathbf{A} nazveme ortogonální, jestliže platí $\mathbf{A}^T \mathbf{A} = \mathbf{A} \mathbf{A}^T = \mathbf{I}_n$.

Definice 1.2. Necht' $\mathbf{A} \in \mathcal{R}^{m,n}$. Zobecněná inverze matice \mathbf{A} je každá matice $\mathbf{A}^- \in \mathcal{R}^{n,m}$ taková, že $\mathbf{A} \mathbf{A}^- \mathbf{A} = \mathbf{A}$.

Nyní si zadefinujeme sloupcový a řádkový prostor matice.

Definice 1.3. Necht' $\mathbf{A} \in \mathcal{R}^{m,n}$. Sloupcový prostor matice \mathbf{A} , který značíme $\mathcal{C}(\mathbf{A})$, je lineární obal sloupcových vektorů matice \mathbf{A} , tj. $\mathcal{C}(\mathbf{A}) = \{\mathbf{y} \in \mathbb{R}^m \mid \exists \mathbf{x} \in \mathbb{R}^n : \mathbf{y} = \mathbf{A}\mathbf{x}\}$. Analogicky definujeme i řádkový prostor matice \mathbf{A} , který značíme $\mathcal{R}(\mathbf{A})$, jako lineární obal řádkových vektorů matice \mathbf{A} , tj. $\mathcal{R}(\mathbf{A}) = \{\mathbf{x}^T \in \mathbb{R}^n \mid \exists \mathbf{y} \in \mathbb{R}^m : \mathbf{x}^T = \mathbf{y}^T \mathbf{A}\}$.

Dimenzi sloupcového prostoru $\mathcal{C}(\mathbf{A})$ nazýváme sloupcovou hodnotí matice \mathbf{A} a podobně řádkovou hodnotí matice \mathbf{A} rozumíme dimenzi řádkového prostoru $\mathcal{R}(\mathbf{A})$. Jestliže se sloupcová a řádková hodnota rovnají, jejich společnou hodnotu nazýváme hodnotí matice a značíme ji $h(\mathbf{A})$. Ekvivalentně lze hodnotu matice definovat následovně.

Definice 1.4. Necht' $\mathbf{A} \in \mathcal{R}^{m,n}$. Hodností matice \mathbf{A} rozumíme číslo $h(\mathbf{A}) \in \mathbb{N}$, které udává maximální počet lineárně nezávislých řádků nebo sloupců matice \mathbf{A} .

Definice 1.5. Necht' $\mathbf{A} \in \mathcal{R}^{m,n}$. Řekneme, že matice \mathbf{A} je plně řádkové hodnosti, jestliže $h(\mathbf{A}) = m$. Analogicky řekneme, že matice \mathbf{A} je plně sloupcové hodnosti, jestliže $h(\mathbf{A}) = n$.

Hodnost matice je využita i v definici regulární matice.

Definice 1.6. Necht' $\mathbf{A} \in \mathcal{R}^{n,n}$. Matice \mathbf{A} je regulární právě tehdy, když $h(\mathbf{A}) = n$.

Než budeme pokračovat s větami, uvedeme následující lemma.

Lemma 1.1. Pro regulární matici \mathbf{A} platí $\mathbf{A}^{-1} = \mathbf{A}^-$.

Nyní uvedeme několik vět, které jsou v práci později využity. Důkaz první z nich může čtenář najít v [9] na straně 125.

Věta 1.1. Necht' jsou dány matice $\mathbf{T} \in \mathcal{R}^{m,p}$, $\mathbf{U} \in \mathcal{R}^{m,q}$, $\mathbf{V} \in \mathcal{R}^{n,p}$ a $\mathbf{W} \in \mathcal{R}^{n,q}$. Definujme matici $\mathbf{Q} = \mathbf{W} - \mathbf{V}\mathbf{T}^- \mathbf{U}$ a předpokládejme, že $\mathcal{C}(\mathbf{U}) \subset \mathcal{C}(\mathbf{T})$ a $\mathcal{R}(\mathbf{V}) \subset \mathcal{R}(\mathbf{T})$. Pak platí

$$h \left(\begin{bmatrix} \mathbf{T} & \mathbf{U} \\ \mathbf{V} & \mathbf{W} \end{bmatrix} \right) = h \left(\begin{bmatrix} \mathbf{W} & \mathbf{V} \\ \mathbf{U} & \mathbf{T} \end{bmatrix} \right) = h(\mathbf{T}) + h(\mathbf{Q}).$$

Navíc platí, že matice

$$\begin{bmatrix} \mathbf{T}^- + \mathbf{T}^- \mathbf{U} \mathbf{Q}^- \mathbf{V} \mathbf{T}^- & -\mathbf{T}^- \mathbf{U} \mathbf{Q}^- \\ -\mathbf{Q}^- \mathbf{V} \mathbf{T}^- & \mathbf{Q}^- \end{bmatrix} = \begin{bmatrix} \mathbf{T}^- & \mathbf{0}_{p,n} \\ \mathbf{0}_{q,m} & \mathbf{0}_{q,n} \end{bmatrix} + \begin{bmatrix} -\mathbf{T}^- \mathbf{U} \\ \mathbf{I}_q \end{bmatrix} \mathbf{Q}^- [-\mathbf{V} \mathbf{T}^-, \mathbf{I}_n],$$

respektive

$$\begin{bmatrix} \mathbf{Q}^- & -\mathbf{Q}^- \mathbf{V} \mathbf{T}^- \\ -\mathbf{T}^- \mathbf{U} \mathbf{Q}^- & \mathbf{T}^- + \mathbf{T}^- \mathbf{U} \mathbf{Q}^- \mathbf{V} \mathbf{T}^- \end{bmatrix} = \begin{bmatrix} \mathbf{0}_{q,n} & \mathbf{0}_{q,m} \\ \mathbf{0}_{p,n} & \mathbf{T}^- \end{bmatrix} + \begin{bmatrix} \mathbf{I}_q \\ -\mathbf{T}^- \mathbf{U} \end{bmatrix} \mathbf{Q}^- [\mathbf{I}_n, -\mathbf{V} \mathbf{T}^-],$$

jsou zobecněné inverze matic $\begin{bmatrix} \mathbf{T} & \mathbf{U} \\ \mathbf{V} & \mathbf{W} \end{bmatrix}$, respektive $\begin{bmatrix} \mathbf{W} & \mathbf{V} \\ \mathbf{U} & \mathbf{T} \end{bmatrix}$.

Abychom mohli využít předchozí větu, je třeba ověřit všechny její předpoklady. K tomu nám může pomoci následující lemma, které lze najít v [9] na straně 30.

Lemma 1.2. *Nechť $\mathbf{T} \in \mathcal{R}^{m,p}$, $\mathbf{U} \in \mathcal{R}^{m,q}$ a $\mathbf{V} \in \mathcal{R}^{n,p}$. Pak $\mathcal{C}(\mathbf{U}) \subset \mathcal{C}(\mathbf{T})$ právě tehdy, když existuje $\mathbf{F} \in \mathcal{R}^{p,q}$ taková, že $\mathbf{U} = \mathbf{T}\mathbf{F}$ a analogicky $\mathcal{R}(\mathbf{V}) \subset \mathcal{R}(\mathbf{T})$ právě tehdy, když existuje $\mathbf{L} \in \mathcal{R}^{n,m}$ taková, že $\mathbf{V} = \mathbf{L}\mathbf{T}$.*

Za předpokladu, že jsou blokové matice uvažované ve Větě 1.1 regulární, lze místo zobecněné inverze uvažovat klasickou inverzi. K ověření, zda jsou tyto matice regulární, můžeme využít následující věty, kterou lze najít v [9] na straně 99.

Věta 1.2. *Nechť $\mathbf{T} \in \mathcal{R}^{m,m}$, $\mathbf{U} \in \mathcal{R}^{m,n}$, $\mathbf{V} \in \mathcal{R}^{n,m}$ a $\mathbf{W} \in \mathcal{R}^{n,n}$ a předpokládejme, že matice \mathbf{T} je regulární. Pak jsou matice*

$$\begin{bmatrix} \mathbf{T} & \mathbf{U} \\ \mathbf{V} & \mathbf{W} \end{bmatrix}, \text{ respektive } \begin{bmatrix} \mathbf{W} & \mathbf{V} \\ \mathbf{U} & \mathbf{T} \end{bmatrix},$$

regulární právě tehdy, když je regulární matice $\mathbf{Q} = \mathbf{W} - \mathbf{V}\mathbf{T}^{-1}\mathbf{U}$.

Nyní již přejdeme ke třetí větě, která je uvedena v [9] na straně 424.

Věta 1.3. *Nechť jsou dány matice $\mathbf{W} \in \mathcal{R}^{n,n}$, $\mathbf{S} \in \mathcal{R}^{n,m}$, $\mathbf{T} \in \mathcal{R}^{m,m}$ a $\mathbf{U} \in \mathcal{R}^{m,n}$. Předpokládejme, že matice \mathbf{W} a \mathbf{T} jsou regulární. Pak $\mathbf{W} + \mathbf{S}\mathbf{T}\mathbf{U}$ je*

regulární právě tehdy, když je regulární matice $\mathbf{T}^{-1} + \mathbf{U}\mathbf{W}^{-1}\mathbf{S}$, a platí

$$(\mathbf{W} + \mathbf{S}\mathbf{T}\mathbf{U})^{-1} = \mathbf{W}^{-1} - \mathbf{W}^{-1}\mathbf{S}(\mathbf{T}^{-1} + \mathbf{U}\mathbf{W}^{-1}\mathbf{S})^{-1}\mathbf{U}\mathbf{W}^{-1}.$$

1.2. Polynomiální splajny

V této části kapitoly si uvedeme základní teorii týkající se splajnu, přičemž pokud není stanoveno jinak, čerpáme z [3] a [4]. Nejprve uvedeme definici sítě uzlů potřebnou k definování polynomického splajnu.

Definice 1.7. Nechť je dán uzavřený interval $[a_1, a_2]$. Sítí uzlů, kterou značíme $\Delta\lambda$, rozumíme body $\lambda_i \in [a_1, a_2]$, kde $i = 0, 1, \dots, g + 1$, které splňují

$$\Delta\lambda : a_1 = \lambda_0 < \lambda_1 < \dots < \lambda_g < \lambda_{g+1} = a_2. \quad (1.1)$$

Definice 1.8. Na síti uzlů $\Delta\lambda$ nazveme funkci $s_k(x)$ splajnem stupně k , kde $k \in \mathbb{N}_0$, jestliže splňuje:

1. $s_k(x)$ je na každém intervalu $[\lambda_i, \lambda_{i+1}]$, $i = 0, \dots, g$, polynom stupně nejvýše k ,
2. $s_k(x)$ je na intervalu $[a_1, a_2]$ spojitá a má na tomto intervalu spojitě derivace až do řádu $k - 1$.

Symbolem $S_k^{\Delta\lambda}[a_1, a_2]$ značíme lineární prostor splajnu $s_k(x)$ stupně k definovaných na intervalu $[a_1, a_2]$ na dané síti uzlů $\Delta\lambda$.

Dimenze tohoto prostoru je dle [4] rovna

$$\dim(S_k^{\Delta\lambda}[a_1, a_2]) = g + k + 1. \quad (1.2)$$

Splajn lze tedy vyjádřit jako lineární kombinaci báзовých funkcí uvažovaného prostoru $S_k^{\Delta\lambda}[a_1, a_2]$. My budeme jako báзовé funkce používat B -splajny, které jsou definované pomocí funkce zvané useknutá mocnina.

Definice 1.9. Funkce $(t - x)_+^k$ proměnné t s daným x , která je definovaná jako

$$(t - x)_+^k = \begin{cases} (t - x)^k, & \text{jestliže } t \geq x \\ 0, & \text{jestliže } t < x, \end{cases}$$

se nazývá useknutá mocnina.

Než se dostaneme k definici B -splajnu, je nutné definovat poměrnou diferenci. V této části jsme opět vycházeli z [3].

Definice 1.10. Nechť je dán interval $[a_1, a_2]$ a na něm síť uzlů $\Delta\lambda$, ve kterých je funkce f definovaná. Poměrná diference k -tého řádu, $k \in \mathbb{N}$, funkce f na síti uzlů $\Delta\lambda$, viz (1.1), je definovaná vztahem

$$f[\lambda_i, \dots, \lambda_{i+k}] = \frac{f[\lambda_{i+1}, \dots, \lambda_{i+k}] - f[\lambda_i, \dots, \lambda_{i+k-1}]}{\lambda_{i+k} - \lambda_i}$$

pro $i = 0, \dots, g - k + 1$, přičemž poměrná diference 1. řádu funkce f na síti uzlů $\Delta\lambda$, viz (1.1), je definovaná jako

$$f[\lambda_i, \lambda_{i+1}] = \frac{f(\lambda_{i+1}) - f(\lambda_i)}{\lambda_{i+1} - \lambda_i}.$$

Pro poměrnou diferenci k -tého řádu $f[\lambda_i, \dots, \lambda_{i+k}]$, $i = 0, \dots, g - k + 1$, platí, že ji lze vyjádřit jako lineární kombinaci funkčních hodnot funkce f

v uzlech λ_{i+j} pro $j = 0, \dots, k$, konkrétně

$$f[\lambda_i, \dots, \lambda_{i+k}] = \sum_{j=0}^k \frac{f(\lambda_{i+j})}{\prod_{\substack{l=0 \\ l \neq j}}^k (\lambda_{i+j} - \lambda_{i+l})}. \quad (1.3)$$

Nyní již můžeme přejít k definici B -splajnu.

Definice 1.11. B -splajn $B_i^{k+1}(x)$ stupně $k \in \mathbb{N}_0$ s uzly $\lambda_i, \lambda_{i+1}, \dots, \lambda_{i+k+1}$ je definován jako funkce

$$B_i^{k+1}(x) = (\lambda_{i+k+1} - \lambda_i)(t - x)_+^k [\lambda_i, \lambda_{i+1}, \dots, \lambda_{i+k+1}].$$

Užitím explicitního vyjádření pro poměrné difference (1.3) lze psát

$$B_i^{k+1}(x) = (\lambda_{i+k+1} - \lambda_i) \sum_{j=0}^{k+1} \frac{(\lambda_{i+j} - x)_+^k}{\prod_{\substack{l=0 \\ l \neq j}}^{k+1} (\lambda_{i+j} - \lambda_{i+l})}.$$

B -splajny mají několik výhodných vlastností, mezi které patří například numerická stabilita, jejich lineární nezávislost nebo konečný nosič. Proto je vhodné tyto funkce použít jako báze funkce pro vyjádření splajnu.

Nechť je dána síť uzlů $\Delta\lambda$, viz (1.1). Na této síti uzlů je možné sestavit pouze $g - k + 1$ lineárně nezávislých B -splajnů $B_i^{k+1}(x)$, kde $i = 0, \dots, g - k$, ale dimenze prostoru $S_k^{\Delta\lambda}[a_1, a_2]$ je rovna $g + k + 1$, viz (1.2). Abychom získali celou bázi tohoto prostoru, je nutné přidat ještě $2k$ lineárně nezávislých funkcí, konkrétně tedy $2k$ lineárně nezávislých B -splajnů. Toho docílíme tak, že přidáme $2k$ tzv. pomocných uzlů, které označíme $\lambda_{-k}, \dots, \lambda_{-1}$

a $\lambda_{g+2}, \dots, \lambda_{g+k+1}$, tak, aby platilo

$$\begin{aligned}\lambda_{-k} &\leq \lambda_{-k+1} \leq \dots \leq \lambda_{-1} \leq \lambda_0, \\ \lambda_{g+1} &\leq \lambda_{g+2} \leq \dots \leq \lambda_{g+k} \leq \lambda_{g+k+1},\end{aligned}$$

čímž získáme tzv. rozšířenou síť uzlů. Tyto pomocné uzly mohou být násobné. Pak platí, že každý splajn $s_k(x) \in S_k^{\Delta\lambda}[a_1, a_2]$ lze vyjádřit jako lineární kombinaci B -splajnů, tj.

$$s_k(x) = \sum_{i=-k}^g b_i B_i^{k+1}(x),$$

což lze maticově zapsat jako

$$s_k(x) = \mathbf{C}_{k+1}(x)\mathbf{b},$$

kde \mathbf{b} je vektor B -splajnových koeficientů b_i , tj. $\mathbf{b} = (b_{-k}, \dots, b_g)^T$, a $\mathbf{C}_{k+1}(x)$ je tzv. kolokační matice, která má pro reálný bod x podobu řádkového vektoru $\mathbf{C}_{k+1}(x) = (B_{-k}^{k+1}(x), \dots, B_g^{k+1}(x))$.

1.3. Splajny v metodě nejmenších čtverců

Nyní již přejdeme k použití splajnů v metodě nejmenších čtverců. Předpokládejme, že jsou dány body x_j , které jsou navzájem různé a leží v uzavřeném intervalu $[a_1, a_2]$, a v nich funkční hodnoty y_j pro $j = 1, \dots, n$. Naším cílem je aproximovat tato data vhodným splajnem.

Uvažujme síť uzlů $\Delta\lambda$ definovanou na intervalu $[a_1, a_2]$, viz. (1.1), a $k \in \mathbb{N}_0$ jako stupeň splajnu. Aby měla aproximace smysl, tj. aby nebylo možné získat splajn, který přesně prochází danými body, předpokládejme, že počet bodů je větší než dimenze prostoru $S_k^{\Delta\lambda}[a_1, a_2]$, tj. $n > g + k + 1$.

Naším úkolem je najít splajn $s_k(x) \in S_k^{\Delta\lambda}[a_1, a_2]$, který minimalizuje vý-

raz

$$\sum_{j=1}^n [y_j - s_k(x_j)]^2, \quad (1.4)$$

který odpovídá součtu čtverců odchylek, proto metodu označujeme jako metodu nejmenších čtverců. Protože budeme uvažovat splajny $s_k(x)$ ve tvaru lineární kombinace B -splajnů, lze tento výraz zapsat jako

$$\sum_{j=1}^n \left[y_j - \sum_{i=-k}^g b_i B_i^{k+1}(x_j) \right]^2$$

a označit jej $\delta(b_{-k}, \dots, b_g)$.

Dále zavedeme značení pro sloupcové vektory $\mathbf{y} = (y_1, \dots, y_n)^T$ a $\mathbf{x} = (x_1, \dots, x_n)^T$, kolokační matici $\mathbf{C}_{k+1}(\mathbf{x})$, která je pro vektor \mathbf{x} ve tvaru

$$\mathbf{C}_{k+1}(\mathbf{x}) = \begin{pmatrix} B_{-k}^{k+1}(x_1) & \cdots & B_g^{k+1}(x_1) \\ \vdots & \ddots & \vdots \\ B_{-k}^{k+1}(x_n) & \cdots & B_g^{k+1}(x_n) \end{pmatrix},$$

a odpovídající vektor B -splajnových koeficientů $\mathbf{b} = (b_{-k}, \dots, b_g)^T$. Potom můžeme pomocí Euklidovské normy minimalizovaný výraz přepsat do tvaru

$$\delta(\mathbf{b}) = \|\mathbf{y} - \mathbf{C}_{k+1}(\mathbf{x})\mathbf{b}\|^2 = (\mathbf{y} - \mathbf{C}_{k+1}(\mathbf{x})\mathbf{b})^T (\mathbf{y} - \mathbf{C}_{k+1}(\mathbf{x})\mathbf{b}). \quad (1.5)$$

Roznásobením pak lze přejít k následujícímu tvaru minimalizovaného výrazu, ze kterého je patrné, že se jedná o kvadratickou funkci proměnné \mathbf{b}

$$\delta(\mathbf{b}) = \mathbf{y}^T \mathbf{y} - 2\mathbf{y}^T \mathbf{C}_{k+1}(\mathbf{x})\mathbf{b} + \mathbf{b}^T \mathbf{C}_{k+1}^T(\mathbf{x})\mathbf{C}_{k+1}(\mathbf{x})\mathbf{b}.$$

Dle [14] platí, že jestliže je matice $\mathbf{C}_{k+1}(\mathbf{x})$ plně sloupcové hodnosti, pak existuje právě jeden splajn daný vektorem \mathbf{b} , který minimalizuje výraz $\delta(\mathbf{b})$,

neboli kvadratická funkce $\delta(\mathbf{b})$ je ryze konvexní. Abychom zjistili, kdy je tato matice plně sloupcové hodnosti, využijeme tzv. Schoenberg-Whitneyho podmínku z [4], dle které platí, že matice $\mathbf{C}_{k+1}(\mathbf{x})$ je plně sloupcové hodnosti právě tehdy, když $n > g + k + 1$ a existuje množina bodů $\{u_{-k}, \dots, u_g\} \subset \{x_1, \dots, x_n\}$, kde $u_i < u_{i+1}$, $i = -k, \dots, g - 1$, taková, že $\lambda_i < u_i < \lambda_{i+k+1}$, $i = -k, \dots, g$. Z toho plyne, že matice $\mathbf{C}_{k+1}(\mathbf{x})$ je plně sloupcové hodnosti právě tehdy, když $n > g + k + 1$, což jsme předpokládali, a když v nosiči každého B -splajnu leží alespoň jeden bod x_j , kde $j = 1, \dots, n$.

Za těchto podmínek tedy existuje jediný splajn daný vektorem $\hat{\mathbf{b}}$, který minimalizuje výraz $\delta(\mathbf{b})$. Vektor $\hat{\mathbf{b}}$ lze určit z rovnice pro stacionární bod

$$-2\mathbf{C}_{k+1}^T(\mathbf{x})\mathbf{y} + 2\mathbf{C}_{k+1}^T(\mathbf{x})\mathbf{C}_{k+1}(\mathbf{x})\hat{\mathbf{b}} = \mathbf{0},$$

kteřou jsme obdrželi tím, že jsme derivaci funkce $\delta(\mathbf{b})$ podle \mathbf{b}^T položili rovnu nulovému vektoru. Řešením této rovnice je

$$\hat{\mathbf{b}} = (\mathbf{C}_{k+1}^T(\mathbf{x})\mathbf{C}_{k+1}(\mathbf{x}))^{-1} \mathbf{C}_{k+1}^T(\mathbf{x})\mathbf{y}, \quad (1.6)$$

přičemž víme, že díky plně sloupcové hodnosti matice $\mathbf{C}_{k+1}(\mathbf{x})$ je matice $\mathbf{C}_{k+1}^T(\mathbf{x})\mathbf{C}_{k+1}(\mathbf{x})$ pozitivně definitní [9], str. 213, a tedy je regulární a existuje k ní inverzní matice. Hledaný splajn, který minimalizuje (1.4), je pak ve tvaru

$$\hat{s}_k(x) = \sum_{i=-k}^g \hat{b}_i B_i^{k+1}(x),$$

kde \hat{b}_i , $i = -k, \dots, g$, jsou prvky vektoru $\hat{\mathbf{b}}$, tj. $\hat{\mathbf{b}} = (\hat{b}_{-k}, \dots, \hat{b}_g)^T$. O takovém splajnu řekneme, že aproximuje daná data ve smyslu metody nejmenších čtverců.

Příklad 1.1. Metodu nejmenších čtverců nyní ilustrujeme na příkladu. Uvažujme situaci, kdy máme dán vektor bodů \mathbf{x} a vektor funkčních hodnot \mathbf{y} , konkrétně

$$\mathbf{x} = [0, 1, 2, 4, 6, 8, 9, 11, 13, 15, 16, 17, 18, 19, 20]^T,$$

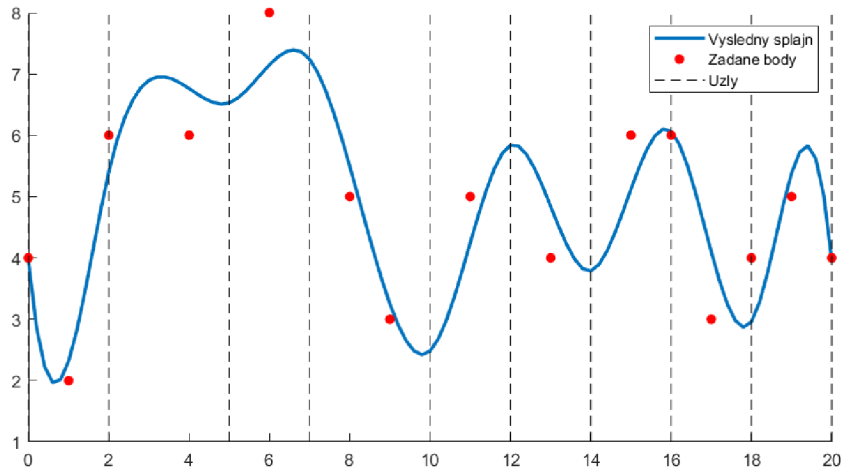
$$\mathbf{y} = [4, 2, 6, 6, 8, 5, 3, 5, 4, 6, 6, 3, 4, 5, 4]^T,$$

tj. $n = 15$, a chceme najít splajn stupně 3, tj. $k = 3$, takový, že aproximuje data ve smyslu metody nejmenších čtverců. Síť uzlů $\Delta\lambda$ zvolíme jako vektor $[0, 2, 5, 7, 10, 12, 14, 16, 18, 20]$, tedy $g = 8$. Jak jsme zmínili v této kapitole, je třeba vytvořit tzv. rozšířenou síť uzlů. V našem případě to znamená přidat 6 uzlů, které mohou být i násobné. K tomu využijeme program MATLAB a jeho funkci `augknt`, pomocí které získáme rozšířenou síť uzlů

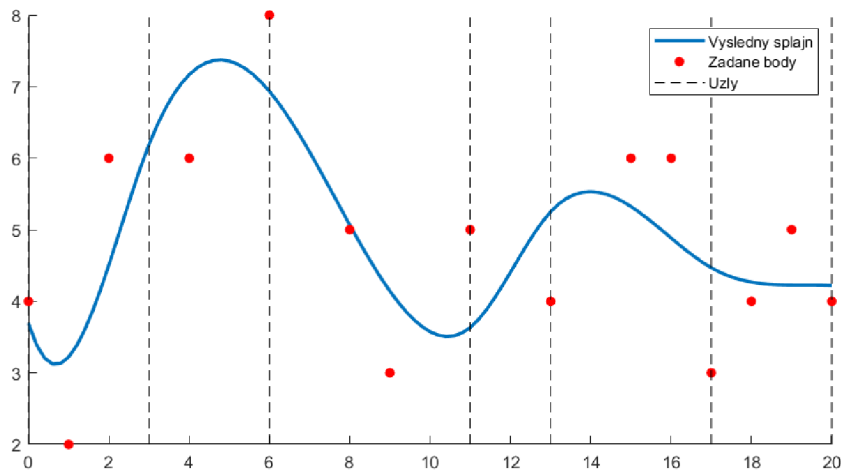
$$[0, 0, 0, 0, 2, 5, 7, 10, 12, 14, 16, 18, 20, 20, 20, 20].$$

Dále využijeme funkci `spsol` k vytvoření kolokační matice $\mathbf{C}_4(\mathbf{x})$, která je typu 15×12 , neboť $g+k+1 = 12$. Protože je splněna Schoenberg-Whitneyho podmínka, je tato matice plně sloupcové hodnosti, což lze ověřit i pomocí funkce `rank`.

Matici $\mathbf{C}_4(\mathbf{x})$ následně použijeme k určení vektoru B -splajnových koeficientů $\hat{\mathbf{b}}$ prostřednictvím rovnice (1.6), neboli $\mathbf{C}_{k+1}^T(\mathbf{x})\mathbf{C}_{k+1}(\mathbf{x})\hat{\mathbf{b}} = \mathbf{C}_{k+1}^T(\mathbf{x})\mathbf{y}$. Řešení získáme pomocí funkce `linsolve`, abychom se vyhnuli přímému počítání inverzní matice. Pomocí funkce `spmak` vytvoříme splajn, na který aplikujeme funkci `fnplt`, čímž získáme grafickou reprezentaci výsledného splajnu, viz obrázek 1. Nakonec by nás mohla zajímat hodnota minimalizovaného výrazu (1.5), ta je v našem případě rovna 6.576.



Obrázek 1: Aproximace splajnem ve smyslu metody nejmenších čtverců



Obrázek 2: Aproximace splajnem ve smyslu metody nejmenších čtverců

Pro porovnání můžeme vyzkoušet jinou volbu uzlů. Uvažujme $g = 5$ a síť uzlů $[0, 3, 6, 11, 13, 17, 20]$. Dále pokračujeme analogicky, výsledný splajn je zobrazen na obrázku 2. Hodnota minimalizovaného výrazu je v tomto případě 15.628, dostali jsme tedy větší hodnotu než s první volbou uzlů, což lze pozorovat i porovnáním grafů, tj. porovnáním velikostí čtverců odchylek.

Celkem jsme tedy na jednoduchém příkladu ukázali, jak volit stupeň splajnu a uzly tak, aby existovalo právě jedno řešení rovnice (1.6), a tedy aby existoval právě jeden splajn, který daná data aproximuje ve smyslu metody nejmenších čtverců. V Příkladu 1.1 jsme navíc ukázali, jak úlohy tohoto typu řešit v programu MATLAB a jak graficky znázornit jejich řešení. Dále jsme také naznačili, že je vhodné vyzkoušet různé volby sítě uzlů a také stupně splajnu, a z nich poté vybrat vhodnou variantu s ohledem na hodnotu minimalizovaného výrazu (1.5).

2. Splajny v penalizované metodě nejmenších čtverců

V této kapitole se zaměříme na použití splajnů v penalizované metodě nejmenších čtverců. Tato metoda slouží k nalezení splajnu, který aproximuje daná data ve smyslu metody nejmenších čtverců stejně, jako tomu bylo v podkapitole 1.3, přičemž na výsledný splajn navíc klademe podmínky na jeho hladkost prostřednictvím penalizačního členu. Tento penalizační člen lze volit různě.

Nechť jsou dány body x_j z intervalu $[a_1, a_2]$ a v nich funkční hodnoty y_j pro $j = 1, \dots, n$. Označme $\mathbf{x} = (x_1, \dots, x_n)^T$ a $\mathbf{y} = (y_1, \dots, y_n)^T$.

Jedna možnost, jak u penalizované metody nejmenších čtverců penalizovat minimalizovaný výraz, je pomocí L_2 normy l -té derivace. Nechť ρ je tzv. vyhlazující parametr, kde $\rho > 0$, a l je uvažovaný řád derivace, kde $l \in \{0, 1, \dots, k-1\}$. Potom hledáme splajn $s_k(x) \in S_k^{\Delta\lambda}[a_1, a_2]$, který minimalizuje výraz

$$\sum_{i=1}^n [y_i - s_k(x_i)]^2 + \rho \int_{a_1}^{a_2} [s_k^{(l)}(x)]^2 dx,$$

přičemž první člen odpovídá minimalizovanému výrazu v metodě nejmenších čtverců a druhý člen slouží pro dosažení požadované hladkosti. Podrobnosti lze najít v [14]. My se dále zaměříme na tzv. P -splajny, které pro penalizaci využívají difference, přičemž budeme vycházet z [6].

2.1. P -splajny

V této části se zaměříme na P -splajny. Naším úkolem je najít splajn $s_k(\mathbf{x}) \in S_k^{\Delta\lambda}[a_1, a_2]$, který minimalizuje výraz

$$\delta(\mathbf{b}) = \|\mathbf{y} - \mathbf{C}_{k+1}(\mathbf{x})\mathbf{b}\|^2 + \rho \mathbf{b}^T \mathbf{P}\mathbf{b}, \quad (2.1)$$

kde první člen odpovídá minimalizovanému výrazu v metodě nejmenších čtverců (1.5) a druhý člen slouží k dosažení požadované hladkosti výsledného splajnu. Parametr $\rho > 0$ je tzv. vyhlazující parametr, který ovlivňuje sílu penalizace na úkor aproximace dat ve smyslu metody nejmenších čtverců. S rostoucím parametrem ρ klademe větší důraz na penalizaci, a výsledná křivka je tedy hladší.

Hladkost křivky je dále ovlivněna tzv. penalizační maticí \mathbf{P} , která je v případě P -splajnu tvořena pomocí difference vektoru \mathbf{b} . Diferencí prvního řádu vektoru $\mathbf{b} = (b_{-k}, \dots, b_g)^T$ rozumíme vektor $\Delta\mathbf{b} = (\Delta b_{-k+1}, \dots, \Delta b_g)^T$, kde $\Delta b_j = b_j - b_{j-1}$, $j = -k + 1, \dots, g$. Diferenci vyšších řádů získáme pomocí opakovaného aplikování operátoru Δ , tedy například difference druhého řádu vektoru \mathbf{b} je vektor $\Delta^2\mathbf{b} = (\Delta^2 b_{-k+2}, \dots, \Delta^2 b_g)^T$, kde $\Delta^2 b_j = \Delta(b_j - b_{j-1}) = \Delta b_j - \Delta b_{j-1} = b_j - 2b_{j-1} + b_{j-2}$, $j = -k + 2, \dots, g$.

Uvažujme nyní obecný řád difference $d \in \{1, 2, \dots, g + k\}$. Diference d -tého řádu vektoru \mathbf{b} je pak vektor $\Delta^d\mathbf{b}$, který je sloupový a délky $g + k - d + 1$. Vektor $\Delta^d\mathbf{b}$ lze maticově zapsat jako součin $\mathbf{D}_d\mathbf{b}$, kde $\mathbf{D}_d \in \mathcal{R}^{g+k-d+1, g+k+1}$. Pro představu nyní uveďme, jak tyto matice vypadají pro $d = 1, 2$

$$\mathbf{D}_1 = \begin{pmatrix} -1 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & -1 & 1 \end{pmatrix}, \quad (2.2)$$

$$\mathbf{D}_2 = \begin{pmatrix} 1 & -2 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 1 & -2 & 1 \end{pmatrix}. \quad (2.3)$$

Obecně lze nenulové prvky matice \mathbf{D}_d vyjádřit pomocí kombinačních čísel.

Pro prvek $\mathbf{D}_d(i, j)$ matice \mathbf{D}_d v i -tém řádku a j -tém sloupci platí

$$\mathbf{D}_d(i, j) = \begin{cases} (-1)^{d+i-j} \binom{d}{j-i}, & \text{jestliže } i \leq j \leq i+d \\ 0, & \text{jinak,} \end{cases} \quad (2.4)$$

kde $i = 1, \dots, g+k-d+1$ a $j = 1, \dots, g+k+1$.

Matice \mathbf{D}_d je pak využita v minimalizovaném výrazu (2.1) tak, že matice \mathbf{P} má pro dané d podobu součinu $\mathbf{D}_d^T \mathbf{D}_d$, přičemž $d \in \{1, \dots, g+k\}$.

Minimalizovaný výraz pak lze přepsat do tvaru

$$\delta(\mathbf{b}) = (\mathbf{y} - \mathbf{C}_{k+1}(\mathbf{x})\mathbf{b})^T (\mathbf{y} - \mathbf{C}_{k+1}(\mathbf{x})\mathbf{b}) + \rho \mathbf{b}^T \mathbf{D}_d^T \mathbf{D}_d \mathbf{b}. \quad (2.5)$$

Jeho úpravou lze dojít k následujícímu tvaru, ze kterého je zřejmé, že se jedná o kvadratickou funkci proměnné \mathbf{b} ,

$$\delta(\mathbf{b}) = \mathbf{y}^T \mathbf{y} - 2\mathbf{b}^T \mathbf{C}_{k+1}^T(\mathbf{x}) + \mathbf{b}^T (\mathbf{C}_{k+1}^T(\mathbf{x})\mathbf{C}_{k+1}(\mathbf{x}) + \rho \mathbf{D}_d^T \mathbf{D}_d) \mathbf{b}. \quad (2.6)$$

Nyní postupujeme analogicky jako v odstavci 1.3, přičemž naším cílem je minimalizovat výraz (2.6). Opět nás zajímá, kdy je funkce $\delta(\mathbf{b})$ ryze konvexní, a tedy kdy existuje právě jedno její minimum.

Již v odstavci 1.3 jsme uvedli, kdy je matice $\mathbf{C}_{k+1}(\mathbf{x})$ plně sloupcové hodnosti, a tedy víme, kdy je matice $\mathbf{C}_{k+1}^T(\mathbf{x})\mathbf{C}_{k+1}(\mathbf{x})$ pozitivně definitní. V tom případě je pozitivně definitní i matice $\mathbf{C}_{k+1}^T(\mathbf{x})\mathbf{C}_{k+1}(\mathbf{x}) + \rho \mathbf{D}_d^T \mathbf{D}_d$. Platí totiž, že matice \mathbf{D}_d je plně řádkové hodnosti, protože její řádky jsou lineárně nezávislé, přičemž víme, že počet řádků této matice je menší než počet sloupců. Hodnost matice \mathbf{D}_d je tedy menší než počet sloupců matice, a proto je součin $\mathbf{D}_d^T \mathbf{D}_d$ pozitivně semidefinitní [9], str. 213. Součet pozitivně definitní a pozi-

tivně semidefinitní matice je matice pozitivně definitní [9], str. 212, a proto je funkce $\delta(\mathbf{b})$ ryze konvexní a existuje právě jedno její minimum.

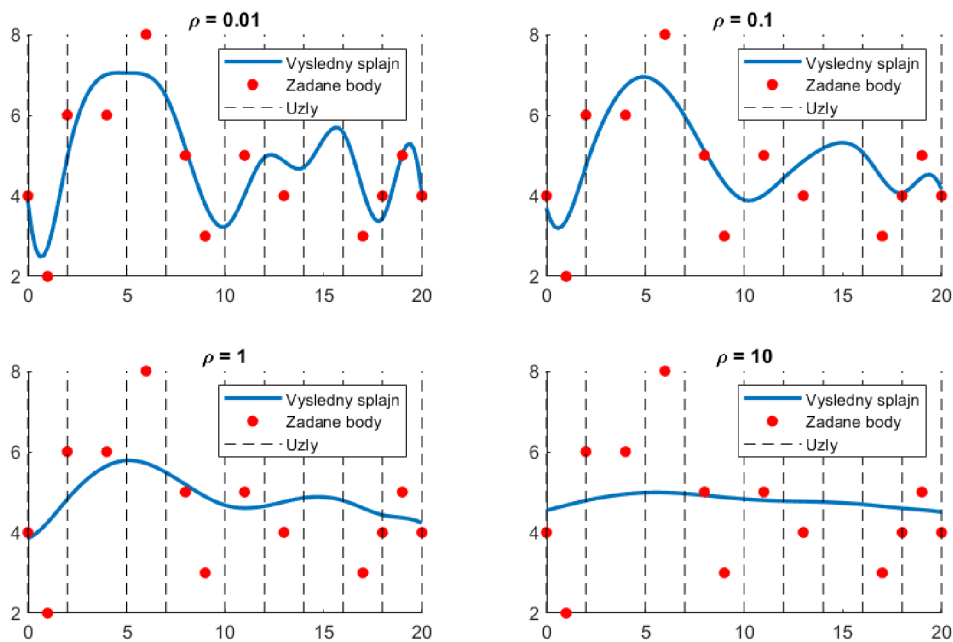
Abychom toto minimum našli, výraz (2.6) zderivujeme podle proměnné \mathbf{b}^T , výslednou derivaci položíme rovnu nulovému vektoru a z této rovnice vyjádříme proměnnou \mathbf{b} . Tím dojdeme ke vztahu

$$\hat{\mathbf{b}} = (\mathbf{C}_{k+1}^T(\mathbf{x})\mathbf{C}_{k+1}(\mathbf{x}) + \rho\mathbf{D}_d^T\mathbf{D}_d)^{-1} \mathbf{C}_{k+1}^T(\mathbf{x})\mathbf{y}, \quad (2.7)$$

přičemž tedy platí, že existuje právě jeden splajn daný vektorem $\hat{\mathbf{b}}$ ve tvaru (2.7), který daná data aproximuje ve smyslu penalizované metody nejmenších čtverců.

Příklad 2.1. Nyní metodu penalizovaných nejmenších čtverců ilustrujeme na příkladě, přičemž použijeme data, stupeň splajnu a první uvažovanou síť uzlů z Příkladu 1.1, tedy $k = 3$, $n = 15$ a $g = 8$. Úlohu budeme opět řešit v matematickém softwaru MATLAB.

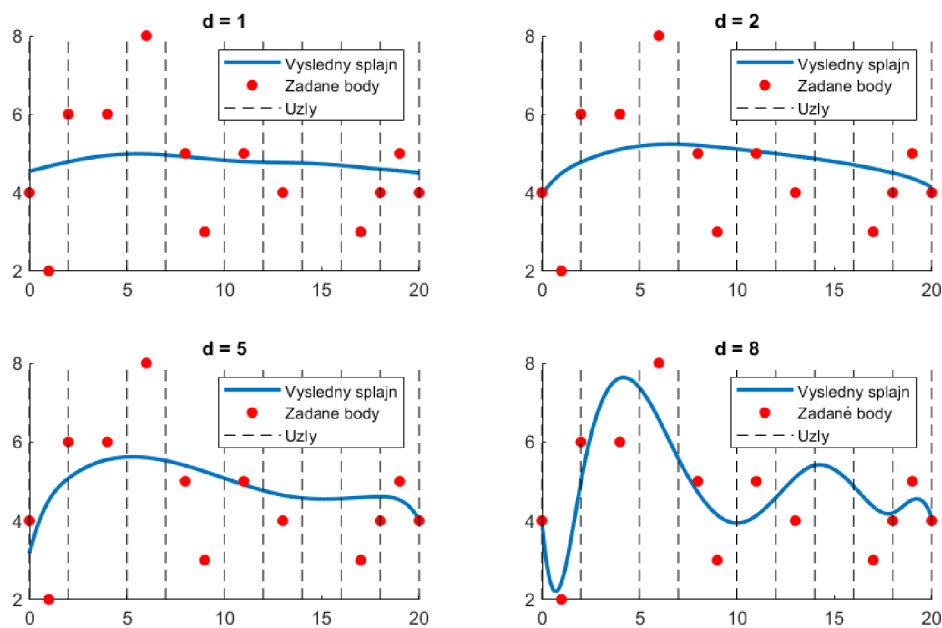
Nejprve budeme uvažovat difference prvního řádu, tj. zvolíme $d = 1$. Matici \mathbf{D}_1 vytvoříme pomocí funkce `diff`, kterou aplikujeme na jednotkovou matici \mathbf{I}_{12} (obecně \mathbf{I}_{g+k+1}). Za parametr ρ postupně volíme hodnoty 0.01, 0.1, 1 a 10. Stejně jako v Příkladu 1.1 platí, že je splněna Schoenberg-Whitneyho podmínka a kolokační matice $\mathbf{C}_4(\mathbf{x})$, kterou jsme vytvořili pomocí funkce `spcol`, je tedy plně sloupcové hodnosti. Poté vyřešíme soustavu $(\mathbf{C}_4^T(\mathbf{x})\mathbf{C}_4(\mathbf{x}) + \rho\mathbf{D}_1^T\mathbf{D}_1) \hat{\mathbf{b}} = \mathbf{C}_4^T(\mathbf{x})\mathbf{y}$ využitím funkce `linsolve`, čímž určíme vektor B -splajnových koeficientů $\hat{\mathbf{b}}$. Výsledné splajny získané za pomocí funkce `spmak` jsou zobrazeny v grafech na obrázku 3. Lze pozorovat, že čím větší je vyhlazující parametr ρ , tím hladší je výsledný splajn. Naopak s klesajícím ρ se splajn blíží splajnu získanému nepenalizovanou metodou nejmenších čtverců, viz obrázek 1.



Obrázek 3: Penalizovaná metoda nejmenších čtverců s různou volbou ρ

Nyní budeme měnit řád diference d , konkrétně uvažujeme hodnoty 1, 2, 5 a 8. Aby byly výsledné grafy jednoduše porovnatelné, zvolíme $\rho = 10$. Pro výrazně menší hodnoty ρ nelze pouhým okem porovnávat výsledné splajny. Matice \mathbf{D}_d určujeme opět pomocí funkce `diff`, přičemž ji na jednotkovou matici aplikujeme d -krát. Například pro $d = 2$ bychom matici \mathbf{D}_2 získali jako `diff(diff(eye(g+k+1)))` nebo lépe pomocí `diff(eye(g+k+1), 2)` (obecně `diff(eye(g+k+1), d)`).

Výsledkem jsou splajny zobrazené na obrázku 4. Pro dostatečně velkou hodnotu vyhlazujícího parametru ρ lze pozorovat, že čím menší je řád diference d v penalizačním členu, tím hladší je výsledný splajn. Tento jev je v souladu s tím, že dle [6] platí, že pro velké ρ se výsledný splajn získaný jako aproximace ve smyslu penalizované metody nejmenších čtverců blíží polynomu stupně $d-1$. Jak již bylo naznačeno, běžně řád diference d neovlivňuje



Obrázek 4: Penalizovaná metoda nejmenších čtverců s různou volbou řádu diference d

výsledný splajn nějak výrazně. Obecně je doporučeno volit $d = 1$ nebo $d = 2$, případně nejvýše $d = 3$.

3. Predikování v penalizované regresii

V této kapitole se dostáváme k hlavnímu tématu práce, což je schopnost predikce pro nová pozorování, přičemž uvažujeme predikování nejen pro pozorování tzv. mimo rozsah, ale i pro chybějící hodnoty z historie, a uvažujeme penalizaci založenou na diferencích, viz kapitola 2.1. Pokud není uvedeno jinak, vycházíme z [1].

3.1. Přístup chybějící hodnoty

V této části se budeme zabývat tzv. přístupem chybějící hodnoty, anglicky *the missing value approach*. Předpokládejme, že jsou dány hodnoty vysvětlující proměnné x_j z intervalu $[a_1, a_2]$ a hodnoty vysvětlované proměnné y_j pro $j = 1, \dots, n$, a označme sloupcové vektory délky n obsahující tyto hodnoty jako \mathbf{x} a \mathbf{y} . Zvolme stupeň splajnu $k \in \mathbb{N}_0$ a ve shodě s článkem [1] uvažujeme síť uzlů

$$\Delta\lambda : a_1 = \lambda_0 < \lambda_1 < \dots < \lambda_{c-k-1} < \lambda_{c-k} = a_2, \quad (3.1)$$

na které je možné zkonstruovat $c - 2k$ lineárně nezávislých splajnů. Protože platí $\dim(S_k^{\Delta\lambda}[a_1, a_2]) = c$, je třeba přidat $2k$ pomocných uzlů, které nyní volíme tak, že k pomocných uzlů přidaných vlevo od sítě uzlů (3.1) může být násobných, ale k pomocných uzlů přidaných napravo od sítě uzlů (3.1) volíme jako nenásobné, abychom je poté mohli využít v síti uzlů uvažované při predikci neznámých hodnot ležících vpravo od intervalu $[a_1, a_2]$, a aby byla splněna Schoenberg-Whitneyho podmínka. Jak konkrétně zvolit těchto k pomocných uzlů popíšeme v následujících odstavcích. Výsledkem je rozšířená

sít' uzlů

$$\lambda_{-k} \leq \dots \leq \lambda_{-1} \leq \lambda_0 < \lambda_1 < \dots < \lambda_{c-k-1} < \lambda_{c-k} < \lambda_{c-k+1} < \dots < \lambda_c,$$

která obsahuje $c+k+1$ uzlů a lze na ní sestavit celou bázi B -splajnů prostoru $S_k^{\Delta\lambda}[a_1, a_2]$. Naším úkolem je pak najít splajn $s_k(x) \in S_k^{\Delta\lambda}[a_1, a_2]$, který aproximuje data (x_j, y_j) , $j = 1, \dots, n$, ve smyslu penalizované metody nejmenších čtverců. Splajn $s_k(x)$ budeme hledat opět jako lineární kombinaci B -splajnů, tj.

$$s_k(x) = \sum_{i=-k}^{c-k-1} b_i B_i^{k+1}(x).$$

Maticově lze tedy pro celý vektor \mathbf{x} psát $s_k(\mathbf{x}) = \mathbf{C}_{k+1}(\mathbf{x})\mathbf{b}$, kde $\mathbf{b} = (b_{-k}, \dots, b_{c-k-1})^T$ je vektor B -splajnových koeficientů a $\mathbf{C}_{k+1}(\mathbf{x}) \in \mathcal{R}^{n,c}$ je kolokační matice, která je ve tvaru

$$\mathbf{C}_{k+1}(\mathbf{x}) = \begin{pmatrix} B_{-k}^{k+1}(x_1) & \dots & B_{c-k-1}^{k+1}(x_1) \\ \vdots & \ddots & \vdots \\ B_{-k}^{k+1}(x_n) & \dots & B_{c-k-1}^{k+1}(x_n) \end{pmatrix}.$$

Na popsanou aproximaci lze nahlížet i tak, že uvažujeme model

$$\mathbf{y} = \mathbf{C}_{k+1}(\mathbf{x})\mathbf{b} + \boldsymbol{\epsilon}, \quad (3.2)$$

kde $\mathbf{C}_{k+1}(\mathbf{x})$ je výše zmíněná kolokační matice a \mathbf{b} je vektor B -splajnových koeficientů a $\boldsymbol{\epsilon}$ je náhodná chyba, která se řídí normálním rozdělením s nulovou střední hodnotou a varianční maticí $\sigma_\epsilon^2 \mathbf{I}_n$, tj. $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma_\epsilon^2 \mathbf{I}_n)$.

Definujme nyní sloupcový vektor \mathbf{y}_p délky n_p , který je tvořen hodnotami vysvětlované proměnné, které chceme predikovat pomocí hodnot vysvětlující proměnné uložených ve vektoru \mathbf{x}_p , který je opět sloupcový stejné

délky. Obecně platí, že jeho hodnoty se nachází buď vpravo (predikce vpřed) nebo vlevo (predikce vzad) od intervalu $[a_1, a_2]$ nebo uvnitř intervalu $[a_1, a_2]$. Jak již bylo naznačeno při vytváření rozšířené sítě uzlů, dále budeme v této teoretické části předpokládat situaci, kdy se hodnoty \mathbf{x}_p nachází vpravo od intervalu $[a_1, a_2]$, konkrétně v intervalu $[a_2, a_3]$. Při tzv. predikci vzad bychom uvažovali rozšířenou síť uzlů

$$\lambda_{-k} < \dots < \lambda_{-1} < \lambda_0 < \lambda_1 < \dots < \lambda_{c-k-1} < \lambda_{c-k} \leq \lambda_{c-k+1} \leq \dots \leq \lambda_c$$

a podobně bychom postupovali i v případě, kdy chceme predikovat chybějící hodnoty uvnitř intervalu $[a_1, a_2]$. Tyto situace později ilustrujeme na příkladech.

Definujme nyní vektor známých hodnot $\mathbf{x}_+ = (\mathbf{x}^T, \mathbf{x}_p^T)^T$ a vektor, který obsahuje pozorované hodnoty vysvětlované proměnné \mathbf{y} a neznámé hodnoty \mathbf{y}_p , které chceme predikovat, tj. $\mathbf{y}_+ = (\mathbf{y}^T, \mathbf{y}_p^T)^T$. Délku těchto vektorů označíme jako n_+ , tedy $n_+ = n + n_p$.

Naším úkolem je aproximovat data (x_j, y_j) pro $j = 1, \dots, n$ ve smyslu penalizované metody nejmenších čtverců, viz odstavec 2.1, a model (3.2) rozšířit na model

$$\mathbf{y}_+ = \mathbf{C}_{k+1}^+(\mathbf{x}_+) \mathbf{b}_+ + \boldsymbol{\epsilon}_+, \quad (3.3)$$

kde $\boldsymbol{\epsilon}_+ \sim \mathcal{N}(0, \sigma_{\boldsymbol{\epsilon}_+}^2 \mathbf{I}_{n_+})$, pomocí kterého budeme predikovat hodnoty vysvětlované proměnné \mathbf{y}_p . Matice $\mathbf{C}_{k+1}^+(\mathbf{x}_+)$ ¹ je nová kolokační matice, která využívá novou síť uzlů, konkrétně původní uzly pokrývající rozsah hodnot x_i pro $i = 1, \dots, n$, tj. síť uzlů (3.1), a $c_p - k + 1$ nových uzlů pokrývajících

¹Horní index $+$ u matic v této kapitole značí, že se jedná o rozšířené matice.

rozsah x_{p_j} , kde $j = 1, \dots, n_p$,

$$\lambda_{c-k+1} < \lambda_{c-k+2} < \dots < \lambda_{c+c_p-2k} < \lambda_{c+c_p-2k+1},$$

kde prvních k uzlů, tj. $\lambda_{c-k+1} < \dots < \lambda_c$, je rovno právě pomocným uzlům přidaným napravo od původní sítě uzlů (3.1) a platí

$$\lambda_{c+c_p-2k} < a_3 \leq \lambda_{c+c_p-2k+1}.$$

Spojením s původní sítí uzlů (3.1) získáme novou síť uzlů

$$a_1 = \lambda_0 < \lambda_1 < \dots < \lambda_{c+c_p-2k} < \lambda_{c+c_p-2k+1} = a_3,$$

kterou je opět třeba rozšířit o $2k$ pomocných uzlů, které mohou být násobné.

Rozšířená síť uzlů s $c + c_p + 2$ uzly je pak dána jako

$$\lambda_{-k} \leq \dots \leq \lambda_0 < \lambda_1 < \dots < \lambda_{c+c_p-2k+1} \leq \lambda_{c+c_p-2k+2} \leq \dots \leq \lambda_{c+c_p-k+1}.$$

Rozšířená kolokační matice je tvaru

$$\mathbf{C}_{k+1}^+(\mathbf{x}_+) = \begin{bmatrix} \mathbf{C}_{k+1}(\mathbf{x}) & \mathbf{0}_{n, c_p} \\ \mathbf{C}_{k+1}^1(\mathbf{x}_p) & \mathbf{C}_{k+1}^2(\mathbf{x}_p) \end{bmatrix} \quad (3.4)$$

a typu $n_+ \times c_+$, kde $c_+ = c + c_p$, protože $\mathbf{C}_{k+1}^1(\mathbf{x}_p) \in \mathcal{R}^{n_p, c}$ a $\mathbf{C}_{k+1}^2(\mathbf{x}_p) \in \mathcal{R}^{n_p, c_p}$. Tyto matice rozšiřují původní kolokační matici $\mathbf{C}_{k+1}(\mathbf{x})$ na matici $\mathbf{C}_{k+1}^+(\mathbf{x}_+)$, která díky využití nových uzlů umožňuje predikci nových n_p hodnot vektoru \mathbf{y}_p . Matici \mathbf{C}_{k+1}^+ odpovídá nový vektor B -splajnových koeficientů $\mathbf{b}_+ = (\mathbf{b}^T, \mathbf{b}_p^T)^T$, který je délky c_+ . Protože uvažujeme aproximaci ve smyslu penalizované metody nejmenších čtverců, je třeba definovat i matici penali-

zace $\mathbf{P}_+ \in \mathcal{R}^{c_+, c_+}$ tvaru

$$\mathbf{P}_+ := \begin{bmatrix} \mathbf{P}_1 & \mathbf{P}_2 \\ \mathbf{P}_2^T & \mathbf{P}_3 \end{bmatrix}, \quad (3.5)$$

kde $\mathbf{P}_1 \in \mathcal{R}^{c, c}$, $\mathbf{P}_2 \in \mathcal{R}^{c, c_p}$ a $\mathbf{P}_3 \in \mathcal{R}^{c_p, c_p}$, kterou lze opět získat pomocí matice $\mathbf{D}_d^+ \in \mathcal{R}^{c+d, c+d}$ d -tých diferencí B -splajnových koeficientů \mathbf{b}_+ , konkrétně

$$\mathbf{P}_+ = (\mathbf{D}_d^+)^T \mathbf{D}_d^+,$$

kde

$$\mathbf{D}_d^+ = \begin{bmatrix} \mathbf{D}_d & \mathbf{0}_{c-d, c_p} \\ \mathbf{D}_d^1 & \mathbf{D}_d^2 \end{bmatrix}, \quad (3.6)$$

přičemž $\mathbf{D}_d \in \mathcal{R}^{c-d, c}$, $\mathbf{D}_d^1 \in \mathcal{R}^{c_p, c}$ a $\mathbf{D}_d^2 \in \mathcal{R}^{c_p, c_p}$. Matici \mathbf{P}_+ tedy lze psát ve tvaru

$$\mathbf{P}_+ = (\mathbf{D}_d^+)^T \mathbf{D}_d^+ = \begin{bmatrix} \mathbf{D}_d^T \mathbf{D}_d + (\mathbf{D}_d^1)^T \mathbf{D}_d^1 & (\mathbf{D}_d^1)^T \mathbf{D}_d^2 \\ (\mathbf{D}_d^2)^T \mathbf{D}_d^1 & (\mathbf{D}_d^2)^T \mathbf{D}_d^2 \end{bmatrix},$$

kde v porovnání se vztahem (3.5) dostáváme

$$\mathbf{P}_1 = \mathbf{D}_d^T \mathbf{D}_d + (\mathbf{D}_d^1)^T \mathbf{D}_d^1, \quad (3.7a)$$

$$\mathbf{P}_2 = (\mathbf{D}_d^1)^T \mathbf{D}_d^2, \quad (3.7b)$$

$$\mathbf{P}_3 = (\mathbf{D}_d^2)^T \mathbf{D}_d^2. \quad (3.7c)$$

Nyní se zaměříme na matici \mathbf{D}_d^2 , přičemž chceme ověřit, že se jedná o regulární matici, protože později budeme pracovat s její inverzí. Ze vztahu (2.4) pro prvky matice \mathbf{D}_d , který lze využít i pro prvky matice \mathbf{D}_d^+ , plyne,

že pro prvek $\mathbf{D}_d^2(i, j)$ matice \mathbf{D}_d^2 v i -tém řádku a j -tém sloupci platí

$$\mathbf{D}_d^2(i, j) = \begin{cases} (-1)^{i-j} \binom{d}{j-i+d}, & \text{jestliže } i - d \leq j \leq i \\ 0, & \text{jinak.} \end{cases}$$

Z toho je zřejmé, že matice \mathbf{D}_d^2 má na diagonále samé jedničky a je to dolní trojúhelníková matice, pro kterou dle [9], str. 180, platí, že její determinant je roven součinu všech prvků na diagonále. Proto platí $|\mathbf{D}_d^2| = 1$, a matice \mathbf{D}_d^2 je tedy regulární a existuje její inverze.

Uveďme nyní, jak konkrétně vypadají submatice matice \mathbf{D}_d^+ pro řády difference zvolené jako $d = 1$ a $d = 2$.

- Pro $d = 1$ je matice $\mathbf{D}_1^+ \in \mathcal{R}^{c+1, c+}$ ve tvaru

$$\mathbf{D}_1^+ = \left(\begin{array}{cccccccc|cccccccc} -1 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & -1 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & \cdots & 0 & 0 & -1 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & -1 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & -1 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & -1 & 1 & 0 \end{array} \right)$$

kde jsou čarami odděleny jednotlivé submatice dle vztahu (3.6), a tedy

platí, že \mathbf{D}_1 je dána vztahem (2.2),

$$\mathbf{D}_1^1 = \begin{pmatrix} 0 \cdots 0 & -1 \\ 0 \cdots 0 & 0 \\ \vdots & \ddots \\ 0 \cdots 0 & 0 \end{pmatrix}, \quad \mathbf{D}_1^2 = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & -1 & 1 \end{pmatrix}. \quad (3.8)$$

- Pro $d = 2$ je matice $\mathbf{D}_2^+ \in \mathcal{R}^{c_+ - 2, c_+}$ ve tvaru

$$\mathbf{D}_2^+ = \left(\begin{array}{cccccccc|cccccccc} 1 & -2 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 & -2 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & -2 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 1 & -2 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 1 & -2 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 1 & -2 & 1 \end{array} \right)$$

a tedy \mathbf{D}_2 je dána vztahem (2.3),

$$\mathbf{D}_2^1 = \begin{pmatrix} 0 \cdots 0 & 1 & -2 \\ 0 \cdots 0 & 0 & 1 \\ 0 \cdots 0 & 0 & 0 \\ \vdots & \ddots & \vdots \\ 0 \cdots 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{D}_2^2 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 1 & -2 & 1 \end{pmatrix}.$$

Abychom získali požadovaný odhad $\hat{\mathbf{b}}_+$ vektoru B -splajnových koeficientů \mathbf{b}_+ , který pak využijeme pro predikci vektoru \mathbf{y}_+ dle vztahu

$$\hat{\mathbf{y}}_+ = \mathbf{C}_{k+1}^+(\mathbf{x}_+) \hat{\mathbf{b}}_+, \quad (3.9)$$

je třeba minimalizovat výraz

$$\delta_+(\mathbf{b}_+) = (\mathbf{y}_+ - \mathbf{C}_{k+1}^+(\mathbf{x}_+)\mathbf{b}_+)^T \mathbf{M} (\mathbf{y}_+ - \mathbf{C}_{k+1}^+(\mathbf{x}_+)\mathbf{b}_+) + \rho \mathbf{b}_+^T \mathbf{P}_+ \mathbf{b}_+, \quad (3.10)$$

kde matice $\mathbf{M} \in \mathcal{R}^{n_+,n_+}$ je diagonální matice s prvky m_j , $j = 1, \dots, n_+$, na diagonále, pro které platí

$$m_j = \begin{cases} 1 & \text{pro } j = 1, \dots, n \\ 0 & \text{pro } j = n + 1, \dots, n_+. \end{cases}$$

Tedy diagonála matice \mathbf{M} je složena z 0 pro chybějící hodnoty, tj. pro \mathbf{y}_p , a 1 pro pozorované hodnoty, tj. pro \mathbf{y} . Matici \mathbf{M} lze tedy zapsat ve tvaru

$$\mathbf{M} = \begin{bmatrix} \mathbf{I}_n & \mathbf{0}_{n,n_p} \\ \mathbf{0}_{n_p,n} & \mathbf{0}_{n_p,n_p} \end{bmatrix}. \quad (3.11)$$

Za účelem minimalizace zderivujeme funkci (3.10) podle proměnné \mathbf{b}_+^T a položením výsledné derivace rovno nulovému vektoru získáme pro \mathbf{b}_+ následující odhad

$$\hat{\mathbf{b}}_+ = \left((\mathbf{C}_{k+1}^+(\mathbf{x}_+))^T \mathbf{M} \mathbf{C}_{k+1}^+(\mathbf{x}_+) + \rho \mathbf{P}_+ \right)^{-1} (\mathbf{C}_{k+1}^+(\mathbf{x}_+))^T \mathbf{M} \mathbf{y}_+. \quad (3.12)$$

Protože se v tomto vztahu vyskytuje inverzní matice, mohlo by dojít k problému s její existencí. Dosazením vztahů (3.4), (3.5) a (3.11) do matice, jejíž

inverze se ve vztahu (3.12) vyskytuje, dojdeme k tvaru

$$\begin{bmatrix} \mathbf{C}_{k+1}^T(\mathbf{x}) & (\mathbf{C}_{k+1}^1(\mathbf{x}_p))^T \\ \mathbf{0}_{c_p, n} & (\mathbf{C}_{k+1}^2(\mathbf{x}_p))^T \end{bmatrix} \begin{bmatrix} \mathbf{I}_n & \mathbf{0}_{n, n_p} \\ \mathbf{0}_{n_p, n} & \mathbf{0}_{n_p, n_p} \end{bmatrix} \begin{bmatrix} \mathbf{C}_{k+1}(\mathbf{x}) & \mathbf{0}_{n, c_p} \\ \mathbf{C}_{k+1}^1(\mathbf{x}_p) & \mathbf{C}_{k+1}^2(\mathbf{x}_p) \end{bmatrix} + \rho \begin{bmatrix} \mathbf{P}_1 & \mathbf{P}_2 \\ \mathbf{P}_2^T & \mathbf{P}_3 \end{bmatrix}.$$

Po vynásobení a následném sečtení matic dostáváme matici tvaru

$$\begin{bmatrix} \mathbf{C}_{k+1}^T(\mathbf{x})\mathbf{C}_{k+1}(\mathbf{x}) + \rho\mathbf{P}_1 & \rho\mathbf{P}_2 \\ \rho\mathbf{P}_2^T & \rho\mathbf{P}_3 \end{bmatrix}, \quad (3.13)$$

do které ještě dosadíme za \mathbf{P}_1 , \mathbf{P}_2 a \mathbf{P}_3 dle vztahu (3.7), tedy

$$\begin{bmatrix} \mathbf{C}_{k+1}^T(\mathbf{x})\mathbf{C}_{k+1}(\mathbf{x}) + \rho\mathbf{D}_d^T\mathbf{D}_d + \rho(\mathbf{D}_d^1)^T\mathbf{D}_d^1 & \rho(\mathbf{D}_d^1)^T\mathbf{D}_d^2 \\ \rho(\mathbf{D}_d^2)^T\mathbf{D}_d^1 & \rho(\mathbf{D}_d^2)^T\mathbf{D}_d^2 \end{bmatrix}. \quad (3.14)$$

O matici \mathbf{D}_d^2 již víme, že je regulární, a proto je dle [9], str. 214, regulární i matice $(\mathbf{D}_d^2)^T\mathbf{D}_d^2$, čímž jsou splněny předpoklady Věty 1.2. Dle této věty je,

za předpokladu regularity matice \mathbf{T} , matice $\begin{bmatrix} \mathbf{W} & \mathbf{V} \\ \mathbf{U} & \mathbf{T} \end{bmatrix}$ regulární právě tehdy,

když je regulární matice $\mathbf{Q} = \mathbf{W} - \mathbf{V}\mathbf{T}^{-1}\mathbf{U}$. V našem případě tedy položíme-li

$$\begin{aligned} \mathbf{W} &= \mathbf{C}_{k+1}^T(\mathbf{x})\mathbf{C}_{k+1}(\mathbf{x}) + \rho\mathbf{D}_d^T\mathbf{D}_d + \rho(\mathbf{D}_d^1)^T\mathbf{D}_d^1, \\ \mathbf{V} &= \rho(\mathbf{D}_d^1)^T\mathbf{D}_d^2, \\ \mathbf{U} &= \rho(\mathbf{D}_d^2)^T\mathbf{D}_d^1, \\ \mathbf{T} &= \rho(\mathbf{D}_d^2)^T\mathbf{D}_d^2. \end{aligned}$$

získáme

$$\begin{aligned} \mathbf{Q} &= \mathbf{C}_{k+1}^T(\mathbf{x})\mathbf{C}_{k+1}(\mathbf{x}) + \rho\mathbf{D}_d^T\mathbf{D}_d + \rho(\mathbf{D}_d^1)^T\mathbf{D}_d^1 - \\ &\quad - \rho(\mathbf{D}_d^1)^T\mathbf{D}_d^2 \left(\rho(\mathbf{D}_d^2)^T\mathbf{D}_d^2 \right)^{-1} \rho(\mathbf{D}_d^2)^T\mathbf{D}_d^1. \end{aligned} \quad (3.15)$$

Využitím vztahů pro inverzní matice platných dle [9], str. 82, lze vztah (3.15) upravit následovně

$$\begin{aligned} \mathbf{Q} = & \mathbf{C}_{k+1}^T(\mathbf{x})\mathbf{C}_{k+1}(\mathbf{x}) + \rho\mathbf{D}_d^T\mathbf{D}_d + \rho(\mathbf{D}_d^1)^T\mathbf{D}_d^1 - \\ & - \rho(\mathbf{D}_d^1)^T\mathbf{D}_d^2 \frac{1}{\rho}(\mathbf{D}_d^2)^{-1} \left((\mathbf{D}_d^2)^T \right)^{-1} \rho(\mathbf{D}_d^2)^T\mathbf{D}_d^1. \end{aligned}$$

Poté využitím definice inverzní matice, viz [9], str. 81, a následným odečtením totožných členů lze dojít ke vztahu

$$\mathbf{Q} = \mathbf{C}_{k+1}^T(\mathbf{x})\mathbf{C}_{k+1}(\mathbf{x}) + \rho\mathbf{D}_d^T\mathbf{D}_d, \quad (3.16)$$

ze kterého je zřejmé, že se za určitých podmínek může jednat o regulární matici. Již v odstavci 1.3 jsme se věnovali tomu, jak zvolit uzly tak, aby byla matice $\mathbf{C}_{k+1}^T(\mathbf{x})\mathbf{C}_{k+1}(\mathbf{x})$ pozitivně definitní, a také již víme, že matice $\rho\mathbf{D}_d^T\mathbf{D}_d$ je pozitivně semidefinitní, viz odstavec 2.1. Součet pozitivně definitní a pozitivně semidefinitní matice je matice pozitivně definitní, a tedy je i regulární.

Existenci inverzní matice obsažené ve vztahu (3.12) máme tedy za daných podmínek zaručenu a využitím vztahu (3.9) lze vztah pro odhad $\hat{\mathbf{b}}_+$ vektoru B -splajnových koeficientů \mathbf{b}_+ (3.12) využít pro predikci $\hat{\mathbf{y}}_+$ hodnot vysvětlované proměnné \mathbf{y}_+ , konkrétně

$$\hat{\mathbf{y}}_+ = \mathbf{C}_{k+1}^+(\mathbf{x}_+) \left((\mathbf{C}_{k+1}^+(\mathbf{x}_+))^T \mathbf{M} \mathbf{C}_{k+1}^+(\mathbf{x}_+) + \rho\mathbf{P}_+ \right)^{-1} (\mathbf{C}_{k+1}^+(\mathbf{x}_+))^T \mathbf{M} \mathbf{y}_+,$$

přičemž pomocí upravené matice (3.13) lze tento vztah přepsat do tvaru

$$\hat{\mathbf{y}}_+ = \mathbf{C}_{k+1}^+(\mathbf{x}_+) \left[\begin{array}{cc} \mathbf{C}_{k+1}^T(\mathbf{x})\mathbf{C}_{k+1}(\mathbf{x}) + \rho\mathbf{P}_1 & \rho\mathbf{P}_2 \\ \rho\mathbf{P}_2^T & \rho\mathbf{P}_3 \end{array} \right]^{-1} (\mathbf{C}_{k+1}^+(\mathbf{x}_+))^T \mathbf{M} \mathbf{y}_+. \quad (3.17)$$

Tento zápis lze ještě zjednodušit využitím tzv. hat matice. Označíme-li

$$\mathbf{H}_+ := \mathbf{C}_{k+1}^+(\mathbf{x}_+) \left((\mathbf{C}_{k+1}^+(\mathbf{x}_+))^T \mathbf{M} \mathbf{C}_{k+1}^+(\mathbf{x}_+) + \rho \mathbf{P}_+ \right)^{-1} (\mathbf{C}_{k+1}^+(\mathbf{x}_+))^T \mathbf{M},$$

pak lze psát

$$\hat{\mathbf{y}}_+ = \mathbf{H}_+ \mathbf{y}_+.$$

Nyní se zaměříme vztah (3.17), konkrétně na součin $(\mathbf{C}_{k+1}^+(\mathbf{x}_+))^T \mathbf{M} \mathbf{y}_+$. Dosadíme do něj vztahy (3.4), (3.11) a vztah pro \mathbf{y}_+ , tj. $\mathbf{y}_+ = (\mathbf{y}^T, \mathbf{y}_p^T)^T$, a vynásobíme. Výslednou matici dosadíme do (3.17), čímž dojdeme ke vztahu

$$\hat{\mathbf{y}}_+ = \mathbf{C}_{k+1}^+(\mathbf{x}_+) \begin{bmatrix} \mathbf{C}_{k+1}^T(\mathbf{x}) \mathbf{C}_{k+1}(\mathbf{x}) + \rho \mathbf{P}_1 & \rho \mathbf{P}_2 \\ \rho \mathbf{P}_2^T & \rho \mathbf{P}_3 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{C}_{k+1}^T(\mathbf{x}) \mathbf{y} \\ \mathbf{0} \end{bmatrix}. \quad (3.18)$$

Protože uvažujeme situaci, kdy je inverzní matice obsažená ve vztahu (3.18) regulární, lze využít Lemma 1.1 i Větu 1.1, pokud ověříme platnost všech jejích předpokladů. Budeme chtít využít toho, že matice

$$\begin{bmatrix} \mathbf{0}_{q,n} & \mathbf{0}_{q,m} \\ \mathbf{0}_{p,n} & \mathbf{T}^- \end{bmatrix} + \begin{bmatrix} \mathbf{I}_q \\ -\mathbf{T}^- \mathbf{U} \end{bmatrix} \mathbf{Q}^- [\mathbf{I}_n, -\mathbf{V} \mathbf{T}^-],$$

kde $\mathbf{Q} = \mathbf{W} - \mathbf{V} \mathbf{T}^- \mathbf{U}$, je inverzní maticí k matici $\begin{bmatrix} \mathbf{W} & \mathbf{V} \\ \mathbf{U} & \mathbf{T} \end{bmatrix}$, viz Věta 1.1, s tím, že hledáme právě inverzní matici obsaženou ve vztahu (3.18), tj. volíme

$$\mathbf{W} = \mathbf{C}_{k+1}^T(\mathbf{x}) \mathbf{C}_{k+1}(\mathbf{x}) + \rho \mathbf{P}_1$$

$$\mathbf{V} = \rho \mathbf{P}_2$$

$$\mathbf{U} = \rho \mathbf{P}_2^T$$

$$\mathbf{T} = \rho \mathbf{P}_3.$$

Nyní využijeme Lemma 1.2 a ověříme, že je splněno $\mathcal{C}(\mathbf{U}) \subset \mathcal{C}(\mathbf{T})$ a $\mathcal{R}(\mathbf{V}) \subset \mathcal{R}(\mathbf{T})$, v našem případě tedy $\mathcal{C}(\mathbf{P}_2^T) \subset \mathcal{C}(\mathbf{P}_3)$ a $\mathcal{R}(\mathbf{P}_2) \subset \mathcal{R}(\mathbf{P}_3)$. Tyto vztahy lze dle (3.7) upravit do tvaru

$$\begin{aligned} \mathcal{C}\left((\mathbf{D}_d^2)^T \mathbf{D}_d^1\right) &\subset \mathcal{C}\left((\mathbf{D}_d^2)^T \mathbf{D}_d^2\right), \\ \text{a } \mathcal{R}\left((\mathbf{D}_d^1)^T \mathbf{D}_d^2\right) &\subset \mathcal{R}\left((\mathbf{D}_d^2)^T \mathbf{D}_d^2\right). \end{aligned} \quad (3.19)$$

O matici \mathbf{D}_d^2 již víme, že je regulární, a tedy existuje její inverze $(\mathbf{D}_d^2)^{-1}$ a proto dle [9], str. 82, existuje i $\left((\mathbf{D}_d^2)^T\right)^{-1}$. Protože pak existují matice $(\mathbf{D}_d^2)^{-1} \mathbf{D}_d^1 \in \mathcal{R}^{c_p, c}$ a $(\mathbf{D}_d^1)^T \left((\mathbf{D}_d^2)^T\right)^{-1} \in \mathcal{R}^{c, c_p}$ takové, že

$$\begin{aligned} (\mathbf{D}_d^2)^T \mathbf{D}_d^1 &= (\mathbf{D}_d^2)^T \mathbf{D}_d^2 (\mathbf{D}_d^2)^{-1} \mathbf{D}_d^1 \\ \text{a } (\mathbf{D}_d^1)^T \mathbf{D}_d^2 &= (\mathbf{D}_d^1)^T \left((\mathbf{D}_d^2)^T\right)^{-1} (\mathbf{D}_d^2)^T \mathbf{D}_d^2, \end{aligned}$$

je dle Lemmatu 1.2 platnost vztahů (3.19) ověřena.

Předpoklady Věty 1.1 jsou tedy splněny a větu můžeme použít k nalezení inverzní matice ze vztahu (3.18). O matici $\mathbf{T} = \rho \mathbf{P}_3 = \rho (\mathbf{D}_d^2)^T \mathbf{D}_d^2$ již víme, že je regulární, a proto dle Lemmatu 1.1 platí, že $\mathbf{T}^- = \mathbf{T}^{-1}$. Hledaná inverzní matice ze vztahu (3.18) je pak ve tvaru

$$\begin{bmatrix} \mathbf{0}_{c, c} & \mathbf{0}_{c, c_p} \\ \mathbf{0}_{c_p, c} & (\rho \mathbf{P}_3)^{-1} \end{bmatrix} + \begin{bmatrix} \mathbf{I}_c \\ -(\rho \mathbf{P}_3)^{-1} \rho \mathbf{P}_2^T \end{bmatrix} \mathbf{Q}^- [\mathbf{I}_c, -\rho \mathbf{P}_2 (\rho \mathbf{P}_3)^{-1}].$$

Nyní opět využijeme vztahy pro inverzní matice platné dle [9], str. 82, a tím dojdeme k tvaru

$$\begin{bmatrix} \mathbf{0}_{c, c} & \mathbf{0}_{c, c_p} \\ \mathbf{0}_{c_p, c} & \frac{1}{\rho} \mathbf{P}_3^{-1} \end{bmatrix} + \begin{bmatrix} \mathbf{I}_c \\ -\mathbf{P}_3^{-1} \mathbf{P}_2^T \end{bmatrix} \mathbf{Q}^- [\mathbf{I}_c, -\mathbf{P}_2 \mathbf{P}_3^{-1}],$$

kde matice \mathbf{Q} je dána vztahem (3.16).

Celkem tedy platí

$$\begin{aligned} & \begin{bmatrix} \mathbf{C}_{k+1}^T(\mathbf{x})\mathbf{C}_{k+1}(\mathbf{x}) + \rho\mathbf{P}_1 & \rho\mathbf{P}_2 \\ \rho\mathbf{P}_2^T & \rho\mathbf{P}_3 \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{0}_{c,c} & \mathbf{0}_{c,c_p} \\ \mathbf{0}_{c_p,c} & \frac{1}{\rho}\mathbf{P}_3^{-1} \end{bmatrix} + \\ & + \begin{bmatrix} \mathbf{I}_c \\ -\mathbf{P}_3^{-1}\mathbf{P}_2^T \end{bmatrix} (\mathbf{C}_{k+1}^T(\mathbf{x})\mathbf{C}_{k+1}(\mathbf{x}) + \rho\mathbf{D}_d^T\mathbf{D}_d)^{-1} [\mathbf{I}_c, -\mathbf{P}_2\mathbf{P}_3^{-1}] \end{aligned}$$

a dosazením do vztahu (3.18) a roznásobením dostaneme

$$\begin{aligned} \hat{\mathbf{y}}_+ &= \mathbf{C}_{k+1}^+(\mathbf{x}_+) \begin{bmatrix} \mathbf{0}_{c,c} & \mathbf{0}_{c,c_p} \\ \mathbf{0}_{c_p,c} & \frac{1}{\rho}\mathbf{P}_3^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{C}_{k+1}^T(\mathbf{x})\mathbf{y} \\ \mathbf{0} \end{bmatrix} + \\ & + \mathbf{C}_{k+1}^+(\mathbf{x}_+) \begin{bmatrix} \mathbf{I}_c \\ -\mathbf{P}_3^{-1}\mathbf{P}_2^T \end{bmatrix} (\mathbf{C}_{k+1}^T(\mathbf{x})\mathbf{C}_{k+1}(\mathbf{x}) + \rho\mathbf{D}_d^T\mathbf{D}_d)^{-1} \quad (3.20) \\ & [\mathbf{I}_c, -\mathbf{P}_2\mathbf{P}_3^{-1}] \begin{bmatrix} \mathbf{C}_{k+1}^T(\mathbf{x})\mathbf{y} \\ \mathbf{0} \end{bmatrix}. \end{aligned}$$

Nyní dosadíme za $\mathbf{C}_{k+1}^+(\mathbf{x}_+)$ dle (3.4) a pro první člen na pravé straně (3.20) dostaneme

$$\begin{bmatrix} \mathbf{C}_{k+1}(\mathbf{x}) & \mathbf{0}_{n,c_p} \\ \mathbf{C}_{k+1}^1(\mathbf{x}_p) & \mathbf{C}_{k+1}^2(\mathbf{x}_p) \end{bmatrix} \begin{bmatrix} \mathbf{0}_{c,c} & \mathbf{0}_{c,c_p} \\ \mathbf{0}_{c_p,c} & \frac{1}{\rho}\mathbf{P}_3^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{C}_{k+1}^T(\mathbf{x})\mathbf{y} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}.$$

Ve druhém členu na pravé straně (3.20) využijeme toho, že

$$[\mathbf{I}_c, -\mathbf{P}_2\mathbf{P}_3^{-1}] \begin{bmatrix} \mathbf{C}_{k+1}^T(\mathbf{x})\mathbf{y} \\ \mathbf{0} \end{bmatrix} = \mathbf{C}_{k+1}^T(\mathbf{x})\mathbf{y}.$$

Tím dojdeme k výslednému vztahu

$$\hat{\mathbf{y}}_+ = \mathbf{C}_{k+1}^+(\mathbf{x}_+) \begin{bmatrix} \mathbf{I}_c \\ -\mathbf{P}_3^{-1}\mathbf{P}_2^T \end{bmatrix} (\mathbf{C}_{k+1}^T(\mathbf{x})\mathbf{C}_{k+1}(\mathbf{x}) + \rho\mathbf{D}_d^T\mathbf{D}_d)^{-1} \mathbf{C}_{k+1}^T(\mathbf{x})\mathbf{y}. \quad (3.21)$$

Využitím (3.9) zjistíme, že z (3.21) plyne, že prvních c prvků vektoru $\hat{\mathbf{b}}_+$ (tj. těch odpovídajícím \mathbf{y}) je rovno

$$\hat{\mathbf{b}} = (\mathbf{C}_{k+1}^T(\mathbf{x})\mathbf{C}_{k+1}(\mathbf{x}) + \rho\mathbf{D}_d^T\mathbf{D}_d)^{-1} \mathbf{C}_{k+1}^T(\mathbf{x})\mathbf{y}. \quad (3.22)$$

Obdrželi jsme tedy stejný vztah jako je (2.7) z kapitoly o penalizované metodě nejmenších čtverců, což znamená, že odhady prvních c B -splajnových koeficientů z modelu (3.3) jsou stejné jako odhady B -splajnových koeficientů v základním modelu (3.2).

Odhady zbylých c_p B -splajnových koeficientů (tj. těch odpovídajících \mathbf{y}_p) jsou dle vztahu (3.21) rovny

$$\hat{\mathbf{b}}_p = -\mathbf{P}_3^{-1}\mathbf{P}_2^T (\mathbf{C}_{k+1}^T(\mathbf{x})\mathbf{C}_{k+1}(\mathbf{x}) + \rho\mathbf{P})^{-1} \mathbf{C}_{k+1}^T(\mathbf{x})\mathbf{y},$$

a tedy užitím vztahu (3.22) lze psát

$$\hat{\mathbf{b}}_p = -\mathbf{P}_3^{-1}\mathbf{P}_2^T\hat{\mathbf{b}}.$$

Využitím vztahu (3.7) a již výše použitých úprav dle [9], str. 82, lze tento vztah přepsat do tvaru

$$\hat{\mathbf{b}}_p = -(\mathbf{D}_d^2)^{-1} \mathbf{D}_d^1\hat{\mathbf{b}}. \quad (3.23)$$

Důsledkem vztahu (3.23) je fakt, že odhady B -splajnových koeficientů odpovídajících neznámým hodnotám \mathbf{y}_p , tj. vektor $\hat{\mathbf{b}}_p$, závisí vždy pouze

na posledních d vypočtených koeficientech, kde d je řád difference použitý v penalizačním členu. To si nyní ukážeme pro $d = 1$ a poté pro $d = 2$. Dosaďme do vztahu (3.23) za matice \mathbf{D}_1^1 a \mathbf{D}_1^2 dle (3.8), tj.

$$\hat{\mathbf{b}}_p = - \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & -1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 0 & \cdots & 0 & -1 \\ 0 & \cdots & 0 & 0 \\ 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 \end{pmatrix} \hat{\mathbf{b}}. \quad (3.24)$$

Rekurzivním algoritmem popsaným v [9] na straně 97 najdeme inverzní matici k dolní trojúhelníkové matici \mathbf{D}_1^2 a zjistíme, že platí

$$(\mathbf{D}_1^2)^{-1} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & -1 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & 1 & \cdots & 1 & 1 \end{pmatrix}.$$

Obdržený tvar matice dosadíme do vztahu (3.24) a matice vynásobíme. Označíme-li i -tý prvek vektoru $\hat{\mathbf{b}}$ jako \hat{b}_i , kde $i = 1, \dots, c$, získáme

$$\hat{\mathbf{b}}_p = - \begin{pmatrix} 0 & \cdots & 0 & -1 \\ 0 & \cdots & 0 & -1 \\ 0 & \cdots & 0 & -1 \\ \vdots & \ddots & \vdots & \vdots \end{pmatrix} \hat{\mathbf{b}} = \hat{b}_c \begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \end{pmatrix}. \quad (3.25)$$

Odhad $\hat{\mathbf{b}}_p$ tedy opravdu závisí pouze na posledním odhadnutém koeficientu \hat{b}_c . Analogicky bychom postupovali pro difference druhého řádu. Protože

platí

$$\hat{\mathbf{b}}_p = - \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 1 & -2 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 0 & \cdots & 0 & 1 & -2 \\ 0 & \cdots & 0 & 0 & 1 \\ 0 & \cdots & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & 0 \end{pmatrix} \hat{\mathbf{b}},$$

přičemž

$$(\mathbf{D}_2^2)^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 1 & -2 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 2 & 1 & 0 & 0 & \cdots & 0 \\ 3 & 2 & 1 & 0 & \cdots & 0 \\ 4 & 3 & 2 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ c_p & c_p-1 & c_p-2 & c_p-3 & \cdots & 1 \end{pmatrix},$$

dospěli bychom ke vztahu

$$\hat{\mathbf{b}}_p = - \begin{pmatrix} 0 & \cdots & 0 & 1 & -2 \\ 0 & \cdots & 0 & 2 & -3 \\ 0 & \cdots & 0 & 3 & -4 \\ \vdots & \ddots & \vdots & \vdots & \vdots \end{pmatrix} \hat{\mathbf{b}} = b_c^* \begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \end{pmatrix} + (b_c^* - b_{c-1}^*) \begin{pmatrix} 1 \\ 2 \\ 3 \\ \vdots \end{pmatrix}, \quad (3.26)$$

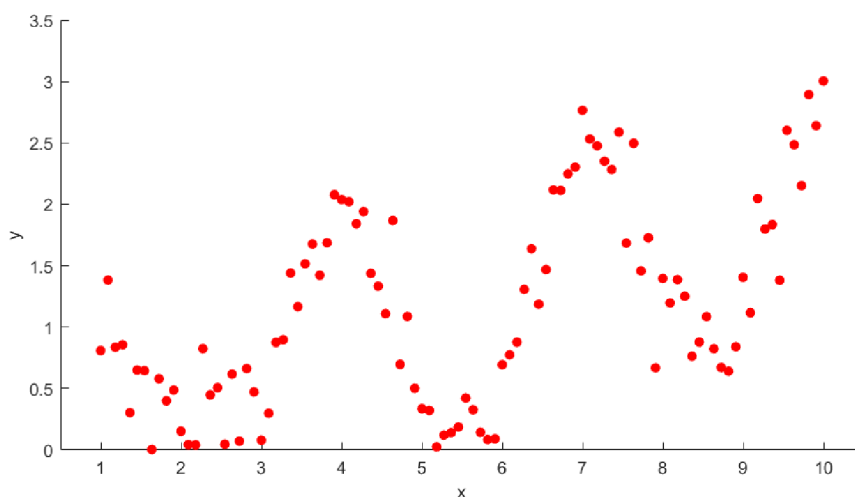
skutečně se tedy jedná o lineární kombinaci posledních dvou odhadnutých koeficientů.

Existuje mnoho případů, kdy výsledný splajn není řádem difference ovlivněn nějak významně, ale jak bylo uvedeno výše, mezi jeho podobou a řádem

diference existuje přímá souvislost.

Predikování v penalizované regresi si ilustrujeme na příkladu.

Příklad 3.1. Necht' jsou dána data zobrazená na obrázku 5. Vektor bodů \mathbf{x} je vygenerován pomocí funkce `linspace` tak, aby obsahoval 100 rovnoměrně rozložených bodů v intervalu $[1, 10]$, tj. $n = 100$, $a_1 = 1$ a $a_2 = 10$. Vektor \mathbf{y} je vytvořen za pomoci funkcí `sin` a `abs` a funkce `randn` pro generování šumu.



Obrázek 5: Zadaná data

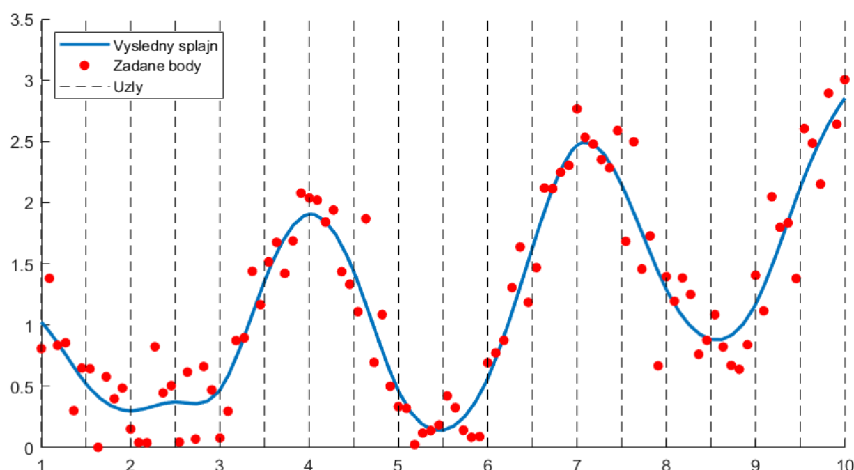
Předpokládejme, že budeme chtít predikovat hodnoty vektoru \mathbf{y}_p příslušející hodnotám vektoru \mathbf{x}_p , jehož složky leží napravo od složek vektoru \mathbf{x} , a který obsahuje 20 hodnot generovaných podobně jako hodnoty vektoru \mathbf{x} , tj. pomocí funkce `linspace` tak, aby byla zachována konstantní vzdálenost mezi body a byl jich požadovaný počet, tj. $n_p = 20$, a proto $n_+ = 120$ a $a_3 = 12$.

Nejprve budeme chtít známá data (x_j, y_j) , $j = 1, \dots, 100$, aproximovat splajnem ve smyslu penalizované metody nejmenších čtverců. Postupujeme tedy analogicky jako v Příkladu 2.1, přičemž volíme stupeň splajnu 3, tj. $k =$

3, a síť uzlů jako posloupnost bodů od 1 do 10 s krokem 0.5, tj. $[1 : 0.5 : 10]$. Nyní je třeba přidat 6 pomocných uzlů, které lze zvolit tak, že rozšířená síť uzlů je dána jako

$$[-0.5, 0, 0.5, 1 : 0.5 : 10, 10.5, 11, 11.5],$$

tj. $c = 21$. Poté pomocí funkce `spsol` vytvoříme kolokační matici $\mathbf{C}_4(\mathbf{x}) \in \mathcal{R}^{100,21}$. Zvolíme vyhlazující parametr $\rho = 0.1$ a řád diference v penalizačním členu jako $d = 1$. Výsledný splajn je zobrazen na obrázku 6.



Obrázek 6: Aproximace dat ve smyslu metody penalizovaných nejmenších čtverců

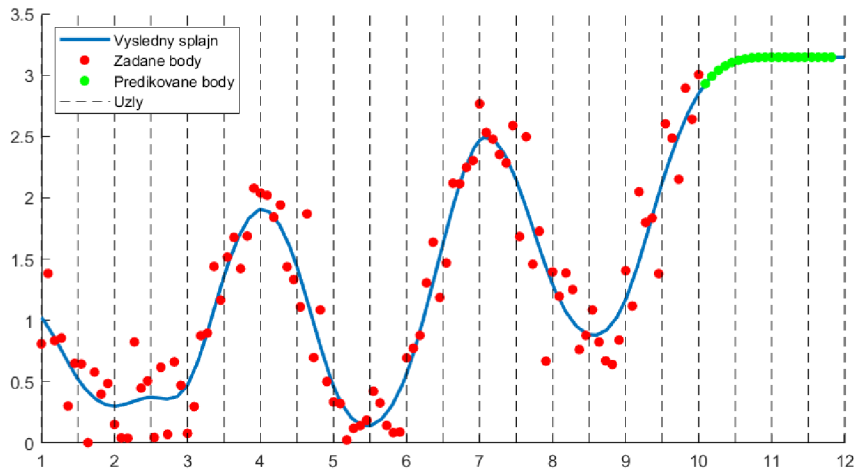
Nyní jsou naším cílem predikce, konkrétně chceme odhadnout neznámé hodnoty \mathbf{y}_p . Nová síť uzlů je zvolena tak, že obsahuje původní uzly $[1 : 0.5 : 10]$ a nové uzly tak, aby pokrývala celý rozsah hodnot vektoru $\mathbf{x}_+ = (\mathbf{x}^T, \mathbf{x}_p^T)^T$, tj. interval $[1, 12]$. Uvažujeme tedy síť uzlů

$$[1 : 0.5 : 10, 10.5, 11, 11.5, 12]$$

a přidáním 6 pomocných uzlů poté vytvoříme rozšířenou síť uzlů

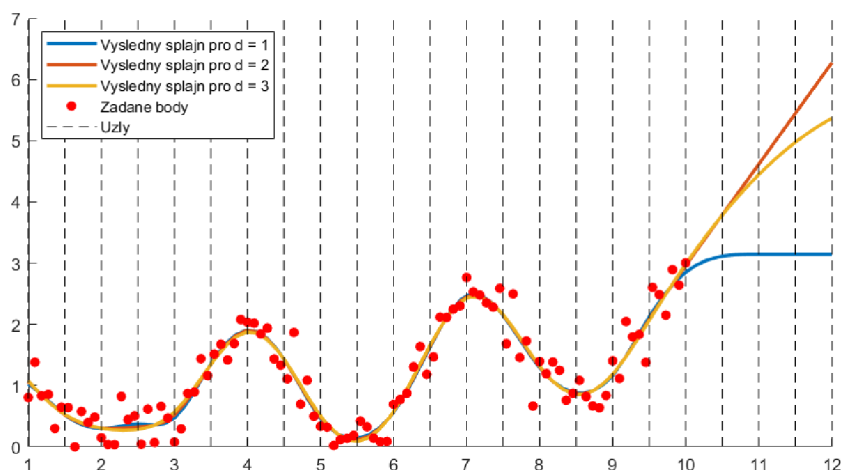
$$[-0.5, 0, 0.5, 1 : 0.5 : 10, 10.5, 11, 11.5, 12, 12.5, 13, 13.5].$$

Pomocí těchto uzlů a vektoru \mathbf{x}_+ vytvoříme kolokační matici $\mathbf{C}_4^+(\mathbf{x}_+) \in \mathcal{R}^{120,24}$ a následně i matici diferencí $\mathbf{D}_1^+ \in \mathcal{R}^{23,24}$, čímž získáme i matice $\mathbf{D}_1^1 \in \mathcal{R}^{3,21}$, $\mathbf{D}_1^2 \in \mathcal{R}^{3,3}$ a $\mathbf{P}_+ = (\mathbf{D}_1^+)^T \mathbf{D}_1^+$. Poté můžeme ověřit, že první blok matice \mathbf{D}_1^+ odpovídá matici \mathbf{D}_1 a stejně tak první blok matice $\mathbf{C}_4^+(\mathbf{x}_+)$ odpovídá kolokační matici $\mathbf{C}_4(\mathbf{x})$ pro původní data. Pomocí funkce `linsolve` pak můžeme prostřednictvím vztahu (3.23) určit odhad $\hat{\mathbf{b}}_p$ vektoru B -splajnových koeficientů \mathbf{b}_p a položit $\hat{\mathbf{b}}_+ = (\hat{\mathbf{b}}^T, \hat{\mathbf{b}}_p^T)^T$. Poznamenejme, že by bylo možné postupovat i tak, že určíme váhovou matici \mathbf{M} a k odhadu $\hat{\mathbf{b}}$ vektoru B -splajnových koeficientů \mathbf{b} využijeme vztah 3.12. Tento postup je praktikován v Příkladu 3.2 a Příkladu 3.3. Aproximované a predikované hodnoty $\hat{\mathbf{y}}_+ = (\hat{\mathbf{y}}^T, \hat{\mathbf{y}}_p^T)^T$ lze poté získat využitím vztahu (3.9), přičemž predikované hodnoty $\hat{\mathbf{y}}_p$ jsou zobrazeny na obrázku 7 včetně výsledného splajnu.



Obrázek 7: Predikce dat v penalizované regresii

Na závěr ukážeme, jak se na výsledném splajnu odrazí jiná volba řádu diferencí v penalizačním členu. Na obrázku 8 jsou vykresleny výsledné splajny pro $d = 1, 2$ a 3 , což jsou obvyklé volby tohoto parametru.



Obrázek 8: Výsledné splajny pro různé d

Dosud jsme uvažovali tzv. predikci vpřed, tedy predikci pro hodnoty \mathbf{x}_p , které se nachází vpravo od vektoru \mathbf{x} , tedy v intervalu $[a_2, a_3]$. V praxi bychom mohli chtít predikovat i hodnoty vlevo od vektoru \mathbf{x} , tj. z intervalu $[a_0, a_1]$, případně chybějící hodnoty z intervalu $[a_1, a_2]$.

Příklad 3.2. V tomto příkladu si ukážeme tzv. predikci vzad, neboli predikci pro hodnoty \mathbf{x}_p nacházející se vlevo od vektoru \mathbf{x} . Využijeme data z Příkladu 3.1, která jsou zobrazena na obrázku 5, a předpokládáme, že chceme predikovat hodnoty vektoru \mathbf{y}_p příslušející hodnotám vektoru \mathbf{x}_p . Vektor \mathbf{x}_p obsahuje opět 20 hodnot generovaných tak, aby byla zachována konstantní vzdálenost mezi body, byl jich požadovaný počet, tj. $n_p = 20$, a aby se jeho složky nacházely vlevo od složek vektoru \mathbf{x} . Platí tedy $a_0 = -1$, $a_1 = 1$ a $a_2 = 10$.

Nalezením splajnu stupně 3, tj. $k = 3$, který aproximuje data (x_i, y_i) , $i = 1, \dots, 100$, ve smyslu penalizované metody nejmenších čtverců pro $d = 1$ a $\rho = 0.1$, jsme se zabývali již v Příkladu 3.1 a výsledný splajn je zobrazen na obrázku 6. Můžeme tedy rovnou přejít k predikování hodnot vektoru \mathbf{y}_p . Nejprve zvolíme síť uzlů, která bude pokrývat rozsah všech hodnot vektoru \mathbf{x}_+ , tedy budeme uvažovat například $[-1, -0.5, 0, 0.5, 1 : 0.5 : 10]$. Nyní je třeba přidat 6 pomocných uzlů, čímž získáme rozšířenou síť uzlů

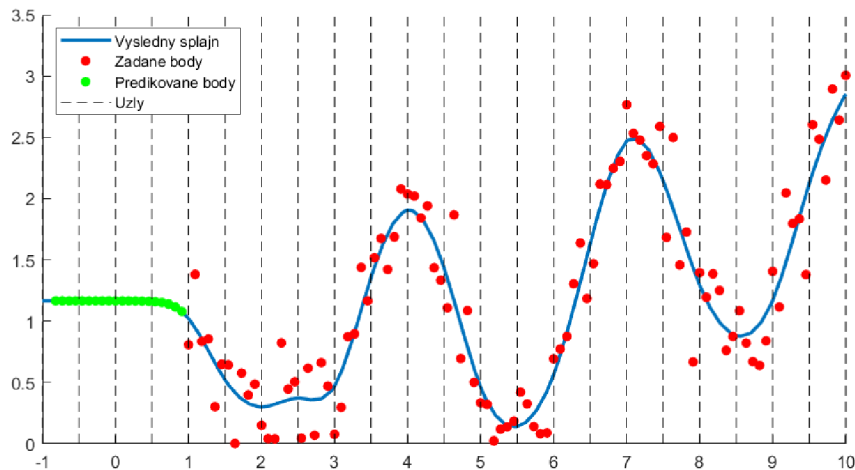
$$[-2.5, -2, -1.5, -1, -0.5, 0, 0.5, 1 : 0.5 : 10, 10.5, 11, 11.5].$$

Dále vytvoříme matici \mathbf{D}_+ aplikováním funkce `diff` na matici \mathbf{I}_{25} a využitím rozšířené sítě uzlů a vektoru \mathbf{x}_+ vytvoříme rozšířenou kolokační matici $\mathbf{C}_4^+(\mathbf{x}_+)$.

Aby bylo možné získat odhad $\hat{\mathbf{b}}_+$ vektoru B -splajnových koeficientů \mathbf{b}_+ dle vztahu (3.12), je třeba ještě určit matici \mathbf{M} , pro kterou opět platí, že její diagonála je složena z 0 pro chybějící hodnoty vektoru \mathbf{y}_p a 1 pro pozorované hodnoty, tj. pro hodnoty vektoru \mathbf{y} . Matice \mathbf{M} je tedy ve tvaru

$$\mathbf{M} = \begin{bmatrix} \mathbf{0}_{20,20} & \mathbf{0}_{100,20} \\ \mathbf{0}_{20,100} & \mathbf{I}_{100} \end{bmatrix}.$$

Poté již můžeme využít vztah (3.12) a pomocí funkce `linsolve` určit vektor $\hat{\mathbf{b}}_+$, přičemž za \mathbf{y}_+ dosadíme vektor $(\mathbf{0}^T, \mathbf{y}^T)^T$, který obsahuje 0 na pozicích odpovídajících neznámým hodnotám \mathbf{y}_p , neboť dle výše popsané teorie víme, že vztah lze upravit tak, že na těchto hodnotách nezávisí. Výsledný splajn vytvořený a vykreslený pomocí funkcí `spmak` a `fnplt` je zobrazen včetně predikovaných hodnot $\hat{\mathbf{y}}_p$ vektoru \mathbf{y}_p na obrázku 9.

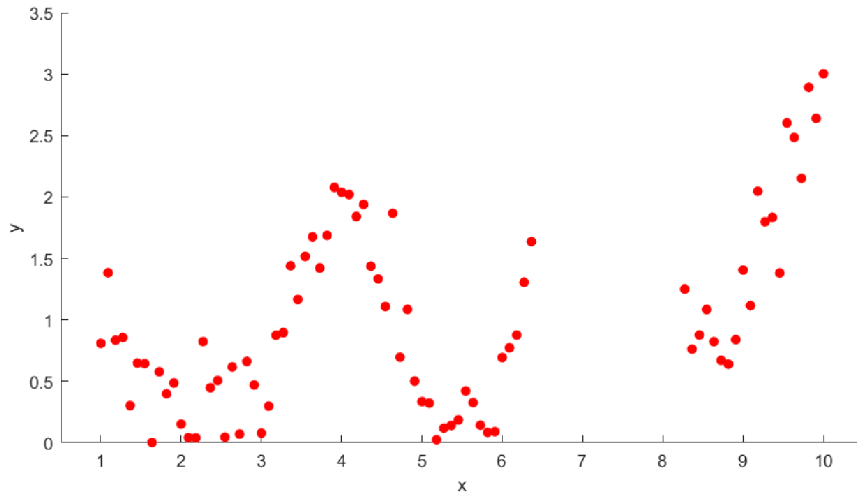


Obrázek 9: Predikce vzad v penalizované regresii

Poslední možností, kterou se v této části práce budeme zabývat, je situace, kdy data obsahují nějaká chybějící pozorování uvnitř intervalu $[a_1, a_2]$. Konkrétně uvažujeme pouze jednu skupinu chybějících pozorování.

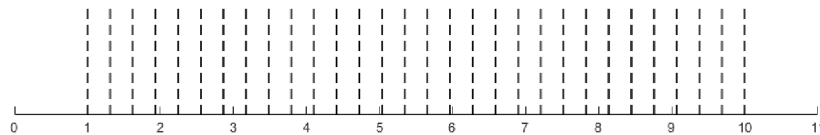
Příklad 3.3. Pro ukázkou predikce chybějících pozorování budeme opět uvažovat data z Příkladu 3.1. Předpokládáme tedy, že je dán vektor \mathbf{x}_+ takový, že obsahuje 100 hodnot generovaných rovnoměrně na intervalu $[1, 10]$, ale pro všechny hodnoty vektoru \mathbf{x}_+ neznáme odpovídající hodnoty, tj. vektor \mathbf{y}_+ , nýbrž známe pouze vektor \mathbf{x} a odpovídající hodnoty \mathbf{y} . Data daná vektory \mathbf{x} a \mathbf{y} jsou zobrazena na obrázku 10, přičemž platí $n = 80$. Naším úkolem je nejprve najít splajn, který tato data aproximuje ve smyslu penalizované metody nejmenších čtverců, a poté predikovat hodnoty \mathbf{y}_p odpovídající hodnotám vektoru \mathbf{x}_p , který je tvořen hodnotami vektoru \mathbf{x}_+ neobsaženými ve vektoru \mathbf{x} .

Nejprve tedy budeme chtít najít splajn, který data aproximuje ve smyslu penalizované metody nejmenších čtverců. Zvolíme stupeň splajnu $k = 3$ a vyjdeme ze sítě uzlů pokrývající celý rozsah vektoru \mathbf{x}_+ , tedy uvažujeme síť uzlů



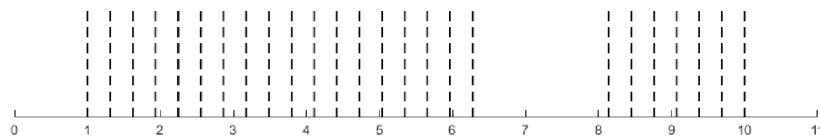
Obrázek 10: Zadaná data

obsahující 30 uzlů rovnoměrně pokrývajících interval $[1, 10]$, která je zobrazena na obrázku 11.



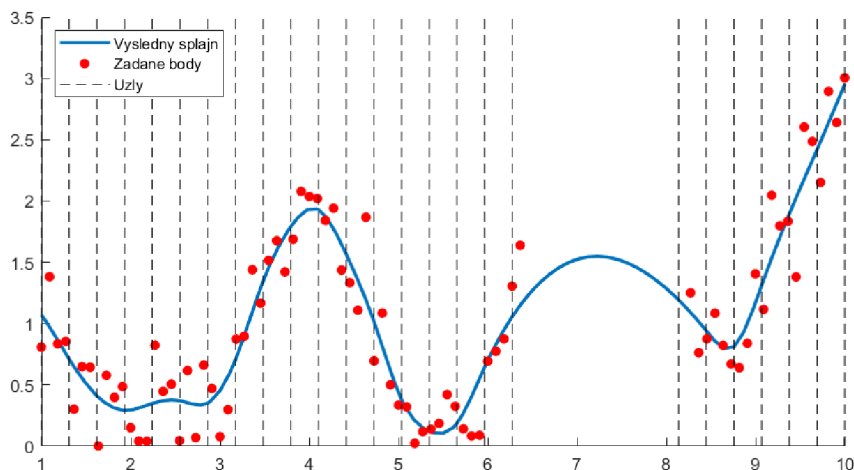
Obrázek 11: Síť uzlů pokrývajících celý rozsah vektoru \mathbf{x}_+

Aby byla splněna Schoenberg-Whitneyho podmínka a kolokační matice byla plně sloupcové hodnosti, musí v nosiči každého B -splajnu ležet alespoň jeden bod x_i , kde $i = 1, \dots, 80$. Proto je třeba z této sítě uzlů odebrat uzly, které pokrývají rozsah vektoru \mathbf{x}_p , a tedy tuto podmínku porušují. Tím získáme síť uzlů zobrazenou na obrázku 12.



Obrázek 12: Síť uzlů pokrývajících hodnoty vektoru \mathbf{x}

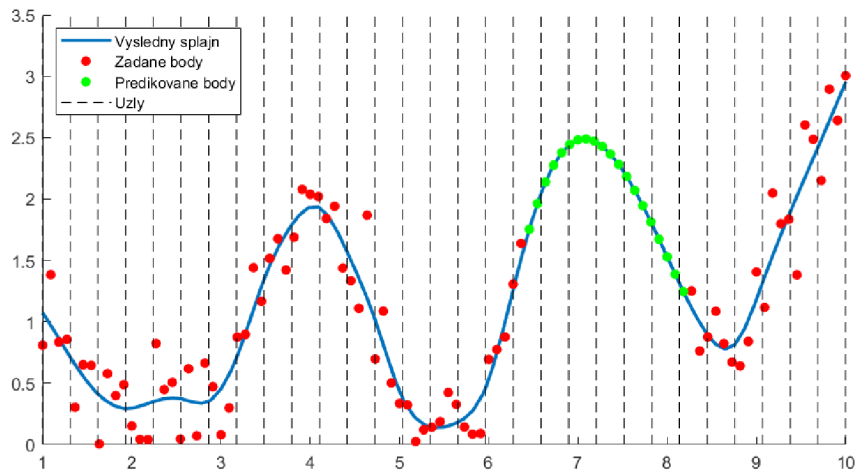
Z této sítě uzlů vytvoříme přidáním 6 pomocných uzlů rozšířenou síť uzlů, kterou využijeme spolu s vektorem \mathbf{x} k vytvoření kolokační matice $\mathbf{C}_4(\mathbf{x})$ za pomoci funkce `spcol`. Po ověření plné sloupcové hodnosti kolokační matice zvolíme řád diference v penalizačním členu $d = 2$ a vyhlazující parametr $\rho = 0.1$ a pomocí funkce `diff` vytvoříme matici $\mathbf{D}_2 \in \mathcal{R}^{25,27}$. Dále již můžeme určit odhad $\hat{\mathbf{b}}$ vektoru B -splajnových koeficientů \mathbf{b} tak, že za pomoci funkce `linsolve` vyřešíme rovnici danou vztahem (2.7). Splajn daný obdrženým vektorem $\hat{\mathbf{b}}$ je zobrazen na obrázku 13.



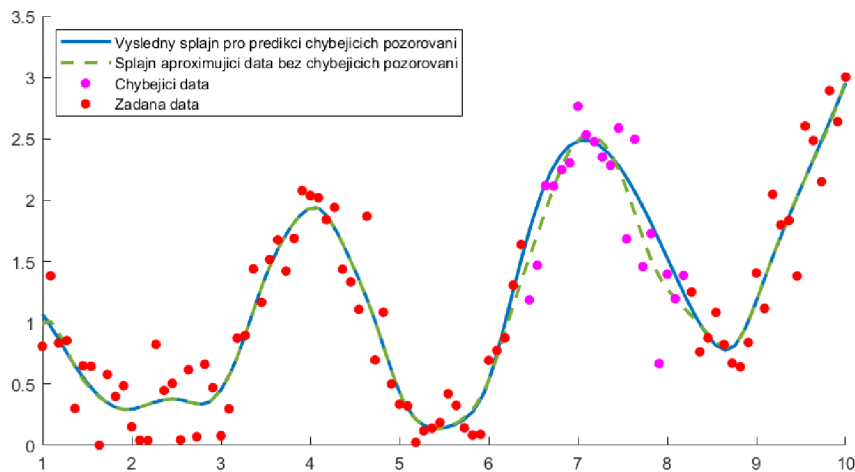
Obrázek 13: Aproximace splajnem ve smyslu penalizované metody nejmenších čtverců

Nyní již přejdeme k predikování požadovaných hodnot. Síť uzlů zvolíme tak, aby pokrývala celý rozsah vektoru \mathbf{x}_+ , tedy síť uzlů zobrazenou na obrázku 11, přičemž je třeba opět přidat 6 pomocných uzlů, čímž získáme rozšířenou síť uzlů. Následně vytvoříme matice $\mathbf{C}_4^+(\mathbf{x})$, \mathbf{D}_d^+ , \mathbf{P}_+ a matici \mathbf{M} , která je diagonální a na diagonále obsahuje 1 na pozici odpovídající hodnotám vektoru \mathbf{x} a 0 na pozici odpovídající hodnotám vektoru \mathbf{x}_p . Poté již můžeme určit odhad $\hat{\mathbf{b}}$ vektoru B -splajnových koeficientů \mathbf{b} dle vztahu (3.12)

za pomoci funkce `linsolve`. Výsledný splajn daný vektorem $\hat{\mathbf{b}}$ je zobrazen na obrázku 14 včetně predikovaných hodnot $\hat{\mathbf{y}}_p$.



Obrázek 14: Predikce pro chybějící pozorování v penalizované regresii



Obrázek 15: Porovnání splajnů

Na závěr by bylo vhodné výsledný splajn z obrázku 14 porovnat se splajnem, který aproximuje ve smyslu penalizované metody nejmenších čtverců

data, která neobsahují žádná chybějící pozorování, tedy data zobrazená na obrázku 5. Tento splajn získáme analogicky jako v Příkladu 2.1, přičemž uvažujeme stejnou síť uzlů, stupeň splajnu, vyhlazující parametr a řád diferencí v penalizačním členu jako doposud. Na obrázku 15 lze pozorovat, že porovnávané splajny jsou velmi podobné, a opět se potvrzuje, že v intervalech známých hodnot jsou dokonce totožné.

4. Predikce pomocí hladkých modelů smíšených efektů

Už před více než třiceti lety bylo zkoumáno, jak by bylo možné propojit penalizované vyhlazování a smíšené modely. Klíčovým bodem tohoto propojení je nastavení vyhlazujícího parametru na hodnotu poměru parametrů rozptylů náhodné chyby ϵ a náhodných efektů α , tedy ve tvaru $\rho = \frac{\sigma_\epsilon^2}{\sigma_\alpha^2}$. Přičemž oba parametry rozptylů lze odhadnout pomocí tzv. restringované metody maximální věrohodnosti (REML).

4.1. Lineární smíšený model

Nejprve si uvedeme, co je to lineární smíšený model neboli lineární model smíšených efektů (anglicky *linear mixed model* nebo *linear mixed-effects model*). Od regresního lineárního modelu $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ (viz [2]) se liší tím, že jeho náhodná složka je rozložena do dvou komponent. Již tedy nepředpokládáme, že variabilita je tvořena pouze náhodnou chybou $\boldsymbol{\epsilon}$, ale přidáváme do modelu další náhodnou složku reprezentovanou tzv. náhodnými efekty (anglicky *random effects*). Vektor regresních koeficientů $\boldsymbol{\beta}$ pak nazýváme vektorem pevných efektů (anglicky *fixed effects*).

Nechť je dáno m subjektů a pro každý z nich vektor pozorování \mathbf{y}_i délky n_i , $i = 1, \dots, m$. Nejprve uvažujme pevnou část modelu, tj. r vysvětlujících proměnných, jejichž hodnoty pro i -té pozorování jsou obsaženy v matici $\mathbf{X}_i \in \mathcal{R}^{n_i, r}$, $i = 1, \dots, m$, přičemž předpokládáme, že všechny tyto matice jsou plně sloupcové hodnosti, tj. obsahují lineárně nezávislé sloupce. Nyní se zaměříme na náhodnou část modelu a obdobně uvažujme matici hodnot s vysvětlujících proměnných pro i -té pozorování $\mathbf{Z}_i \in \mathcal{R}^{n_i, s}$, $i = 1, \dots, m$. Lineární smíšený

model pro i -tý subjekt, $i = 1, \dots, m$, je ve tvaru

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \boldsymbol{\alpha}_i + \boldsymbol{\epsilon}_i, \quad (4.1)$$

kde $\boldsymbol{\beta}$ je vektor pevných efektů délky r , tedy vektor regresních koeficientů příslušejících k r vysvětlujícím proměnným, který je stejný pro všechny subjekty, $\boldsymbol{\alpha}_i$ je vektor náhodných efektů délky s pro i -tý subjekt, tedy vektor regresních koeficientů příslušejících k s vysvětlujícím proměnným a $\boldsymbol{\epsilon}_i$ je vektor reziduí délky n_i . Navíc předpokládejme, že $\boldsymbol{\alpha}_i \sim \mathcal{N}(0, \mathbf{D})$ a $\boldsymbol{\epsilon}_i \sim \mathcal{N}(0, \mathbf{R}_i)$, přičemž varianční matice $\mathbf{D} \in \mathcal{R}^{s,s}$ a $\mathbf{R}_i \in \mathcal{R}^{n_i, n_i}$, $i = 1, \dots, m$, jsou symetrické a pozitivně definitní, a $\boldsymbol{\alpha}_i$ a $\boldsymbol{\epsilon}_i$ jsou nezávislé pro všechna $i = 1, \dots, m$. Pak lze vytvořit lineární smíšený model pro všechny subjekty

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\epsilon}, \quad (4.2)$$

kde $\mathbf{y} = (\mathbf{y}_1^T, \dots, \mathbf{y}_m^T)^T$ je vektor délky $n = \sum_{i=1}^m n_i$, $\mathbf{X} = [\mathbf{X}_1^T, \dots, \mathbf{X}_m^T]^T$ je matice typu $n \times r$, $\mathbf{Z} \in \mathcal{R}^{n, ms}$ je blokově diagonální matice, jejíž diagonála je tvořena jednotlivými maticemi \mathbf{Z}_i , $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^T, \dots, \boldsymbol{\alpha}_m^T)^T$ je vektor náhodných efektů délky ms a $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1^T, \dots, \boldsymbol{\epsilon}_m^T)^T$ je vektor reziduí délky n . Dále víme, že $\boldsymbol{\alpha} \sim \mathcal{N}(0, \mathbf{G})$ a $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{R})$, kde $\mathbf{G} \in \mathcal{R}^{ms, ms}$ je blokově diagonální matice, jejíž diagonála je tvořena m maticemi \mathbf{D} a podobně $\mathbf{R} \in \mathcal{R}^{n, n}$ je blokově diagonální matice, jejíž diagonála je tvořena maticemi \mathbf{R}_i , $i = 1, \dots, m$.

4.1.1. Odhad pevných efektů

Nyní se zaměříme na to, jak odhadnout pevné efekty, tj. vektor $\boldsymbol{\beta}$. Jednou možností je přepsat model (4.2) do tvaru $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}^*$, kde $\boldsymbol{\epsilon}^* = \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\epsilon}$. Protože jsou $\boldsymbol{\alpha}$ a $\boldsymbol{\epsilon}$ nezávislé, dle [13] víme, že varianční matice součtu je součet variančních matic, tj. $\text{Var}(\mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\epsilon}) = \text{Var}(\mathbf{Z}\boldsymbol{\alpha}) + \text{Var}(\boldsymbol{\epsilon})$, kde

$\text{Var}(\mathbf{Z}\boldsymbol{\alpha}) = \mathbf{Z}\text{Var}(\boldsymbol{\alpha})\mathbf{Z}^T$. Pro varianční matici $\boldsymbol{\epsilon}^*$ označenou jako \mathbf{V} tedy platí

$$\mathbf{V} = \text{Var}(\boldsymbol{\epsilon}^*) = \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{R}. \quad (4.3)$$

Protože jsme předpokládali, že matice \mathbf{D} a \mathbf{R}_i jsou pozitivně definitní, jsou dle [9], str. 89, pozitivně definitní i blokově diagonální matice z těchto matic složené, tedy matice \mathbf{G} a \mathbf{R} . Následně využijeme toho, že potom je matice $\mathbf{Z}\mathbf{G}\mathbf{Z}^T$ dle [9], str. 213, také pozitivně definitní. Protože součet dvou pozitivně definitních matic je dle [9], str. 212, matice pozitivně definitní, platí, že varianční matice \mathbf{V} je pozitivně definitní, a tedy i regulární.

Pomocí metody maximální věrohodnosti určíme odhad pevných efektů. Dle [13] víme, že sdružená hustota vektoru $\mathbf{y} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$ je dána jako

$$f_{\mathbf{y}}(\mathbf{y}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\mathbf{V}|} \exp\left(-\frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right)$$

a zlogaritmováním dojdeme k tzv. logaritmické funkci věrohodnosti

$$\ln(f_{\mathbf{y}}(\mathbf{y})) = \ln(1) - \frac{n}{2} \ln(2\pi) - \ln(|\mathbf{V}|) - \frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \quad (4.4)$$

Tuto funkci chceme maximalizovat, a proto ji zderivujeme podle proměnné $\boldsymbol{\beta}^T$ a výslednou derivaci položíme rovnu nulovému vektoru, čímž získáme rovnici

$$\mathbf{X}^T \mathbf{V}^{-1} \mathbf{y} - \mathbf{X}^T \mathbf{V}^{-1} \mathbf{X} \boldsymbol{\beta} = \mathbf{0},$$

a tedy

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{y}. \quad (4.5)$$

Protože je matice \mathbf{V} pozitivně definitní, její inverze \mathbf{V}^{-1} je dle [9], str. 214, také pozitivně definitní. Předpokládali jsme, že matice \mathbf{X}_i , $i = 1, \dots, m$, mají lineárně nezávislé sloupce, a proto platí, že i matice \mathbf{X} má lineárně nezávislé

sloupce, a je tedy plné sloupcové hodnosti. Díky plné sloupcové hodnosti matice \mathbf{X} a pozitivní definitnosti matice \mathbf{V}^{-1} pak platí, že $\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}$ je také pozitivně definitní [9], str. 213, tedy i regulární.

4.1.2. Odhad náhodných efektů

Následně nás zajímá, jak najít odhad náhodných efektů. Za tímto účelem využijeme [2] a zjistíme, že jej lze získat jako nejlepší lineární nestranný odhad (BLUP) jako podmíněnou střední hodnotu $E(\boldsymbol{\alpha} \mid \mathbf{y})$. Poznamenejme ale, že zde je nestrannost myšlena ve smyslu $E(\hat{\boldsymbol{\alpha}}) = E(\boldsymbol{\alpha}) = \mathbf{0}$ a obvykle definovaná nestrannost neplatí, tj. $E(\hat{\boldsymbol{\alpha}}) \neq \boldsymbol{\alpha}$. Nyní uveďme větu, kterou lze včetně důkazu najít v [12].

Věta 4.1. *Nechť $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, kde*

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}, \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}.$$

Pak platí $\mathbf{X}_1 \sim \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$, $\mathbf{X}_2 \sim \mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$ a navíc

$$\mathbf{X}_1 \mid (\mathbf{X}_2 = \mathbf{x}_2) \sim \mathcal{N}(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} (\mathbf{x}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21}).$$

Nejprve je třeba určit kovarianci \mathbf{y} a $\boldsymbol{\alpha}$, tj. $\text{Cov}(\mathbf{y}, \boldsymbol{\alpha})$. Dle [13], str. 129, víme, že platí

$$\text{Cov}(\mathbf{y}, \boldsymbol{\alpha}) = E(\mathbf{y}\boldsymbol{\alpha}) - E(\mathbf{y})E(\boldsymbol{\alpha}), \quad (4.6)$$

a proto lze dokázat, že platí

$$\begin{aligned} \text{Cov}(\mathbf{y}, \boldsymbol{\alpha}) &= \text{Cov}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\epsilon}, \boldsymbol{\alpha}) = \\ &= \text{Cov}(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\alpha}) + \text{Cov}(\mathbf{Z}\boldsymbol{\alpha}, \boldsymbol{\alpha}) + \text{Cov}(\boldsymbol{\epsilon}, \boldsymbol{\alpha}). \end{aligned}$$

Díky nezávislosti $\mathbf{X}\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ a $\boldsymbol{\epsilon}$, $\boldsymbol{\alpha}$ platí, že $\text{Cov}(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\alpha}) = \mathbf{0}$, $\text{Cov}(\boldsymbol{\epsilon}, \boldsymbol{\alpha}) = \mathbf{0}$, a využitím (4.6) dostaneme $\text{Cov}(\mathbf{Z}\boldsymbol{\alpha}, \boldsymbol{\alpha}) = \text{E}(\mathbf{Z}\boldsymbol{\alpha}\boldsymbol{\alpha}^T) - \text{E}(\mathbf{Z}\boldsymbol{\alpha})\text{E}(\boldsymbol{\alpha}^T)$. Protože střední hodnota náhodných efektů je nulová a matice \mathbf{Z} je nenáhodná, platí

$$\text{Cov}(\mathbf{y}, \boldsymbol{\alpha}) = \mathbf{Z}\text{E}(\boldsymbol{\alpha}\boldsymbol{\alpha}^T) = \mathbf{Z}\text{Var}(\boldsymbol{\alpha}) = \mathbf{Z}\mathbf{G}.$$

Pak $\text{Cov}(\boldsymbol{\alpha}, \mathbf{y}) = \mathbf{G}^T\mathbf{Z}^T = \mathbf{G}\mathbf{Z}^T$, a tedy celkem platí

$$\begin{bmatrix} \boldsymbol{\alpha} \\ \mathbf{y} \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{X}\boldsymbol{\beta} \end{bmatrix}, \begin{bmatrix} \mathbf{G} & \mathbf{G}\mathbf{Z}^T \\ \mathbf{Z}\mathbf{G} & \mathbf{V} \end{bmatrix} \right),$$

a tedy dle Věty 4.1 platí, že

$$\boldsymbol{\alpha} \mid \mathbf{y} \sim \mathcal{N}(\mathbf{G}\mathbf{Z}^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \mathbf{G} - \mathbf{G}\mathbf{Z}^T\mathbf{V}^{-1}\mathbf{Z}\mathbf{G}).$$

Odhad náhodných efektů pak lze získat jako

$$\hat{\boldsymbol{\alpha}} = \text{E}(\boldsymbol{\alpha} \mid \mathbf{y}) = \mathbf{G}\mathbf{Z}^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \quad (4.7)$$

přičemž v praxi nahrazujeme $\boldsymbol{\beta}$ jejím odhadem $\hat{\boldsymbol{\beta}}$, viz (4.5).

Maximálně věrohodné odhady $\hat{\boldsymbol{\beta}}$ a $\hat{\boldsymbol{\alpha}}$ lze získat i jinak. Vyjdeme ze sdružené hustoty \mathbf{y} a $\boldsymbol{\alpha}$, tj. $f_{\mathbf{y},\boldsymbol{\alpha}}(\mathbf{y}, \boldsymbol{\alpha})$. Tu lze dle [13], str.122, získat vynásobením hustot podmíněného pravděpodobnostního rozdělení $\mathbf{y} \mid \boldsymbol{\alpha}$ a pravděpodobnostního rozdělení $\boldsymbol{\alpha}$, tj.

$$f_{\mathbf{y},\boldsymbol{\alpha}}(\mathbf{y}, \boldsymbol{\alpha}) = f_{\mathbf{y}|\boldsymbol{\alpha}}(\mathbf{y} \mid \boldsymbol{\alpha})f_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}). \quad (4.8)$$

Nyní se zaměříme na podmíněné rozdělení $\mathbf{y} \mid \boldsymbol{\alpha}$. Víme, že platí

$$\begin{bmatrix} \mathbf{y} \\ \boldsymbol{\alpha} \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \mathbf{X}\boldsymbol{\beta} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{V} & \mathbf{Z}\mathbf{G} \\ \mathbf{G}\mathbf{Z}^T & \mathbf{G} \end{bmatrix} \right),$$

a proto dle Věty 4.1 víme, že

$$\mathbf{y} \mid \boldsymbol{\alpha} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{G}\mathbf{G}^{-1}(\boldsymbol{\alpha} - \mathbf{0}), \mathbf{V} - \mathbf{Z}\mathbf{G}\mathbf{G}^{-1}\mathbf{G}\mathbf{Z}^T),$$

což lze využitím (4.3) upravit do tvaru

$$\mathbf{y} \mid \boldsymbol{\alpha} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha}, \mathbf{R}). \quad (4.9)$$

Využitím (4.9) a $\boldsymbol{\alpha} \sim \mathcal{N}(\mathbf{0}, \mathbf{G})$ lze odvodit hustoty těchto rozdělení

$$\begin{aligned} f_{\mathbf{y}|\boldsymbol{\alpha}}(\mathbf{y} \mid \boldsymbol{\alpha}) &= \frac{1}{(2\pi)^{\frac{n}{2}} |\mathbf{R}|} \exp \left(-\frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{\alpha})^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{\alpha}) \right), \\ f_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}) &= \frac{1}{(2\pi)^{\frac{s}{2}} |\mathbf{G}|} \exp \left(-\frac{1}{2} \boldsymbol{\alpha}^T \mathbf{G}^{-1} \boldsymbol{\alpha} \right). \end{aligned}$$

Dosazením těchto funkcí do vztahu (4.8) a zlogaritmováním dostaneme

$$\begin{aligned} \ln(f_{\mathbf{y},\boldsymbol{\alpha}}(\mathbf{y}, \boldsymbol{\alpha})) &= \ln \left(\frac{1}{(2\pi)^{\frac{n}{2}} |\mathbf{R}|} \right) - \frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{\alpha})^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{\alpha}) \\ &\quad + \ln \left(\frac{1}{(2\pi)^{\frac{s}{2}} |\mathbf{G}|} \right) - \frac{1}{2} \boldsymbol{\alpha}^T \mathbf{G}^{-1} \boldsymbol{\alpha}. \end{aligned}$$

Zderivováním této funkce podle proměnných $\boldsymbol{\beta}$ a $\boldsymbol{\alpha}$ za účelem její maximalizace a položením těchto derivací rovných nule získáme tzv. rovnice smíšeného modelu

$$\begin{bmatrix} \mathbf{X}^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{X}^T \mathbf{R}^{-1} \mathbf{Z} \\ \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\alpha}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T \mathbf{R}^{-1} \mathbf{y} \\ \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{y} \end{bmatrix}, \quad (4.10)$$

ze kterých lze také určit odhady pevných a náhodných efektů pomocí metod pro řešení soustav lineárních rovnic.

4.1.3. Odhad parametrů rozptylů σ_ϵ^2 a σ_α^2

Dosazením odhadu pevných efektů (4.5) do (4.4) získáme při zanedbání konstant funkci proměnné \mathbf{V} , kterou lze upravit do tvaru

$$l_{ML}(\mathbf{V}) = -\frac{1}{2} \ln(|\mathbf{V}|) - \frac{1}{2} \mathbf{y}^T \left(\mathbf{V}^{-1} - \mathbf{V}^{-1} \mathbf{X} (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \right) \mathbf{y}. \quad (4.11)$$

Maximalizací této funkce bychom získali prostřednictvím odhadnuté matice $\hat{\mathbf{V}}$ maximálně věrohodné odhady parametrů rozptylů σ_ϵ^2 a σ_α^2 .

Dle [2] ale platí, že maximálně věrohodné odhady jsou vychýlené směrem dolů, tj. střední hodnota odhadu parametru je menší než skutečná hodnota tohoto parametru, proto se často využívají odhady získané metodou REML, tj. restringovanou metodou maximální věrohodnosti. V této metodě vycházíme z modifikované funkce věrohodnosti, která odpovídá hustotě transformovaných dat. Konkrétně se jedná o transformaci $\mathbf{L}^T \mathbf{y}$, kde matice \mathbf{L} splňuje podmínku $\mathbf{L}^T \mathbf{X} = \mathbf{0}$. Dle [10] pak platí, že odhady rozptylů σ_ϵ^2 a σ_α^2 lze získat metodou REML prostřednictvím maximalizace funkce

$$l_{REML}(\mathbf{V}) = -\frac{1}{2} \ln(|\mathbf{V}|) - \frac{1}{2} \ln(|\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}|) - \frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

Do této funkce můžeme opět dosadit odhad pevných efektů (4.5) a porovnáním s funkcí (4.11) dojdeme ke vztahu

$$l_{REML}(\mathbf{V}) = l_{ML}(\mathbf{V}) - \frac{1}{2} \ln(|\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}|).$$

4.2. Hladké modely smíšených efektů

Nyní se již zaměříme na propojení P -splajnů, viz odstavec 2.1, a lineárních smíšených modelů, tedy na tzv. hladké modely smíšených efektů. Uvažujme lineární smíšený model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\epsilon}, \quad (4.12)$$

kde $\boldsymbol{\alpha} \sim \mathcal{N}(0, \sigma_{\boldsymbol{\alpha}}^2 \mathbf{G})$ a $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma_{\boldsymbol{\epsilon}}^2 \mathbf{I}_n)$. Existuje mnoho možností, jak převést původní model (3.2) na tento smíšený model. Cílem je najít transformaci $\boldsymbol{\Omega}$ takovou, že

$$\mathbf{C}_{k+1}(\mathbf{x}) \boldsymbol{\Omega} = [\mathbf{X} \mid \mathbf{Z}] \quad \text{a} \quad \boldsymbol{\Omega}^T \mathbf{b} = [\boldsymbol{\beta}^T \mid \boldsymbol{\alpha}^T]^T,$$

kde $\boldsymbol{\Omega}$ je ortogonální matice. Transformace je zvolena tak, abychom získali právě $\mathbf{C}_{k+1}(\mathbf{x}) \mathbf{b} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha}$.

Matici $\boldsymbol{\Omega}$ rozdělíme na dvě submatice, které odpovídají pevným, respektive náhodným složkám, tj. $\boldsymbol{\Omega} = [\boldsymbol{\Omega}_f \mid \boldsymbol{\Omega}_r]$. Tedy platí

$$\begin{aligned} \mathbf{X} &= \mathbf{C}_{k+1}(\mathbf{x}) \boldsymbol{\Omega}_f, \\ \mathbf{Z} &= \mathbf{C}_{k+1}(\mathbf{x}) \boldsymbol{\Omega}_r. \end{aligned} \quad (4.13)$$

Existuje více možností, jak zvolit submatici $\boldsymbol{\Omega}_r$, my vyjdeme ze spektrálního rozkladu matice penalizace $\mathbf{P} \in \mathcal{R}^{c,c}$, o které již víme, že je pozitivně semidefinitní, viz odstavec 2.1. Tedy uvažujeme rozklad $\mathbf{P} = \mathbf{U}\tilde{\boldsymbol{\Sigma}}\mathbf{U}^T$, kde $\tilde{\boldsymbol{\Sigma}} \in \mathcal{R}^{c,c}$ je diagonální matice, jejíž diagonála je tvořena vlastními čísly matice \mathbf{P} seřazenými vzestupně, a matice $\mathbf{U} \in \mathcal{R}^{c,c}$ obsahuje odpovídající vlastní vektory. Dle [9] platí, že vlastní čísla semidefinitní matice \mathbf{P} jsou nezáporná a ze strany 551 víme, že počet kladných vlastních čísel matice \mathbf{P} je roven hodnotě matice \mathbf{P} . Hodnota matice \mathbf{P} je dle [9], str. 75, rovna hodnotě matice \mathbf{D}_d , neboť

platí $\mathbf{P} = \mathbf{D}_d^T \mathbf{D}_d$. Protože víme, že matice $\mathbf{D}_d \in \mathcal{R}^{c-d,c}$ je plně řádkové hodnosti, platí, že počet kladných vlastních čísel matice $\mathbf{P} \in \mathcal{R}^{c,c}$ je $c-d$, a tedy počet nulových vlastních čísel matice \mathbf{P} je roven d . Matici $\tilde{\Sigma}$ pak lze zapsat ve tvaru

$$\tilde{\Sigma} = \begin{bmatrix} \mathbf{0}_{d,d} & \mathbf{0}_{d,c-d} \\ \mathbf{0}_{c-d,d} & \Sigma \end{bmatrix},$$

kde $\Sigma \in \mathcal{R}^{c-d,c-d}$ je diagonální matice, jejíž diagonála obsahuje kladná vlastní čísla matice \mathbf{P} , a matici \mathbf{U} lze zapsat ve tvaru $\mathbf{U} = [\mathbf{U}_f, \mathbf{U}_r]$, kde matice \mathbf{U}_f obsahuje vlastní vektory odpovídající nulovým vlastním číslům matice \mathbf{P} a \mathbf{U}_r obsahuje vlastní vektory odpovídající kladným vlastním číslům matice \mathbf{P} . Matici Ω_r pak lze zvolit jako $\Omega_r = \mathbf{U}_r \Sigma^{-\frac{1}{2}}$. Minimalizovaný výraz

$$\delta(\mathbf{b}) = (\mathbf{y} - \mathbf{C}_{k+1}(\mathbf{x}) \mathbf{b})^T (\mathbf{y} - \mathbf{C}_{k+1}(\mathbf{x}) \mathbf{b}) + \rho \mathbf{b}^T \mathbf{P} \mathbf{b}$$

příslušející modelu (3.2), pak lze v kontextu smíšeného modelu (4.12) přepsat do tvaru

$$\delta(\mathbf{b}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{\alpha})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{\alpha}) + \rho \mathbf{b}^T \mathbf{P} \mathbf{b}.$$

Predikce v kontextu smíšených modelů se běžně provádí tzv. dvoufázově, tj. nejprve se vytvoří model a poté se provádí predikce. Stejně se na to zaměříme v kontextu hladkých smíšených modelů. Zde bude platit, že varianční matice náhodných efektů je přímým rozšířením této matice ve vytvářeném modelu. Poté představíme i tzv. jednofázový přístup. V tom dochází k tvoření modelu a predikcím současně a výhodou je, že může být zvolena libovolná transformace.

4.3. Dvoufázový přístup

4.3.1. Standardní metodologie pro predikce

Metodu popsanou v této kapitole označujeme jako přístup smíšeného modelu. K predikci nových pozorování můžeme využít [7], kde je rozebrána predikce v lineárních smíšených modelech. Predikce je dána lineární funkcí nejlepšího lineárního nestranného odhadu náhodných efektů (BLUP) a nejlepšího lineárního nestranného odhadu pevných efektů (BLUE) v modelu. Výsledky jsou založené na následujícím rozšířeném smíšeném modelu

$$\mathbf{y}_+ = \mathbf{X}_+ \boldsymbol{\beta} + \mathbf{Z}_+ \boldsymbol{\alpha}_+ + \boldsymbol{\epsilon}_+, \quad (4.14)$$

který lze rozepsat do tvaru

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{y}_p \end{bmatrix} = \begin{bmatrix} \mathbf{X} \\ \mathbf{X}_p \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \mathbf{Z} & \mathbf{0} \\ \mathbf{Z}_1 & \mathbf{Z}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\alpha}_p \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon} \\ \boldsymbol{\epsilon}_p \end{bmatrix}, \quad (4.15)$$

kde $(\boldsymbol{\epsilon}^T, \boldsymbol{\epsilon}_p^T)^T \sim \mathcal{N}(0, \sigma_\epsilon^2 \mathbf{I}_{n_+})$, $\boldsymbol{\beta}$ jsou pevné efekty z původního modelu (4.12) a $\boldsymbol{\alpha}_+ = (\boldsymbol{\alpha}^T, \boldsymbol{\alpha}_p^T)^T$ je rozšířený vektor náhodných efektů s varianční maticí

$$\text{Var}(\boldsymbol{\alpha}_+) = \sigma_\alpha^2 \mathbf{G}_+ = \sigma_\alpha^2 \begin{bmatrix} \mathbf{G} & \mathbf{G}_{op} \\ \mathbf{G}_{po} & \mathbf{G}_{pp} \end{bmatrix}, \quad (4.16)$$

kde $\sigma_\alpha^2 \mathbf{G}$ je varianční matice náhodných efektů z původního modelu (4.2) pro pozorovaná data, $\sigma_\alpha^2 \mathbf{G}_{op}$ je matice kovariancí náhodných efektů pro pozorovaná data a pro nepozorovaná data ($\sigma_\alpha^2 \mathbf{G}_{po}$ je naopak matice kovariancí náhodných efektů pro nepozorovaná data a pro pozorovaná data) a $\sigma_\alpha^2 \mathbf{G}_{pp}$ je varianční matice náhodných efektů pro nepozorovaná data. Parametry rozptylů σ_ϵ^2 a σ_α^2 jsou odhadnuté pomocí metody REML popsané výše.

Nyní potřebujeme převést rozšířený P -splajnový model (3.3) na rozšířený smíšený model (4.14). Proto definujeme rozšířenou matici transformace

$$\mathbf{\Omega}_r^+ = \begin{bmatrix} \mathbf{\Omega}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega}_{p_r} \end{bmatrix},$$

kde $\mathbf{\Omega}_r$ je matice transformace pro pozorovaná data a $\mathbf{\Omega}_{p_r}$ je matice transformace pro nepozorovaná (tj. predikovaná) data. Pak platí $\mathbf{Z}_+ = \mathbf{C}_{k+1}^+(\mathbf{x}_+) \mathbf{\Omega}_r^+$, tj.

$$\begin{bmatrix} \mathbf{Z} & \mathbf{0} \\ \mathbf{Z}_1 & \mathbf{Z}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{C}_{k+1}(\mathbf{x}) & \mathbf{0} \\ \mathbf{C}_{k+1}^1(\mathbf{x}_p) & \mathbf{C}_{k+1}^2(\mathbf{x}_p) \end{bmatrix} \begin{bmatrix} \mathbf{\Omega}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega}_{p_r} \end{bmatrix},$$

tedy matice \mathbf{Z} je dána vztahem (4.13), $\mathbf{Z}_1 = \mathbf{C}_{k+1}^1(\mathbf{x}_p) \mathbf{\Omega}_r$ a $\mathbf{Z}_2 = \mathbf{C}_{k+1}^2(\mathbf{x}_p) \mathbf{\Omega}_{p_r}$.

Opět existuje více variant, jak $\mathbf{\Omega}_r^+$ zvolit. My ji volíme tak, že $\mathbf{\Omega}_r = \mathbf{U}_r \mathbf{\Sigma}^{-\frac{1}{2}}$, přičemž opět vycházíme ze spektrálního rozkladu penalizační matice $\mathbf{P} = \mathbf{D}_d^T \mathbf{D}_d = \mathbf{U} \tilde{\mathbf{\Sigma}} \mathbf{U}^T$, a matice $\mathbf{\Omega}_{p_r}$ je zvolena jako $\mathbf{\Omega}_{p_r} = (\mathbf{D}_d^2)^{-1}$, kde \mathbf{D}_d a \mathbf{D}_d^2 jsou submatice matice \mathbf{D}_d^+ , viz (3.6). Pomocí rozšířené matice transformace $\mathbf{\Omega}_r^+$ jsme získali varianční matici $\sigma_\alpha^2 \mathbf{G}_+$ rozšířených náhodných efektů α_+ jako rozšíření matice $\sigma_\alpha^2 \mathbf{G}$, viz (4.16).

Ze vztahu (4.15) lze odvodit vztah pro predikované hodnoty

$$\hat{\mathbf{y}}_p = \mathbf{X}_p \hat{\boldsymbol{\beta}} + \mathbf{Z}_1 \hat{\boldsymbol{\alpha}} + \mathbf{Z}_2 \hat{\boldsymbol{\alpha}}_p, \quad (4.17)$$

který bychom rádi upravili tak, aby se v něm vyskytovaly pouze odhady pevných efektů $\hat{\boldsymbol{\beta}}$ a náhodných efektů $\hat{\boldsymbol{\alpha}}$. Proto je třeba odvodit vztah pro $\hat{\boldsymbol{\alpha}}_p$, což provedeme pomocí AI algoritmu popsáném v [8]. Stejně jako v [7] vyjdeme z rozšířeného modelu

$$\mathbf{y}_a = \mathbf{F} \boldsymbol{\psi} + \mathbf{K}_a \boldsymbol{\tau}_a + \boldsymbol{\epsilon}_+,$$

kde

$$\mathbf{y}_a = \begin{bmatrix} \mathbf{y} \\ \mathbf{0} \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} \mathbf{0}_n \\ \mathbf{I}_{n_p} \end{bmatrix}, \quad \mathbf{K}_a = \begin{bmatrix} \mathbf{X} & \mathbf{Z} & \mathbf{0} \\ \mathbf{X}_p & \mathbf{Z}_1 & \mathbf{Z}_2 \end{bmatrix},$$

$$\boldsymbol{\tau}_a = \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha} \\ \boldsymbol{\alpha}_p \end{bmatrix} = \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha}_+ \end{bmatrix}, \quad \boldsymbol{\epsilon}_+ = \begin{bmatrix} \boldsymbol{\epsilon} \\ \boldsymbol{\epsilon}_p \end{bmatrix},$$

$\boldsymbol{\psi}$ je vektor pevných efektů délky n_p , $\text{Var}(\boldsymbol{\alpha}_+) = \sigma_\alpha^2 \mathbf{G}_+$, viz (4.16), a opět platí $\text{Var}(\boldsymbol{\epsilon}_a) = \sigma_\epsilon^2 \mathbf{I}_{n_+}$. Rovnice tohoto rozšířeného smíšeného modelu lze řešit právě pomocí AI algoritmu (tj. pomocí absorpce a zpětné substituce) matice

$$\begin{bmatrix} \mathbf{y}_a^T \mathbf{R}_a^{-1} \mathbf{y}_a & \mathbf{y}_a^T \mathbf{R}_a^{-1} \mathbf{K}_a & \mathbf{y}_a^T \mathbf{R}_a^{-1} \mathbf{F} \\ \mathbf{K}_a^T \mathbf{R}_a^{-1} \mathbf{y}_a & \mathbf{K}_a^T \mathbf{R}_a^{-1} \mathbf{K}_a + \mathbf{G}_a^* & \mathbf{K}_a^T \mathbf{R}_a^{-1} \mathbf{F} \\ \mathbf{F}^T \mathbf{R}_a^{-1} \mathbf{y}_a & \mathbf{F}^T \mathbf{R}_a^{-1} \mathbf{K}_a & \mathbf{F}^T \mathbf{R}_a^{-1} \mathbf{F} \end{bmatrix}, \quad \text{kde } \mathbf{G}_a^* = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_a^{-1} \end{bmatrix}.$$

Absorpcí získáme matici smíšeného modelu

$$\begin{bmatrix} \mathbf{y}^T \mathbf{R}^{-1} \mathbf{y} & \mathbf{y}^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{y}^T \mathbf{R}^{-1} \mathbf{Z} & \mathbf{0} \\ \mathbf{X}^T \mathbf{R}^{-1} \mathbf{y} & \mathbf{X}^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{X}^T \mathbf{R}^{-1} \mathbf{Z} & \mathbf{0} \\ \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{y} & \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{Z} + \mathbf{G}^{oo} & \mathbf{G}^{op} \\ \mathbf{0} & \mathbf{0} & \mathbf{G}^{po} & \mathbf{G}^{pp} \end{bmatrix}, \quad \text{kde } \mathbf{G}_a^{-1} = \begin{bmatrix} \mathbf{G}^{oo} & \mathbf{G}^{op} \\ \mathbf{G}^{po} & \mathbf{G}^{pp} \end{bmatrix}.$$

Z posledních řádků lze dle [7] získat požadovaný vztah

$$\hat{\boldsymbol{\alpha}}_p = -(\mathbf{G}^{pp})^{-1} \mathbf{G}^{po} \hat{\boldsymbol{\alpha}} = \mathbf{G}_{po} \mathbf{G}^{-1} \hat{\boldsymbol{\alpha}}. \quad (4.18)$$

Absorpcí poslední části odpovídající $\hat{\boldsymbol{\alpha}}_p$ bychom došli k matici původního

smíšeného modelu (4.2), tj. k matici

$$\begin{bmatrix} \frac{1}{\sigma_\epsilon^2} \mathbf{y}^T \mathbf{y} & \frac{1}{\sigma_\epsilon^2} \mathbf{y}^T [\mathbf{X} | \mathbf{Z}] \\ \frac{1}{\sigma_\epsilon^2} [\mathbf{X}^T | \mathbf{Z}^T]^T \mathbf{y} & \frac{1}{\sigma_\epsilon^2} [\mathbf{X}^T | \mathbf{Z}^T]^T [\mathbf{X} | \mathbf{Z}] + \frac{1}{\sigma_\alpha^2} \mathbf{G}^* \end{bmatrix}, \text{ kde } \mathbf{G}^* = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{1}{\sigma_\alpha^2} \mathbf{G}^{-1} \end{bmatrix}.$$

Za $\hat{\boldsymbol{\alpha}}_p$ tedy dosadíme dle (4.18) do (4.17) a tím dojdeme ke vztahu

$$\hat{\mathbf{y}}_p = \mathbf{X}_p \hat{\boldsymbol{\beta}} + \mathbf{Z}_{(p)} \hat{\boldsymbol{\alpha}}, \quad (4.19)$$

kde $\mathbf{Z}_{(p)} = \mathbf{Z}_1 + \mathbf{Z}_2 \mathbf{G}_{po} \mathbf{G}^{-1}$ a $\hat{\boldsymbol{\beta}}$ je BLUE a $\hat{\boldsymbol{\alpha}}$ je BLUP odhaduté z pozorovaných dat, viz (4.5) a (4.7). Rovnice smíšeného modelu (4.12) jsou dle (4.10) ve tvaru

$$\begin{bmatrix} \frac{1}{\sigma_\epsilon^2} \mathbf{X}^T \mathbf{X} & \frac{1}{\sigma_\epsilon^2} \mathbf{X}^T \mathbf{Z} \\ \frac{1}{\sigma_\epsilon^2} \mathbf{Z}^T \mathbf{X} & \frac{1}{\sigma_\epsilon^2} \mathbf{Z}^T \mathbf{Z} + \frac{1}{\sigma_\alpha^2} \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\alpha}} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma_\epsilon^2} \mathbf{X}^T \mathbf{y} \\ \frac{1}{\sigma_\epsilon^2} \mathbf{Z}^T \mathbf{y} \end{bmatrix}.$$

Jestliže označíme matici koeficientů jako \mathbf{Q} , tj.

$$\mathbf{Q} = \begin{bmatrix} \frac{1}{\sigma_\epsilon^2} \mathbf{X}^T \mathbf{X} & \frac{1}{\sigma_\epsilon^2} \mathbf{X}^T \mathbf{Z} \\ \frac{1}{\sigma_\epsilon^2} \mathbf{Z}^T \mathbf{X} & \frac{1}{\sigma_\epsilon^2} \mathbf{Z}^T \mathbf{Z} + \frac{1}{\sigma_\alpha^2} \mathbf{G}^{-1} \end{bmatrix},$$

pak pro odhady $\hat{\boldsymbol{\beta}}$ a $\hat{\boldsymbol{\alpha}}$ platí

$$\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\alpha}} \end{bmatrix} = \mathbf{Q}^{-1} \begin{bmatrix} \frac{1}{\sigma_\epsilon^2} \mathbf{X}^T \\ \frac{1}{\sigma_\epsilon^2} \mathbf{Z}^T \end{bmatrix} \mathbf{y}. \quad (4.20)$$

Dle [7] je vhodné vytvořit interval spolehlivosti pomocí varianční matice rozdílu $\hat{\mathbf{y}}_p - \mathbf{y}_p$, tj. $\text{Var}(\hat{\mathbf{y}}_p - \mathbf{y}_p)$, místo obvyklého využití varianční matice $\hat{\mathbf{y}}_p$, tj. $\text{Var}(\hat{\mathbf{y}}_p)$, tedy jako

$$\hat{\mathbf{y}}_p \pm 1.96 \sqrt{\text{diag}(\text{Var}(\hat{\mathbf{y}}_p - \mathbf{y}_p))}, \quad (4.21)$$

kde hodnota 1.96 odpovídá 0.975-kvantilu normovaného normálního rozdělení. Využitím vztahu (4.19) a $\mathbf{y}_p = \mathbf{X}_p\boldsymbol{\beta} + \mathbf{Z}_{(p)}\boldsymbol{\alpha} + \boldsymbol{\epsilon}_p$ získáme

$$\begin{aligned}\text{Var}(\hat{\mathbf{y}}_p - \mathbf{y}_p) &= \text{Var}\left(\left[\mathbf{X}_p \mid \mathbf{Z}_{(p)}\right] \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\alpha}} \end{bmatrix} - \left[\mathbf{X}_p \mid \mathbf{Z}_{(p)}\right] \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha} \end{bmatrix} - \boldsymbol{\epsilon}_p\right) \\ &= \left[\mathbf{X}_p \mid \mathbf{Z}_{(p)}\right] \text{Var}\left(\begin{bmatrix} \hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \\ \hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} \end{bmatrix}\right) \begin{bmatrix} \mathbf{X}_p^T \\ \mathbf{Z}_{(p)}^T \end{bmatrix} - \sigma_\epsilon^2 \mathbf{I}_{n_p}.\end{aligned}\quad (4.22)$$

Nyní je tedy třeba určit $\text{Var}\left(\begin{bmatrix} \hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \\ \hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} \end{bmatrix}\right)$, přičemž si nejprve odvodíme vztah pro $\begin{bmatrix} \hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \\ \hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} \end{bmatrix}$. Dle (4.20) a (4.12) víme, že

$$\begin{bmatrix} \hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \\ \hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} \end{bmatrix} = \mathbf{Q}^{-1} \begin{bmatrix} \mathbf{X}^T \\ \mathbf{Z}^T \end{bmatrix} (\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\epsilon}) - \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha} \end{bmatrix}.$$

Následnými úpravami dojdeme ke vztahu

$$\begin{bmatrix} \hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \\ \hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} \end{bmatrix} = \mathbf{Q}^{-1} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\frac{1}{\sigma_\epsilon^2} \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha} \end{bmatrix} + \mathbf{Q}^{-1} \begin{bmatrix} \mathbf{X}^T \\ \mathbf{Z}^T \end{bmatrix} \boldsymbol{\epsilon},$$

který využijeme k určení varianční matice, tj.

$$\text{Var}\left(\begin{bmatrix} \hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \\ \hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} \end{bmatrix}\right) = \text{Var}\left(\mathbf{Q}^{-1} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\sigma_\epsilon^2 \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha} \end{bmatrix} + \mathbf{Q}^{-1} \begin{bmatrix} \mathbf{X}^T \\ \mathbf{Z}^T \end{bmatrix} \boldsymbol{\epsilon}\right).$$

Tento vztah lze několika úpravami zjednodušit do tvaru

$$\text{Var}\left(\begin{bmatrix} \hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \\ \hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} \end{bmatrix}\right) = \sigma_\epsilon^2 \mathbf{Q}^{-1} \quad (4.23)$$

Pro interval spolehlivosti potom dosazením (4.22) do (4.21) s využitím (4.23) platí

$$\hat{\mathbf{y}}_p \pm 1.96 \sqrt{\text{diag} \left([\mathbf{X}_p \mid \mathbf{Z}_{(p)}] \mathbf{Q}^{-1} \begin{bmatrix} \mathbf{X}_p^T \\ \mathbf{Z}_{(p)}^T \end{bmatrix} \right)}. \quad (4.24)$$

4.3.2. Predikce založená na podmíněném rozdělení $\mathbf{y}_p \mid \mathbf{y}$

Nyní se na tento problém podíváme z hlediska podmíněného rozdělení. Vyjdeme z rozšířeného smíšeného modelu (4.14), tj. uvažujeme sloupcový vektor $\mathbf{y}_+ = (\mathbf{y}^T, \mathbf{y}_p^T)^T$ s varianční maticí, kterou označíme \mathbf{V}_+ , tj.

$$\text{Var}(\mathbf{y}_+) = \mathbf{V}_+ = \sigma_\alpha^2 \mathbf{Z}_+ \mathbf{G}_+ \mathbf{Z}_+^T + \sigma_\epsilon^2 \mathbf{I}_{n_+},$$

přičemž platí, že ji lze zapsat blokově pomocí matic odpovídajících \mathbf{y} , resp. \mathbf{y}_p , jako

$$\mathbf{V}_+ = \begin{bmatrix} \mathbf{V}_{oo} & \mathbf{V}_{op} \\ \mathbf{V}_{po} & \mathbf{V}_{pp} \end{bmatrix},$$

tj.

$$\mathbf{V}_{oo} = \sigma_\alpha^2 \mathbf{Z} \mathbf{G} \mathbf{Z}^T + \sigma_\epsilon^2 \mathbf{I}_n,$$

$$\mathbf{V}_{op} = \mathbf{V}_{po}^T = \sigma_\alpha^2 \mathbf{Z} \mathbf{G} \mathbf{Z}_1^T + \sigma_\alpha^2 \mathbf{Z} \mathbf{G}_{op} \mathbf{Z}_2^T,$$

$$\mathbf{V}_{pp} = \sigma_\alpha^2 \mathbf{Z}_1 \mathbf{G} \mathbf{Z}_1^T + \sigma_\alpha^2 \mathbf{Z}_2 \mathbf{G}_{po} \mathbf{Z}_1^T + \sigma_\alpha^2 \mathbf{Z}_1 \mathbf{G}_{op} \mathbf{Z}_2^T + \sigma_\epsilon^2 \mathbf{I}_{n_p}.$$

Podmíněné pravděpodobnostní rozdělení je pak dle Věty 4.1 ve tvaru

$$\mathbf{y}_p \mid \mathbf{y} \sim \mathcal{N}(\mathbf{X}_p \boldsymbol{\beta} + \mathbf{V}_{po} \mathbf{V}_{oo}^{-1} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}), \mathbf{V}_{pp} - \mathbf{V}_{po} \mathbf{V}_{oo}^{-1} \mathbf{V}_{op}).$$

Jestliže se zaměříme na čtvercovou chybu predikce, zjistíme, že odpovídá podmíněnému rozptylu $\text{Var}(\mathbf{y}_p | \mathbf{y})$, protože

$$\begin{aligned} \mathbb{E} \left((\hat{\mathbf{y}}_p - \mathbf{y}_p)^2 \right) &= \mathbb{E} \left(\mathbb{E} \left((\hat{\mathbf{y}}_p - \mathbf{y}_p)^2 | \mathbf{y} \right) \right) = \mathbb{E} \left(\text{Var}(\mathbf{y}_p | \mathbf{y}) \right) \\ &= \text{Var}(\mathbf{y}_p | \mathbf{y}), \end{aligned}$$

přičemž jsme využili zákon anglicky nazývaný *Law of Total Expectation*, dle kterého podle [16] obecně platí $\mathbb{E}(\mathbf{X}) = \mathbb{E}(\mathbb{E}(\mathbf{X} | \mathbf{Y}))$, a toho, že pro normální rozdělení platí, že podmíněný rozptyl nezávisí na hodnotě podmínky. Predikční interval pak lze zkonstruovat jako

$$\hat{\mathbf{y}}_p \pm 1.96 \sqrt{\text{Var}(\mathbf{y}_p | \mathbf{y})}, \quad (4.25)$$

kde $\text{Var}(\mathbf{y}_p | \mathbf{y}) = \mathbf{V}_{pp} - \mathbf{V}_{po} \mathbf{V}_{oo}^{-1} \mathbf{V}_{op}$. Všimněme si, že u standardní metodologie pro predikce (viz odstavec 4.3.1) bylo možné sestavit jen interval spolehlivosti, zatímco u přístupu pomocí podmíněného rozdělení $\mathbf{y}_p | \mathbf{y}$ lze sestavit predikční interval, který je vhodné použít pro predikci nových hodnot.

4.4. Jednofázový přístup

Tento přístup, narozdíl od dvoufázového přístupu rozebraném v kapitole 4.3, lze použít s libovolnou volbou reparametrizace a pro vhodně zvolené transformace dostaneme stejné výsledky jako použitím zmíněného dvoufázového přístupu. Tuto metodu nazýváme také přístup rozšířeného smíšeného modelu, protože do modelu rovnou zahrnujeme \mathbf{y}_p , ale s nekonečným rozpty-

lem (tj. nulovou váhou). Uvažujeme tedy model

$$\mathbf{y}_+ = \mathbf{X}_+\boldsymbol{\beta} + \mathbf{Z}_+\boldsymbol{\alpha}_+ + \boldsymbol{\epsilon}_+, \quad (4.26)$$

přičemž

$$\boldsymbol{\alpha}_+ \sim \mathcal{N}(0, \sigma_\alpha^2 \mathbf{G}_+) \quad \text{a} \quad \boldsymbol{\epsilon}_+ \sim \mathcal{N}(0, \sigma_\epsilon^2 \mathbf{R}_+),$$

kde $\mathbf{R}_+ \in \mathcal{R}^{n_+, n_+}$ je diagonální váhová matice, jejíž diagonální prvek je roven 1, jestliže odpovídá pozorovaným datům, tj. \mathbf{y} , a jejíž diagonální prvek je roven nekonečnu pro data predikovaná, tj. pro \mathbf{y}_p . Nekonečná hodnota v matici \mathbf{R}_+ pak odpovídá tomu, že o datech, která chceme odhadovat, nemáme žádné informace. Dle [11] jsou pak odhady pevných a náhodných efektů dány vztahy vyplývajícími z rovnic rozšířeného smíšeného modelu. Výsledné odhady jsou dány vztahy²

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \left(\mathbf{X}_+^T \tilde{\mathbf{V}}_+^{-1} \mathbf{X}_+ \right)^{-1} \mathbf{X}_+^T \tilde{\mathbf{V}}_+^{-1} \mathbf{y}_+, \\ \hat{\boldsymbol{\alpha}} &= \sigma_\alpha^2 \mathbf{G}_+ \mathbf{Z}_+^T \tilde{\mathbf{V}}_+^{-1} \left(\mathbf{y}_+ - \mathbf{X}_+ \hat{\boldsymbol{\beta}} \right), \end{aligned} \quad (4.27)$$

kde $\tilde{\mathbf{V}}_+ = \sigma_\alpha^2 \mathbf{Z}_+ \mathbf{G}_+ \mathbf{Z}_+^T + \sigma_\epsilon^2 \mathbf{R}_+$. Protože jsou splněny předpoklady Věty 1.3, můžeme tuto větu využít k určení inverzní matice $\tilde{\mathbf{V}}_+^{-1}$ k matici $\tilde{\mathbf{V}}$. Konkrétně volíme

$$\begin{aligned} \mathbf{S} &= \mathbf{Z}_+, \\ \mathbf{T} &= \sigma_\alpha^2 \mathbf{G}_+, \\ \mathbf{U} &= \mathbf{Z}_+^T, \\ \mathbf{W} &= \sigma_\epsilon^2 \mathbf{R}_+, \end{aligned}$$

²Symbol $\hat{\cdot}$ značí, že odhady jsou založené na tom, že neznámé hodnoty \mathbf{y}_p mají nekonečný rozptyl.

čímž dostaneme

$$\tilde{\mathbf{V}}_+^{-1} = \frac{1}{\sigma_\epsilon^2} \mathbf{R}_+^{-1} - \frac{1}{\sigma_\alpha^4} \mathbf{R}_+^{-1} \mathbf{Z}_+ \left(\frac{1}{\sigma_\alpha^2} \mathbf{G}_+^{-1} + \mathbf{Z}_+^T \frac{1}{\sigma_\epsilon^2} \mathbf{R}_+^{-1} \mathbf{Z}_+ \right)^{-1} \mathbf{Z}_+^T \mathbf{R}_+^{-1}. \quad (4.28)$$

Dále opět platí $\mathbf{Z}_+ = \mathbf{C}_{k+1}^+(\mathbf{x}_+) \boldsymbol{\Omega}_r^+$, kde $\boldsymbol{\Omega}_r^+$ je ortogonální matice transformace taková, že splňuje

$$\begin{aligned} \mathbf{C}_{k+1}^+(\mathbf{x}_+) [\boldsymbol{\Omega}_f^+ \mid \boldsymbol{\Omega}_r^+] &= [\mathbf{X}_+ \mid \mathbf{Z}_+] \\ \text{a } [\boldsymbol{\Omega}_f^+ \mid \boldsymbol{\Omega}_r^+]^T \mathbf{b}_+ &= [\boldsymbol{\beta}^T \mid \boldsymbol{\alpha}_+^T]^T. \end{aligned}$$

Pro případ penalizací založených na diferencích volíme matici $\boldsymbol{\Omega}_r^+$ na základě spektrálního rozkladu penalizační matice $\mathbf{D}_+^T \mathbf{D}_+$. Tím dosáhneme požadované rovnosti $\mathbf{C}_{k+1}^+(\mathbf{x}_+) \mathbf{b}_+ = \mathbf{X}_+ \boldsymbol{\beta} + \mathbf{Z}_+ \boldsymbol{\alpha}_+$. Dále ze vztahu (4.28) plyne, že pro určení pevných a náhodných efektů nepotřebujeme matici \mathbf{R}_+ , ale matici k ní inverzní, tj. matici \mathbf{R}_+^{-1} . Pro \mathbf{R}_+^{-1} dle [9], str. 81, platí, že je také diagonální a na diagonále má 1 na pozici, na které byla 1 v matici \mathbf{R}_+ , a 0 na pozici, na které bylo nekonečno v matici \mathbf{R}_+ .

Abychom dokázali popsat vztah mezi touto jednofázovou metodou a metodou dvoufázovou popsanou v kapitole 4.3, je třeba znát vztah mezi variační maticí náhodných efektů a rozšířenou varianční maticí. S tím nám pomůže následující věta, jejíž důkaz může čtenář najít v dodatku článku [1].

Věta 4.2. *Uvažujme model (3.2) s penalizací založenou na diferencích B-splajnových koeficientů. Jestliže je matice transformace použitá k získání odhadů (4.27) přímým rozšířením původní matice transformace*

$$\boldsymbol{\Omega}_r^+ = \begin{bmatrix} \boldsymbol{\Omega}_r & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_{pr} \end{bmatrix}, \quad (4.29)$$

kde $\mathbf{\Omega}_r = \mathbf{U}_r \mathbf{\Sigma}^{-\frac{1}{2}}$ a $\mathbf{\Omega}_{pr} = (\mathbf{D}_d^2)^{-1}$, viz odstavec 4.3.1, pak jsou aproximované pozorované hodnoty a predikované hodnoty získané pomocí jednofázového přístupu stejné jako ty, které jsou dané standardní metodologií dvoufázového přístupu, viz odstavec 4.3.1.

Zároveň také platí, že odhady parametrů rozptylů σ_α^2 a σ_ϵ^2 získané metodou REML, viz odstavec 4.1.3, jsou stejné pro základní i rozšířený model, tj. maximalizací následujících funkcí bychom získali stejné odhady parametrů rozptylů σ_α^2 a σ_ϵ^2

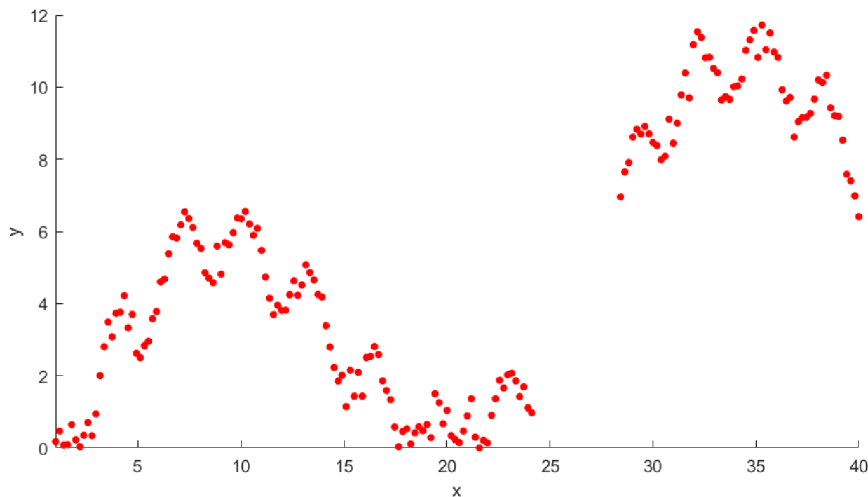
$$l_{REML}(\mathbf{V}) = -\frac{1}{2} \ln(|\mathbf{V}|) - \frac{1}{2} \ln(|\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}|) - \frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}),$$

$$l_{REML}^+(\tilde{\mathbf{V}}_+) = -\frac{1}{2} \ln(|\tilde{\mathbf{V}}_+|) - \frac{1}{2} \ln(|\mathbf{X}^T \tilde{\mathbf{V}}_+^{-1} \mathbf{X}|) - \frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \tilde{\mathbf{V}}_+^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

Stejných výsledků bychom dosáhli i použitím metody maximální věrohodnosti, která má ale výše zmíněné nevýhody.

Důsledkem Věty 4.2 je to, že oba přístupy (jednofázový i dvoufázový) dávají vždy stejné řešení bez ohledu na zvolenou transformaci. Větu 4.2 jsme formulovali pro konkrétní transformaci, abychom obdrželi vztah mezi oběma metodami. Poslední tvrzení ve Větě 4.2 znamená, že odhady parametrů rozptylů použité k predikci jsou stejné jako ty, které byly použity k vytvoření modelu. Jinými slovy, predikce lze nejen provést současně, ale dokonce bude stejný i parametr vyhlazování.

Příklad 4.1. Předpokládejme, že máme dána data generovaná podobně jako v Příkladu 3.3, tj. vektor \mathbf{x} je vytvořen pomocí funkce `linsolve` a vektor \mathbf{y} pomocí funkcí `randn`, `abs` a `sin`. Data jsou zobrazená na obrázku 16, přičemž platí $n = 179$, $a_1 = 1$ a $a_2 = 40$. Naším úkolem je doplnit chybějící pozorování pomocí jednofázového přístupu a poté se budeme věnovat predikcím pro budoucí pozorování.



Obrázek 16: Základní pohled na data

Obecně bychom postupovali následovně. Uvažujme model popsany v [5]

$$E(y_i | x_i) = f(x_i) + \sum_{j=1}^J (g_j(x_i) \cos(j\omega x_i) + h_j(x_i) \sin(j\omega x_i)), \quad (4.30)$$

kde f je hladká funkce popisující trend, g a h jsou hladké funkce zachycující amplitudy kosinových a sinových vln a $\omega = \frac{2\pi}{p}$, kde p je perioda. J je počet uvažovaných funkcí, přičemž se obvykle uvažuje $J = 1$ nebo $J = 2$. Pro volbu $J = 1$, kterou budeme uvažovat v tomto příkladu, lze model (4.30) přepsat do maticového tvaru

$$\mathbf{y} = \check{\mathbf{C}}\check{\mathbf{b}} + \boldsymbol{\epsilon}, \quad (4.31)$$

kde $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma_{\boldsymbol{\epsilon}}^2 \mathbf{I}_n)$, $\check{\mathbf{C}} = [\mathbf{C}_{k+1}(\mathbf{x}) \mid \mathbf{K}\mathbf{C}_{k+1}(\mathbf{x}) \mid \mathbf{S}\mathbf{C}_{k+1}(\mathbf{x})]$ a $\check{\mathbf{b}} = (\mathbf{b}, \mathbf{b}_k, \mathbf{b}_s)$, přičemž \mathbf{K} je diagonální matice, jejíž diagonála je tvořena vektorem $\cos(\omega\mathbf{x})$ a \mathbf{S} je diagonální matice, jejíž diagonála je tvořena vektorem $\sin(\omega\mathbf{x})$, a symbol \otimes značí tenzorový součin těchto matic.

Abychom získali odhad vektoru koeficientů $\check{\mathbf{b}}$, je třeba minimalizovat

funkci

$$\delta(\check{\mathbf{b}}) = (\mathbf{y} - \check{\mathbf{C}}\check{\mathbf{b}})^T (\mathbf{y} - \check{\mathbf{C}}\check{\mathbf{b}}) + \check{\mathbf{b}}^T \check{\mathbf{P}}\check{\mathbf{b}},$$

kde $\check{\mathbf{P}} = \check{\boldsymbol{\rho}} \otimes \check{\mathbf{D}}^T \check{\mathbf{D}}$, přičemž $\check{\boldsymbol{\rho}}$ je diagonální matice, jejíž diagonála je tvořena vektorem (ρ, ρ_k, ρ_s) , a $\check{\mathbf{D}}$ je blokově diagonální matice, jejíž diagonálu tvoří matice $\mathbf{D}_d^T \mathbf{D}_d$ a matice $\mathbf{I}_2 \otimes \mathbf{D}_{d_{ks}}^T \mathbf{D}_{d_{ks}}$. My zvolíme $d = 2$ a $d_{ks} = 1$, neboť je to dle [1] obvyklá volba.

Protože uvažujeme jednofázový přístup, je třeba model (4.31) rozšířit. Definujme tedy matici

$$\check{\mathbf{C}}_+ = [\mathbf{C}_{k+1}^+(\mathbf{x}) \mid \mathbf{K}_+ \mathbf{C}_{k+1}^+(\mathbf{x}_+) \mid \mathbf{S}_+ \mathbf{C}_{k+1}^+(\mathbf{x}_+)], \quad (4.32)$$

kde \mathbf{K}_+ je diagonální matice, jejíž diagonála je tvořena vektorem $\cos(\omega \mathbf{x}_+)$ a \mathbf{S}_+ je diagonální matice, jejíž diagonála je tvořena vektorem $\sin(\omega \mathbf{x}_+)$ a rozšířenou penalizační matici $\check{\mathbf{P}}_+$, která je blokově diagonální s maticemi $\rho (\mathbf{D}_d^+)^T \mathbf{D}_d^+$, $\rho_k (\mathbf{D}_{d_k}^+)^T \mathbf{D}_{d_k}^+$ a $\rho_s (\mathbf{D}_{d_s}^+)^T \mathbf{D}_{d_s}^+$ na diagonále. Poté již můžeme využít vztah (3.12), ve kterém nahradíme matici $\mathbf{C}_{k+1}^+(\mathbf{x}_+)$ maticí $\check{\mathbf{C}}_+$ a matici $\rho \mathbf{P}_+$ maticí $\check{\mathbf{P}}_+$. Pro matice $\check{\mathbf{X}}$ a $\check{\mathbf{Z}}$ platí, že jsou dány jako

$$\check{\mathbf{X}} = [\mathbf{1}_n \mid \mathbf{x} \mid \mathbf{x}^2 \mid \dots \mid \mathbf{x}^{d-1} \mid \cos(\omega \mathbf{x}) \mid \sin(\omega \mathbf{x})], \quad (4.33)$$

$$\check{\mathbf{Z}} = [\mathbf{Z} \mid \mathbf{K}\mathbf{Z}_{ks} \mid \mathbf{S}\mathbf{Z}_{ks}], \quad (4.34)$$

přičemž $\cos(\omega \mathbf{x})$ a $\sin(\omega \mathbf{x})$ jsou sloupcové vektory stejné délky jako vektor \mathbf{x} , pro matici \mathbf{Z} opět platí vztah (4.13) a matici \mathbf{Z}_{ks} lze získat pomocí vztahu

$$\mathbf{Z}_{ks} = \mathbf{C}_{k+1}(\mathbf{x}) \check{\mathbf{\Omega}}_r \quad (4.35)$$

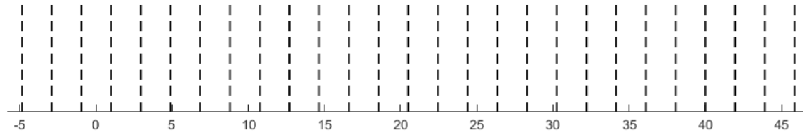
kde $\check{\mathbf{\Omega}}_r = \check{\mathbf{U}}_r \check{\mathbf{\Sigma}}^{-\frac{1}{2}}$ a $\check{\mathbf{U}}_r$ a $\check{\mathbf{\Sigma}}$ jsou opět získány z matic tvořících spektrální rozklad matice $\mathbf{D}_{d_{ks}}^T \mathbf{D}_{d_{ks}}$. Vyhlažující parametry ρ , ρ_k a ρ_s určíme opět jako

poměry parametru rozptylu náhodného členu a parametru rozptylu náhodných efektů σ^2 , σ_k^2 a σ_s^2 . Pro varianční matici $\check{\mathbf{G}}$ platí, že je blokově diagonální s maticemi

$$\begin{aligned}\mathbf{G} &= \sigma^2 \mathbf{I}_{c-d}, \\ \mathbf{G}_k &= \sigma_k^2 \mathbf{I}_{c-1}, \\ \mathbf{G}_s &= \sigma_s^2 \mathbf{I}_{c-1}\end{aligned}\tag{4.36}$$

na diagonále.

Nyní již přejdeme ke konkrétnímu postupu. Uvažujeme stupeň splajnu 3, tj. $k = 3$, a jak již bylo zmíněno výše, zvolíme $d = 2$ a $d_{ks} = 1$. Protože uvažujeme jednofázový přístup, definujeme vektor \mathbf{x}_+ délky 200, který obsahuje složky vektoru \mathbf{x} a složky vektoru \mathbf{x}_p , pro který chceme predikovat neznámé hodnoty \mathbf{y}_p . Dále vytvoříme rozšířenou síť uzlů, která je zobrazena na obrázku 17, a využitím této sítě uzlů, vektoru \mathbf{x}_+ a stupně splajnu 3 vytvoříme



Obrázek 17: Rozšířená síť uzlů pokrývající interval $[1, 40]$

pomocí funkce `spsolve` kolokační matici $\mathbf{C}_{k+1}^+(\mathbf{x}_+) \in \mathcal{R}^{200,23}$. Poté vytvoříme matici \mathbf{D}_d^+ pomocí funkce `diff` aplikované na matici \mathbf{I}_{23} a využitím funkce `svd` určíme spektrální rozklad matice $(\mathbf{D}_d^+)^T \mathbf{D}_d^+$, tj. vytvoříme matice \mathbf{U}_+ a $\tilde{\Sigma}_+$ a tedy i matice \mathbf{U}_r^+ a Σ_+ . Tyto matice poté využijeme k určení matice transformace Ω_r^+ , pomocí které určíme matici \mathbf{Z}_+ .

Dále vytvoříme matici $\check{\mathbf{X}}_+$ dle upraveného vztahu (4.33), přičemž volíme $p = \pi$, tj. $\omega = 2$, a matici $\mathbf{D}_{d_{ks}}^+$, pro kterou budeme chtít opět najít matice, které tvoří spektrální rozklad matice $(\mathbf{D}_{d_{ks}}^+)^T \mathbf{D}_{d_{ks}}^+$. Pomocí nich můžeme opět

vytvořit matici transformace označenou jako $\check{\mathbf{\Omega}}_r^+$ a poté již můžeme určit matici \mathbf{Z}_{ks}^+ dle upraveného vztahu (4.35). Po vytvoření matic \mathbf{K}_+ a \mathbf{S}_+ můžeme určit i matici $\check{\mathbf{Z}}_+$ dle upraveného vztahu (4.34).

Protože již známe matice $\check{\mathbf{X}}_+$ a $\check{\mathbf{Z}}_+$, můžeme k nalezení odhadů parametrů rozptylu pomocí metody REML využít tzv. SOP algoritmus popsany v [15]. K tomu využijeme kód z přílohy článku [1], který je i součástí přílohy diplomové práce, a dojdeme k hodnotám

$$\sigma_\epsilon = 0.0997,$$

$$\sigma^2 = 0.8256,$$

$$\sigma_k^2 = 0.0000,$$

$$\sigma_s^2 = 0.0425,$$

pomocí kterých určíme matici $\check{\boldsymbol{\rho}}$ jako diagonální matici s diagonálními prvky $\rho = \frac{\sigma_\epsilon}{\sigma^2}$, $\rho_k = \frac{\sigma_\epsilon}{\sigma_k^2}$ a $\rho_s = \frac{\sigma_\epsilon}{\sigma_s^2}$. Následně můžeme využitím upravených vztahů (4.36) určit matice \mathbf{G}_+ , \mathbf{G}_k^+ a \mathbf{G}_s^+ a z nich poté vytvořit blokově diagonální matici $\check{\mathbf{G}}_+$. Poté vytvoříme matici $\mathbf{R}_+^{-1} \in \mathcal{R}^{200,200}$ tak, že na diagonále bude mít $\frac{1}{\sigma_\epsilon^2}$ na pozicích odpovídajících složkám vektoru \mathbf{x} a 0 na pozicích odpovídajících složkám vektoru \mathbf{x}_p .

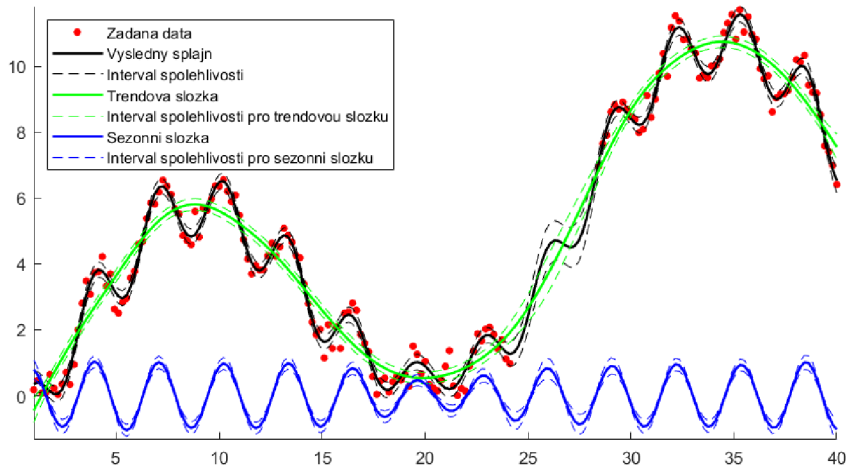
Z rovnic smíšeného modelu (viz (4.10)) určíme odhady pevných a náhodných efektů, které poté využijeme k určení odhadu $\hat{\mathbf{y}}_+$ vektoru \mathbf{y}_+ , konkrétně platí

$$\hat{\mathbf{y}}_+ = [\check{\mathbf{X}}_+ \mid \check{\mathbf{Z}}_+] \begin{bmatrix} \check{\mathbf{X}}_+^T \mathbf{R}_+^{-1} \check{\mathbf{X}}_+ & \check{\mathbf{X}}_+^T \mathbf{R}_+^{-1} \check{\mathbf{Z}}_+ \\ \check{\mathbf{Z}}_+^T \mathbf{R}_+^{-1} \check{\mathbf{X}}_+ & \check{\mathbf{Z}}_+^T \mathbf{R}_+^{-1} \check{\mathbf{Z}}_+ + \check{\mathbf{G}}_+^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \check{\mathbf{X}}_+^T \mathbf{R}_+^{-1} \mathbf{y}_+ \\ \check{\mathbf{Z}}_+^T \mathbf{R}_+^{-1} \mathbf{y}_+ \end{bmatrix}.$$

Poté již můžeme přejít k nalezení intervalu spolehlivosti

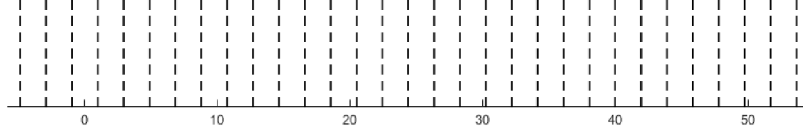
$$\hat{\mathbf{y}}_+ \pm 1.96 \sqrt{\text{diag} \left([\tilde{\mathbf{X}} \mid \tilde{\mathbf{Z}}] \begin{bmatrix} \tilde{\mathbf{X}}_+^T \mathbf{R}_+^{-1} \tilde{\mathbf{X}}_+ & \tilde{\mathbf{X}}_+^T \mathbf{R}_+^{-1} \tilde{\mathbf{Z}}_+ \\ \tilde{\mathbf{Z}}_+^T \mathbf{R}_+^{-1} \tilde{\mathbf{X}}_+ & \tilde{\mathbf{Z}}_+^T \mathbf{R}_+^{-1} \tilde{\mathbf{Z}}_+ + \mathbf{G}_+^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \tilde{\mathbf{X}}_+^T \\ \tilde{\mathbf{Z}}_+^T \end{bmatrix} \right)},$$

který byl vytvořen podobně jako interval spolehlivosti (4.24). Analogicky lze určit i křivky pro samotnou trendovou a sezónní složku a jejich intervaly spolehlivosti. Výsledné křivky včetně zadaných bodů jsou zobrazeny na obrázku 18.



Obrázek 18: Výsledný graf

Nyní bude naším cílem predikovat hodnoty pro budoucí pozorování, konkrétně hodnoty odpovídající 40 novým hodnotám na ose x ležících napravo od intervalu $[1, 40]$, konkrétně v intervalu $[40, 48]$. Tím vzniká nový rozšířený vektor, který označíme jako \mathbf{x}_{ext} délky 240, pomocí kterého vytvoříme rozšířenou kolokační matici $\mathbf{C}_4^{\text{ext}}(\mathbf{x}_{\text{ext}}) \in \mathcal{R}^{240,27}$, přičemž opět uvažujeme $k = 3$ a síť uzlů pokrývající celý interval $[1, 47.8392]$. Přidáním 6 pomocných uzlů vytvoříme rozšířenou síť uzlů zobrazenou na obrázku 19.



Obrázek 19: Rozšířená síť uzlů pokrývající interval $[1, 47.8392]$

Po vytvoření rozšířené kolokační matice $\mathbf{C}_4^{\text{ext}}(\mathbf{x}_{\text{ext}})$ lze ověřit, že první blok matice odpovídá původní kolokační matici $\mathbf{C}_4^+(\mathbf{x}_+)$. Následně můžeme vytvořit váhovou matici $\mathbf{M}_{\text{ext}} \in \mathcal{R}^{240,240}$, která je diagonální a na diagonále obsahuje 1 na pozicích odpovídajících známým hodnotám \mathbf{y} a 0 na pozicích odpovídajících neznámým hodnotám \mathbf{y}_p . Abychom mohli k určení matice $\check{\mathbf{C}}_{\text{ext}} \in \mathcal{R}^{240,81}$ využít upravený vztah (4.32), je třeba ještě určit matice $\mathbf{K}_{\text{ext}} \in \mathcal{R}^{240,240}$ a $\mathbf{S}_{\text{ext}} \in \mathcal{R}^{240,240}$. Ty budou vytvořeny jako diagonální matice s vektory $\cos(\omega \mathbf{x}_{\text{ext}})$, resp. $\sin(\omega \mathbf{x}_{\text{ext}})$, na diagonále. Vytvoříme matice $\mathbf{D}_1 \in \mathcal{R}^{26,27}$ a $\mathbf{D}_2 \in \mathcal{R}^{25,27}$ a poté využijeme vztah (4.29) a vytvoříme matici transformace pro trendovou složku

$$\mathbf{\Omega}_r^{\text{ext}} = \begin{bmatrix} \mathbf{\Omega}_r^+ & \mathbf{0} \\ \mathbf{0} & (\mathbf{D}_2^2)^{-1} \end{bmatrix},$$

kterou využijeme k určení matice $\mathbf{Z}_{\text{ext}} \in \mathcal{R}^{240,25}$ dle vztahu

$$\mathbf{Z}_{\text{ext}} = \mathbf{C}_4^{\text{ext}}(\mathbf{x}_{\text{ext}}) \mathbf{\Omega}_r^{\text{ext}}.$$

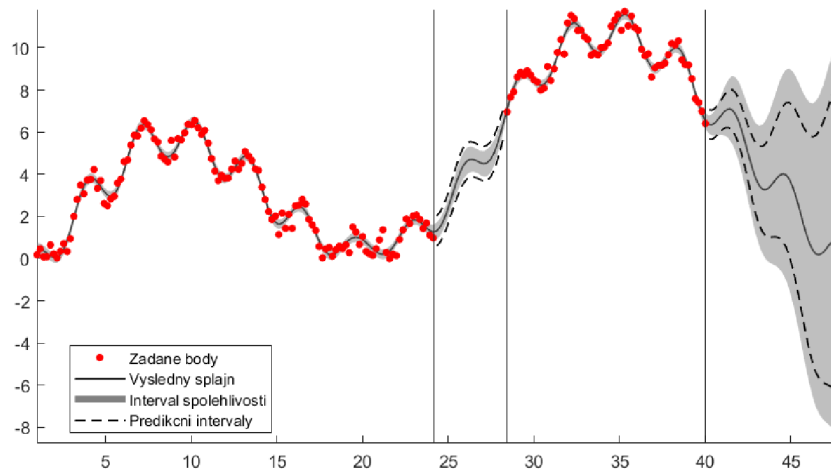
Analogicky bychom vytvořili i matici transformace pro sezónní složku, kterou bychom využili pro určení matice $\mathbf{Z}_{ks}^{\text{ext}}$. Poté již můžeme přejít k určení matice $\check{\mathbf{Z}}_{\text{ext}} \in \mathcal{R}^{240,77}$ dle vztahu $\check{\mathbf{Z}}_{\text{ext}} = [\mathbf{Z}_{\text{ext}} \mid \mathbf{K}_{\text{ext}} \mathbf{Z}_{ks}^+ \mid \mathbf{S}_{\text{ext}} \mathbf{Z}_{ks}^{\text{ext}}]$. Protože uvažujeme $d = 2$, využijeme nyní k určení matice $\check{\mathbf{X}}_{\text{ext}} \in \mathcal{R}^{240,4}$ vztah

$$\check{\mathbf{X}}_{\text{ext}} = [\mathbf{1}_n \mid \mathbf{x}_{\text{ext}} \mid \cos(\omega \mathbf{x}_{\text{ext}}) \mid \sin(\omega \mathbf{x}_{\text{ext}})].$$

Dále vytvoříme blokovou matici $\check{\mathbf{G}}_{\text{ext}} \in \mathcal{R}^{77,77}$, jejíž diagonála je tvořena maticemi $\mathbf{G}_{\text{ext}} \in \mathcal{R}^{25,25}$, $\mathbf{G}_k^{\text{ext}} \in \mathcal{R}^{26,26}$ a $\mathbf{G}_s^{\text{ext}} \in \mathcal{R}^{26,26}$, které vytvoříme na základě vztahu $\mathbf{G}^{-1} = \mathbf{\Omega}_r^T \mathbf{P} \mathbf{\Omega}_r$ platného dle [1]. Dále využijeme váhovou matici \mathbf{M}_{ext} a parametr rozptylu náhodné složky σ_ϵ^2 a vytvoříme matici \mathbf{R}_{ext} a následně i její inverzi $\mathbf{R}_{\text{ext}}^{-1}$.

Nakonec můžeme určit intervaly spolehlivosti a predikční intervaly. Intervaly spolehlivosti určíme podobně jako v předchozí části příkladu, tj. pomocí vztahu (4.24). K určení predikčních intervalů dle vztahu (4.25) je třeba ještě určit varianční matici $\mathbf{V}_{\text{ext}} = \check{\mathbf{Z}}_{\text{ext}} \check{\mathbf{G}}_{\text{ext}} \check{\mathbf{Z}}_{\text{ext}}^T + \sigma_\epsilon^2 \mathbf{I}_{n_{\text{ext}}}$ a jednotlivé bloky, ze kterých je tato matice složena složena.

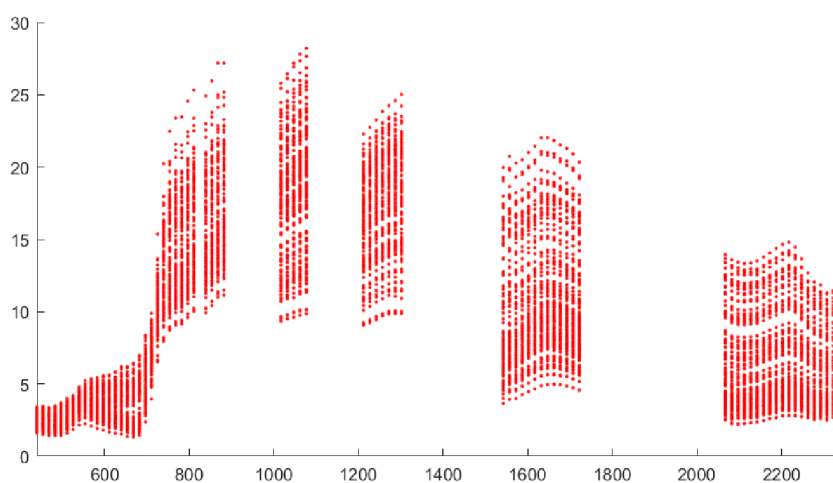
Výsledný splajn, interval spolehlivosti a predikční intervaly jsou pak zobrazeny na obrázku 20, přičemž interval spolehlivosti je nyní znázorněn plochou šedé barvy.



Obrázek 20: Výsledný splajn s intervaly spolehlivosti a predikčními intervaly

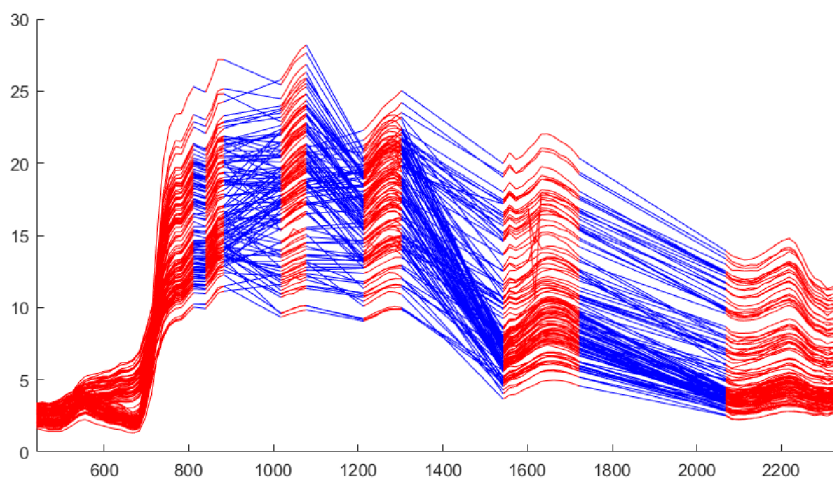
5. Praktická část

V této části přejdeme k praktickému využití výše probrané teorie. V práci budou použita data, která jsou jednorozměrná a mají následující charakter. Data popisují závislost míry reflektance (osa y) na vlnové délce světla (osa x) pro 120 stromů. Na ose x jsou dány hodnoty v nanometrech, konkrétně v mezích od 440.7 do 2337.5, přičemž nejsou dány rovnoměrně, ale data z některých intervalů chybí, což lze pozorovat v grafu na obrázku 21. K těmto hodnotám na ose x máme přiřazeny hodnoty na ose y , a to pro každý ze 120 stromů. Tomu odpovídá 120 křivek zobrazených v grafu 22, které vznikly spojením jednotlivých bodů pomocí úseček. Znamená to, že jsme pro doplnění chybějících částí využili lineární aproximaci, tj. body jsme jednoduše spojili úsečkami, které jsou znázorněny modře.



Obrázek 21: Základní pohled na data

Při bližším pohledu na data si lze všimnout, že průběh funkcí reprezentovaných úsečkami na obrázku 22 není pro všechny objekty stejný. Pro lepší přehlednost při následném vykreslování dat spolu se splajny, které data apro-

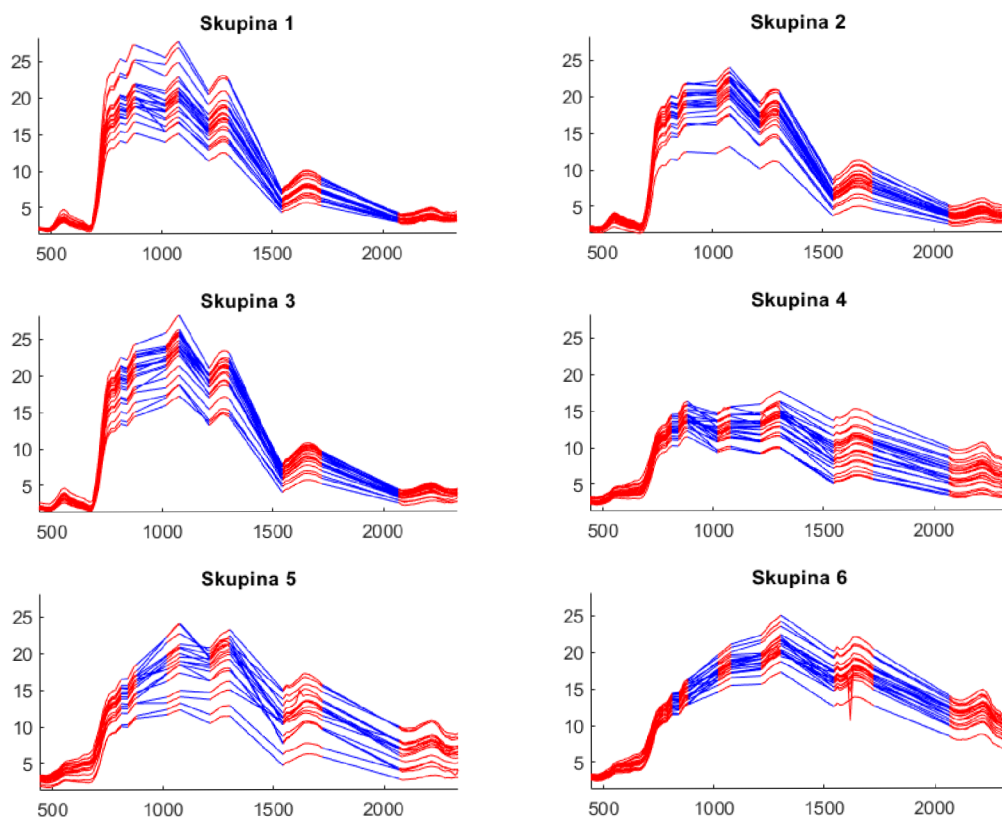


Obrázek 22: Lineární aproximace pro chybějící části

ximují, jsme se rozhodli data intuitivně rozdělit do šesti skupin podle vývoje míry reflektance v závislosti na vlnové délce světla, viz obrázek 23. Toto rozdělení není založeno na přesných matematických kritériích, ale spíše na subjektivním posouzení průběhu funkcí. Poznamenejme, že dělení by mohlo být přesnější a mohlo by být vytvořeno více skupin, ale pro naše účely je toto dělení postačující. Takový pohled na data nám také ukazuje, že u několika objektů (konkrétně například jedno dobře viditelné ve skupině 6) se vyskytují odlehlá pozorování, na které by bylo vhodné se zaměřit, nebo například u penalizované metody nejmenších čtverců přidat jednotlivým pozorováním váhy a tím snížit vliv těchto odlehlých hodnot na výsledný splajn.

5.1. Nalezení splajnu, který data aproximuje ve smyslu penalizované metody nejmenších čtverců

Předpokládejme tedy, že jsou dány sloupcové vektory \mathbf{x} a $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{120}$ délky 75, tj. $n = 75$, a interval $[a_1, a_2] = [440.7, 2337.5]$. Nejprve budeme chtít

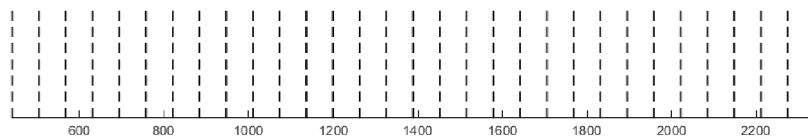


Obrázek 23: Rozdělení dat do skupin

tato data aproximovat splajnem ve smyslu penalizované metody nejmenších čtverců. Zvolíme stupeň splajnu 3, tj. $k = 3$, a vytvoříme síť uzlů, která rovnoměrně pokrývá rozsah hodnot z vektoru \mathbf{x} . Konkrétně uvažujeme uzly, které nabývají hodnot od a_1 do a_2 s krokem $\frac{a_2 - a_1}{30}$, čímž získáme síť uzlů

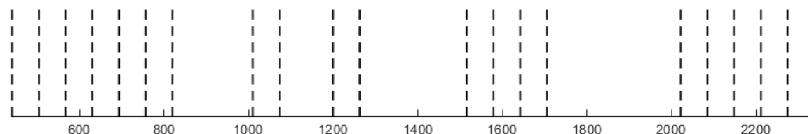
$$[440.7, 503.93, 567.2, \dots, 2211.0, 2274.3, 2337.5], \quad (5.1)$$

která je zobrazena na obázku 24.



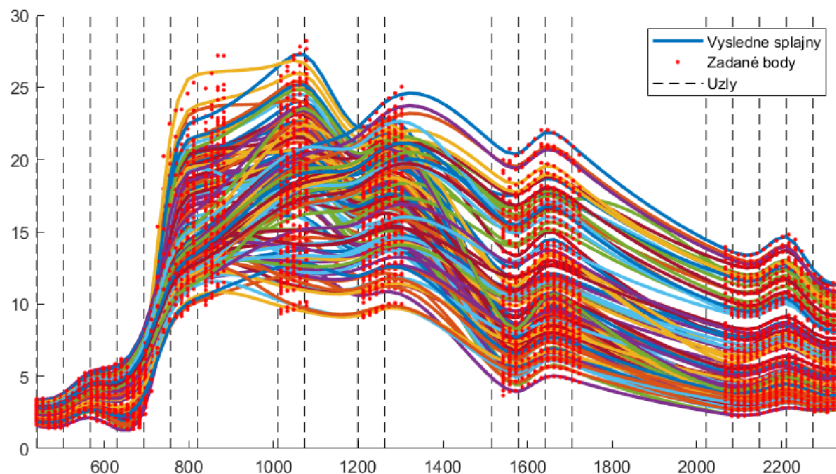
Obrázek 24: Síť uzlů pokrývající celý interval $[a_1, a_2]$

Aby byla kolokační matice plně sloupcové hodnosti, musí v nosiči každého B -splajnu ležet alespoň jeden bod x_i pro $i = 1, \dots, 75$. Je tedy třeba ze sítě uzlů (5.1) odebrat uzly, které tuto podmínku porušují. Tím získáme síť 21 uzlů zobrazenou na obrázku 25, ke které je třeba přidat 6 pomocných uzlů. Tím dostaneme celkem 27 uzlů, a tedy platí $c = 23$. Poté již můžeme pomocí funkce `spcol` vytvořit kolokační matici $\mathbf{C}_4(\mathbf{x})$ a ověřit, že je plně sloupcové hodnosti.



Obrázek 25: Síť uzlů pokrývající hodnoty vektoru \mathbf{x}

Dále zvolíme $\rho = 0.01$ a $d = 2$ a pomocí funkce `diff` aplikované na jednotkovou matici \mathbf{I}_{23} vytvoříme matici \mathbf{D}_2 . Nyní již můžeme přejít k určení odhadů vektorů B -splajnových koeficientů pro každý objekt. Jednotlivé odhady získáme vyřešením soustav $(\mathbf{C}_4^T(\mathbf{x})\mathbf{C}_4(\mathbf{x}) + \rho\mathbf{D}_2^T\mathbf{D}_2)\hat{\mathbf{b}} = \mathbf{C}_4^T(\mathbf{x})\mathbf{y}_i$ pro $i = 1, \dots, 120$ za použití funkce `linsolve`. Odhady B -splajnových koeficientů nyní využijeme pro vytvoření splajnu pomocí funkce `spmak` a jejich vykreslení pomocí funkce `fnplt`. Výsledné splajny pro jednotlivé objekty jsou zobrazeny na obrázku 26. Je zřejmé, že doplnění chybějících hodnot hodnotami danými výslednými splajny by nebylo vhodné. To lze pozorovat především na intervalech, kde je velké množství chybějících hodnot, tedy především na intervalech $[1317.5, 1527.5]$ a $[1737.5, 2052.5]$.



Obrázek 26: Aproximace dat ve smyslu penalizované metody nejmenších čtverců

5.2. Doplnění chybějících hodnot pomocí přístupu chybějící hodnoty

Nyní se chybějící data pokusíme doplnit využitím přístupu chybějící hodnoty popsaného v kapitole 3.1. Začneme tím, že vytvoříme vektor \mathbf{x}_p , který obsahuje hodnoty na ose x , pro které chceme predikovat hodnoty \mathbf{y}_p . Následně budeme uvažovat vektor \mathbf{x}_+ , který obsahuje vzestupně seřazené hodnoty vektoru \mathbf{x} a vektoru \mathbf{x}_p . Síť uzlů nyní volíme tak, aby pokrývala celý rozsah hodnot vektoru \mathbf{x}_+ , tedy uvažujeme síť uzlů (5.1), kterou rozšíříme o 6 pomocných uzlů, tedy platí $c = 33$. Poté již můžeme využitím této sítě uzlů a vektoru \mathbf{x}_+ vytvořit rozšířenou kolokační matici $\mathbf{C}_4^+(\mathbf{x})$, o které je vhodné opět ověřit, že je plně sloupcové hodnosti. Aplikováním funkce `diff` na matici \mathbf{I}_{33} můžeme vytvořit matici \mathbf{D}_2^+ .

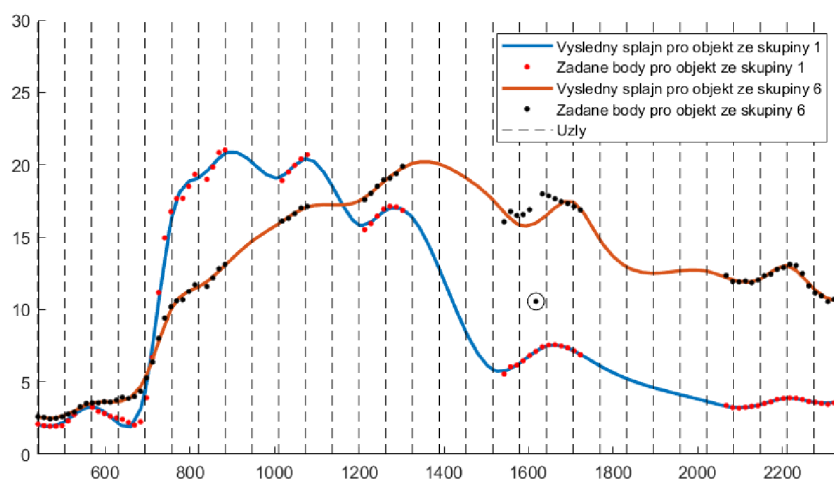
Aby bylo možné pro odhad vektoru B -splajnových koeficientů využít vztah (3.12), je třeba ještě určit váhovou matici \mathbf{M} , která je diagonální a na pozicích odpovídajících známým hodnotám, tj. hodnotám vektoru \mathbf{x} ,

má 1 a na pozicích odpovídajících neznámým hodnotám, tj. hodnotám vektoru \mathbf{x}_p , má 0. Ve vztahu (3.12) se ještě vyskytuje vektor \mathbf{y}_+ , který vytvoříme tak, že bude obsahovat hodnoty na ose y odpovídající hodnotám vektoru \mathbf{x}_+ , přičemž místo neznámých hodnot, které chceme predikovat, bude obsahovat nuly. Pro i -tý prvek $y_+(i)$ vektoru \mathbf{y}_+ , $i = 1, \dots, 129$, tedy platí

$$y_+(i) = \begin{cases} y_j & \text{jestliže existuje } j \in \{1, \dots, 75\} \text{ takové, že } x_{+i} = x_j, \\ 0 & \text{jinak,} \end{cases}$$

kde pro $j = 1, \dots, 75$ je x_j j -tý prvek vektoru \mathbf{x} a y_j je j -tý prvek vektoru \mathbf{y} .

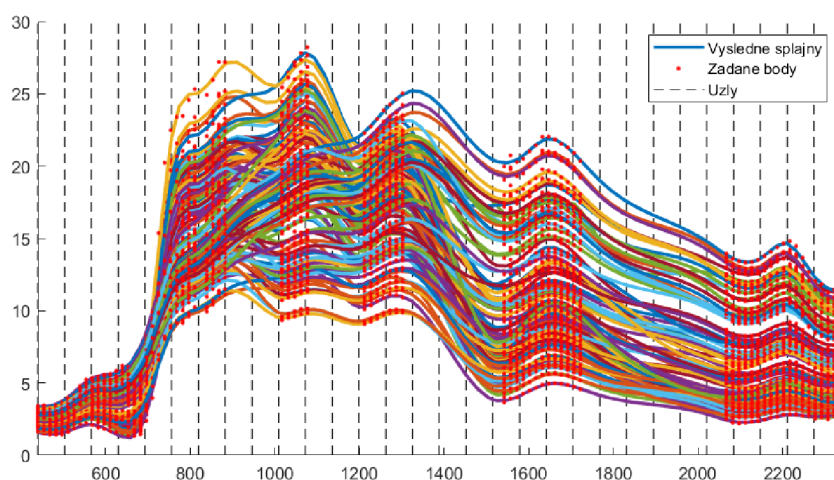
Pak již lze využít vztah (3.12) pro odhad vektoru B -splajnových koeficientů a následně dle vztahu (3.9) určit predikované hodnoty $\hat{\mathbf{y}}_+$. Nejprve se podíváme, jak vypadají výsledné splajny pro dva výrazně odlišné stromy, jeden strom ze skupiny 1 a jeden strom ze skupiny 6. Na obrázku 27 lze poz-



Obrázek 27: Aproximace splajny dle přístupu chybějící hodnoty pro dva vybrané stromy

rovat, že výsledný splajn pro objekt ze skupiny 1 poměrně dobře aproximuje

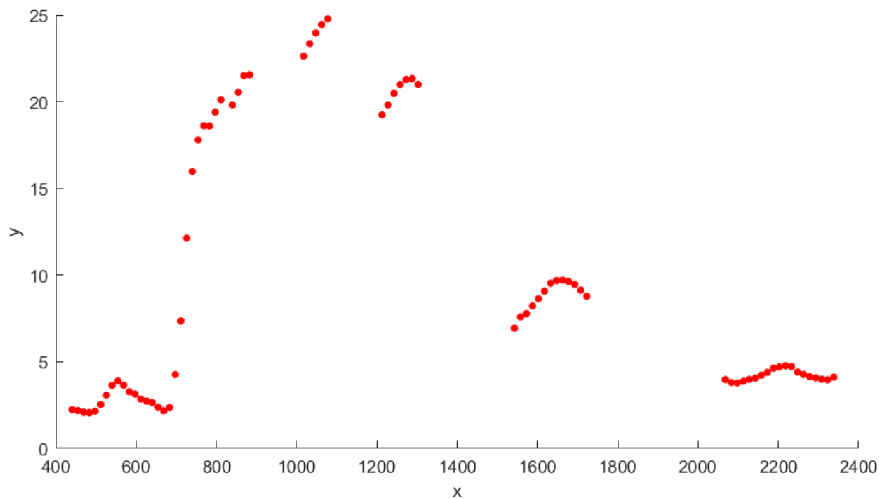
daná data, ale výsledný splajn pro objekt ze skupiny 6 je výrazně ovlivněn odlehlou hodnotou, která je v grafu zakroužkována, proto by bylo vhodné jednotlivým pozorováním přidat váhy a snížit tím vliv odlehlého pozorování. Nakonec se podíváme na výsledné splajny pro všechny objekty, které jsou zobrazeny na obrázku 28.



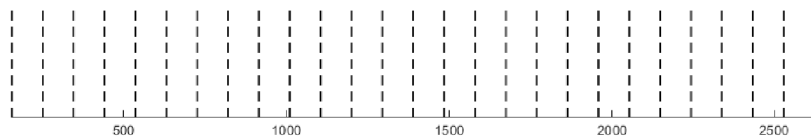
Obrázek 28: Aproximace splajny dle přístupu chybějící hodnoty

5.3. Doplnění chybějících hodnot pomocí hladkých modelů smíšených efektů

Nyní budeme chtít chybějící hodnoty doplnit pomocí hladkých modelů smíšených efektů, konkrétně pomocí jednofázového přístupu, viz část 4.4. Pro názornost uvažujme nejprve data pro první objekt, která jsou zobrazena na obrázku 29. Jsou tedy dány vektory \mathbf{x} a \mathbf{y} délky 75. Uvažujme opět stupeň splajnu 3, tj. $k = 3$, a řád difference v penalizačním členu 2, tj. $d = 2$. Poté postupujme podobně jako v Příkladu 4.1, přičemž uvažujme rozšířenou síť uzlů s 27 uzly z obrázku 30. Pomocí této sítě uzlů a rozšířeného vektoru \mathbf{x}_+



Obrázek 29: Zadaná data pro první objekt



Obrázek 30: Rozšířená síť uzlů

délky 129 vytvoříme kolokační matici $\mathbf{C}_4^+(\mathbf{x}_+) \in \mathcal{R}^{129,23}$. Využitím funkce `diff` vytvoříme matici $\mathbf{D}_2 \in \mathcal{R}^{21,23}$ a najdeme matice $\tilde{\Sigma}$ a \mathbf{U}_r , které tvoří spektrální rozklad matice $\mathbf{D}_2^T \mathbf{D}_2$. Poté určíme matice $\tilde{\Sigma}$ a \mathbf{U} , které využijeme k vytvoření matice transformace $\mathbf{\Omega}_r$. Následně pomocí vztahu (4.13) určíme matici \mathbf{Z} a pomocí vztahu (4.33) matici $\tilde{\mathbf{X}}$, přičemž volíme $p = 200$.

Dále určíme matici $\mathbf{D}_{d_{ks}}$, kde $d_{ks} = 1$, a využitím matic tvořících spektrální rozklad matice $\mathbf{D}_{d_{ks}}^T \mathbf{D}_{d_{ks}}$ vytvořený za pomoci funkce `svd` určíme matici transformace $\tilde{\mathbf{\Omega}}_r$. Poté využijeme vztah (4.34) pro určení matice $\tilde{\mathbf{Z}}$. Pro určení parametrů rozptylů σ_ϵ , σ^2 , σ_k^2 a σ_s^2 metodou REML využijeme tzv. SOP algoritmus blíže popsany v [15], přičemž kód, který je převzatý z přílohy článku [1], lze najít v příloze diplomové práce. Pomocí algoritmu získáme

následující hodnoty

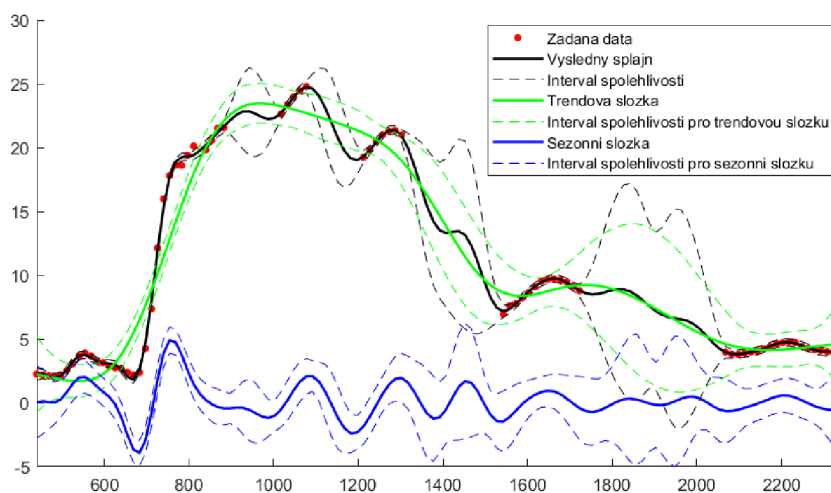
$$\sigma_{\epsilon}^2 = 0.0700,$$

$$\sigma^2 = 18.7797,$$

$$\sigma_k^2 = 3.3348,$$

$$\sigma_s^2 = 16.9877.$$

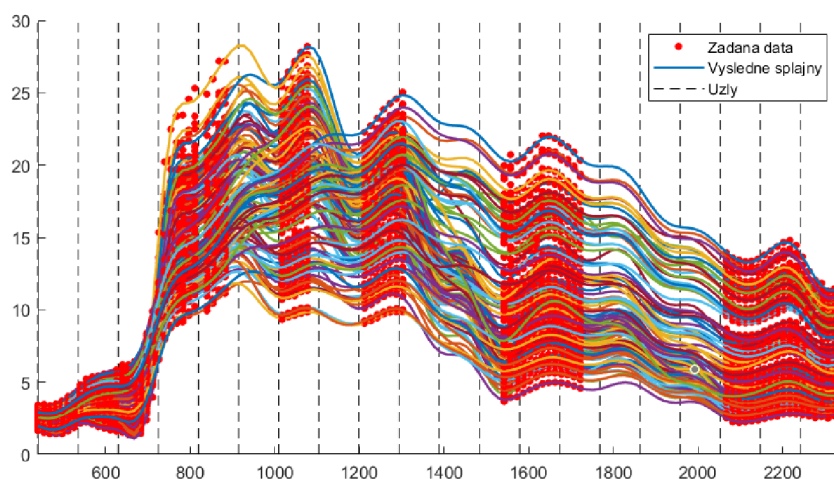
Další postup je analogií k postupu popsánému v Příkladu 4.1, proto rovnou přejdeme k výslednému grafu zobrazenému na obrázku 31. Lze pozorovat,



Obrázek 31: Výsledný splajn, trendová a sezónní složka a intervaly spolehlivosti pro první objekt

vat, že výsledný splajn lze opravdu rozložit na trendovou složku, která se blíží splajnu, který data aproximuje ve smyslu penalizované metody nejmenších čtverců s poměrně velkou hodnotou vyhlazujícího parametru, a na sezónní složku. Protože pracujeme s daty, u kterých není přímo viditelná perioda, jako tomu bylo například v Příkladu 4.1, je vhodné vyzkoušet různé volby periody (tj. parametru p).

Tento graf lze vytvořit pro každý ze 120 objektů, ale zobrazení v jednom výsledném grafu by nebylo přehledné. Podíváme se proto pouze na výsledné splajny pro každý objekt, které jsou pro nás podstatné, viz obrázek 32.



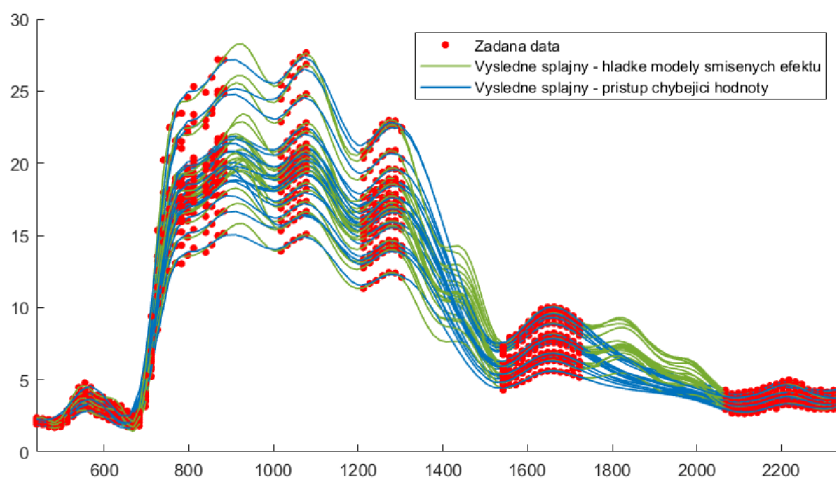
Obrázek 32: Výsledné splajny pro všechny objekty

5.4. Srovnání obou přístupů

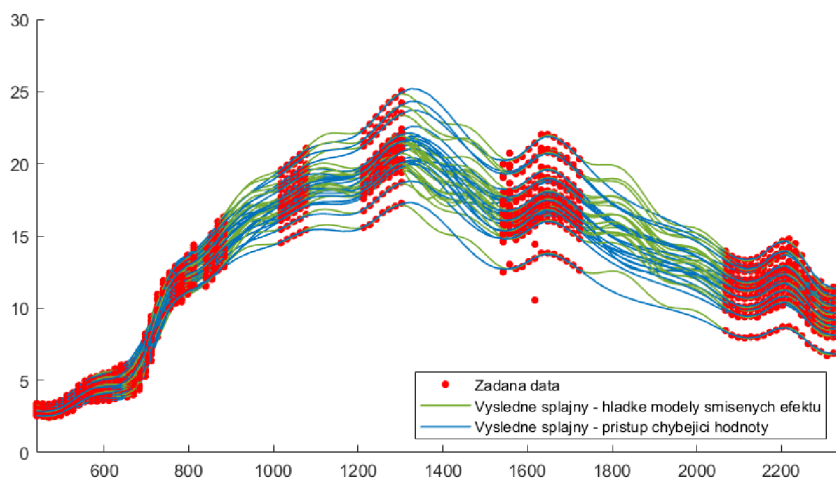
Lze pozorovat, že v porovnání s obrázkem 28, kde jsme chybějící data doplnili pomocí přístupu chybějící hodnoty, jsou výsledné splajny zobrazené na obrázku 32 odlišné především v částech chybějících dat, což je způsobeno přidáním uvažované sezónní složky.

Podívejme se ale nyní na srovnání výsledných splajnů získaných pomocí přístupu chybějící hodnoty a pomocí hladkých modelů smíšených efektů pro objekty z jednotlivých skupin.

Zaměříme se na 2 skupiny nejvíce odlišných objektů, tedy na skupinu 1 a skupinu 6. Pro skupinu 1 jsou zadaná data a výsledné splajny zobrazeny na obrázku 33 a obrázek 34 zobrazuje zadaná data a výsledné splajny pro skupinu 6.



Obrázek 33: Výsledné splajny pro objekty ze skupiny 1



Obrázek 34: Výsledný splajny pro objekty ze skupiny 6

Lze pozorovat, že volbou vyhlazujícího parametru $\rho = 0.01$ jsme získali splajny, které jsou v částech, kde se vyskytují zadaná data, velmi podobné těm získaným pomocí hladkých modelů smíšených efektů. Nejvýraznější odlišnosti mezi splajny získaných těmito dvěma přístupy jsou stále v intervalech

chybějících dat, což lze pozorovat u obou dvou skupin objektů, přičemž největší rozdíly jsou konkrétně na intervalech [1317.5, 1527.5] a [1737.5, 2052.5].

Závěr

V diplomové práci jsem se zaměřila na aplikaci splajnů v klasické metodě nejmenších čtverců a také jejich využití v penalizované metodě nejmenších čtverců pro účely predikce a doplnění chybějících dat. Práce poskytuje přehled teoretických základů potřebných pro pochopení a použití splajnů ve zmíněných metodách, včetně detailnějšího popisu B -splajnů a následně i P -splajnů. Dále jsem se věnovala nastudování algoritmů pro predikci a doplnění chybějících dat, což jsou v dnešní době, kdy data hrají klíčovou roli ve velkém množství různých odvětví, oblasti s bohatým praktickým využitím.

Díky praktické části práce, kde byly popsány metody aplikovány na konkrétní data, jsem se pokusila ukázat, jak mohou splajny efektivně řešit problémy spojené s doplňováním chybějících dat. Výsledky praktické části ukázaly, že splajny představují flexibilní nástroj, který je schopen se přizpůsobit různým strukturám dat a poskytovat predikce i v případech, kdy data nejsou úplná. V příkladech bylo také ilustrováno, jak pomocí splajnů predikovat hodnoty pro budoucí i minulé pozorování.

Použití matematického softwaru Matlab se ukázalo jako vhodný způsob implementace algoritmů. Kódy, které byly v rámci této práce vytvořeny, mohou sloužit jako základ pro další výzkum nebo jako užitečný nástroj pro praktické aplikace v různých oblastech.

Tato práce přináší další pohled na využití splajnů v analýze a zpracování dat a rozšiřuje základní poznatky o penalizované metodě nejmenších čtverců. Během práce jsem narazila na otázky, na které se lze zaměřit detailněji, například jak optimálně volit uzly a jejich počet, jak určit optimální hodnotu vyhlazujícího parametru v penalizované metodě nejmenších čtverců nebo jaké lze využít algoritmy pro určení odhadů parametrů rozptýlů náhodné složky a náhodných efektů v hladkých modelech smíšených efektů pomocí metody

REML. V budoucnu by bylo také možné se zaměřit na otázky, které vedou k rozšíření teorie uvedené v této práci. Jednou z nich je například možnost rozšíření této aplikace splajnů i na vícedimenzionální data.

Závěrem lze říci, že penalizovaná metoda nejmenších čtverců i splajny obecně mají široké spektrum uplatnění a právě některé z nich byly popsány v této práci. Domnívám se, že cíle diplomové práce stanovené v úvodu byly splněny a teoretická část i kódy obsažené v příloze diplomové práce mohou ostatním sloužit k pochopení a aplikování popsaných metod na vlastních datech.

Literatura

- [1] CARBALLO, A., DURBAN, M. a KAUERMANN, G., LEE, D.-J.: *A general framework for prediction in penalized regression*. *Statistical Modelling*, no. 21(4), 293–312, 2021. Dostupné z: <https://doi.org/10.1177/1471082X19896867>. [cit. 2024-04-17].
- [2] COLLINS, D.: *The performance of estimation methods for generalized linear mixed models*. University of Wollongong Thesis Collection, 2009. Dostupné z: <https://ro.uow.edu.au/theses/1737>. [cit. 2024-04-17].
- [3] DE BOOR, C., *A practical guide to splines*, volume 27. Springer, New York, 1978.
- [4] DIERCKX, P.: *Curve and Surface Fitting with splines*. Oxford University Press, New York, 1993.
- [5] EILERS, P., GAMPE, J., MARX, B., RAU, R.: *Modulation models for seasonal time series and incidence tables*. *Statistics in Medicine*, 27, 3430–3441, 2008. Dostupné z: <https://doi.org/10.1002/sim.3188>. [cit. 2024-04-17].
- [6] EILERS, P., MARX, B.: *Practical Smoothing: The Joys of P-splines*. Cambridge: Cambridge University Press, 2021.
- [7] GILMOUR, A., CULLIS, B., WELHAM, S., GOGEL, B., THOMPSON, R.: *An efficient computing strategy for prediction in mixed linear models*. *Computational Statistics and Data Analysis*, 44, 571–586, 2004. Dostupné z: [https://doi.org/10.1016/S0167-9473\(02\)00258-X](https://doi.org/10.1016/S0167-9473(02)00258-X). [cit. 2024-04-17].
- [8] GILMOUR, A., THOMPSON, R., CULLIS, B.: *Average Information REML: An Efficient Algorithm for Variance Parameter Estimation in Linear Mixed Models*. *Biometrics*, 51, 1440–1450, 1995. Dostupné z: <https://doi.org/10.2307/2533274>. [cit. 2024-04-17].
- [9] HARVILLE, D. A.: *Matrix Algebra From a Statistician's Perspective*. Springer, New York, 1997.
- [10] HARVILLE, D. A.: *Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems*. *Journal of the American Statistical Association*, 72(358), 320–338, 1977. Dostupné z: <https://doi.org/10.2307/2286796>. [cit. 2024-04-17].

- [11] HENDERSON, C. R.: *Best Linear Unbiased Estimation and Prediction under a Selection Model*. Biometrics, 31, 423–447, 1975. Dostupné z: <https://doi.org/10.2307/2529430>. [cit. 2024-04-17].
- [12] HOLT, W., NGUYEN, D.: *Essential Aspects of Bayesian Data Imputation*. SSRN Electronic Journal. 2023. Dostupné z: <https://doi.org/10.2139/ssrn.4494314>. [cit. 2024-04-17].
- [13] HRON, K., KUNDEROVÁ, P., VENCÁLEK, O.: *Základy počtu pravděpodobnosti a metod matematické statistiky*. 3. přepracované vydání. Skripta. Olomouc: Univerzita Palackého v Olomouci, 2018.
- [14] MACHALOVÁ, J.: *Optimal interpolating and optimal smoothing spline*. Journal of Electrical Engineering, volume 53, 79 - 82, 2002. Dostupné z: https://www.researchgate.net/publication/268865944_Optimal_interpolating_and_optimal_smoothing_spline. [cit. 2024-04-17].
- [15] RODRÍGUEZ-ÁLVAREZ, M., DURBAN, M., DAE-JIN, L., EILERS, P.: *On the estimation of variance parameters in non-standard generalised linear mixed models: Application to penalised smoothing*. Statistics and Computing, 29, 483-500, 2019.
- [16] SOCH, J.: *Proof: Law of total expectation*. The Book of Statistical Proofs, Proof 310, 2021. Dostupné z: <https://statproofbook.github.io/P/mean-tot.html>, <https://doi.org/10.5281/ZENODO.4305949> [cit. 2024-04-17].

Seznam kódů v příloze

`hladke_modely_smis_ef_prvni_objekt.m` – doplnění chybějících pozorování pro první objekt pomocí hladkých modelů smíšených efektů (viz odstavec 5.3)

`mnc.m` – funkce pro metodu nejmenších čtverců

`penalizovana_mnc.m` – funkce pro penalizovanou metodu nejmenších čtverců

`prakticka_cast.m` – kód k praktické části (viz kapitola 5)

`predikce_uvnitr.m` – funkce pro přístup chybějící hodnoty pro predikci chybějících pozorování uvnitř intervalu

`predikce_vpřed.m` – funkce pro přístup chybějící hodnoty pro predikci budoucích pozorování (predikci vpřed)

`predikce_vzad.m` – funkce pro přístup chybějící hodnoty pro predikci minulých pozorování (predikci vzad)

`priklad1_1.m` – kód k Příkladu 1.1

`priklad2_1.m` – kód k Příkladu 2.1

`priklad3_1.m` – kód k Příkladu 3.1

`priklad3_2.m` – kód k Příkladu 3.2

`priklad3_3.m` – kód k Příkladu 3.3

`priklad4_1.m` – kód k Příkladu 4.1

`rozdel_na_skupiny.m` – funkce pro rozdělení objektů do skupin dle jejich průběhu

`sop.fit.R` – funkce pro SOP algoritmus převzatá z přílohy článku [1]

`vyuziti_SOP_algoritmu.R` – kód k převodu mezi matematickými softwary MATLAB a R a využití SOP algoritmu