

## 9

## INTERPOLACE A APROXIMACE

Interpolace a aproximace funkcí nebo experimentálních dat zahrnuje řadu technik. Účelem je provést náhradu funkce  $f(x)$ , zadané hodnotami  $\{x_i, y_i\}$ ,  $i = 1, \dots, n$ , vhodnou aproximující funkcí  $g(x)$ . Za aproximující funkci  $g(x)$  se často volí lineární kombinace  $m$ -tice elementárních funkcí  $g_j(x)$

$$g(x) = \sum_{j=1}^m c_j g_j(x)$$

Příkladem elementárních funkcí  $g_j(x)$  jsou polynomy, racionální funkce, podíly polynomů, trigonometrické funkce, exponenciální funkce atd. Aproximující funkce souvisí se zadáním dané úlohy a ovlivňuje stupeň aproximace. Ten se obvykle vyjadřuje jako vzdálenost mezi *aproximující funkcí*  $g(x)$  a *aproximovanou funkcí*  $f(x)$ , resp. diskrétními hodnotami  $y_i$ .

Zvláštním případem aproximace je *interpolace*: při interpolaci závislosti se sestruje funkce  $g(x)$  tak, aby procházela zadanými body  $\{x_i, y_i\}$ ,  $i = 1, \dots, n$ , a splňovala přitom podmínky týkající se jejího tvaru.

Při *interpolaci funkcí* musí být v definovaných bodech  $\xi_i$ ,  $i = 1, \dots, n$ , nazvaných *uzlové body interpolace*, funkce  $f(x)$  a  $g(x)$  spojité ve funkčních hodnotách a hodnotách zvolených derivací

$$f^{(j)}(\xi_i) = g^{(j)}(\xi_i), \quad i = 1, \dots, n, \quad j = 0, \dots, r_i$$

Zde  $f^{(j)}$  označuje  $j$ -tou derivaci a  $r_i$  je maximální derivace v  $i$ -tém uzlu, ve které jsou obě, aproximovaná a aproximující funkce, totožné.

**Interpolace** se v technické praxi využívá pro

(a) *zespojiténí tabelárních údajů*, například teplotní závislosti fyzikálně chemických konstant, jako jsou rozpustnost, hustota, iontový součin, relativní permitivita, součin rozpustnosti atd.;

(b) *náhradu složitých funkcí*  $f(x)$  nebo funkcí, které nelze přímo vyčíslit. Příkladem jsou Besselovy funkce, funkce Gamma, nekonečné řady atd.;

(c) *numerickou derivaci a integraci*;

(d) *kreslení grafu* závislosti zadané tabulkou.

Předpokladem interpolace závislosti jsou deterministické hodnoty souřadnice  $x$  a jim odpovídající hodnoty na ose  $y$ . V praxi je však častější případ, kdy  $x$  jsou volené (nastavo-

vané) hodnoty (např. čas, teplota) a  $y_i$  jsou jim odpovídající experimentálně změřené hodnoty (např. koncentrace, absorbance, napětí, proud atd.). Experimentální hodnoty  $y$  jsou pak zatíženy náhodnými chybami.

Při aproximaci závislosti se předpokládá aditivní působení chyb typu

$$y_i = g(x_i) + \varepsilon_i.$$

Pokud je druh funkce  $g(x)$  předem znám, přechází úloha aproximace na úlohu (ne)lineární regrese. Pokud se volí  $g(x)$  ve tvaru lineární kombinace elementárních funkcí, jde o úlohu lineární regrese. Rovněž úloha *aproximace funkce* se převádí na úlohu regrese, kde se však součty nahrazují integrály.

**Aproximace** se v technické praxi využívá k

(a) *vyhlazování závislostí*, tj. k eliminaci náhodných chyb  $\varepsilon$ . Pokud se data  $y_i$  pouze nahrazují hodnotami  $g(x_i)$  a  $x_{i+1} - x_i = \Delta$ ,  $i = 1, \dots, n - 1$ , jde o úlohu *číslicové filtrace*. Vyhlazení se užívají k určení vyhlazených hodnot  $g(x)$  nebo ke kreslení grafů;

(b) *náhradě rozsáhlých souborů dat* funkcemi, obsahujícími méně parametrů, k účelům uchování informací o datech, např. v paměti počítače;

(c) *numerické derivaci a integraci* experimentálních dat, zatížených náhodnými chybami;

(d) *tvorbě speciálních empirických modelů* regresního typu, jako je spline-regrese.

V řadě technických úloh je interpolace a aproximace dílčí částí postupu zpracování dat. V této kapitole jsou uvedeny vybrané, nejvíce užívané techniky aproximace a interpolace funkcí, resp. závislostí. Vedle klasických postupů jsou uvedeny i postupy, které využívají *po částech definovaných funkcí* (piecewise-funkcí).

## 9.1 Klasické interpolační postupy

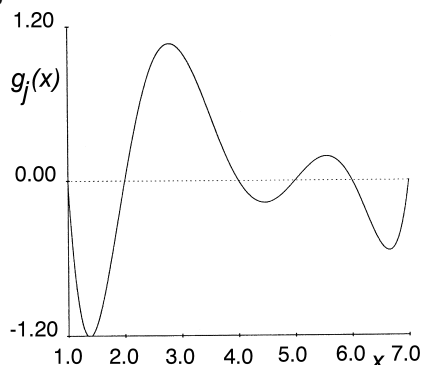
Úlohou interpolace je nalezení funkce  $g(x)$  tak, aby pro  $n$  hodnot  $x_1 < x_2 < \dots < x_n$  (v případě závislostí), nebo pro  $n$  uzlů  $\xi_1 < \xi_2 < \dots < \xi_n$  (v případě funkcí) platila úvodní rovnice. Protože platí  $x_i = \xi_i$ , budeme označovat uzly také symbolem  $x_i$ . Mezi nejznámější postupy patří **polynomická interpolace**, která hledá polynom  $g(x)$ , splňující úvodní podmínku  $f^{(j)}(\xi_i) = g^{(j)}(\xi_i)$ . Hledaný polynom je stupně nejvýše

$$m = \sum_{i=1}^n r_i \leq n - 1.$$

Pokud je požadavkem shoda pouze ve funkčních hodnotách, jsou  $r_i = 0$ ,  $i = 1, \dots, n$ , a  $n$ -tice bodů je interpolována jednoznačně polynomem  $(n - 1)$ ního stupně. Z úvodní podmínky se sestaví  $m$  lineárních rovnic, ze kterých se vypočtou odpovídající koeficienty  $c_j$ .

### 9.1.1 Lagrangeova a Newtonova interpolační formule

Formule se užívají pro případ  $r_i = 0$ , kdy se konstruuje polynom stupně nejvýše  $m = n - 1$ , interpolující  $n$  uzlových bodů, a kdy platí  $y_i = f(x_i) = g(x_i)$ . Interpolační polynom splňující tyto podmínky lze vyjádřit jako lineární kombinaci všech  $n$  hodnot  $L_m(x) = \sum_{j=1}^n y_j g_j(x)$ , kde  $g_j(x)$  jsou polynomy stupně  $(n - 1)$  takové, že  $g_j(x_i) = 0$  pro všechna  $j \neq i$ ,  $g_j(x_j) = 1$ . Tyto podmínky zajišťují, že  $L_m(x)$  je interpolační polynom  $(n - 1)$ . stupně. Funkce  $g_j(x)$  je znázorněna na obr. 9.2.



Obr. 9.2 Jednoduchý Lagrangeův polynom.

Polynomy  $g_j(x)$ , splňující tyto podmínky, lze vyjádřit ve tvaru

$$g_j(x) = w_j \prod_{\substack{k=1 \\ k \neq j}}^n (x - x_k) \prod_{\substack{k=1 \\ k \neq j}}^n (x - x_i) \dots (x - x_{j+1})(x - x_{j+2}) \dots (x - x_n)$$

kde normalizační koeficient  $w_j$  je roven  $w_j = \frac{1}{\prod_{\substack{k=1 \\ k \neq j}}^n (x_j - x_k)}$ . Lagrangeův interpolační

polynom má potom tvar

$$L_m(x) = \sum_{j=1}^n y_j w_j \prod_{\substack{k=1 \\ k \neq j}}^n (x - x_k) = \sum_{j=1}^n y_j \prod_{\substack{k=1 \\ k \neq j}}^n \frac{x - x_i}{x_i - x_j}$$

Tato formulace Lagrangeova interpolačního polynomu se hodí pouze pro malá  $n$  a jednoduché ruční výpočty. Pro výpočty s využitím počítače se doporučuje tzv. *barycentrická reprezentace* ve tvaru

$$L_m(x) = \frac{\prod_{j=1}^n y_j \frac{w_j}{x - x_j}}{\prod_{j=1}^n \frac{w_j}{x - x_j}} \quad \text{pro } x \in \dots x_j ,$$

$$L_m(x) = y_j \quad \text{pro } x = x_j .$$

Tato reprezentace interpolačního polynomu je numericky stabilní a navíc lze pro různá dělení uzlů  $x_i$ ,  $i = 1, \dots, n$ , určit normalizační koeficienty  $w_i$  analyticky<sup>2</sup>.

V případě, že byl Lagrangeův interpolační polynom použit pro interpolaci  $n$ -tice funkčních hodnot  $f(x_i)$  známé funkce  $f(x)$ , platí pro chybu interpolace v libovolném bodě  $x$  vztah

$$f(x) - L_{n+1}(x) = \frac{f^{(n)}(\alpha)}{n!} \prod_{j=1}^n (x - x_j) ,$$

kde  $x_1 < \alpha < x_n$ . Nevýhodou tradičního vyjádření interpolačního polynomu v Lagrangeově tvaru je požadavek na opětovné přepočítání všech členů při přidání dalšího bodu  $x_{n+1}$ ,  $y_{n+1}$ . Z tohoto hlediska je při postupném přidávání uzlů výhodnější *Newtonova interpolační formule*, pro kterou platí

$$P_m(x) = \prod_{j=1}^n a_j \prod_{k=1}^{j-1} (x - x_k) .$$

Přidání bodu  $x_{n+1}$ ,  $y_{n+1}$  pak vede k interpolačnímu polynomu

$$P_{m+1}(x) = P_m(x) + a_{n+1} \prod_{k=1}^n (x - x_k) .$$

Z této rovnice vychází při dosazení za  $x = x_j$  pro koeficient  $a_{n+1}$  vztah

$$a_{n+1} = \prod_{i=1}^{n+1} w_i y_i = \prod_{i=1}^{n+1} \frac{y_i}{\prod_{j=1, j \neq i}^{n+1} (x_i - x_j)} .$$

Pro ostatní koeficienty  $a_k$ ,  $k = 1, 2, \dots, n$  polynomu  $P_{m+1}$  platí

$$a_k = \prod_{i=1}^k w_i y_i \prod_{j=k+1}^{n+1} (x_i - x_j) .$$

Koeficienty  $a_k$  se nazývají postupné diference funkce  $f(x)$  a nultá postupná diference je  $a_1 = y_1$ . Ostatní postupné diference lze definovat rekurentně. Pro první postupnou diferenci platí

$$a_2 = \frac{y_1}{x_1 - x_2} + \frac{y_2}{x_2 - x_1} = \frac{y_2 - y_1}{x_2 - x_1} = [x_2, x_1]f .$$

Pro druhou postupnou diferencí je

$$a_3 = \frac{[x_3, x_2]f + [x_2, x_1]f}{x_3 + x_1} = \frac{\frac{y_3 + y_2}{x_3 + x_2} + \frac{y_2 + y_1}{x_2 + x_1}}{x_3 + x_1}$$

a pro  $i$ -tou postupnou diferencí  $a_{i+1}$  platí

$$a_{i+1} = \frac{[x_{i+1}, \dots, x_2]f + [x_i, \dots, x_1]f}{x_{i+1} + x_1}$$

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_{n+1} \\ a_n \end{bmatrix} = \begin{bmatrix} \binom{n}{j-2} (x_1 + x_j) & 0 & \dots & 0 \\ \binom{n}{j-3} (x_1 + x_j) & \binom{n}{j-3} (x_2 + x_j) & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ (x_1 + x_n) & (x_2 + x_n) & \dots & 0 \\ 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_{n+1} \\ d_n \end{bmatrix}$$

Při sestavování postupných diferencí se s výhodou sestavuje tabulka, jejíž diagonálu tvoří koeficienty  $a_j$ . Tak jsou definovány vztahy mezi koeficienty  $a$  Newtonovy formule a koeficienty  $w_k y_k = d_k$  Lagrangeovy formule. Při znalosti koeficientů  $a_k$  můžeme stanovit koeficienty  $d_k$  řešením soustavy  $n$  lineárních rovnic<sup>3</sup>.

Postup efektivního výpočtu  $d_i$ ,  $i = 1, \dots, n$ , z této soustavy rovnic je popsán v práci<sup>3</sup>. Na jeho základě byl odvozen algoritmus pro efektivní výpočet normalizačních koeficientů  $w_j$ ,  $j = 1, \dots, n$ , který lze vyjádřit posloupností vztahů:

$$\begin{aligned} a_1^{(1)} &= 1, \quad a_k^{(1)} = 0, & k &= 2, \dots, n, \\ a_k^{(i)} &= a_k^{(i-1)} / (x_k - x_i), & k &= 1, 2, \dots, i-1, \\ a_i^{(k+1)} &= a_i^{(k)} - a_k^{(i)}, & i &= 2, 3, \dots, n, \\ w_i &= a_i^{(n)}. \end{aligned}$$

Užitím tohoto schématu je počet operací potřebných pro vyčíslení polynomu v Lagrangeově tvaru shodný s počtem operací pro vyčíslení koeficientů v Newtonově tvaru<sup>3</sup>.

Vytváření postupných diferencí

Data	První diference	Druhá diference	... (n-1)ní diference
$x_1 y_1$			
	$[x_2, x_1]f$		
$x_2 y_2$		$[x_3, x_2, x_1]f$	
	$[x_3, x_2]f$		
$x_3 y_3$		$[x_4, x_3, x_2]f$	
	$[x_4, x_3]f$		

$x_4 y_4$ 

.

 $x_n y_n$  $[x_n, x_{n-1}, \dots, x_1]f$ 

### 9.1.2 Hermitovská interpolace

Při této interpolaci se požaduje, aby interpolační polynom  $H_m$  se svou první derivací souhlasil ve všech uzlových bodech s danou funkcí a její první derivací. To znamená, že  $r_i = 1, i = 1, \dots, n$  a interpolační polynom je stupně  $(2n - 1)$ . Označíme-li hodnoty derivací v uzlových bodech  $x_i$  jako  $y'_i$ , můžeme psát

$$H_m(x) = \sum_{i=1}^n y_i h_i(x) + \sum_{i=1}^n y'_i \bar{h}_i(x),$$

kde  $h_i(x) = [1 + 2(x - x_i)g_i'(x_i)]g_i^2(x)$  a  $\bar{h}_i(x) = (x - x_i)g_i^2(x)$ . V těchto vztazích jsou  $g_i(x)$  elementární Lagrangeovy polynomy. Pro Hermitovskou interpolaci je vhodné použít barycentrického tvaru

$$H_m(x) = \frac{\sum_{i=1}^n \left[ \frac{w_i}{x - x_i} \left( \frac{w_i}{x - x_i} + v_i \right) y_i + w_i \frac{w_i}{x - x_i} y_i \right]}{\sum_{i=1}^n \frac{w_i}{x - x_i} \left( \frac{w_i}{x - x_i} + v_i \right)} \quad \text{pro } x \neq x_j$$

$$H_m(x) = y_j \quad \text{pro } x = x_j.$$

Pro koeficienty  $v_i$  platí<sup>2</sup>  $v_i = 2w_i \prod_{\substack{j=1 \\ j \neq i}}^n \frac{1}{x_i - x_j}$ ,  $i = 1, 2, \dots, n$ . V práci<sup>3</sup> je uveden

efektivní algoritmus simultánního postupného určování koeficientů  $w_i$  a  $v_i$ .

### 9.1.3 Racionální interpolace

Při této aproximaci je interpolující funkce  $R_{m,l}(x)$  definována jako podíl polynomu stupně  $m$  (v čitateli) a polynomu stupně  $l$  (ve jmenovateli)

$$R_{m,l}(x) = \frac{P_m(x)}{P_l(x)}.$$

Tato aproximace nahrazuje klasickou polynomickou interpolaci stupně  $(m + 1)$ . S výhodou se používá *racionální aproximace typu Padé*

$$R(x) = b_1 + \frac{x - x_1}{b_2 - b_1} \frac{x - x_3}{x - x_2} + \frac{x - x_1}{b_2 - b_1} \frac{x - x_2}{b_4 - b_2} \frac{x - x_5}{x - x_4} + \dots$$

Místo tohoto zápisu se používá i zkrácená forma

$$R(x) = b_1 + \frac{x - x_1}{b_2 - b_1} + \frac{x - x_2}{b_3 - b_2} + \dots + \frac{x - x_{n-1}}{b_n - b_{n-1}}$$

Pro určení koeficientů  $b_1, \dots, b_n$  tak, aby  $R(x)$  interpolovala zadanou funkci v  $n$  uzlech, se používá rekurentních formulí

$$\begin{aligned} R_1(x) &= b_1 + \frac{x - x_1}{R_2(x) - b_1} \\ R_2(x) &= b_2 + \frac{x - x_2}{R_3(x) - b_2} \\ &\vdots \\ R_i(x) &= b_i + \frac{x - x_i}{R_{i+1}(x) - b_i} \\ &\vdots \\ R_n(x) &= b_n \end{aligned}$$

Za předpokladu, že  $R_{i+1}(x) \neq 0$ , dostáváme, že

$$b_i = R_i(x_i), \text{ pro } i = 1, \dots, n$$

Pro interpolaci musí ještě platit omezení, že

$$R_1(x_i) = y_i, \quad i = 1, \dots, n$$

Z této rovnice přímo plyne, že  $b_1 = y_1$ . Z předešlých rovnic lze také určit, že platí

$$R_{j+1}(x_i) = \frac{x_i - x_j}{R_j(x_i) - b_j}, \quad i = j + 1, j + 2, \dots, n$$

Využitím této rovnice lze pro  $j = 1, 2, \dots, n - 1$ , počítat  $R_{j+1}(x_i)$  pro  $i = j + 1, j + 2, \dots, n$ , a určovat  $b_{j+1} = R_{j+1}(x_{j+1})$ . Tímto postupem se jednoduše určí koeficienty Padé interpolace. Pokud vyjde  $R_{j+1}(x_j) = 0$  pro  $j = 1, \dots, n - 2$  rovno nule, nelze tento postup použít.

## 9.2 Spline interpolace

Užívání polynomiálních interpolačních formulí má řadu nevýhod. Jsou totiž složeny z elementárních funkcí definovaných na celé reálné ose, což vede u interpolačních formulí

vyšších řádů ke vzniku řady lokálních minim, maxim a inflexních bodů, které neodpovídají průběhu funkce  $f(x)$  či tabelované závislosti  $\{x_i, y_i\}$ ,  $i = 1, \dots, n$ . Při interpolaci fyzikálních závislostí se stává, že chování v jistém intervalu se výrazně liší od jejich chování v intervalech sousedních. Jde o závislost tzv. **neasociativní povahy**. Z těchto úvah plyne, že pro účely interpolace, ale i aproximace, bude výhodnější volit lokálně definované funkce, které budou v místech vzájemného styku, tj. v uzlech, spojité ve funkčních hodnotách a hodnotách zadaných derivací.

Vhodné interpolační funkce tohoto typu jsou složeny z polynomických úseků a platí pro ně, že jsou ze třídy  $C^m[a, b]$ . Obecně jsou funkce třídy  $C^m[a, b]$  na intervalu  $[a, b]$  spojité v prvních  $m$  derivacích a funkčních hodnotách. Na obr. 9.6 jsou schematicky znázorněny funkce třídy  $C^0$ ,  $C^1$ ,  $C^2$  a odpovídající první a druhé derivace. Z obr. 9.6 plyne, že hladké jsou všechny funkce od třídy  $C^1$ . Pro funkce třídy  $C^m$  platí, že  $m$ -tá derivace je lineární lomená závislost,  $(m + 1)$  derivace je po částech konstantní a  $(m + 2)$  derivace je po částech nulová, tj. není definovaná v uzlových bodech  $\xi_j$ .

	$C^0$	$C^1$	$C^2$
$f(x)$	S 	H 	H 
$f(x)^{(1)}$		S 	H 
$f(x)^{(2)}$			S 

Obr. 9.6 Příklady funkcí  $C^0$ ,  $C^1$ ,  $C^2$  a jejich derivací: H značí hladká, S značí spojité křivky.

Využitím uvedených vlastností funkcí ze třídy  $C^m[a, b]$  můžeme definovat obecně polynomický spline  $S_m(x)$  s uzly  $a = \xi_1 < \xi_2 < \xi_3 < \dots < \xi_n = b$ . Tento spline je na každém úseku  $[\xi_j, \xi_{j+1}]$ ,  $j = 1, \dots, n - 1$ , reprezentován polynomem maximálně  $m$ -tého stupně. Pokud je v nějakém bodě  $x_i$  některá derivace  $S_m^{(j)}(\xi_i)$  nespojitá, jde o **defektní spline**. Vlastnosti spline  $S_m(\xi_i)$  závisí na

- řádu polynomu  $m$ , přičemž se obvykle volí kubický spline  $m = 3$ ;
- počtu a polohách uzlů  $\xi_1 < \xi_2 < \dots < \xi_n$ ;
- defektech v uzlových bodech.

Z defektních spline se omezíme na klasické spline, které mají **minimální defekt** roven  $k = 1$ , tj. patří do třídy  $C^{m-1}[a, b]$ . Pro účely Hermitovské interpolace je výhodné použít spline defektu  $k = 2$ , který patří do třídy  $C^{m-2}[a, b]$ . Defekt  $k = 1$  umožňuje zadat podmínky interpolace a defekt  $k = 2$  ještě navíc podmínky týkající se hodnot prvních derivací.

Klasické spline polynomy  $S_m(x)$  ze třídy  $C^{m-1}[a, b]$  je možno definovat několika způsoby. Nejjednodušší je pro každý interval  $I_j \in [\xi_{j+1}, \xi_j]$ ,  $j = 1, 2, \dots, n - 1$ , definovat lokální



polynom  $P_j(x) = c_0 + \sum_{k=1}^m c_k (x - \xi_j)^k$ , pro  $x \in I_j$ . Tento zápis je redundantní, protože

$P_j(x)$  obsahuje v každém  $(m+1)$  intervalu parametrů  $c_k$  a celkově pak  $n(m+1)$  parametrů.

Nejefektivnější je vyjádření spline  $S_m(x)$  jako lineární kombinace bázevých  $B$ -spline s minimální podporou. Bázevých  $B$ -spline  $B_{m,j}$  je definován v  $(m+1)$  uzlových bodech  $\xi_{j-m} < \xi_{j-m+1} < \dots < \xi_j$  jako normalizovaná  $m$ -tá poměrná diference useknutého polynomu  $g(\xi) = (\xi - x)_+^{m-1}$ . Využitím formálního zápisu postupné diference můžeme psát  $B_{m,j} = (\xi_j - \xi_{j+m}) [\xi_{j+m}, \dots, \xi_j] g$ . V praxi se pro výpočet normalizovaných  $B$ -spline používá rekurentní formule

$$B_{m,j}(x) = \frac{x - \xi_{j+m}}{\xi_{j+1} - \xi_{j+m}} B_{m-1,j+1}(x) + \frac{\xi_j - x}{\xi_j - \xi_{j+m}} B_{m-1,j}(x).$$

Začíná se od  $B_{1,j}(x)$ , pro které platí

$$B_{1,j}(x) = \begin{cases} 1 & \text{pro } \xi_{j+1} \leq x \leq \xi_j \\ 0 & \text{jinde} \end{cases}.$$

Bázevých  $B$ -spline  $m$ -tého řádu mají zajímavé vlastnosti:

(a) jsou kladné pouze v intervalu  $\xi_{j-m} < x < \xi_j$  a všude jinde nabývají nulových hodnot,

$$B_{m,j}(x) > 0 \quad \text{pro } \xi_{j+m} < x < \xi_j;$$

$$B_{m,j}(x) = 0 \quad \text{jinde};$$

(b) jsou normalizované, tj.  $\sum_{(j)} B_{m,j}(x) = 1$  pro všechna  $\xi_1 < x < \xi_n$ ;

(c) na intervalu  $(\xi_{j-m}, \xi_j)$  je  $B_{m,j}(x)$  spline polynom stupně  $(m-2)$  s uzlovými body  $(\xi_{j-m}, \xi_{j-m+1}, \dots, \xi_j)$ . To znamená, že v každém intervalu je mezi dvojicí uzlových bodů vyjádřen spoj polynomem stupně maximálně  $(m-1)$  a patří do třídy funkcí  $C^{m-2}[\xi_{j-m}, \xi_j]$ . Tato vlastnost platí, pokud jsou všechny uzlové body  $\xi_j$  navzájem různé.

## 9.2.2 Kubické spline

Kubické spline, nazývané také *křivítkové funkce*, patří k nejznámějším představitelům spline polynomů. Jejich základní výhodou je, že jsou spojité v prvních dvou derivacích, což umožňuje sestavení hladké křivky v první derivaci (obr. 9.6). Používají se hojně ve všech oblastech počítačové grafiky, v systémech CAD/CAM i pro aproximaci funkcí, kde mají řadu vhodných vlastností.

Platí, že při aproximaci funkce  $f(x)$  kubickým splinem  $S_3(x)$  je zajištěna minimální norma  $\int_a^b (f''(x) - S_3''(x))^2 dx$ . Dále  $S_3(x)$  splňují i podmínku

$$\int_a^b [S_3(x)]^2 dx \leq \min,$$

což znamená, že jistým způsobem minimalizují celkovou křivost interpolující funkce. Fyzikálně si lze spline představit jako ideální elastický nosník, se zanedbatelnou hmotností, který je zatížen nebo podepřen v uzlových bodech  $\{x_i, y_i\}$ ,  $i = 1, \dots, n$ . Tento nosník zaujme tvar odpovídající minimu potenciální energie. Interpoláčnı kubický spline lze sestřoit při volbě  $m = 3$  přímo z definice. Numericky nejvýhodnější je však použití  $B$ -spline  $B_4(x)$ , znázorněných na obr. 9.9. Pro ilustraci je výhodně vyjıt přímo z definice  $S_3(x)$  jako polynomické funkce třídy  $C^2[a, b]$ . Kubický spline je pak definován podmínkami

(a) podmínkou interpolace, tj.  $S_3(x_i) = y_i$ ,  $i = 1, \dots, n$ ,

(b) podmínkami spjitosti ve funkčních hodnotách a hodnotách první i druhé derivace. Funkce  $S(x)$ ,  $S^{(1)}(x)$  a  $S^{(2)}(x)$  jsou spjité v celém intervalu  $[a, b]$ ;

(c) podmínkou po částech konstantní třetí derivace  $S^{(3)}(x)$  všude kromě uzlových bodů  $x_i$ ,  $i = 1, \dots, n$ ;

(d) podmínkou nulové čtvrté derivace  $S^{(4)}(x) = 0$  všude kromě uzlových bodů  $x_i$ ,  $i = 1, \dots, n$ .

Z těchto podmınek plyne, že  $S_3(x)$  je v každém intervalu  $I_j$   $[x_{i+1}, x_i]$  definována kubickým polynomem. Pro jednoznačné určení všech čtveřic koeficientů  $c_1$  až  $c_4$  ve všech  $(n - 1)$  intervalech  $I_j$  je potom nutné sestavit  $4(n - 1)$  nezávislých podmınek. Z podmınek spjitosti plyne, že

$$P_j(x_i) = P_{j+1}(x_i), \quad P_j^{(1)}(x_i) = P_{j+1}^{(1)}(x_i), \quad P_j^{(2)}(x_i) = P_{j+1}^{(2)}(x_i)$$

pro všechna  $i = 2, \dots, n - 1$ , což vede na  $3(n - 2)$  vazebných rovnic. Z podmınek interpolace vychází dalších  $n$  rovnic. Celkově vede použití podmınek definice  $S_3(x)$  k sestavení  $(4n - 6)$  lineárních rovnic. Dvě rovnice pro jednoznačné určení  $S_3(x)$  však ještě scházejí. Tyto vazebné rovnice se využívají pro definici *okrajových podmınek* určujících chování spline v místě  $x_1$  a  $x_n$ . Často bývají voleny tzv. přirozené okrajové podmıny  $S_3^{(2)}(x_1) = S_3^{(2)}(x_n) = 0$ . Odpovídající podmıny pro první derivace

$$d_1 = S_3^{(1)}(x_1), \quad d_n = S_3^{(1)}(x_n)$$

mají tvar  $d_1 = 0.5 \left[ \frac{3(y_2 - y_1)}{h_1} + d_2 \right]$ ,  $d_n = 0.5 \left[ \frac{3(y_n - y_{n+1})}{h_{n+1}} + d_{n+1} \right]$ . Název

"přirozené" zde vystihuje fyzikální smysl prostého podepření, kdy vně poslední podpory  $x_1$  a resp.  $x_n$  zaujímá nosník přímkový tvar.

Obecně lze definovat *okrajové podmıny typu I*, kdy se určují  $d_1$ ,  $d_n$  a *okrajové podmıny typu II*, kdy se určují druhé derivace  $S_3^{(2)}(x_1)$  a  $S_3^{(2)}(x_n)$ . Přehled různých typů okrajových podmınek, které ovlivňují chování spline  $S_3(x)$  pouze lokálně, je uveden v práci<sup>12</sup>.

Označme druhé derivace kubického spline  $M_i = S_3^{(2)}(x_i)$  a první derivace  $d_i = S_3^{(1)}(x_i)$ . Pro určení kubického spline postačuje nalezení derivací  $d_2, \dots, d_{n-1}$ . Vyjádříme-li koeficienty  $c_1$  až  $c_4$  pomocí derivací  $d_i, d_{i+1}, M_i, M_{i+1}$ , můžeme z podmıny spjitosti druhých derivací v místě  $x_i$  dospět po úpravách ke vztahu

$$\alpha_i d_{i+1} + \beta_i d_i + \gamma_i d_{i-1} = \delta_i, \quad i = 2, \dots, n - 1,$$

$$\text{kde } \alpha_i = \frac{1}{h_{i+1}}, \quad \beta_i = \frac{2}{h_{i+1} + h_i}, \quad \gamma_i = \frac{1}{h_i}$$

$$\text{a } \delta_i = 3 \left[ \frac{y_i + y_{i+1}}{h_{i+1}^2} - \frac{y_{i+1} + y_i}{h_i^2} \right].$$

Zápis představuje tridiagonální soustavu  $(n - 2)$  lineárních rovnic pro neznámé  $d_1, \dots, d_{n-2}$ . Při použití okrajových podmínek pak dostáváme soustavu  $n$  lineárních rovnic s  $n$  neznámými, kdy je matice koeficientů tridiagonální. Pro její řešení lze využít např. kompaktních algoritmů vycházejících z Gaussovy eliminace<sup>1</sup>.

Základním problémem při použití kubických spline pro rekonstrukci závislosti je jejich tendence k překmitávání při náhlých změnách křivosti interpolované závislosti. Pro odstranění této nevýhody je vhodné použít tzv. *spline pod napětím* (exponenciální spline), která umožňují lokální řízení tvaru interpolující funkce pomocí parametru napětí. Spline pod napětím jsou řešením diferenciální rovnice<sup>14</sup>

$$S_T^{(4)}(x) - \lambda^2 S_T^{(2)}(x) = 0$$

pro  $x = x_i, i = 1, \dots, n$ . Symbol  $\lambda$  označuje *parametr napětí*. Ten může být obecně v každém uzlovém bodě jiný a roven  $\lambda_i$ . Řešením uvedené diferenciální rovnice je spline pod napětím s těmito vlastnostmi:

(a) v každém intervalu  $I_j = [x_{i+1}, x_i]$  je

$$S_T(x) = a + b x + \exp(\lambda x) + \exp(-\lambda x) \dots$$

(b)  $S_T(x)$  je ze třídy funkcí  $C^2[a, b]$ ,

(c) platí podmínka interpolace  $S_T(x_i) = y_i, i = 1, \dots, n$ ,

(d)  $S_T^{(2)}(x_1) = S_T^{(2)}(x_n) = 0$ , tj. platí přirozené okrajové podmínky.

Pro hraniční hodnoty parametru napětí  $\lambda$  platí, že

(a) pro  $\lambda = 0$  je  $S_T(x)$  klasický kubický spline;

(b) pro  $\lambda \neq 0$  je  $S_T(x)$  lineární spline, tj. lomená čára spojující uzlové body.

Pro praktické účely se využívá toho, že rozdíl  $S_T^{(2)}(x) - \lambda^2 S_T(x)$  je v každém intervalu  $I_j$  lineární funkcí  $x$ . Po dvojitě integraci lze pak  $S_T(x)$  vyjádřit ve tvaru

$$S_T(x) = y_{i+1} + t y_i + \frac{M_{i+1}}{\lambda_i^2} \left[ \frac{\sinh(\mu_i t)}{\sinh(\mu_i)} + t \right] + \frac{M_i}{\lambda_i^2} \left[ \frac{\sinh(\mu_i (1 - t))}{\sinh(\mu_i)} + (1 - t) \right]$$

kde  $t = (x - x_{i+1})/h_i$  a  $\mu_i = \lambda_i h_i$ . Symboly  $M_i = S_T^{(2)}(x_i)$  označují druhé derivace. Z této rovnice je patrné, že

(a)  $S_T(x)$  je vzhledem k  $M_i, i = 2, \dots, n - 1$ , lineární;

(b) při znalosti  $M_i$ ,  $i = 2, \dots, n-1$ , je  $S_T(x)$  jednoznačně určeno pro zvolená  $\lambda_i$ ,  $i = 1, \dots, n$ .

Uvedme, že funkce  $\sinh(x)$  a  $\cosh(x)$  lze vyjádřit využitím funkce  $\exp(x)$  ve tvaru

$$\sinh(x) = 0.5 [\exp(x) - \exp(-x)],$$

$$\cosh(x) = 0.5 [\exp(x) + \exp(-x)].$$

Analogicky jako u klasických kubických spline lze z podmínky spojitosti prvních derivací  $S_T'(x)$  v uzlových bodech nalézt tridiagonální soustavu lineárních rovnic

$$\alpha_i M_{i+1} - (\beta_i - \beta_{i+1}) M_i - \alpha_{i+1} M_{i+1} = \delta_i, \quad \text{kde } \alpha_{i+1} = \frac{\sinh(\mu_i) + \mu_i}{\mu_i^2 \sinh(\mu_i)} h_i,$$

$$\beta_i = \frac{\mu_i \cosh(\mu_i) + \sinh(\mu_i)}{\mu_i^2 \sinh(\mu_i)} h_i, \quad \delta_i = \frac{y_{i+1} - y_i}{h_i} + \frac{y_i - y_{i+1}}{h_{i+1}}.$$

Řešení této soustavy rovnic lze provést stejně jednoduše jako u klasických polynomických spline. Samostatným problémem souvisejícím s použitím spline pod napětím je volba parametrů napětí. Rentrop<sup>15</sup> využívá ve své proceduře testu, zda druhé postupné diference  $[x_{i-1}, x_i, x_{i+1}] y = \Delta^2 y_i$  souhlasí co do znaménka s druhou derivací  $M_i$ . Pokud vyjde, že

$$(M_i \Delta^2 y_i) (M_{i+1} \Delta^2 y_{i+1}) > 0, \quad \text{dosazuje se } \lambda_i = 0. \text{ Není-li tato podmínka splněna,}$$

rozlišují se dva případy:

a) je-li  $y_i = y_{i+1} = 0$ , volí se  $\lambda_i = 15/h_i$ ;

b) je-li  $y_i = y_{i+1}$ , volí se  $\lambda_i = \frac{1}{h_i} \left[ 4 - (0.1 - y_{i+1} + y_i^* \max(y_i^*, y_{i+1}^*))^{0.1} \right]$ .

### 9.3 Aproximace funkcí

Při aproximaci funkce  $f(x)$  vhodnou aproximující funkcí  $g(x)$  je třeba řešit dvě základní úlohy: (a) výběr typu funkce  $g(x)$ ; (b) výběr kritéria pro vyjádření blízkosti funkcí  $f(x)$  a  $g(x)$ . S ohledem na jednoduchost zpracování se často volí  $g(x)$  ve tvaru  $g_j(x) = x^{j-1}$ , tj. polynomická aproximace. Blízkost funkcí  $f(x)$  a  $g(x)$  se vyjadřuje pomocí normy ve zvoleném  $L_p$ -prostoru, pro kterou platí

$$S_p = \int_a^b w(x) |f(x) - g(x)|^p dx.$$

V této rovnici je  $w(x)$  vhodná váhová funkce a interval  $a \leq x \leq b$  určuje oblast, ve které se hledá aproximace funkce  $f(x)$  funkcí  $g(x)$ . Koeficienty  $c_j$  se pak hledají tak, aby bylo  $S_p$  minimální. Při volbě  $p = 1$  jde o  $L_1$ -aproximaci a minimalizuje se integrál absolutních odchylek mezi  $f(x)$  a  $g(x)$ ; při volbě  $p = 2$  se minimalizuje integrál čtverců odchylek a jde o  $L_2$ -aproximaci, odpovídající kritériu metody nejmenších čtverců pro diskrétní data.

Konečně při volbě  $p = 4$  jde o *minimaxní (Čebyševovu) aproximaci*, minimalizující kritérium

$$S_4 = \max_{x \in [a, b]} [w(x) |f(x) - g(x)|],$$

Minimalizace kritéria  $S_p$  vede obecně na úlohu nelineární optimalizace. Zde se omezíme na  $L_2$ -normu. Jde o spojitou analogii úlohy lineární regrese, která je podrobně popsána v kap. 6.

Uvažujme pro jednoduchost  $w(x) = 1$ . Pro odhad koeficientů  $c_1$  až  $c_m$  se podobně, jako v diskretním, případě vychází z analytické minimalizace  $S_2$

$$\frac{\delta S_2}{\delta c_j} = 0, \quad j = 1, \dots, m.$$

Po dosazení obdržíme soustavu normálních rovnic ve tvaru

$$\begin{bmatrix} \int_a^b g_1^2 dx \\ \int_a^b g_2^2 dx \\ \vdots \\ \int_a^b g_j^2 dx \\ \vdots \\ \int_a^b g_m^2 dx \end{bmatrix} = \begin{bmatrix} \int_a^b g_1 g_1 dx & \int_a^b g_1 g_2 dx & \dots & \int_a^b g_1 g_m dx \\ \int_a^b g_2 g_1 dx & \int_a^b g_2^2 dx & \dots & \int_a^b g_2 g_m dx \\ \vdots & \vdots & \ddots & \vdots \\ \int_a^b g_1 g_j dx & \int_a^b g_2 g_j dx & \dots & \int_a^b g_j g_m dx \\ \vdots & \vdots & \ddots & \vdots \\ \int_a^b g_1 g_m dx & \int_a^b g_2 g_m dx & \dots & \int_a^b g_m^2 dx \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_j \\ \vdots \\ c_m \end{bmatrix}.$$

Je použito označení  $f = f(x)$  a  $g_j = g_j(x)$ . Pokud máme analytické vyjádření funkce  $f(x)$  a zvolíme vhodně  $g_j(x)$ , můžeme určit jednotlivé integrály analyticky a řešit pak soustavu  $m$  lineárních rovnic o  $m$  neznámých stejně jako v diskretním případě (viz odd. 9.6). Při polynomicke aproximaci je výhodné provést transformaci nezávisle proměnné tak, aby byl integrační obor v rozmezí  $[-1, 1]$ . Toho lze docílit volbou

$$x = \frac{2x' + a + b}{b - a}.$$

Úlohou je pak určení koeficientů polynomu  $g(x) = \sum_{j=1}^m b_j x^{(j-1)}$  tak, aby byla ve smyslu

normy  $L_2$  nejlépe aproximována funkce  $f\left(\frac{a+b}{2} + \frac{b-a}{2} x'\right)$  v intervalu  $[-1, 1]$ . Při

sestavování matice koeficientů uvedené maticové rovnice lze využít známých integrálů

$$\int_{-1}^1 x^{2j} dx = \frac{2}{2j+1}, \quad \int_{-1}^1 x^{2j+1} dx = 0.$$

Pro  $(m-1)$  sudé přechází tato maticová rovnice do tvaru

$$\begin{bmatrix} 1 & 0 & \frac{1}{3} & 0 & \frac{1}{5} & \dots & \frac{1}{m \% 1} \\ 0 & \frac{1}{3} & 0 & \frac{2}{5} & 0 & \dots & 0 \\ \frac{1}{3} & 0 & \frac{1}{5} & 0 & \frac{1}{7} & \dots & \frac{1}{m \% 3} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \dots & \cdot \\ \frac{1}{m \% 1} & 0 & \frac{1}{m \% 3} & 0 & \frac{1}{m \% 5} & \dots & \frac{1}{2m \& 1} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \cdot \\ \cdot \\ b_m \end{bmatrix} = \begin{bmatrix} I_0 \\ I_1 \\ I_2 \\ \cdot \\ \cdot \\ I_{m\&1} \end{bmatrix},$$

$$\text{kde } I_j = 0.5 \int_{-1}^1 x^j f\left(\frac{a \% b}{2} \% \frac{b \& a}{2} x\right) dx.$$

V této rovnici jsou oproti předešlé rovnici všechny koeficienty děleny dvěma tak, aby došlo ke zjednodušení matice koeficientů. Pro zvolený stupeň polynomicke aproximace  $m$  lze určit koeficienty  $b_1$  až  $b_m$  jako lineární kombinace známých integrálů. Obdobně lze postupovat i pro  $(m - 1)$  liché. V tabulce jsou pro  $m = 2, 3, 4, 5$  uvedeny výrazy pro koeficienty  $b_j$ .

Koeficienty aproximačních polynomů  $g(x^*)$  různých stupňů  $(m - 1)$

$m$	$b_1$	$b_2$	$b_3$
2	$I_0$	$3I_1$	0
3	$\frac{3}{4} (3I_0 \& 5I_2)$	$3I_1$	$\frac{15}{4} (3I_2 \& I_0)$
4	$\frac{3}{4} (3I_0 \& 5I_2)$	$\frac{15}{4} (5I_1 \& 7I_3)$	$\frac{15}{4} (3I_2 \& I_0)$
5	$\frac{15}{64} (15I_0 \& 70I_2 \% 63I_4)$	$\frac{15}{4} (5I_1 \& 7I_3)$	$\frac{105}{32} (42I_2 \& 45I_4 \& 5I_0)$

$m$	$b_4$	$b_5$
2	0	0
3	0	0
4	$\frac{35}{4} (5I_3 \& 3I_1)$	0
5	$\frac{35}{4} (5I_3 \& 3I_1)$	$\frac{315}{64} (3I_0 \& 30I_2 \% 35I_4)$

Postačuje-li tedy aproximace polynomem maximálně čtvrtého stupně, lze určit jeho koeficienty přímo z tabulky. Pro vyjádření kvality aproximace se počítá střední kvadratická odchylka<sup>17</sup>

$$SE = \sqrt{\frac{1}{b-a} \int_a^b (f(x) - g(x))^2 dx}$$

$$= \sqrt{\frac{1}{b-a} \int_a^b f^2(x) dx - \sum_{i=1}^m b_i I_{i+1}}$$

## 9.4 Aproximace tabelárních závislostí

Úloha aproximace závislostí, zadaných tabulkou  $\{x_i, y_i\}$ ,  $i = 1, \dots, n$ , se od úlohy aproximace funkcí liší pouze v tom, že místo integrálu se v rovnici k vyjádření kritéria  $S_p$  užívá sumy. Pro známé  $g(x)$  a  $p = 2$  jde o úlohu nelineární nebo lineární regrese, která je podrobně popsána v kap. 8 a kap. 6.

### 9.4.1 Polynomická aproximace

V kap. 6 je pojednáno o odhadu parametrů v polynomických modelech, hledání vhodného stupně polynomu a využití ortogonálních polynomů na dané posloupnosti hodnot  $x_1, x_2, \dots, x_n$ . Jde vždy o metodu nejmenších čtverců odchylek, tj.  $L_2$ -aproximaci. V řadě úloh aproximace metoda nejmenších čtverců odchylek nevyhovuje a požaduje se minimalizace maximální odchylky. Toho lze při polynomické aproximaci docílit využitím *ortogonálních Čebyševových polynomů*. Čebyševovy polynomy  $T_m(x)$  lze generovat podle rekurentní formule

$$T_{m+1}(x) = 2x T_m(x) - T_{m-1}(x),$$

Platí, že  $T_0(x) = 1$  a  $T_1(x) = x$ . Čebyševovy polynomy mají tyto základní vlastnosti<sup>18</sup>:

- (a) koeficient u maximální mocniny  $x^m$  je roven  $2^{m-1} - 1$  pro  $m \geq 1$  nebo 1 pro  $m = 0$ ,
- (b) Čebyševovy polynomy jsou symetrické kolem počátku, tj. platí

$$T_m(-x) = (-1)^m T_m(x),$$

(c) Čebyševův polynom  $T_m(x)$  má v intervalu  $[-1, 1]$  právě  $m$  nulových bodů  $T_m(x) = 0$  v místech  $x_j^*$ , které se nazývají čebyševovské uzlové body.

(d) Čebyševův polynom  $T_m(x)$  má v intervalu  $[-1, 1]$  právě  $(m + 1)$  extrémů  $x_j^+$ , pro které platí

$$x_j^+ = \frac{\cos(j \pi)}{n}; \quad T_m(x_j^+) = (-1)^j \quad \text{pro } j = 0, 1, \dots, m,$$

(e) při zavedení váhové funkce  $w(x) = 1/\sqrt{1-x^2}$  jsou Čebyševovy polynomy vzájemně ortogonální na celém intervalu  $[-1, 1]$ .



(f) pokud se definuje  $(m + 1)$  bodů  $x_j^*$ , které jsou nulovými body Čebyševova polynomu  $T_{m+1}(x)$ , platí, že

$$\int_{j-1}^{m \% 1} T_i(x_j^*) T_k(x_j^*) = \begin{cases} 0 & \text{pro } i \neq k \\ 0.5 (m \% 1) & \text{pro } i = k \neq 0 \\ m \% 1 & \text{pro } i = k = 0 \end{cases}$$

Tato rovnice platí pro všechna  $i, k = 0, \dots, m$ , a ukazuje, že na Čebyševových uzlech jsou Čebyševovy polynomy vzájemně ortogonální.

(g) ze všech polynomů  $m$ -tého stupně, které mají koeficient u mocniny  $x^m$  roven 1, má normalizovaný Čebyševův polynom  $T_m(x)/2^{m-1}$  minimální hodnotu normy  $S_d$  v intervalu  $[-1, 1]$ . Pomocí Čebyševových polynomů lze aproximující funkci  $g(x)$  vyjádřit ve tvaru

$$g(x) = \sum_{j=0}^m c_j T_j(x) \quad \text{pro } -1 \leq x \leq 1$$

Pro zajištění ortogonality funkcí  $T_j(x)$  se nejdříve určí Čebyševovy uzly dvoufázovým postupem:

(a) pro zvolené  $n$  se určí uzlové body  $x_{j+1}^*, j = 0, \dots, n-1$ , v intervalu  $\{-1, 1\}$ ,

(b) využitím vztahu  $Z_j^* = 0.5(a+b) + 0.5(b-a)x_j^*$  se určí Čebyševovy uzly  $Z_j^*$  v intervalu  $[a, b]$ .

Pro hodnoty  $Z_j^*$  se následně určí buď funkční hodnoty  $f(Z_j^*)$ , nebo hodnoty závislosti  $y_j$ . Vzhledem k ortogonalitě jednotlivých  $T_j(x)$  pro Čebyševovy uzly lze určit koeficienty  $c_j$  snadno ze vztahů

$$c_0 = \frac{1}{n} \sum_{j=1}^n f(Z_j^*) \quad \text{a} \quad c_j = \frac{2}{n} \sum_{i=1}^n \frac{T_j(x_i^*) f(Z_i^*)}{T_j(x_i^*)}$$

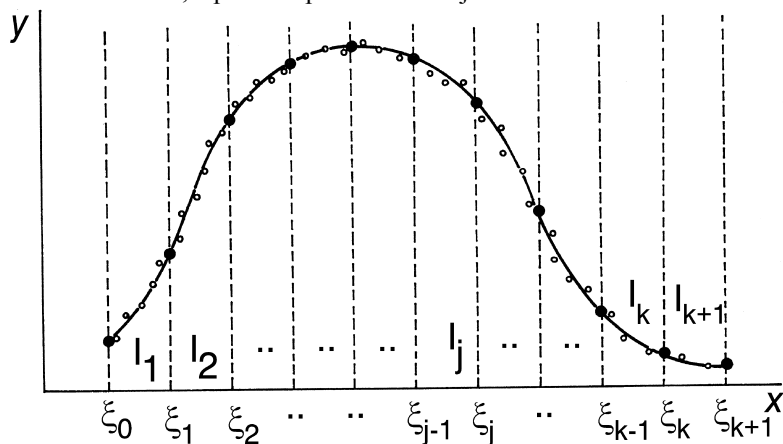
Obě rovnice odpovídají použití klasické metody nejmenších čtverců.

## 9.4.2 Úseková regrese

Použití klasických polynomů jako aproximačních modelů je nevhodné při aproximaci fyzikálních závislostí, které nejsou asociativní povahy. Pro komplikovanější průběhy s několika extrémy mají navíc tendenci oscilovat, či výrazně zkreslovat aproximovanou závislost. V těchto případech je výhodnější použít po částech definovaných funkcí. Kromě zadaných  $n$  bodů  $\{x_i, y_i\}, i = 1, \dots, n$ , kde se předpokládá, že  $y_i$  jsou náhodné veličiny (měřené hodnoty), se ještě určují uzlové body  $\xi_j, j = 1, \dots, k$  (resp. ještě  $\xi_0, \xi_{k+1}$ ). Uzlové body tvoří hranice intervalů, kde jsou definovány jednotlivé funkce. V každém intervalu  $I_j$ , ohraničeném uzlovými body  $\xi_{j-1}, \xi_j$ , lze aproximující funkci  $g(x)$  vyjádřit modelem  $g_j(x)$ , takže platí:

$$\begin{aligned} g(x) &= g_1(x) \quad \text{pro } x \in I_1 \\ &\vdots \\ g(x) &= g_j(x) \quad \text{pro } x \in I_j \\ &\vdots \\ g(x) &= g_{k+1}(x) \quad \text{pro } x \in I_{k+1}. \end{aligned}$$

Funkce  $g_j(x)$  jsou lokálně definovány pouze na intervalech  $I_j$ . Kvalita aproximace závisí na počtu a polohách jednotlivých uzlových bodů  $\xi_j$ , typu funkcí  $g_j(x)$  a na tom, ze které třídy  $C^m$  má být aproximující funkce  $g(x)$ . Úlohu lze převést na úlohu nelineární regrese, kde se hledá počet uzlových bodů, jejich polohy a koeficienty všech lokálně definovaných funkcí  $g_j(x)$  metodou nejmenších čtverců nebo obecněji maximální věrohodnosti. Takto definovaná úloha je značně rozsáhlá, a proto se používá řada zjednodušení.



Obr. 9.16 Zadání úsekové regrese.

Obyčejně si uživatel volí počet uzlových bodů a často i jejich polohy předem na základě průběhu aproximované závislosti. Pokud jsou navíc  $g_j(x)$ ,  $j = 1, \dots, k+1$ , lineární vzhledem k parametrům (polynomy), jde vlastně o úlohu lineární regrese s omezeními, definovanými podmínkami spojitosti ve funkčních hodnotách a hodnotách derivací s ohledem na třídu  $C^m$

$$g_j^{(l)}(\xi_j) = g_{j+1}^{(l)}(\xi_j), \quad j = 1, \dots, k, \quad l = 0, \dots, m.$$

Jak bylo ukázáno v odd. 9.2, splňují podmínku spojitosti ve funkčních hodnotách a hodnotách  $m$  derivací spline polynomy  $(m+1)$  stupně  $S_{m+1}(x)$ , které jsou definovány jako polynomy maximálního stupně  $(m+1)$ . Za aproximující funkci  $g(x)$  lze použít vhodnou definici spline  $S_{m+1}(x)$  a hledat jeho koeficienty metodou nejmenších čtverců.

Pro ilustraci vyjdeme z předpokladu, že funkce  $g(x)$  je požadována ze třídy  $C^0$ . Jako  $g(x)$  použijeme vyjádření lineárního spline ve tvaru useknutého polynomu.

$$g(x) = \beta_1 + \beta_2 x + \sum_{j=1}^k \beta_{j+2} (x - \xi_j)_+^0,$$

Pokud platí aditivní model měření  $y_i = g(x_i) + g_i$ ,  $i = 1, \dots, n$ , a chyby  $g_i$  jsou

nezávislé a stejně rozdělené náhodné veličiny s konstantním rozptylem, lze získat odhady  $b_j$  parametrů  $\beta_j$ ,  $j = 1, \dots, k+1$  minimalizací kritéria metody nejmenších čtverců

$$U = \sum_{i=1}^n [y_i - g(x_i)]^2. \text{ Při znalosti počtu a poloh uzlových bodů } \xi_j \text{ jde o úlohu lineární}$$

regrese. Derivací kritéria  $U$  podle jednotlivých parametrů lze dospět k soustavě normálních rovnic  $\mathbf{M} \mathbf{b} = \mathbf{Z}$ . Soustava představuje  $(k + 2)$  lineárních rovnic vzhledem k hledaným  $b_1, \dots, b_{k+2}$ . Struktura matice  $\mathbf{M}$  a vektoru  $\mathbf{Z}$  je však ovlivněna speciálním typem funkce  $g(x)$ . První řádek matice  $\mathbf{M}$  má tvar

$$[n \quad x_1 \quad x_1^2 \quad (x_1 \& \xi_1)\% \quad \dots \quad (x_1 \& \xi_k)\%] ,$$

a druhý řádek je  $[x_1 \quad x_1^2 \quad x_1(x_1 \& \xi_1)\% \quad \dots \quad x_1(x_1 \& \xi_k)\%]$ .

V dalších řádcích má obecně první prvek tvar

$$M_{j\%2,1} = (x_j \& \xi_j)\% , \quad j = 1, \dots, k ,$$

druhý prvek má tvar  $M_{j\%2,2} = x_j(x_j \& \xi_j)\%$ ,  $j = 1, \dots, k$ .

Další prvky  $M_{l+2,j+2}$  jsou

$$M_{l+2,j+2} = (x_j \& \xi_j)\% (x_l \& \xi_l)\% , \quad j = 1, \dots, k , \quad l = 1, \dots, k .$$

Vektor  $\mathbf{Z}$  má složky  $\mathbf{Z} = [y_1 \quad y_1 x_1 \quad y_1(x_1 \& \xi_1)\% \quad \dots \quad y_1(x_1 \& \xi_k)\%]^T$ .

Jednotlivé složky matice  $\mathbf{M}$  a vektoru  $\mathbf{Z}$  jsou ovlivněny také tím, že se pracuje s useknutými polynomy. Pro zlepšení numerické stability se doporučuje transformace souřadnic  $x$  do intervalu  $[1, 2]$ . Pokud se požaduje aproximace ze třídy  $C^1$ , lze volit kvadratický spline a pro aproximaci ze třídy  $C^2$  kubický spline atd. Pro všechny modely tohoto typu lze sestavit matici  $\mathbf{M}$  i vektor  $\mathbf{Z}$  a nalézt odhady parametrů  $\beta_1, \dots, \beta_{m+k+1}$  aproximačního spline  $S_m(x)$ . Z numerického hlediska však není použití reprezentace spline ve tvaru useknutých polynomů příliš vhodné, protože pro větší počet uzlových bodů je matice  $\mathbf{M}$  špatně podmíněná. Výhodnější je použití  $B$ -spline reprezentace.

Demonstrujme si použití této reprezentace spline na příkladu, kdy má být  $g(x)$  ze třídy  $C^0$ . Při použití lineárních  $B$ -spline je třeba definovat ještě přídavné body  $\xi_0$  a  $\xi_{k+1}$  (viz obr. 9.16). Aproximující model má pak tvar

$$g(x) = \sum_{j=1}^{k\%2} b_j B_{2,j}(x) ,$$

kde  $B_{2,j}(x)$  jsou konkrétně definovány ve vzorové úloze 9.6 a zakresleny na obr. 9.7. Zavedme zkrácené označení  $N_{j,i} = B_{2,j}(x_i)$ . Po dosazení  $g(x)$  do kritéria  $U$  a analytické minimalizaci dospějeme opět k soustavě rovnic  $\mathbf{M} \mathbf{b} = \mathbf{Z}$ . V tomto případě má vektor  $\mathbf{Z}$  složky

$$Z_j = \sum_{i=1}^n y_i N_{j,i} , \quad j = 1, \dots, k\%2 .$$

Matice  $\mathbf{M}$  má vzhledem k lokální definovanosti lineárních  $B$ -spline tridiagonální strukturu. Pro její první řádek platí  $M_{1,1} = \sum N_{1,i}^2$ ,  $M_{1,2} = \sum N_{1,i} N_{2,i}$  a  $M_{1,j} = 0$ ,  $j = 3, \dots, k + 2$ . V  $j$ -tém řádku jsou nenulové pouze diagonální a první poddiagonální resp. naddiagonální prvky, pro

kteřé platí  $M_{j,j} = \sum N_{j,i}^2$ ,  $M_{j,j+1} = \sum N_{j,i} N_{j+1,i}$ ,  $M_{j,j-1} = \sum N_{j,i} N_{j-1,i}$ . Konečně v posledním  $(k+2)$ . řádku jsou nenulové pouze poslední dva prvky

$$M_{k+2,k+1} \quad \dots \quad j \quad N_{k+2,i} \quad N_{k+1,i} \quad \text{a} \quad M_{k+2,k} \quad \dots \quad j \quad N_{k+2,i}^2 \quad \dots$$

Tridiagonální soustava rovnic se dá řešit kompaktními algoritmy. Pro případ  $C^1$ -aproximace vede použití kvadratických  $B$ -spline (viz obr. 9.8) k matici  $M$  s pětidiagonální strukturou. Příklad  $C^2$ -aproximace vede při použití kubických  $B$ -spline (viz obr. 9.9) k matici  $M$  se sedmidiagonální strukturou. Také při volbě  $C^m$ -aproximace pro větší  $m$  je výhodné použití  $B$ -spline reprezentace. Přehled dalších možností použití spline regrese a způsob statistické analýzy těchto speciálních lineárních modelů je popsán v Eubankově práci<sup>20</sup>.

Samostatným problémem je volba uzlových bodů  $\xi_j$ . V programu ADSTAT je možné vybrat mezi 4 alternativami:

- (a) konstantním dělením uzlových bodů,
- (b) umístěním uzlových bodů tak, aby v každém intervalu  $I_j$  byl stejný počet experimentálních bodů,
- (c) volbou poloh uzlových bodů uživatelem,
- (d) hledáním uzlového bodu programem, a to regresní optimalizací.

**Volba uzlových bodů:** Při volbě uzlových bodů uživatelem lze v případě kubické spline regrese, kdy je aproximační funkce ze třídy  $C_2$ , použít následující rámcová pravidla<sup>21</sup>:

I. Nejvhodnější je volit co nejméně uzlových bodů s tím, že v každém intervalu  $I_j$  by mělo být nejméně 4 až 5 bodů.

II. V intervalu  $I_j$  by měl být maximálně jeden extrém (minimum nebo maximum) a jeden inflexní bod.

III. Pokud je v  $I_j$  extrém, měl by ležet přibližně uprostřed.

IV. Pokud je v intervalu  $I_j$  inflexní bod, měl by ležet v blízkosti uzlového bodu.

Jistou nevýhodou modelů ve tvaru spline polynomů je při hledání vhodného počtu a poloh uzlových bodů opakované řešení metody nejmenších čtverců. Tyto modely jsou vhodné při interaktivní práci s obrazovkou počítače, kde uživatel snadno generuje různé způsoby rozmístění uzlových bodů tak, aby byl s výsledkem spokojen. V některých případech se však požaduje jednodušší metoda, kdy se jednotlivé uzlové body zařazují postupně podle zvoleného statistického kritéria: v těchto případech se volí jak třída funkcí  $C^m$ , tak i kritérium regrese. Demonstrujme si takový postup na příkladu, kdy má být aproximující funkce ze třídy  $C^2$  a jsou splněny podmínky pro užití metody nejmenších čtverců. Pro úsekovou regresi se pak používá polynomů stupně  $(m+3)$ , tj. pátého.

Pro zvolené  $\xi_j$  lze odpovídající polynom  $p_1(x)$  vyjádřit ve tvaru

$$p_1(x) = \sum_{k=1}^6 \beta_k (x - \xi_1)^{k-1}$$

Pro odhad parametrů  $\beta_1$  až  $\beta_6$  lze použít metody nejmenších čtverců, což vede k řešení soustavy rovnic  $\beta_1 = H_1^{-1} Z_1$ , kde vektor  $Z_1$  má složky

$$Z_j = \sum_{x_i \in I_j} y_i (x_i - \xi_1)^{j-1}, \quad j = 1, \dots, 6$$

a matice  $H_1$  má prvky

$$H_{lj} = \prod_{x_i \in I_1} (x_i - \xi_1)^{l-1} (x_i - \xi_1)^{j-1}, \quad l = 1, \dots, 6, \quad j = 1, \dots, 6.$$

Při konstrukci ostatních polynomů je třeba vzít v úvahu omezení plynoucí z požadavku spojitosti ve funkčních hodnotách a hodnotách prvních dvou derivací v uzlových bodech. Polynom  $p_k(x)$  pro interval  $(\xi_{k-1}, x, \xi_k) \in I_k$  je nutné vyjádřit ve tvaru

$$p_k(x) = \sum_{l=0}^2 \frac{(x - \xi_{k+1})^l}{l!} \frac{d^l p_{k+1}(x)}{dx^l} + \sum_{r=4}^6 \beta_r (x - \xi_{j+1})^{r-1}.$$

V této rovnici jsou pouze tři neznámé parametry  $\beta_4, \beta_5, \beta_6$ . Parametry lze určit metodou nejmenších čtverců. Označme první sčítance této rovnice jako  $K(x)$ . Formálně lze pak odhad

parametrů  $b_k$  vyjádřit ve tvaru  $b_k = H_k^{-1} Z_k$ , kde  $b_k$  má nyní pouze tři složky.

Vektor  $Z_k$  má prvky

$$Z_j = \prod_{x_i \in I_k} [y_i (x_i - \xi_{k+1})^j + K(x_i)], \quad j = 3, 4, 5,$$

a matice  $H_k$  má prvky  $H_{lj} = \prod_{x_i \in I_k} (x_i - \xi_{k+1})^{l-1} (x_i - \xi_{k+1})^{j-1}$  pro  $j, l = 4, 5, 6$ . Pro

výpočet koeficientů polynomu  $p_k(x)$  postačuje znalost pouze předchozích uzlů  $\xi_1, \dots, \xi_{k-1}$  a volba nového uzlu  $\xi_k$ .

K určení vhodného počtu a polohy  $\xi_k$  existuje řada postupů vycházejících z různých kritérií. Základní statistické kritérium sleduje, aby nevznikl nenáhodný trend v reziduích v intervalu  $I_j$ . V případě, že v reziduích  $e_i$  není trend, mělo by platit

$$\sum_{i=p}^{q-1} \hat{e}_i \hat{e}_{i+1} \approx \sum_{i=p}^q \frac{\hat{e}_i^2}{\sqrt{q} + p},$$

kde  $p$  je index nejmenšího a  $q$  index největšího bodu v intervalu  $I_k$ . Tedy  $x_p = \min(x)$  pro  $x \in I_k$ . Podobně lze definovat i  $x_q$ . Z dalších kritérií, popsanych v kap. 6, u výběru vhodného modelu se často používá *MEP* nebo *AIC* statistika.

## Úseková regrese programem NCSS2000

1. *Model úsekové regrese:* Úseková regrese v často užívaném software NCSS2000 je konstruována kombinací přímek a kvadratických parabol, např. úsekový polynomický regresní model *lineární-lineární* se týká modelu o dvou lineárních rovnicích, když každá platí v jiném úseku proměnné  $x$ . Modelů je celá řada: model, kde větve jsou (a) *lineární-lineární*, (b) *lineární-kvadratická*, (c) *kvadratická-lineární*, (d) *kvadratická-kvadratická*, (e) *lineární-lineární-lineární*.

2. *Uzlové body (body zvratu):* bod zvratu nemusí být uživateli znám, často bývá cílem výpočtu. Přechod jedné větve křivky do druhé v bodu zvratu může být (a) *ostrý*, (b) *vnitřní*, hladký přechod uvnitř průsečíku křivek, (c) *vnější*, hladký, ale vně průsečíku křivek.

3. *Proměnné:* proměnné  $x$  a  $y$  mohou být předem transformovány mocninnou transformací, např. závisle proměnná  $y$  je předem transformována do tvaru  $1/y^2$ ,  $1/y$ ,  $1/y^{0.5}$ ,  $\ln y$ ,  $y^{0.5}$ ,  $y^2$  a nezávisle proměnná  $x$  do tvaru  $1/x^2$ ,  $1/x$ ,  $1/x^{0.5}$ ,  $\ln x$ ,  $x^{0.5}$ ,  $x^2$ .

4. *Tabulka regresních modelů větví prokládané křivky:*

### 1. Model: lineární-lineární větve:

Regresní model:  $y = A + Bx + C(x - D) \text{ sign}(x - D)$

Rovnice:  $y = a_1 + b_1x, x < \xi$

$y = a_2 + b_2x, x > \xi$

Odhadované parametry:

$A = (a_1 + a_2)/2, \quad a_1 = A + DC, \quad a_2 = A - DC,$

$B = (b_1 + b_2)/2, \quad b_1 = B - C, \quad b_2 = B + C,$

$C = (b_2 - b_1)/2, \quad \xi = D$

$D = \xi$

### 2. Model: lineární-kvadratické větve:

Regresní model:  $y = A + Bx + Cx^2 + (x - D) \text{ sign}(x - D)[C(x + D) + E]$

Rovnice:  $y = a_1 + b_1x, x \neq \xi$

$y = a_2 + b_2x + c_2x^2, x > \xi$

Odhadované parametry:

$A = (a_1 + a_2)/2, \quad a_1 = A + DC2 + DE, \quad a_2 = A - DC2 - DE,$

$B = (b_1 + b_2)/2, \quad b_1 = B - E, \quad b_2 = B + E,$

$C = c_2/2, \quad \xi = D, \quad c_2 = 2C$

$D = \xi$

$E = (b_2 - b_1)/2$

### 3. Model: kvadratické-lineární větve:

Regresní model:  $y = A + Bx + Cx^2 + (x - D) \text{ sign}(x - D)[E - C(x + D)]$

Rovnice:  $y = a_1 + b_1x + c_1x^2, x \neq \xi$

$y = a_2 + b_2x, x > \xi$

Odhadované parametry:

$A = (a_1 + a_2)/2, \quad a_1 = A - DC2 + DE, \quad a_2 = A + DC2 - DE,$

$B = (b_1 + b_2)/2, \quad b_1 = B - E, \quad b_2 = B + E,$

$C = c_1/2, \quad \xi = D, \quad c_1 = 2C$

$D = \xi$

$E = (b_2 - b_1)/2$

### 4. Model: kvadratické-kvadratické větve:

Regresní model:  $y = A + Bx + Cx^2 + (x - D) \text{ sign}(x - D)[E(x + D) + F]$

Rovnice:  $y = a_1 + b_1x + c_1x^2, x \neq \xi$

$y = a_2 + b_2x + c_2x^2, x > \xi$

Odhadované parametry:

$$A = (a_1 + a_2)/2, \quad a_1 = A - E D 2 + D F, \quad a_2 = A + E D 2 - D F,$$

$$B = (b_1 + b_2)/2, \quad b_1 = B - F, \quad b_2 = B + F,$$

$$C = (c_1 + c_2)/2, \quad \zeta = D$$

$$D = \zeta \quad c_1 = C - E \quad c_2 = C + E$$

$$E = (c_2 - c_1)/2$$

$$F = (b_2 - b_1)/2$$

##### 5. Model: lineární-lineární-lineární větve:

Regresní model:  $y = A + Bx + C(x - D) \operatorname{sign}(x - D) + E(x - F) \operatorname{sign}(x - F)$

Rovnice:  $y = a_1 + b_1 x, x < \xi_1$

$$y = a_2 + b_2 x, \xi_1 < x < \xi_2$$

$$y = a_3 + b_3 x, x > \xi_2$$

Odhadované parametry:

$$A = (a_1 + a_3)/2, \quad a_1 = A + DC + EF, \quad a_2 = A - DC - EF, \quad a_3 = A - DC + EF,$$

$$B = (b_1 + b_3)/2, \quad b_1 = B - C - E, \quad b_2 = B + C - E, \quad b_3 = B + C + E$$

$$C = (b_2 - b_1)/2, \quad \xi_1 = D, \quad \xi_2 = F$$

$$D = \xi_1$$

$$E = (b_3 - b_2)/2$$

$$F = \xi_2$$

## 9.5 Numerické vyhlazování

Účelem numerického vyhlazování je odstranění náhodných šumů  $g$ . Požaduje se, aby vyhlazující funkce  $g(x)$  měla pouze obecné vlastnosti, jako je spojitost ve zvoleném počtu derivací. Nejde v pravém slova smyslu o nalezení konkrétního funkčního tvaru aproximující funkce  $g(x)$ , ale o nalezení *rekonstruované bezšumové závislosti* především k zobrazení, numerické derivaci a integraci. Pro numerické vyhlazení lze použít především *spline vyhlazování*, pro které je charakteristické, že uzlové body  $\xi_i$  jsou totožné s souřadnicemi  $x$  zadaných experimentálních dat  $\{x_i, y_i\}$   $i = 1, \dots, n$ . S vyhlazujícími spline úzce souvisí tzv. *neparametrická regrese*, kdy funkce  $g(x)$  je vhodně vážená lineární kombinace veličin  $y_i$ ,  $i = 1, \dots, n$ , s vahami závislými na vzdálenostech  $x - x_i$ .

V řadě případů postačuje pouze nalezení posloupnosti vyhlazených hodnot  $g(x_i)$  z původních hodnot  $y_i$ . Pro ekvidistantní dělení experimentálních bodů na ose  $x$ , kdy  $h_i = x_{i+1} - x_i = \text{konst.}$ ,  $i = 1, \dots, n - 1$ , jde o úlohu *číslicové filtrace*. Ta je vhodná pro předzpracování signálů z různých měřicích přístrojů. V technické praxi je výhodné využití těchto postupů všude tam, kde nezávisí na typu vyhlazovací funkce  $g(x)$ , a kde je třeba získat závislost s odstraněnou šumovou složkou (chybami  $g$ ). Často jde o úlohy derivace nebo integrace dat zatížených náhodnými chybami.

### 9.5.1 Spline vyhlazování

Při konstrukci vyhlazujících spline se vychází z požadavku, aby se funkce  $g(x)$  přibližovala co nejvíce k experimentálním datům, a to ve smyslu zvolené  $L_p$ -normy (kap. 6). Jsou-li chyby  $g$  nezávislé a stejně rozdělené náhodné veličiny s konstantním rozptylem, je vhodné volit  $L_2$ -normu, vedoucí ke kritériu metody nejmenších čtverců

$$U(g) = \sum_{i=1}^n w_i [y_i - g(x_i)]^2,$$

kde  $w_i$  jsou váhy jednotlivých bodů, závislé na jejich "přesnosti" vyjádřené např. přes rozptyly. Dalším požadavkem je, aby vyhlazující funkce byla dostatečně hladká a spojitá ve zvoleném počtu derivací. Omezme se na nejčastější případ, kdy se požaduje, aby  $g(x)$  byla dvakrát diferencovatelná, tzn. ze třídy  $C^2[a, b]$ , kde  $a = x_1$  a  $b = x_n$ . Jak bylo uvedeno v odd. 9.2, je možno kritérium hladkosti vyjádřit integrálem

$$I(g) = \int_a^b [g^{(2)}(x)]^2 dx,$$

kde  $g^{(2)}(x)$  je druhá derivace vyhlazující funkce. Integrál  $I(g)$  souvisí s normou druhé derivace a označuje se jako *míra hladkosti v křivosti funkce*  $g(x)$ . Účelem je nalézt takovou funkci  $g(x)$ , která by měla dostatečně malou hodnotu  $U(g)$ , tj. byla v blízkosti experimentálních dat a přitom měla malou hodnotu  $I(g)$ , tj. byla dostatečně hladká a její průběh by neměl být nadměrně zvlněný. Pro stanovení optimální vyhlazující funkce  $g(x)$  můžeme sestavit dvě základní minimalizační úlohy:

I. Jde o minimalizaci modifikovaného součtu čtverců odchylek

$$K_1 = U(g) + \alpha I(g),$$

kde  $0 < \alpha < 4$  je *parametr vyhlazení*, který "řídí" poměr mezi hladkostí  $g(x)$  a jejím přiblížením k experimentálním bodům.

II. Při splnění podmínky  $U(g) = S$  se hledá vyhlazující funkce  $g(x)$  minimalizující integrál  $I(g)$ . Pro dané  $S$  existuje takový *parametr vyhlazení*  $\alpha = \alpha(S)$ , že řešení úlohy I je zároveň i řešením úlohy II. Důvodem zavedení úlohy II je fakt, že parametr  $S$  má význam reziduálního součtu čtverců a souvisí přímo s odhadem rozptylu náhodných chyb  $g$ .

V řadě prací bylo odvozeno, že funkce  $g(x)$  ze třídy  $C^2[a, b]$ , která pro dané  $\alpha$  minimalizuje  $K_1$  z výše uvedené rovnice, má následující vlastnosti<sup>22</sup>:

1. Funkce  $g(x)$  je na každém intervalu  $I_j \in [x_{i+1}, x_i]$  polynomem třetího stupně.

2. Ve všech místech  $x_i$  čili uzlových bodech je funkce  $g(x)$  spojitá ve funkčních hodnotách a hodnotách prvních dvou derivací, což se zapíše rovnicí

$$g^{(k)}(x_i^+) = g^{(k)}(x_i^-), \quad i = 2, \dots, n-1, \quad k = 0, 1, 2,$$

kde  $g(x_i^+)$  je limita v bodě  $x_i$  zprava a  $g(x_i^-)$  je limita v bodě  $x_i$  zleva.

3. Ve třetí derivaci je vyhlazující funkce nespojitá a platí pro ni, že

$$g^{(3)}(x_i^-) - g^{(3)}(x_i^+) = \frac{w_i}{\alpha} [y_i - g(x_i)].$$



4. V rozmezí  $(-4, a]$  a  $[b, 4)$  jsou druhé derivace  $g^{(2)}(x) = 0$ . To znamená, že funkce  $g(x)$  je mimo interval  $(a, b)$  lineární.

Všechny funkce vyhovující těmto podmínkám jsou kubické splíny  $S_3(x)$  s uzly  $x_i$ , které jsou při znalosti parametru vyhlazení  $\alpha$  jednoznačně určeny podmínkou 3, nahrazující klasickou podmínku interpolace u interpolačních splíny. Pokud se v rovnici  $I(g)$  nahradí druhá derivace  $m$ -tou derivací, vychází jako optimální  $g(x)$  ve třídě  $C^{2m-2}[a, b]$  polynomický splín  $S_{2m-1}(x)$  stupně  $(2m-1)$ . Podmínka pak platí pro  $(2k-1)$ . derivaci.

Místo kritéria metody nejmenších čtverců  $U(g)$  lze použít i jiná kritéria. Pokud se místo čtverců použije pomaleji rostoucí funkce (viz kap. 6), rezultují *robustní vyhlazující splíny*. Podobně jako při splíny interpolaci, lze i zde řídit hladkost vyhlazující funkce zavedením *parametrů napětí*  $\rho_i$ . Pro vyhlazující splíny pod napětím lze místo rovnice z podmínky 3 psát

$$[g^{(3)}(x_i) \& \rho_i g^{(1)}(x_i)] \& [g^{(3)}(x_i) \& \rho_i g^{(1)}(x_i)] \quad , \\ , \quad \frac{w_i}{\alpha} [y_i \& g(x_i)].$$

Detailní postup konstrukce vyhlazujících splíny pod napětím je uveden v práci<sup>23</sup>.

Ukažme postup konstrukce kubického vyhlazujícího splíny, jež minimalizuje rovnici  $K_1$  při známé hodnotě *parametru vyhlazení*  $\alpha$ . Nejjednodušší je hledat vyhlazené funkční hodnoty  $g(x_i) = g_i$  a druhé derivace  $g^{(2)}(x_i) = M_i$ . Při znalosti těchto hodnot je možné určit koeficienty kubických polynomů, které procházejí body  $\{x_i, g_i\}$ ,  $i = 1, \dots, n$ . V odd. 9.2 bylo ukázáno, že pro druhé derivace  $M_i$  u klasických interpolačních splíny platí soustava lineárních rovnic

$$M_1 = 0, \quad M_n = 0, \\ \frac{h_{j+1}}{6} M_{j+1} + \frac{h_j}{3} M_j + \frac{h_j}{6} M_{j-1} = \\ = \frac{g_{j+1}}{h_j} + g_j \left( \frac{1}{h_j} + \frac{1}{h_{j+1}} \right) + \frac{g_{j-1}}{h_{j+1}}, \quad j = 2, \dots, n-1.$$

V maticovém zápisu má tato soustava tvar  $AM = Dg$ , kde matice  $A$  i matice  $D$  jsou tridiagonální a vektor  $g = (g_1, \dots, g_n)$  je vektor vyhlazených hodnot. Matice  $A$  má prvky

$$A_{1,1} = A_{n,n} = 1, \quad A_{i,i+1} = \frac{h_{i+1}}{6}, \quad A_{i,i} = \frac{h_i + h_{i+1}}{3}, \quad A_{i,i-1} = \frac{h_i}{6}.$$

Ostatní prvky této matice jsou nulové. Matice  $D$  má první a poslední řádek složen ze samých nul. Dále je

$$D_{i,i+1} = \frac{1}{h_{i+1}}, \quad D_{i,i} = \left( \frac{1}{h_i} + \frac{1}{h_{i+1}} \right), \quad D_{i,i-1} = \frac{1}{h_i}.$$

Ostatní prvky této matice jsou opět nulové. S využitím faktu, že třetí derivace kubického spline v intervalu  $I_j$  je rovna

$$g^{(3)}(x) = \frac{M_{i-1} & M_i}{h_i},$$

je možno podmínku 3. vyjádřit ve tvaru  $\mathbf{g} = \mathbf{y} + \mathbf{v} \mathbf{D}^T \mathbf{M}$ , kde  $\mathbf{v} = (\alpha/w_1, \dots, \alpha/w_p, \dots, \alpha/w_n)^T$  je vektor parametrů vyhlazení a  $\mathbf{y} = (y_1, \dots, y_n)^T$  je vektor původních nevyhlazených hodnot. Po dosazení za  $\mathbf{g}$  můžeme nalézt vektor druhých derivací  $\mathbf{M}$  ze soustavy lineárních rovnic

$$(\mathbf{A} + \mathbf{D} \mathbf{v} \mathbf{D}^T) \mathbf{M} = \mathbf{D} \mathbf{y}.$$

Soustava je pětidiagonální a má jednoznačné řešení. K určení  $\mathbf{M}$  lze použít kompaktních algoritmů podobně jako u spline interpolace. Při znalosti hodnot druhých derivací  $\mathbf{M}$  lze určit vyhlazené hodnoty  $\mathbf{g}$  přímo dosazením dle vztahu

$$g_i = y_i + v_i L_i, \quad i = 1, \dots, n,$$

kde 
$$L_1 = \frac{M_2 + M_1}{h_1}, \quad L_n = \frac{M_n + M_{n+1}}{h_{n+1}}$$

a 
$$L_i = \frac{1}{h_i} (M_{i-1} + M_i) + \frac{1}{h_{i+1}} (M_i + M_{i+1}), \quad i = 2, \dots, n-1.$$

Z této rovnice plyne, že vyhlazující spline je lineární odhad, protože platí

$$\mathbf{g} = \mathbf{H}(\alpha) \mathbf{y}.$$

Matice  $\mathbf{H}(\alpha)$  vyjde dle vztahu

$$\mathbf{H}(\alpha) = [\mathbf{E} + \mathbf{v} \mathbf{D}^T \mathbf{A}^{-1} \mathbf{D}]^{-1}.$$

Reinsch<sup>25</sup> určil matici  $\mathbf{E} - \mathbf{H}(\alpha)$  v kompaktním tvaru

$$\mathbf{E} - \mathbf{H}(\alpha) = \mathbf{Q} (\mathbf{Q}^T \mathbf{Q} + p \mathbf{T})^{-1} \mathbf{Q}^T,$$

kde  $\mathbf{Q}$  je  $n \times (n-2)$  dimenzionální tridiagonální matice, která vznikne vynecháním prvního a posledního sloupce matice  $\mathbf{D}^T$ . Matice  $\mathbf{T}$  je  $(n-2) \times (n-2)$  dimenzionální tridiagonální matice, která vznikne vynecháním prvního a posledního sloupce i řádku matice  $\mathbf{A}$  a vynásobením ostatních prvků matice  $\mathbf{A}$  dvěma. Parametr  $p = 1/\alpha$  platí pro případ stejných vah  $w_i = 1$ ,  $i = 1, \dots, n$ . Matice  $\mathbf{H}(\alpha)$  má řadu vlastností shodných s projekční maticí  $\mathbf{H}$ , definovanou v kap. 6. Pro její diagonální prvky platí dle Eubank<sup>26</sup>  $0 \leq H_{ii}(\alpha) \leq 1$

a mimodiagonální prvky jsou

$$|H_{ij}(\alpha)| \leq 1 \quad \text{pro } j \neq i.$$

Navíc platí, že  $\sum_{j=1}^n H_{ij}(\alpha) = 1$ . Podobně jako u klasické projekční matice, je i zde  $H_{ii}(\alpha)$

$= 1$ , pokud  $H_{ij} = 0$  pro všechna  $i \neq j$ . Chování matice  $\mathbf{H}(\alpha)$  souvisí úzce s projekční maticí  $\mathbf{H}^*$  pro regresní přímkou (viz kap. 6). Projekční matice  $\mathbf{H}^*$  závisí pouze na hodnotách  $x_i$ ,  $i = 1, \dots, n$ . Platí, že

A. Pro  $\alpha \in \mathbb{R}$ :  $H_{ii}(\alpha) \in [0, 1]$  a  $H_{ij}(\alpha) \geq 0$ . Dále plyne, že v tomto případě jsou  $y_i = g(x_i)$ . Vyhlažující funkce  $g(x)$  je totožná s klasickým interpolačním kubickým spline  $S_3(x)$ , který je nejhladší.

B. Pro  $\alpha \in \mathbb{R}^4$ :  $H_{ii}(\alpha) \in [0, 1]$  a  $H_{ij}(\alpha) \geq 0$ . Vyhlažující funkce  $g(x)$  je totožná s regresní přímkou aproximující experimentální body ve smyslu nejmenších čtverců odchylek. Pro případ, že se použije  $m$ -tá derivace a  $\alpha \in \mathbb{R}^4$ , je výsledkem regresní polynom stupně  $(m - 1)$ . Vyhlažující funkce se proto označují jako *zobecněné polynomicke regresní modely*.

**Späthův algoritmus.** Späth<sup>24</sup> použil při konstrukci algoritmu pro vyhlažující kubický spline postup, který vychází z rovnice pro  $\mathbf{g}$ . Vyhlažující spline vyjádřil ve tvaru lokálních kubických polynomů a k řešení pětidiagonální soustavy lineárních rovnic využil kompaktní varianty Choleského metody. V programu SPÄTH jsou použity *lokální parametry vyhlazení*  $\beta_i = w_i/\alpha_i$ , takže platí

(a) pro  $\beta_i \in \mathbb{R}$ ,  $i = 1, \dots, n$ , je potlačena podmínka hladkosti a rezultuje funkce  $g(x)$  jako regresní přímka;

(b) pro  $\beta_i \in \mathbb{R}^4$  prochází vyhlažující spline bodem  $\{x_i, y_i\}$ . Pokud je  $\beta_i \in \mathbb{R}^4$ ,  $i = 1, \dots, n$ , rezultuje funkce  $g(x)$  jako kubický interpolační spline  $S_3(x)$ .

Volbou  $\beta_i$  lze proto řídit lokální přiblížení vyhlažující funkce k experimentálním bodům.

**Reinschův algoritmus:** Reinsch<sup>25</sup> řešil minimalizaci  $I(g)$  za podmínky  $U(g) = S$ , a to využitím metody Lagrangeových multiplikátorů, což vede k minimalizaci funkcionálu

$$K_2 = I(g) + p(U(g) - S),$$

kde  $p$  je Lagrangeův multiplikátor a  $Z$  je pomocná proměnná. Minimalizace funkcionálu  $K_2$  vede ke kubickému vyhlažujícímu spline, což je pro známé  $p$  úloha hledání řešení soustavy lineárních rovnic s pětidiagonální maticí koeficientů. Optimální  $p$  pro zadané  $S$  se hledá iterativním řešením nelineární rovnice Newtonovou metodou. Přesto, že je tento algoritmus komplikovanější, je v praxi rozšířenější. V programu Reinsche<sup>25</sup> se vedle hodnoty  $S$  zadávají i váhy jednotlivých bodů  $w_i$ . Platí, že

- čím je  $S$  větší, tím více se vyhlažující funkce  $g(x)$  blíží k regresní přímce,
- čím jsou váhy  $w_i$  větší, tím více se vyhlažující funkce  $g(x)$  blíží k experimentálním bodům.

**Volba parametru vyhlazení  $\alpha$ .** Samostatným problémem je volba parametru vyhlazení  $\alpha$  a parametru  $S$  s ohledem na to, aby ve zvoleném statistickém smyslu vyhlazující funkce  $g(x)$  co nejlépe aproximovala experimentální data. Jsou-li vhodně vybrány váhy  $w_i$ , jež odpovídají reciprokým hodnotám rozptylů v jednotlivých bodech, je možno volit parametr  $S$  v intervalu

$$(n \% 1) \& \sqrt{2 (n \% 1)} \# S \# (n \% 1) \% \sqrt{2 (n \% 1)} .$$

Dobré výsledky poskytuje volba  $S = n + 1$ . K určení optimálního parametru  $\alpha$  je nejčastěji používána střední kvadratická chyba predikce  $MEP(\alpha)$ , která je definována vztahem

$$MEP(\alpha) = \frac{1}{n} \sum_{i=1}^n \frac{(y_i - g(x_i))^2}{(1 + H_{ii}(\alpha))^2} .$$

Místo kritéria  $MEP(\alpha)$  lze užít zobecněnou střední kvadratickou chybu predikce  $CEP(\alpha)$ , kde se nahrazuje  $H_{ii}(\alpha)$  střední hodnotou

$$T(\alpha) = \frac{1}{n} \sum_{i=1}^n H_{ii}(\alpha) = \frac{1}{n} \text{Tr}(\mathbf{H}(\alpha)) ,$$

kde  $\text{Tr}(\cdot)$  je stopa matice. Rovnice pro  $MEP(\alpha)$  pak přechází na tvar

$$CEP(\alpha) = \frac{1}{n} \frac{\sum_{i=1}^n (y_i - g(x_i))^2}{(1 + T(\alpha))^2} ,$$

Optimální parametr vyhlazení  $\alpha$  je pak takový, pro který nabývá  $CEP(\alpha)$  své minimální hodnoty. S využitím této rovnice lze kritérium  $CEP(\alpha)$  vyjádřit pouze jako funkci hodnot  $\{x_i, y_i\}$ ,  $i = 1, \dots, n$ , ve tvaru

$$CEP(\alpha) = \frac{1}{n} \frac{\sum_{i=1}^n (y_i - g(x_i))^2}{\left[1 + \frac{1}{n} \text{Tr}(\mathbf{H}(\alpha))\right]^2} .$$

Postačuje nalézt pouze matice  $\mathbf{H}(\alpha)$ , protože čítec v této rovnici je reziduální součet čtverců odchylek  $e_i = y_i - g(x_i)$ , který lze pro daný parametr vyhlazení  $\alpha$  snadno vyčíslit. Efektivní postup vyčíslení stopy matice  $\mathbf{E} - \mathbf{H}(\alpha)$  je popsán v Hutchinsonových a de Hoogových pracích<sup>27,28</sup>. Při ekvidistantním dělení bodů na ose  $x$  je možno  $T(\alpha)$  vyjádřit ve tvaru

$$T(\alpha) = \frac{1}{n} \sum_{i=1}^n (1 + \alpha \lambda_i)^{-1} .$$

Konstanty  $\lambda_i$  lze s minimální ztrátou přesnosti vypočítat ze vztahů

$$\lambda_1 = \lambda_2 = 0 ,$$

$$\lambda_i = \left(\frac{\pi}{n}\right)^4 \frac{(i + 1.5)^4}{h^3}, \quad i = 3, 4, \dots, n .$$

Při znalosti  $T(\alpha)$  lze vypočítat i  $CEP(\alpha)$  a vhodnou numerickou metodou hledat jeho minimum. Zobecnění tohoto postupu pro libovolné dělení bodů na ose  $x$  navrhl Silverman<sup>29</sup>. Veličina  $T(\alpha)$  se zde určuje podle vztahu

$$T(\alpha) = \frac{2}{n} \sum_{j=3}^n \left[ 1 - \frac{\pi^4}{n} \alpha (j - 1.5)^4 c_0 \right]^{j-1},$$

kde konstanta  $c_0$  se počítá z přibližného vzorce

$$c_0 = \left[ \int_a^b \hat{f}^{1/4}(t) dt \right]^{8/4} \cdot \left[ \frac{b - a}{32} \sum_{j=1}^{32} \hat{f}^{1/4}(x_j^{(j)}) \right]^{8/4}.$$

Zde  $\hat{f}(t)$  je odhad hustoty pravděpodobnosti, určený z hodnot  $x_i$ ,  $i = 1, \dots, n$ . Postup lze formulovat ve dvou krocích:

(1) Určí se hodnoty  $x_j^{(j)} = a + \frac{(b - a)(j - 0.5)}{32}$ ,  $j = 1, \dots, 32$ . Mezní hodnoty

$[a, b]$  se počítají podle vztahů  $a = x_n - \frac{x_n - x_1}{n}$ ,  $b = x_n + \frac{x_n - x_1}{n}$ .

(2) Vypočtou se odhady hustoty pravděpodobnosti

$$\hat{f}(x_j) = \frac{1}{n \Delta \sqrt{2\pi}} \prod_{i=1}^n \exp\left[-0.5 \left(\frac{x_j - x_i}{\Delta}\right)^2\right].$$

Parametr  $\Delta$  určuje hladkost odhadu hustoty pravděpodobnosti. Pro praktické případy postačuje volba  $\Delta = 1.06 n^{-1/5} s$ , kde  $s$  je směrodatná odchylka počítaná z hodnot  $x_i$ ,  $i = 1, \dots, n$ . Uvedený postup je sice přibližný, ale pro praktické účely postačuje. Parametr  $\alpha$  nezávisí na  $\alpha$  a lze jej určit pouze jednou. V Silvermanově práci<sup>29</sup> je ukázáno, že při volbě  $T(\alpha)$  vychází  $CEP(\alpha)$  větší než při použití přesného vztahu, rozdíl je však výraznější pouze pro malé  $\alpha$ . V jiné Silvermanově práci<sup>31</sup> je uvedeno, jak rozšířit tento přístup i na případ nekonzstantních vah  $w_i$ . Využitím aproximace prvků  $H_{ii}(\alpha)$  ve tvaru<sup>31</sup>

$$H_{ii}(\alpha) = \alpha^{8/4} n^{8/4} 2^{8/2} \hat{f}^{8/4}(x_i)$$

je možné konstruovat i *přibližné pásy spolehlivosti predikce*. Pro *95% pásy spolehlivosti* platí

$$L_{1,2}(x_i) = g(x_i) \pm 1.96 \hat{\sigma} \sqrt{H_{ii}(\alpha)}.$$

Tuto rovnici lze použít pro konstrukci pásů spolehlivosti pro libovolné  $x$ . Vlastně to znamená vyčíslit pouze  $\hat{f}(x)$ . Zbývá ještě nalézt *odhad rozptylu*  $\hat{\sigma}^2$ . Byl navržen vztah

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (y_i - g(x_i))^2}{n(1 - T(\alpha))},$$

kde  $n$  a  $T(\alpha)$  jsou analogicky jako u lineární regrese stupně volnosti, odpovídající vyhlazujícímu spline. Pro určení  $\hat{\sigma}^2$  se dosazuje  $\alpha$ , jež minimalizuje kritérium  $CEP(\alpha)$ . Veličina  $T(\alpha)$  se pak vyčíslí.

## 9.5.2 Neparametrická regrese

*Vyhlazující spline* je lineární kombinací všech měření. Existuje taková váhová funkce  $G_\alpha(x, x_i)$ , pro kterou je

$$g(x) = \frac{1}{n} \sum_{i=1}^n y_i G_\alpha(x, x_i).$$

Váhová funkce závisí na konkrétních hodnotách  $x_i$  a parametru vyhlazení  $\alpha$ . Předpokládejme, že lokální hustota souřadnic na ose  $x$  je  $f(x)$ , takže počet bodů v intervalu  $dx$  je  $f(x) dx$ . Za předpokladu, že  $\alpha$  není ani příliš velké, ani příliš malé, a  $x$  je dostatečně vzdálené od konců intervalu  $[a, b]$ , platí podle Silvermana<sup>31</sup>, že pro dostatečně velkou  $n$  je

$$G_\alpha(x, x_i) = \frac{1}{f(x_i)} \frac{1}{\delta(x_i)} K\left(\frac{x - x_i}{h(x_i)}\right).$$

Symbolem  $K(Z)$  je označena tzv. *jádrová funkce*, která má pro tento případ tvar

$$K(Z) = \frac{1}{2} \exp\left(\frac{\delta Z^*}{\sqrt{2}}\right) \sin\left(\frac{Z^*}{\sqrt{2}} \cdot 14 \pi\right).$$

Parametr  $\delta(x_i)$  určuje *lokální vyhlazení* a platí pro něj vztah

$$\delta(x_i) = \alpha^{1/4} n^{3/4} f^{3/4}(x_i).$$

Z tohoto zápisu vyhlazujícího splínu plynou následující důležité závěry<sup>31</sup>:

(a) Vliv bodu  $\{x_i, y_i\}$  se projevuje pouze na lokálním chování vyhlazující funkce  $g(x)$  pro dostatečně blízká  $x$  k hodnotě  $x_i$ .

(b) Z poslední rovnice pro  $\delta(x_i)$  plyne, že parametr  $\delta(x_i)$  je úměrný čtvrté odmocnině  $\alpha$ . Velké změny  $\alpha$  se proto příliš neprojeví na velikosti lokálního vyhlazení  $\delta(x_i)$ .

V dalším výkladu předpokládáme, že souřadnice experimentálních bodů na ose  $x$  jsou lineárně transformovány, takže  $x_1 = 0$  a  $x_n = 1$ , a dále platí  $x_{i+1} > x_i$ ,  $i = 1, \dots, n-1$ . Pro *ekvidistantně rozdělená data* se vyhlazující neparametrický regresní model vyjadřuje ve tvaru

$$p(x) = \frac{1}{n \delta} \sum_{i=1}^n y_i K\left(\frac{x - x_i}{\delta}\right).$$

2. Pro *neekvidistantně rozdělená data* se používá modifikovaný vyhlazující neparametrický regresní model

$$p(x) = \sum_{i=1}^n y_i \left(\frac{x_i - x_{i-1}}{\delta}\right) K\left(\frac{x - x_i}{\delta}\right),$$

kde jádrová funkce  $K(Z)$  musí mít tyto vlastnosti:

- (a) je nezáporná  $K(Z) \geq 0$ ,
- (b) je symetrická kolem nuly  $K(Z) = K(-Z)$ ,
- (c) má vlastnosti hustoty pravděpodobnosti, tj.

$$\int_{-\infty}^{\infty} K(Z) dZ = 1 \quad \text{a} \quad \int_{-\infty}^{\infty} K^2(Z) dZ < 4.$$

Optimální  $K(Z)$  s ohledem na minimalizaci střední kvadratické chyby predikce je ve tvaru

$$K(Z) = 0.75 (1 - Z^2)_+, \quad \text{kde } (1 - Z^2)_+ \text{ je nenulové jen pro } Z^* < 1. \text{ Přehled dalších}$$

druhů jádrových funkcí je uveden v práci Bendettiové<sup>32</sup>. K určení *optimálního parametru vyhlazení*  $\delta$ , označeného jako *šířka pásu*, lze použít jak kritéria *MEP*, tak i kritéria *CEP* nebo řady dalších<sup>33</sup>. Pro modifikovaný vyhlazující neparametrický regresní model  $p(x)$  má kritérium *střední kvadratické chyby predikce* tvar

$$MEP(\delta) = \frac{1}{n} \sum_{i=1}^n \left[ y_i \left( 1 - \frac{K(0)}{\delta n} \right) + \frac{1}{\delta n} K\left(\frac{1}{\delta n}\right) y_{i-1} + p(x_i) \right]^2.$$

### 9.5.3 Číslicová filtrace

Číslicová filtrace umožňuje průběžnou eliminaci šumové složky ve zpracovávaných signálech. V technické praxi se taková úloha vyskytuje při digitalizaci údajů ze zapisovačů u spektrofotometrů, chromatografických přístrojů, polarografů atd. Vychází se z dat  $y_i$ , měřených po ekvidistantních, a to obvykle časových nebo délkových intervalech  $s = x_{i+1} - x_i$ . Uvažuje se zde aditivní model měření

$$y_i = Z_i + g_i,$$

kde  $Z_i^*$  jsou skutečné deterministické hodnoty a  $g_i$  jsou náhodné chyby. Použitím číslicové filtrace se získá sekvence filtrovaných hodnot  $Z_i$ , které "rekonstruuje" neznámé veličiny  $Z_i^*$ :

1. Lineární číslicový filtr je možno obecně vyjádřit ve tvaru

$$Z_i = \sum_{j=0}^i c_j y_{i-j} + \sum_{j=1}^i d_j Z_{i-j},$$

kde konstanty  $c_j$  a  $d_j$  určují typ filtru.

2. Pro nerekurzivní filtry platí, že všechna  $d_j = 0$ .

3. Pokud je alespoň jedno  $d_j \neq 0$ , jde o filtr rekurzivní.

Klasické digitální filtry, které jsou náhradou analogových filtrů, jsou fyzikálně realizovatelné. Tyto filtry používají pro určení filtrovaných hodnot pouze hodnot  $y_{i-j}$  pro  $j > 0$ , které byly získány až do daného časového okamžiku  $x_i$ . Pro tyto filtry je vždy  $c_j = 0$  pro všechna  $j < 0$ . Pokud se při výpočtu filtrované hodnoty používají i "budoucí" údaje  $y_{i+k}$ ,  $k = 1, 2, \dots$ , jde o fyzikálně nerealizovatelné filtry označované jako "smoothers".

#### Nerekurzivní filtry.

(a) Z nerekurzivních filtrů pro účely předzpracování experimentálních dat doporučuje Marmet<sup>34</sup> opakované použití jednoduchého Marmetova filtru

$$Z_i = \frac{1}{4} (y_{i+1} + 2y_i + y_{i-1}),$$

(b) Dobré vyhlazovací vlastnosti má Hippeho filtr<sup>35</sup>

$$Z_i = \frac{y_{i+2} + y_{i-2} + 4(y_{i+1} + y_{i-1}) + 6y_i}{12},$$

který byl využit pro předzpracování chromatografických měření.

**Rekurzivní filtry.** Rekurzivní filtry se obvykle používají k vyhlazování časových řad a vstupů do číslicových regulátorů.

(a) Nejjednodušší je exponenciální filtr

$$Z_i = K y_i + (1 - K) Z_{i-1},$$

kde  $K$  je stupeň zesílení filtru  $0 < K < 1$ .

(b) Mezi rekurzivní patří také dvoustupňový Holtův filtr, definovaný vztahy



$$Z_i = K q_i \% (1 + K) Z_{i+1} ,$$

$$q_i = K y_i \% (1 + K) q_{i+1} ,$$

kde  $K$  je *konstanta zesílení*. Společnou nevýhodou rekurzivních filtrů je nutnost volby parametru zesílení a dalších konstant.

**Robustní nelineární filtry.** Pro případ, kdy lze v datech očekávat i hrubé nenáhodné chyby (outliers), jsou vhodné *robustní nelineární filtry*. Jsou to varianty robustních vyhlazovacích metod<sup>36</sup>.

(a) Mezi nejjednodušší patří nelineární filtry  $L$ -typu<sup>37</sup>, založené na pohyblivých mediánech. Medián  $S(v, i)$  lichého stupně je definován vztahem

$$S(v, i) = \text{med}(y_{i+u}, \dots, y_i, \dots, y_{i-u}) ,$$

kde  $u = (v - 1)/2$  a symbol  $\text{med}(\cdot)$  označuje střed podle velikosti seříděných hodnot  $y$ . Užívá se mediánu třetího stupně ( $v = 3$ ) a pátého stupně ( $v = 5$ ). Mediány lichého stupně lze kombinovat s pohyblivými aritmetickými průměry.

(b) *Jednoduchý filtr 53H* je dán výrazem

$$Z_i = \frac{S(5, i+2)}{4} \% \frac{S(5, i+1)}{2} \% \frac{S(5, i)}{4} .$$

K zajištění dokonalejšího vyhlazení se mediány používají opakovaně.

(c) Z této skupiny je nejjednodušší *filtr 3T*, pro který je

$$Z_i = \text{med}[S(3, i+2), S(3, i+1), S(3, i)] .$$

Další varianty pohyblivých mediánů obsahuje Vellemannovy práce<sup>38</sup>. Na základě simulační studie bylo zjištěno, že mezi nejvhodnější patří *filtr 53H*, který je dostatečně robustní a přitom neposkytuje "nadměrně" vyhlazené úseky.

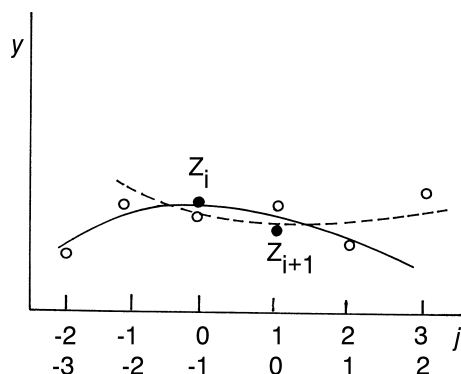
**Lineární regresní filtry Sawitzkého-Golaye.** Pozornost je věnována také lineárním regresním filtrům. Ty jsou často užívány, např. pod názvem *vyhlazení Sawitzkého-Golaye*<sup>45</sup>. Tyto filtry jsou fyzikálně nerealizovatelné a lze je vyjádřit ve tvaru

$$Z_i = \sum_{j=-N}^N c_j y_{i+j} ,$$

kde hodnota  $N$  označuje *řád filtru* a  $(2N+1)$  je *délka filtru*, která určuje počet naměřených dat  $y_{i+j}$ , jež byla užita k rekonstrukci hodnoty  $Z_i$ . *Filtr stupně  $d$*  odpovídá polynomickému regresnímu modelu stupně  $d$ . Pokud je  $d \leq 2N$ , platí, že existuje pouze jeden filtr stupně  $d$ , pro který platí, že  $c_0 = 1$  a ostatní  $c_k = 0$ . To znamená, že  $Z_i = y_i$  a nedochází potom k filtraci. V těchto případech prochází polynomický model všemi hodnotami  $y_{i+j}$ . Pro  $d < 2N$  existuje nekonečně mnoho filtrů řádu  $N$  a stupně  $d$ , a to v závislosti na konkrétních hodnotách  $(2N - d)$  nastavitelných parametrů  $c_k$ .

Za *lineární regresní filtr* pro kritérium nejmenších čtverců odchylek se označuje takový filtr, kterému odpovídá nejmenší součet čtverců koeficientů  $c_j, j = -N, \dots, 0, \dots, N$ . Výsledek

filtrace pomocí tohoto filtru odpovídá postupu, kdy je sekvence  $2N+1$  bodů  $\{j, y_j\}, j = -N, \dots, 0, \dots, N$ , proložena polynomem stupně  $d$  ve smyslu metody nejmenších čtverců a za  $Z_i$  se bere hodnota tohoto regresního polynomu v místě  $j = 0$ . Tento postup se označuje v literatuře jako *pohyblivé nejmenší čtverce* (moving least squares). Schematicky je pro  $d = 2$  (tj. regresní paraboly) a  $N = 2$  (tj. délka filtru rovna 5) znázorněna na obr. 9.32.



Obr. 9.32 Princip činnosti filtru stupně 2 a délky 5.

Pro realizaci číslicových filtrů lze přímo použít metodu nejmenších čtverců a odhadnout koeficienty regresního polynomu a nalézt predikci (vyhlazenou hodnotu)  $Z_i$ . Postup je však jednodušší, protože lze snadno určit koeficienty  $c_j$  vzhledem ke speciální volbě souřadnic  $x$ .

Pokud je  $Z_i$  polynom stupně  $d$  a chyby  $g_j$  jsou stejně rozdělené nezávislé náhodné veličiny s nulovou střední hodnotou a konstantním rozptylem  $\sigma^2$ , platí v souladu s teorií lineární regrese, že

$$E(Z_i) = Z_i \quad \text{a} \quad D(Z_i) = \sigma^2 \sum_{j=-N}^N c_j^2.$$

Výsledky  $Z_i$  lineárních regresních filtrů jsou nevyhýlené odhady s minimálním rozptylem. Hodnota  $Z_i$  odpovídá absolutnímu členu  $b_0$  regresního polynomu

$$Z_i = b_0 + \sum_{k=1}^d b_k j^k,$$

protože v místě  $i$  je  $j = 0$ . Ostatní koeficienty  $b_k, k = 1, 2, \dots, d$ , pak odpovídají hodnotám první, druhé až  $d$ -té derivace dělené faktoriálem  $1!, 2!$  až  $d!$ . S ohledem na speciální sekvenci souřadnic nezávisle proměnné  $j = -N, \dots, 0, \dots, N$  platí, že

$$\sum_{j=-N}^N j^q = 0 \quad \text{pro} \quad q = 2u + 1, \quad \text{kde} \quad u = 0, 1, 2, \dots$$

Důsledkem je, že odhad  $b_0$  pro polynom stupně  $2d$  je totožný s odhadem  $a_0$  pro polynom stupně  $(2d + 1)$ . Regresní filtr sudého stupně je zároveň regresním filtrem větším o jeden lichý stupeň. Thrall<sup>39</sup> ukázal, že pro koeficienty lineárních regresních filtrů  $c_j, j = -N, \dots, 0$ ,

...,  $N$ , stupně  $2d$  platí  $c_j = \sum_{q=0}^d \alpha_q j^{2q}$ . Vektor  $\mathbf{a} = (\alpha_0, \dots, \alpha_d)^T$  je dán řešením soustavy

rovnice  $\mathbf{H} \mathbf{a} = \mathbf{e}$ , kde  $\mathbf{e} = (1, 0, \dots, 0)^T$  a  $\mathbf{H}$  je symetrická matice rozměru  $(d+1) \times (d+1)$  s prvky

$$H_{ik} = \sum_{j=0}^N j^{2i+2k}, \quad i, k = 0, \dots, d.$$

Mezi koeficienty regresních filtrů platí řada vztahů, plynoucích přímo z jejich definice. Tyto vztahy umožňují snadnou kontrolu jejich správnosti. Pro koeficienty  $c_j$  filtru řádu  $N$  a stupně  $d$  platí, že

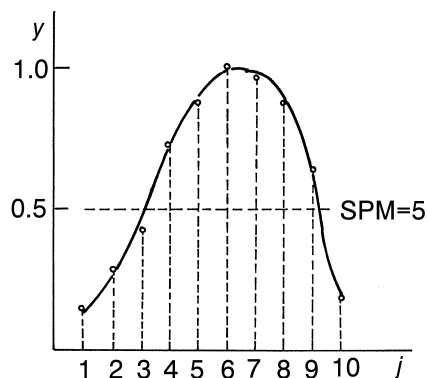
$$\sum_{j=0}^N j^q c_j = \begin{cases} 1 & \text{pro } q = 0 \\ 0 & \text{pro } q = 1, \dots, d \end{cases}.$$

V práci Bromby a Zieglera<sup>40</sup> jsou podrobně rozebrány vlastnosti regresních filtrů. Je ukázáno, že filtry jsou optimální pro signály, které se dají nahradit na délce filtru  $(2N+1)$  Taylorovým rozvojem do stupně  $d$ . Stupeň vyhlazení regresním filtrem bude růst s délkou filtru  $(2N+1)$  a klesat s růstem stupně filtru  $d$ . Při dostatečné délce filtru a nízkém stupni  $d$  lze očekávat i odstranění hrubších chyb.

V technické praxi se často filtruje signál, kde  $Z^*$  je ve tvaru píku, tj. jako součet gaussovských nebo lorentzovských křivek. Pro vyhlazení se používá kvadratických nebo kubických filtrů, kdy je  $d=2$ . Pro optimální vyhlazení je třeba, aby délka filtru  $F = (2N+1)$  byla menší než šířka píku v polovině maxima SPM. Proctor a Sherwood<sup>41</sup> doporučují volit  $F = 0.7 \text{ SPM}$ , kde SPM je udáno v počtech bodů užitých na tuto vzdálenost. Frank<sup>44</sup> však doporučuje volit délku filtru  $F$  při filtraci píku podle vztahu

$$F = A \frac{PM}{\Delta},$$

kde  $\Delta = x_{i+1} - x_i$  je skutečná vzdálenost mezi filtrovanými hodnotami a  $PM$  je šířka píku v polovině maxima v jednotkách  $x$ . Pro gaussovské píky se doporučuje  $A = 1$  a pro lorentzovské  $A = 0.7$ .



Obr. 9.33 Určení délky filtru při filtraci píku.

Detailní analýza výběru vhodné délky filtru pro různé stupně regresních filtrů je uvedena v práci Bromby a Zieglera<sup>40</sup>. Při konstrukci regresních filtrů postačuje pro zadaná  $N$  a  $d$  určit koeficienty  $c_j$ ,  $j = -N, \dots, 0, \dots, N$ . Problémem je, že matice  $\mathbf{H}$  je pro větší  $d$  špatně podmíněná.

Plyne, že vektor  $\mathbf{c}$  koeficientů regresního filtru lze snadno určit na základě koeficientů  $\mathbf{a}$  speciálního regresního polynomického modelu<sup>39</sup>

$$\delta_i = \alpha_0 + \sum_{k=1}^d \alpha_k f_i^k,$$

kde  $\delta_i = 1$  pro  $i = 0$  a  $\delta_i = 0$  pro  $i \neq 0$ . Funkce  $f_i = i^2$ . Soustava rovnic je pak soustavou normálních rovnic, ze které je  $\mathbf{a} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{e}$ , kde matice  $\mathbf{F}$  o rozměru  $(2N+1) \times (d+1)$  má prvky  $F_{ij} = i^{2j}$  pro  $i = -N, \dots, N$  a  $j = 0, 1, \dots, d$ . Místo proměnných  $i^j$  je výhodnější použít ortogonálních polynomů pro dané dělení  $-N, \dots, 0, \dots, N$ . Thrall<sup>39</sup> nahradil pro velká  $N$  tyto polynomy Legendrovými polynomy a odvodil vztahy pro koeficienty regresního filtru stupně  $2d$ . Pro kubické a kvadratické regresní filtry ( $d = 2$ , resp.  $d = 1$ ) lze počítat  $c_j$  v závislosti na velikosti  $N$  podle vztahu<sup>44</sup>

$$c_j = \frac{(3N^2 + 3N + 1) + 5j^2}{(2N + 1)(2N + 1)(2N + 3)/3}$$

nebo přibližně podle Thrallova vztahu<sup>39</sup>

$$c_j \approx \frac{1}{2N + 1} \left[ 1 + \frac{15}{4} \left( \frac{j}{N} \right)^2 + \frac{9}{4} \right].$$

Pro  $d = 4$ , resp.  $d = 5$ , tj. regresní filtr čtvrtého a pátého stupně, je možno použít vztah<sup>44</sup>

$$c_j = \frac{(15N^4 + 30N^3 + 35N^2 + 50N + 12) + 35(2N^2 + 2N + 3)j^2 + 63j^4}{4(2N + 3)(2N + 1)(2N + 1)(2N + 3)(2N + 5)/15}.$$

Regresní filtry lze použít také pro získávání vyhlazených hodnot derivací. Koeficienty kubického filtru pro první derivaci mají tvar

$$c_j^{(1)} = \frac{5[5(3N^4 + 6N^3 + 3N + 1)j + 7(3N^2 + 3N + 1)j^3]}{(2N + 3)(2N + 1)(2N + 1)(N + 2)(N + 1)N(N + 1)}.$$

Analytické vztahy pro případ  $d = 6, 7$  a první až páté derivace jsou uvedeny v práci Maddena<sup>42</sup>.

Regresní filtry neumožňují filtraci prvních  $N$  a posledních  $N$  bodů. To je ovšem při vyhlazování menšího počtu dat nevýhodné. Obvykle však postačuje počítat hodnoty prvních a posledních  $N$  bodů na základě regresních polynomů pro prvních a posledních  $(2N + 1)$  bodů. Dosazují se obecně  $j = 0$ . Pro případ kvadratického regresního filtru byl odvozen<sup>41</sup> pro výpočet  $Z(j)$  v intervalu  $-N \leq j \leq N$  vztah

$$Z(k) = \sum_{j=-N}^N y_j \left\{ \frac{5(3j^2 + N(N+1))k^2(2N+1)2N^3jk}{N(4N^2+1)(2N^3)(N+1)/3} + \frac{N(N+1)[3N(N+1) + 1 + 5j^2]}{N(4N^2+1)(2N^3)(N+1)/3} \right\}$$

Pro prvních  $N$  a posledních  $N$  bodů se počítají hodnoty  $Z(k)$  pro různá  $k$ , tj. od  $-N$  do 0 a od 0 do  $N$ . Ostatní body se vyhlazují dle centrální formule  $Z(0) = Z_0$ . Derivací této rovnice dle  $k$  se získá závislost pro určení prvních derivací. Postup využívající této rovnice se označuje *vyhlazení pomocí klouzavých parabol*.

Při on-line digitální filtraci s využitím jednoduchých programovatelných prostředků je vhodné použít filtrů rekurzivních. Obecný postup konstrukce rekurzivních regresních filtrů z regresních filtrů nerekurzivních je popsán v práci Bromby a Zieglera<sup>43</sup>. Příkladem *rekurzivní verze regresního filtru* pro  $d = 2$  tzv. *kvadratického filtru* o délce filtru 15, tj.  $N = 7$ , je vztah

$$Z_i = Z_{i+3} + 3(Z_{i+2} + Z_{i+1}) + \frac{78}{1105}(y_{i+7} + y_{i+10}) + \frac{221}{1105}(y_{i+6} + y_{i+9}) + \frac{153}{1105}(y_{i+5} + y_{i+8})$$

Regresní filtry lze konstruovat poměrně jednoduše, a to s ohledem na použitý výpočetní prostředek. V literatuře se většinou vychází z původní práce Savitzkého a Golaye<sup>45</sup>, která však obsahuje chyby v numericky vyčíslených koeficientech  $c_j$ . S problematikou regresních filtrů úzce souvisí techniky pohyblivých nejmenších čtverců nebo pohyblivé regrese, kdy se regresní modely určují lokálně pouze z informací o sousedních bodech vyhlazovaného bodu  $\{x_i, y_i\}$ . Tyto techniky se používají při hledání trendů v rozptylových grafech nebo grafech reziduí<sup>46</sup>.

## 9.6 Postup při interpolaci a aproximaci

V první fázi je třeba rozhodnout o tom, zda jde o úlohu *interpolace* či *aproximace*. Pro interpolaci platí, že hodnoty  $y$  jsou nenáhodné veličiny a aproximující funkce prochází všemi zadanými body. V případě aproximace jsou hodnoty  $y$  na ose  $x$  zatíženy náhodnými chybami nebo nepřesné a účelem je nalezení aproximující funkce, která je potlačuje:

### *Postup interpolace a aproximace*

- 1. Interpolace funkcí:** podle požadavků na shodu zadané a aproximující funkce lze volit buď klasickou polynommickou, nebo Hermitovskou interpolaci. Polohy uzlů interpolace je vhodné volit podle formule. Pro libovolné dělení uzlů interpolace je výhodnější použití spline interpolace defektu  $k = 1$  (požadavky na shodu ve funkčních hodnotách) nebo defektu  $k = 2$  (požadavky také na shodu v prvních derivacích).
- 2. Interpolace závislostí:** pro tento případ se doporučuje použití spline interpolace vhodného stupně (defektu  $k = 1$ ) tak, aby byly splněny podmínky spojitosti ve funkčních

hodnotách a hodnotách derivací. Volí se třída funkcí  $C^m$  a dle toho se vybírá vhodná metoda. Pokud není požadována znalost derivací aproximující funkce, je vhodné použít lokální kubické interpolační postupy  $C^1$  s automatickou úpravou lokální monotónnosti. Tyto metody umožňují poměrně kvalitní rekonstrukci závislosti. Pokud jsou požadovány znalosti první derivace rekonstruované závislosti, je vhodné použití kubických spline pod napětím s lokálním řízením tvaru aproximující funkce.

**3. Aproximace funkcí:** rozhodující je volba normy  $S_p$ . Pro případ  $p = 2$  (metoda nejmenších čtverců) nebo  $p = 4$  (Čebyševova minimaxní aproximace) a polynomičké aproximační modely je postup odhadu parametrů poměrně jednoduchý.

**4. Aproximace závislosti:** k řešení této úlohy existuje řada možností. Je možné volit mezi následujícími základními postupy:

- (a) aproximace polynomy pro zvolenou normu  $S_p$ ,
- (b) spline regrese s volbou stupně spline a uzlových bodů,
- (c) úseková regrese,
- (d) spline vyhlazování s volbou parametrů vyhlazení,
- (e) neparametrická regrese s volbou parametrů vyhlazení,
- (f) číslicová filtrace pro ekvidistantní dělení souřadnic  $x$  s volbou stupně a délky filtru.

Výběr vhodného postupu zde závisí na cíli zpracování dat. Odstranění náhodných šumů umožňují prakticky všechny postupy a liší se zejména definicí aproximující funkce a její složitostí. Pro účely tvorby empirických modelů vyhovuje obyčejně nejlépe spline regrese. Spline vyhlazování je účelné pro případy, u kterých se data budou následně numericky derivovat nebo integrovat.

Předběžné zpracování naměřených signálů je účelné provádět využitím robustního vyhlazování nebo regresních filtrů. Pro stejný účel je možné použít i neparametrických regresních modelů.