

Základy numerické matematiky

M. Feistauer

Charles University in Prague

OBSAH

1	Interpolace funkcí	1
1.1	Po částech polynomiální interpolace	1
1.1.1	Konstrukce kubického splinu	2
1.1.2	Odhad chyby	4
1.1.3	Spline s napětím	5
1.1.4	Hermiteův spline	5
2	Numerické řešení obyčejných diferenciálních rovnic	6
2.1	Příklady diskretních metod	7
2.1.1	Eulerova metoda	7
2.1.2	Runge-Kuttova metoda	7
2.1.3	Dvukroková metoda	8
2.2	Obecné jednokrokové metody	8
2.2.1	Konvergence jednokrokových metod	10
2.3	Odvození některých jednokrokových metod	11
2.3.1	Metody založené na přímém použití Taylorova vzorce	11
2.3.2	Runge-Kuttovy metody	12
2.4	Chyby metod	16
2.4.1	Odhad chyby metodou polovičního kroku	17
2.4.2	Zaokrouhlovací chyby	18
2.5	Soustavy lineárních diferenčních rovnic	19
2.5.1	Nalezení fundamentálního systému	21
2.5.2	Nalezení reálného fundamentálního systému	23
2.6	Vícekové metody	24
2.7	Některé vlastnosti obecných vícekové metody	26
2.8	Odvození některých vícekové metody	34
2.8.1	Interpolační polynom a zpětné diference	34
2.8.2	Metody založené na numerické integraci	35
3	Numerické metody optimalizace	38
3.1	Základy konvexní analýzy	39
3.2	Numerické metody hledání minima	43
4	Výsledky cvičení	45
	Rejstřík	46

INTERPOLACE FUNKCÍ

V této části se řeší následující problém: je dán interval $[a, b]$, v něm jsou dány body x_0, \dots, x_n a předepsané hodnoty $f(x_0), \dots, f(x_n)$; hledáme funkci daného typu, která vyhovuje těmto podmínkám. Ze zimního semestru víme, že existuje právě jeden (Lagrangeův) interpolační polynom p_n stupně nejvýše n , pro který je $p_n(x_j) = f(x_j)$, $j = 0, \dots, n$. Pro přesnost aproximace dostatečně hladké funkce máme následující větu:

Věta 1 *Je-li $f \in C^{n+1}[a, b]$, $x_0, \dots, x_n \in [a, b]$, pak pro každé $x \in [a, b]$ existuje $\xi \in (a, b)$ tak, že*

$$f(x) - p_n(x) = \frac{1}{(n+1)!} f^{(n+1)}(\xi) \prod_{j=0}^n (x - x_j).$$

Je-li aproximovaná funkce dokonce třídy C^∞ a má stejně omezené derivace, dostáváme jednoduše

Důsledek 2 *Bud' $f \in C^\infty[a, b]$, $|f^{(k)}| \leq K$ v $[a, b]$. Pak*

$$\max_{x \in [a, b]} |f(x) - p_n(x)| \leq \frac{K^{n+1}(b-a)^{n+1}}{(n+1)!} \xrightarrow{n \rightarrow \infty} 0.$$

Tudíž, $p_n \rightrightarrows f$ v $[a, b]$ pro $n \rightarrow \infty$.

Cvičení 1 *Uvažujme funkci $f(x) = \sin x$ na intervalu $[0, 1]$. Jaké největší chyby se můžeme dopustit, aproximujeme-li f pomocí p_9 ?*

Pokaždé však nedosáhneme tak vynikající aproximace. Příkladem může být funkce $f(x) = \frac{1}{1+x^2}$ na intervalu $[-5, 5]$. Dá se dokázat, že pro interpolační polynomy p_n s uzly v bodech ekvidistantního dělení je posloupnost $(\|f - p_n\|_{C([a, b])})_{n=1}^\infty$ neomezená.

1.1 Po částech polynomiální interpolace

Jednou z nevýhod aproximace interpolačním polynomem je skutečnost, že hodnoty interpolačního polynomu jsou silně ovlivněny i hodnotami funkce ve vzdálených uzlech. Řešením je aproximovat f po částech. Při tomto přístupu metodě je naší úlohou aproximovat funkci f v intervalu $[x_0, x_n]$ pomocí funkce φ takové, že $\varphi|_{[x_i, x_{i+1}]}$ je polynom. Většinou se navíc požaduje, aby φ byla třídy C^k pro dané k . Takovou funkci φ nazýváme *spline*.

Nejjednodušším případem ($k = 0$) je aproximace pomocí funkce po částech lineární, jejímž grafem je lomená čára. V praxi je většinou třeba lepší aproximace. Přijatelným řešením je tzv. *kubický spline*.

Definice 1 Řekneme, že funkce $\varphi : [x_0, x_n] \rightarrow \mathbb{R}$ je *kubický spline*, jestliže

- (i) $\varphi \in C^2[x_0, x_n]$,
- (ii) $\varphi|_{[x_i, x_{i+1}]}$ je polynom stupně nejvýše 3.

Říkáme, že φ je *kubický interpolační spline* k f v bodech x_0, \dots, x_n , jestliže jsou navíc splněny podmínky $\varphi(x_i) = f(x_i), i = 0, \dots, n$.

Restriksi φ na $[x_i, x_{i+1}]$ označíme φ_i . Funkci φ_i lze psát ve tvaru

$$\varphi_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3.$$

Funkce φ je tedy určena celkem $4n$ parametry. Podmínky z definice kubického interpolačního splinu nám však dávají jen $4n - 2$ podmínek. Dá se tedy očekávat, že budeme muset ještě dva parametry zvolit. Nejčastěji se používají okrajové podmínky v x_0 a x_n , a to tří typů:

- (α) $\varphi'(x_0) = f'_0, \varphi'(x_n) = f'_n$;
- (β) $\varphi''(x_0) = f''_0, \varphi''(x_n) = f''_n$;
- (γ) $\varphi''(x_0) = 0, \varphi''(x_n) = 0$.

Kubický interpolační spline určený podmínkou (γ) se nazývá *přirozený spline*.

1.1.1 Konstrukce kubického splinu

Budeme konstruovat spline určený podmínkou (β) (jíž je (γ) speciálním případem). Předpokládejme nejprve, že již známe čísla $M_i = \varphi''(x_i)$, tzv. *momenty splinu*. Funkce φ'' je spojitá a po částech lineární. Označíme-li tedy $h_i = x_{i+1} - x_i$, potom pro $x \in [x_i, x_{i+1}]$ máme

$$\varphi''(x) = M_i + (M_{i+1} - M_i) \frac{x - x_i}{x_{i+1} - x_i} = M_i \frac{x_{i+1} - x}{h_i} + M_{i+1} \frac{x - x_i}{h_i}.$$

Integrováním dostaneme φ'_i a φ_i :

$$\begin{aligned} \varphi'_i(x) &= -M_i \frac{(x_{i+1} - x)^2}{2h_i} + M_{i+1} \frac{(x - x_i)^2}{2h_i} + A_i \\ \varphi_i(x) &= M_i \frac{(x_{i+1} - x)^3}{6h_i} + M_{i+1} \frac{(x - x_i)^3}{6h_i} + A_i(x - x_i) + B_i. \end{aligned}$$

Pomocí (známých) hodnot $\varphi_i(x_i) = f(x_i)$, $\varphi_i(x_{i+1}) = f(x_{i+1})$ určíme konstanty A_i a B_i :

$$\begin{aligned} f(x_i) = \frac{1}{6}M_i h_i^2 + B_i &\implies B_i = f(x_i) - \frac{1}{6}M_i h_i^2 \implies \\ f(x_{i+1}) &= \frac{1}{6}M_{i+1} h_i^2 + A_i h_i + B_i \\ A_i = \frac{1}{h_i} \left(f(x_{i+1}) - \frac{1}{6}M_{i+1} h_i^2 - B_i \right) &= \frac{f(x_{i+1}) - f(x_i)}{h_i} - \frac{h_i}{6}(M_{i+1} - M_i). \end{aligned}$$

Zbývá určit hodnoty momentů: M_0 a M_n máme zadané, ostatní vypočteme ze spojitosti první derivace (derivací funkce φ_i v bodech x_i, x_{i+1} se rozumí příslušná jednostranná derivace).

$$\begin{aligned}\varphi'_{i-1}(x_i-) &= \frac{1}{2}M_i h_{i-1} + A_{i-1} = \\ &= \frac{1}{2}M_i h_{i-1} + \frac{f(x_i) - f(x_{i-1})}{h_{i-1}} - \frac{h_{i-1}}{6}(M_i - M_{i-1})\end{aligned}$$

$$\begin{aligned}\varphi'_i(x_i+) &= \frac{1}{2}M_i h_i + A_i = \\ &= \frac{1}{2}M_i h_i + \frac{f(x_{i+1}) - f(x_i)}{h_i} - \frac{h_i}{6}(M_{i+1} - M_i).\end{aligned}$$

Z rovnosti obou derivací dostaneme po úpravách

$$M_{i-1} \frac{h_{i-1}}{6} + M_i \left(\frac{h_{i-1}}{3} + \frac{h_i}{3} \right) + M_{i+1} \frac{h_i}{6} = \frac{f(x_{i+1}) - f(x_i)}{h_i} - \frac{f(x_i) - f(x_{i-1})}{h_{i-1}};$$

Označíme-li $\lambda_i = \frac{h_{i-1}}{h_{i-1}+h_i}$, $\mu_i = 1 - \lambda_i = \frac{h_i}{h_{i-1}+h_i}$, lze rovnici přepsat ve tvaru

$$\lambda_i M_{i-1} + 2M_i + \mu_i M_{i+1} = g_i,$$

kde g_i je pravá strana původní rovnice, vynásobená výrazem $\frac{6}{h_{i-1}+h_i}$. Dostáváme soustavu

$$\begin{aligned}2M_1 + \mu_1 M_2 &= g_1 - \lambda_1 f_0'' \\ \lambda_2 M_1 + 2M_2 + \mu_2 M_3 &= g_2 \\ \lambda_3 M_2 + 2M_3 + \mu_3 M_4 &= g_3 \\ &\vdots \\ \lambda_{n-1} M_{n-2} + 2M_{n-1} &= g_{n-1} - \mu_{n-1} f_n''.\end{aligned}$$

Dokážeme-li nyní existenci a jednoznačnost řešení, jsme hotovi. Všimněme si, že prvky na diagonále matice soustavy jsou vždy 2, zatímco součet všech ostatních prvků v libovolném řádku je mezi 0 a 1 (s výjimkou prvního a posledního dokonce právě 1). Matice je tedy ostře diagonálně dominantní a tedy i regulární.

Cvičení 2 *Dokažte, že každá ostře diagonálně dominantní matice je regulární.*

Navíc je matice soustavy tzv. třídiagonální matice, na které se poměrně jednoduše provádí eliminace. Soustava vypadá takto (uvažujeme obecný případ – matici $n \times n$):

$$\begin{pmatrix} a_1 & c_1 & 0 & \dots & 0 & 0 \\ b_1 & a_2 & c_2 & \dots & 0 & 0 \\ \vdots & & \ddots & & \vdots & \\ 0 & & \dots & b_{n-1} & a_n \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{pmatrix}$$

Při přímém chodu Gaussovy eliminace (tvorba horní trojúhelníkové matice) nám zmizí všechny b_j a na diagonále se nám postupně objeví členy $A_{jj} = \eta_j$, kde

$$\eta_1 = a_1, \quad \eta_2 = a_2 - \frac{b_1}{\eta_1}c_1, \quad \text{obecně } \eta_i = a_i - \frac{b_{i-1}}{\eta_{i-1}}c_{i-1} \quad (i = 2, \dots, n),$$

a ve vektoru pravých stran analogicky vzniknou ξ_j , kde

$$\xi_1 = d_1, \quad \xi_2 = d_2 - \frac{b_1}{\eta_1}\xi_1, \quad \text{obecně } \eta_i = d_i - \frac{b_{i-1}}{\eta_{i-1}}\xi_{i-1} \quad (i = 2, \dots, n).$$

Dostaneme tedy soustavu

$$\begin{pmatrix} \eta_1 & c_1 & 0 & \dots & 0 & 0 \\ 0 & \eta_2 & c_2 & \dots & 0 & 0 \\ \vdots & & \ddots & & \vdots & \\ 0 & & \dots & 0 & \eta_n & \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{pmatrix}$$

z níž už snadno spočítáme neznámé y_j :

$$y_n = \frac{\xi_n}{\eta_n}, \quad y_{n-1} = \frac{\xi_{n-1} - c_{n-1}y_n}{\eta_{n-1}}, \quad \text{obecně } y_i = \frac{\xi_i - c_i y_{i+1}}{\eta_i} \quad (i = 1, \dots, n-1).$$

Snadno se ukáže, že pro ostře diagonálně dominantní matici vyjdou po eliminaci prvky na diagonále nenulové, takže nemusíme prohazovat řádky a neznámé přímo spočítáme podle uvedených vztahů.

1.1.2 Odhad chyby

Zabývejme se nyní otázkou, jaké chyby se dopustíme, aproximujeme-li funkci interpolačním kubickým splinem. Zhruba řečeno, je-li f dostatečně hladká v $[a, b]$ a dělení intervalu neomezeně zjemňujeme *vhodným způsobem*, pak interpolační kubické spliny konvergují stejnoměrně k f (případně i s některými derivacemi). Přesnou formulaci této myšlenky dává následující věta, kterou uvedeme bez důkazu.

Věta 3 *Bud' $f \in C^4[a, b]$. Pak existuje konstanta $C > 0$ taková, že platí: Nechť $K > 0$ je konstanta. Dále uvažujme dělení D intervalu $[a, b]$, které je tvořeno body $a = x_0 < \dots < x_n = b$ a které splňuje podmínku*

$$\frac{\max h_i}{\min h_i} \leq K,$$

kde $h_i = x_{i+1} - x_i$. Dále uvažujme okrajové podmínky pro interpolační kubický spline $\varphi''(x_0) = f''(x_0)$, $\varphi''(x_n) = f''(x_n)$. Potom

$$\left| f^{(k)} - \varphi^{(k)} \right| \leq CK h^{4-k}, \quad (k = 0, \dots, 3),$$

kde $h = \max h_i$, přičemž v případě $k = 3$ uvažujeme v dělicích bodech derivace zleva nebo zprava.

Důsledek: Máme-li posloupnost dělení D' takových, že $h \rightarrow 0$, dostáváme

$$\varphi^{(k)} \Rightarrow f^{(k)}.$$

Poznámka: Pokud používáme interpolační spline pro interpolaci několika naměřených hodnot, nemůžeme použít okrajové podmínky ve tvaru rovnosti druhých derivací; hodila by se tedy obdobná věta pro přirozený interpolační spline. Nahradíme-li okrajové podmínky ve větě podmínkami $\varphi''(x_0) = 0$, $\varphi''(x_n) = 0$, dostaneme (slabší) odhad

$$|f(x) - \varphi(x)| \leq CK h^2, \quad (x \in [a, b]).$$

1.1.3 Spline s napětím

Ne vždy představuje kubický interpolační spline ideální řešení problému. Například při aproximaci nespojitých funkcí nebo funkcí s nespojitou derivací dostáváme nepříjemnou oscilaci v okolí nespojitosti. Proto se někdy používá modifikovaná konstrukce tzv. *splínu s napětím*. Při této konstrukci opět požadujeme $\varphi \in C^2[a, b]$, $\varphi(x_i) = f(x_i)$, ale požadavek, aby φ byla po částech kubická se nahrazuje požadavkem, aby $\varphi|_{[x_i, x_{i+1}]}$ byla řešením diferenciální rovnice

$$\varphi^{(4)} - \tau\varphi'' = 0,$$

kde $\tau \geq 0$ je tzv. *napěťový parametr*. Pokud bychom položili $\tau = 0$, dostali bychom přesně polynomy stupně nejvýše 3. Pro $\tau > 0$ místo funkcí x^2 a x^3 dostaneme fundamentální řešení $\cosh(t\sqrt{\tau}) - 1$ a $\sinh(t\sqrt{\tau})$. Všimněme si, že při pevné funkci f a dělení D pro $\tau \rightarrow \infty$ dostaneme $\varphi'' \rightarrow 0$. Význam napěťového parametru je tedy takový, že pro velké τ má spline tendenci být „téměř lineární“ (větší napětí).

1.1.4 Hermiteův spline

Další modifikací je *Hermiteův spline*. Při této konstrukci požadujeme pouze $\varphi \in C^1[a, b]$, $\varphi|_{[x_i, x_{i+1}]}$ je opět polynom stupně nejvýše 3 a $\varphi(x_i) = f(x_i)$, $\varphi'(x_i) = f'(x_i)$.

V praxi při interpolaci naměřených hodnot ovšem není derivace f' známa. V takových případech se hodnota $f'(x_i)$ nahrazuje výrazem

$$\frac{f(x_{i+1}) - f(x_{i-1}))}{x_{i+1} - x_{i-1}}$$

(pro $i = 1, \dots, n-1$). Pokud je totiž $f \in C^2[a, b]$, je pro $h \rightarrow 0$

$$f'(x_i) = \frac{f(x_{i+1}) - f(x_{i-1}))}{x_{i+1} - x_{i-1}} + O(h). \quad (*)$$

Cvičení 3 Dokažte platnost vztahu (*).

Cvičení 4 Za předpokladu $f \in C^3[a, b]$ najděte přesnější aproximaci $f'(x_i)$ pro $i = 1, \dots, n-1$ s chybou $O(h^2)$.

NUMERICKÉ ŘEŠENÍ OBYČEJNÝCH DIFERENCIÁLNÍCH ROVNIC

Protože rovnici vyššího řádu lze vždy jednoduše převést na soustavu rovnic prvního řádu, budeme se v této části zabývat pouze rovnicemi prvního řádu, navíc rozřešenými vzhledem k derivaci, tj. rovnicemi tvaru

$$y' = f(x, y),$$

což může reprezentovat buď jednu rovnici nebo celou soustavu – v tom případě řešení $y = (y^1, \dots, y^s) : \mathbb{R} \rightarrow \mathbb{R}^s$ a funkce $f = (f^1, \dots, f^n) : \mathbb{R} \times \mathbb{R}^s \rightarrow \mathbb{R}^s$ jsou vektorové funkce. Numerické metody mají své uplatnění hlavně v situacích, kdy přesné (analytické) řešení nedokážeme najít.

Příklad 1 Pohyb částice v silovém poli

Trajektorii pohybu částice lze popsat jednou vektorovou funkcí $\mathbf{x} = \mathbf{x}(t)$, kde proměnná t má fyzikální význam času a $\mathbf{x}(t)$ je poloha částice v čase t . Newtonův pohybový zákon pak lze formulovat ve tvaru (m_p je hmotnost částice)

$$m_p \mathbf{x}'' = \mathbf{F}(\mathbf{x}', \mathbf{x}, t),$$

což je vlastně soustava tří obyčejných diferenciálních rovnic druhého řádu, kterou lze převést na soustavu šesti obyčejných diferenciálních rovnic prvního řádu. Funkce \mathbf{F} zde vyjadřuje závislost působící síly na čase, poloze částice a její rychlosti. Obecně je však tato závislost natolik složitá, že není šance tuto soustavu vyřešit analyticky. Mnohdy nám však stačí najít numerickými metodami řešení přibližné.

Příklad 2 I u velmi jednoduchých rovnic se nám může stát, že přesné řešení nedokážeme najít. Typickým příkladem je rovnice

$$y' = x^2 + y^2,$$

o níž je dokázáno, že žádné její řešení nelze poskládat z elementárních funkcí ani jejich neurčitých integrálů.

Příklad 3 I v případě, že umíme přesné řešení najít, se může stát, že se bez numerických metod neobejdeme. Rovnice $y' = 1 - 2xy$ s počáteční podmínkou $y(0) = 0$ má řešení

$$y(x) = e^{-x^2} \int_0^x e^{t^2} dt.$$

Chceme-li znát jeho hodnotu v nějakém bodě x , potřebujeme numericky spočítat určitý integrál.

Příklad 4 Rovnice

$$y'' + \frac{A}{x}y' + \left(\frac{B}{x^2} + C + Dx^2\right)y = 0$$

je příkladem rovnice Fuchsova typu. Její řešení lze najít metodou mocninných řad, ale pro většinu argumentů x řada konverguje pomalu. Nahradíme-li Dx^2 členem Dx^3 , nelze metodu mocninných řad použít.

2.1 Příklady diskretních metod

Uvažujme rovnici $y' = f(x, y)$ na intervalu $[a, b]$ s počáteční podmínkou $y(a) = \eta$. Necht' $y : [a, b] \rightarrow \mathbb{R}$ je řešení uvažovaného problému. Uvažujme dělení intervalu $[a, b]$ tak, že $a = x_0 < \dots < x_N \leq b$. Budeme se snažit přiřadit bodům x_i přibližné hodnoty y_i .

2.1.1 Eulerova metoda

Je to nejjednodušší metoda numerického řešení ODR. Předpokládejme, že máme řešení rovnice a v intervalu $[a, b]$ máme dva body $x_n < x_{n+1}$ uvažovaného dělení. Předpokládejme navíc, že řešení je třídy C^2 . Taylorův vzorec nám dává $y(x_{n+1}) = y(x_n) + hy'(x_n) + O(h^2)$. Odtud

$$y'(x_n) = \frac{y(x_{n+1}) - y(x_n)}{h} + O(h),$$

což (y_n mají být přibližné hodnoty řešení) nám dává rovnici

$$\frac{1}{h}(y_{n+1} - y_n) = f(x_n, y_n).$$

Dostáváme tedy rekurentní vztah

$$y_{n+1} = y_n + hf(x_n, y_n), \quad y_0 = \eta.$$

Cvičení 5 Je-li dokonce $y \in C^3[a, b]$, lze derivaci aproximovat přesněji:

$$y'(x_n + \frac{h}{2}) = \frac{y(x_{n+1}) - y(x_n)}{h} + O(h^2).$$

2.1.2 Runge-Kuttova metoda

Vyjdeme-li ze vztahu uvedeného v posledním cvičení a vztahů

$$y'(x_n + \frac{h}{2}) = f(x_n + \frac{h}{2}, y(x_n + \frac{h}{2})),$$

$$y(x_n + \frac{h}{2}) = y(x_n) + \frac{h}{2}y'(x_n) + O(h^2),$$

dostaneme požadavek

$$\frac{1}{h}(y_{n+1} - y_n) = f(x_n + \frac{h}{2}, y_n + \frac{h}{2}f(x_n, y_n));$$

příslušný rekurentní vztah je

$$y_{n+1} = y_n + hf(x_n + \frac{h}{2}, y_n + \frac{h}{2}f(x_n, y_n)), \quad y_0 = \eta.$$

Tato metoda je *Runge-Kuttova metoda druhého řádu*.

2.1.3 Dvoustupňová metoda

Společnou vlastností obou uvedených metod bylo, že hodnota y_{n+1} se vypočítávala pouze z *jedné* předcházející hodnoty y_n (a samozřejmě z x_n a h). U více-
krokových metod používáme rekurentní vyjádření z více předešlých hodnot.

Mějme nyní opět $y \in C^3[a, b]$, x_n, x_{n+1}, x_{n+2} tři po sobě jdoucí body ekvi-
distantního dělení $[a, b]$ (tedy $x_n = a + nh$, kde h je krok metody). Pak máme

$$y(x_{n+1}) = y(x_n) + hy'(x_n) + \frac{1}{2}h^2y''(x_n) + O(h^3), \quad (1)$$

$$y(x_{n+2}) = y(x_n) + 2hy'(x_n) + 2h^2y''(x_n) + O(h^3). \quad (2)$$

Odečtením čtyřnásobku (1) od (2) dostaneme

$$y(x_{n+2}) - 4y(x_{n+1}) = -3y(x_n) - 2hy'(x_n) + O(h^3),$$

odkud dosazením rovnice $y'(x_n) = f(x_n, y(x_n))$ dojdeme k rekurentnímu vztahu

$$y_{n+2} - 4y_{n+1} + 3y_n = -2hf(x_n, y_n),$$

kde $y_0 = \eta$ a y_1 spočítáme pomocí některé jednokrokové metody. Popsaná me-
toda je dvoustupňová – další hodnotu počítáme pomocí dvou předcházejících.

2.2 Obecné jednokrokové metody

Při těchto metodách je dán krok $h > 0$, počáteční podmínka $y(x_0) = \eta$. Uzly jsou
 $x_n = a + nh$, $n \geq 0$. Hodnoty přibližného řešení počítáme podle rekurentního
vztahu

$$y_0 = \eta, \quad y_{n+1} = y_n + h\Phi(x_n, y_n, h).$$

Funkce Φ (která závisí na f) se nazývá *přírůstková funkce*. Například u Eulerovy
metody jsme měli $\Phi(x, y, h) = f(x, y)$, u Runge-Kuttovy metody druhého řádu
je $\Phi(x, y, h) = f(x + \frac{h}{2}, y + \frac{h}{2}f(x, y))$.

Chybou metody (v bodě x_n) rozumíme rozdíl $e_n = y_n - y(x_n)$, tzv. *akumulo-*
vanou diskretizační chybu. Naším cílem je

- (1) Najít odhad e_n v závislosti na h ;
- (2) ukázat, že v jistém smyslu je $e_n \rightarrow 0$ pro $h \rightarrow 0$.

Abychom měli zaručenu existenci a jednoznačnost řešení, budeme předpokládat,
že $f : [a, b] \times \mathbb{R} \rightarrow \mathbb{R}$ je spojitá a že je Lipschitzovská vzhledem k y , tj. že

$$|f(x, y_1) - f(x, y_2)| \leq L|y_1 - y_2| \quad \forall x \in [a, b], \quad y_1, y_2 \in \mathbb{R}.$$

Z Picardovy věty (viz MA) vyplývá, že za těchto předpokladů má úloha

$$y' = f(x, y), \quad y(a) = \eta \quad (*)$$

právě jedno řešení $y : [a, b] \rightarrow \mathbb{R}$. Navíc předpokládejme (budeme se zabývat
jen rozumnými metodami), že i funkce $\Phi : [a, b] \times \mathbb{R} \times [0, h_0] \rightarrow \mathbb{R}$ je spojitá
a splňuje Lipschitzovu podmínku vzhledem k y .

Definice 2 Řekneme, že jednokroková metoda s přírůstkovou funkcí Φ je konvergentní, jestliže platí následující tvrzení: kdykoli je $y : [a, b] \rightarrow \mathbb{R}$ řešením úlohy (*), je

$$\forall x \in [a, b] : \lim_{\substack{h \rightarrow 0+ \\ x_n = x}} y_n = y(x),$$

kde y_n je přibližné řešení v uzlu x_n .

Definice 3 Jiná definice konvergence Krok $h > 0$ nám určuje uzly x_n , k nimž pomocí jednokrokové metody přiřadíme přibližné hodnoty řešení. Označme $e(h)$ maximální chybu, tj.

$$e(h) = \max_{\substack{x_n \in [a, b] \\ x_n = a + nh}} |e_n|.$$

Řekneme, že metoda je konvergentní, jestliže platí následující tvrzení: kdykoli je $y : [a, b] \rightarrow \mathbb{R}$ řešením úlohy (*), je

$$\lim_{h \rightarrow 0+} e(h) = 0.$$

Poznámka: Snadno nahlédneme, že konvergence podle druhé definice implikuje konvergenci podle první definice.

Před odvozením některých jednokrokových metod odvodíme jedno pomocné lemma, které nám bude často užitečné.

Lemma 1 Necht' $A, B \geq 0$, $N \geq 1$ celé a necht' je splněna podmínka

$$|\xi_{n+1}| \leq A|\xi_n| + B, \quad n = 0, \dots, N-1.$$

Poto platí

$$|\xi_n| \leq A^n |\xi_0| + \begin{cases} \frac{A^n - 1}{A - 1} B & \text{pro } A \neq 1, \\ Bn & \text{pro } A = 1. \end{cases} \quad (*)$$

Cvičení 6 Dokažte toto lemma.

Při aplikaci lemmatu (1) je často $A = 1 + \delta$, $\delta > 0$. Použijeme-li nerovnost $1 + \delta < e^\delta$, pak z (*) plyne

$$|\xi_n| \leq e^{n\delta} |\xi_0| + \frac{e^{n\delta} - 1}{\delta} B. \quad (**)$$

Pro $L > 0$ a $x \in \mathbb{R}$ označme

$$E_L(x) = \frac{e^{Lx} - 1}{L}. \quad (***)$$

(E_L je tzv. Lipschitzova funkce.)

2.2.1 Konvergence jednokrokových metod

Předpoklad (P_Φ): Předpokládejme, že přírůstková funkce $\Phi : [a, b] \times \mathbb{R} \times [0, h_0] \rightarrow \mathbb{R}$, $\Phi = \Phi(x, y, h)$, je spojitá a splňuje Lipschitzovu podmínku vzhledem k y s konstantou $L > 0$:

$$|\Phi(x, y, h) - \Phi(x, y^*, h)| \leq L|y - y^*|, \quad x \in [a, b], \quad y, y^* \in \mathbb{R}, \quad h \in [0, h_0]. \quad (+)$$

Definice 4 Řekneme, že jednokroková metoda s přírůstkovou funkcí Φ pro řešení diferenciální rovnice $y' = f(x, y)$ je *konsistentní*, jestliže

$$\Phi(x, y, 0) = f(x, y), \quad x \in [a, b], \quad y \in \mathbb{R}.$$

Věta 4 Jednokroková metoda s přírůstkovou funkcí Φ splňující předpoklad (P_Φ) je konvergentní, právě když je konsistentní.

Důkaz vynecháme, protože nás zajímá odhad chyby metody e_n . Odhad chyby metody provádíme ve dvou krocích.

a) Dosadíme přesné řešení do uvažované metody. Dostaneme vztah

$$y(x_n + h) - y(x_n) - h\Phi(x_n, y(x_n), h) = h\delta_n, \quad x_n, x_n + h \in [a, b], \quad h \in (0, h_0).$$

b) Odhadneme δ_n a z tohoto odhadu učiníme odhad chyby.

Veličinu δ_n nazýváme lokální relativní diskretizační chybou v uzlu x_n . Obecně *lokální relativní diskretizační chybu* (krátce chybu diskretizace) v bodě x definujeme jako výraz

$$\Delta(x, y(x), h) - \Phi(x, y(x), h),$$

kde

$$\Delta(x, y(x), h) = \frac{y(x+h) - y(x)}{h}, \quad x, x+h \in [a, b], \quad h \in (0, h_0),$$

je tzv. *přesný relativní přírůstek*.

Věta 5 Nechť $y : [a, b] \rightarrow \mathbb{R}$ je přesné řešení úlohy $y' = f(x, y)$ v intervalu $[a, b]$, $y(a) = \eta$, y_n pro $x_n \in [a, b]$ jsou hodnoty přibližného řešení vypočtené pomocí jednokrokové metody s přírůstkovou funkcí Φ splňující předpoklad (P_Φ). Nechť navíc existují konstanty $N, p > 0$ takové, že

$$|\Delta(x, y(x), h) - \Phi(x, y(x), h)| \leq Nh^p, \quad x, x+h \in [a, b], \quad h \in (0, h_0). \quad (++)$$

Potom pro chybu metody (tj. akumulovanou diskretizační chybu) $e_n = y_n - y(x_n)$ platí odhad

$$|e_n| \leq Nh^p E_L(x_n - a), \quad x_n \in [a, b], \quad h \in (0, h_0).$$

Důkaz: Zřejmě máme $e_0 = y_0 - y(x_0) = 0$. Dále máme vztahy

$$y_{n+1} = y_n + h\Phi(x_n, y_n, h),$$

$$y(x_{n+1}) = y(x_n) + h\Phi(x_n, y(x_n), h) + h\delta_n,$$

kde

$$\delta_n = \Delta(x_n, y(x_n), h) - \Phi(x_n, y(x_n), h)$$

pro $x_n, x_{n+1} \in [a, b]$, $h \in (0, h_0)$. Odtud vyplývá odečtením druhé rovnice od první

$$e_{n+1} = e_n + h(\Phi(x_n, y_n, h) - \Phi(x_n, y(x_n), h)) - h(\Delta(x_n, y(x_n), h) - \Phi(x_n, y(x_n), h)).$$

Použijeme-li předpoklad (P_Φ) a vztah $(++)$, dostaneme ihned nerovnost

$$|e_{n+1}| \leq (1 + hL)|e_n| + Nh^{p+1}, \quad x_n, x_{n+1} \in [a, b], \quad h \in (0, h_0).$$

Aplikace lemmatu 1 dává odhad

$$|e_n| \leq \frac{(1 + hL)^n - 1}{hL} Nh^{p+1} \leq$$

$$\leq \frac{e^{nhL} - 1}{L} Nh^p = Nh^p E_L(x_n - a), \quad x_n \in [a, b], \quad h \in (0, h_0),$$

Poznámka: Jednokrokové metody mají tu výhodu, že v každém kroku (při přechodu od x_n k x_{n+1}) lze měnit délku kroku. Tzn., že můžeme uvažovat obecné dělení $a = x_0 < x_1 < \dots \leq b$ intervalu $[a, b]$, položit $h_n = x_{n+1} - x_n$ a definovat

$$y_{n+1} = y_n + h_n \Phi(x_n, y_n, h_n).$$

2.3 Odvození některých jednokrokových metod

2.3.1 Metody založené na přímém použití Taylorova vzorce

V tomto odstavci budeme konstruovat jednokrokovou metodu p -tého řádu za předpokladu, že přesné řešení y je třídy C^{p+1} . Napišme si Taylorův vzorec a upravujme:

$$y(x+h) = y(x) + \sum_{k=1}^p \frac{y^{(k)}(x)}{k!} h^k + \frac{y^{(p+1)}(\tilde{x})}{(p+1)!} h^{p+1}$$

$$\frac{1}{h}(y(x+h) - y(x)) = \sum_{k=1}^p \frac{y^{(k)}(x)}{k!} h^{p-1} + \frac{y^{(p+1)}(\tilde{x})}{(p+1)!} h^p.$$

Nyní použijeme rovnici a pokusíme se spočítat několik prvních derivací funkce y :

$$y'(x) = f(x, y(x)),$$

$$y''(x) = \frac{d}{dx} f(x, y(x)) = \frac{\partial f}{\partial x}(x, y(x)) + y'(x) \frac{\partial f}{\partial y}(x, y(x)) =$$

$$= \frac{\partial f}{\partial x}(x, y(x)) + f(x, y(x)) \frac{\partial f}{\partial y}(x, y(x)),$$

$$y'''(x) = \frac{d}{dx} \left\{ \left(\frac{\partial f}{\partial x} + f \frac{\partial f}{\partial y} \right) (x, y(x)) \right\} = \frac{\partial}{\partial x} \{ \dots \} + f \cdot \frac{\partial}{\partial y} \{ \dots \}.$$

Pro zjednodušení (zápisu) definujeme *diferenciální operátor*¹ takto: pro funkci $\varphi \in C^1(G)$, kde $G \subseteq \mathbb{R}^2$, buď

$$D\varphi = \frac{\partial\varphi}{\partial x} + f \cdot \frac{\partial\varphi}{\partial y};$$

mocninou rozumíme skládání, tj. $D^0\varphi := \varphi$, pro $k \geq 0$ buď $D^{k+1}\varphi := D(D^k\varphi)$. S jeho pomocí můžeme snadno vyjádřit derivace y :

$$\begin{aligned} y'(x) &= (D^0 f)(x, y(x)), \\ y''(x) &= (D^1 f)(x, y(x)), \\ y'''(x) &= (D^2 f)(x, y(x)), \\ &\vdots \\ y^{(k)}(x) &= (D^{(k-1)} f)(x, y(x)). \end{aligned}$$

Odtud dostáváme

$$\frac{1}{h}(y(x+h) - y(x)) = \sum_{k=1}^p \frac{(D^{k-1} f)(x, y(x))}{k!} h^{k-1} + \frac{y^{(p+1)}(\tilde{x})}{(p+1)!} h^p,$$

což nás vede k tomu, abychom použili přírůstkovou funkci

$$\Phi(x, y, h) = \sum_{k=1}^p \frac{(D^{k-1} f)}{k!} h^{k-1}.$$

Pak je totiž

$$|\delta(x, h)| = \left| \frac{1}{h}(y(x+h) - y(x)) - \Phi(x, y(x), h) \right| \leq \left| \frac{y^{(p+1)}}{(p+1)!} h^p \right| \leq N h^p,$$

kde

$$N = \max_{x \in [a, b]} \left| \frac{y^{(p+1)}}{(p+1)!} \right|.$$

Cvičení 7 Vypočtěte $D^2 f$, $D^3 f$ a pro $f(x, y) = x^2 + y^2$ vypočtěte $D^4 f$.

2.3.2 Runge-Kuttovy metody

Získali jsme sice jedнокrokovou metodu libovolně vysokého řádu, ale pro praxi je téměř nepoužitelná. Za prvé proto, že (viz cvičení 7), že počítání mocnin operátoru D vede brzy k příliš složitým výrazům. Druhý (a hlavní) důvod spočívá

¹v podstatě nic jiného než zobrazení z jistého prostoru funkcí do nějakého jiného prostoru funkcí

v tom, že v definici funkce Φ vystupuje nejen funkce f sama, ale i její parciální derivace. V dalším označme

$$\tilde{\Phi}(x, y, h) = \sum_{k=1}^p \frac{(D^{k-1}f)(x, y)}{k!} h^{k-1}.$$

Tuto přírůstkovou funkci použijeme k odvození *Runge-Kuttových metod*, u nichž se Φ počítá pouze pomocí hodnot f . Jejich přírůstková funkce bude ve tvaru

$$\Phi(x, y, h) = \sum_{k=1}^P w_i k_i(x, y, h),$$

kde

$$\begin{aligned} k_1(x, y, h) &= f(x, y) \\ k_2(x, y, h) &= f(x + \alpha_2 h, y + \beta_{21} h k_1(x, y, h)) \\ &\vdots \\ k_i(x, y, h) &= f\left(x + \alpha_i h, y + h \sum_{j=1}^{i-1} \beta_{ij} k_j(x, y, h)\right), \end{aligned}$$

kde P , w_i , α_i a β_{ij} jsou vhodně zvolené konstanty (tak, aby výsledná metoda byla řádu p). V dalším se budeme snažit určit hodnoty těchto konstant tak, abychom dostali odhad $|\Phi(x, y, h) - \tilde{\Phi}(x, y, h)| = O(h^p)$.

Příklad 5 Pokusme se odvodit Runge-Kuttovu metodu druhého řádu, tj. pro $p = 2$. Zkusíme počet členů také $P = 2$. Napišme si vztahy pro Φ a $\tilde{\Phi}$:

$$\begin{aligned} \Phi(x, y, h) &= w_1 f(x, y) + w_2 f(x + \alpha_2 h, y + \beta_{21} h f(x, y)) \\ \tilde{\Phi}(x, y, h) &= f(x, y) + \frac{h}{2} (Df)(x, y) = \\ &= f(x, y) + \frac{h}{2} \left(\frac{\partial f}{\partial x}(x, y) + f(x, y) \frac{\partial f}{\partial y}(x, y) \right). \end{aligned}$$

Dále si rozepíšeme funkční hodnotu z prvního vztahu podle Taylorova vzorce:²

²Taylorův vzorec pro funkce n proměnných: buď $f \in C^k(\Omega)$, $\Omega \subseteq \mathbb{R}^n$ otevřená, $a, b \in \Omega$ a předpokládejme, že i celá úsečka s krajními body a, b leží v Ω ; potom

$$f(b) = \sum_{j=0}^{k-1} \frac{1}{j!} \left[\sum_{i=1}^n (b_i - a_i) \frac{\partial}{\partial x_i} \right]^j f(a) + \frac{1}{k!} \left[\sum_{i=1}^n (b_i - a_i) \frac{\partial}{\partial x_i} \right]^k f(a + \theta(b - a))$$

pro nějaké $\theta \in [0, 1]$. Umocňováním hranatých závorek se rozumí umocňování (tedy skládání) tohoto diferenciálního operátoru. K důkazu stačí rozepsat $g(b)$, kde $g(t) = f(a + t(b - a))$, podle jednorozměrného Taylorova vzorce a vyjádřit $g^{(j)}$ pomocí f .

$$\begin{aligned}
f(x + \alpha_2 h, y + \beta_{21} h f(x, y)) &= f(x, y) + \left(\alpha_2 h \frac{\partial}{\partial x} + \beta_{21} h f \frac{\partial}{\partial y} \right) f(x, y) + \\
&+ \frac{1}{2!} \left(\alpha_2 h \frac{\partial}{\partial x} + \beta_{21} h f \frac{\partial}{\partial y} \right)^2 f(x + \theta h, y + \theta \beta_{21} h f(x, y)) = \\
&= f(x, y) + \alpha_2 h \frac{\partial f}{\partial x}(x, y) + \beta_{21} h f(x, y) \frac{\partial f}{\partial y}(x, y) + \\
&+ \frac{1}{2} \left[\alpha_2^2 h^2 \frac{\partial^2 f}{\partial x^2} + 2\alpha_2 \beta_{21} h^2 f(x, y) \frac{\partial^2 f}{\partial x \partial y} + \beta_{21}^2 h^2 f^2(x, y) \frac{\partial^2 f}{\partial y^2} \right] \cdot \\
&\cdot (x + \theta h, y + \theta \beta_{21} h f(x, y)).
\end{aligned}$$

Tento poslední sčítanec ovšem není nic jiného než $O(h^2)$. Dostáváme tedy

$$\Phi(x, y, z) = (w_1 + w_2)f(x, y) + w_2 \alpha_2 h \frac{\partial f}{\partial x}(x, y) + w_2 \beta_{21} h f(x, y) \frac{\partial f}{\partial y}(x, y) + O(h^2).$$

Srovnáme nyní koeficienty u f , $\frac{\partial f}{\partial x}$, $f \frac{\partial f}{\partial y}$ ve vyjádření Φ a $\tilde{\Phi}$. Budou-li stejné, bude i $\Phi - \tilde{\Phi} = O(h^2)$. Získáváme rovnice

$$w_1 + w_2 = 1, \quad w_2 \alpha_2 = \frac{1}{2}, \quad w_2 \beta_{21} = \frac{1}{2},$$

tj. tři rovnice pro čtyři neznámé. Jednu neznámou si tedy můžeme zvolit; zvolme $w_2 = \alpha \neq 0$ (jinak by těžko $\alpha_2 w_2 = \frac{1}{2}$). Dostaneme

$$w_1 = 1 - \alpha, \quad w_2 = \alpha, \quad \alpha_2 = \beta_{21} = \frac{1}{2\alpha}.$$

Tyto hodnoty nám dávají celou třídu Runge-Kuttových metod druhého řádu. Nejpoužívanější hodnoty parametru α jsou 1 , $\frac{3}{4}$ a $\frac{1}{2}$.

$$\begin{aligned}
\alpha = 1: \quad y_{n+1} &= y_n + h f(x_n + \frac{h}{2}, y_n + \frac{h}{2} f(x_n, y_n)) \\
\alpha = \frac{3}{4}: \quad y_{n+1} &= y_n + \frac{h}{4} \left(f(x_n, y_n) + 3f(x_n + \frac{2}{3}h, y_n + \frac{2}{3}h f(x_n, y_n)) \right) \\
\alpha = \frac{1}{2}: \quad y_{n+1} &= y_n + \frac{h}{2} (f(x_n, y_n) + f(x_n + h, y_n + h f(x_n, y_n)))
\end{aligned}$$

V případě, že máme více metod stejného řádu, vybíráme metodu buď tak, aby koeficienty byly co nejjednodušší (jednoduchá metoda), nebo tak, aby konstanta N v odhadu

$$\left| \frac{1}{h} (y(x+h) - y(x)) - \Phi(x, y(x), h) \right| \leq N h^p$$

byla co nejmenší (přesná metoda).

Runge-Kuttovy metody 3. řádu

V tomto případě postačuje $P = 3$, tedy $\Phi(x, y, h) = w_1k_1 + w_2k_2 + w_3k_3$. Uvedeme dva příklady:

(a)

$$y_{n+1} = y_n + h \left(\frac{2}{9}k_1 + \frac{1}{3}k_2 + \frac{4}{9}k_3 \right)$$

$$k_1 = f(x, y)$$

$$k_2 = f\left(x + \frac{h}{2}, y + \frac{h}{2}k_1\right)$$

$$k_3 = f\left(x + \frac{3}{4}h, y + \frac{3}{4}hk_2\right)$$

(b)

$$y_{n+1} = y_n + \frac{h}{6}(k_1 + 4k_2 + k_3)$$

$$k_1 = f(x, y)$$

$$k_2 = f\left(x + \frac{h}{2}, y + \frac{h}{2}k_1\right)$$

$$k_3 = f(x + h, y - hk_1 + 2hk_2)$$

Runge-Kuttovy metody 4. řádu

Metody čtvrtého řádu jsou nejlépe rozpracovány. U řádu 4 máme naposledy $P = p$, tedy $\Phi(x, y, h) = w_1k_1 + w_2k_2 + w_3k_3 + w_4k_4$. Uvedeme tři příklady: *standardní formule:*

$$w_1 = w_4 = \frac{1}{6}, \quad w_2 = w_3 = \frac{1}{3}$$

$$k_1 = f(x, y)$$

$$k_2 = f\left(x + \frac{h}{2}, y + \frac{h}{2}k_1\right)$$

$$k_3 = f\left(x + \frac{3}{4}h, y + \frac{3}{4}hk_2\right)$$

$$k_4 = f(x + h, y + hk_3)$$

tříosminová formule:

$$w_1 = w_4 = \frac{1}{8}, \quad w_2 = w_3 = \frac{3}{8}$$

$$\alpha_2 = \frac{1}{3}, \quad \alpha_3 = \frac{3}{2}, \quad \alpha_4 = 1$$

$$\beta_{21} = \frac{1}{3}, \quad \beta_{31} = -\frac{1}{3}, \quad \beta_{32} = 1$$

$$\beta_{41} = 1, \quad \beta_{42} = -1, \quad \beta_{43} = 1$$

Gillova formule:

$$\begin{aligned} w_1 = w_4 = \frac{1}{6}, \quad w_2 = \frac{1}{3} \left(1 - \frac{1}{\sqrt{2}}\right), \quad w_3 = \frac{1}{3} \left(1 + \frac{1}{\sqrt{2}}\right) \\ \alpha_2 = \alpha_3 = \frac{1}{2}, \quad \alpha_4 = 1 \\ \beta_{21} = \frac{1}{2}, \quad \beta_{31} = \frac{1}{2}(\sqrt{2} - 1), \quad \beta_{32} = 1 - \frac{1}{\sqrt{2}} \\ \beta_{41} = 0, \quad \beta_{42} = -\frac{1}{\sqrt{2}}, \quad \beta_{43} = 1 + \frac{1}{\sqrt{2}} \end{aligned}$$

Gillova formule je sice o něco složitější, ale poskytuje také lepší výsledky, neboť u ní byla provedena minimalizace konstanty N . Odvození Runge-Kuttových metod řádu $p > 4$ studoval Huřa (Bratislava). Pro $p > 4$ vyjde $P > p$ (např. pro $p = 6$ musíme vzít osm členů), metody jsou tedy složitější, a proto se moc nepoužívají.

2.4 Chyby metod

Podívejme se nyní na efektivnost odhadu, který jsme získali. Uvažujme Eulerovu metodu. Ukázali jsme, že pokud je přesné řešení třídy C^2 , je

$$|e_n| \leq N E_L(x_n - a)h,$$

kde N je polovina maxima absolutní hodnoty druhé derivace přesného řešení y na intervalu $[a, b]$. Tento odhad se dá o něco zlepšit, uvědomíme-li si, že pro odhad chyby v bodě x_n stačí řešit rovnici na $[a, x_n]$ (interval $(x_n, b]$ nemá žádný vliv). Můžeme tedy položit $b = x_n$, $N(x_n) = \frac{1}{2} \max_{[a, x_n]} |y''|$ a dostaneme zlepšený odhad

$$|e_n| \leq N(x_n) E_L(x_n - a)h.$$

Uvažujme nyní praktický příklad:

$$\begin{array}{ll} y_0 = 1, \quad y' = y & y_0 = 1 \\ y(x) = e^x & y(x) = e^{-x} \\ L = 1 & L = 1 \\ N(x) = \frac{1}{2}e^x & N(x) = \frac{1}{2} \\ E_1(x_n) = e^{x_n} - 1 & E_1(x_n) = e^{x_n} - 1 \\ |e_n| \leq \frac{1}{2}he^{x_n}(e^{x_n} - 1) & |e_n| \leq \frac{1}{2}h(e^{x_n} - 1) \end{array}$$

Řekněme, že pomocí numerického řešení druhé úlohy chceme spočítat e^{-5} s přesností 10^{-3} . Chceme tedy najít h tak, aby

$$\begin{aligned} \frac{1}{2}h(e^5 - 1) &\leq 10^{-3} \\ h &\leq \frac{2 \cdot 10^{-3}}{e^5 - 1} \doteq \frac{1}{73707}. \end{aligned}$$

Experimentálně však zjistíme, že při volbě $h = 2^{-6} = \frac{1}{64}$ vyjde $e_n = -0,000261$. Vidíme tedy, že odhad z věty je silně nadsazený; při jeho odvozování se v každém kroku počítá s nejhorsí možností. Ve skutečnosti většinou dostáváme výsledky výrazně lepší.

2.4.1 *Odhad chyby metodou polovičního kroku*

Víme, že pokud Φ je L -lipschitzovská v y a pokud

$$\left| \frac{1}{h}(y(x+h) - y(x)) - \Phi(x, y, h) \right| \leq Nh^p,$$

potom platí

$$|y_n - y(x_n)| \leq NE_L(x_n - a)h^p.$$

Dá se ukázat, že pokud jsou f a Φ dostatečně hladké, existuje funkce $e : [a, b] \rightarrow \mathbb{R}$ taková, že

$$|e_n| = h^p e(x_n) + O(h^{p+1}).$$

Tento vztah se nazývá *asymptotický odhad chyby*, protože chybu neopdhaduje, ale spíše aproximuje. Funkce e samozřejmě opět závisí na vlastnostech přesného řešení y .

Předpokládejme, že na tutéž úlohu použijeme postupně metodu s krokem $2h$ a h . Dostaneme dělicí body a přibližná řešení v nich

$$\begin{array}{ll} x_n^{(2h)} = a + n(2h) & y_n^{(2h)} \\ x_n^{(h)} = a + nh & y_n^{(h)} \end{array}$$

Vidíme, že $x_n^{(2h)} = x_{2n}^{(h)}$. V těchto bodech můžeme srovnat přibližná řešení:

$$\begin{aligned} e_n^{(2h)} &= y_n^{(2h)} - y(x_n^{(2h)}) = (2h)^p e(x_n^{(2h)}) + O(h^{p+1}) \\ e_{2n}^{(h)} &= y_{2n}^{(h)} - y(x_{2n}^{(h)}) = h^p e(x_n^{(2h)}) + O(h^{p+1}) \end{aligned}$$

Odtud dostáváme

$$\begin{aligned} y_n^{(2h)} - y_{2n}^{(h)} &= (2^p - 1)h^p e(x_n^{(2h)}) + O(h^{p+1}) \\ h^p e(x_n^{(2h)}) &= \frac{y_n^{(2h)} - y_{2n}^{(h)}}{2^p - 1} + O(h^{p+1}) \\ e_{2n}^{(h)} &= \frac{y_n^{(2h)} - y_{2n}^{(h)}}{2^p - 1} + O(h^{p+1}) \sim \frac{y_n^{(2h)} - y_{2n}^{(h)}}{2^p - 1} \end{aligned}$$

Experimenty ukazují, že v praktických aplikacích tento odhad funguje velice dobře.

Poznámky: U jednokrokových metod můžeme v libovolném bodě změnit délku kroku. Pracujeme tak, že počítáme s krokem $2h$ a h . Pokud zjistíme, že chyba v nějakém bodě překračuje povolenou mez, pokračujeme od tohoto bodu s polovičním krokem. Naopak, pokud je chyba výrazně nižší než naše tolerance, můžeme krok opět zvětšit (abychom ušetřili časovou náročnost).

Existují dokonce *adaptivní metody*, u nichž se automaticky mění délka kroku a řád metody s cílem minimalizovat chybu a časovou náročnost.

2.4.2 Zaokrouhlovací chyby

Uvažujme soustavu dvou lineárních rovnic

$$\begin{pmatrix} 10^{-6} & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

Determinant matice soustavy je $10^{-6} - 1$, což je dostatečně daleko od nuly. Soustava je tedy regulární; použijeme-li Gaussovu eliminaci, vyjde

$$x_1 = (1 - x_2) 10^6, \quad x_2 = \frac{2 - 10^6}{1 - 10^6}.$$

Přibližně $x_2 \doteq 0,999999$. Počítejme nyní na pět platných číslic. Pak ovšem vyjde $x_2 \doteq 1$, a odtud $x_1 = 0$. Při zkoušce dostaneme v první rovnici $0 + 1 = 1$, ale ve druhé $0 + 1 = 1$, což je příliš velká nepřesnost. Pokud bychom ale rovnice prohodili, vyšly by přibližné hodnoty x_1 i x_2 rovny jedné, tedy správně (i zkouška by ve zvolené přesnosti vyšla správně). Obecně je vhodné při řešení soustavy lineárních rovnic prohazovat rovnice tak, aby se na hlavní diagonále neobjevovala příliš malá čísla (*pivotace*).

Zaokrouhlovací chyby u jednokrokových metod

Dosud jsme pracovali při počítání přibližného řešení za předpokladu, že všechny výpočty probíhají přesně. Nyní se zamyslíme nad vlivem zaokrouhlovacích chyb. Zaokrouhlování modelujeme tak, že každému $\alpha \in \mathbb{R}$ přiřadíme jeho zaokrouhlenou hodnotu α^* . Pro jednoduchost předpokládejme, že vstupní hodnoty jsou už zaokrouhlené, tedy $a = a^*$, $h = h^*$, $\eta = \eta^*$. Přibližné řešení se zaokrouhlovacími chybami označme \tilde{y}_n . Pak máme

$$\begin{aligned} \tilde{y}_0 &= \eta = y_0, \\ \tilde{y}_{n+1} &= \tilde{y}_n + h\Phi(x_n, \tilde{y}_n, h) + \varepsilon_{n+1}. \end{aligned}$$

Číslo ε_{n+1} se nazývá *lokální zaokrouhlovací chyba*. Zaveďme ještě *akumulovanou zaokrouhlovací chybu* v uzlu x_n jako $r_n = \tilde{y}_n - y_n$.

Věta 6 *Bud' $\Phi(x, y, h) : [a, b] \times \mathbb{R} \times [a, b] \rightarrow \mathbb{R}$ L -lipschitzovská v y . Nechť platí $|\varepsilon_k| \leq \varepsilon$. Potom*

$$|r_n| \leq \frac{\varepsilon}{h} E_L(x_n - a).$$

Důkaz: Máme $r_0 = 0$ a

$$\begin{aligned} \tilde{y}_{n+1} &= \tilde{y}_n + h\Phi(x_n, \tilde{y}_n, h) + \varepsilon_{n+1} \\ y_{n+1} &= y_n + h\Phi(x_n, y_n, h) \\ |r_{n+1}| &= |r_n + hLr_n| + |\varepsilon_{n+1}| = |r_n|(1 + hL) + \varepsilon \end{aligned}$$

Použijeme-li (stejně jako u věty o odhadu diskretizační chyby) „kumulační lemma“ ze cvičení (5), dostaneme

$$|r_n| \leq \frac{(1+hL)^n - 1}{hL} \varepsilon \leq \frac{e^{nL} - 1}{L} \frac{\varepsilon}{h} = E_L(x_n - a) \frac{\varepsilon}{h}.$$

Velikost horního odhadu samozřejmě o ničem nesvědčí. Jedná se o nejhorší možný scénář, který může nastat. Ukazuje se však, že v praxi je pro malá h opravdu $r_n \approx \frac{1}{h}$. To ve svém důsledku znamená, že zmenšením kroku sice zmenšíme diskretizační chybu metody, ale zvýšíme vliv zaokrouhlovacích chyb.

Hodilo by se tedy najít způsob, jak zjistit velikost zaokrouhlovací chyby. To je možné, pokud můžeme úlohu současně řešit v jednoduché a dvojnásobné přesnosti. Potom bude zaokrouhlovací chyba při dvojnásobné přesnosti zanedbatelná ve srovnání s jednoduchou přesností a máme tedy přibližnou hodnotu diskretizační chyby. Převažuje-li diskretizační chyba, je třeba krok zjemnit, převažuje-li zaokrouhlovací chyba, je třeba krok zvětšit.

2.5 Soustavy lineárních diferenčních rovnic

Definice 5 Buď $k \in \mathbb{N}$, $F_n : \mathbb{R}^{k+1} \rightarrow \mathbb{R}$ (případně $F_n : \mathbb{C}^{k+1} \rightarrow \mathbb{C}$), $n \in \mathbb{N}_0$. Pak systém vztahů

$$F_n(y_n, \dots, y_{n+k}) = 0, \quad n \in \mathbb{N}_0 \quad (\diamond)$$

nazýváme *soustavou diferenčních rovnic*. Řešením soustavy nazveme posloupnost $(y_n)_{n=0}^\infty$ splňující (\diamond) .

Příklad 6 Mějme soustavu rovnic ($k = 3$)

$$y_{n+3}^2 - \left(y_{n+2}^2 - y_{n+1}^2 + \sqrt[3]{y_n^2} \right)^4 = 0, \quad n \in \mathbb{N}_0.$$

Vidíme, že y_{n+3} lze spočítat z předchozích tří členů. Vyjdeme-li z hodnot y_0, y_1, y_2 , můžeme postupně spočítat y_3, y_4, \dots . Čísla y_0, y_1, y_2 se nazývají *počáteční podmínky*.

Cvičení 8 Kolik řešení má tato soustava pro pevně zvolené počáteční podmínky y_0, y_1, y_2 ?

Definice 6 Jsou-li funkce F_n lineární, mluvíme o *soustavě lineárních diferenčních rovnic*. Můžeme ji psát ve tvaru

$$\sum_{\nu=0}^k \alpha_{\nu n} y_{n+\nu} = \gamma_{n+k}; \quad (*)$$

čísla γ_{n+k} se nazývají *pravé strany*. Jsou-li všechny pravé strany nulové, dostaneme soustavu homogenní. Soustavě $(*)$ můžeme vždy přiřadit soustavu

$$\sum_{\nu=0}^k \alpha_{\nu n} y_{n+\nu} = 0, \quad (**)$$

homogenní soustavu příslušnou k $(*)$.

Lemma 2.1 Uvažujme soustavu (*) a k ní příslušnou homogenní soustavu (**). Pak platí:

- (i) množina V všech řešení (**) je lineární podprostor $\mathbb{R}^{\mathbb{N}}$;
- (ii) jsou-li $(y_n)_{n=0}^{\infty}$ a $(z_n)_{n=0}^{\infty}$ řešení soustavy (*), je jejich rozdíl řešením soustavy (**);
- (iii) je-li $(w_n)_{n=0}^{\infty}$ řešením (**) a $(y_n)_n \in V$, pak $(w_n + y_n)_{n=0}^{\infty}$ je řešením (*);
- (iv) je-li $(w_n)_{n=0}^{\infty}$ řešením (**), pak k libovolnému řešení $(z_n)_{n=0}^{\infty}$ soustavy (*) existuje právě jedno $(y_n)_n \in V$ takové, že $z_n = y_n + w_n$; jinak řečeno, označíme-li W množinu řešení (**), je $W = V + w$.

Důkaz si laskavý čtenář rozmyslí sám.

Definice 7 Řekneme, že soustava (*) je řádu k , jestliže koeficienty α_{kn} , $n \in \mathbb{N}_0$, jsou nemulové.

Definice 8 Posloupnosti $\{y_n^\mu\}_{n \in \mathbb{N}_0} \in V$, $\mu = 1, \dots, m$, nazveme lineárně nezávislé, jestliže platí:

$$\sum_{\mu=1}^m a_\mu y_n^\mu = 0, \quad \forall n = 0, \dots, k-1 \Rightarrow a_1 = a_2 = \dots = a_m = 0.$$

Je zřejmé, že tato podmínka je splněna, právě když matice

$$\mathbb{M} = (y_n^\mu)_{\substack{n=0, \dots, k-1 \\ \mu=1, \dots, m}}$$

má hodnotu m . Maximální počet lineárně nezávislých prvků z V je roven k .

Definice 9 Systém k prvků z V lineárně nezávislých nazveme *fundamentálním systémem* (FS) soustavy homogenních rovnic.

Věta 7 Necht' $\{z_n^\mu\}_{n \in \mathbb{N}_0} \in V$, $\mu = 1, \dots, k$, je FS řešení soustavy k -tého řádu a $\{y_n\}_{n \in \mathbb{N}_0} \in V$. Pak posloupnost $\{y_n\}_{n \in \mathbb{N}_0}$ je lineární kombinací posloupností $\{z_n^\mu\}_{n \in \mathbb{N}_0}$, $\mu = 1, \dots, k$.

Důkaz: Z definice FS plyne, že existují konstanty a_1, \dots, a_k takové, že

$$y_n = \sum_{\mu=1}^k a_\mu z_n^\mu \quad \text{pro } n = 0, \dots, k-1.$$

Označme

$$z_n = y_n - \sum_{\mu=1}^k a_\mu z_n^\mu, \quad n \in \mathbb{N}_0.$$

Pak $\{z_n\}_{n \in \mathbb{N}_0} \in V$ a $z_0 = \dots = z_{k-1} = 0$. Poněvadž $\alpha_{kn} \neq 0$ pro všechna $n \in \mathbb{N}_0$, nutně $z_n = 0$ pro všechna $n \in \mathbb{N}_0$ a tudíž,

$$y_n = \sum_{\mu=1}^k a_\mu z_n^\mu, \quad \forall n \in \mathbb{N}_0.$$

2.5.1 Nalezení fundamentálního systému

Definice 10 Soustavou tvaru

$$\sum_{\nu=0}^k \alpha_{\nu} y_{n+\nu} = 0, \quad n \in \mathbb{N}_0, \quad (+)$$

nazveme soustavou lineárních diferenčních rovnic s *konstantními koeficienty*. Tudíž,

$$\alpha_{\nu n} = \alpha_{\nu}, \quad \forall \nu = 0, \dots, k \quad n \in \mathbb{N}_0.$$

Hledejme řešení této soustavy ve tvaru $y_n = \xi^n$, kde $\xi \in \mathbb{C}$. Dosazením do soustavy dostaneme

$$0 = \sum_{\nu=0}^k \alpha_{\nu} \xi^{n+\nu} = \xi^n \sum_{\nu=0}^k \alpha_{\nu} \xi^{\nu}, \quad n \in \mathbb{N}_0,$$

což je ekvivalentní s podmínkou

$$0 = \rho(\xi) = \sum_{\nu=0}^k \alpha_{\nu} \xi^{\nu},$$

kterou nazýváme *charakteristickou rovnicí* a polynom ρ je tzv. *charakteristický polynom*.

Nechť soustava (+) je řádu k , tj. $\alpha_k \neq 0$. Pak ρ má stupeň k . Je zřejmé, že $\{y_n\}_{n \in \mathbb{N}_0} = \{\xi^n\}_{n \in \mathbb{N}_0}$ je řešením soustavy (+), právě když ξ je kořenem charakteristického polynomu ρ . Tento polynom má právě k kořenů, počítáme-li každý tolikrát, kolik činí jeho násobnost.

Rozlišme dva případy

- 1) *Polynom ρ má právě k různých kořenů ξ_1, \dots, ξ_k .*

Pak lze sestavit k řešení soustavy $\{y_n^{\mu}\}_{n \in \mathbb{N}_0}$, $\mu = 1, \dots, k$, kde $y_n^{\mu} = \xi_{\mu}^n$.

Podle definice tvoří tato řešení FS, právě když matice

$$\mathbb{M} = \begin{pmatrix} 1 & \xi_1 & \dots & \xi_1^{k-1} \\ 1 & \xi_2 & \dots & \xi_2^{k-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \xi_k & \dots & \xi_k^{k-1} \end{pmatrix}$$

je regulární. Její determinant (známý jako Vandermondeův) má hodnotu

$$\det \mathbb{M} = \prod_{\substack{i < j \\ j=1, \dots, k}} (\xi_j - \xi_i) \neq 0.$$

Tudíž uvažované posloupnosti tvoří FS.

2) *Obecný případ.*

Věta 8 *Nechť $\xi_1, \dots, \xi_m \in \mathbb{C}$ ($m \leq k$) jsou navzájem různé kořeny charakteristického polynomu ρ s násobnostmi p_1, \dots, p_m (takže $k = \sum_{\mu=1}^m p_\mu$). Pak k posloupností s prvky*

$$(P) \quad \begin{aligned} y_n^{\mu,1} &= \xi_\mu^n, & n \in \mathbb{N}_0, \\ y_n^{\mu,2} &= n\xi_\mu^{n-1}, & n \in \mathbb{N}_0, \\ &\vdots \\ y_n^{\mu,p_\mu} &= n(n-1)\dots(n-p_\mu+2)\xi_\mu^{n-p_\mu+1}, & n \in \mathbb{N}_0, \end{aligned} \quad \mu = 1, \dots, m$$

tvorí FS řešení soustavy (+) řádu k . (Je-li $\xi_\mu = 0$, pak klademe $0 \cdot \xi_\mu^{-\nu} = 0$ pro $\nu > 0$.)

Důkaz: Dokažme, že uvedené posloupnosti jsou řešení uvažované soustavy. Jestliže $\xi_\mu = 0$, pak má charakteristický polynom tvar

$$\rho(\xi) = \sum_{\nu=p_\mu}^k \alpha_\nu \xi^\nu$$

(tudíž $\alpha_0, \dots, \alpha_{p_\mu-1} = 0$) a soustava (+) má tvar

$$\sum_{\nu=p_\mu}^k \alpha_\nu y_{n+\nu} = 0, \quad n \in \mathbb{N}_0.$$

Dosažením snadno ověříme, že posloupnosti (P) mající nenulové prvky pouze na pozicích $n < p_\mu$ jsou řešení.

Nechť $\xi_\mu \neq 0$ je kořen polynomu ρ s násobností p_μ . Pak ξ_μ je kořenem polynomu $f_n(\xi) = \xi^n \rho(\xi)$ násobnosti p_μ pro každé $n \in \mathbb{N}_0$. Tudíž, $p_\mu - 1$ derivací f_n je rovno nule v bodě $\xi = \xi_\mu$, takže

$$\begin{aligned} \sum_{\nu=0}^k \alpha_\nu \xi_\mu^{n+\nu} &= 0, & n \in \mathbb{N}_0, \\ \sum_{\nu=0}^k \alpha_\nu (n+\nu) \xi_\mu^{n+\nu-1} &= 0, & n \in \mathbb{N}_0, \\ &\vdots \\ \sum_{\nu=0}^k \alpha_\nu (n+\nu)(n+\nu-1)\dots(n+\nu-p_\mu+2) \xi_\mu^{n+\nu-p_\mu+1} &= 0, & n \in \mathbb{N}_0, \end{aligned}$$

Tzn., že posloupnosti (P) jsou řešeními soustavy (+).

Dokažme nyní, že řešení (P) jsou lineárně nezávislá, což je ekvivalentní s tím, že příslušná matice \mathbb{M} je regulární. V n -tém sloupci této matice ($n = 0, \dots, k-1$) se nachází prvky

$$\begin{aligned} \xi_1^n, & n\xi_1^{n-1}, \dots, n(n-1)\dots(n-p_1+2)\xi_1^{n-p_1+1}, \dots, \\ \xi_m^n, & n\xi_m^{n-1}, \dots, n(n-1)\dots(n-p_m+2)\xi_m^{n-p_m+1}. \end{aligned}$$

Maticе \mathbb{M} je regulární, právě když její sloupce jsou lineárně nezávislé k -rozměrné vektory. Nechť tomu tak není. Pak existují takové koeficienty a_n , $n = 0, \dots, k-1$, že $\sum_{n=0}^{k-1} |a_n| > 0$ a

$$\begin{aligned} \sum_{n=0}^{k-1} a_n \xi_\mu^n &= 0, \\ \sum_{n=0}^{k-1} a_n n \xi_\mu^{n-1} &= 0, \\ &\vdots \\ \sum_{n=0}^{k-1} a_n n(n-1) \dots (n-p_\mu+2) \xi_\mu^{n-p_\mu+1} &= 0, \end{aligned} \quad \mu = 1, \dots, m,$$

což znamená, že polynom $\sum_{n=0}^{k-1} a_n \xi^n$ stupně $\leq k-1$ má aspoň m různých kořenů ξ_1, \dots, ξ_m s násobnostmi p_1, \dots, p_m , což je spor, protože platí, že $\sum_{\mu=1}^m p_\mu = k$.

2.5.2 Nalezení reálného fundamentálního systému

Uvažujme rovnici

$$\sum_{\nu=0}^k \alpha_\nu y_{n+\nu} = 0$$

s reálnými koeficienty α_j . Charakteristický polynom má pak také reálné koeficienty, jeho kořeny tedy jsou

$$\begin{aligned} \xi_1, \dots, \xi_r &\in \mathbb{R} \\ \gamma_1, \overline{\gamma_1}, \dots, \gamma_s, \overline{\gamma_s} &\in \mathbb{C} - \mathbb{R} \end{aligned}$$

(všechny kořeny opakujeme tolikrát, kolik činí jejich násobnost, tedy $r+2s = k$). Fundamentální systém získaný z naší věty je tvořen posloupnostmi (vyplývá z chování aritmetických operací na \mathbb{C})

$$\begin{aligned} (y_n^\mu)_{n=0}^\infty & \quad \mu = 1, \dots, r & \quad y_n^\mu \in \mathbb{R} \\ (z_n^\mu)_{n=0}^\infty & \quad \mu = 1, \dots, s \\ (\overline{z_n^\mu})_{n=0}^\infty & \quad \mu = 1, \dots, s & \quad z_n^\mu \in \mathbb{C} \end{aligned}$$

Reálné posloupnosti z prvního řádku můžeme převzít do našeho nového systému. Jako zbylých $2s$ posloupností vezmeme

$$(\operatorname{Re} z_n^\mu)_{n=0}^\infty, \quad (\operatorname{Im} z_n^\mu)_{n=0}^\infty, \quad \mu = 1, \dots, s.$$

Ze vztahů $\operatorname{Re} z = \frac{1}{2}(z + \bar{z})$, $\operatorname{Im} z = \frac{1}{2}(z - \bar{z})$ vidíme snadno, že nové posloupnosti jsou lineárními kombinacemi starých, jsou to tedy řešení. K důkazu lineární nezávislosti si stačí uvědomit, že staré posloupnosti lze podobně jednoduše vyjádřit pomocí nových použitím vztahů $z = \operatorname{Re} z + i \operatorname{Im} z$, $\bar{z} = \operatorname{Re} z - i \operatorname{Im} z$.

Poznámky: Kniha [H] obsahuje kapitolu věnovanou soustavám diferencních rovnic. Je vidět, že teorie soustav lineárních diferencních rovnic s konstantními koeficienty se podobá teorii lineárních diferencálních rovnic s konstantními koeficienty. Stejně tak existuje i metoda pro nalezení partikulárního řešení nehomogenní soustavy pomocí fundamentálního systému řešení příslušné homogenní rovnice (*Duhamelův princip*). Tato metoda se podobá variaci konstant, blíže viz [H].

2.6 Vícekrokové metody

Opět budeme hledat přibližné řešení úlohy

$$y' = f(x, y), \quad y(a) = \eta$$

v uzlech $x_n = a + nh$. Na rozdíl od jednokrokových metod budeme tentokrát počítat hodnoty přibližného řešení y_n pomocí několika předchozích hodnot.

Příklad 7 Dříve jsme odvodili metodu

$$-y_{n+2} + 4y_{n+1} - 3y_n = 2hf(x_n, y_n).$$

To nás vede k zobecnění: obecnou k -krokovou metodou rozumíme rekurentní předpis typu

$$\sum_{\nu=0}^k \alpha_\nu y_{n+\nu} = h \sum_{\nu=0}^k \beta_\nu f_{n+\nu}, \quad (+)$$

kde $f_j = f(x_j, y_j)$, za podmínek $\alpha_k \neq 0$ a $|\alpha_0| + |\beta_0| > 0$.

Pomocí předpisu (+) můžeme počítat hodnoty přibližného řešení až od y_k . Hodnoty $y_0 (= \eta), y_1, \dots, y_{k-1}$ se nazývají počáteční podmínky pro k -krokovou metodu. Hodnoty y_1, \dots, y_{k-1} spočítáme některou jednokrokovou metodou. Podle způsobu výpočtu je třeba rozlišit dva případy.

V jednodušším případě $\beta_k = 0$ máme

$$\begin{aligned} \sum_{\nu=0}^k \alpha_\nu y_{n+\nu} &= h \sum_{\nu=0}^{k-1} \beta_\nu f_{n+\nu} = h \sum_{\nu=0}^{k-1} \beta_\nu f(x_{n+\nu}, y_{n+\nu}) \\ y_{n+k} &= \frac{1}{\alpha} \sum_{\nu=0}^{k-1} (h\beta_\nu f_{n+\nu} - \alpha_\nu y_{n+\nu}). \end{aligned}$$

Protože lze hodnotu y_{n+k} přímo spočítat, mluvíme o *explicitní metodě*. Hlavní výhodou těchto (např. oproti jednokrokovým metodám vyšších řádů) je to, že každou z hodnot f_j stačí spočítat jednou (v dalších krocích už si ji pamatujeme). Na jednu hodnotu přibližného řešení tak připadá jedna spočítaná hodnota funkce f .

Ve složitějším případě $\beta_k \neq 0$ máme

$$y_{n+k} = \frac{\beta_k}{\alpha_k} hf(x_{n+k}, y_{n+k}) + \sum_{\nu=0}^{k-1} \left(\frac{\beta_\nu}{\alpha_k} hf(x_{n+\nu}, y_{n+\nu}) - \frac{\alpha_\nu}{\alpha_k} y_{n+\nu} \right).$$

Hodnotu y_{n+k} tedy přímo nespočítáme, dostaneme pouze rovnici pro ni: mluvíme o *implicitní metodě*. Naše rovnice (obecně nelineární) pro y_{n+k} je tvaru $y_{n+k} = F(y_{n+k})$, kterou řešíme iterační metodou: zvolíme počáteční hodnotu y_{n+k}^0 a pro $s \geq 0$ počítáme $y_{n+k}^{s+1} = F(y_{n+k}^s)$. Pokud f je L -lipschitzovská v y ,

pak pro dostatečně malé h je F kontrakce (lipschitzovské zobrazení s konstantou menší než 1) a podle Banachovy věty o kontrakci je

$$y_{n+k} = \lim_{s \rightarrow \infty} y_{n+k}^s$$

pro libovolnou počáteční hodnotu y_{n+k}^0 .

Pro praktické použití je však potřeba odpovědět na dvě otázky: jak volit počáteční hodnotu a kolik iterací provést. Pro volbu počáteční hodnoty je nejjednodušší vzít buď přímo předchozí hodnotu ($y_{n+k}^0 = y_{n+k-1}$) nebo extrapolovat z předchozích dvou ($y_{n+k}^0 = 2y_{n+k-1} - y_{n+k-2}$). Počet iterací se většinou provádí takový, aby metoda byla výhodná. Protože máme poměrně efektivní jednokrokové metody čtvrtého řádu, kde v každém kroku počítáme čtyři hodnoty f , končí se většinou u první, druhé nebo třetí iterace.

Počáteční aproximaci y_{n+k}^0 lze také získat pomocí explicitní metody:

$$\sum_{\nu=0}^k \alpha_{\nu}^* y_{n+\nu} = h \sum_{\nu=0}^{k-1} \beta_{\nu}^* f_{n+\nu}$$

$$y_{n+k}^0 = \frac{1}{\alpha_k^*} \left(h \sum_{\nu=0}^{k-1} \beta_{\nu}^* f_{n+\nu} - \sum_{\nu=0}^{k-1} \alpha_{\nu}^* y_{n+\nu} \right). \quad (1)$$

Potom spočítáme několik aproximací y_{n+k}^s uvažované implicitní metody a klademe $y_{n+k} := y_{n+k}^m$, přičemž obvykle volíme $m = 1$:

$$y_{n+k} = \alpha_k \left(h \beta_k f(x_{n+k}, y_{n+k}^0) + \sum_{\nu=0}^{k-1} \beta_{\nu} f_{n+\nu} - \sum_{\nu=0}^{k-1} \alpha_{\nu} y_{n+\nu} \right). \quad (2)$$

Z rovnice (1) spočítáme y_{n+k}^0 , dosadíme do (2) a spočítáme y_{n+k} . Tato metoda se nazývá *metoda prediktor-korektor*. Explicitní metodě (1) se říká *prediktorová formule*, implicitní metodě (2) *korektorová formule*. Metody prediktor-korektor jsou nejefektivnější vícekrokové metody. V každém kroku se počítají pouze dvě hodnoty funkce f .

Cvičení 9 Uvažujme příklad dvoukrokové metody:

$$\begin{aligned} y_{n+1} &= y_n + hf_n && \text{(prediktor)} \\ y_{n+1} &= y_n + \frac{1}{2}h(f_n + f_{n+1}) && \text{(korektor)}. \end{aligned}$$

Zapište tuto metodu ve tvaru jednoho vzorce pro výpočet y_{n+1} . Ukažte, že výsledkem je Runge-Kuttova metoda druhého řádu.

2.7 Některé vlastnosti obecných více krokových metod

Uvažujme obecnou k -krokovou metodu

$$\sum_{\nu=0}^k \alpha_{\nu} y_{n+\nu} = h \sum_{\nu=0}^k \beta_{\nu} f_{n+\nu}, \quad (*)$$

kde $\alpha_k \neq 0$, $|\alpha_0| + |\beta_0| > 0$, $n = 0, 1, \dots$ (dokud $x_{n+k} \in [a, b]$). Na začátku výpočtu je třeba určit počáteční podmínky

$$y_0 = \eta, y_1, \dots, y_{k-1}. \quad (+)$$

Ty se určí pomocí vhodné jednokrokové metody $y_{\mu} = \eta_{\mu}(h)$, $\mu = 1, \dots, k-1$.

Definice 11 Řekneme, že k -kroková metoda (*) je *konvergentní*, jestliže pro řešení $y : [a, b] \rightarrow \mathbb{R}$ úlohy $y' = f(x, y)$, $y(a) = \eta$ s libovolnou f , která je spojitá v $[a, b] \times \mathbb{R}$ a Lipschitzovská v y a pro libovolné η platí následující tvrzení: jestliže y_n je přibližné řešení získané pomocí metody (*) s počátečními podmínkami (+) takovými, že $\lim_{h \searrow 0} \eta_{\mu}(h) = \eta$, pro $\mu = 0, 1, \dots, k-1$, pak platí

$$\forall x \in [a, b] : \lim_{\substack{h \searrow 0 \\ x_n = x}} y_n = y(x).$$

Definice 12 Říkáme, že metoda (*) je *stabilní* (podle Dahlquist), jestliže všechny kořeny polynomu $\rho(\xi) = \sum \alpha_{\nu} \xi^{\nu}$ jsou v absolutní hodnotě nejvýše rovny jedné a všechny kořeny, jejichž absolutní hodnota je rovna jedné, jsou jednoduché.

Definice 13 Řekneme, že metoda (*) je *řádu* $p \geq 0$, jestliže platí

$$\sum_{\nu=0}^k \alpha_{\nu} = 0$$

a pro $p > 0$ navíc

$$\sum_{\nu=0}^k \frac{\alpha_{\nu} \nu^s}{s!} = \sum_{\nu=0}^k \frac{\beta_{\nu} \nu^{s-1}}{(s-1)!} \quad (s = 1, \dots, p).$$

Má-li metoda (*) řád $p \geq 1$, říkáme, že je *konsistentní*.

Věta 9 Každá konvergentní metoda je stabilní.

Důkaz: Mějme konvergentní metodu (*). Přibližné řešení y_n konverguje za podmínek uvedených v definici k přesnému řešení. Uvažujme úlohu $y' = 0$, $y(0) = 0$: její přesné řešení je $y(x) = 0$ (tato úloha určitě podmínky splňuje). Pak máme

$$\sum \alpha_{\nu} y_{n+\nu} = 0 \quad n = 0, 1, \dots \quad (**)$$

$$\beta_\mu = \eta_\mu(h) \quad \mu = 0, 1, \dots, k-1$$

a za předpokladu $\eta_\mu(h) \rightarrow 0$ pro $h \searrow 0$ víme, že

$$\forall x > 0 : \lim_{\substack{h \searrow 0 \\ x_n = x}} y_n = 0.$$

Vztah (**) je soustava lineárních diferenčních rovnic s konstantními koeficienty a charakteristickým polynomem $\rho(\xi) = \sum \alpha_\nu \xi^\nu$. Nechtě $\xi = re^{i\varphi}$ ($r \geq 0$) je kořen polynomu ρ : pak posloupnost $(\xi^n)_n$ je řešením (**). Protože koeficienty α_ν jsou reálné, víme z teorie, že i $(w_n)_{n=0}^\infty$ a $(z_n)_{n=0}^\infty$, kde $w_n = hr^n \cos n\varphi$ a $z_n = hr^n \sin n\varphi$, jsou řešení (**). Jsou to tedy přibližná řešení vyhovující počátečním podmínkám

$$hr^\mu \cos \mu\varphi \xrightarrow{h \searrow 0} 0, \quad hr^\mu \sin \mu\varphi \xrightarrow{h \searrow 0} 0$$

Odtud pro každé $x > 0$

$$\begin{aligned} \lim_{\substack{h \searrow 0 \\ x_n = x}} z_n = \lim_{\substack{h \searrow 0 \\ x_n = x}} w_n = 0 &\iff \lim_{\substack{h \searrow 0 \\ x_n = x}} |z_n + iw_n| = 0 \iff \lim_{\substack{h \searrow 0 \\ x_n = x}} hr^n = 0 \iff \\ &\iff x \lim_{n \rightarrow \infty} \frac{1}{n} r^n = 0 \iff |\xi| = r \leq 1. \end{aligned}$$

Předposlední ekvivalence vychází z toho, že ta poněkud záhadně vyhlížející limita není ve skutečnosti nic jiného než limita pro $n \rightarrow \infty$ s dosazením $h = \frac{x}{n}$.

Nyní mějme kořen $\xi = re^{i\varphi}$ násobnosti alespoň dvě. Potom je $(n\xi^{n-1})_n$ řešením (**). Stejně jako výše vidíme, že i z_n a w_n , kde $z_n = n\sqrt{h}r^n \cos n\varphi$ a $w_n = n\sqrt{h}r^n \sin n\varphi$ jsou řešení. Jsou to tedy přibližná řešení pro počáteční podmínky

$$\mu\sqrt{h}r^\mu \cos \mu\varphi \rightarrow 0, \quad \mu\sqrt{h}r^\mu \sin \mu\varphi \rightarrow 0 \quad (h \searrow 0).$$

Odtud stejným způsobem jako v první části dostaneme

$$\sqrt{x} \lim_{n \rightarrow \infty} r^n \sqrt{n} = 0 \iff |\xi| = r < 1.$$

Věta 10 *Je-li metoda (*) konvergentní, pak je také konsistentní, tj.*

$$\sum_{\nu=0}^k \alpha_\nu = 0, \quad \sum_{\nu=0}^k \alpha_\nu \nu = \sum_{\nu=0}^k \beta_\nu.$$

Důkaz: Uvažujme úlohu $y' = 0$ v $[0, b]$, $b > 0$, tentokrát s počáteční podmínkou $y(0) = 1$. Pro přibližné řešení opět platí

$$\sum_{\nu=0}^k \alpha_\nu y_{n+\nu} = 0 \quad n = 0, 1, \dots \quad (**)$$

Pro libovolné $h > 0$ volme počáteční podmínky $y_0 = y_1 = \dots = y_{k-1} = 1$. Zřejmě je splněna podmínka $y_\mu = \eta_\mu(h) = 1 \rightarrow 1$ pro $h \searrow 0$. Metoda (*) je konvergentní, tedy

$$\forall x > 0: \lim_{\substack{h \searrow 0 \\ x_n = x}} = 0, \quad (+)$$

přičemž limitu opět chápeme jako limitu pro $n \rightarrow \infty$, kde y_n je hodnota přibližného řešení s krokem $h = \frac{x}{n}$. Počáteční podmínky ani koeficienty rovnice (**) nezávisí na kroku h , nezávisí na něm tedy ani přibližné řešení. Ze vztahu (+)

$$\lim_{n \rightarrow \infty} y_n = 1 \implies \lim_{n \rightarrow \infty} y_{n+\nu} = 1 \quad \nu = 0, \dots, k.$$

Provedeme-li tedy limitní přechod v (**), dostaneme $\sum \alpha_\nu = 0$.

Pro druhou podmínku uvažujme úlohu $y' = 1$, $y(0) = 0$ (řešení je $y(x) = x$). Metoda (*) tady vypadá takto:

$$\sum_{\nu=0}^k \alpha_\nu y_{n+\nu} = h \sum_{\nu=0}^k \beta_\nu \quad n = 0, 1, \dots \quad (***)$$

Pro počáteční podmínky chceme $y_\mu = \eta_\mu(h) \rightarrow 0$ pro $h \searrow 0$, $\mu = 0, 1, \dots, k-1$. Hledejme řešení (***) ve tvaru $y_n = Kx_n = Khn$. Dosaďme do (***):

$$\begin{aligned} Kh \sum \alpha_\nu (n + \nu) &= h \sum \beta_\nu \\ Kh \sum \alpha_\nu n + Kh \sum \alpha_\nu \nu &= h \sum \beta_\nu \\ K \sum \alpha_\nu \nu &= \sum \beta_\nu. \end{aligned}$$

Víme, že $\sum \alpha_\nu = 0$, tedy $\rho(1) = 0$, z předcházející věty vyplývá, že je to kořen jednoduchý. Tedy $\rho'(1) \neq 0$, což znamená $\sum \alpha_\nu \nu \neq 0$, můžeme tedy poslední rovnost touto sumou vydělit:

$$K = \frac{\sum \beta_\nu}{\sum \alpha_\nu \nu}.$$

Protože úpravy byly ekvivalentní, je $y_n = Kx_n$ pro tuto hodnotu K opravdu řešením (***) . Zvolíme-li počáteční podmínky metody $\eta_\mu(h) = K\mu h$ (vyhovují), bude $y_n = Kx_n$ přibližným řešením, získaným metodou (*). Metoda je konvergentní, máme tedy pro každé x

$$\lim_{\substack{h \searrow 0 \\ x_n = x}} y_n = y(x) = x \iff K \lim_{n \rightarrow \infty} n \frac{x}{n} = x \iff K = 1 \iff \sum \alpha_\nu \nu = \sum \beta_\nu.$$

Příklad 8 Představme si pružnou blánu, na okraji upevněnou, na niž působí kolmo nějaká síla. Modelem bude oblast (otevřená souvislá množina) $\Omega \subseteq \mathbb{R}^2$, sílu budeme reprezentovat hustotou jejího rozložení, tedy funkcí $f : \Omega \rightarrow \mathbb{R}$ a výchylku blány hledanou funkcí $u : \bar{\Omega} \rightarrow \mathbb{R}$. Za předpokladu, že jsme dosáhli

stabilního stavu (a že výchylka je malá) bude u splňovat vztah $-\Delta u = f$, kde Δ je takzvaný *Laplaceův operátor* definovaný vztahem

$$\Delta u = \sum_{i=1}^n \frac{\partial^2 u}{\partial x_i^2} \quad (h : \mathbb{R}^n \rightarrow \mathbb{R}).$$

Upevnění na okrajích nám dává okrajovou podmínku $u \upharpoonright \partial\Omega = 0$. Jednodimenzionální analogie: místo Ω máme interval (a, b) , rovnice je $-u'' = f$ a okrajová podmínka $u(a) = u(b) = 0$. Řešíme úlohu přibližně pro $\Omega = (0, 2) \times (0, 1)$. K tomu potřebujeme situaci diskretizovat: obdélník Ω si rozdělme na $2n \times n$ čtverečků o hraně $h = \frac{1}{n}$. Souřadnice uzlů označme $x_i = y_i = ih$. Hodnoty u, f v uzlu (x_i, y_j) označme u_{ij}, f_{ij} . Okrajová podmínka dává $u_{0,j} = u_{2n,j} = u_{i,0} = u_{i,n} = 0$. Podobně jako derivaci funkce jedné proměnné lze aproximovat podle vztahu $y' \approx \frac{1}{h}(y(x+h) - y(x))$, lze aproximovat nesmíšené parciální derivace druhého řádu:

$$\begin{aligned} \frac{\partial^2 u}{\partial x^2}(x_i, y_j) &\approx \frac{1}{h^2} (u(x_{i-1}, y_j) - 2u(x_i, y_j) + u(x_{i+1}, y_j)) \\ \frac{\partial^2 u}{\partial y^2}(x_i, y_j) &\approx \frac{1}{h^2} (u(x_i, y_{j-1}) - 2u(x_i, y_j) + u(x_i, y_{j+1})) \end{aligned}$$

Dosadíme-li tuto diskretizaci do rovnice, dostaneme

$$-u_{i,j-1} - u_{i,j+1} + 4u_{i,j} - u_{i-1,j} - u_{i+1,j} = h^2 f_{i,j}.$$

Dohromady tedy soustava lineárních rovnic. Naskládáme-li neznámé do jednoho vektoru $\mathbf{u} = (u_{1,1}, \dots, u_{1,n-1}, u_{2,1}, \dots, u_{2n-1,n-1})$ (svislého pro formalisty) a stejným způsobem pravé strany do vektoru \mathbf{f} , dostaneme soustavu $\mathbf{A}\mathbf{u} = \mathbf{f}$, kde

$$A = \begin{pmatrix} B & C & 0 & 0 & \dots & 0 \\ C & B & C & 0 & \dots & 0 \\ 0 & C & B & C & \dots & 0 \\ \vdots & & & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & B \end{pmatrix}, \quad B = \begin{pmatrix} 4 & -1 & 0 & 0 & \dots & 0 \\ -1 & 4 & -1 & 0 & \dots & 0 \\ 0 & -1 & 4 & -1 & \dots & 0 \\ \vdots & & & & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 0 \end{pmatrix}$$

a C je $-E$, kde E je jednotková matice. Matice A se skládá z $(2n-1) \times (2n-1)$ bloků typu $(n-1) \times (n-1)$. Matici s takovou strukturou se říká *blokově třídiagonální matice*. Struktura této matice samozřejmě závisí na očíslování „políček“. Popsané metodě se říká *metoda sítě*.

Příklad 9 Metoda $-y_{n+2} + 4y_{n+1} - 3y_n = 2hf_n$ není stabilní, neboť polynom $\rho(\xi) = -\xi^2 + 4\xi - 3$ má kořeny 1 a 3.

Cvičení 10 Určete řád této metody.

Příklad 10 Metoda $y_{n+2} + 4y_{n+1} - 5y_n = h(4f_{n+1} + 2f_n)$ má sice řád 3, ale není stabilní.

Význam stability

Uvažujme rovnici $y' = -y$, $y(0) = 1$ s řešením $y(x) = e^{-x}$. Zkusíme-li některou z předchozích metod, vidíme, že pro rostoucí n hodnoty přibližného řešení kmitají okolo přesného se zvětšujícím se rozptylem.

Význam řádu

Definice řádu nijak nenaznačuje, k čemu je tento pojem dobrý. To vyplyne z následující věty. Mějme obecnou k -krokovou metodu $(*)$, $y : [a, b] \rightarrow \mathbb{R}$ buď přesné řešení naší úlohy. Dosadíme $y(x_n)$ místo y_n v $(*)$:

$$\sum_{\nu=0}^k \alpha_{\nu} y(x_{n+\nu}) = h \sum_{\nu=0}^k \beta_{\nu} f(x_{n+\nu}, y(x_{n+\nu})) + h\delta_n. \quad (**)$$

Veličinu δ_n nazveme *lokální diskretizační chyba* v bodě x_{n+k} .

Věta 11 *Nechť metoda $(*)$ je řádu p a necht' přesné řešení je třídy $C^{p+1}[a, b]$. Pak existují $h_0, K > 0$ tak, že $|\delta_n| \leq Kh^p$ pro všechna n taková, že $x_n, x_{n+k} \in [a, b]$ a pro všechna $h < h_0$.*

Důkaz: Napišme si Taylorův vzorec:

$$\begin{aligned} y(x_{n+\nu}) &= y(x_n + \nu h) = y(x_n) + \sum_{s=1}^p \frac{y^{(s)}(x_n)}{s!} \nu^s h^s + R_{\nu, p+1}, \\ R_{\nu, p+1} &= \frac{y^{(p+1)}(x_{n, \nu})}{(p+1)!} \nu^{p+1} h^{p+1} = O(h^{p+1}). \end{aligned} \quad (+)$$

Protože y je řešení rovnice, máme $f(x_{n+\nu}, y(x_{n+\nu})) = y'(x_{n+\nu})$; opět použijme Taylorův vzorec:

$$\begin{aligned} hf(x_{n+\nu}, y(x_{n+\nu})) &= hy'(x_n + \nu h) = \sum_{s=1}^p \frac{y^{(s)}(x_n)}{(s-1)!} \nu^{s-1} h^s + \tilde{R}_{\nu, p+1}, \\ \tilde{R}_{\nu, p+1} &= \frac{y^{(p+1)}(\tilde{x}_{n, \nu})}{p!} \nu^p h^{p+1} = O(h^{p+1}). \end{aligned} \quad (++)$$

Body $x_{n, \nu}$ a $\tilde{x}_{n, \nu}$ leží v intervalu $[x_n, x_{n+\nu}] \subseteq [a, b]$. Zbytky R, \tilde{R} jsou $O(h^{p+1})$, protože $y^{(p+1)}$ je spojitá, a tedy v $[a, b]$ omezená; označme Y maximum její absolutní hodnoty. Vztahy $(+)$, $(++)$ dosadíme do $(**)$ a upravíme pomocí definice řádu:

$$\begin{aligned} h\delta_n &= y(x_n) \sum_{\nu=0}^k \alpha_{\nu} + \sum_{s=1}^p h^s y^{(s)}(x_n) \left(\sum_{\nu=0}^k \alpha_{\nu} \frac{\nu^s}{s!} - \sum_{\nu=0}^k \beta_{\nu} \frac{\nu^{s-1}}{(s-1)!} \right) \\ &\quad + \sum_{\nu=0}^k \alpha_{\nu} R_{\nu, p+1} - \sum_{\nu=0}^k \beta_{\nu} \tilde{R}_{\nu, p+1} \\ h\delta_n &= h^{p+1} \left(\sum_{\nu=0}^k \alpha_{\nu} \nu^{p+1} \frac{y^{(p+1)}(x_{n, \nu})}{(p+1)!} - \sum_{\nu=0}^k \beta_{\nu} \nu^p \frac{y^{(p+1)}(\tilde{x}_{n, \nu})}{p!} \right) \end{aligned}$$

Protože je $|y^{(p+1)}|$ na $[a, b]$ odhadnuta konstantou Y , dostáváme

$$|\delta_n| \leq Yh^p \left(\sum \frac{\alpha_\nu \nu^{p+1}}{(p+1)!} - \sum \frac{\beta_\nu \nu^p}{p!} \right) = Kh^p$$

pro $x_n, x_{n+\nu} \in [a, b]$, $h \in (0, h_0)$.

Shrnutí: Význam této věty je následující: je-li metoda řádu p a je-li přesné řešení rovnice dostatečně hladké, platí pro chybu $\delta_n = O(h^p)$. Bez důkazu uveďme následující fundamentální výsledek:

Věta 12 *Více kroková metoda je konvergentní, právě když je stabilní a konsistentní.*

Zaveďme nyní opět chybu metody (akumulovanou diskretizační chybu) jako $e_n = y_n - y(x_n)$. Následující věta dává odhad této chyby.

Věta 13 *Nechť $y \in C^{p+1}([a, b])$ je přesným řešením úlohy*

$$y' = f(x, y) \text{ v } [a, b], \quad y(a) = \eta,$$

kde $f \in C([a, b] \times \mathbb{R})$ splňuje Lipschitzovu podmínku s konstantou $L > 0$ a nechť $\{y_n\}_{x_n \in [a, b]}$ je přibližné řešení vypočtené pomocí více krokové stabilní metody řádu $p \geq 1$. Označme

$$\delta = \max_{\mu=0, \dots, b-1} |y(x_\mu) - y_\mu|.$$

Pak existují konstanty $h_1 > 0$, $\tilde{N} > 0$, $\vartheta > 0$ takové, že pro chybu metody $e_n = y_n - y(x_n)$ platí odhad

$$|e_n| \leq e^{(x_n - a)\vartheta} \delta + E_\vartheta(x_n - a) \tilde{N} h^p$$

$$\forall x_n \in [a, b], \quad \forall h \in (0, h_1).$$

Úplný důkaz je uveden ve skriptech [1] věta 1.12.4. Zde se omezíme na jednodušší případ tzv. *silně stabilních metod*.

Definice 14 *Více kroková metoda je silně stabilní jestliže má tvar*

$$y_{n+k} - y_{n+k-1} = h \sum_{\nu=0}^k \beta_\nu f_{n+\nu}. \quad (M)$$

Tzn., že polynom ρ má pouze kořeny 0 a 1 a kořen 1 je jednonásobný.

Nyní dokážeme předchozí větu o odhadu chyby pro případ silně stabilních metod.

Důkaz: Přibližné řešení splňuje vztahy

$$y_{n+k} - y_{n+k-1} = h \sum_{\nu=0}^k \beta_{\nu} f(x_{n+\nu}, y_{n+\nu}), \quad x_n, x_{n+k} \in [a, b].$$

Pokud je metoda (M) řádu p a přesné řešení je $C^{p+1}([a, b])$, pak

$$y_{n+k} - y_{n+k-1} = h \sum_{\nu=0}^k \beta_{\nu} f(x_{n+\nu}, y(x_{n+\nu})) + h\delta_n, \quad x_n, x_{n+k} \in [a, b].$$

a

$$|\delta_n| \leq Nh^{p+1}, \quad h \in (0, h_0).$$

Označíme-li $e_{n+\nu} = y_{n+\nu} - y(x_{n+\nu})$ chybu metody, z předchozích vztahů dostaneme

$$e_{n+k} - e_{n+k-1} = h \sum_{\nu=0}^k \beta_{\nu} \{f(x_{n+\nu}, y_{n+\nu}) - f(x_{n+\nu}, y(x_{n+\nu}))\} - h\delta_n.$$

Poněvadž funkce f splňuje Lipschitzovu podmínku, máme

$$|e_{n+k}| \leq |e_{n+k-1}| + hL \sum_{\nu=0}^k |\beta_{\nu}| |e_{n+\nu}| + h|\delta_n|$$

a tedy

$$|e_{n+k}|(1 - hL|\beta_k|) \leq |e_{n+k-1}|(1 + hL|\beta_{k-1}|) + hL \sum_{\nu=0}^{k-2} |\beta_{\nu}| |e_{n+\nu}| + Nh^{p+1}.$$

Nechť $h_1 > 0$ je takové, že

$$1 - h_1L|\beta_k| \geq \frac{1}{2}.$$

Potom pro $h \in (0, h_1)$ platí:

$$\frac{1 + hL|\beta_{k-1}|}{1 - hL|\beta_k|} = \frac{1 - hL|\beta_k| + hL(|\beta_k| + |\beta_{k-1}|)}{1 - hL|\beta_k|} \leq 1 + 2hL(|\beta_k| + |\beta_{k-1}|),$$

a tudíž,

$$|e_{n+k}| \leq |e_{n+k-1}| + h \sum_{\nu=0}^{k-1} \vartheta_{\nu} |e_{n+\nu}| + \tilde{\vartheta}, \quad (+)$$

kde

$$\vartheta_{k-1} = 2L(|\beta_k| + |\beta_{k-1}|), \quad \vartheta_{\nu} = 2L|\beta_{\nu}|, \quad \nu = 0, \dots, k-2, \quad \tilde{\vartheta} = 2Nh^{p+1}.$$

Nyní použijeme následující výsledek.

Lemma 2 Necht' $u_n \geq 0$ pro $n = 0, \dots, N$, $\vartheta_\nu \geq 0$ pro $\nu = 0, \dots, k-1$, $\tilde{\vartheta} \geq 0$, $\delta \geq 0$, $h > 0$, $u_0, \dots, u_{k-1} \leq \delta$,

$$u_{n+k} \leq u_{n+k-1} + h \sum_{\nu=0}^{k-1} \vartheta_\nu u_{n+\nu} + \tilde{\vartheta} \quad (*)$$

pro $n = 0, \dots, N-k$. Pak

$$u_n \leq e^{nh\vartheta} \delta + \frac{e^{nh\vartheta} - 1}{\vartheta h} \tilde{\vartheta}, \quad n = 0, \dots, N,$$

kde $\vartheta = \sum_{\nu=0}^{k-1} \vartheta_\nu$.

Důkaz:

1) Jestliže

$$\begin{aligned} \xi_{n+1} &= (1 + h\vartheta)\xi_n + \tilde{\vartheta}, \quad n = 0, \dots, N-1, \\ \xi_0 &= \delta, \end{aligned}$$

pak (jak víme)

$$\xi_n \leq e^{nh\vartheta} \delta + \frac{e^{nh\vartheta} - 1}{\vartheta h} \tilde{\vartheta}, \quad n = 0, \dots, N.$$

2) Dokažme indukci, že $u_n \leq \xi_n$, $n = 0, \dots, N$. Máme

$$u_0, \dots, u_{k-1} \leq \delta = \xi_0 \leq \xi_1 \leq \dots \leq \xi_{k-1} \leq \xi_k \leq \dots$$

Necht' $u_j \leq \xi_j$ pro $j = 0, \dots, n+k-1$, $n \geq 0$. Potom z (*) plyne, že

$$\begin{aligned} u_{n+k} &\leq u_{n+k-1} + h\vartheta_{k-1}u_{n+k-1} + \dots + h\vartheta_0u_n + \tilde{\vartheta} \\ &\leq \xi_{n+k-1} + h\vartheta_{k-1}\xi_{n+k-1} + \dots + h\vartheta_0\xi_n + \tilde{\vartheta} \\ &\leq \xi_{n+k-1} \left(1 + h \sum_{\nu=0}^{k-1} \vartheta_\nu \right) + \tilde{\vartheta} \\ &= \xi_{n+k-1}(1 + h\vartheta) + \tilde{\vartheta} = \xi_{n+k}. \end{aligned}$$

3) Použijeme-li výsledek z 1), dostaneme

$$u_n \leq \xi_n \leq e^{nh\vartheta} \delta + \frac{e^{nh\vartheta} - 1}{\vartheta h} \tilde{\vartheta}, \quad n = 0, \dots, N,$$

což jsme chtěli dokázat.

Nyní můžeme aplikovat lemma na nerovnosti (+). Vzhledem k tomu, že $nh = x_n - a$, $\tilde{\vartheta} = Nh^{p+1}$, snadno zjistíme, že

$$\begin{aligned} |e_n| &\leq e^{(x_n-a)\vartheta} \delta + \frac{e^{(x_n-a)\vartheta} - 1}{\vartheta h} \tilde{N}h^{p+1} = \\ &= e^{(x_n-a)\vartheta} \delta + E_\vartheta(x_n - a)\tilde{N}h^p, \quad x_n \in [a, b]. \end{aligned}$$

□

Podobným způsobem lze odvodit odhad pro akumulovanou zaokrouhlovací chybu. Podobně jako u jednokrokových metod budeme symbolem \tilde{y}_n značit přibližné řešení počítané se zaokrouhlováním. Pak lze psát

$$\tilde{y}_{n+k} - \tilde{y}_{n+k-1} = h \sum_{\nu=0}^k \beta_{\nu} f(x_{n+\nu}, \tilde{y}_{n+\nu}) + \varepsilon_n, \quad x_n, x_{n+k} \in [a, b],$$

kde $|\varepsilon_n| \leq \varepsilon$ pro $x_n, x_{n+k} \in [a, b]$ a $\tilde{y}_{\mu} = y_{\mu}$, $\mu = 0, \dots, k-1$. Přibližné řešení počítané přesně bez zaokrouhlování je dáno předpisem

$$y_{n+k} - y_{n+k-1} = h \sum_{\nu=0}^k \beta_{\nu} f(x_{n+\nu}, y_{n+\nu}), \quad x_n, x_{n+k} \in [a, b].$$

Nyní lze aplikovat postup jako v důkazu předchozí věty, kde $\delta = 0$ a $\tilde{\vartheta} = \varepsilon$. Dostaneme pak odhad akumulované zaokrouhlovací chyby $r_n = y_n - \tilde{y}_n$ ve tvaru

$$|r_n| \leq E_{\vartheta}(x_n - a) \frac{\varepsilon}{h}, \quad x_n \in [a, b].$$

Tento odhad nám dává *nejhorší scénář* pro chování chyby r_n v závislosti na h . Numerické experimenty bohužel ukazují, že pro $h \rightarrow 0+$ zaokrouhlovací chyby mohou růst tak, že znehodnotí výpočet.

2.8 Odvození některých vícekrokových metod

2.8.1 Interpolační polynom a zpětné difference

Nechť $J \subset \mathbb{R}$ je interval, $z : J \rightarrow \mathbb{R}$ a necht' $x_0, \dots, x_q \in J$ ($q \geq 1$) jsou navzájem různé body. Pak existuje právě jeden interpolační polynom P takový, že

- a) stupeň $P \leq q$,
- b) $P(x_i) = z(x_i)$, $i = 0, \dots, q$.

Polynom P lze napsat v Lagrangeově vyjádření

$$P(x) = \sum_{i=0}^q \frac{(x-x_0)\dots(x-x_{i-1})(x-x_{i+1})\dots(x-x_q)}{(x_i-x_0)\dots(x_i-x_{i-1})(x_i-x_{i+1})\dots(x_i-x_q)} z(x_i).$$

Pokud $x_i = x_0 + ih$ pro $i = 0, \dots, q$ a $h > 0$, pak lze vyjádřit P pomocí *zpětných diferencí* funkce z , definovaných následujícím způsobem:

$$\nabla^0 z_i = z_i \quad (\text{diference nultého řádu}),$$

$$\nabla z_i = z_i - z_{i-1} \quad (\text{diference prvního řádu}),$$

dále indukci

$$\nabla^r z_i = \nabla(\nabla^{r-1} z_i) \quad (\text{diference } r\text{-tého řádu}),$$

pokud mají výrazy vpravo smysl. Platí:

$$\nabla^2 z_i = (z_i - z_{i-1}) - (z_{i-1} - z_{i-2}) = z_i - 2z_{i-1} + z_{i-2}.$$

Obecně lze dokázat vztah

$$\nabla^r z_i = \sum_{m=0}^r (-1)^m \binom{r}{m} z_{i-m}, \quad r = 0, 1, \dots$$

Existuje i inverzní formule

$$z_{i-r} = \sum_{m=0}^r (-1)^m \binom{r}{m} \nabla^m z_i, \quad r = 0, 1, \dots$$

Lemma 3 *Interpolační polynom P k funkci z v bodech $x_i = x_0 + ih$, kde i nabývá hodnot $p, p-1, \dots, p-q$, lze napsat ve tvaru*

$$P(x) = \sum_{m=0}^q (-1)^m \binom{-s}{m} \nabla^m z_p, \quad (**)$$

kde $s = (x - x_p)/h$ a

$$\binom{-s}{m} = \frac{(-s)(-s-1)\dots(-s-m+1)}{m!}.$$

Důkaz: Je zřejmé, že polynom na pravé straně vztahu je stupně $\leq q$. Je-li $x = x_{p-r}$, potom $s = -r$ a $\binom{r}{m} = 0$ pro $m > r$. Pak podle inverzní formule

$$P(x_{p-r}) = \sum_{m=0}^q (-1)^m \binom{r}{m} \nabla^m z_p = \sum_{m=0}^r (-1)^m \binom{r}{m} \nabla^m z_p = z_{p-r}.$$

2.8.2 Metody založené na numerické integraci

Přesné řešení rovnice $y' = f(x, y)$ splňuje vztah

$$y(x + \tau) - y(x) = \int_x^{x+\tau} f(t, y(t)) dt,$$

pro $x, x + \tau \in [a, b]$. Nyní nahradíme funkci $f(t, y(t))$ interpolačním polynomem majícím v uzlech x_n hodnoty $f_n = f(x_n, y_n)$, kde y_n bylo vypočteno nebo má být právě vypočteno. Za interpolační body vezměme body $x_p, x_{p-1}, \dots, x_{p-q}$. Uvažovaný interpolační polynom má tvar

$$P(x) = \sum_{m=0}^q (-1)^m \binom{-s}{m} \nabla^m f_p, \quad s = \frac{x - x_p}{h}.$$

Zvolme nyní $x = x_p$, $\tau = h$ a aproximujme $y(x) \approx y_p$, $y(x+\tau) \approx y_{p+1}$, $f(t, y(t)) \approx P(t)$. Dostaneme formuli

$$y_{p+1} - y_p = \int_{x_p}^{x_{p+1}} P(t) dt = h \sum_{m=0}^q \gamma_m \nabla^m f_m,$$

kde

$$\gamma_m = (-1)^m \frac{1}{h} \int_{x_p}^{x_{p+1}} \binom{-s}{m} dx = (-1)^m \int_0^1 \binom{-s}{m} ds.$$

α) Takto získaným metodám (pro různá q) se říká *Adams-Bashforthovy metody*. Pro zvolené q dostaneme $(q+1)$ -krokovou metodu: k výpočtu y_{p+1} používáme hodnoty f v x_{p-q}, \dots, x_p ; jedná se tedy o metodu explicitní. Charakteristický polynom $\rho(\xi) = \xi^{q+1} - \xi^q = \xi^q(\xi - 1)$ má q -násobný kořen 0 a jednoduchý kořen 1, metody jsou tedy silně stabilní. Většinou mají tyto metody řád $q+1$.

β) Položme $x = x_{p-1}$, $\tau = h$. Dosaďme do (*):

$$\begin{aligned} y_p - y_{p-1} &= \int_{x_{p-1}}^{x_p} P(x) dx = \int_{x_{p-1}}^{x_p} \sum_{m=0}^p (-1)^m \binom{-s}{m} \nabla^m f_p dx = \\ &= h \sum_{m=0}^q \gamma_m^* \nabla^m f_p, \end{aligned}$$

$$\text{kde } \gamma_m^* = \frac{(-1)^m}{h} \int_{x_{p-1}}^{x_p} x_p \binom{-s(x)}{m} dx = (-1)^m \int_{-1}^0 \binom{-s}{m} ds.$$

V tomto případě dostáváme implicitní q -krokovou metodu. Tyto metody se nazývají *Adams-Moultonovy*.

Hodnoty zpětných diferencí se počítají poměrně snadno, ale přesto můžeme naše metody přepsat do tvaru, kdy místo hodnot $\nabla^m f_p$ používáme přímo hodnoty f_{p-r} podle vztahu

$$\nabla^m f_p = \sum_{\rho=0}^m \binom{m}{\rho} f_{p-\rho}.$$

Dostaneme pak např. Adams-Bashforthovu metodu ve tvaru

$$\begin{aligned} y_{p+1} - y_p &= \sum_{m=0}^q \gamma_m \nabla^m f_p = h \sum_{m=0}^q \gamma_m \sum_{\rho=0}^m (-1)^\rho \binom{m}{\rho} f_{p-\rho} = \\ &= h \sum_{\rho_0}^q f_{p-\rho} \left((-1)^\rho \sum_{m=\rho}^q \binom{m}{\rho} \gamma_m \right). \end{aligned}$$

Příklady některých metod

V následující tabulce jsou uvedeny příklady některých metod, které takto dostaneme.

	metoda	k	p
1.	$y_{n+1} = y_n + hf_n$	1	1
2.	$y_{n+2} = y_{n+1} + \frac{1}{2}h(3f_{n+1} - f_n)$	2	2
3.	$y_{n+3} = y_{n+2} + \frac{h}{12}(23f_{n+2} - 16f_{n+1} - 5f_n)$	3	3
6.	$y_{n+1} = y_n + hf_{n+1}$	1	1
7.	$y_{n+1} = y_n + \frac{h}{2}(f_{n+1} - f_n)$	1	2
8.	$y_{n+2} = y_{n+1} + \frac{h}{12}(5f_{n+2} + 8f_{n+1} - f_n)$	2	3

Metody prediktor-korektor

Budeme-li chtít využít uvedené metody pro použití v metodách prediktor-korektor, je výhodné vzít kombinace 1-7 a 2-8. Z teoretického rozboru totiž vyplývá, že v případě, kdy jako prediktor použijeme metodu řádu o jedna menšího než je řád korektoru, celková metoda bude mít stejný řád jako korektor. Dostaneme tedy jednokrokovou metodu řádu 2 a dvoukrokovou metodu řádu 3.

NUMERICKÉ METODY OPTIMALIZACE

Příklad 11 Prokládání křivek naměřenými daty: máme zadány hodnoty x_i, y_i ($i = 1, \dots, n, x_1 < \dots < x_n$). Hledáme funkci závislou na konečně mnoha parametrech $\alpha_1, \dots, \alpha_r$ tak, aby její hodnoty v x_i byly co nejbližší y_i . Máme tedy funkci $f = f(\alpha_1, \dots, \alpha_r, x)$ a chceme minimalizovat nějaký funkcionál $J(\alpha_1, \dots, \alpha_r)$, například při metodě nejmenších čtverců

$$J(\alpha_1, \dots, \alpha_r) = \sum_{i=1}^n (f(\alpha_1, \dots, \alpha_r) - y_i)^2.$$

Příklad 12 *Optimalizace tvaru radiátoru ústředního topení*

Snažíme se najít co nejvýhodnější tvar průřezu žebra. Tento tvar můžeme reprezentovat uzavřenou křivkou okolo počátku; protože chceme, aby tato křivka byla symetrická, můžeme ji reprezentovat pomocí nezáporné funkce $p : [0, b] \rightarrow [0, +\infty)$. Naše křivka pak vznikne doplněním grafu této funkce o části souměrné s ním sružené podle počátku a obou os. Rozložení teploty T bude splňovat rovnici pro vedení tepla ve tvaru³

$$\frac{d}{dx} \left(p(x) \frac{dT}{dx} \right) = a \sqrt{1 + \left(\frac{dp}{dx} \right)^2} \cdot T$$

s okrajovými podmínkami $T(0) = T_0, T(b) = T_b$, kde a je konstanta závisající na vlastnostech materiálu. Celkové množství tepla uvolněného do okolí je

$$I = 2 \int_0^b kT dx,$$

kde k je opět jakýsi koeficient. Úkolem je najít funkci p takovou, aby hodnota funkcionálu $I(p)$ byla maximální, tedy aby hodnota $F(p) = -I(p)$ byla minimální. Přesněji: hledáme funkci $p^* \in C^1[0, p]$ tak, aby

$$F(p^*) = \min_{p \in C^1[0, p]} F(p)$$

a zároveň $p(b) = 0, p(0) = K > 0$ (K je zadáno) a $p > 0$ na $[0, b)$. Označíme-li U množinu všech takových funkcí (bez požadavku minimality), hledáme prostě

³tento příklad má pouze ilustrovat, jak se složitější praktický problém převádí na standardní situaci hledání minima funkce několika proměnných

minimum F na U . Takto formulovaná úloha spadá do odvětví matematiky zvaného *variační počet*. Pro použití numerických metod se musíme ještě omezit na nějakou množinu $\tilde{U} \subseteq U$ funkcí určenou konečně mnoha parametry u_1, \dots, u_n . Každé takové n -tici parametrů přiřadíme funkci $p \in U$, k té spočítáme teplotu T , hodnoty $I(p)$ a $F(p)$ a položíme $\tilde{F}(u_1, \dots, u_n) := F(p)$. Tím už jsme úlohu převedli na problém minimalizace funkce $\tilde{F} : \mathbb{R}^n \rightarrow \mathbb{R}$.

V obou případech jsme dospěli k úloze minimalizovat reálnou funkci n proměnných. Je vidět, že tato úloha má spoustu praktických aplikací, a proto byly vyvinuty různé algoritmy, které je více či méně úspěšně řeší.

3.1 Základy konvexní analýzy

V této části uvedeme některé základní pojmy a výsledky konvexní analýzy, které se využívají při numerickém hledání extrémů. V dalším bude U podmnožina \mathbb{R}^n a J zobrazení U do \mathbb{R} .

Definice 15 Hodnotu $J(\bar{u})$, $\bar{u} \in U$ nazveme (absolutním, globálním) *minimem* funkce J na U , jestliže $\forall u \in U : J(u) \geq J(\bar{u})$. Bod \bar{u} nazýváme *bod minima*, píšeme $J(\bar{u}) = \min_U J$ nebo $\bar{u} = \arg \min_U J$.

Hodnotu $J(\bar{u})$, $\bar{u} \in U$ nazveme *lokálním minimem* funkce J , jestliže existuje okolí $V(\bar{u})$ takové, že $J(\bar{u})$ je (globálním) minimem na J na $U \cap V(\bar{u})$. Bod \bar{u} se pak nazývá *bodem lokálního minima*.

Věta 14 *Bud' $U \subseteq \mathbb{R}^n$ omezená, uzavřená množina a $J : U \rightarrow \mathbb{R}$ spojitá funkce. Potom J nabývá na U minima.*

Definice 16 Nechť $U \subseteq \mathbb{R}^n$ je neomezená a $J : U \rightarrow \mathbb{R}$. Říkáme, že funkce J je *koercivní*, jestliže

$$\lim_{\substack{\|u\| \rightarrow \infty \\ u \in U}} J(u) = +\infty.$$

Věta 15 *Nechť $U \subseteq \mathbb{R}^n$ je neomezená, uzavřená. Bud' $J : U \rightarrow \mathbb{R}$ spojitá a koercivní. Pak J nabývá na U minima.*

Důkaz: Z podmínky koercivity vyplývá, že existuje takové $R > 0$, že

$$\forall u \in U - B(0, R) : J(u) \geq J(0) + \frac{\pi}{17},$$

kde $B(0, R)$ označuje uzavřenou kouli se středem v počátku a poloměrem R . Označíme-li $K = U \cap B(0, R)$, nabývá J na K podle minulé věty v nějakém bodě \bar{u} minima. Pak je ale pro $u \in U - B(0, R)$

$$J(u) \geq J(0) + \frac{\pi}{17} > J(0) \geq J(\bar{u}),$$

hodnota $J(\bar{u})$ je tedy minimem na celé U .

Označení: Nechť $U \subseteq \mathbb{R}^n$ je otevřená, $J \in C^1(U)$. Pak pro každé $u \in U$, $\varphi \in \mathbb{R}^n$ existuje směrová derivace

$$J'(u, \varphi) = \lim_{\theta \rightarrow 0} \frac{J(u + \theta\varphi) - J(u)}{\theta}.$$

Její hodnota je rovna $\nabla J(u) \cdot \varphi$, kde

$$\nabla J(u) = \text{grad } J(u) = \left(\frac{\partial J}{\partial u_1}(u), \dots, \frac{\partial J}{\partial u_n}(u) \right)^T$$

je *gradient* funkce J .

Definice 17 Buď $U \subset \mathbb{R}^n$ konvexní množina. Řekneme, že $J : U \rightarrow \mathbb{R}$ je *konvexní funkce*, jestliže

$$J(u + \theta(v - u)) \leq J(u) + \theta(J(v) - J(u))$$

pro libovolné $u, v \in U$, $\theta \in [0, 1]$. Řekneme, že J je *ryze konvexní*, jestliže pro $u \neq v$, $\theta \in (0, 1)$ platí ostrá nerovnost.

Věta 16 $U \subset \mathbb{R}^n$ buď otevřená konvexní množina, $J \in C^1(U)$. Pak

- (i) J je konvexní právě tehdy, když $J(v) \geq J(u) + J'(u, v - u)$ pro každé $u, v \in U$;
- (ii) J je ryze konvexní právě tehdy, když $J(v) > J(u) + J'(u, v - u)$ pro každé $u, v \in U$, $u \neq v$.

Důkaz: Nechť je nejprve J konvexní na U . Tzn., že je-li $u, v \in U$, potom

$$J(u + \theta(v - u)) \leq J(u) + \theta(J(v) - J(u)), \quad \forall \theta \in (0, 1).$$

Z toho plyne

$$J(v) - J(u) \geq \frac{J(u + \theta(v - u)) - J(u)}{\theta}, \quad \forall \theta \in (0, 1).$$

Nyní uvažujme $\theta \rightarrow 0+$, pak

$$J(v) - J(u) \geq \lim_{\theta \rightarrow 0+} \frac{J(u + \theta(v - u)) - J(u)}{\theta} = J'(u, v - u).$$

Opačná implikace: nechť je pro každé $u, v \in U$ splněna podmínka

$$J(v) \geq J(u) + J'(u, v - u).$$

Z toho plyne

$$\begin{aligned} J(u) &\geq J(u + \theta(v - u)) - J'(u + \theta(v - u), \theta(v - u)) = \\ &= J(u + \theta(v - u)) - \theta J'(u + \theta(v - u), v - u). \end{aligned}$$

Obdobně

$$\begin{aligned} J(v) &\geq J(u + \theta(v - u)) + J'(u + \theta(v - u), (1 - \theta)(v - u)) = \\ &= J(u + \theta(v - u)) + (1 - \theta)J'(u + \theta(v - u), v - u). \end{aligned}$$

Přenasobením prvního vztahu členem $(1 - \theta)$ a druhého θ a sečtením dostaneme

$$(1 - \theta)J(u) + \theta J(v) \geq J(u + \theta(v - u)) + 0, \quad \forall u, v \in U, \forall \theta \in [0, 1].$$

Nyní budeme dokazovat druhé tvrzení. Nejprve nechť je

$$J(v) > J(u) + J'(u, v - u), \quad \forall u, v \in U, u \neq v, \theta \in [0, 1].$$

Stejným postupem jako jsme právě užili (uvažujeme-li ostré nerovnosti) dostaneme, že f je ryze konvexní.

Předpokládejme naopak, že f je ryze konvexní. Z definice pro $\theta \in (0, 1)$ vyplývá

$$J(u + \theta(v - u)) < J(u) + \theta(J(v) - J(u))$$

a odtud ihned plyne

$$\begin{aligned} J(v) - J(u) &> \frac{J(u + \theta(v - u)) - J(u)}{\theta} \geq \\ &\geq \frac{J'(u, \theta(v - u))}{\theta} = J'(u, v - u). \end{aligned}$$

Věta 17 $U \subset \mathbb{R}^n$ buď otevřená, $J \in C^1(U)$.

- (i) Je-li $\bar{u} \in U$ bodem lokálního minima, je $J'(u, \varphi) = 0$ pro každé $\varphi \in \mathbb{R}^n$ (neboli $\nabla J(\bar{u}) = 0$).
- (ii) Je-li navíc U konvexní, J konvexní, potom jsou následující tvrzení ekvivalentní:
 - 1) \bar{u} je bodem lokálního minima,
 - 2) \bar{u} je bodem minima,
 - 3) $\nabla J(\bar{u}) = 0$.
- (iii) Je-li navíc J ryze konvexní, má J na U nejvýše jeden bod minima.

Důkaz: První tvrzení je známé z matematické analýzy. Pro důkaz druhého předpokládejme $J \in C^1(U)$, J konvexní. Z toho vyplývá

$$J(v) \geq J(\bar{u}) + \underbrace{J'(\bar{u}, v - \bar{u})}_{\forall v \in U=0}$$

$$J(v) \geq J(\bar{u}).$$

Jestliže \bar{u} je bod lokálního minima, potom $J'(\bar{u}, \varphi) = 0$ pro každé $\varphi \in \mathbb{R}^n$, odtud

$$\bar{u} = \arg \min_U J.$$

Pro důkaz třetího tvrzení necht' J je ryze konvexní a předpokládejme, že má $u_1 \neq u_2$ dva body minima v U . Potom ale z konvexity máme

$$J(u_2) > J(u_1) + J'(u_1, u_2 - u_1),$$

příčemž $J'(u_1, u_2 - u_1) = 0$, protože u_1 je bod minima. To je ovšem spor. \square

Necht' $J \in C^2(U)$, $U \subset \mathbb{R}^n$. Pak má J totální diferenciál druhého řádu v libovolném bodě $u \in U$ a směrové derivace druhého řádu $J''(u, \varphi, \psi) = \varphi^T J''(u)\psi$, kde

$$J''(u) = \left(\frac{\partial^2 J}{\partial u_i \partial u_j}(u) \right)_{i,j=1}^n$$

je Hessova matice (symetrická). Taylorův vzorec lze psát ve tvaru

$$J(v) = J(u) + J'(u, v - u) + \frac{1}{2} J''(u + \theta(v - u), v - u, v - u)$$

pro nějaké $\theta \in [0, 1]$. Použitím tohoto vzorce lze dokázat tuto větu:

Věta 18 *Necht' $J : \mathbb{R}^n \rightarrow \mathbb{R}$, $J \in C^2$, necht' $\exists \alpha > 0$ tak, že $J''(u, \varphi, \varphi) \geq \alpha \|\varphi\|^2$ pro každé $u, \varphi \in \mathbb{R}^n$. Potom J je koercivní a ryze konvexní na \mathbb{R}^n .*

Důkaz: Necht' J je dvakrát spojitě diferencovatelná. Nejprve dokážeme koercivitu. Z Taylorova vzorce máme pro $\theta \in [0, 1]$

$$\begin{aligned} J(u) &= J(0) + J'(0, u) + \frac{1}{2} J''(\theta u, u, u) \geq \\ &\geq J(0) - M \|u\| + \frac{\alpha}{2} \|u\|^2 \xrightarrow{\|u\| \rightarrow \infty} +\infty. \end{aligned}$$

Jest totiž podle Cauchyovy nerovnosti

$$|J'(0, u)| = |\nabla J(0) \cdot u| \leq \underbrace{\|\nabla J(0)\|}_{=M} \|u\|$$

a podle předpokladů je

$$\frac{1}{2} J''(\theta u, u, u) \geq \frac{\alpha}{2} \|u\|^2.$$

Tím jsme dokázali koercivitu, zbývá dokázat ryzí konvexitu. Opět uijeme Taylorova vzorce

$$\begin{aligned} J(v) &= J(u) + J'(u, v - u) + \frac{1}{2} J''(u + \theta(v - u), v - u, v - u) \geq \\ &\geq J(u) + J'(u, v - u) + \frac{1}{2} \alpha \|v - u\|^2 > \\ &> J(u) + J'(u, v - u), \end{aligned}$$

čímž je důkaz hotov. \square

3.2 Numerické metody hledání minima

V této části vysvětlíme jeden z numerických minimalizačních algoritmů - *iterační proces*. Nechť $J : \mathbb{R}^n \rightarrow \mathbb{R}$ a je dán počáteční bod $u^0 \in \mathbb{R}^n$. Nechť jsme již určili $u^k \in \mathbb{R}^n$. Pak

$$u^{k+1} = u^k + \rho_k \varphi^k,$$

kde $k \geq 0$, $\varphi^k \in \mathbb{R}^n$, $\rho_k > 0$. Veličinu φ nazýváme *směr spádu*, ρ_k určuje vzdálenost - délku kroku ve směru φ^k . Pro praktické použití je potřeba zodpovědět dvě základní otázky:

- 1) jak volit směr spádu φ_k ,
- 2) jak volit velikosti kroku.

Směr spádu φ^k volíme tak, aby $J(u^{k+1}) < J(u^k)$. Pokud ρ_k zvolíme vhodně (např. $\rho_k > 0$ dost malé), obvykle φ^k volíme tak, že

$$J'(u^k, \varphi^k) = \nabla J(u^k) \cdot \varphi^k < 0. \quad (+)$$

Věta 19 *Nechť je $J \in C^2(\mathbb{R}^n)$ a platí (+). Potom existuje $\tilde{\rho}_0 > 0$ takové, že $J(u^{k+1}) < J(u^k)$, pokud $\rho^k \in (0, \tilde{\rho}_0)$.*

Důkaz:

$$J(u^{k+1}) = J(u^k) + \rho_k J'(u^k, \varphi^k) + \frac{1}{2} \rho_k^2 J''(\beta, \varphi^k, \varphi^k).$$

Nechť $K > 0$. Z toho plyne

$$|J''(\beta, \varphi^k, \varphi^k)| \leq M, \quad \forall \beta u^k + \rho \varphi^k, \varphi \in (0, K].$$

Odtud

$$J(u^{k+1}) = J(u^k) + \rho_k \left(\underbrace{J'(u^k, \varphi^k)}_{<0} + \frac{1}{2} \rho_k \underbrace{J''(\beta, \varphi^k, \varphi^k)}_{|J''| \leq M} \right).$$

Odtud plyne, že existuje $\tilde{\rho}_0 > 0$ tak, že

$$J(u^k, \varphi^k) + \frac{1}{2} \rho_k J''(\beta, \varphi^k, \varphi^k) < 0, \quad \forall \rho_k \in (0, \tilde{\rho}_0).$$

Příklad 13 Směr spádu $\varphi^k =$ směr největšího spádu.

$$\varphi^k = -\nabla J(u^k).$$

Pokud $\nabla J(u^k) \neq 0$, potom

$$J'(u^k, \varphi^k) = -\nabla J(u^k) \cdot \nabla J(u^k) = -\|\nabla J(u^k)\|^2 < 0.$$

Algoritmus I - metoda největšího spádu s konstantním krokem

Nechť je dáno: $u^0 \in \mathbb{R}^n$, $M \geq 1$ celé, $\varepsilon > 0$, $\rho > 0$ (krok). Pro $k = 0, 1, \dots, M$ proveďte:

- 1) $u^{k+1} = u^k - \rho \nabla J(u^k)$
- 2) Jestliže $\|\nabla J(u^k)\| < \varepsilon$, jdi na 3.
- 3) Stop.

Příště ukážeme, že metoda je konvergentní, tj. pro $M \rightarrow \infty$ je $\varepsilon = 0$.

Algoritmus II - metoda největšího spádu s optimálním krokem

Nechť je dáno: $u_0 \in \mathbb{R}^n$, $M \geq 1$ celé, $\varepsilon > 0$. Pro $k = 0, 1, \dots, M$ proveď:

- 1) Vypočti $\varphi^k = -\nabla J(u^k)$.
- 2) Vypočti $\rho_k = \arg \min_{\rho > 0} J(u^k + \rho \varphi^k)$
- 3) $u^{k+1} = u^k + \rho_k \cdot \varphi^k$
- 4) Je-li $\|\nabla J(u^k)\| < \varepsilon$, pak jdi na 5.
- 5) Stop.

Důkazy konvergence metod – viz přednášky!

VÝSLEDKY CVIČENÍ

Cvičení 1 Maximálně $\frac{1}{10!} \doteq 2,76 \cdot 10^{-7}$, není ovšem těžké dokázat (z původního vzorce), že pro ekvidistantní dělení to bude nejvýše $\frac{1}{10!} 2^{-10} \doteq 2,69 \cdot 10^{-10}$.

Cvičení 2 Buď A matice typu $n \times n$. Stačí ukázat, že $\ker A = 0$. Nechť existuje vektor $x \in \mathbb{R}^n$ takový, že $Ax = 0$, $x \neq 0$. Některá složka x je v absolutní hodnotě nejvyšší, nechť je to x_k . Vynásobením vhodnou konstantou dostaneme \tilde{x} takový, že $\tilde{x}_k = 1$, $A\tilde{x} = 0$. Potom k -tá složka $A\tilde{x}$ bude

$$\sum_{j=1}^n A_{kj} \tilde{x}_j \geq A_{kk} - \sum_{\substack{j=1 \\ j \neq k}}^n A_{kj} \tilde{x}_j \geq A_{kk} - \sum_{\substack{j=1 \\ j \neq k}}^n A_{kj} > 0,$$

což je spor.

Cvičení 3 Za chvíli.

Cvičení 4 Za chvíli.

Cvičení 5 Později.

Cvičení 6 To bude nějaká jednoduchá indukce.

Cvičení 7 Přenecháme povolanejším (asi Mathematica).

Cvičení 8 Snadno nahlédneme, že pokud $y_0 = y_1 = y_2 = 0$, jsou i všechny další členy rovny nule, řešení je tedy právě jedno. Předpokládejme tedy, že mezi prvními třemi členy je aspoň jeden nenulový. Čtvrtá moncina závorky je vždy nezáporná, máme tedy v každém kroku aspoň jednu možnost pokračování. Je-li závorka nulová, je i další člen nulový, je-li nenulová, můžeme pokračovat dvěma způsoby (a další člen je nenulový). Tyto dvě možnosti pro další člen se ovšem liší jen znaménkem, neovlivní tedy hodnoty dalších členů (v rekurentním vztahu je všechno v druhé mocnině). Počet řešení je tedy roven 2^a , kde $a = |n \geq 3 : y_n \neq 0|$ pro libovolné řešení $(y_n)_{n=0}^\infty$. Pokud je $a = \omega$, je počet řešení 2^ω (mohutnost kontinua). Pokud ne, musí být v posloupnosti jen konečně mnoho nenulových členů. Pak také existuje n tak, že $y_{n+1} = y_{n+2} = y_{n+3} = 0$. Z rovnice pak snadno dostaneme, že i $y_n = 0$; indukcí dále, že i všechny předchozí členy jsou nulové, tedy i y_0, y_1, y_2 . Pro nulovou počáteční podmínku tedy existuje právě jedno řešení, pro nenulovou 2^ω .

Cvičení 9 Jednoduchým výpočtem (viz definice) ověříme, že řád je 2.

REJSTŘÍK

- bod
 - lokálního minima, 39
 - minima, 39
- chyba
 - akumulovaná diskretizační, 8
 - odhad asymptotický, 17
 - akumulovaná diskretizační, 10, 31
 - diskretizační, 19
 - lokální diskretizační, 10, 30
 - lokální relativní diskretizační, 10
 - metody, 8
 - zaokrouhlovací, 19
 - zaokrouhlovací akumulovaná, 18
 - zaokrouhlovací lokální, 18
- chyby
 - zaokrouhlovací, 18
- derivace
 - směrová, 40
- diference
 - zpětné, 34
- formule
 - Gillova, 16
 - korektorová, 25
 - prediktorová, 25
 - standardní, 15
 - tříosminová, 15
- funkce
 - koercivní, 39
 - konvexní, 40
 - Lipschitzova, 9
 - lipschitzovská, 8
 - přírůstková, 8, 9
 - ryze konvexní, 40
- gradient, 40
- interpolace, 1
- matice, 3
 - blokově třídiagonální, 29
 - Hessova, 42
 - ostře diagonálně dominantní, 3, 4
 - třídiagonální, 3
- metoda
 - obecná k -kroková, 24
 - dvoukroková, 8
 - Eulerova, 7, 8, 16
 - explicitní, 24
 - implicitní, 24
 - konsistentní, 10, 26
 - konvergentní, 9, 26
 - nejmenších čtverců, 38
 - největšího spádu
 - s konst. krokem, 44
 - s optimálním krokem, 44
 - polovičního kroku, 17
 - prediktor-korektor, 25
 - Runge-Kuttova druhého řádu, 7, 8, 13, 25
 - silně stabilní, 31
 - stabilní, 26
 - sítí, 29
 - víceková, silně stabilní, 31
- metody
 - Adams-Bashforthovy, 36
 - Adams-Moultonovy, 36
 - adaptivní, 17
 - diskrétní, 7
 - jednokrokové, 8
 - optimalizace, 38
 - prediktor-korektor, 37
 - Runge-Kuttovy, 12
 - Runge-Kuttovy druhého řádu, 14
 - Runge-Kuttovy třetího řádu, 15
 - Runge-Kuttovy čtvrtého řádu, 15
 - vícekové, 24
- metody, Runge-Kuttovy, 13
- minimum
 - absolutní, globální, 39
 - lokální, 39
- moment splinu, 2
- operátor
 - diferenciální D , 12
 - Laplaceův, 29
- parametr, 5
 - napěťový, 5
- pivotace, 18
- polynom, 1
 - charakteristický, 21, 23
 - Lagrangeův interpolační, 1
- princip
 - Duhamelův, 23
- proces
 - iterační, 43
- přírůstek

- přesný relativní, 10
- rovnice, 6
 - Fuchsova typu, 7
 - obyčejné diferenciální, 6
 - charakteristická, 21
 - prvního řádu, 6
 - vyššího řádu, 6
- směr spádu, 43
- soustava
 - diferenčních rovnic, 19
 - lineárních diferenciálních rovnic, 19
 - homogenní, 19
- splajn, kubický interpolační, 4
- spline, 1
 - Hermiteův, 5
 - kubický, 1
 - kubický interpolační, 2, 4
 - přirozený, 2
 - s napětím, 5
- spline, přirozený, 5
- systém
 - reálný fundamentální, 23
 - fundamentální, 20, 23
- Taylorův vzorec, 13
- variační počet, 39
- věta
 - Banachova o kontrakci, 25
 - Picardova, 8
- řád
 - metody, 26
 - soustavy, 20

LITERATURA

1. Feistauer M.: *Diskrétní metody řešení diferenciálních rovnic*, 1981.
 2. Příkryl P.: *Numerické metody matematické analýzy*, (MVŠT 24), SNTL.
 3. Vitásek E., Prager E. : *Numerická matematika I., II.*
 4. Ralston, A.: *Základy numerické matematiky.*
- [H]. Henrici: *Discrete Variable Methods for ODE's.*